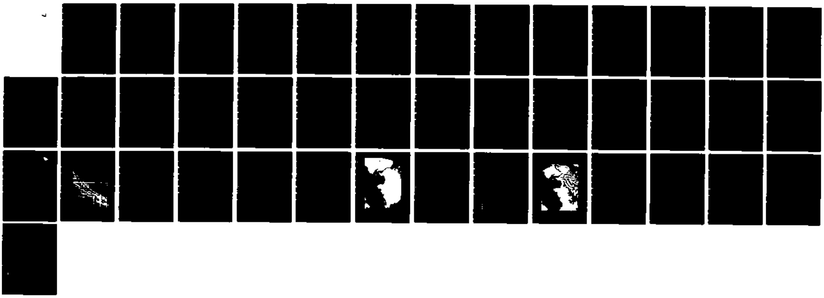
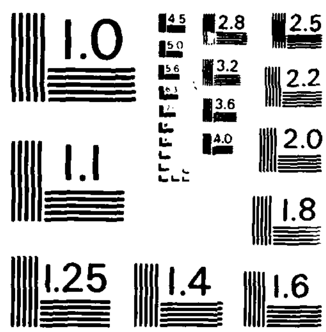
AD-A142 388 DATA INTEGRITY FACTORS AFFECTING THE CONSTRUCTION OF THE MAPPING CHARTING AND GEODESY DATA BASE(U) JET PROPULSION LAB PASADENA CA S Z FRIEDMAN 01 JUL 83 1/1

UNCLASSIFIED JPL-D-879 ETL-0357 F/G 8/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ETL-0357

AD-A142 388

Army Technology Development

Data Integrity Factors Affecting the Construction of the Mapping, Charting, and Geodesy Data Base

S.Z. Friedman

1 July 1983

A

Approved for public release; distribution is unlimited.

Prepared for

U.S. Army Engineer Topographic Laboratories
Information Sciences Division
Fort Belvoir, Virginia

Through an agreement with

National Aeronautics and Space Administration

Prepared by

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL D-879

84 06 26 040

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ETL-0357	2. GOVT ACCESSION NO. AD-A142 388	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DATA INTEGRITY FACTORS AFFECTING THE CONSTRUCTION OF THE MAPPING, CHARTING, AND GEODESY DATA BASE		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 6. PERFORMING ORG. REPORT NUMBER JPL D-879
7. AUTHOR(s) S.Z. Friedman		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Jet Propulsion Laboratory California Institute of Technology Pasadena, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, Virginia 22060		12. REPORT DATE July 1983 13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) National Aeronautics and Space Administration		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Numerous data integrity problems were encountered while processing the digitized map data used for the research task, "An Image Based Approach to Mapping, Charting, and Geodesy" (JPL Task No. RD-182, Amendment 125). Increased interest in those problems has prompted this separate report. Specific data integrity problems are analyzed and methods for their solution are described. Although many of the problems described herein are project specific, this document contains information which could be useful to data base specialists who are concerned with data integrity problems associated with building large cartographic data bases.

Army Technology Development

DATA INTEGRITY FACTORS AFFECTING THE CONSTRUCTION OF THE MAPPING, CHARTING, AND GEODESY DATA BASE

1 July 1983

APPROVED BY:

S. FRIEDMAN
TASK MANAGER

APPROVED BY:

Y. NAKAMURA
PROGRAM MANAGER

APPROVED BY:

E. K. CASANI
DIVISION MANAGER

A-1

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ABSTRACT

Numerous data integrity problems were encountered while processing the digitized map data used for the research task, An Image Based Approach to Mapping, Charting, and Geodesy (JPL Task No. RD-182, Amendment 125). Those problems were not fully addressed within an earlier document (JPL Internal Document 715-153, 1982) describing the Mapping, Charting, and Geodesy (MC&G) project. Increased interest in those problems has prompted the completion of this separate report. In this document, specific data integrity problems are analyzed and methods for their solution are described. Although many of the problems described herein are project specific, this document contains information which could be useful to data base specialists who are concerned with data integrity problems associated with building large cartographic data bases.

ACRONYMS

GIS	Geographic Information System
IBIS	Image Base Information System
IPL	Image Processing Laboratory
JPL	Jet Propulsion Laboratory
MC&G	Mapping, Charting, and Geodesy
pixel	picture element (The finest resolvable element in a digital image)
VICAR	Video Image Communication and Retrieval

CONTENTS

INTRODUCTION

1.0

The Image Processing Laboratory (IPL) of the Jet Propulsion Laboratory (JPL) has been involved in research demonstrating the utility of the Image Based Information System (IBIS) for Mapping, Charting, and Geodesy (MC&G) applications. The research has been funded by the U. S. Army Engineer Topographic Laboratories (USAETL). A document entitled, An Image Based Approach to Mapping Charting and Geodesy (Friedman, 1982), detailed the construction and application of an IBIS data base for MC&G applications. Within the body of that document, numerous problems associated with data capture (the process of entering the data into the data base) were alluded to, but were not described in detail. After publication of that document, the USAETL expressed further interest in data capture problems. This report describes those data capture problems in greater detail. USAETL's motivation is to learn what steps can be taken to ensure the integrity of cartographic data bases. If errors exist in the source materials and remain unedited when they are entered in to the data base, they will most surely affect the results of any analysis derived from the data base.

Most data capture problems were associated with errors in digitizing, although other specialized problems relating to data processing occurred as well. Common digitizing problems included improperly digitized line segments, omitted line segments, mislabeled centroids, and omitted centroids. When most of the digitizing problems were identified, the digitizer operator (a vendor in this case) was requested to make the needed corrections. A few remaining problems which could be easily solved with VICAR (Seidman and Smith, 1979) image processing software were corrected at the IPL. A record of those problems and their solutions was maintained (Table 1) and is the basis for this report.

The purpose of this report is to exemplify common data processing problems associated with data capture and data base construction. Hopefully, it will useful in assessing time frames needed for data capture in future Geographic Information System (GIS) projects. Although vendor-caused problems are discussed in detail in this report, there is no intent to harm the reputation of the vendor. Those problems exemplify common problems which occur frequently with data capture in general, and not confined to the production quality of the specific vendor who digitized data for the MC&G project.

1.1 General Considerations About Data Capture

Perhaps the most important and frequently the most complex and confusing aspect of data base construction is data capture. In order to ensure successful retrieval of information from any data base, the proper data must be selected and entered into the data base in such a manner that the information content of the original data is retained. If an important aspect of source material was not entered, or subsequent data processing leads to alteration or obliteration of the information, it will be unavailable for analysis and the data base may be considered incomplete. Consequently queries to the data base will be inconclusive or totally in error.

For Geographic Information Systems, much information is found in map form. Since map data is available at numerous scales and levels of detail, potential input materials must be screened to ensure that they contain sufficient detail, yet exclude superfluous information. With raster based GIS, such as the Image Based Information System (IBIS), the internal representation of data is by pixel (grid cell). The relative scale of these pixels must be carefully selected to retain exactly the proper level of detail needed in subsequent data base operations. If the scale is too coarse, needed detail will be lost, being too small for representation at the selected cell size. However, if the internal scale is too fine, all needed information will be retained, but data processing costs will become excessive. Consequently, it is essential to maintain the proper balance between too much and too little detail.

Electronic coordinate digitization, the process of tracing selected map elements such as points and line segments to obtain relative Cartesian positions of those features, is a common mode of data capture. As the cursor is traced along selected map features, specialized hardware records every movement of the cursor at a specified level of detail. What appears to be a very simple process: tracing lines, enumerating points, and adding labels to notable features, is frequently beset with numerous errors and is actually very time consuming. Editing digitized data is no simple matter, involving repetitious checking and rechecking of all recorded information. Yet the need to obtain error free data is so important, that editing will often require more time expenditure than the initial digitizing phase of production.

2.0 TYPICAL DIGITIZING PROBLEMS AND THEIR SOLUTIONS

2.1 Digitizing and Preprocessing for the MC&G Project

Source maps for the MC&G task were obtained from USAETL on three stable-base film transparencies (Figure 1). Select thematic information appearing on those maps were identified for inclusion in the data base as the maps were prepared for data capture via electronic coordinate digitization. The IPL does not have a sophisticated coordinate digitizer system with modern features such as real time processing, screen display, and interactive editing. These features make the digitizing process easier. Consequently, a vendor was selected to digitize the data. The decision to utilize the services of a vendor was made with the intent of saving time and maintaining a higher quality product than could be produced with the digitizer at JPL. However, as frequently occurs when time saving measures and/or new technologies are used, the vendor digitized data did not live up to expectations. Data format problems had to be overcome, errors in digitizing had to be corrected, and coordinate scale factors had to be adjusted. It was expected that all digitizing would be completed within one week. However, nearly six weeks elapsed before the vendor could provide data in acceptable form. Still after that time, the data required further editing. An additional eight weeks elapsed until all errors were found and corrected.

Acknowledging that data formatting problems frequently occur when transferring data from one computer system to another, specific data transfer formats were provided to the vendor more than one week before the source maps were to be digitized. The vendor was to provide data in a format compatible with the VICAR image processing system in use at JPL. Furthermore, these data files were to be in a format which could be directly transposed into a graphics data standard used by IBIS. The first digitized data tape was received two weeks after the vendor was supplied with the source maps. When the tape was processed, it was determined that the vendor had not formatted the data properly as the tape could not be read successfully by the software. Both the line segment and centroid files were found to be in error. A second tape was received two days later. From visual analysis of a data listing produced by software, the format for vector data appeared correct. But when processed to produce a test image (Figure 2), the source data was still found to be in error. Subsequent analysis of the data files indicated that alternating data records were improperly formatted. Additionally it was found that centroid files on the same tape were coded as ASCII characters, instead of the prescribed EBCDIC character code. Eight days later, a third data tape was prepared by the vendor. That tape also had similar format errors.

2.2 Editing and Verification

More than four weeks after the start of the digitizing process and eleven days after receipt of the third data tape, a fourth data tape was received. That tape was properly formated and all data files were successfully converted to an IBIS graphics data format. Line segment files from the source maps were transformed from vector to raster representation, and the first verification images were produced (Figure 3). During visual inspection of the imagery, a minor formatting problem was found with some line segments of the land use file. Additionally, several digitizing mistakes were also noted during the visual inspection (Figure 4). A total of 31 errors were identified from the land use file. Only two errors were found on contour image, as two small contours were omitted (Figure 5). Both the land use revision and the hundred year flood plain data sets were correctly digitized. Of course, those two maps were very simple in structure, containing few line segments. The contour data set, which required minimal editing, was to be edited at the IPL. Errors in the land use data set were found to be too extensive to edit at JPL, and the vendor was asked to perform the needed corrections.

A fifth and final data tape containing corrected land use line segments was received from the vendor nearly one month after the vendor was notified of the needed corrections. After vector to raster conversion, the data was found to be correctly digitized except for one small error, a line segment drop out, which was added at JPL (Figure 6).

Once the images that were created from the digitized data appeared to be free of digitizing errors, geographic regions (polygonal features) were checked for closure. An IBIS PAINT image was used for that purpose. Both the hundred year floodplain and the land use revision data planes had no closure problems. However, closure problems were detected for both the land use and contour (Figure 7) data sets. The source of most of those problems was determined to be improper specification for the termination of the right hand edge of the image by three pixel units. The problem was easily corrected by truncating that edge by three pixels. Still, some line segments had to be extended to meet the edge of the image to ensure proper closure.

As previously mentioned, the contour data set was edited at JPL. The two missing contours appeared to be successfully added to the contour data set using an interactive editor contained as a module in the VICAR system. However, when a new test image was produced some line segments appearing on the original contour image and unassociated with the original problem had disappeared, although the two added polygons were properly positioned (Figure 8). A software error with the editor program was found and corrected, and an edited file was produced without error (Figures 9 and 10). Total elapsed time to produce the corrected contour file was one week. Had the VICAR software been working properly, editing would have required less than one day.

More than two and one half months had passed (72 days) between beginning and ending the digitizing process. Although three weeks could be attributed to problems associated with implementing and utilizing a new technology, seven weeks could be directly attributed to the time involved in producing accurate representations of the source maps. The initial digitizing was completed within two weeks. However, it took five weeks to edit the digitized data files before they were correct, and still not in every detail.

2.3 Region Identification

Once a digitized file representing line segments on a map has been edited and verified for accuracy, it is converted to image format a final time. That image is subsequently utilized in various capacities as a data plane within an IBIS data base. It is frequently used as a thematic overlay to provide spatial reference to the result of a data base query. In a similar manner, it can be employed to provide boundary details for choroplethic maps produced from an IBIS data base. By far the most important application of a line segment image plane is for creating a raster region file (e.g., a region-coded data plane or a PAINT image). In that process all geographic regions comprising a data plane, as defined by the structure of line segments contained in a line segment image plane, are encoded with unique gray values. When the process is completed, each region within the data plane has been assigned a unique numerical label to identify it from all other regions comprising the data plane.

For the MC&G project, the raster region identification process worked quite effectively. Afterwhich, a four color mapping algorithm (IBIS program COLOR) was implemented to create images which could be used to verify that all region closure properties were properly maintained. Once closure was verified, the next processing step, centroid matching was begun.

2.4 Label Assignment

By itself, the numerical label assigned to each region by the PAINTing process is limited in utility. These is no implicit association or thematic identity to link the region with the actual feature it represents. IBIS program CTRMATCH produces the needed link by matching digitized labels (centroids) based on their position on a Cartesian reference plane with gray-tone regions comprising the region coded image plane. CTRMATCH generates a label-region pointer file and an error summary. The label-region pointer file is referred to as an interface file because the file contains the needed link, or interface, between a region coded data plane and a tabular file which contains descriptive data about regions comprising the data plane.

Most types of errors associated with digitizing of centroids are either errors of commission or errors of omission on the part of the digitizer operator. Most of those errors are identified during the centroid assignment process. The most frequently occurring errors are: (1) the identification of regions which have not been assigned a centroid, (2) regions which have been assigned multiple centroids, and (3) regions which have been improperly labeled. The first two types of errors are reported by CTRMATCH, but labeling errors can only be identified through combined detailed analysis of (1) the region coded image, (2) the label-region pointer file, and (3) the original base map used as a source for digitization. Essentially that process is straightforward, but it is quite laborious and time consuming. All three products must also be referenced when solving problems associated with omitted centroids or multiply labeled regions, but the investigation area can be limited to the specific region(s) in question and some immediately adjacent regions which were correctly labeled. Frequently region labeling problems occur in pairs. One region without a centroid assigned to it and one with two centroids assigned to it will be found adjacent to each other. Identification and correction of centroid errors can be a time consuming activity. However, in most cases the number of errors are relatively few, are easily corrected with IBIS software, and seldom involve redigitization of the centroid file.

For the MC&G project, centroid matches, were checked for validity interactively, using one of the IPL's raster display devices. Since every region of each region-coded data plane was to be checked for proper label assignment, the interactive method was selected as a time savings option. However, when correcting label omissions and multiple encoded regions as noted by CTRMATCH, a batch method is often utilized instead of the interactive procedure.

The contour centroid data set was analyzed first. Six regions were multiply encoded, while twenty-one regions had no labels assigned. For the land use data plane, eighteen regions contained multiple labels, while nineteen were without centroids. Even the relatively simple land use revision data plane which contained only four regions had one missing label. Only the 100 year flood plain was found to be error free. Although most of the errors were associated with the digitizing process, historically it has been found to be easier to modify the interface file produced by CTRMATCH than to correct those errors through redigitization. Consequently, the vendor was not required to perform any needed corrections to centroid files.

Identification and correction of centroid label errors were easily completed for the MC&G project. All label errors were found and corrected within one week, a far shorter period than that required for editing the line segment files. In addition to correcting errors noted by program CTRMATCH, all polygons were manually checked to ensure that all polygon label assignments were made correctly. Probably half of the time period (3.5 days) was required for that task. But when considering the importance of strict quality control, the time should always be spent when possible.

3.0 THE UNEXPECTED SCALING PROBLEM

3.1 Identification of the Problem

The identification and correction of digitizing problems is a time consuming yet straightforward process. Seldom are the errors so complex or excessive that the original material must be discarded and redigitized. But sometimes inexplicable errors which were deeply rooted in the digitization process might not surface until the data base, including all associated data planes, has been created and is being tested. Although completely unexpected, this type of situation occurred with the MC&G data base.

The problem was encountered after all obvious normal problems were identified and solved. The line segment images were correct, the region identification process completed, and centroid labels were assigned to all regions without error. A master interface file and a composite feature georeference plane were produced and the data base was prepared for query.

The first exercise with the completed data base was to report areal coverage for all land use regions. The computation procedure was completed successfully, but all areal calculations appeared to be underestimated when compared to results obtained by another researcher (Sharpley, et al, 1978) doing the same calculations with their MC&G data base. What made this error very strange was the fact that all regions were underestimated by a standard and very constant error, about 10 percent. Without reference to earlier results by Sharpley, it is doubtful that this error would have been identified very easily. The only possible method for identification would have been through hand-tabulation of areas on the base maps with the aid of a planimeter. That would have been a very time consuming and laborious task. Since this type of error was not expected, the problem may have never been identified if the comparison to the results by Sharpley were not made.

In most cases, areal computation errors can be directly attributed to inaccurate specification of the relative scale of the raster grid cells comprising the data base. However, in the case of the MC&G project, the pixel scale appeared to be accurately calculated. The scale of the data base was originally determined through measurement of several selected line segments on a vendor supplied test plot (Figure 11) which registered exactly to a USGS topographic map of known scale and accuracy. After the areal calculation problem was encountered, the pixel scale was recomputed to ensure that the error did not lie there. Again the scale was determined to be one pixel per 400 square feet (20 x 20 feet). No apparent error was evident in that procedure.

Then, the coordinates that were used to transform the Cartesian-based digitizer data to the image-based coordinate system were analyzed. The transformation appeared to be properly specified as well. A final check was considered: Attempt to overlay the data plane already converted to image format with the test plot produced by the vendor. Since no same-scale copies were available, a Zoom-Transfer Scope was utilized to compare the two maps. The problem was identified at last. The two maps could not be registered at all. When compared to the test plot, the raster map appeared to be expanded along the y-axis (Figure 12). Measurement of the expansion indicated that coordinates were overscaled by ten percent along the y-axis, which was the difference between the expected and measured areal coverage of land use regions.

This type of problem, differential scaling along the two axes, had never been encountered before. Test plots and digitized data provided by the vendor always conformed to each other. Two solutions were possible: (1) to have the vendor rescale the data, or (2) to rescale the data using IBIS software. It was believed that the vendor would take too long to solve the problem. Furthermore, the vendor could not explain the cause of the problem when confronted with the issue. Consequently the digitized data were rescaled using IBIS software.

3.2 Reconstruction of the Data Base

Correction of the scaling error required a complete reconstruction of the data base. An additional coordinate transformation using IBIS program POLYREG was required to convert the digitized data to correctly scaled image-based coordinates. After rescaling, the data was again converted to raster form and region identification was performed on all four data planes. One closure problem which had not occurred previously was found while visually analyzing the new land use image (Figure 13). This problem was corrected by adding a small line segment to close the gap.

After producing clean region-coded data planes, the centroid matching process was repeated. The centroid files were rechecked to ensure that new centroid errors were not caused by the scale change. In most cases, the same errors were encountered as before. However, while analyzing the contour data plane, two regions were found to be mislabeled. (The fact that these two label errors were not identified during the first verification pass is good justification for a multistage quality control process to ensure data integrity when building cartographic data bases.) A total of twenty-seven corrections solved all data base problems.

Twenty-one days elapsed between identification of the scaling problem and complete recreation of the data base. During that period, the new scale of transformation was computed, all digitized data were converted to image form, region coded data sets were created and verified for closure, centroid labels were introduced and checked for proper assignment and edited as needed. The complete data base was reconstructed. In its entirety, it included four line segment data planes, four raster region coded image planes, four associated interface files, a composite feature georeference base, and a master interface file.

Areal calculations were performed with the newly created data base, and accurate results were derived. A query procedure was developed and tested, and all basic goals of the MC&G were completed.

4.0 SUMMARY AND RECOMMENDATIONS

Quality assurance of digitized cartographic data which will be entered into a GIS or other data base is a very important issue. Even with a very small digitizing job as exemplified by the MC&G test data set, numerous errors can, and will, frequently occur. Some of the digitizing errors encountered during the processing of data for the MC&G project could be attributed to utilization of a new technology. The initial data formatting errors are an example. Other problems such as omitted or incorrectly digitized line segments and centroids can only be attributed to carelessness and the monotony of the digitizing process. The act of digitizing is basically very simple and is invariably a very boring process. Those types of errors can be expected to occur frequently. In order to ensure that all such errors are removed before the data is entered into a data base, strict quality control measures must be adhered to. The scaling problem was related to a novel hardware and software situation. The vendor who was contracted to digitize the map data was using a new digitizer system which was installed just prior to beginning the MC&G digitizing job. Perhaps the vendor did not fully understand the subtle features of that new system. Although problems like the scale error are not expected to occur frequently, quality assurance methods must be developed to identify these types of errors as well.

The IBIS MC&G demonstration project involved the construction of a very small data base. Many data verification steps were tried, tested, and proved on a one-time basis only. The establishment of verification procedures which can be implemented with large cartographic data base production projects was not an objective of the MC&G project. However, as a result of the project, it is evident that a well defined and preplanned verification and quality assurance process must be developed to ensure data integrity. Specific tasks and their order of operation should be defined before digitizing specifications are developed and the digitizing is begun. As with all large volume data capture projects, prototyping the entire process will be important to ensure the proper procedures for quality assurance have been developed.

Assuring the integrity of digitized data which will be used in the construction of large cartographic data bases is an extremely important issue. If quality assurance of input data can not be maintained, data bases which utilize that data will not be useful. The verification process will always be labor intensive and time consuming. All digitized products must be carefully compared to their source materials to ensure that information has not been inadvertantly deleted, distorted, or introduced. At least one person, and probably several people, will be involved in verification. The period of verification will most likely encompass from two-to-four times the time period elapsed during digitizing. For the MC&G project, the actual digitizing of the map data was completed within ten days. However, six weeks elapsed before the vendor could provide a usable product. Even then, eight additional weeks passed before a completely corrected and verifiable product was produced. Similar conditions are to be expected with any project involving digitizing of cartographic data.

Table 1

DATA CAPTURE CHRONOLOGY

USAETL Mapping, Charting, and Geodesy

DATE	EVENT
80.10.17	Digitizing plans formulated - file formats and data specifications made - decision to use vendor for digitizing (JPL cannot digitize that quantity of data efficiently)
80.10.20	Potential vendor selected, requested cost estimate
80.10.22	Cost estimate received
80.10.31	Purchase order sent to vendor - requested that vendor be ready to start digitizing in early December 1980
80.11.10	MC&G implementation plan submitted to ETL
80.11.12	Base maps (mylars) for digitizing received from ETL
80.11.20	Film transparencies containing map data photo-copied by JPL photo-lab
..**	START OF DIGITIZING
80.11.25	Film transparencies containing map data submitted to vendor - precise data formatting specifications provided at this time so vendor can produce a VICAR compatible tape - begin coding VICAR tape processor program VMXLOG
80.12.08	First vendor-produced digitizer tape received: unacceptable, as tape was completely unreadable - tape format errors, including improper character representation (ASCII instead of EBCDIC) and improper file structure - vendor also combined floodplain and land use revision data sets instead of creating two separate files
80.12.10	Second data tape received from vendor: also contained numerous errors - internal file structure found to be in error: alternating data records were improperly formatted - character conversions from ASCII to EBCDIC wrong for certain alphabetic characters - general structure of centroid files in error

TABLE 1. DATA CAPTURE CHRONOLOGY

DATE	EVENT
80.12.15	Completed coding of VMXLOG
80.12.18	Third data tape received from vendor - contained EBCDIC coded label files only
80.12.29	Fourth data tape received from vendor - format appears good and has been successfully processed by VMXLOG
..**	BEGIN DIGITIZING VERIFICATION PROCESS
81.01.05	First vector to raster conversion (POLYSCRB) and first region identification test (PAINT). - general feature representation and spatial alignment between data planes appears good - results from PAINT run indicate that line segments do not reach prescribed right and bottom border of image *** must truncate image by three pixels
81.01.08	Visual analysis of image files for digitizing errors - land use: a total of 31 errors noted (includes both digitizing mistakes and edge problems) - contour: 2 omitted features and edge problems - floodplain and land use revisions are OK
81.01.12	Interactive edit of contour data plane using VICAR software to add two omitted features - failed due to software problems
81.01.19	Interactive edit of contour data plane successful
81.02.04	Fifth and final data tape received from vendor. After conversion of the data from vector format to raster format, the vendor product was considered to be acceptable
..**	END DIGITIZING (elasped time: 72 days)
81.02.11	Successful PAINT of land use data plane - verification of digitizing
81.02.12	All PAINTs of thematic and composite image planes are completed successfully - interactive identification of all region coded gray values noted for all four thematic data planes
..**	END DIGITIZING VERIFICATION PROCESS (elapsed tin · 39 days)

TABLE 1. DATA CAPTURE CHRONOLOGY

DATE	EVENT
..**	BEGIN REGION IDENTIFICATION/VERIFICATION
81.02.18	Centroid match (CTRMATCH) on three data planes - contour: 6 regions mapped with multiple centroids 21 regions without centroids - landuse revision: 1 region without a centroid - floodplain: no errors
81.02.19	CTRMATCH of land use data plane - 18 regions mapped with multiple centroids - 19 regions not mapped
..**	BEGIN DATA BASE CREATION
81.02.20	Begin setup of georeference base and definition of thematic keys to be used in data base query
81.02.23	Correction of contour and land use revision centroid errors completed and verified
81.02.24	Correction of land use centroid errors completed and verified
..**	END REGION IDENTIFICATION/VERIFICATION (elapsed time 7 days)
81.02.24	Georeference base assembled
..**	END DATA BASE CREATION (elapsed time 5 days)
81.03.06	Decision to improve registration of land use revision data plane to data base - 39 tiepoints selected interactively - 2 iterations needed to perfect fit
81.03.06 to 81.03.31	Data base processed to obtain preliminary statistics pertaining to areal coverage of thematic features for all data planes. (Note: This time period was not directly related to data capture and associated problems. But as a result of data processing during that period, a major problem with the data base was found).
..**	IDENTIFICATION OF SCALING PROBLEM
81.04.07	Verification of areal computations for all regions indicated a standard and constant error

TABLE 1. DATA CAPTURE CHRONOLOGY

DATE	EVENT
81.04.04	Problem linked to improper scaling of digitized data by vendor - test plot provided by vendor was at scale, but when images were overlayed, they did not register - Y dimension of data was overscaled by approximately 10% - would take too long for vendor to correct, use POLYREG instead
81.04.10	New scale images made - 1 error in land use plane caused by rescaling
81.04.13	Region coding and region identification process completed - PAINT ok, CTRMATCH needs verification
81.04.14	Error with land use data set found and corrected
81.04.22	Recheck contour data sets for proper centroid match - 5 regions mapped with multiple centroids - 16 regions without centroids - 2 regions mislabeled! *** corrected same day
81.04.23	Verification of floodplain and land use revision centroid matches - floodplain ok - land use revision: one error noted and corrected
81.04.24	Verification of land use data plane centroid match - 19 regions mapped with multiple centroids - 14 regions not mapped - 6 regions mislabeled! *** 27 total corrections solve the problems
81.04.27	For all four data sets, centroid matches verified correct, and georeference plane created - first tabulations were derived from the new data base
..**	END PROBLEM (elasped time:21 days)
..**	END DATA CAPTURE AND DATA BASE PREPARATION

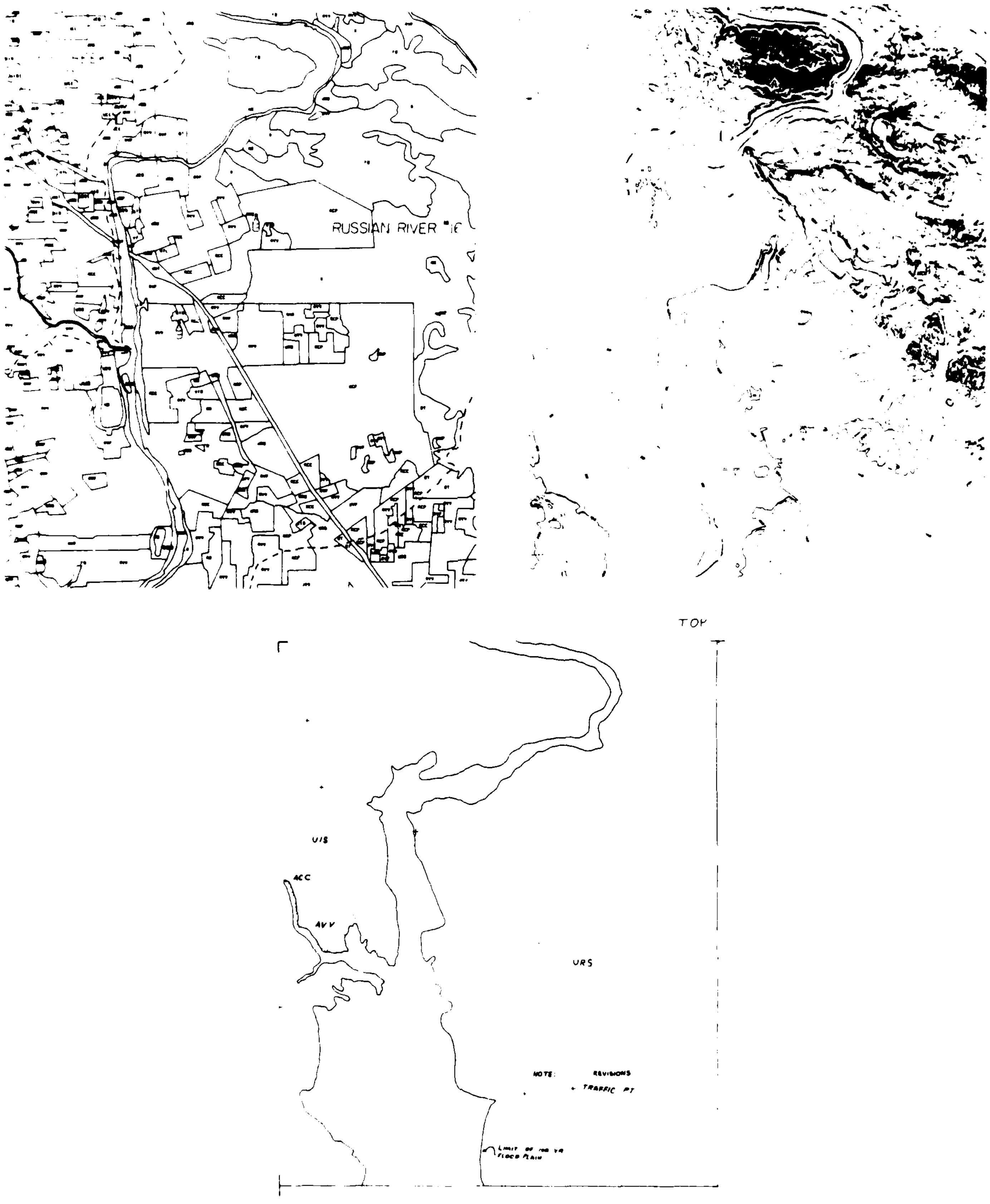


Figure 1: Source maps used for the MC&G project. (Note: These maps have been photographically reduced for depiction in this report.)

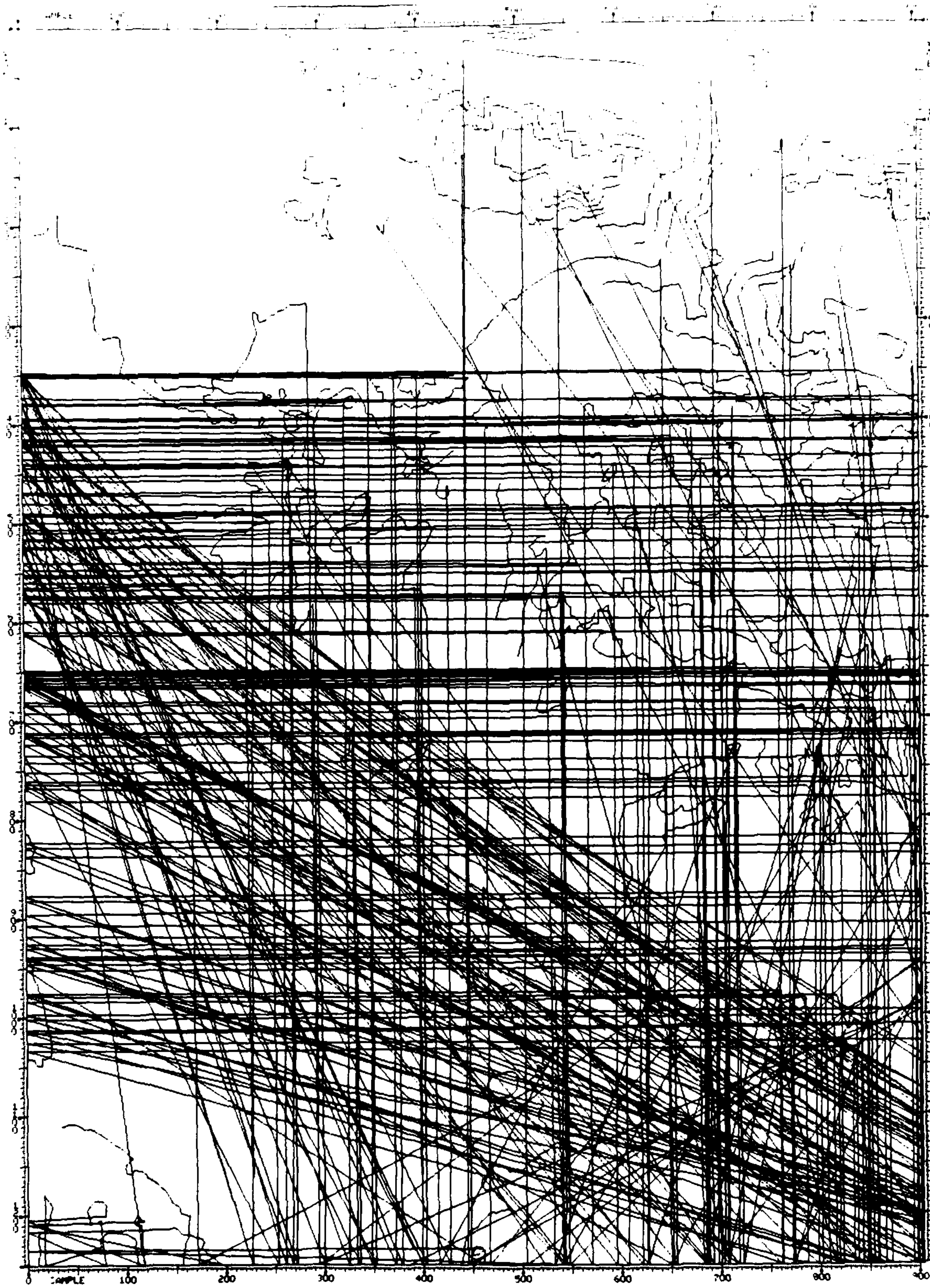


Figure 2: Data formatting problems were quite evident with conversion of the vector data to raster form. In this case, part of the topographic data were converted properly, but every other data record was improperly formatted, causing lines to go astray.

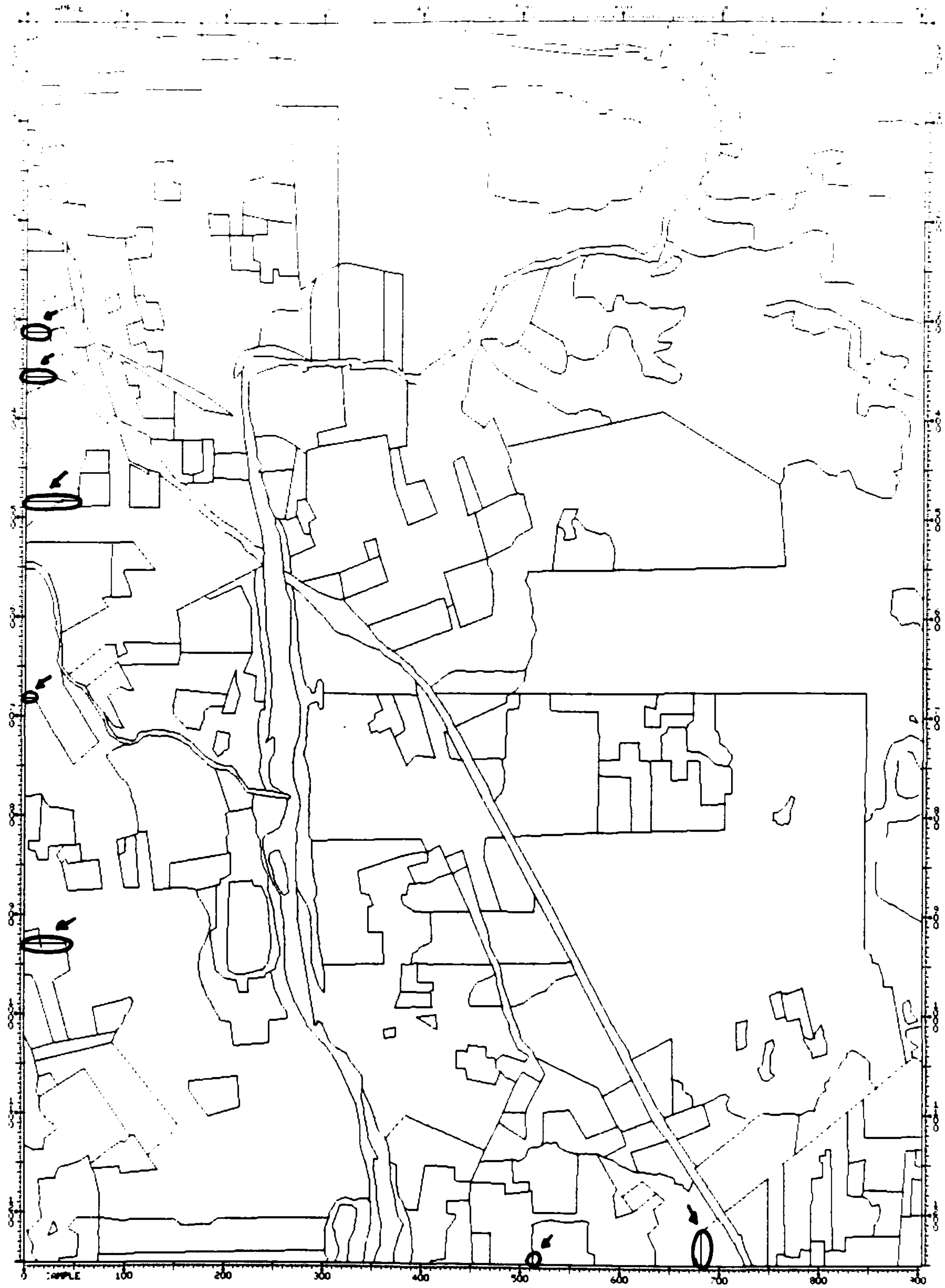


Figure 3: A minor formatting problem was still evident with some line segments of the land use data file after processing the fourth data tape.

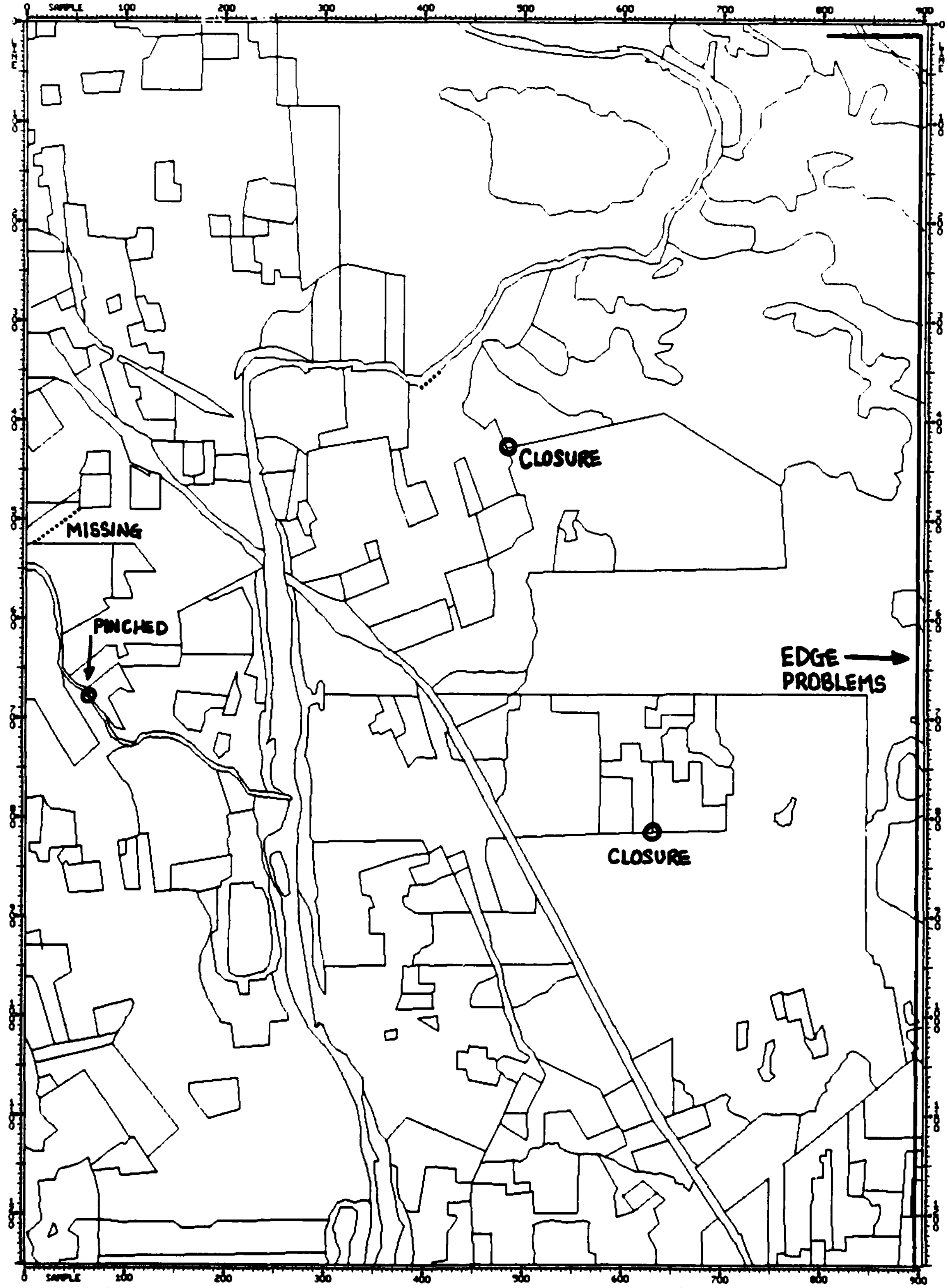


Figure 4: Digitizing errors could be identified once all formatting problems were rectified. Most errors were found on the land use data set shown here.

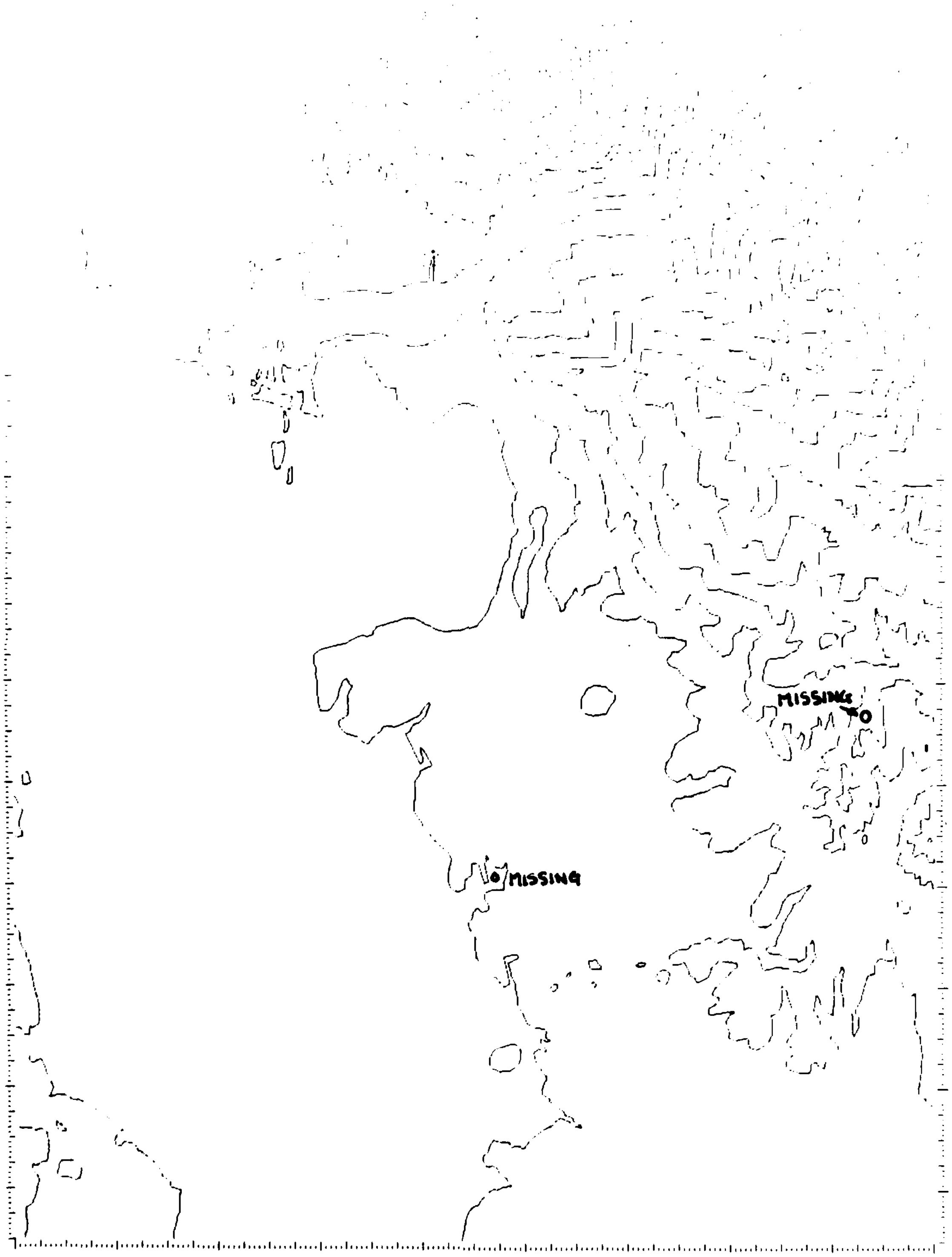


Figure 5: Although no digitizing errors were identified on the contour image, two contours were omitted.

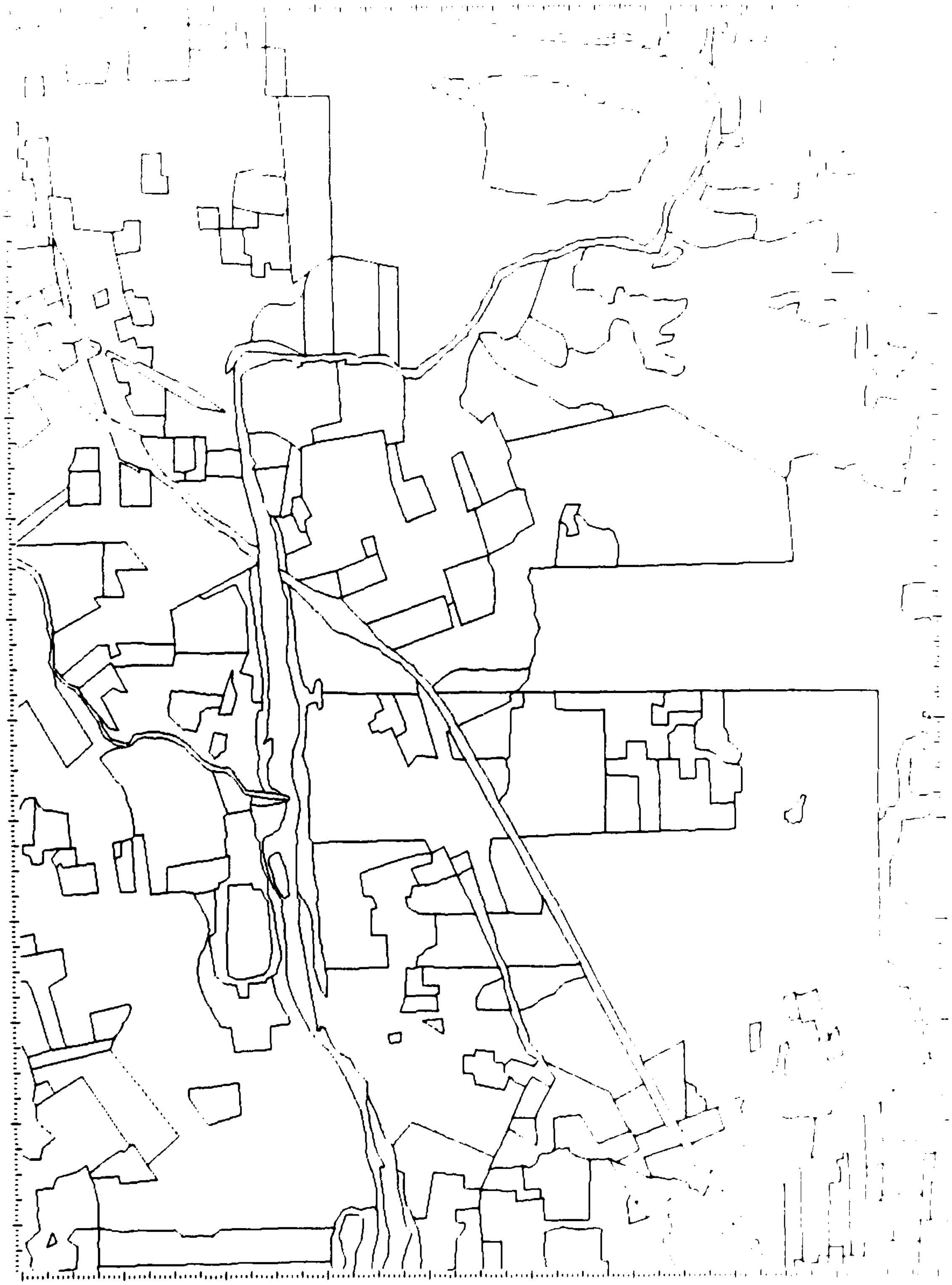


Figure 6: The edited land use image plane.

Figure 7: Closure problems were detected on the contour (shown here) and land use data planes when all regions along the right edge of each data set were encoded with the same gray tone.

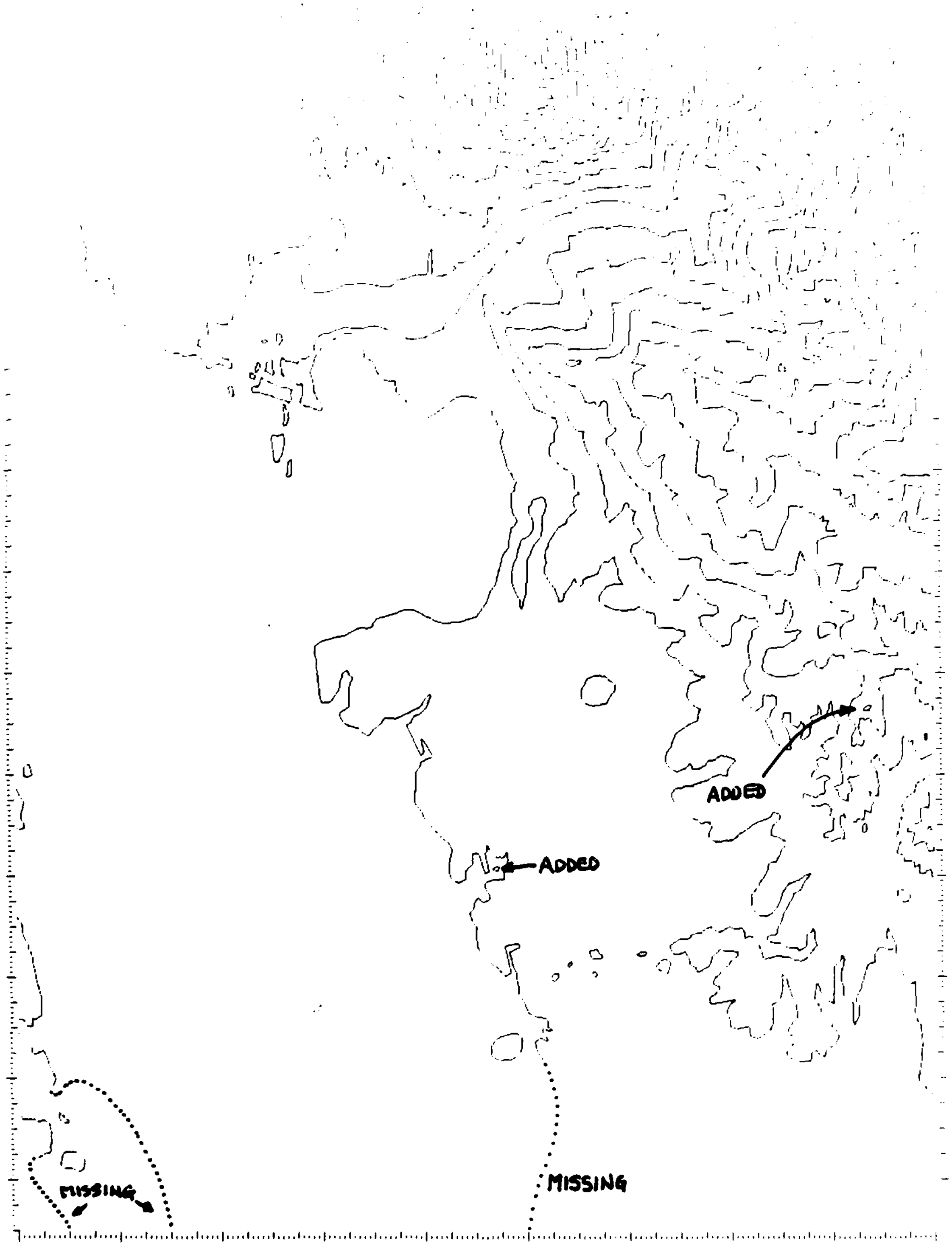


Figure 8: When the two missing contours were added to the contour data set, other line segments which were not edited disappeared. The problem, identified as a software error with the interactive editor, was corrected.

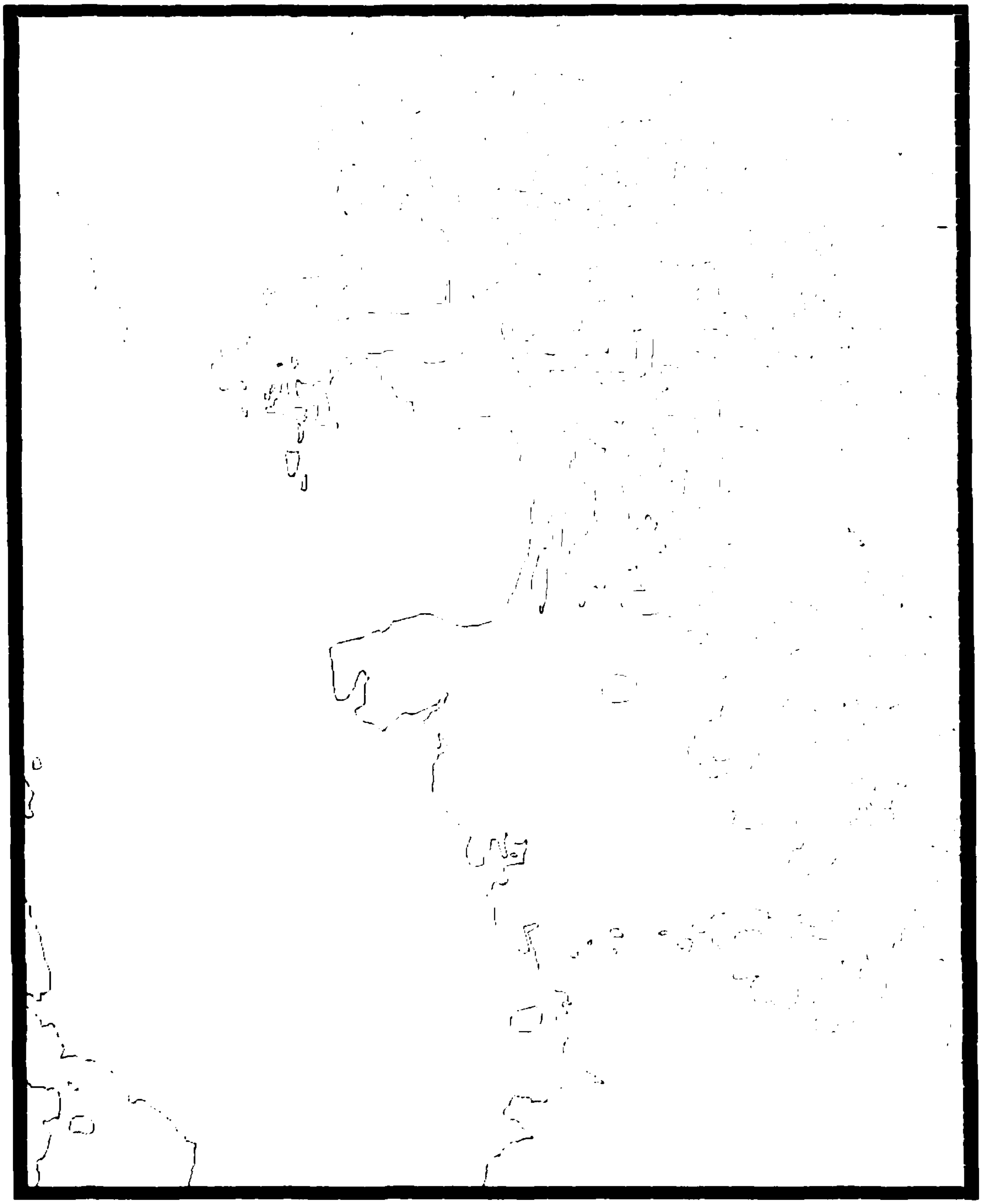


Figure 9: The edited contour image plane. (Note, the horizontal line plotted on the image is not a data error. That line was produced by the film writer.)

Figure 10: A four-color map produced from the region coded data plane of the contour data set was used to verify proper closure and depiction of all regions.

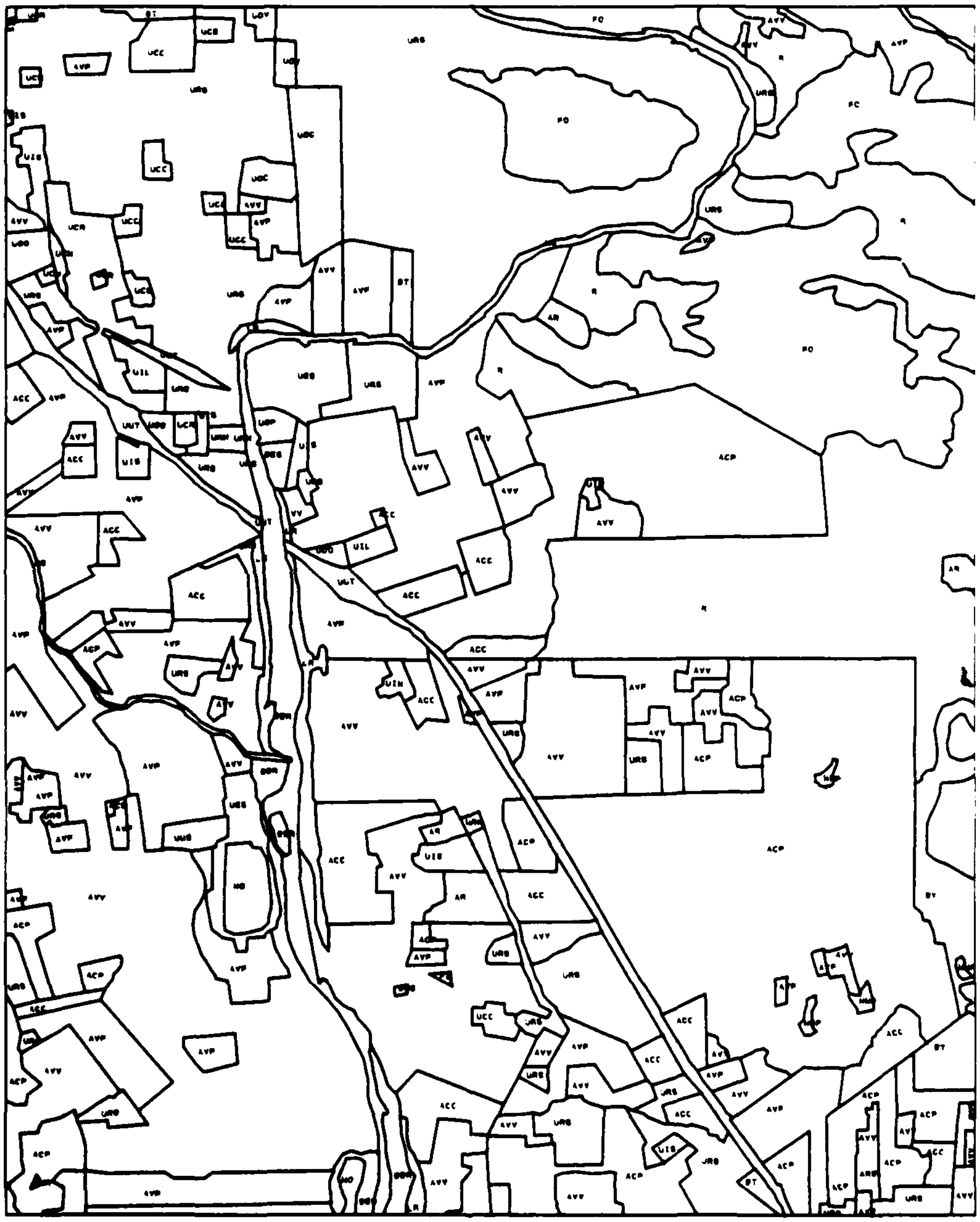


Figure 11: A test plot of the land use data plane produced by the vendor was used to compute pixel scale of the data base. (This photo-reduced version was originally plotted at a scale of 1:62500.)

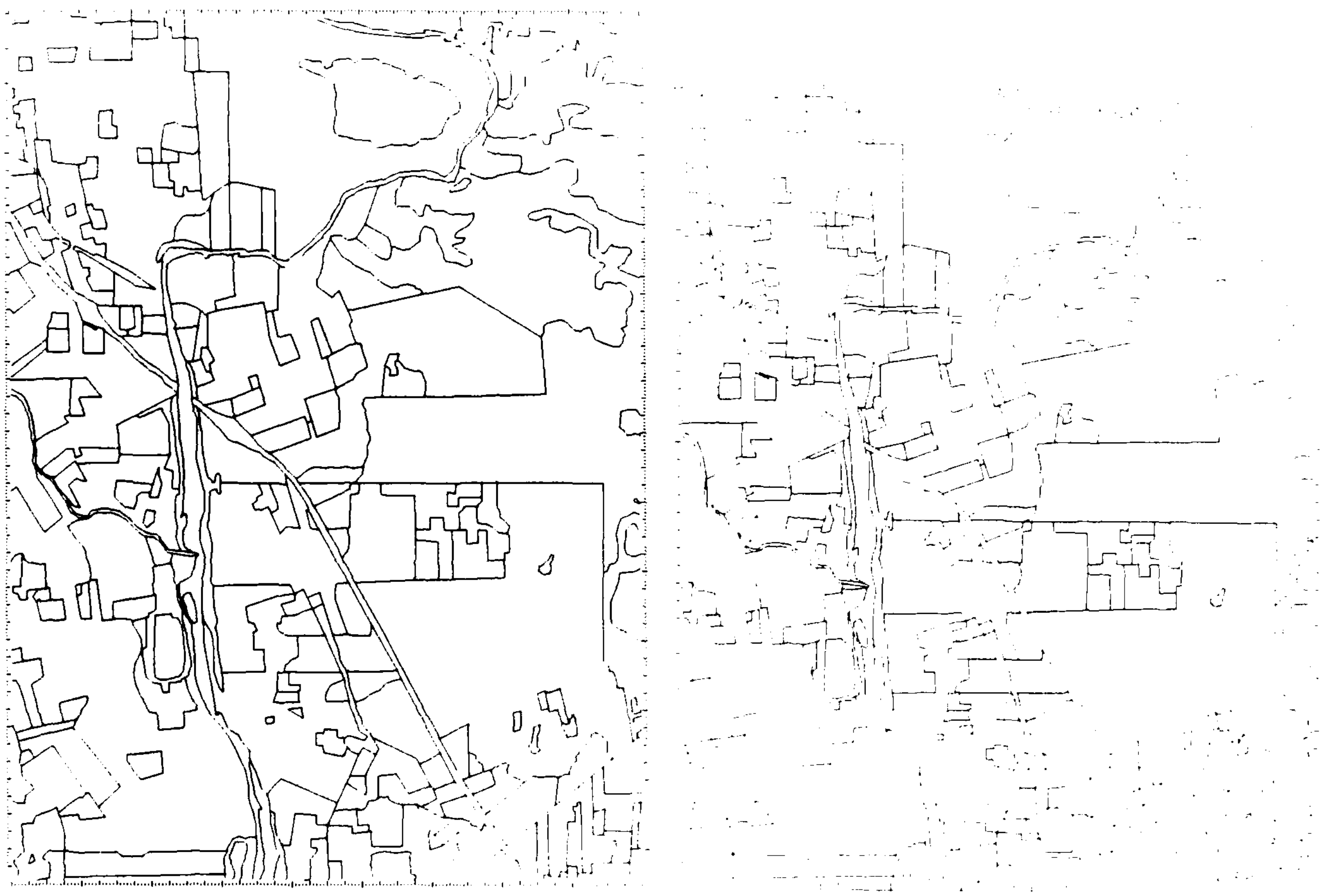


Figure 12: Comparison of scale differences between the improperly scaled data plane (left) and the properly scaled data plane (right).

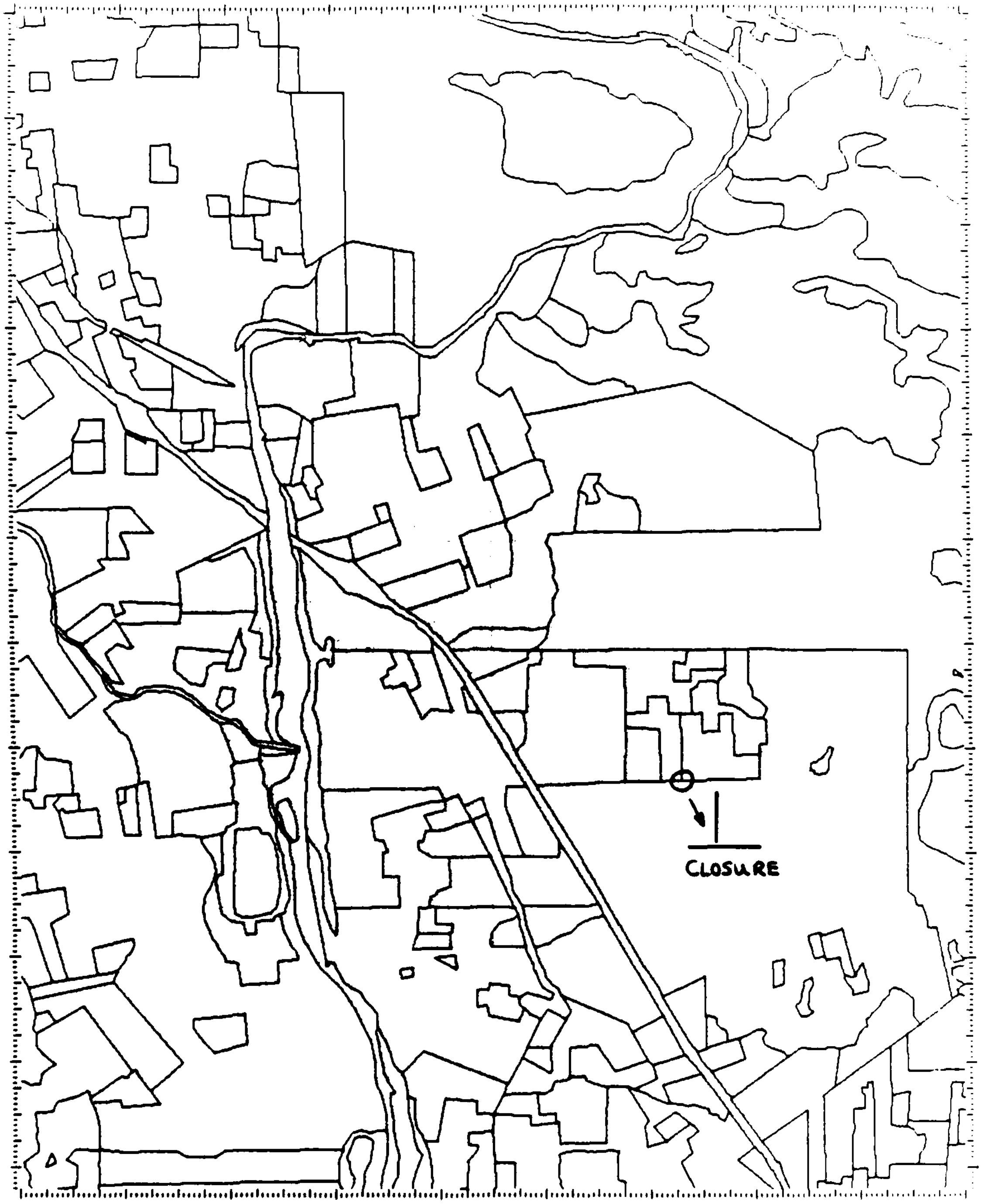


Figure 13: A closure problem was detected after rescaling the land use data plane.

References

Friedman, S. Z., 1982, Image Based Approach to Mapping, Charting, and Geodesy, Internal Publication 715-153, Jet Propulsion Laboratory, Pasadena, California.

Seidman, J. B., and A. Y. Smith, 1979, VICAR Image Processing System: Guide to System Use, Publication 77-37, Revision 1, Jet Propulsion Laboratory, Pasadena, California.

Sharpley, W. K., J. F. Leiserson, and A. H. Schmidt, 1978, A Unified Approach to Mapping Charting and Geodesy (MC&G) Data Base Structure Design, Report ETL-0144, U.S. Army Engineer Topographic Laboratories.

END

FILMED