Leveraging Geospatial Big Data for Smart City Digital Twins: A Framework for 3D Modeling and Solar Energy Assessment

Muhammed Yahya Bıyık[1], Muhammed Oğuzhan Mete[1]

[1] ITU, Civil Engineering Faculty, Maslak, İstanbul, Türkiye (biyik16, metemu)@itu.edu.tr

Keywords: Digital Twin, Smart City, Rooftop Photovoltaics, LoD2 Building Modelling, QGIS, SAGA GIS.

Abstract

This study presents a semi-automated Level of Detail (LoD) 2 building modelling and analysis framework, implemented using fully open-source geographic information system (GIS) software, for the high-accuracy identification of rooftop photovoltaic (PV) potential in smart city digital twins. The LoD 1 building model, traditionally based on two-dimensional building footprint data used in large-scale building modelling, systematically overestimates solar energy potential as it neglects roof pitch, orientation and shading from the immediate surroundings. In this study, analyses were conducted in a high-density urban area of Birmingham, UK, using 1-metre resolution aerial LiDAR point clouds provided by the UK Department for Environment, Food and Rural Affairs (DEFRA). Throughout the process, reliance on proprietary software was completely eliminated; point cloud pre-processing, building boundary extraction using DBSCAN clustering and Concave Hull algorithms, and orthogonalisation filters were carried out entirely within the QGIS environment. Dynamic solar radiation simulations were performed using the SAGA ‘Potential Incoming Solar Radiation’ algorithm. The findings of the study provide a scalable framework for sustainable urban planning by enhancing LoD of modelled buildings and laying the groundwork for the semantic enrichment of urban digital twins. Furthermore, it generates outputs that provide concrete support to end-users, thereby helping achieve global net-zero targets.

1. Introduction

The shift towards decentralised urban energy systems necessitates the accurate identification of rooftop photovoltaic (PV) potential in buildings within the context of smart city planning. Local authorities and energy suppliers must simulate the energy production capacity of each building in urban areas at a microscale to achieve urban carbon-neutral targets (Brito et al., 2012). However, energy models historically used by local authorities have largely relied on two-dimensional building footprints and Level 1 (LoD1) building models (Nouvel et al., 2017). This geometric simplification assumes that roofs are completely flat and unobstructed; it completely disregards roof pitch, aspect orientation, and micro-shading caused by neighbouring buildings and trees (Strzalka et al., 2015). Consequently, analyses based on LoD1 models systematically overestimate urban solar energy potential by significant margins (Waqas et al., 2023).

Three-dimensional building models at LoD2 and above that incorporate roof morphology significantly improve simulation accuracy by eliminating these geometric limitations (Dembski et al., 2020; Biljecki et al., 2016). However, producing LoD2 models and conducting solar analysis using these models involves a multi-stage process and remains an area open to further research and development. In many traditional GIS-based studies, analyses have been conducted using the integrated toolsets of commercial platforms. However, these proprietary platforms create budgetary barriers for local authorities and independent researchers in data production processes, limiting the transparency, reproducibility and scalability of the models produced. Furthermore, multimodal data fusion workflows integrating optical satellite imagery and digital elevation models are frequently plagued by stereoscopic mismatches and geometric distortions.

To overcome these methodological challenges, this study has developed an integrated GIS workflow using the QGIS and SAGA GIS platforms, which remain entirely within the Free and Open-Source Software (FOSS) ecosystem. No proprietary tools were used at any stage of the analysis; all processes, from point cloud to building reconstruction and solar potential simulation, were completed using open-source algorithms. Focusing on a high-density urban area in Birmingham, UK, this study presents quantitative comparisons between LoD1 and LoD2 models and aims to strengthen the geometric and semantic backbone of smart city digital twins (Figure 1).

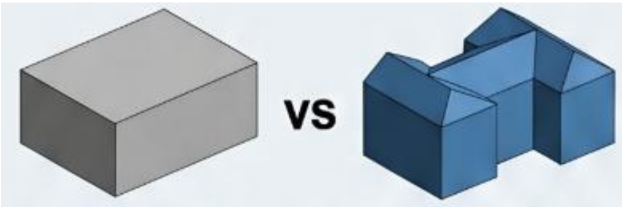

Figure 1. Geometric difference of LoD1 and LoD2 models

2. Study Area and Data Sources

An urban sector in Birmingham, United Kingdom (Figure 2), distinguished by high-density development and morphological heterogeneity, was selected as the application and validation area for this research. The region's urban fabric integrates both historical and contemporary architectural elements, presenting a highly diverse typology of roof structures including pitched, hipped, flat, and mansard forms. This complex physical environment serves as an ideal testing ground for evaluating the linear and non-linear constraints of the proposed spatial framework.

To eliminate geometric discrepancies typically associated with multi-source data fusion, the modeling phase relied exclusively on high-density aerial LiDAR (Light Detection and Ranging) point clouds. This dataset, dated 2024 and featuring a 1-meter spatial resolution, was acquired through the open-access platform of the UK Department for Environment, Food and Rural Affairs (DEFRA) (DEFRA, 2024). Provided in standardized .las and compressed .laz formats, the airborne laser scanning data delivers exceptional point density. Consequently, it effectively captures urban surface geometries, absolute elevation variations, and intricate rooftop micro-structures with centimeter-level precision.

Figure 2. Study area

3. Methodology

The proposed spatial framework consists of four key stages, ranging from processing raw point clouds to deriving a final building-level energy inventory. The entire workflow was carried out within the QGIS framework using local FOSS algorithms.

3.1 Building Boundary Extraction and Normalised Digital Surface Model (nDSM)

The workflow began with the extraction of 2D base building footprints from the classified point cloud (Figure 3), in accordance with automated building extraction and regularisation principles found in the literature (Gilani et al., 2016). In this process, building boundaries were vectorized using DBSCAN clustering and Concave Hull algorithms, and subsequently refined to accurately match architectural forms using orthogonalisation filters.

Following the footprint extraction, in order to create a high-accuracy roof surface, a Digital Surface Model (DSM) was generated from the first-pass reflections of the classified point cloud, whilst a Digital Terrain Model (DTM) was generated from the ground reflections. By calculating the difference between the two raster layers, a normalised Digital Surface Model (nDSM) was obtained, representing the absolute relative elevations of buildings relative to the ground surface.

At this stage, raster interpolation techniques such as "Gaussian smoothing" or "Fill NoData" are commonly used to remove gaps and topological noise caused by sensor angle and shading in LiDAR data, but these were deliberately not applied. Spatial validation analyses have shown that these smoothing algorithms aggressively erode the actual physical building boundaries and artificially distort roof slopes. Therefore, in subsequent stages, raw LiDAR noise and data gaps have been retained in their original form to preserve the actual topological slopes and physical boundaries of the roofs, even at the cost of producing rougher vector boundaries.

3.2 Slope, Aspect and Roof Segmentation

Continuous Slope and Aspect raster layers were derived directly from the unsmoothed original DSM dataset to capture the true morphological complexity of the rooftops. The pixels in the Aspect raster were reclassified into discrete categories representing the primary compass directions based on the orientation of sunlight reception. Subsequently, the 2D orthogonal building footprints were spatially intersected with these aspect boundaries. This geometric intersection effectively partitioned the monolithic building footprints into multi-directional, independent roof segments delineated along the actual roof ridges. Notably, raw LiDAR noise and irregular micro-geometries were deliberately preserved rather than eliminated through smoothing algorithms. This approach ensures that the physical boundaries, topological slopes, and natural surface variations of the roofs remain unaltered, providing a highly realistic geometric baseline for the subsequent LoD2 solar radiation simulations

https://doi.org/10.5194/isprs-archives-L-4-W1-2026-11-2026

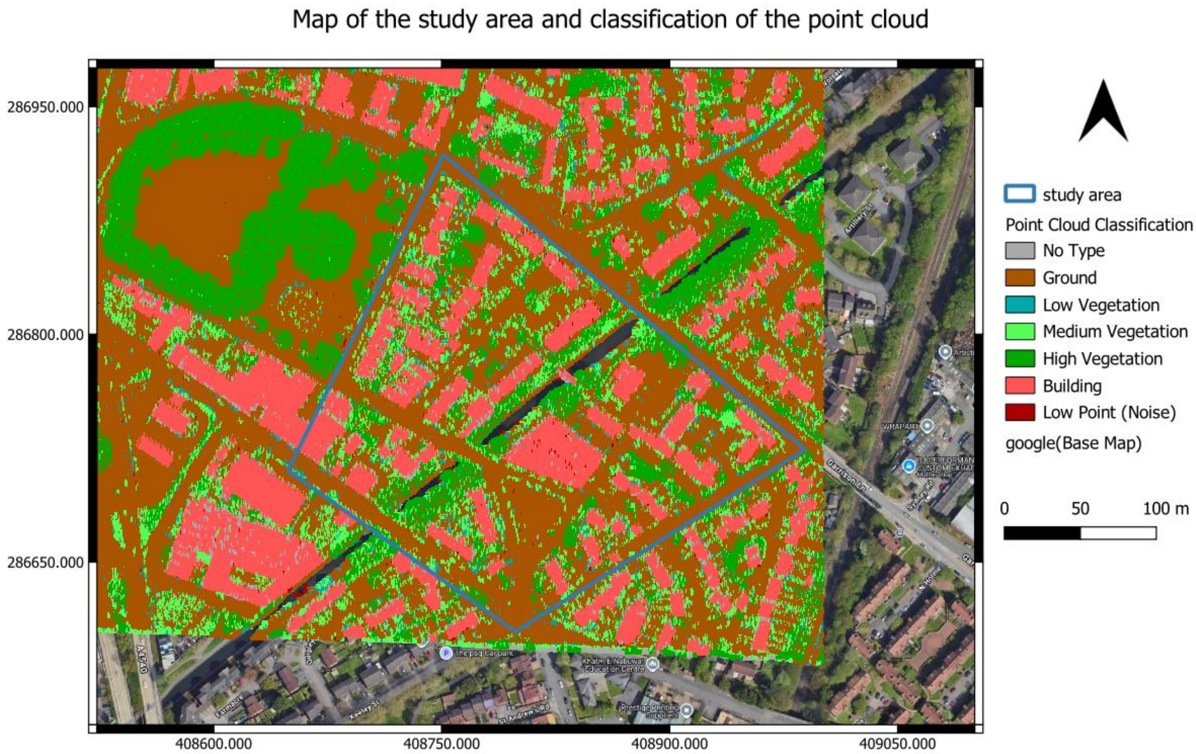

Figure 3. LIDAR data classification

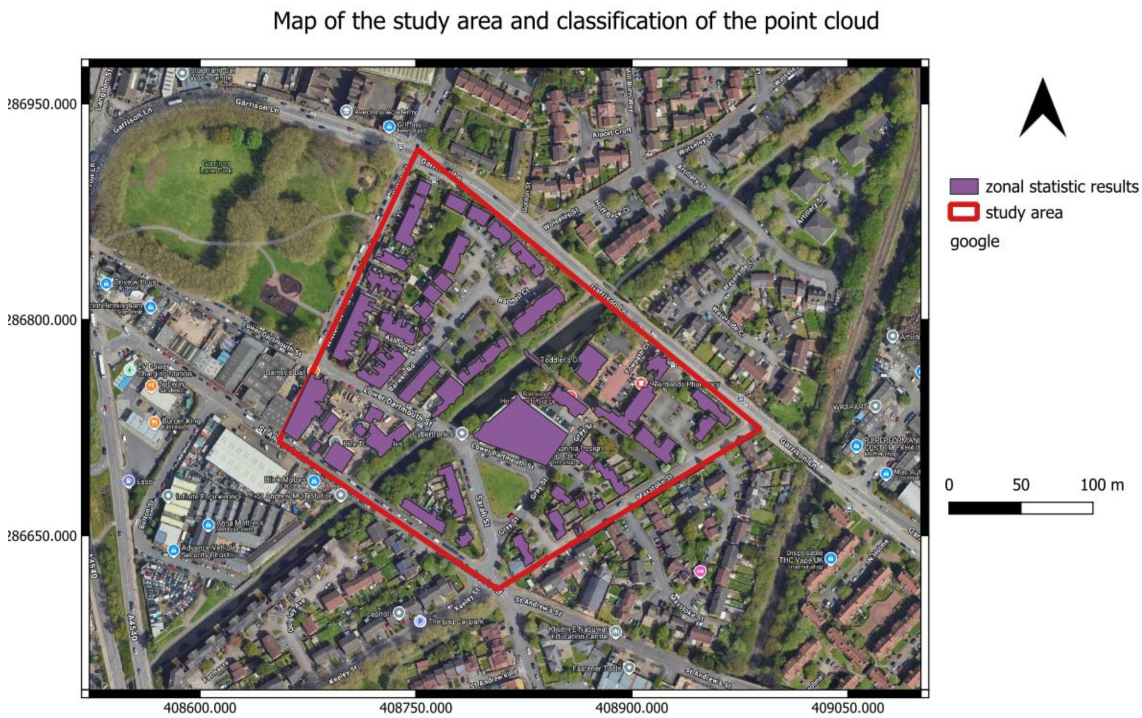

Figure 4. Point cloud based footprints

Following the footprint extraction, a Digital Surface Model (DSM) was generated from the first-pass reflections of the point cloud. To reduce the computational demand of the subsequent solar radiation algorithms, the DSM was interpolated to 2-meter spatial resolution. Furthermore, rather than analyzing the entire continuous terrain, the raster dataset was clipped to the exact extents of the extracted building footprints. This isolation ensured that the SAGA GIS algorithms exclusively processed the relevant roof geometries derived from the extracted building footprints (Figure 4), significantly improving processing efficiency.

3.3 Dynamic Solar Radiation Modelling with SAGA GIS

The fully open-source SAGA GIS (System for Automated Geoscientific Analyses) software was used to simulate solar energy potential in the QGIS environment. The 'Potential Incoming Solar Radiation' module within SAGA GIS (Conrad, 2010) dynamically models the sun's annual and daily orbital movement based on Birmingham's geographical latitude (52.48° N), rather than using static assumptions. The algorithm performs ray-tracing and viewshed analyses at the cellular level, calculating self-shading from building roofs and shadows cast by adjacent tall structures (Böhner & Antonić, 2009).

The direct solar radiation simulation within SAGA GIS is based on the following fundamental equation:

$$S_{TOP} = S_0 \, . \, \mathrm{z} \, . \sin(\mathrm{q}) \tag{1}$$

Here, S_0 denotes the hourly topographic direct radiation, S_0 the solar constant (1367 W/m²), z the binary shadow mask obtained from ray tracing (z=0 = shadow, z=1 = area receiving direct sunlight), and the elevation angle of the sun above the local horizon (Böhner & Antonić, 2009). The Sky View Factor (ws) was used to calculate the component of radiation scattered by the atmosphere. In SAGA GIS, the value of ws is formulated as follows (Böhner & Antonić, 2009):

$$\psi_s = \frac{1}{2\pi} \int_0^{2\pi} \left[\cos\beta \cos^2\varphi + \sin\beta \sin\varphi \cos(\Phi - \alpha) \right] d\Phi \tag{2}$$

Here, β represents the local slope of the roof plane, α represents the aspect angle, and φ represents the zenith angle relative to the local horizon for a specific ϕ azimuth direction. In modelling atmospheric effects, the 'Lumped Atmospheric Transmittance' (LAT) method, which is standard for urban areas and effectively represents humid/continental transitional climates, has been adopted at a rate of 70%.

3.4 Photovoltaic Energy Production Calculation

To quantitatively evaluate the overestimation inherent in traditional models, energy production was calculated separately for both LoD1 and LoD2 scenarios. For the baseline LoD1 model, the roofs were assumed to be entirely flat. The total annual Global Horizontal Irradiance (GHI) for the study area was obtained from the European Commission's Photovoltaic Geographical Information System (PVGIS) tool (European Commission JRC, 2024), using the most recent available meteorological dataset for the 2022–2023 reference period. The annual energy yield for LoD1 (E_{LoD1}) was calculated by multiplying the 2D building footprint area ($A_{footprint}$) by the PVGIS GHI value, the panel efficiency η= 0.18, and the usability coefficient C_U= 0.60. Conversely, for the LoD2 model, the calculated gross solar radiation values incorporating actual slope, aspect, and shading were assigned as attributes to the clipped 3D roof segments using the QGIS Zonal Statistics tool. To transition from theoretical solar irradiation to practical electricity yield, the same constraints were applied. The annual electricity production (E_{LoD2}, kWh/year) was calculated as follows:

$$E_{LoD2} = I_{total} . A_{usable} . \eta . C_U \tag{3}$$

Here, I_{total} represents the annual total global solar radiation intensity extracted via SAGA GIS, and A_{usable} represents the 3D roof segment area. Consistent with the baseline, η represents the 18% efficiency of standard monocrystalline panels (Green Ocean Solar, 2025), and C_U represents the 0.60 usability coefficient accounting for maintenance corridors and system losses and localized shading constraints (Acar et al., 2020)

4. Results

4.1 CityGML and Energy ADE Interoperability

Beyond the quantitative energy estimates, the outputs of this study carry direct implications for smart city data infrastructure. The LoD2 building models and associated photovoltaic attributes obtained in this study serve as both a geometric and a semantic backbone for the digital twins of smart cities. The generated three-dimensional vector layers and attribute tables go beyond merely presenting visual models; they are linked to each building's identification number (ID) and include roof area, slope, orientation, net usable area and annual electricity generation potential (kWh/year) (Mete & Yomralioglu, 2023). Digital twin platforms utilise these semantic structures to monitor urban processes in real time, conduct scenario analyses, and feed data into decision-support mechanisms.

To ensure data integrity and interoperability between systems within digital twin ecosystems, data in the GIS database are structured so they can be converted into the CityGML format, one of the OGC (Open Geospatial Consortium) standards (Biljecki et al., 2016). In particular, the 'Energy Application Domain Extension' (Energy ADE), developed within CityGML, enables the integration of dynamic data such as the physical building materials, heating/cooling demands, and local renewable energy generation capacities of buildings under a single standard schema (Agugiaro et al., 2018). This open-source workflow can directly feed the high-accuracy roof geometries and annual solar energy production capacity attributes that Energy ADE requires.

4.2 Micro-Scale Validation: Single Building Analysis

To explicitly demonstrate the geometric limitations of the baseline assumptions, a micro-scale comparison was conducted on a single large-scale industrial building (approx. 1650 m²) within the study area. As detailed in Table 1, the LoD1 model treats the entire footprint as a single, flat, and unshaded generation surface, resulting in a significant overestimation. In contrast, the proposed LoD2 framework segments the same footprint into true physical roof planes. By identifying North-facing facets and shaded areas, the LoD2 model provides a mathematically rigorous and reduced net potential, accurately reflecting the real-world operational capacity.

As demonstrated in Table 1, the baseline LoD1 model overestimates the total solar energy production by approximately 35.4% (374,727 kWh compared to 241,950 kWh) relative to the LoD2 simulation. This discrepancy primarily arises from the LoD1 model's assumption of a 1,669 m² monolithic flat footprint, whereas the LoD2 approach algorithmically reduces the usable area to 1,637 m² by accounting for structural fragmentation, roof slope, and self-shading. Most importantly, the micro-scale analysis validates the methodological decision to retain North-facing roof segments. Although the South-facing segment exhibits a higher area-specific efficiency (approximately 169.8

kWh/m²), the significantly larger North-facing geometry (828 m²) still captures a substantial yield of 136.8 kWh/m² due to high diffuse atmospheric irradiance. Consequently, the North segment contributes to nearly half of the building's total generation capacity (113,297 kWh). Rather than relying on the conventional assumption. it objectively demonstrates their actual viability.

Model Type	Roof Segment / Aspect	Usable Area (m²)	Estimated Potential (kWh/year)
LoD1 (Baseline)	Entire Footprint (Flat)	1669	374727
LoD2 (Proposed)	South-facing segment	576	97856
	North-facing segment	828	113297
	East-facing segment	69	9314
	West-facing segment	163	21483
	Total LoD2 Yield	1637	241950

Table 1. Assessment of estimated potential between models

4.3 Limitations and Future Work: EPC Data Integration

The next stage of this spatial framework is to establish the urban energy supply-demand balance. The current model accurately maps only the ‘supply’ side of the energy equation using rooftop PV generation. For cities to achieve true energy autonomy and for Positive Energy Districts (PEDs) to be planned, ‘demand’ data must also be incorporated into the system (Agugiaro et al., 2018).

In this regard, it is planned to integrate open-access Energy Performance Certificate (EPC) datasets for buildings in the Birmingham study area into the digital twin platform. By matching the annual electricity and cooling consumption values contained in the EPC data with the LoD2 rooftop PV generation capacities of the buildings, energy self-sufficiency analyses can be carried out at a micro-scale. This will result in a dynamic digital twin simulation tool for balancing loads on the urban grid and making smart urban regeneration decisions.

5. Conclusions

In this study, a repeatable and semi-automated LoD2 building modelling workflow based entirely on open-source GIS tools (QGIS and SAGA GIS) was successfully implemented to determine the rooftop photovoltaic potential in the digital twins of smart cities. The avoidance of proprietary software in GIS processes has demonstrated that the developed method offers a low-cost and accessible alternative for local authorities and independent researchers.

In the proposed modelling approach, the preservation of topological LiDAR noise prevented artificial roof slope distortions and, when combined with SAGA GIS's ray-tracing-based shadow analysis, ensured the most realistic irradiance values were obtained. The spatial distribution of annual solar energy potential under both the LoD1 baseline and the proposed LoD2 framework across the study area are compared in Figures 5 and 6, respectively. The LoD1 map (Figure 5) assigns uniform energy values to entire building footprints, whereas the LoD2 map (Figure 6) reveals the within-building variation driven by actual roof segment orientation, slope, and shading, demonstrating a considerably more nuanced and realistic distribution of generation potential.

This semantic LoD2 database, which is compatible with open international standards such as CityGML and Energy ADE, will serve as a powerful decision-support tool for developing dynamic smart-city digital twins, potentially integrated with EPC consumption data in the future (Agugiaro et al., 2018; Mete & Yomralioglu, 2023).

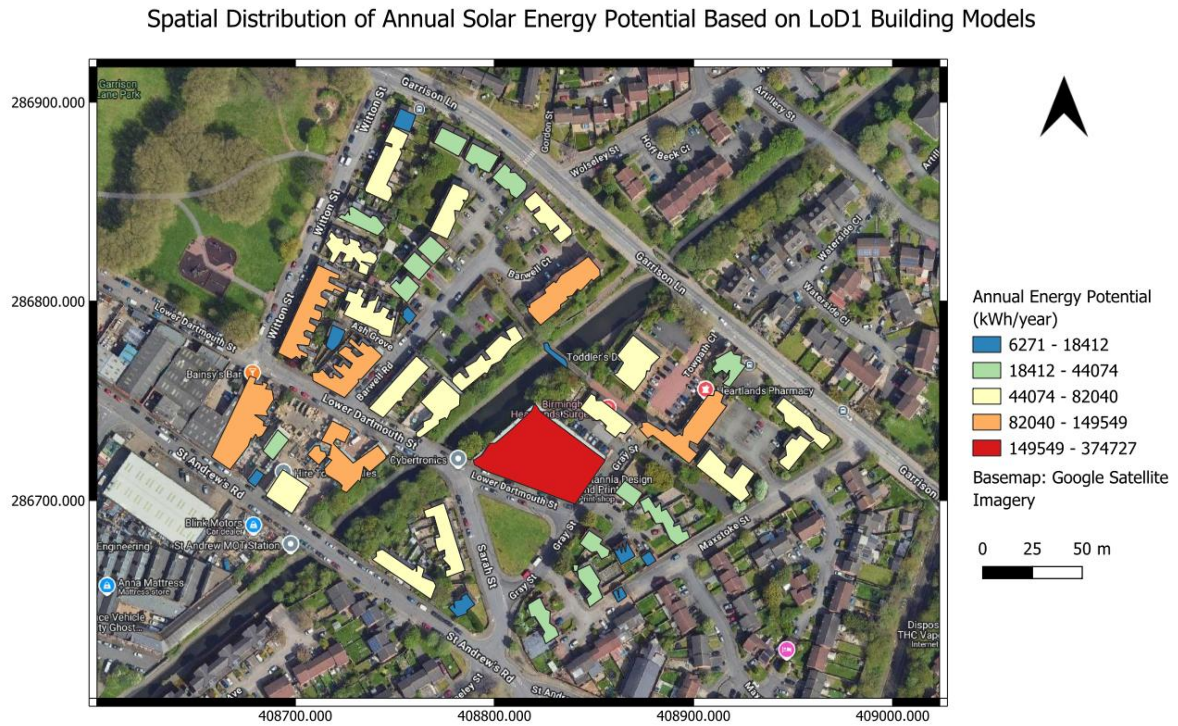

Figure 5. LoD1-based annual solar irradiation potential estimation

Figure 6. LoD2-based annual solar irradiation potential estimation

References

Acar, A., Sarı, A.C., Taranto, Y., 2020. *Rooftop solar energy potential in buildings - financing models and policies for the deployment of rooftop solar energy systems in Turkey*. SHURA Energy Transition Center, Istanbul. ISBN: 978-625-7329-04-0.

Agugiaro, G., Benner, J., Cipriano, P., Nouvel, R., 2018. The CityGML Energy Application Domain Extension for multi-scale urban energy simulation. *Open Geospatial Data, Software and Standards*, 3(1), 2.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25-37.

Brito, M.C., Gomes, N., Santos, T., Tenedório, J.A., 2012. Photovoltaic potential in a Lisbon historical district: Individual budgets and collective assessment. *Applied Energy*, 94, 353-359.

Böhner, J., Antonić, O., 2009. Land Surface Parameters Specific to Topo-Climatology. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry - Concepts, Software, Applications. Elsevier, pp. 195-226.

Conrad, O., 2010. Potential Incoming Solar Radiation. SAGA GIS Module Library Documentation.

DEFRA, 2024. DEFRA Data Services Platform: LiDAR Composite Dataset. Environment Agency.

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., Yamu, C., 2020. Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. *Sustainability*, 12(6), 2307. doi.org/10.3390/su12062307

European Commission JRC, 2024. Photovoltaic Geographical Information System (PVGIS). European Commission, Joint Research Centre. re.jrc.ec.europa.eu/pvg_tools/en/ (24 July 2024).

Gilani, S., Awrangjeb, M., Lu, G., 2016. An automatic building extraction and regularisation technique using airborne LiDAR point cloud data. *Remote Sensing*, 8(3), 258.

Green Ocean Solar, 2025**.** Solar Panel Efficiency Explained: 18% vs 20% vs 23% Made Easy. *Green Ocean Solar Blog*. greenoceansolar.com/blog/solar-panel-efficiency-explained (25 July 2026).

Mete, M.O., Yomralioglu, T., 2023. Development of a multipurpose 3D city model for sustainable smart cities: a case study of Istanbul. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W7-2023, 141–147.

Nouvel, R., Bahu, J.M., Gruzinger, I., Coors, V., Fitzpatrick, P., Peach, L., 2017. Development of the CityGML Energy ADE – Application Domain Extension for Building Energy Simulation. Proceedings of Building Simulation 2017, San Francisco, USA.

Strzalka, A., Bogdahn, J., Coors, V., Eicker, U., 2015. 3D City Modeling for Urban Scale Building Energy Demand Forecasting. *Energy and Buildings*, 90, 11-22.

Waqas, H., Jiang, Y., Shang, J., Munir, I., Khan, F.U., 2023. An Integrated Approach for 3D Solar Potential Assessment at the City Scale. *Remote Sensing*, 15(23), 5616.