

Nature of Geographic Information

Nature of Geographic Information

An Open Geospatial Textbook

David DiBiase

Except where otherwise noted, content on this site is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#).

Contents

About this Book	vi
Acknowledgements	vii
<i>David DiBiase</i>	
Chapter 1	
Data and Information	2
<i>David DiBiase</i>	
Chapter 2	
Scales and Transformations	27
<i>David DiBiase</i>	
Chapter 3	
Census Data and Thematic Maps	67
<i>David DiBiase</i>	
Chapter 4	
TIGER, Topology and Geocoding	88
<i>David DiBiase</i>	
Chapter 5	
Land Surveying and GPS	114
<i>David DiBiase</i>	
Chapter 6	
National Spatial Data Infrastructure I	146
<i>David DiBiase</i>	
Chapter 7	
National Spatial Data Infrastructure II	171
<i>David DiBiase</i>	

Chapter 8	
Remotely Sensed Image Data	203
<i>David DiBiase</i>	
Chapter 9	
Integrating Geographic Data	240
<i>David DiBiase</i>	
About the Author	253

About this Book

Nature of Geographic Information is an open geospatial textbook used at Pennsylvania State University and written by David DiBiase, with contributions by James L. Sloan II, Ryan Baxter, Wesley Stroh, Beth Fletcher King, and many students.

The [B.C. Open Textbook Project](#) has added this book to Pressbooks, an open source authoring platform, in order that files could be generated that allow this textbook to be read on various devices, and to make revision and remixing of this material easier through editable files such as Pressbooks and WordPress XML files, ODT, ePUB, and HTML.

Open textbooks are openly licensed using a [Creative Commons license](#), and are offered free of charge, or as printed books that are available at cost.

If you are an instructor who is using this book for a course, please fill out the [Adoption of an Open Textbook form](#).

Acknowledgements

David DiBiase

An Open Geospatial Textbook

David DiBiase with contributions by James L. Sloan, Ryan Baxter, Wesley Stroh, Beth Fletcher King, and many students

The Pennsylvania State University

The purpose of this text is to promote understanding of the Geographic Information Science and Technology enterprise (GIS&T, also known as “geospatial”). Since I began writing in 1997 it has been a vehicle for me to understand the field better, and to help my students do the same. Originally my students were undergraduates enrolled in the Penn State course GEOG 121 (now 160): Mapping Our Changing World. Later, I developed an online text for students in our [Postbaccalaureate Certificate Program in GIS and Master of GIS \(MGIS\) degree program](#), both offered to adult professionals across the country and around the world through the University’s “World Campus.” A short version that includes ArcGIS exercises appears in Esri’s Virtual Campus as “Understanding Geographic Data.” Now, with the blessings of both Penn State and Esri, I am pleased to share the text with students and teachers everywhere as part of an [Open Educational Resources initiative](#) of Penn State’s John A. Dutton e-Education Institute. You are welcome to use and re-use materials that appear in this text (other than those copyrighted by others) subject to the licensing agreement linked to the bottom of this and every page.

GIS&T is the intersection of professions, institutions, and technologies that produce geographic data and render information from it. It is a rapidly growing and evolving field. Learning is a way of life for all GIS&T professionals. With this in mind, I hope that this text may contribute to your lifelong exploration of how geospatial technologies can be used to improve the quality of life—yours and your neighbors’, locally and globally, now and in the future.

The title of the text is a tribute to the Department of Geography at the University of Wisconsin-Madison, where my exploration of GIS&T began in 1983. There, and through projects that kept us in touch after we had gone our separate ways, I was privileged to work with Barbara Bartz Petchenik and Arthur Robinson. In 1976 Barbara and Robbie co-authored *The Nature of Maps*. Long before that, Richard Hartshorne published the influential (and controversial) *The Nature of Geography* in 1939, a year before moving to Madison. It may be that neither of these works was altogether successful in revealing the nature of their subjects. Neither, perhaps, is this poorer attempt likely to succeed entirely. Still, I hope that its availability in this open format will be useful to those who can’t afford expensive (though sometimes very good) printed texts, and I imagine that Barbara and Robbie would have approved.

I wish to thank colleagues in the Dutton e-Education Institute who help me make this text, and all our online curricula, available. In collaboration with our friends at the [e-Learning Institute](#) of Penn State’s College of Arts and Architecture, Elizabeth Bailey and Martin Gutowski have been particularly instrumental in catapulting our professional practice into the present. I am grateful for the thousands of students who have traversed this text over the past fifteen years, and who continue to challenge me and my team to improve it.

This text serves as an assigned reading for registered students in formal Penn State courses. The Orientation section (under Start Here on the left) familiarizes registered students with the text and with the associated course management system (ANGEL). Links to Chapters 1-9 appear in the Contents menu. Only registered students are entitled to feedback from instructors and to earn academic credit for their participation in the course. For information about how to apply for admission to Penn State’s Postbaccalaureate Certificate and MGIS degree programs see the [Geospatial Education Program Office](#). Information about [Penn State’s Department of Geography](#) appears [here](#).



The National Science Foundation's Digital Libraries in the Classroom program supported portions of this work.

Chapter 1

Data and Information

David DiBiase

1.1. Overview

When I started writing this text in 1997, my office was across the street (and, fortunately, upwind) from Penn State's power plant. The energy used to heat and cool my office is still produced there by burning coal mined from nearby ridges. Combustion transforms the potential energy stored in the coal into electricity, which solves the problem of an office that would otherwise be too cold or too warm. Unfortunately, the solution itself causes another problem, namely emissions of carbon dioxide and other more noxious substances into the atmosphere. Cleaner means of generating electricity exist, of course, but they too involve transforming energy from one form to another. And cleaner methods cost more than most of us are willing or able to pay.

It seems to me that a coal-fired power plant is a pretty good analogy for a geographic information system. For that matter, GIS is comparable to any factory or machine that transforms a raw material into something more valuable. Data is grist for the GIS mill. GIS is like the machinery that transforms the data into the commodity—information—that is needed to solve problems or create opportunities. And the problems that the manufacturing process itself creates include uncertainties resulting from imperfections in the data, intentional or unintentional misuse of the machinery, and ethical issues related to what the information is used for, and who has access to it.

This text explores the nature of geographic information. To study the nature of something is to investigate its essential characteristics and qualities. To understand the nature of the energy produced in a coal-fired power plant, one should study the properties, morphology, and geographic distribution of coal. By the same reasoning I believe that a good approach to understanding the information produced by GIS is to investigate the properties of geographic data and the technologies and institutions that produce it.

Objectives

The goal of Chapter 1 is to situate GIS in a larger enterprise known as Geographic Information Science and Technology (GIS&T), and in what the U.S. Department of Labor calls the “geospatial industry.” In particular, students who successfully complete Chapter 1 should be able to:

1. Define a geographic information system;
2. Recognize and name basic database operations from verbal descriptions;
3. Recognize and name basic approaches to geographic representation from verbal descriptions;
4. Identify and explain at least three distinguishing properties of geographic data; and
5. Outline the kinds of questions that GIS can help answer.

1.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated

quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 1 Checklist (for registered students only)

Chapter 1 Checklist

Step	Activity	Access/Directions
1	Read Chapter 1	This is the second page of Chapter 1. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	Submit quizzes as you come across them in the chapter. Blue banners denote practice quizzes that are not graded. Red banners signal graded quizzes. (Note that Chapter 1 does not include a graded quiz.)	Go to ANGEL > [your course section] > Lessons tab > Chapter 1 folder > [quiz]
3	Perform “ Try This ” activities as you come across them in the chapter. “Try This” activities are not graded.	Instructions are provided for each activity.
4	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

1.3. Data

“After more than 30 years, we’re still confronted by the same major challenge that GIS professionals have always faced: You must have good data. And good data are expensive and difficult to create.” (Wilson, 2001, p. 54)

Data consist of symbols that represent measurements of phenomena. People create and study data as a means to help understand how natural and social systems work. Such systems can be hard to study because they’re made up of many interacting phenomena that are often difficult to observe directly, and because they tend to change over time. We attempt to make systems and phenomena easier to study by measuring their characteristics at certain times. Because it’s not practical to measure everything, everywhere, at all times, we measure selectively. How accurately data reflect the phenomena they represent depends on how, when, where, and what aspects of the phenomena were measured. All measurements, however, contain a certain amount of error.

Measurements of the locations and characteristics of phenomena can be represented with several different kinds of symbols. For example, pictures of the land surface, including photographs and maps, are made up of graphic symbols. Verbal descriptions of property boundaries are recorded on deeds using alphanumeric symbols. Locations determined by satellite positioning systems are reported as pairs of numbers called coordinates. As you probably know, all of these different types of data—pictures, words, and numbers—can be represented in computers in digital form. Obviously, digital data can be stored, transmitted, and processed much more efficiently than their physical counterparts that are printed on paper. These advantages set the stage for the development and widespread adoption of GIS.

1.4. Information

Information is data that has been selected or created in response to a question. For example, the location of a building or a route is data, until they are needed to dispatch an ambulance in response to an emergency. When

used to inform those who need to know “where is the emergency, and what’s the fastest route between here and there?,” the data are transformed into information. The transformation involves the ability to ask the right kind of question, and the ability to retrieve existing data—or to generate new data from the old—that help people answer the question. The more complex the question, and the more locations involved, the harder it becomes to produce timely information with paper maps alone.

Interestingly, the potential value of data is not necessarily lost when they are used. Data can be transformed into information again and again, provided that the data are kept up to date. Given the rapidly increasing accessibility of computers and communications networks in the U.S. and abroad, it’s not surprising that information has become a commodity, and that the ability to produce it has become a major growth industry.

1.5. Information Systems

Information systems are computer-based tools that help people transform data into information.

As you know, many of the problems and opportunities faced by government agencies, businesses, and other organizations are so complex, and involve so many locations, that the organizations need assistance in creating useful and timely information. That’s what information systems are for.

Allow me a fanciful example. Suppose that you’ve launched a new business that manufactures solar-powered lawn mowers. You’re planning a direct mail campaign to bring this revolutionary new product to the attention of prospective buyers. But since it’s a small business, you can’t afford to sponsor coast-to-coast television commercials, or to send brochures by mail to more than 100 million U.S. households. Instead, you plan to target the most likely customers – those who are environmentally conscious, have higher than average family incomes, and who live in areas where there is enough water and sunshine to support lawns and solar power.

Fortunately, lots of data are available to help you define your mailing list. Household incomes are routinely reported to banks and other financial institutions when families apply for mortgages, loans, and credit cards. Personal tastes related to issues like the environment are reflected in behaviors such as magazine subscriptions and credit card purchases. Firms like Claritas amass such data, and transform it into information by creating “lifestyle segments” – categories of households that have similar incomes and tastes. Your solar lawnmower company can purchase lifestyle segment information by 5-digit ZIP code, or even by ZIP+4 codes, which designate individual households.

It’s astonishing how companies like Claritas can create valuable information from the millions upon millions of transactions that are recorded every day. Their products are made possible by the fact that the original data exist in digital form, and because the companies have developed information systems that enable them to transform the data into information that companies like yours value. The fact that lifestyle information products are often delivered by geographic areas, such as ZIP codes, speaks to the appeal of geographic information systems.

TRY THIS

Try out the demo of what Claritas used to call the “You Are Where You Live” tool. The Nielson Company has acquired Claritas and the tool is now called “MyBestSegments.” Point your browser to the [My Best Segments](#) page. Click the button labeled “ZIP Code Look-up.”

Enter your ZIP code then choose a segmentation system. Do the lifestyle segments, listed on the left, seem accurate for your community? If you don’t live in the United States, try Penn State’s Zip code, 16802. Does the market segmentation match your expectations? Registered students are welcome to post comments directly to this page.

1.6. Databases, Mapping, and GIS

One of our objectives in this first chapter is to be able to define a geographic information system. Here's a tentative definition: **A GIS is a computer-based tool used to help people transform geographic data into geographic information.**

The definition implies that a GIS is somehow different from other information systems, and that geographic data are different from non-geographic data. Let's consider the differences next.

1.7. Database Management Systems

Claritas and similar companies use database management systems (DBMS) to create the "lifestyle segments" that I referred to in the previous section. Basic database concepts are important since GIS incorporates much of the functionality of DBMS.

Digital data are stored in computers as files. Often, data are arrayed in tabular form. For this reason, data files are often called **tables**. A **database** is a collection of tables. Businesses and government agencies that serve large clienteles, such as telecommunications companies, airlines, credit card firms, and banks, rely on extensive databases for their billing, payroll, inventory, and marketing operations. **Database management systems** are information systems that people use to store, update, and analyze non-geographic databases.

Often, data files are tabular in form, composed of rows and columns. **Rows**, also known as **records**, correspond with individual entities, such as customer accounts. **Columns** correspond with the various **attributes** associated with each entity. The attributes stored in the accounts database of a telecommunications company, for example, might include customer names, telephone numbers, addresses, current charges for local calls, long distance calls, taxes, etc.

Geographic data are a special case: records correspond with places, not people or accounts. Columns represent the attributes of places. The data in the following table, for example, consist of records for Pennsylvania counties. Columns contain selected attributes of each county, including the county's ID code, name, and 1980 population.

1980 Population Data for PA Counties

FIPS Code	County	1980 Pop
42001	Adams County	78274
42003	Allegheny County	1336449
42005	Armstrong County	73478
42007	Beaver County	186093
42009	Bedford County	47919
42011	Berks County	336523
42013	Blair County	130542
42015	Bradford County	60967
42017	Bucks County	541174
42019	Butler County	152013
42021	Cambria County	163062
42023	Cameron County	5913
42025	Carbon County	56846
42027	Centre County	124812

The contents of one file in a database.

The example is a very simple file, but many geographic attribute databases are in fact very large (the U.S. is made up of over 3,000 counties, almost 50,000 census tracts, about 43,000 five-digit ZIP code areas and many tens of thousands more ZIP+4 code areas). Large databases consist not only of lots of data, but also lots of files. Unlike a spreadsheet, which performs calculations only on data that are present in a single document, database management systems allow users to store data in, and retrieve data from, many separate files. For example, suppose an analyst wished to calculate population change for Pennsylvania counties between the 1980 and 1990 censuses. More than likely, 1990 population data would exist in a separate file, like so:

**1990 Population Data
for PA Counties**

FIPS Code	1990 Pop
42001	84921
42003	1296037
42005	73872
42007	187009
42009	49322
42011	352353
42013	131450
42015	62352
42017	578715
42019	167732
42021	158500
42023	5745
42025	58783
42027	131489

Another file in a database. A database management system (DBMS) can relate this file to the prior one illustrated above because they share the list of attributes called “FIPS Code.”

If two data files have at least one common attribute, a DBMS can combine them in a single new file. The common attribute is called a **key**. In this example, the key was the county FIPS code (FIPS stands for Federal Information Processing Standard). The DBMS allows users to produce new data as well as to retrieve existing data, as suggested by the new “% Change” attribute in the table below.

**Percent Change in Populations for PA Counties
1980-1990**

FIPS	County	1980	1990	% Change
42001	Adams	78274	84921	8.5
42003	Allegheny	1336449	1296037	-3
42005	Armstrong	73478	73872	0.5
42007	Beaver	186093	187009	0.5
42009	Bedford	47919	49322	2.9
42011	Berks	336523	352353	4.7
42013	Blair	130542	131450	0.7
42015	Bradford	60967	62352	2.3
42017	Bucks	541174	578715	6.9
42019	Butler	152013	167732	10.3
42021	Cambria	163062	158500	-2.8
42023	Cameron	5913	5745	-2.8
42025	Carbon	56846	58783	3.4
42027	Centre	124812	131489	5.3

A new file produced from the prior two files as a result of two database operations. One operation merged the contents of the two files without redundancy. A second operation produced a new attribute—" % Change"—dividing the difference between "1990 Pop" and "1980 Pop" by "1980 Pop" and expressing the result as a percentage.

Database management systems are valuable because they provide secure means of storing and updating data. Database administrators can protect files so that only authorized users can make changes. DBMS provide transaction management functions that allow multiple users to edit the database simultaneously. In addition, DBMS also provide sophisticated means to retrieve data that meet user specified criteria. In other words, they enable users to select data in response to particular questions. A question that is addressed to a database through a DBMS is called a **query**.

Database queries include basic set operations, including union, intersection, and difference. The product of a **union** of two or more data files is a single file that includes all records and attributes, without redundancy. An **intersection** produces a data file that contains only records present in all files. A **difference** operation produces a data file that eliminates records that appear in both original files. (Try drawing Venn diagrams—intersecting circles that show relationships between two or more entities—to illustrate the three operations. Then compare your sketch to [the venn diagram example](#).) All operations that involve multiple data files rely on the fact that all files contain a common key. The key allows the database system to relate the separate files. Databases that contain numerous files that share one or more keys are called **relational databases**. Database systems that enable users to produce information from relational databases are called **relational database management systems**.

A common use of database queries is to identify subsets of records that meet criteria established by the user. For example, a credit card company may wish to identify all accounts that are 30 days or more past due. A county tax assessor may need to list all properties not assessed within the past 10 years. Or the U.S. Census Bureau may wish to identify all addresses that need to be visited by census takers, because census questionnaires were not returned

by mail. DBMS software vendors have adopted a standardized language called SQL (Structured Query Language) to pose such queries.

PRACTICE QUIZ

1.8. Mapping Systems

GIS (geographic information systems) arose out of the need to perform spatial queries on geographic data. A spatial query requires knowledge of locations as well as attributes. For example, an environmental analyst might want to know which public drinking water sources are located within one mile of a known toxic chemical spill. Or, a planner might be called upon to identify property parcels located in areas that are subject to flooding. To accommodate geographic data and spatial queries, database management systems need to be integrated with mapping systems. Until about 1990, most maps were printed from handmade drawings or engravings. Geographic data produced by draftspersons consisted of graphic marks inscribed on paper or film. To this day, most of the lines that appear on topographic maps published by the U.S. Geological Survey were originally engraved by hand. The place names shown on the maps were affixed with tweezers, one word at a time. Needless to say, such maps were expensive to create and to keep up to date. Computerization of the mapmaking process had obvious appeal.

Computer-aided design (CAD) CAD systems were originally developed for engineers, architects, and other design professionals who needed more efficient means to create and revise precise drawings of machine parts, construction plans, and the like. In the 1980s, mapmakers began to adopt CAD in place of traditional map drafting. CAD operators encode the locations and extents of roads, streams, boundaries and other entities by tracing maps mounted on electronic drafting tables, or by key-entering location coordinates, angles, and distances. Instead of graphic features, CAD data consist of digital features, each of which is composed of a set of point locations. Calculations of distances, areas, and volumes can easily be automated once features are digitized. Unfortunately, CAD systems typically do not encode data in forms that support spatial queries. In 1988, a geographer named David Cowen illustrated the benefits and shortcomings of CAD for spatial decision making. He pointed out that a CAD system would be useful for depicting the streets, property parcel boundaries, and building footprints of a residential subdivision. A CAD operator could point to a particular parcel, and highlight it with a selected color or pattern. “A typical CAD system”, Cowen observed, “could not automatically shade each parcel based on values in an assessor’s database containing information regarding ownership, usage, or value, however.” A CAD system would be of limited use to someone who had to make decisions about land use policy or tax assessment.

Desktop mapping An evolutionary stage in the development of GIS, desktop mapping systems like Atlas*GIS combined some of the capabilities of CAD systems with rudimentary linkages between location data and attribute data. A desktop mapping system user could produce a map in which property parcels are automatically colored according to various categories of property values, for example. Furthermore, if property value categories were redefined, the map’s appearance could be updated automatically. Some desktop mapping systems even supported simple queries that allow users to retrieve records from a single attribute file. Most real-world decisions require more sophisticated queries involving multiple data files. That’s where real GIS comes in.

Geographic information systems (GIS) As stated earlier, information systems assist decision makers by enabling them to transform data into useful information. GIS specializes in helping users transform geographic data into geographic information. David Cowen (1988) defined GIS as a decision support tool that combines the attribute data handling capabilities of relational database management systems with the spatial data handling capabilities of CAD and desktop mapping systems. In particular, GIS enables decision makers to identify locations or routes whose attributes match multiple criteria, even though entities and attributes may be encoded in many different data files.

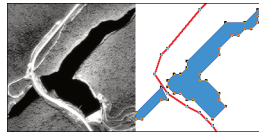
Innovators in many fields, including engineers, computer scientists, geographers, and others, started developing digital mapping and CAD systems in the 1950s and 60s. One of the first challenges they faced was to convert the graphical data stored on paper maps into digital data that could be stored in, and processed by, digital

computers. Several different approaches to representing locations and extents in digital form were developed. The two predominant representation strategies are known as “vector” and “raster.”

1.9. Representation Strategies for Mapping

Recall that data consist of symbols that represent measurements. Digital geographic data are encoded as alphanumeric symbols that represent locations and attributes of locations measured at or near Earth’s surface. No geographic data set represents every possible location, of course. The Earth is too big, and the number of unique locations is too great. In much the same way that public opinion is measured through polls, geographic data are constructed by measuring representative **samples** of locations. And just as serious opinion polls are based on sound principles of statistical sampling, so too do geographic data represent reality by measuring carefully chosen samples of locations. Vector and raster data are, at essence, two distinct sampling strategies.

The **vector** approach involves sampling locations at intervals along the length of linear entities (like roads), or around the perimeter of areal entities (like property parcels). When they are connected by lines, the sampled points form line features and polygon features that approximate the shapes of their real-world counterparts.



Two frames (the first and last) of an animation showing the construction of a vector representation of a reservoir and highway.

TRY THIS

Click the graphic above to download and view the animation file (vector.avi, 1.6 Mb) in a separate Microsoft Media Player window.

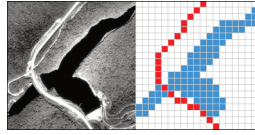
To view the [same animation in QuickTime format \(vector.mov, 1.6 Mb\)](#), [click here](#). Requires the QuickTime plugin, which is available free at [apple.com](#).

The aerial photograph above left shows two entities, a reservoir and a highway. The graphic above right illustrates how the entities might be represented with vector data. The small squares are nodes: point locations specified by latitude and longitude coordinates. Line segments connect nodes to form line features. In this case, the line feature colored red represents the highway. Series of line segments that begin and end at the same node form polygon features. In this case, two polygons (filled with blue) represent the reservoir.

The vector data model is consistent with how surveyors measure locations at intervals as they traverse a property boundary. Computer-aided drafting (CAD) software used by surveyors, engineers, and others, stores data in vector form. CAD operators encode the locations and extents of entities by tracing maps mounted on electronic drafting tables, or by key-entering location coordinates, angles, and distances. Instead of graphic features, CAD data consist of digital features, each of which is composed of a set of point locations.

The vector strategy is well suited to mapping entities with well-defined edges, such as highways or pipelines or property parcels. Many of the features shown on paper maps, including contour lines, transportation routes, and political boundaries, can be represented effectively in digital form using the vector data model.

The **raster** approach involves sampling attributes at fixed intervals. Each sample represents one cell in a checkerboard-shaped grid.



Two frames (the first and last) of an animation showing the construction of a raster representation of a reservoir and highway.

TRY THIS

Click the graphic above to download and view the animation file (raster.avi, 0.8 Mb) in a separate Microsoft Media Player window.

To view the [same animation in QuickTime format \(raster.mov, 0.6 Mb\)](#), [click here](#). Requires the QuickTime plugin, which is available free at [apple.com](#).

The graphic above illustrates a raster representation of the same reservoir and highway as shown in the vector representation. The area covered by the aerial photograph has been divided into a grid. Every grid cell that overlaps one of the two selected entities is encoded with an attribute that associates it with the entity it represents. Actual raster data would not consist of a picture of red and blue grid cells, of course; they would consist of a list of numbers, one number for each grid cell, each number representing an entity. For example, grid cells that represent the highway might be coded with the number “1” and grid cells representing the reservoir might be coded with the number “2.”

The raster strategy is a smart choice for representing phenomena that lack clear-cut boundaries, such as terrain elevation, vegetation, and precipitation. Digital airborne imaging systems, which are replacing photographic cameras as primary sources of detailed geographic data, produce raster data by scanning the Earth’s surface pixel by pixel and row by row.

Both the vector and raster approaches accomplish the same thing: they allow us to caricature the Earth’s surface with a limited number of locations. What distinguishes the two is the sampling strategies they embody. The vector approach is like creating a picture of a landscape with shards of stained glass cut to various shapes and sizes. The raster approach, by contrast, is more like creating a mosaic with tiles of uniform size. Neither is well suited to all applications, however. Several variations on the vector and raster themes are in use for specialized applications, and the development of new object-oriented approaches is underway.

PRACTICE QUIZ

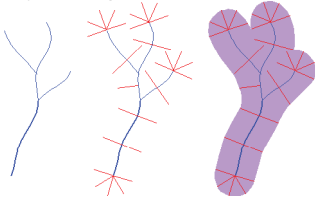
1.10. Automated Map Analysis

As I mentioned earlier, the original motivation for developing computer mapping systems was to automate the map making process. Computerization has not only made map making more efficient, it has also removed some of the technological barriers that used to prevent people from making maps themselves. What used to be an arcane craft practiced by a few specialists has become a “cloud” application available to any networked computer user. When I first started writing this course in 1997, my example was the mapping extension included in Microsoft Excel 97, which made creating a simple map as easy as creating a graph. Ten years later, who hasn’t used Google Maps or MapQuest?

As much as computerization has changed the way maps are made, it has had an even greater impact on how maps can be used. Calculations of distance, direction, and area, for example, are tedious and error-prone operations with paper maps. Given a digital map, such calculations can easily be automated. Those who are familiar with CAD systems know this from first-hand experience. Highway engineers, for example, rely on aerial imagery and

digital mapping systems to estimate project costs by calculating the volumes of rock that need to be excavated from hillsides and filled into valleys.

The ability to automate analytical tasks not only relieves tedium and reduces errors. It also allows us to perform tasks that would otherwise seem impractical. Consider, for example, if you were asked to plot on a map a 100-meter-wide **buffer** zone surrounding a protected stream. If all you had to work with was a paper map, a ruler, and a pencil, you might have a lengthy job on your hands. You might draw lines scaled to represent 100 meters, perpendicular to the river on both sides, at intervals that vary in frequency with the sinuosity of the stream. Then you might plot a perimeter that connects the end points of the perpendicular lines. If your task was to create hundreds of such buffer zones, you might conclude that automation is a necessity, not just a luxury.



Surrounding a protected stream with a buffer polygon.

Some tasks can be implemented equally well in either vector- or raster- oriented mapping systems. Other tasks are better suited to one representation strategy or another. The calculation of slope, for example, or of **gradient**—the direction of maximum slope along a surface—is more efficiently accomplished with raster data. The slope of one raster grid cell may be calculated by comparing its elevation to the elevations of the eight cells that surround it. Raster data are also preferred for a procedure called **viewshed analysis** that predicts which portions of a landscape will be in view, or hidden from view, from a particular perspective.

Some mapping systems provide ways to analyze attribute data as well as locational data. For example, the Excel mapping extension I mentioned above links the geographic data display capabilities of a mapping system with the data analysis capabilities of a spreadsheet. As you probably know, spreadsheets like Excel let users perform calculations on individual fields, columns, or entire files. A value changed in one field automatically changes values throughout the spreadsheet. Arithmetic, financial, statistical, and even certain database functions are supported. But as useful as spreadsheets are, they were not engineered to provide secure means of managing and analyzing large databases that consist of many related files, each of which is the responsibility of a different part of an organization. A spreadsheet is not a DBMS. And by the same token, a mapping system is not a GIS.

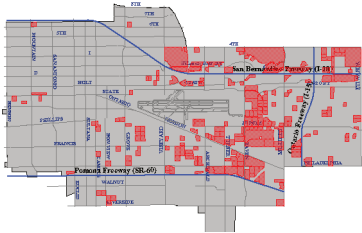
1.11. Geographic Information Systems

The preceding discussion leads me to revise my working definition:

As I mentioned earlier, a geographer named David Cowen defined GIS as **a decision-support tool that combines the capabilities of a relational database management system with the capabilities of a mapping system** (1988). Cowen cited an earlier study by William Carstensen (1986), who sought to establish criteria by which local governments might choose among competing GIS products. Carstensen chose site selection as an example of the kind of complex task that many organizations seek to accomplish with GIS. Given the necessary database, he advised local governments to expect that a fully functional GIS should be able to identify property parcels that are:

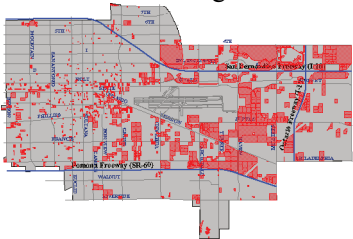
- At least five acres in size;
- Vacant or for sale;
- Zoned commercial;
- Not subject to flooding;
- Located not more than one mile from a heavy duty road; and
- Situated on terrain whose maximum slope is less than ten percent.

The first criterion—identifying parcels five acres or more in size—might require two operations. As described earlier, a mapping system ought to be able to calculate automatically the area of a parcel. Once the area is calculated and added as a new attribute into the database, an ordinary database query could produce a list of parcels that satisfy the size criterion. The parcels on the list might also be highlighted on a map, as in the example below.



The cartographic result of a database query identifying all property parcels greater than or equal to five acres in size. (City of Ontario, CA, GIS Department. Used by permission.)

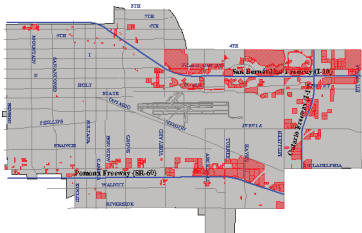
The ownership status of individual parcels would be an attribute of a property database maintained by a local tax assessor's office. Parcels whose ownership status attribute value matched the criteria "vacant" or "for sale" could be identified through another ordinary database query.



The cartographic result of a spatial intersection (or map overlay) operation identifying all property parcels zoned for commercial (C-1) development. (City of Ontario, CA, GIS Department. Used by permission.)

Carstensen's third criterion was to determine which parcels were situated within areas zoned for commercial development. This would be simple if authorized land uses were included as an attribute in the community's property parcel database. This is unlikely to be the case, however, since zoning and taxation are the responsibilities of different agencies. Typically, parcels and land use zones exist as separate paper maps. If the maps were prepared at the same scale, and if they accounted for the shape of the Earth in the same manner, then they could be superimposed one over another on a light table. If the maps let enough light through, parcels located within commercial zones could be identified.

The GIS approach to a task like this begins by digitizing the paper maps, and by producing corresponding attribute data files. Each digital map and attribute data file is stored in the GIS separately, like separate **maplayers**. A fully functional GIS would then be used to perform a **spatial intersection** that is analogous to the overlay of the paper maps. Spatial intersection, otherwise known as **map overlay**, is one of the defining capabilities of GIS.

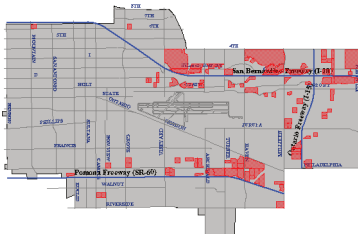


The cartographic result of a buffer operation identifying all property parcels located within a specified distance of a specified type of highway. (City of Ontario, CA, GIS Department. Used by permission.)

Another of Carstensen's criteria was to identify parcels located within one mile of a heavy-duty highway. Such a task requires a digital map and associated attributes produced in such a way as to allow heavy-duty highways to be differentiated from other geographic entities. Once the necessary database is in place, a **buffer** operation can be used

to create a polygon feature whose perimeter surrounds all “heavy duty highway” features at the specified distance. A spatial intersection is then performed, isolating the parcels within the buffer from those outside the buffer.

To produce a final list of parcels that meet all the site selection criteria, the GIS analyst might perform an **intersection** operation that creates a new file containing only those records that are present in all the other intermediate results.



The cartographic result of the intersection of the above three figures. Only the parcels shown in this map satisfy all of the site selection criteria. (City of Ontario, CA, GIS Department. Used by permission.)

I created the maps shown above in 1998 using the Geographic Information Web Server of the City of Ontario, California. Although it is no longer supported, the City of Ontario was one of the first of its kind to provide much of the functionality required to perform a site suitability analysis online. Today, many local governments offer similar Internet map services to current and prospective taxpayers.

TRY THIS

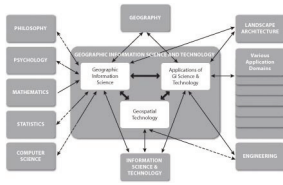
Find an online site selection utility similar to the one formerly provided by the City of Ontario. Registered Penn State students can post a comment to this page describing the site’s functionality, and comparing it with the capabilities of the example illustrated above.

1.12. Geographic Information Science and Technology

So far in this chapter I’ve tried to make sense of GIS in relation to several information technologies, including database management, computer-aided design, and mapping systems. At this point I’d like to expand the discussion to consider GIS as one element in a much larger field of study called “Geographic Information Science and Technology” (GIS&T). As shown in the following illustration, GIS&T encompasses three subfields including:

- **Geographic Information Science**, the multidisciplinary research enterprise that addresses the nature of geographic information and the application of geospatial technologies to basic scientific questions;
- **Geospatial Technology**, the specialized set of information technologies that support acquisition, management, analysis, and visualization of geo-referenced data, including the Global Navigation Satellite System (GPS and others), satellite, airborne, and shipboard remote sensing systems; and GIS and image analysis software tools; and
- **Applications of GIS&T**, the increasingly diverse uses of geospatial technology in government, industry, and academia. This is the subfield in which most GIS professionals work.

Arrows in the diagram below reflect relationships among the three subfields, as well as to numerous other fields, including Geography, Landscape Architecture, Computer Science, Statistics, Engineering, and many others. Each of these fields has influenced, and some have been influenced by, the development of GIS&T. It is important to note that these fields and subfields do not neatly correspond with professions like GIS analyst, photogrammetrist, or land surveyor. Rather, GIS&T is a *nexus* of overlapping professions that differ in backgrounds, disciplinary allegiances, and regulatory status.



The field of Geographic Information Science and Technology (GIS&T) and its relations to other fields. Two-way relations that are half-dashed represent asymmetrical contributions between allied fields. (© 2006 Association of American Geographers and University Consortium for Geographic Information Science. Used by permission. All rights reserved.)

The illustration above first appeared in the *Geographic Information Science and Technology Body of Knowledge* (DiBiase, DeMers, Johnson, Kemp, Luck, Plewe, and Wentz, 2006), published by the University Consortium for Geographic Information Science (UCGIS) and the Association of American Geographers (AAG) in 2006. The *Body of Knowledge* is a community-developed inventory of the knowledge and skills that define the GIS&T field. Like the bodies of knowledge developed in Computer Science and other fields, the *GIS&T BoK* represents the GIS&T knowledge domain as a hierarchical list of knowledge areas, units, topics, and educational objectives. The ten knowledge areas and 73 units that make up the first edition are shown in the table below. Twenty-six “core” units (those in which all graduates of a degree or certificate program should be able to demonstrate some level of mastery) are shown in bold type. Not shown are the 329 topics that make up the units, or the 1,660 education objectives by which topics are defined. These appear in the full text of the *GIS&T BoK*. Unfortunately, the full text is not freely available online. An important related work produced by the U.S. Department of Labor is, however. We’ll take a look at that shortly.

KNOWLEDGE AREAS AND UNITS COMPRISING THE 1ST EDITION OF THE GIS&T BOK

-Knowledge Area AM. Analytical Methods

- Unit AM1 Academic and analytical origins
- Unit AM2 Query operations and query languages
- Unit AM3 Geometric measures
- Unit AM4 Basic analytical operations
- Unit AM5 Basic analytical methods
- Unit AM6 Analysis of surfaces
- Unit AM7 Spatial statistics
- Unit AM8 Geostatistics
- Unit AM9 Spatial regression and econometrics
- Unit AM10 Data mining
- Unit AM11 Network analysis
- Unit AM12 Optimization and location-allocation modeling

-Knowledge Area CF. Conceptual Foundations

- Unit CF1 Philosophical foundations
- Unit CF2 Cognitive and social foundations
- Unit CF3 Domains of geographic information
- Unit CF4 Elements of geographic information
- Unit CF5 Relationships
- Unit CF6 Imperfections in geographic information

-Knowledge Area CV. Cartography and Visualization

- Unit CV1 History and trends
- Unit CV2 Data considerations
- Unit CV3 Principles of map design

- Unit CV4 Graphic representation techniques
- Unit CV5 Map production
- Unit CV6 Map use and evaluation**
 - Knowledge Area DA. Design Aspects**
 - Unit DA1 The scope of GI S&T system design
 - Unit DA2 Project definition
 - Unit DA3 Resource planning
 - Unit DA4 Database design**
 - Unit DA5 Analysis design
 - Unit DA6 Application design
 - Unit DA7 System implementation
 - Knowledge Area DM. Data Modeling**
 - Unit DM1 Basic storage and retrieval structures
 - Unit DM2 Database management systems**
 - Unit DM3 Tessellation data models**
 - Unit DM4 Vector and object data models**
 - Unit DM5 Modeling 3D, temporal, and uncertain phenomena
 - Knowledge Area DN. Data Manipulation**
 - Unit DN1 Representation transformation**
 - Unit DN2 Generalization and aggregation**
 - Unit DN3 Transaction management of geospatial data
 - Knowledge Area GC. Geocomputation**
 - Unit GC1 Emergence of geocomputation
 - Unit GC2 Computational aspects and neurocomputing
 - Unit GC3 Cellular Automata (CA) models
 - Unit GC4 Heuristics
 - Unit GC5 Genetic algorithms (GA)
 - Unit GC6 Agent-based models
 - Unit GC7 Simulation modeling
 - Unit GC8 Uncertainty
 - Unit GC9 Fuzzy sets
 - Knowledge Area GD. Geospatial Data**
 - Unit GD1 Earth geometry**
 - Unit GD2 Land partitioning systems
 - Unit GD3 Georeferencing systems**
 - Unit GD4 Datums**
 - Unit GD5 Map projections**
 - Unit GD6 Data quality**
 - Unit GD7 Land surveying and GPS**
 - Unit GD8 Digitizing
 - Unit GD9 Field data collection
 - Unit GD10 Aerial imaging and photogrammetry**
 - Unit GD11 Satellite and shipboard remote sensing**
 - Unit GD12 Metadata, standards, and infrastructures**
 - Knowledge Area GS. GIS&T and Society**
 - Unit GS1 Legal aspects
 - Unit GS2 Economic aspects
 - Unit GS3 Use of geospatial information in the public sector
 - Unit GS4 Geospatial information as property

- Unit GS5 Dissemination of geospatial information
- Unit GS6 Ethical aspects of geospatial information and technology**
- Unit GS7 Critical GIS
- Knowledge Area OI. Organizational and Institutional Aspects**
- Unit OI1 Origins of GI S&T
- Unit O2 Managing the GI system operations and infrastructure
- Unit OI3 Organizational structures and procedures
- Unit OI4 GI S&T workforce themes
- Unit OI5 Institutional and inter-institutional aspects**
- Unit OI6 Coordinating organizations (national and international)**

Ten knowledge areas and 73 units comprising the 1st edition of the GIS&T BoK. Core units are indicated with bold type. (© 2006 Association of American Geographers and University Consortium for Geographic Information Science. Used by permission. All rights reserved.)

Notice that the knowledge area that includes the most core units is GD: Geospatial Data. This course focuses on the sources and distinctive characteristics of geographic data. This is one part of the knowledge base that most successful geospatial professionals possess. The Department of Labor's Geospatial Technology Competency Model (GTCM) highlights this and other essential elements of the geospatial knowledge base. We'll consider it next.

1.13. Geospatial Competencies and Our Curriculum

A body of knowledge is one way to think about the GIS&T field. Another way is as an industry made up of agencies and firms that produce and consume goods and services, generate sales and (sometimes) profits, and employ people. In 2003, the U.S. Department of Labor (DoL) identified “geospatial technology” as one of 14 “high growth” technology industries, along with biotech, nanotech, and others. However, the DoL also observed that the geospatial technology industry was ill-defined, and poorly understood by the public.

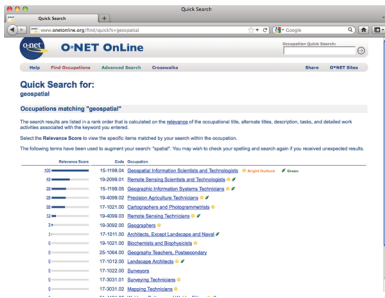
Subsequent efforts by the DoL and other organizations helped to clarify the industry's nature and scope. Following a series of “roundtable” discussions involving industry thought leaders, the Geospatial Information Technology Association (GITA) and the Association of American Geographers (AAG) submitted the following “concensus” definition to DoL in 2006:

The geospatial industry acquires, integrates, manages, analyzes, maps, distributes, and uses geographic, temporal, and spatial information and knowledge. The industry includes basic and applied research, technology development, education, and applications to address the planning, decision making, and operational needs of people and organizations of all types.

In addition to the proposed industry definition, the GITA and AAG report recommended that DoL establish additional occupations in recognition of geospatial industry workforce activities and needs. At the time, the existing geospatial occupations included only Surveyors, Surveying Technicians, Mapping Technicians, and Cartographers and Photogrammetrists. Late in 2009, with input from the GITA, AAG, and other stakeholders, the DoL established six new geospatial occupations: Geospatial Information Scientists and Technologists, Geographic Information Systems Technicians, Remote Sensing Scientists and Technologists, Remote Sensing Technicians, Precision Agriculture Technicians, and Geodetic Surveyors.

TRY THIS

Investigate the geospatial occupations at the [U.S. Department of Labor's “O*Net” database](#). Enter “geospatial” in the search field named “Occupation Quick Search.” Follow links to occupation descriptions. Note the estimates for 2008 employment and employment growth through 2018. Also note that, for some anomalous reason, the keyword “geospatial” is not associated with the occupation “Geodetic Surveyor.”

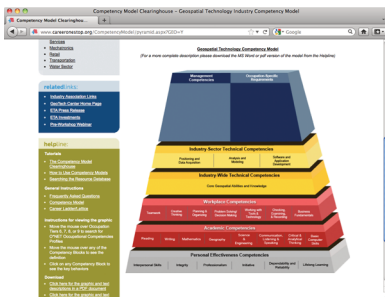


Meanwhile, DoL commenced a “competency modeling” initiative for high-growth industries in 2005. Their goal was to help educational institutions like ours meet the demand for qualified technology workers by identifying what workers need to know and be able to do. At DoL, a *competency* is “the capability to apply or use a set of related knowledge, skills, and abilities required to successfully perform ‘critical work functions’ or tasks in a defined work setting” (Ennis 2008). A *competency model* is “a collection of competencies that together define successful performance in a particular work setting.”

Workforce analysts at DoL began work on a Geospatial Technology Competency Model (GTCM) in 2005. Building on their research, a panel of accomplished practitioners and educators produced a complete draft of the GTCM, which they subsequently revised in response to public comments. Published in June 2010, the GTCM identifies the competencies that characterize successful workers in the geospatial industry. In contrast to *GIS&T Body of Knowledge*, an academic project meant to define the nature and scope of the field, the GTCM is an industry specification that defines what individual workers and students should aspire to know and learn.

TRY THIS

Explore the [Geospatial Technology Competency Model \(GTCM\)](#) at the U.S. Department of Labor’s Competency Model Clearinghouse. Under “Industry Competency Models,” follow the link “Geospatial Technology.” There, the pyramid (as shown below) is an image map which you can click to reveal the various competencies. The complete GTCM is also available as a Word doc and PDF file.



The GTCM specifies several “tiers” of competencies, progressing from general to occupationally specific. Tiers 1 through 3 (the gray and red layers), called Foundation Competencies, specify general workplace behaviors and knowledge that successful workers in most industries exhibit. Tiers 4 and 5 (yellow) include the distinctive technical competencies that characterize a given industry and its three sectors: Positioning and Data Acquisition, Analysis and Modeling, and Programming and Application Development. Above Tier 5 are additional Tiers corresponding to the occupation-specific competencies and requirements that are specified in the occupation descriptions published at O*NET Online, and in a Geospatial Management Competency Model that is in development as of January, 2012.

One way educational institutions and students can use the GTCM is as a guideline for assessing how well

curricula align with workforce needs. The Penn State Online GIS program conducted such an assessment in 2011. Results appear in the spreadsheet linked below.

TRY THIS

Open the [attached Excel spreadsheet](#) to see how our Penn State Online GIS curricula address workforce needs identified in the GTCM.

The sheet will open on a cover page. At the bottom of the sheet are **tabs** that correspond to Tiers 1-5 of the GTCM. Click the tabs to view the worksheet associated with the Tier you want to see.

In each Tier worksheet, **rows** correspond to the GTCM competencies. **Columns** correspond to the Penn State Online courses included in the assessment. Courses that are required for most students are highlighted light blue. Course authors and instructors were asked to state what students actually do in relation to each of the GTCM competencies. Use the **scroll bar** at the bottom right edge of the sheet to reveal more courses.

Open the [attached Flash movie](#) to view a video demonstration of how to navigate the spreadsheet.

By studying this spreadsheet you'll gain insight about how individual courses, and how the Penn State Online curriculum as a whole, relates to geospatial workforce needs. If you're interested in comparing ours to curricula at other institutions, ask if they've conducted a similar assessment. If they haven't, ask why not.

Finally, don't forget that you can preview much of our online courseware through our [Open Educational Resources initiative](#).

1.14. Distinguishing Properties of Geographic Data

The claim that geographic information science is a distinct field of study implies that spatial data are somehow special data. Goodchild (1992) points out several distinguishing properties of geographic information. I have paraphrased four such properties below. Understanding them, and their implications for the practice of geographic information science, is a key objective of this course.

1. Geographic data represent spatial locations and non-spatial attributes measured at certain times.
2. Geographic space is continuous.
3. Geographic space is nearly spherical.
4. Geographic data tend to be spatially dependent.

Let's consider each of these properties next.

1.15. Locations and Attributes

Geographic data represent spatial locations and non-spatial attributes measured at certain times. Goodchild (1992, p. 33) observes that “a spatial database has dual keys, allowing records to be accessed either by attributes or by locations.” Dual keys are not unique to geographic data, but “the spatial key is distinct, as it allows operations to be defined which are not included in standard query languages.” In the intervening years, software developers have created variations on SQL that incorporate spatial queries. The dynamic nature of geographic phenomena complicates the issue further, however. The need to pose spatio-temporal queries challenges geographic information scientists (GIScientists) to develop ever more sophisticated ways to represent geographic phenomena, thereby enabling analysts to interrogate their data in ever more sophisticated ways.

1.16. Continuity

Geographic space is continuous. Although dual keys are not unique to geographic data, one property of the spatial key is. “What distinguishes spatial data is the fact that the spatial key is based on two continuous dimensions” (Goodchild, 1992, p.33). “Continuous” refers to the fact that there are no gaps in the Earth’s surface. Canyons, crevasses, and even caverns notwithstanding, there is no position on or near the surface of the Earth that cannot be fixed within some sort of coordinate system grid. Nor is there any theoretical limit to how exactly a position can be specified. Given the precision of modern positioning technologies, the number of unique point positions that could be used to define a geographic entity is practically infinite. Because it’s not possible to measure, let alone to store, manage, and process, an infinite amount of data, **all geographic data is selective, generalized, approximate.** Furthermore, **the larger the territory covered by a geographic database, the more generalized the database tends to be.**

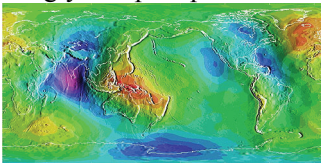
Geographic data are generalized according to scale. Click on the buttons beneath the map to zoom in and out on the town of Gorham. (U.S. Geological Survey). (**Note:** You will need to have the Adobe Flash player installed in order to complete this exercise. If you do not already have the Flash player, you can [download it for free from Adobe.](#))

For example, the illustration above shows a town called Gorham depicted on three different **topographic maps** produced by the United States Geological Survey. Gorham occupies a smaller space on the small-scale (1:250,000) map than it does at 1:62,000 or at 1:24,000. But the relative size of the feature isn’t the only thing that changes. Notice that the shape of the feature that represents the town changes also. As does the number of features and the amount of detail shown within the town boundary and in the surrounding area. The name for this characteristically parallel decline in map detail and map scale is **generalization**.

It is important to realize that generalization occurs not only on printed maps, but in digital databases as well. It is possible to represent phenomena with highly detailed features (whether they be made up of high-resolution raster grid cells or very many point locations) in a single **scale-independent** database. In practice, however, highly detailed databases are not only extremely expensive to create and maintain, but they also bog down information systems when used in analyses of large areas. For this reason, geographic databases are usually created at several scales, with different levels of detail captured for different intended uses.

1.17. Nearly Spherical

Geographic space is nearly spherical. The fact that the Earth is nearly, but not quite, a sphere poses some surprisingly complex problems for those who wish to specify locations precisely.



Differences in elevation between a geoid model and a reference ellipsoid. Deviations range from a high of 75 meters (colored red, over New Guinea) to a low of 104 meters (colored purple, in the Indian Ocean). (National Geodetic Survey, n. d.).

The **geographic coordinate system** of latitude and longitude coordinates provides a means to define positions on a sphere. Inaccuracies that are unacceptable for some applications creep in, however, when we confront the Earth’s “actual” irregular shape, which is called the **geoid**. Furthermore, the calculations of angles and distance that surveyors and others need to perform routinely are cumbersome with **spherical coordinates**.

That consideration, along with the need to depict the Earth on flat pieces of paper, compels us to transform the globe into a plane, and to specify locations in **plane coordinates** instead of spherical coordinates. The set of mathematical transformations by which spherical locations are converted to locations on a plane—called **map projections**—all lead inevitably to one or another form of inaccuracy.

All this is trouble enough, but we encounter even more difficulties when we seek to define “vertical” positions (elevations) in addition to “horizontal” positions. Perhaps it goes without saying that an elevation is the height of a location above some **datum**, such as mean sea level. Unfortunately, to be suitable for precise positioning, a datum must correspond closely with the Earth’s actual shape. Which brings us back again to the problem of the geoid.

We will consider these issues in greater depth in Chapter 2. For now, suffice it to say that geographic data are unique in having to represent phenomena that are distributed on a continuous and nearly spherical surface.

1.18. Spatial Dependency

Geographic data tend to be spatially dependent. Spatial dependence is “the propensity for nearby locations to influence each other and to possess similar attributes” (Goodchild, 1992, p.33). In other words, to paraphrase a famous geographer named Waldo Tobler, while everything is related to everything else, things that are close together tend to be more related than things that are far apart. Terrain elevations, soil types, and surface air temperatures, for instance, are more likely to be similar at points two meters apart than at points two kilometers apart. A statistical measure of the similarity of attributes of point locations is called **spatial autocorrelation**.

Given that geographic data are expensive to create, spatial dependence turns out to be a very useful property. We can sample attributes at a limited number of locations, then estimate the attributes of intermediate locations. The process of estimating unknown values from nearby known values is called **interpolation**. Interpolated values are reliable only to the extent that the spatial dependence of the phenomenon can be assumed. If we were unable to assume some degree of spatial dependence, it would be impossible to represent continuous geographic phenomena in digital form.

PRACTICE QUIZ

19. Geographic Data and Geographic Questions

The ultimate objective of all geospatial data and technologies, after all, is to produce knowledge. Most of us are interested in data only to the extent that they can be used to help understand the world around us, and to make better decisions. Decision making processes vary a lot from one organization to another. In general, however, the first steps in making a decision are to articulate the questions that need to be answered, and to gather and organize the data needed to answer the questions (Nyerges & Gollledge, 1997).

Geographic data and information technologies can be very effective in helping to answer certain kinds of questions. The expensive, long-term investments required to build and sustain GIS infrastructures can be justified only if the questions that confront an organization can be stated in terms that GIS is equipped to answer. As a specialist in the field, you may be expected to advise clients and colleagues on the strengths and weaknesses of GIS as a decision support tool. To follow are examples of the kinds of questions that are amenable to GIS analyses, along with questions that GIS is not so well suited to help answer.

QUESTIONS CONCERNING INDIVIDUAL GEOGRAPHIC ENTITIES

The simplest geographic questions pertain to individual entities. Such questions include:

QUESTIONS ABOUT SPACE

- Where is the entity located?
- What is its extent?

QUESTIONS ABOUT ATTRIBUTES

- What are the attributes of the entity located there?
- Do its attributes match one or more criteria?

QUESTIONS ABOUT TIME

- When were the entity's location, extent or attributes measured?
- Has the entity's location, extent, or attributes changed over time?

Simple questions like these can be answered effectively with a good printed map, of course. GIS becomes increasingly attractive as the number of people asking the questions grows, especially if they lack access to the required paper maps.

QUESTIONS CONCERNING MULTIPLE GEOGRAPHIC ENTITIES

Harder questions arise when we consider relationships among two or more entities. For instance, we can ask:

QUESTIONS ABOUT SPATIAL RELATIONSHIPS

- Do the entities contain one another?
- Do they overlap?
- Are they connected?
- Are they situated within a certain distance of one another?
- What is the best route from one entity to the others?
- Where are entities with similar attributes located?

QUESTIONS ABOUT ATTRIBUTE RELATIONSHIPS

- Do the entities share attributes that match one or more criteria?
- Are the attributes of one entity influenced by changes in another entity?

QUESTIONS ABOUT TEMPORAL RELATIONSHIPS

- Have the entities' locations, extents, or attributes changed over time?

Geographic data and information technologies are very well suited to answering moderately complex questions like these. GIS is most valuable to large organizations that need to answer such questions often.

QUESTIONS THAT GIS IS NOT PARTICULARLY GOOD AT ANSWERING

Harder still, however, are **explanatory questions**—such as *why* entities are located where they are, *why* they have the attributes they do, and *why* they have changed as they have. In addition, organizations are often concerned with **predictive questions**—such as what will happen at *this* location if thus-and-so happens *at that* location? In general, commercial GIS software packages cannot be expected to provide clear-cut answers to explanatory and predictive questions right out of the box. Typically, analysts must turn to specialized statistical packages and simulation routines. Information produced by these analytical tools may then be re-introduced into the GIS database, if necessary. Research and development efforts intended to more tightly couple analytical software with GIS software are underway within the GIScience community. It is important to keep in mind that decision support tools like GIS are no substitutes for human experience, insight, and judgment.

At the outset of the chapter I suggested that producing information by analyzing data is something like producing energy by burning coal. In both cases, technology is used to realize the potential value of a raw material. Also in both cases, the production process yields some undesirable by-products. Similarly, in the process of answering certain geographic questions, GIS tends to raise others, such as:

- Given the intrinsic imperfections of the data, how reliable are the results of the GIS analysis?
- Does the information produced through GIS analysis tend to systematically benefit some constituent groups at the expense of others?
- Should the data used to make the decision be made public?
- Does the use of GIS affect the organization’s decision-making processes in ways that are beneficial to its management, its employees, and its customers?

As is the case in so many endeavors, the answer to a geographic question usually includes more questions.

TRY THIS

Can you cite an example of a “hard” question that you and your GIS system have been called upon to address? Registered Penn State students can post a comment directly to this page.

1.20. Summary

It’s a truism among specialists in geographic information that the lion’s share of the cost of most GIS projects is associated with the development and maintenance of a suitable database. It seems appropriate, therefore, that our first course in geographic information systems should focus upon the properties of geographic data.

I began this first chapter by defining data in a generic sense, as sets of symbols that represent measurements of phenomena. I suggested that data are the raw materials from which information is produced. Information systems, such as database management systems, are technologies that people use to transform data into the information needed to answer questions, and to make decisions.

Spatial data are special data. They represent the locations, extents, and attributes of objects and phenomena that make up the Earth’s surface at particular times. Geographic data differ from other kinds of data in that they are distributed along a continuous, nearly spherical globe. They also have the unique property that the closer two entities are located, the more likely they are to share similar attributes.

GIS is a special kind of information system that combines the capabilities of database management systems with those of mapping systems. GIS is one object of study of the loosely-knit, multidisciplinary field called Geographic Information Science and Technology. GIS is also a profession—one of several that make up the geospatial industry. As Yogi Berra said, “In theory, there’s no difference between theory and practice. In practice

there is.” In the chapters and projects that follow, we’ll investigate the nature of geographic information from both conceptual and practical points of view.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

1.21. Bibliography

Carstensen, L. W. (1986). Regional land information systems development using relational databases and geographic information systems. *Proceedings of the AutoCarto*, London, 507-516.

City of Ontario, California. (n.d.). *Geographic information web server*. Retrieved on July 6, 1999 from <http://www.ci.ontario.ca.us/gis/index.asp> (since retired).

Cowen, D. J. (1988). GIS versus CAD versus DBMS: What are the differences? *Photogrammetric Engineering and Remote Sensing* 54:11, 1551-1555.

DiBiase, D. and twelve others (2010). *The New Geospatial Technology Competency Model: Bringing workforce needs into focus*. *URISA Journal* 22:2, 55-72.

DiBiase, D, M. DeMers, A. Johnson, K. Kemp, A. Luck, B. Plewe, and E. Wentz (2007). *Introducing the First Edition of the GIS&T Body of Knowledge*. *Cartography and Geographic Information Science*, 34(2), pp. 113-120. U.S. National Report to the International Cartographic Association.

Ennis, M. R. (2008). Competency models: A review of the literature and the role of the employment and training administration (ETA). http://www.careeronestop.org/COMPETENCYMODEL/info_documents/OPDRLiteratureReview.pdf.

GITA and AAG (2006). *Defining and communicating geospatial industry workforce demand: Phase I report*.

Goodchild, M. (1992). Geographical information science. *International Journal of Geographic Information Systems* 6:1, 31-45.

Goodchild, M. (1995). GIS and geographic research. In J. Pickles (Ed.), *Ground truth: the social implications of geographic information systems* (pp. of chapter). New York: Guilford.

National Decision Systems. *A zip code can make your company lots of money!* Retrieved on July 6, 1999 from <http://laguna.natdecsys.com/lifequiz> (since retired).

National Geodetic Survey. (1997). *Image generated from 15'x15' geoid undulations covering the planet Earth*. Retrieved 1999, from <http://www.ngs.noaa.gov/GEOID/geo-index.html> (since retired).

Nyerges, T. L. & Gollodge, R. G. (n.d.) *NCGIA core curriculum in GIS*, National Center for Geographic Information and Analysis, University of California, Santa Barbara, Unit 007. Retrieved November 12, 1997, from <http://www.ncgia.ucsb.edu/giscc/units/u007/u007.html> (since retired).

United States Department of the Interior Geological Survey. (1977). [map]. 1:24 000. 7.5 minute series. Washington, D.C.: USDI.

United States Geologic Survey. “Bellefonte, PA Quadrangle” (1971). [map]. 1:24 000. 7.5 minute series. Washington, D.C.:USGS.

University Consortium for Geographic Information Science. Retrieved April 26, 2006, from <http://www.ucgis.org>

Wilson, J. D. (2001). Attention data providers: A billion-dollar application awaits. *GEOWorld*, February, 54.

Worboys, M. F. (1995). *GIS: A computing perspective*. London: Taylor and Francis.

‹ 20. Summary up Chapter 2: Scales and Transformations ›

Chapter 2

Scales and Transformations

David DiBiase

2.1. Overview

Chapter 1 outlined several of the distinguishing properties of geographic data. One is that geographic data are necessarily generalized, and that generalization tends to vary with scale. A second distinguishing property is that the Earth's complex, nearly-spherical shape complicates efforts to specify exact positions on Earth's surface. This chapter explores implications of these properties by illuminating concepts of scale, Earth geometry, coordinate systems, the "horizontal datums" that define the relationship between coordinate systems and the Earth's shape, and the various methods for transforming coordinate data between 3D and 2D grids, and from one datum to another.

Compared to Chapter 1, Chapter 2 may seem long, technical, and abstract, particularly to those for whom these concepts are new. Registered students will notice that we've allotted more time to work through the chapter and associated quizzes. Seven practice quizzes are available in ANGEL to help registered students get a grip on these concepts. Chapter 2 also includes a graded quiz in the same open-book format as the practice quizzes. If you do reasonably well on the practice quizzes, you should do well enough on the graded quiz too.

Objectives

Students who successfully complete Chapter 2 should be able to:

1. Demonstrate your ability to specify geospatial locations using geographic coordinates;
2. Convert geographic coordinates between two different formats;
3. Explain the concept of a horizontal datum;
4. Calculate the change in a coordinate location due to a change from one horizontal datum to another;
5. Estimate the magnitude of "datum shift" associated with the adjustment from NAD 27 to NAD 83;
6. Recognize the kind of transformation that is appropriate to georegister two or more data sets;
7. Describe the characteristics of the UTM coordinate system, including its basis in the Transverse Mercator map projection;
8. Plot UTM coordinates on a map;
9. Describe the characteristics of the SPC system, including map projection on which it is based;
10. Convert geographic coordinates to SPC coordinates;
11. Interpret distortion diagrams to identify geometric properties of the sphere that are preserved by a particular projection; and
12. Classify projected graticules by projection family.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are

discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

2.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 2 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 2	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	<p>Submit eight practice quizzes including:</p> <ul style="list-style-type: none"> • Map Scale • Geographic Coordinate System • Horizontal Datums • Coordinate Transformations • UTM Coordinate System • SPC Coordinate System • Map Projection Properties • Classifying Map Projections <p>Practice quizzes are not graded and may be submitted more than once.</p>	Go to ANGEL > [your course section] > Lessons tab > Chapter 2 folder > [quiz]
3	<p>Perform “Try this” activities Including:</p> <ul style="list-style-type: none"> • Geographic coordinates practice application • Calculate horizontal datum shift • UTM coordinates practice application • Explore SPC zone characteristics • Interactive Album of Map Projections <p>“Try this” activities are not graded.</p>	Instructions are provided for each activity.
4	Submit the Chapter 2 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 2 folder > Chapter 2 Graded Quiz
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

2.3. Scale

You hear the word “scale” often when you work around people who produce or use geographic information. If

you listen closely, you'll notice that the term has several different meanings, depending on the context in which it is used. You'll hear talk about the scales of geographic phenomena and about the scales at which phenomena are represented on maps and aerial imagery. You may even hear the word used as a verb, as in "scaling a map" or "downscaling." The goal of this section is to help you learn to tell these different meanings apart, and to be able to use concepts of scale to help make sense of geographic data.

Specifically, in this part of Chapter 2 you will learn to:

1. Calculate map scale using representative fractions.
2. Describe the general relationship between map scale, detail, and accuracy.

2.4. Scale as Scope

Often "scale" is used as a synonym for "scope," or "extent." For example, the title of an international research project called The Large Scale Biosphere-Atmosphere Experiment in Amazonia (1999) uses the term "large scale" to describe a comprehensive study of environmental systems operating across a large region. This usage is common not only among environmental scientists and activists, but also among economists, politicians, and the press. Those of us who specialize in geographic information usually use the word "scale" differently, however.

2.5. Map and Photo Scale

When people who work with maps and aerial images use the word "scale," they usually are talking about the sizes of things that appear on a map or air photo, relative to the actual sizes of those things on the ground.

Map scale is the proportion between a distance on a map and a corresponding distance on the ground: (D_m / D_g).

By convention, the proportion is expressed as a "representative fraction" in which map distance (D_m) is reduced to 1. The proportion, or ratio, is also typically expressed in the form 1 : D_g rather than 1 / D_g .

The representative fraction 1:100,000, for example, means that a section of road that measures 1 unit in length on a map stands for a section of road on the ground that is 100,000 units long.

If we were to change the scale of the map such that the length of the section of road on the map was reduced to, say, 0.1 units in length, we would have created a **smaller-scale** map whose representative fraction is 0.1:100,000, or 1:1,000,000. When we talk about large- and small-scale maps and geographic data, then, we are talking about the relative sizes and levels of detail of the features represented in the data. In general, the larger the map scale, the more detail is shown. This tendency is illustrated below.

Geographic data are generalized according to scale. Click on the buttons beneath the map to zoom in and out on the town of Gorham. (Adapted from Thompson, 1988)

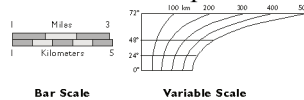
One of the defining characteristics of topographic maps is that scale is consistent across each map and within each map series. This isn't true for aerial imagery, however, except for images that have been **orthorectified**. As discussed in Chapter 6, large scale maps are typically derived from aerial imagery. One of the challenges associated with using air photos as sources of map data is that the scale of an aerial image varies from place to place as a function of the elevation of the terrain shown in the scene. Assuming that the aircraft carrying the camera maintains a constant flying height (which pilots of such aircraft try very hard to do), the distance between the camera and the ground varies along each flight path. This causes air photo scale to be larger where the terrain is higher and smaller where the terrain lower. An "orthorectified" image is one in which variations in scale caused by variations in terrain elevation (among other effects) have been removed.

You can calculate the average scale of an **unrectified** air photo by solving the equation $S_p = f / (H - h_{avg})$, where f is the focal length of the camera, H is the flying height of the aircraft above mean sea level, and h_{avg}

is the average elevation of the terrain. You can also calculate air photo scale at a particular point by solving the equation $S_p = f / (H-h)$, where f is the focal length of the camera, H is the flying height of the aircraft above mean sea level, and h is the elevation of the terrain at a given point. You'll have a chance to practice calculating both map scale and air photo scale in a forthcoming practice quiz.

2.6. Graphic Map Scales

Another way to express map scale is with a graphic (or “bar”) scale. Unlike representative fractions, graphic scales remain true when maps are shrunk or magnified.



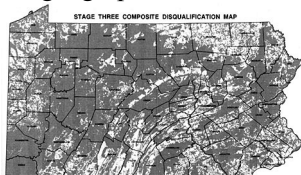
Graphic scales.

If they include a scale at all, most maps include a bar scale like the one shown above left. Some also express map scale as a representative fraction. Either way, the implication is that scale is uniform across the map. In fact, except for maps that show only very small areas, scale varies across every map. As you probably know, this follows from the fact that positions on the nearly-spherical Earth must be transformed to positions on two-dimensional sheets of paper. Systematic transformations of this kind are called **map projections**. As we will discuss in greater depth later in this chapter, all map projections are accompanied by deformation of features in some or all areas of the map. This deformation causes map scale to vary across the map. Representative fractions may therefore specify map scale along a line at which deformation is minimal (**nominal scale**). Bar scales denote only the nominal or average map scale. Variable scales, like the one illustrated above right, show how scale varies, in this case by latitude, due to deformation caused by map projection.

2.7. Map Scale and Accuracy

One of the special characteristics of geographic data is that phenomena shown on maps tend to be represented differently at different scales. Typically, as scale decreases, so too does the number of different features, and the detail with which they are represented. Not only printed maps, but also digital geographic data sets that cover extensive areas tend to be more **generalized** than data sets that cover limited areas.

Accuracy also tends to vary in proportion with map scale. The United States Geological Survey, for example, guarantees that the mapped positions of 90 percent of well-defined points shown on its topographic map series at scales smaller than 1:20,000 will be within 0.02 inches of their actual positions on the map (see the [National Geospatial Program Standards and Specifications](#)). Notice that this “National Map Accuracy Standard” is scale-dependent. The allowable error of well-defined points (such as control points, road intersections, and such) on 1:250,000 scale topographic maps is thus $1 / 250,000 = 0.02 \text{ inches} / D_g$ or $D_g = 0.02 \text{ inches} \times 250,000 = 5,000 \text{ inches}$ or 416.67 feet. Neither small-scale maps nor the digital data derived from them are reliable sources of detailed geographic information.



Areas (in gray) disqualified as potential sites for a low level radioactive waste storage facility depicted on a small scale map (original 1:1,500,000) mask small suitable areas large enough to contain the 500-acre facility (Chem-Nuclear Systems, Inc., 1994).

Sometimes the detail lost on small-scale maps causes serious problems. For example, a contractor hired to

use GIS to find a suitable site for a low level radioactive waste storage facility in Pennsylvania presented a series of 1:1,500,000 scale maps at public hearings around the state in the early 1990s. The scale was chosen so that disqualified areas of the entire state could be printed on a single 11 x 17 inch page. A report accompanying the map included the disclaimer that “it is possible that small areas of sufficient size for the LLRW disposal facility site may exist within regions that appear disqualified on the [map]. The detailed information for these small areas is retained within the GIS even though they are not visually illustrated...” (*Chem-Nuclear Systems, Inc. 1993, p. 20*). Unfortunately for the contractor, alert citizens recognized the shortcomings of the small-scale map, and newspapers published reports accusing the out-of-state company of providing inaccurate documents. Subsequent maps were produced at a scale large enough to discern 500-acre suitable areas.

2.8. Scale as a Verb

The term “scale” is sometimes used as a verb. To scale a map is to reproduce it at a different size. For instance, if you photographically reduce a 1:100,000-scale map to 50 percent of its original width and height, the result would be one-quarter the area of the original. Obviously the map scale of the reduction would be smaller too: $1/2 \times 1/100,000 = 1/200,000$.

Because of the inaccuracies inherent in all geographic data, particularly in small scale maps, scrupulous geographic information specialists **avoid enlarging source maps**. To do so is to exaggerate generalizations and errors. The original map used to illustrate areas in Pennsylvania disqualified from consideration for low-level radioactive waste storage shown on an earlier page, for instance, was printed with the statement “Because of map scale and printing considerations, it is not appropriate to enlarge or otherwise enhance the features on this map.”

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Map Scale.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

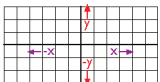
2.9. Geospatial Measurement Scales

The word “scale” can also be used as a synonym for a ruler—a measurement scale. Because data consist of symbols that represent measurements of phenomena, it’s important to understand the reference systems used to take the measurements in the first place. In this section we’ll consider a measurement scale known as the **geographic coordinate system** that is used to specify positions on the Earth’s roughly spherical surface. In other sections we’ll encounter two-dimensional (plane) coordinate systems, as well as the measurement scales used to specify attribute data.

In this section of Chapter 2 you will:

1. Demonstrate your ability to specify geospatial locations using geographic coordinates.
2. Convert geographic coordinates between two different formats.

2.10. Coordinate Systems



A Cartesian coordinate system.

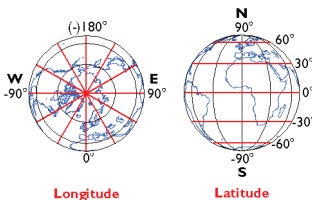
As you probably know, locations on the Earth's surface are measured and represented in terms of **coordinates**. A coordinate is a set of two or more numbers that specifies the position of a point, line, or other geometric figure in relation to some reference system. The simplest system of this kind is a Cartesian coordinate system (named for the 17th century mathematician and philosopher René Descartes). A Cartesian coordinate system is simply a grid formed by juxtaposing two measurement scales, one horizontal (x) and one vertical (y). The point at which both x and y equal zero is called the **origin** of the coordinate system. In the illustration above, the origin (0,0) is located at the center of the grid. All other positions are specified relative to the origin. The coordinate of upper right-hand corner of the grid is (6,3). The lower left-hand corner is (-6,-3). If this is not clear, please ask for clarification!

Cartesian and other two-dimensional (plane) coordinate systems are handy due to their simplicity. For obvious reasons they are not perfectly suited to specifying geospatial positions, however. The **geographic coordinate system** is designed specifically to define positions on the Earth's roughly-spherical surface. Instead of the two linear measurement scales x and y, the geographic coordinate systems juxtaposes two curved measurement scales. The east-west scale, called **longitude** (conventionally designated by the Greek symbol *lambda*), ranges from $+180^\circ$ to -180° . Because the Earth is round, $+180^\circ$ (or 180° E) and -180° (or 180° W) are the same grid line. That grid line is roughly the International Date Line, which has diversions that pass around some territories and island groups. Opposite the International Date Line is the **prime meridian**, the line of longitude defined by international treaty as 0° . The north-south scale, called **latitude** (designated by the Greek symbol *phi*), ranges from $+90^\circ$ (or 90° N) at the North pole to -90° (or 90° S) at the South pole. We'll take a closer look at the geographic coordinate system next.



The geographic (or “geodetic”) coordinate system.

2.11. Geographic Coordinate System



The geographic coordinate system.

Longitude specifies positions east and west as the angle between the **prime meridian** and a second **meridian** that intersects the point of interest. Longitude ranges from $+180^\circ$ (or 180° E) to -180° (or 180° W). 180° East and West longitude together form the International Date Line.

Latitude specifies positions north and south in terms of the angle subtended at the center of the Earth between two imaginary lines, one that intersects the **equator** and another that intersects the point of interest. Latitude ranges from $+90^\circ$ (or 90° N) at the North pole to -90° (or 90° S) at the South pole. A line of latitude is also known as a **parallel**.

At higher latitudes, the length of parallels decreases to zero at 90° North and South. Lines of longitude are not parallel, but converge toward the poles. Thus while a degree of longitude at the equator is equal to a distance of about 111 kilometers, that distance decreases to zero at the poles.

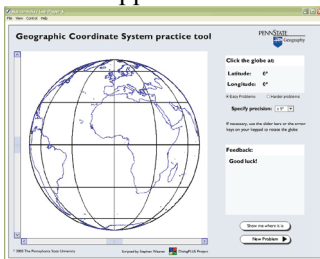
TRY THIS!**GEOGRAPHIC COORDINATE SYSTEM PRACTICE APPLICATION**

Nearly everyone learned latitude and longitude as a kid. But how well do you understand the geographic coordinate system, really? My experience is that while everyone who enters this class has heard of latitude and longitude, only about half can point to the location on a map that is specified by a pair of geographic coordinates. The Flash application linked below lets you test your knowledge. The application asks you to click locations on a globe as specified by randomly generated geographic coordinates.

You will notice that the application lets you choose between “easy problems” and “hard problems.” Easy problems are those in which latitude and longitude coordinates are specified in 30° increments. Since the resolution of the graticule (the geographic coordinate system grid) used in the application is also 30° , the solution to every “easy” problem occurs at the intersection of a parallel and a meridian. The “easy” problems are good warm-ups.

“Hard” problems specify coordinates in 1° increments. You have to interpolate positions between grid lines. **You can consider yourself to have a good working knowledge of the geographic coordinate system if you can solve at least six “hard” problems consecutively and on the first click.**

[Click here to download and launch the Geographic Coordinate System practice application \(5.7 Mb\).](#) (If the globe doesn't appear after the Flash application has loaded, right-click and select “Play” from the pop-up menu.)



Note: You will need to have the Adobe Flash player installed in order to complete this exercise. If you do not already have the Flash player, you can download it for free at the [adobe website](#).

2.12. Geographic Coordinate Formats

Geographic coordinates may be expressed in decimal degrees, or in degrees, minutes, and seconds. Sometimes you need to convert from one form to another. Steve Kiouttis (personal communication, Spring 2002), manager of the Pennsylvania Urban Search and Rescue Program, described one such situation on the course Bulletin Board: “I happened to be in the state Emergency Operations Center in Harrisburg on Wednesday evening when a call came in from the Air Force Rescue Coordination Center in Dover, DE. They had an emergency locator transmitter (ELT) activation and requested the PA Civil Air Patrol to investigate. The coordinates given to the watch officer were 39 52.5 n and -75 15.5 w. This was plotted incorrectly (treated as if the coordinates were in decimal degrees 39.525n and -75.155 w) and the location appeared to be near Vineland, New Jersey. I realized that it should have been interpreted as 39 degrees 52 minutes and 5 seconds n and -75 degrees and 15 minutes and 5 seconds w) and made the conversion (as we were taught in Chapter 2) and came up with a location on the grounds of Philadelphia International Airport, which is where the locator was found, in a parked airliner.”

Here's how it works:

To convert -89.40062 from decimal degrees to degrees, minutes, seconds:

1. Subtract the number of whole degrees (89°) from the total (89.40062°). (The minus sign is used in the decimal degree format only to indicate that the value is a west longitude or a south latitude.)

2. Multiply the remainder by 60 minutes ($.40062 \times 60 = 24.0372$).
3. Subtract the number of whole minutes (24') from the product.
4. Multiply the remainder by 60 seconds ($.0372 \times 60 = 2.232$).
5. The result is $89^\circ 24' 2.232''$ W or S.

To convert $43^\circ 4' 31''$ from degrees, minutes, seconds to decimal degrees:
 $DD = \text{Degrees} + (\text{Minutes}/60) + (\text{Seconds}/3600)$

1. Divide the number of seconds by 60 ($31 \div 60 = 0.5166$).
2. Add the quotient of step (1) to the whole number of minutes ($4 + 0.5166$).
3. Divide the result of step (2) by 60 ($4.5166 \div 60 = 0.0753$).
4. Add the quotient of step (3) to the number of whole number degrees ($43 + 0.0753$).
5. The result is 43.0753°

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about the Geographic Coordinate System.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.13. Horizontal Datums

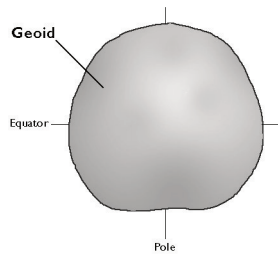
Geographic data represent the locations and attributes of things on the Earth's surface. Locations are measured and encoded in terms of geographic coordinates (i.e., latitude and longitude) or plane coordinates (e.g., UTM). To measure and specify coordinates accurately, one first must define the geometry of the surface itself. To see what I mean, imagine a soccer ball. If you or your kids play soccer you can probably conjure up a vision of a round mosaic of 20 hexagonal (six sided) and 12 pentagonal (five sided) panels (soccer balls come in many different designs, but the 32-panel ball is used in most professional matches. Visit soccerballworld.com for more than you ever wanted to know about soccer balls). Now focus on one point at an intersection of three panels. You could use spherical (e.g., geographic) coordinates to specify the position of that point. But if you deflate the ball, the position of the point in space changes, and so must its coordinates. **The absolute (though not the relative) position of a point on a surface, then, depends upon the shape of the surface.**

Every position is determined in relation to at least one other position. Coordinates, for example, are defined relative to the origin of the coordinate system grid. A land surveyor measures the "corners" of a property boundary relative to a previously-surveyed control point. Surveyors and engineers measure elevations at construction sites and elsewhere. Elevations are expressed in relation to a **vertical datum**, a reference surface such as mean sea level. As you probably know there is also such a thing as a **horizontal datum**, although this is harder to explain and to visualize than the vertical case. **Horizontal datums define the geometric relationship between a coordinate system grid and the Earth's surface.** Because the Earth's shape is complex, the relationship is too. The goal of this section is to explain the relationship.

Specifically, in this section of Chapter 2 you will learn to:

1. Explain the concept of a horizontal datum
2. Calculate the change in a coordinate location due to a change from one horizontal datum to another
3. Estimate the magnitude of "datum shift" associated with the adjustment from NAD 27 to NAD 83

2.14. Geoids



The Earth's shape is defined as a surface that closely approximates global mean sea level, but across which gravity is everywhere equal. The caricature of the geoid shown above is not drawn to scale. Irregularities are greatly exaggerated

(Adapted from Smith, 1988).

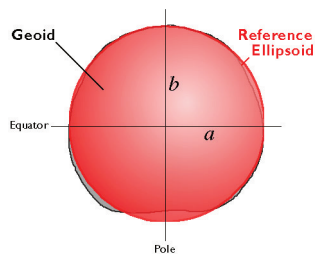
The accuracy of coordinates that specify geographic locations depends upon how the coordinate system grid is aligned with the Earth's surface. Unfortunately for those who need accurate geographic data, defining the shape of the Earth's surface is a non-trivial problem. So complex is the problem that an entire profession, called **geodesy**, has arisen to deal with it.

Geodesists define the Earth's surface as **a surface that closely approximates global mean sea level, but across which gravity is everywhere equal**. They refer to this shape as the **geoid**. Geoids are lumpy because gravity varies from place to place in response to local differences in terrain and variations in the density of materials in the Earth's interior. Geoids are also a little squat. Sea level gravity at the poles is greater than sea level gravity at the equator, a consequence of Earth's "oblate" shape as well as the centrifugal force associated with its rotation.

Geodesists at the [U.S. National Geodetic Survey](#) describe the geoid as an "equipotential surface" because the potential energy associated with the Earth's gravitational pull is equivalent everywhere on the surface. Like fitting a trend line through a cluster of data points, the geoid is a three-dimensional statistical surface that fits as closely as possible gravity measurements taken at millions of locations around the world. As additional and more accurate gravity measurements become available, geodesists revise the geoid periodically. Some geoid models are solved only for limited areas; GEOID03, for instance, is calculated only for the continental U.S.

Recall that horizontal datums define how coordinate system grids align with the Earth's surface. Long before geodesists calculated geoids, surveyors used much simpler surrogates called **ellipsoids** to model the shape of the Earth.

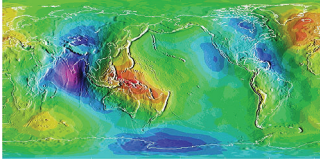
2.15. Ellipsoids



Ellipsoids approximate the geoid (Adapted from Smith, 1988).

An ellipsoid is a three-dimensional geometric figure that resembles a sphere, but whose equatorial axis (a is the illustration above) is slightly longer than its polar axis (b). The equatorial axis of the [World Geodetic System](#) of 1984, for instance, is approximately 22 kilometers longer than the polar axis, a proportion that closely resembles the oblate spheroid that is planet Earth. **Ellipsoids are commonly used as surrogates for geoids so as to simplify**

the mathematics involved in relating a coordinate system grid with a model of the Earth's shape. Ellipsoids are good, but not perfect, approximations of geoids. The map below shows differences in elevation between a geoid model called GEOID96 and the WGS84 ellipsoid. The surface of GEOID96 rises up to 75 meters above the WGS84 ellipsoid over New Guinea (where the map is colored red). In the Indian Ocean (where the map is colored purple), the surface of GEOID96 falls about 104 meters below the ellipsoid surface.



Deviations between an ellipsoid and a geoid (National Geodetic Survey, 1997).

Many ellipsoids are in use around the world. (Peter Dana has published a list at colorado.edu) Local ellipsoids minimize differences between the geoid and the ellipsoid for individual countries or continents. The Clarke 1866 ellipsoid, for example, minimizes deviations in North America. **The North American Datum of 1927 (NAD 27)** associates the geographic coordinate grid with the Clarke 1866 ellipsoid. NAD 27 involved an adjustment of the latitude and longitude coordinates of some 25,000 geodetic control point locations across the U.S. The nationwide adjustment commenced from an initial control point at Meades Ranch, Kansas, and was meant to reconcile discrepancies among the many local and regional control surveys that preceded it.

The North American Datum of 1983 (NAD 83) involved another nationwide adjustment, necessitated in part by the adoption of a new ellipsoid, called **GRS 80**. Unlike Clarke 1866, GRS 80 is a global ellipsoid centered upon the Earth's center of mass. GRS 80 is essentially equivalent to WGS 84, the global ellipsoid upon which the Global Positioning System is based. **NAD 27 and NAD 83 both align coordinate system grids with ellipsoids. They differ simply in that they refer to different ellipsoids.** Because Clarke 1866 and GRS 80 differ slightly in shape as well as in the positions of their center points, the adjustment from NAD 27 to NAD 83 involved a shift in the geographic coordinate grid. Because a variety of datums remain in use, geospatial professionals need to understand this shift, as well as how to transform data between horizontal datums.

2.16. Control Points and Datum Shifts



In the U.S., high-order horizontal control point locations are marked with permanent metal “monuments” like the one shown above. The physical manifestation of datum is a network of control point measurements (National Geodetic Survey, 2004).

Geoids, ellipsoids, and even coordinate systems are all abstractions. The fact that “horizontal datum” refers to a relationship between an ellipsoid and a coordinate system, two abstractions, may explain why the concept is so frequently misunderstood. Datums do have physical manifestations, however.

Shown above is one of the approximately two million horizontal and vertical control points that have been established in the U.S. Although control point markers are fixed, the coordinates that specify their locations are liable to change. The [U.S. National Geodetic Survey](http://www.ngs.noaa.gov) maintains a database of the coordinate specifications of these control points, including historical locations as well as more recent adjustments. One occasion for adjusting control point coordinates is when new horizontal datums are adopted. **Since every coordinate system grid is aligned with an ellipsoid that approximates the Earth's shape, coordinate grids necessarily shift when one ellipsoid is**

replaced by another. When coordinate system grids shift, the coordinates associated with fixed control points need to be adjusted. How we account for the Earth’s shape makes a difference in how we specify locations.

TRY THIS!

Here’s a chance to calculate how much the coordinates of a control point change in response to an adjustment from North American Datum of 1927 (based on the Clarke 1866 ellipsoid) to the North American Datum of 1983 (based upon the GRS 80 ellipsoid). You’ll be asked to interpret your results in an upcoming practice quiz.

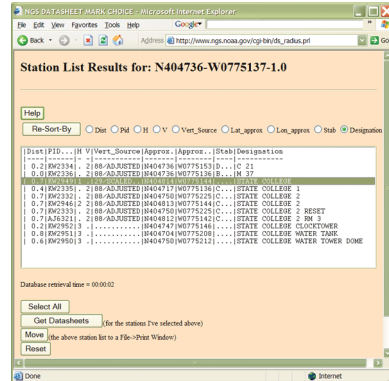
1. Find the geographic coordinates of a populated place

- a. Start at the USGS’ Geographic Names Information System at the [U.S. Board on Geographic Names](#)
- b. Follow the links labeled **Domestic Names**, then **Search** to search place names included in the Geographic Names Information System.
- c. At the Query Form, enter the name of your home town (or other named geographic feature) in the **Feature Name** field, as well as your home **State**. Choose **Populated Place** (or other, as appropriate) for the **Feature Class**.
 - i. If your home is somewhere other than the U.S., enter a place name of interest or fantasy destination (e.g., “Las Vegas” ;-).
- d. Click **Send Query**.
- e. The result should include latitude and longitude coordinates of a **centroid** that represents where the name your town (or other feature) would appear on a map. You’ll need those coordinates for the next step.

2. Find the geographic coordinates of a nearby horizontal control point

- a. Visit the [U.S. National Geodetic Survey home page](#)
- b. Follow the link labeled **Survey Mark Datasheets**
- c. At the NGS Datasheet page, follow the link labeled **Datasheets**.
 - i. You may wish to begin with the “Info Link” labeled “Tell me more about datasheets”
- d. At the NGS Datasheet Retrieval page, follow the link labeled **Radial Search**. (You’re welcome to experiment with another retrieval method if you wish.)
- e. At the NGS Datasheet Point Radius form:
 - i. Enter the latitude and longitude coordinates you looked up in step #1. Pay attention to the input format.
 - i. Specify a **Search Radius**.
 - i. Select **Any Horz. and/or Vert. Control** from the Data Type Desired scrolling field.
 - i. Select **Any Stability** from the Stability Desired scrolling field.
 - i. Click **Submit**.
- f. The result should be a **Station List Results** form that looks like the contents of the window pictured below. These are the results of my search on the centroid coordinates for State College PA. Note that I have highlighted the station that is both nearest to the coordinates I entered and

a first-order control point (see the “1” under the column labeled “H”?)



g. Select the station nearest to the coordinates you specified that is also the highest-order horizontal control point.

h. Click **Get Datasheets**. The system should respond with a [station datasheet like this example](#).

i. In the example linked above, the CURRENT SURVEY CONTROL of the station point is listed as **NAD 83(1992) 40 48 13.83840(N) 077 51 44.25410(W) ADJUSTED**. These are the geographic coordinates of the control point relative to the NAD 83 horizontal datum. In the next step we'll see how much the control point “moved” as a result of the adjustment of those coordinates from the earlier NAD 27 datum. (The geographic coordinates of the control point are specified to 100,000th of a second precision, or approximately 0.3 mm of longitude. Keep in mind, however, the difference between precision and accuracy; the trailing 0 suggests that the accuracy is an order of magnitude less than the precision.)

3. Calculate the datum shift associated with a conversion from one horizontal datum to another

a. Return to the [U.S. National Geodetic Survey home page](#)

b. Follow the link labeled **geodetic tool kit**.

c. At the NGS Geodetic Tool Kit page, follow the link labeled **NADCON** (you'll be taken to an explanatory page, where you'll need to click **NADCON** again to proceed to the utility).

d. At the North American Datum Conversion Utility page, read the introductory paragraphs, then follow the link labeled **Interactively compute a datum shift between NAD 27 and NAD 83**. The link referred to in the previous sentence has recently been removed. Instead click on the **here** link in the *****Notice...** at the top of the page.

e. At the NADCON computations form, under the heading **compute a datum shift for a specific location**:

- i. Select direction of conversion: **NAD 83 to NAD27**
- i. Enter the NAD 83 latitude and longitude coordinates of your control station. Pay attention to format.
- i. Click **Compute Datum Shift for a Single Location**.

f. The result should be a [NADCON Output report like this example](#). In the State College example, the adjustment from NAD 83 to NAD 27 (associated with the replacement of the old Clarke 1866 ellipsoid by the Earth-centered GRS 80 ellipsoid, caused the geographic coordinate system grid to shift nearly 7 meters South and over 23 meters West. That grid shift is reflected

in the adjustment of the coordinates that specify the control point's location. Note that the point didn't move, rather, the grid shifted. How much shift occurred at your location?

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about horizontal datums.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.17. Coordinate Transformations

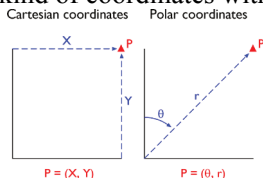
GIS specialists often need to transform data from one coordinate system and/or datum to another. For example, digital data produced by tracing paper maps over a digitizing tablet need to be transformed from the tablet's non-georeferenced plane coordinate system into a georeferenced plane or spherical coordinate system that can be **georegistered** with other digital data "layers." Raw image data produced by scanning the Earth's surface from space tend to be skewed geometrically as a result of satellite orbits and other factors; to be useful these too need to be transformed into georeferenced coordinate systems. Even the point data produced by GPS receivers, which are measured as latitude and longitude coordinates based upon the WGS84 datum, often need to be transformed to other coordinate systems or datums to match project specifications. **This section describes three categories of coordinate transformations: (1) plane coordinate transformations; (2) datum transformations; and (3) map projections.**

Students who successfully complete this section of Chapter 2 should be able to:

1. Recognize the kind of transformation that is appropriate to georegister two or more data sets.

2.18. Plane Coordinate Transformations

Some coordinate transformations are simple. For example, the transformation from non-georeferenced plane coordinates to non-georeferenced polar coordinates shown below involves nothing more than the replacement of one kind of coordinates with another.



The same position specified within two non-georeferenced plane coordinate systems: Cartesian (left) and polar (right) (Adapted from Iliffe, 2000).

Unfortunately, most plane coordinate transformation problems are not so simple. The geometries of non-georeferenced plane coordinate systems and georeferenced plane coordinate systems tend to be quite different, mainly because georeferenced plane coordinate systems are often projected. As you know, the act of projecting a nearly-spherical surface onto a two-dimensional plane necessarily distorts the geometry of the original spherical surface. Specifically, the scale of a projected map (or an unrectified aerial photograph, for that matter) varies from place to place. So long as the geographic area of interest is not too large, however, formulae like the ones described here can be effective in transforming a non-georeferenced plane coordinate system grid to match a georeferenced plane coordinate system grid with reasonable, and measurable, accuracy. We won't go into the math

of the transformations here, since the formulae are implemented within GIS software. Instead, this section aims to familiarize you with how some common transformations work and how they may be used.

SIMILARITY TRANSFORMATION

In the hypothetical illustration below, the spatial arrangement of six control points digitized from a paper map (“before”) are shown to differ from the spatial arrangement of the same points that appear in a georeferenced aerial photograph that referenced to a different plane coordinate system grid (“after”). If, as shown, the arrangement of the two sets of points differs only in scale, rotation, and offset, a relatively simple four-parameter **similarity transformation** may do the trick. Your GIS software should derive the parameters for you by comparing the relative positions of the common points. Note that while only six control points are illustrated, ten to twenty control points are recommended (Chrisman 2002).



Six control point locations before and after a similarity transformation used to correct systematic differences in scale, rotation, and offset between two plane coordinate systems.

TRY THIS!

Click the graphic above to view a Flash animation (transform_sim.swf) in a separate browser window.

Note: You will need to have the Adobe Flash player installed to complete this activity. If you have not already installed the Flash player, you can [download it for free from Adobe](#).

AFFINE TRANSFORMATION

Sometimes a similarity transformation doesn’t do the trick. For example, because paper maps expand and contract more along the paper grain than across the grain in response to changes in humidity, the scale of a paper map is likely to be slightly greater along one axis than the other. In such cases a six-parameter **affine transformation** may be used to accommodate differences in scale, rotation, and offset along each of the two dimensions of the source and target coordinate systems. This characteristic is particularly useful for transforming image data scanned from polar-orbiting satellites whose orbits trace S-shaped paths over the rotating Earth.



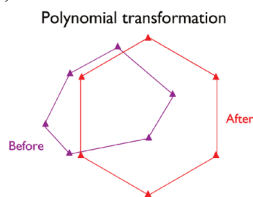
Six control point locations before and after an affine transformation used to correct systematic differences in scale, rotation, and offset between two plane coordinate systems. Notice that the arrangement of points before the transformation is skewed as well as offset and rotated.

TRY THIS!

Click the graphic above to view a Flash animation (transform_aff.swf) in a separate browser window.

SECOND-ORDER POLYNOMIAL TRANSFORMATION

When neither similarity nor affine transformations yield acceptable results, you may have to resort to a twelve-parameter **Second-order polynomial transformation**. Their advantage is the potential to correct data sets that are distorted in several ways at once. A disadvantage is that the stability of the results depend very much upon the quantity and arrangement of control points and the degree of dissimilarity of the source and target geometries (Iliffe 2000).



Six control point locations before and after a second-order polynomial transformation. Notice that the arrangement of points before the transformation is distorted in multiple ways in comparison with the corrected arrangement.

TRY THIS!

Click the graphic above to view a Flash animation (transform_poly.swf) in a separate browser window.

Even more elaborate plane transformation methods, known collectively as **rubber sheeting**, optimize the fit of a source data set to the geometry of a target data set as if the source data were mapped onto a stretchable sheet.

ROOT MEAN SQUARE ERROR

GIS software provides a statistical measure of how well a set of transformed control points match the positions of the same points in a target data set. Put simply, Root Mean Square (RMS) Error is the average of the distances (also known as *residuals*) between each pair of control points. What constitutes an acceptably low RMS Error depends on the nature of the project and the scale of analysis.

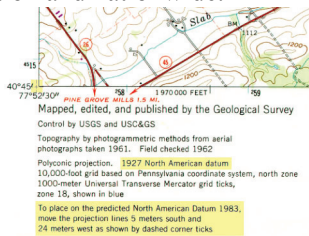
2.19. Datum Transformations

Point locations are specified in terms of (a) their positions relative to some coordinate system grid and (b) their heights above or below some reference surface. Obviously the elevation of a stationary point depends upon the size and shape of the reference surface (e.g., mean sea level) upon which the elevation measurement is based. In the same way, a point's position in a coordinate system grid depends on the size and shape of the surface upon which the grid is draped. **The relationship between a grid and a model of the Earth's surface is called a horizontal datum.** GIS specialists who are called upon to merge data sets produced at different times and in different parts of the world need to be knowledgeable about datum transformations.

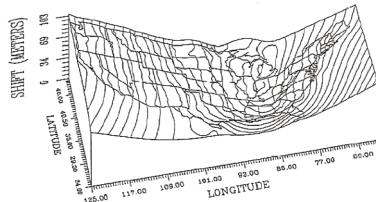
NAD 27 to NAD 83

In the U.S. the two most frequently encountered horizontal datums are the North American Datum of 1927 (NAD 27) and the North American Datum of 1983 (NAD 83). The advent of the Global Positioning System necessitated an update of NAD 27 that included (a) adoption of a geocentric ellipsoid, GRS 80, in place of the Clarke

1866 ellipsoid; and (b) correction of many distortions that had accrued in the older datum. Bearing in mind that the realization of a datum is a network of fixed control point locations that have been specified in relation to the same reference surface, **the 1983 adjustment of the North American Datum caused the coordinate values of every control point managed by the National Geodetic Survey (NGS) to change.** Obviously the points themselves did not shift on account of the datum transformation (although they did move a centimeter or more a year due to plate tectonics). Rather, **the coordinate system grids based upon the datum shifted in relation to the new ellipsoid.** And because local distortions were adjusted at the same time, the magnitude of grid shift varies from place to place. The illustrations below compare the magnitude of the grid shifts associated with the NAD 83 adjustment at one location and nationwide.



A corner of the 1:24,000 scale topographic quadrangle map for State College PA showing the magnitude of grid shift associated with the NAD 83 adjustment. The map is based on NAD 27, but was reprinted with revisions in 1987, including the statement that coordinate system grid lines shift 24 meters west and 5 meters south if NAD 83 coordinates are used instead of NAD 27.



Magnitude of grid shift associated with the NAD 83 adjustment for the continental 48 U.S. states. Shifts range from 10 to 100 meters in the lower 48 (least in upper Midwest states) to over 200 meters in Alaska, and over 400 meters in Hawaii (Dewhurst 1990).

Given the irregularity of the shift, NGS could not suggest a simple transformation algorithm that surveyors and mappers could use to adjust local data based upon the older datum. Instead NGS created a software program called NADCON (Dewhurst 1990, Mulcare 2004) that calculates adjusted coordinates from user-specified input coordinates by interpolation from a pair of 15° correction grids generated by NGS from hundreds of thousands of previously-adjusted control points.

TRY THIS!

Try out the [National Geodetic Survey's NADCON tool](#).

GPS DATA AND WGS 84

The U.S. Department of Defense created the Global Positioning System (GPS) over a period of 16 years at a startup cost of about \$10 billion. GPS receivers calculate their positions in terms of latitude, longitude, and height above or below the World Geodetic System of 1984 ellipsoid (WGS 84). Developed specifically for the Global Positioning System, WGS 84 is an Earth-centered ellipsoid which, unlike the many regional, national, and local ellipsoids still in use, minimizes deviations from the geoid worldwide. Depending on where a GIS specialist may be working, or what data he or she may need to work with, the need to transform GPS data from WGS 84 to some other datum is likely to arise. Datum transformation algorithms are implemented within GIS software as well as in the post-

processing software provided by GPS vendors for use with their receivers. Some transformation algorithms yield more accurate results than others. The method you choose will depend on what choices are available to you and how much accuracy your application requires.

Unlike the plane transformations described earlier, datum transformations involve ellipsoids and are therefore three-dimensional. The simplest is the three-parameter **Molodenski transformation**. In addition to knowledge of the size and shape of the source and target ellipsoids (specified in terms of **semimajor axis**, the distance from the ellipsoid's equator to its center, and **flattening ratio**, the degree to which the ellipsoid is flattened to approximate the Earth's oblate shape), the offset between the two ellipsoids needs to be specified along X, Y, and Z axes. The window shown below illustrates ellipsoidal and offset parameters for several horizontal datums, all expressed in relation to WGS 84.

Datum	Ellipsoid	Delta A	Delta (1/f)x10e4	Delta x	Delta y	Delta z
NAD 27	Clarke 1866	83.4	0.2344503	0	160	176
NAD 83	Clarke 1866	0	0	0	0	0
WGS 84	WGS 84	0	0	0	0	0
Observatio 1966	International	-251	0.14192702	-425	-169	81
Old Egyptian	Hevelius 1905	43	0.00400796	-130	110	-13
Old Hawaiian Kaula	Clarke 1866	83.4	0.23264639	45	-290	-172
Old Hawaiian Maui	Clarke 1866	83.4	0.23264639	65	-290	-180
Old Hawaiian Oahu	Clarke 1866	83.4	0.23264639	56	-294	-181
Old Hawaiian Hawaii	Clarke 1866	83.4	0.23264639	69	-279	-183
Oran	Clarke 1880	-112.145	0.54750714	-346	-1	224
Ordinance Survey of Gr.	Airy	573.804	0.11800222	375	-111	431
Pico de Las Neveas	International	-251	0.14192702	-307	52	127
Pitcairn Auto 1967	International	-251	0.14192702	185	165	42
Provisional South Ameri.	International	-251	0.14192702	-388	175	-395
Provisional South Chiles.	International	-251	0.14192702	16	195	33
Puerto Rico	Clarke 1866	83.4	0.23264639	11	72	-101
Qomona	International	-251	0.14192702	164	138	-189
Quincy National	International	-251	0.14192702	-120	-283	22
Rome 1940	International	-251	0.14192702	-225	85	9
S.E. Asia (Indon)	Everest	860.895	0.28381388	173	750	264
SAD 83 (Brazil)	South American 1...	-23	0.00891204	40	-2	41
Sarawak	International	-251	0.14192702	-203	141	53
Santo (D.S.)	International	-251	0.14192702	170	42	84
Sapper HE 1943	International	-251	0.14192702	-258	15	74
Schwarzsck	Bessel 1841 (Hann.)	653.125	0.10037483	616	87	-251
Sicily	International	-251	0.14192702	97	88	-135
Sri Lanka 1960	WGS 84	0	0	0	0	0
South American 1969 m.	South American 1...	0	0	0	0	0
South Asia	Modified Fischer 1...	-18	0.00489796	7	-10	-26
Southeast Base	International	-251	0.14192702	-489	-349	314
Southwest Base	International	-251	0.14192702	-104	167	-38
Tennessee Observatory	International	-251	0.14192702	-189	-242	-81
Thailand (Indon)	Everest	860.895	0.28381388	214	836	303
Tinian 1945	Everest	860.895	0.28381388	488	691	45
Tokyo mean	Bessel 1841	728.845	0.10037483	-128	481	684
Tulain Auto 1968	International	-251	0.14192702	422	438	609
Utah-Ark. Escalante (N.)	Clarke 1880	-112.145	0.54750714	-943	156	301
Val Levis 1916	Clarke 1880	-112.145	0.54750714	51	391	-36
WGS 84	WGS 84	0	0.00000000	0	0	0
WGS 84	WGS 84	0	0	0	0	0
Wale Swetick 1960	Heugh	-120	0.14192702	101	62	-50
Yacare	International	-251	0.14192702	-155	171	27
Zandén	International	-251	0.14192702	-355	120	-359

Datum list window in the Waypoint+ software utility (Hildebrand 1997). NAD 27, NAD 83, and WGS 84 are highlighted. The ellipsoid associated with each datum is named, and its size and shape specified (Delta A and Delta (1/f)x10e4), along with three offset parameters, in meters, relative to WGS 84 (Delta x, Delta y, Delta z).

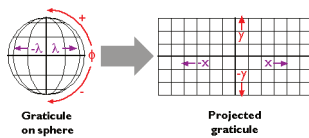
For larger study areas more accurate results may be obtained using a **seven-parameter transformation** that accounts for rotation as well as scaling and offset.

Finally, **surface-fitting transformations** like the NADCON grid interpolation described above yield the best results over the largest areas.

For routine mapping applications covering relatively small geographic areas (i.e., larger than 1:25,000), the plane transformations described earlier may yield adequate results when datum specifications are unknown and when a sufficient number of appropriately distributed control points can be identified.

2.20. Map Projections

Latitude and longitude coordinates specify positions in a more-or-less spherical grid called the **graticule**. Plane coordinates like the eastings and northings in the Universal Transverse Mercator (UTM) and State Plane Coordinates (SPC) systems denote positions in flattened grids. This is why georeferenced plane coordinates are referred to as *projected* and geographic coordinates are called **unprojected**. The mathematical equations used to transform latitude and longitude coordinates to plane coordinates are called **map projections**. Inverse projection formulae transform plane coordinates to geographic. The simplest kind of projection, illustrated below, transforms the graticule into a rectangular grid in which all grid lines are straight, intersect at right angles, and are equally spaced. More complex projections yield grids in which the lengths, shapes, and spacing of the grid lines vary.



Map projections are mathematical transformations between geographic coordinates and plane coordinates.

If you are a GIS practitioner you have probably faced the need to superimpose unprojected latitude and longitude data onto projected data, and vice versa. For instance, you might have needed to merge geographic coordinates measured with a GPS receiver with digital data published by the USGS that are encoded as UTM coordinates. Modern GIS software provides sophisticated tools for projecting and unprojecting data. To use such tools most effectively you need to understand the projection characteristics of the data sets you intend to merge. We'll examine map projections in some detail elsewhere in this lesson. Here let's simply review the characteristics that are included in the "Spatial Reference Information" section of the metadata documents that (ideally!) accompany the data sets you might wish to incorporate in your GIS. These include:

- **Projection Name** Most common in the GIS realm is the Transverse Mercator, which serves as the basis of the global UTM plane coordinate system, the U.K. and proposed U.S. National Grids, and many zones in the U.S. State Plane Coordinate system (SPC). Other SPC zones are based upon the Lambert Conic Conformal projection, which like many projections is named for its inventor as well as its projection category (conic) and the geometric properties it preserves (conformal). Much map data, particularly in the form of printed paper maps, are based upon "legacy" projections (like the Polyconic in the U.S.) that are no longer widely used. A much greater variety of projection types tend to be used in small scale thematic mapping than in large scale reference mapping.
- **Central Meridian** Although no land masses are shown, let's assume that the graticule and projected grid shown above are centered on the intersection of the equator (0 latitude) and prime meridian (0° longitude). Most map projection formulae include a parameter that allows you to center the projected map upon any longitude.
- **Latitude of Projection Origin** Under certain conditions, most map projection formulae allow you to specify different *aspects* of the grid. Instead of the *equatorial aspect* illustrated above, you might specify *apolar aspect* or *oblique aspect* by varying the latitude of projection origin such that one of the poles, or any latitude between the pole and the equator, is centered in the projected map. As you might imagine, the appearance of the grid changes a lot when viewed at different aspects.
- **Scale Factor at Central Meridian** This is the ratio of map scale along the central meridian and the scale at a *standard* meridian, where scale distortion is zero. The scale factor at the central meridian is .9996 in each of the 60 UTM coordinate system zones since each contains two standard lines 180 kilometers west and east of the central meridian. Scale distortion increases with distance from standard lines in all projected coordinate systems.
- **Standard Lines** Some projections, including the Lambert Conic Conformal, include parameters by which you can specify one or two *standard lines* along which there is no scale distortion caused by the act of transforming the spherical grid into a flat grid. By the same reasoning that two standard lines are placed in each UTM zone to minimize distortion throughout the zone to a maximum of one part in 1000, two standard parallels are placed in each SPC zone that is based upon a Lambert projection such that scale distortion is no worse than one part in 10,000 anywhere in the zone.

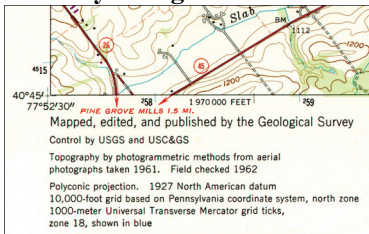
PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about coordinate transformations.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.21. UTM Coordinate System

Shown below is the southwest corner of a 1:24,000-scale topographic map published by the United States Geological Survey (USGS). Note that the geographic coordinates (40° 45' N latitude, 77° 52' 30" W longitude) of the corner are specified. Also shown, however, are ticks and labels representing two plane coordinate systems, the Universal Transverse Mercator (UTM) system and the State Plane Coordinates (SPC) system. The tick labeled “4515” represents a UTM grid line (called a “northing”) that runs parallel to, and 4,515,000 meters north of, the equator. Ticks labeled “258” and “259” represent grid lines that run perpendicular to the equator and 258,000 meters and 259,000 meters east, respectively, of the origin of the UTM Zone 18 North grid. Unlike longitude lines, UTM “eastings” are straight and do not converge upon the Earth’s poles. All of this begs the question, **Why are multiple coordinate system grids shown on the map? Why aren’t geographic coordinates sufficient?**



Southwest corner of a USGS topographic map showing grid ticks and labels for three different coordinate systems, including the UTM coordinate system. (USGS. “State College quadrangle, Pennsylvania”)

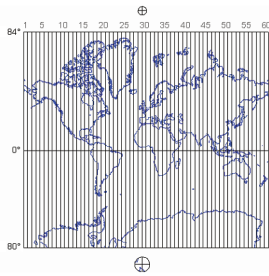
You can think of a plane coordinate system as the juxtaposition of two measurement scales. In other words, if you were to place two rulers at right angles, such that the “0” marks of the rulers aligned, you’d define a plane coordinate system. The rulers are called “axes.” The absolute location of any point in the space in the plane coordinate system is defined in terms of distance measurements along the x (east-west) and y (north-south) axes. A position defined by the coordinates (1,1) is located one unit to the right, and one unit up from the origin (0,0). The UTM grid is a widely-used type of geospatial plane coordinate system in which positions are specified as eastings (distances, in meters, east of an origin) and northings (distances north of the origin).

By contrast, the geographic coordinate system grid of latitudes and longitudes consists of two *curved* measurement scales to fit the nearly-spherical shape of the Earth. As you know, geographic coordinates are specified in degrees, minutes, and seconds of arc. **Curved grids are inconvenient to use for plotting positions on flat maps. Furthermore, calculating distances, directions and areas with spherical coordinates is cumbersome in comparison with plane coordinates.** For these reasons, cartographers and military officials in Europe and the U.S. developed the UTM coordinate system. UTM grids are now standard not only on printed topographic maps but also for the geospatial referencing of the digital data that comprise the emerging U.S. “National Map.”

In this section of Chapter 2 you will learn to:

1. Describe the characteristics of the UTM coordinate system, including its basis in the Transverse Mercator map projection; and
2. Plot UTM coordinates on a map

2.22. The UTM Grid and Transverse Mercator Projection



A Mercator projection of the world showing the 60 UTM coordinate system zones, each divided into north and south halves at the equator. Also shown are two polar coordinate systems used to specify positions beyond the northern and southern limits of the UTM system.

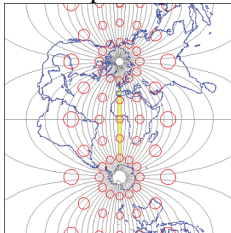
The Universal Transverse Mercator is not really universal, but it does cover nearly the entire Earth surface. Only polar areas—latitudes higher than 84° North and 80° South—are excluded. (Polar coordinate systems are used to specify positions beyond these latitudes.) The UTM system divides the remainder of the Earth’s surface into 60 zones, each spanning 6° of longitude. These are numbered west to east from 1 to 60, starting at 180° West longitude (roughly coincident with the International Date Line).

The illustration above depicts UTM zones as if they were uniformly “wide” from the Equator to their northern and southern limits. In fact, since meridians converge toward the poles on the globe, every UTM zone tapers from 666,000 meters in “width” at the Equator (where 1° of longitude is about 111 kilometers in length) to only about 70,000 meters at 84° North and about 116,000 meters at 80° South.

“Transverse Mercator” refers to the manner in which geographic coordinates are transformed into plane coordinates. Such transformations are called **map projections**. The illustration below shows the 60 UTM zones as they appear when projected using a Transverse Mercator map projection formula that is optimized for the UTM zone highlighted in yellow, Zone 30, which spans 6° West to 0° East longitude (the prime meridian).

As you can imagine, you can’t flatten a globe without breaking or tearing it somehow. Similarly, the act of mathematically transforming geographic coordinates to plane coordinates necessarily displaces most (but not all) of the transformed coordinates to some extent. Because of this, map scale varies within projected (plane) UTM coordinate system grids.

The **distortion ellipses** plotted in red help us visualize the pattern of scale distortion associated with a particular projection. Had no distortion occurred in the process of projecting the map shown below, all of the ellipses would be the same size, and circular in shape. As you can see, the ellipses centered within the highlighted UTM zone are all the same size and shape. Away from the highlighted zone the ellipses steadily increase in size, although their shapes remain uniformly circular. This pattern indicates that scale distortion is minimal within Zone 30, and that map scale increases away from that zone. Furthermore, the ellipses reveal that the character of distortion associated with this projection is that shapes of features as they appear on a globe are preserved while their relative sizes are distorted. Map projections that preserve shape by sacrificing the fidelity of sizes are called **conformal** projections. The plane coordinate systems used most widely in the U.S., UTM and SPC (the State Plane Coordinates system) are both based upon conformal projections.

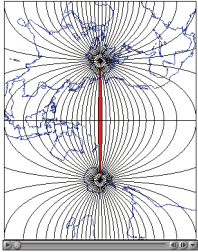


The result of a Transverse Mercator projection of the world centered on UTM Zone 30. Red circles reveal the

scale distortion introduced during the transformation from geographic to projected plane coordinates. On the globe, all the circles would be the same size.

The Transverse Mercator projection illustrated above minimizes distortion within UTM zone 30. Fifty-nine variations on this projection are used to minimize distortion in the other 59 UTM zones. In every case, distortion is no greater than 1 part in 1,000. This means that a 1,000 meter distance measured anywhere within a UTM zone will be no worse than + or – 1 meter off.

The animation linked to the illustration below shows a series of 60 Transverse Mercator projections that form the 60 zones of the UTM system. Each zone is based upon a unique Transverse Mercator map projection that minimizes distortion within that zone. Zones are numbered 1 to 60 eastward from the international date line. The animation begins with Zone 1.



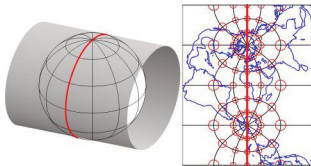
One frame of an animation showing a sequence of the 60 Transverse Mercator projections used as the basis of the UTM coordinate system. Highlighted in red is UTM Zone 01, which spans 180° W to 174° W. A unique projection is used for every UTM zone, so that deformation within each zone is minimized.

TRY THIS!

Click the graphic above to view the animation file (utm.avi, 0.5 Mb) in a separate Microsoft Media Player window.

To view the [same animation in QuickTime format \(utm.mov, 2.9 Mb\)](#), click [here](#). Requires the QuickTime plugin, which is available free at [apple.com](#).

Map projections are mathematical formulae used to transform geographic coordinates into plane coordinates. (Inverse projection formulae transform plane coordinates back into latitudes and longitudes.) “Transverse Mercator” is one of a hypothetically infinite number of such projection formulae. A visual analog to the Transverse Mercator projection appears below. Conceptually, the Transverse Mercator projection transfers positions on the globe to corresponding positions on a cylindrical surface, which is subsequently cut from end to end and flattened. In the illustration, the cylinder is tangent to the globe along one line, called the standard line. As shown in the little world map beside the globe and cylinder, scale distortion is minimal along the standard line and increases with distance from it. The animation linked above was produced by rotating the cylinder 59 times at an increment of 6°.

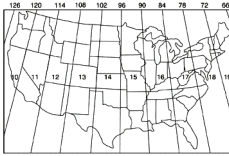


The map above represents a Transverse Mercator projection of the world with a standard meridian at 0° longitude. (Note that because of the very small size of the map, the graticule is shown at 30° resolution.) The globe wrapped in a cylinder is a conceptual model of how the Transverse Mercator projection formula transfers positions on the globe to positions on a plane (The cylinder can be flattened to a plane surface after it is unwrapped from the globe.) The thicker red line on the cylinder and the map is the **standard line** along which scale distortion is zero. As the distortion ellipses on the map indicate, distortion increases with distance from the standard line.

In the illustration above there is one standard meridian. Some projection formulae, including the Transverse Mercator projection, allow two standard lines. **Each of the 60 variations on the Transverse Mercator projection used as the foundations of the 60 UTM zones employ not one, but two, standard lines.** These two standard lines

are parallel to, and 180,000 meters east and west of, each central meridian. This scheme ensures that the maximum error associated with the projection due to scale distortion will be 1 part in 1,000 (at the outer edge of the zone at the equator). The error due to scale distortion at the central meridian is 1 part in 2,500. Distortion is zero, of course, along the standard lines.

So what does the term “transverse” mean? This simply refers to the fact that the cylinder shown above has been rotated 90° from the equatorial aspect of the standard Mercator projection, in which a single standard line coincides with 0° latitude.

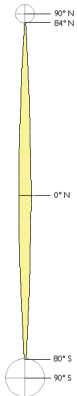


The ten UTM zones that span the conterminous U.S. (U.S. Geological Survey, 2004).

One disadvantage of the UTM system is that multiple coordinate systems must be used to account for large entities. The lower 48 United States, for instance, spreads across ten UTM zones. The fact that there are many narrow UTM zones can lead to confusion. For example, the city of Philadelphia, Pennsylvania is east of the city of Pittsburgh. If you compare the Eastings of centroids representing the two cities, however, Philadelphia’s Easting (about 486,000 meters) is less than Pittsburgh’s (about 586,000 meters). Why? Because although the cities are both located in the U.S. state of Pennsylvania, they are situated in two different UTM zones. As it happens, Philadelphia is closer to the origin of its Zone 18 than Pittsburgh is to the origin of its Zone 17. If you were to plot the points representing the two cities on a map, ignoring the fact that the two zones are two distinct coordinate systems, Philadelphia would appear to the west of Pittsburgh. Inexperienced GIS users make this mistake all the time. Fortunately, GIS software is getting sophisticated enough to recognize and merge different coordinate systems automatically.

2.23. UTM Zone Characteristics

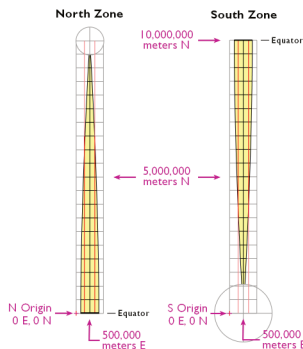
The illustration below depicts the area covered by a single UTM coordinate system grid zone. Each UTM zone spans 6° of longitude, from 84° North and 80° South. Zones taper from 666,000 meters in “width” at the Equator (where 1° of longitude is about 111 kilometers in length) to only about 70,000 meters at 84° North and about 116,000 meters at 80° South. Polar areas are covered by polar coordinate systems. Each UTM zone is subdivided along the equator into two halves, north and south.



Extent of one UTM coordinate system grid zone. Note that although latitudes are used to specify the extent precisely in relation to the globe, they are geographic, not UTM, coordinates.

The illustration below shows how UTM coordinate grids relate to the area of coverage illustrated above. The north and south halves are shown side by side for comparison. Each half is assigned its own **origin**. The north south zone origins are positioned to south and west of the zone. North zone origins are positioned on the Equator,

500,000 meters west of the central meridian. Origins are positioned so that every coordinate value within every zone is a positive number. This minimizes the chance of errors in distance and area calculations. By definition, both origins are located 500,000 meters west of the central meridian of the zone (in other words, the easting of the central meridian is always 500,000 meters E). These are considered “false” origins since they are located outside the zones to which they refer. UTM eastings range from 167,000 meters to 833,000 meters at the equator. These ranges narrow toward the poles. Northings range from 0 meters to nearly 9,400,000 in North zones and from just over 1,000,000 meters to 10,000,000 meters in South zones. Note that positions at latitudes higher than 84° North and 80° South are defined in Polar Stereographic coordinate systems that supplement the UTM system.



UTM coordinate system zone characteristics. Yellow represents areas in which UTM coordinates are valid for a given zone. Red lines parallel to the central meridian represent the two standard lines employed in each Transverse Mercator projection. Each square grid cell in the illustration spans 500,000 meters on each side.

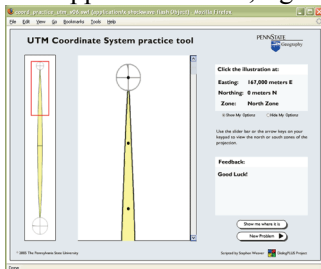
TRY THIS!

UTM COORDINATE SYSTEM PRACTICE APPLICATION

Are you ready to try your hand at positioning within a UTM coordinate system grid? The Flash application linked below lets you test your knowledge. The application asks you to click locations within a grid zone as specified by randomly generated UTM coordinates.

You will notice that the application lets you choose between easy problems (in which the locations of possible solutions are symbolized with dots) or harder problems that require you to **interpolate** solutions. **You can consider yourself to have a good working knowledge of the geographic coordinate system if you can solve at least six “hard” problems consecutively and on the first click.**

[Click here to launch the UTM Coordinate System practice application.](#) If the UTM zone fails to appear after the Flash application loads, right-click the application and select “Play” from the pop-up menu.



Note: You will need to have the Adobe Flash player installed in order to complete this exercise. If you do not already have the Flash player, you can download it for free at adobe.com

See the Bibliography (last page of the chapter) for further readings about the UTM grid system.

2.24. National Grids

The Transverse Mercator projection provides a basis for existing and proposed national grid systems in the United Kingdom and the United States.

In the U.K., topographic maps published by the [Ordnance Survey](#) refer to a national grid of 100 km squares, each of which is identified by a two-letter code. Positions within each grid square are specified in terms of eastings and northings between 0 and 100,000 meters. The U.K. national grid is a plane coordinate system that is based upon a Transverse Mercator projection whose central meridian is 2° West longitude, with standard meridians 180 km west and east of the central meridian. The grid is typically related to the Airy 1830 ellipsoid, a relationship known as the National Grid (OSGB36®) datum. The corresponding UTM zones are 29 (central meridian 9° West) and 30 (central meridian 3° West). One of the advantages of the U.K. national grid over the global UTM coordinate system is that it eliminates the boundary between the two UTM zones.

A similar system has been proposed for the U.S. by the [Federal Geographic Data Committee](#). The proposed “U.S. National Grid” is the same as the Military Grid Reference System (MGRS), a worldwide grid that is very similar to the UTM system. As Phil and Julianna Muehrcke (1998, p.p. 229-230) write in the 4th edition of *Map Use*, “the military [specifically, the U.S. Department of Defense] aimed to minimize confusion when using long numerical [UTM] coordinates” by specifying UTM zones and sub-zones with letters instead of numbers. Like the UTM system, the MGRS consists of 60 zones, each spanning 6° longitude. Each UTM zone is subdivided into 19 MGRS quadrangles of 8° latitude and one (quadrangle from 72° to 84° North) of 12° latitude. The letters C through X are used to designate the grid cell rows from south to north. I and O are omitted to avoid confusion with numbers. [Wikipedia offers a good entry on the MGRS here.](#)

TRY THIS!

INTERACTIVE DEMO OF U.K. NATIONAL GRID

An informative and fun demonstration of the U.K. National Grid is published by the [U.K. Ordnance Survey](#).

Note: You will need to have the Adobe Flash player installed in order to view this demo. If you do not already have the Flash player, you can download it for free at [adobe.com](#)

PRACTICE QUIZ

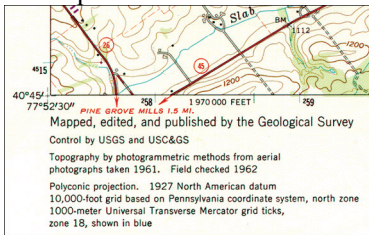
Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about UTM coordinate system.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.25. State Plane Coordinate System

Shown below is the southwest corner of a 1:24,000-scale topographic map published by the United States Geological Survey (USGS). Note that the geographic coordinates (40° 45' N latitude, 77° 52' 30" W longitude) of the corner are specified. Also shown, however, are ticks and labels representing two plane coordinate systems, the Universal Transverse Mercator (UTM) system and the State Plane Coordinate (SPC) system. The tick labeled “1 970 000 FEET” represents a SPC grid line that runs perpendicular to the equator and 1,970,000 feet east of the origin of the Pennsylvania North zone. The origin lies far to the west of this map sheet. Other SPC grid lines, called “northings” (not shown in the illustration), run parallel to the equator and perpendicular to SPC eastings at

increments of 10,000 feet. Unlike longitude lines, SPC eastings and northings are straight and do not converge upon the Earth's poles.



Southwest corner of a USGS topographic map showing grid ticks and labels for three different coordinate systems, including the SPC coordinate system. (USGS. “State College quadrangle, Pennsylvania”)

The SPC grid is a widely-used type of geospatial plane coordinate system in which positions are specified as eastings (distances east of an origin) and northings (distances north of an origin). You can tell that the SPC grid referred to in the map illustrated above is the older 1927 version of the SPC grid system because (a) eastings and northings are specified in feet and (b) grids are based upon the North American Datum of 1927 (NAD27). The 124 zones that make up the State Plane Coordinates system of 1983 are based upon NAD 83, and generally use the metric system to specify eastings and northings.

State Plane Coordinates are frequently used to georeference large scale (small area) surveying and mapping projects because plane coordinates are easier to use than latitudes and longitudes for calculating distances and areas. And because SPC zones extend over relatively smaller areas, less error accrues to positions, distances, and areas calculated with State Plane Coordinates than with UTM coordinates.

In this section you will learn to:

1. Describe the characteristics of the SPC system, including map projection on which it is based; and
2. Convert geographic coordinates to SPC coordinates

2.26. The SPC Grid and Map Projections

Plane coordinate systems pretend the world is flat. Obviously, if you flatten the entire globe to a plane surface, the sizes and shapes of the land masses will be distorted, as will distances and directions between most points. If your area of interest is small enough, however, and if you flatten it cleverly, you can get away with a minimum of distortion. The basic design problem that confronted the geodesists who designed the State Plane Coordinate System, then, was to establish coordinate system zones that were small enough to minimize distortion to an acceptable level, but large enough to be useful.

The State Plane Coordinate System of 1983 (SPC) is made up of 124 zones that cover the 50 U.S. states. As shown below, some states are covered with a single zone while others are divided into multiple zones. Each zone is based upon a unique map projection that minimizes distortion in that zone to 1 part in 10,000 or better. In other words, a distance measurement of 10,000 meters will be at worst one meter off (not including instrument error, human error, etc.). The error rate varies across each zone, from zero along the projection's standard lines to the maximum at points farthest from the standard lines. Errors will accrue at a rate much lower than the maximum at most locations within a given SPC zone. **SPC zones achieve better accuracy than UTM zones because they cover smaller areas, and so are less susceptible to projection-related distortion.**



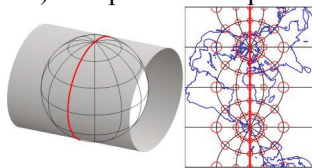
The U.S. State Plane Coordinate system of 1983 consists of 124 zones (Doyle 2004). Each zone is a distinct plane coordinate system. (Alaska and Hawaii not shown).

Most SPC zones are based on either a **Transverse Mercator** or **Lambert Conic Conformal** map projection whose parameters (such as standard line(s) and central meridians) are optimized for each particular zone. “Tall” zones like those in New York state, Illinois, and Idaho are based upon unique Transverse Mercator projections that minimize distortion by running two standard lines north-south on either side of the central meridian of each zone. “Wide” zones like those in Pennsylvania, Kansas, and California are based on unique Lambert Conformal Conic projections that run two standard parallels west-east through each zone. (One of Alaska’s zones is based upon an “oblique” variant of the Mercator projection. That means that instead of standard lines parallel to a central meridian, as in the transverse case, the Oblique Mercator runs two standard lines that are tilted so as to minimize distortion along the Alaskan panhandle.)

The two types of map projections share the property of **conformality**, which means that angles plotted in the coordinate system are equal to angles measured on the surface of the Earth. As you can imagine, conformality is a useful property for land surveyors, who make their livings measuring angles. (Surveyors measure distances too, but unfortunately there is no map projection that can preserve true distances everywhere within a plane coordinate system.) Let’s consider these two types of map projections briefly.

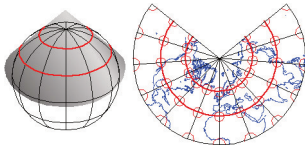
Like most map projections, the Transverse Mercator projection is actually a mathematical transformation. The illustration below may help you understand how the math works. Conceptually, the Transverse Mercator projection transfers positions on the globe to corresponding positions on a cylindrical surface, which is subsequently cut from end to end and flattened. In the illustration, the cylinder is tangent to (touches) the globe along one line, the standard line (specifically, the standard meridian). As shown in the little world map beside the globe and cylinder, scale distortion is minimal along the standard line and increases with distance from it.

The **distortion ellipses** plotted in red help us visualize the pattern of scale distortion associated with a generic Transverse Mercator projection. Had no distortion occurred in the process of projecting the map shown below, all of the ellipses would be the same size, and circular in shape. As you can see, the ellipses plotted along the central meridian are all the same size and circular shape. Away from the central meridian the ellipses steadily increase in size, although their shapes remain uniformly circular. This pattern reflects the fact that scale distortion increases with distance from the standard line. Furthermore, the ellipses reveal that the character of distortion associated with this projection is that shapes of features as they appear on a globe are preserved while their relative sizes are distorted. By preserving true angles, conformal projections like the Mercator (including its transverse and oblique variants) also preserve shapes.



Conceptual model of a Transverse Mercator map projection (left) and the resulting map (right). The thick red lines represent the line of tangency between the globe and the projection surface (the cylinder), and the corresponding standard meridian on the map. Red circles on the map reveal that distortion introduced as a result of the map projection increases with distance from the standard line. On the globe, all the circles would be the same size.

SPC zones that trend west to east (including Pennsylvania’s) are based on unique Lambert Conformal Conic projections. Instead of the cylindrical projection surface used by projections like the Mercator, the Lambert Conformal Conic and map projections like it employ conical projection surfaces like the one shown below. Notice the two lines at which the globe and the cone intersect. Both of these are standard lines; specifically, standard parallels. The latitudes of the standard parallels selected for each SPC zones minimize scale distortion throughout that zone.



Conceptual model of a Lambert Conformal Conic map projection (left) and the resulting map (right). The two thick red lines marking the intersections of the globe and the projection surface (the cone) correspond with two standard parallels on the map. Red circles on the map confirm that map scale is equal along both standard parallels. Distortion increases with distance from the standard parallels everywhere else in the projected map and in the coordinate system on which it is based.

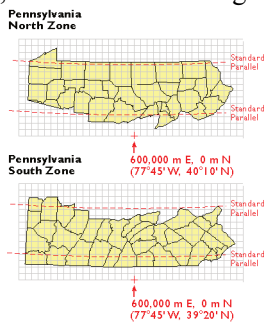
2.27. SPC Zone Characteristics

In consultation with various state agencies, the National Geodetic Survey (NGS) devised the State Plane Coordinate System with several design objectives in mind. Chief among these were:

1. **Plane coordinates** for ease of use in calculations of distances and areas;
2. **All positive values** to minimize calculation errors; and
3. A maximum error rate of **1 part in 10,000**.

Plane coordinates specify positions in flat grids. Map projections are needed to transform latitude and longitude coordinates to plane coordinates. The designers did two things to minimize the inevitable distortion associated with map projections. First, they divided the U.S. into 124 relatively small zones. Second, they used slightly different map projection formulae for each zone. The curved, dashed red lines in the illustration below represent the two standard parallels that pass through each zone. The latitudes of the standard lines are one of the parameters of the Lambert Conic Conformal projection that can be customized to minimize distortion within the zone.

Positions in any coordinate system are specified relative to an origin. **SPC zone origins are defined so as to ensure that every easting and northing in every zone are positive numbers.** As shown in the illustration below, SPC origins are positioned south of the counties included in each zone. The origins coincide with the central meridian of the map projection upon which each zone is based. The easting and northing values at the origins are not 0, 0. Instead, eastings are defined as positive values sufficiently large to ensure that every easting in the zone is also a positive number. The false origin of the Pennsylvania North zone, for instance, is defined as 600,000 meters East, 0 meters North. Origin eastings vary from zone to zone from 200,000 to 8,000,000 meters East.



Schematic view of two State Plane Coordinate System zones, showing the counties that make up each zone (in yellow), the origins of each zone, and the standard parallels of the map projections upon which the zones are based, along which scale distortion is zero.

Try This!

Investigating your local State Plane Coordinate System grid zone

In this activity you will:

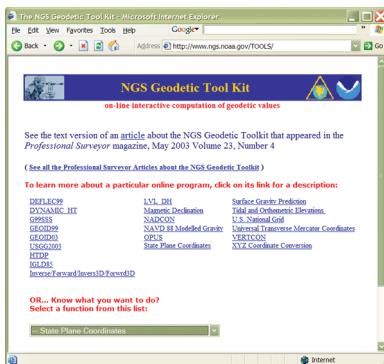
1. Read part of an authoritative manual on State Plane Coordinates;
2. Look up the designation of your local SPC zone (or a would-be zone if you are from a country other than the U.S.);
3. Investigate the parameters of the map projection upon which your local SPC zone is based; and
4. Use a Web-based tool provided by the U.S. National Geodetic Survey to convert geographic coordinates to SPC.

The practice quiz at the end of this section will help you assess your fluency with the SPC system.

1. Read the introduction to the National Geodetic Survey manual *State Plane Coordinate System of 1983* by James E. Stem (1990).

Follow [this link](#) to download the manual in Portable Document Format (PDF). Read pages 1-13 (pages 11-23 of the PDF document). Also see Appendix A, beginning on p. 62 (73), which lists map projections and other parameters of each zone.

2. Look up your local SPC zone.



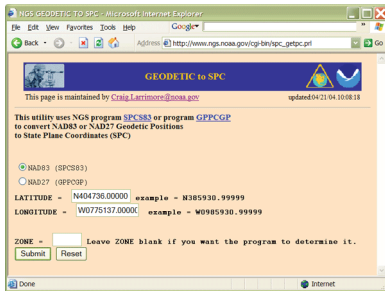
Home page of the National Geodetic Survey toolkit

1. Visit the [National Geodetic Survey's NGS Geodetic Toolkit](#). Note the various programs that NGS supports for the surveying and mapping community.
2. Use the dropdown menu to navigate to the State Plane Coordinates tool.
3. At the [State Plane Coordinates page](#), follow the Interactive Conversions link labeled "Find Zone."
4. Look up your local SPC zone (or adopted zone) by county or position.
5. You might check the result using [Rick King's list](#) or the Stem (1990) manual you downloaded earlier.

3. Look up the map projection and grid origin upon which your local SPC zone is based.

1. Refer to the Stem (1990) manual you downloaded earlier. In particular, see Appendix A, pp. 63-72 (73-83). Upon which map projection is your local (or adopted) zone based?
2. Note that the appendix reports a **central meridian** and **scale factor** for each zone that is based upon a Transverse Mercator projection. **Standard parallels** are listed for zones based upon the Lambert Conformal Conic projection. In the Stem (1990) manual, “scale factor” is expressed in terms of maximum measurement error associated with each zone.
3. Appendix A also lists the SPC coordinates of each zone origin as well as its corresponding geographic coordinates.

4. Use the NGS Toolkit to convert geographic coordinates to SPC coordinates.



National Geodetic Survey State Plane Coordinate toolkit available [here](http://www.ngs.noaa.gov/cgi-bin/gpc_gptpc.pl).

1. You'll need to know the geographic coordinates of a place of interest to complete this part of the activity. To look up the latitude and longitude associated with a U.S. place name, visit the [USGS Geographic Names Information System](#). If you do not reside in the U.S., use the [Getty Thesaurus of Geographic Names](#). Pay attention; be sure to choose the correct instance of the place name. Cities and towns are listed as “Populated Places” in the Geographic Names Information System and as “Inhabited Places” in the Getty Thesaurus.
2. Return to the [National Geodetic Survey's NGS Geodetic Toolkit](#).
3. Use the dropdown menu to navigate to the State Plane Coordinates tool.
4. At the [State Plane Coordinates page](#), follow the Interactive Conversions link labeled “Latitude/Longitude -> SPC”
5. Specify geographic (a.k.a. “geodetic”) coordinates in degrees-minutes-seconds format. You do not need to show five places to the right of the decimal, as in the example illustrated above, but you do need to add a decimal point at the end of the DMS input values; the tool allows you to input a DMS value that is accurate to fractions of a Second and so is programmed to expect the decimal point.
6. A correct result will include a SPC Northing, Easting, zone, convergence (a correction factor for distance calculations that compensates for the tendency of meridians to converge toward the poles), and the scale factor at the specified point.

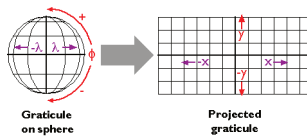
Practice Quiz

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about State Plane Coordinate system.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.28. Map Projections

Latitude and longitude coordinates specify point locations within a coordinate system grid that is fitted to sphere or ellipsoid that approximates the Earth's shape and size. To display extensive geographic areas on a page or computer screen, as well as to calculate distances, areas, and other quantities most efficiently, it is necessary to flatten the Earth.



Map projections are mathematical equations that transform geographic coordinates (conventionally designated by the Greek symbols *lambda* for longitude and *phi* for latitude) into plane coordinates (*x* and *y*). If all the necessary parameters are known, inverse projection equations can be used to transform projected coordinates back into unprojected geographic coordinates.

Georeferenced plane coordinate systems like the Universal Transverse Mercator and State Plane Coordinates systems (examined elsewhere in this lesson) are created by first flattening the graticule, then superimposing a rectangular grid over the flattened graticule. The first step, transforming the geographic coordinate system grid from a more-or-less spherical shape to a flat surface, involves systems of equations called **map projections**.

Many different map projection methods exist. Although only a few are widely used in large scale mapping, the projection parameters used vary greatly. Geographic information systems professionals are expected to be knowledgeable enough to select a map projection that is suitable for a particular mapping objective. Such professionals are expected to be able to recognize the type, amount, and distribution of geometric distortion associated with different map projections. Perhaps most important, they need to know about the parameters of map projections that must be matched in order to merge geographic data from different sources. The pages that follow introduce the key concepts. The topic is far too involved to master in a one section of a single chapter, however. Indeed, Penn State offers an entire one-credit online course in “Map Projections for Geospatial Professionals” (GEOG 861). If you are, or plan to become, a GIS professional, you should own at least one good book on map projections. Several recommendations follow in the bibliography at the end of this lesson. If you care to offer a recommendation of your own, please add it as a comment to the bibliography page.

Students who successfully complete this section should be able to:

1. Interpret distortion diagrams to identify geometric properties of the sphere that are preserved by a particular projection.
2. Classify projected graticules by projection family.

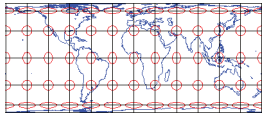
2.29. Geometric Properties Preserved and Distorted

Many types of map projections have been devised to suit particular purposes. No projection allows us to flatten the globe without distorting it, however. **Distortion ellipses** help us to visualize what type of distortion a map projection has caused, how much distortion occurred, and where it occurred. The ellipses show how imaginary circles on the

globe are deformed as a result of a particular projection. If no distortion had occurred in the process of projecting the map shown below, all of the ellipses would be the same size, and circular in shape.

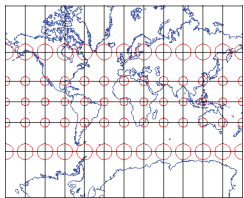
When positions on the graticule are transformed to positions on a projected grid, four types of distortion can occur: distortion of sizes, angles, distances, and directions. Map projections that avoid one or more of these types of distortion are said to preserve certain properties of the globe.

EQUIVALENCE



So-called **equal-area** projections maintain correct proportions in the sizes of areas on the globe and corresponding areas on the projected grid (allowing for differences in scale, of course). Notice that the shapes of the ellipses in the Cylindrical Equal Area projection above are distorted, but the areas each one occupies are equivalent. Equal-area projections are preferred for small-scale thematic mapping, especially when map viewers are expected to compare sizes of area features like countries and continents.

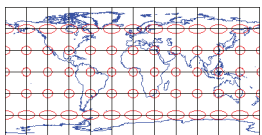
CONFORMALITY



The distortion ellipses plotted on the **conformal** projection shown above vary substantially in size, but are all the same circular shape. The consistent shapes indicate that conformal projections (like this Mercator projection of the world) preserve the fidelity of angle measurements from the globe to the plane. In other words, an angle measured by a land surveyor anywhere on the Earth's surface can be plotted on at its corresponding location on a conformal projection without distortion. This useful property accounts for the fact that conformal projections are almost always used as the basis for large scale surveying and mapping. Among the most widely used conformal projections are the Transverse Mercator, Lambert Conformal Conic, and Polar Stereographic.

Conformality and equivalence are mutually exclusive properties. Whereas equal-area projections distort shapes while preserving fidelity of sizes, conformal projections distort sizes in the process of preserving shapes.

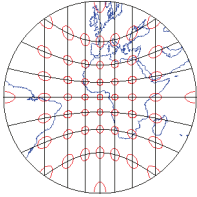
EQUIDISTANCE



Equidistant map projections allow distances to be measured accurately along straight lines radiating from one or two points only. Notice that ellipses plotted on the Cylindrical Equidistant (Plate Carrée) projection shown above vary in both shape and size. The north-south axis of every ellipse is the same length, however. This shows that distances are true-to-scale along every meridian; in other words, the property of equidistance is preserved from the two poles. See chapters 11 and 12 of the online publication *Matching the Map Projection to the Need* to see

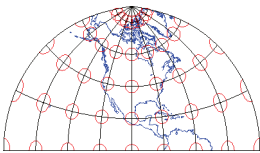
how projections can be customized to facilitate distance measurements and to effectively depict ranges and rings of activity.

AZIMUTHALITY



Azimuthal projections preserve directions (azimuths) from one or two points to all other points on the map. See how the ellipses plotted on the gnomonic projection shown above, vary in size and shape, but are all oriented toward the center of the projection? In this example, that's the one point at which directions measured on the globe are not distorted on the projected graticule.

COMPROMISE



Some map projections preserve none of the properties described above, but instead seek a compromise that minimizes distortion of all kinds. The example shown above is the Polyconic projection, which was used by the U.S. Geological Survey for many years as the basis of its topographic quadrangle map series until it was superseded by the conformal Transverse Mercator. Another example is the Robinson projection, which is often used for small-scale thematic maps of the world.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about the geometric properties of map projections.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.30. Classifying Projection Methods

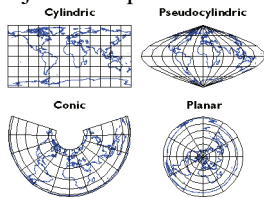
The term “projection” implies that the ball-shaped net of parallels and meridians is transformed by casting its shadow upon some flat, or flattenable, surface. In fact, almost all map projection methods are mathematical equations. The analogy of an optical projection onto a flattenable surface is useful, however, as a means to classify the bewildering variety of projection equations devised over the past two thousand years or more.



Three types of “flattenable” surfaces to which the graticule can be projected: a plane, a cone, and a cylinder. Imagine a model globe that is translucent, and contains a bright light bulb. Imagine the light literally casting

shadows of the graticule, and of the shapes of the continents, onto another surface that touches the globe. [The National Geographic Society](#) has prepared a set of animations that may help you to visualize the analogy.

As you might imagine, the appearance of the projected grid will change quite a lot depending on the type of surface it is projected onto, and how that surface is aligned with the globe. The three surfaces shown above—the disk-shaped plane, the cone, and the cylinder—represent categories that account for the majority of projection equations that are encoded in GIS software. All three are shown in their normal *aspects*. The plane often is centered upon a pole. The cone is typically aligned with the globe such that its line of contact (tangency) coincides with a parallel in the mid-latitudes. And the cylinder is frequently positioned tangent to the equator (unless it is rotated 90°, as it is in the Transverse Mercator projection). The following illustrations shows some of the projected graticules produced by projection equations in each category.



Four categories of map projections

Cylindric projection equations yield projected graticules with straight meridians and parallels that intersect at right angles. The example shown above is a Cylindrical Equidistant (also called Plate Carrée or geographic) in its normal equatorial aspect.

Pseudocylindric projections are variants on cylindrics in which meridians are curved. The result of a Sinusoidal projection is shown above.

Conic projections yield straight meridians that converge toward a single point at the poles, parallels that form concentric arcs. The example shown above is the result of an Albers Conic Equal Area, which is frequently used for thematic mapping of mid latitude regions.

Planar projections also yield meridians that are straight and convergent, but parallels form concentric circles rather than arcs. Planar projections are also called **azimuthal** because every planar projection preserves the property of azimuthality. The projected graticule shown above is the result of an Azimuthal Equidistant projection in its normal polar aspect.

Appearances can be deceiving. It's important to remember that the look of a projected graticule depends on several projection parameters, including latitude of projection origin, central meridian, standard line(s), and others. Customized map projections may look entirely different from the archetypes described above.

TRY THIS!

John Snyder and Phil Voxland (1994) published an *Album of Map Projections* that describes and illustrates many more examples in each projection category. Excerpts from that important work are included in our Interactive Album of Map Projections, which registered students will use to complete Project 1. The [Interactive Album is available here](#). The variety of projections available, as well as users' ability to manipulate projection parameters, is limited to the capabilities of the ArcIMS software platform upon which we developed the Interactive Album.

Flex Projector is a free, open source software program developed in Java that supports many more projections and variable parameters than the Interactive Album. Bernhard Jenny of the Institute of Cartography at ETH Zurich created the program with assistance from Tom Patterson of the US National Park Service. You can download Flex Projector from flexprojector.com

Those who wish to explore map projections in greater depth than is possible in this course might wish to visit an [informative page published by the International Institute for Geo-Information Science and Earth Observation \(Netherlands\)](#), which is known by the legacy acronym ITC.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Classifying Map Projections.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

2.31. Summary

In this lesson we've explored several connotations of the term scale. Scale is synonymous with scope when it is used to describe the extent of a phenomenon. In this sense, "large scale" means "large area." Specialists in geographic information often use the term differently, however. **Map scale** refers to the relative sizes of features on a map and of corresponding objects on the ground. In this context, "large scale" implies "small area." Large scale also implies greater detail and greater accuracy, an important point to keep in mind when using maps as sources for GIS databases. Map scale is defined mathematically as the proportion of map distance to ground distance. I hope you are now prepared to use scale equations to calculate map scale.

Scale can also be thought of as a reference system for measurement. Locations on the globe are specified with reference to the **geographic coordinate system** of latitudes and longitudes. **Plane coordinates** are often preferred over geographic coordinates because they ease calculations of distance, area, and other quantities. Georeferenced plane coordinate systems like **UTM and SPC** are established by first flattening the graticule, then superimposing a plane coordinate grid. The mathematical equations used to transform geographic coordinates into plane coordinates are called **map projections**. Both plane and geographic coordinate system grids are related to approximations of the Earth's size and shape called **ellipsoids**. Relations between grids and ellipsoids are called **horizontal datums**.

Horizontal datum is an elusive concept for many GIS practitioners. It is relatively easy to visualize a horizontal datum in the context of unprojected geographic coordinates. Simply drape the latitude and longitude grid over an ellipsoid and there's your horizontal datum. It is harder to think about datum in the context of a projected coordinate grid like UTM and SPC, however. Think of it this way: First drape the latitude and longitude grid on an ellipsoid. Then project that grid to a 2-D plane surface. Then, superimpose a rectangular grid of eastings and northings over the projection, using control points to georegister the grids. There you have it—a projected coordinate grid based upon a horizontal datum.

Numerous coordinate systems, datums, and map projections are in use around the world. Because we often need to combine georeferenced data from various sources, GIS professionals need to be able to **georegister** two or more data sets that are based upon different coordinate systems, datums, and/or projections. **Transformations**, including coordinate transformations, datum transformations, and map projections, are the mathematical procedures used to bring diverse data into alignment. Characteristics of the coordinate systems, datums, and projections considered in this course are outlined in the following tables.

Coordinate systems referenced in this course (many other national and local systems are in use)

Coordinate systems referenced

Coordinate system	Units	Extent	Projection basis
Geographic	Angles (expressed as degrees, minutes, seconds or decimal degrees).	Global	None
UTM	Distances (meters)	Near-global (8430' N, 80° 30' S)	Unique Transverse Mercator projection for each of 60 zones
State Plane Coordinates	Distances (meters in SPCS 83, feet in SPCS 27)	U.S.	Unique Transverse Mercator or Lambert Conformal Conic projection for each of 123 zones (plus Oblique Mercator for Alaska panhandle)

Datums referenced in this course (many other national and local systems are in use)

Datums referenced

Datum	Horizontal or vertical	Optimized for	Reference surface
NAD 27	Horizontal	North America	Clarke 1866 ellipsoid
NAD 83	Horizontal	North America	GRS 80 ellipsoid
WGS 84	Horizontal	World	WGS 84 ellipsoid
NAVD 88	Vertical	North America	Sea level measured at coastal tidal stations

Map projections referenced in this course (many other national and local systems are in use)

Map projections referenced

Projection name	Properties preserved	Class	Distortion
Mercator	Conformal	Cylindrical	Area distortion increases with distance from standard parallel (typically equator)
Transverse Mercator	Conformal	Cylindrical	Area distortion increases with distance from standard meridian
Lambert Conformal Conic	Conformal	Conic	Area distortion increases with distance from one or two standard parallels
Plate Carrée (sometimes called “Geographic” projection)	Equidistant	Cylindrical	Area and shape distortion increases with distance from standard parallel (typically equator)
Albers Equal-Area Conic	Equivalent	Conic	Shape distortion increases with distance from one or two standard parallels

Compiled from Snyder, 1997

QUIZ

Registered Penn State students should return now to the Chapter 2 folder in ANGEL (via the Resources menu to the left) to take the Chapter 2 graded quiz.

This one counts. **You may take graded quizzes only once.**

The purpose of the graded quizzes is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter's learning objectives. You are free to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 2.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

2.32. Bibliography

3-D Software (2005). *Map projections pages*. Retrieved January 8, 2005, from <http://www.3dsoftware.com/Cartography/>

American Congress on Surveying and Mapping (n. d.). *The North American Datum of 1983. A collection of papers describing the planning and implementation of the readjustment of the North American horizontal network*. Monograph No. 2.

Burkard, R. K. et al. (1959-2002). *Geodesy for the layman*. Retrieved October 29, 2003, from the National Imagery and Mapping Agency Web site http://www.ngs.noaa.gov/PUBS_LIB/Geodesy4Layman/toc.htm

Chem-Nuclear Systems, Inc. (1993). *Site screening interim report: Stage two — regional disqualification*. Harrisburg PA.

Chrisman, N. (2002). *Exploring geographic information systems (2nd ed.)*. New York: John Wiley & Sons.

Clarke, K. (1995). *Analytical and computer cartography (2nd ed.)*. Upper Saddle River, NJ: Prentice Hall.

Dana, P. H. (1998). *Coordinate systems overview. The Geographer's Craft Project*. Retrieved June 25, 2004, from The University of Colorado at Boulder, Department of Geography Web site: <http://www.colorado.edu/geography/gcraft/notes/coordsys/coordsys.html>

Dana, P. H. (1999). *Geodetic datums overview. The Geographer's Craft Project*. Retrieved June 25, 2004, from The University of Colorado at Boulder, Department of Geography Web site: <http://www.colorado.edu/geography/gcraft/notes/datum/datum.html>

Dewhurst, W. T. (1990). *NADCON: The application of minimum-curvature-derived surfaces in the transformation of positional data from the North American datum of 1927 to the North American datum of 1983*. NOAA Technical Memorandum NOS NGS 50. Retrieved January 1, 2005, from http://www.ngs.noaa.gov/PUBS_LIB/NGS50.pdf

Doyle, D. (2004, February). NGS geodetic toolkit, Part 7: Computing state plane coordinates. *Professional Surveyor Magazine*, 24:, 34-36.

Dutch, S. (2003). The Universal Transverse Mercator System. Retrieved January 9, 2008 from <http://www.uwgb.edu/DutchS/FieldMethods/UTMSystem.htm>

Federal Geographic Data Committee. (December 2001). *United States National Grid*. Retrieved May 8, 2006, from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/usng/fgdc_std_011_2001_usng.pdf

Hildebrand, B. (1997). *Waypoint+*. Retrieved January 1, 2005, from <http://www.tapr.org/~kh2z/Waypoint/>

Illiffe, J.C. (2000). *Datums and map projections for remote sensing, GIS and surveying*. Caithness, Scotland: Whittles Publishing. Distributed in U.S. by CRC Press.

Larrimore, C. (2002). *NGS Geodetic Toolkit*. Retrieved October 26, 2004, from <http://www.ngs.noaa.gov/TOOLS>

Muehrcke, P. C. & Muehrcke, J. O. (1992). *Map use (3rd ed.)*. Madison WI: JP Publications.

Muehrcke, P. C. & Muehrcke, J. O. (1998). *Map use (4th ed.)*. Madison WI: JP Publications.

Mulcare, D. M. (2004). The National Geodetic Survey NADCON Tool. *Professional Surveyor Magazine*, February, pp. 28-33.

National Geodetic Survey. (n.d.). *North American datum conversion utility*. Retrieved April 2004, from <http://www.ngs.noaa.gov/TOOLS/Nadcon/Nadcon.html>

National Geodetic Survey. (1997). *Image generated from 15'x15' geoid undulations covering the planet Earth*. Retrieved 1999, from <http://www.ngs.noaa.gov/GEOID/geo-index.html> (since retired).

National Geodetic Survey. (2004). *Coast and geodetic survey historical image collection*. Retrieved June 25, 2004, from <http://www.photolib.noaa.gov/cgs/index.html>

National Geographic Society (1999). *Round earth, flat maps*. Retrieved April 18, 2006, from <http://www.nationalgeographic.com/features/2000/exploration/projections/index.html>

Ordnance Survey (2000). *National GPS network information. 7: Transverse mercator map projections*. Retrieved August 27, 2004, from <http://www.gps.gov.uk/guide7.asp>

Robinson, A. et al. (1995). *Elements of cartography (5th ed.)*. New York: John Wiley & Sons.

Robinson, A. H. & Snyder, J. P. (1997). *Matching the map projection to the need*. Retrieved January 8, 2005, from the Cartography and Geographic Information Society and the Pennsylvania State University web site: <https://courseware.e-education.psu.edu/projection/>

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2005). *Thematic cartography and visualization (2nd ed.)*. Upper Saddle River, NJ: Prentice Hall.

Smith, J.R. (1988). *Basic geodesy*. Rancho Cordova CA: Landmark Enterprises.

Snyder, J. P. (1987). *Map projections: A working manual* (U.S. Geological Survey Professional Paper No. 1395). Washington DC: United States Government Printing Office.

Snyder, J. P. (1987). *Map projections: A working manual*. (USGS Professional Paper No. 1395). Washington DC: U.S. Geological Survey (Electronic versions available at http://pubs.er.usgs.gov/djvu/PP/PP_1395.pdf)

Snyder, J. P. & Voxland P. M. (1989). *An album of map projections* (U.S. Geological Survey Professional Paper No. 1453). Washington DC: United States Government Printing Office.

Snyder, J. P. & Voxland, P. M. (1994). *An album of map projections*. (USGS Professional Paper No. 1453). Washington DC: U.S. Geological Survey. (ordering information published at <http://erg.usgs.gov/isb/pubs/factsheets/fs08799.html>)

Stem, J. E. (1990). *State Plane Coordinate System of 1983 (NOAA Manual NOS NGS 5)*. Rockville, MD: National Geodetic Information Center.

The Large Scale Biosphere-Atmosphere Experiment in Amazonia (1999, July 1). Retrieved July 12, 1999, from http://daacl.ESD.ORNL.Gov/lba_cptec/ (since retired).

United States Geological Survey (2001). *The universal transverse mercator grid. Fact sheet 077-01*. Retrieved June 30, 2004, from <http://mac.usgs.gov/mac/isb/pubs/factsheets/fs07701.html> (since retired).

United States Geological Survey (2003). *National mapping program standards*. Retrieved October 29, 2005, from <http://rockyweb.cr.usgs.gov/nmpstds/nmas647.html>

USGS. "State College Quadrangle" [map]. 7.5 minute series. Washington, D.C.: USGS, 1962.

Van Sickle, J. (2004). *Basic GIS coordinates*. Boca Raton FL: CRC Press.

Wikipedia. The free encyclopedia. (2006). *World geodetic system*. Retrieved May 8, 2006, from <http://en.wikipedia.org/wiki/WGS84>

Wolf, P. R. & Brinker, R. C. (1994) *Elementary Surveying (9th ed.)*. New York NY: HarperCollins.

Chapter 3

Census Data and Thematic Maps

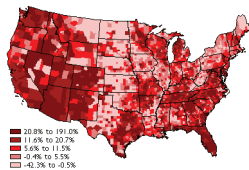
David DiBiase

3.1. Overview

In Chapter 2 we compared the characteristics of geographic and plane coordinate systems that are used to measure and specify positions on the Earth's surface. Coordinate systems, remember, are formed by juxtaposing two or more spatial measurement scales. I mentioned, but did not explain, that **attribute data also are specified with reference to measurement scales**. In this chapter we'll take a closer look at how attributes are measured and represented.

Maps are both the raw material and the product of GIS. All maps, but especially so-called **reference maps** made to support a variety of uses, can be defined as sets of symbols that represent the locations and attributes of entities measured at certain times. Many maps, however, are subsets of available geographic data that have been selected and organized in response to a particular question. Maps created specifically to highlight the distribution of a particular phenomenon or theme are called **thematic maps**. Thematic maps are among the most common forms of geographic information produced by GIS.

A flat sheet of paper is an imperfect but useful analog for geographic space. Notwithstanding the intricacies of map projections, it is a fairly straightforward matter to plot points that stand for locations on the globe. Representing the attributes of locations on maps is sometimes not so straightforward, however. Abstract graphic symbols must be devised that depict, with minimum ambiguity, the quantities and qualities that give locations their meaning. Over the past 100 years or so, cartographers have adopted and tested conventions concerning symbol color, size, and shape for thematic maps. The effective use of graphic symbols is an important component in the transformation of geographic data into useful information.



Population change in the United States, by county, from 1990 to 2000.
(Data from 1990 & 2000 decennial censuses).

Consider the map above, which shows how the distribution of U.S. population changed, by county, from 1990 to 2000. To gain a sense of how effective this thematic map is in transforming data into information, we need only to compare it to a [list of population change rates for the more than 3,000 counties of the U.S.](#) The thematic map reveals spatial patterns that the data themselves conceal.

This chapter explores the characteristics of attribute data used for thematic mapping, especially attribute data produced by U.S. Census Bureau. It also considers how the characteristics of attribute data influence choices about how to present the data on thematic maps.

Objectives

Students who successfully complete Chapter 3 should be able to:

1. Use metadata and the World Wide Web to assess the content and availability of attribute data produced by the U.S. Census Bureau;
2. Discriminate between different levels of measurement of attribute data;
3. Explain the differences between counts, rates, and densities, and identify the types of map symbols that are most appropriate for representing each; and
4. Use quantile and equal interval classification schemes to divide census attribute data into categories suitable for choroplethic mapping.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

3.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 3 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 3	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	Submit five practice quizzes including: <ul style="list-style-type: none"> • Census Attribute Data • Recognizing Levels of Measurement • Levels and Operations • Thematic Map Types • Data Classification for Thematic Mapping Practice quizzes are not graded and may be submitted more than once.	Go to ANGEL > [your course section] > Lessons tab > Chapter 3 folder > [quiz]
3	Perform “ Try this ” activities including: <ul style="list-style-type: none"> • Acquiring U.S. Census data • Acquiring world demographic data “Try this” activities are not graded.	Instructions are provided for each activity.
4	Submit the Chapter 3 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 3 folder > Chapter 3 Graded Quiz
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

3.3. Census Attribute Data

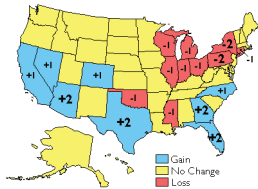
A thematic map is a graphic display that shows the geographic distribution of a particular attribute, or relationships among a few selected attributes. Some of the richest sources of attribute data are national censuses. In the United States, a periodic count of the entire population is required by the U.S. Constitution. Article 1, Section 2, ratified in 1787, states that Representatives and direct taxes shall be apportioned among the several states which may be included within this union, according to their respective numbers . . . The actual Enumeration shall be made [every] ten years, in such manner as [the Congress] shall by law direct.” The U.S. Census Bureau is the government agency charged with carrying out the decennial census.



A portion of the Constitution of the United States of America.

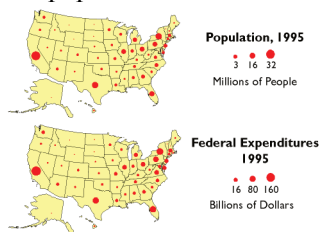
The results of the U.S. decennial census determine states' portions of the 435 total seats in the U.S. House

of Representatives. The map below shows states that lost and gained seats as a result of the **reapportionment** that followed the 2000 census. Congressional voting district boundaries must be redrawn within the states that gained and lost seats, a process called **redistricting**. Constitutional rules and legal precedents require that voting districts contain equal populations (within about 1 percent). In addition, districts must be drawn so as to provide equal opportunities for representation of racial and ethnic groups that have been discriminated against in the past.



Reapportionment of the U.S. House of Representatives as a result of the 2000 census.

Besides reapportionment and redistricting, U.S. census counts also affect the flow of billions of dollars of federal expenditures, including contracts and federal aid, to states and municipalities. In 1995, for example, some \$70 billion of Medicaid funds were distributed according to a formula that compared state and national per capita income. \$18 billion worth of highway planning and construction funds were allotted to states according to their shares of urban and rural population. And \$6 billion of Aid to Families with Dependent Children was distributed to help children of poor families do better in school. The two thematic maps below illustrate the strong relationship between population counts and the distribution of federal tax dollars.



Population and federal expenditures, by state, 1995. (Cartography by Thad Lenker. Data from U.S. Census Bureau, Federal Expenditures by State, <http://www.census.gov/prod/2/gov/fes95rv.pdf>)

The Census Bureau's mandate is to provide the population data needed to support governmental operations including reapportionment, redistricting, and allocation of federal expenditures. Its mission, to be "the preeminent collector and provider of timely, relevant, and quality data about the people and economy of the United States", is broader, however. To fulfill this mission, the Census Bureau needs to count more than just numbers of people, and it does.

TRY THIS!

[The Redistricting Game](#)

3.4. Enumerations versus Samples

Sixteen U.S. Marshals and 650 assistants conducted the first U.S. census in 1791. They counted some 3.9 million individuals, although as then-Secretary of State Thomas Jefferson reported to President George Washington, the official number understated the actual population by at least 2.5 percent (Roberts, 1994). By 1960, when the U.S. population had reached 179 million, it was no longer practical to have a census taker visit every household. The Census Bureau then began to distribute questionnaires by mail. Of the 116 million households to which questionnaires were sent in 2000, 72 percent responded by mail. A mostly-temporary staff of over 800,000 was needed to visit the remaining households, and to produce the final count of 281,421,906. Using statistically reliable estimates produced from exhaustive follow-up surveys, the Bureau's permanent staff determined that the final count

was accurate to within 1.6 percent of the actual number (although the count was less accurate for young and minority residences than it was for older and white residents). It was the largest and most accurate census to that time. (Interestingly, Congress insists that the original **enumeration** or “head count” be used as the official population count, even though the estimate calculated from samples by Census Bureau statisticians is demonstrably more accurate.)

As of this writing, the decennial census of 2010 is still underway. Like 2000, the mail-in response rate was 72 percent. The official 2010 census count, by state, must be delivered to the U.S. Congress by December 31, 2010.

In 1791, census takers asked relatively few questions. They wanted to know the numbers of free persons, slaves, and free males over age 16, as well as the sex and race of each individual. (You can [view replicas of historical census survey forms here](#)) As the U.S. population has grown, and as its economy and government have expanded, the amount and variety of data collected has expanded accordingly. In the 2000 census, all 116 million U.S. households were asked six population questions (names, telephone numbers, sex, age and date of birth, Hispanic origin, and race), and one housing question (whether the residence is owned or rented). In addition, a statistical sample of one in six households received a “long form” that asked 46 more questions, including detailed housing characteristics, expenses, citizenship, military service, health problems, employment status, place of work, commuting, and income. From the sampled data the Census Bureau produced estimated data on all these variables for the entire population.

In the parlance of the Census Bureau, data associated with questions asked of all households are called **100% data** and data estimated from samples are called **sample data**. Both types of data are available aggregated by various **enumeration areas**, including census block, block group, tract, place, county, and state (see the illustration below). Through 2000, the Census Bureau distributes the 100% data in a package called the “Summary File 1” (SF1) and the sample data as “Summary File 3” (SF3). In 2005 the Bureau launched a new project called **American Community Survey** that surveys a representative sample of households on an ongoing basis. Every month one household out of every 480 in each county or equivalent area receives a survey similar to the old “long form.” Annual or semi-annual estimates produced from American Community Survey samples replaced the SF3 data product in 2010.

To protect respondents’ confidentiality, as well as to make the data most useful to legislators, the Census Bureau aggregates the data it collects from household surveys to several different types of geographic areas. SF1 data, for instance, are reported at the block or tract level. There were about 8.5 million census blocks in 2000. By definition, census **blocks** are bounded on all sides by streets, streams, or political boundaries. Census **tracts** are larger areas that have between 2,500 and 8,000 residents. When first delineated, tracts were relatively homogeneous with respect to population characteristics, economic status, and living conditions. A typical census tract consists of about five or six sub-areas called **block groups**. As the name implies, block groups are composed of several census blocks. American Community Survey estimates, like the SF3 data that preceded them, are reported at the block group level or higher.



Relationships among the various census geographies. (U.S. Census Bureau, American FactFinder, 2005, http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pageId=gn7_maps An updated source for the diagram can be found at http://factfinder2.census.gov/faces/nav/jsf/pages/using_factfinder5.xhtml).

TRY THIS!

Acquiring U.S. Census Data via the World Wide Web

The purpose of this practice activity is to guide you through the process of finding and acquiring 2000 census

data from the U.S. Census Bureau data via the Web. Your objective is to look up the total population of each county in your home state (or an adopted state of the U.S.). On January 29, 2013, a redesigned version of the American FactFinder web pages was revealed. Some necessary changes to the steps below are highlighted in green text.

1. Go to the [U.S. Census Bureau site](#).
2. At the Census Bureau home page, hover your mouse cursor over the **Data** tab and select **American FactFinder**. American FactFinder is the Census Bureau's primary medium for distributing census data to the public.
3. Click the **ADVANCED SEARCH** button, and take note of the three steps featured on the page you are taken to. That's what we are about in this exercise.
4. Click the **Topics** search option box. In the **Select Topics** overlay window expand the **People** list. Next expand the **Basic Count/Estimate** list. Then choose **Population Total**. Note that a Population Total entry is placed in the **Your Selections** box in the upper left, and it disappears from the Basic Count/Estimate list.

Close the **Select Topics** window.

5. The list of datasets in the resulting **Search Results** window is for the entire United States. We want to narrow the search to county-level data for your home or adopted state.

Click the **Geographies** search options box. In the **Select Geographies** overlay window that opens, under **Select a geographic type:**, click **County – 050**.

Next select the entry for your state from the **Select a state** list, and then from the **Select one or more geographic areas...** list select **All counties within <your state>**.

Last click **ADD TO YOUR SELECTIONS**. This will place your All Counties... choice in the **Your Selections** box.

Close the **Select Geographies** window.

6. The list of datasets in the Search Results window now pertains to the counties in your state. Take a few moments to review the datasets that are listed. Note that there are SF1, SF2, ACS (American Community Survey), etc., datasets, and that if you page through the list far enough you will see that data from past years is listed. We are going to focus our effort on the **2010 SF1 100% Data**.

7. Given that our goal is to find the population of the counties in your home state, can you determine which dataset we should look at?

There is a **TOTAL POPULATION** entry. Find it, and make certain you have located the 2010 SF1 100% Data dataset. (You can use the **Narrow your search:** slot above the dataset list to help narrow the search.)

Check the box for it, and then click **View**.

In the new **Results** window that opens you should be able to find the population of the counties your chosen state.

Note the row of **Actions:**, which includes **Print** and **Download** buttons. I encourage you to experiment some with the American FactFinder site. Start slow, and just click the **BACK TO ADVANCED SEARCH** button, un-check the TOTAL POPULATION dataset and choose a different dataset to investigate. Registered students will need to answer a couple of quiz questions based on using this site. Pay attention to what is in the **Your Selections** window. You can easily remove entries by clicking the blue circle with the white X. On a search page you might try typing "QT" or "GCT" in the **topic or table**

name slot. QT stands for **Quick Tables** which are preformatted tables that show several related themes for one or more geographic areas. GCT stands for **Geographic Comparison Tables** which are the most convenient way to compare data collected for all the counties, places, or congressional districts in a state, or all the census tracts in a county.

{C}

{C}
{C}
{C}
{C}

1. Go to the U.S. Census Bureau site at <http://www.census.gov>.
2. At the Census Bureau home page, hover your mouse cursor over the **Data** tab and select **American FactFinder**. American FactFinder is the Census Bureau's primary medium for distributing census data to the public.
3. Click the **SEARCH** button, and take note of the three steps featured in the yellow rectangle. That's what we are about in this exercise.
4. Click the **Topics** search option box. In the **Select Topics** overlay window expand the **People** list. Next expand the **Basic Count/Estimate** list. Then choose **Population Total**. Note that a Population Total entry is placed in the **Your Selections** box in the upper left, and it disappears from the Basic Count/Estimate list.

Close the **Select Topics** window.

5. The list of datasets in the resulting **Search Results** window is for the entire United States. We want to narrow the search to county-level data for your home or adopted state.

Click the **Geographies** search options box. In the **Select Geographies** overlay window that opens, under **Geography Filter Options**, click **County**. This will yield a list of **All counties within <your state>** under **Geography Results**.

Check the box next to the entry for your state, and then click **Add**. This will place your All Counties... choice in the **Your Selections** box.

Close the **Select Geographies** window.

6. The list of datasets in the **Search Results** window now pertains to the counties in your state. Take a few moments to review the datasets that are listed. Note that there are SF1, SF2, ACS (American Community Survey), etc., datasets, and that if you page through the list far enough you will see that data from past years is listed. We are going to focus our effort on the **2010 SF1 100% Data**.

7. Given that our goal is to find the population of the counties in your home state, can you determine which dataset we should look at?

There is a **TOTAL POPULATION** entry, probably on page 2. Find it, and make certain you have located the 2010 SF1 100% Data dataset.

Check the box for it and click **View**.

In the new **Results** window that opens you should be able to find the population of the counties your chosen state.

Note the row of **Actions:**, which includes **Print** and **Download** buttons.

I encourage you to experiment some with the American FactFinder site. Start slow, and just click the **BACK TO SEARCH** button, un-check the TOTAL POPULATION dataset and choose a different dataset to investigate. Registered students will need to answer a couple of quiz questions based on using this site.

Pay attention to what is in the **Your Selections** window. You can easily remove entries by clicking the red circle with the white X.

On the **SEARCH** page, with nothing in the Your Selections box, you might try typing “QT” or “GCT” in the **Search for:** slot. QT stands for **Quick Tables** which are preformatted tables that show several related themes for one or more geographic areas. GCT stands for **Geographic Comparison Tables** which are the most convenient way to compare data collected for all the counties, places, or congressional districts in a state, or all the census tracts in a county.

3.5. American Community Survey

Beginning in 2010, the American Community Survey (ACS) replaced the “long form” that was used to collect sample data in past decennial censuses. Instead of sampling one in six households every ten years (about 18 million households in 2000), the ACS samples 2-3 million households every year. The goal of the ACS is to enable Census Bureau statisticians to produce more timely estimates of the demographic, economic, social, housing, and financial characteristics of the U.S. population. You can view a sample ACS questionnaire by entering the keywords “American Community Survey questionnaire” into your favorite Internet search engine.

Acquiring and Understanding American Community Survey (ACS) Data The purpose of this practice activity is to guide your exploration of ACS data and methodology. In the end you should be able to identify the types of geographical areas for which ACS data are available; to explain why 1-year and 3-year estimates are available for some areas and not for others; and to describe how the statistical reliability of ACS estimates vary among 1-year, 3-year and 5-year estimates.

1. Return to the U.S. Census Bureau site at <http://www.census.gov>.
2. With your mouse cursor, hover over the **People** tab and under **Related Content** follow the link to **American Community Survey**. This takes you to the main American Community Survey page. (You can also find a link to American Community Survey by following the **Subjects A to Z** link in the upper right.)
3. Begin by clicking the **Guidance for Data Users** tab and looking through the information available there. Pay particular attention to the **When to use...** section and the descriptions of the various estimates (1-, 3- and 5-year).
You will also find a section on **Comparing ACS Data** to other census data, a section on **Handbooks for Data Users**, and an **E-Tutorial**. (Some of the tutorial is not up to date relative to the new web pages, but you might benefit from **Lesson3: Understanding the American Community Survey**.)
4. Next, look at the content under the **Data & Documentation** tab.
In the **Data Releases** section you will find release dates for the various datasets.
In the **Documentation** section there are links to documentation associated with current and past surveys, and within that section, under **Accuracy of the Data**, links to documents describing the methodology used and the accuracy of the data estimates.
You can download ACS data to make maps and analyses using your own GIS or statistical software. Find download links and pertinent information in the sections titled **Downloadable Data via FTP**, and **Summary File**.
There is also a section pertaining to **Public Use Microdata Sample (PUMS)**. PUMS data are edited, however, to protect the confidentiality of individuals and households.
5. In the remaining steps you will make a map or two, to reinforce the geographies covered by the ACS. You will map data from your home (or adopted) state.
You need to go to the American FactFinder. If you are still on the American Community Survey page, click the **Data & Documentation** tab, then follow the link to the **American FactFinder website**. You should land on the **SEARCH** page with American Community Survey in the **Your Selections** window, and a list of **Search Results** that are ACS-based.
(If you were not already on the American Community Survey page, go to the MAIN American FactFinder site (<http://factfinder2.census.gov>), click the **Topics** search box, then expand the **Program** list and choose **American Community Survey**. Close the Select Topics overlay window.)
6. Click the **Geographies** search options box (on the left) to reveal the **Select Geographies** overlay window. Under **Select a geographic type** click **County – 050**. Next from the **Select a state** list, choose your state. Then from the **Select one or more geographic areas...** list, choose **All Counties within <your state>**. Then click **ADD TO YOUR SELECTIONS**. This will add the All Counties... entry to the **Your Selections** list. Close the **Select Geographies** overlay window.
If the **Search Results** window shows a list of datasets for 2005, advance the page, the list will refresh so that page 1 shows datasets for the most recent years.
7. In the **Search Results** window note that there are many datasets that have 1-, 3- and 5-year estimates entries. Decide upon a *1-Year* dataset to look at and check the box for it. Then click **View**. On the new **Results** page that you land on be sure that the **Create a Map** choice is blue – not grayed out. If it is grayed out click the **BACK TO ADVANCED SEARCH** button and make sure only one dataset box is checked, or make a different choice, then click **View** again.
Click on **Create a Map**. The data values in the table will turn blue and you will be prompted to “Click on a data value in the table.” Clicking a single data value from any *row* will allow you to map the data in that row for all of the counties for which it is available. Click on a blue data value of your choice – *remember which row you choose*. Click on the **SHOW MAP** button in the small popup window that appears.
Are all of the counties in the state symbolized as having data? Why not?
8. Now click the **BACK TO ADVANCED SEARCH** button. Un-check the box for the 1-year dataset and check the box for the 3-year estimate of the same category. Proceed as above to map the data. After the map is refreshed note how many counties now exhibit data.
Take a look at the 5-year estimates for the same dataset if you wish.

Try
This!

Practice Quiz	Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Census Attribute Data. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.
---------------	--

3.6. International Data

The International Data Base is published on the Web by the Census Bureau's International Programs Center. It combines demographic data compiled from censuses and surveys of some 227 countries and areas of the world, along with estimates produced by Census Bureau demographers. Data variables include population by age and sex; vital rates, infant mortality, and life tables; fertility and child survivorship; migration; marital status; family planning; ethnicity, religion, and language; literacy; and labor force, employment, and income. Census and survey data are available by country for selected years from 1950; projected data are available through 2050. The International Data Base allows you to download attribute data in formats appropriate for thematic mapping.

TRY THIS!

Acquiring World Demographic Data via the World Wide Web

The purpose of this practice activity is to guide you through the process of finding and acquiring demographic data for the countries of the world from the U.S. Census Bureau data via the Web. Your objective is to retrieve population change rates for a country of your choice over two or more years.

1. Return to the [U.S. Census Bureau site](#).
2. Hover over or click the **People** tab and choose **International Data Base**.
3. Choose the data theme you are interested in from the **Select Report** pick list. The choices have to do with births and mortality, population change including such things as migration, population by age group, etc.
4. Tables are available by Country or by Region.
From the **Select Country(ies)** selection box you can specify that you want data for a single country, or for a collection of multiple countries, for as many **Year(s)** as you want to select. See the instructions beneath the Submit button on how to select multiple entries from the selection boxes.
From the **Select Region(s)** selection box you can choose from pre-selected groupings of countries.
5. Choose a single country under **Country Search** and two or more years. Then click **SUBMIT**. You will see a summary table of data for your selected country and years.
6. Experiment with the choices in the **Select Region(s)** selection box and the **Aggregation Options** choice list.
7. For your information: to download an Excel (.xls) or text file (.csv) version of the data, find the respective link on the Results page: "Excel" or "CSV"
Download links may not appear when the search has been broad.

3.7. Counts, Rates, and Densities

The raw data collected during decennial censuses are **counts**—whole numbers that represent people and housing units. The Census Bureau aggregates counts to geographic areas such as counties, tracts, block groups and blocks, and reports the aggregate totals. In other cases summary measures, such as **averages** and **medians**, are reported. Counts can be used to ensure that redistricting plans comply with the constitutional requirement that each district

contain equal population. Districts are drawn larger in sparsely populated areas, and smaller where population is concentrated. Counts, averages, and medians cannot be used to determine that districts are drawn so that minority groups have an equal probability of representation, however. For this, pairs of counts must be converted into **rates** or **densities**. A rate, such as Hispanic population as a percentage of total population, is produced by dividing one count by another. A density, such as persons per square kilometer, is a count divided by the area of the geographic unit to which the count was aggregated. In this chapter we'll consider how the differences between counts, rates, and densities influence the ways in which the data may be processed in geographic information systems and displayed on thematic maps.

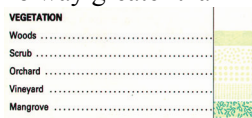
3.8. Attribute Measurement Scales

Chapter 2 focused upon measurement scales for spatial data, including map scale (expressed as a representative fraction), coordinate grids, and map projections (methods for transforming three dimensional to two dimensional measurement scales). You may know that the meter, the length standard established for the international metric system, was originally defined as one-ten-millionth of the distance from the equator to the North Pole. In virtually every country except the United States, the metric system has benefited science and commerce by replacing fractions with decimals, and by introducing an Earth-based standard of measurement.

Standardized scales are needed to measure non-spatial attributes as well as spatial features. Unlike positions and distances, however, **attributes** of locations on the Earth's surface are often not amenable to absolute measurement. In a 1946 article in *Science*, a psychologist named S. S. Stevens outlined a system of four **levels of measurement** meant to enable social scientists to systematically measure and analyze phenomena that cannot simply be counted. (In 1997, geographer Nicholas Chrisman pointed out that a total of nine levels of measurement are needed to account for the variety of geographic data.) The levels are important to specialists in geographic information because they provide guidance about the proper use of different statistical, analytical, and cartographic operations. In the following we consider examples of Stevens' original four levels of measurement: **nominal, ordinal, interval, and ratio**.

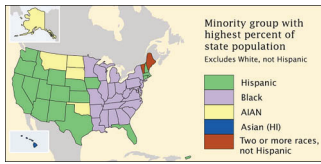
3.9. Nominal Level

Data produced by assigning observations into unranked categories are said to be **nominal** level measurements. Nominal categories can be differentiated and grouped into categories, but cannot logically be ranked from high to low (unless they are associated with preferences or other exogenous value systems). For example, one can classify the land cover at a certain location as woods, scrub, orchard, vineyard, or mangrove. One cannot say, however, that a location classified as "woods" is twice as vegetated as another location classified "scrub." The phenomenon "vegetation" is a set of categories, not range of numerical values, and the categories are not ranked. That is, "woods" is in no way greater than "mangrove," unless the measurement is supplemented by a preference or priority.



Attribute data measured at the nominal level: Selected vegetation categories depicted on USGS topographic maps. (Steger, 1986).

Although census data originate as counts, much of what is counted is individuals' membership in nominal categories. Race, ethnicity, marital status, mode of transportation to work (car, bus, subway, railroad...), type of heating fuel (gas, fuel oil, coal, electricity...), all are measured as numbers of observations assigned to unranked categories. For example, the map below, which appears in the Census Bureau's first atlas of the 2000 census, highlights the minority groups with the largest percentage of population in each U.S. state. Colors were chosen to differentiate the groups, but not to imply any quantitative ordering.



(Brewer & Suchan, 2001).

3.10. Ordinal Level

Like the nominal level of measurement, **ordinal** scaling assigns observations to discrete categories. Ordinal categories are ranked, however. It was stated in the preceding page that nominal categories such as “woods” and “mangrove” do not take precedence over one another, unless an extrinsic set of priorities is imposed upon them. In fact, the act of prioritizing nominal categories transforms nominal level measurements to the ordinal level.

BOUNDARIES	
National	-----
State or territorial	-----
County or equivalent	-----
Civil township or equivalent	-----
Incorporated-city or equivalent	-----
Park, reservation, or monument	-----
Small park	-----

Attribute data measured at the ordinal level: Ranked categories of boundaries depicted on USGS topographic maps.

Examples of ordinal data often seen on reference maps include political boundaries that are classified hierarchically (national, state, county, etc.) and transportation routes (primary highway, secondary highway, light-duty road, unimproved road). Ordinal data measured by the Census Bureau include how well individuals speak English (very well, well, not well, not at all), and level of educational attainment. Social surveys of preferences and perceptions are also usually scaled ordinally.

Individual observations measured at the ordinal level typically should not be added, subtracted, multiplied, or divided. For example, suppose two 640-acre grid cells within your county are being evaluated as potential sites for a hazardous waste dump. Say the two areas are evaluated on three suitability criteria, each ranked on a 0 to 3 ordinal scale, such that 0 = unsuitable, 1 = marginally unsuitable, 2 = marginally suitable, and 3 = suitable. Now say Area A is ranked 0, 3, and 3 on the three criteria, while Area B is ranked 2, 2, and 2. If the Siting Commission was to simply add the three criteria, the two areas would seem equally suitable ($0 + 3 + 3 = 6 = 2 + 2 + 2$), even though a ranking of 0 on one criteria ought to disqualify Area A.

3.11. Interval and Ratio Levels

Interval and **ratio** are the two highest levels of measurement in Stevens' original system. Unlike nominal- and ordinal-level data, which are qualitative in nature, **interval- and ratio-level data are quantitative**. Examples of interval level data include temperature and year. Examples of ratio level data include distance and area (e.g., acreage). The scales are similar in so far as units of measurement are arbitrary (Celsius versus Fahrenheit, Gregorian versus Islamic calendar, English versus metric units). The scales differ in that the zero point is arbitrary on interval scales, but not on ratio scales. For instance, zero degrees Fahrenheit and zero degrees Celsius are different temperatures, and neither indicates the absence of temperature. Zero meters and zero feet mean exactly the same thing, however. An implication of this difference is that a quantity of 20 measured at the ratio scale is twice the value of 10, a relation that does not hold true for quantities measured at the interval level (20 degrees is not twice as warm as 10 degrees).

Because **interval and ratio level data represent positions along continuous number lines, rather than members of discrete categories**, they are also amenable to analysis using inferential statistical techniques. Correlation and regression, for example, are commonly used to evaluate relationships between two or more data

variables. Such techniques enable analysts to infer not only the form of a relationship between two quantitative data sets, but also the strength of the relationship.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Recognizing Levels of Measurement. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

3.12. Levels and Operations

One reason that it's important to recognize levels of measurement is that different measurement scales are amenable to different analytical operations (Chrisman 2002). Some of the most common operations include:

- **Group:** Categories of **nominal** and **ordinal** data can be grouped into fewer categories. For instance, grouping can be used to reduce the number of land use/land cover classes from, say, four (residential, commercial, industrial, parks) to one (urban).
- **Isolate:** One or more categories of **nominal**, **ordinal**, **interval**, or **ratio** data can be selected, and others set aside. As a hypothetical example, consider a range of georeferenced soil moisture readings taken over a farm field. A subrange of readings that are amenable to a particular fertilizer or pesticide might be isolated so that application is limited to the appropriate areas of the field.
- **Cross tab:** Two or more sets of **nominal** or **ordinal** categories can be associated one to another in pairs, triplets, etc. Chrisman (2002) points to the multicharacter codes used in the National Wetland Inventory as an example of a cross tab. Each position in the NWI code represents a particular attribute. Each unique code, therefore, represents a cross tabulation of the possible combinations of attributes.
- **Difference:** The difference of two **interval** level observations (such as two calendar years) results in one **ratio** level observation (such as one age).
- **Other arithmetic operations:** Two or more compatible sets of **ratio** or **interval** level data can be added, subtracted, multiplied, or divided. For example, the per capita (average) income of a census tract can be calculated by dividing the sum of the income of every individual in a census tract (a ratio level variable), by the sum of persons residing in the tract (a second ratio level variable).
- **Classification:** **Interval** and **ratio** data are frequently sorted into ordinal level categories for thematic mapping.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Levels and Operations. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

3.13. Thematic Mapping

Unlike reference maps, thematic maps are usually made with a single purpose in mind. Typically, that purpose has to do with revealing the spatial distribution of one or two attribute data sets.

In this section we will consider distinctions among three types of ratio level data, **counts**, **rates**, and **densities**. We will also explore several different types of thematic maps, and consider which type of map is conventionally used to represent the different types of data. We will focus on what is perhaps the most prevalent type of thematic map, the **choropleth** map. Choropleth maps tend to display ratio level data which have been transformed into ordinal level classes. Finally, you will learn two common data classification procedures, **quantiles** and **equal intervals**.

3.14. Graphic Variables

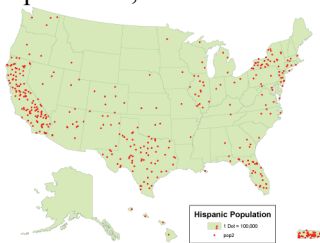
Maps use graphic symbols to represent the locations and attributes of phenomena distributed across the Earth's surface. Variations in symbol **size**, **color lightness**, **color hue**, and **shape** can be used to represent quantitative and qualitative variations in attribute data. By convention, each of these “graphic variables” is used to represent a particular type of attribute data.

3.15. Counts, Rates, and Densities

Ratio level data predominate on thematic maps. Ratio data are of several different kinds, including counts, rates, and densities. As stated earlier, **counts** (such as total population) are whole numbers representing discrete entities, like people. Rates and densities are produced from pairs of counts. A **rate**, such as percent population change, is produced by dividing one count (for example, population in year 2) by another (population in year 1). A **density**, such as persons per square kilometer, is a count divided by the area of the geographic unit to which the count was aggregated (e.g., total population divided by number of square kilometers). It is conventional to use different types of thematic maps to depict each type of ratio-level data.

3.16. Mapping Counts

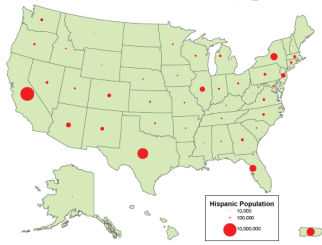
The simplest thematic mapping technique for count data is to show one symbol for every individual counted. If the location of every individual is known, this method often works fine. If not, the solution is not as simple as it seems. Unfortunately, individual locations are often unknown, or they may be confidential. Software like ESRI's ArcMap, for example, is happy to overlook this shortcoming. Its “Dot Density” option causes point symbols to be positioned randomly within the geographic areas in which the counts were conducted. The size of dots, and number of individuals represented by each dot, are also optional. Random dot placement may be acceptable if the scale of the map is small, so that the areas in which the dots are placed are small. Often, however, this is not the case.



A “dot density” map that depicts count data. Cartography by Geoff Hatchard.

An alternative for mapping counts that lack individual locations is to use a single symbol, a circle, square, or some other shape, to represent the total count for each area. ArcMap calls the result of this approach a **Proportional Symbol** map. In the map shown below, the size of each symbol varies in direct proportion to the data value it represents. In other words, the area of a symbol used to represent the value “1,000,000” is exactly twice as great as a symbol that represents “500,000.” To compensate for the fact that map readers typically underestimate symbol size, some cartographers recommend that symbol sizes be adjusted. ArcMap calls this option “Flannery Compensation” after James Flannery, a research cartographer who conducted psychophysical studies of map symbol perception in

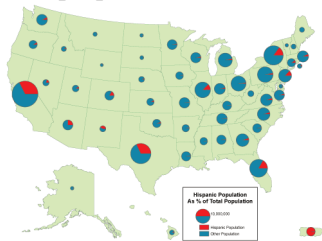
the 1950s, 60s, and 70s. A variant on the Proportional Symbol approach is the **Graduated Symbol** map type, in which different symbol sizes represent categories of data values rather than unique values. In both of these map types, symbols are usually placed at the mean locations, or centroids, of the areas they represent.



A “proportional circle” map that depicts count data. Cartography by Geoff Hatchard.

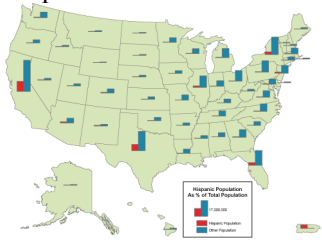
3.17. Mapping Rates and Densities

A **rate** is a proportion between two counts, such as Hispanic population as a percentage of total population. One way to display the proportional relationship between two counts is with what ArcMap calls its **Pie Chart** option. Like the Proportional Symbol map, the Pie Chart map plots a single symbol at the centroid of each geographic area by default, though users can opt to place pie symbols such that they won’t overlap each other (This option can result in symbols being placed far away from the centroid of a geographic area.) Each pie symbol varies in size in proportion to the data value it represents. In addition, however, the Pie Chart symbol is divided into pieces that represent proportions of a whole.



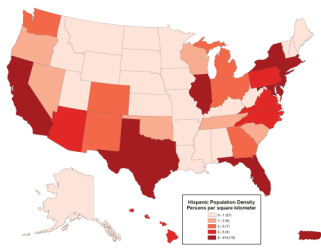
A “pie chart” map that depicts rate data. Cartography by Geoff Hatchard.

Some perceptual experiments have suggested that human beings are more adept at judging the relative lengths of bars than they are at estimating the relative sizes of pie pieces (although it helps to have the bars aligned along a common horizontal base line). You can judge for yourself by comparing the effect of ArcMap’s **Bar/Column Chart** option.



A “bar/column chart” map that depicts rate data. Cartography by Geoff Hatchard.

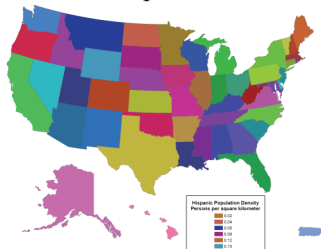
Like rates, **densities** are produced by dividing one count by another, but the divisor of a density is the magnitude of a geographic area. Both rates and densities hold true for entire areas, but not for any particular point location. For this reason **it is conventional not to use point symbols to symbolize rate and density data on thematic maps**. Instead, cartography textbooks recommend a technique that ArcMap calls “Graduated Colors.” Maps produced by this method, properly called **choropleth** maps, fill geographic areas with colors that represent attribute data values.



A “graduated color” (choropleth) map that depicts density data. Cartography by Geoff Hatchard.

Because our ability to discriminate among colors is limited, attribute data values at the ratio or interval level are usually sorted into four to eight ordinal level categories. ArcMap calls these categories **classes**. Users can adjust the number of classes, the **class break values** that separate the classes, and the colors used to symbolize the classes. Users may choose a group of predefined colors, known as a **color ramp**, or they may specify their own custom colors. Color ramps are sequences of colors that vary from light to dark, where the darkest color is used to represent the highest value range. Most textbook cartographers would approve of this, since they have long argued that it is the lightness and darkness of colors, not different color hues, that most logically represent quantitative data.

Logically or not, people prefer colorful maps. For this reason some might be tempted to choose ArcMap’s **Unique Values** option to map rates, densities, or even counts. This option assigns a unique color to each data value. Colors vary in hue as well as lightness. This symbolization strategy is designed for use with a small number of **nominal** level data categories. As illustrated in the map below, the use of an unlimited set of color hues to symbolize unique data values leads to a confusing thematic map.



A “unique values” map that depicts density data. Note that the legend, which in the original shows one category for each state, is trimmed off. Cartography by Geoff Hatchard.

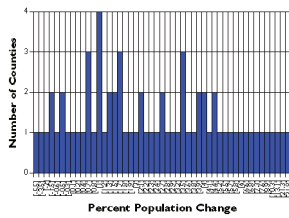
PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Thematic Map Types. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

3.18. Data Classification

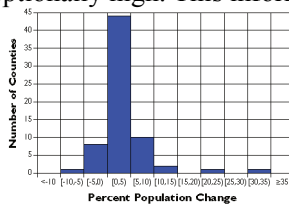
You’ve read several times already in this text that geographic data is always generalized. As you recall from Chapter 1, generalization is inevitable due to the limitations of human visual acuity, the limits of display resolution, and especially by limits imposed by the costs of collecting and processing detailed data. What we have not previously considered is that generalization is not only necessary, it is sometimes beneficial.

Generalization helps make sense of complex data. Consider a simple example. The graph below shows the percent population change for Pennsylvania’s 67 counties over a five-year period. Categories along the x axis of the graph represent each of the 49 unique percentage values (some of the counties had exactly the same rate). Categories along the y axis are the numbers of counties associated with each rate. As you can see, it’s difficult to discern a pattern in these data.



Unclassified population change rates for 67 Pennsylvania counties.

The following graph shows exactly the same data set, only grouped into 7 classes. It's much easier to discern patterns and outliers in the classified data than in the unclassified data. Notice that the mass of population change rates are distributed around 0 to 5 percent, and that there are two counties (x and y counties) whose rates are exceptionally high. This information is obscured in the unclassified data.

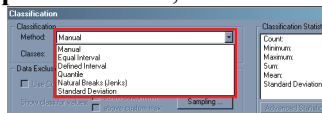


Classified population change rates for 67 Pennsylvania counties.

Data classification is a means of generalizing thematic maps. Many different data classification schemes exist. If a classification scheme is chosen and applied skillfully, it can help reveal patterns and anomalies that otherwise might be obscured. By the same token, a poorly-chosen classification scheme may hide meaningful patterns. The appearance of a thematic map, and sometimes conclusions drawn from it, may vary substantially depending on data classification scheme used.

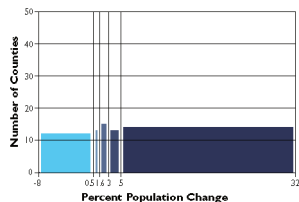
3.19. Two Classification Schemes

Many different systematic classification schemes have been developed. Some produce “optimal” classes for unique data sets, maximizing the difference between classes and minimizing differences within classes. **Since optimizing schemes produce unique solutions, however, they are not the best choice when several maps need to be compared.** For this, data classification schemes that treat every data set alike are preferred.



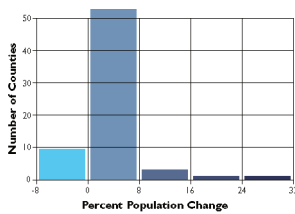
Portion of the ArcMap classification dialog box highlighting the schemes supported in ArcMap 8.2.

Two commonly used schemes are **quantiles** and **equal intervals** (“quartiles,” “quintiles,” and “percentiles” are instances of quantile classifications that group data into four, five, and 100 classes respectively). The following two graphs illustrate the differences.



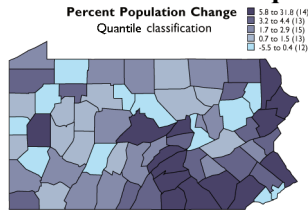
County population change rates divided into five quantile categories.

The graph above groups the Pennsylvania county population change data into five classes, each of which contains the same number of counties (in this case, approximately 20 percent of the total in each). The **quantiles** scheme accomplishes this by varying the width, or range, of each class.

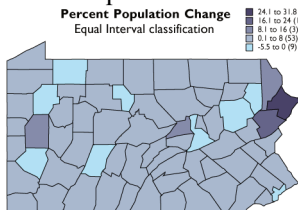


County population change rates divided into five equal interval categories.

In the second graph, the width or range of each class is equivalent (8.5 percentage points). Consequently, the number of counties in each **equal interval** class varies.



The five quantile classes mapped.



The five equal interval classes mapped.

As you can see, the effect of the two different classification schemes on the appearance of the two choropleth maps above is dramatic. The quantiles scheme is often preferred because it prevents the clumping of observations into a few categories shown in the equal intervals map. Conversely, the equal interval map reveals two outlier counties which are obscured in the quantiles map. A good point to take from this little experiment is that it is often useful to compare the maps produced by several different map classifications. Patterns that persist through changes in classification scheme are likely to be more conclusive evidence than patterns that shift.

3.20. Calculating Quantile Classes

The objective of this section is to ensure that you understand how mapping programs like ArcMap classify data for choropleth maps. First we will step through the classification of the Pennsylvania county population change data. Then you will be asked to classify another data set yourself.

Step 1: Sort the data.

Attribute data retrieved from sources like the Census Bureau's Web site are likely to be sorted alphabetically by geographic area. To classify the data set, you need to resort the data from the highest attribute data value to the lowest.

Step 2: Define the number of classes.

There are no absolute rules on this. Since our ability to differentiate colors is limited, the more classes you make, the harder they may be to tell apart. **In general, four to eight classes are used for choropleth mapping.** Use an odd number of classes if you wish to visualize departures from a central class that contains a median (or zero) value.

Step 3: Determine class breaks by dividing the number of observations by the number of classes.

For example, 67 counties divided by 5 classes yields 13.4 counties per class. Obviously, in cases like this the number of counties in each class has to vary a little. Make sure that counties having the same value are assigned to the same class, even if that class ends up with more members than other classes.

Step 4: Assign color symbols to differentiate the categories.

The illustration below shows three iterations of a data table. The first (on the left) is sorted alphabetically by county name. The middle table is sorted by percent population change, in descending order. The third table breaks the re-sorted counties into five quintile categories. Normally you would classify the data and symbolize the map using GIS software, of course. The illustration includes the colors that were used to symbolize the corresponding choropleth map on the preceding page. If you'd like to try sorting the data table illustrated below, [follow this link to open the spreadsheet file](#).

Data classification for choropleth mapping
Breaking a data table into five quintile categories for choropleth mapping.

Practice Quiz

Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Data Classification for Thematic Mapping. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

3.21. Summary

National censuses, such as the decennial census of the U.S., are among the richest sources of attribute data. Attribute data are heterogeneous. They range in character from qualitative to quantitative; from unranked categories to values that can be positioned along a continuous number line. Social scientists have developed a variety of different measurement scales to accommodate the variety of phenomena measured in censuses and other social surveys. The level of measurement used to define a particular data set influences analysts' choices about which analytical and cartographic procedures should be used to transform the data into geographic information.

Thematic maps help transform attribute data by revealing patterns obscured in lists of numbers. Different types of thematic maps are used to represent different types of data. Count data, for instance, are conventionally portrayed with symbols that are distinct from the statistical areas they represent, because counts are independent of the sizes of those areas. Rates and densities, on the other hand, are often portrayed as choropleth maps, in which the statistical areas themselves serve as symbols whose color lightness vary with the attribute data they represent. Attribute data shown on choropleth maps are usually classified. Classification schemes that facilitate comparison of map series, such as the quantiles and equal intervals schemes demonstrated in this lesson, are most common.

The U.S. Census Bureau's mandate requires it to produce and maintain spatial data as well as attribute data. In Chapter 4 we will study the characteristics of those data, which are part of a nationwide geospatial database called "TIGER."

QUIZ

Registered Penn State students should return now to the Chapter 3 folder in ANGEL (via the Resources menu to the left) to access the graded quiz for this chapter. This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter's learning objectives. This quiz consists of ten problems. You are free to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 3.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

3.22. Bibliography

Brewer, C. & Suchan, T., (2001). *Mapping census 2000: The geography of U. S. diversity*. U. S. Census Bureau, Census Special Reports, Series CENSR/01-1. Washington, D. C.: U.S. Government Printing Office.

Chrisman, N. (1997). *Exploring geographic information systems*. New York: John Wiley & Sons, Inc.

Chrisman, N. (2002). *Exploring geographic information systems*. (2nd ed.). New York: John Wiley & Sons, Inc.

Microsoft Corporation. (2006). *MapPoint 2006*. Retrieved April 27, 2006, from <http://www.microsoft.com/mappoint/default.msp>

Monmonier, M. (1995). *Drawing the line: Tales of maps and cartocontroversy*. New York: Henry Holt and Company.

Oregon State University. Information Services. (n. d.). *Government information sharing project*. Retrieved July 19, 1999, from <http://govinfo.kerr.orst.edu> (since retired).

Pennsylvania State University. University Libraries. Social Science Library. Census Extractor and Locator Sites. Retrieved July, 19, 1999, from <http://www.libraries.psu.edu/crsweb/docs/extract.htm> (since retired).

Roberts, S. (1994). *Who we are: A portrait of America based on the latest U.S. census*. New York: Times Books.

Speer, G. (1998). *The metric system*. Retrieved July 19, 1999, from <http://www.essex1.com/people/speer/metric.html> (since retired).

Steger, T. D. (1986). *Topographic maps*. Washington D.C.: U.S. Government Printing Office.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.1

U.S. Census Bureau (n. d.). Retrieved July 19, 1999, from <http://www.census.gov>

U.S. Census Bureau (1996). *Federal expenditures by state for fiscal year 1995*. Retrieved May 9, 2006, from www.census.gov/prod/2/gov/fes95rv.pdf

U.S. Census Bureau (2005). *American FactFinder* Retrieved July, 19, 1999, from <http://factfinder.census.gov>

U.S. Census Bureau (n. d.). *American FactFinder* Retrieved August 2, 2012, from http://factfinder2.census.gov/faces/nav/jsf/pages/using_factfinder5.xhtml

U.S. Census Bureau (2008). *A Compass for understanding and using American Community Survey data: What general users need to know*. U.S. Government Printing Office, Washington DC, 2008.

Chapter 4

TIGER, Topology and Geocoding

David DiBiase

4.1. Overview

In the Chapter 3 we studied the population data produced by the U.S. Census Bureau, and some of the ways those data can be visualized with thematic maps.

In addition to producing data about the U.S. population and economy, the Census Bureau is a leading producer of digital map data. The Census Bureau's Geography Division created its "Topologically Integrated Geographic Encoding and Referencing" (TIGER) spatial database with help from the U.S. Geological Survey. In preparation for the 2010 census, the Bureau conducted a database redesign project that combined TIGER with a Master Address File (MAF) database. **MAF/TIGER** enables the Bureau to associate census data, which it collects by household address, with the right census areas and voting districts. This is an example of a process called address-matching or **geocoding**.

The MAF/TIGER database embodies the vector approach to spatial representation. It uses point, line, and polygon features to represent streets, water bodies, railroads, administrative boundaries, and select landmarks. In addition to the "absolute" locations of these features, which are encoded with latitude and longitude coordinates, MAF/TIGER encodes their "relative" locations—a property called **topology**.

MAF/TIGER also includes attributes of these vector features including names, administrative codes, and, for many streets, address ranges and ZIP Codes. Vector feature sets are extracted from the MAF/TIGER database to produce reference maps for census takers and thematic maps for census data users. Such extracts are called **TIGER/Line Shapefiles**.

Characteristics of TIGER/Line Shapefiles that make them useful to the Census Bureau also make them valuable to other government agencies and businesses. Because they are not protected by copyright, TIGER/Line data have been widely adapted for many commercial uses. TIGER has been described as "the first truly useful nationwide general-purpose spatial data set" (Cooke 1997, p. 47). Some say that it jump-started a now-thriving geospatial data industry in the U.S.

Objectives

The objective of this chapter is to familiarize you with MAF/TIGER and two important concepts it exemplifies: topology and geocoding. Specifically, students who successfully complete Chapter 4 should be able to:

1. Explain how geographic entities are represented within MAF/TIGER;
2. Explain how geometric primitives in MAF/TIGER are represented in TIGER/Line Shapefile extracts;
3. Define topology and explain why and how it is encoded in TIGER;
4. Perform address geocoding; and
5. Describe how TIGER/Line files and similar products can be used for other applications, including routing and allocation.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

Concept Map

You may be interested in seeing the [concept map](#) used to guide development of Chapters 3 and 4.

4.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 4 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 4	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	<p>Submit four practice quizzes including:</p> <ul style="list-style-type: none"> • MAF and TIGER • Shapefiles • Topology • Geocoding <p>Practice quizzes are not graded and may be submitted more than once.</p>	Go to ANGEL > [your course section] > Lessons tab > Chapter 4 folder > [quiz]
3	<p>Perform “Try this” activities including:</p> <ul style="list-style-type: none"> • Explore availability of TIGER/Line Shapefile geographies and features • Download and view a TIGER/Line Shapefile • Geocode your address using a TIGER/Line Shapefile • Compare the geocoding performance of online routing services • Explore resources about the Traveling Salesman Problem <p>“Try this” activities are not graded.</p>	Instructions are provided for each activity.
4	Submit the Chapter 4 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 4 folder > Chapter 4 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

4.3. MAF/TIGER

MAF/TIGER is the Census Bureau’s geographic database system. Several factors prompted the U.S. Census Bureau to create MAF/TIGER: the need to conduct the census by mail, the need to produce wayfinding aids for census field workers, and its mission to produce map and data products for census data users.

CONDUCTING THE CENSUS BY MAIL

As the population of the U.S. increased it became impractical to have census takers visit every household in person. Since 1970, the Census Bureau has mailed questionnaires to most households with instructions that completed forms should be returned by mail. Most but certainly not all of these questionnaires are dutifully mailed—about 72 percent of all questionnaires in 2010. At that rate the Census Bureau estimates that some \$1.6 billion was saved by reducing the need for field workers to visit non-responding households.

This is the official form for all the people at this address. It is quick and easy, and your answers are protected by law.

Use a blue or black pen. **Start here**

The Census must count every person living in the United States on April 1, 2010. Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, too.

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, too:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people =

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? Mark X all that apply.

- Children, such as newborn babies or foster children
- Relatives, such as adult children, cousins, or in-laws
- Nonrelatives, such as nonrelatives or live-in baby sitters
- People staying here temporarily
- No additional people

3. Is this house, apartment, or mobile home — Mark X ONE box.

- Owned by you or someone in this household with a mortgage or loan? Include home equity loans.
- Owned by you or someone in the household free and clear without a mortgage or loan?
- Rented?
- Occupied without payment of rent?

4. What is your telephone number? We may call if we aren't understood or an answer. Area Code + Number

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If that owner or renter lives somewhere else, start with any adult living here. This will be Person 1. What is Person 1's name? First name below.

Last Name First Name MI

6. What is Person 1's sex? Mark X ONE box.

Male Female

7. What is Person 1's age and what is Person 1's date of birth? Please report below on age if other than child or teen (see page 1 year old). Fill in quarters in boxes.

Age on April 1, 2010 Month Day Year of birth

NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- No, not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, whether Hispanic, Latino, or Spanish origin — Please include examples: Argentinian, Colombian, Dominican, Honduran, Salvadoran, Venezuelan, and so on.

9. What is Person 1's race? Mark X one or more boxes.

- White
- Black, African Am., or Negro
- American Indian or Alaska Native — First name of tribal or pretribal.
- Asian Indian
- Japanese
- Native Hawaiian
- Chinese
- Korean
- Guamanian or Chamorro
- Filipino
- Vietnamese
- Samoan
- Other Asian — Please list (for example, Hmong, Laotian, Thai, etc.)
- Other Pacific Islander — Please list (for example, Fijian, Tongan, Tokelauan, and so on).
- Some other race — Please race.

10. Does Person 1 sometimes live or stay somewhere else?

- No
- Yes — Mark X all that apply.
- In college housing
- For child custody
- In the military
- In jail or prison
- At a seasonal or second residence
- In a nursing home or retirement institution
- For another reason

If more people were counted in Question 1, continue with Person 2.

U.S. DEPARTMENT OF COMMERCE
Economic and Statistics Administration
U.S. CENSUS BUREAU

2010 Census questionnaire. For a question-by-question tour, go here.

To manage its mail delivery and return operations, the Census Bureau relies upon a **Master Address File (MAF)**. MAF is a complete inventory of housing units and many business locations in the U.S., Puerto Rico, and associated island areas. MAF was originally built from the U.S. Postal Service's Delivery Sequence File of all residential addresses. The MAF is updated through both corrections from field operations and a Local Update of Census Address (LUCA) program by which tribal, state, and local government liaisons review and suggest updates to local address records. **"MAF/TIGER" refers to the coupling of the Master Address File with the TIGER spatial database**, which together enable the Census Bureau to efficiently associate address-referenced census and survey data received by mail with geographic locations on the ground and tabulation areas of concern to Congress and many governmental agencies and businesses.

It's not as simple as it sounds. Postal addresses do not specify geographic locations precisely enough to fulfill the Census Bureau's constitutional mandate. An address is not a position in a grid coordinate system—it is only one in a series of ill-defined positions along a route. The location of an address is often ambiguous because street names are not unique, numbering schemes are inconsistent, and because routes have two sides, left and right. Location matters, as you recall, because **census data must be accurately georeferenced to be useful for reapportionment, redistricting, and allocation of federal funds**. Thus the Census Bureau had to find a way to assign address referenced data automatically to particular census blocks, block groups, tracts, voting districts, and so on. That's what the "Geographic Encoding and Referencing" in the TIGER acronym refers to.

MAPS FOR CENSUS FIELD WORKERS

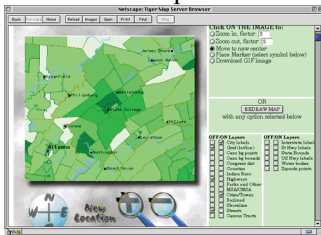
A second motivation that led to MAF/TIGER was the need to help census takers find their way around. Millions of households fail to return questionnaires by mail, after all. Census takers (called "enumerators" at the Bureau) visit non-responding households in person. **Census enumerators need maps showing streets and select landmarks to**

help locate households. Census supervisors need maps to assign census takers to particular territories. Field notes collected by field workers are an important source of updates and corrections to the MAF/TIGER database.

Prior to 1990, the Bureau relied on local sources for its maps. For example, 137 maps of different scales, quality, and age were used to cover the 30-square-mile St. Louis area during the 1960 census. The need for maps of consistent scale and quality forced the Bureau to become a map maker as well as a map user. Using the MAF/TIGER system, Census Bureau geographers created over 17 million maps for a variety of purposes in preparation for the 2010 Census.

DATA PRODUCTS

The Census Bureau's mission is not only to collect data, but also to make data products available to its constituents. In addition to the attribute data considered in Chapter 3, the Bureau disseminates a variety of geographic data products, including wall maps, atlases, and one of the earliest on-line mapping services, the TIGER Mapping Service. You can explore the [Bureau's maps and cartographic data products here](#).



Launched in 1995, the TIGER Mapping Service was one of the earliest Internet map services. Registered students will use its successor, American Factfinder, in Project 2.

MAF/TIGER DATABASE REDESIGN

The Census Bureau conducted a major redesign of the MAF/TIGER database in the years leading up to the 2010 decennial census. What were separate, homegrown database systems (MAF and TIGER) are now unified in the industry-standard Oracle relational database management system. Benefits of this “commercial off-the-shelf” (COTS) database software include concurrent multi-user access, greater user familiarity, and better integration with Web development tools. As Galdi (2005) explains in his white paper “Spatial Data Storage and Topology in the Redesigned MAF/TIGER System,” the redesign “mirrors a common trend in the Information Technology (IT) and Geographic Information System (GIS) industries: the integration of spatial and non-spatial data into a single enterprise data set” (p. 2).

Concurrent with the MAF/TIGER redesign, the Census Bureau also updated the distribution format of its TIGER/Line map data extracts. Consistent with the Bureau's COTS strategy, it adopted the defacto standard Esri “Shapefile” format. The following pages consider characteristics of the spatial data stored in MAF/TIGER and in TIGER/Line Shapefile extracts.

PODCAST

Hear more about [how the Census Bureau's Geography Division uses MAF/TIGER and related tools to create maps for the 2010 Census](#).

4.4. Vector Extracts from MAF/TIGER

The Census Bureau began to develop a digital geographic database of 144 metropolitan areas in the 1960s. By 1990,

the early efforts had evolved into **TIGER**: a seamless digital geographic database that covered the whole of the United States and its territories. As discussed in the previous page, MAF/TIGER succeeded TIGER in the lead-up to the 2010 Census.

TIGER/Line Shapefiles are digital map data products extracted from the MAF/TIGER database. They are freely available from the Census Bureau, and are suitable for use by individuals, businesses and other agencies that don't have direct access to MAF/TIGER.

This section outlines the geographic entities represented in the MAF/TIGER database, describes how a particular implementation of the vector data model is used to represent those entities, and considers the accuracy of digital features in relation to their counterparts on the ground. The following page considers characteristics of the "Shapefile" data format used to distribute digital extracts from MAF/TIGER.

GEOGRAPHIES REPRESENTED IN TIGER AND SHAPEFILE EXTRACTS

The MAF/TIGER database is selective. Only those geographic entities needed to fulfill the Census Bureau's operational mission are included. Entities that don't help the Census Bureau conduct its operations by mail, or help field workers navigate a neighborhood, are omitted. Terrain elevation data, for instance, are not included in MAF/TIGER. A comprehensive list of the "feature classes" and "superclasses" included in MAF/TIGER and Shapefiles can be found in Appendix F of the [TIGER/Line Shapefiles Technical Documentation](#). **Examples of superclasses include:**

- Potential living quarters (e.g., sites of shelters, retirement homes, prisons, dormitories)
- Road/path features (e.g., primary roads, secondary roads, local neighborhood roads)
- Hydrographic features (e.g., stream/river, lake/pond, ocean/sea)
- Miscellaneous linear features (e.g., pipeline, powerline, fence line)
- Tabulation areas (e.g., county or equivalent, tract, block group, block)

Excerpt from TIGER/Line Technical Documentation

MTFCC	FEATURE CLASS	SUPERCLASS	POINT	LINEAR	AREAL	FEATURE CLASS DESCRIPTION
\$1400	Local Neighborhood Road, Rural Road, City Street	Road/Path Features	N	Y	N	Generally a paved non-arterial street, road, or byway that usually has a single lane of traffic in each direction. Roads in this feature class may be privately or publicly maintained. Scenic park roads would be included in this feature class, as would (depending on the region of the country) some unpaved roads.
\$1500	Vehicular Trail (4WD)	Road/Path Features	N	Y	N	An unpaved dirt trail where a four-wheel drive vehicle is required. These vehicular trails are found almost exclusively in very rural areas. Minor, unpaved roads usable by ordinary cars and trucks belong in the \$1400 category.
\$1630	Ramp	Road/Path Features	N	Y	N	A road that allows controlled access from adjacent roads onto a limited access highway, often in the form of a cloverleaf interchange. These roads are unaddressable.

Excerpt from TIGER/Line Technical Documentation (Census Bureau 2012) showing some of the feature classes included in the “Road/Path Features” superclass.

Note also that **neither the MAF/TIGER database nor TIGER/Line Shapefiles include the population data collected through questionnaires and by census takers.** MAF/TIGER merely provides the geographic framework within which address-referenced census data are tabulated.

TRY THIS!

EXPLORING AVAILABLE TIGER/LINE SHAPEFILES

In this Try This (One of 3 dealing with TIGER/Line Shapefiles) you are going to explore which TIGER/Line Shapefiles are available for download at various geographies and what information those files contain. We will be exploring the 2009 and 2010 versions of the TIGER/Line Shapefile data sets. Versions from other years are available. Feel free to investigate those, too.

- Follow [this link to get to the TIGER Products page](#) of the Census Bureau web site, then follow the **TIGER/Line Shapefiles** link found under **Which product should I use?** to get to the Geography page.
- Link to the 2010 TIGER/Line Shapefiles via the **2010** tab link.
- Select **Download**, and then from the expanded list choose **Web Interface**.
- Expand the pick list under **Select a layer type**. Spend some time choosing different entries from the layer pick list and then using the **Submit** button to navigate through the sub layers taking note of when you are offered access to a Download button. Take note of a couple of things. (1) Some of the pick lists make a selection available that allows you to download a shapefile dataset for the entire country. (2) For some of the choices you must navigate to the County level before the Download button is available

As stated above we want you to get a sense of the sorts of data that are available for the various geographies — from the county to the national level. Perusing the various layers as I had you doing above makes it difficult to make an overall assessment of what data there is at a given geographic scale. Fortunately for our purposes the Census has provided a convenient table to help us in this regard.

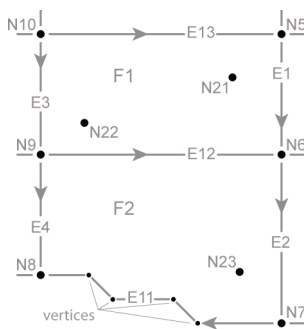
- You should still be on the 2010 TIGER/Line Shapefiles | Select a layer type page. Click on the **Documentation** link in the upper right portion of the page. This will take you back to the Geography page.
- Select the **2010** tab again.
- Select **File Availability**. Study the table that appears.
- Note that there are columns titled *State- and County-based Files*, *Nation-based Files*, and *American Indian Area-based Files*.
- Compare which geographies (the *Layer* column) are available in the *Nation-Based Files* category to those available in the *State-Based Files* category.
What files are available for a state that are not available for the whole nation? Can you think of reasons why these are not available as a single national file? Post a comment below to discuss with your fellow students.

- Now, compare the *State-Based Files* category to the *County-Based Files* category. What files available at the state level are also available at the county-level? Once again, share your thoughts with your peers.

GEOMETRIC PRIMITIVES

Like other implementations of the vector data model, MAF/TIGER represents geographic entities using geometric primitives including nodes (point features), edges (linear features), and faces (area features). These are defined and illustrated below.

- **Nodes** (labeled “N” in the illustration below) are “0-dimensional,” consisting only of a single pair of latitude and longitude coordinates.
 - Nodes N21-23 are **isolated nodes**. That is, they are not end points of edges.
- **Edges** (labeled “E” in the illustration below) are 1-dimensional linear primitives used to represent streets, railroads, pipelines, and rivers.
 - The end points of an edge are called **connecting nodes**.
 - Each edge is assigned a direction, denoted by the arrowheads. The directionality of the edge allows the designation of a **Start Node** and an **End Node**. The Start Node of edge E12 below is N9, and the End Node is N6.
 - An edge may have intermediate points called **vertices** that define its shape.
- **Faces** (labeled “F” in the illustration below) are the 2-dimensional geometric primitives used to represent entities like blocks, counties, and voting districts. A face is a polygon bounded by edges.
 - The directionality of an edge also allows **left and right faces** to be designated. Face F1 is on the left of edge E12 and face F2 is to the right.



Geometric primitives of the Topologically Integrated Geographic Encoding and Referencing (TIGER) database. The figure shows what might be two adjacent Census blocks, with the bottom block bounded on the south by a river. The remaining edges might correspond to streets, and the isolated nodes might be landmarks such as a school, a church and a zoo.

GEOMETRIC ACCURACY

Until recently the geometric accuracy of the vector features encoded in TIGER were notoriously poor (see illustration below). How poor? **Through 2003**, the [TIGER/Line metadata](#) stated that

“

Coordinates in the TIGER/Line files have six implied decimal places, but the positional accuracy of these coordinates is not as great as the six decimal places suggest. The positional accuracy varies with the source materials used, but generally the information is no better than the established National Map Accuracy standards for 1:100,000-scale maps from the U.S. Geological Survey (Census Bureau 2003)

TRY THIS!

Having performed scale calculations in Chapter 2 you should be able to calculate the magnitude of error (ground distance) associated with 1:100,000-scale topographic maps. Recall that the allowed error for USGS topographic maps at scales of 1:20,000 or smaller is 1/50 inch (see the [nationalmap standards pdf](#))



Discrepancy between pre-modernization TIGER/Line file streets (red) and actual geometry of street network shown in an orthorectified aerial image (U.S. Census Bureau n.d).

ACCURACY IMPROVEMENT

Starting in 2002, in preparation for the 2010 census, the Census Bureau commissioned a six-year, \$200 million MAF/TIGER Accuracy Improvement Project (MTAIP). **One objective of the effort was to use GPS to capture accurate geographic coordinates for every household in the MAF. Another objective was to improve the accuracy of TIGER’s road/path features.** The project aimed to adjust the geometry of street networks to align within 7.6 meters of street intersections observed in orthoimages or measured using GPS. The corrected streets are necessary not just for mapping, but for accurate geocoding. Because streets often form the boundaries of census areas, it is essential that accurate household locations are associated with accurate street networks.

MTAIP integrated over 2,000 source files submitted by state, tribal, county, and local governments. Contractors used survey-grade GPS to evaluate the accuracy of a random sample of street centerline intersections of the integrated source files. The evaluation confirmed that most but not all features in the spatial database equal or exceed the 7.6 meter target. Uniform accuracy wasn’t possible due to the diversity of local source materials used, though this accuracy is the standard in the “All Lines” Shapefile extracts. The geometric accuracy of particular feature classes included in particular shapefiles are documented in the metadata associated with that shapefile extract.

MTAIP was completed in 2008. In conjunction with the continuous American Community Survey and other

census operations, corrections and updates are now ongoing. TIGER/Line Shapefile updates are now released annually.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 4 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about MAF and TIGER.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

4.5. Shapefiles

Since 2007, TIGER/Line extracts from the MAF/TIGER database have been distributed in shapefile format. Esri introduced shapefiles in the early 1990s as the native digital vector data format of its ArcView software product. The shapefile format is proprietary, but open; its [technical specifications are published](#) and can be implemented and used freely. Largely as a result of ArcView's popularity, shapefile has become a de facto standard for creation and interchange of vector geospatial data. The Census Bureau's adoption of Shapefile as a distribution format is therefore consistent with its overall strategy of conformance with mainstream information technology practices.

ELEMENTS OF A SHAPEFILE DATA SET

The first thing GIS pros need to know about shapefiles is that **every shapefile data set includes a minimum of three files**. One of the three required files stores the geometry of the digital features as sets of vector coordinates. A second required file holds an index that, much like the index in a book, allows quicker access to the spatial features and therefore speeds processing of a given operation involving a subset of features. The third required file stores attribute data in dBASE© format, one of the earliest and most widely-used digital database management system formats. All of the files that make up a Shapefile data set have the same root or prefix name, followed by a three-letter suffix or file extension. The list below shows the names of the three required files making up a shapefile data set named "counties." Take note of the file extensions.

- counties.shp: The main shape file, containing vector coordinate data
- counties.shx: The index file
- counties.dbf: The dBASE table

Esri lists twelve additional optional files, and practitioners are able to include still others. Two of the most important optional files are the ".prj" file, which includes the coordinate system definition, and ".xml", which stores metadata. (Why do you suppose that something as essential as a coordinate system definition is considered "optional"?)

TRY THIS!

DOWNLOADING AND VIEWING A TIGER/LINE SHAPEFILE

In this *Try This!* (the second of 3 dealing with TIGER/Line Shapefiles), you will download a TIGER/Line Shapefile dataset, investigate the file structure of a typical Esri shapefile, and view it in GIS software.

You can use a free software application called **Global Mapper** (originally known as **dlgv32 Pro**) to investigate TIGER/Line shapefiles. Originally developed by the staff of the USGS Mapping Division at Rolla, Missouri as a data viewer for USGS data, Global Mapper has since been commercialized, but is available in a free trial version.

The instructions below will guide you through the process of installing the software and opening the TIGER/Line data.

1. Downloading TIGER/Line Shapefiles: You are going to use the 2010 TIGER/Line Shapefiles.

- Return to the [2010 TIGER/Line Shapefiles download page](#).
- From the *Select a layer type* pick list, under *Features*, choose **All Lines** and click **submit**. (You are welcome to download and investigate any TIGER/Line Shapefile(s), but we will use an *All Lines* dataset in the geocoding Try This later in the chapter, so your downloading one here will make you more familiar with the content.)
- From the All Lines pick list select a state or territory and click **Submit**.
- Select a County from the next pick list that appears and click **Download**.
- Save the file to your computer.

The file you download should have a name like *tl_2010_42027_edges.zip*. The root name of this file, *tl_2010_42027_edges* in this example, will also be the name of the shapefile dataset.

The *42027* is a federal code that represents Pennsylvania (state 42) and Centre County (county 027). The five-digit code in your file name will depend on which state and county you selected.

- The data are compressed in a .zip archive. Extract the data to a new named folder in a known location. (Within the file hierarchy that is extracted there may be a second .zip file that needs to be uncompressed.)

2. Investigating the shapefile data set:

- Navigate to *within* the folder in which you stored your uncompressed TIGER/Line Shapefile dataset.
- Notice the multiple files which make up the shapefile dataset, including:
 - i. *tl_2010_42027_edges.shp*, containing the vector coordinate data
 - i. *tl_2010_42027_edges.shp.xml*, containing metadata
 - i. *tl_2010_42027_edges.shx*, the index file
 - i. *tl_2010_42027_edges.dbf*, the dBASE file
 - i. *tl_2010_42027_edges.prj*, containing the projection/spatial reference
- All of the files work in concert to store the necessary components of the Esri *shapefile data set*. You may be familiar with some of the individual files types. The contents of three of them can be easily viewed. Let's open those three. You can double click on the file and then select "from a list of installed programs," or you may need to run the suggested application and open the file from within it. Let me know if you need help, or help each other in the ANGEL Chapter 4 Discussion Forum or in the Comments area below.
 - i. Open the **.dbf** file using Microsoft Excel.
Note the typical row-column structure of a flat-file database. Can you find the four columns, or fields, that hold the address range information? Look for LFROMADD, etc. The field name LFROMADD is shorthand for Left From Address. The 10-character length of the field name points up one of the constraints of the dBASE format—field names are limited to 10 characters.
 - i. Open the **.xml** file using your web browser.
You should see the metadata information bracketed by *tags* contained within

directional brackets <>. XML stands for Extensible Markup Language, and is a common set of rules for encoding documents. Can you locate the portion of the document having to do with horizontal spatial accuracy? (Spatial accuracy metadata is available when you've chosen the *All Lines* file as your candidate shapefile.)

- i. Open the **.prj** file using Notepad, or any vanilla text editor.
There are five pieces of information in this file, separated by commas. What are they? They should reinforce some of what you learned in Chapter 2 regarding what defines a geographic coordinate system.
- i. The **.shp** and **.shx** files are proprietary and specific to the functionality of the shapefile data set.
- Discuss what you find with your classmates in comments below.
- Note that one should not alter the contents of any of these files with any application other than a GIS program that is designed for that task.

3. Viewing the shapefile dataset in Global Mapper:

- Download and install the Global Mapper software:
 - i. Navigate to the [Blue Marble Global Mapper site](#).
 - i. Download the trial version of the software
 - i. Double-click on the setup file you downloaded to install the program
 - i. Launch the Global Mapper program
- After opening the Global Mapper software, choose *Open Data File(s)...* under the *File* menu, or click the “Open Your Own Data Files” button in the center of the window. Navigate to the extracted shapefile dataset you downloaded above and open it. (Remember, your complete shapefile data set will have a name similar *totl_2010_42027_edges*. It will show up in the *Open* dialog with a *.shp* extension.)
- You should be able to see all of the line features (the *edges*, from the MAF/TIGER database) contained in your county. If you are using the newest version of Global Mapper you should be able to discern roads from rivers/streams from administrative boundaries, etc. In older versions of the application the default view showed all line features in a single color and line weight, so the user needed to use the symbolization tools to make the different classes of features distinguishable.

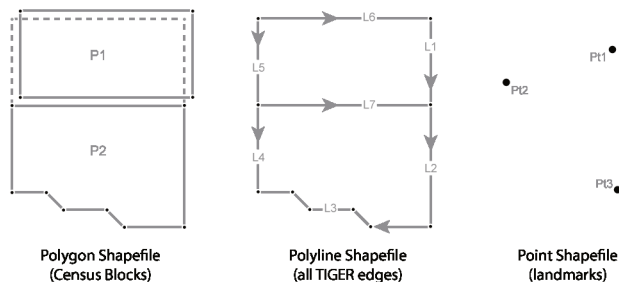
What do you think has to be understood by the mapping application to allow it to automatically symbolize features differently? Post your thoughts below.

SHAPEFILE PRIMITIVES

A single shapefile data set can contain one of three types of spatial data primitives, or features – points, lines or polygons (areas). The technical specification defines these as follows:

- **Points:** A point consists of a pair of double-precision coordinates in the order X,Y.
- **Lines:** More specifically a polyline, is an ordered set of points, or vertices, that consists of one or more **parts**. A part is a connected sequence of two or more points. Parts may or may not be connected to one another. Parts may or may not intersect one another.

- **Polygons:** A polygon consists of one or more **rings**. A ring is a connected sequence of four or more points, or vertices, that form a closed, non-self-intersecting loop.
- **Other:** M (measured; route data) and Z (3D; vertical datum) versions of point, polyline and polygon Shapefile data sets can be created, but are not included in the TIGER/Line Shapefile extracts.



Three Shapefile data sets that could be extracted from the MAF/TIGER data depicted on the preceding page

At left in the illustration above, a **polygon Shapefile data set** holds the Census blocks in which the edges from the MAF/TIGER database have been combined to form two distinct polygons, P1 and P2. The diagram shows the two polygons separated to emphasize the fact that what is the single E12 edge in the MAF/TIGER database (see the diagram on page 4) is now present in each of the Census block polygon features.

In the middle of the illustration, a **polyline Shapefile data set** holds seven line features (L1-7) that correspond to the seven edges in the MAF/TIGER database. The directionality of the line features that represent streets corresponds to address range attributes in the associated dBASE© table. Vertices define the shape of a polygon or a line, and the Start and End Nodes from the MAF/TIGER database are now First and Last Vertices.

Finally, at right in the illustration above, a **point Shapefile data set** holds the three isolated nodes from the MAF/TIGER database.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 4 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Shapefiles.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

4.6. Topology

Topology is different than topography. (You'd be surprised how often these terms get mixed up.) In Chapter 2 you read about the various ways that *absolute* positions of features can be specified in a coordinate system, and how those coordinates can be projected or otherwise transformed. Topology refers to the *relative* positions of spatial features. **Topological relations among features**—such as containment, connectivity, and adjacency—**don't change when a dataset is transformed**. For example, if an isolated node (representing a household) is located inside a face (representing a congressional district) in the MAF/TIGER database, you can count on it remaining inside that face no matter how you might project, rubber-sheet, or otherwise transform the data. Topology is vitally important to the Census Bureau, whose constitutional mandate is to accurately associate population counts and characteristics with political districts and other geographic areas.

As David Galdi (2005) explains in his white paper “Spatial Data Storage and Topology in the Redesigned MAF/TIGER System,” the “TI” in TIGER stands for “Topologically Integrated.” This means that **the various features represented in the MAF/TIGER database**—such as streets, waterways, boundaries, and landmarks (but

not elevation!)—**are not encoded on separate “layers.”** Instead, features are made up of a small set of geometric primitives—including 0-dimensional nodes and vertices, 1-dimensional edges, and 2-dimensional faces—without redundancy. That means that where a waterway coincides with a boundary, for instance, MAF/TIGER represents them both with one set of edges, nodes and vertices. The attributes associated with the geometric primitives allow database operators to retrieve feature sets efficiently with simple spatial queries. The separate feature-specific TIGER/Line Shapefiles published at the county level (such as point landmarks, hydrography, Census block boundaries, and, the “All Lines” file you are using in the multi-part “Try This”) were extracted from the MAF/TIGER database in that way. Notice, however, that when you examine a hydrography shapefile and a boundary shapefile, you will see redundant line segments where the features coincide. That fact confirms that **TIGER/Line Shapefiles**, unlike the MAF/TIGER database itself, **are not topologically integrated**. Desktop computers are now powerful enough to calculate **topology “on the fly”** from shapefiles or other non-topological data sets. However, the large batch processes performed by the Census Bureau still benefit from the MAF/TIGER database’s **persistent topology**.

MAF/TIGER’s topological data structure also benefits the Census Bureau by allowing it to automate error-checking processes. By definition, features in the TIGER/Line files conform to a set of topological rules (Galdi 2005):

1. Every edge must be bounded by two nodes (start and end nodes)..
2. Every edge has a left and right face.
3. Every face has a closed boundary consisting of an alternating sequence of nodes and edges.
4. There is an alternating closed sequence of edges and faces around every node.
5. Edges do not intersect each other, except at nodes.

Compliance with these topological rules is an aspect of data quality called **logical consistency**. In addition, the boundaries of geographic areas that are related hierarchically—such as blocks, block groups, tracts, and counties—are represented with common, non-redundant edges. Features that do not conform to the topological rules can be identified automatically, and corrected by the Census geographers who edit the database. Given that the MAF/TIGER database covers the entire U.S. and its territories, and includes many millions of primitives, the ability to identify errors in the database efficiently is crucial.

So how does topology help the Census Bureau assure the accuracy of population data needed for reapportionment and redistricting? To do so, the Bureau must aggregate counts and characteristics to various geographic areas, including blocks, tracts, and voting districts. This involves a process called “address matching” or “address geocoding” in which data collected by household is assigned a topologically-correct geographic location. The following pages explain how that works.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 4 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Topology.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

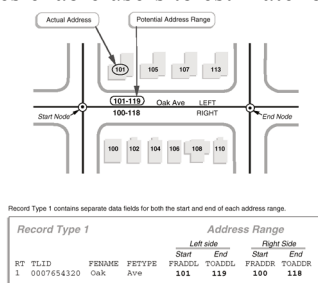
4.7. Geocoding

Geocoding is the process used to convert location codes, such as street addresses or postal codes, into geographic (or other) coordinates. The terms “address geocoding” and “address mapping” refer to the same process. Geocoding address-referenced population data is one of the Census Bureau’s key responsibilities.

However, as you know, it's also a very popular capability of online mapping and routing services. In addition, geocoding is an essential element of a suite of techniques that are becoming known as "business intelligence." We'll look at applications like these later in this chapter, but first let's consider how the Census Bureau performs address geocoding.

ADDRESS GEOCODING AT THE U.S. CENSUS

Prior to the MAF/TIGER modernization project that led up to the decennial census of 2010, the TIGER database did not include a complete set of point locations for U.S. households. Lacking point locations, TIGER was designed to support address geocoding by approximation. As illustrated below, the pre-modernization TIGER database included **address range attributes** for the edges that represent streets. Address range attributes were also included in the TIGER/Line files extracted from TIGER. Coupled with the Start and End nodes bounding each edge, address ranges enable users to estimate locations of household addresses.



How address range attributes were encoded in TIGER/Line files (U.S. Census Bureau 1997). Address ranges in contemporary TIGER/Line Shapefiles are similar, except that "From" (FR) and "To" nodes are now called "Start" and "End". Also, changes have been made to field (column) names in the attribute tables. Compare the names of the address range fields that you looked at in the second Try This exercise to those above.

Here's how it works. The diagram above highlights an edge that represents a one-block segment of Oak Avenue. The edge is bounded by two nodes, labeled "Start" and "End." A corresponding record in an attribute table includes the unique ID number (0007654320) that identifies the edge, along with starting and ending addresses for the left (FRADDL, TOADDL) and right (FRADDR, TOADDR) sides of Oak Avenue. Note also that the address ranges include potential addresses, not just existing ones. This is to make sure that the ranges will remain valid as new buildings are constructed along the street.

A common geocoding error occurs when Start and End designations are assigned to the wrong connecting nodes. You may have read in Galdi's (2005) white paper "Spatial Data Storage and Topology in the Redesigned MAF/TIGER System," that in MAF/TIGER, "an arbitrary direction is assigned to each edge, allowing designation of one of the nodes as the Start Node, and the other as the End Node" (p. 3). If an edge's "direction" happens not to correspond with its associated address ranges, a household location may be placed on the wrong side of a street.

Although many local governments in the U.S. have developed their own GIS "land bases" with greater geometric accuracy than pre-modernization TIGER/Line files, similar address geocoding errors still occur. Kathryn Robertson, a GIS Technician with the City of Independence, Missouri (and a student in the Fall 2000 offering of this course) pointed out how important it is that Start (or "From") nodes and End (or "To") nodes correspond with the low and high addresses in address ranges. "I learned this the hard way," she wrote, "geocoding all 5,768 segments for the city of Independence and getting some segments backward. When address matching was done, the locations were not correct. Therefore, I had to go back and look at the direction of my segments. I had a rule of thumb, all east-west streets were to start from west and go east; all north-south streets were to start from the south and go north" (personal communication).

Although this may have been a sensible strategy for the City of Independence, can you imagine a situation in

which Kathryn's rule-of-thumb might not work for another municipality? If so, and if you're a registered student, please add a comment to this page.

AFTER MAF/TIGER MODERNIZATION

If TIGER had included accurate coordinate locations for every household, and correspondingly accurate streets and administrative boundaries, geocoding census data would be simple and less error-prone. Many local governments digitize locations of individual housing units when they build GIS land bases for property tax assessment, E-911 dispatch and other purposes. The MAF/TIGER modernization project begun in 2002 aimed to accomplish this for the entire nationwide TIGER database in time for the 2010 census. The illustration below shows the intended result of the modernization project, including properly aligned streets, shorelines, and individual household locations, shown here in relation to an orthorectified aerial image.



Intended accuracy and completeness of modernized TIGER data in relation to the real world. TIGER streets (yellow), shorelines (blue), and housing unit locations (red) are superimposed over an orthorectified aerial image. (U.S. Census Bureau n.d.). National coverage of housing unit locations and geometrically-accurate streets and other features were not available in 2000 or before.

The modernized MAF/TIGER database described by Galdi (2005) is now in use, including precise geographic locations of over 100 million household units. However, because **household locations are considered confidential**, users of TIGER/Line Shapefiles extracted from the MAF/TIGER database still must rely upon address geocoding using address ranges.

LEVERAGING TIGER/LINE DATA FOR PRIVATE ENTERPRISE

Launched in 1996, MapQuest was one of the earliest online mapping, geocoding and routing services. MapQuest combined the capabilities of two companies: a cartographic design firm with long experience in producing road atlases, "TripTiks" for the American Automobile Association, and other map products, and a start-up company that specialized in custom geocoding applications for business. Initially, MapQuest relied in part on TIGER/Line street data extracted from the pre-modernization TIGER database. MapQuest and other commercial firms were able to build their businesses on TIGER data because of the U.S. government's wise decision not to restrict its reuse. It's been said that this decision triggered the rapid growth of the U.S. geospatial industry.

Later on in this chapter we'll visit MapQuest and some of its more recent competitors. Next, however, you'll have a chance to see how geocoding is performed using a TIGER/Line data in a GIS.

4.8. Geocoding with TIGER/Line Shapefiles

TRY THIS!

GEOCODING IN A GIS

Part 3 of 3 in the TIGER/Line Shapefile *Try This!* series is not interactive but instead illustrates how the address ranges encoded in TIGER/Line Shapefiles can be used to pinpoint (more or less!) the geographic locations of street addresses in the U.S.

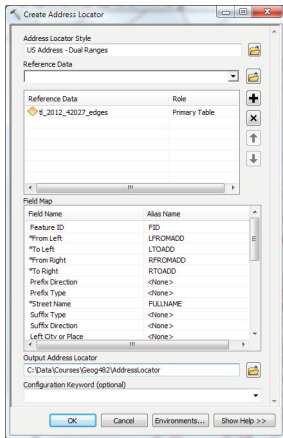
The process of geocoding a location within a GIS begins with a line dataset (shapefile) with the necessary address range attributes. The following image is an example of the attribute table of a TIGER/Line shapefile.

FULLNAME	SMD	LFROMADD	LTOADD	RFROMADD	RTOADD	ZIPL	ZIPR
Fairway Rd	337	700	750	701	750	16003	16003
Fairway Rd	337	600	620	601	631	16003	16003
Fairway Rd	337	500	500	501	500	16003	16003
Fairwood Ln	337			501	500	16003	16003
Fairwood Ln	337	1000	1000			16003	16003
Fairwood Ln	337	2002	2042	2001	2015	16003	16003
Fairwood Ln	337	2044	2000	2017	2000	16003	16003
Fairwood Ln	337	1000	2000	1007	1000	16003	16003
Falcon Ln	337	101	100	100	100	16023	16023

Visible in this image are just a few rows, which represent a handful of road segments and their corresponding address ranges. This shapefile contains over 29,000 road segments in total. Note the names of some of the attributes:

- FULLNAME – The street name of the road segment
- LFROMADD – The address number at the beginning of the road segment on the left side of the street
- LTOADD – The address number at the end of the road segment on the left side of the street
- RFROMADD – The address number at the beginning of the road segment on the right side of the street
- RTOADD – The address number at the end of the road segment on the right side of the street
- ZIPL – The zip code area that is present to the left side of the road segment
- ZIPR – The zip code area that is present to the right side of the street

Next, the GIS software needs to know which of these attributes contains each piece of the necessary address range information. Some shapefiles use different names for their attributes, so the GIS can't always know which attribute contains the Right-Side-From-Address information, for example. In ArcGIS, for example, something called a Locator is configured that maps the attributes in the shapefile to the corresponding piece of necessary address information. The image below illustrates what this mapping looks like:



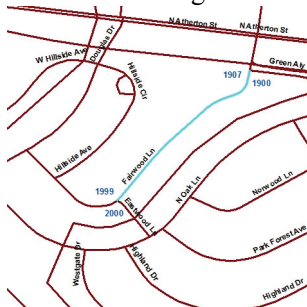
Note the items with an asterisk (*). These are the minimum required attributes that need to be present in the shapfile for the geocoding to work. The items in the “Alias Name” column correspond to attributes in the shapefile.

We are now ready to find a location by searching for a street address! Let’s geocode the location for “1971 Fairwood Lane, 16803”.

When an address is specified, the GIS queries the attribute table to find rows with a matching street name in the correct zipcode. Also, the particular segment of the street that contains the address number is identified. The below image shows the corresponding selection in the attribute table:

FULLNAME	SMD	L*FROMADD	L*TOADD	R*FROMADD	R*TOADD	ZIPL	ZIPR
Fairway Rd	337	700	790	701	799	16803	16803
Fairway Rd	337	650	628	601	631	16803	16803
Fairway Rd	337	500	590	501	599	16803	16803
Fairwood Ln	337			561	1905	16803	
Fairwood Ln	337	1800	1898	591	599	16803	
Fairwood Ln	337	2002	2042	2001	2015	16803	16803
Fairwood Ln	337						
Fairwood Ln	337	2044	2098	2017	2099	16803	16803
Fairwood Ln	337	1900	2000	1907	1999	16803	16803
Fairwood Ln	337						
Falcon Ln	337	101	199	100	198	16823	16823
Falcon Pointe Dr							

The image below shows the corresponding road segment highlighted on a map. The To and From address values for the road segment have been added so you can see the range of addresses.



Finally, the GIS interpolates where along the road segment the value of 1971 occurs and places it on the appropriate side of the street based on the even/odd values indicated in the attribute table. The image below shows the final result of the geocoding process:



The accuracy of a geocoded location is dependent on a number of factors, including the quality of the line work in a shapefile, the accuracy of the address range attributes of each road segment, and the interpolation performed by the software. As you may see in the following section, different geocoding services may provide different location results due to the particular data and procedures used.

4.9. Geocoding Online

No doubt you're familiar with one or more popular online mapping services. How well do they do at geocoding the location of a postal address? You can try it out for yourself at several Web-based mapping services, including MapQuest.com, Microsoft's Bing Maps, and Tele Atlas/TomTom's Geocode.com. Tele Atlas, for example, is a leading manufacturer of digital street data for vehicle navigation systems. To accommodate the routing tasks that navigation systems are called upon to serve, the streets are encoded as vector features whose attributes include address ranges. (In order to submit an address for geocoding at Geocode.com you have to set up a trial account through their EZ-Locate Interactive web tool or download the EZ-Locate software).

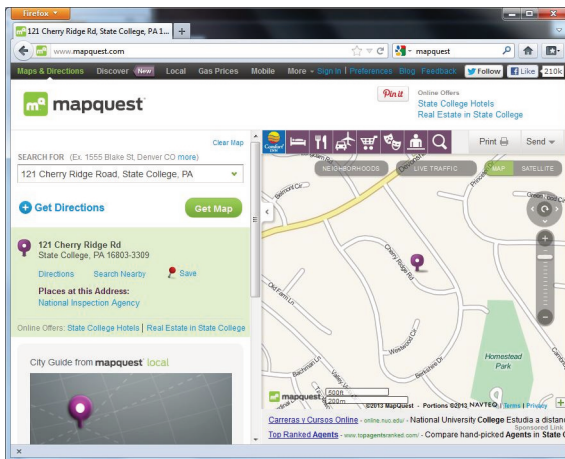
Submitting an address to Tele Atlas' Geocode.com service for geocoding. © 2013 TomTom North America, Inc. All rights reserved.

Shown above is the form by which you can geocode an address to a location in a Tele Atlas street database. The result is shown below.

Matched Address	Location (WGS-84)	2010 Census
121 CHERRY RIDGE RD STATE COLL, PA 16803	Decimal Degrees Long/Min/Sec Lat: 40.796227 40.47146 7778 Lon: -77.806634 77.54131 62W	FIPS State... 42 FIPS County... 027 Tract... 0115.02 Block... 2013 FIPS Place... 96036 FIPS Metro... 26624 MSA... 8650 CBSA... 48300 Metro... 0 Micro... 0
Standard Address	Postal	
121 CHERRY RIDGE RD STATE COLLEGE, PA 16803-3309	Carrier... C075	

Tele Atlas' Geocode.com service estimates the location of the address relative to the address range attributes encoded in its database. © 2013 TomTom North America, Inc. All rights reserved.

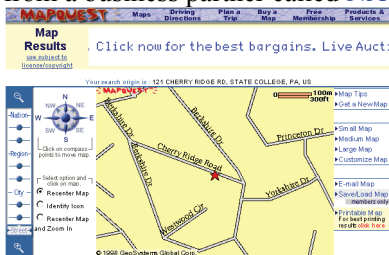
Let's compare the geocoding capabilities of MapQuest.com to locate the address on an actual map.



Address geocoded by MapQuest.com. © 2013 MapQuest.com, Inc. All rights reserved.

The MapQuest.com map from 2013 estimates the address is close to its actual location. Below is a similar MapQuest product created back in 1998, when this course was first being developed. On the older map the same address is plotted on the opposite side of the street. What do you suppose is wrong with the address range attribute in that case?

On the map from 1998, also note the shapes of the streets. The street shapes in the 2011 map have been improved. The 1998 product seems to have been generated from the 1990 version of the TIGER/Line files, which may have been all that was available for this relatively remote part of the country. Now MapQuest licenses street data from a business partner called [NAVTEQ](#).



Same address geocoded by MapQuest.com in 1998. © 1998 MapQuest.com, Inc. (formerly GeoSystems Global Corp.) All rights reserved.

The point of this section is to show that geocoding with address ranges involves a process of estimation. The Census Bureau's TIGER/Line Shapefiles, like the commercial street databases produced by Tele Atlas, Navigation Technologies, and other private firms, represent streets as vector line segments. The vector segments are associated with address range attributes, one for the left side of the street, one for the right side. The geocoding process takes a street address as input, finds the line segment that represents the specified street, checks the address ranges to determine the correct side of the street, then estimates a location at the appropriate point between the minimum and maximum address for that segment and assigns an estimated latitude/longitude coordinate to that location. For example, if the minimum address is 401, and the maximum is 421, a geocoding algorithm would locate address 411 at the midpoint of the street segment.

TRY THIS!

Try one of these geocoding services for your address. Then compare the experience, and the result, with [Google Maps](#), launched in 2005. Apply what we've discussed in this chapter to try to explain inaccuracies in your results, if any. Registered students can log in and post comments directly to this page.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 4 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Geocoding.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

4.10. Applications beyond the Census Bureau

Two characteristics of MAF/TIGER data, address range attributes and explicit topology, make them, and derivative products, valuable in many contexts. Consequently, firms like [NAVTEQ](#) and [Tele Atlas](#) (now owned by TomTom) have emerged to provide data with similar characteristics as MAF/TIGER, but which are more up-to-date, more detailed and include additional feature classes. The purpose of the next section is to sketch some of the applications of data similar to MAF/TIGER data beyond the Census Bureau.

TRY THIS!

A [February 2006 article by Peter Valdes-Dapena in CNNMoney.com](#) describes the work of two NAVTEQ employees. See the link above or search on “where those driving directions really come from”

4.11. Geocoding Your Customers

Geocoded addresses allow governments and businesses to map where their constituents and customers live and work. Federal, state, and local government agencies know where their constituents live by virtue of censuses, as well as applications for licenses and registrations. Banks, credit card companies, and telecommunications firms are also rich in address-referenced customer data, including purchasing behaviors. Private businesses and services must be more resourceful.

Some retail operations, for example, request addresses or ZIP Codes from customers, or capture address data from checks. Discount and purchasing club cards allow retailers to directly match purchasing behaviors with addresses. Customer addresses can also be harvested from automobile license plates. Business owners pay to record license plate numbers of cars parked in their parking lots or in their competitors. Addresses of registered owners can be purchased from organizations that acquire motor vehicle records from state departments of transportation.

Businesses with access to address-referenced customer data, vector street data attributed with address ranges, and GIS software and expertise, can define and analyze the **trade areas** within which most of their customers live and work. Companies can also focus direct mail advertising campaigns on their own trade areas, or their competitors'. Furthermore, GIS can be used to analyze the socio-economic characteristics of the population within trade areas, enabling businesses to make sure that the products and services they offer meet the needs and preferences of target populations.

Politicians use the same tools to target appearances and campaign promotions.

TRY THIS!

Check out the [geocoding system maintained by the Federal Financial Institution's Examination Council](#). The FFIEC Geocoding system lets users enter a street address and get a census demographic report or a street map (Using Tele Atlas data). The system is intended for use by financial institutions that are covered by the Home Mortgage Disclosure Act (HMDA) and Community Reinvestment Act (CRA) to meet their reporting obligation.

4.12. Delivering Products and Services

Operations such as mail and package delivery, food and beverage distribution, and emergency medical services need to know not only where their customers are located, but how to deliver products and services to those locations as efficiently as possible. Geographic data products like TIGER/Line Shapefiles are valuable to analysts responsible for prescribing the most efficient delivery routes. The larger and more complex the service areas of such organizations, the more incentive they have to automate their routing procedures.

In its simplest form, **routing** involves finding the shortest path through a network from an origin to a destination. Although shortest path algorithms were originally implemented in raster frameworks, transportation networks are now typically represented with vector feature data, like TIGER/Line Shapefiles. Street segments are represented as digital line segments each formed by two points, a “start” node and an “end” node. If the nodes are specified within geographic or plane coordinate systems, the distance between them can be calculated readily. Routing procedures sum the lengths of every plausible sequence of line segments that begins and ends at the specified locations. The sequence of segments associated with the smallest sum represents the shortest route.

To compare various possible sequences of segments, the data must indicate which line segment follows immediately after another line segment. In other words, the procedure needs to know about the connectivity of features. As discussed earlier, connectivity is an example of a topological relationship. If topology is not encoded in the data product, it can be calculated by the GIS software in which the procedure is coded.

MAPQUEST Maps Driving Directions Travel Guide Buy a Map Free Membership Products & Services

Starting Address
Please choose from the list of Airports:
San Francisco Intl Airport, Terminal Acc, San Francisco, CA

Enter Destination Address
Select:
City Halls
Address/Intersection:
400 Van Ness Ave
City:
San Francisco
State:
CA

Route Type
Door-to-Door (Optimal)
City-to-City (Faster, avoids road for long trips)

Display Type
Overview Map with Text
Text Only
Turn-by-Turn Maps with Text (for Door-to-Door route type Only)

Input form for an early version of the MapQuest routing utility. © 1998 MapQuest.com, Inc. All rights reserved.

Several online travel planning services, including MapQuest.com and Google Maps, provide routing capabilities. Both take origin and destination addresses as input, and produce optimal routes as output. These services are based on vector feature databases in which street segments are attributed with address ranges, as well as with other data that describe the type and conditions of the roads they represent.

MAPQUEST Maps Driving Directions Travel Guide Buy a Map Free Membership Products & Services

Route and Avoid Preference Options
MapQuest allows you to control some of the options used to compute your route. These options apply to door-to-door directions only.

Route Quality:
Shortest (Selected) Fastest

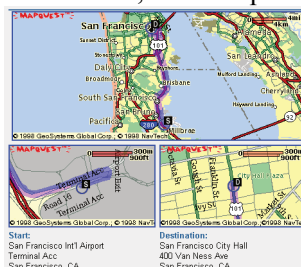
Avoid:
Limited Access Highways, Toll Roads, if possible, Ferry Lanes, if possible (if possible)

Back to Routing Form

For more information, view the [route and avoid preference definitions](#)

An early interface to MapQuest’s routing options. Different algorithms are required to calculate shortest and fastest routes. Specific attributes must be encoded in the database to provide the options to avoid limited access highways, toll roads, and ferry lanes. © 1998 MapQuest.com, Inc. All rights reserved.

The shortest route is not always the best. In the context of emergency medical services, for example, the fastest route is preferred, even if it entails longer distances than others. To determine fastest routes, additional attribute data must be encoded, such as speed limits, traffic volumes, one way streets, and other characteristics.



MapQuest routing solution. © 1998 MapQuest.com, Inc. All rights reserved.

Then there are routing problems that involve multiple destinations—a complex special case of routing called the **traveling salesman problem**. School bus dispatchers, mail and package delivery service managers, and food and beverage distributors all seek to minimize the transportation costs involved in servicing multiple, dispersed destinations. As the number of destinations and the costs of travel increase, the high cost of purchasing up-to-date, properly attributed network data becomes easier to justify.

TRY THIS

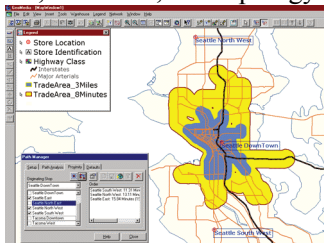
The Georgia Institute of Technology publishes an [extensive collection of resources about the Traveling Salesman Problem](#).

4.13. Delineating Service Areas

The need to redraw voting district boundaries every ten years was one of the motivations that led the Census Bureau to create its MAF/TIGER database. Like voting districts, many other kinds of service area boundaries need to be revised periodically. School districts are a good example. The state of Massachusetts, for instance, has adopted school districting laws that are similar in effect to the constitutional criteria used to guide congressional redistricting. The Framingham (Massachusetts) School District's Racial Balance Policy once stated that "each elementary and middle school shall enroll a student body that is racially balanced. ... each student body shall include a percentage of minority student, which reflects the system-wide percentage of minority students, plus or minus ten percent. ... The racial balance required by this policy shall be established by redrawing school enrollment areas" (Framingham Public Schools 1998). And bus routes must be redrawn as enrollment area boundaries change.

The [Charlotte-Mecklenberg \(North Carolina\) public school district](#) also used racial balance as a districting criterion (although its policy was subsequently challenged in court). Charlotte-Mecklenberg consists of 133 schools, attended by over 100,000 students, about one third of whom ride a bus to school every day. District managers are responsible for routing 3,600 bus routes, traveling a total of 82,000 daily miles. A staff of eight routinely uses GIS to manage these tasks. GIS could not be used unless up-to-date, appropriately attributed, and topologically encoded data were available.

Another example of service area analysis is provided by the City of Beaverton, Oregon. In 1997, Beaverton officials realized that 25 percent of the volume of solid waste that was hauled away to land fills consisted of yard waste, such as grass clippings and leaves. Beaverton decided to establish a yard waste recycling program, but it knew that the program would not be successful if residents found it inconvenient to participate. A GIS procedure called **allocation** was used to partition Beaverton's street network into service areas that minimized the drive time from residents' homes to recycling facilities. Allocation procedures require vector-format data that includes the features, attributes, and topology necessary to calculate travel times from all residences to the nearest facility.



Trade areas defined by 3 miles travel distance (blue) and 8 minutes travel time (yellow). (Francica n.d.). Used by permission.

Naturally, private businesses concerned with delivering products and services are keenly interested in service area delineation. The screen capture above shows two **trade areas** surrounding a retail store location ("Seattle Downtown") in a network database.

Former student Saskia Cohick (Winter 2006), who was then GIS Director for Tioga County, Pennsylvania, contributed another service area problem: “This is a topic that local governments are starting to deal with ... To become Phase 2 wireless capable (that is, capable of finding a cell phone location from a 911 call center within 200 feet of the actual location), county call centers must have a layer called ESZs (Emergency Service Zones). This layer will tell the dispatcher who to send to the emergency (police, fire, medical, etc). The larger problem is to reach agreement between four fire companies (for example) as to where they do or do not respond.”

4.14. Summary

To fulfill its mission of being the preeminent producer of attribute data about the population and economy of the United States, the U.S. Census Bureau also became an innovative producer of digital geographic data. The Bureau designed its MAF/TIGER database to support automatic geocoding of address-referenced census data, as well as automatic data quality control procedures. The key characteristics of TIGER/Line Shapefiles, including use of vector features to represent geographic entities, and address range attributes to enable address geocoding, are now common features of proprietary geographic databases used for trade area analysis, districting, routing, and allocation.

QUIZ

Registered Penn State students should return now to the Chapter 4 folder in ANGEL (via the Resources menu to the left) to access the graded quiz for this chapter. This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter’s learning objectives. You are free to review the chapter during the quiz. Once you’ve submitted the quiz you will have completed Chapter 4.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

4.15. Bibliography

Charlotte-Mecklenberg Public Schools (n. d.). Retrieved July 19, 1999 from <http://www.cms.k12.nc.us>

Cooke, D. F. (1997). Topology and TIGER: The Census Bureau’s Contribution. In T. W. Foresman (Ed.), *The history of geographic information systems: Perspectives from the pioneers*. (pp. 47 – 57). Upper Saddle River, NJ: Prentice Hall.

Dangermond, J. (1982). A Classification of Software Components Commonly Used in Geographic Information Systems. In *Proceedings of the U.S.—Australia Workshop on the Design and Implementation of Computer-Based Geographic Information Systems*, Honolulu, HI, pp. 0-91. In Demers, M.N. (1997) *Fundamentals of Geographic Information Systems*. John Wiley & Sons, Inc.

- Discreet Research (n.d.). Retrieved July 19, 1999 from <http://www.dresearch.com>
- ESRI (1998) Shapefile Technical Description, An ESRI White paper. Environmental Systems Research Institute, Inc. Retrieved October 4, 2010, from <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
- Federal Geographic Data Committee (April 2006). Retrieved July 19, 1999 from <http://www.fgdc.gov>
- Framingham Public Schools (1998). *Racial balance policy: Assignment of students to schools*. Retrieved July 19, 1999 from www.framingham.k12.ma.us/update/0198rbp.html (since retired).
- Francica, J. (n.d.). *Geodezix Consulting*. Retrieved July 19, 1999 from www.geodezix.com (since retired).
- Galdi, D. (2005). Spatial Data Storage and Topology in the Redesign MAF/TIGER System. Retrieved 19 October 2010 from http://www.census.gov/geo/mtep_obj2/topo_and_data_stor.html (since retired).
- MapQuest (n.d. a). Retrieved July 19, 1998 from <http://www.mapquest.com>
- MapQuest (n.d. b). Retrieved January 15, 2013 from <http://www.mapquest.com>
- Marx, R. M. (Ed.). (1990). The Census Bureau's TIGER system. [Special issue]. *Cartography and Geographic Information Systems* 17:1.
- Navigation Technologies Inc. (2006). *Welcome to NavTech*. Retrieved July 19, 1999 from <http://www.navtech.com>
- Rammage, S. and P. Woodsford (2002). The Benefits of Topology in the Database. Retrieved October 6, 2010 from <http://spatialnews.geocomm.com/features/laserscan2/>
- TeleAtlas (2006). *Welcome to TeleAtlas*. Retrieved May 3, 2006 from <http://www.teleatlas.com/Pub/Home> (since retired).
- Theobald, D. M. (2001). Understanding Topology and Shapefiles. *ArcUser* April-June 2001. Retrieved October 5, 2010 from <http://www.esri.com/news/arcuser/0401/topo.html>
- U.S. Census Bureau (1997). *TIGER/Line Files (1997 Technical Documentation)*. Retrieved January 2, 1999 from <http://www.census.gov/geo/tiger/TIGER97C.pdf> (since retired).
- U.S. Census Bureau (2003). TIGER/Line Files, 2003 (metadata). Retrieved February 3, 2008 from <http://www.census.gov/geo/www/tlmetadata/tl2003meta.txt>
- U.S. Census Bureau (n. d.). 21st Century MAF/TIGER Enhancements. Retrieved February 3, 2008 from <http://www.census.gov/geo/mod/overview.pdf> (since retired).
- U.S. Census Bureau (2004). MAF/TIGER Redesign Project Overview. Retrieved October 19, 2010 from http://www.census.gov/geo/mtep_obj2/obj2_issuepaper12_2004.pdf (since retired).
- U.S. Census Bureau (2005). *Geography division map gallery*. Retrieved July 19, 1999 from <http://www.census.gov/geo/www/mapGallery/>
- U.S. Census Bureau (2012). TIGER/Line Shapefiles Technical Documentation. Retrieved June, 2013 from http://www.census.gov/geo/maps-data/data/pdfs/tiger/tgrshp2012/TGRSHP2012_TechDoc.pdf

Chapter 5

Land Surveying and GPS

David DiBiase

5.1. Overview

As you recall from Chapter 1, geographic data represent spatial locations and non-spatial attributes measured at certain times. We defined “feature” as a set of positions that specifies the location and extent of an entity. Positions, then, are a fundamental element of geographic data. Like the letters that make up these words, positions are the building blocks from which features are constructed. A property boundary, for example, is made up of a set of positions connected by line segments.

In theory, a single position is a “0-dimensional” feature: an infinitesimally small point from which 1-dimensional, 2-dimensional, and 3-dimensional features (lines, areas, and volumes) are formed. In practice, positions occupy 2- or 3-dimensional areas as a result of the limited resolution of measurement technologies and the limited precision of location coordinates. Resolution and precision are two aspects of data quality. This chapter explores the technologies and procedures used to produce positional data, and the factors that determine its quality.

Objectives

Students who successfully complete Chapter 5 should be able to:

1. Identify and define the key aspects of data quality, including resolution, precision, and accuracy;
2. List and explain the procedures land surveyors use to produce positional data, including traversing, triangulation, and trilateration;
3. Calculate plane coordinates by open traverse;
4. Calculate elevations by leveling;
5. Explain how radio signals broadcast by Global Positioning System satellites are used to calculate positions on the surface of the Earth;
6. State the kinds and magnitude of error associated with uncorrected GPS positioning; and
7. Identify and explain methods used to improve the accuracy of GPS positioning.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save”

button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

5.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 5 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 5	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	<p>Submit five practice quizzes including:</p> <ul style="list-style-type: none"> • Horizontal Positions • Vertical Positions • GPS Components • GPS Error Sources • GPS Error Correction <p>Practice quizzes are not graded and may be submitted more than once.</p>	Go to ANGEL > [your course section] > Lessons tab > Chapter 5 folder > [quiz]
3	<p>Perform “Try this” activities including:</p> <ul style="list-style-type: none"> • Use trilateration to determine the position of a point in a control network • Investigate the status of the GPS satellite constellation • Visualize the positions and orbits of GPS satellites • Take the Trimble GPS Tutorial • Download and explore the Trimble GPS Planning software • Perform differential correction of GPS coordinates <p>“Try this” activities are not graded.</p>	Instructions are provided for each activity.
4	Submit the Chapter 5 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 5 folder > Chapter 5 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

5.3. Geospatial Data Quality

Quality is a characteristic of comparable things that allows us to decide that one thing is better than another. In

the context of geographic data, the ultimate standard of quality is the degree to which a data set is fit for use in a particular application. That standard is called **validity**. The standard varies from one application to another. In general, however, the key criteria are how much error is present in a data set, and how much error is acceptable.

Some degree of error is always present in all three components of geographic data: features, attributes, and time. Perfect data would fully describe the location, extent, and characteristics of phenomena exactly as they occur at every moment. Like the proverbial 1:1 scale map, however, perfect data would be too large, and too detailed to be of any practical use. Not to mention impossibly expensive to create in the first place!

5.4. Error and Uncertainty

Positions are the products of measurements. All measurements contain some degree of error. Errors are introduced in the original act of measuring locations on the Earth surface. Errors are also introduced when second- and third-generation data is produced, say, by scanning or digitizing a paper map.

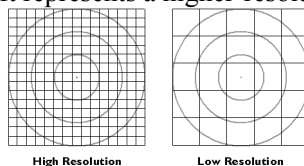
In general, there are three sources of error in measurement: human beings, the environment in which they work, and the measurement instruments they use.

Human errors include mistakes, such as reading an instrument incorrectly, and judgments. Judgment becomes a factor when the phenomenon that is being measured is not directly observable (like an aquifer), or has ambiguous boundaries (like a soil unit).

Environmental characteristics, such as variations in temperature, gravity, and magnetic declination, also result in measurement errors.

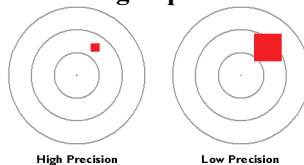
Instrument errors follow from the fact that space is continuous. There is no limit to how precisely a position can be specified. Measurements, however, can be only so precise. No matter what instrument, there is always a limit to how small a difference is detectable. That limit is called **resolution**.

The diagram below shows the same position (the point in the center of the bullseye) measured by two instruments. The two grid patterns represent the smallest objects that can be detected by the instruments. The pattern at left represents a higher-resolution instrument.



Resolution.

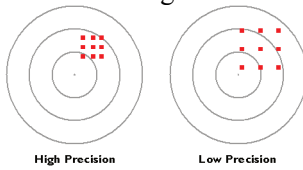
The resolution of an instrument affects the **precision** of measurements taken with it. In the illustration below, the measurement at left, which was taken with the higher-resolution instrument, is more precise than the measurement at right. In digital form, the more precise measurement would be represented with additional decimal places. For example, a position specified with the UTM coordinates 500,000. meters East and 5,000,000. meters North is actually an area 1 meter square. A more precise specification would be 500,000.001 meters East and 5,000,000.001 meters North, which locates the position within an area 1 millimeter square. You can think of the area as a zone of **uncertainty** within which, somewhere, the theoretically infinitesimal point location exists. **Uncertainty is inherent in geospatial data.**



The precision of a single measurement.

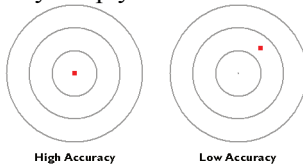
Precision takes on a slightly different meaning when it is used to refer to a number of repeated measurements.

In the illustration below, there is less variance among the nine measurements at left than there is among the nine measurements at right. The set of measurements at left is said to be more precise.



The precision of multiple measurements.

Hopefully you have noticed that resolution and precision are independent from **accuracy**. As shown below, accuracy simply means how closely a measurement corresponds to an actual value.



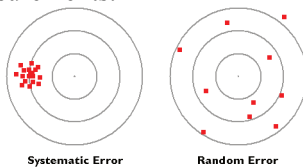
Accuracy.

I mentioned the U.S. Geological Survey's **National Map Accuracy Standard** in Chapter 2. In regard to topographic maps, the Standard warrants that 90 percent of well-defined points tested will be within a certain tolerance of their actual positions. Another way to specify the accuracy of an entire spatial database is to calculate the average difference between many measured positions and actual positions. The statistic is called the **root mean square error (RMSE)** of a data set.

5.5. Systematic vs. Random Errors

The diagram below illustrates the distinction between **systematic** and **random** errors. Systematic errors tend to be consistent in magnitude and/or direction. If the magnitude and direction of the error is known, accuracy can be improved by additive or proportional corrections. **Additive correction** involves adding or subtracting a constant adjustment factor to each measurement; **proportional correction** involves multiplying the measurement(s) by a constant.

Unlike systematic errors, random errors vary in magnitude and direction. It is possible to calculate the average of a set of measured positions, however, and that average is likely to be more accurate than most of the measurements.



Systematic and random errors.

In the sections that follow we compare the accuracy and sources of error of two important positioning technologies: land surveying and the Global Positioning System.

5.6. Survey Control

Geographic positions are specified relative to a fixed reference. Positions on the globe, for instance, may be specified in terms of angles relative to the center of the Earth, the equator, and the prime meridian. Positions in plane coordinate grids are specified as distances from the origin of the coordinate system. Elevations are expressed

as distances above or below a vertical datum such as mean sea level, or an ellipsoid such as GRS 80 or WGS 84, or a geoid.

Land surveyors measure horizontal positions in geographic or plane coordinate systems relative to previously surveyed positions called **control points**. In the U.S., the National Geodetic Survey (NGS) maintains a **National Spatial Reference System (NSRS)** that consists of approximately 300,000 horizontal and 600,000 vertical control stations (Doyle, 1994). Coordinates associated with horizontal control points are referenced to NAD 83; elevations are relative to NAVD 88. In a Chapter 2 activity you may have retrieved one of the datasheets that NGS maintains for every NSRS control point, along with more than a million other points submitted by professional surveyors.



Benchmark used to mark a vertical control point. (Thompson, 1988).

In 1988 NGS established **four orders of control point accuracy**, which are outlined in the table below. The minimum accuracy for each order is expressed in relation to the horizontal distance separating two control points of the same order. For example, if you start at a control point of order AA and measure a 500 km distance, the length of the line should be accurate to within 3 mm base error, plus or minus 5 mm line length error (500,000,000 mm × 0.01 parts per million).

Four orders of control point accuracy

Order	Survey activities	Maximum base error(95% confidence limit)	Maximum Line-length dependent error(95% confidence limit)
AA	Global-regional dynamics; deformation measurements	3 mm	1:100,000,000 (0.01 ppm)
A	NSRS primary networks	5 mm	1:10,000,000 (0.1 ppm)
B	NSRS secondary networks; high-precision engineering surveys	8 mm	1:1,000,000 (1 ppm)
C	NSRS terrestrial; dependent control surveys for mapping, land information, property, and engineering requirements	1st: 1.0 cm 2nd-I: 2.0 cm 2nd-II: 3.0 cm 3rd: 5.0 cm	1st: 1:100,000 2nd-I: 1:50,000 2nd-II: 1:20,000 3rd: 1:10,000

Control network accuracy standards used for U.S. National Spatial Reference System (Federal Geodetic Control Committee, 1988).

Doyle (1994) points out that horizontal and vertical reference systems coincide by less than ten percent. This is because

“

...horizontal stations were often located on high mountains or hilltops to decrease the need to construct observation towers usually required to provide line-of-sight for triangulation, traverse and trilateration measurements. Vertical control points however, were established by the technique of spirit leveling which is more suited to being

“

conducted along gradual slopes such as roads and railways that seldom scale mountain tops. (Doyle, 2002, p. 1)

You might wonder how a control network gets started. If positions are measured relative to other positions, what is the first position measured relative to? The answer is: the stars. Before reliable timepieces were available, astronomers were able to determine longitude only by careful observation of recurring celestial events, such as eclipses of the moons of Jupiter. Nowadays geodesists produce extremely precise positional data by analyzing radio waves emitted by distant stars. Once a control network is established, however, **surveyors produce positions using instruments that measure angles and distances between locations on the Earth’s surface.**

6.7. Measuring Angles

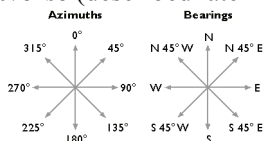
Angles can be measured with a magnetic compass, of course. Unfortunately, the Earth’s magnetic field does not yield the most reliable measurements. The magnetic poles are not aligned with the planet’s axis of rotation (an effect called **magnetic declination**), and they tend to change location over time. Local magnetic anomalies caused by magnetized rocks in the Earth’s crust and other geomagnetic fields make matters worse.

For these reasons land surveyors rely on **transits** (or their more modern equivalents, called **theodolites**) to measure angles. A transit consists of a telescope for sighting distant target objects, two measurement wheels that work like protractors for reading horizontal and vertical angles, and bubble levels to ensure that the angles are true. A theodolite is essentially the same instrument, except that some mechanical parts are replaced with electronics.



Transit. (Raisz, 1948). Used by permission.

Surveyors express angles in several ways. When specifying directions, as is done in the preparation of a property survey, angles may be specified as bearings or azimuths. A **bearing** is an angle less than 90° within a quadrant defined by the cardinal directions. An **azimuth** is an angle between 0° and 360° measured clockwise from North. “South 45° East” and “ 135° ” are the same direction expressed as a bearing and as an azimuth. An **interior angle**, by contrast, is an angle measured between two lines of sight, or between two legs of a **traverse** (described later in this chapter).



Azimuths and bearings.

In the U.S., professional organizations like the American Congress on Surveying and Mapping, the American Land Title Association, the National Society of Professional Surveyors, and others, recommend minimum accuracy standards for angle and distance measurements. For example, as Steve Henderson (personal communication, Fall 2000, updated July 2010) points out, the [Alabama Society of Professional Land Surveyors](#) recommends that errors in angle measurements in “commercial/high risk” surveys be no greater than 15 seconds times the square root of the number of angles measured.

To achieve this level of accuracy, surveyors must overcome errors caused by faulty instrument calibration; wind, temperature, and soft ground; and human errors, including misplacing the instrument and misreading the measurement wheels. In practice, surveyors produce accurate data by taking repeated measurements and averaging the results.

8. Measuring Distances

To measure distances, land surveyors once used 100-foot long **metal tapes** that are graduated in hundredths of a foot. (This is the technique I learned as a student in a surveying class at the University of Wisconsin in the early 1980s. The picture shown below is slightly earlier.) Distances along slopes are measured in short horizontal segments. Skilled surveyors can achieve accuracies of up to one part in 10,000 (1 centimeter error for every 100 meters distance). Sources of error include flaws in the tape itself, such as kinks; variations in tape length due to extremes in temperature; and human errors such as inconsistent pull, allowing the tape to stray from the horizontal plane, and incorrect readings.



Surveying team measuring a baseline distance with a metal (Invar) tape. (Hodgson, 1916).

Since the 1980s, **electronic distance measurement** (EDM) devices have allowed surveyors to measure distances more accurately and more efficiently than they can with tapes. To measure the horizontal distance between two points, one surveyor uses an EDM instrument to shoot an energy wave toward a reflector held by the second surveyor. The EDM records the elapsed time between the wave's emission and its return from the reflector. It then calculates distance as a function of the elapsed time. Typical short-range EDMs can be used to measure distances as great as 5 kilometers at accuracies up to one part in 20,000, twice as accurate as taping.



Total station.

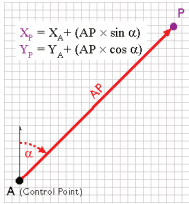
Instruments called **total stations** combine electronic distance measurement and the angle measuring capabilities of theodolites in one unit. Next we consider how these instruments are used to measure horizontal positions in relation to established control networks.

5.9. Horizontal Positions

Surveyors have developed distinct methods, based on separate control networks, for measuring horizontal and vertical positions. In this context, a horizontal position is the location of a point relative to two axes: the equator and the prime meridian on the globe, or x and y axes in a plane coordinate system. Control points tie coordinate systems to actual locations on the ground; they are the physical manifestations of **horizontal datums**. In the following pages we review two techniques that surveyors use to create and extend **control networks** (triangulation and trilateration) and two other techniques used to measure positions relative to control points (open and closed traverses).

5.10. Traverse

Surveyors typically measure positions in series. Starting at control points, they measure angles and distances to new locations, and use trigonometry to calculate positions in a plane coordinate system. Measuring a series of positions in this way is known as “running a traverse.” A traverse that begins and ends at different locations is called an **open traverse**.



An open traverse. (Adapted from Robinson, et al., 1995)

For example, say the UTM coordinates of point A in the illustration above are 500,000.00 E and 5,000,000.00 N. The distance between points A and P, measured with a steel tape or an EDM, is 2,828.40 meters. The azimuth of the line AP, measured with a transit or theodolite, is 45° . Using these two measurements, the UTM coordinates of point P can be calculated as follows:

$$X_P = 500,000.00 + (2,828.40 \times \sin 45^\circ) = 501,999.98$$

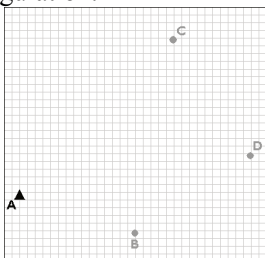
$$Y_P = 5,000,000.00 + (2,828.40 \times \cos 45^\circ) = 5,001,999.98$$

A traverse that begins and ends at the same point, or at two different but known points, is called a **closed traverse**. Measurement errors in a closed traverse can be quantified by summing the interior angles of the polygon formed by the traverse. The accuracy of a single angle measurement cannot be known, but since the sum of the interior angles of a polygon is always $(n-2) \times 180$, it's possible to evaluate the traverse as a whole, and to distribute the accumulated errors among all the interior angles.

Errors produced in an **open traverse**, one that does not end where it started, cannot be assessed or corrected. The only way to assess the accuracy of an open traverse is to measure distances and angles repeatedly, forward and backward, and to average the results of calculations. Because repeated measurements are costly, other surveying techniques that enable surveyors to calculate and account for measurement error are preferred over open traverses for most applications.

5.11. Triangulation

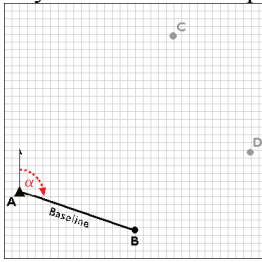
Closed traverses yield adequate accuracy for property boundary surveys, provided that an established control point is nearby. Surveyors conduct **control surveys** to extend and densify horizontal control networks. Before survey-grade satellite positioning was available, the most common technique for conducting control surveys was triangulation.



The purpose of a control survey is to establish new horizontal control points (B, C, and D) based upon an existing control point (A).

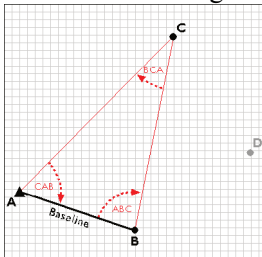
Using a total station equipped with an electronic distance measurement device, the control survey team commences by measuring the azimuth α , and the baseline distance AB. These two measurements enable the

survey team to calculate position B as in an open traverse. Before geodetic-grade GPS became available, the accuracy of the calculated position B may have been evaluated by astronomical observation.



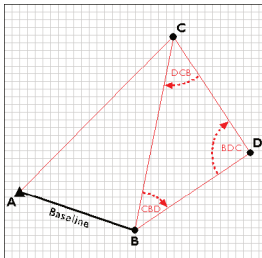
Establishing a second control point (B) in a triangulation network.

The surveyors next measure the interior angles CAB, ABC, and BCA at point A, B, and C. Knowing the interior angles and the baseline length, the trigonometric “law of sines” can then be used to calculate the lengths of any other side. Knowing these dimensions, surveyors can fix the position of point C.



Establishing the position of point C by triangulation.

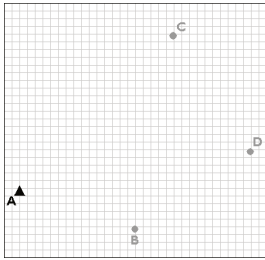
Having measured three interior angles and the length of one side of triangle ABC, the control survey team can calculate the length of side BC. This calculated length then serves as a baseline for triangle BDC. Triangulation is thus used to extend control networks, point by point and triangle by triangle.



Extending the triangulation network.

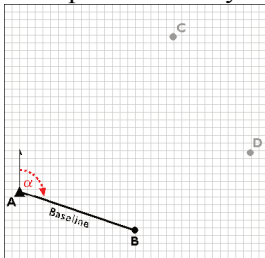
5.12. Trilateration

Trilateration is an alternative to triangulation that relies upon distance measurements only. Electronic distance measurement technologies make trilateration a cost-effective positioning technique for control surveys. Not only is it used by land surveyors, **trilateration is also used to determine location coordinates with Global Positioning System satellites and receivers.**



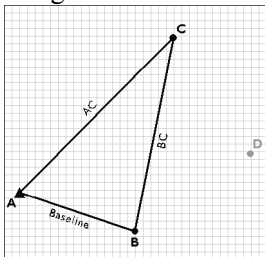
Trilateration is used to extend control networks by establishing new control points (B, C, and D) based upon existing control points (A).

Trilateration networks commence the same way as triangulation nets. If only one existing control point is available, a second point (B) is established by open traverse. Using a total station equipped with an electronic distance measurement device, the survey team measures the azimuth α and baseline distance AB. The total station operator may set up her instrument over point A, while her assistant holds a reflector mounted on a shoulder-high pole as steadily as he can over point B. Depending on the requirements of the control survey, the accuracy of the calculated position B may be confirmed by astronomical observation.



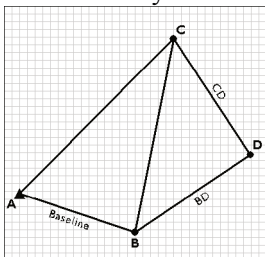
Establishing a second control point (B) in a trilateration network.

Next, the survey team uses the electronic distance measurement feature of the total station to measure the distances AC and BC. Both forward and backward measurements are taken. After the measurements are reduced from slope distances to horizontal distances, the law of cosines can be employed to calculate interior angles, and the coordinates of position C can be fixed. The accuracy of the fix is then checked by plotting triangle ABC and evaluating the error of closure.



Measuring the distances AC and BC.

Next, the trilateration network is extended by measuring the distances CD and BD, and fixing point D in a plane coordinate system.



Fixing control point D by trilateration.

TRY THIS!

USE TRILATERATION TO DETERMINE A CONTROL POINT LOCATION

Trilateration is a technique land surveyors use to calculate an undetermined position in a plane coordinate system by measuring distances from two known positions. As you will see later in this chapter, trilateration is also the technique that GPS receivers use to calculate their positions on the Earth's surface, relative to the positions of three or more satellite transmitters. The purpose of this exercise is to make sure you understand how trilateration works. (*Estimated time to complete: 5 minutes.*)

Note: You will need to have the Adobe Flash player installed in order to complete this exercise. If you do not already have the Flash player, you can [download it for free from Adobe](#).

1. **Display a coordinate system grid:** In this exercise, you will interact with a coordinate system grid. First, [display the coordinate system grid in a separate window](#) so that you can interact with it while you read these instructions. Arrange the coordinate system grid window and this window so that you can easily view both. You may need to make this window more narrow. Two control points, A and B, are plotted in the coordinate system grid. A survey crew has measured distances from the control points to another point, point C, whose coordinates are unknown. Your job is to fix the position of point C. You will find point C at the intersection of two circles centered on control points A and B, where the radii of the two circles equals the measured distances from the control points to point C.
2. **Plot the distance from control point A to point C:** On the coordinate system grid, click on control point A to display the data entry form. (You'll need to click on the actual point, not the "A".) The form consists of a text field in which you can type in a distance, and a button that plots a circle centered on point A. The radius of the circle will be the distance you specify. According to the surveyors' measurements, the distance between control point A and point C is 9400 feet. Enter that distance now and click Plot to plot the circle. [[View result of Step 2](#)]
3. **Plot the distance from control point B to point C:** The measured distance from point B to point C is 7000 feet. Click on point B (on the actual point, not the "B"), enter that distance and plot a circle. [[View result of Step 3](#)]
4. **Plot point C:** Now click within the coordinate grid to reveal the position of point C. You may have to hunt for it, but you should know where to look based on the intersection of the circles. [[View result of Step 4](#)]
5. **Extend the control network further:** Now continue extending the control network by plotting a fourth point, point D, in the coordinate system grid. First plot new circles at points A and C. The measured distance from point A to point D is 9600 feet. The measured distance from point C to point D is 8000 feet. (You may wish to set the radius of the circle centered upon point B to 0.) Finally, click in the coordinate system grid to plot point D. [[View result of Step 5](#)]

Once you have finished viewing the grid, close the popup window.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to

the left) to take a self-assessment quiz about Horizontal Positions. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

5.13. Vertical Positions

A **vertical position** is the height of a point relative to some reference surface, such as mean sea level, a geoid, or an ellipsoid. The roughly 600,000 vertical control points in the U.S. National Spatial Reference System (NSRS) are referenced to the North American Vertical Datum of 1988 (NAVD 88). Surveyors created the National Geodetic Vertical Datum of 1929 (NGVD 29, the predecessor to NAVD 88), by calculating the average height of the sea at all stages of the tide at 26 tidal stations over 19 years. Then they extended the control network inland using a surveying technique called **leveling**. Leveling is still a cost-effective way to produce elevation data with sub-meter accuracy.



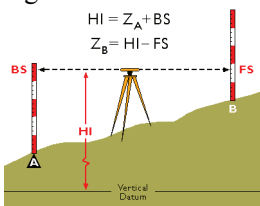
A leveling crew at work in 1916. (Hodgson, 1916).

The illustration above shows a leveling crew at work. The fellow under the umbrella is peering through the telescope of a leveling instrument. Before taking any measurements, the surveyor made sure that the telescope was positioned midway between a known elevation point and the target point. Once the instrument was properly leveled, he focused the telescope crosshairs on a height marking on the rod held by the fellow on the right side of the picture. The chap down on one knee is noting in a field book the height measurement called out by the telescope operator.



A level used for determining elevations.

Leveling is still a cost-effective way to produce elevation data with sub-meter accuracy. A modern leveling instrument is shown in the photograph above. The diagram below illustrates the technique called differential leveling.



Differential leveling. (Adapted from Wolf & Brinker, 1994)

The diagram above illustrates **differential leveling**. A leveling instrument is positioned midway between a point at which the ground elevation is known (point A) and a point whose elevation is to be measured (B). The height of the instrument above the datum elevation is HI. The surveyor first reads a backsight measurement (BS) off of a leveling rod held by his trusty assistant over the benchmark at A. The height of the instrument can be calculated as the sum of the known elevation at the benchmark (Z_A) and the backsight height (BS). The assistant then moves the rod to point B. The surveyor rotates the telescope 180° , then reads a foresight (FS) off the rod at B. The elevation at B (Z_B) can then be calculated as the difference between the height of the instrument (HI) and the foresight height (FS).

Former student Henry Whitbeck (personal communication, Fall 2000) points out that surveyors also use total stations to measure vertical angles and distances between fixed points (prisms mounted upon tripods at fixed heights), then calculate elevations by trigonometric leveling.

HEIGHTS

Surveyors use the term **height** as a synonym for elevation. There are several different ways to measure heights. A properly-oriented level defines a line parallel to the geoid surface at that point (Van Sickle, 2001). **An elevation above the geoid is called an orthometric height.** However, GPS receivers cannot produce orthometric heights directly. Instead, GPS produces heights relative to the WGS 84 ellipsoid. **Elevations produced with GPS are therefore called ellipsoidal (or geodetic) heights.**

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Vertical Positions. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

5.14. Global Positioning System



Positioning signals broadcast from three Global Positioning System satellites are received at a location on Earth. (U.S. Federal Aviation Administration, 2007b)

The Global Positioning System (GPS) employs trilateration to calculate the coordinates of positions at or near the Earth's surface. Trilateration refers to the trigonometric law by which the interior angles of a triangle can be determined if the lengths of all three triangle sides are known. GPS extends this principle to three dimensions.

A GPS receiver can fix its latitude and longitude by calculating its distance from three or more Earth-orbiting satellites, whose positions in space and time are known. If four or more satellites are within the receiver's "horizon," the receiver can also calculate its elevation, and even its velocity. The U.S. Department of Defense created the Global Positioning System as an aid to navigation. Since it was declared fully operational in 1994, GPS positioning has been used for everything from tracking delivery vehicles, to tracking the minute movements of the tectonic plates that make up the Earth's crust, to tracking the movements of human beings. In addition to the so-called **user segment** made up of the GPS receivers and people who use them to measure positions, the system consists of two other components: a **space segment** and a **control segment**. It took about \$10 billion to build over 16 years.

Russia maintains a similar positioning satellite system called **GLONASS**. Member nations of the European Union are in the process of deploying a comparable system of their own, called **Galileo**. The first experimental GIOVE-A satellite began transmitting Galileo signals in January 2006. The goal of the Galileo project is a constellation of 30 navigation satellites by 2020. If the engineers and politicians succeed in making Galileo, GLONASS, and the U.S. Global Positioning System interoperable, as currently seems likely, the result will be a **Global Navigation Satellite System (GNSS)** that provides more than twice the signal-in-space resource that is available with GPS alone. The Chinese began work on their own system, called Beidou, in 2000. At the end of 2011 they had ten satellites in orbit, serving just China, with the goal being a global system of 35 satellites by 2020.

In this section you will learn to:

1. Explain how radio signals broadcast by Global Positioning System satellites are used to calculate positions on the surface of the Earth; and
2. Describe the functions of the space, control, and user segments of the Global Positioning System.

5.15. Space Segment

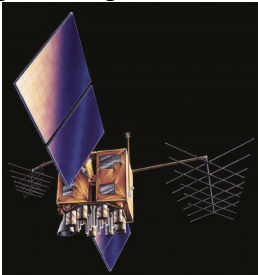
The space segment of the Global Positioning System currently consists of approximately 30 active and spare NAVSTAR satellites (new satellites are launched periodically, and old ones are decommissioned). “NAVSTAR” stands for “NAVigation System with Timing And Ranging.” Each satellite circles the Earth every 12 hours in **sidereal time** along one of six orbital “planes” at an altitude of 20,200 km (about 12,500 miles). The satellites broadcast signals used by GPS receivers on the ground to measure positions. The satellites are arrayed such that at least four are “in view” everywhere on or near the Earth’s surface at all times, with typically up to eight and potentially 12 “in view” at any given time.



The constellation of GPS satellites. Illustration © Smithsonian Institution, 1988. Used by Permission.

TRY THIS!

The U.S. Coast Guard’s Navigation Center publishes status reports on the GPS satellite constellation. Its report of August 17, 2010, for example, listed 31 satellites, five to six in each of the six orbital planes (A-F), and one scheduled outage, on August 19, 2010. [You can look up the current status of the constellation here.](#)



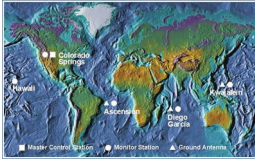
Artist’s rendition of a NAVSTAR satellite (NAVSTAR GPS Joint Program Office, n.d.).

TRY THIS!

Scientific programmers at the U.S. National Aeronautics and Space Administration (NASA) have created an interactive, three-dimensional model of the Earth and the orbits of the more than 500 man-made satellites that surround it. The model is a Java applet called **J-Track 3D Satellite Tracking**. Your browser must have Java enabled to view the applet. Instructions at the site describe how you can zoom in and out, and drag to rotate the model. To view orbits of particular satellites, choose Select from the Satellite menu. The Block IIA and R series are the most current generation of NAVSTAR satellites.

5.16. Control Segment

The control segment of the Global Positioning System is a network of ground stations that monitors the shape and velocity of the satellites' orbits. The accuracy of GPS data depends on knowing the positions of the satellites at all times. The orbits of the satellites are sometimes disturbed by the interplay of the gravitational forces of the Earth and Moon.



The control segment of the Global Positioning System (U.S. Federal Aviation Administration, 2007b).

Monitor Stations are very precise GPS receivers installed at known locations. They record discrepancies between known and calculated positions caused by slight variations in satellite orbits. Data describing the orbits are produced at the **Master Control Station** at Colorado Springs, uploaded to the satellites, and finally broadcast as part of the GPS positioning signal. GPS receivers use this satellite Navigation Message data to adjust the positions they measure.

If necessary, the Master Control Center can modify satellite orbits by commands transmitted via the control segment's ground antennas.

5.17. User Segment

The U.S. Federal Aviation Administration (FAA) estimated in 2006 that some 500,000 GPS receivers are in use for many applications, including surveying, transportation, precision farming, geophysics, and recreation, not to mention military navigation. This was before in-car GPS navigation gadgets emerged as one of the most popular consumer electronic gifts during the 2007 holiday season in North America.

Basic consumer-grade GPS receivers, like the rather old-fashioned one shown below, consist of a radio receiver and internal antenna, a digital clock, some sort of graphic and push-button user interface, a computer chip to perform calculations, memory to store waypoints, jacks to connect an external antenna or download data to a computer, and flashlight batteries for power. The radio receiver in the unit shown below includes 12 channels to receive signal from multiple satellites simultaneously.



Recreation-grade GPS receiver, circa 1998.

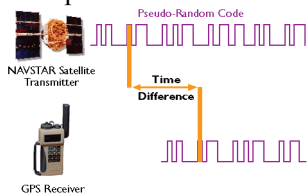
NAVSTAR Block II satellites broadcast at two frequencies, 1575.42 MHz (L1) and 1227.6 MHz (L2). (For sake of comparison, FM radio stations broadcast in the band of 88 to 108 MHz.) Only L1 was intended for civilian use. Single-frequency receivers produce horizontal coordinates at an accuracy of about three to seven meters (or about 10 to 20 feet) at a cost of about \$100. Some units allow users to improve accuracy by filtering out errors identified by nearby stationary receivers, a post-process called "differential correction." \$300-500 single-frequency units that can also receive corrected L1 signals from the U.S. Federal Aviation Administration's Wide Area Augmentation System (WAAS) network of ground stations and satellites can perform differential correction in "real-time." Differentially-corrected coordinates produced by single-frequency receivers can be as accurate as one to three meters (about 3 to 10 feet).

The signal broadcast at the L2 frequency is encrypted for military use only. Clever GPS receiver makers soon figured out, however, how to make dual-frequency models that can measure slight differences in arrival times of the two signals (these are called “carrier phase differential” receivers). Such differences can be used to exploit the L2 frequency to improve accuracy without decoding the encrypted military signal. Survey-grade carrier-phase receivers able to perform real-time kinematic (RTK) differential correction, can produce horizontal coordinates at sub-meter accuracy at a cost of \$1000 to \$2000. No wonder GPS has replaced electro-optical instruments for many land surveying tasks.

Meanwhile, a new generation of NAVSTAR satellites (the Block IIR-M series) will add a civilian signal at the L2 frequency that will enable substantially improved GPS positioning.

5.18. Satellite Ranging

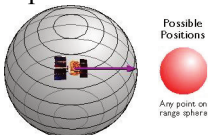
GPS receivers calculate distances to satellites as a function of the amount of time it takes for satellites’ signals to reach the ground. To make such a calculation, the receiver must be able to tell precisely when the signal was transmitted, and when it was received. The satellites are equipped with extremely accurate atomic clocks, so the timing of transmissions is always known. Receivers contain cheaper clocks, which tend to be sources of measurement error. The signals broadcast by satellites, called “pseudo-random codes,” are accompanied by the broadcast ephemeris data that describes the shapes of satellite orbits.



$Distance = Speed\ of\ Light \cdot Time\ Difference$

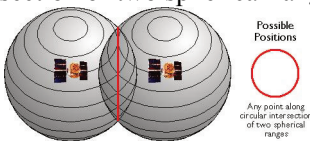
GPS receivers calculate distance as a function of the difference in time of broadcast and reception of a GPS signal. (Adapted from Hurn, 1989).

The GPS constellation is configured so that a minimum of four satellites is always “in view” everywhere on Earth. If only one satellite signal was available to a receiver, the set of possible positions would include the entire range sphere surrounding the satellite.



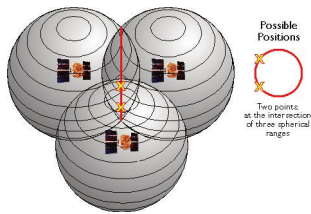
Set of possible positions of a GPS receiver relative to a single GPS satellite. (Adapted from Hurn, 1993).

If two satellites are available, a receiver can tell that its position is somewhere along a circle formed by the intersection of two spherical ranges.



Set of possible positions of a GPS receiver relative to two GPS satellites. (Adapted from Hurn, 1993).

If distances from three satellites are known, the receiver’s position must be one of two points at the intersection of three spherical ranges. GPS receivers are usually smart enough to choose the location nearest to the Earth’s surface. At a minimum, three satellites are required for a two-dimensional (horizontal) fix. Four ranges are needed for a three-dimensional fix (horizontal and vertical).



Set of possible positions of a GPS receiver relative to three GPS satellites. (Adapted from Hurn, 1993).

Satellite ranging is similar in concept to the plane surveying method **trilateration**, by which horizontal positions are calculated as a function of distances from known locations. The GPS satellite constellation is in effect an orbiting control network.

TRY THIS!

Trimble has a tutorial “designed to give you a good basic understanding of the principles behind GPS without loading you down with too much technical detail”. Check it out at http://www.trimble.com/gps_tutorial/. Click “Why GPS?” to get started.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about GPS Components. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

5.19. GPS Error Sources

A thought experiment (Wormley, 2004): Attach your GPS receiver to a tripod. Turn it on and record its position every ten minutes for 24 hours. Next day, plot the 144 coordinates your receiver calculated. What do you suppose the plot would look like?

Do you imagine a cloud of points scattered around the actual location? That’s a reasonable expectation. Now, imagine drawing a circle or ellipse that encompasses about 95 percent of the points. What would the radius of that circle or ellipse be? (In other words, what is your receiver’s positioning error?)

The answer depends in part on your receiver. If you used a hundred-dollar receiver, the radius of the circle you drew might be as much as ten meters to capture 95 percent of the points. If you used a WAAS-enabled, single frequency receiver that cost a few hundred dollars, your error ellipse might shrink to one to three meters or so. But if you had spent a few thousand dollars on a dual frequency, survey-grade receiver, your error circle radius might be as small as a centimeter or less. In general, GPS users get what they pay for.

As the market for GPS positioning grows, receivers are becoming cheaper. Still, there are lots of mapping applications for which it’s not practical to use a survey-grade unit. For example, if your assignment was to GPS 1,000 manholes for your municipality, you probably wouldn’t want to set up and calibrate a survey-grade receiver 1,000 times. How, then, can you minimize errors associated with mapping-grade receivers? A sensible start is to understand the sources of GPS error.

In this section you will learn to:

1. State the kinds and magnitude of error and uncertainty associated with uncorrected GPS positioning; and
2. Use a PDOP chart to determine the optimal times for GPS positioning at a given location and date.

Note: My primary source for the material in this section is Jan Van Sickle's text *GPS for Land Surveyors, 2nd Ed.* If you want a readable and much more detailed treatment of this material, I recommend Jan's book. See the bibliography at the end of this chapter for more information about this and other resources.

5.20. User Equivalent Range Errors

"UERE" is the umbrella term for all of the error sources below, which are presented in descending order of their contributions to the total error budget.

1. **Satellite clock:** GPS receivers calculate their distances from satellites as a function of the difference in time between when a signal is transmitted by a satellite and when it is received on the ground. The atomic clocks on board NAVSTAR satellites are extremely accurate. They do tend to stray up to one millisecond of standard GPS time (which is calibrated to, but not identical to, Coordinated Universal Time). The monitoring stations that make up the GPS "Control Segment" calculate the amount of clock drift associated with each satellite. GPS receivers that are able to make use of the clock correction data that accompanies GPS signals can reduce clock error significantly.

2. **Upper atmosphere (ionosphere):** Space is nearly a vacuum, but the atmosphere isn't. GPS signals are delayed and deflected as they pass through the ionosphere, the outermost layers of the atmosphere that extend from approximately 50 to 1,000 km above the Earth's surface. Signals transmitted by satellites close to the horizon take a longer route through the ionosphere than signals from satellites overhead, and are thus subject to greater interference. The ionosphere's density varies by latitude, by season, and by time of day, in response to the Sun's ultraviolet radiation, solar storms and maximums, and the stratification of the ionosphere itself. The GPS Control Segment is able to model ionospheric biases, however. Monitoring stations transmit corrections to the NAVSTAR satellites, which then broadcast the corrections along with the GPS signal. Such corrections eliminate only about three-quarters of the bias, however, leaving the ionosphere the second largest contributor to the GPS error budget.

3. **Receiver clock:** GPS receivers are equipped with quartz crystal clocks that are less stable than the atomic clocks used in NAVSTAR satellites. Receiver clock error can be eliminated, however, by comparing times of arrival of signals from two satellites (whose transmission times are known exactly).

4. **Satellite orbit:** GPS receivers calculate coordinates relative to the known locations of satellites in space. Knowing where satellites are at any given moment involves knowing the shapes of their orbits as well as their velocities. The gravitational attractions of the Earth, Sun, and Moon all complicate the shapes of satellite orbits. The GPS Control Segment monitors satellite locations at all times, calculates orbit eccentricities, and compiles these deviations in documents called ephemerides. An ephemeris is compiled for each satellite and broadcast with the satellite signal. GPS receivers that are able to process ephemerides can compensate for some orbital errors.

5. **Lower atmosphere:** (troposphere, tropopause, and stratosphere) The three lower layers of atmosphere encapsulate the Earth from surface to an altitude of about 50 km. The lower atmosphere delays GPS signals, adding slightly to the calculated distances between satellites and receivers. Signals from satellites close to the horizon are delayed the most, since they pass through more atmosphere than signals from satellites overhead.

6. **Multipath:** Ideally, GPS signals travel from satellites through the atmosphere directly to GPS receivers. In reality, GPS receivers must discriminate between signals received directly from satellites

and other signals that have been reflected from surrounding objects, such as buildings, trees, and even the ground. Some, but not all, reflected signals are identified automatically and rejected. Antennas are designed to minimize interference from signals reflected from below, but signals reflected from above are more difficult to eliminate. One technique for minimizing multipath errors is to track only those satellites that are at least 15° above the horizon, a threshold called the “mask angle.”

Douglas Welsh (personal communication, Winter 2001), an Oil and Gas Inspector Supervisor with Pennsylvania’s Department of Environmental Protection, wrote about the challenges associated with GPS positioning in our neck of the woods: “...in many parts of Pennsylvania the horizon is the limiting factor. In a city with tall buildings and the deep valleys of some parts of Pennsylvania it is hard to find a time of day when the constellation will have four satellites in view for any amount of time. In the forests with tall hardwoods, multipath is so prevalent that I would doubt the accuracy of any spot unless a reading was taken multiple times.” Van Sickle (2005) points out, however, that GPS modernization efforts and the GNSS may well ameliorate such gaps.

Have you had similar experiences with GPS? If so, please post a comment to this page .

5.21. Dilution of Precision

The arrangement of satellites in the sky also affects the accuracy of GPS positioning. The ideal arrangement (of the minimum four satellites) is one satellite directly overhead, three others equally spaced near the horizon (above the mask angle). Imagine a vast umbrella that encompasses most of the sky, where the satellites form the tip and the ends of the umbrella spines.

GPS coordinates calculated when satellites are clustered close together in the sky suffer from **dilution of precision (DOP)**, a factor that **multiplies the uncertainty associated with User Equivalent Range Errors (UERE** – errors associated with satellite and receiver clocks, the atmosphere, satellite orbits, and the environmental conditions that lead to multipath errors). The DOP associated with an ideal arrangement of the satellite constellation equals approximately 1, which does not magnify UERE. According to Van Sickle (2001), the lowest DOP encountered in practice is about 2, which doubles the uncertainty associated with UERE.

GPS receivers report several components of DOP, including Horizontal Dilution of Precision (HDOP) and Vertical Dilution of Precision (VDOP). The combination of these two components of the three-dimensional position is called PDOP – position dilution of precision. A key element of GPS mission planning is to identify the time of day when PDOP is minimized. Since satellite orbits are known, PDOP can be predicted for a given time and location. Various software products allow you to determine when conditions are best for GPS work.

MGIS student Jason Setzer (Winter 2006) offers the following illustrative anecdote:

“

I have had a chance to use GPS survey technology for gathering ground control data in my region and the biggest challenge is often the PDOP (position dilution of precision) issue. The problem in my mountainous area is the way the terrain really occludes the receiver from accessing enough satellite signals.

“

During one survey in Colorado Springs I encountered a pretty extreme example of this. Geographically, Colorado Springs is nestled right against the Rocky Mountain front ranges, with 14,000 foot Pike’s Peak just west of the city. My GPS unit was easily able to ‘see’ five, six or even seven satellites while I was on the eastern half of the city. However, the further west I traveled, I began to see progressively less of the

“

constellation, to the point where my receiver was only able to find one or two satellites. If a 180 degree horizon-to-horizon view of the sky is ideal, then in certain places I could see maybe 110 degrees.

“

There was no real work around, other than patience. I was able to adjust my survey points enough to maximize my view of the sky. From there it was just a matter of time... Each GPS bird has an orbit time of around twelve hours, so in a couple of instances I had to wait up to two hours at a particular location for enough of them to become visible. My GPS unit automatically calculates PDOP and displays the number of available satellites. So the PDOP value was never as low as I would have liked, but it did drop enough to finally be within acceptable limits. Next time I might send a vendor out for such a project!

TRY THIS!

Trimble, a leading manufacturer of GPS receivers, offers GPS mission planning software for free downloads. This activity will introduce you to the capabilities of the software, and will prepare you to answer questions about GPS mission planning later.

(The mission planning software is a Windows application (.exe). Mac users, as well as Windows users, see beneath the numbered steps.)

1. Visit the [Trimble website](#)
2. Download the Trimble planning software, install it on your computer (note where it is installing its *Common Files*), and launch the application.
3. Install an almanac: In the **Almanac** menu, move your cursor to **Important** in the submenu, choose **Almanac** | navigate to the folder where the Common Files were installed | select **almanac.alm** and click the **Open** button | click **OK**.
(If your Windows operating system is installed on your C-drive, then the path name to the almanac.alm file probably looks like this, with or without the “(x86)”:
C:\Program Files (x86)\Common Files\Trimble\Planning)
4. Choose **File** | **Station...** Choose a location at which you might wish to plan a GPS mission.
5. Choose **Satellite** | **Information** to explore the characteristics of active GPS, GLONASS, and WAAS satellites.
6. Choose **Graphs** | **DOP** | **DOP – position** to see how combined HDOP and VDOP vary throughout a selected 24-hour period at your selected location. Can you determine the best and worst times of day for GPS work?

While Trimble still makes available for free the planning software that you used above, they are not including access to an up to date almanac (information about currently available satellites). You may have noticed that the almanac you loaded was from 2010.

However, if you go [here](#) you will find an interactive interface that gives you access to a more current version of the same functionality as the planning app used above.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about GPS Error Sources. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

5.22. GPS Error Correction

A variety of factors, including the clocks in satellites and receivers, the atmosphere, satellite orbits, and reflective surfaces near the receiver, degrade the quality of GPS coordinates. The arrangement of satellites in the sky can make matters worse (a condition called dilution of precision). A variety of techniques have been developed to filter out positioning errors. Random errors can be partially overcome by simply averaging repeated fixes at the same location, although this is often not a very efficient solution. Systematic errors can be compensated for by modeling the phenomenon that causes the error and predicting the amount of offset. Some errors, like multipath errors caused when GPS signals are reflected from roads, buildings, and trees, vary in magnitude and direction from place to place. Other factors, including clocks, the atmosphere, and orbit eccentricities, tend to produce similar errors over large areas of the Earth's surface at the same time. Errors of this kind can be corrected using a collection of techniques called differential correction.

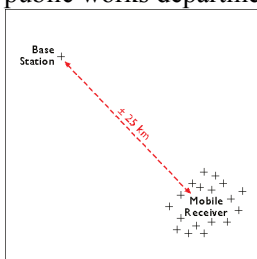
In this section you will learn to:

1. Explain the concept of differential correction and other methods used to improve the accuracy of GPS positioning; and
2. Perform differential correction using data and services of the U.S. National Geodetic Survey.

5.23. Differential Correction

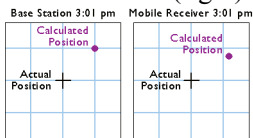
Differential correction is a class of techniques for improving the accuracy of GPS positioning by comparing measurements taken by two or more receivers. Here's how it works:

The locations of two GPS receivers—one stationary, one mobile—are illustrated below. The stationary receiver (or “base station”) continuously records its fixed position over a control point. The difference between the base station's actual location and its calculated location is a measure of the positioning error affecting that receiver at that location at each given moment. In this example, the base station is located about 25 kilometers from the mobile receiver (or “rover”). The operator of the mobile receiver moves from place to place. The operator might be recording addresses for an E-911 database, or trees damaged by gypsy moth infestations, or street lights maintained by a public works department



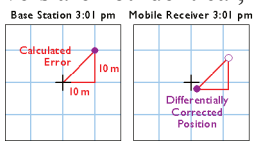
A GPS base station is fixed over a control point, while about 25 km away, a mobile GPS receiver is used to measure a series of positions.

The illustration below shows positions calculated at the same instant (3:01 pm) by the base station (left) and the mobile receiver (right).



Actual and calculated positions of a base station and mobile receiver.

The base station calculates the correction needed to eliminate the error in the position calculated at that moment from GPS signals. The correction is later applied to the position calculated by the mobile receiver at the same instant. The corrected position is not perfectly accurate because the kinds and magnitudes of errors affecting the two receivers are not identical, and because of the low frequency of the GPS timing code.



Error correction calculated at the base station is applied to the position calculated by the mobile receiver.



GPS base station used for differential correction. Notice that the antenna is located directly above a control point monument.

5.24. Real-Time Differential Correction

For differential correction to work, fixes recorded by the mobile receiver must be synchronized with fixes recorded by the base station (or stations). You can provide your own base station, or use correction signals produced from reference stations maintained by the U.S. Federal Aviation Administration, the U.S. Coast Guard, or other public agencies or private subscription services. Given the necessary equipment and available signals, synchronization can take place immediately (“real-time”) or after the fact (“post-processing”). First let’s consider real-time differential.

WAAS-enabled receivers are an inexpensive example of real-time differential correction. “WAAS” stands for **Wide Area Augmentation System**, a collection of about 25 base stations set up to improve GPS positioning at U.S. airport runways to the point that GPS can be used to help land airplanes (U.S. Federal Aviation Administration, 2007c). WAAS base stations transmit their measurements to a master station, where corrections are calculated and then uplinked to two geosynchronous satellites (19 are planned). The WAAS satellite then broadcasts differentially-corrected signals at the same frequency as GPS signals. WAAS signals compensate for positioning errors measured at WAAS base stations, as well as clock error corrections and regional estimates of upper-atmosphere errors (Yeazel, 2003). WAAS-enabled receivers devote one or two channels to WAAS signals, and are able to process the WAAS corrections. The WAAS network was designed to provide approximately 7-meter accuracy uniformly throughout its U.S. service area.

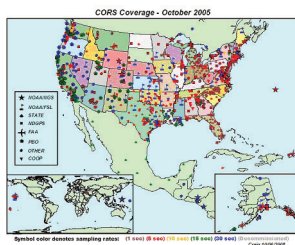
DGPS: The U.S. Coast Guard has developed a similar system, called the **Differential Global Positioning Service**. The DGPS network includes some 80 broadcast sites, each of which includes a survey-grade base station and a “radiobeacon” transmitter that broadcasts correction signals at 285-325 kHz (just below the AM radio band). DGPS-capable GPS receivers include a connection to a radio receiver that can tune in to one or more selected “beacons.” Designed for navigation at sea near U.S. coasts, DGPS provides accuracies no worse than 10 meters. Stephanie Brown (personal communication, Fall 2003) reported that where she works in Georgia, “with a good satellite constellation overhead, [DGPS accuracy] is typically 4.5 to 8 feet.”

Survey-grade real-time differential correction can be achieved using a technique called **real-time kinematic (RTK) GPS**. According to surveyor Laverne Hanley (personal communication, Fall 2000), “real-time kinematic requires a radio frequency link between a base station and the rover. I have achieved better than centimeter accuracy this way, although the instrumentation is touchy and requires great skill on the part of the operator. Several times I found that I had great GPS geometry, but had lost my link to the base station. The opposite has also happened, where I wanted to record positions and had a radio link back to the base station, but the GPS geometry was bad.”

5.25. Post-Processed Differential Correction

Kinematic positioning can deliver accuracies of 1 part in 100,000 to 1 part in 750,000 with relatively brief observations of only one to two minutes each. For applications that require accuracies of 1 part in 1,000,000 or higher, including control surveys and measurements of movements of the Earth’s tectonic plates, **static positioning** is required (Van Sickle, 2001). In static GPS positioning, two or more receivers measure their positions from fixed locations over periods of 30 minutes to two hours. The receivers may be positioned up to 300 km apart. Only **dual frequency, carrier phase differential** receivers capable of measuring the differences in time of arrival of the civilian GPS signal (L1) and the encrypted military signal (L2) are suitable for such high-accuracy static positioning.

CORS and OPUS: The U.S. National Geodetic Survey (NGS) maintains an Online Positioning User Service (OPUS) that enables surveyors to differentially-correct static GPS measurements acquired with a single dual frequency carrier phase differential receiver after they return from the field. Users upload measurements in a standard **Receiver INdependent EXchange format (RINEX)** to NGS computers, which perform differential corrections by referring to three selected base stations selected from a network of continuously operating reference stations. NGS oversees two CORS networks; one consisting of its 600 base stations of its own, another a cooperative of public and private agencies that agree to share their base station data and to maintain base stations to NGS specifications.



The Continuously Operating Reference Station network (CORS) (Snay, 2005)

The map above shows distribution of the combined national and cooperative CORS networks. Notice that station symbols are colored to denote the sampling rate at which base station data are stored. After 30 days, all stations are required to store base station data only in 30 second increments. This policy limits the utility of OPUS corrections to static positioning (although the accuracy of longer kinematic observations can also be improved). Mindful of the fact that the demand for static GPS is steadily declining, NGS’ future plans include streaming CORS base station data for real-time use in kinematic positioning.

TRY THIS!

This optional activity (contributed by Chris Piburn of CompassData Inc.) will guide you through the process of differentially-correcting static GPS measurements using the NGS' Online Positioning User Service (OPUS), which refers to the Continuously Operating Reference Station network (CORS).

The context is a CompassData project that involved a carrier phase differential GPS survey in a remote study area in Alaska. The objective was to survey a set of nine ground control points (GCPs) that would later be used to orthorectify a client's satellite imagery. So remote is this area that no NGS control point was available at the time the project was carried out. The alternative was to establish a base station for the project and to fix its position precisely with reference to CORS stations in operation elsewhere in Alaska.

The project team flew by helicopter to a hilltop located centrally within the study area. With some difficulty they hammered an 18 inch #5 rebar into the rocky soil to serve as a control monument. After setting up a GPS base station receiver over the rebar, they flew off to begin data collection with their rover receiver. Thanks to favorable weather, Chris and his team collected the nine required photo-identifiable GCPs on the first day. The centrally-located base station allowed the team to minimize distances between the base and the rover, which meant they could minimize the time required to fix each GCP. At the end of the day, the team had produced five hours of GPS data at the base station and nine fifteen-minute occupations at the GCPs

As you might expect, the raw GPS data were not sufficiently accurate to meet project requirements. (The various sources of random and systematic errors that contribute to the uncertainty of GPS data are considered elsewhere in this chapter.) In particular, the monument hammered into the hilltop was unsuitable for use as a control point because the uncertainty associated with its position was too great. The project team's first step in removing positioning errors was to post-process the data using baseline processing software, which adjusts computed baseline distances (between the base station and the nine GCPs) by comparing the phase of the GPS carrier wave as it arrived simultaneously at both the base station and the rover. The next step was to fix the position of the base station precisely in relation to CORS stations operating elsewhere in Alaska.

The following steps will guide you through the process of submitting the five hours of dual frequency base station data to the U.S. National Geodetic Survey's Online Positioning User Service (OPUS), and interpreting the results. (For [information about OPUS, go here](#))

1. Download the GPS data file. The compressed RINEX format file is approximately 6 Mb in size and will take about 1 minute to download via high speed DSL or cable, or about 15 minutes via 56 Kbps modem. If you can't download this file, contact me right away so we can help you resolve the problem.

- [WILD282u.zip](#) (5.8 Mb)

GPS receivers produce data in their manufacturers' proprietary formats. NGS requires that the GPS data be converted to the "Receiver INdependent EXchange" format (RINEX) for compatibility with OPUS. Most software packages that come with the GPS units themselves have a built in utility to convert their GPS data to RINEX format. NGS itself uses free conversion software provided by a non-profit, government-sponsored consortium called [UNAVCO](#).

2. Examine the RINEX file.

- Extract the RINEX file "WILD282u.05O" from its ZIP archive.
- Open "WILD282u.05O" with Microsoft Notepad or WordPad. (WordPad does a better job of preserving columns of text in this case. Set the word wrap to off.) Or, use another text editor. (In any open text editor window, in the File > Open dialog, make sure "Files of type:" is set to "All files" so that the target file is listed.)

The RINEX Observation file contains all the information about the signals that CompassData's base station receiver tracked during the Alaska survey. Explaining all the contents of the file is well beyond the scope of this activity. For

now, note the lines that disclose the antenna type, approximate position of the antenna, and antenna height. You'll report these parameters to OPUS in the next step.

3. Submit GPS data to OPUS

- Point your browser to the [OPUS home page](#) and enter the information needed in order to “Upload your data file.”
- (OPUS step 1) Click the Browse button to call up a File Upload window. Navigate to and upload the RINEX file you downloaded earlier. To streamline your submission, choose the compressed archive “WILD282u.zip”.
- (OPUS step 2) Select the antenna type. Refer to the line labeled “ANT # / TYPE” in the RINEX file you opened in your text editor. You should find the antenna type “TPSHIPER_PLUS”. Choose this type from the pull-down list.
- (OPUS step 3) Enter the height of the antenna above the monument. Refer to the line labeled “ANTENNA: DELTA H/E/N” in the RINEX file you opened in your text editor. The first value on that line, “1.0854”, is H, in meters. (See the [OPUS website](#) for more information about antenna height.)
- (OPUS step 4) Enter the email address to which you wish your results to be returned.
- (OPUS step 5) Options. OPUS allows users to specify a State Plane Coordinate system zone, to select or exclude particular CORS reference stations, to request standard or extended output, and to establish a user profile for use with future jobs. For this activity, no changes to the default settings are needed.
- (OPUS step 6) Click the “Upload to STATIC” button to submit the data for differential correction in relation to three CORS reference stations. Depending on how many requests are in NGS’ queue, results may be returned in minutes or hours. When this activity was tested, the queue included only one request (see window below, which appears after requests successfully submitted) and results were returned in just a few minutes, but in the past it has taken up to 10 minutes.



Upload results report, Online Positioning User Service, U.S. National Geodetic Survey.

When you receive your OPUS solution by return email, you will want to discover the magnitude of differential correction that OPUS calculated. To do this you'll need to determine (a) the uncorrected position originally calculated by the base station, (b) the corrected position calculated by OPUS, and (c) the mark-to-mark distance between the original and corrected positions. In addition to the original RINEX file you downloaded earlier, you'll need the OPUS solution and two free software utilities provided by NGS. Links to these utilities are listed below.

4. Determine the position of the base station receiver prior to differential correction.

- Refer to the RINEX file “WILD282u.05O” you opened in your text editor. The ninth line of the RINEX file lists the position of the base station receiver in Earth-Centered Earth-Fixed X, Y, Z coordinates. This is a three-dimensional Cartesian coordinate system whose origin is the Earth’s center of mass (like the NAD 83 datum). -2389892.2740 -1608765.8567 5672855.7386 APPROX POSITION XYZ

NGS provides a conversion utility to transform these X, Y, Z values to more familiar latitude and longitude coordinates and ellipsoidal heights.

- Go to [NGS’ XYZ to GEODETIC conversion](#).
- Enter the X, Y, Z coordinates found in the RINEX file. Your result should be:
 - Latitude = 63°13’53.74280” N
 - Longitude = 146°03’12.12710” W
 - Height = 1349.2248 m

5. Determine the corrected position of the base station receiver. The OPUS solution you receive by email reports corrected coordinates in Earth-Centered Earth-Fixed X, Y, Z, as geographic coordinates, and as UTM and State Plane coordinates. Look for the latitude and longitude coordinates and ellipsoidal height that are specified relative to the NAD 83 datum. They should be very close to:

- Latitude = 63°13’53.73892” N
- Longitude = 146°03’11.98942” W
- Height = 1348.756 m

6. Calculate the difference between the original and corrected base station positions. NGS provides another software utility to calculate the three-dimensional distance between two positions. Unlike the previous XYZ to GEODETIC converter, however, the “invers3d.exe” is a program you download to your computer.

- Download “[invers3d.exe](#)”
- Double-click on the file name to run the program, and choose the **geodetic coordinates** option.
- Paying close attention to the required formats, enter the uncorrected latitude, longitude, and ellipsoidal heights you calculated in step 4 above.
- Next, choose the **geodetic coordinates** option again, enter the corrected coordinates and height you calculated in step 5.
- Among the results, look for the calculated “mark-to-mark distance.” This is the magnitude of the OPUS correction.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about GPS Error Correction. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

5.26. Summary

Positions are a fundamental element of geographic data. Sets of positions form features, as the letters on this

page form words. Positions are produced by acts of measurement, which are susceptible to human, environmental, and instrument errors. Measurement errors cannot be eliminated, but systematic errors can be estimated, and compensated for.

Land surveyors use specialized instruments to measure angles and distances, from which they calculate horizontal and vertical positions. The Global Positioning System (and to a potentially greater extent, the emerging Global Navigation Satellite System) enables both surveyors and ordinary citizens to determine positions by measuring distances to three or more Earth-orbiting satellites. As you've read in this chapter (and may know from personal experience), GPS technology now rivals electro-optical positioning devices (i.e., "total stations" that combine optical angle measurement and electronic distance measurement instruments) in both cost and performance. This raises the question, "If survey-grade GPS receivers can produce point data with sub-centimeter accuracy, why are electro-optical positioning devices still so widely used?" In November 2005 I posed this question to two experts—Jan Van Sickle and Bill Toothill—whose work I had used as references while preparing this chapter. I also enjoyed a fruitful discussion with an experienced student named Sean Haile (Fall 2005). Here's what they had to say:

Jan Van Sickle, author of *GPS for Land Surveyors* and *Basic GIS Coordinates*, wrote:

“

In general it may be said that the cost of a good total station (EDM and theodolite combination) is similar to the cost of a good 'survey grade' GPS receiver. While a new GPS receiver may cost a bit more, there are certainly deals to be had for good used receivers. However, in many cases a RTK system that could offer production similar to an EDM requires two GPS receivers and there, obviously, the cost equation does not stand up. In such a case the EDM is less expensive.

Still, that is not the whole story. In some circumstances, such as large topographic surveys, the production of RTK GPS beats the EDM regardless of the cost differential of the equipment. Remember, you need line of sight with the EDM. Of course, if a topo survey gets too large, it is more cost effective to do the work with photogrammetry. And if it gets really large, it is most cost effective to use satellite imagery and remote sensing technology.

Now, let's talk about accuracy. It is important to keep in mind that GPS is not able to provide orthometric heights (elevations) without a geoid model. Geoid models are improving all the time, but are far from perfect. The EDM on the other hand has no such difficulty. With proper procedures it should be able to provide orthometric heights with very good relative accuracy over a local area. But, it is important to remember that relative accuracy over a local area with line of sight being necessary for good production (EDM) is applicable to some circumstances, but not others. As the area grows larger, as line of sight is at a premium, and a more absolute accuracy is required the advantage of GPS increases.

It must also be mentioned that the idea that GPS can provide cm level accuracy must always be discussed in the context of the question, 'relative to what control and on what datum?'

In relative terms, over a local area, using good procedures, it is certainly possible to say that an EDM can produce results superior to GPS in orthometric heights (levels) with some consistency. It is my opinion that this idea is the reason that it is rare for a surveyor to do detailed construction staking with GPS, i.e. curb and gutter, sewer, water, etc. On the other hand, it is common for surveyors to stake out property corners with GPS on a development site, and other features where the vertical aspect is not critical. It is not that GPS cannot provide very accurate heights, it is just that it takes more time

“ and effort to do so with that technology when compared with EDM in this particular area (vertical component).

It is certainly true that GPS is not well suited for all surveying applications. However, there is no surveying technology that is well suited for all surveying applications. On the other hand, it is my opinion that one would be hard pressed to make the case that any surveying technology is obsolete. In other words, each system has strengths and weaknesses and that applies to GPS as well.

Bill Toothill, professor in the Department of GeoEnvironmental Sciences and Engineering at Wilkes University, wrote:

“ GPS is just as accurate at short range and more accurate at longer distances than electro-optical equipment. The cost of GPS is dropping and may not be much more than a high end electro-optical instrument. GPS is well suited for all surveying applications, even though for a small parcel (less than an acre) traditional instruments like a total station may prove faster. This depends on the availability of local reference sites (control) and the coordinate system reference requirements of the survey.

Most survey grade GPS units (dual frequency) can achieve centimeter level accuracies with fairly short occupation times. In the case of RTK this can be as little as five seconds with proper communication to a broadcasting ‘base’. Sub-centimeter accuracies is another story. To achieve sub-centimeter, which most surveyors don’t need, requires much longer occupation times which is not conducive for ‘production’ work in a business environment. Most sub-centimeter applications are used for research, most of which are in the geologic deformation category. I have been using dual frequency GPS for the last eight years in Yellowstone National Park studying the deformation of the Yellowstone Caldera. To achieve sub-centimeter results we need at least 4-6 hours of occupation time at each point along a transect.

Sean Haile, a U.S. Park Service employee at Zion National Park whose responsibilities include GIS and GPS work, takes issue with some of these statements, as well as with some of the chapter material. While a student in this class in Fall 2005, Sean wrote:

“ A comparison of available products from [one manufacturer] shows that traditional technologies can achieve accuracy of 3mm. Under ideal conditions, the most advanced GPS equipment can only get down to 5mm accuracy, with real world results probably being closer to 10mm. It is true that GPS is often the faster and easier to use technology in the field when compared to electro-optical solutions, and with comparable accuracy levels has displaced traditional methods. If the surveyor needs to be accurate to the mm, however, electro-optical tools are more accurate than GPS.

There is no way, none, that you can buy a sub-centimeter unit anywhere for \$1000-2000. Yes, the prices are falling, but it has only been recently (last three years) that you could even buy a single channel sub-meter accuracy GPS unit for under \$10,000. The units you mention in the chapter for \$1000-2000, they would be ‘sell

“

your next of kin' expensive during that same time period. I am not in the business of measuring tectonic plates, but I deal with survey and mapping grade differential correction GPS units daily, so I can speak from experience on that one.

And Bill's response that GPS is well suited for all survey applications? Well I sincerely beg to differ. GPS is poorly suited for surveying where there is limited view of the horizon. You could wait forever and never get the required number of SVs. Even with mission planning. Obstructions such as high canopy cover, tall buildings, big rock walls... all these things can result in high multi-path errors, which can ruin data from the best GPS units. None of these things affect EDM. Yes, you can overcome poor GPS collection conditions (to an extent) by offsetting your point from a location where signal is good, but when you do that, you are taking the exact measurements (distance, angle) that you would be doing with an EDM except with an instrument that is not suited to that application!

The Global Navigation Satellite System (GNSS) may eventually overcome some of the limitations of GPS positioning. Still, these experts seem to agree that both GPS and electro-optical surveying methods are here to stay.

QUIZ

Registered Penn State students should return now to the Chapter 5 folder in ANGEL (via the Resources menu to the left) to access the graded quiz for this chapter. This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter's learning objectives. You are welcome to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 5.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

5.27. Bibliography

Brinker, R. C. & Wolf, P. R. (1984). *Elementary surveying* (7th ed.). New York: Harper and Row.

Dana, P. H. (1998). Global positioning system overview. *The geographer's craft project*. Retrieved August 2, 1999, from http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html

Doyle, D. R. (1994). Development of the National Spatial Reference System. Retrieved February 10, 2008, from http://www.ngs.noaa.gov/PUBS_LIB/develop_NSRS.html

Federal Geodetic Control Committee (1988). Geometric geodetic accuracy standards and specifications for using GPS relative positioning techniques. Retrieved February 10, 2008, from http://www.ngs.noaa.gov/FGCS/tech_pub/GeomGeod.pdf Retrieved September 14, 2013, from http://docs.lib.noaa.gov/noaa_documents/NOS/NGS/Geom_Geod_Accu_Standards.pdf

Hall, G. W. (1996). *USCG differential GPS navigation service*. Retrieved November 9, 2005, from <http://www.navcen.uscg.gov/pdf/dgps/dgpsdoc.pdf>

Hodgson, C. V. *Measuring base with invar tape. Tape underway. Base line and astro party*, ca. 1916. NOAA Historical Photo Collection (2004). Retrieved on April 20, 2006, from <http://www.photolib.noaa.gov/>.

Hurn, J. (1989). *GPS: A guide to the next utility*. Sunnyvale CA: Trimble Navigation Ltd.

Hurn, J. (1993). *Differential GPS Explained*. Sunnyvale CA: Trimble Navigation Ltd.

Monmonier, M. (1995). Boundary litigation and the map as evidence. In *Drawing the Line: Tales of Maps and Cartocontroversy*. New York: Henry Holt.

National Geodetic Survey (n. d.). Retrieved November 4, 2009, from <http://www.ngs.noaa.gov>

National Geodetic Survey (n.d.). *National Geodetic Survey – CORS, Continuously Operating Reference Station*. Retrieved August 15, 2012, from <http://www.ngs.noaa.gov/CORS/>

NAVSTAR GPS Joint Program Office. Retrieved October 21, 2000, from <http://gps.losangeles.af.mil/>

Norse, E. T. (2004). *Tracking new signals from space – GPS modernization and Trimble R-Track Technology*. Retrieved November 9, 2005, from http://www.trimble.com/survey_wp_gpssys.asp?Nav=Collection-27596 Retrieved September 14, 2013, from http://www.geosystems.co.nz/drupal/files/u1/images/construction/R-Track_technology_and_GPS_Modernization.pdf

Raisz, E. (1948). McGraw-Hill series in geography: *General cartography* (2nd ed.). York, PA: The Maple Press Company.

Robinson, A. et al. (1995). *Elements of cartography* (5th ed.). New York: John Wiley & Sons.

Smithsonian National Air and Space Museum (1998). *GPS: A new constellation*. Retrieved August 2, 1999, from <http://www.nasm.si.edu/gps/>

Snay, R. (2005, September 13). *CORS users forum – towards real-time positioning*. Power point presentation presented at the 2005 CORS Users Forum, Long Beach, CA. Presentation retrieved October 26, 2005, from http://www.ngs.noaa.gov/CORS/Presentations/CORSForum2005/Richard_Snay_Forum2005.pdf

Thompson, M. M. (1988). *Maps for America, cartographic products of the U.S. Geological Survey and others* (3d ed.). Reston, Va.: U.S. Geological Survey.

U.S. Coast Guard Navigation Center (n. d.). *DGPS general information*. Retrieved February 10, 2008, from <http://www.navcen.uscg.gov/?pageName=dgpsMainwww.navcen.uscg.gov/>

U.S. Federal Aviation Administration (2007a). *Frequently asked questions*. Retrieved February 10, 2008, from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/faq/gps/

U.S. Federal Aviation Administration (2007b). *Global Positioning System: How it works*. Retrieved February 10, 2008, from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/gps/howitworks/

U.S. Federal Aviation Administration. (2007c). *Wide Area Augmentation System*. Retrieved February 10, 2008, from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/gps/howitworks/

Van Sickel, J. (2001). *GPS for land surveyors*. New York: Taylor and Francis.

Van Sickel, J. (2004). *Basic GIS coordinates*. Boca Raton: CRC Press.

Wolf, P. R. & Brinker, R. C. (1994). *Elementary surveying* (9th ed.). NY, NY: HarperCollins College Publisher.

Wormley, S. (2006). *GPS errors and estimating your receiver's accuracy*. Retrieved April 20, 2006, from http://www.edu-observatory.org/gps/gps_accuracy.html

Yeazel, J. (2006). *WAAS and its relation to enabled hand-held GPS receivers*. Retrieved October 12, 2005, from <http://gpsinformation.net/exe/waas.html>

Chapter 6

National Spatial Data Infrastructure I

David DiBiase

6.1. Overview

Chapters 6 and 7 consider the origins and characteristics of the **framework data themes** that make up the United States' proposed National Spatial Data Infrastructure (NSDI). The seven themes include geodetic control, orthoimagery, elevation, transportation, hydrography, government units (administrative boundaries), and cadastral (property boundaries). **Most framework data, like the printed topographic maps that preceded them, are derived directly or indirectly from aerial imagery.** Chapter 6 introduces the field of photogrammetry, which is concerned with the production of geographic data from aerial imagery. The chapter begins by considering the nature and status of the U.S. NSDI in comparison with other national mapping programs. It considers the origins and characteristics of the geodetic control and orthoimagery themes. The remaining five themes are the subject of Chapter 7.

Objectives

Students who successfully complete Chapter 6 should be able to:

1. Explain how the distribution of authority for mapping and land title registration among various levels of government affects the availability of framework data;
2. Describe how topographic data are compiled from aerial imagery;
3. Explain the difference between a vertical aerial photograph and an orthoimage;
4. List and describe characteristics and status of the USGS National Map; and
5. Discuss the relationship between the National Map and the NSDI framework.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

6.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated

quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 6 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 6	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	Submit two practice quizzes including: <ul style="list-style-type: none"> National Spatial Data Legacies Photogrammetry Practice quizzes are not graded and may be submitted more than once.	Go to ANGEL > [your course section] > Lessons tab > Chapter 6 folder > [quiz]
3	Perform “Try this” activities including: <ul style="list-style-type: none"> Compare data copyright policies of the U.S. and Britain Search for USGS topographic maps and aerial imagery View and explore a digitally scanned topographic map View and explore a digital orthophoto Assess the availability of digital orthophotos for your area “Try this” activities are not graded.	Instructions are provided for each activity.
4	Submit the Chapter 6 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 6 folder > Chapter 6 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

6.3. National Geographic Information Strategies

In 1998 Ian Masser published a comparative study of the national geographic information strategies of four developed countries: Britain (England and Wales), the Netherlands, Australia, and the U.S. Masser built upon earlier

work which found that countries with relatively low levels of digital data availability and GIS diffusion also tended to be countries where there had been a fragmentation of data sources in the absence of central or local government coordination” (p. ix). Comparing his four case studies in relation to the seven framework themes identified for the U.S. NSDI, Masser found considerable differences in data availability, pricing, and intellectual property protections. Differences in availability of core data, he found, are explained by the ways in which responsibilities for mapping and for land titles registration are distributed among national, state, and local governments in each country.

The following table summarizes those distributions of responsibilities.

Distributions of Responsibilities

	Britain (England & Wales)	Netherlands	Australia	United States
Central government	Land titles registration, small- and large-scale mapping, statistical data	Land titles registration, small- and large-scale mapping, statistical data	Some small-scale mapping, statistical data	Small-scale mapping, statistical data
State/Territorial government	Not applicable	Not applicable	Land titles registration, small- and large-scale mapping	Some land titles registration and small- and large-scale mapping
Local government	None	large-scale mapping, population registers	Some large-scale mapping	Land titles registration, large-scale mapping

Distribution of responsibilities among different levels of government (Masser, 1998).

Masser’s analysis helps to explain what geospatial professionals in the U.S. have known all along — that the coverage of **framework data in the U.S. is incomplete or fragmented because thousands of local governments are responsible for large-scale mapping and land titles registration, and because these activities tend to be poorly coordinated.** In contrast, core data coverage is more or less complete in Australia, the Netherlands, and Britain, where central and state governments have authority over large-scale mapping and land-titles registration.

Other differences among the four countries relate to fees charged by governments to use the geographic and statistical data they produce, as well as the copyright protections they assert over the data. **U.S. federal government agencies, Masser notes, differ from their counterparts by charging no more than the cost of reproducing their data in forms suitable for delivery to customers.** State and local government policies in the U.S. vary considerably, however. Longstanding debates persist in the U.S. about the viability and ethics of recouping costs associated with public data.

The U.S. also differs starkly from Britain and Australia in regards to copyright protection. Most data published by the U.S. Geological Survey or U.S. Census Bureau resides in the public domain and may be used without restriction. U.K. Ordnance Survey data, by contrast, is protected by Crown copyright, and is available for use by others for fees and under the terms of restrictive licensing agreements. **One consequence of the federal government’s decision to release its geospatial data to the public domain, some have argued, was the early emergence of a vigorous geospatial industry in the U.S.**

TRY THIS!

To learn more about the Crown copyright policy of the Great Britain’s Ordnance Survey, search the Internet for “ordnance survey crown copyright.”

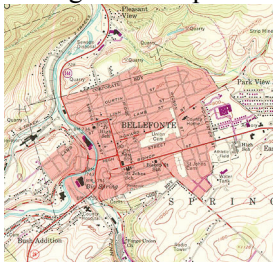
The [USGS policy](#) is explained [here](#) (or search on “acknowledging usgs as information source”)

6.4. Legacy Data: USGS Topographic Maps

“

Since the eighteenth century, the preparation of a detailed basic reference map has been recognized by the governments of most countries as fundamental for the delimitation of their territory, for underpinning their national defense and for management of their resources (Parry, 1987).

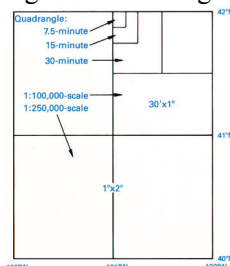
Specialists in geographic information recognize two broad functional classes of maps, reference maps and thematic maps. As you recall from Chapter 3, a thematic map is usually made with one particular purpose in mind. Often, the intent is to make a point about the spatial pattern of a single phenomenon. **Reference maps**, on the other hand, are designed to serve many different purposes. Like a reference book, such as a dictionary, encyclopedia, or gazetteer, reference maps help people look up facts. Common uses of reference maps include locating place names and features, estimating distances, directions, and areas, and determining preferred routes from starting points to a destination. Reference maps are also used as base maps upon which additional geographic data can be compiled. Because reference maps serve various uses, they typically include a greater number and variety of symbols and names than thematic maps. The portion of the United States Geological Survey (USGS) topographic map shown below is a good example.



A typical reference map. A portion of a USGS topographic quadrangle map (USGS, 1971)

The term **topography** derives from the Greek *topographein*, “to describe a place.” Topographic maps show, and name, many of the visible characteristics of the landscape, as well as political and administrative boundaries. Topographic map series provide base maps of uniform scale, content, and accuracy (more or less) for entire territories. Many national governments include agencies responsible for developing and maintaining topographic map series for a variety of uses, from natural resource management to national defense. Affluent countries, countries with especially valuable natural resources, and countries with large or unusually active militaries, tend to be mapped more completely than others.

The systematic mapping of the entire U.S. began in 1879, when the U.S. Geological Survey (USGS) was established. Over the next century USGS and its partners created topographic map series at several scales, including 1:250,000, 1:100,000, 1:63,360, and 1:24,000. The diagram below illustrates the relative extents of the different map series. Since much of today’s digital map data was digitized from these topographic maps, one of the challenges of creating continuous digital coverage of the entire U.S. is to seam together all of these separate map sheets.



Relative extents of the several USGS quadrangle map series. (Thompson, 1988).

Map sheets in the 1:24,000-scale series are known as **quadrangles** or simply quads. A quadrangle is a four-sided polygon. Although each 1:24,000 quad covers 7.5 minutes longitude by 7.5 minutes latitude, their shapes and area coverage vary. The area covered by the 7.5-minute maps varies from 49 to 71 square miles (126 to 183 square kilometers), because the length of a degree of longitude varies with latitude.



Topographer compiling topographic map using a plane table and alidade (NOAA, 2007).

Through the 1940s, topographers in the field compiled by hand the data depicted on topographic maps.

Anson (2002) recalls being outfitted with a 14 inch x 14 inch tracing table and tripod, plus an alidade [a 12 inch telescope mounted on a brass ruler], a 13 foot folding stadia rod, a machete, and a canteen... (p. 1). Teams of topographers sketched streams, shorelines, and other water features; roads, structures, and other features of the built environment; elevation contours, and many other features. To ensure geometric accuracy, their sketches were based upon geodetic control provided by land surveyors, as well as positions and spot elevations they surveyed themselves using alidades and rods. Depending on the terrain, a single 7.5-minute quad sheet might take weeks or months to compile. **In the 1950s, however, photogrammetric methods** involving stereoplotters that permitted topographers to make accurate stereoscopic measurements directly from overlapping pairs of aerial photographs provided a viable and more efficient alternative to field mapping. We'll consider photogrammetry in greater detail later on in this chapter.

By 1992 the series of over 53,000 separate quadrangle maps covering the lower 48 states, Hawaii, and U.S. territories at 1:24,000 scale was completed, at an estimated total cost of \$2 billion. However, by the end of the century the average age of 7.5-minute quadrangles was over 20 years, and federal budget appropriations limited revisions to only 1,500 quads a year (Moore, 2000). As landscape change has exceeded revisions in many areas of the U.S., the USGS topographic map series has become legacy data outdated in terms of format as well as content.

TRY THIS!

Search the Internet on “USGS topographic maps” to investigate the history and characteristics of USGS topographic maps in greater depth. View preview images, look up publication and revision dates, and order topographic maps at “USGS Store.”

6.5. Accuracy Standards

Errors and uncertainty are inherent in geographic data. Despite the best efforts of the USGS Mapping Division and its contractors, topographic maps include features that are out of place, features that are named or symbolized incorrectly, and features that are out of date.

As discussed in Chapter 2, the locational accuracy of spatial features encoded in USGS topographic maps and data are guaranteed to conform to National Map Accuracy Standards. The standard for topographic maps state that horizontal positions of 90 percent of the well-defined points tested will occur within 0.02 inches (map distance) of their actual positions. Similarly, the vertical positions of 90 percent of well-defined points tested are to be true to within one-half of the contour interval. Both standards, remember, are scale-dependent.

Objective standards do not exist for the accuracy of attributes associated with geographic features. Attribute errors certainly do occur, however. A chronicler of the national mapping program (Thompson, 1988, p. 106) recalls

a worried user who complained to USGS that “My faith in map accuracy received a jolt when I noted that on the map the borough water reservoir is shown as a sewage treatment plant.”

The passage of time is perhaps the most troublesome source of errors on topographic maps. As mentioned in the previous page, the average age of USGS topographic maps is over 20 years. Geographic data quickly lose value (except for historical analyses) unless they are continually revised. The sequence of map fragments below shows how frequently revisions were required between 1949 and 1973 for the quad that covers Key Largo, Florida. Revisions are based primarily on geographic data produced by aerial photography.

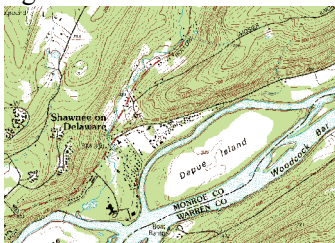
Geographic data quickly lose value if they are not kept up to date. (Thompson, 1988). Select each of the years to view the revised map. Note: You need to have the Adobe Flash player installed in order to see and interact with this illustration. You can download Flash Player free at <http://www.adobe.com/flash>.

TRY THIS!

Investigate standards for data quality and other characteristics of U.S. national map data [here](#) or by searching the Internet for “usgs national map accuracy standards”

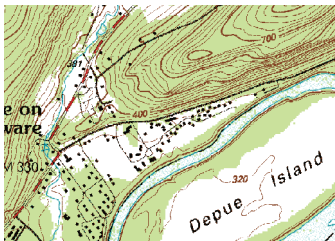
6.6. Scanned Topographic Maps

Many digital data products have been derived from the USGS topographic map series. The simplest of such products are **Digital Raster Graphics (DRGs)**. DRGs are scanned raster images of USGS 1:24,000 topographic maps. DRGs are useful as backdrops over which other digital data may be superimposed. For example, the accuracy of a vector file containing lines that represent lakes, rivers, and streams could be checked for completeness and accuracy by plotting it over a DRG.



Portion of a Digital Raster Graphic (DRG) for Bushkill, PA

DRGs are created by scanning paper maps at 250 pixels per inch resolution. Since at 1:24,000 1 inch on the map represents 2,000 feet on the ground, each DRG pixel corresponds to an area about 8 feet (2.4 meters) on a side. Each pixel is associated with a single attribute: a number from 0 to 12. The numbers stand for the 13 standard DRG colors.



Magnified portion of a Digital Raster Graphic (DRG) for Bushkill, PA

Like the paper maps from which they are scanned, DRGs comply with [National Map Accuracy Standards](#). A subset of the more than 50,000 DRGs that cover the lower 48 states have been sampled and tested for completeness and positional accuracy.

DRGs conform to the Universal Transverse Mercator projection used in the local UTM zone. The scanned

images are transformed to the UTM projection by matching the positions of 16 control points. Like topographic quadrangle maps, all DRGs within one UTM zone can be fit together to form a mosaic after the map “collars” are removed.

To investigate DRGs in greater depth, visit the [USGS Topomaps website](#) or search the Internet on “USGS Digital Raster Graphics”

TRY THIS!

EXPLORE A DRG WITH GLOBAL MAPPER (DLGV32 PRO)

You can use a free software application called **Global Mapper** (also known as **dlgv32 Pro**) to investigate the characteristics of a USGS Digital Raster Graphic. Originally developed by the staff of the USGS Mapping Division at Rolla, Missouri as a data viewer for USGS data, Global Mapper has since been commercialized, but is available in a free trial version. The instructions below will guide you through the process of installing the software and opening the DRG data. Penn State students will later be asked questions that will require you to explore the data for answers. Note: Global Mapper is a Windows application and will not run under the Macintosh operating system. The questions asked of Penn State students that involve the use of Global Mapper are not graded.

GLOBAL MAPPER (DLGV32 PRO) INSTALLATION INSTRUCTIONS

Skip this step if you already downloaded and installed Global Mapper or dlgv32 Pro.

1. Navigate to globalmapper.com or search the Internet for “Global Mapper” or “dlgv32 Pro”
2. Download the trial version of the software.
3. Double-click on the setup file you downloaded to install the program.
4. Launch Global Mapper or dlgv32 Pro.

DOWNLOADING AND EXPLORING DRG DATA IN GLOBAL MAPPER

1. First, create a directory called “USGS Data” on your hard disk, where you can file your course materials if you haven’t done so already.
2. [Download the DRG.zip data archive](#). The ZIP archive is 2.7 Mb in size and will take approximately 35 seconds to download via high speed DSL or cable, or about 9 minutes and 35 seconds minutes via 56 Kbps modem. Registered Penn State students who cannot download the file should contact their assigned teaching assistant for help.
3. Now decompress the archive into a directory on your hard disk.
 - Open the ZIP archive you downloaded.
 - Extract all files in the ZIP archive into a known subdirectory.

The result will be five files that make up one Digital Raster Graphic.

1. Open your DRG in Global Mapper
 - Choose File > Open Data File(s)..., then navigate to the subdirectory into which you extracted the DRG files.
 - Open the file ‘bushkill_pa.tif’

The DRG data correspond with the 7.5 minute quadrangle for Bushkill, PA.

1. Notice that as you **glide the magnifying glass cursor over the DRG**, the UTM (NAD 27) and geographic coordinates of the cursor's position change in the lower right-hand corner of the window. This tells you that the DRG is in fact **georeferenced**.
2. Experiment with Global Mapper's tools. Use the **Zoom** and **Pan** tools to magnify and scroll across the DRG. The Full View button (the one with the house icon) refreshes the initial full view of the data set.
3. The **Measure** tool (ruler icon) allows you to not only measure distance as the crow flies, but also to see the area enclosed by a series of line segments drawn by repeated mouse clicks. Note again the location information that is given to you near the bottom of the application window.

Certain tools, e.g., the 3D Path Profile/Line of Sight tool are not functional in the free (unregistered) version of Global Mapper.

1. To view an excerpt from the DRG metadata, navigate to Tools > Control Center, then click the **Metadata** button.

6.7. Federal Geographic Data Committee

Even before the USGS completed its nationwide 7.5-minute quadrangle series, the U.S. federal government had begun to rethink and reorganize its national mapping program. In 1990 the U.S. Office of Management and Budget issued Circular A-16, which established the Federal Geographic Data Committee (FGDC) as the interagency coordinating body responsible for facilitating cooperation among federal agencies whose missions include producing and using geospatial data. FGDC is chaired by the Department of Interior, and is administered by USGS.

In 1994 President Bill Clinton's Executive Order 12906 charged the FGDC with coordinating the efforts of government agencies and private sector firms leading to a **National Spatial Data Infrastructure (NSDI)**. The Order defined NSDI as "the technology, policies, standards and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data" (White House, 1994). It called upon FGDC to establish a National Geospatial Data Clearinghouse, ordered federal agencies to make their geospatial data products available to the public through the Clearinghouse, and required them to document data in a standard format that facilitates Internet search. Agencies were required to produce and distribute data in compliance with standards established by FGDC. (The Departments of Defense and Energy were exempt from the order, as was the Central Intelligence Agency.)

Finally, the Order charged FGDC with preparing an implementation plan for a National Digital Geospatial Data Framework, the "data backbone of the NSDI" (FGDC, 1997, p. v). The seven core data themes that comprise the NSDI Framework are listed below, along with the government agencies that have lead responsibility for creating and maintaining each theme. Later on in this chapter, and in the one that follows, we'll investigate the framework themes one by one.

NSDI Framework

Geodetic Control	Department of Commerce, National Oceanographic and Atmospheric Administration, National Geodetic Survey
Orthoimagery	Department of Interior, U.S. Geological Survey
Elevation	Department of Interior, U.S. Geological Survey
Transportation	Department of Transportation
Hydrography	Department of Interior, U.S. Geological Survey
Administrative units (boundaries)	Department of Commerce, U.S. Census Bureau
Cadastral	Department of Interior, Bureau of Land Management

Seven data themes that comprise the NSDI Framework and the government agencies responsible for each.

6.8. USGS National Map

Executive Order 12906 decreed that a designee of the Secretary of the Department of Interior would chair the Federal Geographic Data Committee. The USGS, an agency of the Department of Interior, has lead responsibility for three of the seven NSDI framework themes—orthoimagery, elevation, and hydrography, and secondary responsibility for several others. In 2001, USGS announced its vision of a National Map that “aligns with the goals of, and is one of several USGS activities that contribute to, the National Spatial Data Infrastructure” (USGS, 2001, p. 31). A 2002 report of the National Research Council identified the National Map as the most important initiative of USGS’ Geography Discipline at the USGS (NRC, 2002). Recognizing its unifying role across its science disciplines, USGS moved management responsibility for the National Map from Geography to the USGS Geospatial Information Office in 2004. (One reason that the term “geospatial” is used at USGS and elsewhere is to avoid association of GIS with a particular discipline, i.e. Geography.)

In 2001, USGS envisioned the National Map as the Nation’s topographic map for the 21st Century (USGS, 2001, p.1). Improvements over the original topographic map series were to include:

Characteristics of the National Map

Currentness	Content will be updated on the basis of changes in the landscape instead of the cyclical inspection and revisions cycles now in use [for printed topographic map series]. The ultimate goal is that new content be incorporated with seven days of a change in the landscape.
Seamlessness	Features will be represented in their entirety and not interrupted by arbitrary edges, such as 7.5-minute map boundaries.
Consistent classification	Types of features, such as “road” and “lake/pond,” will be identified in the same way throughout the Nation.
Variable resolution	Data resolution, or pixel size, may vary among imagery of urban, rural, and wilderness areas. The resolution of elevation data may be finer for flood plain, coastal, and other areas of low relief than for areas of high relief.
Completeness	Data content will include all mappable features (as defined by the applicable content standards for each data theme and source).
Consistency and integration	Content will be delineated geographically (that is, in its true ground position within the applicable accuracy limit) to ensure logical consistency between related features. For example, ... streams and rivers [should] consistently flow downhill...
Variable positional accuracy	The minimum positional accuracy will be that of the current primary topographic map series for an area. Actual positional accuracy will be reported in conformance with the Federal Geographic Data Committee’s Geospatial Positioning Accuracy Standard.
Spatial reference systems	Tools will be provided to integrate data that are mapping using different datums and referenced to different coordinates systems, and to reproject data to meet user requirements.
Standardized content	...will conform to appropriate Federal Geographic Data Committee, other national, and/or international standards.
Metadata	At a minimum, metadata will meet Federal Geographic Data Committee standards to document ... [data] lineage, positional and attribute accuracy, completeness, and consistency.

Characteristics of the National Map (USGS, 2001, p. 11-13.)

As of 2008, USGS’ ambitious vision has not yet been fully realized. Insofar as it depends upon cooperation by many federal, state and local government agencies, the vision may never be fully achieved. Still, elements of a National Map do exist, including national data themes, data access and dissemination technologies such as the [Geospatial One Stop portal](#) and the [National Map viewer](#), and the [U.S. National Atlas](#). A new Center of Excellence for Geospatial Information Science (CEGIS) has been established under the USGS Geospatial Information Office to undertake the basic GIScience research needed to devise and implement advanced tools that will make the National Map more valuable to end users.

The data themes included in the National Map are shown in the following table, in comparison to the NSDI framework themes outlined earlier in this chapter. As you see, the National Map themes align with five of the seven framework themes, but do not include geodetic control and cadastral data. Also, the National Map adds land cover and geographic names, which are not included among the NSDI framework themes. **Given USGS’ leadership role in FGDC, why do the National Map themes deviate from the NSDI framework?** According to the Committee on Research Priorities for the USGS Center of Excellence for Geospatial Science, “these themes were selected because USGS is authorized to provide them if no other sources are available, and [because] they typically comprise the information portrayed on USGS topographic maps (NRC, 2007, p. 31).

Data Themes

	National Map Themes	NSDI Framework Themes
Geodetic Control	No	Yes
Orthoimagery	Yes	Yes
Land Cover	Yes	No
Elevation	Yes	Yes
Transportation	Yes	Yes
Hydrography	Yes	Yes
Boundaries	Yes	Yes
Structures	Yes	No
Cadastral	No	Yes
Geographic Names	Yes	No

Comparison of data themes included in the National Map and NSDI framework.

The following sections of this chapter, and the one that follows, will describe the derivation, characteristics, and status of the seven NSDI themes in relation to the National Map. Chapter 8, Remotely Sensed Image Data, will include a description of the National Land Cover Data program that provides the land cover theme of the National Map. Registered students used the [USGS Geographic Information Names Information System](#) for a project assignment.

6.9. Theme: Geodetic Control

In the U.S. the National Geodetic Survey (NGS) maintains a national geodetic control network called the **National Spatial Reference System (NSRS)**. The NSRS includes approximately 300,000 horizontal and 600,000 vertical control points (Doyle, 1994). High-accuracy control networks are needed for mapping projects that span large areas; to design and maintain interstate transportation corridors including highways, pipelines, and transmission lines; and to monitor tectonic movements of the Earth's crust and sea level changes, among other applications (FGDC, 1998a).

Some control points are more accurate than others, depending on the methods surveyors used to establish them. [The Chapter 5 page titled "Survey Control"](#) outlines the accuracy classification adopted in 1988 for control points in the NSRS. As geodetic-grade GPS technology has become affordable for surveyors, expectations for control network accuracy have increased. In 1998, the FGDC's Federal Geodetic Control Subcommittee published a set of [Geospatial Positioning Accuracy Standards](#). One of these is the Standards for Geodetic Networks (FGDC, 1998a). The table below presents the latest accuracy classification for horizontal coordinates and heights (ellipsoidal and orthometric). For example, the theoretically infinitesimal location of a horizontal control point classified as "1-Millimeter" must have a 95% likelihood of falling within a 1 mm "radius of uncertainty" (FGDC, 1998b, 1-5).

Accuracy Classifications

Accuracy Classification	Radius of Uncertainty (95% confidence)
1-Millimeter	0.001 meters
2-Millimeter	0.002 meters
5-Millimeter	0.005 meters
1-Centimeter	0.010 meters
2-Centimeter	0.020 meters
5-Centimeter	0.050 meters
1-Decimeter	0.100 meters
2-Decimeter	0.200 meters
5-Decimeter	0.500 meters
1-Meter	1.000 meters
2-Meter	2.000 meters
5-Meter	5.000 meters
10-Meter	10.000 meters

Accuracy classification for geodetic control networks (FGDC, 1998).

If in Chapter 2 you retrieved a NGS datasheet for a control point, you probably found that the accuracy of your point was reported in terms of the 1988 classification. If yours was a “first order” (C) control point, its accuracy classification is 1 centimeter. NGS does plan to upgrade the NSRS, however. Its 10-year strategic plan states that “the geodetic latitude, longitude and height of points used in defining NSRS should have an absolute accuracy of 1 millimeter at any time” (NGS, 2007, 8).

THINK ABOUT IT

Why does the 1998 standard refer to absolute accuracies while the 1988 standard (outlined in Chapter 5) is defined in terms of maximum error relative to distance between two survey points? What changed between 1988 and 1998 in regard to how control points are established?

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 6 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about National Spatial Data Legacies. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

6.10. Theme: Orthoimagery

The Federal Geographic Data Committee (FGDC, 1997, p. 18) defines **orthoimage** as “a georeferenced image prepared from an aerial photograph or other remotely sensed data ... [that] has the same metric properties as a map and has a uniform scale.” Unlike orthoimages, the scale of ordinary aerial images varies across the image,

due to the changing elevation of the terrain surface (among other things). The process of creating an orthoimage from an ordinary aerial image is called **orthorectification**. Photogrammetrists are the professionals who specialize in creating orthorectified aerial imagery, and in compiling geometrically-accurate vector data from aerial images. So, to appreciate the requirements of the orthoimagery theme of the NSDI framework, we first need to investigate the field of **photogrammetry**.

6.11. Photogrammetry

Photogrammetry is a profession concerned with producing precise measurements of objects from photographs and photoimagery. One of the objects measured most often by photogrammetrists is the surface of the Earth. Since the mid-20th century, aerial images have been the primary source of data used by USGS and similar agencies to create and revise topographic maps. Before then, topographic maps were compiled in the field using magnetic compasses, tapes, plane tables (a drawing board mounted on a tripod, equipped with an leveling telescope like a transit), and even barometers to estimate elevation from changes in air pressure. Although field surveys continue to be important for establishing horizontal and vertical control, photogrammetry has greatly improved the efficiency and quality of topographic mapping.

A straight line between the center of a lens and the center of a visible scene is called an **optical axis**. A **vertical aerial photograph** is a picture of the Earth's surface taken from above with a camera oriented such that its optical axis is vertical. In other words, when a vertical aerial photograph is exposed to the light reflected from the Earth's surface, the sheet of photographic film (or an digital imaging surface) is parallel to the ground. In contrast, an image you might create by snapping a picture of the ground below while traveling in an airplane is called an oblique aerial photograph, because the camera's optical axis forms an oblique angle with the ground.



A vertical aerial photograph (National Aerial Photography Program, June 28, 1994).

The **nominal scale** of a vertical air photo is equivalent to f / H , where f is the focal length of the camera (the distance between the camera lens and the film — usually six inches), and H is the flying height of the aircraft above the ground. It is possible to produce a vertical air photo such that scale is consistent throughout the image. This is only possible, however, if the terrain in the scene is absolutely flat. In rare cases where that condition is met, topographic maps can be compiled directly from vertical aerial photographs. Most often however, **air photos of variable terrain need to be transformed, or rectified, before they can be used as a source for mapping.**

Government agencies at all levels need up-to-date aerial imagery. Early efforts to sponsor complete and recurring coverage of the U.S. included the [National Aerial Photography Program](#), which replaced an earlier National High Altitude Photography program in 1987. NAPP was a consortium of federal government agencies that aimed to jointly sponsor vertical aerial photography of the entire lower 48 states every seven years or so at an altitude of 20,000 feet, suitable for producing topographic maps at scales as large as 1:5,000. More recently NAPP has been eclipsed by another consortium called the [National Agricultural Imagery Program](#). According to student Anne O'Connor (personal communication, Spring 2004), who represented the Census Bureau in the consortium

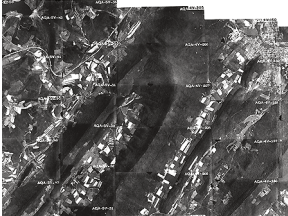
“

A large portion of the country is flown yearly in the NAIP program due to USDA compliance needs. One problem is that it is leaf on, therefore in areas of dense foliage,

“

some features are obscured. NAIP imagery is produced using partnership funds from USDA, USGS, FEMA, BLM, USFS and individual states. Other partnerships (between agencies or an agency and state) are also developed depending upon agency and local needs.

Aerial photography missions involve capturing sequences of overlapping images along many parallel flight paths. In the portion of the air photo mosaic shown below, note that the photographs overlap one another end to end, and side to side. This overlap is necessary for stereoscopic viewing, which is the key to rectifying photographs of variable terrain. It takes about 10 overlapping aerial photographs taken along two adjacent north-south flightpaths to provide stereo coverage for a 7.5-minute quadrangle.



Portion of a mosaic of overlapping vertical aerial photographs. (United States Department of Agriculture, Commodity Stabilization Service, n.d.).

TRY THIS!

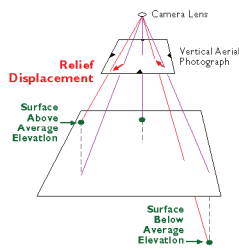
Use the USGS' [EarthExplorer](http://earthexplorer.usgs.gov/) (<http://earthexplorer.usgs.gov/>) to identify the vertical aerial photograph that shows the “populated place” in which you live. How old is the photo? (EarthExplorer is part of a USGS data distribution system.)

Note: The Digital Orthophoto backdrop that EarthExplorer allows you to view is not the same as the NAPP photos the system allows you to identify and order. By the end of this lesson, you should know the difference! If you don't, use the Chapter 6 Discussion Forum to ask.

6.12. Perspective and Planimetry

To understand why topographic maps can't be traced directly off of most vertical aerial photographs, you first need to appreciate the difference between perspective and planimetry. In a **perspective** view, all light rays reflected from the Earth's surface pass through a single point at the center of the camera lens. A **planimetric** (plan) view, by contrast, looks as though every position on the ground is being viewed from directly above. Scale varies in perspective views. In plan views, scale is everywhere consistent (if we overlook variations in small-scale maps due to map projections). Topographic maps are said to be **planimetrically correct**. So are orthoimages. Vertical aerial photographs are not, unless they happen to be taken over flat terrain.

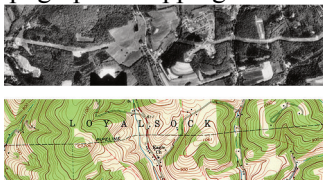
As discussed above, the scale of an aerial photograph is partly a function of flying height. Thus, variations in elevation cause variations in scale on aerial photographs. Specifically, the higher the elevation of an object, the farther the object will be displaced from its actual position away from the principal point of the photograph (the point on the ground surface that is directly below the camera lens). Conversely, the lower the elevation of an object, the more it will be displaced toward the **principal point**. This effect, called **relief displacement**, is illustrated in the diagram below. Note that the effect increases with distance from the principal point.



Relief displacement is scale variation on aerial photographs caused by variations in terrain elevation.

At the top of the diagram above, light rays reflected from the surface converge upon a single point at the center of the camera lens. The smaller trapezoid below the lens represents a sheet of photographic film. (The film actually is located behind the lens, but since the geometry of the incident light is symmetrical, we can minimize the height of the diagram by showing a mirror image of the film below the lens.) Notice the four triangular **fiducial marks** along the edges of the film. The marks point to the **principal point** of the photograph, which corresponds with the location on the ground directly below the camera lens at the moment of exposure. **Scale distortion is zero at the principal point.** Other features shown in the photo may be displaced toward or away from the principal point, depending on the elevation of the terrain surface. The larger trapezoid represents the average elevation of the terrain surface within a scene. On the left side of the diagram, a point on the land surface at a higher than average elevation is displaced outwards, away from the principal point and its actual location. On the right side, another location at less than average elevation is displaced towards the principal point. **As terrain elevation increases, flying height decreases and photo scale increases. As terrain elevation decreases, flying height increases and photo scale decreases.**

Compare the map and photograph below. Both show the same gas pipeline, which passes through hilly terrain. Note the deformation of the pipeline route in the photo relative to the shape of the route on the topographic map. **The deformation in the photo is caused by relief displacement.** The photo would not serve well on its own as a source for topographic mapping.



The pipeline clearing appears crooked in the photograph because of relief displacement.

Still confused? Think of it this way: where the terrain elevation is high, the ground is closer to the aerial camera, and the photo scale is a little larger than where the terrain elevation is lower. Although the altitude of the camera is constant, the effect of the undulating terrain is to zoom in and out. The effect of continuously-varying scale is to distort the geometry of the aerial photo. This effect is called relief displacement.

Distorted perspective views can be transformed into plan views through a process called **rectification**. In a Discussion Forum posting during the Summer 2001 offering of this class, student Joel Hamilton recounted one very awkward way to rectify aerial photographs:

“

“Back in the mid 80’s I saw a very large map being created from a multitude of aerial photos being fitted together. A problem that arose was that roads did not connect from one photo to the next at the outer edges of the map. No computers were used to create this map. So using a little water to wet the photos on the outside of the map, the photos were stretched to correct for the distortions. Starting from the center of the map the mosaic map was created. A very messy process.”

Nowadays, digital aerial photographs can be rectified in an analogous (but much less messy) way, using specialized photogrammetric software that shifts image pixels toward or away from the principal point of each photo in proportion to two variables: the elevation of the point of the Earth's surface at the location that corresponds to each pixel, and each pixel's distance from the principal point of the photo.

Another even simpler way to rectify perspective images is to view pairs of images **stereoscopically**.

6.13. Stereoscopy

If you have normal or corrected vision in both eyes, your view of the world is stereoscopic. Viewing your environment simultaneously from two slightly different perspectives enables you to estimate very accurately which objects in your visual field are nearer, and which are farther away. You know this ability as **depth perception**.

When you fix your gaze upon an object, the intersection of your two optical axes at the object form what is called a **parallaxic angle**. On average, people can detect changes as small as 3 seconds in the parallaxic angle, an angular resolution that compares well to transits and theodolites. The keenness of human depth perception is what makes photogrammetric measurements possible.

Your perception of a three-dimensional environment is produced from two separate two-dimensional images. The images produced by your eyes are analogous to two aerial images taken one after another along a flight path. Objects that appear in the area of overlap between two aerial images are seen from two different perspectives. A pair of overlapping vertical aerial images is called a **stereopair**. When a stereopair is viewed such that each eye sees only one image, it is possible to envision a three-dimensional image of the area of overlap.

In the following page you'll find a couple of examples of how stereoscopy is used to create planimetrically-correct views of the Earth's surface. If you have anaglyph stereo (red/blue) glasses, you'll be able to see stereo yourself. First, let's practice viewing anaglyph stereo images.

TRY THIS!

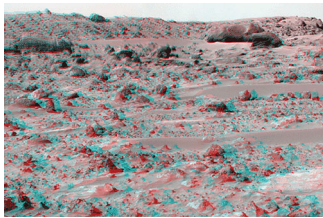
One way to see in stereo is with an instrument called a **stereoscope** (see examples at [James Madison University's Spatial Information Clearinghouse](#)). Another way that works on computer screens and doesn't require expensive equipment is called **anaglyph stereo** (anaglyph comes from a Greek word that means, "to carve in relief"). The anaglyph method involves special glasses in which the left and right eyes are covered by blue and red filters. CPGIS/MGIS registered through the World Campus received anaglyph glasses along with your welcome letters. Penn State students registered at University Park or other campuses should contact their instructor to determine if glasses are available.

The anaglyph image shown below consists of a superimposed stereopair in which the left image is shown in red, and the right image is shown in green and blue. The filters in the glasses ensure the each eye sees only one image. Can you make out the three-dimensional image of the U-shaped valley formed by glaciers in the French Alps?



Anaglyph stereopair by Pierre Gidon showing a scene in the French Alps (the image is used by permission of the author). Requires red/blue glasses.

How about this one: a panorama of the surface of Mars imaged during the Pathfinder mission, July 1997?



(NASA, 1997). Image processing and mosaic by Tim Parker.

To find other stereo images on the World Wide Web, search on “anaglyph.”

6.14. Rectification by Stereoscopy

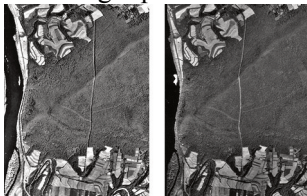
Aerial images need to be transformed from perspective views into plan views before they can be used to trace the features that appear on topographic maps, or to digitize vector features in digital data sets. One way to accomplish the transformation is through stereoscopic viewing.

Below are portions of a vertical aerial photograph and a topographic map that show the same area, a synclinal ridge called “Little Mountain” on the Susquehanna River in central Pennsylvania. A linear clearing, cut for a power line, appears on both (highlighted in yellow on the map). The clearing appears crooked on the photograph due to relief displacement. Yet we know that an aerial image like this one was used to compile the topographic map. **The air photo had to have been rectified to be used as a source for topographic mapping.**



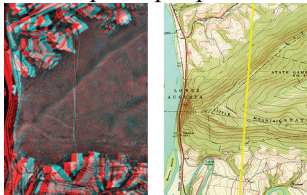
The deformation of the powerline clearing shown in the air photo is caused by relief displacement. (USGS. “Harrisburg East Quadrangle, Pennsylvania”)

Below are portions of two aerial photographs showing Little Mountain. The two photos were taken from successive flight paths. The two perspectives can be used to create a stereopair.



A stereopair: two air photos of the same area taken from different points of view.

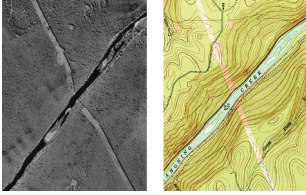
Next, the stereopair is superimposed in an anaglyph image. Using your red/blue glasses, you should be able to see a three-dimensional image of Little Mountain in which the power line appears straight, as it would if you were able to see it in person. Notice that the height of Little Mountain is exaggerated due to the fact that the distance between the principal points of the two photos is not exactly proportional to the distance between your eyes.



An anaglyph (red/blue) stereo image that fuses the stereopair shown in the above figure. When viewed with a red filter over the left eye and a cyan (blue) filter over the right eye, a stereoscopic image is formed. Notice that the

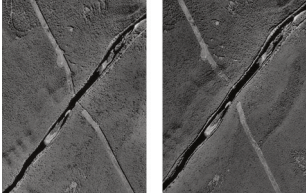
powerline clearing, which appears crooked in both air photos, appears straight in the stereoscopic image. (USGS. “Harrisburg East Quadrangle, Pennsylvania”)

Let’s try that again. We need to make sure that you can visualize how stereoscopic viewing transforms overlapping aerial photographs from perspective views into planimetric views. The aerial photograph and topographic map portions below show the same features, a power line clearing crossing the Sinnemahoning Creek in Central Pennsylvania. The power line appears to bend as it descends to the creek because of relief displacement.



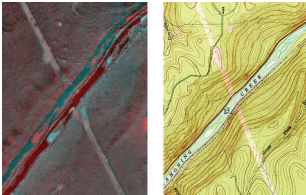
The deformation of the powerline clearing shown in the air photo is caused by relief displacement. (USGS. “Keating Quadrangle, Pennsylvania”).

Two aerial photographs of the same area taken from different perspectives constitute a stereo pair.



A stereopair, two air photos of the same area taken from different points of view.

By viewing the two photographs stereoscopically, we can transform them from two-dimensional perspective views to a single three-dimensional view in which the geometric distortions caused by relief displacement have been removed.



Deformation caused by relief displacement is rectified when the air photos are viewed in stereo. (USGS. “Keating Quadrangle, Pennsylvania”).

Photogrammetrists use instruments called **stereoplotters** to trace, or **compile**, the data shown on topographic maps from stereoscopic images like the ones you’ve seen here. The operator pictured below is viewing a stereoscopic model similar to the one you see when you view the anaglyph stereo images with red/blue glasses. A stereopair is superimposed on the right-hand screen of the operator’s workstation. The left-hand screen shows dialog boxes and command windows through which she controls the stereoplotter software. Instead of red/blue glasses, the operator is wearing glasses with polarized lens filters that allow her to visualize a three-dimensional image of the terrain. She handles a 3-D mouse that allows her to place a cursor on the terrain image within inches of its actual horizontal and vertical position.

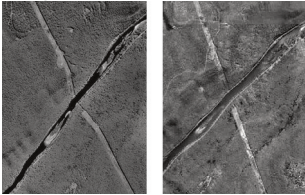


Merri MacKay (graduate of the Penn State Certificate Program in GIS, and employee of BAE Systems ADR), uses an analytic stereoplotter to digitize vertical and horizontal positions from a stereoscopic model. Photo circa

1998, used with permission of Ms. MacKay and ADR, Inc. When she encountered her picture as a student in the class in 2004, Merri wrote “I’ve got short hair and four grandkids now...”

6.15. Orthorectification

An **orthoimage** (or orthophoto) is a single aerial image in which distortions caused by relief displacement have been removed. The scale of an orthoimage is uniform. Like a planimetrically correct map, orthoimages depict scenes as though every point were viewed simultaneously from directly above. In other words, as if every optical axis were **orthogonal** to the ground surface. Notice how the power line clearing has been straightened in the orthophoto on the right below.



Comparison of a vertical aerial photograph (left) and an orthophoto.

Relief displacement is caused by differences in elevation. If the elevation of the terrain surface is known throughout a scene, the geometric distortion it causes can be rectified. Since photogrammetry can be used to measure vertical as well as horizontal positions, it can be used to create a collection of vertical positions called a **terrain model**. Automated procedures for transforming vertical aerial photos into orthophotos require **digital terrain models**.

Since the early 1990s, orthophotos have been commonly used as sources for editing and revising of digital vector data.

6.16. Metadata

Through the remainder of this Chapter and the next we’ll investigate the particular data products that comprise the framework themes of the U.S. National Spatial Data Infrastructure (NSDI). The format I’ll use to discuss these data products reflects the Federal Geographic Data Committee’s **Metadata standard** (FGDC, 1998c). Metadata is data about data. It is used to document the content, quality, format, ownership, and lineage of individual data sets. As the FGDC likes to point out, the most familiar example of metadata is the “Nutrition Facts” panel printed on food and drink labels in the U.S. Metadata also provides the keywords needed to search for available data in specialized clearinghouses and in the World Wide Web.

Some of the key headings included in the FGDC metadata standard include:

1. **Identification Information:** Who created the data, a brief description of its content, form, and purpose; its status, spatial extent, and use restrictions;
2. **Data Quality Information:** Accuracy and completeness of attributes, horizontal and vertical positions, sources, and procedures used to create the data;
3. **Spatial Reference Information:** Projection and/or coordinate system; datum and ellipsoid;
4. **Entity and Attribute Information:** Feature and attribute categories used; and
5. **Distribution Information:** Availability, and how to acquire the data.

FGDC’s Content Standard for Digital Geospatial Metadata is published [here](#). Geospatial professionals understand the value of metadata, know how to find it, and how to interpret it.

6.17. Digital Orthophoto Quadrangle (DOQ)

IDENTIFICATION

Digital Orthophoto Quads (DOQs) are raster images of rectified aerial photographs. They are widely used as sources for editing and revising vector topographic data. For example, the vector roads data maintained by businesses like NAVTEQ and Tele Atlas, as well as local and state government agencies, can be plotted over DOQs then edited to reflect changes shown in the orthoimage.

Most DOQs are produced by electronically scanning, then rectifying, black-and-white vertical aerial photographs. DOQ may also be produced from natural-color or near-infrared false-color photos, however, and from digital imagery. The variations in photo scale caused by relief displacement in the original images are removed by warping the image to compensate for the terrain elevations within the scene. Like USGS topographic maps, scale is uniform across each DOQ.

Most DOQs covers 3.75' of longitude by 3.75' of latitude. A set of four DOQs corresponds to each 7.5' quadrangle. (For this reason, DOQs are sometimes called DOQQs—Digital Orthophoto Quarter Quadrangles.) For its National Map, USGS has edge-matched DOQs into seamless data layers, by year of acquisition.



Portion of a USGS Digital Orthophoto Quad (DOQ) for Bushkill, PA.

DATA QUALITY

Like other USGS data products, DOQs conform to National Map Accuracy Standards. Since the **scale of the series is 1:12,000**, the standards warrant that 90 percent of well-defined points appear within 33.3 feet (10.1 meters) of their actual positions. One of the main sources of error is the rectification process, during which the image is warped such that each of a minimum of 3 control points matches its known location.

SPATIAL REFERENCE INFORMATION

All DOQs are cast on the Universal Transverse Mercator projection used in the local UTM zone. Horizontal positions are specified relative to the North American Datum of 1983, which is based on the GRS 80 ellipsoid.

ENTITIES AND ATTRIBUTES

The fundamental geometric element of a DOQ is the picture element (pixel). **Each pixel in a DOQ corresponds to one square meter on the ground.** Pixels in black-and-white DOQs are associated with a single attribute: a number from 0 to 255, where 0 stands for black, 255 stands for white, and the numbers in between represent levels of gray.

DOQs exceed the scanned topographic maps shown in Digital Raster Graphics (DRGs) in both pixel resolution and attribute resolution. DOQs are therefore much larger files than DRGs. Even though an individual DOQ file covers only one-quarter of the area of a topographic quadrangle (3.75 minutes square), it requires up to 55 Mb of digital storage. Because they cover only 25 percent of the area of topographic quadrangles, DOQs are also known as Digital Orthophoto Quarter Quadrangles (DOQQs).

DISTRIBUTION

USGS DOQ files are in the public domain, and can be used for any purpose without restriction. They are available for free download from the USGS, or from various state and regional data clearinghouses as well as from the [geoCOMMUNITY site](#). Digital orthoimagery data at 1-foot and 1-meter spatial resolution, collected from multiple sources, are available for user-specified areas from the [National Map Viewer site](#), and even higher resolution imagery (HRO) for certain areas is available through the [USGS Seamless Data Warehouse site](#).

To investigate DOQ data in greater depth, including links to a complete sample metadata document, visit [Birthplace of the DOQ](#). You're also welcome to post a comment to this page to describe your source of DOQ data, and how you use it. [FGDC's Content Standard for Digital Orthoimagery is published here](#).

TRY THIS!

EXPLORE DOQS WITH GLOBAL MAPPER (DLGV32 PRO)

Now it's time to use Global Mapper (dlgv32 Pro) again, this time to investigate the characteristics of a set of USGS Digital Orthophoto (Quarter) Quadrangles. The instructions below assume that you have already installed the Global Mapper / dlgv32 Pro software on your computer. (If you haven't, return to [installation instructions](#) presented earlier in Chapter 6).

Note: Global Mapper is a Windows application and will not run under the Macintosh operating system. The questions asked of Penn State students that involve the use of Global Mapper are not graded.

1. First **download one or more DOQ data archives**. Each compressed DOQ is over 37 Mb in size and will take about 8 minutes to download via high speed DSL or cable, or over two hours via 56 Kbps modem.
 - [DOQ_nw.zip](#)
 - [DOQ_ne.zip](#)
 - [DOQ_se.zip](#)
 - [DOQ_sw.zip](#)
2. Next decompress each archive into a directory on your hard disk.
 - Open an archive (e.g., "DOQ_nw.zip").
 - Create a subdirectory called "DOQ" within the directory you are using for class work.
 - Extract all files in the ZIP archive into your new subdirectory.

If you download and extract all four ZIP archives, the end result will be four DOQs that correspond with the Bushkill, PA quadrangle.

3. Launch Global Mapper (dlgv32 Pro).

4. Open a Digital Orthophoto (Quarter) Quadrangle by choosing File > Open Data File(s)..., then navigate to the subdirectory into which you extracted the DOQ data, then open the file 'bushkill_pa_nw.tif'.

5. The trial version of Global Mapper allows you to open and view up to four files at once. Note that you can turn layers on and off, and even adjust their transparency at Tools > Control Center. You might find it interesting to open and compare the DOQ and DRG layers.

6. Use the **Zoom and Pan tools** to magnify and scroll across the DOQ. The "Full View" button (house icon) refreshes the initial full view of the dataset.

7. To view an excerpt of the DOQ metadata, navigate to Tools > Control Center, then click the Metadata button.

TRY THIS!

ASSESS THE AVAILABILITY OF DIGITAL ORTHOIMAGERY VIA THE USGS NATIONAL MAP VIEWER

The National Map Viewer is an Internet Map Server application that provides a browsable map interface to the digital data layers that make up the National Map. The orthoimagery available through this interface has been gathered from several sources in addition to the USGS DOQ collection describe above.

1. On the [National Map home page](#), expand the **Products and Services** list and follow the link to **National Map Viewers** page. This page lists recommended browsers and lets you know that you need the Flash Player in order to use the interface. You are apt to have the Flash Player already installed. Go ahead and follow the **Click here to open viewer** link. This will open the viewer in a new browser tab or window. After the application loads, maximize your browser window. The **Help** link in the upper right of the interface gives you access to a wealth of information about the viewable data as well as user guides. Basic map navigation tools are found on a bar at the top of the map display area, along with a vertical scale change bar.
2. With the **Overlays** tab selected, on the left, and the **Content** sub-tab selected, check the box for the **Imagery** list and expand it. Check both boxes for **1_foot** and **1_meter_imagery_outlines** to see the areal extent of both categories depicted on the map. Take note of the distribution of the two resolutions. You might speculate on the reasons behind the coverage extents of the 1_foot imagery.
3. You can view the actual **1_meter_imagery**, too, by checking the box for it. Go ahead and investigate that.
4. Open the **Help** window, under the **Orthoimagery** entry, and read about the sources drawn from to create the bank of orthoimagery available.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 6 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about Photogrammetry. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

6.18. Summary

Many local, state and federal government agencies produce and rely upon geographic data to support their day-to-day operations. The **National Spatial Data Infrastructure (NSDI)** is meant to foster cooperation among agencies to reduce costs and increase the quality and availability of public data in the U.S. The key components of NSDI include standards, metadata, data, a clearinghouse for data dissemination, and partnerships. The seven **framework data themes** have been described as “the data backbone of the NSDI” (FGDC, 1997, p. v). This chapter and the next review the origins, characteristics and status of the framework themes. In comparison with some other developed

countries, framework data are fragmentary in the U.S., largely because mapping activities at various levels of government remain inadequately coordinated.

Chapter 6 considers two of the seven framework themes: geodetic control and orthoimagery. It discusses the impact of high-accuracy satellite positioning on accuracy standards for the National Spatial Reference System—the U.S.’ horizontal and vertical control networks. The chapter stresses the fact that **much framework data is derived, directly or indirectly, from aerial imagery**. Geospatial professionals understand how photogrammetrists compile planimetrically-correct vector data by stereoscopic analysis of aerial imagery. They also understand how orthoimages are produced and used to help keep vector data current, among other uses.

The most ambitious attempt to implement a nationwide collection of framework data is the USGS’ **National Map**. Composed of some of the digital data products described in this chapter and those that follow, the proposed National Map is to include high resolution (1 m) digital orthoimagery, variable resolution (10-30 m) digital elevation data, vector transportation, hydrography, and boundaries, medium resolution (30 m) land characterization data derived from satellite imagery, and geographic names. These data are to be seamless (unlike the more than 50,000 sheets that comprise the 7.5-minute topographic quadrangle series) and continuously updated. Meanwhile, in 2005, USGS announced that two of its three National Mapping Centers (in Reston, Virginia and Rolla, Missouri) would be closed, and over 300 jobs eliminated. Although funding for the Rolla center was subsequently restored by Congress, it remains to be seen whether USGS will be sufficiently resourced to fulfill its quest for a National Map.

QUIZ

Registered Penn State students should return now to the Chapter 6 folder in ANGEL (via the Resources menu to the left) to access the graded quiz for this chapter. This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter’s learning objectives. You are welcome to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 6.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

6.19. Bibliography

Anson, A. (2002) *Topographic mapping with plane table and alidade in the 1940s*. [CD-ROM] Professional Surveyors Publishing Co.

Doyle, David R. 1994 Development of the national spatial reference system. Retrieved 9 November 2007 from http://www.ngs.noaa.gov/PUBS_LIB/develop_NSRS.html

Federal Geodetic Control Committee (1988). Geometric geodetic accuracy standards and specifications for

using GPS relative positioning techniques. Retrieved March 27, 2013, from http://docs.lib.noaa.gov/noaa_documents/NOS/NGS/Geom_Geod_Accu_Standards.pdf

Federal Geographic Data Committee (1998a). Geospatial positing accuracy standards part 2: standards for geodetic networks. Retrieved February 11, 2008, from http://www.fgdc.gov/standards/standards_publications/

Federal Geographic Data Committee (1998b). Geospatial positing accuracy standards part 1: reporting methodology. Retrieved February 11, 2008, from http://www.fgdc.gov/standards/standards_publications/

Federal Geographic Data Committee (1998c). Content standard for digital geospatial metadata. Retrieved February 19, 2008, from http://www.fgdc.gov/standards/standards_publications/

Gidon, P. (2006). *Alpes_stereo*. Retrieved May 10, 2006, from http://perso.infonie.fr/alpes_stereo/i_index.htm (Expired link.)

Masser, I. (1998). *Governments and geographic information*. London: Taylor & Francis.

Moore, Larry (2000) The U.S. Geological Survey's revision program for 7.5-Minute topographic maps. Retrieved December 14, 2007 from <http://pubs.usgs.gov/of/2000/of00-325/moore.html>

National Aeronautic and Space Administration (1997). Mars pathfinder. Retrieved June 7, 2006, from <http://mars.jpl.nasa.gov/MPF/index0.html>

National Geodetic Survey (2007). The National Geodetic Survey 10 year plan; mission, vision and strategy 2007-2017. Retrieved February 19, 2008 from www.ngs.noaa.gov/INFO/ngs_tenyearplan.pdf

National Oceanic and Atmospheric Administration (2007) NOAA history. Retrieved February 18, 2008, from <http://www.history.noaa.gov/>

National Research Council (2002). *Research opportunities in geography at the U.S. Geological Survey*. Washington DC: National Academies Press.

National Research Council (2007). *A research agenda for geographic information science at the United States Geological Survey*. Washington DC: National Academies Press.

Office of Management and Budget (1990) Circular A-16, revised. Retrieved February 19, 2008, from http://www.whitehouse.gov/omb/circulars_a016_rev

Parry, R.B. (1987). The state of world mapping. In R. Parry & C. Perkins (Eds.), *World mapping today*. Butterworth-Heinemann.

Robinson, A. *et al.* (1995). *Elements of cartography* (5th ed.). New York: John Wiley & Sons.

Thompson, M. M. (1988). *Maps for America, cartographic products of the U.S. geological survey and others* (3d ed.). Reston, Va.: U.S. Geological Survey.

United States Geological Survey (2001). *The National Map: topographic mapping for the 21st century*. Final Report, November 30. Retrieved 11 January 2008 from http://nationalmap.gov/report/national_map_report_final.pdf

White House (1994) Executive order 12906: coordinating geographic data access. Retrieved February 19, 2008, from http://www.fgdc.gov/policyandplanning/executive_order

Chapter 7

National Spatial Data Infrastructure II

David DiBiase

7.1. Overview

Chapters 6 and 7 consider the origins and characteristics of the **framework data themes** that make up the United States' proposed National Spatial Data Infrastructure (NSDI). Chapter 6 discussed the geodetic control and orthoimagery themes. **This chapter describes the origins, characteristics and current status of the elevation, transportation, hydrography, governmental units and cadastral themes.**

Objectives

Students who successfully complete Chapter 7 should be able to:

1. Given a regular or irregular array of spot elevations, construct a triangulated irregular network, interpolate contour intervals and draw contour lines;
2. Compare vector and raster representations of terrain elevation;
3. Acquire and view digital elevation data from the National Elevation Dataset;
4. Calculate an interpolated spot elevation based on neighboring elevations;
5. Contrast the characteristics of three global elevation data products;
6. Describe the characteristics and current status of the NSDI hydrography, transportation, and governmental units themes as implemented in USGS' National Map; and
7. Interpret the size and relative location of a land parcel designated in terms of the U.S. Public Land Survey System.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

7.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated

quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 7 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 7	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	Submit 3 practice quizzes including: <ul style="list-style-type: none"> • Contouring • DLGs and DEMs • Interpolation Practice quizzes are not graded and may be submitted more than once.	Go to ANGEL > [your course section] > Lessons tab > Chapter 7 folder > [quiz]
3	Perform “Try this” activities including: <ul style="list-style-type: none"> • Draw a contour map • Explore Digital Line Graph hypsography • Explore a Digital Elevation Model • Download and view an extract from the National Elevation Dataset “Try this” activities are not graded.	Instructions are provided for each activity.
4	Submit the Chapter 7 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 7 folder > Chapter 7 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

7.3. Theme: Elevation

The NSDI *Framework Introduction and Guide* (FGDC, 1997, p. 19) points out that “elevation data are used in many different applications.” Civilian applications include flood plain delineation, road planning and construction, drainage, runoff, and soil loss calculations, and cell tower placement, among many others. Elevation data are also used to depict the terrain surface by a variety of means, from contours to relief shading and three-dimensional perspective views.

The NSDI Framework calls for an “elevation matrix” for land surfaces. That is, the terrain is to be represented

as a grid of elevation values. The spacing (or resolution) of the elevation grid may vary between areas of high and low relief (i.e., hilly and flat). Specifically, the Framework Introduction states that

Elevation values will be collected at a post-spacing of 2 arc-seconds (approximately 47.4 meters at 40° latitude) or finer. In areas of low relief, a spacing of 1/2 arc-second (approximately 11.8 meters at 40° latitude) or finer will be sought (FGDC, 1997, p. 18).

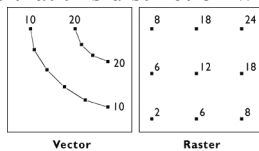
The elevation theme also includes bathymetry—depths below water surfaces—for coastal zones and inland water bodies. Specifically,

For depths, the framework consists of soundings and a gridded bottom model. Water depth is determined relative to a specific vertical reference surface, usually derived from tidal observations. In the future, this vertical reference may be based on a global model of the geoid or the ellipsoid, which is the reference for expressing height measurements in the Global Positioning System (Ibid).

USGS has lead responsibility for the elevation theme. Elevation is also a key component of USGS' National Map. The next several pages consider how heights and depths are created, how they are represented in digital geographic data, and how they may be depicted cartographically.

7.4. Vector and Raster Approaches

The terms **raster** and **vector** were introduced back in Chapter 1 to denote two fundamentally different strategies for representing geographic phenomena. Both strategies involve simplifying the infinite complexity of the Earth's surface. As it relates to elevation data, the raster approach involves *measuring elevation at a sample of locations*. The vector approach, on the other hand, involves *measuring the locations of a sample of elevations*. I hope that this distinction will be clear to you by the end of this chapter.

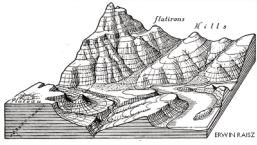


Vector and raster representations of the same terrain surface.

The illustration above compares how elevation data are represented in vector and raster data. On the left are elevation **contours**, a vector representation that is familiar with anyone who has used a USGS topographic map. The technical term for an elevation contour is *isarithm*, from the Greek words for “same” and “number.” The terms *isoline*, *isogram*, and *isopleth* all mean more or less the same thing. (See any cartography text for the distinctions.)

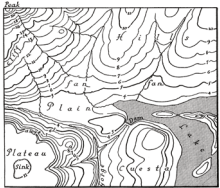
As you will see later in this chapter, when you explore Digital Line Graph hypsography data using Global Mapper or dlgv 32 Pro, elevations in vector data are encoded as attributes of line features. The distribution of elevation points across the quadrangle is therefore irregular. Raster elevation data, by contrast, consist of grids of points at which elevation is encoded at regular intervals. Raster elevation data are what's called for by the NSDI Framework and the USGS National Map. Digital contours can now be rendered easily from raster data. However, much of the raster elevation data used in the National Map was produced from digital vector contours and hydrography (streams and shorelines). For this reason we'll consider the vector approach to terrain representation first.

7.5. Contours



Contour lines trace the elevation of the terrain surface at regularly-spaced intervals (Raisz, 1948. © McGraw-Hill, Inc. Used by permission).

Drawing contour lines is a way to represent a terrain surface with a sample of elevations. Instead of measuring and depicting elevation at every point, you measure only along lines at which a series of imaginary horizontal planes slice through the terrain surface. The more imaginary planes, the more contours, and the more detail is captured.



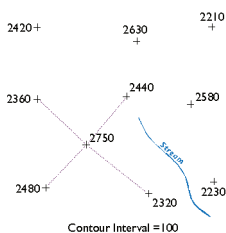
Contour lines representing the same terrain as in the first figure, but in plan view. (Raisz, 1948. © McGraw-Hill, Inc. Used by permission).

Until photogrammetric methods came of age in the 1950s, topographers in the field sketched contours on the USGS 15-minute topographic quadrangle series. Since then, contours shown on most of the 7.5-minute quads were compiled from stereoscopic images of the terrain, as described in Chapter 6. Today computer programs draw contours automatically from the spot elevations that photogrammetrists compile stereoscopically.

Although it is uncommon to draw terrain elevation contours by hand these days, it is still worthwhile to know how. In the next few pages you'll have a chance to practice the technique, which is analogous to the way computers do it.

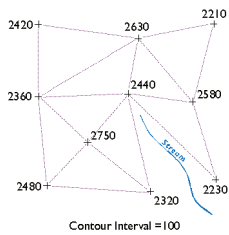
7.6. Contouring By Hand

This page will walk you through a methodical approach to rendering contour lines from an array of spot elevations (Rabenhorst and McDermott, 1989). To get the most from this demonstration, I suggest that you print the [illustration in the attached image file](#). Find a pencil (preferably one with an eraser!) and straightedge, and duplicate the steps illustrated below. A “Try This!” activity will follow this step-by-step introduction, providing you a chance to go solo.



Beginning a triangulated irregular network.

Starting at the highest elevation, draw straight lines to the nearest neighboring spot elevations. Once you have connected to all of the points that neighbor the highest point, begin again at the second highest elevation. (You will have to make some subjective decisions as to which points are “neighbors” and which are not.) Taking care not to draw triangles across the stream, continue until the surface is completely triangulated.



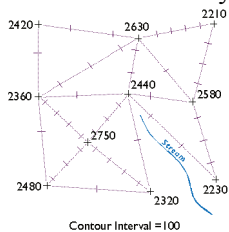
Complete TIN. Note that the triangle sides must not cross hydrologic features (i.e., the stream) on a terrain surface.

The result is a **triangulated irregular network (TIN)**. A TIN is a vector representation of a continuous surface that consists entirely of triangular facets. The vertices of the triangles are spot elevations that may have been measured in the field by leveling, or in a photogrammetrist's workshop with a stereoplotter, or by other means. (Spot elevations produced photogrammetrically are called **mass points**.) A useful characteristic of TINs is that each triangular facet has a single slope degree and direction. With a little imagination and practice, you can visualize the underlying surface from the TIN even without drawing contours.

Wonder why I suggest that you not let triangle sides that make up the TIN cross the stream? Well, if you did, the stream would appear to run along the side of a hill, instead of down a valley as it should. In practice, spot elevations would always be measured at several points along the stream, and along ridges as well. Photogrammetrists refer to spot elevations collected along linear features as **breaklines** (Maune, 2007). I omitted breaklines from this example just to make a point.

You may notice that there is more than one correct way to draw the TIN. As you will see, deciding which spot elevations are “near neighbors” and which are not is subjective in some cases. Related to this element of subjectivity is the fact that the fidelity of a contour map depends in large part on the distribution of spot elevations on which it is based. In general, the density of spot elevations should be greater where terrain elevations vary greatly, and sparser where the terrain varies subtly. Similarly, the smaller the contour interval you intend to use, the more spot elevations you need.

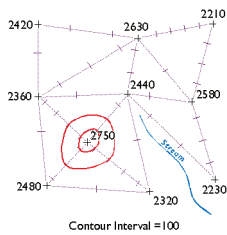
(There are algorithms for triangulating irregular arrays that produce unique solutions. One approach is called **Delaunay Triangulation** which, in one of its constrained forms, is useful for representing terrain surfaces. The distinguishing geometric characteristic of a Delaunay triangulation is that a circle surrounding each triangle side does not contain any other vertex.)



Tick marks drawn where elevation contours cross the edges of each TIN facet.

Now draw ticks to mark the points at which elevation contours intersect each triangle side. For instance, see the triangle side that connects the spot elevations 2360 and 2480 in the lower left corner of the illustration above? One tick mark is drawn on the triangle where a contour representing elevation 2400 intersects. Now find the two spot elevations, 2480 and 2750, in the same lower left corner. Note that three tick marks are placed where contours representing elevations 2500, 2600, and 2700 intersect.

This step should remind you of the equal interval classification scheme you read about in Chapter 3. The right choice of contour interval depends on the goal of the mapping project. In general, contour intervals increase in proportion to the variability of the terrain surface. It should be noted that the assumption that elevations increase or decrease at a constant rate is not always correct, of course. We will consider that issue in more detail later.



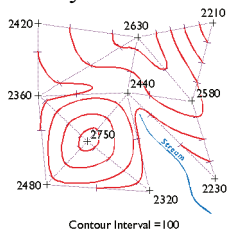
Threading elevation contours through a TIN.

Finally, draw your contour lines. Working downslope from the highest elevation, thread contours through ticks of equal value. Move to the next highest elevation when the surface seems ambiguous.

Keep in mind the following characteristics of contour lines (Rabenhorst and McDermott, 1989):

- Contours should **always point upstream** in valleys
- Contours should **always point downridge** along ridges
- Adjacent contours should **always be sequential or equivalent**
- Contours should **never split into two**
- Contours should **never cross or loop**
- Contours should **never spiral**
- Contours should **never stop in the middle of a map**

How does your finished map compare with the one I drew below?



TRY THIS!

Now try your hand at **contouring on your own**. The purpose of this practice activity is to give you more experience in contouring terrain surfaces.

1. First, view [an image of an irregular array of 16 spot elevations](#).
2. Print the image.
3. Use the procedure outlined in this lesson to draw contour lines that represent the terrain surface that the spot elevations were sampled from. You may find this to be a moderately challenging task that takes about a half hour to do well. TIP: label the tick marks to make it easier to connect them.
4. When finished, [compare your result to an existing map](#).

Here are a couple of somewhat simpler problems and solutions in case you need a little more practice.

- [Practice Problem #1](#)
- [Practice Problem #1 Solution](#)
- [Practice Problem #2](#)
- [Practice Problem #2 Solution](#)

You will be asked to demonstrate your contouring ability again in the Lesson 7 Quiz and in the final exam.

Kevin Sabo (personal communication, Winter 2002) remarked that “If you were unfortunate enough to be hand-contouring data in the 1960’s and 70’s, you may at least have had the aid of a [Gerber Variable Scale](#). After hand contouring in Lesson 7, I sure wished I had my Gerber!”

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 7 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **Contouring**. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

7.7. Digital Line Graph (DLG)

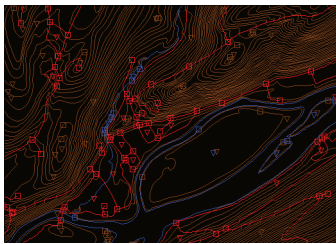
IDENTIFICATION

Digital Line Graphs (DLGs) are vector representations of most of the features and attributes shown on USGS topographic maps. Individual feature sets (outlined in the table below) are encoded in separate digital files. DLGs exist at three scales: small (1:2,000,000), intermediate (1:100,000) and large (1:24,000). Large-scale DLGs are produced in **intiles** that correspond to the 7.5-minute topographic quadrangles from which they were derived.

Description of Digital Line Graph Layers

Layer	Features
Public Land Survey System (PLSS)	Township, range, and section lines
Boundaries	State, county, city, and other national and State lands such as forests and parks
Transportation	Roads and trails, railroads, pipelines and transmission lines
Hydrography	Flowing water, standing water, and wetlands
Hypsography	Contours and supplementary spot elevations
Non-vegetative features	Glacial moraine, lava, sand, and gravel
Survey control and markers	Horizontal and vertical monuments (third order or better)
Man-made features	Cultural features, such as building, not collected in other data categories
Woods, scrub, orchards, and vineyards	Vegetative surface cover

Layers and contents of large-scale Digital Line Graph files. Not all layers available for all quadrangles (USGS, 2006).



Portion of three Digital Line Graph (DLG) layers for USGS Bushkill, PA quadrangle; imaged with Global Mapper (dlgv32 Pro) software. Transportation features are arbitrarily colored red, hydrography blue, and hydrography brown. The square symbols are nodes and the triangles represent polygon centroids.

DATA QUALITY

Like other USGS data products, DLGs conform to National Map Accuracy Standards. In addition, however, DLGs are tested for the logical consistency of the topological relationships among data elements. Similar to the Census Bureau's TIGER/Line, line segments in DLGs must begin and end at point features (nodes), and line segments must be bounded on both sides by area features (polygons).

SPATIAL REFERENCE INFORMATION

DLGs are heterogenous. Some use UTM coordinates, others State Plane Coordinates. Some are based on NAD 27, others on NAD 83. Elevations are referenced either to NGVD 29 or NAVD 88 (USGS, 2006a).

ENTITIES AND ATTRIBUTES

The basic elements of DLG files are nodes (positions), line segments that connect two nodes, and areas formed by three or more line segments. Each node, line segment, and area is associated with two-part integer attribute codes. For example, a line segment associated with the attribute code "050 0412" represents a hydrographic feature (050), specifically, a stream (0412).

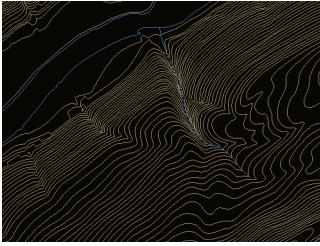
DISTRIBUTION

Not all DLG layers are available for all areas at all three scales. Coverage is complete at 1:2,000,000. At the intermediate scale, 1:100,000 (30 minutes by 60 minutes), all hydrography and transportation files are available for the entire U.S., and complete national coverage is planned. At 1:24,000 (7.5 minutes by 7.5 minutes), coverage remains spotty. The files are in the public domain, and can be used for any purpose without restriction.

Large- and Intermediate -scale DLGs are available for download through [EarthExplorer system](#). You can plot 1:2,000,000 DLGs on-line at the [USGS' National Atlas of the United States](#).

DIGITAL LINE GRAPH HYPSOGRAPHY

In one sense, DLGs are as much "legacy" data as the out-of-date topographic maps from which they were produced. Still, DLG data serve as primary or secondary sources for several themes in the USGS National Map, including hydrography, boundaries, and transportation. DLG hypsography data are not included in the National Map, however. It is assumed that GIS users can generate elevation contours as needed from DEMs. DLG hypsography and hydrography layers are the preferred sources from which USGS DEMs are produced, however.

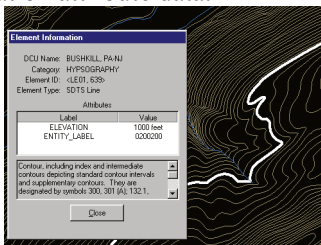


Portion of the hypsography and hydrography layers of a large-scale Digital Line Graph (DLG). USGS Bushkill, PA quadrangle; imaged with Global Mapper (dlgv32 Pro) software.

Hypsography refers to the measurement and depiction of the terrain surface, specifically with contour lines. Several different methods have been used to produce DLG hypsography layers, including:

- Scanning contour lines on photographic film or paper maps, converting the scanned raster data to vectors, then editing and attributing the vector features;
- Manually digitizing and attributing contour lines on photographic film or paper maps; and
- Producing contours by photogrammetric processes.

The preferred method is to manually digitize contour lines in vector mode, then to key-enter the corresponding elevation attribute data.



The highlighted contour line has been selected, and its attributes reported in a Global Mapper window. Notice that the line feature is attributed with a unique Element ID code (LE01, 639) and an elevation (1000 feet).

TRY THIS!

EXPLORING DLGS WITH GLOBAL MAPPER (DLGV32 PRO)

Now I'd like you to use Global Mapper (or dlgv32 Pro) software to investigate the characteristics of the hypsography layer of a USGS Digital Line Graph (DLG). The instructions below assume that you have already installed software on your computer. (If you haven't, return to [installation instructions](#) presented earlier in Chapter 6). First you'll download and a sample DLG file. In a following activity you'll have a chance to find and download DLG data for your area.

1. If you haven't done so already, create a directory called "USGS Data" on your hard disk, where you file your course materials.
2. Next, [Download the DLG.zip data archive](#). The ZIP archive is 1.2 Mb in size and will take approximately 15 seconds to download via high speed DSL or cable, or about 4 minutes and 15 seconds minutes via 56 Kbps modem.
3. Now decompress the archive into a directory on your hard disk.

- Open the archive DLG.zip.
- Create a subdirectory called “DLG” within the directory in which you save data for this class.
- Extract all files in the ZIP archive into your new subdirectory.

The end result will be five subdirectories, each of which includes the data files that make up a DLG “layer,” along with a master directory.

4. Launch Global Mapper or dlgv32 Pro.

5. Open a Digital Line Graph by choosing File > Open as New..., then navigate to the directory “DLG/Hypso.” Open the file ‘Hp01catd.ddf’ (you can open up to four files at once in the trial version of Global Mapper.) The data correspond with the 7.5 minute quadrangle for Bushkill, PA. The file is encoded in Spatial Data Transfer Standard (SDTS) format. For information about SDTS, see the [SDTS Tutorial](#)(PDF format).

6. Global Mapper may ask you to direct it to a ‘Master Data Dictionary’ file. If so, navigate to, and select, the file ‘Dlg/MasterDlg/Dlg3mdir.ddf’

7. Experiment with Global Mapper’s tools. Use Zoom and Pan to magnify and scroll across the DLG. The **Full View button** (the one with the house icon) refreshes the initial full view of the data set.

8. The **Feature Info tool** allows you to query the attributes of a particular feature. Try clicking a single line segment. Note that you can display the attributes of a feature in the lower left portion of the application window by simply hovering over the feature.

9. The **Measure tool** (ruler icon) allows you to not only measure distance as the crow flies, but also to see the area enclosed by a series of line segments drawn by repeated mouse clicks. Note again the location information that is given to you near the bottom of the application window.

10. Certain tools, e.g., the **3D Path Profile/Line of Sight tool** (next to the Feature Info tool) are not functional in the free (unregistered) version of Global Mapper.

11. The trial version of Global Mapper allows you to open and view up to four files at once. You might find it interesting to open and compare the Bushkill DLG hypsography file and the corresponding DRG you viewed in Lesson 6. Note that you can turn layers on and off, and even adjust their transparency at Tools > Control Center. **How do the contours in the DLG compare with those in the DRG? What explains the difference?**

12. Global Mapper provides the metadata you’ll need to answer questions in a practice quiz. **To access the metadata, navigate to Tools > Control Center, then click the Metadata button.**

7.8. Digital Elevation Model (DEM)

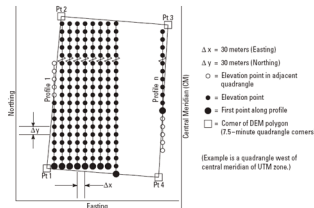
The term “Digital Elevation Model” has both generic and specific meanings. In general, a DEM is any raster representation of a terrain surface. Specifically, a DEM is a data product of the U.S. Geological Survey. Here we consider the characteristics of DEMs produced by the USGS. Later in this chapter we’ll consider sources of global terrain data.

IDENTIFICATION

USGS DEMs are raster grids of elevation values that are arrayed in series of south-north profiles. Like other USGS data, DEMs were produced originally in tiles that correspond to topographic quadrangles. Large scale (7.5-minute

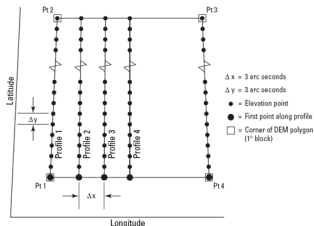
and 15-minute), intermediate scale (30 minute), and small scale (1 degree) series were produced for the entire U.S. The resolution of a DEM is a function of the east-west spacing of the profiles and the south-north spacing of elevation points within each profile.

DEMs corresponding to 7.5-minute quadrangles are available at 10-meter resolution for much, but not all, of the U.S. Coverage is complete at 30-meter resolution. In these large scale DEMs elevation profiles are aligned parallel to the central meridian of the local UTM zone, as shown in the illustration below. **See how the DEM tile in the illustration below appears to be tilted? This is because the corner points are defined in unprojected geographic coordinates that correspond to the corner points of a USGS quadrangle.** The farther the quadrangle is from the central meridian of the UTM zone, the more it is tilted.



Arrangement of elevation profiles in a large scale USGS Digital Elevation Model (USGS, 1987).

As shown below, the arrangement of the elevation profiles is different in intermediate- and small-scale DEMs. Like meridians in the northern hemisphere, the profiles in 30-minute and 1-degree DEMs converge toward the north pole. For this reason the resolution of intermediate- and small-scale DEMs (that is to say, the spacing of the elevation values) is expressed differently than for large-scale DEMs. **The resolution of 30-minute DEMs is said to be 2 arc seconds and 1-degree DEMs are 3 arc seconds.** Since an arc second is 1/3600 of a degree, elevation values in a 3 arc second DEM are spaced 1/1200 degree apart, representing a grid cell about 66 meters “wide” by 93 meters “tall” at 45° latitude.



Arrangement of elevation profiles in a small scale USGS Digital Elevation Model (USGS, 1987).

The preferred method for producing the elevation values that populate DEM profiles is interpolation from DLG hypsography and hydrography layers (including the hydrography layer enables analysts to delineate valleys with less uncertainty than hypsography alone). Some older DEMs were produced from elevation contours digitized from paper maps or during photogrammetric processing, then smoothed to filter out errors. Others were produced photogrammetrically from aerial photographs.

DATA QUALITY

The vertical accuracy of DEMs is expressed as the root mean square error (RMSE) of a sample of at least 28 elevation points. The target accuracy for large-scale DEMs is seven meters; 15 meters is the maximum error allowed.

SPATIAL REFERENCE INFORMATION

Like DLGs, USGS DEMs are heterogenous. They are cast on the Universal Transverse Mercator projection used in the local UTM zone. Some DEMs are based upon the North American Datum of 1983, others on NAD 27. Elevations in some DEMs are referenced to either NGVD 29 or NAVD 88.

ENTITIES AND ATTRIBUTES

Each record in a DEM is a profile of elevation points. Records include the UTM coordinates of the starting point, the number of elevation points that follow in the profile, and the elevation values that make up the profile. Other than the starting point, the positions of the other elevation points need not be encoded, since their spacing is defined. (Later in this lesson you'll download a sample USGS DEM file. Try opening it in a text editor to see what I'm talking about.)

DISTRIBUTION

DEM tiles are available for free download through many state and regional clearinghouses. You can find these sources by searching the [GEODATA portion of the Data.Gov site](#), formerly the separate Geospatial One Stop site.

As part of its National Map initiative, the USGS has developed a “seamless” [National Elevation Dataset](#) that is derived from DEMs, among other sources. NED data are available at three resolutions: 1 arc second (approximately 30 meters), 1/3 arc second (approximately 10 meters), and 1/9 arc second (approximately 3 meters). Coverage ranges from complete at 1 arc second to extremely sparse at 1/9 arc second. An extensive [FAQ on NED data is published here](#). The second of the two following activities involves downloading NED data and viewing it in Global Mapper.

TRY THIS!

EXPLORING DEMS WITH GLOBAL MAPPER (DLGV32 PRO)

Global Mapper time again! This time you'll investigate the characteristics of a USGS DEM. The instructions below assume that you have already installed the software on your computer. (If you haven't, return to [installation instructions](#) presented earlier in Chapter 6). The instructions will remind you how to open a DEM in dlgv32 Pro. In the practice quiz that follows you'll be asked questions require you to explore the data for answers.

1. First [Download the DEM.zip data archive](#). The ZIP archive is 2.5 Mb in size and will take about 30 seconds to download via high speed DSL or cable, or nearly 9 minutes via 56 Kbps modem. If you can't download the file, contact my teaching assistant or me right away so we can help you resolve the problem.
2. Now decompress the archive into a directory on your hard disk.
 - Open the archive **DEM.zip**.
 - Create a subdirectory called “DEM” within the directory in which you save class data.
 - Extract all files in the ZIP archive into your new subdirectory.

The end result will be two subdirectories, one of which contains a 30-meter DEM, the other a 10-meter DEM. These datasets are in the earlier distribution format of USGS DEM data — elevation data in

horizontal (pixel) units of meters and representative of the area covered by a 1:24,000 topo map sheet. In the Try This that follows this one you will see that the distribution format options have expanded.

3. Launch Global Mapper.

4. Open a Digital Elevation Model by choosing **File > Open Data File(s)...**, then navigate to the directory DEM_30m or DEM_10m, then open the file *bushkill_pa.dem*

5. Use the **Zoom** and **Pan** tools to magnify and scroll across the DEM. The **Full View** button (house icon) refreshes the initial full view of the dataset.

6. Global Mapper provides access to the metadata you'll need to answer questions in a practice quiz. To access the metadata, navigate to **Tools > Control Center**, then click the **Metadata** button.

You can change the appearance of the DEM in the Options section of the Control Center. You can also alter the appearance of the DEM by choosing Tools > Configure, and changing the settings in, especially, Vertical Options and Shader Options. To see the DEM data with(out) hill shading, find the Enable/Disable Hill Shading button on the Shader toolbar (it has a sunburst in the lower left corner).

TRY THIS!

DOWNLOAD YOUR OWN NATIONAL ELEVATION DATASET (NED) DATA

1. Go to the [Elevation page of the USGS National Map site](#).

Read any of the information there that you choose.

2. Follow the link to [The National Map Viewer](#).

If you wish, also follow the link to the detailed instructions that you see under The National Map Viewer link. The following instructions in this Try This should also suffice, and perhaps elaborate at bit more. You will not see Elevation data listed in the left hand Overlays pane. The USGS is in the process of adding more visualization options when it comes to elevation data. See the Hill Shade button at the upper right of the map area.

3. Use the GIS tools, found above the map area, to pan to and zoom in on an area of interest. Then, click the **Download Data** button.

Choose a *reference area* from the pick list. The default is the index of the areas covered by the 1:24,000 topo map series.

You could also chose the entire *current map extent*, but depending upon your zoom level that may be a huge dataset.

The instructions found via the Help link mentioned above mention that you can define an area based on creating a *custom polygon*, but at this point I do not see how that is done...

4. Then click on the map to highlight a specific area of interest. A link to the available data sets will appear in the left hand pane under the **Selection** tab. (The **All Results** button will list the multiple areas you click on.) Follow the **Download** link of the area you are interested in.

5. In the **USGS Available Data** window that opens, select **Elevation** from the **Theme** column, and choose a file format from the pick list in the **Format** column.

I know that both the GeoTIFF and ArcGRID formats are compatible with Global Mapper. (ArcGRID is the Esri company's raster format. You'll be using the Esri software in future courses.)

Click the **Next** button.

6. You will be given a list of Elevation *Product* choices.

The choices that show *Dynamic* as the Type will be those that match the Format choice you made in the previous window. The other *Staged* datasets are prepackaged and in the formats stated in the Product column, and apparently listed regardless of the Format choice you made. If multiple resolutions are available they will be listed.

From the Product list, check the box for what you wish to download.

Click the **Next** button.

Go to the **Cart** pane on the right. (It may open automatically.)

Go to **Checkout**, supply your e-mail address, and submit your order via the **Place Order** button.

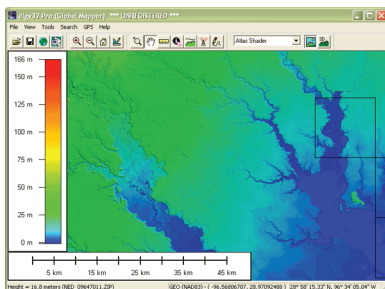
You will receive a message telling you that your order has been placed, and that will soon be followed by an e-mail regarding your order. About an hour and a half after I submitted my request I received a second e-mail containing a download link.

7. The system will produce a ZIP archive that you can save to your hard disk (e.g., “09647011.zip”)

8. Launch Global Mapper and open the ZIP archive. The software can read the data even in its compressed form; you should not need to extract the contents from the .zip file. (It would be a good idea to look at the contents of the .zip archive, though, if only to see the number and type of files included.)

9. An image of the DEM data should appear in the Global Mapper window, similar to what you see shown below (even though the image below is from an older version of Global Mapper).

10. Again, you can view the *metadata* associated with the DEM data via the **Tools > Control Center** menu. Note the PIXEL dimensions reported in *arc degrees*, as opposed to something like meters.

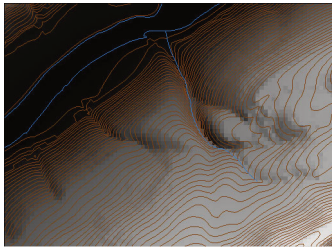


PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 7 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **DLGs and DEMs**. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

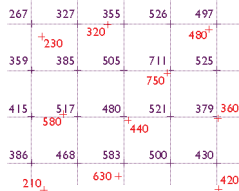
7.9. Interpolation

DEMs are produced by various methods. The method preferred by USGS is to **interpolate** elevations grids from the hypsography and hydrography layers of Digital Line Graphs.



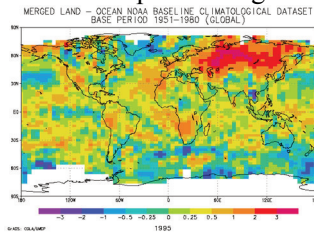
A USGS 7.5-minute DEM and the DLG hypsography and hydrography layers from which it was produced.

The elevation points in DLG hypsography files are not regularly spaced. DEMs need to be regularly spaced to support the slope, gradient, and volume calculations they are often used for. Grid point elevations must be interpolated from neighboring elevation points. In the figure below, for example, the gridded elevations shown in purple were interpolated from the irregularly spaced spot elevations shown in red.



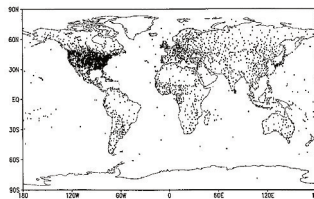
Elevation values in DEMs are interpolated from irregular arrays of elevations measured through photogrammetric methods, or derived from existing DLG hypsography and hydrography data.

Here’s another example of interpolation for mapping. The map below shows how 1995 average surface air temperature differed from the average temperature over a 30-year baseline period (1951-1980). The temperature anomalies are depicted for grid cells that cover 3° longitude by 2.5° latitude.



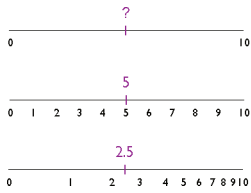
1995 Surface Temperature Anomalies. (National Climatic Data Center, 2005).

The gridded data shown above were estimated from the temperature records associated with the very irregular array of 3,467 locations pinpointed in the map below. In general, **interpolation is the process of estimating an unknown value from neighboring known values.**



The Global Historical Climate Network. (Eischeid et al., 1995).

Elevation data are often not measured at evenly-spaced locations. Photogrammetrists typically take more measurements where the terrain varies the most. They refer to the dense clusters of measurements they take as “mass points.” Topographic maps (and their derivatives, DLGs) are another rich source of elevation data. Elevations can be measured from contour lines, but obviously contours do not form evenly-spaced grids. Both methods give rise to the need for interpolation.

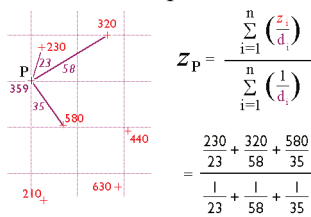


Interpolating an intermediate value on a number line.

The illustration above shows three number lines, each of which ranges in value from 0 to 10. If you were asked to interpolate the value of the tick mark labeled “?” on the top number line, what would you guess? An estimate of “5” is reasonable, provided that the values between 0 and 10 increase at a constant rate. If the values increase at a geometric rate, the actual value of “?” could be quite different, as illustrated in the bottom number line. The validity of an interpolated value depends, therefore, on the validity of our assumptions about the nature of the underlying surface.

As I mentioned in Chapter 1, the surface of the Earth is characterized by a property called **spatial dependence**. Nearby locations are more likely to have similar elevations than are distant locations. Spatial dependence allows us to assume that it’s valid to estimate elevation values by interpolation.

Many interpolation algorithms have been developed. One of the simplest and most widely used (although often not the best) is the **inverse distance weighted** algorithm. Thanks to the property of spatial dependence, we can assume that estimated elevations are more similar to nearby elevations than to distant elevations. The inverse distance weighted algorithm estimates the value z of a point P as a function of the z -values of the nearest n points. The more distant a point, the less it influences the estimate.



The inverse distance weighted interpolation procedure.

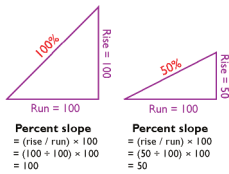
PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 7 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **Interpolation**. You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

7.10. Slope

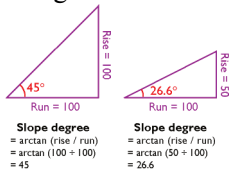
Slope is a measure of change in elevation. It is a crucial parameter in several well-known predictive models used for environmental management, including the Universal Soil Loss Equation and agricultural non-point source pollution models.

One way to express slope is as a percentage. To calculate percent slope, divide the difference between the elevations of two points by the distance between them, then multiply the quotient by 100. The difference in elevation between points is called the rise. The distance between the points is called the run. Thus, percent slope equals (rise / run) x 100.



Calculating percent slope. A rise of 100 feet over a run of 100 feet yields a 100 percent slope. A 50-foot rise over a 100-foot run yields a 50 percent slope.

Another way to express slope is as a slope angle, or degree of slope. As shown below, if you visualize rise and run as sides of a right triangle, then the degree of slope is the angle opposite the rise. Since degree of slope is equal to the tangent of the fraction rise/run, it can be calculated as the arctangent of rise/run.



A rise of 100 feet over a run of 100 feet yields a 45° slope angle. A rise of 50 feet over a run of 100 feet yields a 26.6° slope angle.

You can calculate slope on a contour map by analyzing the spacing of the contours. If you have many slope values to calculate, however, you will want to automate the process. It turns out that slope calculations are much easier to calculate for gridded elevation data than for vector data, since elevations are more or less equally spaced in raster grids.

Several algorithms have been developed to calculate percent slope and degree of slope. The simplest and most common is called the **neighborhood method**. The neighborhood method calculates the slope at one grid point by comparing the elevations of the eight grid points that surround it.



The neighborhood algorithm estimates percent slope in cell 5 by comparing the elevations of neighboring grid cells.

The neighborhood algorithm estimates percent slope at grid cell 5 (Z₅) as the sum of the absolute values of east-west slope and north-south slope, and multiplying the sum by 100. The diagram below illustrates how east-west slope and north-south slope are calculated. Essentially, east-west slope is estimated as the difference between the sums of the elevations in the first and third columns of the 3 x 3 matrix. Similarly, north-south slope is the difference between the sums of elevations in the first and third rows (note that in each case the middle value is weighted by a factor of two).

$$S_5 = (\sqrt{S_{e-w}^2 + S_{n-s}^2}) \times 100$$

$$S_{e-w} = \frac{\begin{matrix} Z_1 & Z_2 & Z_3 \\ 2Z_2 & - & 2Z_3 \\ Z_4 & Z_5 & Z_6 \end{matrix}}{8 \times \text{Grid Spacing}} \quad S_{n-s} = \frac{\begin{matrix} Z_1 & 2Z_2 & Z_3 \\ - & - & - \\ Z_4 & 2Z_5 & Z_6 \end{matrix}}{8 \times \text{Grid Spacing}}$$

The neighborhood algorithm for calculating percent slope.

The neighborhood algorithm calculates slope for every cell in an elevation grid by analyzing each 3 x 3 neighborhood. Percent slope can be converted to slope degree later. The result is a grid of slope values suitable for use in various soil loss and hydrologic models.

7.11. Relief Shading

You can see individual pixels in the zoomed image of a 7.5-minute DEM below. I used dlgv32 Pro's "Gradient Shader" to produce the image. Each pixel represents one elevation point. The pixels are shaded through 256 levels of gray. Dark pixels represent low elevations, light pixels represent high ones.



A digital elevation model in which light pixels represent high elevations, and dark pixels represent low elevations.

It's also possible to assign gray values to pixels in ways that make it appear that the DEM is illuminated from above. The image below, which shows the same portion of the Bushkill DEM as the image above, illustrates the effect, which is called **terrain shading**, **hill shading** or **shaded relief**.



Shaded terrain image produced from the same DEM as shown in the above figures, using dlgv32 Pro's Daylight Shader option, with the Surface Color set to gray.

The appearance of a shaded terrain image depends on several parameters, including **vertical exaggeration**. Click the buttons under the image below to compare the four terrain images of North America shown below, in which elevations are exaggerated 5 times, 10 times, 20 times, and 40 times respectively. (You will need to have the Adobe Flash player installed in order to complete this exercise. If you do not already have the Flash player, you can [download it for free from Adobe](#).)

Effects of vertical exaggeration on a shaded terrain image

Another influential parameter is the **angle of illumination**. Click the buttons to compare terrain images that have been illuminated from the northeast, southeast, southwest, and northwest. Does the terrain appear to be inverted in one or more of the images? **To minimize the possibility of terrain inversion, it is conventional to illuminate terrain from the northwest.**

Effects of illumination angle on a shaded terrain image.

7.12. Lidar

For many applications, 30-meter DEMs whose vertical accuracy is measured in meters are simply not detailed enough. Greater accuracy and higher horizontal resolution can be produced by photogrammetric methods, but precise photogrammetry is often too time-consuming and expensive for extensive areas. Lidar is a digital remote sensing technique that provides an attractive alternative.

Lidar stands for **L**ight **D**etection **A**nd **R**anging. Like radar (RADIO Detecting And Ranging), lidar instruments transmit and receive energy pulses, and enable distance measurement by keeping track of the time elapsed between

transmission and reception. Instead of radio waves, however, lidar instruments emit laser light (laser stands for Light Amplifications by Stimulated Emission of Radiation).

Lidar instruments are typically mounted in low altitude aircraft. They emit up to 5,000 laser pulses per second, across a ground swath some 600 meters wide (about 2,000 feet). The ground surface, vegetation canopy, or other obstacles reflect the pulses, and the instrument's receiver detects some of the backscatter. Lidar mapping missions rely upon GPS to record the position of the aircraft, and upon inertial navigation instruments (gyroscopes that detect an aircraft's pitch, yaw, and roll) to keep track of the system's orientation relative to the ground surface.

In ideal conditions, lidar can produce DEMs with 15-centimeter vertical accuracy, and horizontal resolution of a few meters. Its cost is prohibitive for small missions, but is justified for larger projects in which detail is essential. For example, lidar has been used successfully to detect subtle changes in the thickness of the Greenland ice sheet that result in a net loss of over 50 cubic kilometers of ice annually.

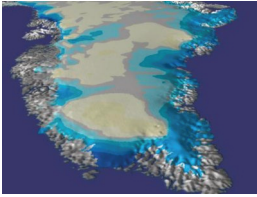


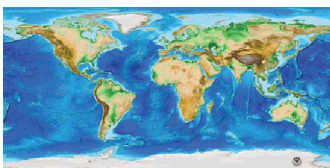
Image of Greenland, viewed from the south, showing changes in ice thickness measured by airborne lidar. Ice sheet thickness decreasing at 40-60 cm per year in darker blue areas (Goddard Space Flight Center, n.d.).

To learn more about the use of lidar in mapping changes in the Greenland ice sheet, visit [NASA's Scientific Visualization Studio](#).

7.13. Global Elevation Data

This page profiles three data products that include elevation (and, in one case, bathymetry) data for all or most of the Earth's surface.

ETOPO1

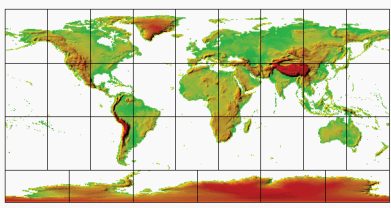


Shaded and colored terrain image produced from ETOPO1 data. (National Geophysical Data Center, 2009).

ETOPO1 is a digital elevation model that includes both topography and bathymetry for the entire world. It consists of more than 233 million elevation values which are regularly spaced at 1 minute of latitude and longitude. **At the equator, the horizontal resolution of ETOPO1 is approximately 1.85 kilometers.** Vertical positions are specified in meters, and there are two versions of the dataset: one with elevations at the "Ice Surface" of the Greenland and Antarctic ice sheets, and one with elevations at "Bedrock" beneath those ice sheets. Horizontal positions are specified in geographic coordinates (decimal degrees). Source data, and thus data quality, vary from region to region.

You can [download ETOPO1 data from the National Geophysical Data Center](#).

GTOPO30



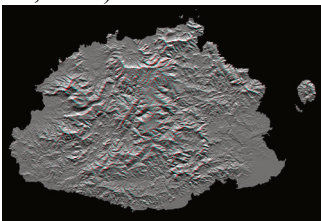
Shaded and colored terrain image produced from GTOPO30 data. Data are distributed as 33 tiles (USGS, 2006b).

GTOPO30 is a digital elevation model that extends over the world's land surfaces (but not under the oceans). GTOPO30 consists of more than 2.5 million elevation values, which are regularly spaced at 30 seconds of latitude and longitude. **At the equator, the resolution of GTOPO30 is approximately 0.925 kilometers — two times greater than ETOPO1.** Vertical positions are specified to the nearest meter, and horizontal positions are specified in geographic coordinates. GTOPO30 data are distributed as tiles, most of which are 50° in latitude by 40° in longitude.

GTOPO30 tiles are available for download from [USGS' EROS Data Center](#). GTOPO60, a resampled and untiled version of GTOPO30, is available through the [USGS' Seamless Data Distribution Service](#).

SHUTTLE RADAR TOPOGRAPHY MISSION (SRTM)

From February 11 to February 22, 2000, the space shuttle Endeavor bounced radar waves off the Earth's surface, and recorded the reflected signals with two receivers spaced 60 meters apart. The mission measured the elevation of land surfaces between 60° N and 57° S latitude. **The highest resolution data products created from the SRTM mission are 30 meters.** Access to 30-meter SRTM data for areas outside the U.S. are restricted by the National Geospatial-Intelligence Agency, which sponsored the project along with the National Aeronautics and Space Administration (NASA). A 90-meter SRTM data product is available for free download without restriction (Maune, 2007).



Anaglyph stereo image derived from Shuttle Radar Topography Mission data (NASA Jet Propulsion Laboratory, 2006).

The image above shows Viti Levu, the largest of the some 332 islands that comprise the Sovereign Democratic Republic of the Fiji Islands. Viti Levu's area is 10,429 square kilometers (about 4000 square miles). Nakauvadra, the rugged mountain range running from north to south, has several peaks rising above 900 meters (about 3000 feet). Mount Tomanivi, in the upper center, is the highest peak at 1324 meters (4341 feet).

Learn more about the Shuttle Radar Topography Mission at Web sites published by [NASA](#) and [USGS](#).

7.14. Bathymetry

The term bathymetry refers to the process and products of measuring the depth of water bodies. The U.S. Congress authorized the comprehensive mapping of the nation's coasts in 1807, and directed that the task be carried out by the

federal government's first science agency, the Office of Coast Survey (OCS). That agency is now responsible for mapping some 3.4 million nautical square miles encompassed by the 12-mile territorial sea boundary, as well as the 200-mile Exclusive Economic Zone claimed by the U.S., a responsibility that entails regular revision of about 1,000 nautical charts. The coastal bathymetry data that appears on USGS topographic maps, like the one shown below, is typically compiled from OCS charts.



”Isobaths” (the technical term for lines of constant depth) shown on a USGS topographic map.

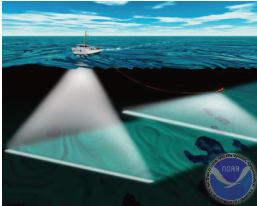
Early hydrographic surveys involved sampling water depths by casting overboard ropes weighted with lead and marked with depth intervals called marks and deeps. Such ropes were called leadlines for the weights that caused them to sink to the bottom. Measurements were called soundings. By the late 19th century, piano wire had replaced rope, making it possible to take soundings of thousands rather than just hundreds of fathoms (a fathom is six feet).



Seaman paying out a sounding line during a hydrographic survey of the East coast of the U.S. in 1916. (NOAA, 2007).

Echo sounders were introduced for deepwater surveys beginning in the 1920s. **Sonar** (SOund NAvigation and Ranging) technologies have revolutionized oceanography in the same way that aerial photography revolutionized topographic mapping. The seafloor topography revealed by sonar and related shipborne remote sensing techniques provided evidence that supported theories about seafloor spreading and plate tectonics.

Below is an artist's conception of an oceanographic survey vessel operating two types of sonar instruments: **multibeam** and **side scan sonar**. On the left, a multibeam instrument mounted in the ship's hull calculates ocean depths by measuring the time elapsed between the sound bursts it emits and the return of echoes from the seafloor. On the right, side scan sonar instruments are mounted on both sides of a submerged “towfish” tethered to the ship. Unlike multibeam, side scan sonar measures the strength of echoes, not their timing. Instead of depth data, therefore, side scanning produces images that resemble black-and-white photographs of the sea floor.

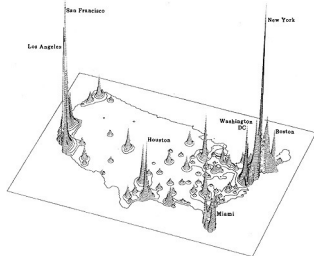


Multibeam and side scan sonar in use for bathymetric mapping. (NOAA, 2002).

A [detailed report of the recent bathymetric survey of Crater Lake, Oregon, USA](#), is published by the USGS [here](#).

7.15. Statistical Surfaces

Strategies used to represent terrain surfaces can be used for other kinds of surfaces as well. For example, one of my first projects here at Penn State was to work with a distinguished geographer, the late Peter Gould, who was studying the diffusion of the Acquired Immune Deficiency Syndrome (AIDS) virus in the United States. Dr. Gould had recently published the map below.



Oblique view of contour lines representing distribution of AIDS cases in the U.S. 1988. (Gould, 1989. © Association of American Geographers. All rights reserved. Reproduced here for educational purposes only).

Gould portrayed the distribution of disease in the same manner as another geographer might portray a terrain surface. The portrayal is faithful to Gould's conception of the contagion as a continuous phenomenon. It was important to Gould that people understood that there was no location that did not have the potential to be visited by the epidemic. For both the AIDS surface and a terrain surface, a quantitative attribute (z) exists for every location (x,y). In general, when a **continuous phenomenon is conceived as being analogous to the terrain surface, the conception is called a statistical surface.**

7.16. Theme: Hydrography

The NSDI Framework Introduction and Reference (FGDC, 1997) envisions the hydrography theme in this way:

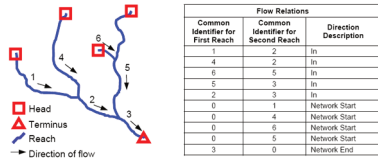
Framework hydrography data include surface water features such as lakes and ponds, streams and rivers, canals, oceans, and shorelines. Each of these features has the attributes of a name and feature identification code. Centerlines and polygons encode the positions of these features. For feature identification codes, many federal and state agencies use the Reach schedule developed by the U.S. Environmental Protection Agency (EPA).

Many hydrography data users need complete information about connectivity of the hydrography network and the direction in which the water flows encoded in the data. To meet these needs, additional elements representing flows of water and connections between features may be included in framework data (p. 20).

IDENTIFICATION

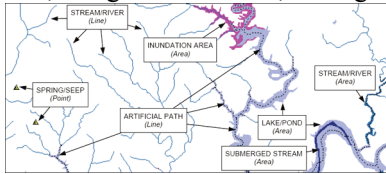
FGDC had the **National Hydrography Dataset (NHD)** in mind when they wrote this description. NHD combines the vector features of **Digital Line Graph (DLG) hydrography** with the **EPA's Reach files**. Reaches are segments of surface water that share similar hydrologic characteristics. Reaches are of three types: transport, coastline, and waterbody. DLG lines features represent the transport and coastline types; polygon features are used to represent waterbodies. Every reach segment in the NHD is assigned a unique reach code, along with a host of other hydrological attributes including stream flow direction (which is encoded in the digitizing order of nodes that make up each segment), network connectivity, and feature names, among others. Because the order of reach codes are sequential from reach to reach, point-source data (such as a pollutant spill) can be geocoded to the affected

reach. Used in this way, reaches comprise **alinear referencing system** comparable to postal addresses along streets (USGS, 2002).



How flow attributes are associated with reaches in the National Hydrography Dataset (USGS, 2000).

NHD parses the U.S. surface drainage network into four hierarchical categories of units: 21 Regions, 222 Subregions, 352 Accounting units, and 2150 Cataloging units (also called Watersheds). Features can exist at multiple levels of the hierarchy, though they might not be represented in the same way. For example, while it might make the most sense to represent a given stream as a polygon features at the Watershed level, it may be more aptly represented as a line feature at the Region or Subregion level. NHD supports this by allowing multiple features to share the same reach codes. Another distinctive feature of NHD is **artificial flowlines**—centerline features that represent paths of water flow through polygon features such as standing water bodies. NHD is complex because it is designed to support sophisticated hydrologic modeling tasks, including point-source pollution modeling, flood potential, bridge construction, among others (Ralston, 2004).



How vector features are used to represent various types of reaches in the National Hydrography Dataset (USGS, 2000).

NHD are available at three levels of detail (scale): medium (1:100,000, which is available for the entire U.S.), high (1:24,000, production of which is underway, “according to the availability of matching resources from NHD partners” (USGS, 2002, p. 2), and local (larger scales such as 1:5,000), which “is being developed where partners and data exist” for select areas (USGS, 2006c; USGS, 2009; USGS 2013).

SPATIAL REFERENCE INFORMATION

NHD coordinates are decimal degrees referenced to the NAD 83 horizontal datum.

DISTRIBUTION

TRY THIS!

DOWNLOAD AND VIEW AN EXTRACT FROM THE NATIONAL HYDROGRAPHY DATASET

1. From the [NHD home page](#) click the **Get Data** link and then follow the link to the **NHD Viewer**. (There is a Help button next to the viewer link.)
2. Use the GIS tools, found above the map area, to pan to and zoom in on an area of interest. Then, click the **Download Data** button.
Choose a *reference area* from the pick list. (You could also chose the entire *current map extent*, but depending upon your zoom level that may be a huge dataset.)

Then click on the map to highlight a specific area of interest. A link to the available data sets will appear in the left hand pane under the **Selection** tab. (The **All Results** button will list the multiple areas you click on.) Follow the **Download** link of the area you are interested in.

3. In the **USGS Available Data** window that opens, select **Hydrography** from the **Theme** column, and choose a file format from the pick list in the **Format** column. I chose **Shapefile** format for my extract, because I know it is compatible with Global Mapper / dlgv32 Pro. If you were working in ArcGIS you could choose the File Geodatabase option. Ralston (2004, p.187) observes that NHD “is precisely the type of information that could benefit from an integrated data model in an object relational database.” Click the **Next** button.

4. From the list of Hydrography Products check the box for what you wish to download. To be assured of getting the data in Shapefile format select a product referred to as a **Dynamic Extract**. Click the **Next** button.

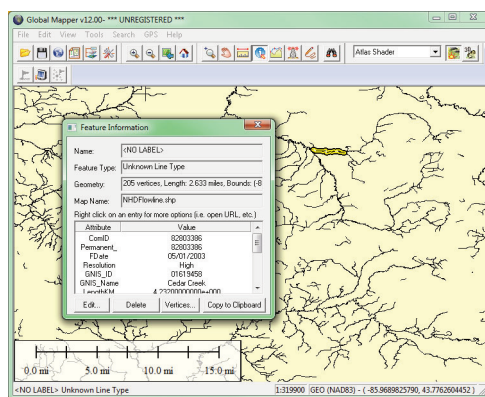
Go to the **Cart** pane on the right. (It may open automatically.)

Go to **Checkout**, supply your e-mail address, and submit your order via the **Place Order** button.

You will receive a message telling you that your order has been placed, and that will soon be followed by an e-mail regarding your order. About an hour and a half after I submitted my request I received a second e-mail containing a download link.

5. Extract the contents of the .zip file and view the Shapefile data set(s) in Global Mapper.

6. Use the Identify pointer tool to reveal attributes of the reaches. In the example below I have highlighted a flowline associated with Cedar Creek in western Michigan.



7.17. Theme: Transportation

Transportation network data are valuable for all sorts of uses, including two we considered in Chapter 4: geocoding and routing. The Federal Geographic Data Committee (1997, p. 19) specified the following vector features and attributes for the transportation framework theme:

Transportation Framework Attributes

Feature	Attributes
Roads	Centerlines, feature identification code (using linear referencing systems where available), functional class, name (including route numbers), and street address ranges
Trails	Centerlines, feature identification code (using linear referencing systems where available), name, and type
Railroads	Centerlines, feature identification code (using linear referencing systems where available), and type
Waterways	Centerlines, feature identification code (using linear referencing systems where available), and name
Airports and ports	Feature identification code and name
Bridges and tunnels	Feature identification code and name

IDENTIFICATION

As part of the National Map initiative, USGS and partners are developing a comprehensive national database of vector transportation data. The transportation theme “includes best available data from Federal partners such as the Census Bureau and the Department of Transportation, State and local agencies” (USGS, 2007).

As envisioned by FGDC, **centerlines** are used to represent transportation routes. Like the lines painted down the middle of two-way streets, centerlines are 1-dimensional vector features that approximate the locations of roads, railroads, and navigable waterways. In this sense, road centerlines are analogous to the flowpaths encoded in the National Hydrologic Dataset (see previous page). Also like the NHD (and TIGER), road **topology** must be encoded to facilitate analysis of transportation networks.

To get a sense of the complexity of the features and attributes that comprise the transportation theme, see the [Transportation Data Model](#) (This is a 36" x 48" poster in a 5.2 Mb PDF file.) [The link to the Transportation Data Model poster recently became disconnected. Instead look at the model diagrams in the [Part 7: Transportation Base](#) of the FGDC Geographic Framework Data Content Standard.]

In the U.S. at least, the best road centerline data is that produced by NAVTEQ and Tele Atlas, which license transportation data to routing sites like Google Maps and MapQuest, and to manufacturers of in-car GPS navigation systems. Because these data are proprietary, however, USGS must look elsewhere for data that can be made available for public use. TIGER/Line data produced by the Census Bureau will likely play an important role after the TIGER/MAF Modernization project is complete (see Chapter 4).

DISTRIBUTION

TRY THIS!

VIEW AND DOWNLOAD NATIONAL MAP TRANSPORTATION DATA

1. [Access the Viewer here.](#)
2. Expand the pane containing the layer options by clicking on Overlays at the upper-left.
3. Under Base Data Layers, click on Transportation. You can expand the Transportation list and sub-select different layers.

4. As you zoom in to larger map scales (using the slider bar at the upper-left of the map), additional transportation layers will become visible.
5. If you wish to download an extract from the transportation database, click the Download Data button in the upper-right of the viewer interface and decide how you wish to extract the data. The Transportation data comes down in ESRI's geodatabase format. Additional information regarding downloading data can be found via the Help button in the upper-right of the viewer interface.

7.18. Theme: Governmental Units

The FGDC framework also includes boundaries of governmental units, including:

- Nation
- States and statistically equivalent areas
- Counties and statistically equivalent areas
- Incorporated places and consolidated cities
- Functioning legal minor civil divisions
- Federal- or state-recognized American Indian reservations and trustlands
- Alaska native regional corporations

FGDC specifies that:

Each of these features includes the attributes of name and the applicable Federal Information Processing Standard (FIPS) code. Features boundaries include information about other features (such as road, railroads, or streams) with which the boundaries are associated and a description of the association (such as coincidence, offset, or corridor. (FGDC, 1997, p. 20-21)

IDENTIFICATION

The USGS National Map aspires to include a comprehensive database of boundary data. In addition to the entities outlined above, the National Map also lists congressional districts, school districts, and ZIP Code zones. Sources for these data include "Federal partners such as the U.S. Census Bureau, other Federal agencies, and State and local agencies." (USGS, 2007).

To get a sense of the complexity of the features and attributes that comprise this theme, see the [Governmental Units Data Model](#) (This is a 36" x 48" poster in a 2.4 Mb PDF file.) [The link to the Governmental Units Data Model poster recently became disconnected. Instead look at the model diagrams in the [Part 5: Governmental unit and other geographic area boundaries](#) of the FGDC Geographic Framework Data Content Standard.]

DISTRIBUTION

TRY THIS!

VIEW AND DOWNLOAD NATIONAL MAP GOVERNMENTAL UNITS DATA

1. [Access the Viewer here.](#)
2. Expand the pane containing the layer options by clicking on Overlays at the upper-left.

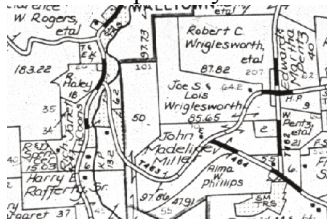
3. Under Base Data Layers, click on Governmental Unit Boundaries. You can expand the this list and sub-select different boundary layers.
4. As you zoom in to larger map scales (using the slider bar at the upper-left of the map), additional boundary layers will become visible.
5. If you wish to download an extract from the Governmental Unit Boundaries database, click the Download Data button in the upper-right of the viewer interface and decide how you wish to extract the data. The Governmental Unit Boundaries data comes down in ESRI's geodatabase format. Additional information regarding downloading data can be found via the Help button in the upper-right of the viewer interface.

7.19. Theme: Cadastral

FGDC (1997, p. 21) points out that:

Cadastral data represent the geographic extent of the past, current, and future rights and interests in real property. The spatial information necessary to describe the geographic extent and the rights and interests includes surveys, legal description reference systems, and parcel-by-parcel surveys and descriptions.

However, no one expects that legal descriptions and survey coordinates of private property boundaries (as depicted schematically in the portion of the **plat map** shown below) will be included in the USGS National Map any time soon. As discussed at the outset of Chapter 6, this is because local governments have authority for land title registration in the U.S., and most of these governments have neither the incentive nor the means to incorporate such data into a publicly-accessible national database.



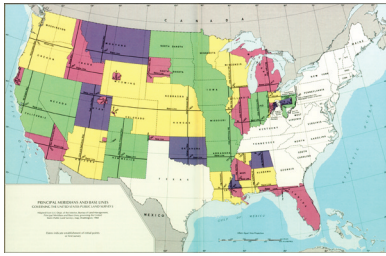
Plat maps are supplementary records that depict property parcel boundaries in graphic form. The geometric accuracy of plats is notoriously poor. The investment required to convert plat maps to properly georeferenced digital data is substantial. Many local governments have converted these records to digital form, or are in the process of doing so.

FGDC's modest goal for the cadastral theme of the NSDI framework is to include:

...cadastral reference systems, such as the Public Land Survey System (PLSS) and similar systems not covered by the PLSS ... and publicly administered parcels, such as military reservations, national forests, and state parks. (Ibid, p. 21)

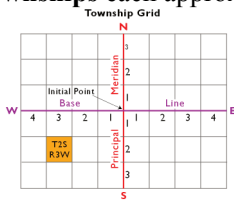
[FGDC's Cadastral Data Content Standard is published here.](#)

The colored areas on the map below show the extent of the United States Public Land Surveys, which commenced in 1784 and took nearly a century to complete (Muehrcke and Muehrcke, 1998). The purpose of the surveys was to partition "public land" into saleable parcels in order to raise revenues needed to retire war debt, and to promote settlement. A key feature of the system is its nomenclature, which provides concise, unique specifications of the location and extent of any parcel.

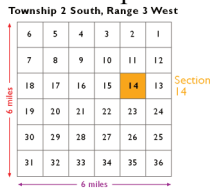


Extent of the U.S. Public Land Survey (Thompson, 1988).

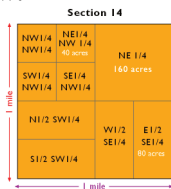
Each Public Land Survey (shown in the colored areas above) commenced from an **initial point** at the precisely surveyed intersection of a **base line** and **principal meridian**. Surveyed lands were then partitioned into grids of **townships** each approximately six miles square.



Townships are designated by their locations relative to the base line and principal meridian of a particular survey. For example, the township highlighted in gold above is the second township south of the baseline and the third township west of the principal meridian. The Public Land Survey designation for the highlighted township is “Township 2 South, Range 3 West.” Because of this nomenclature, the Public Land Survey System is also known as the “township and range system.” Township T2S, R3W is shown enlarged below.



Townships are subdivided into grids of 36 **sections**. Each section covers approximately **one square mile** (640 acres). Notice the back-and-forth numbering scheme. Section 14, highlighted in gold above, is shown enlarged below.



Individual property parcels are designated as shown above. For instance, the **NE 1/4 of Section 14, Township 2 S, Range 3W**, is a 160-acre parcel. Public Land Survey designations specify both the location of a parcel and its area.



The influence of the Public Land Survey grid is evident in the built environment of much of the American Midwest. As Mark Monmonier (1995, p. 114) observes:

The result [of the U.S. Public Land Survey] was an ‘authored landscape’ in which the survey grid had a marked effect on settlement patterns and the shapes of counties and smaller political units. In the typical Midwestern county, roads commonly following section lines, the rural population is dispersed rather than clustered, and the landscape has a pronounced checkerboard appearance.

For more information about the [Public Land Survey System](#), see [this article in the in the USGS’ National Atlas](#).

7.20. Summary

NSDI framework data represent “the most common data themes [that] users need” (FGDC, 1997, p. 3), including geodetic control, orthoimagery, elevation, hydrography, transportation, governmental unit boundaries, and cadastral reference information. Some themes, like transportation and governmental units, represent things that have well-defined edges. In this sense we can think of things like roads and political boundaries as **discrete** phenomena. The **vector** approach to geographic representation is well suited to digitizing discrete phenomena. Line features do a good job of representing roads, for example, and polygons are useful approximations of boundaries.

As you recall from Chapter 1, however, one of the distinguishing properties of the Earth’s surface is that it is **continuous**. Some phenomena distributed across the surface are continuous too. Terrain elevations, gravity, magnetic declination and surface air temperature can be measured practically everywhere. For many purposes, **raster** data are best suited to representing continuous phenomena.

An implication of continuity is that there is an infinite number of locations at which phenomena can be measured. It is not possible, obviously, to take an infinite number of measurements. Even if it were, the mass of data produced would not be usable. The solution, of course, is to collect a sample of measurements, and to estimate attribute values for locations that are left unmeasured. Chapter 7 also considers how missing elevations in a raster grid can be estimated from existing elevations, using a procedure called **interpolation**. The inverse distance weighted interpolation procedure relies upon another fundamental property of geographic data, **spatial dependence**.

The chapter concludes by investigating the characteristics and current status of the hydrography, transportation, governmental units, and cadastral themes. You had the opportunity to access, download, and open several of the data themes using viewers provided by USGS as part of its National Map initiative. In general, you should have found that although neither the NSDI or National Map visions have been fully realized, substantial elements of each is in place. Further progress depends on the American public’s continuing commitment to public data, and to the political will of our representatives in government.

QUIZ

Registered Penn State students should return now to the Chapter 7 folder in ANGEL (via the Resources menu to the left) to access the graded quiz for this chapter. This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter’s learning objectives. You are welcome to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 7.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are

discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

7.21. Bibliography

Federal Geographic Data Committee (1997). *Framework introduction and guide*. Washington DC: Federal Geographic Data Committee.

Eischeid, J. D., Baker, C. B., Karl, R. R., Diaz, H. F. (1995). The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology*, 34, 27-88.

Gould, P. (1989). Geographic dimensions of the AIDS epidemic. *Professional Geographer*, 41:1, 71-77.

Maune, D. F. (Ed.) (2007). *Digital elevation model technologies and applications: The DEM users manual, 2nd edition*. Bethesda, MD: American Society for Photogrammetric Engineering and Remote Sensing.

Monmonier, M. S. (1982). *Drawing the line: tales of maps and cartocontroversy*. New York, NY: Henry Holt.

Muehrcke, P. C. and Muehrcke, J. O. (1998) *Map use*, 4th Ed. Madison, WI: JP Publications.

National Aeronautics and Space Administration, Jet Propulsion Laboratory (2006). *Shuttle radar topography mission*. Retrieved May 10, 2006, from <http://www.jpl.nasa.gov/srtm>

Goddard Space Flight Center, National Aeronautics and Space Administration (n.d.). Greenland’s receding ice. Retrieved February 26, 2008, from <http://svs.gsfc.nasa.gov/stories/greenland/>

National Geophysical Data Center (2010). ETOPO1 global gridded 1 arc-minute database. Retrieved March 2, 2010, from <http://www.ngdc.noaa.gov/mgg/global/global.html>

National Oceanic and Atmospheric Administration, National Climatic Data Center (n. d.). *Merged land-ocean seasonal temperature anomalies*. Retrieved August 18, 1999, from <http://www.ncdc.noaa.gov/onlineprod/landocean/seasonal/form.html> (expired)

National Oceanic and Atmospheric Administration (2002). Side scan and multibeam sonar. Retrieved February 18, 2008, from <http://www.nauticalcharts.noaa.gov/hsd/hydrog.htm>

National Oceanic and Atmospheric Administration (2007). NOAA History. Retrieved February 27, 2008, from <http://www.history.noaa.gov/>

Rabenhorst, T. D. and McDermott, P. D. (1989). *Applied cartography: source materials for mapmaking*. Columbus, OH: Merrill.

Raitz, E. (1948). *General cartography*. New York, NY: McGraw-Hill.

Ralston, B. A. (2004). *GIS and public data*. Clifton Park NY: Delmar Learning.

Thompson, M. M. (1988) *Maps for america*, 3rd Ed. Reston, VA: United States Geological Survey.

United States Geological Survey (1987) *Digital elevation models. Data users guide 5*. Reston, VA: USGS.

United States Geological Survey (1999) *The National Hydrography Dataset. Fact Sheet 106-99*. Reston, VA: USGS. Retrieved February 19, 2008 from <http://erg.usgs.gov/isb/pubs/factsheets/fs10699.html>

United States Geological Survey (2000) *The National Hydrography Dataset: Concepts and Contents*. Reston, VA: USGS. Retrieved February 19, 2008 from http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf

United States Geological Survey (2002) *The National Map – Hydrography. Fact Sheet 060-02*. Reston, VA: USGS. Retrieved February 19, 2008 from <http://erg.usgs.gov/isb/pubs/factsheets/fs06002.html> Retrieved September 22, 2013 from <http://pubs.er.usgs.gov/publication/fs06002>

United States Geological Survey (2006a) *Digital Line Graphs (DLG)*. Reston, VA: USGS. Retrieved February 18, 2008 from <http://edc.usgs.gov/products/map/dlg.html> (In 2010 the site became http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/DLGs)

United States Geological Survey (2006b) GTOPO30. Retrieved February 27, 2008 from <http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html> since moved to <http://www1.gsi.go.jp/geowww/globalmap-gsi/gtopo30/gtopo30.html>

United States Geological Survey (2006c) National Hydrography Dataset (NHD) – High-resolution (Metadata). Reston, VA: USGS. Retrieved February 19, 2008 from http://nhdgeo.usgs.gov/metadata/nhd_high.htm

United States Geological Survey (2007). Vector data theme development of *The National Map*. Retrieved 24 February 2008 from <http://bpgeo.cr.usgs.gov/model/> (expired or moved)

United States Geological Survey (2009) The National Map – Hydrography Dataset. Reston, VA: USGS. Retrieved September 22, 2013 from <http://pubs.usgs.gov/fs/2009/3054/pdf/FS2009-3054.pdf>

United States Geological Survey (2013) National Hydrography Dataset (NHD) – Get NDH Data. Reston, VA: USGS. Retrieved September 22, 2013 from <http://nhd.usgs.gov/data.html>

Chapter 8

Remotely Sensed Image Data

David DiBiase

8.1. Overview

Chapter 7 concluded with the statement that the raster approach is well suited not only to terrain surfaces, but to other continuous phenomena as well. This chapter considers the characteristics and uses of raster data produced with satellite remote sensing systems. Remote sensing is a key source of data for land use and land cover mapping, agricultural and environmental resource management, mineral exploration, weather forecasting, and global change research.

Objectives

The overall goal of the lesson is to acquaint you with the properties of data produced by satellite-based sensors. Specifically, in the lesson you will learn to:

1. Compare and contrast the characteristics of image data produced by photography and digital remote sensing systems;
2. Use the Web to find Landsat data for a particular place and time;
3. Explain why and how remotely sensed image data are processed; and
4. Perform a simulated unsupervised classification of raster image data.

Comments and Questions

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

8.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 8 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 8	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	<p>Submit four practice quizzes including:</p> <ul style="list-style-type: none"> • Nature of Image Data • Visible and Infrared Data • Image Processing • Microwave Data <p>Practice quizzes are not graded and may be submitted more than once.</p>	Go to ANGEL > [your course section] > Lessons tab > Chapter 8 folder > [quiz]
3	<p>Perform “Try this” activities including:</p> <ul style="list-style-type: none"> • Study remote sensing fundamentals and case studies at USGS’ Earthshots site • View the global distribution of Landsat scenes at USGS’ Spacetracks site • Find, preview and acquire Landsat data at the USGS EarthExplorer site • Perform a simulated unsupervised classification of raster image data <p>“Try this” activities are not graded.</p>	Instructions are provided for each activity.
4	Submit the Chapter 8 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 8 folder > Chapter 8 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

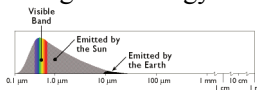
8.3. Nature of Remotely Sensed Image Data

Data, as you know, consist of measurements. Here we consider the nature of the phenomenon that many, though not all, remote sensing systems measure: **electromagnetic energy**. Many of the objects that make up the Earth’s

surface reflect and emit electromagnetic energy in unique ways. The appeal of multispectral remote sensing is that objects that are indistinguishable at one energy wavelength may be easy to tell apart at other wavelengths. You will see that digital remote sensing is a little like scanning a paper document with a desktop scanner, only a lot more complicated.

8.4. Electromagnetic Spectrum

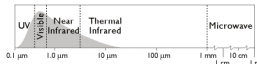
Most remote sensing instruments measure the same thing: electromagnetic radiation. Electromagnetic radiation is a form of energy emitted by all matter above absolute zero temperature (0 Kelvin or -273° Celsius). X-rays, ultraviolet rays, visible light, infrared light, heat, microwaves, and radio and television waves are all examples of electromagnetic energy.



A portion of the electromagnetic spectrum, ranging from wavelengths of 0.1 micrometer (a micrometer is one millionth of a meter) to one meter, within which most remote sensing systems operate. (Adapted from Lillesand & Kiefer, 1994).

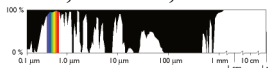
The graph above shows the relative amounts of electromagnetic energy emitted by the Sun and the Earth across the range of wavelengths called the electromagnetic spectrum. Values along the horizontal axis of the graph range from very short wavelengths (ten millionths of a meter) to long wavelengths (meters). Note that the horizontal axis is logarithmically scaled, so that each increment represents a ten-fold increase in wavelength. The axis has been interrupted three times at the long wave end of the scale to make the diagram compact enough to fit on your screen. The vertical axis of the graph represents the magnitude of radiation emitted at each wavelength.

Hotter objects radiate more electromagnetic energy than cooler objects. Hotter objects also radiate energy at shorter wavelengths than cooler objects. Thus, as the graph shows, the Sun emits more energy than the Earth, and the Sun's radiation peaks at shorter wavelengths. The portion of the electromagnetic spectrum at the peak of the Sun's radiation is called the **visible band** because the human visual perception system is sensitive to those wavelengths. Human vision is a powerful means of sensing electromagnetic energy within the visual band. Remote sensing technologies extend our ability to sense electromagnetic energy beyond the visible band, allowing us to see the Earth's surface in new ways, which in turn reveals patterns that are normally invisible.



The electromagnetic spectrum divided into five wavelength bands. (Adapted from Lillesand & Kiefer, 1994).

The graph above names several regions of the electromagnetic spectrum. Remote sensing systems have been developed to measure reflected or emitted energy at various wavelengths for different purposes. This chapter highlights systems designed to record radiation in the bands commonly used for land use and land cover mapping: the visible, infrared, and microwave bands.



The transmissivity of the atmosphere across a range of wavelengths. Black areas indicate wavelengths at which the atmosphere is partially or wholly opaque. (Adapted from Lillesand & Kiefer, 1994).

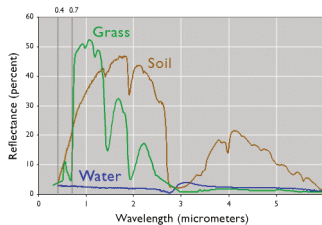
At certain wavelengths, the atmosphere poses an obstacle to satellite remote sensing by absorbing electromagnetic energy. Sensing systems are therefore designed to measure wavelengths within the windows where the transmissivity of the atmosphere is greatest.

8.5. Spectral Response Patterns

The Earth's land surface reflects about three percent of all incoming solar radiation back to space. The rest is either

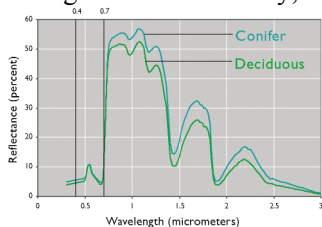
reflected by the atmosphere, or absorbed and re-radiated as infrared energy. The various objects that make up the surface absorb and reflect different amounts of energy at different wavelengths. The magnitude of energy that an object reflects or emits across a range of wavelengths is called its **spectral response pattern**.

The graph below illustrates the spectral response patterns of water, brownish gray soil, and grass between about 0.3 and 6.0 micrometers. The graph shows that grass, for instance, reflects relatively little energy in the visible band (although the spike in the middle of the visible band explains why grass looks green). Like most vegetation, the chlorophyll in grass absorbs visible energy (particularly in the blue and red wavelengths) for use during photosynthesis. About half of the incoming near-infrared radiation is reflected, however, which is characteristic of healthy, hydrated vegetation. Brownish gray soil reflects more energy at longer wavelengths than grass. Water absorbs most incoming radiation across the entire range of wavelengths. Knowing their typical spectral response characteristics, it is possible to identify forests, crops, soils, and geological formations in remotely sensed imagery, and to evaluate their condition.



The spectral response patterns of brownish-gray soil (mollisol), grass, and water. To explore the spectral response characteristics of thousands of natural and man made materials, visit the ASTER Spectral Library at <http://speclib.jpl.nasa.gov/>. (California Institute of Technology, 2002).

The next graph demonstrates one of the advantages of being able to see beyond the visible spectrum. The two lines represent the spectral response patterns of conifer and deciduous trees. Notice that the reflectances within the visual band are nearly identical. At longer, near- and mid-infrared wavelengths, however, the two types are much easier to differentiate. As you'll see later, land use and land cover mapping were previously accomplished by visual inspection of photographic imagery. Multispectral data and digital image processing make it possible to partially automate land cover mapping, which in turn makes it cost effective to identify some land use and land cover categories automatically, all of which makes it possible to map larger land areas more frequently.



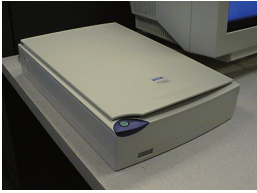
The spectral response patterns of conifer trees and deciduous trees (California Institute of Technology, 1999).

Spectral response patterns are sometimes called **spectral signatures**. This term is misleading, however, because the reflectance of an entity varies with its condition, the time of year, and even the time of day. Instead of thin lines, the spectral responses of water, soil, grass, and trees might better be depicted as wide swaths to account for these variations.

8.6. Raster Scanning

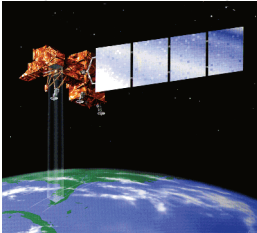
Remote sensing systems work in much the same way as the digital flatbed scanner you may have attached to your personal computer. A desktop scanner creates a digital image of a document by recording, pixel by pixel, the intensity of light reflected from the document. The component that measures reflectance is called the scan head, which consists of a row of tiny sensors that convert light to electrical charges. Color scanners may have three light sources and three sets of sensors, one each for the blue, green, and red wavelengths of visible light. When you push

a button to scan a document, the scan head is propelled rapidly across the image, one small step at a time, recording new rows of electrical signals as it goes. Remotely sensed data, like the images produced by your desktop scanner, consist of reflectance values arrayed in rows and columns that make up raster grids.



Desktop scanner, circa 2000.

After the scan head converts reflectances to electrical signals, another component, called the analog-to-digital converter, converts the electrical charges into digital values. Although reflectances may vary from 0 percent to 100 percent, digital values typically range from 0 to 255. This is because digital values are stored as units of memory called bits. One **bit** represents a single binary integer, 1 or 0. The more bits of data that are stored for each pixel, the more precisely reflectances can be represented in a scanned image. The number of bits stored for each pixel is called the **bit depth** of an image. An 8-bit image is able to represent 28 (256) unique reflectance values. A color desktop scanner may produce 24-bit images in which 8 bits of data are stored for each of the blue, green, and red wavelengths of visible light.



Artist's rendition of the Landsat 7 remote sensing satellite. The satellite does not really cast a four-sided beam of light upon the Earth's surface, of course. Instead, it merely records electromagnetic energy reflected or emitted by the Earth. (NASA, 2001).

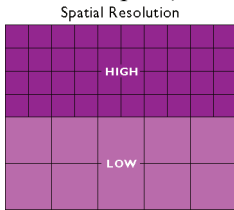
As you might imagine, scanning the surface of the Earth is considerably more complicated than scanning a paper document with a desktop scanner. Unlike the document, the Earth's surface is too large to be scanned all at once, and so must be scanned piece by piece, and mosaicked together later. Documents are flat, but the Earth's shape is curved and complex. Documents lay still while they are being scanned, but the Earth rotates continuously around its axis at a rate of over 1,600 kilometers per hour. In the desktop scanner, the scan head and the document are separated only by a plate of glass; satellite-based sensing systems may be hundreds or thousands of kilometers distant from their targets, separated by an atmosphere that is nowhere near as transparent as glass. And while a document in a desktop scanner is illuminated uniformly and consistently, the amount of solar energy reflected or emitted from the Earth's surface varies with latitude, the time of year, and even the time of day. All of these complexities combine to yield data with geometric and radiometric distortions that must be corrected before the data are useful for analysis. Later in this chapter you'll learn about some of the image processing techniques used to correct remotely sensed image data.

8.7. Resolution

So far you've read that remote sensing systems measure electromagnetic radiation, and that they record measurements in the form of raster image data. The resolution of remotely sensed image data varies in several ways. As you recall, resolution is the least detectable difference in a measurement. In this context, three of the most important kinds are spatial resolution, radiometric resolution and spectral resolution.

Spatial resolution refers to the coarseness or fineness of a raster grid. The grid cells in high resolution data,

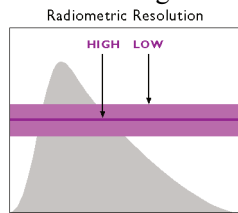
such as those produced by digital aerial imaging, or by the Ikonos satellite, correspond to ground areas as small or smaller than one square meter. Remotely sensed data whose grid cells range from 15 to 80 meters on a side, such as the Landsat ETM+ and MSS sensors, are considered medium resolution. The cells in low resolution data, such as those produced by NOAA's AVHRR sensor, are measured in kilometers. (You'll learn more about all these sensors later in this chapter.)



Spatial resolution is a measure of the coarseness or fineness of a raster grid.

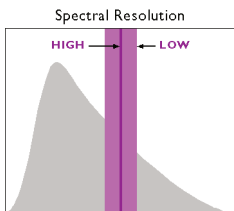
The higher the spatial resolution of a digital image, the more detail it contains. Detail is valuable for some applications, but it is also costly. Consider, for example, that an 8-bit image of the entire Earth whose spatial resolution is one meter could fill 78,400 CD-ROM disks, a stack over 250 feet high (assuming that the data were not compressed). Although data compression techniques reduce storage requirements greatly, the storage and processing costs associated with high resolution satellite data often make medium and low resolution data preferable for analyses of extensive areas.

A second aspect of resolution is **radiometric resolution**, the measure of a sensor's ability to discriminate small differences in the magnitude of radiation within the ground area that corresponds to a single raster cell. The greater the bit depth (number of data bits per pixel) of the images that a sensor records, the higher its radiometric resolution. The AVHRR sensor, for example, stores 210 bits per pixel, as opposed to the 28 bits that the Landsat sensors record. Thus although its spatial resolution is very coarse (~4 km), the Advanced Very High Resolution Radiometer takes its name from its high radiometric resolution.



Radiometric resolution. The area under the curve represents the magnitude of electromagnetic energy emitted by the Sun at various wavelengths. Sensors with low radiometric resolution are able to detect only relatively large differences in energy magnitude (as represented by the lighter and thicker purple band). Sensors with high radiometric resolution are able to detect relatively small differences (represented by the darker and thinner band).

Finally, there is **spectral resolution**, the ability of a sensor to detect small differences in wavelength. For example, panchromatic film is sensitive to a broad range of wavelengths. An object that reflects a lot of energy in the green portion of the visible band would be indistinguishable in a panchromatic photo from an object that reflected the same amount of energy in the red band, for instance. A sensing system with higher spectral resolution would make it easier to tell the two objects apart.



Spectral resolution. The area under the curve represents the magnitude of electromagnetic energy emitted by

the Sun at various wavelengths. Low resolution sensors record energy within relatively wide wavelength bands (represented by the lighter and thicker purple band). High-resolution sensors record energy within narrow bands (represented by the darker and thinner band)

8.8. Site visit to USGS Earthshots

TRY THIS!

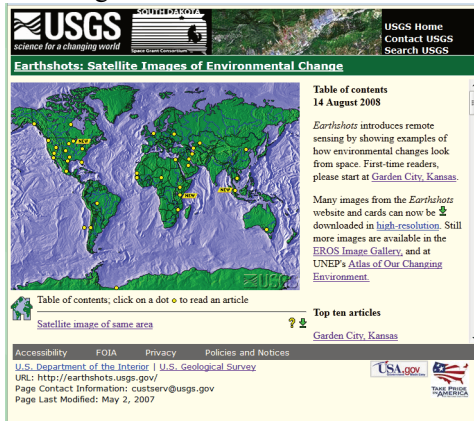
The following exercise involves a site visit to Earthshots, a World Wide Web site created by the USGS to publicize the many contributions of remote sensing to the field of environmental science. There you will view and compare examples of images produced from Landsat data.

The USGS has recently revised the Earthshots website and made it more layman friendly. Unfortunately the new site is much less valuable to our education mission. Fortunately, though, the older web pages are still available. So, after taking you briefly to the new Earthshots homepage, in step 1, I will direct you to the older pages that are more informative for us.

1. To begin, point your browser to the [newer Earthshots site](#). Go ahead and explore this site. Note the information found by following the **About Earthshots** button.

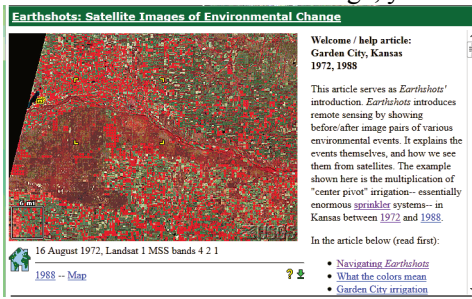


2. Next go to the [USGS Earthshots site](#).

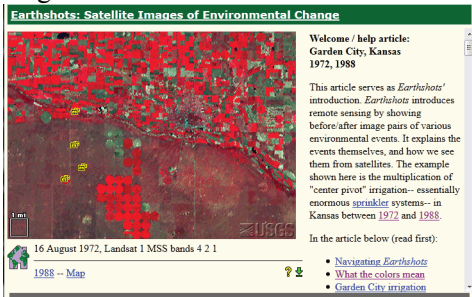


3. **View images produced from Landsat data** Follow the link to the Garden City, Kansas example. You'll be presented with an image created from Landsat data of Garden City, Kansas in 1972. By clicking the date link below

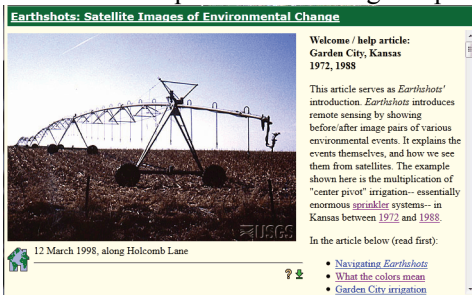
the lower left corner of the image, you can compare images produced from Landsat data collected in 1972 and 1988.



4. Zoom in to a portion of the image Four yellow corner ticks outline a portion of the image that is linked to a magnified view. Click within the ticks to view the magnified image.



5. View a photograph taken on the ground Click on one of the little camera icons arranged one above the other in the western quarter of the image. A photograph taken on the ground will appear.



6. Explore articles linked to the example Find answers to the following questions in the related articles entitled *What the colors mean*, *How images represent Landsat data*, *MSS and TM bands*, and *Beyond looking at pictures*.

1. What is the spectral sensitivity of the Landsat MSS sensor used to capture the image data?
2. Which wavelength bands are represented in the image?
3. What does the red color signify?
4. How was "contrast stretching" used to enhance the images?
5. What is the spatial resolution of the MSS data from which the images were produced?

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 8 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about the **Nature of Image Data**.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

8.9. Visible and Infrared Image Data

Next we explore examples of remotely sensed image data produced by measuring electromagnetic energy in the visible, near-infrared, and thermal infrared bands. Aerial photography is reconsidered as an analog to digital image data. Characteristics of panchromatic data produced by satellite-based cameras and sensors like KVR 1000, IKONOS, and DMSP are compared, as well as multispectral data produced by AVHRR, Landsat, and SPOT. We conclude with an opportunity to shop for satellite image data online.

8.10. Aerial Imaging

Not counting human vision, aerial photography is the earliest remote sensing technology. A Parisian photographer allegedly took the first air photo from a balloon in 1858. If you are (or were) a photographer who uses film cameras, you probably know that photographic films are made up of layers. One or more layers consist of emulsions of light-sensitive silver halide crystals. The black and white film used most often for aerial photography consists of a single layer of silver crystals that are sensitive to the entire visible band of the electromagnetic spectrum. Data produced from photographic film and digital sensors that are sensitive to the entire visible band are called **panchromatic**.

When exposed to light, silver halide crystals are reduced to black metallic silver. The more light a crystal absorbs, the darker it becomes when the film is developed. A black and white panchromatic photo thus represents the intensities of visible electromagnetic energy recorded across the surface of the film at the moment of exposure. You can think of a silver halide crystal as a physical analog for a pixel in a digital image.



A vertical aerial photograph produced from panchromatic film.

Experienced film photographers know to use high-speed film to capture an image in a low-light setting. They also know that faster films tend to produce grainier images. This is because fast films contain larger silver halide grains, which are more sensitive to light than smaller grains. Films used for aerial photography must be fast enough to capture sharp images from aircraft moving at hundreds of feet per second, but not so fast as to produce images so grainy that they mask important details. Thus, the grain size of a photographic film is a physical analog for the spatial resolution of a digital image. Cowen and Jensen (1998) estimated the spatial resolution of a high resolution aerial photograph to be about 0.3-0.5 meters (30-50 cm). Since then, the resolution that can be achieved by digital aerial imaging has increased to as great as 0.05 meters (5 cm). **For applications in which maximum spatial and temporal resolutions are needed, aerial (as opposed to satellite) imaging still has its advantages.**

Earlier in the course you learned (or perhaps you already knew) that the U.S. National Agricultural Imagery Program (NAIP) flies much of the lower 48 states every year. Organizations that need more timely imagery typically hire private aerial survey firms to fly custom photography missions. Most organizations would prefer to have image data in digital form, if possible, since digital image processing (including geometric and radiometric correction) is more efficient than comparable darkroom methods, and because most users want to combine imagery with other data layers in GIS databases. Digital image data are becoming a cost-effective alternative for many applications as the spatial and radiometric resolution of digital sensors increases.

8.11. High Resolution Panchromatic Image Data

Just as digital cameras have replaced film cameras for many of us on the ground, digital sensors are replacing cameras for aerial surveys. This section describes two sources of digital panchromatic imagery with sufficient geometric resolution for some, though certainly not all, large-scale mapping tasks. Also considered is a panchromatic sensing system with sufficient radiometric resolution to detect from space the light emitted by human settlements at night.

KVR-1000 / SPIN-2

High-resolution panchromatic image data first became available to civilians in 1994, when the Russian space agency SOVINFORMSPUTNIK began selling surveillance photos to raise cash in the aftermath of the breakup of the Soviet Union. The photos are taken with an extraordinary camera system called KVR 1000. KVR 1000 cameras are mounted in unmanned space capsules very much like the one in which Russian cosmonaut Yuri Gagarin first traveled into space in 1961. After orbiting the Earth at altitudes of 220 kilometers for about 40 days, the capsules separate from the Cosmos rockets that propelled them into space, and spiral slowly back to Earth. After the capsules parachute to the surface, ground personnel retrieve the cameras and transport them to Moscow, where the film is developed. Photographs are then shipped to the U.S., where they are scanned and processed by Kodak Corporation. The final product is two-meter resolution, georeferenced, and orthorectified digital data called SPIN-2 (Space Information 2-Meter). A firm called Aerial Images Inc. was licensed in 1998 to distribute SPIN-2 data in the U.S (SPIN-2, 1999).



Portion of a SPIN-2 image of Dallas, Texas.

© 2006 Aerial Images, Inc. All rights reserved. Used by Permission.

IKONOS

Also in 1994, when the Russian space agency first began selling its space surveillance imagery, a new company called Space Imaging, Inc. was chartered in the U.S. Recognizing that high-resolution images were then available commercially from competing foreign sources, the U.S. government authorized private firms under its jurisdiction to produce and market remotely sensed data at spatial resolutions as high as one meter. By 1999, after a failed first attempt, Space Imaging successfully launched its Ikonos I satellite into an orbital path that circles the Earth 640 km above the surface, from pole to pole, crossing the equator at the same time of day every day. Such an orbit is called **asun synchronous polar orbit**, in contrast with the **geosynchronous** orbits of communications and some weather satellites that remain over the same point on the Earth's surface at all times.



Portion of a 1-meter resolution panchromatic image of Washington, DC produced by the Ikonos I sensor. © 2001 Space Imaging, inc. All Rights Reserved. Used by permission.

Ikonos' panchromatic sensor records reflectances in the visible band at a spatial resolution of one meter, and a bit depth of eleven bits per pixel. The expanded bit depth enables the sensor to record reflectances more precisely, and allows technicians to filter out atmospheric haze more effectively than is possible with 8-bit imagery. Archived, unrectified, panchromatic Ikonos imagery within the U.S. is available for as little as \$7 per square kilometer, but new orthorectified imagery costs \$28 per square kilometer and up.

A competing firm called ORBIMAGE acquired Space Imaging in early 2006, after ORBIMAGE secured a half-billion dollar contract with the National Geospatial-Intelligence Agency. The merged companies were called GeoEye, Inc. In early 2013 GeoEye was merged into the [DigitalGlobe corporation](#). A satellite named GeoEye-1 was launched in 2008 and provides sub-meter (0.41 meter) panchromatic resolution (GeoEye, 2007). In 2013 a new satellite named GeoEye-2 is due to become operational. It will have a panchromatic resolution of 0.34 meters.

DMSP

The U.S. Air Force initiated its Defense Meteorology Satellite Program (DMSP) in the mid-1960s. By 2001, they had launched fifteen DMSP satellites. The satellites follow polar orbits at altitudes of about 830 km, circling the Earth every 101 minutes.

The program's original goal was to provide imagery that would aid high-altitude navigation by Air Force pilots. DMSP satellites carry several sensors, one of which is sensitive to a band of wavelengths encompassing the visible and near-infrared wavelengths (0.40-1.10 μm). The spatial resolution of this panchromatic sensor is low (2.7 km), but its radiometric resolution is high enough to record moonlight reflected from cloud tops at night. During cloudless new moons, the sensor is able to detect lights emitted by cities and towns. Image analysts have successfully correlated patterns of night lights with population density estimates produced by the U.S. Census Bureau, enabling analysts to use DMSP imagery (in combination with other data layers, such as transportation networks) to monitor changes in global population distribution.



Composite image of the Earth at night showing lights from cities and towns throughout the world. This image was recorded by DMSP satellite sensors. (Geophysical Data Center, 2005).

8.12. Multispectral Image Data

The previous section highlighted the one-meter panchromatic data produced by the IKONOS satellite sensor. Pan data is not all that IKONOS produces, however. It is a multispectral sensor that records reflectances within four other (narrower) bands, including the blue, green, and red wavelengths of the visible spectrum, and the near-infrared band. **The range(s) of wavelengths that a sensor is able to detect is called its spectral sensitivity.**

IKONOS I (1999-)

Spectral Sensitivity	Spatial Resolution
0.45 – 0.90 μm (panchromatic)	1m
0.45 – 0.52 μm (visible blue)	4m
0.51 – 0.60 μm (visible green)	4m
0.63 – 0.70 μm (visible red)	4m
0.76 – 0.85 μm (near IR)	4m

Spectral sensitivities and spatial resolution of the IKONOS I sensor. Wavelengths are expressed in micrometers (millionths of a meter). Spatial resolution is expressed in meters.

This section profiles two families of multispectral sensors that play important roles in land use and land cover characterization: AVHRR and Landsat.

AVHRR

AVHRR stands for “Advanced Very High Resolution Radiometer.” AVHRR sensors have been onboard sixteen satellites maintained by the National Oceanic and Atmospheric Administration (NOAA) since 1979 (TIROS-N, NOAA-6 through NOAA-15). The data the sensors produce are widely used for large-area studies of vegetation, soil moisture, snow cover, fire susceptibility, and floods, among other things.

AVHRR sensors measure electromagnetic energy within five spectral bands, including visible red, near infrared, and three thermal infrared. The visible red and near-infrared bands are particularly useful for large-area vegetation monitoring. The Normalized Difference Vegetation Index (NDVI), a widely used measure of photosynthetic activity that is calculated from reflectance values in these two bands, is discussed later.

AVHRR (1979-)

Spectral Sensitivity	Spatial Resolution
0.58 – 0.68 μm (visible red)	1-4 km*
0.725 – 1.10 μm (near IR)	1-4 km*
3.55 – 3.93 μm (thermal IR)	1-4 km*
10.3 – 11.3 μm (thermal IR)	1-4 km*
11.5 – 12.5 μm (thermal IR)	1-4 km*

Spectral sensitivities and spatial resolution of the AVHRR sensor. Wavelengths are expressed in micrometers (millionths of a meter). Spatial resolution is expressed in kilometers (thousands of meters). *Spatial resolution of AVHRR data varies from 1 km to 16 km. Processed data consist of uniform 1 km or 4 km grids.

The NOAA satellites that carry AVHRR sensors trace **sun-synchronous polar orbits** at altitudes of about 833 km. Traveling at ground velocities of over 6.5 kilometers per second, the satellites orbit the Earth 14 times daily (every 102 minutes), crossing over the same locations along the equator at the same times every day. As it orbits, the AVHRR sensor sweeps a scan head along a 110°-wide arc beneath the satellite, taking many measurements every second. (The back and forth sweeping motion of the scan head is said to resemble a whisk-broom.) The arc corresponds to a ground swath of about 2400 km. Because the scan head traverses so wide an arc, its **instantaneous field of view** (IFOV: the ground area covered by a single pixel) varies greatly. Directly beneath the satellite, the

IFOV is about 1 km square. Near the edge of the swath, however, the IFOV expands to over 16 square kilometers. To achieve uniform resolution, the distorted IFOVs near the edges of the swath must be resampled to a 1 km grid (Resampling is discussed later in this chapter). The AVHRR sensor is capable of producing daily global coverage in the visible band, and twice daily coverage in the thermal IR band.

For more information, visit the [USGS' AVHRR home page](#).

LANDSAT MSS

Television images of the Earth taken from early weather satellites (such as TIROS-1, launched in 1960), and photographs taken by astronauts during the U.S. manned space programs in the 1960s, made scientists wonder about how such images could be used for environmental resource management. In the mid 1960s, U.S. National Aeronautics and Space Administration (NASA) and the Department of Interior began work on a plan to launch a series of satellite-based orbiting sensors. The Earth Resource Technology Satellite program launched its first satellite, ERTS-1, in 1972. When the second satellite was launched in 1975, NASA renamed the program Landsat. Since then, there have been six successful Landsat launches (Landsat 6 failed shortly after takeoff in 1993; Landsat 7 successfully launched in 1999).

Two sensing systems were on board Landsat 1 (formerly ERTS-1): a Return Beam Videcon (RBV) and a Multispectral Scanner (MSS). The RBV system is analogous to today's digital cameras. It sensed radiation in the visible band for an entire 185 km square scene at once, producing images comparable to color photographs. The RBV sensor was discontinued after Landsat 3, due to erratic performance and a general lack of interest in the data it produced.

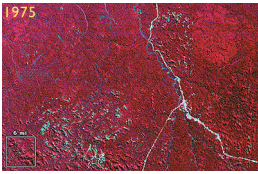
The MSS sensor, however, enjoyed much more success. From 1972 through 1992 it was used to produce an archive of over 630,000 scenes. MSS measures radiation within four narrow bands: one that spans visible green wavelengths, another that spans visible red wavelengths, and two more spanning slightly longer, near-infrared wavelengths.

Landsat MSS (1972-1992)

Spectral Sensitivity	Spatial Resolution
0.5 – 0.6 μm (visible green)	79 / 82 m*
0.6 – 0.7 μm (visible red)	79 / 82 m*
0.7 – 0.8 μm (near IR)	79 / 82 m*
0.8 – 1.1 μm (near IR)	79 / 82 m*

Spectral sensitivities and spatial resolution of the Landsat MSS sensor. Wavelengths are expressed in micrometers (millionths of a meter). Spatial resolution is expressed in meters. *MSS sensors aboard Landsats 4 and 5 had nominal resolution of 82 m, which includes 15 meters of overlap with previous scene.

Landsats 1 through 3 traced near-polar orbits at altitudes of about 920 km, orbiting the Earth 14 times per day (every 103 minutes). Landsats 4 and 5 orbit at 705 km altitude. The MSS sensor sweeps an array of six energy detectors through an arc of less than 12°, producing six rows of data simultaneously across a 185 km ground swath. Landsat satellites orbit the Earth at similar altitudes and velocities as the satellites that carry AVHRR, but because the MSS scan swath is so much narrower than AVHRR, it takes much longer (18 days for Landsat 1-3, 16 days for Landsats 4 and 5) to scan the entire Earth's surface.



Three scenes produced by a Landsat Multispectral Scanner. Images reveal Amazonian rainforest cleared in the Brazilian state of Rondonia between 1975 and 1992. Red areas represent healthy vegetation (USGS, 2001).

The sequence of three images shown above cover the same portion of the state of Rondonia, Brazil. Reflectances in the near-infrared band are coded red in these images; reflectances measured in the visible green and red bands are coded blue and green. Since vegetation absorbs visible light, but reflects infrared energy, the blue-green areas indicate cleared land. Land use change detection is one of the most valuable uses of multispectral imaging.

For more information, visit [USGS' Landsat home page](#)

LANDSAT TM AND ETM+

As NASA prepared to launch Landsat 4 in 1982, it replaced the unsuccessful RBV sensor with a new sensing system called Thematic Mapper (TM). TM was a new and improved version of MSS that featured higher spatial resolution (30 meters in most channels) and expanded spectral sensitivity (seven bands, including visible blue, visible green, visible red, near-infrared, two mid-infrared, and thermal infrared wavelengths). An Enhanced Thematic Mapper Plus (ETM+) sensor, which includes an eighth (panchromatic) band with a spatial resolution of 15 meters, was onboard Landsat 7 when it successfully launched in 1999.

Landsat TM & ETM+ (1982-)

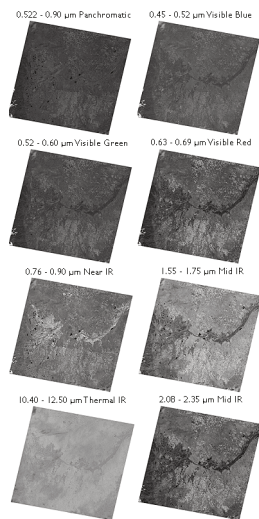
Spectral Sensitivity	Spatial Resolution
0.522 - 0.90 μm (panchromatic)*	15 m*
0.45 - 0.52 μm (visible blue)	30 m
0.52 - 0.60 μm (visible green)	30 m
0.63 - 0.69 μm (visible red)	30 m
0.76 - 0.90 μm (near IR)	30 m
1.55 - 1.75 μm (mid IR)	30 m
10.40 - 12.50 μm (thermal IR)	120 m (Landsat 4-5) 60 m (Landsat 7)
2.08 - 2.35 μm (mid IR)	30 m

Spectral sensitivities and spatial resolution of the Landsat TM and ETM sensors. Wavelengths are expressed in micrometers (millionths of a meter). Spatial resolution is expressed in meters. Note the lower spatial resolution in thermal IR band, which allows for increased radiometric resolution. *ETM+/Landsat 7 only.

The spectral sensitivities of the TM and ETM+ sensors are attuned to both the spectral response characteristics of the phenomena that the sensors are designed to monitor, as well as to the windows within which electromagnetic energy are able to penetrate the atmosphere. The following table outlines some of the phenomena that are revealed by each of the wavelengths bands, phenomena that are much less evident in panchromatic image data alone.

Phenomena revealed by different bands of Landsat TM/ETM data

Band	Phenomena Revealed
0.45 – 0.52 μm (visible blue)	Shorelines and water depths (these wavelenths penetrate water)
0.52 – 0.60 μm (visible green)	Plant types and vigor (peak vegetation reflects these wavelengths strongly)
0.63 – 0.69 μm (visible red)	Photosynthetic activity (plants absorb these wavelengths during photosynthesis)
0.76 – 0.90 μm (near IR)	Plant vigor (healthy plant tissue reflects these wavelengths strongly)
1.55 – 1.75 μm (mid IR)	Plant water stress, soil moisture, rock types, cloud cover vs. snow
10.40 – 12.50 μm (thermal IR)	Relative amounts of heat, soil moisture
2.08 – 2.35 μm (mid IR)	Plant water stress, mineral and rock types



Images produced from 8 bands of Landsat 7 ETM data of Denver, CO. Each image in the illustration represents reflectance values recorded in each wavelength band. False color images are produced by coloring three bands (for example, the visible green, visible red, and near-infrared bands) using blue, green, and red, like the layers in color photographic film.

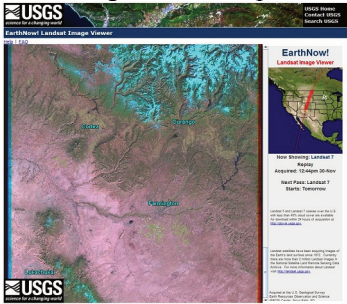
Until 1984, Landsat data were distributed by the U.S. federal government (originally by the USGS's EROS Data Center, later by NOAA). Data produced by Landsat missions 1 through 4 are still available for sale from EROS. With the Land Remote Sensing Commercialization Act of 1984, however, the U.S. Congress privatized the Landsat program, transferring responsibility for construction and launch of Landsat 5, and for distribution of the data it produced, to a firm called EOSAT.

Dissatisfied with the prohibitive costs of unsubsidized data (as much as \$4,400 for a single 185 km by 170 km scene), users prompted Congress to pass the Land Remote Sensing Policy Act of 1992. The new legislation returned responsibility for the Landsat program to the U.S. government. Data produced by Landsat 7 is distributed by USGS at a cost to users of \$600 per scene (about 2 cents per square kilometer). Scenes that include data gaps caused by a "scan line corrector" failure are sold for \$250; \$275 for scenes in which gaps are filled with earlier data.

TRY THIS!

You may choose to visit the USGS' [EarthNow! Landsat Image Viewer](#), which allows you to view live acquisition

of Landsat 5 and Landsat 7 images. This site has a link to the [USGS Global Visualization Viewer](#) where you can search for images based on percent cloud cover.



A snapshot of the live EarthNow! Landsat Image Viewer (USGS, 2011).

8.13. Using EarthExplorer to Find Landsat Data

TRY THIS!

This activity involves using the USGS EarthExplorer system to find Landsat data that corresponds with the scene of the Denver, Colorado area illustrated earlier. At the end of the experiment you can search for data in your own area of interest.

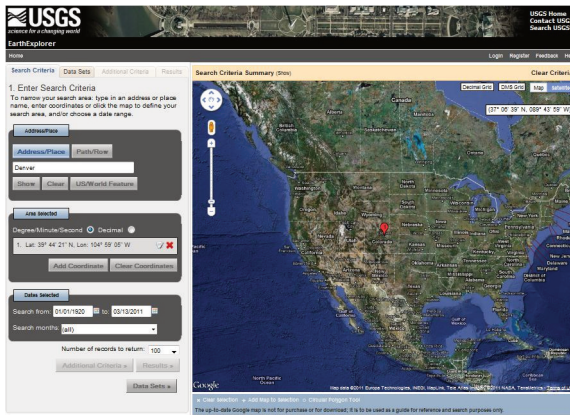
EarthExplorer is a Web application that enables users to find, preview, and download or order digital data published by the U.S. Geological Survey. In addition to Landsat MSS, TM, and ETM+ data, AVHRR, DOQ, aerial photography, and other data are also available from the site.

Begin by pointing your browser to EarthExplorer. (Clicking on this link opens a separate window featuring the EarthExplorer Web site. You may enlarge the window and work within it, or if you prefer, open a separate browser and type in the EarthExplorer Web address.)

- You don't have to register to use EarthExplorer, **unless** you want to download data.
- In order to uncompress the data files that you might choose to download you will need an application that is capable of uncompressing a .tar.gz file. One such application is 7-Zip. You can [download 7-Zip here](#)

1. Enter your search criteria

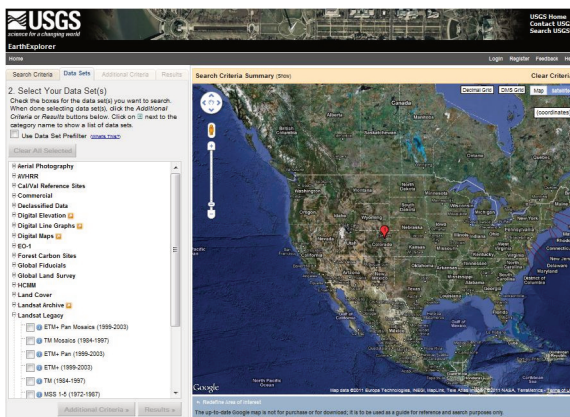
- Enter Address/Place name: **Denver** on the first tab **Search Criteria**
- Click the Address/Place name **Show** button
- “Denver, CO, USA” is returned (as are several other matches on the “Denver” string), along with latitude and longitude coordinates needed to perform a spatial search of EarthExplorer's database. Click on the Denver, CO, USA choice from the list. It may take several seconds, but the display on the left will change to show only the location coordinates for Denver, CO, and a location marker will appear on the map. See below.



EarthExplorer home page, March 2011.

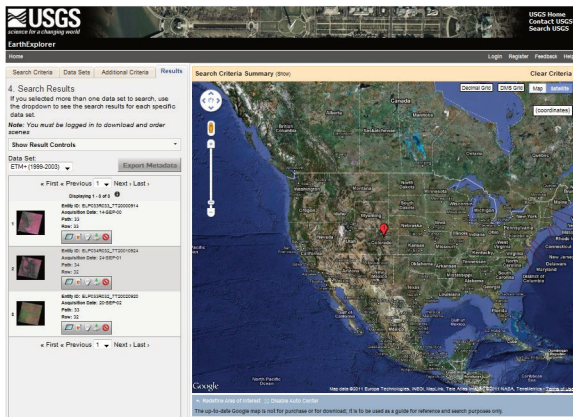
2. Select your Data Set(s) Our objective is to find the Landsat ETM+ data that is illustrated, band by band, in the previous page.

- In the frame in the left side of the window click on the second tab **Data Sets**
- Mark the checkbox to select **ETM+ (1999-2003)** under **Landsat Legacy**.

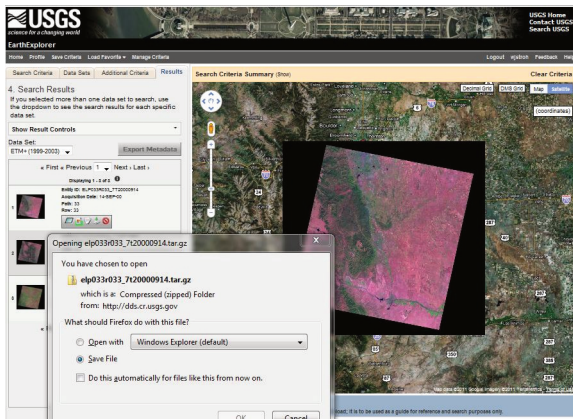


3. View your Results

- The list of **Results** will show up on the fourth tab.

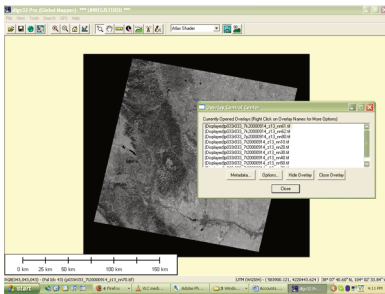


- Click the image icon for one of the results to show a full display of that result, including attribute information.
- Click the “Show Browse Overlay” icon or the “Show Footprint” icon to bring up and overlay or footprint over the Google map.
- Click “Download” to download a specific result. (NOTE: You will need to login or register and then login to the USGS to download data.)



Overlay and footprint image of Landsat scene and download dialog box, EarthExplorer, March 2011.

Downloaded data arrives on your desktop in a double-compressed archive format. For instance, the archive I downloaded is named “elp033r033_7t20000914.tar.gz” To open and view the data in Global Mapper / dlgv32 Pro, I had to first extract the .tar archive from the .gz archive, then extract .tif files from the .tar archive. I used the 7-Zip application, that I mentioned above, to extract the files: right-click on an archive, then choose 7-Zip > Extract files.... The screen capture below shows one of the eight Landsat images (corresponding to the eight ETM+ bands) in Global Mapper.



Landsat scene viewed in Global Mapper software.

Use EarthExplorer to find Landsat data for your own area of interest

- Use the Redefine Criteria link at the bottom of the Results page to begin a new search. Try searching for your hometown or a place you've always wanted to visit.

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 8 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **Visible and Infrared Data**.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

8.14. Multispectral Image Processing

One of the main advantages of digital data is that they can be readily processed using digital computers. Over the next few pages we focus on digital image processing techniques used to correct, enhance, and classify remotely sensed image data.

8.15. Image Correction

As suggested earlier, scanning the Earth's surface from space is like scanning a paper document with a desktop scanner, only a lot more complicated. Raw remotely sensed image data are full of geometric and radiometric flaws caused by the curved shape of the Earth, the imperfectly transparent atmosphere, daily and seasonal variations in the amount of solar radiation received at the surface, and imperfections in scanning instruments, among other things. Understandably, most users of remotely sensed image data are not satisfied with the raw data transmitted from satellites to ground stations. Most prefer preprocessed data from which these flaws have been removed.

GEOMETRIC CORRECTION

You read in Chapter 6 that scale varies in unrectified aerial imagery due to the relief displacement caused by variations in terrain elevation. Relief displacement is one source of geometric distortion in digital image data, although it is less of a factor in satellite remote sensing than it is in aerial imaging because satellites fly at much higher altitudes than airplanes. Another source of geometric distortions is the Earth itself, whose curvature and eastward spinning motion are more evident from space than at lower altitudes.

The Earth rotates on its axis from west to east. At the same time, remote sensing satellites like IKONOS, Landsat, and the NOAA satellites that carry the AVHRR sensor, orbit the Earth from pole to pole. If you were to plot on a cylindrical projection the flight path that a polar orbiting satellite traces over a 24-hour period, you would see

a series of S-shaped waves. As a remote sensing satellite follows its orbital path over the spinning globe, each scan row begins at a position slightly west of the row that preceded it. In the raw scanned data, however, the first pixel in each row appears to be aligned with the other initial pixels. To properly georeference the pixels in a remotely sensed image, pixels must be shifted slightly to the west in each successive row. This is why processed scenes are shaped like skewed parallelograms when plotted in geographic or plane projections, as shown in the illustration you saw earlier that showed eight bands of a Landsat 7 ETM+ scene.

In addition to the systematic error caused by the Earth's rotation, random geometric distortions result from relief displacement, variations in the satellite altitude and attitude, instrument misbehaviors, and other anomalies. Random geometric errors are corrected through a process known as **rubber sheeting**. As the name implies, rubber sheeting involves stretching and warping an image to georegister control points shown in the image to known control point locations on the ground. First, a pair of plane coordinate transformation equations is derived by analyzing the differences between control point locations in the image and on the ground. The equations enable image analysts to generate a rectified raster grid. Next, reflectance values in the original scanned grid are assigned to the cells in the rectified grid. Since the cells in the rectified grid don't align perfectly with the cells in the original grid, reflectance values in the rectified grid cells have to be interpolated from values in the original grid. This process is called **resampling**. Resampling is also used to increase or decrease the spatial resolution of an image so that its pixels can be georegistered with those of another image.

RADIOMETRIC CORRECTION

The reflectance at a given wavelength of an object measured by a remote sensing instrument varies in response to several factors, including the illumination of the object, its reflectivity, and the transmissivity of the atmosphere. Furthermore, the response of a given sensor may degrade over time. With these factors in mind, it should not be surprising that an object scanned at different times of the day or year will exhibit different radiometric characteristics. Such differences can be advantageous at times, but they can also pose problems for image analysts who want to mosaic adjoining images together, or to detect meaningful changes in land use and land cover over time. To cope with such problems, analysts have developed numerous radiometric correction techniques, including Earth-sun distance corrections, sun elevation corrections, and corrections for atmospheric haze.

To compensate for the different amounts of illumination of scenes captured at different times of day, or at different latitudes or seasons, image analysts may divide values measured in one band by values in another band, or they may apply mathematical functions that normalize reflectance values. Such functions are determined by the distance between the Earth and the sun and the altitude of the sun above the horizon at a given location, time of day, and time of year. Analysts depend on metadata that includes the location, date, and time at which a particular scene was captured.

Image analysts may also correct for the contrast-diminishing effects of atmospheric haze. **Haze compensation** resembles the differential correction technique used to improve the accuracy of GPS data in the sense that it involves measuring error (or, in this case, spurious reflectance) at a known location, then subtracting that error from another measurement. Analysts begin by measuring the reflectance of an object known to exhibit near-zero reflectance under non-hazy conditions, such as deep, clear water in the near-infrared band. Any reflectance values in those pixels can be attributed to the **path radiance** of atmospheric haze. Assuming that atmospheric conditions are uniform throughout the scene, the haze factor may be subtracted from all pixel reflectance values. Some new sensors allow "self calibration" by measuring atmospheric water and dust content directly.

Geometric and radiometric correction services are commonly offered by government agencies and private firms that sell remotely sensed data. For example, although the USGS offers raw (Level Zero R) Landsat 7 data for just \$475 per 185 km by 170 km scene, many users find the \$600 radiometrically corrected, orthorectified, and custom projected Level One G data well worth the added expense. GeoEye (formerly Space Imaging) routinely delivers preprocessed IKONOS data that has been radiometrically corrected, orthorectified, projected and

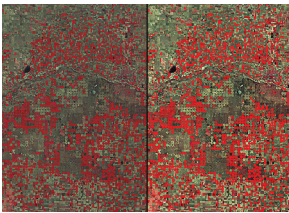
mosaicked to user specifications. Four levels of geometric accuracy are available, from 12 meters (which satisfies National Map Accuracy Standards at 1:50,000 scale) to 1 meter (1:2,400 NMAS).

8.16. Image Enhancement

Correction techniques are routinely used to resolve geometric, radiometric, and other problems found in raw remotely sensed data. Another family of image processing techniques is used to make image data easier to interpret. These so-called image enhancement techniques include contrast stretching, edge enhancement, and deriving new data by calculating differences, ratios, or other quantities from reflectance values in two or more bands, among many others. This section considers briefly two common enhancement techniques, contrast stretching and derived data. Later you'll learn how vegetation indices derived from two bands of AVHRR imagery are used to monitor vegetation growth at a global scale.

CONTRAST STRETCHING

Consider the pair of images shown side by side below. Although both were produced from the same Landsat MSS data, you will notice that the image on the left is considerably dimmer than the one on the right. The difference is a result of contrast stretching. As you recall, MSS data have a precision of 8 bits, that is, reflectance values are encoded as 256 (28) intensity levels. As is often the case, reflectances in the near-infrared band of the scene partially shown below ranged from only 30 and 80 in the raw image data. This limited range results in an image that lacks contrast and, consequently, appears dim. **The image on the right shows the effect of stretching the range of reflectance values in the near-infrared band from 30-80 to 0-255, and then similarly stretching the visible green and visible red bands.** As you can see, the contrast-stretched image is brighter and clearer.



Pair of images produced from Landsat MSS data captured in 1988. The near-infrared band is shown in red, the visible red is shown in green, and the visible green band is shown in blue. The right and left images show the before and after effects of contrast stretching. The images show agricultural patterns characteristic of center-pivot irrigation in a portion of a county in southwestern Kansas. (USGS, 2001a).

DERIVED DATA: NDVI

One advantage of multispectral data is the ability to derive new data by calculating differences, ratios, or other quantities from reflectance values in two or more wavelength bands. For instance, detecting stressed vegetation amongst healthy vegetation may be difficult in any one band, particularly if differences in terrain elevation or slope cause some parts of a scene to be illuminated differently than others. The ratio of reflectance values in the visible red band and the near-infrared band compensates for variations in scene illumination, however. Since the ratio of the two reflectance values is considerably lower for stressed vegetation regardless of illumination conditions, detection is easier and more reliable.

Besides simple ratios, remote sensing scientists have derived other mathematical formulae for deriving useful new data from multispectral imagery. One of the most widely used examples is the Normalized Difference Vegetation Index (NDVI). As you may recall, the AVHRR sensor measures electromagnetic radiation in five wavelengths bands. NDVI scores are calculated pixel-by-pixel using the following algorithm:

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$$

R stands for the visible red band (AVHRR channel 1), while NIR represents the near-infrared band (AVHRR channel 2). The chlorophyll in green plants strongly absorbs radiation in AVHRR's visible red band (0.58-0.68 μm) during photosynthesis. In contrast, leaf structures cause plants to strongly reflect radiation in the near-infrared band (0.725-1.10 μm). NDVI scores range from -1.0 to 1.0. A pixel associated with low reflectance values in the visible band and high reflectance in the near-infrared band would produce an NDVI score near 1.0, indicating the presence of healthy vegetation. Conversely, the NDVI scores of pixels associated with high reflectance in the visible band and low reflectance in the near-infrared band approach -1.0, indicating clouds, snow, or water. NDVI scores near 0 indicate rock and non-vegetated soil.

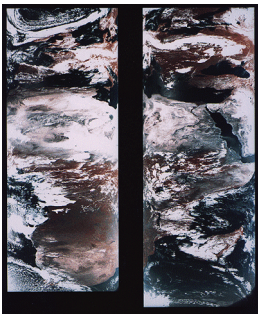
Applications of the NDVI range from local to global. At the local scale, the Mondavi Vineyards in Napa Valley California can attest to the utility of NDVI data in monitoring plant health. In the 1993, the vineyards suffered an infestation of phylloxera, a species of plant lice that attacks roots and is impervious to pesticides. The pest could only be overcome by removing infested vines and replacing them with more resistant root stock. The vineyard commissioned a consulting firm to acquire high-resolution (2-3 meter) visible and near-infrared imagery during consecutive growing seasons using an airborne sensor. Once the data from the two seasons were georegistered, comparison of NDVI scores revealed areas in which vine canopy density had declined. NDVI change detection proved to be such a fruitful approach that the vineyards adopted it for routine use as part of their overall precision farming strategy (Colucci, 1998).

The example that follows outlines the image processing steps involved in producing a global NDVI data set.

8.17. Processing the 1Km Global Land Dataset

The Advanced Very High Resolution Radiometer (AVHRR) sensors aboard NOAA satellites scan the entire Earth daily at visible red, near-infrared, and thermal infrared wavelengths. In the late 1980s and early 1990s, several international agencies identified the need to compile a baseline, cloud-free, global NDVI data set in support of efforts to monitor global vegetation cover. For example, the United Nations mandated its Food and Agriculture Organization to perform a global forest inventory as part of its Forest Resources Assessment project. Scientists participating in NASA's Earth Observing System program also needed a global AVHRR data set of uniform quality to calibrate computer models intended to monitor and predict global environmental change. In 1992, under contract with the USGS, and in cooperation with the International Geosphere Biosphere Programme, scientists at the EROS Data Center in Sioux Falls, South Dakota, started work. Their goals were to create not only a single 10-day composite image, but also a 30-month time series of composites that would help Earth system scientists to understand seasonal changes in vegetation cover at a global scale. This example highlights the image processing procedures used to produce the data set. Further information is available in Eidenshink & Faundeen (1994) and at the [project home page](#).

From 1992 through 1996, a network of 30 ground receiving stations acquired and archived tens of thousands of scenes from an AVHRR sensor aboard one of NOAA's polar orbiting satellites. Individual scenes were stitched together into daily orbital passes like the ones illustrated below. Creating orbital passes allowed the project team to discard the redundant data in overlapping scenes acquired by different receiving stations.



False color images produced from AVHRR data acquired on June 24, 1992 for the 1-Km AVHRR Global Land Dataset project. The images represent orbital passes created by splicing consecutive scenes. The 2400-km wide swaths cover Europe, Africa, and the Near East. Note the cloud cover and geometric distortion. (Eidenshink & Faundeen, 1994).

Once the daily orbital scenes were stitched together, the project team set to work preparing cloud-free, 10-day composite data sets that included Normalized Difference Vegetation Index (NDVI) scores. The image processing steps involved included radiometric calibration, atmospheric correction, NDVI calculation, geometric correction, regional compositing, and projection of composited scenes. Each step is described briefly below.

RADIOMETRIC CALIBRATION

Radiometric calibration means defining the relationship between reflectance values recorded by a sensor from space and actual radiances measured with spectrometers on the ground. The accuracy of the AVHRR visible red and near-IR sensors degrade over time. Image analysts would not be able to produce useful time series of composite data sets unless reflectances were reliably calibrated. The project team relied on research that showed how AVHRR data acquired at different times could be normalized using a correction factor derived by analyzing reflectance values associated with homogeneous desert areas.

ATMOSPHERIC CORRECTION

Several atmospheric phenomena, including Rayleigh scatter, ozone, water vapor, and aerosols, were known to affect reflectances measured by sensors like AVHRR. Research yielded corrections to compensate for some of these.

One proven correction was for **Rayleigh scatter**. Named for an English physicist who worked in the early 20th century, Rayleigh scatter is the phenomenon that accounts for the fact that the sky appears blue. Short wavelengths of incoming solar radiation tend to be diffused by tiny particles in the atmosphere. Since blue wavelengths are the shortest in the visible band, they tend to be scattered more than green, red, and other colors of light. Rayleigh scatter is also the primary cause of atmospheric haze.

Because the AVHRR sensor scans such a wide swath, image analysts couldn't be satisfied with applying a constant haze compensation factor throughout entire scenes. To scan its 2400-km wide swath, the AVHRR sensor sweeps a scan head through an arc of 110°. Consequently, the viewing angle between the scan head and the Earth's surface varies from 0° in the middle of the swath to about 55° at the edges. Obviously the lengths of the paths traveled by reflected radiation toward the sensor vary considerably depending on the viewing angle. Project scientists had to take this into account when compensating for atmospheric haze. The farther a pixel was located from the center of a swath, the greater its path length, and the more haze needed to be compensated for. While they were at it, image analysts also factored in terrain elevation, since that too affects path length. ETOPO5, the most detailed global digital elevation model available at the time, was used to calculate path lengths adjusted for elevation. (You learned about the more detailed ETOPO1 in Chapter 7.)

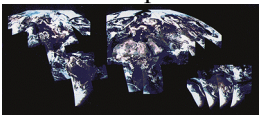
NDVI CALCULATION

The Normalized Difference Vegetation Index (NDVI) is the difference of near-IR and visible red reflectance values normalized over the sum of the two values. The result, calculated for every pixel in every daily orbital pass, is a value between -1.0 and 1.0, where 1.0 represents maximum photosynthetic activity, and thus maximum density and vigor of green vegetation.

GEOMETRIC CORRECTION AND PROJECTION

As you can see in the stitched orbital passes illustrated above, the widerange of view angles produced by the AVHRR sensor results in a great deal of geometric distortion. Relief displacement makes matters worse, distorting images even more towards the edges of each swath. The project team performed both orthorectification and rubber sheeting to rectify the data. The ETOPO5 global digital elevation model was again used to calculate corrections for scale distortions caused by relief displacement. To correct for distortions caused by the wide range of sensor view angles, analysts identified well-defined features like coastlines, lakeshores, and rivers in the imagery that could be matched to known locations on the ground. They derived coordinate transformation equations by analyzing differences between positions of control points in the imagery and known locations on the ground. The accuracy of control locations in the rectified imagery was shown to be no worse than 1,000 meters from actual locations. Equally important, the georegistration error between rectified daily orbital passes was shown to be less than one pixel.

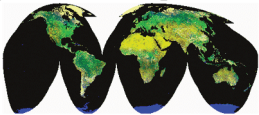
After the daily orbital passes were rectified, they were transformed into a map projection called Goode's Homolosine. This is an equal-area projection that minimizes shape distortion of land masses by interrupting the graticule over the oceans. The project team selected Goode's projection in part because they knew that equivalence of area would be a useful quality for spatial analysis. More importantly, the interrupted projection allowed the team to process the data set as twelve separate regions that could be spliced back together later. The illustration below shows the orbital passes for June 24, 1992 projected together in a single global image based on Goode's projection.



AVHRR orbital passes acquired on June 24, 1992, projected together in a Goode's Homolosine projection of the world. (Eidenshink & Faundeen, 1994).

COMPOSITING

Once the daily orbital passes for a ten-day period were rectified, every one-kilometer square pixel could be associated with corresponding pixels at the same location in other orbital passes. At this stage, with the orbital passes assembled into twelve regions derived from the interrupted Goode's projection, image analysts identified the highest NDVI value for each pixel in a given ten-day period. They then produced ten-day composite regions by combining all the maximum-value pixels into a single regional data set. This procedure minimized the chances that cloud-contaminated pixels would be included in the final composite data set. Finally, the composite regions were assembled into a single data set, illustrated below. This same procedure has been repeated to create 93 ten-day composites from April 1-10, 1992 to May 21-30, 1996.



Ten-day composite AVHRR image. The greenest pixels represent the highest NDVI values. (Eidenshink & Faundeen, 1994).

8.18. Image Classification

Along with military surveillance and weather forecasting, a common use of remotely sensed image data is to monitor land cover and to inform land use planning. The term **land cover** refers to the kinds of vegetation that blanket the Earth's surface, or the kinds of materials that form the surface where vegetation is absent. **Land use**, by contrast, refers to the functional roles that the land plays in human economic activities (Campbell, 1983).

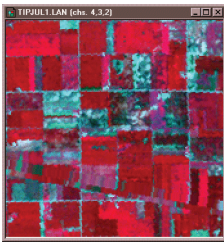
Both land use and land cover are specified in terms of generalized categories. For instance, an early

classification system adopted by a World Land Use Commission in 1949 consisted of nine primary categories, including settlements and associated non-agricultural lands, horticulture, tree and other perennial crops, cropland, improved permanent pasture, unimproved grazing land, woodlands, swamps and marshes, and unproductive land. Prior to the era of digital image processing, specially trained personnel drew land use maps by visually interpreting the shape, size, pattern, tone, texture, and shadows cast by features shown in aerial photographs. As you might imagine, this was an expensive, time-consuming process. It's not surprising then that the Commission appointed in 1949 failed in its attempt to produce a detailed global land use map.

Part of the appeal of digital image processing is the potential to automate land use and land cover mapping. To realize this potential, image analysts have developed a family of image classification techniques that automatically sort pixels with similar multispectral reflectance values into clusters that, ideally, correspond to functional land use and land cover categories. **Two general types of image classification techniques have been developed: supervised and unsupervised techniques.**

SUPERVISED CLASSIFICATION

Human image analysts play crucial roles in both supervised and unsupervised image classification procedures. In supervised classification, the analyst's role is to **specify in advance the multispectral reflectance or (in the case of the thermal infrared band) emittance values typical of each land use or land cover class.**



Portion of Landsat TM scene acquired July 17, 1986 showing agricultural fields in Tippecanoe County, Indiana. Reflectances recorded in TM bands 2 (visible green), 3 (visible red), and 4 (near-infrared) are shown in blue, green, and red respectively. Multispec image processing software © 2001 Purdue Research Foundation, Inc.

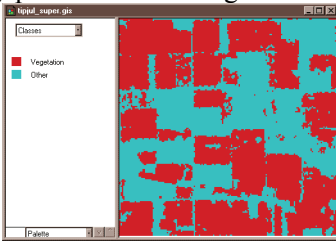
For instance, to perform a supervised classification of the Landsat Thematic Mapper (TM) data shown above into two land cover categories, Vegetation and Other, you would first delineate several training fields that are representative of each land cover class. The illustration below shows two training fields for each class, however, to achieve the most reliable classification possible, you would define as many as 100 or more training fields per class.



Training fields defined for two classes of land cover, vegetation and other. Multispec image processing software © 2001 Purdue Research Foundation, Inc.

The training fields you defined consist of clusters of pixels with similar reflectance or emittance values. If you did a good job in supervising the training stage of the classification, each cluster would represent the range of spectral characteristics exhibited by its corresponding land cover class. Once the clusters are defined, you would apply a classification algorithm to sort the remaining pixels in the scene into the class with the most similar spectral characteristics. One of the most commonly used algorithms computes the statistical probability that each pixel belongs to each class. Pixels are then assigned to the class associated with the highest probability. Algorithms of

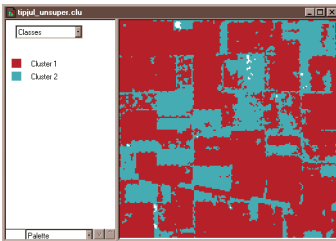
this kind are known as **maximum likelihood classifiers**. The result is an image like the one shown below, in which every pixel has been assigned to one of two land cover classes.



Two-class land cover map produced by supervised classification of Landsat TM data. Multispec image processing software © 2001 Purdue Research Foundation, Inc.

UNSUPERVISED CLASSIFICATION

The image analyst plays a different role in unsupervised classification. They do not define training fields for each land cover class in advance. Instead, they rely on one of a family of statistical clustering algorithms to sort pixels into distinct spectral classes. Analysts may or may not even specify the number of classes in advance. Their responsibility is to **determine the correspondences between the spectral classes that the algorithm defines and the functional land use and land cover categories** established by agencies like the U.S. Geological Survey. The example that follows outlines how unsupervised classification contributes to the creation of a high-resolution national land cover data set.



Two-class land cover map produced by unsupervised classification of Landsat TM data. Multispec image processing software © 2001 Purdue Research Foundation, Inc.

8.19. Classifying Landsat Data for the National Land Cover Dataset

The USGS developed one of the first land use/land cover classifications systems designed specifically for use with remotely sensed imagery. The **Anderson Land Use/Land Cover Classification** system, named for the former Chief Geographer of the USGS who led the team that developed the system, consists of nine land cover categories (urban or built-up; agricultural; range; forest; water; wetland; barren; tundra; and perennial snow and ice), and 37 subcategories (for example, varieties of agricultural land include cropland and pasture; orchards, groves, vineyards, nurseries, and ornamental horticulture; confined feeding operations; and other agricultural land). Image analysts at the U. S. Geological Survey created the USGS Land Use and Land Cover (LULC) data by manually outlining and coding areas on air photos that appeared to have homogeneous land cover that corresponded to one of the Anderson classes.

The LULC data were compiled for use at 1:250,000 and 1:100,000 scales. Analysts drew outlines of land cover polygons onto vertical aerial photographs. Later, the outlines were transferred to transparent film georegistered with small-scale topographic base maps. The small map scales kept the task from taking too long and costing too much, but also forced analysts to generalize the land cover polygons quite a lot. The smallest man-made features encoded in the LULC data are four hectares (ten acres) in size, and at least 200 meters (660 feet) wide at their narrowest

point. The smallest non-man-made features are sixteen hectares (40 acres) in size, with a minimum width of 400 meters (1320 feet). Smaller features were aggregated into larger ones. After the land cover polygons were drawn onto paper and georegistered with topographic base maps, they were digitized as vector features, and attributed with land cover codes. A rasterized version of the LULC data was produced later.

For more information, visit the [USGS' LULC home page](#).

The successor to LULC is the USGS's National Land Cover Data (NLCD). Unlike LULC, which originated as a vector data set in which the smallest features are about ten acres in size, **NLCD is a raster data set with a spatial resolution of 30 meters** (i.e., pixels represent about 900 square meters on the ground) **derived from Landsat TM imagery**. The steps involved in producing the NLCD include preprocessing, classification, and accuracy assessment, each of which is described briefly below.

PREPROCESSING

The first version of NLCD—NLCD 92—was produced for subsets of ten federal regions that make up the conterminous United States. The primary source data were bands 3, 4, 5, and 7 (visible red, near-infrared, mid-infrared, and thermal infrared) of cloud-free Landsat TM scenes acquired during the spring and fall (when trees are mostly bare of leaves) of 1992. Selected scenes were geometrically and radiometrically corrected, then combined into sub-regional mosaics comprised of no more than 18 scenes. Mosaics were then projected to the same **Albers Conic Equal Area projection** (with standard parallels at 29.5° and 45.5° North Latitude, and central meridian at 96° West Longitude) based upon the **NAD83** horizontal datum.

IMAGE CLASSIFICATION

An **unsupervised classification** algorithm was applied to the preprocessed mosaics to generate 100 spectrally distinct pixel clusters. Using aerial photographs and other references, image analysts at USGS then assigned each cluster to one of the classes in a modified version of the Anderson classification scheme. Considerable interpretation was required, since not all functional classes have unique spectral response patterns.

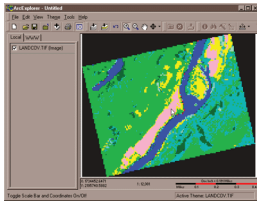
Modified Anderson Land Use/Land Cover Classification

Level I Classes	Level II Classes	
Water	11	Open Water
	12	Perennial Ice/Snow
Developed	21	Low Intensity Residential
	22	High Intensity Residential
	23	Commercial/Industrial/Transportation
Barren	31	Bare Rock/Sand/Clay
	32	Quarries/Strip Mines/Gravel Pits
	33	Transitional
Forrested Upland	41	Deciduous Forest
	42	Evergreen Forest
	43	Mixed Forest
Shrubland	51	Shrubland
Non-Natural Woody	61	Orchards/Vineyards/Other
Herbaceous Upland Natural/Semi-natural Vegetation	71	Grasslands/Herbaceous
Herbaceous Planted/Cultivated	81	Pasture/Hay
	82	Row Crops
	83	Small Grains
	84	Fallow
	85	Urban/Recreational Grasses
Wetlands	91	Woody Wetlands
	92	Emergent Herbaceous Wetlands

Modified Anderson Land Use/Land Cover Classification used for the USGS National Land Cover Dataset.

ACCURACY ASSESSMENT

The USGS hired private sector vendors to assess the classification accuracy of the NLCD 92 by checking randomly sampled pixels against manually interpreted aerial photographs. Results from the first four completed regions suggested that the likelihood that a given pixel is correctly classified ranges from only **38 to 62 percent**. Much of the classification error was found to occur among the Level II classes that make up the various Level I classes, and some classes were much more error-prone than others. **USGS encourages users to aggregate the data into 3 x 3 or 5 x 5 pixel blocks** (in other words, to decrease spatial resolution from 30 meters to 90 or 150 meters), or to aggregate the 21 Level II classes into the nine Level I classes. Even in the current era of high-resolution satellite imaging and sophisticated image processing techniques, there is still no cheap and easy way to produce detailed, accurate geographic data.



An extract from NLCD 92 that corresponds to the same portion of the Bushkill, PA quadrangle mapped in other USGS data files provided with earlier chapters. The data viewer is ESRI's ArcExplorer version 2.

Color Key	RGB Value	Class Number and Name
[Blue]	0, 0, 255	11 Open Water
[White]	255, 255, 255	12 Permanent Snow
[Yellow]	255, 254, 0	21 Low Intensity Residential
[Orange]	255, 163, 0	22 High Intensity Residential
[Red]	255, 0, 0	23 Commercial/Industrial/Transportation
[Light Green]	225, 225, 204	31 Bare Rock/Sand/Clay
[Light Blue]	152, 171, 0	32 Clearwater/Deep Mineral/Gravel Pit
[Dark Green]	255, 0, 255	33 Transitional
[Light Green]	0, 178, 0	41 Deciduous Forest
[Dark Green]	0, 102, 0	42 Evergreen Forest
[Light Green]	0, 178, 178	43 Mixed Forest
[Light Green]	178, 178, 0	51 Shrubland
[Purple]	153, 25, 229	61 Orchards/Vineyards
[Light Green]	225, 254, 153	71 Grasslands/Forage Pasture
[Light Green]	255, 255, 0	81 Pasture/Hay
[Light Green]	255, 179, 204	82 Pasture
[Light Green]	204, 17, 255	83 Small Grains
[Light Green]	178, 225, 178	84 Barrow
[Light Green]	126, 255, 0	85 Urban/Recreational Grasses
[Light Green]	126, 255, 204	91 Woody Wetlands
[Light Green]	0, 255, 255	92 Emergent Herbaceous Wetlands

Map legend for the National Land Cover Dataset.

For more information about NLCD 92 and its successor, **NLCD 2001**, visit the [Multi-Resolution Land Characteristics Consortium](#).

GLOBAL LAND COVER DATA

Meanwhile, the U.S. National Geospatial-Intelligence Agency (formerly the National Imagery and Mapping Agency) hired a company called Earthsat (Now MDA Federal) to produce a 120-meter resolution, 13-class land use / land cover data set for the rest of the world. This product, called **Geocover LC**, is derived from 30-meter Landsat TM imagery from the 1990s and 2000. A team of image analysts assigned 240 clusters produced by **unsupervised classification** into the thirteen land use/ land cover classes (Barrios, 2001). For more information about Geocover LC, visit [MDA Information Systems](#) and [GeoCover LC](#).

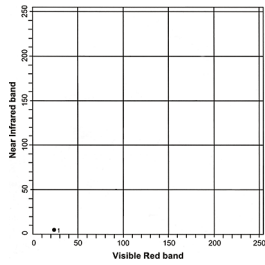
8.20. Unsupervised Classification By Hand

TRY THIS!

This activity guides you through a simulated **unsupervised classification** of remotely sensed image data to create a land cover map. Begin by viewing and printing the [Image Classification Activity PDF file](#).

1. Plot the reflectance values.

The two grids on the top of the second page of the PDF file represent reflectance values in the visible red and near infrared wavelength bands measured by a remote sensing instrument for a parcel of land. Using the graph (like the one below) on the first page of the PDF file you printed, **plot the reflectance values for each pixel and write the number of each pixel (1 to 36) next to its location in the graph**. Pixel 1 has been plotted for you (Visible Red band = 22, Near Infrared band = 6).



- After you have completed filling in the graph you can check your progress by following this link to a [completed graph](#).

2. Identify four land cover classes.

Looking at the completed plot from step one, **identify and circle four clusters (classes) of pixels**. Label these four classes A, B, C, and D.

- After you have circled the four clusters of pixels in the graph you can check your progress by following this link to [view the identified clusters](#).
- **Note:** You may have labeled the clusters differently but the four clusters should contain the same points, more or less, as the example.

3. Complete the land cover map grid.

Using the clusters you identified in the previous step, fill in the land cover map grid with the letter that represents the land use class in which each pixel belongs. The result is a classified image.

Land Cover Map

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

- If you would like to check your land cover map you can follow this link to [view a completed land cover map](#).
- **Note:** If you labeled the clusters in step two differently than the example, your land cover map may look different but the patterns should look similar.

4. Complete a legend that explains the association.

Using the spectral response data provided on the second page of the PDF file, associate each of the four classes with a land use class.

A =	B =
C =	D =

- You can check your results by following this link to [view a completed legend](#).
- **Note:** Depending on how you labeled your clusters, your legend may differ from the example legend, however, when you apply your legend to your Land Cover Map the land use classes on your map should match the example map.

You have now completed the **unsupervised classification activity** in which you used remotely sensed image data to create a land cover map.

PRACTICE QUIZ

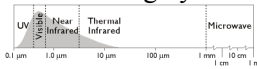
Registered Penn State students should return now to the Chapter 8 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **Image Processing**.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

8.21. Microwave Data

The remote sensing systems you've studied so far are sensitive to the visible, near-infrared, and thermal infrared bands of the electromagnetic spectrum, wavelengths at which the magnitude of solar radiation is greatest. IKONOS, AVHRR, and the Landsat MSS, TM, and ETM+ instruments are all **passive sensors** that only measure radiation emitted by other objects.

There are two main shortcomings to passive sensing of the visible and infrared bands. First, clouds interfere with both incoming and outgoing radiation at these wavelengths. Secondly, reflected visible and near-infrared radiation can only be measured during daylight hours. This is why the AVHRR sensor only produces visible and near-infrared imagery of the entire Earth once a day, although it is capable of two daily scans.



The electromagnetic spectrum divided into five wavelength bands. (Adapted from Lillesand & Kiefer, 1994).

Longwave radiation, or microwaves, are made up of wavelengths between about one millimeter and one meter. Microwaves can penetrate clouds, but the Sun and Earth emit so little longwave radiation that it can't be measured easily from space. **Active remote sensing systems** solve this problem. Active sensors like those aboard the European Space Agency's ERS satellites, the Japanese JERS satellites, and the Canadian Radarsat, among others, transmit pulses of longwave radiation, then measure the intensity and travel time of those pulses after they are reflected back to space from the Earth's surface. Microwave sensing is unaffected by cloud cover, and can operate day or night. Both image data and elevation data can be produced by microwave sensing, as you will discover in the sections on imaging radar and radar altimetry that follow.

8.22. Imaging Radar

One example of active remote sensing that everyone has heard of is radar, which stands for RAdio Detection And Ranging. Radar was developed as an air defense system during World War II and is now the primary remote sensing system air traffic controllers use to track the 40,000 daily aircraft takeoffs and landings in the U.S. Radar antennas alternately transmit and receive pulses of microwave energy. Since both the magnitude of energy transmitted and its velocity (the speed of light) are known, radar systems are able to record either the intensity or the round-trip distance traveled of pulses reflected back to the sensor. **Systems that record pulse intensity are called imaging radars.**

In addition to its indispensable role in navigation, radar is also an important source of raster image data about the Earth's surface. Radar images look the way they do because of the different ways that objects reflect microwave energy. In general, rough-textured objects reflect more energy back to the sensor than smooth objects. Smooth objects, such as water bodies, are highly reflective, but unless they are perpendicular to the direction of the incoming pulse, the reflected energy all bounces off at an angle and never returns to the sensor. Rough surfaces, such as

vegetated agricultural fields, tend to scatter the pulse in many directions, increasing the chance that some back scatter will return to the sensor.



Radar image showing areas inundated by a flood of the Mississippi River in 1993. © 2001 Space Imaging, Inc. All Rights Reserved. Used by permission.

The imaging radar aboard the European Resource Satellite (ERS-1) produced the data used to create the image shown above. The smooth surface of the flooded Mississippi River deflected the radio signal away from the sensor, while the surrounding rougher-textured land cover reflected larger portions of the radar pulse. The lighter an object appears in the image, the more energy it reflected. Imaging radar can be used to monitor flood extents regardless of weather conditions. Passive instruments like Landsat MSS and TM that are sensitive only to shorter wavelengths are useless as long as cloud-covered skies prevail.

8.23. Radar Altimetry

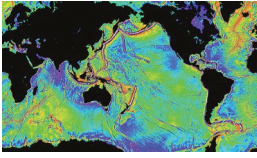
Altimetry is the measurement of elevation. Earlier chapters discussed land survey methods used to calculate terrain elevations in the field (leveling and GPS), and photogrammetric methods used to measure terrain elevations from stereoscopic images produced from pairs of aerial photographs. Land surveys and photogrammetric surveys yield high quality elevation data, but they are also time-consuming and expensive to conduct.

Radar (and laser) altimetry provides more efficient solutions when elevation data are needed for larger areas. For example, you have heard about the **Shuttle Radar Topography Mission (SRTM)**, which used dual radar altimeters to produce 30-meter elevation data as well as stereoscopic terrain imagery for the Earth's land surface between 60° North and South latitude. Next we'll consider how radar altimetry has been used to produce a global seafloor elevation data set.

PREDICTING SEAFLOOR DEPTHS

Detailed maps of the Earth's bathymetry (the topography of the ocean floor) are needed to study plate tectonics, to locate potential offshore oil and mineral deposits, and to route undersea telecommunications cables, among other things. Coarse global data sets (such as ETOPO2, with its 2-minute grid resolution) are inadequate for such purposes. Slow-moving surface vessels equipped with sonar instruments have mapped only a small fraction of the Earth's bathymetry.

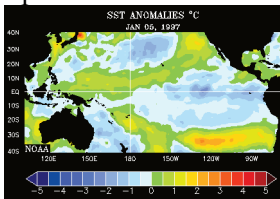
Data produced by radar sensors like ERS-1 have been used to produce global seafloor elevation data. **Radar pulses cannot penetrate the deep ocean, but they can be used to accurately measure the height of the sea surface relative to a global ellipsoid such as WGS 84.** As you know, the geoid is defined as mean sea level adjusted to account for the effects of gravity. Geodesists invent reference ellipsoids like WGS 84 to approximate the geoid's shape with a figure that is easier to define mathematically. Because gravity varies with mass, the geoid bulges slightly above the ellipsoid over seamounts and undersea volcanoes, which often rise 2000 meters or more above the ocean floor. Sea surface elevation data produced by satellite altimeters can thus be used to predict fairly detailed bathymetry, as shown in the map below.



Global bathymetry predicted from sea surface elevations measured by the ERS-1 radar sensing system. The predicted bathymetry reveals seamounts and undersea volcanoes greater than 1000 meters in elevation, more than half of which had not previously been charted. (Sandwell & Smith, 1998).

8.24. Using Radar Altimetry to Monitor El Niño

Remotely sensed data play crucial roles in global change research. One example is data produced by radar altimetry, which are used to monitor the global climatological phenomenon called El Niño. El Niño is the colloquial name for a periodic weakening of the trade winds and warming of the surface layers of the eastern and central equatorial Pacific Ocean. El Niño events occur irregularly at intervals of 2-7 years, and typically last 12-18 months. Poorly understood teleconnections between the ocean and the atmosphere result in altered regional precipitation patterns that sometimes include severe floods. The animation below illustrates the changes in sea surface temperatures that accompanied the 1997-98 El Niño event.



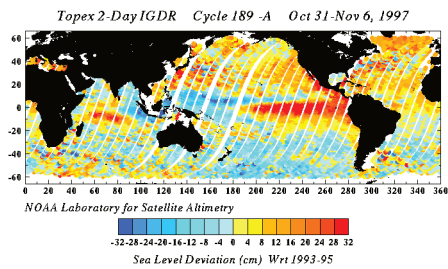
Sea surface temperature anomalies from January 1997 through March 1998, estimated from on-site measurements and sea surface elevation data produced by the TOPEX-POSEIDON satellite altimeter. *NOTE: To replay the animation, hold down the Shift key and reload the page.* (NOAA-CRIES Climate Diagnostics Center, 2006).

Active remote sensing provides an alternative to the expensive and tedious business of measuring sea surface temperatures directly at many locations throughout the Pacific Ocean.



Researchers collect data from one of the 70 buoys in the Tropical Atmosphere Ocean (TAO) array. TAO buoys are equipped with sea surface temperature monitoring instruments. Hourly observations are stored in instrument memory and must be retrieved by operators. (Tropical Atmosphere Ocean Project, n. d.).

From 1992 through 2005, the TOPEX/POSEIDON radar altimeter measured heights of the ocean surface with centimeter accuracy. The sensor transmitted and received longwave energy at 6 km intervals along ground tracks spaced 315 km apart. The satellite that carried the sensor completed a polar orbit every 112 minutes at an altitude of 1,335 km, passing over the same point every 10 days. Sea level deviations (differences between the geoid and measured sea level) were determined from measurements of the height of the ocean surface relative to the sensor, which is calculated from the time difference between the transmission and return of signals reflected from the ocean surface. Sea surface temperature can be inferred from deviations of sea surface heights relative to long-term mean values.



Deviations of sea surface elevations from long-term averages, measured by the TOPEX-POSEIDON radar altimeter. Sea level deviations are used as a surrogate measure of the warming of the eastern equatorial Pacific Ocean that indicates an El Niño event. The S-shaped curves trace the path of the satellite's polar orbit as the Earth rotates beneath it. (NASA, Jet Propulsion Laboratory, 2006).

PRACTICE QUIZ

Registered Penn State students should return now to the Chapter 8 folder in ANGEL (via the Resources menu to the left) to take a self-assessment quiz about **Microwave Data**.

You may take practice quizzes as many times as you wish. They are not scored and do not affect your grade in any way.

8.25. Summary

Most remotely sensed data are derived from measurements of electromagnetic radiation. Aerial photographs are analog forms of data that record intensities of electromagnetic radiation within the visible or near-infrared wavelength bands. Digital sensing systems extend the spectral sensitivity of photographic film far beyond the visible band, enabling users to map phenomena that are otherwise invisible. Because many objects exhibit unique spectral response characteristics across a range of wavelengths, multispectral sensing offers the potential to identify features of interest automatically. Recognizing this potential, analysts in many fields have adopted land remote sensing data for such diverse applications as land use and land cover mapping, geological resource exploration, precision farming, archeological investigations, and even validating the computational models used to predict global environmental change. Once the exclusive domain of government agencies, an industry survey suggests that the gross revenue earned by private land remote sensing firms exceeded \$2.4 billion (U.S.) in 2001 (ASPRS, 2004).

The fact that remote sensing is first and foremost a surveillance technology deployed by government agencies cannot be overlooked. State-of-the-art spy satellites are rumored to be able to detect objects as small as six inches wide. Meanwhile, GeoEye and other private firms have been licensed to build and launch half-meter panchromatic sensors. As the resolution of remotely sensed imagery increases, and its price decreases, privacy concerns are bound to proliferate. For example, remotely sensed data were pivotal in the case of an Arizona farmer who was fined for growing cotton illegally (Kerber, 1998). Was the farmer right to claim that remote sensing constituted unreasonable search? More serious, perhaps, is the potential impact of the remote sensing industry on defense policy of the United States and other countries. In light of an expected \$500 billion investment in commercial satellites (including communications satellites as well as land remote sensing systems) by 2010, some analysts believe that “the military will be called upon to defend American interests in space much as navies were formed to protect sea commerce in the 1700s” (Newman, 1999).

While the ethical implications of remote sensing technologies must not be ignored, neither should their potential to help us to become more knowledgeable, and thus more effective stewards of our home planet. Several challenges must be addressed before remote sensing can fulfill this potential. One is the need to produce cost-effective, high-resolution data suitable for local scale mapping—the scale at which most land use decisions are made. Another is the need to develop more sophisticated image processing algorithms that will enable analysts

to extract vector features from raster source data with minimal intervention. Yet another challenge is to develop automated classification techniques to help derive meaningful patterns from the data produced by a new generation of **hyperspectral** scanners—sensing systems that measure reflected and emitted radiation across 200 or more narrow wavelength bands simultaneously.

QUIZ

Registered Penn State students should return now to the Chapter 8 folder in ANGEL (via the Resources menu to the left) to take a graded quiz for this chapter.

This one counts. **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter's learning objectives. You are free to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to our Chapter 8 Discussion Forum, you will have completed Chapter 8.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under “Post new comment” and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the “Preview” or “Save” button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its “title” in the thread.

8.26. Bibliography

American Society for Photogrammetry and Remote Sensing (2004). ASPRS/NASA ten-year industry forecast. Photogrammetric Engineering and Remote Sensing, 70:1 Originally retrieved March 2, 2008 from <http://www.asprs.org/news/forecast/> (expired). Retrieved November 30, 2011 from <http://www.asprs.org/a/news/forecast/10-year-ind-forecast-exec-summary.pdf>

California Institute of Technology. (2002). *ASTER spectral library*. Retrieved June 3, 2001 from <http://speclib.jpl.nasa.gov>

Campbell, J. B. (1983). *Mapping the land: Aerial imagery for land use information*. Washington, D.C.: Association of American Geographers.

Colucci, J. A. (1998). Uncorking technology: Better wine through remote sensing. *EOM*, May, pp. 32-35.

Cowen, D. J. & Jensen, J. R. (1998). Extraction and Modeling of Urban Attributes Using Remote Sensing Technology. In D. Liverman, E. F. Moran, R. R. Rindfuss & P. C. Stern (Eds.), *People and Pixels: Linking Remote Sensing and Social Science* (pp. 164-188). Washington, D. C.: National Academy Press, National Research Council.

Eidenshink, J. C., & Faundeen, J. L. (1994). The 1-km AVHRR global land data set: First stages in implementation. *International Journal of Remote Sensing*, 15, pp. 3443-3462.

GeoEye (2007). GeoEye imagery products: GeoEye 1. Retrieved March 1, 2008, from <http://www.geoeye.com/products/imagery/geoeye1/default.htm>(expired, or moved to <http://launch.geoeye.com/LaunchSite/about/>)

Geophysical Data Center, National Oceanic and Atmospheric Administration (2005). *EOG Defense*

Meteorological Satellite Program (DMSPP). Retrieved June 3, 2001, from <http://www.ngdc.noaa.gov/dmsp/dmsp.html>

Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective*. Upper Saddle River, N.J.: Prentice Hall.

Jet Propulsion Laboratory. (2006). *Ocean surface topography from space*. National Aeronautics and Space Administration. Retrieved June 20, 2006, from <http://sealevel.jpl.nasa.gov/>

Kerber, R. (1998). When is a satellite photo an unreasonable search? *Wall Street Journal*, January 27, 1998.

Lillesand, T. & Kiefer, R. (1994). *Remote sensing and image interpretation* (3rd ed.). New York, NY: John Wiley and Sons.

Newman, R. J. (1999). The new space race. *U.S. News and World Report*, November 8, pp. 30-38.

National Aeronautics & Space Administration (2001). The Landsat Program. Retrieved June 3, 2001, from <http://landsat.gsfc.nasa.gov>

NOAA-CIRES Climate Diagnostics Center. (2006). Retrieved June 20, 2006, from http://www.cdc.noaa.gov/map/clim/sst_olr/sst_anim.shtml

Sandwell, D. T. & Smith, W. H. F. (1998). *Exploring the ocean basins with satellite altimeter data*. Retrieved December 18, 1998, from www.ngdc.noaa.gov/mgg/bathymetry/predicted/explore.HTML

Steinwand, D.R. (1994). Mapping raster imagery to the interrupted goode homolosine projection. *International Journal of Remote Sensing*, 15:17, pp. 3463-3471.

Tropical Atmosphere Ocean Project. (n. d.). *TAO buoy photo gallery*. Retrieved June 20, 2006, from http://www.pmel.noaa.gov/tao/proj_over/diagrams/buoy.html

United States Geological Survey (2001). *Earthshots: satellite images of environmental change*. Retrieved March 6, 2001, from <http://earthshots.usgs.gov/>

United States Geological Survey. (2005). *Global Land 1-Km AVHRR Project*. Retrieved May 31, 2006, from <http://edcsns17.cr.usgs.gov/1KM/>

United States Geological Survey (2011). *EarthNow! Landsat Image Viewer*. Retrieved November 30, 2011 from <http://earthnow.usgs.gov/>

Chapter 9

Integrating Geographic Data

David DiBiase

9.1. Overview

Geographic data are expensive to produce and maintain. Data often accounts for the lion's share of the cost of building and running geographic information systems. The expense of GIS is justifiable when it gives people the information they need to make wise choices in the face of complex problems. In this chapter we'll consider one such problem: the search for suitable and acceptable sites for low level radioactive waste disposal facilities. Two case studies will demonstrate that GIS is very useful indeed for assimilating the many site suitability criteria that must be taken into account, provided that the necessary data can be assembled in a single, integrated system. The case studies will allow us to compare vector and raster approaches to site selection problems.

The ability to integrate diverse geographic data is a hallmark of mature GIS software. The know-how required to accomplish data integration is also the mark of a truly knowledgeable GIS user. What knowledgeable users also recognize, however, is that while GIS technology is well suited to answering certain well defined questions, it often cannot help resolve crucial conflicts between private and public interests. The objective of this final, brief chapter is to consider the challenges involved in using GIS to address a complex problem that has both environmental and social dimensions. Specifically, in this chapter you will learn to:

Objectives

Chapter 9 should help prepare you to:

- Recognize the characteristics of geographic data that must be taken into account to overlay multiple data layers;
- Compare and contrast vector and raster approaches to site suitability studies;
- Have realistic expectations about what geographic data analysis can achieve.

Comments and Questions?

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

9.2. Checklist

The following checklist is for Penn State students who are registered for classes in which this text, and associated quizzes and projects in the ANGEL course management system, have been assigned. You may find it useful to print this page out first so that you can follow along with the directions.

Chapter 9 Checklist (for registered students only)

Step	Activity	Access/Directions
1	Read Chapter 9	This is the second page of the Chapter. Click on the links at the bottom of the page to continue or to return to the previous page, or to go to the top of the chapter. You can also navigate the text via the links in the GEOG 482 menu on the left.
2	Chapter 9 includes no practice quizzes.	
3	Perform “Try this” activities including: <ul style="list-style-type: none"> • Use Global Mapper / dlgv32 Pro software to create a slope map “Try this” activities are not graded.	Instructions are provided for each activity.
4	Submit the Chapter 9 Graded Quiz	ANGEL > [your course section] > Lessons tab > Chapter 9 folder > Chapter 9 Graded Quiz. See the Calendar tab in ANGEL for due dates.
5	Read comments and questions posted by fellow students. Add comments and questions of your own, if any.	Comments and questions may be posted on any page of the text, or in a Chapter-specific discussion forum in ANGEL.

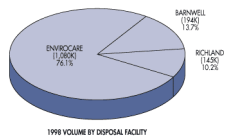
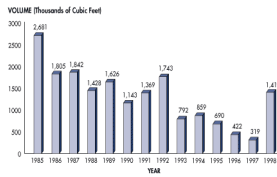
9.3. Context

This section sets a context for two case studies that follow. First, I will briefly define low level radioactive waste (LLRW). Then I discuss the legislation that mandated construction of a dozen or more regional LLRW disposal facilities in the U.S. Finally, I will reflect briefly on how the capability of geographic information systems to integrate multiple data “layers” is useful for siting problems like the ones posed by LLRW.

9.4. Low Level Radioactive Waste

According to the U.S. Nuclear Regulatory Commission (2004), LLRW consists of discarded items that have become contaminated with radioactive material or have become radioactive through exposure to neutron radiation. Trash, protective clothing, and used laboratory glassware make up all but about 3 percent of LLRW. These “Class A” wastes remain hazardous less than 100 years. “Class B” wastes, consisting of water purification filters and ion exchange resins used to clean contaminated water at nuclear power plants, remain hazardous up to 300 years. “Class C” wastes, such as metal parts of decommissioned nuclear reactors, constitute less than 1 percent of all LLRW, but remain dangerous for up to 500 years.

The danger of exposure to LLRW varies widely according to the types and concentration of radioactive material contained in the waste. Low level waste containing some radioactive materials used in medical research, for example, is not particularly hazardous unless inhaled or consumed, and a person can stand near it without shielding. On the other hand, exposure to LLRW contaminated by processing water at a reactor can lead to death or an increased risk of cancer (U.S. Nuclear Regulatory Commission, n.d.).



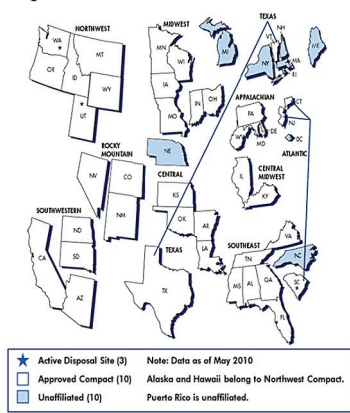
Note: Values are rounded to the nearest thousand cubic feet and percentages are rounded to the nearest tenth of a percent.
 Source: DOE 1998 Data by State Assessment of Low-Level Radioactive Wastes Received at Commercial Disposal Sites (DOE/CLW-25), pages A2 and A3.

Production trends and destinations of low level radioactive waste. (U.S. Nuclear Regulatory Commission, 2005).

Hundreds of nuclear facilities across the country produce LLRW, but only a very few disposal sites are currently willing to store it. Disposal facilities at Clive, Utah, Barnwell, South Carolina, and Richland, Washington, accepted over 4,000,000 cubic feet of LLRW in both 2005 and 2006, up from 1,419,000 cubic feet in 1998. By 2008 the volume had dropped to just over 2,000,000 cubic feet (U.S. Nuclear Regulatory Commission, 2011a). Sources include nuclear reactors, industrial users, government sources (other than nuclear weapons sites), and academic and medical facilities. (We have a small nuclear reactor here at Penn State that is used by students in graduate and undergraduate nuclear engineering classes.)

9.5. Siting LLRW Storage Facilities

The U.S. Congress passed the Low Level Radioactive Waste Policy Act in 1980. As amended in 1985, the Act made states responsible for disposing of the LLRW they produce. States were encouraged to form regional “compacts” to share the costs of locating, constructing, and maintaining LLRW disposal facilities. The intent of the legislation was to avoid the very situation that has since come to pass, that the entire country would become dependent on a very few disposal facilities.



Regional compacts formed by states in response to the LLRW Policy Act (U.S. Nuclear Regulatory Commission, 2011b).

State government agencies and the consultants they hire to help select suitable sites assume that few if any

municipalities would volunteer to host a LLRW disposal facility. They prepare for worst-case scenarios in which states would be forced to exercise their right of eminent domain to purchase suitable properties without the consent of landowners or their neighbors. GIS seems to offer an impartial, scientific, and therefore defensible approach to the problem. As Mark Monmonier has written, “[w]e have to put the damned thing somewhere, the planners argue, and a formal system of map analysis offers an ‘objective,’ logical method for evaluating plausible locations” (Monmonier, 1995, p. 220). As we discussed in our very first chapter, site selection problems pose a geographic question that geographic information systems are well suited to address, namely, which locations have attributes that satisfy all suitability criteria?

9.6. Map Overlay Concept

Environmental scientists and engineers consider many geological, climatological, hydrological, and surface and subsurface land use criteria to determine whether a plot of land is suitable or unsuitable for a LLRW facility. Each criterion can be represented with geographic data, and visualized as a thematic map. In theory, the site selection problem is as simple as compiling onto a single map all the disqualified areas on the individual maps, and then choosing among whatever qualified locations remain. In practice, of course, it is not so simple.

There is nothing new about superimposing multiple thematic maps to reveal optimal locations. One of the earliest and most eloquent descriptions of the process was written by Ian McHarg, a landscape architect and planner, in his influential book *Design With Nature*. In a passage describing the process he and his colleagues used to determine the least destructive route for a new roadway, McHarg (1971) wrote:

“

...let us map physiographic factors so that the darker the tone, the greater the cost. Let us similarly map social values so that the darker the tone, the higher the value. Let us make the maps transparent. When these are superimposed, the least-social-cost areas are revealed by the lightest tone. (p. 34).

“

As you probably know, this process has become known as **map overlay**. Storing digital data in multiple “layers” is not unique to GIS, of course; computer-aided design (CAD) packages and even spreadsheets also support layering. What’s unique about GIS, and important about map overlay, is its ability to generate a new data layer as a product of existing layers. In the example illustrated below, for example, analysts at Penn State’s Environmental Resources Research Institute estimated the agricultural pollution potential of every major watershed in the state by overlaying watershed boundaries, the slope of the terrain (calculated from USGS DEMs), soil types (from U.S. Soil Conservation Service data), land use patterns (from the USGS LULC data), and animal loading (livestock wastes estimated from the U.S. Census Bureau’s Census of Agriculture).

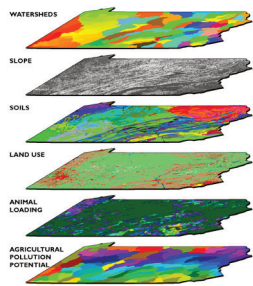
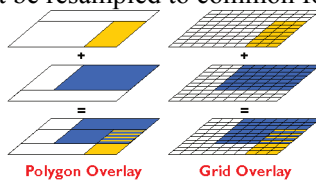


Diagram illustrating the map overlay process used to evaluate potential agricultural pollution by watershed in Pennsylvania.

As illustrated below, map overlay can be implemented in either vector or raster systems. In the vector case, often referred to as **polygon overlay**, the intersection of two or more data layers produces new features (polygons). Attributes (symbolized as colors in the illustration) of intersecting polygons are combined. The raster implementation (known as **grid overlay**) combines attributes within grid cells that align exactly. Misaligned grids must be resampled to common formats.



Map overlay is a procedure for combining the attributes of intersecting features that are represented in two or more georegistered data layers.

Polygon and grid overlay procedures produce useful information only if they are performed on data layers that are properly **georegistered**. Data layers must be referenced to the same coordinate system (e.g., the same UTM and SPC zones), the same map projection (if any), and the same datum (horizontal and vertical, based upon the same reference ellipsoid). Furthermore, locations must be specified with coordinates that share the same unit of measure.

9.7. Pennsylvania Case Study

In response to the LLRW Policy Act, Pennsylvania entered into an “Appalachian Compact” with the states of Delaware, Maryland, and West Virginia to share the costs of siting, building, and operating a LLRW storage facility. Together, these states generated about 10 percent of the total volume of LLRW then produced in the U.S. Pennsylvania, which generated about 70 percent of the total produced by the Appalachian Compact, agreed to host the disposal site.

In 1990, the Pennsylvania Department of Environmental Protection commissioned Chem-Nuclear Systems Incorporated (CNSI) to identify three potentially suitable sites to accommodate two to three truckloads of LLRW per day for 30 years. CNSI, the operator of the Barnwell South Carolina site, would also operate the Pennsylvania site for profit.



Sketch of the proposed Pennsylvania LLRW disposal facility (Pennsylvania Department of Environmental Protection, 1998).

CNSI’s plan called for storing LLRW in 55-gallon drums encased in concrete, buried in clay, surrounded by a polyethylene membrane. The disposal facilities, along with support and administration buildings and a visitors center, would occupy about 50 acres in the center of a 500-acre site. (Can you imagine a family outing to the Visitors

Center of a LLRW disposal facility?) The remaining 450 acres would be reserved for a 500 to 1000 foot wide buffer zone.

The three stage siting process agreed to by CNSI and the Pennsylvania Department of Environmental Protection corresponded to three scales of analysis: statewide, regional, and local. All three stages relied on vector geographic data integrated within a GIS.

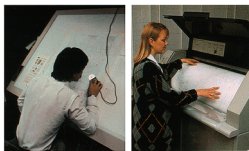
9.8. Vector Approach

CNSI and its subcontractors adopted a vector approach for its GIS-based site selection process. When the process began in 1990, far less geographic data was available in digital form than it is today. Most of the necessary data was available only as paper maps, which had to be converted to digital form. In one of its interim reports, CNSI described two digitizing procedures used, “digitizing” and “scanning.” Here’s how it described “digitizing:”

“

In the digitizing process, a GIS operator uses a hand-held device, known as a cursor, to trace the boundaries of selected disqualifying features while the source map is attached to a digitizing table. The digitizing table contains a fine grid of sensitive wire imbedded within the table top. This grid allows the attached computer to detect the position of the cursor so that the system can build an electronic map during the tracing. In this project, source maps and GIS-produced maps were compared to ensure that the information was transferred accurately. (Chem Nuclear Systems, 1993, p. 8).

One aspect overlooked in the CNSI description is that operators must encode the attributes of features as well as their locations. Some of you know all too well that tablet digitizing (illustrated in the photo below left) is an extraordinarily tedious task, so onerous that even student interns resent it. One wag here at Penn State suggested that the acronym “GIS” actually stands for “Getting it (the data) In Stinks.” You can substitute your own “S” word if you wish.



Vector digitizing with a tablet (left); raster digitizing with a drum scanner (right) (USGS).

Compared to the drudgery of tablet digitizing, electronically scanning paper maps seems simple and efficient. Here’s how CNSI describes it:

“

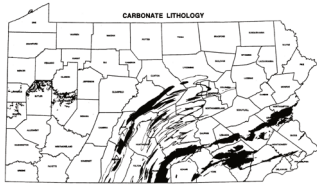
The scanning process is more automated than the digitizing process. Scanning is similar to photocopying, but instead of making a paper copy, the scanning device creates an electronic copy of the source map and stores the information in a computer record. This computer record contains a complete electronic picture (image) of the map and includes shading, symbols, boundary lines, and text. A GIS operator can select the appropriate feature boundaries from such a record. Scanning is useful when maps have very complex boundaries lines that can not be easily traced. (Chem Nuclear Systems, Inc., 1993, p. 8)

I hope you noticed that CNSI’s description glosses over the distinction between raster and vector data. If scanning is really as easy as they suggest, why would anyone ever tablet-digitize anything? In fact, it is not quite so simple to “select the appropriate feature boundaries” from a raster file, which is analogous to a remotely sensed image. The scanned maps had to be transformed from pixels to vector features using a semi-automated procedure called **raster to vector conversion**, otherwise known as “vectorization.” Time-consuming manual editing is required to eliminate unwanted features (like vectorized text), correct digital features that were erroneously attached or combined, and to identify the features by encoding their attributes in a database.

In either the vector or raster case, if the coordinate system, projection, and datums of the original paper map were not well defined, the content of the map first had to be redrawn, by hand, onto another map whose characteristics are known.

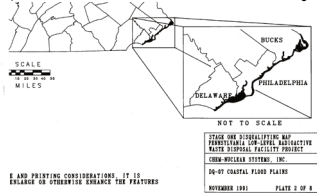
9.9. Stage One: Statewide Screening

CNSI considered several geological, hydrological, surface and subsurface land use criteria in the first stage of its LLRW siting process. [View a table that lists all the Stage One criteria.] CNSI’s GIS subcontractors created separate digital map layers for every criterion. Sources and procedures used to create three of the map layers are discussed briefly below.



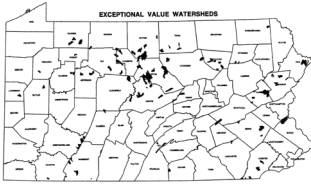
Areas underlain by limestone and other carbonate rocks were digitized from the Pennsylvania Geological Survey’s Geologic Map of Pennsylvania. (Chem-Nuclear Systems, 1991).

One of the geological criteria considered was carbonate lithology. Limestone and other carbonate rocks are permeable. Permeable bedrock increases the likelihood of ground water contamination in the event of a LLRW leak. Areas with carbonate rock outcrops were therefore disqualified during the first stage of the screening process. Boundaries of disqualified areas were digitized from the 1:250,000-scale Geologic Map of Pennsylvania (1980). **What concerns would you have about data quality given a 1:250,000-scale source map?**



Coastal flood plains were digitized from 100-year flood contours compiled from FEMA Flood Insurance Rate Maps onto USGS topographic maps. (Chem-Nuclear Systems, 1991).

Analysts needed to make sure that the LLRW disposal facility would never be inundated with water in the event of a coastal flood, or a rise in sea level. To determine disqualified areas, CNSI’s subcontractors relied upon the Federal Emergency Management Agency’s Flood Insurance Rate Maps (FIRMs). The maps were not available in digital form at the time, and did not include complete metadata. According to the CNSI interim report, “[t]he 100-year flood plains shown on maps obtained from FEMA ... were transferred to USGS 7.5-minute quad sheet maps. The 100-year flood plain boundaries were digitized into the GIS from the 7.5-minute quad sheet maps.” (Chem Nuclear Systems, 1991, p. 11) **Why would the contractors go to the trouble of redrawing the floodplain boundaries onto topographic maps prior to digitizing?**



“Exceptional value watersheds” were delineated on topographic maps, then digitized. (Chem-Nuclear Systems, 1991).

Areas designated as “exceptional value watersheds” were also disqualified during Stage One. Pennsylvania legislation protected 96 streams. Twenty-nine additional streams were added during the site screening process. “The watersheds were delineated on county [1:50,000 or 1:100,000-scale topographic] maps by following the appropriate contour lines. Once delineated, the EV stream and its associated watershed were digitized into the GIS.” (Chem Nuclear Systems, 1991, p. 12) **What digital data sets could have been used to delineate the watersheds automatically, had the data been available?**

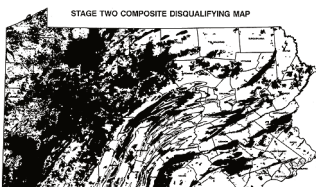
After all the Stage One maps were digitized, georegistered, and overlaid, approximately 23 percent of the state’s land area was disqualified.

9.10. Stage Two: Regional Screening

CNSI considered additional disqualification criteria during the second, “regional” stage of the LLRW siting process. [View a table that lists all the Stage Two criteria.] Some of the Stage Two criteria had already been considered during Stage One, but were now reassessed in light of more detailed data compiled from larger-scale sources. In its interim report, CNSI had this to say about the composite disqualification map shown below:

“

When all the information was entered in to Stage Two database, the GIS was used to draw the maps showing the disqualified land areas. ... The map shows both additions/refinements to the Stage One disqualifying features and those additional disqualifying features examined during Stage Two. (Chem Nuclear Systems, 1993, p. 19)



Composite map showing approximately 46 per cent of the state disqualified as a result of Stages One and Two of the LLRW site selection process. (Chem-Nuclear Systems, 1993).

CNSI added this disclaimer:

“

The Stage Two Disqualifying maps found in Appendix A depict information at a scale of 1:1.5 million. At this scale, one inch on the map represents 24 miles, or one mile is represented on the map by approximately four one-hundreds of an inch. A square 500-acre area measures less than one mile on a side. Printing of such fine detail on the 11" × 17" disqualifying maps was not possible, therefore, **it is possible that small areas of sufficient size for the LLRW disposal facility site may exist within regions**

“

that appear disqualified on the attached maps. [Emphasis in the original document]
The detailed boundary information for these small areas is retained within the GIS even though they are not visually illustrated on the maps. (Chem Nuclear Systems, 1993, p. 20)

As I mentioned back in Chapter 2, CNSI representatives took some heat about the map scale problem in public hearings. Residents took little solace in the assertion that the data in the GIS were more truthful than the data depicted on the map.

11. Stage Three: Local Disqualification

Many more criteria were considered in Stage Three. [[View a table that lists all the Stage Three criteria.](#)] At the completion of the third stage, roughly 75 percent of the state’s land area had been disqualified.

One of the new criteria introduced in Stage Three was **slope**. Analysts were concerned that precipitation runoff, which increases as slope increases, might increase the risk of surface water contamination should the LLRW facility spring a leak. CNSI’s interim report (1994a) states that “[t]he disposal unit area which constitutes approximately 50 acres ... may not be located where there are slopes greater than 15 percent as mapped on U.S. Geological Survey (USGS) 7.5-minute quadrangles utilizing a scale of 1:24,000 ...” (p. 9).

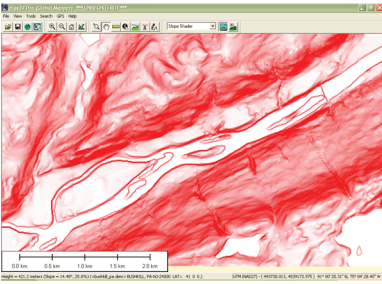
Slope is change in terrain elevation over a given horizontal distance. It is often expressed as a percentage. A 15 percent slope changes at a rate of 15 feet of elevation for every 100 feet of horizontal distance. Slope can be measured directly on topographic maps. The closer the spacing of elevation contours, the greater the slope. CNSI’s GIS subcontractors were able to identify areas with excessive slope on topographic maps using plastic templates called “land slope indicators” that showed the maximum allowable contour spacing.

Fortunately for the subcontractors, 7.5-minute USGS DEMs were available for 85 percent of the state (they’re all available now). Several algorithms have been developed to calculate slope at each grid point of a DEM. As described in chapter 7, the simplest algorithm calculates slope at a grid point as a function of the elevations of the eight points that surround it to the north, northeast, east, southeast, and so on. CNSI’s subcontractors used GIS software that incorporated such an algorithm to identify all grid points whose slopes were greater than 15 percent. The areas represented by these grid points were then made into a new digital map layer.

TRY THIS!

You can create a slope map of the Bushkill PA quadrangle with Global Mapper (dlgv32 Pro) software.

- Launch Global Mapper
- Open the file “bushkill_pa.dem” that you downloaded earlier (either the 10-meter or 30-meter version)
- Change from the default “HSV” shader to the “Slope” shader.



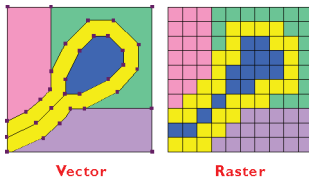
By default, pixels with 0 percent slope are lightest, and pixels with 30 percent slope or more are darkest. You can adjust this at Tools > Configure > Shader Options.

Notice that the slope symbolization does not change even as you change the vertical exaggeration of the DEM (Tools > Configure > Vertical Options).

9.12. Buffering

Several of the disqualification criteria involve **buffer** zones. For example, one disqualifying criterion states that “[t]he area within 1/2 mile of an existing important wetland ... is disqualified.” Another states that “disposal sites may not be located within 1/2 mile of a well or spring which is used as a public water supply.” (Chem-Nuclear Systems, 1994b). As I mentioned in the chapter 1 (and as you may know from experience), **buffering is a GIS procedure by which zones of specified radius or width are defined around selected vector features or raster grid cells.**

Like map overlay, buffering has been implemented in both vector and raster systems. The vector implementation involves expanding a selected feature or features, or producing new surrounding features (polygons). The raster implementation accomplishes the same thing, except that buffers consist of sets of pixels rather than discrete features.



Buffer zones (yellow) surround vector and raster representations of a pond and stream.

9.14. Outcomes

To date, neither Pennsylvania nor New York has built a LLRW disposal facility. Both states gave up on their unpopular siting programs shortly after Republicans replaced Democrats in the 1994 gubernatorial elections.

The New York process was derailed when angry residents challenged proposed sites on account of inaccuracies discovered in the state’s GIS data, and because of the state’s failure to make the data accessible for citizen review in accordance with the Freedom of Information Act (Monmonier, 1995).

Pennsylvania’s \$37 million siting effort succeeded in disqualifying more than three quarters of the state’s land area, but failed to recommend any qualified 500-acre sites. With the volume of its LLRW decreasing, and the Barnwell South Carolina facility still willing to accept Pennsylvania’s waste shipments, the search was suspended “indefinitely” in 1998.

To fulfill its obligations under the LLRW Policy Act, Pennsylvania has initiated a “Community Partnering Plan” that solicits volunteer communities to host a LLRW disposal facility in return for jobs, construction revenues, shares of revenues generated by user fees, property taxes, scholarships, and other benefits. The plan has this to say about the GIS site selection process that preceded it: “The previous approach had been to impose the state’s will

on a municipality by using a screening process based primarily on technical criteria. In contrast, the Community Partnering Plan is voluntary.” (Chem Nuclear Systems, 1996, p. 3)

The New York and Pennsylvania state governments turned to GIS because it offered an impartial and scientific means to locate a facility that nobody wanted in their backyard. Concerned residents criticized the GIS approach as impersonal and technocratic. There is truth to both points of view. Specialists in geographic information need to understand that while GIS can be effective in answering certain well-defined questions, it does not ease the problem of resolving conflicts between private and public interests.

Meanwhile, a Democrat replaced a Republican as governor of South Carolina in 1998. The new governor warned that the Barnwell facility might not continue to accept out-of-state LLRW. “We don’t want to be labeled as the dumping ground for the entire country,” his spokesperson said (Associated Press, 1998).

No volunteer municipality has yet come forward in response to Pennsylvania’s Community Partnering Plan. If the South Carolina facility does stop accepting Pennsylvania’s LLRW shipments, and if no LLRW disposal facility is built within the state’s borders, then nuclear power plants, hospitals, laboratories, and other facilities may be forced to store LLRW on site. It will be interesting to see if the GIS approach to site selection is resumed as a last resort, or if the state will continue to up the ante in its attempts to attract volunteers, in the hope that every municipality has its price. If and when a volunteer community does come forward, detailed geographic data will be produced, integrated, and analyzed to make sure that the proposed site is suitable after all.

TRY THIS!

To find out about LLRW-related activities where you live, use your favorite search engine to search the Web on “Low-Level Radioactive Waste [your state or area of interest]”. If GIS is involved in your state’s LLRW disposal facility site selection process, your state agency that is concerned with environmental affairs is likely to be involved. Add a comment to this page to share your discovery.

9.15. Conclusion

Site selection projects like the ones discussed in this chapter require the integration of diverse geographic data. The ability to integrate and analyze data organized in multiple thematic layers is a hallmark of geographic information systems. To contribute to GIS analyses like these, you need to be both a knowledgeable and skillful GIS user. The objective of this text, and the associated Penn State course, has been to help you become more knowledgeable about geographic data.

Knowledgeable users are well versed in the properties of geographic data that need to be taken into account to make data integration possible. Knowledgeable users understand the distinction between vector and raster data, and know something about how features, topological relationships among features, attributes, and time can be represented within the two approaches. Knowledgeable users understand that in order for geographic data to be organized and analyzed as layers, the data must be both orthorectified and georegistered. Knowledgeable users look out for differences in coordinate systems, map projections, and datums that can confound efforts to georegister data layers. Knowledgeable users know that the information needed to register data layers is found in metadata.

Knowledgeable users understand that all geographic data are generalized, and that the level of detail preserved depends upon the scale and resolution at which the data were originally produced. Knowledgeable users are prepared to convince their bosses that small-scale, low resolution data should not be used for large-scale analyses that require high resolution results. Knowledgeable users never forget that the composition of the Earth’s surface is constantly changing, and that unlike fine wine, the quality of geographic data does not improve over time.

Knowledgeable users are familiar with the characteristics of the “framework” data that make up the U.S. National Spatial Data Infrastructure, and are able to determine whether these data are available for a particular location. Knowledgeable users recognize situations in which existing data are inadequate, and when new data must

be produced. They are familiar enough with geographic information technologies such as GPS, aerial imaging, and satellite remote sensing that they can judge which technology is best suited to a particular mapping problem.

And knowledgeable users know what kinds of questions GIS is, and is not, suited to answer.

QUIZ

Registered Penn State students should return now to the Chapter 9 folder in ANGEL (via the Resources menu to the left) to take the Chapter 9 graded quiz. (Note that this brief chapter included no practice quizzes.) **You may take graded quizzes only once.**

The purpose of the quiz is to ensure that you have studied the text closely, that you have mastered the practice activities, and that you have fulfilled the chapter's learning objectives. You are free to review the chapter during the quiz.

Once you have submitted the quiz and posted any questions you may have to either our discussion forums or chapter pages, you will have completed Chapter 9.

COMMENTS AND QUESTIONS

Registered students are welcome to post comments, questions, and replies to questions about the text. Particularly welcome are anecdotes that relate the chapter text to your personal or professional experience. In addition, there are discussion forums available in the ANGEL course management system for comments and questions about topics that you may not wish to share with the whole world.

To post a comment, scroll down to the text box under "Post new comment" and begin typing in the text box, or you can choose to reply to an existing thread. When you are finished typing, click on either the "Preview" or "Save" button (Save will actually submit your comment). Once your comment is posted, you will be able to edit or delete it as needed. In addition, you will be able to reply to other posts at any time.

Note: the first few words of each comment become its "title" in the thread.

9.16. Bibliography

Associated Press (1998). South Carolina Says Pennsylvania Waste Not Wanted in State. *Centre Daily Times*, , November 28, pp. 1A.

Chem-Nuclear Systems, Inc. (1991). *Pennsylvania low-level radioactive waste disposal facility site screening interim report, stage one — Statewide disqualification*. Harrisburg, PA.

Chem-Nuclear Systems Inc (1993). *Pennsylvania low-level radioactive waste disposal facility site screening interim report stage two — Regional disqualification*. Harrisburg PA.

Chem-Nuclear Systems, Inc. (1994a). *Pennsylvania low-level radioactive waste disposal facility site screening interim report, stage three — local disqualification*. Harrisburg PA.

Chem-Nuclear Systems, Inc. (1994b). *Site selection manual*. S80-PL-007, Rev. 0

Chem-Nuclear Systems Inc. (1996). Community partnering plan: Pennsylvania low-level radioactive waste disposal facility. S80-PL-021, Rev. 0.

Chrisman, N. (1997). *Exploring geographic information systems*. New York: John Wiley & Sons.

McHarg, I. (1971). *Design with nature*. New York: Doubleday / Natural History Press.

Mertz, T. (1993). GIS targets agricultural nonpoint pollution. *GIS World, April*, 41-46.

Monmonier, M. (1995). *Drawing the line: Tales of maps and carto-controversy*. New York: Henry Holt.

Pennsylvania Department of Environmental Protection. (1998). *Proposed model of the PA low-level radioactive waste disposal facility*.

U.S. Nuclear Regulatory Commission. (n. d.). *Radioactive waste: Production, storage, disposal* (Report NUREG/BR-0216).

U.S. Nuclear Regulatory Commission. (2005). *Radioactive Waste Statistics*. Retrieved May 14, 2006, from <http://www.nrc.gov/waste/llw-disposal/statistics.html> (expired)

U.S. Nuclear Regulatory Commission. (2011a). *Low-Level Waste Disposal Statistics*. Retrieved November 30, 2011, from <http://www.nrc.gov/waste/llw-disposal/licensing/statistics.html>

U.S. Nuclear Regulatory Commission. (2011b). *Low-Level Waste Compacts*. Retrieved November 30, 2011, from <http://www.nrc.gov/waste/llw-disposal/licensing/compacts.html>

About the Author

Author: David DiBiase, Senior Lecturer, John A. Dutton e-Education Institute, and Director of Education, Industry Solutions, Esri. Instructors and contributors: Jim Sloan, Research Assistant, Earth and Environmental Systems Institute; and Ryan Baxter, Senior Research Assistant, John A. Dutton e-Education Institute, College of Earth and Mineral Sciences, The Pennsylvania State University.

Penn State Professional Masters Degree in GIS: Winner of the 2009 Sloan Consortium award for Most Outstanding Online Program



© 2014 The Pennsylvania State University

This courseware module is part of Penn State's College of Earth and Mineral Sciences' [OER Initiative](#).

Except where otherwise noted, content on this site is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#).

The College of Earth and Mineral Sciences is committed to making its websites accessible to all users, and welcomes comments or suggestions on access improvements. Please send comments or suggestions on accessibility to the [site editor](#). The site editor may also be contacted with questions or comments about this Open Educational Resource.