Project Sunroof data explorer: a description of methodology and inputs

Last updated: May 2017

Background

Project Sunroof puts Google's expansive data in mapping and computing resources to use for people and organizations interested in solar power, helping illustrate the potential of solar power for a single house, and with the introduction of the data explorer, the potential of solar for zip codes, cities, counties and states. Project Sunroof wants to make it effortless to understand the potential of solar power.

The data explorer estimates the technical potential of solar power for a region chosen by the user. Technical potential of a resource is defined as the amount of energy that the resource can generate, irrespective of financial or societal constraints. Various different definitions of solar technical potential appear, resulting in estimates that can differ by 25% or more.

This detailed methodology description focuses on the data explorer and the approach to arrive at estimates of technical potential. The description here is generally consistent with the approach taken for single-house savings estimator, but this methodology does not cover all elements of the savings estimation at the single-house level.

Methodology

Summary

Project Sunroof estimates the amount of sun hitting a rooftop using 3D models derived from aerial imagery. The 3D models allow estimation of shading for every point on a roof, for each possible position of the sun in the sky. The 3D models also enable the estimation of the amount of available space for solar panels, including the pitch and azimuth of each roof plane. Including typical weather data enables estimates of energy production from panels placed in the viable roof space.

Project Sunroof currently covers roughly 60M buildings in portions of 50 states and Washington DC.

Modeling the roof

1. Input data

Google processes aerial imagery to create a Digital Surface Model (DSM) and co-registered imagery (i.e., each image pixel is aligned with the correct part of the DSM), at high resolution, covering a portion of the US inhabited by >50% of the US population.

2. Energy calculations

The energy received from the sun can be separated into Direct Normal Irradiance (DNI-- energy received directly from the sun), and Diffuse Horizontal Irradiance (DHI-- received from other parts of the sky.) At a particular time, every point on the surface of the earth may receive both DNI and DHI.

As described below, the DNI and DHI for ideal conditions (no obstructions, surface normal to the sun) for a particular location and time of (a typical year) may be obtained from various sources.

To calculate DNI received by a point on the surface, we calculate the shadows that are cast when the sun is at a particular location in the sky. Shading calculations account for trees, buildings and other obstructions within 100 to 150 meters, as well as large obstructions on the roof itself. Shadowed regions receive no DNI. For unshadowed regions, we determine the surface normal of each point P (by fitting a local plane), and derate DNI by the cosine of the angle between the sun and the surface normal.

We estimate DHI with a computationally simple isotropic approximation. Regardless of any other obstacle, DHI is scaled by $(1 + \cos(\theta)) / 2$ where θ is the tilt angle of the array.

Summing DNI and DHI then gives us the total energy received by a given point for a particular hour of a typical year.

We obtained DNI and DHI and temperature values from a commercially available database of weather, which is defined on a 10km grid. In places where that isn't available, we used NREL's database of weather. (We typically use the closest weather station, which can sometimes lead to discontinuities in the dataset.)

To improve efficiency, we divide the sky into a grid with a separation of 7.2 degrees along the azimuth and zenith angles, and (for a given hour) assign the sun to the closest grid point. Over the course of the year, the sun occupies approximately 200 distinct grid points. Computationally expensive operations (e.g., shadowcasting) are performed once per grid point.

Finally, these calculations are repeated for all points in an area of interest. This leads to estimates of total energy received by an arbitrary area.

Building outlines, rooftop segmentation and array layout

Identification of roof outlines begins with the rough building outlines available in Google Maps.

These were further refined using heuristics (e.g., green objects, or surfaces at ground height, are unlikely to be roofs). The result was used as a training data set for a Deep Learning neural net. The resulting model assigns a “roof score” to each pixel in a data set. This roof score is combined with further heuristics to provide roof outlines considerably improved over the initial data set.

Parts of the roof that are situated beneath tree cover are typically classified as not part of a building. While this is unintentional, such areas are usually not suitable for solar and should not lead to material inaccuracies in solar potential estimates. Additionally, the DSM enables the estimation of the locations of some (but not all) chimneys and other objects. Small obstructions that are missing from the DSM can lead to overestimation of the available space for solar panels.

We identify planar segments of roofs by applying a RANSAC algorithm to all surface normals. The azimuth and tilt of each planar segment can then be calculated.

Individual 250W panels are then placed on the roof segment, using a greedy algorithm that maximizes both total energy received and spatial contiguity of the array. We require that all arrays have at least 4 contiguous panels, and the sum of all such segments must total at least 2 KW. (OpenPV data from 1/1/2014 - 8/8/2016 indicate that systems below 2kW make up ~3% of installations and <1% of total installed capacity.)

3. Setbacks

The current version of the placement algorithm does not explicitly include setbacks from the roof edge. However, certain artifacts in the DSM (e.g., a tendency to foreshorten planes near a sharp edge) produce a similar effect, of a magnitude of perhaps 1-2 feet.

4. Conversion to AC

Estimates of total DC energy produced per year are generated for multiple arrays with different sizes and orientations. These are converted into per-array estimates of AC energy using the parameters specified in Table 1.

We also sum up the energy produced as a function of panel azimuth and tilt (both are typically divided into 10 degree bins).

5. Estimating Technical Potential

Technical potential of a resource is defined as the amount of energy that the resource can generate, irrespective of financial or societal constraints. Various different definitions of solar technical potential appear, resulting in estimates that can differ by 25% or more.

We define technical potential to include the energy generation from all solar panels (calculated as described above) that generate at least 75% as much energy as an ideally oriented and unshaded panel (approximated as the single best roof segment in the containing city or county).

For all covered cities, states, zip codes and (soon) census tracts, Data Explorer reports the total technical potential, as well as technical potential broken down by panel azimuth and tilt. The technical potential estimates in the tool are segmented into North, East, South, West and Flat. Roof segments are considered Flat for roofs with a tilt of less than 10%. Each of the cardinal directions are classified by the azimuth of the roof plane, with each cardinal direction consisting of a 90 degree wedge centered around its primary direction (e.g., East is 90 degrees +/- 45 degrees, so it ranges from 45 degrees to 135 degrees). All cardinal directions are included, with the amount of sun hitting that portion of the roof determining whether it should be included. While south-facing roof planes typically receive the most sun, there are many cases where north-facing planes receive more sun than a partially shaded south-facing plane, especially in areas with roofs with lower pitches.

6. Identifying existing installations

To identify existing solar installations, a machine learning algorithm is trained to recognize solar arrays. Such training requires a training data set that includes both positive and negative examples. However, currently available aerial imagery has a relatively sparse presence of rooftop solar arrays, since arrays are installed only on a fraction of houses nationwide.

In order to address this challenge, a training set was generated using an iterative approach. An initial small set of aerial images was obtained from geographical regions with a relatively higher percentage of solar arrays. The images were sent to human labelers, who are tasked with providing labels on whether a rooftop in an image includes a solar array or does not include a solar array. The resulting labeled set was used to train a machine learning algorithm to recognize rooftops as having solar arrays or not.

The machine learning algorithm was used to examine several million rooftop images, generating probability values that a given rooftop has a solar array or not. Rooftops that were identified as having a high probability of a solar array are sent to human labelers to verify that the algorithm correctly determined the presence of a solar array to generate a larger, updated iteration of the training set. In the generation of the training set, a diverse set of aerial images was included, including regions of high solar penetration, regions of low solar penetration, diverse rooftop designs, diverse geographies etc.

The machine learning algorithm picks out both solar photovoltaic installations and solar hot water heating installations, as both types of system often look similar in overhead imagery. In addition, the aerial imagery comes from a variety of flights, with imagery captured at different times, so the installation count is not expected to include the most recent installations.

Summary of Key Assumptions and Criteria

Variable	**Assumed value**
Module efficiency	15.3%
Module dimensions	1.650m x 0.992m
Module power rating	250w
Panel mounting tilt	In the plane of the roof
Maximum installable pitch	60 degrees
Temperature coefficient	-0.5%
Annual power degradation	0.5%
Number of sun positions per year	~50, depending on latitude
DC to AC derate factor	0.85
Minimum subarray size	4 panels
Minimum rooftop capacity for inclusion in technical potential	2kW
Minimum sun threshold	75% of optimum in the county or city
Packing Density	1.0

Definition of terms

Term	Description

Last updated	The last date when the underlying data was updated.
Buildings _% Solar Viable	The percentage of buildings that are viable for solar under the technical potential criteria
_% data coverage	The percentage of buildings with data in Google Maps that is covered by the solar potential data.
Roofs (%)	The percentage of buildings that are viable for solar under the technical potential criteria
Roofs (#)	The number of buildings that are viable for solar under the technical potential criteria
Roof space (square feet)	The roof area where solar panels can be placed using the technical potential criteria
Capacity (MW)	The total capacity that meets the technical potential criteria
Energy generated (MWh per yr)	The total energy production potential of the solar capacity that meets the technical potential criteria
Per roof median, roof space (sq ft)	The amount of roof space that meets the technical potential criteria for the median building
Per roof median, capacity (kW)	The amount of solar capacity potential that meets the technical potential criteria for the median building
Per roof median, energy generated (kWh)	The total energy production potential of the solar capacity that meets the technical potential criteria for the median building
Carbon dioxide (metric tons)	The potential carbon dioxide abatement of the solar capacity that meets the technical potential criteria. The calculation uses eGRID subregion CO2 equivalent non-baseload output emission rates (EPA data).
Passenger cars (taken off the road for 1 yr)	The potential carbon dioxide abatement of the solar capacity that meets the technical potential criteria, stated in terms of the equivalent number of cars taken off the road. The calculation assumes 4.73 metric tons

	CO2E /vehicle/year, based on EPA data.
Tree seedlings (grown for 10 yrs)	The potential carbon dioxide abatement of the solar capacity that meets the technical potential criteria, stated in terms of the equivalent number of trees planted and grown for 10 years. The calculation assumes 0.039 metric ton CO2 per urban tree planted, based on EPA data.

Known Issues

- Project Sunroof data relies on Google's aerial imagery, and not all of the U.S. is covered by this imagery.
- Depending on capture date, aerial imagery may not reflect new construction, the removal or growth of obstructions such as trees, etc.
- Project Sunroof calculates the solar potential on buildings which are present in Google Maps (as denoted by outlines in the Maps view), and not all buildings are included in this data. However, sometimes, Project Sunroof mistakenly calculates viability for large objects that aren't buildings (e.g., bridges).
- Boundaries between weather station sometimes show up as sharp lines when the weather data from the two stations is markedly different