Performance-Based Clustering for Building Stock Management at Regional Level*

Philipp Geyer[1], Arno Schlueter[2]
[1]KU Leuven, Kasteelpark Arenberg 1 - box 2431, 3001 Leuven - Belgium,
p.geyer@kuleuven.be, +32 16 32 69 59
[2]ETH Zürich, Switzerland

Abstract

To facilitate the energy transition, the retrofit of building stocks is a crucial task. A strategy is required to maximize the effect of retrofit to reduce GHG emissions in the given limits of the available investment means. The paper shows that type-age classifications of buildings are not an appropriate grouping for strategy development and proposes an algorithmic clustering as grouping method based on the effect of energy efficiency measures (EEM). This novel clustering method delivers groups of buildings that similarly respond to retrofit measures and thus provide a good basis to develop efficient large-scale retrofit strategies. Besides illustrating the method and its benefits, the paper draws conclusions on the transfer of the method to a regional scale. These conclusions address aspects of the larger heterogeneity of the building stock as well as data availability, scaling and supply structures as well as the utilization of the results for policy making.

1 Introduction

The objective of energy efficiency and emission reduction in urban structures requires the management of and strategy development for building stocks. A central requirement is the determination of the potential of energy efficiency measures (EEM), the related reduction of emissions and the potential for decentralized energy production considering the usual limitations of investment means. As it is, especially at regional scale, not possible to assess each individual building for EEM, building stock models and strategy development tools usually base on type-age classifications of buildings. The groups derived on this basis not necessarily react uniformly on EEM, which is shown at the end of the paper, and therefore do provide not a good basis for strategy development and policy making.

Therefore, the paper presents an approach using algorithmic clustering to derive groups that react similar on EEM. The background for this approach is that due to the digital revolution and smart technologies much more data are available that allow a far better prediction of the effect of EEM. However, the challenge is still to derive strategies to make these data for retrofitting and respective policies usable for decision making. The aim is a management of the building stock to renounce from fossil fuels and to drastically reduce the CO_2 emissions. Therefore, the purpose of applying algorithmic clustering is the derivation of strategies for the building stock, as the local government wants to approach the owners to propose measures and to establish a support program for retrofitting etc. For this purpose, it is key to identify groups of buildings that require the same measures to realize potential energy savings and to be able to find the groups automatically for that an investment has the best effect in emission reduction.

Dealing with large urban buildings stocks and decision making for investments, the pure number of buildings in urban structures provides an obstacle in the development of such retrofit strategies. This problem aggravates at a regional level because the number of buildings increases and even the conditions (energy tariffs, accessibility of energy carriers etc.) may vary. It is purely impossible to examine such numbers of buildings one after the other and proposing retrofit strategies. Also the type age classification is not a good approach for assigning retrofit strategies to building groups, which will be shown at the end of the examination. This situation calls for an advanced computer-aided analysis method.

To address this problem, the paper proposes in Section 2 to use detailed building data for an approach of clustering the building stock by means of hierarchical clustering. This clustering is based on effectiveness

* Parts of the paper have been published in Geyer P, Schlüter A, Cisar S (2014): A Performance-Based Clustering Method for Retrofit Management of Building Stocks, eg-ice Workshop 2014 in Cardiff.

of measures for emission reduction. The method supports the management of the building stock for the planning of large scale retrofit measures.

From this approach groups of buildings result that react similar to EEM. These groups allow for an easy selection of strategies for reducing emissions. Planners and decision makers can select appropriate measures and measure combinations. The reaction on the EEM determined by the method helps in this selection process. Furthermore, scheduling the measures within the development horizon by considering an available investment budget and starting at most effective measures leads to an effective retrofit strategy over time.

So far this clustering approach has been developed for the district and urban scale. Section 2 is therefore a recapture of the existing and previously published method (Geyer et al. 2014). In Section 3 we extrapolate the potential of the method to address the regional scale, which poses additional demands on the management of buildings stock due to the increased number of buildings and more heterogeneous boundary conditions.

1.1 Background

To simulate and propose retrofit measures for building stocks, the type-age classification has been introduced. Type-age classification assumes a correlation between the building age and energy consumption as a basis for retrofit planning. A typical instance of the type-age classification for Germany has been developed by the Institut für Wohnen und Umwelt (IWU) and recently extended to Europe (IWU, 2014). An alternative approach for identifying energy-related parameters within large building stocks is clustering: Santamouris et al. (2007) apply clustering to a database of 320 schools in Greece and build groups based on the energy consumption with climatic normalization. Hernandez et al. (2008) do a similar study. Phil Jones et al. (2001) cluster a building stock by building properties, such as heated ground floor area, facade, window to wall ratio. Gaitani et al. (2010) identify typical building properties / parameters for the clusters of the schools in Greece. Yamaguchi et al. (2007) identify district types and provide typical energy performance by simulating buildings in a representative district. Swan and Ugursal (2009) provide an overview of modelling techniques for the energy consumption of residential building stock. However, none of the previous research project applied clustering based on effectiveness of multiple measures to develop retrofitting strategies for a building stock, which is needed for strategy building and decision making.

1.2 Zernez Case Study

The proposed methods are demonstrated in Section 2 using the case study of Zernez, an applied research project for the transformation of a Swiss alpine village towards zero emissions in building operation. Aim of the research project Zernez Energia 2020 is to develop an action plan for a transformation strategy for the building stock and energy systems to become emission free within the near future. A combination of retrofitting existing building stock, transformation of energy systems for heat, electricity, a model for urban development strategies and sustainable growth of the community including an integrative planning process should enable the municipality to develop and apply measures and policies.

One key part of the project was the research and development on environmentally and economically effective retrofit strategies for the building stock. Since the financial means of building owners and the support by the municipality are limited, it is crucial to identify which measures are most effective for building retrofit. The setting of the project creates a possibility for formulating a bottom-up approach based on real building data. Together with the application patterns, the municipality surveyed and compiled a list of more than 50 parameters for each building including the last retrofit, the condition of the building substance, area, installed heating and domestic hot water system, consumption data if available, images and plans for easy recognition and context. A building database was established to quickly access all building parameters for further analysis and calculation. Each building was identified with an identification number, 3D-modelled and inserted into a GIS model for analysing photovoltaic potential (e.g. on roofs) and associated with a spreadsheet. These data serve as basis for the building-detailed analysis and performance-based clustering described in Section 2.

The analysis of the buildings in this database revealed that there are high variations in the energy consumption by the year of construction. Not the oldest buildings have the highest energy consumption but that one from the 20th century on, as shown in Figure 1 top left exemplarily for one building type, the residential buildings of Zernez. The reduction of the energy consumption by one typical retrofit action, which is the insulation of the façade and the replacement of the heating system, shows a very varying reaction (Column 4 of Figure 1). For this reason, a better adapted approach than the type-age classifications for grouping and strategy development was necessary.

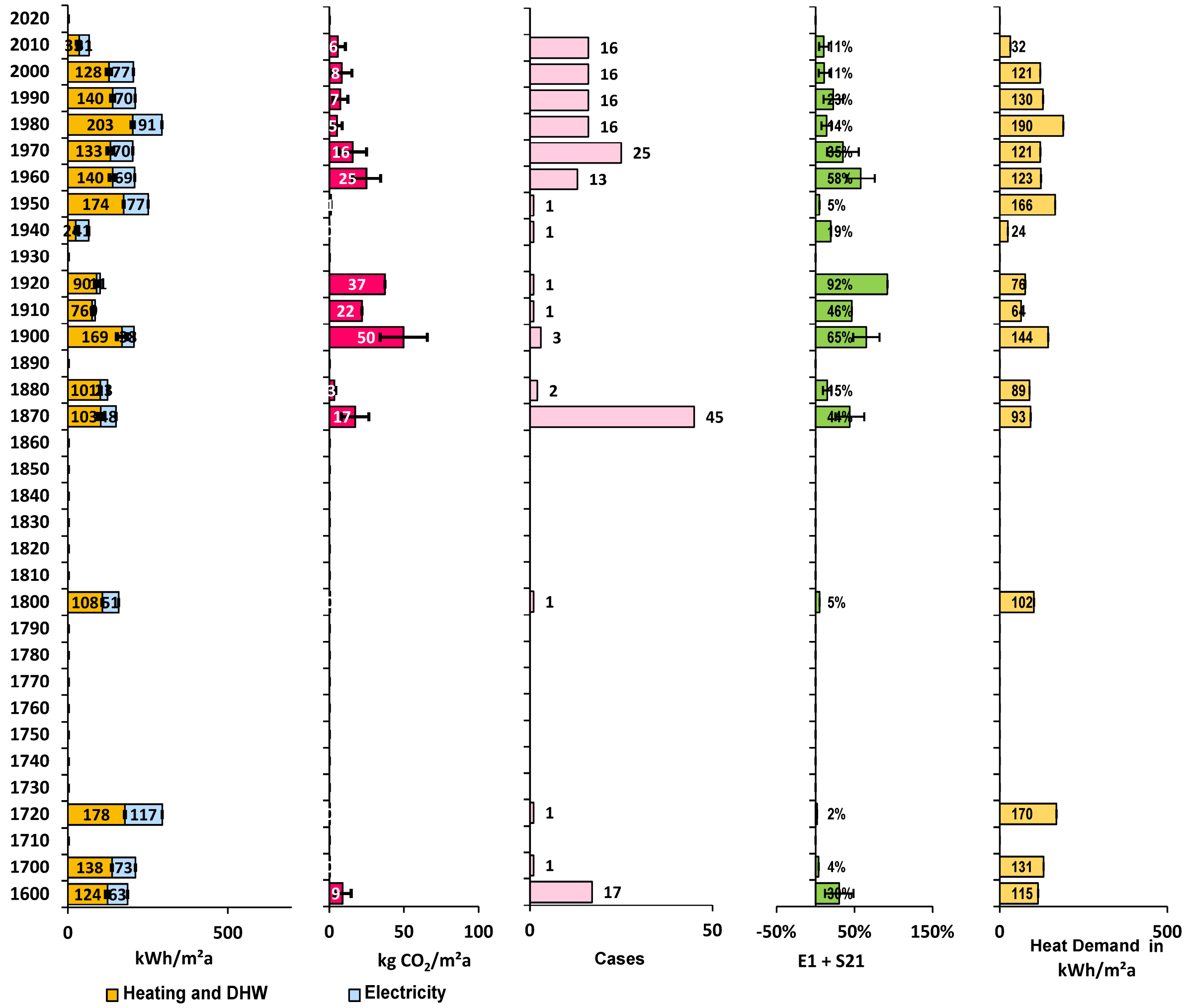


Figure 1: Analysis of the building type “residential” in the building stock of Zernez by construction year.

2 The Method of Performance-Based Clustering for Building Stock Management

The aim of the clustering is to develop groups of buildings that respond similar to the same EEM concerning the reduction of CO_2 emissions. The basis for the clustering is the response of the individual buildings to selected EEMs and measure sets. The clustering method identifies similarities in the responses and groups the buildings to clusters with a specific characteristic. This characteristic directly bases on the impact that the measures have instead of indirect indicators such as building type and age. Figure 2 gives an overview of the data preparation, the pre-processing, the clustering and the post-processing to develop retrofitting strategies. The following subsections provide an outline of the process which is described in detail by Geyer et al. (2014, 2016).

2.1 Effect of retrofit measure effects

Four different retrofit measures for reducing CO2 emissions in operation were defined. These were selected according to best practice examples of successful retrofits of the buildings in the village and include measures to reduce energy demand as well as measures to harness locally produced renewable energy sources. The selected measures included: (1) insulation of the envelope, (2) exchange of the heating systems (exemplarily by a heat pump), (3) decentralized electricity production (by a photovoltaic system) and (4) decentralized heat production (by a solar thermal system). Figure 3 provides an overview of the measures, their combinations and constraining rules.

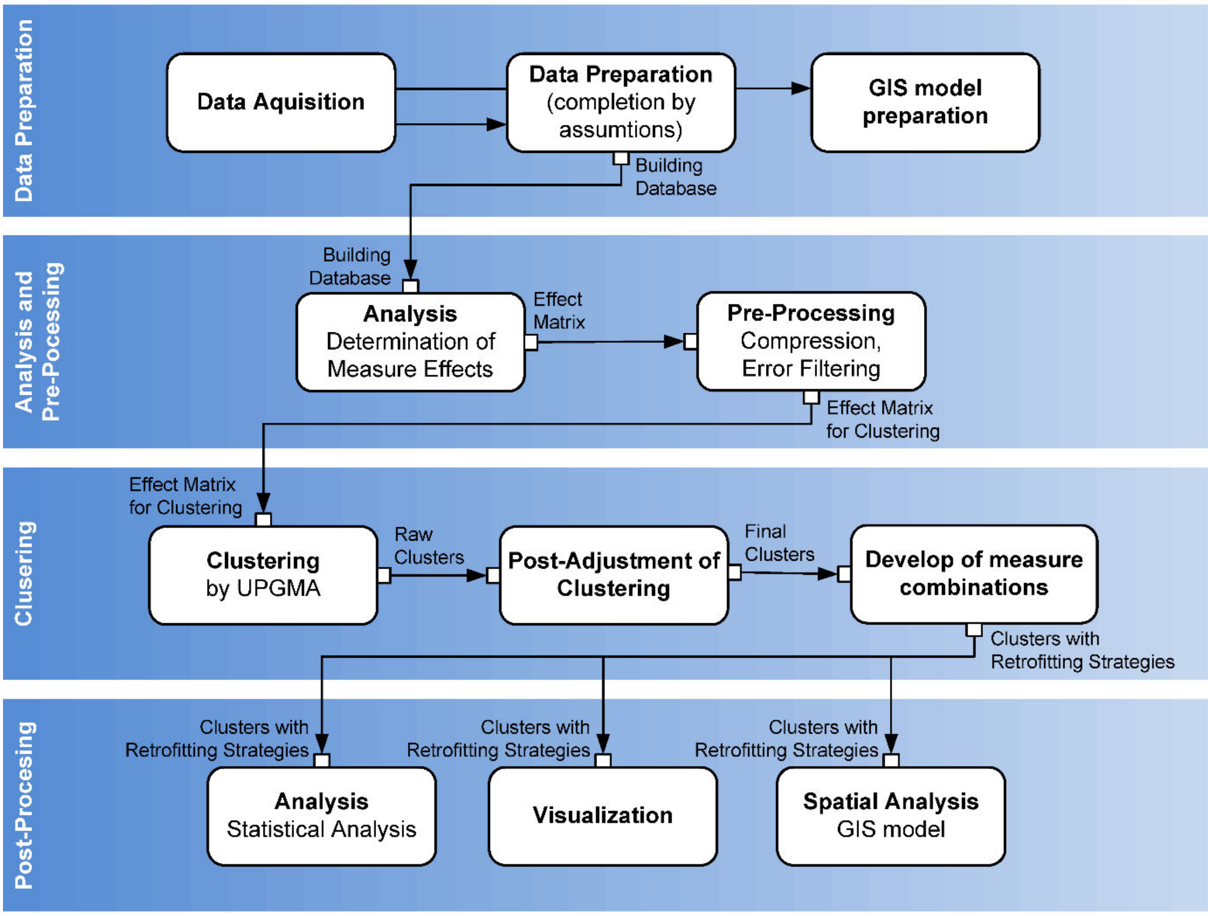


Figure 2. The flow chart shows the method of determining clusters and strategies from the given building stock database (Geyer et al. 2014).

The basis for the clustering is an automated analysis of the effects of the measures and the selected combinations for each building. This analysis uses the detailed database of the current constitution of the building and their energy consumption described in Section 1.2. Each measure was executed for each building in the database using the descriptive parameters of geometry, state of construction and heating system and further constraints, such as historic protection etc. Simplified energy calculations depending on the real energy consumption serve to determine the reduction of the total CO_2 emissions per building by each measure, which forms an effect matrix.

The simplified energy calculations bases on the geometric information from the data base, such as gross floor area, roof form and area, orientation etc., information on the construction, such wall construction, windows area and types etc., and on the energy consumption for heating and the electricity consumption. First, the share used for heating, for domestic hot water and for electricity consumption is constructed; for a part of cases, in which data are missing, assumptions on typical distribution were used. The construction of this share of consumption is required to assess the impact of the measures on the total building emission. As next step, estimation for the missing parameters of the building geometry takes place resulting in the following set of parameters: gross floor area, the areas of envelope, façade, windows, wall, roof and foundation as well as roof area available for photovoltaics and solar thermal systems. Together with the information on the construction and on CO_2 conversion factors describing the energy supply, the calculation of the reduction of emissions by applying the measures is possible. This calculation considers several constraints, e.g., that external insulation is not applicable to a historic façade or that a heat pump system with low temperature radiators should not be combined r with a badly-insulated façade; this is the implementation of the constraining rules shown in Figure 3.

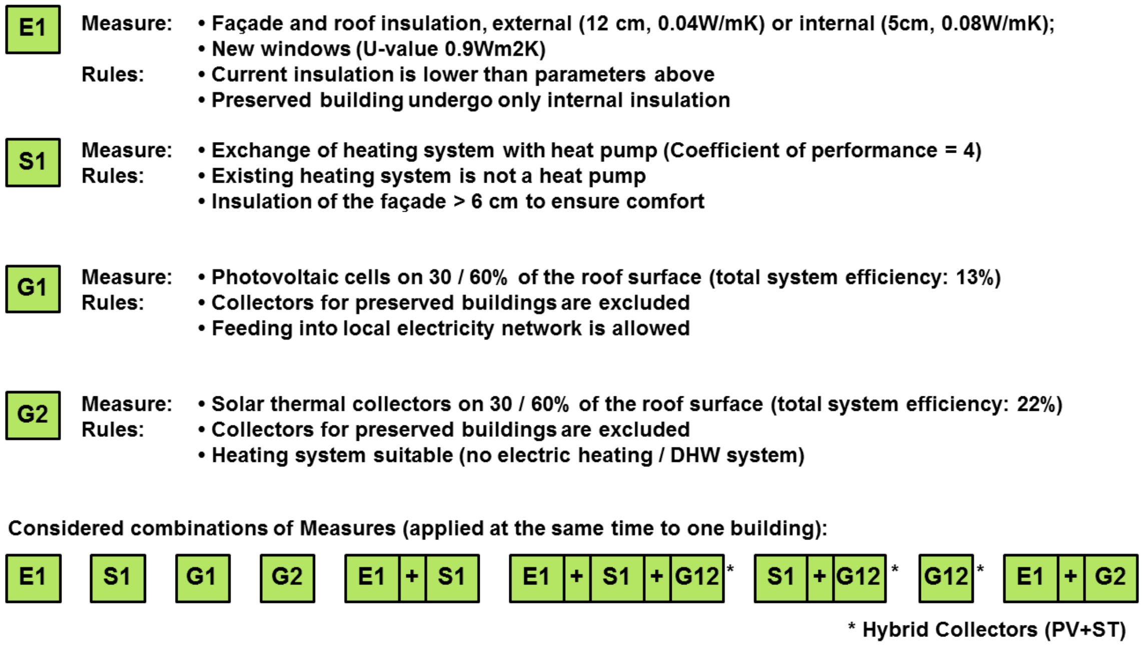


Figure 3: Overview of the measures with constraining rules and their combinations for emission reduction (Geyer et al. 2014).

2.2 Clustering

The next step is the clustering of buildings. The basis for the clustering is the response of each building to each measure of the previous step, which is comprised in the effect matrix. The interpretation of this matrix as multidimensional feature space. In this virtual feature space, buildings that react similar to EEM are located close to each other. This space provides the basis to identify groups of buildings by means of distance-based clustering methods. The both main types of clustering, which are hierarchical and partitioning clustering, perform well for this identification (Geyer et al. 2016).

Figure 4and 5 show exemplary results for the case of Zernez. These results are based on the percentage of emission reduction per building (X and Y axis of all diagrams). Alternatively, other scales are possible and appropriate depending on the objective of the strategy, which we already have shown (Geyer et al. 2016). In the buildings stock of Zernez, four clusters are identified.

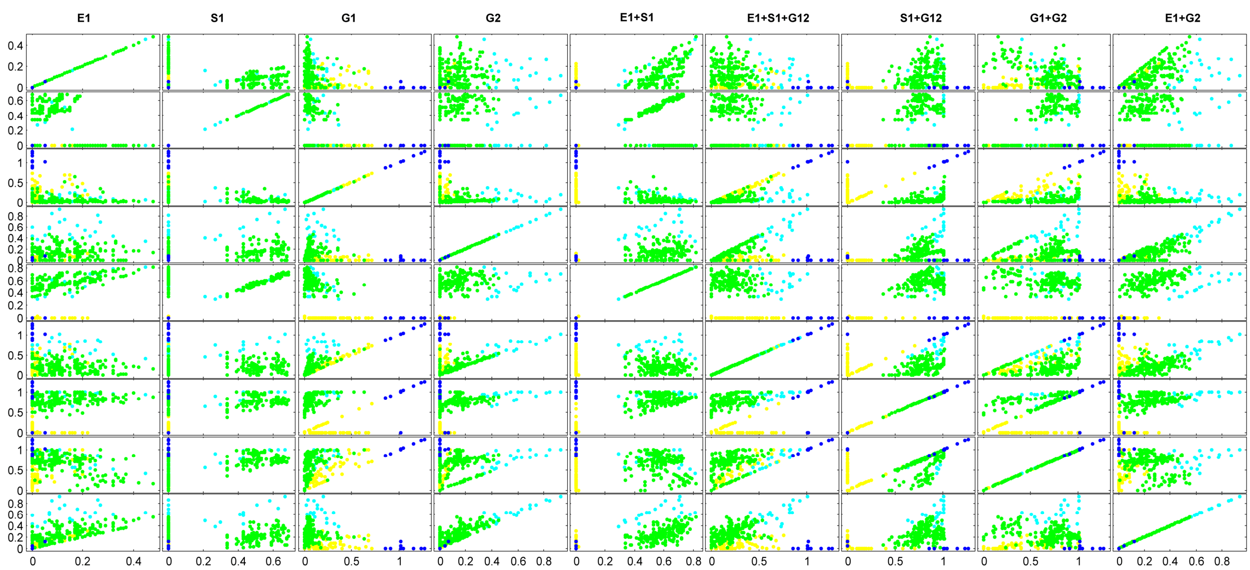


Figure 4: Scatterplots of the four clusters in the nine-dimensional feature space.

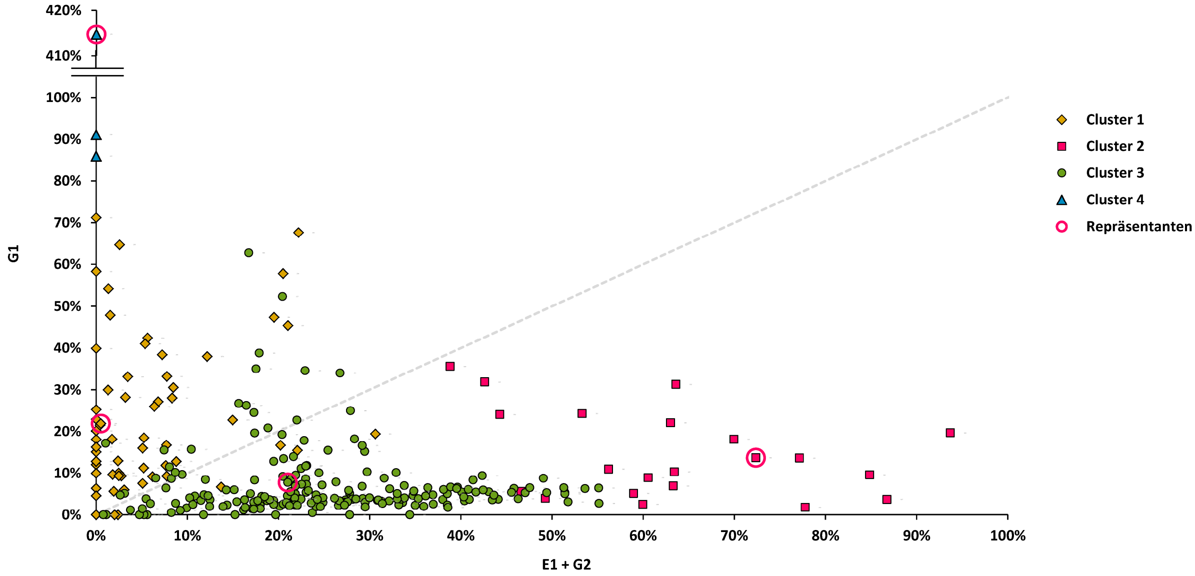
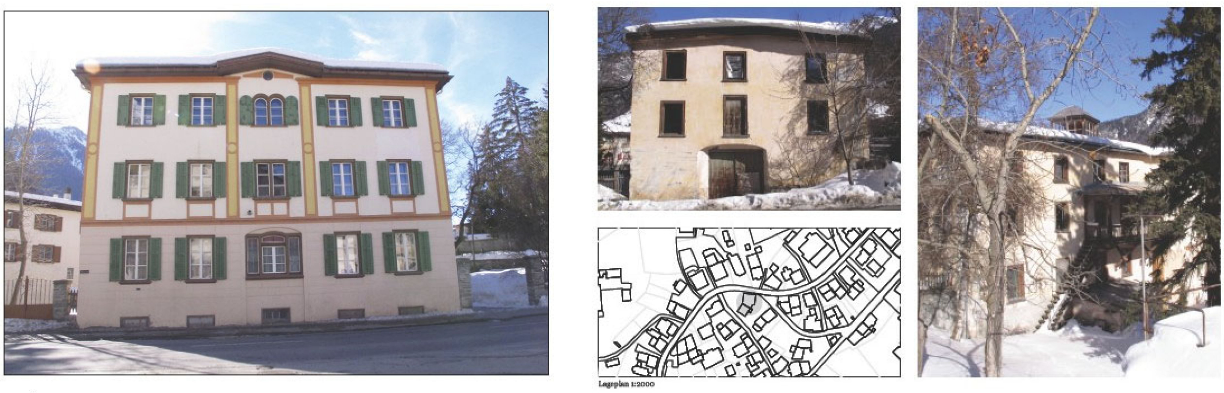


Figure 5: One scatterplot in detail illustrating the four identified clusters (Geyer et al. 2014).

2.3 Post-processing: Result interpretation and strategy development

In post-processing, representative buildings were selected for each of the four clusters to select a combination of measures that is most appropriate. A post-processing algorithm searches the building closest to the centroid but excludes buildings for that the data source is unsure because of missing entries and assumptions or that are marked as special buildings, such as the school or the sports hall. Figure 6 shows such a representative building for Cluster 4. The selection of such buildings supports the development and validation of developed strategies by planning a specific retrofit action.



Via Suot 8

E1: Low effect of facade retrofitting | S1: high heat demand, heat pump makes sense,
G1: high electricity demand, thus low effect of PV | G2: Solar thermal system only limited useful

CLUSTER

Cluster	Anzahl	Menge t/a	S1	G1	G2	E1	E1 +S1	E1 +S1 + G1 + G2	S1 +G1 + G2	G1 + G2	E1 +G2
1	66	3.015	13%	20%	17%	6%	17%	41%	37%	24%	11%
2	28	606	28%	35%	54%	11%	35%	104%	101%	89%	58%
3	41	378	14%	378%	29%	7%	23%	411%	400%	381%	11%
4	174	4.109	46%	12%	21%	14%	53%	85%	72%	19%	20%

REPRÄSENTATIVES GEBÄUDE

Cluster	ID	S1	G1	G2	E1	E1 +S1	E1 +S1 + G1 + G2	S1 +G1 + G2	G1 + G2	E1 +G2	Nutzung	Heizsystei	Zustai	_BGF_	Stockw	ÜP
1	153	0%	31%	20%	7%	15%	63%	51%	51%	25%	Wohnen	Ölheizung	mittel	691	2	1
2	113	35%	25%	36%	8%	40%	87%	83%	61%	42%	Wohnen	Ölheizung	gut	679	2,5	3
3	282	0%	170%	0%	0%	0%	170%	170%	170%	0%	Versorgung	Elektrisch	gut	585	1,5	1
4	177	53%	6%	0%	9%	57%	93%	92%	6%	9%	Wohnen	Elektrisch	gut	483	2,5	3

Figure 6: Representative building in the Zernez building stock for Cluster 4.

Figure 5 shows one scatterplot in detail that allows the exemplary interpretation of the identified clusters. The diagram shows the reaction of the buildings (each symbol is one building from the stock) on the measure photovoltaics (G1, Y axis) and the measure combination insulation of the envelope and solar thermal heat supply (E1+G2). So it shows an electricity measure versus a heat consumption measure. This illustrates the four identified clusters, the measures and their combinations:

Cluster 1: These buildings react more on electricity saving measures (G1) and on heat demand reduction (E1+G2). Therefore, photovoltaics are a good option for them whereas their heat demand is low so that respective measures, especially insulation, have a limited effect.

Cluster 2: This cluster comprises buildings with a high heat demand and/or a badly insulated envelope. Measures on reducing the heat demand have a high effect on reducing CO_2 emissions. Also photovoltaics work well.

Cluster 3: The building in this cluster moderately react on the electricity reduction measures. The reason is that they have only little roof area for photovoltaics due to their compact body. However, due to the high overall CO_2 reduction potential compared to the prices, photovoltaics are applied as well as heating energy reduction measures.

Cluster 4: This cluster includes buildings that have high reduction of CO_2 emissions by photovoltaics going above 100%, which means they are plus energy buildings. Often these are secondary buildings such as garages and storage buildings having low electricity consumption and no heat consumption allowing the production of energy for other buildings in the village.

From these reactions on the measures and by consideration of general cost information, recommended measures per cluster are developed. The following table shows these recommended measure combinations:

Table 1: Measure combinations developed as strategies for the clusters.

Cluster	Measure Combination
C1	S1 + G1 + G2
C2	E1 + S1 + G1 + G2
C3	E1 + S1 + G1 + G2
C4	G1

This exemplary interpretation and the recommended measures show how the clusters can be serve to develop strategies for retrofitting and for policy making for the village's future energy strategy.

2.4 Cluster analysis

Finally, the data analysis of the building stock following the classification shows that the clusters provide a far better classification than the building age with fewer and therefore easier to manage categories. To illustrate this advantage, the effect of one exemplary measure combination, which is E1 + S21, i.e. the insulation of the façade and the exchange of the heating system is compared. On the right of Figure 1, which analyses the residential buildings in the building stock of Zernez by construction year, the reaction on this exemplary measure combination is shown in the fourth column. These reactions exhibit high variation, indicated by the black range symbols, for all larger building groups, i.e. those built in the 1870 and between 1960 and 1990. In contrast, the same analysis for the clusters in Figure 7, forth column exhibits nearly no variation. This means that measures based on cluster analysis are very precisely plannable without a high risk of being inadequate.

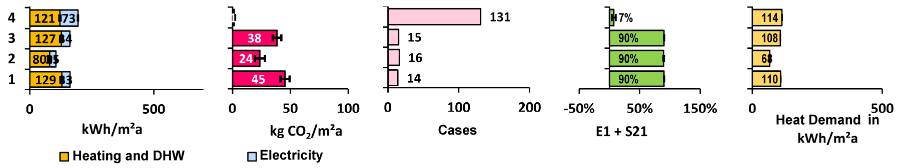


Figure 7. The data analysis of building stock by cluster shows the far better similarity of the buildings than an analysis by construction year.

2.5 Transformation strategy development

Taking a given budget into account, as for the case study given by the available investment means of the community, transformation strategies for a whole region can be developed. For this purpose, investment costs for measures are assessed and the cost effectiveness, which is the amount of CO_2 emission reduced per year per invested Swiss Franc. The cost effectiveness of the measures determines the required investments per cluster. Considering the available budget per year, the duration of the retrofitting of a cluster is known.

From this information, a transformation path is derived such as outlined in Figure 8. The reduction potentials in percent shown in Figures 4 and 5 together with the absolute emission volumes and the required investments for the developed measure combinations (Table 1) form the basis for the transformation path. This path follows the assumption that the retrofit starts with the clusters and measures of the highest cost effectiveness. This means to start from the lowest hanging fruits to quickly achieve reduction of emissions. A further study under publication will illustrate this with real data (Geyer et al. 2016).

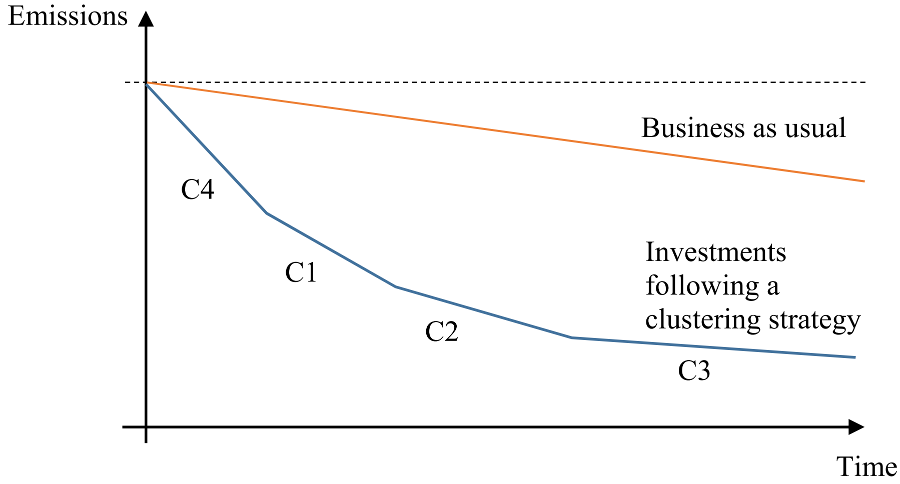


Figure 8: Derived development strategy.

3 Relevance and application of the performance-based clustering in European regions

Residential energy consumption accounts for about a third of EU's total energy consumption (European Commission 2015). The majority of the heating systems still uses fossil fuels (European Commission 2015a). This causes a high volume of anthropogenic GHG emission. Thus, an important goal is the reduction of these emissions by retrofitting the building stock. In this context, the building stock management plays a major role. Building retrofit requires high investments and it is required to invest limited available means with the best effect.

Performance-based clustering allows to identify groups of buildings that have high priority for retrofitting in order to quickly reduce the emissions with the available budget. Especially at regional level with varying external conditions, the approach allows the identification of well-fitting groups and strategies in heterogeneous building stocks. Such varying conditions are caused by large-scale supply structures, such as power plants and district heating systems, local energy prices, changing climate conditions, renewable sources (hydropower, geothermal etc.), economic state of a region etc. The clustering approach dynamically adapts to the situation and thus far better adapts to given conditions than the rigid type-age classification and thus far better meets the individual characteristic especially at a regional level.

These resulting clusters that are adapted to the specific conditions are a very valuable source for regional energy policy development and for the steering of investments. Interpretation allows the derivation of well-fitting retrofitting and investment policies. In the forthcoming study of Zernez (Geyer et al. 2016) it could be shown that with the same investments using the clustering-derived strategy 80% of the emissions of the buildings can be avoided compared to 20% in case of business-as-usual with the same budget.

Up to now, the method has not been applied to very large building stocks, which is planned in future research. However, the following subsections provides an assessment of the application of the method in such a situation.

3.1 Data availability

Compared to the type-age classification, the clustering approach requires more data. Some of these data, such as energy consumption, are connected to economic interests and privacy issues. However, in the context of the smart cities and smart regions paradigm, there is a strong force making these data available. The expectation is that for regional strategy development in near future a far better data basis available. Furthermore, the clustering approach is also applicable on the basis of limited data availability. This would decrease its representation of consumption but still adapt to the specific characteristics of the building stock and support far better policy development and strategic investment than a conventional approach.

3.2 Precision in terms of energy consumption

Compared to the type-age approach, which usually proposes EEM only based on one example building per group and the test of EEM for this specific building by building-level simulation, a far higher matching to the overall group characteristic can be expected by clustering. Although the applied simplified energy calculations have not the precision of the building-level energy simulations, the overall fitting and prediction per group is far better by taking into account all buildings.

Furthermore, in case of Zernez, monitored consumption data from a survey were available and the clustering is based on these data by calculating relative effects of EEM. Such a situation will become increasingly available in future by smart metering etc. The use of such data far better fits retrofitting to the real situation of energy consumption than to assume theoretical data in simulations.

3.3 Scaling of the method

Assuming an automated process of data acquisition and assessment of measure effects in context of smart regions, the number of buildings, which are the instances for clustering procedure, increases significantly. Therefore, a well-scaling behaviour for large number of instances in clustering is therefore of importance. Typically hierarchical clustering has a computational complexity of $O(N^2)$ (Xu and Wunsch 2005), which means that the computation time is proportional to the squared number of buildings. For the Zernez case with 300 buildings, the computation time of clustering did not play a role (it was below a second on a usual PC). However, the quadratic behaviour leads to a significant increase going to 30,000 buildings, which is a usual amount for a region. However, other clustering algorithm behave much better. For instance, the partitioning clustering algorithm k-means performs better with a complexity of $O(NKd)$ with N objects, K clusters and d dimensions (Xu and Wunsch 2005), which means a linear increase of calculation time with the number of buildings. A test on the Zernez data revealed that the results of both clustering methods do not much deviate (Geyer et al. 2016).

3.4 Considering decentralized supply structures

A further aspect of building stock management as a bottom-up strategy for the energy transition needs to go beyond the building level. The energy transition cannot be solved at building level only. Also in this context, clustering can help to identify buildings for measures. However, in this context the spatial aspect has high importance because supply structures and networks strongly rely on location of supply and consumers. In a further forthcoming study, we present a method to identify thermal microgrids in a building stock based on clustering and fuzzy logic (Schlueter et al. 2016). These data mining methods select buildings that are appropriate to join a small local heat network and spatially group them into microgrids. A further potential of similar applications is expected.

3.5 Policy making

Besides the direct approach of building owners to propose retrofit actions, especially at regional level, it is of interest to develop retrofit policies. The proposed clustering method very well supports the activities of policy makers. The groups of buildings derived by clustering may directly serve for the setup of subsidy programs. The method not only delivers effective EEM per group but also, by looking at similarities in building parameters of a group, allows to derive rules for such subsidy programs. For example, the use of a heat pump in Cluster 4 in the Zernez case has a high effect. An examination of the group parameters shows that the predominating electric resistance heating systems are the reason for this behaviour. From this information, a subsidy program for the exchange of such electric heating systems by heat pumps can be derived.

4 Conclusions

Data mining methods, such as clustering, have a high potential to support the planning and strategic development of the retrofit of building stocks. They allow to focus investments on building groups and strategies that have a high effect for a given investment budget. Especially moving to regional level in the context of smart cities and smart regions requires intelligent computer support to gain strategic information from large data bases. The presented approach showed significant better adaption of proposed measures and strategies than conventional classification approaches as the type-age classification.

Upcoming smart cities and regions will lead to large volumes of data available. This will lead to much higher importance of data mining and further intelligent computation. Whereas the type-age approach relies on many unsure assumptions in terms of similarity, typical utilization, construction and performance of the buildings in one group, data mining provides the means to exploit the available data, such as shown for the clustering approach for strategic retrofit development. In the context of smart built structures, we recommend such methods instead of traditional type-age approaches. We expect that such methods will gain increasingly in importance for planning and strategic development of an energy efficient and sustainable built environment.

Acknowledgements

We thank our collaborators of the research project 'Zernez Energia 2020' for the exchange of data related to cost, energy conversion and building energy calculations. This research was partly funded by the Swiss Commission for Technology and Innovation (CTI).

References

Geyer P, Schlüter A, Cisar S (2014): A Performance-Based Clustering Method for Retrofit Management of Building Stocks, eg-ice Workshop 2014 in Cardiff.

Geyer P, Schlueter A, Cisar S (2016): Application of Clustering for the Development of Retrofit Strategies for large Building Stocks, Advanced Engineering Informatics, in print.

European Commission (2015): Statistical pocketbook – energy in figures EU 2015, https://ec.europa.eu/energy/sites/ener/files/documents/PocketBook_ENERGY_2015%20PDF%20final.pdf, accessed Nov 2015.

European Commission (2015a): Communication on an EU strategy for Heating and Cooling – the contribution from heating and cooling to realising the EU's energy and climate objectives, http://ec.europa.eu/smart-regulation/roadmaps/docs/2015_ener_026_heating_cooling_strategy_en.pdf, accessed Nov 2015.

Hernandez P, Burke K, Lewis JO (2008): Development of energy performance benchmarks and building energy ratings for non-domestic buildings: An example for Irish primary schools, Energy and Buildings 40(3), pp. 249-254, http://dx.doi.org/10.1016/j.enbuild.2007.02.020.

IWU (Institut Wohnen und Umwelt) (2012): Typology Approach for Building Stock Energy Assessment. Main Results of the TABULA project, www.building-typology.eu, accessed Dec 2014.

Jones P, Lannon S, Williams J (2001): Modelling building energy use at urban, Proceedings of the 7th International IBPSA conference, Rio de, pp. 175-180.

Santamouris M, Mihalakakou G, Patargias P, Gaitani N (2007): Using intelligent clustering techniques to classify the energy performance of school buildings, Energy and Buildings 39(1), pp. 45-51, http://dx.doi.org/10.1016/j.enbuild.2006.04.018.

Schlueter A, Geyer P, Cisar C (2016): Analysis of Geo-referenced Building Data for the Identification and Evaluation of Thermal Microgrids, IEEE Proceedings Special Issue on Microgrids and Smart Buildings, submitted

Swan LG, Ugursal VI (2009): Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13(8), pp. 1819-1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.

Xu R, Wunsch DI (2005): Survey of clustering algorithms, Neural Networks, IEEE Transactions on 16(3), pp. 645-678.

Yamaguchi Y, Shimoda Y, Mizuno M (2007): Proposal of a modeling approach considering urban form for evaluation of city level energy management, Energy and Buildings 39(5), pp. 580-592, http://dx.doi.org/10.1016/j.enbuild.2006.09.011.