

Getting Started with TIGER: A Tutorial

Gary F. Plazyk, A.T. Kearney, Inc.

What Is TIGER?

TIGER is an acronym for Topologically Integrated Geographic Encoding and Referencing system, the geographic reference data base for the 1990 U.S. Census. The TIGER files provide point and line data for a variety of geographic feature types. These feature types include highways and streets, railroad lines, waterways, political entities such as states and counties, and census-related boundaries such as census tract and block. Additional information includes alternate feature names, address ranges, and ZIP code data.

The Census Bureau plans to offer a variety of products based on the TIGER data base:

- * **TIGER/Line Files.** The first of the TIGER products, this is primarily the coordinates of all the line segments in the TIGER data base. Available initially as the "Prototype", presently the "Precensus", and soon the "Census" versions.
- * **TIGER/Database.** In Federal Spatial Data Transfer Standard format, this will provide more point and area information. To be released in 1991.
- * **TIGER/Boundary.** Similar to the 1980 County Boundary File. To be released in 1991.
- * **TIGER/Comparability.** Shows comparability between the 1980 and 1990 census tracts. To be released in 1991.
- * **TIGER/GICS.** Provides the names and codes for all geographic areas for which census data will be available. To be released in 1991.

For the remainder of this paper, TIGER should be understood to refer to the Precensus TIGER/Line files.

The primary TIGER file data element is a line segment, designated by the latitudes and longitudes of its origin and destination points. The line type of the segment identifies the feature type (such as road or census tract boundary). The line is identified with a name, if applicable (such as street name). The lines are *oriented*, that is, traversed from the origin to the destination point; this allows the identification of areas to the *right* and to the *left* of the line. This allows selection of lines that form the boundaries of areas such as counties or census tracts.

Lines are modified by *shape points*. These are points that provide a finer resolution, such as wiggles in a road, to the portion of the line between the origin and destination.

The TIGER line segments may be read into a SAS* data set and plotted using the SAS/GRAPH ANNOTATE facility. However, they are not suitable for use as map outlines as provided. The segments must be linked to form a closed polygon in order to be used as a SAS/GRAPH map outline. John Blodgett has done the primary work developing the conversion of TIGER line data to SAS/GRAPH map data set format.

TIGER files are grouped by county, identified with the county's FIPS (Federal Information Processing Standard) code. This is a five-digit code; the first two digits

identify a state, the remaining three digits identify a county within a state. These are the same codes that SAS/GRAPH state and county map data sets are keyed to.

Each county may have up to six files, one for each of six TIGER record types:

1. Line segments - the basic individual feature segment data records.
2. Shape points - modify type 1 line segments that are not straight lines.
3. Additional census geographic area codes.
4. Index to alternate feature names.
5. Feature name list.
6. Additional address range and ZIP code data (does not apply for some counties).

The TIGER data provides an accuracy of approximately 50 feet; this resolution will be a factor in determining the suitability of its use for a particular application. TIGER data is suitable for mapping and vehicle routing; it is not adequate for determining legal property boundaries or for remote-controlled vehicle navigation.

TIGER Data Distribution Media

TIGER files are distributed on magnetic tape and CD-ROM (compact disk read-only memory) media. Magnetic tapes are priced by county; CD-ROMs by state. The CD-ROM media is much cheaper than magnetic tape; most states come complete on one CD-ROM, and cost \$250 for the entire state (As of March 1, 1991, the Precensus TIGER/Line files were reduced in price to \$100 apiece). Sometimes more than one state is contained on a single CD-ROM. This discussion will focus on using TIGER data on the CD-ROM media.

A CD-ROM is read using a CD-ROM reader attached to a computer; to the PC, the CD-ROM reader appears to be just another hard drive (albeit read-only). Although the TIGER CD-ROMs conform to ISO Standard 9660, a few older CD-ROM readers may experience difficulty reading the disks. The data files are stored as flat ASCII fixed-format files.

Figure 1 shows a typical TIGER CD-ROM subdirectory tree and file structure. Exploring a CD-ROM disk is easy with an MS-DOS file-review application program like the Norton Commander. A TIGER CD-ROM contains:

- * PRETIGER.DOC, a documentation file containing the TIGER record layouts, field descriptions, and code cross-reference tables.
- * FILECODE.DOC, a cross-reference file of the FIPS codes for states and counties.
- * one subdirectory for each state, identified with the state's 2-digit FIPS code.
- * one sub-subdirectory for each state's county, identified with the county's 3-digit FIPS code. This subdirectory contains one file for each of the six TIGER file types for this county.

These files can be quite large; a typical county may consist of six files, totalling 5 to 10 megabytes. In a few cases, the files are much larger; the files for Cook County, Illinois (Chicago) occupy 58 megabytes. An average state requires about 400 megabytes.

Moving TIGER files can be awkward because of their size. Most of these files are far too large to fit on a single floppy diskette. Several techniques that I have found to help move files from one computer to another are:

- * Copy the files using a direct-connect utility such as LapLink 3.
- * Compress the files using an archiving utility such as the Norton Backup, DS-Backup, or PK-ZIP.
- * Transfer the files to a mainframe computer via a PC-mainframe link, and then back to another PC.

TIGER Record Type 1: Basic Data Record

The TIGER type 1 record provides the basic information about a line segment. These records are fixed-format 228-character ASCII records. Some of the fields that appear on this record are:

- * RECNUM, a unique identifier for this line segment.
- * feature name.
- * CFCC, the census feature class code. This code indicates what type of feature this line segment actually is. The documentation for the file lists 128 different types of features, including 59 types of roads, 20 types of rail lines, 16 types of hydrographic features such as streams and lakes, 20 types of non-visible boundaries, and a variety of other types of boundaries.
- * starting and ending addresses for left and right sides of the line segment.
- * ZIP codes on left and right sides of the line segment.
- * state and county FIPS codes on left and right sides of the line segment.
- * census tract and census block on left and right sides of the line segment.
- * latitude and longitude of starting and ending points.

Latitude and longitude are given in degrees, with six additional places to the right of the decimal point. Latitudes in the northern hemisphere are positive and in the southern hemisphere are negative. Longitudes west of Greenwich to 180 degrees west are negative, east of Greenwich are positive. Thus, the vast majority of points in the United States have positive latitudes and negative longitudes. Negative signs do appear in the data, but the decimal points are implied.

Boundaries between areas are derived from the data by examining the codes to the left and to the right of a line segment. If the codes are different, then this line is a boundary between the two areas. One advantage of this coding scheme is that boundaries for multiple areas can be accommodated on a single record.

TIGER Record Type 2: Shape Coordinate Points

Shape points, specified by TIGER record type 2, are used to modify TIGER record type 1 line segments that are not straight lines. These records are fixed-format 208-character ASCII records.

Figure 2 shows how shape points modify a straight line segment. The shape points are keyed to RECNUM, the unique record identifier of record type 1. There may be up to ten shape points per record, and multiple shape point records for any given RECNUM. Shape points coordinates are given in latitude and longitude.

A Simple Application: Read and Plot Line Type 1 Data

Putnam County, Illinois is used for these examples. It is the smallest county in northeastern Illinois for which I had TIGER file data; it is identified by FIPS code 17-155. Putnam County is located approximately 100 miles southwest of downtown Chicago, on the Illinois River just downstream from Peru and LaSalle, Illinois.

It is relatively simple operation to read selected fields from the TIGER record type 1 file and create an ANNOTATE data set. This data set may be plotted with PROC GPLOT. Although this technique does not retain the correct horizontal and vertical proportions, it has the advantage of being fast and self-scaling. This is a good way to test your TIGER file application programs.

Figure 3 shows the result of plotting Putnam County TIGER type 1 records. Select the features to plot by extracting lines having the CFCC feature code desired; Figure 3 shows highway features and the county border.

Expanding the Application: Add Shape Points

A more complex operation is to add the shape point data from the TIGER type 2 records to the basic data. This may be done by using two SET statements in the DATA step creating the ANNOTATE data set.

Figure 4 shows the result of adding shape point data to the Putnam County highway and county outline features. Observe how the shape points add curves to the roads and the northern county border.

And Getting Back to SAS/GRAPH: Add a SAS/GRAPH Map Outline

Our final application involves using the SAS/GRAPH map data set for a base map. For compatibility with the unprojected SAS/GRAPH map data, it is necessary to multiply the TIGER longitudes (X-coordinates) by -1.

Figure 5 shows water data (including shape points), in the form of an ANNOTATE data set. Solid lines are perennial hydrological features; broken lines are intermittent features. This data has been superimposed on the SAS/GRAPH map for Putnam County, extracted from MAPS.COUNTY.

Compare the county outline extracted from the TIGER data with the SAS/GRAPH map. Note the significantly higher level of detail provided by the TIGER data.

Areas

As noted earlier, the TIGER/Line files provide *line segment* data. This is a different organization of data than the SAS/GRAPH map data sets, which provide *area* data by specifying a series of lines that circumscribe a closed polygon. Closed polygons are required to generate choropleth, or filled-area maps. Linking the TIGER data lines into polygons is a subject beyond the scope of this discussion. For more information on this topic, review John Blodgett's paper noted below, and download the files TIGER.SAS and TIGER.DOC from the SAS Institute Online Customer Support Facility.

Summary

With the release of the 1990 census TIGER line files, computer cartography has taken a great leap forward. As the 1980s were an era of giant strides in computerized business graphics, the 1990s will be an era of extraordinary developments in desktop mapping.

I urge the Institute to help us, its users, by making all of the SAS/GRAPH map data sets available on CD-ROM. The *SAS/GRAPH User's Guide, Release 6.03 Edition* states that "...because of size limitations the map data sets for SAS/GRAPH software Release 6.03 for personal computers do not include the CANADA3, CANADA4, COUNTIES, and STATES map data sets supplied with Version 5 SAS/GRAPH software." I suggest that all of the US, Canada, and world map SAS data sets, as well as 3-digit and 5-digit zip code boundary map data sets, could easily fit on the 650 megabytes of one CD-ROM. Having this data available would lead to a tremendous upsurge in mapping using SAS.

SAS Institute has developed a prototype GIS product for the SAS System that uses TIGER files, demonstrated at SUGI earlier this year. I hope this signals the emergence a new generation of tools for mapping applications driven by the power of the SAS Application System.

Acknowledgements

I appreciate the encouragement given me by Walter Strauss, a Principal at A.T. Kearney. His recognition of the value of geographic analytical techniques and insight into real-world consulting applications was of great value to me.

References

The following publications will be of value to users of TIGER files and SAS:

TIGER: The Coast-to-Coast Digital Map Data Base. U.S. Department of Commerce. Washington. 1990.

TIGER/Line Precensus Files, 1990 Technical Documentation. Prepared by the Bureau of the Census. Washington. 1989.

TIGER Questions and Answers. U.S. Department of Commerce, Bureau of the Census. Washington. 1989.

"TIGER/Line Precensus Files on CD-ROM, 1990", *Data User Services Division Data Developments.* Bureau of the Census. Washington. 1990.

"TIGER on CD-ROM - the Wave of the Future!", *Census and You*, Volume 25, No. 2. U.S. Department of Commerce, Bureau of the Census. Washington. February 1990.

Planned TIGER System Products. Robert A. LaMacchia & Silla G. Tomasi. U.S. Census Bureau Geography Division. Prepared for the 1990 Annual Spring Conference of The Government Publications Librarians of New England, Amherst, Massachusetts, March, 1990.

"Software Tools for Processing U.S. Census Bureau TIGER Line Files", *Proceedings of the Fifteenth Annual SAS Users Group International Conference*, pp. 942-946. John Blodgett & John McIntyre. SAS Institute, Inc. Cary. 1990.

"Using the Census Bureau's TIGER Files with SAS/GRAPH Software" (First in a three-part series), *SAS Communications*, Vol. XV, No. 3, First Quarter 1990, pp. 35-37. SAS Institute, Inc. Cary. 1990.

"Using the Census Bureau's TIGER Files with SAS/GRAPH Software" (Second in a three-part series), *SAS Communications*, Vol. XV, No. 4, Second Quarter 1990, pp. 35-37. SAS Institute, Inc. Cary. 1990.

"Using the Census Bureau's TIGER Files with SAS/GRAPH Software" (Third in a three-part series), *SAS Communications*, Vol. XVI, No. 1, Third Quarter 1990, pp. 35-36. SAS Institute, Inc. Cary. 1990.

TIGER.SAS and TIGER.DOC files, SAS Institute Online Customer Support Facility. June, 1991.

"SUGI 15 - Nashville! You Are Here", Handout from the SAS/GIS demonstration booth at SUGI 15. 1990.

I would be happy to provide copies of the programs used to generate Figures 3, 4, and 5. All the examples in this paper were generated on a Compaq SLT/386 running MS-DOS 3.3 and SAS 6.04. The maps were printed on an Apple Laserwriter II PostScript printer.

Author

Gary F. Plazyk
A.T. Kearney, Inc.
222 S. Riverside Plaza
Chicago, IL 60606
Telephone: (312) 993-8970

Trademark Notice

* SAS and SAS/GRAPH are registered trademarks of SAS Institute, Inc., Cary, NC, USA.

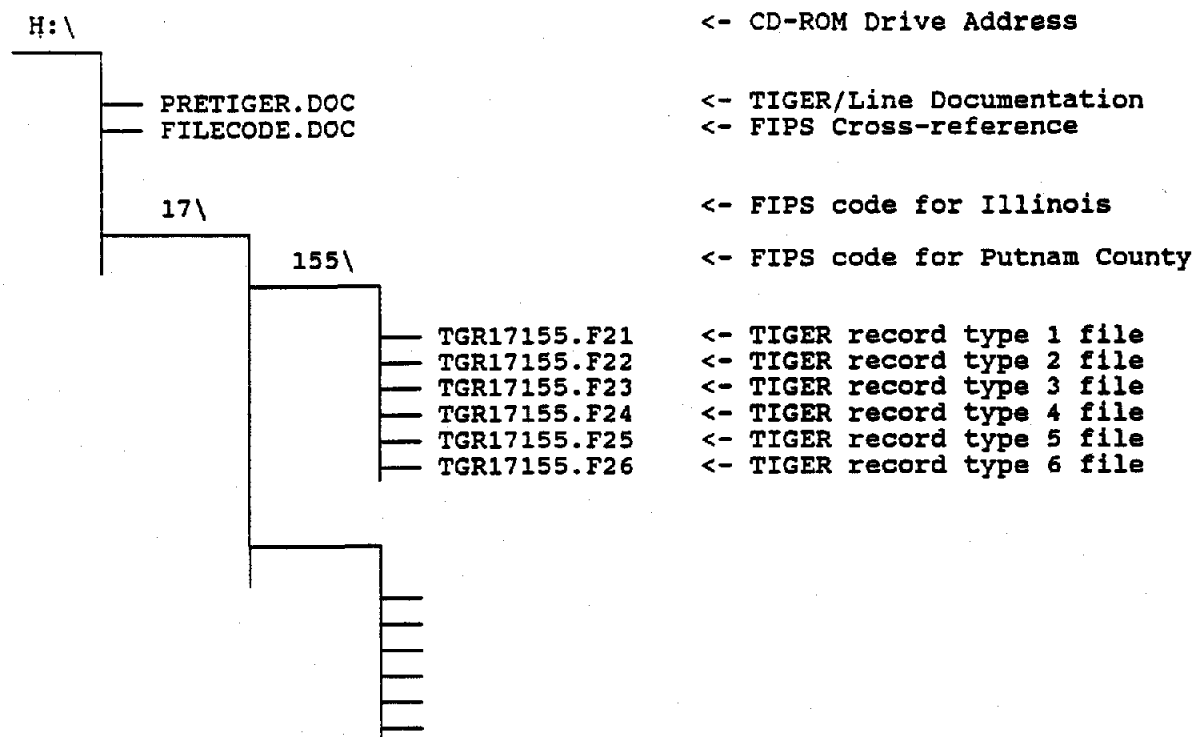


Figure 1. TIGER CD-ROM Subdirectory Tree and File Structure

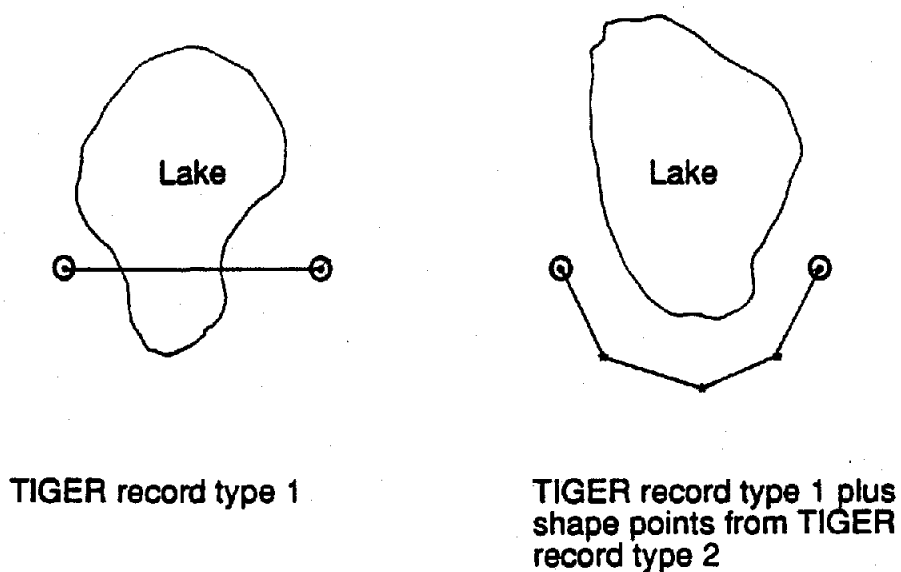
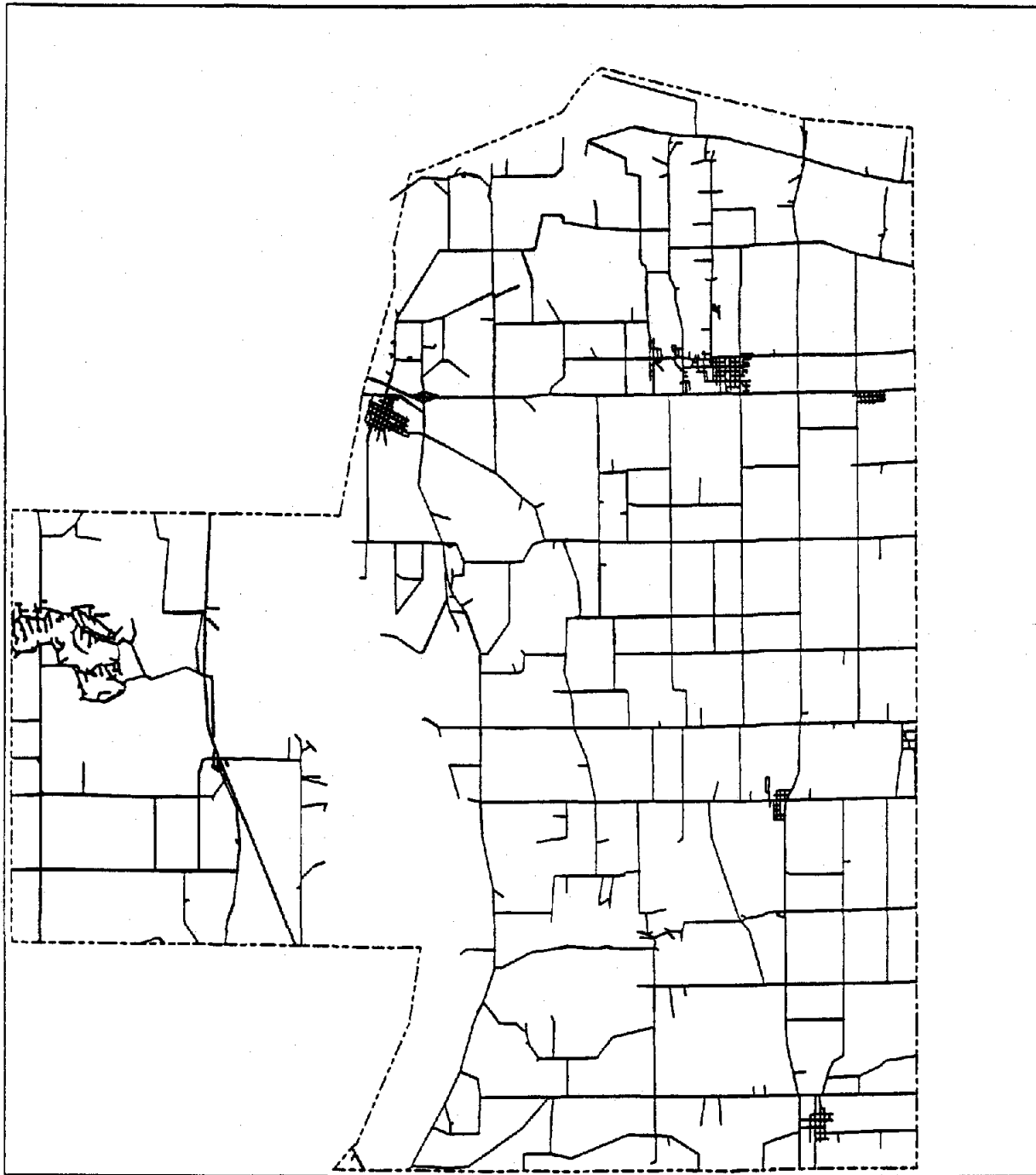


Figure 2. Effect of Shape Points on Basic Line Segment

TIGER File 1 for FIPS 17-155

Putnam County, IL

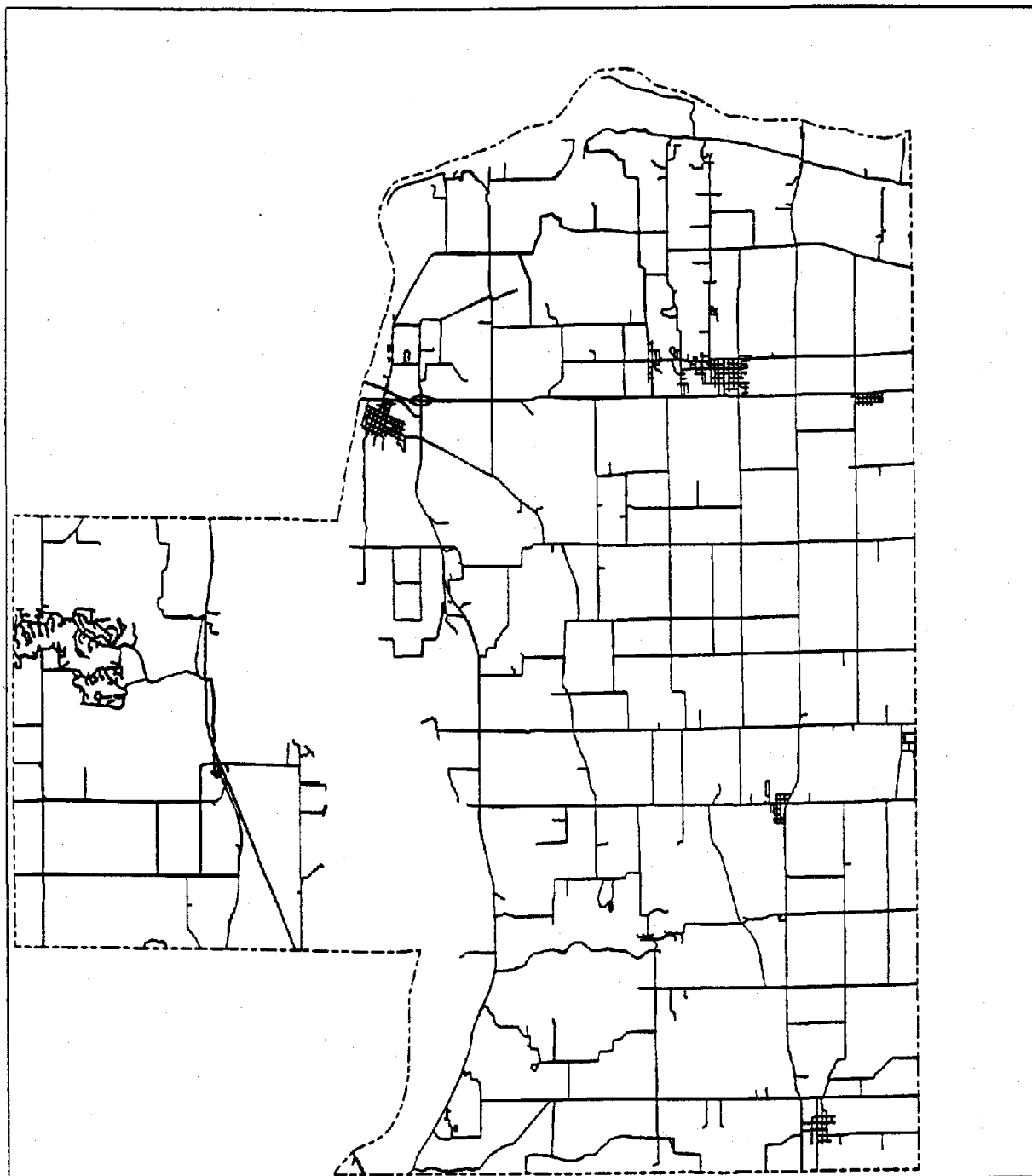


Road Features

Figure 3. Road Features, County Outline From TIGER Type 1 Records

TIGER Files 1 and 2 for FIPS 17-155

Putnam County, IL

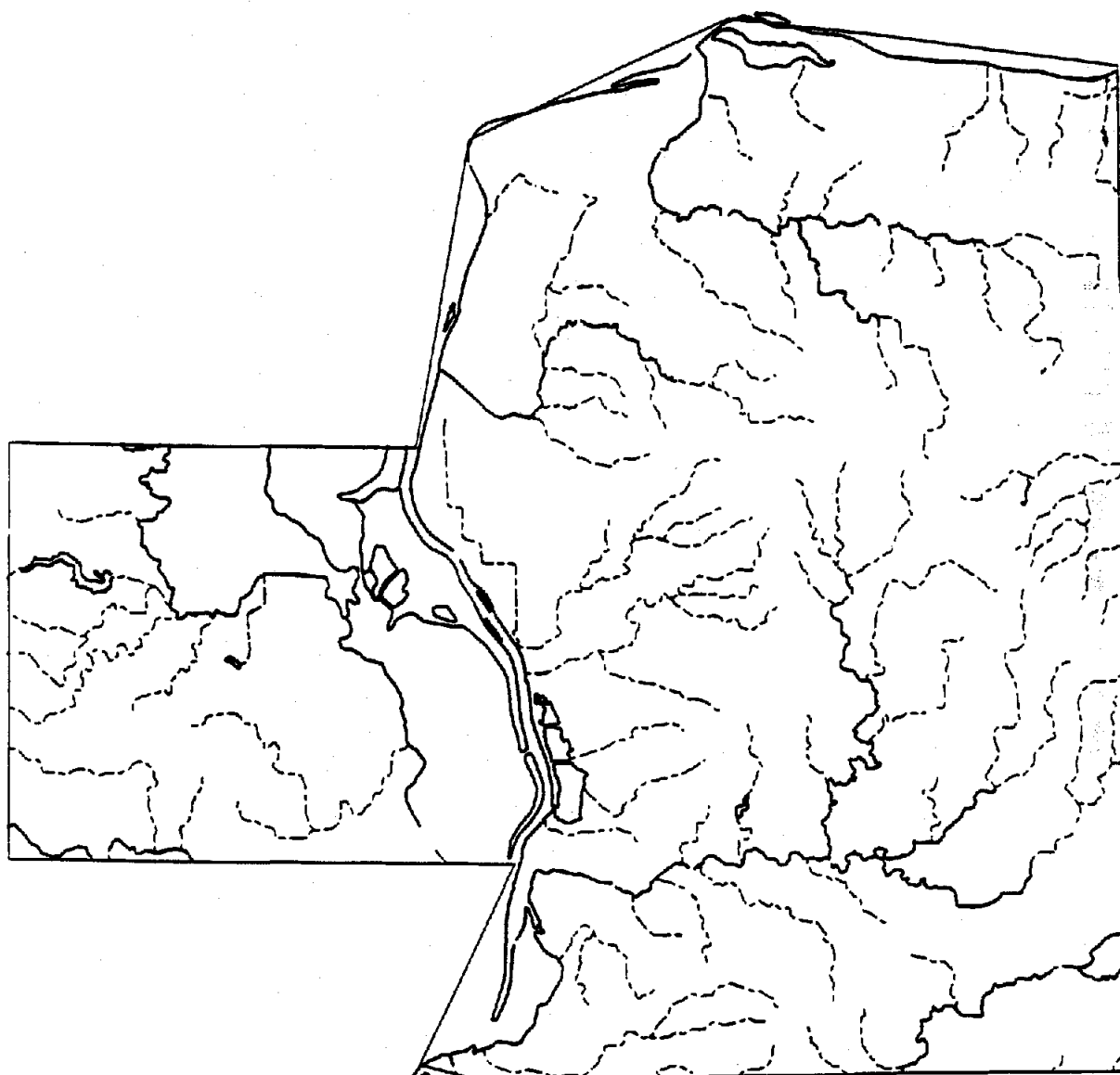


Road Features

**Figure 4. Road Features, County Outline From TIGER Type 1 Records
Modified by TIGER Type 2 Records (Shape Points)**

TIGER Files 1 and 2 for FIPS 17-155

Putnam County, IL



Hydrological Features

**Figure 5. Water Features From TIGER Type 1 Records Modified by
TIGER Type 2 Records (Shape Points), Overlaid on
County Outline Map From SAS/GRAPH MAP.COUNTY**