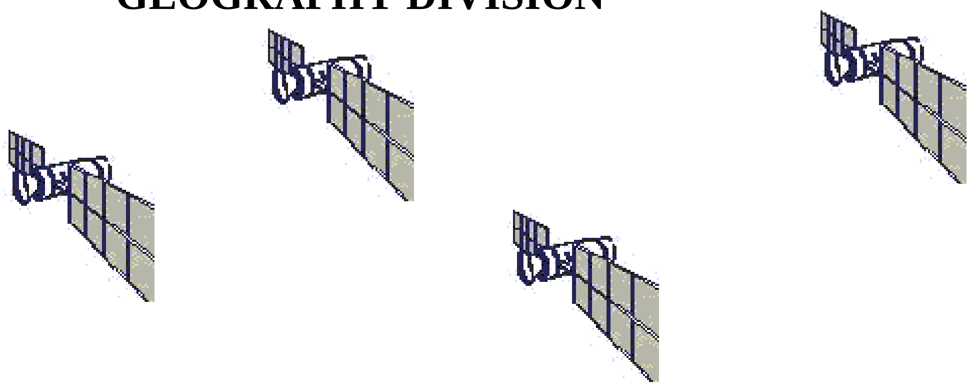
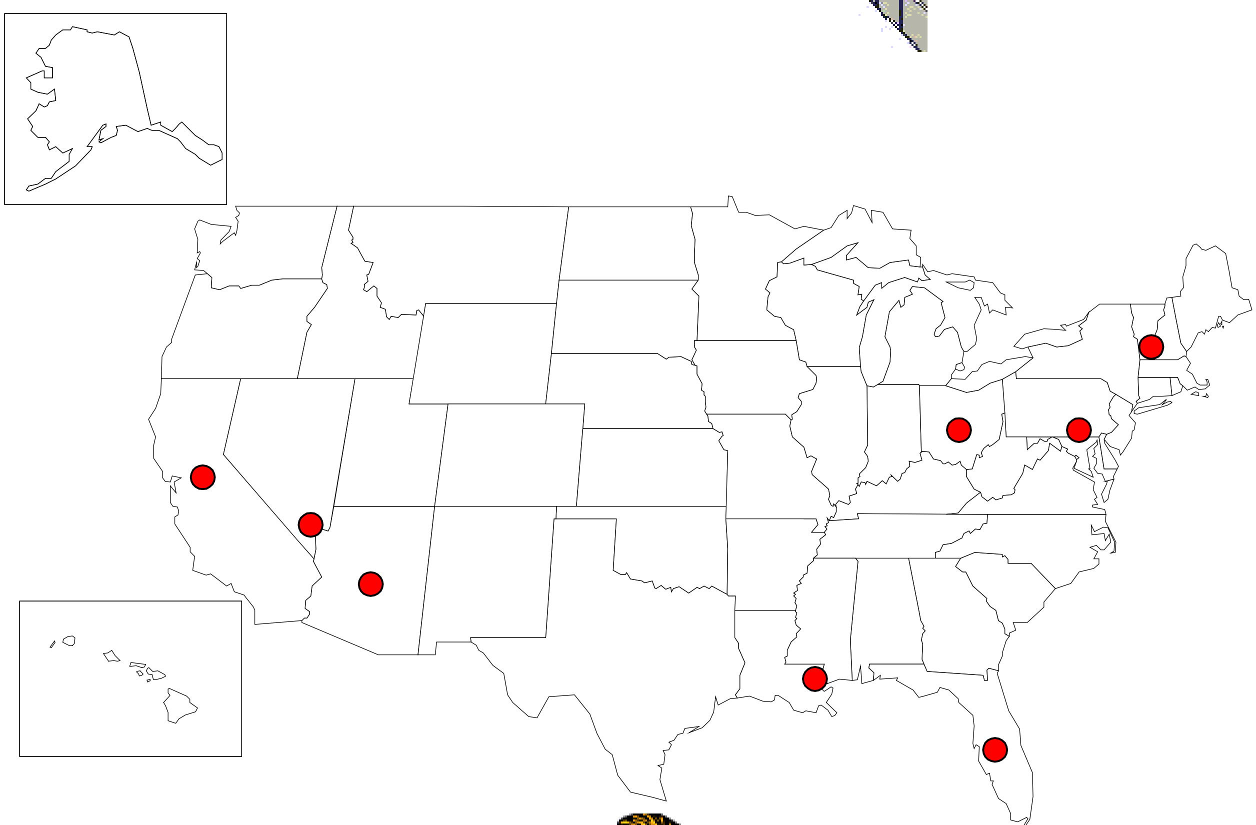
GPS TIGER Accuracy Analysis Tools (GTAAT) Evaluation and Test Results

By John S. Liadis
TIGER Operations Branch
GEOGRAPHY DIVISION





May 24, 2000

TIGER Accuracy Analysis Tools (GTAAT)
Acknowledgements

Joe Knott, Geographic Operations Advisor to the Chief, Geography Division, requested the TIGER coordinate accuracy study and the development of an analysis tool that uses Global Positioning System (GPS) information integrated with automated address listing instrument**.** Significant contributions to this document were made by**: Randy Fusaro**, Chief, TIGER Operations Branch; **Alfred Pfeiffer**, Chief, Address List Review Branch, received a Commerce Pioneer Grant in 1994 that initiated the development of the current GPS field-listing instrument; **Kevin Donnalley,** TIGER Operations Branch, project leader for GPS Field Listing Instrument activities, is the liaison with HTE-UCS, Inc., and administered the Windham, Vermont field test; **Steve Ho**, TIGER Operations Branch, collected field data and administered Windham, VT field test; **Mitch Milligan**, TIGER Operations Branch, wrote the Windham County, VT test requirements and collected field data in the same county; **Alice Blackburn**, software developer from THE-UCS, Inc., main developer of several U.S.Census Bureau software applications, including the GPS Cartographic and GPS TIGER Accuracy Analysis Tool; **Frederick Malkus,** Geospatial Research and Standards Staff, developed anchor point for the GPS Cartographic Instrument and source code link software for the GPS TIGER Accuracy Analysis Tool; **Frederick Broome,** and **Leslie Godwin**, Geospatial Research and Standards Staff.; and **Joan Meiller**, TIGER Systems Branch.

The analysis found hereinafter would not be possible without the staff that gathered the field data. This includes: **Sharon Cochran** and **David Shupe**, TIGER Operations Branch, who collected field data in Clark, NV, Maricopa, AZ, Sacramento, CA, and Windham, VT; **Stuart Irby**, TIGER Operations Branch and **Shawn Radziminski**, Update Operations Branch, who collected field data in Windham, VT and York, PA; **Jon Sperling**, TIGER Operations Branch, **Colleen Carbone**, Address List Review Branch, **Barbara Wallace**, Update Operations Branch, **Michael Fournier**, Geographic Areas Branch, and **Karen Cappiella** and **Mike Weaver**, former Census Bureau employees, who collected field data in Windham, VT.

Table of Contents

Introduction

The Geography Division (GEO) requires an assessment tool to determine the spatial accuracy of the Topologically Integrated Geographic Encoding and Referencing (TIGER®) data base and that of potential geospatial digital data exchange partners. This tool is essential to TIGER modernization as only limited spatial metadata are available in the TIGER data base. A variety of update operations regularly feed into the TIGER data base through partnerships like the Digital Exchange (DEX) program that incorporates non-census geospatial digital data files, via on-going programs such as the Boundary and Annexation Survey (BAS), census updates from the Local Update of Census Addresses (LUCA), local map reference sources, and various field listing and canvassing operations. Each of these feature update operations varies from the next in its collection methodology and its overall spatial accuracy. Additional errors may be introduced at the annotation capture, the processing, and the digitizing phases of each operation. For example, many of the TIGER creation files came from scanned map images with known accuracy standards, most pre-1995 updates to the TIGER data base were made using a digitizing tablet, and finally, most non-DEX updates made after 1995 were digitized freehand. The result is that the coordinate accuracy in TIGER varies greatly by update operation and geography.

The GEO uses the Global Positioning Satellite (GPS) System to assess the spatial accuracy of the TIGER data base in its preparation for TIGER modernization. To this end, the GEO developed a tool that evaluates the spatial accuracy of attributes derived from a variety of operations and sources. This tool is called the GPS TIGER Accuracy Analysis Tool or GTAAT. HTE-UCS, Inc.,[1] a contracting company, created GTAAT to the specifications of the TIGER Operations Branch (TOB) and the Geospatial Research and Standards Staff (GRaSS). The GTAAT utilizes differential correction files created by the GPS Cartographic software (GPSC) also created by HTE-UCS, Inc. and Trimble’s Pathfinder Office 2.51 software. The GTAAT calculates the distance and azimuth difference between the GPS collected point and the equivalent TIGER 0-cell (point.) The results from GTAAT for the eight test sites were used to assess the TIGER data base spatial accuracy described in this document.

[1] *HTE-UCS, Inc. has developed several applications for the Census Bureau that integrate TIGER/Line® map feature data with the Master Address File address data on mobile and desktop computers. The applications allow users to view, edit, add, and delete map feature and address data to support a variety of Census Bureau programs. The applications include Census 2000, Automated Local Update of Census Addresses (LUCA), LUCA Reconciliation, Automated Listing and Mapping Instrument, GPS Cartographic, and GTAAT.

I. Operation Workflow

In order to evaluate the coordinate quality in the TIGER data base, we first needed to gather the Global Positioning Satellite (GPS) data points in the field and compare them to the same points in the TIGER data base. The current operation workflow includes the following steps:

1. The TIGER Operations Branch (TOB) identifies a study area.
2. The TIGER Systems Branch (TSB) creates a TIGER/Line® file used by the GPS Cartographic software (GPSC) with GPS TIGER Accuracy Analysis Tools (GTAAT) Host Application.
3. The Geospatial Research and Standards Staff (GraSS) creates the Common (anchor) point file with source and history codes from the intersecting lines (1-cells) in the TIGER data base.
4. The HTE-UCS, Inc. suite of tools including the GPSC software with GTAAT Host Application, the GPS Cartographic Mapping Application, and Trimble's Pathfinder Office 8 (GPS post-processing differential correction) software were used to produce the data in this analysis.

HTE-UCS, Inc. is currently developing GPSC Tools 2, which will incorporate three elements (steps 2, 3, and 4 above) of the four elements and further facilitate data processing. Figure 1 illustrates the workflow for the GPSC Tools 2 based on the current GPSC/GTAAT software.

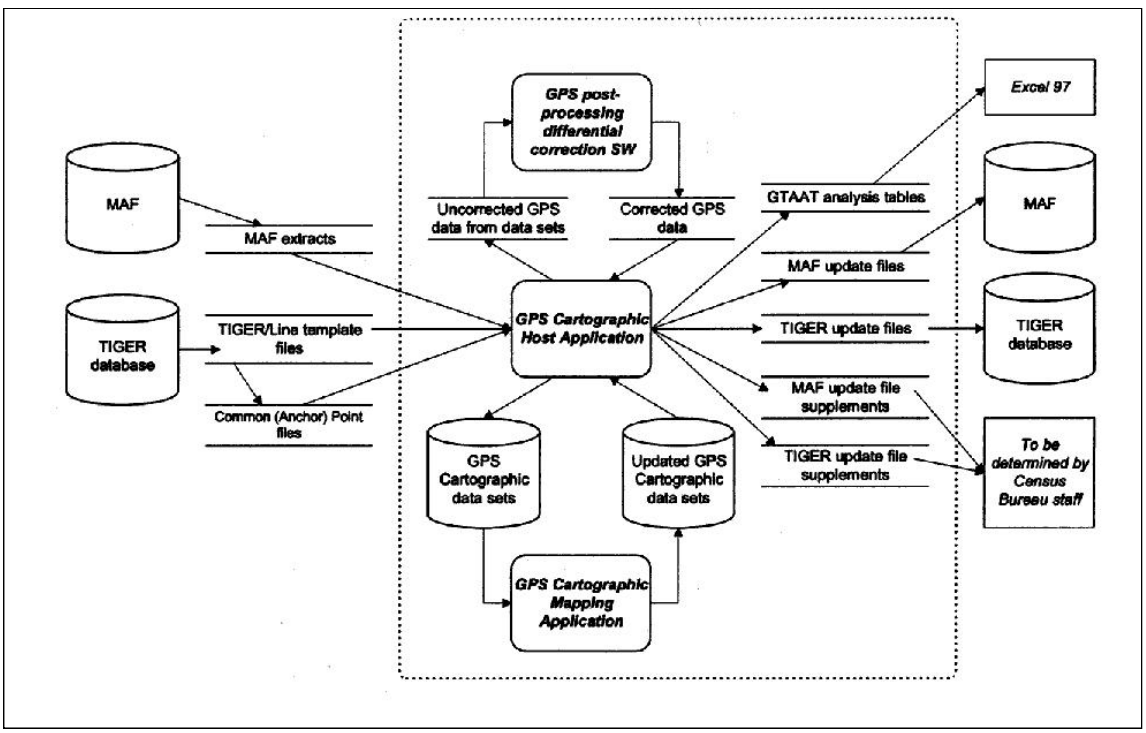


Figure 1. Proposed Workflow of GPS Cartographic Tools 2[2]

[2] This diagram was provided by HTE-UCS, Inc. The operation workflow for GPS Cartographic Tools 2 is still under development and subject to change.

II. Site Selection

For the Windham County field-test, the primary objective was to evaluate the GPSC software and the GPS field coordinate collection methodology while collecting county-wide GPS anchor point data. For the seven additional sites, the experience gained in the Windham County field-test were applied to the procedures used by each data collection team (rover team). The eight test sites met the following criteria:

- Digital files available (potential DEX partners),
- Updates from a variety of operations (source), e.g. Master Address File Geocoding Office Resolution (MAFGOR), Local Update of Census Addresses (LUCA), etc.
- A nearby civil airport to facilitate travel arrangements,
- A varied terrain, tree cover, etc. (to observe the effect on GPS signals, etc.)

The following sites were selected for evaluation:

Site County	State	Primary Update Operations
Maricopa	Arizona	Address List, GBF/DIME Contractor, LUCA Verification, and Digital Pilot
Sacramento	California	GBF/DIME Contractor, MAFGOR, and 1990 Geography Review
Hillsborough	Florida	GBF/DIME Contractor and 1990 Enumerator Updates
St. Tammany	Louisiana	USGS and FCM Updates
Clark	Nevada	1990 Enumerator Updates, Digital Exchange Pilot, and GBF/DIME Contractor
Delaware	Ohio	TIP and GBF/DIME Contractor
York	Pennsylvania	GBF/DIME Contractor and TIP
Windham	Vermont	USGS and FCM Updates

This document highlights the analysis from the GTAAT tables. Diagrams that show the workflow and process of the field tests as background to this evaluation have been included. For a detailed description of preprocessing, field operations and methodologies, and post-processing please refer to the documents listed below.

- *Census 2000 Pen Computer Field Test Project - System Design Document,* UCS, 10.30.1994
- *Census 2000 Pen Computer Field Pilot Test Project - User Screen Document,* UCS, 7.1.1995
- *Census 2000 Field Mopping and Address Listing Host - Operator's Manual,* UCS, 7.1.1995
- *TIGER/Line Files, 1998 Technical Documentation,* Bureau of the Census
- *GPS Cartographic Requirements Analysis Document,* version 1.0, HTE-UCS, 7.16.1999
- *GPS Cartographic System Design Document,* version 1.01, HTE-UCS, 8.30.1999
- *GTAAT Requirements Analysis Document,* version 1.0, HTE-UCS, 9.20.1999
- *GTAAT System Design Document,* version 1.0, HTE-UCS, 10.25.1999

The GPSC stores the GPS coordinates (latitude and longitude) of every captured anchor point, along with the original (TIGER or DEX file) coordinate pairs. The coordinates are stored as one set of attributes linked to the anchor point record. Additional GPS coordinate pairs can be collected for any given anchor point if necessary. The GPSC captures the coordinates for new street features, however, since there are no equivalent TIGER data base coordinate pairs they are not used in this analysis.

The collected GPS coordinates were post-processed to correct for the Selective Availability[3] of the GPS signal, GPS Satellite Clock errors, Ephemeris data errors, tropospheric delays[4], and unmodeled ionospheric delays[5]. Differential corrections in post-processing remove common-mode errors, those errors common to both the reference and remote receivers. They do not, however, correct for multipath[6] or receiver noise. Trimble's Pathfinder Office 2.51 software was used to differentially correct and post-process the captured anchor points. This process requires downloading appropriate data from a GPS base station. It is, therefore, critical to insure that a local base station is available during the data collection period. If a base station is out of service during the data collection period, data are rendered unusable unless data from an alternate local base station are available. We encountered this situation in Windham and thus lost some valuable data. We recommend constant monitoring to insure that data are not lost or unusable for any particular operation in the field; this may require allowing for additional time to recapture lost data.

A. Windham:

The Windham County field test was conducted from October 24 through October 30, 1999. Fifteen staff members from the GRaSS, Address List Review Branch, Update Operations Branch, Geographic Areas Branch, and TOB along with observers from American Community Survey, Geographic Data Technologies, Inc., and HTE-UCS, Inc. participated. Detailed findings from the Windham County test are available in the **GPS Cartographic Software Field Test Report** produced by HTE-UCS, Inc

---

[3] Selective Availability (SA) is the intentional degradation of the Standard Positioning Service signals by a time varying bias. SA is controlled by the DOD to limit accuracy for non-U.S. military and government users. SA was permanently turned of on May 2, 2000, however, the data referenced in this document predates this occurrence and therefore is subject to errors introduced by SA.

[4] The troposphere is the lower part (ground level to from 8 to 13 km) of the atmosphere that experiences the changes in temperature, pressure, and humidity associated with weather changes. Complex models of tropospheric delay require estimates or measurements of these parameters.

[5] The ionosphere is the layer of the atmosphere from 50 to 500 km that consists of ionized air. The transmitted model can only remove about half of the possible 70 nanoseconds of delay leaving a 10 meter un-modeled residual.

[6] Multipath is caused by reflected signals from surfaces near the receiver that can either interfere with or be mistaken for the signal that follows the straight line path from the satellite. Multipath is difficult to detect and sometime hard to avoid.

Footnotes 4, 5 and 6 are from **Global Positioning System Overview** by Peter H. Dana and are found at the following URL: http://www.utexas.edu/depts/grg/gcratt/notes/gps/gps.html.

Windham County, VT was selected for the following reasons[7]:

1. The State of Vermont has provided a file from their E-911 work that is stated to have better coordinates than TIGER.
2. The Geography Division (GEO) has a working relationship with the State of Vermont on other digital exchange activities, including a joint project testing the Road Data Model.
3. The county has three types of enumeration areas (TEAs).
4. The county is mostly rural with only two towns of any size.
5. Because the field test is only to be performed within specified areas, the resources designated for the project appeared to be adequate to perform the test within a 1 week period.

The field test in Windham County, VT had several goals[8]:

1. to develop and refine techniques that will provide a specific quantifiable measure of the coordinate accuracy of TIGER files for different parts of the Nation,
2. to be the first of a sample set of counties in which the coordinate accuracy is measured,
3. to test the contractor provided software for enhancing the Master Address File (MAF) through the collection of housing structure coordinates,
4. to test the parameters of using GPS-computer combinations to collect new features and their attributes, and
5. to provide coordinates for specified TIGER feature intersections and feature center lines as part of current TIGER coordinate improvement research underway in GRaSS.

Some highlights from the Windham County test include:

1. Staff divided the workload (census tract based) and assembled the field collection (rover) teams.
2. Six to seven rover teams collected GPS data for the entire county in 5 days.
3. Staff administered the operation from the home base at the hotel and ran the differential correction software on the collected data daily.
4. The rover teams collected nearly 5,000 anchor points, however some were not useable because the local base station was shutdown during part of the data collection period.

[7] Referenced from September 24, 1999, joint TOB/GraSS Memorandum, **Proposal to Field Test Selected Operational Parameters using GPS Enhanced Pen-based and Laptop Computers to collect Geographic Data in Windham County, VT**.
[8] ibid.

B. Other Sites:

- A single team (driver and data collector) completed selected census tracts within the sampled county.
- It typically took 1 or 2 days to complete each census tract.
- GPS data were differentially corrected when the team returned to headquarters.
- Average site data collection varied from 483 to 981 points.
- Daily communication with headquarters and a backup rover unit became necessary as collection time was limited and we wanted to minimize down time.
- Very few data points were not useable due to base station downtime.

The number of collected anchor points analyzed in this study by site[9]:

Site	Total	Site-by-Site Source Analysis	Source-by-Source Analysis
Maricopa	845	798	829
Sacramento	856	804	840
Hillsborough	614	530	614
St. Tammany	606	562	598
Clark	981	924	962
Delaware	483	335	483
York	804	768	798
Windham	1662	1571	1653
TOTAL:	**6851**	**6292**	**6777**

Not all collected and useable data points were used in this analysis. The less frequently encountered feature update source codes, those with less than 45 sample occurrences, were excluded because they were not statistically reliable for this sample. This is important to note, as data analysis tables and summaries for individual sites may differ from the totals presented in the national data set. Every effort has been made to assure the completeness and accuracy of the data presented in this document.

[9] The numbers in the **Total** column are the total number of GPS captured anchor points by test site.

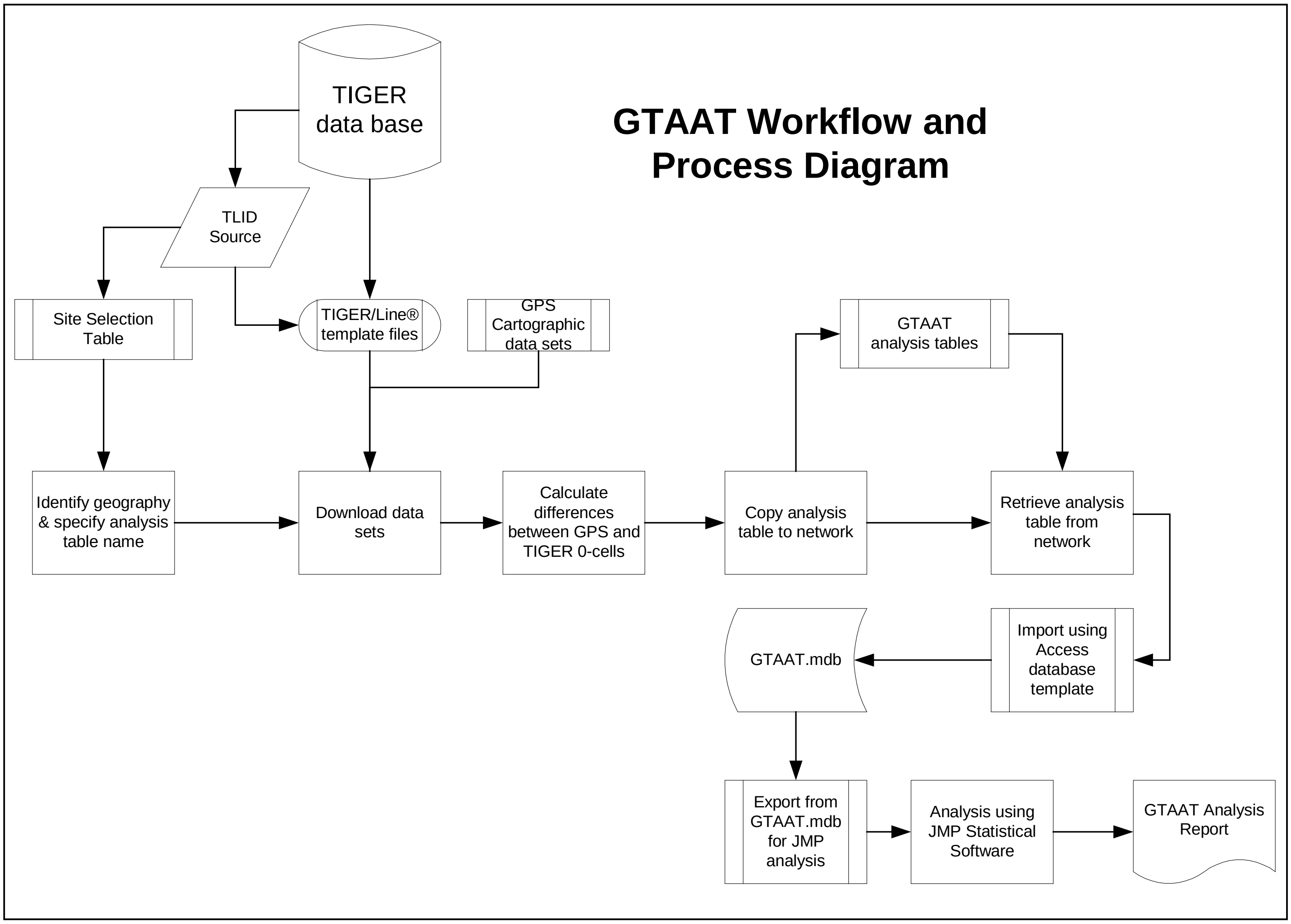
The data in the **Site-by-Site Source Analysis** column are the number of points used for update source analysis within the Site-by-Site section. The last column, "Used for overall Source by Source Analysis" lists the total number of anchor points used in the analysis for the Source-by-Source section. The difference in the counts between the two is due to a smaller site by source sample than the overall source-by-source sample. For example, there are 862 FCM Updates in the universe sample, 18 from Hillsborough, 155 from St. Tammany, and 689 from Windham. Although all 862 anchor points are used in the universe analysis, the 18 in Hillsborough are not used in the Site-by-Site analysis because they fall below the reliability threshold. Some 91.8 percent of the anchor points were used for the Site-by-Site analysis while 98.9 percent were used in the universe.

C. Source/Operation

The following table displays the number of collected GPS anchor points by Source or Update Operation for the entire 8-site universe. The numbers in red reflect source codes that were not used in the Source-by-Source analysis. These operations were included in the Site-by-Site and Tract-by-Tract summary totals.

Source/Operation	**Abbreviation**	**Points**
1980 Boundary Insertion	80BDRY	15
1990 Collection Geography Insertion	90COLL	84
1990 Enumerator Update	90ENUM	488
Census 2000 Address List	ADRLIST	170
1998 Boundary and Annexation Survey	BAS98	80
Census 2000 Block Canvassing	BLK CNVS	62
Census Map Preview	CMP	109
DFB Review of 1990 TIGER Updates	DFBREV	16
Digital Exchange Pilot 1993-1994	DIGPILOT	370
Digital Exchange Review and Correction	DX REV	60
Feature Change Map Update	FCM	862
FID Improvement	FID IMPV	1
GBF-DIME Contractor	GBF-C	1874
1990 Geocoding Operations	GEO90	29
Geography Division Review of Post-census Maps	GEOREV90	137
Unconfirmed Local Updates	LOCAL	53
LUCA Updates	LUCAUPD	88
LUCA Verification Updates	LUCAVERI	110
MAFGOR/DEX Reference	MAF/DEX	13
Master Address File Office Resolution	MAF_ORES	577
Restructure 3 Clean-Up	RSTRUCT3	55
TIGER Improvement Program	TIP	270
United States Geological Survey 100K	USGS	1328
TOTAL:		**6851**

III. The GTAAT



The **GTAAT Workflow and Process Diagram** details the various steps, some manual and some automated, that are part of the GTAAT process. GRaSS developed software that creates a unique anchor point id for each intersecting type "A"[10] Census Feature Classification Code (CFCC) for the front-end processing prior to field GPS data capture. The GRaSS also developed software that links the appropriate source code to each TIGER/Line® Record Identification Number (TLID)[11]. Both are vital components of the GTAAT. Data from the GPSC field collection are imported into the GTAAT software. The GTAAT calculates the differences in distance and azimuth between the TIGER data base 0-cells and their corresponding field "post-processed differentially corrected" GPS anchor

[10] The "A" CFCC indicates it is a road feature.
[11] The TIGER/Line® Record Identification Number is the unique identifier for every line segment in the TIGER data base and TIGER/Line® products.

point coordinates (ground truth.) All distances are reported in meters and azimuths in degrees.

The GTAAT analysis tables are created and exported to an ASCII format file. For each test site, a file is created containing the following fields:

ANCHOR_ID	Anchor Identification Number
ST_CODE	State FIPS Code
CNTY_CODE	County FIPS Code
TRACT	Census Tract Number
DSTAMP	Date Stamp from GPS
LONG1	TIGER Longitude
LAT1	TIGER Latitude
GPS_LONG1	GPS Longitude
GPS_LAT1	GPS Latitude
DISTANCE	Calculated Distance (TIGER 0-cell to GPS) in meters
AZIMUTH	Calculated Azimuth (TIGER 0-cell to GPS) in degrees
TLID1	TLID of 1st Intersecting 1-Cell[12]
TLID2	TLID of 2nd Intersecting 1-Cell
TLID3	TLID of 3rd Intersecting 1-Cell
TLID4	TLID of 4th Intersecting 1-Cell
TLID5	TLID of 5th Intersecting 1-Cell
CFCC1	CFCC[13] of 1st Intersecting 1-Cell
CFCC2	CFCC of 2nd Intersecting 1-Cell
CFCC3	CFCC of 3rd Intersecting 1-Cell
CFCC4	CFCC of 4th Intersecting 1-Cell
CFCC5	CFCC of 5th Intersecting 1-Cell
SRC_CODE1	Source Code of 1st Intersecting 1-Cell
SRC_CODE2	Source Code of 2nd Intersecting 1-Cell
SRC_CODE3	Source Code of 3rd Intersecting 1-Cell
SRC_CODE4	Source Code of 4th Intersecting 1-Cell
SRC_CODE5	Source Code of 5th Intersecting 1-Cell
CONFIDENCE	Data Confidence Code

The GTAAT analysis tables are imported into a Microsoft Access database (GTAAT.mdb) where they are sorted and analyzed for errors. The GTAAT.mdb is linked to the TLID source database to capture the source and history ID codes. The resulting file is exported as a comma delimited file. The comma-delimited file is then imported into the JMP Statistical Software Package and Excel for statistical analysis and graph creation. The graphs, charts, and tables that are found in the analysis section were created in JMP and Excel.

[12] A 1-cell is a line segment composed of a beginning and an ending point (0-cell) and may contain one or more shape points.
[13] The Census Feature Classification Code (CFCC) is a three-character code developed by the Census Bureau to identify the most prominent characteristics of a feature. The first character is a letter describing the feature class; the second and third characters are numbers representing the major and minor categories.

IV. GTAAT Analysis

The eight sites in this test contain a broad spectrum of TIGER update sources, a mixture of urban, suburban, and rural areas, high growth and low growth areas, and varied terrain from mountainous to flat. The test sites are not a statistical sample that represents the nation as a whole. We are confident however, that it represents the plethora of spatial accuracy differences that are found throughout the TIGER data base. This analysis is presented in two distinct sections; one examines the data on a site-by-site basis and the other explores a source-by-source comparison.

A. Site-by-Site Template

This section describes the layout of the Site-by-Site Comparison. The following are found under each test site heading:

- Figure A1: A county map showing the census tract selected for this analysis.
- Figure A2: A source by tract cross-tabulation of collected anchor points.
- Figure A3: Displays the overall distribution of variance (TIGER to ground truth.)
- Figure A4: Displays the distribution of variance by source within the site.
- Figure A5: Displays the distribution of variance by census tract within the site.
- Additional figures are included for selected test sites.

In addition to the figures and tables, there is a brief description of any unique findings for each site. An example of the layout follows:

County name, ST (STCOU)

Test dates: January 1, 2000 to January 3, 2000

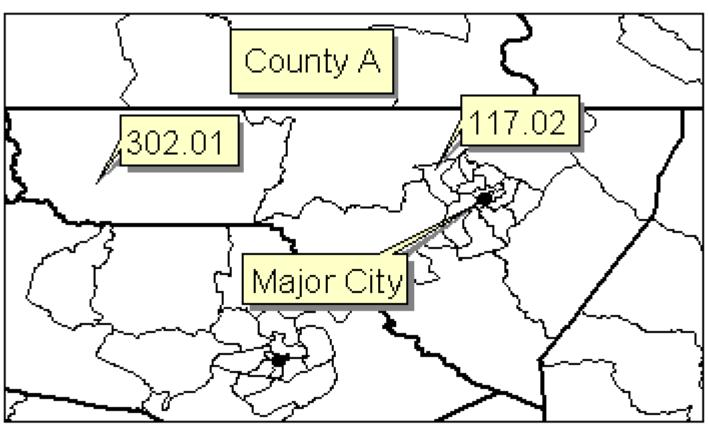
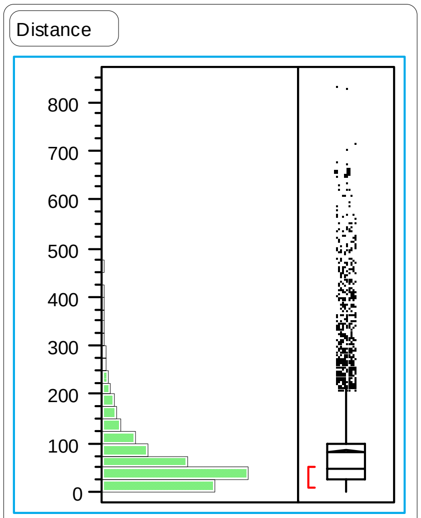


Figure IVA1 displays the location of the selected test census tracts in county A.

County A was selected because it has a variety of update operations including:

Source/Operation	Tract 1	Tract 2	Total	Percentage
90COLL	135	92	227	50.1%
ADRLIST	16	116	132	29.1%
LUCAVERI		51	51	11.3%
MAF_ORES	21		21‘	4.6%
Others	3	19	22	4.9%
TOTAL	175	278	453	100%

Figure IVA2. County A: Source by tract cross-tabulation of collected anchor points.



Quantiles

maximum	100.0%	833.31
	99.5%	555.22
	97.5%	405.46
	90.0%	199.24
quartile	75.0%	101.19
median	50.0%	50.76
quartile	25.0%	27.60
	10.0%	16.57
	2.5%	7.91
	0.5%	3.24
minimum	0.0%	0.13

Moments

Mean	85.662
Std Dev	98.974
Std Error Mean	1.196
Upper 95% Mean	88.006
Lower 95% Mean	83.318
N	6851.000
Sum Weights	6851.000

Figure IVA3. County A

This figure and all subsequent "A3" figures in section V. Site-by-Site, contain four parts:

- The upper left portion contains a histogram[14] of the distance difference in meters between the TIGER data base and the respective collected GPS points (ground truth).
- The upper right portion contains an outlier box plot[15] for this sample.
- The middle portion contains the quantile report[16] for this sample.
- The bottom portion contains the moments report[17] for this sample.

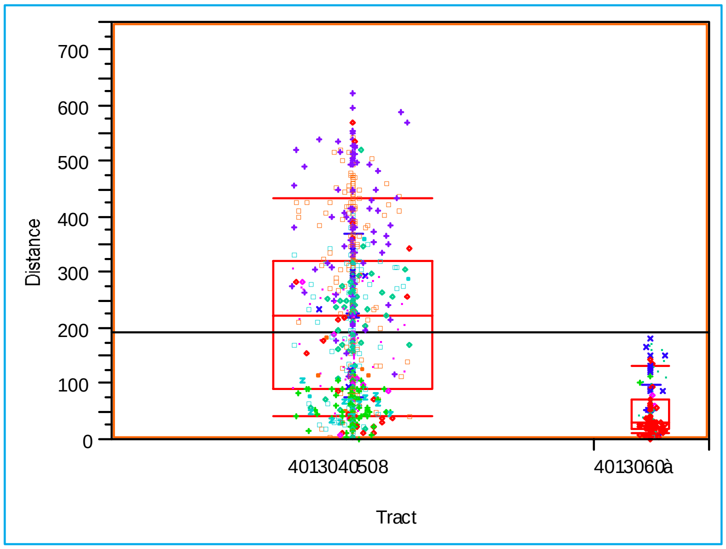
[14] Each bar shows the frequency of occurrence of the value or range of values represented on the axis. If the variable is continuous, the axis is broken into intervals. If the variable is nominal, a bar represents each discrete value.

[15] The Outlier Box Plot is a schematic that lets you see the sample distribution and identify points with extreme values, or outliers. The ends of the box are the 25th and 75th quantiles, also called the quartiles. The difference between the quartiles is the interquartile range. The line across the middle identifies the median sample value. The ends of the whiskers are the outer-most data points from their respective quartiles that fall within the distance computed as 1.5 *(interquartile range). The bracket along the edge of the box identifies the shortest half, which is the most dense 50 percent of the observations.

[16] Quantiles are values that divide a distribution into two groups where the Pth quantile is larger than Percent of the values. For example, half the data are below and half the data are above or equal to the 50th quantile, also called the median.

[17] The Moments report displays the mean, standard deviation, and other summary statistics. If the sample is normally distributed, the values are arranged symmetrically around the mean (arithmetic average). The mean and standard deviation completely define a normal distribution, giving an easy way to summarize data as the sample becomes large.

Distance By Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
4013040508	3.24	42.78	91.245	223.24	324.0975	436.915	626.57
4013060900	2.63	14.52	20.515	33.79	72.225	135.486	183.27

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
4013040508	684	225.516	147.839	5.6527
4013060900	161	54.089	45.942	3.6207

Table IVA4. County A, Distance by Tract.

Table IVA4. and subsequent "A4" figures in the section, V. Site-by Site, section contain three parts:

- A quantile box plot[18] of the collected points by census tract.
- A quantile analysis report[19] by census tract.
- The means and standard deviation report[20] by source code.

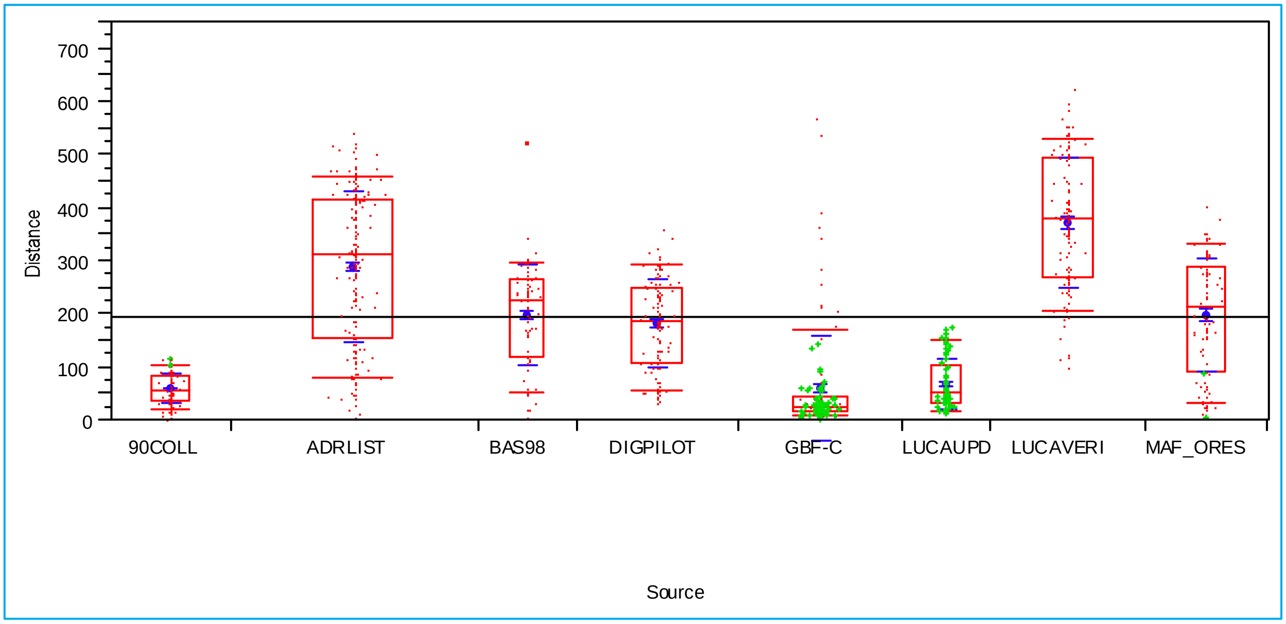
[18] The quantile box plot graphically summarizes the distribution of points at each factor level. The box shows the group median as a line across the middle and the quartiles (25th and 75th percentiles) as its ends. The 10th and 90th quantiles are shown as lines above and below the box. The scatter plot shows the means of each group with a large marker and error bars appear one standard error above and below each group means. The dashed lines above and below the box plot identify one standard deviation above and below the group means.

[19] The quantiles analysis report shows a table listing the minimum, 10 percent, 25 percent, median (50 percent), 75 percent, 90 percent, and maximum values of each group.

[20] The means and standard deviation report displays a table that lists the sample size, mean, standard deviation, and standard error of each group.

The anchor point colors in the "A3" and "A4" type tables represent the source or update operation for the GPS captured anchor point. Attachment A contains a legend depicting the symbols and colors for every source referred in this document.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
90COLL	3.24	20.206	38.36	58.19	87.2	106.238	121.23
ADRLIST	6.4	80.319	158.1725	316.795	420.1775	462.01	546.46
BAS98	5.13	52.64	119.515	227.7	267.635	298.27	523.93
DIGPILOT	33.63	57.204	110.99	186.39	253.54	293.766	364.15
GBF-C	2.63	10.828	17.62	26.35	44.13	171.91	574.04
LUCAUPD	13.91	19.784	34.31	53.15	106.57	154.024	175.99
LUCAVERI	103.04	208.192	273.55	382.66	495.65	533.47	626.57
MAF_ORES	7.47	34.872	91.895	215.04	289.54	335.472	405.72

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
90COLL	83	61.109	30.589	3.358
ADRLIST	170	293.357	145.276	11.142
BAS98	69	199.500	97.636	11.754
DIGPILOT	111	183.984	84.928	8.061
GBF-C	113	61.247	99.551	9.365
LUCAUPD	60	70.252	48.307	6.236
LUCAVERI	107	374.342	123.905	11.978
MAF_ORES	85	201.078	110.493	11.985

Table IVA5. County A, Distance by Source[21]

[21] Source refers to the operation that a feature was added to the TIGER data base. For example, if a BAS official submits a feature update, it is added to TIGER using the BAS option in GusX. Subsequently the feature is tagged with the BAS source code. Source and update operations are used interchangeably in this document. Attachment B contains a list of sources and the approximate dates they were inserted into the TIGER data base.

Table A4. and all subsequent “A4” tables (Distance by Source) in the section V. Site-by Site, contain three parts (similar to the “A3” tables):

- A quantile box plot of the collected points by source code (update operation)
- A quantile analysis report by source code.
- The means and standard deviation report by source code.

The color of the points in the “A4” tables represents the census tract of the GPS captured anchor point. Note this is different from the “source” color scheme.

All quantile box plots and tertiary graphs that compare tracts or counties use the “source” color schema. All quantile box plots and tertiary graphs that compare source data use the “tract” color schema. Attachment F contains a legend defining the symbols and colors for both.

Additional maps, graphs, or tables are included for each test site when necessary to present or clarify important issues.

V. Site-by-Site Comparison

A. Maricopa County, AZ (04013)

Test dates: December 15, 1999 through December 17, 1999

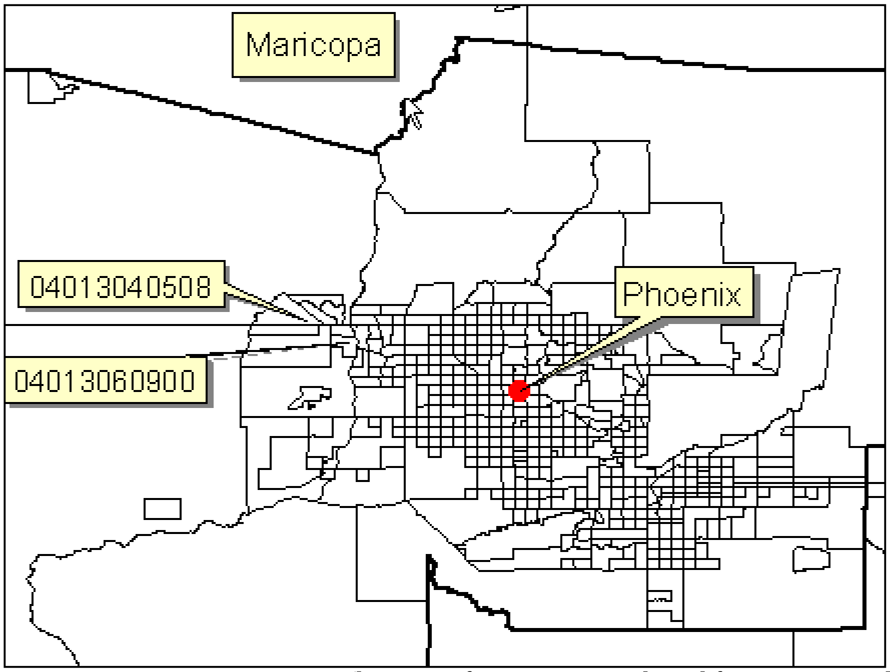


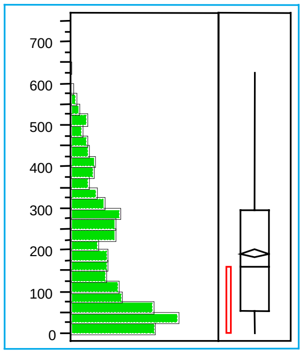
Figure VA1. Maricopa County (location of census tracts selected for GTAAT analysis)

Maricopa County was selected because it has high growth areas and contains features obtained through a variety of update operations including:

Operation	405.08	609	Total	Percentage
0COLL	81	2	83	9.8%
0ENUM	8	12	20	2.4%
DRLIST	170		170	20.1%
AS98	69		69	8.2%
FBREV	16		16	1.9%
IGPILOT	111		111	13.1%
BF-C	30	83	113	13.4%
UCAUPD		60	60	7.1%
UCAVERI	107		107	12.7%
AF_ORES	83	2	85	10.1%
STRUCT3	9	2	11	1.3%
OTAL	684	161	845	100%

Figure VA2. Maricopa County (collected anchor points by source and tract)

Distance



Quantiles

maximum	100.0%	626.57
	99.5%	574.02
	97.5%	519.44
	90.0%	418.67
quartile	75.0%	296.07
median	50.0%	161.52
quartile	25.0%	55.69
	10.0%	25.28
	2.5%	11.49
	0.5%	5.44
minimum	0.0%	2.63

Moments

Mean	192.8538
Std Dev	150.4157
Std Error Mean	5.1745
Upper 95% Mean	203.0103
Lower 95% Mean	182.6973
N	845.0000
Sum Weights	845.0000

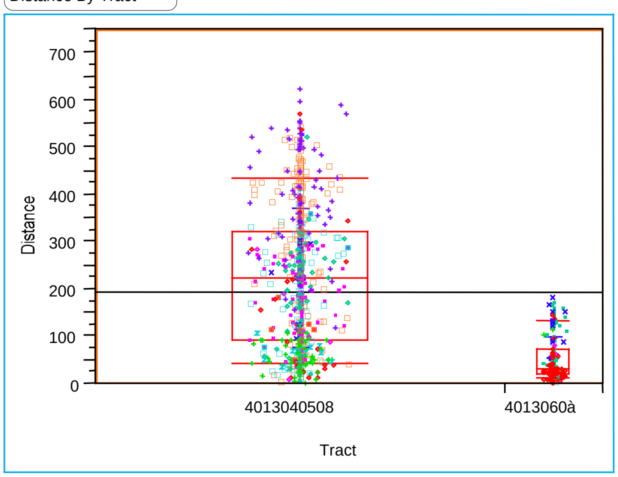
Figure VA3. Maricopa

The two selected tracts have many recent updates from Address List, BAS 98, LUCA, and MAFGOR operations. These updates represent 58 percent of all GPS anchor points collected. Because they are post-1990 updates, they were added to the TIGER data base using free-hand digitizing resulting in a mean distance difference of 196.48 meters from ground truth. The figures that follow display the distribution of collected data points and their variance from ground truth by site and by operation.

Figure, VA3. Maricopa County, examines the distribution of the variance (TIGER to ground truth.) It is an irregular distribution with two variance peaks at 25-50 meters and at 275-300 meters. The lower peak is attributable to updates from the GBF-C, 90 Collection, and LUCA Update operations while the higher peak results from the Address List and LUCA Verification Operations.

Figure VA4. Distance by Tract, compares census tracts 405.08 and 609. The tracts have remarkable differences, the first is a new growth area with an overall **m**ean **d**istance **d**ifference (mdd) of 225.5 meters from ground truth while the second has a 45.9 meter mdd. ***Figure VA5, Maricopa County (Source by Tract)*** shows the largest update source by census tract and the distance from selected places north west of Phoenix. It also shows the location of the test tracts.

Distance By Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
4013040508	3.24	42.78	91.245	223.24	324.0975	436.915	626.57
4013060900	2.63	14.52	20.515	33.79	72.225	135.486	183.27

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
4013040508	684	225.516	147.839	5.6527
4013060900	161	54.089	45.942	3.6207

Figure VA4, Distance by Tract

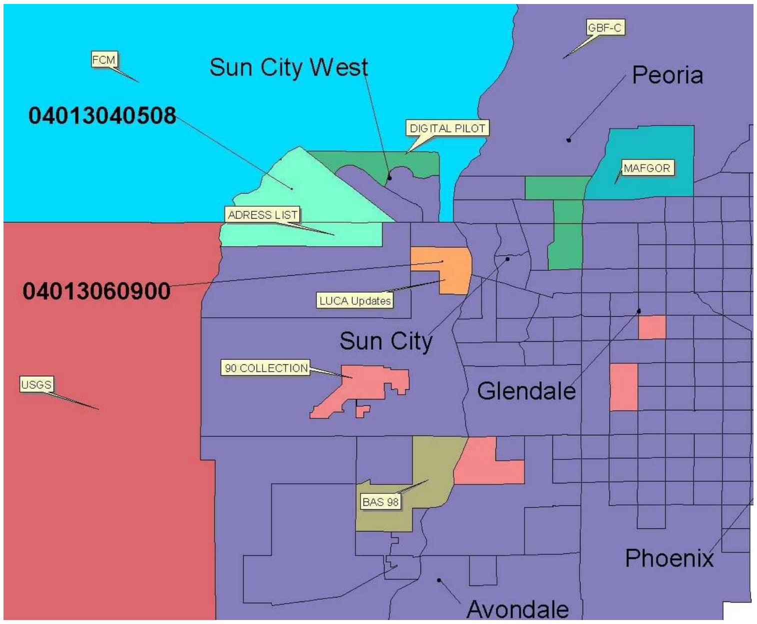
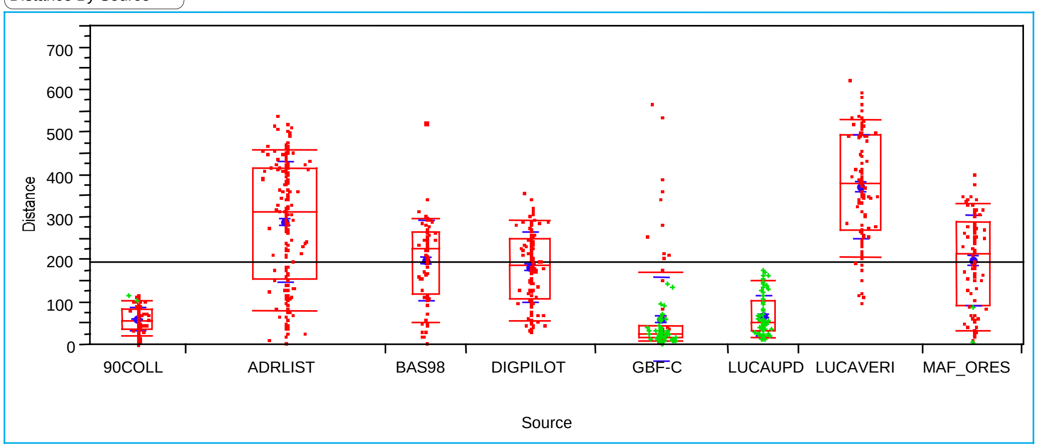


Figure VA5. Maricopa County (Source by tract)

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
90COLL	3.24	20.206	38.36	58.19	87.2	106.238	121.23
ADRLIST	6.4	80.319	158.1725	316.795	420.1775	462.01	546.46
BAS98	5.13	52.64	119.515	227.7	267.635	298.27	523.93
DIGPILOT	33.63	57.204	110.99	186.39	253.54	293.766	364.15
GBF-C	2.63	10.828	17.62	26.35	44.13	171.91	574.04
LUCAUPD	13.91	19.784	34.31	53.15	106.57	154.024	175.99
LUCAVERI	103.04	208.192	273.55	382.66	495.65	533.47	626.57
MAF_ORES	7.47	34.872	91.895	215.04	289.54	335.472	405.72

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
90COLL	83	61.109	30.589	3.358
ADRLIST	170	293.357	145.276	11.142
BAS98	69	199.500	97.636	11.754
DIGPILOT	111	183.984	84.928	8.061
GBF-C	113	61.247	99.551	9.365
LUCAUPD	60	70.252	48.307	6.236
LUCAVERI	107	374.342	123.905	11.978
MAF_ORES	85	201.078	110.493	11.985

Figure VA6. Maricopa County: Distance by Source

Figure VA6, Maricopa County, Distance by Source, compares the variance from source code to source code of anchor points collected in the field. For Maricopa County, the variance ranges from 61 meters for GBF-C and 1990 Collection to over 374 meters for LUCA Verification. Using current accuracy levels, the best one can hope for is a return visit to the wrong house (61 meters = 200 feet) or in the worst case, a return to the wrong block (374 meters =1227 feet).

Maricopa County, like many high growth counties has many updates from operations that utilize the more recent freehand digitizing method for updating. High growth areas with multiple update operations suffer from compound update problems resulting in variable spatial accuracy within a small area. This makes any automated spatial exchange complex and requires interactive manipulation of files. In cases where map spots are used, the problem is grave as it has implications on data allocations. Maricopa County is a good example of the compound update problems faced by high growth areas. The core area built using GBF/DIME files (GBF-C) has a 61-meter average accuracy. The core is expanded with MAFGOR updates, which have 200-meter mdd accuracy. Address List and LUCA Verification are subsequently added to MAFGOR update ring thus further deviating (>350 meters) from ground truth. This pattern will continue until corrective measures like conflation[22] or replacement occurs. To spatially correct the county it may be worthwhile to replace the current file with a local digital file and subsequently add census attributes to it, rather than using a digital file to conflate to using the rubber-sheeting methodology.

[22] Conflation is a global improvement of features in one coverage using the coordinates of features in the other coverage. This conflation may involve the rubber-sheeting of lines, points, shapes or other attributes, the matching of line segments or shapes, the selection of the best features from both coverages, and the transfer of these selected features from one coverage to the other.

B. Sacramento County, CA (06067)

Test dates: December 20, 1999 to December 21, 1999

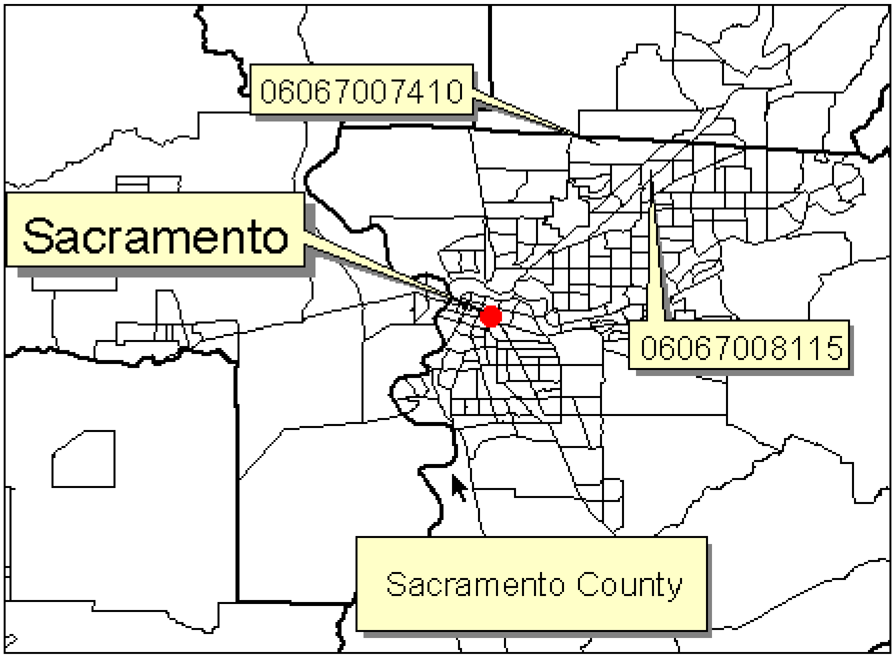


Figure VB1. Sacramento County (location of census tracts selected for GTAAT analysis)

Sacramento County was selected because it is a Census 2000 Dress Rehearsal site and contains modest amounts of updates since 1990 (<33 percent.) The following is the distribution of anchor points by census tract and source:

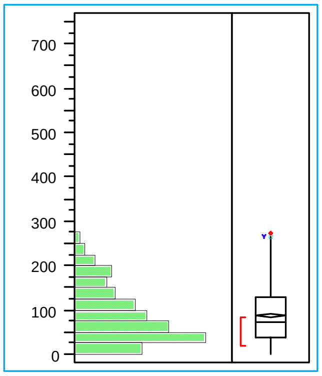
Operation	74.10	81.15	Total	Percentage
90ENUM	22		22	2.6%
BAS98		1	1	0.1%
BLK CNVS	3		3	0.4%
GBF-C	146	190	336	39.3%
GEO90	15	1	16	1.9%
GEOREV90	137		137	16.0%
LOCAL	45		45	5.3%
MAF_ORES	286		286	33.4%
RSTRUCT3	10		10	1.2%
TOTAL	**664**	**192**	**856**	**100%**

Figure VB2. Sacramento County (collected anchor points by source and tract)

(The totals shown in red reflect sources that were not used in the source-by-source analysis in Figure VB5)

Most of the features in Sacramento County were either added by the MAFGOR operation using reference materials and freehand digitizing or by contractors and clerks (GBF-C and GEOREV90) using a digitizing tablet (>55 percent) prior to 1990.

Distance



Quantiles

maximum	100.0%	276.96
	99.5%	267.24
	97.5%	231.29
	90.0%	187.45
quartile	75.0%	130.84
median	50.0%	73.34
quartile	25.0%	38.24
	10.0%	22.94
	2.5%	8.89
	0.5%	3.63
minimum	0.0%	1.72

Moments

Mean	89.1322
Std Dev	62.1120
Std Error Mean	2.1229
Upper 95% Mean	93.2991
Lower 95% Mean	84.9654
N	856.0000
Sum Weights	856.0000

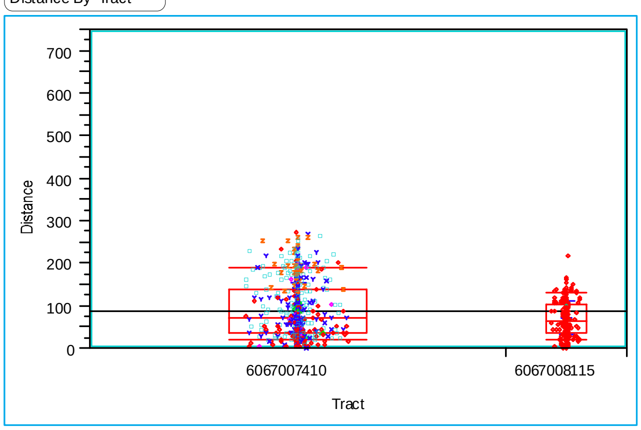
Figure VB3. Sacramento County

The table to the left is ***Figure VB3, Sacramento County,*** and it displays the distribution of variance for Sacramento County. Sacramento County has an 89.10-meter mdd, and like Maricopa County it has cluster peaks, one at the 25-50 meter mark and another smaller peak at the 175-200 meter mark. The first peak is attributable to updates from the Geographic Base File Contractor (GBF-C) updates and the second from Local Unconfirmed Updates (LOCAL) and MAFGOR updates.

Figure VB4, Sacramento County: Distance by Tract, (bottom half of this page) displays the differences between the two test tracts.

The updates for Sacramento County were distributed evenly thus the mdd of the two tracts varies slightly, 93.8 meters and 72.7 meters for census tract 74.10 and census tract 81.15 respectively. Compared to Maricopa County, Sacramento County is a better candidate for conflation because it has fewer and uniform updates. Fewer interactive updates are necessary when a file has a uniform spatial accuracy or inaccuracy. Variable levels of accuracy to either extreme make conflation extremely difficult.

Distance By Tract



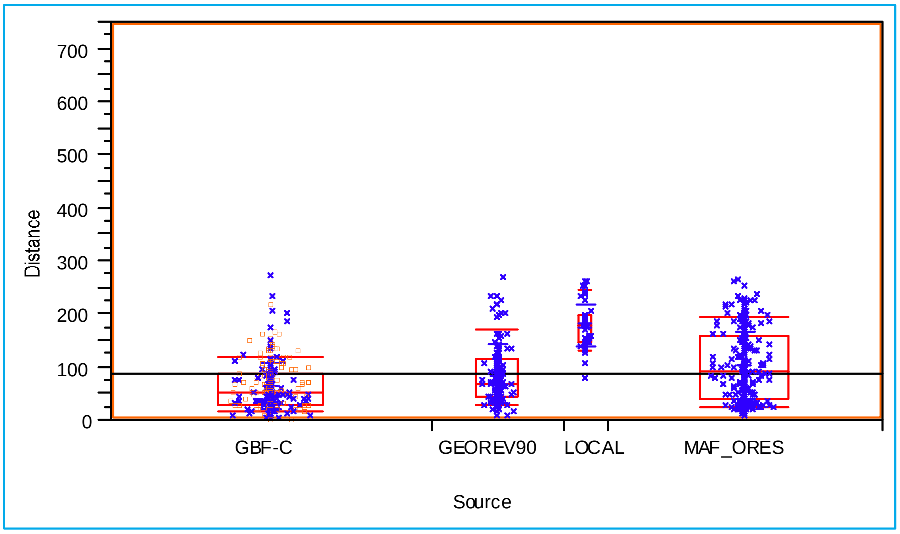
Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
6067007410	2.69	23.73	38.3225	75.71	141.835	195.095	276.96
6067008115	1.72	21.824	37.6925	67.1	107.95	133.432	223.16

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
6067007410	664	93.8694	66.0234	2.5622
6067008115	192	72.7495	42.3154	3.0539

Figure VB4. Sacramento County: Distance by Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
GBF-C	1.72	16.065	31.2475	52.9	91.875	123.154	276.96
GEOREV90	8.85	30.796	45.26	71.8	118.375	173.152	272.11
LOCAL	83.31	135.348	150.7	183.44	201.53	250.774	265.7
MAF_ORES	8.53	27.401	43.4725	92.94	162.795	198.602	270.93

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
GBF-C	336	65.049	45.1013	2.4605
GEOREV90	137	88.594	56.9494	4.8655
LOCAL	45	182.538	41.7156	6.2186
MAF_ORES	286	105.001	66.4269	3.9279

Figure VB5. Sacramento County: Distance by Source

Figure VB5, Sacramento County: Distance by Source, displays the median distance difference between the TIGER data base and ground truth by source code. The unconfirmed local updates, as expected, had the highest mdd with over 182 meters. The MAF_ORES (this is the TIGER data base source abbreviation for MAFGOR updates) operation is a distant second with 105-meter mdd. Both operations were concentrated in census tract 74.10 resulting in a 30 percent increase in the average mdd for that tract.

C. Hillsborough County, FL (12057)

Test dates: November 9, 1999 to November 10, 1999

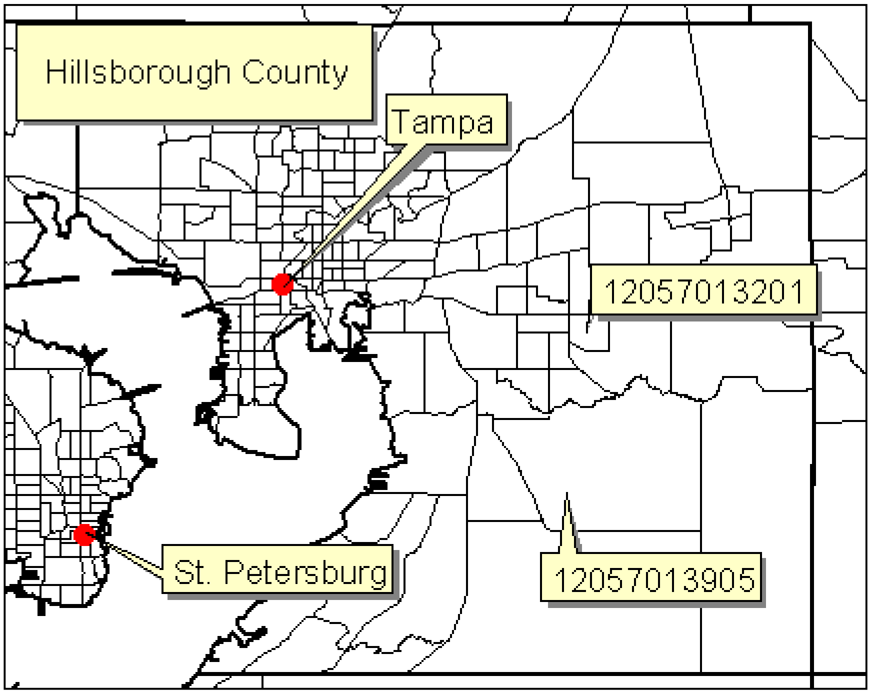


Figure VC1. Hillsborough County (location of census tracts selected for GTAAT analysis)

Hillsborough County, Florida was selected as a test site because the United States Census Bureau has an active DEX partnership with the county and the county actively participates in multiple Census operations. The test tracts had the following source code updates:

Operation	132.01	139.05	Total	Percentage
90ENUM	35	36	71	11.6%
FCM		18	18	2.9%
GBF-C	200	259	459	74.8%
MAF_ORES	19	6	25	4.1%
RSTRUCT3		2	2	0.3%
USGS		39	39	6.4%
TOTAL	**254**	**360**	**614**	**100%**

Figure VC2. Hillsborough County (collected anchor points by source and tract)

Most of the features (points) in the test census tracts were inserted prior to the 1990 census, specifically from the GBF Dime File and 90 Enumerator Updates programs (86.4 percent combined.) Unfortunately the TIGER/Line® file used was from the 1998 benchmark and did not have some of the more recent updates specifically from the

Address List program or from LUCA. These new features were added in the field using the GPS cartographic instrument but were not used for this analysis.

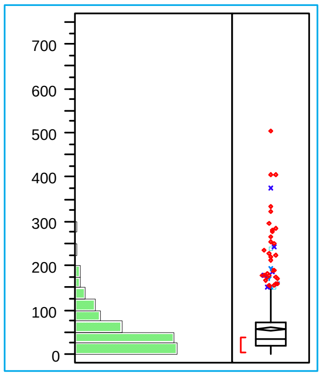
The LUCA participants from Hillsborough County provided digital files that were subsequently used to update features in the TIGER data base. A more recent benchmark of TIGER/Line® file was used during the post processing and the LUCA updates were visually inspected and found to be very consistent with the updates captured using the GPS Cartographic instrument.

The two test census tracts had a 58.4-meter mdd, with a distribution peak at the 0-50 meter range. The census tracts, 132.01 and 139.05 have a mdd of 46.1 meters and 67.1 meters respectively. Interestingly there were many outliers, some with over 100-meter mdd (>10 percent.) Some features in the TIGER data base that came from the U. S. Geological Survey (USGS) no longer exist or are tagged with incorrect CFCCs.

Figure VC3, Hillsborough County, displays the distribution of all collected anchor points for the test tracts. Notice the significant number of outliers shown as color dots above the one standard deviation mark. The majority of them are from census tract 139.05 and may result from bad GBF-DIME fringe digitizing by the contractor.

Figure VC4, Hillsborough County, Distance by Tract, shows

Distance



Quantiles

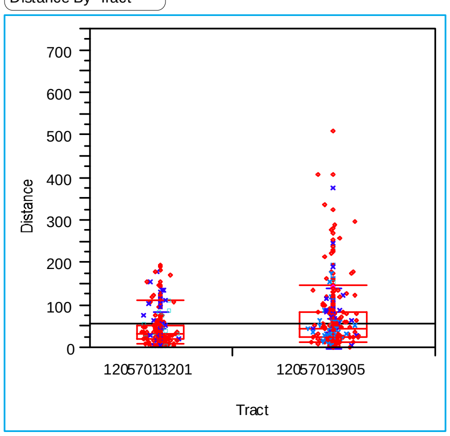
maximum	100.0%	509.43
	99.5%	407.55
	97.5%	238.76
	90.0%	128.79
quartile	75.0%	73.39
median	50.0%	37.46
quartile	25.0%	21.38
	10.0%	10.76
	2.5%	4.79
	0.5%	0.76
minimum	0.0%	0.13

Moments

Mean	58.3911
Std Dev	61.5606
Std Error Mean	2.4844
Upper 95% Mean	63.2701
Lower 95% Mean	53.5121
N	614.0000
Sum Weights	614.0000

Figure VC3. Hillsborough County

Distance By Tract



Quantiles

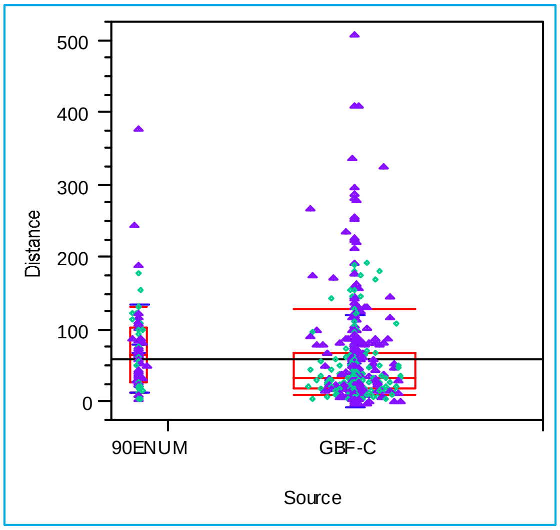
Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
12057013201	3.59	9.04	19.0025	32.415	53.4725	111.995	192.67
12057013905	0.13	12.698	23.785	45.18	83.525	147.585	509.43

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
12057013201	254	46.0608	41.5836	2.6092
12057013905	360	67.0908	71.1964	3.7524

Figure VC4. Hillsborough County: Distance by Tract the mdd comparison of collected anchor points between the two test tracts.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
90ENUM	4.56	14.916	27.48	66.2	104.54	133.388	378.49
GBF-C	0.13	10.3	19.81	35.55	69.23	129.2	509.43

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
90ENUM	71	75.2861	60.8440	7.2209
GBF-C	459	57.2326	63.8195	2.9788

Figure VC5. Hillsborough County: Distance by Source

Figure VC5, Hillsborough County, Distance by Source, provides details of the distribution of the mdd in meters of sampled anchor points by source. The GBF-C and 90 Enumerator updates have significant differences, the first has a 57.24-meter mdd and the second 75.29-meter mdd.

D. St Tammany Parish, LA (22103)

Test dates: November 5, 1999 to November 6, 1999

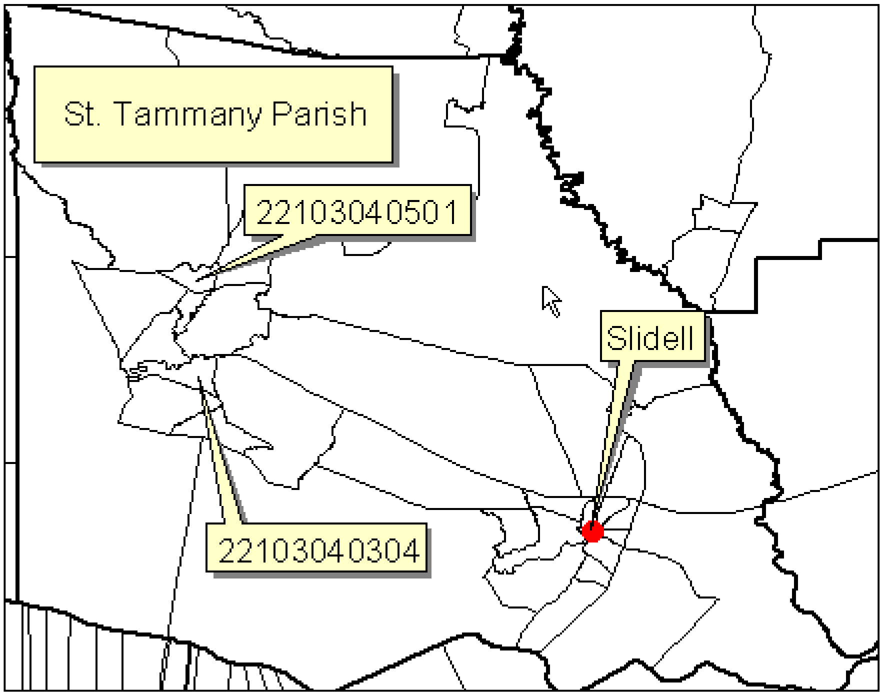


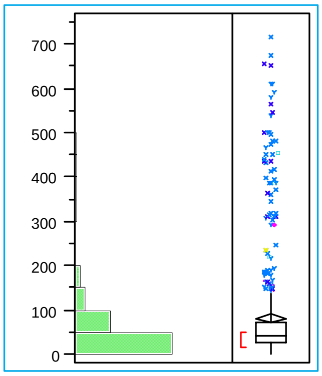
Figure VD1. St. Tammany Parish (location of census tracts selected for GTAAT analysis)

St. Tammany parish was selected because the GEO used a digital computer assisted design (CAD) file to plot maps for digitizing using the GusX software and because it had other significant update operations including:

Operation	403.04	405.01	Total	Percentage
80BDRY	2	4	6	1.0%
90ENUM	18	4	22	3.6%
DIGPILOT	3		3	0.5%
FCM	102	53	155	25.6%
MAF/DEX	2		2	0.3%
MAF_ORES	5	4	9	1.5%
RSTRUCT3	1	1	2	0.3%
USGS	126	281	407	67.2%
TOTAL	259	347	606	100%

Figure VD2. St. Tammany Parish (collected anchor points by source and tract)

Distance



Quantiles

maximum	100.0%	719.19
	99.5%	657.23
	97.5%	495.08
	90.0%	182.73
quartile	75.0%	74.16
median	50.0%	42.56
quartile	25.0%	27.89
	10.0%	15.39
	2.5%	7.11
	0.5%	2.37
minimum	0.0%	1.10

Moments

Mean	83.2933
Std Dev	118.6590
Std Error Mean	4.8202
Upper 95% Mean	92.7598
Lower 95% Mean	73.8268
N	606.0000
Sum Weights	606.0000

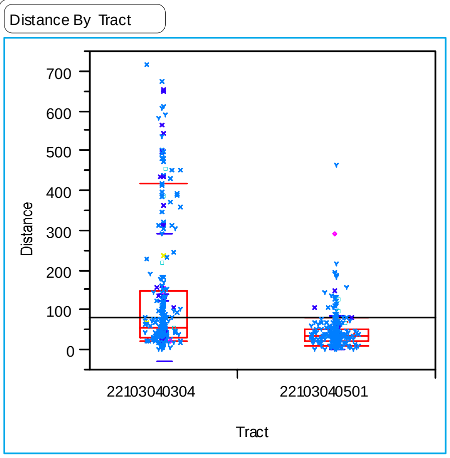
Like Hillsborough County, a 1998 TIGER/Line® benchmark file was used for the anchor point selection. The test file did not include any anchor points from the Address List, Block Canvass, or LUCA updates. The TIGER/Line used with the GPS Cartographic instrument, however, had many of these features as it was more recent and included updates from those operations.

A local street map was necessary for navigation purposes as the post-1990 TIGER updates were not drawn to scale in tract 403.04 and were often more than 500 meters from ground truth. This is reflected in ***Figure VD3, St. Tammany Parish,*** by the large numbers of outlier points with greater than 200-meter mdd.

The two census tracts evaluated in St. Tammany Parish were less than 5 miles apart. One census tract, 405.01, was inside the city of Covington. The other tract, 403.04, was part of a gated community north of the City of Mandeville. Features in the first tract were captured from USGS scanning and for the second mostly from Feature Change Map (FCM) digitized updates. This is reflected in their respective 45.72-meter mdd and 133.63-meter mdd. See ***Figure VD4, St. Tammany Parish, Distance by Tract***, below for details.

Figure VD3. St. Tammany Parish

Distance By Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
22103040304	2.34	22.86	34.15	57.37	150.52	420.42	719.19
22103040501	1.1	12.658	22.93	35.03	53.59	81.736	467.76

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
22103040304	259	133.631	161.508	10.036
22103040501	347	45.721	43.086	2.313

Figure VD4. St. Tammany Parish: Distance by Tract

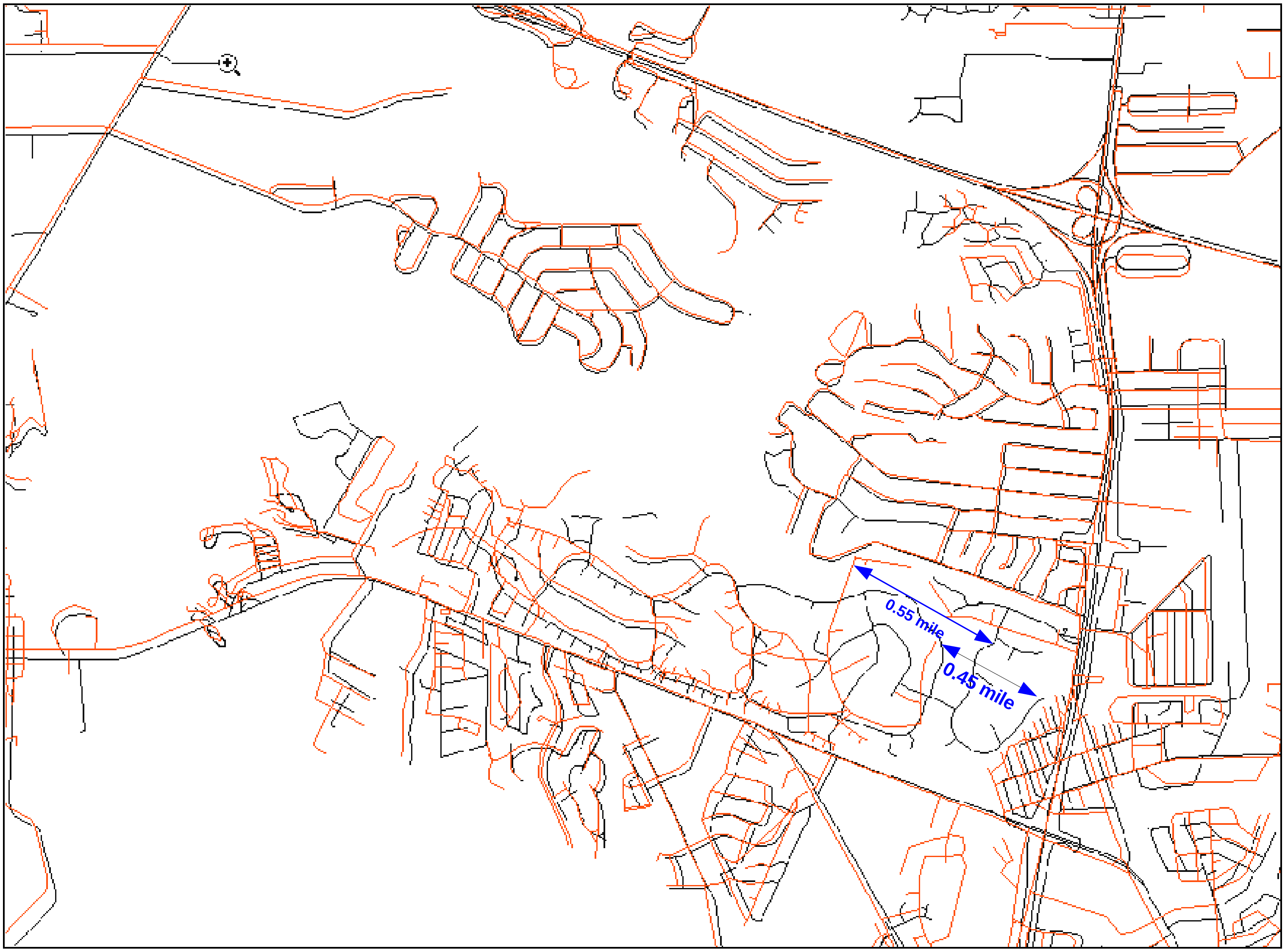


Figure VD5. Map displaying TIGER overlaid on local GIS file for Census Tract 403.04.

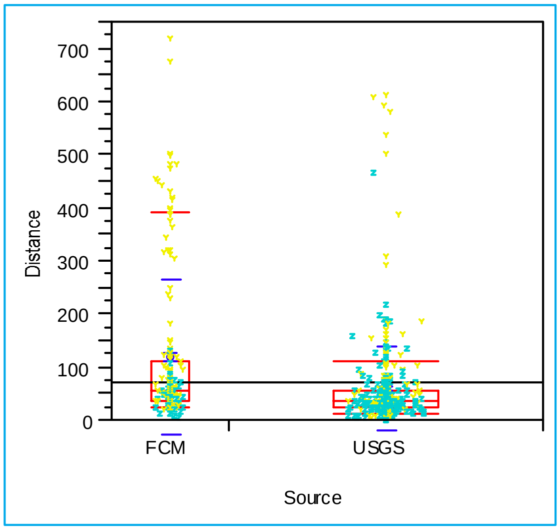
Figure VD5 displays the portion of tract 403.04 that was difficult to navigate through because the TIGER updates were not drawn to scale. In this area, the GPSC instrument indicated the actual location was several streets over from the TIGER map location displayed on the screen. The streets in red ink are those from the TIGER data base, those in black ink are from the local digital file. A return visit using TIGER map products and GPS as navigation devices would be very difficult as would be allocation of housing units based on the current TIGER data base.

Figure VD6. Aerial displaying superimposed TIGER street feature network.

Figure VD6 displays the same area as figure VD5 but using an aerial photograph as the backdrop to the TIGER file. Using the aerial as proof, the local GIS file has a higher spatial accuracy than the TIGER data base, however, the two files are in agreement for the majority of features.

Note: The Address List updates in census tract 403.04 are not included in the data listed in any of the tables, if they were, the results would show a greater variation from ground truth. In general, some of the most severe spatial problems in the TIGER data base are from small clusters of annotations that are digitized either in the wrong location or at a different scale than the surrounding area. They cause great difficulties for allocation of housing units when political boundaries are superimposed. New growth perpetuates the scale/location problem when as it is digitized on the fringes of these features.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
FCM	5.43	23.388	33.66	55.93	109.56	391.996	719.19
USGS	1.1	12.722	23.49	35.93	56.59	108.65	612.27

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
FCM	155	118.489	147.071	11.813
USGS	407	57.204	80.245	3.978

Figure VD7. St. Tammany: Distance by Source

Figure VD7, St Tammany, Distance by Source, shows a significant difference between FCM updates (118.5-meter mdd) and USGS 100K updates (57.2-meter mdd) to the TIGER data base. Note almost all the outliers are from census tract 403.04 (dots shown in yellow.) Interestingly both updates came from USGS based maps, however, the TIGER update methodology was quite different[23].

[23] FCM updates came from maps updated by the Census Bureau's regional office staff to include features not on the original USGS quads (usually 1:24,000 scale.) USGS 100K updates were raster scanned, vectorized, and converted to the TIGER file format.

E. Clark County, NV (32003)

Test dates: December 18, 1999 to December 19, 1999

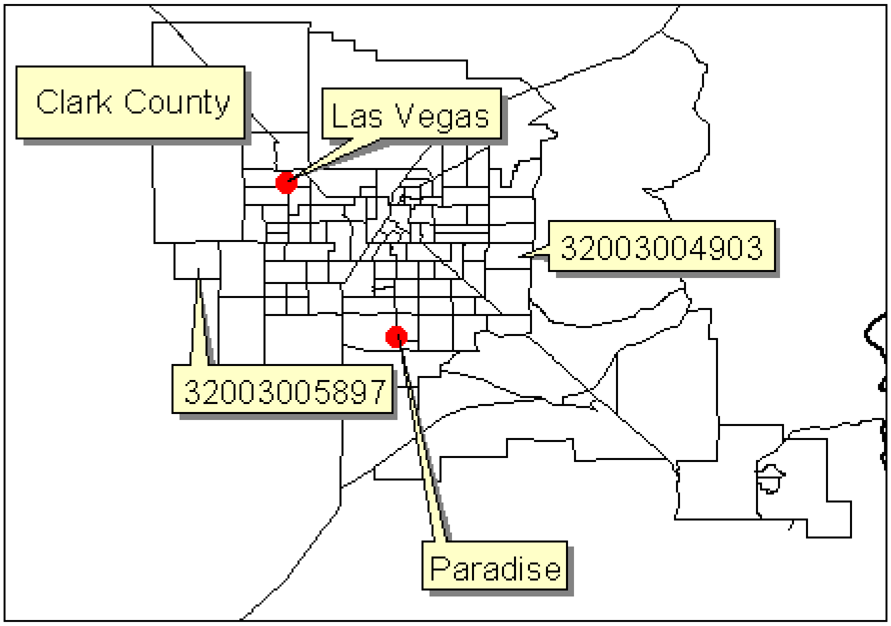


Figure VE1. Clark County (location of census tracts selected for GTAAT analysis)

Clark was selected as a test site because we have a digital GIS file that was used in a pilot program. Feature updates from this pilot program were made to the original Geographic Base File/Dual Independent Map Encoding File (GBF/DIME) and 90 Enumerator update core. The table that follows lists the update operations that were found in the two test census tracts:

Operation	49.03	58.97	Total	Percentage
90COLL		1	1	0.1%
90ENUM		283	283	28.9%
BLK CNVS	10	19	29	3.0%
DIGPILOT	144	112	256	26.1%
DX REV	20	40	60	6.1%
GBF-C	164	91	255	26.0%
GEO90	8		8	0.8%
MAF/DEX		11	11	1.1%
MAF_ORES	27	43	70	7.1%
RSTRUCT3	1	7	8	0.8%
TOTAL	**374**	**607**	**981**	**100%**

Figure VE2. Clark County (collected anchor points by source and tract)

The evaluation tracts were 14 miles apart, the first census tract, 58.97 is 3½ miles north west of Spring Valley and the second, 49.03 is 2¾ miles south of Sunrise Manor, both south of Las Vegas.

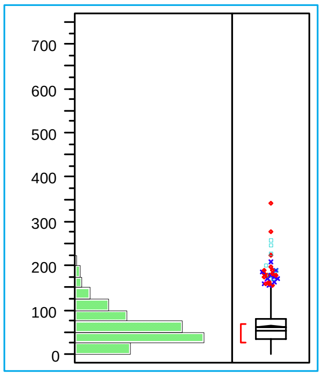
Figure VE3, Clark County, shows the distribution of the mean distance difference for the Clark County test tracts. There is a single peak at the 25-50 meter range with a normal declining distribution as distance increases.

Clark is one of the fastest growing areas in the nation, like other high growth areas, it experiences large amounts of feature updates through programs like Address List, Block Canvass, MAFGOR, LUCA, etc. All of these programs have diluted the original accuracy levels from the USGS or GBF/DIME core areas.

The two census tracts, 49.03 and 58.97 have significant numbers of features added to the TIGER data base through the Digital Pilot and Digital Exchange (DEX) Review programs. Both update operations used a local GIS file. Although both census tracts had updates from the two operations, they formed a larger proportion of the collected GPS anchor points in census tract 49.03. The result is a more accurate mdd (56.3 meters) compared to census tract 58.97 (68 meters.) See ***Figure VE4, Clark County, Distance by Tract***, for details of the distribution of spatial accuracy variance between the two test tracts.

The two census tracts had a variety of update operations, 80 percent of these updates came from 90 Enumerators, the Digital Pilot program, and GBF-C. It is noteworthy that over 1/3 of the anchor points were added to the TIGER data base directly through digital files thus eliminating any transcription or freehand digitizing errors.

Distance



Quantiles

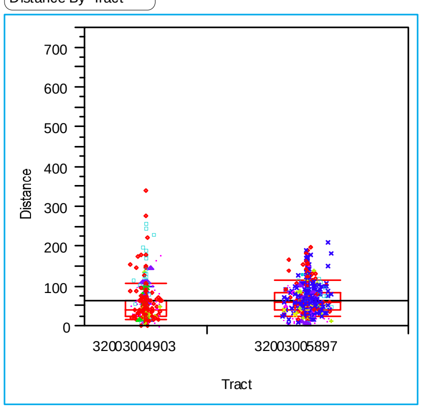
maximum	100.0%	343.27
	99.5%	233.14
	97.5%	174.25
	90.0%	113.60
quartile	75.0%	83.30
median	50.0%	54.91
quartile	25.0%	34.70
	10.0%	21.66
	2.5%	10.49
	0.5%	4.78
minimum	0.0%	2.45

Moments

Mean	63.5518
Std Dev	41.2123
Std Error Mean	1.3158
Upper 95% Mean	66.1340
Lower 95% Mean	60.9696
N	981.0000
Sum Weights	981.0000

Figure VE3. Clark

Distance By Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
32003004903	2.45	17.46	27.835	43.645	67.1025	111.385	343.27
32003005897	4.83	25.89	42.66	60.77	87.42	115.436	211.01

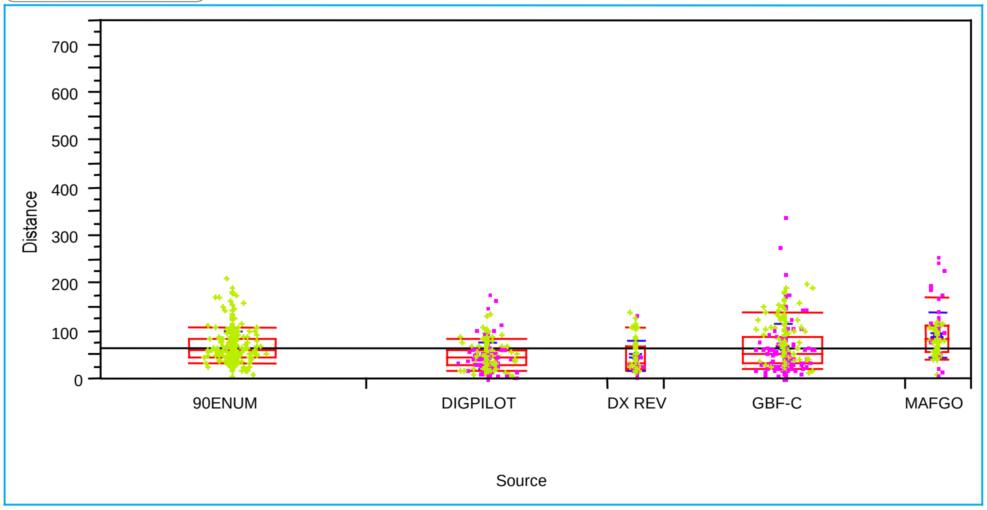
Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
32003004903	374	56.2624	45.9382	2.3754
32003005897	607	68.0431	37.3463	1.5158

Figure VE4. Clark County: Distance by Tract

Figure VE5, Clark County, Distance by Source, is on the following page. It shows the distribution of collected GPS anchor points by source or update operation and reflects the same distribution analyzed by update source. It also clearly makes the case for using spatially accurate digital files to update the TIGER data base.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
90ENUM	7.47	32.64	46.5	62.09	87.14	110.46	211.01
DIGPILOT	2.86	17.027	28.02	44.645	63.1275	86.478	182.75
DX REV	4.22	16.382	22.9175	33.035	71.505	110.199	141.86
GBF-C	2.45	22.874	32.66	53.56	88.67	143.144	343.27
MAF_ORES	10.11	43.063	59.4975	83.9	115.1725	171.881	262.48

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
90ENUM	283	69.6404	34.9746	2.0790
DIGPILOT	256	48.4901	28.6650	1.7916
DX REV	60	49.1563	34.8632	4.5008
GBF-C	255	67.7516	49.9161	3.1259
MAF_ORES	70	94.0711	50.8576	6.0786

The GBF-C and 90 Enumerator updates, which combined make up 55 percent of all 1-cell features in the test tracts, have a 67.8-meter mdd and a 69.6-meter mdd respectively. The two digital exchange updates, DIGPILOT and DX REV make up 32 percent of all updates and have a 48.5-meter mdd and a 49.2-meter mdd respectively. Clark County benefits from a 20-meter mdd accuracy improvement over the original GBF-DIME base by using local GIS files for feature updates. In contrast, the MAF_ORES updates (94.07-meter mdd) reduced the accuracy of the Clark County TIGER data base significantly.

F. Delaware County, OH (39041)

Test date: January 31, 2000

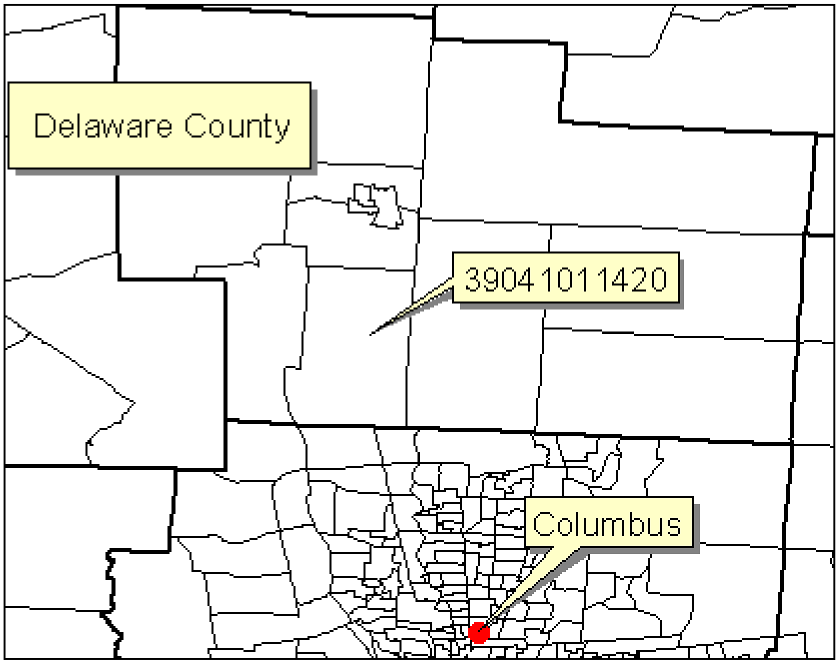


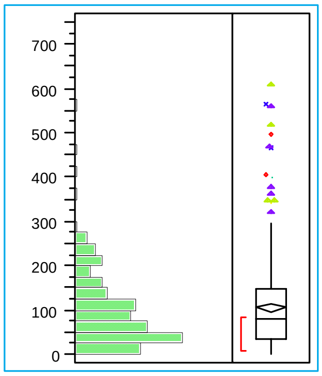
Figure VF1. Delaware County (location of census tracts selected for GTAAT analysis)

Delaware County, Ohio was selected because it was a TIGER Improvement Program (TIP) participant and had significant updates under that operation.

Operation	114.20	Percentage
90ENUM	40	8.3%
BAS98	10	2.1%
BLK CNVS	25	5.2%
CMP	36	7.5%
GBF-C	147	30.4%
LUCAUPD	26	5.4%
LUCAVERI	3	0.6%
MAF_ORES	2	0.4%
RSTRUCT3	6	2.3%
TIP	188	38.1%
TOTAL	**483**	**100%**

Figure VF2. Delaware County (collected anchor points by source and tract)

Distance



Quantiles

maximum	100.0%	612.56
	99.5%	567.64
	97.5%	353.08
	90.0%	225.49
quartile	75.0%	148.80
median	50.0%	81.17
quartile	25.0%	36.27
	10.0%	22.22
	2.5%	12.24
	0.5%	5.69
minimum	0.0%	3.33

Moments

Mean	106.2801
Std Dev	92.8718
Std Error Mean	4.2258
Upper 95% Mean	114.5835
Lower 95% Mean	97.9767
N	483.0000
Sum Weights	483.0000

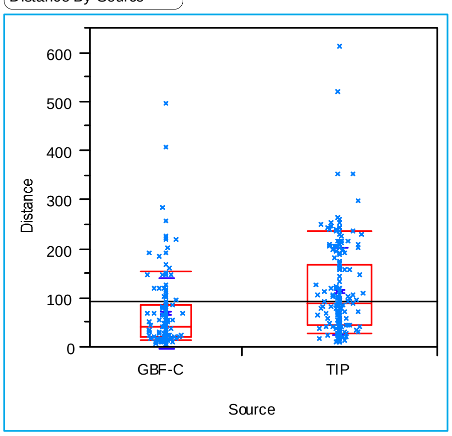
Figure VF3. Delaware County

Figure VF3, Delaware County, displays the distribution of points by distance from ground truth. Like Maricopa County, AZ, the data for Delaware County, OH show two variance peaks. The largest peak is at the 25-50 meter interval (attributable to the GBF-C updates) and a smaller one at the 100-125 meter interval (attributable to the TIP updates.)

Figure VF4, Delaware County, Distance by Source, displays the spatial accuracy differences between the GBF-C and TIP updates in Delaware County. Thirty-eight percent of the updates are pre-1990 (GBF-C, 90 ENUM) and the remaining 60 percent were updates since 1995 (TIP, MAF_ORES, etc.) The GBF-C updates (68.6-meter mdd) have less than half the mean distance difference of the TIP files.

The TIP operation, like other freehand digitizing update operations, has a large variance from ground truth, 114.17 meters or over 350 feet. In most urban or suburban settings, like Delaware County, this is enough to navigate to a wrong housing unit in return visits using a GPS navigation device.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
GBF-C	4.64	14.404	23.76	42.34	86.71	154.144	498.35
TIP	11.77	27.64	46.2325	92.415	170.835	236.201	612.56

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
GBF-C	147	68.949	73.4840	6.0609
TIP	188	114.173	90.3776	6.5915

Figure VF4. Delaware County: Distance by Source

G. York County, PA (42133)

Test dates: February 2, 2000 to February 3, 2000

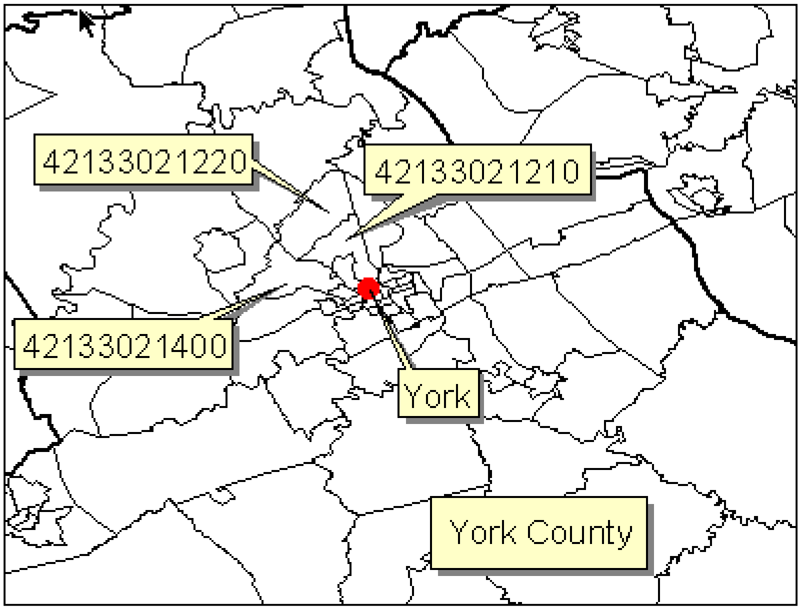


Figure VG1. York County (location of census tracts selected for GTAAT analysis)

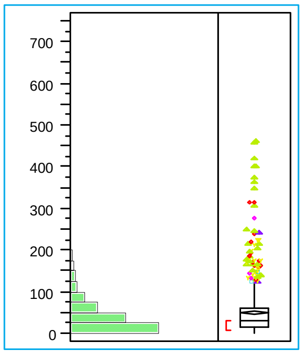
York was selected as a test site for its proximity to Headquarters and the county's participation in a multitude of partnership update operations including:

Operation	212.10	212.20	214	Total	Percentage
90ENUM	4		3	7	0.9%
BLK CNVS	1		4	5	0.6%
CMP	48	15		63	7.8%
FID IMPV		1		1	0.1%
GBF-C	215	105	244	564	70.1%
GEO90		2	3	5	0.6%
LUCAUPD		2		2	0.2%
MAF_ORES	18	4	37	59	7.3%
RSTRUCT3	9	3	4	16	2.0%
TIP		34	48	82	10.2%
TOTAL	**295**	**166**	**343**	**804**	**100%**

Figure VG2. York County (collected anchor points by source and tract)

Over 70 percent of the updates for the three test tracts came from pre-1990 data files, the remainder were from recent updates including TIP, MAFGOR, and CMP. The county has participated in all possible opportunities to improve census coverage including CMP, TIP, and LUCA. York also has a digital file that can be used in future TIGER modernization. With a 51.53-meter mdd, York has the most accurate spatial attributes of all the test sites.

Distance



Quantiles

maximum	100.0%	464.61
	99.5%	404.43
	97.5%	218.82
	90.0%	112.31
quartile	75.0%	62.12
median	50.0%	30.57
quartile	25.0%	18.77
	10.0%	11.36
	2.5%	6.03
	0.5%	1.46
minimum	0.0%	0.87

Moments

Mean	51.5311
Std Dev	59.5204
Std Error Mean	2.0991
Upper 95% Mean	55.6516
Lower 95% Mean	47.4106
N	804.0000
Sum Weights	804.0000

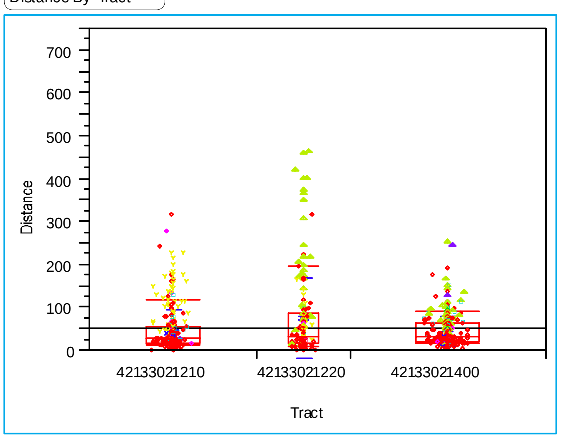
Figure vG3. York County

Figure VG3, York County, reveals that 50 percent of the observations have better than 30-meter mdd, this is attributable to the large number of GBF-C features from the file creation phase prior to 1990. This also supports the theory that most non-digital updates significantly reduce the accuracy of the TIGER data base.

Figure VG4, York County: Distance by Tract, shows a significant difference, between tracts 212.10 and 214 vs. 212.20. They have a 46.9, 44.5 and 74.3 meter mdd respectively.

The differences found in census tract 212.20 are attributed to large number of TIP updates (> 20 percent) with a spatial accuracy of >195 meters mdd (displayed as chartreuse triangles) and significant CMP updates (>10 percent) with 73 meters mdd. Note that the TIP and MAFGOR mdd are significantly higher than the three tract averages. This is partially due to the location of this census tract. Of the three census tracts, 212.20 is the most distant from the center of the city of York, therefore, it is subject to compound updates. As noted previously, compound updates are those feature updates that are added to features from previous update operations and thus have inherent inaccuracies. In this sample, it involves MAF_ORES, TIP, and CMP updates.

Distance By Tract



Quantiles

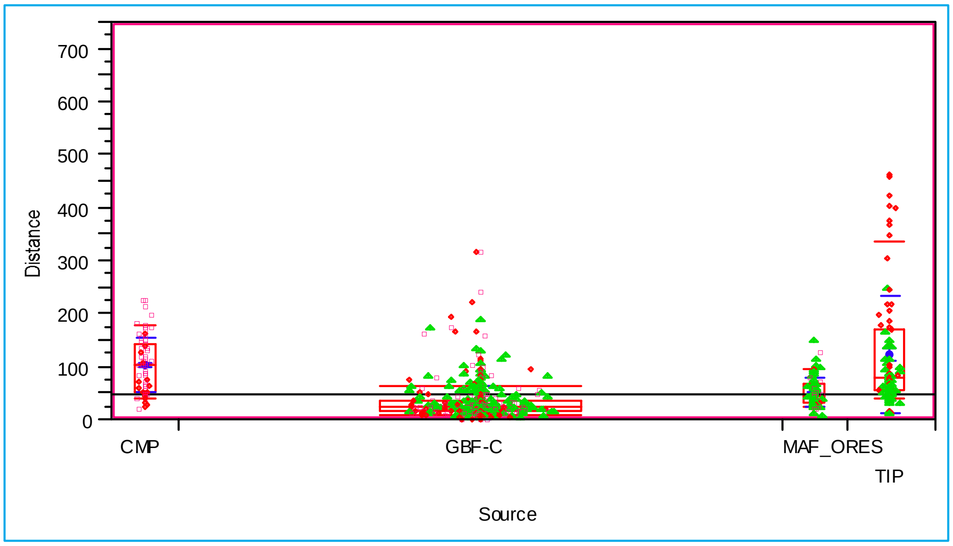
Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
42133021210	1.16	11.368	17.89	25.83	53.14	119.512	316.88
42133021220	0.87	8.496	16.0425	33.5	85.8125	195.923	464.61
42133021400	3.65	14.176	20.38	31.9	61.01	88.576	252.42

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
42133021210	295	46.8763	50.3994	2.9344
42133021220	166	74.3426	97.1911	7.5435
42133021400	343	44.4945	35.4428	1.9137

Figure VG4. York County: Distance by Tract

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
CMP	22.53	41.578	53.76	102.34	145.54	178.066	226.98
GBF-C	0.87	10.21	16.525	23.445	37.8775	66.3	317.63
MAF_ORES	8.29	22.93	32.39	46.03	69.53	97.95	151.36
TIP	11.21	39.011	57.285	81.83	169.52	337.853	464.61

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
CMP	63	103.292	54.474	6.863
GBF-C	564	34.048	35.094	1.478
MAF_ORES	59	54.267	29.875	3.889
TIP	82	125.532	110.697	12.224

Figure VG5. York County: Distance by Source

Figure VG5, York County, Distance by Source, shows the accuracy differences among the four major feature sources in the three test tracts. Clearly, the TIP and CMP updates fall significantly outside the county mean, with 125.5 and 103.3-meters mdd respectively. The MAF_ORES operation faired better with 54.3-meters mdd, however, it was still 60 percent greater than the average of anchor points from the GBF-C operation.

H. Windham County, VT (50025)

Test dates: November 1, 1999 to November 5, 1999

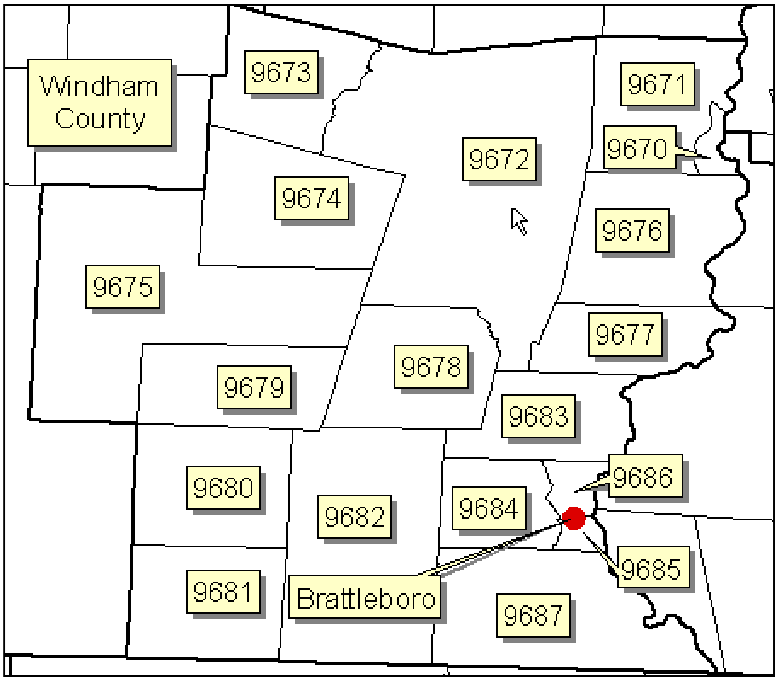


Figure VH1. Windham County (location of census tracts selected for GTAAT analysis)

Windham County was selected as the primary test site for the GPSC instrument, to test our field methodology, and to further develop our knowledge of the GPS technology. The preparation and logistics for this test were the responsibility of the TOB, specifically Kevin Donnalley, Steve Ho, and Mitch Milligan. Please refer to the **GPS Cartographic Software Field Test Report** produced by HTE-UCS, Inc. for detailed report on the Windham County test.

The GEO selected the State of Vermont to be part of a pilot conflation or rubbersheeting program developed by the Geographic Update Systems Branch (GUSB). Vermont has completed a statewide GIS with GPS derived map spots for housing unit coverage. This is an ideal digital file for decennial enumeration and allocation purposes.

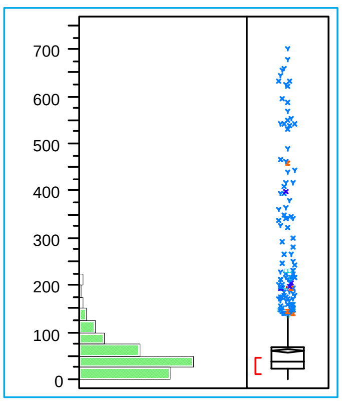
Windham County was not one of the nine (out of 14) Vermont counties run through the rubbersheeting process. This was a major factor in selecting it for this test. The plan is to run Windham County through this process in the future and compare the results with those observed in the field. This allows a comparison between field collected data and local spatial data files, to ensure they meet the minimum spatial accuracy threshold (not yet determined), and helps develop an evaluation mechanism for accepting these files in the future.

Most of the features in the Windham County file came from pre-1990 base development including USGS 100K and FCM updates that when combined provided 95 percent of all features. Other update operations include:

Operation	9670	9672	9677	9683	9684	9685	9686	9687	Total	Percentage
80BDRY				6	1	2			9	0.5%
90ENUM	1	7	4	2	4	2		3	23	1.4%
CMP		8	2						10	0.6%
FCM	21	241	154	79	35	61	37	61	689	41.5%
LOCAL								8	8	0.5%
MAF_ORES					15	13	13		41	2.4%
USGS	29	281	117	88	98	116	86	67	882	53.1%
TOTAL	51	537	277	175	153	194	136	139	1662	100%

Figure VH2. Windham County (collected anchor points by source and tract)

Distance



Quantiles

maximum	100.0%	833.31
	99.5%	636.32
	97.5%	340.87
	90.0%	119.60
quartile	75.0%	69.40
median	50.0%	40.56
quartile	25.0%	23.86
	10.0%	14.79
	2.5%	7.57
	0.5%	3.23
minimum	0.0%	2.44

Moments

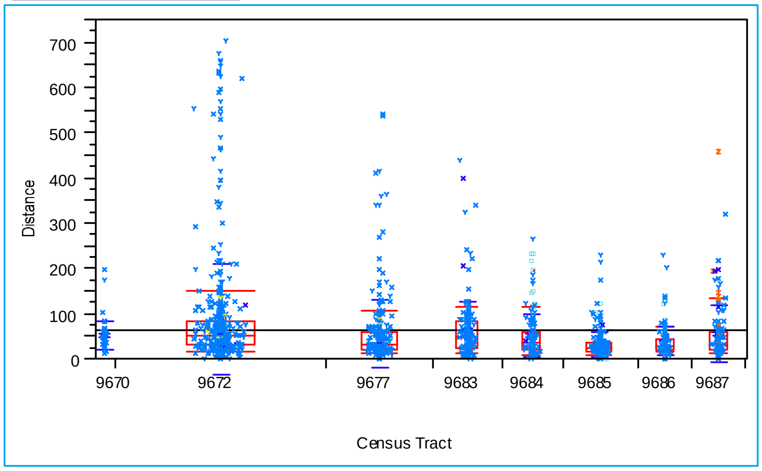
Mean	63.883
Std Dev	87.134
Std Error Mean	2.137
Upper 95% Mean	68.075
Lower 95% Mean	59.691
N	1662.000
Sum Weights	1662.000

Figure VH3. Windham County

The Windham County test was the only countywide test site. It represents an area of diverse terrain, abundant tree cover, and it was the first major GPS field data capture experience using the GPSC software. These factors influenced data capture and contributed to the significant number of data outliers. The Quantiles section of Figure VH3 shows that 10 percent of the observations had greater than a 119-meter mdd. Most observances (bar chart peak), however, are found at the 25-50 meter mdd level, which is more typical of the USGS 100K updates.

The outliers, shown in Figure VH3 by dots floating above the box-plot tail, which represents 1-standard deviation, inflate the data for FCM and USGS source codes in the selected evaluation tracts. Adjusting for these factors, the two USGS based operations and source codes are consistent with the other test sites.

Distance By Census Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
9670	13.32	27.566	39.97	45.84	60.96	80.796	201.68
9672	2.44	19.302	32.39	52.77	84.55	151.812	833.31
9677	3.78	12.258	21.59	34.66	61.435	107.628	543.66
9683	2.73	15.798	25.22	47.85	85.09	118.016	441.52
9684	2.82	10.734	20.16	36.39	62.27	115.66	267.16
9685	2.82	11.28	18.65	27.565	39.0175	66.135	229.52
9686	3.2	17.715	21.1075	29.025	47.3725	74.036	233.31
9687	3.11	14.54	19.94	33.24	61.95	135.37	462.33

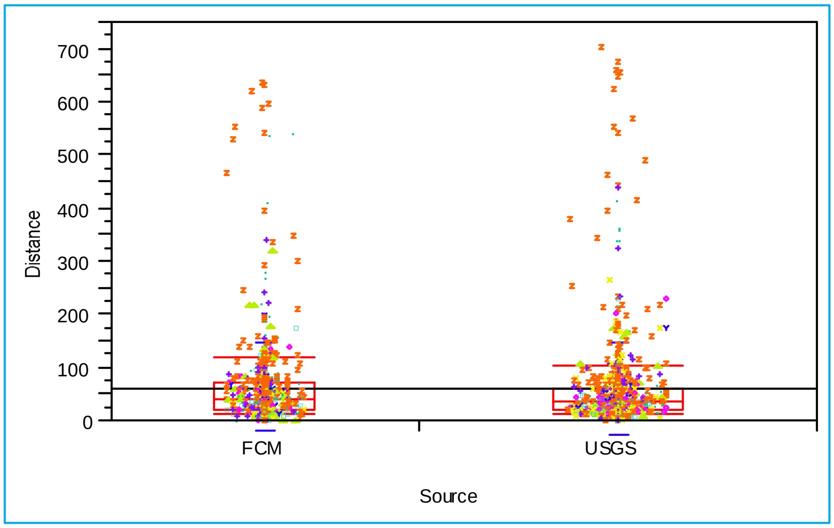
Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
9670	51	54.2816	32.597	4.5644
9672	537	89.2711	124.657	5.3793
9677	277	57.4916	75.543	4.5389
9683	175	64.7605	65.108	4.9217
9684	153	52.4720	50.547	4.0865
9685	194	35.4821	31.502	2.2617
9686	136	40.2950	33.988	2.9144
9687	139	56.2356	63.572	5.3921

Figure VH4. Windham County: Distance by Census

Figure VH4, Windham County, Distance by Census Tract, shows the comparison of the mdd by tract. There is a significant difference when comparing the eight census tracts in this evaluation. The mdd ranges from 35.5-meters for tract 9685 to 89.3-meters for tract 9672. The mdd was lower in the town of Brattleboro and along the Connecticut River (eastern extent of the county) and higher towards the hilly/mountainous interior. The analysis of the Position Dilution of Precision (PDOP) readings for Windham divulged no significant difference between the two areas. It was confirmed with the team that gathered the data for Census Tract 9672 that signal accessibility was a particular problem in the higher elevation and many collection points were deemed not collectable. The only other factor that may attribute to this difference is the distance to the base station, which is located east of Windham County.

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
FCM	2.63	14.78	23.685	43.14	72.46	121.87	636.78
USGS	2.44	14.615	23.0575	37.145	61.5175	106.094	833.31

Means and Std Deviations

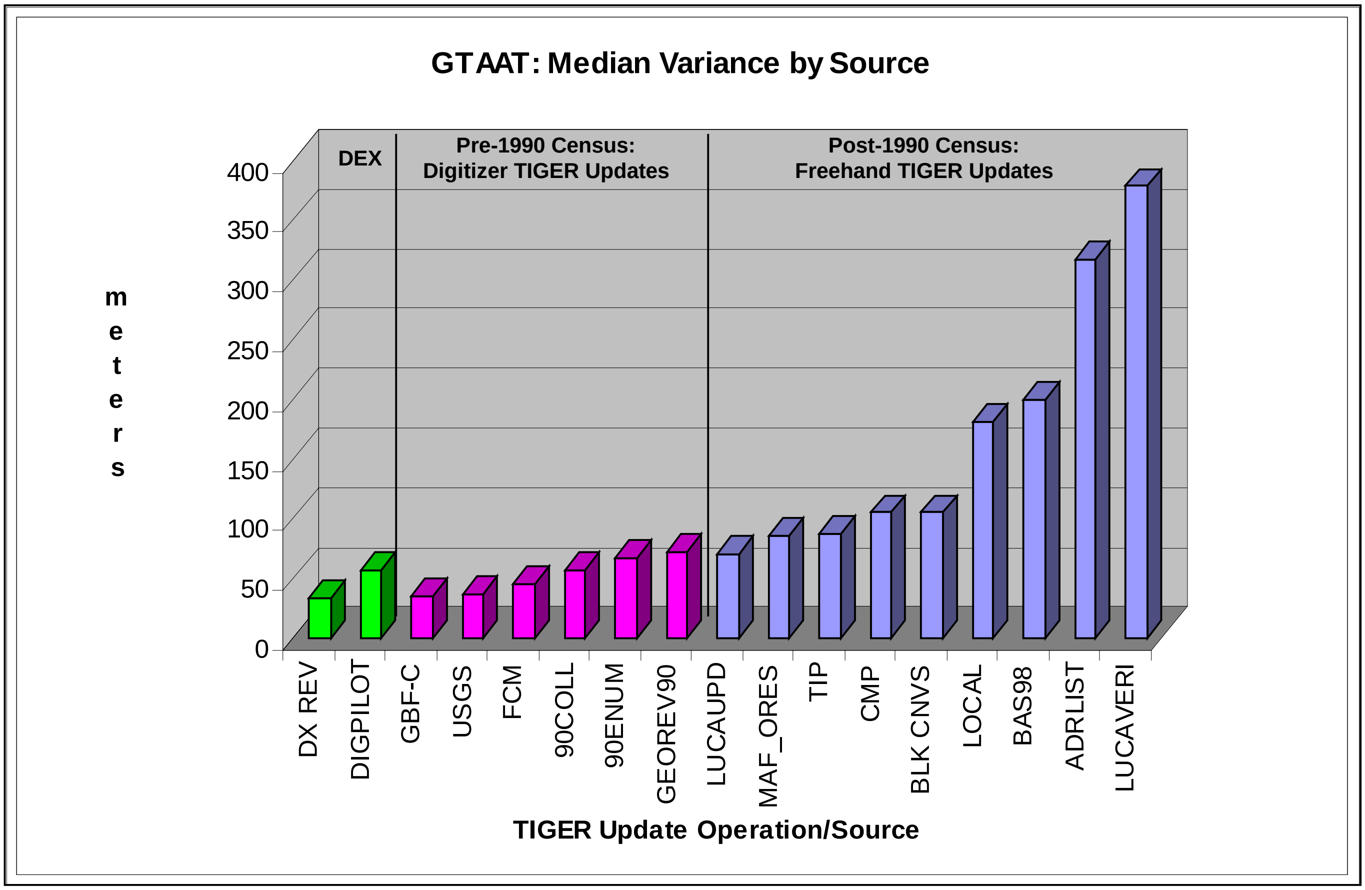
Level	Number	Mean	Std Dev	Std Err Mean
FCM	689	64.9677	83.4649	3.1798
USGS	882	60.8254	90.4042	3.0441

Figure VH5. Windham County: Distance by Source

The FCM and USGS 100K updates come from essentially the same base and, therefore, there are no significant differences in the mdd. See ***Figure VH5, Windham County, Distance by Source***, for details.

VI. Source-by-Source Analysis

Using the GPS to capture data allows us to analyze the TIGER data base across operation/source and history codes. There are currently over 100 different source and history codes in the TIGER data base, and each serve to identify the origin of a feature (source) and the latest update to that feature (history). This is one of the few metadata type fields available in the TIGER data base. We discovered through the GTAAT that there was a large variance in the mean distance difference from TIGER to ground truth based on the source code. This was expected, as there are multiple methods of adding or updating features in the TIGER data base and each source has an inherent positional accuracy.



Graph VI1, GTAAT: Median Variance by Source.

Graph VI1, GTAAT: Median Variance by Source, shows the median distance difference in meters of the selected operations or sources of the GPS anchor points and TIGER features in our sample. The operations displayed in magenta (middle) are mostly from the original TIGER data base building and features added in support of the 1990 Census. Those displayed in green (left) are post-1990 digital file updates like DIGIPILOT and DX REV. The operations on the right (blue) are updates that were made to TIGER in preparation for the 2000 Census.

Typically the pre-1990 and DEX groups are more spatially accurate than the post-1990 group, a major contributing factor is the change in update methodologies used to update the TIGER data base. Examples of these changes between 1990 and 2000 include:

1990 Census:

- Map sources were evaluated in the field and accepted upon passing the evaluation.
- Annotated census maps were registered using a digitizing tablet prior to digitizing.
- Field annotations were traced using a digitizer.
- Secondary sources were used to verify field annotations prior to digitizing.
- USGS scanned maps had known spatial accuracy levels.

2000 Census:

- Map sources were not quality evaluated.
- Annotated census maps were not registered using a digitizing tablet.
- Annotations were digitized freehand using a mouse (also known as heads-up digitizing).
- Rubbersheeting (not evaluated here) and Digital Pilot for limited areas.

The feature capture methodology changes between the two censuses contribute to the decrease in the spatial accuracy of the TIGER data base. The domino effect of successive update operations, each building on the previous operations inaccuracies, is a striking feature of the TIGER data base. A clear example of this can be seen in Maricopa County (see Maricopa County in the Site-by-Site section). The current trend will continue unless corrective measures are taken to assure high levels of accuracy in TIGER. These measures include:

1. utilizing spatially accurate local digital files when they are available;
2. evaluating the source materials we use (digital or otherwise);
3. using new technologies to capture updates (GPS, digital orthophotography, etc.);
4. require verification of feature additions, deletions, or changes to the TIGER data base (digital photography);
5. require metadata for all TIGER features and attributes;
6. provide for enhanced interactive conflation to scanned imagery[24] files in TIGER update software (e.g. GusX); and
7. require minimum spatial accuracy standards to be met prior to acceptance of new features in the TIGER database. All updates are provisional until they meet some minimum spatial accuracy requirements.

[24] Refers to pure digital orthophotography, scanned reference sources (with available known points for spatial reference), or scanned aerial photography. It also may include the creation of a national coordinate pixel grid system using fixed points throughout the country that force features to adhere to a relative or measured location on a digital map base/grid, etc.

VII. A Summary and a Look to the Future

The GEO set out to document the TIGER data base spatial inaccuracies using statistical methods. The GRaSS conducted several pilot tests utilizing GPS technology. With their assistance, the TOB developed a plan to capture GPS data points in the field and use the variance between these points and the TIGER data base for analysis. Until now we only could point to specific examples, usually visual anomalies, or respond to outside inquiries from users. The GEO is interested in finding these spatial anomalies in order to take corrective measures and to establish feature update methodologies that would greatly reduce them in the future. One of our goals is to collect as many data in the field as possible to enable us to establish a spatial accuracy index for each county. This index becomes part of a GIS that includes demographic, topographic, map source, address source, and other attribute source and metadata materials for each county (similar to the Address System Information Survey). Linked to the Geographic Program Participant (GPP) system, it can provide the GEO with an operation and source snapshot for each county and is used like a targeting database for spatial data and attribute capture. Until this system is in place, the data available from this analysis supports the following issues:

With over 6,800 anchor points captured in the field, the GEO has found multidimensional spatial anomalies in all 22 census tracts. The first dimension relates to the operation that a feature is captured, whether it is a scanned USGS image, a digital exchange file, or an annotated map from the Address List operation. This influences the mean distance difference for a given GPS anchor point. The second dimension relates to the digitizing method that is used to add a feature into the TIGER data base and whether a digitizing tablet was used or not. Although this method is greatly correlated to the update operation (first dimension), evidence exists of disparities between similar operations, the 90 Enumerator Update and the Address List operation for Census 2000 for example. A third dimension involves the order and location of updates. This reflects the compound update spatial anomalies that are found throughout the TIGER data base.

How does each of these dimensions affect spatial accuracy in TIGER?

It is easy to categorize or create profiles for counties with multidimensional spatial anomalies. For example, Hillsborough County has scattered updates on a GBF/DIME and USGS base, and this limits the amount of variance new updates have as they most likely build directly on the USGS or GBF/DIME feature base. Maricopa County, on the other hand, has multiple clusters of updates on GBF/DIME and USGS feature base. This has a domino effect on the variance for each successive update operation.

Our good experience with TIGER updates for the 2000 Census is one that demands expansion of partnerships and at the same time have greater control on the update process through new technologies and closer evaluation of source materials. We now have an instrument that is used for evaluating TIGER spatial accuracy and at the same time provides data for county profiling. The GPSC is the listing instrument of choice for inter-decennial field activities and can support Census 2010 decennial activities and the American Community Survey.

As selective availability has been turned off, return visits to housing units in the field become feasible without time-consuming postprocessing. This requires that the TIGER data base be brought within some minimal "acceptable" spatial accuracy levels.

Attached to this document are a series of tertiary graphs,[25] distribution tables, quantile and means tables by Census Tract, Site and Source Code. These graphs and tables summarize the data used for the GTAAT analysis. A brief description of each attachment follows:

1. Attachment A, GTAAT: Distance, Azimuth, and Source by Test Site, displays for each site, the distribution of the mean distance difference between the TIGER data base and the respective GPS anchor points.
2. Attachment B, GTAAT: Distance, Azimuth and Source by Census Tract, displays for each census tract, the distribution of the mean distance difference between the TIGER data base and the respective GPS anchor points.
3. Attachment C, GTAAT: Distance and Azimuth by Source, displays the distribution of the mean distance difference between the TIGER data base and the respective GPS anchor points by Source code.
4. Attachment D, GTAAT: Census Tract Comparison displays a box plot of the distribution of the mean distance difference between the TIGER data base and the respective GPS anchor points for each census tract. Quantiles and means and standard deviations tables are also included.
5. Attachment E, GPS TIGER ACCURACY ANALYSIS TOOLS (GTAAT) Source/Operation Comparison, displays, a box plot of the distribution of the mean distance difference between the TIGER data base and the respective GPS anchor points by source/operation code. Quantiles and means and standard deviations tables are also included

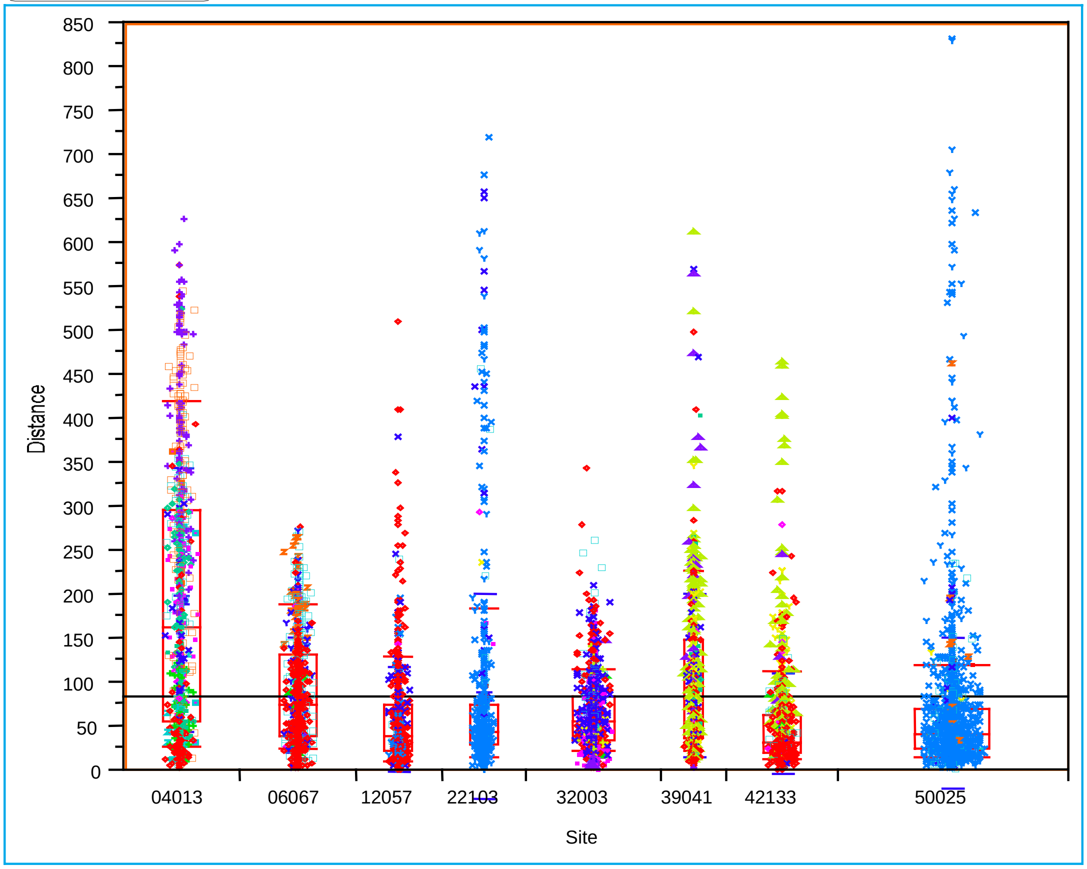
The first three attachments are tertiary graphs with the best data of the distributions at the bottom left of the graph, clustering denotes consistency, and the different colors represent source codes or census tracts depending on the graph. The other two attachments show box plots which are explained in detail in IV GTAAT Analysis. The dot colors for attachments A, B, and D represent the different sources, for attachments C and E they represent tracts. Refer to attachment F for a legend showing both color schemas.

Lastly, three summary tables display by source, by test site, and by census tract, the median distance difference in **feet** between the GPS captured anchor point and the TIGER data base. Note, for all other tables and graphs the mean or median distance is in **meters**.

[25] Tertiary plots or diagrams are a way of displaying the distribution and variability of three-part compositional data such as the proportion of distance, azimuth, and the primary Source code of the anchor point. Data is expressed in proportions, or absolute measures. The Ternary platform converts absolute values to proportions. The ternary display is a triangle with sides scaled from zero to 1. Each side represents one of the three components. A point is plotted so that lines drawn from the point to each leg of the triangle as shown below, intersect at the component values of the point.

GTAAT: Site-to-Site Comparison

Distance By Site



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
04013	2.63	25.28	55.685	161.52	296.07	418.672	626.57
06067	1.72	22.938	38.245	73.345	130.84	187.446	276.96
12057	0.13	10.76	21.385	37.46	73.385	128.79	509.43
22103	1.1	15.387	27.8925	42.555	74.1625	182.733	719.19
32003	2.45	21.662	34.705	54.91	83.3	113.6	343.27
39041	3.33	22.222	36.27	81.17	148.8	225.488	612.56
42133	0.87	11.365	18.7725	30.57	62.1225	112.31	464.61
50025	2.44	14.79	23.86	40.565	69.4025	119.601	833.31

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
04013	845	192.854	150.416	5.1745
06067	856	89.132	62.112	2.1229
12057	614	58.391	61.561	2.4844
22103	606	83.293	118.659	4.8202
32003	981	63.552	41.212	1.3158
39041	483	106.280	92.872	4.2258
42133	804	51.531	59.520	2.0991
50025	1662	63.883	87.134	2.1373

The Median and Mean Distance Difference (in Feet) by Operation/Source and Digitizing Method

OPERATION/SOURCE	Observations	Median	Mean	Process
90 Collection	84	186.12	201.31	Tablet
90 Enumerator Updates	488	216.81	287.51	Tablet
Address List	170	1039.35	962.46	Freehand
BAS 98	80	651.91	593.97	Freehand
Block Canvass	62	342.49	415.79	Freehand
CMP	109	341.40	358.57	Freehand
Digital Pilot	370	185.95	295.37	Direct
DEX Review	60	108.38	161.27	Direct
FCM	862	147.59	243.59	Scanned
GBF-C	1874	114.37	177.98	Scanned
GEO Review 90	137	235.56	290.66	Tablet
LOCAL	53	593.83	584.83	Freehand
LUCA Updates	88	228.92	290.96	Freehand
LUCA Verification	110	1239.86	1205.04	Freehand
MAFGOR	577	279.07	359.25	Freehand
RSTUCT3	55	225.36	288.33	
TIP	270	283.45	385.90	Freehand
USGS	1328	120.36	194.20	Scanned
Others	74	173.59	211.06	Varies
TOTAL	6851	166.53	281.07	Varies

The Median and Mean Distance Difference (in feet) for the Eight Test Sites.

SITE	Observations	Median	Mean	High Tract	Low Tract
Maricopa	845	529.92	632.72	732.41	110.85
Sacramento	856	240.63	293.43	248.55	220.14
Hillsborough	614	122.90	191.57	148.22	106.33
St. Tammany	606	139.62	273.27	188.22	114.93
Clark	981	180.15	208.50	199.37	143.21
Delaware	483	266.30	348.68	N/A	N/A
York	804	100.29	169.06	109.91	84.74
Windham	1662	133.09	209.59	173.13	90.42
TOTAL	6851	166.53	281.07	732.41	84.74

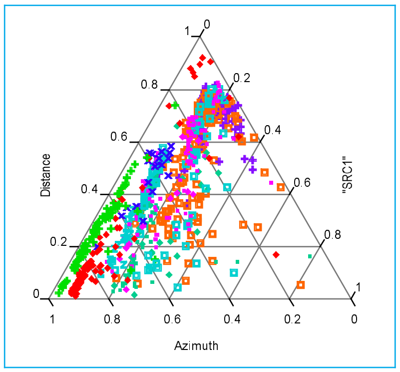
VIII. Conclusion

The Bureau of the Census recognizes the need to modernize the TIGER data base in order to meet the needs of future ongoing and decennial activities including the American Community Survey and the Boundary and Annexation Survey. Our research shows that the current accuracy of point and linear features in TIGER system limit the ability to exchange data digitally through partnerships. The bureau cannot afford to continue doing “business as usual” where updates to the TIGER data base are concerned. The current efforts to redesign TIGER must include strengthening public and private partnerships by promoting digital exchange and improving spatial accuracy standards. The future for TIGER modernization is bright if the ability to exchange data with local and federal partners becomes a trivial matter.

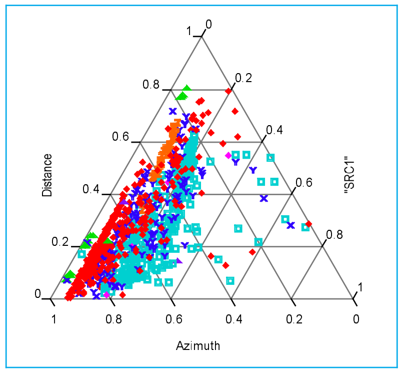
Attachment A

GTAAT: Distance, Azimuth, and Source by Test Site

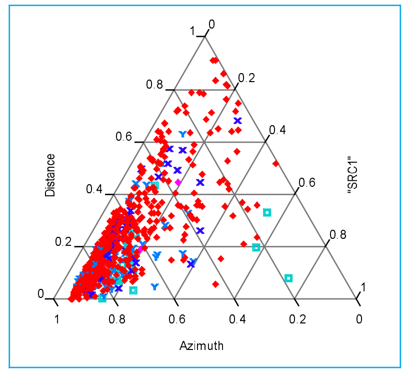
GTAAT3(STCO=04013)



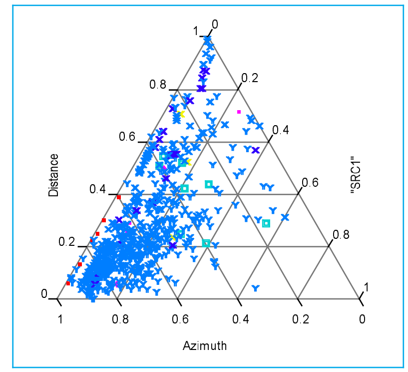
GTAAT3(STCO=06067)



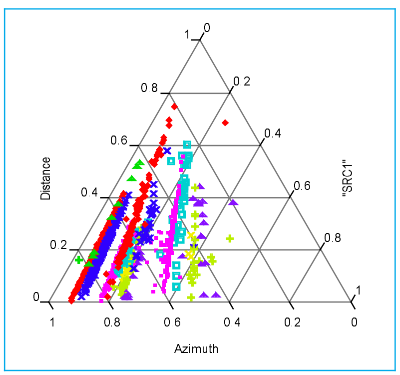
GTAAT3(STCO=12057)



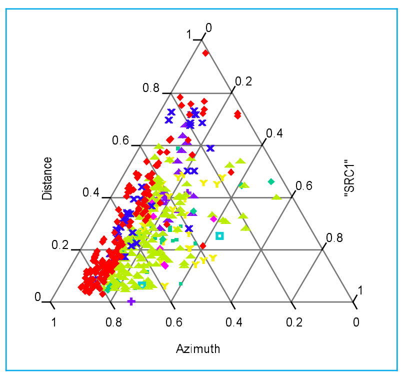
GTAAT3(STCO=22103)



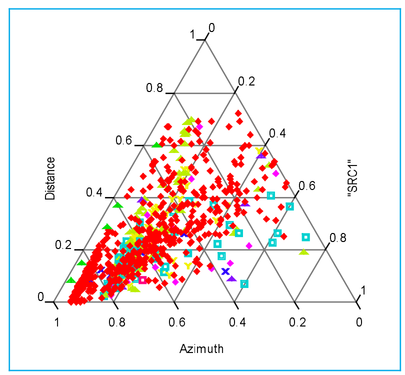
GTAAT3(STCO=32003)



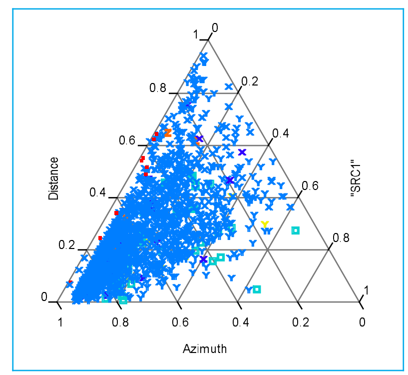
GTAAT3(STCO=39041)



GTAAT3(STCO=42133)

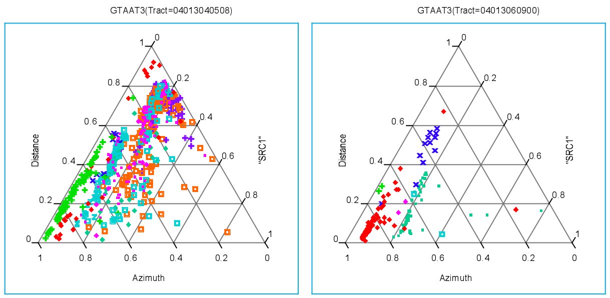


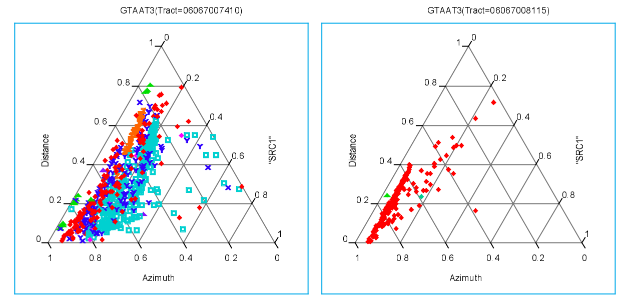
GTAAT3(STCO=50025)

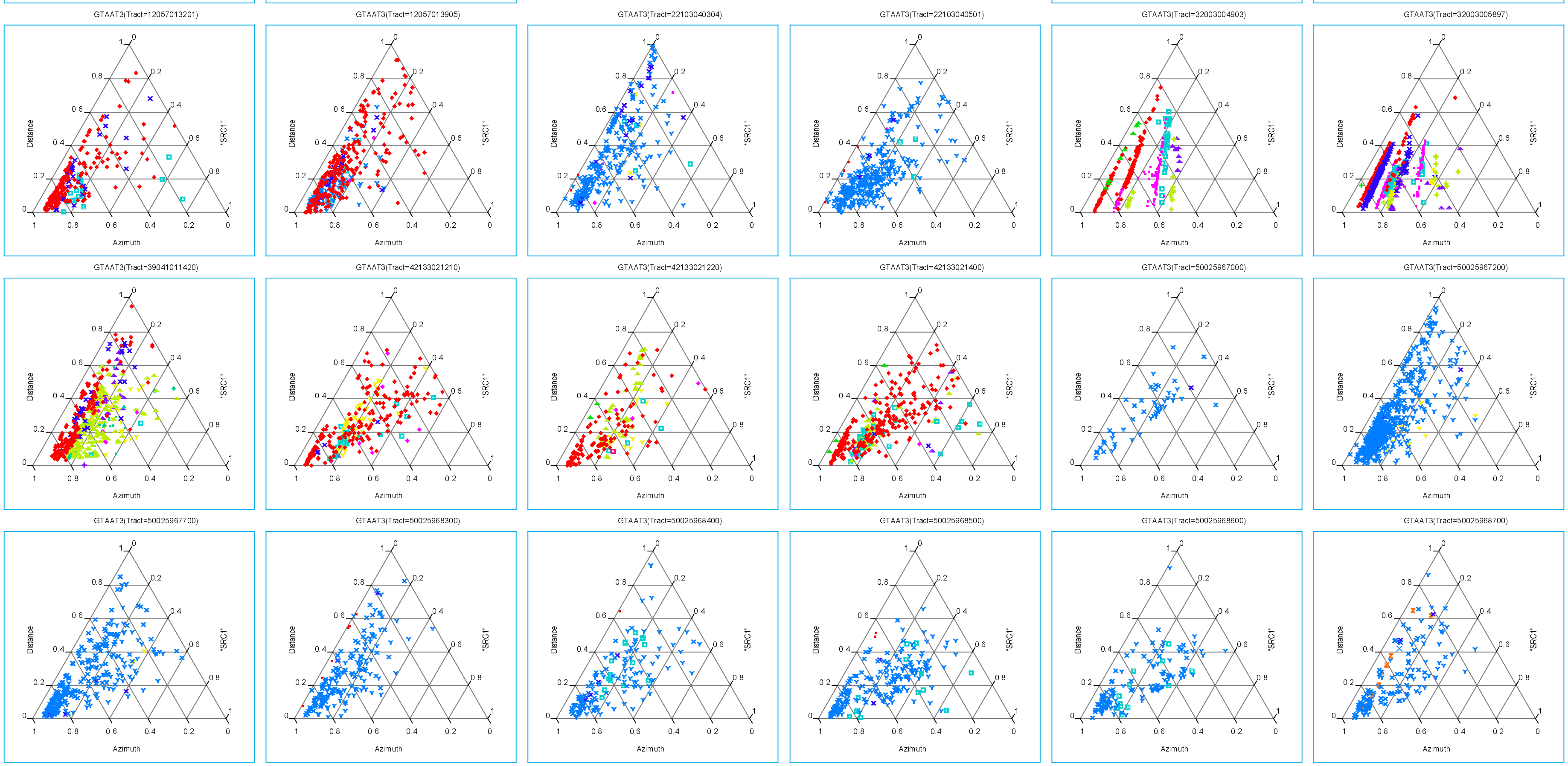


Attachment B

GTAAT: Distance, Azimuth, and Source by Census Tract

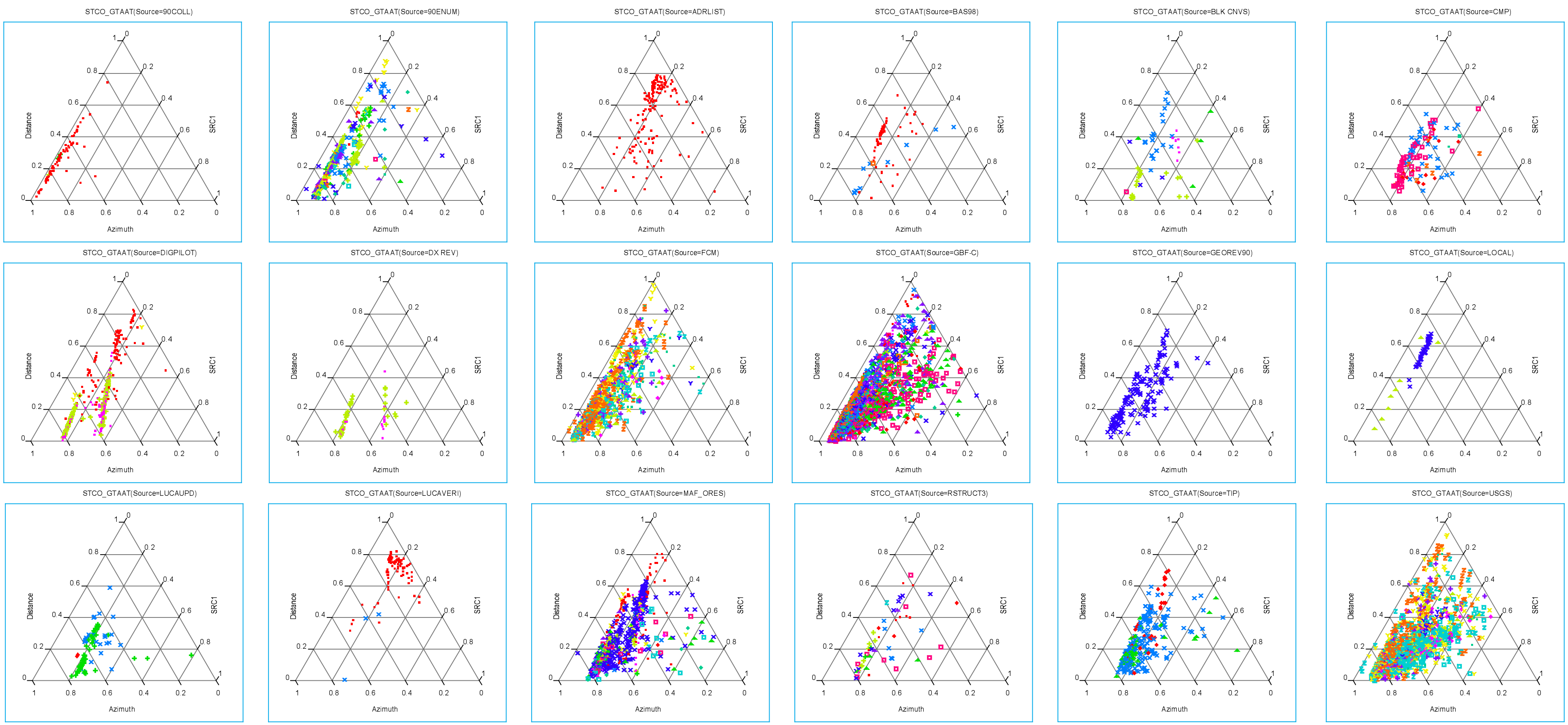






Attachment C

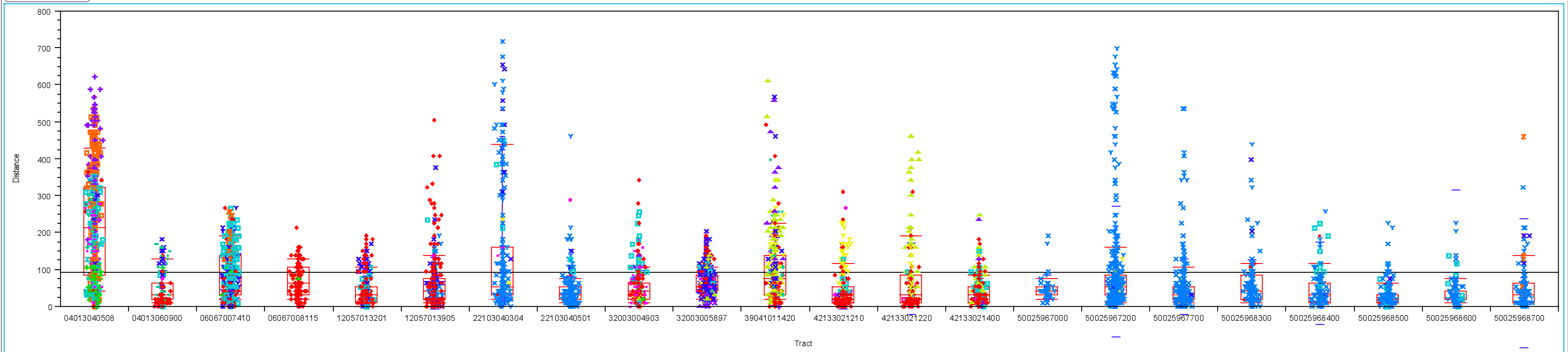
GTAAT: Distance, Azimuth, and Source by Source Code



Attachment D

GTAAT: Census Tract Comparison

Distance By Tract



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
04013040508	3.24	42.78	91.245	223.24	324.0975	436.915	626.57
04013060900	2.63	14.52	20.515	33.79	72.225	135.486	183.27
06067007410	2.69	23.73	38.3225	75.71	141.835	195.095	276.96
06067008115	1.72	21.824	37.6925	67.1	107.95	133.432	223.16
12057013201	3.59	9.04	19.0025	32.415	53.4725	111.995	192.67
12057013905	0.13	12.698	23.785	45.18	83.525	147.585	509.43
22103040304	2.34	23.075	34.18	59.34	160.585	446.94	42895.96
22103040501	1.1	12.658	22.93	35.03	53.59	81.736	467.76
32003004903	2.45	17.46	27.835	43.645	67.1025	111.385	343.27
32003005897	4.83	25.89	42.66	60.77	87.42	115.436	211.01
39041011420	3.33	22.222	36.27	81.17	148.8	225.488	612.56
42133021210	1.16	11.368	17.89	25.83	53.14	119.512	316.88
42133021220	0.87	8.496	16.0425	33.5	85.8125	195.923	464.61
42133021400	3.65	14.176	20.38	31.9	61.01	88.576	252.42
50025967000	13.32	27.566	39.97	45.84	60.96	80.796	201.68
50025967200	2.44	19.336	32.465	52.84	85.21	160.216	2040.76
50025967700	3.78	12.258	21.59	34.66	61.435	107.628	543.66
50025968300	2.73	15.798	25.22	47.85	85.09	118.016	441.52
50025968400	2.82	11.016	20.17	36.74	65.23	120.524	953.12
50025968500	2.82	11.28	18.65	27.565	39.0175	66.135	229.52
50025968600	3.2	17.73	21.235	29.1	47.745	76.714	3094.91
50025968700	3.11	14.561	19.9675	33.43	63.8625	141.58	1985.52

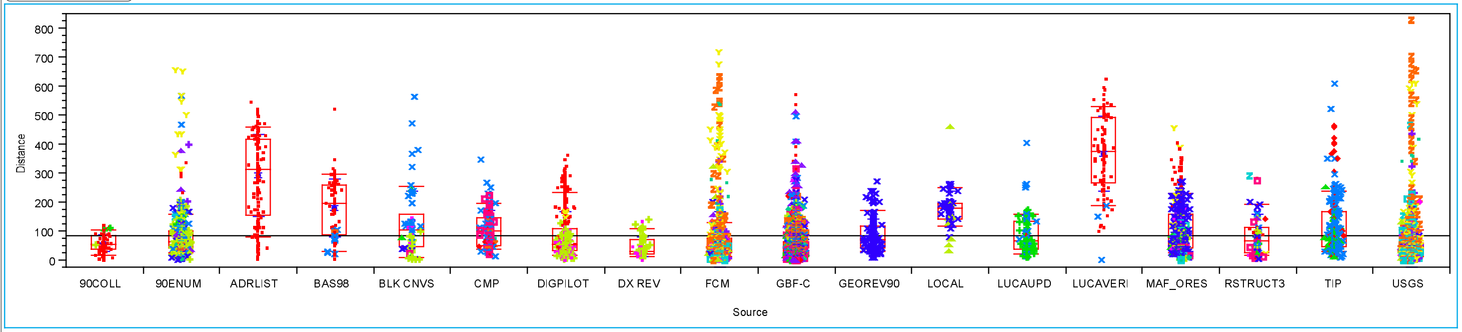
Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
04013040508	684	225.516	147.84	5.65
04013060900	161	54.089	45.94	3.62
06067007410	664	93.869	66.02	2.56
06067008115	192	72.750	42.32	3.05
12057013201	254	46.061	41.58	2.61
12057013905	360	67.091	71.20	3.75
22103040304	264	308.700	2638.01	162.36
22103040501	347	45.721	43.09	2.31
32003004903	374	56.262	45.94	2.38
32003005897	607	68.043	37.35	1.52
39041011420	483	106.280	92.87	4.23
42133021210	295	46.876	50.40	2.93
42133021220	166	74.343	97.19	7.54
42133021400	343	44.495	35.44	1.91
50025967000	51	54.282	32.60	4.56
50025967200	541	100.043	178.16	7.66
50025967700	277	57.492	75.54	4.54
50025968300	175	64.761	65.11	4.92
50025968400	155	64.089	113.63	9.13
50025968500	194	35.482	31.50	2.26
50025968600	137	62.591	263.16	22.48
50025968700	140	70.016	174.93	14.78

Attachment E

GTAAT: Source/Operation Comparison

Distance By Source



Quantiles

Level	minimum	10.0%	25.0%	median	75.0%	90.0%	maximum
90COLL	3.24	20.63	38.4425	56.73	86.7425	106.17	121.23
90ENUM	2.69	26.668	44.27	66.085	103.485	159.319	657.42
ADRLIST	6.4	80.319	158.1725	316.795	420.1775	462.01	546.46
BAS98	5.13	32.702	88.44	198.705	259.5325	296.302	523.93
BLK CNVS	4.83	12.539	46.205	104.39	160.48	255.177	564.28
CMP	15.06	40.7	53.575	104.06	149.59	198.26	345.98
DIGPILOT	2.86	19.93	34.665	56.68	112.5075	236.645	364.15
DX REV	4.22	16.382	22.9175	33.035	71.505	110.199	141.86
FCM	2.63	15.802	25.5175	44.985	75.985	132.123	719.19
GBF-C	0.13	11.885	20.455	34.86	66.745	119.575	574.04
GEOREV90	8.85	30.796	45.26	71.8	118.375	173.152	272.11
LOCAL	32.95	118.438	144.075	181	199.35	252.596	462.33
LUCAUPD	13.91	21.359	40.1775	69.775	135.4525	159.217	406.52
LUCAVERI	3.33	191.564	270.8025	377.91	495.2225	531.771	626.57
MAF_ORES	3.41	26.446	42.81	85.06	160.895	230.91	457.71
RSTRUCT3	6.07	17.716	28.98	68.69	115.45	192.406	293.95
TIP	11.21	30.035	49.8275	86.395	169.52	241.586	612.56
USGS	1.1	13.78	23.0875	36.685	60.1725	105.242	833.31

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
90COLL	84	61.036	30.411	3.318
90ENUM	488	87.634	84.067	3.806
ADRLIST	170	293.357	145.276	11.142
BAS98	80	181.042	102.534	11.464
BLK CNVS	62	126.733	114.541	14.547
CMP	109	109.293	63.236	6.057
DIGPILOT	370	90.030	81.298	4.226
DX REV	60	49.156	34.863	4.501
FCM	862	74.245	99.494	3.389
GBF-C	1874	54.249	57.202	1.321
GEOREV90	137	88.594	56.949	4.866
LOCAL	53	178.255	63.707	8.751
LUCAUPD	88	88.686	67.215	7.165
LUCAVERI	110	367.297	129.990	12.394
MAF_ORES	577	109.498	83.346	3.470
RSTRUCT3	55	87.885	71.757	9.676
TIP	270	117.623	96.930	5.899
USGS	1328	59.191	86.367	2.370