A GIS-Statistical Approach for Assessing Built Environment Energy Use at Urban Scale

A GIS-statistical approach for assessing built environment energy use aturban scale

Sara Torabi Moghadam[a,*], Jacopo Toniolo[a], Guglielmina Mutani[b], Patrizia Lombardi[a]

[a] Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, Viale Mattioli 39, Turin 10125, Italy
[b] Department of Energy, Politecnico di Torino, C.so Duca degli Abruzzi 24, Turin 10129, Italy

A B S T R A C T

Energy consumption modelling at the urban scale is crucial for supporting a transition towards the low-carboncity. Unfortunately, there are not many robust examples or standardised approaches available in the literature for delivering effective low-carbon urban energy planning. In particular, there is a lack of appropriate frame-works or systems which allow an effective and reliable assessment of energy use in the built environment atdistrict-urban scale.

This paper illustrates the development of a geospatial bottom-up statistical model to estimate the energy consumption of a large number of residential building stocks for heating space, considering a wide range ofvariables. The proposed methodology is based on a 2D/3D- Geographic Information System (GIS) and MultipleLinear Regression (MLR), which provides location-based information for each single dwelling to identify cor-relations and assess the demand-side consumption at the urban scale. This framework was tested on a medium-sized Italian city, including around 3600 residential buildings. The results provided by the model are validated using residual analysis and cross-validation. Moreover, the spatial results provided by this study represent a useful tool to aid decision-makers in the urban planning process. These results can help to create future energytransition strategies, implementing energy efficiency and renewable energy technologies in the context of sus-tainable cities. This work is part of a national Smart City & Communities project, named "EEB- Zero Energy Buildings in Smart Urban Districts"; nonetheless, the methodology illustrated in this paper can be generalisedand applied to other European urban contexts.

1. Introduction

Cities are one the most energy consumers in the world (United Nations, 2015), especially, urban buildings account for 60% of total building final energy use (IEA, 2016). In Europe, existing buildings represent the vast majority of the building stock, which are predominantly characterised by low energy performance (BPIE, 2011). Therefore, a rapid transition of urban areas towards a low-carbon scenario is required, taking into account national priorities in the definition of strategic goals at regional and local level (Lombardi et al., 2014).

An accurate diagnosis of the current state of energy consumption is the preliminary key step for the development of low-carbon urban strategies (Howard et al., 2012). Diagnoses require the development of energy consumption modelling to quantify the actual consumption. This will support better urban energy management and the identification of urban energy retrofitting (Howard et al., 2012).

Although different effective energy performance analyses for single buildings are well documented in the literature, comprehensive energy performance analyses at urban-scale have not been developed yet.

According to the rapid growth of data availability sources, statistical techniques can extensively identify the associations and correlations among various variables influencing building energy performance. These methods utilise samples of historical consumption information as a data source for energy modelling, analysing them based on different variables. They are also capable to take into account socioeconomic effects in the consumption equation (Nouvel et al., 2015).

Recently, several appropriate models were presented by means of statistical analysis on energy planning and building stock at the urban scale. In this context, Theodoridou, Papadopoulos, and Hegger (2011) conducted a statistical analysis of the features of the residential building stock relative to energy consumption and potential of energy savings to classify the building typologies in Greece. In 2008, the statistical method for space heating carried out by Caldera, Corgnati, and Filippi (2008) was based on a dataset of 50 multifamily residential buildings, finding out a simple correlation among energy demand for space heating, construction age and thermo-physical and geometrical features. Moreover, Fracastoro and Serraino (2011) developed an analytical method, starting from census data and energy statistics to define the statistical distribution of residential buildings according to energy consumption for heating demand. Another interesting statistical modelling and analysis of energy consumption for the buildings sector were conducted by Hsu (2015). This study highlighted the interactions of several parameters, both technical and non-technical, for developing accurate analysis and policy formulation at the local level.

In 2016, a multivariate linear regression model with numerical predictors and categorical indicator variables to predict energy use intensity was developed by Walter and Sohn (2016). This model measures the contribution of building characteristics and systems to energy use based on the cross-validation approach.

In all aforementioned studies, statistical analysis techniques were applied to identify the most influential variables on energy consumption in buildings. One of the major strengths of statistical analysis is the widespread familiarity of this methodology and its simplicity. On the other hand, there is still an inadequate integration between energy data and spatial planning (Zanon and Verones, 2013).

Since built environment data and information at the local level are significantly scattered among several entities, and there is a lack of interoperability among the data sources, one of the most challenging barriers in developing a robust and detailed analysis is data collection (Caputo & Pasetti, 2015). In this regard, a huge effort is required in order to provide a supportive and comprehensive accessible building stock database, at the local level for different goals and different stakeholders, gathering all the necessary data from various sources (Cajot et al., 2017; Caputo & Pasetti, 2015).

In Italy, information about building stock and their energy performances are derived from different regional and local authorities and often are not homogeneous (e.g., ISTAT, Italian National Institute of Statistics, ARPA, Regional Agency for Environmental Protection, Regional register of energy performance certificates and AEEG Regulatory Authority for Electricity and Gas) (Caputo, Costa, & Ferrari, 2013; Torabi Moghadam, Lombardi, & Mutani, 2017).

Therefore, in order to set up an effective energy planning at the local scale, it is crucial to improve data availability and management. Data availability about buildings energy consumption will be increasingly improved in the future, thanks to smart metering and real-time data monitoring, following recent open data policies (COM n.882 final, 2011).

In this context, Geographic Information Systems (GIS) helps to identify and visualise buildings data and their distribution, supporting decision-making, at urban and regional scale. This approach can manage location-based information, linking alfa-numerical information databases to spatial maps to create dynamic displays. Moreover, GIS highlights the high energy use hotspots that need requalification (Chalal, Benachir, White, & Shrahily, 2016). Although GIS is principally used for buildings' geometrical data, it can significantly assist the energy performance of buildings.

GIS-based methodologies for modelling energy consumption and environmental performance have been progressively increased in the last decade in order to help urban energy planning such as (Caputo & Pasetti, 2017; Carozza, Mutani, Coccolo, & Kaempf, 2017; Cheng & Steemers, 2011; Grassi et al., 2012; Iowerth, Lannon, Waldron, Bassett, & Jones, 2013; La Gennusa, Lascari, Rizzo, Scaccianoce, & Sorrentino, 2011; Li, Quan, Augenbroe, Yang, & Brown, 2015; Yang & Yan, 2016).

An interesting example is the hybrid energy model of building stock developed for the City of Goteborg, based on GIS systems by Tornberg and Thuvander (2005). The energy data were measured at a metering station and were linked to function and age of buildings. The outcomes were presented in energy maps, which provided a very useful overview of the energy performance of the city of Goteborg. Recently, Dall'O', Galante, and Pasetti (2012) introduced a GIS-based methodology for creating the comprehensive framework of the energy performance in buildings and applied it to five municipalities in the province of Milan, considering energy using energy audits of sample buildings. This model used the specific primary energy for space heating data to construct regression lines based on shape factor ratio during different construction periods.

Howard et al. (2012) developed the statistical bottom-up GIS-based model for New York City with the aim at estimating the building sector energy end-use intensity for domestic hot water, space heating, and electricity. In this study, building age factor was not considered and the model was performed by robust multivariate linear regression.

Furthermore, Mastrucci et al. (2014a) developed a bottom-up statistical methodology considering dwelling type, period of construction, floor surface and number of occupants. The Ordinary Least- Squares (OLS) method was used to fit the model.

In 2015, Nouvel et al. (2015) proposed a combined methodology, as a multi-framework for urban scale applications, based on Ordinary Least- Squares (OLS) multiple linear-GIS (Mastrucci et al., 2014a) and an engineering model making use of 3D city models.

Another GIS integrated data mining methodology framework for estimating building performances in the urban scale was proposed by Ma & Cheng (2016). This model is based on 216 building features for a case study of 3640 multi-family residential buildings in New York City and is tested and validated. Recently, Braulio-Gonzalo et al. modelled energy performance of existing residential building stocks based on five parameters using simulation software (Braulio-Gonzalo, Juan, Bovea, & Ruá, 2016).

Furthermore, several recent studies highlight experimental new methodologies for evaluating the energy performance of building stock using GIS with regression methods (Torabi Moghadam, Delmastro, Corgnati, & Lombardi, 2017; Torabi Moghadam, Lombardi et al., 2017; Yeo, Yoon, & Yee, 2013).

Although many studies focused on the development of statistical building stock models, the number of studies which adopted a GIS-statistical methodology is quite limited.

Moreover, the previous studies did not use real data but only predicted values. The main difference between the Urban Energy Modelling (UEM) model proposed in this paper and the previously mentioned studies is that it takes into account various real measured data and a significant number of predicted values. The proposed UEM is useful for reducing time-consuming energy demand estimation processes, supporting urban energy planning. Moreover, the spatial results of this study are a useful tool to help decision-makers in the urban planning process to create future energy transition strategies, implementing energy efficiency and renewable energy technologies in the context of sustainable cities. Additionally, the presented model can be applicable to all cities with comparable building stock.

As Section 2 will better illustrate, the authors developed an urban energy model, which describes the current state of urban energy consumption to support decisional process in evaluating future scenarios. The present work represents a useful tool to estimate the energy demand for space heating of existing residential building stock at the urban scale. The specific goal is to create an energy map of the entire city, integrating multiple linear regression statistical techniques and 2D/3D GIS-based methodologies.

The data used in this study derive from a sample of 290 residential buildings, built in different construction periods. Relationships were searched between the various variables that are appropriately combined to discover statistical relations. The estimated energy demand was validated by splitting the data-set into training and testing subsets. Moreover, the cross-validation was also applied for selecting the features more accurately.

For the development of the GIS database, the input data were composed by:

- climate (external air temperatures);
- geometric data (e.g., surface to volume ratio, floor area, number of

floors);
- typology of the building envelope (class of thermal transmittance U for opaque surfaces, class of U for transparent surfaces);
- period of construction
- ground-floor type (commercial, residential and pilotis, which means open space entrance with pillars that support building on the ground floor);
- roof type (flat, gable);
- building type (residential);
- monthly measured data of space heating consumption (two heating seasons).

This work is part of an ongoing Smart City research, a national project called "EEB-Zero Energy Buildings in Smart Urban Districts" (www.smartcommunitiestech.it/). It represents a primary step for the implementation of future energy analyses at the urban level. The results from this study will aid spatial decision-making processes in performing energy planning and testing how different scenarios affect energy performance and carbon emissions and its relationship as well as maintaining the dynamic context of the smart city.

The rest of the paper is organized as follows: details of the proposed framework are illustrated in Section 2. Section 3 presents the application of the proposed methodology to the case study. This application is used for testing the effectiveness of the proposed framework. Finally, conclusive remarks are discussed in Section 4 and future developments are identified.

2. Methodology framework

In order to create a valid and understandable model for urban energy consumption, a methodology was developed to evaluate space heating of residential building stock in an Italian context. This model is applicable to other similar cities. It represents the spatial distribution of urban building energy consumption to ease the decision-making process to simulate different urban transition energy policies according to local conditions.

The proposed methodology is mainly based on existing census data and real measured district heating (DH) energy consumption data. Moreover, GIS was used to identify the geometrical characteristics, data and information of the building stock. The geo-referencing process assists significantly in managing, analysing and visualizing a huge amount of data to support the participative and collaborative workshops for making the better decisions at the urban scale analysis.

Based on the available data, a regression methodology was applied to estimate the energy demand of city residential building stock. Fig. 1 shows the proposed methodology consisting of three major steps:

- *Step 1- data collection and data integration*: the available data on the existing building stock was collected and analysed. All the collected data were overlapped and integrated at this step. Each building polygon was associated with the relative energy consumption and other data. The building stock was thereby characterised. The goal was to create a city GIS Database framework on the factors influencing building energy consumption.
- *Step 2- Parameter identification, modelling and validation*: Firstly, a pre-processing procedure was performed using "missing value replacement" and "outlier detection." Next, a feature selection procedure was applied to the given dataset to identify the most influencing factors on energy performances. Lastly, a robust Multiple Linear Regression (MLR) was employed to evaluate the energy consumption of building stock. The feature selection process and regression models were integrated with the cross-validation and splitting dataset process to produce more objective and robust outcomes.
- *Step 3- model expansion at urban scale*: the model obtained from the Step 2 was expanded to the urban scale of a medium-sized city, located in North-West of Italy. At this step, the buildings, which were not accurately estimated, were excluded.

The proposed approach could be used by everyone involved in the formulation and optimisation of operation strategy. The methodology is introduced in the sub-sections below.

2.1. Data collection and data integration

The data collection procedure and its main reference sources, (e.g., building stock characterisation and distribution) are fundamental to model the building energy consumption at the urban scale. Although the data collection procedure can be generalised, data and information availability depends strongly on each specific context. The research began with the collection and analysis of the available data of building stock, which affects space heating energy consumption. The proposed methodology (Fig. 1) integrates GIS as a supportive data collection tool, which can join different types of information or datasets by using location as the common feature. For instance, the census datasets consist of demographical and housing information can easily be overlapped to individual buildings which have shape files (Ma & Cheng, 2016).

Since the target is regional and/or local scale, the definition of the building's database is crucial. Table 1 shows the different predictors that principally characterise the heating space energy consumption of buildings with their references. The geometrical data were mostly acquired from the cartographic base using the automatic functions of the GIS tool.

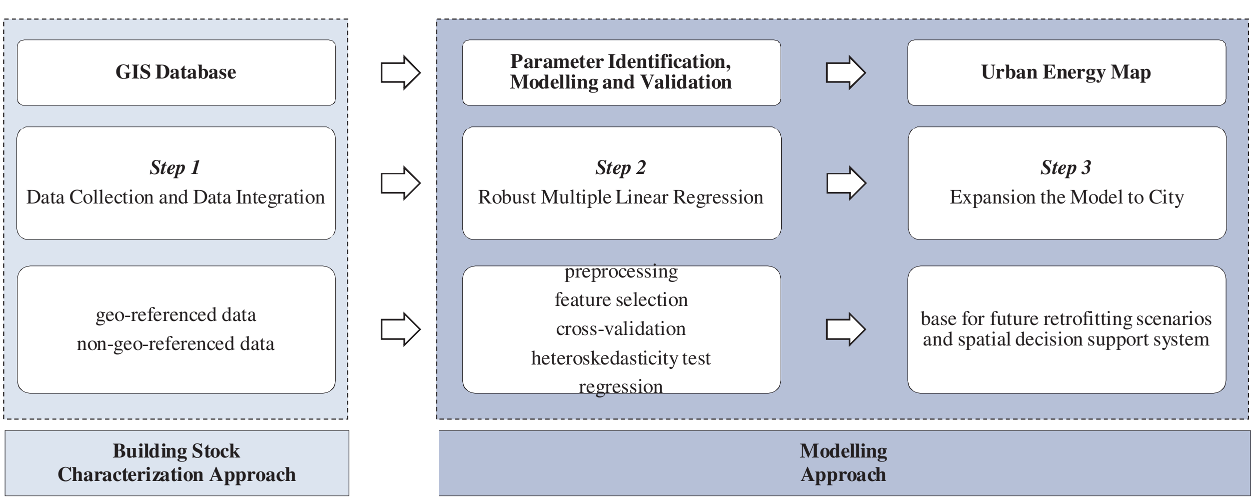

Fig. 1. Proposed methodology framework of the research.

Table 1
Structure of the database and the relative description of the variables.

Data	Raw data	unite	Source of information	Reference
Dispersing Surface	Floor area	m^2	Cartography	Dall'O', Galante, and Pasetti (2012); Fracastoro and Serraino (2011)
	Perimeter		Cartography	
	Height		Derived	
	Contiguity		Derived	
Net floor Area	Gross floor area	m^2	Cartography	Caputo et al. (2013); Fracastoro and Serraino (2011)
	Gross/net ratio		Normative	
Height	Number of floors	m	Cartography	Dall'O', Galante, and Pasetti (2012)
	Floor height		Literature	
Heated Volumes	Net floor area	m^3	Derived	Dall'O', Galante, and Torri (2012), Fracastoro and Serraino (2011)
	Net floor height		Derived	
Number of floors	–	number	Cartography	Dall'O', Galante, & Pasetti (2012)
Perimeter	–	m	Cartography	(Caputo et al., 2013; Dall'O', Galante, & Pasetti, 2012)
Building shape factor	Net floor area	m^{-1}	Derived	Aksoezen, Daniel, Hassler, and Kohler, (2015); Braulio-Gonzalo et al. (2016), Florio and Teissier, (2015), Penna, Prada, Cappelletti, and Gasparella (2015)
	Net floor height			
	Gross floor area			
	Gross/net ratio			
Roof type	–	–	Google earth/In-situ analysis	Dall'O', Galante, & Pasetti (2012)
Period of construction	–	–	ISTAT national census	Aksoezen et al. (2015), Dascalaki, Droutsa, Gaglia, Kontoyiannidis, and Balaras (2010), Florio and Teissier (2015), Theodoridou et al. (2011)
Temperature	Typical meteorological	C°	ARPA	Mastrucci et al. (2014b)
Building occupation ratio	Occupied buildings	%	ISTAT national census	Mutani and Vicentini (2015)
	Empty buildings		ISTAT national census	
Ground floor type	–	–	Cartography	Evans, Liddiard, and Steadman (2015)
Installed power	–	kW	DH Company	–

In this study, the data collected consists in:

First, geo-referenced data: geometrical information on the building stock derived from the digital cartographic technical map of the municipality (perimeter, number of floors, heated volume, and area). Building height (eave height) was determined by multiplying the number of floors by the average height of the floor. The average floor height used depends on the age of a building (Chiara Delmastro, Mutani, & Corgnati, 2016), and consequently, it can be used to calculate gross heated volume. Another interesting approach to determine the height of buildings that it is used in this research just for the validation is to evaluate the height of buildings from the LiDAR Data or DSM (Digital Surface Model) subtracting the DTM (Digital Terrain Model) height data (Normalized Digital Surface Model: NDSM = DSM-DTM), source (Berlin Environmental Atlas, 2014). This process is applicable when the relative data is available.

The prevailing period of construction of a large building stock was extracted from the ISTAT (2011) national Census database, which provides information for each census parcel. This variable implies the typical envelope characteristics of buildings (e.g., roofs, floors and windows) and heating systems efficiencies. According to the Italian national classification, the period of construction can be divided into nine classes characterised by homogeneous features of buildings i.e., age_1: before 1919; age_2: 1919–1945; age_3: 1946–1960; age_4:1961-1970; age_5: 1971 1980; age_6: 1981–1990; age_7: 1981–2000; age_8: 2001–2005; age_9: after 2005. This variable considers the building envelope, such as the percentage of the transparent envelope and a class of U-value ($W.m^{-2}K^{-1}$) for both opaque and transparent surfaces, and the performance of the heating system.

According to Guglielmina Mutani & Todeschi (2017), the Italian periods of construction before 1919–1960 may have an increasing energy consumption. On the other hand, the buildings built during the economic boom period (1961–1980) have higher energy consumption values. Finally, buildings constructed after 1981 show a decreasing energy consumption. This evidently means that Italian building stock is characterised by high-rise energy consumption before the first energy regulation (e.g., Law 373/1976), when any envelope insulation and energy efficient system was required.

Furthermore, from the percentage of the occupied building the building's occupation factor can be identified, which is derived from the ISTAT national census database. Unlike many previous studies, ground-floor typology (R: residential, C: commercial and P: pilotis) was also considered in this study, which is derived from the digital cartographic buildings map of technical departments of the municipality.

The model for determining the space heating consumption of the buildings depends clearly on the surface to volume ratio of the buildings (S/V, dispersing surface/heated volume). This factor represents the non-compactness of the building, and it was determined using GIS, excluding the contiguous surfaces between two heated buildings. In this study, the automatic calculation of the adjacent walls was applied according to Guglielmina Mutani & Vicentini (2013). This procedure permitted the subtraction of this parameter from the gross dispersant surface to obtain the real dispersant surface and also the unheated volumes. Subsequently, the higher dispersant surfaces were considered for typical Italian building archetypes. The surface to volume ratio is classified as Detached House (DH): $S/V \geq 0.8\,m^{-1}$; Terrace House (TH): $0.6 \leq S/V \leq 0.8\,m^{-1}$; Multi-Family House (MFH): $0.4 \leq S/V \leq 0.6\,m^{-1}$; Apartment Block (AB): $S/V \leq 0.4\,m^{-1}$ (TABULA, 2012).

Finally, the roof type (G: gable, F: flat) and the mean daily climate temperature were also added to other information.

On the other hand, the non-georeferenced necessary energy consumption data of buildings were collected, such as measured monthly energy consumption for DH with its installed power information. When the measured real data is not available (in many countries), other methodologies such as building simulation tools can be used to determine the energy consumption.

The database was updated monthly in the two heating seasons 2011–2012 and 2014–2015. The monthly DH energy consumption was given by the district heating company. In this step, these kinds of data were geo-referenced and associated to each building entity using Google maps and in-situ analyses (sometimes it is possible to perform this operation automatically based on buildings address and geocoding).

Finally, in order to create a supportive and strong GIS database, it was necessary to integrate all the data collected at two levels: (a) the individual building level (e.g., the base floor area, the perimeter, the gross volume and eaves height of the buildings, the external surface, DH data) and (b) the ISTAT census cartography level (e.g., the main construction period and the average building occupation percentage).

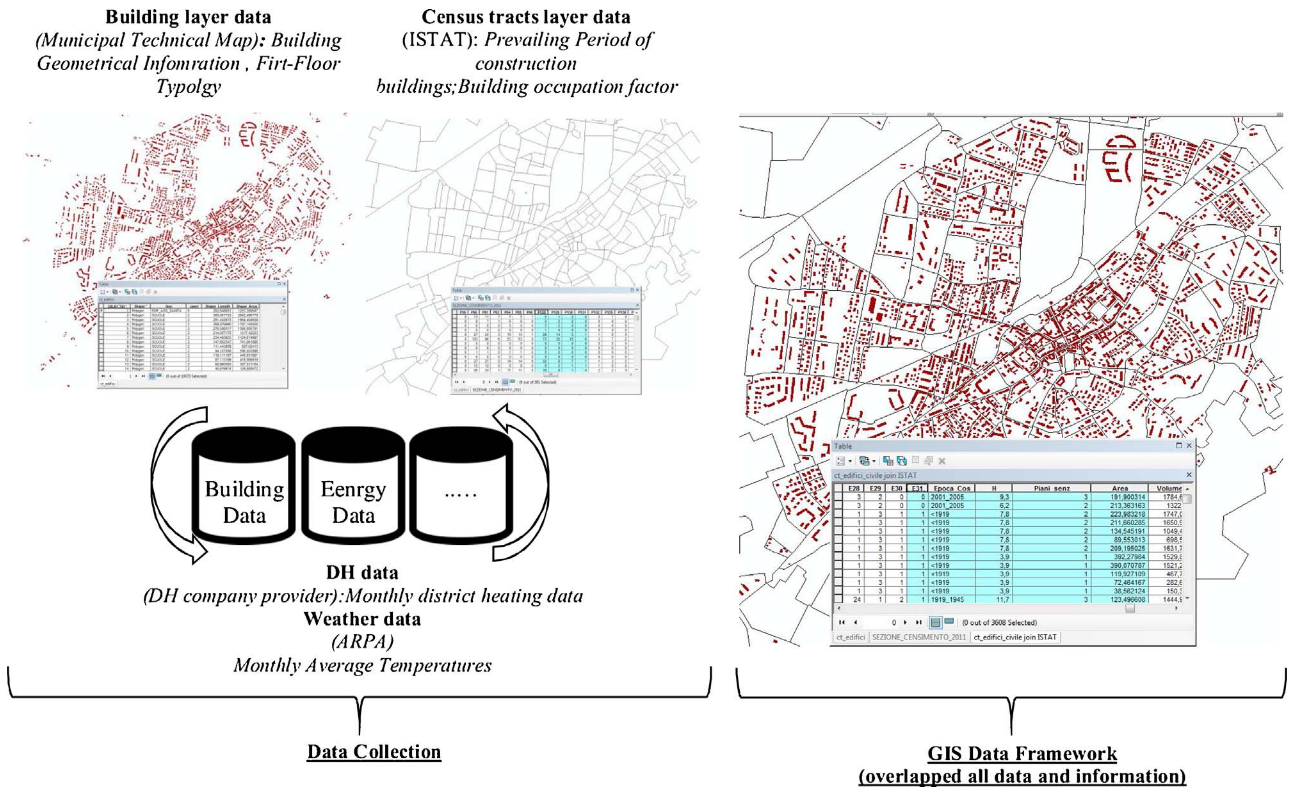

Fig. 2. Conceptual scheme of GIS data platform.

Fig. 2 summarises the procedure of GIS database framework, which consists in all data and information describing each building as a basis for estimating the related energy consumption for space heating. This is a key stage since it is a foundation of the entire calculation process and other substantial actions. In this phase, the stakeholders' involvement should be integrated as well to obtain the existing data and determine relevant sustainable objectives for future planning (Pelzer, Arciniegas, Geertman, & Lenferink, 2015)

2.2. Parameter identification, modelling and validation

The statistical methodology based on a geospatial multiple linear regression model was applied at the urban scale. In the building sector, many different statistical bottom-up methods exist (Torabi Moghadam, Delmastro et al., 2017). From a comparison of regression analysis, decision tree and neural networks it emerged that these methods are comparable in predicting energy consumption with a quite small difference in terms of errors (Tso & Yau, 2007). Using regressions helps in easing the usage and interpretation of the parameters introduced in the analysis (Mastrucci, Baume, Stazi, & Leopold, 2014b). One of the most popular regression algorithms is the multiple linear regression (MLR). Specifically, many researchers have used the MLR method with the aim of predicting energy consumption using a range of different predictors (Section 1). These techniques determine the strength of the relationship between one dependent variable used for numerical prediction. Moreover, the regression models are highly rated due to their simple application (Bassani, Catani, Cirillo, & Mutani, 2016).

Eq. (1) describes a multiple linear regression model with more than one explanatory variable:

$$y = I + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \quad (1)$$

Where:

- y is the output variable;
- I the general model intercept;
- β_i the regression coefficient ($i = 1,2,\ldots,p$);
- x_i the input variables ($i = 1, 2,\ldots,p$)
- ε the random effect (to measure the random difference between the y variables for all buildings and the corresponding prediction for a specific building) and remaining errors.

Pre-processing: There are usually features with missing values in the dataset. In this study, they were replaced by the mean value of that attribute (Han, Kamber, & Pei, 2011). Moreover, a detection procedure was performed to check the presence of outliers, which were subsequently removed.

Feature selection: Many variables for estimating the energy consumption in statistical models could be irrelevant or redundant; therefore, key variables selection is an important step in achieving more accurate predictions (Hsu, 2015). Indeed, those redundant variables lead to reduce the model performance. Hence, a proper feature selection process and identification of the correlation that measures the degree of association between two attributes is fundamental.

Computer science offers a number of different approaches for feature selection. This study used Akaike Information Criterion (AIC) to select features for the linear regression (Akaike, 1973). It selects the attribute with the smallest standardised coefficient in each iteration, removing it and performing another regression (Deshpande, 2012). In order to robustly select the features, the removed correlation variables operation was applied in the proposed methodology to remove the high correlated variables. Correlated attributes are usually removed since they behave in a similar manner and they have the same impact in prediction calculations; therefore, keeping those attributes is redundant and time-space consuming.

Validation: With the aim of assuring prediction accuracy and proper model characteristics, assumptions at the basis of the regression model should be carefully verified. Validation of the statistical model

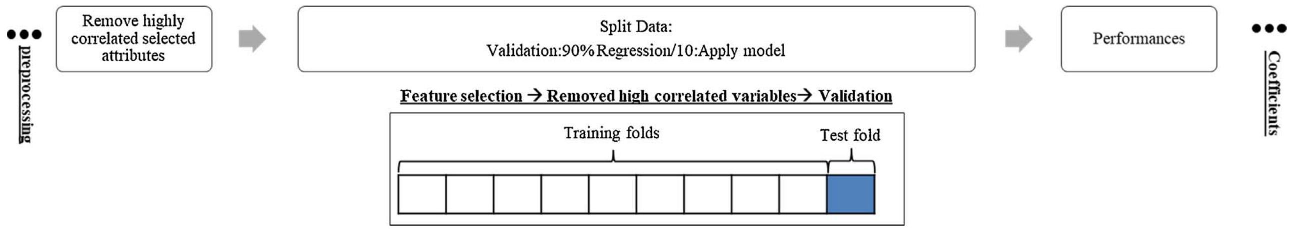

Fig. 3. Validation Process.

can be internally performed, using techniques such as cross-validation.

This study compares the performance of the representative regression approach in these two aspects. To verify the energy demand estimated, the actual energy consumption of the target area was compared with the calculated energy demand. Details are as follows:

First, very high correlated variables were removed before applying the regression model. Then, the dataset was divided into two subsets, training and testing, to assess the model performance. In this way, a model was first trained on a 90% of the dataset, and then that model was applied to the testing partition to validate and identify the reliability of the methodology. The performance showed the difference between training and testing set estimates in terms of fitting (coefficient of determination, R^2) and prediction error (Mean-Root-Squared-Error, MRSE) (Fig. 3).

Second, the feature selection procedure and regression were validated by cross-validation approach to achieve more stable results and to avoid the high risk of overfitting (Han et al., 2011). In the cross-validation process, the dataset was divided into the ten same size folds. A single fold was considered as the testing data set and the remaining nine subsets were used as training data set. The cross-validation process was then repeated ten times, with each of the ten subsets used exactly once as the testing data. By applying cross-validation, the model was able to compare the features selection strategies and identify the regression algorithms (Fig. 4).

The results show (Section 3.2.) that both approaches produce similar performances and coefficients.

Heteroskedasticity test: Homoskedasticity is a significant assumption in regression analysis (Hayes & Cai, 2007). Therefore, the appropriate diagnostics (e.g., Breusch–Pagan test and White test) were performed to carefully check the multiple linear regression model assumptions, verifying the correct specification and accuracy of the model prediction. The homoscedasticity (assumption of homogeneous variance for residuals) was tested through the scatter plot of residuals (or the squared residuals) against predicted values. The initial presence of heteroscedasticity was reduced thorough heteroskedasticity-consistent standard errors (HCSE) (or robust errors) in the Linear Regression Model (White, 1980), which allowed the fitting of a model that contains heteroscedastic residuals. The software SPSS was used additionally for this scope. Finally, no significant heteroskedasticity issues were detected as residuals were randomly scattered.

2.3. *Model expansion into the city*

As previously stated, the goal of the study was the evaluation of a simplified energy consumption model for space heating at the urban scale. All the considered variables needed to be extendable and available for the whole city. Once the statistical analysis using building function was performed, the results were mapped across the city. Since the sample of dataset includes a specific range of heated volume, the buildings that were lower and much higher than this value were excluded. The database information quality and, accordingly, the geo-referenced model can be continuously improved (Ascione et al., 2013). This methodology is flexible enough to add variables according to the data availability and purpose of the analysis, such as occupants' behaviour or buildings renovation ratio information.

3. Case study

The residential building stock of the city of Settimo Torinese, a medium-sized urban area, located in the North-West of Italy in the continental temperature climatic zone, was chosen as a case study to develop and test the methodology. The building stock characterisation, energy consumption profiles and the dataset used are presented in this section. The city is composed of 300 census sections and about 3600 residential buildings with 47,831 inhabitants. The GIS used for this study was ArcGIS 10.3. The total heated volume of the residential buildings is equal to 8.55 Mm^3.

3.1. *Data Set*

The measured energy consumption data available consist of monthly records of DH energy consumption for the residential sector for the heating seasons 2011–2012/2014-2015, with respectively 2597 and 2342 HDD at 20 °C. The data were pre-processed and carefully analysed before being put into the model. In this study, the monthly data energy consumption was elaborated first for each month (from the exact first

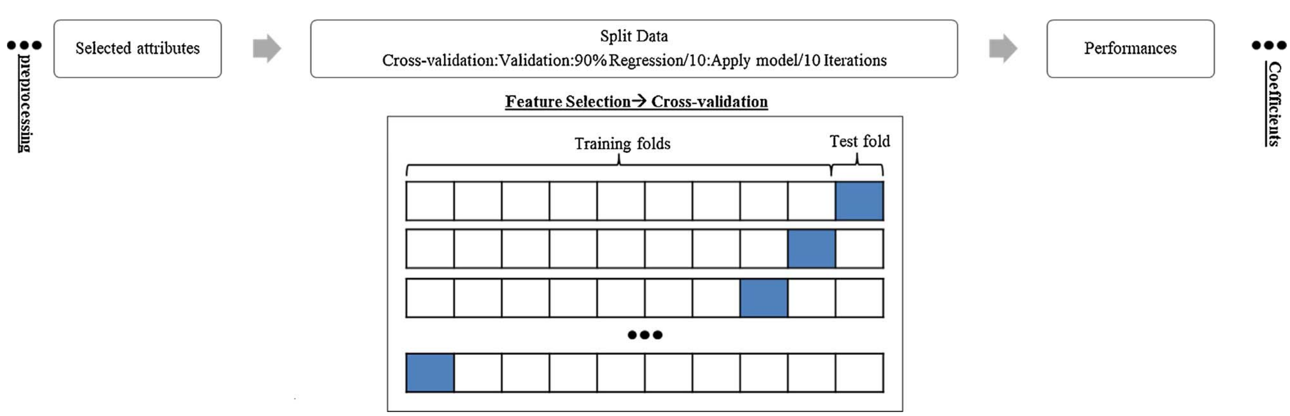

Fig. 4. Cross-validation Process.

day of the month to the last day), and then it was divided by the number of days of each month in order to have a normalised daily energy consumption.

The monthly measured DH energy consumption data of 290 residential buildings were used for the analysis (the local DH Company provided the total number of mixed typology buildings connected to the DH network, which was 350). Almost 50% of the data were excluded before creating the model due to the difficulty of associating the address of the building to its heated volume and due to some differences between the GIS calculated heated volume and the heated volume given by the DH company. Moreover, the GIS data was optimised by comparing the volume of the buildings and the measured volume provided by the DH company, with an acceptable limit of 15% difference. 165 buildings out of a total number of 290 residential buildings connected to the DH were successfully associated and geo-referenced with the polygon of each building on the GIS map. This operation was performed manually using the Google Maps platform and insitu analysis to identify the relative buildings.

In order to create a supportive database and to have enough available data, these 165 residential buildings were considered over 7 months (from October to April) of two heating seasons,taking into account the residential typology; ground floor typology (C: commercial, R: residential, P: pilotis); occupation factor; number of floors; geometrical information of each building (area, perimeter, heated volume, height); the surface to heated volume ratio S/V. Regarding the period of construction, a linear correlation can be hypothesised by dividing the buildings into three clusters, with higher energy consumption for the buildings built from 1961 to 1980 (n. data 1344), lower consumption for the older ones (n. data 678) and for the newer ones (n. data 245).

As a complement to the data set provided directly by the city, daily records of the outdoor air temperature of these two years were made available through the Regional Agency for Environmental Protection (ARPA). The acquired input data set is summarised in Table 2.

By using the GIS tool, it was possible to represent how these variables are distributed in the city of the case study. Fig. 5 shows the urban block distribution according to the year of construction, ground floor type, and S/V as reported in Table 2. Among all the surveyed buildings, 82% of the buildings are characterised by the ground floor as a residential type, and just 14% have a commercial ground floor. Regarding the period of construction, 47 buildings were built in the category of age_3 (1946–1960), accounting for 29%, while 84 buildings were constructed in age_4 (1961–1970), accounting for 51%. This fact is very much proportionate with the reality of the entire building stock and it means that the building stock is mostly characterised by buildings built before the first Italian energy regulation. Fig. 5 shows that the MFH typology is the most widely used, accounting for 68%, followed by TH and AB, accounting for 16%. This fact indicates that the single SFH houses are not connected to the DH network. The results of the MLR analysis strongly depend on the sample of buildings analysed.

3.2. *Results and discussion*

This section discusses the outcomes of the regression analysis and the spatial distribution of the annual energy consumption. The data mining software RapidMiner 7.1 was employed to analyse and model the energy consumption of residential buildings in Settimo Torinese. Furthermore, the influence of every single variable on energy consumption was analysed.

Fig. 6 illustrates the scatter plots for each of the variables with respect to the daily space heating energy consumption. As is shown, there are correlations between the energy-use for space heating and some of the selected variables of buildings, such as the perimeter, heated volume, installed power and area. It should be noted that in this analysis not all the variables were taken into account, due to the fact that data about them were not available, such as the level of renovation of the buildings and the adoption of renewable energy technologies. This may have a slight influence on the dispersion of the results. Moreover, the results of this first analysis are affected by the number of buildings analysed. For instance, for each period of construction and each value of surface to volume ratio (S/V), the number of buildings is not the same. The sample of buildings connected to the DH network mainly consists of big apartment blocks built in 1961-70.

In Table 3 the correlation coefficients between the daily heating energy consumption (kWh) and different single variables of the available sample of data are shown. Some correlations appear to be very rational and intuitive, such as perimeter, surface, area, height, heated volume, installed power, occupation ratio, and air temperature. A particularly interesting result is that some correlations seem to be controversial, taking into account basic thermosphysics of buildings, such as S/V ratio.

Table 3 reports that the S/V ratio correlation is negatively correlated with space heating energy consumption. This fact is explained by the strong correlation between the variables heated volumes and S/V. Due to building geometry, high values of S/V are generally related to small buildings (e.g., semi-detached houses) with low energy consumption, while small S/V values are related to large sie condominiums with higher energy consumption. However, a positive correlation is expected between the S/V and the specific daily energy consumption (kWh/m^2 or kWh/m^3) (see Fig. 7).

The correlations of each of these variables on energy-use for space heating were analysed and a MLR was modelled identifying the best coefficient of determination (R^2).

Table 2
Input sample dataset indicating minimum, maximum, average and standard deviation values.

	Sample of data (165 residential buildings)				Entire Building stock (3600 residential buildings)			
Numerical Input Variable	Min	Max	SD	Ave.	Min	Max	SD	Ave.
Dispersing Surface (m^2)	802.28	11678.66	1618.38	2822.98	190.02	13910.40	1336.19	1384.72
Area (m^2)	147.46	1688.17	270.62	502.89	6.10	2953.72	221.70	230.97
Height (m)	7.00	27.00	5.51	16.49	3.10	27.20	4.35	8.06
Heated volumes (m^3)	1504.11	40178.46	6205.36	8609.83	20.76	51063.76	3902.08	2370.95
Number of floors (number)	2.00	8.00	1.66	4.00	1.00	8.00	1.28	2.35
Perimeter (m)	49.24	348.84	46.91	107.65	9.90	498.90	41.04	66.22
S/V_{real} (m^{-1})	0.34	0.78	0.09	0.51	0.32	2.39	0.31	0.93
Temperature (C°)	−0.30	12.85	4.27	6.81	−0.30	12.85	4.27	6.81
Building occupation ratio (%)	0.00	1.00	0.06	0.95	0.00	1.00	0.20	0.85
Installed power (kW)	50.00	1000.00	126.23	196.43	n.d.	n.d.	n.d.	n.d.
Nominal Input Variable	Least	Most			Least	Most		
Period of construction	> 1919 (1)	1961–1970 (84)	–	–	2001–2005 (82)	1946–1960 (1028)	–	–
Ground floor type	P (6)	R (135)	–	–	P (48)	R (2962)	–	–
Roof type	F (5)	G (160)	–	–	n.d.	n.d.	–	–

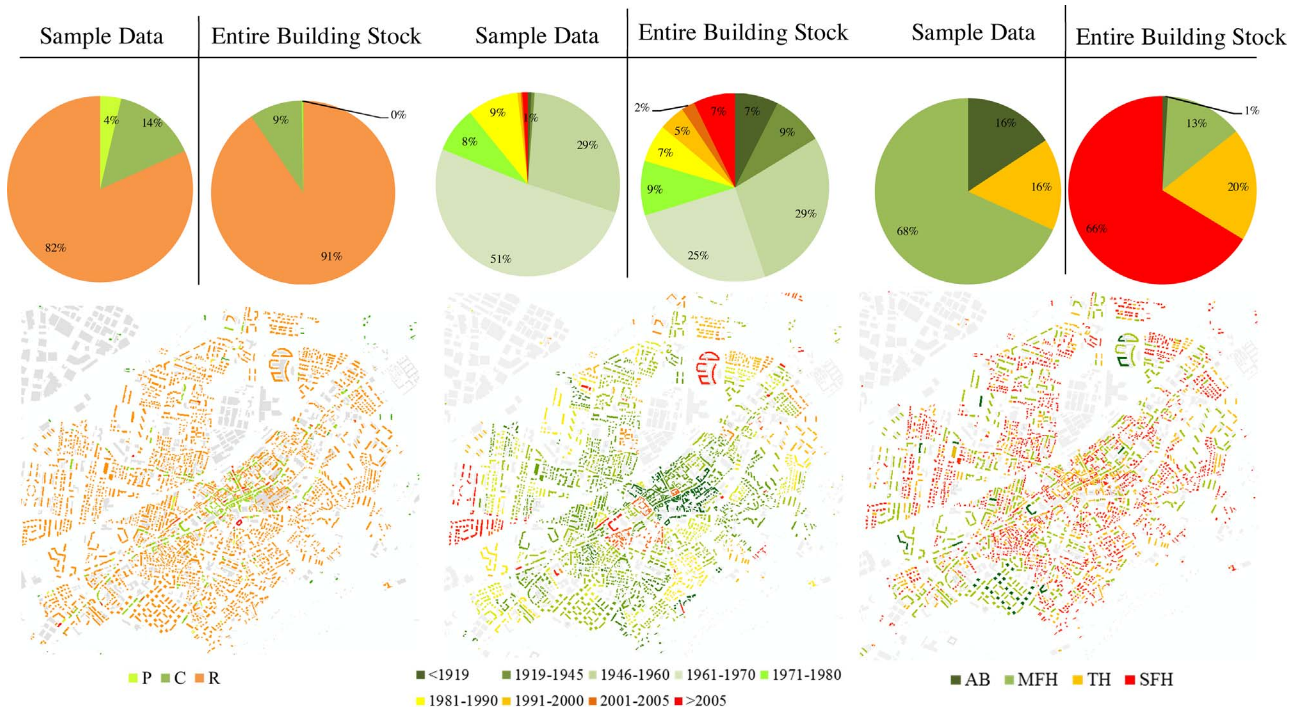

Fig. 5. The comparison between the sample of input data (165 buildings) on the left and the entire building stock (3608 buildings) on the right for ground floor typology, period of construction and S/V respectively.

The estimated coefficients, standard error, and p-values (the probability for a given statistical model) for energy consumption robust linear regression are shown in Table 4 (Model A) and Table 5 (Model B), respectively.

All numerical predictors strongly rejected the null hypothesis for a value of ≤ 0.05 indicating that the estimated intensities are statistically significant. Moreover, the code factor (*) is based directly on p-value, where more stars mean the variable is more significant. In the following paragraphs, the predicted values from the model will be compared to the real data provided by the city.

In the analysis of variance, known as the ANOVA, the F-statistic = 1132.478 (p-value < 0.0001), indicating the results of the regression model is satisfactory.

The correlation coefficient between the predicted and observed values, measured by the mean squared error (MSE) and squared correlation (R^2), are shown Table 4 (Model A) Table 5 (Model B). In the second model (B), the data about the installed power and the type of roof were excluded because they are not available for the entire building stock. As a result, the performance of the second model (B) was slightly diminished with respect to Model A. The coefficients of the regression equation for each variable seem to have the expected trend:

- the highest coefficient for the buildings built in 1961–1980, as they consume more energy, and lower consumption for the older buildings due to their lower percentage of transparent envelope and higher structure thickness and for the newer buildings due to their thermally insulated envelopes;
- a positive coefficient with gable roofs, as the dispersant surface and heated volume is greater with higher energy consumption;
- a negative coefficient for the commercial typology of the ground floor, as it is usually heated autonomously;
- a positive coefficient for the pilotis, as the floor disperses more heat to the outside environment;
- a positive coefficient for bigger buildings, as the high heated volume, number of floors, and perimeter leads to higher energy consumption;
- a positive coefficient for the installed power of the heat exchanger as it depends on the dimensions and the level of energy efficiency of buildings;
- a negative coefficient for the outdoor air temperature, as with lower air temperatures the energy consumption increases;
- a positive coefficient for the occupation factor as the buildings consume more if they are utilised and occupied.

In Figs. 8 and 9, it is possible to notice the good correlation of the models and the correspondence between measured and predicted annual energy consumption. The colours of the points indicate the heated volumes of the buildings and it is shown that the model does not work for massive buildings.

The coefficients of determination R^2 for the two models are of 0.84 and 0.80, meaning a high-performance correlation even without the installed power and the type of roof variables. The precision of a model depends on the availability and the accuracy of data and, mostly, on the typology of the data sample.

Finally, applying the energy regression model to the entire building stock area in Settimo Torinese, through the GIS framework, the energy consumption spatial distribution was represented, creating a visual map. The total annual energy consumption (kWh/m^3) for each individual building is shown in Fig. 10. Moreover, the spatial results of this study help to identify in which neighbourhoods the energy consumption is mostly concentrated. Since the sample of dataset includes heated volumes greater than 1500 m^3 (see Table 2), the building volumes less than this value were excluded (grey polygons). The results show that the residential buildings constructed before 1980 have a mean annual energy consumption of 27.47 kWh/m^3. Indeed, the buildings located in the historic city centre are one of the largest annual energy consumers, as is shown by the dark colours on the map (about 47.70 kWh/m^3). Those constructed after 2005 show a decrease in the heating energy consumption of 10%.

4. Conclusion and future developments

Urban energy efficiency plays a crucial role in the implementation of energy policies in the context of low-carbon cities and smart cities.

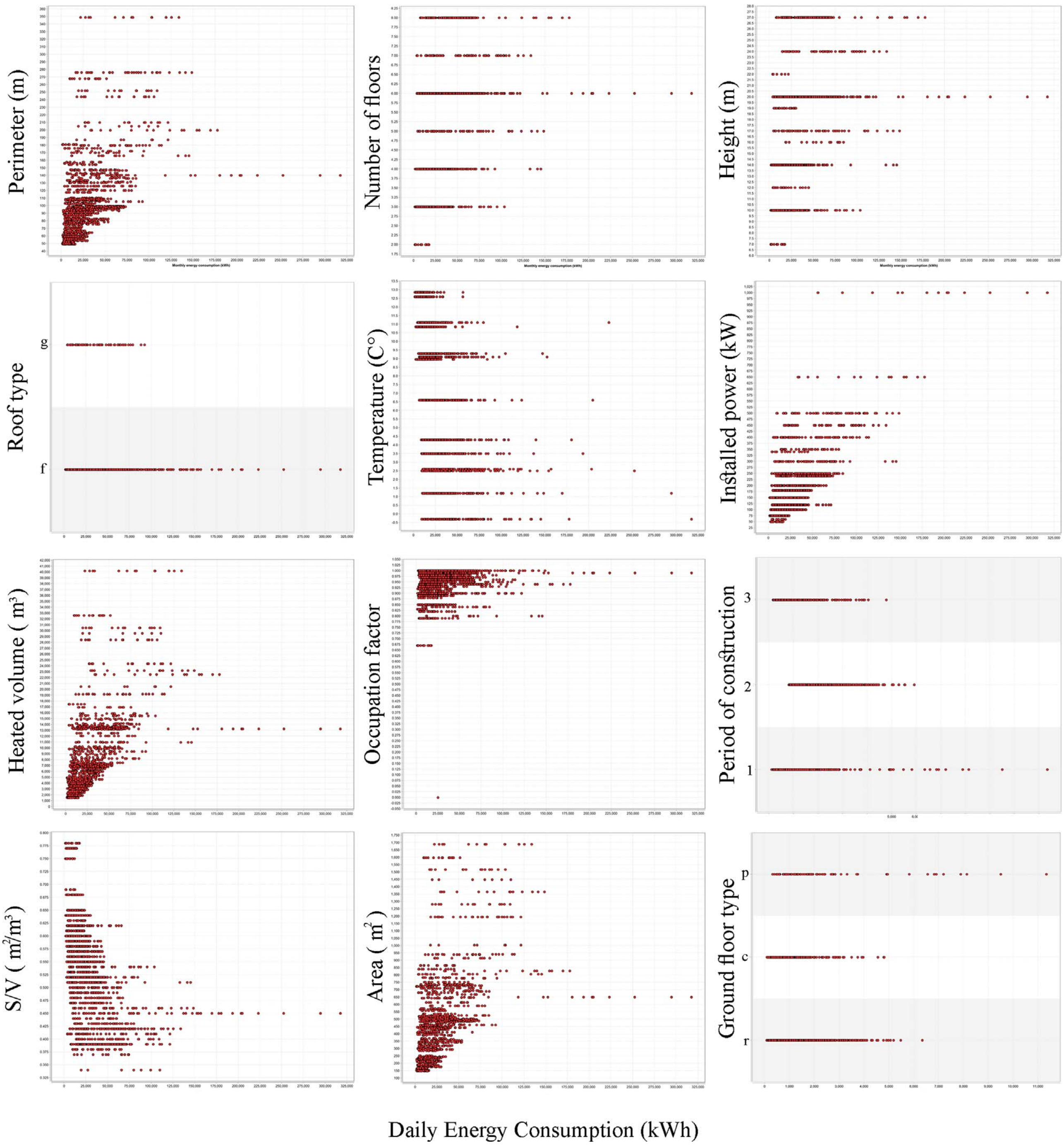

Fig. 6. Scatter plot: DH energy consumption and variables.

The research presented in this paper demonstrates that urban actors (e.g., energy planners, local and public administrators and other stakeholders) can be supported by an appropriate urban scale energy consumption model in delivering the most effective strategies. The whole procedure was successfully tested at the city level for Settimo Torinese and validated. Given available data, the proposed methodology can be applied to any similar city and also to other kinds of energy usage (e.g., electricity, cooling).

Analysis of the available data regarding the existing building stock is crucial to understand the measurements needed to achieve energy efficiency or lower gas emissions in the construction sector. This methodological approach was chosen to determine the energy consumption for space heating of a residential building stock at the urban scale. A framework combining statistical analysis with GIS-based techniques was employed to identify the most appropriate variables influencing energy consumption, using detailed measured building data. Moreover, GIS tools were used to support both the geometrical building stock characterisation and the energy assessment process. The spatial distribution of urban energy consumption in 2D and 3D visualisations represents a useful tool – a new Spatial Decision Support System (SDSS) – to facilitate decision-making process in order to define a variety of urban transition energy policies according to local conditions in a smart cities context.

The Multiple Linear Regression (MLR) analysis applied in this study has highlighted the variables most related to energy consumption, as follows: period of construction, heated volume, type of ground floor,

Table 3
Correlations of the selected variables with energy consumption (kWh).

Attributes	Daily Energy Consumption (kWh)	Association
Numerical variables		
Installed power (kW)	0.789	Strong positive association
Dispersing Surface (m^2)	0.631	Strong positive association
Heated volumes (GIS) (m^3)	0.619	Strong positive association
Perimeter (m)	0.544	Strong positive association
Area (m^2)	0.541	Strong positive association
S/V_{real} (m^{-1})	−0.528	Strong negative association
Height (m)	0.440	Weak positive association
Number of floors (number)	0.434	Weak positive association
Temperature (C°)	−0.383	Weak negative association
Building occupation ratio (%)	0.161	Weak positive association
Non-numerical variables		
Ground-floor type R	−0.074	Weak negative association
Ground-floor type C	−0.057	Weak negative association
Ground-floor type P	0.261	Weak positive association
Roof type F	0.045	Little positive or no association
Roof type G	−0.045	Little negative or no association
Period of construction(Period 1: < 1960)	−0.009	Little negative or No association
Period of construction(Period 2:1961–1980)	0.003	Little positive or No association
Period of construction(Period 3: > 1981)	−0.019	Little negative or No association

occupation factor, air temperature, type of roof and the installed heating power. In case two variables are not available, such as the type of roof and the installed power, the model reaches a determination coefficient of 0.8, but only for buildings of a limited heated volume. Since the building stock is constituted mostly by large condominiums, the models have a higher margin of error on low volume buildings; for the same reason, this model should be utilised only for buildings connected to the district heating network. It is important to remember that the level of uncertainty for a model of this type is strongly dependent on the characteristics of the sample.

Finally, this model makes it possible to evaluate an average consumption of residential buildings for space heating and it can be used to spatially distribute the energy demand, supply and emissions at the urban/local scale.

One of the limitations of this study regards microclimate effects, which were not taken into account in the presented method. In fact, a microclimate model that would give a single value for the whole city for air temperature would not significantly improve the results of the current model presented in this paper. The current paper aimed to provide a statistical relationship for the energy demand. Running a microclimate model would mean including a dynamic and full-year simulation that would be beyond the context of this study. Instead, this research will continue by using a meteorological model developed by Mauree, Coccolo, Kaempf, and Scartezzini (2017) on the city of Settimo Torinese and couple it with the CitySim (Mauree, Lee, Naboni, Coccolo, & Scartezzini, 2017; Robinson et al., 2009) to evaluate the influence of the microclimate on the energy demand of the buildings. In this case, the current MLR-GIS model will be a basis for the future integration evaluating different energy saving scenarios.

Within the EEB project, further development and improvement will be necessary, including other new databases regarding natural gas measured consumption (for a larger part of the city) and regarding building stock characterisation (for each building). Further research will also be necessary for improving this model by taking into account additional elements, such as solar exposure, microclimate effects and urban variables of the surrounding environment (Delmastro, Mutani, Pastorelli, & Vicentini, 2015; Mutani & Todeschi, 2017).

In conclusion, this study represents the first step towards the goal of

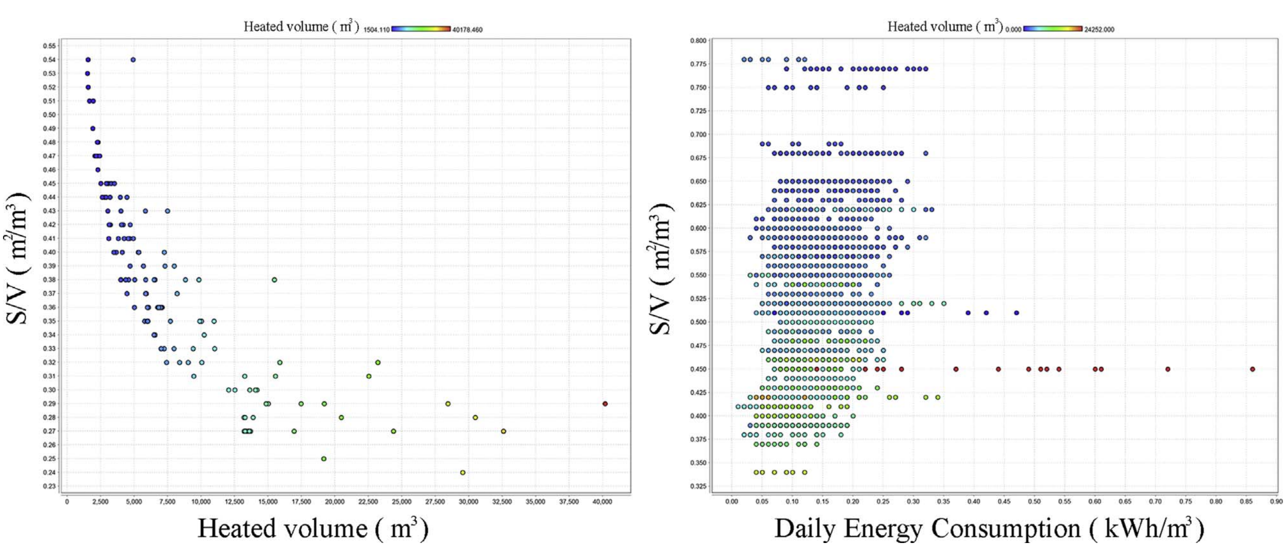

Fig. 7. Relationship between S/V_{real} (m^2/m^3) with heated volume (m^3) and daily energy consumption (kWh/m^3).

Table 4
Linear regression results considering all the influencing variables (Model A).

Attribute	X-Validation				Validation			
	Coef.	Std.error	t-stat	p-value	Coef.	Std.error	t-stat	p-value
Period of construction (< 1960)	−24.32	17.29	−1.41	0.159	−31.39	20.48	−1.53	0.125
Period of construction (1981)	4.87	27.96	0.17	0.861	9.50	33.09	0.29	0.773
Roof = F	−203.95	57.04	−3.58	3.58E-04****	–	–	–	–
Roof = G	–	–	–	–	226.09	59.66	3.79	1.57E-04 ****
Ground floor = R	−14.57	21.25	−0.69	0.493	−14.26	25.53	−0.56	0.576
Ground floor = C	−54.28	22.97	−2.36	0.018**	−55.13	27.28	−2.02	0.043**
Ground floor = P	68.70	49.14	1.40	0.162	69.42	62.74	1.11	0.268
Perimeter (m)	2.29	0.29	7.86	6.11E-15****	3.03	0.36	8.48	0****
Number of floor (Eaves)	54.82	5.78	9.49	0****	76.98	6.95	11.08	0****
Heated volumes-GIS (m^3)	0.04	0.00	13.03	0****	0.03	0.00	7.39	2.28E-13****
Installed power (kW)	1.77	0.12	14.45	0****	1.87	0.15	12.91	0****
Monthly average temperatures (C°)	−63.61	1.86	−34.15	0****	−63.96	2.20	−29.11	0****
Occupation factor	812.97	134.05	6.06	1.57E-09****	865.81	154.45	5.61	2.43E-08****
(Intercept)	−430.42	Infinity	0.00	1	−790.88	Infinity	0.00	1
Performances	root_mean_squared_error: 217.598 ± 22.885 (mikro: 218.930 ± 0.000) squared_correlation: 0.830 ± 0.036 (mikro: 0.829)				root_mean_squared_error: 207.798 ± 0.000 squared_correlation: 0.835			

*Signific.code: < 0.5; ** Signific.code: < 0.01; ***Signific.code: < 0.001.

Table 5
Linear regressions considering the influencing variables that are expandable at the urban scale, removing installed power and the roof type (Model B).

	X-Validation				Validation			
Attribute	Coef.	Std.error	t-stat	p-value	Coef.	Std.error	t-stat	p-value
Period of construction (< 1960)	15.86	17.77	0.89	0.372	10.24	19.23	0.53	0.59
Period of construction (1961–1980)	17.74	16.81	1.06	0.291	18.11	18.19	0.99	0.32
Period of construction (> 1981)	−33.19	28.71	−1.16	0.247	−28.38	31.22	−0.90	0.36
Ground floor_C	−25.88	23.58	−1.10	0.272	−28.45	25.28	−1.12	0.26
Ground floor_P	27.00	50.03	0.54	0.589	23.4	57.90	0.40	0.68
Perimeter (m)	5.77	0.29	19.73	00****	6.44	0.33	19.25	0****
Number of floor (Eaves)	108.43	5.85	18.54	0****	126.62	6.42	19.69	0****
Heated volumes-GIS (m^3)	0.03	0.00	9.97	0****	0.021	0.00	6.15	9.06E-10****
Monthly average temperatures (C°)	−63.34	1.91	−33.11	0****	−63.52	2.05	−30.96	0****
Occupation factor	885.29	137.48	6.44	1.49E-10****	917.00	146.48	6.25	4.79E-10****
(Intercept)	−776.06	Infinity	0.00	1	−873.41	Infinity	0.53	1
Performances	root_mean_squared_error: 234.668 ± 28.713 (mikro: 236.333 ± 0.000) squared_correlation: 0.803 ± 0.058 (mikro: 0.802)				root_mean_squared_error: 216.811 ± 0.000 squared_correlation: 0.826			

*Signific.code: < 0.5; **Signific.code: < 0.01; ***Signific.code: < 0.001.

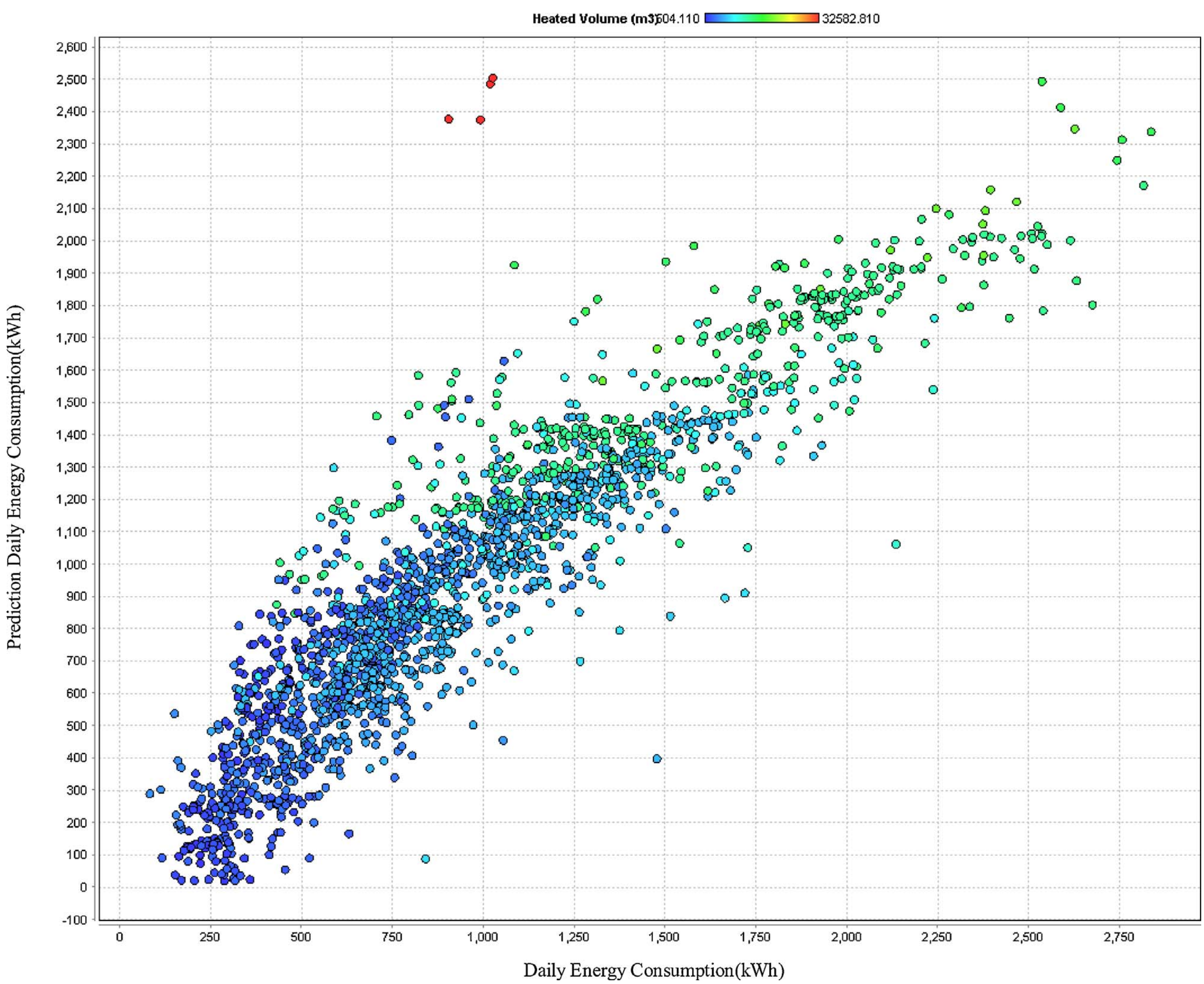

Fig. 8. the comparison between the predicted and the real daily energy consumption (kWh), considering installed power and the roof type (X-validation).

developing future urban energy scenarios through the development of a Spatial Decision Support System (SDSS) (Lombardi, Abastante, & Moghadam, 2017; Torabi Moghadam, Lombardi et al., 2017)

Currently, the work is in progress for developing a powerful Multi-Criteria Spatial Decision Support System (MC-SDSS) visualisation tool to structure group decision-making problems and managing conflicting aspects. The future MC-SDSS will allow the development of scenarios taking into account not only energy consumption but also socio-economic and environmental aspects of future sustainable cities.

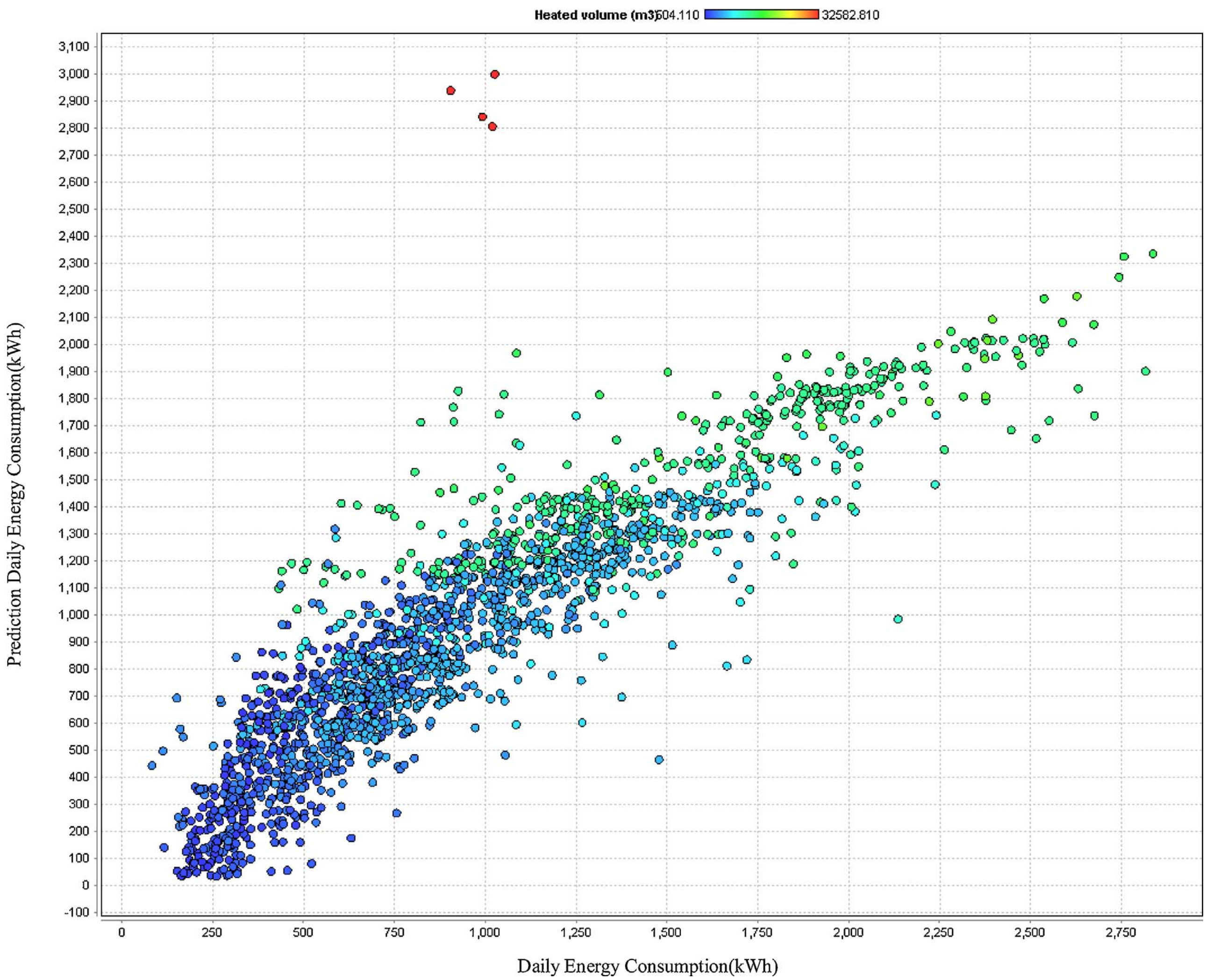

Fig. 9. the comparison between the predicted the real energy consumption (kWh), removing installed power and the roof type.

Fig. 10. Urban Energy Map (2D and 3D); energy consumption for space heating (kWh/m3/y).

Acknowledgements

This work was funded by an on-going National Cluster Smart City & Communities project, named "EEB-Zero Energy Buildings in Smart Urban Districts", National Operational Programme for Research and Competitiveness 2007–2013 (PON R & C), CTN01_00034_594053. Authors would like to also thank Dr. Dasaraden Mauree and the three anonymous reviewers for their constructive and precious comments,

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd international symposium on information theory* (pp. 267 281)

Aksoezen, M., Daniel, M., Hassler, U., & Kohler, N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings, 87*, 74 86. http://dx.doi.org/10.1016/j.enbuild.2014.10.074.

Ascione, F., De Masi, R. F., de Rossi, F., Fistola, R., Sasso, M., & Vanoli, G. P. (2013). Analysis and diagnosis of the energy performance of buildings and districts: Methodology, validation and development of Urban Energy Maps. *Cities, 35*, 270 283. http://dx.doi.org/10.1016/j.cities.2013.04.012.

BPIE (2011). *Europe's buildings uneder the microscope*.

Bassani, M., Catani, L., Cirillo, C., & Mutani, G. (2016). Night-time and daytime operating speed distribution in urban arterials. *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 56 69. http://dx.doi.org/10.1016/j.trf.2016.06.020.

Berlin Environmental Atlas (2014). *Building and vegetation heights.* Retrieved 1 April 2017, from http://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/ed610_03.htm.

Braulio-Gonzalo, M., Juan, P., Bovea, M. D., & Ruá, M. J. (2016). Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference. *Environmental Modelling & Software, 83*, 198 211. http://dx.doi.org/10.1016/j.envsoft.2016.05.018.

COM n.882 final (2011). *Open data An engine for innovation, growth and transparent governance.* European Commission

Cajot, S., Peter, M., Bahu, J.-M., Guignet, F., Koch, A., & Maréchal, F. (2017). Obstacles in energy planning at the urban scale. *Sustainable Cities and Society, 30*, 223–236. http://dx.doi.org/10.1016/j.scs.2017.02.003.

Caldera, M., Corgnati, S. P., & Filippi, M. (2008). Energy demand for space heating through a statistical approach: Application to residential buildings. *Energy and Buildings, 40*(10), 1972 1983. http://dx.doi.org/10.1016/j.enbuild.2008.05.005.

Caputo, P., & Pasetti, G. (2015). Overcoming the inertia of building energy retrofit at municipal level: The Italian challenge. *Sustainable Cities and Society, 15*, 120 134. http://dx.doi.org/10.1016/j.scs.2015.01.001.

Caputo, P., & Pasetti, G. (2017). Boosting the energy renovation rate of the private building stock in Italy: Policies and innovative GIS-based tools. *Sustainable Cities and Society, 34*, 394 404. http://dx.doi.org/10.1016/J.SCS.2017.07.002.

Caputo, P., Costa, G., & Ferrari, S. (2013). A supporting method for defining energy strategies in the building sector at urban scale. *Energy Policy, 55*, 261–270. http://dx.doi.org/10.1016/j.enpol.2012.12.006.

Carozza, M., Mutani, G., Coccolo, S., & Kaempf, J. (2017). Introducing a hybrid energy-use model at the urban scale: The case study of Turin (IT). *3rd IBPSA-Italy conference* ISSN 2531-6702.

Chalal, M. L., Benachir, M., White, M., & Shrahily, R. (2016). Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. *Renewable and Sustainable Energy Reviews, 64*, 761–776. http://dx.doi.org/10.1016/j.rser.2016.06.040

Cheng, V., & Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. *Environmental Modelling and Software, 26*(10), 1186–1198. http://dx.doi.org/10.1016/j.envsoft.2011.04.005

Dall'O', G., Galante, A., & Pasetti, G. (2012). A methodology for evaluating the potential energy savings of retrofitting residential building stocks. *Sustainable Cities and Society, 4*(1), 12–21. http://dx.doi.org/10.1016/j.scs.2012.01.004.

Dall'O', G., Galante, A., & Torri, M. (2012). A methodology for the energy performance classification of residential building stock on an urban scale. *Energy and Buildings, 48*, 211 219. http://dx.doi.org/10.1016/j.enbuild.2012.01.034

Dascalaki, E. G., Droutsa, K., Gaglia, A. G., Kontoyiannidis, S., & Balaras, C. A. (2010). Data collection and analysis of the building stock and its energy performance An example for Hellenic buildings. *Energy & Buildings, 42*(8), 1231–1237. http://dx.doi.org/10.1016/j.enbuild.2010.02.014.

Delmastro, C., Mutani, G., Pastorelli, M., & Vicentini, G. (2015). Urban morphology and energy consumption in Italian residential buildings. *4th international conference on renewable energy research and applications ICRERA-2015, IEEE conference publications*, 1603–1608. http://dx.doi.org/10.1109/ICRERA.2015.7418677

Delmastro, C., Mutani, G., & Corgnati, S. P. (2016). A supporting method for selecting cost-optimal energy retrofit policies for residential buildings at the urban scale. *Energy Policy, 99*, 42 56. http://dx.doi.org/10.1016/j.enpol.2016.09.051.

Deshpande, B. (2012). *How to apply feature selection in linear regression with RapidMiner.* Retrieved 14 February 2017, from http://www.simafore.com/blog/bid/111764/How-to-apply-feature-selection-in-linear-regression-with-RapidMiner.

Evans, S., Liddiard, R., & Steadman, P. (2015). Spatial data for modelling building stock energy needs. A UK perspective for the European Commission Joint Research Centre. H. Bloem, R. Boguslawski, M. T. Borzacchiello, A. Kona, G. Martirano, I. Maschio, & F. Pignatelli (Eds.). *JRC conference and workshop report (EUR 27747)*, 31 45. http://dx.doi.org/10.2790/89997.

Florio, P., & Teissier, O. (2015). Estimation of the energy performance certificate of a housing stock characterised via qualitative variables through a typology-based approach model: A fuel poverty evaluation tool. *Energy and Buildings, 89*, 39–48. http://dx.doi.org/10.1016/j.enbuild.2014.12.024.

Fracastoro, G. V., & Serraino, M. (2011). A methodology for assessing the energy performance of large scale building stocks and possible applications. *Energy and Buildings, 43*(4), 844–852. http://dx.doi.org/10.1016/j.enbuild.2010.12.004.

Grassi, S., Chokani, N., & Abhari, R. S. (2012). Large scale technical and economical assessment of wind energy potential with a GIS tool: Case study Iowa. *Energy Policy, 45*, 73–85. http://dx.doi.org/10.1016/j.enpol.2012.01.061.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques.* Elsevier.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709–722. https://doi.org/10.3758/BF03192961.

Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy & Buildings, 45*, 141–151. http://dx.doi.org/10.1016/j.enbuild.2011.10.061.

Hsu, D. (2015). Identifying key variables and interactions in statistical models ofbuilding energy consumption using regularization. *Energy, 83*, 144–155. http://dx.doi.org/10.1016/j.energy.2015.02.008.

IEA (2016). Energy technology perspectives (Executive summary). *Iea,* 371. http://dx.doi.org/10.1787/energy_tech-2014-en.

ISTAT (2011). *ISTAT.* Retrieved from http://www.istat.it/it/archivio/104317.

Iowerth, H., Lannon, S., Waldron, D., Bassett, T., & Jones, P. (2013). A SAP sensitivity tool and gis-based urban scale domestic energy use model. *Proceedings of BS2013*, 452–459.

La Gennusa, M., Lascari, G., Rizzo, G., Scaccianoce, G., & Sorrentino, G. (2011). A model for predicting the potential diffusion of solar energy systems in complex urban environments. *Energy Policy, 39*(9), 5335–5343. http://dx.doi.org/10.1016/j.enpol.2011.05.031.

Li, Q., Quan, S. J., Augenbroe, G., Yang, P. P.-J., & Brown, J. (2015). Building energy modelling At urban scale: Integration of reduced order energy model with geographical information. *Ibpsa,* 190.

Lombardi, P., Acquaviva, A., Macii, E., Osello, A., Patti, E., & Sonetti, G. (2014). Web and cloud management for building energy reduction: Toward a smart district information modelling. In Z. Sun (Ed.). *Demand-driven web services: Theory, technologies, and applications* (pp. 340–355). Pennsylvania USA: IGI Global. http://dx.doi.org/10.4018/978-1-4666-7230-7.ch079.

Lombardi, P., Abastante, F., & Moghadam, S. T. (2017). *Multicriteria spatial decision support systems for future urban energy retrofitting scenarios.* http://dx.doi.org/10.3390/su9071252.

Ma, J., & Cheng, J. C. P. (2016). Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. *Applied Energy, 183*, 182–192. https://doi.org/10.1016/j.apenergy.2016.08.079.

Mastrucci, A., Baume, O., Stazi, F., Salvucci, S., & Leopold, U. (2014a). A GIS-based approach to estimate energy savings and indoor thermal comfort for urban housing stock retrofitting. *Proceedings of BauSIM 2014* (pp. 190–197).

Mastrucci, A., Baume, O., Stazi, F., & Leopold, U. (2014b). Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. *Energy and Buildings, 75*, 358–367. http://dx.doi.org/10.1016/j.enbuild.2014.02.032.

Mauree, D., Coccolo, S., Kaempf, J., & Scartezzini, J. (2017). Multi-scale modelling to evaluate building energy consumption at the neighbourhood scale. *PLoS ONE, 12*(9), 1–21. http://dx.doi.org/10.1371/journal.pone.0183437.

Mauree, D., Lee, D. S., Naboni, E., Coccolo, S., & Scartezzini, J. (2017). Localized meteorological variables influence at the early design stage. *Energy Procedia, 122*, 325–330.

Mutani, G., & Todeschi, V. (2017). *ScienceDirect Space heating models at urban scale for buildings in the city of Turin.* 0–5 Italy.

Mutani, G., & Vicentini, G. (2013). Analisi del fabbisogno di energia termica degli edifici con software geografico libero. Il caso studio di Torino. *La TERMOTECNICA,* 63–67.

Mutani, G., & Vicentini, G. (2015). Buildings ' energy consumption. *Energy Savings Potential and the Availability of Renewable Energy Sources in Urban Spaces, 2*(11), 1102–1115.

Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V., & Eicker, U. (2015). Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. *Energy and Buildings, 107*, 204–212. http://dx.doi.org/10.1016/j.enbuild.2015.08.021.

Pelzer, P., Arciniegas, G., Geertman, S., & Lenferink, S. (2015). Planning support systems and task-technology fit: A comparative case study. *Applied Spatial Analysis and Policy, 8*(2), 155–175. http://dx.doi.org/10.1007/s12061-015-9135-5.

Penna, P., Prada, A., Cappelletti, F., & Gasparella, A. (2015). Multi-objectives optimization of Energy Efficiency Measures in existing buildings. *Energy and Buildings, 95*, 57–69.

Robinson, D., Haldi, F., Kämpf, J. H., Leroux, P., Perez, D., Rasheed, et al. (2009). CITYSIM: Comprehensive micro-simulation of resource flows for sustainable urban planning. *International IBPSA conference,* 1083–1090.

TABULA (2012). *EU TABLE project.* Retrieved from http://episcope.eu/building-typology/.

Theodoridou, I., Papadopoulos, A. M., & Hegger, M. (2011). Statistical analysis of the Greek residential building stock. *Energy and Buildings, 43*(9), 2422–2428. http://dx.doi.org/10.1016/j.enbuild.2011.05.034.

Torabi Moghadam, S., Delmastro, C., Corgnati, S. P., & Lombardi, P. (2017). Urban energy planning procedure for sustainable development in the built environment: A review of available spatial approaches. *Journal of Cleaner Production, 165*, 811–827. http://dx.doi.org/10.1016/j.jclepro.2017.07.142.

Torabi Moghadam, S., Lombardi, P., & Mutani, G. (2017). A mixed methodology for defining a new spatial decision analysis towards low carbon cities. *Procedia Engineering, 198*, 375–385. http://dx.doi.org/10.1016/j.proeng.2017.07.093.

Tornberg, J., & Thuvander, L. (2005). A GIS energy model for the building stock of Goteborg. *2005 Esri international user conference, (SOU 2004)*. Retrieved from http://proceedings.esri.com/library/userconf/proc05/papers/pap2244.pdf.

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*(9), 1761–1768. http://dx.doi.org/10.1016/j.energy.2006.11.010.

United Nations (2015). *Global initiative for resources efficient cities, Paris, France.* France, Paris: United Nations Environment Programme. Retrieved from http://unep.org/.

Walter, T., & Sohn, M. (2016). A regression-based approach to estimating retrofit savings using the building performance database. *Applied Energy, 179*, 996–1005.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838.

Yang, P. P.-J., & Yan, J. (2016). Modeling urban design with energy performance. *Energy Procedia, 88*, 3–8. http://dx.doi.org/10.1016/j.egypro.2016.06.002.

Yeo, I., Yoon, S., & Yee, J. (2013). Development of an urban energy demand forecasting system to support environmentally friendly urban planning. *Applied Energy, 110*, 304–317. http://dx.doi.org/10.1016/j.apenergy.2013.04.065.

Zanon, B., & Verones, S. (2013). Land Use Policy Climate change, urban energy and planning practices: Italian experiences of innovation in land management tools. *Land Use Policy, 32*, 343–355. http://dx.doi.org/10.1016/j.landusepol.2012.11.009.