IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Data driven economic scenarios for retrofitting residential buildings in a northern Italian region

To cite this article: Fulvio Re Cecconi and Luca Rampini 2023 *IOP Conf. Ser.: Earth Environ. Sci.* **1196** 012113

IOP Conf. Series: Earth and Environmental Science 1196 (2023) 012113 doi:10.1088/1755-1315/1196/1/012113

Data driven economic scenarios for retrofitting residential buildings in a northern Italian region

Fulvio Re Cecconi[1], **Luca Rampini**[2]

Department ABC, Politecnico di Milano, via Ponzio 31, 20133 Milano, Italia

fulvio.rececconi@polimi.it

Abstract. European directives and strategies, such as the 'European Green Deal' and the 'Renovation Wave', point out the importance of the building sector in achieving the climate goals set by the European Union for 2050. However, a higher renovation rate for the existing buildings is required to achieve these goals. Many barriers prevent the renovation rate from growing. Regarding financial barriers, the long payback times of renovation interventions and the high risk perceived by the potential investors make the renovation rate remain low. Based on data from energy performance certificates, this research proposes a data-driven method to create economic retrofit scenarios for residential buildings using Artificial Intelligence techniques and Monte Carlo simulations. Namely, energy savings have been predicted using an Artificial Neural Network on clusters of residential buildings and the Life Cycle Costs forecasted by Monte Carlo simulations taking into account the uncertainty in many of the inputs. Results obtained by applying the method to a region in northern Italy illustrate two scenarios for the energy retrofit of the built environment, one assuming a payback time of fifteen years and the other of twenty-five years. In both cases, the maximum allowable investment, which varies according to the specific characteristics of the buildings, is much lower than the retrofit costs recorded in the same area in recent years.

1. Introduction

The effects of climate change on our planet have created a worldwide consensus on the need for sustainable development. In this context, the European Union (EU) has shown great interest in pointing to drastic pollution cuts. This includes formulating and achieving new and consolidated strategies and directives for net-zero emissions by 2050 [1]. Changes in energy use in the residential sector represent a significant segment of the ongoing low-carbon energy transition process in a multi-dimensional and multi-level process comprising multiple actors [2]. Researchers proved that upgrading building fixtures, equipment, and envelope components could achieve savings beyond 45% in energy and water consumption [2]. Noteworthy, households or the residential sector represents about 26% of final energy consumption or nearly 17% of gross inland energy consumption in the EU [3]

Achieving a low energy standard while being cost-efficient in the existing building is challenging as it is crucial to assess the whole lifecycle in terms of costs and environmental impact [4]. Moreover, the lack of open data about the existing buildings makes it difficult to analyze the building's actual state, measure progress in building decarbonization, and raise public awareness about the importance of refurbishment, design renovation, and maintenance strategies [5].

There are many obstacles hindering the design of the energy retrofit of an existing building. Among them is the need for a deeper and more accurate energy audit process to precisely know the "as-built" situation in terms of envelope and equipment features and building user behavior [6]. Another critical issue preventing a higher building stock renovation rate is the economic one. Barriers such as high investments, long payback periods, and perceived credit risk hamper buildings' energy renovation [6].

Making informed choices about the most suitable energy retrofit policy requires detailed scenarios that specify the buildings to be renovated within certain constraints, e.g., on the total annual budget, renovation rate, or energy savings [7]. Several studies explored approaches and pathways toward low energy consumption for housing stocks without including the energy efficiency costs in the analysis [8]–[11]. Few studies have been undertaken to conduct an economic assessment of energy retrofit on existing buildings. By examining the costs and benefits associated with the retrofit and using economic evaluation methods, such as Net Present Value (NPV), Internal Rate of Return (IRR), and discounted payback period (PB), an assessment of the cost-effectiveness of retrofit investment is performed [12].

Current research methods show the high potential of top-down approaches for retrofitting buildings. These approaches utilize large datasets maintained by public authorities, such as datasets on energy performance certificates that provide a lot of building variables for analysis [13]. Examples of such datasets are the GEAK (Gebäudeenergieausweis der Kantone) in Switzerland and the CENED (Certificazione Energetica degli Edifici) in Italy. These datasets are generated by records of energy certification reports of buildings. These reports are submitted by certified energy consulting firms [13]. The significant challenges for stakeholders using these datasets are reliability, completeness, accuracy, and data consistency [14].

Large-scale identification of the potential of energy efficiency measures would enable mapping of the building stock, revealing cases were economically driven retrofitting is viable [15]. Thus, the main aim of this study is to develop a generalized methodology to optimize regional-scale energy retrofit decisions for residential buildings using data-driven approaches. The methodology will provide information to decision-makers and legislators, facilitating the introduction of new policies supporting the retrofitting market. This paper contributes to the literature on data-driven building energy modeling by introducing the combined use of machine learning (ML) and Monte Carlo (MC) simulations to forecast costs associated with a building energy retrofit. This paper is structured as follows: Section 2 describes a novel methodology; Section 3 evaluates the proposed methodology using an Italian case study; Section 4 concludes this research study and discusses the results.

2. Building retrofit scenarios

The following **Figure** 1 shows the pipeline of the research project. There are three main steps and two main inputs. The latter are: a) an open data collection of information about the energy performances of buildings, described in the next paragraph, and b) some assumptions on buildings' retrofit costs illustrated in paragraph 2.3.

The first main step of the research is to obtain reliable data for input in the data-driven method. Detecting and repairing dirty data is one of the perennial challenges in data analytics, and failure to do so can result in inaccurate analytics and unreliable decisions [16]. A by-product of this step is the clusterization of assets, allowing to group together of buildings built with similar components and systems.

The second step is constructing a predictive model to compute the energy demand of buildings after the retrofit. In recent years, much research has successfully dealt with the prediction of building consumption using artificial intelligence. Most of the works in the literature use artificial neural networks (ANNs), an artificial intelligence technique also adopted in the present research.

The last step of the research is the creation of economic retrofit scenarios. Many of the inputs required for this step are characterized by high uncertainty. This necessitated the use of statistical techniques to work with uncertainty. The most common of such techniques is Monte Carlo simulations [17]. The remaining parts of this paragraph give more details about the research's three steps.

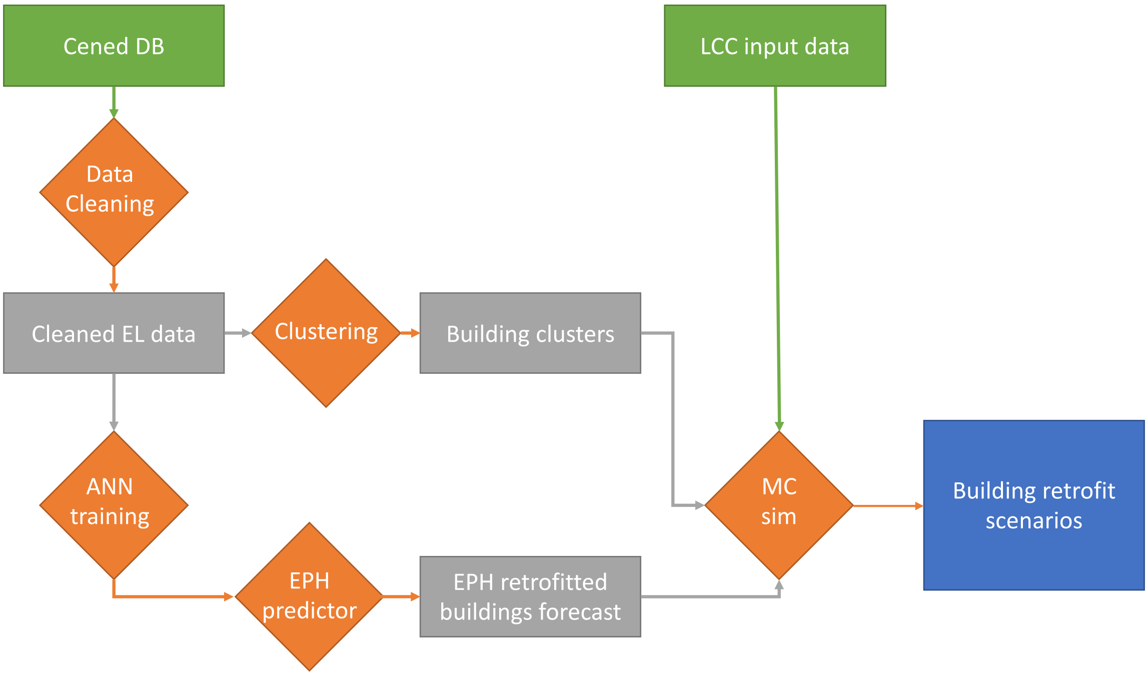

Figure 1. research schema

2.1. Cleaning energy performance certificate data

Some data on the energy certifications of buildings in the Regione Lombardia are available as open data on a portal set up by the Regione itself [18]. The open DB on energy labels includes data on the energy performance of buildings (both primary energy and net energy) and geometric information (e.g., volume, gross and net floor area, window area, etc.) [19]. It contains 1.52 million records, i.e., energy labels, of assets differing in intended use and type. Some residential, industrial, and commercial buildings focus on residential assets, flats, single-family buildings, villas, etc.

The different preparation of individuals who created the energy certificates and the absence of control over the input data resulted in several inaccuracies in the database. Therefore, data must be cleaned before being used. Data cleaning is a standard process when large volumes of data have to be used. In this research, the data cleaning process, described in detail by [11], reduced the database by almost 75%. After that, data from single-family buildings investigated in this research were extracted, resulting in slightly more than 161,000 labels.

The data from the CENED DB were exploited in the subsequent stages of the research using two different artificial intelligence technologies, as in **Figure** 1. The first of the two techniques used is unsupervised learning, in particular clustering, which is the process of grouping similar objects into different groups, or more precisely, the partitioning of a data set into subsets, so that the data in each subset according to some defined distance measure [20]. Using only some of the asset properties recorded in the CENED DB, those related to technological performance and primary energy demand, it was possible to identify eight clusters of similar assets. These are groups of assets characterized by similar technological performance of building envelope components and similar heating energy demand. As a result, it can be reasonably assumed that the building technologies used are similar. The next energy retrofit scenarios will be based on this subdivision into clusters of similar buildings.

2.2. Predicting post-retrofit energy performances

Fundamentally, building energy prediction belongs to the time series forecasting or regression problem, and data-driven methods have drawn more attention recently due to their powerful ability to model complex relationships without expert knowledge. Among those methods, ANNs have proven to be one of the most suitable and potential approaches [21], [22]. ANNs are a subset of Machine Learning (ML) techniques inspired by the biological neural network, which is advantageous in the strong ability to represent and model the nonlinear relationships between inputs and outputs.

Table 1. The statistical description of the numerical features of the dataset used to train the ANN

CENED DB		Gross Volume [m³]	Dispersing surface [m²]	Glass over walls surface ratio	Walls average transmittance [W/m² K]	Roofs average transmittance [W/m² K]	Windows average transmittance [W/m² K]	EPH [kWh / m²y]
Mean		708.71	413.86	0.073	0.873	0.840	3.043	168.24
Standard Deviation		1'090.00	485.00	0.034	0.455	0.508	0.989	78.36
Min		58.60	21.38	0.001	0.010	0.001	0.629	0.02
Quantile	**5%**	174.90	106.04	0.035	0.244	0.205	1.385	40.69
	25%	292.68	187.16	0.052	0.503	0.389	2.302	106.01
	50%	424.00	280.51	0.066	0.810	0.700	3.230	169.46
	75%	642.91	438.66	0.085	1.197	1.300	3.382	235.37
	95%	2'284.02	1'213.87	0.131	1.634	1.700	4.920	287.35
Max		9'999.00	11'330.62	0.540	2.604	6.897	6.478	300.00

The first step in the implementation of an ANN model is the selection of meaningful features. These features must be a logical input set for the model; therefore, knowledge domain is a fundamental skill in this phase. In the CENED DB there are 44 other features, including the proposed ANN output, EPH. Many of them are not necessary to reliably predict the primary energy demand of the building. Describing the feature selection process is not in the scope of this article; here are the ones chosen: a) city name; b) year of construction; c) gross heated volume; d) dispersing surface; e) glass over walls surface ratio; f) walls average transmittance; g) roofs average transmittance; h) windows average transmittance. The first and the second are categorical variables, while the others are numerical variables whose statistical description is given in Table 1.

The definition of the ANN architecture is the second step, and it consists of an iterative process in which an attempt is made to optimize the depth and density of the network layers while monitoring performance. Moreover, the model's performance depends on a set of hyper-parameters (optimizer, activation function, batch size) that were tuned to reach the minimum error level. Overall, the model used in this research consists of 6 layers and is described in Table 2.

Table 2. ANN architecture.

Layer (type)	Output shape	Parameters
Normalization	(None, 8)	17
Dense	(None, 256)	2304
Dense	(None, 128)	32896
Dense	(None, 128)	16512
Dense	(None, 64)	8256
Dense	(None, 1)	65

2.3. *Simulating economic retrofit scenarios*

The ANN described in the previous paragraph provides the potential savings in terms of lower EPH that can be obtained from an energy retrofit intervention. This is the first but not the only data needed to create economic scenarios. These were created using the Life Cycle Costs (LCC) method as described in ASTM [23], [24], ISO [25], and EN [26] standards. The ASTM standards, in particular, provide the formulae for calculating the LCC and other economic Key Performance Indicators (KPIs), including the one used in this research: the payback period (PB), i.e., the period required for the annual savings produced by the retrofit to equal the initial costs.

LCC analysis is an economical method for evaluating a project or alternatives over a designated study period. A considerable amount of research uses LCC methods to assess the economic impacts of several energy efficiency measures for building design and renovation [6]. The method entails computing the LCC for alternative building designs or system specifications having the same purpose and then comparing them to determine which has the lowest LCC over the study period [23]. The LCC of an asset may be computed according to the following equation (1):

$$pvLCC = pvIC + pvM + pvR + pvF - pvS \tag{1}$$

Where:

$pvLCC$ is the present value of the Life Cycle Costs LCC
$pvIC$ is the present value of the Initial Cost IC
pvM is the present value of maintenance and repairs (M) costs
pvR is the present value of the replacement (R) costs
pvF is the present value of the fuel (F) costs
pvS is the present value of the resale value (S)

In Equation (1) all costs are discounted to the base time, i.e., their present value is used to compare similar objects. Two project alternatives, such as the pre and post-retrofit situation of a building, may also be compared through the payback period (PB), i.e., the time required for the cumulative benefits from an investment to pay back the investment cost and other accrued costs considering the time value of money [27]. The PB may be computed by solving equation (2):

$$\sum_{t=1}^{PB} \frac{(B_t - \tilde{C}_t)}{(1+i)^t} = C_0 \tag{2}$$

Where:

$(B_t - \tilde{C}_t)$ is the net cash flow in year t computed as the difference between the dollar value of benefits in year t B_t minus the dollar value of costs in year t $\tilde{C}_t$.
i is the discount rate per time period
C_0 are the initial project investment costs

Some simplifying assumptions can be adopted in calculating LCC and PB, given the goal of creating economic scenarios. The first simplification is to assume that the annual maintenance and repair costs pvM of the building in the post-retrofit condition are not very dissimilar to the pre-retrofit condition. This can be justified by considering that the higher costs of new systems maintenance are compensated by the decrease in maintenance costs of the building parts and the repairs of wear and tear damage of the old systems. A second simplification is obtained by considering the replacement costs pvR equal in the two configurations, pre-, and post-retrofit. Finally, the residual value pvS is always considered zero.

The above assumptions allow the annual net cash flow to be computed as the difference between the pre-retrofit fuel cost minus the post-retrofit fuel cost. The fuel cost can be calculated as the product of EPH and the cost of fuel per kWh. Calculating the annual net cash flow and assumed values for the necessary economic and financial parameters needed (discount rate and inflation rate), it is possible to derive from equation (2) the maximum value of the initial investment (MI) such that the PB is as long as desired. If the actual retrofit cost for an asset is more than the computed MI the savings will be lower than the costs in the desired PB; otherwise, the benefit will pay back the retrofit cost during the PB. For each asset in the CENED DB, the maximum investments resulting in a PB of 15 or 25 years were computed to create the reference scenarios.

All inputs required to calculate MI are characterized by uncertainty, and many examples of representations of such inputs with probability distributions instead of deterministic values can be found in the literature. Monte Carlo (MC) simulations [28] can calculate MI by accounting for the effects of uncertainty in the input data. MC simulations may compute a probability distribution of MI; thus, a reliability value may be given for a predicted MI value.

The economic retrofit scenarios may be defined by choosing a type of energy retrofit intervention and a PB. In this article, a retrofit intervention was chosen to bring all buildings to the best level in the CENED DB, i.e., energy class A, and MI was calculated for a PB of 15 years and one of 25 years. The following section details the results obtained.

3. Results

This section presents the results obtained in the second and third steps that form the research method, as detailed in the previous section. The step of training and testing the ANN for post-retrofit EPH calculation is, at least in this context, insignificant since the results are the input for the next step and do not allow to derive useful information for the retrofit economic scenarios directly; here, the step is merely reported in very synthetic terms.

Table 3. Clusters' description

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
U_{walls} Mean	1.296	0.943	0.683	0.343	1.148	1.211	1.205	1.186
U_{walls} Std Deviation	0.434	0.341	0.280	0.152	0.333	0.432	0.377	0.379
U_{roofs} Mean	1.065	0.970	0.654	0.307	1.162	1.071	1.130	1.226
U_{roofs} Std Deviation	0.519	0.433	0.330	0.177	0.450	0.520	0.473	0.490
$U_{windows}$ Mean	3.529	3.335	2.831	1.708	3.738	3.495	3.665	3.715
$U_{windows}$ Std Deviation	0.724	0.653	0.599	0.498	0.781	0.830	0.739	1.001
EPH pre-retrofit Mean	214.961	194.420	143.757	62.828	230.946	210.282	223.055	170.851
EPH pre-retrofit Std Deviation	56.813	56.766	52.160	32.479	48.765	65.212	55.475	60.230
EPH post-retrofit Mean	41.254	35.689	35.212	35.670	36.592	39.902	38.438	25.248
EPH post-retrofit Std Deviation	5.519	4.454	4.562	6.544	5.063	5.458	5.564	4.222

The economic scenario creation is based on the convenience of a similar building cluster partitioning of the entire CENED DB. Each cluster, obtained with an unsupervised ML algorithm, represents a group of buildings with similar technological characteristics and performances. Table 3 shows, for each cluster, the average values and standard deviation of: a) average transmittance of the opaque envelope (W/m^2K); b) average transmittance of roofs (W/m^2K); c) average transmittance of windows and doors (W/m^2K); d) pre-retrofit EPH (kWh/m^2y); e) post-retrofit EPH (kWh/m^2y) computed using the trained ANN. In Table 3, the considerable reduction in primary energy demand achieved by the retrofit can be seen.

The second step is a practical application of LCC methodologies and relies on data quality and long-term forecasts. Data uncertainty is a well-recognized matter associated with LCC methods [29]–[31] and affects both the data coming from the CENED DB and the saving forecasts made by the ANN in the previous step. Focusing on EPH forecasts, if the error terms (difference between observed and predicted) are normally distributed, the standard deviation of ANN prediction is related to the MAE measured in the training phase as in Equation (3).

$$\sigma = MAE \times \sqrt{\frac{2}{\pi}} \tag{3}$$

The intrinsic uncertainty in the other economic parameters was modeled with probability distributions chosen according to the criteria set out in [6]. Fuel cost has been modeled by a Weibull distribution described with two parameters assumed as follows: scale = 0.12; shape = 1.5. The discount rate has been described using a triangular distribution as follows: min = 1%; mode = 4%; max = 7%. The inflation has been modeled by a Gaussian distribution with the following parameters: mean = 3%; standard deviation 1%.

With the assumptions illustrated above regarding the input variables and using Equation (2) as described in section 2.3, an MC simulation was carried out for each asset in the eight clusters providing a statistical distribution of maximum investment values for the retrofit intervention in two payback period cases, 15 and 25 years. This MI value represents the maximum expenditure threshold for the energy retrofit intervention on a building to break even after 15 or 25 years. The statistical distributions of MI allow for the construction of economic scenarios for retrofit interventions. The number of assets for which the intervention is profitable—that is, whose payback period is at most 15 or 25 years—can be calculated for various thresholds of the cost of the energy retrofit intervention by selecting from the statistical distribution of MI a value with a probability of error equal to a specific threshold (in the research, this threshold is 15%, i.e., a value of MI such that the reliability is 85%).

The result of this analysis that led to the definition of economic scenarios is illustrated in Figure 2 for a payback period of 15 years and in Figure 3 for a payback period of 25 years. Both figures are divided into two parts. In part a) the X-axis shows the value of MI, and the Y-axis the number of assets for which energy retrofit is worthwhile for each value of MI. Eight curves are represented in the graph, one for each building cluster. Part b) shows for each cluster (X-axis) the number of assets that it is convenient to retrofit set ten MI thresholds. Thresholds are identified with a color that changes from the blue for the lowest threshold to red for the threshold with the highest MI value through green and yellow for intermediate thresholds. The threshold values are: A – MI < 21.9 Euro/m^2; B – MI < 43.8 Euro/m^2; C – MI < 65.7 Euro/m^2; D – MI < 87.6 Euro/m^2; E – MI < 109.5 Euro/m^2; F – MI < 131.4 Euro/m^2; G – MI < 153.3 Euro/m^2; H – MI < 175.2 Euro/m^2; I – MI < 197.1 Euro/m^2; L – MI <= 219 Euro/m^2. It is noticeable that for an MI greater than 219 Euro/m^{2,} no energy retrofit intervention on a building in the Regione Lombardia building stock has a payback period of fewer than 25 years.

The curves illustrating the trend in the number of assets for which energy retrofit is cost-effective as a function of MI, part a) of Figure 2 and Figure 3, show that for some clusters, e.g., Cluster 2 and 3, the number of assets that can be retrofitted varies rapidly as a function of MI, i.e., the slope of the curves is stepping in the first part. These clusters contain buildings with high energy performance, i.e., with a low EPH, so with high MI values, there is no economic viability of retrofitting. For example, for Cluster 3, almost 90% of the assets have an MI below 60 Euro/m^2 when considering a PB of 15 years. This means that an energy retrofit is not cost-effective if it costs more than 60 Euro/m^2 for 90% of the assets in the cluster, i.e., more than 33'600 assets in the whole region. Conversely, part b) of Figure 2 and Figure 3 shows that for some clusters, e.g., clusters 1, 5, and 6, even with high MIs, energy retrofit remains cost-effective for certain buildings. In cluster 1, for example, there are 1549 buildings for which an energy retrofit is cost-effective up to an expenditure threshold of G, i.e., 105 Euro/m^2. In part a) of the two figures representing economic retrofit scenarios with PB equal to 15 and 25 years, there is a grey area starting from an MI equal to 120 Euro/m^2. This threshold value is significant for the research as it corresponds to 10% of the maximum eligible expenditure to benefit from the economic incentives granted by the Italian Government for retrofit interventions. From the figures, it can be seen that there is no economic advantage for any asset, regardless of the PB value, if the retrofit cost is equal to the maximum eligible expenditure (1,200 Euro/m^2).

Eventually, Table 4 summarises the research results by providing the basic parameters of the computed economic scenarios. It shows, for each cluster and the two payback periods examined, the number of assets with a maximum energy retrofit cost of 120 Euro/m2, both as an absolute value and a percentage value. For example, it is cost-effective to retrofit 20'204 buildings, equal to 73.5% of cluster 5 if PB is 15 years, and 10'325 equal to 33.1% if PB is 25 years. Because the cumulative value of the benefits, the term on the left in Equation 2, is higher at the same final performance of the building, i.e. at the same EPH post retrofit, the number of buildings for which an energy retrofit costing 120 Euro/m2 is higher than for longer PBs, 25 years in the case of the table, means that the same retrofit can pay off higher costs, i.e. have a higher MI.

Table 4. Assets worth retrofitting according to an investment of 120 Euro/m^2 and a payback period of 15 and 25 years

	Cluster 1	Cluster 3	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
PBP 15	10841	33941	36986	19858	20204	4879	12667	5344
	75.6%	100.0%	100.0%	100.0%	64.7%	73.5%	100.0%	100.0%
PBP 25	6766	20452	34824	19858	10325	3012	12617	5332
	47.2%	60.3%	94.2%	100.0%	33.1%	45.3%	99.6%	99.8%

4. Discussion and conclusions

Data-driven methods are gaining more and more interest from scientists and practitioners in the construction industry as the digitization of the sector makes available a large amount of data that could not be obtained before. Among these, data on Energy Performance Certificates (EPC), which all European states are collecting, are certainly among the most interesting. The European Union is attempting to homogenize these data, ensuring that the suggested approach may be used in other parts of the EU. However, the method suffers greatly from data quality like any data-driven method; therefore, the data cleaning process must be strengthened to extend to other regions or countries.

The case study showed that the proposed data-driven method works but highlighted that, especially in the post-retrofit EPH prediction phase, the more buildings are used to train the neural network, the more accurate the data must be. Large datasets such as the one in the case study put the predictive capabilities of ANNs to the test if the input data are unreliable. Therefore, before extending the dataset for training the network to other Italian regions or states, it is necessary to homogenize and improve the data quality; otherwise, the mean absolute error (MAE) is likely to increase disproportionately.

In conclusion, this article describes an innovative data-driven method for generating economic reference scenarios for energy retrofits of the large building stock. This method is classified as Top-Down in the scientific literature because it uses large datasets created and maintained by public authorities. The method is based on machine learning techniques: unsupervised learning for clustering buildings and supervised learning, specifically an artificial neural network, to predict the primary energy demand of buildings after the retrofit. The economic scenarios are generated based on life cycle cost and consider the uncertainty associated with the inputs using Monte Carlo simulations to calculate the maximum cost threshold for a retrofit intervention to have the desired payback period. The method proved valid when applied to a large building stock, namely single-family residential buildings in the Lombardy Region.

A by-product of the case study is the proof that no energy retrofit intervention will ever have a break even if the investment costs are close to the cost threshold set by the Italian Government to benefit the economic incentives, i.e., when incentives will end the energy retrofit market in Italy will stop.

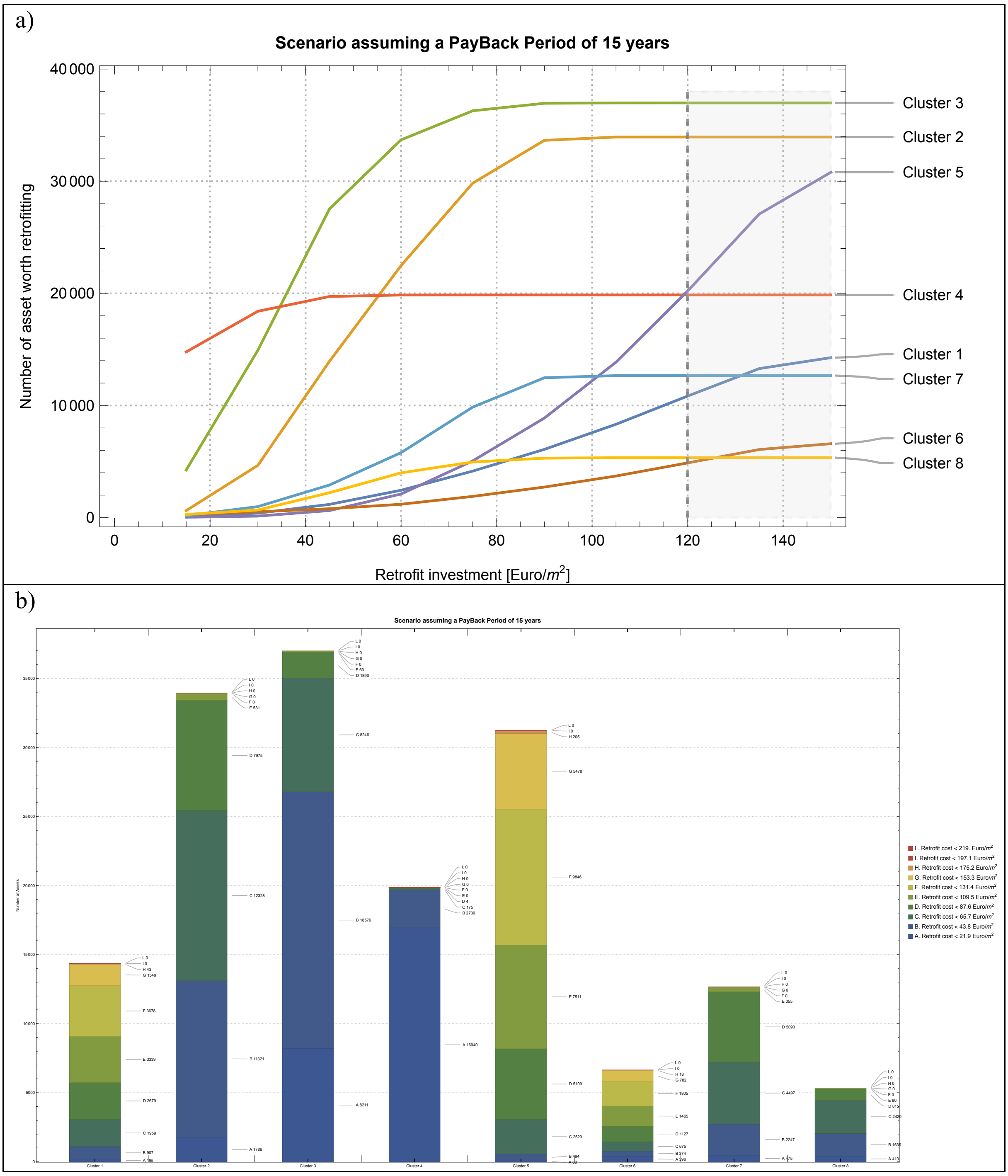

Figure 2 retrofit scenarios assuming a payback period of 15 years. Figure shows the number of assets worth retrofitting according to the maximum investment as a) a cumulative line or b) a bar highlighting 10 investment thresholds

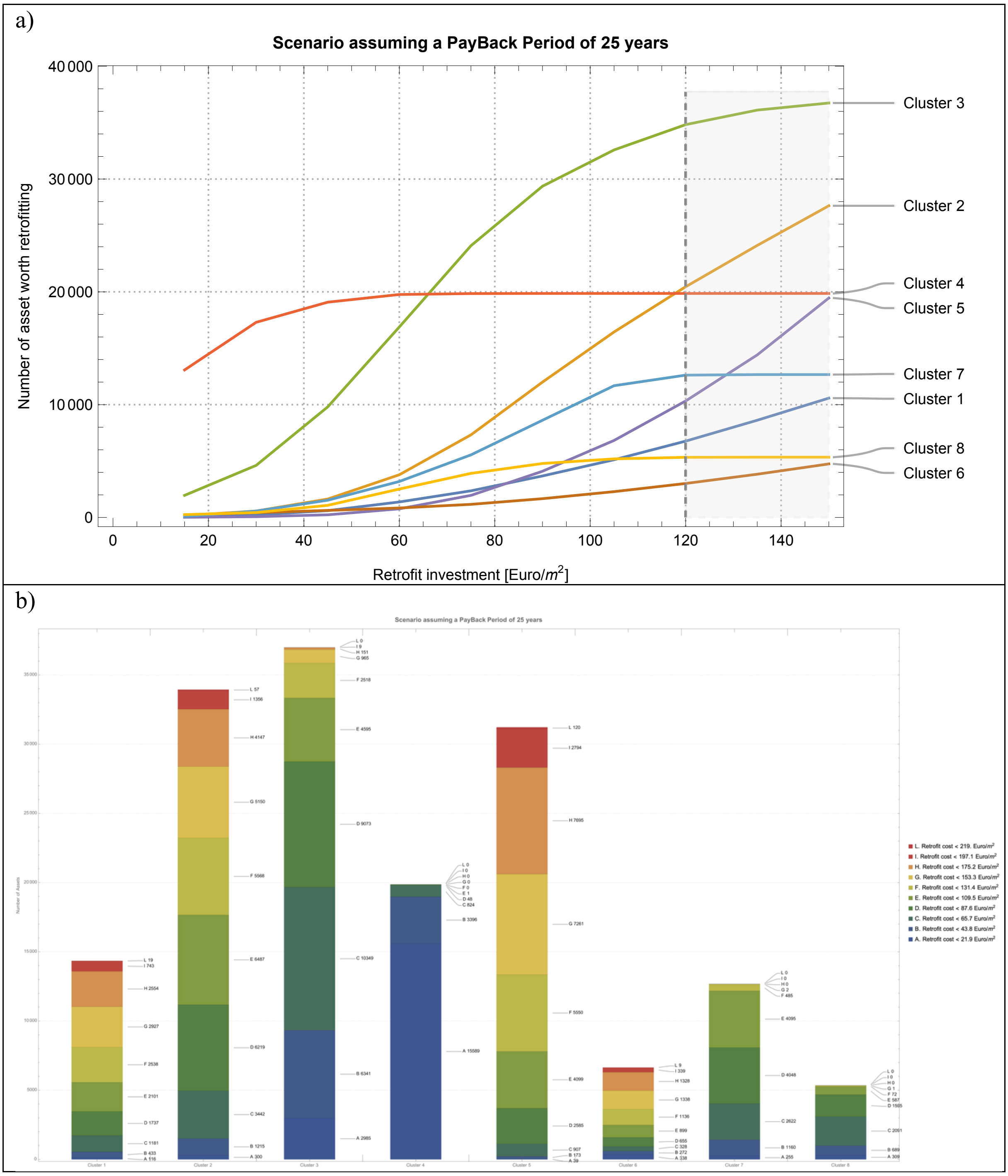

Figure 3. Retrofit scenarios assuming a payback period of 25 years. Figure shows the number of assets worth retrofitting according to the maximum investment as a) a cumulative line or b) a bar highlighting 10 investment thresholds

References

[1] European Commission, "A Clean Planet for all A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy," 2018. Accessed: Nov. 14, 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0773

[2] M. Ghafoori and M. Abdallah, "Innovative Optimization Model for Planning Upgrade and Maintenance Interventions for Buildings," *Journal of Performance of Constructed Facilities*, vol. 36, no. 6, Dec. 2022, doi: 10.1061/(asce)cf.1943-5509.0001757.

[3] B. Frantál and P. Dvořák, "Reducing energy poverty in deprived regions or supporting new developments in metropolitan suburbs? Regional differences in the use of subsidies for home energy efficiency renovations," *Energy Policy*, vol. 171, 2022, doi: 10.1016/j.enpol.2022.113250.

[4] A. Galimshina *et al.*, "Statistical method to identify robust building renovation choices for environmental and economic performance," *Build Environ*, vol. 183, Oct. 2020, doi: 10.1016/j.buildenv.2020.107143.

[5] M. Gómez-Gil, A. Espinosa-Fernández, and B. López-Mesa, "Review and Analysis of Models for a European Digital Building Logbook," *Energies (Basel)*, vol. 15, no. 6, 2022, doi: 10.3390/en15061994.

[6] E. di Giuseppe, M. Iannaccone, M. Telloni, M. D'Orazio, and C. di Perna, "Probabilistic life cycle costing of existing buildings retrofit interventions towards nZE target: Methodology and application example," *Energy Build*, vol. 144, pp. 416–432, Jun. 2017, doi: 10.1016/j.enbuild.2017.03.055.

[7] U. Filippi Oberegger, R. Pernetti, and R. Lollini, "Bottom-up building stock retrofit based on levelized cost of saved energy," *Energy Build*, vol. 210, Mar. 2020, doi: 10.1016/j.enbuild.2020.109757.

[8] A. Mastrucci, O. Baume, F. Stazi, and U. Leopold, "Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam," *Energy Build*, vol. 75, pp. 358–367, 2014, doi: https://doi.org/10.1016/j.enbuild.2014.02.032.

[9] E. G. Dascalaki, C. A. Balaras, S. Kontoyiannidis, and K. G. Droutsa, "Modeling energy refurbishment scenarios for the Hellenic residential building stock towards the 2020 & 2030 targets," *Energy Build*, vol. 132, pp. 74–90, 2016, doi: https://doi.org/10.1016/j.enbuild.2016.06.003.

[10] M. Aksoezen, M. Daniel, U. Hassler, and N. Kohler, "Building age as an indicator for energy consumption," *Energy Build*, vol. 87, pp. 74–86, 2015, doi: https://doi.org/10.1016/j.enbuild.2014.10.074.

[11] F. Re Cecconi, A. Khodabakhshian, and L. Rampini, "Data-driven decision support system for building stocks energy retrofit policy," *Journal of Building Engineering*, vol. 54, p. 104633, 2022, doi: https://doi.org/10.1016/j.jobe.2022.104633.

[12] Y. Liu, T. Liu, S. Ye, and Y. Liu, "Cost-benefit analysis for Energy Efficiency Retrofit of existing buildings: A case study in China," *J Clean Prod*, vol. 177, pp. 493–506, Mar. 2018, doi: 10.1016/j.jclepro.2017.12.225.

[13] C. Deb and A. Schlueter, "Review of data-driven energy modelling techniques for building retrofit," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 110990, 2021, doi: https://doi.org/10.1016/j.rser.2021.110990.

[14] U. Ali *et al.*, "A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings," *Appl Energy*, vol. 267, p. 114861, 2020, doi: https://doi.org/10.1016/j.apenergy.2020.114861.

[15] O. Pasichnyi, F. Levihn, H. Shahrokni, J. Wallin, and O. Kordas, "Data-driven strategic planning of building energy retrofitting: The case of Stockholm," *J Clean Prod*, vol. 233, pp. 546–560, 2019, doi: https://doi.org/10.1016/j.jclepro.2019.05.373.

[16] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data Cleaning: Overview and Emerging Challenges," in *Proceedings of the 2016 International Conference on Management of Data*, 2016, pp. 2201–2206. doi: 10.1145/2882903.2912574.
[17] C. Mooney, *Monte Carlo Simulation*. 2455 Teller Road, Thousand Oaks California 91320 United States of America : SAGE Publications, Inc., 1997. doi: 10.4135/9781412985116.
[18] Regione Lombardia, "Open Data Regione Lombardia." https://dati.lombardia.it (accessed Nov. 18, 2022).
[19] F. Re Cecconi, N. Moretti, and L. C. Tagliabue, "Application of artificial neutral network and geographic information system to evaluate retrofit potential in public school buildings," *Renewable and Sustainable Energy Reviews*, vol. 110, pp. 266–277, 2019, doi: https://doi.org/10.1016/j.rser.2019.04.073.
[20] T. S. Madhulatha, "An Overview on Clustering Methods," *CoRR*, vol. abs/1205.1117, 2012, [Online]. Available: http://arxiv.org/abs/1205.1117
[21] J. Seo, S. Kim, S. Lee, H. Jeong, T. Kim, and J. Kim, "Data-driven approach to predicting the energy performance of residential buildings using minimal input data," *Build Environ*, vol. 214, p. 108911, 2022, doi: https://doi.org/10.1016/j.buildenv.2022.108911.
[22] C. Lu, S. Li, and Z. Lu, "Building energy prediction using artificial neural networks: A literature survey," *Energy Build*, vol. 262, May 2022, doi: 10.1016/j.enbuild.2021.111718.
[23] ASTM International, "ASTM E917 Standard Practice for Measuring Lifecycle Costs of Buildings and Building Systems." 2017. doi: 10.1520/E0917-17E01.
[24] ASTM International, "ASTM E1121 Standard Practice for Measuring Payback for Investments in Buildings and Building Systems," 2020, doi: 10.1520/E1121-15R20E01.
[25] ISO International Organization for Standardization, "ISO 15686-5 Buildings and constructed assets - Service life planning - Part 5: Life-cycle costing." 2017.
[26] CEN European Committee for Standardisation, "EN 15459 Energy performance of buildings - Economic evaluation procedure for energy systems in buildings." 2017.
[27] ASTM International, "ASTM E833 Standard Terminology of Building Economics." 2021. doi: 10.1520/E0833-14R21.
[28] C. Mooney, *Monte Carlo Simulation*. 2455 Teller Road, Thousand Oaks California 91320 United States of America : SAGE Publications, Inc., 1997. doi: 10.4135/9781412985116.
[29] N. Wang, Y.-C. Chang, and A. A. El-Sheikh, "Monte Carlo simulation approach to life cycle cost management," *Structure and Infrastructure Engineering*, vol. 8, no. 8, pp. 739–746, 2012, doi: 10.1080/15732479.2010.481304.
[30] P. Gluch and H. Baumann, "The life cycle costing (LCC) approach: a conceptual discussion of its usefulness for environmental decision-making," *Build Environ*, vol. 39, no. 5, pp. 571–580, 2004, doi: https://doi.org/10.1016/j.buildenv.2003.10.008.
[31] C. C. Menassa, "Evaluating sustainable retrofits in existing buildings under uncertainty," *Energy Build*, vol. 43, no. 12, pp. 3576–3583, 2011, doi: https://doi.org/10.1016/j.enbuild.2011.09.030.