ORIGINAL ARTICLE

Understanding the energy behavior of households in the mountainous town of Metsovo, Greece

Anastasios Balaskas · **Ioanna Karani** · **Nikolas Katsoulakos** · **Dimitris Damigos** · **Dimitris Kaliampakos**

Abstract This article is a methodical attempt to understand the factors that influence energy consumption in households in the mountainous settlement of Metsovo, Greece. So far, most of the research on the settlement has indirectly approached the investigation of the factors that shape the energy behavior of households. In the present research, the identification of factors is directly approached through linear regression and clustering methods. Income, heating system, and household size were identified as the main factors influencing household energy expenditure. Since mountain areas are plagued by energy poverty, the study of household energy behavior inevitably highlights aspects of this phenomenon. By highlighting these factors and the spatial dimension of energy consumption (i.e., higher thermal energy needs in mountain areas), it was possible to suggest more targeted measures specifically designed for mountain areas, complementing the existing energy policy.

Keywords Energy profile · Energy poverty · Metsovo · Mountain areas · Clustering · Regression

A. Balaskas (✉) · I. Karani · D. Damigos ·
D. Kaliampakos
School of Mining and Metallurgical Engineering, National Technical University of Athens, 9 Heroon Polytechneiou Str, 15780 Zographos, Athens, Greece
e-mail: balaskas@metal.ntua.gr

I. Karani
e-mail: iwannakaranh@gmail.com

D. Damigos
e-mail: damigos@metal.ntua.gr

D. Kaliampakos
e-mail: dkal@central.ntua.gr

N. Katsoulakos
Academy of Merchant Marine, Aspropyrgos, Attica, Greece
e-mail: katsoulakos@metal.ntua.gr

Introduction

Energy consumption is closely related to social and economic development and is considered a crucial factor in humanity's evolution. The continuous industrialization of modern society, combined with economic growth and population growth, has led to a 44% increase in global energy consumption from 1971 to 2014 (Eren et al., 2019).

The residential sector accounts for an important share of energy consumption worldwide. In China, which has a large industrial sector, the residential sector is the second largest energy consumer (Wang et al., 2021). In the case of Europe, in 2019, the residential sector accounted for 26.3% of total energy consumption and an increase to 27% was recorded in 2020 (Eurostat, 2020a). At the European level the total household energy consumption is predominantly shaped by electricity and thermal energy consumption (Eurostat, 2023). The trend among Greek households aligns with the rest of Europe, as from the total energy consumption, 72% counts for thermal energy

and 28% for electricity (Hellenic Statistical Authority, 2013) and they represent the 39% of the total national energy consumption, followed by transport (36%) and industry (25%) for the year 2021 according to the International Energy Agency (IEA)[1].

It is estimated that the global population will increase by 66% by 2050 (UN, 2015) and the total global energy consumption will increase by 50% between 2020 and 2050 (EIA, 2021). The significant contribution of energy consumption regarding the social welfare in the light of the increase of energy demand underlines the need for further understanding of the factors shaping households' energy behavior, that can be classified under three main groups: socioeconomic data, dwellings characteristics and weather or/and climatic data. Weather and climatic data mainly affect the spatial differential of energy consumption (Auffhammer and Mansur, 2014).

A particular aspect of the spatial differentiation of energy consumption constitutes the mountainous areas. Especially Greece is classified among the most mountainous areas of Europe (Schuler et al., 2004). The unfavorable position of mountainous Greece regarding residential energy welfare has been identified through a dedicated survey for energy poverty (Papada & Kaliampakos, 2016a, 2016b). Expanding the research focusing on Greek households in mountainous areas, the present paper aims at enhancing knowledge in the following ways:

- Studying a typical mountainous settlement of Greece, Metsovo (1100 m altitude), using the method of personal interviews to achieve robust results.
- Adopting an integrated regression model for determining the factors forming energy consumption dedicated in a Greek mountainous settlement.
- Using established methods, such as k-Prototype clustering algorithm and the LSTM neural network, to enhance the results of the regression model and to deepen our understanding of energy behavior in Metsovo.
- Highlighting individual aspects of energy consumption.

- Supporting specific and effective energy policy measures for Greek mountainous areas.

The key contribution of the research is the use of metered electricity consumption data. As reflected in the Table 1, the studies that use metered consumption data in Greece are limited and none of them focus on mountainous areas. In this research, the results of the analysis of metered data are compared to the results of the analysis of the data collected through the questionnaires. The regression analysis identified the determinants that shape the level of household energy expenditure as (a) the type of main heating system, (b) the hours of operation of the heating system, and (c) the size of the dwelling. In addition, it highlights the inelasticity of energy expenditure to income, due to the objective need for heating in mountain areas. Of particular interest are the results of the clustering, whereby in both cases how income relates to energy expenditure is passed through the choice of fuel. The highest-income households—which form the smallest cluster—choose to heat with oil, compared to biomass burning—which is a less healthy but more economical option—chosen by lower-income households. This finding highlights the specific social and economic characteristics of mountain communities.

Overall, the use of the measurement data, the comparison of the analysis of the two data sets for the first time in the case of Greece, and the convergence of their results reinforces the validity of the questionnaire survey, which has been used as a priority over time in this area of research in the country.

Literature review

The factors that determine electricity consumption are of immense importance for understanding household energy consumption and for designing effective energy policies. A review of the literature reveals the multifactorial dimension of electricity consumption.

Determinant factors of residential energy consumption

In household energy consumption, the demographic characteristics of the households' members constitute a shaping element. Age and education level are highlighted through related work as an endogenous

---

[1] Greece 2023, Energy Policy Review of the International Energy Agency (IEA). https://www.iea.org/reports/greece-2023

Table 1 Summarized table of empirical studies on energy households' sector in Greece

Survey	Sample Data	Methodology	Key Findings
Botetzagias et al., 2014	285 multi-person households / Telephone survey/Athens 2012	Hierarchical binary logistic regression modeling	Age, gender and perceived behavioral control are statistically significant predictors of most electricity curtailment behaviors
Vogiatzi et al., 2018	451 questionnaires/ Athens 2016	Principal Component Analysis	Education, age and presence of economic crisis key elements of efficient energy consumption
Boemi et al., 2017	762 questionnaires/ mainly Western and Central Macedonia Greece 2015–2016	Structural Equation Modelling	Educational level, income & building characteristics are related with energy consumption and energy poverty
Santamouris et al., 2013	598 questionnaires / Greece 2010–12	Cluster Analysis	Income is a crucial factor for energy consumption
Sardianou, 2007	586 questionnaires/ Athens 2003	Regression Analysis	Income, family size, electricity expenditures, age and environmental awareness affect the energy conservation behavior of the households
Kostakis, 2020	Household Budget Survey and NUTS II region resolution, Greece 2017	Region-fixed effects OLS, Quantile Regression Models	Income, age, number of employed members in the household, and gender affect household electricity consumption
Sardianou, 2008	586 questionnaires/ Athens 2003	Regression Analysis	Age of the respondent, family size, households' annual income, dwelling's size and rate of occupancy affect heat consumption
Chatzikonstantinou et al., 2022	1.061 questionnaires/ Telephone survey/ Greece 2022	Descriptive and deductive statistical analysis	Income, employment status, disabilities, occupancy status affects the energy behavior of the households
Kotsila & Polychronidou, 2021	1801 questionnaires and real electricity consumption data of the households from a Greek electricity provider/ Greece 2017	OLS Regression Analysis	Determinants influencing the electricity consumption are the number of occupants, the size of the dwelling, the heating type, the heating and cooling hours, and the weather conditions
Santamouris et al., 2007	1110 questionnaires/ major Athen's area 2004	Cluster Analysis	Income and dwelling characteristics affect the energy expenditures of the households
Assimakopoulos & Domenikos, 1991	Greek household expenditure surveys 1974 and 1982	Principal Component Analysis	Socio-economic characteristics have a great influence in consumption (e.g. property status, education, income)
Papada & Kaliampakos, 2016a	400 Greek households	Descriptive and deductive statistical analysis	The findings showed that, under the objective expenditure-based method, 58% of Greek households are energy poor. Among households under the poverty threshold, the energy poverty rate exceeds 90%

Table 1 (continued)

Survey	Sample Data	Methodology	Key Findings
Papada & Kaliampakos, 2016b	400 Greek households in mountainous areas	Descriptive and deductive statistical analysis	The surveys are based on objective financial and expenditure data as well as on subjective views about housing conditions. The energy poverty rate skyrocketed to 73.5%. The subjective indicators reveal significant aspects of the problem
Balaskas et al., 2021	300 households in Metsovo	Descriptive and deductive statistical analysis	9/10 households in Metsovo are exposed to energy poverty, based on the objective 10% index. The subjective indicators revealed other interesting aspects of the problem. Over one-third of the local population live in houses that are inadequately heated and/or present damp problems. Almost two-thirds of the local population have been forced to restrict other basic needs, to respond to high energy costs

determinant of behavior, is known to affect an individual's vitality utilization through different organic, mental, financial, and social components. Considers have appeared that more seasoned people tend to require more warming administrations amid winter, whereas more youthful individuals tend to spend more time online and utilize electronic gadgets (Li et al., 2023). More specifically, the elderly household members (Ota et al., 2018) and the number of household member (Kotsila & Polychronidou, 2021) are related with higher energy consumption, while the higher education level is inversely connected with energy consumption (Ali et al., 2021).

Income is one of the most crucial factors determining energy expenses and therefore the consumption of a household (Wen-Hsiu Huang, 2015). In addition, the fuel price and the electricity price per kWh have a substantial impact on the amount of energy consumed (Ye et al., 2018). In conjunction with the abovementioned economic factors, the type of the main heating system used by a household and the operation hours significantly determines the energy cost, considering both the efficiency of the system and the price of the relative fuel (Bedir et al., 2013).

Dwellings' characteristics, such as type of dwelling, the area of the dwelling and the ownership status have been identified as determinants of the energy behavior of a household (Su, 2019). The building year of a dwelling is also related to the energy consumption, in the light of the less effective insulation and older heating systems (González et al., 2013). Especially for electricity consumption the household appliances are related with the consumption of the household (Zhou & Teng, 2013).

Except the socioeconomic and technical characteristics of the household there are neutral elements that form the energy behavior of the households, i.e. climatic and weather conditions. Cooling and heating degree days or cooling and heating calorimeters equivalently are measures of the actual cooling and heating energy needs and are positively related with higher energy thermal and electricity consumption (Kotsila & Polychronidou, 2021; Wang et al., 2021). Considering the climatic conditions' changes, heating energy demand is expected to be reduced, while cooling energy demand is expected to be increased, in Europe and USA, unveiling the dynamic profile of energy consumption (Petri & Caldeira, 2015).

The spatial dimension of energy consumption

Energy consumption has a spatial dimension, not only between different countries but also within each country. Nakagami et al. (2008) conducted a survey of household energy consumption in eighteen countries. Within this survey, energy consumption for China, India, and Thailand is divided into urban and rural categories. In all three cases, the total energy consumption in the rural part of the country was recorded to be increased, while in the case of China, where both urban and nonurban thermal needs are noted, these needs are significantly higher in rural China. Higher population densities are linked to lower levels of HEC, according to numerous studies (Chen et al., 2022). Urbanity, or population density, is a factor that contributes to lower energy use, according to a comparative study on the HEC of households in Brazil, Australia, Japan, Denmark, and India that revealed the HEC of large metropolitan areas to be significantly lower than that of rural areas (Lenzen et al., 2006). A study conducted on fourteen distinct statistical zones in Sydney demonstrates a significant correlation between a zone's lower energy intensity and its higher levels of urbanity, or population density.

In the European Union, the residential sector plays a significant role in overall energy consumption and is expected to maintain this importance in the coming thirty years, as projected by the EU Commission in their EU Reference Scenario. This projection relies on the assumption that the EU Commission's energy efficiency initiatives, such as the Energy Efficiency Directive (EED) and the Energy Performance of Buildings Directive (EPBD), are effectively implemented.

Many researchers have used Geographic Information Systems (GIS) to analyze the spatial relationships between energy consumption and cities by studying urban forms (Quan & Li, 2021), land-use patterns (Zhao et al., 2017), morphological and functional aspects (Chen et al., 2022), and built environment characteristics (Balaskas et al., 2021). These studies highlight the importance of using data at different spatial scales to address energy consumption concerns. However, there is a lack of research on simpler approaches to studying interactions within complex urban and/or rural systems. Understanding the spatial dimensions of energy consumption is crucial for addressing climate change and transitioning to sustainable energy practices in urban areas.

The spatial differentiation of energy consumption is already mentioned and results from both climate and socioeconomic factors. Per capita heating consumption has strong latitude correlation and heating energy needs decrease from north to south (Ma et al., 2021). Altitude has been identified as a determinant factor for the energy needs of a household and especially heating energy consumption. It is estimated that a typical dwelling has almost double the thermal energy needs at 1200 m compared to 200 m, regardless of its energy efficiency class (Papada & Kaliampakos, 2016a, 2016b), while heating degree days are linearly related with altitude (Katsoulakos and Kaliampakos, 2016).

Mountainous areas are not defined exclusively by altitude, but also by inclination and isolation. Terrain inclination sets barriers -neutral and economical- at the expansion of fuel grids and the energy availability mix in mountainous areas is relatively restricted. The fuel prices increase because of the delivery services (Gohari et al., 2018). Fuel delivery services cover moving fuel from refineries to consumers. The cost depends on how far the fuel is needed to travel and how it gets there. Mountain settlements are often far away from major energy production and distribution centers, and as a result the fuel prices are higher comparatively to the lowland areas.

It is common behavior among households in mountainous or rural areas to use biomass-based heating systems, which are less energy efficient, due to inability to afford more environmentally friendly heating systems. From December 2010 to February 2011, incidents of illegal logging were recorded in mountainous regions of Greece (Peklaris, 2010), as many people could not afford diesel or other conventional fuel type and were desperate to find how to meet its thermal energy needs.

To explore their relationship with the energy sector, the main characteristics of mountain areas: altitude, slope, remoteness, lack of productive activities, and old buildings/landscape must be considered. Mountainous areas have common characteristics that make them highly vulnerable to energy poverty as many mountain economies are not particularly strong (Funnell & Parish, 2005), mountain settlements face difficulties in sitting energy projects in the restricted usable mountainous terrain.

Energy poverty and its spatial dimension

Energy poverty, which describes the inability of households to afford adequate access to energy services, is an issue that can have a significant effect on the quality of life and even the state of health of individuals and, subsequently, on the overall living standard of a country.

Boardman ([1991]) offers a broader definition according to which a household is fuel poor if it cannot attain adequate energy services for less than 10 percent of its net income. Apart from the 10% objective indicator, qualitative indicators that aim to capture the subjective perception of households regarding the effects of the phenomenon on their everyday lives are widely used. The most common indicators of this category are the three indicators monitored by Eurostat ([2020a], [2020b]), i.e. Inability to keep home adequately warm, Arrears on utility bills and Leakages/damp walls/mold problems. A complex phenomenon like energy poverty is recommended to be approached using both objective and complementary subjective indicators, in order to illuminate its different aspects (Roberts et al., [2015]).

The mountainous areas of Greece experience energy poverty more severely, highlighting household energy consumption as a direct problem in these areas, as being burdened with considerably increased energy costs versus lowlands, accompanied by lower incomes, at the same time. Indicatively, Papada and Kaliampakos ([2016a], [2016b]) calculated based on the conventional 10% energy poverty index the percentage of mountain households classified as energy poor, which amounts to 73.5% compared to 58% for the country. Balaskas et al., ([2021]) calculated that the percentage of households classified as energy-poor in the settlement of Metsovo, a mountainous settlement in Greece, based on the 10% conventional index is 90%. Karani et al. ([2022]) studying energy poverty in the municipality of Agrafa, one of the most mountainous municipalities in Greece, estimated the percentage of energy-poor households at 62.4% based on the conventional energy poverty index and 91.3% based on the required energy consumption.

Therefore, the discrete study of household energy consumption for mountainous areas is important both in the short term, in the sense of immediately meeting their energy needs more efficiently and improving the living standards of the inhabitants through targeted policies, and in the longer term in the broader context recorded, especially in mountainous countries such as Greece, which is characterized as mountainous in more than 70% (Schuler et al., [2004]).

The households' energy consumption and behavior research in Greece

In Table [1], a non-exhaustive overview of the research concerning energy consumption of the Greek residential sector and its various aspects through the last twenty years is presented. The sample data, the methodology used, and the key findings are listed as well.

Methodologies applied in energy consumption prediction and research.

Various approaches and procedures have been proposed for determining the energy consumption of building stocks using different energy performance modeling tools. Machine learning algorithms, Neural Networks models, and classical statistics methods -especially ARIMA model and models based on ARIMA- are used to predict energy consumption.

Several comparative studies have been conducted, which examine the prediction of future time series values using neural networks and statistical models. Indicatively, Dubey et al. ([2021]) used data from smart meters and used the ARIMA, SARIMA, and LSTM models to predict future energy consumption prices. They used data from 5,567 London households with a reference period of November 2011 to February 2014 and the electricity consumption data has a period of half an hour. The statistical measures RMSLE, RMSE, MASE, and MAPE were used to compare the performance of the models, which together indicated the superiority of the LSTM.

Annual data on Thailand's energy consumption from 1986 to 2010 were used from Kandananond ([2011]) for the prediction of the country's future energy consumption using the ARIMA model, ANN of the MLP architecture, and multiple linear regression. The MAPE statistical measure indicated the neural networks as the most efficient, while the regression method recorded the worst performance.

Nugaliyadde et al. ([2019]) used recurrent neural networks (RNN) and the LSTM model to predict the energy consumption of individual houses and entire blocks. The results of these two methods

were compared with the results of ARIMA, ANN, and DNN, using the statistical measure RMSE. The methods were tested for both short, medium, and long-term forecasting and it was shown that in all three cases, RNN and LSTM outperformed the others, while only in the case of short-term forecasting ARIMA showed similar performance.

Guenoukpati et al. (2021) used data on hourly household energy consumption from 1 January 2015 to 31 December 2019 and used ARIMA, LSTM, ANN, SVR (an adaptation of the SVM machine learning algorithm to the regression problem) and the multiple linear regression method to predict future values. The fit of the models was evaluated using the statistical measures RMSE, nRMSE, MAPE and R2-value and overall, it was found that the neural network and machine learning models outperform the classical ARIMA and multiple linear regression models.

Overall, neural network and machine learning models can predict future time series values more accurately in the case of energy consumption compared to statistical models, while the comparison of their performance gives similar results and flexibility of choice.

Based on the above, the LSTM model, which is based on Convolutional Neural Networks (CNN), was selected in this paper to predict the future prices of electricity consumption, due to the proven well performance of the model compared to other popular methods on the electricity consumption prediction and the ease of configuration and implementation.

Apart from predicting the energy consumption of households, unveiling the factors shaping their energy consumption is a significant field of survey. Linear regression is widely used to predict, model and understand household energy consumption not only in case of Greece as stated in Table 1, but worldwide (e.g. Al-Ghandoor et al., 2009; Filippini & Pachauri, 2004; McLoughlin et al., 2012). In addition to linear regression, clustering techniques have been applied to understand the energy behavior of households, which aim to divide the population or data into clusters (groups), so that the subjects in the same cluster have more similarities with each other than with the objects/subjects in the other groups. The main algorithms that have been used for this purpose are k-mode for qualitative variables (e.g. Dorman & Maitra., 2022), k-means for qualitative variables cases

(e.g. Dent et al., 2014), but also methods for mixed data such as Principal Component Analysis—PCA (e.g. Vogiatzis et al. 2018) and Hierarchical Clustering Anaysis (e.g. Santamouris et al. 2007).

The previous mentioned statistical methods involve a number of variables for modeling and clustering: socio-economic (household size, population, population, household type, ownership status, GDP, income, energy prices), climatic (temperature, humidity, solar radiation, climate zones, climate change, heating and cooling calorimeters), technical building characteristics (year of construction, floor area of dwelling, building envelope, type of dwelling, home appliances and energy heating systems), as already highlighted in the previous sections.

Given the above both linear regression analysis and clustering techniques were applied on the data sets used in this survey, to decode the energy consumption behavior of the participants and produce results comparable to the existing studies. Multiple linear regression model was used, due to its simplicity and explanatory effectiveness, while concerning clustering, the k-Prototype algorithm was chosen – a combination of k-means and k-modes algorithms for mixed data- as no other study implementing this algorithm in the case of Greece is stated.

Summarizing, the methodology used to approach the data sets described in Section 3.1 are:

a) Electricity consumption prediction with LSTM model where needed (3.2.1)
b) Thermal energy needs modelling (3.2.2)
c) Clustering with k-Prototype algorithm (3.2.3)
d) Construction of linear regression model (3.2.4)

Both the data and the methods used are described in detail in Section 3.

Data and analytical methods

Data

Both the determinants of household energy consumption and the common characteristics of households are examined to reveal the energy profile of households in Metsovo. Two data sources were used for this scope. The first data source comes from a socioeconomical survey in the region. The survey

was conducted during 2019–2020 using a detailed questionnaire circulated among households and their responses were collected through personal interviews. The sample size was estimated at 300 households, with a 95% confidence interval and a maximum margin of error of 5%. The selection of participants was random. Respondents were selected on a random basis at different locations of Metsovo and at different parts of the day to ensure a cross-section of residents. According to Hellenic Statistical Authority (2011) the settlement of Metsovo includes 1409 residences, of which approximately 890 are permanently occupied. The interviews were carried out door to door. The questionnaire consisted of 24 'closed-ended' questions on living conditions, housing infrastructure, heating systems, subjective perception of energy coverage and quality of life, quantitative data on energy expenditure and household energy behavior and 17 questions on demographic and income data (such as age of family members, gender, education, occupation, family income, expenditure on various household needs and savings. Data regarding electricity and thermal energy expenditures were collected directly through the answers of the participants to the relevant questions.

In addition, a second survey with the same structure and methodology as described above was carried out on 60 households where smart meters were installed. The monitoring equipment, which was installed, included the following:

- Indoor temperature and humidity meters (meteorological monitors) with external sensors, located in three different rooms.
- Electricity consumption meters, measuring in real time the electricity consumption of households.

The monitoring equipment allowed recording of the above parameters (i.e., electricity consumption, temperature, and humidity) in the form of actual time-series data. The data were also used to predict the annual energy consumption of households in order to use the total electricity consumption in the cases of households where the measurements that were recorded did not represent a whole year, due to technical problems of the smart energy meters. The metering data of the 60 households were recorded in 2 stages. The first, involving the initial 30 households, took place from April 2019 to October 2019. The next one that took place on the subsequent 30 households, was conducted from November 2019 to February 2021. The thermal energy consumption of the households was estimated following the methodology presented in paragraph 2.2.2. The annual energy expenses for each household were calculated using the above-mentioned consumptions and the prices mentioned in paragraph 2.2.2.

Energy consumption prediction

Deep LSTM model for estimating households' electricity consumption

To predict the annual electricity consumption a deep LSTM (or stacked LSTM) model -which results from the stacking of LSTM layers- was run individually for the timeseries of energy consumption of each household, that did not cover a whole year of electricity consumption measurements. The proposed LSTM model structure is presented in Fig. 1:

Deep LSTM model was chosen as the preferred model to predict the annual energy consumption considering the evidence described in paragraph 1.1.4. LSTM neural networks are classified as recurrent neural networks (RNNs), with the main difference observed in the complexity of the building blocks used in the hidden layers, which include an input gate, a selective gate, an output gate, and a memory cell (Zhang et al., 2018).

In general, the structure of LSTMs can vary depending on the number of inputs and outputs (one to one, one to many, many to one, many to many). Each cell in the LSTM network receives information modeled by previous cells, as well as the current data x_t. The influence of the previous data on the current state is determined through a series of logic gates. Each gate determines which information is retained and which is not, depending on the value the function takes, with values close to 0 being equivalent to rejection and values close to 1 being equivalent to acceptance. The input gate determines which new information is used to upgrade the internal memory of the current unit ct, which is upgraded based on the internal memory of the previous time, the selective gate, and the new hidden state. In the final stage, the output gate determines the output signal, which will be the input signal for

the next hidden state (Hochreiter and Schmidhuber, 1997).

At the first level, the stationarity of the time series was checked. The condition of stationarity is particularly important in the problem of predicting the future values of a time series (Kugiumtzis, 1999). The Dickey-Fuller statistical test was used for this purpose. This test is based on the empirical value of the t-statistic from a simple regression, but the comparison for the acceptance or rejection of H0 is not with values from the t-distribution but with values empirically determined by MacKinnon (1992). The null hypothesis was rejected for all the time series under consideration and therefore in their entirety they are characterized as stationary.

Then, the gaps in the time series were filled by the linear interpolation method, a simple and popular method for creating a full timeseries (Lepot et al., 2017). The change in consumption of the immediately preceding and immediately following measurement is used to estimate the intermediate, blank value, assuming each time that these values are linearly related. After the timeseries were fully completed, the data were normalized using the MinMaxScaler function.
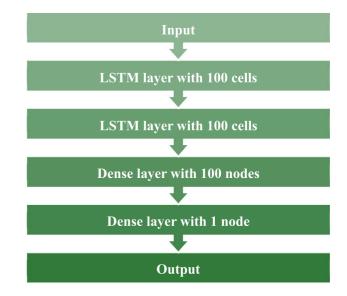
Finally, before training and running the model, a training data set and a control data set were defined for each time series. The volume of data corresponding to each set varies for each timeseries and in each case was determined through testing and error estimation.

To train the model, a widely used algorithm -ADAM algorithm- was used. In the ADAM algorithm, a training rate is maintained for each parameter (weight) of the network and is adopted separately in the weights as the training progresses. In practice it calculates individual training rates for each parameter and does so by approximating the first and second order moments of the derivatives (gradients).

The statistical measure "Mean Squared Error— MSE" was used to evaluate the results of the model. The MSE is calculated as the average of the difference between the actual and the estimated value squared. It is one of the most widely used statistical evaluation measures for prediction in machine learning, along with "Mean Absolute Error- MAE" and "Mean Absolute Percentage Error -MAPE" (Botchkarev, 2018) and the smaller and closer to 0 the values of MSE, the better the model performance is considered.

The parameters used in the proposed model are presented in Table 2.

For those households for which there was sufficient data, and it was necessary to predict annual energy consumption, the model with the architecture described above was used. In predicting the future values of each time series, the training size, lookback, batch size, epochs and early stopping patience were adjusted separately for each time series through empirical tests. In each case the MSE value is smaller than 0.5, indicating a good fit of the model to the data. The model was then used to predict the hourly consumption of households.

Fig. 1 The architecture of the proposed LSTM model

The limitations of the proposed LSTM model can be attributed to the volume of training and control data. The available data for prediction were multiples of the future data that each model attempted to predict each time, and in this case the model was in most cases asked to calculate a larger number of future values than the available prior values. Longer data series would provide more accurate results to avoid overfitting because of the small size of the dataset.

Modeling thermal energy needs of the households

The total energy needs of households require the calculation of their thermal needs in addition to their electricity consumption. To calculate the estimated energy needs, several parameters related to housing characteristics were used, such as year of construction, floor area, type, building envelope characteristics (e.g., roof and type of windows), cooling, heating, hot water systems, and lighting. The energy consumption of the building for heating, cooling, and domestic hot water (DHW) needs and the corresponding costs were calculated by applying energy estimation methods and rules. Specifically, based on the above data, the lateral surface area of the building and the surface area of walls and openings to the external environment are calculated.

The total heat transfer (H_{tot}) coefficient is then estimated, which is the sum of the heat transfer coefficient due to thermal conductivity the building elements (H_{tr}) and the heat transfer coefficient due to ventilation (H_{ven}). Some basic equations of applied thermodynamics are presented in the following lines for determining the heat transfer coefficients and then the energy demand. The core of these equations has been included in the Greek Regulation of Buildings' Energy Performance (acronym KENAK).

The heat transfer coefficient due to thermal conductivity is calculated as follows:

$$H_{tr} = \sum_i A_i \bullet U_i \tag{1}$$

where,

A_i the surface of the building element i.

U_i the thermal conductivity coefficient of the building element i.

Regarding the heat transfer coefficient due to ventilation, the following mathematical expression is used:

$$H_{ven} = \rho_{air} \bullet C_{air} \sum_k A_k \bullet q_k + \rho_{air} \bullet C_{air} \bullet q_{ven} \tag{2}$$

where,

ρ_{air} the density of atmospheric air.

C_{air} the specific heat capacity of atmospheric air.

A_k the surface of window/ door k.

q_k the flow rate of air from window/ door k into the house.

q_{ven} the flow rate of air due to air ventilation systems and natural ventilation.

Once the heat transfer coefficient has been calculated, together with the Heating Degree Days—HDD (which are derived from the location of the building), the energy demand for heating is (in kWh):

$$E_{th,d} = H_{tot} \bullet HDD \bullet \frac{24}{1000} \tag{3}$$

This mathematical expression of heating demand is based on a stable situation approach (heating degree method), which is the simplest way of estimating energy demand. The heating degree method is considered particularly reliable and well documented (Büyükalaca et al., 2001, Matzarakis and Balafoutis, 2002). Even though more sophisticated models give better estimations of energy demand, studies aiming to gain a generic view of energy demand and approach policy issues can be based on simple models. This leads to faster results which can be more easily edited and managed (Moschou, 2011).

Consequently, the energy consumption for heating (in kWh) is given by:

Table 2 LSTM layers parameters

Layer	Parameters	Layer	Parameters
1st LSTM layer	Units = 100	2nd LSTM layer	Units = 100
	Activation = tanh		Activation = tanh
	Recurrent Activation = sigmoid		Recurrent Activation = sigmoid
	Unit_forget_bias = True		Unit_forget_bias = True
	Kernel_regularizer = None		Kernel_regularizer = None
	Bias_regularizer = None		Bias_regularizer = None
	Activity_regularizer = None		Activity_regularizer = None
	Kernel_constraint = None		Kernel_constraint = None
	Recurrent_constraint = None		Recurrent_constraint = None
	Bias_constraint = None		Bias_constraint = None
	Dropout = 0		Dropout = 0
	Recurrent dropout = 0		Recurrent dropout = 0
	Return_sequenses = True		Return_sequenses = True
	Return_state = False		Return_state = False
	Go_Backwards = False		Go_Backwards = False
	Stateful = False		Stateful = False
	Use_bias = True		Use_bias = True
	Unroll = False		Unroll = False
	Time_major = False		Time_major = False
1st Dense layer	Units = 100	2nd Dense layer	Units = 1
	Activation = None		Activation = None
	Use_bias = False		Use_bias = False
	Kernel_initializer = None		Kernel_initializer = None
	Bias_initializer = None		Bias_initializer = None
	Kernel_regularizer = None		Kernel_regularizer = None
	Activity_regularizer = None		Activity_regularizer = None
	Kernel_constraint = None		Kernel_constraint = None
	Bias_constraint = None		Bias_constraint = None

$$E_{th,c} = \frac{E_{th,d}}{n} \quad (4)$$

where, the parameter n is the efficiency of the heating system used.

As far as energy for producing domestic hot water (DHW) is concerned, the methodology proposed by KENAK is followed. More specifically, the energy demand for DHW is calculated as follows:

$$E_{DHW,d} = V_d \cdot C_w \cdot \rho_w \cdot \Delta T \quad (5)$$

where:

V_d the annual volume of DHW necessary (according to KENAK this is equal to 27.38 m³/year for each bedroom).

C_w the specific heat capacity of water.

ρ_w the density of water.

ΔT the temperature difference between DHW and water from the water supply.

And so, the corresponding energy consumption for DHW is:

$$E_{DHW,c} = \frac{E_{DHW,d}}{n} \quad (6)$$

where the parameter n is the efficiency of the DHW production system used. If a solar water heater is

present, then the final consumption of DHW is 40% of the one calculated based on the previous equations.

Finally, based on the above calculations regarding energy consumption, the energy cost is estimated. In particular:

For systems using oil, natural gas/ LPG, pellets, firewood, the equation is as follows:

$$\text{Cost} = \frac{EnergyConsumption}{CalorificValue} \cdot \text{Fuel cost} \qquad (7)$$

For systems using electricity the equation is as follows:

$$\text{Cost} = \text{Energy Consumption} \cdot \text{Electricity Price} \qquad (8)$$

It is not easy to make specific assumptions for energy prices. The liberalized energy market, combined with geopolitical developments, leads to significant fluctuations in energy prices. Since the primary surveys in Metsovo were conducted between December 2018 and March 2019, the energy prices used in the present paper correspond to the situation of the Greek energy market in this period. Regarding diesel oil price and LPG price, according to the Observatory of Fuel Prices,[2] a representative average is 0,8 €/lit for diesel oil and 0,5 €/lit for LPG. According to Eurostat,[3] electricity prices for households were about 0.11 €/kWh in Greece in 2018–2019. It is more realistic to add to the energy price some extra charges and levies, which increase the electricity price to the value of 0.2 €/kWh. Finally, as far as pellets are concerned, the price of 250 €/tn was assumed to be a representative average. This was estimated during the conduction of the primary surveys by a short market survey in the region of Epirus. The estimation seems to be in accordance with relevant analyses in energy websites.[4]

k-Prototype clustering algorithm

The multivariate nature of the household energy consumption phenomenon requires the use of both quantitative and qualitative variables (mixed data) and therefore it was not possible to use traditional statistical clustering methods that are mainly specialized in one type of data. This paper uses the k-Prototypes algorithm, a combination of k-means (quantitative data) and k-modes (quantitative data) methods.

The k-prototypes algorithm was developed by Huang in 1997 and is a clustering technique based on the k-means algorithm, extending its use beyond numerical data (Huang, 1998).

The same statistical distance measure as the k-means algorithm is used to handle numerical data, i.e., Euclidean distance, and the distance function is defined as a measure of similarity between two objects. Regarding non-numerical variables, different values of these variables also imply a reduction in the similarity of a pair of observations (Akay & Yüksel, 2018).

In practice, the operation of the algorithm is based on three processes: the initial selection of the models, the initial allocation of observations and their reallocation. The process is carried out in the following four steps:

(1) Selection of k original prototypes from a dataset X.
(2) Allocate each element of X to the cluster with whose prototype it has the highest similarity based on Eq. (8).
(3) After all elements are partitioned, their similarity to the prototypes as defined for each iteration is rechecked. If an element shows greater similarity to the prototype of a class other than the one in which it has been placed, then the element is placed in the correct class and the prototypes of the two classes are recomputed.
(4) Repeat step 3, until there is no change in the classes for a complete check of the similarity of X elements and prototypes.

Linear regression model for households' annual energy consumption

In simple linear regression there is only one independent variable x and one dependent variable y,

---

approximated as a linear function of x. The value yi of y, for each value xi of x, is given by:

$$Y = a + b * Xi + e\_i \tag{9}$$

The linear regression problem is to find the parameters a and b that best express the linear dependence of y on x. Each pair of values (a,b) defines a different linear relationship expressed geometrically by a straight line and the two parameters are defined as follows:

- The constant a is the value of Y for $Xi = 0$
- The coefficient b of Xi is the slope of the line or the regression coefficient. It expresses the change in the variable y when the variable x changes by one unit.

In the case where the independent variable y depends linearly on more than one variable, it is referred to as multiple linear regression and is practically an extension of the simple linear regression model.

The equation describing the relationship between dependent and independent variables is:

$$\Upsilon_i = b_o + b_1 * X_{1i} + b_2 * X_{2i} + b_3 * X_{3i} + \cdots + b_k * X_{ki} + \varepsilon_i \tag{10}$$

The estimation of multiple linear regression is not substantially different from the single linear regression model. What is required to ensure is that the independent variables are zero correlated ($\rho(X_i, X_j) \rightarrow 0j$), to identify if all or a subset of the dependent variables will be used for the calculation of the model.

In linear regression the parameters are estimated by the least squares method, i.e. the coefficients are computed such that the sum of the squares of the differences between the observed and the estimated is the minimum.

Moreover, logarithmic transformation was used as an alternative of the multiple linear regression, to strengthen the regression results. The logarithmic transformation is a widely used data manipulation in the case of linear regression, as it allows the linearization of the relationship between the independent and dependent variable in case it does not

exist in the first place, or the possibility of transforming the asymmetric distribution of a variable into a distribution that better approximates the normal distribution through its logarithmic transformation, thus ensuring the reliability of the model (F). Logarithmic transformation can be applied to both the dependent and independent variables. In the case where both the dependent variable and one or more of the independent variables are logarithmically transformed, it is possible to calculate the elasticity of the dependent variable, i.e. the percentage change in the dependent variable caused by the unit percentage change in the independent variable.

The method chosen to calculate the model is the stepwise regression method. This is a method of selecting a "good" subset of independent variables. At each step, the null hypothesis H0: $\beta j = 0$, i.e. the coefficient of the independent variable is tested for all possible independent variables to exclude those for which the values of the statistical function are less than the predefined critical level. In this case the rejection and acceptance criteria are defined by the statistical package as follows:

- The variable is excluded from the model if the p-value of the statistical function F is greater than or equal to 0.1.
- The variable is entered into the model if the p-value of the statistical function F is less than or equal to 0.05.

Moreover, in the case of 60 households, the results of the linear regression model are presented because although it was implemented for a small set of data fits well and supports the results of the 300 questionnaires.

Results

Multiple regression on the social survey

Multiple regression was applied to identify the main social and economic factors that affect the annual energy expenses of the households in Metsovo. In

Table 3, the abbreviations for each potential factor are presented.

Categorical variables, namely, TPHS and HOHS, were converted into dummy variables, representing the choice of each household for the primary heating system and the hours of operation of the heating system. The skewness and kurtosis of the distribution of each variable were examined to determine whether they satisfactorily approximated the normal distribution. The AH variable was detected to be highly skewed (skewness $= 1.5 \pm 0.14$); therefore, the variable was log-transformed using the decimal logarithm. The logAH was found to be acceptably skewed (skewness $= 0.10 \pm 0.14$).

Before exploring the regression analysis, the correlations among the variables needed to be evaluated. In Table 4, the results of the examination of multicollinearity are listed.

As observed in Table 4, high multicollinearity occurs between the variables HOHS3 and HOHS4 (|Pearson correlation| $= 0.875 > 0.8$). Therefore, the variable HOHS3 is excluded from the linear model calculation. Table 5 presents the regression coefficients (B), F values, and p values of the statistically significant factors.

According to the multiple regression analysis, the best model for estimating annual energy expenditure (AEE) is as follows:

$$
\begin{aligned}
AEE = &-1799.654 + 615.078 * TPHS_1 + 1903.936 \\
&* lohArea + 4036.758 * TPHS_3 + 0.020 * AI \\
&- 435.786 * TPHS_4 + 408.096 * HOHS_4 \\
&+ 970.246 * HOHS
\end{aligned}
\tag{11}
$$

The best fitted model was selected by the algorithm itself as stated at Section 3.2.4.

Furthermore, the model was validated through the analysis of the residuals. The assumptions of independence, homoscedasticity, linearity, and normality of the residuals were tested.

For the independence hypothesis, the Durbin-Watson statistic was calculated. The Durbin Watson (DW) statistic is a test for autocorrelation in the residuals from a statistical model or regression analysis. The Durbin-Watson statistic will always have a value ranging between 0 and 4. A value of 2.0 indicates there is no autocorrelation detected in the sample. Values from 0 to less than 2 points

to positive autocorrelation and values from 2 to 4 means negative autocorrelation (Evans, 2010). The value of the statistic meter for the model is equal to 2.006, and therefore, the independence hypothesis is satisfied.

The core premise of multiple linear regression is the existence of a linear relationship between the dependent (outcome) variable and the independent variables. This linearity can be visually inspected using scatterplots, which should reveal a straight-line relationship rather than a curvilinear one. The scatter plot, where the horizontal axis represents the independent variable and the vertical axis represents the residual values, was displayed. In this case, at most $0.05*233 \approx 12$ points are expected to be outside the interval [-2,2]. The diagram shows that 10 points lie outside the above interval, and thus, it is concluded that the assumption of linearity is also satisfied. In addition, the relatively stable vertical range of the residuals indicates that the assumption of homoscedasticity for multiple linear regression is not violated. Finally, the Q−Q plot of the residuals suggests near-perfect linearity and hence normality of the residuals. This is also reflected in the histogram of the residuals. The distribution of the residuals closely approximates the normal distribution, showing a slight positive asymmetry.

The Variance Inflation Factor (VIF) is a diagnostic tool used to detect multicollinearity in a regression model. VIF measures the degree of collinearity between each predictor and the other predictors in the model (Pérez et al., 2009). Specifically, the VIF of a predictor quantifies how much the variance of the estimated regression coefficient is inflated due to collinearity with other predictors.

Table 3 Abbreviations for each factor used in multiple regressions

FACTOR	ABBREVATION
Annual income	AI
Year of house construction	YHC
Type of primary heating system	TPHS
Area of the house	AH
Hours of operation of heating system	HOHS
Number of households members	HM
Number of rooms in the house	HR
Annual energy expenses	AEE

Table 4 Correlations among factors

FACTORS	TPHS1	TPHS2	TPHS3	TPHS4	TPHS7	TPHS9	TPHS10	TPHS11	TPHS12	HOHS1	HOHS2	HOHS3	HOHS4	AI	YHC	HM	HR	logArea
TPHS1	1	-0.111	-0.557	-0.368	-0.07	-0.049	-0.086	-0.058	-0.111	0.018	0.103	0.086	-0.127	0.072	-0.004	-0.216	-0.084	-0.057
TPHS2		1	-0.095	-0.061	-0.011	-0.007	-0.013	-0.017	-0.017	-0.015	-0.024	-0.065	0.075	-0.011	0.054	0	-0.033	-0.05
TPHS3			1	-0.345	-0.6	-0.042	0.066	-0.95	-0.041	0.036	-0.136	-0.076	0.118	0.073	0.19	0.264	0.133	0.127
TPHS4				1	-0.038	-0.027	-0.047	-0.061	-0.061	-0.055	-0.039	-0.024	0.053	-0.163	-0.251	-0.082	-0.07	-0.113
TPHS7					1	-0.005	-0.008	-0.011	-0.011	-0.009	0.213	-0.041	-0.048	0.077	0.045	-0.038	0.019	-0.009
TPHS9						1	-0.006	-0.007	-0.007	-0.007	-0.011	0.115	-0.1	0.078	0.19	0.001	-0.128	-0.011
TPHS10							1	-0.013	-0.013	-0.012	0.168	-0.05	-0.02	0.102	0.003	0.002	-0.058	0.065
TPHS11								1	-0.018	-0.15	-0.24	-0.65	0.075	-0.004	0.036	0.021	0.069	-0.012
TPHS12									1	-0.15	-0.24	0.129	-0.106	0.003	0.067	0.074	0.018	0.085
HOHS1										1	-0.021	-0.058	-0.202	-0.114	0.073	-0.017	-0.086	-0.036
HOHS2											1	-0.093	-0.322	-0.077	-0.075	-0.109	-0.029	0.013
HOHS3												1	-0.875	-0.179	0.048	-0.162	-0.053	-0.037
HOHS4													1	0.228	-0.032	0.2	0.085	0.039
AI														1	0.161	0.369	0.075	0.184
YHC															1	0.101	0.111	-0.02
HM																1	0.162	0.186
HR																	1	0.634
logArea																		1

Table 5 Regression model results

Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig	Tolerance	VIF
Constant	-1799.654	945.864		-1.903	0.058		
TPHS1	615.078	128.028	0.287	4.804	0.000	0.813	1.230
logArea	1903.936	479.097	0.221	3.974	0.000	0.940	1.063
TPHS9	4036.758	886.910	0.250	4.551	0.001	0.963	1.039
AI	0.020	0.007	0.162	2.788	0.006	0.859	1.165
TPHS4	-435.796	161.505	-0.163	-2.698	0.008	0.799	1.251
HOHS4	408.096	148.597	0.165	2.746	0.007	0.807	1.239
HOHS2	970.246	356.038	0.157	2.726	0.007	0.877	1.141

The values of the statistical measures of multicollinearity are not suggestive of harmful multicollinearity since the values of VIF are all close to 1 and the tolerance values are all discretely greater than 0.2. In addition, all variables are statistically significantly correlated with the dependent variable, AEE, since a p value < 0.05 has been obtained for all variables.

Overall, all four hypotheses of linear regression are satisfied.

Through the regression analysis the great influence of the primary heating system on the annual energy expenses was highlighted. More specifically, the dummy variables TPHS1 and TPHS9 show a positive correlation with the variable AEE and thus the choice of oil burner or heat accumulator, respectively, as the main heating system increases the total annual energy expenditure of the household. The dummy variable TPHS4 shows a negative correlation with the AEE, indicating that the choice of a firewood stove or pellets as the main heating system reduces the total household energy expenditure. TPHS9 (B = 4036.758), representing heating accumulators as primary heating system, was identified as the most crucial factor shaping AEE and highlighted the thermal fuel pricing as major issue for the households.

Apart from the type of primary heating system, the area of the house, (represented by the logarithmic transformed variable logArea (B = 1903.936)), has a significant impact on AEE of the households. A 10% increase in dwelling area would lead to an increase in expected annual energy expenditure of nearly 80€. Moreover, the influence of the operation hours of the heating systems on the AEE, was confirmed through the dummy variables HOHS1 and HOHS2 also has a positive correlation with the dependent variable and therefore the increase of the hours of the heating operation system lead to an increase in the household's annual energy expenditure.

Less influential compared to the other independent variables appears to be the annual household income (AI) (B = 0.020). The reduced influence of this variable can be attributed to the climatic characteristics of Metsovo since it is a mountainous settlement with an extended heating season, and therefore, heating is primarily a component of survival rather than a choice. In other words, in a cold town such as Metsovo, heating demand creates inelastic expenditures.

The model explains approximately 34% of the variation in the actual annual energy expenditure (R^2adj = 33.8%). Although the value of R^2adj seems low, it is in the range of the previous relevant studies, e.g., Bedir et al. (2013) ($R^2 = 50\%$), Sardianou (2007) ($R^2 = 10\%$), Kotsila and Polychronidou (2021) ($R^2 = 30\%$), Wiesmann et al. (2011) ($R^2 = 33\%$), and Meier and Redhanz (2010) ($R^2 = 30\%$ in general). In general, issues such as energy consumption, which is inherently multidimensional, and composites cannot be approached by models with high R^2.

Multiple regression on the energy monitoring data

Following the methodology applied on the social survey data, a linear regression model was developed for the case of 60 households. The annual energy expenditure of the households was used as the dependent variable, and the independent variables were the type of the main heating system and the hours of operation of the heating system, which were entered into the model in the form of dummy variables,

- the annual income logarithmically transformed,
- the area of the dwelling logarithmically transformed,
- the year of construction of the dwelling,
- the number of members making up the household,
- and the number of rooms in the house,

in addition to the abovementioned independent variables, which were also used in the case of the 300 households, the area of the dwelling heated logarithmically transformed, as well as the average temperature of the dwelling during the day.

By choosing the stepwise method to construct the optimal linear regression model, it was found that the variables expressing the oil boiler as the main heating system, income, log-transformed dwelling area, and hours of operation of the heating system explained approximately 57% of the changes in total annual energy expenditure ($R^2 = 57.33\%$, R^2 (adj) $= 52.3\%$).

The oil boiler as the main heating system appears to be the most important factor, followed by the hours of operation of the heating system and the number of rooms in the dwelling, while all independent variables show a positive correlation with the independent variable. The equation of the linear regression model is as follows:

$$AEE = 693.929 + 1494{,}421 * TPHS1 + 1364.652$$
$$* HOHS3 + 583.315 * HR$$

$$(12)$$

To conclude, the key findings of the multiple regression on the energy monitoring data tend to confirm the findings of the survey data model, underlining the role of type of the primary heating system, the operation hours of the heating system and the area of the house as shaping factors of AEE.

The absence of the variable of households' income results from the inelastic necessity of heating in Metsovo, as it was identified through Eq. 7 and the small influence of income on AEE.

It should be mentioned that, although the explanatory power of the linear regression model increases in the case of 60 households, the results are not considered fully reliable due to the small sample size (Gatsonis & Sampson, 1989). The findings still robust the key points of the social survey.

K-prototype clustering for the 300 households

Except the identification of AEE influencing factors, to identify households with similar energy behavior, several socioeconomic variables were introduced in the K-Prototypes clustering algorithm, listed in Table 6.

The heuristic elbow method is used to determine the optimal number k of classes. In this clustering problem, the elbow method was applied with a range of clusters from 1 to 10 (Fig. 2).

As observed in Fig. 2, the number of 4 clusters fits the data quite well. Since the SSE has decreased significantly, the "elbow" effect appears in the graph, and for several clusters larger than 4, the SSE tends to stabilize. The k-prototypes algorithm was then fitted to the data. A total of 3 iterations of the algorithm were performed until the final clustering of households was obtained. The centers of each variable for each of the four clusters are listed in Table 7.

Once the clustering was calculated, the 10% objective indicator of energy poverty, as well as three subjective indicators of energy poverty, were recorded through the survey, namely, "Winter Comfort: Comfort in terms of warmth inside the dwelling during the winter months", "Moisture: Moisture or mold on floors, walls, and/or ceilings", and "Bill Delay: Inability to pay energy bills on time" for each cluster, and the results are presented in Table 8.

As there is insufficient income data available for households in fourth class, it is not possible to draw conclusions on the extent of energy poverty for households in this class.

It is worth noting that although oil heating systems are predominant in all clusters, if the percentages of biomass-based heating systems (central heating with a firewood boiler, fireplaces, and wood stoves) are summed, they count 52% for the first cluster, 56% for the second cluster, 43% for the third cluster, and 57.8% for the fourth cluster. The percentages for heating oil consumption are 45%, 41%, 37%, and 36%, respectively. Therefore, considering the heating system, central heating with oil prevails, but considering the fuel, biomass is the most often used, confirming the significant dependence of households on it. Both oil and biomass form almost 100% of the fuels used for heating purposes in Metsovo.

A direct relationship between income and energy expenditure can be determined. The third cluster includes households with higher incomes, who spend higher amounts to meet their energy needs, followed by the first cluster with the next highest reported disposable income and, correspondingly, the next highest energy expenditure. Cluster 1 includes medium-income households with medium energy expenditures and older dwellings than Cluster 3. Finally, cluster 2 includes households with distinctly lower incomes, lower energy expenditures, and smaller dwellings than in the two previous classes, while the choice of firewood or
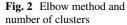
biomass heating systems dominates over the choice of oil heating systems.

Regarding energy poverty, most of the households in clusters 1 (medium income) and 2 (low income) are energy poor according to the 10% objective indicator, while the proportion of energy poor households in cluster 3 (high income) is significantly lower but still high. The percentage of arrears in utility bills remains low, among all clusters, due to the strict regime of energy providers in relation to arrears and interruptions of their services, which highlights a possible squeeze on other household needs to cope financially with energy bills. The fact that almost 1 in 2 households in cluster 2 are not comfortable in terms of heat during the winter months, while at the same time they are classified as energy poor, indicates, first, the compression of their energy needs and the severity of the energy poverty phenomenon, since despite this compression, they are almost all classified as energy poor.

Of particular interest is the fact that even among cluster 3, households which report the highest incomes, the highest energy expenditure, and the lowest rates of energy poverty according to the 10% index, almost 3 out of 10 households report a lack of thermal comfort during the winter months, a higher rate than in cluster 1, where 89% of households are considered energy poor. The common characteristic for all four clusters is the age of the main heating systems in the settlement of Metsovo and therefore their reduced efficiency, as well as the age of the housing stock, with most of the houses being built before 1980 (Greek Thermal Insulation Regulation of Buildings, 1979), when no thermal insulation measures were established, thus leading to higher heat losses. The lack of thermal comfort, which also indicates underperformance of heating systems, are also reflected in households' statements regarding mold and dampness problems. Even though Cluster 3 households reside in comparatively newer dwellings, mold and dampness problems are at the same level as the corresponding percentages for Cluster 2 households. Overall, even in the classes with the highest rates of thermal

Table 6 Socioeconomic variables are used in clustering

Variable	Abbreviation
Type of Residence	TR
Year of House Construction	YHC
Area of the house	AH
Number of rooms of the household	HR
Type of primary heating system	TPHS
Number of household members	HM
Annual Income	AI
Annual Energy Expenses	AEE
Hours of operation of heating system	HOHS

Fig. 2 Elbow method and number of clusters

comfort, there is a significant proportion of households that are not comfortable in terms of heat in the winter, which is particularly important due to the high and inelastic thermal needs in study area.

Interpreting the previous findings, the role of income is highlighted through the clustering process, in contrast with the results of the regression models. Higher income is linked with higher AEE. The higher AEE do not result from more operating hours of the heating system, as stated by Table 5 (HOHS > 10 in all four clusters), but from the chosen type of primary heating system. Households with lower income choose cheaper fuels—biomass, to meet inelastic thermal needs in Metsovo, and thus their AEE is lower. The income and type of primary heating system/fuel reflects the impact of income on households' energy behavior. Biomass heating systems are less energy efficient, with a clear impact on the thermal comfort of the households. Furthermore, households with lower AEE in clusters 1 and 2, are experiencing extended energy poverty considering 10% indicator, despite the operating hours of the heating system.

K-prototype clustering for the 60 households

The k-prototypes algorithm, as described above, was also applied to the data concerning the 60 households where smart meters were installed. For

Table 7 Centers of each variable for each of the four clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
TR	Apartment	Apartment	Apartment	Apartment
YHC	1973	1972	1985	1973
AH	105	90	100	100
HR	4	4	3	4
TPHS	Oil boiler	Oil boiler	Oil boiler	Oil boiler
HM	3	2	4	3
AI	18733	9933	31545	"Do not know/ No answer"
AEE	2953	2540	3260	2810
HOHS	> 10 h	> 10 h	> 10 h	> 10 h

Table 8 The 10% objective and the subjective indicators of energy poverty for each cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Objective Energy Poverty Indicator				
10% Energy Poverty Indicator	90%	96%	41%	-
Subjective Energy Poverty Indicators				
Inability to keep home adequately warm	20%	42%	27%	33%
Dwellings with leakages and damp/mold problems	40%	30%	27%	50%
Inability to pay energy bills in time	18.7%	11%	9%	9%

the clustering of the households in this case, the same variables were used as input variables for the clustering of the 300 households, as well as variables corresponding to the internal temperature of the house, the temperature of the thermostat, and the use of the kitchen, boiler, and washing machine. These factors, as described above, influence electrical and total energy consumption. Applying the elbow method, four classes were selected as the optimal number of classes. The arguments entered in the algorithm are the same as in the case of 300 households, and a total of 3 iterations of the algorithm were performed until the final classes were obtained. After running the algorithm, 13 households were classified in the first and second classes, 19 households in the third cluster, and 10 households in the fourth class. The objective and subjective indicators of energy poverty that were monitored in the case of 300 households were also monitored in this case.

Briefly, the main drivers of household clustering seem to be the floor area of the dwelling, income, and the date of construction of the dwelling. Compared to the clustering obtained for the case of 300 households, it appears that households with more members and higher incomes are more likely to choose oil as a heating fuel and to spend the most on meeting their energy needs. In both cases, these are the smallest in population classes. In relation to the rates of energy poverty based on the 10% objective indicator, it is observed that for the first three classes, approximately 50–60% of households are classified as energy poor in each case (54% for classes 1 and 2 and 58% for cluster 3), while for cluster 4, the corresponding percentage reaches 70%. Additionally, while almost all households in each cluster — 300 households — do not report an inability to pay their energy bills on time, in the case of the 60 households, the corresponding percentages increase significantly, ranging from 10 to 46%. At the same time, however, there is also a decrease in the percentage that feels comfortable in terms of heat inside the dwelling, with this percentage estimated at 46% for clusters 1 and 2, 42% for class 3, and 60% for class 4, while the corresponding range for the case of 300 households is from 80

to 58%. The inverse trend observed between the two subjective indicators in these two cases may indicate a compression of the energy needs of the 60 households. It should be noted, however, that the sample of 60 households cannot be considered statistically representative.

Comparisons with previous studies

Although in most of the surveys examining the household's energy consumption the key factor is the income, in mountain areas such as Metsovo the income is not crucial because heating is not a matter of choice. On the other hand, the type of heating system which is a shaping factor for Annual Energy Expenses in Metsovo is absent from most of the key findings listed in Table 1. The main reason for this differentiation is the fact that people in Metsovo and in others mountain towns are using biomass which is plentiful and easily accessible due to the surrounding forest environment. Only the survey using metered data (Kotsila & Polychronidou, 2021) highlighted the impact of the heating type and the hours of the operation of the heating system as important determinant of electricity consumption. These two factors emerged also in the case study of Metsovo. The surveys with metered data seem to be able to better highlight the complex issue of household energy sector. The importance of the variable "area of the house" has been also highlighted in alignment withal the above-mentioned surveys, but the lack of influence of rest of the socioeconomics, such as age, results from the unique characteristics of the mountainous area which seems to be more crucial.

Conclusion and policy implications

The average electricity consumption of households in the settlement of Metsovo is 3,700 kWh, while the consumption of thermal energy (space heating and DHW) is 30,700 kWh. At the country level, the average electricity consumption was calculated at 3,750 kWh, and the thermal energy consumption was calculated at 10,244 kWh (Hellenic Statistical

Authority, 2013). The households of Metsovo follow the trend of the rest of the country in terms of electricity consumption, while the average thermal energy consumption is almost three times higher than at the country level, confirming the observed increase in thermal needs due to the increase in altitude. More specifically, the total average energy consumption of households in Metsovo is 245% higher than the average total consumption of households in the country. Approximately 90% of the consumption in Metsovo is related to thermal energy, with the corresponding percentage at the country level being 73% (Hellenic Statistical Authority, 2013).

The particularly high thermal needs of households are also highlighted by the fact that 73% of households in Metsovo spend more than 10% on heating, making these households vulnerable to energy poverty. The two linear regression models were fitted to the social survey and the monitored house samples to identify the factors that shape the total annual energy expenditure of households. Both models imply that the use of oil burners as the main heating system leads to increased total annual energy expenditure, and the same is concluded for the hours of operation of the heating system. In addition, the surface area of the house is positively correlated with the increase in household energy costs. For instance, a 10% increase in the floor area of the dwelling would lead to an increase in the expected annual energy expenditure of approximately 80 €. Income seems to be the least influential factor. According to the social survey regression model (this variable proved to be statistically insignificant in the monitored houses regression model), an increase in income by 1,000€ would lead to an increase in annual energy expenditure by 20€. The small or even non-existent participation of income in shaping household energy expenditure highlights the fact that in mountainous areas, such as Metsovo, heating is a necessity and not a choice.

Concerning clustering, the most populous cluster is the one with the lowest average income, while the cluster with the highest average income is the smallest. In addition, the poorest households predominantly opt for main heating systems based on firewood. They feel less comfortable in terms of warmth during the winter and are almost universally classified as energy poor. Interestingly, even in the highest income classes, the proportions of energy-poor households remain relatively high (approximately 40% of households), while the subjective indicators measuring thermal comfort and humidity also show relatively high proportions. The discrepancy between high expenditure and actual satisfaction of needs highlights the high energy requirements in mountainous areas. Combined with the old housing stock, limited energy upgrade measures, and old, obsolete heating systems, they exacerbate the phenomenon of energy poverty in mountainous areas.

Although electricity savings are important and can reduce overall energy costs, the aim of an energy policy for mountain areas should be to reduce the costs of households to meet their thermal needs, considering that these needs amount to 70% of their total energy expenditure. The Greek government should intend to develop a national plan to alleviate energy poverty, including short- and long-term measures, with a particular focus on increasing per capita income and saving energy. Indicatively, tax incentives, such as tax credits and soft loans, apart from the existing allowances, for households that adopt energy efficiency measures could be implemented. In the case of Metsovo, in which approximately 1 out of 2 households use oil central heating systems, a special tariff policy for oil heating fuel should be applied in the context of mitigating the consequences of the global recession that affects the energy sector. In addition, there should be a national target to improve the energy efficiency of heating systems, aiming at improving indoor air quality, reducing household energy expenditure, and reducing greenhouse gas emissions.

Greece, although a predominantly mountainous country, does not have a specific national mountain policy; this sets a major obstacle to effective and sustainable spatial policy. At the same time, a major obstacle is the particularly limited research on the mountainous areas of the country, not only

on the issues of energy needs and energy poverty but also on other characteristics of these areas, such as demographic and economic issues. A better understanding of households' energy behavior through research at many different geographical levels, as well as the establishment of effective and long-term policies, are key pillars in addressing both the energy issues of local communities and the wider development of mountain areas. Alleviation of energy poverty in mountainous areas is a more complex issue than eradication of economic poverty, as it involves the development of major infrastructure and a large amount of funding.

Continuing the effort of this research, it is proposed at the first level to reinstall the smart meters with the aim of the longitudinal evolution of data recording by households so that it is safer to draw conclusions and to be able to integrate the changes of the energy market, which as we have seen in the last weather is characterized by significant fluctuations. Moreover, the same survey could take place in different mountainous areas with the urpose to best fit the model and draw safer conclusions.

In addition, the weather factors that will be measured inside the household, such as humidity and temperature, could be integrated into the final model. This model that will be exported would be particularly useful to compare with corresponding models of mountainous and non-mountainous areas. It will thus be possible to create a model for the entire country, which will have arisen respecting the spatial differences within the country, increasing its validity and imprecision.

Appendix

Table 9 Variable table of social survey (300 questionnaires)

VARIABLE NAME	TYPE	UNIT
Thermal comfort	String	-
Arrears on energy bills	String	-
Rooms of house being heated	String	-
Gender	String	-
Age	Numeric	years
Area of house	Numeric	m^2
Year of house construction	Numeric	ear
Household members	Numeric	-
Type of household	String	-
Total annual income	Numeric	€
Humidity, mold or leaks in the house	String	-
Health problems linked to poor heating	String	-
Reduction of the operating hours of the heating system	String	-
Restriction of other essential needs to meet heating energy needs	String	-
Type of primary heating system	String	-
Type of supplementary heating system(s)	String	-
Type of the building	String	-
Monthly cost of electricity during heating period	Numeric	€
Monthly cost of electricity out of heating period	Numeric	€
Annual cost of fuel for heating purposes	Numeric	€
Heating subsidy	String	-
Ownership of the house	String	-
10% index (EPR)	Numeric	-
Annual energy expenses	Numeric	€
Number of rooms in the house	Numeric	-
Hours of operation of the heating system	Numeric	hours

Table 10 Variable table of households with metered equipment (60 questionnaires)

VARIABLE NAME	TYPE	UNIT
Thermal comfort	String	-
Arrears on energy bills	String	-
Rooms of house being heated	String	-
Gender	String	-
Age	Numeric	years
Area of the house heated	Numeric	m^2
Area of house	Numeric	m^2
Year of house construction	Numeric	years
Number of rooms of the household	Numeric	-
Household members	Numeric	-
Type of household	String	-
Total annual income	Numeric	€
Hours of operation of the heating system	Numeric	hours
Humidity, mold or leaks in the house	String	-
Health problems linked to poor heating	String	-
Type of primary heating system	String	-
Type of supplementary heating system(s)	String	-
Type of the building	String	-
Indoor household's temperature	Numeric	ºC
Heating subsidy	String	-
Ownership of the house	String	-
10% index (EPR)	Numeric	-
Annual cost of electricity	Numeric	€
Annual modeled energy consumption for thermal needs	Numeric	kWh
Annual cost of the model energy consumption for thermal needs	Numeric	€
Annual energy expenses	Numeric	€

Declarations

Conflict of interest All authors declare that they have no conflicts of interest. Tables 9 and 10

References

Akay, Ö., & Yüksel, G. (2018). Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms. *Communications in Statistics-Simulation and Computation, 47*(10), 3031–3041. https://doi.org/10.1080/03610918.2017.1367806

Al-Ghandoor, A. J. J. O., 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

Ali, S. S. S., Razman, M. R., Awang, A., Asyraf, M. R. M., Ishak, M. R., Ilyas, R. A., & Lawrence, R. J. (2021). Critical determinants of household electricity consumption in a rapidly growing city. *Sustainability, 13*(8), 4441. https://doi.org/10.3390/su13084441

Assimakopoulos, V., & Domenikos, H. G. (1991). Consumption preferences structure of Greek households. *Energy Economics, 13*(3), 163–167. https://doi.org/10.1016/0140-9883(91)90017-T

Auffhammer, M., & Mansur, E. T. (2014). Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Economics, 46*, 522–530. https://doi.org/10.1016/j.eneco.2014.04.017

Balaskas, A., Papada, L., Katsoulakos, N., Damigos, D., & Kaliampakos, D. (2021). Energy poverty in the mountainous town of Metsovo Greece. *Journal of Mountain Science, 18*(9), 2240–2254. https://doi.org/10.1007/s11629-020-6436-1

Bedir, M., Hasselaar, E., & Itard, L. (2013). Determinants of electricity consumption in Dutch dwellings. *Energy and Buildings, 58*, 194–207. https://doi.org/10.1016/j.enbuild.2012.10.016

Boardman, B. (1991). Fuel poverty is different. *Policy Studies, 12*(4), 30–41. https://doi.org/10.1080/01442879108423600

Boemi, S. N., Avdimiotis, S., & Papadopoulos, A. M. (2017). Domestic energy deprivation in Greece: A field study. *Energy and Buildings, 144*, 167–174. https://doi.org/10.1016/j.enbuild.2017.03.009

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint* arXiv:1809.03006. *https://doi.org/10.48550/arXiv.1809.03006*

Botetzagias, I., Malesios, C., & Poulou, D. (2014). Electricity curtailment behaviors in Greek households: Different behaviors, different predictors. *Energy Policy, 69*, 415–424. https://doi.org/10.1016/j.enpol.2014.03.005

Büyükalaca, O., Bulut, H., & Yılmaz, T. (2001). Analysis of variable-base heating and cooling degree-days for Turkey. *Applied Energy, 69*(4), 269–283. https://doi.org/10.1016/S0306-2619(01)00017-4

Chatzikonstantinou, E., Katsoulakos, N., & Vatavali, F. (2022). Housing and energy consumption in Greece. Households' experiences and practices in the context of the energy crisis. *In IOP Conference Series: Earth and Environmental Science, 1123*(1), 012043. https://doi.org/10.1088/1755-1315/1123/1/012043

Chen, Y., Guo, M., Chen, Z., Chen, Z., & Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. *Energy Reports, 8*, 2656–2671. https://doi.org/10.1016/j.egyr.2022.01.162

Dent, I., Craig, T., Aickelin, U., & Rodden, T. (2014). Variability of behaviour in electricity load profile clustering; Who does things at the same time each day?. In Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16–20, 2014. Proceedings 14 (pp. 70–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-08976-8_6

Dorman, K. S., & Maitra, R. (2022). An efficient k-modes algorithm for clustering categorical datasets. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 15*(1), 83–97. https://doi.org/10.1002/sam.11546

Dubey, A. K., Kumar, A., García-Díaz, V., Sharma, A. K., & Kanhaiya, K. (2021). Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments, 47*, 101474. https://doi.org/10.1016/j.seta.2021.101474

EIA,International Energy Outlook 2021 U.S. Energy Information Administration, Washington D.C, USA (2021). (Accessed on 22/06/2023)

Eren, B. M., Taspinar, N., & Gokmenoglu, K. K. (2019). The impact of financial development and economic growth on renewable energy consumption: Empirical analysis of India. *Science of the Total Environment, 663*, 189–197. https://doi.org/10.1016/j.scitotenv.2019.01.323

Eurostat (2020) EU statistics on income and living conditions (EU-SILC) methodology – economic strain. <http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_statistics_on_income_and_living_conditions_(EU-SILC)_methodology_-_economic_strain#Main_tables> (Accessed on 25/05/2023)

Eurostat, 2020, Inability to keep home adequately warm - EU-SILC survey, https://ec.europa.eu/eurostat/databrowser/view/ilc_mdes01/default/table?lang=en,(accessed on:23/12/2020)

Eurostat (2023) Energy consumption in households. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households (Accessed on 25/06/2023)

Evans, W. (2010). Durbin-Watson significance tables. University of Notre Dame

Filippini, M., & Pachauri, S. (2004). Elasticities of electricity demand in urban Indian households. *Energy Policy, 32*(3), 429–436. https://doi.org/10.1016/S0301-4215(02)00314-2

Funnell, D., & Parish, R. (2005). *Mountain environments and communities*. Routledge.

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: exact power and sample size calculations. *Psychological Bulletin, 106*(3), 516. https://doi.org/10.1037/0033-2909.106.3.516

Gohari, A., Matori, N., Yusof, K. W., Toloue, I., & Myint, K. C. (2018). Effects of the fuel price increase on the operating cost of freight transport vehicles. *In E3S Web of Conferences, 34*, 01022. https://doi.org/10.1051/e3sconf/20183401022

González-Aguilera, D., Lagueela, S., Rodríguez-Gonzálvez, P., & Hernández-López, D. (2013). Image-based thermographic modeling for assessing energy efficiency of buildings façades. *Energy and Buildings, 65*, 29–36. https://doi.org/10.1016/j.enbuild.2013.05.040

Greek Thermal Insulation Regulation of Buildings, 362/Δ, 4.7.1979. (Accessed on 25/05/2023)

Guenoukpati, A., Salami, A. A., Birregah, B., & Bakpo, Y. A (2021) A Novel Approach for Electric Load Prediction Using Convolutional Lstms Networks with Sorted Wavelet Transform Coefficient. *Available at SSRN 4775353*. https://doi.org/10.2139/ssrn.4775353

Hellenic Statistical Authority, 2011. Cencus of population – residences 2011. http://www.statistics.gr/el/2011-census-pop-hous (Accessed on 22/04/2023)

Hellenic Statistical Authority. (2013). *Press Release, Research of energy consumption in households, 2011-2012*. https://www.statistics.gr/documents/20181/e74d6134-8c02-404e-a02baa6d959219e3. Accessed 22 Jun 2023.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery, 2*(3), 283–304.

Huang, W. H. (2015). The determinants of household electricity consumption in Taiwan: Evidence from quantile regression. *Energy, 87*, 120–133. https://doi.org/10.1016/j.energy.2015.04.101

Kandananond, K. (2011). Forecasting electricity demand in Thailand with an artificial neural network approach. *Energies, 4*(8), 1246–1257. https://doi.org/10.3390/en4081246

Karani, I., Papada, L., & Kaliampakos, D. (2022). Energy poverty signs in mountainous Greek areas: The case of Agrafa. *International Journal of Sustainable Energy, 41*(10), 1408–1433. https://doi.org/10.1080/14786451.2022.2055029

Katsoulakos, N. M., & Kaliampakos, D. C. (2016). Mountainous areas and decentralized energy planning: Insights from Greece. *Energy Policy, 91*, 174–188. https://doi.org/10.1016/j.enpol.2016.01.007

Kostakis, I. (2020). Socio-demographic determinants of household electricity consumption: Evidence from Greece using quantile regression analysis. *Current Research in Environmental Sustainability, 1*, 23–30. https://doi.org/10.1016/j.crsust.2020.04.001

Kotsila, D., & Polychronidou, P. (2021). Determinants of household electricity consumption in Greece: A statistical analysis. *Journal of Innovation and Entrepreneurship, 10*, 19. https://doi.org/10.1186/s13731-021-00161-9

Kugiumtzis, D. (1999). Test your surrogate data before you test for nonlinearity. *Physical Review E, 60*(3), 2808. https://doi.org/10.1103/PhysRevE.60.2808

Lenzen, M., Wier, M., Cohen, C., Hayami, H., Pachauri, S., & Schaeffer, R. (2006). A comparative multivariate analysis of household energy requirements in Australia, Brazil, India and Japan. *Energy, 31*(2–3), 181–207. https://doi.org/10.1016/j.energy.2005.01.009

Lepot, M., Aubin, J. B., & Clemens, F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water, 9*(10), 796. https://doi.org/10.3390/w9100796

Li, X., Smyth, R., Xin, G., & Yao, Y. (2023). Warmer temperatures and energy poverty: Evidence from Chinese households. *Energy Economics, 120*, 106575. https://doi.org/10.2139/ssrn.4166337

Ma, C., Zhang, Y., & Zhao, W. (2021). Influence of latitude on raw material consumption by biomass combined heat and power plants: Energy conservation study of 50 cities and counties in the cold region of China. *Journal of Cleaner Production, 278*, 123796. https://doi.org/10.1016/j.jclepro.2020.123796

MacKinnon, J. G. (1992). Model specification tests and artificial regressions. *Journal of Economic Literature, 30*(1), 102–146.

Matzarakis, A. & Balafoutis, Ch. (2002). Geographical distribution of Heating Degree Days in Greece for Use in Energy Calculations. *6th Pan-hellenic Conference of Meteorology, Climatology and Atmospheric Physics* (pp. 156–163). Ioannina, B.D. Katsoulis. (in Greek) https://doi.org/10.1002/joc.1107

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

Meier, H., & Rehdanz, K. (2010). Determinants of residential space heating expenditures in Great Britain. *Energy Economics, 32*(5), 949–959. https://doi.org/10.1016/j.eneco.2009.11.008

Moschou, Ch. (2011). *Calculation of energy loads for buildings' energy design using mathematical programming*. BSc Thesis. Athens, National Technical University of Athens, School of Chemical Engineering. (In Greek)

Nakagami, H., Murakoshi, C., & Iwafune, Y. (2008). International comparison of household energy consumption and its indicator. *Proceedings of the 2008 ACEEE Summer Study on Energy Efficiency in Buildings, 8*, 214–224.

Nugaliyadde, A., Somaratne, U., & Wong, K. W. (2019). Predicting electricity consumption using deep recurrent neural networks. *arXiv preprint* arXiv:1909.08182. https://doi.org/10.48550/arXiv.1909.08182

Ota, T., Kakinaka, M., & Kotani, K. (2018). Demographic effects on residential electricity and city gas consumption in the aging society of Japan. *Energy Policy, 115*, 503–513. https://doi.org/10.1016/j.enpol.2018.01.016

Papada, L., & Kaliampakos, D. (2016). Developing the energy profile of mountainous areas. *Energy, 107*, 205–214. https://doi.org/10.1016/j.energy.2016.04.011

Papada, L., & Kaliampakos, D. (2016). Measuring energy poverty in Greece. *Energy Policy, 94*, 157–165. https://doi.org/10.1016/j.enpol.2016.04.004

Peklaris V. (2010). The crisis creates woodcutters [in Greek]. Agelioforos Newspaper 28 December 2010. http://www.agelioforos.gr/default.asp?pid=7&ct=1&artid=73039. Accessed 10 Dec 2023.

Pérez, N. F., Ferré, J., & Boqué, R. (2009). Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemometrics and Intelligent Laboratory Systems, 95*(2), 122–128. https://doi.org/10.1016/j.chemolab.2008.09.005

Petri, Y., & Caldeira, K. (2015). Impacts of global warming on residential heating and cooling degree-days in the United States. *Scientific Reports, 5*(1), 12427. https://doi.org/10.1038/srep12427

Quan, S. J., & Li, C. (2021). Urban form and building energy use: A systematic review of measures, mechanisms, and methodologies. *Renewable and Sustainable Energy Reviews, 139*, 110662. https://doi.org/10.1016/j.rser.2020.110662

Roberts, D., Vera-Toscano, E., & Phimister, E. (2015). Fuel poverty in the UK: Is there a difference between rural and urban areas? *Energy Policy, 87*, 216–223. https://doi.org/10.1016/j.enpol.2015.08.034

Santamouris, M., Kapsis, K., Korres, D., Livada, I., Pavlou, C., & Assimakopoulos, M. N. (2007). On the relation between the energy and social characteristics of the residential sector. *Energy and Buildings, 39*(8), 893–905. https://doi.org/10.1016/j.enbuild.2006.11.001

Santamouris, M., Paravantis, J. A., Founda, D., Kolokotsa, D., Michalakakou, P., Papadopoulos, A. M., ... & Servou, E. (2013). Financial crisis and energy consumption: A household survey in Greece. *Energy and Buildings, 65*, 477–487. https://doi.org/10.1016/j.enbuild.2013.06.024

Sardianou, E. (2007). Estimating energy conservation patterns of Greek households. *Energy Policy, 35*(7), 3778–3791. https://doi.org/10.1016/j.enpol.2007.01.020

Sardianou, E. (2008). Estimating space heating determinants: An analysis of Greek households. *Energy and Buildings, 40*(6), 1084–1093. https://doi.org/10.1016/j.enbuild.2007.10.003

Schuler, M., Stucki, E., Roque, O., & Perlik, M. (2004). Mountain Areas in Europe: Analysis of mountain areas in EU member states, acceding and other European countries

Su, Y. W. (2019). Residential electricity demand in Taiwan: Consumption behavior and rebound effect. *Energy Policy, 124*, 36–45. https://doi.org/10.1016/j.enpol.2018.09.009

United nations, Sustainable development goals, Available at https://unric.org/el/17-%CF%83%CF%84%CE%BF%CF%87%CE%BF%CE%B9-%CE%B2%CE%B9%CF%89%CF%83%CE%B9%CE%BC%CE%B7%CF%83-%CE%B1%

CE%BD%CE%B1%CF%80%CF%84%CF%85%CE%BE%CE%B7%CF%83/ (Accessed: 24/4/2023)

Vogiatzi, C., Gemenetzi, G., Massou, L., Poulopoulos, S., Papaefthimiou, S., & Zervas, E. (2018). Energy use and saving in residential sector and occupant behavior: A case study in Athens. *Energy and Buildings, 181*, 1–9. https://doi.org/10.1016/j.enbuild.2018.09.039

Wang, Z., Bui, Q., Zhang, B., Nawarathna, C. L. K., & Mombeuil, C. (2021). The nexus between renewable energy consumption and human development in BRICS countries: The moderating role of public debt. *Renewable Energy, 165*, 381–390. https://doi.org/10.1016/j.renene.2020.10.144

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

Ye, Y., Koch, S. F., & Zhang, J. (2018). Determinants of household electricity consumption in South Africa. *Energy economics, 75*, 120–133. https://doi.org/10.1016/j.eneco.2018.08.005

Zhang, Y., Liu, Q., & Song, L. (2018). Sentence-state LSTM for text representation. *arXiv preprint* arXiv:1805.02474. *https://doi.org/10.48550/arXiv.1805.02474*

Zhao, J., Thinh, N. X., & Li, C. (2017). Investigation of the impacts of urban land use patterns on energy consumption in China: a case study of 20 provincial capital cities. *Sustainability, 9*(8), 1383. https://doi.org/10.3390/su9081383

Zhou, S., & Teng, F. (2013). Estimation of urban residential electricity demand in China using household survey data. *Energy Policy, 61*, 394–402. https://doi.org/10.1016/j.enpol.2013.06.092