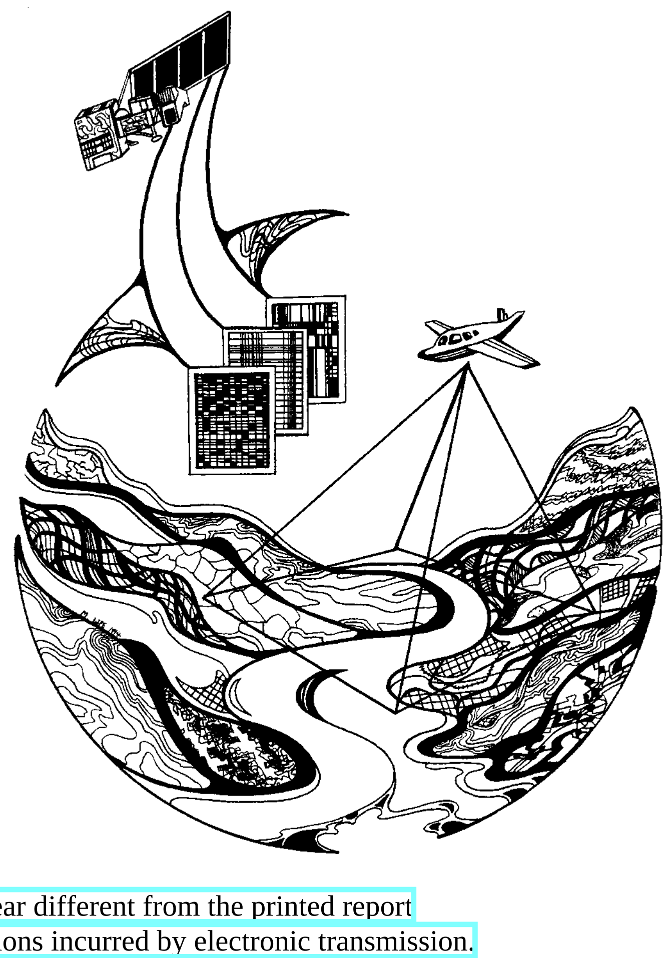
Long Term Resource Monitoring Program

Technical Report

96-T002

Geospatial Application:

Estimating the Spatial Accuracy of Coordinates Collected Using the Global Positioning System



April 1996

Geospatial Application:

Estimating the Spatial Accuracy of Coordinates Collected Using the Global Positioning System

by

Thomas Owens
National Biological Service
Environmental Management Technical Center
575 Lester Avenue
Onalaska, Wisconsin 54650

and

David McConville
Saint Mary's University of Minnesota
700 Terrace Heights
Winona, Minnesota 55987

April 1996

LTRMP Technical Reports provide Program partners with scientific and technical support.

All reports in this series receive anonymous peer review.

Cover graphic by Mi Ae Lipe-Butterbrodt

The National Biological Service . . . gathering, analyzing, and sharing the biological information necessary to support the wise stewardship of the Nation's natural resources.

Printed on recycled paper

Contents

Page

Preface v

Abstract 1

Introduction 1

Methods 4
  Study Site 4

Results and Discussion 7

Conclusion 13

Acknowledgment 14

References 14

Appendix. Equations and Calculations A-1

Tables

Table 1. Coordinates of survey and Global Positioning System measurements 9

Table 2. The effects of different sample days 12

Table 3. Testing for the effects of different sample days 13

Table A-1. Calculations for the χ^2 goodness-of-fit test for the northing distances A-2

Table A-2. Calculations for the χ^2 goodness-of-fit test for the easting distances A-3

Table A-3. Calculations for the correlation coefficient A-5

Figures

Figure 1. Distribution of distances between the Global Positioning System (GPS) coordinates and survey coordinates, and the ellipses, which describe the sample, accuracy, and probability distribution function 5

Figure 2. Location of the study area 6

Figure 3. Location of Global Positioning System sample points 8

Figure 4. Regression of the number of sample points taken at each Global Positioning System location versus euclidian distance ($R^2 = 0.031$) . 11

Preface

The Long Term Resource Monitoring Program (LTRMP) was authorized under the Water Resources Development Act of 1986 (Public Law 99-662) as an element of the U.S. Army Corps of Engineers' Environmental Management Program. The LTRMP is being implemented by the Environmental Management Technical Center, a National Biological Service Science Center, in cooperation with the five Upper Mississippi River System (UMRS) States of Illinois, Iowa, Minnesota, Missouri, and Wisconsin. The U.S. Army Corps of Engineers provides guidance and has overall Program responsibility. The mode of operation and respective roles of the agencies are outlined in a 1988 Memorandum of Agreement.

The UMRS encompasses the commercially navigable reaches of the Upper Mississippi River, as well as the Illinois River and navigable portions of the Kaskaskia, Black, St. Croix, and Minnesota Rivers. Congress has declared the UMRS to be both a nationally significant ecosystem and a nationally significant commercial navigation system. The mission of the LTRMP is to provide decision makers with information for maintaining the UMRS as a sustainable large river ecosystem given its multiple-use character. The long-term goals of the Program are to understand the system, determine resource trends and effects, develop management alternatives, manage information, and develop useful products.

This report supports Task 4.3.1.2, *Develop, Maintain, and Enhance Geographic Information Systems and Remote Sensing Analysis Capabilities*, of the Operating Plan for the Upper Mississippi River System Long Term Resource Monitoring Program (USFWS 1993). This Task involves collecting and managing spatial datasets, including quality assurance and quality control. This report was developed with funding provided by the Long Term Resource Monitoring Program.

Additional copies of this report may be obtained from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161 (1-800-553-6847 or 703-487-4650).

This report may be cited:

Owens, T., and D. McConville. 1996. Geospatial Application: Estimating the spatial accuracy of coordinates collected using the Global Positioning System. National Biological Service, Environmental Management Technical Center, Onalaska, Wisconsin, April 1996. LTRMP 96-T002. 13 pp. + Appendix.

Geospatial Application:

Estimating the Spatial Accuracy of Coordinates Collected Using the Global Positioning System

By Thomas Owens and David McConville

Abstract

Evaluating the accuracy of spatial data is important to determine the appropriate use of these data. However, a good method has not been documented to measure locational accuracy. The Global Positioning System (GPS) reduces the difficulty of measuring the location of objects and enables non-surveyors to determine their location with relative ease. This study applied a straight-forward, repeatable, and statistically sound method of estimating the horizontal accuracy of GPS-derived location data. We concentrated on the spatial accuracy of points because points represent simple locations and not cartographic abstractions such as lines or polygons.

When GPS coordinates are taken at surveyed locations, the quantity of interest is the difference from the surveyed (assumed true) coordinates. This difference in coordinates is a bivariate quantity and the probability distribution function (PDF) can be described by an ellipse with the center at $\overline{X}$ and $\overline{Y}$. An ellipse is an appropriate shape for a PDF; it has two dimensions but is not rectangular because the joint probability of points occurring in the corners is very small, and it is generally not circular because *X* and *Y* are not necessarily the same. There are three ellipses of interest: the standard ellipse, the confidence ellipse, and the tolerance ellipse. The standard ellipse is a descriptive tool used to visualize the ellipse's shape and orientation. It contains about 40% of the sample, is not dependent on the sample size, and cannot be used for statistical inference. The other two ellipses have identical shapes and orientation but different major and minor axes. The confidence ellipse is an estimate of accuracy; the sample mean is or is not significantly different from the survey locations at a given α. The tolerance ellipse is an estimate of precision; a given percentage of the population sampled is enclosed in the tolerance ellipse at a given α.

Thirty-six locations were measured and compared to surveyed locations. The average offset was -1.13 m in the northing (*Y*) direction and 0.18 m in the easting (*X*) direction. Hotelling's one-sample test determined that H_0 (no significant departure from the survey locations exists) was rejected at the 0.05 level, which indicates there was a systematic error in the sample in the south and east directions. Ninety-five percent of the population sampled (at the 0.05 level) was contained in an ellipse that was centered on 0.18, -1.13, and had a major axis of 7.49 m, and a minor axis of 5.12 m with an angle of 87.74°. Thus, if an additional point were taken, we are 95% confident that it would fall within this tolerance ellipse.

Introduction

Evaluating the accuracy and precision of spatial data in geographic information systems (GIS) is important in determining the appropriate use of these data. For example, if the measured locations of sampling points were accurate to within 100 m of their actual location, the data could be used for general assessment of trends for an area, but not for site-specific analysis. However, a good method for measuring the accuracy of location measurements has not been documented because of the complexity of the problem and the difficulty of obtaining accurate location information with traditional survey techniques.

It is conceptually easier to determine the spatial accuracy of points than of lines or polygons. A point may represent a specific location, such as a survey point, or it may represent an area displayed at a small scale, such as a city displayed on a world map. If a point represents a real, single location, it is possible to measure its displacement from its "true" location. Lines and polygons are complex abstract objects and measuring their spatial accuracy is more complex than measuring that of points (Goodchild and Gopal 1989). For example, a polygon representing a vegetation bed is a cartographic abstraction with uncertainty associated with its defining boundaries in addition to its positional accuracy. For this reason, we concentrate on the spatial accuracy of points and leave the spatial accuracy of polygons for later study.

The advent of the Global Positioning System (GPS) has reduced the expense and complexity of traditional survey techniques for measuring the location of objects on the earth. The GPS enables nonsurveyors to determine the location of objects with relative ease and accuracy. However, a simple, clear, repeatable, and statistically sound method of determining the accuracy of GPS-generated data has not been readily available. We have devised a straightforward, repeatable, and statistically sound method of estimating the horizontal accuracy of GPS-derived coordinates. Two measures of accuracy, the root mean square and the circle error probable, are widely used. Root mean square (RMS) error is the square root of the sum of the square of the euclidian distances of the error, divided by the number of observations. It has two main drawbacks: (1) it contains no information about the direction of the error, and (2) it is only a measure of accuracy, with no information about the variability (precision) of the measurements; no confidence interval may be associated with the RMS. Circle error probable (CEP) is a circle that will enclose 50% of the dot print. The CEP has the same disadvantages as the RMS.

It is now necessary to define error, accuracy, and precision. Although some authors use these terms interchangeably, distinguishing among them is important. Error is the difference between a quantity's correct value and its estimated value (Thapa and Bossler 1992). Accuracy is the nearness of quantities to their true values (Bolstad et al. 1990). In a sample of coordinates representing the same position, accuracy is the deviation of the mean coordinate from the true point. Errors may be classified into three types: (1) gross errors or blunders, (2) systematic errors, and (3) random errors. Gross errors are caused by the carelessness or mistakes of the observer using the equipment. Measurements laden with gross errors are worthless; the elimination of gross errors is essential. Systematic errors are caused by environmental factors, instrumental imperfections, and human limitations; they introduce bias into the estimate. Even after all gross and systematic errors have been eliminated, some small random errors will remain. Random errors are caused by flaws of equipment and observer and cannot be modeled. Random errors have the following characteristics: (1) positive and negative errors occur with the same frequency, (2) small errors occur more often than large errors, and (3) large errors rarely occur (Thapa and Bossler 1992).

Precision is the similarity of repeated measurements among themselves. In a sample of coordinates representing the same position, precision is the deviation of the points from the mean coordinate. Total error for a particular measurement is the sum of the mean (systematic) and deviation (random) errors. Average error (accuracy) and distribution about this mean (precision) can characterize the positional uncertainty of digital spatial data, which is related to the statistical distribution of the errors and can be modeled with a probability distribution function (PDF). For normally distributed data, the mean error characterizes the accuracy, and the standard deviation characterizes the precision (Bolstad et al. 1990).

When assessing the accuracy of test data by using field data, the control data must be surveyed with an accuracy of at least an order of magnitude higher than that of the test data. The average difference between the test and control data estimates the accuracy, whereas the frequency distribution establishes the theoretical PDF, which contains information on both accuracy and precision.

The GPS contains several sources of error: (1) satellite clock error, (2) receiver error, (3) atmospheric–ionospheric effects, (4) satellite geometry, and (5) selective availability (S/A). The GPS works by determining how long it took a signal to travel from a satellite to a receiver; if the clocks are inaccurate in the satellite or the receiver, error will be introduced. More expensive GPS receivers have more accurate clocks and less noisy circuitry. Water vapor in the atmosphere can affect the speed of the signal traveling from the satellite, as can the charged particles in the ionosphere. These atmospheric–ionospheric effects are the second largest source of error, after S/A. If the satellites used to calculate the position of the receiver are close together in the sky, the error will be higher, and if they are spread apart, the error will be lower. This error is expressed as the position dilution of precision (PDOP), which is a unitless number that describes the satellite geometry. A PDOP of 5 or less is considered good. The intentional lessening of accuracy by the Department of Defense (the custodian of the GPS) is S/A. The use of differential GPS can reduce many of these errors and subsequently improve the accuracy of a point estimate from 100 to 1–2 m from the true point (Hurn 1993).

The GPS receiver used in the study used the Standard Positioning Service (SPS). The SPS uses the single-frequency coarse/acquisition (C/A) code, which has less accuracy than the Precise Positioning Service (PPS). The C/A code is subject to S/A; the Y-code is not subject to S/A. The PPS is the military system using the Y-code, which is becoming available to Federal civilian agencies.

August et al. (1994) assessed the quality of horizontal position data obtained from an inexpensive GPS receiver. Their study was not designed to identify or measure different sources of error that degrade GPS data. The basic unit of analysis was the distance between two first-order survey points and the computed locations from the GPS receiver rover taken on different days. They found that 95% of the GPS derived locations were within 73 m without differential correction and within 6 m with differential correction. They also concluded that averaging replicate fixes significantly improved accuracy; 50 or more replicates markedly improves accuracy, and significant day-to-day variation in accuracy existed.

When a sample of GPS coordinates is taken at surveyed locations, the quantity of interest is the difference in GPS coordinates from the surveyed (assumed true) coordinates. The X and Y difference between GPS coordinate estimates and the surveyed coordinate is a bivariate quantity, and the joint PDF can be described by a probability ellipse (Batschelet 1981) with the center at $\overline{X}$ and $\overline{Y}$. An ellipse is an appropriate shape for a joint PDF because it has two dimensions; it is not rectangular because the joint probability of points occurring in the corners is small, and it is generally not circular, because X and Y are not necessarily the same.

If the individual dimensions (X and Y) are linear and normally distributed, they can be described by two statistics: (1) the mean, and (2) standard deviation. The ellipse can be described with five statistics: the sample means of (1) X, and (2) Y, and the sample standard deviations of (3) X, (4) Y, and (5) the correlation coefficient between X and Y. The quantities X and Y are jointly distributed, where X and Y depend on each other, but different pairs are independent of each other in a sample. From these statistics, ellipses can be constructed.

There are three ellipses of interest: the standard ellipse, the confidence ellipse, and the tolerance ellipse. The standard ellipse is a descriptive tool used to visualize the shape of the ellipse and its orientation. It is readily calculated and contains about 40% of the sample population. It is not dependent on the sample size and cannot be used for statistical inference. The other two ellipses have an identical shape and orientation but have different major and minor axes. The Hotelling's confidence ellipse covers the sample's center with a given accuracy and estimates accuracy. In addition, Hotelling's one-sample test can be used to determine if the bivariate sample mean is significantly different than zero. The tolerance ellipse is an estimate of precision. It is an estimate of the confidence (e.g., 95%) of a percentage of the population sampled (e.g., 95%) that is enclosed in the tolerance ellipse. Or in other words, you are 95% confident that 95% of your points fall within the ellipse (Chew 1966). The ellipses are shown in Figure 1, representing the test data; the methods for calculating them and Hotelling's one-sample test are described in the Appendix.

This elliptical method of describing spatial accuracy and precision is valid because (1) it is intuitive, (2) it is understandable graphically, (3) it is statistically sound, (4) it is sufficient (i.e., it describes all relevant parameters), and (5) it is concise.

Methods

Study Site

The study included Navigation Pools 7 and 8 of the Upper Mississippi River near La Crosse, Wisconsin (Fig. 2). This reach of the river is a complex of islands, channels, and backwaters surrounded by 500-ft-high bluffs.

Forty-one surveyed ground locations were used as test points. Thirty-six third-order points were surveyed by the U.S. Army Corps of Engineers and were distributed along the main channel of the river. Five first-order points that were located at road intersections in the floodplain were surveyed by the Wisconsin and Minnesota Departments of Highways. All survey points were tied to the U.S. Geological Survey High Accuracy Reference Network. These points were surveyed for purposes unrelated to this study and should not have introduced bias into the sample. A third-order point has a relative accuracy of at least one part in 5,000 between directly connected points. This means that the point is at least within 1 m of its true location if the previous point in the survey was 5,000 m away. A first-order point has an accuracy of one part in 100,000 (Brinker and Wolf 1984).

A Trimble Pathfinder Basis Plus GPS receiver was used in the field as the rover. Global Positioning System measurements were taken with the rover on June 8, 9, and 17, 1994. The PDOP mask was set to 5. A mask is a screening value that eliminates values higher or lower than the mask. For example, a PDOP mask of 5 eliminates all readings above 5, whereas a horizon mask of 10 eliminates all readings received from satellites less than 10^o above the horizon. The receiver was set to take one position per second and the horizon mask was set at 10^o. The rover elevation mask angle of 10^o was considered adequate because the sample locations were of the base (the manufacturer recommends a base station mask of 10^o and a rover mask of 15^o). The unit received signals until about 200 readings were taken at each point (180 readings is the minimum recommended by Trimble), usually about 5 min at each site. The receiver was in the 3D position fix mode, which provides more accurate X and Y coordinates than the 2D position fix mode.

Survey measurements were recorded in the State Plane Coordinate System, Minnesota South zone, and converted to the Universal Transverse Meter (UTM) projection, zone 15. The GPS measurements were taken in WGS-84 and converted to the UTM projection, zone 15, NAD27.

A Trimble Pathfinder Professional receiver was established at a surveyed point to serve as the base station. The base was set to take positions once every 5 s, and had a PDOP mask of 5 and a horizon mask of 10^o. The measurements were postprocessed to introduce differential correction by using Trimble's PFinder software on a personal computer. The software uses the code phase position correction method.

The X (easting) and Y (northing) differences between the survey GPS measurements were tested to determine if their distribution was

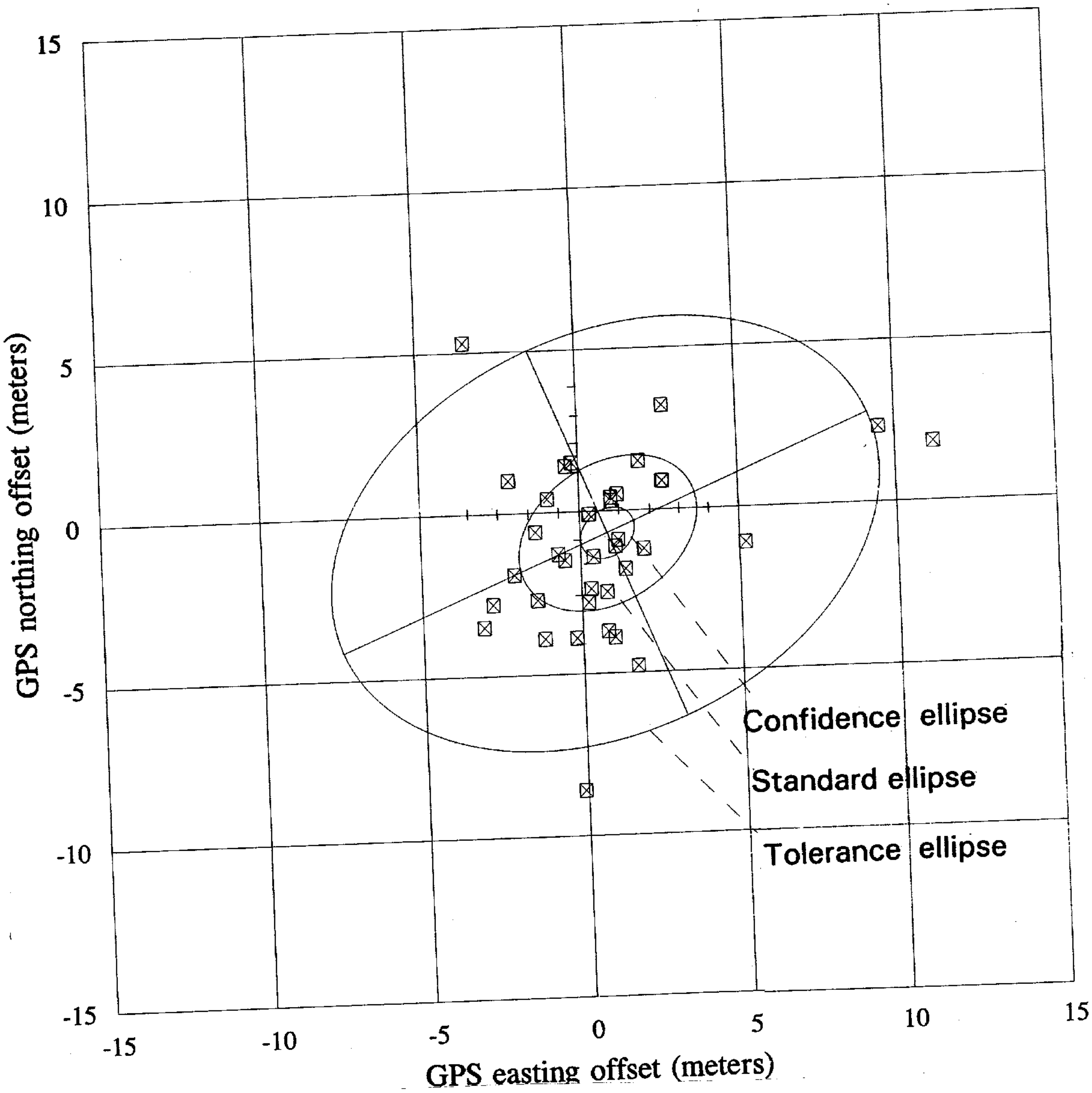


Figure 1. Distribution of distances between the Global Positioning System (GPS) coordinates and survey coordinates, and the ellipses, which describe the sample, accuracy, and probability distribution function.

Figure 2. Location of the study area. L&D = lock and dam.

normal using the chi-square (χ^2) test (Zar 1984). If the easting and northing differences were normally distributed, it is possible to use bivariate normal tests to test for systematic bias from zero and to describe the PDF with ellipses (Batschelet 1981). Hotelling's confidence ellipse, the standard ellipse, and the tolerance ellipse (Chew 1966) were calculated to described the PDF (the Appendix contains detailed descriptions of the equations and calculations used in the study). Hotelling's one-sample test was run to determine if the sample average was significantly different than zero.

Results and Discussion

Thirty-five usable locations were differentially corrected from the 41 locations measured (Fig. 3). Five of the original locations were not usable because the base station and rover did not receive positions from the same satellites and so the GPS readings were not corrected during differential postprocessing (possibly due to setting the mask angle to 10° on both the rover and the base station). One surveyed point had incorrect coordinates and was discarded. The results are presented in Table 1. It is apparent from Table 1 that some sample points had significantly less than 200 readings corrected. Three sample points had less than 10 readings corrected. This is less than the recommended amount and would seem to have a detrimental effect on accuracy (Fig. 4).

The average distance from the survey points was -1.13 m in the northing (Y) direction and 0.15 m in the easting (X) direction.

The easting and northing distances were tested for normality. H_0 (that the distributions were normal) was not rejected at the 0.05 level. More formally: H_0: This sample came from a normal distribution; H_A: This sample did not come from a normal distribution. Thus, we could calculate the descriptive and statistical ellipses and test to determine if there was a significant difference from zero for the mean distance.

The standard ellipse was centered on the sample means of (0.15, -1.13), and had a major axis of 2.60 m and a minor axis of 1.77 m with an angle from the X axis of 87.74° (Fig. 1). The confidence ellipse had a major axis of 1.12 m and a minor axis of 0.77 m, with the same angle of 87.74°. Hotelling's one-sample test was used to determine if the sample means were significantly different than zero. The result was that H_0 (that the population means equal zero) was rejected at the 0.05 level. The tolerance ellipse, which was calculated to contain 95% of the population at the 95% confidence level, had a major axis of 7.49 m and a minor axis of 5.12 m. The ellipses can be seen in Figure 1.

The confidence ellipse does not encompass the origin (0, 0), which graphically shows that the sample mean is different than zero. Thus, we can say that there is a systematic error in the sample and that it is offset in the south and east directions. We can also say that we are 95% confident that 95% of the population we sampled is contained in an ellipse that is centered on 0.15, -1.13 and has a major axis of 7.49 m and a minor axis of 5.12 m with an angle of 87.74° at the 0.05 level.

August et al. (1994) stated that 50 or more replicates markedly improves accuracy and that significant day-to-day variation in the accuracy existed. These statements were tested with the sample data. The euclidian distance between the GPS coordinate and the survey coordinate for each sample point was calculated (where euclidian distance is the straight line distance between two points and is the square root of the sum of the squared distances along the X and Y axes) and used as a measure of accuracy. A simple linear regression was calculated with the euclidian distance as the dependent variable and the number of points at each sample location as the independent variable. The resultant regression equation was the following:

$$Y = 4.079 + -0.061X$$

with a coefficient of determination R^2 of 0.031, which is a measure of the portion of total

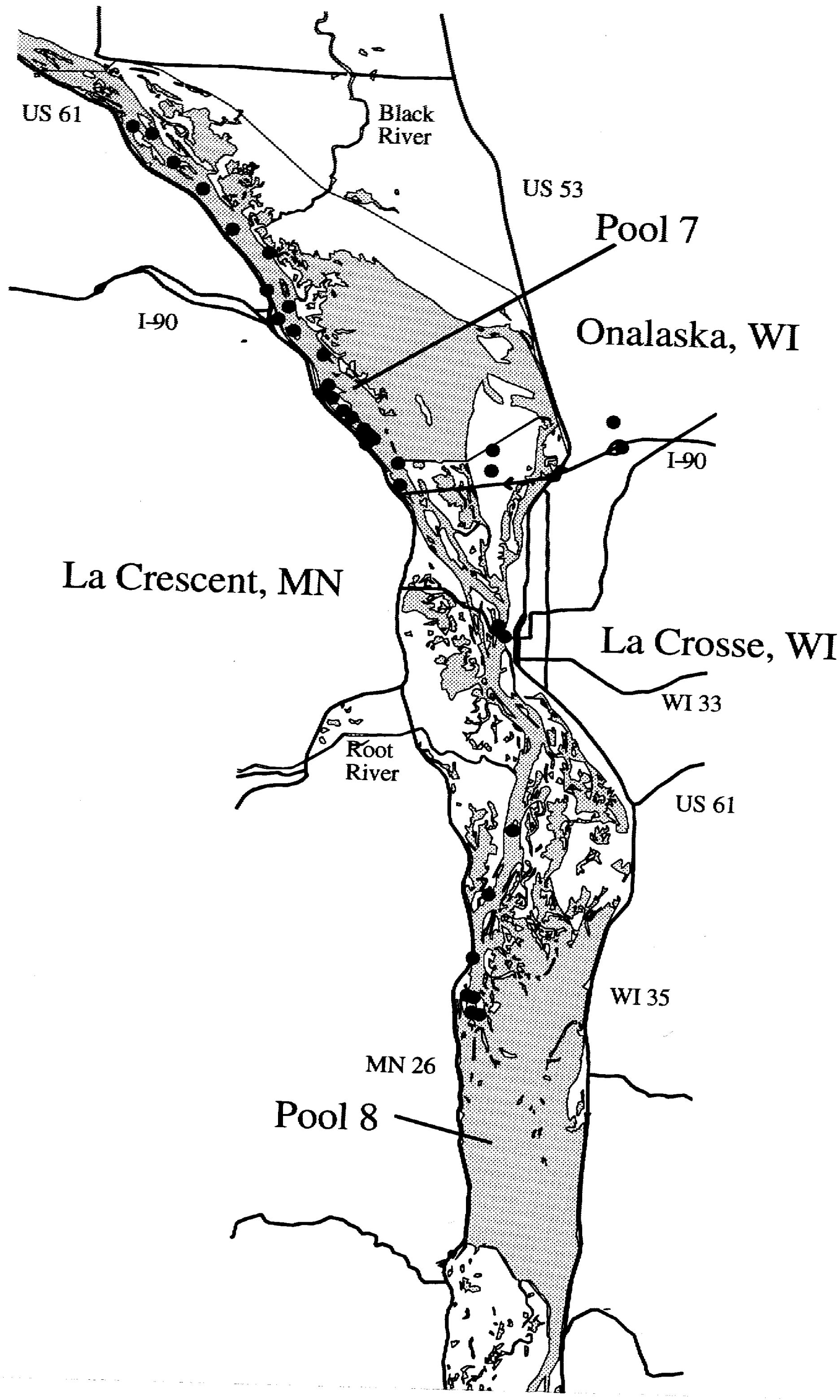


Figure 3. Location of Global Positioning System sample points.

Table 1. Coordinates of survey and Global Positioning System measurements.

Sample ID[a]	Survey N[b]	Survey E[c]	GPS N[d]	GPS E[e]	Number of points[f]	Diff N[g]	Diff E[h]
6-7-94	4,859,230.0	644,048.2	4,859,229.5	644,047.0	199	0.5	1.2
C060814A	4,858,379.5	636,135.4	4,858,384.3	636,133.7	97	-4.8	1.7
C060815C	4,859,275.3	635,251.5	4,859,273.9	635,251.9	211	1.4	-0.4
C060815D	4,859,085.9	634,984.5	4,859,089.6	634,983.7	31	-3.7	0.8
C060815E	4,859,431.0	635,076.8	4,859,427.8	635,074.1	65	3.2	2.7
C060815F	4,859,649.2	634,880.5	4,859,650.5	634,881.2	89	-1.3	-0.7
C060816A	4,860,059.5	634,440.8	4,860,063.4	634,442.0	192	-3.9	-1.2
C060816B	4,860,320.0	634,122.0	4,860,319.7	634,121.0	18	0.3	1.0
C060816C	4,860,783.3	633,716.9	4,860,781.8	633,715.0	108	1.5	1.9
C060816C	4,858,203.5	639,455.7	4,858,206.2	639,457.1	213	-2.7	-1.4
C060816D	4,860,937.9	633,383.0	4,860,936.4	633,383.2	202	1.5	-0.2
C060816E	4,861,266.3	633,564.3	4,861,269.1	633,567.1	6	-2.8	-2.8
C060817B	4,862,391.5	633,353.6	4,862,393.9	633,353.3	137	-2.4	0.3
C060818C	4,863,273.4	632,276.3	4,863,275.2	632,274.9	197	-1.8	1.4
C060818D	4,864,183.3	632,063.9	4,864,184.4	632,062.8	152	-1.1	1.1
C060819A	4,864,778.4	631,268.1	4,864,778.0	631,269.1	199	0.4	-1.0
C060819B	4,866,196.8	631,304.8	4,866,198.0	631,302.8	186	-1.2	2.0
C060819C	4,867,038.2	629,982.2	4,867,040.1	629,984.3	49	-1.9	-2.1
C060819D	4,868,511.4	628,882.8	4,868,511.0	628,881.8	163	0.4	1.0
C060820A	4,869,470.5	627,815.1	4,869,471.4	627,813.9	204	-0.9	1.2
C060820B	4,870,533.0	627,021.8	4,870,541.6	627,021.9	196	-8.6	-0.1
C060916A	4,860,187.8	643,713.6	4,860,190.6	643,713.4	5	-2.8	0.2
C060916B	4,858,982.4	639,482.6	4,858,983.5	639,477.4	8	-1.1	5.2
C060916C	4,870,785.4	626,330.0	4,870,784.5	626,327.5	125	0.9	2.5
C060917A	4,857,562.2	636,223.2	4,857,559.9	636,223.7	48	2.3	0.5
C060917B	4,838,307.4	639,225.1	4,838,311.3	639,224.1	152	-3.9	1.0
C060918A	4,852,335.8	639,895.9	4,852,336.4	639,897.3	55	-0.6	-1.4
C060918B	4,851,924.9	640,129.0	4,851,927.4	640,128.2	179	-2.5	0.8
C062113A	4,837,588.9	639,706.6	4,837,589.0	639,706.3	195	-0.1	0.3
C062113B	4,837,665.2	639,401.0	4,837,669.1	639,401.2	220	-3.9	-0.2
C062113C	4,838,235.7	639,474.1	4,838,239.2	639,477.2	63	-3.5	-3.1
C062114A	4,839,710.6	639,400.1	4,839,712.0	639,399.7	198	-1.4	0.4
C062114D	4,842,116.9	639,875.0	4,842,115.9	639,877.2	196	1.0	-2.2
C062114E	4,844,581.9	640,649.0	4,844,583.4	640,649.5	145	-1.5	-0.5
C062115C	4,851,189.9	640,048.4	4,851,184.6	640,051.9	72	5.3	-3.5
Sample mean						-1.13	0.15
Sample SD					2.56	1.75	

[a]Sample ID = The identification number of the location.
[b]Survey N = The northing Universal Transverse Mercator (UTM) coordinate (in meters) from survey.
[c]Survey E = The easting UTM coordinate (in meters) from survey.
[d]GPS N = The northing UTM coordinate from Global Positioning System (GPS) measurement.
[e]GPS E = The easting UTM coordinate from GPS measurement.
[f]Number of points = Number of points corrected in the differential calculation.
[g]Diff N = The difference in meters between the survey northing coordinate and the GPS northing coordinate.
[h]Diff E = The difference in meters between the survey easting coordinate and the GPS easting coordinate.

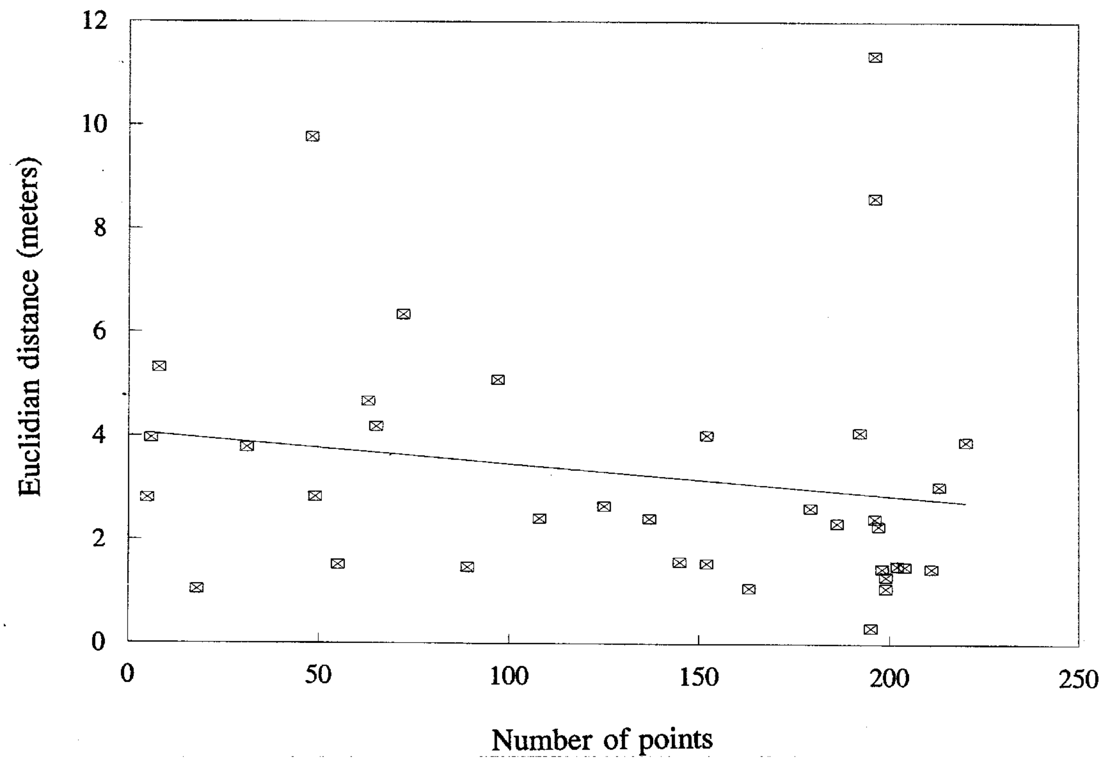


Figure 4. Regression of the number of sample points taken at each Global Positioning System location versus euclidian distance ($R^2 = 0.031$).

variation that is explained by the relation (where an R^2 of 1 is a perfect relation and an R^2 of 0 is no relation). The Y in this equation is the euclidian distance for each sample distance, and X is the number of points taken at each GPS location. The regression is shown in Figure 4. The slightly negative slope suggests a negative relation between the number of points and the euclidian distance. We would expect greater accuracy at more points. However, visual inspection of the distribution of points shows no apparent pattern; locations that had only a few points were 1 m away, whereas locations with 200 points may have been as far as 11 m away. The R^2 of 0.031 indicates that little of the variability in the euclidian distance is explained by the number of points. We tested the relation between the euclidean distance and the number of points (H_0: $\beta = 0$, where β is the regression coefficient and indicates that the euclidian distance has no dependence on the number of sample points taken at each GPS location), where the t statistic is the following:

$$t = (-0.0061 - 0)/0.0058 = -1.05$$

with 33 (n-2) degrees of freedom. At the 0.05 level and 33 df, the critical level is $t_{0.05,\ (2)\ 34} = 2.035$, and so H_0 is not rejected; there is no significant relation.

The data were pooled over several days so that enough sample points were taken to produce statistically significant results. However, since the GPS points were taken over a period of 3 days, we tested to see if there was an effect caused by different days. Table 2 summarizes the effects of different days.

Table 3 shows the results of using Hotelling's test for the coordinate matrix $\boldsymbol{\mu} = (\mu_x\ \mu_y)$ H_0 **:** $\boldsymbol{\mu} = \mathbf{0}$.

We can see from these tests that on June 8 there was a systematic error to the south and east, while on June 9 and 21 there was no bias. Although the sample size was small for the last two dates (seven samples each day), this result indicates that there was a systematic effect that occurred on one day.

Table 2. The effects of different sample days.

Date	**Number of GPS[a] samples taken each day**	**Average Diff N[b]**	**Average Diff E[c]**
6/08/94	21	-1.33	0.30
6/09/94	7	-1.10	1.26
6/21/94	7	-0.59	-1.26
All points	35	-1.13	0.18

[a]GPS = Global Positioning System.

[b]Diff N = The difference in meters between the survey northing coordinate and the GPS northing coordinate.

[c]Diff E = The difference in meters between the survey easting coordinate and the GPS coordinate.

Table 3. Testing for the effects of different sample days.

Date	Number of GPS[a] points taken each day	T^2 calc	T^2 critical	H_0: μ = 0
6/08/94	21	13.45	6.7	reject, bias in S and E
6/09/94	7	4.61	13.9	do not reject, no bias
6/21/94	7	6.67	13.0	do not reject, no bias

[a]GPS = Global Positioning System

Conclusion

The purpose of our study was to apply a method for measuring the spatial accuracy of point data. The method used was Hotelling's one-sample test to make inferences about the sample data's means and to use tolerance ellipse to describe the PDF of the population. The confidence ellipse is an intuitive graphic that shows whether systemic error exists in the sample. This method is theoretically sound and provides intuitive and concise statistics about a population. We can now describe the accuracy of the sample data generated by GPS described in the study as having a mean that is 0.18, -1.13 m and has systematic error in the east and south directions. We can further state that 95% of the population is contained in an ellipse (at the 95% confidence level) with a center at 0.18, -1.13 m, having a major axis of 7.49 m, a minor axis of 5.12 m, and an angle of translation from the horizontal of 87.74°.

This method for measuring and describing the spatial accuracy of point data is useful for three reasons. First, the method is repeatable; similar results will be obtained for a given set of points when the method is repeated and accuracy statements can be compared for different sets of data. Second, it is statistically valid; this method provides a probability statement about the accuracy statement. Third, it is sufficient. Merely stating that the GPS-derived points are "within 5 m of their true location" is not sufficient because no probability statement is associated with the accuracy statement. In addition, *X* and *Y* coordinates are related, but do not vary in a perfect one-to-one relation; it is necessary to describe how both *X* and *Y* vary, along with the confidence ellipse, which describes the accuracy, and the tolerance ellipse, which describes the precision.

We have proposed a method for testing the accuracy and precision of specific GPS arrangement. This method could be applied to other systems to test their accuracy and precision. Once the accuracy and precision parameters of a particular system are known—independent of manufacturer claims—the system can be used with confidence for collecting location data.

Accuracy requirements are not rigid for biological applications. The particular application will drive the accuracy needs. As a general rule, the accuracy of location data must be sufficient so there is no ambiguity in its application; for example, if you need to locate a nest site in a vegetation patch that is 1 mi square, an accuracy within 100 m is probably sufficient, whereas if you need to locate a sample site within a half-acre patch, accuracy on the order of 1-2 m is necessary.

The value of the study was that it applies a statistically valid method of estimating accuracy. This method is useful for any scale or resolution of data; whether the data are global in extent or cover 1 ha, the method of measuring accuracy is still applicable.

Acknowledgment

We thank S. Gutreuter at the Environmental Management Technical Center, Onalaska, Wisconsin, for his statistical guidance.

References

August, P., J. Michaud, C. Labash, and C. Smith. 1994. GPS for environmental applications: Accuracy and precision of locational data. Photogrammetric Engineering and Remote Sensing 60(1):41-45.

Batschelet, E. 1981. Circular statistics in biology. Academic Press, New York. 310 pp.

Brinker, R., and P. Wolf. 1984. Elementary surveying. Harper & Row, New York. 450 pp.

Bolstad, P., P Gessler, and T. Lillesand. 1990. Positional uncertainty in manually digitized map data. International Journal Geographic Information Systems 4:399-412.

Chew, V. 1966. Confidence, predictive, and tolerance regions for the multivariate normal distribution. American Statistical Association Journal 61:605-617.

Goodchild, M., and S. Gopal. 1989. Preface. Pages xi-xv *in* M. Goodchild and S. Gopel, editors. National Center for Geographic Information and Analysis. Accuracy of Spatial Database. Taylor & Francis, New York.

Hurn, J. 1993. Differential GPS explained. Trimble Navigation Ltd., Sunnyvale, California. 55 pp.

Thapa, K., and J. Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogrammetric Engineering and Remote Sensing 58:835-841.

U.S. Fish and Wildlife Service. 1993. Operating Plan for the Upper Mississippi River System Long Term Resource Monitoring Program. Environmental Management Technical Center, Onalaska, Wisconsin, Revised September 1993. EMTC 91-P002R. 179 pp. (NTIS #PB94-160199)

Zar, J. 1984. Biostatistical analysis. Prentice Hall, Englewood Cliffs, New Jersey. 620 pp.

Appendix. Equations and Calculations

Chi-Square Goodness-of-Fit Test for Normality

The χ^2 goodness-of-fit test for normality (Zar 1984) compares the expected versus observed frequencies of the X and Y distances between the Global Positioning System and survey points. The distances are put into classes. For a normally distributed population, there are expected frequencies that can be calculated for each class. If the expected and actual frequencies are similar, H_0 (the distribution is normal) is not rejected.

Formally, the hypothesis that we are testing is as follows:

H_0: The data are normally distributed
H_A: The data are not normally distributed

First, the sample mean and the sample standard deviation for X and Y are calculated as estimates of the population mean and standard deviation. These are presented at the bottom of Table A-1. The rest of the calculations are shown in Table A-1 for the northing (Y) differences. The χ^2 statistic's degrees of freedom is the number of classes (k) minus the number of constants necessary to calculate the statistic. Three constants ($\bar{X}$, the sample standard deviation, and the sample size) were used, so the degrees of freedom is 15-3 = 12. Consulting a table of the critical values for the χ^2 distribution, we find that the calculated χ^2 of 7.49 for 12 df is $0.75 < P < 0.90$. This tells us that we would expect to get results which deviate from expected frequencies 75% to 90% of the time by chance alone when H_0 is true, so we would not reject H_0 here and thus we conclude that the sample is normally distributed.

Table A-1. Calculations for the χ^2 goodness-of-fit test for the northing distances.

Diff class[a]	**Diff(x)**[b]	**Obs freq (f_i)**[c]	***Z* calc**[d]	***Z* table**[e]	***P*(x)**[f]	**Exp freq (F_i)**[g]	**$(f\text{-}F)^2/F$**[h]
< -6.5		1			0.017	0.60	0.28
-6.5 to -5.5	-6	0	-2.12	0.017	0.0248	0.87	0.87
-5.5 to -4.5	-5	1	-1.73	0.0418	0.0483	1.69	0.28
-4.5 to -3.5	-4	5	-1.34	0.0901	0.081	2.84	1.65
-3.5 to -2.5	-3	4	-0.95	0.1711	0.1166	4.08	0.00
-2.5 to -1.5	-2	4	-0.56	0.2877	0.1409	4.93	0.18
-1.5 to -0.5	-1	7	-0.18	0.4286	0.1546	5.41	0.47
-0.5 to 0.5	0	5	0.21	0.4168	0.1425	4.99	0.00
0.5 to 1.5	1	5	0.60	0.2743	0.1132	3.96	0.27
1.5 to 2.5	2	1	0.99	0.1611	0.0773	2.71	1.08
2.5 to 3.5	3	1	1.38	0.0838	0.0454	1.59	0.22
3.5 to 4.5	4	0	1.77	0.0384	0.023	0.81	0.81
4.5 to 5.5	5	1	2.16	0.0154	0.01	0.35	1.21
5.5 to 6.5	6	0	2.55	0.0054	0.0038	0.13	0.13
> 6.5		0	2.94	0.0016	0.0016	0.06	0.06
Sum		35.00			1.00	35.00	7.49

Sample mean -1.13
Sample SD 2.59

[a]Diff class = The class range of the frequency distribution.

[b]Diff (x) = The midpoint of the frequency classes.

[c]Obs freq (f_i) = The number of observations that fall into each class or the observed frequency of that class.

[d]Z calc = The estimate of the Z value or normal deviate for that class. It is the low end of the range minus $\bar{X}$ divided by the standard deviation. It is the estimate used to derive the Z table value.

[e]Z table = The proportion of the normal curve that lies beyond the normal deviate for that class. For example, the calculated Z value for the class -6.5 to -5.5 is -2.12; the proportion that falls in that class is 0.017.

[f]$P(x)$ = The proportion of the normal curve that lies within that class. This is the absolute value of the difference between the Z table values in that class and the one below it.

[g]Exp freq (F_i) = The proportion of the normal curve in that class times the sample size n. This is called fitting the normal distribution to the sample.

[h]$(f\text{-}F)^2/F$ = The observed frequency minus the expected frequency squared divided by the expected frequency which, when summed over all the classes, is the χ^2 statistic used for the goodness-of-fit test.

The calculations to test the normality of the X or easting offsets sample are shown in Table A-2.

Table A-2. Calculations for the χ^2 goodness-of-fit test for the easting distances.

Diff class[a]	**Diff(*x*)**[b]	**Obs freq (f_i)**[c]	***Z* calc**[d]	***Z* table**[e]	***P*(*x*)**[f]	**Exp freq (F_i)**[g]	**$(f\text{-}F)^2/F$**[h]
< -3.5		0		0.0495	1.73	1.73	
-3.5 to -2.5	-3	3	-1.46	0.0721	0.0614	2.15	0.34
-2.5 to -1.5	-2	2	-1.11	0.1335	0.0872	3.05	0.36
-1.5 to -0.5	-1	6	-0.77	0.2207	0.1129	9.95	1.06
-0.5 to 0.5	0	8	-0.43	0.3336	0.1345	4.71	2.30
0.5 to 1.5	1	10	-0.08	0.4681	0.1859	6.51	1.88
1.5 to 2.5	2	4	0.26	0.3974	0.1265	4.43	0.04
2.5 to 3.5	3	1	0.61	0.2709	0.0998	3.49	1.78
> 3.5		1	0.95	0.1711	0.1423	4.98	3.18
Sum		35			1.00	35.00	12.67

Sample mean 0.18
Sample SD 2.59

[a]Diff class = The class range of the frequency distribution.

[b]Diff (*x*) = The midpoint of the frequency classes.

[c]Obs freq (f_i) = The number of observations that fall into each class or the observed frequency of that class.

[d]*Z* calc = The estimate of the *Z* value or normal deviate for that class. It is the low end of the range minus $\bar{X}$ divided by the standard deviation. It is the estimate used to derive the *Z* table value.

[e]*Z* table = The proportion of the normal curve that lies beyond the normal deviate for that class. For example, the calculated *Z* value for the class -6.5 to -5.5 is -2.12; the proportion that falls in that class is 0.017.

[f]*P*(*x*) = The proportion of the normal curve that lies within that class. This is the absolute value of the difference between the *Z* table values in that class and the one below it.

[g]Exp freq (F_i) = The proportion of the normal curve in that class times the sample size *n*. This is called fitting the normal distribution to the sample.

[h]$(f\text{-}F)^2/F$ = The observed frequency minus the expected frequency squared divided by the expected frequency which, when summed over all the classes, is the χ^2 statistic used for the goodness-of-fit test.

The χ^2 statistic's degrees of freedom are 9-3 = 6. Consulting a table of the critical values for the χ^2 distribution, we find that the calculated χ^2 of 12.67 for 6 degrees of freedom is $0.05 < P < 0.10$. This tells us that we would expect to get results which deviate from expected frequencies 5% to 10% of the time with H_0 being true, so we would not reject H_0 and thus we conclude that the sample is normally distributed.

Thus, we have shown that both X and Y are normally distributed, and that a bivariate sample from a normal population exists.

Standard Ellipse

The standard ellipse (Batschelet 1981) is a descriptive tool used to visualize the distribution and serves the same function as $\bar{x} \pm s$ does for univariate statistics. About 40% of the individual pairs are contained within the standard ellipse. Five statistics are required to construct the ellipse: $\bar{X}$, $\bar{Y}$, s_x^2, s_y^2, and r (r = Cov(X,Y)/$s_x s_y$ and Cov(X,Y) = $1/(n\text{-}1)\sum(X_i - \bar{X})(Y_i - \bar{Y})$; is also the correlation coefficient), which measures the joint behaviors of X and Y. The center of the ellipse is ($\bar{X}$, $\bar{Y}$) and the equation of the ellipse is as follows:

$$s_y^2(X - \bar{X})^2 - 2rs_xs_y(X - \bar{X})(Y - \bar{Y}) + s_x^2(Y - \bar{Y})^2 = (1 - r^2)s_y^2s_x^2.$$

This equation says that for any *X* coordinate on the ellipse, we can solve for the corresponding *Y* coordinate and vice versa. The first step is to calculate the correlation coefficient as shown in Table A-3.

Table A-3. Calculations for the correlation coefficient.

Sample ID[a]	**Difference easting (*X*)**	**Difference northing (*Y*)**	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$
6-7-94	0.5	1.2	0.46	1.55	0.72
C060814A	-4.8	1.7	0.96	-3.75	-3.60
C060815C	1.4	-0.4	-1.14	2.45	-2.79
C060815D	-3.7	0.8	0.06	-2.65	-0.16
C060815E	3.2	2.7	1.96	4.25	8.34
C060815F	-1.3	-0.7	-0.25	-1.44	0.36
C060816A	-3.9	-1.2	-1.94	-2.85	5.52
C060816B	0.3	1.0	0.26	1.35	0.35
C060816C	1.5	1.9	1.16	2.55	2.96
C060816C	-2.7	-1.4	-2.14	-1.65	3.52
C060816D	1.5	-0.2	-0.94	2.55	-2.40
C060816E	-2.8	-2.8	-3.54	-1.75	6.18
C060817B	-2.4	0.3	-0.44	-1.35	0.59
C060818C	-1.8	1.4	0.66	-0.75	-0.49
C060818D	-1.1	1.1	0.36	-0.05	-0.02
C060819A	0.4	-1.0	-1.74	1.45	-2.53
C060819B	-1.2	2.0	1.26	-0.15	-0.19
C060819C	-1.9	-2.1	-2.84	-0.85	2.41
C060819D	0.4	1.0	0.26	1.45	0.38
C060820A	-0.9	1.2	0.46	0.15	0.07
C060820B	-8.6	-0.1	-0.84	-7.55	6.33
C060916A	-2.8	0.2	-0.54	-1.75	0.94
C060916B	-1.1	5.2	4.46	-0.05	-0.21
C060916C	0.9	2.5	1.76	1.95	3.44
C060917A	2.3	0.5	0.32	3.43	1.09
C060917B	-3.9	1.0	0.26	-2.85	-0.74
C060918A	-0.6	-1.4	-2.14	0.45	-0.97
C060918B	-2.5	0.8	0.06	-1.45	-0.09
C062113A	-0.1	0.3	-0.44	0.95	-0.42
C062113B	-3.9	-0.2	-0.94	-2.85	2.67
C062113C	-3.5	-3.1	-3.84	-2.45	9.40
C062114A	-1.4	0.4	-0.34	-0.35	0.12
C062114D	1.0	-2.2	-2.94	2.05	-6.03
C062114E	-1.5	-0.5	-1.24	-0.45	0.55
C062115C	5.3	-3.5	-4.23	6.35	-26.87
Sample mean	-1.05	0.74		Sum	4.96
Sample SD	2.57	2.91		Cov(*X*,*Y*)	0.14
				Cor coef	0.03

[a]Sample ID = The identification number of the location.

For ease of calculation, some coefficients are shortened:

$$A = s_y^2,\ B = -rs_xs_y,$$

$$C = s_x^2\ ,\ \text{and}\ D_s = (1 - r^2)s^2_xs^2_y.$$

An ellipse is defined by a major axis (a), a minor axis (b), and an angle θ of translation, which is the angle that the major axis is offset from the X axis (and the angle the minor axis is offset from the Y axis).

Another coefficient must be calculated to arrive at the ellipse parameters:

$$R = [(A - C)^2 + 4B^2]^{1/2}.$$

Then we can calculate the ellipse parameters:

$$a = [2D/(A + C - R)]^{1/2},$$

$$b = [2D/(A + C + R)]^{1/2},\ \text{and}$$

$$\theta = \arctan[2B/(A - C - R)].$$

The coefficients are as follows:

$$A = s_y^2 = 6.73,\ B = -\mathrm{Cov}(X,Y) = -0.14,$$

$$C = s_x^2 = 3.15,\ D_s = (1 - r^2)s_xs_y = 21.18,$$

$$R = [(A - C)^2 + 4B^2]^{1/2} = 3.59,$$

$$a = [2D/(A + C - R)]^{1/2} = 2.60,$$

$$b = [2D/(A + C + R)]^{1/2} = 1.77,$$

$$\theta = \arctan[2B/(A - C - R)] = 1.53,$$

$$\text{and when converted from radians to degrees} = 87.74^o.$$

The ellipse is centered on -1.13 m on the Y-axis and 0.18 m on the X-axis, the major axis is 2.60 m, the minor axis is 1.77 m, and the angle of translation is 87.74°. The standard ellipse can be seen in Figure 1.

Hotelling's Confidence Ellipse

Hotelling's confidence ellipse (Batschelet 1981) is a statistical tool used for statistical inference. It serves the same function as $\bar{x} \pm tsn^{-1/2}$ does for univariate statistics. The confidence ellipse is a region that contains the population mean at a given probability with the center of the ellipse at ($\bar{X}$, $\bar{Y}$). The equation of the ellipse is as follows:

$$s_y^2(X - \bar{X})^2 - 2rs_xs_y(X - \bar{X})(Y - \bar{Y}) + s_x^2(Y - \bar{Y})^2 = (1 - r^2)s_y^2s_x^2\, n^{-1}T^2.$$

Note that the right side of the expression has an added term, T^2, which is based on the familiar F value for the univariate solution where

$$T^2 = 2[(n - 1)/(n - 2)]F_{2,n-2}.$$

$F_{\alpha/1,\, n-2}$ denotes the critical value from the F one-tailed distribution with n-2 degrees of freedom and a significance level of α. In our example, we have two variables, 36 samples, and are using $\alpha = 0.05$ (95% confidence level), so $F_{.05/1,\, 34} = 3.28$. Thus $T^2 = 6.75$.

The shortened coefficients are as follows:

$$A = s_y^2,\ B = -rs_xs_y,$$

$$C = s_x^2,\ D_C = (1 - r^2)s^2{}_x s^2{}_y n^{-1}T^2, \text{ and}$$

$$R = [(A - C)^2 + 4B^2]^{1/2}.$$

Again, the ellipse parameters are as follows:

$$a = [2D/(A + C - R)]^{1/2},$$

$$b = [2D/(A + C + R)]^{1/2},$$

$$\theta = \arctan[2B/(A - C - R)],$$

$$A = s_y^2 = 6.73,\ B = -\mathrm{Cov}(x,y) = -0.14,$$

$$C = s_x^2 = 3.15,\ D_s = (1 - r^2)s_xs_yn^{-1}T^2 = 3.97,$$

$$R = [(A - C)^2 + 4B^2]^{1/2} = 3.59,$$

$$a = [2D/(A + C - R)]^{1/2} = 1.12,$$

$$b = [2D/(A + C + R)]^{1/2} = 0.77,$$

$$\theta = \arctan[2B/(A - C - R)] = 1.53,$$

and when converted from radians to degrees = 87.74°.

The ellipse is centered on -1.13 m on the Y-axis and 0.18 m on the X-axis, the major axis is 1.12 m, the minor axis is 0.77 m, and the angle of translation is 87.74°. The standard ellipse can be seen in Figure 1. Note that the ellipse does not include the origin of the graph (0, 0). This graphically shows that a systematic difference exists between the sample and the expected result of no difference. This can be statistically tested using Hotelling's one-sample test.

Hotelling's test is based on the assumption that the sample is normal bivariate and tests whether the sample center deviates significantly from the origin. Thus, for the coordinate matrix $\boldsymbol{\mu} = (\mu_x\ \mu_y)$ $H_0 : \boldsymbol{\mu} = \mathbf{0}$ and $H_A : \boldsymbol{\mu} \neq \mathbf{0}$. The test statistic is T^2, which is equal to the following:

$$T^2 = n/(1 - r^2)(\bar{X}^2/s_x^2 - 2r\bar{Y}\bar{X}/s_x s_y + \bar{Y}/s_y^2).$$

For the given significance level, α, $F_{\alpha/1,\, n-2}$ can be found and the $T^2(\alpha)$ can be calculated by the same equation used above. Thus, T^2 for $\alpha = 0.05$; with two variables and 35 samples $T^2(\alpha)$ is 6.75 as before. The decision rule is that if $T^2 > T^2(\alpha)$, reject H_0. If $T^2 \leq T^2(\alpha)$, there is no reason to reject H_0. $T^2 = 7.37$, which is greater than the $T^2(\alpha)$ of 6.75, so we reject H_0 at $\alpha = 0.05$. Therefore, the sample center deviates significantly from zero (i.e., systematic bias exists).

Tolerance Ellipse

We now calculate the tolerance ellipse (Batschelet 1981), which allows us to make inferences about the population. The tolerance ellipse is a region that contains a given percent of the population at a given probability with the center of the ellipse at $(\bar{x}, \bar{y})$. The equation of the ellipse is

$$s_y^2(X - \bar{X})^2 - 2rs_x s_y(X - \bar{X})(Y - \bar{Y}) + s_x^2(Y - \bar{Y})^2 = (1 - r^2)s_y^2 s_x^2\, H.$$

The right side of the expression has the term H, which is based on the noncentral χ^2 distribution and is approximated by Table 2 in Chew (1966). Thus, at $\alpha = 0.05$, for $n = 35$, and $H = 8.34$ for 95% of the population. Using the same coefficients as above,

$$D_s = (1 - r^2)s_x^2 s_y^2 H = 176.44,$$

$$a = [2D/(A + C - R)]^{1/2} = 7.49,$$

$$b = [2D/(A + C + R)]^{1/2} = 5.17,$$

$$\theta = \arctan[2B/(A - C - R)] = 1.53,$$

and when converted from radians to degrees = 87.74°.

The ellipse is centered on -1.13 m on the Y-axis and 0.18 m on the X-axis, the major axis is 7.49 m, the minor axis is 5.17 m, and the angle of translation is 87.74°. The tolerance ellipse can be seen in Figure 1.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, D.C. 20503			
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 1996	3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE Geospatial Application: Estimating the spatial accuracy of coordinates collected using the Global Positioning System			5. FUNDING NUMBERS
6. AUTHOR(S) Thomas Owens[1] and David McConville[2]			
7. PERFORMING ORGANIZATION NAME AND ADDRESS [1]National Biological Service, Environmental Management Technical Center, 575 Lester Avenue , Onalaska, Wisconsin 54650 [2]Saint Mary's University of Minnesota, 700 Terrace Heights, Winona, Minnesota 55987-1399			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Biological Service Environmental Management Technical Center 575 Lester Avenue Onalaska, Wisconsin 54650			10. SPONSORING/MONITORING AGENCY REPORT NUMBER 96-T002
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Release unlimited. Available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161 (call toll free 1-800-553-6847)			12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Evaluating the accuracy of spatial data is important to determine the appropriate use of these data. However, a good method has not been documented to measure locational accuracy. The Global Positioning System (GPS) reduces the difficulty of measuring the location of objects and enables non-surveyors to determine their location with relative ease. This study applied a straight-forward, repeatable, and statistically sound method of estimating the horizontal accuracy of GPS-derived location data. We concentrated on the spatial accuracy of points because points represent simple locations and not cartographic abstractions such as lines or polygons.

When GPS coordinates are taken at surveyed locations, the quantity of interest is the difference from the surveyed (assumed true) coordinates. This difference in coordinates is a bivariate quantity and the probability distribution function (PDF) can be described by an ellipse with the center at $\bar{X}$ and $\bar{Y}$. An ellipse is an appropriate shape for a PDF; it has two dimensions but is not rectangular because the joint probability of points occurring in the corners is very small, and it is generally not circular because X and Y are not necessarily the same. There are three ellipses of interest: the standard ellipse, the confidence ellipse, and the tolerance ellipse. The standard ellipse is a descriptive tool used to visualize the ellipse's shape and orientation. It contains about 40% of the sample, is not dependent on the sample size, and cannot be used for statistical inference. The other two ellipses have identical shapes and orientation but different major and minor axes. The confidence ellipse is an estimate of accuracy; the sample mean is or is not significantly different from the survey locations at a given α. The tolerance ellipse is an estimate of precision; a given percentage of the population sampled is enclosed in the tolerance ellipse at a given α.

Thirty-six locations were measured and compared to surveyed locations. The average offset was -1.13 m in the northing (Y) direction and 0.18 m in the easting (X) direction. Hotelling's one-sample test determined that H_0 (no significant departure from the survey locations exists) was rejected at the 0.05 level, which indicates there was a systematic error in the sample in the south and east directions. Ninety-five percent of the population sampled (at the 0.05 level) was contained in an ellipse that was centered on 0.18, -1.13, and had a major axis of 7.49 m, and a minor axis of 5.12 m with an angle of 87.74°. Thus, if an additional point were taken, we are 95% confident that it would fall within this tolerance ellipse.

14. SUBJECT TERMS ellipse, Global Positioning System, locational accuracy, spatial accuracy			15. NUMBER OF PAGES 13 pp. + Appendix
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

The Long Term Resource Monitoring Program (LTRMP) for the Upper Mississippi River System was authorized under the Water Resources Development Act of 1986 as an element of the Environmental Management Program. The mission of the LTRMP is to provide river managers with information for maintaining the Upper Mississippi River System as a sustainable large river ecosystem given its multiple-use character. The LTRMP is a cooperative effort by the National Biological Service, the U.S. Army Corps of Engineers, and the States of Illinois, Iowa, Minnesota, Missouri, and Wisconsin.