Automatic Estimation of Building Construction Year and Height from Earth Observation Data for Urban Risk Assessment

Mateo Gašparović [1], Filip Radić [2], Iva Gašparović [3], Mario Uroš [4]

[1] Chair of Photogrammetry and Remote Sensing, Faculty of Geodesy, University of Zagreb, Kačićeva 26, Zagreb, Croatia – mateo.gasparovic@geof.unizg.hr
[2] Chair of Photogrammetry and Remote Sensing, Faculty of Geodesy, University of Zagreb, Kačićeva 26, Zagreb, Croatia – filip.radic@geof.unizg.hr
[3] State Geodetic Administration, Gruška 20, Zagreb, Croatia – iva.gasparovic@dgu.hr
[4] Department of Engineering Mechanics, Faculty of Civil Engineering, University of Zagreb, Kačićeva 26, Zagreb, Croatia – mario.uros@grad.unizg.hr

Keywords: Remote sensing, Building variable extraction, Cloud-based processing, Multispectral time-series analysis, Very high-resolution satellite imagery (VHRSI).

Abstract

Reliable urban risk assessment requires accurate and up-to-date information on building characteristics, particularly construction year and height, which are often incomplete or unavailable in existing databases. This study presents a cloud-based methodology for the automatic estimation of these parameters using multispectral and very high-resolution Earth Observation (EO) data. The proposed approach integrates temporal analysis of multispectral satellite imagery (Sentinel-2 and Landsat) with photogrammetric processing of very high-resolution stereo imagery (Pléiades). Building construction year is estimated by detecting temporal changes in spectral indices using spline-based modeling and discrete-difference analysis, achieving an accuracy of better than ±3 years. Building height is derived from digital surface models generated from satellite stereo imagery, with a mean accuracy of less than 2 m relative to LiDAR reference data (~1.40 m). The methodology was implemented in a cloud computing environment (Google Earth Engine and Google Colab) and tested in the City of Zagreb, Croatia. Validation results show robust performance, with an F1-score of 0.819 for construction year estimation and strong agreement between EO-derived and LiDAR-based height values. The results demonstrate the potential of EO-based methods for scalable, reliable extraction of building information, thereby supporting improved urban risk assessment and decision-making.

1. Introduction

Rapid urbanization and increasing exposure to natural hazards have intensified the need for reliable and spatially detailed urban risk assessment (Rentschler et al 2023). A critical component of such assessments is the availability of accurate building-related information, particularly construction year (Hu et al. 2024) and building height (Esch et al. 2022). These attributes play a central role in understanding structural vulnerability, estimating population exposure, and modeling potential impacts of disasters. However, in many regions, building inventories remain incomplete, outdated, or inconsistent, limiting their usefulness for risk analysis and decision-making.

Construction year provides insight into the structural standards and regulations applied at the time of building construction, enabling estimates of vulnerability to seismic, hydrological, or climatic hazards (Vagtholm et al. 2023). Building height is directly linked to the number of floors and potential population exposure, both of which are fundamental variables in modeling disaster impacts and emergency response scenarios (Li et al. 2022). Despite their importance, these parameters are often unavailable at large scales or lack the temporal consistency required for dynamic risk assessment (Ma et al. 2023).

Recent advances in Earth Observation (EO) technologies offer new opportunities to address these data gaps. The increasing availability of long-term multispectral satellite archives (Frantz et al. 2023) and very high-resolution satellite imagery (Zhao et al. 2023) enables the extraction of detailed urban information in a consistent and scalable manner. In particular, time-series analysis of multispectral data enables detection of urban growth and land-cover changes (Gašparović 2020, Das & Angadi 2022), while photogrammetric processing of high-resolution imagery supports the derivation of three-dimensional urban structure (Li et al. 2023, Farhan et al. 2024).

In this context, this study presents a cloud-based methodology for the automatic estimation of building construction year and height using EO data. The proposed approach integrates temporal analysis of multispectral imagery with photogrammetric techniques applied to very high-resolution satellite data. By leveraging large-scale satellite archives and cloud computing platforms, the methodology enables efficient and scalable processing across extensive urban areas (Xu et al. 2022).

To address the aforementioned data gaps, this study proposes a cloud-based methodology for the automated estimation of key building characteristics (construction year and building height) using multispectral and very high-resolution satellite imagery. The approach combines temporal analysis of multispectral data (Sentinel-2, Landsat) with photogrammetric processing of stereo and multi-view imagery (e.g., Pléiades, WorldView), enabling the detection of urban development patterns and the extraction of three-dimensional building characteristics. The developed algorithm for construction year estimation leverages spectral changes and building appearance signals, achieving an accuracy better than ±3 years when validated against reference data.
Building height estimation is performed by generating digital surface models from very high-resolution imagery and extracting

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-589-2026

per-building height values within defined footprints, achieving a mean accuracy of 2 m, corresponding to approximately one building floor. The entire workflow is implemented in a cloud computing environment (Google Earth Engine and Google Colab), allowing scalable processing across large urban areas without the need for local infrastructure (Farhan et al., 2024). Validation across multiple test sites confirms the robustness and reliability of the proposed approach, highlighting its potential for supporting urban risk assessment and decision-making processes.

2. Materials and Methods

2.1 Study Area and Data

The City of Zagreb is the capital of the Republic of Croatia and a regional center of exceptional importance. It houses significant educational, cultural, artistic, and health institutions, industrial plants, and cultural heritage of tremendous national value. The City of Zagreb has approximately 770,000 inhabitants, according to the 2021 census (close to 20% of Croatia's population), while the wider metropolitan area exceeds 1 million residents (Croatian Bureau of Statistics, 2021). Zagreb represents a complex urban environment characterized by heterogeneous building stock, ranging from historic masonry structures in the city center to more recent reinforced-concrete residential developments. This diversity makes it particularly suitable for testing methodologies for estimating the year of building construction and height. The city has also been affected by significant seismic activity, most notably the 2020 Zagreb earthquake, which highlighted the vulnerability of older buildings built before modern seismic regulations. Consequently, accurate and up-to-date information on building age and height is essential for improving urban risk assessment and resilience planning in this area.” The study area of this research is in the City of Zagreb (Figure 1), covers ~500 ha, and is known as “Trešnjevka sjever.”

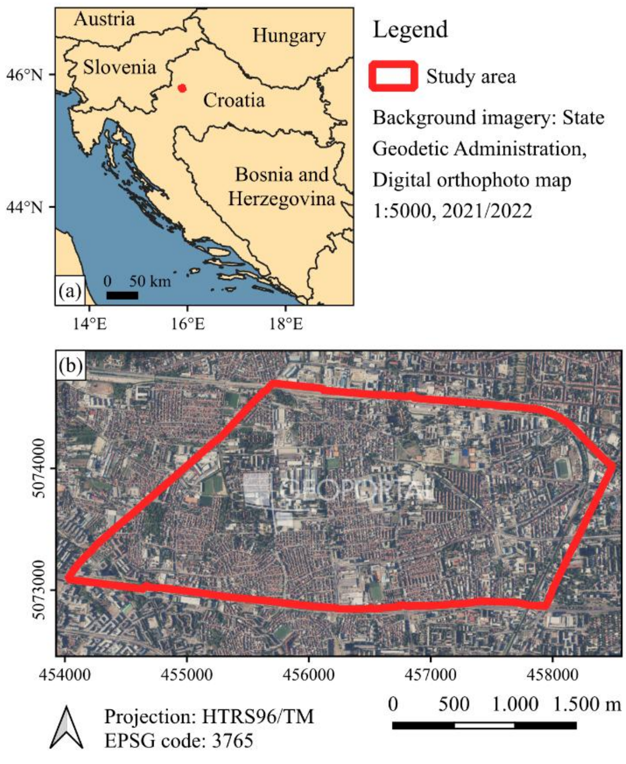

Figure 1. Study area (background imagery: State Geodetic Administration – SGA, Digital orthophoto map – DOF 2021/2022).

Obtaining imagery data is automated using the Google Earth Engine API platform, with ease of use it is possible to create a spatial database for collecting obtained satellite imagery. Imagery data were sourced from both the Sentinel-2 and Landsat constellations to construct a long-term chronological record based on Surface Reflectance (SR) products. Sentinel-2 data were obtained from the 'COPERNICUS/S2_HARMONIZED' dataset, spanning 2016 through 2023. To extend the temporal depth of the study, a harmonized Landsat collection was synthesized from the 'C02/T1_L2' archives of the Landsat 5, 7, 8, and 9 missions, covering the period from 1984 to 2023. All scenes were filtered to exclude those with cloud cover exceeding 10%, and a standard cloud-masking algorithm was applied to the remaining imagery to ensure data integrity.

The primary dataset utilized in this study for building height estimation consists of high-resolution satellite imagery acquired by the Pléiades-1A (PHR1A) sensor on June 9, 2017. The imagery is a Pansharpened (PMS) product, fusing a 0.5-meter panchromatic band with four 2-meter multispectral bands (Blue, Green, Red, and Near-Infrared) to provide a 0.5-meter-resolution color output. Geometric correction was performed using the provided Rational Polynomial Coefficients (RPCs, Gašparović et al. 2019), enabling precise orthorectification and terrain alignment.

Building footprints are obtained from OpenStreetMap (OSM) auxiliary data; each data type labeled “building” is filtered to exclude smaller, non-residential structures.

2.2 Automatic Estimation of Building Construction Year

The automatic estimation of the building construction year is based on temporal analysis of multispectral satellite imagery acquired from Sentinel-2 and Landsat Earth Observation Data. The developed algorithm detects temporal signals related to building appearance, spectral changes, and urban growth dynamics by analyzing dense time series of satellite data at the level of individual building footprints. By leveraging long-term image archives, the method identifies the most probable year of construction using change detection and temporal consistency analysis.

For the purposes of this study, several vegetation indices were used, including the Enhanced Vegetation Index 2 (EVI2) (1), Soil Adjusted Vegetation Index (SAVI) (2), and Normalized Difference Vegetation Index (NDVI) (3), Normalized Difference Built-up Index (NDBI) (4) and the Built-up Index (BUI) (5). These indices were calculated using the following formulas:

$$EVI2 = 2.5 * (NIR - RED) * ((NIR + RED) + 1), \tag{1}$$
$$SAVI = ((NIR - RED) / (NIR + RED + L)) * (1 + L), \tag{2}$$

$$NDVI = (NIR - RED) * (NIR + RED), \tag{3}$$
$$NDBI = (SWIR - NIR)/(SWIR + NIR), \tag{4}$$
$$BUI = (NDVI - NDBI), \tag{5}$$

where NIR – Near Infrared band
NIR – Red band
L – soil brightness correction factor (L=0.5)
SWIR – Short-wave infrared.

By analyzing the spatio-temporal trends of the above-mentioned spectral indices (Figure 2), spatial changes in satellite imagery can be clearly detected. A strong relationship between the indices is observed: periods of a sharp decrease in the vegetation index

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-589-2026 |

coincide with pronounced positive peaks in the building index, indicating building development.

To better represent the temporal behavior of the indices, a regression model based on spline approximation is applied. This allows smoothing of the time series and clearer identification of underlying trends. To further emphasize abrupt changes, the n-th discrete difference along the temporal axis is computed. These differences are scaled by the relative range between minimum and maximum values, resulting in a percentage change of the index.

Based on the analysis of these percentage values, threshold levels are defined to distinguish significant changes. The construction year is then estimated as the year in which the percentage of the discrete difference reaches its maximum.

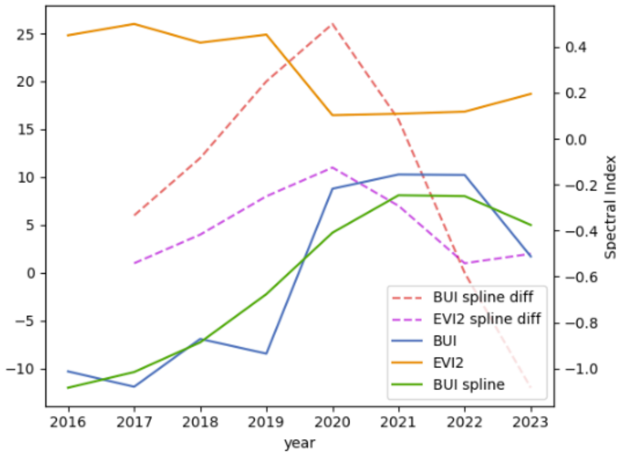

Figure 2. Spectral index time series and discrete differences for construction year detection.

2.3 Automatic Estimation of Building Height

The automatic estimation of building height is performed using very high-resolution stereo satellite imagery from Pléiades-1A via a dedicated photogrammetric workflow with Agisoft Metashape software. Geometric correction was facilitated via the provided Rational Polynomial Coefficients (RPC). Digital surface models (DSMs) are generated from stereo or multi-view imagery and used to extract per-building height values within predefined building footprints. The method incorporates photogrammetric routines and DSM filtering to approximate roof heights while minimizing noise and artifacts.

To ensure the accuracy of the extracted building heights, an Interquartile Range (IQR) filter was applied to mitigate the influence of rooftop outliers. In the study area, numerous residential and commercial structures feature antennas, chimneys, HVAC systems, and transmitters that create localized elevation spikes. To isolate the representative roof height and minimize noise, particularly in the more variable EO-derived DSM, a percentile-based mask was applied. For the satellite DSM, pixels within the 70th to 80th percentile range were retained, while a broader 70th to 95th percentile range was used for the more precise SGA LiDAR data (State Geodetic Administration). Following this outlier removal, the mean elevation of the remaining pixels within each building footprint was calculated to define the final building height.

Both algorithms (explained in sections 2.2 and 2.3) were implemented and tested in cloud computing environments (Google Earth Engine and Google Colab), enabling scalable processing across large urban areas without the need for local computational infrastructure (Farhan et al. 2024).

2.4 Accuracy Assessment

Accuracy assessment of development methods was conducted using ground-truth maps derived from field measurement data for building-year validation (Figure 2) and higher-resolution LiDAR data (State Geodetic Administration – SGA) for building height validation (Figure 3).

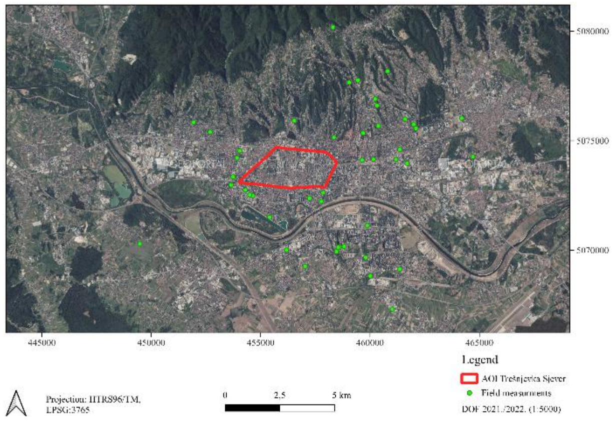

Figure 3. Building year validation ground truth map (background imagery: SGA DOF 2021/2022).

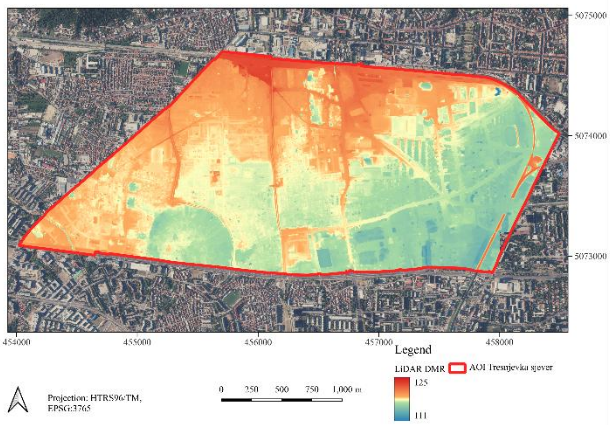

Figure 4. Building height validation ground-truth map created using SGA LiDAR data (background imagery: SGA DOF 2021/2022).

The accuracy of building construction year estimation was evaluated against a Building year validation ground truth map derived from field-measured reference data, with a target temporal frame of ±3 years. The performance was assessed using the following metrics: precision (6), accuracy (7), recall (8), and F1 score (9). The performance was evaluated based on its ability to correctly identify construction events within a permissible temporal tolerance (TP – true positive) versus instances of inaccurate dating (FP – false positive), missed detections (FN – false negative), or the correct exclusion of activity outside the satellite mission's lifespan (TN – true negative).

$$P = \frac{TP}{TP+FP} \tag{6}$$

$$A = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

$$R = \frac{TP}{TP+FN} \tag{8}$$

$$F1 = \frac{2*TP}{2*TP+FP+FN} \tag{9}$$

The accuracy of the derived building heights was validated against the SGA LiDAR reference dataset using descriptive statistics, including the mean, standard deviation, minimum, and maximum. Digital Surface Models (DSM) and Digital Terrain Models (DTM) were generated from stereo Earth Observation (EO) satellite imagery via photogrammetric reconstruction. The relative height for each building footprint was determined by subtracting the DTM from the DSM (normalized DSM). Subsequently, the aforementioned descriptive statistics were calculated for each footprint to assess vertical accuracy.

3. Results

Figure 4 presents the values of the aforementioned spectral indices derived from Sentinel-2 for the period 2016–2025. Figure 5 shows the values of the same spectral indices derived from combined Landsat missions for the period 1984–2025.

By comparing Figures 4 and 5 for the overlapping period (2016–2025), the difference in spatial resolution between the two missions becomes clearly evident. Sentinel-2, as a higher-spatial-resolution mission, produces smoother data-value curves, while Landsat provides values less affected by abrupt fluctuations.

Higher spatial resolution results in fewer boundary pixels representing transitions between built-up and non-built-up areas or vegetation. Consequently, noise caused by elements such as shrubs or tree canopies near buildings does not dominate the final pixel value, leading to a more realistic representation. In contrast, lower-spatial-resolution imagery covers a larger area within a single pixel, which may include a mixture of different objects and land cover types. As a result, boundary pixels intended to represent transitions between objects are highly variable, as is clearly visible in the figures presented.

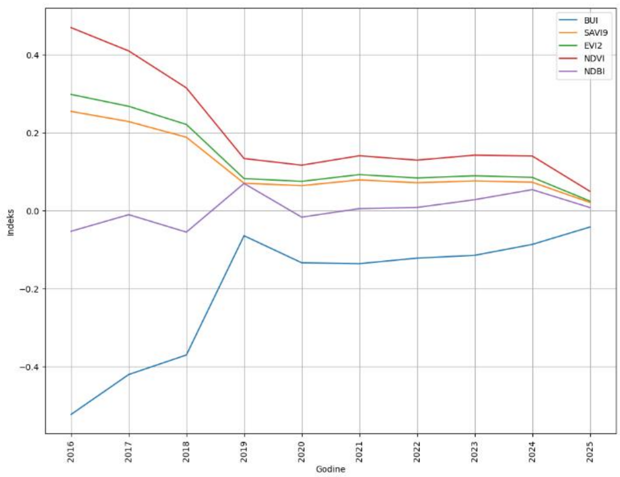

Figure 4. Values of spectral indices derived from Sentinel-2 imagery for the building (Lopatinečka ul. 4, Zagreb, Croatia).

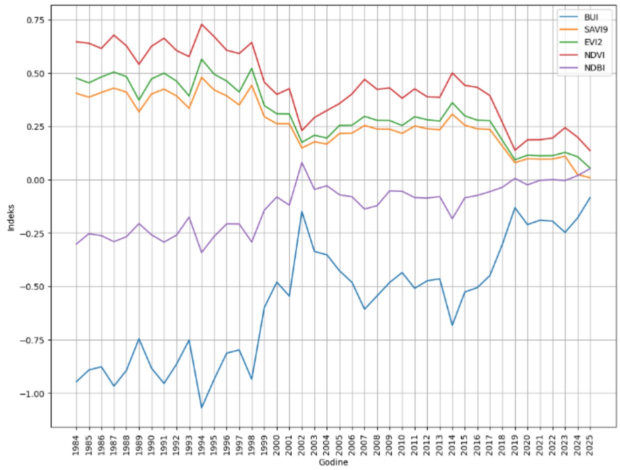

Figure 5. Values of spectral indices derived from Landsat imagery for the building (Lopatinečka ul. 4, Zagreb, Croatia).

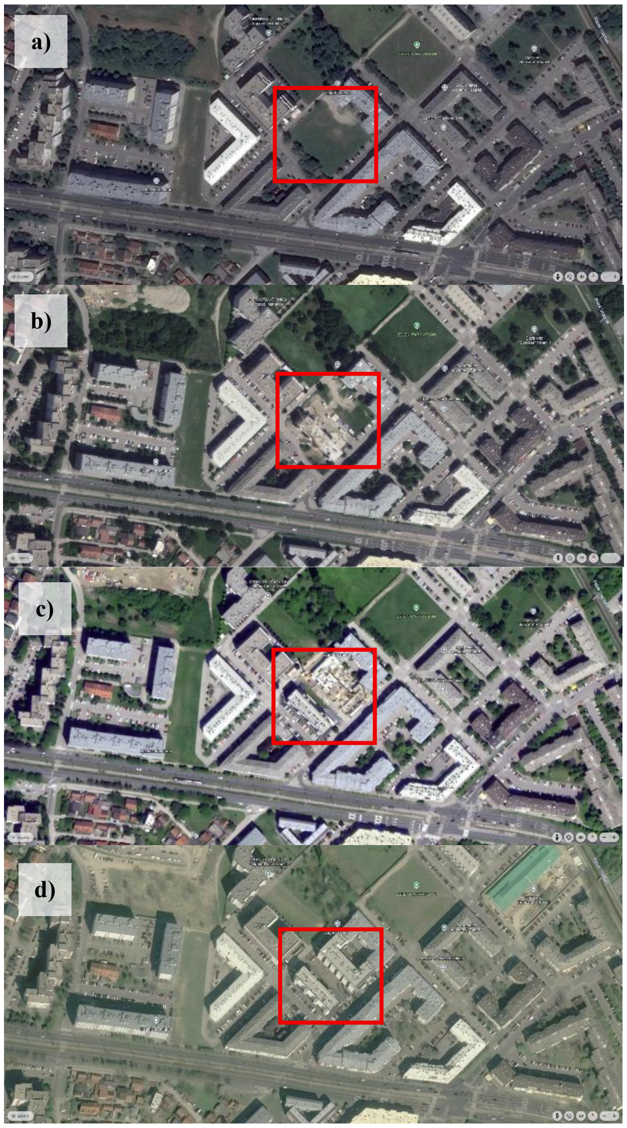

Figure 6. Historical overview of the Google Earth platform for building Lopatinečka ul. 4, Zagreb, Croatia (red square): (a) 2017, (b) 2018, (c) 2019, (d) 2020.

Figure 7 presents the final results of the automatic estimation of building heights from Earth Observation data, showing the spatial distribution of the estimates.

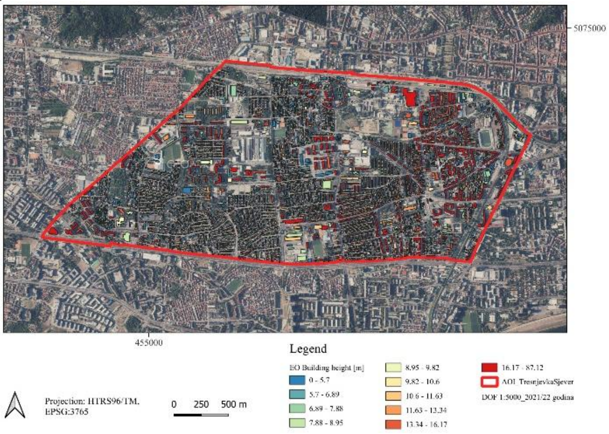

Figure 7. Automatic Estimation of Building Height from Earth Observation Data (background imagery: SGA DOF 2021/2022).

The results presented in Table 1 provide descriptive statistics for validating building height estimates derived from Earth Observation (EO, Figure 7) data against LiDAR-based reference models. The mean EO-derived height is 127.52 m, closely matching the LiDAR DSM mean of 128.29 m, while the LiDAR DMR mean is significantly lower (117.77 m), reflecting ground elevation. The average difference between EO heights and the DMR is 9.74 m, which is consistent with the DSM–DMR difference of 10.52 m, indicating consistent estimates of building heights. The relatively small mean difference (Δ=1.40 m) between EO-derived and LiDAR-derived heights confirms good agreement between methods. However, the higher standard deviation and the presence of extreme values suggest occasional outliers, likely due to noise or mismatches in complex urban structures. Overall, the results demonstrate that EO-based height estimation provides reliable and accurate building height information suitable for urban risk assessment applications.

	EO [m]	LiDAR DSM [m]	LiDAR DMR [m]	ΔEO-DMR [m]	ΔDSM-DMR [m]	Δ [m]
count	7633	7633	7633	7633	7633	7633
mean	127.52	128.29	117.77	9.74	10.52	1.40
std	8.37	5.157	1.241	8.28	5.00	6.83
min	115.72	114.67	113.71	-3.07	-2.26	0.00
25%	124.11	125.25	116.94	6.31	7.36	0.26
50%	126.72	127.69	117.70	8.84	9.82	0.59
75%	129.31	130.08	118.53	11.55	12.34	1.23
max	469.62	204.05	123.45	350.14	87.12	339.12

Table 1. Building height validation descriptive statistics.

The accuracy assessment of the building year estimation algorithm, conducted on the training dataset, is illustrated in Figure 8, while Figure 9 shows a histogram of performance based on F1-score evaluation.

It is important to note that the validation dataset exclusively contains construction events occurring within the satellite mission's lifespan; consequently, the results include no True Negative (TN) outcomes. Furthermore, no False Negative (FN) outcomes were recorded during this assessment phase. The evaluation yielded 34 true positives (TP) and 15 false positives (FP).

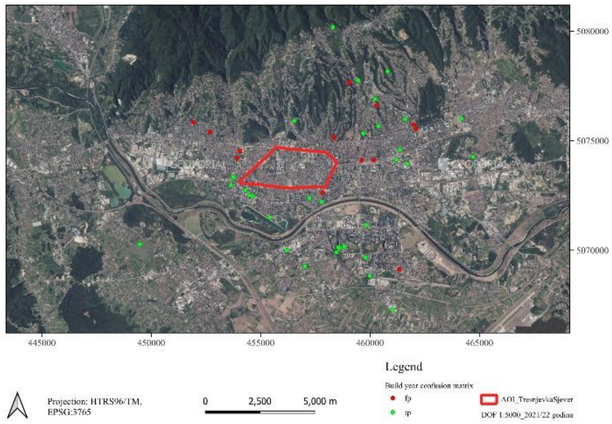

Figure 8. Building year validation ground truth map results after confusion-matrix computation (background imagery: SGA DOF 2021/2022).

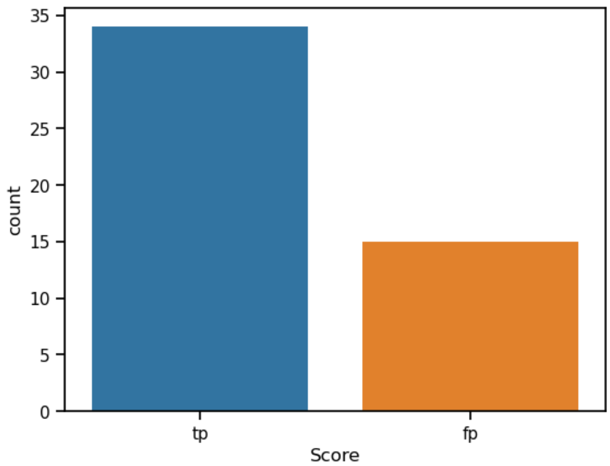

Figure 9. Building year validation F1 test.

Detailed performance metrics for the developed algorithm are summarized in Table 2. The model achieved a Precision and Accuracy score of 0.694, while the Recall reached 1.000 due to the absence of False Negatives. The resulting F1-score of 0.819 represents a strong performance, consistent with established benchmarks for machine learning applications in urban remote sensing.

	Score
Precision	0.694
Accuracy	0.694
Recall:	1.000
F1 score:	0.819

Table 2. Building year validation F1 test.

4. Conclusions

This study presented a cloud-based methodology for the automatic estimation of the year of building construction and height using multispectral and very high-resolution Earth Observation data. The proposed approach enables scalable and consistent extraction of key building characteristics across urban areas.

The results show that construction year can be estimated with an accuracy of better than ±3 years, while building height estimation achieves a mean accuracy of less than 2 m relative to LiDAR reference data (~1.40 m). Validation confirms the robustness and reliability of the methods, despite the inherent complexity of urban environments.

The presented methodology demonstrates that Earth Observation data can reliably provide two essential variables for urban risk assessment building age and height automatically and at scale. By exploiting freely available satellite archives and cloud-based processing, the approach supports continuous monitoring of urbanization, improves the completeness of building inventories, and enhances models for disaster preparedness, vulnerability analysis, and urban resilience planning.

Future work will focus on improving model generalization, reducing sensitivity to data noise, and extending the methodology to larger and more diverse urban areas.

Acknowledgements

This study was supported by the Croatian Science Foundation for the ALCAR project “Assessment of the Long-term Climatic and Anthropogenic Effects on the Spatio-temporal Vegetated Land Surface Dynamics in Croatia using Earth Observation Data” (Grant No. HRZZ IP-2022-10-5711).

References

Croatian Bureau of Statistics, 2021. Population Census. https://dzs.gov.hr/vijesti/objavljeni-konacni-rezultati-popisa-2021/1270 (1 March 2026)

Das, S., Angadi, D. P., 2022. Land use land cover change detection and monitoring of urban growth using remote sensing and GIS techniques: A micro-level study. *GeoJournal*, *87*(3), 2101-2123.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., ..., Zeidler, J., 2022. World Settlement Footprint 3D-A first three-dimensional survey of the global building stock. *Remote sensing of environment*, 270, 112877. doi.org/10.1016/j.rse.2021.112877.

Farhan, M., Wu, T., Amin, M., Tariq, A., Guluzade, R., Alzahrani, H., 2024. Monitoring and prediction of the LULC change dynamics using time series remote sensing data with Google Earth Engine. *Physics and Chemistry of the Earth, Parts a/b/c*, *136*, 103689.

Frantz, D., Rufin, P., Janz, A., Ernst, S., Pflugmacher, D., Schug, F., Hostert, P., 2023. Understanding the robustness of spectral-temporal metrics across the global Landsat archive from 1984 to 2019–a quantitative evaluation. *Remote Sensing of Environment*, 298, 113823.

Gašparović, M., Dobrinić, D., Medak, D., 2019. Geometric accuracy improvement of WorldView-2 imagery using freely available DEM data. *The Photogrammetric Record*, *34*(167), 266-281.

Gašparović, M., 2020: *Urban growth pattern detection and analysis*. In Urban ecology (pp. 35-48). Elsevier.

Hu, T., Zhang, M., Li, X., Wu, T., Ma, Q., Xiao, J., ..., Liu, D., 2024. Extraction of Building Construction Time Using the LandTrendr Model With Monthly Landsat Time Series Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 18335-18350. doi.org/10.1109/JSTARS.2024.3409157.

Ma, X., Zheng, G., Chi, X., Yang, L., Geng, Q., Li, J., Qiao, Y., 2023. Mapping fine-scale building heights in urban agglomeration with spaceborne lidar. *Remote Sensing of Environment*, *285*, 113392.

Rentschler, J., Avner, P., Marconcini, M., Su, R., Strano, E., Vousdoukas, M., Hallegatte, S., 2023. Global evidence of rapid urban growth in flood zones since 1985. *Nature*, *622*(7981), 87-92.

Li, S., He, S., Jiang, S., Jiang, W., Zhang, L., 2023. WHU-stereo: A challenging benchmark for stereo matching of high-resolution satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-14.

Li, J., Zheng, B., Bedra, K. B., Li, Z., Chen, X., 2022. Effects of residential building height, density, and floor area ratios on indoor thermal environment in Singapore. *Journal of Environmental Management*, *313*, 114976.

Vagtholm, R., Matteo, A., Vand, B., Tupenaite, L., 2023. Evolution and current state of building materials, construction methods, and building regulations in the UK: Implications for sustainable building practices. *Buildings*, 13(6), 1480.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., ..., Guo, H., 2022. Cloud-based storage and computing for remote sensing big data: a technical review. *International Journal of Digital Earth*, *15*(1), 1417-1445.

Zhao, L., Wang, H., Zhu, Y., Song, M., 2023. A review of 3D reconstruction from high-resolution urban satellite images. *International Journal of Remote Sensing*, 44(2), 713-748.