REVIEW OF EDGEMATCHING PROCEDURES FOR DIGITAL CARTOGRAPHIC DATA USED IN GEOGRAPHIC INFORMATION SYSTEMS (GIS)

By Douglas D. Nebert

U.S. GEOLOGICAL SURVEY
Open-File Report 89-579

Prepared in cooperation with
BONNEVILLE POWER ADMINISTRATION

Portland, Oregon
1989

DEPARTMENT OF THE INTERIOR
MANUEL LUJAN, Jr., Secretary
U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director

For additional information write to:

Oregon Office Chief
U.S. Geological Survey
10615 S.E. Cherry Blossom Drive
Portland, Oregon 97216

Copies of this report can be purchased from:

U.S. Geological Survey
Books and Open-File Reports Section
Federal Center, Bldg. 810
Box 25425
Denver, Colorado 80225

CONTENTS

ILLUSTRATIONS

CONVERSION FACTORS AND ABBREVIATIONS

For the convenience of readers who may prefer to use metric (International System) units rather than inch-pound units used in this report, values may be converted by using the following factors:

Multiply inch-pound units	By	To obtain metric units
	<u>Length</u>	
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)

REVIEW OF EDGEMATCHING PROCEDURES FOR DIGITAL CARTOGRAPHIC DATA USED IN GEOGRAPHIC INFORMATION SYSTEMS (GIS)

--

By Douglas D. Nebert

--

ABSTRACT

In the process of developing a continuous hydrographic data layer for water resources applications in the Pacific Northwest, map-edge discontinuities in U.S. Geological Survey 1:100,000-scale digital data that required application of computer-assisted edgematching algorithms were identified. The resulting spatial data sets require line features that match well enough across map boundaries to ensure full line topology when adjacent files are joined in the computer. Automated edgematching techniques are evaluated as to their effects on positional accuracy. Interactive methods such as selective node-matching and on-screen editing also are reviewed. Interactive methods are described that complement the automated methods by allowing supervision of the edgematching procedures in a cartographic and hydrologic context. Common edge conditions encountered in the preparation of the Northwest Rivers data base are described, as are recommended processing solutions. Suggested edgematching procedures for 1:100,000-scale hydrography data are included in an appendix to encourage consistent processing of this theme on a national scale.

INTRODUCTION

The development of statewide or regional spatial data bases for use in Geographic Information Systems (GIS) usually requires map data to be continuous across map edges Q both in position and in characteristics. A primary objective of a cooperative project between the U.S. Geological Survey and Bonneville Power Administration (BPA) has been to develop a regional data base of rivers and their characteristics for the Northwest Rivers Study. In 1985, BPA convened a task force of State and Federal experts in GIS to provide advice on the most efficient way to implement a hydrographic reference data layer and to manage the extensive, associated attribute data bases. The task force recommended that BPA acquire the U.S. Environmental Protection Agency's (EPA) 1:250,000-scale hydrography data base known as the River Reach system and move its hydrologic attributes into the recently completed 1:100,000- scale data base of digital hydrography from the U.S. Geological Survey (Bonneville Power Authority, unpub. data, 1985).

Purpose and Scope

In 1987, the Oregon Office of the U.S. Geological Survey began implementing a set of procedures to process the 1:100,000-scale hydrography data for the Pacific Northwest for assignment with the EPA River Reach codes. The issue of matching the edges of the digital maps was an early consideration in the project and is the subject of this preliminary report. A full description of the processing details used in the project other than edgematching are beyond the scope of this paper.

This paper evaluates alternatives and provides guidelines for other digital data users to follow when edgematching intermediate-scale spatial data, given that no standards for edgematching adjacent data in digital spatial data bases are in common use. Suggested guidelines for edgematching 1:100,000-scale DLG (Digital Line Graph) digital data sets are included in Appendix A.

Definition of Terms used in Spatial Data Processing

An arc is a series of contiguous line segments that have a common identity. It may also be known as a chain. Its endpoints are, by definition, nodes.

Attributes are the numeric or text characteristics which describe a given spatial feature. Attributes are commonly stored in a relational data base structure. Common attributes in a hydrographic data base of streams would include stream name, length, gradient, direction, and stream codes.

Edgematching is the procedure used to adjust features along one or both sides of a common map edge to create a continuous or seamless product when the maps are joined. Edgematching is usually done prior to the joining of adjacent maps in the computer into a single data file so as to minimize automated, unsupervised adjustment of features. Using commercially available computer software, this usually means comparing the edge coordinates of two adjacent spatial data files, verifying if the edge coordinates are within a stated distance (referred to as the match tolerance), and changing the coordinates of one or both sides. The results of edgematching features along the edges of two adjacent maps by applying a predetermined match tolerance are shown in figure 1.

Nodes are the endpoints of an arc. They occur where arcs terminate or at an intersection of two or more arcs.

The process of rubber sheeting is defined as the adjustment of all spatial features in a spatial data file with respect to known registration points. Commercially-available software may allow rubber sheeting to be performed on a complete data file or within a specified part of a data file. All spatial features within the area to be rubber sheeted are globally, rather than selectively, adjusted to match the known registration points. The registration points also may be referred to as control points.

A spatial feature is a cartographic, addressable element in a digital data file. Points, arcs (lines), nodes, and polygons (areas) are examples of spatial features.

Thematic cartography is the application of mapping procedures that result in a contextually correct cartographic product for use in illustrations. The standards of accuracy are dictated by the user and the method of map production.

Topology is defined as the explicit connectivity of spatial features. Topology is required in an arc-node data model (such as a DLG [digital line graph]) for proper associations to be made among features (such as arcs, nodes, points, and polygons) and their attributes.

A vertex is an internal coordinate along an arc that defines the individual line segments.

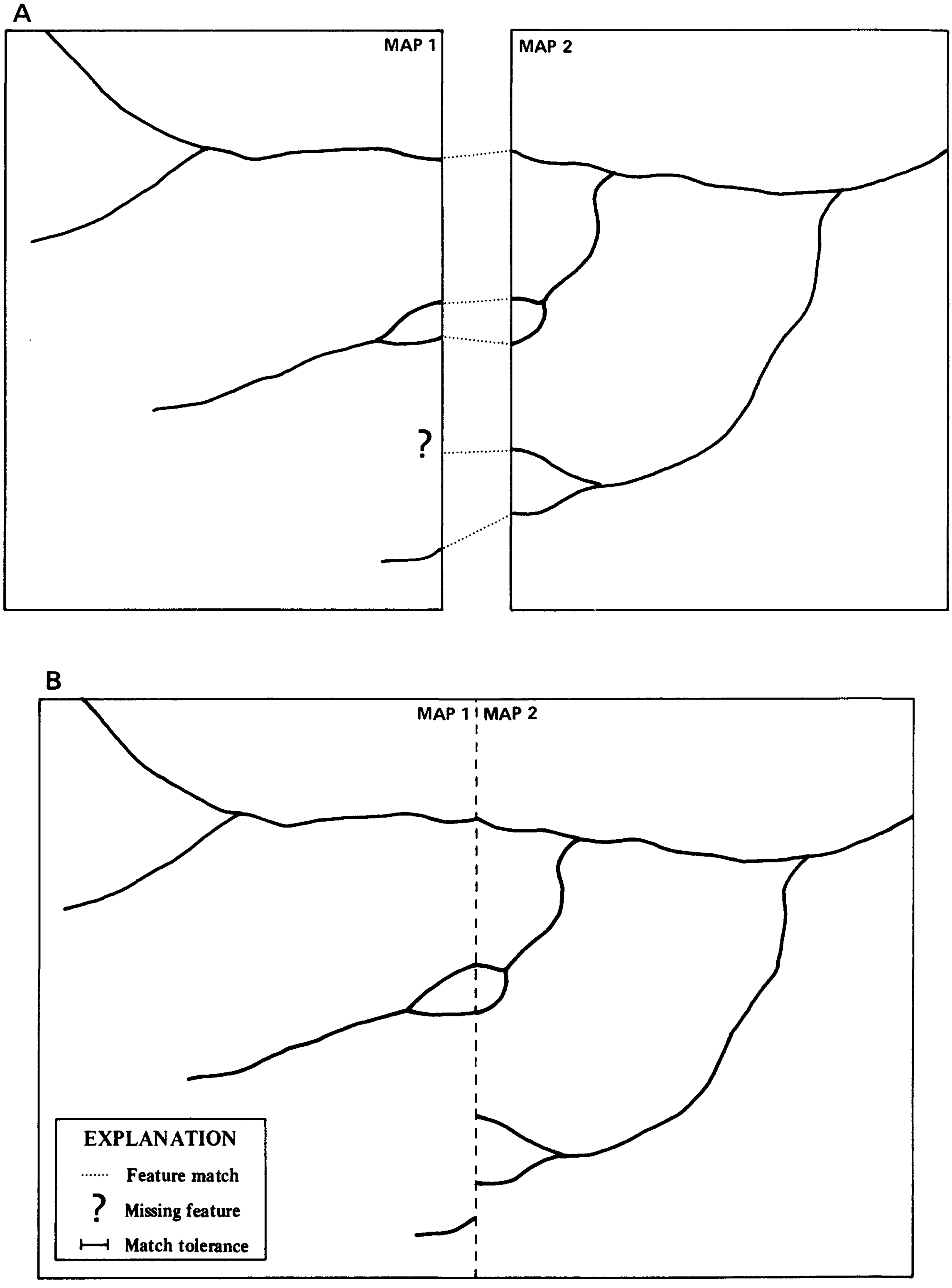

Figure 1.--Typical automated edgematching procedure to adjust edge features within a specified tolerance. Diagram A illustrates possible matches across a map edge. Diagram B shows the effect of applying a match tolerance when matching. Note the lowermost possible match was beyond the match tolerance and was not joined.

REVIEW OF EDGEMATCHING PROCEDURES

Edgematching of adjacent digital map files is a consideration whenever a GIS application study area larger than a single map file is undertaken. Errors due to rounding of coordinates, resolution or date of base materials, and processing digital data with a default software tolerance or coordinate precision may contribute to mismatches along the edges of maps. A large part of these mismatches can be corrected using applications programs that match node features within a tolerance less than or equal to the map accuracy standard for the series. Applying an automated matching procedure will usually correct errors due to rounding, where edges of base materials truly match. However, such a procedure will not compensate for more severe mismatches of features caused by differences in cartographic interpretation, alignment, or age of source materials used.

Several procedures are available for use in edgematching data for contiguous data themes in GIS applications. The effects of each procedure need to be weighed with respect to the desired result. The following categories of edgematching are available to most users through editing or edgematching routines in GIS software packages:

1. Pervasive rubber sheeting
2. Localized rubber sheeting
3. Node-moving or "snapping"
4. Adding connector segments

The four procedures described herein define a continuum of effects from most to least extensive with respect to the number of and degree to which the original features are shifted from their original position. When adjusting the edge features that are beyond a specified match tolerance, the alteration of whole features that are near or connected to the edge-- whose positions are "correct"--should be minimized. As a general rule, the fewer features moved from their original locations, the better.

Pervasive Rubber sheeting

This procedure is one of the most commonly applied edgematching routines and may be used on one or both adjacent edges. The term "pervasive" implies that effects of adjustment may reach far into the interior of a map. A tolerance can be set whereby any feature--not just edge features--may be adjusted. The process is typically run interactively under supervision and ad hoc correction, but may be run as a batch procedure with default tolerances and control points.

The overall result of this procedure is to produce a cartographically pleasing contiguous match across the map edge by adjusting a large number of internal feature coordinates. In some cases, intersecting features far away from the edge may become separated at nodes. This effect is due to coordinate rounding. Although this separation can be corrected by node-matching programs, it is an indication of how pervasive the effects may be. If effects are restricted to features on one of the adjacent maps, the tolerance required to match the edge may be unacceptably high. If both edges undergo rubber sheeting, then the tolerances to achieve the same match rate may be approximately one-half the single-edge value.

Pervasive rubber sheeting should not be used on data sets for which a high degree of positional accuracy is required unless the tolerances used are well within a selected positional accuracy tolerance. Mismatches that do not fall within this tolerance may be corrected through interactive editing.

Localized Rubber Sheeting

Localized rubber sheeting is a slight modification to the previous procedure wherein the adjustment effects are limited to a user-defined zone along the edge of one or both maps. This process is usually done interactively, allowing the user to review the results of the adjustment as they occur.

Once again, if features on only one map are adjusted to match their neighbor, fewer overall features must be adjusted. A higher tolerance would be required to match features than would be needed if both edges were altered, but with fewer features being adjusted. Again, coordinate rounding may separate some features, requiring adjustment. The procedure should yield a data set in which the zone of alteration was intentionally limited to optimize the continuity and position of features.

Localized rubber sheeting may be applied to data sets used for thematic cartography or GIS applications provided that the zone in which the adjustment occurs is small. Given that no standards are available, a zone of approximately 10 percent or less of the map dimension perpendicular to the edge being matched is suggested for the edit region. On a map whose west edge is being adjusted and whose east-west dimension is 14 inches, a zone of 1.4 inches or less would meet this suggested tolerance.

Node Moving or Snapping

Moving the node features of one map to match an adjacent map, whether through editing techniques or generalized node "snapping" routines, is one of the least drastic techniques. The option may exist to adjust one or both edges in the process, but the effect of the processing results in the movement of nodes only and does not effect the coordinates of other features including vertexes and distant nodes. Two procedures widely used are global node matching using the adjacent data set as a control to which nodes are "snapped" and interactive node moving done in an edit session.

Snapping of the nodes in one map to match the neighboring nodes can be done as a batch procedure with small tolerances to catch many potential match errors. Although it may not catch all of the mismatches, snapping is a good process to run before an interactive review and edit session. In some dense data files, however, it may take as much time to correct the results of an ambitious automated match as to edit the files interactively at the outset.

Interactive node moving allows a user to select and reposition selected nodes, with respect to the adjacent map file, shown graphically on the terminal. The position of adjacent nodes may be referenced by some programs allowing an absolute match between selected nodes. As the edges are edited, other errors may be caught and corrected while still in the edit session.

A combination of node snapping and interactive map editing should be used on GIS applications data sets for which absolute positional accuracy is important. Errors such as mismatch of feature type (one stream line becomes two banks at map edge), missing features, and features that were snapped incorrectly by automated routines also can be evaluated and corrected in an interactive edit session.

Adding Connector Segments

The addition of artificial segments to a map to connect nodes to neighbor nodes is a technique that preserves all the positional accuracy of all features on both map edges. This option may be applied by users who do not have access to programs that move nodes or perform rubber sheeting or who do not wish to change the position of any features. The results may not be as cartographically pleasing but, if done properly, can guarantee topologic continuity across the map boundaries.

For application of cartographic symbology, perhaps based on DLG codes, appropriate attributes must be assigned to these new segments. Users wishing to calculate lengths along streams or roads should be aware that the lengths of features that traverse the edge will be distorted slightly more than with other methods because of the angular nature of the artificial connectors. Alternately, the connector arcs can be flagged so that, although they provide topological continuity, they are neither plotted nor are included in length calculations.

Because of the potential for jagged features along map edges, connector segments should not be added where a "pleasant" cartographic product is desired. The results of this technique may not be desirable for transportation DLG's where abrupt changes in straight roads can be easily seen. This technique should be used only if other techniques are not available or occasionally applied in those cases where other techniques fail.

EDGEMATCHING CONSIDERATIONS

Selecting a Match Tolerance

The process of edgematching causes the positions of features in a digital map to be modified. The distance that features are moved can be controlled in the edgematching process by setting a match tolerance. Features that fall within this distance will be adjusted in the edgematching process. Computer software for edgematching often provides a means of setting a match tolerance.

Map accuracy standards may be used as an initial guide for selecting a suitable match tolerance. For the preparation of the topographic and thematic map series published by the U.S. Geological Survey, a positional standard exists that should be considered in the development of data layers used in GIS applications. For maps published at scales between 1:20,000 and 1:100,000, no more than 10 percent of sampled locations with good horizontal control may be more than 0.02 map inches from their true location to meet National Map Accuracy Standards (U.S. Bureau of the

Budget, 1947). This measure equates to the following dimensions on the ground for each scale (shown to three significant digits):

- 1:24,000 scale = 40.0 feet
- 1:62,500 scale = 104 feet
- 1:100,000 scale = 167 feet

For example, on maps published at 1:100,000 scale, 90 percent of randomly-selected yet well-referenced positions on the ground--such as benchmarks or road intersections--must fall within 167 feet of their true positions in order to meet National Map Accuracy Standards. Other features that are not as well referenced--which translates to most features on a map--do not have explicit accuracy standards because there is no practical means of measuring the error. As a result, it is suggested the National Map Accuracy Standard at a given scale should be used as a maximum match-tolerance for edge features.

National Map Accuracy Standards do not exist for data produced at smaller than 1:100,000 scale. This includes the Army Mapping Service 1:250,000-scale maps and state maps at 1:500,000 scale. Consequently, digital data from these sources may have less spatial resolution than is implied by the standards given above.

For determining a target match tolerance for use at these smaller scales it is suggested that an extrapolation of the National Map Accuracy Standards be used as a guide. At 1:250,000 scale, 0.02 map inches equals 416 feet; at 1:500,000 scale it equals 832 feet. These figures are given as a guide to what match tolerance to select at a given scale below 1:100,000. Some published map bases may actually be more accurate than their scale might imply. Features on other maps, especially older maps with poor horizontal control, may be significantly less accurate than their scale would imply. The user must evaluate the results of applying various tolerances visually to determine the optimal tolerance to use for a given theme at a given scale.

Data layers, such as transportation and hydrography, that are going to be used in routing or network analysis must be managed as a single contiguous data layer for use within a given study area. This may also be an issue for the management of area data if an analyst is concerned with queries of adjacency among polygons that span map boundaries. Users interested in generating a thematic cartographic product that does not need to register with other data (or to be spatially verified), could use a matching tolerance that ignores the positional accuracy implied at the original scale.

Selecting Which Features to Move

Most edgematching software allows users to adjust features along one edge to match an adjacent stationary edge. The advantage of this type of program is that the effect is limited to only one edge, such that about one-half the features are modified. Other software allows users to "split the difference" between the adjacent edges and modify both. The apparent advantage to this type of program is that the effects are distributed across many features and produce visually clean edges.

As adjustments are made to features that are beyond the implicit positional accuracy standards, the following observations apply--sometimes in opposition to one another:

- the more localized the effect the better,
- the fewer features moved from their original location, the better, and
- the simpler (lower dimension) the feature moved, the better.

The simpler the feature moved (nodes as compared with arcs or polygons), the lesser the overall effects of the adjustment. By moving only nodes, only the last node-to-vertex segment location has been modified. If one moves whole arcs, many or all of the vertex locations are modified, and potentially more positions may fall outside the error tolerance.

Regardless of which approach is taken, it is important to document which edges have been adjusted as part of a data-layer history file so that the data product is used correctly in the future. Although one may choose to edit all four edges of selected maps--in a "checkerboard" pattern--it is simpler to apply a systematic rule, such as adjusting the north and west edges of the 1:100,000-scale maps to match their east and south counterparts.

PREPARATION OF THE NORTHWEST RIVERS DATA BASE

The 1:100,000-scale digital hydrography data layer was used as the spatial base for the Northwest Rivers Data Base. The NMD (National Mapping Division) of the U.S. Geological Survey had compiled the 1:100,000-scale hydrography and transportation series to satisfy the need of the Census Bureau for nationwide coverage for the 1990 Census. NMD prepared the 1:100,000-scale base from the best available materials, relying on the existing 1:24,000- and 1:62,500-scale topographic bases. The features kept in the 1:100,000-scale base were selected by length and relative importance to produce a less-dense cartographic product. In some cases, the features along adjacent edges were compared and matched by those who compiled the 1:100,000 scale base, subject to the availability of the completed 1:100,000-scale base sheets.

The 1:100,000-scale base was digitized using a high-resolution raster scanner. The results were edited and converted to a vector file for tagging and sale as a DLG. A DLG may contain reference to spatial features such as points, lines, and areas, and maintains an associated series of cartographic feature codes. These codes describe the types of water features in the DLG so that one can generate appropriate symbology to create a map using mapping software. Hydrologic users of the data sets should be aware that the DLG codes may not necessarily describe features that a hydrologist or water resource manager might expect, and additional attributes may be necessary for water-resources applications. This was a primary motive in adding the EPA River Reach attributes to the 1:100,000-scale base in the BPA project.

With respect to edgematching, it has been the policy of the NMD to add flags (codes in the digital file) to their DLG products that describe which edges have been matched and the potential reasons for non-alignment (U.S. Geological Survey, 1985). Thus far, very few 1:100,000-scale DLG data sets have been processed through these edge verification routines. If mismatches are encountered, the flag is set as to whether the errors present are due to non-alignment of source material and (or) to discontinuity in attributes. No attempt has yet been made to correct the apparent "errors" along the edges of the digital materials. A philosophical dilemma exists--if both map manuscripts are correct, and prepared to National Map Accuracy Standards, which map should be adjusted? One way to correct the series would be to recompile the base materials. Although a revision schedule is planned for some of the 1:100,000-scale maps, it will take many years to complete. For users of digital cartographic data who have an immediate need to append these data for use in GIS applications, the philosophical issues are overridden by the desire to prepare a continuous data layer using appropriate techniques.

Problems Encountered Along Edges

In the course of processing the hydrography data in the Pacific Northwest the following situations were observed with respect to edge features:

1. Features have the same DLG code combinations, and are within a predetermined processing (match) tolerance of one another.

2. Features do not have the same DLG code combinations, but are within a predetermined match tolerance of one another. An example would be a feature described as a canal on one map and as a stream on its neighbor.

3. Features have the same DLG code combinations, but are in excess of a predetermined processing tolerance of one another. For data at 1:100,000 scale, an acceptable match tolerance would be less than 167 feet; in some cases, gaps will exceed this distance.

4. Feature is represented differently on adjacent map sheets. A good example is where a stream shown as a single line becomes a double-line stream upon crossing the map edge.

5. Feature is present on one edge but absent on the adjacent sheet.

The first situation is the ideal edge-match condition and usually can be remedied using software that matches node coordinates. The second example is topologically the same condition as the first. The user may wish to process the joined coverage later and keep nodes that signify a change in arc attributes. Situations 3 and 4 are the most common and challenging, but can be remedied using a combination of supervised editing and automated matching processes to be described in detail later. The fifth situation requires no intervention but indicates a potential discontinuity in the hydrography data set.

Techniques used in the Northwest Rivers Data Base

Data in the Northwest Rivers Data Base were edge matched using a combination of automated node matching and interactive, on-screen editing of individual node to vertex segments. The default edgematching tolerance used in automated node-matching routines was 131 feet--on the conservative side of the 167 feet guideline. This was applied to the internal edges of the (up to 32) DLG files that compose a 1:100,000-scale map sheet and to the external edges of the 1:100,000-scale files. External edges were then reviewed and edited using interactive editing techniques which provided the user the opportunity to assure the proper association of features. Where possible, interactive editing was limited to node adjustment or adding connector arcs to complete the continuity of the data across the map edge. Where connector segments were added, their Minor DLG code was calculated to 999 to signify an artificial arc. Connector arcs were frequently added where a hydrographic feature was discontinuous (broken stream segment), paralleled or straddled the edge of a map sheet, or changed from a single-line to a double-line feature.

SUMMARY AND CONCLUSIONS

Although it is desirable to maintain spatial digital data to a set of scale-sensitive positional map accuracy standards, the realities of digital-spatial-data processing often require the user to make some modifications. Where possible, techniques should be applied that minimize the changes with respect to the number, extent, and dimension of features.

Of the techniques available, one that allows for the supervised movement of nodes is the most desirable--along one or both edges. Techniques that employ a highly constrained rubber sheeting may be appropriate for some applications, but the zone of influence and the match tolerance must be carefully and consistently defined. After any of the procedures--whether supervised or unsupervised--it is suggested that the results be reviewed in an interactive edit session and changes made based on base data and professional judgment.

To improve communication about the quality of the data, regardless of the technique used, it is important that a user document (1) the technique used as described in this or other techniques papers, (2) whether one or both edges were altered, (3) tolerance(s) used, and (4) any noteworthy exceptions to the rule encountered in the matching process. Such information may be preserved with each digital data set in a data-history file along with resolution, source, automation purpose, projection, and other basic data.

REFERENCES CITED

Bonneville Power Authority, 1985, Geographic Information System Task Force Report and Recommendations, unpublished report, 18 p. [Report on file at Bonneville Power Administration, Portland, Oregon.]

U.S. Bureau of the Budget, 1947, United States national map accuracy standards, revised June 17, 1947.

U.S. Geological Survey, 1985, Digital line graphs from 1:100,000-scale maps: U.S. GeoData Data Users Guide 2, 74 p.

APPENDIX A

Suggested Steps for Edgematching 1:100,000-scale Hydrography Data

1. Acquire and load the various map files that compose a 1:100,000-scale map sheet into the GIS software system. The number of files for a given theme will range between 4 and 32 panels per 1:100,000-scale map sheet.

2. Associate the feature attributes for lines and polygons with the features in the attribute-management data base for each panel. One may optionally build attributes for point and node features, too.

3. Based on the DLG attributes, remove neatline arcs from each panel. These may be placed into a separate map file for later use, if the user desires.

4. Append the panels for a given theme into a single, contiguous digital map file, maintaining both line and polygon attributes. Node and point features may also be preserved and merged at this step.

5. Run Steps 1 through 4 on all adjacent map sheets to be used in the project area.

6. Evaluate map projection. DLG data are maintained in the Universal Transverse Mercator (UTM) projection. It is possible that the study area will span a UTM Zone boundary if it has an east-west extent of hundreds of miles. If this is the case, one should select a projection coordinate system that covers the entire study area and optimizes those map characteristics that are required by the study (such as correct area, length, or direction). Most state governments reference spatial data to a State Plane coordinate system--which uses a specific combination of projection type, coordinate units, x- or y-shift, and other projection parameters. Studies that span state boundaries should not use the State Plane coordinate system. Studies of national scope tend to use the Albers Equal Area projection. The Albers projection was used for the Northwest Rivers Study to provide a contiguous and highly accurate spatial reference.

7. If necessary, project map data from UTM into selected projection system.

8. Apply a node match tolerance of between 131 and 164 feet to match line features internal to the 1:100,000-scale map sheet. This should be all that is required to connect internal features.

9. Establish an edge of a map sheet to be "snapped" to, and snap the current sheet to it using the same node match tolerance above (131-164 feet). As a rule, the Northwest Rivers Study held the east and south edges constant, while the north and west edges or adjacent maps were modified.

10. Review the results of the snapping in an interactive editor. Make changes as needed, code DLG Minor 1 code to 999 for connector arcs added.

GPO 792-108/23017