An exploratory study on the impact of physical and geospatial characteristics of the urban built environment on the buildings annual electricity usage

Mohammadi, Saleh; de Vries, Bauke; Rafiee, Azarakhsh; Esfandiari, Masoud; Dias, Eduardo

DOI
10.1016/j.jobe.2021.102359

Publication date
2021

Document Version
Final published version

Published in
Journal of Building Engineering

Journal of Building Engineering 40 (2021) 102359

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: http://www.elsevier.com/locate/jobe

An exploratory study on the impact of physical and geospatial characteristics of the urban built environment on the buildings annual electricity usage

Saleh Mohammadi [a,b,*], Bauke de Vries [c], Azarakhsh Rafiee [d], Masoud Esfandiari [e], Eduardo Dias [d]

[a] Department of Architectural Engineering and Technology, Faculty of Architecture and the Built Environment, Delft University of Technology, Julianalaan 134, 2628, BL Delft, the Netherlands
[b] Department of Architecture, Faculty of Art and Architecture, University of Kurdistan, Sanandaj, Iran
[c] Eindhoven University of Technology, Department of Built Environment, Vertigo 9.09, P.O. Box 513, 5600MB, Eindhoven, the Netherlands
[d] Department of Spatial Economics/SPINlab, VU University Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands
[e] Department of Architecture, Faculty of Architecture and Urban Planning, Shahid Beheshti University, Tehran, Iran

A R T I C L E I N F O

Keywords:
Electricity demand modelling
Built environment physical And geospatial characteristics
Multiple linear regression
Ordinary least Squares technique

A B S T R A C T

Applying any sustainable intervention in the urban energy system requires fundamental knowledge of the energy demand dynamics. Only when we can predict the users' energy demand at any given time with accuracy, we can redesign the urban energy system. Accordingly, the main objective of this paper is to determine the annual electricity usage of the building connections in the urban built environment. In this paper firstly through a literature review, the important electricity usage explanatory variables of the built environment are recognized. For each building, besides the annual electricity usage, three major categories of explanatory variables, including physical, socioeconomic, and geospatial characteristics are determined. Based on the available data sources, a building electricity usage database is created. The database is categorized based on the two most frequently used building sectors including residential and non-residential. Ordinary Least Squares (OLS) technique is applied to the constructed database to estimate the predicting model parameters establishing a relationship between the annual electricity usage as a dependent variable and physical, socioeconomic, and geospatial variables as independent variables. In this research, to determine the contribution of geospatial characteristics in the annual electricity usage variability, regression analysis is performed in two consecutive steps. In the first step only, the geospatial characteristics were implemented in the multiple linear regression analysis. Following that, in the second step, the other categories including physical and socioeconomic characteristics are added to the model. The result revealed that in both building sectors most of the predictors are statistically significant at the 0.05 level. While for the residential buildings the geospatial characteristics account for 9.7% of the electricity usage variation, these values for the service and industry sub-sectors are 9.9% and 8.7% respectively. In total, all variables explain 28.1%, 39.4%, and 42.9% of the electricity usage variability of residential, service, and industrial buildings respectively.

1. Introduction

Urban energy demand models seek to determine the energy requirement of the built environment as a function of the input parameters. Models can be applied for different reasons, however, the most common use is determining energy supply requirements for a specific area or evaluation of the changes in the energy consumption of the built environment due to the upgrade or addition of new technology [1]. The main energy forms considered in the literature are electricity and thermal energy and the most frequently considered building types are residential, service, and industrial buildings, varying from small rooms to big estates [2]. The modelling energy demand of the built environment

* Corresponding author. Department of Architectural Engineering and Technology, Faculty of Architecture and the Built Environment, Delft University of Technology, Julianalaan 134, 2628, BL Delft, the Netherlands.
E-mail address: saleh.mohammadi@tudelft.nl (S. Mohammadi).

https://doi.org/10.1016/j.jobe.2021.102359
Received 20 September 2020; Received in revised form 7 February 2021; Accepted 26 February 2021
Available online 2 March 2021

is significantly complex, as the energy types and building types vary greatly and it is influenced by various factors such as physical, constructional, and behavioural characteristics [3]. Due to the complexity of the problem, often a precise prediction is not feasible. In recent years, a plethora of demand prediction models have been proposed and applied to a broad range of problems which vary considerably in terms of data input requirement, disaggregation levels, the socio-technical assumptions, and accordingly the type of results and scenarios that they can assess and predict [4]. Broadly speaking, there are two fundamental classes of modelling methods used to predict and analyse the energy demand of the built environment: the top-down and bottom-up approaches [1,4,5]. The terminology is related to the hierarchal position of the input data compared to the built environment as a whole [1]. While top-down models attribute the total energy demand of the built environment to the characteristics of the entire building stock, bottom-up techniques calculate the energy demand of individual or groups of buildings and then extrapolate these results to the whole area [1]. These models are constructed based on various simulation approach and levels of details and their spatial and temporal application scales are significantly different. The level of detail of the model's inputs depends on the model aim, availability of data, approach of analysis, and underlying assumptions [1]. The bottom-up models are built up based on the data on an individual level to investigate the contribution of end-user on energy usage in the urban and regional levels. These models can be used for simulation of energy use for individual occupants, building, or groups of buildings and then extrapolate the results to represent the city or region, based on the representative weight of the modelled samples [1,6]. Bottom-up models work on a micro-level, therefore for evaluating the energy use, they need extensive databases of empirical data to support the description of each user [7]. The physical characteristics of the built environment such as geometry, envelope fabric, appliances, indoor temperatures, and occupancy and equipment schedules are major common input data for the bottom-up models [1]. In this approach a high level of detail is used which provides modelling of the many technological and technical changes in the built environment. The main strength of the bottom-up approach is the capability of determining the total energy demand of the whole sector without relying on historical data [1] however, as they work at a disaggregated level, they need extensive databases of the empirical data to support the description of each component [4]. As these models evolve, they may be used to estimate the energy consumption of the buildings representative of the built environment and subsequently, results can be extrapolated to the entire urban or regional level.

Recently spatial approaches have been introduced and integrated with the bottom-up energy demand models to incorporate new data collection and extraction techniques. Using these approaches, the energy demand of the urban built environment can be determined more efficiently and effectively without costly on-site measurements [8]. It also provides the possibility to integrate a new category of characteristics in addition to the commonly used physical and socioeconomic characteristics to predict the annual electricity usage of the built environment. In this approach, the demand models are integrated with GIS (Geographic Information System) and RS (Remote Sensing) techniques for spatial data extraction and management. These techniques are applied to facilitate the acquisition and extraction of building data and spatial parameters from the building footprints and fulfil the large data requirements of the urban bottom-up energy demand models without the need of visual inspection and long survey of the properties [9]. [8] introduced a prototype building energy modelling approach to estimate the baseline energy consumption of the built environment, demonstrating how buildings constructional and geometrical data, including form, area, perimeter, exposed size, and orientation can be extracted effectively from aerial maps. A hybrid GIS-based approach is applied to calculate the urban built environment residential heating demand to explore the impact of urban form on the demand levels in the city of London by Ref. [10]. Results revealed that the outputs of the hybrid model are comparable with the top-down energy model. In the last decades, aerial photogrammetric techniques have been widely used to produce highly accurate and detailed 2D and 3D (stereo photos) maps [11]. The availability of 3D data sources enables researchers to use 3D models to analyse the energy demand of the urban built environment. Beginning in the early 1990s, the first 3D city models were built and deployed for the representation of a city in urban planning and still, the deployment of a comprehensive 3D models is the focus of many researchers [12]. Today by employing high-quality remote sensing data, technology capability has reached a level where 3D models can be obtained more readily and cheaply. Besides cloud computing infrastructures and a vast number of software applications has intensified the 3D applications in urban energy demand modelling [3,13]. Consequently, 3D city models can be utilized as a powerful tool for energy evaluation at the large urban scale applications [14,15]. developed an urban energy tool that enables automatic extraction of the building's heating volume from a Geo-information system and LIDAR data.

In this paper, the urban spatial techniques are integrated with the bottom-up demand models to determine the annual electricity usage of the building (electrical) connections in the urban built environment by considering the most important electricity usage explanatory variables. In addition to the conventional constructional and socioeconomic characteristics, the geospatial explanatory variables are also considered in the model. LIDAR technology as an airborne mapping technique is utilized to extract the building's characteristics such as height, building orientation, and solar radiation. The finding of the study demonstrates the significance of geospatial characteristics besides other conventional explanatory variables to predict and explain the electricity usage of the urban built environment. To address this, the paper is structured as follows; after an introduction in Sec. 1, Sec. 2 will briefly describe the applied methodology. The model setup, including identifying the electricity usage explanatory variables, caste study, the building electricity usage database, and pre-processing the data is presented in Sec. 3. This follows by Sec. 4 explaining the MLR analysis and the construction of the electricity predicting formulas. The results are discussed in Sec. 5. Lastly, Sec. 6 summarizes and concludes the paper.

2. Methodology

The main objective of this paper is to determine the annual electricity usage of the building connections by considering the most important electricity usage explanatory variables. Accordingly, a methodology composed of several specific steps is designed and implemented to fulfil this goal. Firstly, through a literature review, the important electricity usage explanatory variables are identified. For each building, besides the annual electricity usage, three major categories of explanatory variables, including physical, socioeconomic, and geospatial characteristics are determined (Sec. 3.2). Case study explanations is presented in the Sec. 3.3. Based on the identified variables and available data sources the building electricity usage database comprising of the electricity usage of buildings and related descriptive variables is constructed (Sec.3.4). The created database is categorized based on the two most frequently used building sectors including residential and non-residential (NR). Subsequently, the pre-processing step involving identifying missing data and outlier detection is performed to prepare data for the statistical analysis (Sec. 3.5). Multiple linear regression (MLR) is applied to the created database using the Ordinary Least Square (OLS) technique to estimate the parameters of the models that establish a relationship between the annual electricity usage and the electricity usage explanatory variables (Sec. 4). The models are utilized to predict the annual electricity usage of three different building connections. In this research, the statistical analysis is carried out by IBM SPSS® Statistics, Version 22, and GIS operations are carried out with the spatial analysis tools, FME® Version 2014-build 14235, QGIS version 2.0.1 and ArcGIS version 10.3. FME is utilized for data transformation and translation and QGIS and ArcGIS are applied to store, retrieve, manage, display and analyse geographical and spatial data.

3. Model setup

3.1. Demand modelling technique

In this research, a statistical/regression bottom-up model is employed which is particularly useful when a large dataset exists and as it is based on real data, this gives a good understanding of the energy usage behaviour [16]. The energy form considered is electricity and the target building sectors are residential and NR. The energy assessment of the urban built environment can be performed more efficiently and effectively by spatial tools such as GIS and RS. Therefore, in this research, the GIS and RS are utilized to facilitate the acquisition and extraction of the geospatial characteristics such as building area, density, and urbanization degree. LIDAR technology is also utilized to generate the 3D model of the analysed area and extraction of features such as height, solar intensity, and roof type, and area.

3.2. Electricity usage explanatory variables

Building electricity usage is influenced by a variety of factors, such as physical properties, the behaviour of occupants, and the surrounding environment. Due to the complex interrelation of these factors often a precise prediction of demand is not feasible [2]. Building type and the activities performed in the building are the main factors that influence electricity usage. The most widely used building sectors are residential and NR buildings which vary from small rooms to large estates [2]. Based on this fact that the building type has a significant effect on building electricity usage, in this section the determination of the electricity usage explanatory variables for each sector will be carried out separately. In the following subsections, through the literature review, the important explanatory variables are identified and subsequently will be utilized to create the building electricity usage database.

3.2.1. Residential sector

Residential electricity demand modelling is a heavily studied subject [5] which resulted in a large body of research. Literature has identified important explanatory variables that have been used to model the residential electricity demand [5,16–22]. Based on the studies reviewed, Table 1 shows a list of explanatory variables that have been used in the electricity demand modelling of residential buildings ranked on the number of citations in the reviewed literature.

Additionally, Table 1 also shows that type of building, income, the number of occupants, floor area, building age and location are some of the most used variables for modelling the residential electricity demand.

Table 1
List of explanatory variables which have been used to model the residential electricity demand.

Variable	Number of Citations	Variable	Number of Citations
Type of building	11	*No. of rooms*	4
Income	10	*Employment status*	4
Number of occupants	10	*Insulation variable*	4
Floor area	10	*Tenure type*	3
Homeowner age	9	*Degree days*	3
Building age	9	*Dwelling value*	3
Weather (temperature)	9	*Electricity price*	2
Appliance holdings	9	*Social group*	2
Heating type	8	*No. of bedrooms*	2
Location	8	Educational level	2
Household composition	8	Period of residency	1
Appliance rating	8		
Time use	7		

Using a questionnaire survey, supported by annual gas and electricity meter data and floor-area estimates [23], suggests that the floor area of dwellings, total occupancy, and homeowner age are the major indicators of the residential electricity demand. According to Ref. [24], type of dwelling, location, size, household appliances, and attributes of the occupants, including the number of occupants and age have differing but significant impacts on the electricity demand. Also, a strong correlation was found between the average annual electricity demand and floor area [24]. It should be noted that the higher frequency of indicators is not necessarily related to their performance, but may have other reasons such as data availability, the ease with which data can be collected, or the analytical approach of the study. For example, data related to the building type or age can be extracted from the national census with relative ease. Other variables such as floor area may be overlooked due to the difficulty of gathering this information [16,22]. In agreement with Table 1 [25,26], showed that residential electricity demand is strongly influenced by the income and floor area, whereas the age of family members and the level of education has limited impact [24]. showed that residential electricity demand is correlated with the location of individual dwellings. Applying the linear regression model on the Irish National Survey of Housing Quality revealed that dwelling features include location and value and that household features such as income, the period of residency, social status, and tenure type have significant correlations with the residential electricity demand [27].

3.2.2. Non-residential sector

NR sector, including service and industrial sub-sectors, consume a significant portion of the total urban electricity demand, however, compared with the residential sector very few studies have been conducted to investigate their electricity consumption behaviour. It is primarily for the reason that the residential sector broadly is considered as the dominant consumer of the electricity within the building stock. Secondly, the large-scale assessment of the NR sector is often infeasible or difficult due to the extreme diversity of uses, activities, and ownership structures within this sector. Hence, comprehensive information and detailed data about electricity modelling in this sector are considerably limited [28,29]. A recent study of European countries found the lack of available data, format inconsistencies of available data, and the overall difficulty in collecting necessary data for this sector [29].

Efforts have been made on tackling the complex task of characterizing electricity consumption in the NR building stock. In the UK the attempts began in the 1990s to develop a national NR building stock database to understand electricity use. This dataset provides building floor, age, and activity information that is supplemented by street surveys [30,31]. By breaking down the NR sector into the 588 premises [32,33] reported electricity-use patterns for different types of NR buildings. While [34] used the non-domestic building stock (NDBS) dataset of the UK to estimate the carbon emissions from the main building types [35,36], applied the NDBS dataset to develop a technique for dealing with the heterogeneity of the non-domestic building stock combined with the floor area model for each property type to predict total electricity consumption [9]. quantified the electricity usage of NR buildings by assigning published energy benchmark values to relevant property types [6]. developed a district clustering method to calculate the electricity use of the commercial sector of the Osaka City based on representative building types [29]. elaborated on the approach to determine the potential energy conservation in the Hellenic non-residential building stock [37]. developed simulation models with EnergyPlus for two office buildings in an R&D centre in Shanghai, China to evaluate the electricity and heating cost savings of green building design options compared with the baseline building [38]. develop a multiple regression model to predict the annual energy consumption in the Spanish banking sector and determine how energy efficient a bank branch is depending on its construction characteristics and climatic area. Specifically, in the industrial sub-sector, most of the investigations to analyse the electricity demand have been carried out since the middle of the seventies at the aggregate level without considering activity/industry details. Recently [39–41] have analysed industrial companies' electricity demand by micro cross-section data and repeated cross-section data to investigate how the characteristics of the industrial companies such as size, area, type, and electricity intensity influence the electricity demand. An econometric analysis of electricity demand based on the panel data of 2949 Danish industrial companies over 13 years has been presented in Ref. [41]. In this study, attention has been devoted to the variation in electricity demand, according to the characteristics of the industrial companies such as size, type of industrial sub-sector and electricity intensity in production.

3.3. Case study

To demonstrates the impact of physical, socioeconomic and geospatial characteristics of the urban built environment on the annual electricity usage and in-depth study of patterns and relationships, a case study method has been applied. This provides a ground to support an empirical inquiry to investigates a phenomenon in depth and within its real-life context [42]. For this research, Eindhoven has been chosen as the study area. Eindhoven is the fifth-largest city of the Netherlands with a population of 234,401 in 2020 [43] and total electricity connections of 111410. This large number of connections provide data for the bottom-up statistical approach as this model require extensive databases of empirical data to be able to infer and explain the patterns. The result of this large-scale modelling can also be applied in the regional and national level in which the collected data can be utilized for electricity usage benchmarking and long-term decision making in respect to identifying the saving measures to achieve a sustainable cities ambitions.

Eindhoven has become a "Brainport" city, the centre of high-tech industries in the Netherlands, since 2008. This has led to attracting skilled immigrants and the rapid increase of the population and urban density [44]. [45,46] have implied rapid urbanization is associated with increased energy demand. To investigate this issue and generalize this study to other cities with rapid urbanization, Eindhoven can be a good example. In fact, the developed method for this study regardless of the local and contextual data can be applied for a variety of cities to analyses electricity demand. In additions, many cities have restricted regulations to access data such as geospatial or energy demand data; however, in this case, these data are open and accessible to researchers, providing this study with essential information to conduct the analysis. Considering the fact that Eindhoven is situated in the northern European oceanic climate, this case can also be a representative of the cities in this climate with the same population size.

3.4. Building electricity usage database

Since the bottom-up statistical demand models work on a micro-level, for examining the annual electricity usage on the building level, these models need extensive databases of historical data. In these models, the basic step to perform the annual electricity usage analysis is creating the building electricity usage database by incorporating different data sources. To achieve this goal different sources have been investigated and several organizations have been asked for the required data for analysis. In this research, two main types of data, including energy and spatial data are required and for this purpose, numerous sources are explored. The most important source of energy data was Endinet groep BV. Endinet as a local distribution system operator (DSO) operates the power grid in the Eindhoven. The electricity billing data of the Eindhoven region are retrieved as it was registered in the company SAP[2] system. This resulted in the list of 111410 electricity connections. For each connection in addition to the electricity usage, the building address (including street, house number, house letter, and six-digit postal-code), ground area, woz-value[3] and physical capacity of the connection (amperage/connection phase-type) are specified. The woz-value is a valuation of the property and it is estimated every year by the municipality. For the demand module, woz-value is used as an indication of the property value. The annual electricity usage of the building connections is determined by the average consumption of the building over the period between 2005 and 2009. The electricity data of the connections were available in the time intervals that were not necessarily entire years; therefore, a pre-processing analysis to determine the annual electricity usage per connection is performed. The second main data source of this research was the BAG[4] dataset. This dataset was the main source of the spatial data of this research. Several main building characteristics, including building footprint, location coordination, function, roof and parcel area, building age, building type, and roof type are acquired from this dataset. Row, apartment, and detached buildings were the main considered building types for residential buildings and with respect to the roof types, the flat and pitched roofs were the dominant types. Records in the Energy (Endinet) and spatial (BAG) datasets are connected by the postal code and street number. A combination of six-digit postal code and street number generates a unique indicator for each building which is utilized to connect different datasets. By merging the datasets, the building energy characteristics are associated with building spatial characteristics. Connecting the datasets result in the final list of connection records, comprising 110232 building connections in the study area (1178 records were disregarded as they were not referenced in the BAG dataset). As expected, a majority of the building connections were residential buildings with 89.12% (98237) of the total records. 9.45% (10414) and 1.43% (1581) of the records were services and industrial buildings respectively.

As described, in this study to determine the annual electricity usage of the building connections, geospatial characteristics besides the physical and socioeconomic are also applied. These characteristics cannot be obtained directly from the open datasets and need to be determined which includes the following variables:

- ***Building height:*** It is estimated by overlaying the detailed height data of the study area on the 2D footprints of the buildings. The height data, called AHN2[5], is captured through the airborne LIDAR technique for the entire study area. Fig. 1 shows the AHN2 point cloud data of the part of the study area. By intersecting the LIDAR point cloud of each building with its 2D footprint using their planar coordinates, a statistical average height was estimated for each building.
- ***Urbanization degree*****:** It is introduced to account for different levels of urbanization. For each building, it is estimated using the land use map of the region, which is divided into two categories, namely, urban and non-urban areas. Urban areas consist of urbanized elements, such as buildings and roads, and non-urban areas, on the contrary, composes of natural elements such as forests, parks, and agricultural areas. For each building, the urbanization degree is estimated as the proportion of urban/non-urban areas in its proximity. A buffer with a 3 km radius is defined for each building to calculate the urban/non-urban land use share for the building in that radius. Fig. 2 (left) presents the land use map of part of the study area which is classified into the urban (red) and non-urban (green) land uses. The resulting urbanization degree for each building is presented in Fig. 2 (right). The right part of the image consists of buildings with more green colours. These are the buildings with a low urbanization degree. Moving to the left part of the image the buildings colour changes to yellow, orange, and red. This is the result

[1] Systems Applications Products.
[2] Waardering Onroerende Zaken (Real Property Valuation).
[3] Basisregistraties Adressen en Gebouwen
[4] The AHN2 was collected and became available by the Dutch Water Boards and Rijkswaterstaa as a part of the Ministry of Infrastructure and the Environment.

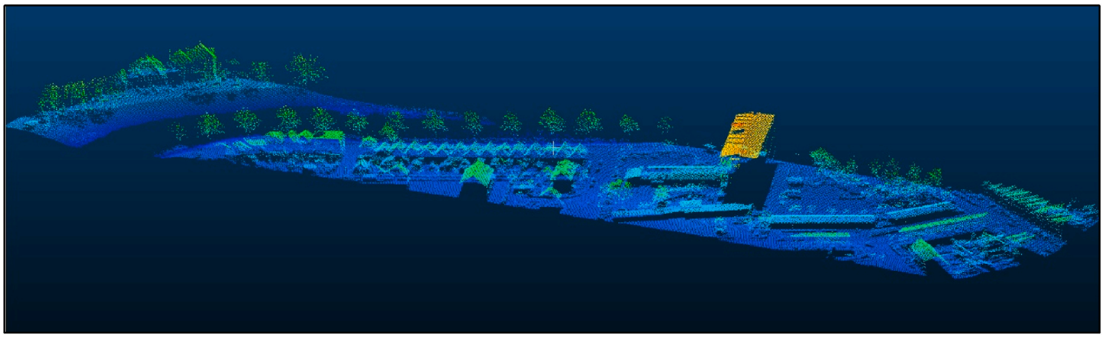


Fig. 1. AHN2 point cloud data of the part of the study area used to calculate the heights of the building.

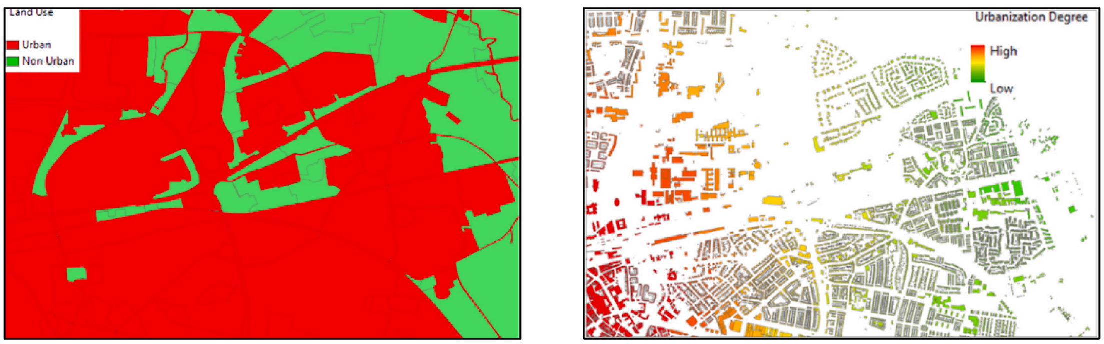


Fig. 2. (left) Land use map of part of the study area which is classified into two categories of urban (red) and non-urban (green) land uses (right) The resulting urbanization degree for each building. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of the urban area increase from right to left, which can be seen in Fig. 2 (left).

- ***Urban density:*** It is estimated as the total number of buildings in the proximity of each building. To estimate the local urban density, nearby each building, a buffer of 100 m is defined. Using the 2D footprint of the buildings of the study area, the number of buildings overlaying each buffer is calculated and assigned to the building for which the buffer is defined. The result of urban density calculations for part of the study area is shown in Fig. 3.

Fig. 3. Urban density analysis of part of the study area; the low-density areas are depicted in red and high-density areas are in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- ***Solar intensity:*** The total amount of solar radiation on a specific location has been estimated for the entire year using ArcGIS software. This software estimates the solar insolation based on a hemispherical viewshed algorithm developed by Ref. [47] and improved by Ref. [48]. The input geospatial data for the calculation of solar insolation is the elevation model as well as the latitude of the study area. Fig. 4 presents an example of the height model and the calculated solar insolation of the study area for the whole year.

3.5. Data pre-processing and outlier detection

Through a review of electricity usage explanatory variables in the residential and NR sectors (Sec. 3.2.1 and Sec. 3.2.2) and based on the constructed database (Sec.3.4), for each building connection besides the *annual electricity usage*, three major categories of explanatory variables are identified as follows:

- **Physical characteristics** (*CONSTRUCTION_YEAR*, *GROUND_AREA*, *ROOF_AREA*, *PARCEL_AREA*, *CONNECTION_TYPE*, *BUILDING_TYPE,* and *ROOF_TYPE*)
- ***Geospatial characteristics*** (*URBANIZATION_DEGREE*, *HEIGHT*, *SOLAR_INTENSITY,* and *DENSITY*)
- ***Socioeconomic characteristics*** (*WOZ_VALUE*; only applicable for the residential sector)

Table 2 illustrates the list of the electricity usage explanatory variables which is employed in this study to analyse the annual electricity usage of the urban built environment. It is observed that several variables in the database have missing values such as *WOZ_VALUE* (12.45%), *CONNECTION_TYPE* (~0.00%), *BUILDING_TYPE* (1.6%), *ROOF_TYPE* (9.57%), *PARCEL_AREA* (~0.00%), *HEIGHT* (4.5%) and *SOLAR_INTENSITY* (3.49%). However, as seen, the number of missing values compared to the valid values is not significant, therefore these values are treated as missing data in the subsequent statistical analysis and are replaced by 999 in the case of continuous variables and by NA for the categorical variables. Details of the missing data are shown in Table 2 for each sector separately.

After specifying the missing data, the next step of the data pre-processing is outlier detection and removal of the outliers. The outliers are typical, infrequent observations that appear to be inconsistent with the majority of the observations in a database [49]. In the building electricity usage database, it is important to identify the outliers and adjust them based on the available rules. An outlier identification approach base on the *outlier labelling rule* is applied to detect outliers in the database. The *boxplot outlier labelling rule* was introduced by Ref. [50] and later has been developed further by Refs. [51,52]. It determines observations as outliers if they appear outside the following boundaries which will be determined by the following equations:

$$upper\_boundary = q_3 + (g \times (q_3 - q_1)) \quad (1)$$

$$lower\_boundary = q_1 - (g \times (q_3 - q_1)) \quad (2)$$

where q_1 and q_3 are the lower and upper quartiles of the variable respectively. The common choices for g are 1.5 or 2.2 depending on the distribution of data and research context. The outlier detection survey has been performed on the building electricity usage database and the results are presented in Table 3. As shown, although most of the continuous variables in the database have outliers, however, compared to the total number of records is not significant and in the subsequent analysis, they are excluded. As demonstrated, except for the solar light intensity all other continuous variables have upper boundary outliers.

Several variables in the constructed database are categorical variables. For applying these variables in the subsequent analysis, they have to be transformed into the dummy variables. A dummy variable is a qualitative variable that can only take 0 or 1 and is used in the regression analysis instead of the categorical variable. While the dummy variable levels of the *BUILDING_TYPE* are ROW and APARTMENT (the detached house is considered as a base value), the levels of the *CONSTRUCTION_YEAR* are a stepwise increase of the oldest level up to the most recent level (1992 is considered as a base value). *CONNECTION_TYPE* is categorized into three levels based on the connection amperage and phase types (3 × 100 amp connection type is considered as a base value). *ROOF_TYPE* is also categorized into the flat and pitched roofs with FLAT as a dummy variable. Table 4 shows the categorical variables and corresponding assigned dummy variables.

Table 5 summarizes the final list of the explanatory variables as it is used for the subsequent statistical analysis. As mentioned, the database is categorized based on the main building sectors comprising residential

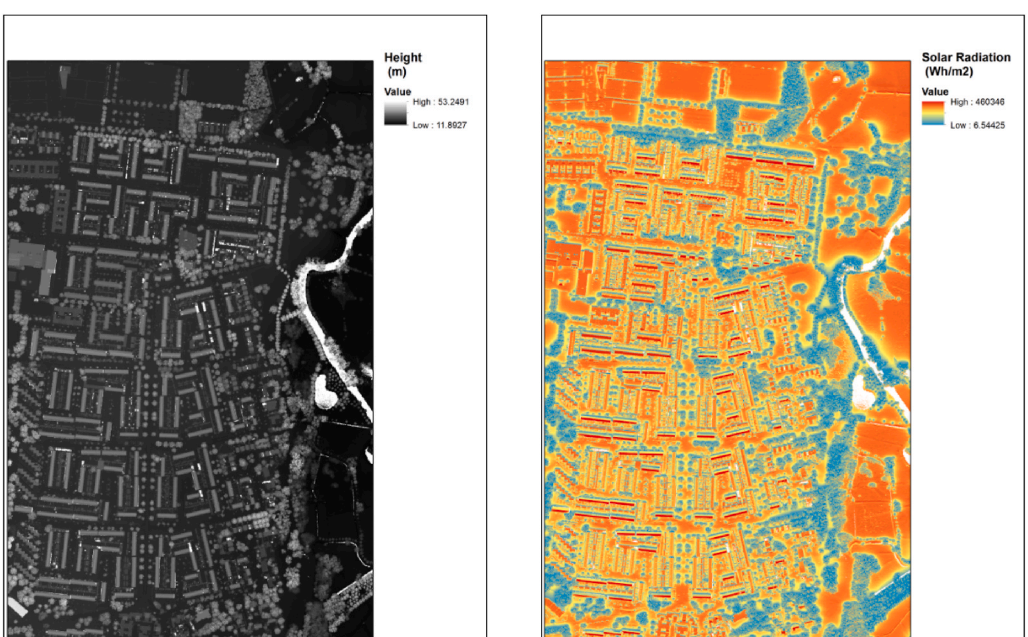


Fig. 4. (left) An example of the height model of part of the study area and (right) Calculated solar radiation of the area for the entire year.

Table 2
List of the explanatory variables and number of valid and missing values.

Categories	Variables	Sources	Res.	NR	N. of Valid	N. of Missing
	annual electricity usage	*Endinet*	×	×	*Res: 98237* *Ser: 10414* *Ind: 1581*	*Res: 0* *Ser: 0* *Ind: 0*
Socioeconomic characteristics	*WOZ_VALUE*	*Endinet*	×		*Res: 90362* *Ser: 6264*	*Res: 7875* *Ser: 4150*
Physical characteristics	*CONNECTION_TYPE*	*Endinet*	×	×	*Res: 98222* *Ser: 10388* *Ind: 1573*	*Res:15* *Ser: 26* *Ind:8*
	BUILDING_TYPE	*BAG*	×	×	*Res: 96467* *Ser: 10414* *Ind:1581*	*Res: 1770* *Ser: 0* *Ind: 0*
	CONSTRUCTION_YEAR	*BAG*	×	×	*Res: 98237* *Ser: 10414* *Ind: 1581*	*Res: 0* *Ser: 0* *Ind: 0*
	GROUND_AREA	*BAG*	×	×	*Res: 98237* *Ser: 10414* *Ind: 1581*	*Res: 0* *Ser: 0* *Ind: 0*
	ROOF_TYPE	*BAG*	×	×	*Res: 89644* *Ser: 8720* *Ind: 1317*	*Res: 8593* *Ser: 1694* *Ind: 264*
	PARCEL_AREA	*BAG*	×	×	*Res: 98166* *Ser: 10414* *Ind: 1581*	*Res: 71* *Ser: 0* *Ind: 0*
	ROOF_AREA	*BAG*	×	×	*Res: 98237* *Ser: 10414* *Ind: 1581*	*Res: 0* *Ser: 0* *Ind: 0*
Geospatial characteristics	*URBANIZATION_DEGREE*	*BAG*	×	×	*Res: 98237* *Ser: 10414* *Ind: 1581*	*Res: 0* *Ser: 0* *Ind: 0*
	HEIGHT	*AHN2*	×	×	*Res: 94017* *Ser:9744* *Ind:1507*	*Res:4220* *Ser:670* *Ind:74*
	SOLAR_INTENSITY	*AHN2*	×	×	*Res: 95209* *Ser:9695* *Ind:1482*	*Res:3028* *Ser:719* *Ind:99*
	DENSITY	*BAG*	×	×	*Res: 98237* *Ser:10414* *Ind:1581*	*Res: 0* *Ser: 0* *Ind: 0*

Res: Residential, NR: Non-residential, Ser: Service, Ind: Industry.

and NR.

4. Multiple linear regression analysis

Multiple linear regression (MLR) is applied to the building electricity usage database to develop a model that establishes the relationship between the annual electricity usage and the electricity usage explanatory variables to predict the annual electricity usage of building connections in the urban environment. Traditionally, regression analysis has been the most popular modelling technique in predicting electricity usage in the built environment [53–57]. Regression analysis is especially valuable when a large database exists as it is based on real data and provides a good understanding of the electricity usage behaviours [16]. The main reasons for the popularity of the regression models may be explained by the interpretability of model parameters and results, however, the major limitation of these techniques is that one can only ascertain the relationship but cannot be sure about the underlying causal mechanism [57]. It also can be costly and inconvenient to implement and may suffer from multicollinearity (hereafter referred to as collinearity) between independent variables [16]. MLR seeks to establish a relationship between a dependent variable and two or more independent variables or predictors that may be written as [57,58]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad (3)$$

Where y is the dependent variable, β_i the regression coefficients ($i = 1, 2, \ldots, n$), x_i the predictive variables ($i = 1, 2, \ldots, n$) and ε the random error term. The regression coefficients are estimated based on a record of observations which is normally carried out by curve fitting based on the ordinary least square method (OLS) to minimize the difference between the observed and estimated values [58]. Once the coefficients are estimated, a prediction equation can be used to predict the value of a dependent variable as a function of the independent predictors [57]. The independent variables should also have little or no correlation with each other to prevent collinearity problems during data analysis [58]. The random error term, ε, is used to test the overall significance of the regression equation and the significance of the estimated coefficients which have to be normally and independently distributed with a mean of zero [58]. The coefficient of determination (R^2), which ranges between 0 and 1, is another important output of the regression analysis which indicates the goodness of fit of a regression model and can be interpreted as the proportion of the variance in the dependent variable that can be explained by the independent variables [58]. In MLR analysis the choice of the functional form to describe the dependent variable is somewhat arbitrary and there is no real consensus about the best form [5]. While [59] assumes a linear relationship [60], considers a logarithmic relationship, and a mixture of both is assumed by Ref. [61]. In this research, the regression analysis is performed with the IBM SPSS Statistics software package, version 22. For this application to predict the annual electricity usage of building connections, a stepwise regression model is applied as it ranks the variables based on their importance and to minimize the collinearity between explanatory variables it adds the variables to the model in sequentially [20]. For the application of stepwise selection model, the inclusion p-value is set to 0.05 and exclusion p-value is set to 0.10. In the next section the MLR analysis will be applied to the created building electricity usage database and analysis is conducted separately for each building sector. Following that the

Table 3
Results of the performing outlier's detection analysis ($g = 2.2$)

Variables	Mean	Lower boundaries	Upper boundaries	Number of modified cases
GROUND_AREA	*Res[1]: 139.6658 Ser[2]: 349.7361 Ind[3]: 942.5648*		*Res: 248.8 Ser: 358 Ind: 2314.4*	*Res: 3047 (3.1%) Ser: 1345 (12.91%) Ind: 155 (9.8%)*
ROOF_AREA	*Res: 325.9423 Ser: 2402.6524 Ind: 1399.1065*		*Res: 422.884 Ser: 5545.222 Ind: 4807.242*	*Res: 15323 (15.6%) Ser: 850 (8.16%) Ind: 71 (4.49%)*
PARCEL_AREA	*Res: 6749.2425 Ser: 3754.5836 Ind: 5650.4132*		*Res: 10845.2 Ser: 8886.8 Ind: 13564*	*Res: 3991 (4.6%) Ser: 753 (7.23%) Ind: 93 (5.88%)*
annual electricity usage	*Res: 4488.38 Ser: 24044.32 Ind: 52463.53*		*Res: 9454.4 Ser: 27745.2 Ind: 68565.2*	*Res: 2542 (2.59%) Ser: 1432 (13.75%) Ind: 189 (11.95%)*
WOZ_VALUE	*Res: 371605.85 Ser: 980323.91*		*Res: 808400 Ser: 1059300*	*Res: 4775 (4.86%) Ser: 695 (6.67%)*
HEIGHT	*Ind: 51.8623*		*Ind: 12.76*	*Ind: 18 (5%)*
SOLAR_INTENSITY	*Res: 330913.06 Ind: 348409.26*	*Res: 78762.2 Ind: 228533.2*		*Res: 1398 (6.3%) Ind: 117 (13.9%)*
URBANIZATION_DEGREE	*Ser: 5633302.86*		*Ser: 12882562.4*	*Ser: 41 (0.7%)*
DENSITY	*Res: 192.74 Ser: 193.71*		*Res: 503.6 Ser: 527.2*	*Res: 1425 (1.45%) Ser: 204 (1.95%)*

Res: Residential, Ser: Service, Ind: Industry.

Table 4
Categorical variables in the building electricity usage database and assigned dummy variables.

Variables	Base values	Dummy variables	Explanations
BUILDING_TYPE	*Detached house*	*ROW*	*Row buildings*
		APARTMENT	*Apartments*
CONNECTION_TYPE	*Over 3×100 Amp*	*CONNECTIONI*	*1×25 Amp and 1×35 Amp*
		CONNECTIONII	*1×35 Amp*
		CONNECTIONIII	*3×25 Amp*
CONSTRUCTION_YEAR	*Over 1992*	*AGEI*	*To 1945*
		AGEII	*1946–1964*
		AGEIII	*1965–1974*
		AGEIV	*1975–1991*
ROOF_TYPE	*Pitched roof*	*FLAT*	*Buildings with flat roofs*

results in each sector are described in detail.

4.1. Residential sector

A stepwise MLR analysis is performed to determine the statistically significant predictors (explanatory variables) for the annual electricity

Table 5
Descriptive statistics of the final list of the electricity usage explanatory variables.

Variables	Unit	Modalities (code) or range	Mean	Std. Dev.
annual electricity usage	*kWh*	***Res:** 100- 9454 **Ser:**12-27735 **Ind:** 13-67899*	***Res:** 3072.95 **Ser:** 4460.91 **Ind:** 10328.5*	***Res:**1753.15 **Ser:** 5607.90 **Ind:** 13918.88*
WOZ_VALUE	*€*	*Res: 65000-808000 Ser:73000-1056000*	***Res:** 302547 **Ser:** 332070*	***Res:** 134464.6 **Ser:** 209047.4*
CONNECTION_TYPE		***Res:** 1×25 Amp (1) [75.7%];3×25 Amp (2) [22.8%]; 3×35 Amp (3) [1.1%]; Over 3×100 Amp (4) [0.3%] **Ser:** 1×25 Amp (1) [49%];3×25 Amp (2) [31.1%]; 3×35 Amp (3) [14.4%]; Over 3×100 Amp (4) [5.5%] **Ind:** 1×25 Amp (1) [23.1%];3×25 Amp (2) [35.9%]; 3×35 Amp (3) [29.1%]; Over 3×100 Amp (4) [12%]*		
BUILDING_TYPE		***Res:** Apartment (1) [29%]; Row (3) [65.8%]; Detached (4) [5.2%] **Ser:** Apartment (1) [92.2%]; Row (3) [3.7%]; Detached (4) [4.1%]*		
CONSTRUCTION_YEAR		***Res:** to 1945 [16.9%]; 1946–1964 [24%]; 1965–1974 [18.3%]; 1975–1991 [21%]; over 1992 [19.8%] **Ser:** to 1945 [18.7%]; 1946–1964 [17.7%]; 1965–1974 [13.8%]; 1975–1991 [22.4%]; over 1992 [27.4%] **Ind:** to 1945 [9.5%]; 1946–1964 [13.4%]; 1965–1974 [12.9%]; 1975–1991 [23.8%]; over 1992 [40.4%]*		
GROUND_AREA	*m^2*		***Res:** 117.07*	

(*continued on next page*)

Table 5 (*continued*)

Variables	Unit	Modalities (code) or range	Mean	Std. Dev.
		Res: *10-249* ***Sir:*** *11- 358* ***Ind:****15- 2293*	***Ser:*** *105.99* ***Ind:*** *420.70*	***Res:*** *38.32* ***Ser:*** *62.70* ***Ind:*** *457.14*
ROOF_TYPE		***Res:*** *Flat (1) [18.4%]; Pitched (2) [81.6%]* ***Sir:*** *Flat (1) [6.2%]; Pitched (2) [93.8%]* ***Ind:*** *Flat (1) [38%]; Pitched (2) [62%]*		
PARCEL_AREA	m^2	***Res:*** *11-10794* ***Sir:*** *5 - 8875.00* ***Ind:****23 - 12311.00*	***Res:*** *1950.63* ***Ser:*** *1903.24* ***Ind:*** *2698.51*	***Res:*** *2477.10* ***Ser:*** *2134.67* ***Ind:*** *2971.89*
ROOF_AREA	m^2	***Res:*** *10- 423.26* ***Sir:*** *12.04-5459.46* ***Ind:****7.35-4644.43*	***Res:*** *90.21* ***Ser:*** *1142.66* ***Ind:*** *1035.27*	***Res:*** *65.98* ***Ser:*** *1291.17* ***Ind:*** *1071.2*
URBANIZATION_DEGREE		***Res:*** *3192787-17152876* ***Ser:****3191325-12781883* ***Ind:*** *3274455-16884071*	***Res:*** *7555977* ***Ser:*** *5600457* ***Ind:*** *7568679*	***Res:*** *2575696* ***Ser:*** *2038455* ***Ind:*** *2871998*
HEIGHT	*m*	***Res:*** *2.30-18.50* ***Ser:*** *2.2-52.4* ***Ind:*** *2.40-13.40*	***Res:*** *5.8940* ***Ser:*** *4.5400* ***Ind:*** *5.308*	***Res:*** *2.06866* ***Ser:*** *3.40607* ***Ind:*** *1.8247*
SOLAR_INTENSITY		***Res:*** *78773-458388* ***Ser:*** *145-457002* ***Ind:*** *229080-457799*	***Res:*** *335346* ***Ser:*** *288700* ***Ind:*** *365807*	***Res:*** *78260.05* ***Ser:*** *105342.2* ***Ind:*** *40423.32*
DENSITY		***Res:*** *1-503* ***Sir:*** *2-527* ***Ind:*** *4-523*	***Res:*** *186.77* ***Ser:*** *193.37* ***Ind:*** *194.74*	***Res:*** *90.33* ***Ser:*** *94.39* ***Ind:*** *101.13*

Res: Residential, Ser: Service, Ind: Industry.

usage of the residential building connections and construct the prediction equation. As mentioned in Section 3.5, three main categories of predictors are applied in the regression analysis, including physical, geospatial, and socioeconomic characteristics. MLR analysis is applied to determine the influence of each predictive category on the annual electricity usage prediction. Descriptive statistics of the continuous predictors such as mean and standard deviation values are presented in Table 5. A full list of categorical variables that are applied in the MLR analysis is shown in Several variables in the constructed database are categorical variables. For applying these variables in the subsequent analysis, they have to be transformed into the dummy variables. A dummy variable is a qualitative variable that can only take 0 or 1 and is used in the regression analysis instead of the categorical variable. While the dummy variable levels of the *BUILDING_TYPE* are *ROW* and *APARTMENT* (the detached house is considered as a base value), the levels of the *CONSTRUCTION_YEAR* are a stepwise increase of the oldest level up to the most recent level (1992 is considered as a base value). *CONNECTION_TYPE* is categorized into three levels based on the connection amperage and phase types (3 × 100 amp connection type is considered as a base value). *ROOF_TYPE* is also categorized into the flat and pitched roofs with *FLAT* as a dummy variable. Table 4 shows the categorical variables and corresponding assigned dummy variables.

Table 4 with both the base variable and corresponding dummy variables. Before the MLR analysis is performed, the correlations analysis and collinearity diagnostics should be investigated. Independent variables which have a high correlation with the dependent variable is preferred in the MLR analysis. The parametric and nonparametric correlations between the dependent variable and continuous and categorical predictors in the residential sector are presented in Table 12 in the Appendix. For the continuous variables, while the correlations between the *annual electricity usage* and *GROUND_AREA*, *WOZ_VALUE*, *PARCEL_AREA*, *HEIGHT,* and *URBANIZATION_DEGREE* are significant, the correlation coefficients between *ROOF_AREA*, *SOLAR_INTENSITY,* and *DENSITY* with *annual electricity usage* is relatively low. For the categorical variables, the non-parametric correlations between the *annual electricity usage* and *CONNECTION_TYPE*, *ROOF_TYPE,* and *BUILDING_TYPE* are significant. The residential *annual electricity usage* variations in terms of categorical variables, including *BUILDING_TYPE* and *CONNECTION_TYPE* are presented in Fig. 5. As shown, detached buildings consume more electricity than other building types, with the mean value of 4770.47 kWh/yr. more than double of the apartments (2025.35 kWh/yr.) and higher than the average annual electricity usage of the residential buildings in the Netherlands (~3500 kWh/yr.). The graphs also show the standard deviation of annual electricity usage for each level of the above-mentioned variables.

The other important aspect of the correlation analysis is the investigation of the collinearity between the predictors. To have a more acceptable outcome of the regression analysis the correlation effects between the predictors should be minimized. The collinearity analysis between the predictors in the residential sector is presented in Table 13 in the Appendix. As the table shows there is a relatively low level of correlations between the predictors, therefore it is expected that most of them will be employed in the MLR analysis. The only exceptions are the *HEIGHT* and *BUILDING_TYPE* that show relatively strong correlations with other independent variables. The *BUILDING_TYPE* shows strong collinearity with the *WOZ_VALUE*, *GROUND_AREA,* and *ROOF-AREA* with the (Pearson) correlation coefficients of 0.32, 0.41, and 0.61 respectively. This can be attributed to the fact that building types such as detached or semi-detached have higher ground areas which consequently lead to the higher woz-value and roof area. *HEIGHT* also shows strong collinearity with the *BUILDING_TYPE*, *ROOF_AREA,* and *GROUND_AREA* with the (Pearson) correlation coefficients exceeding 0.45 in three cases. In order to determine the contribution of geospatial characteristics in the annual electricity usage variability, the regression analysis is performed in two steps. In the first step only, the geospatial characteristics are implemented in the MLR analysis (Model 1). Following that, in the second step, all the variable categories are implemented in the model (Model 2). Table 6 shows the results of the regression analysis in the residential sector for the two models.

According to Table 6, the geospatial characteristics account for 9.7% of the total variability in building electricity usage. In total, all variables explain 28.1% of the electricity usage variance. The results also revealed that the independent error assumption has been met and the Durbin-Watson value was within the specified boundaries. A 95% confidence interval for β showed that the model is reliable. The model also seems not to have collinearity, because tolerance values and variance inflation factors (VIF) are within the limits. The standardized residuals and normal probability plots compare the distribution of the standardized residuals to the normal distribution. Fig. 6 indicates that the histogram of the standardized residuals is normally and independently distributed, with a mean of almost zero and a constant variance which accordingly we can fairly conclude that the regression model is accurate.

In Table 7 the unstandardized coefficients including coefficients of β and standard error of β and the standardized coefficients of all variables are presented.

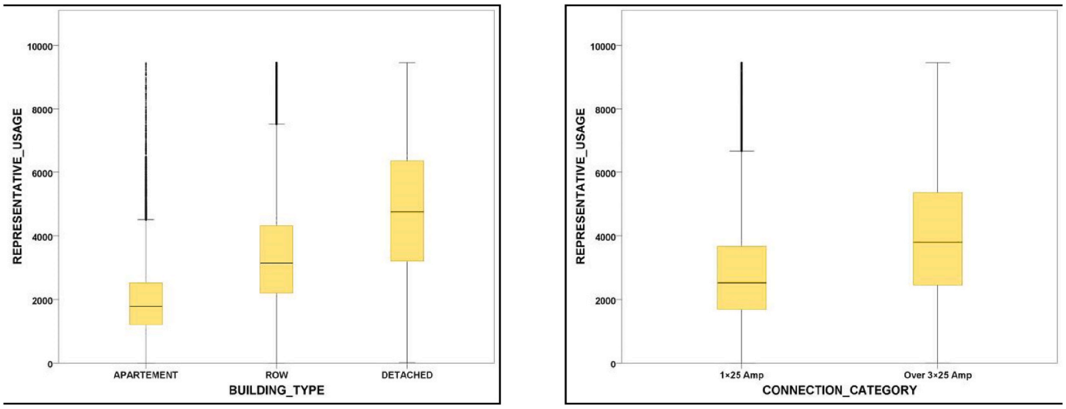


Fig. 5. Variations in the annual electricity usage of the residential buildings in terms of the categorical variables, (left) *BUILDING-TYPE* and (right) *CONNECTION-TYPE*.

Table 6
Results of the regression models in the residential sector.

Model	Variable categories	R	R Square	Adjusted R Square	Std. Error of the Estimate
Model 1	*Geospatial characteristics*	*0.312*	*0.097*	*0.097*	*1650.927*
Model 2	*All variable categories*	*0.530*	*0.281*	*0.281*	*1394.675*

a. Predictors: (Constant), HEIGHT, URBANIZATION_DEGREE, SOLAR_RADIATION, DENSITY.
b. Predictors: (Constant), GROUND_AREA, CONNECTIONI, APARTMENT, PARCEL_AREA, AGEIV, SOLAR_RADIATION, AGEIII, AGEII, AGEII, WOZ-VALUE, DENSITY, URBANIZATION_DEGREE, HEIGHT.

The results indicate that most of the predictors are employed in the regression model except for *ROW (BUILDING-TYPE)*, *FLAT* (ROOF-TYPE), and *ROOF-AREA*. It also revealed that all the applied variables are statistically significant at the 0.038 level. The sign of the *GROUND_AREA* and *URBANIZATION_DEGREE* is following what was expected. The standardized coefficients show that *GROUND_AREA* has the highest effect on the annual residential electricity demand. Using dummy coding to analyse the effect of *CONSTRUCTION_YEAR* on the residential electricity demand, it can be seen that building which constructed before 1990 have a lower electricity usage. The standardized coefficients also show that *DENSITY* and *WOZ_VALUE* have a very low effect on annual electricity usage. Using dummy coding to analyse the effect of the *BUILDING_TYPE* on the annual residential electricity usage, and taking a detached dwelling as a reference type, it can be seen that less electricity is used in apartments. It has been observed that *HEIGHT* has also a low significant effect on *residential electricity usage*.

Finally, based on the estimated coefficients (Table 7) the regression model predicting the annual electricity usage in the residential sector can be summarized as:

$$\boldsymbol{Annual\,residential\,electricity\,usage}\,(\boldsymbol{kWh}\,/\,\boldsymbol{year}) = 2500.268$$

$$+13.740\,GROUND\_AREA\,(m^2) - 522.495\,connectionI - 621.934\,apartment$$
$$-0.052\,PARCEL\_AREA\,(m^2) - 117.807\,ageIV$$
$$-0.001\,SOLAR\_INTENSITY(kWh\,/\,m\hat{}2\,) - 396.599\,ageIII\; - 413.845\,ageII$$

$$-458.878\,ageI + 0.000442\,WOZ\_VALUE(€) - 0.237\,DENSITY$$
$$+0.000005\,URBANIZATION\_DEGREE + 11.245\,HEIGHT \tag{4}$$

4.2. Non-residential sector

The NR sector comprises two main groups of buildings: service and industry. The regression analysis of this sector for both building groups are explained in detail in this section. A stepwise MLR analysis is performed to determine the statistically significant predictors of the annual electricity usage of the NR building connections and construct the prediction equations. As mentioned in Section 3.5, for NR buildings, two main categories of predictors are employed for the regression analysis, including physical and geospatial characteristics (Several variables in the constructed database are categorical variables. For applying these

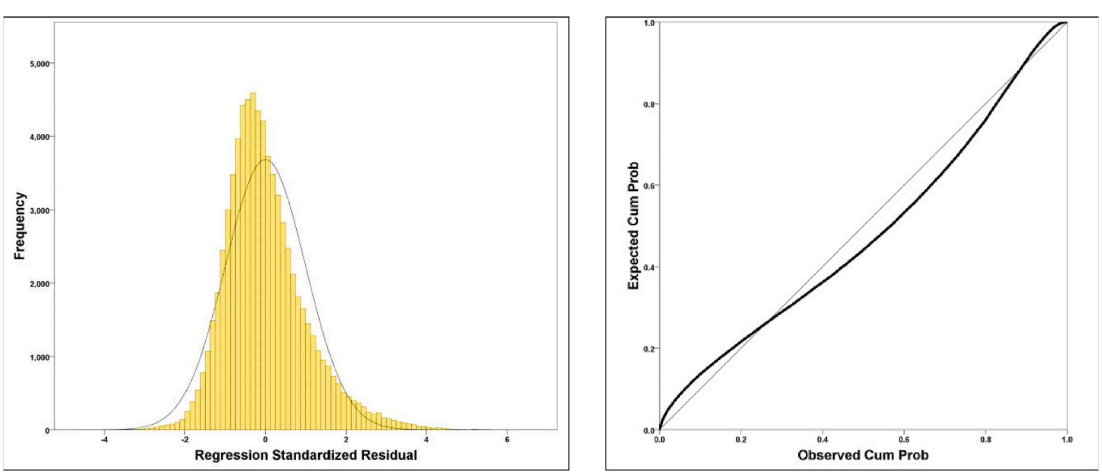


Fig. 6. Histogram of the standardized residuals and normal probability plots in the residential sector.

Table 7
Unstandardized and standardized coefficients of the regression model in the residential sector.

Coefficients[a,b]					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	β	Std. Error	Beta		
(Constant)	*2500.268*	*58.427*		*42.793*	*0.000*
GROUND_AREA	*13.740*	*0.232*	*0.289*	*59.259*	*0.000*
CONNECTIONI	*−522.495*	*13.652*	*−0.126*	*−38.272*	*0.000*
APARTMENT	*−621.934*	*24.958*	*−0.165*	*−24.919*	*0.000*
PARCEL_AREA	*−0.052*	*0.002*	*−0.080*	*−23.923*	*0.000*
AGEIV	*−117.807*	*18.302*	*−0.029*	*−6.437*	*0.000*
SOLAR_INTENSITY	*−0.000582*	*0.000070*	*−0.027*	*−8.362*	*0.000*
AGEIII	*−396.599*	*18.015*	*−0.098*	*−22.016*	*0.000*
AGEII	*−413.845*	*18.012*	*−0.112*	*−22.976*	*0.000*
AGEI	*−458.878*	*21.062*	*−0.100*	*−21.787*	*0.000*
WOZ-VALUE	*0.000442*	*0.000054*	*0.034*	*8.226*	*0.000*
DENSITY	*−0.237*	*0.061*	*−0.012*	*−3.902*	*0.000*
URBANIZATION_DEGREE	*5.497E-6*	*0.000002*	*0.009*	*2.524*	*0.012*
HEIGHT	*11.245*	*5.421*	*0.013*	*2.074*	*0.038*

[a] Dependent Variable: REPRESENTATIVE-USAGE.
[b] Excluded Variables: ROW, FLAT, ROOF-AREA.

variables in the subsequent analysis, they have to be transformed into the dummy variables. A dummy variable is a qualitative variable that can only take 0 or 1 and is used in the regression analysis instead of the categorical variable. While the dummy variable levels of the *BUILDING_TYPE* are *ROW* and *APARTMENT* (the detached house is considered as a base value), the levels of the *CONSTRUCTION_YEAR* are a stepwise increase of the oldest level up to the most recent level (1992 is considered as a base value). *CONNECTION_TYPE* is categorized into three levels based on the connection amperage and phase types (3 × 100 amp connection type is considered as a base value). *ROOF_TYPE* is also categorized into the flat and pitched roofs with *FLAT* as a dummy variable. Table 4 shows the categorical variables and corresponding assigned dummy variables.

Tables 4 and 5). MLR analysis is applied to determine the influence of predictive categories on the annual electricity usage variations. As the residential sector, before the MLR analysis is performed, the correlations analysis and collinearity diagnostics should be investigated. The parametric and nonparametric correlations between the dependent variable and the continuous and categorical predictors for both service and industrial buildings are presented in Table 14 in the Appendix. Regarding the continuous variables, for both service and industrial buildings while the correlations between the annual electricity usage and *GROUND-AREA* and *HEIGHT* are significant, the correlation coefficients between *PARCEL-AREA*, *SOLAR-INTENSITY,* and *DENSITY* with *annual electricity usage* is relatively low. Although the annual electricity usage of the service buildings has a relatively high correlation with the *PARCEL-AREA* and *ROOF-AREA*, however for the industrial buildings, it is insignificant. Annual electricity usage of the industrial buildings has a high correlation with the *URBANIZATION-DEGREE* which can be attributed to the fact that most of the high electricity-consuming industrial buildings are located at the urban peripheries as shown in Fig. 7.

With relation to the categorical variables, for both service and industrial buildings, the non-parametric correlations between the *annual*

Fig. 7. Visual representations of the correlation between *annual electricity usage* of the industrial buildings and *URBANIZATION-DEGREE*; the dots are representing the industrial buildings in and around the city, the colour spectrum from yellow (low) to red (high) depicts the *annual electricity usage* of the industrial buildings. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

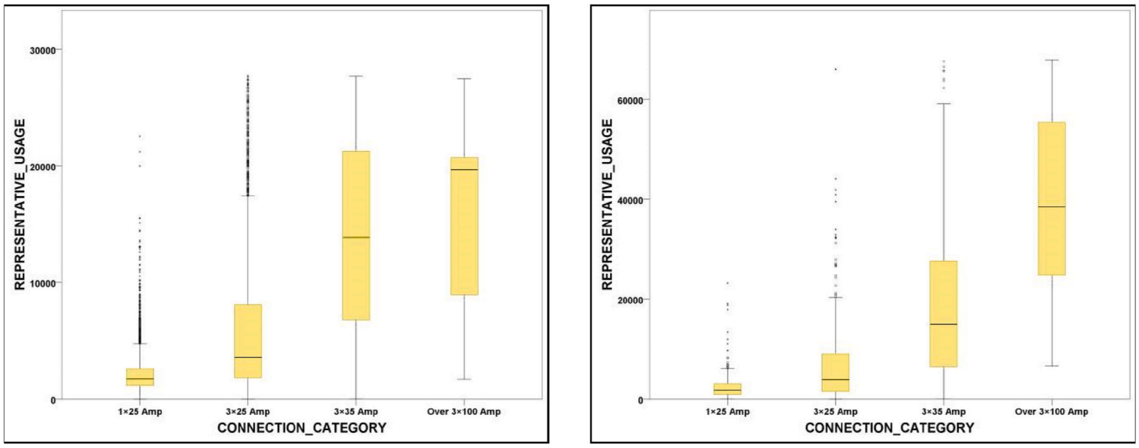


Fig. 8. Variations in the annual electricity usage of the non-residential sector based on the connection type (left) service buildings, (right) industrial buildings.

electricity usage and *CONNECTION-TYPE* are significant. Moreover, service building electricity usage has also relatively high correlations with the *BUILDING-TYPE* and *CONSTRUCTION-YEAR*. The annual electricity usage variations based on the *CONNECTION-TYPE* for both service and industrial buildings are presented in Fig. 8. As the graphs show for both sub-sectors the standard deviations of electricity usage per connection types are relatively large, particularly in medium and high voltage connections. The graphs also show that in this sector the annual electricity usage per connection types is significantly higher than the residential sector, particularly for the industrial connections.

The collinearity analysis between the predictors for both the service and industrial sub-sectors is presented in Tables 15 and 16 respectively in the Appendix. For both subsectors except for a limited number of variables, there is a relatively low level of correlations between the predictors which accordingly it is expected that most of the predictors will be employed in the MLR analysis. For both subsectors, *CONNECTION-TYPE* has a high degree of correlations with the *GROUND-AREA* and *HEIGHT* with (Pearson) correlation coefficients exceeding 30%. This can be explained by the fact that NR buildings with the higher ground area and height are connected to the high voltage electrical connections. In the service sub-sector *HEIGHT* also shows strong collinearity with the *BUILDING-TYPE* with the (Pearson) correlation coefficient exceeding 70%. One striking result of collinearity analysis in the industry sub-sector is the high correlations of the *URBANIZATION-DEGREE* with the other explanatory variables, such as *GROUND-AREA*, *ROOF-AREA*, *PARCEL-AREA*, *CONSTRUCTION-YEAR*, *CONNECTION-TYPE,* and *HEIGHT*. This again can be attributed to the fact that most of the industrial buildings with the higher ground area, height, and voltage electricity connections are located on the periphery of urban areas. Fig. 9 evidently demonstrates these correlations as most of the industrial buildings with the stated characteristics are located on the periphery of the Eindhoven municipality.

In this sector also to determine the contribution of geospatial characteristics in the electricity usage variability, the regression analysis is performed in two steps. In the first step only, the geospatial characteristics are implemented in the MLR analysis, and in the second step, all variable categories are employed. According to Table 8 for the service sub-sector, the geospatial characteristics account for 9.9% of the total variability in the building electricity usage. In total, all variables explain 39.4% of the total electricity usage variance in this sub-sector. In the industry sub-sector (Table 9) geospatial characteristics account for 8.7% of the total variability in the building electricity usage. In total in this

Table 8
Results of the regression models in the service subsector.

Model	Variable categories	R	R Square	Adjusted R Square	Std. Error of the Estimate
Model 1	*Geospatial characteristics*	*0.315*	*0.099*	*0.099*	*5318.056*
Model 2	*All variable categories*	*0.628*	*0.394*	*0.393*	*3268.018*

a. Predictors: (Constant), HEIGHT, SOLAR-INTENSITY.
b. Predictors: (Constant), CONNECTIONIII, CONNECTIONI, APARTEMENT, GROUND-AREA, AGEI, ROOF-AREA, PARCEL-AREA, ROW, SOLAR-INTENSITY, AGEII, AGEIV.

Fig. 9. Visual representations of the correlations between *URBANIZATION-DEGREE* with (left) GROUND-AREA, (right) HEIGHT, the dots are representing the industrial buildings in and around the city, the colour spectrum from yellow (low) to red (high) depicts ground area (left) and height (right) of the industrial buildings. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 9
Results of the regression models in the industry subsector.

Model	Variable categories	R	R Square	Adjusted R Square	Std. Error of the Estimate
Model 1	*Geospatial characteristics*	*0.294*	*0.087*	*0.085*	*13730.367*
Model 2	*All variable categories*	*0.655*	*0.429*	*0.426*	*10268.235*

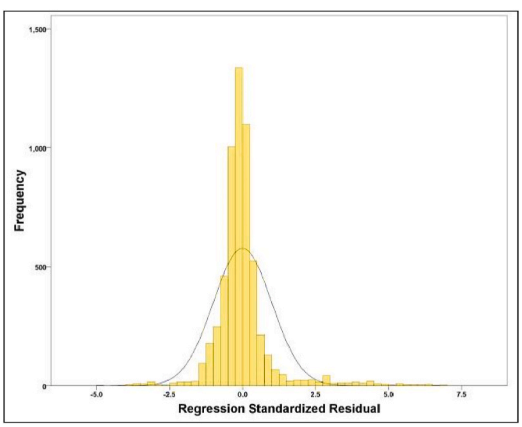
a. Predictors: (Constant), HEIGHT, URBANIZATION-DEGREE.
b. Predictors: (Constant), GROUND-AREA, CONNECTIONIII, CONNECTIONI, CONNECTIONII, URBANIZATION-DEGREE, AGEII.

sub-sector, all variables explain 42.9% of the total electricity usage variance.

The results revealed that in both sub-sectors the independent error assumptions have been satisfied and the Durbin-Watson values are within the specified boundaries. A 95% confidence intervals for β show that the models are reliable. The models also seem not to have collinearity, because tolerance values and VIFs are within the limits. Fig. 10 demonstrates that the standardized residuals' histograms in both service and industrial sub-sectors are normally and independently distributed, with a mean of almost zero and a constant variance. Accordingly, as the residential sector, we can conclude that the resulting models are fairly accurate.

The unstandardized coefficients including β and standard error of β and the standardized coefficients of regression variables in service and industry sub-sectors are presented in Table 10 and Table 11 respectively.

The results show that in the service sub-sector, most of the predictors are deployed in the regression model with the exceptions of *AGEIII*, *CONNECTIONII*, *FLAT*, *HEIGHT*, and *URBANIZATION_DEGREE*. It also revealed that all the deployed variables are statistically significant at the 0.037 level. For the industrial buildings as Table 11 shows only *GROUND_AREA*, *CONNECTIONIII*, *CONNECTIONI*, *CONNECTIONII*, *URBANIZATION_DEGREE,* and *AGEII* are deployed in the regression model which all are statistically significant at the 0.046 level. For the service buildings, the signs of *CONNECTION_TYPE* and *BUILDING_TYPE* are following what was expected. The standardized coefficients show that *CONNECTION_TYPE* has the highest effect on the *annual electricity usage* in the service buildings. Using dummy coding to analyse the effect of the *BUILDING_TYPE* on the annual electricity usage, it can be seen that less electricity is used in the apartments followed by the row buildings similar to the pattern observed in the residential sector. The *CONSTRUCTION_YEAR* dummy variables all have a positive sign showing that they have positive effects on electricity usage. The coefficients also show that the old buildings consume more electricity compared with the new ones. With one of the lowest beta values, the *SOLAR_INTENSITY* has a very low effect on the service sub-sector electricity usage, which was not expected. For the industrial buildings as Table 11 shows *GROUND_AREA*, *CONNECTION_TYPE*, *CONSTRUCTION_YEAR,* and *URBANIZATION_DEGREE* are the main variables that influence the industry sub-sector electricity usage. As expected, buildings with high voltage connections have higher electricity usage compared with the medium and low voltage connections. As discussed, the *URBANIZATION_DEGREE* has a positive effect on the



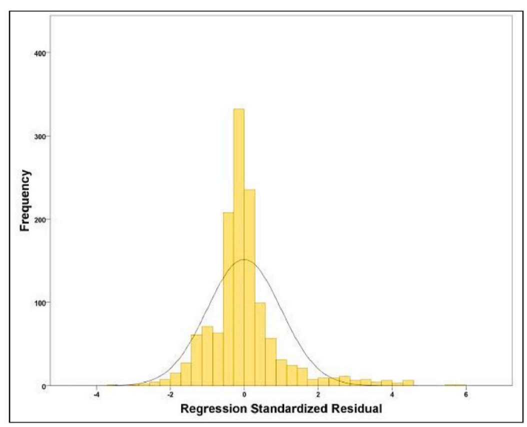


Fig. 10. Histogram of the standardized residuals in the non-residential sector, left: service, right: industry.

Table 10
Unstandardized and standardized coefficients of the regression model in the service sub-sector.

Coefficients[a b]					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	β	Std. Error	Beta		
(Constant)	*4594.254*	*314.230*		*14.621*	*0.000*
CONNECTIONIII	*9115.594*	*234.322*	*0.441*	*38.902*	*0.000*
CONNECTIONI	*−1901.681*	*104.696*	*−0.216*	*−18.164*	*0.000*
APARTEMENT	*−1257.921*	*234.961*	*−0.086*	*−5.354*	*0.000*
GROUND-AREA	*7.996*	*0.977*	*0.094*	*8.181*	*0.000*
AGEI	*978.516*	*144.354*	*0.088*	*6.779*	*0.000*
ROOF-AREA	*−0.280*	*0.046*	*−0.093*	*−6.057*	*0.000*
PARCEL-AREA	*0.132*	*0.028*	*0.075*	*4.744*	*0.000*
ROW	*1018.793*	*302.988*	*0.051*	*3.362*	*0.001*
SOLAR-INTENSITY	*−0.001*	*0.000447*	*−0.037*	*−3.334*	*0.001*
AGEII	*435.609*	*129.808*	*0.043*	*3.356*	*0.001*
AGEIV	*253.573*	*121.233*	*0.026*	*2.092*	*0.037*

[a] Dependent Variable: REPRESENTATIVE_USAGE.
[b] Excluded Variables: AGE3, CONNECTION2, FLAT, HEIGHT, URBANIZATION-DEGREE.

Table 11
Unstandardized and standardized coefficients of the regression model in the industry sub-sector.

Coefficients[a,b]					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	β	Std. Error	Beta		
(Constant)	*30689.562*	*2028.423*		*15.130*	*0.000*
GROUND-AREA	*6.631*	*0.807*	*0.202*	*8.222*	*0.000*
CONNECTIONIII	*−18551.249*	*1841.355*	*−0.622*	*−10.075*	*0.000*
CONNECTIONI	*−31790.741*	*1934.478*	*−1.049*	*−16.434*	*0.000*
CONNECTIONII	*−29049.912*	*1871.797*	*−1.052*	*−15.520*	*0.000*
URBANIZATION-DEGREE	*0.000390*	*0.000106*	*0.081*	*3.682*	*0.000*
AGEII	*−1685.244*	*841.959*	*−0.044*	*−2.002*	*0.046*

[a] Dependent Variable: REPRESENTATIVE_USAGE.
[b] Excluded Variables: ROOF-AREA, AGEIII, AGEIV, DENSITY, AGEI, SOLAR-INTENSITY.

electricity usage of industrial buildings.

Finally, based on the estimated coefficients, we can predict the amount of annual electricity which is consumed in each NR sector. The regression model predicting the annual electricity usage in the service sub-sector can be summarized as:

$$\textbf{\textit{Annual electricity usage in the service subsector}}\ (\textbf{\textit{kWh}}\ /\ \textbf{\textit{year}}) = 4594.254$$

$$\begin{aligned} &+9115.594\ CONNECTIONIII - 1901.681\ CONECCTIONI \\ &- 1257.921\ APARTEMENT + 7.996\ GROUND\_AREA\left(m^2\right) \\ &+\ 978.516\ AGEI - 0.280\ ROOF\_AREA\left(m^2\right) + 0.132\ PARCE\_AREA\left(m^2\right) \\ &+ 1018.793\ ROW \\ &- 0.001\ SOLAR\_INTENSITY(kWh\ /\ m2) + 435.609\ AGEII + 253.573\ AGEIV \end{aligned} \tag{5}$$

Also, the regression model predicting annual electricity usage in industrial sub-sector can be summarized as follows:

$$\textbf{\textit{Annual electricity usage in the industry subsector}}\ (\textbf{\textit{kWh}}\ /\ \textbf{\textit{year}}) = 30689.562$$

$$\begin{aligned} &+6.631\ GROUND\_AREA\left(m^2\right) - 18551.249\ CONNECTIONIII \\ &- 31790.741\ CONNECTIONI - 29049.912\ CONNECTIONII \\ &+ 0.000390\ URBANIZATION\_DEGREE\ - 1685.244\ AGEII \end{aligned} \tag{6}$$

5. Discussion

This study investigated the physical, geospatial and socioeconomic explanatory variables contributing to the electricity usage of the urban built environment in the municipality of Eindhoven. Due to the significant impact of building types on electricity usage, for each building sector, the analysis has been performed separately. The results in both sectors revealed the sizable effects of these characteristics on the annual electricity usage of building connections.

Within the physical characteristics of the residential sector, *ground area*, *building type*, *construction year* and *connection type* contributed the most to the annual electricity usage of the building connections, whereas within the geospatial characteristics, *height*, *solar intensity*, *urban density*, and *urbanization degree* were the significant variables. In this sector, all physical and geospatial characteristics account for 28.1% of the annual electricity usage variability which is in agreement with the research of [16]. Physical characteristics in overall demonstrate the considerable correlations with the annual electricity usage. According to Table 7 *Ground area* is the most important variable among the physical characteristics category which also has been reported in the literature by Refs. [20,23,60,62,63]. This can be explained by the effect of building size on the absolute electricity usage of buildings. This finding is also in agreement with research that considers the number of rooms and bedroom and their effects on electricity usage. According to Ref. [64] the effect of *ground area* is mainly connected to the building electricity usage for lighting, heating, and cooling, in particular in ones which their main heating sources are from electric boilers. According to the regression results, the effect of *building type* on residential sector electricity usage is significant (Table 7 and Fig. 5) which was in line with the research of [5,16,20,64–68]. Some previous studies reported that the electricity usage of residential dwellings will increase with the level of the detachment of the buildings [64]. This indicates that detached houses use more electricity compared with the row houses and apartments. These are in agreement with the result of this research since considering detached dwelling as a base value, other types of dwelling have a decreasing impact on electricity usage. Following the [16,20] apartments have the lowest demand as a result of their small size followed by row and detached houses. It is observed that the *construction year* has an increasing effect on the annual electricity usage, which is in accordance with the previous studies that mentioned the higher electricity usage is consumed in new houses [23,64,67]. According to Ref. [64], this can be explained by the fact that the penetration rate of new and high electricity consuming appliances such as heat pumps and electric vehicle in the new dwelling is higher. While [20] observed an insignificant effect of the construction year on the annual electricity usage, other studies report a decrease in dwelling electricity demand for newer houses due to the improvement in the construction technology, better thermal isolation and use of more efficient appliances for lighting and conditioning [5,64,65]. Others argue that the two above mentioned forces have cancelled out each other's impact, resulting in a uniform trend between dwelling electricity usage and the age of the buildings [20]. Results show that *connection type* also has a significant increasing correlation with the electricity usage of residential sector nonetheless it is not as significant as non-residential sectors since in this sector buildings are mostly connected to the low voltage connections and the connection type variations are not high (Table 5).

The electricity demand of residential buildings is also significantly influenced by socioeconomic factors such as *woz-value*. As mentioned in this research *woz-value* is applied as an indication of the household income level. It is observed that *woz-value* has a relatively significant positive correlation with the annual electricity usage, following the [64] which indicates that electrical energy consumption increases significantly with income. According to Refs. [16,20] the effect of income on household electricity usage can be explained by the tendency of high-income people to live in larger dwellings and having a greater number of electrical appliances.

Concerning the geospatial characteristics of residential buildings while *building height* and *urbanization degree* show a significant correlation with the annual electricity usage, the *urban density,* and *solar intensity* show lower correlations. In total the effects of geospatial characteristics on the annual electricity usage of the dwelling were lower than the physical characteristics, however since these characteristics are not considered in the previous studies, accordingly, examining their impacts is important. Since lighting is one of the main contributors to the electricity consumption of residential buildings and according to Ref. [16] nearly all dwellings will have some degree of lighting, it is expected that the solar intensity will have a significant correlation with the electricity demand. The results of this research revealed the negative (decreasing effect) correlation of solar intensity with the annual electricity usage of the residential buildings which corroborates the initial assumption. As mentioned, this relationship can be attributed to the availability of natural light in dwelling with higher solar intensity and consequently reduction in their electricity usage. The height of the buildings shows a strong increasing correlation with residential electricity usage. This can be explained by the relationship between height and number of building floors which consequently affects building electricity usage as has also been observed by Ref. [19]. *Urbanization degree* shows a significant positive correlation with the residential electricity usages. This can be attributed to the effect of temperature increase in urban areas compared to the urban surroundings due to the absorption of solar energy in urban areas which results in a higher temperature, also known as Urban Heat Island (UHI) [69–71]. The effect of u*rbanization degree* on residential electricity usage in this research is following the [70] which claims a significant role of UHI on household residential energy use. Finally, residential sector analysis revealed that higher density leads to lower electricity usage that means building in the high-density area use less electricity. This corresponds with the result of [17] which reported that urban density is one of the main factors influencing the energy use of cities.

The findings also show that both the physical and geospatial characteristics have significant effects on annual electricity usage. While in the service sub-sector all variables account for 39.4% of the annual electricity usage, in industrial buildings 42.6% of the building electricity variations can be attributed to the applied explanatory variables. In the service sub-sector, physical characteristics such as *connection type*, *ground area*, *building type*, *construction year,* and *roof and parcel area* contributed the most to the annual electricity usage, whereas within the geospatial characteristics (although all variables have significant correlations with the electricity usage) only the *solar intensity* is considered in the regression model (Table 10). The c*onnection type* is the most significant variable which largely describes the electricity usage variability in the service sub-sector. It can be possibly explained by the agreement between the physical capacity of the electrical connections and the amount of electricity which is used in the buildings. Service buildings mostly are connected to the low and medium voltage electrical connections which consequently lead to higher usage variability compared to the residential buildings. The building g*round area* is the second most important variable among the physical characteristics of service buildings. As the residential sector, this also can be explained by the effect of building size and scale on the absolute value of electricity usage. *Building type* also has a significant influence on the electricity usage of service buildings. Again, the electricity will increase with the level of the detachment of buildings which indicates that the stand-alone buildings consume more electricity compared with the row or flats. Contrary to the residential buildings, *construction year* have decreasing effects on the annual electricity usage of service buildings. This can be explained by the improvement in construction technology, better thermal isolation, and the use of more efficient appliances for lighting and conditioning. For the geospatial characteristics, as mentioned only solar intensity is included in the regression analysis of service buildings. Although *building height* had a strong correlation with the electricity usage, however, due to its strong multicollinearity with physical characteristics such as *building* and *connection type*, it was discarded from the regression analysis. It was expected that the solar intensity will have a decreasing correlation with the electricity demand. The results revealed the negative correlation of *solar intensity* with the electricity usage of service buildings which verifies the initial assumption. As mentioned, this relationship can be attributed to the availability of natural lighting in building with higher solar intensity and subsequently a reduction in electricity usage.

In the industrial sub-sector *connection type* and *ground area* appear to be the most important predictors of electricity usage followed by *urbanization degree* and *construction year*. Correlation analysis shows that in this sub-sector *connection type* has the highest correlation with the electricity usage variations as shown in Fig. 10. This can be attributed to the high variability of connection types in this sub-sector compared to the residential and service buildings. As Table 5 shows there is a high variation of connection types in this sector and buildings are also mostly connected to high voltage connections due to their higher electrical demand. As other building types, *Ground area* also explains high electricity variations in this sub-sector. As mentioned in Sec. 4.2 annual electricity usage of the industrial buildings have a high correlation with the *urbanization degree* which shows most of the high consuming industrial buildings are located in the non-urban areas (Fig. 7).

6. Conclusions

The main objective of this research paper was to examine the impacts of physical, geospatial and socioeconomic characteristics of the urban built environment on the annual electricity usage of building connections in both residential and non-residential sectors by integrating the urban spatial techniques with the statistical/regression bottom-up demand models. The findings in both sectors revealed the significant effects of these characteristics on the buildings annual electricity usage. The identified explanatory variables for both sectors can be summarized as follows:

- *Residential sector:*
  1. Physical and socioeconomic categories including ground area, building type, construction year, connection type and woz-value
  2. Geospatial category including building height, solar intensity, urban density, and urbanization degree
- *Non-residential sector*
  - o Service buildings:
    1. Physical category including connection type, ground area, building type, construction year, and roof and parcel area
    2. Geospatial category including height, urban density, and urbanization degree
  - o Industrial buildings:
    1. Physical category including connection type, ground area, and construction year
    2. Geospatial category including urbanization degree

Although the contribution of physical characteristics on the annual electricity usage is dominant, however, the geospatial variables such as solar intensity, urbanization degree, and urban density are also significant determinants (9.7% of the variation in electricity usage in the residential sector and 9.9% and 8.5% in the service and industrial buildings, respectively). This indicates the necessity to consider these characteristics besides the physical characteristics as well. This can also have significant policy implications to devise any energy-saving interventions and measures in the built environment towards a sustainable built environment and cities. Moreover, this has been intensified in recent years due to the transition toward the electrification of heating and cooling systems and the introduction of the EVs. Therefore, geospatial characteristics will play an important role in developing any future energy transition plans and it is recommended to consider these variables in the future studies more thoroughly.

In this study, the MLR method applied to predict the annual electricity demand on the urban scale. It is recommended that besides the linear method, nonlinear methods such as an artificial neural network (ANN) or real-time PCR as an alternative way will be deployed to further examine the impact of these variables on the electricity usage of the built environment and compare their forecasting performance. Furthermore, it is also possible to use the linear and nonlinear hybrid models and univariate time series to forecast electricity consumption. Another aspect that needs to be considered in future studies is the electricity usage load profile of building connections and explore the effects of these explanatory variables on the peak and off-peak hours. This has become more significant as the large-scale implementation of renewable energy technologies in the urban energy system leads to the growth in the intermittent generation of electricity which to balance these volatilities, identifying the peak times become very critical. Recent technological developments in the integration of digital processing and communication technologies such as smart meters with the electricity grid in both low and medium voltage levels have provided the possibility to collect large amounts of the time-series data. This data allows considering the varieties of physical and behavioural determinant factors to perform a detailed analysis of electricity load profiles of different building connections.

CRediT authorship contribution statement

Saleh Mohammadi: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Supervision, Project administration. **Bauke de Vries:** Writing – review & editing, Supervision. **Azarakhsh Rafiee:** Formal analysis, Investigation, Resources. **Masoud Esfandiari:** Formal analysis. **Eduardo Dias:** Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 12
Parametric and non-parametric correlations between annual electricity usage and continuous and categorical variables in the residential sector

Variables	Correlation coefficients	REPRESENTATIVE_USAGE
WOZ_VALUE	*Pearson Correlation*	*0.289***
	Sig. (2-tailed)	*0.000*
	N	*84648*
PARCEL_AREA	*Pearson Correlation*	*-.251***
	Sig. (2-tailed)	*0.000*
	N	*91810*
ROOF_AREA	*Pearson Correlation*	*−0.042***
	Sig. (2-tailed)	*0.000*
	N	*81116*
GROUND_AREA	*Pearson Correlation*	*0.47***
	Sig. (2-tailed)	*0.000*
	N	*93725*
HEIGHT	*Pearson Correlation*	*0.282***
	Sig. (2-tailed)	*0.000*
	N	*91638*
SOLAR-INTENSITY	*Pearson Correlation*	*−0.077***
	Sig. (2-tailed)	*0.000*
	N	*91485*
URBANIZATION-DEGREE	*Pearson Correlation*	*0.108***
	Sig. (2-tailed)	*0.000*
	N	*95695*
DENSITY	*Pearson Correlation*	*−0.013***
	Sig. (2-tailed)	*0.000*
	N	*94355*
ROOF_TYPE	*Spearman Coefficient*	*0.117***
	Sig. (2-tailed)	*0.000*
	N	*87490*
CONNECTION_TYPE	*Spearman Coefficient*	*0.259***
	Sig. (2-tailed)	*0.000*
	N	*95695*
BUILDING_CATEGORY	*Spearman Coefficient*	*0.443***
	Sig. (2-tailed)	*0.000*
	N	*89623*
CONSTRUCTION_YEAR	*Spearman Coefficient*	*0.037***
	Sig. (2-tailed)	*0.000*
	N	*95695*

**Correlation is significant at the 0.01 level (2-tailed).

Table 13
Collinearity analysis between the predictors in the residential sector

		WZ[a]	PA[b]	RA[c]	AR[d]	HE[e]	SR[f]	UD[g]	DE[h]	RT[i]	CT[j]	BT[k]
WZ[a]	Pearson Cor.											
	Sig. (2-tailed)											
	N											
PA[b]	***Pearson Cor.***	*-.225***										
	Sig. (2-tailed)	*.000*										
	N	*83450*										
RA[c]	***Pearson Cor.***	*.118***	*.013***									
	Sig. (2-tailed)	*.000*	*.000*									
	N	*73983*	*79992*									
AR[d]	***Pearson Cor.***	*.516***	*-.297***	*-.044***								
	Sig. (2-tailed)	*.000*	*.000*	*.000*								
	N	*84629*	*91368*	*80539*								
HE[e]	***Pearson Cor.***	*.183***	*-.193***	*-.621***	*.468***							
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*							
	N	*82865*	*90473*	*81004*	*91155*							
SR[f]	***Pearson Cor.***	*-.070***	*.036***	*-.022***	*-.095***	*-.037***						
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*						
	N	*82473*	*90214*	*81369*	*91007*	*91885*						
UD[g]	***Pearson Cor.***	*-.029***	*-.045***	*-.070***	*.162***	*.087***	*-.066***					
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*					
	N	*85587*	*94175*	*82914*	*95190*	*94004*	*93811*					
DE[h]	***Pearson Cor.***	*.004*	*.023***	*-.087***	*.000*	*.079***	*-.020***	*.008**				
	Sig. (2-tailed)	*.233*	*.000*	*.000*	*.947*	*.000*	*.000*	*.013*				
	N	*84673*	*92869*	*81784*	*93815*	*92832*	*92577*	*96812*				
RT[i]	***Pearson Cor.***	*.113***	*-.127***	*-.072***	*.178***	*.116***	*-.165***	*.051***	*.019***			
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*			
	N	*79307*	*86356*	*78180*	*87057*	*89629*	*87696*	*89644*	*88524*			
CT[j]	***Pearson Cor.***	*.243***	*-.163***	*.196***	*.306***	*.013***	*-.054***	*.051***	*-.004*	*.080***		
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.237*	*.000*		
	N	*85587*	*94173*	*82914*	*95190*	*93989*	*93796*	*98222*	*96797*	*89637*		
BT[k]	***Pearson Cor.***	*.322***	*-.235***	*-.415***	*.611***	*.733***	*-.081***	*.215***	*.058***	*.050***	*.179***	
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	
	N	*80803*	*87525*	*76777*	*89163*	*88386*	*87210*	*91263*	*90110*	*84670*	*91261*	
CY[l]	***Pearson Cor.***	*-.062***	*-.031***	*.197***	*.104***	*-.178***	*-.007**	*.158***	*-.003*	*.040***	*.079***	*-.095***
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.028*	*.000*	*.341*	*.000*	*.000*	*.000*
	N	*85587*	*94175*	*82914*	*95190*	*94004*	*93811*	*98237*	*96812*	*89644*	*98222*	*91263*

[a]WOZ_VALUE.
[b]PARCEL_AREA.
[c]ROOF_AREA.
[d]GROUND_AREA.
[e]HEIGHT.
[f]SOLAR_INTENSITY.
[g]URBANIZATION_DEGREE.
[h]DENSITY.
[i]ROOF_TYPE.
[j]CONNECTION_TYEP.
[k]BUILDING_TYPE.
[l]CONSTRUCTION_YEAR.
**Correlation is significant at the 0.01 level (2-tailed).

Table 14
Parametric and non-parametric correlations between annual electricity usage and continuous and categorical explanatory variables in the non-residential sectors

Variables	Correlation coefficients	REPRESENTATIVE_USAGE	
		Service	Industry
PARCEL_AREA	*Pearson Correlation*	*−0.150***	*−0.013*
	Sig. (2-tailed)	*0.000*	*0.642*
	N	*8471*	*1337*
ROOF_AREA	*Pearson Correlation*	*−0.191***	*0.069**
	Sig. (2-tailed)	*0.000*	*0.011*
	N	*8338*	*1354*
GROUND_AREA	*Pearson Correlation*	*0.328***	*0.454***
	Sig. (2-tailed)	*0.000*	*0.000*
	N	*8343*	*1329*
HEIGHT	*Pearson Correlation*	*0.319***	*0.293***
	Sig. (2-tailed)	*0.000*	*0.000*
	N	*8367*	*1318*
SOLAR_INTENSITY	*Pearson Correlation*	*−0.021***	*0.061**
	Sig. (2-tailed)	*0.000*	*0.036*
	N	*8329*	*1190*
URBANIZATION_DEGREE	*Pearson Correlation*	*−0.011***	*0.169***
	Sig. (2-tailed)	*0.000*	*0.000*
	N	*8960*	*1392*
DENSITY	*Pearson Correlation*	*0.012***	*−0.010*
	Sig. (2-tailed)	*0.000*	*0.708*
	N	*8795*	*1384*
ROOF_TYPE	*Spearman Coefficient*	*−0.103***	*0.007*
	Sig. (2-tailed)	*0.000*	*0.820*
	N	*7423*	*1143*
CONNECTION_TYPE	*Spearman Coefficient*	*0.487***	*0.613***
	Sig. (2-tailed)	*0.000*	*0.000*
	N	*8982*	*1392*
BUILDING_CATEGORY	*Spearman Coefficient*	*0.226***	
	Sig. (2-tailed)	*0.000*	
	N	*7268*	
CONSTRUCTION_YEAR	*Spearman Coefficient*	*−0.148***	*0.064**
	Sig. (2-tailed)	*0.000*	*0.016*
	N	*8982*	*1392*

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 15
Collinearity analysis between the predictors in the service sub-sector

		PA[b]	RA[c]	AR[d]	HE[e]	SR[f]	UD[g]	DE[h]	RT[i]	CT[j]	BT[k]
PA[b]	Pearson Cor.										
	Sig. (2-tailed)										
	N										
RA[c]	***Pearson Cor.***	*.691***									
	Sig. (2-tailed)	*.000*									
	N	*9299*									
AR[d]	***Pearson Cor.***	*-.142***	*-.155***								
	Sig. (2-tailed)	*.000*	*.000*								
	N	*8486*	*8309*								
HE[e]	***Pearson Cor.***	*-.072***	*-.138***	*.334***							
	Sig. (2-tailed)	*.000*	*.000*	*.000*							
	N	*9032*	*8898*	*8473*							
SR[f]	***Pearson Cor.***	*-.109***	*-.096***	*.005*	*.004*						
	Sig. (2-tailed)	*.000*	*.000*	*.660*	*.693*						
	N	*8978*	*8847*	*8442*	*9546*						
UD[g]	***Pearson Cor.***	*.242***	*.162***	*-.055***	*-.047***	*.045***					
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*					
	N	*9632*	*9525*	*9052*	*9706*	*9656*					
DE[h]	***Pearson Cor.***	*-.066***	*-.126***	*-.021*	*-.012*	*-.123***	*-.113***				
	Sig. (2-tailed)	*.000*	*.000*	*.053*	*.239*	*.000*	*.000*				
	N	*9459*	*9360*	*8891*	*9541*	*9492*	*10169*				
RT[i]	***Pearson Cor.***	*.031***	*.113***	*-.084***	*-.058***	*-.139***	*-.126***	*.034***			
	Sig. (2-tailed)	*.006*	*.000*	*.000*	*.000*	*.000*	*.000*	*.001*			
	N	*8042*	*8110*	*7564*	*8720*	*8526*	*8683*	*8538*			
CT[j]	***Pearson Cor.***	*-.131***	*-.114***	*.370***	*.360***	*-.031***	*-.033***	*.085***	*-.115***		
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.002*	*.001*	*.000*	*.000*		
	N	*9640*	*9542*	*9067*	*9718*	*9669*	*10347*	*10185*	*8694*		
BT[k]	***Pearson Cor.***	*-.186***	*-.225***	*.211***	*.744***	*.007*	*.024**	*-.001*	*-.081***	*.138***	
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.558*	*.035*	*.940*	*.000*	*.000*	
	N	*7246*	*7032*	*7570*	*7314*	*7276*	*7847*	*7694*	*6478*	*7845*	
CY[l]	***Pearson Cor.***	*.360***	*.345***	*-.064***	*-.129***	*-.149***	*.054***	*.062***	*-.044***	*-.063***	*-.284***
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*
	N	*9661*	*9564*	*9069*	*9744*	*9695*	*10373*	*10210*	*8720*	*10388*	*7852*

[b]PARCEL_AREA.
[c]ROOF_AREA.
[d]GROUND_AREA.
[e]HEIGHT.
[f]SOLAR_NTENSITY.
[g]URBANIZATION_DEGREE.
[h]DENSITY.
[i]ROOF_TYPE.
[j]CONNECTION_TYPE.
[k]BUILDING_TYPE.
[l]CONSTRUCTION_YEAR.
**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Table 16
Collinearity analysis between the predictors in the industry subsector

		PA[b]	RA[c]	AR[d]	HE[e]	SR[f]	UD[g]	DE[h]	RT[i]	CT[j]
PA[b]	Pearson Cor.									
	Sig. (2-tailed)									
	N									
RA[c]	***Pearson Cor.***	*.648***								
	Sig. (2-tailed)	*.000*								
	N	*1445*								
AR[d]	***Pearson Cor.***	*.004*	*.106***							
	Sig. (2-tailed)	*.879*	*.000*							
	N	*1378*	*1404*							
HE[e]	***Pearson Cor.***	*.102***	*.169***	*.393***						
	Sig. (2-tailed)	*.000*	*.000*	*.000*						
	N	*1411*	*1446*	*1352*						
SR[f]	***Pearson Cor.***	*-.113***	*.131***	*.091***	*.274***					
	Sig. (2-tailed)	*.000*	*.000*	*.001*	*.000*					
	N	*1281*	*1312*	*1223*	*1354*					
UD[g]	***Pearson Cor.***	*.281***	*.274***	*.126***	*.361***	*.048*				
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.077*				
	N	*1488*	*1510*	*1426*	*1500*	*1365*				
DE[h]	***Pearson Cor.***	*.073***	*-.045*	*-.035*	*-.099***	*-.211***	*-.042*			
	Sig. (2-tailed)	*.005*	*.085*	*.192*	*.000*	*.000*	*.094*			
	N	*1477*	*1499*	*1419*	*1490*	*1354*	*1570*			
RT[i]	***Pearson Cor.***	*.108***	*.177***	*-.054*	*−0.062**	*-.085***	*-.052*	*.093***		
	Sig. (2-tailed)	*.000*	*.000*	*.066*	*.024*	*.004*	*.059*	*.001*		
	N	*1230*	*1264*	*1181*	*1310*	*1177*	*1317*	*1307*		
CT[j]	***Pearson Cor.***	*-.079*	*.032*	*.543***	*.496***	*.161***	*.201***	*-.054**	*-.044*	
	Sig. (2-tailed)	*.003***	*.209*	*.000*	*.000*	*.000*	*.000*	*.031*	*.113*	
	N	*1480*	*1507*	*1424*	*1494*	*1359*	*1573*	*1562*	*1312*	
CY[l]	***Pearson Cor.***	*.356***	*.343***	*.125***	*.181***	*.025*	*.348***	*-.065**	*-.223***	*.122***
	Sig. (2-tailed)	*.000*	*.000*	*.000*	*.000*	*.364*	*.000*	*.011*	*.000*	*.000*
	N	*1488*	*1510*	*1426*	*1500*	*1365*	*1581*	*1570*	*1317*	*1573*

[b]PARCEL_AREA.
[c]ROOF_AREA.
[d]GROUND_AREA.
[e]HEIGHT.
[f]SOLAR_INTENSITY.
[g]URBANIZATION_DEGREE.
[h]DENSITY.
[i]ROOF_TYPE.
[j]CONNECTION_TYPE.
[l]CONSTRUCTION_YEAR.
**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

References

[1] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835, https://doi.org/10.1016/j.rser.2008.09.033.

[2] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (2012) 3586–3592, https://doi.org/10.1016/j.rser.2012.02.049.

[3] D. Robinson, C. Giller, F. Haldi, F. He, J. Kämpf, 520 : towards Comprehensive Simulation and Optimisation for More Sustainable Urban Design, 2008.

[4] M. Kavgic, a. Mavrogianni, D. Mumovic, a. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (2010) 1683–1697, https://doi.org/10.1016/j.buildenv.2010.01.021.

[5] D. Wiesmann, I. Lima Azevedo, P. Ferrão, J.E. Fernández, Residential electricity consumption in Portugal: findings from top-down and bottom-up models, Energy Pol. 39 (2011) 2772–2779, https://doi.org/10.1016/j.enpol.2011.02.047.

[6] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Proposal of a modeling approach considering urban form for evaluation of city level energy management, Energy Build. 39 (2007) 580–592, https://doi.org/10.1016/j.enbuild.2006.09.011.

[7] L. Shorrock, J. Dunster, The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock, Energy Pol. 25 (1997) 1027–1037, https://doi.org/10.1016/S0301-4215(97)00130-4.

[8] R. Rylatt, S. Gadsden, K. Lomas, Methods of predicting urban domestic energy demand with reduced datasets: a review and a new GIS-based approach, Build. Serv. Eng. Technol. 24 (2003) 93–102, https://doi.org/10.1191/0143624403bt061oa.

[9] P. Jones, J. Patterson, S. Lannon, Modelling the built environment at an urban scale—energy and health impacts in relation to housing, Landsc. Urban Plann. 83 (2007) 39–49, https://doi.org/10.1016/j.landurbplan.2007.05.015.

[10] A. Mavrogianni, M. Davies, M. Kolokotroni, I. Hamilton, A GIS-based bottom-up space heating demand model of the London domestic stock, in: Int. Build. Perform. Simul. Assoc. Build. Simul., 2009, pp. 1061–1067. Glasgow, UK, http://discovery.ucl.ac.uk/1302452/.

[11] T. Santos, N. Gomes, S. Freire, M.C. Brito, L. Santos, J.a. Tenedório, Applications of solar mapping in the urban environment, Appl. Geogr. 51 (2014) 48–57, https://doi.org/10.1016/j.apgeog.2014.03.008.

[12] A. Lippold, 3D city-scale modeling applied to sustainable design: history, example, and future, in: Proc. Policity Conf., Stuttgart's Wagenhallen, 2010.

[13] D. Robinson, N. Campbell, W. Gaiser, K. Kabel, a. Le-Mouel, N. Morel, J. Page, S. Stankovic, a. Stone, SUNtool – a new modelling paradigm for simulating and optimising urban sustainability, Sol. Energy 81 (2007) 1196–1211, https://doi.org/10.1016/j.solener.2007.06.002.

[14] A. Strzalka, M.Z.H. Siddiquee, U. Eicker, Potential of roof top PV-systems for supplying electricity in residential area Scharnhauser Park, 2009 1st, Int. Conf. Dev. Renew. Energy Technol. (2009) 1–4, https://doi.org/10.1109/ICDRET.2009.5454212.

[15] A. Strzalka, J. Bogdahn, V. Coors, U. Eicker, 3D City modeling for urban scale heating energy demand forecasting, HVAC R Res. 17 (2011) 526–539.

[16] F. McLoughlin, A. Duffy, M. Conlon, Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: an Irish case study, Energy Build. 48 (2012) 240–248, https://doi.org/10.1016/j.enbuild.2012.01.037.

[17] I. Larivière, G. Lafrance, Modelling the electricity consumption of cities: effect of urban density, Energy Econ. 21 (1999) 53–66, https://doi.org/10.1016/S0140-9883(98)00007-3.

[18] D. Ndiaye, K. Gabriel, Principal component analysis of the electricity consumption in residential dwellings, Energy Build. 43 (2011) 446–453, https://doi.org/10.1016/j.enbuild.2010.10.008.

[19] C. Bartusch, M. Odlare, F. Wallin, L. Wester, Exploring variance in residential electricity consumption: household features and building properties, Appl. Energy 92 (2012) 637–643, https://doi.org/10.1016/j.apenergy.2011.04.034.

[20] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy 55 (2013) 184–194, https://doi.org/10.1016/j.energy.2013.03.086.

[21] A. Soares, Á. Gomes, C.H. Antunes, Categorization of residential electricity consumption as a basis for the assessment of the impacts of demand response actions, Renew. Sustain. Energy Rev. 30 (2014) 490–503, https://doi.org/10.1016/j.rser.2013.10.019.

[22] J. Torriti, A review of time use models of residential electricity demand, Renew. Sustain. Energy Rev. 37 (2014) 265–272, https://doi.org/10.1016/j.rser.2014.05.034.

[23] K.J. Baker, R.M. Rylatt, Improving the prediction of UK domestic energy-demand using annual consumption-data, Appl. Energy 85 (2008) 475–482, https://doi.org/10.1016/j.apenergy.2007.09.004.

[24] Y.G. Yohanis, J.D. Mondol, A. Wright, B. Norton, Real-life energy use in the UK: how occupancy and dwelling characteristics affect domestic electricity use, Energy Build. 40 (2008) 1053–1059, https://doi.org/10.1016/j.enbuild.2007.09.001.

[25] K. Gram-Hanssen, C. Kofod, K.N. Petersen, Different everyday lives - different patterns of electricity use, Proc. 2004 Am. Counc. Energy Effic. Econ. Summerstudy Build. (2004) 13.

[26] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M.N. Assimakopoulos, R. Mitoula, S. Zerefos, Using intelligent clustering techniques to classify the energy performance of school buildings, Energy Build. 39 (2007) 45–51, https://doi.org/10.1016/j.enbuild.2006.04.018.

[27] J. O'Doherty, S. Lyons, R.S.J. Tol, Energy-using appliances and energy-saving features: determinants of ownership in Ireland, Appl. Energy 85 (2008) 650–662, https://doi.org/10.1016/j.apenergy.2008.01.001.

[28] R. Choudhary, Energy analysis of the non-domestic building stock of Greater London, Build, Environ. Times 51 (2012) 243–254, https://doi.org/10.1016/j.buildenv.2011.10.006.

[29] A.G. Gaglia, C.a. Balaras, S. Mirasgedis, E. Georgopoulou, Y. Sarafidis, D.P. Lalas, Empirical assessment of the Hellenic non-residential building stock, energy consumption, emissions and potential energy savings, Energy Convers. Manag. 48 (2007) 1160–1175, https://doi.org/10.1016/j.enconman.2006.10.008.

[30] F.E. Brown, P.A. Rickaby, H.R. Bruhns, P. Steadman, Surveys of nondomestic buildings in four English towns, Environ. Plann. Plann. Des. 27 (2000) 11–24, https://doi.org/10.1068/b2571.

[31] H. Bruhns, P. Wyatt, A data framework for measuring the energy consumption of the non-domestic building stock, Build. Res. Inf. 39 (2011) 211–226, https://doi.org/10.1080/09613218.2011.559704.

[32] N.D. Mortimer, a. Ashley, M. Elsayed, M.D. Kelly, J.H.R. Rix, Developing a database of energy use in the UK non-domestic building stock, Energy Pol. 27 (1999) 451–468, https://doi.org/10.1016/S0301-4215(99)00044-0.

[33] N.D. Mortimer, M.A. Elsayed, J.F. Grant, Patterns of energy use in nondomestic buildings, Environ. Plann. Plann. Des. 27 (2000) 709–720, https://doi.org/10.1068/b2573.

[34] C.H. Pout, N.- Deem, The national nondomestic buildings energy and emissions model, Environ. Plann. Plann. Des. 27 (2000) 721–732, https://doi.org/10.1068/bst12.

[35] H. Bruhns, P. Steadman, H. Herring, A database for modeling energy use in the non-domestic building stock of England and Wales, Appl. Energy 66 (2000) 277–297, https://doi.org/10.1016/S0306-2619(00)00018-0.

[36] H. Bruhns, P. Steadman, L. Marjanovic, A preliminary model of non-domestic energy use for England and Wales. http://eprints.ucl.ac.uk/13095/, 2006. (Accessed 7 January 2015).

[37] Y. Pan, R. Yin, Z. Huang, Energy modeling of two office buildings with data center for green building design, Energy Build. 40 (2008) 1145–1152, https://doi.org/10.1016/j.enbuild.2007.10.008.

[38] A. Aranda, G. Ferreira, M.D. Mainar-Toledo, S. Scarpellini, E. Llera Sastresa, Multiple regression models to predict the annual energy consumption in the Spanish banking sector, Energy Build. 49 (2012) 380–387, https://doi.org/10.1016/j.enbuild.2012.02.040.

[39] T. Doms, M.E. Dunne, Energy intensity, electricity consumption, and advanced manufacturing technology usage, Center Econ. Stud., U.S. Census Bureau (1993). http://ideas.repec.org/p/cen/wpaper/93-9.html. (Accessed 8 January 2015).

[40] A.D. Woodland, A micro-econometric analysis of the industrial demand for energy in NSW, Energy J. 14 (1993) 57–90. http://econpapers.repec.org/RePEc:aen:journl:1993v14-02-a04. (Accessed 8 January 2015).

[41] T.B. Bjørner, M. Togeby, H.H. Jensen, Industrial companies' demand for electricity: evidence from a micropanel, Energy Econ. 23 (2001) 595–617, https://doi.org/10.1016/S0140-9883(00)00084-0.

[42] S. Ebneyamini, M.R. Sadeghi Moghadam, Toward developing a framework for conducting case study research, Int. J. Qual. Methods 17 (2018), https://doi.org/10.1177/1609406918817954, 160940691881795.

[43] Bevolking | eindhoven (n.d.), https://www.eindhoven.nl/bevolking. (Accessed 7 February 2021).

[44] K. Nabielek, D. Hamers, D. Evers, Cities in The Netherlands, Facts and figures on cities and urban areas, PBL Netherlands Environ, Assess. Agency, 2016, pp. 1–18. https://www.pbl.nl/sites/default/files/downloads/PBL-2016-Cities-in-the-Netherlands-2470.pdf.

[45] R. Madlener, Y. Sunak, Impacts of urbanization on urban structures and energy demand: what can we learn for urban energy planning and urbanization management? Sustain. Cities Soc. 1 (2011) 45–53, https://doi.org/10.1016/j.scs.2010.08.006.

[46] Y. Yu, N. Zhang, J.D. Kim, Impact of urbanization on energy demand: an empirical study of the Yangtze River Economic Belt in China, Energy Pol. 139 (2020) 111354, https://doi.org/10.1016/j.enpol.2020.111354.

[47] P.M. Rich, {Using Viewshed models to calculate intercepted solar radiation: applications in ecology}, Am. Soc. Photogramm. Remote Sens. Tech. Pap. (1994) 524–529.

[48] P. Fu, P.M. Rich, A geometric solar radiation model with applications in agriculture and forestry, Comput. Electron. Agric. 37 (2002) 25–35, https://doi.org/10.1016/S0168-1699(02)00115-1.

[49] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, P. Patargias, Using principal component and cluster analysis in the heating evaluation of the school building sector, Appl. Energy 87 (2010) 2079–2086, https://doi.org/10.1016/j.apenergy.2009.12.007.

[50] J.W. Tukey, exploratory data analysis, Addison-Wesley Publishing Company, 1977. http://books.google.nl/books/about/Exploratory_Data_Analysis.html?id=UT9dAAAAIAAJ&pgis=1. (Accessed 12 January 2015).

[51] D.C. Hoaglin, B. Iglewicz, J.W. Tukey, Performance of some resistant rules for outlier labeling, J. Am. Stat. Assoc. 81 (1986) 991–999, https://doi.org/10.2307/2289073.

[52] D.C. Hoaglin, B. Iglewicz, Fine-tuning some resistant rules for outlier labeling, J. Am. Stat. Assoc. 82 (1987) 1147–1149, https://doi.org/10.2307/2289392.

[53] G. Tso, A study of domestic energy usage patterns in Hong Kong, Energy 28 (2003) 1671–1682, https://doi.org/10.1016/S0360-5442(03)00153-1.

[54] A. Al-Garni, A regression model for electric-energy-consumption forecasting in Eastern Saudi Arabia, Energy 19 (1994) 1043–1049, https://doi.org/10.1016/0360-5442(94)90092-2.
[55] M. Ranjan, V.K. Jain, Modelling of electrical energy consumption in Delhi, Energy 24 (1999) 351–361, https://doi.org/10.1016/S0360-5442(98)00087-5.
[56] F. Egelioglu, A.A. Mohamad, H. Guven, Economic variables and electricity consumption in Northern Cyprus, Energy 26 (2001) 355–362, https://doi.org/10.1016/S0360-5442(01)00008-1.
[57] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (2007) 1761–1768, https://doi.org/10.1016/j.energy.2006.11.010.
[58] M.R. Braun, H. Altan, S.B.M. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the UK, Appl. Energy 130 (2014) 305–313, https://doi.org/10.1016/j.apenergy.2014.05.062.
[59] M.F. Poulsen, J. Forrest, Correlates of energy use: domestic electricity consumption in Sydney, Environ. Plan. A. 20 (1988) 327–338, https://doi.org/10.1068/a200327.
[60] M. Filippini, S. Pachauri, Elasticities of electricity demand in urban Indian households, Energy Pol. 32 (2004) 429–436, https://doi.org/10.1016/S0301-4215(02)00314-2.
[61] F. Halicioglu, Residential electricity demand dynamics in Turkey, Energy Econ. 29 (2007) 199–210, https://doi.org/10.1016/j.eneco.2006.11.007.
[62] L. Nielsen, How to get the birds in the bush into your hand, Energy Pol. 21 (1993) 1133–1144, https://doi.org/10.1016/0301-4215(93)90263-F.
[63] S. Zhou, F. Teng, Estimation of urban residential electricity demand in China using household survey data, Energy Pol. 61 (2013) 394–402, https://doi.org/10.1016/j.enpol.2013.06.092.
[64] R.V. Jones, A. Fuertes, K.J. Lomas, The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings, Renew. Sustain. Energy Rev. 43 (2015) 901–917, https://doi.org/10.1016/j.rser.2014.11.084.
[65] E. Leahy, S. Lyons, Energy use and appliance ownership in Ireland, Energy Pol. 38 (2010) 4265–4279, https://doi.org/10.1016/j.enpol.2010.03.056.
[66] I.G. Hamilton, P.J. Steadman, H. Bruhns, A.J. Summerfield, R. Lowe, Energy efficiency in the British housing stock: energy demand and the homes energy efficiency database, Energy Pol. 60 (2013) 462–480, https://doi.org/10.1016/j.enpol.2013.04.004.
[67] P. Tiwari, Architectural, demographic, and economic causes of electricity consumption in Bombay, J. Pol. Model. 22 (2000) 81–98, https://doi.org/10.1016/S0161-8938(98)00003-9.
[68] M. Bedir, E. Hasselaar, L. Itard, Determinants of electricity consumption in Dutch dwellings, Energy Build. 58 (2013) 194–207, https://doi.org/10.1016/j.enbuild.2012.10.016.
[69] A.G. Touchaei, M. Hosseini, H. Akbari, Energy savings potentials of commercial buildings by urban heat island reduction strategies in Montreal (Canada), Energy Build. 110 (2016) 41–48, https://doi.org/10.1016/j.enbuild.2015.10.018.
[70] S. Arifwidodo, O. Chandrasiri, Urban heat island and household energy consumption in bangkok, Thailand, Energy Procedia 79 (2015) 189–194, https://doi.org/10.1016/j.egypro.2015.11.461.
[71] H. Radhi, S. Sharples, E. Assem, Impact of urban heat islands on the thermal comfort and cooling energy demand of artificial islands—a case study of AMWAJ Islands in Bahrain, Sustain. Cities Soc. 19 (2015) 310–318, https://doi.org/10.1016/j.scs.2015.07.017.