Exploratory Study on Predicting Glazed Surfaces from the Spanish Cadastre and Energy Performance Records

Daniel Garcia-Gonzalez, Jorge Rodríguez-Álvarez, Miguel R. Luaces, and Susana Ladra

Universidade da Coruña, Centro de Investigación CITIC, A Coruña, Spain
d.garcia2@udc.es, jorge.ralvarez@udc.es, miguel.luaces@udc.es, susana.ladra@udc.es

Abstract

This study examines predicting glazed surfaces in buildings by combining open data from the Spanish Cadastre and energy performance certificates. After processing both sources, a two-step modeling approach was applied: first, a classifier to detect glazing in each orientation, then a regressor to estimate the surface proportion where glazing was present. The performance of the model was limited by the quality and detail of the data, and weak regression results reveal challenges in capturing the proportion of glazing from the available attributes. Nevertheless, the work highlights both opportunities and obstacles in applying AI to open urban data, highlighting the value of linking cadastral and energy records for future large-scale building envelope characterization.

1 Introduction

Building energy performance is influenced by envelope characteristics such as insulation, material composition, and the proportion of glazed surfaces [1]. In particular, glass façades affect heating and cooling demands; however, information about their extent and orientation is often missing in open urban databases [2]. For that reason, accurate city-wide analyses require such details to estimate solar gains, daylighting, and heat losses.

This study explores the potential of open data to infer façade glazing in Spanish buildings. To this end, we integrate geometric information from the Spanish Cadastre [3], which provides footprints and heights, with Energy Performance Certificates (EPCs) [2], which offer aggregated envelope and energy indicators. By linking these sources, our aim is to predict, for each orientation, both the presence and proportion of glazed surfaces.

Rather than aiming for full predictive accuracy, the objective is to evaluate the feasibility of using existing datasets. A lightweight modeling pipeline is adopted, highlighting data gaps and methodological challenges for future urban energy mapping initiatives.

2 Study

Two main data sources were used in this study. The first was the Spanish Cadastre, which provides detailed building geometries and descriptive attributes. From these data, wall segments were derived by decomposing building footprints and height information, distinguishing between external and internal façades, and categorizing each segment by orientation (N, NE, E, SE, S, SW, W, NW). The second source was the regional EPCs, which report overall building energy demands and, in some cases, include the proportion of glazed surfaces by orientation.

Both datasets were linked using the cadastral reference code and, when necessary, other specific geometric features. The integration process retained only those buildings with complete

geometric descriptions and valid EPC information. From the combined dataset, a range of features was derived, including external perimeter by orientation, total floor area, construction year, and aggregated energy indicators. The dependent variable for modeling was the proportion of glazed surface per orientation.

A two-step modeling approach was implemented. In the first step, a classification model was trained to predict whether glazing was present on each façade. In the second step, façades identified as glazed were used to train regression models estimating the proportion of glazed surface. Both models employed geometric and energy-related predictors.

Model performance was evaluated through cross-validation. Classification results showed high accuracy, averaging around 0.90 across all orientations, while regression performance was comparatively limited, with R^2 around 0.15. These results reflect the aggregated and façade-independent nature of EPC data, as well as the limited level of detail in cadastral attributes.

Despite these limitations, the integration of cadastral and EPC data provided an initial framework for estimating façade glazing characteristics at the city scale. The analysis highlights several key challenges, including incomplete EPC coverage, heterogeneity in data quality, and lack of façade-level descriptors. Nevertheless, the proposed two-step approach demonstrates the potential of open urban data for large-scale building envelope characterization and lays the groundwork for future improvements using higher-resolution datasets.

3 Conclusions

This study examined the feasibility of predicting façade glazing proportions by combining Spanish Cadastre and EPC data through a two-step pipeline: first detecting the presence of glazing, then estimating its extent for each building orientation.

While classification performed well, regression was weak due to limited data availability and façade-level detail. Nonetheless, the study demonstrates the potential of open datasets for large-scale envelope characterization. Future work could use higher-resolution sources, such as LiDAR façades or building information models (BIM), to improve predictive accuracy.

In conclusion, open urban data can support AI-based energy analyses; however, robust façade-level modeling requires finer-grained, standardized information. These findings highlight current limitations and provide guidance for future data collection and integration efforts.

Acknowledgments

This work was partially funded by Ministry for Digital Transformation and Civil Service and "NextGenerationEU"/PRTR under grant TSI-100925-2023-1. CITIC, as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, has been partially funded by GAIN/Xunta de Galicia: [ED431G 2023/01].

References

[1] Jongwook Jeon, Jaehyuk Lee, and Youngjib Ham. Quantifying the impact of building envelope condition on energy use. *Building Research & Information*, 47(4):404–420, 2019.

[2] Galicia's EPCs. Available at: https://appsinega.xunta.es/rgeee/web/consultaPublica.php. Accessed: 26/09/2025.

[3] Spanish cadastral datasets. Available at: https://www.sedecatastro.gob.es/Accesos/SECAccDescargaDatos.aspx. Accessed: 26/09/2025.