Exploitation of Energy Performance Certificate Database in Urban Energy Modelling

Sebastiano Anselmo – Politecnico di Torino, Italy – sebastiano.anselmo@polito.it
Maria Ferrara – Politecnico di Torino, Italy – maria.ferrara@polito.it
Piero Boccardo – Politecnico di Torino, Italy – piero.boccardo@polito.it
Stefano Paolo Corgnati – Politecnico di Torino, Italy – stefano.corgnati@polito.it

Abstract

Cities are crucial for the energy transition, as recognized by the European Union in policies such as the Fit for 55 package and the Climate-Neutral and Smart Cities mission. The former calls for the revision of several directives, among which the Energy Performance of Buildings Directive (EPBD) plays a major role, targeting the phasing out of fossil fuels and the achievement of minimum performance objectives for all existing buildings. It reinforced the role of the Energy Performance Certificate (EPC) as a shared evaluation schema. However, considering the 30-50% coverage of EPCs in the European building stock, new methodologies and models are required to assess the building stock extensively. Considering the valuable data contained in EPCs, these can be used to train Urban Building Energy Models which leverage the potential of Earth Observation. This study proposes a new method to segment the building stock according to thermographic pictures, resorting to EPC information for the energy class distribution analysis. Thermographic values are used to assess thermal losses and replicate the energy class distribution accordingly. Different EPC data are assessed in order to understand the best configuration both in terms of share between training and validation data and of the need for potential pre-filtering. The method appears to be reliable – with 66% of buildings classified correctly on average – yet simple, thus being attractive for policymakers to define retrofitting campaigns able to meet European requirements. With simplicity and flexibility being the main strengths of the method, it is also possible to consider additional inputs and make the model more complex to improve the accuracy.

1. Introduction

The energy transition challenge is to be addressed at first in cities, which have the potential to effectively cut two-thirds of their energy-related emissions (IEA, 2016). Moreover, initiatives such as the EU Climate-Neutral and Smart Cities mission (European Commission, 2023) – providing funding to local administrations for the implementation of measures for climate neutrality – fostered the debate on the most effective ways to reduce impacts and mitigate the effects of climate change.

The energy performance of the building stock, together with the possibility to cover the total demand as much as possible from Renewable Energy Sources, has been considered in international policies, issued by the European bodies, thus favouring an increase in the yearly deep renovation rate of the building sector from the current 0.2% up to the target of 3%.

The Energy Performance of Buildings Directive (EPBD) recast (Council of the European Union, 2023b) – framed into the Fit for 55 package (Council of the European Union, 2023a), which aims to align the EU with its decarbonisation goals – set a minimum performance objective in the upgrade of all buildings falling into the energy performance class G and defined a path for the complete decarbonisation of the building stock. Both processes are based on the Energy Performance Certificate (EPC) – required before renting or selling and after deep renovation processes – in which an auditor compares the real performance to the ones of a reference building and provides a comprehensive evaluation. The relevance of energy performance is pointed out

Part of
Pernigotto, G., Ballarini, I., Patuzzi, F., Prada, A., Corrado, V., & Gasparella, A. (Eds.). 2025. Building simulation applications BSA 2024. bu,press.
https://doi.org/10.13124/9788860462022

also by the recast of the Energy Efficiency Directive (European Parliament and Council, 2023), which introduced the "energy efficiency first" principle, requiring all policies – both energy-centred and not – to consider their implications on this topic.
In this framework it is necessary to elaborate tools able to support energy policy by assessing the building stock on a wide scale, starting from the identification of the worst-performing buildings. Indeed, EPCs cover only a small portion of the building stock e.g., in Italy 10% (Pagliaro et al., 2021), making it necessary to model the remaining share through energy models and algorithms (Johari et al., 2020). Still, EPCs can be used to gather the inputs to define archetypes in Urban Building Energy Models, which consider both geometrical and energy-related parameters to estimate energy consumptions, thus going beyond the problem of unbundled data gathering – costly in terms of time and money (Deng et al., 2023). Data homogenisation – referring the information to entire buildings – can be performed through Geographic Information Systems (GISs), powerful tools in energy modelling thanks to the possibility to consider different layers and spatial resolutions (Yu et al., 2021).
This paper – starting from the potential to use EPCs in training datasets for Urban Energy Models recognised in existing studies (Conticelli et al., 2024; Johari et al., 2023) – aims to leverage infrared thermography to estimate energy consumptions of the *Barriera di Milano* neighbourhood in Turin, Italy. A review by Martin et al. (2022) has highlighted the potential of thermography in quantifying thermal losses, one of the principal indicators for energy classification. Therefore, from the assessment of the surface temperature difference between buildings – considering a constant internal temperature as set by law – it is possible to classify buildings according to their thermal signature.

2. Simulation

This research aims to improve a previously established method for buildings' energy classification based on the combination of EPCs and infrared thermography (Anselmo et al., 2023). The process – whose workflow is pictured in Fig 1 – can be divided into two main components: 1) data gathering and pre-processing of both inputs – the EPC dataset and the thermographic picture; 2) attribution of a class itself. This second step, in which the class distribution observed in the EPCs is replicated based on thermal losses, is better detailed in Section 2.3.1. A sample of the EPCs is not used for the training, being kept for validation.

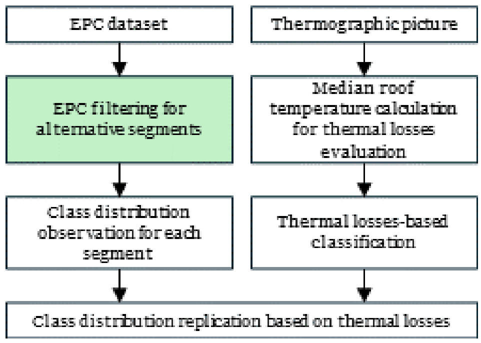

Fig. 1 – Workflow of the proposed methodology

While a first element to consider for the refinement of the drafted methodology is the segmentation of the building stock, this paper focuses on the optimal selection of training data from the EPCs database, highlighted in green in Fig. 1. It is believed that by guaranteeing the trustworthiness of the training dataset, final results can be highly improved.

2.1 Data Preparation

EPCs are gathered from the open portal of the Piedmont Region (Osservatorio ITC della Regione Piemonte, 2024) divided by section and filtered based on cadastral sheets in order to keep only those on the Area of Interest (AoI). Based on the available information, selection and aggregation criteria are defined, considering whether the same parameter is repeated in different sections – and therefore in different tables. All data – processed through DB Browser for SQLite – are merged first according to a univocal ID to return a single entry for each EPC and then by considering the address to have one value for each volumetric unit, the minimum unit of geospatial data derived from the Municipal Technical Map.
Also thermal values derived from Infrared Thermography are referred to volumetric units. The QGIS zonal statistics tool was run to obtain the

median temperature value – chosen instead of the average to mitigate the relevance of extreme values – from a thermal orthophoto produced with pictures acquired through a FLIR A8581 MWIR HD camera.

2.2 Correlation Analysis

In order to understand the most critical indicators to be considered when estimating the energy performance from EPCs, it is necessary to first look for correlations between the energy class and specific characteristics of the buildings.

According to Conticelli et al. (2024), the period of construction is the principal information among building parameters, because it mirrors the evolution of the technical and normative standards in the building industry. Therefore, this analysis started from the year of construction. It was also decided to consider the year of the last renovation in order to check whether refurbishments correspond to significant increases in energy performance by making the building compliant with legislation.

Finally, the potential correlation with the Surface-to-Volume (SV) ratio is explored, so as to understand whether there is a building typology whose consumptions are generally lower.

2.3 Energy Classification

2.3.1 Class attribution

This thermography-based methodology applies the class distribution observed in the EPCs dataset to the volumetric units based on roof temperature values. By assuming a constant internal temperature – defined by law to be 20 °C for residential buildings – the temperature of the envelope returns the thermal losses. Therefore, by ordering volumetric units according to the temperature value, it can be assumed that the order goes from the least to the most performing building.

The class distribution can be observed for the whole EPC database or filtering the entries according to specific fields. Therefore, different samples are extracted. First, the use of the building will be considered: as the premises are based on residential buildings, it can be assumed that the selection of residential buildings only would refine the results. The other two elements to base the selection on are the motivation for the EPC issue and the scale of the analysis. EPCs are released both before selling or renting a building or building unit and after a deep renovation; it can be assumed that the EPCs released after refurbishing are more reliable, based on technical sheets of newly-installed technology – as the thermal transmittance of fixtures – while the obsolescence of existing systems in units to be rented or sold may cause data gaps. As for the scale of analysis, it is relevant to note that this method uses the volumetric unit as the minimum unit of analysis. What needs to be explored is the difference between considering a single EPC issued for the whole building or multiple EPCs for single building units.

In all cases, how the classification changes when increasing the share of buildings used for training, keeping the remaining for validation will be assessed. In particular, combinations of 20% for training and 80% for validation, 50/50% and 80/20% will be considered. As a result, a total of 15 simulations will be carried out:

- Three without EPC filtering.
- Three based on EPCs covering whole buildings only.
- Three extracting only the EPCs issued for residential units or buildings.
- Three for EPCs issued after deep renovations.
- Three for EPCs issued before renting or selling the unit.

2.3.2 Validation

As previously mentioned, pre-defined shares of EPCs will be used when training the estimation model, keeping the remaining for validation, carried out by comparing correct and incorrect class attributions. True positives and false positives will be considered to compute the Receiver Operating Curve (ROC), which plots on the X axis the share of misallocations and on the Y axis the share of correct classifications. In this research, all volumetric units classified within a ±1 deviation from the exact class are considered as True Positives. Variations based on alternative filters and sampling on the full EPC dataset are assessed.

3. Discussion and Result Analysis

3.1 Data Preparation

The four tables concerning the four EPC sections required to gather the necessary information were processed through SQL code in DB Browser for SQLite and aggregated according to the address in order to return one value for each civic number, to be attached to volumetric units. As a result, 346 out of 688 volumetric units in the AoI (50%) were characterized with EPC data.

On the other hand, the median roof temperature – crucial to define the performance of the building envelope – was computed for each volumetric unit, with values ranging from -1.09 °C to 15.94 °C.

3.2 Correlation Analysis

The first analysis considered the potential correlations between specific information considered into EPCs, starting with the year of construction. However, a clarification is needed: out of 346 buildings for which the correlation is assessed, 266 (77.1%) are in the two least performing classes; it is not surprising to observe a widespread presence of these two classes in all the observed categories.

Although there is not a direct relationship between the year of construction and the energy performance class, in Fig. 2 it can be observed that the most recent buildings are included in the most performing classes, thus remarking the relevance of recent energy policies, especially from 1990s. A similar trend could be observed also on the opposite side of the graph: with the exception of class D, it can be noticed that by increasing the energy class, the oldest building of every class has been realised progressively later. On the other hand, it is not possible to observe any correlation between the year of the last renovation – reported in 79 cases (23%) – and the energy performance. Therefore, by looking at the wide presence of buildings refurbished after 2010 in classes F and G, it can be assumed that most renovations are not related to energy performance.

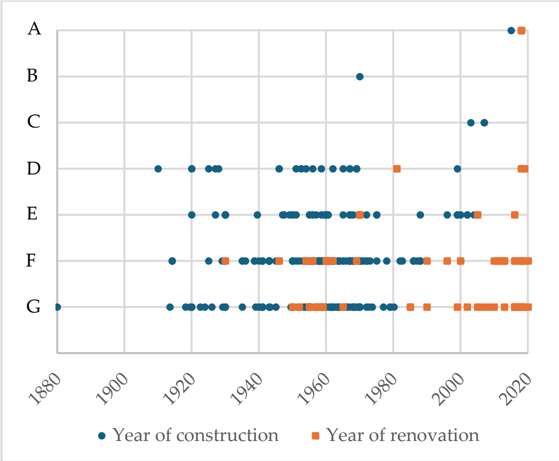

Fig. 2 – Correlation between year of construction and year of renovation and the energy performance class

A similar reasoning can derive from the observation of the relationship between the energy classification and the SV ratio, with a homogeneous distribution of the values. It is only possible to mention a slight tendency towards a worsening of the class when increasing the SV ratio – standing for less compact building typologies.

3.3 Energy Classification

3.3.1 Class distribution observation

Together with the roof temperature, the class distribution which can be observed in the EPC dataset is the crucial input. In this study, this component is varied by filtering the EPCs according to five criteria and further subdividing with different shares between the training and validation sets. From this, it results that different class distributions are observable in the AoI, plotted in Fig. 3.

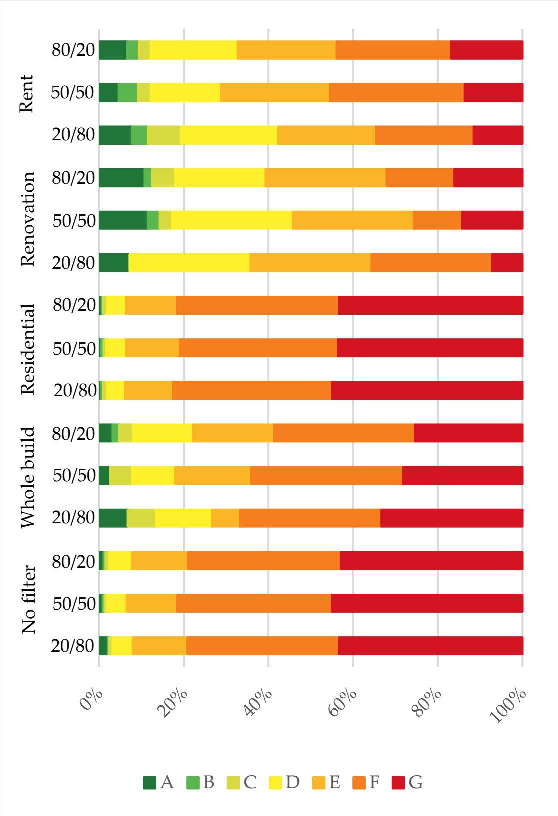

Fig. 3 – Class distribution according to different filtering

The prevalence of the two least performing classes – mentioned above – can be widely observed, with most classifications including more than 50% of the values in these two. This is valid especially when no filter is applied and for residential buildings, since more than 90% of buildings belong to classes E to G. On the other hand, it can be observed that by considering only renovated buildings there is a wider incidence of the higher classes – especially A – despite the lack of correlation observed in Section 3.2. A final element to be highlighted is the scarce presence of class B buildings: in the full database, only seven volumetric units – approximately half of the ones falling in classes A and C – are included in this class.

3.3.2 Class attribution

The class attribution is based on the roof temperature; thus all classifications are comparable in terms of the area where least performing buildings are located. This is the Western part, around Respighi Square, generally characterised by a prevalence of class G buildings. On the contrary, it can be observed that there are homogeneous urban portions, in particular in the Eastern part – around Cravero Street – where there is a stronger incidence of classes D and E. Due to the low share of buildings included in the higher classes, it is not possible to observe a clear concentration of volumetric units in classes from A to C – unless by observing the classification according to EPCs following a renovation, where there is a cluster of 11 buildings in the two most performing classes.

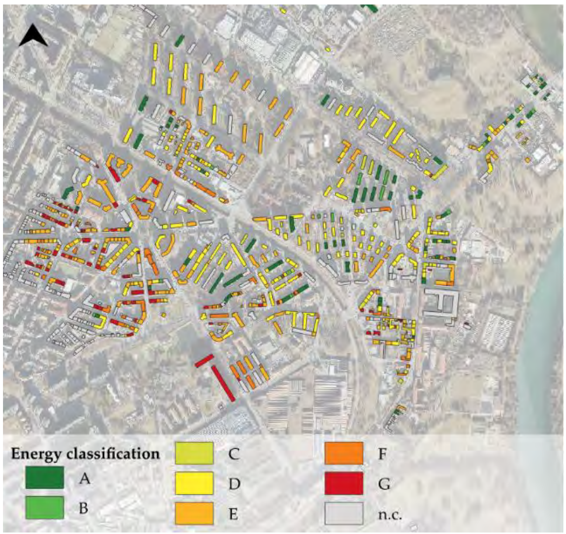

Fig. 4 – Energy classification based on EPCs after renovation

Moving to a comparison between the different filtering categories, it is strongly related to the observations on the distribution of the different classes. The highest performances can be observed by segmenting the building stock according to renovation EPCs – shown in Fig. 4 –, while the opposite results from the residential EPCs – plotted in Fig. 5. Nevertheless, the latter is comparable with the full dataset, with no filtering applied: most buildings are residential, so this segment significantly influences the trends of the entire district. Also EPCs deriving from renting or selling and the dataset considering whole buildings show similar values, despite the former having a slight tendency to higher classes.

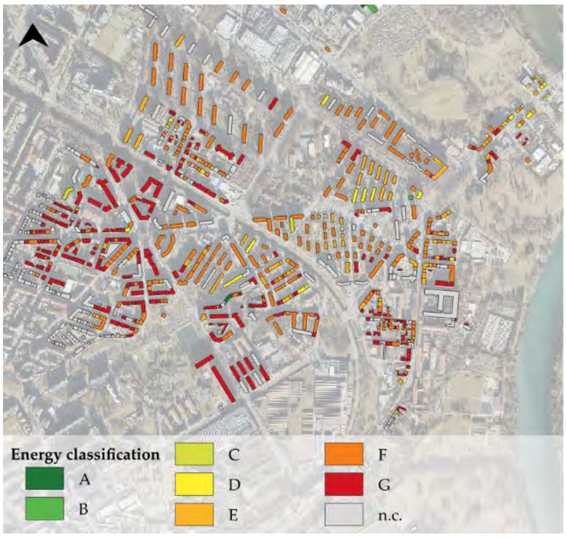

Fig. 5 – Energy classification based on EPCs of residential buildings

When considering the EPCs covering the whole building, it can be observed that in two cases – out of the three training/validation combinations – a small homogenous area located between Cravero and Ancina Streets and Taranto Avenue has buildings falling in the same class, D; in the remaining case, three buildings are classified as A. This is particularly relevant for the draft of energy policies: it can be foreseen that the whole portion is to be renovated simultaneously, taking advantage of the homogeneity of the district – which simplifies the preliminary phases.

3.3.3 Validation

The different classifications were validated against subsets of the EPC database – as described previously – in order to check the class differences, both positive and negative.

There is a general tendency towards overestimating the energy class, especially when testing the results of the estimation based on residential EPCs. However, there are no areas in which a high number of buildings are misclassified. As for the characteristics of the misclassified buildings, there is not a direct correlation between errors and the period of construction, while most differences are observed for buildings with an SV ratio lower than 0.4 – 70% in the 20/80 combination of residential EPCs classification.

For plotting the ROC – Fig. 6, which compares true and false positives, all classifications within a ±1 class difference were considered as correct (true positives). Results are satisfactory, with most classifications proving to be above the chance level – indicating the results given by a random classification (with True Positives probability equal to 50%) –, thus making the validation results trustworthy. The only classification which proved to be wrong in every configuration is the one carried out according to the EPCs issued after renting or selling, confirming the assumptions stated in the introduction about a lower reliability deriving from problems in gathering the necessary data. Further problems can be observed when considering the 20%/80% configuration of the EPCs issued for residential buildings and the 80%/20% configuration of the renovation dataset; the renovation dataset in the 20/80% configuration corresponds to the chance level. Problems in classifying according to EPCs issued after renovation show a lack of correlation observed in Section 3.2.

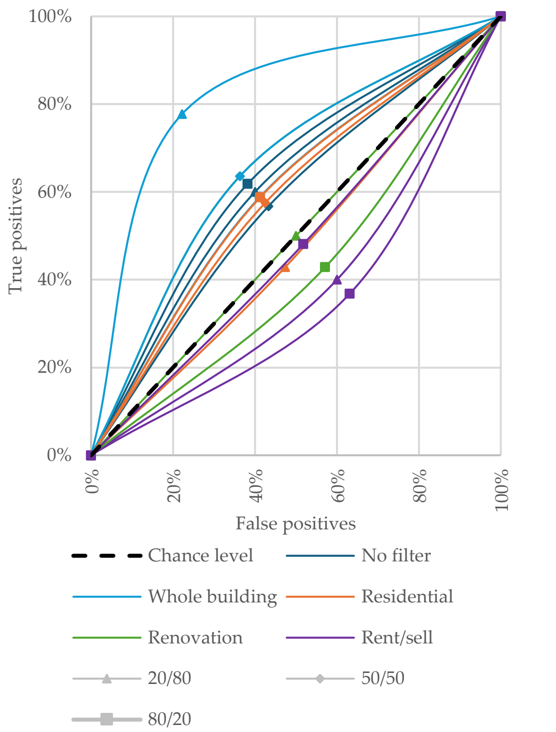

Fig. 6 – Receiver Operating Curves

Synthetic observations can be observed by looking at the 40 and 60 percentile and the median value of all the curves. For true positives, these correspond – respectively – to 53%, 59% and 57%, making the general results satisfactory. The best results are observable for classifications based on the EPCs issued for whole buildings – which do not require any data aggregation –, further confirming the initial assumptions. However, this result is downsized considering the lower number of volumetric units used for validation, 12 on average in the three combinations.

4. Conclusion

This study refined a methodology presented in a previous publication (Anselmo et al., 2023) which combines aerial thermography and data mining from Energy Performance Certificates for energy performance classification. EPCs were used to both train and validate the model, with different filters and combinations, allowing the identification of worst performing buildings.

The results proved the efficiency of the method and the incremental nature of the resulting energy performance evaluation accuracy, with an accuracy which reached 80% in some cases. Once having gathered the necessary inputs – mainly EPCs and temperature values – the methodology can be fully automated, not requiring any human resource to perform the classification: this could be appealing for Public Administrations in their work to make the building stock compliant with European Directives, in particular the EPBD (Council of the European Union, 2023b). The automation could consider also the implementation of an Artificial Intelligence algorithm, automatically correlating input data to an energy class: this would be particularly relevant when extending the AoI and therefore increasing the amount of training data.

From the scientific point of view, this work paves the way for the realisation of highly accurate UBEMs, simulating building performance on a district or urban scale by integrating remotely sensed pictures and EPC data for detailed archetyping. A detailed refinement of such methodology on building scale could eventually lead to the possibility of issuing EPCs without the need for extensive on-site surveys, estimating the U-value and the necessary geometric parameters accurately.

Nevertheless, some weaknesses emerged. First, the limited extension of the AoI, constraining the heterogeneity of the training samples – unfavoured by the homogeneity of the district too. Second, it would be relevant to perform the classification on limited segments of the building stock, in order to have a specific characterisation based on the building age and type. Finally, further information could be added such as the case of the window-to-wall ratio, a crucial indicator when assessing thermal losses; this could be derived from both the segmentation of a 3D model and the share between sunlit and useful surfaces in the EPCs.

Acknowledgement

This paper reports part of the work developed within the project NODES, which has received funding from the MUR - M4C2 1.5 of PNRR with grant agreement no. ECS00000036. Maria Ferrara's activity was funded by Italian MUR within the PON "Ricerca e Innovazione" 2014–2020, Asse IV "Istruzione e ricerca per il recupero"—Azione IV.4—"Dottorati e contratti di ricerca su tematiche dell'innovazione" and Azione IV.6—"Contratti di ricerca su tematiche Green. Further, this work was developed in accordance with the framework agreement between the City of Turin and the Polytechnic of Turin, signed on 9th February 2023, for the realisation of pilot projects towards the implementation of a Digital Twin.

Nomenclature

Symbols

AoI	Area of Interest
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
EU	European Union
GIS	Geographic Information System
ROC	Receiver Operating Curve
SV	Surface-to-Volume
UBEM	Urban Building Energy Model

References

Anselmo, S., M. Ferrara, S.P. Corgnati, and P. Boccardo. 2023. "Aerial Urban Observation to Enhance Energy Assessment and Planning towards Climate-Neutrality: A Pilot Application to the City of Turin." *Sustainable Cities and Society* 99. https://doi.org/10.1016/j.scs.2023.104938

Conticelli, Elisa, Stefania Falcioni, Giulia Marzani, Gian Luca Morini, and Simona Tondelli. 2024. "Assessing Energy Efficiency at Urban Scale through the Use of Energy Performance Certificates: An Application in the Emilia-Romagna Region, Italy." *Cities* 145: 104728. https://doi.org/10.1016/j.cities.2023.104728

Council of the European Union. 2023a. "Fit for 55." 2023. https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/

Council of the European Union. 2023b. "Proposal for a Directive of the European Parliament and of the Council on the Energy Performance of Buildings (Recast) - Analysis of the Final Compromise Text with a View to Agreement." Brussels. https://data.consilium.europa.eu/doc/document/ST-16655-2023-INIT/en/pdf

Deng, Zhang, Yixing Chen, Jingjing Yang, and Francesco Causone. 2023. "AutoBPS: A Tool for Urban Building Energy Modeling to Support Energy Efficiency Improvement at City-Scale." *Energy and Buildings* 282: 112794. https://doi.org/10.1016/j.enbuild.2023.112794

European Commission. 2023. "EU Mission: Climate-Neutral and Smart Cities." 2023. https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/climate-neutral-and-smart-cities_en#documents

European Parliament and Council. 2023. "Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on Energy Efficiency and Amending Regulation (EU) 2023/955 (Recast)."

IEA. 2016. "Energy Technologies Perspectives 2016. https://www.iea.org/reports/energy-technology-perspectives-2016

Johari, F., G. Peronato, P. Sadeghian, X. Zhao, and J. Widén. 2020. "Urban Building Energy Modeling: State of the Art and Future Prospects." *Renewable and Sustainable Energy Reviews* 128: 109902. https://doi.org/10.1016/j.rser.2020.109902

Johari, Fatemeh, Farshid Shadram, and Joakim Widén. 2023. "Urban Building Energy Modeling from Geo-Referenced Energy Performance Certificate Data: Development, Calibration, and Validation." *Sustainable Cities and Society* 96: 104664. https://doi.org/10.1016/j.scs.2023.104664

Martin, Miguel, Adrian Chong, Filip Biljecki, and Clayton Miller. 2022. "Infrared Thermography in the Built Environment: A Multi-Scale Review." *Renewable and Sustainable Energy Reviews* 165: 112540. https://doi.org/10.1016/j.rser.2022.112540

Osservatorio ITC della Regione Piemonte. 2024. "Datipiemonte." 2024. https://www.dati.piemonte.it/#/home

Pagliaro, Francesca, Francesca Hugony, Fabio Zanghirella, Rossano Basili, Monica Misceo, Luca Colasuonno, and Vincenzo Del Fatto. 2021. "Assessing Building Energy Performance and Energy Policy Impact through the Combined Analysis of EPC Data – The Italian Case Study of SIAPE." *Energy Policy* 159: 112609. https://doi.org/10.1016/j.enpol.2021.112609

Yu, Hang, Meng Wang, Xiaoyu Lin, Haijin Guo, He Liu, Yingru Zhao, Hongxin Wang, Chaoen Li, and Rui Jing. 2021. "Prioritizing Urban Planning Factors on Community Energy Performance Based on GIS-Informed Building Energy Modeling." *Energy and Buildings* 249: 111191. https://doi.org/10.1016/j.enbuild.2021.111191