Preprint, the final version, of the article that has been transferred to the journal's production team.

To cite this article: Afaifia, M., Djiar, K. A., Bich-Ngoc, N., & Teller, J. (2021). An energy consumption model for the Algerian residential building's stock, based on a triangular approach : Geographic Information System (GIS), regression analysis and hierarchical cluster analysis. *Sustainable Cities and Society*, 103191. https://doi.org/10.1016/j.scs.2021.103191

An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic Information System (GIS), regression analysis and hierarchical cluster analysis.

Abstract

Modelling residential energy consumption (REC) represents a key step towards the implementation of energy transition policies for more sustainable cities. Developing such policies requires considering the characteristics of the residential building stock (RBS). In the literature, REC modelling is generally applied on a single or set of cities-provinces, through limited approaches, using data from one typical year. In this paper, an energy consumption model for the entire Algerian RBS is developed through a triangular approach combining Geographic Information System, regression analysis and hierarchical clustering, applied to all provinces from 1995 to 2018. This allows mapping the spatial-temporal distribution of REC and RBS, developing a REC model, and dividing all provinces into clusters based on their REC behaviour. Provinces are aggregated into four clusters and four provinces are identified as archetypes. The results highlight that, besides the size of the RBS, REC is highly dependent on the electricity and gas connection rates. However, the influence of GDP and urban density only play a minor role. This can be explained by the evolving demands in thermal comfort associated with access to energy networks. The likely impact of increased gas and electricity connection represents a crucial factor in the design of energy policies.

Keywords

Residential energy consumption (REC); housing stock; energy modelling; Multiple Linear Regression (MLR); spatialisation ; Clustering; Algeria

Nomenclature

APRUE	Agence Nationale pour la Promotion et la Rationalisation de l'Utilisation de l'Energie
CNERIB	Centre National d'Etudes et de Recherches Intégrées du Bâtiment
DTR	Document Technique Réglementaire
GIS	Geographic Information System
Log	Logarithme
MHUV	Ministère de l'Habitat, de l'Urbanisme et de la Ville
QGIS	Quantum GIS
R^2	coefficient of determination
RBS	Residential building stock
REC	Residential energy consumption
Toe	Tonne of oil equivalent

1. Introduction

According to the final energy consumption decomposition established by IEA (2020), the residential buildings consumes 20% of total energy consumption. In Algeria, the largest country of Maghreb with a population of over 43 million inhabitants in 2019 (ONS, 2018), the RBS consumes an average of 50.4 million tons oil equivalent (Mtoe). This figure represents 36 % of total primary energy, followed by 30.6 % and 22.7 % absorbed by transport and industry sectors respectively (Ministère de l'Energie, 2020). If RBS represents the most energy-consuming sector in Algeria, it is also one of the main potential source of energy savings. To manage REC, the Algerian government implemented a national energy management policy through the ECO-BAT housing programme. From 2011, this programme involved the construction of 600 new housing units, with high-energy performance, in all climate zones (APRUE, 2015a). For the Algerian government, it represents a first initiative to developing a large-scale strategy dedicated to improving energy efficiency.

Yet, in addition to building new high-energy performance housings, it is necessary to improve the performances of the existing RBS. The 2030 energy efficiency programme is supposedly set toward improving the thermal insulation of buildings (APRUE, 2015b). It focuses on developing solar water heaters, generalising the use of low-energy lamps, and renovating 4000 dwellings. However, the thermal renovation was rapidly abandoned given the absence of adequate energy planning to orient the actions. The energy transition of the residential sector indeed relies on the ability of policy makers, energy planners, and energy assessors to develop strategies based on an analysis of REC at national and local scale (Swan & Ugursal, 2009; Howard et al, 2012). In Algeria, data on REC is not readily available at a provinces' scale but it is grouped together with the energy consumptions of other sectors.

In order to bridge this gap, our research proposes a REC model that combines a regression analysis with GIS and hierarchical clustering, to guide the development of energy efficiency scenarios at the national and local level. It is based on estimated RBS and REC for all Algerian provinces from 1995 to 2018.

The main aims of this paper are:

1- To develop a database of REC and RBS covering the whole Algerian territory, between 1995-2018, considering the different energy sources consumed by RBS ;

2- To provide a spatial-temporal analysis of REC and RBS distribution at a provincial scale, using GIS ;

3- To identify the main variable influencing REC through a Multiple Linear Regression (MLR) ;

4- To characterize the REC profile of each province and identify archetypical provinces to use for meso-scale planning and analysis, through a hierarchical clustering.

The originality of this research lies in the use of a combined approach for energy modelling, in line with the recommendations of Torabi Moghadam (2017). The originality of the present study is based on a thorough work of collection, synthesis, quantification and analysis of the most relevant spatial and statistical REC data in Algeria.

Our paper is divided into six main sections. The first section sets the background of the study. The second section presents the literature review based on the examination of different approaches to model REC. The third part explains the research methodology. Fourthly, the paper discusses the results of our triangular approach. The fifth section validates the REC model. Finally, the last part presents the conclusion and gives an outline of the future work.

2. Literature review

REC studies require a serious effort of modelling to understand energy consumption patterns at different scales, in order to support energy transition strategies. Swan & Ugursal (2009) reviewed techniques for REC modelling by grouping them into two types of approaches, top-down and bottom-up approaches. The use of one or another approach depends on the available input information and the desired output scenarios. Both approaches may be based on three methods for modelling energy consumption: the physics-based, data-driven and hybrid methods (Bourdeau et al., 2019). As our research is designed to develop and process a large RBS/REC database, this paper will mainly focus on data-driven methods used for REC modelling (Foucquier et al., 2013). The literature is classified into three groups:

- Studies based on statistical approaches to model REC
- Studies based on spatial analyses using GIS
- Studies that combine several approaches and methods

Fig 1 summarises the review of the most relevant energy modelling methods, based primarily on the sample of data used, the case studies scale, the energy type consumption, and the employed method.

		Sample and scale of data				Scale of case studies				Response variable (Energy type)			
Authors	Statistical approaches	Representative sample	Neighbourhoods	City/Province	National	Representative day	Month	Typical year	Study period	1	2	3	4
(Al-Ghandoor et al. 2009)	Multiple linear regression								1985–2006				
(Catalina et al., 2013)	Multiple linear regression												
Bianco et al. (2013)	Multiple linear regression								1970–2007				
Kavousian et al. (2013)	Multiple linear regression	1628 households											
Štreimikienė (2014)	Correlation analysis								2001–2009				
Tso & Guan (2014)	Multiple linear regression	10838 cases						2009					
Rhodes et al. (2014)	Clustering and Probit regression	103 households		Austin				2012-2013					
Fumo & Rafe Biswas (2015)	Linear regression and LMR	TxAIRE research house					June 2013						
Hsu (2015)	Penalized regression	3941 multifamily housing		New York City				2010 Census and 2011 Survey					
Walter & Sohn (2016)	Multiple linear regression	742,500 residential building						One full year					
Chen et al. (2016)	Correlation analysis				Several countries				2001–2008				
Williams & Gomez (2016)	Linear regression Regression trees Multiple linear regression	426,305 single-family houses		Texas			January 2010 to December 2013						
Biswas et al. (2016)	Neuronal network	TxAIRE pilot houses				Summer days		2013					
Sepehr et al. (2018)	Mathematical models	149 residential subscribers				Daily		2016's summer					
Kim et al. (2020)	Multiple linear regression Decision tree	71 apartments		Seoul				October 2017 to September 2018					
Iraganaboina & Eluru (2021)	Multiple level regression	4000 households						2015					

1: Electricity
2:Natural gas
3:Fuels
4: All energy use

Fig. 1. The summarised review of the most relevant energy modelling methods

The first group is mainly based on statistical approaches to model REC, analyse interactions, forecast REC, and provide recommendations to policy-makers and researchers. The most significant works in this category are (Al-Ghandoor et al., 2009; Biswas et al., 2016; Catalina et al., 2013; Chen et al., 2016; Fracastoro & Serraino, 2011; Fumo & Rafe Biswas, 2015; Hsu, 2015; Iraganaboina & Eluru, 2021; Kavousian et al., 2013; Kim et al., 2020; Rhodes et al., 2014; Sepehr et al., 2018; Štreimikienė, 2014; Theodoridou et al., 2011; Tso & Guan, 2014; Walter & Sohn, 2016; Williams & Gomez, 2016).

Multiple Linear Regression is the main method used in statistical approach for REC modelling. This observation is also highlighted by Torabi Moghadam et al. (2017), and very recently by Mata et al. (2021). REC depends on several independent variables such as the socio-economic, housing, appliance, climate and occupant behaviour factors (Jones et al., 2015; Mora et al., 2018). Table 1 identifies the main used variables in REC analysis. These variables can be grouped along five main dimensions influencing REC: socio-demographic, economic, climate, building and urban dimensions.

Table 1

A summary of the main independent variables used in statistical approaches

Dimensions	Independent variables	References
Socio-demography	Population	(Al-Ghandoor et al., 2009; Bianco et al., 2013; Chen et al., 2016; Theodoridou et al., 2011)
	Household size	(Iraganaboina & Eluru, 2021; Kim et al., 2020; Tso & Guan, 2014)
	Occupant density	Kim et al. (2020)
	Households characteristics	Iraganaboina & Eluru (2021)
	The number of residents	Sepehr et al. (2018)
	Homeowner	Rhodes et al. (2014)
Economy	GDP and GDP/capita	Bianco et al. (2013)
	Income level	Kavousian et al. (2013)

	Energy price	Al-Ghandoor et al. (2009)
Climate	Temperature	(Al-Ghandoor et al., 2009; Biswas et al., 2016; Catalina et al., 2013; Fumo & Rafe Biswas, 2015; Kim et al., 2020; Williams & Gomez, 2016)
	Humidity	Williams & Gomez (2016)
	Solar radiation	(Biswas et al., 2016; Fumo & Rafe Biswas, 2015)
	Climate zone	Walter & Sohn (2016)
	Climate change	Chen et al. (2016)
	Heating degree days/ Cooling degree days	(Iraganaboina & Eluru, 2021; Kavousian et al., 2013)
Building	Year of construction (age)	(Hsu, 2015; Kavousian et al., 2013; Kim et al., 2020; Theodoridou et al., 2011)
	Floor area	(Fracastoro & Serraino, 2011; Kavousian et al., 2013; Kim et al., 2020; Theodoridou et al., 2011)
	Building envelop	(Catalina et al., 2013; Theodoridou et al., 2011; Williams & Gomez, 2016)
	Housing type/size	(Hsu, 2015; Tso & Guan, 2014)
	Energy system and appliance use	(Kim et al., 2020; Sepehr et al., 2018; Štreimikienė, 2014; Tso & Guan, 2014)
Urban	Urbanization rate Location characteristics Settlement pattern Density and mixed use	(Chen et al., 2016; Dujardin et al., 2014; Iraganaboina & Eluru, 2021)

The second group of literature is mainly based on GIS approaches that create and analyse REC's spatial dimension, considering a number of variables as for instance urban density or climate zone. In this context, several studies based on spatial approaches associated with energy modelling have progressively developed in recent years, in order to help energy planning such as (Caputo et al., 2013; Caputo & Pasetti, 2017; de Santoli et al., 2019; Dujardin et al., 2014; Evola et al., 2016; Fichera et al., 2016; Groppi et al., 2018; Howard et al., 2012; Mattinen et al., 2014; Österbring et al., 2016). In this second group, GIS are used to visualise REC distribution at different scales. They also help analysing the influence of spatial factors, like urban density, mixed-use, age and typology of buildings, on REC.

The third group involves studies that combine several methods and approaches, for multi-disciplinary analysis and complementarity to provide adequate information. Accordingly, studies that combined more than one approach to residential energy modelling are as follows (Ahmed Gassar et al., 2019; Aydinalp-Koksal & Ugursal, 2008; Brounen et al., 2012; Delmastro et al., 2016; Fonseca & Schlueter, 2015; Gianniou et al., 2018; Mastrucci et al., 2014; Nishimwe & Reiter, 2021; Nouvel et al., 2015; Reiter & Marique, 2012; Sachs et al., 2019; Salari & Javid, 2016; Torabi Moghadam et al., 2018; Zhu et al., 2013).

The combination between Multiple Linear Regression and GIS approaches is the most widely used in REC modelling. This observation is also highlighted by Torabi Moghadam, Delmastro, Corgnati, and Lombardi (2017). The combination of hierarchical clustering with GIS is often used on energy modelling, as it allows organising the REC according to different patterns.

In the Algerian context, REC modelling is not yet very common. One of the few researches on energy consumption was conducted by Bélaïd & Abderrahmani (2013), who examined the relationship between electricity consumption, oil prices and economic growth between 1971 and 2010. The results show that there is a bi-directional causal relationship between electricity consumption and GDP. However, this consumption is mainly focused on one energy source, across all sectors. A few years later, Ghedamsi et al. (2016) developed a REC model based on a bottom-up approach. Their estimation of the annual REC is based on annual costs of energy consumption for heating and cooling. The evaluation and visualisation is conducted according to degree-day method with GIS, on data collected in 2008. The results show that the number of dwellings is the main effective factor in REC. In 2017, Amri (2017) used the cointegration test to examine the impact of Algeria's economic growth on energy consumption from 1980 to 2012. The results show that Algeria's economic development is strongly linked to the abundance of non-renewable fossil resources such as natural gas and oil products. One year later, Boukarta & Berezowska-Azzag (2018) highlighted the influence of housing and households characteristics on REC at the Algerian province of Djelfa, using MLR and GIS.

The present paper is based on the following four characteristics, based on the main shortcomings of REC modelling in previous studies described in the literature.

Longitudinal data sets. Most models only analysed a partial REC data sets, i.e. the data is usually focuses on one typical year, month, or a representative day. Longitudinal database is required for analysing the influence of long-term structural factors, like GDP or the growth of the RBS on REC over time. This research proceeded to the estimation of the evolution of REC and RBS over a 23-year period.

National scale analysis. Many studies model the REC at the scale of a neighbourhood, district, city or a town, but few studies model the national REC across all its representative units. Considering the entire national territory is essential in those countries, like Algeria, where energy planning and policies are highly centralized in order to transfer results from research to the implementation of energy policies. Accordingly this paper proceeds to REC modelling across all the 48 provinces that make up the Algerian territory.

Triangulation of approaches. A number of REC models are based on mono- or coupled approaches. This paper proposes the implementation of a triangulation of regression analysis, GIS and hierarchical clustering. This will help to address challenges and recommendations made by (Torabi Moghadam et al., 2017) and identify regional archetypes where to test possible energy planning policies.

Integration of all energy sources. Mata et al. (2021) highlights that most REC models focus on electricity consumption, while RBS may consume other energy source such as gas, LPG, etc. The present paper proposes to address all energy sources consumed in Algerian RBS, considering their evolution over time.

Therefore, this paper sets itself the objective to answer three main research questions:

1. How to analyse and spatialize the REC based on all energy source used in Algerian RBS, at the level of the 48 provinces?
2. What is the relative impact of GDP, the growth of the RBS, climate, connection to energy networks (gas and electricity), household size and urban density on REC at the province level between 1995 and 2018?
3. How to classify the 48 Algerian provinces into clusters based on their REC profile over time?

3. Methodology

The present study is based on a triangulation of (1) Geographic Information System (GIS), (2) regression analysis and (3) hierarchical cluster analysis, applied to the REC of Algeria's 48 provinces between 1995-2018. It goes through three main steps as follows:

a- It calculates and creates maps of spatial-temporal variability of REC and RBS using GIS, and analyses the energy consumption per housing unit.

b- It develops a REC model using Multiple Linear Regression.

c- It presents a hierarchical clustering of provinces according to their REC profile.

The conceptual framework includes data collection, data processing, development model, and finally, validation. Fig. 2 shows the conceptual framework, which can be divided into four main steps. Each step is described in detail in the following sections.

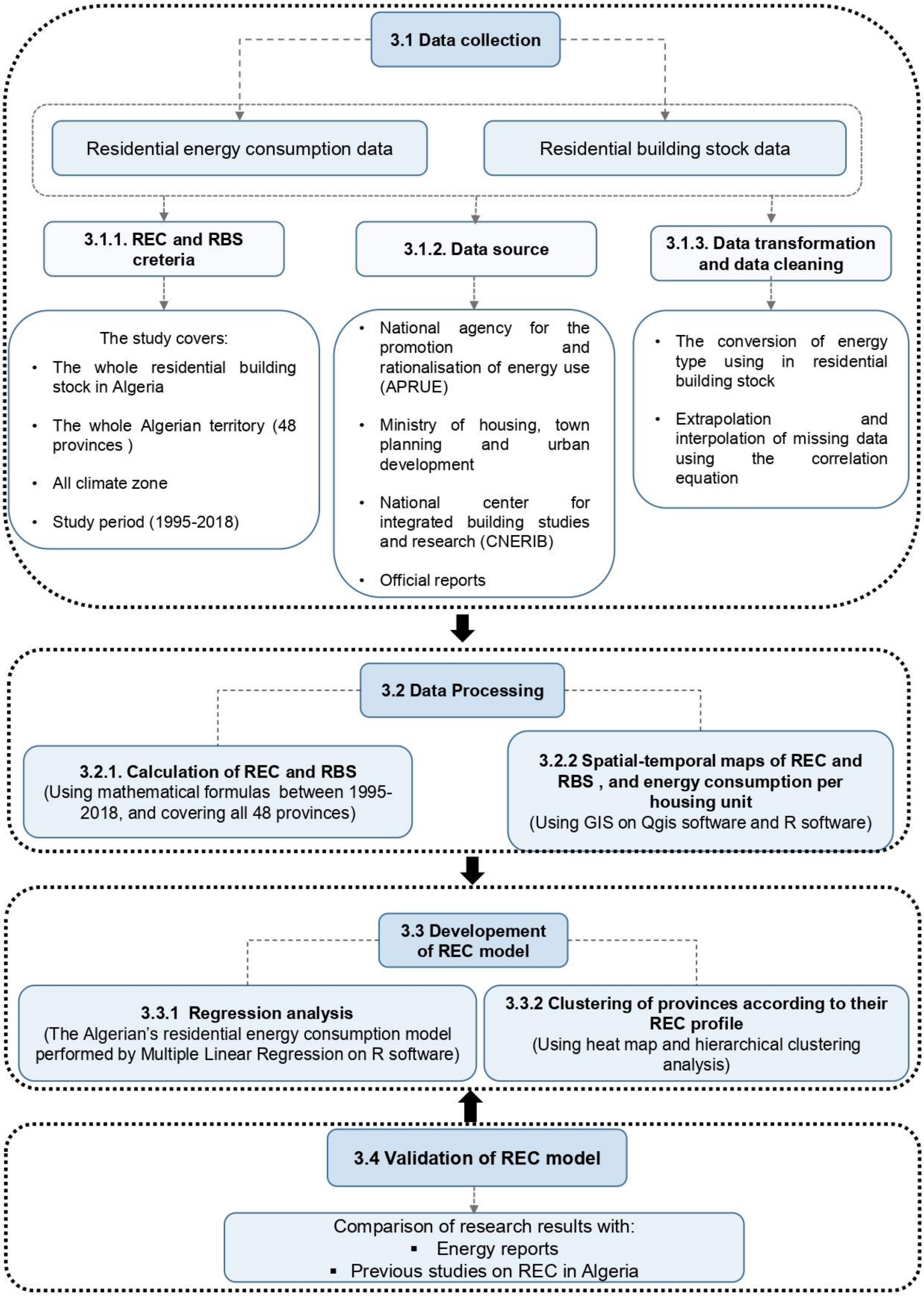


Fig. 2. The study conceptual framework.

3.1. Data collection

The first key step in this study is data collection. It allows the identification of important criteria and the selection of the most relevant sources for data collection.

3.1.1. REC and RBS criteria

Firstly, this research covers the whole RBS in Algeria that is composed of more than 8.5 million housings in 2017 (APRUE, 2017). Secondly, the study covers all 48 provinces that represent the entire Algerian territory (2.382 million km^2) (see Fig. 3). According to the thermal regulation applicable to residential buildings, the Algerian territory is divided into six climatic zones. Fig 4 summarises the description of characteristics climate zones, extracted from the APRUE database.

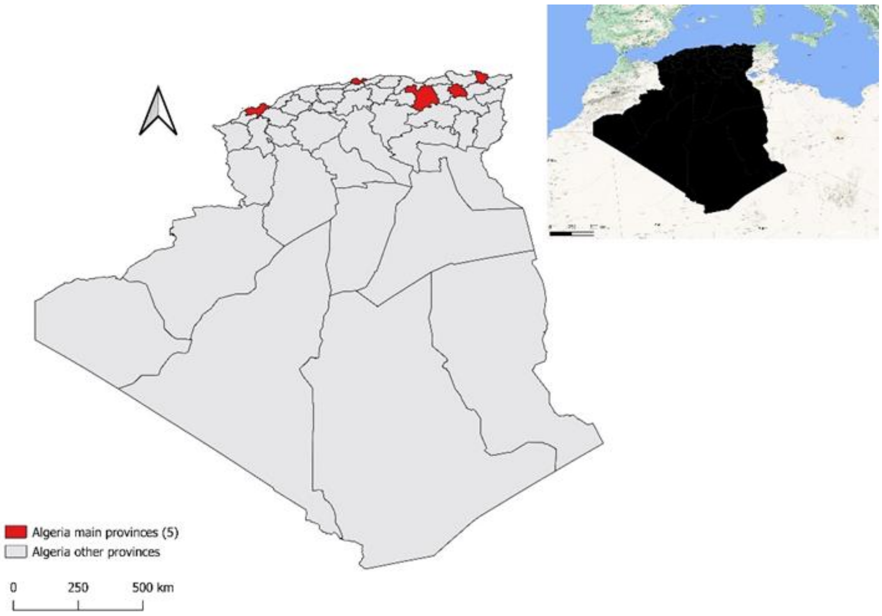
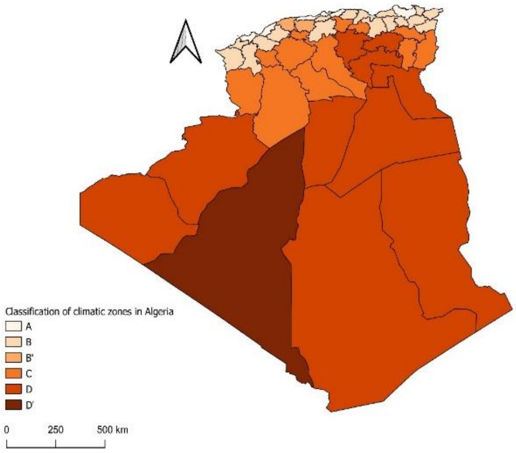


Fig. 3. Algeria geographic situation and presentation of its administrative division. Algeria is the northern country and the biggest one in the Maghreb. It is composed of 48 provinces.



Climate characteristics data of each zone

Climate zone	Geographical location	Winter (dry-bulb temperature)°C	Summer (dry-bulb temperature /wet bulb temperature/ Temperature of dew) °C		
A	The Northern Algeria with coastal band	5	35	25	25
B	The plain behind the sea shore	2	41	23	21
B'	Sub-zones of zone B	2	38	24	23
C	The Hauts Plateaux, between the Tellian and Saharan Atlas	-2	39	24	16
D	The North, east, southwest, and central Sahara	4.5	45	27	18
D'	Sub-zones of zone D	7.5	48	29	19

Fig. 4. Classification of official Algerian climate zones based on winter and summer climate temperatures
Data source: (CNERIB, 1997, APRUE database)

The distribution of provinces by climate zones, shown in Fig. 4 represents the third parameter that defines this database. Finally, the study is carried out using a large database, which is spread over a large period of twenty-three years, which corresponds to the construction of the largest housing programmes in the country.

3.1.2. Data source

The first type of data is collected from the National Agency for the Promotion and Rationalisation of the Use of Energy (APRUE), which includes statistical references of various economic sectors (residential, tertiary and industrial). The present study focuses on the residential sector. The database comprises five key parameters: energy consumption by energy source, electrification rate, gas connection rate, socio-economic indicators such as number of inhabitants and GDP. These parameters were gathered by APRUE from several official bodies. The second set of data was collected from the Algeria Ministry of Housing, Urbanism and City (MHUV). It contains all statistics about the RBS. Both sets of data represent the most recently available statistics on REC and RBS in Algeria.

3.1.3. Data transformation and data cleaning

As mentioned above, the first step in cleaning data is to extract data relevant to the residential sector from APRUE database. Secondly, several units of measurement that concern energy consumption are used in this database such as (kw/h, thermie, m^3). These are converted into a universal unit of measurement that is the tons oil equivalent (Toe) using energy conversion factors. Finally, in order to calculate the REC for the entire RBS, 7.6 % of missing data are interpolated and extrapolated.

3.2. Data processing

In order to develop an Algerian REC model, data processing involves, firstly, the calculation of REC and RBS in 48 provinces from 1995 to 2018, and secondly, the mapping of the spatial-temporal distribution of REC and RBS through GIS, and finally, the statistical analysis of the energy consumption per housing.

3.2.1. Calculation of REC and RBS

There are several interpretations of energy consumption in buildings: energy consumption in the construction industry, energy consumption in the construction sector, energy consumption during the life cycle of building, and energy consumption in buildings to meet their needs (Huo et al., 2018).

In this study, the emphasis is placed upon the quantity of energy consumption during its exploitation at the level of Algerian RBS. There is no data that matches the various interpretations of REC which are cited by Huo et al. (2018) at the level of statistical yearbooks and official databases. Moreover, data that are related to REC are combined with statistics from other sectors through the consumption of different types of energy. It is the reason why the present study proposes to fill in the gaps in terms of REC data at the level of all provinces. The aim is to create a larger database that can be used later as a reference by policy-makers and researchers, by using calculation methods as found in the work of Y. Wang et al. (2020), who quantifies REC-RBS in some Chinese cities.

Based on the APRUE database, the present paper calculates REC using the five energy types that were used in the Algerian RBS between 1995 and 2018: electricity, natural gas, liquefied petroleum gas (LPG), gasoil, and fuels. From the percentages of energy consumption by each energy type in the residential sector, three formulas are selected and applied for three periods:

From 1995 to 2005, the formula is as follows:

$$REC^{P}_{t} = 35\ \%\ E^{p}_{electricity,t} + 41\%\ E^{P}_{natural\ gaz,t} + 92\ \%\ E^{P}_{LPG,t} + 3.6\ \%\ E^{P}_{fuels,\ t} + 0.10\%\ E^{p}_{gasoil,t} \qquad (1)$$

From 2006 to 2008, the formula is as follows:

$$REC^{P}_{t} = 35\ \%\ E^{p}_{electricity,t} + 41\%\ E^{P}_{natural\ gaz,t} + 92\ \%\ E^{P}_{LPG} + 0.1\ \%\ E^{P}_{fuels,\ t} \qquad (2)$$

From 2009 to 2018, the formula is as follows:

$$REC^{P}_{t} = 35\ \%\ E^{p}_{electricity,t} + 55\ \%\ E^{P}_{natural\ gaz,t} + 92\ \%\ E^{P}_{LPG} + 0.1\ \%\ E^{P}_{fuels,\ t} \qquad (3)$$

Where REC^{p}_{t} indicates annual residential energy consumption, the $E^{P}_{electricity,\ t}$ represents annual electricity consumption, the $E^{P}_{natural\ gaz,\ t}$ corresponds to annual natural gas consumption, and the $E^{P}_{LPG,t}$, $E^{P}_{fuels,\ t}$, $E^{p}_{gasoil,t}$ refer to annual consumption of liquefied petroleum gas (LPG), domestic fuels and gasoil, respectively. These entire energy sources are provided at the level of the province P during the year t. The presentation of the formulas according to three different periods corresponds to the percentage of energy consumption as extracted from the APRUE database. The calculation of REC by province over the study period is obtained by the sum of all types of energy consumptions.

The RBS data is obtained from (MHUV) database. The base year 2008 corresponds to the last national census. The RBS for the years following the base year are then equal to the sum of the previous year's building and the newly constructed stock (Eq.(4)). The RBS for the years before the base year are obtained by subtracting the newly constructed stock from the following year (Eq. (5)) as follow:

$$RBS^{p}_{t} = RS^{P}_{t-1} + NBS^{p}_{t},\ (t > \text{base year}) \quad (4)$$

$$RBS^{p}_{t} = RS^{P}_{t+1} - NBS^{p}_{t+1},\ (t < \text{base year}) \quad (5)$$

Where RBS^{p}_{t} refers to residential building stock in province p during year t, and RBS^{p}_{t+1} represents residential building stock after the following base year, while RBS^{p}_{t-1} corresponds to residential building stock of previous year. Finally, the NBS^{p}_{t} indicates the newly constructed stock in province p during year t.

In this study, the demolished building stock is not taken into account in the estimation of RBS. The demolition rate is very low given that a large share of the Algerian housing stock is recent and detailed statistics relating to demolitions are missing in both databases and official reports.

3.2.2. Spatial-temporal maps of REC and RBS, and energy consumption per housing unit

The spatial-temporal distribution of REC and RBS at province scale is mapped with quantum GIS (QGIS, 2021). The results are visualized along specific dates to visualize the evolution between provinces through a distribution of six classes, whose limits were defined according to the presence of significant differences between them, through natural breaks of Jenks.

In order to provide more information on REC, the energy intensity, which is an energy efficiency indicator, that measures residential energy efficiency, is used to analyse the variation of energy consumption between different climate zones. In the absence of data on built surface areas of RBS in official statistics, this article proposes to calculate energy consumption per housing unit according to the following formula:

$$\text{Energy consumption per housing unit} = REC^{P}_{t} / RBS^{P}_{t} \quad (6)$$

Where REC^p_t and RBS^p_t represents residential energy consumption and residential building stock in province p during year t.

3.3. Development of REC model

The third section corresponds to the statistical REC model. It is based on a mixed approach, which combines regression analysis and hierarchical clustering. The regression allows understanding the way that the REC evolves as a function of independent variables, whereas the clustering aims to cluster similar provinces according to their energy profiles.

3.3.1. Regression analysis

In order to study the relationship between the REC and its independent variables, an analysis is performed using Multiple Linear Regression on R software (R, 2021) to develop an energy consumption model for Algerian RBS. The REC model is used as a tool for the analysis and identification of the influence of independent variables on REC. This model is developed through several steps:

The first step is the selection of variables to be used in regression analysis. These variables are based on crosscheck between literature review and Algerian context, according to the availability of data from major national statistical databases during the period 1995 to 2018, for all provinces and all climate zones.

Table 2 shows the description of independents variables integrated into REC model. These variables cover the four main dimensions identified in our literature review (socio-demographic, economic, climate, building and urban dimensions). Three of these variables are described in the above section, such as the REC, which is our response variable, the RBS and the climatic zones. The population variable is directly extracted from APRUE database. The other variables, household size, urban density, Gas connection rate, Electrification rate and GDP were calculated according to the following formulas:

$$\textit{Household size}^{p}_{t} = P^{p}_{t}/RBS^{p}_{t} \quad (7)$$

$$\textit{Urban density}^{p}_{t} = P^{p}/\textit{Province area}^{p} \quad (8)$$

$$GDP^{p}_{t} = GDP/capita^{N}_{t} \times P^{p}_{t} \quad (9)$$

$$\textit{Gas Connection Rate}^{p}_{t}\ \% = \textit{Nb. of subscribers (Gas)}\ {}^{p}_{t}/RBS^{P}_{t} \quad (10)$$

$$\textit{Electricity Connection Rate}^{p}_{t}\ \% = \textit{Nb. of subscribers (Electricity)}^{p}_{t}/RBS^{P}_{t} \quad (11)$$

Table 2
Description of independents (exploratory) variables integrated in regression analysis.

Designating variables*	**Type of variable**	References
RBS**	• Discrete quantitative variable	(Hsu, 2015; Kavousian et al., 2013; Kim et al., 2020; Theodoridou et al., 2011)
Population ******	• Discrete quantitative variable	(Al-Ghandoor et al., 2009; Bianco et al., 2013; Chen et al., 2016; Theodoridou et al., 2011)
Households size**	• Continuous quantitative variable	(Iraganaboina & Eluru, 2021; Kim et al., 2020; Tso & Guan, 2014)
Urban Density **	• Continuous quantitative variable	(Chen et al., 2016; Dujardin et al., 2014; Iraganaboina & Eluru, 2021)
Electricity and gas Connection rate **	• Discrete quantitative variable	(Kim et al., 2020; Sepehr et al., 2018; Štreimikienė, 2014; Tso & Guan, 2014)
Climatic zone ***	• Nominal qualitative variables	(Al-Ghandoor et al., 2009; Biswas et al., 2016; Catalina et al., 2013; Fumo & Rafe Biswas, 2015; Kim et al., 2020; Williams & Gomez, 2016, Walter & Sohn, 2016)
GDP ******	• Continuous quantitative variable	Bianco et al. (2013)

*All variables are analysed across 48-province scale from 1995 to 2018
** Its variables have been calculated from the previous sections.
*** This variable was extracted from (CNERIB, 1997)
**** Its variable was extracted from APRUE database

Household size$^{p}_{t}$ represents the average number of persons in a housing. The RBS$^{p}_{t}$ corresponds to the residential building stock. It has been calculated above using formulas (4) and (5). Urban density$^{p}_{t}$ refers to number of inhabitants per province level area. The GDP$^{p}_{t}$ is the estimated Gross Domestic Product. The Gas Connection Rate$^{p}_{t}$ % corresponds to the rate of dwellings connected to natural gas network, according to the number of subscribers of gas companies. The Electricity Connection Rate$^{p}_{t}$ % corresponds to the rate of dwellings connected to electricity network, according to the number of subscribers of electricity companies. All these variables are provided at the level of the province P for the year t.

The second step consists in the pre-processing of data. The population variable has been removed from the matrix in order to avoid multicolinearity, due to its presence in the calculation of three variables. Secondly, levels of qualitative variable (Climatic zone) are reordered, so that climatic zone A is used as the reference level. Finally, logarithmic transformation of all variables is done to satisfy the normality and homoscedasticity assumptions of linear regression.

The third step focuses on the execution of Multiple Linear Regression using R software, in order to detect the most significant variables in REC model, in the form of the following equation:

$$\mathrm{Log}(REC) = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m + \beta_{m+1}\log(X_{m+1}) + \ldots + \beta_n \log(X_n) + \varepsilon \quad (12)$$

Where REC refers to the response variable, which represents residential energy consumption. $X_1, \ldots, X_m$ are the qualitative variables while $X_{m+1}, \ldots, X_n$ are the quantitative ones cited above, with n as the number of variables. $\beta_0, \beta_1, \ldots \beta_n$ represents the regressions coefficients, and ε is the error of the model, which expresses the remaining variation in Y that cannot be explained by the independent variables $X_1, X_2, \ldots\ldots X_n$.

The fourth step consists in validating underlying assumptions of multiple linear regression. We first check that there is a linear relationship between the dependent variable and the independent variables, through scatterplots and R^2, of the following formula:

$$\mathbf{Log(\mathit{REC}) = \beta_0 + \beta_n log(\mathit{X_n}) + \mathcal{E}} \quad \textbf{(13)}$$

Where REC refers to the response variable, β_n represents the regressions coefficients of each independent variable, and Ɛ is the error of the simple model.

We further perform a multivariate normality test, using the residual quantile-quantile plot, and a homoscedasticity test by a graphic diagnosis of standardized residuals versus predicted values. Both assumptions are tested from the regression equation (12). We finally check for multicolinearity between independent variables through the Variance Inflation Factor (VIF).

3.3.2. Clustering of provinces according to their REC profile

In order to identify province's REC behaviour, a heat map is developed by post-processing the REC data from the database. The objective is to extract data from the study span 1995-2018 along three periods. These periods are chosen based on Algeria's macro-social and political developments. Furthermore, formula 14 is used to calculate the evolution of REC on an annual basis for each period.

$$REC^p_t \,\% = (REC^P_t / REC^p_{t-n}) exp^{1/n-1} \quad \textbf{(14)}$$

Where REC^p_t % represents the evolution of REC for Province P over the period t. REC^P_t refers to the ending residential energy consumption per province P. REC^p_{t-n} corresponds to the beginning residential energy consumption per province P, n is the number of years, which represents a period of study.

A hierarchical cluster analysis is then performed to group provinces with similar REC profiles. This analysis is executed using R software according to the following steps:

Step 1: The data is first processed into a data matrix with rows representing 48 provinces and columns being the changes in REC for three studied periods (1995-2002, 2002-2009, 2009-2016).

Step 2: Hierarchical ascending classification is then carried out using Euclidean distances between observations. The cluster agglomeration method employed in this study is Ward's minimum variation method that minimizes the total within-cluster variance. The resulted dendrogram and inertial gain plots (Fig. 14) is used to determine the optimal number of clusters (Husson et al., 2017).

Step 3: The final clusters are compared and interpreted using all variables as well as the variable that contributed the most in class separation.

Step 4: A representative province is selected from each cluster to serve as a pilot province for the government to test the feasibility of energy transition scenarios. The selected province of each cluster is the closest one to the gravity center of the cluster.

3.4. **Validation of REC model**

The fourth and final section in this study corresponds to the validation of REC model. This validation is carried out through comparison of research results with values found in official reports and previous researches.

4. Results and discussion

4.1. REC and RBS of the Algerian territory

The total REC significantly increased from 3.57 million tons oil equivalent (Mtoe) in 1995 to 9.55 (Mtoe) in 2018, with an average annual growth rate of 2.7% (see Fig 5). The annual growth rate of the REC varies between -9.68 % and 22.82 %.

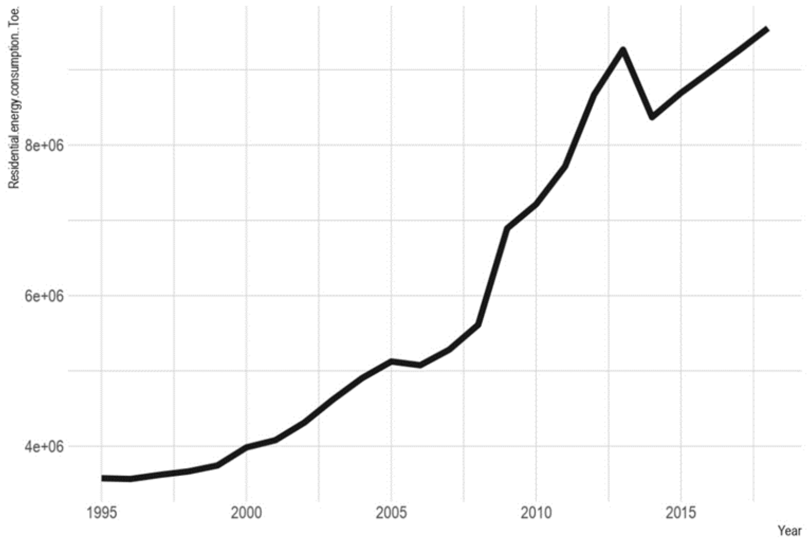


Fig. 5. Residential energy consumption (REC) of Algeria.

The detailed values of the estimated REC and RBS are provided in Appendix A and B.

In the same time, the total RBS increased from 4.36 million housing units in 1995 to 9.6 million housing units in 2018 (see Fig 6). The evolution rate varies from 1.86 % to 10.97 % over the same period, with an average annual growth of 3.5%.

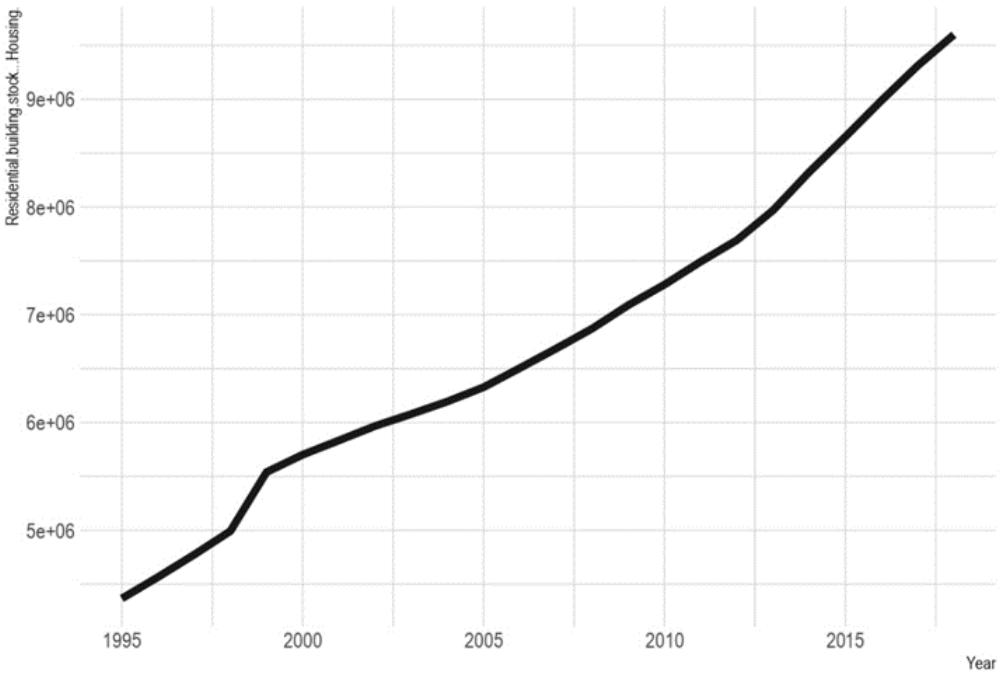


Fig. 6. Residential building stock (RBS) of Algeria.

Obviously the growth of REC is positively related to the growth of RBS. It is a result of the high demand for housing which the Ministry of Housing is trying to address in order to meet a demand driven by population growth, which has increased from 28 million in 1995 to 40 million in 2018, with an average annual growth of 2 % over the last ten years (ONS, 2020). Still the average annual growth rate of RBS is higher than the one of REC, which means that the residential energy consumed by housing is progressively decreasing in Algeria.

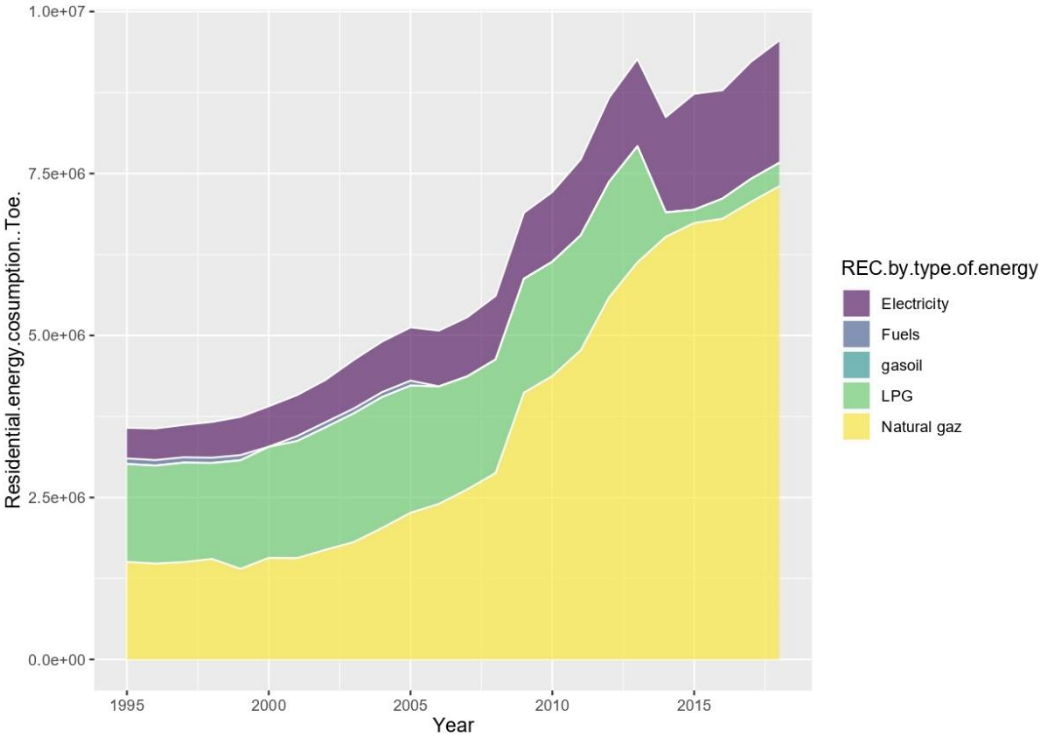


Fig. 7. Algeria's 1995-2018 REC structure.

The REC structure can be grouped into three categories based on the evolution of energy sources consumed by RBS from 1995 to 2018. Fig. 7 shows that the gasoil and fuels represent the first category of REC, which is characterised by the two least used energy sources by RBS, with a complete disappearance of gasoil from the 2000s onwards and a decrease in the use of fuels, which is tending to disappear from 2014. The latter are generally used in the rural, remote areas in the south of country, which covers climatic zones D and D', and that because of the low percentage of electricity and gas connections compared to the rest of the country, induced by the low rate of urban density in these areas, where gasoil was used for lighting, and domestic fuels were used for heating. The petroleum products tend to disappear from REC since 2014, in response to the evolution of gas and electricity connection rate, in the southern part of the country from 4.5% to 22.5% and 82% to 100% percentage respectively.

The LPG represents the second category of REC, which is characterised by a strong growth between 1995 and 2008, to meet heating and cooking needs in some rural areas, with low gas connection rates. The latter started to decrease from 2008 onwards, in response to national connection rate evolution, which evolved from 36 % to 56 over the last 10 years (APRUE, 2015c; APRUE, 2017). Natural gas and electricity both represent the third category of REC, characterized by the most common source that continues, up until today, to be consumed. Natural gas corresponds to the most consumed energy in RBS.

The decrease in REC recorded in 2014 (see Fig. 5), can be explained by the decrease in consumption of LPG and fuels, which decreased by 78.9% and 88.5% between 2013 and 2014 respectively. This decrease is linked on the one hand to the evolution of gas connection rate in provinces located mainly in the south of Algeria, covering climatic zones D and D', passing from 15.8% to 22.2% between 2010-2013 (APRUE , 2015c). The evolution in the rate of connection to natural gas is supported by the increase in the number of low voltage subscribers (households) of 7.3% between 2013 and 2014 (Ministère de l'Energie, 2014). Gas connection rate is slowly increasing nationwide, and still does not cover the whole RBS, as in 2017, it recorded 56% of gas connection rate, compared to 99% of electrification (APRUE, 2017). On the other hand, electricity consumption has seen a sharp increase of 32.8% between 2013-2015 to meet the energy demand during this period of decline, (See Fig. 7).

4.2. Spatial-temporal distribution of RBS and REC per province, and dispersion of energy consumption per housing unit

The spatial-temporal distribution of RBS is presented in Fig. 8. The northern part of the country, which represents the climatic zone (A) contains the highest number of RBS. It is followed by the highland that represent the climatic zones (B), (B'), and (C).

Finally, the climatic zones (D) and (D’), which refers to southern Algeria has the lowest number of RBS. At the provincial level, the total RBS varies from 2000 to 830 000 housing units. Furthermore, the number of RBS is the highest in the province of Algiers, followed by Oran, Setif, Tizi Ouzou and Batna. In contrast, the provinces of Tindouf, Illizi, Naama, Tamanghasset and El Bayad present the lowest number of RBS on a national scale. Fig. 9 reinforces the observations resulting from the spatial-temporal distribution of RBS. It indicates that climatic zone (A) gathers the provinces with the highest RBS values, whereas climatic zone (D') refers to the provinces with the lowest RBS values. It is worth noting that the highest RBS value is located at the level of climate zone (A), whereas the lowest RBS value is located at the level of climate zone (D).

The distribution of RBS through provinces is directly related to the housing demand induced by the variability of the number of inhabitants per province. The Algerian territory is characterised by a very high population concentration index, with 90% of the population concentrated in the north of the country, while only a small share of the inhabitants are located in the south of country. The decrease in population density from north to south is related to the presence of agricultural land, the abundance of water resources and the concentration of transport and communication infrastructure along the northern strip of the country (Kateb, 2003).

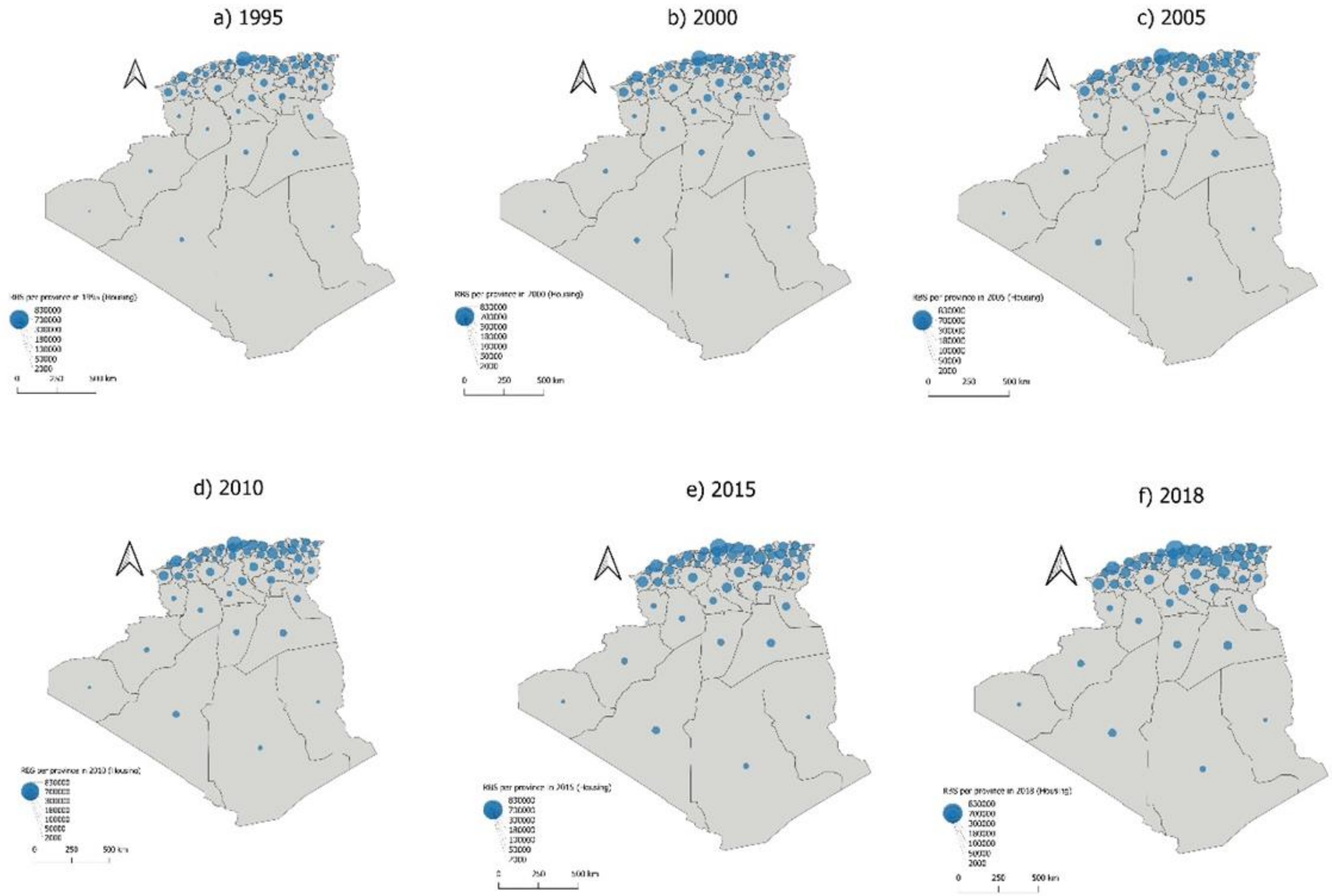


Fig. 8. Spatial-temporal distribution of RBS by 48 provinces in selected years.

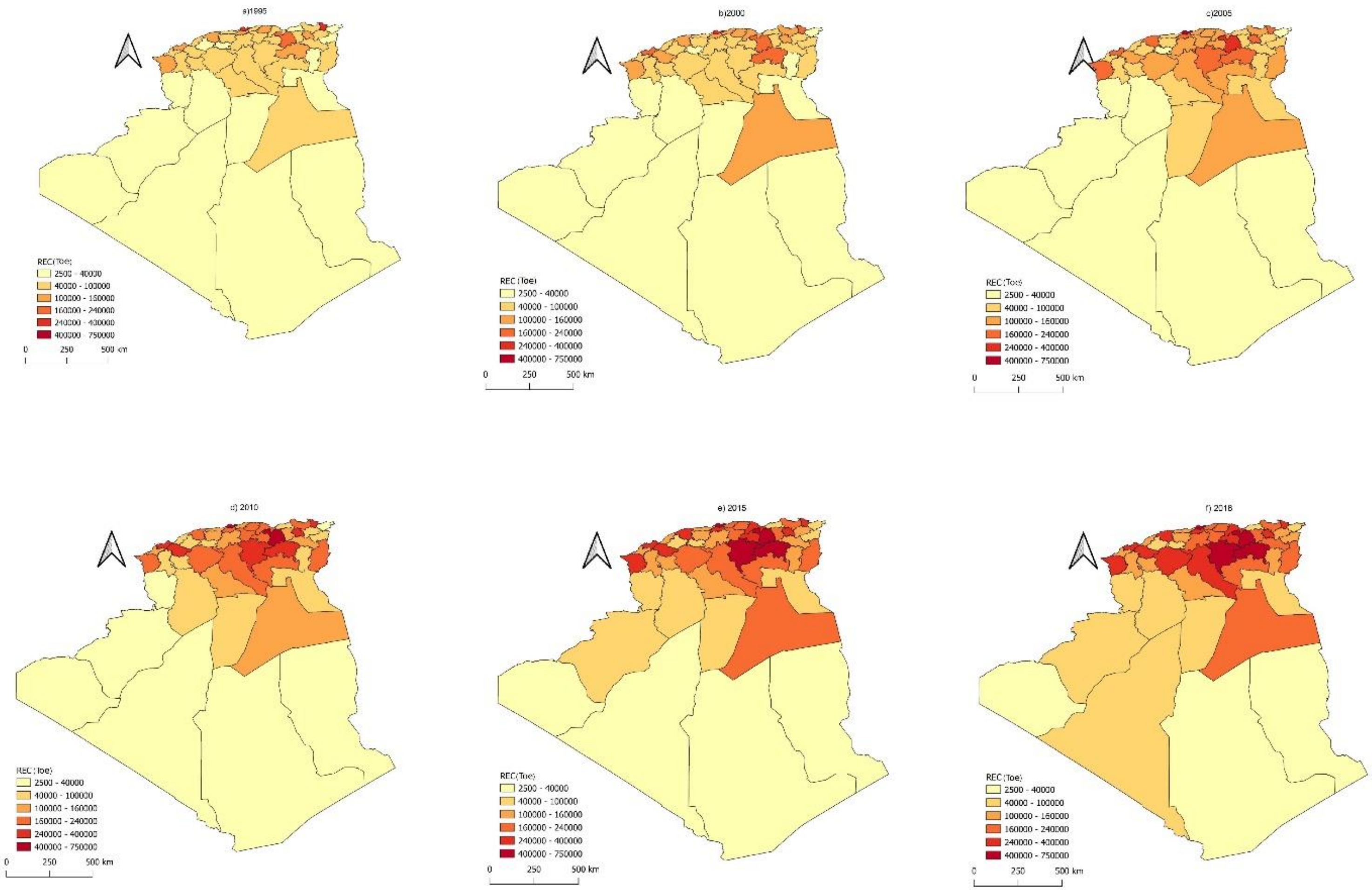


Fig. 10. Spatio-temporal distribution of REC by 48 provinces in selected years.

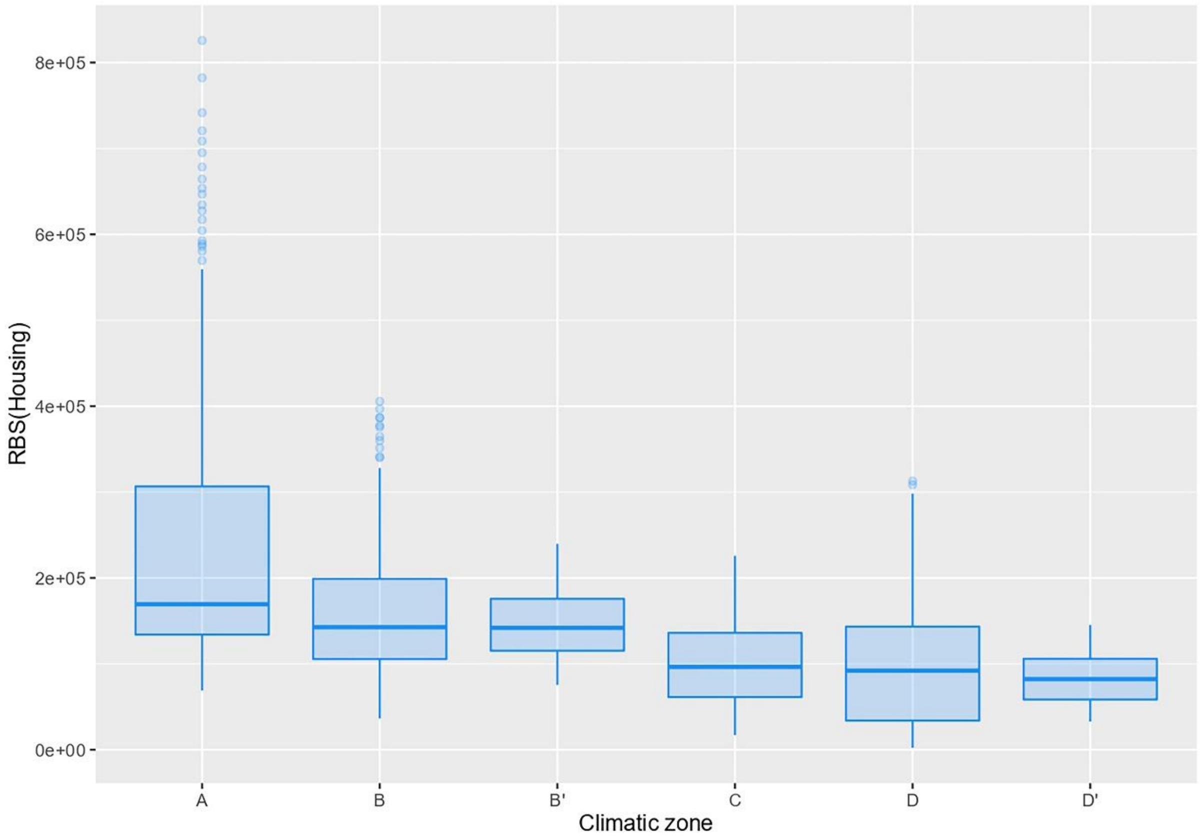


Fig. 9. Dispersion of RBS per climate zones.

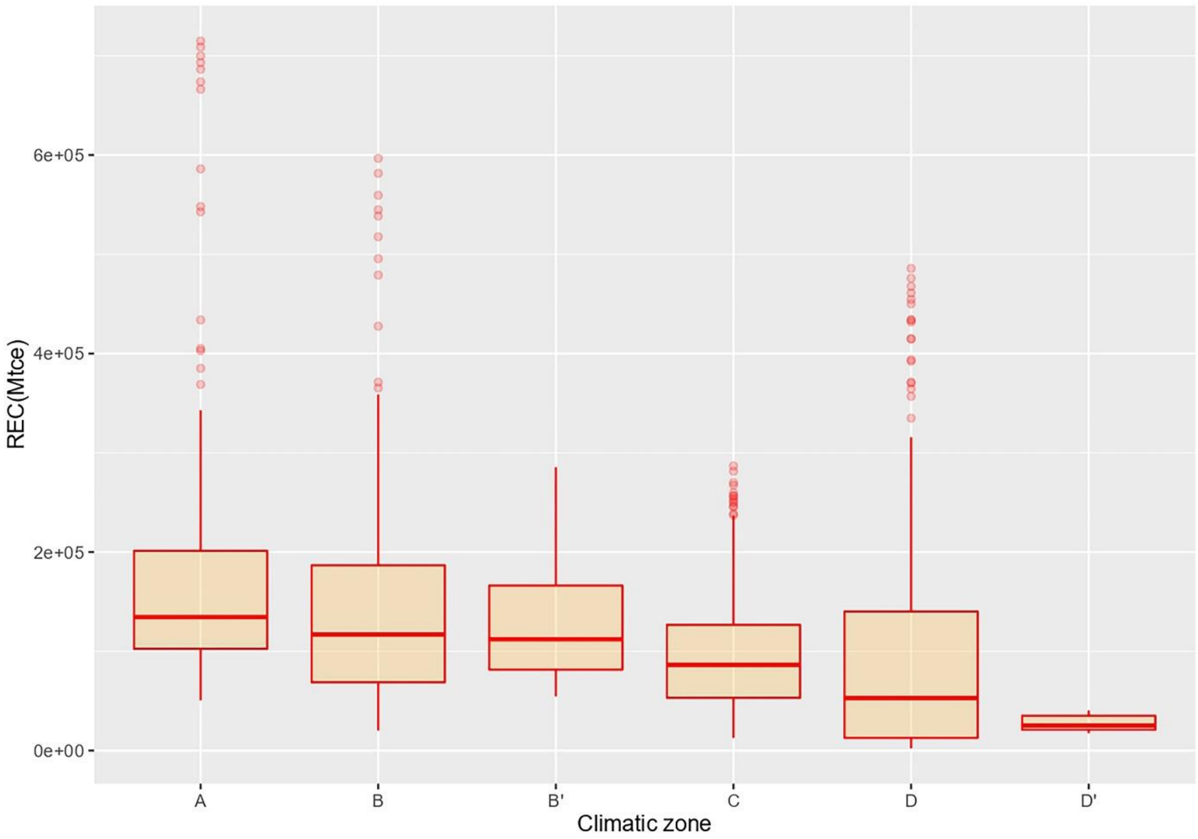


Fig. 11. Dispersion of REC per climate zones.

The spatial-temporal distribution of estimated REC in Algeria's 48 provinces indicates that REC is higher in the northern provinces such as Algiers, followed by Setif, Batna, M'sila, and Constantine (see Fig. 10). These are located in the climatic zones (A), (B) and (C). In contrast, the southern provinces have the lowest REC. These provinces are located on the climatic zones (D) and (D'). The total REC varies from 2500 to 750000 (Toe). The spatial distribution of REC largely follows the same pattern as the one of RBS.

The highest REC is observed in the Northern provinces with a high residential density. The south of Algeria is characterized by a lower REC. Fig. 11 shows that the climate zone (A) includes the provinces with the highest REC values, while the climatic zone (D') includes the provinces with the lowest REC values. The highest REC value is located in the climate zone (A), whereas the lowest REC value is located in the climate zone (D).

In order to guide policy makers and energy planners in their decision-making, the energy consumption per housing unit was obtained by dividing REC by RBS according to the formula (6). The box plot in fig 12 materializes the results.

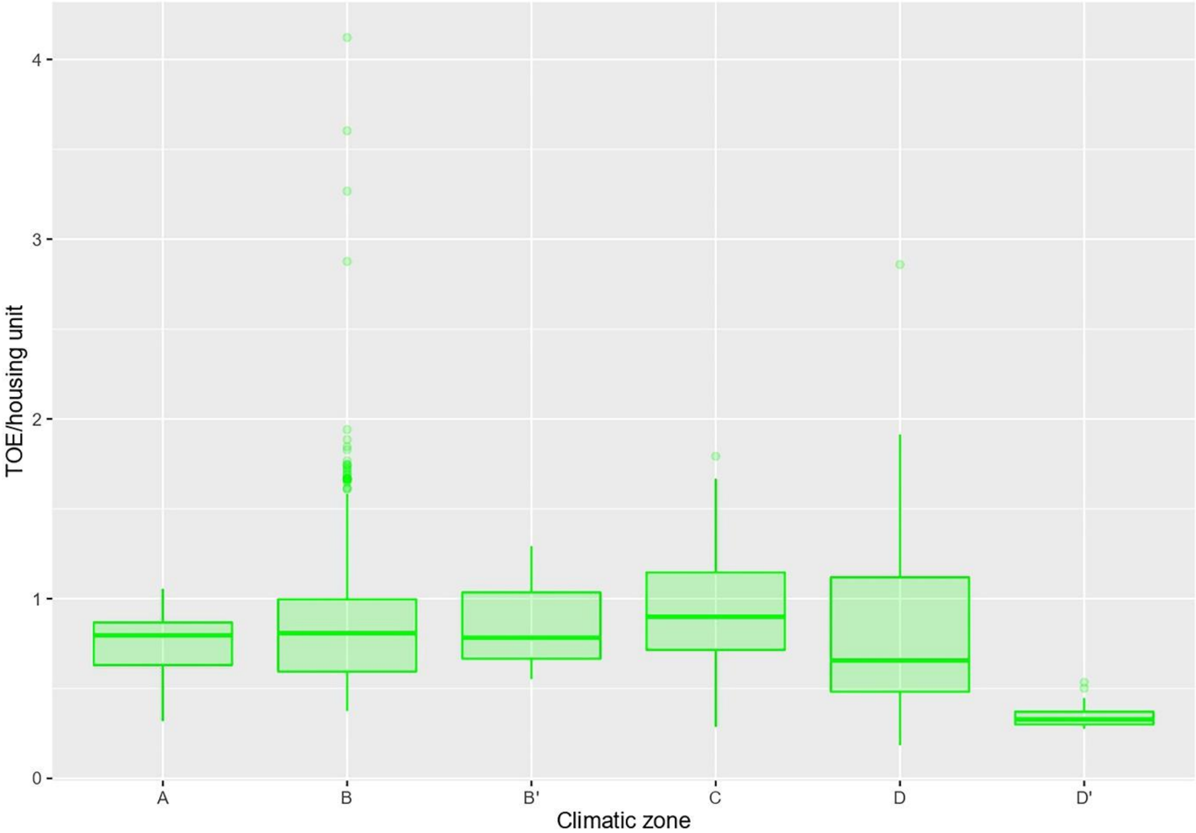


Fig. 12. Box plot representing the dispersion of energy consumption per housing unit at climate zones level.

Fig. 12 shows the dispersion of energy consumption per housing unit across the six climate zones. Climate zone C and D are the one where maximum energy consumption per housing unit is the highest, as it is characterised by hot summers and cold winters (see Fig. 4). For climate zones B and B', energy consumption per dwelling unit is very similar, from a homogeneity and median point of view. Zone A has a very low energy consumption compared to other climate zones, due to the Mediterranean climate. On the other hand, climate zone D' is completely out of line with all zones, despite having very similar climate characteristics that zone D. It seems that other parameters affect energy consumption per housing unit.

4.3. Regression analysis

4.3.1. Multiple linear regression model for REC

For the Multiple linear regression model of the form of Eq. (12), the estimated REC model executed on 10,368 values[1], is found to be:

$$\log(\text{REC}) = -1.73179 + 0.06070\,\text{Climatic Zone}_{B} + 0.13385\,\text{Climatic Zone}_{B'} + 0.15441\,\text{Climatic Zone}_{C} + 0.17433\,\text{Climatic Zone}_{D} - 0.01789\,\text{Climatic Zone}_{D'} + 0.36455\,\log(\text{Electrification rate}) + 0.24866\,\log(\text{Gas connection rate}) + 0.10763\,\log(\text{GDP}) + 0.10433\,\log(\text{Household size}) + 0.85921\,\log(\text{RBS})\ 0.06045\,\log(\text{Urban density}) + \varepsilon \qquad (15)$$

Table 3 summarizes the results of REC model with the estimated coefficients, standard error, and the p-values. Based on this table, it can be noted that all climatic zones, except climatic zone D', the RBS, the gas connection, the electrification rate, and finally the urban density are significant in explaining the REC, with p-values less than 0.05. Household size in Algeria does not have a great significance in REC model.

Table 3
Regression summary outputs for the Algerian REC model.

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	-1.73179	0.18817	-9.203	< 2e-16 ***
Climatic Zone B	0.06070	0.01314	4.621	4.29e-06 ***
Climatic Zone B'	0.13385	0.02298	5.824	7.67e-09 ***
Climatic Zone C	0.15441	0.01749	8.827	< 2e-16 ***
Climatic Zone D	0.17433	0.02281	7.642	4.85e-14 ***
Climatic Zone D'	-0.01789	0.04229	-0.423	0.6724
Z.Electrification rate (%)	0.36455	0.07109	5.128	3.49e-07 ***
Z.Gas connection (%)	0.24866	0.01678	14.815	< 2e-16 ***
Z.GDP ($)	0.10763	0.02210	4.870	1.29e-06 ***
Z.Household (size persons)	0.10433	0.04315	2.418	0.0158 *
Z.RBS (Housing)	0.85921	0.03788	22.683	< 2e-16 ***
Z.Urban density (Inha/Km2)	0.06045	0.01244	4.859	1.36e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1282 on 1042 degrees of freedom
Multiple R-squared: 0.868, Adjusted R-squared: 0.8667
F-statistic: 623.2 on 11 and 1042 DF, p-value: < 2.2e-16
Note: Variable “Climatic Zone A” assumed as reference.

[1] The matrix is provided in appendix C

The coefficients of determination R^2 and adjusted R^2 of REC model are 0.86. The p-value, which is associated with REC linear regression, is 2.2e-16. This means that the REC model is statistically significant.

On the level quantitative variables, 1% increase in RBS, Electricity connection rate, Gas connection rate, GDP, and urban density respectively lead to an increase of 0.85%, 0.36%, 0.24%, 0.10%, and 0.06% in REC. The RBS is obviously the quantitative variable that most affects the REC. Still the model highlights that, when considering all factors altogether, electricity and gas connection rates have a much stronger influence than GDP on REC. Besides, this urban density is positively correlated with REC. This goes against part of the literature in this field where density is usually associated with lower REC. That may be related to an evolution of thermal comfort expectations of the inhabitants. However, the household size in Algeria does not have a great significance in REC model, probably due to the socio-political context.

With regard to the qualitative variables, and using the climatic zone A as a reference variable, climatic zone D, C, B and B' in this order, have a determining impact on REC. This is in line with observations made in the previous section.

In order to validate REC model, the assumptions mentioned in the methodology section have been tested and validated. The outputs of the different tests illustrated in different tables and graphs can be consulted in Appendix D. Therefore, REC model assumptions are carefully verified. The model is statistically significant and the null hypothesis can be almost certainly rejected.

4.3.2. REC clusters

To provide additional visualisation, Fig. 13 shows the heat map with the vertical axis representing each province and the horizontal axis representing the study period. The intensity of colour in the boxes indicates the evolution of REC over the relevant period. The red colour represents the provinces with a strong REC evolution, i.e. between 5 and 10% of evolution on an annual basis. The green colour represents provinces with a declining REC evolution, i.e. between 0 and -5%. Provinces with an average REC evolution of, i.e. between 0 and 5 %, are represented by a colour gradation between red and green.

The majority of provinces had a strong evolution of REC during the 2002-2009 period. This period corresponds to the launch of the five-year housing programmes. On the other hand, a downward trend in the REC is observed during the period 2009-2016 except for the province of Setif, which is undergoing rapid urbanisation because it represents an economic hub in northeastern Algeria.

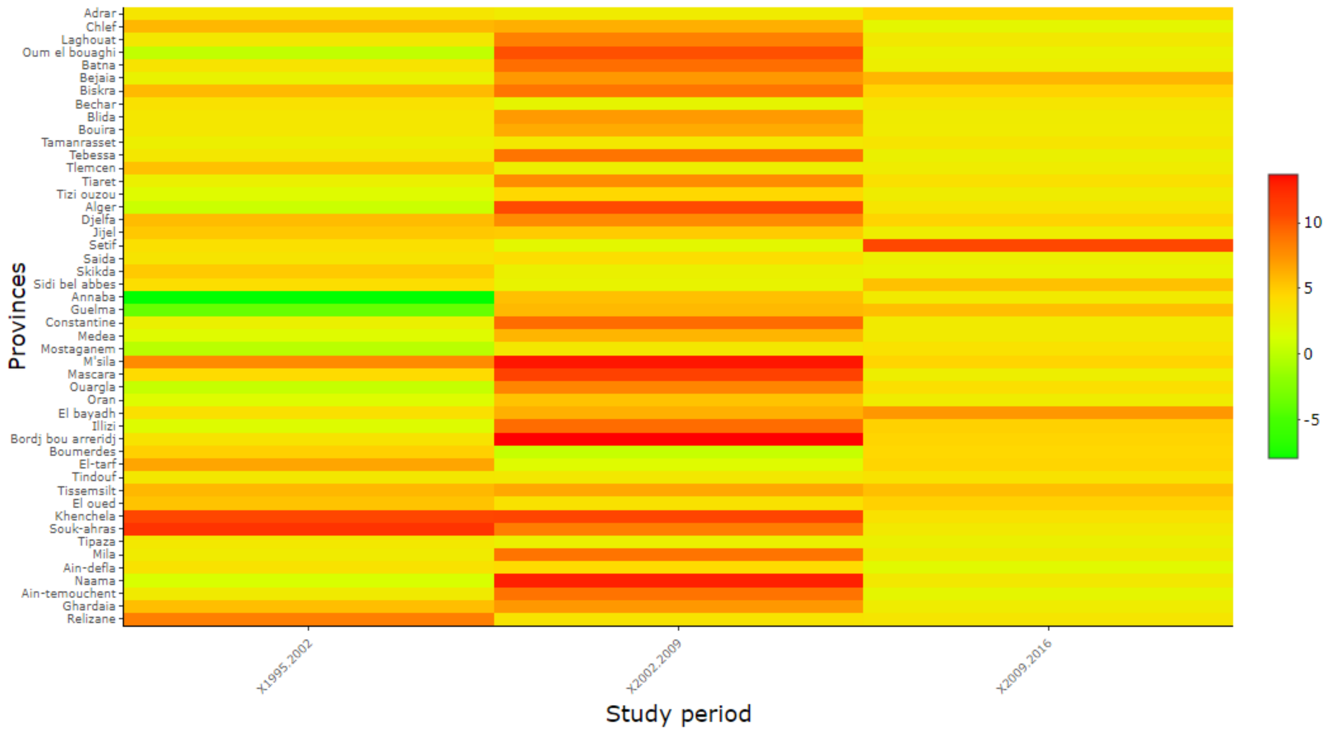


Fig. 13. Heat map showing REC evolution by 48 provinces per period.

Fig. 14 shows the hierarchical tree used to group together provinces with similarities in REC profile. Four clusters are identified based on the ward method and adopted the

Euclidean distance as a vector distance. Based on REC pattern across all periods (see Table 4), these four clusters are created as follows:

- The **cluster 1** gathers provinces such as Annaba and Guelma. This group is characterized by a strong decrease in 1995-2002 and stable or minor increases from 2002 to 2016.
- The **cluster 2** gathers provinces such as Bechar, Skikda, Oran, etc. This group is characterized by quite stable RECs with minor increases or decreases in some periods.
- The **cluster 3** gathers provinces such as Algiers, Ouargla, Mascara, etc. This group is characterized by a rather stable trend in 1995-2002 with minor increases or decreases in some provinces. Strong increases can be observed in 2002-2009. The trends come back to stable from 2009-2016.
- The **cluster 4** gathers provinces such as M'sila, Souk-Ahras and Khenchela. This group is characterized by average to strong increases in 1995-2009, which later become stable in 2009-2016.

Based on the observed REC profiles, and especially over the (2009-2016) period, energy policies for cluster 1 and cluster 3, with a decreasing REC, should consider an optimisation of thermal comfort in the housing stock. Cluster 2 with a rising REC caused by the strong growth of the housing stock rather requires energy efficiency strategies with a specific focus on the construction of high-performance buildings and the use of energy retrofit. Energy policies for cluster 4, with a stable REC, should then favour measures to reduce energy consumption through energy renovation.

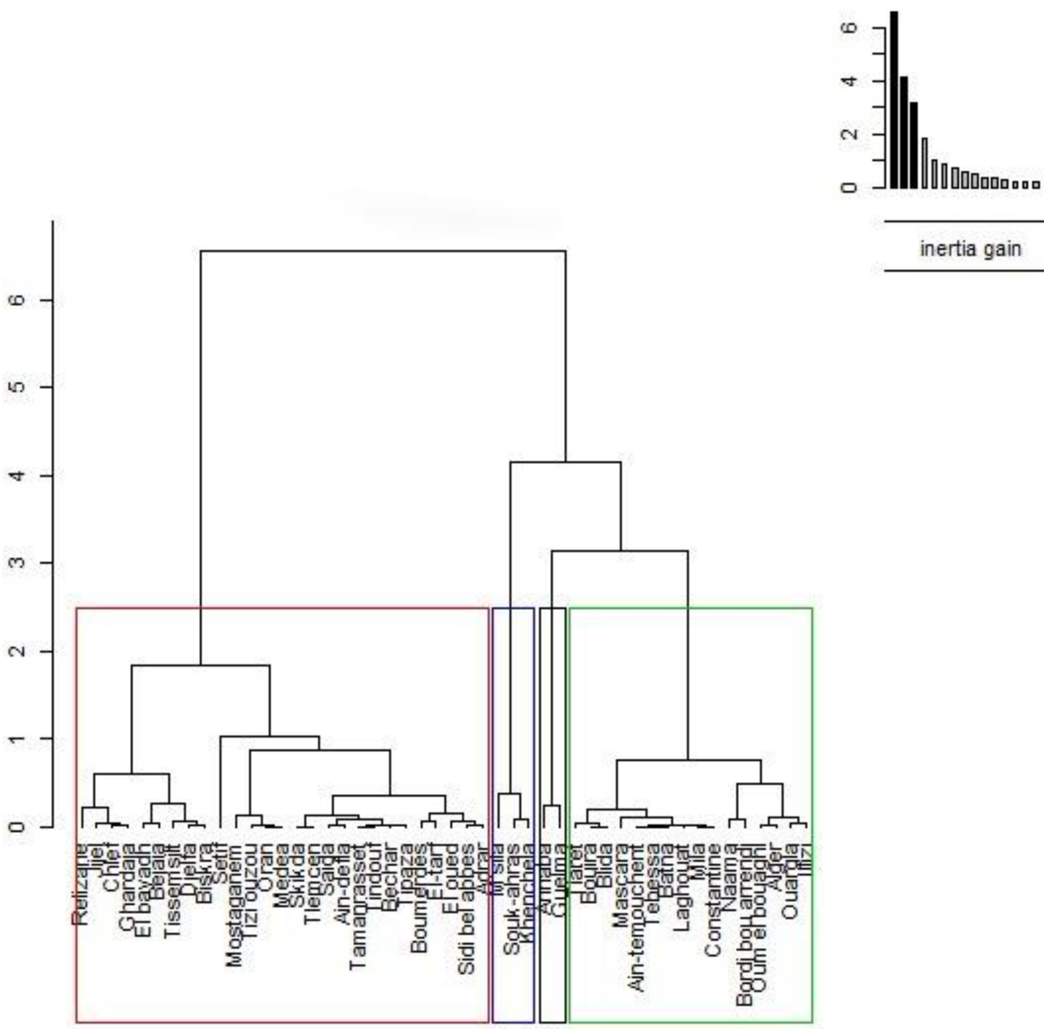


Fig. 14. Hierarchical clustering tree applied on REC evolution data of 48 provinces. The colored rectangles represent the cluster according to the ward’s method.

Table 5 shows pilot provinces that are extracted from each cluster based on their proximity to the barycentre. They represent characteristic provinces that can be used by energy planners, to test the socio-economic and environmental effectiveness of adaptation measures before being replicated at a larger scale to other provinces of the same cluster.

Table 4
The average annual growth rate of each cluster in each period.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
REC (1995-2002) %	-4.32	1.5	-1.66	3.73
REC (2002-2009) %	-0.35	-5.18	4.37	2.93
REC (2009-2016) %	-0.518	1.42	-1.75	0.078

Table 5
Extraction of pilot provinces of each cluster based on the distance from the barycentre.

Clusters	Pilot provinces (Parangons)	Distance from the barycentre
Cluster 1	Annaba	2.407224
Cluster 2	Tindouf	1.408166
Cluster 3	Consantine	0.2547846
Cluster 4	Khenchela	0.5745982

5. Validation of the developed REC Model

Tables, 6 and 7 indicates a comparison of national REC in selected years as calculated in this study, with the statistics that are gathered on the APRUE report and the values that are found in the annual REC estimations developed by Ghedamsi et al. (2016). Tables, 8 and 9 show a comparison between the national RBS results of the present study and the official reports of the Ministry of Housing as well as the results of the RBS estimations that are presented by Ghedamsi et al. (2016).

Table 6
Comparison of national REC results with those of energy reports.

Year	This study (total annual national REC (Toe)	APRUE total annual national REC (Toe)	Relative difference
2008 (latest census)	5615309.32	5039724.51	**10.25 %**
2013	9269638.95	8264981.78	**10.84 %**

Table 7
Comparison of national REC results with those of previous study.

Year	This study (total annual national REC (Toe)	R. Ghedamsi et al. (2016) total annual national REC (Toe)	Relative difference
2008	5615309.32	6210662.08	**-9.58 %**

Our results are larger (+10%) than the ones provided by the APRUE report and smaller (-9%) than the ones of Ghedamsi et al (2016). The REC calculated in the present study is situated between the two references, which confirms that it does not exceed the admitted limit set between ± 5% and ± 15%.

In addition, Tables 8 and 9 shows that RBS calculated in this study is slightly larger (+7% to +9 %) than the one provided in ministerial reports. It is very comparable to the one (+2 % to +7 %) provided by Ghedamsi et al. (2016).

Table 8
Comparison of national RBS results with those of ministerial reports.

	This study (total national RBS (housing)	MHUV total annual national RBS (housing)	Relative difference
2000	5704402	5279257	**7.45 %**
2008(latest census)	6874549	6276517	**8.70 %**
2013	7971998	7200000	**9.68 %**

Table 9
Comparison of national RBS results with those of previous study.

	This study (total national RBS (housing)	Ghedamsi et al. (2016) total annual national RBS (housing)	Relative difference
2008	6874549	6720000	**2.25 %**
2010	7216538	7040000	**2.45 %**
2015	8696308	8080000	**7.09 %**

6. Conclusion

6.1. Main findings and recommendations

This paper presents an Algerian REC model as a reference that would inform policy makers in the view of designing energy policies. The modelling task was developed through a triangular approach combining GIS, Multiple Linear Regression and hierarchical clustering.

The study shows that the REC and RBS have increased significantly with an average annual growth of 2.7 % and 3.5 % respectively, caused mainly by the combined pressure of demography and socio-economic development. Natural gas and electricity represent the most common fuels used in the Algerian RBS, where natural gas remains the most consumed source of energy.

The energy consumption per housing unit is higher in climate zones with hot summers and cold winters, especially when compared to Mediterranean climate of zone.

The Multiple Linear Regression highlights that the size of the RBS is the primary factor that influences REC. Besides this factor, electricity and gas connection rates are the most important ones, much more important than GDP or urban density. Climate zone influences the REC, especially in provinces located in zones C, D, B and B'. Even though modest, the influence of urban density on REC is positive, which means that provinces with a higher urbanization rate are associated with higher energy consumptions. That may be related to an evolution of thermal comfort expectations of the inhabitants. These are important factors to consider and monitor as urbanisation rate, and associated levels of gas and electricity connections, are expected to rise in the near future.

The analysis of hierarchical clustering allowed classifying all provinces into four main clusters, associated with decreasing (1 and 3), stable (4) or rising (2) REC over time. These patterns are important to consider when designing adequate energy policies for each of these situations.

Based on the results of the triangular approach, we propose these recommendations:

- Promoting the energy transition of the RBS through the introduction of the energy mix in the energy consumption, through the large-scale exploitation of the solar energy, thanks to the sunshine potential of Algeria.
- Developing the energy mix taking into account the evolution rate of natural gas and electricity, the impact of climatic zones on residential energy consumption per housing unit, and the energy consumption profile of each province.

- Designing energy policies according to the energy consumption profile of each province, associated with the actual rates of natural gas and electricity consumption.

Other recommendations can reinforce the orientations of the triangular approach:

- Combining results of triangular approach with technical policy-oriented studies to examine the feasibility and impact of energy efficiency scenarios at the scale of residential buildings, to mitigate the impact of outdoor temperatures, especially in climate zone C, D, B and B'.
- Extend the reflection by analysing REC models based on pilot provinces by including other variables that may affect REC, such as building typology, age, building characteristics, HVAC system, energy cost.

6.2. Strength and limitations of the study

The strength of this study relates to the use of a triangular approach, combining GIS, MLR and clustering, on a longitudinal basis. This original study is the first Algerian research that calculates, spatializes and analyses the REC and RBS across the entire Algerian territory, covering all climatic zones and over a long study period. The knowledge and analysis of the REC and RBS at the provincial level as the basic unit constitutes an important reference for policy makers. This research remains limited because a few studies on REC modelling have been conducted, taking Algeria as case study, which means that the housing typology has not been considered yet.

6.3. Implication on practice and future research

The results of this research provide a lever for policy makers for the revision of national energy management strategy and to broaden its scope. Actions that are appropriate to energy behaviour of each province can be applied, to evaluate the relevance of energy efficiency scenarios such as energy renovation and new low-energy housing. The objective is to achieve the target of reducing energy consumption in RBS, through the ratification of most of the agreements on climate change aimed at producing more sustainable cities. Future research will focus on the extension of REC model through the introduction of typology and characteristics variables of the Algerian RBS.

Appendix A: Estimated RBS data from 1995 to 2018 across all provinces

Appendix B: Estimated REC data from 1995 to 2018 across all provinces

Appendix C: Matrix of all variables integrated in Multiple Linear Regression model

7. References

Ahmed Gassar, A. A., Yun, G. Y., & Kim, S. (2019). Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy*, *187*, 115973. https://doi.org/10.1016/j.energy.2019.115973

Al-Ghandoor, A., Jaber, J. O., Al-Hinti, I., & Mansour, I. M. (2009). Residential past and future energy consumption: Potential savings and environmental impact. *Renewable and Sustainable Energy Reviews*, *13*(6–7), 1262–1274. https://doi.org/10.1016/j.rser.2008.09.008

Amri, F. (2017). The relationship amongst energy consumption (renewable and non-renewable), and GDP in Algeria. *Renewable and Sustainable Energy Reviews*, *76*, 62–71. https://doi.org/10.1016/j.rser.2017.03.029

APRUE. (2015a). *Programme Eco-Bat: Efficacité énergétique dans le bâtiment*. Retrieved from http://www.aprue.org.dz/prg-eco-bat.html. Accessed May 19, 2021

APRUE. (2015b). *Programme de développement de l 'efficacité énergétique à l'horizon 2030*. Retrieved from http://www.aprue.org.dz/documents/prog.develp.energ-2030.pdf. Accessed October 6, 2020

APRUE. (2015c). *Publications énergie dans les zones*. Retrieved from http://www.aprue.org.dz/documents/publication%20energie%20dans%20les%20zones-edition%202015.pdf . Accessed May 19, 2021

APRUE. (2017). *La consommation énergétique finale*. Retrieved from http://www.aprue.org.dz/documents/Consommation%20%C3%A9nerg%C3%A9tique%20finale.pdf. Accessed May 17, 2021

Aydinalp-Koksal, M., & Ugursal, V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, *85*(4), 271–296. https://doi.org/10.1016/j.apenergy.2006.09.012

Bélaïd, F., & Abderrahmani, F. (2013). Electricity consumption and economic growth in Algeria: A multivariate causality analysis in the presence of structural change. *Energy Policy*, *55*, 286–295. https://doi.org/10.1016/j.enpol.2012.12.004

Bianco, V., Manca, O., & Nardini, S. (2013). Linear Regression Models to Forecast Electricity Consumption in Italy. *Energy Sources, Part B: Economics, Planning, and Policy*, *8*(1), 86–93. https://doi.org/10.1080/15567240903289549

Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016). Prediction of residential building energy consumption: A neural network approach. *Energy*, *117*, 84–92. https://doi.org/10.1016/j.energy.2016.10.066

Boukarta, S., & Berezowska-Azzag, E. (2018). Assessing Households' Gas and Electricity Consumption: A Case Study of Djelfa, Algeria. *Quaestiones Geographicae*, *37*(4), 111–129. https://doi.org/10.2478/quageo-2018-0034

Bourdeau, M., Zhai, X. qiang, Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, *48*, 101533. https://doi.org/10.1016/j.scs.2019.101533

Brounen, D., Kok, N., & Quigley, J. M. (2012). Residential energy use and conservation: Economics and demographics. *European Economic Review*, *56*(5), 931–945. https://doi.org/10.1016/j.euroecorev.2012.02.007

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

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

Catalina, T., Iordache, V., & Caracaleanu, B. (2013). Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*, *57*, 302–312. https://doi.org/10.1016/j.enbuild.2012.11.010

Chen, H., Huang, Y., Shen, H., Chen, Y., Ru, M., Chen, Y., Lin, N., Su, S., Zhuo, S., Zhong, Q., Wang, X., Liu, J., Li, B., & Tao, S. (2016). Modeling temporal variations in global residential energy consumption and pollutant emissions. *Applied Energy*, *184*, 820–829. https://doi.org/10.1016/j.apenergy.2015.10.185

CNERIB. (1997). *D.T.R.C.3.2. Réglementation thermique des bâtiments d'habitation*. Retrieved from http://www.cnerib.edu.dz/. Accessed October 26, 2020

Corgnati, S. P., Fabrizio, E., Filippi, M., & Monetti, V. (2013). Reference buildings for cost optimal analysis: Method of definition and application. *Applied Energy*, *102*, 983–993. https://doi.org/10.1016/j.apenergy.2012.06.001

de Santoli, L., Mancini, F., & Astiaso Garcia, D. (2019). A GIS-based model to assess electric energy consumptions and usable renewable energy potential in Lazio region at municipality scale. *Sustainable Cities and Society*, *46*, 101413. https://doi.org/10.1016/j.scs.2018.12.041

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

Dujardin, S., Marique, A.-F., & Teller, J. (2014). Spatial planning as a driver of change in mobility and residential energy consumption. *Energy and Buildings*, *68*, 779–785. https://doi.org/10.1016/j.enbuild.2012.10.059

ONS. (2018). *Démographie algérienne 2008*. 32. Retrieved from https://www.ons.dz/spip.php?rubrique34. Accessed May, 8, 2021

Evola, G., Fichera, A., Gagliano, A., Marletta, L., Nocera, F., Pagano, A., & Palermo, V. (2016). Application of a Mapping tool to Plan Energy Saving at a Neighborhood Scale. *Energy Procedia*, *101*, 137–144. https://doi.org/10.1016/j.egypro.2016.11.018

Fichera, A., Inturri, G., La Greca, P., & Palermo, V. (2016). A model for mapping the energy consumption of buildings, transport and outdoor lighting of neighbourhoods. *Cities*, *55*, 49–60. https://doi.org/10.1016/j.cities.2016.03.011

Fonseca, J. A., & Schlueter, A. (2015). Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*, *142*, 247–265. https://doi.org/10.1016/j.apenergy.2014.12.068

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288. https://doi.org/10.1016/j.rser.2013.03.004

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

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, *47*, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Ghedamsi, R., Settou, N., Gouareh, A., Khamouli, A., Saifi, N., Recioui, B., & Dokkar, B. (2016). Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach. *Energy and Buildings*, *121*, 309–317. https://doi.org/10.1016/j.enbuild.2015.12.030

Gianniou, P., Liu, X., Heller, A., Nielsen, P. S., & Rode, C. (2018). Clustering-based analysis for residential district heating data. *Energy Conversion and Management*, *165*, 840–850. https://doi.org/10.1016/j.enconman.2018.03.015

Groppi, D., de Santoli, L., Cumo, F., & Astiaso Garcia, D. (2018). A GIS-based model to assess buildings energy consumption and usable solar energy potential in urban areas. *Sustainable Cities and Society*, *40*, 546–558. https://doi.org/10.1016/j.scs.2018.05.005

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

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

Huo, T., Ren, H., Zhang, X., Cai, W., Feng, W., Zhou, N., & Wang, X. (2018). China's energy consumption in the building sector: A Statistical Yearbook-Energy Balance Sheet based splitting method. *Journal of Cleaner Production*, *185*, 665–679. https://doi.org/10.1016/j.jclepro.2018.02.283

Husson, F., Le, S., & Pagès, J. (2017). *Exploratory Multivariate Analysis by Example Using R* (2nd ed.). CRC Press (Chapter 4).

IEA. (2020). *Key World Energy Statistics 2020*. 81. Retrieved from https://webstore.iea.org/download/direct/4093?fileName=Key_World_Energy_Statistics_2020.pdf . Accessed November 23, 2020

Iraganaboina, N. C., & Eluru, N. (2021). An examination of factors affecting residential energy consumption using a multiple discrete continuous approach. *Energy and Buildings*, *240*, 110934. https://doi.org/10.1016/j.enbuild.2021.110934

Jones, R. V., Fuertes, A., & Lomas, K. J. (2015). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*, *43*, 901–917. https://doi.org/10.1016/j.rser.2014.11.084

Kateb, K. (2003). Population et organisation de l'espace en Algérie. *L'Espace géographique*, *32*(4), 311–331. Cairn.info. https://doi.org/10.3917/eg.324.0311

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

Kim, Y.-J., Lee, S.-J., Jin, H.-S., Suh, I.-A., & Song, S.-Y. (2020). Comparison of linear and nonlinear statistical models for analyzing determinants of residential energy consumption. *Energy and Buildings*, *223*, 110226. https://doi.org/10.1016/j.enbuild.2020.110226

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

Mata, É., Wanemark, J., Cheng, S. H., Ó Broin, E., Hennlock, M., & Sandvall, A. (2021). Systematic map of determinants of buildings' energy demand and CO_2

emissions shows need for decoupling. *Environmental Research Letters*, *16*(5), 055011. https://doi.org/10.1088/1748-9326/abe5d7

Mattinen, M. K., Heljo, J., Vihola, J., Kurvinen, A., Lehtoranta, S., & Nissinen, A. (2014). Modeling and visualization of residential sector energy consumption and greenhouse gas emissions. *Journal of Cleaner Production*, *81*, 70–80. https://doi.org/10.1016/j.jclepro.2014.05.054

Ministère de l'Energie. (2014). *Bilan des Réalisations du Secteur de l'Energie et des Mines 2014*. Restrieved from https://www.energy.gov.dz/Media/galerie/bilan_des_realisations_du_secteur_2014_5b43758484b17.pdf . Accessed May 19, 2021

Ministère de l'Energie. (2020). *Bilan énergétique national - Année 2019*. Retrieved from https://www.energy.gov.dz/?article=bilan-energetique-national-du-secteur. Accessed November 23, 2020

Mora, D., Carpino, C., & De Simone, M. (2018). Energy consumption of residential buildings and occupancy profiles. A case study in Mediterranean climatic conditions. *Energy Efficiency*, *11*(1), 121–145. https://doi.org/10.1007/s12053-017-9553-0

Nishimwe, A. M. R., & Reiter, S. (2021). Building heat consumption and heat demand assessment, characterization, and mapping on a regional scale: A case study of the Walloon building stock in Belgium. *Renewable and Sustainable Energy Reviews*, *135*, 110170. https://doi.org/10.1016/j.rser.2020.110170

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

ONS. (2020). Démographie Algérienne 2018. Retrieved from https://www.ons.dz/spip.php?rubrique34. Accessed October 26, 2020

Österbring, M., Mata, É., Thuvander, L., Mangold, M., Johnsson, F., & Wallbaum, H. (2016). A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model. *Energy and Buildings*, *120*, 78–84. https://doi.org/10.1016/j.enbuild.2016.03.060

QGIS. (2021). *QGIS Geographic Information System*. Retrieved from https://www.qgis.org. Accessed January 6, 2019

R. (2021). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org/. Accessed January 6, 2019

Reiter, S., & Marique, A.-F. (2012). Toward Low Energy Cities: A Case Study of the Urban Area of Liége, Belgium. *Journal of Industrial Ecology*, *16*(6), 829–838. https://doi.org/10.1111/j.1530-9290.2012.00533.x

Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471. https://doi.org/10.1016/j.apenergy.2014.08.111

Sachs, J., Moya, D., Giarola, S., & Hawkes, A. (2019). Clustered spatially and temporally resolved global heat and cooling energy demand in the residential sector. *Applied Energy*, *250*, 48–62. https://doi.org/10.1016/j.apenergy.2019.05.011

Salari, M., & Javid, R. J. (2016). Residential energy demand in the United States: Analysis using static and dynamic approaches. *Energy Policy*, *98*, 637–649. https://doi.org/10.1016/j.enpol.2016.09.041

Sepehr, M., Eghtedaei, R., Toolabimoghadam, A., Noorollahi, Y., & Mohammadi, M. (2018). Modeling the electrical energy consumption profile for residential buildings in Iran. *Sustainable Cities and Society*, *41*, 481–489. https://doi.org/10.1016/j.scs.2018.05.041

Štreimikienė, D. (2014). Residential energy consumption trends, main drivers and policies in Lithuania. *Renewable and Sustainable Energy Reviews*, *35*, 285–293. https://doi.org/10.1016/j.rser.2014.04.012

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, *13*(8), 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033

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

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

Torabi Moghadam, S., Toniolo, J., Mutani, G., & Lombardi, P. (2018). A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society*, *37*, 70–84. https://doi.org/10.1016/j.scs.2017.10.002

Tso, G. K. F., & Guan, J. (2014). A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption. *Energy*, *66*, 722–731. https://doi.org/10.1016/j.energy.2014.01.056

Walter, T., & Sohn, M. D. (2016). A regression-based approach to estimating retrofit savings using the Building Performance Database. *Applied Energy*, *179*, 996–1005. https://doi.org/10.1016/j.apenergy.2016.07.087

Wang, S., Liu, H., Pu, H., & Yang, H. (2020). Spatial disparity and hierarchical cluster analysis of final energy consumption in China. *Energy*, *197*, 117195. https://doi.org/10.1016/j.energy.2020.117195

Wang, Y., Wu, T., Li, H., Skitmore, M., & Su, B. (2020). A statistics-based method to quantify residential energy consumption and stock at the city level in China: The case of the Guangdong-Hong Kong-Macao Greater Bay Area cities. *Journal of Cleaner Production*, *251*, 119637. https://doi.org/10.1016/j.jclepro.2019.119637

Williams, K. T., & Gomez, J. D. (2016). Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach. *Energy and Buildings*, *128*, 1–11. https://doi.org/10.1016/j.enbuild.2016.06.076

Zhu, D., Tao, S., Wang, R., Shen, H., Huang, Y., Shen, G., Wang, B., Li, W., Zhang, Y., Chen, H., Chen, Y., Liu, J., Li, B., Wang, X., & Liu, W. (2013). Temporal and spatial trends of residential energy consumption and air pollutant emissions in China. *Applied Energy*, *106*, 17–24. https://doi.org/10.1016/j.apenergy.2013.01.040