ENCLUDE
Energy Citizens for Inclusive Decarbonization

D4.1 – Report on qualified clustering input attributes

WP4 – Identification of citizens' clusters for decarbonization

10/06/2022

Sobhan Naderian, Anastasia Ioannou, Gioia Falcone

University of Glasgow

Version: 1.0

www.encludeproject.eu

Sobhan Naderian	University of Glasgow
sobhan.naderian@glasgow.ac.uk	https://www.gla.ac.uk/schools/engineering/staff/sobhannaderian/

Anastasia Ioannou	University of Glasgow
Anastasia.Ioannou@glasgow.ac.uk	https://www.gla.ac.uk/schools/engineering/staff/anastasiaioannou/

Gioia Falcone	University of Glasgow
Gioia.Falcone@glasgow.ac.uk	https://www.gla.ac.uk/schools/engineering/staff/gioiafalcone/

Disclaimer

Copyright Message

The ENCLUDE project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101022791

ENCLUDE PROJECT & DELIVERABLE PROFILE

Project Acronym and Full Name:	ENCLUDE - Energy Citizens for Inclusive Decarbonization
Grant Agreement No.:	101022791
Programme:	H2020-EU.3.3.6. - Robust Decision Making and Public Engagement
Topic:	LC-SC3-CC-1-2018-2019-2020 – Social Sciences and Humanities (SSH) aspects of the Clean-Energy Transition
Funding Type:	RIA - Research and Innovation Action
Deliverable:	D4.1 – Report on qualified clustering input attributes
Work Package:	WP4 – Identification of citizens' clusters for decarbonization
Deliverable Due Date:	Project month M9 (28/02/2022)
Actual Date of Submission:	10/06/2022
Dissemination Level:	Public (pending approval by the European Commission)
Lead Beneficiary:	UOG
Responsible Author:	Sobhan Naderian (UOG)
Contributor(s):	Sobhan Naderian (UOG), Anastasia Ioannou (UOG), Gioia Falcone (UOG)
Internal Reviewers:	Ilias Tsopelas (UPRC), Stephan Schwarzinger (Joanneum Research)

Preface

The overall vision of ENCLUDE is to help the EU to fulfil its promise of a just and inclusive decarbonization pathway through sharing and co-creating new knowledge and practices that maximize the number and diversity of citizens who are willing and able to contribute to the energy transition. Motivated by achieving an equitable and sustainable future and the fulfilment of individual potential, ENCLUDE will contribute to the upcoming transformation of energy use by: (1) Assembling, aligning, and adapting disparate energy citizenship concepts for diverse communities of citizens and for different scales of policy making, lowering the barrier for action. (2) Operationalizing the energy citizenship concept at all scales of policy making for decarbonization. (3) Catalyzing a chain reaction of decarbonization actions across the EU.

1. Changes with respect to the DoA

This deliverable was originally due at M9. However, the planned post-doctoral research associate post was not in place until the end of November 2021, and only at 0.5 FTE until March 2022.

2. Dissemination and uptake

This deliverable constitutes the basis for further research within the project, aiming to convert findings from WPs 2 and 3, and to enable the statistical approaches and modelling instruments of WP4 and WP5. It will also contribute to training modules for the ENCLUDE Academy for Energy Citizen Leadership, by offering a critical review of the state-of-the-art in relevant attributes for clustering energy behavior (and the associated impact) at individual and collective levels. This review will therefore also support future investigations outside of the project, by researchers interested in the topic of qualified attributes for clustering energy-related actions.

3. Short Summary of results

This report captures the results of an extensive literature review of studies that cluster citizens in terms of their energy/environmental behaviors. The report maps the factors that might be used in the literature to create clusters for decarbonization under the work of WP4. Outputs of the review are presented at two levels, according to whom the data clusters refer to, namely individual and collective.

- At an *individual* level, major variables for clustering energy behaviors were categorized as socio-economic and demographic, psychological, energy consumption/environmental patterns across different areas of life (housing, transport, etc.), and other contextual variables.
- At a *collective* level, major variables were categorized as socio-economic and demographic, energy infrastructure variables, energy consumption profiles, environmental performance, and other contextual factors.

Establishing clusters of citizens based on their individual attributes leads to their distinct grouping, which can provide insights regarding their energy behavior and lifestyle and can assist the development of policies targeting specific groups of citizens. However, when considering spatially targeted policies, “aggregated level” data (e.g., at neighborhood or even at building level) might be more appropriate than household level data (Reyna *et al.*, 2016). In the context of this study, clustering data for a single person or household fall under the individual level (although the house could be occupied by one or more persons), while clustering data collected at any scale bigger than household is considered collective. Finally, a potentially insightful way to cluster citizens and groups of citizens may be based on their needs/priorities (affordability, access to energy, sustainability, efficiency).

4. Evidence of accomplishment

This report serves as evidence of accomplishment.

LIST OF PARTICIPANTS

	Participant Name	Short Name	Country	Logo
1	TECHNISCHE UNIVERSITEIT DELFT	TUD	Netherlands	TUDelft
2	UNIVERSITY OF PIRAEUS RESEARCH CENTER	UPRC	Greece	TEESlab
3	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETHZ	Switzerland	ETH zürich
4	UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK.	UCC	Ireland	UCC
5	UNIVERSITY OF GLASGOW	UOG	United Kingdom	University of Glasgow
6	JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH	JR	Austria	JOANNEUM RESEARCH LIFE
7	TH!NK E	THNK	Belgium	Th!nk E
8	UNIVERSITEIT UTRECHT	UU	Netherlands	Utrecht University
9	GREEN PARTNERS SRL	GP	Romania	Green Partners
10	ZDRUZENIE CENTAR ZA ISTRAZUVANJE I INFORMIRANJE ZA ZIVOTNA SREDINA EKO-SVEST	Eko-svest	North Macedonia	Еко-свест
11	MISSIONS PUBLIQUES	MP	France	MISSIONS PUBLIQUES
12	HOLISTIC IKE	HOLISTIC	Greece	HOLISTIC
13	UNIVERSITY OF VICTORIA	Uvic	Canada	University of Victoria

Executive Summary

This deliverable aims to identify key variables relevant to the emergence and fostering of energy citizenship that will accordingly enable the strategic clustering of citizens for decarbonization.

To this end, an in-depth literature review of articles and EU-funded project reports (e.g., ECHOES and ENTRUST) has been carried out to map the factors that have been used to create clusters of citizens within the wider research area of energy decarbonization and create a database of attributes and other required variables such as the size of the samples, the level of analysis, the number and types of clusters, along with the methods of clustering.

A large portion of literature has used big data analytics to establish clusters of citizens based on their individual or household energy consumption patterns and lifestyles across different areas of life (housing, mobility, consumption, diet, etc.). Other researchers have focused on the building level (rather than the individual or household), collecting data about outdoor/indoor environment variables, as well as the aggregated energy consumption (heating and electricity) data and discover common trends and hidden knowledge from large datasets (more details in Section 3). To this end, outputs of the review are presented under two levels, according to whom the data clusters refer to, namely individual or collective level. In the context of this study, clustering data for a single person or household fall under the individual level (although the house could be occupied by one or more persons), while clustering data collected at any scale bigger than household is considered collective.

On the one hand, establishing clusters of citizens based on their individual clustering attributes leads to their distinct grouping based on their energy behavior/environmental lifestyles and assists the development of policies targeting specific groups of citizens. On the other hand, certain variables cannot be accurately estimated unless contextual factors are taken into consideration. For example, an occupant of a flat often does not have the power to change the central heating system of the multi-apartment building. In such cases, data on electricity/heating consumption, along with data on the carbon emissions profile may be preferable to be collected at an aggregated level (e.g., building), as this would result to a more representative set of clusters, enabling the development of spatially targeted policies for energy decarbonization.

Figure 1 illustrates the main categories of clustering variables that have been identified through the literature review process at the individual and collective level as follows:

- Socio-economic & demographic variables: Age, income, education, employment, gender, income, number of family members
- Geographic variables: Urban, rural environment
- Psychological variables: Personal norms, preferences, values, lifegoals
- Energy/environmental lifestyle and infrastructure: Energy usage data for different types such as electricity and heat at both individual and collective levels across different areas of life, including electricity load, heat/electricity load. Environmental lifestyle, including CO_2 emissions and carbon impact at a collective level
- Other factors, such as management of finances, social networks and community, and experience of local political networks

Existing studies have focused on establishing distinct clusters of citizens based on their energy behavior, their environmental lifestyle, as well as their motivation to become early adopters of innovate technologies. Other researchers focus on deriving load profiles of buildings to improve accuracy of load forecasting (electricity/heating/cooling) and/or to support the design of energy efficiency initiatives, policies, and roadmaps for long-term resource/energy system planning (Figure 1).

An interesting and potentially insightful way to cluster citizens and groups of citizens is based on their needs/priorities (affordability, access to energy, sustainability, efficiency). Especially in neighborhoods/areas/countries suffering from lack of access to clean and affordable energy services (heating/cooling, electricity), factors such as the low income, the high energy costs, and the house energy performance are becoming increasingly important. In such cases, clustering of citizens based on their needs/priorities will potentially lead to the creation of quite different groupings and enable policy makers to emphasize on inclusive policies that can further promote decarbonization across different segments of the population.

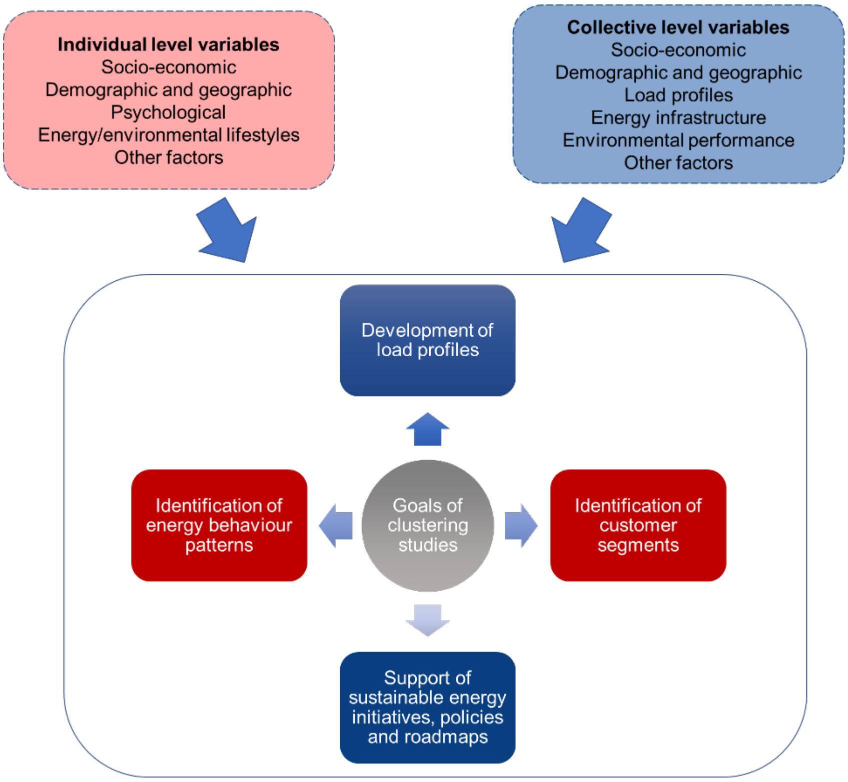

Figure 1 Common types of clusters and key clustering variables at both individual and collective level

Contents

Figures

Tables

1 Introduction

1.1 Background

The overall vision of the Horizon 2020 (H2020) project Energy Citizens for Inclusive Decarbonization (ENCLUDE) is to help the EU to fulfil its promise of a just and inclusive decarbonization pathway through sharing and co-creating new knowledge and practices that maximize the number and diversity of citizens who are willing and able to contribute to the energy transition.

In the transformation of the energy system, the role of citizens is becoming increasingly important through their engagement, involvement, and shaping of the future energy landscape. This important role of citizens for energy consumption is reflected in the term "energy citizen". ENCLUDE aims to share new knowledge and motivate the broadest possible population to contribute to the energy transition. Citizens consume significant amounts of energy each year and their total energy consumption is increasing because of the population growth. On the other hand, new appliances are more efficient; for example, the electric vehicles reduce the amount of energy consumed by conventional fossil-fuel based vehicles. In this regard, they play a key role in energy consumption and CO_2 emissions, too. It is crucial to create a framework to measure citizens' CO_2 impact and form a pathway to decrease their influence on carbon emissions. To measure the shift of citizens' energy behavior, the energy usage and CO_2 emissions dataset needs to be considered; accordingly, an efficient clustering method is required to categorize them and define an effective policy to facilitate behavior change.

1.2 Context

Within ENCLUDE, the aim of Work Package (WP) 4 is to identify groupings of citizens (clusters for decarbonization) at different scales of analysis, so that they can be more effectively mobilized by policies to accelerate energy system decarbonization.

Individual clusters for decarbonization are not necessarily groups of citizens with common demographic characteristics; rather, they may involve demographically diverse groups sharing common characteristics of energy behavior, including readiness to embrace energy citizenship actions, which in turn result in each carbon footprint. Such groupings can inform policy makers of the citizens' profiles that are expected to be more responsive to energy citizenship policy initiatives, as these should be targeted due to their decarbonization potential.

The first task of WP4 is to qualify, in collaboration with WP3, key clustering variables relevant to the emergence and fostering of energy citizenship, to enable effective data extraction from the project's case studies and the successful development of clustering models (e.g., through application of Machine Learning (ML) and deep learning methods).

1.3 Structure of the report

This report summarizes the findings on qualified clustering input attributes, based on results from previous H2020 projects, such as ECHOES (Schwarzinger *et al.*, 2019; Bird *et al.*, 2020) and ENTRUST (Axon *et al.*, 2018), along with results from extended literature to identify energy decarbonization-related clustering attributes (Tobin De Fuentes & Dunphy, 2018).

The report is structured as follows: First, the categorization of clustering studies as presented in this report is introduced in Section 2. Next, the main findings (clustering attributes, clustering approaches and key outputs) from the review of clustering studies are presented at individual and collective levels in Section 3. In Section 4, key attributes for efficient individuals and collective clustering are presented and discussed. Finally, summary and conclusions are presented in Section 5.

2 Classification of clustering studies

2.1 Introduction

Increasing concerns about global warming and climate change have placed "citizens at the heart of energy transitions" (COMETS *et al.*, 2022; Blasch *et al.*, 2022) As such, the role of citizens' energy behavior and shift to more sustainable paradigms towards achieving the net zero emissions target have drawn researchers' attention around the world. As already mentioned, it is crucial to categorize people into different clusters to create proper policies and roadmaps for each group. Several studies are conducted, and methods are developed for clustering at different levels to identify separable patterns, lifestyles, energy profiles and energy behaviors, especially during the last two decades.

Different levels of clustering have been considered in literature, varying from individual to country level. In this report, clustering variables are presented on two analytical levels: *individual* and *collective*. This classification of clustering studies is chosen for the sake of simplicity, as well as to identify how clustering variables differ among different scales of analysis and to increase the understanding about the role of collective factors in shifting to more sustainable lifestyles. In the context of this study, clustering data for a single person or household fall under the individual level (although the house could be occupied by one or more persons), while clustering data collected at any scale bigger than household is considered collective. The categorization of clustering studies is presented in Figure 2.

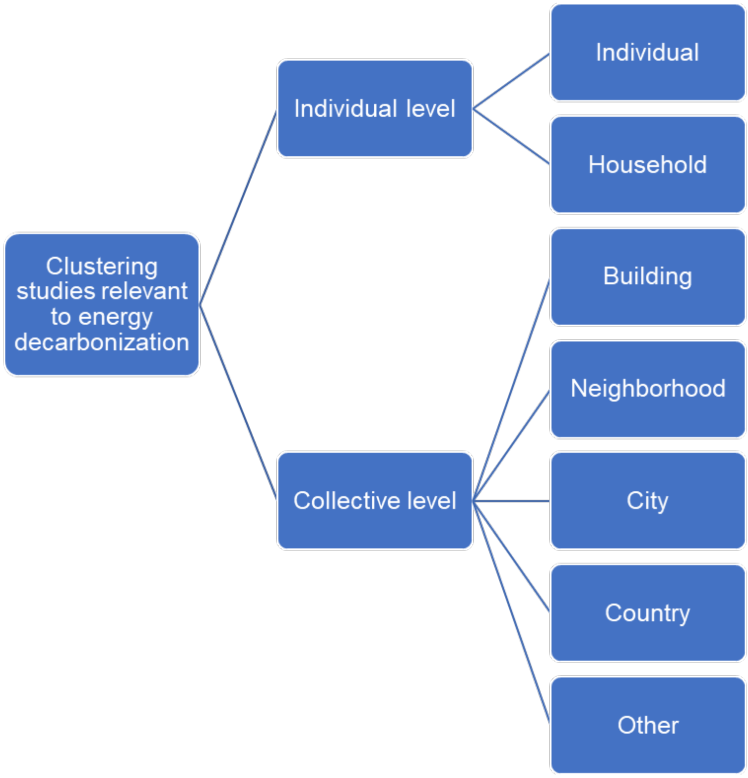

Figure 2 Categorization of clustering studies at individual and collective level

2.2 Individual level

Clustering energy behaviors at the individual level has been carried out in several research studies, considering different variables for clustering and pattern identification within a dataset. In this regard, studies that use data collected at individual level (e.g., individual energy usage, sociodemographic, psychological factors) as input for clustering are gathered under this category. The data may have been gathered via surveys, or energy usage measurements. There are, however, cases when data at the individual level are not directly known, e.g., when energy usage is only measurable on a per-household basis. To simplify this uncertainty in this deliverable, studies that used surveys data for individuals who live in households are considered as individual level (Seebauer *et al.*, 2017). Most studies focus on energy usage of individuals and use the data to develop energy consumption (electricity/heating/cooling) profiles and categorize people into different clusters (more details in section 3.2), with the aim to support policy decision making towards energy decarbonization pathways.

Some other studies use socio-demographics and psychological variables to categorize people into separable clusters. Common examples are age, gender, income, location of residency and other socio-demographic and geographic variables. Motivation, willingness to change and mental barriers are considered as psychological variables besides socio-demographic variables to create a vision for policy makers and help create better roadmaps for behavior change among individuals.

2.3 Collective level

Several studies have been carried out to evaluate the effects of environmental and net zero policies on groups (rather than individuals), such as apartments, cities, and countries. Different clustering methods are applied to categorize collective entities into distinct clusters and to consider their large-scale attributes such as energy behavior and CO_2 emissions.

Different types of energy usage in buildings, such as electricity consumption, district heat demand, and electricity demand profiles, are considered to find meaningful patterns in data using clustering methods. To investigate the energy behavior at a larger scale than at individual level, different variables such as energy demand, renewable energy installations, and CO_2 emissions are taken into account. Variables at a collective level could create a clearer vision for decision makers and help form roadmaps for net zero carbon. Other important variables that have been evaluated for energy behavior changes are greenhouse gas emissions and clean energy usage. The amount of CO_2 emissions is explored at varying levels, from buildings to country level, and collected data are used to recognize different clusters.

In the next Section, published articles and studies are reviewed based on the classification proposed in this Section.

3 Review of clustering attributes

3.1 Introduction

In this Section, references and articles relating to citizens' behavior clustering are reviewed and evaluated at two different levels of analysis: individual and collective. During the review process, over 140 articles were identified using a set of keywords, including "citizens", "energy behavior", "emissions", "decarbonization" and "clustering". 54 articles were found to be directly relevant to the scope of ENCLUDE and they are reviewed in this section, taking into consideration quantitative variables, qualitative variables, scale of the datasets, and clustering methods.

In what follows, the individual level of clustering is addressed first, with an assessment of the studies that have considered different thematic types of variables, such as energy behavior, socio-demographic and psychological factors. The collective level is then discussed, and the clustering variables identified are thematically grouped under energy consumption, clean energy, and CO_2 emissions for subsequent evaluation.

3.2 Clustering attributes at individual level

3.2.1 Socio-economic, demographic, and geographic factors

Socio-demographic factors such as age, gender, education, and income have been widely as clustering variables in individual level clustering studies (Schwarzinger et al., 2019; Diao et al., 2017).

Boucher *et al.* (2018) investigate the correlation between education and income with the energy awareness, using energy audits as a proxy. An ordinary least squares method has been applied to 1670 postcodes in New York. The results show that audits are ubiquitous for individuals with a prestigious education and mediocre income as well as individuals with lower income inhabiting zip code areas with a higher average age (more than 45 years-old) in big cities. This study emphasizes on spatial and social stratification of pro-environmental behaviors. Furthermore, living in the household for a long time could increase the chance of effectiveness of behavior change schemes.

The willingness to adopt innovative and potentially more sustainable technologies has also been investigated in literature. Individuals' criteria and behavior attributes to adopt Fuel Cell Electric Vehicles (FCEV) have been studied in (Moon *et al.*, 2021) to identify early adopters in South Korea. Consumers are clustered according to criteria that affect their choice when purchasing a vehicle (fuel type, max price, brand, etc.). Accordingly, socio-economic and demographic characteristics of consumers belonging to each cluster are analyzed, including age, income, number of family members, number of children in a family, driving distance, value on car ownership, perception regarding the uncertainty of low emissions vehicle (LEV), and necessity of LEV). Citizens have been classified into 6 clusters using the k-means clustering method. It was found that 44.9% of consumers consider FCEV as potential choice, with the 'innovative and luxurious' consumer group showing the highest likelihood of buying FCEV. Interestingly, this study considered sustainable technology initiative and behavior changes simultaneously, hence highlighting the potential for technology to influence behavior.

3.2.2 Psychological factors

The relation between households' characteristics and residential electricity usage patterns has been discussed by (Fang *et al.* 2021), using both socio-demographic and psychological attributes. Socio-demographic variables like age, income, number of adults in each house, and psychological variables like lack of convenience were considered. The proposed method used a dataset of 3117 consumers in China and grouped them into 7 groups of daily energy. Similar to other psychological studies on climate-related behavior, this work takes both sociodemographic and psychological attributes into consideration (Moon *et al.*, 2021; Mi *et al.* 2021).

The effect of psychological and socio-economic variables on pro-environmental behavior and energy usage has been investigated by (Moser & Kleinhückelkotten, 2018). In this study, two approaches for pro-environmental behavior are explored: intent-oriented and impact-oriented. Research that focused on intent emphasized on motivation, while impact-based approaches focused on socio-economic variables. To this extent, data from a survey of 1012 individuals from Germany were gathered, and the impact of variables such as age, income, gender, education, home owning and pro-environmental behavior (intent-related attribute) for each individual, was measured. It was found that focusing on impact-related variables have a significant effect on pro-environmental behaviors, which could represent a vital insight for policy makers.

Behavior changes in adolescents by improving awareness about energy savings and the impact of greenhouse gas emissions has been studied by (Cornelius *et al.* 2014). In this research, the principles of behavior change have been evaluated, including self-efficacy, key learning processes, and motivation factors, to improve knowledge, awareness, and attitude amongst 165 adolescent students in the USA. Students were divided into two groups: control and treatment. It was found that the treatment group improved its energy savings behavior significantly (by means of switching off appliances) as opposed to the control group. Their theory-based intervention showed significant effects on energy saving behavior.

Psychological variables such as motivation and willingness to change are qualitative factors, and they must be considered for behavior change analysis. Measurement of psychological attributes after applying behavior change initiatives is difficult due to the complex nature of human response to a contextual change. It is difficult to create guidelines and roadmaps using psychological variables, but it is possible to cluster individuals based on the similarity in their behaviors. Overall, few studies have considered psychological variables and it is a research area that is still developing to increase the understanding of its implications on energy behavior change.

3.2.3 Energy consumption

Domestic UK consumers have been clustered into 2 groups based on their electricity consumption peak time and off-peak time by (Yao *et al.*, 2018). The study suggests that it is difficult to derive a behavior change policy by examining daily energy usage; rather, hourly electricity usage would lead to more accurate energy behavior patterns. A novel method to recognize usage patterns of AC for residential buildings and development of a new data mining framework for data clustering is proposed by (An *et al.* 2018). AC usage data from 301 apartments in Zhengzhou, China have been clustered using the fuzzy k-means method, leading to the development of 4 clusters. This study is applied to three-bedroom apartments located in a specific region in China with cold climate to create more efficient and accurate energy simulations. The consideration of data on AC usage was suggested to approximate well consumers' energy habits and can assist policy makers to create targeted policy schemes.

The residency owners' mental model for the identification of appliance similarity has been proposed by (Gabe-Thomas *et al.*, 2016) to categorize energy consuming appliances based on their energy use pattern. In this study, the daily energy behavior was associated with the types of appliances used. The dataset of energy usage from Exeter in the UK has been successfully clustered into three distinct appliance groups (kitchen, entertainment, and other appliances) to create a vision for future policies and decisions.

Dynamic and multiscale relationships of hourly electricity consumption for 13 residencies in Florida is proposed by (Knowles *et al.*, 2018). In this research, Ward's minimum variance method has been used for clustering data into 5 unique consumer groups. The main advantage of this study is its higher resolution of hourly energy data, leading to more accurate pattern recognition for creating efficient and accurate behavior change policies. Also, considering household data could help more effective policies for carbon emissions reduction. Supervised and unsupervised clustering for categorization of consumers' energy usage and behavior have been developed by (Jiang *et al.*, 2018). In this study, fuzzy consumer categorization and k-means clustering have been developed to categorize consumers into 13 clusters. Real world results confirmed that the proposed method could improve accuracy of clustering and provide significant data for consumers' electricity usage pattern. Past studies only took residential consumers into account and neglected industrial or business consumers. The main advantage of this study is that they used smart metering data for non-residential consumers in the USA. Another study in China has developed a new framework for short term load prediction (Fu *et al.*, 2018). In this study, five unique clusters have been identified amongst 533 households using the fuzzy C-means clustering method. The clusters are distinguished in terms of their sensitivity to high temperature and their demand profile. The consideration of both demand and sensitivity to temperature resulted in more representative clusters than those identified in studies that considered demand only, which introduced bias in the output clusters. Temperature sensitivity could generate better insights regarding habits and behavior. Application of new classification methods such as support vector machine (SVM) to identify energy usage patterns has been proposed by (Singh & Yassine 2018; Singh & Yassine 2018) to create a data mining framework which was used to visualize, analyze, and predict energy consumption time series. In this study, the UK domestic appliance-level electricity dataset has been used and 50 clusters have been recognized. The UK-DALE dataset records electricity usage of five houses every six seconds at appliances level. The main outcome of this study is the clustering of user's appliances in terms of their energy consumption. It was found that laptops, monitors, and washing machines are the top three appliances whose more efficient use can yield higher energy savings.

Power load data of USA users have been used to identify energy user behavior by developing an "Abnormal User Detection" approach based on power load multi-step clustering with multiple time scales method (Lin *et al.*, 2019). In this research, the k-means clustering method is applied to categorize people and identify 10 separate clusters. The main outcome of this research of commercial buildings energy usage data was that abnormal users (i.e., outliers with unusual energy behaviors) have a stable distribution of time and space dimensions, which is a helpful insight into energy behavior change.

A novel model that can identify an individual's use of mechanical ventilation systems in households has been proposed (Ren *et al.*, 2019). In this study, a data mining platform based on k-means clustering method has been developed to classify usage patterns of 10 households in the Netherlands by using indoor and outdoor temperature and energy consumption data. The output of this research can help to create a high-performance simulation model for buildings, as well as to properly define energy behavior change schemes by considering the Air-Conditioners (AC) usage data (noting that AC adds to electricity consumption). Identification of an individual's lifestyle by their energy impact is proposed by (Schwarzinger *et al.* 2019) to develop policies that are based on elaborated social science data. Housing, mobility, consumption, diet, information, and other/leisure activities are the different aspects comprising the individual's lifestyle and the associated energy impact is calculated for each of them. The proposed analytical framework classified individuals into 5 separate clusters using k-means clustering method. The ECHOES dataset has been used in this research and it consists of individuals' data collected across 31 European countries (Reichl *et al.*, 2019; Schwarzinger, Bird, & Skjølsvold, 2019). A new data-driven method is proposed by (Calikus *et al.* 2019) that can categorize patterns of heat energy usage for district heating network in a city located in southern Sweden. A novel method based on k-shape clustering was developed to categorize 1222 households into three different groups. It is suggested that households with different customer categories can behave in the same way. This finding is crucial regarding energy behavior in buildings, as it proved that a building may have a specific energy and carbon footprint regardless of its occupants.

To reduce the dimension of the load profile, Shi *et al.* (2020) have applied a piecewise aggregate approximation. In this study, the University of California Irvine (UCI) dataset was used to validate that the proposed method could significantly enhance efficiency and quality of clustering. Results indicated the existence of 4 different energy usage profiles for the group of individuals. For example, it was found that type 1 and type 3 consumers are similar and unimodal (less energy usage during the day and more during night), while type 4 consumers have flat energy profiles. This insight could help to identify consumers with similar habits and form a proper behavior change scheme.

A new deep reinforcement learning framework is proposed by (Zhong *et al.* 2021) for dynamic pricing demand response using the k-means clustering method for more than 1000 consumers in Tianjin, China. In this study, two types of users have been identified: price-sensitive and price-insensitive, which provides a promising starting point for policy makers because it is easier to create effective and maybe expensive policies for price-insensitive consumers and change their energy routines. Furthermore, the cluster analysis of 4422 consumers located in Xuzhou, Jiangsu, China, using the Expectation Maximization algorithm has been proposed by (Mi *et al.* 2021) to evaluate energy reduction motivations of consumers with three levels of electricity consumption: high, medium, and low. In this study, the theory of planned behavior has been used to find effective differences in psychological variables that change electricity consumption in buildings. The psychological approach could help to develop an efficient behavior change scheme by generating proper insights for policy makers. The clustering of 23 under-privileged households in Bangladesh using a hybrid combination of quantitative and qualitative data has been carried by (Neto-Bradley *et al.*2021) to recognize the controlling socio-economic factors and behaviors preventing communities from clean cooking. For input, they used socio-cultural, demographic, education, and economic indicators as quantitative data and household energy consumption preferences and practices, management of finances, social networks, and experience of local political networks as qualitative data. Subsequent analysis split the households into five different groups by applying the agglomerative clustering method. They successfully identified four pathways to overcome barriers in the process of clean cooking transformation and proposed required strategies.

A novel method to classify occupancy patterns of energy consumers in residential buildings is proposed by (Panchabikesan *et al.* 2021). In this study, they consider eight apartments in Lyon, France, and identified five typical clusters for occupants using the k-shape clustering method. As in previous studies, they found different patterns of energy usage over a 24-hour period. Different energy consumption patterns during weekdays and weekends were also observed, while most studies done previously had focused only on the variation of electricity usage on an hourly basis. A novel semi-supervised automatic clustering method based on a self-adaptive metric learning process, is proposed to recognize distinctive electricity usage patterns from 5,566 households in London. This new method classifies households into 6 clusters and proved its efficiency against other conventional methods (Zhang *et al.*, 2022). Similar to other load profile analyses, each cluster has unique electricity usage over a 24-hour period. The study also tried to address challenges in electrical load clustering by applying semi-supervised learning to create more meaningful patterns for policy makers. Real data of 71 residencies in the USA have been used to create a hybrid deep learning model to forecast and cluster electricity consumption behavior (Yang *et al.*, 2022). In this study, the convolutional long short-term memory (Co-LSTM) deep learning method has been proposed to predict load and to consider behaviors at different time resolutions (hourly, daily, and weekly). The k-means clustering method was used to group data into two distinctive clusters. The main advantage of this study is its consideration of both shorter- and longer-term behavior, as past studies have tended to investigate short-term behavior only. The inclusion of long-term behavior analysis could help create a more meaningful vision of behavior change.

Considering energy consumption and usage profiles for clustering individuals has advantages and disadvantages. Energy usage data can be potentially accessible, providing a simple and straightforward information source for decision makers. Assessing different temporal resolutions of energy usage could help improve the understanding of an individual's habits and routine. This approach is considered in the ECHOES and the ENTRUST projects. Both projects highlighted that there are knowledge gaps between current practice and proposed solutions for behavior changes. Furthermore, using hourly electricity usage data could help decision makers to identify individuals' daily habits and to define required action plans for decreasing their CO_2 impact. Lower scale resolution like daily and monthly could help policy makers to develop roadmaps and pathways for reducing individuals' carbon footprint in the long-term and to plan for permanent behavior change.

Heat energy demand and usage could help in behavior identification and carbon footprint, too. Using heat energy usage data could provide further insights as most studies consider electricity demand only. The identification of heat energy usage patterns can help policy makers to identify potential CO_2 sources and try to define incentives for individuals to change their behavior and reduce the use of heating technologies with large carbon impact, such as gas and oil boilers.

However, using energy usage data at an individual level to induce behavior change may have some drawbacks as both electricity and heat information is required, and these are difficult to access due to privacy issues. In addition, the output dataset would be enormous and require huge computational power to recognize valid patterns. So, while consideration of electricity and heat energy usage data could help to create a better vision of behavior change, at present it is impractical. Finally, the analysis of energy consumption data at a collective level may ignore certain circumstances, e.g., split incentive issues between owners and occupants associated with the ownership status of the residence/building (Castellazzi *et al.*, 2017)

3.3 Clustering attributes at collective level

3.3.1 Socio-economic, demographic, and geographic

Socio-economic, demographic and geographic variables are relevant also at collective level referring to the number of people in a building/neighborhood/city/country, average education level, occupation, income and expenditure.

It is highlighted in (Reyna *et al.* 2016) how clustering at meso-scale (i.e., buildings at neighborhood level) can improve the energy efficiency assessment of urban buildings. It is suggested that by re-defining the spatial boundaries, policies targeting neighborhood/mesoscale geographies can be formulated for energy efficiency and conservation, resulting to higher decarbonization impact. The authors support that assessment at this scale can provide sufficient detail for understanding energy consumption patterns. As such, they use demographic, building characteristics, and energy consumption data to create clusters of buildings with data collected from surveys of communities in Los Angeles and New York, USA.

The effect of tax on energy consumption and behavior to design an economic model has been investigated by (Zaharia *et al.* 2017). In this study, factors like share of environmental, energy, transport and pollution taxation in Gross Domestic Product (GDP) have been considered for 28 countries in Europe. A hierarchical method was used to cluster the countries into 7 separate groups. The results confirmed that using energy taxation can help improve competitiveness at EU-28 level. This study proved that efficient taxation on energy could assist in energy behavior change.

Assessment of key players in Europe's energy infrastructure and deep insight of human energy behavior is developed by (Axon *et al.*, 2018). In this study, technological systems (energy saving solutions) and socio-demographic factors are accounted for simultaneously. This study was part of the ENTRUST project outputs and created new insights for energy-related practices to improve stakeholder engagement in Europe's energy transition. One of the important outcomes of the ENTRUST project is the identification of barriers to behavior change initiatives. It is pointed out that there are "knowledge gaps between what is known to work to engage individuals in behavioral change, and what is currently being applied in practice." (Kennedy *et al.*, 2019).

3.3.2 Energy consumption

A novel clustering algorithm has been proposed by (Jafari-Marandi *et al.*, 2016) based on a self-organizing map to categorize buildings into different clusters based on their energy usage, while they also introduce a homogeneity index to evaluate the clusters' heterogeneity. The authors compare building clusters based on their energy profile heterogeneity index and suggest different systems (shared battery and renewable energy) depending on this index. Energy consumption data for 30 buildings has been used and buildings have been categorized into 4 clusters.

Heating and energy consumption profiles for education buildings have been analyzed by (Ma *et al.*, 2017) to recognize conventional daily heating usage patterns. In this study, partitioning was achieved using the medoids method using data from 19 education buildings in Norway to identify heat demand patterns. It was observed that some buildings had peak heat demand between 7AM and 6PM while others were flat during a 24-hour period. While several researchers have focused on electricity demand, this study considers heat demand for energy behavior analysis.

A novel approach, based on the agglomerative hierarchical clustering-based strategy, considers shared nearest neighbors to recognize the patterns of daily electricity consumption of two library buildings at the University of Wollongong (Li *et al.*, 2019). The authors identified 10 distinct clusters of building energy usage; 3 clusters had flat energy profiles for almost whole 24 hours while others were unimodal. Consideration of non-residential buildings is one of the novelties of this study; past research discussed in this report focused exclusively on residential buildings when identifying energy behavior patterns. Similarly, another study focused on university buildings in Italy to identify typical and extreme days for multi-energy systems design optimization (Zatti *et al.*, 2019).

In a further study, the gaussian mixture model (GMM) algorithm has been proposed for clustering the heat usage of 561 consumers and identified 5 clusters. In this study, the authors suggest that the ambient temperature has a significant effect on heat demand (Wang *et al.*, 2019). According to this study and that of (Fu *et al.* 2018), temperature plays a significant role in the identification of energy habits and behaviors, and it can create a better vision regarding behavior change schemes. (Ashouri *et al.* 2019) proposes a novel method to identify how occupants of 80 buildings in Japan behave in terms of their electricity usage. Using the k-means clustering method, two different energy consumption behaviors were recognized in the dataset. Authors assign ranking for each building to assess building energy consumption (high or low energy usage). This approach is beneficial for energy behavior change as it helps decision makers to form separate schemes for each building regardless of occupant.

In another study on energy consumption clustering at a building level (Culaba *et al.*, 2020a), the authors use machine learning to characterize and forecast energy consumption for multi-application buildings. One of the important points about these studies is that they considered the energy usage at a building level, rather than at the individual level, to provide better insights for policy makers because of the wider and clearer vision it offers compared to the individual level. The relationship between GDP, global competitiveness index, and usage of renewable energy has been discussed by (Simionescu *et al.* 2020), in which a dataset of 28 European countries was clustered into two groups using the k-means clustering method, based on their share of renewable energy sources in their gross final energy consumption.

Application of new deep neural networks like Generative Adversarial Network (GAN) for creating realistic electricity usage patterns in buildings is proposed by (Wang & Hon 2020). In this study, the authors applied GAN on the Building Data Genome Project dataset from North America and identified 14 clusters for load profile using the dataset. Five factors are used to quantify load shape of buildings: base load, peak load, peak load duration, and rise and fall times. Like other studies, flat and unimodal load profiles were derived, and it is concluded that generated load profiles are significantly comparable to real load profiles. Generated and real load profiles can provide significant insights about energy behavior by disclosing energy usage patterns in buildings and it is one of the important outcomes of this research.

Identification of energy consumption patterns in university students' dormitories has been developed by (Zhou *et al.* 2021). In this research, the k-means clustering algorithm has been used to categorize data into five clusters that exhibit distinctive energy consumption. This study provided a detailed vision of electricity usage patterns in university dormitories but found no meaningful correlation between outside temperature and electricity. The study concluded that there is a strong relationship between the academic calendar and electricity usage in the dormitories.

Assessment of energy consumption at a larger scale than individuals could be effective in terms of energy savings, because it enables decision makers to consider the contextual factors affecting their decisions and helps them to design more comprehensive roadmaps. Different clustering algorithms have been applied to identify energy consumption patterns at different spatial levels (e.g., from buildings to countries). At a building level, it appears that energy usage patterns do not differ substantially across different occupants' lifestyles, especially as far as heating systems are concerned. Designing policies for buildings with the same characteristics such as age, heating systems, and region could potentially have a higher impact on cutting CO_2 emissions than targeting an individual's behavior change. As such, it may be more effective if policies for energy behavior are designed to apply at a collective level rather than at an individual level. For example, an occupant of a flat often does not have the power to change the central heating system of the multi-apartment building; policies introduced at a building level, along with the necessary incentives, may prove to be more effective. Furthermore, in some regions, it is difficult or impossible to change the heating systems in building due to lack of infrastructure such as old heating systems and difficulties to access funding.

3.3.3 Clean energy and greenhouse gas emissions

Variables related to clean energy usage and CO_2 emissions have also been considered at a collective level. The usage of renewable/cleaner energy resources as well as the carbon emissions intensity -have been considered as attributes for clustering. For example, the amount of CO_2 emissions at country level has been considered to categorize countries in terms of their carbon intensity and study energy behavior at a larger scale.

The overall penetration of Photo Voltaic (PV) installation in Germany has been evaluated using an artificial neural network model, with PV installation historic data and socio-economic factors (Lück & Moser, 2019). The number of PV installations during 2000-2018, full load hours, income, and share of green party voters were clustered using the k-means method and the neural network proved to give more accurate results than the fundamental model that only considered single variables such as PV installation. The energy demand, energy production, and sociodemographic variables were used to identify clusters of households with the aim of providing more robust insights to policy makers.

The effect of environmental policies in China has been discussed at provincial level by Gong *et al.*, (2019). Chinese provinces are evaluated against different variables such as energy consumption, CO_2 emissions, and GDP during the last 20 years. The study identified five separable clusters and confirmed that the proposed algorithm can separate provinces by their relative impact on the environment and energy consumption. It is noteworthy that this study investigates CO_2 emissions and behavior changes together at provincial level.

Greenhouse gas emissions, energy consumption, energy efficiency, and energy import of European countries are compared by Beken (2019). In this research, 30 European countries are evaluated and classified using data over the period from 2009 to 2017, using the k-means clustering method to compare them with Turkey. The outcome of this study could be helpful in decision-making at national level regarding energy behavior change.

Uncertain decarbonization pathways were discussed by (Li *et al.*, 2020), who applied a clustering method for long term characterization. More than 600 synthetic decarbonization pathways have been evaluated and classified using the k-means clustering method to identify representative pathways for 2030 and 2050 in the UK. In this study, 5 clusters were identified and each of them has a specified focus on a certain energy sector. For example, cluster 2 focused on electricity technologies by 2050, while cluster 4 focused on hydrogen. The outcome of this study could be helpful regarding large scale generation schemes for energy behavior change by suggesting the most efficient pathways for decarbonization.

Energy behavior changes have been studied based on renewable energy installations, using clean energy for cooking and CO_2 impact measurement. Different levels of analysis have been considered, but most research has focused on buildings and countries. Regarding, renewable energy usage and clean energy for heating, most of the studies emphasized on buildings. They used different clustering methods to identify patterns and create distinct groups that share common characteristics. Also, countries are clustered based on renewable energy installations and their CO_2 emissions profiles.

Studying countries with respect to their CO_2 emissions could offer insights for designing policies and support decision makers as they can provide a rough estimate of the status of CO_2 emissions and create decarbonization pathways at a national level. However, clustering at this more holistic level of analysis may be less accurate and potentially neglect some groups of people, such as the more vulnerable to energy poverty and energy access. The effect of individuals on CO_2 emissions is not negligible, so measurement of it could be beneficial for behavior identification although it is very difficult to calculate the exact amount of emissions for a person. Also, considering the CO_2 impact at a country level would be helpful and policy makers could reduce it using renewable energy installations and electric buses. Policies targeted at the mid-scale collective level, such as buildings, may potentially have a higher impact on the behavior change and the CO_2 impact of individuals. Changing building heating systems and providing incentives to improve the energy performance of their houses could be more efficient.

4 Key Clustering Variables

4.1 Introduction

Following the literature review (Section 3) of studies carrying out clustering of energy and environmental behaviors/energy consumption patterns at different levels of analysis, this Section, first, discusses common goals of clustering studies relevant to energy decarbonization and then summarizes key clustering factors identified as relevant for deriving clusters for decarbonization. As mentioned above, two levels of clustering have been considered: individual and collective; key variables will be introduced, based on this categorization (Figure 3).

Clustering variables	Individual and household level	Collective level (building, district, city, country)
Socio-economic, demographic and geographic data	Age, income, education, employment, gender, income, number of family members	Average income, education, employment, number of individuals in the group
Psychological	Personal norms, preferences, values, life-goals and awareness of consequences associated with energy savings	-
Energy lifestyles & consumption	Energy consumption preferences and practices across different areas of life, such as housing, mobility, cconsumption & diet (energy lifestyles)	Electricity/heating/cooling load data at building, district, province, etc. level
Building/devices physical characteristics	-	Infrastructure characteristics (buildings' age, type, size, area, energy performance, centralized/decentralized electricity/heating infrastructure) Space/water heating/cooling technologies, other devices

Goals

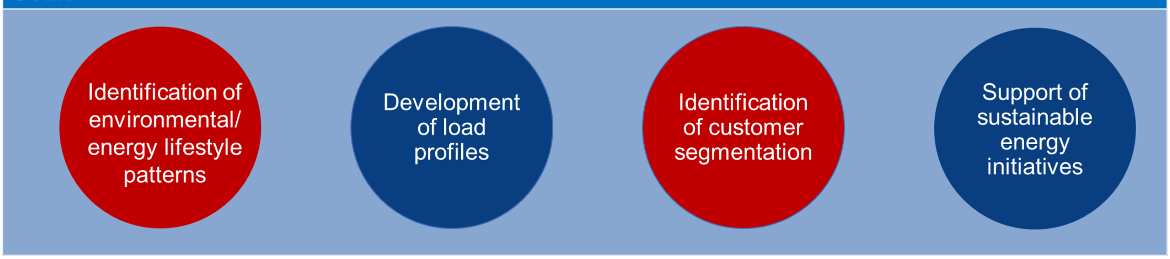

Figure 3 Mapping of key clustering variables at individual and collective level

4.2 Goals of clustering studies

Typical goals of clustering studies include one or more of the following:

1) To derive load profiles to improve accuracy of load forecasting (electricity/heating/cooling)
2) To identify socio-technical patterns within a population (relationships between socio-demographics, energy lifestyles and technical characteristics of the energy infrastructure) for effective energy system assessment and management
3) To identify customer segments (e.g., for the adoption of innovative technologies) for developing more relevant and targeted policies
4) To support the design of sustainable energy initiatives, policies and roadmaps for long-term resource/energy system planning

Developing realistic load profiles can improve understanding of the building energy efficiency, identification of unnecessary waste as well as increase the accuracy of load forecasting (Wang and Hong, 2020). Load profiles have been developed by numerous researchers using smart meter data (mostly for electrical power consumption) collected at high temporal granularities at both collective and individual level. For example, Wang and Hong (2020) have used smart meter data of 156 office buildings to develop building load profiles, while(Czétány *et al.*, 2021) used smart meter data from residential buildings and proposed four different electric user profiles. The latter also examined the impact of specific parameters (day of the week, seasonality, geography, housing type, etc.) on the consumption of occupants. Clustering load data can be used to develop prediction models to characterize and forecast the energy consumption at different scales of analysis (Culaba *et al.*, 2020). Accurate forecasting of energy consumption is essential not only for economic and sustainability reasons (i.e., energy conservation and decarbonization), but also to facilitate power system planning and stable grid operations (Khan *et al.*, 2021). Clustering under different spatial boundaries has been cited to be useful for improving predictions and forecasts of short-term and mid-term future electricity consumption, as well as for sub- and cross-city energy studies or different spatial scales (such as apartment, building, and floor level) (Reyna et. Al., 2016; Khan *et al.*, 2021). Load forecasting plays an essential role in modern power system planning, scheduling, operation, maintenance, and control, and it has received a great deal of attention from the research community (Fu *et al.*, 2018; Testi *et al.*, 2020).

Apart from smart metering/load data, some studies have also tried to map trends in the energy behavior of building occupants and uncover their temporal energy consumption patterns. Goal of such studies has been the identification of typical energy usage profiles/consumption patterns and lifestyles (Schwarzinger, et al., 2019; Lu et al., 2019; Bogin et al., 2021; Moon et al., 2021). Such studies analyze patterns (customers' energy-use habits) for effective operation and management by energy system companies (e.g., district heating companies) to optimize their operations, to implement new control strategies or to optimize a smart grid (Calikus et al., 2019). They can enable utilities perform online/real-time energy predictions and engage consumers upon realizing consumption changes for an improved smart grid energy saving program (Singh & Yassine, 2018). Clustering consumers in terms of their characteristics and preferences towards identifying customer segments that can be potential innovators and early adopters of innovative energy technologies, can support the design of suitable marketing strategies to accelerate the diffusion of the technology (Moon et al., 2021). Pattern identification has been also useful to evaluate buildings' energy-saving potential through occupants' contribution, as well as ranking buildings in terms of achieved and potential savings (Ashouri et al., 2019). Together with the identification of energy consumption patterns, a key objective of clustering studies has been the identification of the effect of individual lifestyles on their energy consumption. Identification of groups based on the impact of their climate and energy-related behavior using a lifecycle energy assessment in six main areas of life (“Housing”, “Mobility”, “Diet”, “Consumer Goods”, “Recreational Hobbies” and “Information) has also been proposed, aiming at providing a basis towards target group- oriented and impact-oriented policy design (Schwarzinger, *et al.*, 2019).

Furthermore, at building level, evaluating the heterogeneity across multiple buildings in terms of their energy consumption can allow to determine which types of buildings should form a cluster to share energy and exchange information in the context of a smart grid. Clustering buildings can contribute to the reduction of energy consumption, improving the sustainability and resilience of the smart grid, leading to significant energy savings (Jafari-Marandi *et al.*, 2016).

Cluster analysis in terms of energy profiles can also assist the identification of dominant characteristics (socio-economic, psychological, etc.) that can act as enablers and barriers towards sustainable energy transition pathways, as well as behavior change initiatives (Mi *et al.*, 2021; Axon *et al.*, 2018).

Common purpose of clustering studies in the context of climate change and sustainable energy transition is to support decision makers to take informed decisions in the design of policies, initiatives and roadmaps for decarbonizing the energy system. Varying levels of decision support include marketing strategies (customized to customer segments) for the diffusion of innovative sustainable energy technologies (Moon *et al.*, 2019), impact-oriented policy design (Schwarzinger *et al.*, 2019), electricity network planning and operations policies according to consumption patterns (Zhou *et al.*, 2021) , district heating operation and management (Calikus *et al.*, 2019b) and demand-side management policy design (Khan *et al.*, 2021; Fu *et al.*, 2018; Jiang *et al.*, 2018), among others.

Given the urgent need to a rapid low carbon energy system transition, policy priorities are increasingly focused on energy efficiency, greenhouse gas emissions mitigation, energy savings and reduction of energy intensity (deLlano-Paz *et al.*, 2016). However, governments also need to account for other policy priorities including access to energy, energy affordability and sustainability. Therefore, the national agendas need to balance or prioritize the different aspects of this transition while considering people’s vision, priorities, and needs.

Although energy behavior characteristics (at both collective and individual level) have been extensively analyzed by the research community, limited studies have explored community visions of energy transition (Morrisey *et al.*, 2020). In the context of the EU Horizon 2020 ENTRUST project (Axon *et al.*, 2018) a cluster analysis approach of community’s visions of energy change pathways was carried out (Morrissey *et al.*, 2020a). Researchers investigated the visions of community residents’ energy system transition through a survey in Liverpool UK, what issues they prioritize and the role of various technologies on this transition. In their study, they identified distinct energy visions, including community affordability, energy security and environmental sustainability, among others. Mapping community’s vision for the future can provide an essential foundation for long-term policies and strategies.

In the context of ENCLUDE, a key priority is to ensure the inclusive involvement of citizens to “engage ‘hard to reach’ citizens, and to understand how their aspirations and perceptions can be mapped onto the requirements, or opportunities of a low-carbon transition”. According to this approach, a different way to develop groupings of citizens is based on their *needs and priorities*. People’s priorities can be affected by their socio-economic status and geographical locations. People in certain geographic areas suffer from lack of access to clean heating or electricity services and energy poverty. In such cases, policies tailored to the specific conditions of such citizens should be developed and to this end, distinct categories of consumers' whose priorities/needs are different should be distinguished. As such, clustering citizens in terms of their needs could include factors such as affordability, access to energy services, efficiency, sustainability. As far as energy poverty is concerned, there are several qualitative and quantitative indicators that need to be considered. Qualitative indicators refer to a self-assessed situation of households such as the condition of the house (presence of leaks, damp, or rot), ability to pay utility bills or achieve thermal comfort, while quantitative indicators are based on the citizens' income, expenditures, and the share of energy utility bills in income (Price *et al.*, 2012).

4.3 Individual level

For studies focusing on clustering at the individual level, the following key variables enabling the meaningful grouping in terms of energy/environmental behavior have been identified:

- *Socio-economic and demographic* variables, including age, gender, income, education, employment, etc. have been widely used to approximate the energy and environmental profile of the individual**.** In general, individuals with higher incomes and households larger in size tend to use more energy. However, higher income individuals also have relatively more capacity to adopt (costly) energy-saving measures and innovative energy-saving technologies, such as the installation of in-home insulation or the purchase of electric vehicles, which could result to high decarbonization impact (Abrahamse & Steg, 2009; Newton & Meyer, 2013). Geographic factors (urban, rural) were also found to be important in determining energy consuming behaviors.
- *Psychological* variables concern factors such as personal norms, preferences, values, lifegoals and awareness of consequences associated with energy savings. Numerous studies have highlighted what is known as the "attitude-action" gap, suggesting that high environmental awareness including attitudes and intentions towards sustainability does not necessarily imply corresponding behavior or result to high ecological impact (Newton & Meyer, 2013; Schwarzinger *et al.*, 2019). As such, factors like environmental self-identity play an ambiguous role in predicting actual environmental impacts, since people with high pro-environmental self-identity tend to behave in an ecologically responsible way but often focus on actions with relatively small ecological benefits (Moser & Kleinhückelkotten, 2018). Furthermore, some studies suggest that there are factors that outweigh attitudes, opinions and intentions as indicators of consumer behavior; such factors include access to information, organization and finance (Newton & Meyer, 2013).
- *Energy consumption/Environmental lifestyles* require the consideration of further specific variables that concern the individual's energy behavior and use of equipment. Factors related to the individual's lifestyle across different areas of life (e.g., housing, mobility, consumption, diet, information, and other/leisure activities). Indicatively:
    - Housing: consideration of the building characteristics (age, type, size, area, energy performance), space and water heating/cooling technologies as well as other electricity consuming devices.
    - Mobility: transportation used, travelling habits.
    - Consumption & Diet: durable and non-durable goods, nutrition habits.

  According to the review outputs, housing and mobility appear to have a higher impact on the development of distinct clusters in relation to other areas of life.

4.4 Collective level

Research studies at a collective level typically focus on developing energy profiles at building, neighborhood, city, or country level.

- *Socio-economic and demographic variables* are relevant also at collective level, including the average age, number of people in a building/neighborhood/city/country, average level of education, occupation, income and expenditure of the group of people.
- *Energy infrastructure* is associated with availability and access to sustainable energy services/infrastructure, such as the availability of low carbon centralized or decentralized electricity and heating/cooling (of space) infrastructure (e.g., existence of district heating) that can support the transition to decarbonization. Another relevant example is the ownership or intention to purchase an electric vehicle (EV). The decision to adopt this technology depends on the existence of charging facilities within the country.
- *Energy consumption patterns* relate to electricity/heating loads, fuel use by activity (i.e., housing, mobility, and consumption at collective level), factors influencing energy-relevant decisions, time of use, etc. Numerous studies focus on the specification of energy behavior patterns at a building level and consider different variables for categorization of the buildings' energy footprints. Clustering methods have been used to identify building electricity and heat load patterns. Like research at an individual level, fewer studies have considered heat energy consumption, when clustering buildings in terms of their energy footprint, while majority has focused on electricity consumption profiles. It is recommended that both electricity and heating loads are used to identify energy behavior patterns to create useful insights for policy makers. Most studies have focused on residential buildings when trying to characterize the building's energy behavior, and only a few have considered different types of buildings such as educational structures or the tertiary sector.
- The *environmental performance* at collective level is associated with the measurement of CO_2 emissions footprint at a building or country level. The environmental performance is correlated with both the energy infrastructure (electricity, heating/cooling) and behavior patterns across individuals' different areas of life, as previously discussed. Consideration of the infrastructure could provide insights to policy makers about the expected energy behavior of the population at a specific scale, assisting targeted interventions and planning for CO_2 footprint reduction. Relatively few studies have assessed CO_2 emissions at a building level; most of them focus on country level.
- Clustering based on citizens' *priorities/needs* (affordability, sustainability, efficiency, etc.) is also relevant at collective level. Grouping of buildings and neighborhoods based on the identified priorities of the local population is expected to provide useful insights regarding the areas that should be targeted by policies and to inform a policy-mix with multiple simultaneous objectives, for example simultaneous prioritization of affordability, security and sustainability aspects (Morrissey *et al.*, 2020).

5 Summary and conclusions

This report maps different clustering variables that can be used to derive citizens' clusters for decarbonization at two different levels of analysis: individual and collective. In the context of this work, the distinction between the two levels was on whether data were clustered on an individual or an aggregated level: an individual level would be everything from individual to household level and collective level would be data collected at a building, district or another spatial level.

Citizens' clusters for decarbonization may not necessarily involve citizens with common demographic characteristics; rather, they may involve demographically diverse groups sharing common characteristics of energy behavior, or common priorities/needs, allowing the identification of potential "hidden" commonalities. Indicative examples include readiness to embrace energy citizenship actions, carbon emissions profiles, response to energy citizenship initiatives, as well as access to sustainable energy services/sources and energy poverty.

The key motivation to identify groupings of people sharing common characteristics is to enable the efficient design/implementation of policy measures, as well as to capture the key determining factors characterizing different groups. Such groupings can inform policy makers on the citizens' profiles and the profiles of groups of citizens that should be targeted to maximize energy savings and carbon emissions reduction potential, as well as to support decisions on the appropriate policy initiatives that should be put forward to tap into this decarbonization potential.

At an *individual* level, major variables for clustering energy behaviors were categorized as socio-economic and demographic, psychological, energy consumption/environmental lifestyles and other contextual variables. Most studies found in literature have focused on electricity consumption patterns, while fewer considered heating/cooling energy.

At a *collective* level, major variables were categorized as socio-economic and demographic, energy infrastructure variables, energy consumption patterns, environmental performance, and other contextual factors. Adopting a collective approach can contribute to the formulation of spatially targeted policies for energy efficiency and conservation.

Key conclusions derived from this critical review are summarized below:

- Most studies focused on deriving clusters to distinguish groups based on their energy consumption patterns (Abrahamse & Steg, 2009), their pro-environmental lifestyle (Newton & Meyer 2013) and their motivation to adopt innovative technologies (Moon *et al.*, 2021).
- The sociodemographic variables such as age, location of residency, income, education, as well as the psychological ones, e.g., personal values, willingness to change, and motivation were found to be key clustering variables. Contextual factors such as access to information, organization and finance can also be important and can outweigh attitudes, opinions, and intentions for behavior change.
- A key attribute at both individual and collective level appeared to be the energy consumption patterns across specific sectors, namely housing (electricity, heating/cooling) and mobility. Different temporal resolutions for energy consumption have been considered. Hourly, daily, and weekly energy usage data could help policy makers to recognize patterns and form proper schemes for changing citizens' behavior in the short- and the long-term. A lot of studies focus on electricity consumption by

end users (also due to the availability of smart metering data), while relatively fewer consider heating/cooling consumption patterns (mainly based on survey-based estimations). A cross-sectoral consideration can provide more accurate energy usage patterns and potentially contribute to the introduction of targeted energy behavior change schemes.

- Data on carbon emissions and electricity/heating consumption may be preferable to be collected at a building level, as this would result to more representative energy use patterns to be devised. Taking the relevant building variables into consideration could result to a more realistic estimation of the amount of energy consumed. For example, variables such as building type, age, heating/cooling energy systems, can assist to identify the energy intensity of its residents and address suitable schemes for building upgrades. It is unlikely that the carbon footprint of residents of an old, inefficient building can be reduced substantially by only changing the residents' energy behavior; rather, this is an infrastructure issue, which needs to be addressed at a building level (e.g., through the installation of insulation systems). Furthermore, most studies have focused on residential buildings, while fewer have considered different sectors, such as the educational or the tertiary sector.
- An interesting and potentially insightful way to cluster citizens and groups of citizens is based on their needs/priorities (affordability, access to energy, sustainability, efficiency). Especially in neighborhoods/areas/countries suffering from lack of access to clean and affordable energy services (heating/cooling, electricity), factors such as the low income, the high energy costs, and the house energy performance are becoming increasingly important. In such cases, there are distinct categories of citizens whose priorities/needs are different.

From the review of literature, it became apparent that there is need to continue studying energy and environmental lifestyles by looking at both individual and collective level clustering variables (individual, household, building, neighborhood, country), as aimed by a number of previous studies (Ashouri et al., 2019; Calikus et al., 2019; Reyna et al., 2016). Aggregating data at a higher level of analysis, while taking into consideration the relationship with demographics, building characteristics and energy behaviors across different areas of life (as usually individuals are influenced by the activities of others located in the same area), can be effective for the development of spatially targeted policies for energy savings and decarbonization.

Finally, although there is an urgent need to accelerate the transition to a low carbon energy system, urging governments to develop policies focused on cleaner energy technology diffusion, greenhouse gas emissions mitigation, energy savings and reduction of energy intensity, people's needs and priorities (energy poverty reduction, access to cleaner energy services, environmental sustainability, among others) should also be taken into consideration whilst this transition takes place (deLlano-Paz, 2016).

References

Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of Economic Psychology*, *30*(5), 711–720. https://doi.org/10.1016/j.joep.2009.05.006

An, J., Yan, D., & Hong, T. (2018). Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings. *Energy and Buildings*, *174*, 214–227. https://doi.org/10.1016/j.enbuild.2018.06.035

Ashouri, M., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2019). Development of a ranking procedure for energy performance evaluation of buildings based on occupant behavior. *Energy and Buildings*, *183*, 659–671. https://doi.org/10.1016/j.enbuild.2018.11.050

Axon, S., Morrissey, J., Aiesha, R., Hillman, J., Revez, A., Lennon, B., Salel, M., Dunphy, N., & Boo, E. (2018). The human factor: Classification of European community-based behaviour change initiatives. *Journal of Cleaner Production*, *182*, 567–586. https://doi.org/10.1016/j.jclepro.2018.01.232

Blasch, J., Barnes, J. P., & Palm, J. (2022). *Putting People at the Heart of Energy Transitions Social Innovation in Energy: four projects shine a light on the path forward*. https://www.researchgate.net/publication/360218763

Bogin, D., Kissinger, M., & Erell, E. (2021). Comparison of domestic lifestyle energy consumption clustering approaches. *Energy and Buildings*, *253*, 111537. https://doi.org/10.1016/j.enbuild.2021.111537

Boucher, J. L., Araújo, K., & Hewitt, E. (2018). Do education and income drive energy audits? A socio-spatial analysis of New York State. *Resources, Conservation and Recycling*, *136*, 355–366. https://doi.org/10.1016/j.resconrec.2018.05.009

Buttitta, G., Turner, W. J. N., Neu, O., & Finn, D. P. (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings*, *198*, 84–99. https://doi.org/10.1016/j.enbuild.2019.05.056

Caballero, V., Vernet, D., & Zaballos, A. (2020). A heuristic to create prosumer community groups in the social internet of energy. *Sensors (Switzerland)*, *20*(13), 1–26. https://doi.org/10.3390/s20133704

Calikus, E., Nowaczyk, S., Sant'Anna, A., Gadd, H., & Werner, S. (2019a). A data-driven approach for discovering heat load patterns in district heating. *Applied Energy*, *252*. https://doi.org/10.1016/j.apenergy.2019.113409

Calikus, E., Nowaczyk, S., Sant'Anna, A., Gadd, H., & Werner, S. (2019b). A data-driven approach for discovering heat load patterns in district heating. *Applied Energy*, *252*. https://doi.org/10.1016/j.apenergy.2019.113409

Campos, I., & Marín-González, E. (2020). People in transitions: Energy citizenship, prosumerism and social movements in Europe. *Energy Research and Social Science*, *69*. https://doi.org/10.1016/j.erss.2020.101718

Castellazzi, L., Bertoldi, P., Economidou, M. Overcoming the split incentive barrier in the building sectors: unlocking the energy efficiency potential in the rental & multifamily sectors, EUR 28058 EN, Luxembourg: Publications Office of the European Union, 2017, ISBN 978-92-79-58837-2, doi:10.2790/912494, JRC101251

Cornelius, M., Armel, K. C., Hoffman, K., Allen, L., Bryson, S. W., Desai, M., & Robinson, T. N. (2014). Increasing energy- and greenhouse gas-saving behaviors among adolescents: A school-based cluster-randomized controlled trial. *Energy Efficiency*, *7*(2), 217–242. https://doi.org/10.1007/s12053-013-9219-5

Creutzig, F., Callaghan, M., Ramakrishnan, A., Javaid, A., Niamir, L., Minx, J., Müller-Hansen, F., Sovacool, B., Afroz, Z., Andor, M., Antal, M., Court, V., Das, N., Díaz-José, J., Döbbe, F., Figueroa, M. J., Gouldson, A., Haberl, H., Hook, A., … Wilson, C. (2021). Reviewing the scope and thematic focus of 100 000 publications on energy consumption, services and social aspects of climate change: A big data approach to demand-side mitigation. In *Environmental Research Letters* (Vol. 16, Issue 3). IOP Publishing Ltd. https://doi.org/10.1088/1748-9326/abd78b

Culaba, A. B., del Rosario, A. J. R., Ubando, A. T., & Chang, J. S. (2020a). Machine learning-based energy consumption clustering and forecasting for mixed-use buildings. *International Journal of Energy Research*, *44*(12), 9659–9673. https://doi.org/10.1002/er.5523

Culaba, A. B., del Rosario, A. J. R., Ubando, A. T., & Chang, J. S. (2020b). Machine learning-based energy consumption clustering and forecasting for mixed-use buildings. *International Journal of Energy Research*, *44*(12), 9659–9673. https://doi.org/10.1002/er.5523

Czétány, L., Vámos, V., Horváth, M., Szalay, Z., Mota-Babiloni, A., Deme-Bélafi, Z., & Csoknyai, T. (2021). Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy and Buildings*, *252*, 111376. https://doi.org/10.1016/j.enbuild.2021.111376

deLlano-Paz, F., Martínez Fernandez, P., & Soares, I. (2016). Addressing 2030 EU policy framework for energy and climate: Cost, risk and energy security issues. *Energy*, *115*, 1347–1360. https://doi.org/10.1016/j.energy.2016.01.068

Diao, L., Sun, Y., Chen, Z., & Chen, J. (2017). Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings*, *147*, 47–66. https://doi.org/10.1016/j.enbuild.2017.04.072

Fang, H., Wang, Y. W., Xiao, J. W., Cui, S., & Qin, Z. (2021). A new mining framework with piecewise symbolic spatial clustering. *Applied Energy*, *298*. https://doi.org/10.1016/j.apenergy.2021.117226

Fu, X., Zeng, X. J., Feng, P., & Cai, X. (2018). Clustering-based short-term load forecasting for residential electricity under the increasing-block pricing tariffs in China. *Energy*, *165*, 76–89. https://doi.org/10.1016/j.energy.2018.09.156

Gabe-Thomas, E., Walker, I., Verplanken, B., & Shaddick, G. (2016). Householders' mental models of domestic energy consumption: Using a sort-and-cluster method to identify shared concepts of appliance similarity. *PLoS ONE*, *11*(7). https://doi.org/10.1371/journal.pone.0158949

Gong, B., Zheng, X., Guo, Q., & Ordieres-Meré, J. (2019). Discovering the patterns of energy consumption, GDP, and CO2 emissions in China using the cluster method. *Energy*, *166*, 1149–1167. https://doi.org/10.1016/j.energy.2018.10.143

IEEE Staff. (2019). *2019 8th International Conference on Renewable Energy Research and Applications (ICRERA)*. IEEE.

Jafari-Marandi, R., Hu, M., & Omitaomu, O. F. A. (2016). A distributed decision framework for building clusters with different heterogeneity settings. *Applied Energy*, *165*, 393–404. https://doi.org/10.1016/j.apenergy.2015.12.088

Jiang, Z., Lin, R., & Yang, F. (2018). A hybrid machine learning model for electricity consumer categorization using smart meter data. *Energies*, *11*(9). https://doi.org/10.3390/en11092235

Kennedy, R., Numminen, S., Sutherland, J., & Urpelainen, J. (2019). Multilevel customer segmentation for off-grid solar in developing countries: Evidence from solar home systems in Rwanda and Kenya. *Energy*, *186*. https://doi.org/10.1016/j.energy.2019.07.058

Khan, A. N., Iqbal, N., Rizwan, A., Ahmad, R., & Kim, D. H. (2021). An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings. *Energies*, *14*(11). https://doi.org/10.3390/en14113020

Knowles, H. S., Hostetler, M. E., & Liebovitch, L. S. (2018). Describing the dynamics, distributions, and multiscale relationships in the time evolution of residential building energy consumption. *Energy and Buildings*, *158*, 310–325. https://doi.org/10.1016/j.enbuild.2017.09.071

Krayem, A., al Bitar, A., Ahmad, A., Faour, G., Gastellu-Etchegorry, J. P., Lakkis, I., Gerard, J., Zaraket, H., Yeretzian, A., & Najem, S. (2019). Urban energy modeling and calibration of a coastal Mediterranean city: The case of Beirut. *Energy and Buildings*, *199*, 223–234. https://doi.org/10.1016/j.enbuild.2019.06.050

Lara Lück, & Albert Moser. (2019). Artificial neural networks modeling photovoltaic power system allocation – Can artificial intelligence beat a fundamental approach? *2019 16th International Conference on the European Energy Market (EEM)*.

Li, K., Yang, R. J., Robinson, D., Ma, J., & Ma, Z. (2019). An agglomerative hierarchical clustering-based strategy using Shared Nearest Neighbours and multiple dissimilarity measures to identify typical daily electricity usage profiles of university library buildings. *Energy*, *174*, 735–748. https://doi.org/10.1016/j.energy.2019.03.003

Li, P. H., Pye, S., & Keppo, I. (2020). Using clustering algorithms to characterise uncertain long-term decarbonisation pathways. *Applied Energy*, *268*. https://doi.org/10.1016/j.apenergy.2020.114947

Lin, R., Yang, F., Gao, M., Wu, B., & Zhao, Y. (2019). AUD-MTS: An abnormal user detection approach based on power load multi-step clustering with multiple time scales. *Energies*, *12*(16). https://doi.org/10.3390/en12163144

Lu, Y., Tian, Z., Peng, P., Niu, J., Li, W., & Zhang, H. (2019). GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system. *Energy and Buildings*, *190*, 49–60. https://doi.org/10.1016/j.enbuild.2019.02.014

Ma, Z., Yan, R., Li, K., & Nord, N. (2018). Building energy performance assessment using volatility change based symbolic transformation and hierarchical clustering. *Energy and Buildings*, *166*, 284–295. https://doi.org/10.1016/j.enbuild.2018.02.015

Ma, Z., Yan, R., & Nord, N. (2017). A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. *Energy*, *134*, 90–102. https://doi.org/10.1016/j.energy.2017.05.191

Mi, L., Xu, T., Sun, Y., Yang, H., Wang, B., Gan, X., & Qiao, L. (2021). Promoting differentiated energy savings: Analysis of the psychological motivation of households with different energy consumption levels. *Energy*, *218*. https://doi.org/10.1016/j.energy.2020.119563

Moon, H. bin, Park, S. Y., & Woo, J. R. (2021). Staying on convention or leapfrogging to eco-innovation?: Identifying early adopters of hydrogen-powered vehicles. *Technological Forecasting and Social Change*, *171*. https://doi.org/10.1016/j.techfore.2021.120995

Morrissey, J., Schwaller, E., Dickson, D., & Axon, S. (2020). Affordability, security, sustainability? Grassroots community energy visions from Liverpool, United Kingdom. *Energy Research & Social Science*, *70*, 101698. https://doi.org/10.1016/j.erss.2020.101698

Moser, S., & Kleinhückelkotten, S. (2018). Good Intents, but Low Impacts: Diverging Importance of Motivational and Socioeconomic Determinants Explaining Pro-Environmental Behavior, Energy Use, and Carbon Footprint. *Environment and Behavior*, *50*(6), 626–656. https://doi.org/10.1177/0013916517710685

Neto-Bradley, A. P., Rangarajan, R., Choudhary, R., & Bazaz, A. (2021). A clustering approach to clean cooking transition pathways for low-income households in Bangalore. *Sustainable Cities and Society*, *66*, 102697. https://doi.org/10.17863/CAM.5

Panchabikesan, K., Haghighat, F., & Mankibi, M. el. (2021). Data driven occupancy information for energy simulation and energy use assessment in residential buildings. *Energy*, *218*. https://doi.org/10.1016/j.energy.2020.119539

Reichl, J., Cohen, J., Kollmann, A., Azarova, V., Klöckner, C., Royrvik, J., Vesely, S., Carrus, G., Panno, A., Tiberio, L., Fritsche, I., Masson, T., Chokrai, P., Lettmayer, G., Schwarzinger, S., & Bird, N. (2019, November 1). *International survey of the ECHOES project*.

Ren, X., Zhang, C., Zhao, Y., Boxem, G., Zeiler, W., & Li, T. (2019). A data mining-based method for revealing occupant behavior patterns in using mechanical ventilation systems of Dutch dwellings. *Energy and Buildings*, *193*, 99–110. https://doi.org/10.1016/j.enbuild.2019.03.047

Reyna, J. L., Chester, M. v., & Rey, S. J. (2016). Defining geographical boundaries with social and technical variables to improve urban energy assessments. *Energy*, *112*, 742–754. https://doi.org/10.1016/j.energy.2016.06.091

Schwarzinger, S., Bird, D. N., Lettmayer, G., Henriksen, I. M., Skjølsvold, T. M., Olaeta, X. U., Alvarez, L. P., Velte, D., Iturriza, I. J., Biresselioglu, M. E., Demir, M. H., Dimitrova, E., Tasheva, M., Tiberio, L., Panno, A., Carrus, G., & Costa, S. (2019). *Comparative Assessment of European Energy Lifestyles*.

Schwarzinger, S., Bird, D. N., & Skjølsvold, T. M. (2019). Identifying consumer lifestyles through their energy impacts: Transforming social science data into policy-relevant group-level knowledge. *Sustainability (Switzerland)*, *11*(21). https://doi.org/10.3390/su11216162

Seebauer, S., Fleiß, J., & Schweighart, M. (2017). A Household Is Not a Person: Consistency of Pro-Environmental Behavior in Adult Couples and the Accuracy of Proxy-Reports. *Environment and Behavior*, *49*(6), 603–637. https://doi.org/10.1177/0013916516663796

Shi, Y., Yu, T., Liu, Q., Zhu, H., Li, F., & Wu, Y. (2020). An Approach of Electrical Load Profile Analysis Based on Time Series Data Mining. *IEEE Access*, *8*, 209915–209925. https://doi.org/10.1109/ACCESS.2020.3019698

Simionescu, M., Păuna, C. B., & Diaconescu, T. (2020). Renewable energy and economic performance in the context of the european green deal. *Energies*, *13*(23). https://doi.org/10.3390/en13236440

Singh, S., & Yassine, A. (2018). Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, *11*(2). https://doi.org/10.3390/en11020452

Song, K., Anderson, K., Lee, S. H., Raimi, K. T., & Sol Hart, P. (2020). Non-invasive behavioral reference group categorization considering temporal granularity and aggregation level of energy use data. *Energies*, *13*(14). https://doi.org/10.3390/en13143678

Testi, D., Franco, A., Conti, P., & Bartoli, C. (2020). Clustering of educational building load data for defining healthy and energy-efficient management solutions of integrated HVAC systems. *E3S Web of Conferences*, *197*. https://doi.org/10.1051/e3sconf/202019703001

Waddams Price, C., Brazier, K., & Wang, W. (2012). Objective and subjective measures of fuel poverty. *Energy Policy*, *49*, 33–39. https://doi.org/10.1016/j.enpol.2011.11.095

Wang, C., Du, Y., Li, H., Wallin, F., & Min, G. (2019). New methods for clustering district heating users based on consumption patterns. *Applied Energy*, *251*. https://doi.org/10.1016/j.apenergy.2019.113373

Wang, Z., & Hong, T. (2020). Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). *Energy and Buildings*, *224*. https://doi.org/10.1016/j.enbuild.2020.110299

Yan, L., & Liu, M. (2021). Predicting household air conditioners' on/off state considering occupants' preference diversity: A study in Chongqing, China. *Energy and Buildings*, *253*. https://doi.org/10.1016/j.enbuild.2021.111516

Yang, L., Wu, D., Xu, S., Zhang, G., & Cai, Y. (2018). Social-energy-aware user clustering for content sharing based on D2D multicast communications. *IEEE Access*, *6*, 36092–36104. https://doi.org/10.1109/ACCESS.2018.2849204

Yang, W., Shi, J., Li, S., Song, Z., Zhang, Z., & Chen, Z. (2022). A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. *Applied Energy*, *307*. https://doi.org/10.1016/j.apenergy.2021.118197

Yao, B., Xu, Y., Pang, Y., Jin, C., Tan, Z., Zhou, X., & Su, Y. (2018). Electricity consumption model analysis based on sparse principal components. *ICPRAM 2018 - Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, *2018-January*, 590–596. https://doi.org/10.5220/0006715405900596

Zaharia, M., Pătraşcu, A., Gogonea, M. R., Tănăsescu, A., & Popescu, C. (2017). A cluster design on the influence of energy taxation in shaping the new EU-28 economic paradigm. *Energies*, *10*(2). https://doi.org/10.3390/en10020257

Zatti, M., Gabba, M., Freschini, M., Rossi, M., Gambarotta, A., Morini, M., & Martelli, E. (2019). k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization. *Energy*, *181*, 1051–1063. https://doi.org/10.1016/j.energy.2019.05.044

Zhang, X., Ramírez-Mendiola, J. L., Li, M., & Guo, L. (2022). Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study. *Applied Energy*, *308*, 118335. https://doi.org/10.1016/j.apenergy.2021.118335

Zhong, S., Wang, X., Zhao, J., Li, W., Li, H., Wang, Y., Deng, S., & Zhu, J. (2021). Deep reinforcement learning framework for dynamic pricing demand response of regenerative electric heating. *Applied Energy*, *288*. https://doi.org/10.1016/j.apenergy.2021.116623

Zhou, Y., Sun, L., Hu, X., & Ma, L. (2021). Clustering and statistical analyses of electricity consumption for university dormitories: A case study from China. *Energy and Buildings*, *245*. https://doi.org/10.1016/j.enbuild.2021.110862

Appendix

Table 1 Summary of key information of reviewed articles

Author	Individual level	Collective level	Clustering variables	Number of Clusters	Clustering method	Dataset name and size	Region
(Cornelius *et al.*, 2014)	✓		Psychological	2	cluster-randomized controlled trial	high school students (n=165)	USA
(Reyna *et al.*, 2016)		✓	Sociodemographic, electricity usage, technical variables on building and appliance characteristics	5	max-p clustering	American community survey	LA and NY social and technical profiles
(Gabe-Thomas *et al.*, 2016)	✓		Energy usage	3	k-means	data of social residents housing from Exeter, UK	UK
(Jafari-Marandi *et al.*, 2016)		✓	Energy usage	4	self-organizing map based	monthly energy consumption data for 30 buildings	-
(Zaharia *et al.*, 2017)		✓	Psychological, energy usage	7	hierarchical clustering methodology	GDP of eu-28	EU-28 Member States
(Ma *et al.*, 2017)		✓	Energy usage	11	partitioning around medoids	3 years hourly heating energy usage 19 education buildings	Norway
(Fu *et al.*, 2018)	✓		Energy usage	5	fuzzy c-means (FCM) clustering algorithm	533 households	China
(Diao *et al.*, 2017)	✓		Energy usage	10	k-modes clustering and PNN	ATUS records	USA
(Knowles *et al.*, 2018)	✓		Energy usage	5	ward's minimum variance	13 households	1
(Yao *et al.*, 2018)	✓		Energy usage	2	k-means and affinity propagation	energy demand research project (EDRP) (aecom,2011)	UK

Author	Individual level	Collective level	Clustering variables	Number of Clusters	Clustering method	Dataset name and size	Region
(L. Yang *et al.*, 2018)	✓		Sociodemographic, energy usage	N/A	CH selection algorithm	synthetic dataset for number of users	-
(Boucher *et al.*, 2018)	✓		Sociodemographic	N/A	ordinary least squares	1670 zip codes in NY	USA
(Jiang *et al.*, 2018)	✓		Energy usage	13	fuzzy consumer categorization and k-means	US. non-residential consumers	USA
(Singh & Yassine, 2018)		✓	Energy usage	50	SVM	UK-dale and AMPDS2	UK
(Ma *et al.*, 2018)		✓	Energy usage	7	symbolic transformation	3 years hourly heating energy usage 19 education buildings	Norway
(An *et al.*, 2018)	✓		Energy usage	4	k-means and fuzzy clustering	301 apartments	Zhengzhou, China
(Beken *et al.*, 2019)		✓	Energy usage	2	k-means	2009-17 for 29 countries	30
(Schwarzinger, Bird, & Skjølsvold, 2019)	✓		Sociodemographic, energy usage		impact-based lifestyle research framework	ECHOES	31
(Caballero *et al.*, 2020)	✓		Energy usage	6	clustering-by-compatibility	200 prosumers real and synthetic date	-
(Lara Lück & Albert Moser, 2019)		✓	Sociodemographic, energy usage	N/A	k-means	installed PV units	Germany
(Calikus *et al.*, 2019b)	✓		Energy usage	2	k-shape clustering	1222 buildings	south of Sweden
(Ren, 2019)	✓		Energy usage	4	k-means	10 dwellings	Dutch community
(K. Li *et al.*, 2019)		✓	Energy usage	10	agglomerative hierarchical clustering	two university library buildings	University of Wollongong
(Lin *et al.*, 2019)	✓		Energy usage	10	k-means	power load data of American users	USA
(Ashouri *et al.*, 2019)		✓	Energy usage	2	k-means	80 buildings	Japan
(Buttitta *et al.*,	✓		Energy usage	5	k-means	UK national building stock	UK

Author	Individual level	Collective level	Clustering variables	Number of Clusters	Clustering method	Dataset name and size	Region
2019)							
(Gong *et al.*, 2019)		✓	Energy usage	5	GMM	provincial data on CO_2 and GDP	China
(Lu *et al.*, 2019)		✓	Energy usage	6	GMM	energy station in Tianjin	China
(Zatti *et al.*, 2019)		✓	Energy usage	5	k-MILP	the university campus case study	Italy
(Kennedy *et al.*, 2019)	✓		Energy usage	6	linear methods and k-means clustering	BBOXX'S customer	Rwanda and Kenya
(C. Wang *et al.*, 2019)	✓		Energy usage	5	GMM	Swedish dh company heat usage for 561 users	Sweden
(Krayem *et al.*, 2019)	✓		Energy usage, sociodemographic	4	k-mean	2311 buildings	Beirut, Lebanon
(P. H. Li *et al.*, 2020)		✓	Energy usage	5	k-means	600 pathways	UK
(Testi *et al.*, 2020)		✓	Energy usage	6	k-means	MERRA-2	worldwide
(Simionescu *et al.*, 2020)		✓	Energy usage	2	k-means	EU-28 countries	28
(Culaba *et al.*, 2020b)		✓	Energy usage	2	k-means and SVM	30 mixed-use buildings	OpenEI and NREL
(Shi *et al.*, 2020)		✓	Energy usage	4	k-means	UCI test dataset	University of California
(Z. Wang & Hong, 2020)		✓	Energy usage	19	k-means	the building data genome project	North America Europe
(Song *et al.*, 2020)	✓		Energy usage	7	k-means, hierarchical cluster analysis, SOM	2248 households	Holland, Michigan USA.
(Neto-Bradley *et al.*, 2021)	✓		Energy usage, sociodemographic	5	agglomerative clustering	23 households	Bangalore
(Creutzig *et al.*, 2021)		✓	Energy usage, sociodemo-	4	double-stacked expert, machine learning	121 165 publications	worldwide

Author	Individual level	Collective level	Clustering variables	Number of Clusters	Clustering method	Dataset name and size	Region
			graphic				
(Fang *et al.*, 2021)	✓		Psychological	7	piecewise symbolic spatial clustering	3117 records	China
(Khan *et al.*, 2021)		✓	Energy usage	6	k-means	four multi-storied residential buildings	Seoul, South Korea
(Zhou *et al.*, 2021)		✓	Energy usage	5	k-means	data of undergraduate and PhD dormitories	China
(Bogin *et al.*, 2021)	✓		Psychological	5	hierarchical cluster analysis	146 households	Beer Sheva, Israel
(Panchabikesan *et al.*, 2021)	✓		Energy usage	5	k-shape clustering	eight apartments	Lyon France
(Zhong *et al.*, 2021)	✓		Energy usage	2	k-means	1000 users	Tianjin, China
(Yan & Liu, 2021)		✓	Energy usage	3	k-means	1,274 ACS	Chongqing, China
(Mi *et al.*, 2021)	✓		Psychological	3	expectation maximization (EM) algorithm	4422 households	Xuzhou, Jiangsu, China
(W. Yang *et al.*, 2022)	✓		Energy usage	2	DBSCAN and k-means	actual load data of 71 single household	American Pecan Street Energy Project
(Zhang *et al.*, 2022)	✓		Energy usage	6	DLDA and AP	5566 households	London

PARTICIPANTS

ETH zürich · TEESlab University of Piraeus Research Center · TU Delft · UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh

University of Glasgow · JOANNEUM RESEARCH LIFE · Th!nk E · Utrecht University

Green Partners environmental consulting · Eko-cBecT · MISSIONS PUBLIQUES Bringing citizens into policy · HOLISTIC

University of Victoria

ENCLUDE project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 101022791

ENCLUDE
Energy Citizens for Inclusive Decarbonization

@encludeproject @encludeproject

www.encludeproject.eu