DIGITIZED HYDROLOGIC UNITS FOR THE UNITED STATES AT A SCALE OF 1:2,500,000

Kenneth J. Lanfear and James C. Schornick, Jr.

U.S. Geological Survey Open-File Report 85-92

ABSTRACT

Boundaries of the hydrologic cataloging units for the United States were digitized from a 1:2,500,000 scale map. The data set produced is continuous across the conterminous United States, and should be very useful for national and regional studies of water resources and displays of choropleth maps of data applicable to the units.

INTRODUCTION

The U.S. Geological Survey (USGS) in cooperation with the U.S. Water Resources Council (WRC) established a uniform, nationally consistent set of maps delineating the hydrologic boundaries of the major United States river basins (USGS, 1982). This four-level hierarchical delineation consists of 21 WRC regions (18 in the conterminous United States, and 1 each in Alaska, Hawaii and the Caribbean), 222 WRC planning subregions, 352 National Water Data Network accounting units, and 2,149 cataloging units. Each hydrologic unit is identified by a unique 8-digit code: the first 2 digits identify the region, the first 4 the subregion, the first 6 the accounting unit, and all 8 the cataloging unit. The hydrologic units were reviewed extensively by all principal Federal, regional, and State water-resources agencies, and these hydrologic unit codes and basin names were adopted as a Federal Information Processing Standard in 1983. All four levels are published on a 2-piece 1:2,500,000 scale map (USGS, 1980).

A digital representation of this map is needed by USGS project personnel for use in computerized mapping systems. These mapping systems can show, for example, the hydrologic units overlaying county boundaries and compute the areas of the intersecting polygons; this makes it possible to relate data aggregated at the county level (such as population data) to hydrologic units. Choropleth maps of the hydrologic units can also be produced to show each unit in a color that is related to the value of a selected parameter. For example, all cataloging units where a water quality constituent exceeds some standard could be colored red.

The digitized map is in arc-node format, with each digitized line (arc) representing a segment of a cataloging unit boundary that begins at one node and ends at another (possibly the same) node with no intervening nodes. Any number of arcs can come together at a node. Any boundary at any hierarchical level can be constructed from the appropriate cataloging unit arcs.

A digitized national map should be continuous over the entire United States, with no gaps or "slivers" in the digitization. To accomplish this, all arcs coming together at a node have to be given precisely the same starting (or ending) point. This ensures an exact closure for all polygons and allows a precise tracing of any border. Since the source map was in two parts, each too large for a digitizing tablet, the digitizing technique had to be able to "edge match" numerous digitized files.

METHODS

The main hardware used for this project was a Tektronix (1) 4956 graphics tablet driven by a Compustar microcomputer. The graphics tablet, or digitizer, was 20 x 20 inches in size, with an accuracy of 0.004 inches, and was connected to the host microcomputer by an RS-232 connection. A 10 megabyte hard disk was attached to the microcomputer, although 5 1/4 inch floppy disks could also be used. A Tektronix 4662 plotter was also connected, so that digitized data could be displayed.

The software for digitizing was a set of programs written in the PL/1 language. These were developed in-house for earlier projects, and

---

1/ Use of trade names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

needed few modifications. The map digitizing principle used by this software is that of the Spatial Environmental Data Digitizing System (SEDDS) developed by Lanfear and Nakassis (1980): given a map projection and at least 3 reference points (expressed in latitude and longitude), a transformation can be found for converting all digitized points to latitude and longitude. The first program, called SETREF, orients a map on the digitizer by establishing the map projection and the reference point locations. The main digitizing program, DIGALLL1 (Digitize Albers to latitude/longitude), prompts the user through the digitizing process and records the data on a disk, after converting all points to latitude and longitude. A final program, CONVDIG, checks the data and removes excess points.

A key to the success of this system is its ability to immediately match the end points of arcs to particular nodes. Each node is assigned a unique number; the operator must enter the numbers of the "from" node and the "to" node before digitizing an arc. The first time a node is encountered, it's latitude and longitude is stored in memory; alternately, the location can be read from a file prior to digitizing. Each time the node is subsequently encountered, its location is automatically adjusted to the location stored in memory. If the adjustment is too large -- indicating an operator error -- the record is flagged.

Because of their size, the two map halves were digitized in 76 page-size sections. Each page was photocopied from a paper copy of the original map. (Linear distortions introduced in the copying process are removed by program SETREF.) Local node numbers were assigned to each

node on the page. For nodes on the boundary with another page (pages were allowed to overlap), a node file was prepared with the latitudes and longitudes of previously defined nodes. After establishing the reference points with program SETREF, the operator started program DIGALLL1 and was prompted to enter, for each arc, the "from" node, the "to" node, the 8-digit code for the cataloging unit on the left of the arc, and the 8-digit code for the area on the right. The arc was then traced with the crosshairs; between 3 and 15 points per second were recorded. Upon completion, the identifying information above was saved on a header record, followed by the latitudes and longitudes of all digitized points, one point per record.

The raw digitized files were stored in an ASCII format, and could be read and changed with a text editor. This, plus the ability to insert a remark in the files during digitizing, made correcting mistakes relatively simple. However, the raw data contained many more points than were needed. Program CONVDIG read the raw data and reduced each arc to the minimum number of points needed for an acceptable representation. It accomplished this using a data reduction algorithm by Douglas and Peucker (1973), which removed points lying within 0.5 km of a straight line. This algorithm proved particularly good at removing points from long, nearly straight lines, yet preserving detail in crooked lines. The number of points was usually reduced by a factor of 5 to 10.

The reduced files were transferred to a Prime minicomputer of the Water Resources Division's Distributed Information System (DIS). The left/right hydrologic unit code indicators stored with each arc were

checked against a list of valid codes to flag bad numbers and to ensure that digitized codes were present for all cataloging units. Each non-coastal cataloging unit was checked for closure by ensuring that all arcs for a unit formed one or more (e.g. islands) closed polygons with endpoints of connecting arcs matching exactly. If closure was prevented by incorrect left/right indicators, this information was displayed in the terminal for corrective action. Coastal units were not closed, but were extended a short distance from the shore so that they can be "clipped" to other data sets of the coastline.

After some additional editing, the files were merged into one large file for distribution. To save space, the data were stored unformatted rather than as an ACSII file. The example FORTRAN77 program in Appendix A shows how to read these files.

RESULTS

The final digital data set for the 1:2,500,000 scale cataloging unit boundaries in the conterminous United States contains 68,737 points in 6,179 arcs. This data set was installed in the DATAGRAF (Sigma Data, 1983) software package and plotted. Boundaries of the 18 water resource regions are shown in figure 1, subregions in figure 2, accounting units in figure 3, and the cataloging units in figure 4.

The digital data file, with coastal arcs "unclipped" and coastal polygons left open, is available on a Prime computer located in Reston, Virginia (DIS node "QVARSA"). Selected portions may be extracted and transmitted over the DIS network. Magnetic tape copies are available

for transmittal to other locations.

REFERENCES

Douglas, D., and Peucker, T., 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, Vol. 10, No. 2, December, 1973.

Lanfear, K., and Nakassis, A., Documentation and users' guide to the spatial environmental data digitizing system, SEDDS, U.S. Geological Survey Open-file report 80-871, 1980.

Sigma Data, 1983, DATAGRAF reference manual, MA-COM Sigma Data, Inc., Rockville, Maryland, 1983.

U.S. Geological Survey, 1980, Hydrologic unit map of the United States, 1:2,500,000 scale.

U.S. Geological Survey, 1982, Codes for the identification of hydrologic units in the United States and the Caribbean outlying areas, Geological Circular 878-A, 1982.

ACKNOWLEDGEMENTS

The digitized hydrologic units described in this report were prepared cooperatively among the Office of National Water Summary and Long Range Planning, the Water Quality Branch, and the Office of Water Data Coordination of the U.S. Geological Survey. The help of Jon Bradley, Steven Flammey, and Linda Stein in operating the digitizing equipment is greatly appreciated.

APPENDIX A

```
c     Example FORTRAN77 program to show how to read the digitized
c        cataloging unit files.
c
      real dlon(500),dlat(500)
c
      open (10,file='huc.dat',form='unformatted')
c

c     ***** Read the process the arcs one-by-one.
      do 100 i = 1, 100000
c
c        ***** Read the header for the arc.
c              n   -- number of points
c              ial -- 8-digit code of area on the left
c              iar -- 8-digit code of area on the right
         read (10,end=900) n,ial,iar
c
c        ***** Read each point in the arc.
         do 110 j = 1, n
            read (10) dlon(i), dlat(i)
  110       continue
c
c        ***** You have the n points of the arc in storage, and
c              can plot it, save it, or do whatever else you want.
c
c        ***** Go back for another arc.
  100    continue
c
c     ***** End of the file.
  900 continue
      stop
      end
```

Figure 1. Map showing water resource regions of the conterminous United States

Figure 2. Map showing the hydrologic subregions of the conterminous United States.

Figure 3. Map showing the hydrologic accounting units of the conterminous United States.

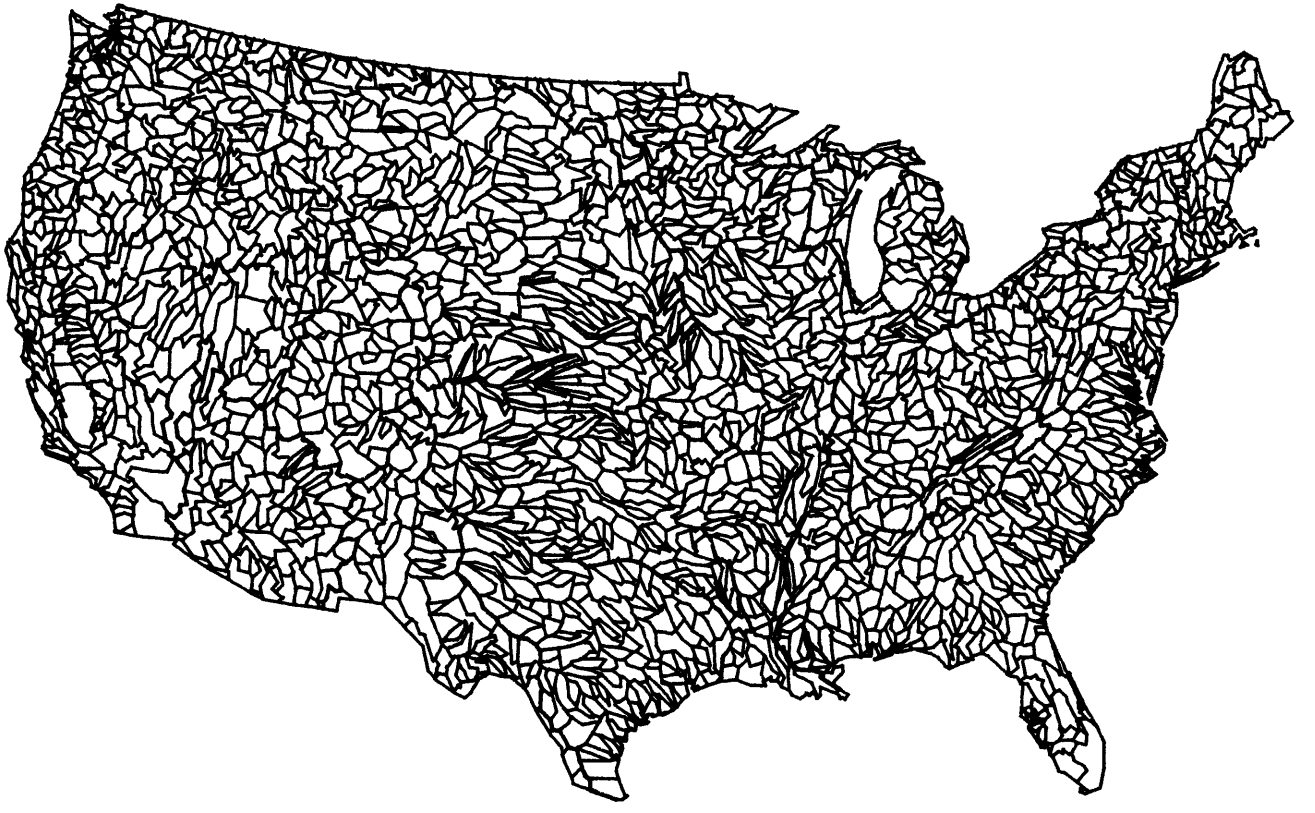

Figure 4. Map showing the hydrologic cataloging units of the conterminous United States.