MARYLAND
DEPARTMENT OF
NATURAL RESOURCES

Robert L. Ehrlich, Jr.
Governor

Michael S. Steele
Lt. Governor

C. Ronald Franks
Secretary

W.P. Jensen
Deputy Secretary

LIDAR 2002
Quality Assurance Report

Authored by Dewberry LLC
2004

Tawes State Office Building • 580 Taylor Avenue • Annapolis, Maryland 21401
410.260.8DNR or toll free in Maryland 877.620.8DNR • TTY users call via Maryland Relay

The initial review of the Worcester County LIDAR data flown in the summer of 2002 did not yield the results as required per contractual specifications. With either the standard Root Mean Square Error (RMSE) method or the modified RMSE method accepted by FEMA for Phase 1 of the North Carolina Floodplain Mapping Program (NCFMP), the dataset errors greatly exceeded the 18.5 cm RMSE errors expected for elevation datasets equivalent to 2 ft contours. Although the data did not meet specifications, it was clear that some of the data was in fact sufficient. This indicated that with further processing and analysis, the data could potentially meet the desired accuracy. This was ascertained by reviewing and analyzing the data not only on the whole of the dataset but by looking at each land cover category in great detail.

This report will outline the analysis of the LIDAR data as reprocessed by Computational Consulting Services (CCS) for the Spatial Systems and Associates team during 2003. It is acknowledged that the initial data collection and post processing exhibited large errors. However, the intent of this report is to not explicitly identify the cause of these errors but rather review how this data could improved to better meet mandatory and/or desired specifications. The techniques utilized for the review process follow the same procedures as outlined in the initial Quantitative and Qualitative report submitted in the fall of 2002. These steps include verification against ground-truthed surveys, statistical analysis, and interpretative analysis based on the level of cleanliness of the data. This report will also examine the re-fly area and compare edge matching to ensure that this particular data is homogenous to the whole of the dataset.

Figure 1 illustrates the tile locations for re-assessment and the type of additional analysis. These tiles typically contain QA/QC survey checkpoints except for the area of re-flight.

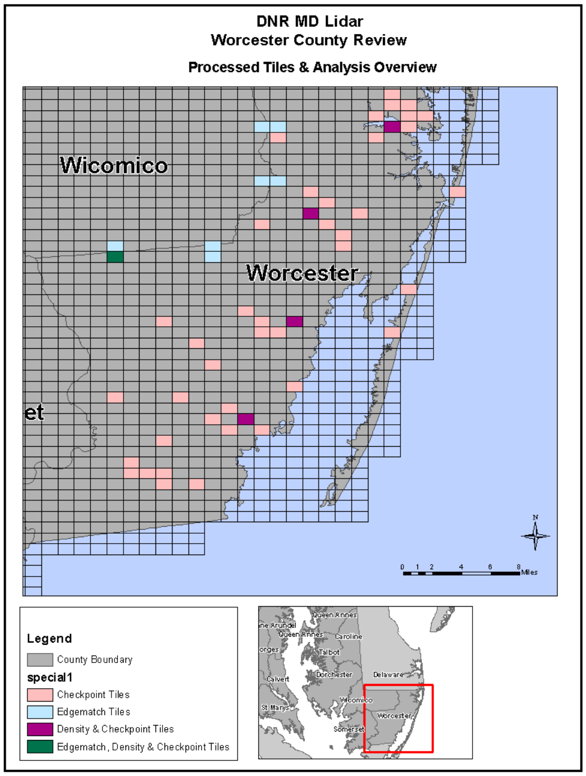


Figure 1 – Overview of project area.

With the review of LIDAR data, two fundamental questions must be asked;

1. Did the LIDAR system perform to specifications?
2. Did the vegetation removal process yield desirable results for the intended product?

Firstly, the basis was to establish whether or not the system obtained accurate elevation data, hence: did the LIDAR system perform to specification? In order to assess if the raw data were in fact accurate, only data collected in open terrain could be evaluated. This would ensure that the LIDAR "returns" were not influenced by artifacts (vegetation or man-made structures) or that the vegetation removal processes were in error. The basic principle is: if the data were to be measured in open terrain, the pulse of energy emitted by the sensor would be detected as a strong peak in reflected light. Since the laser light would not be influenced by the filtering through vegetation, the mathematics could easily identify the "last peak pulse" return of the laser, thereby obtaining an accurate delta elevation between the sensor and the target. Using the geo-referenced position of the aircraft, coupled with that of the sensor data, an accurate elevation is obtained.

The first report indicated that the open terrain land cover category of "Grass" data had an RMSE of 15.9 cm. This was a clear indication that the sensor performed to specification. The overall premise is: if the LIDAR exhibited accurate returns in open terrain, then it is capable of obtaining accurate returns in vegetated areas as long as the LIDAR could penetrate to the bare-earth terrain being measured. It should be noted that although the land cover category of "Built-up" could be considered open terrain, since this includes sidewalks and roadways, it is not open terrain. This is due to the wavelength of the LIDAR system and the ability of asphalt to absorb the laser light yielding slightly lower elevations. Also built-up areas that include structures can introduce some multi-path of the LIDAR near building edges, again lowering the elevations slightly.

It should be noted that although the data met specification within open terrain, which proved the sensor was obtaining adequate elevations, the LIDAR was not penetrating certain vegetation groups. The cause of non-penetration can be from a host of problems from the size of the laser footprint to a non-sensitive receiving mirror. For example, as the LIDAR is pulsed from the sensor it must pass through a beam diverter which converts the coherent light (laser) to a cone-like shape. This allows a larger target area to be measured, but it also allows this larger pulse of light to filter through openings in vegetation, allowing a better chance that the light may pass through to the ground. The trade-off for beam divergence is the ability to clearly identify the peak pulse of the last return. To measure this return value, typically only a few micro joules of energy are needed; however if the receiver is not fine tuned, it may be incapable of measuring the faint returns of light in heavily vegetated areas. Other factors such as extremely dense vegetation may never allow penetration which could be true in some cases with this dataset.

Since the data exhibited accurate results for open terrain, the potential was that the data could be reprocessed utilizing the proprietary algorithms of CCS. This reprocessed data would then re-address fundamental question number 2; did the vegetation removal process yield desirable results for the intended product? To ascertain if the data was not only accurate enough but also "clean" of artifacts for a true bare-earth product, the data was analyzed with a quantitative and qualitative approach.

Checkpoint Survey – A Quantitative Analysis

As outlined in the initial proposal and assessment report, the vertical accuracy of the LIDAR data (ground-truthing) was to be performed by surveying checkpoints in strategic locations. These checkpoint surveys were to follow the locational criteria as set forth by the FEMA Guidelines and Specifications for Flood Hazard Mapping Partners (Section A.6.4 of Appendix A: Guidance for Aerial Mapping and Surveying), by the testing methods of the National Digital Elevation Program (NDEP), and by methods developed by Dewberry (for both these programs). The first part of this process is to base the number of checkpoints on the number of major land cover categories representative of the county. The example given was that if 5 categories represented the major land cover categories, then a minimum of 20 checkpoints would be measured for each of these land cover categories, for a total of 100 checkpoints.

The checkpoint survey submitted to Dewberry by the Independent Surveyor (IS) does not meet guidelines regarding the number of checkpoints needed for assessment. The current checkpoint survey has a total of 60 checkpoints in five categories but does not have the required 20 checkpoints for each land cover type as previously required. Since some checkpoints were not in major land cover categories as defined, they were re-classed to fit one of the other major types for a total of four land cover types (see Table 1). Even with this, the minimum still does not satisfy the initial requirement. However, by utilizing additional checkpoints as supplied by the Department of Natural Resources (DNR), the minimum requirement is met.

A total of 36 survey checkpoints from DNR were used to supplement the LIDAR verification. From these 36 checkpoints, 31 were located in the project area with the remainder just outside. No supplemental data was provided with these checkpoints to qualify their accuracy or to determine whether or not they satisfied locational guidelines specified by FEMA or the NDEP. Since the checkpoints were very specific to the land cover type, they were re-classed into the more generalized land cover categories as used by the independent surveyed checkpoints. Within this list, 11 checkpoints did not have any classification so they were classed as "unknown". The following vertical accuracy assessments are based on using the combined checkpoints from the Independent Surveyor and from DNR.

Figure 2 illustrates the location and agency responsible for the survey checkpoints within the county. Table 1 lists the location, QA/QC survey elevation, LIDAR elevation as derived from a Triangulated Irregular Network (TIN), original land cover classification, re-classed land cover classification, and the associated slope of the terrain of the checkpoints. Care must be taken to assess the slope of the checkpoints locations since the checkpoints are verifying the LIDAR. Checkpoints located on a high slope could falsely accuse the LIDAR data of being inaccurate. The outline for the Independent Surveyor was to establish checkpoints on as level terrain as possible within a 5 meter radius. The secondary criteria was that the slope be less than 20% (preferably less than 10%) and at least 5 meters away from breaklines, as specified in section A.6.4, Appendix A to FEMA's Guidelines and Specifications; this same criteria for selection and location of checkpoints has been adopted by the National Digital Elevation Program (NDEP) which has submitted its recommendations to the Federal Geographic Data Committee (FGDC) for adoption in the next revision to the National Standard for Spatial Data Accuracy (NSSDA). If the derived slope value is high and there is confidence that the checkpoint is on fairly level ground, this could indicate an error within the LIDAR. For the 91 checkpoints, 4 stations are greater than 10% but less than 14% with three being in vegetated areas and one in grass. Since these values are within specifications and they are not the largest outliers, no further action is warranted.

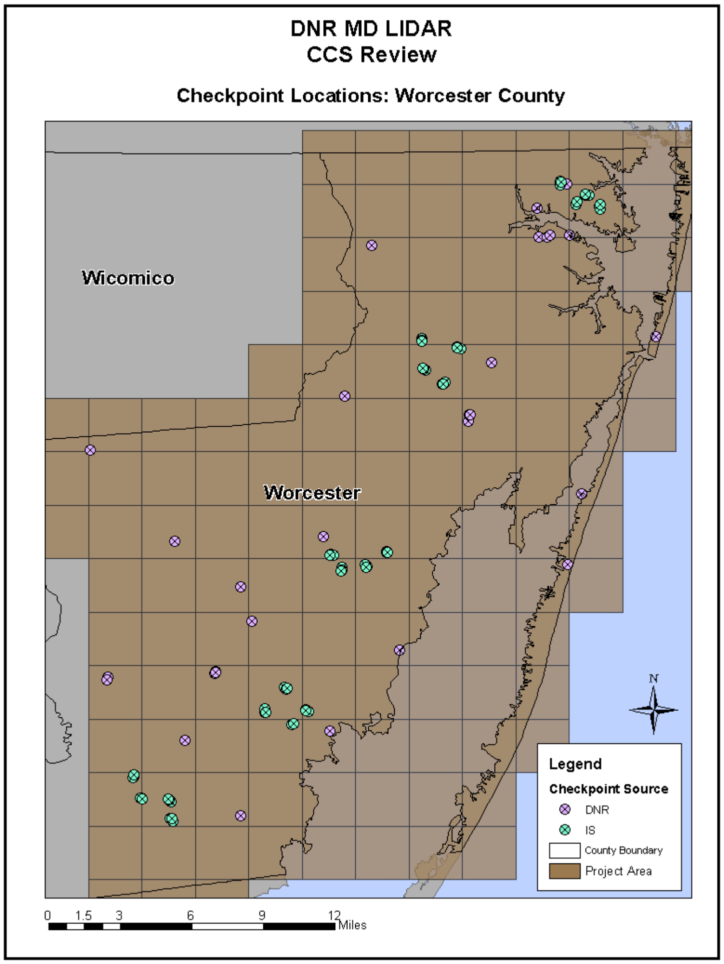


Figure 2- Survey checkpoint locations

Survey Checkpoints With Associated Slope							
STATION	EASTING	NORTHING	QA_QC Elev (meters)	LIDAR Elev (meters)	Land Cover Original	Land Cover Re-classed	Slope %
1	561679.834	82959.388	2.315	2.250	Urban	Built-up	1.5
2	560334.056	82919.008	4.476	3.610	Deciduous Forest	Forest	2.5
3	559610.023	82793.864	3.653	3.920	Open Field	Grass	5.1
4	559441.944	84724.338	0.202	0.630	Veg. Wetland	Weeds/Crop	8.1
5	561491.005	86368.293	5.157	5.070	Road Bed	Built-up	1.0
6	567451.759	76119.936	1.140	1.190	Urban	Built-up	3.6
7	562489.439	65539.275	0.242	0.270	Veg. Wetland	Weeds/Crop	1.2

8	556373.415	74344.147	8.580	8.710	Road Bed	Built-up	1.1
9	554872.541	70881.000	8.259	8.210	Pine Forest	Forest	1.0
10	546474.986	72094.919	8.273	8.150	Cropland	Weeds/Crop	13.0
11	535038.884	62325.839	9.530	9.580	Cropland	Weeds/Crop	3.5
12	529366.575	68470.561	13.605	13.650	Open Field	Grass	12.8
15	530560.872	53222.754	8.312	8.340	Pine Forest	Forest	1.4
16	540210.860	56931.386	4.424	4.240	Urban	Built-up	2.4
17	545043.734	62625.462	7.286	7.380	Open Field	Grass	3.5
18	550226.250	55031.752	0.895	0.690	Road Bed	Built-up	1.1
19	545525.532	49562.785	0.158	0.190	Veg. Wetland	Weeds/Crop	5.2
20	537805.481	53587.939	4.242	4.070	Deciduous Forest	Forest	3.3
21	535709.507	48935.525	8.668	9.300	Cropland	Weeds/Crop	7.3
22	539478.540	43875.629	8.205	8.060	Road Bed	Built-up	1.7
97	539512.844	59276.809	6.199	6.220	FURNACE	Unknown	2.7
98	561509.582	60750.798	1.460	1.390	NORTH BEACH 2	Unknown	3.1
99	548333.117	82237.774	10.577	11.200	PITTSVILLE EAST	Unknown	9.0
102	560330.407	82950.279	3.949	4.140	D02A	Unknown	1.5
109	554967.120	70840.999	9.043	8.900	D09A	Unknown	2.4
115	530545.575	53201.079	9.035	9.090	D15A	Unknown	8.4
120	537803.298	53534.089	5.226	5.220	D20A	Unknown	3.0
202	560136.606	82806.528	6.002	5.530	D02B	Unknown	1.7
209	554843.824	70387.649	9.507	9.420	D09B	Unknown	3.8
215	530505.629	53032.196	7.885	8.000	D15B	Unknown	9.2
220	537716.451	53465.909	6.494	6.290	D20B	Unknown	10.4
101	563672.030	84889.623	1.525	1.460	Pavement	Built-up	3.5
102	563694.588	84681.853	1.447	1.810	Crop Field	Weeds/Crop	1.2
103	563693.472	85027.891	1.650	1.660	Mixed Woods	Forest	2.0
104	562149.500	85204.827	2.411	2.230	Pavement	Built-up	4.2
105	562061.655	85045.715	1.940	2.360	Tall Weeds	Weeds/Crop	4.9
106	562065.070	84996.893	1.788	2.010	Mixed Woods	Forest	1.5
107	561050.803	86539.876	5.458	5.530	Pavement	Built-up	6.4
108	561024.285	86333.621	5.309	5.380	Crop Field	Weeds/Crop	1.4
109	561117.183	86529.644	5.227	5.360	Mixed Woods	Forest	6.7
110	562959.016	85627.473	1.219	1.170	Pavement	Built-up	2.7
111	562739.514	85608.407	1.108	1.160	Tall Grass	Weeds/Crop	1.8
112	562719.568	85679.223	0.840	0.790	Mixed Woods	Forest	3.0
201	554335.402	75305.306	10.431	10.310	Bare ground	Grass	1.6
202	554070.700	75402.795	10.076	9.940	Grass	Grass	10.0
203	554059.685	75373.282	10.210	9.740	Mixed Woods	Forest	5.7
204	551682.612	75979.943	7.904	7.620	Pavement	Built-up	1.5
205	551667.884	75847.698	8.283	8.110	Tall Grass	Weeds/Crop	0.1
206	551674.193	75801.668	8.390	8.850	Crop Field	Weeds/Crop	1.6
207	551968.710	73872.385	9.030	8.890	Bare ground	Grass	13.8
208	551816.859	73907.217	8.495	8.360	Tall Grass	Weeds/Crop	1.6
209	551779.575	73970.949	9.259	8.750	Mixed Woods	Forest	8.2
210	553079.811	72908.926	9.872	9.870	Pavement	Built-up	1.2
211	553237.521	73050.738	9.648	9.800	Tall Weeds	Weeds/Crop	1.6
212	553165.121	72896.330	9.753	9.990	Mixed Woods	Forest	7.7

301	545534.226	61479.443	7.423	7.560	Crop Field	Weeds/Crop	6.6
302	545770.249	61407.103	9.441	9.430	Bare ground	Grass	1.9
303	545513.684	61406.221	7.886	7.810	Mixed Woods	Forest	2.8
304	546340.254	60565.367	10.591	10.510	Bare ground	Grass	1.4
305	546298.655	60418.141	10.766	10.580	cut over	Weeds/Crop	2.5
306	546239.341	60317.214	10.592	10.540	Mixed Woods	Forest	1.4
307	547990.159	60665.920	11.520	11.290	Pavement	Built-up	2.0
308	547875.424	60752.873	10.569	10.680	Crop Field	Weeds/Crop	10.4
309	547966.306	60614.009	10.993	10.300	tall weeds/tree	Forest	0.9
310	549332.814	61645.096	12.045	11.850	grass	Grass	2.0
311	549311.010	61570.311	13.241	14.020	Crop Field	Weeds/Crop	4.0
312	549401.780	61574.156	10.638	10.720	tall weeds	Weeds/Crop	2.4
401	542428.798	52516.384	9.877	9.650	Pavement	Built-up	2.1
402	542564.705	52415.151	10.379	10.500	Tall Grass	Weeds/Crop	3.9
403	542592.237	52457.496	10.394	10.160	Mixed Woods	Forest	3.1
404	543863.878	50997.904	11.576	11.520	low grass	Grass	3.0
405	544057.094	50851.028	11.203	11.260	Crop Field	Weeds/Crop	1.2
406	543863.924	50975.404	11.838	11.840	Mixed Woods	Forest	3.5
407	542904.351	50016.494	10.193	9.970	Pavement	Built-up	2.6
408	543030.875	50062.614	10.013	10.280	Crop Field	Weeds/Crop	2.5
409	543084.418	50039.682	10.229	10.320	Mixed Woods	Forest	2.5
410	541116.690	50810.485	11.551	11.200	low grass	Grass	3.6
411	541127.669	51069.356	11.381	10.690	Crop Field	Weeds/Crop	1.4
412	541173.960	50817.495	11.517	11.220	Mixed Woods	Forest	1.6
501	532288.686	46515.909	4.192	4.400	grass	Grass	4.4
502	532250.216	46409.751	4.216	4.700	Crop Field	Weeds/Crop	7.8
503	532320.093	46619.852	4.524	4.490	Mixed Woods	Forest	2.8
504	532884.714	44967.181	8.398	8.570	Crop Field	Weeds/Crop	1.5
505	532785.274	45062.306	7.532	7.690	grass	Grass	1.7
506	532746.562	45077.845	6.801	6.870	Mixed Woods	Forest	3.7
507	534752.702	43664.557	9.995	10.290	tall weeds	Weeds/Crop	2.4
508	534896.467	43494.702	10.355	10.370	crop field	Weeds/Crop	2.8
509	534834.684	43667.954	10.041	10.120	mixed woods	Forest	3.0
510	534773.773	44816.959	11.641	11.730	bare ground	Grass	2.3
511	534643.018	44930.735	10.347	10.830	crop field	Weeds/Crop	6.1
512	534613.214	45015.608	10.457	10.400	mixed woods	Forest	1.8

Table 1 – Combined Survey Checkpoints from the Independent Surveyor and DNR Checkpoints with the associated slope.

Vertical Accuracy Assessment using RMSE Methodology

The first method of testing vertical accuracy is to use the Root Mean Square Error (RMSE) approach which is valid when errors follow a normal distribution. This methodology measures 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 points. The vertical accuracy assessment compares the measured survey checkpoint elevations with those of the Triangulated Irregular Network (TIN) as generated from the LIDAR. The survey checkpoint's X/Y location is overlaid on the TIN and the interpolated Z value is recorded. This interpolated Z value is then compared to the survey checkpoint Z value and this difference represents the amount of error between the measurements. The following graphs and tables outline the vertical accuracy and the statistics of the associated errors.

Table 2 summarizes the RMSE using:

- 100% of the IS and DNR checkpoints (method used by FEMA when errors are assumed to follow a normal distribution)
- 95% of the checkpoints ("95% clean" methodology used in Phase I of the NCFMP, where errors are still assumed to follow a normal distribution but where 5% of the errors were assumed to fall in "uncleaned" areas)
- Checkpoints categorized by land cover type

RMSE by Land Cover				
%	**RMSE (cm)**	**# of Points**	**Land Class**	**RMSE Criteria (cm)**
100	**26.9**	91	All	18.5 (FEMA methodology)
95	**21.3**	86	**All**	**18.5 (NCFMP Phase I methodology)**
15	**16.5**	14	**Grass**	
29	**24.9**	25	**Weeds/Crop**	
21	**20.2**	20	**Forest**	
18	**16.0**	16	**Built-up**	
12	**25.9**	11	**Unknown**	

Table 2 – RMSE of LIDAR based on QA/QC survey checkpoints

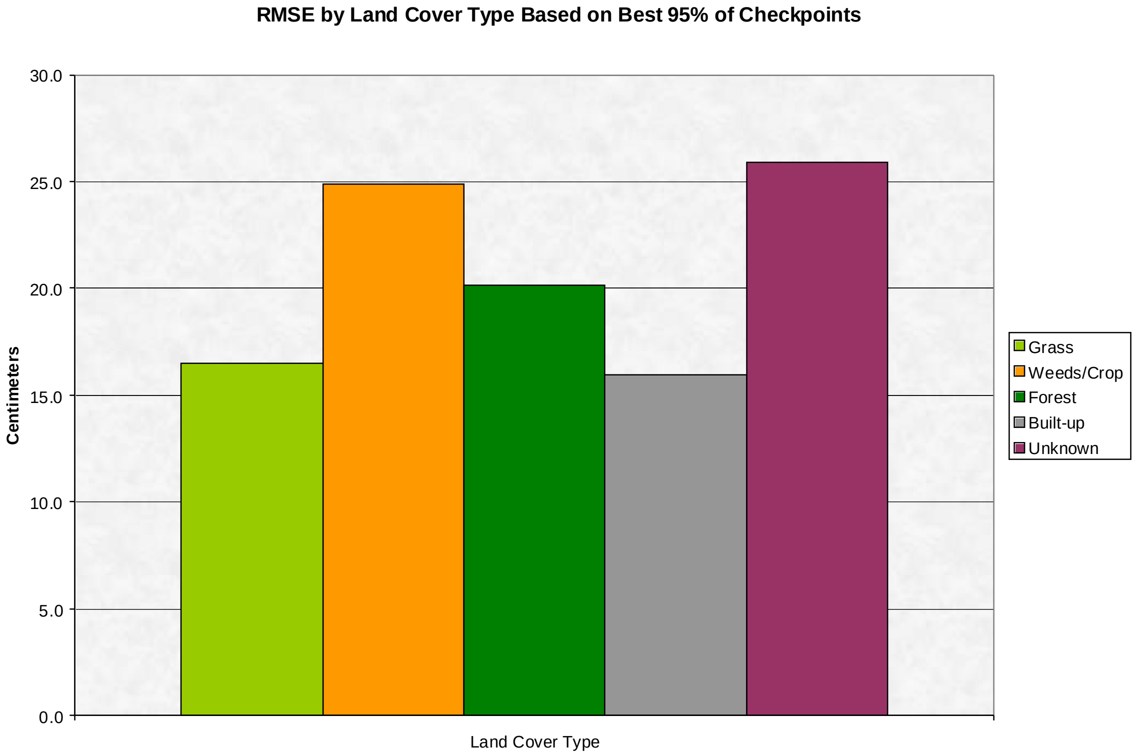


Figure 3 – RMSE by specific land cover type

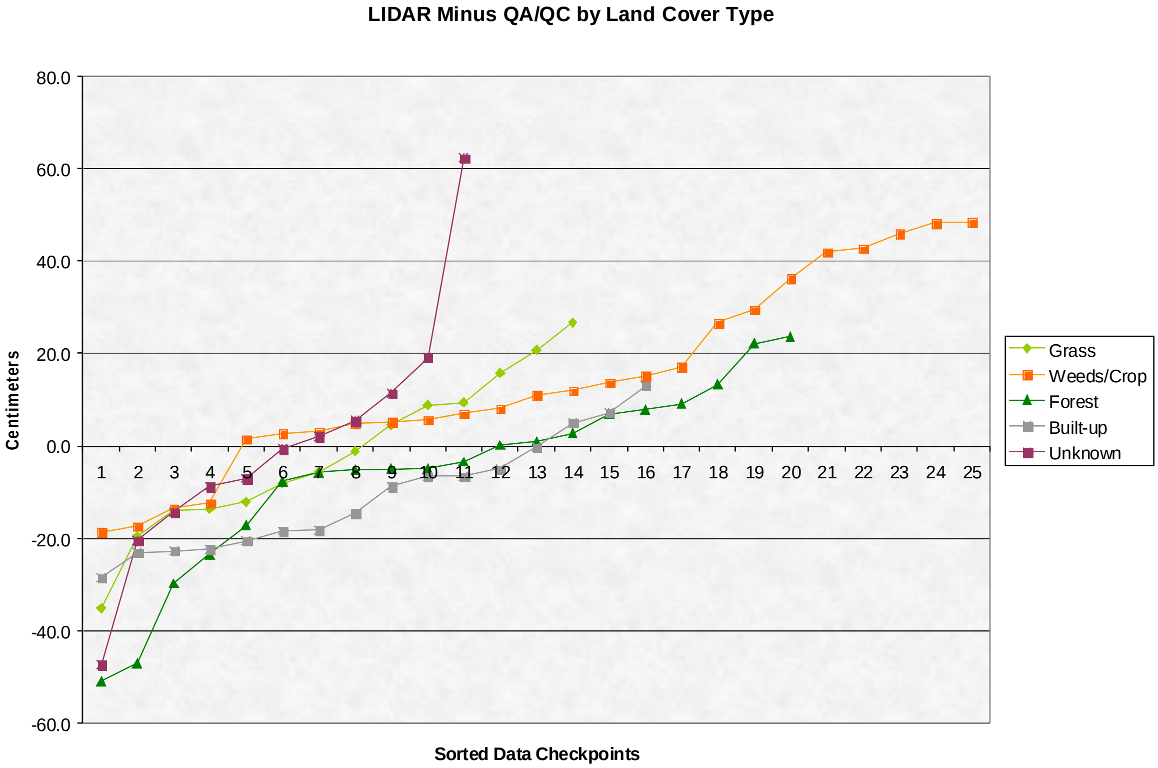


Figure 4 - Illustrates the magnitude of differences between the checkpoints and LIDAR data by specific land cover type and sorted from lowest to highest.

Table 3 summarizes all the descriptive statistics referenced in FEMA guidelines.

Overall Descriptive Statistics								
	RMSE (cm)	**Mean (cm)**	**Median (cm)**	**Skew (cm)**	**Std Dev (cm)**	**# of Points**	**Min (cm)**	**Max (cm)**
100% Pts	**26.9**	-0.2	0.2	-0.1	27.1	91	-86.6	77.9
95% Pts	**21.3**	0.7	0.6	0.3	21.4	86	-50.9	62.3
Grass	**16.5**	-1.6	-3.4	-0.1	17.1	14	-35.1	26.7
Weeds/Crop	**24.9**	14.7	11.1	0.2	20.5	25	-18.6	48.4
Forest	**20.2**	-5.6	-4.1	-0.9	19.9	20	-50.9	23.7
Built-up	**16.0**	-10.6	-11.6	0.5	12.3	16	-28.4	13.0
Unknown	**25.9**	0.2	-0.6	0.8	27.2	11	-47.2	62.3

Table 3 – Overall descriptive statistics.

Figure 5 and Figure 6 illustrates the histogram of the associated delta errors between the interpolated LIDAR TIN and the survey checkpoint. It is interesting to note that the errors do not follow a normal distribution. Even when the 5% largest errors are removed, the errors still do not follow a normal distribution. With this scenario where some errors do not follow a normal distribution, invalidates the RMSE methodology, the NDEP recommends that alternative criteria be used to determine the Fundamental Vertical Accuracy (mandatory) and Supplemental and Consolidated Vertical Accuracies (optional).

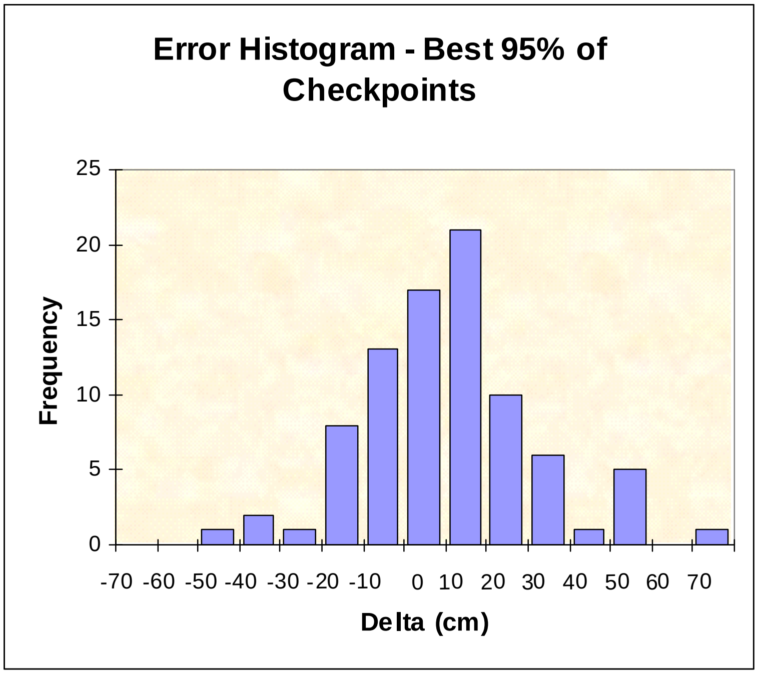


Figure 5 – Error Histogram of the best 95% of data checkpoints. This illustrates that the errors do not follow a normal distribution even when the top 5% of outliers are removed, because there are nine outliers larger than 41.7 cm (1.9600 X RMSE) whereas there should only be four (actually 4.3) such outliers if errors had followed a normal distribution.

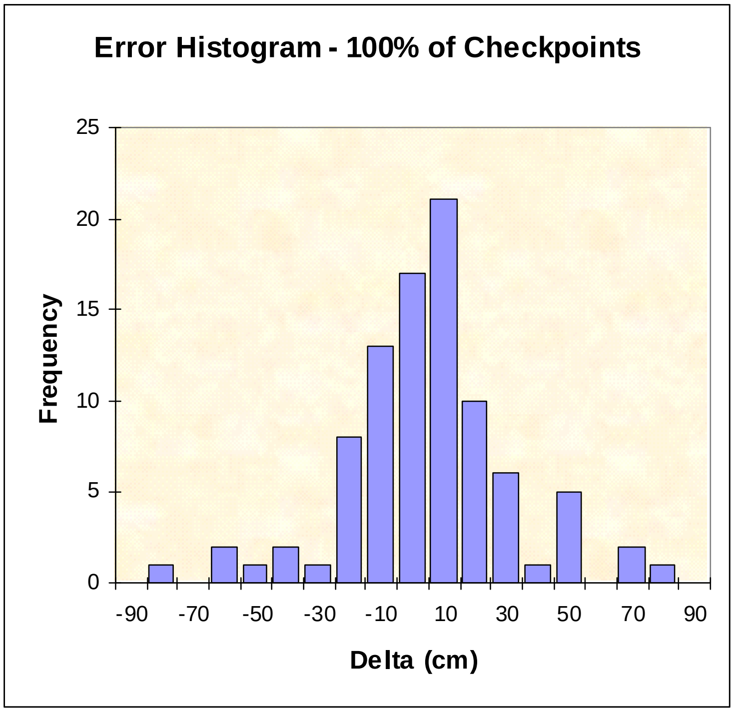


Figure 6 – Error histogram of 100% of the data checkpoints. Note the data does not have a normal distribution of errors. Here also, there are more outliers than expected for the RMSE.

Vertical Accuracy Assessment using NDEP Methodology

The Fundamental Vertical Accuracy at the 95% confidence level equals 1.9600 times the RMSE in open terrain only; in open terrain, there is no valid excuse why errors should not follow a normal error distribution, for which RMSE methodology is appropriate. Supplemental Vertical Accuracy at the 95% confidence level utilizes the 95th Percentile error individually for each of the other land cover categories which may have valid reasons (e.g., problems with vegetation removal) why errors do not follow a normal distribution. Similarly, the Consolidated Vertical Accuracy at the 95% confidence level utilizes the 95th Percentile error for all land cover categories combined. This NDEP methodology is used on all 100% of the checkpoints and not just on the best 95% of those checkpoints.

The target objective for this project in Worcester County was to achieve bare-earth elevation data with an accuracy equivalent to 2 ft contours, which equates to an RMSE of 18.5 cm when errors follow a normal distribution. With this criteria, the Fundamental Vertical Accuracy of 36.3 cm must be met. Furthermore, it is desired that the Consolidated Vertical Accuracy and each of the Supplemental Vertical Accuracies also meet the 36.3 cm criteria to ensure that elevations are also accurate in vegetated areas. As summarized in **Table 4**, this data does satisfy the NDEP's mandatory Fundamental Vertical Accuracy criteria for 2 ft contours, but it fails the NDEP's optional Consolidated and Supplemental Vertical Accuracy criteria for 2 ft contours in weeds/crops, forests, and land covers specified as "unknown." Although vastly improved from the original dataset, the reprocessed data still has more outliers than desired in these land cover categories, indicating the elevations in vegetated areas are less accurate than desired when assuming that all checkpoints are correct and all errors are attributed to the reprocessed LIDAR data.

Vertical Accuracy at 95% Confidence Level Based on NDEP Methodology for 2 ft contours				
Land Cover Category	**# of Points**	**Fundamental Vertical Accuracy (mandatory) 36.3 cm**	**Consolidated Vertical Accuracy (optional) 36.3 cm**	**Supplemental Vertical Accuracy (optional) 36.3 cm**
Open Terrain/Grass	**14**	**32.3 cm**		
Built Up	**16**			**24.4 cm**
Weeds/Crops	**28**			**67.0 cm**
Forests	**22**			**68.4 cm**
Unknown	**11**			**54.7 cm**
Total Combined	**91**		**62.8 cm**	

Table 4 - Vertical Accuracy per NDEP Methodology

In order to further assess this reprocessed data, it must be compared to the previous processing to validate it's improvement not only statistically but also qualitatively. Table 5 compares the original processing RMSE values with the new values.

Comparison of Original Processing vs. Re-Processed			
	Original	**Re-Processed**	
Land Class	**RMSE (cm)**	**RMSE (cm)**	**# of Points**
All 100%	**36.6**	**26.9**	91
All 95%	**28.0**	**21.3**	86
Grass	**15.9**	**16.5**	14
Weeds/Crop	**28.2**	**24.9**	25
Forest	**40.1**	**20.2**	20
Built-up	**20.5**	**16.0**	16
Unknown	**23.4**	**25.9**	11

Table 5 - Comparison of original process vs. Re-processed data.

Table 5 clearly shows a vast improvement from the original process, especially in Forest. The two land cover categories that did not change significantly is that of Weeds/Crop and Unknown. As previously indicated the LIDAR sensor was not able to penetrate the dense agricultural areas partly due to the crop type, stage of growth and potentially a sensor that was not ideal for this time of year with full leaf-on vegetation. It should also be noted that although the sensor did not perform optimally for these types of areas, it was still able to obtain adequate results that no other sensor would be capable of. If this area were flown using traditional photogrammetric techniques, these areas (including forest) would be marked with dashed contours indicating they are obscured as the photogrammetrist would not be able to "see" (in stereo) the ground being measured. Photogrammetry requires a view of bare-earth terrain points from two distinctly different perspective views (normally obscured by vegetation in agricultural fields and forests), whereas LIDAR only needs to penetrate the vegetation with a single pulse between or through the trees, crops and weeds.

From further review of the data for "Unknown," it can be clearly seen that one outlier can easily skew the results. For example, one checkpoint from the DNR group of points has a delta difference of 62.3 cm, yet the next largest error is 47.2 and the remaining nine (9) apparent errors are less than 20 cm. By removing this one point of 62.3 we can see that the RMSE of Unknown can improve to a RMSE of 18.7 cm. Since there is no metadata associated with DNR points pertaining to how they are collected, and not knowing whether or not these checkpoints satisfied the FEMA and NDEP criteria for their selection and location, some caution must be used to assess the derived values. This can also be seen with Weeds/Crop whereby one or two outliers can skew the RMSE, especially if the apparent outlier errors might be caused by improperly-located checkpoints or actual errors in checkpoint coordinates.

In Phase I of the NCFMP, the purpose of utilizing the best 95% of points is to address data that does not follow a normal distribution for errors and to allow for 5% of the checkpoints to fall in "uncleaned" areas since the NCFMP's criteria was for the dataset to be 95% clean. Normally, removing the worst 5% of the checkpoints removes most outliers, but due to the factors affecting this project in Worcester County, such as dense agricultural growth, there are a greater number of outliers which skew the distribution. Since the data does not follow a normal distribution the methodology for reporting accuracy utilizes the 95 percentile method as endorsed by the National Digital Elevation Program (NDEP), the methodology currently being used for Phase II of the NCFMP.

Consistent with the NDEP, in order for elevations in all land cover categories combined to be equivalent to 2 ft contours, 95% of the elevation errors at checkpoints (86 of 91 checkpoints) should be accurate within 36.3 cm, i.e., the 95th Percentile error used for Consolidated Vertical Accuracy should be 36.3 cm or less. However, for the 91 checkpoints used in this evaluation, 14 of those checkpoints (15.4%) had errors larger than 36.3 cm. Thus, it can be said that the dataset tested as equivalent to 2 ft contours only at the 84.6% confidence level rather than the 95% confidence level. The 95th Percentile error is 62.8 cm; the 90th Percentile error is 47.2 cm; the 85th Percentile error is 39.2 cm; and the 80th Percentile error is 28.4 cm.

Statistically the reprocessed data has improved considerably from original data reduction. Although it does not meet the ideal criteria of two foot contours, the data still should be sufficient for FEMA's needs since FEMA guidelines allow flexibility in using the best available data. The criteria of 18.5 cm was derived from the guidelines and are not strict specifications. Since they are guidelines, FEMA typically will accept data that exceeds this amount. One of the mandates of FEMA is to use the "best available data" without having to spend great amounts of money to obtain higher accuracy. If the data is statistically tested and explicitly states the calculated accuracy, it most likely will suffice for FEMA's and the State's needs. For example, for Phase II of the NCFMP, FEMA accepts LIDAR bare-earth data, designed for 2 ft contours, but tested to achieve a Fundamental Vertical Accuracy specification of 36.3 cm in open terrain, a Consolidated Vertical Accuracy specification of 49.0 cm in all land cover categories combined, and Supplemental Vertical Accuracy "targets" of 49.0 cm in each of the five individual land cover categories assessed.

Re-Fly Area

Figure 7 illustrates the area that had been re-flown by Airborne 1 in order to fulfill the missing data as per the contract for Phase 1. Since the re-flight was with a different company using a different sensor, verification was required. For the re-flight area no ground-truth surveys were available so different techniques were utilized to look at the relative accuracy between the adjacent tiles of new and old data. By having tested the old data for absolute accuracy the relative accuracy should yield results extrapolating the absolute accuracy of the new data.

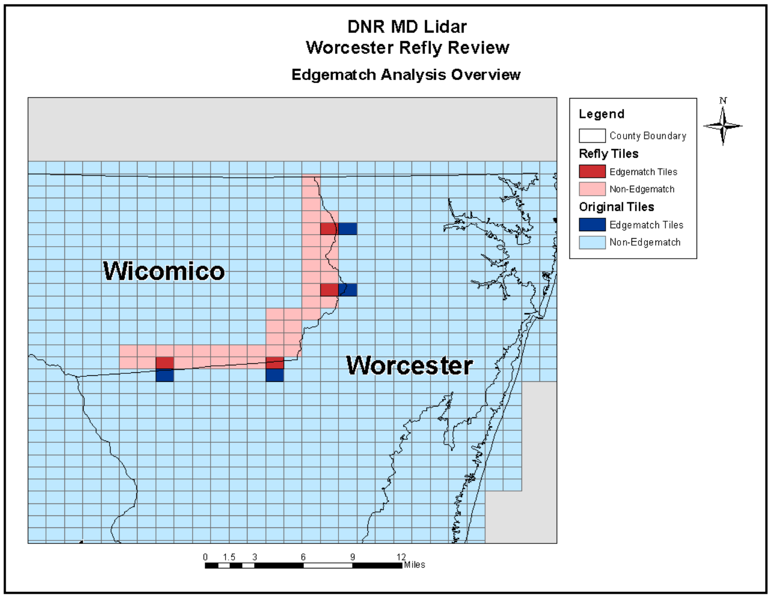


Figure 7 – Map of Re-fly area

To assess the relative accuracy, TIN's were created for adjacent tiles from both the new and old data. The TIN's were visually inspected looking for anomalies that may indicate a vertical shift between the two data sets. Additionally hillshade and slope maps were created to aid in this process. The overall consistency of the new re-flight data contrasts slightly with the re-processed data due to the issues of the original sensor and the time of year the data was flown. The re-flight area was flown during leaf free conditions and with crops in their starting growth stage. Figure 8 illustrates the two adjacent TIN's and the locations of the cross sections. It can be noted that the re-fly area to the north indicates a smoother consistent terrain model. Figure 9 shows the slope between the adjacent TIN's, again showing the re-fly area to be more consistent.

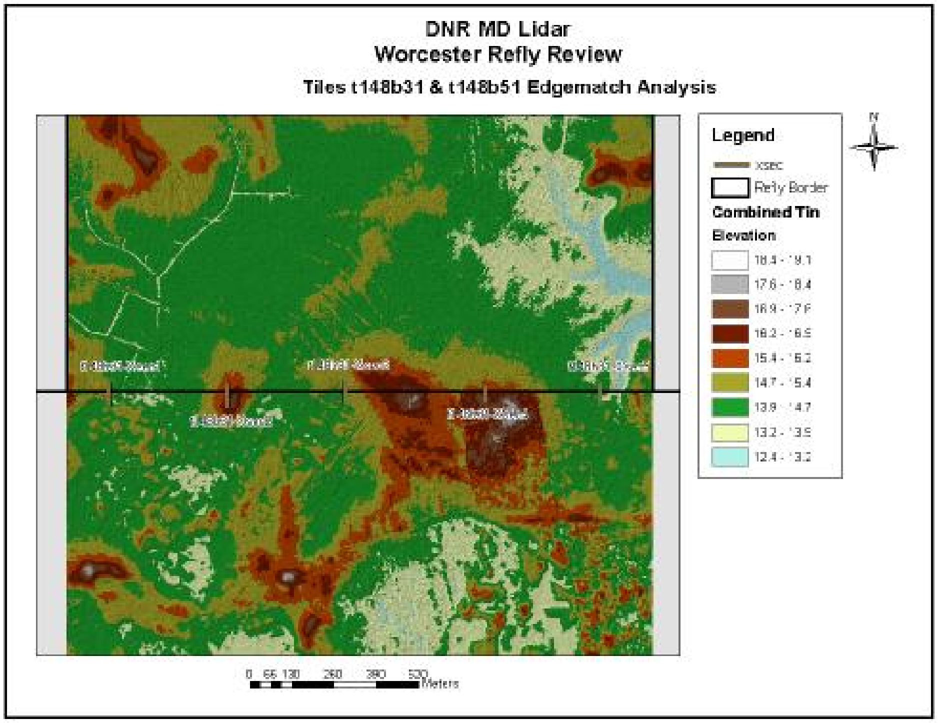


Figure 8 - Edge matching of TIN's between the re-flight area (top area) and the existing re-processed data (bottom area)

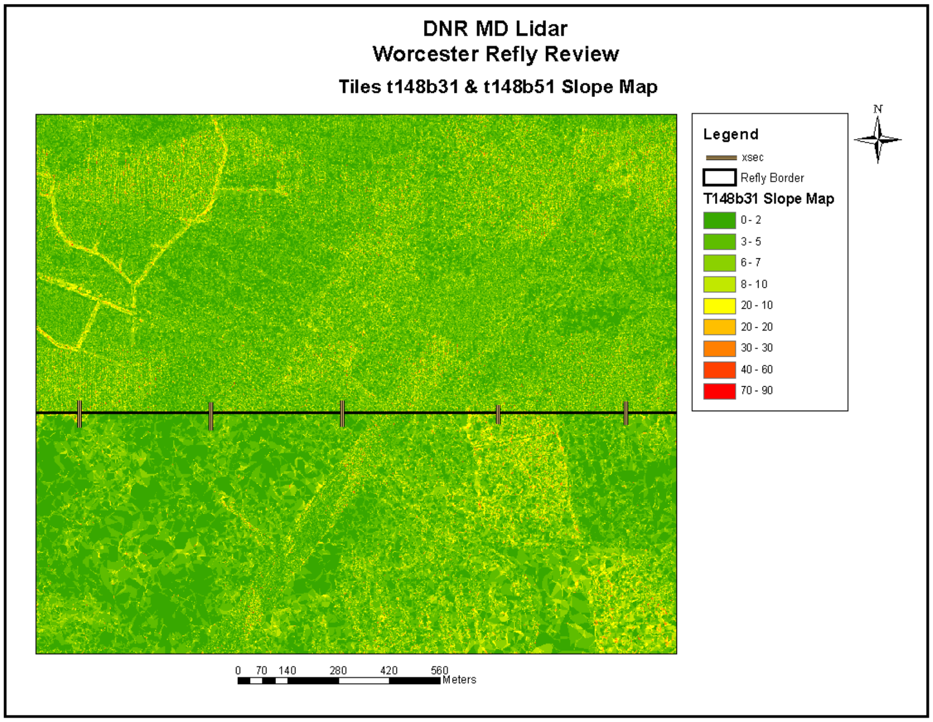
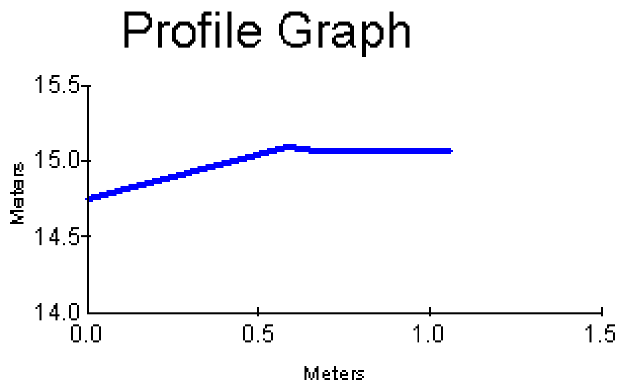
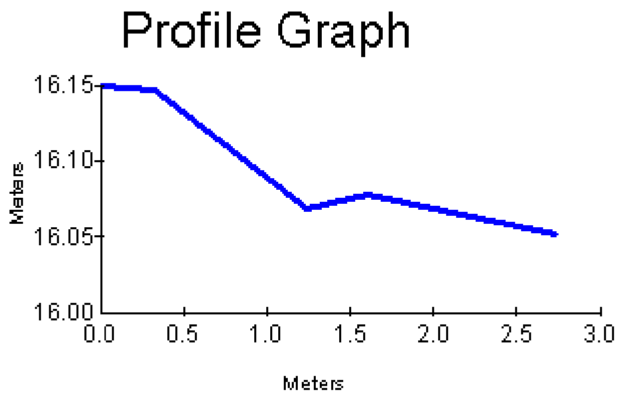


Figure 9 – Slope map between the re-flight area (top area) and the existing re-processed data (bottom area)

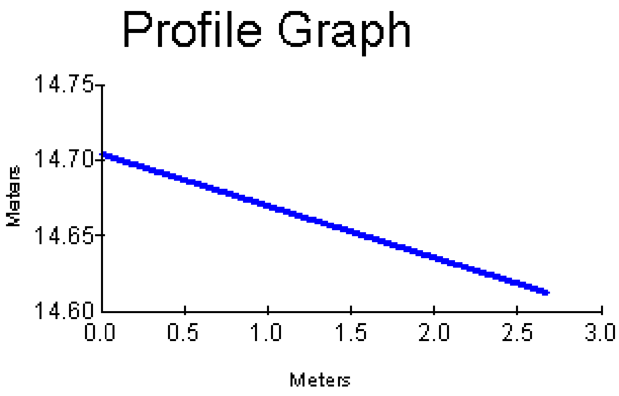
The graphs in Figure 10 illustrate cross sections (Figure 9 for locations) between the two dataset TIN's. Overall the data fits relatively well with a slight miss-matching between edges. In areas of gradual slope there are no apparent large vertical shifts but further analysis is recommended.



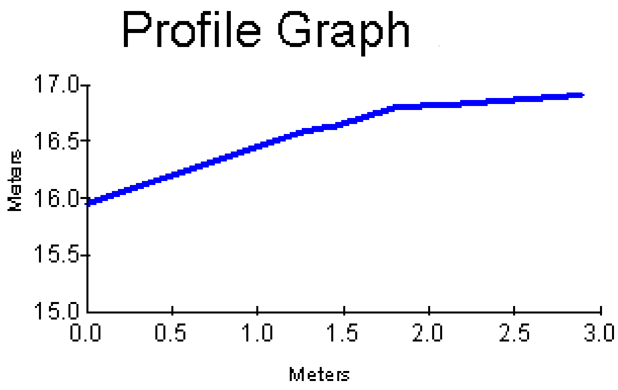
X-Sec 1



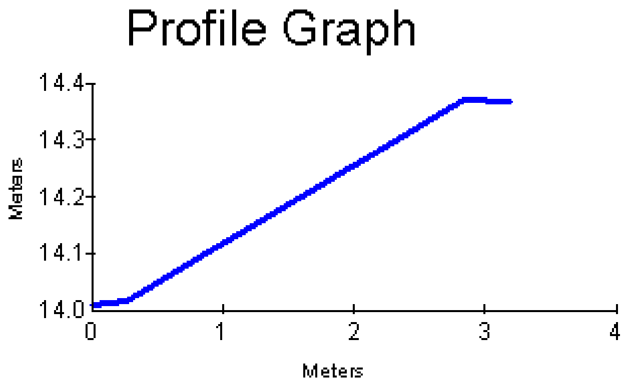
X-Sec 2



X-Sec 3



X-Sec 4



X-Sec 5

Figure 10 – Cross sections between re-flight area and re-processed data

Qualitative Analysis

The original data exhibited a high degree of data points due to the high repetition laser system (50 kHz). However since the data did not perform as well in penetrating the vegetation canopy, a large number of points were removed from the dataset with the CCS processing. This lower number of data points could impact the data quality if there are insufficient points to support the intended product. The following figures of density maps examine the original processing with it's associated density and the re-processed data with it's newly computed densities. Figure 11 illustrates the original processing of the data, Figure 12 illustrates re-processed data, and Figure 13 the delta between two processing methods. It is clearly seen that the CCS methodology has fewer data points in the vegetated areas and equal value in the open terrain while still having a higher degree of accuracy than the original processing.

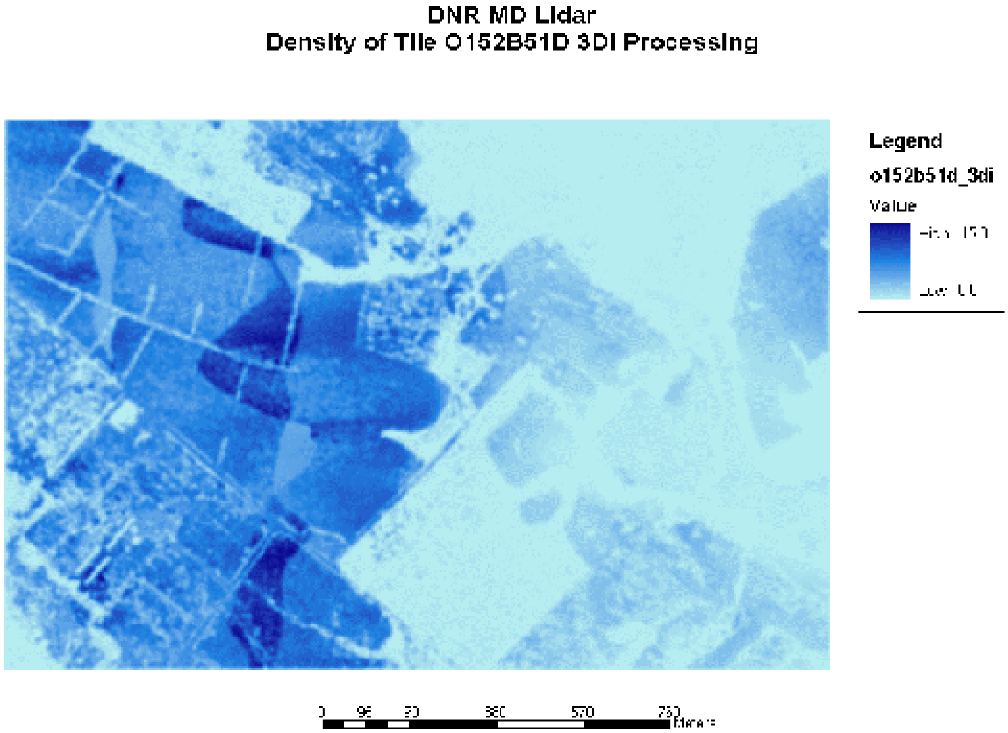


Figure 11 – Density of data points from original processing (3Di)

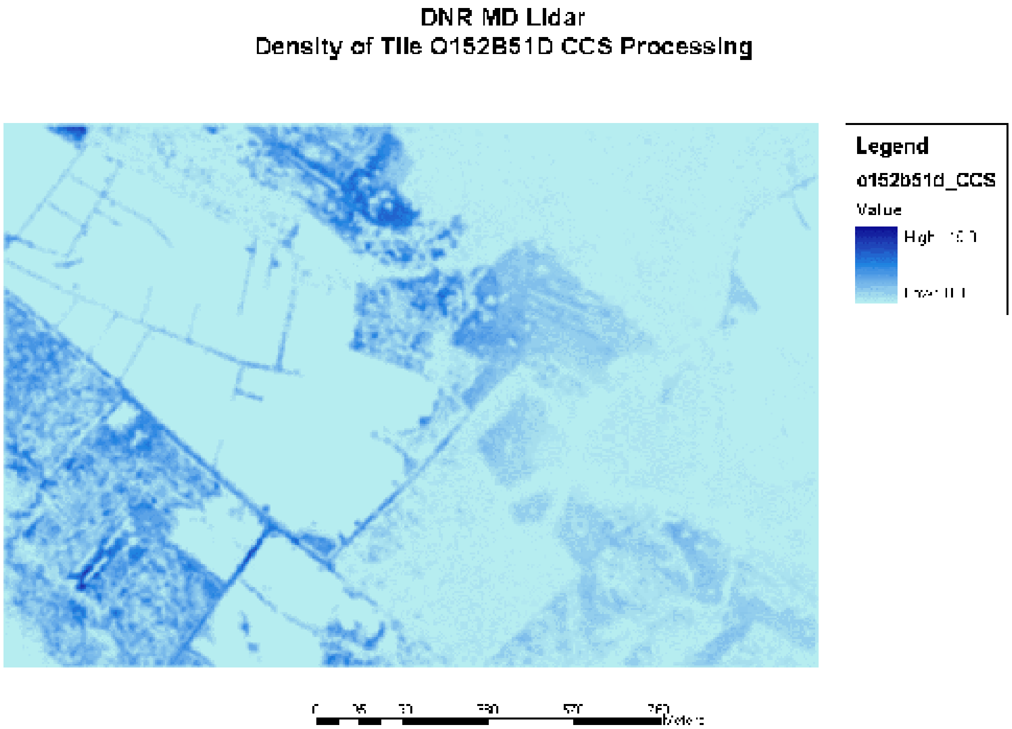
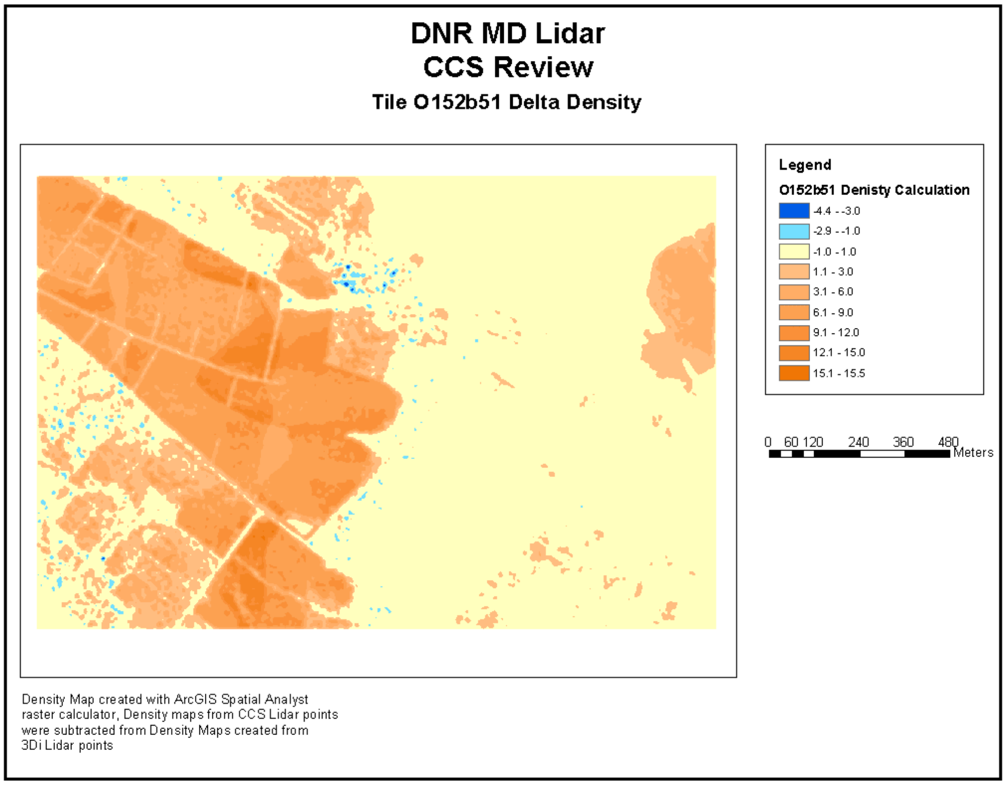


Figure 12 - Density of data points from re-processing (CCS)



Density Map created with ArcGIS Spatial Analyst raster calculator, Density maps from CCS Lidar points were subtracted from Density Maps created from 3Di Lidar points

Figure 13 – Delta of density maps (3Di minus CCS)

The areas of less data points are easily correlated to that of Weeds/Crop whereby the LIDAR had the most difficult time penetrating to the surface. This is especially true in the agricultural fields. In this scenario having less points may yield a better answer but it also introduces large data voids. In some areas these voids are quite large (>5 acres – Figure 14) and will have to be addressed on a product-need basis. This example identifies data voids but these voids are located in vegetated fields where hydraulic cross sections would normally not be located for flood studies. Since these fields are regularly plowed, it is safe to assume that they are relatively flat with slight modulations.

FEMA guidelines state; data voids may be acceptable if they are outside the floodplain and if interpolation methods can be used to fill the voids (i.e. hydrologic modeling). If the data void is within the floodplain where hydraulic modeling is performed, then the size and location will have a bearing on whether additional surveys will be required to fill the voids. In this scenario, voids greater than 1 acre where representative cross sections may need to be cut may need supplemented surveys unless other equally acceptable areas exist for cutting of representative cross sections. This decision process would normally include the FEMA Lead to decide if the additional costs are warranted or if the alternative methods are acceptable.

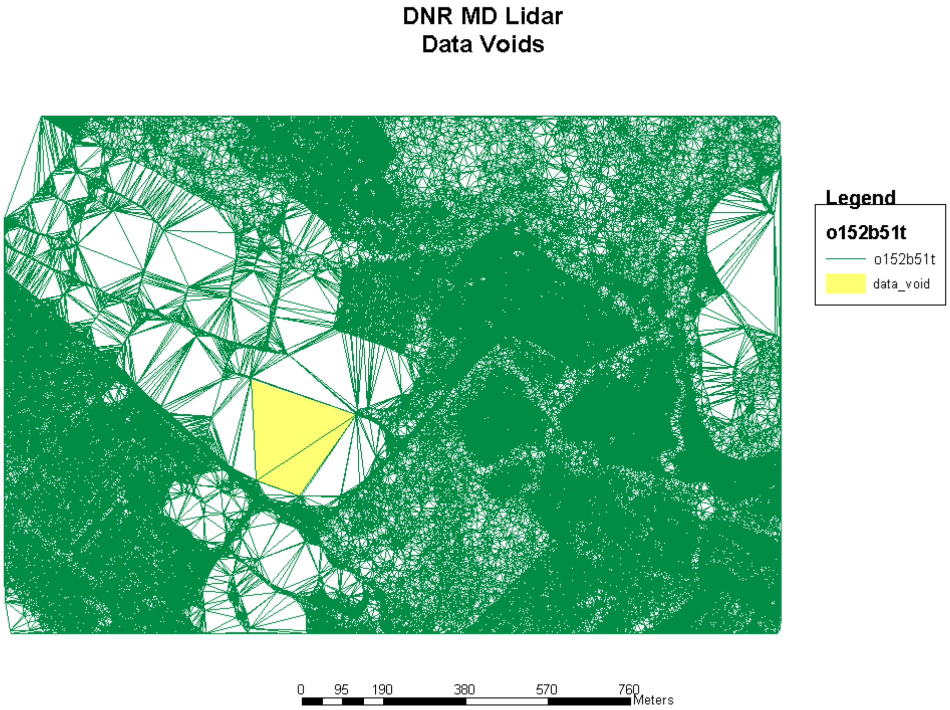


Figure 14 - TIN edges illustrating density and lack of density leading to data voids. The area in yellow is approximately 5.6 acres.

Data Cleanliness

With the original processing, data cleanliness was a major concern. The LIDAR did not penetrate the vegetation canopy as well as expected but also the algorithms used to produce a bare-earth terrain did not perform adequately. As stated previously, the data exhibited suitable accuracy in bare-earth terrain but not in vegetated areas. The CCS methodology has addressed the cleanliness issue from an algorithmic software point of view, but the data still exhibits a level of inherent "noise" which appears to be sensor related. This noise level can be seen specifically in vegetated areas where the relative accuracy of LIDAR shot to shot is slightly higher than typically expected. This relative accuracy could again potentially be from a sensor that is not fine tuned to record and identify the peak of the laser pulse. If the laser return pulse comes back slightly flat (due to the weak amount of energy reflected from the terrain), then the elevation can be slightly off but still within specification.

Although there is an inherent noise level, the data is vastly improved from the first processing methods. Figure 15 illustrates a Hillshade with a TIN Grid image used to help identify the surface roughness through illumination. This methodology aids in identifying potential artifacts by using the intensity of the grid with the aspect of the surface. Figure 16 illustrates the same tile viewed in 3 dimensions. Here the noise level in the vegetated areas is slightly higher (see profile graphs Figure 17 and Figure 18), and the stream section can be clearly identified.

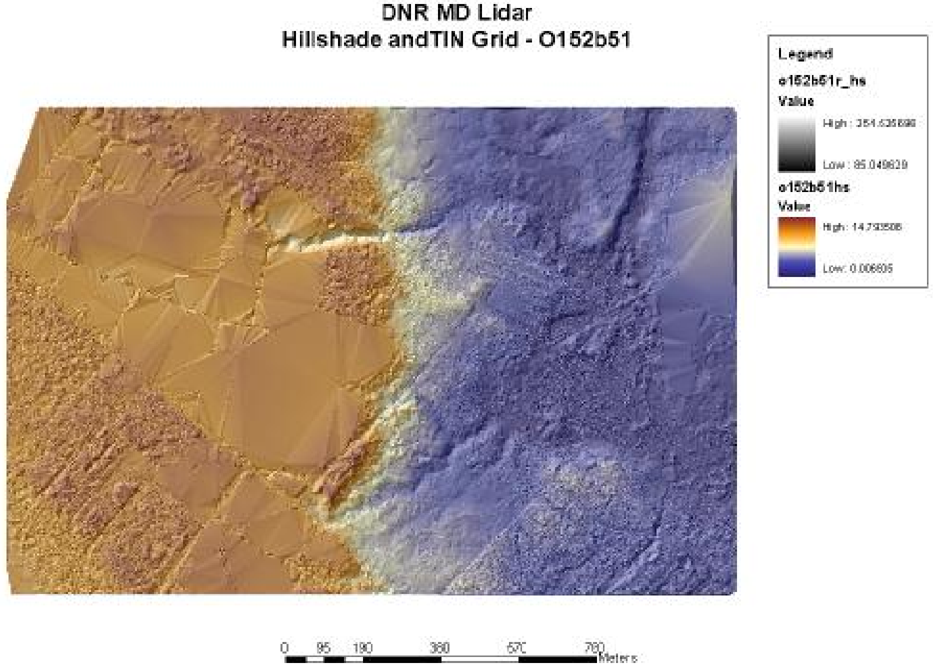


Figure 15 - Hillshade overlaid on TIN Grid

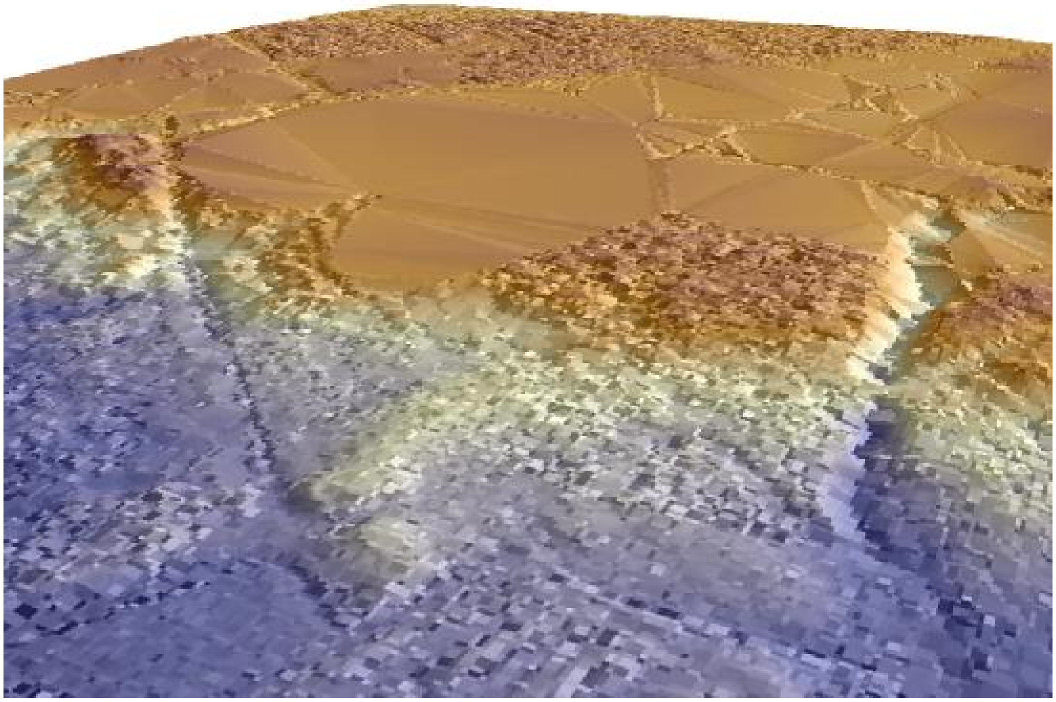


Figure 16 – Hillshade overlaid on TIN Grid viewed in 3D.

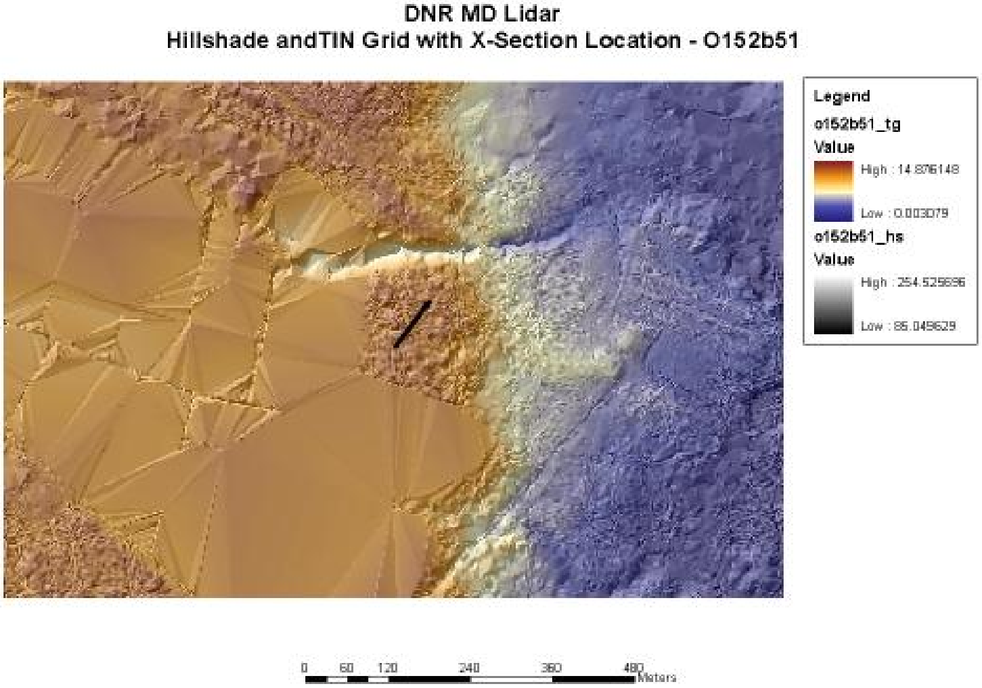


Figure 17 – Cross section location to test "noise level" of LIDAR

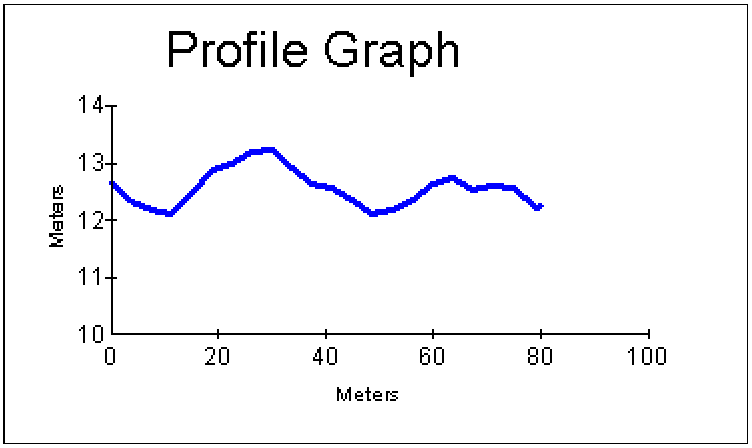


Figure 18 – Profile Graph of vegetated area.

Figure 19 and Figure 20 reiterate the findings of the review of this data. The stream channels are clearly identified while the noise level is slightly elevated within the 15 cm level.

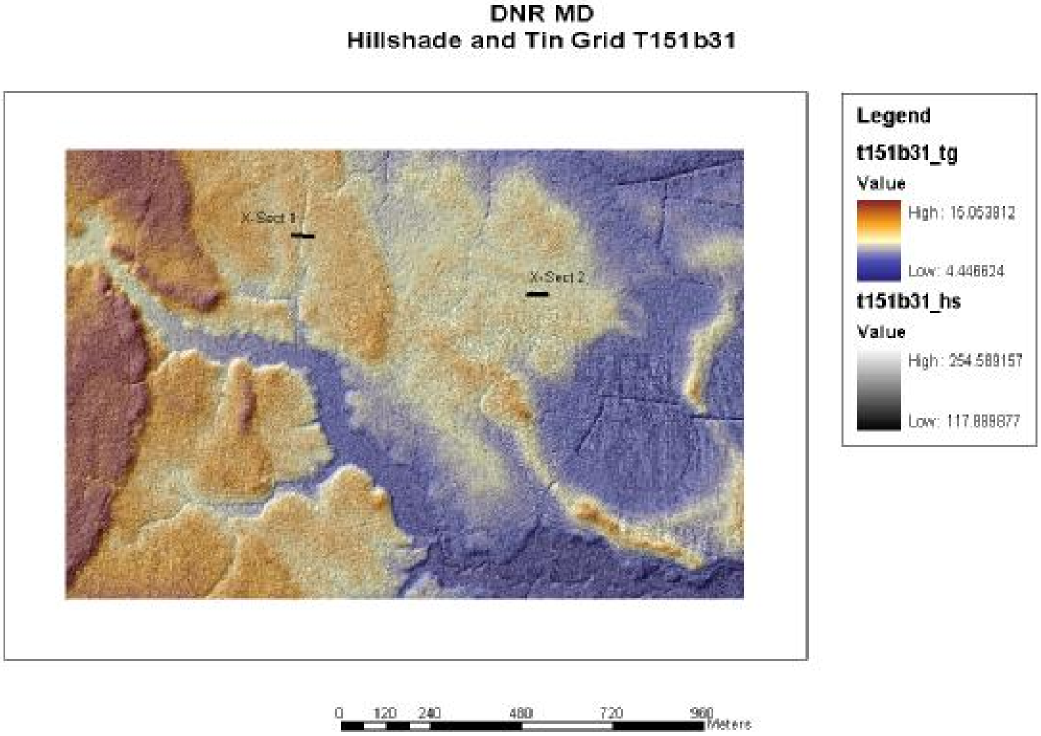


Figure 19 - Cross section locations to test stream channel geometry and "noise level" of LIDAR (Tile T151B31)

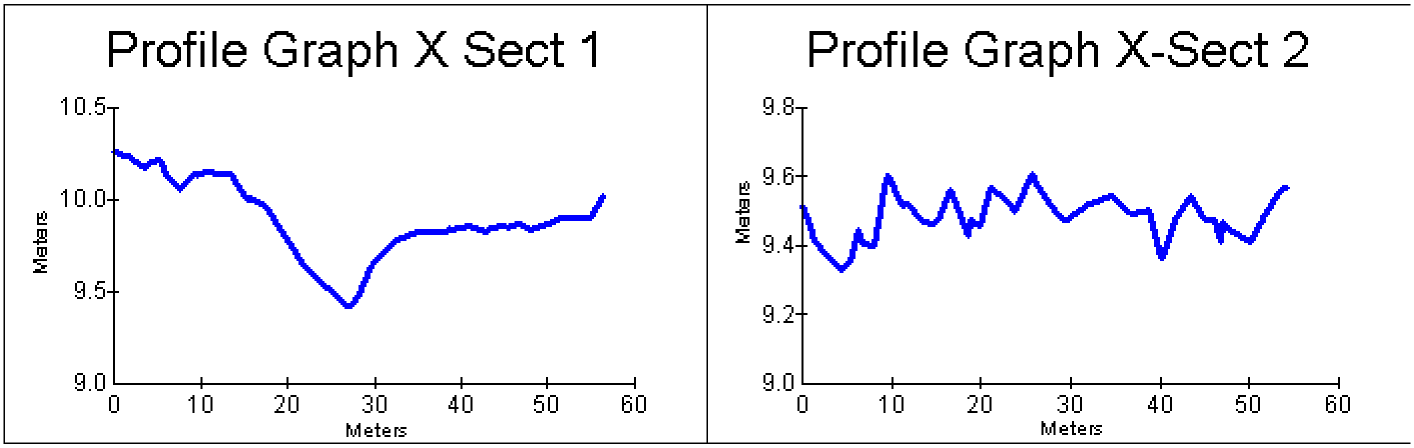


Figure 20 - Cross section of stream channel and vegetated area

Conclusion

The LIDAR data as processed by CCS is greatly improved from the original data set. Although the data does not fully satisfy the 18.5 cm RMSE criteria or the NDEP criteria in vegetated categories, the data should be suitable for most topographic needs. This would include contour mapping (with the appropriate interval) and floodplain mapping for both hydrology and hydraulics. Statistically the data is skewed due to a higher number of outliers. Further review of the accuracy of the ground truth QA/QC surveys may yield better results from the DNR and the Independent Surveyor's checkpoints. Additional checkpoints would also be beneficial for further assessment if a higher accuracy is desired with this dataset. The LIDAR data also contains a higher noise level than is typically seen but can be further smoothed through additional processes. It is highly recommended that the metadata reflect the intended use of the data collection and it's accuracy as well as outlining the associated issues with the data collection and re-processing of the data to a bare-earth model.