Journal of Coastal Research	17	3	611–619	West Palm Beach, Florida	Summer 2001

An Error Assessment of Vector Data Derived from Scanned National Ocean Service Topographic Sheets

Richard C. Daniels† and Robert H. Huxford‡

†Department of Transportation
Information Technology
P.O. Box 47340
Olympia, WA 98504-7430
E-Mail:
danielri@wsdot.wa.gov

‡Department of Natural Resources
Information Technology Division
P.O. Box 47031
Olympia, WA 98504-7031
E-Mail:
robert.huxford@wsdnr.gov

ABSTRACT

DANIELS, R.C.; HUXFORD, R.H., 2001. An Error Assessment of Vector Data Derived from Scanned National Ocean Service Topographic Sheets. *Journal of Coastal Research,* 17(3), 611–619. West Palm Beach (Florida), ISSN 0749-0208.

The National Oceanic and Atmospheric Administration, National Ocean Service (NOS) has undertaken a data-rescue project to convert historical topographic sheets (T-sheets) from paper or cloth to a digital format. The original maps have been scanned and saved as raster images. These images have been registered and the line work vectorized using heads-up digitizing methods to obtain X, Y coordinate pairs that delineate the location of the shorelines depicted on the original maps. A methodology is described here for obtaining error estimates for the derived vector data. The error assessment methodology uses the coordinates obtained for survey stations digitized from the scanned T-sheets and compares these coordinates to those published by the National Geodetic Survey. Differences between published coordinates for survey stations and those measured from the T-sheets have been calculated and descriptive statistics obtained. In southwest Washington and northwest Oregon the mean error for 1926/27 and 1950-era T-sheets is ±3 m for 1:10,000 scale sheets and ±6 m for 1:20,000 scale sheets.

ADDITIONAL INDEX WORDS: *National Ocean Service, U.S. Coast & Geodetic Survey, historical shoreline delineation, shoreline position, mean high water line, Oregon, Washington*

INTRODUCTION

The National Ocean Service (formerly the U.S. Coast & Geodetic Survey) has produced topographic (T-sheets) and hydrographic (H-sheets) surveys since 1834, when the first topographic and hydrographic surveys for the Great South Bay of Long Island, New York were completed. T-sheets are detailed records of the surveys that were conducted to provide shorelines for use on navigation charts issued by the National Ocean Service (NOS) (SHALOWITZ, 1964). The surveys conducted by NOS and it predecessors comprise the most complete, consistent, and accurate record of historical shoreline position available for the United States. As such, these historical surveys are often the only source of information available to determine shoreline change or position for periods prior to the advent of controlled air photography.

The National Oceanic and Atmospheric Administration (NOAA) has undertaken a data-rescue project to convert historical T-sheets into digital formats (NATIONAL OCEAN SERVICE, 1999a). In this data-rescue project the original paper or cloth T-sheets are being scanned in black and white at 400 dots per inch on a large format scanner located at NOS headquarters in Silver Spring, Maryland. The scanned data are then saved to CD-ROM for archive. This scanning process has been successful in saving these important resources from further physical deterioration and possible loss. However, these scanned maps are just images, the true value of these images will not be realized until the images are registered to a coordinate system, shorelines digitized, and attribute information added.

NOAA has partnered with several state agencies across the United States to assist in extracting attribute information and line work from the digital images (*e.g.,* HUXFORD and DANIELS, 1998). Under these cooperative arrangements NOAA provides the scanned raster images to their partners. The partners then register and vectorize the shorelines and other relevant features from the raster T-sheets and return the data to NOAA in a vector format suitable for use in a geographic information system (GIS) (NATIONAL OCEAN SERVICE, 1999b).

As the coastal zone management agency for the state of Washington, the Department of Ecology is currently involved in a detailed analysis of shoreline change within the Columbia River Littoral Cell—which covers the ocean coast of Southwest Washington and Northwest Oregon (KAMINSKY *et al.,* 1998; KAMINSKY *et al.,* 1999). As such, the Department of Ecology has undertaken the task of digitizing the NOS

99127 received 13 December 1999; accepted in revision 3 July 2001.

shorelines for Washington and the northwest ocean coast of Oregon. To complete our analysis of shoreline change within the Columbia River Littoral Cell an accuracy assessment of the data derived from the historical T-sheets was required. Described herein are the statistical methods used to obtaining error estimates for the vector data derived from the historical T-sheets.

TOPOGRAPHIC SURVEYS AND T-SHEETS

NOS traditionally conducted two kinds of surveys in support of their nautical chart production work, topographic and hydrographic. NOS nautical charts are recompilations of information from one or more topographic and hydrographic surveys and information from other sources. As such, the scale of a chart may be significantly different from the topographic or hydrographic surveys used in its compilation (*e.g.,* 1:80,000 vs. 1:10,000).

Topographic surveys produced T-sheets as their primary product at scales ranging from 1:5,000 to 1:20,000 (smaller scale sheets are often associated with lower order reconnaissance or preliminary surveys and are referred to as preliminary charts or sketches). T-sheets constructed for Washington and Oregon prior to the 1930s were plotted using plane table methods, after which photogrammetric methods were used. Note that this change did not modify the required mapping accuracy standards used by NOS.

Three different sets of T-sheets are available (1850–1890, 1926–27 and 1950–1958) for southwest Washington and northwest Oregon. The information content of these T-sheets varied over time. At a minimum, each T-sheet contained the mean high water line as derived from field survey (1850/90 and 1926/27) or by photogrammetric methods (1950-era), survey station locations used to provide control for the field work, a coordinate system formed by longitude and latitude projection lines, and a graphic representation of the vegetation inland from the shore. It should be noted that many of the older T-sheets contain additional longitude and latitude lines (projection lines) and datum correction graticules. These depict changes between spheroid models and datums over time. Some of the oldest T-sheets in Washington State contain as many as three sets of projection lines. Most of the T-sheets available for the Columbia River Littoral Cell are at a scale of 1:10,000 or 1:20,000.

EXTRACTION OF VECTORS FROM RASTER IMAGES

NOS T-sheets have been shown to be accurate in previous studies (BYRNES *et al.,* 1991; ANDERS and BYRNES, 1991; CROWELL *et al.,* 1991). However, there are few well-documented methods for independently determining the statistical accuracy of the three-step (paper-raster-vector) data conversion process used here. In this project several improvements in the data conversion process have been implemented. First, line work was digitized from raster images scanned directly from the original paper or cloth document (not second or third generation copies). Secondly, the maps are registered using the original longitude and latitude projection lines on each sheet –as many as 60 registration points were digitized per T-sheet. Thirdly, heads-up digitizing software was used with auto-tracing capabilities. This method allowed unlimited ability to zoom into features on the map and obtained accuracy's unobtainable with manual digitizing methods. Lastly, survey stations on the map were digitized, but not used, in the initial registration process.

By digitizing the survey stations, but not using them in the initial registration of the map, we were able to use the survey stations as an independent check of the accuracy of the complete registration, datum conversion, and projection process. This was made possible by the fact that the original longitude and latitude projection lines and the survey stations were drawn on the original map with the same relative accuracy (ELLIS, 1978; SHALOWITZ, 1964).

The process of converting raster images to a vector format is a four-step process. First, the projection and datum used on the source map [*i.e.,* United States Standard Datum/North American Datum (NAD), North American Datum of 1927 (NAD 27), or North American Datum of 1983 (NAD 83)] must be identified. Second, the raster image must be registered to the longitude and latitude projection lines on the original map. Third, the coastline data on the raster image are digitized and attributes added to obtain the final line work that is saved to a vector data file. Lastly, the vector data is transformed or projected into the desired datum and coordinate system (in this case Washington State Plain, meters, NAD 83).

Several factors may introduce error into the final vector data. The primary error sources are as follows:

1. The accuracy of the longitude and latitude projection lines drawn on the map,
2. ability of the computer operator to identify the original longitude and latitude projection lines and to determine the source datum, projection, and spheroid of the map,
3. non-uniform shrinkage of the original map,
4. variations in the speed of movement of the media under the scanner during the scanning process,
5. the transformation method or program used to convert between datum's, projections, and spheroid models,
6. ability of the computer operator to accurately trace the line work on the raster image and save the data to a vector or line based file.

The raster images provided by NOS were registered by Ecology based on the latitude and longitude projection lines drawn on the map by the original cartographer and survey teams. The line work on each image (*e.g.,* survey stations, shorelines, aids to navigation, etc.) was digitized using the heads-up digitizing method in conjunction with an automated line vectorization program (ArcScan™, Environmental Sciences Research Institute, Inc., Redlands, California,) and a set of ArcView and ArcInfo GIS scripts provide by the NOAA Coastal Services Center, Charleston, South Carolina.

The vectorization process used allowed the operator to digitized to within half the width of a line, a ground distance of 0.79 m (2.6 ft) for a 1/32 inch line on a 1:20,000 scale map or 0.39 m (1.3 ft) on a 1:10,000 scale map (HUXFORD and DANIELS, 1998). During the digitization process the operator traced and stored X, Y coordinate pairs that described the lines of interest into a digital vector file. Vectors derived from maps completed in the 1950's were in NAD 27 and were easily projected into the desired datum, in this case NAD 83, using the NGS program NADCON (HUXFORD and DANIELS, 1998). Older maps required additional processing to shift between datum's (*e.g.,* NAD to NAD 27) or to correct for different spheroid models (*e.g.,* Bessel 1841 to Clark 1866).

Because the longitude and latitude projection lines on the scanned T-sheet were used to register the raster image, they can not be used to assess the accuracy of the data transformation process. Since the original longitude and latitude projection lines and the survey stations shown on the map were drawn with the same accuracy (ELLIS, 1978; SHALOWITZ, 1964), we may use the published coordinates of these survey stations to provide an independent check of the entire registration, projection, and data extraction process.

In this case study the shoreline and location of twenty-nine National Geodetic Survey (NGS) survey stations were digitized from nine T-sheets in mapping project PH-62 and PH-155 compiled between 1950 and 1957. Twenty-eight NGS survey makers were digitized from four T-sheets compiled in 1926 and 1927 from project PH-46C. The markers selected are uniformly spread over the landward portion of each sheet. The published location of these survey stations were obtained from the NGS and used to create a separate GIS point coverage. These two digital data sets were then overlaid for comparison and error analysis.

For the most part the NGS survey marks utilized in this analysis have horizontal coordinates that are second or third order. As stated by John Bossler (1993), "when a horizontal control point is classified with a particular order and class, NGS certifies that the . . . coordinates of that control point bear a relation of specific accuracy to the coordinates of all other points in the horizontal control network." This relationship is expressed as a ratio, where the ratio is a measure of the maximum positional error between two points related to the horizontal separation of the points. By definition, first-order horizontal control stations have geodetic coordinates that are within 10 ppm (1 cm + 1:100,000) of their true location at the time of survey; second-order, class I and class II, stations have published coordinates that are within 15 ppm and 18 ppm; and third-order, class I and class II, stations have published coordinates that are within 20 ppm or 25 ppm of their true location (BOSSLER, 1993).

The order and class of a NGS station is meaningful only when considered in relationship to the separation between pairs of stations. In this study the survey markers depicted on each T-sheet were digitized and compared to coordinates obtained from the National Geodetic Survey for each station. The digitized coordinate and published coordinate for each station may be seen as such a station pair, and assuming a normal distribution the maximum potential separation (99.74% probability) between the such a pair is 13.83 m (45.4 ft) for the T-sheets studied.

In the worst case scenario the maximum positional error inherent in the published coordinates of a third-order, class II station, is ~0.007 m (0.02 ft) over 13.83 m (45.4 ft). Recall that the heads-up digitization process used in this study was able to digitize within half the width of a line, or a ground distance of 0.79 m (2.6 ft) with a 1/32 inch line on a 1:20,000 scale map. Note that accuracy issues related to the class and order of a given survey station is smaller that the ability of the cartographer or surveyor to place the station on the T-sheet, and thus may be safely ignored in this analysis.

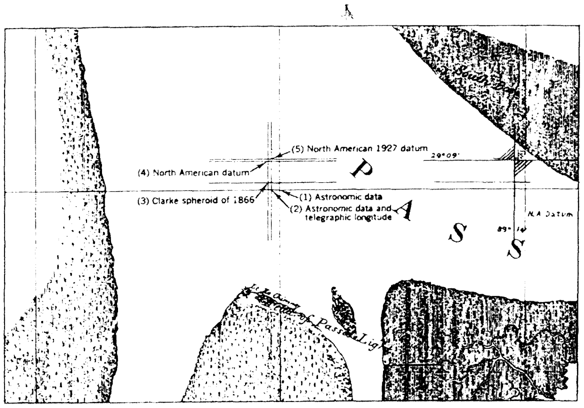
Ellipsoids, Projections, and Datums

If we are to accurately complete an error analysis of historic maps, it is necessary for each map to be transformed to a common datum, ellipsoid model, and projection. For this error analysis the common datum used was NAD 83 with a Lambert Conformal Conic projection (with the Washington State Plane South coordinate system).

Before 1853 the U.S. Coast and Geodetic Survey (CGS) used the Bonne projection for most of its charting activities. In 1853, the CGS adopted its own polyconic projection. This projection was invented by Ferdinand Hassler, the first superintendent of the U.S. Coast Survey, and is a derivative of the French Bonne projection (SNYDER, 1987). This new CGS projection has been referred to in CGS literature by various names. Names commonly used for this projection include ordinary, simple, coast survey, or the American Polyconic projection. CGS adopted this projection because map projection lines could be easily constructed from published projection tables while in the field (SHALOWITZ, 1964). This projection became a standard for much of the official mapping of the United States including U.S. Geological Survey (USGS) quadrangle sheets (SNYDER, 1987). After the development of the State Plane Coordinate Systems (SPCS) in the early 1950's the USGS and CGS ceased using the American Polyconic projection in favor of the Mercator or Lambert Conformal Conic projection.

There were only two known ellipsoid models used by the CGS for T-sheets made within the state of Washington. The first is the Bessel Spheroid of 1841. It was used between 1844 and 1880. The second is the Clark Spheroid of 1866. The Clark Spheroid was not adopted for use by the CGS until about 1880. The Clark model was adopted because it incorporated newer, more accurate earth radius values for the equator and poles (SNYDER, 1987). These more accurate values allowed for the construction of more precise topographic maps.

Unlike the ellipsoid models and projections, there was a proliferation of local datum's that were developed by CGS cartographers in the agencies early years. As settlements spread across the continent, each region developed its own system of triangulation based on one or more astronomical determination of latitude, longitude, and azimuth. Because of errors in astronomical methods, each settlement essentially had its own independent datum. After the transcontinental arc of triangulation was completed in 1899 it became apparent that many local datum's contained significant errors (some over 1000 meters). In 1901 the United States Standard Datum was established. The datum recalculated the position of every triangulation station within the United States except for the coordinate of a station located at Meades Ranch, Kansas, and the azimuth from this station to station Waldo. In 1913 the triangulation network was connected with the Mexican and Canadian networks and the datum was renamed the



Notes:

1. Astronomical Datum. (Used on maps made prior to 1901.) Digitized. (A small portion of the astronomical datum near the other correction lines needs to be digitized so adjustments can be made between it and the other datum's/corrections.)
2. Telegraphic Longitude Correction. (Only appears on some maps, generally on the east coast.) Digitized if it appears.
3. Clark Spheroid of 1866 Correction. (Must be applied to most maps made before 1880. Corrects maps made between 1844 and 1880 that use the Bessel spheroid of 1841.) Digitized if it appears.
4. North American Datum Correction. (Must be applied to maps made before 1901, possible some shortly after that time. This is used to correct between locally established datum's and the North American Datum / U.S. Standard Datum.) Digitized if it appears.
5. North American 1927 Datum Correction. (Must be applied to maps made before 1927, possible some shortly after that time. This is used to correct between the North American Datum / U.S. Standard Datum and the North American 1927 Datum.) Digitized if it appears.

* All the above corrections are digitized as lines!

** It should be noted that any time a spheroid model is changed (i.e.: astronomical, Bessel spheroid of 1841, Clark Spheroid of 1866) a XY shift only approximates the correction for a map over a small localize area. To be technically correct, the map needs to be reprojected, otherwise small distortions persist. These errors become apparent when appending adjacent maps together or applying a single correction to a large area. Reprojecting a map using manual cartographic methods is a daunting task. Fortunately, modern Geographical Information Systems (GIS) make this process possible for both vector and image data.

Figure 1. Topographic Sheet corrections used to correct older T-sheets. Portion of an early topographic sheet (reduced scale) showing corrections to the projection lines as a result of changes in the geographic data (from SHALOWITZ 1964).

North American Datum to reflect its international character. The United States Standard Datum and the North American Datum are identical within the continental United States (SNYDER, 1987).

With further additions to the triangulation network it became necessary to conduct another adjustment of the network in 1927 (NAD 27). Starting at that time and continuing until 1932, all primary triangulation data were adjusted to reflect this new datum. All three of these datums were based on a common point (Meades Ranch), and a common ellipsoid (Clark Spheroid of 1866). Figure 1 shows how these datum shifts are shown on historical T-sheets. The most modern datum, NAD 83, the one being used as a base line for this error analysis, differs from NAD 27 in that it is based on an earth-centered model of the earth (GRS 80) rather than being based on an initial point and azimuth.

Contemporary T-sheets (1950 era) to NAD 83

Contemporary T-sheets are considered to be those produced after 1930 that utilized the NAD 27 datum. As these T-sheets were digitized, the coordinates for the intersections of the latitude and longitude lines are entered into the GIS system. These coordinates are then used to register the map in NAD 27. The final step in the process is to take the NAD 27 data and project in the GIS into NAD 83 using NADCON (a datum conversion program written by the NGS).

Table 1. *Methods tested to covert 1920 era T-sheets from the North American Datum to NAD 27 and their resulting errors.*

Correction Method	Error (m)	Error (ft)
Graphical	10 to 12	32.8 to 39.4
Tabular	10	32.8
Photographic	10 to 15	32.8 to 39.2
Mathematical	6	19.7

Turn of the Century T-sheets (1920 era) to NAD 83

T-sheets made between 1901 and 1930 on the Pacific Coast were constructed using either the United States Standard Datum (1901 to 1913), or the North American Datum (1913 to 1927). Each datum used the same spheroid (Clark 1866) and projection (American Polyconic) as used in the 1950 era T-sheets. The only difference is that the coordinates of the origin point of the datum shifted during the NAD 1927 adjustment. These adjustments, shown in Figure 1 (items 4 and 5), were calculated for each T-sheet and then processed in the same manner as contemporary T-sheets. The major difference being that a XY shift was made to the entire map prior to the projecting from NAD 27 to NAD 83.

Four methods were tested to determine the optimum shift to apply to the T-sheets. The four techniques and their resulting errors are shown in Table 1. In the graphical technique, the digitized map elements were moved as a unit such that the digitized longitude and latitude projection lines would matched the NAD 27 corrections drawn on the original T-sheet in the 1930's by NOS cartographers. In the tabular method, a XY shift was made based on X and Y values extracted from CGS datum difference tables (PATTON, 1985). In the photographic technique, digital orthophotos with an accuracy of ±1.5 m were used as a backdrop while the map was moved to match up with known features that appeared on both the map and the photo (*e.g.,* a lighthouse). In the final mathematical technique, a mean error was calculated for all the survey stations on each T-sheet and an average error or offset calculated and a XY shift made. This proved to be the most accurate method and the only one that brought the maximum error to within the NOS standard of ±6.0 M for 1:20,000 scale topographic maps.

T-sheets and Charts Made Prior to the North American Datum

Corrections for CGS charts and reconnaissance surveys produced prior to 1920 are more problematic since additional adjustments for astronomical azimuth errors and transformation from the Bessel 1841 to the Clark 1866 spheroid are needed. GIS methods to deal with these transformations are still under development.

At this time fourteen historical CGS reconnaissance surveys or charts that cover Washington and northwest Oregon have been examined. Preliminary results indicated that the

Table 2. *Comparison of published (by the NGS) and extracted coordinates for seven survey stations shown on topographic sheet T-9521. Coordinates in Washington State Plane, South, NAD83, meters.*

					Difference (m)		
Station Name	X-Map	Y-Map	X-NGS	Y-NGS	X	Y	XY
LAST	224500.59	176397.72	224502.58	176397.70	1.99	−0.02	1.99
ISLAND	228860.78	174855.20	228858.36	174855.21	−2.42	0.01	2.42
BERT	227876.19	174080.40	227879.70	174079.37	3.51	−1.03	3.66
DIKE	227326.55	173952.92	227333.32	173951.33	6.77	−1.59	6.95
ROBIN	227862.80	173660.43	227865.44	173661.49	2.64	1.06	2.94
GRAY	225576.82	170529.83	225583.41	170529.86	6.59	0.03	6.59
FIRST	225756.67	166744.23	225758.32	166745.76	1.65	1.53	2.25
				Median	2.64	0.01	2.84
				Standard Dev.	3.15	1.08	2.09
				Minimum	−2.42	−1.59	1.99
				Mean	2.96	0.01	3.82
				Maximum	6.77	1.53	6.95

Data sheets with coordinates for NGS and U.S. C&GS survey marks may be obtained from the World Wide Web at http://www.ngs.noaa.gov.

older maps have rotational errors ranging from −0.84° (skewed counter clockwise) to +0.25° (clockwise). These errors are not uniform over time or space. It is believed these deflections in magnetic north may be due to magnetic anomalies within the region (ATWATER, 1991) that were not corrected for during these initial reconnaissance surveys. Other possibilities include the "less stringent" astronomical observation methods used to calculate magnetic variation for reconnaissance surveys (SHALOWITZ, 1964), or distortion caused by the conversion between spheroid models.

The spheroid problem can only be resolved if the central meridian and standard parallels used on the chart are known. The only place this information may exist is in the descriptive reports that accompanied each survey when it was submitted to the superintendent of CGS or NOS for review and acceptance. At the present time these reports are not available from NOAA, NOS, or the US national archives.

Perhaps the greatest barrier to evaluating the overall accuracy of these surveys is the lack of survey stations with known coordinates. Only five to ten percent of the survey stations shown on the Washington and northwest Oregon charts had documented coordinates. These charts were registered by the Department of Ecology using survey stations (when available) and coordinates for locations which had not moved significantly over the last century (e.g., offshore rocks, stream meander bends). This registration process has achieved linear distances accurate to within ±15 m between known points that were over five kilometers apart. Due to the limitations described above, no attempt was made to apply the detailed error assessment methodology described here to these charts.

ACCURACY ASSESSMENT FOR INDIVIDUAL T-SHEETS

The accuracy assessments that may be conducted vary based on the number of survey stations that are recovered on a given T-sheet. 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. If five or more survey markers are available, and they are distributed throughout the map, then a trend and skew analysis may be conducted.

The skew statistic may be used to determine if the mean error obtained is skewed toward the minimum or maximum value of the sample. If the sample is skewed to the left (negative), then a majority of the sample values are less than the mean. Conversely, right (positive) skew indicates that a majority of the values are greater than the mean. Large skew values may indicate that the entire map is shifted in the indicated direction. Skew may also be used to identify problematic survey markers. For example, since the completion of the original survey the survey station may have been destroyed and a new one of the same or similar name installed. The NGS data sheet for survey markers flagged by the skew statistic should be reviewed to assist in the identification and removal of problematic stations from the analysis.

A trend analysis may be conducted to determine if the measured errors (between known coordinates and those derived from the digitized data) are systematic. Linear regression methods may be used to determine if a trend exists in the X or Y coordinate. A systematic error in the X or Y coordinate may be an indicator of errors in the scanning or projection process. For example, variation in roller speed (the speed at which the paper map traveled under the scanning head) during the scanning process may have resulted in a stretch along the Y coordinate that resulted in increased error from the bottom to the top of the map.

An Example Using T-Sheet 9521

The following example is based on T-sheet 9521 (1951), scale 1:20,000, of Grayland, Washington from project PH-62. This T-sheet was transformed from NAD 27 to NAD 83 and the line work on the raster image extracted and saved (HUXFORD and DANIELS, 1998). During the digitizing process, crosshairs representing the location of third order or higher horizontal survey stations on the sheet were digitized. The coordinates of each survey station was obtained and saved to a spreadsheet and compared to published coordinates obtain from the NGS for these same stations. In Table 2, the extracted coordinates are compared with those obtained from the NGS.

Simple Statistical Tests

Utilizing the information in Table 2, the minimum, mean, and maximum error associated with sheet T-9521 was determined. The standard deviation and median of the X difference is 3.15 m (10.3 ft) and 2.64 m (8.7 ft). The standard deviation and median of the Y difference is 1.08 m (3.5 ft) and 0.01 m (0.03 ft), respectively.

The Pearson's coefficient of skew may be used to determine if there is a tendency for the values to be larger or smaller than the mean. Skew values between ±0.5 m (1.6 ft) indicate that the sample is symmetrical and for most practical applications that the sample may be considered as representative of a population with a normal distribution (RUNYON and HABER, 1984). In the above example, *sk* was 0.30 m (1.0 ft) for the X difference and −0.03 m (−0.1 ft) for the Y difference.

Based on the assumption of a normal distribution, the maximum error (99.74% probability) may be calculated as follows:

$$\text{Max}^{99.74\%} = 3 \times \text{Standard Deviation} + \text{Mean} \quad (1)$$

The combined error in the X and Y has a mean XY error of 3.82 m (12.5 ft) and a standard deviation of 2.09 m (6.8 ft). The error in this sample ranged from a low of 1.99 m (6.5 ft) to a high of 6.95 m (22.8 ft). Assuming a normal distribution, there is a 99.74% probability that the maximum positional error on this T-sheet is less than 10.09 m (33.1 ft).

Regression Tests

In cases where more than five survey markers are available for a given area, a linear regression model may be constructed to assist in identifying potential systematic offsets or errors within a individual T-sheet. The model hypothesizes that the actual X or Y coordinate (from the NGS) may be calculated based on the measured values obtained from the T-sheets and a one-dimensional slope factor with a value of one and a intercept value of zero. In this example the following equation for a line is fitted to the data:

$$X_{\text{actual}} = X_{\text{measured}} \times m + b \quad (2)$$

where m is the slope of the line and b is the y intercept. In a case where no linear error exists, the m value would equal one and b would equal zero. If the calculated m is significantly different from one then there is a systematic error in the given coordinate. If the b value is significantly different from zero and is larger than the standard error for b, then the entire map may be offset from its origin by the given amount. The regression statistics calculated from the X and Y coordinates shown in Table 2 are as follows. The m and b value for the X coordinate is 0.999 and 152.4 m (500.0 ft), respectively. The standard error estimate for b is 193 m (633.2 ft). The m and b value for the Y coordinate is 0.999 and 32 m (105.0 ft), respectively, with a standard error estimate of 21 m (68.9 ft). These values indicate that the measured X coordinate for T-9521 does not have an offset in the X origin. The origin of the Y coordinate may have a small positive offset. The closeness of both m values to 1 shows that a linear systematic error is not evident in the data.

To provide an example of how a systematic error would affect the regression analysis, the X coordinates shown in Table 2 were modified by adding 15 m (49.2 ft) to the X coordinate value of each survey station for every 1000 m (3280.8 ft) the station was west of survey marker "LAST" (the east most survey marker in the table). This procedure introduced a linear systematic error into the measured X coordinates [*i.e.,* the error increased by 15 m (49.2 ft) for every 1000 m (3280.8 ft) traveled west].

The linear regression analysis was repeated for the X coordinate using the modified data. The calculated m was 0.980 and the b, or intercept, was 4516 m (14818.2 ft) with a standard error for b of 1769 m (5803.8 ft). As expected, when a systematic error was introduced the slope (m) value became less than 0.99 and the standard error for b became much smaller than the b value itself. The change in the slope (m) value indicated that the desired 1 to 1 relationship between the X and Y coordinate was weak, while the large standard error of b indicated that the X coordinate was offset from the origin.

ACCURACY ASSESSMENT FOR PROJECTS

A typical NOS project is conducted over several years and may involve several 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 each sheets. Based on this assumption, an error assessment may be made for a project as a whole, as well as for individual T-sheets. Combining information for an entire project increases the total number of survey markers used in the analysis and improves the results obtained by the statistical tests described in this paper. Note that the regression methodology previously described is not applicable to an entire project, and will not be utilized here.

Project PH-62 and PH-155 (1950 era T-sheets)

Eight 1:10,000 scale T-sheets and one 1:20,000 scale T-sheet from project PH-62 and PH-155 were used in the analysis. The mean, standard deviation and median for the nine sheets were calculated for differences in X, Y, and XY and are shown in Table 3. The XY difference mean is 3.15 m (10.33 ft) with a standard deviation of 1.48 m (4.8 ft). The error in this sample varies between 0.25 m (0.8 ft) to 6.95 m (22.8 ft), with a 99.74% probability that the error is less than 7.50 m (24.6 ft). Both the X and Y coordinate have positive skews of 1.95 m (6.4 ft) and 1.87 m (6.1 ft), respectively, which may indicate a non-normal distribution of the data. This relatively large skew statistic is a result of the use of two different map scales in project PH-62 and PH-155, that introduced a non-linear discontinuity between the adjoining 1:10,000 and 1:20,000 scale T-sheets.

Project PH-46C (1926/27 T-sheets)

The same analysis as described above was conducted for PH-46C for four sheets. This project consists of 1:20,000 scale

Table 3. *Comparison of published (by the NGS) and extracted coordinates for projects PH-62 and PH-155 conducted in the 1950s. Coordinates in Washington State Plane, South, NAD83, meters.*

						Difference (m)		
Station Name	T-sheet	X-Map	Y-Map	X-NGS	Y-NGS	X	Y	XY
BURNT	T-10344	223964.43	110011.55	223963.64	110008.83	0.79	2.72	2.83
McKENZIE	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	T-10344	224448.75	113573.92	224449.87	113572.27	−1.12	1.65	1.99
BAKER WB	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
WB 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
				Median		0.93	0.75	3.15
				Standard Dev.		2.65	1.99	1.48
				Minimum		−4.30	−3.99	0.25
				Mean		0.89	0.27	3.06
				Maximum		6.77	3.23	6.95

Note: Station names in italics are from 1:20,000 scale T-sheets. All others are from 1:10,000 scale T-sheets.

T-sheets. The calculated mean XY for this project is 5.10 m (16.7 ft) (Table 4) with a standard deviation of 2.91 m (9.5 ft). The X and Y coordinates have a skew of −0.13 m (0.42 ft) and −0.11 m (0.36 ft), respectively. The error in this sample varied from a minimum of 0.24 m (0.8 ft) to a maximum of 10.73 m (35.2 ft). Assuming a normal distribution, there is a 99.74% probability that the maximum error in the project is less than 13.83 m (45.4 ft).

Assessment Results

When the 1926/27 T-sheets in PH-46C are compared to the 1950-era sheets in PH-62 and PH-155, one finds that the older project has a larger mean error (5.10 m vs. 3.06 m). The larger mean error was expected. However, this larger error is not due to the reasons most commonly thought of (*i.e.,* age of the document, equipment, surveying techniques), but is related to the mapping scale used in the project. Recall that the stated accuracy of well-defined points on the original 1950 era NOS T-sheets varies based on scale from ±3 m (9.8 ft) for 1:10,000 T-sheets to ±6 m (19.6 ft) for 1:20,000 T-sheets. Because PH-46C is comprised of 1:20,000 T-sheets, whereas PH-62 and PH-155 consists mostly of 1:10,000 sheets, we should have expected the mean error for PH-46C to be approximately twice that of PH-155 and PH-62.

CONCLUSION

Converting the raster images to vector data was a multistep process, where each step may have introduced error into the final data product. The registration procedure used in this study utilized the original longitude and latitude projection lines drawn on each T-sheet to register the image to real world coordinates. To obtain an independent assessment of the error of the final data, a method was developed and describe here that compares measured (*i.e.,* from the digitized data) and published coordinates obtained from a third party for known points on the map. The following conclusions have been made about the vector shoreline data derived from the scanned T-sheets during this data conversion process:

1. The methods used by NOS to scan the T-sheets are sound and induced no identifiable error into the raster images.
2. Shrinkage or warping of the original paper T-sheets was corrected by the registration and rectification process.
3. The digitizing process used is accurate to within half the

Table 4. *Comparison of published (by the NGS) and extracted coordinates for project PH-46C conducted in 1926 and 1927. Coordinates in Washington State Plain, South, NAD83, meters.*

Station Name	T-sheet	X-Map	Y-Map	X-NGS	Y-NGS	Difference (m)		
						X	Y	XY
TAHOLAH 1914	T 4306	233310.87	211489.92	233304.20	211494.05	6.67	−4.12	7.84
QUINAULT	T 4306	230576.43	213163.28	230570.66	213160.67	5.78	2.62	6.34
NORTH 1927	T 4306	226466.44	214321.91	226467.82	214327.71	−1.38	−5.79	5.96
WRECK	T 4306	222160.03	217889.83	222162.02	217885.89	−2.00	3.93	4.41
PIER 1927	T 4306	218763.92	218536.42	218756.47	218539.54	7.45	−3.12	8.08
HIGHLANDS	T 4306	216144.87	219218.46	216144.28	219224.94	0.59	−6.49	6.51
BLUFF	T 4306	212742.33	219743.41	212744.36	219743.52	−2.03	−0.11	2.03
HEAD	T 4306	208257.68	220237.51	208263.94	220228.79	−6.26	8.72	10.73
COPALIS ROCK	T 4305	208410.91	219883.66	208415.65	219881.96	−4.74	1.71	5.04
HEAD	T 4305	208263.11	220237.86	208263.94	220228.79	−0.83	9.07	9.11
CONNOR	T 4305	199475.91	221382.48	199475.20	221389.92	0.71	−7.44	7.48
SAMPSON	T 4305	196320.80	221489.53	196326.47	221492.00	−5.67	−2.47	6.18
HUT	T 4305	193904.26	221498.98	193903.44	221498.45	0.82	0.53	0.98
BROWN	T 4305	186883.66	221026.49	186878.68	221026.16	4.99	0.33	5.00
KLIPSAN	T 4251	131883.37	226991.26	131880.78	226994.95	2.59	−3.69	4.51
GREEN	T 4251	127521.29	226902.04	127530.43	226899.90	−9.14	2.24	9.41
BONNIE	T 4251	123510.75	226753.59	123518.11	226749.41	−7.36	4.17	8.46
SEAVIEW	T 4251	117526.69	226198.76	117528.62	226197.96	−1.93	0.80	2.09
HOLMAN	T 4251	116279.86	226016.67	116282.06	226015.31	−2.20	1.35	2.58
TURN	T 4251	116247.61	226911.48	116245.92	226915.85	1.69	−4.38	4.69
LAKE	T 4251	115177.48	227254.51	115176.10	227257.50	1.37	−2.99	3.29
HILL	T 4251	114906.79	227177.06	114901.95	227179.27	4.84	−2.21	5.32
ILWACO	T 4251	114311.24	227149.12	114311.84	227150.73	−0.60	−1.61	1.72
START	T 4251	114060.33	227217.85	114057.24	227215.18	3.09	2.67	4.09
DOCK	T 4251	113419.79	227418.56	113413.66	227418.62	6.13	−0.06	6.13
POINT 2	T 4251	113160.56	226680.21	113158.28	226677.00	2.27	3.21	3.93
EAST BATTERY	T 4251	111246.57	226874.29	111247.24	226873.99	−0.67	0.30	0.73
NORTH HEAD LH	T 4251	113573.77	224448.90	113573.77	224448.66	0.00	0.24	0.24
PARK 2	T 4252	134381.43	227219.32	134382.52	227222.04	−1.09	−2.72	2.93
ALICE 2	T 4252	137888.52	227114.81	137885.91	227108.99	2.61	5.82	6.38
OYSTER 2	T 4252	141172.46	227160.52	141168.86	227160.96	3.60	−0.44	3.63
SAND	T 4252	144278.19	227035.55	144281.08	227036.78	−2.89	−1.23	3.14
BEACH 2	T 4252	145835.62	226845.48	145833.94	226855.60	1.68	−10.12	10.26
DIG 2	T 4252	147951.16	226813.62	147949.62	226815.62	1.54	−2.00	2.52
BETTER	T 4252	147709.88	228654.39	147710.73	228653.07	−0.85	1.32	1.57
LEAD 3	T 4252	151630.56	227606.25	151635.17	227596.88	−4.61	9.37	10.44
				Median		0.29	0.09	4.84
				Standard Dev.		3.98	4.40	2.91
				Minimum		−9.14	−10.12	0.24
				Mean		0.12	−0.07	5.10
				Maximum		7.45	9.37	10.73

Note: All T-sheets in this project are at 1:20,000 scale.

width of a line, a ground distance of 0.79 m (2.6 ft) for a 1/32 inch line on a 1:20,000 scale map.

4. The published coordinates for NGS survey stations that are shown on historical T-sheets may be use to independently verify the accuracy of the digital line work produced by the registration and digitizing process.
5. The vector data obtained from the 1950-era T-sheets meet published NOS accuracy standards for the original data source. The 1926/27 T-sheets are slightly less accurate, yet on average still meet the 1950-era accuracy standards.
6. By extension, the mean high water line (shoreline) digitized from the historical T-sheets are within ±12 m of their actual location during the time of survey for 1:20,000 scale T-sheets and ±9 m for 1:10,000 scale T-sheets (assuming the original survey crew was able to visually identify the mean high water line to within ±6 m).

The procedures demonstrated in this paper utilized published coordinates for NGS survey stations in combination with coordinates digitized from scanned T-sheets to obtain an independent accuracy assessment of the vector data derived during the digitization process. Statistical methods were used to search for and identify linear and systematic errors in the final vector data that may have been introduced during the data conversion/extraction process.

ACKNOWLEDGMENTS

This research was supported in part by the National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS), Pacific Coast Program and the Washington State Department of Ecology, Coastal Monitoring and Analysis Program. This work was supported in part by the U.S. Geodetic Survey as a component of the Southwest Washington Coastal Erosion Study. The authors thank the NOAA

Coastal Services Center for developing several ArcView Avenue™ and ArcInfo AML™ scripts used by Ecology during the data conversion process. Special thanks go to David McKinnie and Peter Leon, NOS Pacific Coast Program, and Greg Fromm of the National Geodetic Survey for their assistance in scanning and delivering the raster T-sheets to Ecology. Thanks are also extended to Nicholas Kraus, U.S. Corps of Engineers, who provided a detailed review of an early version of this document.

LITERATURE CITED

ANDERS, F.J. and BYRNES, M.R., 1991. Accuracy of shoreline change rates as determined from maps and aerial photographs. *Shore and Beach,* 59, 17–26.

ATWATER. T., 1991. Tectonics of the northeast Pacific. *Transactions of the Royal Society of Canada,* Series I, Vol. 1, pp. 295–318.

BOSSLER, J.D., 1993. *Standards and Specifications for Geodetic Control Networks.* NOAA, National Geodetic Information Branch, Rockville, Maryland.

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.

CASTAGNA, R.G.; THORNTON, L.L., and TYRAWSKI, J.M., 1999. Where is ELLIS Island? GIS and coastal boundary disputes. *ArcUser,* October–December, pp. 67–69.

CROWELL, M.; LEATHERMAN, S.P., and BUCKLEY, M.K., 1991. Historical shoreline change: Error analysis and mapping accuracy. *Journal of Coastal Research,* 7, 839–852.

ELLIS, M.Y., 1978. *Coastal Mapping Handbook.* U.S. Department of Commerce, National Ocean Survey, Washington, DC.

HUXFORD, R.H. and DANIELS, R.C., 1998. Historical map recovery using multiple integrated ESRI programs. *ESRI Eighteenth Annual International User Conference Proceedings,* ESRI, Redlands, California.

KAMINSKY, G.M.; RUGGIERO, P., and GELFENBAUM, G., 1998. Monitoring coastal change in southwest Washington and northwest Oregon during the 1997/98 El Niño. *Shore and Beach,* 66, 42–51.

KAMINSKY, G.M.; DANIELS, R.C.; HUXFORD, R.H.; MCCANDLESS, D., and RUGGIERO, P., 1999. Mapping of erosion hazard areas in Pacific County, Washington. *Journal of Coastal Research,* Special Issue No. 28, 158–170.

NATIONAL OCEAN SERVICE, 1999a. *Annual Operating Plan, Fiscal Year 1999.* U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Silver Spring, Maryland.

NATIONAL OCEAN SERVICE, 1999b. *FY 1999 Vector Shoreline Acquisition and Maintenance Plan.* U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Marine Chart Division, Silver Spring, Maryland.

PATTON, R.S., 1985. *Datum Differences: Atlantic, Gulf, and Pacific Coast of the United States.* U.S. Department of Commerce, U.S. Coast and Geodetic Survey, Washington, DC.

RUNYON, R.P. and HABER, A., 1984. *Fundamentals of Behavioral Statistics.* New York: Newbery Award Records, Random House, Inc.

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

SNYDER, J.P., 1987. *Map Projections—A Working Manual.* USGS Professional Paper 1395. U.S. Government Printing Office, Washington, DC.