Politecnico di Torino
Università di Torino
Dipartimento Interateneo di Scienze, Progetto e Politiche del Territorio
Eccellenza MIUR 2018-2022
ScuDo
Scuola di Dottorato – Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Urban and Regional Development (37th Cycle)

Renewable Energy Communities:
A Preference Learning Approach to Evaluate Differential Participation in Cities

By

Giulio Cavana

******

Supervisors:
Prof. Marta Bottero, Supervisor
Prof. Cristina Becchio, Co-Supervisor
Prof. Giovanna Fancello, Co-Supervisor

Doctoral Examination Committee:
Prof. Ivan Blečić, Referee, University of Cagliari
Prof. Matteo Brunelli, Referee, University of Trento
Prof. Irene Pluchinotta, Referee, University College of London
Prof. Eleftherios Siskos, Referee, Technical University of Crete
Prof. Andrea Lanzini, Referee, Polytechnic of Turin

Politecnico di Torino
2025

Politecnico di Torino - Castello del Valentino – Viale Mattioli 39, 10125 - Torino, Italia
Ph.D. in URBAN AND REGIONAL DEVELOPMENT
Ph: +39 011.090. 6439
dottorato.d.ist@polito.it - http://dottorato.polito.it/urb

Dipartimento Interateneo di Scienze, Progetto e Politiche del Territorio
Eccellenza MIUR 2018-2022

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Giulio Cardullo

2025

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Alla mia famiglia, la mia mangrovia

Acknowledgment

Vorrei innanzitutto esprimere la mia gratitudine alle mie supervisors: Marta Bottero, Cristina Becchio e Giovanna Fanciello. La loro costante guida, la loro conoscenza e l'incoraggiamento ricevuto durante questo percorso sono stati fondamentali per la realizzazione di questa tesi.

Un ringraziamento speciale va ad Alexis Tsoukiàs per la sua enorme disponibilità e l'immenso bagaglio scientifico che è sempre stato pronto a condividere durante le innumerevoli occasioni in cui mi ha supportato nel lavoro di dottorato (e anche per le incredibili conversazioni sui viaggi in giro per il mondo).
Vorrei anche ringraziare il gruppo di lavoro con cui ho avuto la possibilità di collaborare in tanti progetti: Caterina, Giulia D., Vanessa, Federico, Sebastiano, Anthony, Marco V., Sara V., Giulia V., Ilaria, dai quali ho imparato tantissimo.

A huge thank you also goes to the friends I made during my PhD, the Sottopalco people, and especially the "White Truffles," for all the time we spent together and for always being there during the most complicated moments. I am always amazed at how fortunate I have been to meet so many wonderful people by chance at the same time and in the same place.
Grazie anche alle mie fantastiche coinquiline Glo, Paula, Carlotta, Andreha, Mati e Ale, e con loro Bert, Dario e Fra. Via Cernaia è stata casa anche e soprattutto grazie a loro.

Grazie a Sofia per tutto il suo affetto, per il tempo passato insieme, e anche per essere stata vicina nei momenti di sconforto… e grazie anche per tutta la pazienza che ha avuto.

Passando agli amici di Modena sono assolutamente da ringraziare Fillo, Fre, Vero, Cate, Stefi ed Ale. Anche se non essere li fisicamente limita le occasioni di vedersi, sanno sempre farsi sentire presenti (oltre che rimettere tutto in carreggiata quando serve…ognuno coi propri modi).

Infine (solo come posizione nell'elenco sia chiaro) ringrazio la mia famiglia tutta. In particolare, mia Madre e mio Padre, per il loro affetto e per tutti i sacrifici che hanno fatto (qualsiasi ringraziamento in forma scritta è riduttivo). Grazie a Carlo e Simone, la Chicca e la Sammy, e a quegli uragani chiamati Martina, Francesco,

Andrea e Federico (e tra qualche tempo anche Irene), tornare alla base è sempre una festa grazie a loro. Grazie anche ai nonni, quelli presenti, quelli che ci sono stati e quelli con cui non ci siamo incontrati per un soffio, e poi agli zii, e a Mauro, per essere sempre disponibile. (chiedo scusa se mi sono dilungato ma non l'avevo fatto nella tesi di magistrale quindi ho colto l'occasione).

Forse non sono i ringraziamenti più professionali che avrei potuto fare, ma è grazie a voi che questi anni sono stati occasione di crescita personale prima ancora che professionale.

Abstract

Urban areas are of particular significance in addressing the environmental impacts of human activity, particularly with regard to energy consumption and greenhouse gas emissions. As they account for over 70% of global energy-related CO_2 emissions, cities are central to climate and energy strategies. Within the European Union (EU), the pivotal function of urban areas in fostering decarbonization is exemplified by initiatives such as the "EU Mission: 100 Climate-Neutral and Smart Cities by 2030," which underscores inclusivity and citizen participation. This mission-oriented approach underscores the imperative for sustainability transitions to be socially inclusive, economically feasible, and environmentally ambitious.

Among urban sectors, the building sector is of particular significance due to its substantial decarbonization potential. Within the European Union, the construction sector accounts for approximately 40% of final energy consumption and 36% of energy-related carbon dioxide emissions. However, the adoption of clean energy at the urban scale continues to progress at a gradual pace, with regulatory, economic and behavioral factors acting as impediments to this process.

There is an emergent paradigm shift towards citizen-centered energy systems, in which individuals assume active roles in the generation and management of energy. This phenomenon is exemplified by the emergence of Renewable energy communities, as delineated in the EU Renewable Energy Directive (RED II). Renewable energy communities facilitate joint production, distribution, and governance of renewable energy by citizens, local governments, and small enterprises with an emphasis on the environmental, economic and social benefits that they could bring.

This research project explores the socio-demographic and performance-related factors influencing citizen participation in Renewable energy communities. In addition, it investigates the potential for their urban-scale diffusion based on preference models, spatial variations in participation, and financial mechanisms to align private actions with public sustainability goals.

To answer the above, a survey was conducted to collect data on citizens' willingness to participate in different Renewable energy community configurations and their preference rankings among these alternatives. In accordance with the respondents stated preferences, two models were developed: a moderated logistic regression model and an additive model using value functions estimated through a UTA method. These models were then subjected to a process of generalization through the utilization of supervised classification algorithms and subsequently linked to

respondent profiles defined by socio-demographic vectors. A synthetic population was generated for a selected case study area using statistical data. Finally, the share of the population likely to accept participation in a Renewable energy communities, as well as preferences for specific alternatives, was estimated.

The findings of the study indicate that six socio-demographic factors significantly affect individuals' propensity to participate in Renewable energy communities and their preferences regarding different Renewable energy communities performance attributes. The utilization of these variables determined the categorization of 144 distinct types of individuals, with each type being associated with the two distinct models: the logistic model for the estimation of participation likelihood and the additive model for the ranking of alternative options. The analysis estimated the acceptability levels of various Renewable energy communities configurations, revealing that the most environmentally effective option corresponded to the lowest potential acceptance rate. In order to enhance its acceptability and elevate its standing in relation to competing alternatives, it was estimated that a comprehensive financial support package amounting to €3.9 billion would be required at the urban scale. The study tested the possibility to apply a formalized tool to combine individual subjective information with objective characteristics (socio-demographic attributes) to evaluate and generalize individuals' system of values. The application to a real case study should be taken as a test bed to evaluate the potentiality of the method rather than providing quantitative results (especially due to the limitation in data collection). Finally, the method has been applied specifically to estimate the acceptability of Renewable energy communities but demonstrated promising potential for its application to other urban transformation challenges.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The European Union is committed to achieving a competitive economy with zero net greenhouse gas (GHG) emissions by 2050. This milestone was reaffirmed in the European Green Deal [1] in 2019, when it was clearly stated that with the current policies in place, the overall GHG reduction of the European Union would have reached only 60% in 2050 and would not be in line with its ambitious target. In this context, the integration of renewable energy, energy efficiency and sustainable solutions in all sectors should be strengthened.

Urban areas are considered strategic in the attempt to reduce the environmental impact of society. Indeed, urban areas account for approximately 75% of global primary energy consumption and a significant proportion of total GHG emissions [2]. Furthermore, as the majority of the world's population lives in cities [3], they are seen as central to climate change mitigation, and also as an appropriate arena in which to address some of the problems of our time. Indeed, several of the Sustainable Development Goals (SDGs) could find a direct translation in the attempt to achieve a just green transition at the urban level, from the simple creation of sustainable cities and communities (SDG 11), to the need to provide clean and affordable energy (SDG 7), to the commitment to climate action (SDG 13). Cities are also specifically targeted by European policies. For example, the EU Mission 100 climate neutral cities by 2030 - by and for the citizens [4] with the specific need to address Scope 1 (direct emissions from buildings, industry, transport, waste treatment, agriculture, forestry and other activities within city boundaries) and Scope 2 (GHG from grid-supplied energy) greenhouse gas emissions from cities [5].

Among Scope 1 emissions, reducing energy consumption and associated GHG emissions from the buildings sector has long been considered a pressing issue [6]. This sector is responsible for 40% of energy consumption and 36% of energy-related carbon dioxide (CO_2) emissions at European level [7]. The European Commission has been particularly prolific in developing policies targeting the building sector (specifically or as part of a broader perspective) [1,8], highlighting the emission reductions that can be achieved in the short term [9].

In particular, the Energy Performance of Buildings Directive (EPBD) [10,11] introduces a new definition for buildings aiming at a high standard of sustainability, namely the zero-emission building. As defined in the Directive, the Zero Emission Building is a very high-performance building with a very low energy demand, to be covered by energy from renewable sources at the local level, with a specific focus at the district or community scale.

The possibility of achieving a large use of renewable energy in the urban context has been widely discussed [12–15], and its role in the possible reduction of the environmental impact of buildings has been demonstrated [16,17]. Moreover, energy efficiency measures and electrification of consumption have been considered as the most promising way to achieve a decarbonized future for the sector [2,18–20], representing an active field of research [21].

Several tools have been introduced to raise public awareness regarding building efficiency, such as the Energy Performance Certificates [22] or the Green Building Certificates such as LEED, BREEAM, DGNB, to name just a few of the more well-known ones. In addition, some experiments [4] seek to act as lighthouses to stimulate debate on the benefits of a more sustainable urban environment.

The European Union is particularly clear on the need to involve civil society in the joint effort to reduce the negative impacts of the urban environment [7] and on the possible wider objectives that could be achieved through participatory frameworks [23].

Finally, in terms of scale, it has already been demonstrated that the adoption of cooperative behavior at a scale larger than that of the building, such as the sharing of locally produced energy [24] or the adoption of technological features such as energy storage systems [25,26], can generate further benefits than considering each building individually.

Aggregation of consumers in participatory schemes is not a new concept. Extending energy analysis to the neighborhood scale [24] to exploit the potential of different building typologies and consumption curves [27,28] has been explored using a range of conceptual frameworks representing different carbon neutrality targets and temporal discrepancies between energy production and consumption [29,30].

Currently, the transition to clean energy use at the urban scale is slow [31], while inherent technological issues remain unresolved, including the generation-consumption mismatch [32] and the non-programmability of RES [33]. To address the slow pace of this transition, citizen-led bottom-up initiatives [34] are gaining attention due to their ability to integrate citizens as active actors in the energy market [32,35,36], to increase awareness of energy consumption [37,38], and to improve individuals' "energy literacy" [39]. Among these, the European Union has recently introduced Renewable energy communities as legal entities to formalize and support the integration of citizens as active players in the energy market, institutionalizing cooperative forms of participation.

1.1.1 RECs and CECs

Recent developments in European legislation have given rise to two novel forms of energy-related associative entities, namely Renewable energy communities (RECs) and Citizens energy communities (CECs). RECs are defined by the recast of the EU Renewable energy directive (RED II) [40], while CECs are introduced by the EU Internal electricity market directive [3,41]. Both entities can engage in different activities along the energy supply chain, such as production, consumption, storage, sharing among members, and selling. RECs and CECs emphasize the change in citizens' role, from passive consumers to active prosumers, and the need to consider factors beyond the mere financial motivations in their implementation [42,43]. While exhibiting similarities, RECs and CECs also differ in their formulation. CEC is intended as a more general concept [44], encompassing the possibility to exploit non-RES [45,46], without geographical restriction, whereas RECs are limited to the production of energy only from RES, and by a proximity criterion [47,48], and are thus considered as "Community of place and interest" [49]. These differences substantiate the hypothesis that the RECs may align more directly with the decarbonization target of the urban building stock, especially considering the potential multifaceted positive impacts that the REC formulation is ought to bring to the local community in which it operates.

In accordance with RED II, EU Member States must establish an "enabling framework" with the aim of facilitating the formation of RECs and ensuring the inclusion of low-income and vulnerable consumers [50]. This framework is required to facilitate access to finance and information, as well as to promote collaboration with distribution operators (DSO) [40]. It has been asserted that the correspondence of RECs impacts with the outcomes and benefits envisaged by the EU directives is contingent on the attractiveness and coherence of the enabling framework [48]. Furthermore, consumers' willingness to participate has been stressed as fundamental in the deployment of RECs [51], as well as their social acceptance [37]. In this direction, the enabling framework should consider the motivations behind actors' decision to participate or not, and their potential incapability to do so.

In the broader context of community energy [31], numerous studies have highlighted the significance of both external (e.g. costs, policies, energy models) and internal factors (e.g. financial and attitudinal) in influencing citizen participation [52–54]. Consequently, a thorough examination of the motivations, and external influences that drive the adoption of RECs is paramount, in conjunction with technical design considerations [54,55]. However, there is a paucity of approaches to evaluate how motivations vary across consumers [56,57], as well as a lack of a comprehensive evaluation framework able to consider the different potential benefits brought by the implementation of a REC [44,58,59].

RECs have been considered aligned with several of the United Nations' Social development goals as outlined in the Agenda 2030 (affordable and clean energy, sustainable cities and communities, responsible consumption and production, climate action) [3], suggesting their capability to confer a range of benefits to various stakeholders. From an environmental perspective, RECs may encourage the acceptance of RES generation in urban areas [46,60], thereby contributing to the achievement of national decarbonization targets [58], reducing GHG emissions, and implementing actions to mitigate the impact of cities [42,46]. Furthermore, RECs might increase energy security and independence [61–64], and enhance territorial resiliency [28]. They can provide flexibility services to the main grid [36,65,66], offsetting violations in voltage and capacity limits [67], and reduce its necessity for upgrades [68,69]. In urban environments, RECs could also promote energy efficiency at building level, decreasing energy consumption and emissions [11,42,51], enhancing energy self-sufficiency [34], and overcoming technical constraint at single building level [70,71]. The economic sphere, as well, is affected by the implementation of RECs. New jobs could be created [60,72,73], and market opportunities may emerge, offering prospects for profit-driven operators. Furthermore, members could benefit from energy bills savings and energy self-consumption.

In addition to the economic and environmental advantages, both the European legislator and the scientific community place significant emphasis on the positive social impacts associated with the implementation of RECs. In conjunction with the promotion of citizen engagement and innovative forms of participation [3], RECs have been regarded as capable of combating energy poverty [3,46,47,74], with the ability to reduce consumption and facilitate affordable energy tariffs [28,75,76] for demographic groups that are typically excluded from renewable energy investments and face split incentives issues [48,77]. Moreover, it has been documented that the establishment of energy communities in socio-economically disadvantaged and mistrusted neighborhood has been associated with an enhancement in the civic sense of residents [78], leading to a reduction in the stigma experienced by marginalized groups [79].

Notwithstanding the aforementioned benefits, there are several challenges that may emerge if the impacts of the implementation of RECs is not given full consideration. This calls for a need to guide their implementation both from a design [76] and a policy point of view [63]. For instance, the wide spread of distributed generation could bring issues to the stability of the energy system [29,32,36] with excessive stress on grid components determined by energy flux inversion due to generation-consumption mismatch [80]. Furthermore, there might be a misalignment of goals between RECs and other operators [58], determining potentially negative impacts of the community itself [75]. For instance, private members might target profits maximization, the DSO might value the avoidance of grid update necessity, while local authorities might be interested in the social and environmental benefits generated by RES upscaling [64]. The presence of such potentially conflicting goals calls for a concertation of actions to avert adverse outcomes for specific actors, with the possibility to evaluate the introduction of policies and price signals that could direct a coordinated deployment of RES [81]. Finally, numerous barriers may be faced by citizens willing to participate in a REC, including space limitation for distributed generation [82], high investment cost, lack of access to finance [50,58], issues related to socio-economic factors [31,83] knowledge [44] and regulation [32]. These factors might result in an uneven distribution of RECs across cities and socio-economic groups [48]. This uneven distribution argument could be further substantiated by the fact that, in the process of designing RECs, a variable that could be optimized is the selection of participants so to achieve higher financial results [84]. The focus on maximizing profits in the implementation of RECs may result in a neglect of areas where greater social, economic, and environmental benefits could be achieved [46]. This could lead to a harmful polarization effect between classes residing in sustainable areas and disadvantaged classes living in more polluted areas [85].

In light of the above, and given the emphasis that the European policy framework places on the role that RECs could play in the decarbonization pathway for the built environment, assessing the impact of such a new legal entity is an open and fundamental question [23,86]. In particular, as will be shown in Chapter 2, the scientific contribution analyzing the combined effect of the reduction of energy consumption due to energy efficiency measures and the implementation of RECs at the urban level is still scarce. Moreover, the rules of behavior of actors involved or not in an initiative at the community level represent an interesting avenue of research [2,53,87]. In this respect, understanding the conditions for setting up a REC could have important policy implications.

The analysis of the economic and financial drivers that motivate an actor to participate could be used to analyze the possible expected penetration of RECs in the urban context. In particular, the preference of potential members for different performance indicators of alternative scenarios (such as minimizing costs or maximizing emission reductions), and the trade-offs between them, could lead to the implementation of sub-optimal solutions in terms of the potential environmental benefits achievable compared to interventions aimed at minimizing the environmental impact of the building sector.

Finally, by analyzing the different distribution of preferences for participation in an energy community among different socio-economic groups, and the distribution of these groups at the urban level, it would be possible to identify the areas where the establishment of a REC does not meet these preferences, and to consider possible modifications to the enabling framework to promote their adoption.

1.2 Objectives and Research questions

The present work is based on the contextual background outlined previously and aims to investigate the extent to which participation in a REC could become widespread among urban residents. It is worth noting that the research is specifically tailored to achieve this objective, but the method used to answer the questions could also be adapted to the different cases of urban interventions in which the assessment of residents' preferences and acceptance of urban transformation may be required (and it could be argued that this is the case in most interventions aimed at transforming the urban environment). It is also important to stress that the application of the method to the specific case study has to be considered as a test to prove the functioning of the method, and the results should not be taken as a precise quantitative estimation.

The aim of this research could be summarized in four main bullet points:

- determine citizens' preferences in the participation to the energy transition of the building stock, depending on socio-demographic factors;
- evaluate the potential performance results of Renewable energy communities at the city level, depending on the solution preferred by the inhabitants.
- evaluate the possible intervention of third parties (municipalities, ESCOs) to promote supporting measures that could increase stakeholder participation and align the private initiative with public objectives.

The final objective of this work will be to contribute to the body of research aiming at identifying the optimized actions to be taken in order to achieve a more sustainable building stock. In particular, it will explore the possibility of achieving a better allocation of financial resources and better environmental outcomes through stakeholder collaboration. Another objective will be to take into account the attitudes of citizens towards this type of issues in order to better assess their real feasibility and scalability, not only from a technical point of view, but also from a broader socio-environmental-economic perspective.

1.2.1 Specific research questions

The bullet points indicated as objective of the research in the previous subsection can be translated into the research questions that the present work will tackle.

Q1. What are the socio-demographic characteristics of individuals and the performance attributes of the potential intervention scenarios that influence inhabitants' decision to accept to participate in a Renewable energy community?

The answer to this question will aim to evaluate which are the parameters and thresholds that different individuals take into consideration when confronted with the decision to engage or not in the participation to a REC. Most of the study that evaluate the adoption of energy efficiency measures uses different parameters such as financial (Investment Costs, Pay-back Period, etc.) or environmental ones (avoided GHG emissions, self-sufficiency, etc.) but the trade-offs between those parameters and, especially, the decision model followed by different actors have to be further investigated in order to evaluate to which extent it would be possible to expect the private sector to take actions toward the achievement of a more sustainable building stock, and which is the extension of the acceptability space in terms of RECs performances.

Q2. To which extent could different renewable energy communities proliferate at the city scale, and which are the most probable solutions to be implemented based on the inhabitants' preference models?

Building on the previous research question (Q1), different individuals might evaluate in different ways the possibility to participate in a REC depending on their personal characteristics and the performance that different solutions could achieve. By analyzing how the parameters potentially influencing the willingness to participate in a REC solution characterize the population in a certain area, it would be possible to estimate the probable level of acceptability of different solutions at the local scale, thus evaluating the possible outcomes that the widespread adoption of RECs at the city scale might have.

Q3. What could be the expected variation in the participation to different configurations of renewable energy communities among the inhabitants in different parts of the city?

Considering the different parts of the urban environment, it might be reasonable that different RECs would achieve various level of performances in different areas of intervention. Furthermore, different actors might have different level of acceptance in taking the decision to participate in an Energy community or not depending on its performances. A consequence of the latter could be that some areas at the urban level will be less likely to be involved in the process of sustainable transition. This could be determined by a series of factors both technical (limited availability of suitable areas for the generation of renewables, or limited possibilities of energy consumption reduction) and socio-demographic (related to the specific personal inclinations toward the participation in such schemes). By means of a spatialization of the characteristics influencing the probability to participate in a REC into thematic maps, the policy-maker could individuate such critical areas in which the current policy framework is less likely to result in the adoption of energy efficiency measures.

Q4. What are the financial supporting interventions from third-parties (e.g., public entities, ESCOs) that could foster the alignment of the private initiative with the public perspective to maximize the positive impacts that the implementation of RECs could have?

The previous analysis would likely raise the issue that some areas would see a low potential level of participation in the process to achieve a more sustainable use of energy at the building scale due to hindering conditions that would restrain the stakeholders from taking actions. Another possible issue might be the possibility that different specific preference toward the potential performance that a REC should have to become acceptable by the majority of inhabitants might result in the adoption of alternatives not aligned with the public point of view (as it could be the minimization of the environmental impact of the energy use at the building scale). This question would address this shortcoming by highlighting the specific actions and incentives that the local policy-maker could put in place in order to increase the desirability of specific interventions. In particular, the rules that the different stakeholders follow in their decision process, depending also on their socio-economic characteristics (socio-demographic profiles), could be used to analyze the different strategies to support intervention scenarios more appealing to a wider number of individuals in order to direct their actions toward more public oriented goals.

1.2.2 The thesis in a glimpse

The research presented in this thesis is structured in six chapters (Figure 1). In this first chapter, the context in which the research is grounded and the aims of the thesis have been described. The second chapter provides a literature review of the academic production on RECs. This literature review presents the common approaches to assess the feasibility of REC implementations, together with a meta-analysis of a sub-sample of documents presenting the analysis of Italian case studies. The latter was instrumental in analyzing the set of attributes best suited to describe RECs and in assessing the homogeneity of the criteria used. The meta-analysis was also instrumental to identify a number of potential alternatives for RECs interventions to be used in a public survey administered to a sample of the population. In particular, this chapter resulted in an already published paper [88]. The third chapter presents the method used in the thesis. The proposed method approximate the preference models of a sample of individuals towards the participation in REC alternatives, and generalize these models to the entire urban population, pivoting on socio-demographic vectors of attributes. In Chapter 4 and 5, the application of the method to the case study of the city of Turin is presented. In the fourth chapter, the individuals' decision models are approximated and generalized based on socio-demographic characteristics. In the same chapter the population of the city of Turin is characterized using these socio-demographic characteristics to generate a synthetic population. In the fifth chapter, the match between the preferences of the population and the alternative performance of RECs is discussed in order to evaluate the different options that are more likely to be implemented, and different possible perspectives (i.e. public and private) are compared. In the sixth and final chapter, the conclusions are drawn from the application of the method to the case study and possible avenues for further research are outlined.

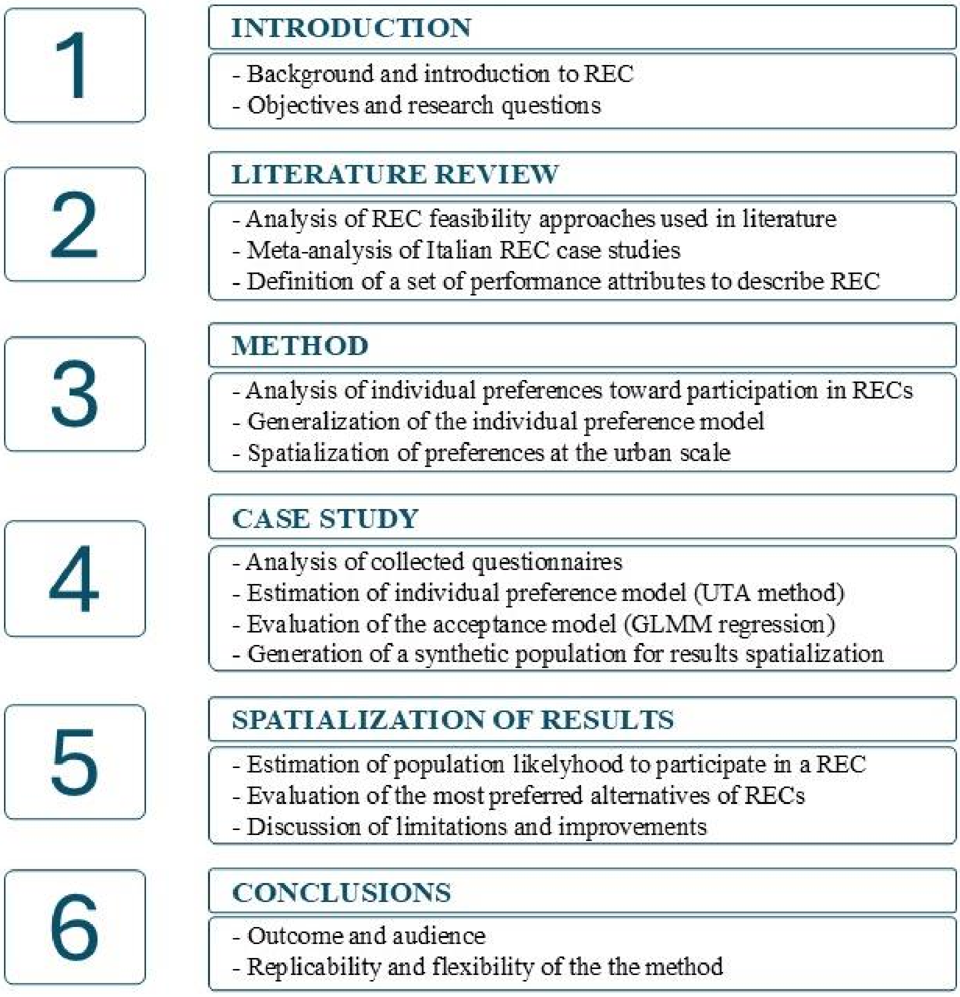


Figure 1 Overview of the structure of the thesis

It is worth noting that the proposed method has been designed to assess the acceptability of different REC alternatives at the urban scale. A potential decision-maker interested in assessing individuals' preferences for specific other forms of urban intervention could use the same method, with only minor adjustments in the procedure adopted to collect individual preferences (i.e., adapting the alternatives between which the surveyed respondent could choose to the specific case under investigation; see Appendix A for a preliminary study on assessing inhabitants preferences to gather insights regarding the design of a new urban park in the city of Turin).

Chapter 2

Feasibility analysis of RECs

In this chapter, a literature review on RECs is presented [88]. The objective of that is to map how the feasibility of RECs is evaluated by different scientific approaches, from both a technical and behavioral perspective, in order to evaluate potential cross-contaminations among sectors. Furthermore, a meta-analysis of the Italian subset of case studies is performed to evaluate inconsistencies among the indicator used, and a homogenized set of KPIs is proposed. Finally, potential hindering and promoting factors in the constitution and governance of RECs are examined, those being usually neglected in the predominantly technical feasibility approaches.

The remaining of the section is organized as follows: Section 2.1 presents the methodology; Section 2.2 classifies the reviewed articles based on their approaches and main goals; Section 2.3 analyses the nexus between energy generation and consumption reduction; Section 2.4 presents insights on the governance dynamics of RECs; and Section 2.5 compares the subset of Italian case studies and proposes a set of KPIs. Finally, Section 2.6 draws conclusions on the lesson learnt.

2.1 Methodology

In general, there are three main approaches to conduct a literature reviews: a systematic, a semi-systematic, and an integrative one. Among these, systematic reviews are the most accurate and rigorous approaches collecting the most relevant information related to a narrow research question, usually with the aim of directly comparing results by meta-analysis. Systematic literature reviews allow for transparency and replicability of the study. Several frameworks and guidelines are available for conducting them, more or less tailored for specific disciplinary fields (e.g., the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, the Cochrane Handbook, the Meta-analysis Of Observational Studies in Epidemiology, the Joanna Briggs Institute standard, the Enhancing Transparency in Reporting the Synthesis of Qualitative Research, the Reporting Standards for

Systematic Evidence Syntheses). Among these alternatives, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was adopted in this work, given its widespread use and broad acceptance across research communities. The PRISMA framework is particularly helpful tool in reporting the selection process of articles retained and discharged, and in guiding the process to perform meta-analysis, thus guaranteeing a high level of transparency in the collected and reported data.

The research was conducted on the Scopus database. The primary keyword employed was “energy community”, accompanied by a series of geographical constraints keywords, including “urban”, “city”, “neighborhood”, “district”, “block”, and “municipality”. A further limitation was imposed to include only journal papers in the English language, yielding a total of 349 documents. Figure 2 presents the temporal distribution of the selected papers. The introduction of the European directives between 2018 and 2019 is evident in the number of academic products on the topic, especially in the European economic area countries. For this reason, only papers published after 2018 have been further analyzed.

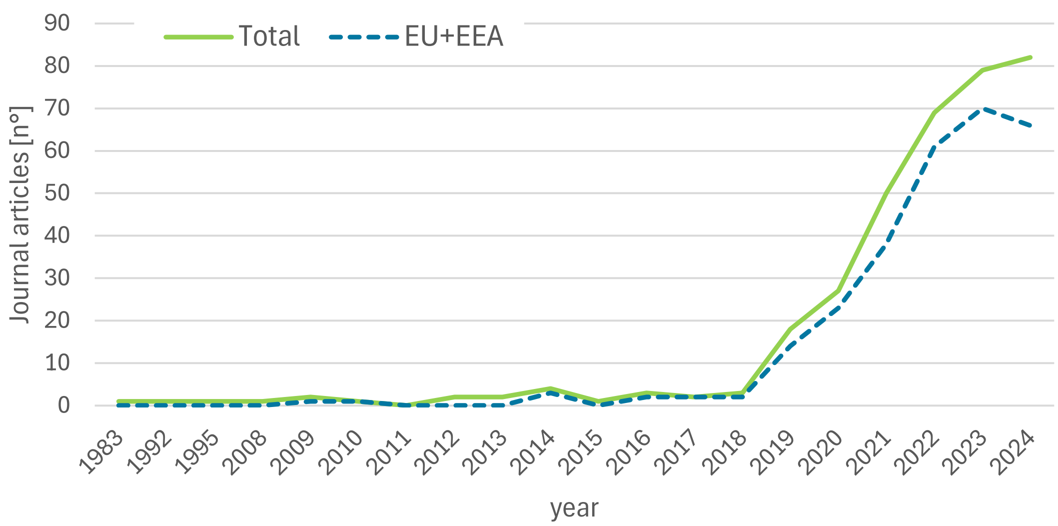


Figure 2 Publication trend of the collected articles worldwide, and in the European+ European economic area (EEA). The increase in the number of paper produced after 2018 is evident.

It is important to underline that the decision to prefer the term “energy community” over “Renewable energy community” is primarily due to the fact that a broad array of appellations has been employed in literature to define groups of individuals engaged in energy-oriented endeavors, not necessarily constrained to financial motivations [29,30,42]. Moreover, the utilization of the terminology introduced by the European Union would have likely constituted a Eurocentric limitation. Consequently, a number of studies that refer to configurations not precisely aligned with the definition of REC have been included in the analysis, for example those from the community energy field.

A second skim of documents has been conducted based on article title and abstract, including only contribution considering case studies, excluding those focusing on technical aspects of RECs operation (such as. blockchain technologies or energy exchange management with the grid), or those focusing on technical analysis of grid components. The resulting 179 documents have been further analyzed according to the goals set for this contribution. Among these, 129 were considered relevant (thus with 50 documents further excluded). The flow diagram used for the selection of articles is presented in Figure 3. It is worth highlighting that while all the 129 papers were used in the analysis, only those referring to Italian case studies were used for the “Scenario reconstruction” sub-section.

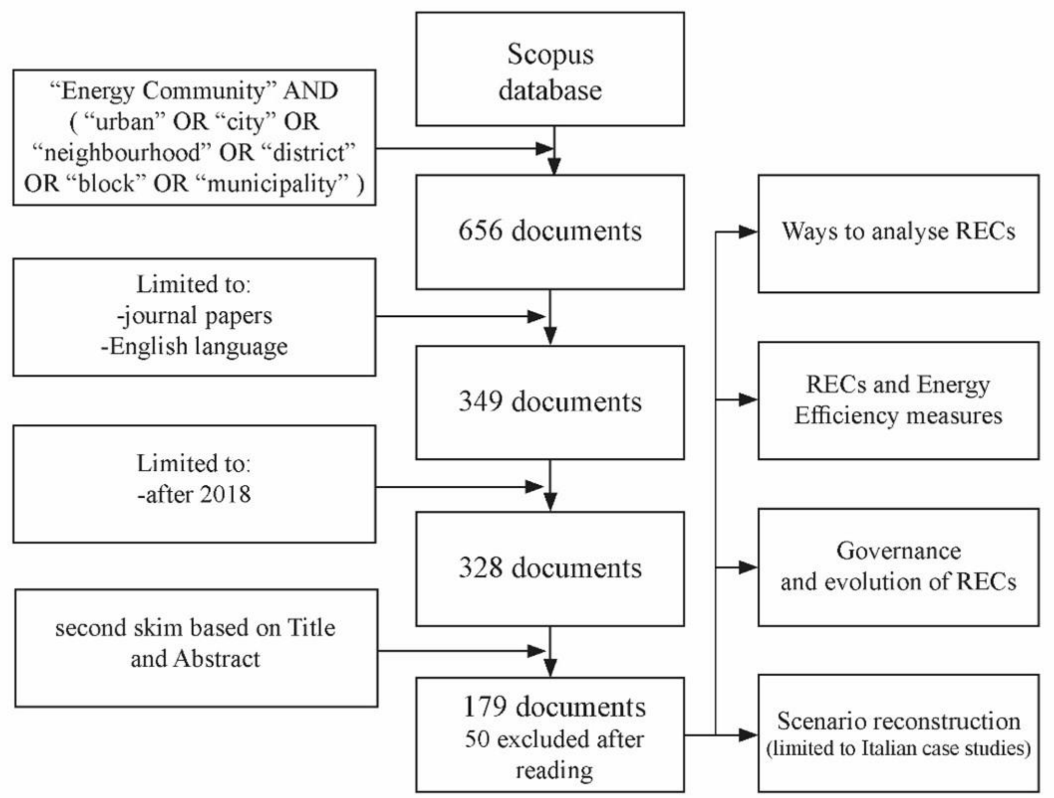


Figure 3 Paper selection flow of the literature review.

2.2 Feasibility analysis of a REC

With regard to the feasibility and potential establishment of RECs within an urban environment, three primary approaches have been identified, and they are described in this section. The proposed clusters do not refer to specific sub-fields or disciplines; instead, they are defined according to the similarities in the viewpoint used to analyse REC formation, as well as citizens' attitudes towards them. The aforementioned clusters are defined in a loose manner as follows: (i) "Scenario comparison", (ii) "Portfolio optimization", and (iii) "Behaviour analysis". The first two clusters are characterised by the aim of defining the most suitable REC layout for specific areas, evaluating the availability of renewable sources, and examining alternative technologies and configurations. The first cluster is based on a discrete space of solutions (scenarios) and considers different KPIs to support decision-makers in selecting the preferred solution. The second cluster employs optimisation models operating on a continuous space of solutions to identify the optimal portfolio of technologies in accordance with pre-determined objectives. A degree of overlap can be identified between the two clusters in studies that compare different scenarios following initial optimisation (post-optimal comparison). In distinction to the previous two groups, the "Behaviour analysis" cluster seeks to identify individuals' motivators and to model their propensity to participate in RECs. The analysis of these three clusters is focused on identifying the specificity of research goals, with the aim of understanding the potential for the cross-sectorial approach to RECs analysis [89].

2.2.1 Scenario comparison

The first cluster groups studies that aim to compare different alternatives of intervention for the constitution of a REC. This approach is particularly targeted to support decision makers when potentially conflicting or incommensurable performance indicators are present [90,91]. The production in this cluster includes several works that have evaluated alternative technology combinations to address district/municipality energy demand [30,33,92,93], the introduction of energy efficiency measures [94–97], alternative system controlling strategies [33,98], as well as REC size and configuration options [32,50,97,99]. A growing body of research has focused on the benefit associated with synergies among end-users' aggregating consumption patterns [27,66], also leveraging on residual energy such as from industrial processes [100]. Another line of research focuses on the potential of alternative Power-to-X strategies [29,30,92,101] such as the integration of electric (PtP), thermal (PtH), vehicles (PtV), and hydrogen (PtG) storage strategies of locally produced energy surplus [29,92]. Finally, several works focus on the relations between distributed generation and the impact that it could have on the main grid.

Table 1 provides an overview of the indicators used in the selected studies to evaluate the competing alternatives, and shows the combinations of KPIs in the four typically evaluated dimensions (i.e. economy, energy, environment, and social).

Table 1 overview of indicators used by the papers in the "Scenario building" cluster.

	Economy							**Energy**							**Env.**		**Social**		
Source	NPV	PBP	IRR	Investment cost	Cost & benefits	LCoE	Technical complexity	SSR	SCR	Energy generation	Grid interaction	Energy consumption	Seasonal COP	Seasonal EER	CO_2 emissions	Embodied carbon	"10%" indicator	Comfort level	EPHI
[93]	X	X																	
[102]	X	X	X		X			X	X						X				
[32]	X							X			X								
[33]	X				X			X	X			X			X				
[27]	X		X		X			X	X						X				
[29]					X				X										
[30]	X	X						X	X						X				
[47]*																			
[50]	X				X			X	X						X				
[60]								X	X		X								
[64]	X	X		X											X				
[65]*																			
[66]								X	X	X		X							
[71]	X	X	X					X	X						X				
[74]		X						X	X			X			X		X		
[76]	X	X	X												X				
[79]		X										X							
[92]	X	X			X	X		X				X			X				
[94]	X	X						X				X			X	X			

[95]	X		X						X						X				
[96]	X							X	X						X				
[97]					X			X	X						X				
[98]								X	X			X							
[103]							X			X								X	
[99]		X			X			X							X				
[101]					X	X		X	X						X				
[104]	X	X	X		X										X				
[105]								X	X			X	X	X					
[106]	X	X			X														
[107]									X			X			X				
[108]*																			
[109]*	X							X	X						X				X

* case studies for which a first optimization was performed (post-optimal comparison)

Among the evaluated papers, the economic dimension is the most analysed one, with an array of indicators employed, including payback period (PBP), net present value (NPV), internal rate of return (IRR), levelized cost of energy (LCoE), benefit/cost ratio, and annual cost. In addition, the energy dimension, typically evaluated through the self-sufficiency ratio (SSR) and the self-consumption ratio (SCR), is considered in more than half of the cases, in accordance with the necessity for RECs to deliver benefits that extend beyond financial ones. With respect to the multi-objective nature of RECs, numerous studies utilise a combination of indicators, with 15 documents evaluating REC feasibility from environmental, economic and energy perspectives. As Gjorgievski et al. [58] previously observed, the social dimension is rarely addressed in REC analyses, with only a few studies attempting to factor it into the evaluation framework. For instance, Ceglia et al. [74] analysed the effect of the establishment of a REC on the “10%” indicator usually used to quantify energy poverty. Cutore et al. [47] propose the energy poverty help indicator (EPHI), to assess the number of families in energy poverty conditions benefitting from the distribution of revenues from energy surplus. In this article, the dependency of energy poverty alleviation performance on members’ goals is used to raise the argument regarding potentially conflicting outcomes of the proposed REC.

These initial results underscore the multidimensional and multi-actor nature of RECs, marked by the existence of incomparable KPIs and a multitude of perspectives in the evaluation of their impacts. Decision-making problems of this nature are specifically addressed by multi-criteria analysis methods, intended to rank alternatives and to prioritize criteria. Among the analysed studies, Efthymiou et al. [76] combined multi-criteria and SWOT analysis to support the selection of the layout of a municipality-led REC, together with the most appropriate business model to be implemented. Sibilla and Abanda [103] utilised PROMETHEE method to prioritize retrofit alternatives for a community of school buildings, against qualitative and quantitative indicators. Torabi et al. [51] assess retrofit scenarios of an educational complex against environmental, economic, technical, and social criteria, being the latter the visual quality after intervention.

Three principal limitations can be identified regarding the set of KPIs used in this first cluster. Firstly, the selection of KPIs varies depending on the focus of each study (municipality, DSO, or inhabitants' perspective). Secondly, the conceptualization of indicators is problematic. For instance, the social impact represented by energy poverty alleviation is related to the internal redistribution of revenues or incentives granted, thus only representing a different definition of an economic aspect. Finally, certain indicators pose problems in their calculation, as they are expressed in a qualitative manner [47,103], or in a scenario-dependent way. For instance, the calculation of SSR is sensitive to energy vector and final uses considered (e.g. accounting only for electric consumption or including also thermal loads in case of the introduction of heat pumps).

2.2.2 Portfolio optimization

The second approach involves the analysis of continuous spaces of solutions to select the optimal portfolio of technologies by minimising or maximising target performance. A range of optimisation techniques are available, with the most widely used being Mixed integer linear programming (MILP) [68]. An optimisation process comprises three main components: (i) variables, (ii) constraints, and (iii) objective functions [110]. The first component, “variables”, refers to the alternative technologies to be selected, taking a binary (absence/presence of a specific technology), or continuous form (sizing of a technology) [108]. Their specification allows for the optimisation of the displacement of DES in the most suitable locations [111], the evaluation of hybrid energy systems [65,112], and the enhancement of the complementarity of energy production and storage systems [48]. The second component, “constrains”, represents the boundaries of the space of solutions, which can be categorised as either internal (e.g., equipment sizes, availability of space for installation) or external to the portfolio (e.g., limitation to thresholds violations such as transformer overload, or comfort degradation [68,80].

Finally, the objective function represents the performance to be either minimised or maximised.

The selection of the optimization function is indicative of the objectives and perspective of the decision-maker, as evidenced by the predominance of studies that target the minimisation or maximisation of economic costs and benefits for REC members. Such studies generally account for investment (CAPEX) and operational (OPEX) costs or utilise economic indicators such as NPV or PBP. For example, Monsberger et al. [113] utilise MILP to investigate the optimal REC layout, varying the objective functions according to the business model selected by a simulated initial investor. Other studies assume the DSO's viewpoint. For instances, Simoiu et al. [70] set an objective function that minimises the net-energy exchanged with the grid, while Delarestaghi et al. [69] optimise grid reinforcement intervention and consumers' investments by minimising investment and operational costs for both utility and consumers. Liu et al. [114] introduce a penalty cost to account for the extra burden to the grid caused by off/on-peak energy import/export.

In several cases, the formulation of the optimisation problem considers also the environmental and energy dimensions while selecting the technology portfolio, with this attempt being performed in three possible ways [115]. The first approach (i) involves converting GHG emissions into economic terms and minimising the total cost [65] in the target function formulation. The second approach, (ii) referred to as epsilon constraint method, allows the required performance to be specified as a further constraint. This approach has been used to guarantee a minimum of primary energy savings [116] and carbon emission reduction [115], as well as to constrain SSR and SCR while minimising costs [28]. Other works consider the impacts of RECs on the grid by applying a constraint to peak power exchange [117], and to energy surplus fed into the grid [80]. Finally, a third approach (iii) involves the introduction of an additional objective function, transforming the optimization problem into a multi-objective one [59], often presented as n-dimensional Pareto fronts [73,112,118].

The Pareto front representation of solutions enables decision-makers to specify their trade-offs concerning different performance levels, allowing for weak sustainability compensatory mechanisms among criteria [118,119]. In such a way, different criteria can be perceived as not equally important by the decision-maker, who can assign them different weights in the process of selecting among solutions. This latter approach poses several questions related to subjectivity and the sensitivity of outcomes to weights elicitation [120]. Another approach is represented by the minimum distance to utopia. This method identifies a hypothetical alternative that optimises all evaluation criteria simultaneously (corresponding to the origin of the axis in a Cartesian representation of solutions) and selects the closest attainable solution in Euclidean terms from that utopic unreachable point [119]. Among these studies, Liu et al.[114] produce a three-dimensional Pareto front considering SSR, SCR, and the time-of-use grid penalty cost (thus considering grid relief potential provided by the REC). Ascione et al. [121] adopt this approach to discuss three suboptimal points in the solution space representing the preferences of private individuals (economic driven), municipality preferences (environmental driven), and the minimum distance from utopia as a mediated one.

In general, the mathematical definition of the optimization problem requires the prioritization of one perspective over others by specifying in advance the target performance to be optimized. Some studies overcome this limitation by first optimizing the technology portfolio and then comparing the obtained scenarios against sets of indicators (post optimal selection) [47,65,108,109].

Table 2 presents an overview of the parameters specified in the optimisation functions utilised in the analysed documents, as well as the combinatory strategy employed to address the multiple dimensions of the problem. In particular, the variables that are monetized when accounted in the objective function are marked with Xm, while the variable for which a minimum or maximum value to be achieved (ϵ-constrained) are marked with Xϵ in the table. It is noteworthy that economic indicators predominate, and there is a relatively frequent attempt to introduce environmental and energy dimensions into the optimisation functions. Only a few studies evaluate these two dimensions independently, and only one introduces prosumers' comfort by monetising disutility due to the application of demand-response strategies.

Table 2 overview of the papers in the “Portfolio optimization” cluster. Xm represents monetized variables, while Xϵ the ϵ-constrained variables

	CAPEX	OPEX	peak use costs	SCR	SSR	energy balance	grid impacts	environmental impact	disutility cost	target min/max	Combinatorial strategy	Source
Single objective optimization	X	X								max (NPV)	-	[62]
		X	X							min		[63]
		X						Xm		min	CO_2 monetization	[37]
				X						max	-	[37]
	X	X								min	-	[68]
	X	X	X						Xm	min	Users’ disutility monetization	[69]
							X			min		[70]
	X	X	X							min		[100]
	X	X								min	-	[110]
	X	X								max (NPV)	-	[111]
	X	X								min	-	[122]
	X	X								min		[123]
	X	X	X				Xm	Xm		min	CO_2 monetization, slack penalties	[124]
ϵ-constraint method	X	X		Xϵ	Xϵ					min	ϵ-constraint (SCR, SSR)	[28]
	X	X					Xϵ			min	ϵ-constraint (positive net load)	[80]
	X, Xϵ	X, Xϵ								min	ϵ-constraint (alternatively CAPEX and OPEX)	[81]
	X	X, Xϵ								max (gains)	ϵ-constraint (residents’ energy cost)	[113]
	X, Xϵ	X					Xϵ	Xϵ		min	ϵ-constraint (maximum energy withdraw, CO_2, breakable investment constraint)	[115]
	X	X				Xϵ				min	ϵ-constraint (primary energy consumption)	[116]
	X	X	X				Xϵ			min	ϵ-constraint (peak power exchange)	[117]

Post-optimal comparison	X	X							max (NPV)	Post-optimal evaluation (SCR, SSR, CO_2 emissions, EPHI)	[47]
	X	X						Xm	min	CO_2 monetization, Post-optimal evaluation (energy self-consumed, purchased, and used; grid capacity, storage capacity, CO_2 emissions)	[65]
	X	X							min	Post-optimal evaluation (LCoE, SSR)	[108]
	X	X							max (NPV)	Post-optimal evaluation (SCR, SSR, CO_2 emissions, EPHI)	[109]
Multi objective optimization	X	X					X		min	Multi-objective cost-grid sustainability index	[59]
	X	X						X	min min	Multi-objective NPC-ReCiPe indicator, post-optimal evaluation (renewable energy fraction, Energy Performance Indicators, economic and environmental PBP, employment opportunities)	[73]
	X	X				X			min min	Multi-objective Life-cycle costs-imported electricity	[112]
			X	X	X				min max max	Multi-objective grid penalty-SCR-SSR (Post-optimal evaluation NPV, CO_2 emissions)	[114]
	X	X						X, Xm Xϵ	min min	Multi-objective cost-emissions (further ϵ-constraint on CO_2, and introduction of carbon taxes)	[118]
	X	X						X	min min	Multi-objective global cost-CO_2 emissions	[121]

A potential limitation of this cluster could be seen in the arbitrary definition of the geographical boundaries of the case studies examined. The optimisation of portfolios in a certain area might result in a sub-optimal solution if the investigation area is expanded, especially in cases where a regulatory framework imposes limits on the geographical configuration of RECs (e.g., the Italian context). Interesting approaches to address this potential issue are those that attempt to evaluate the matching between energy demand and consumption at the city scale using GIS tools [125,126]. Furthermore, it is important to consider the evolution of RECs over time, as the optimised solution might limit the desirability of including new members after the first implementation of the community [62]. Additionally, the evolution of consumption pattern (i.e. due to electrification of consumption and energy demand reduction thanks to the introduction of energy efficiency measures) may modify the boundary conditions on which the optimisation is based.

2.2.3 Behavior analysis

The final cluster of studies encompasses papers that concentrate on the motivations behind inhabitants' participation in a REC, in addition to the potential evolution of REC formation. With regard to the evaluation of drivers and motivation in REC participation, real-word case studies have been utilised as sources of information. By conducting interviews with members of both existing and emerging energy communities, several authors have noted a preference for independence from the distribution grid [127], with a tendency towards autarkic discourses, even when the proposed solution was not economically or environmentally advantageous [128]. This regain of control has been posited as a means to tentatively mitigate various forms of uncertainties associated with geopolitical conflicts, escalating energy prices, and environmental challenges [129,130]. The environmental factor has been identified too as a motivator for engagement in both existing communities [31] and experimental simulations [131], with the presence of grants, subsidies, and other forms of incentives being recognised as a prerequisite [31]. The interest in the environmental impacts of RECs has also been confirmed by the application of discrete choice experiment methods, with the social aspect also deemed an important driver [132], even if sometimes considered secondary to the environmental one [133], especially in cases where context factors (e.g., legal barriers, lack of incentives) fail to foster RECs formation [134]. The application of discrete choice experiment has yelled insights into the characteristics and configurations of RECs, which have the potential to either impede or encourage individual participation. Among the characteristics that have been observed and linked to individuals' socio-demographic and attitudinal attributes are ownership regimes of the energy generation systems [87], the involvement of municipalities and firms, governance characteristics such as voting rules [132], investment options, risks and losses, and other co-benefits [133].

The nexus between individual attitudes and characteristics, and the diffusion of RECs is the focal point of a body of research that utilises agent-based models (ABM) to simulate individuals' decisions regarding renewable generation projects. ABMs have commonly been employed to evaluate the evolution of a system under specific conditions, capturing the dynamic interaction among individuals [54,135], factoring in the time variable in the evaluation of the system [136]. In ABMs, agents are modelled as autonomous entities capable of deciding whether or not to act, considering various characteristics including attitudinal factors, socio-economic parameters, motivations, and interactions with other agents [2,54]. In the initialisation stage of an ABM (time 0), each agent is assigned an internal state represented by variables describing the propensity to act or not. At each time step, an agent determines whether to act or not, based on a set of rules defined by the modeller [54]. These rules delineate the manner in which the internal state of the agents are modified over time and through interaction with other agents [2]. The way in which these behavioural rules are specified by the modeller constitutes a pivotal step in the implementation of ABM, necessitating the collection of extensive data to ensure their accurate determination [135].

Of the studies analysed, Mittal et al. [54] modelled the willingness of agents to participate in a “community solar” project, defining a set of NPV-based behavioural rules moderated by individuals’ attitudes towards systems ownership and environmental concerns. Schiera et al. [2] integrated Theory of planned behaviour, Relative agreement and Small-world network to simulate the diffusion of distributed generation systems. Fouladvand et al. [137] simulated the behaviour of inhabitants confronted with the possibility to join a forming REC or an existing one, or the conditions under which the individual would decide to drop-off from a community after being a member. Building on Social value orientation, Fouladvand et al. [136] modelled the motivations and concerns of four person types (i.e. altruistic, cooperative, individualistic, competitive) deciding to join a REC. In this model, the primary motivations of the agents in joining the REC were energy independence, a sense of community, environmental concern and economic benefits.

ABMs present limitations in the way they describe the world [136]. Indeed, there is a certain degree of discretion in determining agents’ behaviour according to predefined sets of mathematical rules and thresholds [2]. Studies employing this method may utilise the results and methodological approach of research based on real case ex-post evaluation, as well as those employing choice experiments, in order to expand the analysis of consumers’ motivations and trade-offs toward the adoption of renewable energy solutions [54,136], to quantify and test the decision models specified through the decision rules as well as the assignment of inhabitants to different clusters of initial attitudinal states [136,137].

The studies belonging to this last cluster could be useful in evaluating the possible spread of RECs as well as their contribution to the energy transition [134]. However, further analysis of people’s motivations and values (also in quantitative terms) is necessary to support the role of policymakers in taking actions fostering citizens’ participation [2].

2.3 RECs and energy efficiency measures

The positive implications of a large penetration of RECs are manifold, with the alleviation of energy poverty frequently highlighted among the social benefits. Energy poverty affects around 8% of the European population [74], with a link, especially among vulnerable groups, to poor conditions and low energy efficiency of buildings [51,103]. Such interconnection between efficiency and distributed generation has been already stressed at European level [8]. It has been noted that RECs mainly focus on production from RES and not on energy consumption reduction [138]. While some research addresses this by integrating RECs in advanced district heating networks to avoid major refurbishments [139], few articles in the revised literature on RECs evaluate simultaneously energy production and consumption reduction. This still represents a limitation in the state of the art on the topic [121]. Amongst this limited number of articles, Mutani & Usta [96] considered the application of a REC to two condominiums in Italy, evaluating it with and without retrofit interventions. The study found that the implementation of a REC led to a reduction in energy consumption by 28% compared to the existing situation. Additionally, it reported a 26% increase in SSR and a noteworthy 11% increase in SCR. Sougkakis et al. [95] evaluated the financial benefits of performing a renovation of an existing neighbourhood in Greece, acting individually or as a community of inhabitants. Mutani et al. [97] evaluated the constitution of a REC in a small Italian town, pooling together several types of actors, among which the municipality. They evaluated the different performance of various scenarios, considering the possibility to reduce the energy consumption of public lighting, reporting, as predictable, an increase in SSR for more energy efficient interventions. Ascione et al. [121] analysed four mainly residential buildings in Italy, integrating the installation of roof-mounted PV panels with various EEMs, while also taking into account the impact of incentive schemes. Through an examination of the results on a multi-objective Pareto front, the authors identified the variation in the suitability of alternatives due to public grants. They concluded that such incentive schemes should prioritise the promotion of PV adoption over EEMs. It could be argued that this conclusion might be specific to the case under consideration and that incentivising distributed energy generation without simultaneously promoting the curtailment of energy consumption could reduce the environmental benefits associated with RECs [97], as well as cause potential problems to the main grid due to a generation/consumption mismatch [96,117].

There is a broad consensus that the combination of RECs and EEMs adoption at the building scale can have significant economic and environmental benefits.

However, it should be noted that, to the best of the authors' knowledge, all the articles considering energy efficiency in their analysis belong to the comparison and optimisation clusters, as previously identified. This imply that no real-world case studies have been found in the current literature, representing the combination of EEMs and distributed generation still a niche topic.

It could be argued that neglecting retrofit constitutes a sub-optimal approach to the debate on RECs contribution to the decarbonisation of the building sector, and also represent a misalignment between proposed solutions and the range of benefits (and potential risks) that a REC is ought to bring. In addition, given the progressive electrification of thermal uses [129], retrofit interventions are likely to play a pivotal role in ensuring a higher level of self-sufficiency, while limiting negative effects on the electricity grid (by reducing the generation/consumption mismatch). Finally, behavioural and demand management strategies regarding the willingness to modify individuals' consumption patterns [124,131] could represent a significant research avenue for optimising the use of locally generated energy.

2.4 Governance and evolution of RECs

The analysis of numerous real-world cases of energy communities has revealed a variety of contextual factors capable of promoting or hindering the diffusion of RES [31], as well as conditions that might determine their successful implementation and scalability [140,141], also leading to different forms of association [141]. In this regard, the role of institutional actors, such as municipalities, has been recognised as pivotal in the establishment and evolution of RECs, as well as in their governance [142]. In particular, it has been observed that individuals' attitude toward the establishment of energy communities are positively influenced by the participation of institutional actors in the process [140,143], or by the institutionalisation of the community itself [141]. Despite being constrained by a lack of control over local energy infrastructures and fiscal austerity [142], local authorities can play a pivotal role in fostering community acceptance [144]. Municipality-led RECs are usually favoured over enterprise-led ones by individuals [132,145]. Furthermore, the necessity to avoid uncoordinated development [117,146,147] could be mitigated by the integration of municipalities in the formation and operation of a REC, with some studies proposing the inclusion of RECs in municipal planning, further prioritising areas characterised by high levels of energy poverty [46]. However, the role of municipalities in the governance of a REC, and especially the interaction between them and the local population, has been interpreted in conflicting ways in different studies analysing real-world examples. For instance, the supportive role of municipalities in the inclusion of residents in an active manner under all the aspects of REC operation has been highlighted [89,127].

Conversely, both the high investment costs and the level of professionalisation needed to manage and upscale such projects could require a deeper involvement of the central authority as well as specialised actors [142,144], resulting in a decrease in the local dimension and sense of community ownership [89,148], ultimately relegating residents to passive roles [142]. Nevertheless, the inclusion of local authorities within RECs has the potential to yield notable social benefits, such as facilitating the deployment of generation technologies (i.e. on public buildings) in disadvantaged areas [149], and serving as a guarantor in the allocation of profits [129].

The issue of profit redistribution is intrinsically linked to that of voting model within the community itself. Indeed, it has been posited that actors should be enabled to self-determine the internal organization of the community [140]. However, specific arrangements may have counterproductive effects, such as discouraging small investors when a capital-based voting rule is introduced [33]. This poses a further balancing dilemma between the involvement of the local community and the necessary access to financial capital.

Finally, real-word RECs have been regarded as dynamic structures, in which the original scope and activities carried out evolve over time [129,140]. This emphasises again the importance of a coordinated development at different planning levels, as well as the proper integration of institutional and local actors and stakeholders, balancing the voting model between financial and participatory instances.

2.5 Scenario reconstruction

The analysis of the academic literature on REC feasibility has revealed a broad spectrum of indicators employed, along with inconsistencies in their conceptualisation and calculation. In this section, a meta-analysis is conducted, encompassing case studies within the Italian context. The objective of the analysis is to compare the performance achieved in these case studies, using a harmonised set of KPIs capable of providing a common description of such performance.

Table 4 provides a schematic representation of the boundaries and energy fluxes entering and exiting a REC. In light of the prevailing academic literature, three boundaries are delineated: one considering solely electrical uses, one encompassing both electrical and thermal uses, and a final one expanding the scope of analysis to include mobility uses.

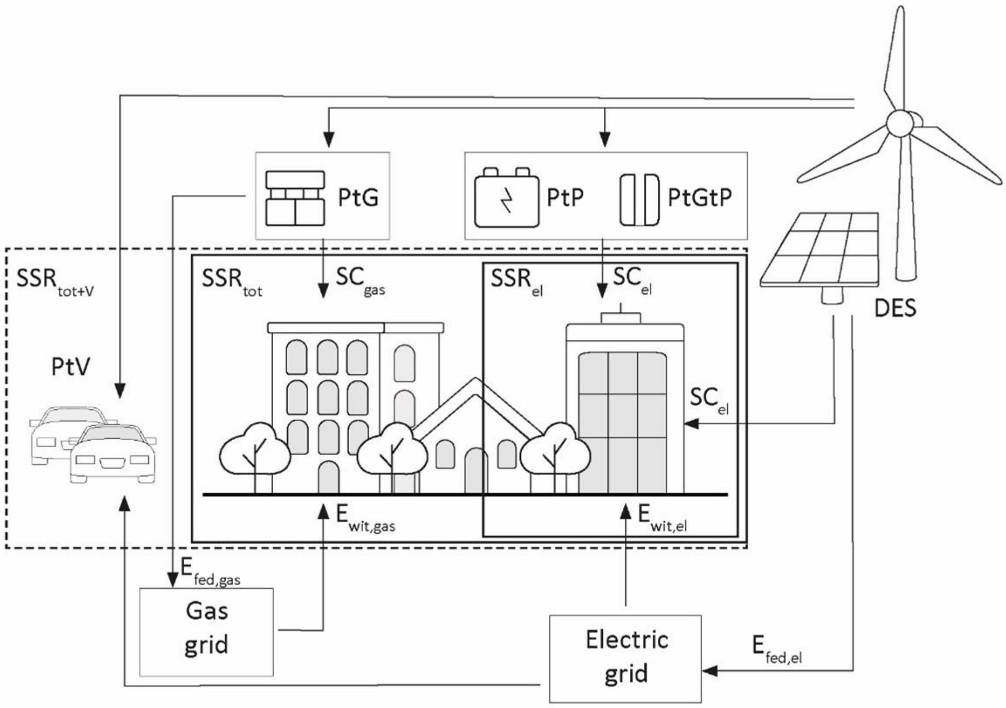


Figure 4 boundaries and energy fluxes considered in the meta-analysis.

In terms of descriptors, the performance of RECs could be evaluated by a limited set of indicators that outline the community feasibility and sustainability from a variety of viewpoints. These KPIs are summarised in Table 3.

Table 3 KPIs used to describe the RECs case studies.

KPI	Definition
$E_{wit,I,t}$	energy withdrawn from the grid for each energy vector *i* at time *t* (kWh/t)
$E_{fed,I,t}$	energy fed into the grid for each energy vector *i* at time *t* (kWh/t)
SSR_{tot}	self-sufficiency ratio (%)
SCR_i	self-consumption ratio for each energy vector *i* (%)
C_{inv}	initial investment cost for the establishment of the REC (€)
$CO_{2,emis,t}$	CO_2 equivalent emitted by the REC at time *t* (kg_{CO2e}/t)
$C_{en,t}$	community expenditure for energy purchase at time *t* (€/t)

In particular, the environmental sustainability of RECs could be described by its CO_2 equivalent emissions ($CO_{2,emis}$). Its interaction with the grid (a relevant aspect for the DSO) could be described by taking into consideration the energy taken from ($E_{wit,i,t}$), and fed into ($E_{fed,i,t}$), the national grid at each time *t*, for each energy vector *i* (i.e., controlling for its impacts on grid components [147]). The self-consumption ratio (SCR_i) is another significant indicator in the description of the interaction between the community and the grid, given the risk posed by excessive injection of locally produced energy on system stability [101]. The decision to utilize energy fed and not generated is substantiated by the need to consider potential transformations of energy from one vector to another, particularly in scenarios where Power-to-X strategies are implemented. The economic sustainability of the initiative is expressed by the initial investment cost (C_{inv}) and the expenses sustained by the community for energy purchase ($C_{en,t}$). Depending on the internal redistribution of financial benefits among members, the cost for energy purchase can also be interpreted as a social indicator when pursuing the goal of aiding disadvantaged segments of the population. It should be stressed here that this set of KPIs is among the most frequently used in the current production on RECs. Potential enhancements of the analysis could include the extension of the environmental scope, for example by incorporating life-cycle assessment approaches, or the addition of end-uses such as mobility ones (i.e., SSR_{tot+V}).

Finally, SSR is employed to assess the impact that the REC could have on the decentralization of the energy system and the independence that the configuration could achieve from uncertainties such as price fluctuation and energy supply risks. A key limitation of the current production on REC (especially in the Italian case studies analyzed) is the incomparability of the SSR indicator, defined as self-consumed over consumed energy. The boundaries within which this indicator is calculated are subject to variations depending on the energy produced and consumed for different energy vectors in the REC. To illustrate, in case studies where the electric vector covers only uses such as lighting and appliances, SSR considers only the share of consumption for those services [28], thus achieving high values of performance. Conversely, when the electric vector also covers thermal uses [98], SSR takes a more comprehensive nature, achieving lower values. Furthermore, in scenarios proposing a transition from fossil fuel generators to electric ones, the augmentation of the electric uses determines an incommensurability between SSR among the compared alternatives [102]. Finally, a further source of discrepancy has been identified in the consideration of alternative storage systems, such as those employed in PtV strategies, or the conversion of electricity produced locally into other energy sources. This is exemplified by the production of hydrogen, which could be sold, blended into the national grid (PtG), or converted back to electricity (PtGtP).

In order to mitigate the aforementioned issues, this analysis has constrained the scope of the proposed total self-sufficiency ratio (SSR_{tot}), limiting its application to building-level uses such as heating, lighting, and appliances (thus excluding PtV strategies). The cooling service is seldom included in the analyzed studies, as well as the domestic hot water one. While this might constitute a limitation, it could be argued that the considered uses are the most impacting ones in the existing Italian building stock. SSR_{tot} quantifies the energy used across the various vectors employed by a REC, according to equation 1:

$$SSR_{tot} = \frac{\sum SC_{i,t} * f_{p,i}}{\sum C_{i,t} * f_{p,i}} \quad (1)$$

where $SC_{i,t}$ is the self-consumed energy for vector i at time t, and $f_{p,i}$ is the conversion factor to primary energy for vector i. Self-consumption ($SC_{i,t}$) is calculated as in equation 2 in order to avoid double-counting of energy conversion in the REC (i.e., PtG or PtGtP):

$$SC_{i,t} = C_{i,t} - E_{wit,i,t} \quad (2)$$

where C_i is the energy consumption for energy vector i at time t, and $E_{wit,I,t}$ is the energy that the REC is not able to produce in its premises. It is necessary to highlight that in this meta-analysis $SC_{i,t}$ does not directly account for the energy exchange between grid and REC as it would be in a virtual self-consumption (VSC) configuration [47]. In case studies considering electricity to hydrogen conversion, the blending ratio within the national grid has been limited to 10% [29]. Consequently, natural gas self-consumption ($SC_{gas,t}$) is calculated according to equation 3:

$$SC_{gas,t} = C_{gas,t} * 0.1 \quad (3)$$

As mentioned above, in order to calculate SSR_{tot}, it is necessary to take into account all the energy consumed at the building level. This is not only to determine the full impact of RECs, but also to simultaneously consider the decarbonization of uses and the prior reduction of consumption, a strategy that, as highlighted, is rarely accounted in the current literature on RECs.

Several assumptions were made in order to reconstruct the scenarios and obtain comparable information from the meta-analysis. Firstly, the focus of this analysis is on the residential sector and the potential inclusion of occupants, so the normalization of the indicators was carried out considering the dwelling as the basic unit (with some possible inconsistencies related to the different apartment dimensions (m^2), which were mitigated by averaging different sizes on a larger scales). In case of additional non-residential members, each KPI has been referred to the residential members (excluding SSR_{tot} and SCR_{tot}, which were kept equal to those reported for the total REC), based on the energy consumption according to equation 4:

$$KPI_{x,resid} = KPI_{x,REC} * \frac{C_{resid}}{C_{REC}} \tag{4}$$

where $KPI_{x,REC}$ is the performance reported in the case study, $\frac{C_{resid}}{C_{REC}}$ is the ratio of energy consumed by the residential uses to the total energy consumed by the REC, and $KPI_{x,resid}$ is the KPI limited to the residential members.

A second assumption is the inclusion of space heating use. Several studies do not take space heating into account, focusing mainly on PV-generated electricity. To overcome this limitation, typical space heating consumptions for different geographical locations have been derived from Canova et al. [102], and have been added to the energy consumption reported by those case studies neglecting their calculation.

Finally, it is worth mentioning that some of the analyzed documents did not provide all the information necessary to reconstruct the scenarios according to the selected range of performance attributes. For this reason, some integrations of data were necessary to solve this issue. Table 4 provides an overview of these integrations. Looking at "Space heating" column, it is possible to notice that several documents did not consider space heating service, and data from Canova et al. [102] have been used. The same source has been used in case of missing CO_2 emission factors (as shown in CO_2 emission column). Three of the analyzed studies did not provide investment costs; for those, the equation provided by [29,101,150] to calculate investment cost as function of EEM size has been used. Finally, the studies not considering space heating service did not provide an energy price for natural gas. For these studies, data from [29,150] were used.

Table 4 Data integration strategies and parameters used in the reconstructed studies

Source	**Space heating**	**Investment costs**	**CO_2 emissions**	**SSR_{tot}**	**Energy costs**
[29]	-	PV: 715 €/kW_p EES: 491-530 €/kWh TES 4042 €/m^3 *	[102]	Calculated (not provided in the study)	Electricity: 184 €/MWh Natural gas: 86€/MWh
[98]	-	[29, 101, 150] PV: 759-850 €/kW_p	[102]	-	[102]
[101]	[102]	PV: 799-1001 €/kW_p *	[102]	Constructed†	Electricity: 249 €/MWh Natural gas: 86 €/MWh
[102]	-	PV: 1450 €/kW_p HP: 750 €/kW	-	Recalculated**	Electricity: 220 €/MWh Natural gas: [29, 150]
[60]	[102]	[29, 101, 150] PV: 1155 €/kW_p	[102]	Constructed†	Electricity: [102] Natural gas: [29, 150]
[47]	[102]	PV: 1500 €/kW_p	$CO_{2,NG}$ conversion factors [102]	Constructed†	Electricity: 145 €/MWh Natural gas: [29, 150]
[28]	[102]	PV: 1000 €/kW_p EES: 500 €/kW	[102]	Constructed†	Electricity: 220 €/MWh Natural gas: [29, 150]
[237]	[102]	PV: 1250 €/kW_p	[102]	Constructed†	Electricity: 220 €/MWh Natural gas: [29, 150]
[151]	-	PV: 1200 €/kW_p EES: 1000 €/kW	-	-	disaggregated value not provided
[150]	-	PV: 620-742 €/kW_p HP_{TH}: 571 €/kW HP_{DHW}: 994 €/kW EES: 447-546 €/kWh *	[102]	-	Electricity: 184 €/MWh Natural gas: 86€/MWh
[62]	[102]	PV: 1500 €/kW_p	[102]	Constructed†	Electricity: [102] Natural gas: [29, 150]

*: these studies used the equation provided by [29, 101, 150] to calculate the investment costs of different EEMs as function of their size; **: SSR provided in the document was not considering space heating consumption when the energy vector used to satisfy this service was natural gas; †: the study did not considered space heating, therefore SSR_{tot} has been recalculated adding the consumption for this service based on data from [102].

In Appendix B the reconstructed cases are reported. Figure 4 to Figure 7 plot the recalculated performance against the investment cost (C_{inv}) from a residential perspective, assuming a physical self-consumption (PSC) configuration. SCR_i and $E_{fed,i}$ are excluded from the plots as their maximization or minimization may constitute a conflicting objective for different actors. In order to better compare these scenarios, C_{en}, and $CO_{2,emis}$ are given in terms of savings, according to equations 5 and 6:

$$CO_{2,sav} = CO_{2,emis,b} - CO_{2,emis,s} \tag{5}$$

$$C_{en,sav} = C_{en,b} - C_{en,s} \tag{6}$$

where the subscript *sav* refers to savings, *b* and *s* refer to the baseline and specific scenarios in each study, respectively. It should be noted that $CO_{2,sav}$, $C_{en,sav}$, and SSR_{tot} are presented on an annual basis.

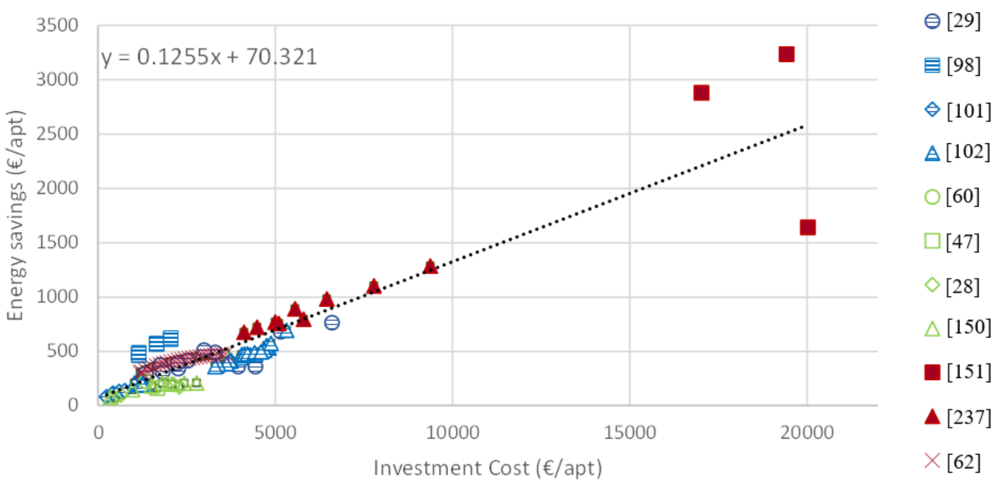


Figure 5 energy savings as function of investment cost in the analyzed papers.

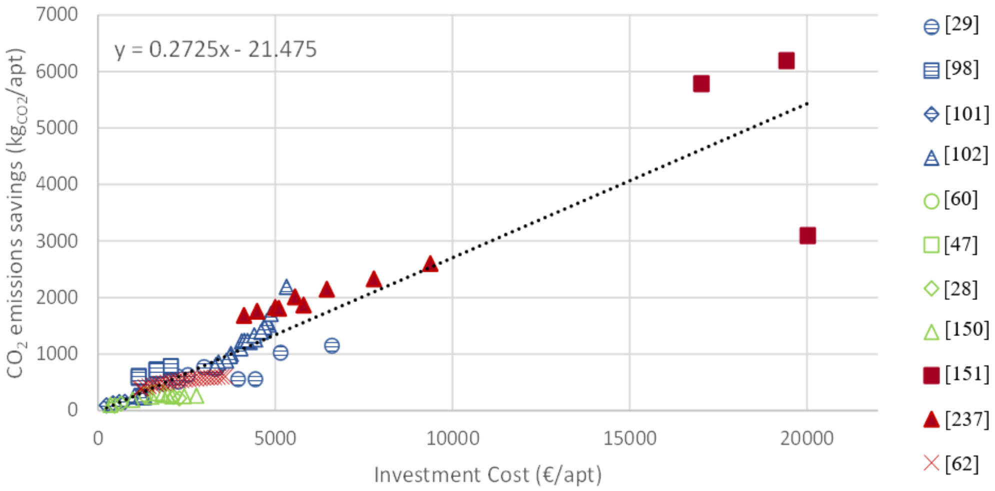


Figure 6 CO_2 savings as function of investment cost in the analyzed papers.

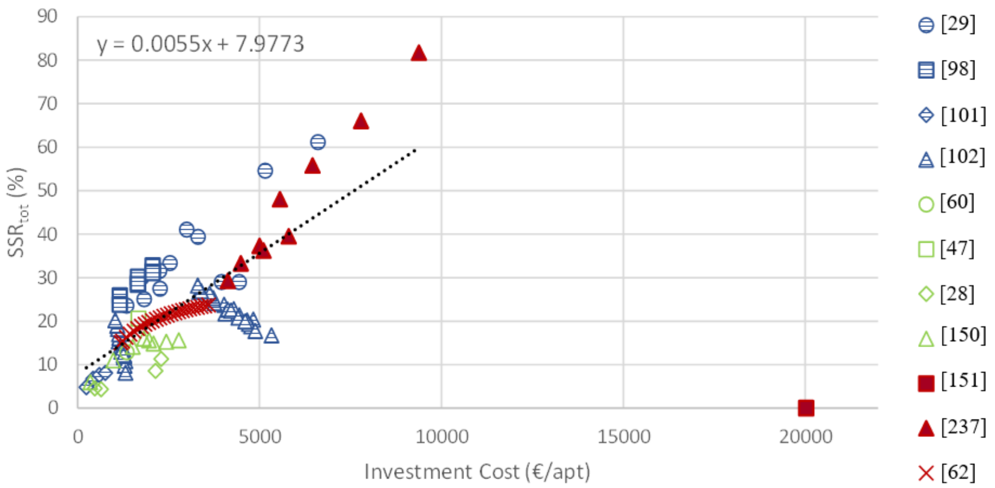


Figure 7 SSR_{tot} as function of investment cost in the analyzed papers. It is evident how the scenario from Aruta et al. [151] marked with a red square constitutes a peculiar case due to the fact that this scenario was not contemplating any active EEM.

All three figures show a direct correlation between investment costs and the three other indicators. Considering Figure 5, the cost of energy used by most of the study was quite similar, leading to a quite linear relation between the energy savings and the size of the installed systems (especially for energy generation systems). The same trend is mirrored in Figure 6 presenting the emission savings; here higher reduction of CO_2 emissions were due to larger size intervention measures with the same conversion factors used for national grid electricity and natural gas.

Most studies report relatively low investment costs and low benefits in terms of energy savings and avoided CO_2 emissions. It is worth noting that the analyzed studies mostly focus on the electricity produced by the installed PV panels and express SSR in terms of the electricity vector only, thus achieving low levels of SSR_{tot}. In particular, SSR_{tot} is the performance for which the larger variation is registered (as per Figure 7). This variation depends on the type of EEMs used in the scenarios. For instance, the introduction or not of storage systems affects this performance substantially, with studies like the one by Canova et al. achieving low levels of SSR_{tot} when the proposed scenarios contemplate only PV installation, with a slight improvement when energy self-consumption is increased by the introduction of an HP. On the other hand, Pastore et al. [29] introduce EES with a relatively low unitary price, scoring higher performances with similar investment costs. The highest SSR_{tot} performance are those reported by Spazzafumo & Raimondi [237]; in these scenarios also the natural gas consumption for space heating is tackled, achieving the highest values of SSR_{tot} even if in view of higher investment costs.

It is worth noting in Figure 7 that the scenario proposed by Aruta et al. [151] constitutes an outlier considering it only passive strategies and not active ones, thus presenting a null SSR_{tot} being the implemented measures uncapable of producing energy.

In general, all the best performing scenarios (without considering the investment cost) are those that introduce a heat pump ([29,98,150,151]) and consider different Power-to-X strategies. This is particularly evident in Canova et al. [102], where all the scenarios where heat pumps are introduced outperform those where only PV panels are installed, with an increase in SSR_{tot} between 7% and 8%. The same trend can be seen in Franzoi et al. [98], where the high performing scenarios also result in relatively low investment costs due to the presence of heat pumps in the status quo.

Finally, the best performing scenarios in terms of CO_2 and energy savings are by far those studied by Aruta et al. [151]. Here the strategy has been to set the investment costs similar to those for a deep renovation of the envelope. The best performing (from the environmental perspective) REC among all the studies belongs to this work, where the authors introduced a hybrid configuration with RES and EEMs (window replacement). It is worth noting that this solution refers to a rather moderate climate (a typical residential building in the same region consumes 89.04 kWh/m^{2}y [102], while the same building in different Italian regions presents energy consumption ranging from 59.2 kWh/m^{2}y to 238.8 kWh/m^{2}y [102]), which implies a potential different feasibility of other EEMs in colder zones.

In general, it is possible to argue that no substantial difference could be noted between scenarios considering various strategies when analyzing energy and emission savings performance, while a higher variation could be seen in the SSR_{tot} achieved.

Figure 5 has been generated considering a physical self-consumption (PSC) configuration (except for Aruta et al. [151], where it was not possible to extrapolate the energy self-consumption). Although the authors recognize that this is a simplification and an overestimation given the current Italian legislation, which takes into account the use of the public grid as a buffer in determining energy tariffs and incentives, it is nevertheless useful to limit the assumptions for comparison purposes. A further analysis to mitigate this problem was carried out by considering 100% of the energy self-consumed by the REC in a virtual self-consumption configuration, as shown in Figure 8. The figure shows how the reduction in benefits associated with self-consumption affects the results, making the introduction of EEMs even more attractive and effective.

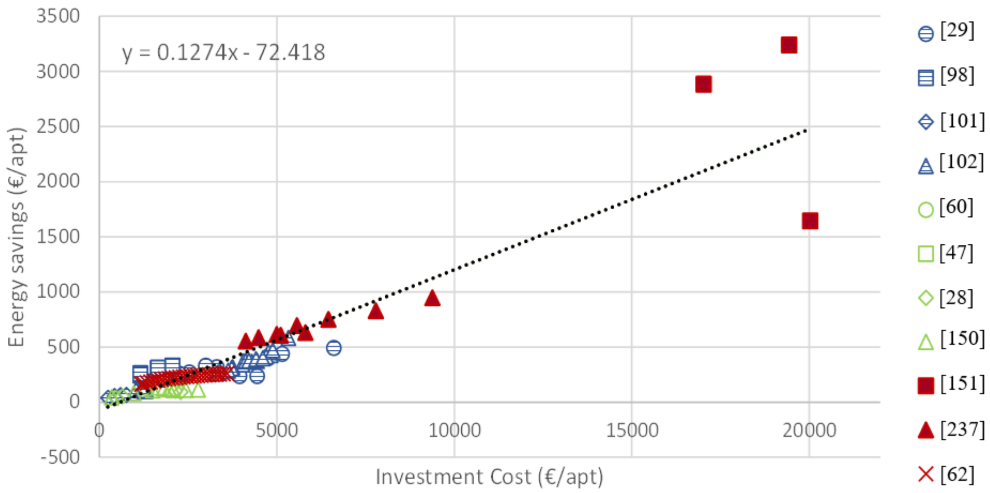


Figure 8 energy savings as function of investment cost in the analyzed papers in VSC configuration.

Several studies did not report SSR, but it has been calculated here on the basis of self-consumption and production data provided. For those that calculated SSR in their analyses, it is worth noting the difference in results when considering total energy consumption (recalculated according to equation 1) instead of just electricity consumption (as presented in the analyzed documents), with values 35.6% to 83.8% lower than those originally reported. Figure 9 shows this difference.

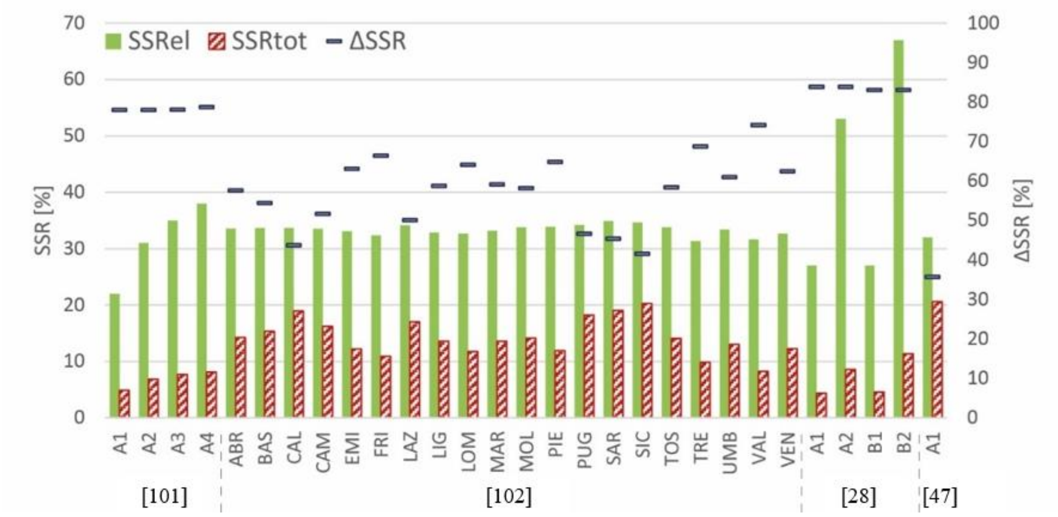


Figure 9 SSR_{el} variation as reported in the studies, and SSR_{tot} recalculated according to equation 1.

2.6 Lessons learnt

Renewable energy communities (RECs) could have a major impact on the energy transition in the built environment, mitigating its negative impact on GHG emissions in the urban context and thus representing a potential driver for urban renewal. According to RED II, RECs could bring several benefits to people and society, which are not only limited to the financial sphere, but also have a positive impact on the environmental and social dimensions. This literature review evaluated the common approaches used to analyze RECs in the feasibility phase as well as from the perspective of individual willingness to act, allowing for cross-sectional readings. Firstly, the study highlighted that the recognized lack of a comprehensive evaluation framework to assess RECs is also complicated by the inhomogeneity of the KPIs used (such as the economic or social aspect of energy cost reduction), and by the definition of the boundary set for their calculation (particularly for SSR_{el} and SSR_{tot} assessment). This inconsistency between boundaries can be related to the bias towards the productivity aspect of RECs, with less attention paid to consumption reduction. This is reflected in the tendency to rarely consider efficiency measures together with distributed generation solutions, thus limiting the potential (and neglecting risks) of RECs in supporting energy transition and reducing energy poverty. It is interesting to note that the studies that include energy efficiency measures in their scope of investigation belong to the feasibility analysis clusters, meaning that the introduction of energy efficiency measures in real-world case studies is a niche topic that is rarely explored.

In addition, this work analyzed the literature findings on the governance and internal organization of RECs, as well as their evolution over time revealing the need to further explore the interactions between different typologies of actors, which can potentially promote or discourage citizens' willingness to participate.

Furthermore, REC emerges as a multi-criteria, multi-stakeholder decision problem, which has to take into account different trade-offs and preferences at these two multi-dimensional levels. For instance, the relationship between the distribution system operator and civil society is usually designed to minimize and prevent the inversion of energy flows, rather than to optimize the bidirectional efforts of the two actors. Here, another research trend could be identified in the design of sustainable business models, able to exploit the desirability of solutions from the trade-offs between the parties' objectives.

It is clear from this review that participation in RECs is most often explained through the economic dimension. However, there are several other dimensions that are also taken into account by analysts and REC members. Environmental concerns and the goal of reducing environmental impacts are among the most important reasons for participation, albeit secondary to cost savings, along with reasons for self-sufficiency and energy security. Transparency of data and management, as well as bridging the knowledge gap, are also important for participation in the community. The development of RECs over time is characteristic of only part of the approaches used to study the subject, as researchers have mostly focused on feasibility. Feasibility over time could be another important avenue of research, considering the different times at which such experiences could start to work and how they could interact synergically with each other. Moreover, the choice of different solutions at different times could lead to sub-optimal or even negative outcomes for certain stakeholders (e.g. a large penetration of distributed generation from unpredictable renewable sources followed by a reduction in energy demand, leading to an increase in the production/consumption mismatch). In this sense, a well-established multi-stakeholder preference assessment framework may be needed to evaluate the feasibility of this promising instrument and to monitor its effectiveness throughout its life cycle.

The authors acknowledge some limitations in the scope of the analysis carried out, for example, a specific exploration of the business models adopted, as well as the evaluation of internal organization strategies or financial benefits and burden redistribution, could provide further insights in the academic production on RECs, benefit the synergies between actors and supporting policy makers in producing tailored solutions. These extensions of the horizon of analysis will be further explored in future studies.

A useful extension of the present study would be to include feasibility studies located in other countries, in order to compare the performance of RECs in different locations, as well as to include those entities aligned with the European definition that are currently being established and for which no data is yet available. Such an extension could provide an interesting opportunity to evaluate different business models, as well as redistribution mechanisms and economic feedback within them. Furthermore, broadening the scope of the scale of the analysis to include other uses (e.g. mobility) or other actors (e.g. municipalities, industry, tertiary, commercial, etc.) could represent further optimization to achieve higher decarbonization targets at the urban scale. Finally, as the evaluation of the social benefits of RECs remains an unresolved issue, an agreed evaluation framework would be required to better design the establishment of RECs in line with the idea of a just transition.

Chapter 3

Spatialization of preferences

3.1 Overview of the method

As previously stated, the European Union recognizes the critical role that civic participation can play in the reduction of its environmental impact, and in the energy transition of its society [152]. Despite this, bottom-up initiatives at the urban scale remain sluggish [31]. When analyzing individual participation in energy projects, it is necessary to consider both the actions that individuals might be capable of implementing, and the actual behavior that they decide to implement. This process is consistent with Sen's Capability approach [153,154], which emphasizes the individual's decision in converting a potential set of capabilities into functioning, with the non-negligible mediation role played by individuals' preferences.

The Capability Approach differentiates between two primary components: functionings, which represent observable states of being and actions derived from an individual's resources, and capabilities, the spectrum of potential functionings available to an individual. The conversion of capabilities into functionings, or the selection of one specific functioning against another [153], is neither automatic nor uniform, underscoring the pivotal role that personal preferences play.

In the context of urban environment, the capability approach has been further specified in the Urban capability approach, with urban functioning referring to the choices individuals make, or the identities they adopt within the myriads of options (or urban capabilities) available to them [155–158]. This set of capabilities is shaped by the interplay between an individual's inherent abilities and the external conditions in which they operate [159]. As articulated by Chiappero-Marinetti [160], the transformation of means into actual states is not a mechanistic process but is instead influenced by a plurality of intrinsic and extrinsic conditions, with the space itself functioning as a potential conditioning factor. In this sense, the Capability approach implies the necessity to describe the real opportunities an individual has to use the city and to function autonomously and freely [161].

This perspective positions individuals as central agents in processes of transformation [162], while also shedding light on scenarios where the interaction between a city and its residents might be suboptimal, or where only a fraction of the urban population might be effectively engaged [157,161].

Despite the provision of an expanded set of capabilities to urban residents, individual preferences have been shown to critically influence the translation of these opportunities into functionings [157,161]. For instance, while energy projects may enhance individuals' capabilities [163–165], personal priorities on how individual and community well-being are conceived might determine different opinions on the conversion from capabilities to functionings [163]. In the context of the energy transition in urban areas, it becomes essential to assess how individuals evaluate, prioritize and value different outcomes and co-benefits of their participation in energy projects [164]. Indeed, different value systems or interpretations of what constitutes a meaningful sustainable contribution may result in the prioritization of alternatives that are not in line with the overall community objectives or leads to conflicts [165]. Therefore, the investigation of the preferences in converting the possible option that individuals have in terms of access to energy projects in actual functionings could be a way to promote the alignment of private actions with the public interest.

This section describes the method used to estimate two aspects of RECs potential diffusion: (i) the probability of individuals to accept to participate in a REC configuration, and (ii) their preferences among potential alternative RECs (thus the process of converting capabilities into functionings). Figure 10 gives an overview of the method.

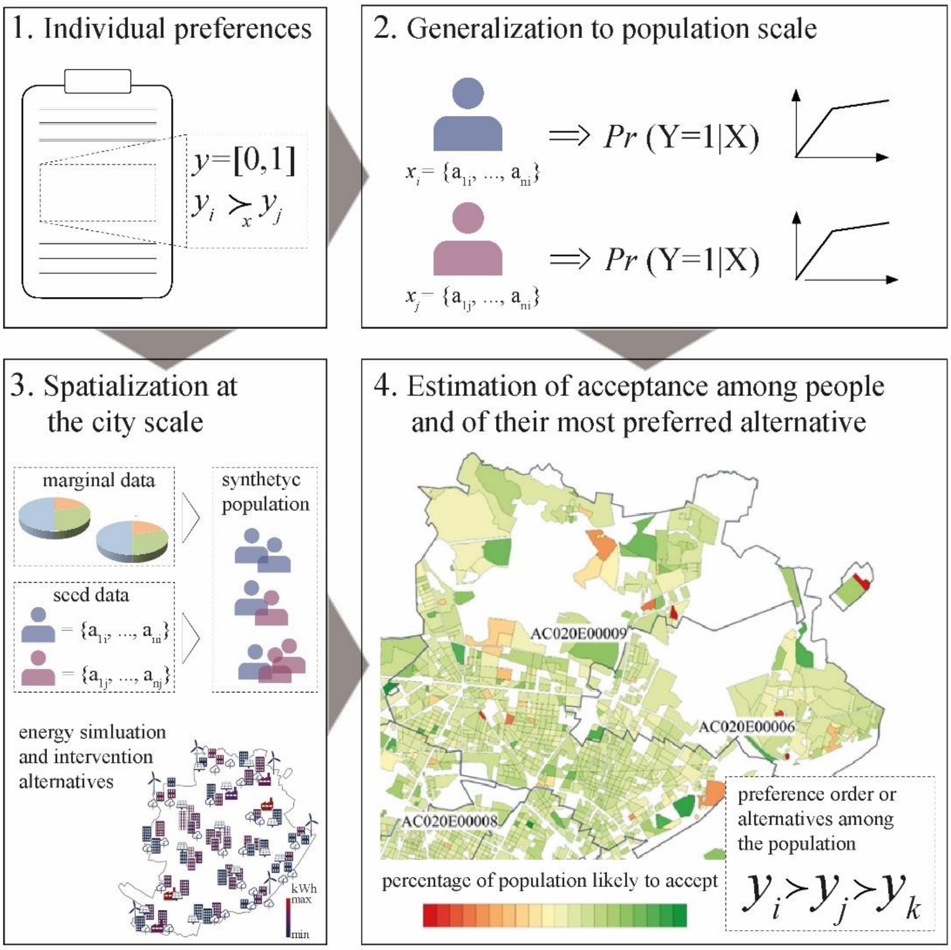


Figure 10 Graphical overview of the method. In step 1 the questionnaire is distributed; in step 2 the preferences are generalized; in step 3 the preferences are spatialized at the urban scale. In the last step the percentage of the population likely to accept a REC is evaluated, together with the ranking among the proposed alternatives.

In particular, the method is divided into three main parts. The first part (Figure 10.1) is dedicated to the analysis of individual preferences for RECs and is divided into the following steps:

- In the first step, a questionnaire is distributed to a sample of the population to collect information regarding the willingness to participate or not in different RECs alternatives. The survey also collects the preference of respondents towards these alternatives in the form of rankings.
- In the second step, a UTA method is applied on the preference rankings to estimate each respondent's decision model, expressed as a set of value functions.

- In the third step, two regression models are applied. One aims to estimate the probability of a respondent accepting a REC (generalized linear mixed model) based on socio-demographic profiles of individuals, while the other analyses (linear mixed model) the influence of these socio-demographic characteristics on the preferences expressed by individuals (i.e., the shape of the estimated value functions).

The second part of the method (Figure 10.2) aims to generalize the results of the previous analysis from the individual to the population scale. In particular, this part of the method is divided into the following steps:

- First, types of individuals are constructed based on the results of the regression models. These types of individuals are defined by vector of socio-demographic attributes.
- Second, the previously estimated value functions are clustered using an unsupervised algorithm. The representative value functions of the clusters are extracted.
- Third, a supervised classification algorithm is used to assign the sets of value functions to the different types of individuals.

In the third part (Figure 10.3), the generalized decision model and the possible outcomes of RECs alternatives are spatialized at the city level. This is done by:

- generating a synthetic population at the urban scale characterized into the types of individuals estimated previously;
- estimating the performance level of different possible alternative REC configurations.

Finally, the probability of the inhabitants accepting (using the results of the generalized linear mixed model), and their preferred alternative (by means of the value functions estimated for the types of individuals) are estimated (Figure 10.4).
In the following section, after a brief introduction of the theoretical foundations of preference analysis, the above steps are described in more detail.
It should be noted that the method presented is specifically tailored to investigate the potential acceptance of RECs and the preferability among alternatives. It is worth noting that this could be interpreted as one of several instances representing the set of urban capabilities of the inhabitants. It could be argued that the same method could be used to study individuals' preferences for several other types of potential processes of translating capabilities into functionings.

3.2 Preference analysis

3.2.1 Preference Learning theoretical framework

Formally, the preference model of an individual can be expressed by means of binary relations [166]. Binary relations are used in order theory to formally express relations among elements in terms of relative positions in a ranking or to compare them. For instance, considering a set of elements A and a binary relation $\succcurlyeq$ among its inhabitants, it is possible to define $\succcurlyeq$ as a subset of the cartesian product A x A, thus describing the relative order among the elements of A. In other words, considering two elements a,b $\in$ A, a preference could be expressed as a binary relation inducing an order a $\succcurlyeq$ b such that a is weakly preferred to b (i.e., it occupies at least the same position as b or a higher one in the order of preference). Among the possible order structures that could be described, the preorder is a specific type of binary relation satisfying the property of:

- reflexivity: a $\succcurlyeq$ a, and
- transitivity: if a $\succcurlyeq$ b, and b $\succcurlyeq$ c, then a $\succcurlyeq$ c

A preorder that is also antisymmetric is called a partial order, thus:

- if a $\succcurlyeq$ b and b $\succcurlyeq$ a then a = b.

Furthermore, a total order in which any two elements are comparable such as:

- a $\succcurlyeq$ b or b $\succcurlyeq$ a

is called a totally ordered set [167].

Binary relations and orders are commonly used in Preference Learning as input data for the solution of "learning to rank" problems. Three main ranking problems have been identified in the attempt to predict (i) the order that the user would give to instances in a set (object ranking), (ii) the rank of possible labels (among the elements of a set of labels) with which these instances could be classified (label ranking), and (iii) the classification of instances in ordered classes (instance ranking). These problems are usually solved by learning a ranking function that could be useful to identify covariance between preferences and socio-economic information [168] in case of similarities among users [169]. In particular, the object-ranking family of problems attempts to learn a function that, given a potentially infinite set of objects y $\in$ Y and training information in form of exemplary rankings or pairwise preferences, produces a permutation τ of the set of objects. The ranking of objects y $\in$ Y is determined by assigning a score and sorting them accordingly [170]. The score assigned to the objects of set Y could be explained with a value-based representation of preferences in the framework of Multi-Attribute Value Theory (MAVT) developed by Keeney and Raiffa [170,171].

MAVT represents alternatives in a decision problem as composed by a set of attributes and interprets preferences by means of value functions associated with these attributes. The total value of an alternative is given by the sum of the marginal values for the attributes the alternative is composed of. Therefore, by means of criteria aggregation such as the additive one, it is possible to express the total value of the alternative *y* (7):

$$V_x(y) = \sum_{a=1}^{n} w_{xa} v_{xa}(y) \tag{7}$$

where V_x(y) is the overall value of the alternative y for individual x, v_{xa} are value functions for individual x on each attribute a that describe the alternative y, normalized between 0 and 1, and w_{xa} represents the trade-offs among the attributes for individual x. In this way, we can define the preference of individual x over two objects (the ranking function for x over Y to be learned) as (8):

$$y_i \succcurlyeq_x y_j \Leftrightarrow \sum_{a=1}^{n} w_{xa} v_{xa}(y_i) \geq \sum_{a=1}^{n} w_{xa} v_{xa}(y_j) \tag{8}$$

meaning that, individual x weakly prefers (i.e., either y_i is preferred over y_j or the two are indifferent) if the weighted sum of attribute values for y_i is equal or exceeds that for y_j, according to x's own preferences. To evaluate the preferences of individuals it becomes paramount, then, to estimate the shape of the value functions that the specific individual assign to each attribute that describe the different alternatives at their disposal. In the operational research and artificial intelligence communities, the interest around learning and predicting preferences has grown [169], and, in recent years, Preference Learning (PL) has emerged as a branch of Machine learning referring to the task of learning the preferences of individuals or classes of individuals from information about a set of items [172]. Roughly speaking, PL aims to induce preference models from data regarding explicit preference information [173,174], with a common attempt being to approximate the utility function of a decision maker in multi-criteria and multi-attribute settings.

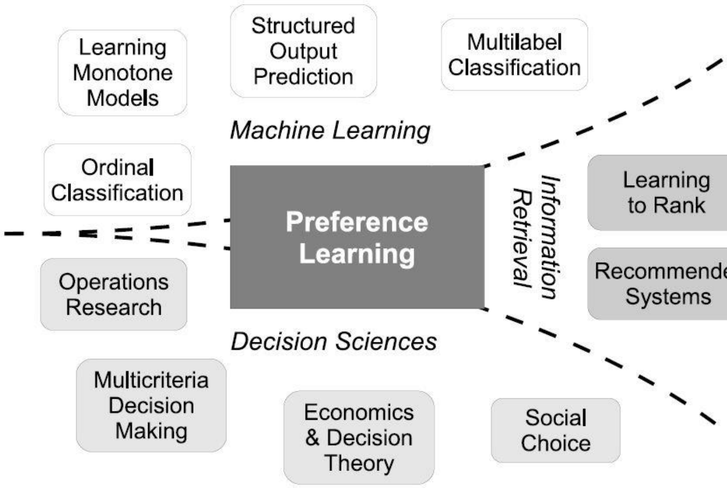


Figure 11 Fields of applications of Preference Learning (source: (Hüllermeier & Fürnkranz, 2013))

PL has been widely used in several fields of applications (Figure 11) such as prediction of customer design choices [175], travel modal choices [176], logistics, information retrieval, collaborative filtering [173].

3.2.2 Survey of population preferences: the questionnaire

The proposed questionnaire has been constructed using the online survey tool limesurvey [177], and is composed of three main sections. A first introductory section (i) devoted to a general set of questions based on a 4-point Likert scale on the themes of the research, in order to gather some introductory data regarding respondents' inclination toward RECs (the respondents' "Vies of the World") and their possible economic, environmental, and social benefits [23]. A second section (ii) in which the respondents were asked to state if they would accept-maybe accept-not accept to participate in several REC alternatives. In the same section, respondents were asked to rank the alternatives they defined as acceptable and maybe acceptable. The third section of the questionnaire (iii) collected the socio-demographic data regarding the respondent and their households. In Table 5 the logical structure of the questionnaire is presented (in Appendix C screenshots of the online questionnaire are provided).

Table 5 Logical structure of the questionnaire distributed to the population sample. The "Condition" column define questions that were presented or not to the respondents depending on the answer to previous questions.

Section 1			
Code	**Question**	**Modalities**	**Condition**
1.1	Collaboration between people is essential to reduce energy consumption and all be more sustainable	4-Point Likert scale	-
1.2	Reducing energy consumption is necessary to combat climate change	4-Point Likert scale	
1.3	Adopting energy efficiency measures in my home seems expensive to me, and it is not a priority for me	4-Point Likert scale	
1.4	Climate change can worsen social inequalities between people	4-Point Likert scale	
1.5	Promoting the well-being of all individuals is important, even if this may limit individual profit	4-Point Likert scale	
1.6	The State must commit to reducing society's environmental impact, including by introducing taxes to discourage the use of fossil fuels	4-Point Likert scale	
1.7	Are you already member of a REC	Y/N	
1.8a	Participating in the Energy Community has allowed me to reduce my impact on the environment	4-Point Likert scale	If 1.7 = Y
1.9a	Participating in the Energy Community has allowed me to have a positive impact on my quality of life and that of other people as well	4-Point Likert scale	If 1.7 = Y
1.10a	Participating in the Energy Community has allowed me to save on my energy bill	4-Point Likert scale	If 1.7 = Y
1.11a	Participating in the Energy Community has allowed me to be less dependent on changes in the price of energy	4-Point Likert scale	If 1.7 = Y
1.11a	Participating in the Energy Community required a significant initial expense	4-Point Likert scale	If 1.7 = Y
1.8b	I am interested in the social benefits that an Energy Community can guarantee for families in economic difficulty	4-Point Likert scale	If 1.7 = N
1.9b	I am interested in the savings on my energy bill that an Energy Community can guarantee	4-Point Likert scale	If 1.7 = N

1.10b	I am interested in the reduction of polluting emissions that an Energy Community can guarantee by producing clean energy or reducing consumption	4-Point Likert scale	If 1.7 = N
1.11b	I am interested in the possibility of self-producing part of the energy I consume in order to be more self-sufficient	4-Point Likert scale	If 1.7 = N
1.12b	If I decided to become a member of an Energy Community, I would also be willing to invest money	4-Point Likert scale	If 1.7 = N

Section 2

Code	Alternative selection/Ranking	Modalities	Conditions
2.1-2.8	Would you participate in a REC option (y)? (With y∈Y, and Y={A, B, C,…,H})	Yes/No/Maybe	
2.9	Rank the options you stated you would agree to participate in	Complete order	∀y∈Y, | ⟨y,Yes⟩
2.10	Rank the options you stated you might agree to participate in	Complete order	∀y∈Y, | ⟨y,Maybe⟩

Section 3

Code	Question	Modalities
3.1	Sex	2
3.2	Age group	13
3.3	Educational level	5
3.4	Profession	10
3.5	Citizenship	2
3.6	Household size	7
3.7	Questions 3.1 to 3.5 for each member of the household (as stated in 3.6)	-
3.8	Property regime of the household	2
3.9	Annual income of the household	5
3.10	Who has more influence on household decisions	4
3.11	Household location	Pin on map

The modalities of the questions define the possible answers that the respondents could select. In the first section of the questionnaire (question 1.1 to 1.12b) the respondents could select their level of agreement with the question statement on a 4-Point Likert scale. In the third section the respondents could select from different classes of socio-demographic characteristics (i.e., age = {<15, 15-19, 20-24, …, >75 years old}, sex = {Male, Female, Other}, etc.) in accordance with the modalities collected by Italian census surveys [178].

With reference to the Income level, the classes among which the respondent could have selected are based on those provided by the Italian tax agency. The choice to allow for class selection and not to treat the variable as a numerical one (i.e., allowing the respondent to directly input a value) has been taken to reduce the risk of mistakes in the filling of the questionnaire by the respondents (for data gathering purposes the reliability of data has been prioritized over the granularity of the data).

In this section of the questionnaire, it was also possible to indicate the postcode (in Italian: Codice di avviamento postale, CAP) as well as the location of residents' household. The second option allowed the respondents to select a point of interest in proximity to the location of their dwelling. This was meant to protect their privacy and increase the chances of obtaining a response.

Focusing on the second part of the questionnaire, the respondents were asked to express two types of information: (i) the alternative RECs that they would consider acceptable, maybe acceptable, or not acceptable, and (ii) the preferential order among the alternatives that they would accept or maybe accept. The set of alternatives Y for which the respondents had to express their acceptability and preference was extracted from the dataset of case studies analyzed and reconstructed in the literature review on RECs (Chapter 2). Each alternative y has been described as a vector of four performance attributes: $y_i \in Y \mid y = \langle \beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i} \rangle$, where β_1 is the Investment cost, β_2 is the energy cost savings, β_3 is the avoided CO_2 emissions, and β_4 is the energy Self-sufficiency of the alternative. The list of attributes used to describe each alternative has been selected from those commonly found in the literature on energy communities as analyzed in Chapter 2. While it is not possible to exclude the presence of other important attributes in the decision related to REC participation (for instance the management form adopted, or the presence of public bodies among the members), those selected cover the range of quantifiable performance that could affect the residential stakeholders (target of the study). It is not possible to exclude the presence of hidden preferences related to RECs such as, for example, the capability of the configuration to maximize its social impact. Nevertheless, this possibility is registered by the energy cost reduction attribute that could be seen as a benefit allocated to the whole community, or to specific members (as suggested by the explanation videos included in the questionnaire). Finally, the set of attributes have been tested for redundancy and separability to determine their relevance in modelling preferences towards REC participation for residential stakeholders.

To define the set Y to be presented to the respondent a K-means algorithm has been applied to the dataset of case studies. K-means is an unsupervised clustering Machine learning algorithm that uses Euclidean distance as a metric to assign data to one of a pre-specified number of clusters. The algorithm first defines the position of each cluster centroid and calculates the distance between each point of the dataset and the cluster's centroid, assigning the points to the clusters based on the minimum distance between points and centroids. Then, the cluster centroids are recalculated on each cluster's points. The procedure is iterated until convergence. The number of clusters to be specified is not given, therefore several techniques could be used to determine the proper number of clusters. Among these techniques, the Silhouette method tests the model for several number of clusters and calculates an a-dimensional score from 0 to 1 for each of the runs, where 0 is the worst performing model and 1 is the best performing one. Figure 12 presents the Silhouette score results for several specified number clusters:

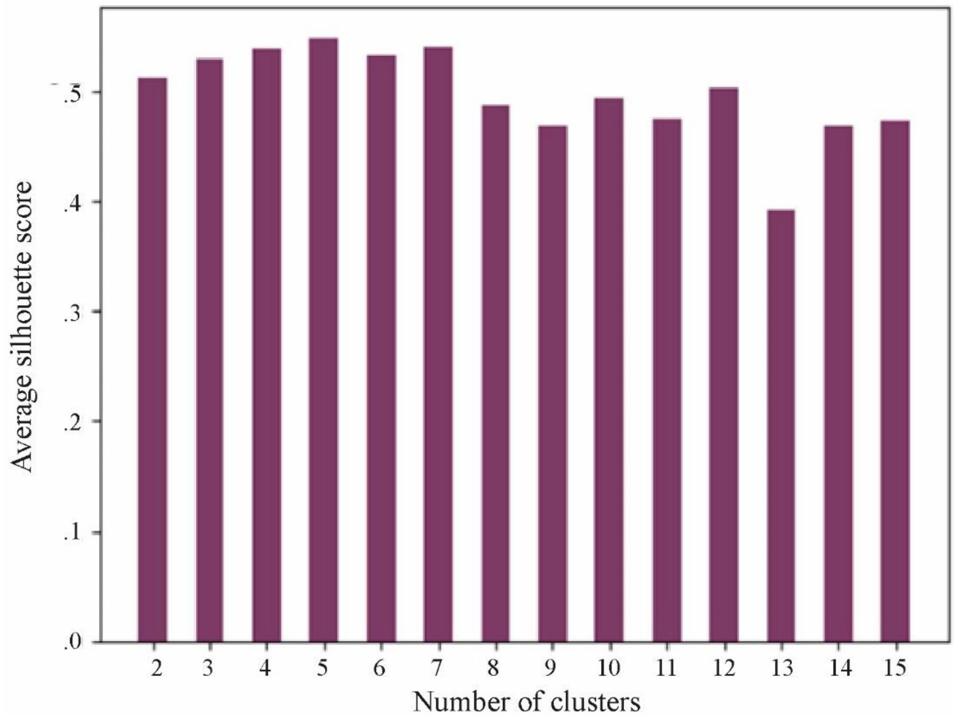


Figure 12 Silhouette scores for different number of clusters specified in K-Means classification algorithm

By using the Silhouette score it is possible to see that the best number of clusters is five, while after seven clusters, the performance is reduced. Nonetheless, the two models specifying five and seven clusters present similar results (0.549, and 0.541 Silhouette score respectively). Therefore, it has been decided not to limit the number of alternatives in the questionnaire to five (while not overwhelming the respondent with an excessive number of options to process), but to perform the K-Means clustering algorithm with seven clusters. These clusters were then analyzed to extract one representative alternative from each of them. This step required a manual analysis of the clusters to define a proper set of alternatives not allowing the simple definition of a characteristic alternative from the clusters (e.g., the alternative geometrically closest to the centroid of the cluster). In fact, the set of alternatives had to be sufficiently differentiated, but also had to include alternatives with some degree of similarity in order to better extrapolate the preferences of the respondents. For this definition a try and error procedure has been used. It is also worth noting that one of the reconstructed scenarios from Chapter 2 presents an Investment cost value far higher than the others due to the inclusion of the retrofit of the envelope among the energy efficiency measures [151]. For this reason, such point of the dataset has been at first excluded from the clusters' definition and reinserted at the end to compose the final set of eight alternatives to be included in the questionnaire: In Table 6 the set of alternatives is presented.

Table 6 Final set of alternatives included in the online questionnaire

Scenario	**Investment cost [€]**	**avoided CO_2 emissions [kg_{CO2}/year]**	**energy-bill savings [€/year]**	**Self-sufficiency [%]**
P8	6,600	1,147	764	61
C17	1,237	250	199	14
C12	4,650	1,419	507	20
Pb2	4,996	1,827	769	37
Pb8	9,376	2,603	1,285	82
C38	5,322	2,190	694	17
A1	20,024	3,095	1,643	0*
C28	3,395	854	382	27
P7	2,992	769	512	41

*this alternative presented only passive strategies therefore, even if it is the most expensive one (also achieving the highest performance in terms of avoided emissions and energy savings), it has no capability to produce energy locally, thus having no self-sufficiency.

For each of the alternatives, a respondent was allowed to state either if she would be willing or not to participate in the constitution of a REC with such characteristics (Yes/No), or if she was uncertain (Maybe).

In Figure 13, an example of the question is displayed (note that the values reported for each attribute have been simplified and rounded to reduce the effort required from the respondent to interpretate the data).

Alternativa X	Costo iniziale	4.000 €
	Risparmio in bolletta	300 €/anno
	Riduzione emissioni	600 kg_{CO2}/anno
	Autosufficienza	40 %

Figure 13 Example of tabular description of a REC alternative as presented in the online questionnaire. “Costo iniziale” is the Investment cost, “Risparmio in bolletta” is Energy cost savings, Riduzione emissioni is Emission reduction, and “Autosufficienza” is Self-sufficiency.

The latter step results in the clustering of the alternatives proposed in the questionnaire into three sub-sets by each respondent: (i) the acceptable sub-set ($Y_x^{(i)}$), (ii) the maybe acceptable sub-set ($Y_x^{(j)}$), and (iii) the not acceptable sub-set ($Y_x^{(k)}$). After this subdivision, each respondent x was asked to express two distinct rankings, one among the alternatives for which she stated the willingness to participate (thus a permutation $\tau_x(Y_x^{(i)})$), and one among the alternatives for which she was uncertain regarding accepting or not to participate ($\tau_x(Y_x^{(j)})$).

3.2.3 Individual value function estimation: UTA methods

The rankings of the alternatives extracted from the questionnaire are used as input data for the Preference Learning phase to construct an additive preference model coherent with the respondents’ judgement [171].

Among the available methods, Additive Utility Functions (UTA) methods [179,180] take global statements upon set of alternatives and subsequently infer the value functions that best fit such statements [181,182]. The methods approximate value functions by minimizing the errors between the given preference structure over the set of alternatives, and the one computed by calculating the total values of alternatives given by the estimated value functions.

UTA methods use linear programming to infer a set of piece-wise linear value functions that match a learning set (i.e., information about preferences) provided by a DM [171]. In the context of MAUT, this set of value functions allows one to calculate the marginal utility provided to the DM by the level of performance of an alternative in each of the attributes used to describe it. Finally, the aggregation of the marginal utility into a total utility permits the comparison between different possible alternatives. In the application, the UTAstar algorithm, evolution of the UTA algorithm, has been used [180].

To perform the analysis, the learning set (i.e., the order of preferred alternatives) for each respondent was retrieved from the questionnaire. At first, the two ordered set of alternatives provided by the respondent have been combined: (i) those for which the respondent has expressed a willingness to participate ($\tau_x(Y_x^{(i)})$) and (ii) those for which the respondent was not fully sure about acceptance ($\tau_x(Y_x^{(j)})$) as seen in Section 3.2.2 and expressed in equation (9):

$$\tau_x\left(Y_x^{(i)}\right) + \tau_x\left(Y_x^{(j)}\right) = (\tau_1^{(i)} \succ \tau_n^{(i)} \succ \tau_1^{(j)} \succ \tau_m^{(j)}) \quad (9)$$

being:

$\tau_x\left(Y_x^{(i)}\right) = (\tau_1^{(i)}, \dots, \tau_n^{(i)})$ the alternatives for which the respondent x would accept participation ($Y_x^{(i)}$) ordered from the most preferred to least preferred one, and

$\tau_x\left(Y_x^{(j)}\right) = (\tau_1^{(j)}, \dots, \tau_m^{(j)})$ the alternatives for which the respondent x would be unsure regarding participation ($Y_x^{(j)}$) ordered from the most preferred to least preferred one.

Furthermore, the subset of alternatives $Y_x^{(k)} = (y_1^{(k)}, \dots, y_q^{(k)})$ for which the respondent expressed the refusal to participate, have been introduced in the combined set as less preferred in comparison to the alternative ranking last ($\tau_m^{(j)}$), and indifferent among each other (i.e., they were not ordered by the respondent). The latter results in the final total rank of all the initial 8 alternatives proposed in the questionnaire:

$$\tau_x\left(Y_x^{(i)}\right) + \tau_x\left(Y_x^{(j)}\right) + Y_x^{(k)} = (\tau_1^{(i)} \succ \tau_n^{(i)} \succ \tau_1^{(j)} \succ \tau_m^{(j)} \succ y_1^{(k)} \sim y_q^{(k)}) \quad (10)$$

Considering that several respondents expressed only one alternative for which they would have been interested in participating (both answering "yes" or "maybe"), a necessary threshold of two minimum preferences expressed has been set to have at least one pair of ordered alternatives on which to train the algorithm. It could be argued that inferring one individual set of value functions from only one pair of ordered alternatives is questionable. Nevertheless, the aim of the analysis was to estimate a generalized model at the urban scale; hence the decision to retain these rather small learning sets. This aspect is discussed further in Section 5.1.4.

3.2.4 Individuals' acceptance and preferences

General linear and linear mixed models have been used to estimate (i) the decision of the respondents to accept a REC alternative, and (ii) the variation of the marginal utility of the performance attributes for different types of individuals. The goal is to test the influence of socio-demographic characteristics on these two types of information.

Linear mixed-effects models (LMM) are a family of tools that differ from traditional regression models by the fact that they include random effects in addition to fixed effects [183]. Fixed effect represents the determinant used to explain a specific output variable under study. For instance, in the present analysis these are the alternative performance and individuals' socio-demographic characteristics associated with the probability to accept the alternative REC [183]. Random effects, on the other hand, capture correlations among observations that arise from grouping or clustering in the data (such as repeated measurements from the same individual). In this context, each respondent evaluates the opt-in possibility for multiple alternatives, and these responses are influenced by respondent-specific factors (e.g., stable preferences or biases). This induces statistical dependence among their responses. Introducing random effects allows the model to account for this within-respondent correlation, ensuring valid inference.

The analytical representation of linear mixed-effect models is as follows:

$$y = X\beta + Zu + \varepsilon \tag{11}$$

where y is the vector of observations, X is the matrix of explanatory variables, and β the vector of fixed-effects coefficients; Z is the matrix of random effects, and u the vector of random-effects (complement to β); finally, the term ε represent the vector of the residuals.

Besides the introduction of the random effect to correct for autocorrelation among clusters, another effect could be introduced in the model to account for interactions among different fixed effects. For this purpose, the moderator effect could be introduced in the model. In a linear model, a moderator variable M affects the relationship that exists between a determinant x and a dependent variable y, describing the conditions that might vary the strength and direction of such relationship [184,185].

Figure 14 presents a graphical representation of the concept behind the effect of moderators in linear models.

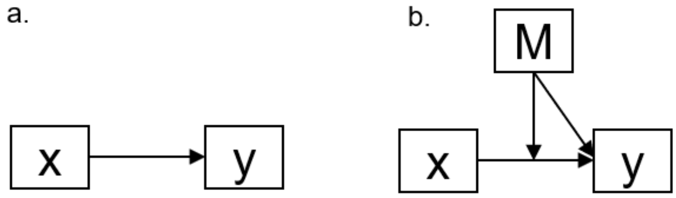


Figure 14 linear model (a) and linear model with moderator (b).

With reference to the above, the model including interaction (Figure 14.b) could be expressed as follows:

$$y = \beta_0 + \beta_1 x + \beta_2 M + \beta_3 xM + \varepsilon \tag{12}$$

with β_0 being the intercept of the regression model, β_1 and β_2 the regression coefficients for variable x and M, and β_3 the effect of the interaction between the two variables in explaining the dependent variable y (it is worth noticing that the identification of the moderator is theoretical, and the model treats the two variables in a similar way). As usual in regression analysis, in order to check for moderation between the variables, it is necessary to verify if the interaction term β_3 is associated in the model.

LMM are used to estimate a continuous dependent variable based on associated determinants (the variation in marginal utility in the present analysis). On the other hand, generalized linear mixed models (GLMM) are commonly used to determine the likelihood of a studied condition or behavior to occur or not, thus when the dependent variable assumes either the value 1 or 0 [186].

In this study GLMM are used to estimate the likelihood of respondents to participate in a REC. The investigated question regards the probability of respondent x to participate in a REC as function of the vector of socio-demographic attributes A of the respondent, and the vector of performance attributes P of the alternative REC:

$$Pr_x(Y = 1) = f(A_x, P_{REC}) \tag{13}$$

GLMM account for nonlinear relationships between determinants and estimated outcome of a dependent variable. This relationship is represented in a linear form by means of mathematical transformation allowing for its interpretation as linear combination of the determinant specified in the model [187]. In logistic regression, the logit function is one of the link functions that allows such linear interpretation of the regression model. The logit function is the natural log of the odds (i.e., the ratio between the probability π of an event to occur and the probability 1- π of the event not to occur) and is equal to the linear combination of the predictors and the coefficients estimated by the regression.

$$ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \tag{14}$$

and by exponentiating the two sides of the equation:

$$odds = \left(\frac{\pi}{1-\pi}\right) = e^{\beta_0} * e^{\beta_1 x_1} * \ldots * e^{\beta_n x_n} \tag{15}$$

e^{β_1} represent the odd ratio and could be interpreted in a similar way as in linear regression, as the e^{β_1}increase in the odds of the event to occur for every 1-unit increase in the independent variable x, holding all the other conditions unchanged. Odd ratios could range from 0 to positive infinity, with 1 representing the indifference in the variation of odd related to the variation of the determinants, a direct relation with the determinants in case of values greater than 1, and an indirect relation between odds and determinants in case of values in the range 0 to 1 [187]. As already stated, GLMM are used to estimate the probability of respondents to participate or not in an alternative REC. The perceptions of individual profiles over RECs performance are further considered by introducing the moderator effect of socio-demographic characteristics on the association between REC performance and acceptation. The tested model is specified as follows:

$$Y \sim \beta_0 + \beta_n P + \beta_n A + \beta_m P * A + (1|seed) \tag{16}$$

Where Y is the Boolean acceptance of the REC alternative, P is the vector of performance attribute of the alternative REC, A is the vector of socio-demographic characteristics of the respondent, and β are the coefficients of the linear regression model. Finally, a random intercept was included at the individual level to account for within-subject correlation (1|seed).
Regarding LMM, they have been used to estimate the socio-demographic characteristics having an influence on the respondents' decision models (i.e., homogeneous cluster of individuals sharing similar socio-demographic characteristics having an influence on the variation in the marginal utility associated with different REC performance attributes).
The goal is to define a set of socio-demographic vectors defining types of individuals with similar marginal utility for different levels of performance of the

RECs. These types will be successively used to characterize the resident in order to generalize the decision models approximated by the UTA method.
To do so, LMM have been run for each of the four performance attributes, setting the marginal utility of the attribute as the continuous dependent variable *Y*, and both the performance of the REC for that specific attribute and the socio-demographic characteristics of the respondents as determinant variables. The general form of the fixed slope LMM specified in each of the following models was as follows:

$$U_a \sim \beta_0 + \beta_1 a + \beta_n X + \beta_m a * X + (1|seed) \tag{17}$$

Where U_a is the marginal utility for each performance attribute of the alternative REC, a is one of the performance attributes of the alternative REC (i.e., investment cost, cost savings, emission savings, and self-sufficiency), X is the vector of socio-demographic attributes of the respondent, and β are the coefficients of the linear regression model. Again, the interaction terms are included to assess for moderation effects of socio-demographic characteristics over the performance characteristics of the REC associated with the marginal utility of its attributes.
In the selection and comparison of both GLMM and LMM models, the common marginal R^2 and conditional R^2 metrics to describe the overall fit of the model have been used [188], while the Akaike information criterion (AIC) [189], and Bayesian information criterion (BIC), [190] are used to compare and select the most parsimonious models [191]. Finally, concerning the selection of the determinants associated with a value of the dependent variable, Wald-test and Confidence intervals have been used to exclude the null hypothesis.
Finally, the model has been tested against multicollinearity using variance inflation factor (VIF), setting a threshold of 5 to allow for moderate correlation.
Both GLMM and LMM analysis have been performed in R environment using lme4 package [192,193].

3.3 Generalization of the individual models

Once the socio-demographic factors influencing the probability to accept an alternative REC and the preference model related to the ranking among alternatives has been evaluated, the method aims to generalize such information. The goal is to assess the probability of types of individuals to accept an alternative, and to determine which is the most preferred alternative among the available ones (i.e., the alternative most likely to be converted into functioning among the available set of capabilities at the disposal of the population). First, types of individuals are constructed by combinations of socio-demographic attributes. Second value functions are clustered and the characteristic value function for each cluster is identified. Finally, the socio-demographic attributes of the types of individuals are used to match them with the characteristic value function for each attribute.

3.3.1 Definition of types of individuals

The first step of the generalization of the results from the individual scale is to construct a set of types of individuals sharing similar behaviors. For this purpose, the socio-demographic attributes associated with the probability to accept a REC from GLMM, and having an influence on the preference model of individuals as from LMM models have been used. These socio-demographic characteristics have been combined to determine the types of individuals.

In a second step, the value functions estimated by means of the UTA method have been clustered to determine sets of characteristic value function in order to generalize the decision model estimated at the individual level.

To determine the sets of characteristic value functions the k-medoids algorithm, also called partitioning around medoids (PAM), has been used [194].

K-medoids is an unsupervised clustering algorithm that uses median points to determine clusters. The advantage of this method is the possibility to center the cluster on an actual object of the dataset instead of a virtual one (as in the case of k-means that uses centroids). The algorithm is composed of two phases: a build phase, and a swap phase. In the first one, the first medoid is selected as the one with the smaller distance to all the other points (i.e., the central point in the dataset). The other medoids are then selected by calculating the distance of the candidate point to all the other points and selecting the one with the minimum cost (i.e., the sum of distances to the other points) as the next medoid. Finally, in the swap phase the algorithm calculates the cost of swapping the medoid with all the other non-medoid points, performing the swap with the lowest cost (the sum of distances) if its cost is lower than the current medoid candidate.

The clustering with PAM has been performed using fpc package in R environment [195]. The partition has been based on the marginal utility value (on the y-axis) of the four discontinuity points of each of the piece-wise value functions. As in k-means, the number of clusters must be specified in advance. The pamk function in the same R package has been used to estimate the most suitable number of clusters based on the silhouette metric.

3.3.2 Association of value functions to the types of individuals

Once the clusters are defined, a supervised classification algorithm uses the value functions classes resulting from PAM as labels to calculate the probability of different types of individuals to belong to one cluster of value functions or another. Supervised learning constitutes a category of machine learning in which the classification algorithm is trained with labelled datasets constituted of input-output pairs. In this context, the inputs are represented by the socio-demographic characteristics of individuals, and the outputs correspond to the class of value functions to which the respondents belong. The fundamental goal of the algorithm is to construct a mapping from inputs to outputs, thus enabling it to make predictions or classifications on novel, unseen data. The efficacy of these models is contingent upon the availability and quality of labelled data, which are prerequisites for achieving optimal performance of supervised learning models [196].
Several algorithms are available, being the selection of the proper algorithm a self-standing field of research [197]. In this study, nine classification algorithms have been performed using the caret package in R as a general framework [198]. Caret package makes it possible to perform an automatic grid search for the optimization of algorithms hyperparameters and to select the best ones based on the selected metric [197]. In Table 7 the tested algorithms are shown, together with the R libraries used form caret to run them, and the tuning parameters.

Table 7 Summary of the supervised classification algorithms tested. In the "Library" column the packages called by Caret library are displayed, while in "Tuning parameters" column the hyperparameters for each algorithm are reported.

Algorithm	**Library**	**Tuning parameters**
Naive Bayes	naivebayes	laplace, usekernel, adjust
Support Vector Machine (linear)	LiblineaR	cost, Loss, weight
Support Vector Machine (Polynomial)	kernlab	degree, scale, tau
Rule-Based Classifier (JRip)	RWeka	NumOpt, NumFolds, MinWeights
Rule-Based Classifier (PART)	RWeka	threshold, pruned
C5.0	C50, plyr	trials, model, winnow
Single5.0Tree	C50	None
C4.5-like Trees J48	RWeka	C, M
Penalized Multinomial Regression	nnet	decay

To measure the accuracy of the different models in predicting the expected outcome, a set of metrics could be extrapolated from a confusion matrix. The confusion matrix is a n*n table that shows the number of correct and incorrect predictions compared with the actual classifications in the dataset, where n is the number of classes.

Table 8 Example of a confusion matrix in a two-class classification problem showing the four possible outcomes.

		True class	
		i	j
Predicted class	i	True Positive (TP)	False positive (FP)
	j	False negative (FN)	True negative (TN)

As shown in Table 8, taking a two-label classification problem with the *i* class as reference, the prediction performed by the model can lead to four possible outcomes displayed by the confusion matrix:

- True Positives (TP): observations labelled as *i* class that are predicted in that class by the model;
- True Negatives (TN): observations not labelled as *i* class that the model predicts in the *j* class;
- False Positives (FP): observations misclassified by the model in the *i* class, referred as type 1 errors;
- False Negatives (FN): observations labelled as *i* class that the model classifies in *j* class, referred as type 2 errors.

From the outputs of the confusion matrix, four metrics could be gathered: Accuracy, Precision, Recall, and F-1 score. Accuracy is the ratio of the correctly predicted classification over the total classification made by the model, and it is calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{18}$$

Precision focuses on the ability of the model to distinguish positive from negative classifications, and is calculated as:

$$Precision = \frac{TP}{TP+FP} \quad (19)$$

Recall focuses on the ability of the model to predict positive samples and is calculated as:

$$Recall = \frac{TP}{TP+FN} \quad (20)$$

Finally, the F-1 score combines both Precision and Recall and is calculated as:

$$F-1\ score = \frac{2*Precision*Recall}{Precision+Recall} \quad (21)$$

F-1 score ranges from 0 to 1, with values close to one representing better performing models both in Recall and Precision. The calculation of Precision, Recall, and F-1 score depends on the reference class considered as positive. For multi-class classification problems, it is possible to calculate their values using a one-vs-all approach translating the multi-class problem in a sum of binary-classification models. In this way, each time a class is taken as the reference positive one, and the confusion matrix is generated using all the other classes as the negative class. In Figure 15 a graphical representation of the one-vs-all approach is displayed.

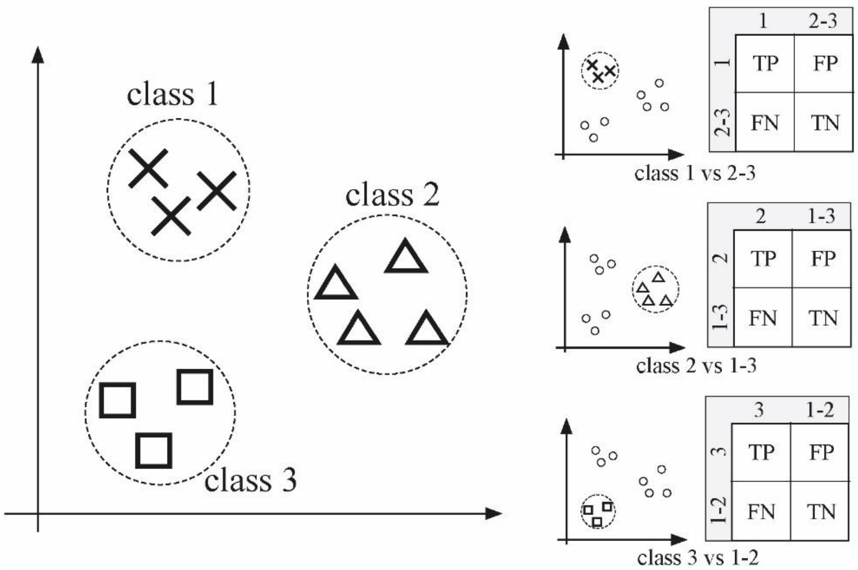


Figure 15 Graphical representation of the translation of the 3-class classification problem in 3 binary classification problems following the one-vs-all approach.

With such an approach, the calculation of Precision and Recall could be calculated as an average of the values for each class taken each time as the positive one:

$$Mean\ Precision = \frac{\sum_{i=1}^{n} Precision_i}{i} \quad (22)$$

$$Mean\ Recall = \frac{\sum_{i=1}^{n} Recall_i}{i} \quad (23)$$

Where $Recall_i$ and $Precision_i$ are the two metrics calculated for each of the *n* classes using the *i*-th class as the reference true class. Regarding F-1 score, in case of an unbalanced number of observations in each class the weighted F-1 score could be calculated by weighing each class specific F-1 score by the number of samples in that class as expressed in the following equation:

$$Weighted\ F-1\ score = \sum_{i=1}^{n} w_i * F - 1score_i \quad (24)$$

With w_i the ratio between the number of samples in class *i* and the total samples in the dataset, and F-1 $score_i$ the value of F-1 score calculated with class *i* as the true reference class.
Accuracy provides an overall evaluation of the predictive capacity of the model, while F-1 score is a single metric combining Precision and Recall. In this study the selection of the best performing classifier is based on the average of these two metrics.

3.4 Spatialization of acceptability and preferences

Once the generalized decision model (the probability of being assigned a specific set of characteristic value function or another) has been estimated for different individuals type, the distribution of these types across the city is estimated by means of the generation of a synthetic population. Further on, different RECs alternatives could be generated at the urban scale. Finally, by matching the probability to accept a REC and the preference models estimated in the previous section with the distribution of types of individuals across the city, the percentage of population likely to accept to participate in the generated RECs as well as the most preferred alternative by the overall population could be estimated.

3.4.1 Generation of a synthetic population

First of all, the distribution of the types of individuals is simulated by generating a synthetic population coherent with the socio-demographic distribution of the real population in the area. Popular in the transportation research field, Population synthesis refers to the process of creating a simplified microscopic representation of an actual population that matches the aggregated statistical measures of the actual population [2,199]. Population data are usually provided with a coarser aggregation at a higher spatial resolution (marginal distributions), while disaggregated at the individual level at a lower spatial resolution such as the regional or country scale. The population synthesis process allows to reconstruct the individual vectors of socio-demographic characteristics (gender, age, education level, etc.) of a population from the marginal distribution of socio-demographic characteristics of the population in a defined area as provided by census data. To generate a synthetic population, a Population synthesizer (the software that implements the process) requires two types of input data: (i) a disaggregate sample of the population's vectors of socio-demographic attributes at the individual level, and (ii) the aggregated marginal distributions of socio-demographic characteristics of the population in the area of analysis [200]. By applying data fusion techniques, the synthetic population is generated expanding the disaggregated sample dataset to match the marginal distribution of socio-demographic characteristics in the area (Figure 16). The population's aggregated marginal distributions represent the constraints for the population synthesis expansion procedure and might include social, demographic, and economic variables useful to describe the population living in a specific area under analysis (Census Block, Block Group, Census Tract, County, Metropolitan Statistical Area, or at any specified geography for which data are available).

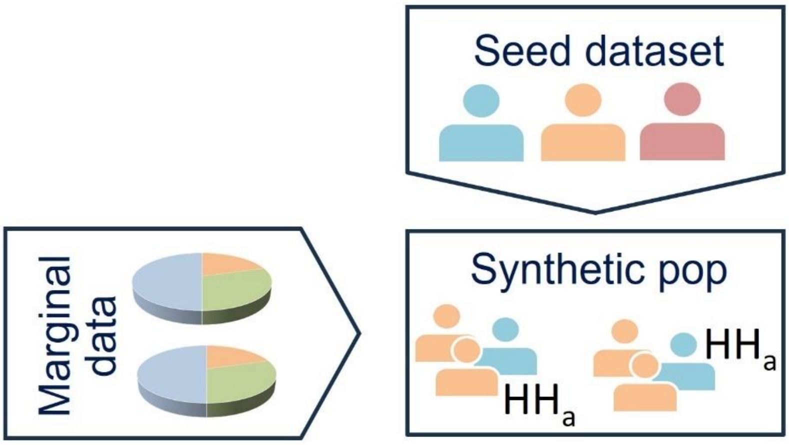


Figure 16 Graphical representation of the construction of the synthetic population expanding the seed dataset to match the marginal distribution of the population in the area

The disaggregated sample at the individual level is typically obtained by census data and other specific surveys to describe the population according to the goal of the study for which a synthetic population needs to be produced. This sample should contain socio-demographic characteristics corresponding to the marginal distribution used as controlling variables (age, sex, occupation, etc.), but it could also contain information that are not registered in the marginal distribution dataset, thus that are not used as controlling variables.
Numerous are the software available to generate synthetic populations, among them, PopulationSim [200], is an open-source Python-based tool that is part of the collaborative framework ActivitySim [201]. PopulationSim accepts four types of comma-separated values (csv) files:

- **persons and households seeds**. These two datasets contain the individual-level sample of socio-demographic characteristics vectors of the population derived from census survey at a regional scale. In the Italian context this statistical microdata could be found among the publicly available data provided by the National census institute (ISTAT);
- **marginal controls**. This is the dataset of aggregated marginal distributions of socio-demographic variables presented at a more granular resolution. National census data could be used to gather this data;
- **geographic crosswalk file**. This file specifies the hierarchy of geographies used to allocate the population specified at an upper

geography (persons and households seeds) to a lower geography (marginal controls);

- **controls file**. This file is used to link the seeds to the marginal controls by specifying the match between variables in the datasets and by setting the importance of the marginal control with respect to each other.

The main output of the synthetic population is an artificial population whose distribution is coherent with the actual distribution of the real population living in a specified area with individual-level granularity of data.
Regarding the persons and household seeds, the publicly available "Aspect of daily life" ("Aspetti della vita quotidiana", AVQ) dataset provided by ISTAT has been used. AVQ dataset is a yearly survey conducted on a sample of about 20,000 households and 50,000 individuals, with its last release referring to the 2021 survey [202]. The survey is part of an integrated system of multi-scope analysis on households and focuses on several aspects of families and individuals' daily life, collecting a total of 743 different variables provided as individual's vectors of socio-demographic characteristics.
As marginal controls dataset, the data collected during the last national census of 2021 have been used [203].
In terms of granularity, the seeds dataset is aggregated at the regional level, while the marginal control is referred to the census track scale (in Italian "Sezione censuaria"). The allocation of the population from the upper geography (seeds at the regional scale) to the lower one (marginal controls at the census track scale) follows the hierarchical structure (from right to left) expressed in the geographic crosswalk, a sample of which is presented in Table 9.

Table 9 Sample of the geographic crosswalk csv file from smaller census track (SEZ21_ID) to Regional scale (REG). Turin is divided in 6915 census tracks.

Progressive number	**SEZ21_ID**	**COM**	**REG**
1	12720000001	272	1
2	12720000002	272	1
3	12720000004	272	1
…	…	…	…
6915	12728888888	272	1

As displayed in Table 9, and according to Italian census data, region (REG) 1 refers to the Piedmont and Valle d'Aosta regions, municipality (COM) 272 is the identifier for the city of Turin, while SEZ21_ID refers to revised tessellation of the national territory in the census tracks according to the last national census survey (2021), with a total of 6915 census tracks in the city of Turin.
Finally, the controls file allows to specify the link between seeds and marginal controls datasets. A sample of a controls csv file is reported in Table 10.

Table 10 Sample of the controls csv file specifying the links between seeds and marginal controls datasets. The number of individuals in the seeds ("seed table") are expanded to match each variable distribution in the marginal controls dataset ("control field") at a specific scale ("geography") with different matching priorities among variables ("importance").

target	**geography**	**seed table**	**importance**	**control field**	**expression**
Num_HH	SEZ21_ID	households	1000000	HH_BASE	(households.wgtp > 0) & (households.wgtp < np.inf)
Num_pp	SEZ21_ID	persons	50000	Pop_Base	(persons.wgtp > 0) & (persons.wgtp < np.inf)
HH_size_1	SEZ21_ID	households	10000	HH_size1	households.NP == 1
HH_size_2	SEZ21_ID	households	10000	HH_size2	households.NP == 2

With reference to Table 10, each row specifies a link between variables in the seeds datasets (persons or households) with those in the marginal controls one to be matched in the seeds expansion. In particular, the controls file specifies the marginal distribution of the specific socio-demographic characteristic to be used as a constrain to the expansion ("control field" column), the spatial scale ("geography" column) at which the marginal distribution of the variable refers, and the seeds dataset ("seed table" column) to which the control is applied during the expansion of the seed dataset itself. The controls file also allows to determine the weight of priority ("importance" column) with which PopulationSim attempts to match the control: a higher importance control will be attempted to be matched with higher priority than a lower importance one, with more deviations allowed in the final distribution of lower importance controls specifications. In this sense, higher importance controls could be interpreted as harder constraints, while lower importance ones as softer constraints. Finally, in the "expression" column the link between the marginal controls to the households or persons seeds is expressed in the Python syntax. For instance, taking into consideration the 3rd row in Table 10, according to this control input, PopulationSim would take the households composed by one person (households.NP == 1) from the households dataset and try to populate the synthetic population produced as an output with a number of households of this size as close as possible to the amount registered in the marginal control dataset (variable HH_size1) in each specific census track (SEZ21_ID). Furthermore, in the process of expanding the individual vectors to match the marginal distributions, the algorithm will give higher priority in matching first the total number of households (first row of control with importance 10^6), then the total amount of population (second row of control with importance 50000), and finally the number of households composed by one person, allowing for more flexibility in matching such controls with a lower importance.

3.4.2 Simulation of alternatives at the city scale

To calculate the potential reduction of buildings energy consumption at the urban level, a first evaluation of the status quo has to be performed. To do so, different strategies have been developed in literature: some studies use a top-down approach starting from aggregated data at urban or national level and then disaggregate them by means of statistical methods [204], while others opt for a bottom-up approach in which data from individual buildings are extrapolated and elaborated [205]. The bottom-up approach usually takes advantages of the generation of archetypes in order to define the parameters used in the modelling such as construction materials, insulation values (U-values), window-to-wall ratio (WWR), infiltration levels, and mechanical system efficiencies [205,206]. These archetypes are largely based on the period of construction and building typology [205,207,208] and are built on several possible sources such as EU projects [209], national building libraries [210], and expert opinions and/or past studies. The use of archetypes is a generalization used by energy modelers to overcome inconsistencies and lack of data in available datasets, especially to retrieve non-geometrical data such as building assembly, system information and occupancy schedules [211].
Once the status quo is modelled, alternative intervention scenarios could be proposed using different approaches as mentioned in Section 2 (such as linear programming and scenarios comparisons).
In the present study, the new alternatives have been generated based on the alternatives used in the questionnaire. The performance of the best performing

alternative in terms of CO_2 emission savings has been partially modified in order to evaluate the degree of acceptability of public oriented solutions.

3.4.3 Estimation of acceptability and of the preferred alternative

In the last step of the method, the probability to accept a REC alternative as well as the preferences of the types of individuals are matched with the newly constructed alternatives at the city scale. In this way, the percentage of population that might accept different REC configurations, as well as the most preferred alternative among them is assessed.
In particular, the percentage of population likely to accept a REC solution has been estimated according to equation (25):

$$P_i(Y = 1) = \frac{\sum_{x=1}^{n} Pr_{x,i}}{n} \tag{25}$$

Where P_i is the percentage of the entire population likely to accept the alternative i, $Pr_{x,i}$ is the probability of individual x to accept alternative i, and n is the number of individuals (total population). Furthermore, based on individual probability, it is also possible to estimate the probability density function (PDF) of participation following Poisson generalization of binomial distribution.
To calculate the preferred alternative among the possible solutions, a probabilistic approach has been used. In particular, the use of supervised classification algorithms (Section 3.3.2) allowed to estimate the probability of each type of individual to be assigned a characteristic set of value functions. Therefore, the total utility of a REC alternative for different types of individuals has been calculated by applying equation (26):

$$U_{tot,i,x} = \sum Pr_{c,a,x} * U_{m,c,a,i} \tag{26}$$

Where $U_{tot,i}$ is the total utility of alternative i for an individual x, $Pr_{c,a,x}$ is the probability for the individual to be assigned one of the characteristic value function c, relative to attribute a, while $U_{m,c,a,i}$ is the marginal utility for attribute a of alternative i, calculated using value function c. In this way, the total utility of an alternative for a specific type of individual is the weighted sum of the marginal utilities, where the weights are probabilities of participation to one clusters of characteristic value functions. The most preferred alternative for each individual is the one scoring the highest total utility. Consequently, the most preferred alternative across the population is the one ranking first for most of the inhabitants.

This last step of the method aims to highlight the areas of the city in which it could be expected that a lack of interest (or desirability) in joining a REC will arise.
Furthermore, the degree of support from a third party (either public bodies or ESCOs) necessary to promote the preferability of these alternatives over those potentially preferred from a private initiative point of view could be estimated. Such results will support the consideration of possible variation of the enabling framework aiming at promoting the adoption of RECs at the local level in order to increase the acceptability of such interventions. A second possibility could be to identify possible sub-optimal solutions from the public interest point of view, such as the preference of inhabitants for more profit-oriented solutions. Again, the results provided could support the Decision-maker intervention in order to mitigate this less desirable effect in favor of solution that might enhance public utility, such as the maximization of GHG emission reduction.

Chapter 4

Case study

The method proposed in Chapter 3 has been applied to the case study of the city of Turin in the northwest of Italy (Figure 17). This section presents the results of the application.

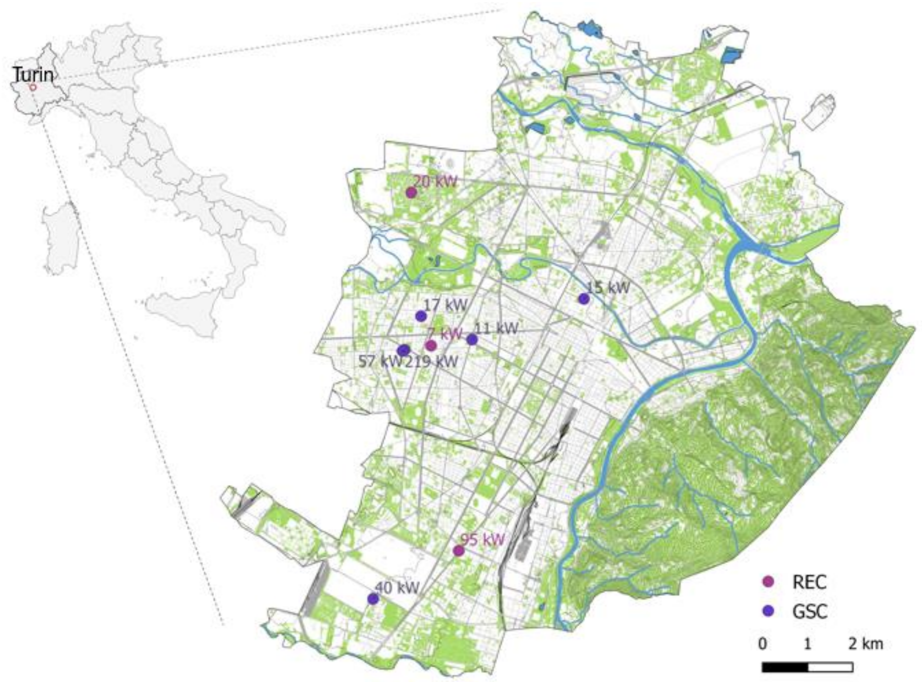


Figure 17 Turin and RECs already formed in the city. Renewable energy community (REC) and Group of self-consumers (GSC). In the labels the installed kW until December 2024 (source: author’s elaboration on [212]).

In recent years Turin has been selected as one of the showcase cities for the EU Mission 100 climate-neutral cities by 2030 – by and for the citizens [4,5]. The Mission specifically targets Scope 1 (direct emissions from buildings, industry, transport, waste treatment, agriculture, forestry, and other activities within the city boundaries), and Scope 2 (GHG from grid-supplied energy) greenhouse gases emissions from cities [5], with the aim to support the transition of the selected cities toward climate neutrality as experimental hubs. In this attempt the EU Mission highlights the necessity to promote bottom-up initiatives from citizens as well as new forms of governance to coordinate different streams of funding (public and private) to facilitate a systemic transformation to climate neutrality [4]. Among the potential innovative bottom-up initiatives that could help direct these streams, REC could play an important role as stressed in Chapter 2.

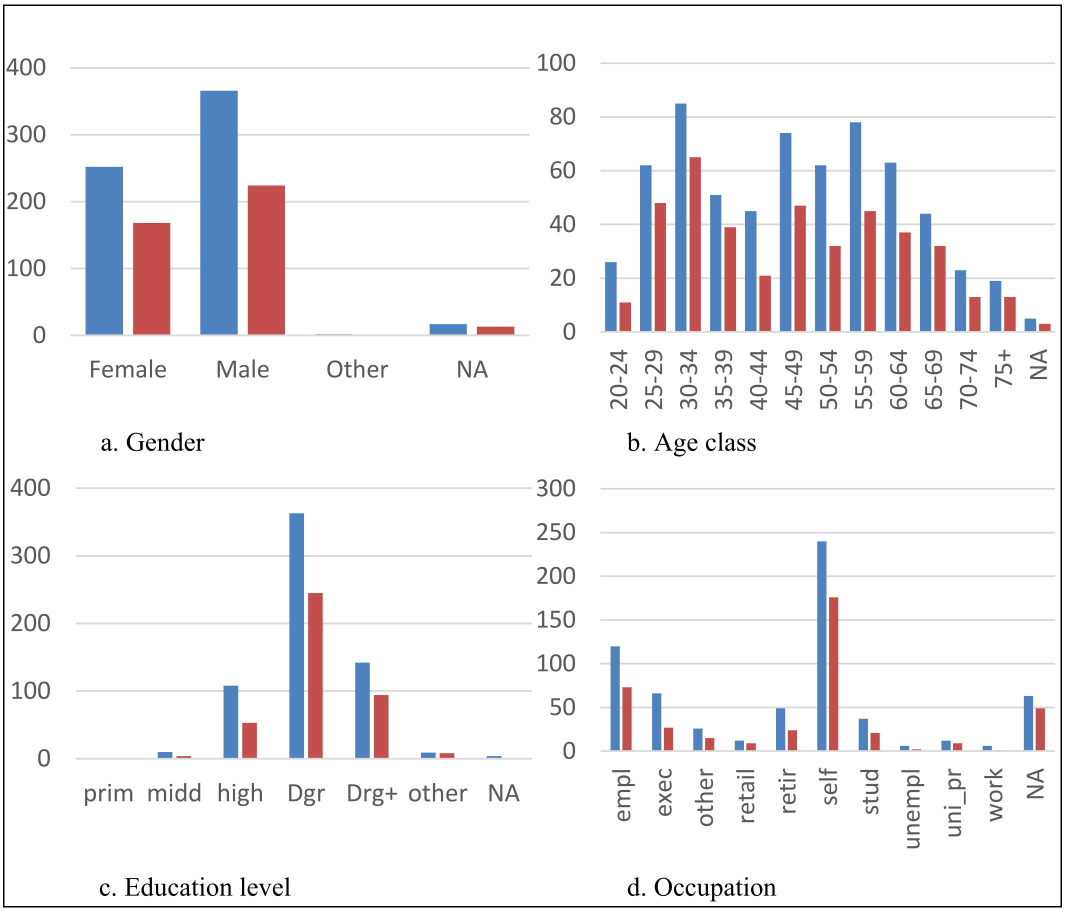
From their early legal definition, several RECs configurations have been already implemented in Turin as shown in Figure 17 [212]. The selection of this case study could then represent a good choice to examine the potential synergy between the public and the private sphere in the promotion of RECs penetration.

This section is organized as follows. Section 4.1 presents the results of the survey campaign. In Section 4.2 the respondents' value functions are estimated using the UTA method. In Section 4.3 GLMM is used to estimate the individuals' probability to accept a REC based on its performance and the respondents' socio-demographic attribute. In Section 4.4 the socio-demographic characteristics having an influence on individuals' value functions are analyzed. In Section 4,5 the socio-demographic characteristics individuated in the previous section (4.4) are used to construct the type of individuals to which the characteristic value functions are assigned in order to generalize the preference model. Finally, in Section 4.6 the synthetic population is generated according to the marginal distribution of socio-demographic characteristics.

4.1 Collected responses from the population survey

The questionnaire was distributed from October 2024 to June 2025 by email. Various channels have been used to distribute it (word of mouth, QR code on flyers, associations, social network groups, public professional registers) using a convenience sampling technique. Using this approach, participants in the survey were not selected using a random or systematic method, but rather based on availability and ease of data collection, at the expense of the possibility of generalizing the results of the analysis [213]. Although this being recognized as a limitation, the selection of the convenience sampling technique was justified by budget and time constraints, as well as the pilot nature of the study, which was intended to test the method. It is also worth noticing that the participation to the survey was completely voluntary and no forms of rewards nor compensation has been given to the respondents, with a high chance of increasing the reliability of the responses on one hand, but a substantial limitation in the response ratio on the other hand (on top of that, a certain number of missing data are present in the collected observations). It could be expected that the introduction of a rewarding mechanism in the administration of the survey might result in a higher response ratio.

A total of 1479 questionnaires were collected (the exact total number of people exposed to the questionnaire could not be estimated due to the way it was distributed). The incomplete questionnaires were excluded (i.e., those questionnaires in which the respondents did not reach the final page, meaning that questionnaires with missing data were accepted) yielding a total of 637 valid responses collected. In Figure 18 some preliminary analyses are presented to describe the sample of respondents, both for total number of questionnaires collected (blue bars) and focusing on those responses coming from individuals residing in the city of Turin (red bars).



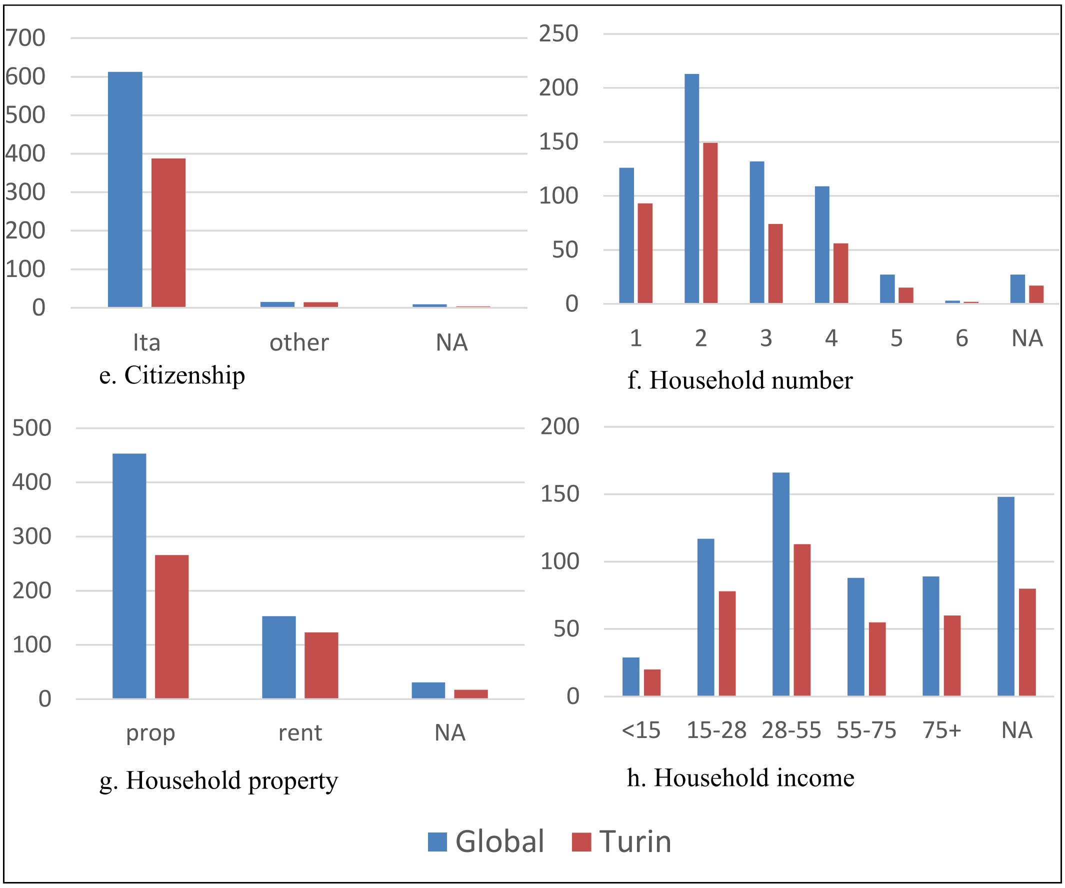


Figure 18 Distribution of socio-demographic characteristics of the respondents to the population survey

Regarding the information about the location of their households, a total of 380 respondents reside in the city of Turin for which other information could be retrieved, such as the neighborhood where they live and the primary HV/MV substation to which they belong. The spatial distribution of these 380 respondents living in the city of Turin is shown in Figure 19 and Figure 20, subdivided in neighborhoods and HV/MV substations respectively.

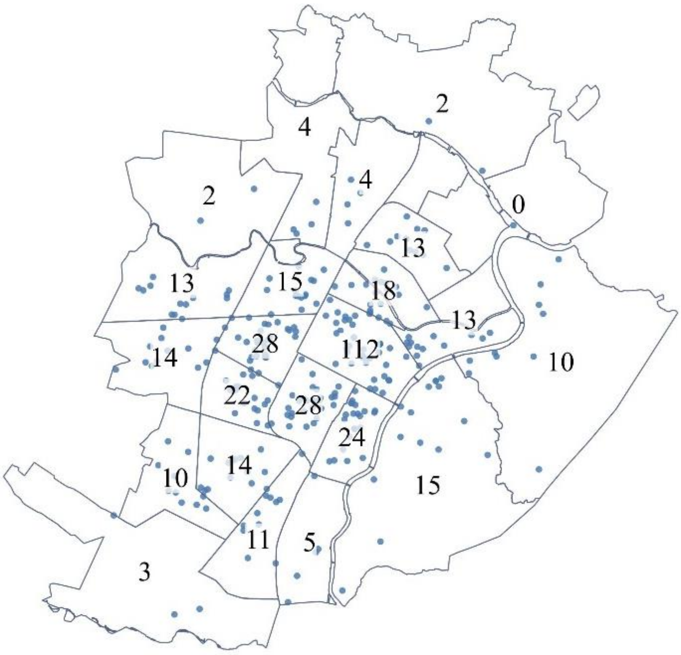


Figure 19 Spatial distribution of the 380 respondents living in Turin and number of respondents in each neighborhood.

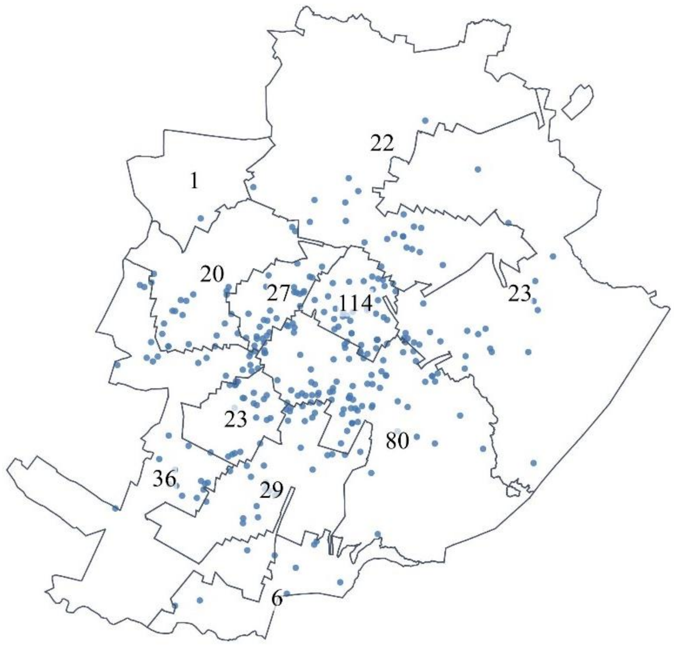


Figure 20 Spatial distribution of the 380 respondents living in Turin and number of respondents in each HV/MV substation.

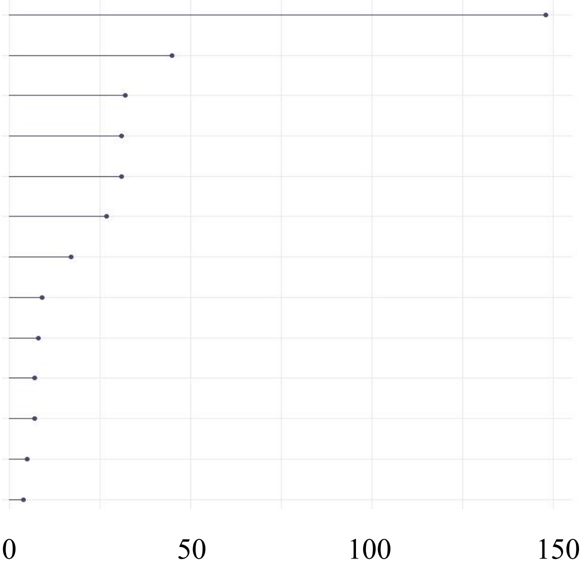
Most of the questionnaires were collected from respondents residing in the central areas of the city of Turin, while a lower number of questionnaires were collected from respondents residing in the peripheral neighborhoods, and from HV/MV substations serving these areas. This is a direct consequence of the convenience sampling technique resulting in an unbalanced number of respondents from different areas of the city. Considering the experiment as a pilot case to test the application of the method, such limitation was accepted at the expense of the possibility to generalize the results obtained.
Finally, 26 respondents did not give their place of residence (or a point close to it) but agreed to give their postcode. This made it possible to assign them to a neighborhood and, consequently, to one of the eleven HV/MV substations present in the city of Turin. Including these last 26 respondents, the final number of questionnaires collected in the city was 406. For these respondents the HV/MV substation was identified by means of the online tool provided by the Energy services manager mapping the substations on the national territory [212]. In particular, it is worth recalling that the Italian legal framework on RECs requires the energy production system and the delivery point of the same configuration to be located under the same HV/MV substation [214].

4.1.1 Imputation of missing data

A relevant aspect to be checked in the dataset of respondents to the questionnaire is the presence of missing socio-demographic data. With the necessity to choose the best approach to solve the problem without reducing the statistical power of the analyses carried out on it. A simple strategy would have been to delete the incomplete observations, but this would have significantly reduced the number of entries in the already rather small dataset of responses collected. The number and frequency of missing data for each socio-demographic variable among respondents is presented in Table 11.

Table 11 Occurrence of missing values in the database of respondents

	Missing values	
variable	(%)	(num)
HH_income	23.2	148
HH_age	7.06	45
Location	5.02	32
HH_property	4.87	31
HH_type	4.87	31
HH_size	4.24	27
Gender	2.67	17
Citizenship	1.41	9
Occupation	1.26	8
HH_choice	1.10	7
HH_retrofit	1.10	7
Age_class	0.78	5
Edu_level	0.63	4



Out of the 637 collected questionnaires, 430 respondents (67.5%) provided the complete set of socio-demographic information requested. The majority of variables exhibited a percentage of missing values lower than 8%. The only exception to this is the Income level of the Household (HH_income) variable, for which 23.2% of the respondents did not disclose this information.
As previously mentioned, the questionnaire permitted respondents to specify either their postcode or to select a point on the map in close proximity to their approximate location (or both). In Table 11, the variable named "Location" was used to indicate whether the respondent provided the postcode or pointed a location on the map (or provided both).
The presence of missing data represents a substantial problem for further analysis. It is common practice to include in the analysis only those observations for which no missing data are present, thus performing complete case analysis [215,216], with the risk of reducing the statistical power of the analysis [216]. This straightforward strategy, commonly referred to as "case deletion" [217] is deemed unsuitable for the case study under investigation, due to the already limited size of the dataset of respondents. In order to evaluate alternative approaches in dealing with the presence of missing data, it is first necessary to determine whether particular patterns exist between variables, and whether particular types of respondents were likely to not provide one type of information or another.

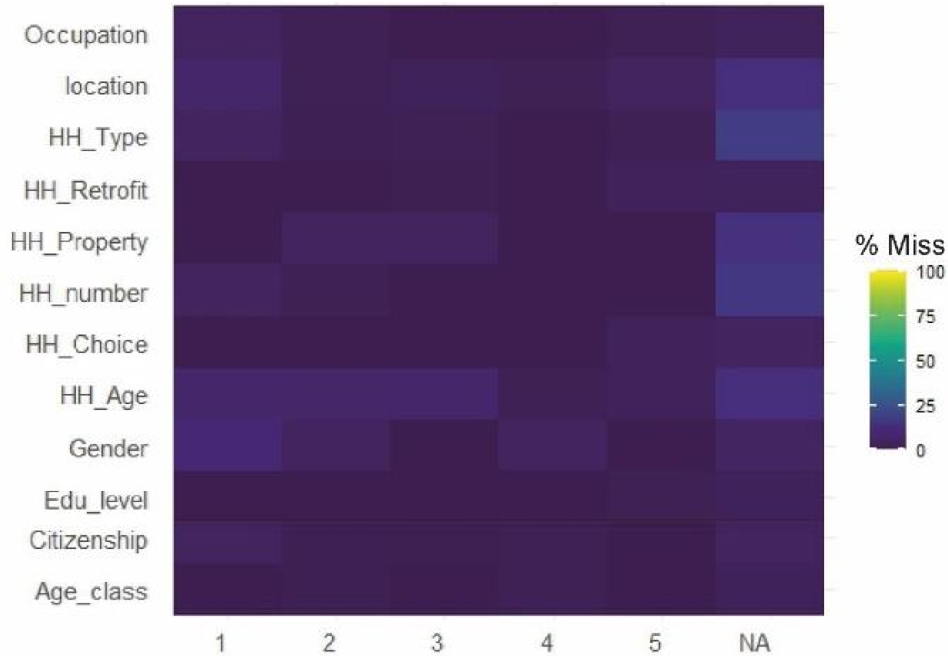
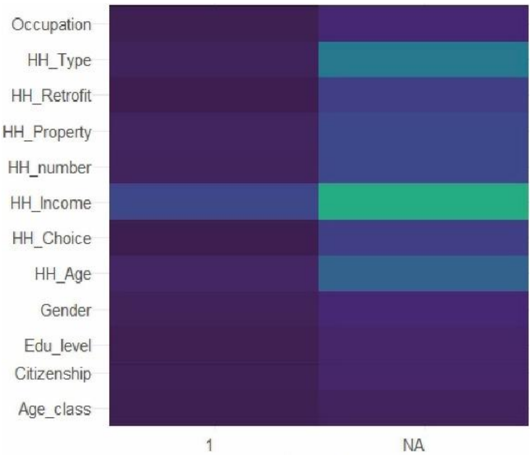
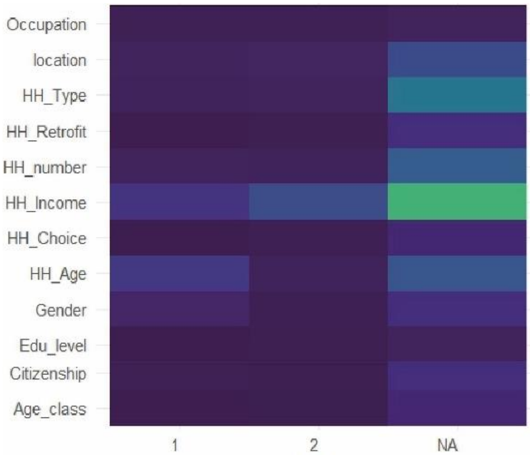


Figure 21 Co-occurrence of household income missing data and the remaining socio-demographic characteristics of the respondents (5 modalities/not provided)

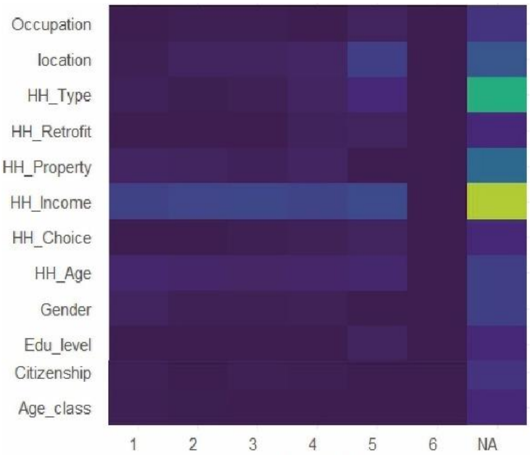
For this purpose, Figure 21 presents the rate of item nonresponse [217] for the various household income classes (HH_income) against the other variables. The figure shows no specific nonresponse pattern between the different levels of income of the household and other socio-demographic variables of the respondents.
Furthermore, Figure 22 provides a series of heatmaps to illustrate the occurrence of nonresponses for each categorical value of one socio-demographic variable compared to the others.



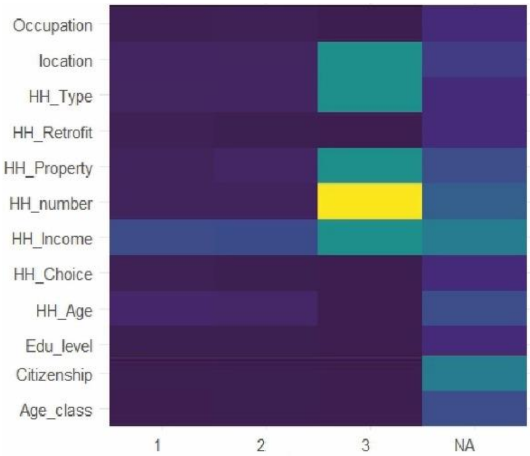
a. Location of the household
(provided/not provided)



b. Property of the household
(owned/rented/not provided)

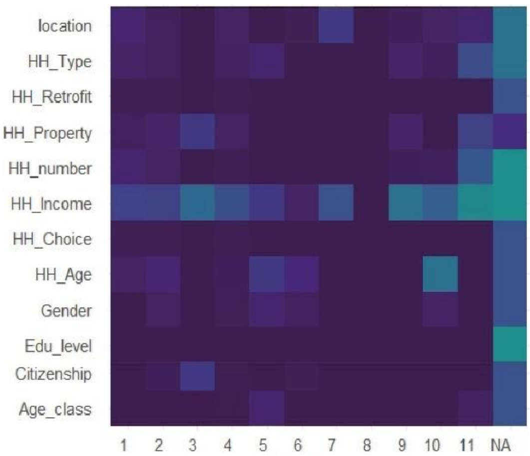


c. Number of people in the household
(number/not provided)

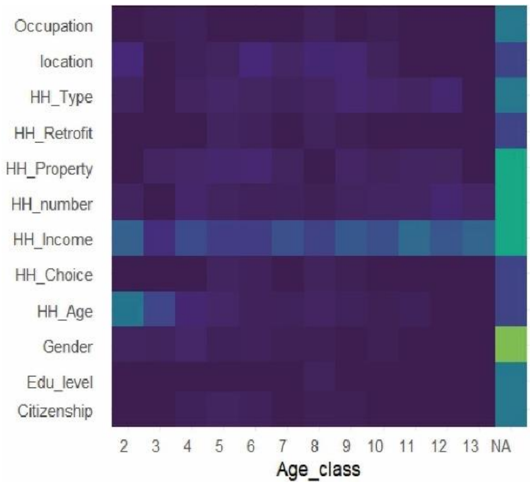


d. Gender
(male/female/other/not provided)

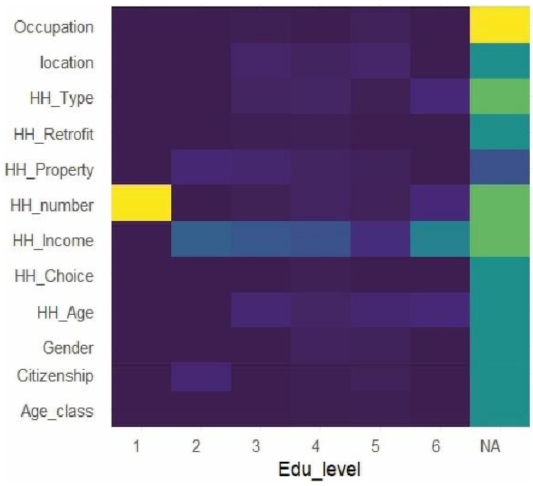
e. Citizenship
(Italian/other/not provided)



f. Occupation
(11modalities/not provided)



g. Age class
(13 modalities/not provided)



h. Education level
(6 modalities/not provided)

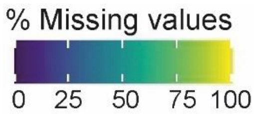


Figure 22 Heatmaps of the percentage of missing values co-occurrence for the different socio-demographic variables

The highest ratio of missing values is recognizable in the "Non-Available" column (NA), meaning that the respondents that did not provide information regarding one socio-demographic variable (the "NA" on the x-axis in Figure 22) also did not disclose several other information. This could be interpreted as indicative of a general concern for privacy reasons, as opposed to being indicative of a tendency among certain respondents to omit specific socio-demographic data. Only exceptions to this pattern are represented by two high ratios of missing values for Education level class 1 (elementary) and Gender class 3 (non-binary) who present high values of missingness in combination with the information regarding the size of the household (HH_number) in Figure 22.d and Figure 22.h. In this case it is important to consider the limited number of respondents belonging to these two classes. These few respondents did not provide information also for the size of the household, explaining the high ratio of missingness but a low absolute number of missing data compared to the total of the responses. For this reason, this anomaly could be considered negligible, not representing a specific pattern in the distribution of missing data.
Finally, focusing on the distribution of missing data for different classes of variables in combination with household income (row named HH_income in Figure 22), it is possible to note that the ratio of missingness is fairly distributed among the classes.

The highest values are again observed in the NA column for each variable, thereby confirming the privacy motivation behind the choice to not disclose socio-demographic information in general and not for specific aspects. In light of these considerations, it is possible to exclude the existence of a specific pattern in the refusal of certain respondents to disclose specific information depending on their other socio-demographic characteristics.

It is common to classify missing data [216,217] as (i) missing completely at random (MCAR), (ii) missing at random (MAR), and (iii) missing not at random (MNAR). In the case under investigation, it is possible to interpret the missing data concerning respondents' socio-demographic characteristics as MCAR [218] on the condition that there is no systematic difference between the missing and observed data.

Under this assumption, it is possible to use multiple imputation methods to estimate missing values for a given unit (a respondent) based on the values observed for that given individual and similar pattern of data observed for other individuals [219]. This approach enables the analysis to proceed without discarding incomplete observations, thereby preventing the loss of statistical power that results from reducing the number of unit responses and preserving the variability within the data [215,217,218].

Multivariate imputation by chained equations (MICE), also referred as sequential regression multivariate imputation [220], is one popular and well-known technique [216,220,221] belonging to the family of multiple imputation methods [220,222]. MICE procedure (for a more detailed description of the steps followed by the algorithm the reader might refer to [219,220]) starts with a simple imputation of missing values in the dataset interpreted as placeholders; then the algorithm fits a series of regression models for a specific variable containing missing values by treating it as a dependent variable and the others as independent ones. This process is looped across the other variables containing missing data and the probable values are predicted to generate a complete dataset. Following the standard multiple imputation procedure, the imputing process is then repeated m-times to generate m-datasets of plausible complete observations [217]. These m-datasets are then used to perform the desired complete case analysis, pooling and combining results from all of the m-dataset following Rubin approach [220,222].

Regarding the dataset under analysis, the MICE package in R [222,223] has been used to perform the multiple imputation of missing data. The first step of the process was to select the features to be used in the imputation, thus the variable used by the algorithm to recognize patterns of similarity among the individuals. Including auxiliary variables that are not further used in the subsequent analysis could improve the performance of the imputation [219,224]. For this reason, all the socio-

demographic variables have been included together with the "Views of the World" set of qualitative responses to account for common perceptions among individuals. Other variables have been excluded from the imputation process as they were considered not related to socio-demographic missing data (i.e., they were not used to recognize the patterns in the similarities among respondents). These excluded variables were household type (HH_type), household age (HH_age), and if the dwelling had undergone a retrofit intervention (HH_retr). In order to account for potential differences or similarities across the respondents' various locations, the postcodes were included in the analysis. This latter variable was filled with a dummy 00000 value in case no information was available (i.e., the respondent did not specify neither the postcode nor the geographical coordinates of the residence). Finally, the other socio-demographic variables included in the process were: (i) Gender, (ii) Age, (iii) Education level, (iv) Household number, (v) Household property, (vi) Household income, (vii) Occupation, (viii) Citizenship.
Further specification to be defined in the algorithm is the number of m-datasets to be generated through the process. Regarding this choice, literature does not agree on a common selection strategy. According to [225], the process has been repeated 5 times to produce 5 different datasets. In Figure 23, the percentage distribution of the categorical variables with the highest number of missing values is presented for the original dataset (purple bar) and for the 5 imputed datasets (the variables with less than 2% of missing values are not showed due to their limited variation across different imputed datasets).

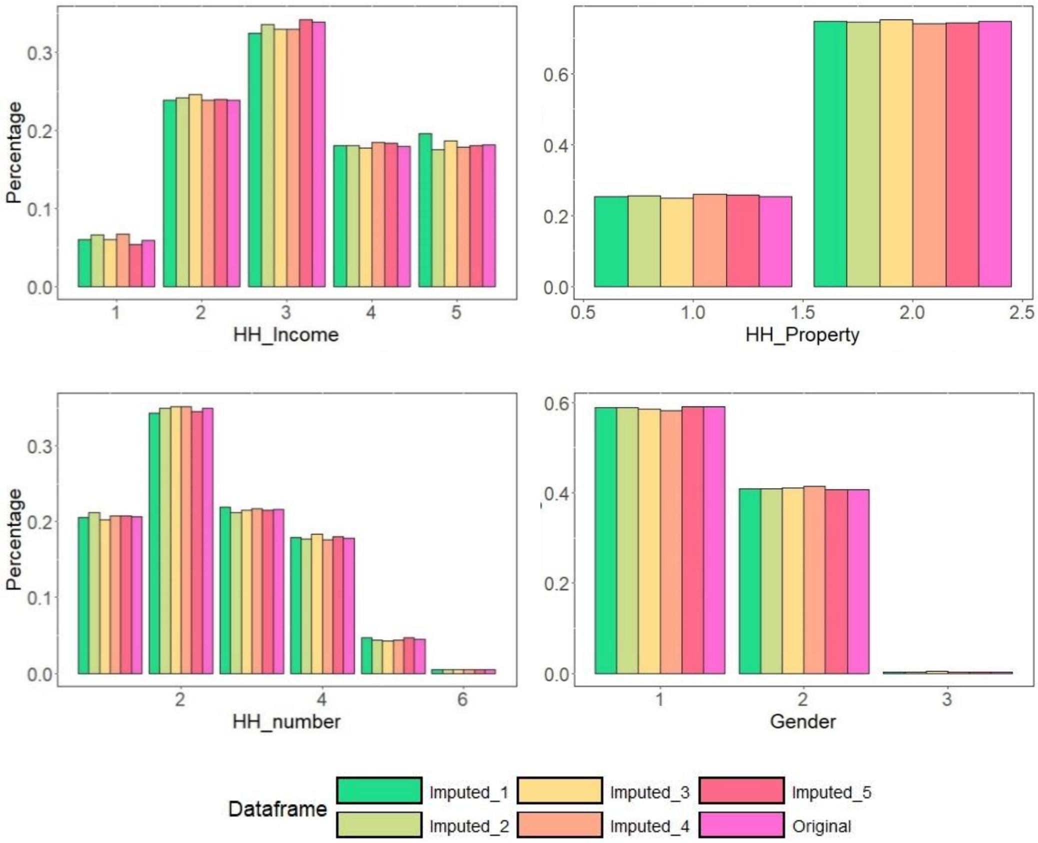


Figure 23 Comparison of frequency of imputed variables in the imputed datasets compared to the original one (with missing values).

Figure 23 shows an overall similarity between the frequency of occurrence of different classes of variables in the 5 imputed (presenting only complete observations) datasets and in the original one containing missing values. As anticipated, the most significant discrepancy was observed in the household income variable, attributable to its highest proportion of missing values among the original dataset's variables. The remaining variables exhibited minimal disparities in their overall distributions. Considering this similarity in the distributions of imputed data across variables and the increase in computational burden related to the MICE process of pooling and combining results from several analysis on a multiple number of datasets without sensibly increasing the meaningfulness of results, the study has been limited by selecting only one dataset on which to further perform the data analysis.

4.2 Individuals' value functions estimation

The estimation of respondents' value function has been performed by means of the MCDA library in Python environment [226]. As stated previously, the learning set to train the algorithm was built based on the responses gathered from the questionnaires for each individual respondent. Considering that several of them expressed only one alternative for which they would have been interested in participating either answering "Yes" or "Maybe", a necessary threshold of two minimum preferences expressed has been set to have at least one pair of ordered alternatives on which to train the algorithm. This limitation resulted in the reduction of the number of useful questionnaires to 556 valid ones.

The number of alternatives considered acceptable provided by several respondents (i.e., for which they answered "Yes" or "Maybe in the possibility to participate) was found to be quite limited, thus reducing the size of the learning set on which the algorithm would have been trained (i.e. a minimum of 2). In order to mitigate this issue, a number of additional dummy alternatives were introduced to enhance the initial learning set provided by each respondent.

It is worth noticing that, even though these dummy alternatives do not add additional information regarding the preference of the respondents, they have been introduced for modelling reasons. In particular, UTAstar method is based on linear programming technique, with as objective function the minimization of the sum of two errors introduced to account for overestimation (σ^+) and underestimation (σ^-) when calculating the global value of the alternatives of the learning sets. Being such error estimated over pairs of consecutive instances in the learning set [181] its expansion by including additional alternatives (some of which dominated) is instrumental to refine the estimated shape of the value functions.

These dummy alternatives were generated in a way that the respondent's preference for these alternatives over others in the data set would have been irrational, therefore, their introduction in the learning set would not have been in contradiction with preferences expressed by the respondents. In total, 11 dummy alternatives have been generated and introduced in each respondent's learning set: 5 "dominated" alternatives, 4 "worsened" alternatives, and 2 alternatives generated with the best and worst performance levels in each attribute, respectively the "Best" and "Worst" alternatives. The "dominated" alternatives (i.e., those demonstrating substandard performance across all the descriptive criteria) for each of the 8 initial alternatives utilized in the questionnaire were identified from the list of case studies analyzed in the literature review (i.e., the same from which the eight alternative subsets used in the questionnaire were extracted). These alternatives were then introduced in the preference order expressed by the respondent. In particular, for each alternative y_n the dominated alternative y_{n-d} was identified and added in the learning set by specifying a preference relation $y_n \succ_x y_{n-d}$. Furthermore, the four additional "worsened" alternatives have been introduced by duplicating the last ranking alternative in the learning set of selected alternatives ($y_I \succ_x y_J$) and worsening one performance attribute at a time to the minimum level of performance across the 8 original set of alternatives of the questionnaire. This step generated four additional dummy alternatives denoted as (i) W_inv when the investment cost attribute was worsened, (ii) W_csav for the dummy alternative for which the cost saving attribute was worsened (iii) W_esav for the emission saving attribute detriment, and (iv) W_ssi for the alternative in which the self-sufficiency attribute was set to the minimum. Finally, the two additional alternatives "Best" and "Worst" have been introduced as the most and least preferred alternatives (being them dominating-all and dominated-by-all the alternatives respectively).

In Figure 24, the process of constructing the final learning set from the original one, retrieved from the questionnaire for a random respondent is illustrated.

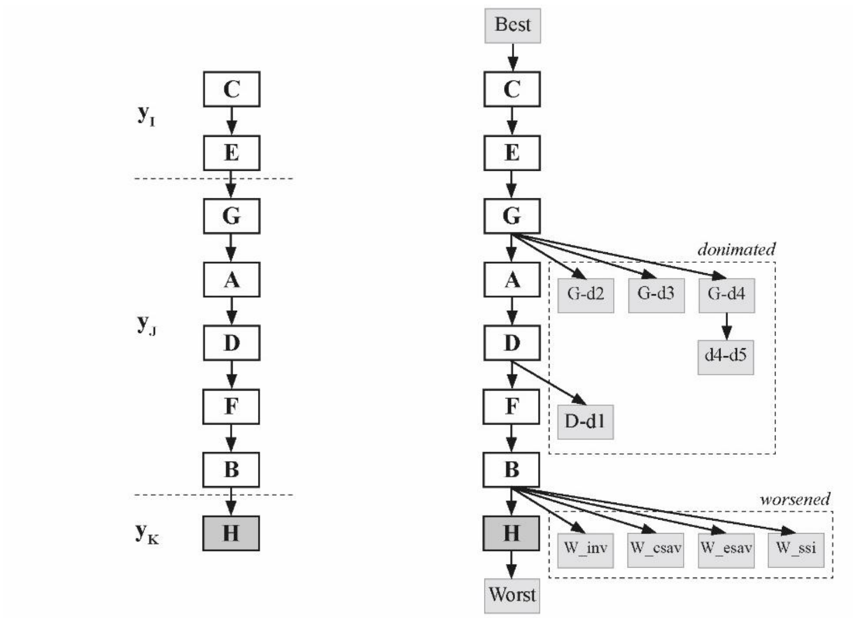


Figure 24 Original (left) and enlarged (right) learning set for a random respondent. The preference order shows that this respondent selected 7 alternatives out of the 8 included in the questionnaire (only excluding alternative H). This original learning set was expanded with the 5 dominated alternatives ("dominated"). Moreover, the 4 dummy worsened alternatives ("worsened") were added, together with the "Worst" and the "Best" alternatives occupying the first and the last positions.

For the same respondent, the performance of the alternatives used to train the UTAstar algorithm are presented in Table 12. It is worth noting that in the table, the performance of alternatives W_inv, W_csav, W_esav, and W_ssi depends on the specific order expressed by the respondent who ranked alternative B as the least preferred one (thus the performance levels of these dummy alternatives were based on the original alternative B). In Table 12 the dominated alternatives are denoted as "dominating alternative-dominated alternative" (e.g., alternative D-d1 is the alternative extracted from the dataset of all the case studies analyzed in Chapter 2. This alternative is dominated by Alternative D in the 8-alternative subset that has been presented to the respondent in the questionnaire).

Table 12 Complete set of alternatives used to estimate the value functions. Alternative A to H were those included in the questionnaire.

Alternative	**Investment Cost [€]**	**Cost Savings [€]**	**Emission Savings [kg_{CO2}]**	**Selfsufficiency [%]**
original				
A	9400	1300	2600	82
B	4650	500	1400	20
C	6600	750	1150	61
D	5300	700	2200	17
E	5000	750	1850	37
F	3400	400	850	27
G	2550	400	650	33
H	20000	1650	3100	0
dominated				
D-d1	5322	694	2190	17
G-d2	3950	359	554	29
G-d3	4439	360	554	29
G-d4	2778	210	264	16
d4-d5	20000	210	264	15
Best/Worst				
Best	2550	1650	3100	82
Worst	20000	210	264	0
worsened				
W_inv	20000	500	1400	20
W_csav	4650	210	1400	20
W_esav	4650	500	264	20
W_ssi	4650	500	1400	0

The algorithm then requires providing four additional parameters to estimate the value functions. For this purpose (i) the four criteria, (ii) their range, (iii) direction, and (iv) number of segments of the positive, non-decreasing, monotonous value function that the algorithm is asked to estimate have to be declared. The parameters used for the estimation are reported in Table 13.

Table 13 Parameters used in the UTASTAR algorithm

Criterion	**Range**	**UoM**	**Direction**	**N segments**
Investment cost	2550-20000	€	min	3
Cost savings	210-1650	€	max	3
Emission savings	264-3100	Kg_{CO2}	max	3
Self-sufficiency	0-82	%	max	3

The number of segments to define the value function as been set to 3 for all the criteria, resulting in a set of piecewise-linear functions, each of them of length $(max_{range}-min_{range})/3$.

In Figure 25, the set of value functions coherent with the order expressed by the same random respondent seen above are presented.

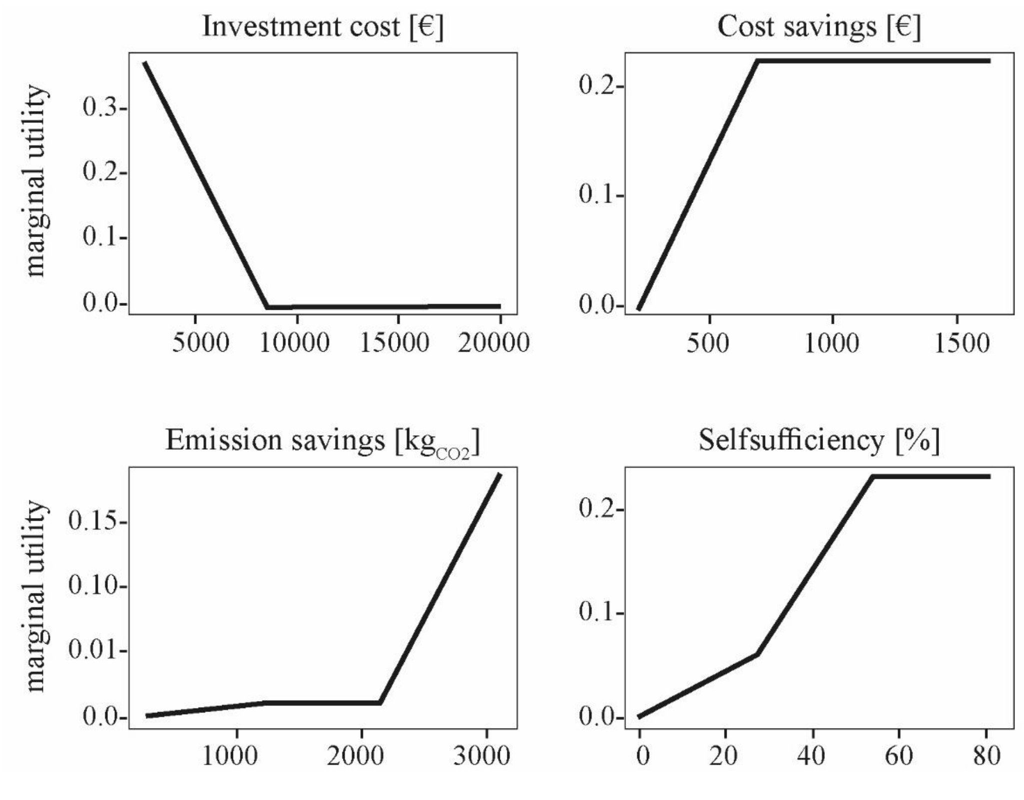


Figure 25 Example of the set of 3-segment piecewise value functions estimated for a random respondent in the dataset

It could be noted that several value functions present some flat parts. While this does not constitute a general mistake, it is related to some degree of model uncertainty due to lack of discriminatory information regarding the variation of the attribute level in these flat ranges. This is a consequence of the performance attributes of the alternatives ranked in the learning sets provided by the respondents. This aspect is further analysed in the discussion section.
Finally, the marginal utility and the total utility perceived by each respondent for each alternative is estimate. As an example, in Table 14 the utilities for the same respondent as for Figure 25 is reported.

Table 14 Performance, and estimated marginal and total utility of each of the alternatives according to the learning set provided by the random respondent.

Alternative	Investment Cost [€]	Cost Savings [€]	Emission Savings [kg_{CO2}]	Selfsufficiency [%]	Inv	C_Sav	E_Sav	SSI	U_Total
C	6600	750	1150	61	0.111	0.220	0.009	0.232	0.573
E	5000	750	1850	37	0.212	0.220	0.100	0.121	0.563
G	2550	400	650	33	0.366	0.087	0.004	0.095	0.553
A	9400	1300	2600	82	0.000	0.220	0.091	0.232	0.543
D	5300	700	2200	17	0.193	0.220	0.018	0.037	0.469
F	3400	400	850	27	0.313	0.087	0.006	0.059	0.465
B	4650	500	1400	20	0.234	0.133	0.010	0.044	0.421
H*	20000	1650	3100	0	0.000	0.220	0.181	0.000	0.401

*this alternative was rejected by the respondent

Beside the possibility to calculate the marginal and total utility for the known alternatives, this resulting set of value function allows to calculate the total perceived utilities for new alternatives outside of the set (e.g., real scenarios of RECs configurations), thus to compare them against each other and to order them in terms of preferability coherently with the preference model estimated for the specific respondent.

4.3 Estimation of individuals' acceptance

This subsection describe the process followed to estimate the likelihood of respondents to participate or not in a REC alternative based on its perfromance, and the socio-demographic characteristics of the respondent.
The models are specified including a random intercept at the individual-level to account for within-subject correlation, thus the slopes coefficients (β_n) of the fitting line were kept constant (fixed slopes), while the intercept regression coefficient (β_0) was allowed to vary between different respondents (defined in R environment with the common syntax (1|seed) as reported in each model specification and captured by the output τ_{00}).

4.3.1 Test of the presence of a Turin/non Turin specificity

A first analysis has been performed to test the hypothesis of the possible presence of a difference in the pattern of choices collected from respondents living inside or outside of Turin. This allows to evaluate whether to include all the observation present in the dataset or if it was necessary to exclude the respondents living outside of the city.
In order to allow for an easier interpretation of results, in all the following models the variables representing RECs performance are expressed in thousands of Euros for Investment cost [Inv_cost(k€)] and cost savings [Cost_sav(k€)], tons of CO_2 for the avoided emissions [Emis_sav(tCO_2)], while the 1-unit increase in self-sufficiency is set to a 10% increase [SSuff(%)]. Moreover, in terms of notation used in the tables reporting results of the regression analysis, OR stands for Odd Ratios, CI for Confidence interval, and p for p-value. Finally, the associated variables are marked in boldface.
As specified in the method chapter, the best models have been selected using marginal and conditional R^2 indicators, comparing AIC and BIC to account for parsimony of the model, while the Wald-test (p-value) and Confidence intervals (CI) have been used to select the determinants to be included.
In Table 15, Model_1 and Model_2 are displayed. Model_1 estimates the odds of participants to the survey to answer "Yes" or "Maybe" to the proposal of an alternative REC as a function of the performance level of the alternatives in the four attributes used to describe it [investment cost (Inv_cost), cost savings (Cost_sav), avoided emissions (Emis_sav), and self-sufficiency (SSuff)]. Model_2 estimates the same association between determinants and observed output, but adds a Boolean variable to distinguish the respondents living in Turin from those living outside of

the city. Table 15 displays the two estimated GLMM models, as well as the results of the performed regression:

Table 15 Association of RECs performance with the odds to accept a REC (Yes vs No), without (Model_1) and with (Model_2) the introduction of living in Turin as a determinant factor. Associated variables are in boldface.

Models equations:

Model_1: Y ~ Inv_cost + Cost_sav + Emis_sav +SSuff + (1|seed)

Model_2: Y ~ Inv_cost + Cost_sav + Emis_sav + SSuff + Turin + (1|seed)

Regression results:

	Model 1			Model 2		
Predictors	OR	CI 95%	p	OR	CI 95%	p
Numerical						
Intercept	**0.22**	0.18-0.26	<0.001	**0.22**	0.18-0.29	<0.001
Inv_cost (k€)	**0.76**	0.75-0.77	<0.001	**0.76**	0.75-0.77	<0.001
Cost_sav (k€)	**3.94**	2.99-5.19	<0.001	**3.49**	2.99-5.19	<0.001
Emis_sav(tCO_2)	**1.65**	1.50-1.83	<0.001	**1.65**	1.50-1.83	<0.001
SSuff(%)	**1.42**	1.39-1.46	<0.001	**1.42**	1.39-1.46	<0.001
Categorical						
Turin (no)				ref.		
Turin (yes)				0.95	0.73-1.24	0.701
Random Effects						
σ^2	3.29			3.29		
τ_{00}	2.18			2.18		
ICC	0.40			0.40		
Evaluation metrics						
AIC	10357			10359		
BIC	10401			10410		
Marginal R^2 / Conditional R^2	0.511 / 0.706			0.511 / 0.706		

As shown in Table 15, it is possible to notice that all the determinant factors describing the performance of REC alternatives are associated (marked in bold in the table) and follow an expected trend. For instance, every 1-unit increase in the Investment cost required to implement the alternative is associated with a decrease in the odd ratio of accepting to participate. On the contrary, performance representing benefits arising from joining the REC determine an increase in the odds to accept the alternative proposed (odd ratios of 3.94 for cost savings, 1.65 for emission savings, and 1.42 for self-sufficiency from the grid). In particular, it is worth noting how the 1-unit increase in cost savings determines almost a 4-time

increase in the odds to participate, revealing the sensitivity of respondents to this aspect in the implementation of a REC.
Considering the results of Model_2, it is noticeable that these do not substantially differ from those of Model_1. The variable distinguishing the respondents that live in Turin from those living outside of the city is not associated with a variation of the odd to participate in a REC (CI including 1, and p-value > 0.05). This is also confirmed by the odd ratio resulting from the analysis for this determinant, showing a value close to one (thus having no effect on the dependent variable in a logistic regression model).
A second set of analysis has been performed to assess the moderation effect of the location of the respondent on the association between determinants (i.e., the performance of the alternative RECs) and the odds to accept to participate in a REC (Model_3 to Model_6 in Table 16). Four different GLMM models have been estimated by adding the "residency in Turin" as a moderator term of each of the RECs performance variables one at a time. This allows to estimate the slope of the RECs performances at the specific value of the moderator (for respondents living in Turin or not in Turin). The four models have been specified as shown in Table 15, where the summary of the estimates of these models is displayed.

Table 16 Moderation effect of the location of the respondent (Turin/not Turin) on the association between RECs performance and the odd to participate. CI not displayed but in accordance with registered p-values (CI not including 1 for p-values<0.05). Associated variables are in boldface.

Models equations:

Model_3:	Y ~ Inv_cost*Turin + Cost_sav + Emis_sav + SSuff + (1|seed)
Model_4:	Y ~ Inv_cost + Cost_sav*Turin + Emis_sav + SSuff + (1|seed)
Model_5:	Y ~ Inv_cost + Cost_sav + Emis_sav *Turin + SSuff + (1|seed)
Model_6:	Y ~ Inv_cost + Cost_sav + Emis_sav + SSuff*Turin + (1|seed)

Regression results:

	Model 3		Model 4		Model 5		Model 6	
Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	**0.18**	<0.001	**0.20**	<0.001	**0.19**	<0.001	**0.21**	<0.001
Inv_cost (k€)	**0.78**	<0.001	**0.76**	<0.001	**0.76**	<0.001	**0.76**	<0.001
Cost_sav (k€)	**3.96**	<0.001	**3.49**	<0.001	**5.02**	<0.001	**3.95**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.83**	<0.001	**1.66**	<0.001	**1.65**	<0.001
SSuff(%)	**1.42**	<0.001	**1.42**	<0.001	**1.42**	<0.001	**1.45**	<0.001
Categorical								
Turin (no)	ref.		ref.		ref.		ref.	
Turin (yes)	1.31	0.092	1.15	0.375	1.20	0.282	1.08	0.652

Interactions								
Inv_cost (k€) *Turin (yes)	**0.96**	<0.001						
Cost_sav (k€) *Turin (yes)			**0.86**	0.031				
Emis_sav(tCO_2) *Turin (yes)					**0.69**	0.026		
SSuff(%) *Turin (yes)							0.97	0.165
Random Effects								
σ^2	3.29		3.29		3.29		3.29	
τ_{00}	2.20		2.18		2.19		2.18	
ICC	0.40		0.40		0.40		0.40	
Evaluation metrics								
AIC	10347		10356		10356		10359	
BIC	10405		10415		10414		10417	
Marginal R^2 / Conditional R^2	0.517 / 0.710		0.512 / 0.707		0.512 / 0.707		0.511 / 0.706	

The results shows that in three models out of four the presence of an interaction effect of the geographical location of the respondents' household on the association between the performance of the RECs an the odd to accept the offered alternative (only the moderating effect of the residency of the respondent being located in the city and the self-sufficiency performance is not associated with a variation in the odd to participate). In terms of parsimony of the models and their explanatory power, it is possible to notice a small increase in the conditional R^2 compared to Model_1 where no interaction was introduced (conditional $R^2 = 0.706$ in Model 1), while the AIC value sees a decrease in all the models except for Model_6, thus marking an improvement in the performance of the models. On the contrary, BIC registers a small increase in all the models with interactions compared to Model_1, therefore, highlighting a small detriment of models' performance (it is worth nothing that this is due to a tendency of BIC to penalize the model complexity more than the AIC metric). In Figure 26 to Figure 28 the moderating effect of living outside/inside of the city on the probability of accepting a REC offer is displayed. In these figures the variation in odd to participate is plotted against the increase (or decrease in case of investment cost) of the performance attribute of the REC. In each plot, the variation in the slope of the curve is determined by the value assumed by the moderator, thus when the respondent resides in Turin (the dashed line) or outside of the city (the continuous line). Considering each REC performance attribute, Figure 26 shows that each 1-unit increase in Investment cost (1 k€) determines a reduction in the odd to participate (OR=0.76). For individuals living in Turin this is further reduced (OR=0.96 when the moderator is present). This trend is similar also for the remaining two performance attributes of the REC. Indeed, for every thousand euros saved in energy expenditure (Figure 27), the odd to accept increases (OR=3.49), with participants living in Turin showing a less steep increase in this association (OR of the interaction with the moderator = 0.86). Finally, also for the emission saving performance attribute (Figure 28) the increase of the odd to accept (OR=1.66) is moderated by the location of the respondent (OR=0.69), showing a smaller effect of this attribute for the city residents.

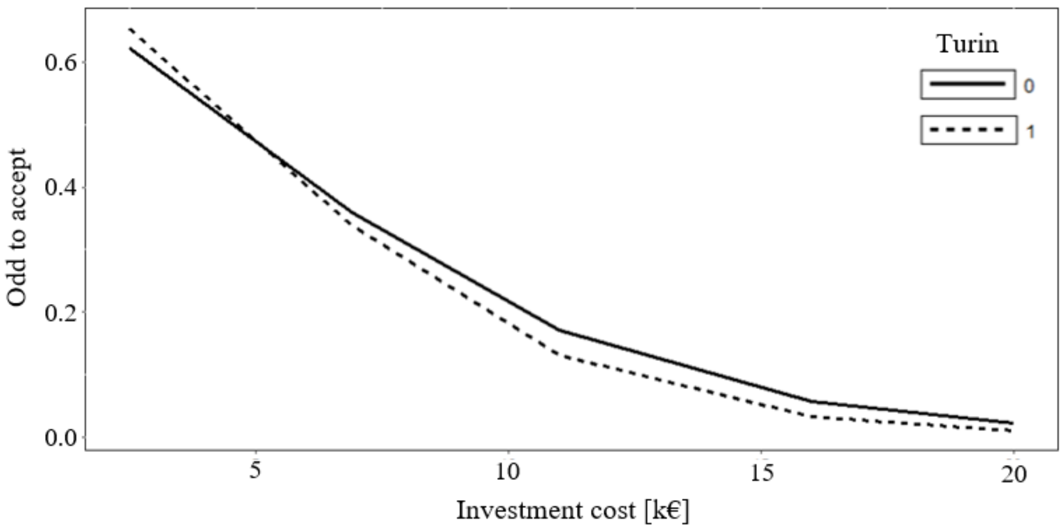


Figure 26 Association between Investment cost performance attribute and the odd to accept a REC with the moderating effect of the respondents' location

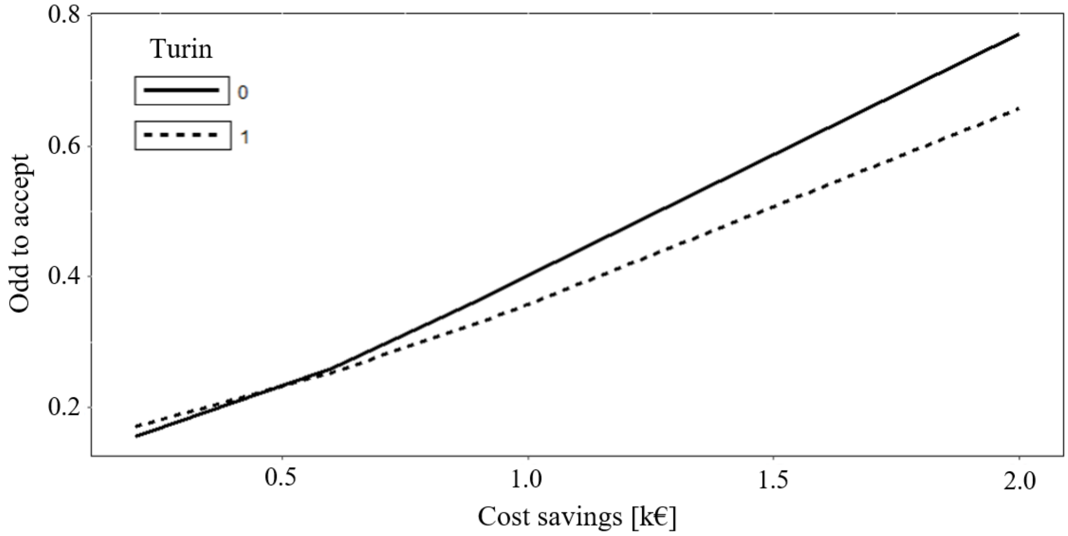


Figure 27 Association between Cost savings performance attribute and the odd to accept a REC with the moderating effect of the respondents' location

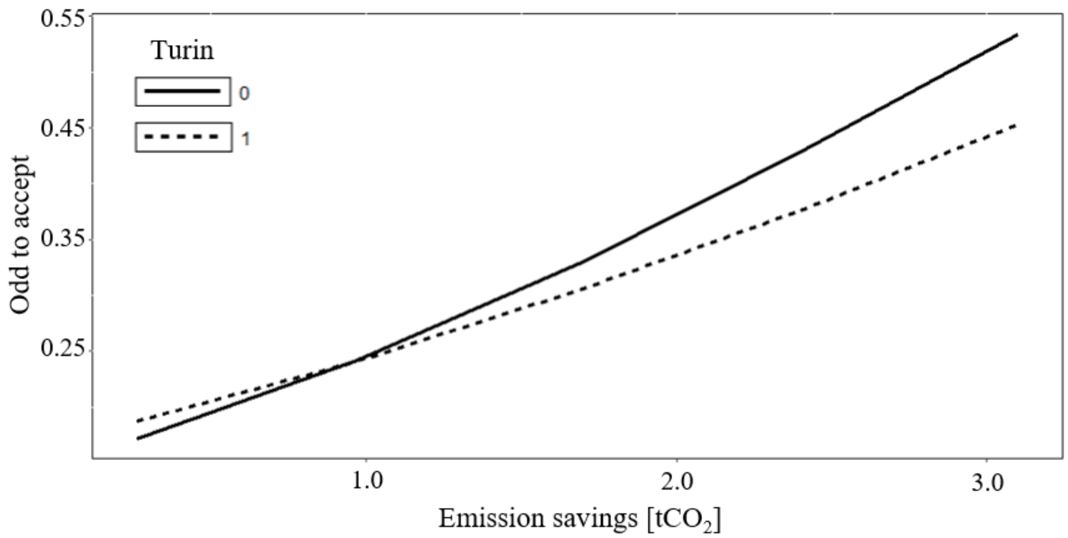


Figure 28 Association between Emission savings performance attribute and the odd to accept a REC with the moderating effect of the respondents' location

In light of the above, it could be concluded that the location of the respondent inside or outside of the city acts as a moderator on the association between the determinants constituted by the performance attributes and the probability to accept to participate in a REC. This confirms a difference in the pattern of response from the two clusters of participants to the survey (those living in Turin from those living in other areas). For this reason, the observations referring to respondents located outside of the city have been excluded from the next analysis limiting the dataset only to the respondents living in Turin.

4.3.2 Estimation of likelihood to accept of Turin inhabitants

After excluding the observations related to respondents living in locations other than the city, the analysis on the respondents living in Turin has proceeded to evaluate the socio-demographic variables that could have influenced the likelihood of accepting or not a possible REC alternative. Together with the respondents not located in the city of Turin, few other removals of specific observations have been performed on the dataset of respondents. This was necessary for two main reasons: (i) the presence of few cases for a specific class of respondence (i.e., Gender = non-binary has been excluded due to the few number of respondents that provided this answer), and (ii) the impossibility to categorize the respondent in one of the available classes used for the national census survey (as stated in the method, the classes used to cluster the population have to be coherent with those provided by the census survey in order to link the different estimated decision models with the resident population in the city in the phase of spatialization of results). Furthermore, only those respondents who specified at least two preferences over RECs in the questionnaire (meaning that the respondent has provided a rank between at least two alternatives) were retained. This guarantees coherence with the next section (Section 4.3.2) and the possibility to include also the dummy alternatives described in Section 4.3, thus expanding the dataset of observations.

The final dataset comprises 349 respondents living in the city Turin, with 19 observations for each of them, for a total of 6631 observations.

A number of preliminary analysis has been conducted to aggregate the classes of variables proposed in the questionnaire to limit their number. In this aggregation, particular care has been taken to ensure that the aggregated classes match those of the national census data in order to allow for the spatialization of the decision model estimated in this section and in the next one. This has determined the necessity to aggregate the classes of specific socio-demographic variables (i.e., Age and Education level of the respondents) coherently with the two datasets used for the generation of the synthetic population (see Section 3.3.2). For instance, concerning the education level of the respondents the possible matching classes were: (i) no title/elementary school, (ii) middle school, (iii) High school, (iv) degree and higher according to the classes registered by the census survey. For the age of the respondents, the modalities used in the two census survey datasets [202,203] are shown in Figure 29, together with a remap into matching classes ranges ("possible common modalities" in the figure). For instance, a common modality between the two datasets is the one clustering age classes 1 (0-2 y.o.) to 7 (18-19 y.o.) of AVQ dataset, and classes P14 (<5 y.o.) to P17 (15-19 y.o.) of census survey dataset. The higher age classes in the two datasets are quite consistent, with some pairs of the classes of the census survey dataset having an aggregated range equal to a single class in the AVQ dataset, thus presenting a higher granularity of the data.

code	AVQ modalities [y]	possible common modalities	census modalities [y]	code
1	0-2	1	<5	P14
2	3-5		5-9	P15
3	6-10	2	10-14	P16
4	11-13		15-19	P17
5	14-15	3	20-24	P18
6	16-17		25-29	P19
7	18-19	4	30-34	P20
8	20-24	5	35-39	P21
9	25-34		40-44	P22
10	35-44	6	45-49	P23
11	45-54		50-54	P24
12	55-59	7	55-59	P25
13	60-64		60-64	P26
14	65-74	8	65-69	P27
15	75+	9	70-74	P28
			>74	P29

Figure 29 Matching between age classes modalities in the AVQ and census datasets

In light of the above, six socio-demographic characteristics of the inhabitants have been used as categorical determinants. Two of them specified in 3 classes each: (i) age of the respondent (Age) and (ii) household income in thousands of Euros (HH_inc), and 4 dichotomous variables: (i) if the size of the household is bigger or smaller than 3 inhabitants (HH_num); (ii) if the respondent has a degree level or not (Edu_lev); (iii) the gender of the respondent (Gender); (iv) the regime of property of the dwelling (HH_prop). In Figure 30 the original occurrence of the socio-demographic characteristics of the 349 respondents in the classes available in the online survey is displayed, together with their distribution in the remapped classes. In Appendix E the distribution of socio-demographic characteristics of this subset of respondents is compared to the marginal distribution in the city of Turin according to the last census survey.

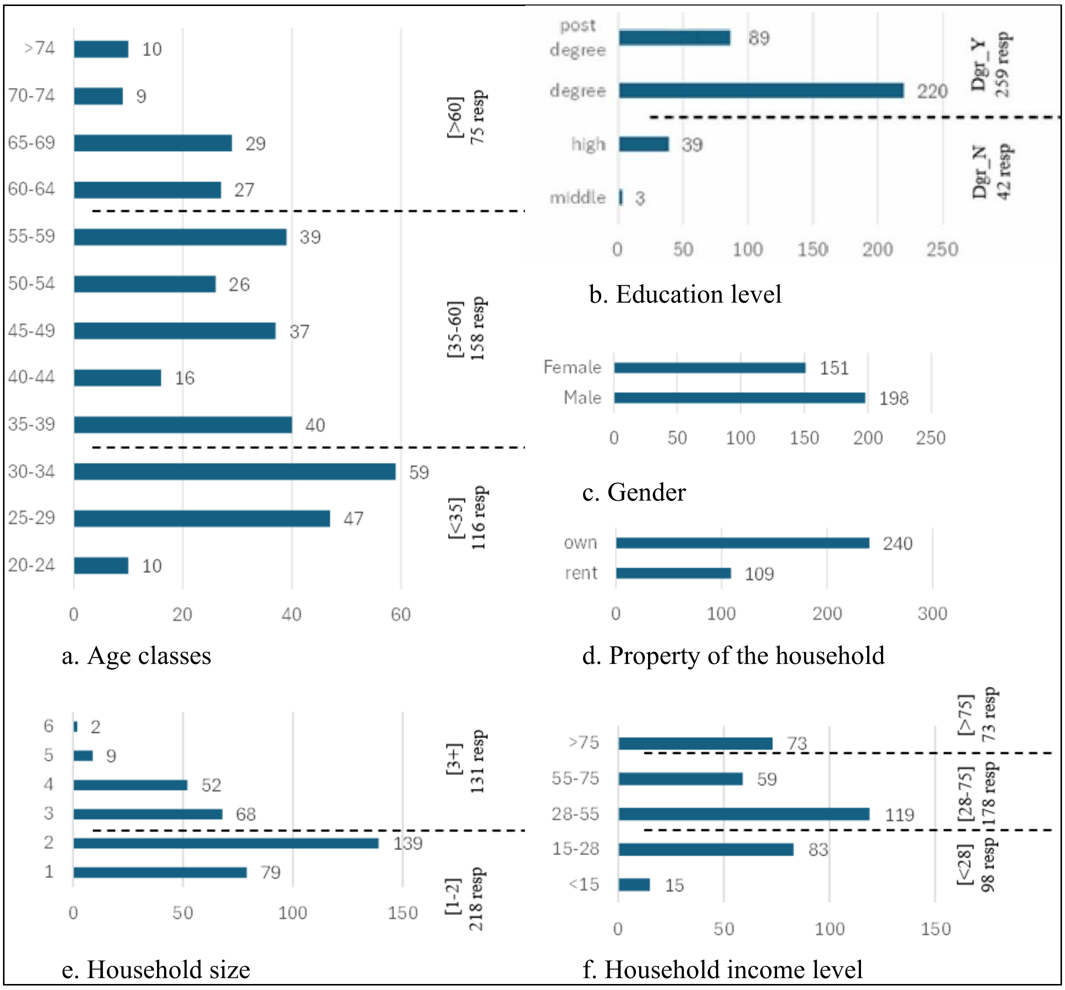


Figure 30 Occurrence of the socio-demographic characteristics of the 349 respondents in the questionnaire and in the remapped classes

As mentioned in the construction of the questionnaire (Section 3.2.2), the respondents were enabled to specify the Income level of the households by selecting from pre-specified classes. This was justified by the necessity to reduce potential mistakes in the filling of the questionnaire at the expense of some degree of granularity. This limitation caused a rough reparameterization of the classes, resulting in a large range of values included in the same income level classes associated with the estimated likelihood to participate in a REC. A more granular set of data (i.e., the definition of this variable as numeric one instead of categorical) could result in a more refined modellization.

Table 17 shows the GLMM model with the final aggregation of the socio-demographic determinants (Model_7), together with the results of the regression.

Table 17 Association of RECs performance and socio-demographic attributes of the respondents with the odds to accept a REC (Yes vs No). Associated variables are in boldface.

Models equations:

Model_7: Y ~ Inv_cost + Cost_sav + Emis_sav + SSuff + socio-demographic attributes + (1|seed)

Regression results:

Predictors	OR	CI 95%	p
Numerical			
Intercept	0.65	0.37-1.12	0.122
Inv_cost (k€)	**0.75**	0.73-0.76	<0.001
Cost_sav (k€)	**4.57**	3.20-6.52	<0.001
Emis_sav(tCO_2)	**1.66**	1.47-1.89	<0.001
SSuff(%)	**1.42**	1.38-1.47	<0.001
Categorical			
Age [< 35]	ref.		
Age [35-60]	**0.62**	0.43-0.90	0.012
Age [60+]	**0.38**	0.24-0.60	<0.001
HH_num [>3]	ref.		
HH_num [3+]	**0.69**	0.50-0.95	0.022
HH_inc [<28k]	ref.		
HH_inc [28-75k]	1.27	0.90-1.80	0.178
HH_inc [75k+]	**1.65**	1.04-2.62	0.034
Edu_lev [Dgr_N]	ref.		
Edu_lev [Dgr_Y]	0.80	0.51-1.25	0.332
Gender [Fem]	ref.		
Gender [Mal]	0.82	0.62-1.10	0.191
HH_prop [rent]	ref.		
HH_prop [own]	0.82	0.58-1.16	0.256
Random Effects			
σ^2	3.29		
τ_{00}	1.41		
ICC	0.30		
Evaluation metrics			
AIC	6079		
BIC	6174		
Marginal R^2 / Conditional R^2	0.590 / 0.713		

Considering the results, all the determinants related to the performance of REC are associated with the odd to accept, and follow an expected trend as in the model tested for the Turin/non-Turin effect (every 1-unit increase of the investment cost decreases the odds to participate, while for each 1-unit increase in the remaining performances of the REC the odds to accept increases). Regarding the socio-demographic attributes, the increase in age of the respondent is associated with a reduction in the odds to participate, suggesting a more positive attitude of younger respondents. Moreover, the increase in the size of the household (HH_num) decreases the odds of accepting the REC. The last determinant that shows an association with the odds to accept is the level of income of the household. Regarding this, the results seem to show an increase in the willingness to participate when the level of income increases, this potentially being explained by the higher availability of money. Finally, the other 3 socio-demographic variables (Education level, Gender, property of the household) do not seem to influence the odds of participating in a REC. Nevertheless, they have been retained in the final model to further test for possible interactions as moderators.

Model_7 has been tested against multicollinearity, showing Variance inflation factors (VIFs) lower than 3 for two independent variables (Inv_cost and Cost_sav), and lower than 2 for the others, thus presenting no issues in this sense.

A further analysis was conducted to investigate the possible moderating effect of socio-demographic characteristics on the association between performance attributes and the participation to a REC and to evaluate potential improvement in the explanatory power of the model.

For this analysis, each socio-demographic attributes has been inserted as a moderator interacting with the association between the REC performance attribute and the odd to accept the alternative REC. Each interaction has been tested one at a time.

These exploratory analysis are reported in Appendix D, while Table 18 shows a summary of the individuated moderating effect of socio-demographic attributes on the association of REC performances with the odds to accept.

Table 18 summary of the individuated interactions between socio-demographic attributes and REC performance attributes on the odds to participate in a REC alternative.

Socio-demographic moderators	Performance attributes	Evaluation metrics			
		AIC	BIC	Cond. R^2	Marg. R^2
HH_inc	Inv_cost	No moderating effect			
	Cost_sav	6077.679	6186.471	0.714	0.591
	Emis_sav	6078.590	6187.382	0.714	0.591
	SSuff	6078.639	6187.432	0.713	0.591
Edu_lev	Inv_cost	No moderating effect			
	Cost_sav	No moderating effect			

	Emis_sav	No moderating effect			
	SSuff	No moderating effect			
Age	Inv_cost	No moderating effect			
	Cost_sav	No moderating effect			
	Emis_sav	No moderating effect			
	SSuff	No moderating effect			
HH_prop	Inv_cost	No moderating effect			
	Cost_sav	No moderating effect			
	Emis_sav	No moderating effect			
	SSuff	6068.165	6170.157	0.714	0.591
Gender	Inv_cost	6074.961	6176.954	0.714	0.591
	Cost_sav	6079.584	6181.577	0.713	0.590
	Emis_sav	No moderating effect			
	SSuff	No moderating effect			
HH_num	Inv_cost	6073.430	6175.422	0.713	0.590
	Cost_sav	6069.916	6171.909	0.713	0.590
	Emis_sav	6076.573	6178.565	0.713	0.589
	SSuff	6070.553	6172.546	0.714	0.591

The associated interactions resulting from the models (summarized in Table 18) have been implemented in a full model shown in Table 19. This full model (Model_8a) has been simplified by excluding the less associated interactions to reach a stable and efficient model able to predict the probability to accept or not a specific REC. In the same Table 18 the final simplified model (Model_8b) is displayed as well.

Table 19 Moderating effects of the socio-demographic caracteristics on the association between RECs performance and the odd to accept it including all the potential moderators (Model_8a) and only those associated (Model_8b). Associated variables are in boldface.

Models equations:

Model_8a: Y ~ REC performance attributes * full set of moderators + Socio-demographic characteristics + (1|seed)

Model_8b: Y ~ REC performance attributes * associated moderators + Socio-demographic characteristics + (1|seed)

Regression results:

	Model_8a			Model_8b		
Predictors	OR	CI 95%	p	OR	CI 95%	p
Numerical						
Intercept	1.47	0.77-2.81	0.244	1.25	0.68-2.28	0.473
Inv_cost (k€)	**0.73**	0.70-0.76	<0.001	**0.75**	0.73-0.76	<0.001

Cost_sav (k€)	**2.99**	1.46-6.13	0.003	**2.60**	1.64-4.10	<0.001
Emis_sav(tCO_2)	**1.57**	1.22-2.02	<0.001	**1.68**	1.48-1.90	<0.001
SSuff(%)	**1.28**	1.19-1.37	<0.001	**1.29**	1.22-1.37	<0.001
Categorical						
Age [< 35]	ref.					
Age [35-60]	**0.63**	0.43-0.91	0.014	**0.63**	0.43-0.91	0.014
Age [60+]	**0.38**	0.24-0.60	<0.001	**0.38**	0.24-0.60	<0.001
HH_num [>3]	ref.					
HH_num [3+]	**0.35**	0.22-0.57	<0.001	**0.35**	0.22-0.55	<0.001
HH_inc [<28k]	ref.					
HH_inc [28-75k]	1.10	0.66-1.84	0.713	1.27	0.89-1.81	0.180
HH_inc [75k+]	1.04	0.52-2.07	0.917	**1.66**	1.04-2.64	0.034
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	0.79	0.50-1.25	0.320	0.80	0.51-1.26	0.330
Gender [Fem]	ref.					
Gender [Mal]	**0.59**	0.39-0.87	0.009	**0.60**	0.41-0.89	0.010
HH_prop [rent]	ref.					
HH_prop [own]	**0.59**	0.39-0.90	0.015	**0.57**	0.38-0.87	0.009
Interactions						
Inv_cost (k€) *Gender [Mal]	1.03	0.98-1.08	0.253			
Inv_cost (k€) *HH_num [3+]	1.01	0.96-1.06	0.760			
Cost_sav (k€) *HH_inc [28-75k]	1.10	0.52-2.31	0.810			
*HH_inc [75k+]	1.10	0.43-2.81	0.839			
Cost_sav (k€) *Gender [Mal]	1.35	0.78-2.33	0.288	**1.71**	1.11-2.62	0.014
Cost_sav (k€) *HH_num [3+]	1.62	0.77-2.33	0.288	**1.96**	1.25-3.06	0.003
Emis_sav(tCO_2) *HH_inc [28-75k]	1.02	0.75-1.37	0.909			
*HH_inc [75k+]	1.23	0.84-1.80	0.286			
Emis_sav(tCO_2) *HH_num [3+]	1.05	0.80-1.37	0.733			
SSuff (%) *HH_inc [28-75k]	1.02	0.94-1.10	0.641			
*HH_inc [75k+]	1.05	0.94-1.16	0.392			
SSuff (%) *HH_prop [own]	**1.10**	1.03-1.18	0.006	**1.11**	1.04-1.18	0.002

SSuff (%) *HH_num [3+]	**1.08**	1.01-1.16	0.027	**1.08**	1.01-1.16	0.020
Random Effects						
σ^2	3.29			3.29		
τ_{00}	1.44			1.44		
ICC	0.30			0.30		
Evaluation metrics						
AIC	6065			6053		
BIC	6249			6175		
Marginal R^2 / Conditional R^2	0.594 / 0.718			0.593 / 0.717		

Model_8a shows that all the independent variables describing REC performance are associated, while, among the socio-demographic ones, the age and gender of the respondent, as well as the size and property of the household are associated. Among the possible interactions between the socio-demographic characteristics and the vector of REC performance, only the size of the household and being or not the owner of the household are associated, with a p-value <0.05. It is worth noting that, even if its use is debated [227], a threshold of 0.05 for the p-value has been used in this phase to select the variables to be included or excluded from the model. Using this threshold, a backward step regression has been used to remove step by step the hypothesized interaction effects presenting the highest p-value over the imposed threshold: (i) Cost_sav (k€)*HH_inc (CI 0.52-2.31, p-value 0.810, and CI 0.43-2.81, p-value 0.839 for the two classes respectively); (ii) Inv_cost (k€) *HH_num [3+] (CI 0.96-1.06, p-value 0.759); (iii) Emis_sav(tCO_2)*HH_num [3+] (CI 0.80-1.36, p-value 0.755); (iv) SSuff (%)*HH_inc (CI 0.95-1.10, p-value 0.568, and CI 0.95-1.16, p-value 0.341 for the two classes respectively); (v) Emis_sav(tCO_2)*HH_inc (CI 0.86-1.28, p-value 0.655, and CI 0.97-1.62, p-value 0.085 for the two classes respectively); (vi) Cost_sav (k€) *Gender [Mal] (CI 0.81-2.41, p-value 0.231). In Table 20 the two models without (Model_7) and with interactions (Model_8b) are displayed to better compare their results.

Table 20 Association of RECs performance and socio-demographic attributes of the respondents with the odds to accept a REC (Yes vs No). Comparison between the model without (Model_7) and with the socio-demographic moderating effect (Model_8). Associated variables are in boldface.

Models equations:

Model_7: Y ~ Inv_cost + Cost_sav + Emis_sav + SSuff + socio-demographic attributes + (1|seed)

Model_8b: Y ~ Inv_cost + Cost_sav + Emis_sav + SSuff + Cost_sav*Gender + Cost_sav*HH_num + SSuf*HH_prop + SSuff*HH_num + (1|seed)

Regression results:

	Model_7			Model_8b		
Predictors	OR	CI 95%	p	OR	CI 95%	p
Numerical						
Intercept	0.65	0.37-1.12	0.122	1.25	0.68-2.28	0.473
Inv_cost (k€)	**0.75**	0.73-0.76	<0.001	**0.75**	0.73-0.76	<0.001
Cost_sav (k€)	**4.57**	3.20-6.52	<0.001	**2.60**	1.64-4.10	<0.001
Emis_sav(tCO_2)	**1.66**	1.47-1.89	<0.001	**1.68**	1.48-1.90	<0.001
SSuff(%)	**1.42**	1.38-1.47	<0.001	**1.29**	1.22-1.37	<0.001
Categorical						
Age [< 35]	ref.					
Age [35-60]	**0.62**	0.43-0.90	0.012	**0.63**	0.43-0.91	0.014
Age [60+]	**0.38**	0.24-0.60	<0.001	**0.38**	0.24-0.60	<0.001
HH_num [>3]	ref.					
HH_num [3+]	**0.69**	0.50-0.95	0.022	**0.35**	0.22-0.55	<0.001
HH_inc [<28k]	ref.					
HH_inc [28-75k]	1.27	0.90-1.80	0.178	1.27	0.89-1.81	0.180
HH_inc [75k+]	**1.65**	1.04-2.62	0.034	**1.66**	1.04-2.64	0.034
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	0.80	0.51-1.25	0.332	0.80	0.51-1.26	0.330
Gender [Fem]	ref.					
Gender [Mal]	0.82	0.62-1.10	0.191	**0.60**	0.41-0.89	0.010
HH_prop [rent]	ref.					
HH_prop [own]	0.82	0.58-1.16	0.256	**0.57**	0.38-0.87	0.009
Interactions						
Cost_sav (k€) *Gender [Mal]				**1.71**	1.11-2.62	0.014
Cost_sav (k€) *HH_num [3+]				**1.96**	1.25-3.06	0.003
SSuff (%) *HH_prop [own]				**1.11**	1.04-1.18	0.002
SSuff (%)				**1.08**	1.01-1.16	0.020

*HH_num [3+]		
Random Effects		
σ^2	3.29	3.29
τ_{00}	1.41	1.44
ICC	0.30	0.30
Evaluation metrics		
AIC	6079	6053
BIC	6174	6175
Marginal R^2 / Conditional R^2	0.590 / 0.713	0.593 / 0.717

Model_8b presents a higher conditional R^2 (0.717), compared to Model_7 without interactions included (R^2 = 0.713), Also AIC shows a better result, while BIC presents a small but negligible detriment in performance. Finally, Model_8b has been tested against multicollinearity showing acceptable values (VIF < 5).

The results of Model_8b show that all the independent variables related to the performance of the REC are associated with the likelihood of the respondents to accept or decline to participate (a hindering effect for Investment cost, and an enabling effect played by the other three).

Considering the socio-demographic variables, three of them are included in the model as covariates: age of the respondent, income level of the household, and education level of the respondent. The increase of age of the respondent seems to be accompanied by a reduction in the acceptance of participating compared to respondents younger than 35 years old. Regarding the income level of the household, both the two higher-level classes present an increase in the odd ratios of participating in a REC compared to the income level class lower than 28k€ a year. The increasing trend seems rational also in its magnitude, increasing from 1.27 for the second class (with income ranging from 28 to75k€) to 1.66 for the third class (income higher than 75k€). This might be justifiable with an increased probability to accept when the household has higher money availability. For this determinant, it is worth noting that only the third class (income higher than 75k euros) presents a p-value lower than 0.05, with the impossibility to exclude the null option for the class of income in a range from 28k€ to 75 k€. Finally, the education level of the respondents resulted as not associated with the probability of the respondent opting for acceptance. Also considering the odd ratio resulting from the logistic regression, the increase in education level seems to have a hindering effect when the respondent holds at least a degree compared to one with a lower education level. This latter might be counterintuitive, with the expectation that highly educated individuals might be more likely to be willing to accept the offer to participate. This effect

could be explained by the limited number of respondents, and especially by the unbalanced nature of the dataset for highly educated individuals. To test for any major changes in results with the exclusion of the education level of the respondents a further analysis has been run, showing no major differences in results.
The analysis of the moderating effect of the socio-demographic variables on the association between performance attributes and odd to participate to a REC showed that four main interactions have been confirmed from the full set: (i) cost savings performance moderated by gender (Cost_sav (k€)*Gender [Mal]), and (ii) by the size of the household (Cost_sav (k€)*HH_num [3+]), and (iii) the self-sufficiency achievable by the REC moderated by the regime of property of the household (SSuff (%)*HH_prop [own]), and (iv) by the size of the household (SSuff (%)*HH_num [3+]). It is interesting to notice how these four interactions sort of counterbalance the effect of the Boolean socio-demographic independent variables. For instance, it could be seen how males seem to be less prone to accept the participation in a REC with respect to females (0.60 odd ratio), but they are more sensitive to the increase in monetary savings achievable by the REC (1.71 odd ratio for the interaction of the two variables). In the same way, bigger households are less prone to accept participating in a REC (0.35 odd ratio), while this independent socio-demographic variable presents a quite high moderating effect on the cost savings variable (1.96 odd ratio) and a limited one on the self-sufficiency performance of the REC (1.08 odd ratio). This reasoning also helps to explain the partially counterintuitive coefficient related to the regime of property (HH_prop). Indeed, the evidence suggests that owning the dwelling in which the household resides appears to reduce the propensity to participate. However, this is counterbalanced by the moderating effect that ownership of the dwelling exerts on self-sufficiency performance, thereby increasing the likelihood of acceptance of the alternative (1.11 odd ratio for the interaction of the two variables). As for the previous analysis on the effect of the location of the respondents inside or outside of Turin, this interaction effect between independent variables (interchangeably interpretable as moderator or moderated variables) is graphically displayed in Figure 31 to Figure 34.

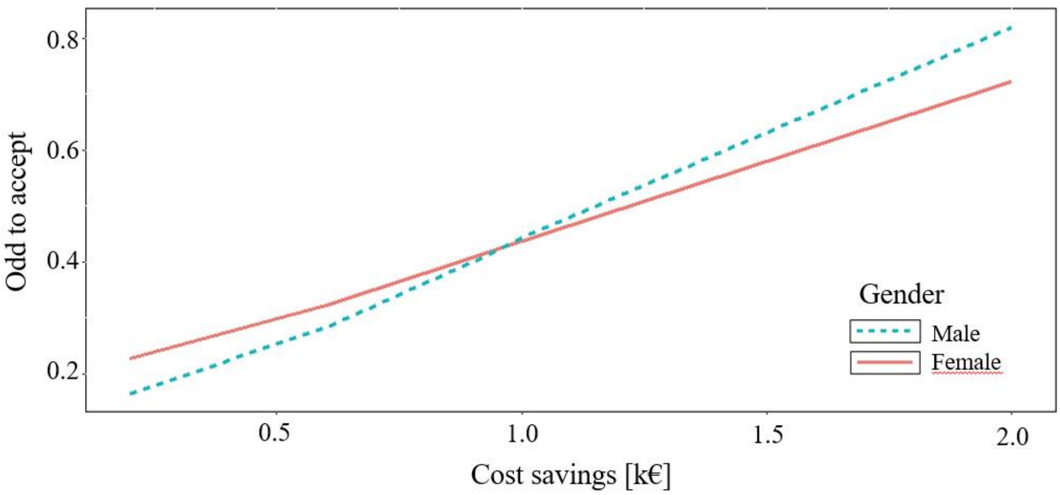


Figure 31 Association between Cost savings performance attribute and the odd to accept a REC with the moderating effect of Gender

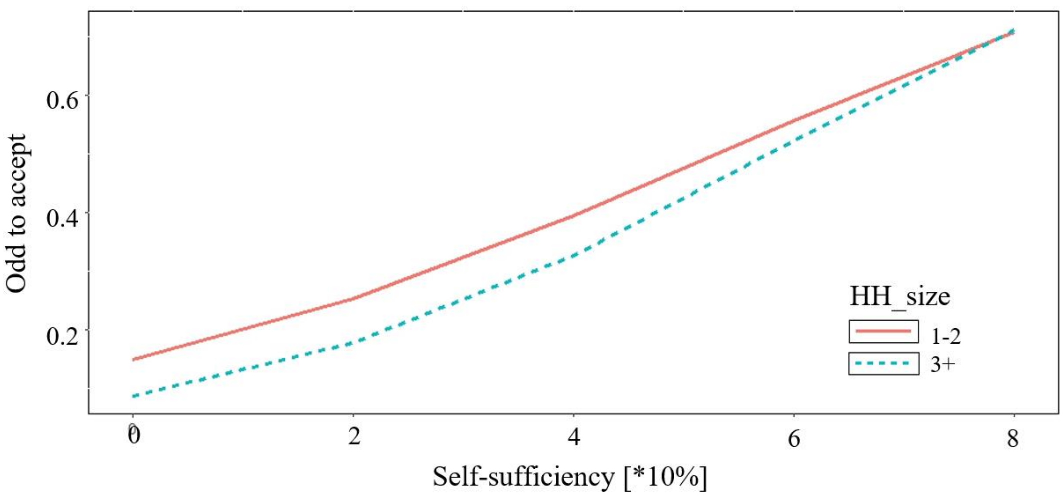


Figure 32 Association between Self-sufficiency performance attribute and the odd to accept a REC with the moderating effect of the size of the Household

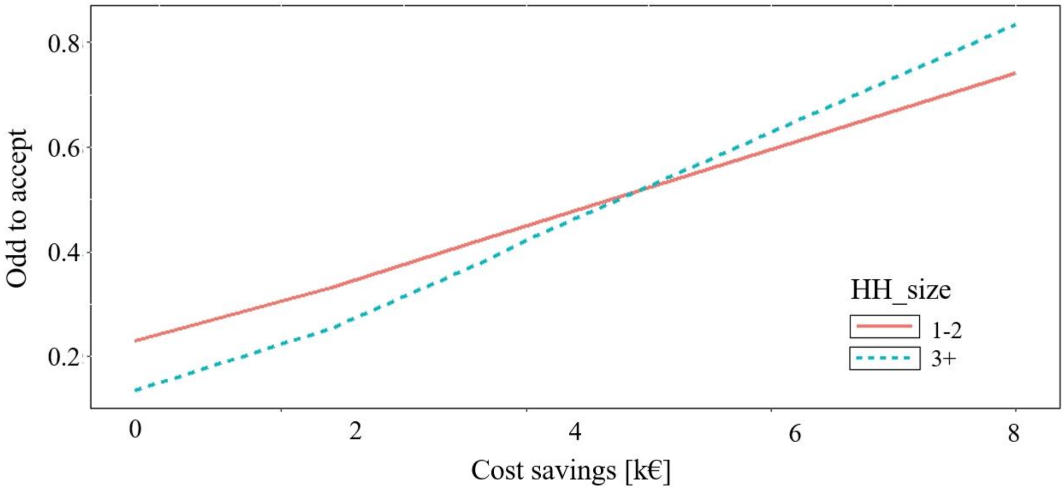


Figure 33 Association between Cost savings performance attribute and the odd to accept a REC with the moderating effect of the size of the Household

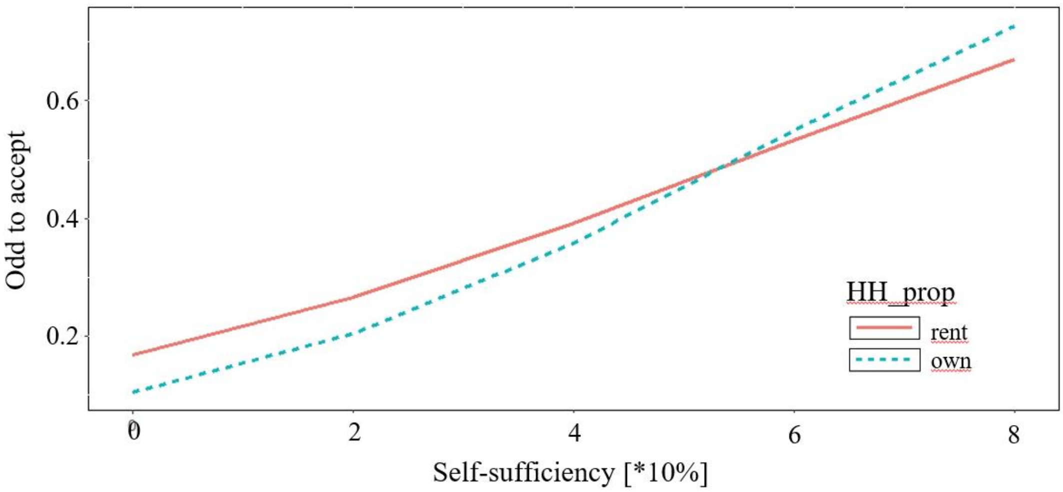


Figure 34 Association between Self-sufficiency performance attribute and the odd to accept a REC with the moderating effect of the Household property condition

Considering the results of Model_8b, it is possible to transform the odd ratios in regression coefficients by calculating the natural logarithm of odd ratios. In this way it is possible to express the probability of participating in a REC (Y=1) depending on REC performance attributes, socio-demographic characteristics of the respondents, and the moderating effect of the latter.

$$Pr(y=1) = logit^{-1}(-0.288 * \text{Inv\_cost} + 0.955 * \text{Cost\_sav} + 0.519 * \text{Emis\_sav} + 0.255 * \text{SSuff} - 0.462 * \text{Age}^{[35-60]} - 0.968 * \text{Age}^{[60+]} - 1.050 * \text{HH\_num}^{[3+]} + 0.507 * \text{HH\_inc}^{[75k+]} - 0.511 * \text{Gender}^{[Mal]} - 0.562 * \text{HH\_prop}^{[own]} + 0.536 * \text{Cost\_sav} * \text{Gender}^{[Mal]} + 0.673 * \text{Cost\_sav} * \text{HH\_num}^{[3+]} + 0.104 * \text{SSuff} * \text{HH\_prop}^{[own]} + 0.077 * \text{SSuff} * \text{HH\_num}^{[3+]})$$

A further analysis was conducted including the neighborhood and the HV/MV substation to which each respondent belongs as additional random factors in the regression. This was done to investigate for potential differences in the behavior of respondents residing in different parts of the city. Although the results showed no effect on the likelihood to accept to participate in a REC (no variation in σ^2 and ICC metrics), a potential effect could not be excluded due to the limited size of the sample and the unbalanced distribution of respondents within the city (as it was showed in Figure 19 and Figure 20).

4.4 Socio-demographic influence on value functions

As stated, Linear mixed models (LMM) are used to identify the set of socio-demographic vectors constituting the types of individuals in order to characterize the resident population and generalize the decision models approximated by the UTA method.

The analysis has been conducted discretizing the value functions for each respondent into equidistant segments. The coordinates of the resulting points have been used as observations on which to perform the analysis. This has ensured the reduction of any possible bias determined by higher density of observations around specific performance level. In Figure 35 an example is provided for the subdivided value function for investment cost attribute showing the points associated with specific marginal utility levels for the individuals in the dataset.

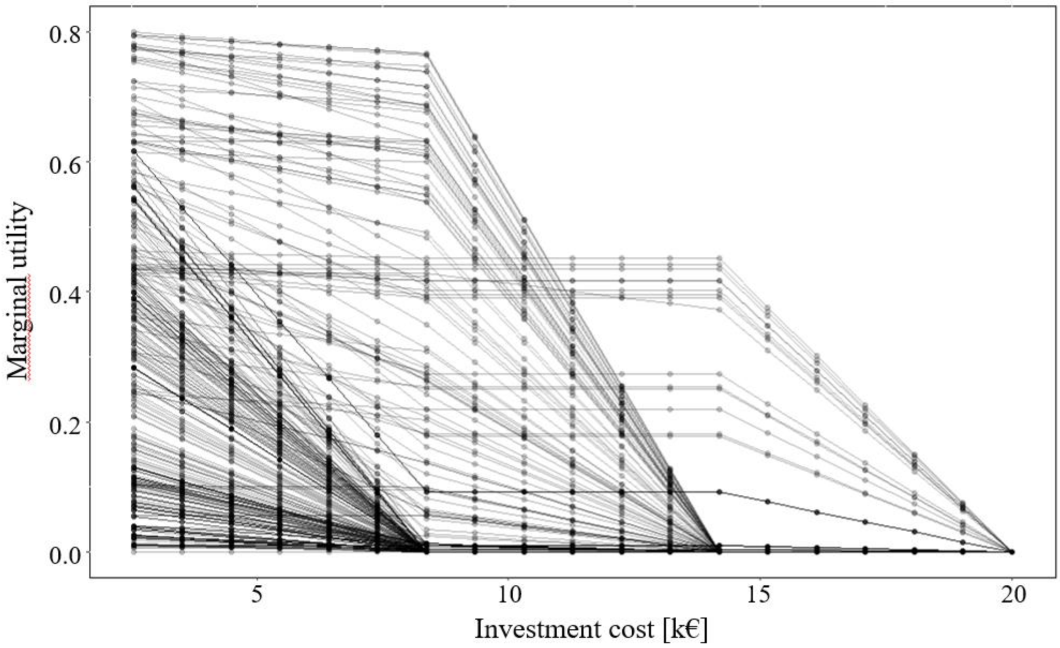


Figure 35 Value functions of the 349 respondents for investment cost performance attribute. The figure displays 248 unique value functions, 22 of them are shared by 2 individuals, 5 by 3 individuals, 4 by 4 individuals, while 3 are repeated 5 times. Four value functions are repeated 6, 10, 12, and 21 times respectively, while the remaining ones are only expression of single individuals.

For consistency, the same classes used in the previous section to identify the probability of accepting to participate have been used to reparametrize the socio-demographic attributes of the respondents. For each performance attribute, a first model tested the relation of the REC performance and socio-demographic characteristics of the individual with the marginal utility related to each attribute without including any moderating effect. Then, a full model for each attribute has been tested to account for moderating effects of the socio-demographic characteristics on the REC performance. From this full model, the less associated interaction effects have been removed following a backward step regression based on p-value and confidence intervals CI (the independent variable with the highest p-values in each iteration and with coefficient intervals including 0 were excluded). AIC and BIC values have been used to select the most parsimonious models. Each resulting model includes only the interaction effects having an influence on the association between the performance level of the REC in each specific attribute and the marginal utility related to that attribute, as well as the socio-demographic variables that are associated with a variation in marginal utility. The four models are displayed in Table 21, while the preparatory analysis are reported in Appendix F together with the indicators used to select the best models.

Table 21 Association of performance attributes and socio-demographic characteristics with the marginal utility of RECs attributes, with socio-demographic attributes acting as moderators. Associated variables are in boldface.

Models equations:

Model_9a: Y ~ Inv_cost+ socio-demographic attributes + Inv_cost * (Age + HH_num + HH_inc + Gender + HH_prop) + (1|seed)

Model_9b: Y ~ Cost_sav + socio-demographic attributes + Cost_sav * (HH_num + HH_inc + Gender + HH_prop) + (1|seed)

Model_9c: Y ~ Emis_sav + socio-demographic attributes + Emis_sav * (HH_num + HH_inc + Edu_lev + HH_prop) + (1|seed)

Model_9d: Y ~ Cost_sav + socio-demographic attributes + Cost_sav * (HH_num + HH_inc + Gender + HH_prop) + (1|seed)

Regression results:

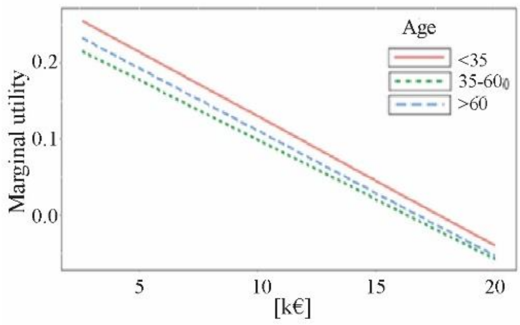
	Model_9a		Model_9b		Model_9c		Model_9d	
Predictors	β	p	β	p	β	p	β	p
Numerical								
Intercept	0.38	<0.001	-0.00	0.982	-0.02	0.341	0.07	0.016
Inv_cost (10k€)	**-0.22**	<0.001						
Cost_sav (k€)			**0.05**	<0.001				
Emis_sav(tCO_2)					**0.07**	<0.001		
SSuff (%)							**0.03**	<0.001
Categorical								
Age [< 35]	ref.							
Age [35-60]	**-0.04**	0.006	**-**		**-**		**-**	
Age [60+]	-0.02	0.214	**-**		**-**		**-**	
HH_num [>3]	ref.							
HH_num [3+]	**-0.05**	<0.001	-0.01	0.537	0.00	0.942	0.03	0.208
HH_inc [<28k]	ref.		ref.		ref.			
HH_inc [28-75k]	0.00	0.943	-0.01	0.414	0.01	0.633	-0.02	0.538
HH_inc [75k+]	**-0.05**	0.008	**-0.03**	0.042	-0.01	0.498	-0.02	0.556
Edu_lev [Dgr_N]	ref.							
Edu_lev [Dgr_Y]	-		-		-0.01	0.428	-	
Gender [Fem]	ref.							
Gender [Mal]	-0.01	0.240	-0.00	0.672	-		0.01	0.579
HH_prop [rent]	ref.							
HH_prop [own]	**-0.06**	<0.001	-0.01	0.461	0.01	0.491	0.04	0.113
Interactions								
* Age [35-60]	**0.01**	0.028						
* Age [60+]	0.01	0.464						
* HH_num [3+]	**0.03**	<0.001	**0.02**	<0.001	**-0.01**	0.002	**0.01**	<0.001

* HH_inc [28-75k]	0.00	0.995	**0.02**	<0.001	**0.01**	<0.001	**-0.00**	0.001
* HH_inc [75k+]	**0.03**	<0.001	**0.08**	<0.001	**0.01**	<0.001	**-0.01**	<0.001
* Edu_lev [Dgr_Y]	-		-		**0.01**	<0.001		
*Gender [Mal]	**0.02**	0.002	**0.01**	0.010	-		**0.00**	0.025
* HH_prop [own]	**0.04**	<0.001	**0.02**	<0.001	**-0.02**	<0.001	**0.01**	<0.001
Random Effects								
σ^2	0.01		0.01		0.00		0.01	
τ_{00}	0.01		0.01		0.01		0.05	
ICC	0.50		0.60		0.64		0.81	
Evaluation metrics								
Marginal R^2 / Conditional R^2	0.287 / 0.641		0.163 / 0.666		0.247 / 0.728		0.167 / 0.841	

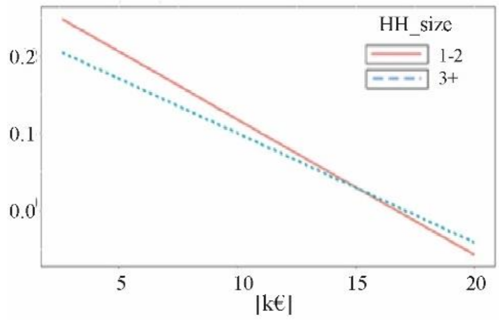
*Investment cost has been remapped to better grasp the effects of the performance. ** AIC and BIC not reported in the table, as no comparison among model is needed.

Considering Model_9a, the size of the household and the property of the dwelling are associated with the estimated marginal utility (β=-0.05 and β=-0.06 in Model_9a), while age and income seems associated with only one class out of two as fixed factors (age [35-60] with β=-0.04, and HH_inc [75k+] with β=-0.05).
The analysis reveals an interaction between the size (HH_num [3+]) and the property of the dwelling (HH_prop [own]) on the association between the performance attribute and the estimated marginal utility perceived. In particular, the positive coefficient associated to the interaction between household size and the investment cost (β =0.03) is balanced by the fixed effect of the household size in the marginal utility perceived by 1-unit increase in investment cost (β =-0.03), counterbalancing this counterintuitive association. The same trend is noticeable for the property of the dwelling, with owners associated with lower value of marginal utility as a fixed effect cost (β =-0.06 in Model_9a) but presenting a less steep descending curve in 1-unit increase thanks to the moderating effect (β =0.04). These effects are visible in Figure 36 where the association between Investment cost attributes and marginal utilities are plotted accounting for the moderating effects of the size and property of the households. In these curves it is possible to notice how the marginal utility perceived for both smaller size households and renters is higher in comparison with larger size households and owners, but due to the moderating effects the slope of their curve is steeper than those of the other two classes.
Particularly focusing on the moderating effect that income level has on investment cost (Figure 36.c), it is possible to notice that the effect is only associated for the

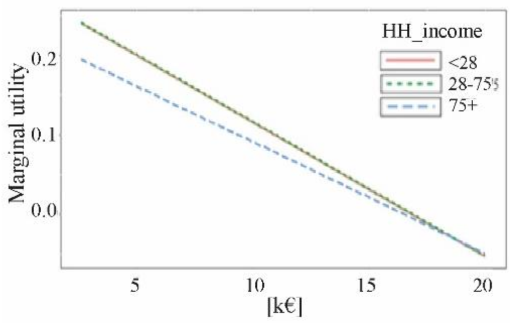
highest class (more than 75k euros), while the intermediate class (from 28k€ to 75k€) does not seem to be associated (p-value >0.05, CI=-0.01/0.01)



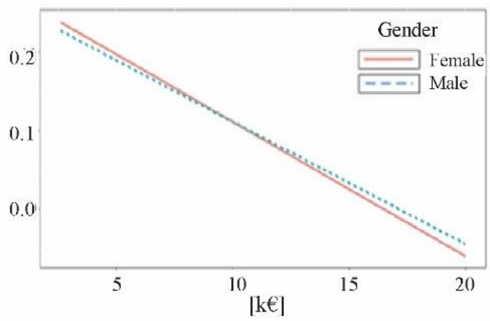
a. Association between Investment cost and odd to accept moderated by Age



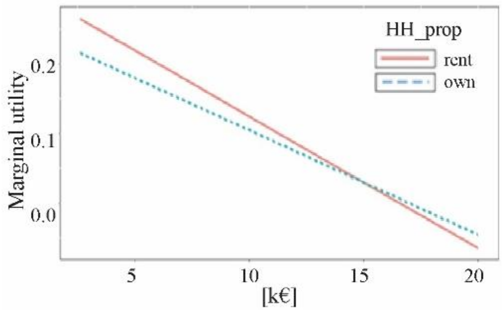
b. Association between Investment cost and odd to accept moderated by Household size



c. Association between Investment cost and odd to accept moderated by income level



d. Association between Investment cost and odd to accept moderated by Gender



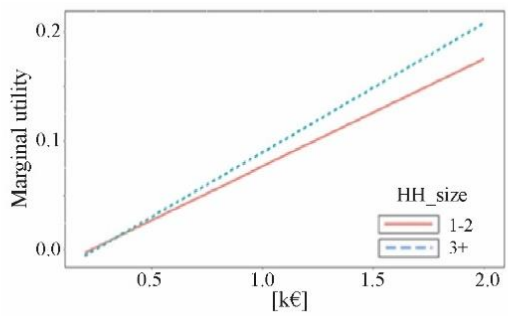
e. Association between Investment cost and odd to accept moderated by property condition

Figure 36 Association between Investment cost performance attribute and the odd to accept a REC with the moderating effect of associated socio-demographic caracteristics.

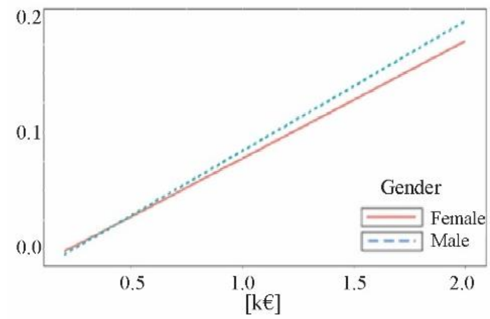
Gender seems to be a moderator too for this performance attribute (Figure 36.d), while age acts as moderator only for one class (Age [35-60]), decreasing the slope of the marginal utility curve for this socio-demographic class (Figure 36.a).

In Figure 37 the associations between cost saving performance attribute and marginal utility moderated by the socio-demographic attributes are displayed.

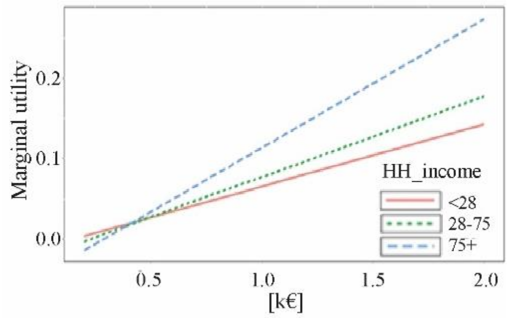
Also for this attribute (Model_9b) it is highlighted an interaction between the size (HH_num [3+] in Figure 37.a) and the property of the dwelling (HH_prop [own] in Figure 37.d) on the association with the estimated marginal utility perceived (β=0.02 for both socio-demographic characteristics).
Furthermore, households with higher incomes (Figure 37.c) show increasing marginal utility levels for higher performance of the RECs in this performance attribute (β=0.02 and β=0.08 for the class [28-75k], and class [75k+] respectively). Gender seems to be a moderator too (Figure 37.b), with males receiving slightly more marginal utility for each thousands of euros of cost savings granted by the REC.



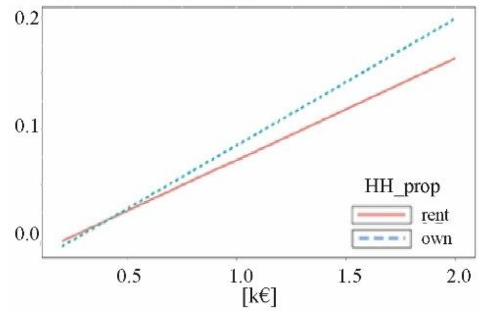
a. Association between Cost savings and odd to accept moderated by Household size



b. Association between Cost savings and odd to accept moderated by Gender



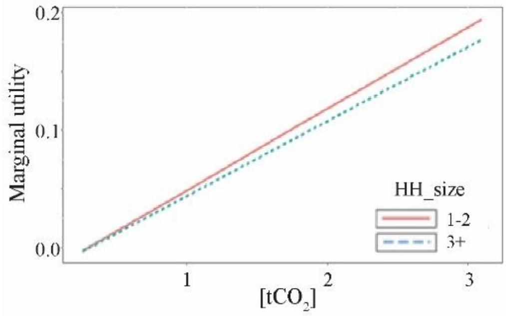
c. Association between Cost savings and odd to accept moderated by income level



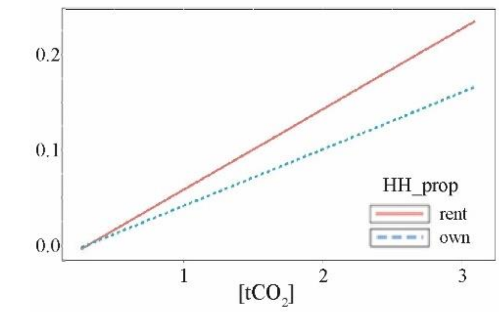
d. Association between Cost savings and odd to accept moderated by property condition

Figure 37 Association between cost savings performance attribute and the odd to accept a REC with the moderating effect of associated socio-demographic caracteristics.

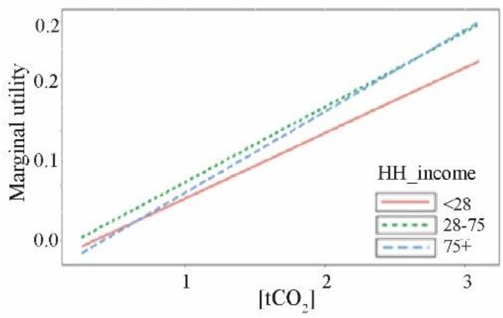
Regarding emission savings (Model_9c), the moderating effect of household size and the property of the dwelling is again present on the association between the performance attribute and the marginal utility received from it (Figure 38.a and Figure 38.b respectively), but in this case the smaller households as well as the renters seems to gain more marginal utility for increase in this performance attribute (β =-0.01 and β =-0.02 for HH_num [3+] and HH_prop [own] respectively). Again, the income level of the household act as a moderator (Figure 38.c) increasing the perceived marginal utility due to each ton of CO_2 emission spared (β =0.01 for both classes). Finally, unique case among all the performance attributes, the education level of the respondents acts as a moderator on the association between this attribute and related marginal utility (Figure 38.d), with higher educated individuals having a steeper curve compared to individuals not holding a degree (β =0.01).



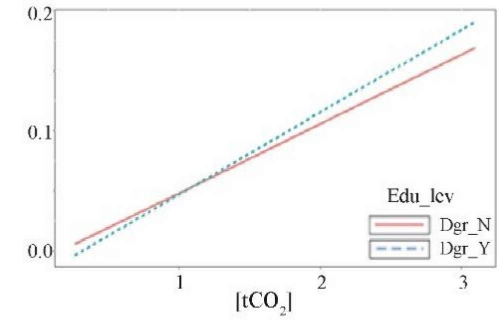
a. Association between Emission savings and odd to accept moderated by Household size



b. Association between Emission savings and odd to accept moderated by property condition



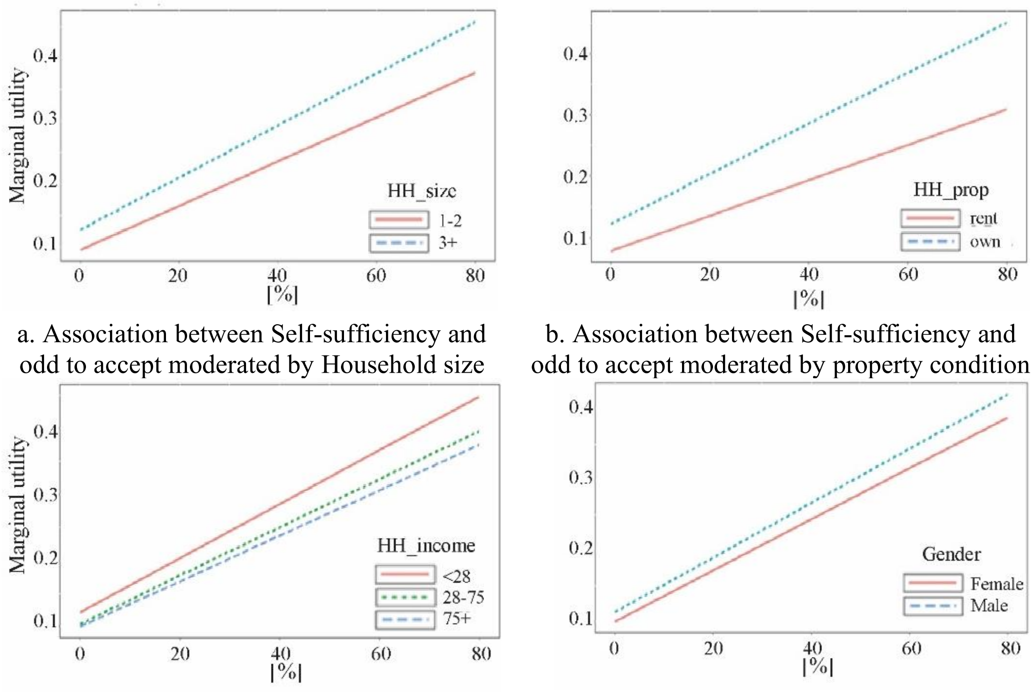
c. Association between Emission savings and odd to accept moderated by income level



d. Association between Emission savings and odd to accept moderated by education level

Figure 38 Association between Emission savings performance attribute and the odd to accept a REC with the moderating effect of associated socio-demographic caracteristics.

Finally, considering Model_9d, The association between Self-sufficiency performance attribute of RECs and marginal utility seems to be moderated by the household size (β =0.01 for households composed by more than 3 people as displayed in Figure 39.a), property of the household (β =0.01 for owners shown in Figure 39.b), and income level of the household (β =-0.00 and β =-0.01 for the two classes respectively as in Figure 39.c). The interaction of gender (Figure 39.d) seems to be associated too, but with a negligible coefficient (β =0.00).



a. Association between Self-sufficiency and odd to accept moderated by Household size

b. Association between Self-sufficiency and odd to accept moderated by property condition

c. Association between Self-sufficiency and odd to accept moderated by income level

d. Association between Self-sufficiency and odd to accept moderated by Gender

Figure 39 Association between Self-sufficiency performance attribute and the odd to accept a REC with the moderating effect of associated socio-demographic caracteristics.

Based on the previous analysis, it is possible to affirm that the six selected socio-demographic characteristics have a moderating effect on the association between the four performance attribute of the REC alternatives and their marginal utility. This result is summarized in Table 22 where the presence of an associated interaction between performance attribute of a REC and socio-demographic caracteristics is highlighted (marked with X), together with the number of modalities for each socio-demographic independent variable.

Table 22 Summary of the estimated interactions between socio-demographic characteristics and REC performance attributes ("X" represents an associated interaction).

	Age 3 classes	Gender 2 classes	HH_num 2 classes	HH_inc 3 classes	HH_prop 2 classes	Edu_lev 2 classes
Inv_cost (10k€)	X	X	X	X	X	-
Cost_sav (k€)	-	X	X	X	X	-
Emis_sav(tCO_2)	-	-	X	X	X	X
SSuff (%)	-	X	X	X	X	-

4.5 Generalization of the preference model

The six socio-demographic characteristics influencing the marginal utility highlighted at the end of the previous section have been combined to build the set of types of individuals. The number of types of similar individuals is a combination of classes (three for age and income of the household, and two for the other four socio-demographic characteristics) as follows:

$$\text{Number of types} = 2^4 * 3^2 = 144 \text{ types of individuals}$$

For instance, taking the first class of each socio-demographic characteristics, the description of one type of individual will be a “less than 35 years old female without a degree, living in a rented apartment with up to another person, with an annual income of less than 28k€”. On the opposite side of the spectrum (taking all the last classes of the socio-demographic characteristics), another type will be a “male older than 60 years old, with a degree title, living in an owned household with at least 2 other persons, and having an annual income of more than 75k€”.
In this following sub-section, the value functions approximated for each respondent by using UTA method (Section 4.3) are clustered and the characteristic sets of value functions for each cluster are extracted. These sets are then assigned to each of the 144 types of individuals. By matching the population in the city of Turin with the types of individuals, it will be then possible to spatialize the preference model at the urban scale. The final output of this section will be to assign the most probable sets of characteristics value functions to the different types of individuals in order to generalize the decision models estimated at the individual scale. The resulting match between decision models and socio-demographic variables will be the basis for the spatialization of the individuals’ decision models across the city of Turin. This section is composed of two parts. In the first part, an unsupervised classification algorithm is used to extrapolate clusters of value functions for each performance attribute and to identify the characteristic value function (the central one) for each cluster. In the second part, the types of individuals are assigned to one of the value function clusters by means of the application of a supervised classification algorithm.

4.5.1 Definition of characteristic value functions

The identification of the set of characteristic value functions has been performed by using k-medoids algorithm (Partitioning Around Medoids, PAM) on the sets of respondents' value functions approximated by UTA method. The clustering with PAM has been performed using fpc package in R environment [195], selecting the most suitable number of clusters based on the silhouette metric. In Figure 40 to Figure 43 the full range of value functions is displayed divided in the identified clusters. For each of these clusters the central value function (i.e., the medoid of the cluster) is represented with a thicker line.

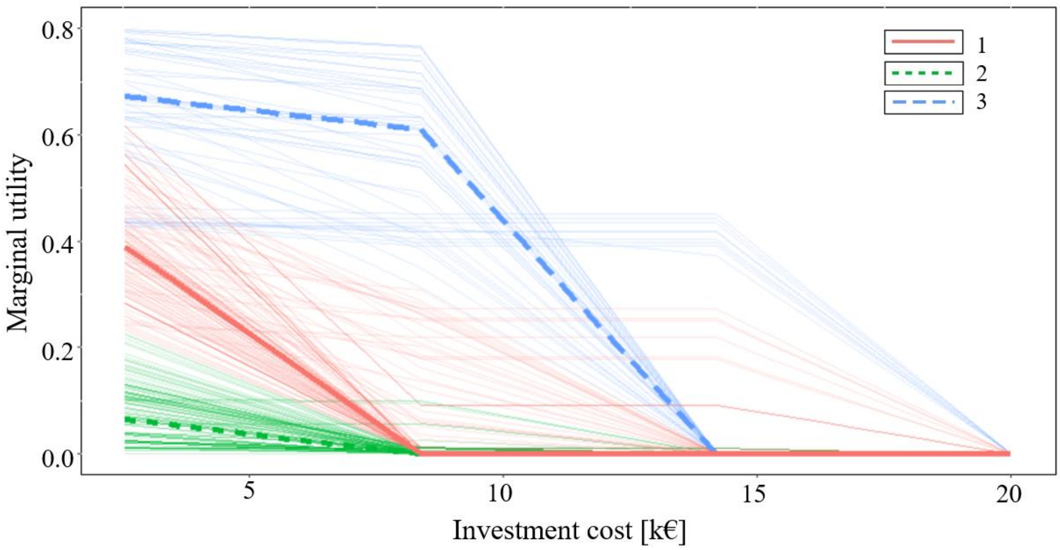


Figure 40 Clusters of investment cost value functions and identification of the characteristic value function for each of the clusters (thicker line).

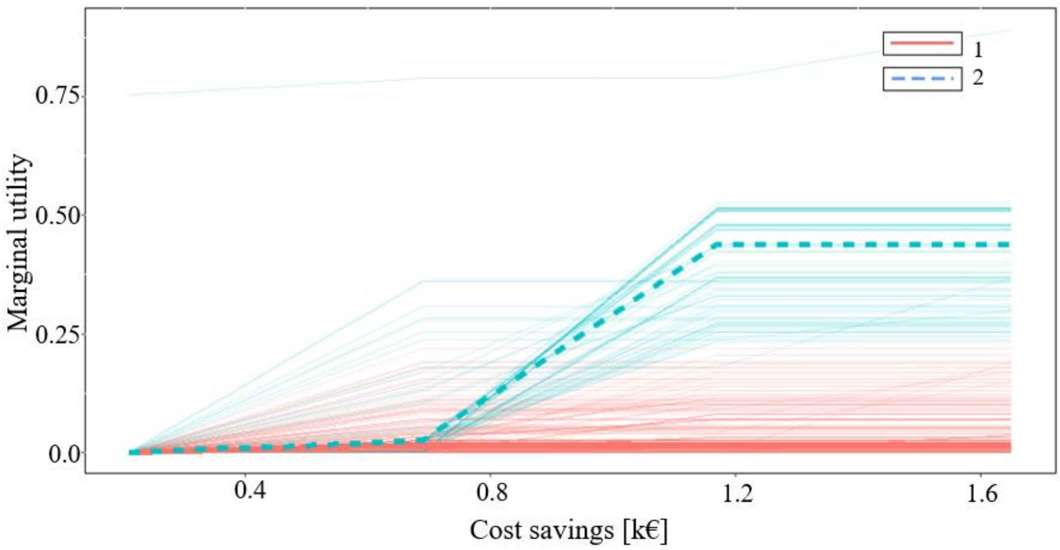


Figure 41 Clusters of cost savings value functions and identification of the characteristic value function for each of the clusters (thicker line).

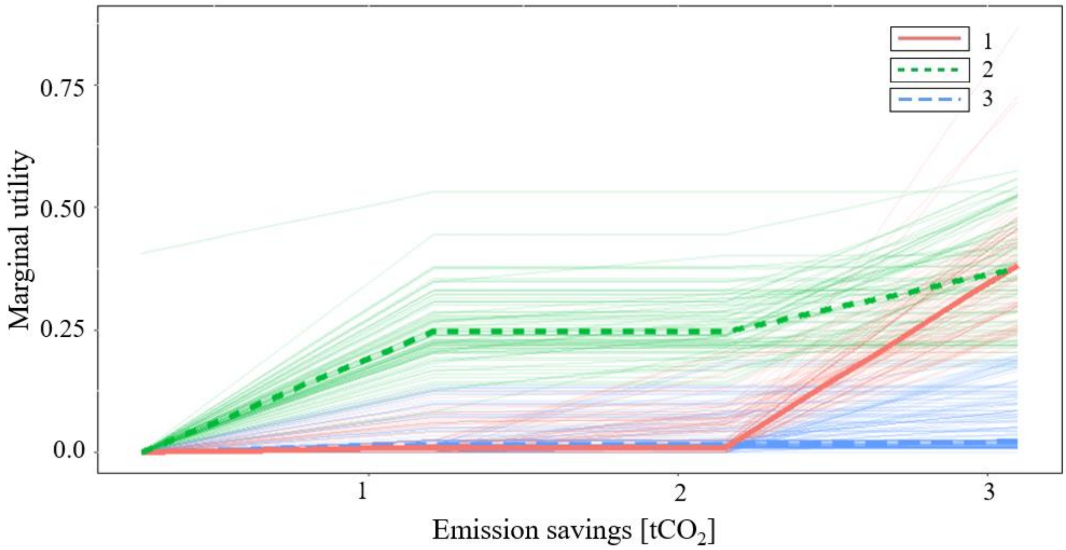


Figure 42 Clusters of emission savings value functions and identification of the characteristic value function for each of the clusters (thicker line).

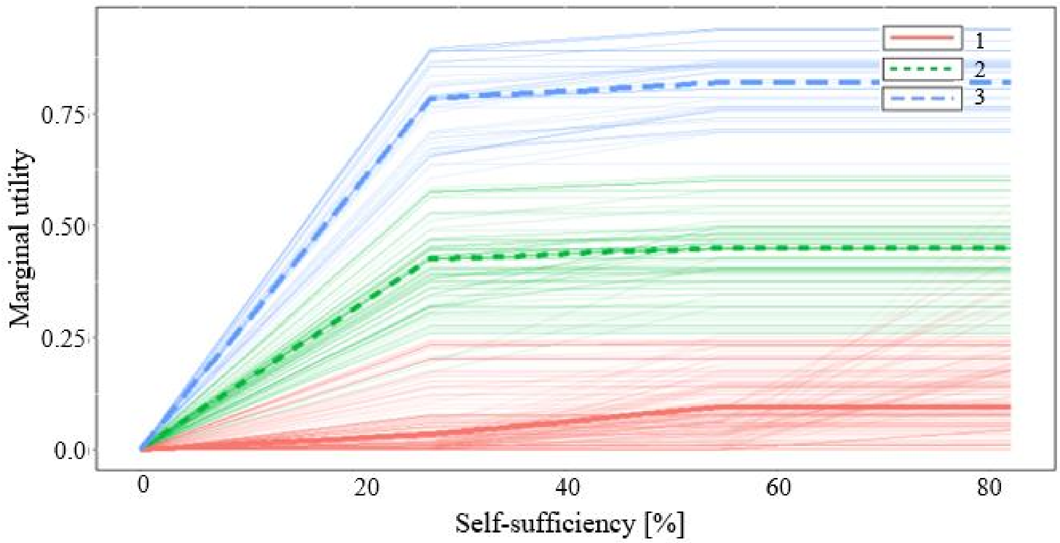


Figure 43 Clusters of self-sufficiency value functions and identification of the characteristic value function for each of the clusters (thicker line).

As shown in the figures, the most efficient number of partitioning clusters (and therefore of classes of characteristics value functions) according to the silhouette method is 3 for all attributes with the exception of cost savings for which the optimal number of clusters is 2. It is worth noting that such approach for the clustering of value functions into characteristic classes might result in a reduction of specificity of the information related to the description of the value functions. For instance, the low number of clusters determines that all of the classes of characteristics value functions (the medoid of the cluster) for the investment cost attribute assign a zero marginal utility for investment cost higher than 14k€. This determines some degree of information lost compared to the original value functions estimated for a number of respondents. The second issue is determined by the initial parametrization of the value function using only three segments to define them. While the first issue (number of clusters) might be solved by specifying a larger number of clusters, the second one could be easily solved by increasing the number of segments to define the piece-wise value functions. Due to the limited number of respondents, it would have been not appropriate to specify more clusters (with the risk of having clusters populated by a low number of types of individuals). Regarding the parametrization of the value function, this has been kept with 3 segments in their definition in order to test the methodological approach to the problem.

A sensitivity test has been conducted to evaluate the detriment between the ranking accuracy of the decision model approximated by the use of UTA method and the ranking accuracy determined by using the characteristic value functions (resulting from the PAM clustering) for the member of each specific cluster. In particular, for each respondent the total utility of the sub-set of alternatives selected and ranked by the same respondent have been calculated using the new set of value functions represented by the medoids of the clusters, and the number of pair-wise errors with respect to the original preference relation structure provided by the respondent have been calculated. Table 23 presents an example of the comparison between the original and the recalculated ranking for a random respondent.

Table 23 Example of an error matrix between the alternative ranking provided by a respondent and the one recalculated with the respontent assigned set of characteristic value functions.

Original ranking	**Recalculated U_{tot}**	**Recalculated ranking**	**Error matrix**						
			G	**F**	**B**	**E**	**D**	**C**	**A**
G	0.447	1	0	0	0	0	0	0	0
F	0.379	2	0	0	0	0	0	0	0
B	0.294	5	0	0	1	1	0	0	0
E	0.308	3	0	0	0	0	1	0	0
D	0.271	6	0	0	0	1	0	0	0
C	0.240	7	0	0	0	1	0	0	0
A	0.297	4	0	0	0	0	1	1	1

As shown in Table 23, the recalculated total utility for the respondent led to some error with respect to the original preference structure expressed. For instance, alternative B originally ranked 3rd by the respondent, ranks 5th in the recalculated rank. This means that both the preference relation B $\succ$ E, and B $\succ$ A are violated,

and obviously the reciprocal E $\nprec$ B, and A$\nprec$ B. This led to a total error with respect to the preference structure provided of 4. This could also be seen as the number of lower ranking indexes (intended as further from the preferred) preceding a certain object in the new ranking. For instance, taking alternative E, this was originally ranked 4th, while in the new order it was placed in the 3rd position. This means that by ordering the new ranking to the original ranking position, the index 3 is preceded by a lower ranking element (the index 5 in this case), this producing one error in the error count. It is worth noticing that the two ways of counting errors lead to a proportional number of errors (the first approach counting twice as many errors as the second one).
Comparing the errors committed by the ranking calculated applying the individual specific value functions and the clustered ones, it is possible to observe an increase in the number of ranking mistakes of 22.9%. These mistakes are concentrated in the individuals on the border of the clusters. Accepting some percentage of errors might be necessary in order to generalize the preference models to types of individuals at the city scale. Nonetheless, in order to reduce the number of errors it might be advisable to optimize the selection of the number of clusters including the minimization of the errors in the pair-wise preference structure as specified above.

4.5.2 Assignment of value function to types

The second part of the section aims to assign the characteristic value functions (the medoids of the clusters) to the 144 types of individuals. In this attempt a probabilistic approach has been adopted rather than a deterministic one, thus the probability of a type of individual being assigned a class of value functions has been calculated. Several supervised clustering algorithms have been tested in order to define the one most fitting for the above specified goal.
To properly evaluate the performance of the classification model, a 10-fold cross validation has been used [228]. K-fold cross validation requires splitting the dataset in k-parts (defined as folds) and to train the model on all the subsets but one that is used for the testing. The process is then iterated k-times using each time a different subset of size 1/k to perform the testing, and the performance of the k-models are averaged. Such a procedure is suitable to better assess the performance of the model on unseen data and avoid overfitting. The Caret package in R environment allows to perform an automatic grid search for the fine tuning of the hyperparameters used in each classification algorithm and to select the best ones based on the selected metric [197]. In the specified models, the fine-tuning parameter has been set to 5, with a total number of models tested equal to 5^p for each algorithm, where p is the

number of hyperparameters that could be specified for each classification algorithm, and the selection metric being the accuracy of the model.

In Table 24 the tested algorithms are displayed together with the calculated performance metrics. The selection of the best performing classifier is based on the average of Accuracy and F_1score metrics across the four attributes of performance of REC (investment cost, cost savings, emission savings, and self-sufficiency).

Table 24 Performance metrics of the supervised classification algorithm tested.

		Attribute			
Algorithm	**Metric**	**Inv_cost**	**Cost_sav**	**Emis_sav**	**SSuff**
Naive Bayes	Accuracy	0.5091	0.765	0.4613	0.5119
	Average Precision	0.3378	0.3825	0.3599	0.2546
	Average Recall	0.3947	0.5	0.3387	0.3317
	Weighted F_1score	0.4692	0.6632	0.3380	0.3526
Support Vector Machine (linear)	Accuracy	0.4776	0.765	0.4776	0.5043
	Average Precision	0.3159	0.3825	0.2979	0.3077
	Average Recall	0.3678	0.5	0.3829	0.3517
	Weighted F_1score	0.4318	0.6632	0.4073	0.4175
Support Vector Machine (Polynomial)	Accuracy	0.5043	0.7746	0.4699	0.5234
	Average Precision	0.3327	0.6810	0.1566	0.3313
	Average Recall	0.3892	0.5612	0.3333	0.3745
	Weighted F_1score	0.4595	0.7220	0.3004	0.4472
Rule-Based Classifier (JRip)	Accuracy	0.5043	0.7689	0.4699	0.5072
	Average Precision	0.3353	0.6621	0.1111	0.2881
	Average Recall	0.3869	0.5363	0.3337	0.3350
	Weighted F_1score	0.4473	0.7013	0.3029	0.3755
Rule-Based Classifier (PART)	Accuracy	0.5005	0.7727	0.4527	0.5072
	Average Precision	0.4010	0.6695	0.3881	0.3915
	Average Recall	0.3912	0.5726	0.3776	0.3638
	Weighted F_1score	0.4657	0.7291	0.4134	0.4355
C5.0	Accuracy	0.5091	0.765	0.4728	0.5244
	Average Precision	0.3400	0.6335	0.3806	0.3383
	Average Recall	0.3960	0.5056	0.3752	0.3711
	Weighted F_1score	0.4704	0.6702	0.3911	0.4430
Single5.0Tree	Accuracy	0.5033	0.7622	0.4737	0.5119
	Average Precision	0.3383	0.5496	0.3640	0.3766
	Average Recall	0.3913	0.5023	0.3722	0.3647
	Weighted F_1score	0.4610	0.6670	0.3838	0.4370
C4.5-like Trees J48	Accuracy	0.4976	0.7736	0.4976	0.5158
	Average Precision	0.3297	0.6768	0.4510	0.1719
	Average Recall	0.3831	0.5605	0.4020	0.3333
	Weighted F_1score	0.4494	0.7213	0.4207	0.3510
Penalized Multinomial Regression	Accuracy	0.4909	0.7593	0.4661	0.4986
	Average Precision	0.3252	0.3818	0.3981	0.4824
	Average Recall	0.3802	0.4962	0.3696	0.3502
	Weighted F_1score	0.4514	0.6604	0.3998	0.4101

From the performance displayed in Table 24 it could be noted that the overall accuracy of all the tested algorithms is quite limited. This could be explained mainly due to the limited size of the dataset and its imbalance across different classes (among the reason for the low performance is that some classes would be less represented in the partitioning of the original dataset into training and testing sub-sets when performing the 10-fold cross validation). The issue related to the size of the dataset and the consequent limitations for the 10-fold cross validation justifies the choice made in the previous sub-section to limit the number of clusters used in the partition around medoids. In fact, increasing the number of classes that the supervised classification algorithm has to predict would have further reduced the performance of the classifier (with the same small dataset). Again, it could be argued that a larger dataset of respondents (and a better constructed one) would have reduced the number of constraints that the application of the method had to face.

The best classifier in terms of accuracy averaged across the four value functions for each performance attribute is C4.5-like Trees J48 algorithm (0.571 averaged Accuracy), followed by the polynomial Supporting Vector Machine (0.568 averaged Accuracy); while the best one in terms of weighted F-1 score averaged across the four value functions is the Rule-Based Classifier (PART) with an average weighted F_1score of 0.510, followed by the C5.0 (0.494 averaged weighted F-1). Finally, by averaging the two metrics (Accuracy and F_1 score), it results that the best performing classifier is Rule-Based Classifier (PART) algorithm (0.532) followed by C5.0 (0.531) and the C4.5-like Trees J48 (0.528).

PART algorithm is a rule-based algorithm that uses partial decision trees to produce classification rules [229] and is implemented in Caret by calling the RWeka package in Rstudio [230]. Among the advantages of rule-based algorithms is the enhanced interpretability of the model results. In fact, it is possible to examine the rule path followed by the algorithm to classify an instance in one class rather than another. In the following Table 25 the different rules that associate each type of respondent based on the combination of socio-demographic attributes with a characteristic value function (the medoid of each cluster in previous sub-section) are reported.

Table 25 PART algorithm rules for the assignment of characteristic value fuctions to the types of respondents

Rules		**class**	
ID	Investment cost If…		
1	HH_prop is owner & HH_num is 3+ & Age between 35 and 60y.o.	2	(78/30)
2	HH_inc is above 75k€ & HH_prop is rent & HH_num is 1 or 2	1	(3/0)
3	HH_inc is not above 75k€ & HH_num is 1 or 2 & Gender is female	1	(102/52)
4	HH_inc is above 75k€	2	(46/18)
5	HH_inc is not 28k€-75k€ & Age more than 60y.o.	1	(5/1)
6	Age is not over 60y.o. & Edu_lev is with a degree & HH_inc is 28k€-75k€ & HH_prop is rent	1	(32/14)
7	Age is not over 60y.o. & Edu_lev with a degree & HH_inc is not 28k€-75k€	1	(36/19)
8	Age is not in range 35-60y.o. & HH_inc is in range 28k€-75k€ & Edu_lev with a degree	2	(31/11)
9	None of the above	1	(16/7)
ID	Cost savings If…		
1	HH_inc is less than 28k€	1	(98/13)
2	HH_prop is owner & HH_inc is 28k€-75k€ & Edu_lev with a degree & Age is not over 60y.o.	1	(83/18)
3	HH_prop is rent	1	(60/8)
4	Gender is male & Edu_lev with a degree & HH_inc is not 28k€-75k€ & Age is not over 60y.o.	1	(26/9)
5	Age is not in range 35-60y.o. & HH_num is 1 or 2 & Edu_lev with a degree & HH_inc is 28k€-75k€	1	(17/5)
6	Gender is male	1	(38/13)
7	Edu_lev with a degree	2	((22/7)
8	None of the above	1	(5/1)
ID	Emission savings If…		
1	HH_prop is owner & HH_num is 3+	3	(102/41)
2	Age more than 60y.o. & HH_prop is owner & HH_inc is not above 75k€ & Edu_lev with a degree	1	(22/11)
3	Age more than 60y.o. & HH_prop is owner	3	(31/12)
4	Age is not over 60y.o. & HH_inc is not 28k€-75k€	3	(95/49)
5	HH_inc is in range 28k€-75k€ & Edu_lev with a degree & HH_num is 1 or 2 & Age is in range 35-60y.o.	1	(31/16)
6	HH_inc is in range 28k€-75k€ & HH_prop is rent	2	(43/22)
7	Age is not over 60y.o	3	(21/9)
8	None of the above	1	(4)

ID	Self-sufficiency If…		
1	HH_prop is rent & HH_inc is not above 75k€ & Age is not over 60y.o. & HH_inc is in range 28k€-75k€	1	(51/15)
2	HH_inc is above 75k€ & Edu_lev with a degree & HH_prop is owner & Gender is female	2	(19/9)
3	HH_inc is above 75k€ & Edu_lev with a degree & Gender is male & HH_prop is owner & Age is over 60y.o. & HH_num is 1 or 2	1	(13/6)
4	HH_inc is above 75k€ & Edu_lev with a degree	2	(34/15)
5	HH_prop is rent & Age is not over 60y.o. & HH_num is 3+ & Edu_lev with a degree	1	(12/2)
6	HH_num is 1 or 2	1	(144/65)
7	Age is not over 60y.o.	1	(66/39)
8	None of the above	2	(10/4)

*for a more clear interpretation, the classes of socio-demographic characteristics used by the algorithm to define the classification rules are expressed in a more readable way. For instance, rule 1 of investment cost performance attribute is HH_prop=[own] & HH_num=[3+] & Age=[35-60].

From the rules displayed in Table 25 it is possible to notice that the classification model is not well performing in predicting class 3 for both the Investment cost and the self-sufficiency performance attribute. Indeed, no rule leads to the assignment of this characteristic value functions to any combination of socio-economic attribute. Also for the cost saving performance attribute the classifier tends to be biased toward the first class of value functions, being seven out of eight rules pointing to class 1. This tendency could be determined both by the limited size of the dataset of respondents and the unbalanced frequency of classes. This latter aspect supports the argument in favor of opting for a probabilistic assignment of value functions to types of individuals, rather than using a deterministic approach. Figure 44 to Figure 47 graphically shows this issue by plotting the frequency of respondents in each profile of type of individuals defined by the vector of socio-demographic attributes clustered by class of value function. In the same figures, the type of individuals among the 144 ones individuated by the combination of socio-demographic attributes that were not observed in the respondents' dataset are displayed in grey color.

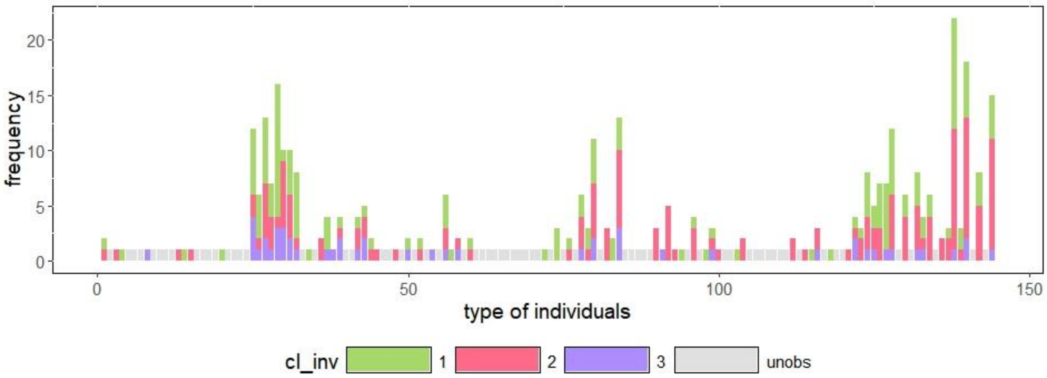


Figure 44 Subdivision of the 349 respondents in types of individuals and clusters of investment cost value functions

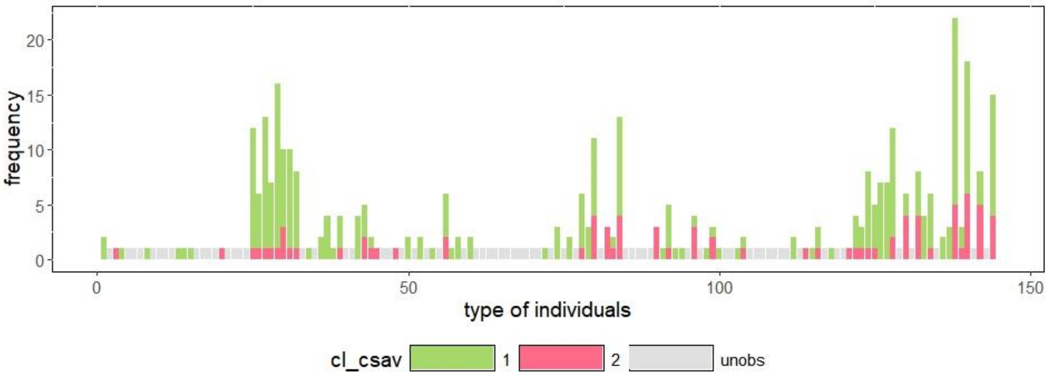


Figure 45 Subdivision of the 349 respondents in types of individuals and clusters of cost savings value functions

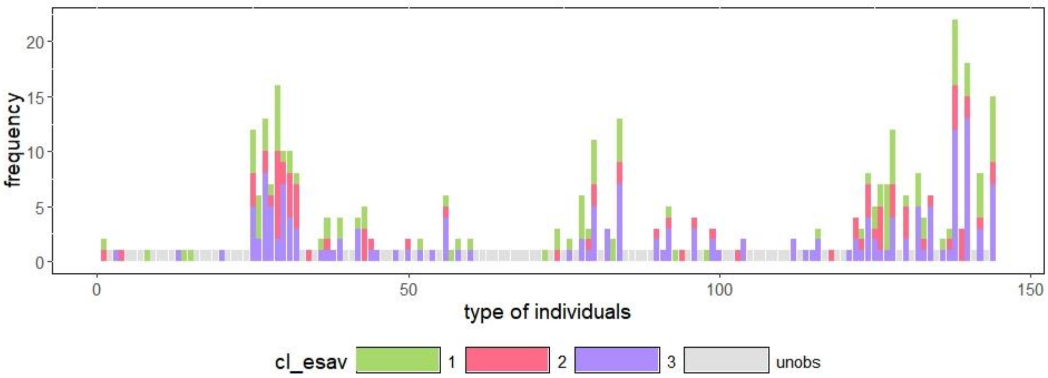


Figure 46 Subdivision of the 349 respondents in types of individuals and clusters of emission savings value functions

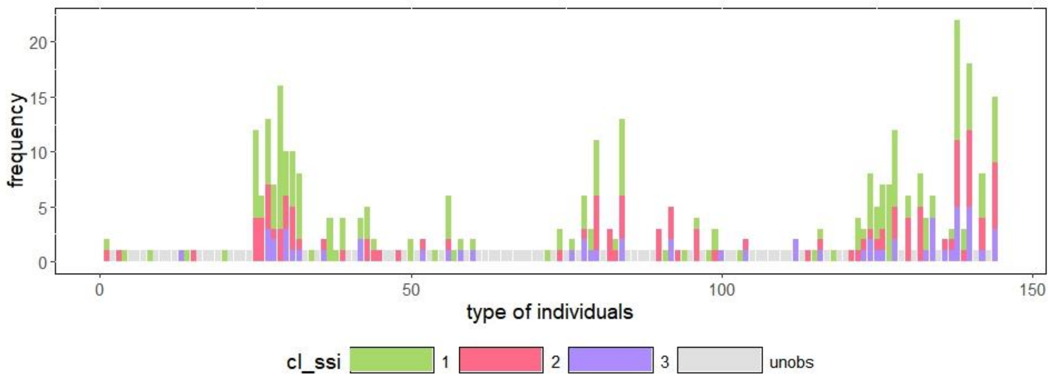


Figure 47 Subdivision of the 349 respondents in types of individuals and clusters of self-sufficiency value functions

Looking at the figures above it is again possible to notice the unbalance of the dataset toward certain classes of value functions. For instance, class 3 is less represented compared to class 1 and 2 in the investment cost performance attribute (Figure 44), while class 1 in predominant in the cost saving performance attribute (Figure 45). Moreover, the different types of respondent present a quite dispersed propension toward one class of value function with respect to the others, with different probability to be in a cluster of characteristic value function. By looking at the figures it is also possible to notice the quite high incidence of unobserved types of respondents in the dataset: out of the 144 theorical types of individuals constructed by the combination of socio-economic attributes, 67 were not observed in the dataset containing the responses to the questionnaire.
To generalize the decision model to all the different types of individual, the probability of both observed and unobserved types to be assigned one value function or another has been estimated using the outcome of the classification algorithm to the totality of the 144 types of individuals. In Table 26 the probability of belonging to one class or another of caracteristic value functions is reported for eight examples of both observed and unobserved types of individuals.The types of individuals reported in the table share the same level of socio-demographic classes for education level, size of the household, and income of the households, while varying in age. The first four individuals are in the “<35 years old” class, while the last four in age class “35-60 years old”. Finally the profiles are completed with the combination of the last two socio-demographic attributes determining the vector of socio-demographic attributes (gender and property of the household).

Table 26 Probability of a subset of types of individuals to be assigned to one cluster of value functions. The most probable class for each type is displayed in boldface.

Type	Age, Edu_lev, HH_size, HH_inc	Gender	HH_prop	Inv_cost		Cost_sav		Emis_sav		SSuff	
				Class	%	Class	%	Class	%	Class	%
	Age = <35, Edu_lev = Dgr_N, HH_size = <3, HH_inc = <28k€			**1**	**56.2**	**1**	**86.7**	1	29.5	**1**	**54.9**
1		fem	rent	2	37.5	2	13.3	2	22.1	2	27.1
				3	6.2	-		**3**	**48.4**	3	18.1
				1	**56.2**	**1**	**86.7**	1	29.5	**1**	**54.9**
2*		fem	own	2	37.5	2	13.3	2	22.1	2	27.1
				3	6.2	-		**3**	**48.4**	3	18.1
				1	**49.0**	**1**	**86.7**	1	29.5	**1**	**54.9**
3		mal	rent	2	37.2	2	13.3	2	22.1	2	27.1
				3	13.7	-		**3**	**48.4**	3	18.1
				1	**49.0**	**1**	**86.7**	1	29.5	**1**	**54.9**
4		mal	own	2	37.2	2	13.3	2	22.1	2	27.1
				3	13.7	-		**3**	**48.4**	3	18.1
	Age = 35-60, Edu_lev = Dgr_Y, HH_size = 3+, HH_inc = 28-75k€			**1**	**56.2**	**1**	**86.7**	1	27.9	**1**	**70.6**
113*		fem	rent	2	37.5	2	13.3	**2**	**48.8**	2	23.5
				3	6.2	-		3	23.2	3	5.9
				1	32.0	**1**	**80.0**	1	24.5	**1**	**40.9**
114		fem	own	**2**	**61.5**	2	20.0	2	15.7	2	27.2
				3	6.5	-		**3**	**59.8**	3	31.8
				1	**56.2**	**1**	**86.7**	1	27.9	**1**	**70.6**
115		mal	rent	2	37.5	2	13.3	**2**	**48.8**	2	23.5
				3	6.2	-		3	23.2	3	5.9
				1	32.0	**1**	**65.8**	1	24.5	**1**	**40.9**
116		mal	own	**2**	**61.5**	2	34.2	2	15.7	2	27.2
				3	6.5	-		**3**	**59.8**	3	31.8

*unobserved type

Once the probability of each type of individual to be assigned a class of value functions is calculated, it is possible to spatialize the decision models by matching the type of individuals with the occurrency of the same types at the urban scale. In the next section the generation of the synthetic population will be used to determine the number of individuals in the urban area belonging to one or the other of the 144 identified types.

4.6 Population synthesis

As mentioned in the method section (Section 3.4.1), the population synthesis process allows to reconstruct the individual-level socio-demographic characteristics of a population in an area by expanding a sample of disaggregated data (the seeds dataset), to match the marginal distribution of population's characteristics (the marginal control dataset). The output of the synthetic population is an artificial population with individual-level granularity of socio-demographic data whose distribution is coherent with that of the real population living in a specified area.
Together with variables used to expand the seed to match the marginal distribution, the seeds sample could contain further information not controlled via the marginal distribution of variables. In the application under investigation this was the case for two socio-demographic information contained in the seeds dataset: dwelling property, and source of income of the household. Regarding the property of the dwelling (HH_prop) this is provided in the seeds dataset, but it is not controlled by a marginal distribution. Regarding the source of income of the household, this variable has been included to estimate the households' level of income (described further on), and its marginal distribution is controlled only at the employed/unemployed level.
As stated previously, the generation of the synthetic population has been performed using PopulationSim library in Python environment [200], using population data from the Italian national census institute: the aspect of daily life (Aspetti della vita quotidana, AVQ) and the 2021 census survey datasets.
The AVQ provides information at the individual level (the seeds dataset), while the census survey dataset provides the marginal distribution of socio-demographic data at the census tracks scale. A dedicated Python script has been written (Appendix G) to split the AVQ dataset into the two required datasets (Persons and Households seeds datasets).
Finally, the controls file allows to specify the link between seeds and marginal controls datasets. The specified controls csv file is reported in Table 27.

Table 27 Defined controls.csv file for the allocation of seeds population. The number of individuals in the seeds ("seed table") are expanded to match each variable distribution in the marginal controls dataset ("control field") in a specific census track ("geography") with different matching priorities among variables ("importance").

target	**geography**	**Seed table**	**importance**	**Control field**	**expression**
Num_HH	SEZ21_ID	households	1000000	HH_BASE	(households.WGTP > 0) & (households.WGTP < np.inf)
Num_pp	SEZ21_ID	persons	50000	Pop_Base	(persons.wgtp > 0) & (persons.wgtp < np.inf)
HH_size_1	SEZ21_ID	households	10000	HH_size1	households.NP == 1
HH_size_2	SEZ21_ID	households	10000	HH_size2	households.NP == 2
HH_size_3	SEZ21_ID	households	10000	HH_size3	households.NP == 3
HH_size_4	SEZ21_ID	households	10000	HH_size4	households.NP == 4
hh_size_5	SEZ21_ID	households	10000	HH_size5	households.NP == 5
hh_size_6+	SEZ21_ID	households	10000	HH_size6	households.NP >= 6
Males	SEZ21_ID	persons	500	P_Male	persons.Gender==1
Females	SEZ21_ID	persons	500	P_Female	persons.Gender==2
Degree_Y	SEZ21_ID	persons	500	P_degree	persons.Edu_lev==1
p_Age19	SEZ21_ID	persons	500	AGE_19	persons.AGE_p<=7
p_Age24	SEZ21_ID	persons	500	AGE_24	persons.AGE_p==8
p_Age34	SEZ21_ID	persons	500	AGE_35	persons.AGE_p==9
p_Age44	SEZ21_ID	persons	500	AGE_44	persons.AGE_p==10
p_Age54	SEZ21_ID	persons	500	AGE_54	persons.AGE_p==11
p_Age59	SEZ21_ID	persons	500	AGE_59	persons.AGE_p==12
p_Age64	SEZ21_ID	persons	500	AGE_64	persons.AGE_p==13
p_Age74	SEZ21_ID	persons	500	AGE_74	persons.AGE_p==14
p_Age75	SEZ21_ID	persons	500	AGE_75	persons.AGE_p==15
p_Occ	SEZ21_ID	persons	500	OC_TOT	persons.OCCUP==1

As specified in Table 27, the variables for which the algorithm had to attempt a closer match with the marginal distribution, thus those with a higher importance value are the total number of households (importance 1000000), and the amount of population (importance 50000). Then, the sizes of the households are tried to be matched (HH_size_1 to HH_size_6) with a slightly lower importance. Finally, the gender of the individuals, their education level (Degree_Y registering the individual with at least a degree), their age (p_Age19 to p_Age75), and the employment status (p_Occ) are those variables for which the larger deviations from the marginal distribution of socio-demographic characteristics are allowed.

The AVQ dataset and the census dataset share the same classes for most of the socio-demographic variables (household sizes from 1 to 6+, gender, education level, and employment status), while the classes of age variable needed a first remapping to match the partitions in the census survey datasets. With specific reference to the occupation variable, it is worth noting that this variable was not used in the construction and assignment of the decision models in the previous sections (Section 4.3.2 and Section 4.5.2). Nevertheless, this variable has been included to further constrain the expansion of the seeds dataset to match the marginal data, better fitting the distribution of population across census tracks. Furthermore, the inclusion of occupation status in the controlled variables was useful to provide information regarding individuals recipient of an income or not to then estimate the income level of the households. In Table 28 a sample of the synthetic population generated is displayed.

Table 28 Example of the final result of the synthetic population generation process. The first 8 rows represent 5 households in the first census track identified by National tessellation code (SEZ21_ID=12720000001). The last two rows represent the last generated household (HH_ID = 433,469) in the last census track (SEZ_ID=12728888888).

SEZ21_ID = 12720000001							
HH_ID	**Person number**	**Age (y.o.)**	**Gender**	**Education level**	**Employed**	**Income source**	**HH-property**
1	1	30-34	Male	High school	Yes	Employee	Owned
1	2	25-29	Female	High school	No	Allowance	Owned
2	1	50-54	Male	Secondary	No	Pension	Owned
3	1	55-59	Female	Primary	No	Allowance	Owned
4	1	30-34	Female	High school	Yes	Self-empl	Owned
5	1	35-39	Male	Degree/+	Yes	Employee	Owned
5	2	35-39	Female	Degree/+	Yes	Self-empl	Owned
5	3	11-13	Male	Primary	N.A.	N.A.	Owned
…							
SEZ21_ID = 12728888888							
HH_ID	**Person number**	**Age (y.o.)**	**Gender**	**Education level**	**Employed**	**Income source**	**HH-property**
…							
433469	1	25-29	Female	High school	Yes	Support	Owner
433469	2	3-5	Male	N.A.	N.A.	N.A.	Owner

As from Table 28, the first column represents the ID of the household in the generated synthetic population, while the second one identifies the progressive number of the person in the household. Columns three to six are the socio-demographic characteristics controlled by the marginal control dataset, while the last two columns (income source and property regime of the household) are individuals' characteristics registered in the seeds dataset that are not controlled by the marginal distribution provided at the census track level.

A total population of 837,559 individuals distributed in 433,469 households has been generated for the whole city of Turin. In the following figures (from Figure 48 to Figure 52) the differences between the generated synthetic population and the real marginal distribution of a sub-set of socio-economic variables have been plotted and spatialized in order to check the reliability of results and to quantify the magnitude of errors committed by the algorithm.

In the figures, the left-hand side shows the absolute difference between the marginal data and the generated synthetic population ($\text{census data}_i - \text{sythetic population value}_i$ for variable i), while the right-hand side shows the relative difference calculated as the ratio between the absolute difference and the census data ($\text{absolute difference}_i$ / census data_i for variable i). It is possible to notice that the difference between the census marginal data and the generated synthetic population is generally limited. Some exceptions are present in a few of the census tracks for which the synthetic population presents a lower number compared to the marginal data provided by the census survey, but always limited considering the absolute population.

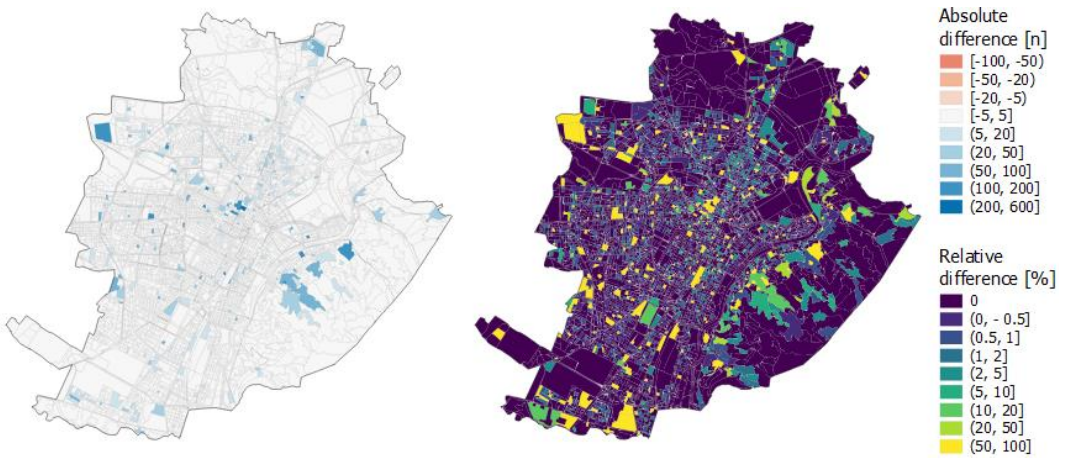


Figure 48 Population difference. 4916 census tracks show no difference between census data and simulated population. 1 census track shows a difference of 584 people, 49 a difference between 50 and 186 people.

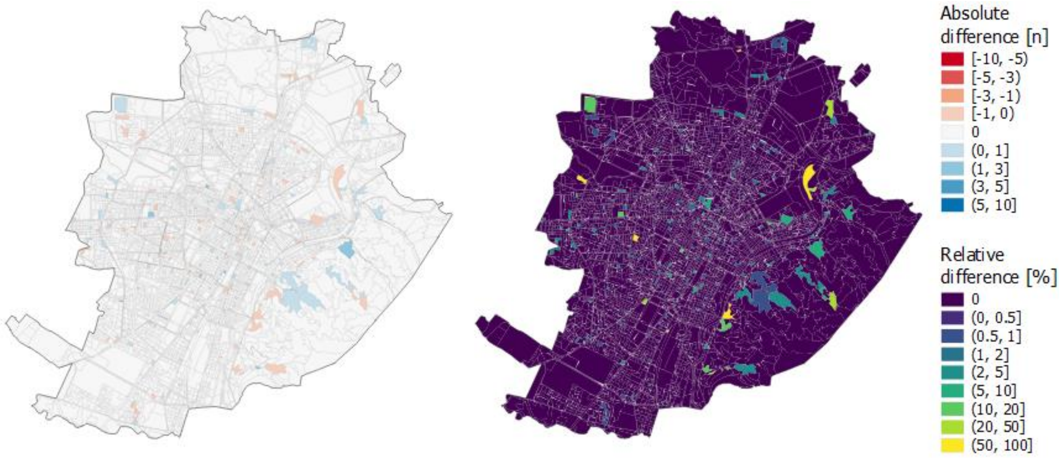


Figure 49 Difference in the number of one person households. 6640 census tracks show no difference compared to the marginal data. 8 census tracks show an error between 2 and 3 households less than the marginal data.

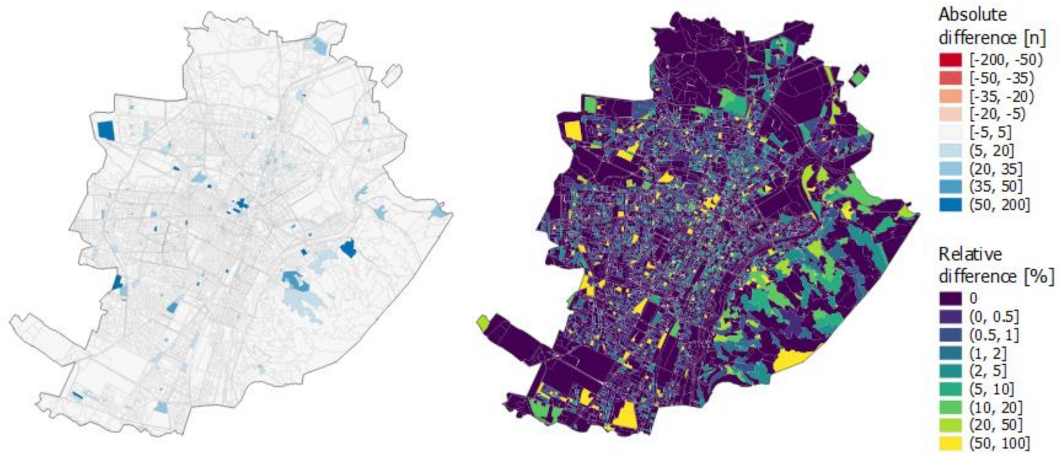


Figure 50 Difference in the number of males. 4731 census tracks show no difference with the marginal data. 3 census tracks differ by a number between 106 and 166 people less in the synthetic population. 100 by a number between 10 to 100 less, while 67 census tracks result in 2 or 3 males more than the marginal data.

Figure 51 Difference in the number of people with a degree. 4955 census tracks show no difference. 1 census track show 52 people with a degree less than the marginal data, while 12 show a number of people with a degree between 10 and 27 less. All the other census tracks show errors within -4 and 10 people.

Figure 52 Difference between population in age class less than 10 years old. 4895 census tracks show no difference compared to the marginal data. 1 census track shows 53 people less in this age class compared to marginal data, while all the other errors are in a range from -4 to 48 difference in number of people.

Neither the AVQ dataset, nor the census marginal data provide information regarding the income of inhabitants. This represents a problem in the attempt to spatialize the decision model evaluated in the previous sections, being this socio-demographic characteristic among the variables influencing both the willingness to participate in alternative RECs, and the estimated rank of different alternative configurations. To solve this issue the information regarding the tax declaration of the population provided yearly by the national income revenue authority ("Agenzia delle entrate") has been used. Such dataset contains the average gross incomes of individuals for difference source of income aggregated at the postcode scale [231], thus, the households' income level has been estimated by matching the source of income of individuals in the synthetic population generated with the information provided by the tax declaration dataset. First, the different income levels by type of source from the postcode scale has been assigned to each census track. In particular, in the city of Turin there are 33 postcode zones in total, thus determining a quite coarse resolution in the assignment of income levels. In Figure 53 it is possible to see the different spatial resolution of postcodes and census tracks in the city of Turin, with the displayed number showing the number of census tracks belonging to the same postcode.

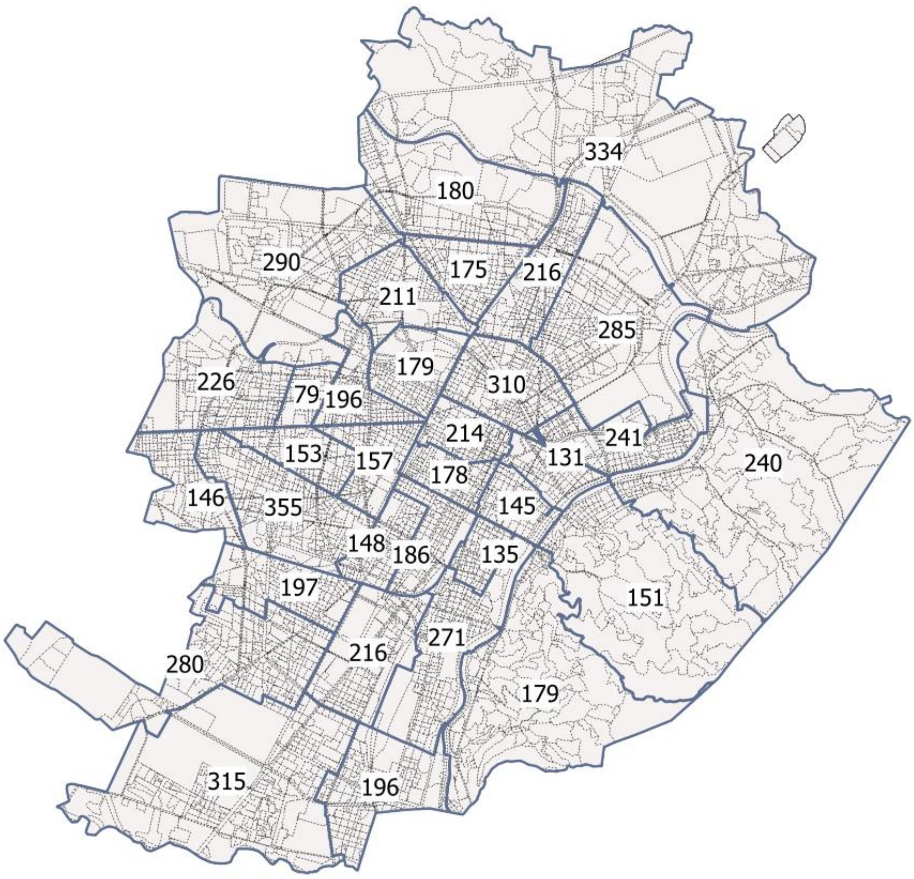


Figure 53 Number of census tracks in each of the 33 postcode zones in the city of Turin.

After having assigned the average gross income by source to the different census tracks, the net income has been calculated by applying the Italian tax rate [232]. The two sources of data (the aspect of daily life dataset and the income revenue authority one) present a mismatch in the way the source of incomes is clustered, with the necessity to establish a strategy to match the income sources as presented in the income revenue authority dataset, with those classified in the census survey one. Such logic is presented in Table 29.

Table 29 Matching logic between the AVQ dataset and the tax declaration dataset to assign the income source. The first three source of income are the same in the two datasets, while the last four necessitated a remapping.

AVQ dataset	Tax declaration dataset
Employee work	Employee work
Self-employed	Self-employed
Pension	Pension
Allowance	0
Revenue from assets	Average income
Family support	0
N.A.	0

As specified in Table 29, for the first three source of income a direct correspondence could be established between the two datasets, while for the revenues from assets in the AVQ dataset it has been preferred to assign the average value for income from the tax declaration dataset due to a lack of information (i.e., related to magnitude of this type of income over the total income of an individual). Furthermore, it was not possible to establish a clear value for allowance and family support. Nonetheless, it might be argued that individuals relying on this source of income are unlikely to be able to allocate part of the income to energy related investments (in particular, out of the 23,654 individuals for which allowance was the source of income, 11,159 belong to the category of employment “inactive”, and 12,396 to category “in search of occupation”, those two sources representing a minority of the population). For this reason, a value of zero has been assigned to these individuals. Finally, the individuals for which no data was available (N.A.) have been assigned a value of zero as well. In particular, it is worth noting that the majority of individuals for which no data was available are also those in age classes below 13 years old, thus the assignment of no income has been considered as rational. Once the income at the individual scale has been assigned, it has been then aggregated at the household scale by performing a simple sum among the individuals belonging to the same household.
Finally, it is possible to compare the distribution of respondents to the questionnaire into the characteristics value functions with the distribution of types of individuals as generated with the synthetic population procedure (from Figure 54 to Figure 57). In particular, it is possible to notice that certain frequently occurring profiles do not see a proper representation in the number of observations collected with the

questionnaire, while other less frequently occurring ones according to the marginal distribution across the city are quite overrepresented in the observations present in the respondents' dataset. This does not constitute a reduction of the applicability of the methodology, but even if the necessity to easily collect data has justified the use of a convenience sampling technique in this first experimentation, it would be advisable to use a more reliable sampling technique in case of a real application of the method to achieve more accurate results and provide more informed recommendations to Decision-makers.

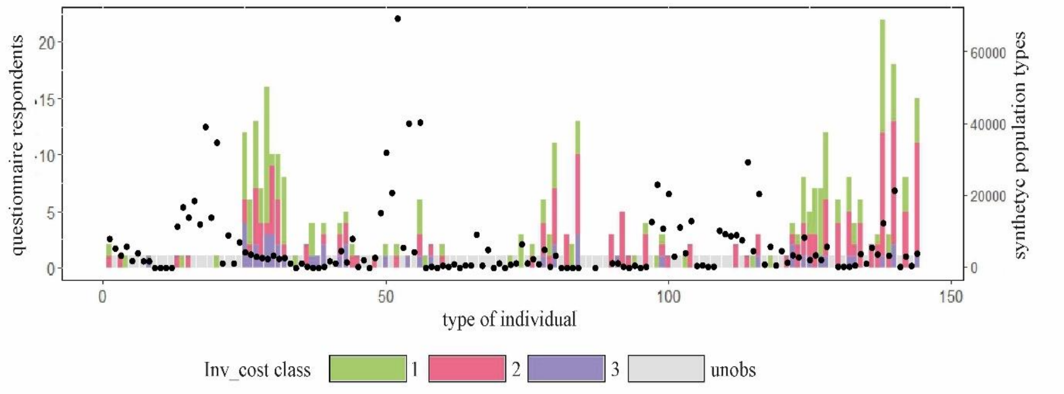


Figure 54 Comparison between the occurrence of the observed types of individuals (349 respondents) subdivided by clusters of investment cost value function and the occurrency of the types of individuals in the synthetic population

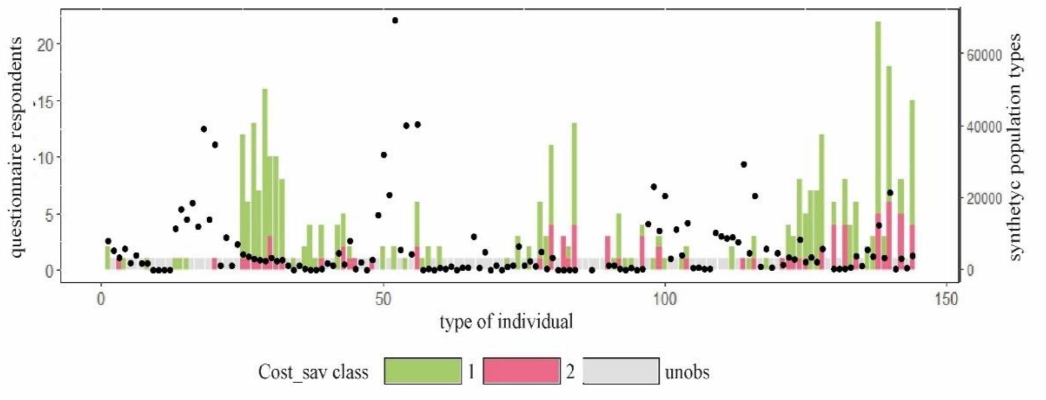


Figure 55 Comparison between the occurrence of the observed types of individuals (349 respondents) subdivided by clusters of cost savings value function and the occurrency of the types of individuals in the synthetic population

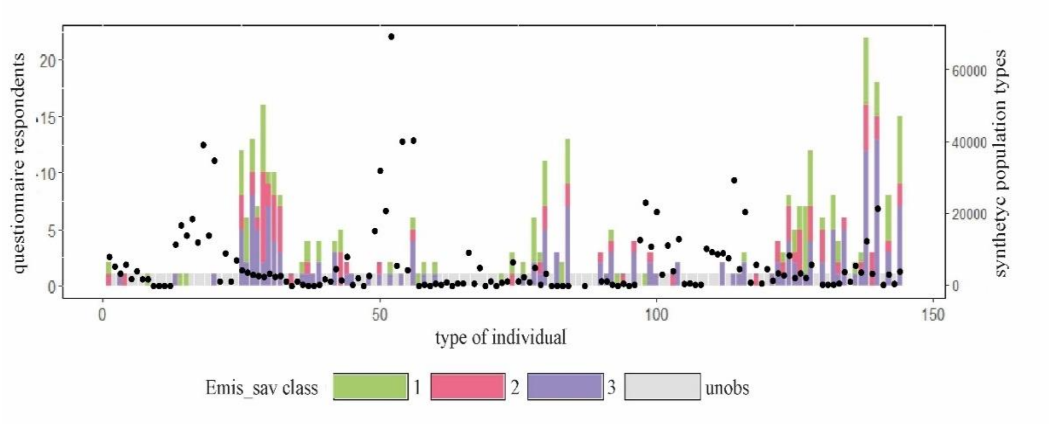


Figure 56 Comparison between the occurrence of the observed types of individuals (349 respondents) subdivided by clusters of emission savings value function and the occurrency of the types of individuals in the synthetic population

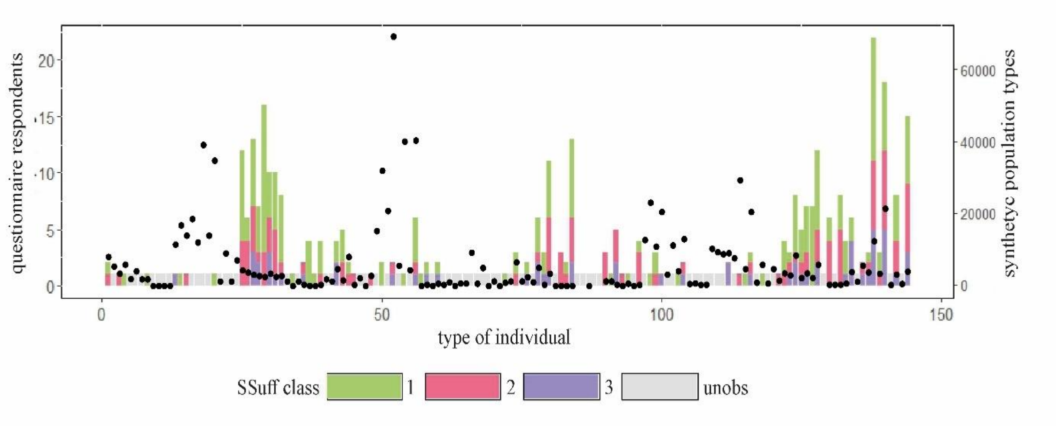


Figure 57 Comparison between the occurrence of the observed types of individuals (349 respondents) subdivided by clusters of self-sufficiency value function and the occurrency of the types of individuals in the synthetic population

Chapter 5

Spatialization of the results

5.1 Application of the decision model to RECs alternatives

In this chapter, the two models defined in the previous chapter (Section 4.3.2 and 4.4.2) are spatialized over the synthetic population distributed across the city and combined in order to evaluate both the propensity of the population to accept or not different alternative RECs, and to determine which of the alternatives might be the most preferred ones among the simulated population.
As stated before, the Italian legislation regulating the constitution of RECs requires that both the energy production systems from renewable sources and the point of delivery (POD) to be located in the same HV/MV substation for each configuration of a REC. For this reason, the results of the spatialization of the probability to accept and the preferences among the population are at first presented at the scale of the HV/MV substation.
In Figure 58 the areas of Turin served by each HV/MV substation in the city are displayed.

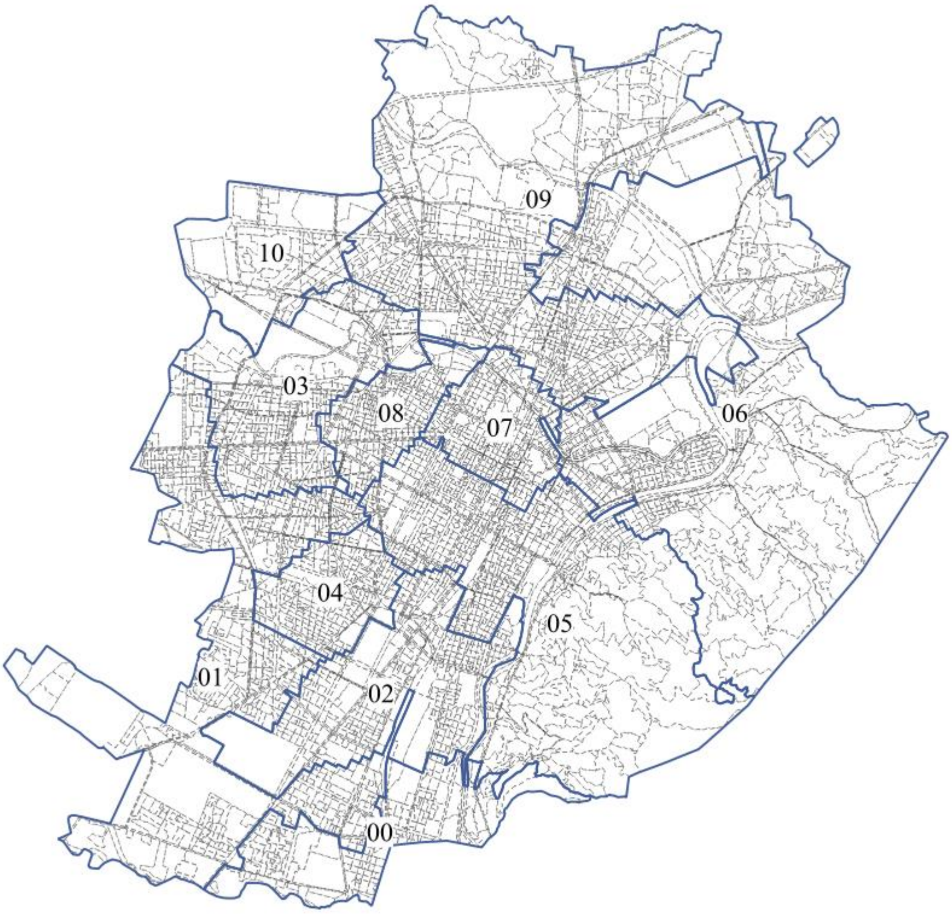
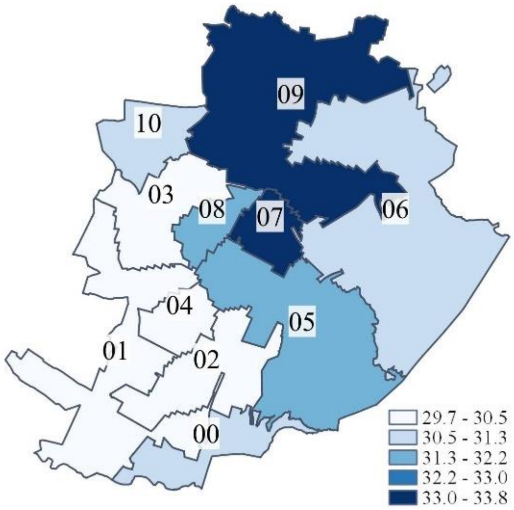
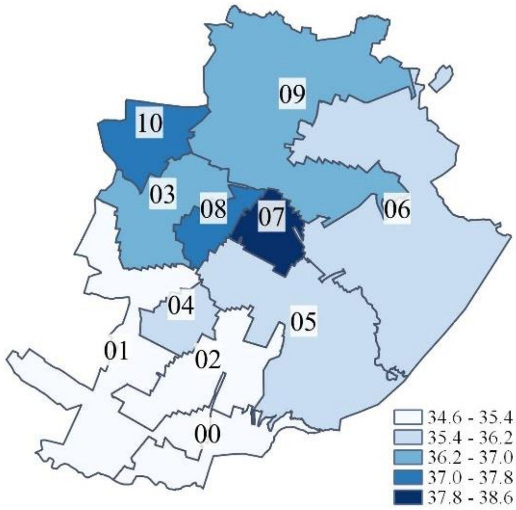


Figure 58 Map of the 11 HV/MV sub-stations in the city of Turin (elaboration from GSE data).

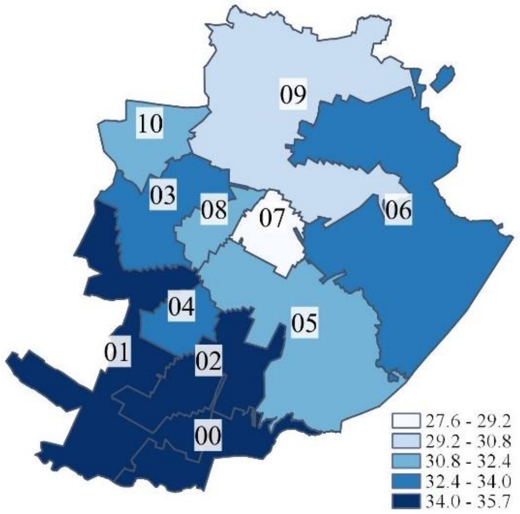
The same aggregation scale has been used to map the percentage distribution of the socio-demographic characteristics of the synthetic population generated in Section 4.5. In Figure 59 the percentage of population belonging to one class or the other of the socio-demographic attributes used to generate the 144 types of individuals are presented (for dichotomic variables only one of the two levels is displayed while the other can be easily calculated as the remaining percentage).



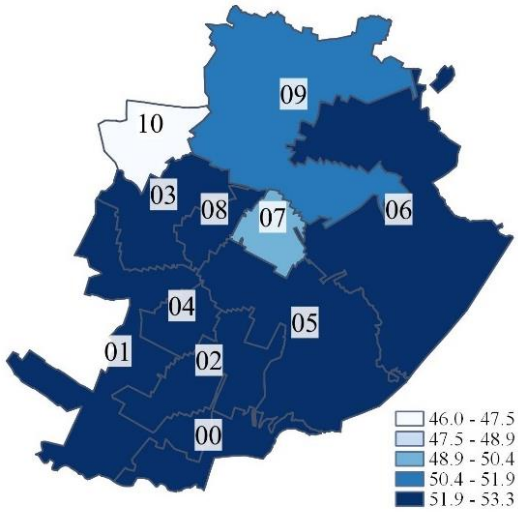
a. Percentage of population in age class [<35yo]



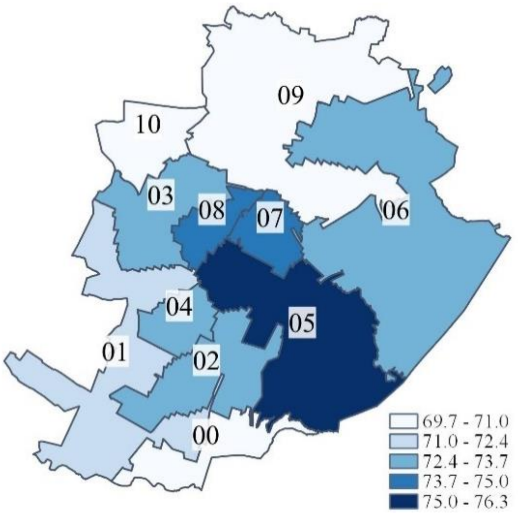
b. Percentage of population in age class [35-60yo]



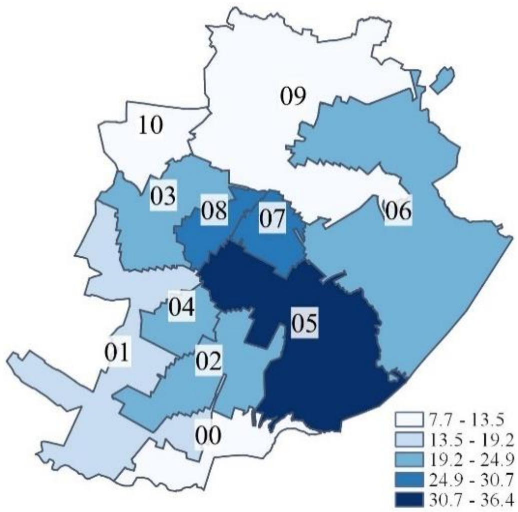
c. Percentage of population in age class [>60yo]



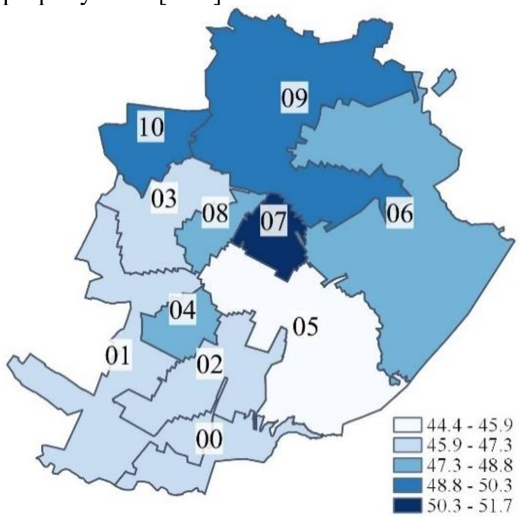
d. Percentage of population in gender class [Male]



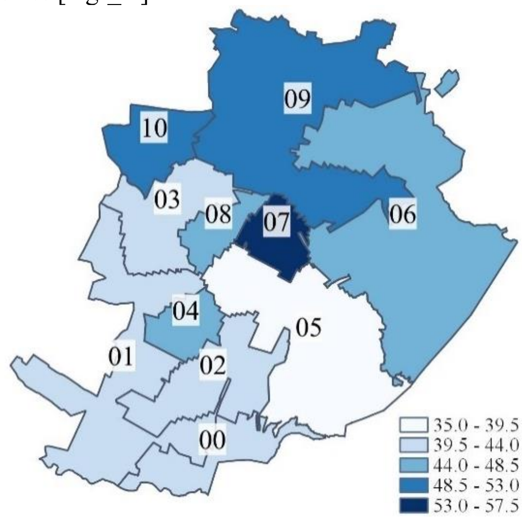
e. Percentage of population in Household property class [own]



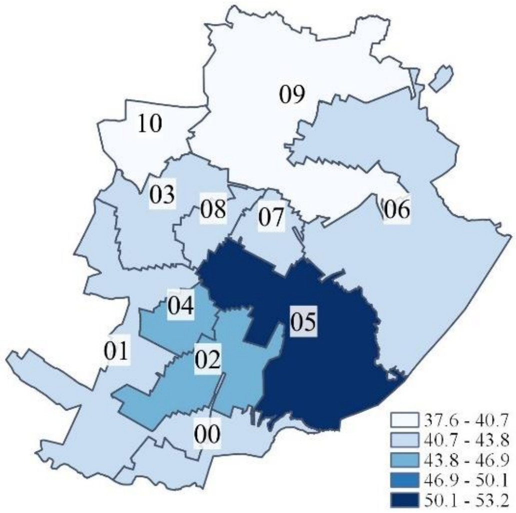
f. Percentage of population in Education level class [Dgr_Y]



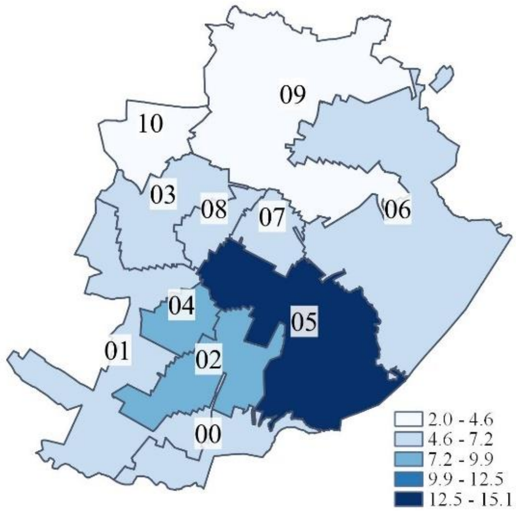
g. Percentage of population in Household size class [3+]



h. Percentage of population in Household income class [<28k€]



i. Percentage of population in Household property class [28-75k€]



j. Percentage of population in Household property class [+75k€]

Figure 59 Distribution of the synthetic population in the socio-demographic classes

By analyzing the distribution of population over the classes defining the types of individuals it is possible to notice that most of these classes are quite equally distributed. For instance, (d.) gender, (e.) property of the household, (g.) size of the household, and (a., b., c.) age variable vary in a percentage below 10% across the different substations. The major difference in the distribution of socio-demographic characteristics across substations is registered for the two socio-demographic variables of education level of the respondent and income level of the household. Concerning the first one, substations 09 and 10 are those covering the areas of the city characterized by a majority of inhabitants with an education level lower than the degree. On the contrary, the highest number of inhabitants with at least a degree is located in the central areas served by substation 07 and 08, and those served by substation 05. The first two coincide with the areas with a high incidence of young strata of the population [233], the last one covers the areas characterized by households with the highest level of income and the highest housing prices in the city [234]. Households in the lowest class of income are more frequent in the substations serving the central areas and the northern one (substations 07 and 09 respectively). These are the areas in which the presence of young and foreigner population is more significative [233].

It must be highlighted here that the coarse resolution of the aggregation at the HV/MV substations scale has an averaging effect on the variation of the distribution of socio-demographic characteristics across different census tracks at a finer scale resolution [233].

5.1.1 Estimation of percentage of accepting population

The percentage of population that might accept to participate in a REC was calculated based on REC performance attributes and the socio-demographic characteristics of the generated synthetic population, according to the results of the logistic model.

Figure 72 displays the percentage of the population likely to accept one or the other of the alternative RECs proposed in the questionnaire at the scale of the HV/MV substation.

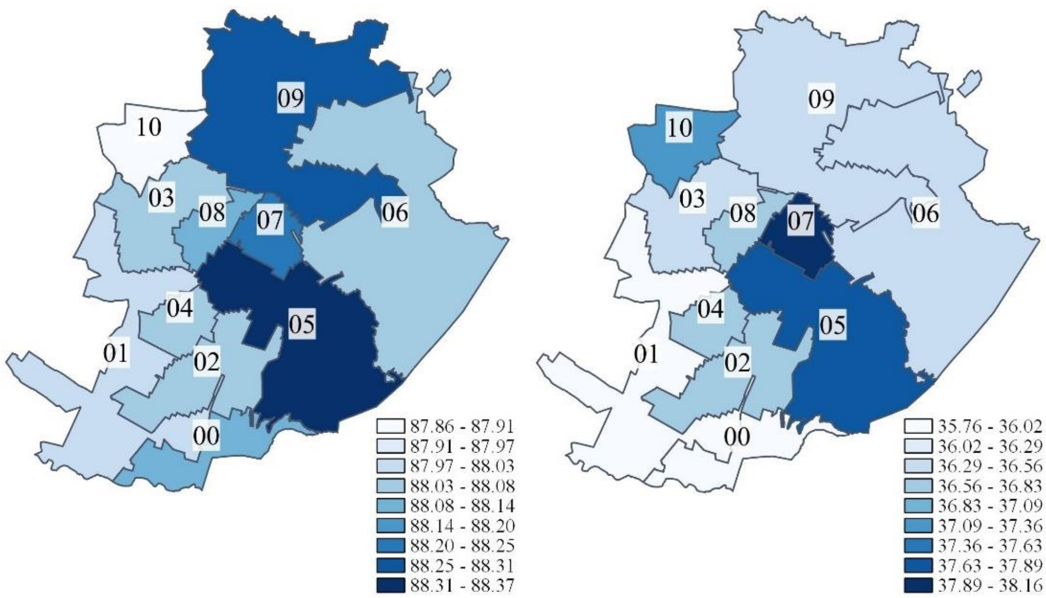
By recalling the equation estimating the probability to participate (Section 4.3.2), it is possible to analyze the spatialized results from three points of view. In particular, the same equation could be interpreted as the composition of three main parts:

$$Pr(y=1) = logit^{-1}\left(\left(\sum_{p=1}^{n} \beta_p * x_p\right) + \left(\sum_{s=1}^{m} \beta_s * x_s\right) + \left(\sum_{i=1}^{q} \beta_i * x_i\right)\right)$$

Where x_p denotes the performance of the alternative RECs, and β_p the relative regression coefficients, x_s and β_s the socio-demographic attributes and the relative coefficients, and x_i and β_i the same for the moderating effect that the socio-demographic attributes have on the association between performance attributes and the odds to participate.

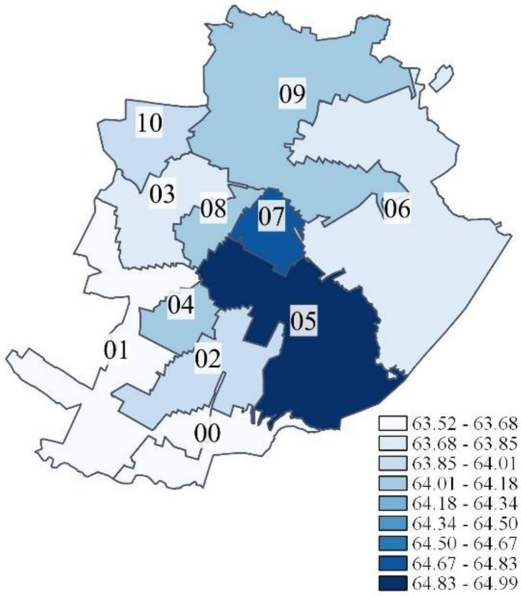
It is possible to affirm that the first part of the equation (relative to x_p) estimates the odds of one alternative REC to be accepted based on its performance, thus independently from the area in which the acceptability is estimated. Therefore, in Figure 60 it would be possible to compare the percentage of population likely to accept one alternative or another in the same area (the same HV/MV sub-station). The second part (relative to x_s) associates the acceptance of the alternative with the socio-demographic characteristics of the population, therefore, in each figure it is possible to evaluate the areas in which the same alternative has a higher or lower percentage of population likely to accept it. Finally, the third part (relative to x_i) is responsible for variation in patterns of the percentage of population likely to be willing to accept a specific alternative across different areas. In particular, the variation of percentage of accepting population for different alternatives in different areas is expected to be proportional (if alternative X_1 sees twice as much percentage

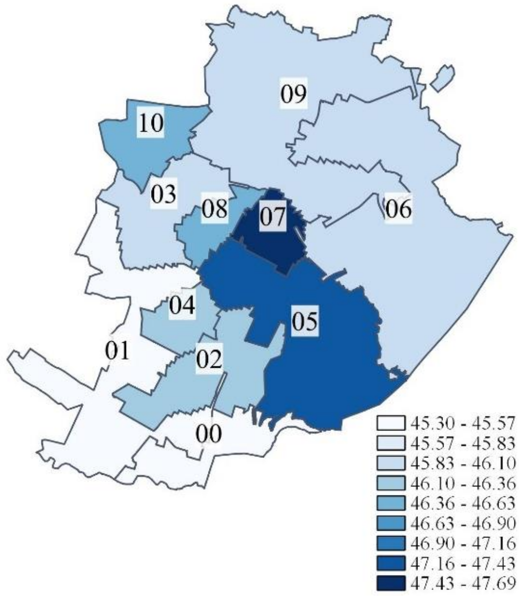
of population likely to accept it in substation S_1 compared to substation S_2, alternative X_2 would see the same proportion of population likely to accept it in the same two areas). When this proportionality is not preserved, the effect could be explained by the moderating effect of the socio-demographic characteristics on the performance attributes.
To better clarify this, let us consider the first two of the following figures at random (Figure 60.a and Figure 60.b), and to focus on the two HV/MV substations in the northern part of Turin (substations 09 and 10). In both areas alterative B (Figure 60.b) has a lower percentage of population likely to accept to participate (36-37%), compared to alternative A (Figure 60.a showing a percentage of population likely to accept of 88-89%). Looking at Figure 60.a, it is possible to notice that in the first substation (09) the percentage of population likely to accept alternative A is 88.3%, slightly higher than the one estimated in the adjacent substation (10) where the percentage of population likely to accept the same alternative is 87.9%. If no interaction is to be considered we would expect to observe a similar pattern also in Figure 60.b, being the distribution of population unchanged in the same area, and the only variation being the performance of RECs A and B respectively. On the contrary, we observe a variation in the pattern from Figure 60.a to Figure 60.b. Indeed, Figure 60.b shows that in substation 09 the percentage of population likely to accept the alternative is 36.4%, while in substation 10 the percentage is 37.2%. This variation could be explained by the moderating effect of the socio-demographic characteristics of the population on the performance attributes of the alternative.



a. Percentage of population likely to accept alternative A

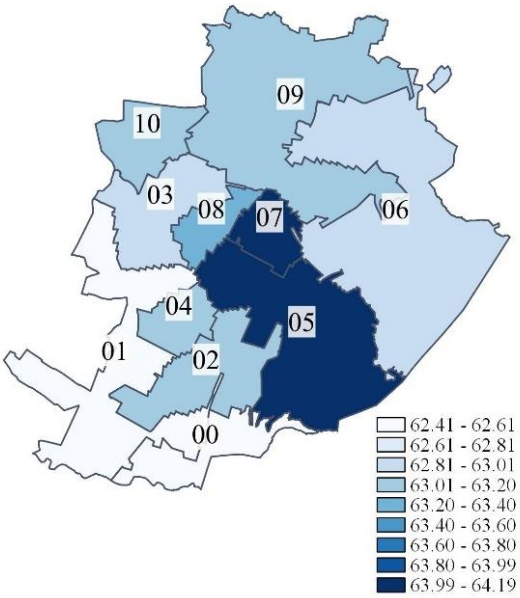
b. Percentage of population likely to accept alternative B

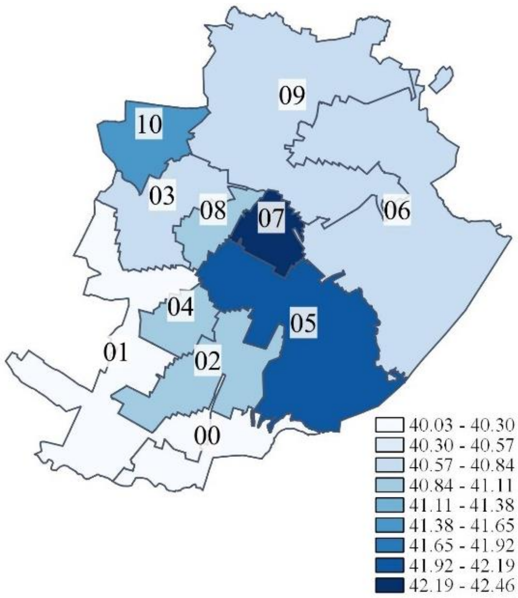




c. Percentage of population likely to accept alternative C

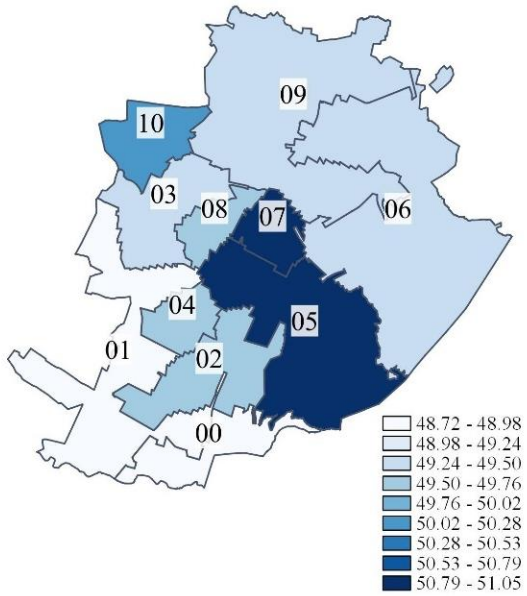
d. Percentage of population likely to accept alternative D



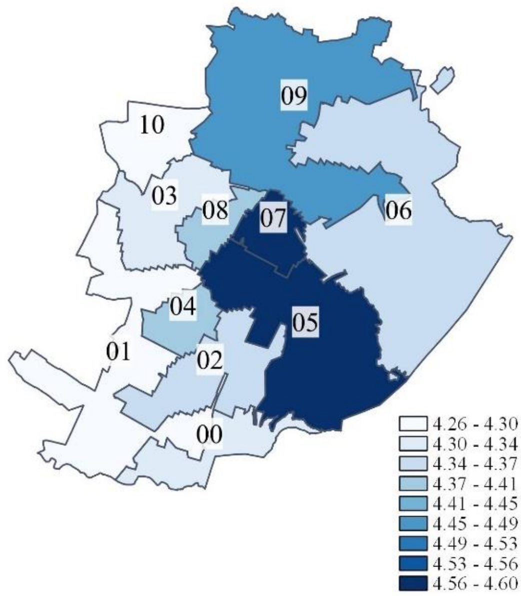


e. Percentage of population likely to accept alternative E

f. Percentage of population likely to accept alternative F



g. Percentage of population likely to accept alternative G



h. Percentage of population likely to accept alternative H

Figure 60 Estimated percentage of simulated population likely to accept each of the 8 alternatives presented in the questionnaire

Table 30 summarizes the percentages of population likely to accept the different alternatives in each of the HV/MV substation, according to the spatialized results of the logistic regression model, together with the estimated Confidence intervals. In particular, due to the large number of individuals it is unfeasible to calculate the exact probability distribution function of acceptability. To solve this issue a Monte Carlo method could be used to repeatedly simulate the acceptance behavior of the entire population to estimate the distribution of acceptance percentages together with Confidence intervals.

Table 30 Percentage of population likely to accept the different REC alternatives in each HV/MV substation according to the spatialized results of the GLMM. Confidence intervals are displayed in square brackets.

HV/MV sub-stations	Percentage of accepting population in each alternative [%]							
	A	B	C	D	E	F	G	H
00	88.10 [87.73, 88.46]	35.90 [35.36, 36.45]	63.62 [63.07, 64.17]	45.50 [44.95, 46.07]	62.57 [62.03, 63.12]	40.15 [39.60, 40.71]	48.83 [48.26, 49.39]	4.33 [4.10, 4.56]
01	88.02 [87.83, 88.20]	35.75 [35.48, 36.03]	63.52 [63.24, 63.79]	45.30 [45.02, 45.58]	62.41 [62.14, 62.69]	40.03 [39.76, 40.31]	48.72 [48.44, 49.00]	4.26 [4.15, 4.38]
02	88.04 [87.80, 88.27]	36.59 [36.25, 36.94]	63.95 [63.61, 64.30]	46.13 [45.77, 46.49]	63.02 [62.68, 63.37]	40.89 [40.54, 41.24]	49.55 [49.19, 49.91]	4.37 [4.22, 4.52]

03	88.08 [87.84, 88.31]	36.32 [35.97, 36.67]	63.83 [63.47, 64.18]	45.87 [45.51, 46.24]	62.85 [62.49, 63.20]	40.60 [40.24, 40.96]	49.27 [48.90, 49.64]	4.33 [4.18, 4.48]
04	88.05 [87.76, 88.34]	36.75 [36.32, 37.19]	64.02 [63.58, 64.46]	46.30 [45.85, 46.75]	63.14 [62.70, 63.58]	41.03 [40.59, 41.48]	49.68 [49.23, 50.13]	4.41 [4.22, 4.60]
05	88.37 [88.14, 88.59]	37.79 [37.45, 38.13]	64.99 [64.66, 65.33]	47.35 [47.00, 47.70]	64.10 [63.76, 64.44]	42.13 [41.78, 42.48]	50.80 [50.45, 51.16]	4.60 [4.46, 4.75]
06	88.06 [87.82, 88.30]	36.31 [35.95, 36.66]	63.81 [63.46, 64.17]	45.87 [45.51, 46.24]	62.83 [62.48, 63.19]	40.58 [40.22, 40.94]	49.25 [48.88, 49.61]	4.36 [4.21, 4.51]
07	88.24 [87.93, 88.57]	38.16 [37.68, 38.64]	64.83 [64.36, 65.30]	47.69 [47.20, 48.18]	64.19 [63.71, 64.66]	42.46 [41.97, 42.94]	51.05 [50.56, 51.54]	4.59 [4.38, 4.79]
08	88.12 [87.83, 88.40]	36.82 [36.40, 37.24]	64.09 [63.67, 64.51]	46.37 [45.94, 46.81]	63.22 [62.79, 63.64]	41.10 [40.67, 41.53]	49.75 [49.31, 50.19]	4.40 [4.22, 4.58]
09	88.31 [88.14, 88.48]	36.40 [36.15, 36.65]	64.04 [63.79, 64.29]	46.08 [45.82, 46.33]	63.06 [62.81, 63.32]	40.64 [40.38, 40.89]	49.32 [49.06, 49.57]	4.45 [4.35, 4.56]
10	87.86 [87.42, 88.29]	37.20 [36.56, 37.84]	63.93 [63.29, 64.57]	46.54 [45.88, 47.21]	63.17 [62.52, 63.81]	41.52 [40.87, 42.17]	50.09 [49.42, 50.75]	4.26 [3.99, 4.53]
Average	88.14 [88.07, 88.21]	36.59 [36.48, 36.70]	64.03 [63.92, 64.14]	46.16 [46.05, 46.27]	63.08 [62.97, 63.19]	40.87 [40.76, 40.98]	49.53 [49.42, 49.65]	4.40 [4.35, 4.45]

To increase the readability of this part, the performance of the 8 alternatives presented in the questionnaire are reported in Table 31.

Table 31 Performance of the 8 alternatives included in the questionnaire.

Alternative	**Investment Cost [€]**	**Cost Savings [€]**	**Emission Savings [kg_{CO2}]**	**Selfsufficiency [%]**
A	9400	1300	2600	82
B	4650	500	1400	20
C	6600	750	1150	61
D	5300	700	2200	17
E	5000	750	1850	37
F	3400	400	850	27
G	2550	400	650	33
H	20000	1650	3100	0

Focusing on the percentage of population likely to accept the different alternatives, it is possible to notice that alternative A sees an average of 88% of accepting population. This is mainly due to its second to best performance level among all the alternatives in cost savings and emission savings, while being by far the best one in terms of self-sufficiency achieved. Alternatives C and E follows with percentage of population likely to accept them around 64% and 63% respectively. These two alternatives present the same level of cost savings performance, with quite good performance in emission savings and self-sufficiency respectively. This balanced performance seems to determine the higher percentage of population interested in these two alternatives compared to others with lower investment costs but also lower cost savings (e.g., alternative G) and self-sufficiency performance (e.g., alternative D). Moreover, the same lower performance in cost savings and self-sufficiency seems to determine a low acceptability of alternative B and F. Finally, the high investment cost and low self-sufficiency performance of alternative H determines its almost null likelihood to be accepted even if it is the alternative with the best performance in terms of cost-savings and emission savings.

Overall, the difference in the percentage of population interested in each alternative does not vary much across different HV/MV substations, with a variation of less than 1%. This is linked to the recognized limited variability of socio-demographic characteristics averaged across the coarse aggregation on HV/MV sub-stations (see Figure 59). Also in this regards, it is worth recalling that it was not possible to test for the presence of a random effect related to different areas of the city in the analysis conducted in Section 4.3.2. Due to the small size of the respondents' dataset and its uneven distribution across the urban area, it is not possible to exclude the possible presence of such an effect. Nonetheless, a variation in the pattern of acceptability across the different HV/MV substations could be highlighted due to the distribution of socio-demographic characteristics and their moderating effect on RECs performance.

Focusing attention on the moderating effects, in all the substations except substation 07 and substation 10 the interaction effect between gender and cost savings increases the odds to participate for each 1-unit increase in thousands of euros of cost savings ($e^{0.536*Cost\_sav(k€)*Gender[Mal]}$) compared to these two substations. Also considering the effect that the property regime of the household has on the probability of accepting the alternative with higher level of self-sufficiency ($e^{0.104*SSuff(\%)*HH\_prop[own]}$), it is possible to notice that the higher percentage of owners in the southern part of Turin, as well as in the eastern and western parts increases the percentages of accepting population for alternatives A and C. Compared to these areas, the percentages results reduced in HV/MV substations 07, 09, and 10 where the number of owners is the lowest in the city. Finally, a consistent interaction effect is represented by the moderation of household size on cost savings performance achieved by the REC alternative ($e^{0.673*Cost\_sav(k€)*HH\_num[3+]}$). Indeed, this effect is particularly evident in sub-stations 09. In this area the percentage of households of 3 people or more is higher than the other parts of Turin. This peculiarity explains the different pattern of acceptability of those alternatives for which the cost savings performance was among the highest, such as alternative A and H.

As a general trend, the central and south-eastern part of the city (HV/MV substations 05 and 07) show higher percentage of population likely to accept any of the possible alternative RECs compared to other areas. This percentage radially decreases moving toward the outskirts of the city, with an exception represented by alternative A due to the interaction between the alternative performance and the socio-demographic moderators.

The small variation in the percentage of population willing to accept an alternative or not could be explained by the large aggregation made at the scale of the HV/MV substation. A further analysis has been conducted at the single census track level for alternative A (being it the one with the highest percentage of population likely to participate) to observe whether the pattern of acceptability would vary at a finer scale. Two focuses have been made, one in the northern part of the city (HV/MV substation 09 and northern part of substation 06), and one in the central and south-eastern part (HV/MV substation 05 and southern part of substation 06). In Figure 61 and Figure 62 these two focuses are mapped.

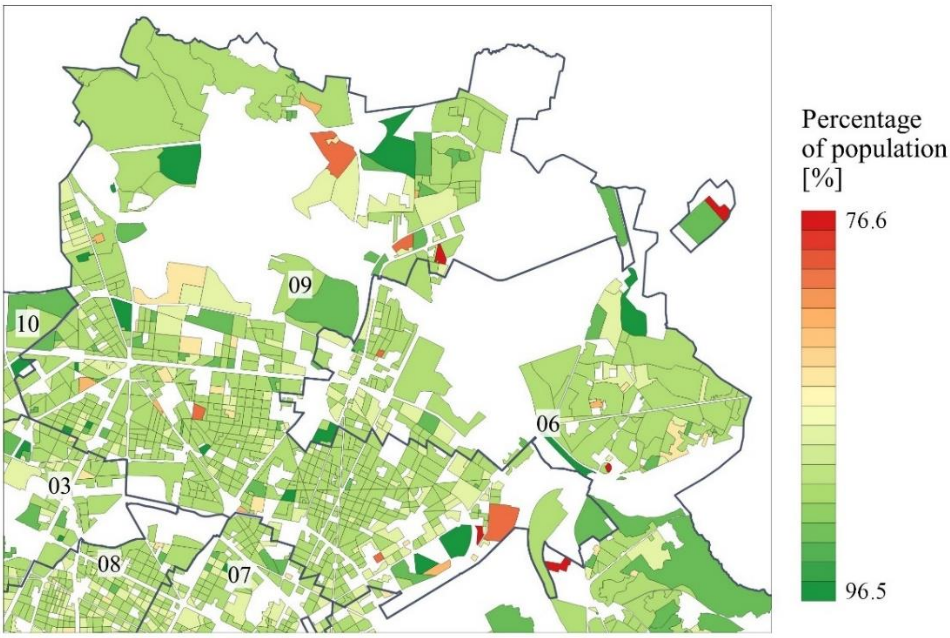


Figure 61 Focus on the percentage of simulated population likely to accept Alternative A at the census track scale in the northern part of Turin.

Figure 62 Focus on the percentage of simulated population likely to accept Alternative A ate the census track scale in the eastern part of Turin.

From the first focus (Figure 61) it is possible to notice a fair increase in the variation of percentage of population likely to accept the alternative across the different census tracks clustered in the same HV/MV substation. These values range between 76.6% and 96.5%, with some areas in the northern part presenting values close to the lowest among those estimated at the city level (76.6%), compensated by higher percentages of participation especially in some of the census tracks closer to the central part of the city. In the second focus (Figure 62) it is possible to notice a more homogeneous distribution of the percentage of population likely to accept alternative A, with the majority of areas presenting average values. Few areas present medium-to-low percentage of population likely to accept, these coinciding with large areas with few resident population (two census tracks showing values close to the lowest percentages as well). This analysis highlights a difference in the potential distribution of REC participation within the same HV/MV substation masked by the coarse data resolution at this scale. This is particularly evident in the northern areas of the city (Figure 61) where higher socio-demographic imbalances were highlighted.

5.1.2 Combination of the two models to estimate the acceptability of public oriented alternatives

The final level of analysis combines the two approaches used to (i) evaluate the percentage of population likely to accept to participate in a REC, and (ii) to assess the most probably preferred alternative across the population of the city. The goal is to identify the trade-offs between the performance of a potentially widely accepted alternative and those of a "public oriented" one. Here "public oriented" alternatives could simply be defined as the ones prioritizing the reduction of GHG emissions from the use of energy, but other public goals such as the different distribution of financial benefits across the population (such as the Energy help indicator) could also be envisioned

Second aim is to calculate the potential support from third parties (e.g., ESCOs or public authorities) to ensure the preference of the private bottom-up initiative for "public oriented" alternatives, thus prioritizing those solutions maximizing the positive outcomes for the society as a whole. In particular, considering the results of the two models (the likelihood to participate or not and the preference ranking of alternatives), this subsection evaluated the level of acceptability of new "public oriented" alternatives characterized by an high public desirability outcome [121], and the probability of those to become also the most preferred ones among individuals (i.e., ranking first among possible competing alternatives).

With this purpose, the focus has been put on alternative H as the one with the highest performance in terms of CO_2 savings among the RECs in the dataset, thus the one that could be considered the most preferred one from a public point of view [121]. Therefore, it is worth investigating the performance changes in the alternative to increase its desirability (i.e., increasing the percentage of population likely to accept it), as well as to increase its chance of becoming the preferred alternative over the others. First of all, it is worth recalling that alternative H implemented only passive retrofit solutions (i.e., no renewable energy production measure was included) determining a self-sufficiency performance of zero. Therefore, a first new alternative (public_H) has been generated by increasing the self-sufficiency performance of the original alternative H to the average of the alternatives (excluding the initial value of alternative H). This to assume the inclusion of renewable energy production among the energy efficiency measures implemented in the alternative, without changing its investment cost (it is therefore hypothesized that at the same cost fewer passive energy efficiency measures were implemented, while some energy production measures were proposed). To evaluate the increase in the alternative desirability, the initial investment cost has been

progressively reduced by steps of 5%. This cost reduction was introduced to account for potential public intervention to stimulate the constitution of alternatives (public_H) maximizing the community-oriented outcomes (i.e., the reduction of GHG emissions).
A further new alternative (esco_H) has been proposed by further increasing the self-sufficiency performance while reducing the cost savings achieved by the original alternative H. This again tests the extent to which the intervention of an external investor (such as an ESCO) with a different business model (e.g., financing the initial costs while using part of the cost saving to cover the investment) could increase the acceptability of the proposed alternative. The initial performance (with investment cost = 20,000€ as in the original alternative H) of the two newly generated alternatives (public_H and esco_H) are reported in Table 32, together with the performance of alternative A (reported as the one with the highest percentage of accepting population).

Table 32 Performance of Alternative A, and the two newly proposed alternative public_H and esco_H maximizing the reduction of GHG emissions. The newly generated alternatives (public_H and esco_H) have the same investment cost of Alternative H (20,000 €), while for both alternatives self-sufficiency performance was increased. Alternative esco_H presents a reduced cost-saving performance too.

Alternative	**Investment Cost [€]**	**Cost Savings [€]**	**Emission Savings [kg_{CO2}]**	**Selfsufficiency [%]**
A	9400	1300	2600	82
public_H	20000	1650	3100	40
Esco_H	20000	1300	3100	60

In Figure 63 the percentage of the population likely to accept the participation in the two newly generated alternatives reducing the investment cost of the alternatives by steps of 5% is presented.

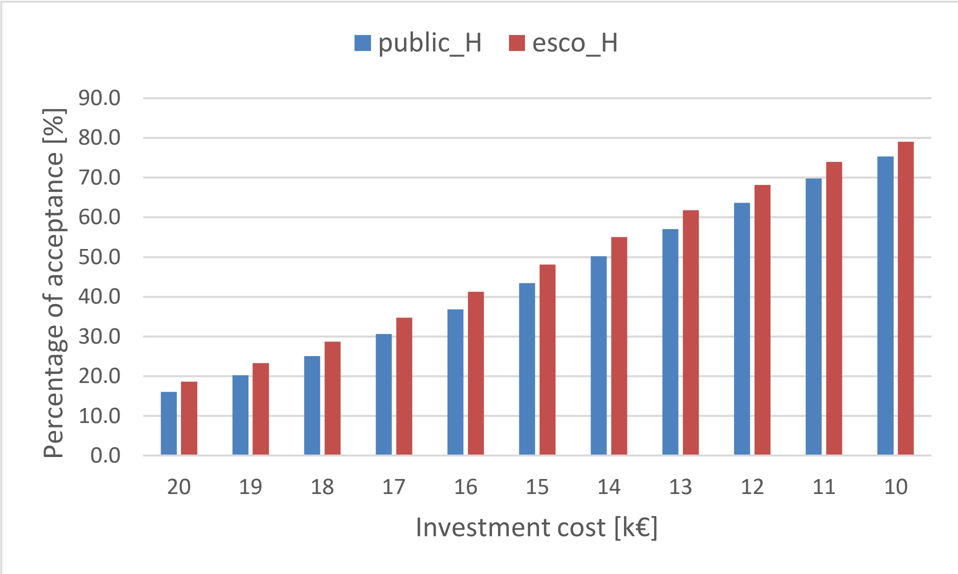
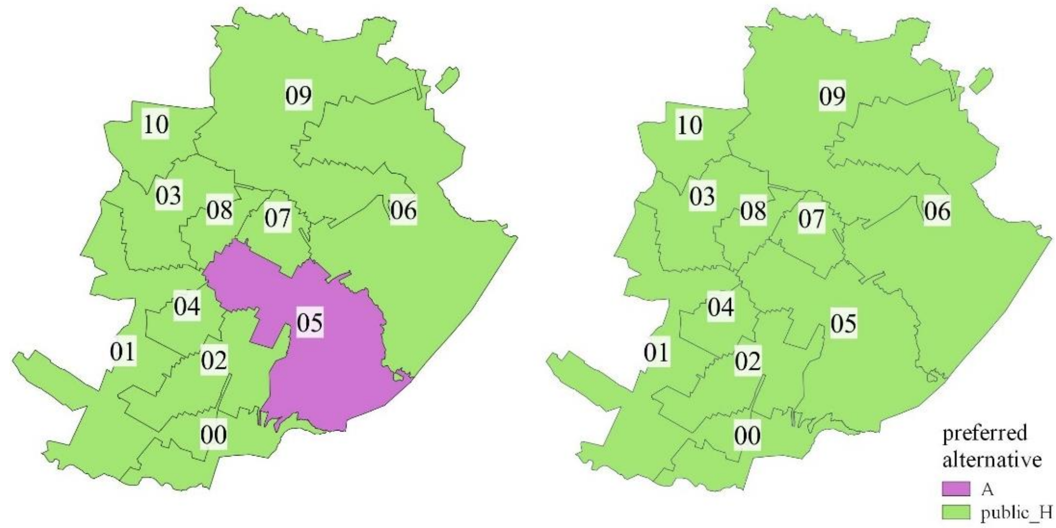


Figure 63 Percentage of population likely to accept alternative public_H and esco_H at each 5% reduction of investment cost.

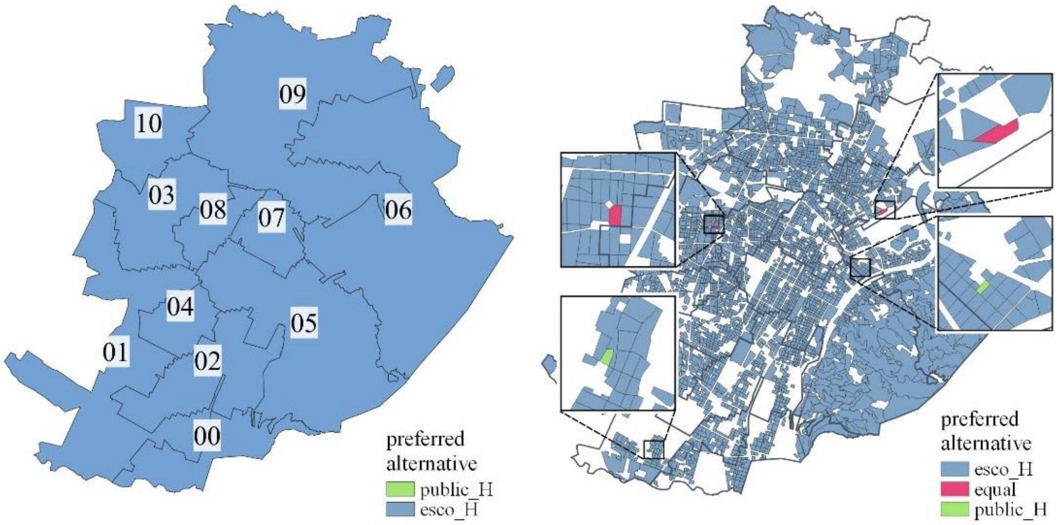
Observing Figure 63, it results clear the impact that a higher self-sufficiency performance has on the acceptability of the alternatives. For instance, the acceptability of both alternative public_H and esco_H with the same investment cost of 20,000€ as the original alternative H increases on average from 4.4% to 16.1% and 18.6% respectively. As expected, by reducing the initial investment cost of the alternatives, and thus assuming an increasing involvement of potential external financial support, the percentage of population likely to participate increases. What is interesting to notice here is how the reduction of cost savings (even though by a small amount) from alternative public_H to esco_H does not affect negatively the participation, with potential for external investors to propose different potential business models to the population (such as the one described previously). On the other hand, the emphasis on self-sufficiency is confirmed, with a higher level of performance in this attribute determining a higher percentage of people likely to accept solution esco_H compared with alternatives public_H with the same investment cost.

In terms of spatialization of the preferences, it is also possible to establish which alternative might be preferred in each HV/MV substation by analyzing the preferability of the two new alternatives compared to alternative A (being it the one with the highest percentage of population likely to accept) and with each other. This will give an idea of the minimal intervention of the public body (or of external investor in case of alternative esco_H) necessary to ensure that the publicly preferred alternative coincides with the private individuals' one. As specified in the description of the method, this is done by calculating the total utility of the alternatives and calculating which ranks first for the majority of the population. Figure 64 shows the most preferred alternatives in each HV/MV sub-station for different reductions of initial investment costs.



a. Preferred alternative between option A and public_H (investment cost=14k€)

b. Preferred alternative between option A and public_H (investment cost=13k€)



c. Preferred alternative between option public_H (investment cost=14k€) and esco_H (investment cost=14k€)

d. Preferred alternative between option public_H (investment cost=14k€) and esco_H (investment cost=14k€) at the census track scale

Figure 64 Estimated preference among different versions of public oriented alternatives and alternative A

In particular, Figure 64.a compares alternative A with alternative public_H characterized with an investment cost of 14k€, while Figure 64.b compares alternative A with alternative public_H characterized by an investment cost of 13k€. Figure 64.c compares the latter (alternative public_H with 13k€ investment cost) with alternative esco_H characterized by 14k€ investment cost. Finally, Figure 64.d compares the same alternatives as in Figure 64.c but displays the preferences at the census track scale.

By observing Figure 64.a, it is possible to conclude that comparing alternative A with alternative public_H requiring an investment of 14k€ (thus with a public intervention of 6k€), this latter would be preferred in the majority of the HV/MV substations with the exception of substation 05. Figure 64.b shows that in case of alternative public_H with investment cost equal to 13k€ (thus increasing the public intervention to 7k€) all the substations would prefer this alternative over alternative A on average. Moreover, comparing alternative public_H requiring 13k€ investment cost with esco_H with investment cost equal to 14k€ (Figure 64.c.), it results that the latter is preferred over alternative public_H in all the substations. This confirms the importance given to self-sufficiency performance by the population, determining the preference for a more self-sufficient alternative even if this performs worse in the other financial aspects (investment cost and cost savings performance).

As shown in Figure 64.d, the analysis confirms the preference for alternative esco_H over alternative public_H in all the census tracks but four. Two of them still prefer alternative public_H (in green color), while the two alternatives are equally preferred in the remaining two census tracks (in magenta in the figure).

In light of the above, it would be possible to calculate the necessary support to ensure that the public point of view is taken from the majority of the individual, and the alternative with the highest level of CO_2 reduction potential is preferred over others. In the specific case, the support to ensure that alternative public_H would be the preferred one could be determined as the difference between the original investment cost for alternative H (20k€) and the alternative public_H requiring 13k€ to be implemented. Therefore, with a needed financial support of 7000€ per apartment and a total of 433,469 apartments in the city, the total expenditure for the third-part actor would amount to 3,034,283,000€.

On the other hand, in case of alternative esco_H with a required investment cost of 14k€ this support is reduced to 6000€/apartment, therefore resulting in a total financial support by a third party of 2,600,814,000€ (6000€/apt*433,469apartments in the city). These values are theoretical, and consider only one alternative to be implemented (either public_H or esco_H). Indeed, several different configurations could be proposed, especially considering site-specific REC scenarios. Moreover, further analysis could be performed to estimate the preferred configurations (in

terms of performance levels) based on types of individual giving more importance to one or the other performance parameter with the possibility to optimize and tailor the alternatives on their preferences (for instance combining the financing strategies of the analyzed public and ESCO subsidized alternatives).
It is worth noting that alternative H was the most expensive among all the others with comparable performance with respect to alternative A. For this reason, this value must be taken as a proof of concept rather than a case specific value, with the necessity to calculate the specific potential performance of different alternatives of interventions in the different areas of the city. It might also be argued that the value functions as parametrized in this experimental analysis are not so fitted in describing the preferences of the population for expensive alternatives (in case of the value functions related to the investment cost attribute, most of the times these give a null marginal utility for values greater than 14k€) due to their coarse parametrization with only three segments. This aspect will be further discussed in the next section, considering possible refinement of the proposed method.

5.1.3 Discussion of limitations and improvements

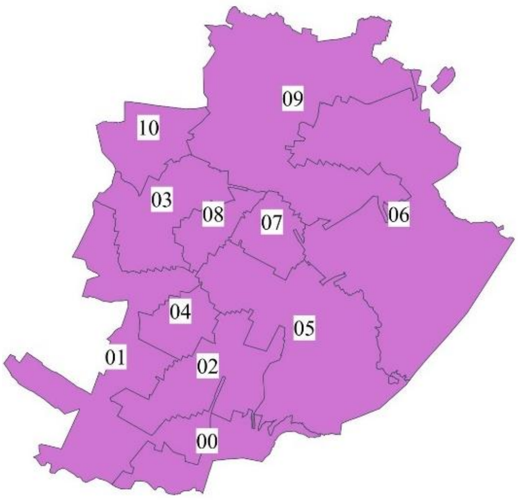
From the analysis described in the previous subsection, we can see that the two models provide different but complementary information. In particular, the first model identifies the percentage of the population likely to accept a particular alternative at different scales (HV/MV substation or census tracks). Although it can be argued that an alternative characterized by a high percentage of acceptability may also be the one preferred by the majority of people, in the case of two alternatives with similar acceptance rates, it is not possible to say which of the two will be preferred (and thus potentially realized). The second model estimating the people's preferences complements the previous one by providing information on which, among the alternatives potentially accepted, is also the one that is likely to rank first in the majority of individuals' preference ranking.
In this way, the two models can be integrated to answer the question on what are the sufficient conditions (high percentage of participation) and necessary conditions (position in the ranking of preferences) for the individual initiative to be in line with the public purpose of prioritizing solutions that maximize the beneficial impacts for the community (in the specific case, those that guarantee a high level of greenhouse gas reduction). It is worth noting that different objectives can be translated into the same performance attributes used to define RECs. In fact, the reduction in energy cost savings seen for the esco_H alternative can also be seen as a way of freeing up economic resources that can be used for community aiding purposes (such as the case of the Energy Poverty Help Indicator, EPHI [47]). The integration of the two

approaches may therefore prove to be a way of channeling individual private action towards the goals of a just environmental transition.

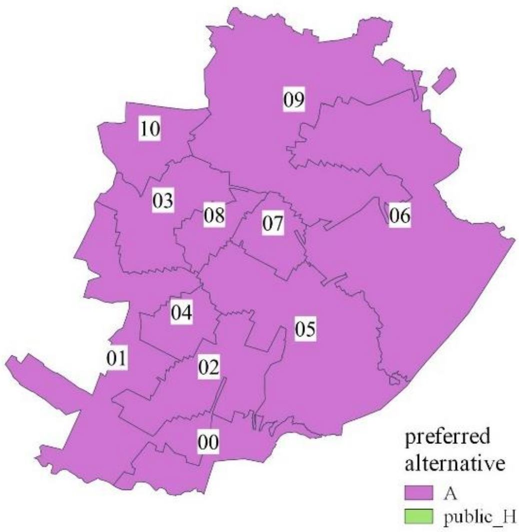
Although the application is promising, it is necessary to recognize the limitations found in the results obtained. First of all, on average alternative A and public_H with an investment cost of 14k€ are estimated to have a percentage of accepting population of 50% and 88% respectively according to the GLMM model. Considering the second model (preference ranking), alternative public_H is preferred in almost all the substations compared to alternative A. Even if it is accepted that both alternatives A and public_H are present in the individual preference set each time, it is hardly credible that alternative public_H is the one preferred on average in all the substations of the city. This discrepancy in results could be imputed to the choice of parameterizing the value functions with 3 segments. In particular, this coarse parametrization limits the ability of the value functions to discriminate between competing alternatives with high initial investment costs. In fact, it can be seen that all three value functions defined as characteristics for the investment cost attribute estimate a null marginal utility for the investment cost performance of the alternative falling in the third segment of the value function. This means that the ranking of alternatives for expensive solutions is estimated based on the remaining performance attributes due to (estimated) indifference for investment costs above 14k€. It is useful to note that this result cannot be considered reliable or unreliable a priori, given the lack of information on this subject resulting from the answers obtained from the questionnaire proposed to the respondents (in particular due to a limited number of alternative in this range of performance inserted in the questionnaire).

This potential discrepancy between the results obtained by the two models can be further analyzed by recognizing that the first one, based on logistic regression, suffers from less severe limitations due to the initial input data. On the contrary, the second model might present a potential issue in the choice of the initial parameters for the construction of the value functions (in particular their piece-wise definition in three segments). In order to test for a possible increase in the reliability of the results, the value functions were re-estimated using a five-segment parametrization with the inclusion of additional dummies alternatives (dominated by alternatives A and H), with investment costs in the 9.4k€-20k€ range (the investment cost of alternative A and H respectively).

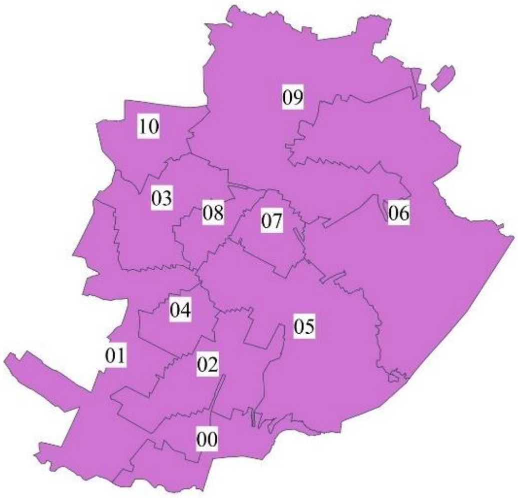
The maps in Figure 65 show the comparison between alternative A and public_H, and between alternative A and esco_H applying the decision model with the reparametrized value functions as 5-segment piecewise curves. This in order to evaluate if this refined construction of the value functions would increase the interpretability of the complementary results from the two models constructed (the one estimating the percentage of population likely to accept a solution, and the one estimating the preference among competing alternatives). Differently from the comparison shown in the previous subsection, in Figure 65 both alternative (public_H and esco_H) are compared to alternative A due to the fact that none of the simulated versions of public_H (even the one with the maximum reduction of investment cost) are preferred over alternative A.



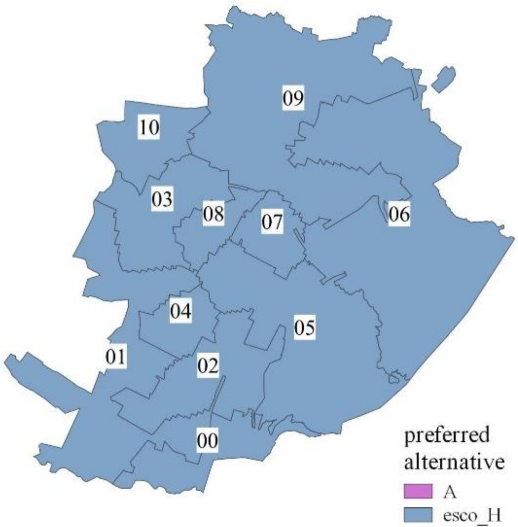
a. Preferred alternative between option A and public_H (investment cost=13k€)



b. Preferred alternative between option A and public_H (investment cost=10k€)



c. Preferred alternative between option A and esco_H (investment cost=12k€)



d. Preferred alternative between option A and esco_H (investment cost=11k€)

Figure 65 Estimated preference among different versions of public oriented alternatives and alternative A

In Figure 65.a alternative public_H with an investment cost of 13k€ (the one resulting preferred across all the substation from the previous parametrization of value functions in 3 segments) is not preferred over alternative A. Moreover, Figure 65.b shows that not even the versions of alternative public_H with the lowest simulated investment cost (10k€) is able to surpass alternative A in the preference ranking of the population. On the contrary, alternative esco_H is preferred over alternative A when its investment cost is reduced to 11k€ (Figure 65.d).
These final results do not seem to contradict the previous ones achieved with a coarser parametrization of the value functions, but confirm the pattern highlighted previously. In particular, the preference for higher level of self-sufficiency holds true in the last results (as confirmed by the impossibility of alternative public_H to become the preferred one when compared to alternative A). Moreover, alternative esco_H characterized by the same level of cost savings, comparable self-sufficiency performance, and maximizing the reduction of GHG emissions become the preferred alternative even when its investment cost is reduced only to 11k€ (1600€ more than alternative A). This highlights the presence of different levers apart from financial one in preferring a REC alternative over another. Finally, the total amount of financial support from third-party to guarantee that alternative esco_H becomes the preferred one is 3,901221,000€ (9000€/apartments*433,469 apartments).
To complete the analysis, in Figure 66 the comparison between alternative A and esco_H with investment cost 12k€ is displayed at the census track scale, while Figure 67 shows the comparison between alternative A and esco_H with investment cost of 11k€ at the same granularity level.

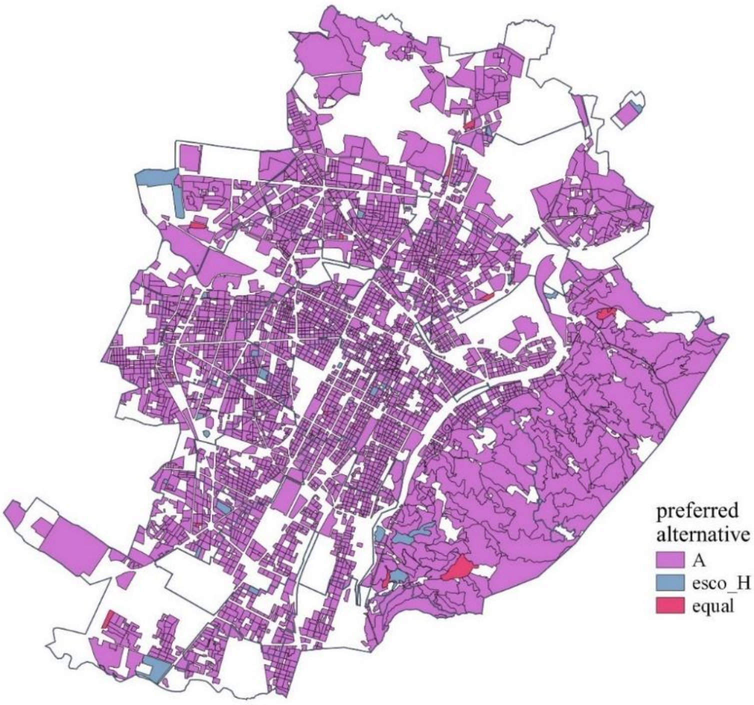


Figure 66 Estimated preference between alternative A and esco_H with an investment cost performance of 12k€ at the census tracks scale.

Figure 67 Estimated preference between alternative A and esco_H with an investment cost performance of 11k€ at the census tracks scale.

Clearly, the majority of preferences towards alternative A is confirmed in Figure 66 with the results presented at the HV/MV substation aggregation scale, as it is for preferences toward alternative esco_H in Figure 67. Again, it is possible to notice a small variation within different census tracks toward one or the other alternative This last analysis highlights how the variation in the pattern of preference across different options confirms the analysis made in the previous subsection. It could be sustained that the discrepancy between the results from the two models (one estimating the percentage and the other the preferences of the population) have been reduced by a finer definition of the value functions. In particular, it is worth remembering that the last two alternatives compared (alternative A and alternative esco_H with 11k€ investment cost) showed similar percentages of population estimated to accept them (on average 89% and 74% respectively). These considerations confirm the possibility to complement the information gathered by the application of one model with those resulting from the other.

5.1.4 Possible improvements of the workflow

Examining the method, it is possible to identify potential areas in which this could be improved. This will enhance the quality of results provided to Decision-Makers when considering possible interventions on the policy framework in place to support the alignment of private individual actions with the public utility. In the previous subsection the issue related to the parametrization of value functions has been discussed and their refined formulation has been tested. Two main areas of improvement could be identified (i) in the construction and distribution of the questionnaires, and (ii) in the way the collected data have been analyzed. Regarding the first area, an improvement could be identified in the size of the dataset and in the population sampling technique used. From the descriptive analysis conducted in Section 4.1 on the collected data, and the synthetic population generated in Section 4.5, it results clear that the responses retrieved from the survey cannot be considered as explanatory of the actual population present in the city. Moreover, by plotting the distribution of the respondents clustered in types of individuals against the distribution of the same types across the synthetic population generated it was clear how several types of individuals were overrepresented in the sampled population, while others were underrepresented in comparison with their distribution at the city scale (Figure 54 to Figure 57). Furthermore, the alternatives used to construct the choice and ranking experiment were selected from a metanalysis of case studies from academic production over RECs in the Italian context. Even though these alternatives could be considered suitable, a better selection might have been done among alternatives constructed using specific data from Turin context (in terms of performance achievable), with the possibility to better define the range and distribution of performance among the alternatives (even if the adopted approach has been deemed as convenient in terms of time consumed and replicability of the study). In particular, the selection of the alternatives through the metanalysis had a direct effect on the estimation of the sets of value function with the UTAstar method. In particular, the uneven distribution of the attribute performance level of the alternative across the overall range of performance determined a lack of discriminatory information in the modeling, and thus the presence of flat parts in the estimated sets of value functions. Again, the use of site-specific data could reduce this issue, reducing also the performance range to that achievable by real case alternatives specific of the area under investigation (being the inclusion of unachievable performance levels useless for the scope of the application). Furthermore, it is necessary to keep in mind the final goal of the application, and in particular the determination of the tradeoffs to ensure that the most desirable RECs from a public point of view (i.e., those prioritizing public benefits) would be favored over private oriented ones (i.e., those maximizing private benefits). In this sense, the focus on those two competitive clusters of scenarios would also determine the best set of alternatives to be included in a real case application of the presented method.

Regarding the filling in of the questionnaire, the approach used did not require any further interaction with the respondents after the first administration of the questionnaire. This resulted in a faster collection of data at the expense of possible further information collected from the respondents. In terms of the cognitive effort required by the respondents, the number of alternatives among which they had to make a choice was limited, thus lessening such effort. An improvement in the collection of learning sets from the questionnaires could be seen in the possibility for the respondent to specify more relation structures between alternatives (i.e., weak preference, indifference [235]). Nonetheless, limiting the possible answer to only the strict preference relation was deemed convenient to reduce the cognitive effort required from the respondent, as well as the risk of misunderstanding in the absence of interaction with the respondent.

In the area of improvements of the analysis of data, the first improvement could be seen in the estimation of value functions for each individual. Regarding that, the first argument that could be raised relates to the identification of one specific set of value functions among all the possible sets coherent with the learning set provided by the respondents. In particular, for small learning sets (i.e., when the respondent ranked few alternatives, thus expressing only one or few pairwise comparisons on which to train the UTA algorithm used) several sets of possible value functions could have been identified, representing a limitation in the presented workflow. While acknowledging that, it could be argued that the presented application represents a proof of concept on the possibility to generalize the individual values to a population by pivoting on socio-demographic characteristics. For this reason, the main focus of the application was not to identify the best value function describing the individual behavior, but to estimate common values among group of individuals. That justified the retention of the learning sets provided by respondents that performed the ranking among few selected alternatives. A possible relaxation of this limitation could be seen in the extension of the method including procedures to estimate a number of sets of value functions for each individual starting from the provided learning sets and to identify similarities among the value functions provided by different individuals in order to select those shared by several individuals to be further generalized at the city level.

Another limitation related to the modeling of the value function has been previously discussed regarding the flat parts due to the lack of information in the learning sets. Even though this issue could be partially reduced by the inclusion of more even distributed alternatives across the performance range, a further improvement could be, for instance, to include curvature bounds in the estimation. Even if this method could have reduced the presence of plateaus in the value functions, the lack of information on which basing this refinement justified to avoid its inclusion.

A further improvement is related to the clustering of the individual value functions performed with the unsupervised algorithm (partitioning around medoids). The definition of the characteristic value function might be performed by minimizing the error calculated with the error matrix described in Section 4.4.1 in order both to better shape the characteristic value function and to determine the optimum number of clusters selected. It is worth noting that the limited number of clusters specified in the application of the method was also necessary due to the limited size of the dataset. Indeed, specifying a higher number of clusters would have been incompatible with the aim to generalize the decision model to the city level due to a higher dispersion of data from the limited number of respondents in the dataset.

The low performance achieved by the supervised clustering algorithm (PART) is directly linked to the previous limitation and the small size of the dataset of respondents. As showed, the unbalanced distribution of classes across the type of respondents has deeply affected the capability of the algorithm to predict the classes to be assigned to one or the other type of individuals. In Appendix H a speculation has been done inflating the dataset to check if better performance could be achieved.

Finally, due to the unbalanced distribution of the collected questionnaire across the city it was not possible to test for potential variation in both the acceptability and preferences toward RECs specifically related to the residency location of the respondents (i.e., the introduction of neighborhood and HV/MV as random effects in the GLMM model). This latter represents an interesting further development of this study to check for influence of the geographic location of the respondents as an intra-clusters effect.

Chapter 6

Conclusions

6.1 Outcomes and audience

The European goal of achieving climate neutrality by 2050 requires profound structural changes across sectors and scales. Urban areas are key arenas for climate action due to their high environmental impact (especially due to their high emission levels) on a global scale. This led to several initiatives aiming to demonstrate the possibility to achieve climate neutrality at the city scale, such as the European Mission 100 Climate Neutral City. Furthermore, in the urban context, reducing carbon emissions from the building stock has been identified as a priority, primarily because of the building sector's impact on the overall emissions budget of cities, but also because of the mitigation potential that the sector offers in the short term.
The role of civil society has been increasingly recognized as an integral part of the energy transition. Bottom-up initiatives and participatory models offer a promising way to foster citizen engagement, improve energy literacy and democratize access to clean energy. These developments signal a paradigmatic shift in the governance of urban energy systems - one that moves away from a technocratic model towards a more inclusive, citizen-centered approach.
Academic and policy discourse has emphasized the need to scale up interventions beyond the individual building level to community and neighborhood approaches, enabling synergies across building typologies and consumption patterns.
Despite the potential that the building sector could have in driving the urban energy transition, both from a technical perspective by reducing its greenhouse gas emissions, and from a civil society perspective by changing the role that citizens can play in the energy market, the transition to more sustainable energy use at the building scale is still slow.
The relatively recent introduction of the Renewable energy communities legal framework could specifically target this slowness by encouraging the creation of bottom-up initiatives to promote more sustainable energy use and the diffusion of technologies capable of producing renewable energy locally. The benefits that the Renewable energy community should bring to its members and to the areas in which

they operate cover several dimensions, such as economic, environmental and social. If considered as a specific realization among the possible options that individuals have at their disposal (borrowing from the formalization made by the urban capability approach), it is then important to investigate individuals' prioritization of outcomes in the conversion process from opportunities to their realization. This can eventually highlight the mismatch between private bottom-up initiatives and the public utility in the implementation of such legal entities. In particular, different actors might pursue more private objectives when deciding if and which alternative participation scheme they would prefer when comparing different options, thus missing the opportunity to unlock the maximization of positive impacts for the community as a whole (such as not opting for initiatives that would maximize the level of GHG emissions reductions, even if this alternative might not be the most convenient one in terms of possible financial outcome for the promoter).
In light of the above and considering the four research questions posed at the beginning of the study, the present work has proposed a combined method to estimate (i) the percentage of the population that might participate in a Renewable energy community at the local level, and (ii) the preferential order of possible solutions among the offered ones.
The first question investigated the *"socio-economic characteristics of individuals and the performance attributes of the potential intervention scenarios that influence citizens' decision to accept to participate in a Renewable Energy Community"*. In this perspective, a logistic and a linear regression models have been used to extrapolate the association between socio-demographic characteristics of individuals and both the probability to accept a REC option and the preference related to its performance attributes. This has highlighted that five socio-demographic characteristics (age, gender, and size, income level and property of the household) are associated with the odds to participate in a REC, while three of them (gender, size and property of the household) also act as moderators in the association between RECs performance and the odds to accept it. Furthermore, the individuals' value functions have been estimated by applying a UTA method, and the association between the socio-demographic characteristics of individuals with the marginal utility perceived from performance attributes of the RECs have been investigated. Six influential characteristics have been individuated (age, gender, education level, and size, income and property of the household). These characteristics have been combined into 144 types of individuals to which sets of characteristic value functions have been assigned.
Directly related to the first research question, the second one investigated "*to which extent could different renewable energy communities proliferate at the city scale,*

and which are the most probable solutions to be implemented based on the inhabitants' preference models". To this end, the distribution of the 144 types of individuals at the city scale has been simulated by means of a population synthesizer. Then, the two previously defined models (logistic regression and the preference model based on value functions) have been spatialized pivoting on the distribution of the types of individuals. Focusing on the application of the logistic regression model, the percentage of population that might accept the participation into the alternative configurations of renewable energy community have been estimated matching the results from this model with the socio-demographic characteristics of the types of individuals.

In this manner, the expected percentage of accepting population for the alternatives proposed in the questionnaire has been calculated at the city scale. This has highlighted the interest of individuals for alternatives characterized by a high level of independency from the grid (such as alternative A with a self-sufficiency performance of 82% estimated to be accepted by 88% of the population on average), while a lower penetration capacity of alternatives maximizing GHG emission reductions and cost savings due to a sharp increase in their investment cost (alternative H with percentage of accepting population of 4.4% on average).

The pote*ntial variation in the participation to different configurations of Renewable energy communities among the inhabitants in different parts of the city* has been posed by the third research question. This has been examined by spatializing the results first at the High-Voltage/Medium-Voltage substation, and then at the different census track. In particular, an overall similarity in terms of percentages of the population likely to accept different RECs alternatives was estimated across the different substations of the city of Turin (with a variation of few percentage points across them). At the finer scale of the census track, this variation increases fairly, with a variation of around 20% in the percentage of population accepting a REC solution from areas with less interest to areas characterized by a larger amount of population likely to accept (such as in the northern part of the city). A further analysis introducing the geographical location of the individual as a random factor in the logistic regression model has been tried without showing meaningful results. It is worth stressing here that it was not possible to exclude the existence of such an effect due to the limited spatial distribution in the data collected during the survey.

Finally, the fourth question specifically addressed the possibility of the public authority and third parties to *"foster the alignment of the private initiative with the public perspective".* This has been targeted in the last part of the research where the two estimated models (logistic regression and value function sets) have been combined. In particular, two new RECs (public_H and esco_H) have been

simulated based on the alternative of the questionnaire with the maximum reduction of GHG emissions, and applying some variation to the other two beneficial outputs (self-sufficiency and cost savings). These alternatives have been considered as those more aligned with the public goal of maximizing the positive environmental impact of the implementation of a possible REC solution. For these alternatives, both the percentage of population likely to accept them, and the degree to which they would be preferred over other alternatives have been investigated by applying the two models (the logistic one for the percentage estimation, and the preference model for the definition of the ranking among different alternatives).
The final aim of this analysis was to calculate the external support needed to ensure that the preferred alternative among residents (i.e. the one most likely to be adopted by the private initiative) coincides with the one consistent with the public perspective of a more sustainable use of energy at the building scale.
This has led to the calculation of a total amount of financial support from the public sector (or from energy service providers) of 3,901,221,000€ to ensure that the alternative that maximizes the reduction of GHG emissions would become the most preferred one compared to others across the city. Furthermore, the trade-offs in the acceptability levels of performance among different alternatives have been also evaluated, with the possibility to propose different business models to compensate across different levels of performance of alternatives.
It must be stressed that these simulations must take into consideration several limitations and possible refinements in the application of the method (as discussed in Section 4.6.3), while alternatives based on case specific energy simulations will lead to different results across different parts of the city. Nonetheless, these results could assist the Decision-makers in tailoring and fine-tuning the enabling framework to promote the acceptance of renewable energy communities and participation in such schemes. Considering the application of the proposed methodology to the question of the diffusion of Renewable energy communities, both public authorities and potential independent investors could benefit from the information provided by this simulation. In particular, the estimation of the space of acceptable solutions at the urban scale could be informative for the construction of alternatives considered satisfactory by the majority of the population, and also for the consideration of performance trade-offs in the allocation of resources and benefits.
Moreover, it is possible to consider the method as a predictive tool with the aim to forecast decision outcomes in hypothetical settings (i.e., different scenarios of implementation of a REC), supporting the Decision-maker in estimating the possible realization of one or the other of the possible scenario at the disposal of the

individuals. Similar to this, the proposed method could also represent a tool to estimate the elasticity of the demand. As demonstrated in Section 5.1.2 for instance, the method has been used to analyze the variation of the probability of the individuals to participate at specific configuration of a REC varying the price for their implementation, thus estimating an elasticity of the demand for such services. In this sense, a possible extension of the study could consist in further analyzing the acceptability of the different alternatives by the different types of individuals whose preference system has been estimated. This analysis would be an extension of the test performed in comparing the acceptability of the two alternatives public_H and esco_H, evaluating the different propensity of groups of individuals towards specific performance levels of the same REC proposed with different business models. For example, by analyzing the trade-offs between investment costs and energy savings for different groups of individuals (defined through their preference models), it would be possible to propose customized participatory models in order to optimize the economic resources involved while maximizing the probability of participation.

It should be stressed that the present application represents a proof of concept for the applicability of the proposed method. Even if the method could be replicated in other urban areas, the results of this specific application could not be directly generalized. In particular, in the preliminary analysis conducted to determine if all the respondents could have been included in the analysis or not (Section 4.3.1), a substantial difference among the preferences of the respondents living in Turin and those living outside the city has been highlighted. For this reason, it may be concluded that the results cannot be generalized. Nevertheless, extending the scope of the research including respondents from different areas (thus enlarging the sample of respondents) some preference patterns at higher scales could be estimated.

Two further considerations are necessary regarding the reliability of the results. The primary concern pertains to the utilization of stated preferences as the foundation for the study's empirical analyses. This particular aspect renders it difficult to eliminate the possibility of a discrepancy between the choice behavior that respondents claim to exhibit and the actual behavior they would demonstrate when confronted with the tangible prospect of allocating their resources. The integration of the proposed method with an analysis of revealed preferences could facilitate a comparison between the results obtained with the two approaches. However, it is important to note that the current lack of widespread diffusion of REC configurations limits the possibility of applying methods related to the revealed preference family. In any case, at present, the participation in RECs through

bottom-up initiatives by private citizens is still not a widespread phenomenon. Therefore, the main result of the study could be seen in the assessment of the individuals' trade-offs and preferences with regard to the performance of RECs depending on socio-demographic characteristics.
Finally, the analysis performed combined a series of methods based on different assumptions (e.g., linearity of predictors in regression methods, parameters for determining value functions and absence of interaction between criteria in the UTAstar method, choice of socio-demographic characteristics, etc.) with the consequent risk of reducing the reliability of the final results. Once again, the study focused on testing the applicability of a formal model for analyzing individuals' subjective characteristics (preferences) in relation to objective (socio-demographic characteristics) linked to urban transformations (in this specific case, the reduction of energy consumption in the built environment). A future development of the study could be to analyze specifically the individual methods used and the related assumptions, such as the possibility of evaluating multi-additive models in the case of interactions between selection criteria.
For such reasons, it is recommended that the results of this study are interpreted with a degree of caution.

6.2 Replicability and flexibility of the method

The specific application of the method has focused on demonstrating its possible use in estimating the degree of acceptability of Renewable energy communities in the city of Turin. In particular, all the population data used to estimate both the percentage of the population likely to accept one configuration or another, and to determine the distribution of decision models in the urban environment, came from official open-source repositories provided by public authorities. The analysis of the data (from combining different datasets to generate the simulated population in the area, the spatialization of the results, and application of the regression models and the UTA method) was carried out using open-source scripts (either in Python language or in RStudio environment) and software (QGIS). This ensures the possibility to replicate the same method in different contexts, with only the need to carry out the population survey to collect data specific to the environment under study. In addition, the comparison of different results in different contexts could also prove useful in considering similarities and differences between individuals, with the possibility of studying possible intervention alternatives tailored to each specific case.
It could also be emphasized that the same method could be useful in cases where the decision problem is different from the one analyzed in the present research. In

particular, by adapting the survey to case-specific situations in which the decision problem may affect the urban community, and thus the acceptability of the proposed intervention may be in line with (or in conflict with) the trade-offs of the population, it may be possible to study the population's preference for urban transformation interventions. This may help the public Decision-Maker in choosing the option that could be supported and accepted by the vast majority of the population, and also to consider possible corrective actions to accommodate divergent preferences in the commonly agreed solution.

Appendices

A. Case study: “A preference Learning-based approach for the design of green infrastructures”

The application of the Preference Learning method has been already performed as a preparatory study in the determination of citizens’ preference toward the design of a new urban park in the city of Turin[1]. In the present section an extract of the study is presented. Aim of the study is to analyze the relation between socio-economic attributes of users and urban park features in order to support Decision-makers in the identification of different design alternatives for new urban parks. The research uses Preference Learning techniques in order to profile different park users based on their specific socio-economic attributes and to retrieve their preferences toward existing parks. The goal is to provide DMs with a data-driven supporting tool to analyze urban population’ demand in terms of green areas configuration and provided services (such as leisure infrastructures and facilities) in order to better design the layout of urban green areas. The investigation of park users’ profile could also give information about the possibility to establish participatory process tailored on the different categories of users. The proposed methodology extracts preferences from perceived characteristics recognized by interviewees when stating their preference for specific urban park. A case study concerning the introduction of a new urban park in the city of Turin (Italy) is proposed.

A.2 Application

The study focuses on a downgraded area called Basse di Stura (Figure A.1), located in the city of Turin (Italy). Due to the natural potentialities of the site and for its strategical location, local public administration has expressed its interest for the area, and has planned the construction of a new park in it.

[1] The present research has been conducted by Professor Marta Bottero, the candidate, Professor Alexis Tsoukias and Professor Brice Mayag, at the Laboratoire d’Analyse et de Modélisation de Systèmes pour l’Aide à la Décision (LAMSADE) - University Paris Dauphine.

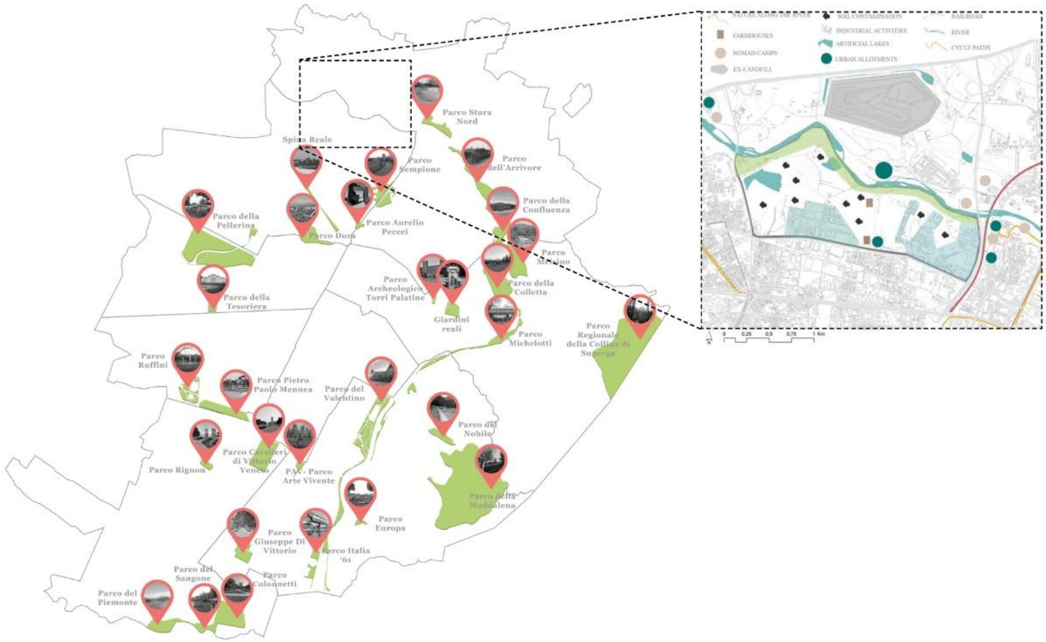


Figure A.1 Existing parks in Turin and Basse di Stura study area (elaborated from Bottero et al., 2020)

Grounding on a previously conducted survey, the present study, analyses city users' preferences toward the possible configuration of the forecasted park. The research focused on retrieve information about respondent's preferred existing urban parks in the city, the perceived characteristics of that parks, and the socio-economic attributes of the interviewee (Table A.1). A total of 408 valid questionnaires have been collected, forming the dataset for the analysis conducted in the present research.

Table A.1 Socio-economic attributes of respondents

Attribute	Type	Range	Number of attribute possible values
Declared gender	nominal	[male; female]	2
Age	cardinal	[18; 87]	
Nationality	nominal	-	
Family with kids	dichotomic	[Yes; No]	2
Single person family	dichotomic	[Yes; No]	2
Elder living with the family	dichotomic	[Yes; No]	2
Study level	ordinal	[elementary; PhD]	
Profession	nominal	-	
Income	ordinal	[<600;>10000]	
Weekly working hours	ordinal	[<20; >40]	
Residency neighbourhood	nominal	[1;23]*	23
Working neighbourhood	nominal	[1;23]*	23

*numeric ID of the city neighborhoods

A.2.1 Process scaffolding

The process to determine the users' preferences toward the new park characteristics is divided in two subsequent steps by means of the construction of different matrixes (Figure A.2). The first one (Matrix of the users' profile) will display each respondent's socio-economic attributes and her preferred existing urban park (this will answer the question "who is the typical user?"); the second matrix (Matrix of parks characteristics) displays the existing characteristics of the preferred urban parks (answering the question "what are the characteristics on which the users base their preferences?"). A third matrix (Global profile matrix) will be constructed by the combination of the previous two. This matrix will contain the characteristics of each preferred park, and the typical profile of the respondents that prefer that park. By analyzing this last matrix, it will be possible to determine the socio-economic characteristics of each park user and to support the DM in defining the layout of the new urban green area according to different users' preferences. In Figure A.2 it is displayed the process and analysis conducted to determine the three matrixes. The analysis on the socio-economic characteristics of the respondents have been performed using several algorithms embedded in the WEKA software.

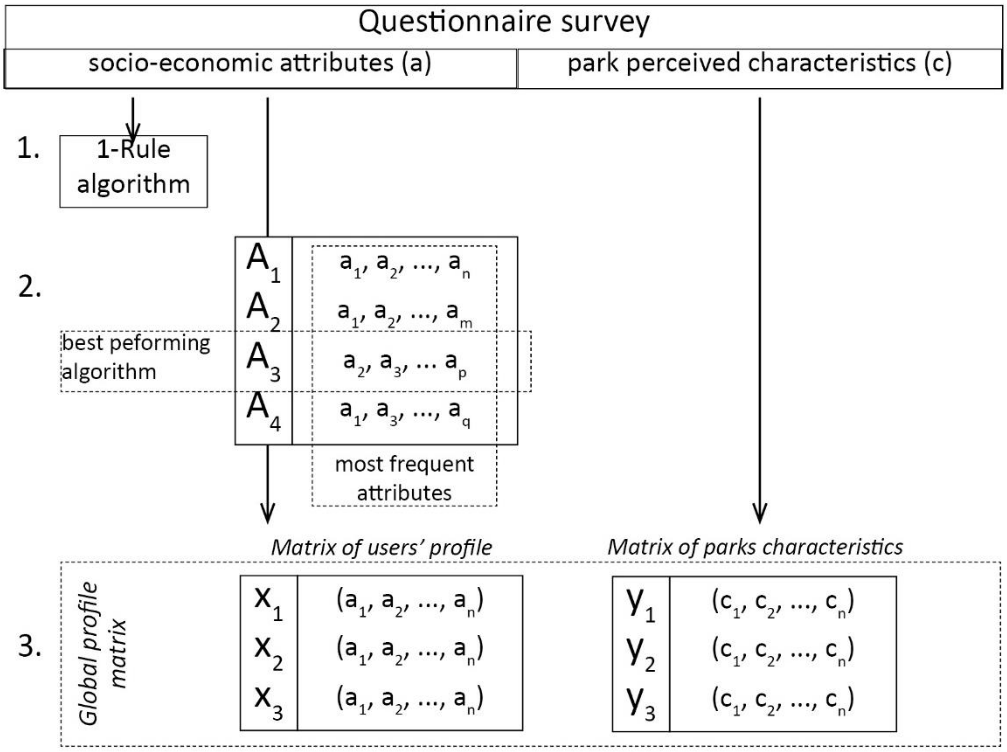


Figure A.2 Research workflow

> 1-Rule algorithm

A preliminary analysis has been conducted applying a simple 1-Rule Algorithm (OneR) that classify instances using a single attribute among the set in order to provide a benchmark for further analysis. Applying the algorithm to the dataset, it chose the one socio-economic attribute that performs best in predicting the respondent preferred park. The analysis has been repeated recursively excluding the best performing attribute at each repetition.

> Algorithms and attributes subset selection

A second round of analysis aimed at investigating different algorithms and sub-sets of socio-economic attributes to predict respondents' preference. The performance of each method has been evaluated considering the percentage of correct prediction and AUC value. The same dataset as been used both for training the algorithm and to validate it using a 10-fold-cross validation splitting method. Different subsets of attributes have been tested by adding or removing them using different wrapping techniques.

> Matrixes construction

The *Matrix of the users' profile* is constructed using the best performing algorithm, together with the subsets of attributes most frequently used in the previous step. The *Matrix of parks characteristics* has been built by calculating the frequency with which each characteristic has been noted as present in each preferred park by means of the following equation:

$$F_{P_i,x_i} = \frac{\sum n_{x_i}}{N_{P_i}} \tag{2}$$

where F_{P_i,x_i} is the frequency with which a specific attribute x_i is perceived as present in the park P_i, $\sum n_{x_i}$is the total number of questionnaires in which the respondent affirm that the attribute x_i is present in park P_i, and N_{P_i} is the total number of questionnaires in which park P_i is the preferred choice.

A.3 Results

In order to apply PL techniques to the dataset, it is necessary to set a minimum threshold of expressed preferences. In fact, using the same dataset for both training and testing the algorithms, the inclusion in the evaluation of less frequently preferred parks would have likely reduced the performance of the models. Further on, considering very specific preferences would have been outside of the scope of providing guidelines for the future park features design. Following these hypothesis, a minimum threshold of 30 preferences has been set, reducing the number of valid questionnaires to 328 ones, and the number of considered parks to 5, namely 1) Valentino, 2) Pellerina, 3) Dora park, 4) Ruffini, and 5) Giardini Reali.

A.3.1 Gravitation Effect

The first analysis highlighted that the OneR model based on the attribute "Neighbor of residency" was the best performing one (57.68% precision and 0.70 AUC). The resulting preference are mapped in Figure A.3, dots mark the barycenter of neighbourhoods, and arrows the preferred park of respondents living there.

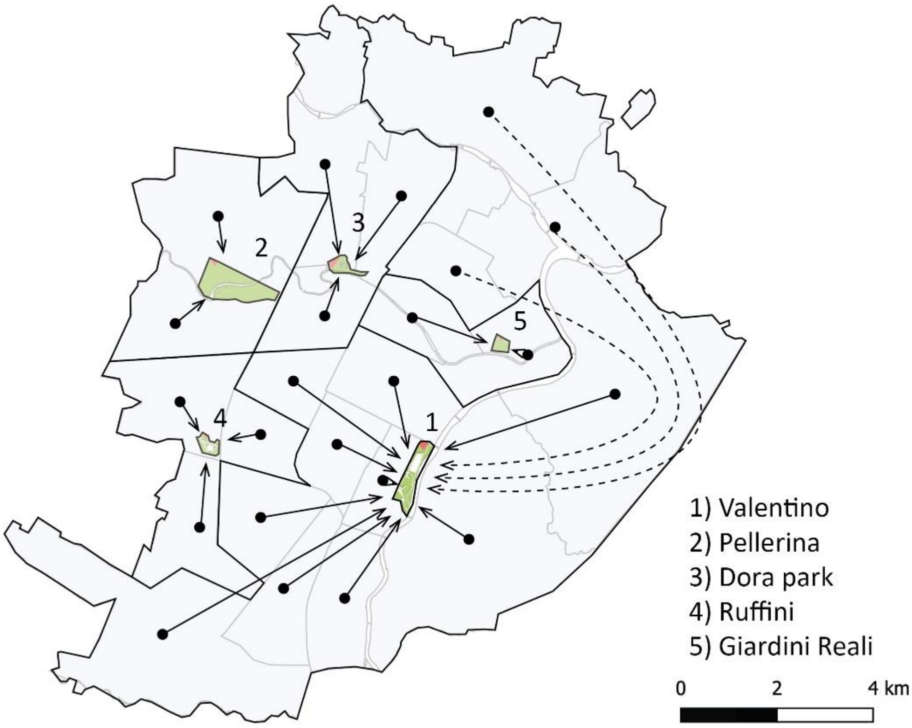


Figure A.3 preferred parks depending on the Neighborhood of residency

The second and third best performing OneR model is the one using the working neighborhood (51.54% precision) and age (47.44% precision) to base the prediction. The other models perform close to the classification of instances assigning them always to the most popular class (i.e., Valentino park), with a 41.77% of precision (see supplementary materials). The low performance of the latter models might reside in the number of possible values of the attributes used (Table A.1), being them fewer than the possible classes of preferred parks. Nonetheless, this first analysis highlights a certain degree of proximity effect, being the park closest to a respondent house her preferred one.

A.3.2 Matrix of park characteristics

The features that each respondent could have stated as present in her favorite park is displayed in Table A.2, and their frequency has been calculated according to (2). The results have been discretized in 5 possible clusters of frequency: (i) 0-25%, (ii) 26-50%, (iii) 51-75%, (iv) 76-100%, while complete absence of an attribute is assigned when no respondent registered its presence.

Table A.2 Matrix of parks characteristics

Chosen park	n. choices	Farm school	Educational tour	Social centre	Urban gardens	Natural trails	Birdwatching	Butterfly farm	Horse farm	No food crop	Phytodepuration	Bio-energy areas	Research centre	Sport fields	Skatepark	Outdoor equipment	Cycle path	Children playground	Dog park	Study areas	Organized beach
1) Valentino	137	0-25%	51-75%	0-25%	0-25%	0-25%	0-25%	Absent	0-25%	0-25%	0-25%	0-25%	0-25%	51-75%	26-50%	51-75%	76-100%	51-75%	51-75%	51-75%	0-25%
2) Pellerina	66	0-25%	0-25%	0-25%	0-25%	0-25%	0-25%	Absent	0-25%	Absent	Absent	0-25%	Absent	51-75%	0-25%	26-50%	51-75%	51-75%	51-75%	0-25%	0-25%
3) Dora	58	0-25%	0-25%	0-25%	0-25%	0-25%	Absent	0-25%	0-25%	0-25%	0-25%	0-25%	0-25%	51-75%	0-25%	26-50%	51-75%	51-75%	51-75%	26-50%	0-25%
4) Ruffini	37	Absent	Absent	0-25%	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	76-100%	51-75%	76-100%	76-100%	76-100%	76-100%	51-75%	Absent
5) Giardini R.	30	Absent	26-50%	0-25%	0-25%	Absent	Absent	0-25%	0-25%	0-25%	Absent	Absent	Absent	0-25%	0-25%	0-25%	0-25%	0-25%	26-50%	26-50%	Absent

Legend
Absent
0-25%
26-50%
51-75%
76-100%

From Table A.2, it appears clear that there is an overall homogeneity among the parks perceived features, with some specificity. The "sport oriented" features (Sport fields, Outdoor equipment, Cycle path, Children playground and Dog park) are the most registered ones, with specific concentration in the Ruffini park: here almost all the respondents affirm the presence of such characteristics without noticing almost any other characteristic. Valentino park and Giardini Reali one are the only two in which the presence of historical features is recognized, testified by the presence of "Educational tour" feature. The other characteristics are perceived by a low amount of respondents justifying the assumption that they are either not present, or their presence is not perceived.

A.3.3 Matrix of users' profile

Eight algorithms have been tested together with three attribute selection methods (backward, forward and bi-directional selection) and their performance evaluated regarding their prediction accuracy and AUC value. Results are shown in Table A.3.

Table A.3 algorithms and set of attributes used in the different models

Algorithm	Selection method: Backward	Forward	Bi-directional	Correctly predicted	Weighted AUC	Selected attributes: Gender	Age	Nationality	Kids (Y/N)	Single person	Elder (Y/N)	Study level	Profession	Income	Working h	Neig_Res.	Neig_Work
Naïve Bayes	X			47.26	0.729												
		X		46.04	0.725												
			X	46.65	0.729												
SMO	X			47.26	0.656												
		X		48.48	0.648												
			X	49.70	0.657												
Multilayer Perceptron	X			46.65	0.646												
		X		48.48	0.674												
			X	48.78	0.682												
IBK	X			55.18	0.733												
		X		57.93	0.777												
			X	57.93	0.777												
PART	X			54.88	0.741												
		X		55.18	0.761												
			X	54.88	0.764												
JRip	X			55.18	0.675												
		X		53.05	0.683												
			X	55.18	0.677												
J48	X			56.90	0.737												
		X		59.15	0.735												
			X	58.54	0.739												
Random Forest	X			58.54	0.794												
		X		56.40	0.793												
			X	58.23	0.799												
Total appearance						16	9	7	13	9	8	10	7	8	7	23	8

The best performing algorithm is the J48 (decision tree), with an accuracy ranging from 56.90% to 59.15%, and an AUC from 0.735 to 0.739. Due to this good performance the J48 has been further used to extract the park users' profile. To do so, four J48 models have been tested using a 10-fold cross validation approach repeated 25 times each, and the results averaged. The four algorithms (Table A.4) used different subsets of attributes: (i) the first model used the same subset as the most accurate J48, while the other three add (ii) "Age" and (iii) "Study Level" and (iv) both attributes. In particular, the choice of subsets is derived by Table A.3, being "Age" and "Study level" the two most frequent attributes other than the subset used by the best performing J48.

Table A.4 Decision tree models applied to construct the users' profile

Model	Gender	Age	Family with Kids	Single person family	Elder	Study level	Residency neighbourhood	Working neighbourhood	Average Percentage of correct	Average AUC
1st	X		X	X	X		X	X	62.04	0.772
2nd	X	X	X	X	X		X	X	55.32	0.729
3rd	X		X	X	X	X	X	X	54.72	0.727
4th	X	X	X	X	X	X	X	X	55.88	0.719

All the models yelled quite good AUC values (higher than 0.5). The first model has the highest percentage of correctness but resulted in a pretty big tree (tree size = 71) making it hard to interpret the results. The same could be said about the third model (tree size = 49). The second model has the second to highest AUC, with a percentage of corrected prediction aligned with the second to best one, while keeping the dimension of the tree manageable (tree size = 25). The third model has a quite good prediction capacity but lacks information regarding one of the parks: Giardini Reali is not represented in any leaf of the tree, making it impossible to determine a user profile for this park. Therefore, the second model has been used to extract the users' profile and its graphical representation is presented in Figure A.4.

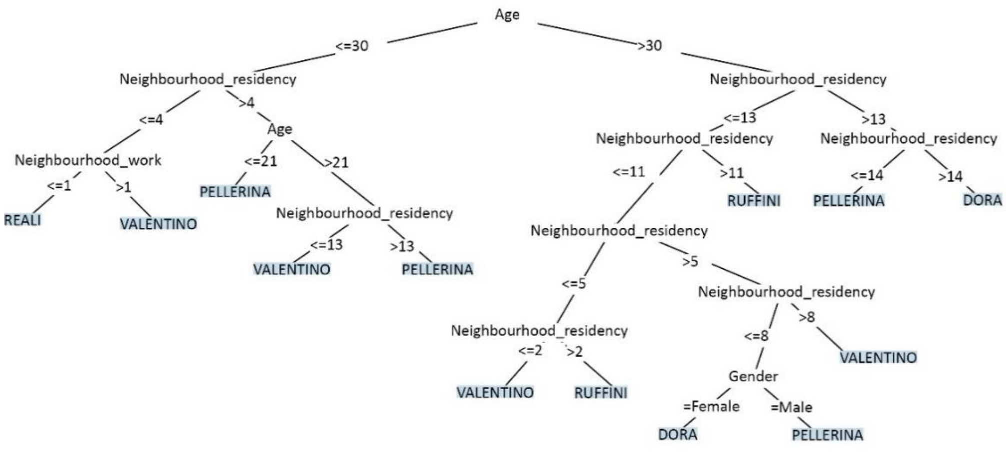


Figure A.4 Decision tree of the 2nd model

The Decision Tree in Figure A.4 could be analyzed to define the rules that lead to the preference for a specific park. In particular, starting from a leaf, it is possible to reconstruct unique socio-economic user's profile in the form of "if…, then…" rules, following the splitting conditions up until the first one (in our case being "Age" either lower or greater than 30). Taking the left outermost leaf, for example, a set of rules is extracted: “if the respondent (i) works in neighborhood 1, and (ii) she lives in a neighborhood with ID in the range 1 to 4, and (iii) she is younger than 30 years old, then she will prefer Giardini Reali Park”. The same procedure is repeated for each leaf to build the *Matrix of the users’ profile*, each set of rules representing the socio-economic profile of a user preferring a specific park over all the others.

A.3.4 Global profile matrix

Combining the two previously generated matrixes, the *Global profile matrix* is constructed (Table A.5). Each row of this matrix identifies the socio-economic profile of a park user (*Matrix of the user’s profile*), and the attributes that are present (or perceived as so) in her preferred park (*Matrix of park characteristics*).

Table A.5 Global profile matrix

		Park characteristics																					
Chosen park	n. choices	Farm school	Educational tour	Social centre	Natural trails	Urban gardens	Birdwatching	Butterfly farm	Horse farm	No food crop	Phytodepuration	Bio-energy areas	Research centre	Sport fields	Skatepark	Outdoor equipment	Cycle path	Children playground	Dog park	Study areas	Organized beach	Rule ID	Users’ profile characteristics
Valentino	137																					A B C	Age<=30, & Work>1, & Res∈[1;3] Age∈[21;30], & Res∈[4;13] Age>30,& Res∈[1;2]U[9;13]
Pellerina	66																					D E F G	Age<=21, & Res∈[5;23] Age∈[21;30], & Res∈[13;23] Age>30, & Gen=M, AND Res∈[6;8] Age>30, & Res=14
Dora	58																					H I	Age>30, & Gen=F, AND Res∈[6;8] Age>30, & Res∈[15;23]
Ruffini	37																					L	Age>30, & Res∈[3;5]U[12;13]
Giardini Reali	30																					M	Age<=30, & Work=1, & Res∈[1;4]

By spatializing the identified rules (Figure A.5), it is possible to highlight which are the preferences of the respondents depending on their Neighborhood of residency, and, considering the proximity effect, the preference of the individuals that are most likely to become the most frequent users of the forecasted park.

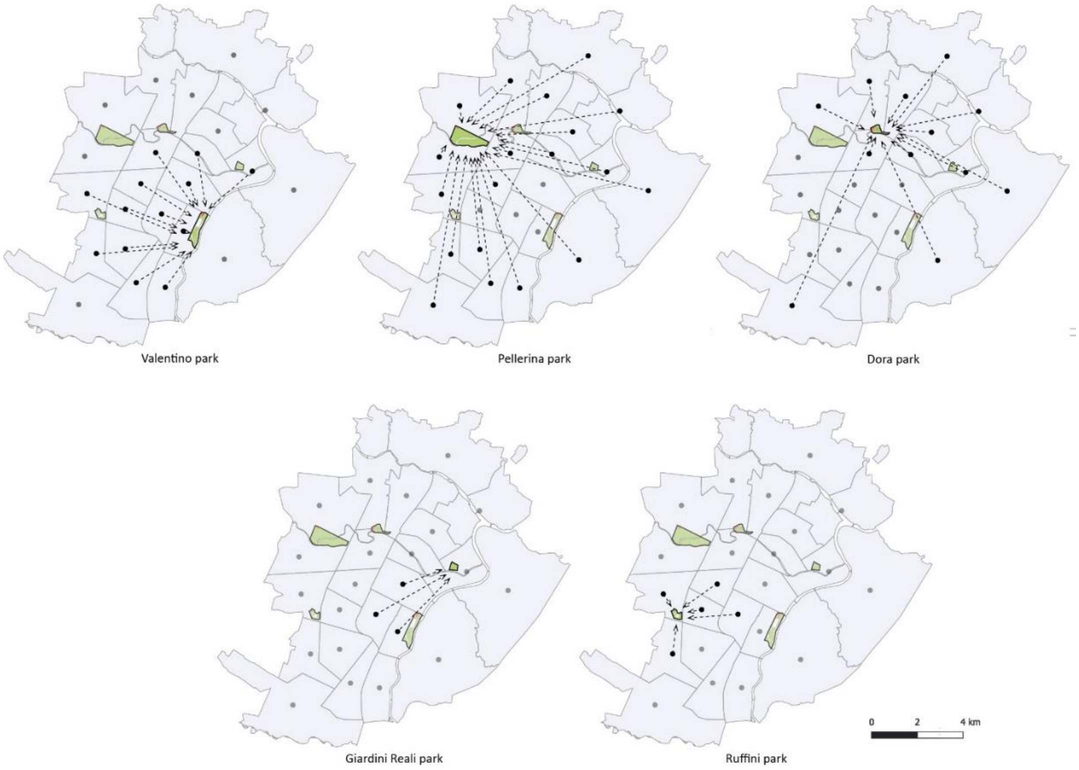


Figure A.5 Spatialization of the Decision rules

In particular, it appears clear that the respondents located in the gravitational area of the forecasted new park follow rules D, E and I, with a main preference for Pellerina Park and Dora Park, the main predictors being the age of the respondents. Furthermore, Table A.6 displays a comparison between the original Matrix of park characteristics, and a second one constructed by limiting the application of equation (2) to the profiles that satisfy the highlighted Decision Rules (D, E, and I).

Table A.6 Matrix of park characteristics according to the Decision Rules

Chosen park	n. choices	Farm school	Educational tour	Social centre	Urban garden	Natural trails	Birdwatching	Butterfly farm	Horse farm	No food crop	Phytodepuration	Bio-energy areas	Research centre	Sport fields	Skatepark	Outdoor equipment	Cycle path	Children playground	Dog park	Study areas	Organized beach
Pellerina	66	.05	.02	.08	.05	.05	.06	0	.02	0	0	.02	0	.65	.23	.47	.53	.64	.64	.23	.03
Pellerina limit	20	.05	.05	.10	.05	.10	.05	0	0	0	0	.05	0	.75	.10	.55	.40	.65	.60	.20	.05
ΔPellerina			.03	.02		.05	-.01		-.02			.03		.10	-.13	.08	-.13	.01	-.04	-.03	.02
Dora	58	.03	.19	.02	.05	.03	0	.03	.02	.02	.22	.03	.16	.55	.69	.66	.60	.71	.60	.26	.03
Dora limit	45	.04	.24	0	.04	.04	0	.04	.02	.02	.22	.04	.18	.56	.69	.67	.60	.71	.58	.24	.02
ΔDora		.01	.05	-.02	-.01	.01		.01				-.01	.02	.01		.01			-.02	-.02	-.01

The perceived frequency of park feature remains quite stable, with some prominent variations limited to the Pellerina park, where “Sport fields” and “Outdoor equipment” rise by 10 and 8 percentage points respectively, while “Skatepark” and “Cycle path” reduce their perceived frequency by 13 percent.

A.4 Discussions

> Urban green area accessibility and services provision

The first analysis conducted using a simple 1-rule algorithm (OneR) has highlighted the presence of a gravitational effect confirming the propensity to prefer a park close to the residency neighborhood. This trend seems to be confirmed by the spatialization of rules A, G, and L, while rule F and H might be explained with a potential deficit in the provision of park characteristics: these two profiles prefer to commute longer distances to enjoy spending time in a park further away. The same holds true for rule D, E and I when spatialized in the eastern and southern part of the city, and thus distant form the preferred parks located in the north-west part of the city. Also rule B follows the same pattern: with its corresponding user’s profiles (age between 21 and 30 years old) preferring the Valentino park located in the eastern part of the city, rather than parks closer to their neighborhood of residency (mostly located in the western part). It is worth noting that Valentino Park is located near one University, and this preference might reflect this aspect. In fact, “Working neighborhood” attribute did not specifically take into consideration faculties locations as working places for university students for example (aspect that might be worth investigating in the future).

> Socio-economic predictors

The present study found that several attributes could be used as predictors of the individual preference towards urban parks such as gender, neighborhood of residency, and attributes related to family composition (presence of kids or elder people or a single person family). This outcome is consistent with other studies related to landscape and park preferences based on gender and family composition. Marital status influences on park attributes perception is also highlighted together with a negative correlation between the distance from respondents’ home to an Urban Green Area.

Similar to other studies, the analysis of the different algorithms (Table A.2) has shown that other attributes such as “Age” and “Study level” have some potential in preferences prediction. The influence of these two socio-economic attributes is consistent with several sources). The high predictive potential of the Residency Neighborhood (used in 23 out of 24 of the models) seems coherent with the fact

that people tend to consider important to have urban green areas close to their home while it also gives important insight of the possible lack of green area provision and their specific characteristics in certain parts of the city. Rather than being related to specific park amenities, the gender attribute might be related to safety perception. Regarding this, no specific question has been designed in the questionnaire to evaluate this aspect, representing a possible further expansion of the analysis.
Also the highlighted attributes describing family composition seem reasonable to influence the preference: families with fragile members such as elder people or kids might prefer specific parks, while single person might favor them depending on social interactions possibilities. Finally, considering the study level, this attribute has been often associated with vegetation perception. While it might be difficult to establish a correlation between Study Level and park attributes, it has been argued that the interactions among demographic variables should not be neglected. In light of this, for example, we might consider some sort of overlapping between age, educational level, and the family composition. A further investigation of this aspect might be beneficial in order to better understand the predictive capacity of such attribute.

> Design of the new park

As already showed, rules D, E and I are the one that are spatialized in the area where the new park is meant to be established. Especially rule D and E might be interpreted as a lack of the provision of recreational activities in the area: the matrix of parks' characteristics (Table A.2), in fact, shows an increase of the frequency of respondents noticing the presence of such characteristics in the park that they have chosen as their preferred one ("Sport field" +10%, "Outdoor equipment" +8%). Considering the low score of the majority of characteristics (most of them less than 10%), and the reduction of other frequently perceived ones such as "Cycle path" (-13%), and "Dog park" (-4%) both actually present in the Pellerina park2 it might be said that there is a demand for outdoor sport facilities in the area among the young population.
Regarding the presence of such facilities in parks, there is no agreement over their influence on user perception. Some Authors recognize a low importance attributed to furniture and recreational facilities, while others highlight the presence of trade-offs between recreational facilities and natural characteristics.
Our study confirms the importance of recreational facilities, suggesting that one of the principal characteristics that determine the preference of parks is their presence in the park. For these reasons, it is desirable that the design of the new park to be

[2] http://www.comune.torino.it/verdepubblico/parco-della-pellerina/

established in the Basse di Stura area would focus explicitly on the provision of these types of infrastructures (i.e., sport fields, outdoor equipment). This strategy would likely attract the younger population more interested in these characteristics (rule D and E).
Considering rule I, the preference towards Dora park seems to confirm the demand for recreational facilities, being them the most perceived features. Limiting the frequency of park attributes to the users that have expressed their preference for this park (Table A.6), the scores of each attribute remains quite stable, with the highest difference registered for the presence of "Educational tours" (+5%). Indeed, the Dora park is deeply informed by its industrial heritage features that constitutes the most iconic element of the park and this is acknowledged by the respondents (19% and 24% frequency of "Educational tour"). Also the Pellerina park is characterized by the presence of cultural heritage buildings, but this aspect is less noted by the respondents. Indeed, it might be argued that cultural attractions are less valued in park use motivations. The results suggest that the cultural heritage preservation might not be a priority in the design of the new park. This outcome could suggest the possibility for the DMs to envision adaptive reuse of the cultural heritage in the area, with the necessity to further investigate this specific aspect.
Also other human-nature interaction characteristics have been highlighted, especially in the Pellerina park. Here, the presence of natural trails and the possibility for birdwatching are not neglected by respondents' perception, requiring a specific attention to be given to the balance of natural-artificial aspects of the new park design.

A.5 Conclusions

The provision of green spaces has been stressed as fundamental to enhance the quality of life of urban population, providing them with a large variety of benefits. Considering the preferences of final users towards Urban Parks characteristics is fundamental to increase their satisfaction and the public support of urban green infrastructure administration.
In this research, the Preference Learning approach has been used to determine possible design guidelines of a new urban park in Turin by deriving people's preference for existing parks in the city.
First of all, a "gravitational effect" on park preferences has been confirmed by the use of a simple Machine-Learning algorithm (OneR). Furthermore, socio-economic characteristics like gender, neighborhood of residency, family composition, age, and educational level could be used as predictors to determine individual

preferences toward different urban parks characteristics. Finally, the combination of park users' profile and the preferred parks perceived characteristics in the Global profile matrix has allowed to determine the demand for certain park characteristics in different areas of the city. Specifically considering the area under investigation, the recognition of the features that are most frequently perceived by the proximity users of the future park reveals the demand for recreational and sport facilities. At the same time, the respondents do not seem to have preferences towards the possible preservation of the historical building heritage of the area. This information could be useful to help Decision Makers to define strategies in defining the most successful layout of the "Basse di Stura" regeneration project.

This study proves that the application of Machine Learning technique could be used as a decision supporting tool in regeneration processes. Furthermore, such techniques might enhance participation by individuating possible stakeholders potentially interested in the process, as well as their preferences. The followed workflow could be useful to reduce the bias of respondents in expressing their preferences toward attributes and the limitation of other methods (i.e. the limited number of characteristics in a Choice Experiment).

This study is fundamentally explorative in nature, and it might be extended in order to provide a wider spectrum of insight on the design of urban green areas alternatives. First of all, a more complex Decision Tree might be used to generate more rules and thus identify more specific park users' profiles. Furthermore, the expansion of the number of respondents might be welcome to include some existing urban parks excluded in the data pre-processing phase. This last possibility might also help solve one of the limitation of the current study: the absence of several facilities in existing urban parks, in fact, limits the spectrum of possible features and characteristics that could be envisioned for the new urban area to the one existing in the 5 analyzed ones. Further considering the less frequently perceived green areas characteristics, it might be possible to deepen their analysis, as well as to focus on others that are not present in the existing parks of the city by combining different evaluation methods (Choice experiment, Social Media reviews etc.), thus ensuring the meeting of the population demands and proposing more innovative alternative uses and amenities.

B. Meta-analysis of Italian case studies

Table B.1 Meta-analysis of the reconstructed Italian case studies analyzed in the literature review

Source/ scenario	$E_{wit,el}$ [kWh_{el}/apt/y]	$E_{wit,ng}$ [kWh_{gas}/apt/y]	$E_{fed,el}$ [kWh_{el}/apt/y]	SSR_{tot} [%]	SCR_{el} [%]***	CO_{inv} [€/apt]	CO_{2emit} [kg_{CO2}/apt/y]	CO_{en} [€/apt/y]
[29]†								
Baseline	5,760	1,900	0	0.0	0.0	0	1,975	1,223
1	4,160	1,900	1,250	23.6	56.1	1,341	1,533	929
2	3,495	1,900	585	33.4	79.5	2,533	1,349	806
3	3,720	1,900	810	30.1	71.6	1,760	1,412	848
4	3,085	1,900	175	39.4	93.9	3,303	1,236	731
5	3,895	1,900	3,835	27.5	32.7	2,261	1,460	880
6	2,055	1,900	1,995	54.6	65.0	5,149	952	542
7	2,975	1,900	2,915	41.1	48.9	2,992	1,206	711
8	1,610	1,900	1,550	61.2	72.8	6,600	829	460
9	4,160	1,710	355	25.1	87.5	1,824	1,495	913
10	3,720	1,710	155	31.6	94.6	2,243	1,373	832
11	3,895	1,710	405	29.0	92.9	3,950	1,422	864
12	3,895	1,710	165	29.0	97.1	4,439	1,422	864
[28]*								
Baseline(Crocetta)	1,660	15,978	0	0.0	0.0		3,686	1,739
1_Cr	1,212	15,978	192	4.4	70.0	644	3,562	1,641
2_Cr	780	15,978	377	8.6	70.0	2,136	3,443	1,546
Baseline(Arquata)	1,180	10,754	0	0.0	0.0		2,498	1,184
1_Ar	861	10,754	137	4.6	70.0	472	2,410	1,114
2_Ar	389	10,754	339	11.3	70.0	2,291	2,280	1,011
[98]								
Baseline(TN)	8,974	0	0	0.0	0.0	0	2,470	1,967
1_Tn	6,836	0	1,745	24.0	55.3	1,151	1,889	1,504
2_Tn	6,365	0	3,224	28.8	44.4	1,651	1,759	1,400
3_Tn	6,156	0	5,011	31.5	36.1	2,056	1,701	1,354
Baseline(PD)	8,646	0	0	0.0	0.0	0	2,383	1,898
1_Pd	6,401	0	2,000	25.9	52.8	1,151	1,769	1,408
2_Pd	6,020	0	3,642	30.2	41.7	1,651	1,663	1,324
3_Pd	5,830	0	5,674	32.8	33.4	2,056	1,611	1,283
Baseline(Bl)	9,196	0	0	0.0	0.0	0	2,533	2,017
1_Bl	7,018	0	1,862	23.9	54.2	1,151	1,939	1,544
2_Bl	6,564	0	3,382	28.4	43.5	1,651	1,814	1,444
3_Bl	6,338	0	5,220	31.1	35.4	2,056	1,751	1,394
[101]*								
Baseline	1,470	10,073	0	0.0	0.0	0	406	366
1	1,136	10,073	33	4.9	91.0	238	314	283
2	1,000	10,073	265	6.8	64.0	425	276	249
3	941	10,073	573	7.7	48.0	597	260	234

4 [102]	911	10,073	911	8.1	38.0	759	252	227
Baseline(Pi)	2,750	9,410	0	0.0	0.0	0	2,661	1,414
1_Pi	1,819	9,410	224	11.9	80.6	1,281	2,403	1,209
2_Pi	3,869	344	270	20.4	79.4	4,825	1,138	881
Baseline(Ab)	2,750	6,939	0	0.0	0.0	0	2,161	1,202
1_Ab	1,827	6,939	217	14.2	81.0	1,194	1,906	999
2_Ab	3,208	323	261	22.8	79.3	4,288	952	733
Baseline(Ba)	2,750	6,089	0	0.0	0.0	0	1,990	1,129
1_Ba	1,824	6,089	215	15.4	81.2	1,125	1,734	925
2_Ba	3,026	256	255	23.8	79.5	4,023	888	688
Baseline(Cal)	2,750	3,966	0	0.0	0.0	0	1,561	946
1_Cal	1,825	3,966	208	18.9	81.7	1,083	1,305	742
2_Cal	2,526	272	248	26.8	79.8	3,336	753	579
Baseline(Cam)	2,750	5,454	0	0.0	0.0	0	1,861	1,074
1_Cam	1,829	5,454	219	16.2	80.8	1,138	1,607	871
2_Cam	2,856	347	252	24.2	79.4	3,754	859	658
Baseline(ER)	2,750	8,722	0	0.0	0.0	0	2,522	1,355
1_Er	1,841	8,722	227	12.2	80.0	1,277	2,271	1,155
2_Er	3,696	402	253	20.2	79.7	4,651	1,102	848
Baseline(FVG)	2,750	10,086	0	0.0	0.0	0	2,797	1,472
1_Fvg	1,861	10,086	214	10.9	80.6	1,312	2,552	1,277
2_Fvg	3,921	648	251	18.9	79.8	4,761	1,214	918
Baseline(La)	2,750	5,124	0	0.0	0.0	0	1,795	1,046
1_La	1,811	5,124	215	17.0	81.3	1,129	1,536	839
2_La	2,765	311	259	25.4	79.4	3,725	827	635
Baseline(Li)	2,750	7,273	0	0.0	0.0	0	2,229	1,230
1_Li	1,846	7,273	207	13.6	81.4	1,237	1,979	1,032
2_Li	3,263	488	245	21.8	80.0	4,055	1,000	760
Baseline(Lo)	2,750	9,106	0	0.0	0.0	0	2,599	1,388
1_Lo	1,852	9,106	219	11.7	80.4	1,261	2,351	1,191
2_Lo	3,875	279	235	19.4	80.5	4,701	1,127	876
Baseline(Ma)	2,750	7,390	0	0.0	0.0	0	2,253	1,241
1_Ma	1,838	7,390	228	13.5	80.0	1,247	2,001	1,040
2_Ma	3,370	320	246	21.4	79.7	4,443	996	769
Baseline(Mo)	2,750	7,093	0	0.0	0.0	0	2,193	1,215
1_Mo	1,822	7,093	222	14.1	80.7	1,173	1,936	1,011
2_Mo	3,237	325	265	22.7	79.1	4,124	960	740
Baseline(Pu)	2,750	4,458	0	0.0	0.0	0	1,660	988
1_Pu	1,810	4,458	216	18.2	81.3	1,098	1,401	782
2_Pu	2,641	240	255	26.4	79.6	3,626	778	602
Baseline(Sa)	2,750	4,243	0	0.0	0.0	0	1,617	970
1_Sa	1,791	4,243	216	19.1	81.6	1,084	1,352	759
2_Sa	2,569	264	247	26.9	80.2	3,395	763	588
Baseline(Si)	2,750	3,625	0	0.0	0.0	0	1,492	917
1_Si	1,797	3,625	207	20.3	82.1	1,033	1,229	707
2_Si	2,443	208	253	28.3	79.9	3,300	717	555
Baseline(To)	2,750	7,155	0	0.0	0.0	0	2,205	1,220
1_To	1,822	7,155	224	14.1	80.6	1,217	1,949	1,016
2_To	3,212	427	261	22.5	79.3	4,213	974	743
Baseline(TAA)	2,750	11,224	0	0.0	0.0	0	3,027	1,570
1_Taa	1,888	11,224	187	9.8	82.2	1,283	2,789	1,381
2_Taa	4,278	652	234	17.8	81.1	4,879	1,314	997

Baseline(Um)	2,750	7,990	0	0.0	0.0	0	2,374	1,292
1_Um	1,833	7,990	221	13.0	80.6	1,190	2,120	1,090
2_Um	3,515	352	245	20.8	79.9	4,410	1,042	804
Baseline(VDA)	2,750	14,627	0	0.0	0.0	0	3,714	1,863
1_Vda	1,879	14,627	183	8.2	82.6	1,308	3,474	1,671
2_Vda	5,075	607	271	16.8	80.1	5,322	1,525	1,169
Baseline(Ve)	2,750	8,495	0	0.0	0.0	0	2,476	1,336
1_Ve	1,852	8,495	216	12.3	80.6	1,253	2,228	1,138
2_Ve	3,711	272	235	20.0	80.4	4,598	1,080	840
[60]*†								
Baseline	2,902	9,410	0	0.0	0.0	0	2,703	1,448
1	1,876	9,410	571	12.9	64.3	1,347	2,419	1,222
[47]*								
Baseline	3,524	3,625	0	0.0	0.0	0	1,706	823
1	2,396	3,625	851	20.6	57.0	1,671	1,394	659
[236]								
Baseline	2,373	6,939	0	0.0	0.0	0	2,057	1,119
1	2,008	6,939	0	6.0	100.0	347	1,956	1,038
2	1,723	6,895	4	11.0	99.5	972	1,869	972
3	1,609	6,752	7	14.1	99.3	1,509	1,809	935
4	1,546	6,659	142	16.0	90.3	1,816	1,772	913
5	1,500	6,789	694	15.6	62.1	1,960	1,786	914
6	1,465	6,939	1,287	14.9	41.4	2,083	1,806	919
7	1,437	6,939	1,625	15.3	36.5	2,431	1,799	913
8	1,416	6,939	1,970	15.7	32.7	2,778	1,793	908
[151]								
Baseline	36,000**		0	0.0	0.0	0	8,429	5,357
1	23,143**		0	0.0	0.0	20,024	5,333	3,714
[150]*†								
Baseline	3,826	9,234	0	0.0	0.0	0	2,922	1,498
1	4,472	0	463	29.3	80.0	4,121	1,236	823
2	3,962	0	2,270	37.4	51.0	4,996	1,095	729
3	3,823	0	4,447	39.5	36.0	5,799	1,056	703
4	4,032	0	23	36.3	99.0	5,104	1,114	742
5	4,217	0	208	33.3	91.0	4,483	1,165	776
6	2,146	0	454	66.1	90.2	7,784	593	395
7	3,286	0	1,593	48.0	65.6	5,560	908	605
8	1,155	0	1,779	81.7	74.4	9,376	319	213
9	2,795	0	3,418	55.8	50.8	6,452	772	514
[62]*†								
Baseline	4,097	9,410	0	0	0	0	3,033	1,711
1	2,718	9,410	57	15	96	1,187	2,652	1,407
2	2,681	9,410	74	15	95	1,232	2,642	1,399
3	2,603	9,410	112	16	93	1,335	2,620	1,382
4	2,536	9,410	173	17	90	1,438	2,601	1,367
5	2,475	9,410	221	18	88	1,541	2,585	1,354
6	2,421	9,410	296	18	85	1,643	2,570	1,342
7	2,372	9,410	379	19	82	1,746	2,556	1,331
8	2,328	9,410	442	19	80	1,849	2,544	1,321
9	2,289	9,410	540	20	77	1,951	2,533	1,313
10	2,253	9,410	615	20	75	2,054	2,523	1,305
11	2,220	9,410	694	20	73	2,157	2,514	1,298
12	2,190	9,410	779	21	71	2,259	2,506	1,291

13	2,161	9,410	911	21	68	2,362	2,498	1,285
14	2,134	9,410	967	21	67	2,465	2,490	1,279
15	2,109	9,410	979	22	67	2,568	2,483	1,273
16	2,085	9,410	1,084	22	65	2,670	2,477	1,268
17	2,062	9,410	1,195	22	63	2,773	2,471	1,263
18	2,041	9,410	1,315	22	61	2,876	2,465	1,258
19	2,021	9,410	1,384	23	60	2,978	2,459	1,254
20	2,003	9,410	1,517	23	58	3,081	2,454	1,250
21	1,985	9,410	1,593	23	57	3,184	2,449	1,246
22	1,969	9,410	1,672	23	56	3,286	2,445	1,242
23	1,953	9,410	1,826	23	54	3,389	2,440	1,239
24	1,938	9,410	1,915	24	53	3,492	2,436	1,236
25	1,923	9,410	2,007	24	52	3,595	2,432	1,232

*for these data, the energy consumption for heating from Canova et al. [102] have been taken,

**only given as primary energy consumption,

***all the electricity self-consumed within the community is considered (also the one used for energy conversion).

†: for these studies SSR was not given, but it has been calculated based on self consumption and production data.

C. The questionnaire

The following images show the online questionnaire that the respondents were asked to fill in. These were preceded by an introductory part explaining the questionnaire and its aims. A short video was included to provide information regarding the REC concept

Figure C.1 Screenshot of the explanatory video presented in the questionnaire

	NO	più NO che si	più SI che no	SI	Non so
La collaborazione tra persone è fondamentale per ridurre i consumi energetici ed essere tutti più sostenibili	○	○	○	○	○
Ridurre il consumo di energia è necessario per contrastare il cambiamento climatico	○	○	○	○	○
Adottare misure di efficientamento energetico nella mia abitazione mi sembra costoso, e non è una mia priorità	○	○	○	○	○

Figure C.2 Introductory 4-point Likert scale questions to gather introductory data regarding respondent’s inclination toward RECs and their possible economic, environmental, and social benefits.

Figure C.3 Selection of the alternatives that the respondent would have accepted, maybe accepted, or rejected. It is worth noting that the respondent was exposed to all the alternatives in the same webpage (this part was explained with a short video to explain the meaning of the different performance attributes of the offered REC).

Figure C.4 First ranking of alternatives in which the respondent was asked to assign an order of preference to the alternatives for which the answer "YES" was given. This part was introduced by a short video to explain how to proceed with the ranking of the alternative

Figure C.5 Second ranking of alternatives in which the respondent was asked to assign an order of preference to the alternatives for which the answer “MAYBE” was given.

Figure C.6 Set of questions to gather socio-demographic information regarding respondents and their household

Figure C.7 Final question regarding the position -or in proximity- of the respondents' household of residency. Alternatively, respondents were asked to specify their postcode.

D. GLMM models

In the following tables the preparatory analysis to the definition of the Generalized linear mixed model are presented

Table D 1 First analysis including all the variable classes from the questionnaire

Predictors	OR	CI 95%	p
Numerical			
Intercept	3.70	0.67-20.52	0.135
Inv_cost (k€)	**0.75**	0.73-0.76	<0.001
Cost_sav (k€)	**4.58**	3.21-6.53	<0.001
Emis_sav(tCO_2)	**1.66**	1.47-1.89	<0.001
SSuff(%)	**1.43**	1.38-1.47	<0.001
HH_num	0.87	0.75-1.00	0.052
Categorical			
Age [20-24]	ref.		
Age [25-29]	0.98	0.36-2.67	0.970
Age [30-34]	0.83	0.31-2.24	0.716
Age [35-39]	0.57	0.20-1.58	0.281
Age [40-44]	**0.31**	0.10-0.97	0.045
Age [45-49]	0.60	0.22-1.68	0.332
Age [50-54]	0.57	0.20-1.65	0.305
Age [55-59]	0.57	0.21-1.57	0.276
Age [60-64]	**0.30**	0.10-0.86	0.025
Age [65-69]	**0.34**	0.12-0.98	0.045
Age [70-74]	0.51	0.14-1.88	0.310
Age [75+]	**0.26**	0.07-0.92	0.037
Age [< 35]			
Age [35-60]			
Age [60+]			
HH_num [>3]			
HH_num [3+]			
HH_inc [<15k]	ref.		
HH_inc [15-28k]	0.98	0.43-2.20	0.954
HH_inc [28-55k]	1.35	0.60-3.03	0.472
HH_inc [55-75k]	1.26	0.54-2.97	0.596
HH_inc [75k+]	1.72	0.72-4.11	0.220
HH_inc [<28k]			
HH_inc [28-75k]			
HH_inc [75k+]			
Edu_lev [middle]	ref.		
Edu_lev [high]	**0.21**	0.05-0.92	0.039
Edu_lev [Dgr]	**0.20**	0.05-0.87	0.032
Edu_lev [Dgr+]	**0.16**	0.04-0.71	0.016
Edu_lev [Dgr_N]			

Edu_lev [Dgr_Y]	
Gender [Fem]	
Gender [Mal]	
HH_prop [rent]	
HH_prop [own]	
Random Effects	
σ^2	3.29
τ_{00}	1.36
ICC	0.29
Evaluation metrics	
AIC	6093
BIC	6277
Marginal R^2 / Conditional R^2	0.595 / 0.714

Table D.2 Interaction between household income level and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.71	0.245	0.78	0.407	0.73	0.280	0.75	0.336
Inv_cost (k€)	**0.73**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.57**	<0.001	**3.38**	<0.001	**4.58**	<0.001	**4.58**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.67**	<0.001	**1.52**	<0.001	**1.66**	<0.001
SSuff (%)	**1.43**	<0.001	**1.43**	<0.001	**1.43**	<0.001	**1.36**	<0.001
Categorical								
Age [< 35]	ref.		ref.		ref.		ref.	
Age [35-60]	**0.62**	0.012	**0.62**	0.013	**0.62**	0.013	**0.62**	0.013
Age [60+]	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001
HH_num [>3]	ref.		ref.		ref.		ref.	
HH_num [3+]	**0.69**	0.022	**0.69**	0.022	**0.69**	0.021	**0.69**	0.021
HH_inc [<28k]	ref.		ref.		ref.		ref.	
HH_inc [28-75k]	1.19	0.430	1.06	0.819	1.15	0.512	1.08	0.736
HH_inc [75k+]	1.27	0.405	1.07	0.827	1.19	0.544	1.19	0.540
Edu_lev [Dgr_N]	ref.		ref.		ref.		ref.	
Edu_lev [Dgr_Y]	0.80	0.335	0.80	0.325	0.80	0.327	0.80	0.321
Gender [Fem]	ref.		ref.		ref.		ref.	
Gender [Mal]	0.82	0.189	0.82	0.191	0.82	0.191	0.83	0.195
HH_prop [rent]	ref.		ref.		ref.		ref.	
HH_prop [own]	0.82	0.254	0.82	0.258	0.82	0.259	0.82	0.261
Interactions								
Inv_cost (k€)								
*HH_inc [28-75k]	1.01	0.641						
*HH_inc [75k+]	1.04	0.096						
Cost_sav (k€)								
*HH_inc [28-75k]			1.36	0.224				
*HH_inc [75k+]			**2.05**	0.022				
Emis_sav(tCO_2)								
*HH_inc [28-75k]					1.08	0.451		

*HH_inc [75k+]					**1.30**	0.041		
SSuff (%)								
*HH_inc [28-75k]							1.05	0.181
*HH_inc [75k+]							**1.10**	0.042
Random Effects								
σ^2	3.29		3.29		3.29		3.29	
τ_{00}	1.42		1.42		1.41		1.41	
ICC	0.30		0.30		0.30		0.30	
Evaluation metrics								
AIC	6080		6078		6079		6079	
BIC	6189		6186		6187		6187	
Marginal R^2 / Conditional R^2	0.590 / 0.714		0.591 / 0.714		0.591 / 0.714		0.591 / 0.713	

Table D.3 Interaction between education level and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.64	0.155	0.63	0.159	0.63	0.148	0.70	0.261
Inv_cost (k€)	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.57**	<0.001	**4.81**	<0.001	**4.57**	<0.001	**4.57**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.66**	<0.001	**1.70**	<0.001	**1.66**	<0.001
SSuff (%)	**1.42**	<0.001	**1.42**	<0.001	**1.42**	<0.001	**1.39**	<0.001
Categorical								
Age [< 35]	ref.		ref.		ref.		ref.	
Age [35-60]	**0.62**	0.012	**0.62**	0.012	**0.62**	0.012	**0.62**	0.012
Age [60+]	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001
HH_num [>3]	ref.		ref.		ref.		ref.	
HH_num [3+]	**0.69**	0.022	**0.69**	0.022	**0.69**	0.022	**0.69**	0.022
HH_inc [<28k]	ref.		ref.		ref.		ref.	
HH_inc [28-75k]	1.27	0.178	1.27	0.178	1.27	0.178	1.27	0.178
HH_inc [75k+]	**1.65**	0.034	**1.65**	0.034	**1.65**	0.034	**1.65**	0.035
Edu_lev [Dgr_N]	ref.		ref.		ref.		ref.	
Edu_lev [Dgr_Y]	0.82	0.472	0.83	0.538	0.82	0.493	0.73	0.269
Gender [Fem]	ref.		ref.		ref.		ref.	
Gender [Mal]	0.82	0.192	0.82	0.255	0.82	0.255	0.82	0.192
HH_prop [rent]	ref.		ref.		ref.		ref.	
HH_prop [own]	0.82	0.256	0.82	0.255	0.82	0.255	0.82	0.257
Interactions								
Inv_cost (k€) *Edu_lev Dgr_Y	1.00	0.915						
Cost_sav (k€) *Edu_lev Dgr_Y			0.94	0.858				
Emis_sav(tCO_2) *Edu_lev Dgr_Y					0.98	0.870		
SSuff (%) *Edu_lev Dgr_Y							1.03	0.586
Random Effects								
σ^2	3.29		3.29		3.29		3.29	

τ_{00}	1.41	1.41	1.41	1.41
ICC	0.30	0.30	0.30	0.30
Evaluation metrics				
AIC	6081	6081	6081	6081
BIC	6183	6183	6183	6183
Marginal R^2 / Conditional R^2	0.590 / 0.713	0.590 / 0.713	0.590 / 0.713	0.590 / 0.713

Table D.4 Interaction between age and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.61	0.091	0.61	0.094	0.62	0.097	0.66	0.150
Inv_cost (k€)	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.58**	<0.001	**5.06**	<0.001	**4.57**	<0.001	**4.54**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.66**	<0.001	**1.73**	<0.001	**1.67**	<0.001
SSuff (%)	**1.43**	<0.001	**1.43**	<0.001	**1.43**	<0.001	**1.42**	<0.001
Categorical								
Age [< 35]	ref.		ref.		ref.		ref.	
Age [35-60]	0.70	0.112	0.72	0.162	0.65	0.054	0.68	0.080
Age [60+]	**0.39**	0.001	**0.38**	0.001	**0.44**	0.003	**0.31**	<0.001
HH_num [>3]	ref.		ref.		ref.		ref.	
HH_num [3+]	**0.69**	0.022	**0.69**	0.022	**0.69**	0.022	**0.69**	0.022
HH_inc [<28k]	ref.		ref.		ref.		ref.	
HH_inc [28-75k]	1.27	0.176	1.27	0.178	1.27	0.177	1.27	0.183
HH_inc [75k+]	**1.65**	0.034	**1.65**	0.034	**1.65**	0.034	**1.64**	0.036
Edu_lev [Dgr_N]	ref.		ref.		ref.		ref.	
Edu_lev [Dgr_Y]	0.80	0.328	0.80	0.330	0.80	0.333	0.80	0.322
Gender [Fem]	ref.		ref.		ref.		ref.	
Gender [Mal]	0.83	0.196	0.82	0.191	0.82	0.190	0.82	0.189
HH_prop [rent]	ref.		ref.		ref.		ref.	
HH_prop [own]	0.82	0.253	0.82	0.253	0.82	0.255	0.82	0.256
Interactions								
Inv_cost (k€)								
*Age [35-60]	0.98	0.354						
*Age [60+]	1.00	0.869						
Cost_sav (k€)								
*Age [35-60]			0.79	0.342				
*Age [60+]			1.01	0.969				
Emis_sav(tCO_2)								
*Age [35-60]					0.97	0.740		
*Age [60+]					0.89	0.355		
SSuff (%)								
*Age [35-60]							0.98	0.521
*Age [60+]							1.06	0.198
Random Effects								
σ^2	3.29		3.29		3.29		3.29	
τ_{00}	1.41		1.41		1.41		1.41	
ICC	0.30		0.30		0.30		0.30	
Evaluation metrics								

AIC	6082	6082	6082	6079
BIC	6191	6190	6191	6188
Marginal R^2 / Conditional R^2	0.592 / 0.714	0.591 / 0.714	0.591 / 0.714	0.590 / 0.713

Table D.5 Interaction between household property and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.74	0.299	0.72	0.272	0.61	0.090	0.85	0.575
Inv_cost (k€)	**0.73**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.56**	<0.001	**3.80**	<0.001	**4.56**	<0.001	**4.48**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.66**	<0.001	**1.75**	<0.001	**1.67**	<0.001
SSuff (%)	**1.42**	<0.001	**1.42**	<0.001	**1.43**	<0.001	**1.32**	<0.001
Categorical								
Age [< 35]	ref.							
Age [35-60]	**0.62**	0.013	**0.63**	0.013	**0.62**	0.012	**0.63**	0.013
Age [60+]	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001
HH_num [>3]	ref.							
HH_num [3+]	**0.69**	0.023	**0.69**	0.022	**0.69**	0.022	**0.69**	0.021
HH_inc [<28k]	ref.							
HH_inc [28-75k]	1.27	0.185	1.27	0.180	1.27	0.178	1.27	0.178
HH_inc [75k+]	**1.65**	0.035	**1.65**	0.034	**1.65**	0.034	**1.66**	0.034
Edu_lev [Dgr_N]	ref.							
Edu_lev [Dgr_Y]	0.80	0.337	0.80	0.334	0.80	0.332	0.80	0.330
Gender [Fem]	ref.							
Gender [Mal]	0.82	0.189	0.82	0.191	0.82	0.191	0.83	0.196
HH_prop [rent]	ref.							
HH_prop [own]	0.67	0.065	0.65	0.105	0.89	0.598	**0.55**	0.004
Interactions								
Inv_cost (k€) *HH_prop [own]	1.03	0.107						
Cost_sav (k€) *HH_prop [own]			1.31	0.239				
Emis_sav(tCO_2) *HH_prop [own]					0.93	0.432		
SSuff (%) *HH_prop [own]							**1.13**	<0.001
Random Effects								
σ^2	329		329		329		329	
τ_{00}	1.42		1.42		1.42		1.42	
ICC	0.30		0.30		0.30		0.30	
Evaluation metrics								
AIC	6078		6079		6080		6068	
BIC	6180		6181		6182		6170	
Marginal R^2 / Conditional R^2	0.588 / 0.712		0.589 / 0.712		0.590 / 0.713		0.591 / 0.714	

Table D.6 Interaction between gender and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.76	0.350	0.76	0.351	0.71	0.229	0.69	0.201
Inv_cost (k€)	**0.73**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.57**	0.012	**3.48**	<0.001	**4.57**	<0.001	**4.55**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.66**	<0.001	**1.55**	<0.001	**1.66**	<0.001
SSuff (%)	**1.43**	<0.001	**1.43**	<0.001	**1.42**	<0.001	**1.40**	<0.001
Categorical								
Age [< 35]	ref.							
Age [35-60]	**0.62**	0.012	**0.63**	0.013	**0.62**	0.013	**0.63**	0.013
Age [60+]	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001
HH_num [>3]	ref.							
HH_num [3+]	**0.69**	0.021	**0.69**	0.022	**0.69**	0.022	**0.69**	0.022
HH_inc [<28k]	ref.							
HH_inc [28-75k]	1.26	0.188	1.27	0.180	1.27	0.178	1.27	0.176
HH_inc [75k+]	**1.64**	0.036	**1.65**	0.035	**1.65**	0.035	**1.65**	0.034
Edu_lev [Dgr_N]	ref.							
Edu_lev [Dgr_Y]	0.80	0.332	0.80	0.332	0.80	0.332	0.80	0.333
Gender [Fem]	ref.							
Gender [Mal]	**0.63**	0.012	**0.62**	0.014	0.70	0.055	0.73	0.083
HH_prop [rent]	ref.							
HH_prop [own]	0.82	0.257	0.82	0.245	0.82	0.255	0.82	0.255
Interactions								
Inv_cost (k€) *Gender [Mal]	**1.05**	0.016						
Cost_sav (k€) *Gender [Mal]			**1.62**	0.025				
Emis_sav(tCO_2) *Gender [Mal]					1.14	0.144		
SSuff (%) *Gender [Mal]							1.04	0.251
Random Effects								
σ^2	3.29		3.29		3.29		3.29	
τ_{00}	1.42		1.42		1.42		1.42	
ICC	0.30		0.30		0.30		0.30	
Evaluation metrics								
AIC	6075		6076		6079		6080	
BIC	6177		6178		6181		6182	
Marginal R^2 / Conditional R^2	0.591 / 0.714		0.590 / 0.713		0.590 / 0.713		0.590 / 0.713	

Table D.7 Interaction between education level and REC performance on the association with the odd to accept

Predictors	OR	p	OR	p	OR	p	OR	p
Numerical								
Intercept	0.73	0.280	0.76	0.351	0.71	0.221	0.74	0.286

Inv_cost (k€)	**0.73**	<0.001	**0.75**	<0.001	**0.75**	<0.001	**0.75**	<0.001
Cost_sav (k€)	**4.57**	<0.001	**3.46**	<0.001	**4.57**	<0.001	**4.49**	<0.001
Emis_sav(tCO_2)	**1.66**	<0.001	**1.66**	<0.001	**1.55**	<0.001	**1.67**	<0.001
SSuff (%)	**1.42**	<0.001	**1.42**	<0.001	**1.42**	<0.001	**1.37**	<0.001
Categorical								
Age [< 35]	ref.							
Age [35-60]	**0.62**	0.012	**0.62**	0.012	**0.62**	0.012	**0.62**	0.013
Age [60+]	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001	**0.38**	<0.001
HH_num [>3]	ref.							
HH_num [3+]	**0.51**	0.001	**0.44**	0.001	**0.55**	0.002	**0.48**	<0.001
HH_inc [<28k]	ref.							
HH_inc [28-75k]	1.27	0.182	1.27	0.181	1.27	0.179	1.27	0.179
HH_inc [75k+]	**1.64**	0.036	**1.65**	0.035	**1.65**	0.035	**1.65**	0.034
Edu_lev [Dgr_N]	ref.							
Edu_lev [Dgr_Y]	0.80	0.319	0.80	0.328	0.80	0.332	0.80	0.338
Gender [Fem]	ref.							
Gender [Mal]	0.82	0.176	**0.82**	0.190	0.82	0.193	0.83	0.207
HH_prop [rent]	ref.							
HH_prop [own]	0.82	0.263	0.82	0.260	0.82	0.257	0.82	0.253
Interactions								
Inv_cost (k€) * HH_num [3+]	**1.05**	0.006						
Cost_sav (k€) * HH_num [3+]			**2.09**	0.001				
Emis_sav(tCO_2) * HH_num [3+]					**1.21**	0.037		
SSuff (%) * HH_num [3+]							**1.11**	0.001
Random Effects								
σ^2	3.29		3.29		3.29		3.29	
τ_{00}	1.41		1.41		1.41		1.41	
ICC	0.30		0.30		0.30		0.30	
Evaluation metrics								
AIC	6073		6070		6077		6070	
BIC	6175		6172		6179		6172	
Marginal R^2 / Conditional R^2	0.590 / 0.713		0.590 / 0.713		0.589 / 0.713		0.591 / 0.714	

E. Socio-demographic distribution of the sample

The following figures show the distribution of socio-demographic characteristics of the 349 restrained respondents to the questionnaire living in Turin is compared to the marginal distribution in the city according to the last census survey.

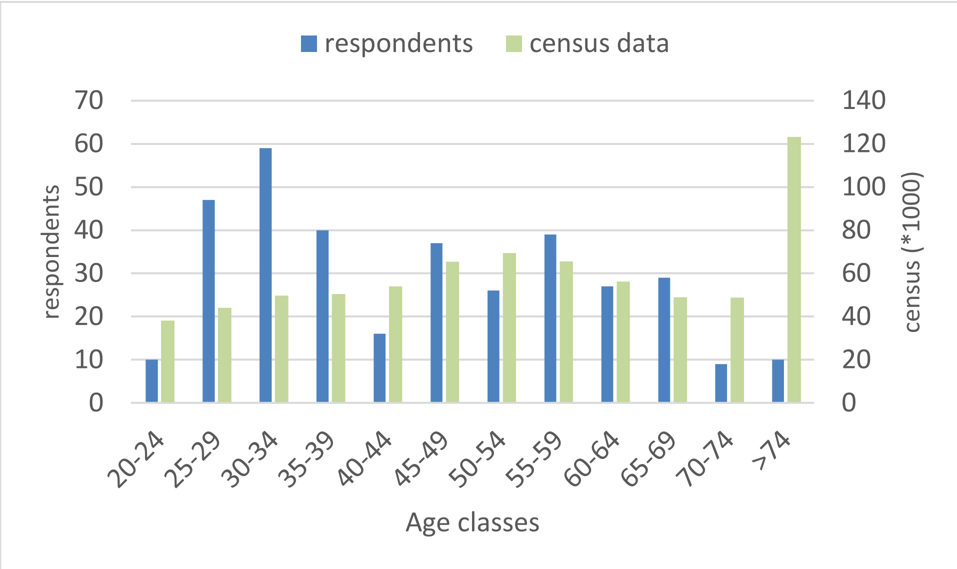


Figure E.1 Age class distribution comparison between respondents and census data. No respondents younger than 20 years old responded to the questionnaire

Table E.1 Age class distribution comparison between respondents and census data in absolute values and percentage.

	Respondents		Census data	
Age	number	%	number	%
20-24	10	2.9	38157	5.3
25-29	47	13.5	44023	6.2
30-34	59	16.9	49686	7.0
35-39	40	11.5	50410	7.1
40-44	16	4.6	54040	7.6
45-49	37	10.6	65442	9.2
50-54	26	7.4	69388	9.7
55-59	39	11.2	65547	9.2
60-64	27	7.7	56168	7.9
65-69	29	8.3	48894	6.9
70-74	9	2.6	48819	6.8
>74	10	2.9	123170	17.3

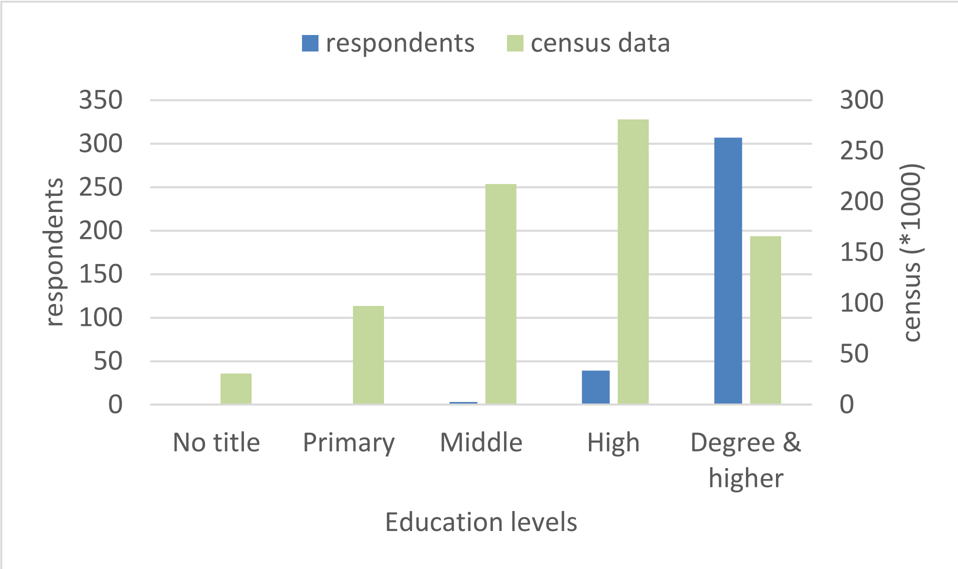


Figure E.2 Education level distribution comparison between respondents and census data. In the questionnaire the respondents were able to differentiate between holding a degree and higher education level. These respondents are aggregated coherently with the available marginal data provided by the census survey

Table E.2 Education level distribution comparison between respondents and census data in absolute values and percentage.

	Respondents		Census data	
Education level	number	%	number	%
No title	0	0.0	30826	3.9
Primary	0	0.0	97165	12.3
Middle	3	0.9	217537	27.4
High	39	11.2	281335	35.5
Degree & higher	307	88.0	165934	20.9

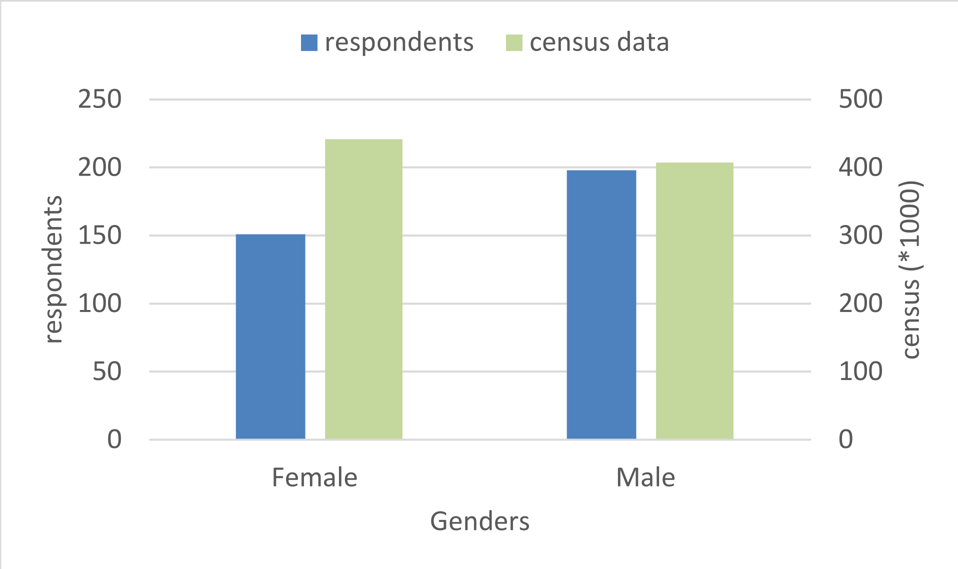


Figure E.3 Gender distribution comparison between respondents and census data.

Table E.3 Gender distribution comparison between respondents and census data in absolute values and percentage.

	Respondents		Census data	
Gender	number	%	number	%
Female	151	43.3	441686	52.0
Male	198	56.7	407062	48.0

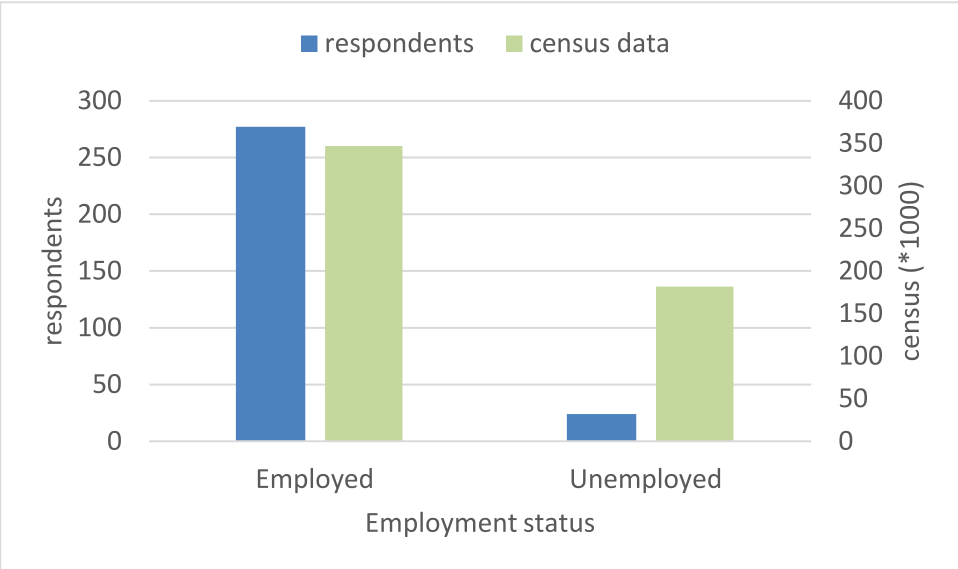


Figure E.4 Employment status distribution comparison between respondents and census data. The census data are provided for individuals in a age range from 15 to 64 years old. For the sake of comparability, the data for respondents have been truncated to eliminate respondents older than 64 years old.

Table E.4 Employment status distribution comparison between respondents and census data in absolute values and percentage.

Employment status	Respondents		Census data	
(15-64 y.o.)	number	%	number	%
Employed	277	92.0	346734	65.6
unemployed	24	8.0	181778	34.4

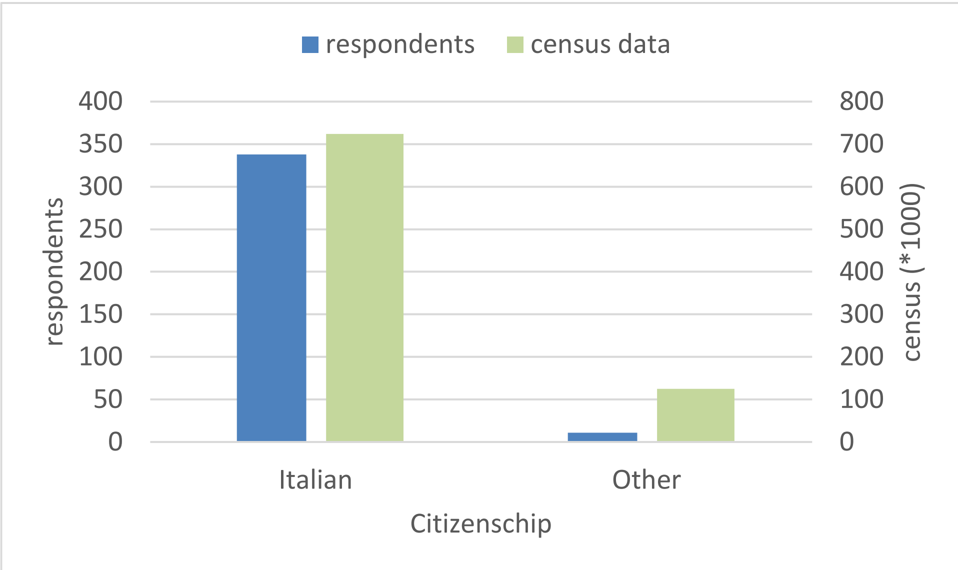


Figure E.5 Citizen distribution comparison between respondents and census data.

Table E.5 Citizen distribution comparison between respondents and census data in absolute values and percentage.

	Respondents		Census data	
Citizenship	number	%	number	%
Italian	338	96.8	724163	85.3
Other	11	3.2	124585	14.7

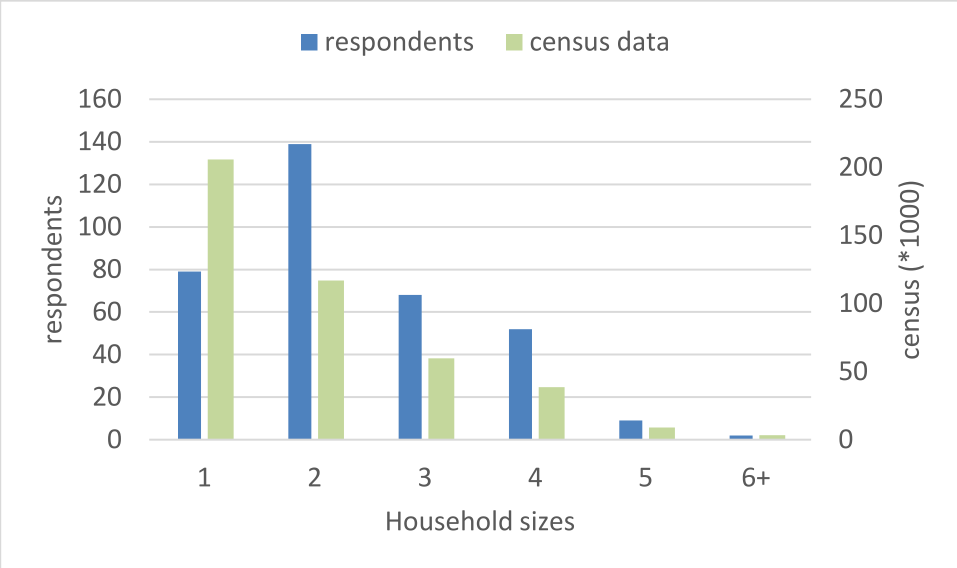


Figure E.6 Households size distribution comparison between respondents and census data.

Table E.6 Households size distribution comparison between respondents and census data in absolute values and percentage.

	Respondents		Census data	
HH_size	number	%	number	%
1	79	22.6	205865	47.5
2	139	39.8	117005	27.0
3	68	19.5	59793	13.8
4	52	14.9	38548	8.9
5	9	2.6	8996	2.1
6+	2	0.6	3262	0.8

The comparisons between socio-demographic characteristics collected with the survey and the census data demonstrate the non-representativeness of the collected sample.

F. LMM models

In the following tables the preparatory analysis to the definition of the Linear mixed model are presented

Table F.1 Association between socio-demographic attributes and investment cost with marginal utility

Predictors	β	CI 95%	p
Numerical			
Intercept	0.33	0.28/0.37	<0.001
Inv_cost (10k€)	**-0.16**	-0.17/-0.01	<0.001
Categorical			
Age [< 35]	ref.		
Age [35-60]	-0.01	-0.06/0.01	0.228
Age [60+]	**-0.03**	-0.06/0.00	0.041
HH_num [>3]	ref.		
HH_num [3+]	-0.01	-0.04/0.01	0.274
HH_inc [<28k]	ref.		
HH_inc [28-75k]	0.00	-0.02/0.03	0.807
HH_inc [75k+]	-0.02	-0.05/0.02	0.334
Edu_lev [Dgr_N]	ref.		
Edu_lev [Dgr_Y]	-0.02		0.298
Gender [Fem]	ref.		
Gender [Mal]	0.00	-0.02/0.02	0.921
HH_prop [rent]	ref.		
HH_prop [own]	-0.01	-0.04/0.01	0.287
Random Effects			
σ^2	0.01		
τ_{00}	0.01		
ICC	0.49		
Evaluation metrics			
AIC	-10,083		
BIC	-10,002		
Marginal R^2 / Conditional R^2	0.276 / 0.629		
VIF	[1.00;1.62]		

Table F.2 Association between socio-demographic attributes and investment cost with marginal utility, with moderating effects included.

Predictors	β	CI 95%	p	β	CI 95%	p
Numerical						
Intercept	**0.40**	0.35/0.44	<0.001	**0.38**	0.35/0.41	<0.001
Inv_cost (10k€)	**-0.23**	-0.24/-0.21	<0.001	**-0.22**	-0.24/-0.21	<0.001
Categorical						
Age [< 35]	ref.					

Age [35-60]	-0.03	-0.07/0.01	0.153	-0.02	-0.06/0.01	0.214
Age [60+]	**-0.05**	-0.08/-0.01	0.005	**-0.04**	-0.08/-0.01	0.006
HH_num [>3]	ref.					
HH_num [3+]	**-0.05**	-0.08/-0.02	<0.001	**-0.05**	-0.08/-0.03	<0.001
HH_inc [<28k]	ref.					
HH_inc [28-75k]	0.00	-0.03/0.03	0.808	0.00	-0.03/0.03	0.943
HH_inc [75k+]	**-0.05**	-0.09/-0.01	0.015	**-0.05**	-0.09/-0.01	0.008
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	-0.02	-0.06/0.02	0.270			
Gender [Fem]	ref.					
Gender [Mal]	-0.02	-0.04/0.01	0.208	-0.01	-0.04/0.01	0.240
HH_prop [rent]	ref.					
HH_prop [own]	**-0.06**	-0.09/-0.03	<0.001	**-0.06**	-0.09/-0.03	<0.001
Interactions						
Inv_cost (k€)						
*Age [35-60]	**0.01**	0.00/0.03	0.026	**0.01**	0.00/0.03	0.028
*Age [60+]	0.01	-0.01/0.02	0.429	0.01	-0.01/0.02	0.464
*HH_num [3+]	**0.03**	0.02/0.04	<0.001	**0.03**	0.02/0.04	<0.001
*HH_inc [28-75k]	-0.00	-0.01/0.01	0.957	0.00	-0.01/0.01	0.995
*HH_inc [75k+]	**0.03**	0.01/0.04	<0.001	**0.03**	0.01/0.04	<0.001
*Edu_lev [Dgr_Y]	0.00	-0.01/0.02	0.702			
*Gender [Mal]	**0.02**	0.01/0.02	0.002	**0.02**	0.01/0.02	0.002
*HH_prop [own]	**0.04**	0.03/0.05	<0.001	**0.04**	0.03/0.05	<0.001
Random Effects						
σ^2	0.01			0.01		
τ_{00}	0.01			0.01		
ICC	0.50			0.50		
Evaluation metrics						
AIC	-10,219			-10,236		
BIC	-10,083			-10,114		
Marginal R^2 / Conditional R^2	0.288 / 0.641			0.287 / 0.641		

Table F.3 Association between socio-demographic attributes and cost savings with marginal utility

Predictors	β	CI 95%	p
Numerical			
Intercept	**-0.05**	-0.09/-0.01	0.007
Cost_sav (k€)	**0.11**	0.10/0.11	<0.001
Categorical			
Age [< 35]	ref.		
Age [35-60]	-0.00	-0.03/0.02	0.726
Age [60+]	-0.01	-0.04/0.02	0.616
HH_num [>3]	ref.		
HH_num [3+]	0.01	-0.01/0.03	0.312
HH_inc [<28k]	ref.		

HH_inc [28-75k]	0.01	-0.01/0.04	0.360
HH_inc [75k+]	**0.05**	0.01/0.08	0.007
Edu_lev [Dgr_N]	ref.		
Edu_lev [Dgr_Y]	-0.00	-0.03/0.03	0.958
Gender [Fem]	ref.		
Gender [Mal]	0.01	-0.01/0.03	0.555
HH_prop [rent]	ref.		
HH_prop [own]	0.01	-0.01/0.04	0.272
Random Effects			
σ^2	0.01		
τ_{00}	0.01		
ICC	0.59		
Evaluation metrics			
AIC	-13,766		
BIC	-13,684		
Marginal R^2 / Conditional R^2	0.147 / 0.651		
VIF	[1.00;1.62]		

Table F.4 Association between socio-demographic attributes and cost savings with marginal utility, with moderating effects included.

Predictors	β	CI 95%	p	β	CI 95%	p
Numerical						
Intercept	-0.00	-0.04/0.04	0.935	-0.00	-0.03/0.03	0.982
Cost_sav (k€)	**0.05**	0.04/0.07	<0.001	**0.05**	0.04/0.06	<0.001
Categorical						
Age [< 35]	ref.					
Age [35-60]	-0.00	-0.03/0.02	0.763			
Age [60+]	-0.01	-0.05/0.02	0.450			
HH_num [>3]	ref.					
HH_num [3+]	-0.01	-0.03/0.02	0.497	-0.01	-0.03/0.02	0.537
HH_inc [<28k]	ref.					
HH_inc [28-75k]	-0.01	-0.04/0.02	0.483	-0.01	-0.04/0.02	0.414
HH_inc [75k+]	-0.03	-0.07/0.00	0.085	**-0.03**	-0.07/-0.00	0.042
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	0.00	-0.03/0.04	0.872			
Gender [Fem]	ref.					
Gender [Mal]	-0.00	-0.03/0.02	0.752	-0.00	-0.03/0.02	0.672
HH_prop [rent]	ref.					
HH_prop [own]	-0.01	-0.01/0.02	0.655	-0.01	-0.03/0.02	0.461
Interactions						
Cost_sav (k€)						
*Age [35-60]	-0.00	-0.01/0.01	0.946			
*Age [60+]	0.01	-0.01/0.02	0.427			
*HH_num [3+]	**0.02**	0.01/0.03	<0.001	**0.02**	0.01/0.03	<0.001
*HH_inc [28-75k]	**0.02**	0.01/0.03	<0.001	**0.02**	0.01/0.03	<0.001

*HH_inc [75k+]	**0.08**	0.07/0.10	<0.001	**0.08**	0.07/0.10	<0.001
*Edu_lev [Dgr_Y]	-0.00	-0.02/0.01	0.562			
*Gender [Mal]	**0.01**	0.00/0.02	0.017	**0.01**	0.00/0.02	0.010
*HH_prop [own]	**0.02**	0.01/0.03	<0.001	**0.02**	0.01/0.03	<0.001
Random Effects						
σ^2	0.01			0.01		
τ_{00}	0.01			0.01		
ICC	0.60			0.60		
Evaluation metrics						
AIC	-13,984			-14,039		
BIC	-13,848			-13,944		
Marginal R^2 / Conditional R^2	0.162 / 0.668			0.163 / 0.666		

Table F.5 Association between socio-demographic attributes and emission savings with marginal utility

Predictors	β	CI 95%	p
Numerical			
Intercept	-0.01	-0.04/0.03	0.752
Emis_sav(tCO_2)	**0.07**	0.07/0.07	<0.001
Categorical			
Age [< 35]	ref.		
Age [35-60]	0.00	-0.02/0.02	0.956
Age [60+]	-0.01	-0.04/0.02	0.357
HH_num [>3]	ref.		
HH_num [3+]	-0.01	-0.03/0.01	0.232
HH_inc [<28k]	ref.		
HH_inc [28-75k]	**0.02**	0.00/0.05	0.047
HH_inc [75k+]	0.02	-0.01/0.05	0.188
Edu_lev [Dgr_N]	ref.		
Edu_lev [Dgr_Y]	0.00	-0.03/0.03	0.880
Gender [Fem]	ref.		
Gender [Mal]	-0.01	-0.02/0.01	0.551
HH_prop [rent]	ref.		
HH_prop [own]	**-0.03**	-0.05/-0.01	0.006
Random Effects			
σ^2	0.00		
τ_{00}	0.01		
ICC	0.63		
Evaluation metrics			
AIC	-15,976		
BIC	-15,894		
Marginal R^2 / Conditional R^2	0.241 / 0.721		
VIF	[1.00;1.62]		

Table F.6 Association between socio-demographic attributes and emission savings with marginal utility, with moderating effects included.

Predictors	β	CI 95%	p	β	CI 95%	p
Numerical						
Intercept	-0.01	-0.05/0.03	0.539	-0.02	-0.05/0.02	0.341
Emis_sav(tCO_2)	**0.07**	0.06/0.08	<0.001	**0.07**	**0.06/0.08**	<0.001
Categorical						
Age [< 35]	ref.					
Age [35-60]	-0,01	-0.03/0.02	0.582			
Age [60+]	-0.01	-0.04/0.02	0.585			
HH_num [>3]	ref.					
HH_num [3+]	0.00	-0.02/0.02	0.912	0.00	-0.02/0.02	0.942
HH_inc [<28k]	ref.					
HH_inc [28-75k]	0.01	-0.02/0.03	0.554	0.01	-0.02/0.03	0.633
HH_inc [75k+]	-0.01	-0.04/0.02	0.659	-0.01	-0.04/0.02	0.498
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	-0.01	-0.04/0.02	0.373	-0.01	-0.04/0.02	0.428
Gender [Fem]	ref.					
Gender [Mal]	-0.00	-0.02/0.02	0.804			
HH_prop [rent]	ref.					
HH_prop [own]	0.01	-0.01/0.03	0.385	0.01	-0.01/0.03	0.491
Interactions						
Cost_sav (k€)						
*Age [35-60]	0.00	-0.00/0.01	0.055			
*Age [60+]	-0.00	-0.01/0.00	0.295			
*HH_num [3+]	**-0.01**	-0.01/-0.00	<0.001	**-0.01**	-0.01/-0.00	<0.002
*HH_inc [28-75k]	**0.01**	0.00/0.01	<0.001	**0.01**	0.00/0.01	<0.001
*HH_inc [75k+]	**0.02**	0.01/0.02	<0.001	**0.01**	0.00/0.01	<0.001
*Edu_lev [Dgr_Y]	**0.01**	0.00/0.02	0.001	**0.01**	0.01/0.02	<0.001
*Gender [Mal]	-0.00	-0.01/0.00	0.312			
*HH_prop [own]	**-0.03**	-0.03/-0.02	<0.001	**-0.02**	-0.03/-0.02	<0.001
Random Effects						
σ^2	0.00			0.00		
τ_{00}	0.01			0.01		
ICC	0.64			0.64		
Evaluation metrics						
AIC	-16,070			-16,121		
BIC	-15,934			-16,026		
Marginal R^2 / Conditional R^2	0.249 / 0.729			0.247 / 0.728		

Table F.7 Association between socio-demographic attributes and self-sufficiency with marginal utility

Predictors	β	CI 95%	p
Numerical			
Intercept	0.04	-0.05/0.12	0.408

SSuff (%)	**0.04**	0.04/0.04	<0.001
Categorical			
Age [< 35]	ref.		
Age [35-60]	0.02	-0.04/0.08	0.548
Age [60+]	0.03	-0.04/0.11	0.381
HH_num [>3]	ref.		
HH_num [3+]	**0.06**	0.01/0.11	0.030
HH_inc [<28k]	ref.		
HH_inc [28-75k]	-0.04	-0.10/0.02	0.178
HH_inc [75k+]	-0.06	-0.13/0.01	0.137
Edu_lev [Dgr_N]	ref.		
Edu_lev [Dgr_Y]	-0.01	-0.08/0.07	0.886
Gender [Fem]	ref.		
Gender [Mal]	0.02	-0.03/0.07	0.399
HH_prop [rent]	ref.		
HH_prop [own]	**0.08**	0.03/0.14	0.004
Random Effects			
σ^2	0.01		
τ_{00}	0.05		
ICC	0.81		
Evaluation metrics			
AIC	-9185		
BIC	-9104		
Marginal R^2	/0.164 /		
Conditional R^2	0.838		
VIF	[1,00;1.62]		

Table F.8 Association between socio-demographic attributes and emission savings with marginal utility, with moderating effects included.

Predictors	β	CI 95%	p	β	CI 95%	p
Numerical						
Intercept	0.07	-0.02/0.16	0.123	0.07	0.01/0.13	0.016
Emis_sav(tCO_2)	**0.03**	0.03/0.03	<0.001	**0.03**	0.03/0.03	<0.001
Categorical						
Age [< 35]	ref.					
Age [35-60]	0.02	-0.04/0.08	0.596			
Age [60+]	0.03	-0.05/0.10	0.498			
HH_num [>3]	ref.					
HH_num [3+]	0.03	-0.02/0.08	0.242	0.03	-0.02/0.08	0.208
HH_inc [<28k]	ref.					
HH_inc [28-75k]	-0.02	-0.08/0.04	0.482	-0.02	-0.07/0.04	0.538
HH_inc [75k+]	-0.03	-0.11/0.05	0.459	-0.02	-0.09/0.05	0.556
Edu_lev [Dgr_N]	ref.					
Edu_lev [Dgr_Y]	-0.00	-0.08/0.07	0.947			
Gender [Fem]	ref.					
Gender [Mal]	0.01	-0.04/0.06	0.643	0.01	-0.03/0.06	0.579

HH_prop [rent]	ref.					
HH_prop [own]	0.04	-0.02/0.09	0.225	0.04	-0.01/0.10	0.113
Interactions						
Cost_sav (k€)						
*Age [35-60]	0.00	-0.00/0.00	0.737			
*Age [60+]	0.00	-0.00/0.01	0.303			
*HH_num [3+]	**0.01**	0.00/0.01	<0.001	**0.01**	0.00/0.01	<0.001
*HH_inc [28-75k]	**-0.00**	-0.01/-0.00	0.001	**-0.00**	-0.01/0.00	0.001
*HH_inc [75k+]	**-0.01**	-0.01/-0.00	<0.001	**-0.01**	-0.01/-0.00	<0.001
*Edu_lev [Dgr_Y]	-0.00	-0.00/0.00	0.676			
*Gender [Mal]	**0.00**	0.00/0.00	0.040	**0.00**	0.00/0.00	0.025
*HH_prop [own]	**0.01**	0.01/0.01	<0.001	**0.01**	0.01/0.01	<0.001
Random Effects						
σ^2	0.01			0.01		
τ_{00}	0.05			0.05		
ICC	0.81			0.81		
Evaluation metrics						
AIC	-9235			-9293		
BIC	-9099			-9198		
Marginal R^2 / Conditional R^2	0.168 / 0.842			0.167 / 0.841		

G. Households and person seeds generation

```
CE="AVQ_2021.xlsx"
data=pd.read_excel(CE)
Piemonte=data.loc[data['REGMf'] == 10]
Piemonte.to_excel("AVQ_Piemonte_2021.xlsx")
Piemonte=pd.read_excel("AVQ_Piemonte_2021.xlsx")

# retained variables
Piemonte_sub=Piemonte[["PROFAM", "NCOMP", "RELPAR", "ETAMi",
"SESSO", "STCIVMi", "CONDMi", "POSIZMi", "COEFIN", "CITTMi"]]
Piemonte_struct=Piemonte_sub.drop(columns=["RELPAR", "ETAMi",
"SESSO", "STCIVMi", "CONDMi", "POSIZMi",
CITTMi"]).groupby(by="PROFAM").mean()

# max number of component in household
max_HH = Piemonte_sub["NCOMP"].max()

# extract the list of households ID
ID_famiglie=Piemonte_sub['PROFAM'].drop_duplicates().to_list()

dummy=pd.DataFrame()
for h in ID_famiglie:
    HH=h
    pro_test=Piemonte_sub.loc[Piemonte_sub['PROFAM'] == HH]
    d=pd.DataFrame([[0]],index=[HH],columns=["NaN"])
    df_test = d.copy(deep=True)
    len(pro_test)
    for i in range(0,len(pro_test)):
        prs=pro_test.iloc[i:i+1]
        v_1, v_2, v_3, v_4, v_5, v_6, v_7 = prs.iat[0,2], prs.iat[0,3],
prs.iat[0,4],
            prs.iat[0,5], prs.iat[0,6], prs.iat[0,7], prs.iat[0,9]
        var_1, var_2, var_3, var_4, var_5, var_6,
var_7="RELPAR_P"+str(i+1),
            "ETAMi_P"+str(i+1), "SESSO_P"+str(i+1),
        "STCIVMi_P"+str(i+1),
```

```
                        "CONDMi"+str(i+1), "POSIZMi"+str(i+1),
                        "CITTMi"+str(i+1)
            df_test[[var_1, var_2, var_3, var_4, var_5, var_6, var_7]]=[v_1,
v_2, v_3, v_4,
            v_5, v_6, v_7]
            dummy = pd.concat([dummy,df_test])

# reconstructed households
HH = pd.merge(Piemonte_struct.reset_index(),
dummy.reset_index().rename(columns =
{'index':'PROFAM'}).drop(columns=["NaN"]), how='left', on='PROFAM')
HH.to_excel("/content/gdrive/My
Drive/POPSIM/elaborazioni/SEED_HH.xlsx")
```

H. Inflation of the dataset

In this appendix an inflation of the original dataset has been performed to test for improvements in the performance of the supervised classification algorithm.
It is worth noting that this process is NOT advised in real case situations, therefore this part has to be considered only a speculation to account for the limitation discussed regarding the 10-fold cross validation procedure on the small size dataset available in the application of the method.
In particular, the original dataset has been multiplied 2-, 3-, 5-, 10-, and 20-times to check for improvements in the accuracy of the PART algorithm performance.
In the following Figure H.1 the accuracy results are plotted for different dataset inflation, while in Table H.1 the relative confusion matrices are displayed.

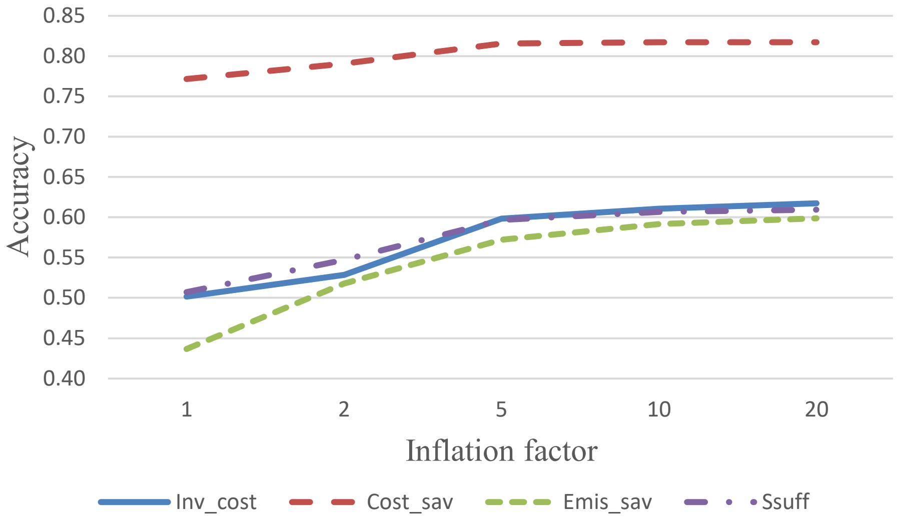


Figure H.1 accuracy results are plotted for different dataset inflation

Table H.1 Confusion matrices for different inflation factors of the original dataset

	Dataset inflation																			
	*1				*2				*5				*10				*20			
Investment cost	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3
	1	243	180	87	1	491	312	165	1	1296	566	303	1	2699	1119	637	1	5586	2320	1272
	2	185	282	63	2	351	601	120	2	822	1702	313	2	1521	3419	586	2	2844	6751	1136
	3	4	3	0	3	22	17	15	3	42	57	134	3	100	112	277	3	210	229	592
Cost savings	n	1	2		n	1	2		n	1	2		n	1	2		n	1	2	
	1	769	207		1	1536	372		1	3890	849		1	7750	1652		1	15522	3323	
	2	32	39		2	66	120		2	115	381		2	260	808		2	498	1597	

Emission savings	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3
	1	60	42	92	1	199	67	166	1	608	147	351	1	1262	276	596	1	2501	560	1110
	2	55	55	58	2	103	161	94	2	222	503	225	2	443	1034	427	2	916	2128	823
	3	203	140	342	3	334	246	724	3	760	535	1884	3	1475	1060	3897	3	2943	2052	7907
Self-sufficienc	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3	n	1	2	3
	1	440	233	147	1	864	397	237	1	2315	933	577	1	4666	1800	1142	1	9232	3522	2180
	2	84	85	27	2	176	233	74	2	333	674	187	2	643	1414	385	2	1342	2912	801
	3	16	9	6	3	40	24	49	3	52	28	136	3	91	56	273	3	226	106	619

References

[1] European Commission. The European Green Deal [COM(2019) 640 final COMMUNICATION]. vol. 53. 2019.

[2] Schiera DS, Minuto FD, Bottaccioli L, Borchiellini R, Lanzini A. Analysis of Rooftop Photovoltaics Diffusion in Energy Community Buildings by a Novel GIS- and Agent-Based Modeling Co-Simulation Platform. IEEE Access 2019;7. https://doi.org/10.1109/ACCESS.2019.2927446.

[3] Cappellaro F, D'Agosta G, De Sabbata P, Barroco F, Carani C, Borghetti A, et al. Implementing energy transition and SDGs targets throughout energy community schemes. Journal of Urban Ecology 2022;8. https://doi.org/10.1093/jue/juac023.

[4] European Commission. EU missions. 100 climate-neutral and smart cities. 2022.

[5] Suppa AR, Cavana G, Binda T. Supporting the EU Mission "100 Climate-Neutral Cities by 2030": A Review of Tools to Support Decision-Making for the Built Environment at District or City Scale. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13380 LNCS, 2022, p. 151–68. https://doi.org/10.1007/978-3-031-10542-5_11.

[6] Buildings Performance Institute Europe (BPIE). Europe's buildings under the microscope. 2011.

[7] European Commission. A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives [COM(2020) 662 final COMMUNICATION]. 2020.

[8] European Commission. COM(2021) 802 final. Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the energy performance of buildings (recast) 2021.

[9] Petersdorff C, Boermans T, Harnisch J. Mitigation of CO2 Emissions from the EU-15 Building Stock. Beyond the EU Directive on the Energy Performance of Buildings (9 pp). Environ Sci Pollut Res Int 2006;13:350–8. https://doi.org/10.1065/espr2005.12.289.

[10] European Parliament. Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings (recast) 2024.

[11] European Commission. for a Directive of the European Parliament and of the Council on the energy performance of buildings (recast) (COM/2021/802 final). Brussels. n.d.
[12] Pavani P, Bak-Jensen B, Pillai JR. Maximizing the self-consumption of Solar-PV using Battery Energy Storage System in Samso-Marina. Proceedings of 2019 IEEE PES Innovative Smart Grid Technologies Europe, ISGT-Europe 2019, 2019. https://doi.org/10.1109/ISGTEurope.2019.8905498.
[13] Gagliano A, Nocera F. Analysis of the performances of electric energy storage in residential applications. International Journal of Heat and Technology 2017;35. https://doi.org/10.18280/ijht.35Sp0106.
[14] Struth J, Leuthold M, Aretz A, Bost M. PV-Benefit: a critical review of the effect of grid integrated PV-storage-systems. 8th International Renewable Energy Storage Conference and Exhibition 2013.
[15] Perea-Moreno MA, Hernandez-Escobedo Q, Perea-Moreno AJ. Renewable energy in urban areas: Worldwide research trends. Energies (Basel) 2018;11. https://doi.org/10.3390/en11030577.
[16] Camilo FM, Castro R, Almeida ME, Pires VF. Economic assessment of residential PV systems with self-consumption and storage in Portugal. Solar Energy 2017;150. https://doi.org/10.1016/j.solener.2017.04.062.
[17] Urge-Vorsatz D, Khosla R, Bernhardt R, Chan YC, Verez D, Hu S, et al. Advances toward a net-zero global building sector. Annu Rev Environ Resour 2020;45. https://doi.org/10.1146/annurev-environ-012420-045843.
[18] Calise F, D'Accadia MD, Barletta C, Battaglia V, Pfeifer A, Duic N. Detailed modelling of the deep decarbonisation scenarios with demand response technologies in the heating and cooling sector: A case study for Italy. Energies (Basel) 2017;10. https://doi.org/10.3390/en10101535.
[19] Leibowicz BD, Lanham CM, Brozynski MT, Vázquez-Canteli JR, Castejón NC, Nagy Z. Optimal decarbonization pathways for urban residential building energy services. Appl Energy 2018;230. https://doi.org/10.1016/j.apenergy.2018.09.046.
[20] Viesi D, Crema L, Mahbub MS, Verones S, Brunelli R, Baggio P, et al. Integrated and dynamic energy modelling of a regional system: A cost-optimized approach in the deep decarbonisation of the Province of Trento (Italy). Energy 2020;209. https://doi.org/10.1016/j.energy.2020.118378.
[21] Grillone B, Danov S, Sumper A, Cipriano J, Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renewable and Sustainable Energy Reviews 2020;131. https://doi.org/10.1016/j.rser.2020.110027.
[22] European Commission. COM(2021) 802 final. Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the energy performance of buildings (recast) 2021.

[23] DIRECTIVE (EU) 2018/2001. Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources (recast) 2018.
[24] Marique AF, Reiter S. A simplified framework to assess the feasibility of zero-energy at the neighbourhood/community scale. Energy Build 2014;82. https://doi.org/10.1016/j.enbuild.2014.07.006.
[25] Resch M, Bühler J, Schachler B, Kunert R, Meier A, Sumper A. Technical and economic comparison of grid supportive vanadium redox flow batteries for primary control reserve and community electricity storage in Germany. Int J Energy Res 2019;43. https://doi.org/10.1002/er.4269.
[26] Barbour E, Parra D, Awwad Z, González MC. Community energy storage: A smart choice for the smart grid? Appl Energy 2018;212. https://doi.org/10.1016/j.apenergy.2017.12.056.
[27] Mansó Borràs I, Neves D, Gomes R. Using urban building energy modeling data to assess energy communities' potential. Energy Build 2023;282. https://doi.org/10.1016/j.enbuild.2023.112791.
[28] Todeschi V, Marocco P, Mutani G, Lanzini A, Santarelli M. Towards energy self-consumption and self-sufficiency in urban energy communities. International Journal of Heat and Technology 2021;39. https://doi.org/10.18280/ijht.390101.
[29] Pastore LM, Lo Basso G, Ricciardi G, de Santoli L. Synergies between Power-to-Heat and Power-to-Gas in renewable energy communities. Renew Energy 2022;198. https://doi.org/10.1016/j.renene.2022.08.141.
[30] Li M, Cao S, Zhu X, Xu Y. Techno-economic analysis of the transition towards the large-scale hybrid wind-tidal supported coastal zero-energy communities. Appl Energy 2022;316. https://doi.org/10.1016/j.apenergy.2022.119118.
[31] Ruggiero S, Busch H, Hansen T, Isakovic A. Context and agency in urban community energy initiatives: An analysis of six case studies from the Baltic Sea Region. Energy Policy 2021;148. https://doi.org/10.1016/j.enpol.2020.111956.
[32] Mehta P, Tiefenbeck V. Solar PV sharing in urban energy communities: Impact of community configurations on profitability, autonomy and the electric grid. Sustain Cities Soc 2022;87. https://doi.org/10.1016/j.scs.2022.104178.
[33] Ancona MA, Baldi F, Branchini L, De Pascale A, Gianaroli F, Melino F, et al. Comparative Analysis of Renewable Energy Community Designs for District Heating Networks: Case Study of Corticella (Italy). Energies (Basel) 2022;15. https://doi.org/10.3390/en15145248.
[34] Kazmi H, Munné-Collado Í, Mehmood F, Syed TA, Driesen J. Towards data-driven energy communities: A review of open-source datasets, models and tools. Renewable and Sustainable Energy Reviews 2021;148. https://doi.org/10.1016/j.rser.2021.111290.

[35] Yaqoot M, Diwan P, Kandpal TC. Review of barriers to the dissemination of decentralized renewable energy systems. Renewable and Sustainable Energy Reviews 2016;58. https://doi.org/10.1016/j.rser.2015.12.224.
[36] Wu Y, Wu Y, Cimen H, Vasquez JC, Guerrero JM. Towards collective energy Community: Potential roles of microgrid and blockchain to go beyond P2P energy trading. Appl Energy 2022;314. https://doi.org/10.1016/j.apenergy.2022.119003.
[37] Zwickl-Bernhard S, Auer H. Citizen participation in low-carbon energy systems: Energy communities and its impact on the electricity demand on neighborhood and national level. Energies (Basel) 2021;14. https://doi.org/10.3390/en14020305.
[38] Ryszawska B, Rozwadowska M, Ulatowska R, Pierzchała M, Szymański P. The power of co-creation in the energy transition—dart model in citizen energy communities projects. Energies (Basel) 2021;14. https://doi.org/10.3390/en14175266.
[39] DeWaters J, Powers S, Graham M. Developing an energy literacy scale. ASEE Annual Conference and Exposition, Conference Proceedings, 2007. https://doi.org/10.18260/1-2--2076.
[40] European Union. Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources (recast) (Text with EEA relevance.) 2018.
[41] European Union. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU (recast) (Text with EEA relevance.) 2019.
[42] Haji Bashi M, De Tommasi L, Le Cam A, Relaño LS, Lyons P, Mundó J, et al. A review and mapping exercise of energy community regulatory challenges in European member states based on a survey of collective energy actors. Renewable and Sustainable Energy Reviews 2023;172. https://doi.org/10.1016/j.rser.2022.113055.
[43] REScoop.eu n.d. https://www.rescoop.eu/ (accessed February 4, 2023).
[44] Bukovszki V, Magyari Á, Braun MK, Párdi K, Reith A. Energy modelling as a trigger for energy communities: A joint socio-technical perspective. Energies (Basel) 2020;13. https://doi.org/10.3390/en13092274.
[45] Sokołowski MM. Renewable and citizen energy communities in the European Union: how (not) to regulate community energy in national laws and policies. Journal of Energy and Natural Resources Law 2020;38. https://doi.org/10.1080/02646811.2020.1759247.
[46] Gerundo R, Marra A. A Decision Support Methodology to Foster Renewable Energy Communities in the Municipal Urban Plan. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su142316268.
[47] Cutore E, Volpe R, Sgroi R, Fichera A. Energy management and sustainability assessment of renewable energy communities: The Italian

context. Energy Convers Manag 2023;278. https://doi.org/10.1016/j.enconman.2023.116713.
[48] Lowitzsch J, Hoicka CE, van Tulder FJ. Renewable energy communities under the 2019 European Clean Energy Package – Governance model for the energy clusters of the future? Renewable and Sustainable Energy Reviews 2020;122. https://doi.org/10.1016/j.rser.2019.109489.
[49] Baigorrotegui G, Lowitzsch J. Institutional aspects of consumer (Co-) ownership in RE energy communities. Energy Transition: Financing Consumer Co-Ownership in Renewables, 2019. https://doi.org/10.1007/978-3-319-93518-8_28.
[50] Chaudhry S, Surmann A, Kühnbach M, Pierie F. Renewable Energy Communities as Modes of Collective Prosumership: A Multi-Disciplinary Assessment Part II—Case Study. Energies (Basel) 2022;15. https://doi.org/10.3390/en15238936.
[51] Torabi Moghadam S, Di Nicoli MV, Manzo S, Lombardi P. Mainstreaming energy communities in the transition to a low-carbon future: A methodological approach. Energies (Basel) 2020;13. https://doi.org/10.3390/en13071597.
[52] Rai V, Robinson SA. Agent-based modeling of energy technology adoption: Empirical integration of social, behavioral, economic, and environmental factors. Environmental Modelling and Software 2015;70. https://doi.org/10.1016/j.envsoft.2015.04.014.
[53] Magnani N, Osti G. Does civil society matter? Challenges and strategies of grassroots initiatives in Italy's energy transition. Energy Res Soc Sci 2016;13. https://doi.org/10.1016/j.erss.2015.12.012.
[54] Mittal A, Krejci CC, Dorneich MC, Fickes D. An agent-based approach to modeling zero energy communities. Solar Energy 2019;191. https://doi.org/10.1016/j.solener.2019.08.040.
[55] 'Hledik R, 'Tsuchida B, 'Palfreyman J. Beyond Zero Net Energy? Alternative Approaches to Enhance Consumer and Environmental Outcomes. 2018.
[56] Koirala BP, Koliou E, Friege J, Hakvoort RA, Herder PM. Energetic communities for community energy: A review of key issues and trends shaping integrated community energy systems. Renewable and Sustainable Energy Reviews 2016;56. https://doi.org/10.1016/j.rser.2015.11.080.
[57] Bucking S. Energy modeling methodology for community master planning. ASHRAE Conference-Papers, vol. 124, 2018.
[58] Gjorgievski VZ, Cundeva S, Georghiou GE. Social arrangements, technical designs and impacts of energy communities: A review. Renew Energy 2021;169. https://doi.org/10.1016/j.renene.2021.01.078.
[59] Afzali P, Rashidinejad M, Abdollahi A, Salehizadeh MR, Farahmand H. A stochastic multi-objective model for energy efficiency and renewable resource planning in energy communities: A sustainably cost-effective trade-

off. IET Renewable Power Generation 2023;17. https://doi.org/10.1049/rpg2.12662.
[60] Ghiani E, Trevisan R, Rosetti GL, Olivero S, Barbero L. Energetic and Economic Performances of the Energy Community of Magliano Alpi after One Year of Piloting. Energies (Basel) 2022;15. https://doi.org/10.3390/en15197439.
[61] Mutani G, Beltramino S, Forte A. A clean energy atlas for energy communities in piedmont region (Italy). International Journal of Design and Nature and Ecodynamics 2020;15. https://doi.org/10.18280/ijdne.150308.
[62] De Santi F, Moncecchi M, Prettico G, Fulli G, Olivero S, Merlo M. To Join or Not to Join? The Energy Community Dilemma: An Italian Case Study. Energies (Basel) 2022;15. https://doi.org/10.3390/en15197072.
[63] Sudhoff R, Schreck S, Thiem S, Niessen S. Operating Renewable Energy Communities to Reduce Power Peaks in the Distribution Grid: An Analysis on Grid-Friendliness, Different Shares of Participants, and Economic Benefits. Energies (Basel) 2022;15. https://doi.org/10.3390/en15155468.
[64] Cirone D, Bruno R, Bevilacqua P, Perrella S, Arcuri N. Techno-Economic Analysis of an Energy Community Based on PV and Electric Storage Systems in a Small Mountain Locality of South Italy: A Case Study. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su142113877.
[65] Zwickl-Bernhard S, Auer H. Open-source modeling of a low-carbon urban neighborhood with high shares of local renewable generation. Appl Energy 2021;282. https://doi.org/10.1016/j.apenergy.2020.116166.
[66] Caballero V, Briones A, Coca-Ortegón A, Pérez A, Barrios B, de la Mano M. Analysis and simulation of an Urban-Industrial Sustainable Energy Community: A use case in San Juan de Mozarrifar using photovoltaic energy. Energy Reports 2023;9. https://doi.org/10.1016/j.egyr.2022.12.059.
[67] Paladin A, Das R, Wang Y, Ali Z, Kotter R, Putrus G, et al. Micro market based optimisation framework for decentralised management of distributed flexibility assets. Renew Energy 2021;163. https://doi.org/10.1016/j.renene.2020.10.003.
[68] Novoa L, Flores R, Brouwer J. Optimal renewable generation and battery storage sizing and siting considering local transformer limits. Appl Energy 2019;256. https://doi.org/10.1016/j.apenergy.2019.113926.
[69] Maleki Delarestaghi J, Arefi A, Ledwich G, Borghetti A. A distribution network planning model considering neighborhood energy trading. Electric Power Systems Research 2021;191. https://doi.org/10.1016/j.epsr.2020.106894.
[70] Simoiu MS, Fagarasan I, Ploix S, Calofir V. Sizing and management of an energy system for a metropolitan station with storage and related district energy community. Energies (Basel) 2021;14. https://doi.org/10.3390/en14185997.
[71] Moretti E, Stamponi E. The Renewable Energy Communities in Italy and the Role of Public Administrations: The Experience of the Municipality of

Assisi between Challenges and Opportunities. Sustainability (Switzerland) 2023;15. https://doi.org/10.3390/su151511869.
[72] Cielo A, Margiaria P, Lazzeroni P, Mariuzzo I, Repetto M. Renewable Energy Communities business models under the 2020 Italian regulation. J Clean Prod 2021;316. https://doi.org/10.1016/j.jclepro.2021.128217.
[73] Abokersh MH, Gangwar S, Spiekman M, Vallès M, Jiménez L, Boer D. Sustainability insights on emerging solar district heating technologies to boost the nearly zero energy building concept. Renew Energy 2021;180. https://doi.org/10.1016/j.renene.2021.08.091.
[74] Ceglia F, Marrasso E, Samanta S, Sasso M. Addressing Energy Poverty in the Energy Community: Assessment of Energy, Environmental, Economic, and Social Benefits for an Italian Residential Case Study. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su142215077.
[75] Kyriakopoulos GL. Energy Communities Overview: Managerial Policies, Economic Aspects, Technologies, and Models. Journal of Risk and Financial Management 2022;15. https://doi.org/10.3390/jrfm15110521.
[76] Efthymiou EN, Yfanti S, Kyriakarakos G, Zervas PL, Langouranis P, Terzis K, et al. A Practical Methodology for Building a Municipality-Led Renewable Energy Community: A Photovoltaics-Based Case Study for the Municipality of Hersonissos in Crete, Greece. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su141912935.
[77] Lowitzsch J. Investing in a Renewable Future – Renewable Energy Communities, Consumer (Co-)Ownership and Energy Sharing in the Clean Energy Package. Renewable Energy Law and Policy Review 2023;9. https://doi.org/10.4337/relp.2019.02.02.
[78] Cavallaro E, Sessa MR, Malandrino O. Renewable Energy Communities in the Energy Transition Context. International Journal of Energy Economics and Policy 2023;13. https://doi.org/10.32479/ijeep.14230.
[79] Parreño-Rodriguez A, Ramallo-González AP, Chinchilla-Sánchez M, Molina-García A. Community energy solutions for addressing energy poverty: A local case study in Spain. Energy Build 2023;296. https://doi.org/10.1016/j.enbuild.2023.113418.
[80] Bartolini A, Carducci F, Muñoz CB, Comodi G. Energy storage and multi energy systems in local energy communities with high renewable energy penetration. Renew Energy 2020;159. https://doi.org/10.1016/j.renene.2020.05.131.
[81] Terrier C, Loustau JRH, Lepour D, Maréchal F. From Local Energy Communities towards National Energy System: A Grid-Aware Techno-Economic Analysis. Energies (Basel) 2024;17:910. https://doi.org/10.3390/en17040910.
[82] van Zalk J, Behrens P. The spatial extent of renewable and non-renewable power generation: A review and meta-analysis of power densities and their application in the U.S. Energy Policy 2018;123. https://doi.org/10.1016/j.enpol.2018.08.023.

[83] Inês C, Guilherme PL, Esther MG, Swantje G, Stephen H, Lars H. Regulatory challenges and opportunities for collective renewable energy prosumers in the EU. Energy Policy 2020;138. https://doi.org/10.1016/j.enpol.2019.111212.
[84] Volpato G, Carraro G, Cont M, Danieli P, Rech S, Lazzaretto A. General guidelines for the optimal economic aggregation of prosumers in energy communities. Energy 2022;258. https://doi.org/10.1016/j.energy.2022.124800.
[85] Nielsen BF, Baer D, Lindkvist C. Identifying and supporting exploratory and exploitative models of innovation in municipal urban planning; key challenges from seven Norwegian energy ambitious neighborhood pilots. Technol Forecast Soc Change 2019;142. https://doi.org/10.1016/j.techfore.2018.11.007.
[86] Caramizaru Aura, Uihlein Andreas. Energy communities : an overview of energy and social innovation. European Commission Joint Research Centre 2020;EUR 30083 EN.
[87] Mihailova D, Schubert I, Martinez-Cruz AL, Hearn AX, Sohre A. Preferences for configurations of Positive Energy Districts – Insights from a discrete choice experiment on Swiss households. Energy Policy 2022;163. https://doi.org/10.1016/j.enpol.2022.112824.
[88] Cavana G, Becchio C, Bottero M. Feasibility and evolution studies on renewable energy communities in cities. Renewable and Sustainable Energy Reviews 2025;213:115477. https://doi.org/10.1016/j.rser.2025.115477.
[89] Teladia A, van der Windt H. Citizen participation gaps and challenges in the heating transition: Learning from Dutch community initiatives. Renewable and Sustainable Energy Reviews 2024;189:113975. https://doi.org/10.1016/j.rser.2023.113975.
[90] Stadler M, Groissböck M, Cardoso G, Marnay C. Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel. Appl Energy 2014;132. https://doi.org/10.1016/j.apenergy.2014.07.041.
[91] Bottero M, Comino E, Duriavig M, Ferretti V, Pomarico S. The application of a Multicriteria Spatial Decision Support System (MCSDSS) for the assessment of biodiversity conservation in the Province of Varese (Italy). Land Use Policy 2013;30. https://doi.org/10.1016/j.landusepol.2012.05.015.
[92] Kim MH, Kim D, Heo J, Lee DW. Techno-economic analysis of hybrid renewable energy system with solar district heating for net zero energy community. Energy 2019;187. https://doi.org/10.1016/j.energy.2019.115916.
[93] Ceglia F, Marrasso E, Roselli C, Sasso M, Coletta G, Pellegrino L. Biomass-Based Renewable Energy Community: Economic Analysis of a Real Case Study. Energies (Basel) 2022;15. https://doi.org/10.3390/en15155655.
[94] Fouad MM, Iskander J, Shihata LA. Energy, carbon and cost analysis for an innovative zero energy community design. Solar Energy 2020;206. https://doi.org/10.1016/j.solener.2020.05.048.

[95] Sougkakis V, Lymperopoulos K, Nikolopoulos N, Margaritis N, Giourka P, Angelakoglou K. An investigation on the feasibility of near-zero and positive energy communities in the greek context. Smart Cities 2020;3. https://doi.org/10.3390/smartcities3020019.
[96] Mutani G, Usta Y. Design and Modeling Renewable Energy Communities: A Case Study in Cagliari (Italy). International Journal of Sustainable Development and Planning 2022;17. https://doi.org/10.18280/ijsdp.170401.
[97] Mutani G, Santantonio S, Beltramino S. Indicators and representation tools to measure the technical-economic feasibility of a renewable energy community. The case study of villar pellice (italy). International Journal of Sustainable Development and Planning 2021;16. https://doi.org/10.18280/ijsdp.160101.
[98] Franzoi N, Prada A, Verones S, Baggio P. Enhancing PV self-consumption through energy communities in heating-dominated climates. Energies (Basel) 2021;14. https://doi.org/10.3390/en14144165.
[99] Vivian J, Chinello M, Zarrella A, De Carli M. Investigation on Individual and Collective PV Self-Consumption for a Fifth Generation District Heating Network. Energies (Basel) 2022;15. https://doi.org/10.3390/en15031022.
[100] Berg K, Stefanussen Foslie S, Farahmand H. Industrial Energy Communities: Energy Storage Investment, Grid Impact and Cost Distribution. SSRN Electronic Journal 2024. https://doi.org/10.2139/ssrn.4703806.
[101] Pastore LM, Lo Basso G, Quarta MN, de Santoli L. Power-to-gas as an option for improving energy self-consumption in renewable energy communities. Int J Hydrogen Energy 2022;47. https://doi.org/10.1016/j.ijhydene.2022.06.287.
[102] Canova A, Lazzeroni P, Lorenti G, Moraglio F, Porcelli A, Repetto M. Decarbonizing residential energy consumption under the Italian collective self-consumption regulation. Sustain Cities Soc 2022;87. https://doi.org/10.1016/j.scs.2022.104196.
[103] Sibilla M, Abanda FH. Multi-Criteria Decision Making Optimisation Framework for Positive Energy Blocks for Cities. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su14010446.
[104] Li X, Chalvatzis KJ, Stephanides P. Innovative energy islands: Life-cycle cost-benefit analysis for battery energy storage. Sustainability (Switzerland) 2018;10. https://doi.org/10.3390/su10103371.
[105] Bordignon S, Quaggiotto D, Vivian J, Emmi G, De Carli M, Zarrella A. A solar-assisted low-temperature district heating and cooling network coupled with a ground-source heat pump. Energy Convers Manag 2022;267. https://doi.org/10.1016/j.enconman.2022.115838.
[106] Zeyad M, Ahmed SMM, Hasan S, Mahmud DM. Community microgrid: an approach towards positive energy community in an urban area of Dhaka, Bangladesh. Clean Energy 2023;7:926–39. https://doi.org/10.1093/ce/zkad027.

[107] Pastore LM. Combining Power-to-Heat and Power-to-Vehicle strategies to provide system flexibility in smart urban energy districts. Sustain Cities Soc 2023;94:104548. https://doi.org/10.1016/j.scs.2023.104548.
[108] Luz GP, E Silva RA. Modeling energy communities with collective photovoltaic self-consumption: Synergies between a small city and a winery in Portugal. Energies (Basel) 2021;14. https://doi.org/10.3390/en14020323.
[109] Cutore E, Fichera A, Volpe R. A Roadmap for the Design, Operation and Monitoring of Renewable Energy Communities in Italy. Sustainability (Switzerland) 2023;15. https://doi.org/10.3390/su15108118.
[110] De Souza R, Nadalon E, Casisi M, Reini M. Optimal Sharing Electricity and Thermal Energy Integration for an Energy Community in the Perspective of 100% RES Scenario. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su141610125.
[111] Fina B, Auer H, Friedl W. Profitability of PV sharing in energy communities: Use cases for different settlement patterns. Energy 2019;189. https://doi.org/10.1016/j.energy.2019.116148.
[112] Rehman H ur, Reda F, Paiho S, Hasan A. Towards positive energy communities at high latitudes. Energy Convers Manag 2019;196. https://doi.org/10.1016/j.enconman.2019.06.005.
[113] Monsberger C, Fina B, Auer H. Profitability of energy supply contracting and energy sharing concepts in a neighborhood energy community: Business cases for austria. Energies (Basel) 2021;14. https://doi.org/10.3390/en14040921.
[114] Liu J, Chen X, Yang H, Shan K. Hybrid renewable energy applications in zero-energy buildings and communities integrating battery and hydrogen vehicle storage. Appl Energy 2021;290. https://doi.org/10.1016/j.apenergy.2021.116733.
[115] Alabi TM, Lu L, Yang Z, Zhou Y. A novel optimal configuration model for a zero-carbon multi-energy system (ZC-MES) integrated with financial constraints. Sustainable Energy, Grids and Networks 2020;23. https://doi.org/10.1016/j.segan.2020.100381.
[116] Comodi G, Bartolini A, Carducci F, Nagaranjan B, Romagnoli A. Achieving low carbon local energy communities in hot climates by exploiting networks synergies in multi energy systems. Appl Energy 2019;256. https://doi.org/10.1016/j.apenergy.2019.113901.
[117] Weckesser T, Dominković DF, Blomgren EMV, Schledorn A, Madsen H. Renewable Energy Communities: Optimal sizing and distribution grid impact of photo-voltaics and battery storage. Appl Energy 2021;301. https://doi.org/10.1016/j.apenergy.2021.117408.
[118] Fleischhacker A, Lettner G, Schwabeneder D, Auer H. Portfolio optimization of energy communities to meet reductions in costs and emissions. Energy 2019;173. https://doi.org/10.1016/j.energy.2019.02.104.

[119] Lee U, Park S, Lee I. Robust design optimization (RDO) of thermoelectric generator system using non-dominated sorting genetic algorithm II (NSGA-II). Energy 2020;196. https://doi.org/10.1016/j.energy.2020.117090.
[120] Molyneaux A, Leyland G, Favrat D. Environomic multi-objective optimisation of a district heating network considering centralized and decentralized heat pumps. Energy 2010;35. https://doi.org/10.1016/j.energy.2009.09.028.
[121] Ascione F, Bianco N, Mauro GM, Napolitano DF, Vanoli GP. Comprehensive analysis to drive the energy retrofit of a neighborhood by optimizing the solar energy exploitation – An Italian case study. J Clean Prod 2021;314:127998. https://doi.org/10.1016/J.JCLEPRO.2021.127998.
[122] Isaac S, Shubin S, Rabinowitz G. Cost-optimal net zero energy communities. Sustainability (Switzerland) 2020;12. https://doi.org/10.3390/su12062432.
[123] Alaifan B, Azar E. Potential for Net-Zero Energy Communities in Kuwait: An Empirical Techno-Economic Modeling and Optimization Approach. Buildings 2023;13. https://doi.org/10.3390/buildings13082096.
[124] Leprince J, Schledorn A, Guericke D, Dominkovic DF, Madsen H, Zeiler W. Can occupant behaviors affect urban energy planning? Distributed stochastic optimization for energy communities. Appl Energy 2023;348:121589. https://doi.org/10.1016/j.apenergy.2023.121589.
[125] Vecchi F, Berardi U. Solar analysis for an urban context from GIS to block-scale evaluations. Energy Policy 2024;184. https://doi.org/10.1016/j.enpol.2023.113884.
[126] Orlando M, Bottaccioli L, Quer S, Poncino M, Vinco S, Patti E. A Framework for Economic and Environmental Benefit Through Renewable Energy Community. IEEE Syst J 2023;17. https://doi.org/10.1109/JSYST.2023.3290941.
[127] De Franco A, Venco E, De Lotto R, Pietra C, Kutzner F, Bielig M, et al. Drivers, Motivations, and Barriers in the Creation of Energy Communities: Insights from the City of Segrate, Italy. Energies 2023;16. https://doi.org/10.3390/en16165872.
[128] Kojonsaari AR, Palm J. Distributed Energy Systems and Energy Communities Under Negotiation. Technology and Economics of Smart Grids and Sustainable Energy 2021;6. https://doi.org/10.1007/s40866-021-00116-9.
[129] Hartmann K, Palm J. The role of thermal energy communities in Germany's heating transition. Frontiers in Sustainable Cities 2023;4. https://doi.org/10.3389/frsc.2022.1027148.
[130] Jans L, Goedkoop F, Perlaviciute G, Hamann K, Masson T, Burgerhof B. How bottom-up and top-down governance of community energy initiatives affects citizens' perceptions, acceptability, and willingness to join. Energy Policy 2024;195:114389. https://doi.org/10.1016/j.enpol.2024.114389.
[131] Brakovska V, Vanaga R, Bohvalovs G, Fila L, Blumberga A. Multiplayer game for decision-making in energy communities. International Journal of

Sustainable Energy Planning and Management 2023. https://doi.org/10.54337/ijsepm.7549.
[132] Guetlein M-C, Schleich J. Empirical insights into enabling and impeding factors for increasing citizen investments in renewable energy communities. Energy Policy 2024;193:114302. https://doi.org/10.1016/j.enpol.2024.114302.
[133] Guetlein M-C, Schleich J. Understanding citizen investment in renewable energy communities. Ecological Economics 2023;211:107895. https://doi.org/10.1016/j.ecolecon.2023.107895.
[134] Broska LH, Vögele S, Shamon H, Wittenberg I. On the Future(s) of Energy Communities in the German Energy Transition: A Derivation of Transformation Pathways. Sustainability (Switzerland) 2022;14. https://doi.org/10.3390/su14063169.
[135] Zhang H, Vorobeychik Y. Empirically grounded agent-based models of innovation diffusion: a critical review. Artif Intell Rev 2019;52. https://doi.org/10.1007/s10462-017-9577-z.
[136] Fouladvand J, Ghorbani A, Sarı Y, Hoppe T, Kunneke R, Herder P. Energy security in community energy systems: An agent-based modelling approach. J Clean Prod 2022;366. https://doi.org/10.1016/j.jclepro.2022.132765.
[137] Fouladvand J, Mouter N, Ghorbani A, Herder P. Formation and continuation of thermal energy community systems: An explorative agent-based model for the netherlands. Energies (Basel) 2020;13. https://doi.org/10.3390/en13112829.
[138] Blumberga A, Vanaga R, Freimanis R, Blumberga D, Antužs J, Krastiņš A, et al. Transition from traditional historic urban block to positive energy block. Energy 2020;202. https://doi.org/10.1016/j.energy.2020.117485.
[139] Calise F, Cappiello FL, Cimmino L, Dentice d'Accadia M, Vicidomini M. Thermo-economic analysis and dynamic simulation of a novel layout of a renewable energy community for an existing residential district in Italy. Energy Convers Manag 2024;313:118582. https://doi.org/10.1016/j.enconman.2024.118582.
[140] Otamendi-Irizar I, Grijalba O, Arias A, Pennese C, Hernández R. How can local energy communities promote sustainable development in European cities? Energy Res Soc Sci 2022;84. https://doi.org/10.1016/j.erss.2021.102363.
[141] Mahzouni A. The role of institutional entrepreneurship in emerging energy communities: The town of St. Peter in Germany. Renewable and Sustainable Energy Reviews 2019;107. https://doi.org/10.1016/j.rser.2019.03.011.
[142] Bonfert B. 'We like sharing energy but currently there's no advantage': Transformative opportunities and challenges of local energy communities in Europe. Energy Res Soc Sci 2024;107:103351. https://doi.org/10.1016/j.erss.2023.103351.

[143] Meister T, Schmid B, Seidl I, Klagge B. How municipalities support energy cooperatives: Survey results from Germany and Switzerland. Energy Sustain Soc 2020;10. https://doi.org/10.1186/s13705-020-00248-3.
[144] Hartmann K, Palm J. The role of thermal energy communities in Germany's heating transition. Frontiers in Sustainable Cities 2023;4. https://doi.org/10.3389/frsc.2022.1027148.
[145] Jans L, Goedkoop F, Perlaviciute G, Hamann K, Masson T, Burgerhof B. How bottom-up and top-down governance of community energy initiatives affects citizens' perceptions, acceptability, and willingness to join. Energy Policy 2024;195:114389. https://doi.org/10.1016/j.enpol.2024.114389.
[146] Terrier C, Loustau JRH, Lepour D, Maréchal F. From Local Energy Communities towards National Energy System: A Grid-Aware Techno-Economic Analysis. Energies (Basel) 2024;17. https://doi.org/10.3390/en17040910.
[147] Dimovski A, Moncecchi M, Merlo M. Impact of energy communities on the distribution network: An Italian case study. Sustainable Energy, Grids and Networks 2023;35:101148. https://doi.org/10.1016/j.segan.2023.101148.
[148] Hartmann K, Palm J. The role of thermal energy communities in Germany's heating transition. Frontiers in Sustainable Cities 2023;4. https://doi.org/10.3389/frsc.2022.1027148.
[149] Gaddi R, Mastrolonardo L. The energy of internal areas: a systemic approach in Taranta Peligna. TECHNE - Journal of Technology for Architecture and Environment 2023:142–50. https://doi.org/10.36253/techne-14474.
[150] Pastore LM, Lo Basso G, Ricciardi G, de Santoli L. Smart energy systems for renewable energy communities: A comparative analysis of power-to-X strategies for improving energy self-consumption. Energy 2023;280. https://doi.org/10.1016/j.energy.2023.128205.
[151] Aruta G, Ascione F, Bianco N, Iovane T, Mastellone M, Maria Mauro G. Optimizing the energy transition of social housing to renewable nearly zero-energy community: The goal of sustainability. Energy Build 2023;282. https://doi.org/10.1016/j.enbuild.2023.112798.
[152] Angrisano M, Nocca F. Urban Regeneration Strategies for Implementing the Circular City Model: The Key Role of the Community Engagement. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 14108 LNCS, 2023, p. 359–76. https://doi.org/10.1007/978-3-031-37117-2_25.
[153] Sen A. The Idea of Justice — Amartya Sen | Harvard University Press. Harvard University Press 2009.
[154] Sen A. Equality of what? The Tanner Lecture on Human Values. 1979.
[155] Blečić I, Bibo Cecchini A, Talu V. The capability approach in urban quality of life and urban policies: Towards a conceptual framework. City Project and Public Space, 2013, p. 269–88. https://doi.org/10.1007/978-94-007-6037-0_17.

[156] Basta C. On Marx's human significance, Harvey's right to the city, and Nussbaum's capability approach. Planning Theory 2017;16:345–63. https://doi.org/10.1177/1473095216641153.
[157] Talu V. Qualità della vita urbana e approccio delle capacità. ARCHIVIO DI STUDI URBANI E REGIONALI 2013:52–73. https://doi.org/10.3280/asur2013-107004.
[158] Alkire S. Valuing Freedoms. 2002. https://doi.org/10.1093/0199245797.001.0001.
[159] Nussbaum MC. Creating Capabilities. 2022. https://doi.org/10.2307/j.ctt2jbt31.
[160] Chiappero-Martinetti E, Moroni S. An analytical framework for conceptualizing poverty and re-examining the capability approach. Journal of Socio-Economics 2007;36:360–75. https://doi.org/10.1016/j.socec.2006.12.002.
[161] Blečić I, Cecchini A, Talu V. Capability approach and urban planning. Fertile urban capabilities and quality of urban life of the most disadvantaged inhabitants. Archivio Di Studi Urbani e Regionali 2018;48:269–88. https://doi.org/10.3280/ASUR2018-122003.
[162] Apsan Frediani A, Boni A, Gasper D. Approaching Development Projects from a Human Development and Capability Perspective. J Human Dev Capabil 2014;15:1–12. https://doi.org/10.1080/19452829.2013.879014.
[163] Velasco-Herrejón P, López Toledo E, Romo López JA, Antonio Escobar TK, Carrasco Morales E, Hernández Ramírez W, et al. A capability approach to analyse well-being impacts of wind energy infrastructure. Environmental Research: Infrastructure and Sustainability 2024;4. https://doi.org/10.1088/2634-4505/ad269c.
[164] Oliveira RV, Thorseth M. Ethical implications of a co-benefits rationale within climate change mitigation strategy. Etikk i Praksis 2016;10. https://doi.org/10.5324/eip.v10i2.1942.
[165] de Wildt TE, Chappin EJL, van de Kaa G, Herder PM, van de Poel IR. Conflicted by decarbonisation: Five types of conflict at the nexus of capabilities and decentralised energy systems identified with an agent-based model. Energy Res Soc Sci 2020;64. https://doi.org/10.1016/j.erss.2020.101451.
[166] Öztürk M, Tsoukiàs A. Bipolar preference modeling and aggregation in decision support. International Journal of Intelligent Systems 2008;23:970–84. https://doi.org/10.1002/int.20302.
[167] Vembu S, Gärtner T. Label ranking algorithms: A survey. Preference Learning, 2011, p. 45–64. https://doi.org/10.1007/978-3-642-14125-6_3.
[168] Yu PLH, Wan WM, Lee PH. Decision tree modeling for ranking data. Preference Learning, 2011, p. 83–106. https://doi.org/10.1007/978-3-642-14125-6_5.

[169] Pigozzi G, Tsoukiàs A, Viappiani P. Preferences in artificial intelligence. Ann Math Artif Intell 2016;77:361–401. https://doi.org/10.1007/s10472-015-9475-5.
[170] Keeney RL, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. IEEE Trans Syst Man Cybern 1979;9. https://doi.org/10.1109/TSMC.1979.4310245.
[171] Auriau V, Belahcène K, Malherbe E, Mousseau V. Learning Multiple Multicriteria Additive Models from Heterogeneous Preferences, 2025, p. 207–24. https://doi.org/10.1007/978-3-031-73903-3_14.
[172] Fürnkranz J, Hüllermeier E. Preference learning: An introduction. Preference Learning, 2011, p. 1–17. https://doi.org/10.1007/978-3-642-14125-6_1.
[173] Hüllermeier E, Fürnkranz J. Editorial: Preference learning and ranking. Mach Learn 2013;93:185–9. https://doi.org/10.1007/s10994-013-5414-z.
[174] Gehrlein J, Miebs G, Brunelli M, Kadziński M. An active preference learning approach to aid the selection of validators in blockchain environments. Omega (United Kingdom) 2023;118. https://doi.org/10.1016/j.omega.2023.102869.
[175] Burnap A, Pan Y, Liu Y, Ren Y, Lee H, Gonzalez R, et al. Improving Design Preference Prediction Accuracy Using Feature Learning. Journal of Mechanical Design 2016;138:071404. https://doi.org/10.1115/1.4033427.
[176] Ali Aden W, Zheng J, Ullah I, Safdar M. Public Preferences Towards Car Sharing Service: The Case of Djibouti. Front Environ Sci 2022;10:889453. https://doi.org/10.3389/fenvs.2022.889453.
[177] LimeSurvey Gmbh. LimeSurvey n.d. LimeSurvey Gmbh (accessed December 18, 2024).
[178] Istituto Nazionale di Statistica (ISTAT). Italian general censuses of population and housing 2011.
[179] Jacquet-Lagreze E, Siskos J. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. Eur J Oper Res 1982;10:151–64. https://doi.org/10.1016/0377-2217(82)90155-2.
[180] Siskos Y, Grigoroudis E, Matsatsinis NF. UTA methods. International Series in Operations Research and Management Science, vol. 78, 2005. https://doi.org/10.1007/0-387-23081-5_8.
[181] Fancello G, Congiu T, Tsoukiàs A. Mapping walkability. A subjective value theory approach. Socioecon Plann Sci 2020;72:100923. https://doi.org/10.1016/j.seps.2020.100923.
[182] Patiniotakis I, Apostolou D, Mentzas G. Fuzzy UTASTAR: A method for discovering utility functions from fuzzy data. Expert Syst Appl 2011;38. https://doi.org/10.1016/j.eswa.2011.06.014.
[183] Gomes DGE. Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? PeerJ 2022;10:e12794. https://doi.org/10.7717/peerj.12794.

[184] Aguinis H, Gottfredson RK. Best-practice recommendations for estimating interaction effects using moderated multiple regression. J Organ Behav 2010;31:776–86. https://doi.org/10.1002/job.686.
[185] Van Vegchel N, De Jonge J, Landsbergis PA. Occupational stress in (inter)action: The interplay between job demands and job resources. J Organ Behav 2005;26. https://doi.org/10.1002/job.327.
[186] McCabe CJ, Halvorson MA, King KM, Cao X, Kim DS. Interpreting Interaction Effects in Generalized Linear Models of Nonlinear Probabilities and Counts. Multivariate Behav Res 2022;57:243–63. https://doi.org/10.1080/00273171.2020.1868966.
[187] Coxe S, West SG, Aiken LS. Generalized Linear Models. Oxford University Press; 2013. https://doi.org/10.1093/oxfordhb/9780199934898.013.0003.
[188] Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 2013;4. https://doi.org/10.1111/j.2041-210x.2012.00261.x.
[189] Akaike H. Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd international symposium on information theory. Second International Symposium on Information Theory 1973.
[190] Schwarz G. Estimating the dimension of a model. The annals of statistics. Ann Stat 1978.
[191] Burnham K, Anderson D. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edn. Springer, Berlin. Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN 2002.
[192] Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67. https://doi.org/10.18637/jss.v067.i01.
[193] Bates D, Maechler M, Bolker B, Walker S. lme4: Linear Mixed-Effects Models using "Eigen" and S4. CRAN: Contributed Packages 2003. https://doi.org/10.32614/CRAN.package.lme4.
[194] Partitioning Around Medoids (Program PAM), 1990, p. 68–125. https://doi.org/10.1002/9780470316801.ch2.
[195] Hennig C. fpc: Flexible Procedures for Clustering. CRAN: Contributed Packages 2003. https://doi.org/10.32614/CRAN.package.fpc.
[196] Huu PN, Van DP, Xuan TH, Ilani MA, Trong LN, Thanh HH, et al. Review: enhancing Additive Digital Manufacturing with supervised classification machine learning algorithms. The International Journal of Advanced Manufacturing Technology 2024;133:1027–43. https://doi.org/10.1007/s00170-024-13781-8.
[197] Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning Research 2014;15.
[198] Kuhn M. caret: Classification and Regression Training. CRAN: Contributed Packages 2007. https://doi.org/10.32614/CRAN.package.caret.

[199] Massano M, Macii E, Lanzini A, Patti E, Bottaccioli L. A GIS Open-Data Co-Simulation Platform for Photovoltaic Integration in Residential Urban Areas. Engineering 2023;26:198–213. https://doi.org/10.1016/j.eng.2022.06.020.
[200] Paul BM, Doyle J, Stabler B, Freedman J, Bettinardi A. Multi-level Population Synthesis Using Entropy Maximization-Based Simultaneous List Balancing. Transportation Research Board 93rd Annual Meeting January 12-16, Washington, DC 2018;4457.
[201] Metropolitan Planning Organizations (MPOs). ActivitySim n.d. https://activitysim.github.io/ (accessed February 26, 2024).
[202] ISTAT. Aspetti della vita quotidiana 2022.
[203] ISTAT. Basi territoriali e variabili censuarie n.d. https://www.istat.it/notizia/basi-territoriali-e-variabili-censuarie/ (accessed January 10, 2024).
[204] González-García S, Dias AC. Integrating lifecycle assessment and urban metabolism at city level: Comparison between Spanish cities. J Ind Ecol 2019;23. https://doi.org/10.1111/jiec.12844.
[205] De Wolf C, Cerezo C, Murtadhawi Z, Hajiah A, Al Mumin A, Ochsendorf J, et al. Life cycle building impact of a Middle Eastern residential neighborhood. Energy 2017;134:336–48. https://doi.org/10.1016/J.ENERGY.2017.06.026.
[206] Di Turi S, Stefanizzi P. Energy analysis and refurbishment proposals for public housing in the city of Bari, Italy. Energy Policy 2015;79:58–71. https://doi.org/10.1016/J.ENPOL.2015.01.016.
[207] García-Pérez S, Sierra-Pérez J, Boschmonart-Rives J. Environmental assessment at the urban level combining LCA-GIS methodologies: A case study of energy retrofits in the Barcelona metropolitan area. Build Environ 2018;134:191–204. https://doi.org/10.1016/J.BUILDENV.2018.01.041.
[208] García-Pérez S, Sierra-Pérez J, Boschmonart-Rives J, Lladó Morales G, Romero Calix A. A characterisation and evaluation of urban areas from an energy efficiency approach, using geographic information systems in combination with life cycle assessment methodology. International Journal of Sustainable Development and Planning 2017;12. https://doi.org/10.2495/SDP-V12-N2-294-303.
[209] Delmastro C, Mutani G, Corgnati SP. A supporting method for selecting cost-optimal energy retrofit policies for residential buildings at the urban scale. Energy Policy 2016;99:42–56. https://doi.org/10.1016/J.ENPOL.2016.09.051.
[210] Harter H, Weiler V, Eicker U. Developing a roadmap for the modernisation of city quarters – Comparing the primary energy demand and greenhouse gas emissions. Build Environ 2017;112:166–76. https://doi.org/10.1016/J.BUILDENV.2016.11.031.
[211] HosseiniHaghighi SR, de Uribarri PMÁ, Padsala R, Eicker U. Characterizing and structuring urban GIS data for housing stock energy

modelling and retrofitting. Energy Build 2022;256. https://doi.org/10.1016/j.enbuild.2021.111706.
[212] Gestore Servizi Energetici. Mappa interattiva delle cabine primarie n.d. https://www.gse.it/servizi-per-te/autoconsumo/mappa-interattiva-delle-cabine-primarie (accessed November 1, 2024).
[213] Ahmed SK. How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. Oral Oncology Reports 2024;12:100662. https://doi.org/10.1016/j.oor.2024.100662.
[214] Ministero dell'Ambiente e dell'Energia. Decreto CER [DM n.414, 17.12.2023] n.d.
[215] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393–b2393. https://doi.org/10.1136/bmj.b2393.
[216] Mongin D, Lauper K, Turesson C, Hetland ML, Klami Kristianslund E, Kvien TK, et al. Imputing missing data of function and disease activity in rheumatoid arthritis registers: what is the best technique? RMD Open 2019;5:e000994. https://doi.org/10.1136/rmdopen-2019-000994.
[217] Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychol Methods 2002;7. https://doi.org/10.1037/1082-989X.7.2.147.
[218] Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? Int J Epidemiol 2014;43. https://doi.org/10.1093/ije/dyu080.
[219] Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? Int J Methods Psychiatr Res 2011;20. https://doi.org/10.1002/mpr.329.
[220] Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. Surv Methodol 2001;27.
[221] van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007;16. https://doi.org/10.1177/0962280206074463.
[222] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011;45. https://doi.org/10.18637/jss.v045.i03.
[223] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations. CRAN: Contributed Packages 2006. https://doi.org/10.32614/CRAN.package.mice.
[224] Schafer JL. Multiple imputation in multivariate problems when the imputation and analysis models differ. Stat Neerl 2003;57. https://doi.org/10.1111/1467-9574.00218.
[225] Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science 2007;8. https://doi.org/10.1007/s11121-007-0070-9.

[226] Nicolas Duminy. PyMCDA n.d.
[227] Andrade C. The P value and statistical significance: Misunderstandings, explanations, challenges, and alternatives. Indian J Psychol Med 2019;41. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19.
[228] Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. 2016.
[229] Frank E, Witten IH. Generating accurate rule sets without global optimization. Proceedings of the Fifteenth International Conference on Machine Learning 1998.
[230] Hornik K. RWeka: R/Weka Interface. CRAN: Contributed Packages 2006. https://doi.org/10.32614/CRAN.package.RWeka.
[231] Agenzia delle Entrate. Redditi e principali variabili Irpef su base comunale [Sub-Comunale(CAP)] 2024.
[232] DECRETO LEGISLATIVO n.216, 30.12.2023 2023.
[233] ISTAT. Sicurezza e stato di degrado delle città e delle loro periferie. 2024.
[234] Agenzia delle entrate. Quotazioni immobiliari n.d. https://www.agenziaentrate.gov.it/portale/schede/fabbricatiterreni/omi/banche-dati/quotazioni-immobiliari (accessed February 10, 2025).
[235] Greco S, Kadziński M, Mousseau V, Słowiński R. Robust ordinal regression for multiple criteria group decision: UTAGMS-GROUP and UTADISGMS-GROUP. Decis Support Syst 2012;52:549–61. https://doi.org/10.1016/j.dss.2011.10.005.
[236] Spazzafumo G, Raimondi G. Economic assessment of hydrogen production in a Renewable Energy Community in Italy. E-Prime - Advances in Electrical Engineering, Electronics and Energy 2023;4. https://doi.org/10.1016/j.prime.2023.100131.