TESTS TO ESTABLISH THE QUALITY OF DIGITAL CARTOGRAPHIC DATA: SOME EXAMPLES FROM THE DANE COUNTY LAND RECORDS PROJECT

Alan P. Vonderohe and Nicholas R. Chrisman
University of Wisconsin - Madison
Madison, WI 53706

ABSTRACT

Tests, particularly those based on independent sources of higher accuracy, are crucial to the evaluation of the quality of digital cartographic data. This paper reports interim results of tests applied to positional accuracy of USGS Digital Line Graph data with additional reference to logical consistency. Procedures to carry out tests are discussed.

BACKGROUND: QUALITY STANDARDS

While advances in some aspects of digital cartography have been rapid, studies of data quality have lagged behind. Recently, issues of data quality have begun to attract attention with the work on data standards in a number of countries (Canadian Council on Surveying and Mapping, 1982; National Committee for Digital Cartographic Data Standards [NCDCDS], 1985). This work coincides with the overhaul of US geodetic control standards (Federal Geodetic Control Committee [FGCC], 1984) and a new specification for the accuracy of large-scale maps (American Society of Photogrammetry [ASP], 1985). Although the need for standards is clear, and the professions and agencies have focussed attention on the problem, there has been remarkably little research published that investigates alternatives for quality standards. The published literature particularly lacks examples drawn from prevalent data products. Perhaps agencies or companies perform such tests internally, but the procedures used and the results obtained are not reported.

Before the cartographic standards take on a final form, there should be a fully informed debate on the nature and purpose of quality tests. The overall mission of the NCDCDS covers quality testing, but, as a volunteer committee, it cannot afford to undertake a full testing project. This paper hopes to contribute to a debate on the procedures for testing digital data quality by reporting some interim results of recent research projects.

DANE COUNTY LAND RECORDS PROJECT

The tests were performed as a part of the Dane County Land Records Project (Chrisman and others, 1984), a two year cooperative effort addressing a broad range of issues in modernizing land information. Within the broad goal, the project is directed towards information needs for soil erosion control planning. The cooperators in the project include US Soil Conservation Service, Wisconsin Departments of Natural Resources and Agriculture, Trade and Consumer Protection, Dane County Land Conservation Committee and Land Records and Regulation Department, along with University of Wisconsin faculty. For certain aspects, Wisconsin Department of Transportation, US Geological Survey and Bureau of Land Management have cooperated as well.

COMPONENTS OF QUALITY

The NCDCDS Working Group on Data Set Quality has discerned six interrelated components of quality: lineage, positional accuracy, attribute accuracy, logical consistency, completeness and currency (Chrisman, 1984). Of these, lineage is not testable because it contains a summary of the source material and transformations that lead to the data. Currency can masquerade as each of the other errors, and is difficult to test for separately. The other four components can be tested to different degrees. This paper will focus on positional accuracy, but the intent is to view quality as a whole.

TESTS OF POSITIONAL ACCURACY

The ASP committee and the NCDCDS agree that testing forms a firmer basis for quality assessment than any other method. The NCDCDS interim proposed standard has adopted the draft ASP specification as the basis for conducting a test (Merchant, 1982; ASP, 1985; NCDCDS, 1985). This testing procedure has been applied to digital products of a project preceding the Dane County project (Petersohn and Vonderohe, 1982). In this study we will apply this test to a digital data product of national significance, the Public Land Survey [PLSS] layer in the USGS National Digital Cartographic Data Bank.

USGS Public Land Survey section data

One of the most popular digital data products in the US is the description of the Public Land Survey System. Demand for this information is described in the National Research Council [NRC] report (1982) on the modernization of the PLSS. The states of Minnesota, Illinois and Montana (at least) have digitized this information for state information systems (NRC, 1982, p. 25). Phillips Petroleum (1984) has digitized 30,000 quadrangles in many states. US Geological Survey has also been active. A large proportion of their Digital Line Graph products consist of the PLSS layer, along with related political boundaries. A major recommendation of the NRC report is that USGS complete a comprehensive coverage of the PLSS to contribute to integrated mapping and leading towards a national basis for multipurpose land information systems.

The Dane County Land Records Project is a pilot project for integrating land information, with particular attention to SCS soil surveys and local property records. To help provide a basis for areas of Dane County with limited survey control, we entered into cooperation with USGS National

Mapping Division [NMD]. NMD digitized the PLSS and political boundaries for the 34 seven-minute quadrangles that cover Dane County. On the cost-sharing basis of such projects, costs were competitive with digitizing the data ourselves.

Digitizing was done at the Mid-Continent Mapping Center using graphics workstations followed by conversion into the DLG topological structure (Allder and Elassal, 1983). Unlike some digitizing operations, NMD has access to stable base versions of the maps, precluding the large errors associated with paper maps. This study is not directed at the internal procedures used by NMD, because the proof is in testing the result.

There is a difference between the legal layout of the PLSS and its cartographic representation. PLSS section lines are defined by the section corners (including closing corners) and the quarter section corners. Each quarter corner was supposed to be set on a straight line between two section corners, hence in cartographic representation, particularly at 1:24000, there may be no visible deflection at quarter corners. For this reason, this study will exclude quarter corners. To promote a truly multipurpose system, beyond the cartographic representation, USGS should consider including more explicit recognition of quarter corners.

Comparison Surveys

The quality standards call for tests against an independent source of higher accuracy. Both independence and higher accuracy must be stressed.

Independence. Independent sources would seem to require separate derivation of information, for instance a separate survey. With map information, completely separate information is quite rare because all maps and surveys usually relate to the same geodetic reference network. Independence in statistics refers not to procedures, but to results. Independent processes produce uncorrelated errors. Typically, the use of different technology (for example photogrammetry versus field survey) will suffice to produce uncorrelated errors.

Higher Accuracy. Clearly a test is most useful if one source can be treated as the "truth". The ASP specifies that the "nominal positional accuracy" of the check survey be "three times that required" of the product to be tested (Section 2.4.1). This section also specifies, somewhat redundantly, FGCC second order horizontal control surveys. For the intended large-scale maps the latter may be appropriate, but for the 1:24000 maps the former seems more in order.

Sources of Check Surveys. The digital products tested were dictated by availability of appropriate check surveys. The ASP specification calls for twenty test points distributed in a sheet. The cost of second order surveying for twenty points would be rather high; this project relied on surveys performed for other reasons. One extensive ground survey covers the remonumented section corners of the

Town of Oregon in Dane County. This survey was performed by a registered land surveyor and his measurements have been adjusted as a part of our project (Krohn, 1984). The survey was controlled by three second order stations in the National Geodetic Network, having an approximate six mile spacing. Two of these have no direct connection in the National Network. Coordinate comparisons between this survey and an inertial survey, recently conducted by Bureau of Land Management (BLM) and Dane County Land Records Project personnel, indicate a mean positional discrepancy of 0.433 meters. A great deal of this discrepancy may be attributable to the difference in control for the two surveys. The inertial survey was tied to three second order, class I control stations established by Global Positioning System technology and two stations in the National Network, only one of which was common with the ground survey. In any case, the quality of the ground survey was deemed more than adequate to perform the analysis reported herein. It provides a sufficient sample of check points for one USGS sheet (Oregon) and a smaller sample for the Attica quad.

The other area of testing is in the Town of Westport in the Waunakee quad. For this report, only those surveys performed by University of Wisconsin groups to second and third order standards have been used. All surveys were tied to a second order, class I network established during an earlier project (Crossfield and others, 1983). The number of section corners available in Westport falls below the ASP threshold, but provides additional information to evaluate the Oregon test.

Coordinates. The Digital Line Graph encodes coordinates for section corners as integers. The distance represented by one count is .61 meter on the ground. The integer coordinates are related to UTM through a transformation specified in double precision. On obtaining the DLG data, the coordinates were converted to UTM in double precision (64 bits). Due to the roundoff, there should be at least .61 meter of uncertainty.

The check surveys were computed using the Wisconsin South Zone of the State Plane Coordinate system, in feet. The coordinates in State Plane were converted to UTM using the geodetic software distributed by the National Geodetic Survey in double precision. This software may differ in some details of map projection from that used by NMD, but the results should be identical to a level of precision below .6 meters. A national effort should standardize the software used for these projections to remove any uncertainties. Though coordinates of one map projection may be readily transformed into another, it would seem wise to adopt a national standard, agreed to by the states and the user community, and thereby avoid the process altogether.

Results

The DLG coordinates were subtracted from the survey coordinates to yield the measured discrepancies. A number of summary statistics can be calculated from this raw data. The ASP (1985, Appendix A) suggests bias (mean discrepancy) and precision (standard deviation) calculated separately for each coordinate axis. For some purposes it is more customary to report Root Mean Square Error which measures absolute deviation, without accounting separately for bias. As another option, the National Map Accuracy Standards are specified in terms of straight line positional discrepancy without regard to coordinate axis. The table below shows mean distance, and gives the distance of the ninetieth percentile of the distance distribution. These figures (the basis of the National Map Accuracy Standard) are difficult to interpret, because they lack a statistical basis. It has become common practice to convert the accuracy standard to a function of standard deviation (Circular Map Accuracy Standard) (ACIC, 1962). A few of these measures are reported in Table 1 for the sheets described above. In the table, eleven of the check points in the Oregon sheet contained redundancy in the ground survey and twenty-three were established by sideshots with no redundancy. The row labelled "all data" includes points with and without redundancy. In the Attica quad, one of the points was established with redundancy and the others were sideshots. All eleven points in the Waunakee sheet contained redundancy in their ground survey.

Table 1: Summary of positional accuracies (figures in meters UTM)

Sheet	Points	Bias		Std. Dev.		RMSE		Distance	
		X	Y	X	Y	X	Y	ave.	90%
Oregon									
redundant	11	-1.6	-1.2	6.7	3.6	6.6	3.7	6.2	10.5
all data	34	-1.8	-1.2	5.4	3.5	5.6	3.6	5.7	10.2
Attica	12	1.4	3.0	5.4	3.3	5.3	4.3	5.5	9.7
Waunakee	11	-1.4	6.2	5.7	7.5	5.6	9.2	9.9	16.9

The lack of points for the Attica and Waunakee sheets makes statistical statements less applicable than in the Oregon case. The bias detected in the Oregon sheet is statistically significant, but low on a substantive scale. The standard deviations for all the sheets are lower than might have been expected. All but the Waunakee sheet pass the National Map Accuracy threshold. The Waunakee sheet is strongly affected by the large Y bias and a few large outliers. All but one of the eleven points in the Waunakee sheet showed the survey points on the positive side of the mapped points. Such errors, if they can be explained, can be easily removed.

The Oregon sheet not only passes the ASP Class 1 standard at 1:24000, it passes at a scale of about 1:18000. It must be remembered that the test covers the digitizing and computer roundoff effects, as well as the inherent error in the map. The largest error detected, in fact, may be due to a difference in information available at different times. The southwest corner of section 13 showed an error of 17.4 meters in the X direction, but the 1960 topographic survey could not have included the legal position of the section corner which was discovered a foot below the pavement in

1975. Another large error (13.5 meters) occurs at a section corner where no evidence was found in 1977 and a new monument had to be set. Without more detailed notes on the methods used by USGS to compile the maps, the erroneous assumption has to be made that section corners are perfectly known.

Discussion of Positional Test

The positional accuracy of the USGS Digital Line Graph that we tested was better than the figures usually quoted. In the NRC (1983) report and in other places, the figure of 40 feet (16.2 meters) is quoted freely. This number is derived from the outer tolerance permitted by the National Map Accuracy Standards. Although it proves to be a relatively valid description of the data, the figures on bias and precision take more of the distribution of errors into account.

The NRC (1982) report recommends that a PLSS data bank be constructed by BLM based on the USGS Digital Line Graph. If all the Digital Line Graphs are as accurate as the Oregon sheet, such a data bank might serve some functions in geographic information systems. However, such information systems could support only those purposes having the coarsest accuracy requirements, such as natural resource inventory. If, ultimately, such information systems are intended to meet the needs of all users, and thus be truly "multipurpose", the positions assigned to PLSS corners should not be considered static. The data bank should be improved as new measurements are obtained, through highway construction, property survey, control densification or other processes. We plan to provide a revised copy of the DLG data for placement in the national archive. In the few locations where comprehensive remonumentation and control surveys of the PLSS have been conducted, as in Racine County Wisconsin (Bauer, 1976), the data bank should be constructed directly without digitizing the topographic map. The maintenance of a dynamic geometric framework poses a significant challenge to software development (Chrisman, 1983).

LOGICAL CONSISTENCY

The PLSS DLG's have a topological data structure which encodes relationships between objects on the map. This provides a verifiable structure for each layer of the map, but does not deal directly with relationships between layers. Software can verify the logical consistency of each layer, and USGS is committed to distribute only clean data (so-called DLG-3). Many systems are able to produce logically consistent structures for a single layer, but few are able to handle interrelationships of features.

In this case, the boundary and section data were constructed together. One set of nodes were created for a quadrangle for use with both layers. This procedure implicitly ensures that lines intended to be identical would overlay directly on each other. The geometric relationship is as far as the structure goes. The complex attribute scheme provides for coding a line as coincident with another feature, but it does not permit naming that other feature. Furthermore, this scheme is not always used.

In one case in the Oregon sheet, a section line and a political boundary which should have been the same were separated by a 5.7 meter gap. This created a sliver polygon where a section was placed outside its township. Strangely, the proper line was present, but encoded with the same section on either side.

Another major consistency problem is created by water bodies. Due to public trust doctrines, the PLSS does not cover navigable waters. The topographic maps show section lines running up to water but not across it. With a traditional product, the section lines are in red, and the hydrography is in blue, but the blue lines are needed to make the sections topologically complete. In the digital representation there are lines to represent water boundaries with codes to show that they are intended to coincide with hydrographic features. However, there is no explicit reference possible in the structure to the actual line intended. This structural problem has been noted by others (Dangermond, personal communication, 1984). In our case, the hydrographic layer has not yet been digitized. The PLSS data contains straight lines that connect section lines together, instead of following the shore. This representation robs a rather accurate product of its ability to produce reasonable area estimates and convincing graphics. Furthermore, when the hydrography is encoded, there will be no easy way to determine which pieces to be used. We suggest that a digital line graph requires a new perspective on its purpose. The current PLSS data can be used as a useful source of section corner locations, but not as a description of sections as areas.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to a number of parties. Don Doyle performed the fieldwork and furnished the data for the ground survey in Oregon Township. David Krohn and James LaVerriere performed the least squares adjustment of Doyle's data. Nancy VonMeyer and Mr. Krohn assisted in the inertial survey which provided a check on the Oregon adjustment. The Bureau of Land Management, Cadastral Surveys Division furnished equipment and personnel for the inertial survey. Partial support for this research comes from US Department of Agriculture Hatch and University of Wisconsin University Industry Research funds.

REFERENCES

Aeronautical Chart and Information Center (ACIC) 1966, Principles of Error Theory and Cartographic Applications: Report 96, Washington DC

Allder, W.R. and Elassal, A.A. 1984, Digital Line Graphs from 1:24000-Scale Maps, in USGS Digital Cartographic Data Standards, Geological Survey Circular 895-C, Reston VA

American Society of Photogrammetry (ASP) 1985, Accuracy Specifiations for Large Scale Line Maps, Photogrammetric Engineering and Remote Sensing forthcoming

Bauer, K.W. 1976, Integrated Large-scale Mapping and Survey Control Program Completed by Racine Co. Wis., Surveying and Mapping, vol. 36, no. 4

Canadian Council on Surveying and Mapping 1982, Standards for the Quality Evaluation of Digital Topographic Data, Energy, Mines and Resources, Ottawa

Chrisman, N.R. 1983, The Role of Quality Information in the Long-Term Functioning of a Geographic Information System, Proc. AUTO-CARTO 7, vol. 1, p. 303-312

Chrisman, N.R. 1984, Alternatives for Specifying Quality Standards for Digital Cartographic Data, in Report 4, National Committee for Digital Cartographic Data Standards, Columbus OH

Chrisman, N.R., and others, 1984, Modernization of Routine Land Records in Dane County, Wisconsin: Implications to Rural Landscape Assessment and Planning, URISA Professional Papers Series 84-1

Crossfield, J.K., Mezera, D.F., and Vonderohe, A.P. 1983, The Westport Experience, Proc. ACSM-ASP Fall Meeting, p. 18-27

Federal Geodetic Control Committee, 1984, Standards and Procedures for Geodetic Control Surveys, Rockville MD

Krohn, D.K. 1984, The Least Squares Adjustment of Oregon Township: An Interim Report for the Dane County Land Records Project, University of Wisconsin-Madison

National Committee for Digital Cartographic Data Standards (NCDCDS), 1985, An Interim Proposed Standard for Digital Cartographic Data Quality, in Report 6, Columbus OH

National Research Council, 1982, Modernization of the Public Land Survey System, National Academy Press, Washington DC

Petersohn, C. and Vonderohe, A.P. 1982, Site-Specific Accuracy of Digitized Property Maps, Proc. AUTO-CARTO 5, p. 607-619

Phillips Petroleum 1984, Digitized Base Map Data Using USGS Quads (marketing brochure)