Censustigeresriuc030706

Use of Conflation Software in US Census Bureau's TIGER Coordinate Correction Process

Prepared for ESRI User Conference 2003
San Diego, CA

Prepared by Bernard Catalinotto
Geo-Information Services, JV
Alexandria, VA

July 10, 2003

OVERVIEW

This paper covers the following topics:

1. Background
2. Coordinate Correction Process
3. Coordinate Correction Tasks
4. What is Conflation?
5. What is MapMerger?
6. Vendor 1 – Standard Approach to MapMerger
7. Vendor 2 – Creative Approach to MapMerger
8. Results and Next Steps

1. BACKGROUND

The US Census Bureau is in the process of correcting the positional accuracy of the TIGER/Line coordinates of all 3300 counties and non-county geographic units in the USA. The Bureau has developed the following programs to correct the TIGER database:

1. Initial Programs
    a. R&D, the "ITS-088" program, 77 files
    b. Interim development, the "Task 4/Geo-Serv" program, 190 files
2. Major Program
    a. The MTAIP program, Census Bureau working with major contractor .Harris Corporation, ~3000 files

This paper focuses on some of the conflation methods used in the interim Task 4/Geo-Serv program.

2. COORDINATE CORRECTION PROCESS

The Task 4/Geo-Serv process of correcting 190 county files includes the following subtasks:

1. Provision of TIGER and Local GIS Files.....Census
2. Data Preparation of Local File……Geo-Serv/Census
3. Coordinate Correction……..……Vendors
4. Geo-Serv QC……Geo-Serv
5. Final Fixes……Geo-Serv
6. Census QC……Census
7. Upload into TIGER DB……Geo-Serv/Census

The remainder of this report focuses on the third sub-task listed above, coordinate correction.

3. COORDINATE CORRECTION TASKS

Each vendor has created an independent production method for coordinate correction. One vendor has created a platform that uses powerful custom software to implement the matches. Two vendors have chosen MapMerger as a key element of their process. They have supplemented MapMerger with custom software that tailors the results to meet the Census Bureau's specifications. The major steps in the processes created by the two "MapMerger" vendors are as follows:

1. Conflate TIGER to Local file
    - Match all matchable features
2. Find and Fix Errors
    - Flag issues, resolve, fix, flag action
3. Re-Format, Prepare Deliverables, Final Check

The balance of this paper focuses on the Step 1, the MapMerger conflation step, as it is performed by the two "MapMerger vendors."

5. WHAT IS CONFLATION?

GIS conflation is the process on combining the graphics and attributes of two or more GIS files to create a third, hopefully better, file.

In a typical conflation project, the major contributor of attributes is called the source file and the major contributor of graphics is called the target file. The source features are matched to the target features and the source attributes are transferred to the matching target graphics. In the most simple case, the "brains" of the source file is combined with the "beauty" of the target file to yield a high quality "result" file. Unmatched features from the source file (sometimes called "orphans") are rubber-sheeted and aligned with the target file to create a result file that includes features from both files.

6. WHAT IS MAPMERGER?

MapMerger is an ArcMap extension for automated conflation. It provides tools for automatic and manual matching of source files to target files and enables the user to create a conflated "result" file in an efficient manner.

The standard method of using MapMerger is to first match the nodes of each file and then match the lines (arcs) of each file. The two MapMerger vendors have taken different approaches to using this tool, as described below.

7. STANDARD APPROACH

The standard approach, as noted above, is to first match all nodes and then match all lines. Vendor 1 uses the standard approach but has added the beneficial step of

performing the node matches on separate layers, which are easier to match, rather than trying to match the entire source file to the entire target file, as follows:

Node Match

Vendor 1 divides the TIGER and local files into separate "roads," "hydrology" and "other" layers. Nodes that indicate intersections with different layers – for example bridges where hydro crosses roads -- are maintained as pseudo-nodes in each layer.

The layers are matched in two rounds. The initial round uses a maximum search distance of 45 meters; the second uses 60 meters. After each round, the technician conducts a visual check and fixes the errors manually. MapMerger allows the user to mark bad matches so they will be deleted and not recur

To complete the node matching process, the node matches from each layer are combined and are applied to the combined TIGER and local files, and the result is edited manually in MapMerger. It is important to create a perfect node match before moving on to the line match in order to avoid generating mismatches in the result layer.

Line Match

The line match feature matches vectors in the source file, TIGER, to vectors in the target local file.

Line matches are also implemented in two rounds. The initial round uses a maximum search distance between lines of 45 meters; the second round uses 60 meters. After each round, the technician conducts a visual check and fixes the errors manually..

The result of this approach is to perform approximately 80-90% of the possible matches while in MapMerger, taking full advantage of its editing tools.

Vendor 1 reports that MapMerger takes about half the time needed for the coordinate correction task. The remaining time is used to run custom software that finds potential errors and fixes them automatically or manually (or, if the potential errors prove upon further investigation to be correct, they are flagged as correct so they will not be found again during subsequent programmatic checks).

8. CREATIVE APPROACH

Vendor 2 has taken a very different approach to using MapMerger in the coordinate correction process, what we consider a very creative approach.

This vendor skips the node matching and starts out with line matching. They do not work by layer, they work only on the complete TIGER and combined local files. They start with line matching, working in rounds.

They start out by running the automatic line matcher with the incredibly small search area of one meter (compared to 45 meters for the standard approach), and check the results. If there are no errors, they add one meter to the search area and run the automatic line matcher for a second round and so on until they encounter a mismatch.

This vendor has processed a wide variety of files. They find that they usually encounter matching errors at 7 meters. They then use the manual matching tools in MapMerger to match the remaining lines.

They intensely use Census's QC software while still in the MapMerger phase of production to identify errors. They also run their own checkers and fix the errors manually in MapMerger.

As a result they achieve 90-100% of all possible matches in MapMerger.

They then run their final automatic and visual QC checks and make their final fixes outside of MapMerger, and then prepare the final deliverables.

9. RESULTS & NEXT STEPS

Although the methods used are different, the results in both cases are an efficient use of the MapMerger conflation tool, in combination with custom GIS software.

All three vendors are making significant contributions to the success of the Task 4/Geo-Serv project. The next steps are to continue to apply these procedures with the goal of correcting coordinates of all 190 counties by the end of the fiscal year, September 30, 2003.