

Application of Artificial Neural Network and Geographic Information System to Evaluate Retrofit Potential in Public School Buildings

Re Cecconi, F.^a, Moretti, N.^{a,*}, Tagliabue, L.C.^b

^a Politecnico di Milano, Department of Architecture, Built Environment and Construction Engineering, Via G. Ponzio, 31, 20133, Milano, Italy

^b University of Brescia, Department of Civil, Environmental, Architectural Engineering and Mathematics, Via Branze, 43, 25123 Brescia, Italy

ABSTRACT

School buildings in Italy are outdated, in critical maintenance conditions and they often perform below acceptable service levels and quality standards. Nevertheless, data supporting renovation policies are missing or very expensive to be obtained. The paper presents a methods for evaluating building's energy savings potential, using the Building Energy Certification (Certificazione Energetica degli Edifici - CENED) open database. The aim of the research concerns the development of a data-driven set of methods, based on the use of open data, machine learning (ML) and Geographic Information Systems (GIS) to support regional energy retrofit policies on school buildings. The main advantage concerns the possibility to predict the post-retrofit energy savings, avoiding the expensive on-site Condition Assessment (CA) phase. Data have been first clustered to identify the most common thermo-physical properties of the envelope, then three retrofit scenarios have been defined to allow the retrofit of homogeneous types of buildings. The energy saving potentials have been evaluated through the implementation of eight Artificial Neural Networks. Ultimately, data have been geolocated and further processed to support the definition of the energy retrofit policies for the most critical regional areas. The Lombardy region has been chosen as case study to test the robustness of the proposed methods. The results of the case study proved that school buildings energy retrofit policies can be defined and compared using available open data, ML and GIS. The future developments of the research concern the further integration of GIS for retrofit cost assessment and scenario analysis.

HIGHLIGHTS

- Open-data, machine learning and spatial analyses support regional energy policies
- Eight Neural Networks are used to compute energy savings in three retrofit scenarios
- Data are geolocated and processed to guide the regional retrofit policy
- The retrofit policy is defined avoiding expensive on-site Condition Assessment

KEYWORDS

Energy retrofit, school buildings, open data, Artificial Neural Networks (ANN), Geographical Information System (GIS), data-driven process.

WORD COUNT: 5.634 (including the abstract of 246 words)

LIST OF ABBREVIATIONS

* Corresponding author

Phone: +39 02 2399 5174; E-mail address: nicola.moretti@polimi.it (N. Moretti)

AECO	Architecture, Engineering, Construction and Owner-operated
AM	Asset Management
ANN	Artificial Neural Network
API	Application Programming Interface
CENED	Database encompassing data about buildings energy labelling in Regione Lombardia
DB	Database
DM	A type of Italian Law (Decreto Ministeriale)
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificates
EPh	Primary energy factor for heating
ETh	Thermal energy factor for heating
EU	European Union
GIS	Geographic Information Systems
ICTs	Information and Communication Technologies
ML	Machine Learning
U-value	Thermal transmittance [W/K m ²]

1. INTRODUCTION

School buildings in the Lombardy region, northern Italy, are obsolete and in critical maintenance conditions, therefore their performances are below acceptable service levels and quality is a very serious issue [1]. Among unsatisfactory performances, those related to thermal comfort are extremely critical since they are strictly connected to the pupils' learning ability [2] affecting their cognitive performance [2] and to the (over)use of fossil fuels for energy supply [4]. Thus, under-performing buildings give rise to waste of public resources and air pollution, strongly contributing to climate change dynamics and increasing health problems for the citizens [5]. Moreover, high costs for data gathering and analysis often force public administrations to make strategic decisions on the refurbishment of the school building stock based on limited information and consolidated but not optimised procedures [6]. To make this process more efficient, additional time and resources are required: a further burden for public expenditure, already in a critical condition. As an example, approximately one third of maintenance costs are used inefficiently as a result of improper and unnecessary maintenance activities at national level.

The administration of the Lombardy region provides a public database of Energy Performance Certificates (EPC), in Italian the Certificazione ENergetica degli EDifici (CENED) [7], which encompasses data about building energy performances (i.e. both primary energy and net energy), geometric information (e.g. volume, gross and net surface, windows area, etc.) and adopted technologies (i.e. mainly the average thermal transmittance of building components and information about thermal plants' global efficiency).

These data have been used as driver for the definition of the most effective retrofit measures to be implemented on school buildings. Moreover, the proposed methodology, allows to make viable strategic decisions, not impacting on the costs needed to acquire the knowledge of the current energy and technical performance of buildings, since it does not employ a condition assessment of the physical assets, allowing consistent time and cost savings.

1.1 School buildings stock in Italy

According to the Governmental vision, the energy efficiency is a key driver for improvement of school buildings, in order to provide efficient and satisfactory educational facilities [8]. The school building stock amounts to 35% of the entire national building stock [9] and, most of the

cases, it requires deep refurbishment and maintenance interventions. Considering the whole national territory, 75% of school building dates before any Italian energy laws: 33% before Law 373/76 [10] and 25% before Law 10/91 [11], both concerning the definition of the national plan for energy use and saving. Moreover, glancing at the current period, this stock must be compliant with more recent standards [12] and European Union (EU) Directives [13] which are boosting the energy renovation and promoting higher efficiency levels with advanced technologies and cost-optimal criteria. The school building stock counts 45,000 public schools out of 62,000, which overtake public housing sector with an amount of energy consumption reaching approximatively 1 million Tonnes of Oil Equivalent (toe) per year (70% heating and 30% electricity). Therefore, remarkable improvement could be achieved through the optimisation of the energy consumptions, not only thanks to retrofitting interventions on buildings, but also promoting an energy behavioural awareness of the users: it had been estimated that a better energy behaviour, could lead to the reduction of the actual energy consumption by 20%, in school buildings [14]. On the other hand, improvement interventions on the envelope and thermal plants can heavily reduce energy consumption and associated running costs, though they generate additional investment costs [15]. Nevertheless, approximately 40% of school buildings in Italy needs refurbishment interventions, therefore the running cost mark-up gained thanks to energy improvements could compensate the overall costs for refurbishment interventions. These interventions could decrease the actual average energy consumption of public schools (calculated as around 180 kWh/m²year) towards those required for new constructions (30-40 kWh/m²year) introduced by national regulations since 2009 and promoted at European level [16].

1.2 Artificial Neural Networks to forecast energy demands

Building energy management is one of the most relevant topics for the definition of an effective AM strategy [17]. In the scientific literature, three main approaches for addressing the issue can be spotted: the white-box approach corresponding to the engineering approach; the black box approach, namely the machine learning approach and the grey box approach [18], which can be intended as an hybrid solution between the former two [17]. Despite being all widely exploited, the black box approach allows to reach the objective of prediction in a faster and highly precise way compared to the other two methodologies[20], especially when the dataset to be analysed is rather extended and complex [19]. The availability of information is a primary driver for the implementation of Artificial Neural Networks (ANN) [22], a subset of ML instruments. The use of ICTs in management of the built environment is rapidly gaining momentum [23], thanks to the increasing trend concerning the digitisation of a great amount of information related to the physical assets. This information represent a valuable resource for the development of data-driven tools for management of the built environment [24]. ANNs, whose behaviour is similar to the biological neural networks, allows to predict values with high accuracy against a low amount of input variables due to its ability to exploit a set of equations characterised by two functions: the “activation function” and the “transfer function”. These functions trigger a sequence of connected nodes, namely the neurons of the ANN [25]. Typically, the network is formed by an input layer of information to be processed, an output layer of values and a set of intermediate layers, called hidden, where the activation and transfer functions take place [26]. The number of hidden layers can vary as well as the number of artificial neurons in each layer, if the number of hidden layers is small then the network is called shallow, otherwise, if the number of hidden layers is high, the net is usually referred to as a deep neural network. The main advantage in the use of ANN concerns the possibility to forecast extremely reliable values using few input parameters without using time consuming engineering simulations. Several studies have been carried out for energy prediction exploiting ANN: from the first studies concerning the prediction of utility loads forecasting in 1990s, until more recent researches on

the energy saving potential of refurbished buildings [27]. Some studies demonstrate the possibility to predict energy consumptions, training the network on historical data series, demonstrating a good accuracy of the prediction carried out on a limited number of variables [28]. Backpropagation neural networks have been applied for the prediction of energy needs of buildings, according to the variation of the thickness of the insulation layer, orientation and transparency ratio [29]. ANNs have been used for the prediction of primary energy consumption in office buildings, the thermal comfort of the occupants and to develop retrofit scenarios [30]. ANNs have been employed for the prediction of the air-conditioning loads integrating multiple sensor data or related information as input for the network [31]. Moreover, ANNs have been employed for the development of an energy benchmarking model, including as variables physical properties, occupancy and climate [32] and for the identification of factors that mostly impact on the energy profile of buildings [33].

These are only some of the application of the neural networks to engineering field, which are limited to the energy prediction tasks. Nevertheless, many other applications can be found in literature, especially when the scale field of analysis concerns not only the buildings [24], but also the built environment and the cities [34].

1.3 Geographic Information Systems supporting strategic decisions

Geographic Information Systems (GIS) are software platforms capable to handle a great amount of data, coupling quantitative and qualitative information with geographic one [35]. These system have been exploited for the analysis of georeferenced data since 1960s when they have been employed by the Canadian Government for the implementation of the Canadian Land Inventory [36]. Moreover, being traditionally considered tools for strategic decisions, they are suitable for synthesis, representation and location-based analysis. Different applications of GIS can be found in literature [24]. Some of the primary advantages enabled by the use of these tools concerns the advanced data management functionalities that they allow and the possibility to effectively represent the outcomes of the data processing [37]. These capabilities can be exploited for energy modelling and planning, especially at the urban scale. Some application in this field concern topics as energy performance simulation [38], renewable energy modelling and planning [39], optimisation of urban energy systems and smart grid management [40].

In Italy, and in particular in the Lombardy region, urban planning decisions at the municipal and regional level must be taken considering the support of geographic information compliant with directive INSPIRE [41], therefore according to an homogeneous and harmonised framework for information management [18]. GIS database for the Lombardy region is rather developed and rich [43] and can be considered as a crucial resource for analysis and representation of phenomena at the regional level. Among data provided by Regione Lombardia, energy performance data of buildings within the region can be downloaded freely, despite they do not belong to datasets available through the GIS portal, therefore they are not compliant with specifications of directive INSPIRE [41].

More recently, the use of GIS coupled with prediction capabilities provided by the ANN is increasing, despite a knowledge gap can be identified for what concerns the application of GIS and ANN for supporting energy retrofit policies. Some applications of GIS and ANN for energy modelling concern the definition of a decision support tool to select the most suitable areas for the installation of large ground-mounted Photovoltaic Plants (PV) [44] and the development of a tools for site location of urban energy supply plants [45].

1.4 Aim of the research work

The research sits in the wider context of the digitisation of the Asset Management (AM) processes, specifically related to the sustainability and energy management: the proposed set of methods aim at leading strategic decisions on retrofitting interventions on school buildings,

harnessing innovative Information and Communication Technologies (ICTs) for management of the built environment and optimisation of digital based AM processes. Indeed, ML approaches and GIS tools are well documented in scientific literature, despite a knowledge gap has been identified concerning their joint use for supporting portfolio energy retrofit policies. Therefore, a Machine Learning (ML) approach has been coupled with the spatial analysis allowed by Geographic Information Systems (GIS). ANNs predictions, GIS representations and analyses for energy saving estimations at the regional level has seldom been adopted to develop an energy policy. The proposed methodology has been employed for supporting the development of a portfolio management strategy for school buildings energy retrofit in Lombardy Region in Italy.

2. METHODS

The implementation of the set of methods proposed through this research, starts from a specific dataset: the CENED DB provided by Regione Lombardia. This DB includes all the energy labels provided by accredited professional in the Lombardy Region. The process of the energy certification in Italy is often carried out without the necessary in-depth studies and verifications and there is not an extensive or automated assessment of reliability of input data about building certifications in the DB. Therefore, as a first step it is appropriate to implement a screening on the data collected by CENED in order to identify prominent discrepancies from reliable data and exclude the inconsistent values, before the analysis of the significant sample. After this first step of data cleaning, data have been processed and clustered to acquire a good knowledge of the building stock. After this process they have been used to train a set of multi-layer feed-forward ANN [26], that proved to be reliable instruments to forecast energy performance of school buildings [27]. Database cleaning is an operation with primary importance, because of the huge number of clearly wrong data in the DB: almost half of the energy labels in the DB has been discarded for evident inconsistency.

School buildings have been analysed and classified according to their age (in an overall time span of more than one century) and the performance of their envelope. This process gave as outcome the definition of homogeneous classes of comparable school buildings. For each class, specific retrofit strategies, suitable with their characteristics, have been defined and the potential energy savings have been computed through the trained ANN. The output data have been imported in a GIS environment, through which it has been possible to carry out a spatial analysis for the whole Lombardy Region territory. Figure 1 represents the main steps carried out in the research. The results of this analysis are presented, highlighting how a low-cost analysis can affect decisions on more than 1,500 school buildings.

Although there is a gap between actual and ANN-computed performance, the proposed process balances the reliability of energy savings forecasts, with the necessity of saving time and resources to carry out the estimation. A more precise energy demand estimation method would be certainly too expensive to be applied on a such extended school buildings stock. Moreover, the proposed method allows to easily spot the most convenient retrofit strategy for the whole portfolio, even in the very early stage of the decision-making process.

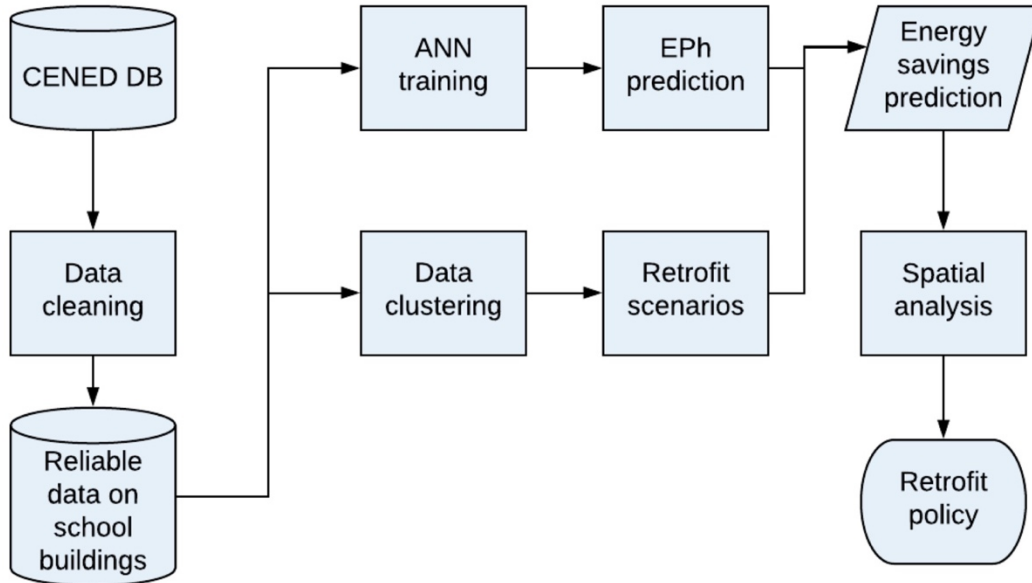


Figure 1: Framework of the research methodological steps.

2.1 Database cleaning

The first operation carried out on the DB concerns the cleaning, namely the removal of incomplete or clearly wrong records. It should be considered that in Italy energy assessors have very different backgrounds and work experiences (e.g. architects, civil and building engineers but also chemical engineers, naval engineers, etc.), which may lead to non-standardised in-depth analysis on the buildings. Therefore, not all the records in the CENED DB have a sufficient reliability. On the contrary, often the DB presents wrong or inconsistent data and, in the worst case, incomplete records which do not allow the application of the proposed set of methods for the computation of the energy savings. The DB cleaning phase allowed to avoid massive errors in the following steps of the research.

The CENED DB includes a very extensive set of information concerning energy performances of buildings in Lombardy Region. Therefore, the first operation that was accomplished concerns the filtering of records only related to school buildings. For each building 47 parameters are recorded in CENED DB. Some of them are needed to identify the building (address, cadastre data, type of buildings, ...), others to assess the dimensions of the building and its parts (volume, surface, façade and windows area, ...), to describe components and systems characteristics (thermal transmittance of building envelope components, type of heating system, photovoltaic energy production, ...) and to understand building overall energy performances (primary energy for heating, for cooling and for the production of domestic hot water, energy label, ...). Incomplete records (i.e. records in which some of the 47 values were missing), have been deleted, since they do not fully describe the energy behaviour of the buildings. After these two very first operations, the database has been further refined, according to the parameters listed in Table 1 [26].

Table 1. Constraints to spot unreliable data in the CENED database.

N°	CENED database parameter	U.M.	Threshold
1	Heated gross surface	m ²	< 250
2	Heated gross volume / heated gross surface	m	< 2.5
3	Building envelope surface	m ²	< 5
4	Walls or Roofs thermal transmittance	W/m ² K	< 0.05 or > 17

5	Windows thermal transmittance	W/m ² K	< 0.1
6	EPh/ETH	-	< 0.5 or >1.5

The first parameter in Table 1 has been selected to limit the research scope to entire buildings. This operation allowed to exclude schools located inside large residential buildings: a rather frequent situation in Italy. The threshold value set to 250 m² assures that small school inside other buildings are excluded. The second (Heated gross volume / heated gross surface) and the third (Building envelope surface) parameters are used to exclude records that are surely wrong because they describe a school with an average floor height lower than 2.5 m (in Italy the minimum floor height for a school is 2.7 m) or school with an envelope surface smaller than 5 m² (i.e. the sum of roofs surface, windows surface and external walls surface is lower the 5 m²). The thresholds set for thermal transmittance (U-values) of components of the envelop (fourth and fifth parameters), allow to exclude records where the characteristics of components exceed the minimum or maximum most likely values found on the Italian market. Eventually, the ratio Primary Energy factor for heating divided by Thermal Energy factor for heating (EPh/Eth) excludes buildings where the heating system has an efficiency too high or too low according to the Italian standards.

After the DB cleaning step, almost 50% of the whole database related to school buildings has been removed. Form the initial number of 2,915 rows representing the school buildings, a reduced DB of 1,632 rows has been obtained and further analysed.

The following step of the research concerns a set of analyses carried out to acquire a deep knowledge of the DB. Buildings in the CENED DB are organised by construction year, the representation of the energy labels by construction year shows a peak for classes 1961-1976 and 1977-1992, as showed in Figure 2. These two classes are characterised not only by the high number of buildings belonging to them, but also by low performances in terms of EPh consumption and U-values of walls, windows and roofing systems (Figure 3). Noteworthy, before 1976 there was no Law in Italy on energy performances of buildings. Therefore, these two classes are likely to be the ones of buildings which need the heaviest retrofitting interventions.

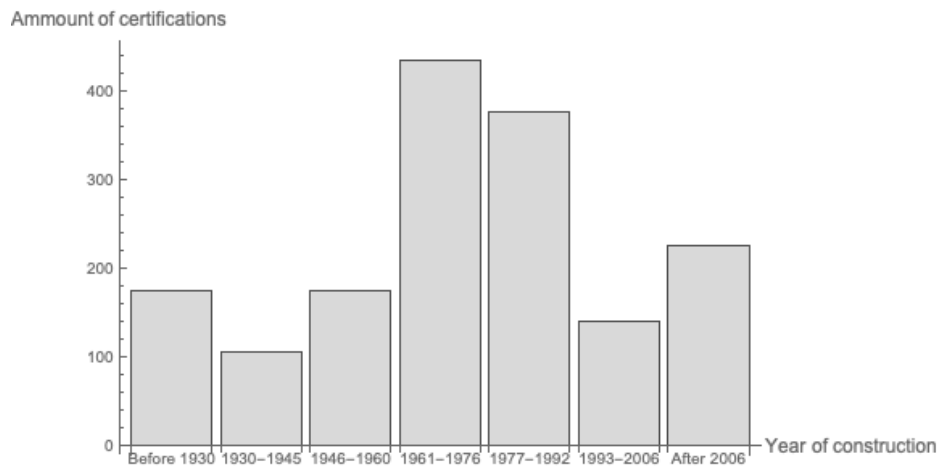


Figure 2. Labels groups according to the year of construction (7 groups).

Nevertheless, the most critical class concerns the schools built before 1930. This is rather easy to understand, since it is likely that those buildings, due to the typology of systems installed, the effects of time, the constant use of the facilities and the spaces and the inherent degradation of their components, show a low energy performance level.

a)

b)

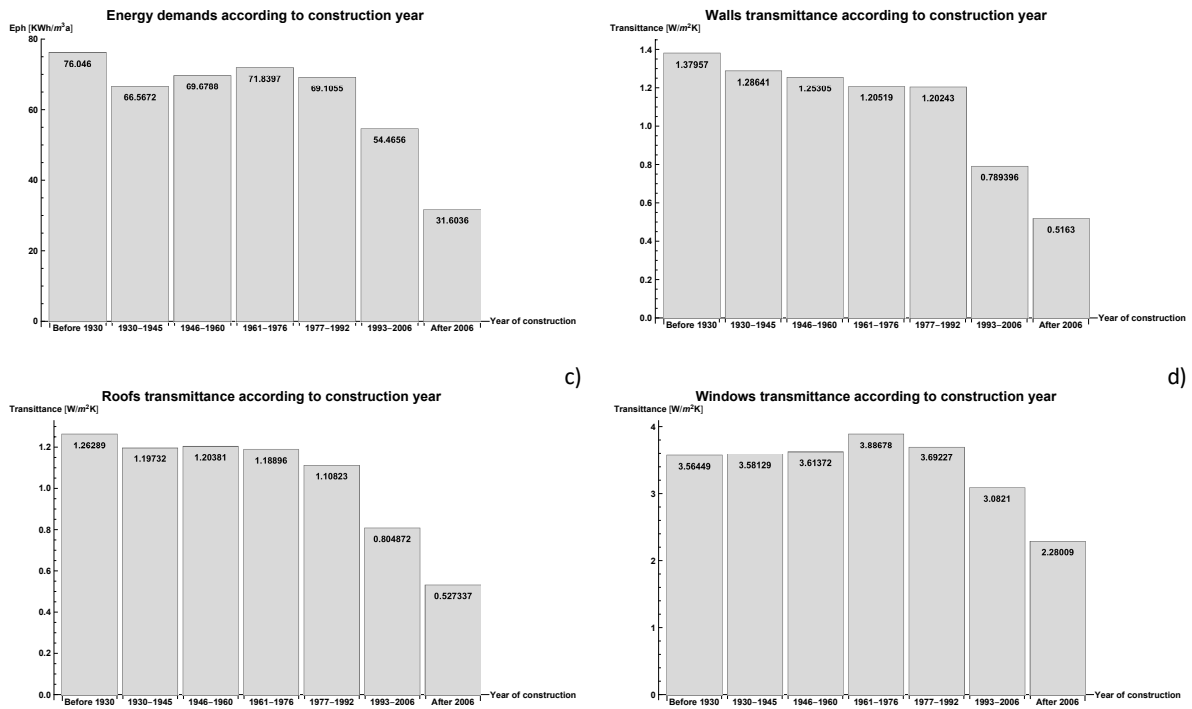


Figure 3. Average EPh (a) and thermal transmittance of building envelope components (b - walls, c - roofs, d - windows) according to the year of construction.

Finally, Figure 4 represents the heating demand of primary energy consumptions (EPh) related to the U-value of the three main technological units of the envelope: windows, roof and walls. The representation in Figure 4 is quite effective for the description of the overall trend of the whole filtered and cleaned database.

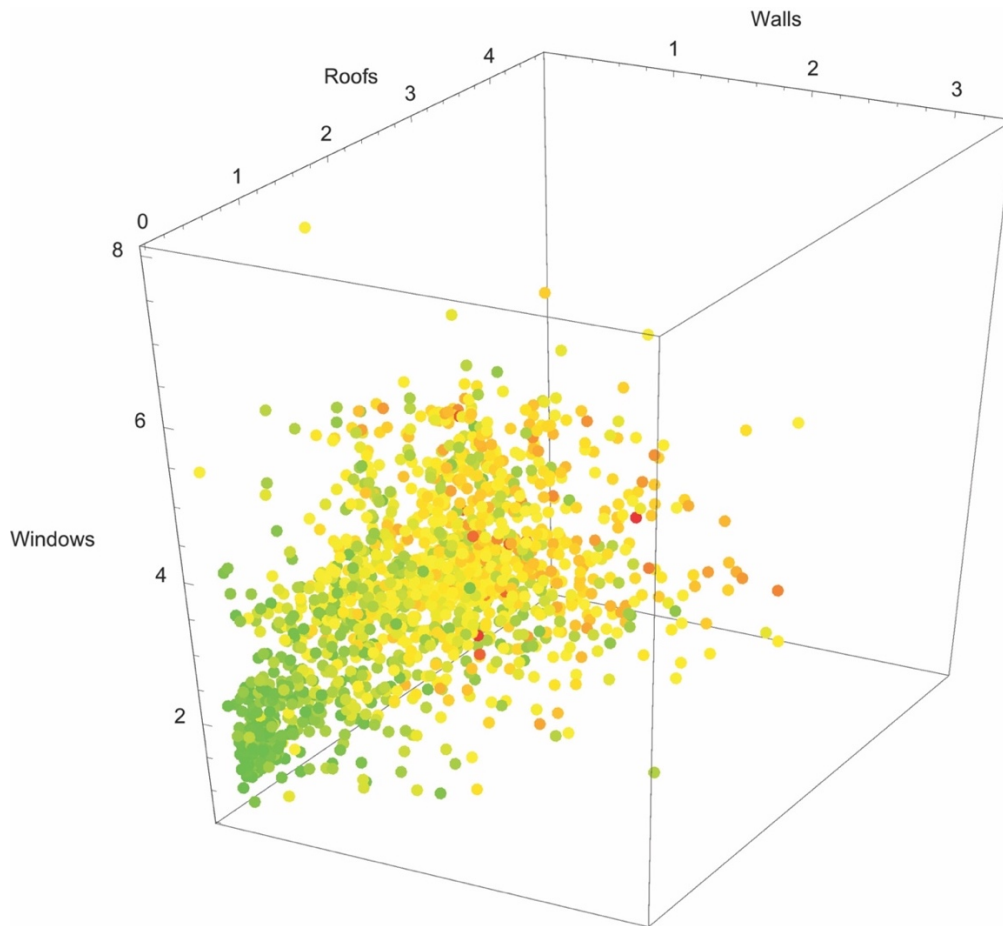


Figure 4. EPh according to thermal transmittance of walls, windows and roofs.

2.2 Clustering building technologies

For the definition of an appropriate retrofit strategy, it is necessary to identify homogeneous classes of buildings, on which to implement retrofit interventions. The aim of this step is to classify objects in the same class when they are as similar as possible, whereas in different classes when they are as dissimilar as possible. For this purpose, many different clustering methods can be employed. Clustering is an unsupervised learning [46] task that aims at decomposing a given set of objects into subgroups based on similarity. The most known methods to identify clusters are Hierarchical, K-means, K-medoid, etc. In this research most of the cluster have been identified using K-means method [47], i.e. a partitioning methods [48] consisting in moving instances from one cluster to another, starting from an initial partitioning. For the identification of homogeneous groups of elements, according to the U-value for each of the three main technological units (i.e. walls, roofs and windows), a clustering algorithm has been run. This gave as output, in most of the cases, a clusterisation of U-values in three groups. In Figure 5 are described clusters obtained for the technological unit walls according to the construction year. The U-value thresholds identified for the clusterisation can change according to the sample taken into account, thus colours only represent the belonging to a specific cluster and not the width of the cluster. Figure 5 represents only data concerning walls, since the representation and the description of 21 clusterisation analysis for the three main technological units would have been redundant in this article. The lower cluster, for all the components and classes of construction years is considered as the best performing situation, especially for those buildings belonging to the older construction year classes. Nevertheless, Table 2 shows the mean of each cluster for the three components of the school building envelope. As it can be

seen, roofs are responsible for the largest energy flows, thus their U-Values are highly affecting buildings' energy demand.

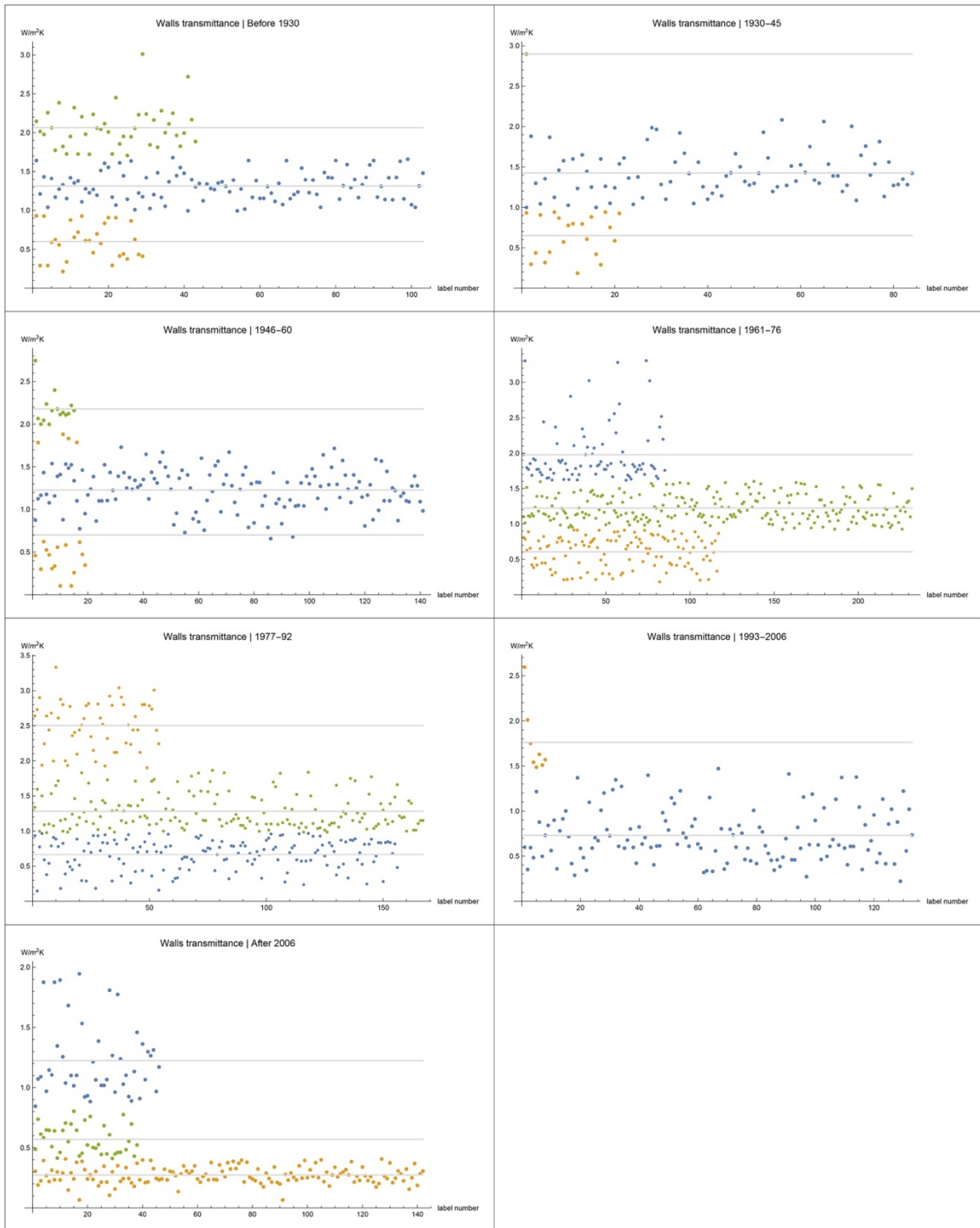


Figure 5. Clusters of walls' thermal transmittance according to construction year.
 Table 2. Mean U-values for each cluster and for the three components of the envelope.

		Before 1930	1930- 45	1946-60	1961-76	1977- 92	1992-06	After 2006
Walls	Cluster 1	2.063	2.898	2.180	1.978	2.504		1.224
	Cluster 2	1.314	1.426	1.229	1.224	1.283	1.761	0.569
	Clustre 3	0.600	0.651	0.702	0.606	0.666	0.731	0.273

Windows	Cluster 1	4.888	5.064		5.214	5.200	4.760	4.967
	Cluster 2	3.608	3.630	6.751	3.630	3.640	3.520	3.186
	Clustre 3	2.211	2.152	3.596	2.097	2.257	2.346	1.736
Roofs	Cluster 1	2.832		1.935	3.133	2.597		1.621
	Cluster 2	1.518	1.274	1.397	1.524	1.528	1.463	0.679
	Clustre 3	0.563	0.254	0.587	0.669	0.668	0.562	0.257

2.3 Computing savings with Artificial Neural Networks

Once analysis employed for achieving the knowledge of the database have been carried out, the energy savings have been predicted through the ANNs. In the following paragraphs the steps for achieving these results are described.

2.3.1 Artificial Neural Network training

All the records in the CENED DB are divided in 7 classes according to the year of construction or major refurbishment of the building. The time span goes from before 1930 (class 1) to after 2006 (class 7). Accordingly, seven ANNs have been trained on each CENED construction year class. In addition, one ANN has been trained on the whole cleaned CENED DB. The depth of the network (number of layers) and the type of each layer (type of equations implemented by the artificial neurons) have been chosen according to a trial-based process, which allowed to define the most suitable parameters. The number of neurons (in Table 3 the number below the type of function) of each layer has been optimised with an automatic process. Figure 6 shows the performance of the Total network, the one trained on the whole CENED dataset, according to the number of neurons on the layers. Being a black box, the ANN does not need input parameters directly related to the physical model in order to make reliable previsions. Therefore, input parameters have been defined empirically, as the ones that help the net to better interpolate the EPh values in the training phase. For instance, though the EPh takes into account also the building's systems, among the selected parameters the efficiency of the system is not considered. Parameters selected for the total ANN are:

1. winter degree days;
2. construction year;
3. gross surface [m²];
4. gross volume [m³];
5. dispersant surface [m²];
6. ratio between glazing surface and dispersant surface;
7. ratio between opaque surface and dispersant surface;
8. average U-value of walls [W/m² K]
9. average U-value of roof [W/m² K]
10. average U-value of windows [W/m² K]
11. average U-value of basement [W/m² K].

Obviously, the construction year has been used as input parameter only in the net trained on the whole dataset. In Table 3, the main characteristics of the eight ANNs are presented.

Functions employed for the computation process can be classified as:

- Linear is the function characterising the layers with dense connections computing $w \cdot x + b$;
- Tanh – net layer applies a unary function f to every element of the input tensor, in this case the function is the hyperbolic tangent;

- Ramp – net layer applies a unary function f to every element of the input tensor, in this case the function, gives x if $x \geq 0$ otherwise 0 .

Table 3: ANNs characteristics

	<1930	1931-45	1946-60	1961-76	1977-92	1993-06	>2006	Total
Layer 1	Linear 9	Linear 9	Linear 9	Linear 9	Linear 9	Linear 9	Linear 9	Linear 10
Layer 2	Linear 126	Linear 406	Linear 196	Linear 412	Linear 158	Linear 68	Linear 236	Linear 476
Layer 3	Tanh 126	Tanh 406	Tanh 196	Tanh 412	Tanh 158	Tanh 68	Tanh 236	Tanh 476
Layer 4	Linear 126	Linear 406	Linear 196	Linear 412	Linear 158	Linear 68	Linear 236	Linear 476
Layer 5	Tanh 126	Tanh 406	Tanh 196	Tanh 412	Tanh 158	Tanh 68	Tanh 236	Tanh 476
Layer 6	Linear 63	Linear 203	Linear 98	Linear 206	Linear 79	Linear 34	Linear 118	Linear 238
Layer 7	Ramp 63	Ramp 203	Ramp 98	Ramp 206	Ramp 79	Ramp 34	Ramp 118	Ramp 238
Layer 8	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1
Layer 9	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1	Linear 1

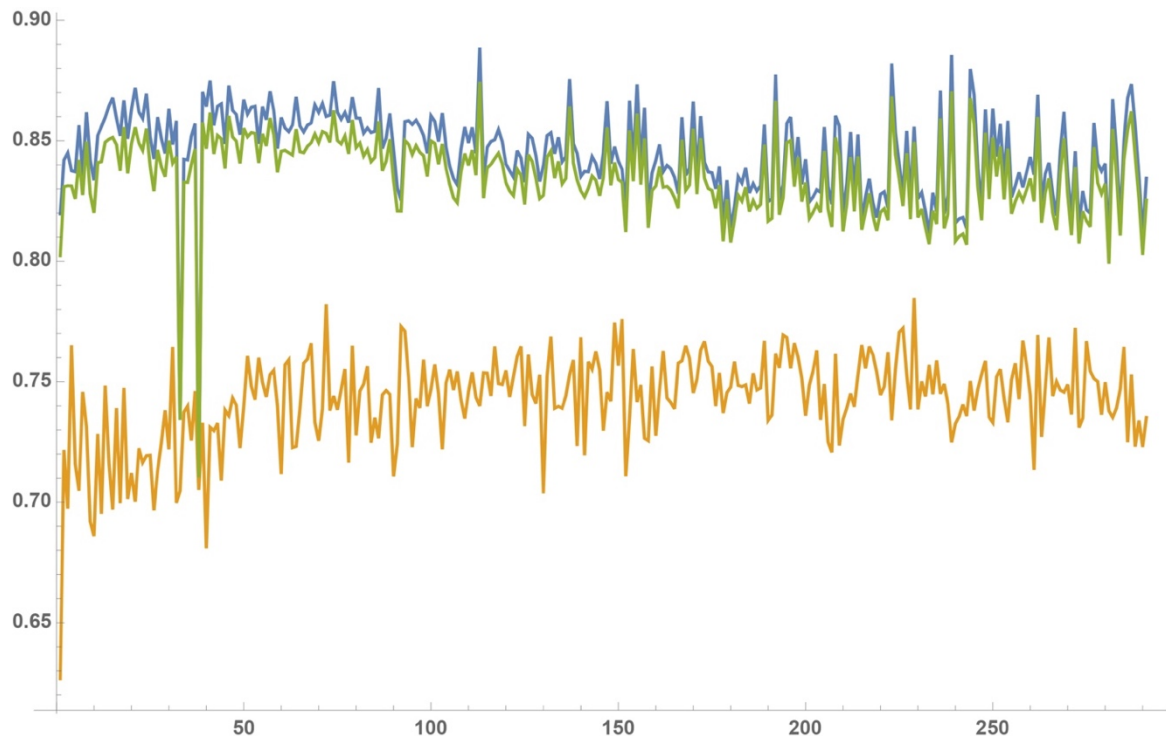


Figure 6. ANN performance according to the number of neurons (total dataset).

The best forecasts are provided by the mean of the prediction given by the ANN for the specific time span and the forecast by the net trained on the whole DB, using this mean the correlation between CENED data and predicted values is equal to 0.948329. Figure 7 compares predicted Eph to the actual one in the CENED DB.

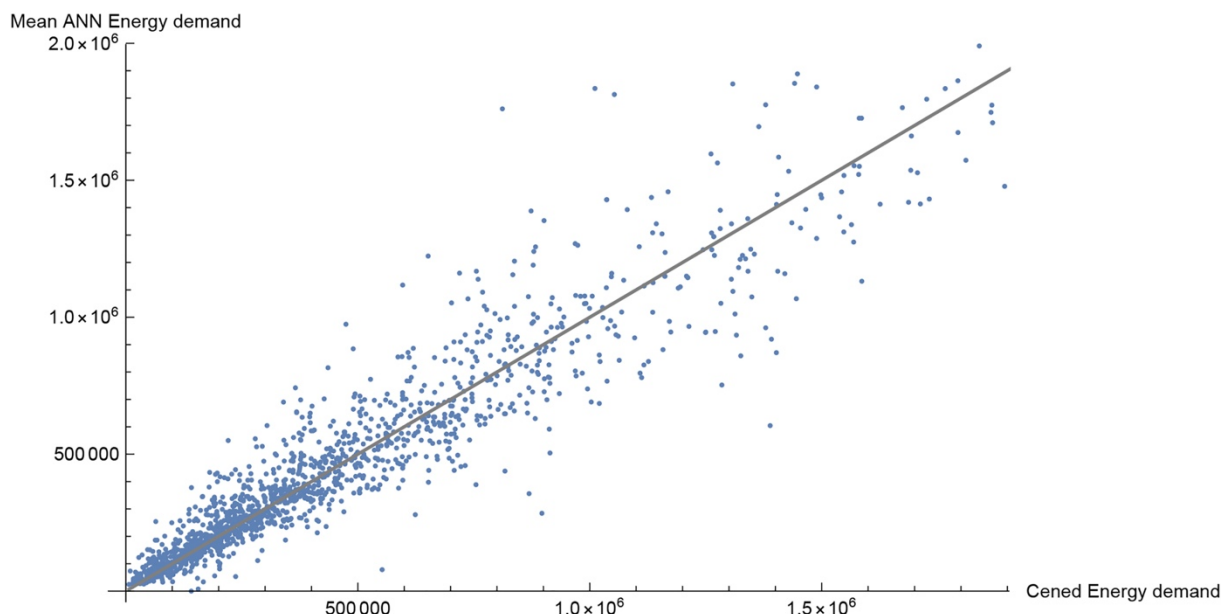


Figure 7. Correlation between the CENED Energy Demand and the one forecasted by the ANN.

2.3.2 Use of Artificial Neural Networks to compute energy savings

Once the ANNs have been defined and trained, they have been employed for the prediction of energy savings, according to the three retrofit strategies described in Table 4. The prediction provides as outcome the energy consumption of buildings, after the improvement of the performance of the envelope's components, within the parameters defined in Table 4. The parameters (U-values for walls, roofs and windows) used for the characterisation of the three retrofit cases have been chosen according to the clusterisation process represented in Figure 5. For the Retrofit case 1, U-values, respectively, for walls, roofs and windows are set to 1.0, 1.0, and 3.5 W/m²K. These values represent the thresholds between the worst class, i.e. the one including components with the highest U-values, and the intermediate class of the clusterisation (Figure 5). The aim is not to retrofit school buildings already performing well, but to address the most part of interventions on those which sit in a worst energy performance condition. Retrofit case 3 has been defined in order to refurbish the whole amount of schools whose envelope has a thermal transmittance higher than the threshold set by Italian and regional Laws. Italian Decreto Ministeriale (DM) 26/06/2015 [49] defines, for the Lombardy region, limits of transmittance values for walls, roofs and windows. These limits are set to 0.26, 0.22 and 1.4 W/m²K for buildings in climate zone E and to 0.24, 0.2 and 1.1 for buildings in climate zone F. [49]. Retrofit case 2 has been defined to provide an intermediate scenario between the two described above. Table 4 summarises key values for the three retrofit scenarios. Although retrofit cases are described by parameters that do not consider the plants, ANNs have been trained from a database where buildings with low levels of transmittance have efficient plants, thus prediction of energy savings already embed an improvement of the plants.

Table 4: Retrofit cases.

Case 1	Case 2	Case 3
--------	--------	--------

	U-value limit [W/m ² K]	Components below the limit	U-value limit [W/m ² K]	Components below the limit	U-value limit [W/m ² K]	Components below the limit
Walls	1	58%	0.5	82%	0.26/0.24	95%
Roofs	1	52%	0.5	78%	0.22/0.2	92%
Windows	3.5	49%	2	87%	1.4/1.1	97%

3. RESULTS

The computed energy savings are presented below for each retrofit case, highlighting the total energy consumption pre-retrofit interventions, total energy consumption post-retrofit and total savings for the whole school building stock. To assure a more homogeneous comparison between pre and post-retrofit energy demands, even the pre-retrofit values have been computed using the ANNs. This approximation does not affect the results since the difference between the computed values and the CENED values is small (Figure 7).

Table 5, Table 6 and Table 7 present the total savings, the average savings and the number of retrofitted schools according to the type of retrofit needed. The type of retrofit intervention has been named according to main building components refurbished (Complete retrofit, Walls & Roof, Walls & Windows, etc.).

3.1 Retrofit case 1

Results obtained running the ANN on the whole school building stock are:

- total energy consumption in current status (pre-retrofit interventions): $8.9712 \cdot 10^5$ MWh/y;
- total energy consumption post-retrofit: $5.25874 \cdot 10^5$ MWh/y;
- 1,216 refurbished schools, out of 1,634 in the database for a total saving of $3.57953 \cdot 10^5$ MWh/y. In average, for each school a saving of 294.370 MWh/y has been computed.

Table 5: Number of retrofit schools, total and average savings by type of intervention (Case 1).

Type of retrofit	Complete retrofit	Walls & Roofs	Walls & Windows	Roofs & Windows	Only walls	Only Roofs	Only windows
Number of retrofit	462	227	153	68	98	88	120
Tot. savings [MWh/y]	185,228.92	57,387,267	53,250.59	25,433.07	9,107.18	8,153.68	19,392.92
Average savings [kWh/y]	400,928.40	252,807.34	348,043.07	374,015.67	92,930.41	92,655.45	161,607.68

3.2 Retrofit case 2

Results obtained running the ANN on the whole school building stock are:

- total energy consumption in current status (pre-retrofit interventions): $8.9712 \cdot 10^5$ MWh/y;
- total energy consumption post-retrofit: $4.19596 \cdot 10^5$ MWh/y;
- 1,487 refurbished schools, out of 1,634 in the database for a total saving of $3.93793 \cdot 10^5$ MWh/y. In average, for each school a saving of 264.824 MWh/y has been computed.

Table 6: Number of retrofit schools, total and average savings by type of intervention (Case 2).

Type of retrofit	Complete retrofit	Walls & Roofs	Walls & Windows	Roofs & Windows	Only walls	Only Roofs	Only windows
Number of retrofit	1,166	38	118	56	19	14	76
Whole savings [MWh/y]	366,628.14	3,343.63	13,315.08	5,635.82	484.47	297.08	4,088.77
Average savings [kWh/y]	314,432.37	87,990.37	112,839.63	100,639.62	25,498.61	21,219.80	53,799.58

3.3 Retrofit case 3

Results obtained running the ANN on the whole school building stock are:

- total energy consumption in current status (pre-retrofit interventions): $8.9712 \cdot 10^5$ MWh/y;
- total energy consumption post-retrofit: $3.44654 \cdot 10^5$ MWh/y;
- 1,620 schools refurbished, out of 1,634 in the database for a total saving of $3.94719 \cdot 10^5$ MWh/y. In average, for each school a saving of 243.654 MWh/y has been computed

Table 7: Number of retrofit schools, total and average savings by type of intervention (Case 3).

Type of retrofit	Complete retrofit	Walls & Roofs	Walls & Windows	Roofs & Windows	Only walls	Only Roofs	Only windows
Number of retrofit	1,497	20	25	36	4	12	26
Whole savings [MWh/y]	390,862.86	948.02	751.22	808.34	80.60	603.90	664.68
Average savings [kWh/y]	261,097.43	47,400.81	30,048.61	22,453.93	20,151.04	50,324.96	25,564.48

Figure 8 shows a comparison between the number of retrofit interventions in each retrofit scenario and the corresponding average savings per school. The graph shows that increasing the number of buildings to be refurbished, the average savings decrease (Retrofit case 3). This trend is due to the fact that lowering the thresholds of retrofitting parameters (walls, roof and windows transmittance), the ANN produces more complete retrofit forecasts than in the other two cases. Therefore, also schools with better energy performances are included in the retrofit scenario. These buildings are the ones where lower savings can be achieved. The retrofit case 1 can be considered the most viable among the three, presenting higher average post-retrofit savings, though the percentage of total retrofit is much lower than in case 2 and case 3. This suggests that the average retrofit cost will be the lowest among the three cases. Despite the highly precise forecasting provided by the ANN, it must be considered that the database on which savings have been calculated is the cleaned one, therefore, some school buildings might not be present.

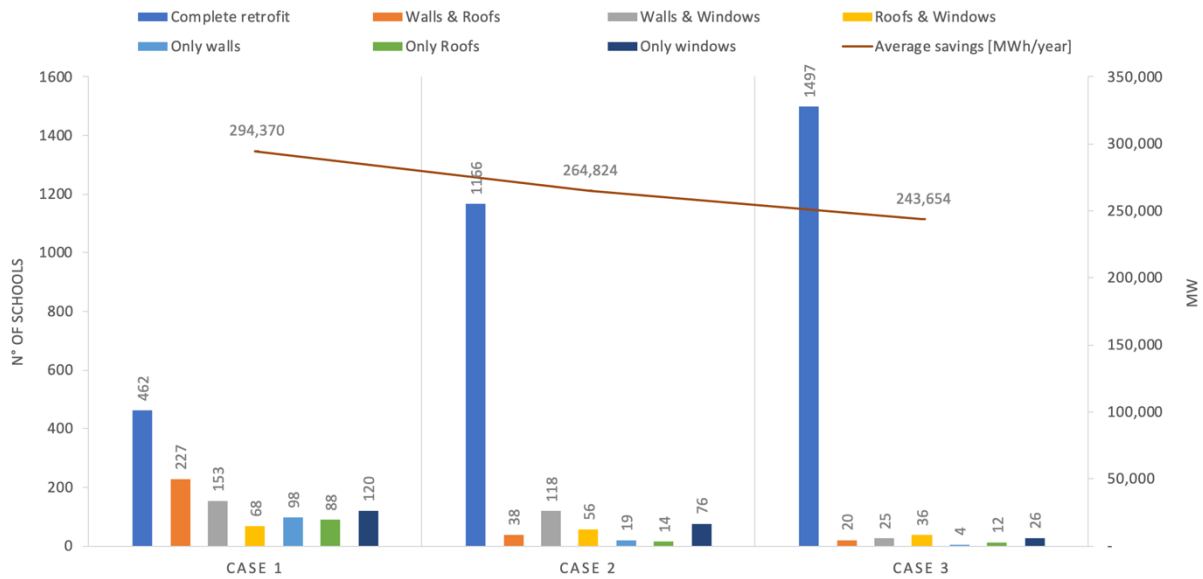


Figure 8. Number of retrofit works divided by type compared to the potential average savings.

4. GEOGRAPHIC INFORMATION SYSTEM INTEGRATION FOR PORTFOLIO MANAGEMENT

Results of the previous phases have been imported and processed in GIS. GIS analysis and representations are here presented only for one of the three retrofit scenarios: case 1. In order to integrate energy savings computed with the ANNs with spatial information it has been necessary to run a geocoding script [50]. The address field in the CENED cleaned DB has been corrected according to the format required by the Google Application Programming Interface (API), since the geocoding process has been carried out exploiting the online open geocoding service provided by Google. The geocoding process, gave as outcome a point layer, whose attribute table contain values obtained as outcome of the three retrofit cases. This process gave an error log of 47 addresses not found, corresponding to 2.9% of the whole cleaned database. Values in the map have been categorised and represented according to the predicted energy demands (dimension of the marker) and according to the type of intervention for achieving the energy retrofitting strategy (colour of the marker). Figure 9 represents the school buildings retrofitted according to the Case 1.

The geographic representation of data obtained through ANN prediction allows to spot most critical areas in the Lombardy region, namely where through the retrofitting interventions it is possible to achieve the highest energy saving. These data, combined with other concerning, for instance, cost for retrofit interventions, age of the assets, demographic trends etc. can be exploited by the public administrations for making informed decisions on the school building portfolio.

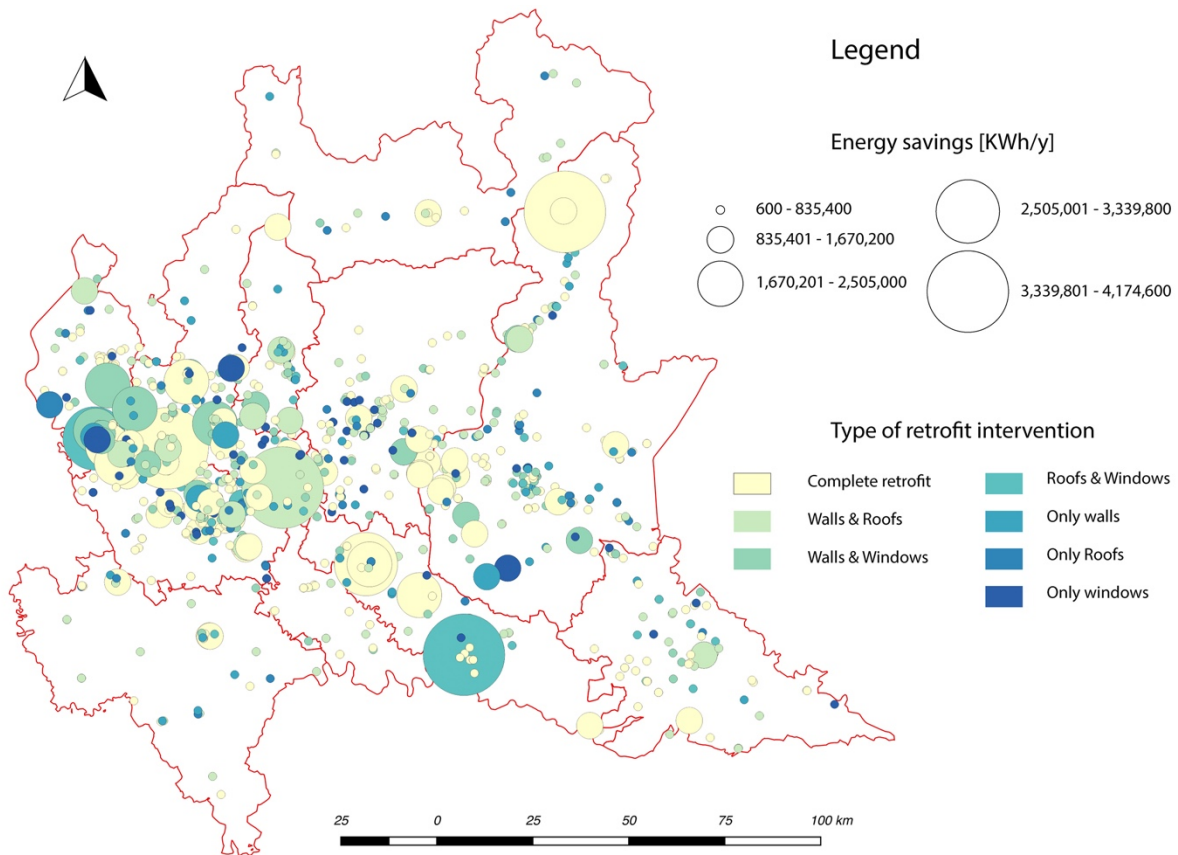


Figure 9. ANN computed energy savings by type of intervention in the Lombardy region (Case 1).

5. DISCUSSION AND CONCLUSIONS

In a period of public spending review, the public administration needs cost-effective methods to find the way to address energy related issues of buildings stocks. This research demonstrated that the integrated use of open-data, ANN and GIS satisfies this need providing efficient methods to set regional energy policy. The primary advantage of the proposed approach concerns the possibility to compute buildings' post-retrofit Eph without an on-site inspection and using parameters which are easily retrieved or computed. The second important outcome of the research is that it is further demonstrated the feasibility and usefulness of employing GIS tools in energy policies. Focusing on the case study, a third result of the research is that it is demonstrated that increasing the number of retrofitted buildings (selecting a lower U-value retrofit threshold for the envelope), average energy savings decrease. This suggests that the retrofit of the whole school buildings stock in Lombardy region is not cost-effective.

Energy savings are rather important for the development of an effective asset and portfolio management strategy [51], though when dealing with school buildings, other parameters related to the learning performance should be taken into account [52]. For this reason, in order to prioritise retrofitting interventions, it is relevant, for instance, to take into account also compliance of the building to the contemporary standards (e.g. flexibility of the spaces, dimensions of the classrooms, fire safety codes etc.). Thus, the proposed methodology should be encompassed in a wider framework.

Moreover, it must be considered that the methods presented in this article are not the only tools to be adopted in decision making. The demographic trends, for instance, have not been considered in the analysis. Nevertheless, they could be crucial for the definition of a fair policy. Moreover, GIS tools has been employed mainly for representing effectively the results of the computation of the retrofit scenarios, despite other insights could be achieved if this information would be combined

with other spatial data which could describe further phenomena at the regional level. For instance, a further research insight can be carried out for the definition of the most suitable U-values thresholds to be adopted when the focus of the refurbishment is an historical or a protected building. Indeed, in this case, the modification of the envelope is way more critical and, due to the complexity of the intervention, it may cause additional costs, therefore it needs special attention. Nevertheless, the proposed methods must be considered within the context of high-level portfolio management strategy definition. Therefore, once the retrofit policy is developed and implemented a retrofit project must be defined case by case.

Moreover, a further integration of the ANNs and GIS will be achieved through a future development of the research, concerning the calculation of the retrofitting costs of the school buildings. Regione Lombardia provides free datasets concerning the main geometric characteristics of the buildings, in compliance with the EU INSPIRE Directive [41]. These data can be used to estimate the surfaces and volumes of buildings subject to retrofitting, supporting the cost assessment of different types interventions. Once the geometries are computed, a following step concerns the identifications of the costs related to the retrofit interventions, this will be done through the identification of walling, roofing and glazing technologies according to the year of construction and to the thermal transmittance. This requires further research efforts, testing of the methods and integration with additional open spatial datasets. Moreover, despite the research has been carried out at the Lombardy region level, it is possible to apply the same methods to further datasets at the national or European level, since member states must be compliant to Energy Performance Building Directive (EPBD) [5].

Declaration of interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6. REFERENCES

- [1] Dall'O' G, Sarto L. Potential and limits to improve energy efficiency in space heating in existing school buildings in northern Italy. *Energy Build* 2013;67:298–308. doi:10.1016/J.ENBUILD.2013.08.001.
- [2] Zhang Y, Barrett P. Findings from a post-occupancy evaluation in the UK primary schools sector. *Facilities* 2010;28:641–56. doi:http://dx.doi.org/10.1108/MRR-09-2015-0216.
- [3] Chatzidiakou L, Mumovic D, Dockrell J. The Effects of Thermal Conditions and Indoor Air Quality on Health, Comfort and Cognitive Performance of Students. 2014. doi:10.1002/psc.385.
- [4] Gupta R, Gregg M. Empirical evaluation of the energy and environmental performance of a sustainably-designed but under-utilised institutional building in the UK. *Energy Build* 2016;128:68–80. doi:10.1016/j.enbuild.2016.06.081.
- [5] European Parliament and European Council. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast). *Off J Eur Union* 2010;13–35. doi:doi:10.3000/17252555.L_2010.153.eng.
- [6] Vanier D, Tesfamariam S, Sadiq R, Lounis Z. Decision models to prioritize maintenance and renewal alternatives. *Jt. Int. Conf. Comput. Decis. Mak. Civ. Build. Eng. Montr.*, 2006, p. Pages 2594-2603.
- [7] Infrastrutture Lombarde. Infrastrutture Lombarde S.p.A. - Dati CENED n.d. http://www.cened.it/focus_ceer (accessed March 5, 2018).
- [8] Yılmaz Y, Koçlar Oral G. An approach for an educational building stock energy retrofits through life-cycle cost optimization. *Archit Sci Rev* 2018;61:122–32.

- doi:10.1080/00038628.2018.1447438.
- [9] MIUR Ministero dell'Istruzione dell'Università e della R. Linee Guida per le architetture interne delle scuole 2013.
 - [10] Legge 373/1976 n.d.
 - [11] Legge 9 gennaio 1991, n. 10. Norme per l'attuazione del Piano energetico nazionale in materia di uso nazionale dell'energia, di risparmio energetico e di sviluppo delle fonti rinnovabili di energia 1991.
 - [12] Basarir B, Diri BS, Diri C. Energy efficient retrofit methods at the building envelopes of the school buildings 2012.
 - [13] European Parliament and European Council. Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency. Off J Eur Union 2018;2018.
 - [14] ENEA-FIRE. Guida per il contenimento della spesa energetica nelle scuole n.d.
 - [15] Lohse R, Staller H, Riel M. The economic challenges of deep energy renovation - Differences, similarities, and possible solutions in central Europe: Austria and Germany. ASHRAE Trans., vol. 122, Amer. Soc. Heating, Ref. Air-Conditioning Eng. Inc.; 2016, p. 69–87.
 - [16] dPR 59/2009 (eff.energ.edilizia) n.d.
 - [17] Minoli D, Sohraby K, Occhiogrosso B. IoT Considerations, Requirements, and Architectures for Smart Buildings—Energy Optimization and Next-Generation Building Management Systems. IEEE Internet Things J 2017;4:269–83. doi:10.1109/JIOT.2017.2647881.
 - [18] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92. doi:10.1016/J.RSER.2012.02.049.
 - [19] Cecconi FR, Manfren M, Tagliabue LC, Ciribini ALC, De Angelis E, Re Cecconi F, et al. Probabilistic behavioral modeling in building performance simulation: A Monte Carlo approach. Energy Build 2017;148:128–41. doi:10.1016/j.enbuild.2017.05.013.
 - [20] Azadeh A, Babazadeh R, Asadzadeh SM. Optimum estimation and forecasting of renewable energy consumption by artificial neural networks. Renew Sustain Energy Rev 2013;27:605–12. doi:10.1016/J.RSER.2013.07.007.
 - [21] Khosravani H, Castilla M, Berenguel M, Ruano A, Ferreira P. A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building. Energies 2016;9:57. doi:10.3390/en9010057.
 - [22] Reich Y. Machine learning techniques for civil engineering problems. Comput Civ Infrastruct Eng 1997;12:295–310. doi:10.1111/0885-9507.00065.
 - [23] Tang S, Sheldon DR, Eastman CM, Pishdad-Bozorgi P, Gao X. A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. Autom Constr 2019;101:127–39. doi:10.1016/j.autcon.2019.01.020.
 - [24] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;82:1027–47. doi:10.1016/j.rser.2017.09.108.
 - [25] Jovanović R, Sretenović AA, Živković BD. Multistage ensemble of feedforward neural networks for prediction of heating energy consumption. Therm Sci 2016;20:1321–31. doi:10.2298/TSCI150122140J.
 - [26] Khayatian F, Sarto L, Dall'O' G. Application of neural networks for evaluating energy performance certificates of residential buildings. Energy Build 2016;125:45–54. doi:10.1016/j.enbuild.2016.04.067.
 - [27] Hawkins D, Hong SM, Raslan R, Mumovic D, Hanna S. Determinants of energy use in

- UK higher education buildings using statistical and artificial neural network methods. *Int J Sustain Built Environ* 2012;1:50–63. doi:10.1016/j.ijbsbe.2012.05.002.
- [28] Ferlito S, Atrigna M, Graditi G, De Vito S, Salvato M, Buonanno A, et al. Predictive models for building’s energy consumption: An Artificial Neural Network (ANN) approach. 2015 XVIII AISEM Annu. Conf., IEEE; 2015, p. 1–4. doi:10.1109/AISEM.2015.7066836.
- [29] Ekici BB, Aksoy UT. Prediction of building energy consumption by using artificial neural networks. *Adv Eng Softw* 2009;40:356–62. doi:10.1016/J.ADVENGSOFT.2008.05.003.
- [30] Ascione F, Bianco N, De Stasio C, Mauro GM, Vanoli GP. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy* 2017;118:999–1017. doi:10.1016/j.energy.2016.10.126.
- [31] Hou Z, Lian Z, Yao Y, Yuan X. Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. *Appl Energy* 2006;83:1033–46. doi:10.1016/J.APENERGY.2005.08.006.
- [32] Yalcintas M, Aytun Ozturk U. An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. *Int J Energy Res* 2007;31:412–21. doi:10.1002/er.1232.
- [33] Hong S-M, Paterson G, Mumovic D, Steadman P. Improved benchmarking comparability for energy consumption in schools. *Build Res Inf* 2014;42:47–61. doi:10.1080/09613218.2013.814746.
- [34] Rodrigues E, Gomes Á, Gaspar AR, Henggeler Antunes C. Estimation of renewable energy and built environment-related variables using neural networks – A review. *Renew Sustain Energy Rev* 2018;94:959–88. doi:10.1016/j.rser.2018.05.060.
- [35] Boyes GA, Ellul C, Irwin D. Exploring BIM for operational integrated asset management - a preliminary study utilising real-world infrastructure data. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2017;4:49–56. doi:10.5194/isprs-annals-IV-4-W5-49-2017.
- [36] Hill LL, Dalkiliç MM, Medjahed B, Ouzzani M, Elmagarmid AK, Hellerstein JM, et al. Geographic Information System. *Encycl. Database Syst.*, Boston, MA: Springer US; 2009, p. 1231–6. doi:10.1007/978-0-387-39940-9_178.
- [37] Ellul C, Stoter J, Harrie L, Shariat M, Behan A, Pla M. INVESTIGATING THE STATE OF PLAY OF GEOBIM ACROSS EUROPE. *ISPRS - Int Arch Photogramm Remote Sens Spat Inf Sci* 2018;XLII-4/W10:19–26. doi:10.5194/isprs-archives-XLII-4-W10-19-2018.
- [38] Koo C, Hong T, Lee M, Kim J. An integrated multi-objective optimization model for determining the optimal solution in implementing the rooftop photovoltaic system. *Renew Sustain Energy Rev* 2016;57:822–37. doi:10.1016/j.rser.2015.12.205.
- [39] Resch B, Sagl G, Törnros T, Bachmaier A, Eggers J-B, Herkel S, et al. GIS-Based Planning and Modeling for Renewable Energy: Challenges and Future Research Avenues. *ISPRS Int J Geo-Information* 2014;3:662–92. doi:10.3390/ijgi3020662.
- [40] Belyakov S, Savelyeva M, Samoylov L. Geoinformation Models in Smart Grid Management. 2018 3rd Russ. Conf. Comput. Technol. Appl., IEEE; 2018, p. 1–5. doi:10.1109/RPC.2018.8482218.
- [41] EU. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) 2007.
- [42] Territorio PDEL. “Legge per il governo del territorio”. 2005.
- [43] Regione Lombardia. Geoportale Regione Lombardia 2017. <http://www.geoportale.regione.lombardia.it/> (accessed February 3, 2017).
- [44] Borgogno Mondino E, Fabrizio E, Chiabrando R. Site Selection of Large Ground-

- Mounted Photovoltaic Plants: A GIS Decision Support System and an Application to Italy. *Int J Green Energy* 2015;12:515–25. doi:10.1080/15435075.2013.858047.
- [45] Yeo I-A, Yee J-J. A proposal for a site location planning model of environmentally friendly urban energy supply plants using an environment and energy geographical information system (E-GIS) database (DB) and an artificial neural network (ANN). *Appl Energy* 2014;119:99–117. doi:10.1016/j.apenergy.2013.12.060.
- [46] Hastie T, Tibshirani R, Friedman J. *Unsupervised Learning*. *Elem. Stat. Learn.*, Springer; 2009. doi:doi.org/10.1007/978-0-387-84858-7_14.
- [47] Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett* 2010;31:651–66. doi:10.1016/j.patrec.2009.09.011.
- [48] Berkhin P. A Survey of Clustering Data Mining Techniques. *Group. Multidimens. data*, Springer; 2006, p. 127–60. doi:doi.org/10.1007/3-540-28349-8_5.
- [49] Decreto interministeriale 26 giugno 2015 - Applicazione delle metodologie di calcolo delle prestazioni energetiche e definizione delle prescrizioni e dei requisiti minimi degli edifici n.d.
- [50] Cetl V, Kliment T, Jogun T. A comparison of address geocoding techniques—case study of the city of Zagreb, Croatia. *Surv Rev* 2018;50:97–106. doi:10.1080/00396265.2016.1252517.
- [51] ISO. BS ISO 55000:2014 Asset management Overview, principles and terminology 2014.
- [52] Determan J, Akers MA, Williams I, Hohmann C, Martin-Dunlop C. *Learning Space Design for the Ethnically Diverse Undergraduate Classroom*. 2015.