Chapter 1

Introduction to Geodesy

According to the Geodetic Glossary (NGS 2009), **geodesy** can be defined as “The science concerned with determining the size and shape of the Earth” or “The science that locates positions on the Earth and determines the Earth’s gravity field.” It might be hard to see anything too interesting or exciting in these definitions. After all, generally speaking, the Earth is round, a fact known at least since the time of the ancient Greeks. The earliest known definitive work on this was done by Aristarchus of Samos (ca. 310-230 B.C.E.), an astronomer who worked out a geometric method to determine ratios of the distance between the Sun and the Earth with the distance between the Moon and the Earth, a method that required knowing that all three bodies were round (Maor 1998, p. 63). But what is not obvious from the definition of geodesy is that much modern technology is possible only due to geodesy. For example, without modern geodesy, navigation might still be done with ancient methods (e.g., don’t lose sight of the shore) and high-accuracy maps of moderate or larger regions could not exist. Perhaps surprisingly, geodesy plays a central role in a wide variety of cutting-edge sciences, such as geophysics, astronomy, and climatology. The Earth’s warming due to climate change is expected to produce rising sea levels in many places, and sea level rise cannot be observed and monitored without geodesy. The observable change in sea level is due to a combination of a change in the ocean’s mean surface level and any subsidence of the shore. Geodesy is necessary to tease these two apart. Global warming is also generally reducing the size of ice sheets and glaciers as well as dramatically increasing how quickly they move. Space-age positioning methods make it possible to determine not only the location of an ice sheet, but also its velocity, shape, and size. Geodesy is also a key component of geophysics by making it possible to observe tectonic plate motion, for example.

Geodesy is usually subdivided into geometrical geodesy, physical geodesy, and satellite geodesy, although additional subdivisions are recognized as well. **Geometrical geodesy** is concerned with describing locations in terms of geometry. Consequently, coordinate systems are one of the primary products of geometrical geodesy. **Physical geodesy** is concerned with determining the Earth’s gravity field, which is necessary for establishing heights. **Satellite geodesy** is concerned with using orbiting satellites to obtain data for geodetic purposes.

To understand the role of geodesy, it might be useful to look at how maps are often drawn. A **map** is an abstract, scaled, two-dimensional image of some region typically on the Earth’s surface, showing the features of interest in their correct relative locations, sizes, and orientations. According to this definition, a globe is not a map because it is not two-dimensional. Neither are directions to the neighborhood gas station drawn on a napkin for a friend. Although entirely adequate for its purpose, such a rendering is a sketch, not a map, because, while schematically correct, it is not an accurate rendition of the features. Maps are drawn essentially the same way whether they are drawn on paper or on a computer screen. The process begins with someone determining positions for all the features of interest. A **position** is (usually) a pair or triplet of numbers, called **coordinates**. A **location** is a place in the real world, whose spatial placement is described by a position. A **coordinate system** specifies how coordinates are assigned to locations.

In order to draw, say, a building on a map, we need to determine the positions of its defining points, like its corners. Linear coordinates, such as the familiar x, y, z, are descriptions of offsets from some point of reference, called an **origin**, each in the direction of its axis. Suppose one corner of the building was found to be 150.25 meters (m) east, 95.51 m north, and 10.59 m above the origin. The coordinates are $x = 150.25$, $y = 95.51$ and $z = 10.59$, and the position is written $(150.25, 95.51, 10.59)$. After the positions have been determined, appropriately scaled symbols of the locations can be drawn on the paper to produce a map (see section 3.5.1 on page 28). The building's edges can then be drawn by "connecting the dots." This example used a **Cartesian coordinate system**. Cartesian coordinate systems have an origin at the intersection of several straight lines, the axes, that are mutually perpendicular.

Linear coordinates are distances as opposed to locations on a number line. Some might ask, "Don't numbers on a number line always represent distances?" The answer is "not necessarily." In a Cartesian coordinate system they represent distances, but in other types of coordinate systems they do not. To illustrate this point, think about street addresses. Street addresses are really just symbols, same as numbers on a number line. They don't *necessarily* mean anything spatial. In Tokyo, street addresses were assigned sequentially in time, not in space, so an address in Tokyo indicates *when* the building was built but not *where* it is located! In contrast, a Cartesian coordinate represents a linear distance from an origin in a specified direction. Geodesy uses many types of coordinate systems, including ones whose coordinates are *angles*, such as latitude and longitude, rather than linear distances. However, regardless of whether a coordinate is an angle or a linear distance, it is always a separation from an origin in a particular direction.

But where do coordinates come from? What coordinate system are they in? Is there more than one coordinate system of a particular type? In the above example with the building, we assumed that it was possible to determine in what direction the building's corner lay from the origin. But how is that possible? Who decides in what direction is north and, once that has been decided, how do we determine other directions for ourselves? By now you may have guessed that geodesy provides the answers to these questions. Some of the primary products of geometrical geodesy are coordinate systems within which realistic geospatial coordinates can be derived.

1.1 Everyday Geodesy

Until recently, most people had not even heard of geodesy and probably even fewer cared. Up to the end of the 20th century, geodesists worked more-or-less quietly behind the scenes, producing the spatial frameworks used by surveyors and cartographers, enabling them to do things like delineate property, design roads, and produce road maps. Things went along as they always had until two technologies changed geodesy's background role: geographic information systems (GIS) and global navigation satellite systems (GNSS). The U.S. NAVSTAR Global Positioning System (GPS) is the latest U.S. GNSS.

Before GNSS, high-accuracy determinations of latitude and longitude were very hard to come by;

Figure 1.1: A total station is a theodolite and an electromagnetic distance measurement (EDM) instrument packaged together.

only trained mapping professionals and geodesists knew how to do it, and it was a time-consuming process. GNSS removed this obstacle. GNSS has made it possible for anyone to determine the latitude and longitude of (almost) anywhere on Earth, typically in a matter of seconds, simply by turning on an inexpensive instrument the size of a pocket calculator. This single fact constituted a cartographic revolution and was utterly unimaginable before the advent of the space age, radio communications, the high-speed digital computer, and the atomic clock.

Although technologically marvelous, the ability to rapidly determine latitude and longitude coordinates, in itself, is probably not much more than a novelty to most people. After all, few people need or want to know the latitude and longitude of some place they are trying to navigate to; latitude and longitude are usually not helpful to most people trying to get somewhere. Car navigation usually depends on road maps or written descriptions; knowing the coordinates of the destination is practically useless to a typical driver. However, GIS changed this, too. If a car is equipped with a GNSS and that GNSS communicates with a computer with access to a database of digital maps all compiled in latitude and longitude, then search algorithms can instantly solve routing problems and tell the driver the best way to get to a destination.

Before GIS, if a map of a private property parcel (**plat**) was compiled in any coordinate system at all, it was usually compiled in a **local coordinate system** (see chapter 3). The coordinates of the origin and the orientation of local maps are arbitrary, so local coordinate system maps convey only relative information of their features, such as their separation and size, but not their location in a global sense. Such maps are readily compiled from measurements made using instruments such as total stations (see Fig. 1.1 on the previous page). A total station is a combination of two instruments, one that measures horizontal and vertical angles, called a **theodolite**, and an electronic distance measuring (EDM) instrument. Total stations cannot determine their own global location or orientation. Plats in local coordinate systems are entirely acceptable to landowners and tax assessors who use these maps individually. However, to see how, say, a city's parcels fit together, then local coordinate system maps must be compiled at the same scale, which is not always the case, and the plats would have to be manually fit together like a jigsaw puzzle. The coordinates of such maps are not useful in figuring out which parcels abut one another because it would be an extraordinary coincidence if two side-by-side local coordinate system maps happened to be in commensurate coordinate systems. GIS solved these problems. A GIS can display a map at almost any scale, rescaling as desired. A GIS will also automatically display independently compiled maps in their correct locations (up to the accuracy of the data) if they were compiled in the same coordinate system. The desirability of having all a city's parcels being available for display and analysis as a whole provided the economic impetus to create maps compiled in global coordinate systems, which are a product of geodesy.

Nowadays, whether someone is aware of it or not, geodesy is having an impact on our daily lives. Emergency responders are beginning to navigate using GNSS. Travelers renting cars in unfamiliar cities can navigate like natives thanks to the GNSS navigation system in rental cars. Many people use Web-based mapping services to produce maps with step-by-step directions from their door to their destination, on demand. Ships can sail faster in inclement weather and ride lower in the water without running aground thanks to GNSS and high-accuracy navigation channel surveys. Aircraft can operate in all visibility conditions, and pilots can land with confidence that they will touch down on the runway and not hit any obstacles during their descent. Cell phones can broadcast coordinates in an emergency. Geophysicists are determining the inner structure of the Earth using GNSS both to measure the motion of the tectonic plates and to, essentially, perform computed axial tomography (CAT) scans of its interior. GIS and GNSS are central to a broad range of technological applications, both mundane and exotic, and none of these applications would be possible without geodesy.

1.2 Why Learn about Geodesy?

The GIS revolution created a new generation of cartographers who often are expert computer users but not necessarily trained in the concepts underlying spatial data or the processes by which spatial data are created. This can lead to confusion and errors. Furthermore, GIS analyses always depend upon map projections of the latitude and longitude data from which they were compiled. All map projections distort one or more of length, area, shape and direction, so any analysis that ignores these distortions is less accurate than it could be because, with the knowledge of geodesy, these distortions can be computed and then either eliminated or reduced to negligible levels. Geomaticists need to understand that all geospatial data contain error, either from their measurement or from distortions introduced by geospatial manipulations.

There are many kinds of geospatial coordinate systems and many kinds of map projections. Geospatial data that refer to different coordinate systems cannot be mixed together in a meaningful way any more than temperatures in degrees Fahrenheit can be used with temperatures in degrees Celsius. Spatial data are routinely collected in many different coordinate systems, and using them together in a naïve way results in unrealistic maps. Knowledge of geodesy and cartography is necessary to solve this problem correctly. There are tools to transform the data so that they are all in the same coordinate system. With these tools and the proper knowledge, the geomaticist is free to mix and match data with an understanding of the compromises and trade-offs inherent in each decision made along the way. In extreme cases, ignorance of these distortions and other geodetic topics could lead to maps that could endanger public safety. It's entirely possible, for example, to depict a fire hydrant in the wrong place on a map by not using the proper units for its position. Such a mistake might hinder firefighters in locating the hydrant in an emergency. Gunther Greulich, past president of the American Congress on Surveying and Mapping, stated somewhat facetiously that, "Geography without geodesy is a felony," which underlines the importance of geodesy for all practitioners of geomatics.

1.2.1 The need for flatness

Cartographic products are flat, whether they are paper maps or images on a computer monitor. Maps are flat for many reasons, some rooted in convenience and others stemming from psychology. Humans tend to think of distances as being straight-line distances. We are comfortable with the idea of a horizontal distance and a vertical distance, and we know that they are measured with, say, a measuring tape pulled taught in a straight line. We tend to think of our environment as being flat, and we are very comfortable with (and in fact we prefer), depictions of space that conform to that perception.

The convenience of a flat model of space is clear: globes don't fit in the glove box of a car, and they would have to be very big to show important details of small regions. We usually want to determine distances by laying a ruler on the map and multiplying the measured distance by a (constant) map scale factor. In fact, for distances measured with a ruler on a map to be valid in all circumstances *requires* that the Earth actually be flat.

1.2.2 Round realities

Regardless of our perceptions, the Earth is round. Perhaps the most apparent contradiction of the flat-Earth hypothesis comes from watching ships as they approach land. If the Earth were flat, then ships would suddenly pop into view and be visible from their waterline to the top of their masts all at once. This does not happen. The tops of the masts come into view first with the hulls appearing last. Sailors describe ships as being "hull down" if their masts are visible but the hull is below the horizon. Of course, images of the Earth seen from space are pretty compelling, too.

For the purposes of making maps can't we just pretend that the Earth is flat? Could we, for example, just treat latitude and longitude coordinates, which are angles, as if they were Cartesian coordinates, which are linear distances? For that to work, geographic coordinates would need to have the same properties as Cartesian coordinates, but they do not. For example, every point in a plane has a unique pair of Cartesian coordinates (x, y). But what are the coordinates of the North Pole? Its latitude is 90°N, but what is its longitude? All meridians converge at the poles (Fig. 1.2), so the North Pole apparently is on *all* of the meridians at the same time, meaning any longitude from 0° to just less than 360° seems equally correct. Unlike Cartesian (x, y) coordinates on a plane, longitude and latitude for points on a sphere are not unique in some cases.

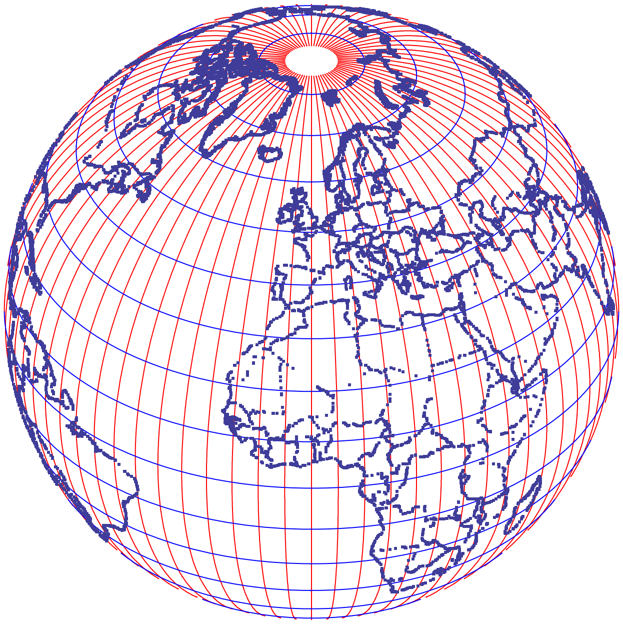

Figure 1.2: The orthographic map projection portrays the Earth as it would be seen from space. Meridians of longitude are shown converging towards the North Pole with parallels of latitude running perpendicular to the meridians (data courtesy of NGA).

A map is a planar depiction of a round reality and it's a mathematical fact that such a depiction cannot be free of distortion. Planes are fundamentally different than spheres. It is impossible to concoct some magical mathematical formula to convert one to the other without distorting their geometrical relationships in a significant way. It might be said that the role of geometric geodesy is to bring the world of the round and the world of the flat together.