Testing the Spatial Accuracy of GIS Data

Rj Zimmer, LS

A common mistake among GIS folks who do not have a surveying background, is to rely completely on the GPS user's manual to tell them what the accuracy of their data will be if they follow certain procedures. For instance, the user's manual may state that in order to achieve sub-meter accuracy you must have a PDOP of four, a SNR of four, track at least five satellites, and perform a differential correction. Many GIS folks read the reverse in that statement and assume that if the conditions are met, then the data will be sub-meter. That is similar to thinking that if you used a steel tape that is graduated to hundredths of a foot, all of your measurements with that tape will be accurate to a hundredth of a foot regardless of how you measure. Of course this is not necessarily true. There are many factors that affect the quality and accuracy of field measurements. Some of those factors are acknowledged and quantifiable, some can be controlled, and some factors are beyond human control or knowledge. Another issue, with respect to GPS coordinates, is the difference between relative accuracy and absolute accuracy. Most resource grade GPS software will provide an error report for a data set. Yet, typically, what they are reporting is the relative error of the measurements, not the absolute error. That is a report stating an accuracy of ±1 meters is reporting that the error of that set of measurements is within a meter of each other. That does not necessarily mean that those coordinates are within a meter of the true (absolute) coordinate. In order to test the absolute error, a point with known higher accuracy coordinates must be observed.

Figure 1: Example of GPS road centerline vs. ortho-photography

GIS data may be created from GPS, by digitizing source documents, performing field surveys, aerial mapping, address matching, and other methods. Regardless of the way in which a GIS layer is obtained or acquired, there will be positional errors in the data set. Determining the magnitude of the positional (or location) error is important because the usability of the data set may be dependent upon its spatial accuracy. Metadata, often referred to as the data about the data, is essential for providing potential users with the information needed to determine a GIS data set's usability for an intended purpose. One of the metadata content components is a statement of spatial accuracy. Spatial accuracy is probably the issue that surveyors focus on most when criticizing GIS. However, although spatial accuracy is a big concern to surveyors, it's important to keep in mind that it is not necessarily the most important issue to all GIS users.

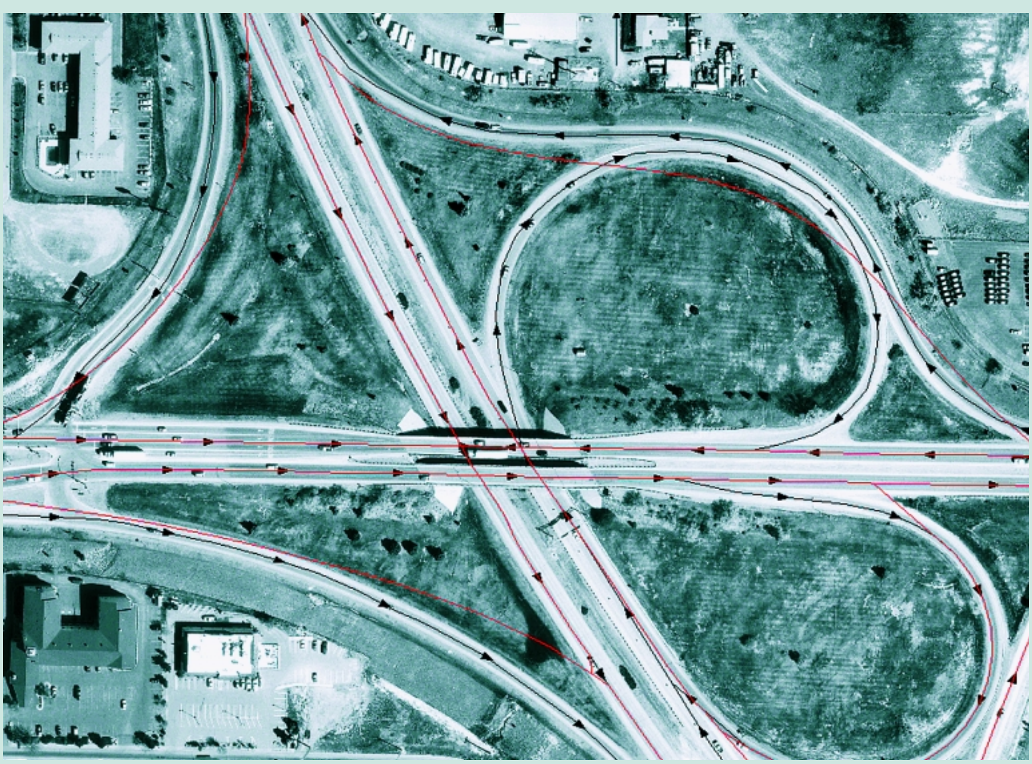
Level of Accuracy

For instance, when emergency responders need information on house locations to aid in developing evacuation plans, their priority is focused on getting the information as quickly as possible. They do not care whether the houses were mapped to an accuracy of ±10 meters, 30 meters, etc. Additionally, wildlife biologists studying land use patterns of a watershed may use small scale mapping as low as 1:100,000 which, according to National Mapping Accuracy Standards, would require accuracy of ±50 meters. Nevertheless, spatial accuracy reporting is important because it does provide the potential data user with the information needed to determine whether or not a data set will work for his or her intended use. Although guidelines for spatial and/or mapping accuracy do exist (see sidebar), the data creator may or may not choose to follow those guidelines. In any case, when data is mapped or converted there is usually some type of mapping or spatial accuracy goal that the project must achieve. This article suggests some methods for testing and validating the accuracy of GIS data sets.

Types of GIS Data

GIS data can be described as points, lines, areas, and raster data. Each of these data types has its unique requirements for spatial accuracy testing. However, for all types of data there are some common considerations to observe. In order to determine the spatial accuracy of a GIS data set the following tasks must be considered: determining what to test, deciding how to test it, (procedure, sample size, sample method), analyzing the sample data, and reporting the results.

Figure 2: Geometry error



How to Test

There are a variety of ways to perform the spatial accuracy testing. Typically the procedures are to use some sort of measurement or test that is extrinsic to the data set, such as an independent data source or computation. Independent sources should be of a higher accuracy than the data set to be tested. Some examples include existing digital or hard copy map data, GPS, or terrestrial survey data. In lieu of extrinsic data, estimates can be computed from intrinsic sources such as knowledge of the accuracy of the source document, map registration and digitizing accuracy (based on scale and methods used), and so forth. Creating independent measurements assure the highest reliability of the accuracy determination.

The methods selected should depend on the objectives and the availability of existing data. If the GIS objective is to fit the data set into other existing (higher accuracy) data, then the new set may be tested against the existing data. For example, the State of Texas created a state-wide GIS layer of public roads with the requirement that it correlate with existing digital orthorectified photography (see **Figure 1**). In that case, the accuracy of GPS data can be tested by overlaying the road network on the photography, then measuring, on-screen, the difference between the image of a road segment and the GPS road segment. For instance, a road intersection on the photography and the GPS centerline would be the test. The distance between them would be a single sample. Natural resources data, such as vegetation coverages, are more difficult to test because such data sets do not describe well-defined points.

Another way to perform testing is to identify points in the data set that can be physically measured in the field, then measuring those samples with higher accuracy methods. For example, if the accuracy of a manhole inventory was being checked, then a sample set of manholes would be re-measured with higher accuracy equipment and methods. If the GIS requirement was to map those manholes to a one-meter level of accuracy, then a sample set of manholes should be tested using methods that yield better than a one-meter accuracy. The difference between the original measurements, and the test measurements will be the accuracy of the manhole data set.

Some Sample Methods

Points are the simplest GIS features to test. Points have only a location (coordinates), so the method would be to obtain the coordinates of the test point, then compare that to the coordinates of the same point as defined by a higher accuracy source (such as GPS). Lines and areas, however, are more complex features to test, because their geometry is more complex. The consideration for the more complex geometry data sets is to test the geometry as well as the coordinates of discrete points.

FGDC Accuracy Standards:
www.fgdc.gov/standards/status/sub1_3.html

"The NSSDA is intended to replace the 1947 National Map Accuracy Standard (NMAS). The applicability of NMAS is limited to graphic maps, as accuracy is defined by map scale. The NSSDA was developed to report accuracy of digital geospatial data that is not constrained by scale."

Testing lines, for example, requires testing the accuracy of the end points of the lines such as at road intersections, and the geometry of the line between the endpoints (see **Figure 2** geometry error). The geometry is how the line behaves between the end points. This raises such questions as: Does it go in a straight line? or, Does it curve left or right?, and so forth. Testing the end points of the lines is the same as testing a simple point. However, testing the accuracy of the geometry of the line requires obtaining a coordinate for some point or points along that geometry. Since a road centerline created from parametric modeling (such as typing in the curve data) would create a more mathematically correct geometry than a GPS representation of that same curve, the difference can be difficult to differentiate. A sharp curve would require more intermediate GPS points than a flatter curve.

Testing the more complex geometry of a polygon requires finding discrete points on the edges of the polygon, such as vertices, which can be correlated to similar points on the data set of higher accuracy. If the polygon were a parcel boundary, for instance, one of the checks for spatial accuracy would be to test a corner of the parcel for accuracy.

Sample Size

An important consideration when testing spatial accuracy is selecting an appropriate size sample set. There must be a large enough sample size to produce a statistically valid result. On the other hand, the sample size is constrained by the cost and the time required to perform the sampling. Surveyors well understand the value of redundant measurements, but there is a point after which the value of more measurements diminishes. The balance between the number of measurements and the value of the measurements is unique for each project and should be dealt with individually. Some of the factors to consider are the size of the data set (10 points or 10,000 points), the distribution of the data, and the importance of the spatial accuracy. If the spatial accuracy is of high importance, then that may be incentive to test a large sample set than the minimum required. There are statistical programs available for calculating sampling sizes for known and unknown data set sizes.

For photography, a common sample size is 20 points per image, distributed randomly throughout the image. Line data, such as river or road networks, could be tested several different ways. A certain portion of the total length of the data set could be tested, or a percentage of characteristics, such as junctions or angle points, could be tested. Additionally, a random set (such as 1 percent) of the points along the network could be tested, or measurements could be taken at predetermined intervals (X distance or Y percent along the network). For example, if a linear network were 100 kilometers long, then the sample set may consist of points every 1/2 km, or 1 km along the route for a standard interval. Alternatively, samples may be taken every 5 percent, or 20 percent along the way.

Reporting the Results

One method of reporting the spatial accuracy is the FGDC Metadata Content Standard for Spatial Data Accuracy. The standard requires a quantitative and qualitative statement of accuracy. Additionally, most survey adjustment software and GPS software provide reports in a variety of proprietary formats (some are customizable). The important things that are generally helpful to the end user are a quantitative statement of accuracy and some information about how that was determined.

An abridged example is shown below, taken from the North Texas GIS Consortium Metadata for Pavement (*www.ntgisc.org/warehouse/metadata/roadedge.html*).

Positional Accuracy

Horizontal Positional Accuracy

Horizontal Positional Accuracy Report: Horizontal positional accuracy is 1.0 meter defined by the root mean square error (RMSE) method. This requires that two-thirds of all photo-identifiable arc features fall within the stated accuracy of 1.0 meter and that 90 percent of all arcs must fall within twice the distance specified (*i.e.*, 2.0 meters). A final inspection and acceptance process was completed by several North Texas Consortium members.

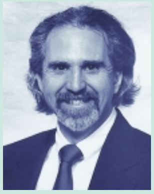
Quantitative Horizontal Positional Accuracy Assessment

Horizontal Positional Accuracy Value: 3.2ft Horizontal Positional Accuracy Explanation: Resolution as reported.

Vertical Positional Accuracy

Vertical Positional Accuracy Report: Vertical positional accuracy of 1 meter RMSE.

As the push to share geographic information increases, the need to verify and document the spatial accuracy of data sets becomes ever more important. Spatial accuracy assessments of GIS data is a task that surveyors are well-suited for, therefore they should lend their expertise to those that can benefit from it.



RJ Zimmer *is the GIS Coordinatior for Lewis & Clark County and the city of Helena, Montana, and a Contributing Writer for the magazine.*