A critique of the National Standard for Spatial Data Accuracy

Paul Zandbergen
Department of Geography
University of New Mexico
Warning: Statistics Ahead!

\[ s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \]

\[ X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \]

\[ \rho = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{X_i - \mu_X}{\sigma_X} \right) \left( \frac{Y_i - \mu_Y}{\sigma_Y} \right) \]
Spatial Data Quality Matters!
Plane took off on a wrong runway and crashed in flames because the pilot used an outdated airport map.

49 persons were killed
A U.K. woman sent her Mercedes flying into a river, trusting the car's optimistic GPS guidance instead of the road signs warning of impending doom. Matters were made worse as the river was swollen from recent heavy rains, which caused the vehicle to be swept some 200 meters downstream before the woman was able to escape.
Spatial Data Quality Standards

- *National Map Accuracy Standards* (NMAS)
- *National Standards for Spatial Data Accuracy* (NSSDA) of the Federal Geographic Data Committee (FGDC)
- *Guidelines for Digital Elevation Data* of the National Digital Elevation Program (NDEP)
- *Minimum Standard Detail Requirements* of the American Land Title Association (ALTA) and the American Congress on Surveying and Mapping (ACSM)
- Various standards and guidelines of the American Society for Photogrammetry and Remote Sensing (ASPRS)
National Map Accuracy Standards (1947)

- Based on the error resulting from paper maps
  - i.e. what is the error in a very thin line on a paper map?
- Standard:
  - 90% of well-defined points that are tested must fall within a specified tolerance
  - For map scales larger than 1:20,000, the NMAS horizontal tolerance is 1/30 inch, measured at publication scale.
  - For map scales of 1:20,000 or smaller, the NMAS horizontal tolerance is 1/50 inch, measured at publication scale.
- Typical statement:
  - “Data is estimated to be compliant with the National Map Accuracy Standards for 1:12,000, estimated +/- 33.3 feet”
National Standard for Spatial Data Accuracy (1998)

- Current standard for all spatial data in the US
- Adopted by the Federal Geographic Data Committee (FGDC)
- Foundation for most other recent standards
- Outlines a testing methodology
  - Minimum of 20 sample points
  - Distributed across quadrants, not too close together
  - Calculate Root Mean Square Error (RMSE)
  - Compute 95% confidence interval by multiplying RMSE with 1.960 for vertical error and 1.7308 for horizontal error
- Provides suggested reporting format

- NSSDA is a *procedural* standard ➔ does not define thresholds
Federal Geographic Data Committee
Geospatial Positioning Accuracy Standards
Part 3: National Standard for Spatial Data Accuracy

3.2 Testing Methodology And Reporting Requirements

3.2.1 Spatial Accuracy

The NSSDA uses root-mean-square error (RMSE) to estimate positional accuracy. RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points1.

Accuracy is reported in ground distances at the 95% confidence level. Accuracy reported at the 95% confidence level means that 95% of the positions in the dataset will have an error with respect to true ground position that is equal to or smaller than the reported accuracy value. The reported accuracy value reflects all uncertainties, including those introduced by geodetic control coordinates, compilation, and final computation of ground coordinate values in the product.
Accuracy Assessment
Accuracy Assessment

- true coordinates, $x_t$, $y_t$
- data coordinates, $x_d$, $y_d$
- error distance, $e$

$$e = \sqrt{(x_t - x_d)^2 + (y_t - y_d)^2}$$
Root Mean Square Error (RMSE)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e^2}
\]
Resources to Implement NSSDA

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point number</td>
<td>Point description</td>
<td>x (independent)</td>
<td>x (test)</td>
<td>diff in x</td>
<td>(diff in x)^2</td>
<td>y (independent)</td>
<td>y (test)</td>
<td>diff in y</td>
<td>(diff in y)^2</td>
<td>(diff in x)^2 + (diff in y)^2</td>
</tr>
</tbody>
</table>

| sum | average | RMSE | NSSDA |
Assumptions of NSSDA

• Data is free of systematic errors
• Errors of spatial data are normally distributed
• Errors in X and Y direction are similar and independent
• Spatial autocorrelation of error is limited

• Therefore:
  – RMSE is used as basis for calculating 95% confidence interval
  – A sample of 20 points is sufficient
  – Stratified random sampling is sufficient
Research Question

Are positional errors of spatial data normally distributed?

If not, what does that mean for accuracy metrics, like RMSE?

Should we revise the NSSDA?
Four Data Types Considered

• Autonomous GPS position fixes
  – Collected using a Trimble Geo XM unit, 1 second logging for 8-hours, random sample of 1,000 locations
  – Reference: surveyed benchmark

• Geocoding
  – Home addresses of public school students (163,886) geocoded against 1:5,000 street centerlines, random sample of 1,000 locations
  – Reference: parcel centroids

• TIGER Roads 2000
  – Intersections of TIGER road data (2000), random sample of 1,000 locations
  – Reference: 1-meter orthophotography

• Lidar elevation data
  – Processed bare earth lidar elevation points
  – Reference: field accuracy reports obtained using RTK GPS
GPS Position Fixes

Legend
- ▲ Surveyed Benchmark
- • GPS Position Fixes

Meters
Geocoded Addresses
TIGER Road Network Intersections
Lidar Elevation Data

LIDAR Elevation Error (m)
- < -0.30
- -0.30 - -0.15
- -0.15 - 0.00
- 0.00 - 0.15
- 0.15 - 0.30
- > 0.30

Neuse / Tar-Pamlico Basins
North Carolina Counties
# Positional Errors

<table>
<thead>
<tr>
<th>Error statistic</th>
<th>GPS (m)</th>
<th>Geocoding (m)</th>
<th>Roads (m)</th>
<th>LIDAR (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90th Percentile</td>
<td>4.478</td>
<td>68.145</td>
<td>74.734</td>
<td>0.293</td>
</tr>
<tr>
<td>95th Percentile</td>
<td>5.158</td>
<td>90.337</td>
<td>95.623</td>
<td>0.372</td>
</tr>
<tr>
<td>99th Percentile</td>
<td>6.439</td>
<td>194.596</td>
<td>167.045</td>
<td>0.663</td>
</tr>
<tr>
<td>90% RMSE</td>
<td>2.395</td>
<td>27.245</td>
<td>33.986</td>
<td>0.135</td>
</tr>
<tr>
<td>95% RMSE</td>
<td>2.574</td>
<td>31.925</td>
<td>38.370</td>
<td>0.151</td>
</tr>
<tr>
<td>100% RMSE</td>
<td>2.848</td>
<td>50.821</td>
<td>53.521</td>
<td>0.218</td>
</tr>
<tr>
<td>NSSDA</td>
<td>4.930</td>
<td>87.961</td>
<td>92.635</td>
<td>0.428</td>
</tr>
</tbody>
</table>
Normality Testing

• Examining the distribution
  – Skewness
  – Kurtosis

• Normality testing
  – Lillefors
  – Shapiro-Wilk

• Visualization
  – Histogram
  – Q-Q plot
GPS Locations – Horizontal (XY)
GPS Locations – X and Y Components

X component

Y component
GPS Error Approximates a Rayleigh Distribution
Geocoding Errors

![Histogram of Error Distribution](image1)

![Quantile-Quantile Plot](image2)
Geocoding Errors – Log Transformed
TIGER Roads Errors

![Histogram of Errors](left)

- Frequency vs. Error (m)

![Normality Test](right)

- Expected Normal vs. Observed Value
- Red line represents the expected normal distribution
TIGER Roads Errors – Log Transformed
Lidar Elevation Errors
bare earth and low grass

Expected Normal vs. Observed Value

brush lands and low trees

Expected Normal vs. Observed Value

fully forested

Expected Normal vs. Observed Value

urban areas

Expected Normal vs. Observed Value
Findings About Normality

- Positional errors are not normally distributed
- Some come close, e.g. X and Y components of GPS
- Some approximate log-normality
- Outliers and skew are the norm!
What Does This Mean for Determining the Accuracy of My Data?

• Test Protocol:
  – Use sample of 1,000 as the “true” distribution
  – Derive accuracy metrics from this distribution
  – Take a sample of 20 points, as per NSDDA protocol
  – Determine accuracy metrics
  – Repeat 100 times, i.e. 100 samples of 20 points
  – Determine robustness of metrics

• Metrics employed
  – RMSE, “trimmed” RMSE, 90th and 95th percentile
TIGER Roads – 1,000 points

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>38.5</td>
</tr>
<tr>
<td>Median</td>
<td>29.9</td>
</tr>
<tr>
<td>RMSE</td>
<td>53.5</td>
</tr>
<tr>
<td>90\textsuperscript{th}</td>
<td>95.6</td>
</tr>
<tr>
<td>NSSDA</td>
<td>92.6</td>
</tr>
</tbody>
</table>
# TIGER Roads – 20 points

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>24.2</td>
</tr>
<tr>
<td>Median</td>
<td>21.3</td>
</tr>
<tr>
<td>RMSE</td>
<td>30.1</td>
</tr>
<tr>
<td>90&lt;sup&gt;th&lt;/sup&gt;</td>
<td>44.7</td>
</tr>
<tr>
<td>NSSDA</td>
<td>52.2</td>
</tr>
</tbody>
</table>
### TIGER Roads – 20 points

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>40.0</td>
</tr>
<tr>
<td>Median</td>
<td>32.1</td>
</tr>
<tr>
<td>RMSE</td>
<td>57.7</td>
</tr>
<tr>
<td>90(^{th})</td>
<td>82.7</td>
</tr>
<tr>
<td>NSSDA</td>
<td>99.8</td>
</tr>
</tbody>
</table>
Distribution of NSSDA Error Statistic

“True” value 92.6 m

Outliers of 175 and 256 m
Comparing Error Statistics

• Are they unbiased?
  – Which metrics consistently over or underestimate?
  – Compare average value of metrics of 100 20-point samples to “true” metric for 1,000 points

• Are they reliable?
  – Which metrics reveal very high or very low variability?
  – Compare standard deviation of metrics of 100 20-point samples to average value of metrics
Non-normal data:
1. RMSE and percentiles poor
2. Trimmed RMSE best

Close-to-normal data:
1. Metrics close in performance
2. RMSE appears OK

<table>
<thead>
<tr>
<th></th>
<th>Avg/True Ratio</th>
<th>Rank</th>
<th>SDev/Avg Ratio</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.928</td>
<td>4</td>
<td>0.151</td>
<td>4</td>
</tr>
<tr>
<td>95th Percentile</td>
<td>0.927</td>
<td>5</td>
<td>0.156</td>
<td>5</td>
</tr>
<tr>
<td>90% RMSE</td>
<td>0.992</td>
<td>2</td>
<td>0.133</td>
<td>3</td>
</tr>
<tr>
<td>95% RMSE</td>
<td>0.996</td>
<td>1</td>
<td>0.126</td>
<td>2</td>
</tr>
<tr>
<td>100% RMSE</td>
<td>0.983</td>
<td>3</td>
<td>0.120</td>
<td>1</td>
</tr>
<tr>
<td><strong>Geocoding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.880</td>
<td>3</td>
<td>0.284</td>
<td>3</td>
</tr>
<tr>
<td>95th Percentile</td>
<td>0.857</td>
<td>5</td>
<td>0.344</td>
<td>4</td>
</tr>
<tr>
<td>90% RMSE</td>
<td>0.992</td>
<td>2</td>
<td>0.253</td>
<td>1</td>
</tr>
<tr>
<td>95% RMSE</td>
<td>1.002</td>
<td>1</td>
<td>0.254</td>
<td>2</td>
</tr>
<tr>
<td>100% RMSE</td>
<td>0.870</td>
<td>4</td>
<td>0.433</td>
<td>5</td>
</tr>
<tr>
<td><strong>Roads</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.881</td>
<td>4</td>
<td>0.258</td>
<td>3</td>
</tr>
<tr>
<td>95th Percentile</td>
<td>0.882</td>
<td>5</td>
<td>0.342</td>
<td>5</td>
</tr>
<tr>
<td>90% RMSE</td>
<td>1.013</td>
<td>1</td>
<td>0.167</td>
<td>1</td>
</tr>
<tr>
<td>95% RMSE</td>
<td>1.014</td>
<td>2</td>
<td>0.197</td>
<td>2</td>
</tr>
<tr>
<td>100% RMSE</td>
<td>0.926</td>
<td>3</td>
<td>0.307</td>
<td>4</td>
</tr>
<tr>
<td><strong>LIDAR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.905</td>
<td>3</td>
<td>0.242</td>
<td>3</td>
</tr>
<tr>
<td>95th Percentile</td>
<td>0.860</td>
<td>5</td>
<td>0.285</td>
<td>4</td>
</tr>
<tr>
<td>90% RMSE</td>
<td>0.989</td>
<td>2</td>
<td>0.216</td>
<td>2</td>
</tr>
<tr>
<td>95% RMSE</td>
<td>0.991</td>
<td>1</td>
<td>0.215</td>
<td>1</td>
</tr>
<tr>
<td>100% RMSE</td>
<td>0.900</td>
<td>4</td>
<td>0.474</td>
<td>5</td>
</tr>
</tbody>
</table>
Effect of Outliers and Trimming

![Graph showing the effect of outliers and trimming on RMSE for different land cover types. The graph plots RMSE (meters) against the percentile of the error distribution (%). The y-axis ranges from 0.00 to 0.30, and the x-axis ranges from 75 to 100. The graph includes lines for bare earth and low grass, high grass, weeds and crops, shrubs and low trees, fully forested, and urban areas.]
What Metrics to Use?

- RMSE on complete dataset performs poorly
- Percentiles (90th, 95th) perform even worse
- “Trimmed” RMSE values perform better
- Alternative: median
General Recommendations

• Never assume your data is normal!
  – Expect outliers
  – Expect skew, log-normal behavior etc.
  – Test for normality if sample size allows

• Increase sample size
  – Still effect of outliers on RMSE
  – Metrics become more robust

• Properly characterize the error distribution
  – Multiple metrics
  – Curves/graphs

• Do not just fill out the NSSDA worksheet without examining your data!
How To Improve NSSDA

1. Embrace alternative approaches to error characterization
2. Do not calculate the 95% confidence interval from RMSE
3. Increase minimum sample size
4. Consider multiple error components (X, Y, vertical)
5. Consider some common alternative distributions
6. Data trimming might be one reasonable alternative
7. Expand beyond point features – add line and area
Final Take Home Lessons

• Positional errors in spatial data do not behave normally!

• NSSDA is still **THE** standard we should all be using

• When using the NSSDA protocol, carefully examine your data

• Learn from others who are working with the same data types
Further Reading....


Contact Information

Paul Zandbergen
Department of Geography
University of New Mexico

zandberg@unm.edu

www.paulzandbergen.com