Abstract: A new method for wide-area urban roof assessment of suitability for solar photovoltaics is introduced and validated. Knowledge of roof geometry and physical features is essential for evaluation of the impact of multiple rooftop solar photovoltaic (PV) system installations on local electricity networks. This paper begins by reviewing and testing a range of existing techniques for identifying roof characteristics. It was found that no current method is capable of delivering accurate results with publicly available input data. Hence a different approach is developed, based on slope and aspect using LiDAR data, building footprint data, GIS tools and aerial photographs. It assesses each roof’s suitability for PV installation. That is, its properties should allow the installation of at least a minimum size photovoltaic system. In this way the minimum potential solar yield for region or city may be obtained. The accuracy of the new method is then established, by ground-truthing against a database of 886 household systems. This is the largest validation of a rooftop assessment method to date. The method is flexible with few prior assumptions. It is based on separate consideration of buildings and can therefore generate data for various PV scenarios and future analyses.

Keywords: solar; LiDAR; rooftop photovoltaics; building characteristics; wide-area solar yield.

1. Introduction

1.1 Significance of 3D rooftop attributes for photovoltaic system installation and yield

Precise estimation of the solar energy resource on pitched roofs is crucial for modelling photovoltaic (PV) installation in residential scenarios. However, there is no national database of building characteristics in the UK. This is also a common omission in other countries. The EU Buildings Database [1] contains information on gross floor area and roof insulation type but has no details of roof inclination, orientation or pitched dimensions. The US is slightly better provided for with a report which assesses the rooftop solar PV potential of 23% of buildings nationwide [2]. In this paper, previous work described in [3–5] is re-visited and expanded. That is, automated extraction of building roof plane characteristics over wide areas, shading techniques and the influence of module orientation on yield are studied. A grid square (pixel)-based approach to estimation of solar energy potential over pitched roofs is developed. This is achieved by combining publicly available building outline maps with aircraft-based LiDAR (Light Detection and Ranging) data. These are analysed statistically within a Geographical Information Systems (GIS) environment.
The new approach is validated against data from a selection of approximately 2000 citywide PV systems currently installed in Nottingham, UK. Up to the present, no other rooftop PV capacity studies have been as rigorously validated.

1.2 PV in the built environment

There is strong growth of PV in the built environment. To date in the UK, rooftop PV has been deployed faster than expected. 20% of total installed solar PV capacity (2.6 GW) comes from small scale 0 to 4 kW installations [6]. Net-metering to encourage self-consumption is taking place in the UAE, Lebanon, Chile and parts of India [7]. In the Netherlands household installations comprise 90% of PV capacity [8]. In China, rooftop distributed installations reached 15 GW in 2017, a rise of 400-500% compared to 2016 [9].

Modern living space offers a range of challenges and opportunities for solar panel installation. Many neighbourhoods display complex amalgams of roof pitch (tilt) and building orientation (azimuth). Roof features such as chimneys, vent pipes, aerials, roof lights, cross-gables, and dormers may further reduce possible system size.

Roof pitch is linked to building age and roofing material. These in turn are related to geographic regional variation in construction methods. Building orientation (azimuth) is largely dictated by road layout which is a reflection of local topography.

1.3 The influence of tilt and azimuth on rooftop solar irradiation

The tilt and azimuth of a PV system have two main influences on energy yield. First, there is an increase or decrease in annual total yield depending on how well the roof pitch and azimuth match the average sun position over the year [10]. Second, the daily or seasonal timing of peak energy generation is influenced [11].

Solar panels may capture the maximum solar radiation by inclination at an optimal angle dependent on sun path (site latitude) and typical weather (diffuse fraction of solar radiation), which is about 38 degrees for most of the UK. Roofs which have higher or lower pitches than this optimal value receive less irradiation.

Existing housing stock does not always allow the use of this optimum and compromises in deployment are necessary. Similar mismatches occur in other countries. Whereas the traditional UK roof pitch (40-50°) [12] tends to be steeper than the optimum angle for PV, in Mediterranean regions traditional tile roofs (20-25°) [13] are frequently shallower than the optimum 30-33°. Azimuth may also be less than ideal. Although total annual yield is lowered by non-optimal building orientation there may be positive side effects such as higher morning, evening and winter generation.

Analysis of roof characteristics and their impact on PV output and timing for an individual house is relatively straightforward. This research provides an efficient method for analysing areas too large to investigate manually.

1.4 Research Methodology in Brief

There has been substantial previous research into computerised recognition of three-dimensional structural features. [14–16] provide excellent literature reviews. The authors of [14] divide rooftop area estimation methods into three. First, the constant value methods approach the problem collectively by scale-up e.g. [17]. Although quick and easy, they employ broad assumptions and produce generalized results. Second, manual selection e.g. [18] which is time-consuming. Third, GIS-based methods e.g. [3] which deliver detail and may be automated but require substantial computing power. No technique is widely accepted as definitive. This paper concentrates on GIS-based methods.

This area of research is challenging in terms of both data quality and the sheer size of LiDAR datasets. Additionally, 3D feature extraction is non-trivial. First, existing methods using both LiDAR and aerial photography as inputs are tested. The advantages and disadvantages of these techniques are reported, and results presented. These methods include model driven, peak detection, iterative
voting, LiDAR edge detection, image edge detection, image recognition and hill shading with ambient occlusion.

LiDAR datasets are a grid of height values from an aircraft flying at constant altitude pulsing a laser to Earth and timing the returns. The number of returns per square metre determines the resolution of the data. LiDAR supplies detailed heights of objects (e.g. buildings and vegetation), as well as terrain surface. In the UK, LiDAR data is supplied by the Environment Agency [19]. Several resolutions are available for limited areas. 1 m resolution was chosen as the best compromise between accuracy and availability. It covers approximately 70% of England. Aerial photography was obtained from GoogleEarth [20].

Since none of the existing methods was found to be adequate using the available input data, a new approach is elaborated. This is expedient for medium resolution LiDAR. Previous methods may result in imprecise values unless high resolution data is available. Rather than attempting to obtain exact roof areas, each roof is assessed to discover whether it is suitable for PV installation. Suitability here is defined as a roof to which at least a minimum size photovoltaic system may be fitted (8 m$^2$ of roof area corresponding to a 1 kW system, assuming a panel size of 1.6 m$^2$ [14]), with an azimuth East through South to West and tilt of between 15$^\circ$ and 60$^\circ$. (Most UK homeowners install a 1 kW to 4 kW solar panel system on their roof [21].) Thus, the minimum potential solar yield for region or city may be obtained. The need to generate accurate roof areas and PV system sizes from inexact data is avoided because the goal is the minimum requirements for domestic PV only. Nor is this method intended to separately identify multiples of 1 kW. This is not possible with the available input data and domestic systems are almost entirely comprised of smaller multiples e.g. 3.76 kW.

LiDAR heights on roof tops only are selected by clipping them out using building outlines (from OS MasterMap Topography Layer [22]) as patterns or “cookie cutters”. The tilt and azimuth of each roof pixel is calculated by weighted least squares fit of a plane to a 3 × 3 pixel neighbourhood, centred on each LiDAR point (see [3] for details). Due to LiDAR inaccuracies, flight paths and chance, this will result in a unique value for each pixel. Roofs slope evenly, so a statistical technique smoothing azimuth pixel values into groups is identified. Next the roof is divided into separate planes according to the grouping. The size, tilt and azimuth of each plane is known from the calculation just carried out. A check is performed to ascertain whether a system of 8 m$^2$ can be mounted on any of the roof faces.

Two case studies are used to test the methods. The first is the Wollaton Park area of Nottingham, UK. This was selected for the variety of architectural styles displayed by its houses. The second is a set of about 2000 housing association domestic installations in Nottingham. Locations, system sizes, and installer records of the tilt and azimuth of each of these systems have been gathered from a monitoring portal.

2. Review and test of existing methods of rooftop PV estimation

2.1 Simple Roof slope and azimuth extraction

Trigonometry may be applied to the LiDAR grid of roof point heights to calculate tilt and azimuth of every grid square or pixel (2 m, 1 m, 50 cm, 25 cm or 15 cm, depending on the resolution available in the area of interest). The traditional approach is to group pixel values obtained from a 3 × 3 neighbourhood by compass point bands to produce realistic roof planes [23]. However, Figure 1 illustrates the problems which may occur. The building in the aerial photograph has a simple roof layout, comprising one north and one south-facing roof section. (Note: the overhead perspective images in this paper are of poor quality. This is a reflection of the data which is publicly available and is part of the problem which this paper seeks to address.) Whilst the azimuth diagram generated from LiDAR reproduces the two sections, there are many spurious small roof planes pointing in various directions. These result from the presence of chimneys, TV aerials etc, as well as overhanging trees and surrounding structures such as garages. In the case of more complex roof structures, these pseudo planes can be difficult to distinguish from genuine dormers and porches.
Extraction of roof geometry from LiDAR has been the subject of extensive research over the last ten years. Existing solutions are categorised, reviewed and tested in the following paragraphs.

2.2 Model Driven Methods

This approach comprises the matching of the irregular roof segment shapes obtained from LiDAR to the best-fitting model in a library of basic building shapes. Jacques et al (2014) [24] utilise it to classify small buildings in the city of Leeds, UK, using a restricted catalogue of common roof profiles (gabled, hipped, flat, complex, or unclassified). Model-based methods do not work well for multifaceted roof shapes and intricate building construction. Looking at the topic from a country-wide perspective, there are numerous possible building types. Internationally, roof type is just as varied [25]. Some authors list as many as 50 categories with multiple sub-categories. The Geograph Britain and Ireland Project [26], which collects representative photographs for every square kilometre of the nation, has captured examples of over 25 different roof profiles. Some have very different forms (e.g. flat, round or hipped dormer). Due to the multiplicity of possible model shapes, this line of research was not pursued.

2.3 Histogram discrimination / peak detection

This approach is perhaps the oldest and simplest. Peaks are searched for in elevation (above ground level), tilt or azimuth histograms and used to segment the data. Spatial planes are fitted for each segment. Theoretically, simple gabled roofs should display a rectangular height histogram and that of hip roofs should resemble a trapezium [27]. In fact, these ideals cannot be achieved with real data, as Figure 2 explains. The Wollaton Park (Nottingham) house in the Figure 2 example is a complex but not unusual structure, comprising a hipped roof with a porch and dormers. Its elevation histogram should slope gently straight down (in the shape of a trapezium [27]) but in reality is concave (see dotted and solid lines in Figure 2 top right). The building in question is known to have a roof tilt of 38° (the same for each of the two major front and rear planes). Actually, it is barely possible to distinguish the 35-40 degree bin as the most frequently occurring in the tilt histogram. (This is obvious on a simpler roof form.) The azimuth histogram is a little clearer in as far as the major front (south-facing, 180°) and rear (north-facing, 360°) planes may be discerned. Then again, the east (90°) and west (270°) planes cannot be extricated from random noise in the data.
Figure 2. Elevation, Tilt and Azimuth histograms from 1 m LiDAR for a complex roof.
Probably because of these kind of problems, the peak-fitting method seems to have been largely replaced by other techniques. Furthermore, it analyses each height point in isolation. The spatial relationship between points is not considered, although all the points on a plane will have related values until an edge is reached. Newer methods refine peak detection with iterative voting (e.g. region-growing [28], random sample consensus algorithm (RANSAC) [29] and Hough Transform [30]. These are all region dependent i.e. they account for the spatial location of each height point with reference to its neighbour. The Hough Transform is the computationally fastest of these techniques.

Hough plane detection has two stages: edge detection, followed by grouping of the points inside the edges to generate the planes.

2.4 Edge Detection

Initially, a Canny edge detector [31] was applied to the 1 m LiDAR data for the Nottingham house in Figure 2 (GRASS software, i.edge [32]). The Canny edge detector is well known and often used to process both LiDAR data and images. It works by marking local maxima in the LiDAR as edges. However, in the case of the Nottingham building in Figure 2, the Canny algorithm completely failed to discriminate any edges (roof ridges), due to noise and the relatively coarse resolution of the data. When tested on several of the smaller housing association properties, edges were detected but not all correctly (Figure 3). On some homes the roof ridge is identified but on others an edge perpendicular to the expected position is located. With no clear or consistent pattern to these errors, further algorithms were trialled with the aim of improving reliability.

These included the simple (moving window) filter of SAGA GIS [33], and ArcGIS [34,35] low pass (3 × 3 cell area mean), majority (3 × 3 cell area mode) and high pass (3 × 3 cell area weighted) filters. There was no improvement in results. Roof ridges appeared too wide or were not detected. The problem appears to be the resolution of the input LiDAR data. Roof features are too small to be easily perceptible in 1 m data. Figure 4 illustrates the LiDAR data for the example Nottingham house in the form of a simple graph. The larger the circle, the higher the roof elevation of the 1 m grid cell it represents. As may be seen, even with manual intervention, not all roof features are visible in 1 m LiDAR.
Higher resolution LiDAR is only publicly available for small areas of the UK and not for the Nottingham test area. For this reason, tests were carried out with aerial photography instead of LiDAR.

2.5 Edge Detection using Google Earth Images

Images captured from Google Earth were utilised because they are readily available and cover all areas. Several filters available in GIMP software were investigated [36,37], including the low pass, Sobel (horizontal and vertical moving windows) and Laplace (high pass). The basis of all of them is gradient calculation, with edges being defined when a threshold value is exceeded, similar to the Canny edge detector. The best results were achieved with the Laplace filter preceded by a 10 pixel blur to prevent false edges (Figure 5).

![Figure 5. Laplace (weighted high pass) filter with 10 pixel blur applied to Google Earth image of single complex roof in Nottingham](image)

When a wider area was investigated (the Wollaton Park suburb surrounding the example house), it became obvious that only two planes of four-plane roofs were being identified (Figure 6). This image shows the south and west plane as one, and the north and east plane as one, for four-plane roofs. This is possibly due to the aerial photograph being taken in the afternoon and the filter merely distinguishing the sunny/less sunny sides of the roof. The next step was to investigate image recognition techniques.
Figure 6. Laplace (weighted high pass) filter with 10 pixel blur applied to Google Earth image of a residential area in Nottingham

2.6 Image Recognition as a method of extracting roof planes

Initially, an unsupervised technique (i.e. image classification without the analyst’s intervention) [38] was tried on the example Nottingham house. The ArcGIS software automatically groups image pixels with similar values into statistically distinct classes using iterative clustering around the mean (iso cluster algorithm). In this instance, the outcome was unusable. Almost every pixel in the image was treated as a separate roof plane, the exception being areas of shade which were well distinguished because they are much darker than the rest of the roof. Several supervised methods were then tested. That is, training areas representative of separate roof planes were created by manually digitising polygons. Next, these training samples were used to categorise all other pixels in the image via a classification algorithm. Training examples were digitised for all directions of the example Nottingham roof which may be identified manually (north, east, south and west). Areas of shade on the roof and chimneys were also digitised for recognition as separate features. Two classification methods delivered reasonable results (Figure 7):

- Class probability which employs Bayesian statistics to segment the image.
- Maximum likelihood classification which also uses Bayes theorem but weights classes if they are more likely to occur.
It may be seen that again there are problems with spurious features being identified. Having said that, the main difficulty is that the north plane is hard to distinguish from the east, and the south plane cannot be separated from the west. This is similar to findings from edge detection of Google Earth images. It appears that what is needed for accurate roof plane segmentation, is aerial photographs taken at different times of the day. These images would show different roof directions in slightly different colours, as the sun lights each one in sequence on its daily path. A composite from several images would then deliver accurate results. However, multiple daily photographs are not available from Google Earth or any other freely available image source. Therefore the decision was taken to re-examine LiDAR as a data source.

(N.B. Google Earth’s Voyager 3D Cities layer [39] is generated from multiple sources, including Sketch-up models and stereoscopic imagery. There is no 3D geometry currently accessible for download, which eliminates this resource at present.)

2.7 Hill Shading with Ambient Occlusion applied to LiDAR as a roof segmentation method

The previous sections discovered a need for images captured at successive times during the day. This was achieved by applying the hill shading with ambient occlusion module from SAGA software [4,40–42] to 1 m LiDAR data for the residential area in Nottingham. Hill shading models beam radiation from a single direction. Ambient occlusion adds the diffuse component of sunlight. It samples a hemisphere around each LiDAR height point and ascertains what proportion of that hemisphere is blocked by higher surrounding points. The pixel is shaded to suit. The combined technique was used to generate shading patterns on roofs at different times during the day. Preliminary results appear encouraging (Figure 8). North and west-facing roof planes are shaded (darker) in the morning simulation (sun in southeast), north only at 2 pm, and north and east-facing segments are in shadow slightly later in the afternoon (sun in southwest).
Figure 8. Hillshading with Ambient occlusion applied to a house in Nottingham at three time periods

Nonetheless, this technique has some shortcomings. It does not allow for beam reflection and transmission e.g. through thin cloud and therefore is not completely realistic. In addition, it is slow [41].

2.8 Review of Progress

All the techniques covered so far are based on grouping the unique values allocated to each LiDAR grid height point or Google Earth image pixel colouration to produce realistic roof planes. That is, roof segmentation traditionally precedes estimation of solar potential on building roofs. This may be the standard approach but, as illustrated above, there are many difficulties, summarised in Table 1. None of the above methods works well with the data resolution available in the UK (1 m for the most part). The following sections present an alternative methodology to conventional rooftop PV models.

<table>
<thead>
<tr>
<th>Method</th>
<th>Data Input</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Model driven</td>
<td>LiDAR 1 m</td>
<td>Too many model types in UK (&gt; 50)</td>
</tr>
<tr>
<td>2 Peak driven</td>
<td>LiDAR 1 m</td>
<td>Hard to distinguish peaks (noisy data)</td>
</tr>
<tr>
<td>3 Iterative voting e.g. region-growing, RANSAC, Hough</td>
<td>LiDAR 1 m</td>
<td>Require initial edge detection</td>
</tr>
<tr>
<td>4 Edge detection e.g. Canny, high pass filter</td>
<td>LiDAR 1 m</td>
<td>Fails due to noise and low resolution of data</td>
</tr>
<tr>
<td>5 Image detection e.g. Gaussian, Sobel, Laplace (supervised &amp; unsupervised)</td>
<td>Aerial photographs i.e. GoogleEarth</td>
<td>Only two planes of four-plane roofs distinguished. Need photos at different times of day.</td>
</tr>
<tr>
<td>6 Image supervision</td>
<td>Aerial photographs i.e. GoogleEarth</td>
<td>Only two planes of four-plane roofs distinguished. Need photos at different times of day.</td>
</tr>
<tr>
<td>7 Hill shading with ambient occlusion</td>
<td>LiDAR 1 m</td>
<td>Generates shading patterns at different times of day. Most promising of these methods but not completely realistic.</td>
</tr>
</tbody>
</table>

3. Method to discover whether roofs are suitable for minimum size PV installation

Instead of beginning by segmenting roofs into planes, this method takes the following question as its premise: “Is this roof suitable for PV?”. The suitability checklist has three elements: (1) azimuth East through South to West; (2) space for at least a minimum size photovoltaic system (8 m² of roof area for a 1 kW system); (3) pitched roof tilt of between 15⁰ and 60⁰ (flat roofs are treated separately).
This new approach comprises the subsequent steps which are displayed graphically in Figure 9. ArcGIS commands to complete the steps are listed in the Appendix.

1. Extract building height points only from the LiDAR grid, using OS Mastermap Topography layer [22] as a “cookie cutter”. That is, heights of all other objects e.g. trees, cars, bus shelters are removed (Figure 9, step 1).

2. The tilt and azimuth of each roof pixel is calculated using the Slope and Aspect functionality of ArcGIS software (Figure 9, step 2).

3. Calculate the mean of tilt for the whole roof. All major planes are assumed to have the same tilt and using the whole roof improves accuracy. Exclude all roofs with tilts not between 15° and 60°. Those roofs with a tilt of less than 15° are too shallow for accurate azimuth estimations and are re-classified as “flat”.

4. Calculate the mode of azimuth for all the pixels contained by each roof boundary.

5. At the outset, all houses are assumed to have simple two plane roofs but the area is visually checked using GoogleEarth and if more complex building forms are present, extra steps are implemented:

a. Two plane houses: taking due North as zero degrees, if the mode is greater than 270° and less than 90°, swap by 180° to obtain the south-facing plane suitable for PV. Theoretically a two-plane house should have two azimuth “modes” but by chance (and inaccuracies in LiDAR) one will prevail. Every non-flat building must have at least two opposite aspects. (It is only necessary to find one azimuth peak, not a minimum of two as in traditional peak detection.)

b. Four plane houses: if the mode is greater than or equal to 90° and less than or equal to 180°, then add 90°. If the number of west-facing pixels is greater than the number of east-facing pixels, take the west-facing ones. These deliver a higher solar yield and it is unlikely both roof planes will have PV installed due to the cost.
c. Three plane houses: as for four. However, one aspect will be missing. If no actual pixel values are within $10^9$ of the swapped mode, the swap is abandoned.

d. More than four major planes – this research does not attempt to include complex roof formats because these are considered unsuitable for PV.

6. Pick out roofs in the southern half of the compass only: East through South to West.

7. Select pixels within half a standard deviation of the mode (Figure 9, step 3).

8. Perform a Rook’s Case connectivity check to eliminate roof areas connected diagonally (by the corners) because solar panels cannot be installed in this situation (Figure 9, step 5).

9. Apply a minimum 10 pixel (10 of 1 × 1 m grid squares) filter to the selected pixels to remove small areas (Figure 9, step 6).

10. Carry out a boundary clean to remove dangling pixels etc.

11. Size of the roof patches may be computed (see Appendix). However, all patches selected now meet the minimum requirements for PV, which is the aim of this approach.

Note: for speed or in very large areas, the default of two plane roofs may be accepted. This is the most common roof type for houses of all ages (see photographs by [43]). Gabled (two plane) roofs are also found on terraced houses which comprise large areas of industrial cities.

The decision was taken to use the azimuth rather than the tilt to check for minimum PV system size. Experience proved the azimuth to be subject to less minor variations than the tilt, hence it was easier to aggregate pixels around a statistical value. An experiment on ten houses where the azimuth could be measured revealed the mode to be the most successful statistic for aggregation. (As opposed to mean, maximum etc). There is less skewing effect from errors.

In order to group roof pixels into areas which may be checked for minimum PV size requirements, the following statistical methods were tested. These all select azimuth pixels around the mode:

- Equal interval +/- 45 degrees.
- Jenks Natural Breaks [44]
- Half standard deviation of mode. This collects one third of roof data (68% std/2).
- One third standard deviation of mode. This collects about a quarter of roof data (68% std/3 = 23%).
- One quarter standard deviation of mode. This collects about one sixth of roof data (68% std/4 = 17%).

These five techniques were tried on a database of housing association homes with PV installed (see Section 4). 886 of the homes are covered by LiDAR flights, making them usable as test cases. System size of each installation is known, so solar panel area may be calculated (1 kW = 8 m²). The horizontal roof patch area selected as suitable for PV in each case was corrected to tilt area with the cosine rule (see Appendix). It was found that the half standard deviation method delivered the most accurate results. It failed to identify roofs as suitable for PV installation for only 2.5% of the housing association homes which are already fitted with systems. The other four techniques failed about twice as frequently. Manual comparison of the more complex houses in the Wollaton Park case study with aerial photography also found the half standard deviation method to be preferable.

This method is compatible with the available LiDAR resolution and is achievable using a standard desktop PC. No specialist software is required, other than GIS. The process relies on data processing. Automation is possible, but not essential. Sample results for the Wollaton Park area of Nottingham are illustrated in Figure 10. 40 roofs are identified as suitable for PV. Some complex roofs are wrongly identified in the top right of the image. These are inappropriate for solar panels because of dormers and cross gables. However, compared to the methods detailed in Section 2, this new approach performs very well. A detailed validation is described in the next section.
4. Results and Validation of half standard deviation of mode method

The new method is validated against data from a selection of approximately 2000 domestic citywide PV systems currently installed in Nottingham, UK. These are part of database of housing association homes with PV installed. It was possible to obtain address (and therefore latitude/longitude), system size and installers’ values of tilts and azimuths for these systems. 886 of the homes are covered by LiDAR, so it is possible to compare modelled results to actual on-the-ground measurements. The results are summarised in Table 2 and illustrated in Figures 11 and 12. Figure 11 graphs the percentage of actual systems within 5 degree bins of the modelled value of tilt/azimuth. Figure 12 charts the under/over-estimation of roof plane size.

4.1 Tilt

61% of the LiDAR estimated tilts were found to be within 5° of the installer’s values (Table 2). 87% of the LiDAR estimated tilts were within 10° of the installer’s values (Figure 11). The Mean Bias Error (MBE) is 4° and the Root Mean Square Error (RMSE) is 7°, most frequently occurring error 4°. Given that homogeneous houses vary by 3° [1], these are acceptable results. In addition, the installers’ figures are thought to be “rule of thumb” and not measured e.g. by inclinometer. 10° tilt variation between the traditional UK roof pitches of 40°-50° will only make a 1% difference to average annual plane-of-array irradiation received [3].

4.2 Azimuth

33% of the LiDAR estimated tilts were found to be within 5° of the installer’s values (Table 2). 66% of the LiDAR estimated tilts were within 15° of the installer’s values. 100% of the LiDAR estimated tilts were within 45° of the installer’s values (Figure 11). 45° azimuth variance impacts plane-of-array irradiation by 15%. The MBE is 14% and RMSE 18%, most frequently occurring error 5°. Again, these figures are considered to be satisfactory.

Tolerable results have been achieved despite the fact that difficulties were noted with the housing association dataset. Visual checks using GoogleEarth discovered cases where the LiDAR derived figure is correct and the installers’ value is not. Tilts and azimuths appear to have been transposed in the database in some instances. Additionally, the installers appear to have estimated azimuth by the position of the sun without allowing for its annual path.
432 4.3 Roof Patch Area

97.5% of established systems used in the validation process were correctly identified as being suitable for at least a minimum potential 1 kW system (Figure 12). 80% had an area at least the size of the actual installed system. Comparing LiDAR derived values and values calculated from the system sizes, the MBE is 6% and RMSE 9%.

Table 2. Results of validation of half standard deviation of mode method.

<table>
<thead>
<tr>
<th>Roof Characteristic</th>
<th>Total Systems n=886</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage</td>
</tr>
<tr>
<td>Tilt</td>
<td></td>
</tr>
<tr>
<td>Model within 0-5 degrees</td>
<td>61%</td>
</tr>
<tr>
<td>Model within 5-10 degrees</td>
<td>27%</td>
</tr>
<tr>
<td>Model within 10-15 degrees</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 15-20 degrees</td>
<td>2%</td>
</tr>
<tr>
<td>Azimuth</td>
<td></td>
</tr>
<tr>
<td>Model within 0-5 degrees</td>
<td>33%</td>
</tr>
<tr>
<td>Model within 5-10 degrees</td>
<td>19%</td>
</tr>
<tr>
<td>Model within 10-15 degrees</td>
<td>13%</td>
</tr>
<tr>
<td>Model within 15-20 degrees</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 20-25 degrees</td>
<td>6%</td>
</tr>
<tr>
<td>Model within 25-30 degrees</td>
<td>6%</td>
</tr>
<tr>
<td>Model within 30-35 degrees</td>
<td>5%</td>
</tr>
<tr>
<td>Model within 35-40 degrees</td>
<td>4%</td>
</tr>
<tr>
<td>Model within 40-45 degrees</td>
<td>5%</td>
</tr>
<tr>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>Modelled results show minimum area</td>
<td>98%</td>
</tr>
<tr>
<td>Model within 0-10% of actual size</td>
<td>12%</td>
</tr>
<tr>
<td>Model within 10-20% of actual size</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 20-30% of actual size</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 30-40% of actual size</td>
<td>11%</td>
</tr>
<tr>
<td>Model within 40-50% of actual size</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 50-60% of actual size</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 60-70% of actual size</td>
<td>7%</td>
</tr>
<tr>
<td>Model within 70-80% of actual size</td>
<td>9%</td>
</tr>
<tr>
<td>Model within 80-90% of actual size</td>
<td>6%</td>
</tr>
<tr>
<td>Model within 90-100% of actual size</td>
<td>6%</td>
</tr>
<tr>
<td>Model within 100-200% of actual size</td>
<td>13%</td>
</tr>
</tbody>
</table>
4.4 Review of Validation

Table 2 and Figures 11 and 12 show that the majority of the modelled tilt and azimuth values fall within 15-20 degrees (6%) of the installers' values. However, the spread of size value differences is much wider. Only 12% of results are within 10% of the ground-measured value. Therefore this research concentrates on identifying suitability for a minimum size PV system only, with 98% accuracy.

4.5 Context of half standard deviation of mode method.

Comparable techniques have been developed by [16] and NREL [14,45,46]. This section discusses the similarities and differences between previous work and the current method.
[16] bin individual pixel values from LiDAR into seven slope (tilt) and five aspect (azimuth) classes. The results are smoothed by replacing pixel values based on the majority of a 3 × 3 neighbourhood to the present pixel (Majority Filter). Prior classification means that the fine detail of the original input LiDAR is lost during the roof segmentation. This work is validated visually by matching 150 rooftops in Philadelphia to aerial imagery.

Similar to [16], NREL categorize each 1 × 1 m roof square into nine azimuth classes. For each distinct roof plane creating by azimuth classification, the mean tilt was determined (Zonal Mean). A PV installer’s data set containing the location, tilt and azimuth of 205 assembled PV arrays was used to validate the results of this analysis.

[16] report that their method gives the most precise results when applied to simple roof structures. NREL’s technique has virtually the same accuracy as the half standard deviation of mode method. 89% of NREL modelled results were within 10° of the actual slope compared to 87% obtained by the current authors. 96% of NREL’s modelled results have the same azimuth as the actual azimuth set against the 100% accuracy obtained by the current authors (allowing for categorization into compass bands to be compatible with NREL).

The techniques of [16] (prior classification with Majority Filter) and NREL (single tilt value for each unique roof plane obtained from the azimuth) were tested with UK data. It was found that the Majority Filter did not add any accuracy. Straightforward categorization into compass band as previously carried out by the current authors [3] gives greater accuracy in the UK. This is due to the low number of LiDAR pixels which fall inside the boundary of a typical UK home. Likewise, calculating the tilt value for every roof plane generated flawed results because segmenting small buildings gives inaccurate results due to lack of data points.

The methods of [16] and NREL are reported as working well for the larger homes of the US. The method presented here (entire building average for tilt and half standard deviation of mode for azimuth planes) is suitable for the smaller houses of the UK and other European and Asian countries. It is validated against more actual buildings’ data than any previous method.

5. Research Summary and Discussion

The tool developed here can be a powerful resource for investigating the deployment of rooftop PV. It can assist network operators in understanding how much energy the UK’s potential minimum number of solar panels can produce and improve the efficiency of the electricity network.

It does not focus on obtaining accurate values of tilt, azimuth and roof area but simply asks, “is this roof suitable for PV installation?”. Thus, the minimum PV capacity for any city region may be estimated and hence minimum solar yield. The maximum sized systems may not be installed on houses in any event, due to cost, aesthetics or fairness between rented properties.

The new method works on the basis of selecting pixels within half a standard deviation of the azimuth mode. The mode is the value at which the peak of the distribution curve occurs. It is a flexible approach to handling non-ideal data, where standard peak finding algorithms cannot cope with the noise. The end result is a map of roofs suitable for PV system installation; size at least 1 kW, known tilt and azimuth. These results can be aggregated by region to calculate minimum potential yield per area.

This technique has been comprehensively validated using two techniques. Firstly, by a check for wrongfully selecting inappropriate roofs as suitable for PV. This was carried out by manually matching 50 rooftops in Wollaton Park to GoogleEarth imagery (Figure 10). Two roofs were incorrectly selected as opposed to 50 correctly categorized. Secondly, by a check for missing suitable roofs by comparison against the biggest installation database used by any analogous research to date.

5. Conclusions

This method is useful, effective and functions correctly with the data publicly available in the UK (predominantly 1 m LiDAR, GoogleEarth images and MasterMap building outlines, Section 1.4). In
it provides a valuable contribution to the scientific field because the methods tested and reviewed in Section 2 require higher resolution input data than can be provided to produce usable results.

The unique attribute of the method presented here is that it is twice validated, by extensive ground-truthing against a database of 886 installations, and against aerial photography. This makes it the most thoroughly validated method to date.

The aim of any PV roof-area estimation method is to provide data for further analysis. This method is flexible. It allows individual houses as well as large numbers of properties to be examined, depending upon later requirements. Unlike some methods, it makes no assumptions when larger numbers are involved and does not rely on compass band classification. An example of use of the data generated would be a study of the relationship between azimuth and self-consumption. A range of PV-related research is enabled.

Author Contributions: Conceptualization, Diane Palmer, Ian Cole and Ralph Gottschalg; Data curation, Diane Palmer and Elena Koubli; Formal analysis, Diane Palmer; Funding acquisition, Ralph Gottschalg; Investigation, Diane Palmer; Methodology, Diane Palmer; Project administration, Diane Palmer; Resources, Diane Palmer; Software, Diane Palmer; Supervision, Ralph Gottschalg and Thomas Betts; Validation, Diane Palmer; Visualization, Diane Palmer; Writing – original draft, Diane Palmer; Writing – review & editing, Thomas Betts.

Funding: This work has been conducted as part of the research projects ‘PV2025 - Potential Costs and Benefits of Photovoltaic for UK Infrastructure and Society’ and ‘Joint UK-India Clean Energy Centre (JUICE)’ which are funded by the RCUK’s Energy Programme (contract no: EP/K02227X/1 and contract no: EP/P003605/1). The projects funders were not directly involved in the writing of this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

ArcGIS Commands for method to discover whether roofs are suitable for minimum size PV installation:

1. Preparation: Subtract Environment Agency digital terrain (DTM) LiDAR from the digital surface model (DSM) to obtain building height above ground level. Delete pixels with a lower height than 2 m. This removes rogue values whilst allowing for low eaves.

2. Cut out buildings only. Prepared LiDAR DSM-DTM grid and buildings from OS Mastermap Topography layer as inputs: ArcToolbox > Spatial Analyst > Extraction > Extract by Mask

3. Spatial Analyst > Surface > Slope / Aspect. Create integer rasters to enable mode to be computed in the next step.

4. Spatial Analyst Tools > Zonal > Zonal Statistics > Mean

5. Spatial Analyst Tools > Zonal > Zonal Statistics > Mode

6. Mode may be in the north, so carry out some swaps:
   a. Swap mode raster values by 180:
      Raster calculator:
      i. Values between 90 and 270 are alright:
         Con("aspectmode" <= 270) & ("aspectmode" >= 90), "aspectmode", 0)
      ii. Values between 0 and 90:
         Con("aspectmode" <= 90) & ("aspectmode" >=-1), ("aspectmode" +180), 0)
      iii. Values between 270 and 360:
         Con("aspectmode" <= 360) & ("aspectmode" >=270), ("aspectmode" -180), 0)
      iv. "PVMode" = "Con1" + "Con2" + "Con3"
   b. Optionally, switch by 90 east to west for 4-plane houses:
      Con(("Con4" >= 90) & ("Con4" <= 160), ("Con4" +90), "Con4")
   c. If the west mode generates a bigger polygon than the east, take that.

7. Standard Deviation Bands: Spatial Analyst Tools > Zonal > Zonal Statistics > Std
Con("intaspect" >= "PVMode" - "StdAspect" / 2) & ("intaspect" <= "PVMode" + "StdAspect" / 2),1,0)  
This makes a 1,0 raster of cells half std around the mode.

8. Connectivity: ArcToolbox > Spatial Analyst Tools > Generalisation > Region Group  
Four neighbours (for edges only, Rooks Case), Cross – exclude zero (“0”).

9. Select Large Enough Areas  
From Count because 1 × 1 m pixels. Reclassify as in Table A1 below and discard highest number which is areas not suitable for PV.

Table A1. Reclassification of pixel values to enable selection of roof area of at least 8m².

<table>
<thead>
<tr>
<th>Old Values</th>
<th>New Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>NoData</td>
</tr>
<tr>
<td>9-19</td>
<td>2</td>
</tr>
<tr>
<td>19-30</td>
<td>3</td>
</tr>
<tr>
<td>30-44</td>
<td>4</td>
</tr>
<tr>
<td>44-63</td>
<td>5</td>
</tr>
<tr>
<td>63-96</td>
<td>6</td>
</tr>
<tr>
<td>96-175</td>
<td>7</td>
</tr>
<tr>
<td>175-302</td>
<td>8</td>
</tr>
<tr>
<td>&gt; 302</td>
<td>NoData</td>
</tr>
</tbody>
</table>

10. Clean: Spatial Analyst Tools > Generalisation > Boundary Clean  
No sort, run twice.

11. Measure Roof Patch with homogeneous aspect (azimuth):  
Zonal statistics sum points in raster.  
(Add all the “1”s, not zeroes because “1”s are 1 m squares).

12. To calculate a more accurate area allowing for the roof tilt:  
Slope distance = horizontal distance/cosine(Tilt in degrees)  
E.g. Slope distance = 21.2 m / cos [32°]

References


https://www.nrel.gov/docs/fy16osti/65298.pdf


https://www.researchgate.net/publication/324728347_TRENDS_2017_IN_PHOTOVOLTAIC_APPLICATIONS


30. Tarsha-Kurdi F, Landes T, Grussenmeyer P, Koehl M. Model-Driven and Data-Driven Approaches Using
34. ESRI. ArcGIS Desktop: Release 10.3.2014.
39. Romaszewicz, M. How are the 3D models of buildings generated on Google Earth? Where did they get all
this data, and how was it compiled? n.d.
y-get-all-this-data-and-how-was-it-compiled (accessed January 11, 2018).
41. Tarini M., Cignoni P., Montani C. Ambient Occlusion and Edge Cueing for Enhancing Real Time Molecular
Visualization. IEEE Trans Vis Comput Graph 2006;12:1237–44. doi:10.1109/TVCG.2006.11
September 7, 2018).
44. Conrad O., Wichmann V. SAGA-GIS Module Library Documentation (v2.2.1) 2013:1.
45. http://www.saga-gis.org/saga_tool_doc/2.2.1/ta_lighting_0.html
2018)