NOAA, Topographic Sheets: Vectorization and Error Analysis

By Robert H. Huxford and Richard C. Daniels
Coastal Monitoring and Analysis Program
Department of Ecology
Olympia, Washington 98504-7600

Abstract

The National Oceanic and Atmospheric Administration (NOAA) and the Washington Department of Ecology has undertaken a data rescue project to convert historical and contemporary topographic sheets (T-sheets) from paper or cloth to a digital format. The original maps have been scanned at 400 dpi, saved as raster images, and vectorized to obtain X, Y coordinate pairs that describe the location of historical shorelines depicted on the original maps. A methodology is developed and explained here for vectorization of NOAA topographic sheets and for the analysis of errors on the sheets. The error assessment methodology utilizes coordinates obtained for survey markers shown on the scanned T-sheets and compares these coordinates to those published by the National Geodetic Survey for the same markers. Differences between the measured and published coordinates are then computed and several descriptive statistics obtained.

Introduction

The National Oceanic and Atmospheric Administration (NOAA) has undertaken a data rescue project to convert historical and contemporary topographic sheets (T-sheets) from paper to a digital format. These T-sheets are detailed survey maps that were produced to provide coastlines for use on navigation charts issued by the National Ocean Service (formerly the U.S. Coast and Geodetic Survey). The information content of T-Sheets varies. At a minimum, each T-sheet contains the following: the mean high water line as derived from field survey; survey marker locations used to provide control for the field work; a coordinate system formed by longitude and latitude lines; and graphic representation of the vegetation inland from the coast. Most of the T-Sheets done for the state of Washington are at a scale of 1:10,000 or 1:20,000. In addition, many maps include topography and additional longitude and latitude lines that represent translations between old and new datums or projections.

In this data rescue project the original paper or cloth maps were scanned in black and white at 400 dots per inch on a large format scanner by NOAA. The scanned data was then saved to CD-ROMs for archive. This scanning process has been successful in saving these important informational resources from loss. However, the scanned images have no attribute information – they are just images of a map. To extract and assign attribute information from the scanned image, it is necessary to convert the information on the raster image to vector lines (arcs) that can be described by X, Y coordinate pairs.

Analysis Purpose

Although the accuracy of NOAA topographic sheets has been shown be very good (Byrnes et al. 1991, Anders and Byrnes 1991), there are few methods defined to determine the statistical accuracy of the data conversion process. Because our research involved quantitative analysis of changes in configuration of the Washington shoreline and its change in position over time, it was essential that we be able to quantify the accuracy of the data conversion process.

Our goal was to induce the minimal amount of error possible in the scanning or vectorization process and to correct for any shrinking or warping that might have occurred in the original maps. In comparison, U.S. Geological Survey 1:24000 topographic maps are only accurate to approximately 40 feet, or about 12.2 meters, for a well defined point. It was our hope that the T-sheet conversion process would produce accuracy's one half of that value. The additional accuracy is needed to be able to measure erosion or accretion quantitatively.

Methodology for Extraction Vectors from Raster Images

The process of converting the raster data to vectors is a four-step process. First, the projection and datum used on the source map [i.e., North American Datum (NAD), North American Datum of 1927 (NAD27), or North American Datum of 1983 (NAD83)] must be identified. Second, the raster image must be registered to the longitude and latitude lines on the original map. Third, the image is transformed or projected into NAD83. Last, the coastline data on the raster image must be vectorized and attributes must be added to obtain the final line work that is saved in a digital format.

Several sources may introduce error in the final digital data. The primary sources are map shrinkage, locational accuracy of the longitude and latitude, scanner errors, datum accuracy, datum transformation errors, and digitizing errors.

The methodology currently being used by the NOAA data rescue project is to register the scanned map images using the longitude and latitude lines that were drawn on the source maps by the original cartographer and survey teams. These original datums are then converted to the most currently used datum, NAD83, using available transformation algorithms. These original datums were in local datums, the North American Datum (NAD) (historical maps older than the 1927) or the North American Datum 1927 (NAD27) (contemporary maps). The survey teams also plotted the location of the benchmarks or survey marks used during their fieldwork. Some maps contained local datums, and NAD longitude and latitude lines and update marks in NAD27. For the NOAA data rescue project, all maps were converted to the newest datum NAD83 so maps from different dates could be compared with each other and the NGS benchmarks/survey marks on these maps could be compared with current NGS data sheets.

According to National Ocean Service (NOS) guidelines, map features critical to safe marine navigation are to be mapped to accuracy stricter than national standards. More specifically, the shoreline is mapped to within 0.5 mm of (at map scale) of true position. With a 1:10,000 scale, this is 5.0 m on the ground. Fixed aids to navigation and objects charted as landmarks must be located within 3.0 m at this scale. Because longitude and latitude lines are aids to navigation, the assumption is made that they and the survey markers were drawn on the original map with the same 3.0 m accuracy. If this is correct, then the coordinates of these survey markers can be used to provide an independent check of the accuracy of the registration and data extraction process (Ellis 1978, Shalowitz 1964).

As a test of accuracy of our methodologies, the shoreline and locations of 29 National Geodetic Survey (NGS) survey markers were digitized from eight topographic sheets from mapping project PH-62, conducted between 1950 and 1951 by the Coast and Geodetic Survey (now the National Ocean Service). The digitizing was done using ArcView®, Spatial Analyst®, ArcInfo® and ArcScan® software. The ArcScan software allowed lines to be digitized to within ½ the width of what was seen as the raster image of the line or approximately two pixel widths. At 400 dpi resolution for scanned images on a 1:10,000 map, the location of lines was estimated to be within 1.3 m of true.

The bench marks selected for our error analysis were evenly spread over the entire mapped area and all had sub-meter accuracy. The published location of these same marks were obtained from the NGS and used to create a separate GIS point coverage. These two digital data sets were then overlaid for comparison.

Error Analysis of the Extracted Line Work

The accuracy assessments that may be conducted vary based on the number of survey markers that are recovered on a given map and their known positional accuracy. In all cases, calculation of the minimum, mean, and maximum values will give an overall accuracy assessment of the data conversion process if three or more markers are available and they are well distributed over the land portion of the map.

Overall Error Assessment for Project PH-62

A given NOAA survey project is conducted over a one to three year period and may involve from one to many individual T-sheets. Since the same personnel work on a project throughout its lifetime, it can be assumed that the same (or similar) procedures were followed for construction of the T-sheets. Based on this assumption, an error assessment may be made for a project as a whole as well as for individual T-sheets.

For PH-62, the mean, standard deviation and median were calculated (Table 1) for differences in X, Y, and XY. For XY mean is 3.06 m (10 ft). Based on the normal distribution assumption, the error in this project varied between 0.25 m to 6.95 m.

Table 1. Comparison of published and extracted coordinates for PH-62

Marker Name	Sheet	X-Map	Y-Map	X-NGS	Y-NGS	X Diff	Y Diff	XY Diff
BURNT	T-10344	223964.43	110011.55	223963.64	110008.83	0.79	2.72	2.83
McKENZIE HEAD	T-10344	225360.68	111869.24	225361.94	111866.35	1.26	2.89	3.15
DEADMAN	T-10344	224599.66	112231.37	224603.96	112228.90	4.30	2.47	4.96
NORTH HEAD LH 1909	T-10344	224448.75	113573.92	224449.87	113572.27	1.12	1.65	1.99
BAKER WEST BASE	T-10340	229465.08	115093.09	229468.01	115091.62	2.93	1.47	3.28
LAKE	T-10340	227257.22	115176.20	227256.29	115172.97	0.93	3.23	3.36
TURN	T-10340	226915.97	116245.84	226914.31	116242.69	1.66	3.15	3.56
APEX	T-10340	230253.14	118727.98	230256.05	118726.66	2.91	1.32	3.20
TIOGA RESET	T-10340	226500.91	120421.00	226497.76	120419.81	3.15	1.19	3.37
BONNIE	T-10649	226749.26	123517.93	226748.93	123516.53	0.33	1.40	1.44
GREEN RESET	T-10649	226899.68	127530.55	226897.84	127528.75	1.84	1.80	2.57
SNAKE 2	T-10649	229550.84	128681.48	229546.62	128682.10	4.22	0.62	4.27
DOANE 2	T-9637S	229554.55	137083.37	229555.82	137082.62	1.27	0.75	1.47
OYSTER 2	T-9637S	227160.89	141168.77	227158.61	141171.03	2.28	2.26	3.21
GOULTER 2	T-9637S	229760.29	141539.65	229758.58	141538.88	1.71	0.77	1.88
MESS	T-9637N	229982.42	144909.53	229983.85	144910.32	1.43	0.79	1.63
BETTER	T-9637N	228653.27	147710.59	228649.96	147713.08	3.31	2.49	4.14
GRASSY 1939	T-9634S	228954.83	150373.61	228953.86	150377.30	0.97	3.69	3.82
LEAD 4	T-9634S	228196.88	151487.00	228196.31	151490.99	0.57	3.99	4.03
WILLAPA BAY LIGHT	T-9634N	226814.61	160922.42	226814.73	160922.64	0.12	0.22	0.25
LARKIN	T-9634N	228934.47	162353.00	228935.28	162353.30	0.81	0.30	0.86
BEACH 2	T-9634N	225942.13	163121.74	225942.72	163124.39	0.59	2.65	2.71
FIRST	T-9521	225758.32	166745.76	225756.67	166744.23	1.65	1.53	2.25
GRAY	T-9521	225583.41	170529.86	225576.82	170529.83	6.59	0.03	6.59
ROBIN 1940	T-9521	227865.44	173661.49	227862.80	173660.43	2.64	1.06	2.84
DIKE	T-9521	227333.32	173951.33	227326.55	173952.92	6.77	1.59	6.95
BERT 1940	T-9521	227879.70	174079.37	227876.19	174080.40	3.51	1.03	3.66
ISLAND	T-9521	228858.36	174855.21	228860.78	174855.20	2.42	0.01	2.42
LAST	T-9521	224502.58	176397.70	224500.59	176397.72	1.99	0.02	1.99
					Mean	2.21	1.62	3.06

Marker Name	Sheet	X-Map	Y-Map	X-NGS	Y-NGS	X Diff	Y Diff	XY Diff
					Stnd Dev	1.67	1.14	1.48
					Median	1.71	1.47	3.15
					Minimum	0.12	0.01	0.25
					Maximum	6.77	3.99	6.95

Conclusions

The process of converting the raster data to vectors is a multi-step process. Each step, or the quality of the original data source, may introduce error into the final digitized data. The conversion process herein utilized the existing longitude and latitude coordinate system on the map. Thus, to obtain an independent assessment of the error within the final data product, a method was needed that could compare measured (i.e., from the digitized data) and published coordinates for known points on the map.

The preceeding examples demonstrate how the coordinates for survey markers, published by National Geodetic Survey, may be used in combination with coordinates measured from the vectorized T-sheets to obtain error assessments for the map conversion process. The statistical methods described here may be used to find errors in the vectorized data. The following conclusions can be made about the methodologies and procedures used in this data conversion process:

1. The average accuracy of the benchmarks we measured is 3.06 m (10 ft) and this meets the NOS guidelines for fixed aids to navigation and objects charted as landmarks. This accuracy is stricter than national standards and four times the accuracy of current USGS 1:24,000 scale topographic maps. Because of this, the original T-sheets can be assumed to also meet NOS guidelines and to be very accurate in their depiction of the shoreline that existed at the time of the surveys.

2. The methods used by NOAA to scan the T-sheets are sound and induce no identifiable error into the vectorization process.

3. Any shrinkage or warping of the original paper maps has been corrected by the registration and rectification process.

4. The vectorization process using ArcScan® and ArcEdit® in ArcInfo® is accurate to within half the width of a line (0.79 m with a 1/32 inch line on a 1:20,000 scale map).

5. The NGS benchmarks can be effectively used to verify the accuracy of the digital data produced by the data rescue project.

References

Anders, F.J. and Byrnes, M.R. 1991. Accuracy of Shoreline Change Rates as Determined from Maps and Aerial Photographs. **Shore and Beach**, Vol. 59, pp. 17-26.

Byrnes, M.R., Hiland, M.W., McBride, R.A., and Westphal, K.A. 1991. **Pilot Erosion Rate Data Study: Harrison County, Mississippi**. Federal Emergency Management Administration, Office of Risk Assessment, Washington, DC.

Ellis, M.Y. (editor). 1978. **Coastal Mapping Handbook**. U.S. Department of Commerce, National Ocean Survey, Washington, DC.

Schalowitz, A.L. 1964. **Shoreline and Sea Boundaries, Volume 2**. U.S. Department of Commerce, Coast and Geodetic Survey, Washington, DC.