Lecture 8: GIS Data Error & GPS Technology

A. Introduction

We have spent the beginning of this class discussing some basic information regarding GIS technology. Now that you have a grasp of the basic terminology associated with GIS we can move on. The remainder of this course will focus on specific GIS applications and data sources that are used by most users of this technology. This lecture will focus on primary data collection, which is possible through the use of Global Positioning Systems (GPS) and supports the on-campus laboratory assignment that you will have next week. Before discussing some specifics associated with GPS we need to review some basic information that defines the quality of spatial data.

B. Spatial Data Quality

Imagine you want to find a specific point by using a map. The actual location of that point compared to its location on the map is an expression of the quality of the map. There are several different ways of expressing uncertainty in spatial data. You need to be able to use the correct terminology.

Precision refers to the variance of repeated measurements of a single entity.

• Precision can be expressed in terms of standard deviation

Accuracy compares the average of a set of measures with the established "true" value. The USGS has established a "National Map Accuracy Standard," which we have previously mentioned. The basis of this standard is that 90% of well-defined points tested will be within a certain tolerance of their actual position.

For example, with a very large-scale map with a 1:2,400 fractional scale, corresponding to the map scale to be used in next week's assignment, the standard is that 90% of the points will be within 0.02 in of their actual position on the map. Let's us work through the real-world distance represented by this standard. This is based in manipulating the fractional scale, which is a topic that some still have issues with, as follows:

	Fractional scale is 1:2400
SO	1 inch on the map $= 2400$ inches in the real world
if	permitted error on the map is 0.02 inches
then	0.02 inches on the map = 0.02×2400 inches in the real world
or	48 inches or 4 feet in the real world

Typically, the accuracy of a data set can be expressed by the comparison of the average of a data set with the actual position and it can be quantified with the **Root Mean Square Error (RMSE)**, where:

RMSE = $(\Sigma (X-t)^2 / N)^{0.5}$

X is an estimator of the central tendency of your observed data set and t represents the "true" value. N represents the number of observations.

Precision and accuracy have specific meanings and cannot be used interchangeable. Figure 1 below graphically illustrates the difference between these two concepts.



Figure 1. Classic illustration of accuracy and precision through indicating clusters of shots relative to a bulls eye.

This figure illustrates that there are four possible combinations of accuracy and precision for a data set with eight measurements.

• High precision, high accuracy - indicates a tight clustering of data that is on the bulls eye - i.e. corresponds with the true value (upper left).

• High precision, low accuracy - indicates a tight clustering of data that is not on the bulls eye indicating that there is a systematic bias in the data set (upper right).

- Low precision, high accuracy indicates a widely dispersed clustering of data whose average is close to the bullseye (lower left).
- Low precision, low accuracy indicates a widely dispersed clustering of data whose average does not correspond with the bullseye (lower right).

In general, decreased accuracy and precision can result from different types of errors. **Systematic error** results from inaccuracies that tend to be consistent in magnitude and direction (Figure 1 - upper right). Conversely, **random errors** vary in magnitude and direction and are difficult to correct. When collecting geospatial data, you need to learn to recognize both of these potential sources of error. Errors associated with primary data collection with a GPS unit can result from three distinct sources:

(1) Human errors can result from incorrect data collection and entry or be related to differences in how geographic entities are conceptualized. Double and triple checking your data can mitigate data entry mistakes and if you are in the field with another operator have that person crosscheck your results. Conceptualization issues can be addressed via communication and planning within a geospatial collection team. An example of how divergent conceptualizations can cause problems goes back to example previously discussed in terms of how to represent a road. Two geospatial technicians are assigned the task to develop a street layer for a small city. Technician A defines the streets as centerlines and technician B defines the streets as polygon as represented by their edges. So basically, the two technicians have collected data that is not compatible with each other and one data set will have to be redone.

(2) GPS and surveying equipment can be influenced by a host of **environmental factors** that can introduce error into a dataset, which we will discuss at the end of the lecture. Additionally, highly precise surveying can be influenced by local variations in gravity as well as magnetic declination.

(3) Instrument error can also cause problems with GPS and surveying activities. A GPS unit obtains radio signals from orbiting satellites to determine its position. Random noise generated within the electronics of the devise that can cause biases in your measurements. This problem is quantified with the signal to noise ratio, which is illustrated in Figure 2.



Figure 2. Illustration of a strong signal and noise.

Think of signal and noise in the context of a radio transmission. The signal is the transmission associated with a specific station and the noise is the white noise you get when your radio is not tuned to a specific station. Figure 2 indicates that the signal is significantly above the noise level indicating a strong signal-to-noise ratio. If the signal has amplitude (or magnitude) that is close to the level of the noise then this is indicative of a weak signal-to-noise ratio - *i.e.* a signal from a far-off radio station that you can barely make out.

C. GPS Systems

An additional primary method of collecting data is with a Global Positioning System (**GPS**) or with Light Detection and Ranging (**LIDAR**). The current GPS system has three major components, which are referred to as segments.

- User segment Individual users
- Space segment US system consists of 24 satellites with other systems around the World

• Control segment - Maintains accurate information on satellite position and adjust orbits of satellites when needed

Typically, in one spot on the planet there are 4 to 7 satellites that are 15 degrees above the horizon. Satellites that are too close to the horizon have their signals attenuated by atmospheric inferences. Obviously, the higher the satellites are above the horizon the stronger the signal. Three satellites need to be in the line of sight of the user in order for a determination of horizontal location (latitude-longitude; northing, easting). With four or more satellites both horizontal and vertical location (elevation) can be determined (Fig. 3).



Figure 3. **Trilateration** determination of latitude, longitude, and elevation with four GPS satellites.

The satellite geometry is one of the most important controls on the accuracy of GPS measurements, which can be expressed as the positional dilution of position (PDOP). Most ideally satellites are spread out in different locations above the horizon (Figure 4), which will result in a low PDOD and greater accuracy. Typically, a PDOP value of less than 6 is considered acceptable. By knowing the orbits of the GPS satellites, you can determine periods when PDOP will be minimized and plan your field work accordingly.

The physical basis of satellite range finding is based on a pseudo-random code that is embedded into the signal sent out to the ground by the satellite (Fig. 5). To make use of this code the GPS receiver must be able to tell precisely when the signal was transmitted and when it was received. Knowing the time that it takes a GPS signal to travel allows for the calculation of range or

distance between the satellite and earth's surface. GPS Satellites have atomic clocks so that the timing of transmission is always known. The quality of the receiver clock on the ground is less robust, which is a major source of measurement error.



Figure 4. PDOP plotted over a 24-hour period. Note the spike in PDOP values around 3:00. During this period unacceptable results would be obtained.

Location errors specifically associated with GPS units can be caused by a number of errors, which can manifest themselves as both systematic and random errors, which can degrade the determination of a location using a GPS unit.

- Instrument error
 - The aforementioned asynchrony between satellite and receiver clocks
 - Departures from expected satellite orbits
 - Electronic noise in the receiver, which was discussed previously
- Environmental error
 - Changes in atmospheric conditions from the "standard atmosphere"
 - Multi-path errors from signal scattering by buildings parking lots and trees



Distance = Speed of Light • Time Difference

Figure 5. Basis for GPS is the time difference between the satellite and receiver that can be used to calculate distance based on the speed of radio waves.

There are several methods that can improve accuracy of GPS data based on measurements derived from single-receiving positioning. The most common of these is the Wide Area Augmentation system (WAAS; http://www8.garmin.com/aboutGPS/waas.html). This system was developed by the FAA to improve the georeferencing of plane location for navigation purposes. The WAAS system uses data from 25 ground stations across the country to correct for problems associated satellite clocks, trajectory, and atmospheric conditions.

Another approach to improving the quality of GPS georeferencing is through differential correction positioning. This method involves using the location of a fixed base station to correct measurements made with a mobile receiver. Using the GPS location of a base station then a range correction can be applied to measurements made using a mobile GPS unit.

Finally, carrier phase tracking provides accuracy by measuring not only the arrival times of GPS signals, but also differences in the phase of the radio waves received at both the base station and mobile receiver providing an additional correction.

There is a strong correlation between the horizontal accuracy of a GPS receiver and its cost!

	Cost	Accuracy
Carrier phase tracking	\$10's K	sub centimeter
Differential positioning	\$1's K	sub meter (Fig. 5.3)
Single-receiving positioning- WAAS	high \$100's	1's meter
Single-receiving positioning	low \$100's	10 meter (Fig. 5.20)

Readings

GIS Commons webpage; Chapters 2. DiBiase, D., 2014, Nature of Geographic Information Systems. Sections 3.11; 3.13 to 3.24.

Terms

Precision	Accuracy	RMSE	Systematic Error
National Map A	ccuracy Standard	Random Error	Signal to Noise Ratio
GPS	LIDAR	Trilateration	PDOP
WAAS	Pseudo-Random Code	Differential GPS	

Concepts

Know the difference between precision and accuracy

Know how to calculate precision based on the National Map Accuracy Standard and accuracy with RMSE

What is the difference between systematic and random error

Outline the different specific types of errors that can degrade a GPS signal

Be able to describe the different segments within the GPS system

Be able to explain how PDOP is related to the number and orientation of GPS satellite overhead

Know about how different GPS receivers operate and how they vary in terms of accuracy

HOMEWORK

1. Describe how a data set can be imprecise, and yet accurate. Would you trust this data?

2. Describe the relationship between PDOD and the number of GPS satellites overhead.

3. Select the most appropriate type of GPS for the following application

General navigation across the country	
Delineation of property boundaries	
Aircraft operations	
Mapping of soil zones	

4. Knowing the precise locations of the GPS satellites in orbit is critical for this system to work.

True or False

5. In Mexico a lawyer is using a relatively small-scale (1:250,000) to locate an abandoned oil well on a client's property for legal purposes. He is having a difficult time finding this old well because it is overgrown with brush. Assuming that 90% of the points will be within 2 mm of their actual position on the map calculate the relative accuracy of features plotted on this map in meters.

6. You go outside and with a GPS measure the location of a tree on ten separate occasions. Later you plot your points using ArcGIS Pro over a reference map layer (the topic for lecture 9). You notice the following offsets in this data from the true value (see table below). Calculate the precision of your data using standard derivation and the accuracy of your data using RMSE in meters. Show your work.

Object	Offset (m)	
1	3.1	
2	3.4	
3	2.7	
4	3.0	
5	2.5	
6	3.8	
7	0.5	Standard Dev
8	3.3	
9	3.2	RMSE
10	2.4	