Brief Introduction

Building footprints represent the perimeter outline of each building, with a description of the size, shape, and location of its foundation. Building footprints data have applications in various kinds of research, especially in risk analysis. For instance, building footprints assist in testing the building location and footprint against flood extents and other hazards, allowing people to accurately locate, analyze, and visualize risk exposure.

A dataset has been prepared that contains building footprints derived from high-resolution digital elevation models from LiDAR. The building footprints were created by extracting points with Classification 6 buildings of the Illinois LiDAR LAS-format files using the LP360 software. LP360 includes Point Cloud Tasks (PCTs) for performing building classification and extraction from a point cloud. The spatial coordinate system is NAD1983 (EPSG: 4269). Data are available for the 36 counties in Illinois listed in Table 1 and shown in Figure 1, whose LiDAR collection dates range between 2012 and 2017. Counties with data before 2012 are not included in this dataset because they did not include Classification 6 buildings, and counties with LiDAR data after 2017 were not available at the time of processing.

Table 1. Available Counties Data with Year and File Size

County	Year	File Size (bytes)	County	Year	File Size (bytes)
CASS	2017	3,133,188	**HARDIN**	2014	1,466,228
GREENE	2017	4,279,964	**JACKSON**	2014	9,647,748
HANCOCK	2017	6,709,452	**KANKAKEE**	2014	16,228,252
MACOUPIN	2017	11,907,476	**MADISON**	2014	43,203,956
MONTGOMERY	2017	7,889,312	**PERRY**	2014	5,546,236
BOND	2015	4,314,676	**POPE**	2014	2,212,500
BUREAU	2015	8,831,960	**WILL**	2014	76,802,512
CLINTON	2015	9,091,412	**GALLATIN**	2012	1,384,140
FORD	2015	3,544,468	**HAMILTON**	2012	2,403,944
IROQUOIS	2015	7,895,528	**JOHNSON**	2012	1,499,972
JEFFERSON	2015	8,744,560	**MASSAC**	2012	2,932,188
LIVINGSTON	2015	8,769,624	**MONROE**	2012	6,545,172
MARION	2015	8,732,956	**PULASKI**	2012	1,422,192
PIKE	2015	4,443,892	**RANDOLPH**	2012	10,135,516
SCOTT	2015	1,419,564	**SALINE**	2012	5,261,600
WASHINGTON	2015	6,336,104	**ST. CLAIR**	2012	44,641,384
CHRISTIAN	2014	8,581,984	**WHITE**	2012	3,270,096
FRANKLIN	2014	97,820,880	**WILLIAMSON**	2012	10,918,996

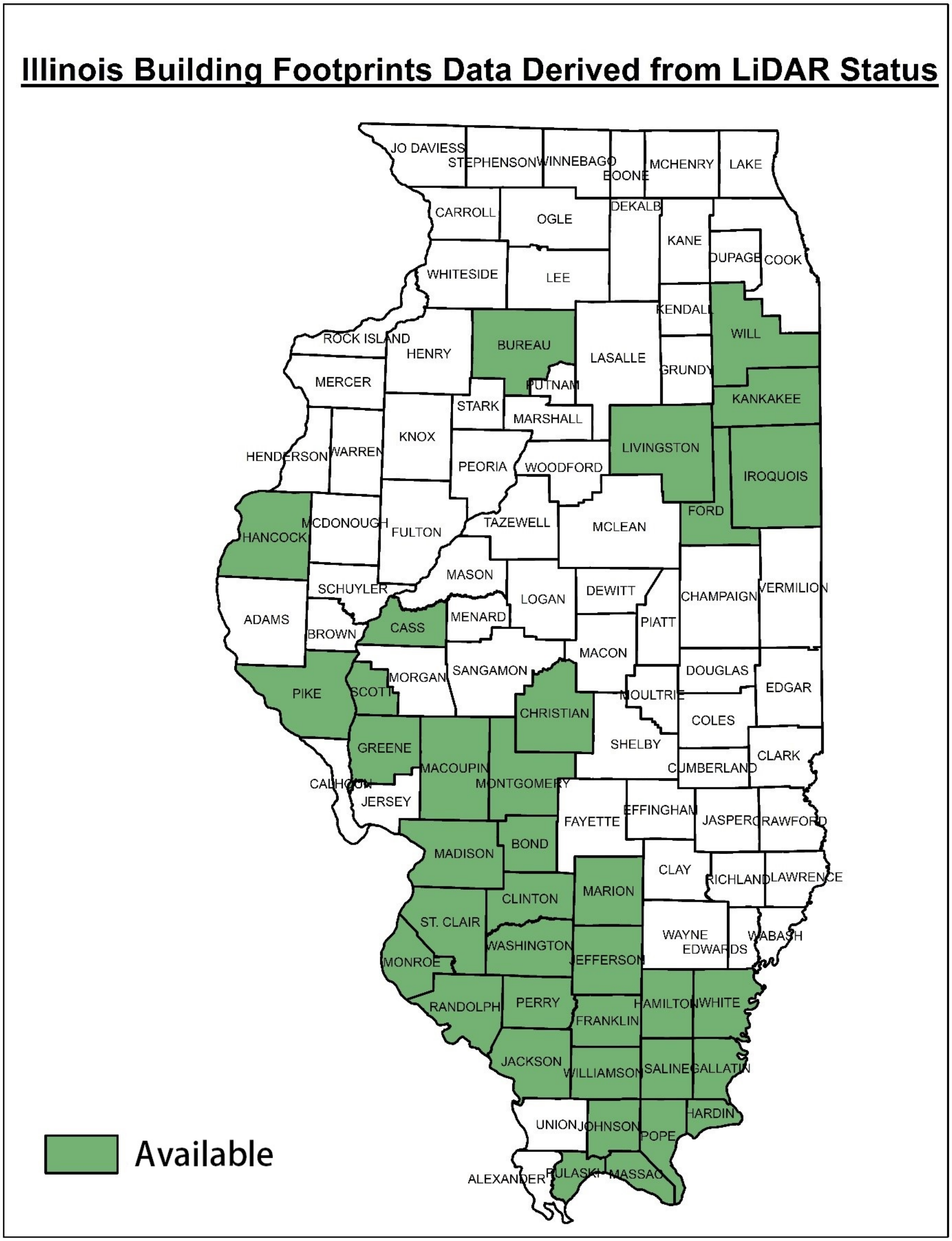

Figure 1. Thirty-six counties of Illinois have building footprints extracted from LiDAR

Data Details and Processing

Data Details

Data were derived from LiDAR data in LAS format and were converted into vector data in *shapefile* format. This dataset contains the default attributes created by the LP360 software, as shown in Table 2.

Table 2. Attribute Table of the Shapefile

LP 360: Attribute Table		
Field Name	**Data Type**	**Definition**
FID	Object ID	Object ID generated by ArcGIS
Shape	Geometry	Feature geometry
ID	Long Integer	Object ID generated by LP360
Area	Double	The area of the polygon
RmsErr	Double	Root Mean Square Error
MaxErr	Double	Maximum Error
ForceFit	Long Integer	Undefined

Table 3. Spatial Reference of the Shapefile

LP 360: Spatial Reference			
Projected Coordinate System	*NAD_1983_2011_StatePlane_Illinois_West_FIPS_1202_Ft_US*	Bond, Bureau, Cass, Christian, Clinton, Greene, Hancock, Jackson, Macoupin, Madison, Monroe, Montgomery, Perry, Pike, Randolph, Scott, St. Clair, Washington	
	NAD_1983_2011_StatePlane_Illinois_East_FIPS_1201_Ft_US	Ford, Franklin, Hardin, Iroquois, Jefferson, Johnson, Kankakee, Livingston, Marion, Massac, Pope, Pulaski, Will	
	NAD_1983_UTM_Zone_16N	Gallatin, Hamilton, Saline, White, Williamson	
Projection	Transverse_Mercator	**Linear Unit**	Foot_US
Geographic Coordinate System	GCS_NAD_1983_2011		
Datum	D_NAD_1983_2011	**Angular Unit**	Degree
Coordinates have Z values		Yes	
Coordinates have measures		Yes	

Data Quality

1. The building footprints data were derived from LiDAR data directly without modifying specific building footprint boundaries. Only sample locations were inspected regarding proper identification, size, location, and shape. This inspection was conducted with the intent to adjust the parameters used in the LP360 software for a better overall product, but not with the intent to fix deficiencies in specific building footprints. A thorough inspection of all building footprints was not done.
2. Building footprints represent the perimeter outline of each building, but the building outlines can only roughly represent the buildings. Sometimes one single polygon may include many buildings that are close to each other.
3. The squaring function was performed to produce an approximation of the roof outlines of buildings by squaring the traced building outlines. Thus the extraction detected rectangular buildings very effectively; the buildings with other shapes may not be extracted as their real shapes. There was no classification within the LiDAR point cloud that differentiated between the types of structure, so round features, such as storage tanks, could not be discerned as different from square buildings without a visual inspection. As a result, the squaring function was mistakenly applied to features of all shapes (Figure 2). The data user, such as a community, could manually digitize building footprints to replace these features.

Figure 2. A sample dataset showing errors from squaring buildings from Will County, Illinois

4. The LP360 software point group tracing and squaring task extracts the building footprints from the LiDAR point cloud. This task includes these parameters: Grow Window, Trace Window, Minimum Area, and Minimum Points. Each parameter affects the identification, size, location, and shape of the extracted building footprint. The Minimum Area and Minimum Points were set as defaults. Trace Window and Grow Window are the main parameters that determine the outcome of extraction in LP360. Trace Window controls the "jaggedness" of the polygon outline. The smaller the value, the more detailed the edge. Grow Window controls clustering. The larger value makes courser clusters. Each county was assigned a different parameter value to receive the best outcome. The value of the parameter and data details are listed in Table 4.

Table 4. Data Details with Value of Parameters for Extraction

County	Number of Buildings	Area (sq ft)	Grow Window /Trace Window
BOND	18827	35426974.9	7,10
BUREAU	36170	75426124.3	6,8
CASS	12782	25964947	6,8
CHRISTIAN	34908	67849064.6	6,8
CLINTON	40584	87562263.6	6,12
FORD	14054	30475120.9	6,8
FRANKLIN	34512	63858481.7	7,12
GALLATIN	6068	1073883.77	6,12
GREENE	17668	30075469	6,8
HAMILTON	9302	1588593.14	6,8
HANCOCK	27506	55066263.7	6,8
HARDIN	6190	9662789.63	5,8
IROQUOIS	31225	65476191.8	6,8
JACKSON	39320	79680443.7	7,10
JEFFERSON	34077	69329907.5	6,8
JOHNSON	6796	10202605.7	6,12
KANKAKEE	64795	161302493	6,8
LIVINGSTON	34500	82722746.7	6,8
MACOUPIN	47718	86227302.4	6,8
MADISON	158195	418645449	6,8
MARION	37178	72100909.6	7,10
MASSAC	12187	17841116.6	7,10
MONROE	25893	46259931	6,10
MONTGOMERY	32072	58694022	6,8
PERRY	23330	41122892.9	7,10

PIKE	18475	40865389.3	6,8
POPE	9255	15512723.9	6,8
PULASKI	5970	8704399.99	7,10
RANDOLPH	35645	51538395.1	6,8
SALINE	19553	3402983.17	6,8
SCOTT	5792	13154603.7	6,8
ST. CLAIR	142105	260885360	6,8
WASHINGTON	24753	52289139.6	6,8
WHITE	13321	2405697.67	7,10
WILL	295703	927956711	6,8
WILLIAMSON	42830	9027262.19	7,10

Comparison with Microsoft Building Footprints Data

In 2018, after ISWS' work to extract building footprints was underway, Microsoft released approximately 125 million building footprint polygon geometries from all 50 U.S. States in GeoJSON format. The building footprints were generated by training computer vision algorithms to recognize building geometries on aerial imagery of the U.S. The coordinate reference system is WGS84 (EPSG: 4326).

Information about the Microsoft data is provided in Tables 5 and 6. More information can be found at https://github.com/Microsoft/USBuildingFootprints.

Table 5. Attribute Table of the Data

Microsoft: Attribute Table		
Field Name	**Data Type**	**Definition**
OBJECTID	Object ID	Object ID generated by ArcGIS
Shape	Geometry	Feature geometry
Shape_Length	Double	Object ID generated by LP360
Shape_Area	Double	The area of the polygon

Table 6. Spatial Reference of the Data

Microsoft: Spatial Reference	
Geographic Coordinate System	GCS_WGS_1984
Datum	D_WGS_1984
Angular Unit	Degree
Coordinates have Z values	No
Coordinates have measures	No

Extraction Method

Microsoft developed a method that approximates the prediction pixels into polygons, making decisions based on the whole prediction feature space. This process is different from the LP360 extraction method. Results obtained from LP360 had to be smoothed to decrease the occurrence of noisy points near the building edges. This was performed using the squaring function, which produces smoother and neater polygons of buildings by squaring the traced outlines (Figure 3).

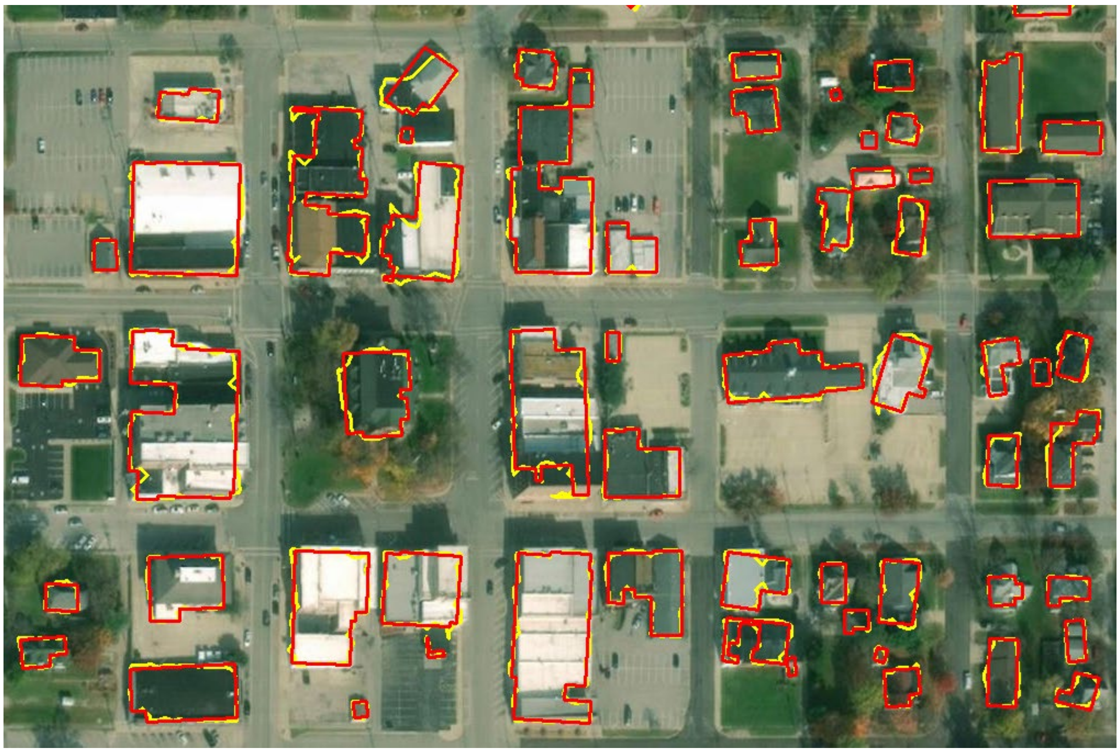

Figure 3. A sample area from Bond County, Illinois showing buildings before and after applying the squaring function in LP360

The example illustrated in Figures 4 and 5 shows the difference between these two extraction methods. The building footprints extracted from LP360 may not detect buildings with shapes other than rectangular very effectively, but result in clean outlines. The Microsoft building footprints can approximately trace the building outlines, but lead to an excessive curvature in polygons and ignore the occurrence of noisy points near the building edges.

Figure 4. LP360 extraction: A sample dataset from Will County, Illinois

Figure 5. Microsoft building footprints: A sample dataset from Will County, Illinois

Comparison of Quantity

Comparing these two datasets, there were more building footprints per county extracted by ISWS using LP360 than the number of building footprints that Microsoft released. Johnson County in Illinois is the only one of the 36 counties processed where the Microsoft release number is greater; however, the total area of ISWS-extracted building footprints was not greater for all counties. The total area was greater in only about 60% (23 of the 36) of counties compared. This comparison does not readily point to reasons for the difference in quantity.

Initially, the reason for the increase was thought to be because the LiDAR data used for the extraction are in LAS tile format. The default tiling scheme in this format is a series of square tiles that include areas of adjacent counties beyond the county boundary; therefore, building footprints of adjacent counties were extracted, which inflated the count. Subsequently, building footprints outside of the county boundary were removed from each county using GIS processes; the results in Table 7 show that decrease, but all counties, except Johnson, still have a greater number of ISWS-extracted building footprints than Microsoft-released building footprints.

Table 7. Comparison between LP360 Building Footprints and Microsoft Building Footprints

Building Footprints Dataset Comparison							
	Number of Buildings				Area of Buildings (sqft)		
County	LP360 with county overlap	LP360	Microsoft	Ratio (Microsoft /LP360)	LP360 (no county overlap)	Microsoft	Ratio (Microsoft /LP360)
BOND	18,827	16,625	12,752	77%	31,800,820	30,027,191	94%
BUREAU	36,170	35,009	28,908	83%	71,802,329	67,547,607	94%
CASS	12,782	12,480	9,817	79%	25,198,289	22,413,468	89%
CHRISTIAN	34,908	32,085	23,737	74%	63,292,995	56,041,368	89%
CLINTON	40,584	32,243	24,740	77%	72,013,318	66,798,643	93%
FORD	14,054	14,043	11,211	80%	30,740,095	27,514,941	90%
FRANKLIN	34,512	32,897	25,362	77%	60,359,288	55,329,073	92%
GALLATIN	6,068	5,652	5,194	92%	10,954,705	11,885,311	108%
GREENE	17,668	16,510	12,680	77%	28,364,341	27,087,360	95%
HAMILTON	9,302	8,687	8,363	96%	16,025,343	18,315,073	114%
HANCOCK	27,506	25,133	18,226	73%	49,010,990	41,674,229	85%
HARDIN	6,190	5,423	4,308	79%	8,560,075	9,553,666	112%
IROQUOIS	31,225	30,351	25,448	84%	63,805,643	62,056,607	97%
JACKSON	39,320	36,369	28,659	79%	74,308,304	73,462,687	99%
JEFFERSON	34,077	33,246	22,280	67%	67,716,187	56,787,933	84%
JOHNSON	6,796	6,558	9,005	137%	9,826,685	19,135,818	195%
KANKAKEE	64,795	62,486	53,903	86%	156,677,942	145,024,553	93%
LIVINGSTON	34,500	32,379	26,103	81%	78,594,096	69,677,071	89%
MACOUPIN	47,718	46,286	37,062	80%	83,810,701	82,749,213	99%
MADISON	158,195	138,099	130,983	95%	360,373,318	335,819,210	93%
MARION	37,178	35,104	24,406	70%	68,397,680	58,935,444	86%
MASSAC	12,187	12,038	10,524	87%	17,624,955	24,158,565	137%
MONROE	25,893	24,095	19,774	82%	43,087,105	49,653,321	115%
MONTGOMER	32,072	30,697	22,985	75%	56,303,750	53,245,773	95%
PERRY	23,330	19,561	15,619	80%	34,825,540	35,289,338	101%
PIKE	18,475	17,642	15,795	90%	38,031,988	35,093,799	92%
POPE	9,255	6,286	4,799	76%	10,088,868	9,423,064	93%
PULASKI	5,970	5,652	4,380	77%	8,285,754	9,918,113	120%
RANDOLPH	35,645	33,388	22,379	67%	47,374,253	52,715,958	111%
SALINE	19,553	19,133	17,745	93%	35,924,321	40,033,742	111%
SCOTT	5,792	5,510	5,085	92%	12,372,181	11,186,132	90%
ST. CLAIR	142,105	139,733	119,887	86%	254,339,674	300,523,748	118%
WASHINGTON	24,753	19,447	14,941	77%	41,705,853	39,868,482	96%
WHITE	13,321	12,577	12,065	96%	24,787,010	27,383,961	110%
WILL	295,703	246,275	236,927	96%	782,212,765	740,268,108	95%
WILLIAMSON	42,830	42,641	39,720	93%	97,823,917	103,717,496	106%

Pope County provides an example showing the impact of GIS processing to remove building footprints of adjacent counties. In Figure 6, the picture on the left shows the original data tiles. It is obvious that the original data overlap the boundary of Pope County. The picture on the right shows the Microsoft-released data, where building footprints are within the county boundary.

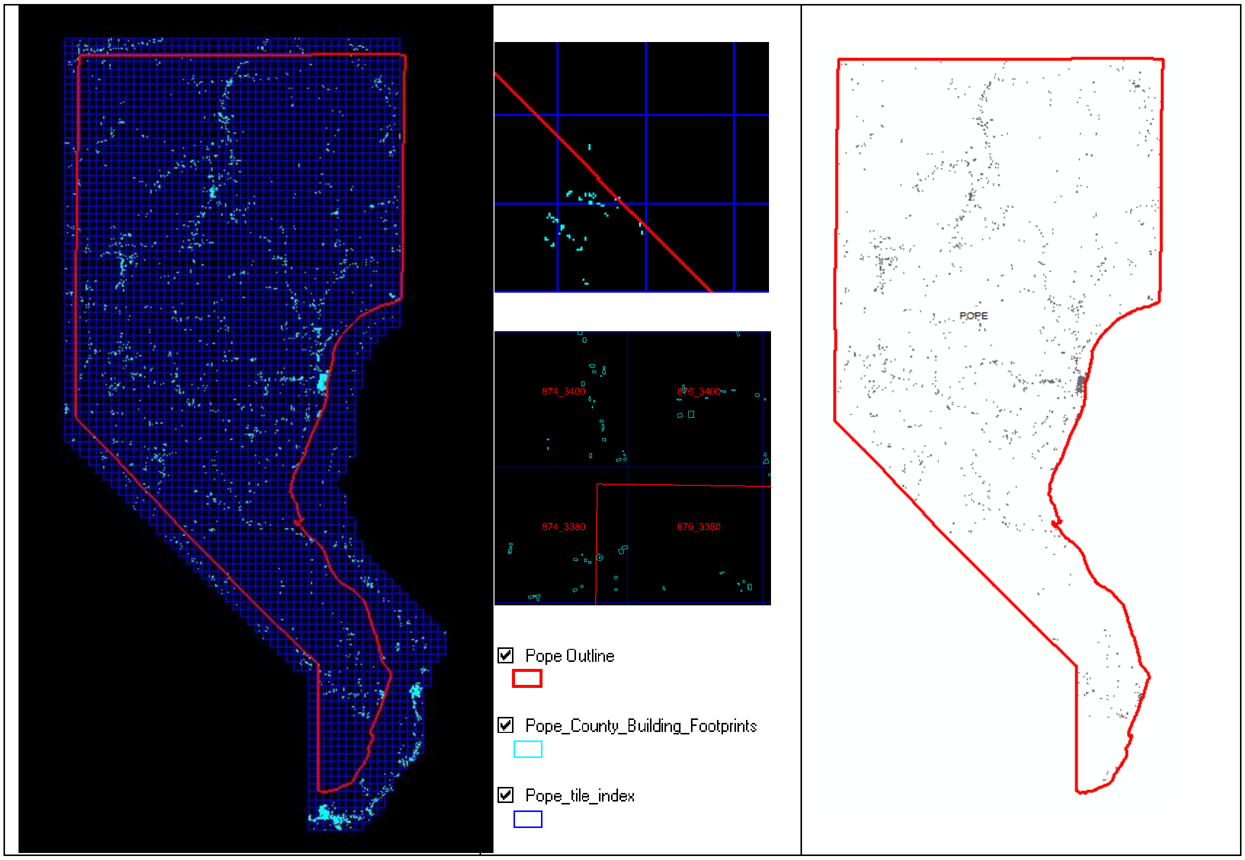

Figure 6. Comparision of building footprints in Pope County, Illinois

Left: Some parcels outside the county boundary are included in the Building Footprints Data extracted by LP360.
Right: Building Footprints Data released by Microsoft are within the boundary.

Overlap with Land-use Data

Another possible reason for the difference in quantity of building footprints between the two methods is that LiDAR has the capacity to penetrate the forest canopy and sense structures that may not be visible in aerial images. Thus the Microsoft image-derived building footprints would exclude structures in forested areas. To test this idea, the tallies of building footprints within forested areas were compared for the two methods. A simple comparison was made between these two building footprints datasets using the 36 counties data in Illinois. The land cover data are from the National Land Cover Database (NLCD) Land Cover Collection, 2011. The outcome of the comparison shows an increase in building footprints extracted from LiDAR in forested areas. However, the increase of 0.7% using the LiDAR is far less than the observed

9.2% increase of building footprints occurring in agricultural areas (planted or cultivated). Agricultural areas don't have canopy penetration considerations, so another possible reason for the difference may be related to urbanization. Unfortunately, the Microsoft-released data do not include metadata noting the dates of the aerial imagery used; however, if the imagery data are older than the LiDAR data, the difference in building counts is possibly due to urbanization. The general practice of new development in Illinois is to expand communities into neighboring agricultural lands. Between 2007 and 2015, rural land in Illinois decreased by 95,000 acres and developed land increased by 106,700 acres.

Table 8. Percentage of Buildings Overlapping with Land-use Data

Overlapping building Percentage*	Forest	Water	Planted or Cultivated	Wetland	Grassland or Herbaceous	Shrubland	Barren
LP360	3.7612%	0.3723%	24.9162%	0.1413%	1.3146%	0.0582%	0.5178%
Microsoft	3.0584%	0.4210%	15.6815%	0.1579%	0.9633%	0.0149%	0.3263%

*percentage was calculated with 7 kinds of land use and 36 counties

More details about the definition and classification of the land cover can be found at https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend.

Without a more thorough investigation, the reason for the difference in quantity between the Microsoft-released and the ISWS LiDAR-extracted building footprints is not fully understood, and may result from a mixture of reasons, including the different extraction methods, differences in data age, or something else.

Bibliography

https://github.com/Microsoft/USBuildingFootprints

https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend
GeoCue Group Support, LP 360
http://kb.geocue.com/wp-content/uploads/2014/11/Automatic-Building-Extraction.pdf

Illinois Height Modernization Program (ILHMP): LiDAR Data
https://www.nrcs.usda.gov/Internet/NRCS_RCA/reports/nri_il.html

Prerna, R., and C.K. Singh. 2016. Evaluation of LiDAR and image segmentation based classification techniques for automatic building footprint extraction for a segment of Atlantic County, New Jersey. Geocarto International 31(6):694-713, DOI: 10.1080/10106049.2015.1076060