An Accuracy Assessment of SAMB Elevation Data

Evan J Fedorko
West Virginia GIS Technical Center
20 April 2005

As a check of SAMB elevation point and breakline data quality, we compared the raw elevation points to LIDAR DEM surfaces generated in Gilmer County, West Virginia. We based this accuracy assessment on the statistical method for estimating the accuracy of spatial data based on ground positions of higher accuracy as outlined by the National Standards for Spatial Data Accuracy (FGCS 1998).

The NSSDA provides a statistical methodology to compute a vertical accuracy statistic for a dataset. First, we selected two study areas, both in Gilmer County, West Virginia. In YEAR, LIDAR elevation data was collected and processed to produce tiled DEMs. We used eleven tiles of new elevation data and twelve LIDAR DEM tiles for the comparison. First, using the ET Geowizards toolset, we converted the POINT ZM type shapefiles to POINT type shapefiles which contained an attribute of the elevation (feet) in the table. We then projected these from West Virginia state plane south to UTM Zone 17 North, to match the LIDAR elevation data. It is required that the elevation points match the projection of the LIDAR DEMs for the next step to work properly. Using a Visual Basic script from ERSI's technical support website, we added the elevation from the underlying LIDAR DEMs to each point. Next, in order to weed out some of the potential error in the LIDAR data, we used DOQQs to digitize several bare earth or open field areas and took a subset from the elevation points using those polygons.

From here, we shifted to Microsoft Excel to complete the statistical analysis. We completed two statistical analyses, one for all data points in our test area, and one for only those data points that intersect bare earth or open fields. We calculated several statistics for this analysis, but only two warrant an explanation. First, we calculated the root mean square difference (RMSD). We use the word "difference" instead "error" because what we are measuring isn't really error so much as it is a statistical difference measured between two datasets. As we are making the assumption that LIDAR data is essentially more accurate than the SAMB point and breakline data, and not actually citing any conclusive accuracy of the LIDAR data, we feel it is necessary to make this distinction. The RMSD was used to calculate the NSSDA accuracy statistic. The root mean square statistic is calculated via the following equation:

$$\mathrm{RMSD}_z = \sqrt{\frac{\sum (X_{groundvalue_i} - X_{testvalue_i})^2}{n}}$$

Where

$groundvalue_i$: ground truth point of the ith point in the dataset

$testvalue_i$: test point of the ith point in the dataset

$\sum^2$: sum of the set of squared differences between the ground and test data

n : total number of test points

This value is then multiplied by 1.96, the NSSDA constant for a 95% confidence accuracy statistic. This number tells us, in map units, the expected plus or minus accuracy of the data (FGCS, 1988). The table below contains the statistics we calculated.

Test Group	*n*	Mean Error (m)	Mean Error (ft)	RMSD (m)	NSSDA (m)	NSSDA (ft)
ALL	9288	-0.6819	-2.2372	1.7714	3.4720	11.3910
BARE EARTH	463	-0.5822	-1.9101	1.0669	2.0912	6.8609

Our target accuracy is 10 feet. Our initial examination of error gave us pause as the returned NSSDA value is over 11 feet. We were concerned, however, that this level of dissimilarity between the two datasets could be a result of errors in the post processing of the LIDAR data associated with vegetation canopy returns. For this reason, we used a bare earth/open field subset. This returned a more satisfactory value of 6.9 feet. While this error is still somewhat large, it is within the desirable margin of error for an elevation dataset of this resolution.

REFERENCES

Federal Geodetic Control Subcommittee, Federal Geographic Data Committee. *Geospatial Positioning Accuracy Standards*. Report FGDC-STD-007.1-1998. FGDC, 1998.