QUALITY ASSURANCE AND POTENTIAL APPLICATIONS OF A HIGH DENSITY LiDAR DATA SET FOR THE CITY OF NEW YORK

Sean C. Ahearn, Director
Hyo Jin Ahn, Post Doctorate
Center for Advanced Research of Spatial Information (CARSI)
Hunter College-City University of New York
New York, New York 10065
sahearn@hunter.cuny.edu
hyojin@hunter.cuny.edu

ABSTRACT

What is perhaps the highest density data acquisition by a Light Detection and Ranging (LiDAR) system was made for the New York City in the spring of 2010. The area coverage 937 square kilometers, and the total number of collection points were 15 billion with an average point density of 8-12 points per square meters. The specified accuracy of the data was 9.5 cm in x, y and z. Four different assessments of quality were conducted as part of this project: vertical accuracy, horizontal accuracy, point density, and classification accuracy. Vertical quality assurance of these data were conducted by comparing the LiDAR point cloud to 1,772 survey level ground points with a 2 cm root mean square error (RMSE). Horizontal accuracy was conducted using photogrammetry derived data of building outlines. Density was analyzed with a GIS and classification was analyzed through visual examination, building overlays, image intensity and ortho-image analysis. Studies have shown that careful selection of points from the point cloud for comparison with the Ground Survey Points is necessary to insure comparisons are made with point on the ground. This paper presents the quality assurance techniques for evaluating the quality of LiDAR data; results of that analysis; and a discussion of several promising applications including: solar potential mapping, 3D building modeling, urban surface mapping, contour mapping, storm surge control, forest management and emergency management.

KEY WORDS: LiDAR, quality assurance, vertical accuracy, horizontal accuracy, 3-D point cloud, DEM, DSM, nDSM

INTRODUCTION

LiDAR technology provides 3-dimensional geospatial data that make possible to generate accurate position of features, structures, and infrastructures over large areas in a high density level. In near future features on line environments and the urban modeling and planning tools will be based on liDAR (Harrup and Lato; Jochem et al., 2009).

For the Solar New York Project run out of the Center for Sustainable Energy, the City University of New York, the air borne LiDAR data acquisitions were made between April 14th, 2010, and May 1st, 2010 by Sanborn Inc., Colorado Springs CO. The project area is approximately 937 square kilometers (362 square miles) and is subdivided based on the (New York Digital Orthoimagery Program) tile index as shown in Figure 1. The collected LiDAR raw data consisted of 15 billion points collected for a nominal point spacing of 8-12 points per square meters based on 4 returns per pulse along with an intensity value for each return. This data set marks one of the highest density LiDAR data ever made in the United States. The quality review for all unclassified raw data (binary LAS format version 1.2 format), classified data (binary LAS format version 1.2 format), and the bare earth data (ASCII x, y, z, and Intensity) as well as testing potential model constructions such as Digital Elevation Model (DEM), Digital Surface Model (DSM), Normalized Digital Surface Model (nDSM), Building Extraction model were carried out by the Center for Analysis and Research of Spatial Information (CARSI), Hunter College-The City of New York.

The primary task of CARSI includes the assessment of the LiDAR data in both quantitative and qualitative checks to ensure accuracy. The quantitative assessment focuses on the vertical accuracy of LiDAR data. The qualitative assessment included a completeness of tile coverage, density, qualitative horizontal accuracy, and the classification accuracy delivered from Sanborn. In the part one of this paper, these quality assessment techniques that CARSI lab used will be presented. In part two, potential derivable LiDAR models and its applications will be discussed.

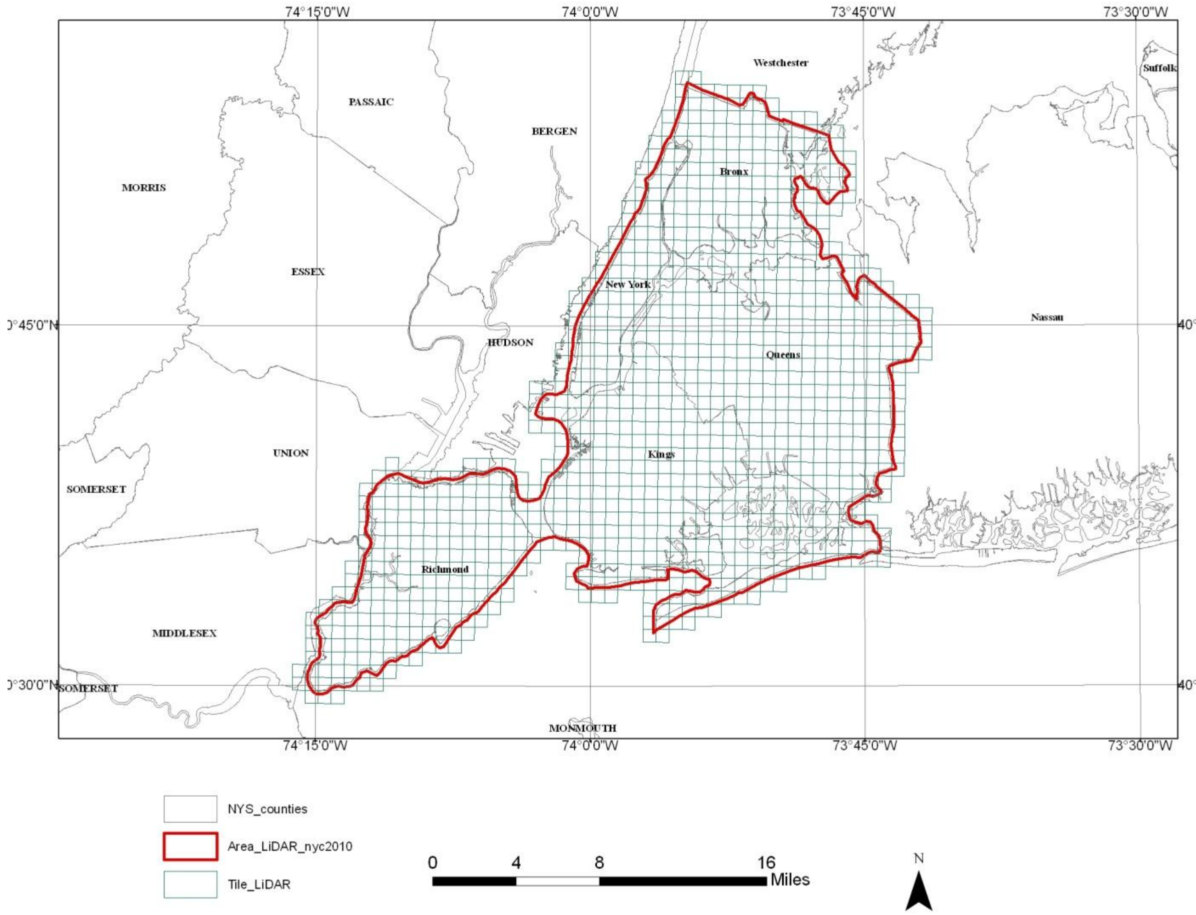


Figure 1. The project area of New York City (red polygon) and Lidar tiles.

QUALITY ASSURANCE ASSESSMENT TECHNIQUES

1.Vertical Accuracy

The method chosen for vertical accuracy assessment was to compare the elevation z value from laser point cloud bare earth data set to ground survey checkpoints. criteria for ground checkpoints for vertical accuracy test required them to be at least three times more accurate than the data being tested (National Geodetic Survey, National Oceanic & Atmospheric Administration, 2009; ASPRS 2004). The ground checkpoints used for our assessment have a 2 cm RMSE which meets these requirements given that the requirements for the LiDAR are that they have an RMSE of 9.5 cm.

To assess the vertical accuracy, CARSI used a total of 1,772 ground survey checkpoints provided by ConEdison Inc. The process of the vertical accuracy assessment and generation of statistics were accomplished by an analysis tool *Point Query Utility* in the LiDAR processing software Quick Terrain Modeler. The survey checkpoint's X and Y location was overlaid on the bare-earth point clouds, and the LiDAR Z value and the difference between the ground checkpoint elevation and LiDAR Z value were recorded.

In Table 1 summary statistics of RMSEs, mean differences, and the number of ground checkpoints used for each borough is presented. The smallest RMSE shows is in Manhattan at 6.69 cm based on 196 checkpoints, and the largest RMSE appears in Brooklyn with 8.95 cm based on 721 checkpoints.

The scatter diagrams shown in Figure 2 depict the distribution of Δz and the statistics of the associated errors in each borough of New York City.

Table 1. RMSE and mean of Δz between the ground checkpoints and LiDAR z values for NYC 5 boroughs.

NYC 5 Boroughs	RMSE (cm)	Mean of Elevation Differences (cm)	# of Ground Checkpoints	
Manhattan	6.69	-2.15	196 out of 198	PASS
Queens	6.80	-3.41	574 out of 590	PASS
Staten Island	7.64	2.02	100 out of 100	PASS
Bronx	7.30	-3.27	131out of 132	PASS
Brooklyn	8.95	-5.23	721 out of 737	PASS
	MEAN=7.48	MEAN =-2.41	TOTAL# used =1722 out of 1757	PASS

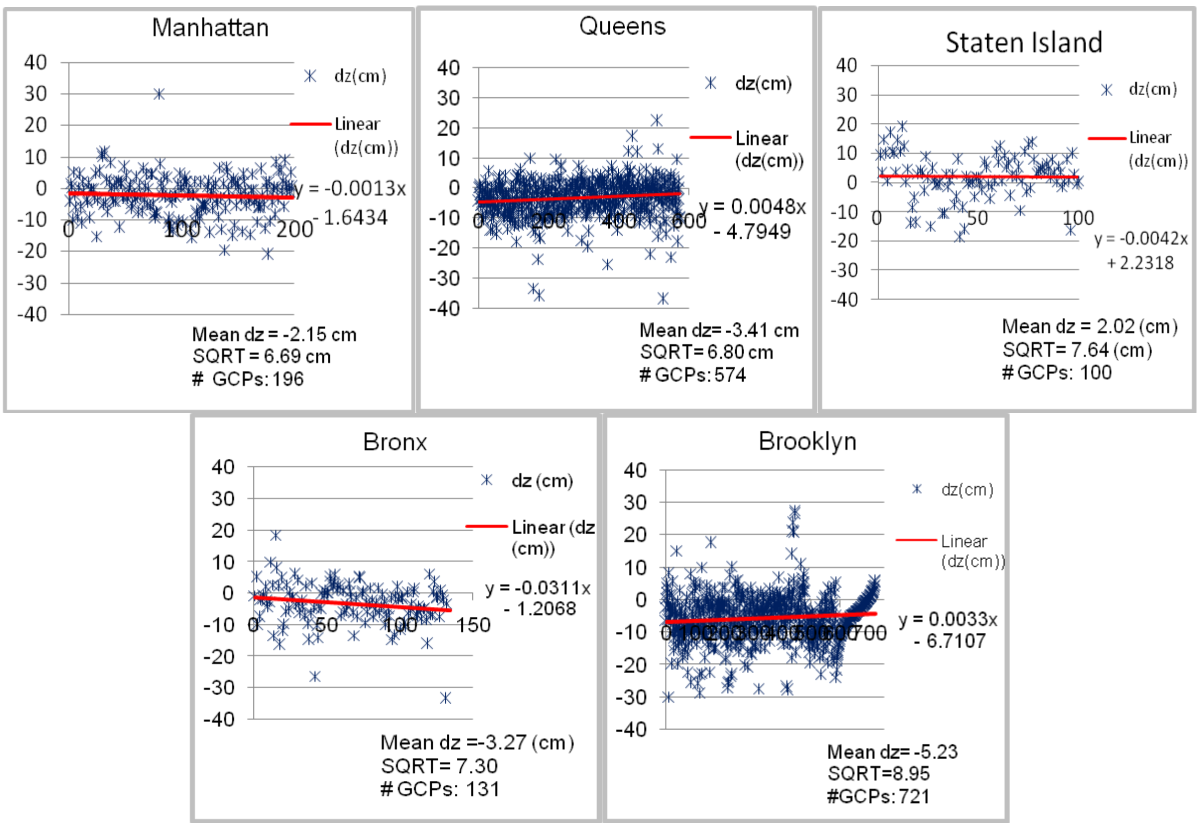


Figure 2. Scatter diagrams of the mean of Δz between the ground checkpoints and the LiDAR z values for each borough in NYC.

Vertical Accuracy Conclusion

Based on the results from comparison with the ground checkpoints, CARSI's review of the vertical accuracy of the Sanborn LiDAR data confirms that the dataset meets and exceeds the accuracy requirement as specifications.

2. Density Assessment for LiDAR Tiles

Assessing the required point density, each tile was queried using all return points of unclassified raw data to extract statistics for the whole tile and then area statistics which excluded water bodies. This process was completed using the analysis tool in Quick Terrain Modeler. The analysis results were transferred to the ArcGIS software for mapping. The data show the anticipated size range at 8-12 points per square meter except the areas near river or coast lines (Figure 3). However, most of these areas with the exception of two tiles, also met the range when the further analysis carried out with the areas containing coastal lines.

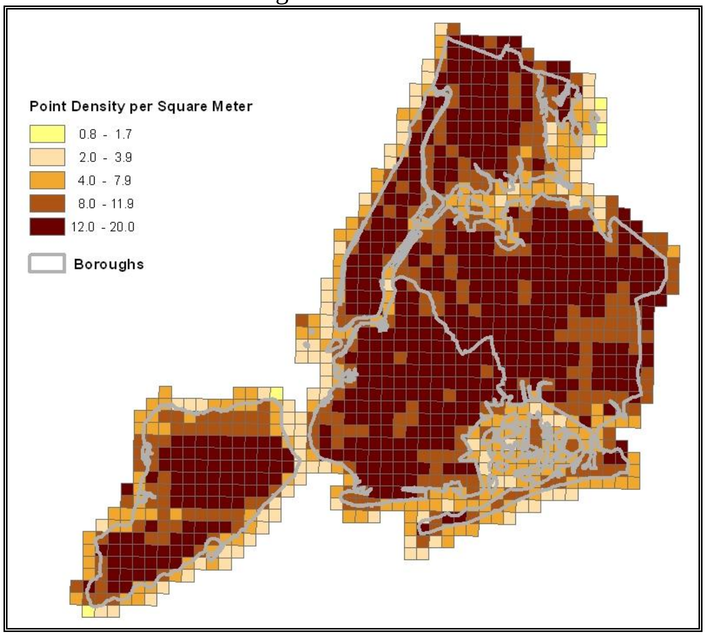


Figure 3. LiDAR point density per square meter in NYC.

3. Horizontal Accuracy

Testing horizontal accuracy is known to be more difficult than vertical error assessment. This is because the resolution of the reference data is coarser than LiDAR data or that the land surface often lacks well-defined topographic features (ASPRS, 2005).

The horizontal accuracy report is often less often required, however, a few attempts have shown new assessment techniques or module development for automatic assessment of the horizontal errors and for effective adjustments (Toth et al., 2007; Ray and Graham, 2008). Their approaches were to use pavement markings as ground survey control points, and to identify them through the overlay on the intensity image of the LiDAR data as a LiDAR ortho. This situation is analogous to the accuracy assessment of traditional mapping, but the difference is identifying the precise location of the paint marks that were placed with ground survey records in advance. In addition to using the intensity of LiDAR data, there are other strategies also recommended to achieve reasonable horizontal accuracies. One of methods is overlaying the intensity image on digital orthophotos if they exist for the project area. However, to be valid in an accuracy assessment, the digital orthophotos must be of sufficiently higher accuracy than the project requirements for horizontal accuracy (Meade, 2008).

CARSI decided to use the photogrammetric building elevation data released in January 2010 by the Department of Information Technology and Telecommunications (DOITT) of New York City (NYC) as a reference data set for the horizontal accuracy analysis. The following Figure 4 presents the random distribution of total 50 selected tiles over the project area. The total 200 building corner points were used to test the horizontal errors from the selected tiles which are composed of 4 points per each tile and 10 tiles per each borough. The building elevation data set acquired from the DOITT NYC 2010 is in a geographic information system (GIS) of vector format. These GIS vector polygons were draped over the LiDAR point cloud image using the software Quick Terrain Modeler. In the selection of building corners from the LiDAR point cloud image as well as from the overlaid vector polygon, some rules were applied especially for the selection on LiDAR point cloud image:

- ✓ Choose the building corner has no raised edge or wall boundary
- ✓ Choose the building corner having overlap pass (see Figure 5 (a))
- ✓ Choose the building edge line appeared in straight line, not curved or round
- ✓ Avoid choosing the building has overhanging near the roof top or side (see Figure 5 (b) & (c))

Figure 5 shows examples of how to choose a point for the horizontal accuracy check.. Figure (a) shows several warehouse buildings that are relatively simple roof tops with low heights. In this case, the yellow circle with overlap passing area (bright yellow as ground, and bright white as building roof top) has higher density points than the black circle corner which has less density without overlap pass. Therefore, the building corner selection at yellow circle is a better accuracy test case. Figure 5 (b) and (c) are the same site, but one is 2-D and the other is 3-D These depict the case of multiple heights of buildings with overhanging near roof edge. This type of building edge has been observed very often over Manhattan metropolitan area in NYC. These sites need to be avoided.

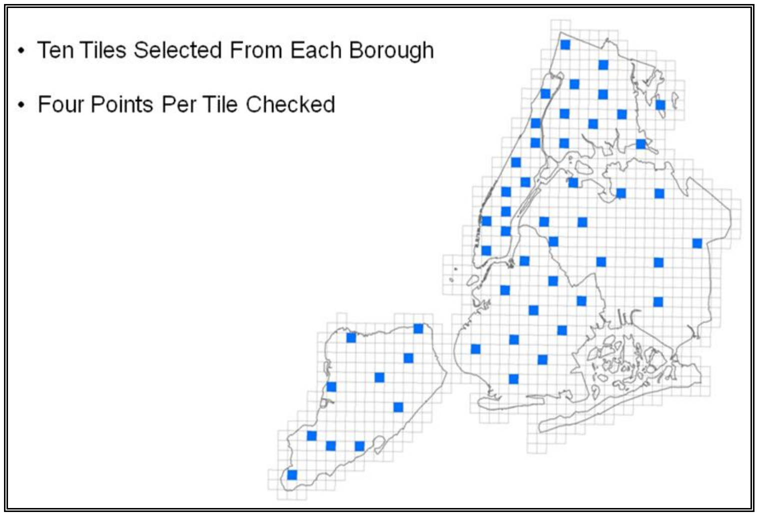


Figure 4. The selected LiDAR tiles for horizontal accuracy assessment in .

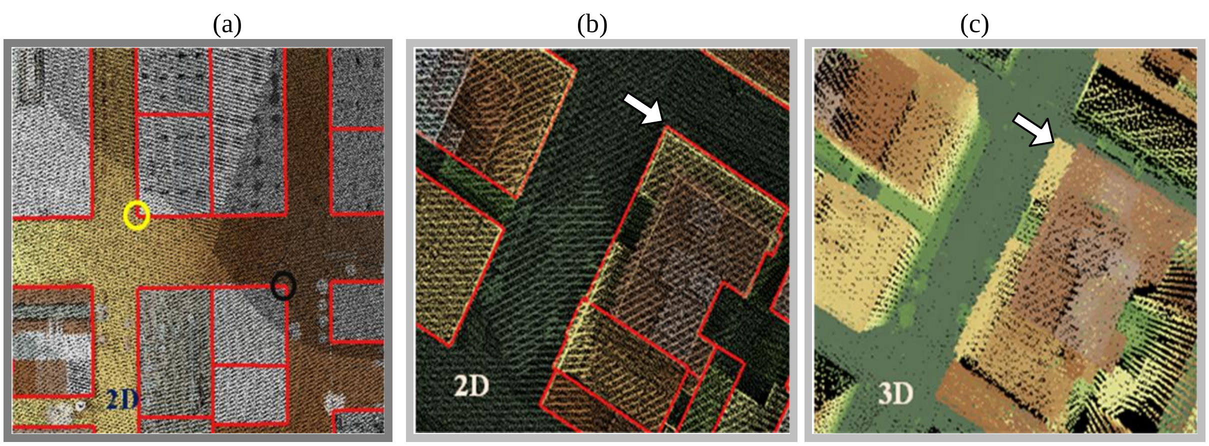


Figure 5. Building polygons overlaid on the LiDAR point cloud image: (a) the yellow circled corner over the bright yellow area having overlaps represents a better choice for x, y, and z collection corner than the black circled corner where the point density is less (brown color region). (b) & (C) are the same building sites, but (b) is a 2D point cloud with vector overlay, and (c) shows a 3D of point cloud image. The building corner pointed by white arrow is a bad choice case of irregular edge line.

The results of the horizontal accuracy analysis for each borough are summarized in Table 3. The overall RMSE for entire city depicts 33.08 cm which is close to 1 ft, and the average of distance difference between LiDAR point and the building corner point of photogrammetric data shows 27.75 cm. The smallest distance difference depicts in Queens borough with 13.17 cm, respectively. The greatest distance difference appears in Manhattan borough with 34.55cm, which can be explained by most of tall buildings that generated errors in photogrammetric data. The accuracy of the photogrammetric data was specified at 45 cm or better, so these results are the best that can be expected.

Table 3. RMSE and Mean of Distance Differences between Building Footprint and LiDAR Data for Each Borough.

NYC Borough	RMSE (cm)	Mean of Distance Differences (cm)
Manhattan	39.72	34.55
Queens	15.07	13.17
Staten Island	37.53	33.96
Bronx	37.42	33.10
Brooklyn	29.29	23.97
NYC	MEAN=33.08	MEAN =27.75

The following histogram in Figure 6 shows the frequencies of distance differences in 12 groups. The highest frequency appears in the group of 15 cm difference in X and 23 cm difference in Y.

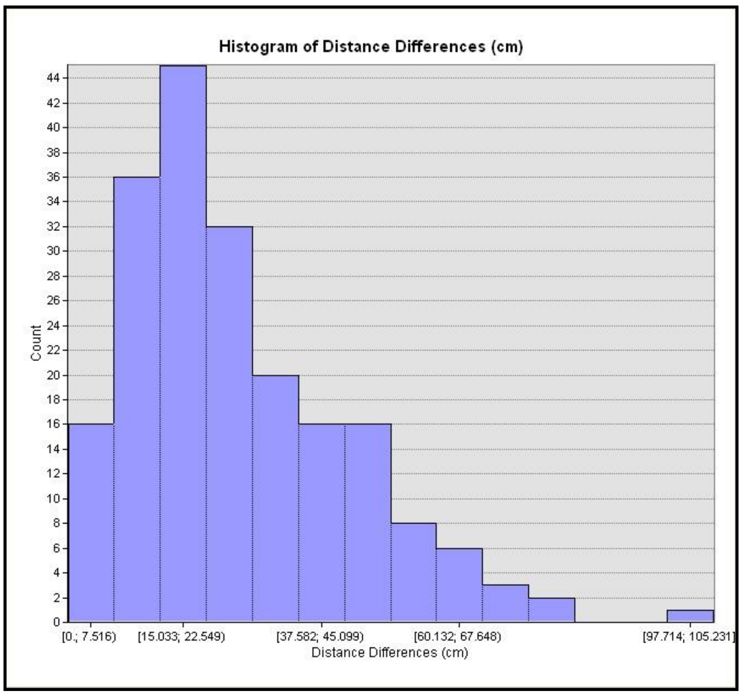


Figure 6. Frequency distribution histogram of distance differences in X and Y coordinates.

4. Classification Accuracy

Among the LiDAR data deliverables , one of data set called “Classified LAS All Returns” contains 4 classes: Class 1= Unclassified (i.e. include things above ground surface like buildings, bridges, trees, and other manmade structures), Class 2 = Bare earth (ground), Class 7=low point noise, and Class 12=overlap.

The first quality check was to find any missing spots (omission error) in bare earth through visual examination. Figure 7 shows two examples of omission errors found over the Manhattan area. In this figure, the left column displays only bare earth class point cloud image with the red box as an indication of missing ground part. The right column shows the same site as the left column but with unclassified Class 1 and intensity image. The second quality check was to assess whether bare earth classes included any non-ground portions (i.e. commission error of Class 2). Some examples of commission errors found are presented in the blue box of Figure 8.

Among those randomly tested tiles, about 5 out of 100 tiles showed either omission or commission error in the Bare earth class. Considering the coverage of the error area in comparison with the whole tile is minimal, the accuracy was within those specified in the contract. Though further editing is necessary to get an accurate DEM in these problematic areas.

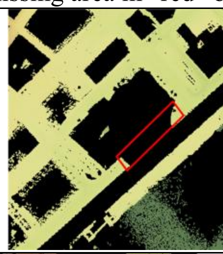
Bare earth (Class 2): missing area in "red" box	Unclassified (Class 1) with Intensity image
	
	

Figure 7. Bare earth misclassification: examples of omission error.

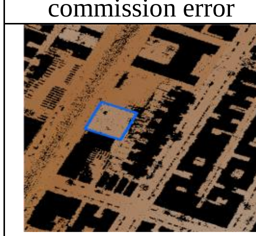
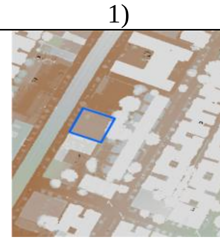
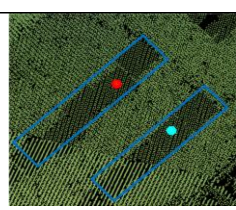
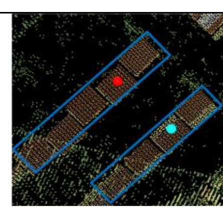
Bare earth (Class 2) commission error	Unclassified (Class 1)
	
	

Figure 8. Bare earth commission errors.

5. Derived Products

With the quality assured LiDAR data, we can generate a wide range of derived products. These include: a digital elevation model (DEM) , a digital surface model (DSM) , a normalized DSM (nDSM), and an intensity image. Derived DSM or nDSM is useful for building extraction or rooftop segmentation through analyses of slope, aspect, and z deviation. Further applications can be for flood simulation modeling, or emergency management through generating possible helicopter landing zone or convoy detections.

Our primary application was to use the DSM to generate the solar insolation for the entire city at a resolution of 1 meter. This insolation map was then used to calculate solar potential for all 1 million buildings in the City of New York as part of the US Department of Energy Sustainable Cities initiative. The Solar Insolation Map was an input to an application on a NYC SOLAR MAP project web porthole (Figure 9).

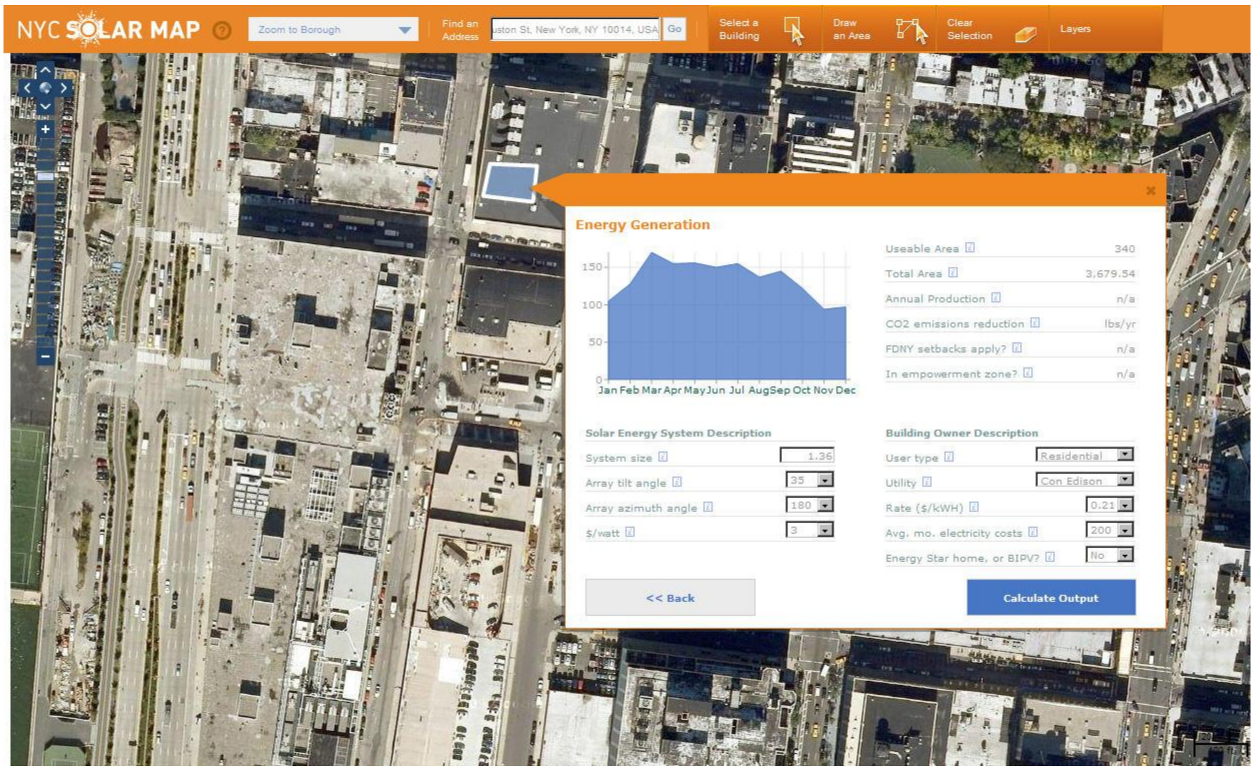


Figure 9. NYC SOLAR MAP showing energy profile for a user-generated polygon on a selected building

ACKNOWLEDGEMENTS

This project was supported by grants from the US Department of Energy and the City of New York Department of City-wide Administrative Services (DCAS).

REFERENCES

ASPRS Guidelines: Vertical Accuracy Reporting for Lidar Data, version 1.0, 2004.

ASPRS LIDAR GUIDELINES: Horizontal Accuracy Reporting, March 7, 2005. http://www.asprs.org/society/committees/standards/Horizontal_Accuracy_Reporting_for_Lidar_Data.pdf.

Harrup, R. and M. Lato. An Overview of LIDAR for Urban Applications. http://geol.queensu.ca/faculty/harrap/pdf/WhatIsLIDAR_release1.pdf.

Jochem, A., B. H¨ofle, M. Rutzinger, and N. Pfeifer, 2009. Automatic Roof Plane Detection and Analysis in

Airborne Lidar Point Clouds for Solar Potential Assessment, *Sensors* 2009, *9*, 5241-5262; doi:10.3390/s90705241.
Meade, M.E., 2008. From the ground up: Horizontal accuracy assessment in LiDAR. http://www.pobonline.com/CDA/Articles/Column/BNP_GUID_9-5-2006_A_10000000000000361586
NDEP Guidelines for Digital Elevation Data, version 1.0, National Digital Elevation Program (NDEP), 2004.
National Geodetic Survey, National Oceanic & Atmospheric Administration, U.S. Department of Commerce, 2009. Light Detection and Ranging (LiDAR) RequirementsNGS.
Ray, J. A., and L. Graham,2008. New horizontal accuracy assessment tools and techniques for LiDAR data. ASPRS 2008 Annual Conference, Portland, Oregon, April 28-may 2, http://geocue.com/resources/geocue_LIDARHorizontalAccuracy.pdf.
Toth, C., E. Paska, and D. Brzezinska, 2007. Using pavement markings to support the QA/QC of LiDAR data. In: Stilla U et al (Eds) PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (3/W49B).