The Positional Accuracy of MAF/TIGER

Kristen O'Grady
Leslie Godwin
U.S. Department of Commerce, Bureau of the Census
Geography Division
Geospatial Research and Standards Staff
Washington, DC 20233-7400
kogrady@geo.census.gov / lgodwin@geo.census.gov

Abstract

The Bureau of the Census initially developed the Topologically Integrated Geographic Encoding and Referencing (TIGER) database, a main component of the Master Address File/TIGER System (MAF/TIGER System), to provide geographic services and products in support of the 1990 decennial and other Census Bureau statistical programs. Since its inception, the public has used the TIGER/Line Files (public-released extracts of the TIGER database) in a variety of activities, many of which require a positional accuracy beyond the Bureau's initial needs. These many additional uses, coupled with the Bureau's anticipated positional accuracy requirements for successfully merging the functionality of the MAF and TIGER into a fully integrated system for use in future monthly programs and censuses, have made improving the positional accuracy of TIGER a Bureau goal to be accomplished "beyond-2000."

TIGER was built and has been continuously updated using a wide variety of source materials and techniques, including the GBF/DIME files, USGS 1:100,000-scale topographic maps, local and tribal maps, and enumerator updates of differing positional accuracy. Because of its varied update history, answering the question "What is the positional accuracy of MAF/TIGER" is not as straightforward as one might hope.

This paper discusses the ways by which the positional accuracy of TIGER can be quantitatively measured and reported. Understanding the existing positional accuracy of TIGER is the first step toward attempting to improve the positional accuracy of TIGER. The Bureau is considering and testing courses of action for improving the positional accuracy of TIGER, focusing on differing data collection methodologies. These methodologies are reviewed.

Positional Accuracy: What is it?

<u>A Historic Perspective</u>

Positional accuracy is the most widely accepted and reported of the components used for describing the quality of spatial data. Positional accuracy was initially viewed in terms of the cartographic representations of datasets. The U.S. Bureau of the Budget's[1] United

[1] Now the Office of Management and Budget.

States National Map Accuracy Standard (1947) developed methods for measuring and reporting the accuracy of both location (the latitude and longitude of a point) and elevation (the altitude above sea level) of maps at particular publication scales.

The Spatial Data Transfer Standard (SDTS) identified positional accuracy as an element of spatial data quality (in 1992) in the larger context of its relevancy to other spatial datasets[2]. Although a substantive definition of positional accuracy was not offered, the SDTS identified positional accuracy as having a quality report that includes "the degree of compliance to the spatial registration standard" or the mechanisms by which spatial data are related to locations on the Earth's surface.

In 1982, positional accuracy was identified in the Federal Geographic Data Committee's (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM) as an important component of the metadata describing a dataset[3]. The CSDGM defined the positional accuracy metadata element as "an assessment of the accuracy of the positions of spatial objects." The CSDGM identified two aspects of positional accuracy in keeping with the United States National Map Accuracy Standard's location and elevation components. Horizontal positional accuracy (location) is "an estimate of accuracy of the horizontal positions of the spatial objects." Vertical positional accuracy (elevation) is "an estimate of accuracy of the vertical positions in the data set." The CSDGM's metadata for positional accuracy can be quantitative (for example, a statistic such as 'root mean square error at a 90% confidence level') or qualitative (for example, a subjective analysis such as 'we believe the accuracy of this dataset to be pretty good'). In allowing for both quantitative and qualitative statements of positional accuracy, the CSDGM acknowledged that, although quantitative quality measures are desired, often only qualitative quality information is available. Providing any quality statement is better than providing none.

In 1991, the International Cartographic Association's Commission on Spatial Data Quality published the *Elements of Spatial Data Quality*. In keeping with its professional scope, this publication offered a description of positional accuracy within the context of the mapping sciences: "In the mapping sciences the position of a real world entity is described by values in an appropriate coordinate system. Positional accuracy represents the nearness of those values to the entity's 'true' position in that system."

A surprising amount of time elapsed before the United States National Map Accuracy Standard was replaced by a standard that reflected the need of the geographic community for a methodology of describing the positional accuracy of digital spatial datasets that can

---

[2] Work on the concepts reflected in the SDTS actually began much earlier than 1992. The American Congress on Surveying and Mapping (ACSM) sponsored the Moellering Committee in the early 80's. The Moellering Committee published a series of reports; the final report released in 1987 contained the essence of the SDTS. The SDTS was issued as Federal Information Processing Standard Publication (FIPS PUB) 173 by the U.S. Department of Commerce, National Institute of Standards and Technology (NIST) in August 1992. The American National Standards Institute (ANSI) approved a modified version of the SDTS as ANSI/NCITS 320.1998 in June 1998. Current users of this standard should reference ANSI/NCITS 320.1998.

[3] Version 2.0 of the CSDGM was released by the FGDC as FGDC-STD-001-1998 in 1998. Current users of this standard should reference FGDC-STD-001-1998.

be represented at many scales. In June 1998, the FGDC endorsed the National Standard for Spatial Data Accuracy (NSSDA). The NSSDA defined a statistical and testing methodology for positional accuracy and provided a common language for reporting this accuracy. The NSSDA acknowledged the requirement for testing the positional accuracy of a dataset using a control dataset acquired separately from the dataset being tested and being more accurate. In fact, it stated the control dataset should be "of the highest accuracy available."

A Current Perspective

In 1990, the International Organization for Standardization, Technical Committee 211 (ISO/TC 211) Geographic Information/Geomatics began developing a suite of standards, the 19100 series[4], for geographic information systems. The standards are scheduled for completion by 2002. Two of ISO/TC 211's standards address the quality of digital geographic data. ISO 19113 Quality Principles (19113) identifies five specific quantitative components of data quality. Components of data quality are called data quality elements. One identified data quality element is positional accuracy. Positional accuracy "shall describe the accuracy of the position of features."

19113 notes all data quality elements have unique aspects of quality. The unique aspects of a data quality element are its data quality subelements. The data quality element of positional accuracy has three identified data quality subelements: absolute or external accuracy, relative or internal accuracy and gridded data position accuracy.

This paper's discussion on the positional accuracy of the Census Bureau's Master Address File/Topologically Integrated Geographic Encoding and Referencing System (the MAF/TIGER System, hereafter referred to as TIGER) and the TIGER/Line files that are publicly available extracts of TIGER, is based upon 19113's data quality element/data quality subelement of positional accuracy/absolute or external accuracy. (Although not explicitly stated, absolute or external accuracy was implied in all the publications discussed in the Historic Perspective section of this paper.) 19113 describes the data quality subelement of absolute or external accuracy as the "closeness of reported coordinate values to values accepted as or being true." Truth, for TIGER, is considered by the Bureau as being actual ground truth or the position of a feature on the earth's surface. Although absolute or external accuracy can be reported for both the horizontal and vertical dimensions of a dataset, TIGER currently contains only X, Y coordinate values. Therefore, this paper will address only horizontal absolute or external accuracy.

19113 identifies six descriptors of a data quality subelement to be used in describing and reporting relevant quality information minimally as metadata. The six descriptors are:

- data quality scope - extent or characteristics of the data for which quality information is reported,

[4] ISO/TC 211 initially began developing a multi-part standard. The International Standard was numbered 15046 with each specific part noted by a hyphenated extension (for example, 15046-13). The International Standard has recently been reorganized as a suite of standards to be numbered as a 19100 series (for example, 19113).

- data quality measure - type of quality test applied,
- data quality evaluation procedure - operations used in applying the data quality measure to the data quality scope,
- data quality date - date the data quality measure is applied,
- data quality result - value or set of values resulting from applying the data quality measure, and
- data quality value type - value type or unit for reporting the data quality result.

The descriptors are necessary for complete reporting and subsequent interpretation of quality information. Neither 19113 nor ISO 19114, Quality Evaluation Procedures (19114), identify a specific statistic to be used in reporting a data quality result (such as standard deviation or root mean square error) or a minimum testing methodology for obtaining a data quality result as is identified in the NSSDA. This is partially because the ISO/TC 211 quality standards address the range of data quality elements and their data quality subelements from completeness/commission through logical consistency/ topologic consistency (to name two of the 17 identified data quality element/data quality subelement combinations). The complexity of identifying a single statistic and minimum testing methodology for the diversity of national perspectives on spatial data quality components should be evident. This paper accepts the FGDC-endorsed NSSDA as the methodology to be used to establish the positional accuracy/absolute or external accuracy of a dataset.

Positional Accuracy Information for the TIGER/Line Files

The initial purpose of TIGER was to automate and integrate all cartographic and geographic processes for the collection, tabulation and dissemination needs of the 1990 decennial census. TIGER enabled the Bureau to produce its geographic products from a single computer database in support of its operations and, thereby, know that map lists of geographic areas, and representations of feature names would be consistent in all products. Accuracy was crucial but only in a relational sense. It was not necessary to produce a positionally accurate file. Of main concern was the location of geographic features in relation to other geographic features.

Historically, there have been four main uses of TIGER within the Bureau, including:

- assigning geographic location codes to residential and business addresses for data collection,
- recording Census-recognized geographies (i.e. counties, places, tracts, census blocks etc.),
- providing geographic structure (i.e. the relationship of one geographic area to another), and
- map production and dissemination, for use by census enumerators and data users respectively.

TIGER and the TIGER/Line Files share the same positional accuracy/absolute or external accuracy. In this respect, the quality information for positional accuracy/absolute or external accuracy is unlike quality information for other data quality elements and their data quality subelements. For example, TIGER and the TIGER/Line Files have unique formats; therefore, quality information for the data quality element/data quality subelement of logical consistency/format consistency is unique to each. A format error could occur in extracting the TIGER/Line Files that is totally unrelated to and irrelevant when speaking of the logical consistency/format consistency of TIGER. Because the positional accuracy/absolute or external accuracy quality information for both TIGER and the TIGER/Line Files is identical and TIGER/Line Files are released to the public, this paper will refer to the positional accuracy of the TIGER/Line Files in the next few sections.

The Bureau's objective is to adopt the use of the 19113 data quality elements and their data quality subelements and report quality information for those determined to be relevant to the TIGER/Line Files using the six descriptors of a data quality subelement. Coordinates for features are a major component of the TIGER/Line Files. This coupled with the fact that the Bureau's initial requirement for minimal positional accuracy is rapidly becoming a requirement for increased positional accuracy, makes quality information for the positional accuracy/absolute or external accuracy of the TIGER/Line Files very relevant.

Current Positional Accuracy Statements

Data quality scope often is viewed as a beginning point for organizing quality information. The Bureau currently reports positional accuracy for a data quality scope equaling its entire database (or the entire series of TIGER/Line Files). TIGER/Line File coverage equals the United States, Puerto Rico, and the Island Areas of American Samoa, Guam, the Commonwealth of Northern Mariana Islands, and the Virgin Islands of the United States, with one TIGER/Line File being produced per county or statistically equivalent entity.

Most users of the TIGER/Line Files who are unfamiliar with the lineage or source of the data they contain initially expect a high degree of positional accuracy. This is because the coordinates in the TIGER/Line Files have six implied decimal places of precision. The six-decimal place precision is useful when producing maps, as it allows the proper relative placement of features that are next to each other on the ground without overlap.

The following statement on positional accuracy first appeared in the 1995 TIGER/Line File documentation and refutes the accuracy implied by the six-decimal place precision of coordinates:

> "The positional accuracy varies with the source materials used, but at best meets the established National Map Accuracy standards (approximately +/- 167 feet) where 1:100,000-scale maps from the USGS are the source. The Census Bureau cannot specify the accuracy of feature updates added by its field staff or of features derived from the GBF/DIME-Files or other map sources. Thus, the level of positional accuracy in the 1995 TIGER/Line files is not suitable for high-precision measurement applications such as engineering problems, property

transfers, or other uses that might require highly accurate measurements of the earth's surface."

In terms of the six descriptors of a data quality subelement, the Bureau provides quantitative quality information for a data quality scope equaling the features within the TIGER/Line Files that share a common source of USGS 1:100,000-scale maps. The data quality result is +/- 167 and the data quality value type is feet. The data quality measure and data quality evaluation procedure for this data quality scope, though not specifically stated, can be obtained by referencing the testing methodology described in the 1947 United States National Map Accuracy Standard. The data quality date is not provided and is difficult to do for a dynamic file.

The quality statement for the remaining features within the TIGER/Line Files (those having other sources) is qualitative, rather than quantitative, in nature.

<u>Desired Positional Accuracy Information</u>

The intent of 19113 is for a data producer to identify and report quality information for all data quality scopes thought to share a uniform quality. A data quality scope can be any subset of data within a dataset that has common characteristics, including belonging to a feature type, sharing data collection criteria, sharing original source, or being within a specified geographic or temporal extent. Several data quality scopes can be identified within the TIGER/Line Files, all based on source. All TIGER/Line features are assigned a source code[5] that represents the operation creating the geographic object and implies its geometric properties. The currently utilized source code values and their descriptions are:

Source Code	
Value	Description
Blank	Not Documented Elsewhere
A	Updated 1980 GBF/DIME-File
B	USGS 1:100,000-Scale DLG-3 File
C	Other USGS Map
D	Census Bureau Update Prior to 1990 Enumeration Operations
E	Census Bureau 1990 Enumerator Update
F	Census Bureau Update from Other 1990 Operations
G	Unconfirmed Local Official Updates
H	Census Bureau Update Post-1990 Operations
I	Census Address List/TIGER Linkage Operations

TIGER began life as a patchwork quilt of data sources and subsequent update operations have exacerbated the situation. Operations codes are updated whenever a feature is 'touched' by an operation. Touching often means a change in attribution rather than a change in coordinate position. Because of this, it is difficult to definitively identify the source of the position of a feature. There are numerous operations codes (approximately

5 Revisions to source codes have been proposed to add more specificity and for better understanding for TIGER/Line users.

one hundred). Operations codes are categorized into types and summarized into the source codes listed above.

The Bureau currently provides quality information for a single data quality scope (the entire nation). This data quality scope has been shown to have variable rather than uniform quality in terms of positional accuracy/absolute or external accuracy. The Bureau would like to provide quality information for, at a minimum, each group of features sharing the same source code. Optimally, the Bureau would like to provide quality information at the individual feature level. The Bureau's challenge in providing this information lies in the diversity of the TIGER/Line File feature sources, the limited information maintained on the positional accuracy of each source, and the difficulty in identifying the true lineage of a feature's coordinates; at this point in the life of the TIGER/Line Files doing so is not possible. Until TIGER is restructured to maintain a positional accuracy quality attribute for each individual feature, the Bureau will be able to provide only quantitative quality statements for the few data quality scopes that share common source codes. The quality information becomes more relevant if it is accompanied with the knowledge of the percentage of features belonging to each data quality scope that occurs within a particular TIGER/Line File. Addressing several data quality scopes could provide a more complete picture of positional accuracy, in a deduced rather than actually measured sense, of the overall accuracy of an area based on the percentages of features associated with each type of source or update operation.

The Bureau's goals after completing Census 2000 include improving the positional accuracy of TIGER. Resources will not likely be expended to improve the quality information available for previously released TIGER/Line Files. The Bureau has actively participated in the development of the ISO/TC 211 quality standards and is aware of the importance in reporting quantitative quality information. To this end, the Bureau is focusing on both the documentation and reporting aspects of quality information as it investigates methodologies for improving the positional accuracy of TIGER. The remainder of this paper discusses the methodologies being considered for successfully improving the positional accuracy of TIGER.

Improving the Positional Accuracy of TIGER

Three needs contribute to the requirement of improving the positional accuracy of TIGER:

- internal needs,
- a desire to use local, and tribal files for updates, and
- a desire to facilitate data exchange.

Internal needs arise from a technological requirement. One of the Bureau's goals for the next decade is to capture the latitude and longitude coordinate for living quarters and to equip each field interviewer with portable computers equipped with Global Positioning System (GPS) technology to more effectively locate living quarters requiring a visit. However, to integrate the more accurate coordinates that GPS can provide for living quarters (and the streets they are along) with the existing MAF/TIGER System[6], current TIGER features must have an equivalent level of positional accuracy. Herein lies the challenge. As GPS field-collected data become 'truth,' what is the best way to improve TIGER feature positions to match the positional accuracy of GPS data? Additionally, the Bureau has found in its efforts to use the many files that local, and tribal agencies offer for update purposes which have a greater positional accuracy, the current positional accuracy of TIGER is a limiting factor. The Bureau's desire to form partnerships with local, and tribal agencies to maintain an up-to-date database relies upon improving the positional accuracy of TIGER.

<u>Formulating a Plan</u>

An extreme method of improving the positional accuracy of TIGER would be to re-collect all database features. Such an approach would be neither prudent nor practical, however, because of the extensive attribute information currently associated into the existing feature information. The Geography Division believes there is a viable alternative to this all out effort. An example of this alternative would be the limited re-collection of a series of points, already in TIGER, that could then be used to realign both the original points and additionally all other features around them through transformation (that will maintain existing topological relationships). This particular method does not address the needs to both spatially enhance existing features (by improving their shape fidelity) or improve TIGER by adding new features.

How can valid transformation points be identified? Valid points must exist in TIGER and be easily locatable in the field (or on imagery). TIGER contains five basic types of line features, including roads, railroads, hydrography, miscellaneous transportation features (including selected power lines and pipe lines), and boundaries (which align with often non-visible features). Line features consist of one or more line segments. A line segment is comprised of two nodes (a beginning and end node) and may additionally contain shape points.

The Geography Division has named its valid transformation points "anchor points." An anchor point is a node representing an intersection of three or more nodes of TIGER linear features, with only roads, railroads and hydrographic features acceptable as an intersecting linear feature. A further requirement for an anchor point is that at least two of the intersecting linear features <u>must</u> be roads. The quality of each anchor point is important and depends upon the line segments that comprise it. A quality-rating scheme has been developed from the existing source codes assigned each linear segment in TIGER for documenting the quality of anchor points.

<u>Capture Techniques</u>

How to best capture new coordinates for existing anchor points? Methodologies being considered are the use of GPS technology in the field and the use of imagery in the office.

---

[6] The TIGER/Line Files will not contain latitude and longitude coordinates for living quarters. Title 13 U.S.C. prohibits the disclosure of confidential information and it has been determined that an address, when attached to coordinates, makes the coordinates confidential.

The Geospatial Research and Standards Staff (GRaSS) of the Geography Division is performing tests investigating the pros and cons of each methodology.

The use of GPS has some distinct advantages over the use of imagery. For example, GRaSS has collected, along with anchor points, road centerlines with the objective of additionally improving the shape fidelity of roads as well as updating the inventory of TIGER road features. A two-week field project involved six teams of two individuals, each team equipped with a vehicle, laptop, GPS receiver and GPS software. Each team collected anchor points, road centerlines, feature attributes and other optional information. A key attribute collected was the anchor point identification number, ensuring a link back to existing anchor points in TIGER.

GRaSS is experimenting with Digital Orthophoto Quadrangles (DOQ) as a preferred imagery media. DOQs are commonly used as source data for collecting digital information in many other GIS applications. GRaSS selected DOQs for their availability, extensive geographic coverage, and ease of use. With the DOQ serving as the base, the object of the test is limited to capturing the coordinates of anchor points only. Again, a key attribute collected is the anchor point identification number that ensures a link to TIGER.

The use of satellite imagery with a pixel resolution of 1 to 5 meters also is being considered.

Concurrent with anchor point data collection, GRaSS is investigating both "in-house" development and commercial coordinate transformation software. Eventually, all three data collection techniques will be evaluated to determine which is a more efficient and feasible method of data collection to be considered for use on a nationwide basis.

Conclusion

Whereas relational accuracy once was adequate for Census activities, changing goals and technologies have caused the Bureau to acknowledge the future need for a positionally more accurate TIGER. An understanding of this need has been accompanied by the desire to more thoroughly understand the existing positional accuracy of TIGER.

As the Census Bureau explores methodologies for collecting and improving the coordinates in TIGER, it does so with the added goal of documenting and recording quality information about the positional accuracy of TIGER using the ISO/TC 211 suite of standards as its guide. More specifically, the Bureau seeks to incorporate the concepts of the six descriptors of a data quality subelement to provide complete quantitative quality information. The Bureau is working toward both improving positional accuracy and providing positional accuracy statements that ultimately will express the positional accuracy of individual features. By so doing both the data producer (the Census Bureau) and the data user will be able to more effectively assess the quality of TIGER in terms of their specific needs and requirements.

References

American National Standards Institute. 1998. *Information Technology – Spatial Data Transfer Standard (SDTS)* ANSI-NCITS 320. New York, NY. http://mcmcweb.er.usgs.gov/sdts/

American Society for Photogrammetry and Remote Sensing (ASPRTS) Specifications and Standards Committee. 1990. "ASPRS Accuracy Standards for Large-Scale Maps." Photogrammetric Engineering & Remote Sensing. Washington, D.C. v56, no 7, pp 1068-1070.

Bureau of the Budget. 1947. *United States National Map Accuracy Standards*. U.S. Bureau of the Budget. Washington, D.C.

Federal Geographic Data Committee. 1998. *Content Standard for Digital Geospatial Metadata (version 2.0)* FGDC-STD-001-1998. FGDC. Washington, D.C. http://www.fgdc.gov/standards/status/standards.html

Federal Geographic Data Committee. 1998. *Geospatial Positioning Accuracy Standards; Part 3: National Standard for Spatial Data Accuracy* FGDC-STD-007.3. FGDC. Washington, D.C. http://www.fgdc.gov/standards/status/standards.html

Hartung, Valerie L. 1998. Bureau of the Census' GPS Research Project: Setup and Execution of Field Activities. In: Proceedings, GIS/LIS '98, Annual Conference and Exposition.

International Organization for Standardization/Technical Committee (ISO/TC) 211, *N621 Committee Draft (CD) 15046-13.2: Geographic information - Part 13: Quality principles* (not available to the public).

International Organization for Standardization/Technical Committee (ISO/TC) 211, N665 CD 15046-14: Geographic information - Part 14: Quality evaluation procedures (not available to the public).

LaMaccia, Robert. 1989. Census Data Available for Transportation Planners.

Minnesota Governor's Council on Geographic Information GIS Standards Committee. October, 1998. *Implementing the National Standard for Spatial Data Accuracy*. http://www.Imic.state.mn.us/gc/gc.htm.

Morrison, Joel L. and Guptill Stephen C., Editors. 1995. *Elements of Spatial Data Quality*, published on behalf of the International Cartographic Association by Elsevier Science.

TIGER/Line™ Files, 1995 [machine-readable data files]/prepared by the Bureau of the Census. - Washington, DC: the Bureau [producer and distributor], 1995.