COMPUTING SOLAR ENERGY POTENTIAL OF URBAN AREAS 
USING AIRBORNE LIDAR AND ORTHOIMAGERY

Ryan Hippenstiel  
The Pennsylvania State University  
John A. Dutton e-Education Institute  
2217 Earth & Engineering Sciences Building  
University Park, PA 16802  
Email: rdh154@psu.edu

Jeffrey R. S. Brownson  
Department of Energy and Mineral Engineering  
The Pennsylvania State University  
University Park, PA 16802  
Email: nanomech@psu.edu

ABSTRACT

Alternative energy sources are a necessity, as is the ability to analyze them. Photovoltaic (PV) potential is commonly studied over large areas, yet implementation is often desired at a local scale. To support research and tools developed to study PV potential in urban areas, this project analyzed remotely sensed data, specifically LiDAR (Light Detection and Ranging) and orthoimagery, to extract 3-dimensional building and vegetation features for use in existing modeling tools. LiDAR and orthoimagery will allow a more efficient and geo-referenced way for users to compute solar potential for individual or clusters of locations in their selected areas of interest. This project has tested different extraction tools and concepts, identifying those that can easily be incorporated into a Geographic Information System (GIS). Parameters of feature extraction were tailored to facilitate shading analysis and eliminate areas unsuitable for PV systems. Extraction of buildings and high vegetation, and creation of 3D models of usable areas were investigated. From this, a reliable workflow is being developed to serve as a tool for future use. The direction of this project is important to analysts desiring accurate, geo-referenced data for input into various models, but will specifically support the on-going research in inter-building shadowing effects for energy simulations and solar technology deployment.

1. INTRODUCTION

According to the U.S. Energy Information Agency, between 2004 and 2008, the United States’ energy consumption produced by solar energy nearly doubled (1). Unfortunately, it was still only 0.1% of overall consumption. With the sun continuing to rise each day and a continuing need to develop and produce alternative energy sources, many groups and governments are investing time, money, research, and development into solar energy. One of the biggest challenges these groups face is that current processes and technology limit the cost-effectiveness of solar energy use. This makes it very difficult to efficiently design photovoltaic (PV) systems for single homes or buildings. As the size of a geographic area rises, it makes it easier to design systems that will actually have a reasonable return on investment for the users. One of the biggest challenges of completing PV analysis at regions of this size is that much of the existing data and analysis systems are at too large of a scale to provide reliable data at the smaller level. Because photovoltaic potential can be influenced by local conditions, this type of wide-scale modeling does not always provide information accurate enough to analyze customized areas. So there is a strong demand for tools to easily and effectively analyze data at a regional level chosen by the user. In response to the need for smaller-area modeling methods, many simulation tools have been developed with more concentration on increasing the amount of data inputs available. Even these do not take into account local changes in solar radiation, climate, and topography on the
specific geographic region selected. A need for effectively using accurate, geo-referenced data in the development of urban models to compute PV potential is very high. A great example of this is LiDAR (Light Detection and Ranging) collection.

2. **LIDAR DATA**

Light Detection and Ranging is a process of collecting geographic data by using a series of laser pulses and their subsequent returns from reflection of surfaces they strike. This technology uses many types of scanners and processing routines, but simply, the returns are compared to the original pulse and origin to create a “point cloud” of features of the scan area.

LiDAR data is collected by a sensor emitting a laser pulse in the near-infrared band of the electromagnetic spectrum which is reflected by the surfaces it strikes. Using the speed of light and the reflected pulse’s time of return, the position of the object struck relative to the sensor can be computed. The sensor is mounted on an aircraft equipped with a Global Positioning System (GPS) receiver to record its location. In addition to GPS, the aircraft is also equipped with an inertial measurement unit (IMU) to record the orientation of the aircraft while flying. This includes recording the pitch, roll, and heading which in combination of the GPS can provide speed and acceleration. These two units overall determine the exact position and pointing angle of the sensor at any given time. Using the previously mentioned calculation, the geographic position of the laser source can be used to calculate the geographic position of the objects the pulse is striking. See Figure 1 for a basic configuration diagram.

The accuracy of which these positions are collected varies based on project specifications, flight conditions, and the surfaces they are striking. In the case of this project, the source of the information for this project will be the PAMAP Program, through Pennsylvania Spatial Data Access (PASDA). The PAMAP data to be used has been processed and published in a few formats. Users can download 2-foot contours and breaklines (produced using corresponding orthoimagery), both in shapefile format. A 1 meter bare-earth DEM in a GeoTIFF format and a classified *.las (standard format for lidar data) file are also available. The LiDAR data will be the dataset used for building models of the urban area to be analyzed. The LiDAR data has an overall vertical RMSE of 0.61 ft, and a Fundamental Vertical Accuracy of 1.19 feet at a 95% confidence level. All data is referenced to the PA State Plane Coordinate System, NAD83, and the units are in feet.

In addition to comparatively good accuracies versus other remotely sensed data from airborne platforms, LiDAR also has some strong advantages. One of the more interesting is the ability to capture data at night. This helps increase the availability of flight windows, a benefit when considering that LiDAR collection can be hampered by weather conditions such as fog or clouds. A major advantage, and one that helps define the data processing, is the ability of LiDAR to collect data in tree canopy. Depending on density of vegetation, the laser pulses can travel through holes between leaves, branches, or individual trees. Therefore, whole laser pulses can strike levels of vegetative canopy or parts of a pulse can hit multiple level of canopy being recorded as first, second, third…”returns”. These returns are a valuable piece of information as it pertains to data processing. The level of a pulse’s return is very informational about the surfaces it is striking. These various returns can be sorted and filtered to provide various ranges and visualizations to enhance data. Along with return values, sensors often can collect the intensity of pulses. The value is dependent on many factors including the composition of and the angle at which a surface is struck. A highly reflective surface will have a high intensity while a paved surface may have a lower value. These qualities help in automated and human processing into standard classifications.

This point cloud is separated into main classes such as *bare earth*, *buildings*, *high vegetation*, etc. LiDAR data has been standardized so that processing produces a header file for referencing information and points in a classification
system that has been defined by the American Society for Photogrammetry and Remote Sensing (ASPRS) (3). This format ensures that LiDAR files (*.las format) are transferrable and readable with a multitude of remote sensing, classification, and GIS software packages. Table 1 shows the current standard and classes.

**TABLE 1: ASPRS STANDARD LIDAR POINT CLASSES**

<table>
<thead>
<tr>
<th>Classification Value (bits 0-4)</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Created, never classified</td>
</tr>
<tr>
<td>1</td>
<td>Unclassified</td>
</tr>
<tr>
<td>2</td>
<td>Ground</td>
</tr>
<tr>
<td>3</td>
<td>Low Vegetation</td>
</tr>
<tr>
<td>4</td>
<td>Medium Vegetation</td>
</tr>
<tr>
<td>5</td>
<td>High Vegetation</td>
</tr>
<tr>
<td>6</td>
<td>Building</td>
</tr>
<tr>
<td>7</td>
<td>Low Point (noise)</td>
</tr>
<tr>
<td>8</td>
<td>Model Key-point (mass point)</td>
</tr>
<tr>
<td>9</td>
<td>Water</td>
</tr>
<tr>
<td>10</td>
<td>Reserved for ASPRS Definition</td>
</tr>
<tr>
<td>11</td>
<td>Reserved for ASPRS Definition</td>
</tr>
<tr>
<td>12</td>
<td>Overlap Points</td>
</tr>
<tr>
<td>13-31</td>
<td>Reserved for ASPRS Definition</td>
</tr>
</tbody>
</table>

3. PAST RESEARCH

There is a history of past research efforts to extract buildings from LiDAR data. Because the use of this type of data is relatively new, there was a lack of automated processing within software programs. Instead, earlier researchers developed ways to extract certain shapes by identifying patterns in the point cloud. You et al. described a process in which a 3D mesh model is created from the LiDAR points (4). The various general shapes of buildings are then classified into plane, sphere, slope, etc., dictating what parametric model will be used to finalize a shape for that building segment. After the shapes are completed, a fitting routine is used to create the model of the building. Rottensteiner & Briese describe subtracting a DTM (digital terrain model) from a DSM (digital surface model) using threshold values to determine building locations (5). Regions of points within these locations are grouped together into planar patches. The patches are then formed together to create building models. Haithcoat, et al. applied height and size thresholds to determine building locations and remove smaller planar objects such as cars or benches (6). In addition, differential geometry was used to determine shape characteristics of features, allowing trees to be separated from buildings.

4. SOFTWARE INTERACTION

These and many other research efforts created a basis for the software that is available today. LiDAR is a relatively new technology, so in turn so is the ability to manipulate and process it. Much of the software on the market is tailored to specific uses, but all of it can handle the millions of points found in a LiDAR file amazingly well. In addition, some products function within or as extensions of ESRI’s ArcGIS. LP360 by QCoherent (7) is an example of this, and is a very straight-forward software for LiDAR visualization.

![Fig. 2 – LiDAR file displayed by elevation with zoomed in detail showing points as viewed in LP360](image)

Another helpful tool to visualization and classification is adding orthoimagery to the LiDAR data. From a visual standpoint, the imagery can provide a great background clarifying areas that may be difficult to interpret from the raw LiDAR points. In addition some software packages contain the ability to use imagery as a classification tool by identifying patterns in the color returned by surfaces. At a minimum, manual classification can be performed more easily when the user is able to see the landscape in natural color as the human eye would normally perceive it as opposed to deciphering the point cloud in raw form.

5. DATA PROCESSING

The overall goal of this project was to improve the abilities of photovoltaic analysis in urban areas, which is a multi-step approach. The first step was testing existing extraction tools. This included ease of use, functionality, and the output data available from the extraction. The standardization of *.las files has made it easier for the user to pass files between software packages. Each *.las file has a header section with the metadata needed for
georeferencing and units of measure for proper import and processing. Figure 3 is only a portion of a header file but shows data extents, datum, and projection information.

LAS version 1.2
MinX: 2680000.000 MinY: 210000.010
MaxX: 2689999.990 MaxY: 220000.000
Height: -77.190 to 5163.200
Database is monochromatic
Units: Feet in US Survey Standard
GTypeGeoKey: 1
GeographicTypeGeoKey: 4269
GeogCitationGeoKey: GCS_North_American_1983
GeogGeodeicDatumGeoKey: 6269
GeogPrimeMeridianGeoKey: 8901
GeogAngularUnitsGeoKey: 9102
GeogAngularUnitSizeGeoKey: 0.017453
GeogEllipsoidGeoKey: 7019
GeogSemiMajorAxisGeoKey: 6378137.000000
GeogInvFlatteningGeoKey: 298.257222

Fig. 3 – *.las header file

The first task was to extract the buildings and high vegetation from the overall point cloud. The previously described past research efforts led to this process being partially automated depending on the quality of the input dataset. First, building footprints are defined by analysis of the breaks in the elevations of the point. Generally, these elevations are based on a bare earth digital elevation model (DEM) that is created by the software. This model consists of ground and mass points either pre-classified or as determined by the software.

In the case of the test area, there are complex roof shapes that lead to manual editing being necessary. This type of editing is common in remote sensing with LiDAR and aerial photography and is improved by the processor’s knowledge of the site or use of other imagery sources. As an example, see Figure 4 & 5 below illustrating how comparative information is extremely valuable when processing and cleaning LiDAR data. Figure 4 shows a profile view of LiDAR points closely depicting ground, vegetation, and a building. But there is also a row of points under the roof line that seems to be out of place. Figure 5 shows the same area from the top, displaying the LiDAR points with aerial imagery behind them.

The imagery clearly shows two windows in the roof of the building. The points in the profile shown below the roofline are actually pulses that traveled through the window and were returned to the sensor. This type of situation needs to be remedied during the data processing to accurately depict the buildings. In the case of solar energy studies, it might be necessary to remove this area from the equations altogether as it wouldn’t be suitable for placement of photovoltaic panels.

Fig. 4 – Profile view as viewed in LP360

Fig. 5 – Aerial imagery and LiDAR points of building

Depending on the software, various parameters control how building edges and planes of roofs are determined. Generally, they have a setting of how far between adjacent points is acceptable to still consider them in the same plane. This setting must be leveraged based on the LiDAR dataset’s point spacing, in order to include or exclude necessary areas. There are also similar recognitions of point patterns to identify or filter trees based on canopy shapes. The user can generally define height and size
restrictions based on the study area to narrow down and improve this process.

With this basic level of extraction complete, a 3D model of both buildings and vegetation is completed. Solids are created for individual buildings and tree canopies based on user inputs. A sample of a completely processed dataset can be seen below as processed in E3De (8) by ITT Visual Information Systems.

Fig. 6 – Processed LiDAR scene in E3De

The solar analysis can begin once the 3D models are created by utilizing existing tools with GIS software, such as the Spatial Analyst extension with ESRI’s ArcGIS (9) suite.

6. FUTURE EFFORTS

As LiDAR data improves and processing technology becomes more advanced, a continued development of a workflow and tailored parameters will be important. This will also be true as demands of solar analysis change. To facilitate this process, complete documentation of a final workflow should be produced. Documentation should include recordation of extraction parameters and what effect they have on output data. This workflow will recommend what the user should be inputting on the front end, and then what manipulation, manual or automated, is completed to generate the necessary format of data that will be pushed to the PV analysis. But it should be kept in mind that the testing done during this project highlighted the fact that ideal extraction parameters and techniques are different based on data quality and even individual buildings. Therefore, a stringent routine most likely cannot, and probably should not, be mandated.

7. CONCLUSION

In conclusion, it is apparent that airborne LiDAR data can be used to improve the accuracy and efficiency in which urban areas are analyzed for solar potential. The combination of true georeferenced positions of buildings along with accurate modeling of characteristics such as slope, orientation, and area provide a powerful combination of data to employ in solar analysis. While this method may be limited by the availability of LiDAR information, the amount of datasets is growing as the technology improves and becomes more popular. As a whole, it appears that using LiDAR data to facilitate urban solar analysis is a beneficial collaboration.

8. REFERENCES