Dewberry

8401 Arlington Boulevard
Fairfax, Virginia 22031-4666

703 849 0100
703 849 0118 fax
www.dewberry.com

Report of Vertical Accuracy Testing 2005 LiDAR Bare-Earth Dataset for Baltimore County, Maryland

Date: June 7, 2006

References: A — Part 3: *National Standard for Spatial Data Accuracy (NSSDA)*, "Geospatial Positioning Accuracy Standards," published by the Federal Geographic Data Committee (FGDC), 1998
B — Appendix A, *Guidance for Aerial Mapping and Surveying*, "Guidelines and Specifications for Flood Hazard Mapping Partners," published by the Federal Emergency Management Agency (FEMA), April 2003
C — *Guidelines for Digital Elevation Data*, Version 1.0, published by the National Digital Elevation Program (NDEP), May 10, 2004
D — *ASPRS Guidelines, Vertical Accuracy Reporting for Lidar Data*, published by the American Society for Photogrammetry and Remote Sensing (ASPRS), May 24, 2004
E — Appendix F, *Acceptance Criteria for DTMs and Contours*, "Quality Plan for Baltimore County," Version 1.2, 2005

Background

FEMA guidelines (reference B) recommend the survey of 20 QA/QC checkpoints in each major land cover categories representative of the area being tested, with a minimum of three land cover categories, i.e., a minimum of 60 checkpoints. References C and D also recommend a minimum of 60 checkpoints, with up to 100 points preferred. Baltimore County identified five land cover categories, with the goal of surveying 20 x 5 = 100 QA/QC checkpoints if sufficient checkpoints could be reasonably identified.

The LiDAR accuracy assessment for Baltimore County was performed in accordance with References C and D which assumes that LiDAR errors in some land cover categories may not follow a normal error distribution. This assessment was also performed in accordance with References A and B which assumes that LiDAR bare-earth datasets errors do follow a normal error distribution. Comparisons between the two methods help determine the degree to which *systematic errors* may exist in Baltimore County's five major land cover categories: (1) open terrain, (2) weeds and crops, (3) scrub and bushes, (4) forests, and (5) built-up areas. When a LiDAR bare-earth dataset passes testing by both methods, compared with criteria specified in Reference E, the dataset clearly passes all vertical accuracy testing criteria for a digital terrain model (DTM) suitable for generation of 2-ft contours in Baltimore County. The relevant criteria from Reference E are summarized in Table 1. Criteria in yellow refer to references A and B (NSSDA and FEMA); criteria in green refer to references C and D (NDEP and ASPRS).

Table 1 — DTM Acceptance Criteria from the Quality Plan for Baltimore County

Quantitative Criteria	Measure of Acceptability
$RMSE_z$ = NSSDA vertical accuracy statistic at 68% confidence level	0.60 ft for all land cover categories combined
$Accuracy_z$ = NSSDA vertical accuracy statistic at the 95% confidence level	1.19 ft ($RMSE_z$ x 1.9600) for all land cover categories combined
Fundamental Vertical Accuracy (FVA) in open terrain only = 95% confidence level	1.19 ft ($RMSE_z$ x 1.9600) for open terrain only
Supplemental Vertical Accuracy (SVA) in individual land cover categories = 95% confidence level	1.19 ft (based on 95th percentile per category; this is a target value only, not mandatory)
Consolidated Vertical Accuracy (CVA) in all land cover categories combined = 95% confidence level	1.19 ft (based on combined 95th percentile)

Vertical Accuracy Testing in Accordance with NDEP and ASPRS Procedures

References C and D specify the mandatory determination of Fundamental Vertical Accuracy (FVA) and the optional determination of Supplemental Vertical Accuracy (SVA) and Consolidated Vertical Accuracy (CVA). FVA determines how well the LiDAR sensor performed in category (1), open terrain, where all errors are presumed to be random; whereas SVA determines how well the vegetation removal algorithms worked in land cover categories (2), (3) and (4) where LiDAR elevations are often higher than surveyed elevations, and how much the LiDAR penetrated into asphalt in land cover category (5) where LiDAR elevations are often lower than surveyed elevations if acquired when asphalt is hot.

FVA is determined with check points located only in land cover category (1), open terrain (grass, dirt, sand, and/or rocks), where there is a very high probability that the LiDAR sensor will have detected the bare-earth ground surface and where random errors are expected to follow a normal error distribution. The FVA determines how well the calibrated LiDAR sensor performed. With a normal error distribution, the vertical accuracy at the 95% confidence level is computed as the vertical root mean square error ($RMSE_z$) of the checkpoints x 1.9600, as specified in Appendix 3-A of the National Standard for Spatial Data Accuracy (NSSDA), FGDC-STD-007.3-1998. For Baltimore County, the FVA standard is 1.19 feet (ft) at the 95% confidence level, equivalent to the $Accuracy_z$ required for 2 ft contours.

CVA is determined with all checkpoints in all land cover categories combined where there is a possibility that the LiDAR sensor and post-processing may yield elevation errors that do not follow a normal error distribution. CVA at the 95% confidence level equals the 95th percentile error for all checkpoints in all land cover categories combined. The CVA is accompanied by a listing of the 5% outliers that are larger than the 95th percentile; these are always the largest outliers that may depart from a normal error distribution.

SVA is determined separately for each individual land cover category, recognizing that the LiDAR sensor and post-processing may yield elevation errors that do not follow a normal error distribution, and where discrepancies can be used to identify the nature of systematic errors by land cover category. For each land cover category, the SVA at the 95% confidence level equals the 95th percentile error for all checkpoints in each individual land cover category. SVA statistics are calculated individually for open terrain, weeds and crops, scrub, forests, and built-up areas in order to facilitate the analysis of the data based on each of these land cover categories that exist within Baltimore County. The SVA criteria in Table 1 are target values only and are not mandatory; it is common for some SVA criteria to fail individual target values, yet satisfy the mandatory CVA criterion.

The primary Quality Assurance/Quality Control (QA/QC) steps used by Dewberry were as follows:

1. Dewberry’s LiDAR subcontractor (Sanborn Mapping) acquired the raw LiDAR data in the spring of 2005 and performed post-processing to derive the bare-earth digital terrain model (DTM). Sanborn also performed in-house QA/QC of its data.
2. Dewberry surveyed "ground truth" QA/QC vertical checkpoints in accordance with guidance in references A, B, C and D. Figure 1 shows the location of “cluster areas” where Dewberry attempted to survey two QA/QC checkpoints in each of the five land cover categories. Some cluster areas did not include all land cover categories. The final totals were 22 checkpoints in open terrain; 17 checkpoints in weeds and crops; 14 checkpoints in scrub; 18 checkpoints in forests; and 20 checkpoints in built up areas, for a total of 91 checkpoints.
3. Next, Dewberry interpolated the bare-earth LiDAR DTM to provide the z-value for each of these checkpoint coordinates.
4. Dewberry then computed the associated z-value differences between the interpolated z-value from the LiDAR data and the ground truth survey checkpoints and computed the FVA, CVA and SVA values.
5. The data were analyzed by Dewberry to assess the quantitative quality of the data. The review process examined the various accuracy parameters as defined by NDEP and ASPRS guidelines. Also, the

overall descriptive statistics of each dataset were computed to assess any trends or anomalies. The following tables, graphs and figures illustrate the data quality.

Figure 1 shows the location of the QA/QC checkpoint clusters within Baltimore County, symbolized to reflect the five land cover categories used. However, most of the symbols for the different land cover categories overlay each other and are not individually visible.

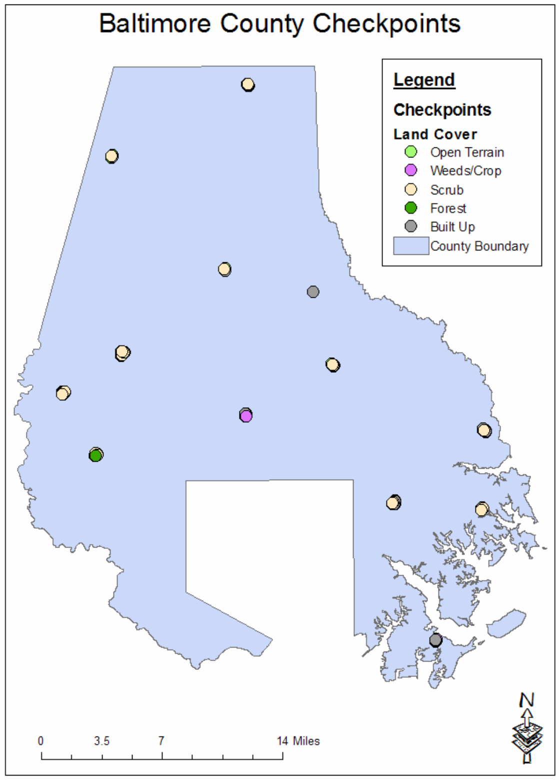


Figure 1 — Location of QA/QC Checkpoint Clusters

Table 2 summarizes the vertical accuracy by fundamental, consolidated and supplemental methods:

Table 2 — FVA, CVA and SVA Vertical Accuracy at 95% Confidence Level

Land Cover Category	# of Points	FVA Fundamental Vertical Accuracy Spec = 1.19 (ft)	CVA Consolidated Vertical Accuracy Spec = 1.19 (ft)	SVA Supplemental Vertical Accuracy Target = 1.19 (ft)
Total Combined	91		**0.76 ft**	
Open Terrain	22	**0.72 ft**		0.59 ft
Weeds/Crops	17			0.58 ft
Scrub	14			1.03 ft
Forest	18			0.54 ft
Built Up	20			0.78 ft

The LiDAR data of Baltimore County <u>meets the specifications</u> as per the following vertical accuracy tests.

Compared with the 1.19 ft FVA specification, FVA tested 0.72 ft at the 95% confidence level in open terrain, based on $RMSE_z$ x 1.9600. The NSSDA specifies that vertical accuracy at the 95% confidence level equals $RMSE_z$ x 1.9600; the NDEP and ASPRS state that this method is valid only when random errors follow a normal error distribution, as in open terrain.

Compared with the 1.19 ft CVA specification, CVA tested 0.76 ft at the 95% confidence level in open terrain, weeds and crops, scrub, forests, and built-up areas combined, based on the 95th Percentile. NDEP and ASPRS guidelines specify that vertical accuracy at the 95% confidence level equals the 95th percentile when random errors may not follow a normal error distribution, as in vegetated areas. Table 3 lists the 5% outliers larger than the 95th percentile (0.76 ft).

Table 3 — 5% Outliers Larger than 95th Percentile

Land Cover Category	Elev. Diff (ft)	
Open Terrain	0.77	Five points had errors larger than the 95th percentile error. However, there were no errors larger than the CVA standard (1.19 ft) which permits up to 5% of the checkpoints, normally 5 of 100, to be larger than 1.19 ft.
Scrub	-1.19	
Scrub	-0.95	
Built Up	0.78	
Built Up	0.85	

Compared with the 1.19 ft SVA target values, SVA tested 0.59 ft at the 95% confidence level in open terrain; 0.58 ft in weeds and crops; 1.03 ft in scrub; 0.54 ft in forests; and 0.78 ft in built-up areas, based on the 95th Percentile. These values exceed their target values. We did not survey the full 20 checkpoints desired in each land cover category, since these values exceeded the target value, Dewberry saw no need to survey a few additional checkpoints.

Figure 2 illustrates the SVA by specific land cover category.

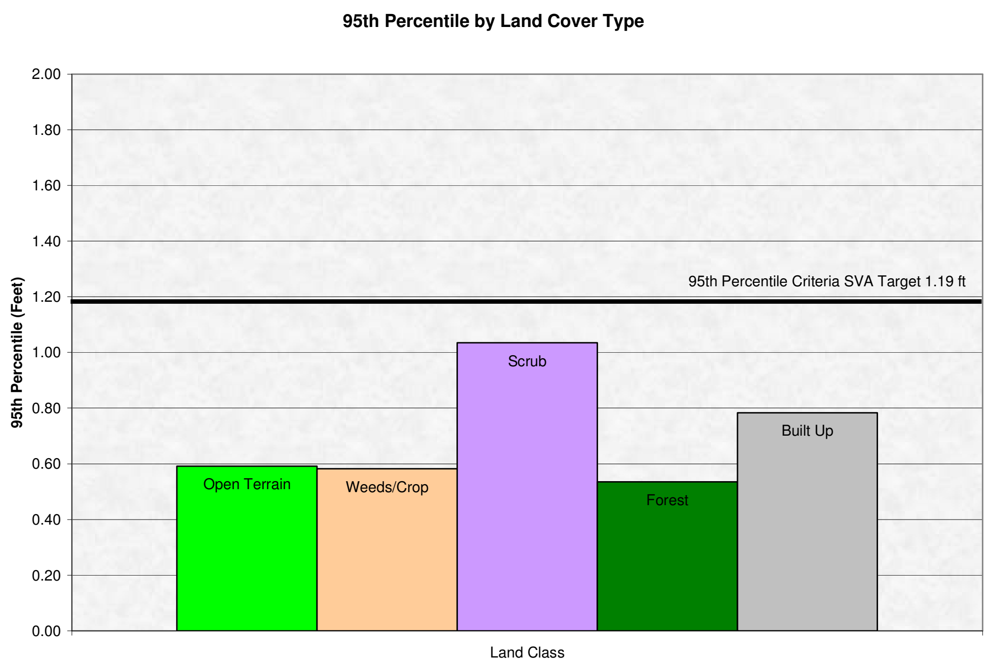


Figure 2 — Graph of SVA Values by Land Cover

Figure 3 illustrates the magnitude of the differences between the QA/QC checkpoints and LiDAR data by specific land cover category and sorted from lowest to highest. Whereas 95% of the checkpoints should be accurate within ±1.19 ft as shown in Figure 3, all 100% of the checkpoints met this criterion. Overall, there is a small positive bias (+0.14 ft) to the LiDAR data, as shown by the discrepancies in Open Terrain that vary between -0.48 ft and +0.77 ft. If the bare-earth elevations in Open Terrain followed a normal error distribution exactly, the mean would be zero, the maximum discrepancy would be approximately +0.62 ft and the minimum discrepancy would be approximately – 0.62 ft.

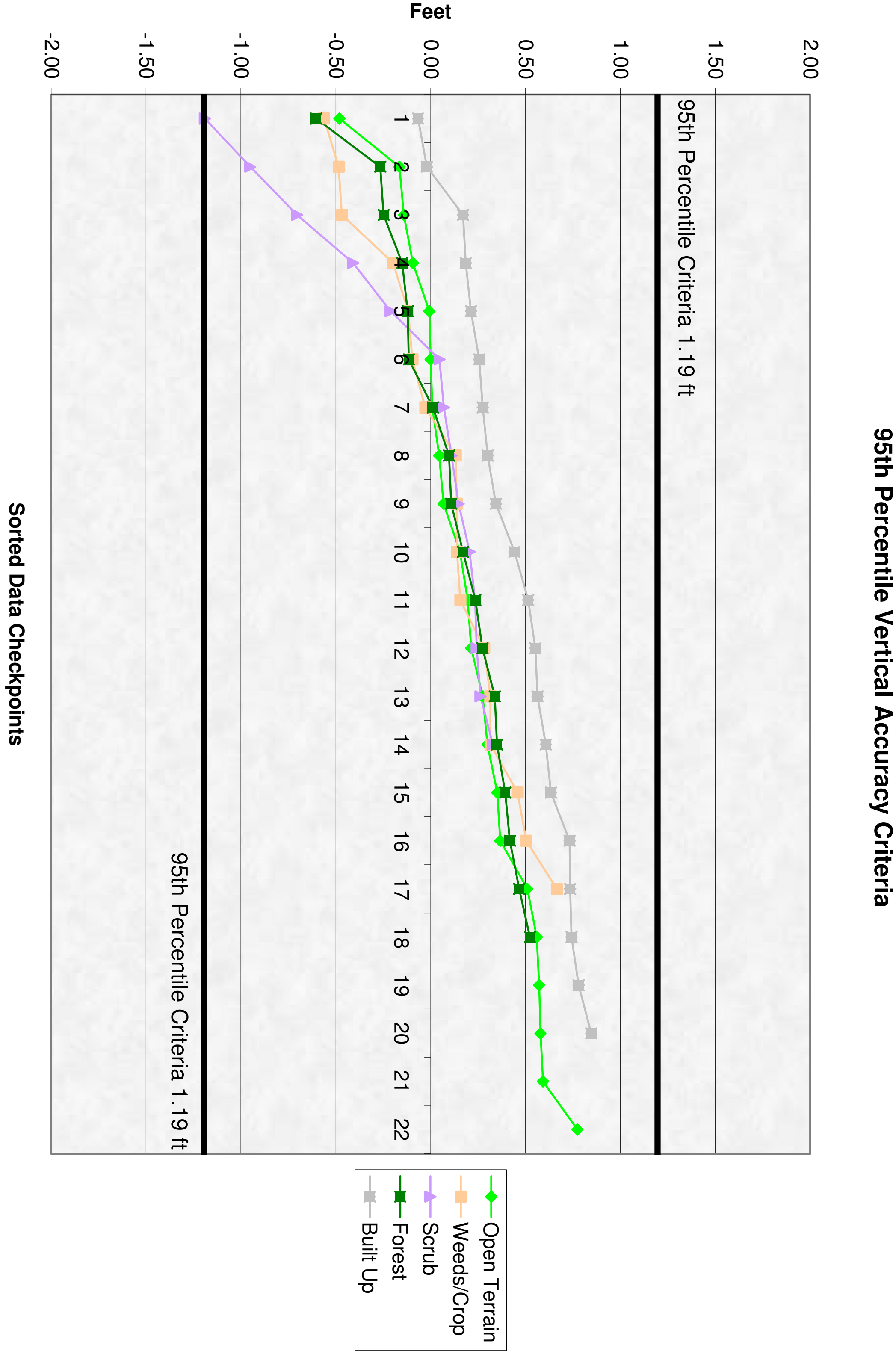


Figure 3 — Magnitude of Elevation Discrepancies, Sorted from Largest Negative to Largest Positive

Frequently Asked Questions

Q1: Why are errors in some land cover categories larger than others? Why can't the data be processed to be more consistent?

A1: The reason the LiDAR community (ASPRS and NDEP) converted from RMSE to 95th percentile is because the previously-used RMSE process assumes all errors are random and follow a normal error distribution (bell curve). Whereas the LiDAR community has proven that the LiDAR sensor itself yields random errors that follow a normal error distribution, and that DTM errors in Open Terrain also follow a normal error distribution, other land cover categories often have some forms of systematic errors that are not normal:

- In vegetated areas, for example, many LiDAR pulses never reach the ground. Where there are openings in the vegetation, the first LiDAR return may in fact measure the ground; but where the vegetation is dense, even the last return may measure the top of the vegetation. This is known to be a problem in dense weeds, wheat and corn fields, sawgrass, and forests where someone walking through the forest would have difficulty seeing the sky above. But this problem is even worse with all other remote sensing technologies. Photogrammetry, for example, needs to see the same points on the ground from two different aerial photos, with different perspectives, whereas LiDAR only needs a single pulse to penetrate between or through the vegetation.
- In LiDAR bare-earth post-processing, the "point cloud" elevations are filtered to remove all points that are believed to be above the ground surface. This leaves data voids where elevations have been deliberately deleted, creating larger TIN triangles from LiDAR mass points that remain. When x/y coordinates of a QA/QC checkpoint fall inside such dense vegetation, the LiDAR z-value (elevation) will be interpolated from a larger TIN triangle, causing larger interpolation errors that are unavoidable. LiDAR checkpoint elevations interpolated from larger TIN triangles almost always (unfairly) depict errors that are larger than when interpolated from small TIN triangles where an actual last return is perhaps only one foot away. When QA/QC checkpoints are interpolated from points that are 3 ft, 5 ft, 10 ft, 20 or more feet away, the interpolation errors will get increasingly larger.
- Some LiDAR analysts have recommended that QA/QC checkpoints only be used if they fall within one foot of a LiDAR last return elevation; but that would essentially dictate that checkpoints NOT be located in dense vegetation, and that would defeat the purpose for testing LiDAR datasets in vegetated areas in the first place. Furthermore, such surveys would be much more expensive to conduct.
- The compromise reached by the NDEP and ASPRS is that the Fundamental Vertical Accuracy (FVA) standard is mandatory for Open Terrain — that Supplemental Vertical Accuracy (SVA) is optional and only has "target values" for each land cover category — and that individual SVA values can exceed their "target value," especially if the Consolidated Vertical Accuracy (CVA) meets its standard. Some users specify FVA standards only, whereas other users treat CVA as mandatory. To closely align with FEMA requirements, Baltimore County treated CVA as mandatory; therefore individual SVA values could exceed their "target values" so long as the CVA standard is satisfied. However, in the case of this dataset for Baltimore County, all SVA "target values" were satisfied as well. In fact, the individual SVA values are much more consistent than in most other datasets evaluated by Dewberry.
- LiDAR bare-earth DTM elevations in vegetated areas often have a positive bias (higher than actual surveyed ground elevations), but individual errors may be negative because of the interpolation process described for larger TIN triangles.
- LiDAR returns in the Built-Up land cover category often have a negative bias (lower than actual surveyed ground elevations). This is because many checkpoints are on asphalt where (if hot), the LiDAR pulses tend to penetrate into the asphalt, providing returns that are a few centimeters lower than actual surveyed elevations.

Q2: If there is a small bias, can't we just adjust the data accordingly?

A2: Such an adjustment could be performed, of course, but the LiDAR community has determined that producers can adjust the elevations to better fit the checkpoint elevations only if they are able to detect and correct a systematic error. For example, one countywide LiDAR dataset had an apparent positive bias of 0.7 ft. The client subsequently determined that there was a 0.5 ft discrepancy between the elevation of the GPS base station used for the QA/QC checkpoint surveys and the GPS base station at the airport used for airborne GPS control of the LiDAR aircraft. That discrepancy had been unknown until both survey monuments were surveyed relative to each other. No one knew for sure which benchmark was more accurate, but this identified a systematic reason why the LiDAR errors would appear to have a positive bias of 0.5 ft. They could appropriately adjust for the known 0.5 ft systematic error, but they could not justifiably correct for the remaining 0.2 ft bias that remained.

The 0.15 ft bias in Baltimore County is so small that it does not justify the expense necessary to research and attempt to identify a systematic error of that magnitude. Such expenditure of time and money would only be justified for datasets that fail — not for datasets that pass.

Vertical Accuracy Testing in Accordance with NSSDA and FEMA Procedures

The NSSDA and FEMA guidelines were both published before it was recognized that LiDAR errors do not always follow a normal error distribution. Future changes to these FGDC and FEMA documents are expected to follow the lead of the NDEP and ASPRS. Nevertheless, to comply with FEMA's current guidelines in Reference C, $RMSE_z$ statistics were computed in all five land cover categories, individually and combined, as well as other statistics that FEMA recommends to help identify any unusual characteristics in the LiDAR data. These statistics are summarized in Figures 4 and 5 and Table 4 below, consistent with Section A.8.6.3 of Reference B. Table 4 also shows that the mean and median values are skewed on the high side of a normal error distribution.

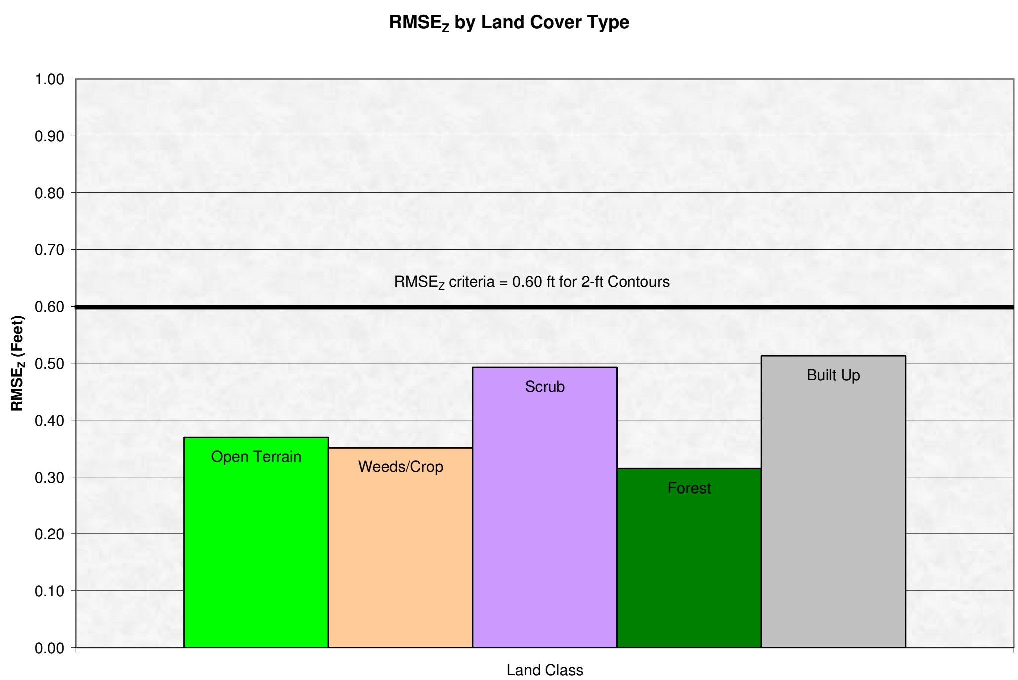


Figure 4 — $RMSE_z$ statistics by Land Cover Category

Table 4 — Overall Descriptive Statistics by Land Cover Category and Consolidated

Land Cover Category	$RMSE_z$ (ft)	Mean (ft)	Median (ft)	Skew	Std Dev (ft)	# of Points	Min (ft)	Max (ft)
Consolidated	0.41	0.16	0.21	-0.88	0.38	91	-1.19	0.85
Open Terrain	0.37	0.21	0.20	-0.17	0.31	22	-0.48	0.77
Weeds/Crops	0.35	0.07	0.14	-0.32	0.36	17	-0.56	0.66
Scrub	0.49	-0.13	0.09	-1.21	0.49	14	-1.19	0.33
Forest	0.31	0.10	0.14	-0.65	0.31	18	-0.60	0.52
Built Up	0.51	0.44	0.48	-0.27	0.27	20	-0.07	0.85

Figure 5 illustrates a histogram of the associated elevation discrepancies between the QA/QC checkpoints and elevations interpolated from the LiDAR triangulated irregular network (TIN). The frequency shows the number of discrepancies within each band of elevation differences. Although the discrepancies vary between a low of -1.19 ft and a high of +0.85 ft, the histogram shows that the majority of the discrepancies are skewed on the positive side of what would be a "bell curve," with mean of zero, if the data were truly normally distributed. This histogram is typical of all LiDAR datasets evaluated by Dewberry for hundreds of counties nationwide, because discrepancies in vegetation are typically positive.

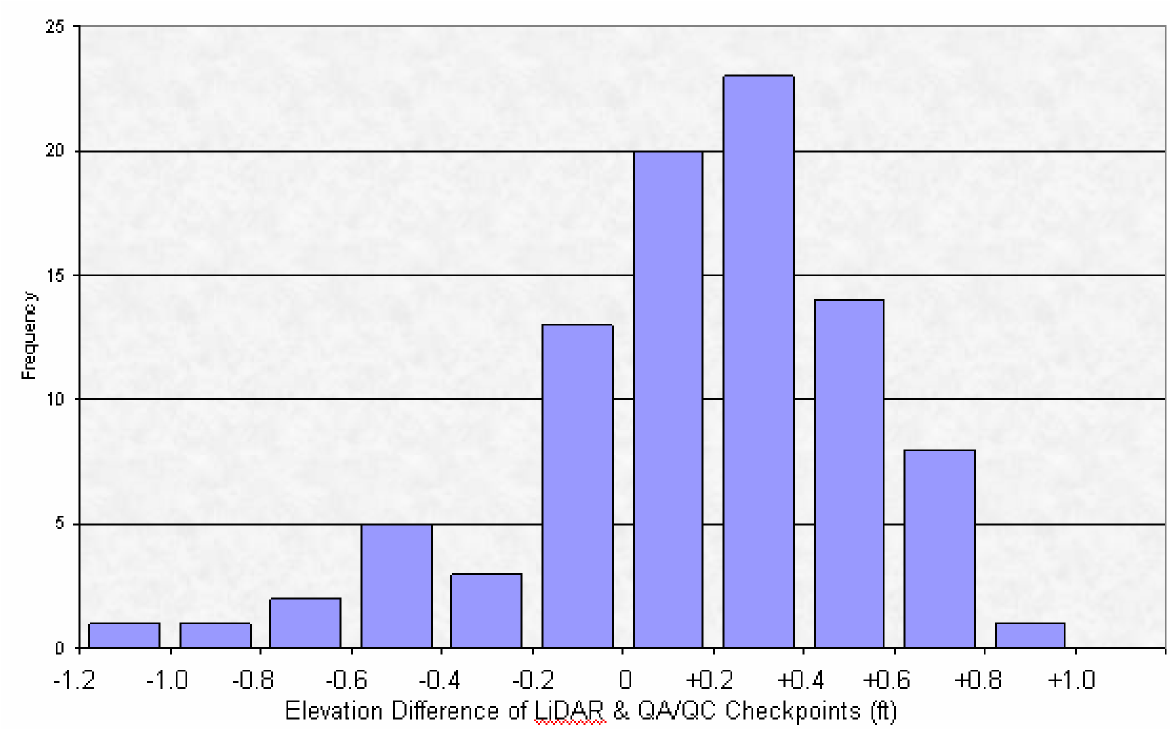
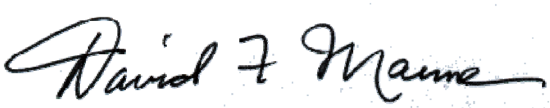


Figure 5 — Histogram of Elevation Discrepancies within 0.2 ft Bands

Conclusions

Based on the vertical accuracy testing methodology and the number of checkpoints, the LiDAR data have excellent vertical accuracy and are well suited for production of 2 ft contours.

- **Based on NSSDA and FEMA methodology: Tested 0.80 ft vertical accuracy at 95% confidence level.**
- **Based on NDEP and ASPRS methodology: Tested 0.76 ft vertical accuracy at 95% confidence level.**
- **These values greatly exceed the 1.19 ft vertical accuracy standard required for digital elevation data to support the generation of 2 ft contours**



David F. Maune, Ph.D., PS, GS, CP
QA/QC Manager