DISS. ETH NO. 25083

Data Mining Meets Life Cycle Assessment: Towards Understanding and Quantifying Environmental Impacts of Individual Households

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

ANDREAS DANIEL FRÖMELT

MSc ETH Env Eng, ETH Zurich

born on 20.04.1983

citizen of

Stettfurt (TG), Switzerland

accepted on the recommendation of

Prof. Dr. Stefanie Hellweg, examiner

Prof. Dr. Edgar Hertwich, co-examiner

Prof. Dr. Bernhard Steubing, co-examiner

2018

To Melanie, Ronja and Maira

ACKNOWLEDGEMENTS

The following lines are dedicated to all the people, colleagues, friends and family who have accompanied my journey through this PhD thesis. The support and encouragement I have received from all these people are unbelievable. Although it is difficult to adequately reward the aid I have received, I will try to express my gratitude especially to the following persons.

First, I would like to thank Stefanie Hellweg for her great supervision, encouragements and support during my dissertation. I am also thankful for her patience with me, for the openness to my ideas, and especially also for the liberties she granted to me, not only with regard to my research, but also with regard to pursuing a work-life balance. It goes without saying that the last point is of utmost importance for a father of two girls.

I would also like to thank Edgar Hertwich and Bernhard Steubing for co-examining my PhD thesis. Although it was not his official role, Bernhard was an important mentor to me in the beginning of my dissertation. Besides introducing me to the world of life cycle assessment, I will never forget the many interesting discussions and cheerful chats with him. As well, I am very grateful for having had the chance to do my PhD at the Chair of Ecological Systems Design (ESD). I cannot think of a better research group. The working atmosphere is just amazing and the high level of research very much inspiring. ESD is a great team and the spirit goes beyond work; it is actually like working in a big family and I am thankful for all the uncounted occasions of interesting talks and hilarious chats. I apologize for only naming a few, but I would like to express my thanks especially to the ones spending most time with me in the group: Carl Vadenbo, Stephan Pfister, Christopher Mutel, my officemate Mélanie Haupt, Christie Walker, Catherine Raptis, Adrian Haas, Christopher Oberschelp, Thomas Sonderegger, Franziska Stössel, Claudio Beretta, Laura Scherer, Florian Suter, and Barbara Dold. Of course, I also would like to thank Dominik Saner for sparking ignition with his great ideas and Niko Heeren who was not only a great fellow-PhD-candidate but has also become a good friend. In this regard, I also say thank you to René Buffat from the Chair of Geoinformation Engineering at ETH Zurich. Starting from a loose expression of intentions, Niko, René and I have evolved from a very fruitful and motivating collaboration to real friendship.

Moreover, thanks go to Manuela Mauchle, my first Master student, who finally even became a co-author. A special thanks also to David Dürrenmatt for the valuable discussions and inputs with regard to data mining and machine learning techniques. Furthermore, I am thankful for having had the chance to be part of the team of the Laboratory for Environmental Engineering. Especially, I would like to thank this team's leader Daniel Braun for the motivating work in the lab, the many challenging discussions and particularly for giving me the opportunity to come back to ETH.

Furthermore, I would like to express my thankfulness to all the different project-teams I have been part of: *Zernez Energia 2020*, THELMA and SCCER Mobility. In this context, I give a special mention to the Zernez-project coordinator Michael Wagner, Kristina Orehounig, Georgios

Mavromatidis and the municipality of Zernez. Additionally, I am very grateful for the pleasant and fruitful collaboration with the Institute for Transport Planning and Systems Group at ETH Zurich. Here, I would like to offer my thanks especially to Rashid Waraich and Francesco Ciari who patiently gave me an understanding of MATSim.

Finally, I will also attempt to thank my dear friends and family for their love and support during my dissertation. The encouragements and aid I have received especially from my parents Christine and Hubert, but also from my sister Carmen, my sister-in-law Stefanie, my parents-in-law Doris and Peter and my best and closest friends are just amazing. I actually do not know how I would have survived the last few years without them.

The last point is particularly also true for my wife Melanie and my two daughters Ronja and Maira. There is not enough space and there are no adequate words to express my deepest gratefulness to Melanie. With your love and your patience, you have always encouraged and motivated me, even in the most difficult times. I would have never managed to accomplish this dissertation without your support. I am thankful for you being by my side and I am looking forward to embarking together on many future adventures that life has got in store for us.

Last but not least, I definitely also would like to thank my two little girls Ronja and Maira. Your love, joy and laugh have always been a big motivation, inspiration, a source of strength and just a great reason to love life. Although I unfortunately could not always spend the time I wished with you, I would like to thank you for distracting me from work once in a while and for leading me into your kids' world full of fantasy.

ABSTRACT

To reduce adverse impacts on nature, thus enabling future generations to lead a decent life, deep changes in present human behavior are urgently needed. Policymakers can assume a key role and aim at creating an environment that enables producers and consumers to move towards more sustainable behavioral patterns. However, in order to successfully implement policy interventions and to efficiently invest time and money in the most promising fields of action, policymakers are in need of a highly detailed level of quantitative information on prevailing consumption patterns and production systems.

The goal of this dissertation was to investigate and develop new approaches that would be able to provide a thorough information base to support the design, prioritization and implementation of effective environmental policies. Thereby, a special focus was laid on the exploitation of Big Data and the application of data mining and machine learning techniques to extract new information tailored to the respective policymakers' areas of influence.

In order to achieve this goal, a two-track approach was pursued. Initially, building upon in-depth surveys and data collection, a database providing accurate data of local actors and activities was established for the municipality of Zernez, a Swiss mountain village, in the scope of a transdisciplinary research project. Subsequently, a comprehensive spatially resolved modeling framework was developed to estimate similarly detailed data for data-scarce regions.

The in-depth analysis of the current carbon footprint in Zernez provided an effective planning basis for the research team to develop a concrete action plan. Yielding a greenhouse gas (GHG) reduction potential of 80%, the building stock was identified as a reasonable first step to devise GHG mitigation strategies in Zernez. The proposed actions could then lead to a reduction of 13% and 17% of the municipality's total carbon footprint from a consumption and production perspective, respectively. The experiences gained in this project demonstrated the importance of understanding and quantifying the variability of local actors (producers and consumers) to develop and prioritize targeted GHG reduction measures.

In order to provide other municipalities with similarly comprehensive information without laborious data collection, an extensive modeling framework was established in a second stage. The models elaborated in this dissertation follow three principles. First, they are built from the *bottom-up*, which allows for reproducing a realistic picture of the variability of local actors and for aggregating simulation results on any desired regional scale. Second, the models source data from publicly accessible databases and third, they adopt a life cycle consumption-based perspective, which means that they assess resource uses and emissions induced by the consumer demand in a certain area. The overall modeling framework takes thus individual households as central modeling elements and aims at deriving a realistic environmental profile for each household within a certain region. This resolution is important because purchase decisions are made on the level of households.

The overall modeling framework encompasses three sub-models: a physically-based building energy model, a data-driven consumption model and a mobility model building upon the results of an agent-based simulation. The overall model was applied to the whole of Switzerland in order to demonstrate its practical feasibility. Still, the concepts of the sub-models are generic enough to be applicable to other countries with comparable data.

A global sensitivity analysis was applied to study the building energy model's internal structure, and the database of Zernez allowed for an in-depth evaluation with primary data. Based on these insights, the model was then improved by integrating comprehensive large-scale geographic data, including the use of nationwide laser-scanning data to derive 3D-building geometries. The final model is able to provide estimates of energy demand for each residential building in Switzerland. A thorough evaluation of the model results with reported data concluded that this model can approximate a realistic picture of the overall characteristics of a certain building stock's energy demand.

The consumption model embarked on a novel approach to assess the variability of lifestyle-induced environmental impacts. Based on an extensive use of data mining techniques, prevailing consumption patterns were studied and 28 consumption-based archetypes derived. These archetypes were further investigated and revealed different behavior patterns within similar socio-economic groups. Furthermore, archetypical behavior deviating from macro-trends, such as increased environmental impacts with higher income, could be detected. The proposed archetype-approach can thus be regarded as a promising basis to foster the understanding of current consumption patterns and to contemplate effective policy measures to reduce consumption-induced environmental impacts. Moreover, these archetypes can serve as building blocks to model the demand of food, services, consumables and other goods of households within a specified region. For this, the archetypes were assigned to real Swiss households by means of a newly developed probability-based classification framework which simultaneously interlinks with both the building energy model and the mobility sub-model. The latter estimates the households' transport demands based on the results of an agent-based simulation framework that reproduces typical mobility behavior of the Swiss population.

The overall model predicts the demands in about 400 different consumption areas for all approximately four million real Swiss households by taking into account the given circumstances of a specific household. A hybrid life cycle assessment (LCA) framework then assesses and subdivides the environmental impacts associated with these demands into more than 200 different categories. The applied LCA allows for computing different environmental indicators, and found that the estimated average consumption-based carbon footprint of Switzerland amounts to 9.5 tons CO_2-equivalents per person per year. Besides interesting differences among household consumption archetypes, the large-scale application of the overall model also reveals regional distinctions. For instance, mobility demands in rural areas tend to induce higher GHG emissions than their urban peers. Such differences should be further investigated to identify potential drivers of environmental impacts.

The high resolution of the overall model and its ability to quantify the variabilities in both household- and region-specific behavior renders it a powerful information tool to understand locally occurring consumption patterns. It may thus help to find hotspots, identify areas of action, and allow for the designing of impactful environmental measures tailored to specific household groups to reduce their impacts. The model can further be used as a platform to evaluate policy scenarios in upcoming research. The physically- and component-based approach of the building energy model as well as the link to an agent-based mobility model in particular will allow for the analysis of future scenarios in the context of total household consumption.

This dissertation demonstrated how Big Data and its analysis techniques can be employed to create a comprehensive knowledge base to inform environmental policymaking at different regional scales. The developed model is a starting point for more detailed investigations and it is open for further developments, improvements and extensions in future.

ZUSAMMENFASSUNG

Tiefgreifende Veränderungen des menschlichen Verhaltens sind dringend nötig, um weitere Schäden an der Natur zu vermindern und somit zukünftige Generationen nicht zu beeinträchtigten. Hierbei können insbesondere politische Entscheidungsträger eine wichtige Rolle spielen, indem sie Rahmenbedingungen gestalten, die einen Anreiz für nachhaltigere Verhaltensmuster für Produzenten und Konsumenten geben. Um entsprechende Massnahmen erfolgreich zu implementieren und sowohl Zeit als auch Geld effizient in die vielversprechendsten Aktionsfelder zu investieren, benötigen diese Entscheidungsträger detaillierte quantitative Informationen über das vorherrschende Konsumverhalten und die Produktionssysteme.

Das Ziel dieser Dissertation bestand daher darin, neue Ansätze zur Erstellung einer entsprechend umfassenden Informationsbasis zu untersuchen und zu entwickeln. Diese Informationsbasis soll dabei so ausgelegt sein, dass sie die Gestaltung, Priorisierung und Implementierung von wirkungsvollen Umweltstrategien effektiv unterstützen kann. Ein spezieller Fokus wurde dabei auf die Nutzung von heute zur Verfügung stehenden grossen Datenmengen (*Big Data*) und die Anwendung von Techniken aus den Bereichen Data-Mining und maschinellem Lernen gelegt.

Für die Erreichung des Ziels der Dissertation wurde ein zweigleisiger Ansatz verfolgt. Einerseits wurde im Rahmen eines transdisziplinären Forschungsprojekts mit Hilfe von detaillierten Erhebungen eine Datenbank für das kleine Schweizer Bergdorf Zernez aufgebaut, die Daten über lokale Akteure und deren umweltrelevanten Aktivitäten umfasst. Andererseits wurde eine umfangreiche räumlich aufgelöste Modellplattform entwickelt, die ähnlich detaillierte Daten für beliebige Regionen abschätzen kann.

Die vertiefte Analyse des heutigen CO_2-Fussabdrucks von Zernez wurde vom Forschungsteam als Planungsgrundlage genutzt, um einen konkreten Aktionsplan zu erarbeiten. Der Gebäudepark wies ein Treibhausgas-Reduktionspotential von 80% auf und wurde daher als ein vielversprechendes erstes Aktionsfeld für die Konzipierung von Strategien zur Senkung von Treibhausgasemissionen identifiziert. Die vorgeschlagenen Massnahmen könnten schliesslich den totalen CO_2-Fussabdruck der Gemeinde um 13% (Konsumperspektive) beziehungsweise um 17% (Produktionsperspektive) reduzieren. Die Erfahrungen, die in diesem Projekt gesammelt wurden, zeigten auf, wie wichtig es für die Herleitung und Priorisierung von Massnahmen ist, dass der Vielfalt der lokalen Akteure (Produzenten und Konsumenten) Rechnung getragen und dabei die Variabilität in ihren Verhaltensmustern verstanden und quantifiziert wird.

In einer zweiten Phase wurde ein umfangreiches Modell erarbeitet, welches ohne aufwendige Datenakquisition auch für andere Gemeinden Daten auf einem ähnlichen Detaillierungsniveau generiert. Die Entwicklung der Modellmodule dieser Dissertation basierte auf drei Grundprinzipien: 1. Um ein realistisches Abbild der Variabilität der Verhaltensmuster von lokalen Akteuren zu reproduzieren und um die Modellresultate auf beliebigen geographischen Ebenen aggregieren

zu können, wurden sämtliche Modelle *bottom-up* („von unten nach oben"[1]) aufgebaut. 2. Die Modellmodule beziehen ihre Eingabedaten von öffentlich zugänglichen Datenbanken, damit sie ohne grossen Datenerhebungsaufwand universell einsetzbar sind. 3. Sie nehmen eine konsumbasierte Lebenszyklusperspektive ein, das heisst, sie bewerten Ressourcenverbräuche und Emissionen, die aufgrund der Konsumnachfrage in einem Gebiet verursacht werden. Insgesamt bedeutet dies, dass das Gesamtmodell darauf abzielt, realistische Umweltprofile für jeden einzelnen Haushalt in einer bestimmten Region herzuleiten. Dies ist wichtig, weil Kaufentscheide auf der Haushaltsebene gefällt werden.

Die Gesamtmodellplattform setzt sich aus drei Modellmodulen zusammen: ein physikalisch-basiertes Gebäudeenergiemodell, ein rein datenbasiertes Konsummodell und ein Mobilitätsmodell, das sich auf die Resultate einer agentenbasierten Simulation stützt. Das Gesamtmodell mit den verknüpften Modellmodulen wurde auf die gesamte Schweiz angewendet, um die praktische Umsetzbarkeit der entwickelten Ansätze zu demonstrieren. Die Konzepte der Modellmodule wurden jedoch generisch gehalten, so dass sie auch auf andere Länder mit vergleichbaren Daten angewendet werden könnten.

Mit Hilfe einer globalen Sensitivitätsanalyse wurde die interne Struktur des Gebäudeenergiemodells beleuchtet und zudem eine Modellevaluierung mittels Primärdaten aus der Zernezer Datenbank durchgeführt. Basierend auf den Erkenntnissen aus dieser Untersuchung wurde das Modell verbessert, indem umfangreiche räumliche Daten integriert wurden. Unter anderem wurden mit Hilfe von nationalen *Lidar*-Daten[2] dreidimensionale Gebäudegeometrien bestimmt. Das endgültige Gebäudeenergiemodell kann schliesslich für jedes Wohngebäude in der Schweiz den Energiebedarf abschätzen. Eine detaillierte Gegenüberstellung von Modellresultaten und Messdaten zeigte, dass das Modell ein realistisches Bild des Wärmebedarfs eines Gebäudeparks nachzeichnen kann.

Das Konsummodell verfolgt einen neuartigen Ansatz, um die Variabilität der durch unterschiedliche Lebensstile hervorgerufenen Umweltauswirkungen zu untersuchen. Der intensive Einsatz von Data-Mining-Techniken ermöglichte die Analyse von heutzutage vorkommenden Konsummustern und darauf aufbauend die Bestimmung von 28 konsumbasierten Archetypen. Diese Archetypen zeigten, dass unterschiedliche Verhaltensweisen innerhalb ähnlicher sozio-ökonomischer Gruppen vorkommen können. Des Weiteren wurde auch archetypisches Verhalten gefunden, das von Makrotendenzen, wie beispielsweise erhöhte Umweltfussabdrücke mit steigendem Einkommen, abweicht. Der vorgeschlagene Archetypenansatz kann als eine vielversprechende Grundlage verstanden werden, um heutige Konsummuster besser zu verstehen und wirkungsvolle Strategien zur Reduktion von konsuminduzierten Umweltauswirkungen auszuarbeiten. Darüber hinaus können die Archetypen auch als Modellbausteine verwendet werden, um den Haushaltsbedarf an Nahrungsmitteln, Dienstleistungen, Konsum- und weiteren Gütern innerhalb einer bestimmten Region abzuschätzen. Aus diesem Grund wurden die Archetypen rea-

[1] Im konkreten Fall bedeutet dies, dass kleine Einheiten (z. B. einzelne Haushalte) einzeln modelliert werden und später zu einem Gesamtbild aggregiert werden können.
[2] *Light detection and ranging*: Höhenmessung mittels Laserstrahlen.

len Schweizer Haushalten mit Hilfe eines neu entwickelten wahrscheinlichkeitsbasierten Klassifikationsverfahrens zugewiesen. Mit der Archetypenzuweisung verknüpft dieses Verfahren auch das Konsummodell mit dem Gebäudeenergiemodell und dem Mobilitätsmodell. Letzteres ermittelt den Mobilitätsbedarf von einzelnen Haushalten basierend auf den Resultaten einer agentenbasierten Transportsimulation, die das Mobilitätsverhalten der Schweizer Bevölkerung zu reproduzieren versucht.

Das Gesamtmodell berechnet schliesslich die Nachfrage in ca. 400 unterschiedlichen Konsumbereichen für alle ungefähr vier Millionen Schweizer Haushalte, indem es den gegebenen Umständen spezifischer Haushalte Rechnung trägt. Ein hybrides Ökobilanzierungsverfahren bewertet sodann die mit dieser Nachfrage einhergehenden Emissionen und Ressourcenverbräuche mit Hilfe von mehr als 200 Prozessmodellen. Die angewendete Ökobilanzierung erlaubt die Berechnung verschiedener Umweltindikatoren. So beträgt zum Beispiel der durchschnittliche konsumbasierte CO_2-Fussabdruck der Schweiz 9.5 Tonnen CO_2-Äquivalente pro Person und Jahr. Neben interessanten Unterschieden zwischen Haushaltskonsumarchetypen, zeigt die grossmassstäbliche Anwendung des Gesamtmodells auch regionale Differenzen. Beispielsweise tendieren ländliche Gebiete dazu, mit ihrem Mobilitätsverhalten höhere Treibhausgasemissionen zu verursachen als ihre städtischen Gegenstücke. Solche Unterschiede sollten zukünftig weiter erforscht werden, um mögliche Treiber von Umweltauswirkungen zu eruieren.

Die hohe Auflösung des Gesamtmodells und seine Fähigkeit, die Variabilität von haushalts- und regionen-spezifischen Verhaltensweisen zu quantifizieren, erlauben die Erstellung einer nützlichen Informationsbasis, um lokale Konsummuster zu verstehen. Das Modell kann daher helfen, Umwelt-*Hotspots* und Handlungsbereiche zu identifizieren, sowie wirksame Massnahmen zur Reduktion von Umweltauswirkungen, die auf spezifische Haushaltsgruppen zugeschnitten sind, zu erarbeiten. Im Rahmen zukünftiger Arbeiten, kann das Modell schliesslich auch als Plattform für die Bewertung von Massnahmenszenarien dienen. Insbesondere der physikalisch- und komponentenbasierte Ansatz des Gebäudeenergiemodells sowie die Verknüpfung mit einem agentenbasierten Mobilitätsmodell erlauben die Analyse einer Vielfalt an Zukunftsszenarien im Kontext des Gesamthaushaltskonsums.

Diese Dissertation demonstrierte wie *Big Data* und die entsprechenden Analysetechniken angewendet werden können, um eine umfassende Wissensgrundlage zu generieren, die die Gestaltung von wirkungsvollen Umweltstrategien auf verschiedenen geographischen Ebenen unterstützen kann. Das entwickelte Gesamtmodell stellt einen Ausgangspunkt für detailliertere Analysen dar und ist offen für zukünftige Entwicklungen, Verbesserungen und Erweiterungen.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Melting glaciers, sea level rise, droughts compromising food supply, severe losses of biodiversity, mudslides and floods threatening human settlements as well as heatwaves and air pollution affecting human health are just some of the environmental problems that are perceived by the public and make it to the headlines of newspapers year after year. In fact, there is also a large body of scientific reports and articles which underpins the increase of anthropogenic emissions, the use of resources and the consequential environmental problems. For instance, the Intergovernmental Panel on Climate Change (IPCC) scientifically confirms the existence of global warming, its potential implications such as glacial melting, mudslides or sea level rise and points out the influence of human activities on the climate system [1]. The IPCC also warns of further warming and changes in the climate system if greenhouse gas (GHG) emissions continue. However, to limit the consequences on the climate system, substantial reductions of anthropogenic GHG emissions are necessary according to IPCC. Steffen and colleagues [2] reveal that environmental problems other than climate change also need attention. Besides global warming, they identify the effects of mankind in the areas of biosphere integrity, biogeochemical flows, and land-system change as exceeding the carrying capacity of the Earth system. Living beyond the "planetary boundaries" in these environmental spheres might substantially threaten the functioning of the Earth system and thus pose a severe risk for modern societies and mankind [2]. Facing these findings, a large consensus among scientists exists that today's consumption and production patterns are unsustainable and deep changes in human behavior, society and economy are urgently needed (see the calls in e.g. [3–10]). The scientific evidence has also triggered the action of politicians and different international agreements have been established and signed. In the recent past, the Paris Agreement [11] and the United Nations' Sustainable Development Goals[1] (SDGs) [12] clearly demonstrate the political willingness of nations to take action and to find new pathways towards more sustainable interaction with the environment while simultaneously keeping or increasing the level of human well-being.

Despite the indisputable importance of international agreements and declarations of intents, a real change in present-day consumption behavior and production structure is an almost insurmountable challenge for all involved parties. By being embedded in complex and interacting economic, societal and cultural systems and driven by economic profit, personal and social values, shifting towards more sustainable forms of consumption and production is difficult for individual consumers and producers [5, 13]. Against this background, and regardless of the question about who should assume responsibility, policymakers can play a key role in creating an enabling environment for change [3, 5, 7, 13, 14]. They are not only in the position to reach and objectively inform consumers and producers about environmental implications, but they may also devise policies, set standards, enforce rules or create incentives. In this regard, local initiatives in particular have gained much attention in recent years and their importance has been widely acknowl-

[1] First and foremost SDGs 6 "Clean Water and Sanitation", 7 "Affordable and Clean Energy", 11 "Sustainable Cities and Communities", 12 "Responsible Consumption and Production", 13 "Climate Action", 14 "Life below Water", and 15 "Life on Land".

edged [15]. This is likely due to the fact that local authorities and initiatives are close to the wants and needs of the individual stakeholders, who are needed to achieve substantial reductions of environmental impacts. However, to successfully derive, prioritize and implement a reasonable set of measures which aims at mitigating environmental problems, quantitative information is needed on prevailing consumption patterns and local production systems.

A common first step to provide an adequate information base for policymakers is to environmentally assess the current emissions and resource uses for a certain area. This usually involves coupling statistics on material and energy flows with environmental background data on emissions and resource uses. In view of the importance of such analyses and given today's priority on climate change issues, a large amount of carbon footprint studies, mainly focusing on nations or cities, have been conducted (see e.g. [16–23]). An important aspect of such environmental assessments constitutes the system boundaries to adequately consider emissions and resource uses. In this regard, different frameworks have been established and comprehensive overviews of state-of-the-art balancing approaches are, for instance, given in [14, 24–26]. In spite of the many accounting schemes, two fundamentally different approaches can be distinguished: the production and the consumption perspective. While production-based accounting focuses on the direct emissions caused by actors within certain geographical boundaries, the consumption perspective accounts for environmental impacts independent of their geographical occurrence but induced by the inhabitants residing within the considered territorial boundaries. Consequently, the two perspectives overlap in the consideration of domestic production for domestic demand and differ in accounting for imports to satisfy the final demand of the study area's inhabitants (consumption perspective) and for export-oriented production within the geographical system boundaries (production perspective). Several researchers emphasize the complementarity of both accounting schemes [14, 24, 26–28]. In fact, both frameworks provide important insights for local policymakers from different angles. The consumption perspective might be more suited for tackling consumption behavior and might pinpoint potential shifts of environmental burdens by adopting a life cycle perspective. In contrast, the production perspective considers all actors within the area of influence of local authorities and thus includes also the entire local trade and industry (comprising production for domestic demand and export from the study area).

Even though the importance of both accounting perspectives is unquestionable, the consumption-oriented viewpoint takes a special role in countries like Switzerland. The two emission-intensive primary and secondary sectors contribute little to Switzerland's Gross Value Added (0.7% and 25.9% respectively [29]). Hence, the territorial inventory for GHG emissions, with its approximately 58 million tons CO_2-equivalents, is clearly below the estimated consumption-based footprint of about 91 million tons CO_2-equivalents [30]. Jungbluth and colleagues [30] estimate that 50% of the GHG emissions induced by the average Swiss person are embodied in trade, meaning that they were caused by the final demand within Switzerland but occurred outside the national borders (see also [19, 21, 30–32]). Therefore, a life cycle perspective as provided by the consumption-based accounting is essential for such countries to prevent problem shifts (also known as "carbon leakage" in the case of GHG emissions). Furthermore, the focus on households is another important aspect of the consumption perspective. Households are main drivers of economy. Their demand for products and services initiates diverse economic activities up- and downstream of the supply chains serving the households. In this respect, household consumption is estimated to cause 65% of global GHG emissions and 50% to 80% of global land, material, and water use [32]. The remaining shares can mainly be attributed to governmental consumption (5% – 7%) and gross capital formation (19% – 37%) [32]. Yet, the investments (gross capital formation) can be charged to the production of new goods and services and are thus induced by the final demand of households or governments. Additionally, it can be argued that governments' final aim is to serve households. Households can thus be regarded as ultimately responsible for environmental impacts associated with economic and governmental activities that aim at satisfying their needs. Many studies also go a step further by investigating priority fields of household consumption and identify housing, mobility and food as the consumption areas contributing most to environmental impacts (e.g. [6, 21, 32, 33]).

Most of the above-mentioned environmental assessment studies [16–23] deploy a top-down perspective. This means that they use aggregated data for the whole study area. Such assessments are very insightful to find environmental hotspots, to define priority areas or to reveal general tendencies. However, providing only total results or averaged data for the whole study area is too coarse to identify targeted environmental policies. Few consumption-based modeling approaches exist which attempt to estimate and map environmental footprints at spatially higher resolved levels than the study area as a whole: e.g. Baiocchi et al. [9], Minx et al. [34], Druckman and Jackson [35], and Jones and Kammen [36]. However, by focusing on the derivation and assessment of the average household of smaller sub-areas, these approaches still remain on aggregated levels to some extent. In order to derive effective measures which are tailored to the specific actors in the study area, more specific knowledge on and an understanding of the local processes and consumption patterns are required. In a consumption perspective, this means that individual households living in the study area should be the central element of consideration. It is at this level that purchase decisions are taken and thus the right level to search for solutions to mitigate household environmental impacts. However, household behavior is diverse and influenced by many different factors, ranging from living conditions and financial constraints to personal preferences, just to name a few. Therefore, to offer an in-depth knowledge base for policymaking, the variability of households or industrial actors needs to be captured in the case of a consumption-based or a production-based approach, respectively; especially if aspects of behavioral economics or psychology shall be involved in the development of policies [13, 37]. Although some of the above-mentioned spatially resolved modeling approaches [9, 34–36] rely on different household types or household characteristics data, none of these takes individual households as central elements and thus attempts to explicitly model the variability of household behavior within the study region. Furthermore, some of these models are either partly based on commercial and nontransparent data or are limited in scope; for instance by only considering certain consumption areas. Although Girod and De Haan [38, 39] assess the behavior of individual households on a national level, their analysis neither allows for a regional model nor for subsequent scenario analysis, which is important when investigating the effects of measures. The approaches of Saner and colleagues [40, 41] come closest to the suggestions above, but by modeling different consumption areas independent of each other, these models do not preserve the context of total household consumption and have not yet been applied on large scale. Nevertheless, these models can be regarded as a promising basis for further developments towards a comprehensive knowledge database for policymakers.

Finally, it is a matter of fact, that nowadays more and more data is available. This huge amount of information even resulted in the term "Big Data", which, according to Oxford Dictionaries [42], can be described as: "Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions." In the context of providing information for policymakers, not only the availability of large masses of data is of interest, but specifically also the techniques which are used to exploit these datasets. Such techniques are often summarized in terms like "Data Mining"[2] or "Machine Learning"[3] and may help to locate and interpret patterns in the data and also to make predictions for unknown variables [43, 44]. Therefore, considerable contributions to effectively support policymakers can be expected from Big Data and its analysis approaches. However, the application of the powerful tools of these data science disciplines requires careful and thorough investigations.

1.2 Goal and Research Questions

The goal of this dissertation is to **investigate and provide new approaches to support impactful environmental policymaking**. Thereby, the focus lies on the **retrieval and provision of comprehensive information** for – particularly local – decision- and policymakers. The provided information base shall be tailored to local conditions and thus support the derivation, prioritization and implementation of measures to abate adverse environmental impacts. However, it is beyond the scope of this dissertation to exhaustively explore and evaluate possible environmental policies.

More specifically, the present dissertation aims to address the following research questions (RQ):

RQ 1 What kind of information on which level of detail can serve as a basis to derive targeted measures aimed at mitigating environmental impacts?

RQ 2 What are efficient ways to provide this information, especially in view of constrained financial budgets to gather data and data scarcity in many sub-national regions?

RQ 3 How can Big Data and machine learning techniques contribute to the support of environmental policymaking?

[2] Oxford Dictionaries [42]: „The practice of examining large pre-existing databases in order to generate new information."

[3] Oxford Dictionaries [42]: „The capacity of a computer to learn from experience, i.e. to modify its processing on the basis of newly acquired information."

RQ 4 Specifically in a consumption-based scope: how can household consumption patterns be modeled to capture the context of total consumption and how can the variability of these patterns be regionalized and thus transferred to larger scales?

RQ 5 What are the requirements and how should a framework be designed to evaluate and investigate large-scale effects of planned environmental measures?

1.3 METHODOLOGICAL APPROACH

In order to set up and prepare a detailed information base for policymaking, basically two different pathways can be pursued: Either detailed data is collected locally by means of in-depth surveys or data is estimated by modeling approaches. Accordingly, a dual approach is chosen for this dissertation. On the one hand, accurate data of actual actors in a small municipality is gathered in a **transdisciplinary research project**. In addition, this project enables for direct contact to local policymakers and thus particularly helps to address RQ1. On the other hand, building upon the insights and experiences gained in this project and with a view to responding to RQ2, a comprehensive spatially resolved modeling framework is developed. By adopting a consumption-based perspective, **the goal of this bottom-up household consumption model is to derive a realistic environmental profile for each household within a certain region**. Thereby, a special focus is laid on the use of national statistics, publicly accessible data[4] and transparent, well-established databases as input to the sub-models. The sub-models themselves are all developed with open-source software. Furthermore, the modeling framework is applied to the whole of Switzerland and thus provides estimates of the environmental footprints for all approximately four million individual Swiss households. In spite of this concrete demonstration, the proposed methods and developed approaches are generic enough to be transferable to other countries which have similar datasets available.

An overview of the modeling approach is provided in Figure 1.1 in the form of a simplified flow scheme. The overall model consists of three sub-models: a physically-based building energy model, a model to assess mobility behavior and a data-driven consumption model.

Building energy model The housing energy demand model of Saner and colleagues [40] estimates the energy needs of each residential building within a specified region. Additionally, it represents an interesting compromise between computational load and data requirements and is thus a promising approach for building stock investigations on large scales. In a first step, the model of Saner et al. [40] is subjected to an in-depth analysis. Thereby, a global sensitivity analysis is applied for an internal model evaluation, while empirical data is used for an external model evaluation. Building upon these insights, the model is improved and its data basis is enhanced with large-scale geographic information. The final model is able to assess the space heating demand for individual buildings based on simplified heat bal-

[4] This means that accessing the data is in principle open to everyone. However, it is possible that the data acquisition is not free of charge and might also require data protection contracts.

ances and to roughly estimate electricity demand and hot water production. The allocation of the housing energy use to individual real households is straight forward since the national census data [45] indicate in which building a household lives.

Mobility model Also based on the ideas of Saner et al. [40], the mobility behavior of households is modeled by means of MATSim simulation results (Multi-Agent Transport Simulation) [46]. MATSim is an agent-based modeling framework and delivers spatially and temporally resolved information on traffic modes and routes chosen by the simulated agents. However, this traffic simulation framework has been improved since the publication of [40] – especially with regard to modeling public transportation – and new simulation runs based on more recent mobility data [47] have become available for the whole of Switzerland [48]. While the concepts of [40] to assign simulated agents to real household members is only slightly improved, the achievements of this dissertation lie in integrating new data and scaling up the modeling approach to the whole of Switzerland.

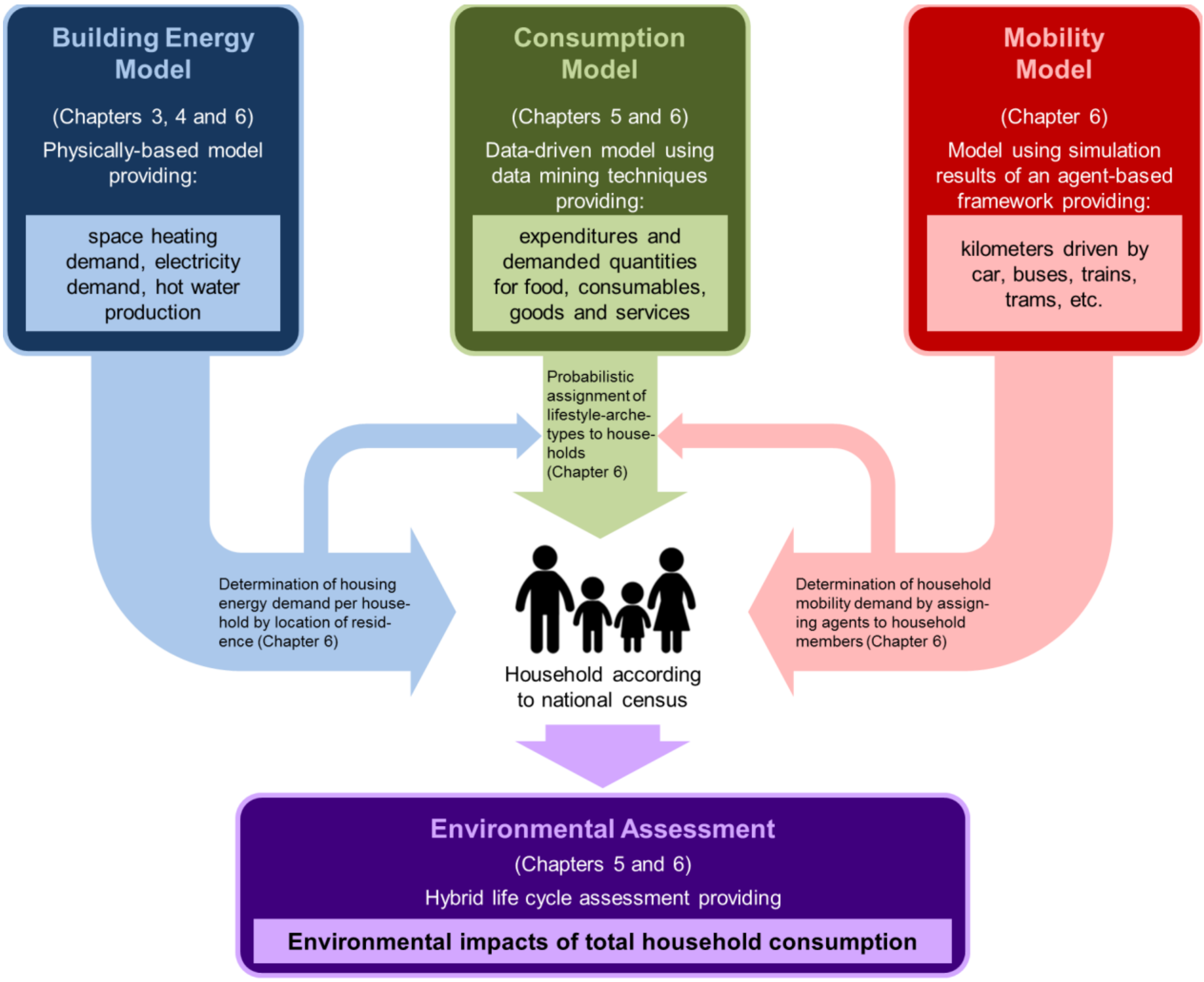


Figure 1.1: Simplified flow scheme providing an overview of the modeling approach developed in this dissertation. Note that the results of the building energy model and the mobility model are used in the assignment of the consumption model's archetypes to households in order to interlink the three models.

Consumption model In order to model the demand and associated environmental impacts of food, services, consumables, and other goods, novel approaches based on an extensive use of data mining techniques are developed. Thereby, the data of the Swiss Household Budget Survey [49] is explored to recognize consumption patterns and to identify consumption-based archetypes in a first step. These archetypes are then allocated to individual census households within a newly developed probability-based classification framework in a second step. Since the archetypes also provide estimates for mobility and housing behavior, the assignment of this archetypical behavior to households can be based on the results of the two other models, as well. Consequently, classifying a household as a certain archetype implicitly interlinks all three models and thus accounts for preserving the context of total household consumption.

Environmental assessment After having quantified all consumption areas of an individual household by means of the three sub-models, the environmental impacts associated with the household's demands need to be determined. For this purpose, the life cycle assessment (LCA) methodology is used. LCA systematically assesses the environmental impacts induced by all resource uses and emissions over the whole life cycle of an activity (e.g. a product, process or a service) comprising the stages of resource extraction, production, use phase, and disposal [6, 50, 51]. More specifically, a hybrid life cycle assessment (LCA) framework is established for this dissertation's modeling approach. This means that life cycle environmental impacts are computed using two methodologically different approaches: Environmentally-extended input-output models (EEIOM, see e.g. [6, 10]) and process-based LCA (see e.g. [6, 50]). The first takes a top-down perspective and focuses on financial transactions between industry sectors of national economies or even between industry sectors of different countries in the case of multi-regional input-output models. It computes environmental impacts of final demand by coupling environmental accounts of economic sectors with these inter-sectoral financial flows. In contrast, process-based LCA goes in the direction of a bottom-up approach: Life cycle inventory databases provide detailed information on emissions and resource uses in up- and downstream processes of an activity under consideration. While the specificity of process-based LCA is much higher, EEIOM is usually more comprehensive in scope since all industry sectors of economy and their interrelationships are considered.

This assessment of environmental impacts shall also enable for the computation of different environmental indicators and not only GHG emissions. Considering a range of environmental pressures or a cumulative impact assessment method (e.g. ReCiPe [52]) may also help to avoid unintended shifts of environmental problems from one compartment to another.

1.4 STRUCTURE OF THE DISSERTATION

This PhD thesis is a cumulative dissertation and encompasses five scientific articles. Thereby, Chapters 2 to 5 constitute articles which are published in peer-reviewed scientific journals. Chapter 6 corresponds to a conference proceeding which was accepted in a peer-review process. In the beginning of each chapter, the information on the respective publication is indicated.

Chapter 2 summarizes our contribution in the transdisciplinary project *Zernez Energia 2020*. The municipality of Zernez, a small Swiss village in the Alps, initiated this project to find effective ways to decrease its greenhouse gas emissions. In the scope of the project, a thorough information database is established in a laborious data collection process. Based on this database and as a starting point for deriving measures and policies, the current carbon footprint of this rural community is assessed by employing both a consumption and a production perspective.

From Chapter 3 on, the modeling approach to provide a detailed information base for policymaking without antecedent excessive data collection is presented. Figure 1.1 supports the outline of the dissertation's modeling part by depicting which sub-model parts are described in which chapters. Because Chapter 2 identifies the building sector as a first step towards a reduction of greenhouse gas emissions for Zernez, **Chapter 3** starts with the evaluation of a bottom-up building energy model. On the one hand, the model's internal structure is scrutinized in the scope of a global sensitivity analysis. On the other hand, the detailed database of the research project in Zernez is used for an evaluation of the model results with primary data. Building upon these internal and external evaluations, **Chapter 4** aims to overcome flaws of the investigated building model by integrating comprehensive geographic information and developing new approaches for large-scale building stock modeling. Thereby, the uncertainty of model results is assessed in a Monte Carlo simulation framework. The improved building energy model is finally evaluated again and provides simulation of space heating as well as estimates for hot water production and electricity demand for residential buildings.

Chapter 5 elaborates a novel approach to study lifestyle-induced environmental impacts. By employing data mining techniques, consumption behavior patterns are studied and consumption-based archetypes are identified. These archetypes quantify household needs in the context of total household consumption. Furthermore, a hybrid LCA framework is established to assess environmental impacts associated with these archetypical behavior patterns.

In addition to the direct support to identify targeted environmental measures, the archetypes of Chapter 5 can also be used as building blocks of a large-scale bottom-up household consumption model. Such a model is described in **Chapter 6** and applied to the whole of Switzerland. Thereby, a probability-based classification approach is developed to assign the archetypes to actual households while interlinking the archetypes with the building energy model of Chapter 4 and a mobility model. The latter model estimates mobility demand of households based on the simulation results of an agent-based traffic simulation framework. Moreover, the hybrid LCA framework of

Chapter 5 is extended in order to also encompass the housing energy demands provided by the building energy model as well as the estimates of the household mobility behavior.

Finally, the conclusions in **Chapter 7** provide a synthesis of the whole dissertation. In addition, this chapter considers the dissertation in a broader context by discussing its scientific and practical relevance and finally gives an outlook on future research work.

1.5 REFERENCES

[1] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013.

[2] W. Steffen, K. Richardson, J. Rockström, S. Cornell, I. Fetzer, E. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M. Persson, R. Veerabhadran, B. Reyers, and S. Sörlin, *Planetary boundaries: Guiding human development on a changing planet*, *Science*, vol. 347, no. 6223, 2015.

[3] K. Schanes, S. Giljum, and E. Hertwich, *Low carbon lifestyles: A framework to structure consumption strategies and options to reduce carbon footprints*, *Journal of Cleaner Production*, vol. 139, pp. 1033–1043, 2016.

[4] C. M. Jones and D. M. Kammen, *Quantifying carbon footprint reduction opportunities for U.S. households and communities*, *Environ. Sci. Technol.*, vol. 45, no. 9, pp. 4088–4095, 2011.

[5] J. H. Spangenberg and S. Lorek, *Environmentally sustainable household consumption: From aggregate environmental pressures to priority fields of action*, *Ecological Economics*, vol. 43, no. 2–3, pp. 127–140, 2002.

[6] E. G. Hertwich, *Life Cycle Approaches to Sustainable Consumption: A Critical Review*, *Environ. Sci. Technol.*, vol. 39, no. 13, pp. 4673–4684, 2005.

[7] A. Tukker, M. J. Cohen, K. Hubacek, and O. Mont, *The Impacts of Household Consumption and Options for Change*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 13–30, 2010.

[8] A. Druckman and T. Jackson, *The bare necessities: How much household carbon do we really need?*, *Ecological Economics*, vol. 69, no. 9, pp. 1794–1804, 2010.

[9] G. Baiocchi, J. Minx, and K. Hubacek, *The Impact of Social Factors and Consumer Behavior on Carbon Dioxide Emissions in the United Kingdom. A Regression Based on Input-Output and Geodemographic Consumer Segmentation Data*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 50–72, 2010.

[10] J. C. Minx, T. Wiedmann, R. Wood, G. P. Peters, M. Lenzen, A. Owen, K. Scott, J. Barrett, K. Hubacek, G. Baiocchi, A. Paul, E. Dawkins, J. Briggs, D. Guan, S. Suh, and F. Ackerman, *Input-output analysis and carbon footprinting: An overview of applications*, *Economic Systems Research*, vol. 21, no. 3, pp. 187–216, 2009.

[11] United Nations Framework Convention on Climate Change (UNFCCC), *The Paris Agreement*, 2018. [Online]. Available: http://unfccc.int/paris_agreement/items/9485.php. [Accessed: 15-Feb-2018].

[12] Secretary-General (United Nations), *Progress towards the Sustainable Development Goals*, E/2017/66; United Nations: New York, 2017.

[13] D. O'Rourke and N. Lollo, *Transforming Consumption: From Decoupling, to Behavior Change, to*

System Changes for Sustainable Consumption, *Annual Review of Environment and Resources*, vol. 40, pp. 233–259, 2015.

[14] A. Ramaswami, A. Chavez, J. Ewing-Thiel, and K. E. Reeve, *Two Approaches to Greenhouse Gas Emissions Foot-Printing at the City Scale*, *Environ. Sci. Technol.*, vol. 45, no. 10, pp. 4205–4206, 2011.

[15] UN-HABITAT, *Cities and Climate Change: Global Report on Human Settlements*, London, UK: Earthscan, 2011.

[16] C. A. Kennedy, N. Ibrahim, and D. Hoornweg, *Low-carbon infrastructure strategies for cities*, *Nature Climate Change*, vol. 4, no. 5, pp. 343–346, 2014.

[17] C. Kennedy, J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havranek, D. Pataki, A. Phdungsilp, A. Ramaswami, and G. V. Mendez, *Methodology for inventorying greenhouse gas emissions from global cities*, *Energy Policy*, vol. 38, no. 9, pp. 4828–4837, 2010.

[18] C. Kennedy, J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havránek, D. Pataki, A. Phdungsilp, A. Ramaswami, and G. Villalba Mendez, *Greenhouse Gas Emissions from Global Cities*, *Environ. Sci. Technol.*, vol. 43, no. 19, pp. 7297–7302, 2009.

[19] R. Frischknecht, C. Nathani, S. Büsser Knöpfel, R. Itten, F. Wyss, and P. Hellmüller, *Entwicklung der weltweiten Umweltauswirkungen der Schweiz. Umweltbelastung von Konsum und Produktion 1996 bis 2011*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2014.

[20] H. N. Larsen and E. G. Hertwich, *Identifying important characteristics of municipal carbon footprints*, *Ecological Economics*, vol. 70, no. 1, pp. 60–66, 2010.

[21] E. G. Hertwich and G. P. Peters, *Carbon Footprint of Nations: A Global, Trade-Linked Analysis*, *Environ. Sci. Technol.*, vol. 43, no. 16, pp. 6414–6420, 2009.

[22] T. Hillman and A. Ramaswami, *Greenhouse Gas Emission Footprints and Energy Use Benchmarks for Eight Cities*, *Environ. Sci. Technol.*, vol. 44, no. 6, pp. 1902–1910, 2010.

[23] A. Ramaswami, T. Hillman, B. Janson, M. Reiner, and G. Thomas, *A Demand-Centered, Hybrid Life-Cycle Methodology for City-Scale Greenhouse Gas Inventories*, *Environ. Sci. Technol.*, vol. 42, no. 17, pp. 6455–6461, 2008.

[24] J. Lin, Y. Hu, S. Cui, J. Kang, and A. Ramaswami, *Tracking urban carbon footprints from production and consumption perspectives*, *Environmental Research Letters*, vol. 10, no. 5, 2015.

[25] A. Chavez and A. Ramaswami, *Articulating a trans-boundary infrastructure supply chain greenhouse gas emission footprint for cities: Mathematical relationships and policy relevance*, *Energy Policy*, vol. 54, pp. 376–384, 2013.

[26] T. M. Baynes and T. Wiedmann, *General approaches for assessing urban environmental sustainability*, *Current Opinion in Environmental Sustainability*, vol. 4, no. 4, pp. 458–464, 2012.

[27] A. Ramaswami, A. Chavez, and M. Chertow, *Carbon Footprinting of Cities and Implications for Analysis of Urban Material and Energy Flows*, *Journal of Industrial Ecology*, vol. 16, no. 6, pp. 783–785, 2012.

[28] T. Baynes, M. Lenzen, J. K. Steinberger, and X. Bai, *Comparison of household consumption and regional production approaches to assess urban energy use and implications for policy*, *Energy Policy*, vol. 39, no. 11, pp. 7298–7309, 2011.

[29] Bundesamt für Statistik (BFS), *Bruttowertschöpfung nach Branchen*, 2015. [Online]. Available: https://www.bfs.admin.ch/bfs/de/home/statistiken/querschnittsthemen/wohlfahrtsmessung/alle-indikatoren/wirtschaft/branchenstruktur.assetdetail.3442051.html. [Accessed: 19-Feb-2018].

[30] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[31] A. Tukker, T. Bulavskaya, S. Giljum, A. de Koning, S. Lutter, M. Simas, K. Stadler, and R. Wood, *The Global Resource Footprint of Nations: Carbon, water, land and materials embodied in trade and final consumption calculated with EXIOBASE 2.1*, Leiden (NL), Delft (NL), Vienna (AT), Trondheim (NO), 2014.

[32] D. Ivanova, K. Stadler, K. Steen-Olsen, R. Wood, G. Vita, A. Tukker, and E. G. Hertwich, *Environmental Impact Assessment of Household Consumption*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[33] A. Tukker and B. Jansen, *Environmental Impacts of Products: A Detailed Review of Studies*, *Journal of Industrial Ecology*, vol. 10, no. 3, pp. 159–182, 2006.

[34] J. Minx, G. Baiocchi, T. Wiedmann, J. Barrett, F. Creutzig, K. Feng, M. Förster, P.-P. Pichler, H. Weisz, and K. Hubacek, *Carbon footprints of cities and other human settlements in the UK*, *Environmental Research Letters*, vol. 8, no. 3, 2013.

[35] A. Druckman and T. Jackson, *Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model*, *Energy Policy*, vol. 36, no. 8, pp. 3177–3192, 2008.

[36] C. Jones and D. M. Kammen, *Spatial Distribution of U.S. Household Carbon Footprints Reveals Suburbanization Undermines Greenhouse Gas Benefits of Urban Population Density*, *Environ. Sci. Technol.*, vol. 48, no. 2, pp. 895–902, 2014.

[37] E. R. Frederiks, K. Stenner, and E. V. Hobman, *Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour*, *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1385–1394, 2015.

[38] B. Girod and P. De Haan, *GHG reduction potential of changes in consumption patterns and higher quality levels: Evidence from Swiss household consumption survey*, *Energy Policy*, vol. 37, no. 12, pp. 5650–5661, 2009.

[39] B. Girod and P. De Haan, *More or Better? A Model for Changes in Household Greenhouse Gas Emissions due to Higher Income*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 31–49, 2010.

[40] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[41] D. Saner, C. Beretta, B. Jäggi, R. Juraske, F. Stoessel, and S. Hellweg, *FoodPrints of households*, *The International Journal of Life Cycle Assessment*, vol. 21, no. 5, 2016.

[42] Oxford University Press, *Oxford Dictionaries*. [Online]. Available: https://www.oxforddictionaries.com/. [Accessed: 16-Feb-2018].

[43] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer Dordrecht, Heidelberg / New York / London, 2013.

[44] M. North, *Data Mining for the Masses*, A Global Text Project Book, 2012.

[45] Bundesamt für Statistik (BFS), *STATPOP 2013 - Statistik der Bevölkerung und Haushalte*, Neuchâtel, Switzerland, 2014.

[46] A. Horni, K. Nagel, and K. W. Axhausen, *The Multi-Agent Transport Simulation MATSim*, London, UK: Ubiquity Press, 2016.

[47] Bundesamt für Statistik (BFS) and Bundesamt für Raumentwicklung (ARE), *Mobilität in der Schweiz. Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2012.

[48] S. Hirschberg (ed.), C. Bauer, B. Cox, T. Heck, J. Hofer, W. Schenler, A. Simons, A. Del Duce, H.-J. Althaus, G. Georges, T. Krause, M. Vayá González, F. Ciari, R. A. Waraich, B. Jäggi, A. Stahel, A. Froemelt, and D. Saner, *Opportunities and challenges for electric mobility: An interdisciplinary assessment of passenger vehicles*, Villigen, Dübendorf and Zurich, Switzerland, 2016.

[49] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[50] S. Hellweg and L. Mila i Canals, *Emerging approaches, challenges and opportunities in life cycle assessment*, *Science*, vol. 344, no. 6188, pp. 1109–1113, 2014.

[51] R. Heijungs and S. Suh, *The Computational Structure of Life Cycle Assessment*, vol. 11, Dordrecht: Springer Netherlands, 2002.

[52] M. Goedkoop, R. Heijungs, M. Huijbregts, A. De Schryver, J. Struijs, and R. van Zelm, *ReCiPe 2008*, Ministrie van VROM: Den Haag, Netherlands, 2012.

CHAPTER 2

GREENHOUSE GAS EMISSIONS QUANTIFICATION AND REDUCTION EFFORTS IN A RURAL MUNICIPALITY

Andreas Froemelt*, *Manuela Mauchle, Bernhard Steubing and Stefanie Hellweg*

Journal of Industrial Ecology (2018) 22 (1) 92 – 105

* The individual contribution of Andreas Frömelt consisted of collecting and preparing the data, conducting the analyses and preparing the manuscript for publication.

SUMMARY

The present chapter aims to determine the current carbon footprint (CF) of Zernez, a Swiss mountain village, and to identify reduction potentials of greenhouse gas (GHG) emissions. For this purpose, material and energy flows were assessed mainly based on detailed household surveys, interviews and energy bills, but also by means of other information sources, for example, national statistics, traffic censuses, and literature values. To set up the GHG-balance, special attention was paid to the consistent definition of system boundaries by adopting two fundamentally different perspectives: purely geographical accounting (PGA) and the consumption-based footprint (CBF) method. Each of these two perspectives total approximately 10 tonnes of carbon dioxide equivalents per capita per year. The PGA revealed that 70% of the direct emissions in Zernez are caused by agricultural activities, whereas no consumption area dominated the consumption-induced CF. For the identification of targeted measures, both perspectives were considered in a complementary manner. The building stock and its underlying energy supply system showed a GHG reduction potential of 80%. The building sector was thus detected as a reasonable first step for the municipality to adopt GHG mitigation strategies. In the case of Zernez, building-stock-related measures are predicted to decrease the current CF by 13% (CBF) and 17% (PGA) respectively.

2.1 INTRODUCTION

Urban and rural settlements constitute the heart of human activities and are thus central sources of anthropogenic greenhouse gas (GHG) releases. Local initiatives to reduce GHG emissions from human settlements are essential to abate climate change (UN-HABITAT 2011 [1]). In view of the important role that GHG accounting plays in the planning process of mitigation measures, a large body of carbon footprint (CF) studies on different scales has evolved, among them Kennedy and colleagues (2009, 2010, 2014) [2–4], Frischknecht and colleagues (2014) [5], Minx and colleagues (2013) [6], Goldstein and colleagues (2013) [7], Jungbluth and colleagues (2011) [8], Larsen and Hertwich (2010) [9], Hertwich and Peters (2009) [10], Hillman and Ramaswami (2010) [11], and Ramaswami and colleagues (2008) [12]. Most of these studies concentrate on nations or cities. Given that around half of the global population currently lives in urban areas, with this share rapidly increasing in the future (UN 2012 [13]), many authors (e.g. Kennedy et al. 2009, 2010, 2014 [2–4]; Goldstein et al. 2013 [7]; Baynes et al. 2011 [14]) emphasize that a focus on cities is crucial. Whereas the present and future importance of urban settlements is unquestionable, it should not be overlooked that a large part of the world's population continues to live in the countryside. This is also true for both developing and industrialized nations. For instance, according to the "Degree of Urbanization (DEGURBA)"-definition of the European Union (Eurostat 2016 [15]), approximately 65% of all Swiss municipalities are classified as "thinly populated" and are home to about 25% of the Swiss population (BFS 2014a, 2014b [16, 17]). However, CF studies and local GHG mitigation initiatives in rural contexts are, to the best of our knowledge, less common (Minx et al. 2013 [6]).

All of the abovementioned CF studies had to deal with an important question: how to set the system boundaries to adequately account for GHG emissions. This is an ongoing debate and different frameworks have emerged. Lin and colleagues (2015) [18], Chavez and Ramaswami (2013) [19], Baynes and Wiedmann (2012) [20], and Ramaswami and colleagues (2011, 2012) [21, 22] provide comprehensive overviews of state-of-the-art carbon balancing approaches. According to these studies, there are basically two fundamentally different accounting perspectives (some aforementioned articles distinguish more): the consumption-oriented carbon footprint and the production-based approach. The first corresponds to a life cycle perspective and focuses on global GHG releases induced by the consumption behavior of the inhabitants living within the study area. In contrast, emissions related to activities and processes taking place within the study boundaries are the center of focus in production-based accounting schemes. The narrowest scope of a production-based accounting method accords with a territorial approach and concentrates solely on direct emissions within the geographical boundaries of the study area. This balancing scheme is often encountered in national-scale footprints and conforms to the so-called Scope 1-perspective of the World Resources Institute/World Business Council for Sustainable Development (WRI/WBCSD) (2004) [23]. Some studies (Ramaswami et al. 2008 [12]; Kennedy et al. 2009, 2010, 2014 [2–4]; Hillman and Ramaswami 2010 [11]; Lin et al. 2015 [18]) broaden this territorial inventory viewpoint to a “geographic-plus” (Ramaswami et al. 2011 [21]; Baynes and Wiedmann 2012 [20]) or “trans-boundary community-wide infrastructure” footprint (Chavez and Ramaswami 2013 [19]; Lin et al. 2015 [18]) by including Scope 2 (indirect emissions from electricity production) and Scope 3 emissions (indirect emissions from other key flows into the study area) (WRI/WBCSD 2004 [23]). By concentrating on key infrastructure, this approach is especially useful for urban planners and facility managers.

The choice of a GHG accounting scheme depends on the question one seeks to answer. However, it is important to be consistent within the chosen approach. As stated by several researchers (Ramaswami et al. 2011, 2012 [21, 22]; Baynes et al. 2011 [14]; Baynes and Wiedmann 2012 [20]; Lin et al. 2015 [18]), consumption-based and production-based accounting methods are complementary to each other. The consumption perspective supports the identification of measures without problem shifting by pursuing a life cycle perspective. Moreover, it facilitates the derivation of policies aimed at influencing consumption behavior (Baynes and Wiedmann 2012 [20]). By quantifying direct emissions within the geographical system boundaries, a purely territorial inventory pinpoints areas of action for local authorities since it takes into account the local conditions. Both accounting frameworks hence provide important insights for local policy making from different viewpoints.

The Swiss mountain village of Zernez initiated the research project *Zernez Energia 2020* to identify ways to reduce its GHG emissions using both a consumption-based footprint and a purely geographical accounting. The assessment of the municipality’s CF is an important first step in the project and is presented in this chapter. The analysis of the current CF allows for identifying the most relevant actors and sectors regarding GHG emissions as well as for discussing GHG reduction potentials. It also serves as a planning basis for the municipality of Zernez with the aim of developing a concrete action plan to implement GHG mitigation measures. This study can provide a baseline for similar studies in other municipalities and fills a knowledge gap by delivering insights into the CF of a village in an industrialized country. Additionally, there are advantages in focusing on a small village. Whereas many large-scale studies struggle with data scarcity, the limited size of Zernez allows, for example, for data collection through detailed surveys of all households and buildings in the village. Surveys with 100% coverage, as in the present study, are extremely rare. Therefore, the present study provides also a unique occasion for an in-depth analysis of a rural community.

The goal of the present chapter is twofold: it aims to assess the CF of Zernez as well as to discuss its GHG reduction potentials. Moreover, it intends to provide an insight into a rural area's CF based on a unique data set.

2.2 METHODOLOGY

2.2.1 System Boundaries

The municipality of Zernez, which was home to 1,140 inhabitants (BFS 2014c [24]) in the year 2010, is situated in the Swiss Alps at an average altitude of 1,471 meters with an area of 203.85 square kilometers (see Appendix A) (swisstopo 2014 [25]). Hosting large parts of the Swiss National Park, Zernez belongs to the largest municipalities in Switzerland in terms of land area.

The geographical system boundaries focused on the settlement area of the village's core with 1,003 residents. Additionally, all agricultural and forest areas within the municipal borders and managed by these persons or by the municipality were included in the analysis. Hamlets and remote single houses located within the municipality's borders as well as the Swiss National Park were excluded from consideration.

In view of the specific benefits of the two GHG accounting schemes, the CF of Zernez was assessed simultaneously by a production-based territorial inventory (only Scope 1/direct emissions) and from a consumption-based perspective (life cycle emissions from final consumption). Following the definitions given by Lin and colleagues (2015) [18], these applied accounting methods will be called henceforth "purely geographical accounting" (PGA) and "consumption-based footprint" (CBF), respectively.

Computations and analyses of the present chapter refer, whenever possible, to the year 2010.

2.2.2 Overview of the Applied Greenhouse Gas Accounting Approach

In a preparatory step, all processes, activities, materials, and energy flows that are relevant for the municipality's CF were identified. This systematic identification process was conducted in close collaboration with the municipality and built upon similar studies (Jungbluth et al. 2011, 2012 [8, 26]; BFS 2014d [27]). All processes and flows were then classified into nine categories (left part of Figure 2.1).

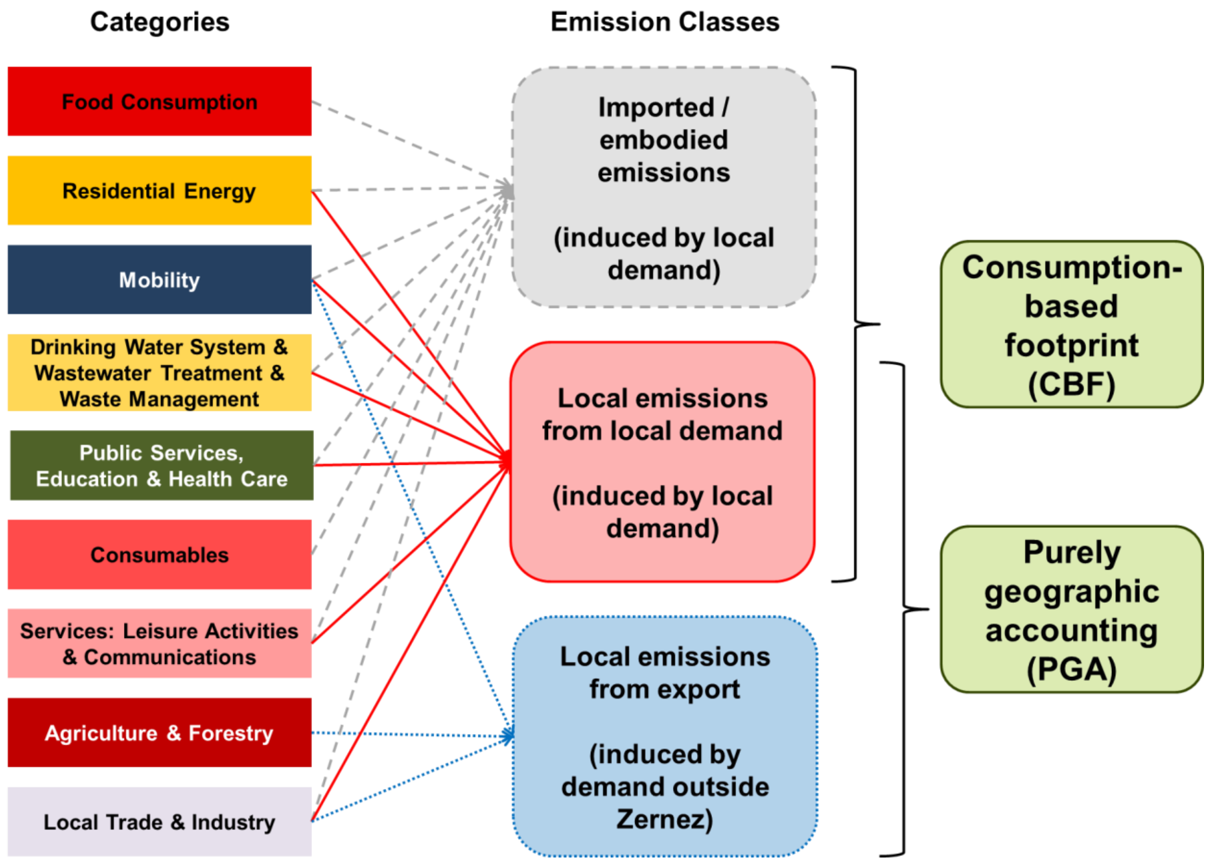


Figure 2.1: **Schematic overview of the allocation of GHG emissions from different categories to emission classes ("imported emissions", "local emissions from local demand" and "local emissions from export"). The figure also shows how emission classes add up to CBF and PGA. [GHG = greenhouse gas]**

In order to set up the GHG balance according to CBF and PGA, a three-tiered approach was applied. First, the magnitude of all flows belonging to the above-mentioned nine categories was quantified. For example, in the category *residential energy*, this includes the amounts of electricity and fuels needed to provide buildings with heat. Second, the GHG emissions of these quantified entities were estimated based on life cycle inventory (LCI) databases, such as ecoinvent (ecoinvent Center 2013 [28]). Thereby, the CF of an activity is assessed on the basis of the global warming potentials published by the Intergovernmental Panel on Climate Change (IPCC) (2007) [29] for a time horizon of 100 years.

In the third step, the GHG emissions of each of the modeled activities were classified into the following three *emission classes* (Figure 2.1, center): emissions outside Zernez caused by the consumer demand of the inhabitants ("imported/embodied emissions"); emissions released within the geographical system boundaries that occur on account of the consumption behavior of the people in Zernez ("local emissions from local demand"); and emissions released within the geographical system boundaries, but induced by the demand of consumers outside Zernez ("exported emissions").

The use of background data from LCI databases results in life cycle GHG emissions for the modeled processes. For the aforementioned classification into emission classes, which is schematically illustrated in Figure 2.1, this assessment of life cycle GHG emissions had thus to be split into local emissions and emissions occurring upstream in the supply chain. Thereby, the latter is allocated to the embodied emissions given that these emissions occur outside the village. However, imported/embodied emissions do not include re-exported embodied emissions because these GHG are not induced by the inhabitants of Zernez. An overview of the most important assumptions with regard to this classification can be found in Appendix A.

The distinction into three emission classes allowed for appropriately summing up to CBF and PGA, respectively (Figure 2.1, right).

All of these three steps (quantification of flows, assessment of GHG emissions, and allocation of emissions to emission classes) are explained in more detail in the subsequent section. Note that, in this chapter, "imports" and "exports" do not refer to a national level, but describe goods entering and leaving the municipal boundaries.

2.2.3 Quantification and Modeling of Activities

A bottom-up approach was followed in the data collection for the quantification of material and energy flows. Within the scope of the project, surveys and interviews were conducted and energy bills and municipal statistics were gathered. A large part of the collected data was compiled into a building database containing comprehensive information about energy consumption as well as specific data for all buildings in the municipality (Wagner et al. 2015a [30]). This database is unique in terms of full coverage of buildings and represents one of the most important data sources for the present study. The combination of this database with energy bills and a census of enterprises (see Appendix A) provided data on final energy demand of all households and enterprises in the village. The building database scheme as well as further details on other bottom-up data used for the present study (e.g., operation information of district heating network, waste statistics, and information from the forestry administration) are described in Wagner and colleagues (2015a) [30] and in Appendix A. In addition to these data, further interviews were conducted to gain detailed information on the operation of specific facilities in the municipality (e.g., a biogas plant). The whole data set was finally amended by statistics of federal and cantonal offices, federal surveys, traffic censuses, and literature values.

After quantifying the energy and material flows within the municipality, GHG emissions were assessed by adding suitable background data from process-based LCI databases (e.g., ecoinvent Center 2013 [28]).

Table 2.1 provides a brief overview of the used data sources. Further, a detailed list of all modeled activities and data sources used for the estimation of GHG emissions as well as the most important assumptions for the subdivision of emissions into emission classes are documented in Appendix A.

Table 2.1: **Overview of the data sources used for the quantity and GHG assessment in the different categories.**

Category	Quantity assessment: Data source	Quantity assessment: Scale of data	GHG assessment (background data)
Food consumption	HBS (BFS 2012 [31])	National[a]	Saner et al. (2016) [32]
	GastroSuisse (2012) [33]	National[a]	ecoinvent Center (2013) [28]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	
Residential energy	Building Database (Wagner et al. 2015a [30])	Zernez	ecoinvent Center (2013) [28]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	
	OI DHN (Wagner et al. 2015a [30])	Zernez	
	Forestry Administration (Wagner et al. 2015a [30])	Zernez	
	OI for Biogas Plant (Grass 2014 [34])	Zernez	
Mobility	Mobility and Transport Microcensus (ARE et al. 2012 [35])	Cantonal[b]	ecoinvent Center (2013) [28]
	Local Age Structure (AWT 2010 [36])	Cantonal[c]	
	Automatic Traffic Counters (TBA 2011 [37], ASTRA 2014 [38])	Zernez	
	Timetables of PT (PostAuto 2014 [39], SBB 2014 [40])	Zernez	
Drinking water and wastewater system and waste management	Building Database (Wagner et al. 2015a [30])	Zernez	ecoinvent Center (2013) [28]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	
	OI WWTP (Filli 2014 [41])	Zernez	
Public services and health care	Building Database (Wagner et al. 2015a [30])	Zernez	ecoinvent Center (2013) [28]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	Jungbluth et al. (2011) [8]
	OI DHN (Wagner et al. 2015a [30])	Zernez	
	Census of Enterprises (see Appendix A)	Zernez	
	Jungbluth and colleagues (2011) [8]	National[d]	
Consumables	Jungbluth and colleagues (2011) [8]	National[d]	Jungbluth et al. (2011) [8]
Services: Leisure activities and communications	Building Database (Wagner et al. 2015a [30])	Zernez	ecoinvent Center (2013) [28]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	König et al. (2014) [42]
	OI DHN (Wagner et al. 2015a [30])	Zernez	Jungbluth et al. (2011) [8]
	Mobility and Transport Microcensus (ARE et al. 2012 [35])	Cantonal[b]	
	Survey of Travels of Swiss Residents (BFS 2011 [43])	National[a]	
	Local Age Structure (AWT 2010 [36])	Cantonal[c]	
	OI Swiss Post (Saner 2014 [44])	Zernez	
	Census of Enterprises (see Appendix A)	Zernez	
	Jungbluth and colleagues (2011) [8]	National[d]	
Agriculture and forestry	Census of Enterprises (see Appendix A)	Zernez	ecoinvent Center (2013) [28]
	Farm Structure Survey (BFS 2014a [24])	National[c]	Nielsen et al. (2003) [45]
	Forestry Administration (Wagner et al. 2015a [30])	Zernez	
Local Trade & Industry	Building Database (Wagner et al. 2015a [30])	Zernez	ecoinvent Center (2013) [28]
	Census of Enterprises (see Appendix A)	Zernez	Jungbluth et al. (2011) [8]
	Electricity Bills (Wagner et al. 2015a [30])	Zernez	
	OI DHN (Wagner et al. 2015a [30])	Zernez	
	OI Gravel Quarry (SOSA GERA SA 2014 [46])	Zernez	
	OI Swiss Post (Saner 2014 [44])	Zernez	
	Jungbluth and colleagues (2011) [8]	National[d]	

[a]Adjusted for Zernez (e.g., by taking local age structure into account; see text for more information).
[b]Adjusted to represent the canton's rural areas and adjusted for Zernez by taking into account the local age structure.
[c]Specifically retrieved for Zernez from a national or cantonal database.
[d]Scaled down to Zernez on a per-capita basis
HBS = Swiss Household Budget Survey; OI = operation information; DHN = district heating network;
PT = public transportation; WWTP = wastewater treatment plant.

In addition, an independent alternative approach was followed by applying Saner and colleagues' (2013) [47] model to Zernez in order to assess the consumption-related CF in the areas of residential energy and mobility (see Appendix A). This model is able to determine housing and land-based mobility demand of individual households. Because it is mainly based on national databases and simulation results of MATSim (Multi-Agent Transport Simulation) (Balmer et al. 2006 [48]; Meister et al. 2010 [49]), no extra data acquisition is needed beforehand. MATSim, which is the basis for the mobility sub-model, was calibrated and validated for Switzerland (Balmer et al. 2006 [48]; Meister et al. 2010 [49]), whereas the housing sub-model was evaluated by Froemelt and Hellweg (2017) [50] (see Chapter 3).

2.2.3.1 Food Consumption

All liquid and solid food has to be "imported" from outside the municipality given that no industrial food processing takes place in Zernez. Because sales data of the two largest local retailers could not be accessed, food consumption was estimated by means of the Swiss Household Budget Survey (HBS) (BFS 2012 [31]). Among other things, the HBS provides detailed insight into the quantities of more than 100 different food categories bought by an average Swiss household. This also includes farm gate sales and food produced in gardens. These two food production processes were allocated to "imported/embodied emissions" instead of "local emissions from local demand" because they were assumed to play a minor role and detailed data were lacking. Food consumed in restaurants was estimated by combining the HBS with a survey of the Association for Hotels and Restaurants in Switzerland (GastroSuisse 2012 [33]) and the menu cards of local restaurants.

Preparation of food at home or in the restaurants is not accounted for in this category, but covered by the categories residential energy and services: leisure activities and communications (see corresponding sections below).

The life cycle GHG emission factors for food consumption were retrieved from Saner and colleagues (2016) [32], who consolidated different databases and studies (e.g., ecoinvent Center 2013 [28]; Thrane 2006 [51]; Büsser and Jungbluth 2009 [52]; Stoessel et al. 2012 [53]). However, these factors include neither transport to the shops nor electricity and space heating demand of the shops. These processes were additionally modeled by electricity bills of the local retailers and by transport distances from the food distribution centers of the respective company to the shop in Zernez.

2.2.3.2 Residential Energy

In the scope of the research project *Zernez Energia 2020*, a survey was conducted in order to collect data on the quantities of energy carriers used, including fuel oil, wood chips, firewood, district heating, and electricity. Besides interviews with all households, the municipality provided detailed electricity and district heating energy bills for all households and enterprises. In addition to final energy consumption data, the municipality and the inhabitants supplied also detailed information on installed heating systems and on different building characteristics. This enabled the setup of a unique database with full coverage of all buildings (Wagner et al. 2015a [30]). More information on this building database as well as on data from electricity bills and the district heating network is presented in Appendix A.

The municipality draws its electricity from two main sources: a power company that delivers certified electricity from a run-of-river power plant through the national grid and a reservoir power plant located partly in the municipality. The latter is therefore regarded as locally produced energy. Local electricity is also produced by photovoltaic (PV) systems and a biogas plant; however, this power is exported to the national grid and, consequently, only considered in the PGA. Using a combined heat and power plant, the biogas plant operating on agricultural waste produces heat in addition to electricity. Detailed plant operation information was provided by the owner (Grass 2014 [34]).

Data on wood chip supplies for the local district heating network were provided by the operators, the municipality, and the forestry administration (see section 2.2.3.8 (*Agriculture and Forestry*) below as well as Appendix A). Imported firewood was calculated as the difference between used amounts of log wood and the amount of firewood sold by the forestry administration (Wagner et al. 2015a [30]).

In order to model the energy supply for residential buildings in detail, suitable ecoinvent processes were chosen and adjusted (ecoinvent Center 2013 [28].). The building database, as well as the information above, was also used to model building-energy-related parts of the categories *public* se*rvices*, *education, and health care*, *services*: *leisure activities and communications*, and *local trade and industry* (see sections below).

2.2.3.3 Mobility

The Mobility and Transport Microcensus of the canton of Grisons provides comprehensive insight into the mobility behavior of the canton's population (ARE et al. 2012 [35]). Because the municipality of Zernez is located in a rural area of Grisons, the canton's only region with an urban character (the city of Chur) was separated and removed from the results of the Microcensus. Details on chosen traffic modes, distances driven, as well as purpose and character of trips were deduced for different age cohorts representing the canton's rural population. To set up the carbon balance for the CBF, these results were applied to Zernez by taking into account the local age structure (AWT 2010 [36]).

A totally different approach had to be chosen for the PGA. In this view, the mobility demands of the inhabitants are not of interest, but instead the traffic volume within the municipality's borders. For this purpose, automatic traffic counters (operated by the cantonal administration) were analyzed (TBA 2011 [37]; ASTRA 2014 [38]). Public transportation within the municipality was assessed by means of timetables and route information (PostAuto 2014 [39]; SBB 2014 [40]).

For the allocation of GHG to emission classes according to Figure 2.1, both approaches needed to be combined. The information provided by the Microcensus includes data on distances, duration, traffic mode, as well as purpose and number of trips. This allowed for estimations of trips within the municipal geographical boundaries and therefore to assess "local emissions from local demand." These emissions were then subtracted from the above-described total mobility CBF and total mobility PGA, leaving the remainder as the "imported/embodied emissions" and the "local emissions from export," respectively.

2.2.3.4 Drinking Water System, Wastewater Treatment and Waste Management

The municipality's drinking water system is composed of two spring water catchments, one groundwater pumping station and one reservoir to which the water has to be pumped. No further treatment steps are needed. Together with the electricity demand of the pumps in 2010, the drinking water system could be modeled with ecoinvent activities (ecoinvent Center 2013 [28]).

In order to represent the wastewater treatment plant (WWTP), ecoinvent activities (ecoinvent Center 2013 [28]) were adjusted and remodeled according to interviews with operators and operating data (Wagner et al. 2015a [30]; Filli 2014 [41]). Biogas produced in an anaerobic digester was used to satisfy parts of the WWTP's heating demand. Usually, excess biogas is converted into electricity and exported to the national grid. However, in the year 2010, the WWTP heating demand exceeded biogas production and was supplemented by burning fuel oil.

Detailed data were available to represent the waste management system (see Appendix A) (Wagner et al. 2015a [30]). Municipal solid waste (MSW), cardboard, paper, and waste glass are separated by the citizens and then collected regularly by a garbage truck. After trans-shipping to a freight train, MSW is transported to an MSW incinerator, whereas cardboard, paper, and waste glass are conveyed to recycling processes. Background data were taken from ecoinvent (ecoinvent Center 2013 [28]).

2.2.3.5 Public Services, Education and Health Care

Public services subsume all services from governmental institutions, excluding waterworks, wastewater system, and waste management (already covered by the section above). Local emissions stemming from the operation of public buildings could be assessed based on the building database (cf. section 2.2.3.2 (*Residential Energy*)) and a census of enterprises (cf. Appendix A). However, many of the emissions from the public sector, for example, those caused by the operation of federal offices or by the national defense, have to be allocated to all Swiss inhabitants. Therefore, Swiss average values were applied from Jungbluth and colleagues (2011) [8] to cover remaining emissions of this category, all of which were assumed to take place outside the village. Swiss average values were also used for emissions from education and health care (Jungbluth et al. 2011 [8]).

2.2.3.6 Consumables

Very little data were available for consumables such as clothing, furniture, and domestic appliances. Just as comestible goods, these products are mostly imported to Zernez. Therefore, Swiss average values retrieved from Jungbluth and colleagues (2011) [8] were used for this consumption area.

2.2.3.7 Services: Leisure Activities and Communications

This section encompasses all services consumed that are not already covered in other categories. For instance, recreational traffic is accounted for in *mobility*, *food consumption* includes eating out, and the *consumables* category already accounts for leisure equipment.

An estimation of the residents' leisure behavior was facilitated by the Microcensus (ARE et al. 2012 [35]) from which information on purpose and destination of the residents' trips could be retrieved. A combination of this information with the estimation of trips within the geographical system boundaries (see section 2.2.3.3 (*Mobility*)), a census of enterprises (cf. Appendix A), and further interviews (e.g., with the operators of the indoor hot spa swimming pool regarding the number of sold tickets) allowed for a rough assessment of leisure services consumed within Zernez by residents and thus the computation of "local emissions from local demand."

Vacations are an important area not included in other categories. The Swiss Federal Statistical Office published data on the travel patterns of the Swiss population, including number of trips, duration of trips, destinations, and number of hotel overnight stays (BFS 2011 [43]). Because this statistical analysis was performed for different age groups, the data could be used to determine the vacations of people from Zernez by considering the local age structure (AWT 2010 [36]). A study that assessed GHG emissions for different vacation scenarios (König et al. 2014 [42]) was adjusted and used for GHG intensity factors for hotel overnight stays.

Further services considered in this category are newspaper consumption, which was modeled by ecoinvent activities (ecoinvent Center 2013 [28]) and data from the Swiss Post (Saner 2014 [44]), as well as communications. For the latter, Swiss average values from Jungbluth and colleagues (2011) [8] were applied because no specific data were available for Zernez and because it is plausible to assume that people in Zernez behave similarly and cause similar emissions in terms of communications as the Swiss average citizen.

2.2.3.8 Agriculture and Forestry

Just as in most other Alpine regions, livestock farming dominates Zernez's agricultural sector.

Arable farming was negligibly small (total arable area in Zernez is only 4.2 hectares) (BFS 2014c [24]) and not further considered. A list of all animals kept in the municipality was obtained from the Farm Structure Survey of the Swiss Federal Statistical Office (BFS 2014c [24]). Building upon this information and in combination with a census of enterprises (cf. Appendix A), the agricultural sector was modeled by adjusting appropriate processes of ecoinvent (ecoinvent Center 2013 [28]) and the life cycle assessment food database (Nielsen et al. 2003 [45]).

The municipal forestry administration provided detailed data on sales of firewood, trunk wood, wood chips, and other unspecified wood (see Wagner et al. [2015a] [30] and Appendix A). Ten percent of the trunk wood is processed by the forestry administration themselves, whereas the rest is sent to different sawmills in Switzerland, Austria, and Italy. Operating records on amounts of diesel and petrol used were available. The forestry sector was then modeled by adapting ecoinvent activities (ecoinvent Center 2013 [28]) accordingly.

2.2.3.9 Local Trade and Industry

According to the census of enterprises (cf. Appendix A), local trade comprises mostly restaurants, hotels, offices, and local retailers. Direct emissions induced by tourism, an important economic sector for Zernez, are thus accounted for in this category. Most emissions caused by the aforementioned industries are determined by modeling the respective energy use based on combining the building database (cf. section 2.2.3.2 (*Residential Energy*)) with the census of enterprises (cf. Appendix A).

Two important enterprises, a large gravel quarry and the office of the Swiss Post, were assessed in more detail. For the post office, detailed data regarding information about amounts of letters, parcels, transport distances, and means of transportation were available from interviews with the sustainability department of the Swiss Post (Saner 2014 [44]). This allowed for a detailed modeling of the Post's activities. For modeling the gravel quarry (SOSA GERA SA 2014 [46]), an appropriate ecoinvent activity (ecoinvent Center 2013 [28]) was chosen.

Housing construction is also included in this category. Though local companies are often responsible for on-site building construction, most emissions from these activities are generated in the supply chain outside of the municipal geographical boundaries through the production of construction materials. Because of a lack of more-detailed data, the modeling of housing construction was based on Swiss average values (Jungbluth et al. 2011 [8]). Building upon the findings of Zhang and colleagues (2013) [54], we assumed that 10% of these construction emissions are released locally.

2.3 Results and Discussion

Figure 2.2 (top) shows the split of the total GHG into embodied emissions, direct emissions induced by local demand, and exported emissions. These emissions are then summed appropriately in order to represent the CBF and the PGA, respectively. In Figure 2.2 (bottom), the results for these two accounting methods were divided by the number of inhabitants. This normalization facilitates the comparison of the results with other CF studies.

Even though the two viewpoints are fundamentally different, the result of the consumption perspective (9.9 tonnes of carbon dioxide equivalents per person per year [t CO_2-eq/(cap·yr)]) is – by coincidence – similar to the production perspective (10.6 t CO_2-eq/(cap·yr)) in the case of Zernez. Thereby, embodied emissions, taking place outside the municipal boundaries of Zernez, account for around 85% of the total consumption GHG, whereas exported emissions represent also approximately 85% of the total territorial GHG emissions. These figures confirm that this small Alpine village is not self-sustaining, but highly integrated in the Swiss and global economy.

2.3.1 Purely Geographical Accounting

The PGA quantifies direct emissions in Zernez and therefore identifies the most important local GHG emitters. Figure 2.2 (bottom) reveals that the *agriculture and forestry* sector clearly dominates the PGA and causes 70% of the direct GHG releases. Thereby, agricultural activities are responsible for more than 99% of the total GHG emissions in this category, especially attributed to large enteric methane (CH_4) emissions and nitrous oxide (N_2O) releases stemming from livestock farming. This means that the territorial and the consumption perspective differ also in the composition of GHGs. Whereas the former shows a large share of CH_4 and N_2O emissions, the latter is rather dominated by CO_2 releases.

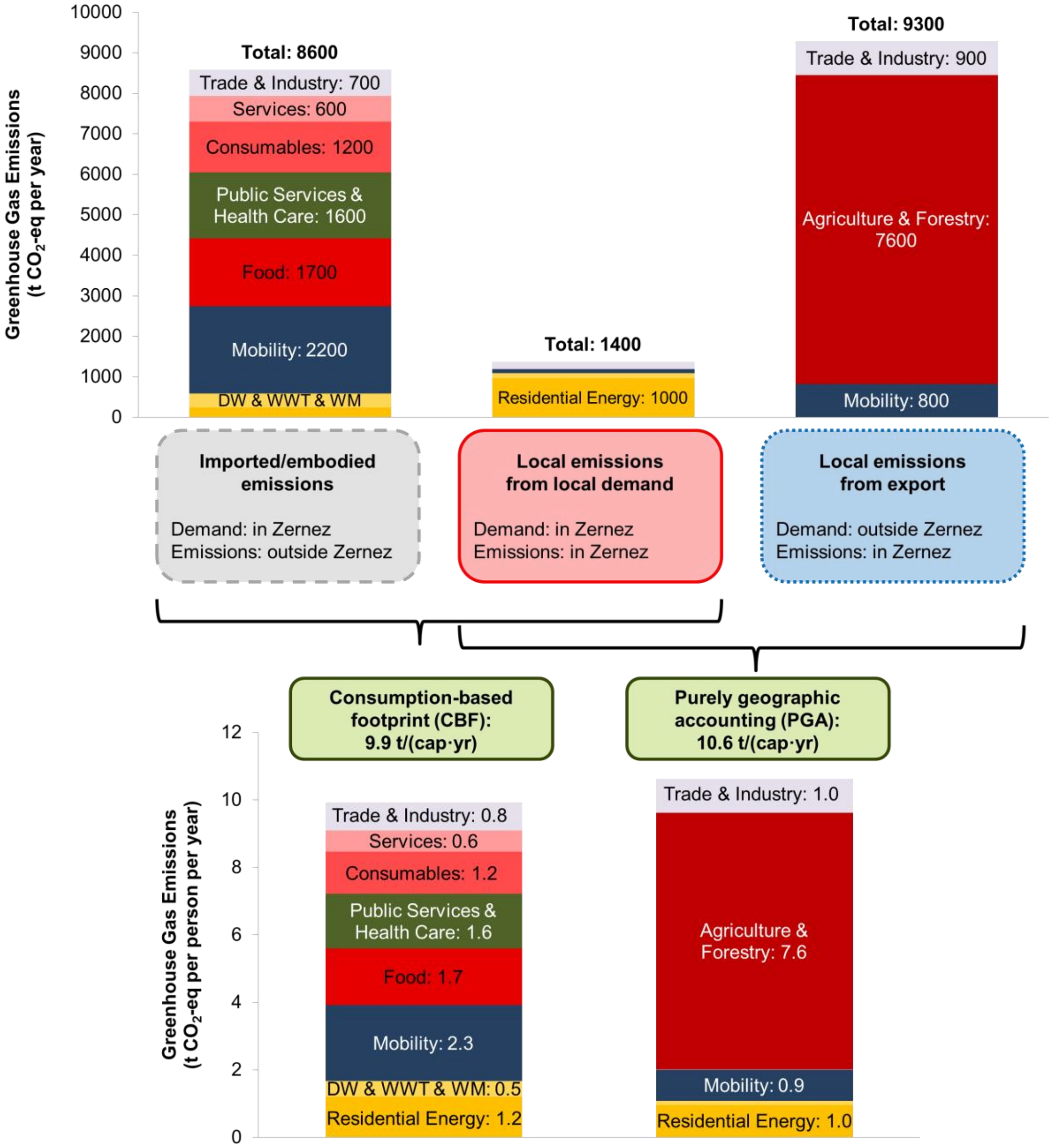


Figure 2.2: **Top: Total GHG emissions divided into "embodied emissions," "local emissions from local demand," and "exported emissions." Bottom: Per-capita carbon footprint of Zernez according to the CBF and the PGA, respectively. Values are rounded to 100 t CO_2-eq/yr or 0.1 t CO_2-eq/(cap·yr). The category *services* includes leisure activities and communications. [DW & WWT & WM = drinking water system, wastewater treatment, and waste management; GHG = greenhouse gas; t CO_2-eq/yr = tonnes of carbon dioxide equivalents per year; t CO_2-eq/(cap·yr) = tonnes of carbon dioxide equivalents per capita per year]**

2.3.2 Consumption-Based Footprint

The CBF provides a life cycle perspective of the consumer behavior. It encompasses all GHG emissions worldwide induced by the demand of the people living within the geographical system boundaries. The consumption-caused total life cycle GHG emissions in Zernez are approximately 22% below the Swiss average of 12.8 t CO_2-eq/(cap·yr) according to Jungbluth and colleagues (2012) [26] (please see Appendix A for a graphical comparison of Zernez's CBF with the national CBF of Switzerland).

Though *mobility* spearheads GHG emissions, all categories are of similar importance (see Figure 2.2). Many studies (Jungbluth et al. 2007, 2011, 2012 [8, 26, 55]; Hertwich and Peters 2009 [10]) identify food consumption, mobility, and housing (here, the sum of *residential energy* and *drinking water system*, *wastewater treatment*, *and waste management*) as the most important consumption domains from an environmental viewpoint. These three categories are responsible for more than 50% of Zernez's GHG emissions; however, they do not outstrip the other categories.

Mobility GHG emissions in Zernez (2.3 t CO_2-eq/(cap·yr)) correspond well to the Swiss average of 2.4 t CO_2-eq per person per year (Jungbluth et al. 2012 [26]). Motorized private transportation is responsible for more than 70% of the Swiss mobility emissions; however, its contribution in Zernez amounts to more than 85%.

Despite the prevailing cold climate (mean annual air temperature is approximately 5°C) (Orehounig et al. 2014 [56]), housing-induced GHG emissions (comprising the categories of *residential energy* and *drinking water system*, *wastewater treatment*, *and waste management*) in Zernez average out to 1.7 t CO_2-eq/(cap·yr) and are considerably below the national mean of 3.0 t CO_2-eq/(cap·yr) (Jungbluth et al. 2012 [26]). This large deviation can be explained by two main reasons: First, oil boilers cause around 90% of the *residential energy* GHG, but only supply around 40% of the local heat demand for space heating and hot water. A large amount of the local heat demand is covered by wood-based and direct electric heating systems (around 30% and 20%, respectively), both of which involve low GHG emissions. Second, electricity is exclusively supplied by hydropower plants, which exhibit very low carbon intensities.

Part of the deviation of Zernez's CF from the national average may also be explained by different methodological approaches, such as different GHG emission factors or different reference years. The areas of housing and mobility were double-checked with the model of Saner and colleagues (2013) [47] and revealed to be consistent with the model results (for further results, see Appendix A).

2.3.3 Identification of Greenhouse Gas Reduction Potentials

Although the political system in Switzerland grants some liberties to municipalities, their ability to implement measures to reduce GHG emissions is nevertheless limited. They may extend their own infrastructure, provide advice to consumers in different areas, and are free to set financial incentives. But taxes or mandatory measures are usually introduced at national or cantonal levels and only exceptionally at the municipal stage. Making laws stricter is possible for municipalities in some domains, but depends strongly on the canton. Facing this restrictive scope of action, areas susceptible to municipal policies, consumer-driven areas, and areas of supraregional importance are distinguished in order to discuss sectors where the municipality could intervene. On account of their complementary nature, the following discussion examines simultaneously both footprints, PGA and CBF.

2.3.3.1 Areas Susceptible to Municipal Policies

The municipality has the ability to reduce *mobility* emissions through awareness-raising campaigns (e.g., promoting the use of bikes or public transportation or the avoidance of air travels), financial incentives (e.g., for energy efficient and electric vehicles), or an extension of public transportation possibilities. However, decreasing GHG emissions through improved public transport services needs a careful optimization of timetables in order to increase the occupancy of trains and buses. This requires a supraregional perspective, usually lying with the respective railway/bus companies and not directly with the municipality. Implementation of a car sharing system could reduce GHG stemming from commuting.

In contrast, technical options for the building stock are available and can be implemented at the municipal level for both the building energy supply and the building stock materials. Possible measures comprise the extension of the existing district heating network or financial support for private refurbishment initiatives aimed at both replacing fossil-fuel-based heating systems or at decreasing heating demand by improved insulation. The possibility of tightening the regulations for building design to some extent might also be taken into consideration. Changing the building energy system could affect around 70% of the direct emissions from local *residential energy* demand (see Figure 2.2), but also affects embodied, direct, and exported GHG in the areas of *services: leisure activities and communications*, *local trade and industry* (e.g., hotels, restaurants, and offices), as well as *public services*, *education*, *and health care*.

2.3.3.2 Consumer-Driven Areas

Consumer behavior dictates large parts of the GHG emissions in different areas of the CBF, including *mobility*, *food consumption* (nutrition), *consumables*, *services: leisure activities and communications*, and *residential energy*. In view of Figure 2.2, a special focus should be laid on the areas of *mobility*, *food consumption*, and *residential energy*.

Influencing and changing behavioral patterns can be challenging, but will be necessary in order to achieve a sustainable level of GHG emissions. Some options, such as raising awareness and financial incentives, have already been mentioned. Raising awareness of sustainable consumption can be achieved by organizing information events, distributing leaflets or by setting up an information center. A successful reduction of GHG emissions depends on local consumers assuming responsibility for their consumption patterns.

2.3.3.3 Areas of Supraregional Importance

As mentioned above, many aspects are beyond the control of a Swiss municipality. For instance, agriculture belongs to domains that are mainly regulated at the national level. Regardless, mitigating GHG from agricultural activities in Zernez would be difficult, given that Alpine livestock farming is the nearly only feasible and economically viable agricultural activity in Zernez. In addition, eliminating meat and milk production in Zernez in an effort to reduce local GHG emissions may simply lead to the relocation of said activities rather than to overall emission savings, as long as the demand for meat and dairy products do not change. However, introducing regulations to alter livestock production practice could be an option from a pure GHG perspective. Nguyen and colleagues (2010) [57] and Peters and colleagues (2010) [58] found that intensive meat production systems tend to perform more efficiently than extensive ones with regard to GHG. However, a deepened analysis would be required to assess whether these study results also apply to the prevailing circumstances in Zernez. Nguyen and colleagues (2010) [57] further indicate that targeted feeding strategies might minimize enteric CH_4 emissions and improved manure management could decrease N_2O emissions.

The municipality of Zernez and its inhabitants have only a limited scope of action available to reduce GHG in domains such as public services or health care. Local authorities account for only a minor contribution to GHG emissions in this area. Therefore, the largest potential is at cantonal and national levels, for example, through federal programs (Schweizerische Eidgenossenschaft 2013 [59]; RUMBA 2015 [60]).

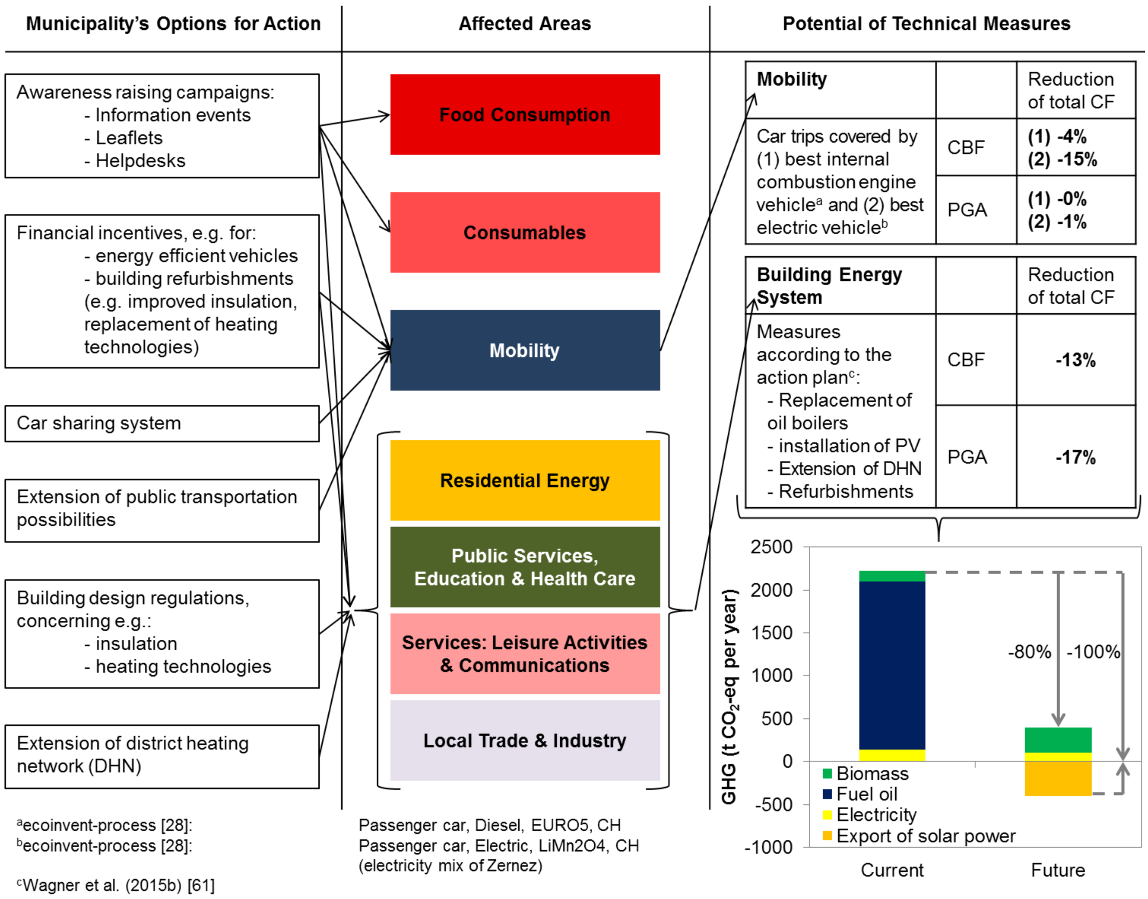


Figure 2.3: Overview of the municipality's areas of action, which are mentioned in the text. Rough estimates for two extreme mobility scenarios were conducted (replacing all car trips by the best internal combustion engine vehicle and the best electric vehicle, respectively). The reduction potential of technical measures in the building sector was extracted from Wagner and colleagues (2015a, 2015b) [30, 61]. Bottom right corner: Simplified illustration of the current and the predicted community-wide infrastructure footprint of the building stock's energy demand (see text for more information) and contributing energy carriers (Wagner et al. 2015b [61]). [CBF = consumption-based footprint; CF = carbon footprint; PGA = purely geographical accounting; PV = photovoltaics]

2.3.3.4 Overview of Areas of Action

Figure 2.3 gives an overview of all above-mentioned options for action. In addition, rough estimates of theoretical reduction potentials, which can be achieved by implementing technical measures, are also shown.

The discussion above reveals that it is generally difficult to adopt GHG reduction strategies at the municipal level given that many GHG areas are behavior driven or need to be tackled at a supraregional scale. The municipality of Zernez decided to focus on the CF of the building stock's energy demand as a first step. This is a reasonable starting point given that Swiss municipalities are in the position of setting financial incentives, providing helpdesks, and also have the ability to adopt new building design regulations. In addition, they often own energy utilities (e.g., electrical power supply or district heating networks).

2.3.4 Derivation of Targeted Measures

For the assessment of the CF and for the identification of GHG reduction potentials, the present study focused on the PGA and CBF because these two accounting methods allow for consistently assessing each category and complement each other in an easily understandable way (cf. Figure 2.2). However, for the derivation of measures targeted at the optimization of infrastructure, such as the building energy system, a "trans-boundary community-wide infrastructure" footprint (CIF) is more appropriate (Chavez and Ramaswami 2013 [19]; Lin et al. 2015 [18]). Appendix A provides information on the determination of the CIF for the building energy system in Zernez as well as a detailed flow scheme and a table showing the current energy flows related to the building stock.

Taking the CIF as a concrete planning basis to reduce GHG from the operation of the building stock, the municipality and its inhabitants entered an extensive stakeholder engagement process in the scope of the research project *Zernez Energia 2020* (Wagner et al. 2015a, 2015b [30, 61]). The project used an interdisciplinary approach to integrate building optimization options, renewable energy production, and urban planning. Additionally, the municipal authorities, the residents, and private companies were involved to include socioeconomic aspects of technical measures. Based on this transdisciplinary process, an action plan was developed for Zernez to mitigate GHG emissions caused by building energy demand (Wagner et al. 2015b [61]). Measures include replacement of oil boilers, extension of the district heating network, installation of PV cells, and targeted refurbishments of specific buildings and building components. The implementation of all of the proposed actions is forecasted to decrease the CIF of the building stock from the current value of approximately 2,225 to approximately 400 t CO_2-eq/yr for an 80% emission reduction (see bottom right corner in Figure 2.3). The project *Zernez Energia 2020* intends to compensate for these remaining emissions by exporting low carbon solar power to the national grid. By achieving a "CO_2-neutrality" for the building stock, total embodied emissions could decrease by 3%, total direct emissions from local demand by 76%, and exported emissions by 8% (percentages are related to Figure 2.2). This equals a reduction of the total footprint by 13% for the CBF and by 17% for the PGA, respectively (Figure 2.3).

2.3.5 Limitations of the Study

The presented approach to quantify the CF of Zernez involves many assumptions. Although the assessment of the GHG emissions was conducted with due care, the presented results contain uncertainties for example, attributed to simplifications, estimations, or inaccurate data. However, a lack of uncertainty information in the collected data hindered performing a detailed uncertainty analysis. Nevertheless, it has to be emphasized that the present study is based on a data set that is unique in its detail. This data set allowed for deducing results that provide a clear order of magnitude of the different categories and uncover how these different areas perform compared to one another.

Further, global warming is just one environmental concern of several. Besides different global-level indicators, local environmental impacts also need consideration given that they directly affect local living conditions and can also provide feedback to the functioning of the Earth system as a whole (Steffen et al. 2015 [62]; Rockström et al. 2009 [63]). GHG emissions illustrate the main focus of the project *Zernez Energia 2020* because global warming tops the political discussion regarding environmental impacts. However, recommendations for sustainable measures require a holistic perspective and should thus take different environmental indicators into account. For example, a large share of buildings in Zernez are heated by direct electricity or wood-based systems (wood logs or wood chips), both of which are low-carbon technologies. But the former wastes high-quality energy for low-quality purposes and thus prevents a more-efficient energy use, whereas the latter may cause significant emissions of particulate matter (Szidat et al. 2007 [64]), which may lead to serious health damage ranging from asthma to respiratory illness and lung cancer. Technical measures to reduce these emissions need to be implemented as well. Moreover, biogenic CO_2 emissions from wood-based heating systems were not considered in the presented CF, although they might have an impact on the global climate in the short term (Cherubini et al. 2011 [65]).

2.4 CONCLUSIONS AND OUTLOOK

The transformation of the building stock, especially of the energy supply system, was identified as a field of action for Swiss municipalities. The implementation of the action plan (Wagner et al. 2015b [61]) will lead to a significant reduction of GHG in Zernez and will increase the shares of renewable energy sources. Albeit these building-stock-oriented measures need to be followed by further action in other sectors, it is already a good step forward and might act as an example for other rural or Alpine municipalities. In the present case of Zernez, the current CF of the building stock is already below the national average, meaning the reduction potential may be higher in other Swiss municipalities.

However, the purpose of this chapter is not to compare Zernez with other municipalities. Detailed household- and building-level data are available for Zernez, but no other municipality possesses a comparable database. Whereas this unique data set enabled the exemplification of how a rural community can quantify and reduce its GHG emissions, we attempt to develop a detailed bottom-up consumption model based on nation-wide data in follow-up research. This model will build upon the ideas of Saner and colleagues (2013) [47] and on the detailed insights of the present study. The planned model will allow for computing the CBF on different aggregation levels and will thus enable comparisons between different municipalities and cities. Further, it shall support the development of targeted measures also in Swiss municipalities with less data available than Zernez.

2.5 REFERENCES

[1] UN-HABITAT, *Cities and Climate Change: Global Report on Human Settlements*, London, UK: Earthscan, 2011.

[2] C. A. Kennedy, N. Ibrahim, and D. Hoornweg, *Low-carbon infrastructure strategies for cities*, *Nature Climate Change*, vol. 4, no. 5, pp. 343–346, 2014.

[3] C. Kennedy, J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havranek, D. Pataki, A. Phdungsilp, A. Ramaswami, and G. V. Mendez, *Methodology for inventorying greenhouse gas emissions from global cities*, *Energy Policy*, vol. 38, no. 9, pp. 4828–4837, 2010.

[4] C. Kennedy, J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havránek, D. Pataki, A. Phdungsilp, A. Ramaswami, and G. Villalba Mendez, *Greenhouse Gas Emissions from Global Cities*, *Environ. Sci. Technol.*, vol. 43, no. 19, pp. 7297–7302, 2009.

[5] R. Frischknecht, C. Nathani, S. Büsser Knöpfel, R. Itten, F. Wyss, and P. Hellmüller, *Entwicklung der weltweiten Umweltauswirkungen der Schweiz. Umweltbelastung von Konsum und Produktion 1996 bis 2011*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2014.

[6] J. Minx, G. Baiocchi, T. Wiedmann, J. Barrett, F. Creutzig, K. Feng, M. Förster, P.-P. Pichler, H. Weisz, and K. Hubacek, *Carbon footprints of cities and other human settlements in the UK*, *Environmental Research Letters*, vol. 8, no. 3, 2013.

[7] B. Goldstein, M. Birkved, M. Quitzau, and M. Hauschild, *Quantification of urban metabolism through coupling with the life cycle assessment framework: concept development and case study*, *Environmental Research Letters*, vol. 8, no. 3, pp. 35024–35024, 2013.

[8] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[9] H. N. Larsen and E. G. Hertwich, *Identifying important characteristics of municipal carbon footprints*, *Ecological Economics*, vol. 70, no. 1, pp. 60–66, 2010.

[10] E. G. Hertwich and G. P. Peters, *Carbon Footprint of Nations: A Global, Trade-Linked Analysis*, *Environ. Sci. Technol.*, vol. 43, no. 16, pp. 6414–6420, 2009.

[11] T. Hillman and A. Ramaswami, *Greenhouse Gas Emission Footprints and Energy Use Benchmarks for Eight Cities*, *Environ. Sci. Technol.*, vol. 44, no. 6, pp. 1902–1910, 2010.

[12] A. Ramaswami, T. Hillman, B. Janson, M. Reiner, and G. Thomas, *A Demand-Centered, Hybrid Life-Cycle Methodology for City-Scale Greenhouse Gas Inventories*, *Environ. Sci. Technol.*, vol. 42, no. 17, pp. 6455–6461, 2008.

[13] United Nations, *World Urbanization Prospects: The 2011 Revision*, United Nations: New York, 2012.

[14] T. Baynes, M. Lenzen, J. K. Steinberger, and X. Bai, *Comparison of household consumption and regional production approaches to assess urban energy use and implications for policy*, *Energy Policy*, vol. 39, no. 11, pp. 7298–7309, 2011.

[15] Eurostat, *RAMON - Reference And Management of Nomenclatures*, 2016. .

[16] Bundesamt für Statistik (BFS), *STATPOP 2013 - Statistik der Bevölkerung und Haushalte*, Neuchâtel, Switzerland, 2014.

[17] Bundesamt für Statistik (BFS), *Die Raumgliederungen der Schweiz 2014*, Neuchâtel, Switzerland, 2014.

[18] J. Lin, Y. Hu, S. Cui, J. Kang, and A. Ramaswami, *Tracking urban carbon footprints from production and consumption perspectives*, *Environmental Research Letters*, vol. 10, no. 5, 2015.

[19] A. Chavez and A. Ramaswami, *Articulating a trans-boundary infrastructure supply chain greenhouse gas emission footprint for cities: Mathematical relationships and policy relevance*, *Energy Policy*, vol. 54, pp. 376–384, 2013.

[20] T. M. Baynes and T. Wiedmann, *General approaches for assessing urban environmental sustainability*, *Current Opinion in Environmental Sustainability*, vol. 4, no. 4, pp. 458–464, 2012.

[21] A. Ramaswami, A. Chavez, J. Ewing-Thiel, and K. E. Reeve, *Two Approaches to Greenhouse Gas Emissions Foot-Printing at the City Scale*, *Environ. Sci. Technol.*, vol. 45, no. 10, pp. 4205–4206, 2011.

[22] A. Ramaswami, A. Chavez, and M. Chertow, *Carbon Footprinting of Cities and Implications for Analysis of Urban Material and Energy Flows*, *Journal of Industrial Ecology*, vol. 16, no. 6, pp. 783–785, 2012.

[23] World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD), *The Greenhouse Gas Protocol - A Corporate Accounting and Reporting Standard*, World Resources Institute (WRI) / World Business Council for Sustainable Development (WBCSD): Geneva, Switzerland and Washington, DC, 2004.

[24] Bundesamt für Statistik (BFS), *STAT-TAB: Die interaktive Statistikdatenbank*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2014.

[25] Bundesamt für Landestopografie (swisstopo), *Website;*, 2014. [Online]. Available: http://www.swisstopo.admin.ch. [Accessed: 01-Apr-2014].

[26] N. Jungbluth, R. Itten, and M. Stucki, *Umweltbelastungen des privaten Konsums und Reduktionspotenziale*, Uster, Switzerland, 2012.

[27] Bundesamt für Statistik (BFS), *Physische Konten - Materialflüsse*, 2014. [Online]. Available: http://www.bfs.admin.ch/bfs/portal/de/index/themen/02/05/blank/data/03/01.html. [Accessed: 01-Apr-2014].

[28] ecoinvent Center, *Ecoinvent data, version 2.2*, http://www.ecoinvent.org/database/, 1998-2013, 2013.

[29] IPCC, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2007.

[30] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[31] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung 2010: Kommentierte Ergebnisse und*

Tabellen, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2012.

[32] D. Saner, C. Beretta, B. Jäggi, R. Juraske, F. Stoessel, and S. Hellweg, *FoodPrints of households*, *The International Journal of Life Cycle Assessment*, vol. 21, no. 5, 2016.

[33] GastroSuisse, *Branchenspiegel 2012*, GastroSuisse: Zurich, Switzerland, 2012.

[34] F. Grass, *Operator of the biogas plant in Zernez, Switzerland, Personal Communication, 2014*, 2014.

[35] Amt für Raumentwicklung, Amt für Energie und Verkehr, Amt für Natur und Umwelt, Amt für Wirtschaft und Tourismus, Tiefbauamt, and Verkehrsplanung Stadt Chur, *Mobilität in Graubünden - Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Chur, Switzerland, 2012.

[36] Amt für Wirtschaft und Tourismus (AWT), *Ständige Wohnbevölkerung nach Eckwerten, Gemeinden*, Chur, Switzerland, 2010.

[37] Tiefbauamt Graubünden, *Permanente Strassenverkehrszählung 2010*, Chur, Switzerland, 2011.

[38] Bundesamt für Strassen (ASTRA), *Verkehrsdaten*, 2014. .

[39] PostAuto, *Liniennetze*, 2014. [Online]. Available: www.postauto.ch/en/node/31788. [Accessed: 08-May-2014].

[40] SBB (Schweizerische Bundesbahnen), *SBB: Der grösste Bahnhof der Schweiz*, 2014. [Online]. Available: www.sbb.ch/home.html. [Accessed: 01-May-2014].

[41] I. A. Filli, *Operator of the waste water treatment plant in Zernez, Switzerland, Personal Communication, 2014*, 2014.

[42] A. König, N. Jungbluth, and G. Doublet, *Treibhausgasbilanz verschiedener Ferienszenarien*, Zurich, Switzerland, 2014.

[43] Bundesamt für Statistik (BFS), *Reisen der Schweizer Wohnbevölkerung 2010*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2011.

[44] D. Saner, *Politics and Social Responsibility, Post CH Ltd., Switzerland, Personal Communication, 2014*, 2014.

[45] P. H. Nielsen, A.-M. Nielsen, B. P. Weidema, R. Dalgaard, and Halberg N, *LCA food data base*, www.lcafood.dk, 2003.

[46] SOSA GERA SA, *SOSA GERA SA Zernez*, 2014. [Online]. Available: www.sosagera.ch/php/sosa-gera-kisswerk-zernez-schweiz.php. [Accessed: 02-Apr-2014].

[47] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[48] M. Balmer, K. W. Axhausen, and K. Nagel, *Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations*, *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1985, no. 1, pp. 125–134, 2006.

[49] K. Meister, M. Balmer, F. Ciari, A. Horni, M. Rieser, R. A. Waraich, and K. W. Axhausen, *Large-Scale Agent-Based Travel Demand Optimization Applied to Switzerland, Including Mode Choice*, in *12th World Conference on Transportation Research*, 2010.

[50] A. Froemelt and S. Hellweg, *Assessing Space Heating Demand on a Regional Level: Evaluation of a Bottom-Up Model in the Scope of a Case Study*, *Journal of Industrial Ecology*, vol. 21, no. 2, pp. 332–343, 2017.

[51] M. Thrane, *LCA of Danish fish products - New methods and insights*, *The International Journal of Life Cycle Assessment*, vol. 11, no. 1, pp. 66–74, 2006.

[52] S. Büsser and N. Jungbluth, *The role of flexible packaging in the life cycle of coffee and butter*, *The International Journal of Life Cycle Assessment*, vol. 14, no. 1, pp. 80–91, 2009.

[53] F. Stoessel, R. Juraske, S. Pfister, and S. Hellweg, *Life Cycle Inventory and Carbon and Water FoodPrint of Fruits and Vegetables: Application to a Swiss Retailer*, *Environ. Sci. Technol.*, vol. 46, no. 6, pp. 3253–3262, 2012.

[54] X. Zhang, L. Shen, and L. Zhang, *Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong*, *Renewable and Sustainable Energy Reviews*, vol. 17, pp. 160–169, 2013.

[55] N. Jungbluth, R. Steiner, and R. Frischknecht, *Graue Treibhausgas-Emissionen der Schweiz 1990-2004. Erweiterte und aktualisierte Bilanz*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2007.

[56] K. Orehounig, G. Mavromatidis, R. Evins, V. Dorer, and J. Carmeliet, *Towards an energy sustainable community: An energy system analysis for a village in Switzerland*, *Energy and Buildings*, vol. 84, pp. 277–286, 2014.

[57] T. L. T. Nguyen, J. E. Hermansen, and L. Mogensen, *Environmental consequences of different beef production systems in the EU*, *Journal of Cleaner Production*, vol. 18, no. 8, pp. 756–766, 2010.

[58] G. M. Peters, H. V Rowley, S. Wiedemann, R. Tucker, M. D. Short, and M. Schulz, *Red Meat Production in Australia: Life Cycle Assessment and Comparison with Overseas Studies*, *Environ. Sci. Technol.*, vol. 44, no. 4, pp. 1327–1332, 2010.

[59] Schweizerische Eidgenossenschaft, *Leitlinien und Ziele für das Ressourcen- und Umweltmanagement der Bundesverwaltung (RUMBA)*, Schweizerische Eidgenossenschaft: Bern, Switzerland, 2013.

[60] Koordinationsgruppe RUMBA and Fachgruppe RUMBA, *Umweltbericht 2015 der Bundesverwaltung. Berichtsperiode 2013 - 2014*, Schweizerische Eidgenossenschaft: Bern, Switzerland, 2015.

[61] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Aktionsplan*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010525998, 2015.

[62] W. Steffen, K. Richardson, J. Rockström, S. Cornell, I. Fetzer, E. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M. Persson, R. Veerabhadran, B. Reyers, and S. Sörlin, *Planetary boundaries: Guiding human development on a changing planet*, *Science*, vol. 347, no. 6223, 2015.

[63] J. Rockström, W. Steffen, K. Noone, A. Persson, F. Stuart Chapin III, E. F. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, B. Nykvist, C. A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. A. Foley, *A safe operating space for humanity*, *Nature*, vol. 461, no. 7263, pp. 472–475, 2009.

[64] S. Szidat, A. S. H. Prévôt, J. Sandradewi, M. R. Alfarra, H.-A. Synal, L. Wacker, and U. Baltensperger, *Dominant impact of residential wood burning on particulate matter in Alpine valleys during winter*, *Geophysical Research Letters*, vol. 34, no. 5, 2007.

[65] F. Cherubini, G. P. Peters, T. Berntsen, A. H. Strømman, and E. Hertwich, *CO2 emissions from biomass combustion for bioenergy: atmospheric decay and contribution to global warming*, *GCB Bioenergy*, vol. 3, no. 5, pp. 413–426, 2011.

CHAPTER 3

ASSESSING SPACE HEATING DEMAND ON A REGIONAL LEVEL: EVALUATION OF A BOTTOM-UP MODEL IN THE SCOPE OF A CASE STUDY

Andreas Froemelt* *and Stefanie Hellweg*

Journal of Industrial Ecology (2017) 21 (2) 332 – 343

* The individual contribution of Andreas Frömelt consisted of collecting and preparing the data, conducting the analyses and preparing the manuscript for publication.

SUMMARY

The residential sector constitutes a major energy consumer, particularly on account of its needs for space heating. Offering a high leverage potential, this sector is a suitable starting point for greenhouse gas mitigation policies. By providing predictions of the energy demand of building stocks, bottom-up building energy models represent a first step toward deriving strategies for abatement of detrimental effects related to housing energy use.

This chapter aims at evaluating the performance of a simplified bottom-up housing energy model. A global sensitivity analysis was performed to study the model's structure and the impact of individual model parameters. Moreover, an extensive final energy consumption data set allowed for an in-depth comparison of this model with primary data in the scope of a case study in a Swiss municipality.

On an individual building scale, the model fails to accurately simulate the energy demand. Deviations can be attributed to a range of factors, such as variability in occupants' behavior and problems of representativeness in the underlying statistical database. Nevertheless, such under- or overestimations level off on an aggregated scale. In particular, the model reproduces the overall characteristics of the residential building stock's heating demand well. It is therefore well suited as a building stock model and provides a promising basis for an extended assessment of housing energy demands. In future research work, we will apply this model to a larger region in order to study various types of settlements from a life cycle perspective and to derive targeted measures aimed at reducing environmental impacts.

3.1 INTRODUCTION

Energy is an integral component of economic and societal development. However, the dependence of today's energy systems on fossil or other nonrenewable energy sources not only engenders economic and societal consequences, but also causes a range of adverse environmental impacts, among which is climate change. Approximately 65% of the global anthropogenic greenhouse gas (GHG) emissions are estimated to be energy related (Herzog 2009 [1]).

Although the figures vary from country to country, the residential sector, with its associated heating and cooling loads as well as its electricity demand, is responsible for an estimated 24% (Lucon et al. 2014 [2]) to 30% (Saidur et al. 2007 [3]) of world-wide final energy consumption. Therefore, the residential building stock presents a high potential leverage for GHG mitigation initiatives. However, this sector is a complex system with several characteristics that influence energy consumption, such as building geometries, component materials, heating systems, and human behavior. Therefore, a holistic, life cycle-based approach is needed in order to identify and evaluate strategies for a sustainable development of urban settlements and for abatement of negative energy-related effects. Many important decisions are taken on a household or building level, which calls for a bottom-up approach. In addition to this, from an urban planning and policymaking point of view, it is important to obtain energy predictions for the whole building stock of a region and not only for individual buildings. It is at these larger scales that regulations for building design or the construction of district heating networks are planned. Consequently, there is a need for regionalized bottom-up models to support effective political decision making.

Estimates of the housing energy demand are prerequisites for a life cycle assessment (LCA) of the residential sector as well as for subsequent planning and exploring of different policy scenarios. Currently, a wide variety of different building energy models exist. For example, Swan and Ugursal (2009) [4] as well as Kavgic and colleagues (2010) [5] give a comprehensive overview of current residential building stock models. In both studies, these models are categorized into top-down and bottom-up approaches.

Top-down models regard the housing sector as an energy sink and relate the total energy demand of the building stock to certain – usually socio-economic – input variables based on historical data. Although such models are attractive because of their simplicity and are well suited for regional supply analyses, they only provide cumulative estimates and are generally incapable of assessing improvement options, such as the introduction of new technologies (Swan and Ugursal 2009 [4]; Kavgic et al. 2010 [5]).

In contrast, bottom-up models examine smaller sections of the residential sector, such as single buildings, groups of dwellings, or different energy end users, which are then aggregated to project the energy demand of the whole building stock. Bottom-up approaches cover a broad spectrum of different degrees of detail and hierarchical levels. Swan and Ugursal (2009) [4] distinguish two bottom-up mainstreams: statistical and engineering methods. Similar to top-down models, statistical bottom-up models often build upon historical data and regression analysis to relate energy needs to econometric indicators. Other techniques include conditional demand analysis and neural network methods. In contrast to top-down techniques, statistical bottom-up models allocate the energy consumption to particular end uses in order to attain a disaggregated analysis. Statements about the total energy demand of the residential stock are only possible by weighted combination and extrapolation of the model's forecasts. A further advantage of these kinds of bottom-up models is the implicit consideration of the large variation in occupants' behavior by utilizing statistical data. These methods are therefore suitable for the determination of typical energy demand patterns and for analysis of the contribution of different end uses to overall energy consumption (Swan and Ugursal 2009 [4]). However, statistical models are always bound to the data source they were derived from. Unreported energy uses or effects going beyond the observed historical data remain uncaptured.

Engineering models are based on building physics and explicitly compute energy demands of different end-use categories. The engineering approach then makes it possible to evaluate the effects of physical measures, such as component refurbishments or replacement of heating systems. These models feature the highest degree of flexibility and capability with regard to new technologies (Swan and Ugursal 2009 [4]; Kavgic et al. 2010 [5]). Yet, the determination of energy demands by solving complex physical equations requires a large amount of input data and is often accompanied by a heavy computational burden. Further, these models are unable to relate energy consumption to economic factors and show difficulties in reproducing "unreasonable" energy use originating from behavioral aspects (Swan and Ugursal 2009 [4]; Kavgic et al. 2010 [5]).

This chapter regards building stock models as tools to support the identification and exploration of environmental impact reduction strategies from a life cycle point of view. Mitigation measures are very likely to involve the deployment of new technologies and materials. Therefore, this study focuses on physical bottom-up models. A range of such bottom-up building energy models exist. Kavgic and colleagues (2010) [5] compared nine bottom-up building stock models, five of which focus on the building stock in the UK. All of these UK models, including the Community Domestic Energy Model *CDEM* developed by Firth and colleagues (2010) [6], use *BREDEM* (The Building Research Establishment's Domestic Energy Model) or modified versions of it (e.g., Dickson et al. (1996) [7]) as the core calculation engine. For annual or monthly predictions of an individual dwelling's energy demand, this engine relies on heat balance equations and empirical relationships. Other thermal energy calculation engines include *HOT2000* (CanmetENERGY 2011 [8]), which is used by *CREEM* (the Canadian Residential Energy End-use Model) (Farahbakhsh et al. 1998 [9]), or the software *TRNSYS* (The University of Wisconsin 2007 [10]). The latter is applied in the Belgrade Domestic Energy Model (*BEDEM*), which was developed and evaluated by Kavgic and colleagues (2013) [11]. A large-scale application of an engineering modeling approach is pursued by the *TABULA* project (Typology Approach for Building Stock Energy Assessment) and its successor: the *EPISCOPE*-project (Energy Performance Indicator Tracking Schemes for the Continuous Optimisation of Refurbishment Processes in European Housing Stocks). Both projects investigate the residential building stock of 16 European countries based on building typologies (TABULA Project Team 2012 [12]; IWU 2015 [13]). In Switzerland, an important dynamic city simulation tool is *CitySim* (Robinson et al. 2009, 2011 [14, 15]; Kämpf 2009 [16]; Kämpf and Robinson 2007 [17]). This complex modeling tool is based on dynamic heat transfer calculations. However, just as with most of the aforementioned bottom-up engineering models, *CitySim* requires an extensive amount of input information and the effects of occupants' behavior have to be assumed.

The simplified housing energy demand model of Saner and colleagues (2013) [18] is an interesting compromise between computational effort and data requirements and hence a promising approach for a rough assessment of urban energy flows in general. Based on the Swiss Standard SIA 380/1 (SIA 2009 [19]), this model provides estimates of space heating, domestic hot water, and electricity use for each residential building in Switzerland without antecedent excessive data acquisition requirements. For example, heating demand is quantified using heat balances, accounting for heat losses (ventilation, thermal transmission) and heat gains (solar heat gains and internal gains) in a simplified manner. The model by Saner and colleagues (2013) [18] can be used to perform LCAs of individual households' housing demand. However, this model has not yet been properly validated, so an in-depth evaluation is needed before it is applied to a larger region, for example, to a Swiss canton or even to the whole of Switzerland.

The goal of this chapter is therefore to evaluate the model performance of this simplified bottom-up housing energy demand model. For this purpose, the model is applied to a case study. A global sensitivity analysis is performed in a first step in order to understand the interactions between model parameters and how these parameters affect model results. In a second step, the model results are compared with an extensive data set of measured heating loads. This comparison is conducted on two different scales: A detailed analysis on an individual building scale shall reveal the strengths and flaws of the model, whereas an aggregated perspective shall investigate the model's suitability as a building stock model.

Housing energy demand consists in principle of three major end-use groups (space heating/cooling, domestic hot water, and electricity demand for appliances and lighting). However, this chapter only focuses on space heating because it is the most important building-specific energy consumption component. Hot water use and electricity demand for appliances depend on the users, rather than on the buildings, and are considered by means of standard values in the simplified model.

3.2 Data and Methodology

3.2.1 Case Study and Energy Demand Data Collection

Zernez is a small municipality located in the Swiss Alps at an altitude of approximately 1,471 meters above sea level (swisstopo 2014 [20]). Annual global horizontal solar irradiance amounts to 1,170 kilowatt-hours (kWh) per square meter, and the average annual ambient temperature was 4.8°C for the time period from October 2010 to September 2011 (Orehounig et al. 2014 [21]).

In 2010, 1,140 inhabitants lived in Zernez (BFS 2014 [22]) in 279 residential buildings (BFS 2013 [23]) erected between 1250 and 2010 (BFS 2013 [23]; ETHZ and Zernez 2015 [24]). In 1872, a major fire devastated 117 of the 154 buildings existing at that time. Most buildings were reconstructed in a short time. In order to protect the village from future fires, the buildings were re-erected in an architectural style that is atypical for the region. However, just as the traditional buildings, these edifices feature thick stone walls that do not correspond to an average building of this time in Switzerland.

Presently, 39% of all residential buildings in Zernez are heated by burning fuel oil, 24% possess direct electrical heating systems, and 21% use log wood (BFS 2013 [23]). A small district heating network, which is fed by a wood chip furnace, supplies heat to another 5% of the residential buildings (BFS 2013 [23]). The remaining 11% use air-source heat pumps, ground-source heat pumps, or energy carriers that are not further specified (e.g., solar thermal collectors).

One of the goals of the municipality of Zernez is to replace fossil energy carriers for space heating and hot water production with renewable energy sources until 2020. For this purpose, the research project *Zernez Energia 2020* was initiated. In the context of this project, a survey was conducted within the municipality in order to gather data on fuel oil consumption, on the amount of wood chips and logs used, as well as on district heating and electricity demand. The main data sources were energy bills and interviews of inhabitants. Further, the municipality supported the establishment of a database containing detailed information on installed heating systems and on different building characteristics, such as age, type, construction method, and insulation quality. However, the survey was limited to the core part of the village and excluded hamlets and remote single houses located within the municipality's borders. The final database comprises information on all 309 buildings of the village's core (ETHZ and Zernez 2015 [24]). One hundred ninety-four of these edifices are residential buildings (ETHZ and Zernez 2015 [24]).

For the evaluation of the model results with primary data, the amounts of energy carriers collected in the survey were converted to net energy demand required for space heating. Thereby, a problem was encountered given that electricity bills did not distinguish between electricity used for appliances and for resistance heaters or heat pumps, respectively. In order to split the total electricity consumption into the various uses, typical electricity demands for appliances and lighting were derived for some buildings in the scope of the project (ETHZ and Zernez 2015 [24]). These typical electricity demands were then extrapolated to buildings with electricity-driven heating technologies and deducted from the total electricity consumption to derive the electricity demand for heating (ETHZ and Zernez 2015 [24]).

In a next step, the amounts of energy carriers were translated to final energy demand for heating. The electricity bills and the operator of the district heating network provide data in kWh or megajoules (MJ). For buildings, relying on fuel oil or firewood as an energy source, the energy contents were assumed to be 36.7 MJ per liter and 10,080 MJ per cubic meter, respectively, in order to compute the net calorific value (Orehounig et al. 2014 [21]; ETHZ and Zernez 2015 [24]). The resulting amount of energy corresponds to the delivered energy entering a building (final energy). In order to derive the actual heating demand (useful energy), the efficiency of the installed heating system was considered in a third step. The great variety of different heating systems found in Zernez rendered it almost impossible to determine the exact efficiency value for each system. Therefore, standard conversion efficiencies related to the net calorific value were assigned to the different energy carriers (electricity: 95%; fuel oil: 75% to 104%; district heating: 95%; wood: 75% to 93%) (Orehounig et al. 2014 [21]; ETHZ and Zernez 2015 [24]; Burri and Knecht 2014 [25]). The resulting amounts of used energy represent the sum of the heat demanded for hot water production and for space heating. In order to subtract the energy needs for domestic hot water (DHW), typical energy uses for hot water were assumed (Nipkow et al. 2007 [26]). Given the large share of holiday flats and secondary homes in Zernez and given the fact that DHW was already separated for some buildings, the subtraction of these typical DHW energy requirements results in an underestimation of the effective space heating demand.

In order to cover the uncertainty in the heating demands derived from measurements, two values of space heating demand were calculated to form a range. For the minimum, the heating demand with subtracted DHW needs and with low conversion efficiencies was computed. For the maximum, the heating demand with high conversion efficiencies was calculated. This range is referred to henceforth as "empirical database range." A schematic description of the applied procedure to derive this empirical database range can be found in Appendix B.

3.2.2 Model Description

The underlying idea of Saner and colleagues' (2013) [18] model is the establishment of simplified energy balances for each building according to the Swiss Standard SIA 380/1 (SIA 2009 [19]), published by the Swiss Society of Engineers and Architects (SIA). The core of the hourly energy balance to estimate the heating energy demand (Q_h) for a specific building in the time period from t_{begin} to t_{end} is formed by equation (3.1).

$$Q_h = \sum_{t=t_{begin}}^{t_{end}} (Q_{T,t} + Q_{V,t}) - \eta_g (Q_{s,t} + Q_{iP,t} + Q_{iEl,t}) \tag{3.1}$$

In this formula, the heating energy demand is computed as the difference between thermal losses and thermal gains. The latter includes solar gains (Q_s) through windows, internal gains from the presence of people (Q_{iP}), and internal gains attributed to the use of electricity (Q_{iEl}). The sum of all thermal gains is multiplied by η_g, which is the degree of utilization for heat gains and depends on the thermal storage capacity of the building mass (SIA 2009 [19]). Thermal losses comprise transmission losses through the building's envelope (Q_T) as well as ventilation losses (Q_V).

The input data for this energy balance are retrieved from three main sources: climatic data produced by the software *Meteonorm* (METEOTEST 2012 [27]), the Swiss Federal Register of Buildings and Dwellings (FRBD) (BFS 2013 [23]), which contains up-to-date building-specific data about each residential building in Switzerland, and different building-specific statistics (e.g., from Wallbaum et al. [2010] [28]). Although the data sources for this simplified model are very comprehensive, some important information could only be derived by means of assumptions. These include, among others, the window area share, the renovation year of a specific building component, the roof type and inclination, as well as the building orientation. The distributions of each one of these parameters was simulated by Latin Hypercube sampling, so as to gain a better grasp on the variability of these parameters (a full list of all stochastically modeled parameters can be found in Appendix B). In the scope of the case study in Zernez, 1,000 simulation runs were performed. A discussion of the model results for individual buildings can be found in Appendix B. For the comparison with the municipal data set, the model results were averaged for each building (see Appendix B).

A more detailed description of this simplified building energy demand model can be found in Saner and colleagues (2013) [18]. A simplified flow chart in Appendix B illustrates the interactions of the different model components, and how the input data and the stochastically modeled parameters enter the computations to finally derive the heating demand according to equation (3.1).

3.2.3 Global Sensitivity Analysis

Before comparison of the model predictions with primary data, a global sensitivity analysis was performed in order to learn more about model inherent effects and investigate the impact of individual model parameters on the model results. Only model parameters subjected to Latin Hypercube sampling and therefore associated with a probability distribution were part of this global sensitivity analysis, whereas the variation of parameters represented by default values (e.g., internal gains from the presence of people [Q_{iP}]) and the input data (e.g., measured outdoor temperature) were not taken into account in the quantitative analysis and are only discussed qualitatively. The global sensitivity analysis followed the density-based approach introduced by Plischke and colleagues (2013) [29] using Borgonovo's moment-independent δ as a measure for the global sensitivity of the model results. This approach derives global sensitivity indices from given data at the minimum computational cost and is thus predestinated for investigating models with underlying Monte Carlo or Latin Hypercube sampling.

The density-based moment-independent measure (δ) used to express the global sensitivity is normalized between 0 and 1. If the model output is independent of parameter i, a bias reduction filter based on a Kolmogorov–Smirnov test sets δ_i to zero (Plischke et al. 2013 [29]). In contrast, the larger δ_i, the higher is the impact of parameter i on the model result.

3.2.4 Model Evaluation with Empirical Data

The geographical boundaries of this study are the borders of the case study municipality of Zernez. The simplified housing energy demand model was applied to the whole building stock resulting in heating energy demand predictions for each of the 279 residential buildings within the municipal borders. The database, which was established in collaboration with the inhabitants and the municipality, contains information on 194 residential buildings. However, the information in the database is not equally reliable for all buildings. For instance, the consumption of wood logs was not determined by measured quantities, but estimated by the owners of the buildings using firewood for space heating. Another example: No data were collected on energy savings from solar thermal systems.

Further difficulties arose when results of the model had to be compared to buildings in the database. The model produced heating demand estimates for entries in the FRBD, which did not always correspond to database buildings. For instance, two houses attached to each other featured sometimes only one entry in the FRBD, but two entries in the municipal database and vice versa. Whereas such problems could be tackled in the majority of cases, there were also cases where no clear juxtaposition was possible. Facing these problems, it was thus decided to consider only those residential buildings with reliable and unambiguous data entries for the evaluation of the model. In the end, 133 residential buildings were used for a building-wise evaluation of the model by the measured heating loads of the municipal database.

The temporal resolution of the model is on an hourly basis. However, data collected for the energy consumption of the municipality of Zernez contains annual values only. Therefore, the comparisons were conducted on a yearly aggregated basis. Because the empirical dataset of the case study was established from October 2010 to September 2011, the inputs for the model were referred to this time period wherever possible.

Classical descriptive statistical indicators were calculated to compare the sum of all model estimates and the measured total annual heating demand of the building stock considered, namely, the average, standard deviation, correlation coefficient, minimum, maximum, as well as absolute and relative difference. The detailed building-wise comparison focused primarily on absolute and relative differences.

3.3 RESULTS AND DISCUSSION

3.3.1 Internal Model Evaluation: Global Sensitivity Analysis

The global sensitivity analysis evaluates the model from an internal perspective and reveals the importance of individual model parameters within the model structure. Given that the parameter sampling for a certain building depends on the specific situation and on specific building characteristics, the same parameter took different δs for different buildings, pointing out that a certain parameter is not of the same importance for all buildings. The application of Plischke and colleagues' approach [29] to the 133 case-study buildings therefore resulted in a distribution of δs, which is shown in Figure 3.1 as box plots. It is important to note that this figure presents the results for only some of the parameters to facilitate the interpretation of the sensitivity analysis. For instance, results for the time of wall refurbishment are not presented, but results for the choice of U-values – which is directly dependent on time of wall refurbishment – are indeed presented and have an impact on model results, which is easier to understand. Details on the global sensitivity analysis results can be found in Appendix B.

Figure 3.1 clearly shows that effective room temperature and U-values for walls are the two most influential parameters at the building stock level. This outcome, especially the importance of indoor temperature, is also in line with the findings of other sensitivity studies (Firth et al. 2010 [6]; Kavgic et al. 2013 [11]). Old buildings are dominated by U-values given that their long lifetime enables the Latin Hypercube sampling to choose many different years of refurbishment, resulting in a large spectrum of different U-values. Contrary to that, U-values for newer buildings are restricted to a narrow distribution. The most extreme case in this regard would be a building erected in the year 2010, for which the Latin Hypercube sampling is only allowed to choose one U-value. In such cases, the indoor temperature becomes the predominant factor influencing the model results.

Evidently, parameters such as deviation from the South, thermal storage capacity, floor U-values, roof type, and roof inclination do not have a big influence on overall heating demand. The latter two are used for the determination of roof U-values and for the effect of solar thermal collectors. However, in the present case study, no solar collector systems are considered and U-values for slanted and flat roofs are very similar.

Compared to the other components, floor U-values seem to be much less important. According to SIA 380/1, various reduction factors are applied to the chosen U-value on account of the floor's direct connection to the soil. By rendering similar U-values for all floors, these factors make up for any impact of different floor ages.

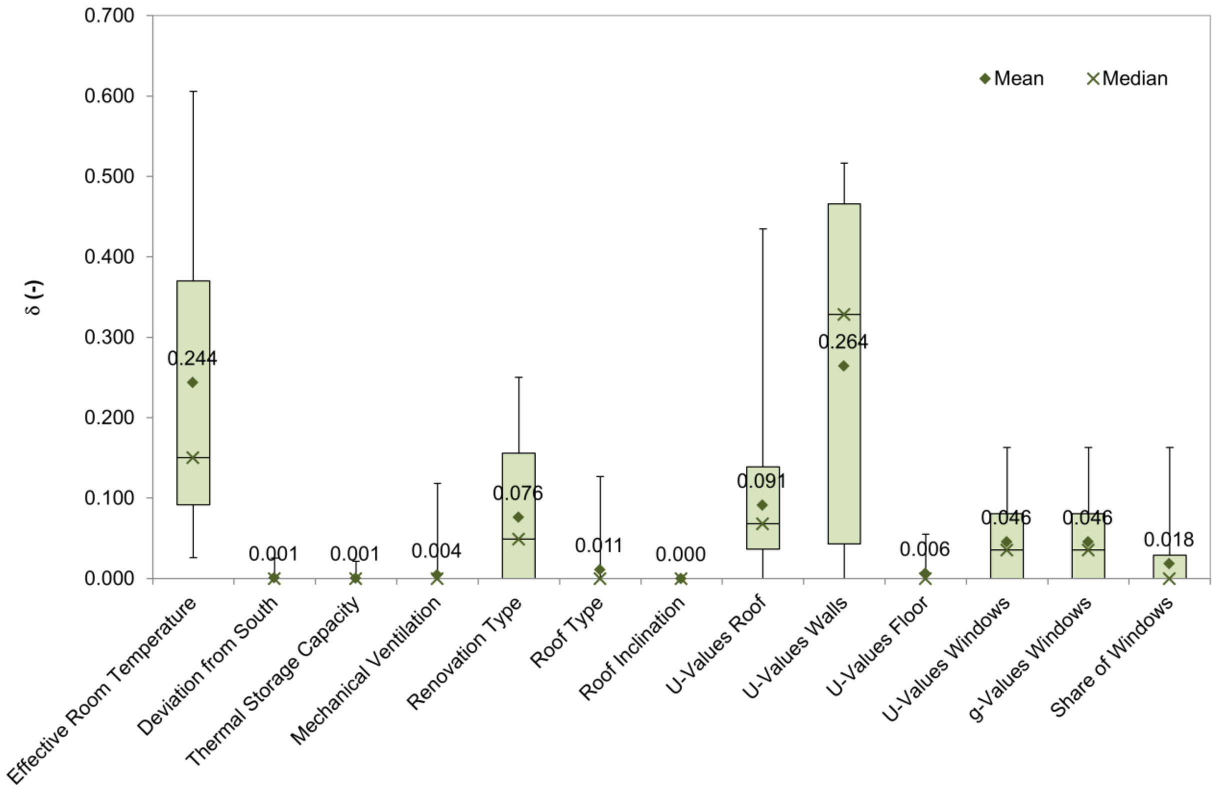


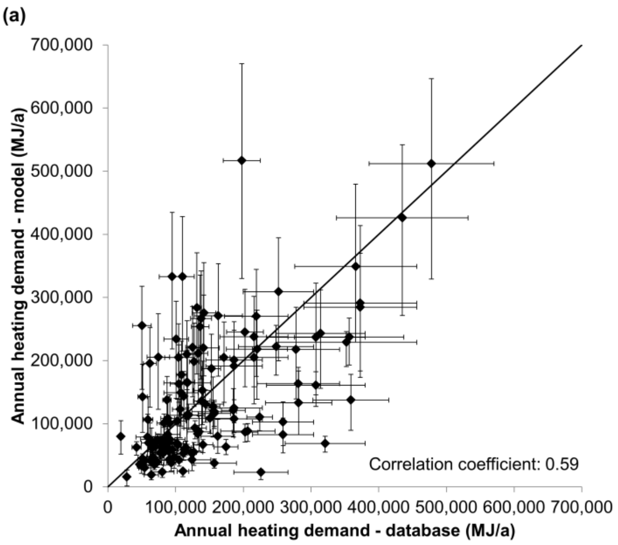
Figure 3.1: **Box plots of the density-based sensitivity measure δ for different model parameters after applying Plischke and colleagues' approach (2013)** [29] **to the 133 case-study buildings. More information about the applied stochastic modeling and presented results is given in Appendix B. [U-values = heat transfer coefficients; g-values = solar energy transmittance]**

The global sensitivity analysis permitted to take into consideration only those parameters with a distribution simulated by Latin Hypercube sampling. This should not hide the fact that also other factors can exhibit a crucial impact on the model predictions. For example, Firth and colleagues (2010) [6] and Kavgic and colleagues (2013) [11] found their models to be very sensitive to outdoor temperature, which was here considered as a deterministic parameter from temperature records. In addition to climatic input data, Heeren and colleagues (2015) [30] identified the ventilation rate to be essential for their model, which is another parameter not examined in our global sensitivity analysis (in fact, ventilation rate was modeled individually per building, but it was not varied in the sensitivity analysis). However, comparisons with other sensitivity studies are only possible to a limited degree. None of these other analyses performed a density-based global sensitivity analysis, and the building characteristics as well as the building models varied between studies. For example, Heeren and colleagues (2015) [30] applied their detailed model to a hypothetical new building and not to a building stock.

3.3.2 External Model Evaluation: Comparison of the Model with Primary Data

3.3.2.1 Individual Building Scale

Figure 3.2 presents the building-wise comparison of model results and reported heating loads. Figure 3.2a relates the average model results to the mean of the empirical database range, whereas Figure 3.2b compares the 95% confidence intervals of the model results with the empirical database range. For most of the subsequent analyses, only the average model results are considered.



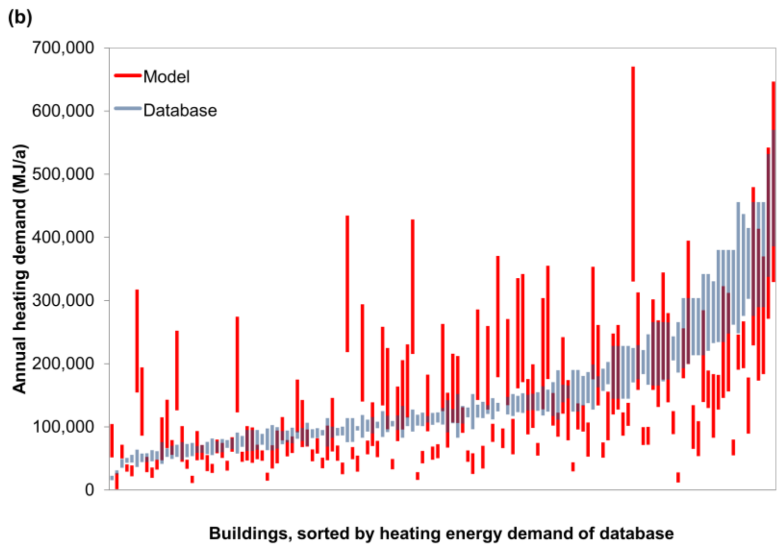


Figure 3.2: (a) Building-wise comparison of annual heating demand estimated by the model and reported heating loads according to the empirical database. Error bars in the y-direction represent the 95% confidence intervals of the model results, whereas error bars in the x-direction show the empirical database range. Straight line corresponds to a 1:1-relationship (perfect match). (b) Building-wise comparison of the model's 95% confidence intervals with the empirical database range.

As can be seen in Figure 3.2a, the data points scatter around the 1:1 line and the correlation coefficient is equal to 0.59. The model's 95% confidence intervals overlap with the empirical database range for 65 of 133 buildings (Figure 3.2b). Yet, this figure shows clearly that simulated heating demands largely deviate from database entries for some individual buildings. In order to understand the origin of such deviations, the building characteristics were examined in the cases where the model performed poorly. Two building properties turned out to be most interesting in this regard: the year of construction and the amount of space heating demand according to the empirical database. Figure 3.3 displays, for all 133 case-study buildings, absolute and relative deviations between average model results and database entries sorted by recorded heating energy demands and construction year, respectively. Larger absolute differences occur for larger measured space heating demands, whereas relative deviations are larger for smaller database entries (Figure 3.3a and 3.3b). It can be deduced from Figure 3.3a that the model tends to underestimate larger effective energy demands. This might arise from not adequately reproducing the occupants' behavior or from an intrinsic model structure problem that prevents the model from predicting extreme values. For instance, if the U-value of a building component is above the maximum U-value of the model's range and causes a large heating demand, the model will not be capable of reproducing this building's large heating load. Another example concerns the model simplifications regarding the building's shape (see Appendix B): Here, we assumed buildings to be cubes, which is an optimistic assumption, because it results in a minimum envelope to volume ratio. This leads to an underestimation of thermal transmission losses for elongate structures.

Figure 3.3c and 3.3d shows that heat demands for older buildings are often overestimated, whereas the model tends to underestimate newer buildings. Further, in Figure 3.3d, it can be seen that relative deviations are higher for older buildings and lower for newer buildings. The largest absolute as well as the largest relative deviations are observed for the period of reconstruction after the fire in 1872. This is also supported by Figure 3.4, in which the relative deviations are separately depicted for the reconstructed and all other buildings. The box plots illustrate that the width of relative deviations is larger for the re-erected buildings compared to that of other buildings. As mentioned earlier, the dwellings built after the fire show an atypical style for this time period. In addition to these buildings, the traditional buildings do not conform to the Swiss average of their time either. On account of harsh winter conditions, the traditional style of buildings in the Engadin valley dictates the need for thick stone walls as well as few and small windows. However, being bound to the statistical findings of Wallbaum and colleagues (2010) [28], the model can be expected to perform well only if the simulated buildings correspond to the average of buildings of a certain time period. Yet, in the case of traditional and postfire buildings, the statistical basis here predicts worse insulation than encountered in reality and thus overestimates thermal transmission losses (Figure 3.3c,d).

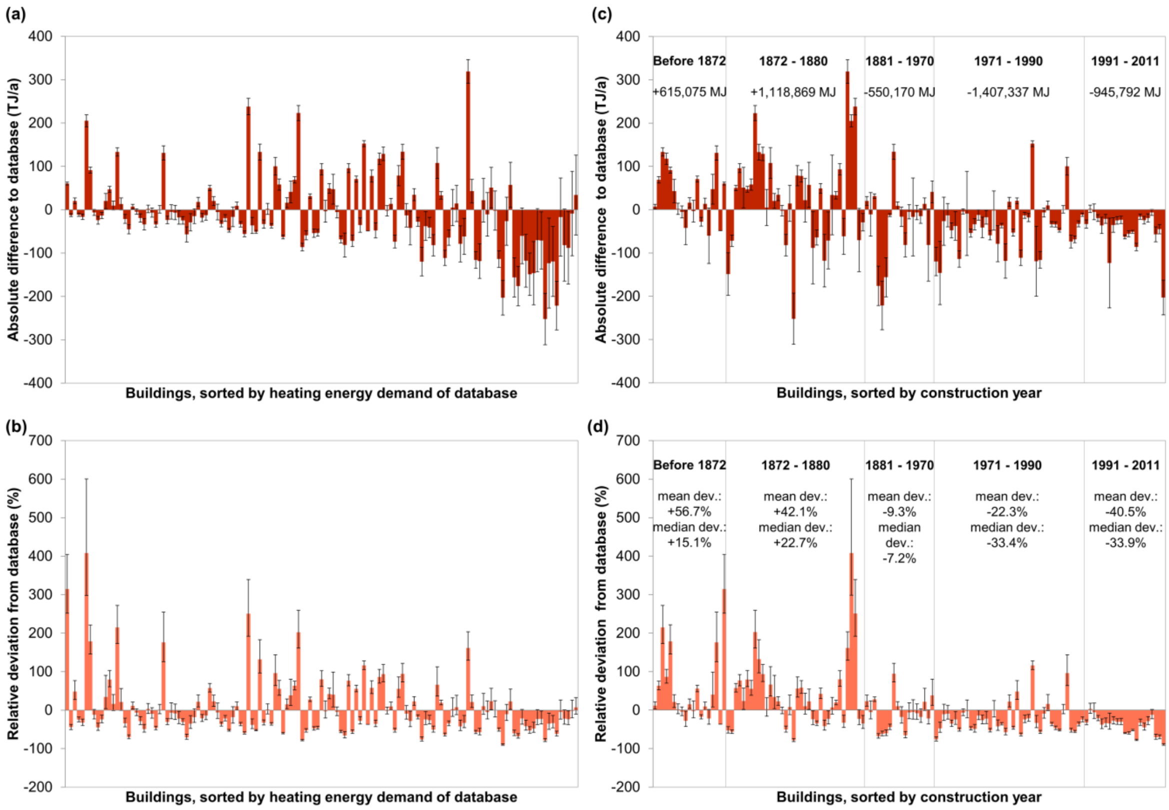


Figure 3.3: **Absolute [(a) and (c)] and relative [(b) and (d)] differences between model results and empirical database, sorted by measured space heating demand [(a) and (b)] and construction year [(c) and (d)]. Error bars indicate variation of the differences that arise because of referencing once to the maximum and once to the minimum database estimations.**

In addition to the problems of statistical representativeness of specific buildings, there are other reasons that explain the large deviations for some single buildings. Approximately 18% of the apartments in Zernez are holiday flats or secondary homes and are therefore only occasionally occupied. Long periods of absence during which the homes are not – or only partially – heated are not accounted for in this simplified model. In general, the occupants' behavior poses a major difficulty for bottom-up engineering models. Several articles emphasize that the large variation in behavior of inhabitants can impact the total energy demand of a certain home by 100% or more (Nipkow et al. 2007 [26]; Swan and Ugursal 2009 [4]; Kavgic et al. 2013 [11]; Robinson et al. 2007 [31]; Haldi and Robinson 2011 [32]). The influence of the dwellers' behavior on a building's energy demand emerges in the present case also from the results of the global sensitivity analysis. For instance, Figure 3.1 highlights the importance of the indoor temperature, which is chosen by the occupants.

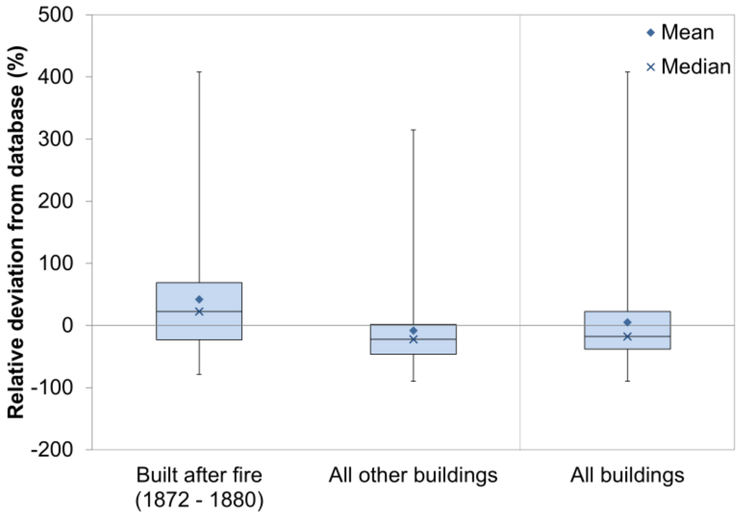


Figure 3.4: **Box plots illustrating relative deviations of model predictions from the empirical database. Relative deviations refer to the mean of the empirical database heating demand range.**

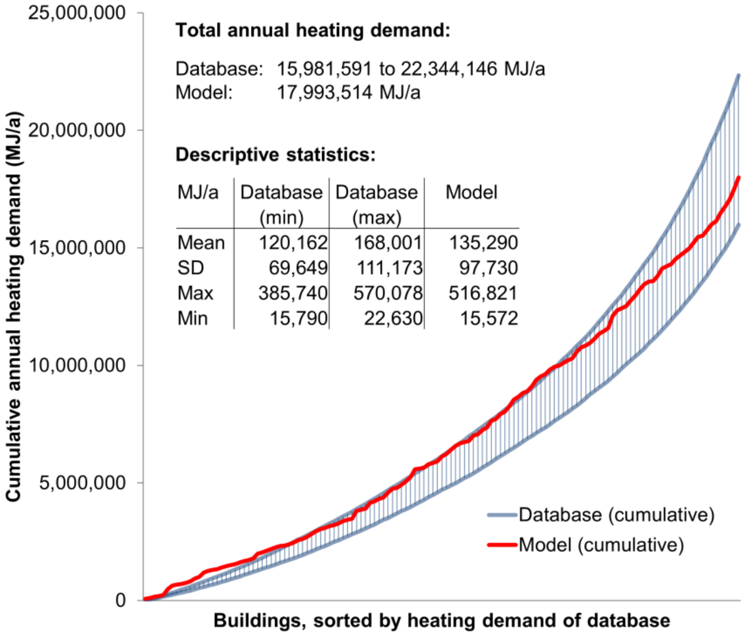
3.3.2.2 Building Stock Perspective

One of the main intentions of this chapter is to investigate the suitability of this simplified housing energy demand model as a building stock model. For this purpose, Figure 3.5 compares the model results with the empirical database range on an aggregated level.

The total annual heating demand of all 133 buildings is appropriately simulated by the applied model. The relative deviation of the cumulated model results from the building cluster's heating demand ranges from an underestimation of 20% to an overestimation of 13%. From the descriptive statistical indicators (see table in Figure 3.5) the model seems to adequately emulate the overall characteristics of the residential building stock's heating demand.

Figure 3.5 considers cumulative curves of the heating demands because this presentation gives a better overview of the stock's behavior than that on an individual basis (cf. Figure 3.2). Two important properties of the diagram in Figure 3.5 may help to evaluate the model from a building stock perspective: The proximity and the slope of the curves. First, if the model's cumulative curve follows the empirical database curve closely, this suggests that the total annual heating demand of the building stock is met on different levels of aggregation and not only for the whole building stock. Second, and more important, the model's performance can directly be judged by comparing the similarity in the slopes of the two curves. Discontinuities indicate a bad model performance for one or a group of buildings. But, if after the discontinuity, the curve regains a trajectory parallel to the empirical one, this means that the model reproduces the database entries well for the subsequent buildings.

Analyzing Figure 3.5 in detail, three areas can be distinguished: The cumulative model curve exceeds the maximum of the empirical database for low heating demands, follows then the edge of the maximum space heating demand derived from the empirical database, and settles down between the minimum and the maximum of the empirical database range for large heating demands (this behavior can be explained by means of Figure 3.2b and Figure 3.3a). Overall, both the proximity of the lines and the slopes point to a good model performance: The model's line is, in general, plotted closely, and more or less parallel, to the reported cumulative annual heating demand. In Figure 3.5, the buildings are ranked according to the heating demand of the empirical database because this facilitates the assessment of the slopes. Additional diagrams showing the buildings sorted in a different manner are presented in Appendix B.



MJ/a	Database (min)	Database (max)	Model
Mean	120,162	168,001	135,290
SD	69,649	111,173	97,730
Max	385,740	570,078	516,821
Min	15,790	22,630	15,572

Figure 3.5: **Cumulative annual heating demand of buildings sorted by ascending heating demand according to the mean of the empirical database range. The cumulative curve of the model (red) was built by accumulating the simulated heating demands corresponding to the sorted buildings. Analogously, the lower blue line shows the cumulated minimum of the empirical database range whereas the upper blue line denotes the cumulated maximum of the empirical database range. [SD = standard deviation; Max = maximum value; Min = minimum value; Database (min) = minimum of empirical database range; Database (max) = maximum of empirical database range]**

3.4 Conclusion and Outlook

The extensive data set of reported heating loads of this study presented an excellent opportunity to evaluate the performance of housing energy demand models. In our evaluation procedure, we aimed to quantitatively account for many sources of uncertainty introduced by both the empirical data and model assumptions, respectively. However, some potential sources of error were not part of the present analysis, such as inaccuracy of climatic input parameters, correctness of data from the FRBD (BFS 2013 [23]), and some default values (e.g., internal gains from the presence of people). These unconsidered uncertainties as well as those induced by the assumptions made to derive heating demands from the empirical database hinder exact statements about the model's accuracy, but it was possible to identify general trends and directions as well as general flaws and strengths of the model. The comparison of primary data and model results revealed that the model of Saner and colleagues (2013) [18] is not very accurate in predicting heating demand of single buildings. However, it was rather strong in simulating the overall characteristics of the residential building stock's heating demand.

The analysis of the traditional and the postfire buildings showed that the statistical building stock data of Wallbaum and colleagues (2010) [28] was not representative for the old buildings in the municipality investigated here and led to systematic model errors. The global sensitivity analysis supports this conclusion by revealing the important impact of U- and g-values on the model outputs. Whereas the model especially tends to deviate for old buildings, the heating demand of newer edifices is generally better reproduced. This suggests that the model may perform better for building stocks with newer buildings.

Model assumptions on occupants' behavior are another major source of uncertainty. In the case of Zernez, this is particularly important because of the high ratio of holiday flats and secondary homes, which are irregularly heated. An interesting analysis on the impacts of dwellers on energy demands by their interactions with windows and shading devices is given by Haldi and Robinson (2011) [32]. Given the high relevance of behavioral patterns, large differences between model results and the municipal database for some single buildings are no surprise, because variation in user behavior was only partially considered. The global sensitivity analysis confirmed the significance of behavioral aspects by highlighting the important impact of the chosen indoor temperature on model results. The comparison of the model results with primary data leads, however, to the conclusion that the standard and default values, which were applied to cover individual behavior, work well on a building stock level where under- and overestimations cancel each other out. Nevertheless, future research should focus on a better representation of the interaction of dwellers with buildings.

Further improvement of the model could be achieved by replacing the simplified assumptions regarding the building's shape by three-dimensional building geometries. Current work is attempting to derive such information from nation-wide laser-scanning data.

By providing estimates of the energy demand of an entire building stock, bottom-up building energy models, such as the one examined in this chapter, constitute powerful tools for urban planners and local authorities, who need a district or municipality perspective. Many important decisions are made at this level, related to regulations for building design, the construction of district heating networks, or to financial support for private refurbishment initiatives, to name a few. There are numerous fields of application of such models: They can support scenario planning and identify strategies for the abatement of environmental impacts related to housing energy use. For example, in the present case study, several recommendations were made for the municipality; among them a detailed list with indications on the buildings and the specific building components that should be given priority for refurbishment, as well as a concrete proposition to extend the district heating network (ETHZ and Zernez 2015 [24]).

Another possibility is to combine the housing stock model with optimization techniques, as demonstrated by Saner and colleagues (2014) [33], who applied mixed integer linear programming to the simplified energy demand model examined in this chapter. Further, the model may be integrated in household consumption studies (Saner et al. 2013 [18]). Regionalized LCAs of household consumption might support the search for holistic and sound environmental impact mitigation strategies.

We conclude that the simplified model of Saner and colleagues (2013) [18] performed well on an aggregated level. Therefore, this model provides a promising basis for the investigation of heating energy demand of the residential building stock, for instance, in the context of decision-supporting scenario analyses or regionalized LCA studies. The findings of this study, and the fact that this model features a fast performance and does not require antecedent excessive data acquisition, provide motivation to continue our work with this model. In a first step, we will further improve the model by integrating the aforementioned three-dimensional building geometries. Subsequently, we will apply the model to a larger region, such as a whole canton or the whole of Switzerland. This will allow for an in-depth analysis of housing patterns on a large scale and an assessment of different urban settlement typologies related to building energy demand.

In principle, the applicability of the model itself is not restricted to Switzerland, but it could also be used for other countries. Yet, the availability of adequate input data has to be scrutinized, especially with regard to building-specific information (here provided by the FRBD [BFS 2013 [23]]) and thermophysical parameters (here provided by Wallbaum et al. [2010] [28]).

3.5 REFERENCES

[1] T. Herzog, *World Greenhouse Gas Emissions in 2005*, World Resources Institute, Washington, 2009.

[2] O. Lucon, D. Ürge-Vorsatz, A. Zain Ahmed, H. Akbari, P. Bertoldi, L. F. Cabeza, N. Eyre, A. Gadgil, L. D. D. Harvey, Y. Jiang, E. Liphoto, S. Mirasgedis, S. Murakami, J. Parikh, C. Pyke, and M. V. Vilariño, *Buildings*, in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, O. Edenhofer et al., Ed. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014.

[3] R. Saidur, H. H. Masjuki, and M. Y. Jamaluddin, *An application of energy and exergy analysis in residential sector of Malaysia*, *Energy Policy*, vol. 35, no. 2, pp. 1050–1063, 2007.

[4] L. G. Swan and V. I. Ugursal, *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques*, *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.

[5] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, *A review of bottom-up building stock models for energy consumption in the residential sector*, *Building and Environment*, vol. 45, no. 7, pp. 1683–1697, 2010.

[6] S. K. Firth, K. J. Lomas, and A. J. Wright, *Targeting household energy-efficiency measures using sensitivity analysis*, *Building Research & Information*, vol. 38, no. 1, pp. 25–41, 2010.

[7] C. M. Dickson, J. E. Dunster, S. Z. Lafferty, and L. D. Shorrock, *BREDEM: Testing monthly and seasonal versions against measurements and against detailed simulation models*, *Building Services Engineering Research and Technology*, vol. 17, no. 3, pp. 135–140, 1996.

[8] CanmetENERGY, *HOT2000*, Natural Resources Canada, 2011.

[9] H. Farahbakhsh, V. I. Ugursal, and A. S. Fung, *A residential end-use energy consumption model for Canada*, *International Journal of Energy Research*, vol. 22, no. 13, pp. 1133–1143, 1998.

[10] The University of Wisconsin, *TRNSYS*, The University of Wisconsin, Madison, United States, Madison, United States, 2007.

[11] M. Kavgic, D. Mumovic, A. Summerfield, Z. Stevanovic, and O. Ecim-Djuric, *Uncertainty and modeling energy consumption: Sensitivity analysis for a city-scale domestic energy model*, *Energy and Buildings*, vol. 60, pp. 1–11, 2013.

[12] TABULA Project Team, *Typology Approach for Building Stock Energy Assessment. Main Results of the TABULA project*, Darmstadt, Germany, 2012.

[13] Institut Wohnen und Umwelt (IWU), *EPISCOPE*, 2015. [Online]. Available: episcope.eu. [Accessed: 01-Dec-2015].

[14] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, and U. Wilke, *CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning*, in *Proc. Building Simulation, Eleventh International IBPSA Conference*, 2009.

[15] D. Robinson, A. Rasheed, M. Bruse, F. Haldi, J. H. Kämpf, D. Perez, K. W. Axhausen, F. Flourentzou, and M. Batty, *COMPUTER MODELLING FOR SUSTAINABLE URBAN DESIGN - Physical Principles, Methods and Applications*, London, Washington DC: ed. D. Robinson, Earthscan, 2011.

[16] J. H. Kämpf, *On the Modelling and Optimisation of Urban Energy Fluxes*, PhD Thesis, EPFL, Lausanne, 2009.

[17] J. H. Kämpf and D. Robinson, *A simplified thermal model to support analysis of urban resource flows*, *Energy and Buildings*, vol. 39, no. 4, pp. 445–453, 2007.

[18] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[19] SIA, *SIA 380/1: Thermische Energie im Hochbau*, Zurich, Switzerland: Swiss Society of Engineers and Architects, 2009.

[20] swisstopo, *Schweizer Landeskarte*, Wabern, Switzerland: Bundesamt für Landestopographie (swisstopo), 2014.

[21] K. Orehounig, G. Mavromatidis, R. Evins, V. Dorer, and J. Carmeliet, *Towards an energy sustainable community: An energy system analysis for a village in Switzerland*, *Energy and Buildings*, vol. 84, pp. 277–286, 2014.

[22] Bundesamt für Statistik (BFS), *STAT-TAB: Die interaktive Statistikdatenbank*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2014.

[23] Bundesamt für Statistik (BFS), *Eidgenössisches Gebäude- und Wohnungsregister*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[24] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[25] R. Burri and K. Knecht, *Tool zur Lebenszyklusanalyse von Wärmeerzeugungssystemen*, Master Thesis, ETHZ, Zurich, 2014.

[26] J. Nipkow, S. Gasser, and E. Bush, *Der typische Haushalt-Stromverbrauch. Energieverbrauch von Haushalten in Ein- und Mehrfahmilienhäusern*, Bulletin SEV/VSE 19/2007, 2007.

[27] METEOTEST, *Meteonorm, version 7*, METEOTEST, Bern, Switzerland, 2012.

[28] H. Wallbaum, N. Heeren, M. Jakob, G. Martius, and N. Gross, *Gebäudeparkmodell*, Stadt Zürich, Zurich, Switzerland, 2010.

[29] E. Plischke, E. Borgonovo, and C. L. Smith, *Global sensitivity measures from given data*, *European Journal of Operational Research*, vol. 226, no. 3, pp. 536–550, 2013.

[30] N. Heeren, C. L. Mutel, B. Steubing, Y. Ostermeyer, H. Wallbaum, and S. Hellweg, *Environmental Impact of Buildings - What Matters?*, *Environ. Sci. Technol.*, vol. 49, no. 16, pp. 9832–9841, 2015.

[31] D. Robinson, N. Campbell, W. Gaiser, K. Kabel, A. Le-Mouel, N. Morel, J. Page, S. Stankovic, and A. Stone, *SUNtool – A new modelling paradigm for simulating and optimising urban sustainability*, *Solar Energy*, vol. 81, no. 9, pp. 1196–1211, 2007.

[32] F. Haldi and D. Robinson, *The impact of occupants' behaviour on building energy demand*, *Journal of Building Performance Simulation*, vol. 4, no. 4, pp. 323–338, 2011.

[33] D. Saner, C. Vadenbo, B. Steubing, and S. Hellweg, *Regionalized LCA-Based Optimization of Building Energy Supply: Method and Case Study for a Swiss Municipality*, *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7651–7659, 2014.

CHAPTER 4

BIG DATA GIS ANALYSIS FOR NOVEL APPROACHES IN BUILDING STOCK MODELLING

René Buffat, **Andreas Froemelt***, *Niko Heeren, Martin Raubal and Stefanie Hellweg*

Applied Energy (2017) 208 277 – 290

* The individual contribution of Andreas Frömelt consisted of significant inputs for building upon and improving a precursor model, co-developing the final model, code-contributions, and important inputs to the analysis and to the preparation of the manuscript.

ABSTRACT

Building heat demand is responsible for a significant share of the total global final energy consumption. Building stock models with a high spatio-temporal resolution are a powerful tool to investigate the effects of new building policies aimed at increasing energy efficiency, the introduction of new heating technologies or the integration of buildings within an energy system based on renewable energy sources. Therefore, building stock models have to be able to model the improvements and variation of used materials in buildings. In this chapter, we propose a method based on generalized large-scale geographic information system (GIS) to model building heat demand of large regions with a high temporal resolution. In contrast to existing building stock models, our approach allows to derive the envelope of all buildings from digital elevation models and to model location dependent effects such as shadowing due to the topography and climate conditions. We integrate spatio-temporal climate data for temperature and solar radiation to model climate effects of complex terrain. The model is validated against a database containing the measured energy demand of 1845 buildings of the city of St. Gallen, Switzerland and 120 buildings of the Alpine village of Zernez, Switzerland. The proposed model is able to assess and investigate large regions by using spatial data describing natural and anthropogenic land features. The validation resulted in an average goodness of fit (R^2) of 0.6.

4.1 INTRODUCTION

Energy consumption of buildings plays a significant role in global energy demand. Lucon et al. [1] estimated that in 2014 buildings were responsible for 32% of the world-wide final energy consumption. In view of the large contribution of fossil-fuel based energy systems to climate change, the building stock yields a high reduction potential of global anthropogenic greenhouse gas (GHG) emissions.

However, new challenges emerge with the integration of renewable energy sources into the existing energy system. For instance, electricity production becomes fluctuating and requires load balancing capabilities [2, 3]. Research studies demonstrate that buildings can play a major role for the integration of renewable energy production with novel concepts and technologies. Examples are distributed swarms of combined heat and power (CHP) plants [4, 5] or time controlled operation of heat pumps [6, 7].

To investigate the potential contribution to load balancing and GHG mitigation of these technologies, as well as planning their implementation, it is necessary to model the heat demand of buildings. Such models are typically referred to as building stock models or urban building energy models. In the past, such models were developed for different applications, such as to estimate future development of energy consumption of the building stock [8, 9], costs [10], retrofit scenarios [11, 12] or environmental impact [13, 14]. A review of different building stock models can be found in Swan and Ugursal [15], Kavgic et al. [16], Keirstead et al. [17] or Reinhart et al. [18].

In general, the models can be divided into either top-down or bottom-up. Top-down models use aggregated data as input, which are a fit of historical time series of national energy consumption or CO_2 emissions data [16]. However, top-down models do not allow for investigating decentralized energy production in detail. Bottom-up models use disaggregated data as input, and thus need extensive databases of empirical data to describe each component of a building [16]. To model larger areas, typically only one or multiple cohorts of buildings with simulated or measured heat demand consumption are used [16, 19]. Major limiting factors to model all buildings individually include the dependency on local data sources for the required input parameters of the heat demand simulation tools, the computational performance of the heat demand simulation and data processing or the limited availability of data. Current studies that model buildings individually rely on data sources that are not available on a large scale. In Zucker et al. [20] it was reported that the simulation of 72 buildings using EnergyPlus required 45 min. Therefore, this method is not suitable to perform simulations of large areas.

Current geographic information systems (GIS) [21] allow for using large spatial datasets, which have become more and more available in recent years, and thus for an intensive and detailed spatio-temporal analyses of different natural and anthropogenic processes and phenomena. However, in building stock modeling GIS is not currently used to its potential. Rather, it is primarily used for retrieving data of individual buildings from existing spatial databases or to visualize results. In Heiple and Sailor [22] building energy demand on a parcel level is estimated for large cities in the USA. Other studies describe modeling of the building heat demand in Switzerland [9, 14, 23] using spatial datasets of the official registry of buildings. In most of the studies [9, 14, 22–24] the use of GIS is limited to governmental spatial databases, and includes data such as net dwelling areas, number of floors, building age or building category (such as single or multifamily houses). Typically, these datasets are associated with the approximate location of a building and do not contain detailed information about its geometry and surfaces. These surfaces, which are the input data of physically-based heat demand models, are estimated in various ways: for instance, in [14] they are estimated using net dwelling areas combined with the assumption that all buildings have a cube shape. On the other hand, in [25] additional building footprint data, in combination with the number of floors, retrieved from registry data, are used to estimate building dimensions. In Froemelt and Hellweg [26] (see Chapter 3), transmission losses of exterior walls were identified as the most important influencing factor when modeling building heat demand and, thus, the estimate of the building surface is a crucial parameter, bringing into question the validity of the simplified building geometry assumptions in [14]. Ma and Cheng [27] use machine learning techniques to model building energy demand in New York City. Due to their different modeling approach, the use of GIS is not aimed at deriving building features such as surfaces but to determine predictors such as the distance to the coast line or nearest subway entrance. In Alhamwi et al. [28] GIS is used to process OpenStreetMap building data.

For energy related applications specific requirements need to be met. For instance, buildings need to be spatially linked to the electricity distribution grid. This can only be done if the exact position of each building is known. Both the production and the consumption of energy are fluctuating and can only be approximated but not accurately predicted [29, 30]. Despite the thermal inertia of buildings as well as integrated short-term heat storage (e.g. hot water tanks) that can attenuate the temporal variation of heat demand, many energetic applications require models with at least daily or intra daily temporal resolutions. This requires the usage of climate data of similar temporal resolution. Furthermore, external factors such as the characteristics of the surrounding topography (e.g. mountains) and objects (e.g. other buildings or vegetation) can have a significant effect on solar gains of windows through shadowing or on local weather phenomena such as sun duration and exposition. In addition, the topography can cause temperature differences in relatively small areas due to elevation differences, which is common, for example, in mountainous regions. Thus, a high spatial resolution of the climate data is required in order to detect such variations that have an impact on the energy demand of each building.

Table 4.1 shows an overview of recent physical based bottom up building stock models. Existing bottom-up models are limited by their ability to simulate large regions or use archetypes buildings to simulate the building stocks heat demand. Thus, their use is limited for assessing the load-balancing capabilities of buildings on a large scale.

Table 4.1: Comparison of different bottom up building stock models.

	Individual buildings	Building dimensions	Shading	Spatial climate data	Uncer-tainty analysis	Scale	Heat model	Validation
This work	✓	✓[a]	✓[e]	✓	✓	Country	SIA 380/1	1845 + 120
Tuominen et al. (2014) [8]						Country	IDA ICE	
Mata et al. (2013) [10]						Country	ECCABS	1400
Exner et al. (2017) [11]						Regions	EN 832	
Balaras et al. (2016) [12]						Regions	TEE-KENAK	
Saner et al. (2013) [14]	✓	(✓)[c]	(✓)[d]		✓	Regions	SIA 380/1	126 [26]
Heiple et al. (2008) [22]						City	DOE2	
Österbring et al. (2016) [25]	✓	✓[b]	(✓)[d]			City	ECCABS	433
Nageler et al. (2017) [31]	✓	✓[b]	(✓)[d]			Town	IDA ICE	69
Zucker et al. (2016) [20]	✓		(✓)[d]			Neighbor-hoods	EnergyPlus	

[a] Digital surface model.

[b] 2.5D vector dataset.

[c] Cube shape assumption.

[d] Simplified buildings.

[e] Complete surrounding topography.

The aim of this chapter is to present a method for building stock modeling over large regions using GIS and spatio-temporal data: the proposed model is intended to be applicable for different purposes in the energy field such as deriving new energy policies, spatial distribution of heat demand, detailed analyses of refurbishment scenarios, assessment of building-related environmental impacts, combined response of power grid analyses and heat demand at different time resolution. In this work, the developed model is based on studies carried out by Saner et al. [14] but with an extensive use of GIS. To simulate the heat demand of each individual building the SIA 380/1 [32] model is used. The model relies entirely on nationwide available datasets, such as registry data, digital elevation models and spatial climate data. We present novel approaches to derive building geometries and volumes from digital elevation models. By extensive use of GIS methods in combination with spatial datasets, such as digital surface models and building footprints, we overcome the issue of making assumptions about the shape of buildings. Furthermore, we are able to model the effect of shared walls between adjacent heated buildings. The model is generic and can be applied to any region where the aforementioned data are available.

Using spatial climate datasets allows for a time resolution of up to half an hour for the past 30 years. This enables for simulation of arbitrary periods, long-term averages or extrema periods. These are typically referred to as typical meteorological year (TMY) or future meteorological year (FMY) [33]. The digital surface models, including natural and anthropological objects above the ground, allow for estimating solar gains through windows, including shading effects by the surrounding topography and structures. Furthermore, since building registry data are known to not be spatially accurate [11] (i.e., not matching the corresponding buildings), we propose an algorithm to improve the joining of registry data with building datasets. Uncertainties due to gaps in the available data are accounted for by performing a Monte Carlo simulation for each building. This is especially important when the model is used for combined response of power grid analyses when the building energy demand can be the limiting factor [5]. At present, no work has demonstrated its ability to estimate, with such level of detail, the spatio-temporal distribution of heat demand of each individual building over large regions. To simulate the heat demand of each individual building the Swiss standard to determine space heat demand (SIA 380/1 [32]) is used, as shown in Saner et al. [14] This physically-based model depicts a reasonable compromise between the required input data as well as computation performance. The method for heat demand is applied to Switzerland and validated against 1845 buildings in the city of St. Gallen and 120 buildings in the alpine town of Zernez.

We first present the datasets used (section 4.2), followed by the discussion of our method to derive the heat demand for each individual building from these datasets (section 4.3). This is followed by a comparison of the simulated heat demand with measured heat demand (section 4.4). Finally, we conclude this chapter with the discussion of the obtained results (section 4.5).

4.2 Data

In the following, the different datasets used for this research are described.

4.2.1 Building Footprints

The building footprint is used to define the boundaries of a building, relate other spatial datasets such as climate data and derive building dimensions with the help of digital elevation models. Thus, the quality of the building footprints is crucial. In general, the building footprints of the cadastral survey have the best quality, as they are measured in the field by professionals. As the cadastral survey does not yet contain a digital record of every buildings footprint (coverage of 89%), OpenStreetMap [34] as well as the Swiss cartographic SwissTLM dataset [35] were used to fill data gaps, adding 7% and 4% of the total number of buildings, respectively. Concerning the latter, only footprints not intersecting a footprint from another dataset were included, to avoid double counting. The final dataset contains all buildings as of 2013, including residential buildings, offices, and other buildings.

4.2.2 Digital Elevation Models

Digital elevation models represent the topographic surface [36]. Typically, the elevation is either represented as a regularly spaced grid (or raster data) or as triangulated irregular networks (TIN). Digital surface models (DSM) represent the surface including all objects above ground, such as canopies of vegetation or building roofs, while digital terrain models (DTM) model the elevation at ground level.

The Swiss Federal Office of Topography (swisstopo) provides both a DSM and DTM raster dataset with a spatial resolution of two meters for Switzerland. Both datasets are derived from the same LIDAR (LIght Detection And Ranging) point cloud. The point cloud was created between 2001 and 2008 and has as a low point density of approximately one point per square meter. Due to the applied smoothing and the low spatial resolution the DSM does not represent buildings accurately. Buffat [37] introduced a new DSM with a resolution of 0.5 m, which was derived from the raw LIDAR data using an optimized interpolation method for buildings. The method uses error tolerant fitted planes to interpolate building roofs and improves the accuracy considerably [37].

For the cantons Basel-Landschaft, Bern, Geneva, Glarus, Schaffhausen, Solothurn, Zug and Zurich newer DSMs exist in a 0.5-m resolution. For buildings located in these cantons, the new data is used.

4.2.3 Climate Data

The Swiss weather service MeteoSwiss produces raster datasets with daily mean free-air temperature two meters above the ground for Switzerland [38]. This dataset contains data series from 1961 onwards and has a spatial resolution of 0.02 degree in latitude and longitude (equal to a resolution of roughly 1.6 km in longitude and 2.3 km in latitude). It was spatially interpolated using the method from Frei et al. [39] from the measurements of between 70 and 110 weather stations that measure the temperature two meters above the ground. The mean absolute error of the dataset is in the range from 0.5 °C over flat and hilly terrain in summer to 1.5 °C in the Alps in winter [39].

The Satellite Application Facility on Climate Monitoring (CM SAF) Surface Solar Radiation Data Set - Heliosat (SARAH) [40] provides solar radiation data for the whole of Switzerland (i.e., ±65 degree longitude and ±65 degree latitude with a spatial resolution of 0.05 degrees, corresponding to a resolution of roughly 3.8 km in longitude and 5.6 km in latitude). The dataset provides time series from 1983 to 2013 with a temporal resolution of 30 min. For each day and for each half an hour the dataset provides global irradiance (i.e., total irradiance on a flat surface), the direct normal irradiance (direct beam component of the global irradiance normalized with the cosine of the solar zenith angle) as well as the cloud albedo (amount of light reflected by the atmosphere). In [41], the dataset was validated for Switzerland with measured data from 104 weather stations. The dataset showed a mean bias of 0.18 ± 8.54 W/m^2 for stations below 1000-m elevation. With higher elevations, the dataset underestimates the solar irradiance. Around 87% of all Swiss buildings are located below a 1000-m elevation [41].

4.2.4 Buildings and Dwellings Statistics

The Swiss Federal Statistical Office maintains the Federal Register of Buildings and Dwellings (FRBD) [42]. For each included building, the register collects attributes such as a unique building identification number, parcel identification number, occupation (e.g. single family houses, multi-family houses or houses with mixed residential usage), building construction period, building footprint area, number of floors, coordinates of the building and type of the building heating system for space heating as well as warm water. It is mandatory that every Swiss residential building is included in the register.

Before 2001, the data of the register was collected by means of questionnaires that are sent to each household as part of the national census [43]. Since then the municipalities are responsible for maintaining the dataset. Kulawik [44] found that (at least for the canton of Lucerne) data quality is limited. For instance, coordinates of buildings may be outside of building footprints, and the footprint area can deviate from the area derived from the cadastral survey. Furthermore, there is no systematic quality control of the entries, which leads to differing regional data quality. In Appendix C, a comparison of the building footprint area between the cadastral building footprints and the FRBD building footprint area can be found. As the FRBD is the only comprehensive dataset available in Switzerland, it is widely used for building heat demand estimation [5, 14, 45]. Furthermore, Froemelt and Hellweg [26] (see Chapter 3) showed that the bottom-up building energy demand model of Saner et al. [14], which is partly based on the FRBD, achieved reasonable results for residential building stocks. However, it cannot be ruled out that differences between measured energy demand and the modeled result are due to inaccurate area attributes of the FRBD. Therefore, where possible the other available data sources presented in section 4.2.1 were used, leaving only building type, building construction period and, for the model validation, the type of energy carrier of the FRBD attributes.

4.3 Methods

Physically-based building heat models need a wide range of different input parameters to simulate the heat demand of one building [16, 46]. These parameters range from building dimensions (surfaces of walls, floors, windows, roofs) to physical properties of materials (thermal transmittance of walls, windows), user behavior (ventilation behavior, room temperature) and climate data (solar radiation, temperature).

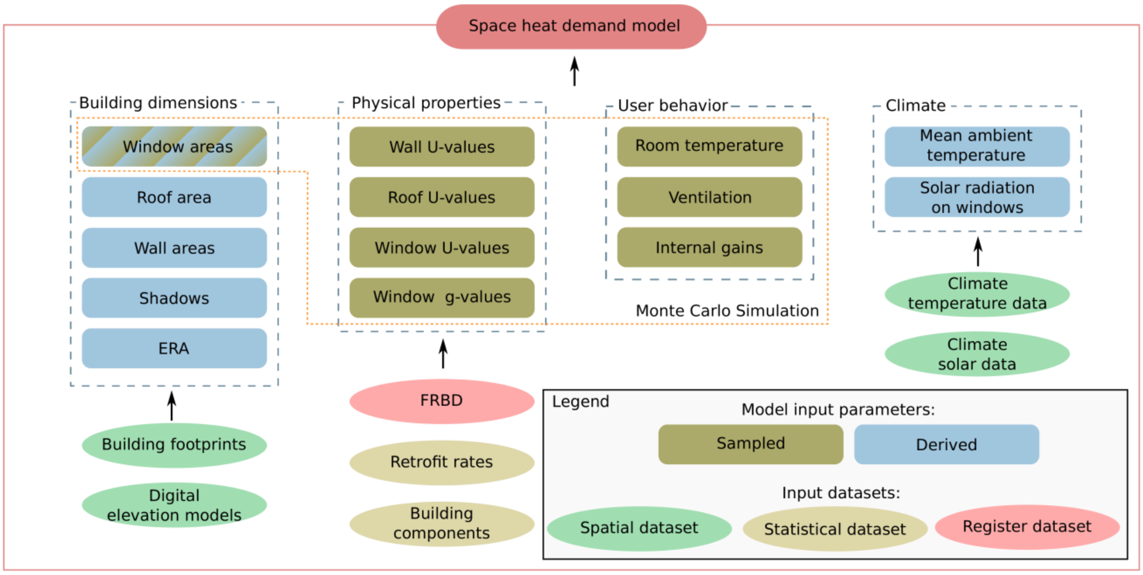


Figure 4.1: **Simplified workflow of presented method. Input parameters of the heat model are derived from spatial data where possible. If not possible, input parameters are sampled based on assumed probability distributions or statistical data. [ERA: Energy reference area, FRBD: Federal Register of Buildings and Dwellings]**

Commonly, these data are not available in a single database and need thus to be extracted from different data sources rendering the collection of data for each building not trivial. Therefore, an integral part of this section is the derivation of the required input parameters for each individual building. Figure 4.1 shows an overview of the developed workflow to derive the necessary input parameters from the heterogeneous data sources presented in section 4.2 including spatial datasets, registry datasets and statistical data. While many parameters, such as building dimensions and climate, can be derived from spatial datasets, some parameters, such as the physical properties of building components, need to be derived differently. For instance, no large scale available dataset records the building materials used to construct each building or the user behavior of the residents. However, these data gaps can be derived from statistical information or literature-informed assumptions [9, 26]. For example, the probability distributions of the facade to window ratios or the room temperature can be collected from a smaller number of buildings. A value can then be sampled from these distributions to provide an absolute value as an input to simulate the heat demand with a physically-based heat model. Thus, the simulated heat demand is subject to uncertainty. To account for this uncertainty we perform a Monte Carlo simulation with 2000 iterations, sampling from these distributions in each run. Table 4.2 lists all distributions used for the simulation. Thus, we do not calculate only one heat profile for each building, but as many as there are iterations of the Monte Carlo simulation. This results in a distribution of the heating demand for each simulated building. Figure 4.2 shows such a distribution for a randomly selected building.

In the following, we describe our approach in more detail.

Table 4.2: Parameters sample in the Monte Carlo simulation. [SFH: Single-family house, MFH: Multi-family house]

Name	Distribution	Parameters	Description
Facade to window ratio for construction period:<1919 [%]	Normal	$\mu = 15.4, \sigma = 4.6$ ($min = 5, max = 30$)	Based on [47]
Facade to window ratio for construction period: 1919–1970 [%]	Normal	$\mu = 13.6, \sigma = 4.2$ ($min = 6, max = 25$)	Based on [47]
Facade to window ratio for construction period: 1971–1980 [%]	Normal	$\mu = 15.2, \sigma = 4.1$ ($min = 10, max = 25$)	Based on [47]
Facade to window ratio for construction period: 1981–2000 [%]	Normal	$\mu = 18.5, \sigma = 4.1$ ($min = 10, max = 28$)	Based on [47]
Facade to window ratio for construction period:>2000 [%]	Normal	$\mu = 16.4, \sigma = 3.9$ ($min = 10, max = 28$)	Based on [47]
Room temperature [°C]	Normal	$\mu = 20.0, \sigma = 1.5$	Based on SIA 380/1 [32]
Surcharge if room temperature controller [K]		0 if hydronic system newer than 2006; Random selection of {1, 2}	Based on SIA 380/1 [32]
Heat gains - electricity [MJ/m²]	Normal	SFH: $\mu = 100.0, \sigma = 7.0$, MFH: $\mu = 80.0, \sigma = 7.0$	Based on SIA 380/1 [32]
Shadow factor ($F_{S,2}$)	Uniform	(0.9, 1.0)	Based on SIA 380/1 [32]
Thermal storage capacity [MJ/(m²·K]	Triangular	$left = 0.1, mode = 0.4, right = 0.5$	From Saner et al. [14]
Time of refurbishment - hydronic system			According to Wallbaum et al. [48]
Time of refurbishment - roof (flat+tilted)			According to Wallbaum et al. [48]
Time of refurbishment - floor			According to Wallbaum et al. [48]
Time of refurbishment – walls			According to Wallbaum et al. [48]
Time of refurbishment - windows			According to Wallbaum et al. [48]
Ventilation rate Dec-Feb [h^{-1}]	Lognormal	$\mu = -0.798, \sigma = 0.673$	From Murray and Burmaster et al. [49]
Ventilation rate March-May [h^{-1}]	Lognormal	$\mu = -1.177, \sigma = 0.807$	From Murray and Burmaster et al. [49]
Ventilation rate Jun-Aug [h^{-1}]	Lognormal	$\mu = -0.588, \sigma = 0.612$	From Murray and Burmaster et al. [49]
Ventilation rate Sept-Nov [h^{-1}]	Lognormal	$\mu = -1.173, \sigma = 0.540$	From Murray and Burmaster et al. [49]

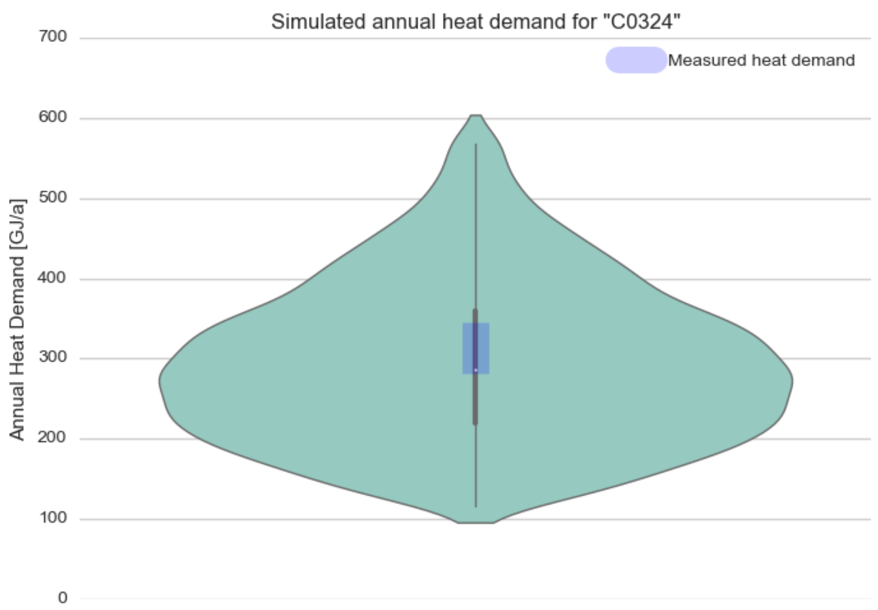


Figure 4.2: Violin plot of the results of the Monte Carlo simulation for one building. In this plot, the heat demand profiles of each Monte Carlo iteration are aggregated to yearly values. The blue area indicates the measured heat demand with respect to the uncertainty from the efficiency of the heating system.

4.3.1 Space Heating Demand Model

The SIA 380/1 Norm [32] is a building heat model used in Switzerland to verify that new and renovated buildings satisfy heat insulation obligations. It was developed by the Swiss Society of Engineers and Architects (SIA). The model is based on the EN ISO 13790 norm [50] and uses a monthly steady-state method to estimate heat demand based on equations modeling the heat balance of a building [32]. This algorithm has also been implemented in calculating heat demand in building stocks, since it offers a promising compromise between accuracy and computational effort [9, 14].

The model defines the heat demand of a building (Q_h) as the sum of the heat losses subtracted by the utilized heat gains as in equation (4.1). Heat losses and gains are thereby calculated for different time periods t.

$$Q_h = \sum_t Q_{T,t} + Q_{V,t} - \eta_{g_t} \cdot (Q_{S,t} + Q_{iP,t} + Q_{iEL,t}) \tag{4.1}$$

In equation (4.1) $Q_{T,t}$, are the transmission heat losses, $Q_{V,t}$, ventilation heat losses, η_{gt} the degree of utilization of the heat gains, $Q_{S,t}$, solar heat gains, $Q_{iP,t}$ heat gains from inhabitants, and $Q_{iEL,t}$ heat gains from electric devices.

Transmission heat losses ($Q_{T,t,c}$) include the heat losses of different building components (c) such as walls, roof, ceilings, floors, windows and thermal bridges and are calculated as shown in equation (4.2). These losses are modeled based on the temperature differences between heated and unheated areas (ΔT), the area (or in case of thermal bridges the length) of the components (A_c) and the thermal transmittance (U-values, respectively ψ-values for thermal bridges) of the materials used. Reduction factors (b_c) account for surfaces with reduced thermal loss, for example, walls against unheated rooms or soil.

$$Q_{T,t,c} = A_c \cdot U_c \cdot \Delta \mathrm{T} \cdot b_c \tag{4.2}$$

Heat gains from inhabitants ($Q_{iP,t}$) are modeled based on the number of inhabitants and the heat produced per inhabitant. The heat gains of electric devices ($Q_{iEL,t}$) are calculated based on the energy reference area (ERA) and typical mean yearly electricity heat gains per ERA. We sample the typical heat gains using a normal distribution with the mean heat gain from [32] and an assumed standard deviation of 7 MJ/m^2. Solar heat gains are modeled based on equation (4.3) using the window area ($A_{w,\alpha}$) for a specific orientation (α), the global solar radiation $\bar{I}_{sis,t,\alpha,\beta=90°}$ of period t (see section 4.3.4), a reduction factor F_{soil} of 0.9 for soiling, the total energy conductivity of a window (g-values, $g_\perp$), the reduction factor for window frames (F_F = 0.7 from [32]) and the reduction factor for shadowing ($F_{S,\alpha}$, see section 4.3.2). Just as the U-values, the g-values are sampled from probability distributions distinguished by building type and age [14, 51] (cf. section 4.3.3 for more information).

$$Q_{s,\alpha,t} = \bar{I}_{sis,t,\alpha,\beta=90°} \cdot A_{w,\alpha} \cdot F_{soil} \cdot g_\perp \cdot F_F \cdot F_{S,\alpha} \tag{4.3}$$

The original model uses the monthly mean radiation data for a set of weather stations in Switzerland. This dataset provides only the long-term average monthly radiation for four orientations

[52]. As we are able to derive location-specific radiation data for arbitrary orientations (see section 4.3.4) we increase the number of orientations to 24 by grouping the facade orientations into 15 degree wide segments.

4.3.2 Building Dimensions

Equations (4.2) and (4.3) need the areas of building components as input. Therefore, the accurate determination of building geometries is crucial for modeling building energy demand. We propose a novel method for doing so.

4.3.2.1 Wall and Roof Areas

The SIA 380/1 model distinguishes between two types of wall areas: exterior walls against ambient air and exterior walls against neighboring buildings. All areas of walls and roofs are derived using the building footprint in combination with the digital surface models. The process is divided into two steps. In the first step, the wall areas located on the building perimeter are determined. For the second step, the inside of the building footprint is processed to derive the roof as well as exterior wall surfaces located within a building footprint.

For each segment of the perimeter of a building footprint the corresponding wall area is calculated using the height difference between the DSM and the DTM. For adjoining buildings with touching perimeters, the wall area shared between both buildings needs to be derived. This is achieved by determining the building height of both buildings using a buffer region. The buffer is constructed using a depth of two meters as shown in Figure 4.3. This area is then used to sample the average height of each building within the buffer and subsequently the shared wall area for this segment.

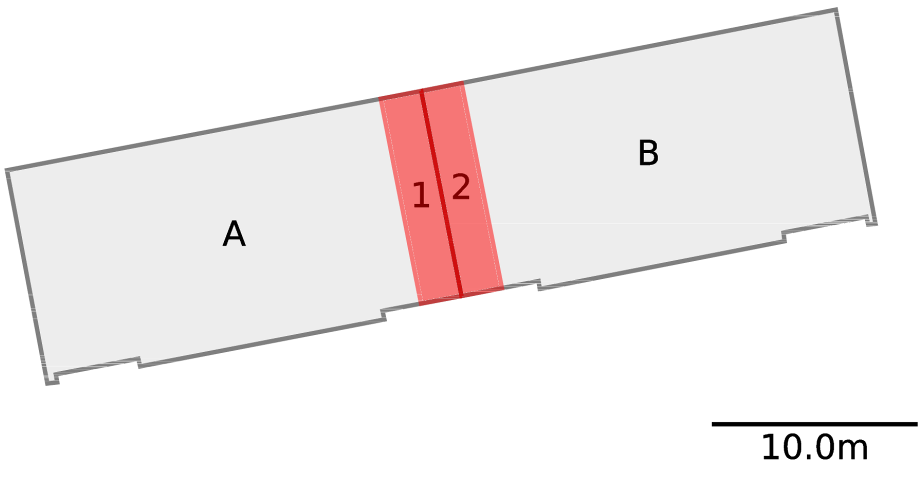


Figure 4.3: **The shared area between building A and B is determined using the difference of the average elevation of region one and two.**

In the second step, the area within a building footprint is processed. Each cell of the digital elevation models is partitioned into 8 triangles connecting to the midpoints of the neighboring cells as shown in Figure 4.4a, as was done similarly in [53]. These triangles can either be classified as roof triangles or exterior wall triangles as shown in Figure 4.4b. The classification is based on the slope angle (α) of the triangle. As the distance between the middle points of two neighboring cells is determined by the resolution of the dataset and is always greater than zero, the slope of a triangle belonging to a wall will always be less than 90 degrees but higher than typical roof slopes. We classify wall triangles as triangles with a slope greater than 60 degrees. This threshold was empirically determined using a sample of buildings. The area of roof triangles contributes only to the roof area (A_R), whereas the area of wall triangles is split into a vertical wall (A_W) and horizontal roof component.

The SIA 380/1 model differentiates between flat and slanted roofs. To classify the roofs, slope maps of each building are created using Horns formula [54] and the DSM. A building is classified as having a flat roof if the median slope of the cells within the building footprint is less than five degrees.

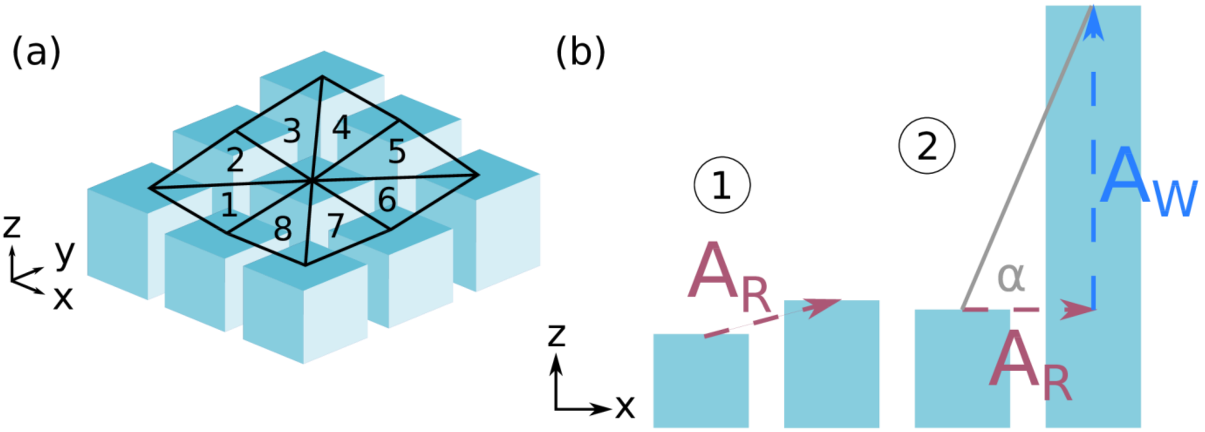


Figure 4.4: **(a) Vectorization (black lines) used to derive surface areas from the elevation raster data (blue). (b) The area of individual faces can, depending on the slope (α) of the face, contribute to only the roof area (A_R) (case 1) or both the roof as well as the wall area (A_W) (case 2).**

4.3.2.2 Window Areas and Shadows

To estimate the window areas of each facade we use facade to window area ratios. These ratios can vary between building type, construction period and region. In Ramallo-González et al. [55] facade to window ratios were surveyed for six large, culturally varied cities across the globe using Google StreetView. It was found that the ratio differs for each city. In Meier [56], a bachelor thesis supervised by the authors of this study, a similar approach was used to analyze Swiss buildings using Google StreetView. Over 539 were analyzed and a mean ratio of 12.3% was found. This corresponds well with the results from [55]. However, the study was not able to validate their method and results as no validation data was available.

In this work, we used a dataset with manually measured ratios from buildings of the village of Zernez [47]. From this, we estimated the mean and variance for the different construction periods listed in Table 4.2. Furthermore, we assumed that the ratios follow a normal distribution. To avoid extreme outliers, we assume a minimum and maximum ratio. The sampling is repeated until a ratio within these bounds is selected. These distributions are based on a sample of buildings in the village of Zernez [47].

The length of the thermal bridges of windows is derived by multiplying the window area with a factor of three, which is recommended in [32]. Based on the threshold value for thermal bridges of windows defined in [57], a ψ value of 0.1 is assumed.

The solar gain of windows is reduced for periods when the sun is blocked by obstacles, such as mountains or neighboring buildings. To account for this, the SIA 380/1 heat model provides three reduction factors ($F_{S,\alpha} = F_{S,\alpha,1} \cdot F_{S,\alpha,2} \cdot F_{S,\alpha,3}$) for the solar gain based on the horizon angle and orientation from the viewpoint of a window. The first reduction factor, $F_{S,\alpha,1}$, accounts for the effects of the topography and neighboring buildings. As exact positions of windows are unknown, horizon angles are sampled from the midpoint of a building at its average height. For each of the 24 orientations, a five-kilometer-long straight line starting from the midpoint of the building is used to sample the horizon angle every 2 m. All sample points within the current building footprint are ignored. This approach is reasonable to account for the influence of the topography - which plays an important role in Switzerland - and to some extent the influence of neighboring buildings. However, the effect of neighboring buildings might not be modeled accurately for buildings with a large footprint area, as the windows may be located too far away from the midpoint of the building.

The second and third reduction factors ($F_{S,\alpha,2}$, $F_{S,\alpha,3}$) are for shadows due to overhangs such as balconies and other elements at the side of the window. As we do not know if a building has a balcony we sample the second reduction factor uniformly between 0.9 and 1.0. A reduction factor of 0.9 corresponds to an overhang of 30 degrees. We do not model the influence of side elements and therefore set the third reduction factor to 1.0.

4.3.2.3 Energy Reference Area

The so-called energy reference area (ERA) is defined in [58] and is used in [32] to calculate the energy demand of electric devices and occupancy. It is estimated by deriving the area of each floor as shown in Figure 4.5. A ceiling height of 2.8 m based on [5] is assumed. To calculate the building height, the DSM is subtracted from the DTM. For each floor, the area of the floor with a ceiling height of at least one meter (as defined by [32]) is added to the ERA. This reduces the ERA for floors below a slanted roof.

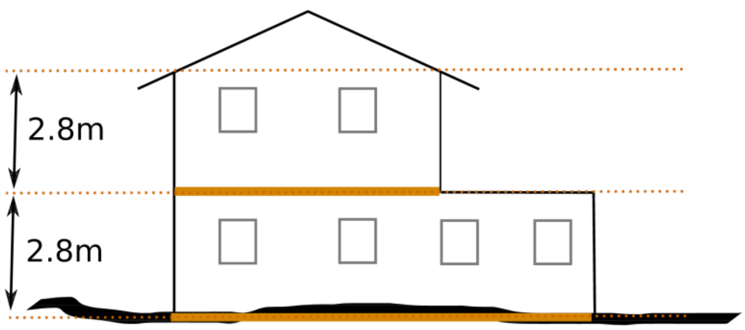


Figure 4.5: Estimation of ERA by aggregating the floor area.

4.3.3 Physical Properties

As shown in equation (4.2), heat transmission losses of building elements such as walls or floors depend on the surface areas as well as on the U-values of the materials. In equation (4.3) solar energy transmittance values (g-values) for windows are required. While surface areas can be derived from the building footprints and digital elevation models for each building individually (see section 4.3.2), U- and g-values are unknown on a building level. However in [9, 48] typical Swiss U- and g-values and retrofit rates were collected for different time periods. As in [14], the U- and g-values of a building are sampled based on the construction period and the retrofit rate.

4.3.4 Climate Data

In [14] synthetic climate data was used, which introduces additional uncertainty [59]. The interpolation of climate data is especially challenging in complex terrain, as found in the Alps [39, 60]. Thus, spatial climate raster datasets are used to obtain accurate, location specific climate data with a high temporal resolution. Building the model upon spatial climate data with a long time series has multiple benefits. For example, the ability to simulate specific time periods is especially useful if validation data is only available for certain time periods. Additionally, it allows simulating scenarios such as local historic extreme periods. In building simulation, standardized weather scenarios, such as the typical meteorological year (TMY) or future meteorological year (FMY) are used. In Herrera et al. [61] an overview of different such weather scenarios used in building simulations are given.

To estimate the long-term outside temperature we use the daily mean temperature datasets described in section 4.2.3. As the global mean temperature is rising, only the period from 1994 to 2013 is used to calculate the mean temperatures.

The amount of solar irradiance on tilted surfaces is dependent on the date, time of the day, and atmospheric condition, as well as orientation and slope of the surface [62]. To estimate the solar radiation on facades monthly mean radiation datasets were created for 24 different angles of azimuth (α), each separated by 15 degrees, and a slope (β) of 90 degrees. For each half hour (t), the SARAH data set provides the direct irradiance ($I_{sid,t}$) as well as the global irradiance ($I_{sis,t}$) on horizontal surfaces. The diffuse irradiance ($I_{dif,t}$) is then calculated with equation (4.4).

$$I_{dif,t} = I_{sis,t} - I_{sid,t} \tag{4.4}$$

Direct irradiance on tilted surfaces ($I_{sid,t,\alpha,\beta}$) can be calculated geometrically [62]. Different models exist to estimate diffuse radiation on inclined surfaces ($I_{dif,t,\alpha,\beta}$) [63]. We use the model of Perez [64] as it showed good results in [63]. Reflected solar radiation on inclined surfaces ($I_{ref,t,\alpha,\beta}$) is calculated using [63]. Global irradiance on tilted surfaces is then derived using equation (4.5):

$$I_{sis,t,\alpha,\beta} = I_{dif,t,\alpha,\beta} + I_{sid,t,\alpha,\beta} + I_{ref,t,\alpha,\beta} \tag{4.5}$$

Solar irradiance on tilted surfaces was calculated for every half hour and azimuth angle. The half hour irradiance was then used to calculate the mean irradiance $\bar{I}_{sis,t,m,\alpha,\beta}$ for each month (m) and half hour interval (t). This results in a typical daily profile for every month, which was then aggregated to retrieve the solar radiation for each month using equation (4.6).

$$\bar{I}_{sis,m,\alpha,\beta} = days_m \cdot 0.5 \sum_t \bar{I}_{sis,t,m,\alpha,\beta} \tag{4.6}$$

4.3.5 User Behavior

Apart from the buildings physical properties the user-specific parameters (e.g. indoor room temperature, heat gains from electric devices and air exchange rate) enter the heat demand calculation. These are also sampled from probability distributions based on [14, 26, 49, 51]. The parameters used can be found in the Appendix C.

Ventilation heat losses are modeled based on the specific heat storage capacity of air, derived by the elevation above sea level of the building, the air exchange rate and the volume of a building. The volume of a building above ground is derived using the DSM and DTM. However, this volume includes also unheated spaces, such as stair cases. The net volume relevant for the air exchange is then calculated using a reduction factor of 0.8, or 0.76 for residential buildings with less than 3 floors based on [65]. In Murray and Burmaster [49] empirical air exchange rates are determined for four climatic regions differentiated by the number of heating days and four seasons of the continental US. Air exchange rates can heavily depend on the behavior of the inhabitants (e.g. by leaving windows open). Murray and Burmaster [49] reflect the dependence of air exchange rates on inhabitants by providing fitted probability distributions to the measurements rather than single values. We use the distributions of the second coldest climatic region as the heating days of this region correspond best to Switzerland. However, due to Murray and Burmaster low number of measurements for the summer season, distributions of the next warmest region are utilized for this season. Despite the fact that no distributions are available for Europe, the distributions used, with a median of 0.41 h^{-1} and higher air exchange rates in summer compared to winter are comparable to European data [66–68].

4.3.6 Data Integration

In our model, a building is defined by its building footprint. We determine these by combining different datasets (section 4.2). Integrating the building footprints with the raster data is a straightforward process as the spatial relation can be used to connect the data. For example, a building footprint is joined with the temperature dataset by using the cell of the temperature raster which contains the centroid of a building.

This is different for the FRBD dataset. While this dataset includes both a unique building identifier as well as coordinates for each building, this identifier is often not present in the footprint dataset. As shown in Figure 4.6, where the blue dots represent the coordinates of the FRBD buildings and the dark grey polygons the building footprints of one parcel (light grey), the FRBD coordinates are not always within the corresponding building footprints. This makes matching of FRBD records to building footprints difficult. An example is shown in Figure 4.6, where matching a FRBD record simply to the nearest building footprint would lead to wrong matches.

We developed an algorithm to match building footprints and the FRBD dataset. This algorithm uses multiple techniques to match records of both datasets, including the detection of typographical errors using the Damerau-Levenshtein distance [69] and minimum cost matching [70]. A detailed description of the algorithm can be found in Appendix C.

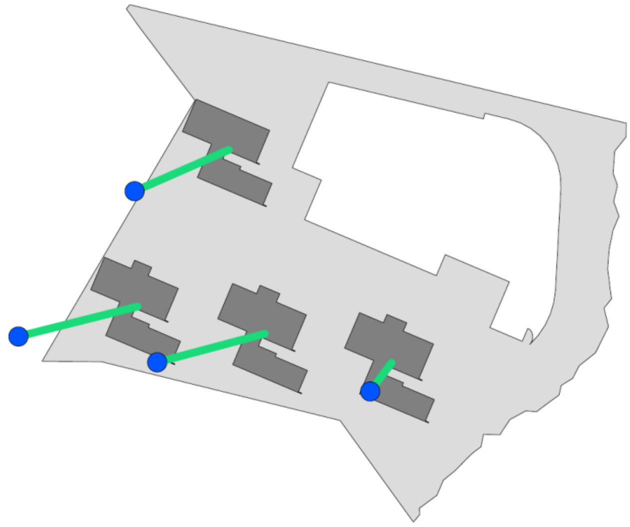


Figure 4.6: Matching of FRBD building coordinates (blue) to building footprints (dark grey) for one parcel (light grey). The green lines show the assignment of FRBD records to building footprints.

4.4 Results and Validation

For validation, the model was applied to the city of St. Gallen, located in eastern Switzerland, as well as the Alpine village of Zernez. Due to the small size of the validation dataset of Zernez, its results are only shown in Appendix C. St. Gallen is the eighth largest city in Switzerland with a population of roughly 75,000 inhabitants. It holds the label of Energy city and has set the target to reduce its energy consumption from 2005 to 2020 by 15%. To monitor the progress, the city collects data about the yearly energy consumption of its building stock. Specifically, it collects final energy consumption of all buildings connected to either the gas grid or district heating networks. The dataset contains the energy measured at so-called energy reference points, which are either the locations of gas boilers or heat meters for district heating networks. A reference point can measure the energy demand of one building but also multiple adjoined buildings or neighboring buildings connected with a micro district heating network. In order to avoid ambiguity issues, only reference points that correspond to exactly one simulated building were used in the validation.

Figure 4.7 shows the distributions of normalized heat consumptions for different building types and construction periods. The consumed heat was normalized using the above ground volume as this number can be accurately calculated using the building footprints and digital elevation models. The number of buildings of the validation dataset is not evenly distributed for the different categories of construction periods and building types. For example, the dataset contains significantly more buildings built before 1946 than newer buildings. This influences the results when the age of the buildings is not considered. The heat consumption is highest for buildings built between 1920 and 1990, however it is lower for buildings constructed outside of these years. This pattern is present for all building types and is in accordance with existing studies [71]. It should be noted that the variation of the heat consumption per category (construction period and building type) is rather large.

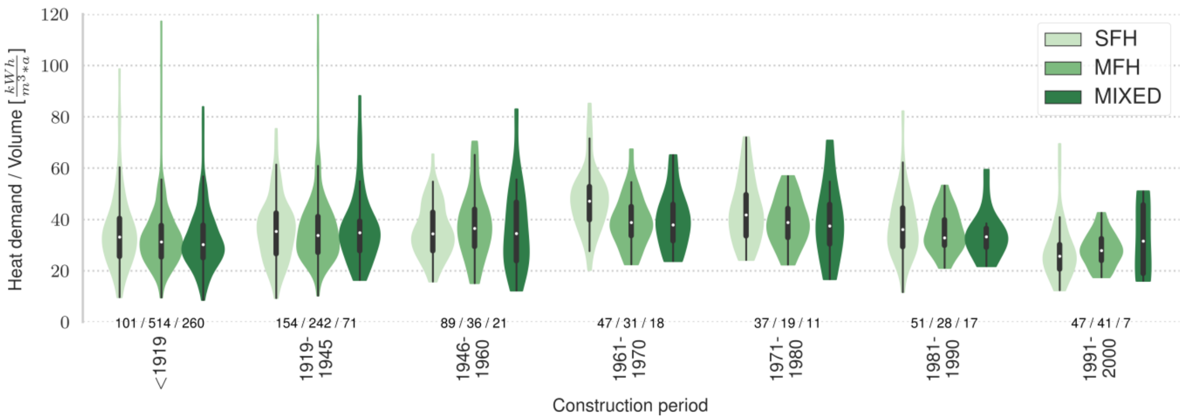


Figure 4.7: **Violin plots of measured heat consumption of space heat (including warm water) normalized by the above ground building volume for different construction periods and the building types single-family houses (SFH), multi-family houses (MFH), and mixed residential usage (MIXED). Numbers below the violins represent the sample size.**

To compare the simulated heat demand, the final energy consumption (measured energy demand) was converted to the useful heating energy by using the efficiency of the heating system as was done in Froemelt and Hellweg [26] (see Chapter 3). As the efficiency of the heating system of a particular building is not known, a low, average and high efficiency, depending on the energy source (gas or district heating), is assumed. The consumed energy of a reference point is not necessarily used exclusively for space heating. It can be also used to produce domestic hot water or for cooking in the case of gas. Based on the FRBD dataset we can distinguish between the energy source of the heating system for space heating and warm water. If the energy source is not the same for both heating systems, we can assume that the energy is exclusively used for space heat. Domestic hot water heat demand is responsible of roughly 10% of the total heat demand for buildings constructed before 1980 (see Appendix C). In [72], the energy demand for cooking was estimated to be roughly six times smaller than the warm water energy demand. As the energy sources of stoves are not known and the heat demand for cooking is in the lower single-digit percentage range we neglect the amount of gas that is used for cooking.

For the validation, we used two different sets of buildings. The first set only considers buildings with a different energy source for space heating and warm water. With this set we eliminate the uncertainty of the warm water modeling but this considerably limits the number of samples. The second set of buildings includes all buildings. For this set we take also warm water into account. Warm water heat demand (Q_{ww}) is modeled by equation (4.7) with the number of inhabitants (P) as well as a typical annual hot water demand per inhabitant of 833 kWh taken from [32].

$$Q_{ww} = P \cdot 833 \tag{4.7}$$

Figure 4.8 shows boxplots of the relative errors between the median simulated heat demand of the Monte Carlo simulation and the measured energy demand converted to the useful heating demand by using the median efficiency of the heating system for different construction periods and building types. It can be seen that the model overestimates the heat demand of single family houses. The relative error is smaller for multi-family houses and buildings with mixed residential usages. A more detailed figure, also showing the different construction periods, can be found in Appendix C. This figure shows that energy demand of single family houses older than 1961 are predominantly underestimated while for newer houses the relative error of the model is much smaller. As 82% of the single-family houses of the validation dataset are built before 1961, the results of Figure 4.8 are skewed. The opposite is the case for multi-family houses. While the median relative error is close to zero for buildings built before 1961, the heat demand of newer multi-family houses built after 1960 are underestimated.

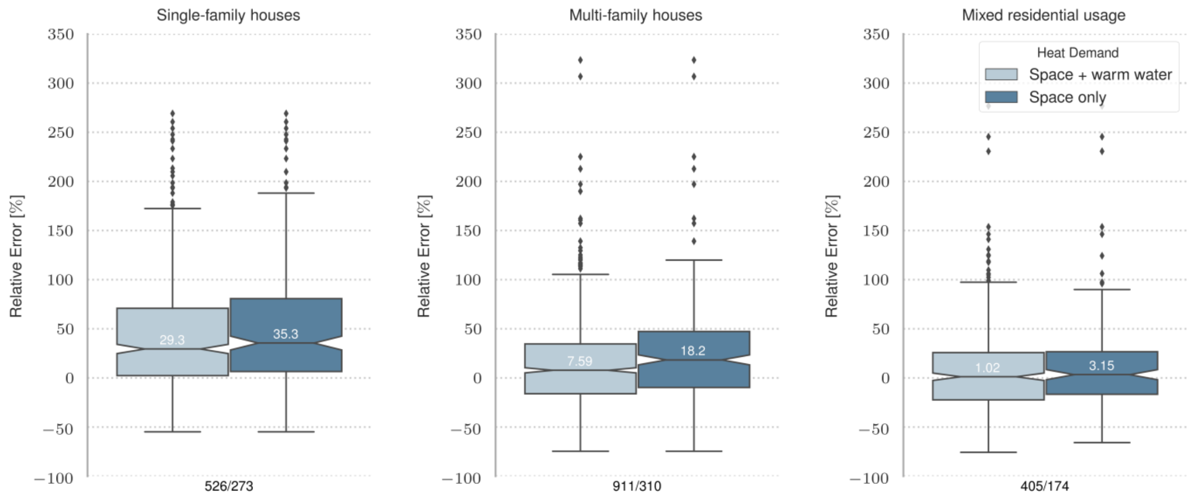


Figure 4.8: **Relative error of simulated heat demand (only space heating and space heating+warm water) to the measured consumed energy demand for different building types. Sample size n is given below the plots. The boxes indicate the interquartile range (IQR) between the first and the third quartile. The whiskers extend until 1.5·*IQR*. The notches in the boxes indicate the confidence interval for the median. The numbers colored in white represent the median values.**

Figure 4.9 illustrates scatter plots between the measured and simulated space heating demand, distinguishing between the different building types. Again, median simulated heat demands and average heat system efficiencies are used. Single family houses typically have a smaller volume and thus also a smaller heat demand. Especially for houses with a heat demand of less than 40 MWh/a the results show a high variance. The smaller volume of single family houses makes their estimated heat demand more sensitive to heated basements or unheated spaces above ground such as garage. The model performs better when simulating larger single-family houses. In general, buildings with a large energy demand above 200 MWh/a tend to be underestimated. However, most of these buildings are of mixed residential usage, thus can have additional heat demands that are not accounted for. It is possible that these samples are outliers, but due to the small sample size no statements can be made. A version of Figure 4.9 that also distinguishes between different construction periods can be found in Appendix C. Overall the model fits the measured heat demand quite well with an R^2 of over 0.5. Exceptions are the construction periods of 1961–1970 and 1971–1979 as well as single-family houses.

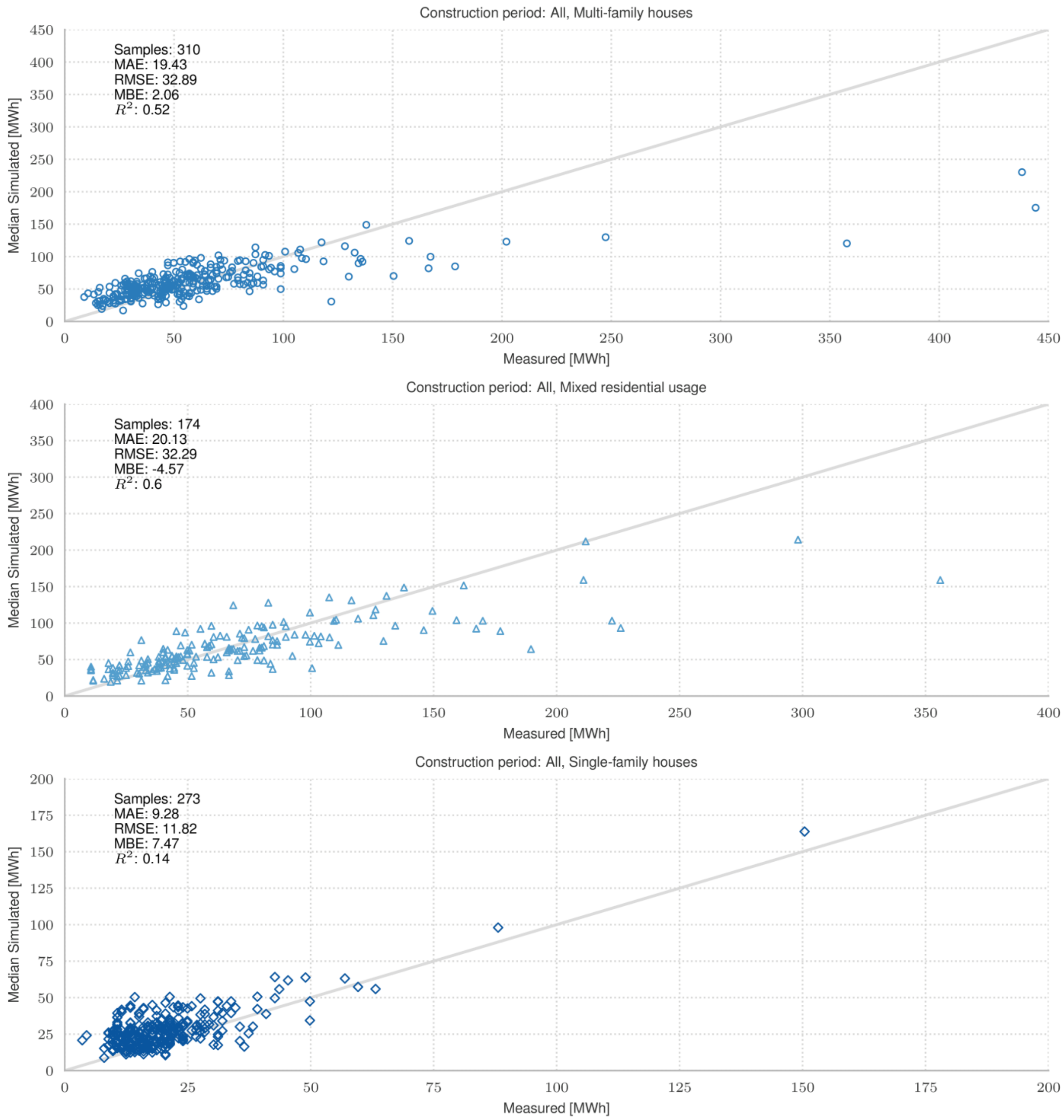


Figure 4.9: **Comparison of simulated and measured space heating demand for the building types single-family houses (SFH), multi-family houses (MFH), and mixed residential usage (MIXED).**

More statistical metrics which investigate the goodness of the model can be found in Appendix C, including mean absolute error (MAE), root mean squared error (RMSE), mean relative error (MRE), mean bias error (MBE) as well as the coefficient of determination (R^2) for the different building categories and building construction periods. As the average heat demand within each category can differ significantly the MAE, RMSE and MBE are given as percentage to the average heat demand of their category (building type, construction period). The formulas used to calculate the different statistics are given in Appendix C.1.

For Figures 4.8 and 4.9 only the median simulated heat demand is used. This ignores the modeled uncertainty of the Monte Carlo simulation. Especially for old buildings, the uncertainty can be high, due to a wide range of possible retrofit scenarios. In Figure 4.10, both the modeled uncertainties as well as the uncertainties of the validation data are visualized in a manner similar to [26] (see Chapter 3). This plot compares for each building type both the measured (green), as well as the simulated (blue) warm water and space heating demands. The x axis contains all buildings sorted by their measured heat demand while the y axis shows the cumulative heat demand of all buildings with a smaller heat demand than the current building. The uncertainty of the measured heat demand results from the minimum and maximum efficiency of the heating system of a building. For the simulated heat demand both the 50% percentile range (dark blue, using the 25% and 75% percentiles) as well as the 90% percentile range (light blue, using the 5% and 95% percentiles) are shown.

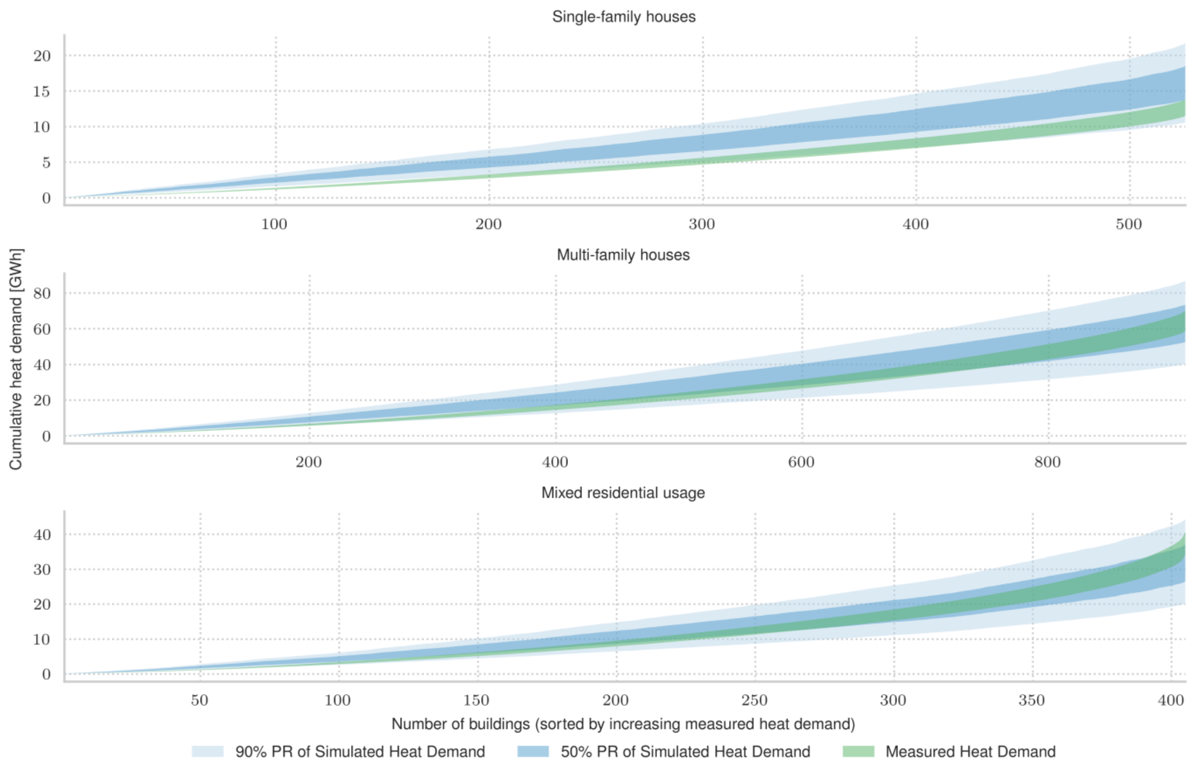


Figure 4.10: Cumulative space heating and warm water demand of simulated (blue) and measured (green) heat demand. The 50% and 90% percentile range (PR) of simulated heat demand, indicating the interval between the 25% and 75%, respectively 5% and 95% percentiles are shown.

The model shows a better fit for the multi-family houses as well as mixed residential usage when considering all buildings. Again, in Appendix C a version of Figure 4.10 can be found also distinguishing between the different construction periods. It can be seen that buildings with low agreement between modeled and measured demand level themselves out over the study area. The cumulative simulated energy demand for single family houses is still within the 90% percentile range of the model, while it overlaps with the 50% percentile range for the other building types.

For single-family and multi-family houses the slope of the measured and simulated heat demands are generally similar, which means that the error of the model is small. Thus, the error of the model is found mainly with buildings of either small or large heat demand. For mixed residential usage, we see an increase of the slope for buildings with higher measured energy demand while the simulated and measured heat demand overlap. This indicates that although these buildings can have heat demands other than space heat or warm water, or different usage patterns, a fair share of the mixed residential buildings can be accurately modeled.

4.5 DISCUSSION

4.5.1 Building Data

The heat demand of a building is dependent on many parameters. In [26, 51] the most sensitive parameters for space heating demand were room temperature and U-values (see Chapter 3). From equation (4.2) it can be seen that an error in the estimation in a components area contributes equally to the error of heat demand estimation than the U-value of this component. This means that, for example, a 10% error in the estimation of the wall area is equal to a 10% error of the U-value of this wall. Likewise, thermal losses due to ventilation are highly dependent on the building volume. By using precise building footprints of the cadastral survey as well as high resolution digital elevation models the surface areas and building volumes can be derived more accurately compared to previous work such as [14, 26]. Knowing the spatial location of buildings enables us to incorporate the effect of shadowing into the estimation of the solar gains as well as to derive location specific climate data for each building.

Digital elevation models and building footprints do not allow for differentiating between heated and unheated floor area. Building wings such as staircases or garages located within the building footprint remain usually unheated. While staircases typically comprise only a small part of the volume of buildings, other unheated parts (e.g. elevated cellars or attics) might be more important. However, there might also be heated space below ground which will not be considered by our model resulting in an underestimation of the effective heating demand Therefore we had to make assumptions concerning the estimation of heated floor area (ERA, see section 4.3.2), which introduced a source for errors. However, the influence of the ERA parameter in the SIA 380 heat model is limited since it is only used to sample the heat gains due to inhabitants and electric devices. Both were not identified as significant parameters in [26] (see Chapter 3).

4.5.2 User Behavior

It is well known that user behavior is one of the major factors contributing to the discrepancy between simulation prediction and real energy use [73]. This increases the uncertainty of the model. However, improving the model for this aspect will be difficult as no better data are available or could easily be collected on a large scale. It is known that the occupants behavior can impact the energy demand by more than 100% [15, 16, 74, 75]. Causes are the choice of the room temperature, non-optimal ventilation patterns (e.g. leaving windows or doors open), changing occupancy of buildings during certain time periods (e.g. holidays or secondary residences) or the number of inhabitants can be larger or smaller compared to a typical household. Variation induced by user behavior was covered by Monte Carlo simulations as far as possible.

4.5.3 Climate Data

The importance of using spatial climate datasets arises when large regions are analyzed. For small regions, such as a village or town, the change in temperature and solar radiation can be assumed to be equal, allowing the use of a single weather station as source for the climate data. However, for a country like Switzerland with its complex topography, where in mountainous regions a distance of a few kilometers can result in an elevation change of several 1000 m, the spatial resolution of the climate data is important. A 1000-m elevation change results in a temperature change of several degrees, which is relevant for building heat demand modeling [76].

4.5.4 Energy Demand Model

The validation showed that energy demand of single-family houses is overestimated by the model. This overestimation can have many possible origins, such as different estimations of the ERA or an increased level of renovations. Additionally, as discussed earlier, the impact of unheated spaces is likely to be higher in small buildings such as single-family houses. The heat demand of multi-family houses built after 1960 is underestimated. Both effects could be a consequence of an over-optimistic or, respectively, pessimistic selection of U-values or user behavior.

As buildings are renovated over their lifetime their energy efficiency changes. Building components, such as windows, can be replaced multiple times during the lifespan of a building. Many different scenarios are possible in particular for older buildings. We model these uncertainties by using statistical retrofit rates. Thus, the variance in the comparison using only the median simulated value, as in Figure 4.8 is higher for old buildings compared to recent buildings with no renovations. Nevertheless, we account for this as can be seen in Figure 4.10.

The model seems to have a systematic error for space heating demand. For instance, old single-family houses are consistently overestimated and newer multi-family houses are underestimated. Such an error suggests that certain assumptions should be investigated further in the future. One possible source could be faulty or outdated data in the typology that was used [48]. Furthermore, thermal bridges have a higher relative impact on highly insulated buildings. U-values of buildings as well as retrofit rates might have regional dependencies that are not accounted for. This requires additional datasets with measured heat demand.

4.5.5 Data Availability

While some datasets we used are specific to Switzerland, similar datasets can be found for many other countries. For example, we use building footprints from the cadastral survey. Such footprints are surveyed in many other industrialized countries, such as Germany, France or the Netherlands. An alternative source are the widely available building footprints of OpenStreetMap. Several countries exist that have a complete coverage of LIDAR data (Finland, Netherlands, Slovenia) or are in the process of doing so (Latvia, Poland, Spain, Sweden, England) [77]. Google maps demonstrates that in the near future, 3D buildings models created from aerial photos could be available worldwide. Building registries similar to the FRBD are available for many different countries with a national census, such as Germany [78] or Austria [79]. In fact, the building classification is standardized among the countries of the European Union [80] and Switzerland. The CM SAF SARAH solar dataset covers Europe and most of Africa. As weather satellites cover most of the inhabited regions worldwide, similar dataset exists for other parts of the world, such as North America [81]. Similarly, spatial temperature datasets can be computed for regions where readings from land based measurement stations are available. This is the case for most of the industrial nations. Thus, the applicability of the developed method is not limited to Switzerland.

4.5.6 Applications

The number of applications of a physically-based bottom-up building energy stock model is manifold. Recent studies begin to study the interaction between buildings and the electricity grid [5, 82, 83]. Our model allows large scale analysis of such concepts for Switzerland on different scales ranging from neighborhoods to the whole nation.

Furthermore, the model can be used as a basis for an energetic retrofit decision support system for policy makers or building owners [84]. As our model includes each residential building in Switzerland, users of such a system could use the modeled building data to test different retrofit scenarios for their environmental, economic and energetic performance.

4.6 Conclusions

This chapter presents new approaches for building stock modeling and illustrates their implementation. Using spatial datasets such as building footprints and digital elevation models allows for the derivation of wall and roof areas individually for each building. This reduces model uncertainties compared to previous models, such as Saner et al. [14].

Replacing synthetic climate data with spatial datasets allows simulating heat demand with location specific data. Due to their high temporal resolution of a half hour respectively daily the time resolution of the model can be increased. The high temporal resolution of the data used facilitates also the application of the model for complex optimizations such as the sizing and operation strategies of existing and novel energy systems [5]. Additionally, using historical data allows for the simulation of extrema periods and facilitates validation with measured data.

The model was validated against an extensive database revealing that the model performance varies with building types and construction periods. Single-family houses especially are less accurately modeled compared to multi-family houses and buildings with mixed usage. However, given the uncertainty due to data gaps, the results can be considered good with an overall goodness of fit R^2 of 0.6. For some categories, the goodness of fit reaches up to an R^2 of 0.81.

We intend to further improve the model. As determined in Froemelt and Hellweg [26] (see Chapter 3), the age of buildings and thus the large span of possible renovations is responsible for a large part of the uncertainties. A dataset with a sufficiently large number of samples would allow the model to derive more up-to-date input parameters through statistical analysis. We also intend to include environmental assessments in the future. Furthermore, we developed a construction material model of the Swiss building stock, which can be integrated with the current model in a future step [85]. This will reduce uncertainty concerning thermal resistance of the building envelope as we account for actual construction elements.

The spatial and temporal resolution of the model allows for a coupling with other models, with the Multi-Agent Transport Simulation (MATSim) [86] model being an especially interesting option. This would allow for a better understanding of the occupancy of buildings and would thus provide better insights into the energy consumption of households.

4.7 REFERENCES

[1] O. Lucon, D. Ürge-Vorsatz, A. Zain Ahmed, H. Akbari, P. Bertoldi, L. F. Cabeza, N. Eyre, A. Gadgil, L. D. D. Harvey, Y. Jiang, E. Liphoto, S. Mirasgedis, S. Murakami, J. Parikh, C. Pyke, and M. V. Vilariño, *Buildings*, in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, O. Edenhofer et al., Ed. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014.

[2] A. Castillo and D. F. Gayme, *Grid-scale energy storage applications in renewable energy integration: A survey*, *Energy Conversion and Management*, vol. 87, pp. 885–894, 2014.

[3] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari, *Review of energy system flexibility measures to enable high levels of variable renewable electricity*, *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 785–807, 2015.

[4] P. Vögelin, B. Koch, G. Georges, and K. Boulouchos, *Heuristic approach for the economic optimisation of combined heat and power (CHP) plants: Operating strategy, heat storage and power*, *Energy*, vol. 121, pp. 66–77, 2017.

[5] P. Vögelin, G. Georges, F. Noembrini, B. Koch, K. Boulouchos, R. Buffat, M. Raubal, G. Beccuti, T. Demiray, E. Panos, and R. Kannan, *System modelling for assessing the potential of decentralised biomass-CHP plants to stabilise the Swiss electricity network with increased fluctuating renewable generation*, Bern-Ittigen, Switzerland, 2016.

[6] C. Protopapadaki and D. Saelens, *Heat pump and PV impact on residential low-voltage distribution grids as a function of building and district properties*, *Applied Energy*, vol. 192, pp. 268–281, 2017.

[7] E. Georges, B. Cornélusse, D. Ernst, V. Lemort, and S. Mathieu, *Residential heat pump as flexible load for direct control service with parametrized duration and rebound effect*, *Applied Energy*, vol. 187, pp. 140–153, 2017.

[8] P. Tuominen, R. Holopainen, L. Eskola, J. Jokisalo, and M. Airaksinen, *Calculation method and tool for assessing energy consumption in the building stock*, *Building and Environment*, vol. 75, pp. 153–160, 2014.

[9] N. Heeren, M. Jakob, G. Martius, N. Gross, and H. Wallbaum, *A component based bottom-up building stock model for comprehensive environmental impact assessment and target control*, *Renewable and Sustainable Energy Reviews*, vol. 20, pp. 45–56, 2013.

[10] É. Mata, A. S. Kalagasidis, and F. Johnsson, *A modelling strategy for energy, carbon, and cost assessments of building stocks*, *Energy and Buildings*, vol. 56, pp. 100–108, 2013.

[11] D. Exner, V. D'Alonzo, G. Paoletti, R. Pascual, and R. Pernetti, *Building-stock analysis for the definition of an energy renovation scenario on the urban scale*, Springer International Publishing, 2017.

[12] C. A. Balaras, E. G. Dascalaki, K. G. Droutsa, and S. Kontoyiannidis, *Empirical assessment of calculated and actual heating energy use in Hellenic residential buildings*, *Applied Energy*, vol. 164, pp. 115–132, 2016.

[13] A. Mastrucci, A. Marvuglia, U. Leopold, and E. Benetto, *Life Cycle Assessment of building stocks from urban to transnational scales: A review*, *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 316–332, 2017.

[14] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[15] L. G. Swan and V. I. Ugursal, *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques*, *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.

[16] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, *A review of bottom-up building stock models for energy consumption in the residential sector*, *Building and Environment*, vol. 45, no. 7, pp. 1683–1697, 2010.

[17] J. Keirstead, M. Jennings, and A. Sivakumar, *A review of urban energy system models: Approaches, challenges and opportunities*, *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3847–3866, 2012.

[18] C. F. Reinhart and C. Cerezo Davila, *Urban building energy modeling - A review of a nascent field*, *Building and Environment*, vol. 97, pp. 196–202, 2016.

[19] F. Vásquez, A. N. Løvik, N. H. Sandberg, and D. B. Müller, *Dynamic type-cohort-time approach for the analysis of energy reductions strategies in the building stock*, *Energy and Buildings*, vol. 111, pp. 37–55, 2016.

[20] G. Zucker, F. Judex, M. Blöchle, M. Köstl, E. Widl, S. Hauer, A. Bres, and J. Zeilinger, *A new method for optimizing operation of large neighborhoods of buildings using thermal simulation*, *Energy and Buildings*, vol. 125, pp. 153–160, 2016.

[21] P. Longley, M. Goodchild, D. Maguire, and D. Rhind, *Geographic information systems and science*, Wiley, 2010.

[22] S. Heiple and D. J. Sailor, *Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles*, *Energy and Buildings*, vol. 40, no. 8, pp. 1426–1436, 2008.

[23] M. Jakob, H. Wallbaum, G. Catenazzi, G. Martius, C. Ngeli, and B. Sunarjo, *Spatial building stock modelling to assess energy-efficiency and renewable energy in an urban context*, in *Proceedings of CISBAT*, 2013.

[24] A. Mastrucci, O. Baume, F. Stazi, and U. Leopold, *Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam*, *Energy and Buildings*, vol. 75, pp. 358–367, 2014.

[25] M. Österbring, É. Mata, L. Thuvander, M. Mangold, F. Johnsson, and H. Wallbaum, *A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model*, *Energy and Buildings*, vol. 120, pp. 78–84, 2016.

[26] A. Froemelt and S. Hellweg, *Assessing Space Heating Demand on a Regional Level: Evaluation of a Bottom-Up Model in the Scope of a Case Study*, *Journal of Industrial Ecology*, vol. 21, no. 2, pp. 332–343, 2017.

[27] J. Ma and J. C. P. Cheng, *Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology*, *Applied Energy*, vol. 183, pp. 182–192, 2016.

[28] A. Alhamwi, W. Medjroubi, T. Vogt, and C. Agert, *GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas*, *Applied Energy*, vol. 191, pp. 1–9, 2017.

[29] M. De Felice, A. Alessandri, and P. M. Ruti, *Electricity demand forecasting over Italy: Potential benefits using numerical weather prediction models*, *Electric Power Systems Research*, vol. 104, pp. 71–79, 2013.

[30] J. Pascual, J. Barricarte, P. Sanchis, and L. Marroyo, *Energy management strategy for a renewable-based residential microgrid with generation and demand forecasting*, *Applied Energy*, vol. 158, pp. 12–25, 2015.

[31] P. Nageler, G. Zahrer, R. Heimrath, T. Mach, F. Mauthner, I. Leusbrock, H. Schranzhofer, and C. Hochenauer, *Novel validated method for GIS based automated dynamic urban building energy simulations*, *Energy*, vol. 139, pp. 142–154, 2017.

[32] SIA, *SIA 380/1: Thermische Energie im Hochbau*, Zurich, Switzerland: Swiss Society of Engineers and Architects, 2009.

[33] K. Hubbard, K. Kunkel, A. DeGaetano, and K. Redmond, *Sources of uncertainty in the calculation of the design weather conditions in ASHRAE Handbook-Fundamentals. ASHRAE Research Project RP-1171, Final Report*, Atlanta.

[34] H. Fan, A. Zipf, Q. Fu, and P. Neis, *Quality assessment for building footprints data on OpenStreetMap*, *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 700–719, 2014.

[35] E. Schmassmann and R. Bovier, *Topografisches Landschaftsmodell TLM: swissTLM3D*, *Geomatik Schweiz*, vol. 108, no. 9, p. 407, 2010.

[36] M. Worboys and M. Duckham, *GIS: a computing perspective, Section Edition*, CRC Press, 2004.

[37] R. Buffat, *Feature-aware surface interpolation of rooftops using low-density Lidar data for photovoltaic applications*, in *Geospatial Data in a changing world, lecture notes in geoinformation and cartography*, T. Sarjakoski, M. Santos, and L. Sarjakoski, Eds. Cham: Springer, 2016, pp. 337–50.

[38] MeteoSchweiz, *Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD*, 2013.

[39] C. Frei, *Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances*, *International Journal of Climatology*, vol. 34, no. 5, pp. 1585–1605, 2014.

[40] R. Müller, U. Pfeifroth, C. Träger-Chatterjee, R. Cremer, J. Trentmann, and R. Hollmann, *Surface Solar Radiation Data Set - Heliosat (SARAH) - Edition 1*, EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF), 2015.

[41] R. Buffat and S. Grassi, *Validation of CM SAF SARAH solar radiation datasets for Switzerland*, in *Proceedings of 2015 IEEE International Renewable and Sustainable Energy Conference, IRSEC 2015*, 2015.

[42] Bundesamt für Statistik (BFS), *Eidgenössisches Gebäude- und Wohnungsregister - Merkmalskatalog*, Neuchâtel, Switzerland, 2012.

[43] Bundesamt für Statistik (BFS), *Eidgenössische Volkszählung 2000, Gebäude, Wohnungen und Wohnverhältnisse*, Neuchâtel, Switzerland, 2004.

[44] M. Kulawik, *Grundlagenstudie Energetisch relevante Gebäudedaten in Luzerner Gemeinden*, 2014.

[45] M. Erb and S. Gutzwiller, *Ausbau von WKK in der Schweiz WKK-Standortevaluation auf Basis einer GIS-Analyse*, 2011.

[46] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, *State of the art in building modelling and energy performances prediction: A review*, *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 272–288, 2013.

[47] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[48] H. Wallbaum, N. Heeren, M. Jakob, G. Martius, and N. Gross, *Gebäudeparkmodell*, Stadt Zürich, Zurich, Switzerland, 2010.

[49] D. M. Murray and D. E. Burmaster, *Residential Air Exchange Rates in the United States: Empirical and Estimated Parametric Distributions by Season and Climatic Region*, *Risk Analysis*, vol. 15, no. 4, pp. 459–465, 1995.

[50] International Organization for Standardization (ISO), *Technical Committee 163/SC 2, Calculation methods. Energy performance of buildings – calculation of energy use, vol. 13790 for space heating and cooling*, Geneva, Switzerland, 2008.

[51] N. Heeren, C. L. Mutel, B. Steubing, Y. Ostermeyer, H. Wallbaum, and S. Hellweg, *Environmental Impact of Buildings - What Matters?*, *Environ. Sci. Technol.*, vol. 49, no. 16, pp. 9832–9841, 2015.

[52] Schweizerischer Ingenieur- und Architekten-Verein (SIA), *Klimadaten für Bauphysik, Energie- und Gebäudetechnik, Vol. 2028, ed. 2010*, SIA, Zurich, Switzerland, 2010.

[53] J. Jenness, *Calculating landscape surface area from digital elevation models*, *Wildlife Soc Bull*, vol. 32, no. 3, pp. 829–39, 2004.

[54] B. K. P. Horn, *Hill Shading and the Reflectance Map*, *Proceedings of the IEEE*, vol. 69, no. 1, pp. 14–47, 1981.

[55] A. P. Ramallo-González, M. Vellei, M. Brown, and D. A. Coley, *Remote facade surveying of windows characteristics*, *Energy Procedia*, vol. 78, pp. 925–930, 2015.

[56] E. Meier, *Extraktion von Gebäudeeigenschaften aus Google StreetView zur Berechnung von Gebäudewärme*, ETH Zurich, Switzerland, 2016.

[57] Konferenz Kantonaler Energiefachstellen, *Vollzugshilfe Wärmeschutz von Gebäuden*, 2007.

[58] SIA, *SIA 416:2003: Flächen und Volumen von Gebäuden*, no. 1993, 2003.

[59] M. A. Silva-Pérez, J. M. Barea-García, M. Larrañeta, S. Moreno-Tejera, and I. Lillo, *Analysis of the distribution of measured and synthetic DNI databases and its effect on the expected production of a parabolic trough plant*, *Energy Procedia*, vol. 49, pp. 2512–2520, 2014.

[60] J. Trentmann, C. Träger-Chatterjee, and R. Müller, *user manual - meteosat solar surface irradiance and effective cloud albedo climate data records*, 2014.

[61] M. Herrera, S. Natarajan, D. A. Coley, T. Kershaw, A. P. Ramallo-González, M. Eames, D. Fosas, and M. Wood, *A review of current and future weather data for building simulation*, *Building Services Engineering Research and Technology*, vol. 38, no. 5, pp. 602–627, 2017.

[62] J. Duffie and W. Beckman, *Solar engineering of thermal processes*, Wiley-Blackwell, 2013.

[63] C. Demain, M. Journée, and C. Bertrand, *Evaluation of different models to estimate the global solar*

radiation on inclined surfaces, *Renewable Energy*, vol. 50, pp. 710–721, 2013.

[64] R. Perez, R. Seals, P. Ineichen, R. Stewart, and D. Menicucci, *A new simplified version of the perez diffuse irradiance model for tilted surfaces*, *Solar Energy*, vol. 39, no. 3, pp. 221–231, 1987.

[65] Cabinet of Germany, *Verordnung über energiesparenden Wärmeschutz und energiesparende Anlagentechnik bei Gebäuden (Energieeinsparverordnung - EnEV)*, 2007.

[66] G. Bekö, S. Gustavsen, M. Frederiksen, N. C. Bergsøe, B. Kolarik, L. Gunnarsen, J. Toftum, and G. Clausen, *Diurnal and seasonal variation in air exchange rates and interzonal airflows measured by active and passive tracer gas in homes*, *Building and Environment*, vol. 104, pp. 178–187, 2016.

[67] C. Dimitroulopoulou, *Ventilation in European dwellings: A review*, *Building and Environment*, vol. 47, pp. 109–125, 2012.

[68] S. Langer and G. Bekö, *Indoor air quality in the Swedish housing stock and its dependence on building characteristics*, *Building and Environment*, vol. 69, pp. 44–54, 2013.

[69] F. J. Damerau, *A technique for computer detection and correction of spelling errors*, *Communications of the ACM*, vol. 7, no. 3, pp. 171–176, 1964.

[70] J. Munkres, *Algorithms for the Assignment and Transportation Problems*, *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–8, 1957.

[71] Baudirektion Kanton Zürich (Abteilung Energie), *Energiekennzahl Wohnbauten*, 2014.

[72] A. Huser, T. Grieder, and K. Wiederkehr, *Geräteausstattung und Stromverbrauch von Schweizer Haushalten*, *Bulletein SEV/VSE*, vol. 4, pp. 1–8, 2006.

[73] K. U. Ahn, D. W. Kim, C. S. Park, and P. de Wilde, *Predictability of occupant presence and performance gap in building energy simulation*, *Applied Energy*, vol. 208, pp. 1639–1652, 2017.

[74] J. Nipkow, S. Gasser, and E. Bush, *Der typische Haushalt-Stromverbrauch. Energieverbrauch von Haushalten in Ein- und Mehrfahmilienhäusern*, Bulletin SEV/VSE 19/2007, 2007.

[75] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, and U. Wilke, *CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning*, in *Proc. Building Simulation, Eleventh International IBPSA Conference*, 2009.

[76] B. Givoni, *Climate considerations in building and urban design*, Wiley, 1998.

[77] Wikipedia, *National lidar dataset*, 2017. [Online]. Available: http://en.wikipedia.org/w/index.php?title=National lidar dataset&%0Aoldid=784664971. [Accessed: 28-Aug-2017].

[78] Statistisches Bundesamt, *Zensus 2011, Gebäude und Wohnungen*, Bundesrepublik Deutschland, 2013.

[79] Bundesanstalt Statistik Österreich, *Mikrozensus ab 2004 Arbeitskräfte- und Wohnungserhebung*, 2017.

[80] Eurostat, *Klassifikation der Bauwerke*, 1997.

[81] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, *A new operational model for satellite-derived irradiances: Description and validation*, *Solar Energy*, vol. 73, no. 5, pp. 307–317, 2002.

[82] E. Cubi, O. Akbilgic, and J. Bergerson, *An assessment framework to quantify the interaction between the built environment and the electricity grid*, *Applied Energy*, vol. 206, pp. 22–31, 2017.

[83] D. Patteeuw, G. P. Henze, and L. Helsen, *Comparison of load shifting incentives for low-energy buildings with heat pumps to attain grid flexibility benefits*, *Applied Energy*, vol. 167, pp. 80–92, 2016.

[84] R. Buffat, L. Schmid, N. Heeren, A. Froemelt, M. Raubal, and S. Hellweg, *GIS-based Decision Support System for Building Retrofit*, *Energy Procedia*, vol. 122, pp. 403–408, 2017.

[85] N. Heeren and S. Hellweg, *Tracking construction material over space and time: Prospective and geo-referenced modeling of building stocks and construction material flows*. [submitted for publication]

[86] A. Horni, K. Nagel, and K. W. Axhausen, *The Multi-Agent Transport Simulation MATSim*, London, UK: Ubiquity Press, 2016.

CHAPTER 5

USING DATA MINING TO ASSESS ENVIRONMENTAL IMPACTS OF HOUSEHOLD CONSUMPTION BEHAVIORS

Andreas Froemelt*, *David J. Dürrenmatt and Stefanie Hellweg*

Environmental Science & Technology (2018)

* The individual contribution of Andreas Frömelt consisted of conceptualizing the analysis framework, developing and coding the required models, conducting the analyses and preparing the manuscript for publication.

ABSTRACT

Household consumption is a main driver of economy and might be regarded as ultimately responsible for environmental impacts occurring over the life cycle of products and services. Given that purchase decisions are made on household levels and are highly behavior-driven, the derivation of targeted environmental measures requires an understanding of household behavior patterns and the resulting environmental impacts. To provide an appropriate basis in support of effective environmental policymaking, we propose a new approach to capture the variability of lifestyle-induced environmental impacts. Lifestyle-archetypes representing prevailing consumption patterns are derived in a two-tiered clustering that applies a Ward-clustering on top of a preconditioning self-organizing map. The environmental impacts associated with specific archetypical behavior are then assessed in a hybrid life cycle assessment framework. The application of this approach to the Swiss Household Budget Survey reveals a global picture of consumption that is in line with previous studies, but also demonstrates that different archetypes can be found within similar socio-economic household types. The appearance of archetypes diverging from general macro-trends indicates that the proposed approach might be useful for an enhanced understanding of consumption patterns and for the future support of policymakers in devising effective environmental measures targeting specific consumer groups.

5.1 INTRODUCTION

Households are major drivers of the economy. Their consumption behavior triggers a multitude of economic activities along the supply chain of each product and service, which subsequently involves the use of resources and the release of emissions. Household consumption is estimated to be responsible for 65% of global greenhouse gas emissions and 50% to 80% of total land, material, and water use [1]. The United Nation's Sustainable Development Goal 12 ("Sustainable Consumption and Production") [2] demonstrates a large consensus that today's consumption patterns are unsustainable and changes in consumer behavior are urgently needed [3–11]. However, changing household consumption behavior is challenging [4, 6, 12], as it is embedded in complex economic, social, technological and cultural systems. In addition to informing households about their environmental impacts, policymakers should frame an enabling environment for individuals to change toward more sustainable lifestyles [4, 6, 8, 12].

Several studies quantified environmental impacts induced by household consumption (see e.g., [7, 8, 13–15] for reviews). While many studies focus on a national average household and on identifying environmental priorities of different consumption areas [1, 3, 7, 13, 14, 16, 17], several researchers acknowledge the importance of investigating the environmental consequences of different household groups (e.g., [5, 7, 10, 11, 18–25]). Being highly influenced by socio-cultural and economic factors, as well as driven by individual preferences, household behavior and consequential purchase decisions are diverse, and "one-size-fits-all"-recommendations are likely to fail

[5, 11, 26]. Therefore, understanding the variability of consumption patterns and associated environmental impacts is required for devising targeted environmental policies.

Studies attempting to capture this lifestyle-induced variability usually combine household budget surveys (also called consumer expenditure surveys) with environmentally-extended input-output models (EEIOMs) and then assess the environmental impacts of different socio-economic cohorts [3, 5, 9–11, 17, 20, 22, 23, 26], or fit regression models with socio-economic characteristics as explanatory variables [24, 27]. The findings of both (sometimes combined) approaches are very insightful, especially if the applied regression models aim at explaining the drivers of environmental consequences [10, 16, 20, 22, 23, 25, 26]. However, Girod and De Haan [19] found that there might still be significant variability of behavior within investigated household types. This raises the question if the use of household segments that are pre-defined solely based on socio-economic characteristics might prevent recognition of important behavioral patterns by assuming all households within a segment behave similarly. How could the support for policymaking be improved, especially in view of recent calls to increase the involvement of behavioral economics and psychology when deriving environmental measures [4, 12, 28]? Building upon the mentioned lifestyle-studies, we propose a new approach: Instead of pre-defined household segments, we suggest deriving clusters of households which are not only based on socio-economic aspects but also on real observed behavior. The proposed two-stage clustering allows for studying behavior-associated environmental impacts in the context of total consumption and is simultaneously able to capture non-linear effects. This approach thus allows for investigating the nature and implications of different household behaviors in detail. The emergence of archetypes might then form a new information basis to derive environmental policies tailored to actual consumption patterns.

The goal of this chapter is twofold: First, we will demonstrate the clustering of household behavior by applying our approach to the Swiss Household Budget Survey (HBS). Second, this application will result in ready-to-use consumption archetypes with associated environmental modeling for Switzerland. While the transferability of the Swiss archetypes to other countries is unclear, the proposed methods can definitely be applied to different expenditure surveys.

5.2 METHODOLOGY

Grouping households based on their characteristics and on their consumption behavior represents the core of our approach to studying household environmental impacts. Utilizing groups of households is also important because of the so-called “infrequency of purchase problem” [22, 25] that is encountered when working with expenditure surveys in which households participate only during a limited time period. To obtain a representative picture of a certain population group’s consumption behavior, in which infrequent or season-specific purchases average out, several similar households need to be clustered. In previous lifestyle-studies, this is implicitly solved by averaging over pre-defined household segments. In our approach (Figure 5.1), we first form groups of similarly behaving households and then in a second step derive archetypes representing the average behavior of these groups. Finally, the environmental impacts of these archetypes will be assessed by means of a hybrid life cycle assessment (LCA) that sources environmental background data from both EEIOMs and process-based life cycle inventory databases. With a hybrid LCA framework, known issues with EEIOMs (e.g., the "product quality problem" [7, 9, 18, 20, 22, 25]) can be partly overcome by using physical functional units, and the EEIOMs can complement missing data of process-based LCA [7]. Furthermore, the applied LCA-modeling will also allow for the computation of different impact assessment methods and not only carbon footprints or energy requirements as done in previous studies [3, 5, 9, 10, 13, 16–20, 22, 24, 29]. Thus, the chosen LCA-approach might also help to reveal potential burden shifts induced by planned environmental measures. Although suggested by several authors [1, 3, 13, 15], the application of hybrid LCA is, to our knowledge, rare. That said, it would still be possible to directly couple the archetypes with EEIOMs or even more sophisticated macro-economic models as used, for example, in [21, 30].

All computation steps will be described in more detail below (see also Figure 5.1).

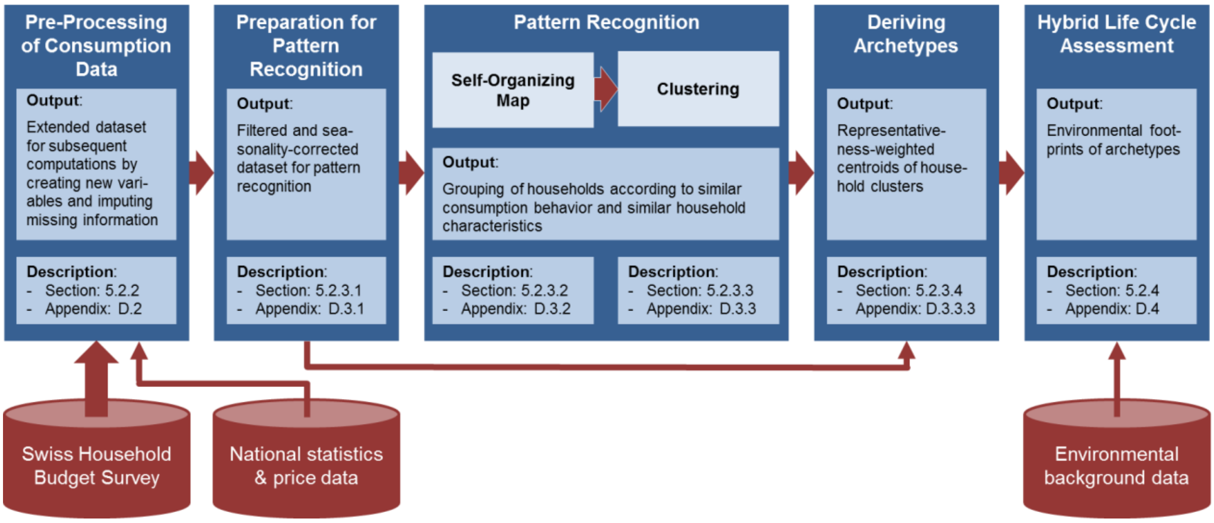


Figure 5.1: **Simplified flow scheme providing a synopsis of the whole modeling approach. For each step, sections of Chapter 5 and Appendix D are indicated in which more detailed descriptions can be found.**

5.2.1 Consumption Data

The main data source of this study is the Swiss HBS [31] (2009-2011) which provides detailed information on the characteristics and consumption behavior for 9734 households. Households participating in the survey report on daily expenditures, income, and quantities of bought goods (e.g., for food) during one month. In addition, they also report on periodic expenditures (e.g., newspaper subscriptions), possession of durable goods and on extraordinary purchases or revenues in the last few months (e.g., buying a car within the last year). Data on household characteristics and on household members are also collected. For each household, the final data set comprises statistics on 20 different durable goods, 8 income categories (plus 4 on aggregated levels), 19 household variables, 6 attributes for each household member and 356 (plus 175 on aggregated levels) consumption categories classified based on the United Nations' "Classification of Individual Consumption according to Purpose" (COICOP) [32]. For consumption areas, purchased amounts in liters or kilograms are available for 92 categories (plus 14 on aggregated levels). Further information on the categorization and a list of survey attributes is provided in Appendix D.

5.2.2 Pre-Processing of Consumption Data

The present section describes how variables needed for the computations in sections 5.2.3 (pattern recognition) and 5.2.4 (LCA) were created based on the original HBS-data.

Besides converting categorical variables to numerical data by a set of binary variables, several count-statistics were created to better compare household member information (see Appendix D). Moreover, for some categories, more detailed information than specified in the HBS-data is required to apply LCA (cf. section 5.2.4). This particularly concerns housing-related categories and public transport demand. With regard to the first, especially tenants typically do not have full information on the exact breakdown of their utility bills into refuse and sewage collection, water supply, electricity, and heating. To impute this missing information, a modeling approach using K-Nearest-Neighbor-Regression [33], Random-Forest-Regression [34, 35] and LASSO-Regression [36] was employed (see Appendix D). The predicted data was then converted to quantities by means of price data [37–40]. These prices were retrieved as close as possible to the specific circumstances of each household by taking into account household type, location, and survey year. The pre-processed HBS-data was then validated against national statistics [40–44] (cf. Appendix D).

The demand for public transport is a second issue. While kilometers driven by car can be estimated with liters of fuel purchased, HBS-information on public transport mainly relates to season tickets and travel card expenditures, thereby lacking information on effectively driven kilometers by public transport. Therefore, this demand was estimated for each household using data from the Swiss Mobility Microcensus 2010 [45], which provides detailed information on the mobility behavior of the Swiss population (see Appendix D).

5.2.3 Pattern Recognition and Clustering of Households

5.2.3.1 Preparation for Pattern Recognition

Because of seasonality and storage effects, the survey month might bias the results of individual households in the HBS [31]. Consequently, the determination of household clusters with similar characteristics and behavior requires the data set to be pre-processed and filtered for household features which make similar behavior identifiable independent of the month in which the survey took place.

A flow scheme in Appendix-section D.3.1 visualizes the following preparatory step. Each HBS-attribute was judged separately whether it shall be included for pattern recognition or not. Note that exclusion only concerns the pattern recognition steps in sections 5.2.3.2 (*Self-Organizing Map (SOM))* and 5.2.3.3 (*Clustering*), while the deduction of archetypes in section 5.2.3.4 (*Deriving Consumption Archetypes*) will resort to all available attributes. Household characteristics including the count-statistics of section 5.2.2, durable goods statistics, periodic expenditures, and revenues were all considered in the pattern recognition. In contrast, daily purchase attributes were only included if they are bought on a regular basis and not stored for more than a month. Otherwise, it was appraised if the attribute's inclusion within an aggregated level is reasonable (e.g., if an aggregated attribute such as "fruits" shall rather be included than "grapes" which are rarely bought from January to June in Switzerland).

In a next step, all candidate attributes were checked for seasonality. The application of ANOVA [46] and the Kruskal-Wallis-tests [47] revealed if the influence of the survey month is statistically significant. In the case of statistical significance, the respective attribute was corrected for seasonality (see Appendix-section D.3.1).

The final data set for pattern recognition comprises 157 attributes in total, thereof 85 consumption and income categories.

5.2.3.2 *Self-Organizing Map (SOM)*

A two-tiered approach was applied to find patterns in the HBS-data based on which the households shall be grouped. In a first stage, a self-organizing map (SOM) [48, 49] pre-conditioned and reduced the data set, which was then clustered in a second stage. Such two-step clusterings demonstrated to perform well and are robust even in the case of noisy and high-dimensional data sets [50–53].

The SOM was proposed by Kohonen [48, 49] and belongs to the class of unsupervised artificial neural networks. It generally combines vector quantization and non-linear projection to a lower-dimensional space; usually to a discrete 2D-lattice of neurons. Thereby, the SOM learns the patterns in the data set in an ordered fashion and is thus able to preserve the topology of the data. This means that neighboring neurons in the map have similar characteristics. A prototype vector with the same dimension as the vectors of the input data set is associated with each neuron. During the training phase of the SOM, this set of prototype vectors is optimized to become a representative substitute for the original data set. The resulting map is thus a reduced, but still representative, data set that is not only smoothed with regard to noise but also facilitates the recognition of patterns, be it for subsequent clustering algorithms or visually for the human eye (see component maps in Appendix D).

For the present study, the SOM was tuned by generating several SOMs with different parameters (number of neurons, arrangement of neurons, initial and final neighborhood radius, neighborhood function, and number of epochs) based on literature recommendations [48, 49, 54–56] and then by choosing the model with a topographic error close to zero and the lowest quantization error [56]. The topographic error evaluates the order of the map (e.g., if it is twisted) and represents the share of samples in the input data set for which the first and second closest neurons are not adjacent in the map [56], while the quantization error judges the representativeness and thus the accuracy of the map [56]. The final map consists of 987 neurons arranged in a 21:47-planar-lattice (see Appendix D for more tuning information and a short introduction to SOMs).

5.2.3.3 *Clustering*

In the second stage of pattern recognition, we apply clustering algorithms on top of the SOM to form groups of neurons. Since each clustering technique has its strengths and weaknesses, two well-known approaches, which differ in their basic principles, were tested and evaluated: K-Means [57–59] and agglomerative clustering [60, 61]. While *k*, the number of clusters to be built, was the only tuning parameter for K-Means, agglomerative clustering was run in different combinations of affinity metrics (e.g., Euclidean distance, L1-norm) and linkage criteria (Ward [62] and average).

The evaluation of clusters – and thus the "best" choice of clustering techniques and associated parameters – is not a trivial task if the ground truth is unknown [63]. For the present study, we mainly focused on two performance methods: Silhouette coefficients (*S*) [64] and U-Matrix. *S* relates the distances between one point in a cluster and all other points in the same cluster to the distances from that point to all points in the second closest cluster. The so-called U-Matrix (unified distance matrix) is an important visualization of a SOM and supports clustering on top of the map. At the map position of each neuron, an U-Matrix depicts the sum of distances in the high-dimensional space between the prototype vector of the respective neuron and the prototype vectors of all adjacent neurons [52, 65, 66]. Large U-heights indicate that a neuron's prototype vector is distant from others, while small U-heights are associated with prototypes that are surrounded by other vectors in the data space [66]. The U-Matrix thus suggests visually which neurons should be grouped together to form clusters.

The quest for the best clustering was subdivided into two parts (see Appendix D for details): First, a pre-selection of parameters within each approach was conducted mainly based on *S*. Afterward, a detailed comparison of the two alternatives and fixing the number of clusters for both was based on the U-Matrix. For this, the cluster borders were projected on the U-Matrix and the clustering methods were judged by their ability to redraw the visible groupings of neurons in the U-Matrix. Additionally, two other criteria broadened the information basis for the final decision: ANOVA-tests were run for each attribute to obtain an impression if reasonable results are produced. Second, the number of households per cluster was determined to get some indication about the representativeness of the clusters. Considering all these criteria, an agglomerative clustering technique with Ward-linkage, which produces 34 clusters, was finally selected. Furthermore, the applied agglomerative clustering implementation [67] allows for including connectivity constraints, meaning that only clusters which are adjacent on the map can be merged by the algorithm. This ability was seen as another advantage over K-Means.

5.2.3.4 Deriving Consumption Archetypes

Considering the statistical basis of the analyses of the Federal Statistical Office [31], we assumed a cluster to be representative for a population's group if at least 130 HBS-households are member of this cluster. Since not all clusters built in section 5.2.3.3 (*Clustering*) comply with this criterion, some post-processing in the sense of manually merging clusters was needed. Indeed, this merging is also justified by the dendrogram in Appendix D which shows a blur between 34 and 24 clusters, indicating that some clusters are close to each other. Cutting the dendrogram at different positions might be reasonable in the presence of sub-clusters [51] (see Appendix D for more reasons). Therefore, starting with 34 clusters, all clusters with less than 130 households were merged with adjacent clusters if these merges happen in the dendrogram between 24 and 34 clusters. This resulted in 26 fair clusters and 2 clusters with less than 130 households. These two clusters will continue to be part of the subsequent analysis, but they will be marked accordingly.

The behavior-archetypes are now constituted by the centroids of the clusters. Note that the computation of the clusters' centroids follows the averaging-procedure of the Federal Statistical Office and includes a representativeness-weight [31]. Furthermore, the attributes which were filtered out in section 5.2.3.1 (*Preparation for Pattern Recognition*) were now reused and also entered the vector of the archetypes.

5.2.4 LCA-Modeling

The final modeling step comprised the coupling of the archetypes' demands with detailed life cycle background data in order to quantify the environmental impacts of the archetypes' consumption behavior. The overall functional unit for the LCA was chosen to be one year of household consumption. The life cycle inventory data were extracted from three well-known and transparent databases in the following priority order: ecoinvent v3.3 [68], Agribalyse v1.2 [69], and EXIOBASE v2.2 [70, 71].

The functional units of ecoinvent- and Agribalyse-activities require quantities instead of expenditures. While section 5.2.2 prepared housing- and public-transport-related categories for this purpose, conversions based on price data (e.g., [37–39, 72]) or further information [73] were necessary for other consumption areas. Fortunately, for almost all food categories, quantities are directly available in the HBS-data set. The process models for processed food closely followed the modeling of Walker et al. [74], but adjusted to Swiss conditions. Generally, we always attempted to adapt the process models as close as possible to the domestic conditions of Switzerland. For instance, Swiss market activities for food were constructed based on the import statistics provided by Scherer and Pfister [75] and car fleets and energy mixes for heating technologies were based on national statistics [40, 76].

The creation of process models with EXIOBASE-sectors including the conversion of purchaser-prices (HBS) to basic-prices (EXIOBASE) generally followed the suggestions of Steen-Olsen et al. [3]

Finally, it needs mentioning that only environmental impacts directly associated with a certain household's behavior were considered. For instance, only direct spending on education or health care and thus retraceable to a specific household were taken into account. This leads to an underestimation of impacts from "health" and "education" since the Swiss education system is largely financed by the government, while health expenses are usually covered by insurances whose premiums are not necessarily indicative of the health care utilization of households and thus of the effectively caused impacts. Furthermore, governmental consumption was not re-distributed to households as done in other studies [77, 78]. Details of the LCA-modeling are disclosed in Appendix D.

5.3 Results and Discussion

5.3.1 Drivers of Clustering

The U-Matrix and the final clustering are presented in Figure 5.2. The two clusters not reaching full representativeness are named "OA" and "OB", while all other clusters are labeled with random letters. The illustration of clusters on top of the SOM together with the distance information given in the U-Matrix also reveals relationships between clusters. For instance, "N" will be more similar to the adjacent "O" than to "J".

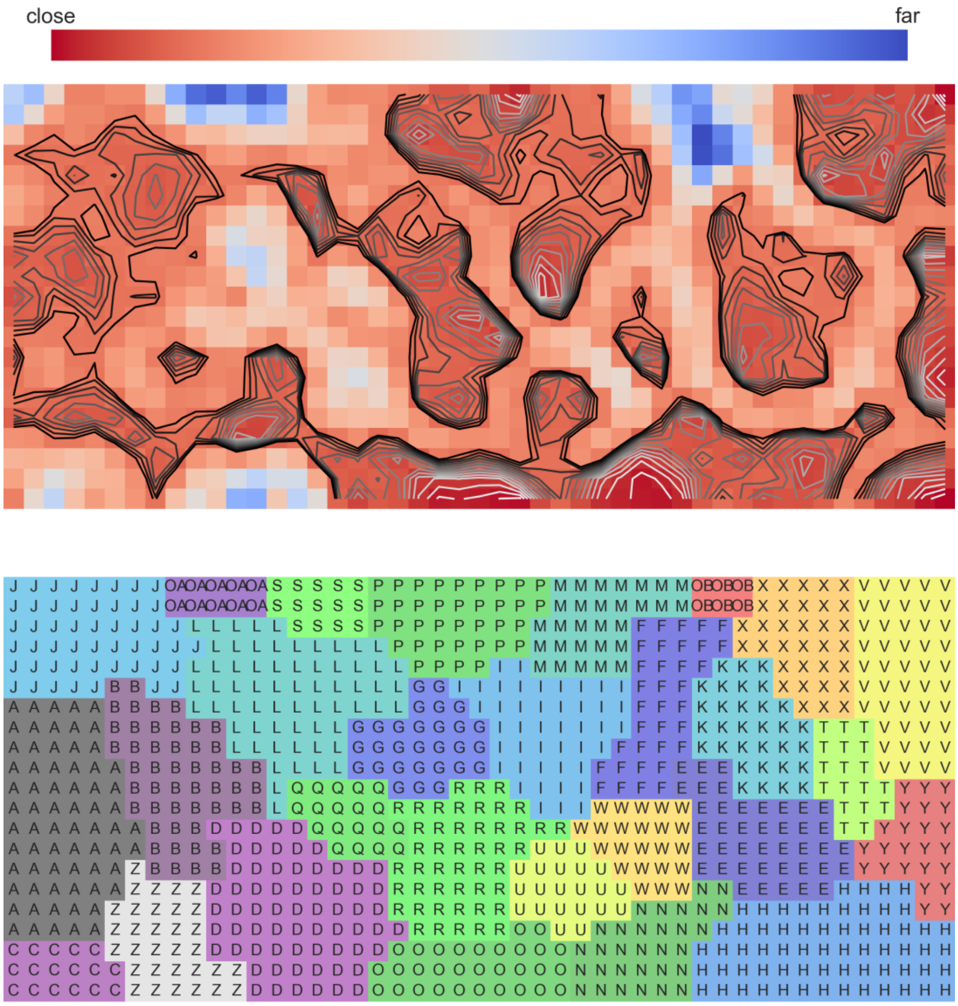


Figure 5.2: **(Top) U-Matrix of the SOM. Contours were inserted to improve visibility of distances. Note that a pixel in the map corresponds to the map position of a neuron. (Bottom) Final clustering on top of the SOM. The clusters "OA" and "OB" are the clusters with less than 130 households.**

Building upon the clustering selection procedure in 5.2.3.3 (*Clustering*), the quality of the final clustering was further analyzed by a visual appraisal of the 95%-confidence intervals of the clusters' means of each attribute and by the ANOVA-tests (see section 5.2.3.3 (*Clustering*) and Appendix D). These tests indicated that for all attributes there is at least one cluster significantly differing from the others (largest p-value: 2.45 x 10^{-14}). ANOVA and the qualitative visual judgment support the reasonability of the final clustering. In addition, ANOVA provides a possibility to analyze the underlying drivers of the clustering process. An attribute's test statistics (F-values) can be considered as a measure of "distinction power". Among the top 20 distinctive attributes, there are, in particular, geographic variables and employment status, but also variables related to household size, consumption areas, and durable goods statistics. Attributes which are rated to be least important for forming the clusters are either variables which concern only few households (e.g., university fees) or consumption areas for which obviously less distinct patterns could be found (e.g., consumption of "coffee, tea and cocoa"). All results can be found in Appendix D.

5.3.2 Individual Archetypes and Their Interrelations

A simplified description of all archetypes is available in Table 5.1. In this table, the clusters' centroids are categorized according to income and average number of persons per household. However, note that the clusters emerged from many different – including also non-socio-economic – attributes. The attributes presented in Table 5.1 and those used in the following analysis are thus just of an indicative nature meant to help better grasp the archetypes (see also additional details and results in Appendix D).

The environmental impacts of the individual archetypes are presented in Figure 5.3. Even though the applied LCA-modeling allows for the computation of all environmental indicators that are supported by the background life cycle inventory databases, only greenhouse gas emissions (GHG) according to IPCC 2013 (100a) [79] and total endpoints of ReCiPe 2008 (H,A) [80] are discussed in the following (see Appendix D for results of the Ecological Scarcity method [81]). Besides allowing for comparison to other studies, GHG have the highest priority in political discussions due to climate change, while the ReCiPe-endpoints shall help to calculate the overall environmental relevance of household consumption.

In the scope of this chapter, only a few comparisons between some of the archetypes are presented. Thereby, selected examples in the aspects of income, household size, age structure, and footprint composition will be discussed. To support these comparisons, Figure 5.3 resumes information from Table 5.1 and shows some general characteristics of the archetypes such as income, number of persons, and average age of adults and children.

In the upper part of Figure 5.3, a correlation between household income and total impacts can be observed. However, a few high-income households (e.g., C ("family with babies"), D ("young, unmarried couples"), and Q ("divorced, middle-aged males")) are an exception to this. In addition to income, household size is crucial for total emissions. This becomes apparent in the per-capita-bar plots in the lower part of Figure 5.3 when we observe how the order of archetypes changes. The archetypes with the largest incomes (OA ("very high-income family"), OB ("very high-income retired couple"), S ("small families with self-employed persons")) still spearhead the per-capita figure. But OB outstrips OA, and except for OA and S, most other family archetypes (e.g., A ("family with primary school children"), B ("single-parent-families"), and Z ("young adults with babies")), including some with high-income (e.g., J ("family with teenagers") or C), can be found on the bottom of the per-capita-impact-ranking. In contrast, single-person households (e.g., H, N, O, and R) show low total annual impacts, while in a per-capita perspective these archetypes move to the middle field. H ("old, widowed females"), together with V ("low-income, retired couple") and Y ("low-income, very old couple"), also reveal a certain generation gap with regard to footprint composition: all three of these clusters – which are also neighbors in the SOM – are relatively older in age, have low communications and mobility impacts, and have higher housing energy and health impacts compared to other archetypes. This is for instance in contrast with the "young" D archetype, which exhibits high transport, but low housing, impacts.

Table 5.1: Simplified categorization of archetypes (clusters' centroids) along the axes of income (vertical) and number of persons per household (horizontal)[a].

	Single-person households (1.1 - 1.3)	Small households (1.5 - 1.7)	Two-persons households	Small families (2.4 - 3.2)	Families (3.4 - 3.6)	Large families (>4)
Very high income (>19000)			**OB** (around retirement)	**S** (self-employed; some with primary school kids/ teenagers)	**OA** (homeowners with primary school kids/ teenagers)	
High income (13000 - 19000)				**L** (with primary school kids/ teenagers)	**C** (rather young adults with babies)	**J** (homeowners with teenagers; "technophile")
Medium-high income (9500 - 13000)		**Q** (divorced middle-aged males)	**D** (young, unmarried; w/o children)	**K** (Swiss-Italian; with primary school kids/ teenagers; homeowners)	**F** (Swiss-French; with teenagers; tenants)	**A** (homeowners with primary school kids)
			M (Swiss-French; around retirement)			
			P (Swiss-German; around retirement)			
			X (Swiss-German; Zurich; around retirement, older than P)			
Medium income (7800 - 9500)		**G** (Northwestern Switzerland; middle-aged adults w/o children)			**Z** (very young adults with babies)	
		U (Central Switzerland, well-mixed-not-retired adults with children)			**B** (tenants with primary school kids/ teenagers; depend on other households; by trend: 1 adult and 2 children)	
Medium-low income (5000 - 7800)	**O** (Swiss-German; Zurich; well-mixed-not-retired adults)	**I** (Swiss-French; well-mixed-not-retired adults w/o children)	**Y** (very old people)			
	R (Eastern Switzerland; well-mixed-not-retired adults)	**W** (Swiss-French; slightly older middle-aged adults w/o children)	**T** (Swiss-Italian; retired)			
	N (Swiss-German; Bern; well-mixed-not-retired adults)	**E** (Swiss-Italian; middle-aged adults w/o children)	**V** (Swiss-German; retired)			
Low income (< 5000)	**H** (old, widowed females)					

[a]The numbers in parentheses show income in Swiss Francs per month and average numbers of persons per household respectively.

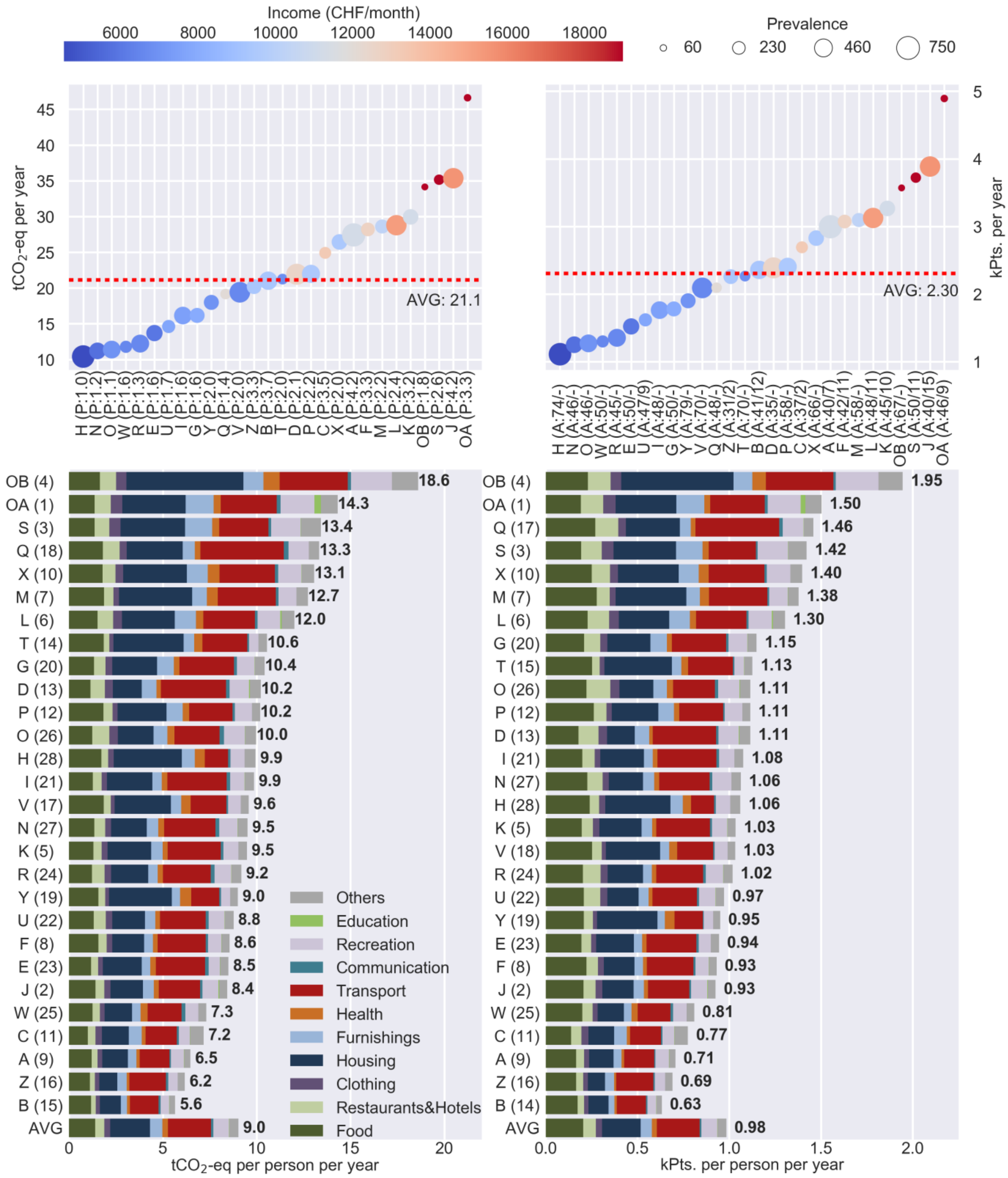


Figure 5.3: (Top) Total annual impacts per archetype; colors illustrate income, while size of the markers represents prevalence of the archetypes (number of households per cluster). The red dashed line depicts the prevalence-weighted average (AVG). Indicative archetype characteristics are given in parentheses: P stands for average number of persons per archetype, while A provides the average age of adults/children. (Bottom) Bar plots showing the composition of the environmental impacts on a per-capita basis. GHG values (IPCC 2013, 100a) are shown in the left part of the figure and results of ReCiPe 2008 (total endpoints, H,A) are on the right-hand side. Ranks of total footprints are given in parentheses to facilitate the comparison of top and bottom figures.

Interestingly, there are also archetypes with similar socio-economic conditions but different total emissions and footprint compositions. O and R are adjacent in the SOM and both are single-person households with similar age and income range. However, archetype O is related to urban households (Zurich), while R has its origin in the more rural Eastern Switzerland. R's mobility impacts are higher on account of larger emissions from car-driven distances. This is, however, partly compensated by O's large demand for air travel and taxi rides. Furthermore, O has larger impacts induced by eating out.

Finally, the comparison of J and F reveals another interesting aspect. Both archetypes are similar in regard to size, income, age ("families with teenagers"), per-capita footprints, and compositions. However, for devising targeted measures to abate housing-induced emissions, there is one especially important difference between the two clusters: J households are homeowners, whereas F households are tenants.

5.3.3 Archetypes in the Collective and Environmental Hotspots

The analysis of the archetypes can help policymakers in identifying target groups. For instance, OA and OB have large footprints and have money available to invest in environmentally beneficial technologies to cover, for example, their transport and housing demand. However, even though the signaling effect to target these groups may be large, they are not very prevalent. Consequently, it is also important to look at the overall picture and investigate how the different archetypes contribute to cumulative impacts; which is in the case of OA and OB together less than 2.5%. Since HBS-data can be assumed to be representative for Switzerland, we used the number of households per cluster as an indication of prevalence. This information is also illustrated by the size of the markers in the scatterplots of Figure 5.3. The upper part of Figure 5.4 depicts the prevalence-weighted contributions of both the archetypes and the consumption areas. The lower part of Figure 5.4 then relates the archetypes' prevalence-weighted total impact shares to their per-capita-emissions in a simplified attempt to group the archetypes based on the prospects to target them. It becomes apparent that the three family archetypes A, J, and L, which are ranked first, fifth, and sixth by frequency, are together responsible for about 27% of total impacts. While the per-capita-emissions are rather low for types A and J, their high contribution to total impacts is due to their size (>4 persons per household on average) and their sheer abundance. In contrast, type L is not only frequent, but also causes large per-capita and total footprints. These small, high-income and home-owning families with teenagers could be an interesting target group for reducing consumption impacts. However, this is just one example since type L's contribution is about 8% of total emissions and further archetypes need to be targeted as well. An analysis of the main consumption areas in more detail shows that V ("low-income, retired couple") and D ("young couple") appear to be important. Being among the three archetypes with largest per-capita-emissions for air travel and car trips, type D has particularly large transport impacts and is also responsible for high emissions in the category "restaurants and hotels". In contrast, type V shows high health and housing impacts in addition to large per-capita food emissions. However, while the lower part of Figure 5.4 classifies D as a "promising target", V seems to be a borderline case which requires more careful consideration.

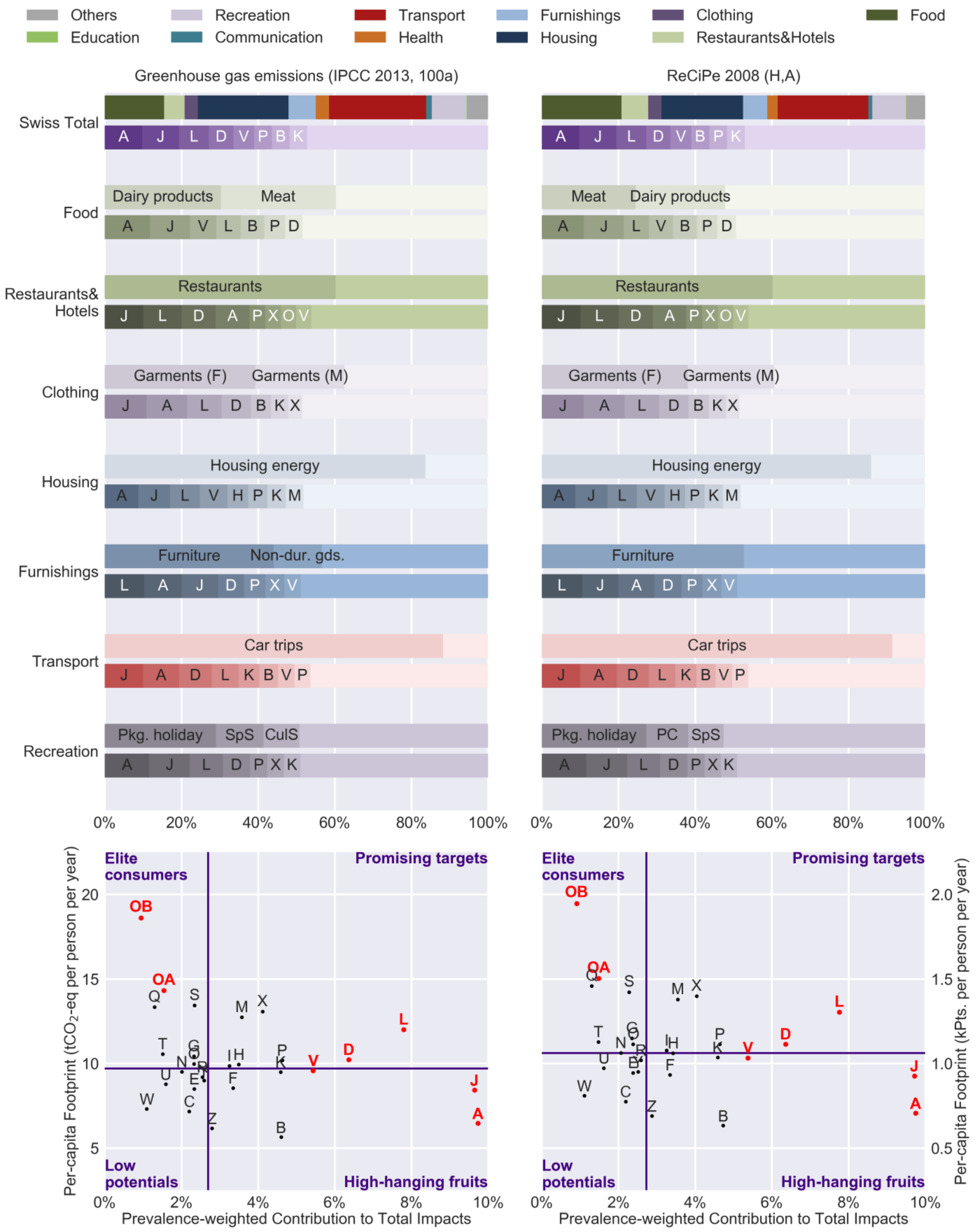


Figure 5.4: (Top) Prevalence-weighted contributions of the archetypes and main consumption areas to total environmental impacts as well as prevalence-weighted contributions of archetypes and consumption-subcategories within the main consumption areas (ordered by magnitude of contribution). Except for the Swiss total, only contributions with a cumulative effect of about 50% are displayed. (Bottom) Per-capita-emissions in relation to prevalence-weighted contributions to total impacts. The divisive horizontal and vertical lines correspond to the medians of the *x*- and *y*-values, respectively. Archetypes mentioned in the text are highlighted in red. [Non-dur gds.: Non-durable household goods; Pkg. holiday: Package holiday; SpS: Sporting services; CulS: Cultural services; PC: Computer, office appliances and other peripherals]

The second perspective provided in Figure 5.4 (upper) identifies, just as in many previous studies [1, 3, 7, 14–16], food, housing, and transport as the most important consumption categories. While transport and housing each account for about 25% of total GHG emissions, transport takes the lead in ReCiPe-endpoints with 24%, closely followed by food and housing (each 21%). Thereby, transport is clearly dominated by car trips (about 90%) and housing by energy use (about 85%). For food, dairy products (especially semi-hard/hard cheese) and meat (mainly beef) are of similar importance; each with shares around 30% in the GHG-perspective and about 25% with regard to ReCiPe-endpoints.

The computed prevalence-weighted average for Switzerland totals at 9.0 t CO_2-eq/person/year. This is between the top-down study of Jungbluth et al. [77] with 11.0 t CO_2-eq/person/year and 8.6 t CO_2-eq/person/year found by Girod and De Haan [18] in another HBS-based study. Note that both studies refer to a prior time period and that the original average of [77] (12.8 t CO_2-eq/person/year) was adjusted to account for our study's underestimation of health care and educational services, and for not redistributing governmental consumption.

5.3.4 Limitations of the Study

The presented archetypes were formed based on behavior and characteristics of households. Thereby, the goal of our approach was to provide the most generalized basis for further investigations. Yet, within these post-analyses – such as the application of LCA with a specific environmental indicator – some clusters become seemingly very similar. Depending on the policy goals, it may thus be reasonable to further group the archetypes according to their environmental impacts.

Another limitation of this study pertains to the underreporting in the HBS. Although participating households are closely supervised and receive advice from specialists [82], previous studies that have made use of consumer expenditure surveys revealed that underreporting is a common issue [3, 22, 29]. This could also explain the slightly lower total carbon footprint assessed here in comparison to the economy-wide study of Jungbluth et al. [77].

Finally, the applied hybrid LCA-modeling was done with great care and adjusted as much as possible to Swiss conditions. Still, LCA-modeling always requires assumptions, average mixes, and simplifications that might affect the final results. Hereby, the uncertainties induced by the conversion of expenditures to functional units via price data, the uncertainties arising from converting HBS-purchaser-prices to EXIOBASE-basic-prices [3, 9, 10, 29], and the limited number of biosphere flows in EXIOBASE need special mentioning. The latter is discussed in Appendix D and may lead to a slight under-accounting of ReCiPe-endpoints. Follow-up research aims at developing a framework to capture uncertainties involved in the archetype-approach.

5.4 Outlook

The proposed archetype-approach is meant to deliver important insights into households' consumption behavior for policymakers to derive and prioritize targeted measures. In this context, our approach can also be used for identifying drivers of environmental impacts as done in previous studies [10, 16, 20, 22, 25, 26]. To demonstrate this application, a univariate correlation analysis between environmental impacts and main household characteristics is presented in Appendix D. The computation of such correlations, as well as studying the drivers for clustering and the prevalence-weighted analyses, help to capture the big picture and support thus the identification of general tendencies, hotspots, and potential target groups of households. However, since the overall trends do not necessarily apply to particular archetypes, the deduction of targeted actions requires an in-depth understanding of the target groups and should thus be done on the basis of individual archetypes. For this, the archetype-approach offers a predestined basis by allowing for backtracking environmental impacts to the living conditions of real households and observable consumption behavior.

While this study aimed to provide the basis for identifying strategies, the specific analyses need to be done by local environmental policymakers according to their political agendas. Thereby, they could follow the comprehensive framework proposed by Schanes et al. [4] and also consider suggestions from related research in behavioral economics and psychology [28]. The input from these disciplines, which aim to understand motivational factors and cognitive biases, could provide support to go even beyond "conventional" measures, such as taxes, regulations, or subsidies, and allow for profiling households. On this basis, personalized messages and measures could be formulated which directly address different consumer groups in order to effectively raise awareness and to encourage them to change toward more sustainable consumption patterns. But before investigating how people could be motivated to lower their environmental impacts, it is of high importance to know which archetypes prevail in the local policymakers' sphere of influence. The prevalence-weighted analyses illustrates a first attempt in this direction at a national level, but the archetypes can further be used as a basis for a regionalized model. In follow-up research, the archetypes will be assigned to real households based on the national census within a probabilistic-classification approach. In addition, this use of archetypes will simultaneously be combined with other nationwide bottom-up models such as the building energy model of Buffat et al. [83] (see Chapter 4) and agent-based mobility models as used by Saner et al. [84]. The final model will estimate a realistic environmental profile for each real household in Switzerland and thus provide information for analyses on any desired regional scale.

5.5 REFERENCES

[1] D. Ivanova, K. Stadler, K. Steen-Olsen, R. Wood, G. Vita, A. Tukker, and E. G. Hertwich, *Environmental Impact Assessment of Household Consumption*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[2] Secretary-General (United Nations), *Progress towards the Sustainable Development Goals*, E/2017/66; United Nations: New York, 2017.

[3] K. Steen-Olsen, R. Wood, and E. G. Hertwich, *The Carbon Footprint of Norwegian Household Consumption 1999-2012*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[4] K. Schanes, S. Giljum, and E. Hertwich, *Low carbon lifestyles: A framework to structure consumption strategies and options to reduce carbon footprints*, *Journal of Cleaner Production*, vol. 139, pp. 1033–1043, 2016.

[5] C. M. Jones and D. M. Kammen, *Quantifying carbon footprint reduction opportunities for U.S. households and communities*, *Environ. Sci. Technol.*, vol. 45, no. 9, pp. 4088–4095, 2011.

[6] J. H. Spangenberg and S. Lorek, *Environmentally sustainable household consumption: From aggregate environmental pressures to priority fields of action*, *Ecological Economics*, vol. 43, no. 2–3, pp. 127–140, 2002.

[7] E. G. Hertwich, *Life Cycle Approaches to Sustainable Consumption: A Critical Review*, *Environ. Sci. Technol.*, vol. 39, no. 13, pp. 4673–4684, 2005.

[8] A. Tukker, M. J. Cohen, K. Hubacek, and O. Mont, *The Impacts of Household Consumption and Options for Change*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 13–30, 2010.

[9] A. Druckman and T. Jackson, *The bare necessities: How much household carbon do we really need?*, *Ecological Economics*, vol. 69, no. 9, pp. 1794–1804, 2010.

[10] G. Baiocchi, J. Minx, and K. Hubacek, *The Impact of Social Factors and Consumer Behavior on Carbon Dioxide Emissions in the United Kingdom. A Regression Based on Input-Output and Geodemographic Consumer Segmentation Data*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 50–72, 2010.

[11] J. C. Minx, T. Wiedmann, R. Wood, G. P. Peters, M. Lenzen, A. Owen, K. Scott, J. Barrett, K. Hubacek, G. Baiocchi, A. Paul, E. Dawkins, J. Briggs, D. Guan, S. Suh, and F. Ackerman, *Input-output analysis and carbon footprinting: An overview of applications*, *Economic Systems Research*, vol. 21, no. 3, pp. 187–216, 2009.

[12] D. O'Rourke and N. Lollo, *Transforming Consumption: From Decoupling, to Behavior Change, to System Changes for Sustainable Consumption*, *Annual Review of Environment and Resources*, vol. 40, pp. 233–259, 2015.

[13] E. G. Hertwich, *The Life Cycle Environmental Impacts of Consumption*, *Economic Systems Research*, vol. 23, no. 1, pp. 27–47, 2011.

[14] A. Tukker and B. Jansen, *Environmental Impacts of Products: A Detailed Review of Studies*, *Journal of Industrial Ecology*, vol. 10, no. 3, pp. 159–182, 2006.

[15] M. Di Donato, P. L. Lomas, and Ó. Carpintero, *Metabolism and Environmental Impacts of Household Consumption: A Review on the Assessment, Methodology, and Drivers*, *Journal of Industrial Ecology*, vol. 19, no. 5, pp. 904–916, 2015.

[16] E. G. Hertwich and G. P. Peters, *Carbon Footprint of Nations: A Global, Trade-Linked Analysis*, *Environ. Sci. Technol.*, vol. 43, no. 16, pp. 6414–6420, 2009.

[17] S. Bin and H. Dowlatabadi, *Consumer lifestyle approach to US energy use and the related CO2 emissions*, *Energy Policy*, vol. 33, no. 2, pp. 197–208, 2005.

[18] B. Girod and P. De Haan, *More or Better? A Model for Changes in Household Greenhouse Gas Emissions due to Higher Income*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 31–49, 2010.

[19] B. Girod and P. De Haan, *GHG reduction potential of changes in consumption patterns and higher quality levels: Evidence from Swiss household consumption survey*, *Energy Policy*, vol. 37, no. 12, pp. 5650–5661, 2009.

[20] M. Wier, M. Lenzen, J. Munksgaard, and S. Smed, *Effects of household consumption patterns on CO2 requirements*, *Economic Systems Research*, vol. 13, no. 3, pp. 259–274, 2001.

[21] R. Duarte, K. Feng, K. Hubacek, J. Sánchez-Chóliz, C. Sarasa, and L. Sun, *Modeling the carbon consequences of pro-environmental consumer behavior*, *Applied Energy*, vol. 184, pp. 1207–1216, 2016.

[22] C. L. Weber and H. S. Matthews, *Quantifying the global and distributional aspects of American household carbon footprint*, *Ecological Economics*, vol. 66, no. 2, pp. 379–391, 2008.

[23] A. Druckman and T. Jackson, *Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model*, *Energy Policy*, vol. 36, no. 8, pp. 3177–3192, 2008.

[24] C. Jones and D. M. Kammen, *Spatial Distribution of U.S. Household Carbon Footprints Reveals Suburbanization Undermines Greenhouse Gas Benefits of Urban Population Density*, *Environ. Sci. Technol.*, vol. 48, no. 2, pp. 895–902, 2014.

[25] M. Büchs and S. V. Schnepf, *Who emits most? Associations between socio-economic factors and UK households' home energy, transport, indirect and total CO2 emissions*, *Ecological Economics*, vol. 90, pp. 114–123, 2013.

[26] M. Lenzen, M. Wier, C. Cohen, H. Hayami, S. Pachauri, and R. Schaeffer, *A comparative multivariate analysis of household energy requirements in Australia, Brazil, Denmark, India and Japan*, *Energy*, vol. 31, no. 2, pp. 181–207, 2006.

[27] D. Saner, C. Beretta, B. Jäggi, R. Juraske, F. Stoessel, and S. Hellweg, *FoodPrints of households*, *The International Journal of Life Cycle Assessment*, vol. 21, no. 5, 2016.

[28] E. R. Frederiks, K. Stenner, and E. V. Hobman, *Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour*, *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1385–1394, 2015.

[29] D. Ivanova, G. Vita, K. Steen-Olsen, K. Stadler, P. C. Melo, R. Wood, and E. G. Hertwich, *Mapping the carbon footprint of EU regions*, *Environmental Research Letters*, vol. 12, no. 5, 2017.

[30] M. Sommer and K. Kratena, *The Carbon Footprint of European Households and Income Distribution*, *Ecological Economics*, vol. 136, pp. 62–72, 2017.

[31] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[32] United Nations Statistics Division, *COICOP (Classification of Individual Consumption According to Purpose)*, 2017. [Online]. Available: https://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5. [Accessed: 22-Nov-2017].

[33] N. S. Altman, *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*, *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[34] L. Breiman, *Random Forests*, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[35] R. Genuer, J.-M. Poggi, and C. Tuleau, *Random Forests: some methodological insights*, Institut National de Recherche en Informatique et en Automatique: Le Chesnay Cedex, France, 2008.

[36] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, *Journal of the Royal Statistical Society. Series B: Statistical (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[37] Federal Department of Economic Affairs Education and Research, *Preisüberwachung*, 2017. [Online]. Available: http://www.preisvergleiche.preisueberwacher.admin.ch. [Accessed: 12-Apr-2017].

[38] Federal Electricity Commission (ElCom), *Strompreis-Webseite*, 2017. [Online]. Available: https://www.strompreis.elcom.admin.ch. [Accessed: 20-Jan-2017].

[39] Federal Statistical Office (BFS), *Landesindex der Konsumentenpreise. Durchschnittspreise Energie*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2017.

[40] Bundesamt für Energie (BFE), *Schweizerische Gesamtenergiestatistik 2015*, Bundesamt für Energie (BFE): Bern-Ittigen, Switzerland, 2016.

[41] Schweizerischer Verein des Gas- und Wasserfaches (SVGW), *Trinkwasser*, 2017. [Online]. Available: http://wasserqualitaet.svgw.ch/index.php?id=874. [Accessed: 28-Nov-2017].

[42] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2009 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2010.

[43] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2010 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2011.

[44] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2011 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2012.

[45] Bundesamt für Statistik (BFS) and Bundesamt für Raumentwicklung (ARE), *Mobilität in der Schweiz. Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2012.

[46] R. A. Fisher, *Statistical methods for research workers*, 14th ed., New York, NY, USA: Hafner, 1973.

[47] W. H. Kruskal and W. A. Wallis, *Use of ranks in one-criterion variance analysis*, *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.

[48] T. Kohonen, *The Self-Organizing Map*, *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[49] T. K. Kohonen, *Self-Organizing Maps*, vol. 30, Springer Berlin Heidelberg, 2001.

[50] D. J. Dürrenmatt and W. Gujer, *Identification of industrial wastewater by clustering wastewater treatment plant influent ultraviolet visible spectra*, *Water Science & Technology*, vol. 63, no. 6, pp. 1153–1159, 2011.

[51] J. Vesanto and E. Alhoniemi, *Clustering of the Self-Organizing Map*, *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.

[52] J. Vesanto and M. Sulkava, *Distance matrix based clustering of the Self-Organizing Map*, in *Artificial Neural Networks - ICANN 2002*, 2002.

[53] L. Canetta, N. Cheikhrouhou, and R. Glardon, *Applying two-stage SOM-based clustering approaches to industrial data analysis*, *Production Planning and Control*, vol. 16, no. 8, pp. 774–784, 2005.

[54] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, *SOM Toolbox for Matlab 5*, Technical Report A57; Helsinki University of Technology: Helsinki, Finland, 2000.

[55] J. Vesanto, *Neural network tool for data mining: SOM toolbox*, in *Proceedings of symposium on tool environments and development methods for intelligent systems (TOOLMET 2000)*, 2000, pp. 184–196.

[56] H. Sen Tan and S. E. George, *Investigating learning parameters in a standard 2-D SOM model to select good maps and avoid poor ones*, in *Australasian Joint Conference on Artificial Intelligence*, 2004, pp. 425–437.

[57] J. MacQueen, *Some methods for classification and analysis of multivariate observations*, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281–297, 1967.

[58] S. Lloyd, *Least squares quantization in PCM*, *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[59] D. Arthur and S. Vassilvitskii, *K-Means++: the Advantages of Careful Seeding*, in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1035.

[60] M. R. Anderberg, *Cluster analysis for applications*, New York, NY, USA: Academic Press, 1973.

[61] J. A. Hartigan, *Clustering algorithms*, no. Wiley series in probability and mathematical statistics. Applied probability and statistics, New York, NY, USA: Wiley, 1975.

[62] J. H. Ward, *Hierarchical Grouping to Optimize an Objective Function*, *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[63] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, *Understanding and enhancement of internal clustering validation measures*, *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 982–994, 2013.

[64] P. J. Rousseeuw, *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*, *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[65] J. Lötsch and A. Ultsch, *Exploiting the Structures of the U-Matrix*, in *Advances in Self-Organizing Maps and Learning Vector Quantization*, Cham: Springer, 2014, pp. 249–257.

[66] A. Ultsch, *U*-Matrix: a Tool to visualize Clusters in high dimensional Data*, University of Marburg: Marburg, Germany, 2003.

[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2012.

[68] G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, and B. Weidema, *The ecoinvent database version 3 (part I): overview and methodology*, *The International Journal of Life Cycle Assessment*, vol. 21, no. 9, pp. 1218–1230, 2016.

[69] P. Koch and T. Salou, *AGRIBALYSE: Methodology Report - Version 1.2*, ADEME: Angers, France, 2015.

[70] A. Tukker, A. de Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, J. M. Rueda Cantuche, M. Bouwmeester, J. Oosterhaven, T. Drosdowski, and J. Kuenen, *Exiopol - Development and Illustrative Analyses of a Detailed Global MR EE SUT/IOT*, *Economic Systems Research*, vol. 25, no. 1, pp. 50–70, 2013.

[71] R. Wood, K. Stadler, T. Bulavskaya, S. Lutter, S. Giljum, A. de Koning, J. Kuenen, H. Schütz, J. Acosta-Fernández, A. Usubiaga, M. Simas, O. Ivanova, J. Weinzettel, J. H. Schmidt, S. Merciai, and A. Tukker, *Global Sustainability Accounting-Developing EXIOBASE for Multi-Regional Footprint Analysis*, *Sustainability (Switzerland)*, vol. 7, no. 1, pp. 138–163, 2015.

[72] Federal Statistical Office (BFS), *Landesindex der Konsumentenpreise. Durchschnittspreise ausgewählter Produkte*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2017.

[73] B. P. Weidema, C. Bauer, R. Hischier, C. Mutel, T. Nemecek, J. Reinhard, C. O. Vadenbo,

and G. Wernet, *Overview and methodology - Data quality guideline for the ecoinvent database version 3*, ecoinvent report No. 1(v3); ecoinvent: St. Gallen, Switzerland, 2013.

[74] C. Walker, E. R. Gibney, and S. Hellweg, *Comparison of Environmental Impact and Nutritional Quality among a European Sample Population – findings from the Food4me study*, *Scientific Reports*, vol. 8, no. 1, 2018.

[75] L. Scherer and S. Pfister, *Global Biodiversity Loss by Freshwater Consumption and Eutrophication from Swiss Food Consumption*, *Environ. Sci. Technol.*, vol. 50, no. 13, pp. 7019–7028, 2016.

[76] Federal Statistical Office (BFS), *STAT-TAB: Die interaktive Statistikdatenbank*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2016.

[77] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[78] A. Froemelt, M. Mauchle, B. Steubing, and S. Hellweg, *Greenhouse Gas Emissions Quantification and Reduction Efforts in a Rural Municipality*, *Journal of Industrial Ecology*, vol. 22, no. 1, pp. 92–105, 2018.

[79] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013.

[80] M. Goedkoop, R. Heijungs, M. Huijbregts, A. De Schryver, J. Struijs, and R. van Zelm, *ReCiPe 2008*, Ministrie van VROM: Den Haag, Netherlands, 2012.

[81] R. Frischknecht and S. Büsser Knöpfel, *Swiss Eco-Factors 2013 according to the Ecological Scarcity Method. Methodological fundamentals and their application in Switzerland*, Environmental studies no. 1330; Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2013.

[82] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung 2011: Kommentierte Ergebnisse und Tabellen*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2013.

[83] R. Buffat, A. Froemelt, N. Heeren, M. Raubal, and S. Hellweg, *Big data GIS analysis for novel approaches in building stock modelling*, *Applied Energy*, vol. 208, pp. 277–290, 2017.

[84] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

CHAPTER 6

ASSESSING ENVIRONMENTAL IMPACTS OF INDIVIDUAL HOUSEHOLDS: A LARGE-SCALE BOTTOM-UP LCA-MODEL FOR SWITZERLAND

Andreas Froemelt*, *René Buffat, Niko Heeren, Stefanie Hellweg*

Presented and published at SETAC Europe 28th Annual Meeting, 13-17 May, Rome, Italy, 2018

* The individual contribution of Andreas Frömelt consisted of conceptualizing the modeling framework, developing the model, coding, running the model, conducting the analyses and preparing the manuscript for publication.

SHORT ABSTRACT

Besides governmental consumption, household consumption is the main driver of economy, and is thus ultimately responsible for the environmental impacts that occur over the whole life cycle of products and services that households consume. Therefore, assessing environmental footprints of households is an important basis to identify environmental policies. This study aimed to develop a comprehensive regionalized bottom-up model for Switzerland that is able to assess the environmental impacts induced by individual households. The purpose of this model is to provide a virtual platform for detailed scenario analysis which shall support effective political decision making on different scales.

Three existing bottom-up models were merged: a building stock energy model, an agent-based transport simulation and a household consumption model. All of them were tested and evaluated beforehand. The physically-based building energy model establishes simplified energy balances for each residential building based on spatially and temporally resolved climate data, building characteristics and 3D-geometries. It provides estimates of space heating, hot water and electricity demand for each Swiss household. The mobility sub-model builds upon the results of an agent-based traffic simulation framework which was applied to Switzerland and reproduces mobility patterns of Swiss inhabitants in space and time. The third sub-model pursues a data-driven approach and enables the quantification of consumption of food, consumables, and other goods and services for each Swiss household by means of data mining techniques. Linking these sub-models with environmental background data allowed for computing an environmental profile for each household in Switzerland.

The application of this model to the current situation of Switzerland reveals interesting differences between individual households, different regions and different consumption areas. By covering the variability of household behavior and quantifying the demands and environmental footprints of households within a certain area, the model delivers important insights for local policymakers to derive targeted environmental strategies tailored to the specific problems and household types in a region. Furthermore, the high resolution of all three sub-models permits testing of policies and in-depth analyses of scenarios, ranging from detailed building refurbishment programs to future mobility solutions such as autonomous vehicle systems.

6.1 INTRODUCTION

Households are the main drivers of the economy by triggering a multitude of activities along the supply chain of products and services they consume. Therefore, household consumption can be regarded as ultimately responsible for any environmental impacts that occur over the life cycle of products and services. However, most household consumption studies apply top-down approaches and provide estimates of the impacts of the average national household. This is not an appropriate method to support local policymakers in their quest for specific strategies to reduce environmental impacts in their particular regions. The deduction and prioritization of targeted measures requires an understanding and a quantification of the variability in behavior of individual households within a certain area. A few approaches for such bottom-up models exist, but most are either limited in scope or do not capture the context of total household consumption. The goal of this study is to develop a comprehensive regionalized bottom-up household consumption model which is able to derive a realistic environmental profile for each household in a region. The presented approach was applied to the whole of Switzerland.

6.2 METHODOLOGY

Three existing bottom-up models were merged in the new regionalized household consumption model: a building stock energy model, an agent-based transport simulation and a data-driven consumption model. The physically-based building energy model [1] estimates space heating, hot water and electricity demand for each residential building in Switzerland based on simplified energy balances as a function of time, site, climate data, building characteristics, surrounding topography and 3D-geometries derived from laser-scanning data. Since the Swiss national census [2] indicates in which building a household lives, these housing energy estimates can be directly allocated to individual households. The mobility sub-model builds upon the simulation results of MATSim [3], an agent-based traffic simulation framework. The application of MATSim to Switzerland [4] reproduced the mobility behavior of the Swiss population and provides spatio-temporal information on chosen traffic modes and driven routes for each agent. Building upon spatial information and a number of personal characteristics, the simulated agents and their associated mobility demands were assigned to household members by means of a partially randomized optimization approach. The third sub-model [5] identified consumption patterns based on the Swiss Household Budget Survey [6] and derived 28 different consumption-based archetypes through extensive data mining techniques. These archetypes quantify the consumption behavior for food, consumables, and other goods and services for different clusters of households. To assign these archetypes to the households and thus to cover the remaining parts of consumption, a Random-Forest-Classifier was trained based on geographic information and household characteristics, as well as on housing energy and mobility demand, for merging the consumption model with the building energy model and the mobility sub-model. Because the national census [2] does not provide all information that is necessary to classify precisely a specific household as a certain archetype, we used the calibrated classifier to compute the probabilities with which the different archetypes can be assigned to a census household and then randomly sampled among these archetypes in question. This probabilistic assignment shall ensure the reproduction of a realistic variability of household behavior within a certain area. In a final step, the estimated housing, mobility and consumption demands were coupled with detailed life cycle background data in order to assess the resource uses and emissions along the whole supply chain. Pursuing a hybrid life cycle assessment approach, we included data from ecoinvent v3.3, Agribalyse v1.2 as well as EXIOBASE v2.2 to compute the environmental footprints of individual Swiss households.

6.3 Results and Discussion

The interlinked model assesses the current environmental footprints for all 4 million Swiss households as a realistic estimate taking into account the given circumstances of a particular household (Figure 6.1b shows the distribution of all household carbon footprints). The results of bottom-up models can be aggregated on any desired regional scale and thus, for instance, provide benchmarking maps of municipalities as shown in Figure 6.1a. In addition, different spatial structures can be compared. In Figure 6.1c, it becomes obvious that different degrees of urbanization exhibit similar total emissions per capita. However, the compositions of the footprints reveal that rural areas tend to cause larger mobility GHG emissions per person than urban regions. This is due to larger mobility demands and higher shares of car-driven kilometers. But even more detailed analyses of compositions are possible: more than 200 different consumption areas can be investigated in the model's highest resolution. Moreover, our model is able to apply all life cycle impact assessment methods supported by the background databases. However, we only show carbon footprints in Figure 6.1 to enable for comparison with existing top-down studies. For instance, the composition of the estimated Swiss average carbon footprint shown in Figure 6.1d is comparable with [7] while the absolute amount deviates by less than 15%. Housing, mobility and food are identified as the most important consumption areas in both studies.

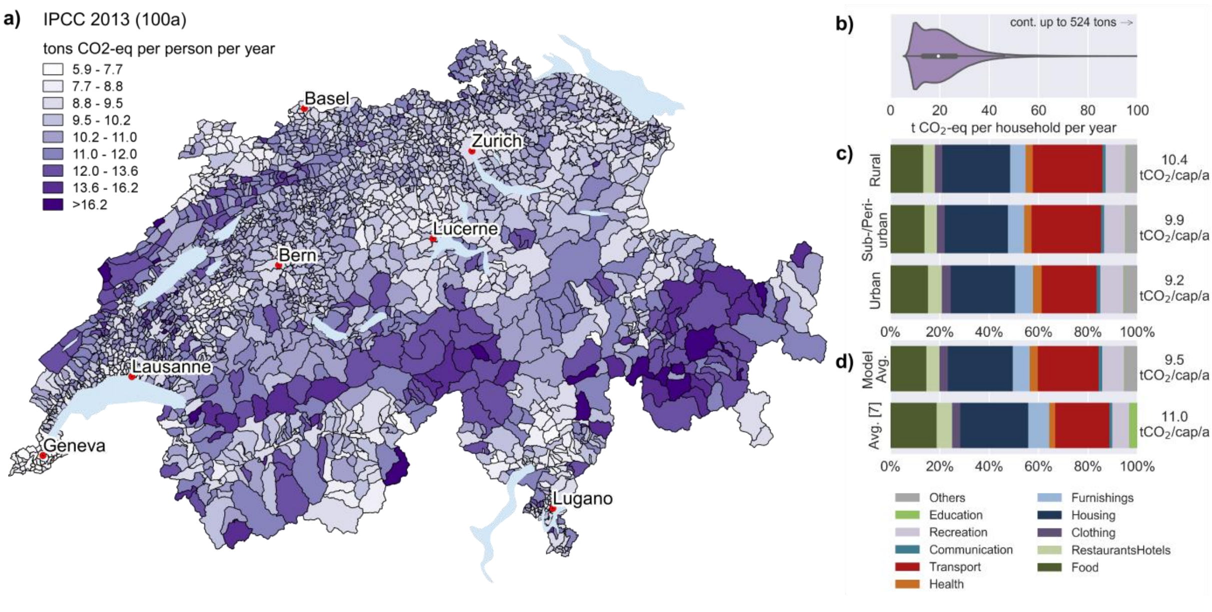


Figure 6.1: **Results for GHG: a) Average per capita emissions per municipality, b) Violinplot of the carbon footprints of all Swiss households, c) Comparison of different spatial structures on a per capita basis, d) Comparison of Swiss average with [7].**

6.4 Conclusions and Outlook

This highly resolved model enables for the comparison of households, regions, and different consumption areas. It provides insights into the specific problems of a region and allows for the analysis of consumption patterns within this area. The model may not only be used to derive targeted incentives for more sustainable consumption, but also to investigate in detail future scenarios and thus effects of planned measures. For instance, the component-based approach of the building model facilitates the analysis of detailed refurbishment scenarios, while the link to MATSim allows for including future mobility scenarios such as electric car penetration, increased home office activities or even autonomous vehicle systems. The model might thus be regarded as a virtual platform to evaluate policy scenarios aimed at lowering environmental impacts from household consumption. In the future, it could also serve as a basis for a complete agent-based model for Switzerland in which agents can manage their expenditures as well as interact with buildings and the mobility system. This improved model can then be used for analyzing dynamic scenarios such as diffuse penetrations of new technologies and might even capture associated rebound effects.

6.5 References

[1] R. Buffat, A. Froemelt, N. Heeren, M. Raubal, and S. Hellweg, *Big data GIS analysis for novel approaches in building stock modelling*, *Applied Energy*, vol. 208, pp. 277–290, 2017.

[2] Bundesamt für Statistik (BFS), *STATPOP 2013 - Statistik der Bevölkerung und Haushalte*, Neuchâtel, Switzerland, 2014.

[3] A. Horni, K. Nagel, and K. W. Axhausen, *The Multi-Agent Transport Simulation MATSim*, London, UK: Ubiquity Press, 2016.

[4] S. Hirschberg (ed.), C. Bauer, B. Cox, T. Heck, J. Hofer, W. Schenler, A. Simons, A. Del Duce, H.-J. Althaus, G. Georges, T. Krause, M. Vayá González, F. Ciari, R. A. Waraich, B. Jäggi, A. Stahel, A. Froemelt, and D. Saner, *Opportunities and challenges for electric mobility: An interdisciplinary assessment of passenger vehicles*, Villigen, Dübendorf and Zurich, Switzerland, 2016.

[5] A. Froemelt, D. J. Dürrenmatt, and S. Hellweg, *Using Data Mining To Assess Environmental Impacts of Household Consumption Behaviors*, *Environ. Sci. Technol.*, 2018.

[6] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[7] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

CHAPTER 7

CONCLUSIONS AND OUTLOOK

7.1 SYNTHESIS

The goal of this dissertation was to examine and develop new approaches to support effective environmental policymaking. The focus was placed on the collection, preparation and supply of data for decision- and policymakers on different scales. While it was not the goal to exhaustively investigate or to evaluate potential environmental policies, this dissertation aimed to provide information tailored to the decision-makers' spheres of influence and suited for facilitating the derivation, prioritization and implementation of targeted measures to mitigate negative impacts on the environment.

In Chapter 1, the following five research questions (RQ) were formulated:

RQ 1 What kind of information on which level of detail can serve as a basis to derive targeted measures aimed at mitigating environmental impacts?

RQ 2 What are efficient ways to provide this information, especially in view of constrained financial budgets to gather data and data scarcity in many sub-national regions?

RQ 3 How can Big Data and machine learning techniques contribute to the support of environmental policymaking?

RQ 4 Specifically in a consumption-based scope: how can household consumption patterns be modeled to capture the context of total consumption and how can the variability of these patterns be regionalized and thus transferred to larger scales?

RQ 5 What are the requirements and how should a framework be designed to evaluate and investigate large-scale effects of planned environmental measures?

In order to address these research questions, a dual approach was pursued. On the one hand, local policymakers were informed by processing detailed survey data within the scope of a transdisciplinary project in a small rural municipality. On the other hand, modeling approaches were developed to provide similarly detailed information for regions where in-depth surveys are missing. These two lines of research also reflect the two opposing possibilities to establish a comprehensive database for local policymakers: either data is gathered in surveys or generated by models.

While responses to **RQ1** can be found throughout the whole thesis, the research project *Zernez Energia 2020* was especially insightful in this regard (our contribution to this project is presented in Chapter 2). There it became apparent that a detailed comprehensive database should be the starting point for any policy tailored to local conditions. Based on an extensive stakeholder engagement process, the project derived targeted measures and developed a concrete action plan to reduce greenhouse gas emissions from the operation of the building stock [1, 2]. Thereby, the action plan foresees an extension of the district heating network, the installation of photovoltaic cells, and proposes replacements of heating systems and refurbishments of specific building components for actual buildings in the municipality [2]. The highly resolved database, which was established in the course of this project, not only enabled for the identification of these targeted actions, but was also an important tool to estimate the impact of such measures and thus to communicate with both the municipal authorities and the residents of Zernez. The experiences gained in this project confirm the importance of thorough databases as planning tools. The knowledge of the actors and processes within the geographical system boundaries especially proved its usefulness, be it e.g. for the intended placement of photovoltaic cells or for suggestions with regard to increased insulations in specific buildings. Understanding and quantifying the variability of the local actors also helped to prioritize measures. Furthermore, the analyses in Chapter 2 reveal the suitability of considering the environmental impacts of a geographical region from two angles, the production-based and the consumption-based perspective. Both accounting frameworks provided insights from different viewpoints and allowed for identifying areas of actions for the municipal authorities on account of their complementarity.

However, the laborious data collection process in Zernez also made obvious that such an effort is not a feasible option for other municipalities, particularly in view of constrained financial budgets for data collection. Furthermore, even the unique project database of Zernez with 100% coverage of all buildings will face future challenges pertaining to updating and maintaining the data[1]. This clears the stage for **RQ2** which asks for more efficient ways to provide comprehensive information without the need for antecedent excessive data acquisition. In response to this research question, a model framework was elaborated in Chapters 3 to 6 (see also flow scheme in Figure 1.1 of Chapter 1). Based on the insights with regard to RQ1 in the Zernez-project and the reasoning in Chapter 1, mainly three basic principles were followed while developing the models: 1. The models were constructed **bottom-up**. This means that the starting point of the models are individual units or entities in a study area (e.g. individual buildings or individual households). This allows for capturing the variability among different actors and thus supports the quest for answers to questions such as: Who should be targeted? What are promising measures for these target groups? What is the reduction potential for individual actors, but also with regard to total impacts? Furthermore, in Chapter 2, it was also revealed that the abilities to implement measures are limited for municipal authorities. There, we indicated that some fields of actions are of supraregional importance and need to be tackled by cantonal or national authorities. In this regard, another advantage of bottom-up models is their capability to aggregate results on any desired regional scale. They are thus not only tailored to a municipal scope, but can also be used by regional, cantonal or national authorities. 2. As a second principle, the input data of the models was sourced from **national statistics and publicly accessible databases** which are transparent and well-established. The motivation for this was threefold. First, many – especially rural – municipalities lack detailed data. In such data-scarce situations, national statistics and federal registers provide a promising starting point for models attempting to generate more specific data. Additionally, these databases are regularly updated and help to keep the models up-to-date. Second, the use of such data also ensures a consistent modeling over a large area and thus allows for comparing and benchmarking different regions or municipalities. Third, similar registers and statistics are also maintained in other countries. Although the application of the developed models was demonstrated for the case of Switzerland, the approaches could also be used for other nations. 3.

[1] Note that one of the measures in the action plan [2] also comprises updating the database.

As mentioned in Chapter 1, consumption-oriented accounting is particularly important for countries like Switzerland given the large share of embodied emissions in the total impacts induced by Swiss residents. For this dissertation, the scope of the models was thus restricted to providing a **consumption-based life cycle perspective**. In combination with the first principle (bottom-up modeling), this means that the central modeling element of the overall model framework were individual households. However, for the respective sub-models which serve the overall model, other units could be considered (e.g. individual buildings in the case of the building energy models in Chapters 3 and 4).

The model approaches presented in Chapters 3 to 6 build upon the above principles and offer answers to RQ2. Additionally, our approaches were not kept in a theoretical stage but their practical applicability was demonstrated on large scale for the case of Switzerland. With respect to the physically-based building energy model, this required an evaluation of the model performance in advance. For this purpose, Chapter 3 scrutinizes the building stock model of Saner and colleagues [3]. The evaluation of this model with the detailed building database of Zernez showed that it is generally well suited for building energy modeling and is able to provide a realistic picture of a municipality's building stock. Furthermore, it complies with the above requirements by building upon publicly accessible databases and providing estimates for individual buildings, while simultaneously allowing for a municipal or district view. The comparison with primary data and the global sensitivity analysis in Chapter 3 also reveals improvement potential. This could be attributed in particular to the used heat transfer coefficients and the behavior of occupants. In an attempt to overcome these flaws and to take advantage of the possibilities of Big Data, Chapter 4 enhanced the building energy model in several aspects by integrating comprehensive geographic data. Apart from several new approaches including the consideration of shadowing effects from topography, the building energy model of Chapter 4 also allows for deducing the envelope of individual buildings from digital elevation models. This reduces model uncertainties and immediately addresses one of the main weaknesses pointed out in Chapter 3; since the areas of walls, roofs and windows are directly multiplied with heat transfer coefficients (see equation (4.2)), improved estimates of the envelope have a similar effect as updated coefficients. Hence, the achieved improvements in Chapter 4 demonstrate the potential of using Big Data and also respond to **RQ3**.

The train of thought driven by RQ3 is further investigated in Chapter 5. In this chapter, we propose a novel approach to assess environmental impacts induced by household consumption behavior. By employing data mining techniques, we exploit the data of the Swiss Household Budget Survey (HBS) [4]. While missing information was successfully imputed with regression models, the application of a two-staged clustering enabled for the recognition of consumption patterns and consequently for deriving consumption-based archetypes. These archetypes constitute insightful building blocks for policymakers. They provide the quantification and environmental assessment of a manageable set of typical household consumption patterns based on real observable behavior and thus capture and preserve the context of total consumption (first part of **RQ4**). Previous studies have also aimed at quantifying the variability of environmental impacts of different household groups based on consumer survey data [5–14]. Indisputably, these studies have delivered valuable insights for the understanding of household consumption, but all of them used pre-defined household segments based on solely socio-economic characteristics. Several of these studies point out that there might still be an important variability of behavior within the considered household types. Girod and De Haan [15] confirm this assumption quantitatively. The application of clustering techniques in our approach allowed for including consumption data along with socio-economic parameters and thus forming new household groups with comparable living conditions and similar behavior concurrently. This resulted in the recognition of patterns that can be regarded as archetypical behaviors and whose environmental impacts could then be investigated. In fact, the analyses in Chapter 5 show that households in similar socio-economic circumstances might differ in their consumption behavior and consequently in their environmental footprints, be it with regard to total impacts or with regard to composition. Additionally, archetypes that deviate from general macro-trends emerge. For instance, Chapter 5 observes generally increased impacts with higher income; however, few archetypes diverge from this by causing comparably low impacts with continued high income. Understanding the nature and implications of such different household behavior is the first step towards the derivation and prioritization of policies tailored to specific consumer groups. The proposed archetype-approach allows for analyses at household level and forms thus a promising basis for further investigations. In this context, it should be noted that recent studies emphasize that behavioral economics and psychology ought to be involved in the development of successful measures [16–18]. In conclusion, Chapter 5 provides needed information (RQ1) in a modeling approach (RQ2) which seizes household consumption in a coherent context (RQ4) using machine learning and data mining techniques to gain new information for environmental policymaking (RQ3).

The second part of **RQ4** goes a step further and asks not only how household behavior can be modeled and analyzed, but also how this variability of household consumption can be regionalized and thus applied to estimate the environmental impacts of real households. This is of high importance to local policymakers. They are capable of devising and prioritizing environmental strategies only if they know the consumption patterns and the associated environmental hotspots in their regions. For this purpose, Chapter 6 embarks on a novel approach by employing again techniques from the data mining toolbox (RQ3). A direct assignment of archetypes to real households of the national census [19] in the scope of a classification framework was considered to not deliver reasonable results; the census only provides socio-demographic information and thus lacks important variables which were used to derive the archetypes in Chapter 5. A reliable matching of census households with archetypes could thus not be expected. Therefore, in a first step, we deployed a Random-Forest-Classifier [20] to compute the probabilities that a census household would belong to a specific archetype. In a second step, an archetype was allocated to this census household in a random sampling process based on these probabilities. Furthermore, this probabilistic classification approach comprises data not only from the national census, but also from the building energy model (Chapters 3 and 4) and the mobility sub-model in order to maintain the interrelations of consumption areas (see also Figure 1.1). The mobility model introduced in Chapter 6 builds upon the preliminary works of Saner et al. [3] and Froemelt in [21]. Based on the results of an agent-based transport simulation of Switzerland with MATSim (Multi-Agent Transport Simulation) [22], this sub-model assesses mobility demand of households by assigning simulated agents to household members. The overall, interlinked model provides realistic estimates of the environmental footprints for all households in Switzerland. Therefore, it answers RQ4 and additionally proves the feasibility of large-scale bottom-up models (RQ2) supplying in-depth knowledge for environmental decision-makers (RQ1). In fact, the model is highly resolved in space and detail by estimating the demands in almost 400 consumption areas for about four million real households and by subdividing the resulting environmental impacts into more than 200 categories. Consequently, this model constitutes a predestined knowledge base to analyze the status quo, find hotspots, identify fields of actions and target groups of households for reducing consumption impacts, and finally to start developing targeted environmental policies on any desired regional scale. The bottom-up character of the model offers a high flexibility to policymakers in terms of the provided level of detail and delivers insights into locally occurring consumption patterns.

Bearing the experiences from the Zernez-project in mind and coming back to RQ1, another important piece of needed information are assessments of the consequences of planned actions. Therefore, an important question has this far remained open; how should a framework be designed to enable a thorough evaluation of intended environmental policies (**RQ5**)? For the models elaborated in the scope of this dissertation, this aspect was already considered in the development phase. For instance, a physical engineering model was favored over a statistical model for estimating building energy demand (see also the reasoning in Chapter 3). These models build upon physical principles and are thus capable of evaluating the effects of physical measures such as the refurbishment of building components [23]. Indeed, the bottom-up and component-based structure of the building energy model (Chapter 4) follows also the ideas of Heeren and colleagues [24] and will allow for evaluating detailed building refurbishment scenarios. For instance, the model is not only able to consider the effects of refurbishing specifically chosen individual buildings, but also the effects of retrofitting only particular components of these edifices (e.g. what are the effects of only exchanging windows and insulating walls, but not improving the roof?). Similarly, the link to the implementation of the agent-based transport simulation framework MATSim [22] facilitates the analysis of future mobility scenarios [21]. Being interlinked with the estimates of archetypical behavior, all scenarios analyzed in our modeling framework will not only be considered in the context of total household consumption, but may also be evaluated with regard to potential burden shifts between consumption areas (rebounds). Additionally, the consumption model enables the simulation of further scenarios such as changes in diets. Further application possibilities for scenario analysis are contemplated in section 7.3.

Last but not least, we would like to point out that we see our modeling efforts in the light of being supportive for policymakers. The modeling framework shall provide realistic estimates to bridge knowledge gaps and thus complement available information. However, the model's high level of detail shall not obscure the fact that it remains a model, meaning that it is an approximation of reality and accordingly needs to be treated with common sense. Or to conclude with a famous quotation of George Box (1978): "All models are wrong but some are useful". In this sense, we hope that our model will be useful for future environmental policymaking.

7.2 SCIENTIFIC AND PRACTICAL RELEVANCE

The main scientific contribution of this dissertation is the development and the application of a comprehensive large-scale bottom-up model to assess demands and associated environmental impacts from individual households. Thereby, novel approaches were elaborated in different aspects, but especially in the light of how the exploitation of Big Data and the application of data mining techniques might create insights in support of environmental policymaking. Key innovations range from the use of digital elevation models to derive building envelopes and to consider location-specific shadowing including effects from topography (Chapter 4), the identification of household behavior archetypes and their environmental impacts (Chapter 5) to the large-scale assignment of these archetypes to real households (Chapter 6). We hope that these ideas may either help other modelers in solving problems encountered or spark ignition to come up with creative thoughts and with better models. As will be discussed in section 7.3, our model allows for further investigations in many ways. For instance, the modular structure of the overall model facilitates the coupling with models of other researchers. In this sense, the modeling framework of this dissertation could serve as an input to other scientific models or vice versa: other models could supply more detailed information to our model. Both ways could be fruitful and deliver more insights for effective environmental policymaking. Moreover, the availability of the ready-to-use archetypes for Switzerland (Chapter 5) could also offer an interesting basis to analyze consumption behavior for economists, psychologists or sociologists. This could finally support fostering the efforts to achieve a better understanding of household consumption behavior. Last but not least, the input datasets for all models in this dissertation are often also available in other countries. Following our concepts, it would be possible to set up similar models for countries other than Switzerland. This would not only be interesting for policymakers in these nations, but also to compare and study consumption patterns on an international level.

This dissertation is not only relevant for science, but it has specifically also aimed at being important for practice. First, we would like to emphasize its practical relevance with regard to the research project in Zernez. Unfortunately, current administrative processes in connection with reorganizations and the fusion with neighboring municipalities have caused some delay for applying the developed action plan. Nevertheless, the municipality still intends to implement the planned measures which would lead to a significant reduction of greenhouse gas emissions in Zernez. On the one hand, this will contribute to the Sustainable Development Goal 13 ("Climate Action") [25] and thus to the global fight against climate change. On the other hand, Zernez could also serve as a role model for other comparable municipalities.

The overall model developed in the context of this dissertation provides a highly resolved data basis which can be used either by consultants or directly by policymakers in order to develop impactful environmental strategies. In addition, the current model results, as well as the computation of scenarios of planned actions, can help to communicate intended measures to stakeholders, such as inhabitants, local companies, utility operators, or authorities on a different level (e.g. national or cantonal agencies). In this regard, the outcomes of this dissertation could effectively lead to a reduction of adverse environmental impacts. Although these direct model uses are in principle possible, it should not be overlooked that the current model is still in a rather academic state. Consequently, in order to achieve practical relevance, the challenge of disseminating our approaches and results needs to be tackled. In respect thereof, first attempts have been undertaken. For instance, a prototype of a web-based decision support tool based on the building energy model of Chapter 4 was presented in September 2017 at the CISBAT-conference [26]. This tool is designed for homeowners and allows them to select and compare different refurbishment options for their buildings. In an adjusted format, this tool could also inform policymakers about the effects of different building retrofit scenarios. Additionally, ongoing research in the scope of the SCCER Mobility (Swiss Competence Center for Energy Research: Efficient Technologies and Systems for Mobility) [27] aims at designing a decision support system building upon the overall model of this dissertation – but with a special focus on the mobility sub-model – which shall help to identify optimal mobility portfolios for municipalities and potential incentive schemes. Last but not least, it is also envisaged to provide the model results in the form of a footprint atlas for Switzerland, similar to the data visualizations which have been developed for the Carbon Footprint of Nations [28].

7.3 CRITICAL APPRAISAL AND OUTLOOK

7.3.1 Model Evaluation and Uncertainty Analysis

A model can only be considered useful if it meets the initial intentions within a certain accuracy range. The goal of the modeling framework developed in this dissertation is to provide a realistic picture of variability in environmental footprints induced by individual households. Chapter 3 and 4 subject the building energy model to an in-depth evaluation and appraise this sub-model as suitable for the present purpose. Also, the performance of MATSim applications to Switzerland have been investigated including juxtapositions with automated traffic counters [29]. Additionally, we analyzed the mobility demands resulting from the assignment of agents to actual household members in the context of the overall model; in an attempt to evaluate how well the computed person-kilometers reproduce the real variability, differently aggregated results were compared with the Swiss Mobility and Transport Microcensus 2010 [30]. A selection of these comparisons is presented in Appendix E. In conclusion, the mobility sub-model was deemed reasonable for the overall model's aim. The archetypes in Chapter 5 were derived from the Swiss Household Budget Survey [4] which can be regarded as representative for Switzerland [31]. However, as long as the ground truth is unknown, a clustering procedure is difficult to evaluate [32]. Nevertheless, in Chapter 5, we carefully followed good modeling practice and used internal evaluation measures in several steps [33]. For instance, the regression models used to impute data were subjected to

10-fold cross-validation in the training phase and their results were afterwards compared to overall national statistics. The clustering itself used different performance indicators. In addition, it took a minimum number of households into account which can be regarded as representative for household groups according to the Federal Statistical Office [31]. The question remains if the extrapolation of the archetypes to real households in Chapter 6 still provides a realistic picture of household consumption. First, it needs mentioning that the applied Random-Forest-Classifier [20] was again tuned in an internal cross-validation process based on a 90% training-set and the probabilities were subsequently calibrated with the remaining 10% of the dataset [33–35]. Second, to prevent problems with class imbalance and thus to account for prevalence, a stratified splitting was used during the cross-validation procedure and the clusters themselves were weighted within the Random-Forest-Classifier according to their frequency in the original dataset [33, 34]. These measures were successful and resulted in correlation coefficients of 0.86 (Pearson) and 0.87 (Spearman) for the archetypes' prevalence in the HBS [4] and the overall model, respectively. This means that the assignment of the archetypes to real households yielded a similar frequency distribution of archetypes in the model as in the original HBS-data. Correspondingly, the comparisons of differently aggregated expenditures and revenues of the model with the original HBS resulted in a good agreement (see Appendix E). Additionally and similar to Appendix D.2.4 the national statistics for housing-related data could be satisfactorily reproduced. Finally, Chapter 5 and 6 both relate the resulting environmental impacts of the model to other consumption-based studies of Switzerland [36, 37]. This further underpins the plausibility and the reasonability of the model results.

We would like to point out that – apart from the above-mentioned internal validation-mechanisms (e.g. cross-validation) – no possibility exists to evaluate the model in a household-by-household comparison since external primary data in such a high resolution is not available. However, the goal of the model was to provide a realistic – but not necessarily an accurate – estimate of the consumption-induced environmental impacts for individual households and a realistic – but not necessarily an accurate – picture of the variability in household footprints within a certain area. The internal cross-validations as well as the above-presented attempts to analyze the model's ability to reproduce the overall characteristics of national statistics confirm that the model points in the right direction and is able to calculate realistic assessments of consumption footprints. Nevertheless, we see it as an open task to look for more data and to further improve our validation attempts.

Especially in situations facing difficulties to perform in-depth external validations, it is important to include uncertainty analyses. Thereby, we would like to emphasize that a comprehensive Monte Carlo simulation is incorporated in the building energy model of Chapter 4. In spite of the fact that uncertainty data is lacking for the Swiss Household Budget Survey [4], current research aims at developing an uncertainty framework also for the archetypes. As a possible way to capture uncertainties of the archetypes' centroid vectors, we conceive to conduct bootstrap-sampling[2]

[2] See for instance [33] for a description of bootstrap-sampling.

within each cluster. Furthermore, Monte Carlo simulations can directly be implemented in the allocation of archetypes to households as well as in the assignment of MATSim-agents to household members. The latter has already been done in the preliminary works of Saner and colleagues [3]. In addition to capturing uncertainties introduced by the modeling of demands, also uncertainty originating from the environmental background databases shall be included in future research.

7.3.2 Insights for Deriving Targeted Measures

The results of the individual archetypes provide a valuable basis for identifying strategies tailored to the different archetypical behaviors. One interesting direction of future work could comprise a systematic development of targeted measures for all 28 archetypes. Schanes and colleagues [16] elaborated a methodical framework to deduce strategies aiming at reducing consumption-induced carbon footprints. The application of such frameworks to the consumption-based archetypes could come up with ready-made measures for different target groups. Further insights are also conceivable with regard to prioritization of strategies, e.g. in the case some measures turn out to be more impactful than others or some policies apply to more than only one archetype. Moreover, as mentioned in Chapter 5, drivers of environmental impacts could – and should – be investigated based on the archetype-approach. In this regard, it would also be interesting to apply similar clustering attempts to other countries or larger datasets. This could further improve our understanding of today's consumption patterns. Finally, the derivation of successful measures needs to be accompanied by additional input from other disciplines including economics and psychology to achieve an effective and long-lasting change of consumption patterns. For instance, Frederiks and colleagues [17] argue that people sometimes do not respond in an expected nor desired manner to incentives (be it rewards or sanctions). While the archetypes provide quantitative information for the development of interventions, it is thus also crucial to understand motivational factors and cognitive biases of households that should be encouraged changing towards more sustainable consumption.

7.3.3 Scenario Analyses: Effects of Planned Measures

The overarching model of this dissertation constitutes a comprehensive information base and can also function as a virtual platform for detailed scenario analysis to seize the effects of planned measures. As mentioned above, the highly resolved bottom-up structure of the model allows for setting up and analyzing detailed policy scenarios. For instance, the physically- and component-based approach of the building energy model enables computing the effects of retrofitting specific building components (i.e. roofs, walls, windows, and floors) of individual buildings. Consequently, various refurbishment programs can be investigated encompassing best case scenarios or scenarios targeting only specific buildings, specific geographical regions or specific homeowners (e.g. target groups which were identified due to their consumption behavior and living conditions).

The coupling with the MATSim framework opens up a plethora of possibilities to analyze future mobility scenarios. In the scope of the THELMA-project [21] (<u>Tec</u><u>h</u>nology-centered <u>El</u>ectric

Mobility Assessment), future scenarios including increased home office activities, the impacts of teleshopping and different scenarios of electric car penetrations were already computed and could directly be fed to our model. Additionally, efforts have been made to analyze the introduction of autonomous vehicles [38, 39] with MATSim. Such groundbreaking systems have gained a lot of attention in recent years and it would be important to analyze their environmental implications in a broader picture, e.g. in the scope of this dissertation's modeling approach.

The consumption sub-model enables the simulation of scenarios such as changes in diets as well as the consideration of policy scenarios in the context of total household consumption. In this regard, follow-up research envisages capturing rebound effects of policy measures. For instance, if policies lead to demand reductions (e.g. less heating demand on account of better insulation) and consequently to lower expenditures in one consumption area for a certain household, this could lead to a change in the archetypical behavior of this household. The most basic approach in this regard could be to assign another archetype to the respective household. However, it would also be possible to elaborate more sophisticated models which build upon the variability of household behavior within the clusters of Chapter 5 and which thus allow for capturing also small behavior changes.

In addition to the computation of future scenarios, we also see further research options in the analysis of the current situation. Chapter 6 already provides a first rudimentary comparison of different regions in Switzerland. However, an in-depth study of different municipalities, cities, and interrelations of consumption areas could deliver important insights to deriving successful policies in addition to the efforts mentioned in section 7.3.2. Such a spatial comparison and benchmarking analysis could help municipalities to learn from each other.

7.3.4 Model Improvement and Extension

The future research possibilities discussed so far relate to direct applications of the developed approaches and models. Yet, a multitude of opportunities to improve and extend the existing model already exist. For instance, Chapter 3 reveals improvement potential for the building energy model with regard to heat transfer coefficients and the consideration of occupants' behavior. While the first issue is addressed in Chapter 4, the second one remains unsolved. Possible improvements regarding occupational behavior modeling could be achieved either by deriving archetypical behavior patterns for housing energy[3] or by reinforcing the connection to the agent-based MATSim-model. The latter possibility could even be pursued until a complete merger of the two frameworks resulting in a full-fledged agent-based model. In such a model, agents would not only interact with the mobility system, but also with buildings and even manage their financial budgets and purchase decisions. The highly dynamic structure would allow for analyzing policy scenarios in an unprecedented detail and capturing complex, non-linear effects of planned measures. In particular, temporal aspects such as the diffuse penetration of new technologies could be anticipated.

[3] This could be done analogous to the procedure in Chapter 5 but this time applied to e.g. smart meter data.

Another line of possible model extensions concerns the coupling with macro-economic models. Econometric models enable for further investigating supply chains and additionally for tracking system-wide impacts induced by policy measures. Hereby, the consideration of rebound effects plays an important role. Even though section 7.3.3 discusses a straight-forward implementation for this in the scope of the current model, the application of sophisticated econometric models as e.g. used by Sommer and Kratena [40] or by Duarte et al. [41] could provide more comprehensive insights in this respect. These two studies coupled input-output tables with computable general equilibrium models and could thus reproduce economy-wide feedbacks and rebounds of measures.

Last but not least, we would like to challenge the adopted system perspective in the developed model. As is argued above and in Chapter 1, a consumption-based perspective is justified for our model. The responsibility of households is conceptually reasonable and they can definitely be considered as key actors when it comes to reducing environmental impacts. However, it should not be ignored that households do not have full control of the supply chains serving them [18, 42]. They are able to send indirect signals by their purchase choices, but they cannot directly change economic activities or economic systems. Consequently, only focusing on households will not suffice to reduce emissions and resource uses to more sustainable levels. The environmental impacts of local industry and trade need to be tackled as well and producers must accept their responsibility [43]. Consequently, a useful complement to the consumption-focused approaches would be a bottom-up model which reproduces a production-based view of a particular region. To provide a realistic impression of the variability of local actors, this "industry model" should take individual companies or enterprises as the central modeling elements. The idea of such a model is also in line with the findings of Chapter 2 which highlight the complementarity of both system perspectives and its importance to derive effective environmental measures. The availability of both perspectives in a single model would provide deeper insights into interventions along the production-consumption dichotomy [43]. Even though such a model is not directly evolvable from the presented approaches, the successful application of data mining techniques and the exploitation of Big Data in this dissertation are encouraging that a production-oriented bottom-up model building upon national statistics and registers could be feasible.

Although improvement potential exists and interesting enhancements should be envisaged, we would like to conclude that the present consumption-based model provides already now a thorough knowledge base which helps local policymakers to understand and quantify prevailing consumption patterns in their sphere of influence. It is not only able to support the identification of strategies tailored to specific regions and different household types, but can also be regarded as a virtual platform for scenario analysis. Finally, it also constitutes a starting point for more detailed investigations to understand today's consumption patterns and is open for further developments and extensions.

7.4 REFERENCES

[1] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[2] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Aktionsplan*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010525998, 2015.

[3] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[4] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[5] M. Wier, M. Lenzen, J. Munksgaard, and S. Smed, *Effects of household consumption patterns on CO2 requirements*, *Economic Systems Research*, vol. 13, no. 3, pp. 259–274, 2001.

[6] C. M. Jones and D. M. Kammen, *Quantifying carbon footprint reduction opportunities for U.S. households and communities*, *Environ. Sci. Technol.*, vol. 45, no. 9, pp. 4088–4095, 2011.

[7] A. Druckman and T. Jackson, *The bare necessities: How much household carbon do we really need?*, *Ecological Economics*, vol. 69, no. 9, pp. 1794–1804, 2010.

[8] A. Druckman and T. Jackson, *Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model*, *Energy Policy*, vol. 36, no. 8, pp. 3177–3192, 2008.

[9] K. Steen-Olsen, R. Wood, and E. G. Hertwich, *The Carbon Footprint of Norwegian Household Consumption 1999-2012*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[10] M. Lenzen, M. Wier, C. Cohen, H. Hayami, S. Pachauri, and R. Schaeffer, *A comparative multivariate analysis of household energy requirements in Australia, Brazil, Denmark, India and Japan*, *Energy*, vol. 31, no. 2, pp. 181–207, 2006.

[11] S. Bin and H. Dowlatabadi, *Consumer lifestyle approach to US energy use and the related CO2 emissions*, *Energy Policy*, vol. 33, no. 2, pp. 197–208, 2005.

[12] J. C. Minx, T. Wiedmann, R. Wood, G. P. Peters, M. Lenzen, A. Owen, K. Scott, J. Barrett, K. Hubacek, G. Baiocchi, A. Paul, E. Dawkins, J. Briggs, D. Guan, S. Suh, and F. Ackerman, *Input-output analysis and carbon footprinting: An overview of applications*, *Economic Systems Research*, vol. 21, no. 3, pp. 187–216, 2009.

[13] G. Baiocchi, J. Minx, and K. Hubacek, *The Impact of Social Factors and Consumer Behavior on Carbon Dioxide Emissions in the United Kingdom. A Regression Based on Input-Output and Geodemographic Consumer Segmentation Data*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 50–72, 2010.

[14] C. L. Weber and H. S. Matthews, *Quantifying the global and distributional aspects of American household carbon footprint*, *Ecological Economics*, vol. 66, no. 2, pp. 379–391, 2008.

[15] B. Girod and P. De Haan, *GHG reduction potential of changes in consumption patterns and higher quality levels: Evidence from Swiss household consumption survey*, *Energy Policy*, vol. 37, no. 12, pp. 5650–5661, 2009.

[16] K. Schanes, S. Giljum, and E. Hertwich, *Low carbon lifestyles: A framework to structure consumption strategies and options to reduce carbon footprints*, *Journal of Cleaner Production*, vol. 139, pp. 1033–1043, 2016.

[17] E. R. Frederiks, K. Stenner, and E. V. Hobman, *Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour*, *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1385–1394, 2015.

[18] D. O'Rourke and N. Lollo, *Transforming Consumption: From Decoupling, to Behavior Change, to System Changes for Sustainable Consumption*, *Annual Review of Environment and Resources*, vol. 40, pp. 233–259, 2015.

[19] Bundesamt für Statistik (BFS), *STATPOP 2013 - Statistik der Bevölkerung und Haushalte*, Neuchâtel, Switzerland, 2014.

[20] L. Breiman, *Random Forests*, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] S. Hirschberg (ed.), C. Bauer, B. Cox, T. Heck, J. Hofer, W. Schenler, A. Simons, A. Del Duce, H.-J. Althaus, G. Georges, T. Krause, M. Vayá González, F. Ciari, R. A. Waraich, B. Jäggi, A. Stahel, A. Froemelt, and D. Saner, *Opportunities and challenges for electric mobility: An interdisciplinary assessment of passenger vehicles*, Villigen, Dübendorf and Zurich, Switzerland, 2016.

[22] A. Horni, K. Nagel, and K. W. Axhausen, *The Multi-Agent Transport Simulation MATSim*, London, UK: Ubiquity Press, 2016.

[23] L. G. Swan and V. I. Ugursal, *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques*, *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.

[24] N. Heeren, M. Jakob, G. Martius, N. Gross, and H. Wallbaum, *A component based bottom-up building stock model for comprehensive environmental impact assessment and target control*, *Renewable and Sustainable Energy Reviews*, vol. 20, pp. 45–56, 2013.

[25] Secretary-General (United Nations), *Progress towards the Sustainable Development Goals*, E/2017/66; United Nations: New York, 2017.

[26] R. Buffat, L. Schmid, N. Heeren, A. Froemelt, M. Raubal, and S. Hellweg, *GIS-based Decision Support System for Building Retrofit*, *Energy Procedia*, vol. 122, pp. 403–408, 2017.

[27] SCCER Mobility, *Swiss Competence Center for Energy Research: Efficient Technologies and Systems for Mobility (SCCER Mobility)*, 2018. [Online]. Available: http://www.sccer-mobility.ch/. [Accessed: 26-Feb-2018].

[28] Norwegian University of Science and Technology (NTNU), *Carbon Footprint of Nations*, 2018. [Online]. Available: carbonfootprintofnations.com. [Accessed: 26-Feb-2018].

[29] K. Meister, M. Balmer, F. Ciari, A. Horni, M. Rieser, R. A. Waraich, and K. W. Axhausen, *Large-Scale Agent-Based Travel Demand Optimization Applied to Switzerland, Including Mode Choice*, in *12th World Conference on Transportation Research*, 2010.

[30] Amt für Raumentwicklung, Amt für Energie und Verkehr, Amt für Natur und Umwelt, Amt für Wirtschaft und Tourismus, Tiefbauamt, and Verkehrsplanung Stadt Chur, *Mobilität in Graubünden - Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Chur, Switzerland, 2012.

[31] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung 2011: Kommentierte Ergebnisse und Tabellen*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2013.

[32] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, *Understanding and enhancement of internal clustering validation measures*, *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 982–994, 2013.

[33] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer Dordrecht, Heidelberg / New York / London, 2013.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2012.

[35] A. Niculescu-Mizil and R. Caruana, *Predicting good probabilities with supervised learning*, in *Proceedings of the 22nd international conference on Machine learning - ACM*, 2005, pp. 625–632.

[36] B. Girod and P. De Haan, *More or Better? A Model for Changes in Household Greenhouse Gas Emissions due to Higher Income*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 31–49, 2010.

[37] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[38] P. M. Bösch, F. Ciari, and K. W. Axhausen, *Required autonomous vehicle fleet sizes to serve different levels of demand*, *Transportation Research Board 95th Annual Meeting*, no. 16–2682, 2016.

[39] R. Frischknecht, C. Bauer, A. Froemelt, S. Hellweg, K. Biemann, T. Buetler, B. Cox, P. De Haan, S. Hoerl, R. Itten, N. Jungbluth, Y. Ligen, N. A. Mathys, S. Schiess, S. Schori, P. van Loon, J. Wang, and S. Wettstein, *LCA of mobility solutions: approaches and findings — 66th LCA forum, Swiss Federal Institute of Technology, Zurich, 30 August, 2017*, *The International Journal of Life Cycle Assessment*, vol. 23, no. 2, pp. 381–386, 2018.

[40] M. Sommer and K. Kratena, *The Carbon Footprint of European Households and Income Distribution*, *Ecological Economics*, vol. 136, pp. 62–72, 2017.

[41] R. Duarte, K. Feng, K. Hubacek, J. Sánchez-Chóliz, C. Sarasa, and L. Sun, *Modeling the carbon consequences of pro-environmental consumer behavior*, *Applied Energy*, vol. 184, pp. 1207–1216, 2016.

[42] J. H. Spangenberg and S. Lorek, *Environmentally sustainable household consumption: From aggregate environmental pressures to priority fields of action*, *Ecological Economics*, vol. 43, no. 2–3, pp. 127–140, 2002.

[43] A. Tukker, M. J. Cohen, K. Hubacek, and O. Mont, *The Impacts of Household Consumption and Options for Change*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 13–30, 2010.

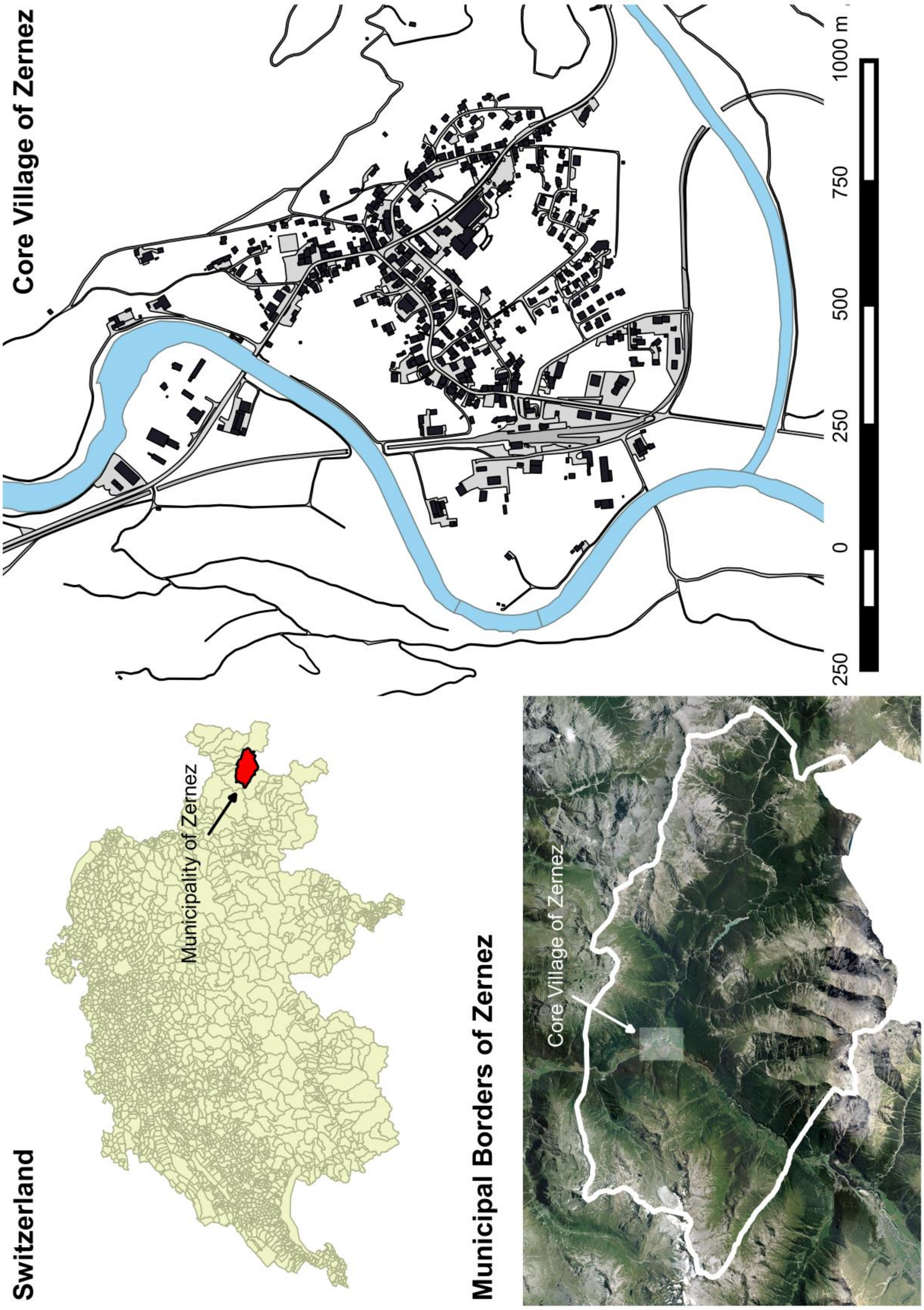
APPENDICES

Appendix A

Greenhouse Gas Emissions Quantification and Reduction Efforts in a Rural Municipality

A.1 MAP OF THE MUNICIPALITY OF ZERNEZ

The map below shows how the municipality of Zernez is situated within Switzerland and details the area of Zernez's core village. Map information is provided by swisstopo (2014) [1] and Wagner et al. (2015) [2].



A.2 Household Consumption Model of Saner et al. (2013) [3]

A.2.1 Model Description

Figure A.1 illustrates a simplified flow chart of the household consumption model according to Saner et al. (2013) [3]. The model estimates the space heating demand for each residential building by means of simplified heat balances according to the Swiss Standard SIA 380/1 (SIA 2009) [4] and based on the Swiss Federal Register of Buildings and Dwellings (BFS 2013) [5], building specific statistics (e.g. Wallbaum et al. 2010) [6] as well as on climatic data (METEOTEST 2012) [7]. Domestic hot water needs and electricity demand are covered by default values provided by the SIA 380/1.

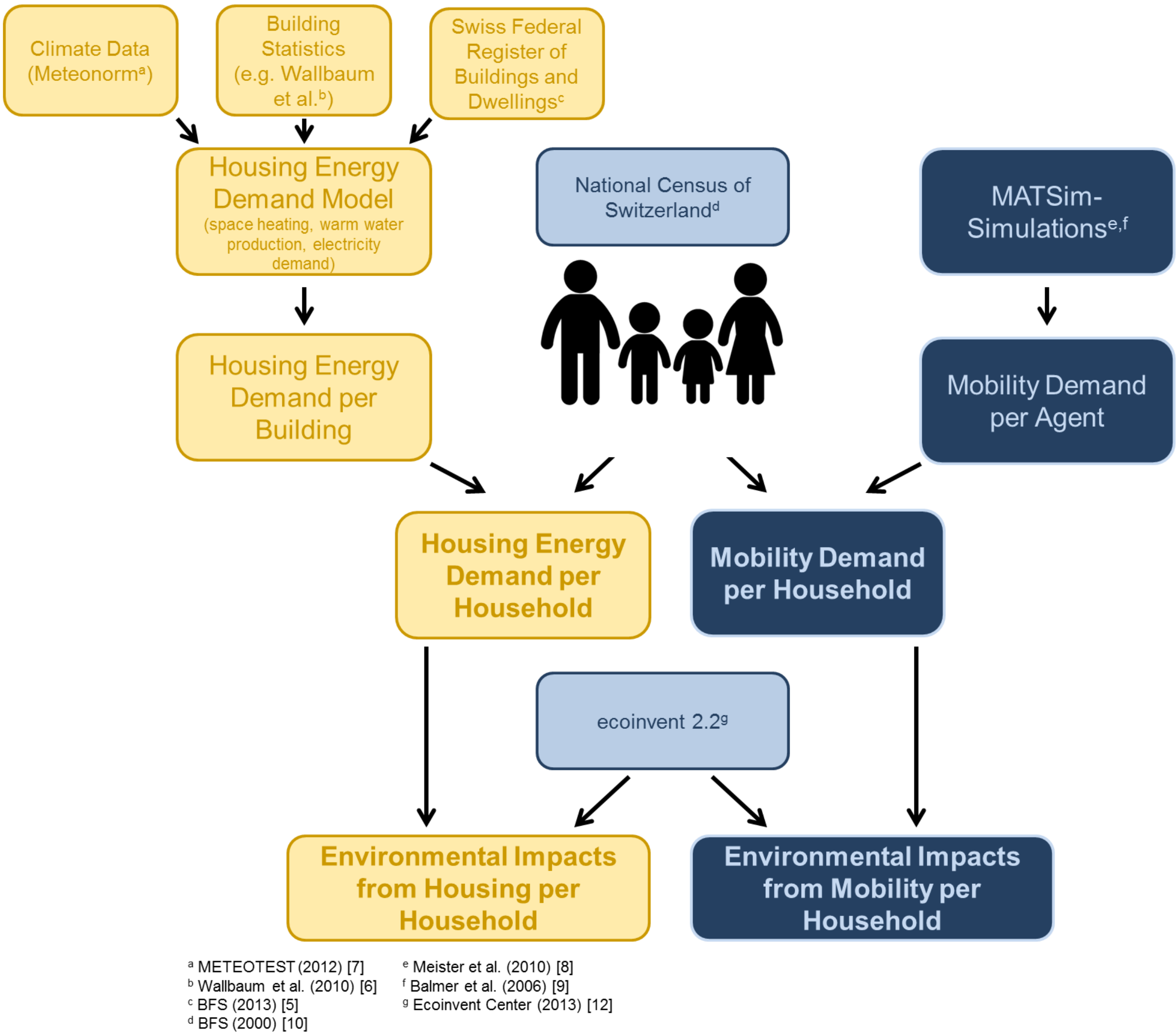


Figure A.1: Simplified flow chart of Saner et al. (2013)'s household consumption model [3].

The land-based mobility demand is assessed based on the simulation results of MATSim (Multi-Agent Transport Simulation) (Meister et al. 2010 [8]; Balmer et al. 2006 [9]). MATSim is an agent-based traffic model which simulates the mobility behavior of the Swiss population by means of data from the National Census (BFS 2000) [10] and the Mobility and Transport Microcensus of

Switzerland (BFS and ARE 2012) [11]. The MATSim-results provide information on chosen traffic modes and driven kilometers for each agent.

The agents are then matched with household members based on the personal characteristics provided by the National Census (BFS 2000) [10] in order to derive the household's mobility demand. Furthermore, the model assigns households to appropriate apartments in order to associate housing demand with a specific household.

Finally, environmental impacts of households are assessed based on the ecoinvent-database (ecoinvent Center) [12].

A.2.2 Results of Saner et al. (2013)'s Model [3]

By quantifying the environmental impacts for each household, this model is capable of capturing the variability of individual households' behavior. Figure A.2 reveals that in Zernez approximately 26% of the households with the largest impacts are responsible for about 50% of the greenhouse gas (GHG) emissions stemming from housing and mobility.

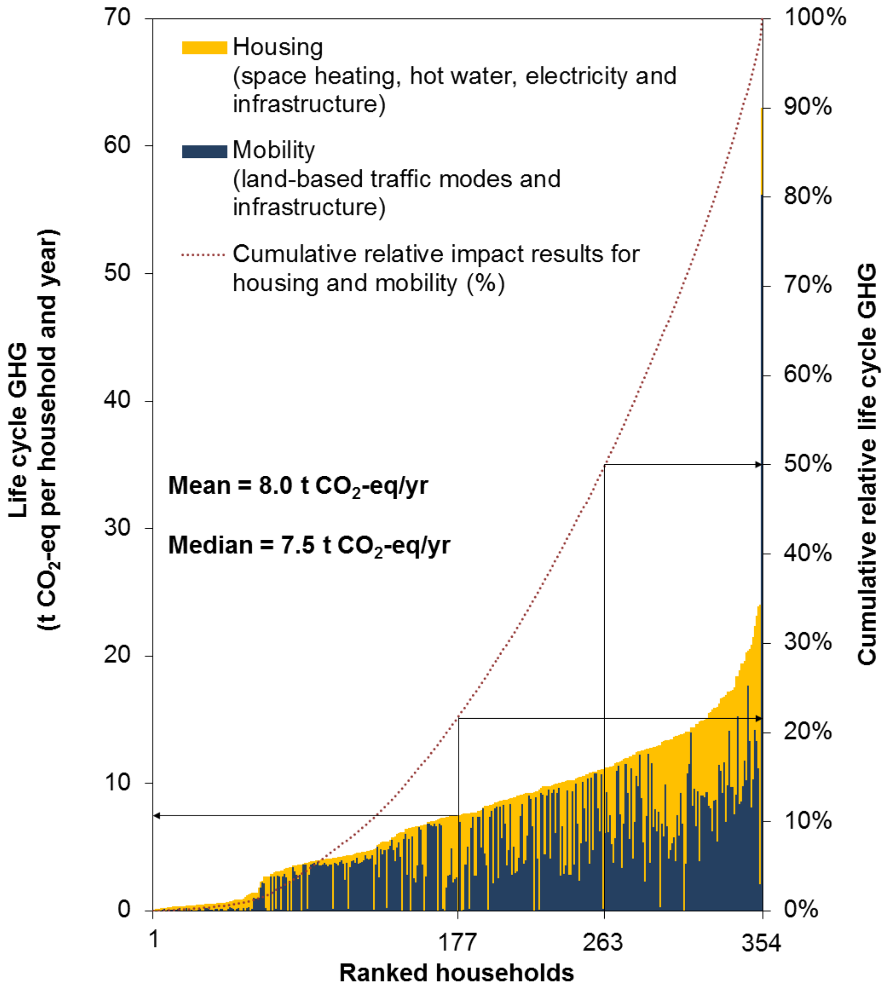


Figure A.2: Life cycle GHG emissions induced by the mobility and housing demand of individual households according to Saner et al. (2013)'s model [3].

Figure A.3 depicts separately the results of the application of Saner et al. (2013)'s model [3] to Zernez for housing and mobility.

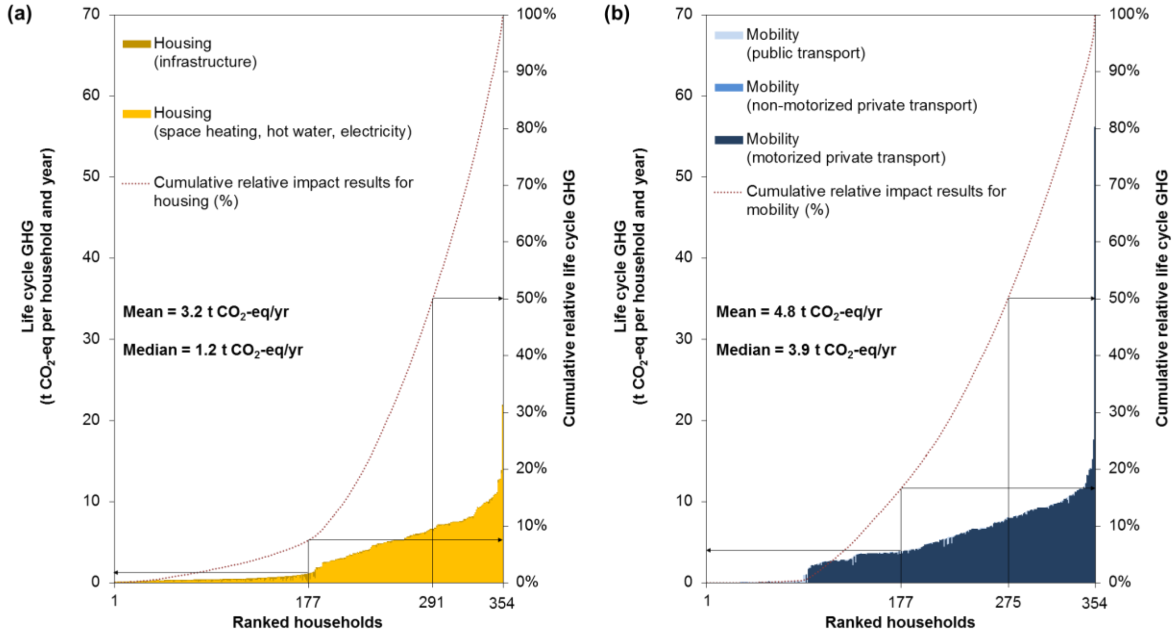


Figure A.3: Life cycle GHG emissions of individual households according to Saner et al. (2013)'s model [3] caused by the consumption areas of housing (a) and mobility (b).

In Figure A.4, the model results presented in Figure A.2 and Figure A.3 are normalized by the respective household size. These allows for a more intuitive comparison of the behavior of individual households.

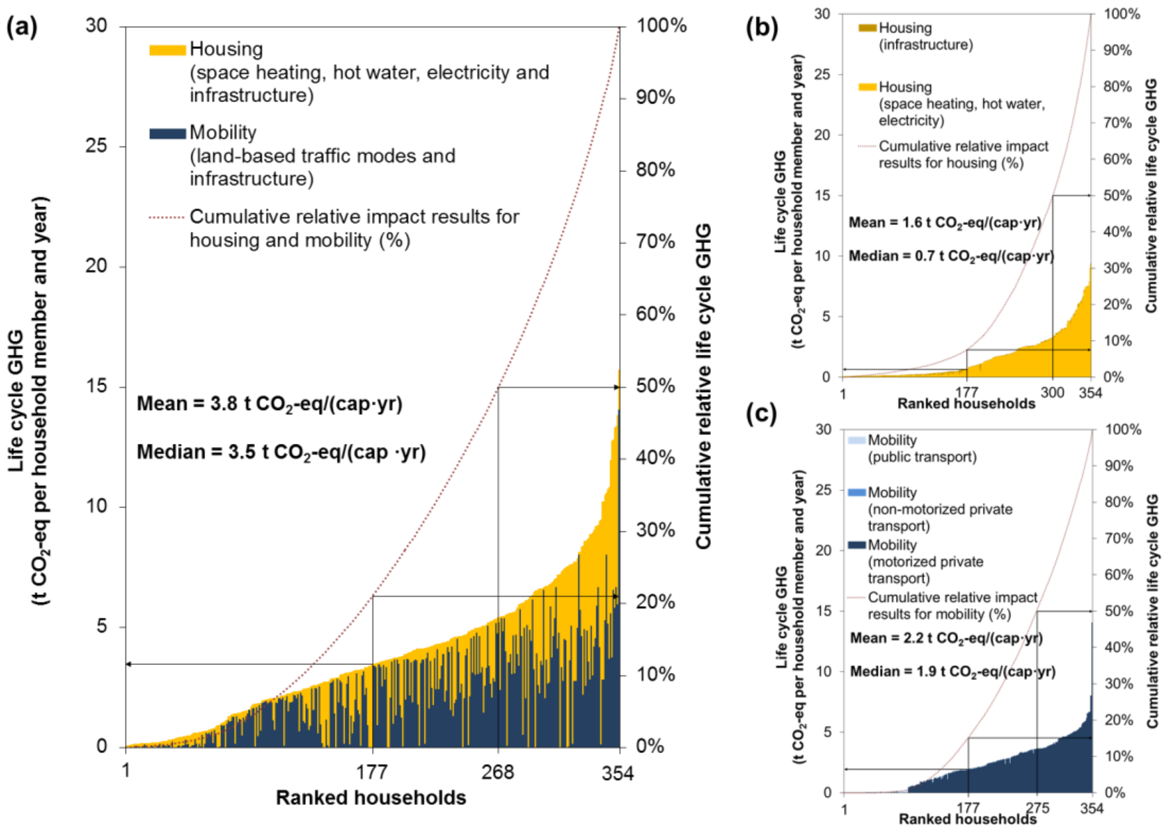
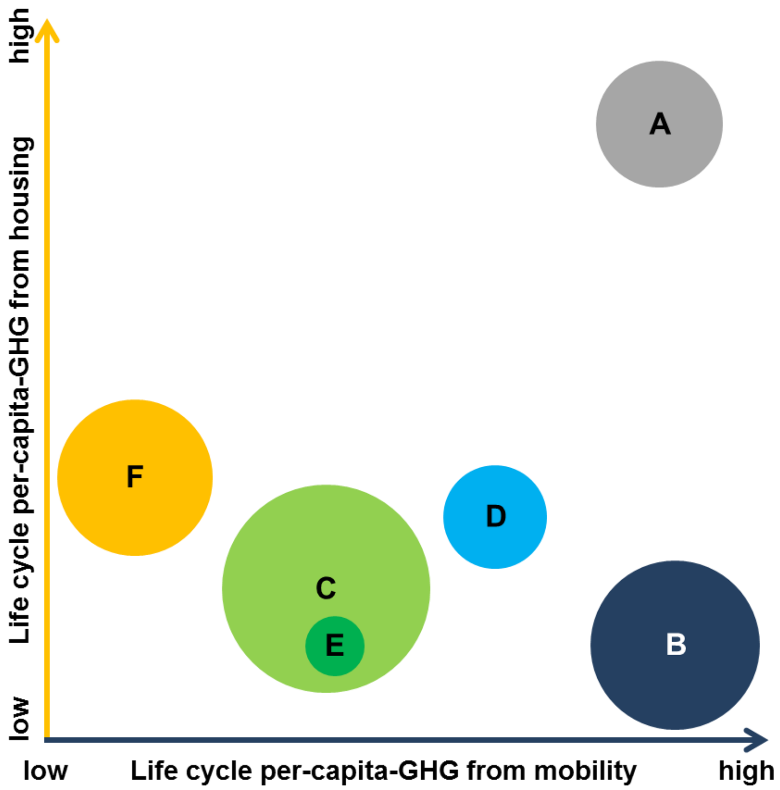


Figure A.4: Per-capita-life cycle GHG emissions of individual households induced by housing and mobility (a) and separately presented for housing (b) and mobility (c).

The model of Saner et al. (2013) [3] estimates the per capita consumption-induced mobility GHG to be 2.2 t CO_2-eq per year (Figure A.4). The model only considers land-based mobility. Subtracting air travels, mobility GHG estimated for Zernez and for Switzerland (Jungbluth et al. 2012 [13]) both amount to about 2.1 t CO_2-eq per person and year. This accordance of the study results and the model prediction gives confidence to use Saner et al. (2013)'s [3] approach for further analyses.

A similar conclusion can be drawn by the comparison of housing energy emissions. The model of Saner et al. (2013) [3] computes 1.6 t CO_2-eq per year as the average residential energy GHG emissions per person, while the presented carbon footprint study found 1.2 t CO_2-eq/(cap·yr) (cf. Figure 2.2 in Chapter 2).



Cluster	Type of Household	Housing Situation
A	Retired persons; Single or two-persons households	Apartments heated by fuel oil; Large per-capita living area
B	Just before retirement; Single or two-persons households	Apartments and buildings mostly heated by direct electric heating; Large per-capita living area
C	Young families	Mainly buildings heated by fuel oil; Small per-capita living area
D	Mainly middle-aged couples	Apartments and buildings mostly heated by direct electric heating; Medium per-capita living area
E	Small families, partly with adult children	Buildings heated by fuel oil, electricity or by wood-based technologies; Small per-capita living area
F	Mainly middle-aged single-person households	Mostly buildings or apartments heated by fuel oil or by electricity; Large per-capita living area

Figure A.5: **Simplified illustration of the outcomes of the cluster analysis applied to results of Saner et al. (2013) [3]'s model. The size of the circles is proportional to the number of households in a certain cluster. It has to be pointed out that the clusters A – F are only of an indicative nature suggesting how the households tend to behave in a certain cluster.**

We took further advantage of the model's ability to assess environmental impacts on a household level and conducted a cluster analysis on the basis of the model results. This cluster analysis was carried out analogously to the one presented in Saner et al. (2013) [3] and is illustrated in a simplified manner in Figure A.5.

According to the model-based cluster analysis, oil boilers and apartment area per person are the most important factors regarding housing GHG. In terms of mobility, the GHG emissions of clusters B and D are mainly induced by commuting. Cluster A shows high mobility GHG as well. This cluster exclusively comprises retired persons which cover their large demand for recreational traffic and for shopping by car.

It has to be pointed out that the model might deviate strongly from reality for some individual households, but it is capable of reproducing the statistical properties of the behavior of the entirety of all households.

A.3 Comparison of the Consumption-Based Carbon Footprint (CBF) of Zernez with the national CBF of Switzerland

The consumption-based carbon footprint (CBF) of Zernez is compared to the Swiss national CBF according to Jungbluth et al. (2012) [13] in Figure A.6. Note that in Figure A.6 the original categories used by Jungbluth et al. (2012) [13] were re-ordered and renamed in order to match the categories used in the present study. Even though the subdivision into different emission categories is similar in both studies, two major differences have to be kept in mind for an in-depth comparison. First, Jungbluth et al. (2012) [13] did not distinguish between *residential energy* and *drinking water system, wastewater treatment and waste management*. For Zernez, these two categories were thus summed up in order to form the category *housing*.

Second, Jungbluth et al. (2012) [13] had a separate category "hotel and restaurant industry". In Figure A.6 this category was fully assigned to the category *services* for the Swiss CBF. While vacations and leisure activities can also be found in the category *services* for Zernez, food eaten in a restaurant or hotel are allocated to the category *food* in the CBF of Zernez.

The biggest differences are for *housing*, which is discussed in Chapter 2, followed by *food* and *services*. Those are – just as mentioned in Chapter 2 – partly explainable by different methodological approaches, but also by the aforementioned differences in allocation of emissions to emission categories. In contrast to all other categories, these reasons play indeed an important role for *services*. However, there are more reasons for the differences. For instance, the age group of 18- to 24-aged persons is – compared to the national average – underrepresented in Zernez (AWT 2010 [14]). However, according to BFS (2011) [15] this age cohort covers the largest distances and often undertakes air travels. Therefore, emissions due to vacations are lower for Zernez compared to the Swiss average (see also section 2.2.3.7 (*Services: Leisure Activities and Communications*) in Chapter 2 for a description of how emissions from vacations were determined).

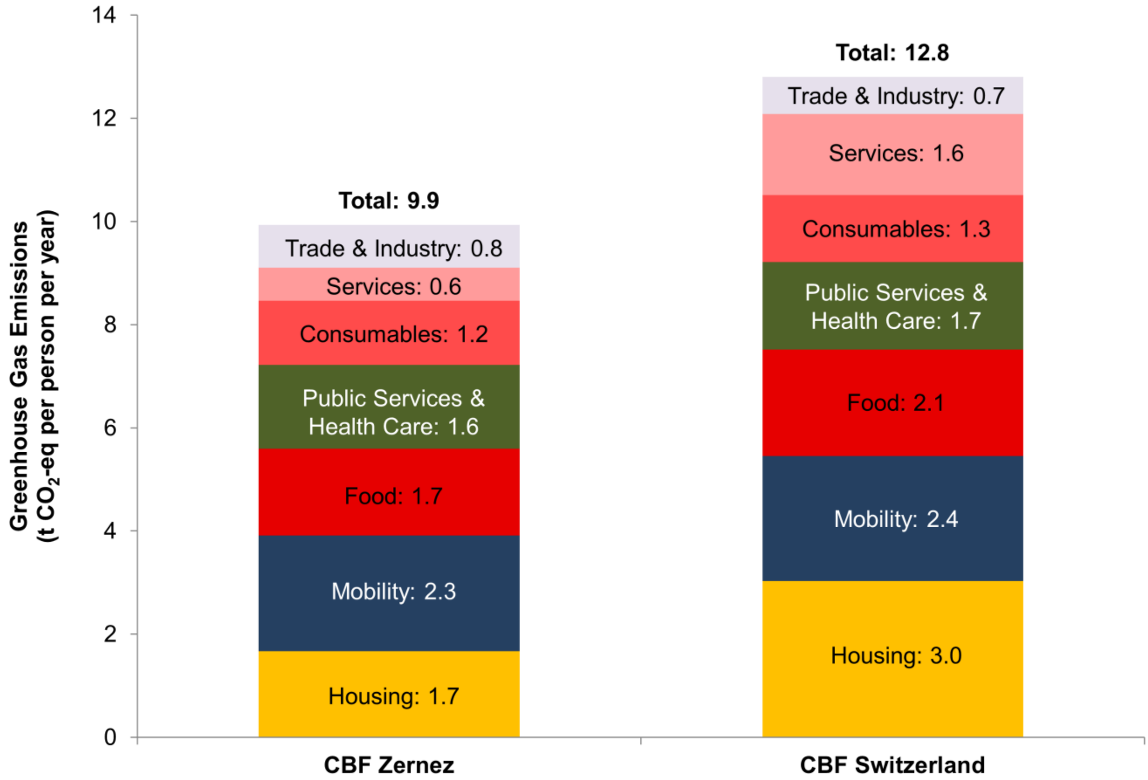


Figure A.6: **Comparison of the per-capita consumption-based carbon footprint (CBF) of Zernez with the national CBF of Switzerland according to Jungbluth et al. (2012)** [13]**. Please note that the difference in the category** ***consumables*** **is due to rounding differences.**

A.4 BUILDING ENERGY SYSTEM

A.4.1 Energy Flows of the Building Stock

Table A.1 and Figure A.7 are based on the building database (Wagner et al. 2015) [2] which was established during the project *Zernez Energia 2020* (see also section A.5 *Description of Bottom-Up Data from Wagner et al.* [2] in this Appendix) and show the energy flows within the building stock of Zernez in detail.

Table A.1: Energy balance for the building stock of Zernez based on Wagner et al. (2015) [2]**. An illustration of these energy flows is given by a Sankey-diagram in Figure A.7. [WWTP = Wastewater Treatment Plant].**

Imports	**kWh/a**	**Local Production**	**Input (kWh/a)**	**Output (kWh/a)**	**Local Demand**	**kWh/a**
Run-of-River Power Plant	8,702,003	Electric Heating	2,848,051	2,705,648	Lighting & Electrical Appliances	7,328,513
Fuel Oil	6,981,390	Hydro Power Plant	1,888,889	1,700,000		
Wood Chips	2,024,793	Air Source Heat Pumps	20,370	20,370	Space Heating & Hotwater Production	14,620,093
Firewood	1,085,067	Ground Source Heat Pumps	874,596	874,596		
Total:	**18,793,253**	Oil Boilers	6,981,390	5,934,182	**Total:**	**21,948,606**
		Wood Chips Furnaces	4,912,062	3,291,860		
Local Resources	**kWh/a**	Wood Log Heating	2,241,400	1,568,980	**Exports**	**kWh/a**
Hydropower	1,888,889	Solar Thermal Systems	105,094	36,783	Photovoltaic Systems	11,000
Solar Power	178,428	Photovoltaic Systems	73,333	11,000	Biogas Plant / WWTP	32,640
Ambient Heat	669,527	Biogas Plant / WWTP	229,914	220,314	**Total:**	**43,640**
Wood Chips	2,887,269	**Total:**	**20,175,100**	**16,363,733**		
Firewood	1,156,333					
Organic Waste	229,914					
Total:	**7,010,359**					

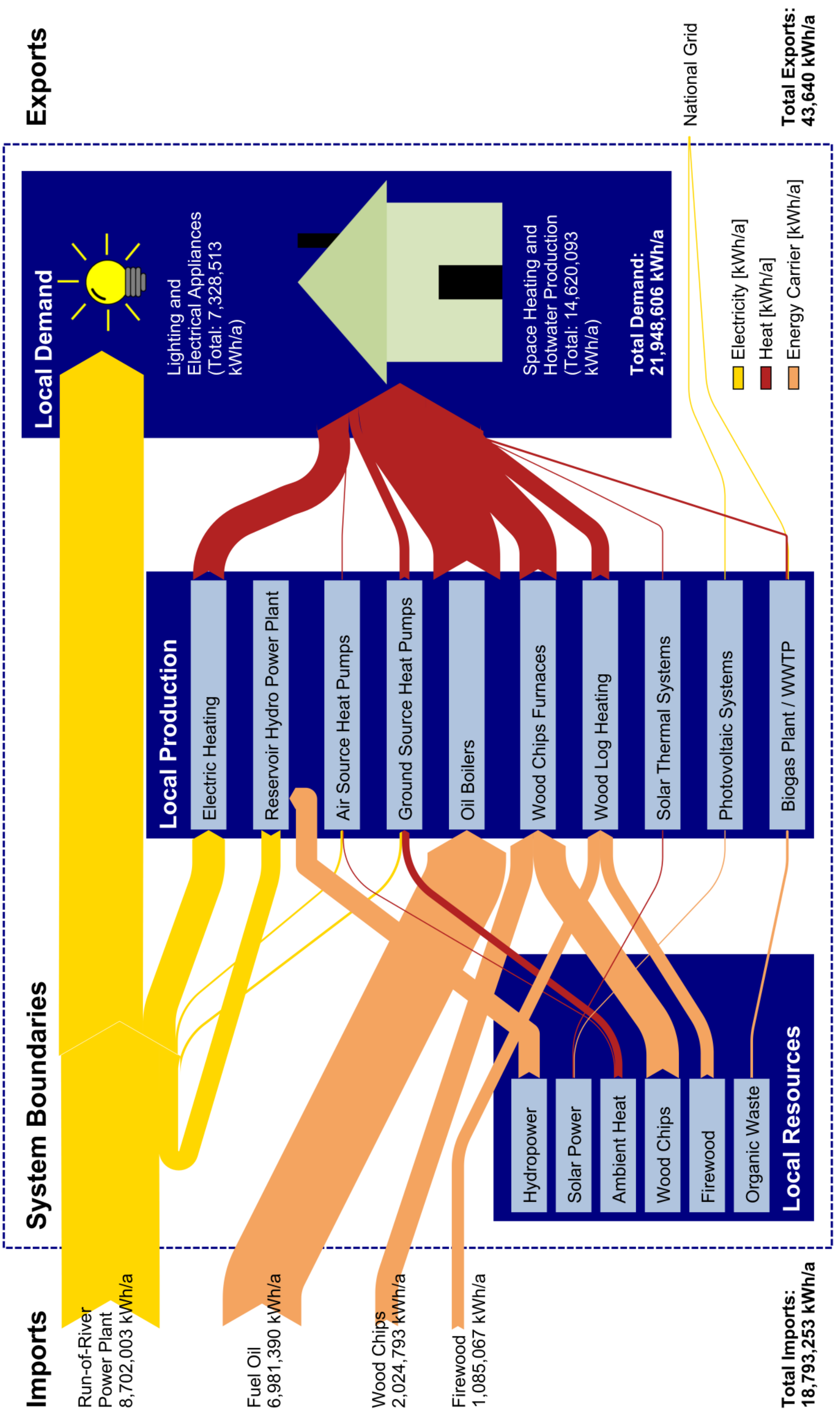


Figure A.7: Sankey-diagram illustrating the energy flows within Zernez's building stock based on Wagner et al. (2015) [2]. Detailed information on the energy flows is provided in Table A.1. [WWTP = Wastewater Treatment Plant].

A.4.2 Derivation of the Trans-Boundary Community-Wide Infrastructure Footprint of the Building Energy System

A "trans-boundary community-wide infrastructure footprint" (CIF) provides urban planners and infrastructure managers with a holistic view by considering local activities (similar to purely geographical accounting (PGA)) but from a life cycle perspective (similar to CBF). This approach is useful for infrastructure, which is not only constructed to serve residents but also local businesses and industries (Chavez and Ramaswami 2013) [16]. Having identified the building stock as an important area of action for the municipality of Zernez, a CIF of the building energy system constitutes thus an important planning basis.

In order to deduce the CIF for the building energy system for Zernez, embodied, direct and exported emissions of the *residential energy* category as well as of all building-related activities of the categories *public services, education and health care*, *services: leisure activities and communications*, and *local trade and industry* were summed up (cf. also corresponding sections Chapter 2). Finally, re-exported emissions induced by the operation of buildings were added. The resulting CIF is presented in the bottom right corner of Figure 2.3 in Chapter 2.

A.5 DESCRIPTION OF BOTTOM-UP DATA FROM WAGNER ET AL. (2015) [2]

The present study is largely based on the bottom-up data which is described and presented in aggregated form in Wagner et al. (2015) [2]. From this data collection, we specifically used the following five datasets: the building database, electricity bills, operation information of the district heating network, waste statistics, and the data from the forestry administration. The available data and the attributes of these datasets will be explained in the following.

The building database was the most important data source. This database comprises all buildings in Zernez and was established based on interviews, surveys, electricity bills, and operation information of the district heating network. Table A.2 provides an overview of the attributes available in the building database.

Even though the operation information of the district heating network and the electricity bills were integrated in the building database, we still needed the original information from these two sources for the quantification of some activities in Zernez. For instance, the distinction of energy consumption by residential and different commercial activities could only be retrieved from the electricity bills and the operation information of the district heating network, but not from the building database which supplies data on a building level. The available operation information of the district heating network and the attributes of the electricity bills are described in Table A.3 and Table A.4. The final energy consumption of commercial activities was further used to set up a census of enterprises (see A.6 *Census of Local Enterprises* in this Appendix).

Furthermore, available statistics pertaining to the municipal waste management is described in Table A.5, while data from the forestry administration is characterized in Table A.6.

Table A.2: **Original and translated attributes available in the building database according to Wagner et al. (2015) [2]. One building was randomly chosen and is presented in this table in order to illustrate the entries of the database. Due to data privacy protection some attributes were replaced by "xxx".**

Attributes (original name)	Attributes (translated)	Building XXX (example)	Building …
ID	Identification Number of Building	xxx	…
Bemerkung_Allgemein	Remarks	-	…
Nutzung	Use	Wohnen	…
Adresse	Address (Street)	xxx	…
Bezeichnung	Name	xxx	…
Gebäude_Nummer	Building Number	xxx	…
Nummer	Address (Number)	xxx	…
Bewohnt	Occupied [yes/no]	ja	…
Anzahl_Einwohner	Number of Inhabitants	7	…
Anzahl_Wohneinheiten	Number of Apartments	3	…
Stockwerke_ohne_UG_	Number of Floors (w/o Basement Floor)	1.5	…
UG_Untergeschoss_	Basement Floor [0/1]	1	…
BGF_Bruttogeschossfläche__m2____0_10_Konstruktion_und_ohne_UG_	Gross Floor Area (w/o Basement Floor) in m^2	540	…
EBF_Energiebezugsfläche_m2_	Energy Reference Area in m^2	324	…
System_für_Warmwasser_WW_	Type of Domestic Hot Water System	Boiler - Solarthermie + Elektrisch + Holz	…
Lage_des_WW_Systems	Position of Hot Water System	zentral - beim Heizsystem	…
Einbaujahr_Heizsystem	Year of Installation of Heating System	1981	…
Heizsystem	Type of Heating System	Elektrisch oder Elektrisch/Holz	…
Wärme_Abgabesystem_zum_Beispiel_Heizkörper_Fussbodenheizung_	Type of Heat Distribution System	Elektroheizung	…
Verbrauch_Öl_Liter_a_	Amount of Fuel Oil Used in liters/a	-	…
Verbrauch_Holz_Pellets_m3_a_	Amount of Wood Chips Used in m^3/a	10	…
Verbrauch_Holz	Amount of Log Wood Used in m^3/a	-	…
Produktion_Wärmepumpe_Falls_ablesbar_in_kWh_a_	Heat Production by Heat Pumps in kWh/a	-	…
Zusätzliche_Wärmeproduktion_mit_Kamin_Ofen	Additional Heat Production (e.g. Solar Thermal Systems, Wood Log Heating)	Solar, Ofen	…
Winter_Hochverbrauch_Strom__kWh_a_	Electricity Consumption During Wintertime in kWh/a	44212	…
Sommer_Niederverbrauch_Strom__kWh_a_	Electricity Consumption During Summertime in kWh/a	10906	…
Gesamt_Stromverbrauch_10_2010_10_2011_kWh_a_	Total Electricity Consumption	55118	…
Eigene_Stromproduktion_Photovoltaik_Blockheizkraftwerk_Biogas_in_kWh_a_	Electricity Production (e.g. Photovoltaic Cells, CHP at Biogas Plant)	Keine	…
Verbrauch_Wärmepumpe_Falls_ablesbar_in_kWh_a_	Electricity Consumption by Heat Pumps in kWh/a	-	…
Verbrauch_Gas_m3_a_	Gas Consumption in m^3/a	-	…
Lüftung_WRG_Wärmerückgewinnung_	Type of Ventilation System	Keine	…
Baujahr_1872_Brand_	Year of Construction	1981	…
Denkmalschutz_Erhaltungsbereich	Class of Monument Protection	Nein	…
Zustand	Condition (of the Built Volumes)	gut	…
Bauweise	Construction Type / Architectural Style	Massivbauweise	…
Traufhöhe_m_	Eaves Height in m	5+7	…
Fassadekonstruktion	Type/Materials of Facades	Verputzt - Mauerwerk	…
Fassaden_Dämmung_cm_	Insulation of Facades in cm	10	…
Fassadenfarbe	Color of Facades	rose	…
Letzte_Sanierung_Jahr_	Last Refurbishment	2010	…
Sanierte_Bauteile	Refurbished Building Components	Heizsystem	…
Fensteranteil_Prozent_	Share of Window Area in %	15	…
Fensterart	Type of Windows	zweifach verglast	…
Fensterrahmen	Type of Window Frame	Holz	…
Anschluss_Strom	Connection to Electricity [yes/no]	ja	…
Anschluss_Wasser	Connection to Water [yes/no]	ja	…
Anschluss_Fernwärme	Connection to District Heating Network [yes/no]	nein	…
Anschluss_Gas	Connection to Gas [yes/no]	nein	…
Dachform	Type of Roof	Giebeldach	…
Dacheindeckung	Roof Covering	Eternit - Schindel	…
Fläche_Solarthermie_WW_m2_	Area of Solar Thermal System in m^2	-	…
Verbrauch_Fernwärme_für_Zeitraum_04_2010_04_2011__kWh_a_	Consumption of Heat from District Heating Network in kWh/a	-	…
Dachneigung_°_geschätzt_	Roof Inclination (estimated)	30	…
Dachausrichtung_Sonne	Roof Orientation to Sun	Südost	…
Gesammte_Dachfläche_nach_Himmelsrichtung_m2_	Total Roof Area Depending on Orientation in m^2	100	…
Gebäudetyp	Building Type	Wohnhaus	…
Anteil_Konstruktion	Share of Construction in Gross Floor Area	17	…
Nettogeschossflaeche	Net Floor Area		…
Eigentümer	Owner	xxx	…
Adresse_Eigentümer	Address of Owner	xxx	…
Nummer_Eigentümer	Identification Number of Owner	31D	…
Wohnort_Eigentümer	Owner's Place of Residence	xxx	…
GGF_Gebäudegrundfläche_Gesamte_Parzelle__m2_	Total Area of Building Footprints in the Parcel in m^2	256	…
Parzellen_Nummer	Identification Number of Parcel	xxx	…
Parzellenfläche_m2_	Area of Parcel in m^2	2081	…
Bauzone	Construction Zone	W3 - Wohnzone 3	…
Letzte_Bearbeitung	Last Modification (of Entry)	xxx	…
Letzte_Aktualisierung	Last Update (of Entry)	xxx	…

Table A.3: Original and translated attributes of the operation data of the district heating network according to Wagner et al. (2015) [2].

Supply of Wood Chips

Attributes (original name)	Attributes (translated)
data quint	Date
furnitur	Name of Wood Chips Supplier
import	Costs
quantità	Supplied Amount of Wood Chips

Supply of Consumers with Heat

Attributes (original name)	Attributes (translated)
nom	Name of Consumer
kWh	Supplied Heat in kWh/a
tariffa	Price per kWh
total	Total Price

Table A.4: Original and translated attributes of the available electricity bills according to Wagner et al. (2015) [2].

Attributes (original name)	Attributes (translated)
Abo-Nr	Identification Number
Name	Name
Objekt	Object
Strasse	Address (Street)
Ort	Location
Hoch-Verbrauch	Consumption of Electricity (High-Rate Tariff)
Nieder-Verbrauch	Consumption of Electricity (Low-Rate Tariff)
Total-Verbrauch	Consumption of Electricity (Total)

Table A.5: Original and translated attributes of the municipal solid waste statistics according to Wagner et al. (2015) [2].

Attributes (original name)	Attributes (translated)
Gemeinde	Municipality
Jahr	Year
Gewicht Kehricht (to)	Weight of Municipal Solid Waste in Tons
Sammelzeit (min)	Collection Time in Minutes
km	Distance Driven for Collection in km
Karton (to)	Cardboard in Tons
Papier (to)	Waste Paper in Tons
Glas (to)	Waste Glass in Tons
Totalgewicht (to)	Total Weight in Tons

Table A.6: Original and translated attributes of the statistics provided by the forestry administration according to Wagner et al. (2015) [2].

Attributes (original name)	Attributes (translated)
Waldfläche (nur Hochwald)	Forest Area (only timber forest)
Zuwachs Brutto (m^3/ha/Jahr) (Waldinventur 2010/11)	Gross Growth Rate in m^3/ha/yr
Zuwachs Brutto (m^3/Jahr)	Gross Growth Rate in m^3/yr
Vorratsabbau	Decrease in Stock
Zuwachs Schaftderbholz in Rinde	Gain of Merchantable Wood Under Bark
Verkauf	Sales (in m^3)
Eigengebrauch	Wood for Own Use (in m^3)
Liegengelassen	Left in Forest (in m^3)
Total	Total (in m^3)
Verbrauch an Diesel	Amount of Diesel Used
Verbrauch an Petrol	Amount of Petrol Used

A.6 CENSUS OF LOCAL ENTERPRISES

In the scope of the present study, a detailed census of all enterprises which are located within the geographical system boundaries was compiled. This census encompasses the following information:

- Name of company
- Building identification number (to match with the building database (see Table A.2))
- Electricity consumption (see also Table A.4)
- Fuel oil consumption (see also Table A.2)
- Heat bought from the district heating network (see also Table A.3)

Table A.7 provides an overview in terms of number and types of different enterprises in Zernez.

Table A.7: Statistical overview of different enterprise types in Zernez.

Type of Enterprise	Number of Companies
Agricultural Enterprises	20
Forestry	1
Mining and Quarrying	1
Manufacturing of Wood Products, Metal Products, Furniture and Other Products	4
Electricity Supply	1
Wastewater Treatment	1
Waste Management	1
Construction Companies (incl. Specialized Companies and Architectural/Engineering Enterprises)	16
Trade and Repair of Motor Vehicles	2
Wholesale Trade	3
Retail Trade	10
Transport Companies and Transport-Related Companies (incl. Service Stations)	5
Postal and Courier Activities	3
Accommodation (e.g. Hotels)	13
Restaurants	2
Financial Services	3
Travel Agencies	2
Public Administration	5
Education	2
Human Health Activities	2
Cultural Activities	2
Churches	1
Other Services	3

A.7 Detailed List of Modeled Activities for the Assessment of Greenhouse Gas Emissions

Category	Process Name According to Respective Database	Database	Quantity[a]	Unit	Adjusted[b]	Remarks
Food Consumption	Rice	Saner (2013) [17], Saner et al. (2016) [18]	3.8	kg/(pers·yr)		
Food Consumption	Pasta	Saner (2013) [17], Saner et al. (2016) [18]	9.9	kg/(pers·yr)		
Food Consumption	Bread	Saner (2013) [17], Saner et al. (2016) [18]	23.5	kg/(pers·yr)		
Food Consumption	Flour	Saner (2013) [17], Saner et al. (2016) [18]	4.7	kg/(pers·yr)		
Food Consumption	Other Flours, Semolinas and Cereal Grains	Saner (2013) [17], Saner et al. (2016) [18]	2.7	kg/(pers·yr)		
Food Consumption	Other Products with Cereals	Saner (2013) [17], Saner et al. (2016) [18]	6.5	kg/(pers·yr)		
Food Consumption	Fresh Potatoes, not steamed	Saner (2013) [17], Saner et al. (2016) [18]	19.0	kg/(pers·yr)		
Food Consumption	Products made out of Potatoes and Other Tuber Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	4.5	kg/(pers·yr)		
Food Consumption	Lettuce and Leaf Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	8.6	kg/(pers·yr)		
Food Consumption	Stem Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	3.0	kg/(pers·yr)		
Food Consumption	Soy	Saner (2013) [17], Saner et al. (2016) [18]	3.4	kg/(pers·yr)		
Food Consumption	Cabbage Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	4.4	kg/(pers·yr)		
Food Consumption	Tomatoes	Saner (2013) [17], Saner et al. (2016) [18]	7.6	kg/(pers·yr)		
Food Consumption	Beans and Peas	Saner (2013) [17], Saner et al. (2016) [18]	2.1	kg/(pers·yr)		
Food Consumption	Other Fruiting Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	8.2	kg/(pers·yr)		
Food Consumption	Onions	Saner (2013) [17], Saner et al. (2016) [18]	3.4	kg/(pers·yr)		
Food Consumption	Garlic	Saner (2013) [17], Saner et al. (2016) [18]	0.4	kg/(pers·yr)		
Food Consumption	Carrots and Other Root Vegetables	Saner (2013) [17], Saner et al. (2016) [18]	11.9	kg/(pers·yr)		
Food Consumption	Fresh Mushrooms	Saner (2013) [17], Saner et al. (2016) [18]	0.9	kg/(pers·yr)		
Food Consumption	Dried Vegetables and Mushrooms	Saner (2013) [17], Saner et al. (2016) [18]	0.4	kg/(pers·yr)		
Food Consumption	Vegetables and Mushrooms in Cans or differently prepared	Saner (2013) [17], Saner et al. (2016) [18]	6.8	kg/(pers·yr)		
Food Consumption	Nuts	Saner (2013) [17], Saner et al. (2016) [18]	2.4	kg/(pers·yr)		
Food Consumption	Vegetable and Fruit Juices	Saner (2013) [17], Saner et al. (2016) [18]	27.5	kg/(pers·yr)		
Food Consumption	Lemons	Saner (2013) [17], Saner et al. (2016) [18]	1.6	kg/(pers·yr)		
Food Consumption	Oranges and Other Citrus Fruits	Saner (2013) [17], Saner et al. (2016) [18]	11.3	kg/(pers·yr)		
Food Consumption	Bananas	Saner (2013) [17], Saner et al. (2016) [18]	8.4	kg/(pers·yr)		
Food Consumption	Apples	Saner (2013) [17], Saner et al. (2016) [18]	13.0	kg/(pers·yr)		
Food Consumption	Pears and Quinces	Saner (2013) [17], Saner et al. (2016) [18]	3.2	kg/(pers·yr)		
Food Consumption	Stone Fruits	Saner (2013) [17], Saner et al. (2016) [18]	8.3	kg/(pers·yr)		
Food Consumption	Berries	Saner (2013) [17], Saner et al. (2016) [18]	3.0	kg/(pers·yr)		
Food Consumption	Grapes	Saner (2013) [17], Saner et al. (2016) [18]	3.5	kg/(pers·yr)		
Food Consumption	Melons and Water Melons	Saner (2013) [17], Saner et al. (2016) [18]	3.8	kg/(pers·yr)		
Food Consumption	Other Exotic Fruits	Saner (2013) [17], Saner et al. (2016) [18]	4.1	kg/(pers·yr)		
Food Consumption	Other Dried Fruits	Saner (2013) [17], Saner et al. (2016) [18]	1.0	kg/(pers·yr)		
Food Consumption	Canned Fruits	Saner (2013) [17], Saner et al. (2016) [18]	1.4	kg/(pers·yr)		
Food Consumption	Margarine	Saner (2013) [17], Saner et al. (2016) [18]	0.9	kg/(pers·yr)		
Food Consumption	Other Vegetable Oils	Saner (2013) [17], Saner et al. (2016) [18]	0.2	kg/(pers·yr)		
Food Consumption	Olive Oil	Saner (2013) [17], Saner et al. (2016) [18]	1.6	kg/(pers·yr)		
Food Consumption	Other Vegetable Oils and Animal Fats	Saner (2013) [17], Saner et al. (2016) [18]	2.6	kg/(pers·yr)		
Food Consumption	Beef	Saner (2013) [17], Saner et al. (2016) [18]	4.9	kg/(pers·yr)		
Food Consumption	Veal	Saner (2013) [17], Saner et al. (2016) [18]	1.5	kg/(pers·yr)		
Food Consumption	Pork	Saner (2013) [17], Saner et al. (2016) [18]	5.4	kg/(pers·yr)		
Food Consumption	Mutton and Goatmeat	Saner (2013) [17], Saner et al. (2016) [18]	0.6	kg/(pers·yr)		
Food Consumption	Horsemeat	Saner (2013) [17], Saner et al. (2016) [18]	0.2	kg/(pers·yr)		
Food Consumption	Poultry	Saner (2013) [17], Saner et al. (2016) [18]	7.2	kg/(pers·yr)		
Food Consumption	Rabbit Meat and Game	Saner (2013) [17], Saner et al. (2016) [18]	0.4	kg/(pers·yr)		
Food Consumption	Organ Parts	Saner (2013) [17], Saner et al. (2016) [18]	2.3	kg/(pers·yr)		
Food Consumption	Sausages, Sausage Products and Pastries	Saner (2013) [17], Saner et al. (2016) [18]	8.2	kg/(pers·yr)		
Food Consumption	Ham, Bacon and other Salted or Smoked Meat	Saner (2013) [17], Saner et al. (2016) [18]	4.1	kg/(pers·yr)		
Food Consumption	Other Cooked, Dried, Salted or Smoked Meat	Saner (2013) [17], Saner et al. (2016) [18]	0.6	kg/(pers·yr)		
Food Consumption	Canned Meats and Products made out of Meat	Saner (2013) [17], Saner et al. (2016) [18]	0.4	kg/(pers·yr)		
Food Consumption	Full Cream Milk	Saner (2013) [17], Saner et al. (2016) [18]	38.5	kg/(pers·yr)		
Food Consumption	Milk and Skimmed Milk	Saner (2013) [17], Saner et al. (2016) [18]	23.7	kg/(pers·yr)		
Food Consumption	Yoghurt	Saner (2013) [17], Saner et al. (2016) [18]	16.5	kg/(pers·yr)		
Food Consumption	Cheese	Saner (2013) [17], Saner et al. (2016) [18]	13.6	kg/(pers·yr)		
Food Consumption	Butter	Saner (2013) [17], Saner et al. (2016) [18]	3.3	kg/(pers·yr)		

Category	Process Name According to Respective Database	Database	Quantity[a]	Unit	Adjusted[b]	Remarks
Food Consumption	Cream	Saner (2013) [17], Saner et al. (2016) [18]	8.0	kg/(pers·yr)		
Food Consumption	Curd	Saner (2013) [17], Saner et al. (2016) [18]	1.7	kg/(pers·yr)		
Food Consumption	Other Dairy Products and Milk Substitute Drinks	Saner (2013) [17], Saner et al. (2016) [18]	6.6	kg/(pers·yr)		
Food Consumption	Eggs	Saner (2013) [17], Saner et al. (2016) [18]	5.1	kg/(pers·yr)		
Food Consumption	Fish	Saner (2013) [17], Saner et al. (2016) [18]	5.3	kg/(pers·yr)		
Food Consumption	Sugar	Saner (2013) [17], Saner et al. (2016) [18]	5.2	kg/(pers·yr)		
Food Consumption	Jam	Saner (2013) [17], Saner et al. (2016) [18]	2.4	kg/(pers·yr)		
Food Consumption	Honey	Saner (2013) [17], Saner et al. (2016) [18]	0.8	kg/(pers·yr)		
Food Consumption	Chocolate	Saner (2013) [17], Saner et al. (2016) [18]	4.6	kg/(pers·yr)		
Food Consumption	Ice Cream	Saner (2013) [17], Saner et al. (2016) [18]	3.2	kg/(pers·yr)		
Food Consumption	Mineral Water	Saner (2013) [17], Saner et al. (2016) [18]	77.4	kg/(pers·yr)		
Food Consumption	Soft Drinks	Saner (2013) [17], Saner et al. (2016) [18]	54.6	kg/(pers·yr)		
Food Consumption	Syrup for Beverages	Saner (2013) [17], Saner et al. (2016) [18]	3.3	kg/(pers·yr)		
Food Consumption	Wines	Saner (2013) [17], Saner et al. (2016) [18]	22.7	kg/(pers·yr)		
Food Consumption	Spirits	Saner (2013) [17], Saner et al. (2016) [18]	0.9	kg/(pers·yr)		
Food Consumption	Alcoholic and Nonalcoholic Liqueurs, Aperitifs with Liqueur	Saner (2013) [17], Saner et al. (2016) [18]	0.5	kg/(pers·yr)		
Food Consumption	Alcoholic and Nonalcoholic Beer	Saner (2013) [17], Saner et al. (2016) [18]	17.7	kg/(pers·yr)		
Food Consumption	Pure and Ground Coffee	Saner (2013) [17], Saner et al. (2016) [18]	4.3	kg/(pers·yr)		
Food Consumption	Powdered Coffee and Coffee Surrogates	Saner (2013) [17], Saner et al. (2016) [18]	0.4	kg/(pers·yr)		
Food Consumption	Tea, Herbal Tea and Surrogates	Saner (2013) [17], Saner et al. (2016) [18]	0.5	kg/(pers·yr)		
Food Consumption	Cocoa Drinks	Saner (2013) [17], Saner et al. (2016) [18]	0.7	kg/(pers·yr)		
Food Consumption	Transport, lorry 20-28t, fleet average/CH U	ecoinvent Center (2013) [12]	316.3	tkm		
Residential Energy	Heat, at flat plate collector, one-family house, for hot water/CH U	ecoinvent Center (2013) [12]	36783	kWh		
Residential Energy	Electricity, at cogen with biogas engine, agricultural covered, alloc. exergy/CH U	ecoinvent Center (2013) [12]	32640	kWh		Adjusted according to interviews with operators and operating data (Grass 2014) [19]
Residential Energy	Heat, at cogen with biogas engine, agricultural covered, allocation exergy/CH U	ecoinvent Center (2013) [12]	53760	kWh		Adjusted according to interviews with operators and operating data (Grass 2014) [19]
Residential Energy	Electricity, production mix photovoltaic, at plant/CH U	ecoinvent Center (2013) [12]	11000	kWh		
Residential Energy	Heat, at air-water heat pump 10kW/CH U	ecoinvent Center (2013) [12]	23086	kWh		
Mobility	Transport, bicycle/CH U	ecoinvent Center (2013) [12]	215.5	pkm		Consumption-based footprint
Mobility	Transport, passenger car/CH U	ecoinvent Center (2013) [12]	10091.5	pkm		Consumption-based footprint
Mobility	Transport, metropolitan train, SBB mix/CH U	ecoinvent Center (2013) [12]	457.5	pkm		Consumption-based footprint
Mobility	Transport, long-distance train, SBB mix/CH U	ecoinvent Center (2013) [12]	1129.2	pkm		Consumption-based footprint
Mobility	Transport, regional train, SBB mix/CH U	ecoinvent Center (2013) [12]	645.0	pkm		Consumption-based footprint
Mobility	Transport, regular bus/CH U	ecoinvent Center (2013) [12]	446.3	pkm		Consumption-based footprint
Mobility	Transport, coach/CH U	ecoinvent Center (2013) [12]	245.1	pkm		Consumption-based footprint
Mobility	Transport, scooter/CH U	ecoinvent Center (2013) [12]	123.5	pkm		Consumption-based footprint
Mobility	Transport, aircraft, passenger, intercontinental/RER U	ecoinvent Center (2013) [12]	1776.8	pkm		Consumption-based footprint
Mobility	Transport, bicycle/CH U	ecoinvent Center (2013) [12]	592.2	pkm/d		Purely geographic accounting
Mobility	Transport, coach/CH U	ecoinvent Center (2013) [12]	1450.5	pkm/d		Purely geographic accounting
Mobility	Transport, passenger car/CH U	ecoinvent Center (2013) [12]	12444.0	pkm/d		Purely geographic accounting
Mobility	Transport, scooter/CH U	ecoinvent Center (2013) [12]	327.0	pkm/d		Purely geographic accounting
Mobility	Transport, lorry 16-32t, EURO4/RER U	ecoinvent Center (2013) [12]	2197.0	pkm/d		Purely geographic accounting
Drinking Water & Wastewater System & Waste Management	Disposal, municipal solid waste, 22.9% water, to municipal incineration/CH U	ecoinvent Center (2013) [12]	264.52	t		
Drinking Water & Wastewater System & Waste Management	Corrugated board, recycling fibre, single wall, at plant/CH U	ecoinvent Center (2013) [12]	28.25	t		
Drinking Water & Wastewater System & Waste Management	Paper, recycling, with deinking, at plant/RER U	ecoinvent Center (2013) [12]	76.62	t		
Drinking Water & Wastewater System & Waste Management	Packaging glass, green, at regional storage/CH S	ecoinvent Center (2013) [12]	82.24	t		
Drinking Water & Wastewater System & Waste Management	Operation, lorry 3.5-20t, fleet average/CH U	ecoinvent Center (2013) [12]	912.7	km		
Drinking Water & Wastewater System & Waste Management	Transport, freight, rail/CH U	ecoinvent Center (2013) [12]	29740.0	tkm		
Drinking Water & Wastewater System & Waste Management	Operation, lorry >28t, empty, fleet average/CH U	ecoinvent Center (2013) [12]	82.6	km		
Drinking Water & Wastewater System & Waste Management	Operation, lorry >28t, full, fleet average/CH U	ecoinvent Center (2013) [12]	82.6	km		
Drinking Water & Wastewater System & Waste Management	Treatment, sewage, to wastewater treatment, class 5/CH U	ecoinvent Center (2013) [12]	290701	m3	X	Adjusted according to interviews with operators and operating data (Wagner et al. 2015 [2]; Filli 2014 [20])
Drinking Water & Wastewater System & Waste Management	Heat, at cogen with biogas engine, allocation exergy/CH U	ecoinvent Center (2013) [12]	133914	kWh		
Drinking Water & Wastewater System & Waste Management	Electricity, at cogen with biogas engine, allocation exergy/CH U	ecoinvent Center (2013) [12]	133914	kWh		
Drinking Water & Wastewater System & Waste Management	Electricity, at cogen with biogas engine, allocation exergy/CH U	ecoinvent Center (2013) [12]	1103	kWh		
Public Services & Health Care	Public consumption, C_GOV, government consumption/year/CH U	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Public Services & Health Care	Public consumption, INV_OBLD, government investment/year/CH U	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Public Services & Health Care	Private consumption, C05, health care/year	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Public Services & Health Care	Private consumption, C09, education/year/Zernez	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Consumables	Private Ponsumption, C04, furnishings, household equipment/year	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Consumables	Private consumption, C02, clothing/year	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Consumables	Private consumption, C11, miscellaneous goods/year	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez

Category	Process Name According to Respective Database	Database	Quantity[a]	Unit	Adjusted[b]	Remarks
Services: Leisure Activities and Communications	Private consumption, C07, communication/year	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Services: Leisure Activities and Communications	Paper, newsprint, at regional storage/CH U	ecoinvent Center (2013) [12]	25276	kg		
Services: Leisure Activities and Communications	Vacation, hotel overnight stays	König et al. (2014) [22]	1	yr	X	Adjusted to fit the annual travel demand according to the local age structure
Agriculture & Forestry	Milking/CH U	ecoinvent Center (2013) [12]	522600	kg	X	
Agriculture & Forestry	Label housing system, pig, operation/CH U	ecoinvent Center (2013) [12]	4	porks		
Agriculture & Forestry	Loose housing system, cattle, operation/CH U	ecoinvent Center (2013) [12]	732	cows		
Agriculture & Forestry	Sheep for slaughtering, live weight, at farm/US	ecoinvent Center (2013) [12]	21882	kg		Sheep and goats included
Agriculture & Forestry	Chicken	LCA Food DK (Nielsen et al. 2013 [23])	144	kg	X	
Agriculture & Forestry	Pork (farm type 8)	LCA Food DK (Nielsen et al. 2013 [23])	800	kg	X	
Agriculture & Forestry	Beef (farm type 23)	LCA Food DK (Nielsen et al. 2013 [23])	549000	kg	X	
Agriculture & Forestry	Round wood, softwood, under bark, u=70% at forest road/RER U	ecoinvent Center (2013) [12]	2849.23	m^3		
Agriculture & Forestry	Residual wood, softwood, under bark, u=140%, at forest road/RER U	ecoinvent Center (2013) [12]	32.98	m^3		
Agriculture & Forestry	Diesel, burned in building machine/GLO U	ecoinvent Center (2013) [12]	159179.64	MJ		
Agriculture & Forestry	Naphtha, APME mix, at refinery/RER U	ecoinvent Center (2013) [12]	881.25	kg		
Agriculture & Forestry	Petrol, unleaded, at regional storage/CH U	ecoinvent Center (2013) [12]	26.87	kg		
Trade & Industry	Transport, electric scooter/CH U	ecoinvent Center (2013) [12]	645858	pkm	X	Adjusted to represent the activities of the Swiss Post (Wagner et al. 2015 [2], Saner 2014 [24])
Trade & Industry	Transport, scooter/CH U	ecoinvent Center (2013) [12]	258343	pkm	X	Adjusted to represent the activities of the Swiss Post (Wagner et al. 2015 [2], Saner 2014 [24])
Trade & Industry	Transport, lorry 16-32t, EURO4/RER U	ecoinvent Center (2013) [12]	8668	tkm	X	Adjusted to represent the activities of the Swiss Post (Wagner et al. 2015 [2], Saner 2014 [24])
Trade & Industry	Transport, lorry 3.5-7.5t, EURO4/RER U	ecoinvent Center (2013) [12]	3620	tkm	X	Adjusted to represent the activities of the Swiss Post (Wagner et al. 2015 [2], Saner 2014 [24])
Trade & Industry	Private consumption, INV_RES, residential construction/year/CH U	Jungbluth et al. (2011) [21]	1.00	yr	X	Based on Swiss input output database; scaled down to Zernez
Trade & Industry	Mine, gravel/sand/CH/I U	ecoinvent Center (2013) [12]	1	p	X	Adjusted to fit the annual output of 34000 tons
Residential Energy / Trade & Industry	Heat, borehole heat exchanger, at brine-water heat pump 10kW/CH U	ecoinvent Center (2013) [12]	897010	kWh		
Residential Energy / Trade & Industry	Heat, softwood logs, at wood heater 6kW/CH U	ecoinvent Center (2013) [12]	1568980	kWh		
Residential Energy / Trade & Industry / Services	Heat, softwood chips from forest, at furnace 1000kW/CH U	ecoinvent Center (2013) [12]	3291860	kWh	X	District heating network modelled according to Frischknecht et al. (2012) [25] and KBOB (2012) [26]
Residential Energy / Agriculture & Forestry / Services	Wood chips, softwood, u=140%, at forest/RER U	ecoinvent Center (2013) [12]	1405.91	m^3	X	Adjusted to fit wood chips supply in Zernez (imported and produced chips within the municipality) (Wagner et al. 2015 [2])
Residential Energy / Trade & Industry / Services	Heat, light fuel oil, at boiler 10kW, non-modulating/CH U	ecoinvent Center (2013) [12]	5934182	kWh		
Residential Energy / Trade & Industry / Services / Drinking Water & Wastewater System & Waste Management / Food Consumption	Electricity, low voltage, consumer mix, at grid	ecoinvent Center (2013) [12]	10402003	kWh	X	Electricity mix modelled for Zernez

[a]For food consumption category: average value for in-house consumption

[b]The electricity input was changed to the local electricity mix for all processes using local electricity

A.8 Overview of Assumptions for the Emission Classification

Table A.8 summarizes the most important assumptions which were used to allocate the emissions of a certain category to either “imported/embodied emissions”, “local emissions from local demand” or “local emissions from export” (cf. Figure 2.1 in Chapter 2). This table is meant to be a complement to the explanations in section 2.2.3 *Quantification and Modeling of Activities* in Chapter 2. It shall be pointed out that only the most important data sources which are helpful to understand the emission classification approach are mentioned in Table A.8 (see Table 2.1 in Chapter 2 for a full list of data sources which were used for the quantification and greenhouse gas assessment).

Table A.8: Overview of the most important assumptions for the classification of emissions of different categories into "imported/embodied emissions", "local emissions from local demand" and "local emissions from export".

Category	Emission Classes		
	Imported / embodied emissions	**Local emissions from local demand**	**Local emissions from export**
Food Consumption	- Assumption: food consumed by residents is imported	-	-
Residential Energy	- Upstream emissions according to ecoinvent Center (2013) [12]	- Local emissions estimated by BDB, OI DHN, EB (Wagner et al. 2015 [2]) and according to ecoinvent Center (2013) [12] - Local production of wood chips - Local electricity production - Heat production by biogas plant	- Electricity production by photovoltaic cells
Mobility	- Computation of total mobility CBF based on Microcensus (ARE et al. 2012 [27]) minus local emissions from local demand	- Rough estimation of trips within system boundaries by means of Microcensus (ARE et al. 2012 [27])	- Computation of total mobility PGA based on automatic traffic counters minus local emissions from local demand
Drinking Water & Wastewater System & Waste Management	- Upstream emissions according to ecoinvent Center (2013) [12] - Emissions from waste treatment	- Local emissions estimated by OI WWTP (Filli 2014 [20]), BDB, EB (Wagner et al. 2015 [2]) and according to ecoinvent Center (2013) [12]	-
Public Services & Health Care	- Total CBF of public services & health care estimated by Jungbluth et al. (2011) [21] minus local emissions from local demand	- Local emissions estimated by BDB, OI DHN, EB (Wagner et al. 2015 [2]), census of enterprises and according to ecoinvent Center (2013) [12]	-
Consumables	- Assumption: all consumables consumed by residents are imported	-	-
Services: Leisure Activities & Communications	- Upstream emissions according to ecoinvent Center (2013) [12] - Vacations - Assumption: emission from newspaper production are all released outside Zernez - Assumption: emission from communications are all released outside Zernez	- Local emissions estimated by combination of information from Microcensus (ARE et al. 2012 [27]), census of enterprises, interviews and BDB, OI DHN, EB (Wagner et al. 2015 [2])	-
Agriculture & Forestry	-	- Local emissions from production of wood chips for residents is already considered in the *residential energy* category	- All forestry and agricultural products are exported
Local Trade & Industry	- 90% (assumption based on Zhang et al 2013 [28]) of emissions from housing construction according to Jungbluth et al. (2011) [21] - Upstream emissions from the activities of the Swiss Post according to ecoinvent Center (2013) [12]	- Local emissions from the activities of the Swiss Post according to ecoinvent Center (2013) [12] - 10% (assumption based on Zhang et al 2013 [28]) of emissions from housing construction according to Jungbluth et al. (2011) [21]	- Local emissions estimated by BDB, OI DHN, EB (Wagner et al. 2015 [2]) and census of enterprises - Emissions from gravel quarry

Definitions:
BDB = building database; OI = operation information; DHN = district heating network; EB = electricity bills; CBF = consumption-based footprint; PGA = purely geographic accounting; WWTP = wastewater treatment plant;

A.9 References

[1] Bundesamt für Landestopografie (swisstopo), *Website;*, 2014. [Online]. Available: http://www.swisstopo.admin.ch. [Accessed: 01-Apr-2014].

[2] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[3] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[4] SIA, *SIA 380/1: Thermische Energie im Hochbau*, Zurich, Switzerland: Swiss Society of Engineers and Architects, 2009.

[5] Bundesamt für Statistik (BFS), *Eidgenössisches Gebäude- und Wohnungsregister*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[6] H. Wallbaum, N. Heeren, M. Jakob, G. Martius, and N. Gross, *Gebäudeparkmodell*, Stadt Zürich, Zurich, Switzerland, 2010.

[7] METEOTEST, *Meteonorm, version 7*, METEOTEST, Bern, Switzerland, 2012.

[8] K. Meister, M. Balmer, F. Ciari, A. Horni, M. Rieser, R. A. Waraich, and K. W. Axhausen, *Large-Scale Agent-Based Travel Demand Optimization Applied to Switzerland, Including Mode Choice*, in *12th World Conference on Transportation Research*, 2010.

[9] M. Balmer, K. W. Axhausen, and K. Nagel, *Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations*, *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1985, no. 1, pp. 125–134, 2006.

[10] Bundesamt für Statistik (BFS), *Eidgenössische Volkszählung (VZ)*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2000.

[11] Bundesamt für Statistik (BFS) and Bundesamt für Raumentwicklung (ARE), *Mobilität in der Schweiz. Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2012.

[12] ecoinvent Center, *Ecoinvent data, version 2.2*, http://www.ecoinvent.org/database/, 1998-2013, 2013.

[13] N. Jungbluth, R. Itten, and M. Stucki, *Umweltbelastungen des privaten Konsums und Reduktionspotenziale*, Uster, Switzerland, 2012.

[14] Amt für Wirtschaft und Tourismus (AWT), *Ständige Wohnbevölkerung nach Eckwerten, Gemeinden*, Chur, Switzerland, 2010.

[15] Bundesamt für Statistik (BFS), *Reisen der Schweizer Wohnbevölkerung 2010*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2011.

[16] A. Chavez and A. Ramaswami, *Articulating a trans-boundary infrastructure supply chain greenhouse gas emission footprint for cities: Mathematical relationships and policy relevance*, *Energy Policy*, vol. 54, pp. 376–384, 2013.

[17] D. Saner, *Towards Sustainable Supply of Household Consumption: Demand Modeling and Environmental Assessment*, PhD Thesis, ETH Zurich, Switzerland, 2013.

[18] D. Saner, C. Beretta, B. Jäggi, R. Juraske, F. Stoessel, and S. Hellweg, *FoodPrints of house-*

holds, *The International Journal of Life Cycle Assessment*, vol. 21, no. 5, 2016.

[19] F. Grass, *Operator of the biogas plant in Zernez, Switzerland, Personal Communication, 2014*, 2014.

[20] I. A. Filli, *Operator of the waste water treatment plant in Zernez, Switzerland, Personal Communication, 2014*, 2014.

[21] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[22] A. König, N. Jungbluth, and G. Doublet, *Treibhausgasbilanz verschiedener Feriensznarien*, Zurich, Switzerland, 2014.

[23] P. H. Nielsen, A.-M. Nielsen, B. P. Weidema, R. Dalgaard, and Halberg N, *LCA food data base*, www.lcafood.dk, 2003.

[24] D. Saner, *Politics and Social Responsibility, Post CH Ltd., Switzerland, Personal Communication, 2014*, 2014.

[25] R. Frischknecht, M. Stucki, K. Flury, R. Itten, and M. Tuchschmid, *Primärenergiefaktoren von Energiesystemen*, Uster, 2012.

[26] KBOB, *Ökobilanzdaten im Baubereich 2009/1*, 2012.

[27] Amt für Raumentwicklung, Amt für Energie und Verkehr, Amt für Natur und Umwelt, Amt für Wirtschaft und Tourismus, Tiefbauamt, and Verkehrsplanung Stadt Chur, *Mobilität in Graubünden - Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Chur, Switzerland, 2012.

[28] X. Zhang, L. Shen, and L. Zhang, *Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong*, *Renewable and Sustainable Energy Reviews*, vol. 17, pp. 160–169, 2013.

Appendix B

Assessing Space Heating Demand on a Regional Level: Evaluation of a Bottom-Up Model in the Scope of a Case Study

B.1 COMPUTATION OF THE EMPIRICAL DATABASE RANGE

In order to evaluate the housing energy demand model of Saner et al. (2013) [1], the model results were compared with final energy consumption data collected within the project *Zernez Energia 2020* (ETHZ and Zernez 2015 [2]). For this comparison, the amounts of energy carriers had to be converted to net energy demand required for space heating. Figure B.1 explains in detail how we proceeded for this conversion per building. This procedure was applied to all 133 case-study buildings.

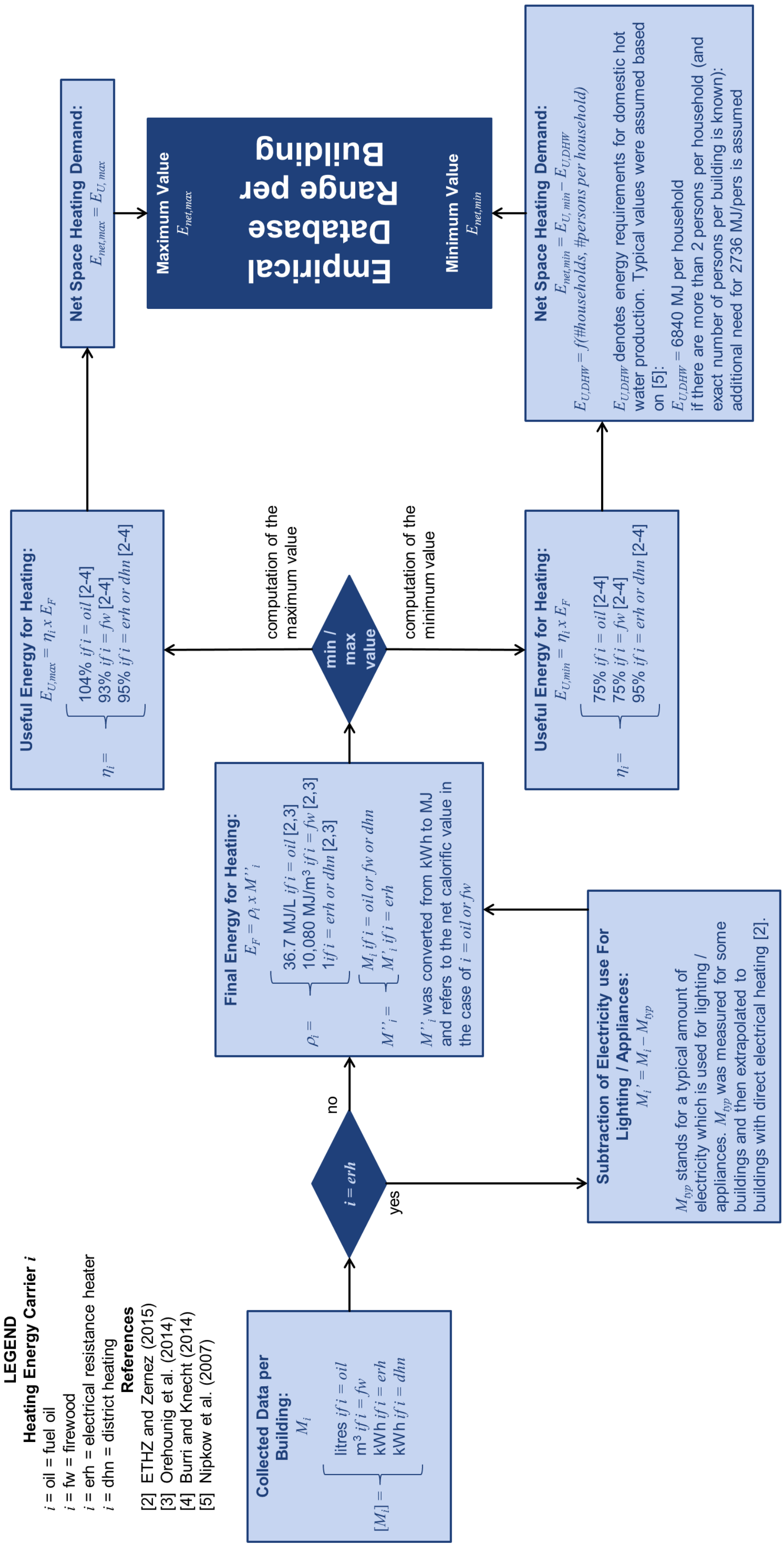


Figure B.1: Schematic description of the derivation of the empirical database range for one building. This procedure was applied to all 133 case-study buildings. Amounts of energy carriers and energy demands refer to annual values.

B.2 MODEL DESCRIPTION

The housing energy demand model of Saner et al. (2013) [1] is based on the Swiss Standard SIA 380/1 (SIA 2009 [6]) and establishes heat balances for each building according to equation (3.1) in Chapter 3.

Besides SIA 380/1, three main sources supply input data to the model: Meteonorm (METEOTEST 2012 [7]), the Swiss Federal Register of Buildings and Dwellings (FRBD) (BFS 2013 [8]) and building-specific statistics such as the analyses of Wallbaum et al. (2010) [9]. *Meteonorm* version 7 (METEOTEST 2012 [7]) is capable of generating climatological data of a typical year for any location in the world. This software was thus used for data on hourly outdoor temperatures as well as direct horizontal and diffuse horizontal radiation in Zernez. The FRBD contains up-to-date building-specific data about each residential building in Switzerland, including information about geographical coordinates of the building, construction year, construction period, building area, number of stories, number of apartments, area of each apartment, number of rooms of each apartment, heating and hot water system as well as corresponding energy carriers. The data from FRBD was combined and enhanced with statistical data provided by Wallbaum et al. (2010) [9], who assessed various aspects of four different building components (roof, walls, floor and windows) encompassing refurbishment rates, U-values (heat transfer coefficients), and g-values (solar energy transmittance of windows) and referred their findings to the construction year of a building as well as to the year of renovation of the respective building component. In order to determine the different terms of the energy balance in equation (3.1), the surface areas of these four components (roof, walls, floor and windows) have to be known for each building. Unfortunately, information about the shape of buildings is not available in the FRBD. Therefore, each building is assumed to be a cube with a base area corresponding to the sum of all apartment areas divided by the number of stories. The building surface is then derived by the multiplication of this bottom area with a correction factor and a building envelope factor (Wallbaum et al. 2010 [9]; Dettli et al. 2007 [10]). The subtraction of floor and roof areas, which both equal the base area, from the building surface results in the façade area. Finally, the share of window area is assumed to be 18% (Jagnow et al. 2002 [11]) in order to split the façade area into window and wall area.

Based on these assumptions and data sources, the different components of equation (3.1) can be derived. While Q_{iP} and Q_{iEl} are estimated by standard values provided by SIA 380/1, the determination of the solar gains is more complex: The total area of the windows is deduced by the aforementioned assumptions, glass properties are extracted from Wallbaum et al. (2010) [9] by means of the construction year which is provided by the FRBD, and shading of nearby buildings is taken into account by geographical coordinates given in the FRBD and according to the SIA 380/1. Finally, relevant climatic data was extracted from the software *Meteonorm* (METEOTEST 2012 [7]). The sum of all thermal gains is multiplied by η_g which is the degree of utilization for heat gains and depends on the thermal storage capacity of the building mass (SIA 2009 [6]).

Table B.1: Overview of stochastically modeled parameters for the computations of the space heating demand in the simplified model (table adapted from Saner et al. (2013) [1]**). [CoV = Coefficient of variation]**

Variable name	Remarks	Distribution type	Parameter	CoV
Room temperature	Average room temperature	normal	μ=20°C, σ=2°C	0.1
Deviation from south	Deviation of a building facing towards south	normal	μ=0°, σ=9°	0.2
Thermal storage capacity	Thermal storage capacity of a building mass (SIA 2009 [6])	triangular	mean=0.4, min=0.1, max=0.5	
Temperature control	Addition to room temperature due to inadequate control of room temperature (SIA 2009 [6])	triangular	mean=1°C, min=0°C, max=2°C	
Mechanical ventilation	Share of buildings with mechanical ventilation	according to Salvi et al. (2010 [12])		
Renovation type	Determines the refurbished component for buildings where a renovation period is indicated in the FRBD	uniform	min=0, max=1	
Roof type	Slanted roof or flat roof	discrete uniform	min=0, max=1	
Roof inclination	Inclination of slanted roofs	normal	μ=30°, σ=6°	0.2
Time of refurbishment of roof		according to Wallbaum et al. (2010) [9]		
Time of refurbishment of walls		according to Wallbaum et al. (2010) [9]		
Time of refurbishment of floor		according to Wallbaum et al. (2010) [9]		
Time of refurbishment of windows		according to Wallbaum et al. (2010) [9]		
Share of window area	Window area divided by total façade area (Jagnow et al. 2002 [11])	normal	μ=0.18, σ=0.036	0.2

Transmission losses are estimated by means of heat transfer coefficients (U-values) for walls, roof, floor, and windows, the corresponding component area and the temperature gradient through the building envelope. The U-values are retrieved from Wallbaum et al. (2010) [9], while the areas of the components are estimated by the FRBD combined with the above mentioned assumptions. Finally, calculations for ventilation losses (Q_V) draw on hourly differences between outdoor and ambient room temperature as well as on standard values for hourly air exchange flows (SIA 2009 [6]).

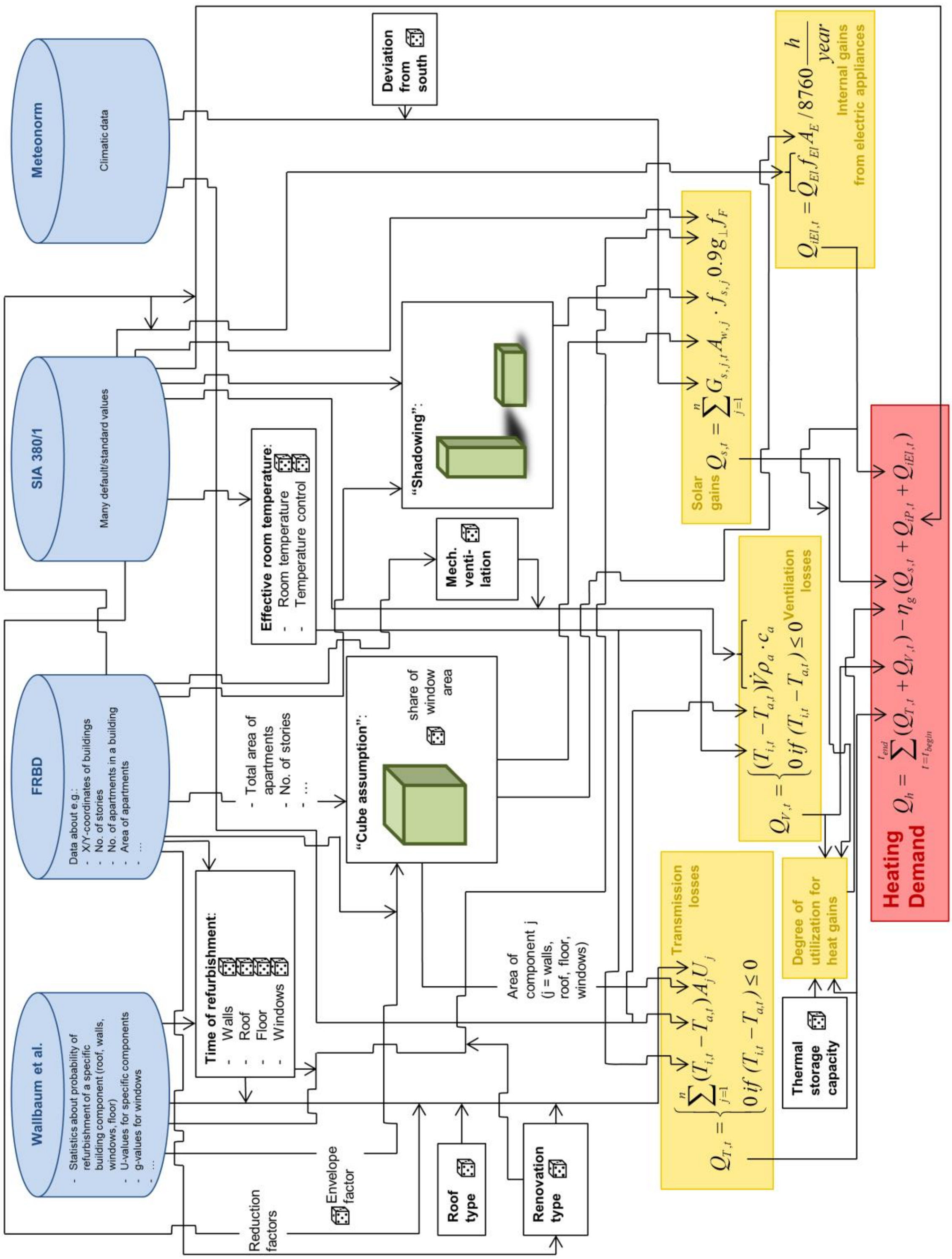


Figure B.2: **Simplified flow chart of the housing energy demand model of Saner et al. (2013)** [1]**. The dices stand for the stochastically modeled parameters of Table B.1. The variables denote the following: Q_h = space heating demand; Q_T = transmission losses; Q_V = ventilation losses; η_g = degree of utilization for heat gains; Q_s = solar gains; Q_{iP} = internal gains due to occupancy; Q_{iEl} = internal gains due to electricity use; T_i = indoor temperature; T_a = outdoor temperature; A_j = area of component j; U_j = heat transfer coefficient of component j; V = hourly air exchange; ρ_a = density of air; c_a = specific heat capacity of air; $G_{s,j}$ = hourly global solar radiation for façade j; $A_{w,j}$ = window area of façade j; $f_{s,j}$, f_F = reduction factors for shading of buildings and window frame; g = solar energy transmittance of windows; Q_{El} = electricity demand; f_{El} = reduction factor; A_E = heated floor area.**

As mentioned in Chapter 3, Latin Hypercube sampling was applied to capture the uncertainty and variability of parameters for which assumptions had to be taken. Table B.1 was adapted from Saner et al. (2013) [1] and gives an overview of all parameters that were modeled stochastically. Contrary to Saner et al. (2013) [1], Table B.1 shows only variables used for modeling space heating demand and lists neither parameters used for estimating hot water production, electricity demand nor for assigning demands to specific households.

The flow chart in Figure B.2 illustrates in a simplified manner how the different model components interact in order to calculate the space heating demand of a certain building. The dices symbolize the stochastically modeled parameters listed in Table B.1.

B.3 Model Results for Single Buildings

The Latin Hypercube sampling results in distributions of the estimated space heating demand for each single building. Three examples for randomly chosen buildings are depicted in Figure B.3.

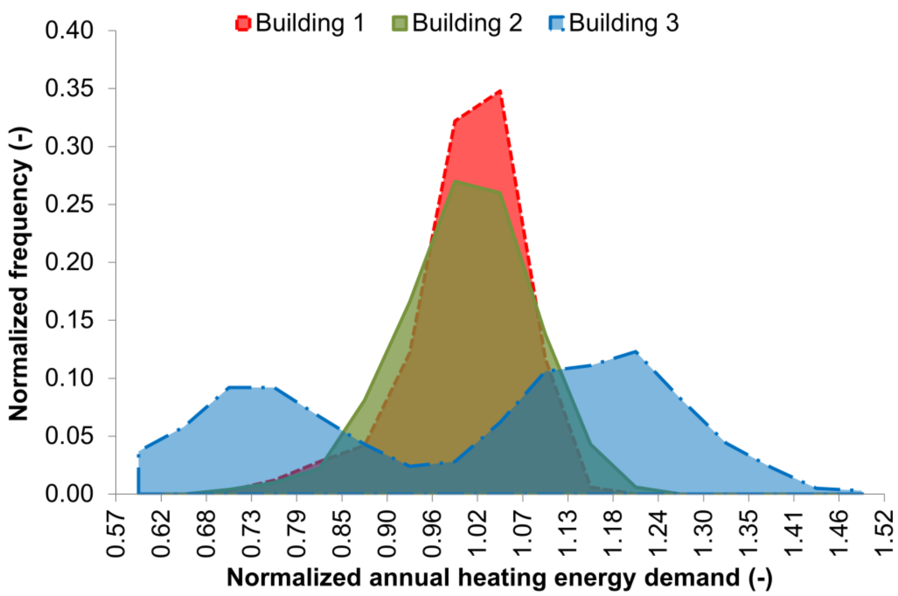


Figure B.3: **Normalized frequency distributions of the space heating demand estimations for three randomly chosen buildings (applying "Sturges' Rule" (Sturges 1926 [13])). The simulation results of each building were normalized with the respective average prediction.**

Figure B.3 reveals two important aspects. On the one hand, large variations in the predictions can be observed for some buildings (e.g. building 3 in Figure B.3), whereas the frequency distribution of others is rather narrow (e.g. building 2 in Figure B.3). A slight tendency for newer buildings to feature lower coefficients of variation than older ones was observed. This might arise from the fact that the time of refurbishment of a building component is modeled stochastically. Thereby, older buildings may introduce larger uncertainties by offering a larger time span in which refurbishments could take place compared to recently erected buildings. This conclusion is also supported by the results of the global sensitivity analysis. The two peaks in the frequency distribution of building 3 are probably also due to this effect since this building was constructed in the year 1600. On the other hand, the frequency distributions illustrate that the demand predictions are generally not normally distributed, although the building-wise comparison of the average with the median of the sample estimates revealed that these two indicators are close to each other. The maximum deviation of the median from the average is 10%, while in 90% of the cases the deviation amounts to less than 5% (mostly even less than 1%). Therefore, the conclusions are likely to stay the same regardless of the choice between these two statistical indicators. For the comparisons with primary data, it was decided to take the average of the space heating demand estimates.

B.4 GLOBAL SENSITIVITY ANALYSIS

Figure B.4 gives a full overview of the results of the global sensitivity analysis. In this figure, green box plots symbolize the results for parameters which were explicitly sampled by the Latin Hypercube simulation and correspond to Table B.1. Red boxes however stand for parameters which are not directly varied by the model but which are in direct dependence of model parameters and which can help to better understand the results of the global sensitivity analysis. For instance, room temperature as well as temperature control were sampled. The sum of these two parameters is the effective indoor temperature which actually enters the model computations and which is much easier to interpret than its separated constituents. Similarly, this applies to the time of refurbishment of a certain component (roof, floor, walls, and windows). While the time of refurbishment is actually subjected to Latin Hypercube sampling, this parameter is harder to interpret from an intuitive point of view than for instance a U-value which directly depends on the time of refurbishment.

Figure B.4 suggests that the model results are most sensitive to effective room temperature and U-values of walls. Though, according to the box plots, δ of the wall's U-value can become zero and δ of room temperature is close to zero for some buildings. The question arises whether one of these two parameters is always important to the model result of a certain building. Figure B.5 shows that there is a negative correlation between the δ of these two parameters and that in all cases at least one of these two parameters takes a relatively large value compared to all other parameters. In Figure B.6, the δ-values of the wall's U-value and of the effective room temperature are sorted by the age of the corresponding building. It becomes obvious that the model results of old buildings are dominated by the U-values of walls while effective room temperature is more important for heat demand predictions for new buildings.

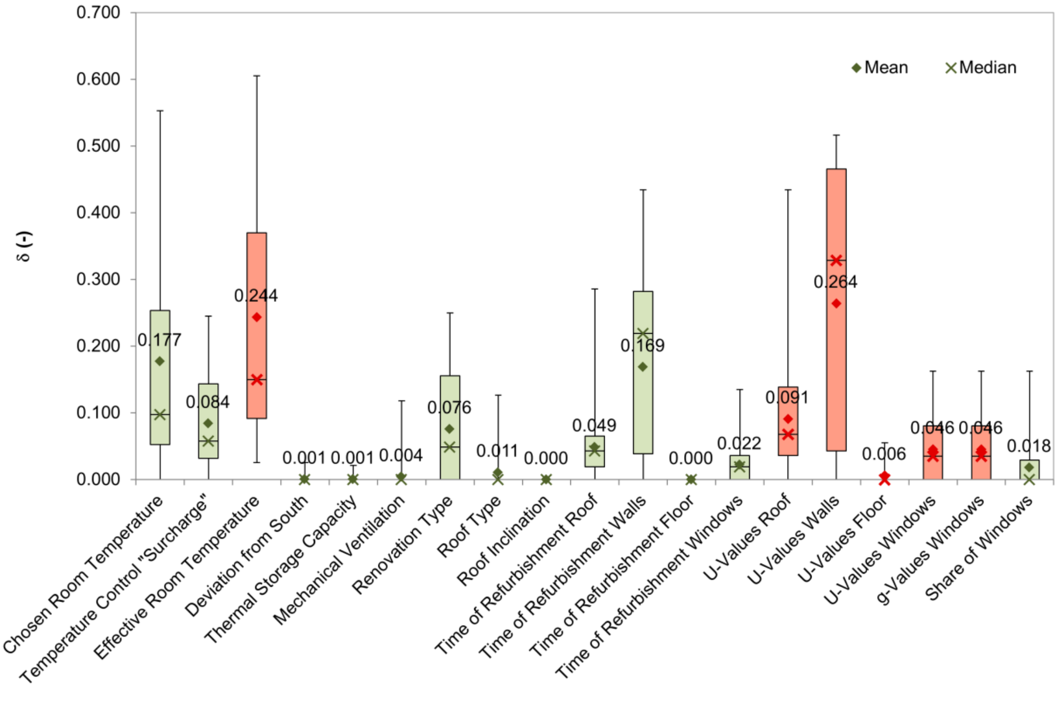


Figure B.4: **Resulting distributions of the density-based sensitivity measure δ for different model parameters after applying Plischke et al.'s approach (Plischke et al. 2013 [14]) to the 133 case-study buildings. Green boxes correspond to the model parameters of Table B.1 (directly varied by Latin Hypercube sampling), while red boxes symbolize derived parameters which directly depend on model parameters.**

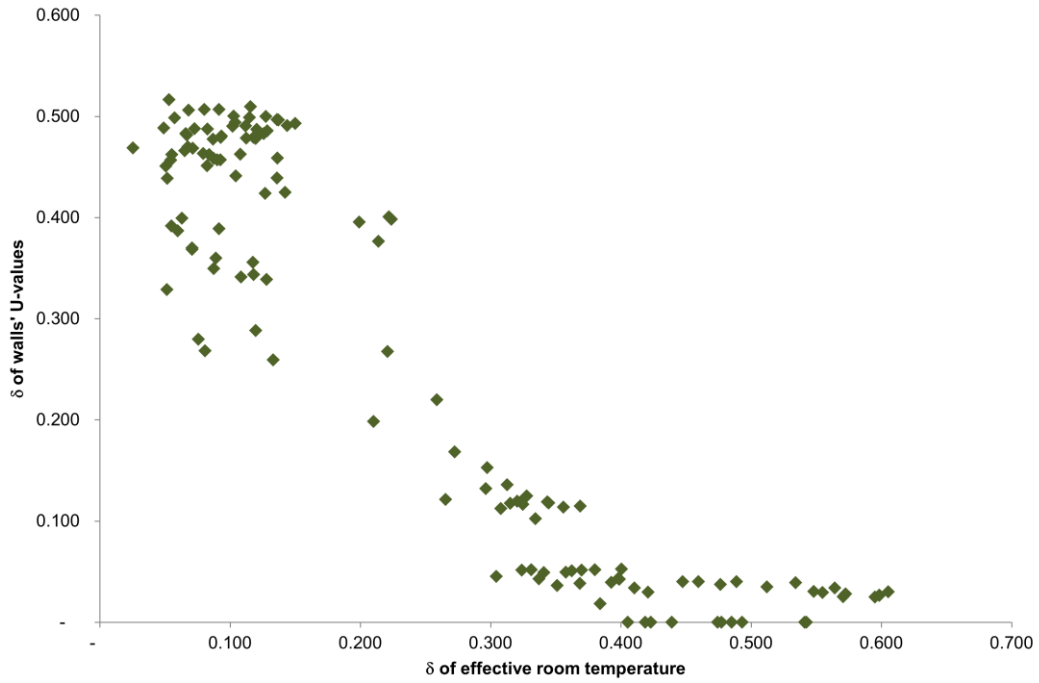


Figure B.5: **δ-values for U-values of walls against δ-values for the effective room temperature. See text, Table B.1 and Figure B.4 for more information.**

The δs of the U- and g-values of all components are depicted in Figure B.7, again sorted by the building's age. In this figure, one can observe that δ decreases in distinct steps. These drops are in line with the steps of U-values in the statistical database of Wallbaum et al. (2010) [9]. Old buildings offer a large time span in which refurbishments can take place. Therefore, U-values will spread over a large range of values. For new buildings however, the possible range of U-values is much more restricted. The most extreme case in this regard is a building constructed in the year 2010. For such a building, no refurbishment can be modeled and hence the U-values will stay the same for all 1000 model runs. As a consequence, U-values are not thought to influence the model predictions of new buildings and the corresponding δ becomes zero.

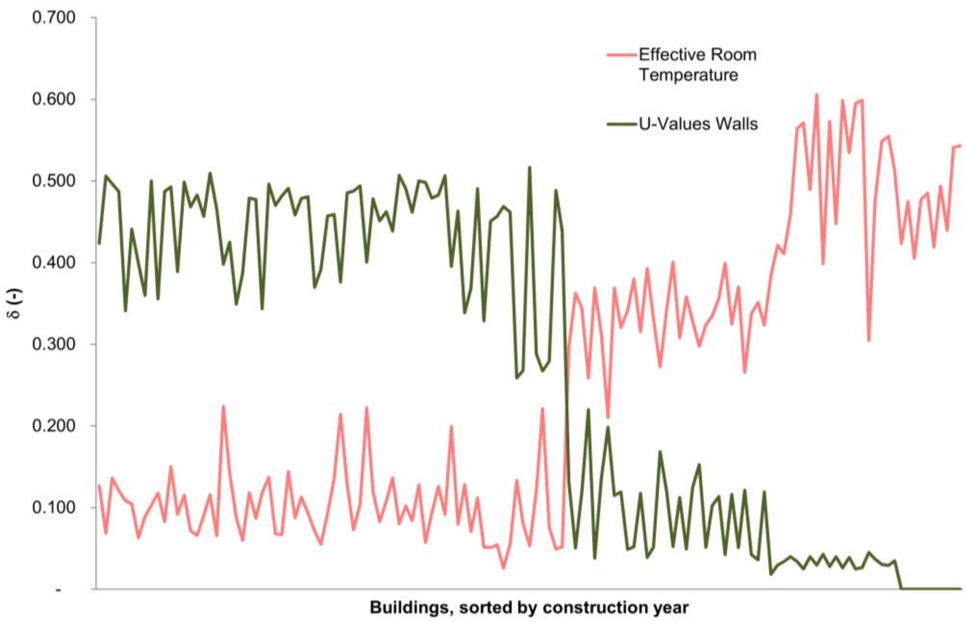


Figure B.6: **δ-values for U-values of walls and for effective room temperature sorted by the construction year of the corresponding buildings.**

As a conclusion, one can state that the applied restriction of the Latin Hypercube sampling to building-specific properties has direct consequences on which parameters are important for the model predictions of which building. This effect cannot only be observed for U-values and effective room temperature, but also for other parameters. For instance, mechanical ventilation is just important for newer buildings since this parameter is only allowed to be non-zero for buildings constructed after 2001. Another example is given by the parameter "renovation type". This parameter only plays a role for buildings for which a renovation period is indicated by the FRBD. The specification of the renovation period is not mandatory which is why this entry is empty for many buildings in the FRBD. However for those buildings with a renovation period indicated, no further information is given about what component was refurbished. The uniformly distributed random variable "renovation type" determines therefore the component that was refurbished in the indicated renovation period. Consequently, δ takes only a non-zero value for buildings where a renovation period is given in the FRBD. For these buildings however, δ can become large as "renovation type" directly affects the modeled U-values of components.

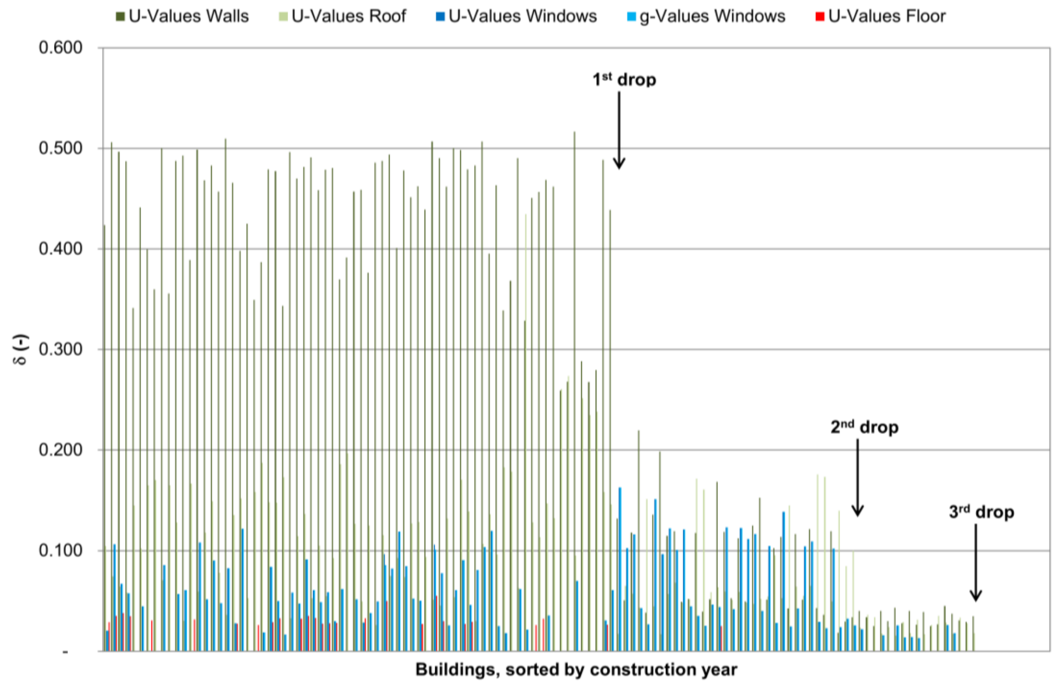


Figure B.7: δ-values for U-values of roof, walls, floor, and windows as well as δ-values for the g-values of windows, sorted by the age of the buildings.

B.5 Further Results

Figure B.8 supports the conclusions drawn in the context of Figure 3.5 in Chapter 3. Irrespective of how the buildings are sorted, the cumulative annual heating demand curves are close to one another and more or less parallel to each other.

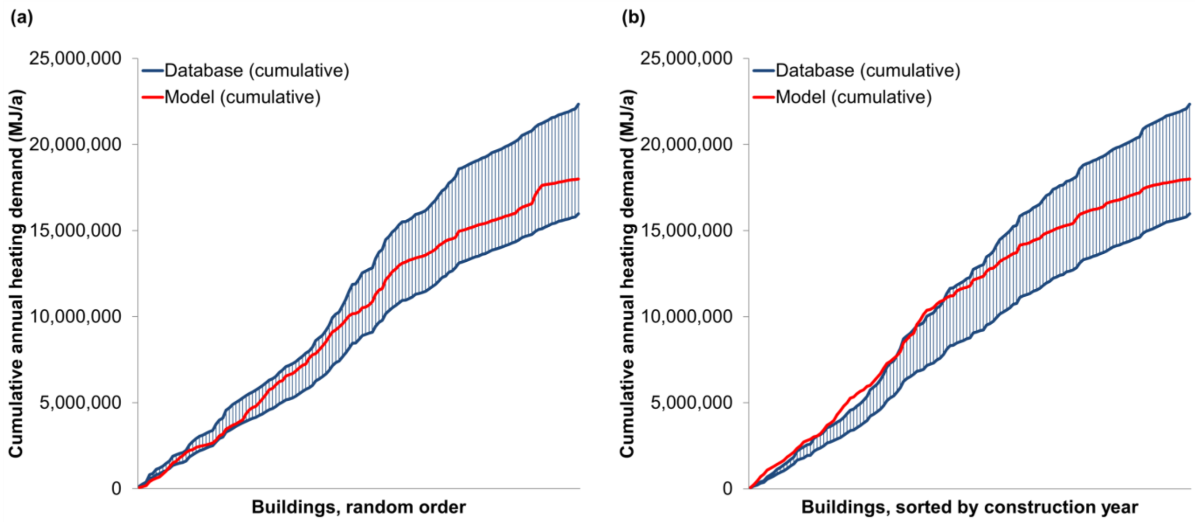


Figure B.8: Cumulative annual heating demand of buildings (empirical database range in dark blue). Buildings are in random order in (a), while in (b), buildings are sorted by construction year (from oldest to newest date). The cumulative curve of the model (red) was built by accumulating the simulated heating demands corresponding to the sorted buildings.

B.6 References

[1] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[2] M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, A. Frömelt, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

[3] K. Orehounig, G. Mavromatidis, R. Evins, V. Dorer, and J. Carmeliet, *Towards an energy sustainable community: An energy system analysis for a village in Switzerland*, *Energy and Buildings*, vol. 84, pp. 277–286, 2014.

[4] R. Burri and K. Knecht, *Tool zur Lebenszyklusanalyse von Wärmeerzeugungssystemen*, Master Thesis, ETHZ, Zurich, 2014.

[5] J. Nipkow, S. Gasser, and E. Bush, *Der typische Haushalt-Stromverbrauch. Energieverbrauch von Haushalten in Ein- und Mehrfahmilienhäusern*, Bulletin SEV/VSE 19/2007, 2007.

[6] SIA, *SIA 380/1: Thermische Energie im Hochbau*, Zurich, Switzerland: Swiss Society of Engineers and Architects, 2009.

[7] METEOTEST, *Meteonorm, version 7*, METEOTEST, Bern, Switzerland, 2012.

[8] Bundesamt für Statistik (BFS), *Eidgenössisches Gebäude- und Wohnungsregister*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[9] H. Wallbaum, N. Heeren, M. Jakob, G. Martius, and N. Gross, *Gebäudeparkmodell*, Stadt Zürich, Zurich, Switzerland, 2010.

[10] R. Dettli, S. Bade, A. Baumgartner, and M. Bleisch, *Vorstudie zur Erhebung von Energiekennzahlen von Wohnbauten*, Bundesamt für Energie (BFE), Berne-Ittigen, Switzerland, 2007.

[11] K. Jagnow, D. Horschler, and S. Wolff, *Die neue Energieeinsparverordnung 2002*, DWD-Verlag, Köln, Germany, 2002.

[12] M. Salvi, A. Horehàjovà, and J. Neeser, *Der Minergie-Boom unter der Lupe*, Zürich, 2010.

[13] H. A. Sturges, *The choice of a class interval*, *J. Am. Stat. Assoc.*, vol. 21, no. 153, pp. 65–66, 1926.

[14] E. Plischke, E. Borgonovo, and C. L. Smith, *Global sensitivity measures from given data*, *European Journal of Operational Research*, vol. 226, no. 3, pp. 536–550, 2013.

Appendix C

Big Data GIS Analysis for Novel Approaches in Building Stock Modelling

C.1 STATISTICAL FORMULAS

For the following formulas x is the measured value whereas y is the simulated value.

Mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{C.1}$$

Root mean squared error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{C.2}$$

Mean relative error:

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - x_i}{x_i} \cdot 100.0 \tag{C.3}$$

Mean bias error:

$$MBE = \frac{1}{n}\sum_{i=1}^{n} y_i - x_i \tag{C.4}$$

Coefficient of determination:

$$R^2 = 1.0 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{C.5}$$

C.2 WARM WATER

In [1] a typical heat demand of warm water of 21 kWh/m^2 for single-family houses and 14 kWh/m^2 for multi-family houses is specified. Table C.1 shows the percentage of the warm water heat demand to the total building heat demand estimated in [2].

Table C.1: Percentage of warm water heat demand to the total heat demand for different building construction periods of single-family houses and multi-family houses.

Construction period	SFH [%]	MFH [%]
< 1919	13.9	9.9
1919-1945	11.9	9.0
1946-1960	13.9	9.9
1961-1970	13.9	10.3
1971-1980	14.9	10.7
1981-1985	16.7	12.1
1986-1990	17.4	13.9
1991-1995	18.1	14.6
1996-2000	18.9	17.4
2001-2005	19.8	17.4
2006-2010	23.1	21.7
2011-2015	37.9	39.7

C.3 Detailed Results

Table C.2: **Overview of normalized mean absolute error (MAE%), root mean squared error (RMSE%), mean bias error (MBE%) and coefficient of determination (R_2) for the different building types and construction periods for both space heating demand, and space heating demand and warm water heat demand.**

		Room heat demand						Room heat + warm water demand					
	Age	#	MAE%	RMSE%	MRE	MBE%	R^2	#	MAE%	RMSE%	MRE	MBE%	R^2
Single-family houses	<1919	50	50.1	59.3	71.4	46.6	0.09	101	45.8	54.1	60.8	38.6	0.21
	1919-1945	92	49.0	60.9	63.4	42.0	-0.49	155	45.4	56.0	55.9	36.8	-0.20
	1946-1960	45	54.3	66.4	70.5	54.1	0.58	89	56.7	70.1	67.5	56.2	0.12
	1961-1970	8	30.8	37.2	-15.6	-16.7	-0.27	47	18.7	25.3	-0.4	-2.7	0.00
	1971-1980	17	28.0	38.4	7.2	6.2	-0.26	37	24.5	39.5	-0.4	-4.8	0.40
	1981-1990	30	27.8	38.3	21.8	11.2	-0.16	51	23.3	31.0	14.7	6.5	0.28
	1991-2000	31	35.5	45.6	37.0	25.4	-0.66	48	29.6	38.2	32.6	20.0	0.35
	All	273	44.7	56.9	52.6	35.9	0.14	528	39.8	51.8	43.7	28.9	0.13
Multi-family houses	<1919	199	32.0	41.1	29.6	13.6	0.40	513	26.1	33.1	20.8	7.9	0.53
	1919-1945	75	30.5	63.4	29.1	6.8	0.53	241	25.3	41.8	19.3	4.5	0.62
	1946-1960	16	27.2	44.1	2.4	-17.4	0.60	36	24.8	37.1	-2.2	-18.0	0.67
	1961-1970	2	63.3	82.2	-54.3	-63.3	-0.22	31	35.3	51.3	-24.3	-32.9	0.22
	1971-1980	3	41.5	49.0	-14.3	-28.9	0.18	19	28.6	38.2	-16.1	-23.2	0.62
	1981-1990	5	39.1	55.8	-35.2	-39.1	0.46	29	32.0	51.1	-24.3	-31.2	0.57
	1991-2000	9	38.5	53.4	-21.1	-33.9	0.08	41	25.4	34.3	-17.2	-22.7	0.34
	All	309	32.7	55.3	24.6	3.4	0.52	910	26.7	40.3	14.0	-1.6	0.61
Mixed residential usage	<1919	121	26.6	38.4	20.3	3.3	0.59	261	26.9	37.9	15.7	0.0	0.64
	1919-1945	31	30.7	55.4	1.7	-15.5	0.61	71	26.0	43.0	7.2	-7.6	0.67
	1946-1960	9	26.9	35.6	-1.8	-15.0	0.41	23	51.7	117.1	-16.1	-45.1	0.36
	1961-1970	5	34.5	48.7	-31.2	-34.5	-0.29	18	42.0	71.4	-36.1	-42.0	0.22
	1971-1980	2	18.8	25.8	44.5	17.6	0.81	11	38.9	57.9	-17.7	-37.1	0.39
	1981-1990	6	41.6	58.5	-14.5	-36.9	0.45	16	43.2	72.3	-19.9	-41.0	0.49
	1991-2000	1	47.1	47.1	-47.1	-47.1	0.00	7	25.3	33.5	-14.7	-15.5	0.57
	All	175	28.6	46.2	13.1	-5.9	0.59	407	33.0	82.0	7.3	-15.6	0.51
All residential buildings	<1919	370	31.0	41.6	32.2	11.5	0.56	875	27.3	36.4	23.9	6.6	0.64
	1919-1945	198	34.9	67.0	40.7	8.6	0.63	467	28.5	46.3	29.6	6.7	0.70
	1946-1960	70	34.9	54.1	45.6	3.8	0.76	148	43.4	142.7	37.5	-16.8	0.52
	1961-1970	15	43.6	87.6	-26.0	-41.7	0.42	96	36.2	83.9	-14.8	-33.1	0.55
	1971-1980	22	29.6	40.5	7.7	1.0	0.20	67	31.6	62.5	-7.7	-24.1	0.68
	1981-1990	41	36.7	83.2	9.6	-24.5	0.72	96	34.1	73.9	-2.8	-28.0	0.68
	1991-2000	41	37.6	69.0	22.2	-10.7	0.61	96	26.1	41.5	7.9	-15.0	0.77
	All	757	33.2	56.5	32.0	5.4	0.63	1845	30.2	66.5	21.1	-2.8	0.60

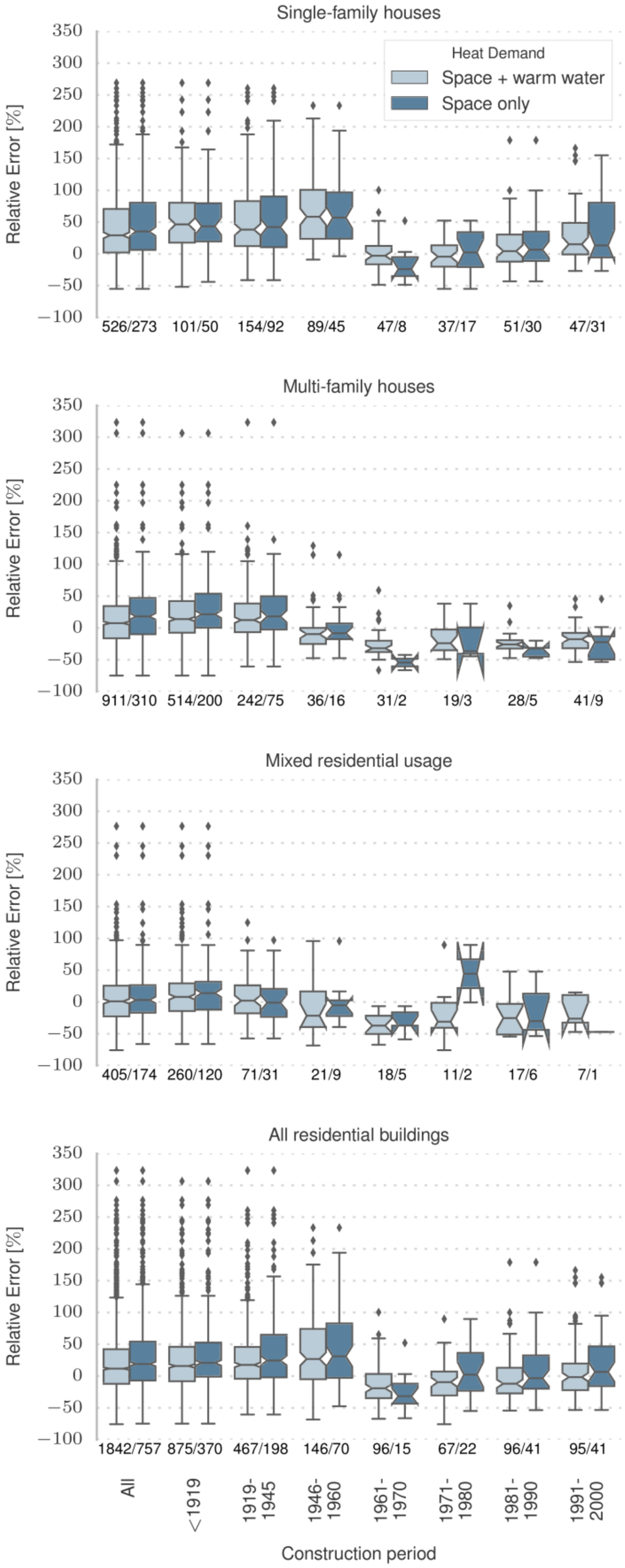


Figure C.1: **Relative error of simulated heat demand (only space heating and space heating + warm water) to the measured consumed energy demand for different building types. Sample size *n* is given below the plots. The boxes indicate the interquartile range (IQR) between the first and the third quartile. The whiskers extend until 1.5·*IQR*. The notches in the boxes indicate the confidence interval for the median.**

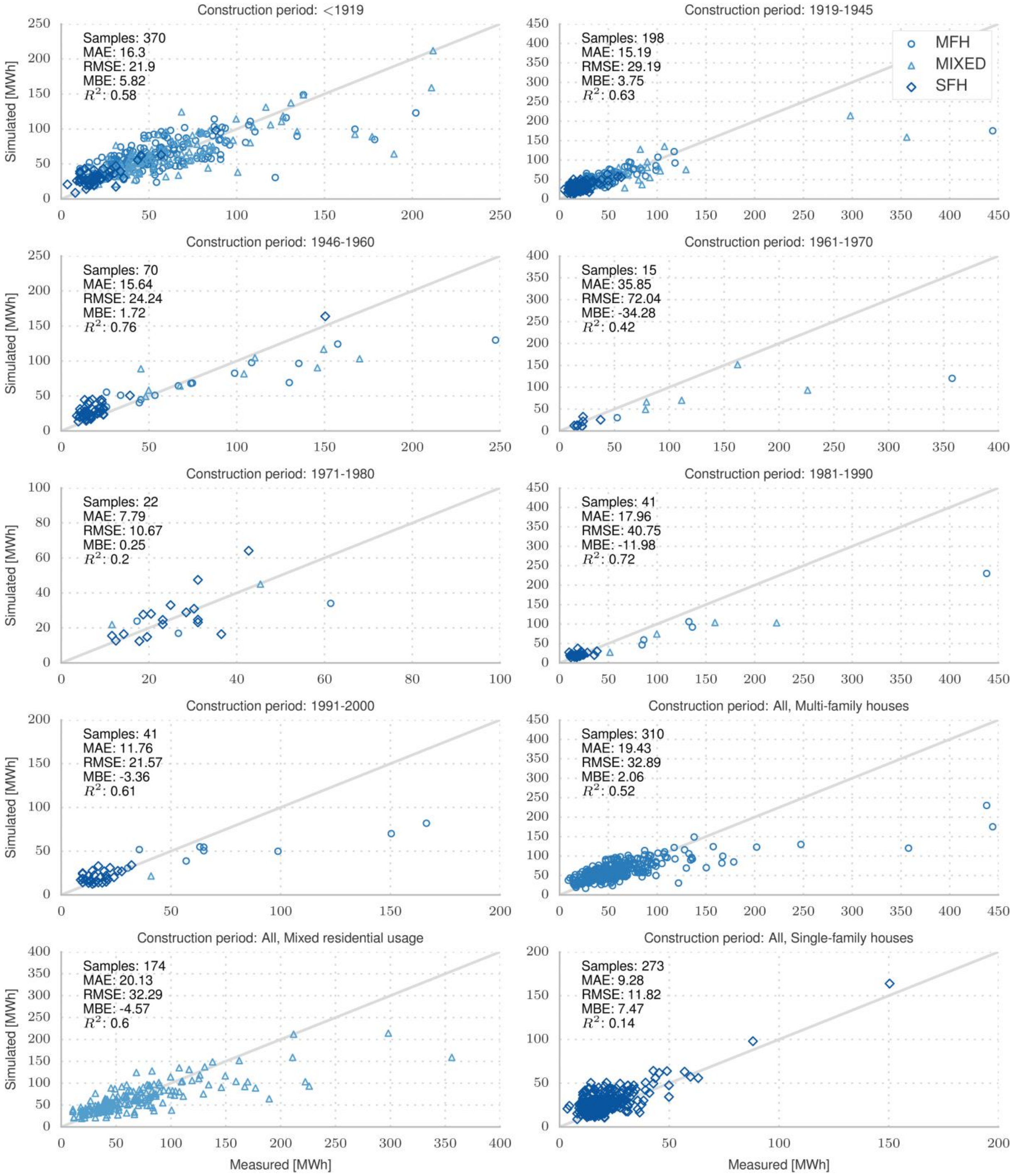


Figure C.2: **Comparison of simulated and measured space heating demand for the building types single-family houses (SFH), multi-family houses (MFH), and mixed residential usage (MIXED).**

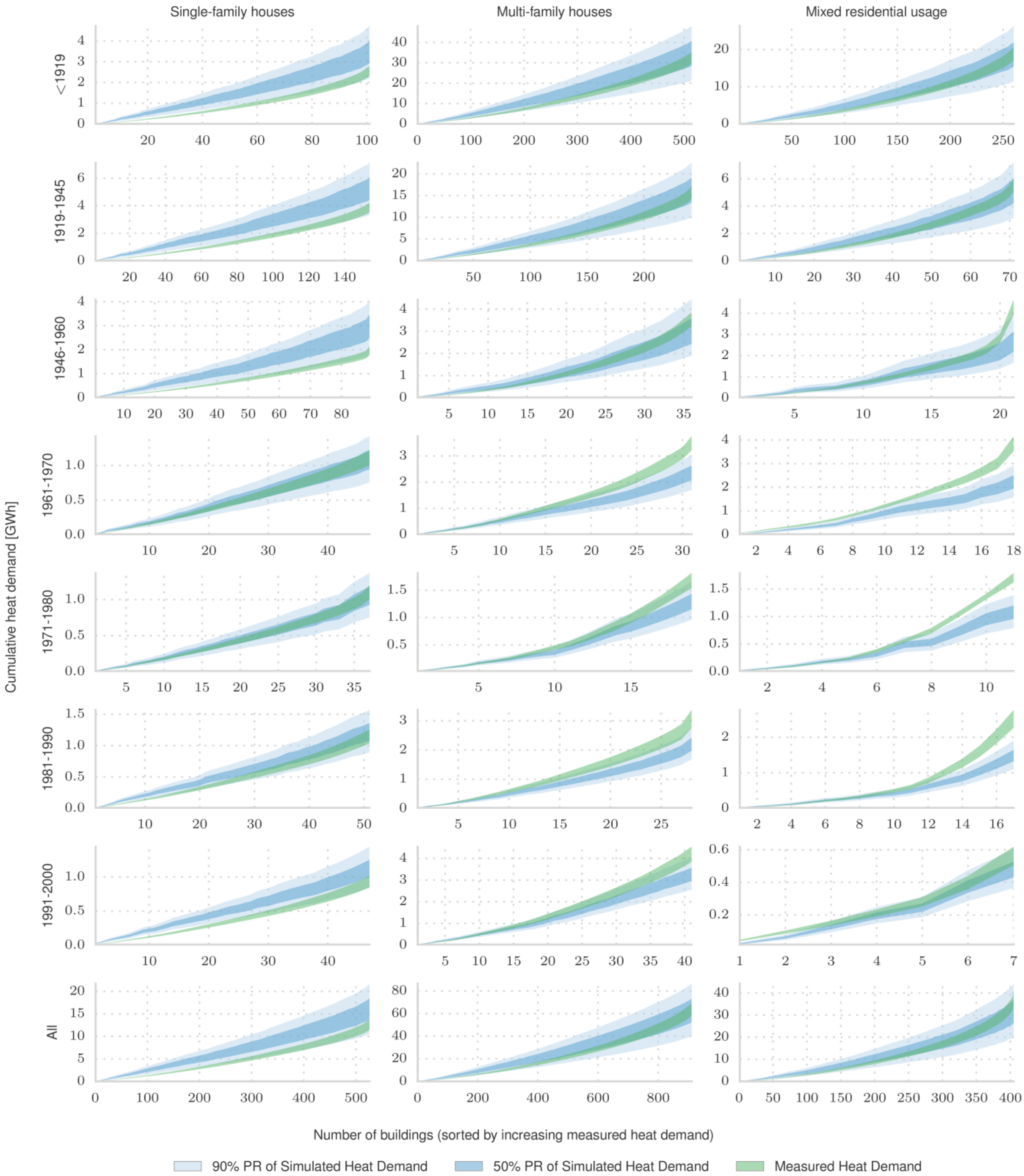


Figure C.3: **Cumulative space heating and warm water demand of simulated (blue) and measured (green) heat demand. The 50% and 90% percentile range (PR) of simulated heat demand, indicating the interval between the 25% and 75%, respectively 5% and 95% percentiles are shown.**

C.4 ZERNEZ

The community of Zernez is located in the Swiss Alps at an elevation of roughly 1500-meter elevation. In Froemelt & Hellweg [3] (Chapter 3), the model of Saner et al. [4] was validated with buildings located in Zernez. For completeness we also evaluated our model with this dataset.

Figure C.4 shows the overview of the measured heat demand normalized by the building volume for different construction periods and building types. Thereby the same categories as in Chapter 4 are used. In Figures C.5 and C.6 it can be seen that the fit of the model is good for construction periods after 1960 and not so good for buildings built before 1945. In our model we estimate physical properties of building based on the construction period from Wallbaum et al. [5]. It is likely that the cohort of buildings used in Wallbaum et al. considerably deviates from the buildings found in Zernez. Furthermore, old buildings, especially in rural locations, can include unheated spaces such as old stables within the building footprint. This can lead to a significant overestimation of our model.

In Figure C.7 the simulated heat of both our model and the model of Saner et al is compared to the measured heat demand. Due to the low sample size it is impossible to make a statement which model performs better.

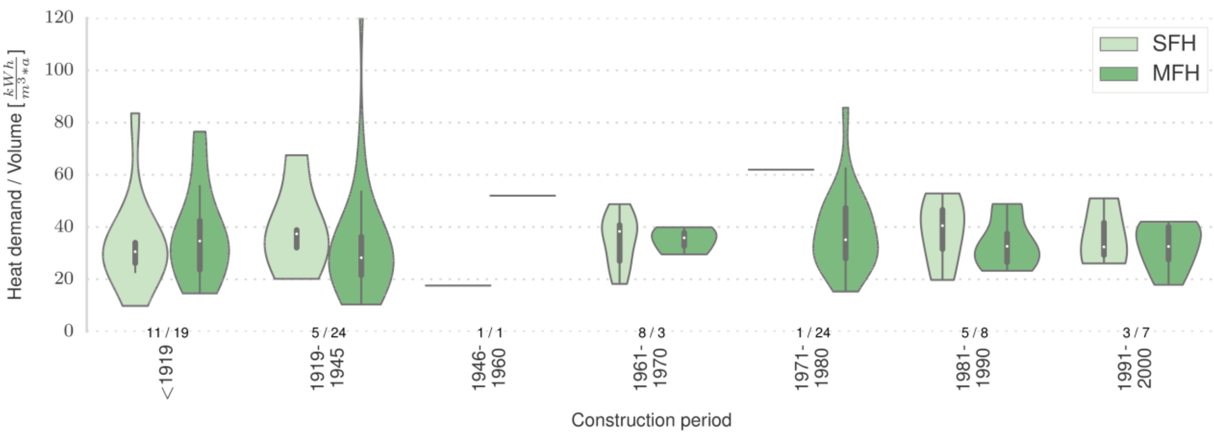


Figure C.4: **Violin plots of measured heat consumption of space heat including warm water normalized by the building volume above ground for different construction periods and the building types single-family houses (SFH), multi-family houses (MFH), and mixed residential usage (MIXED). Numbers below the violins represent the sample size.**

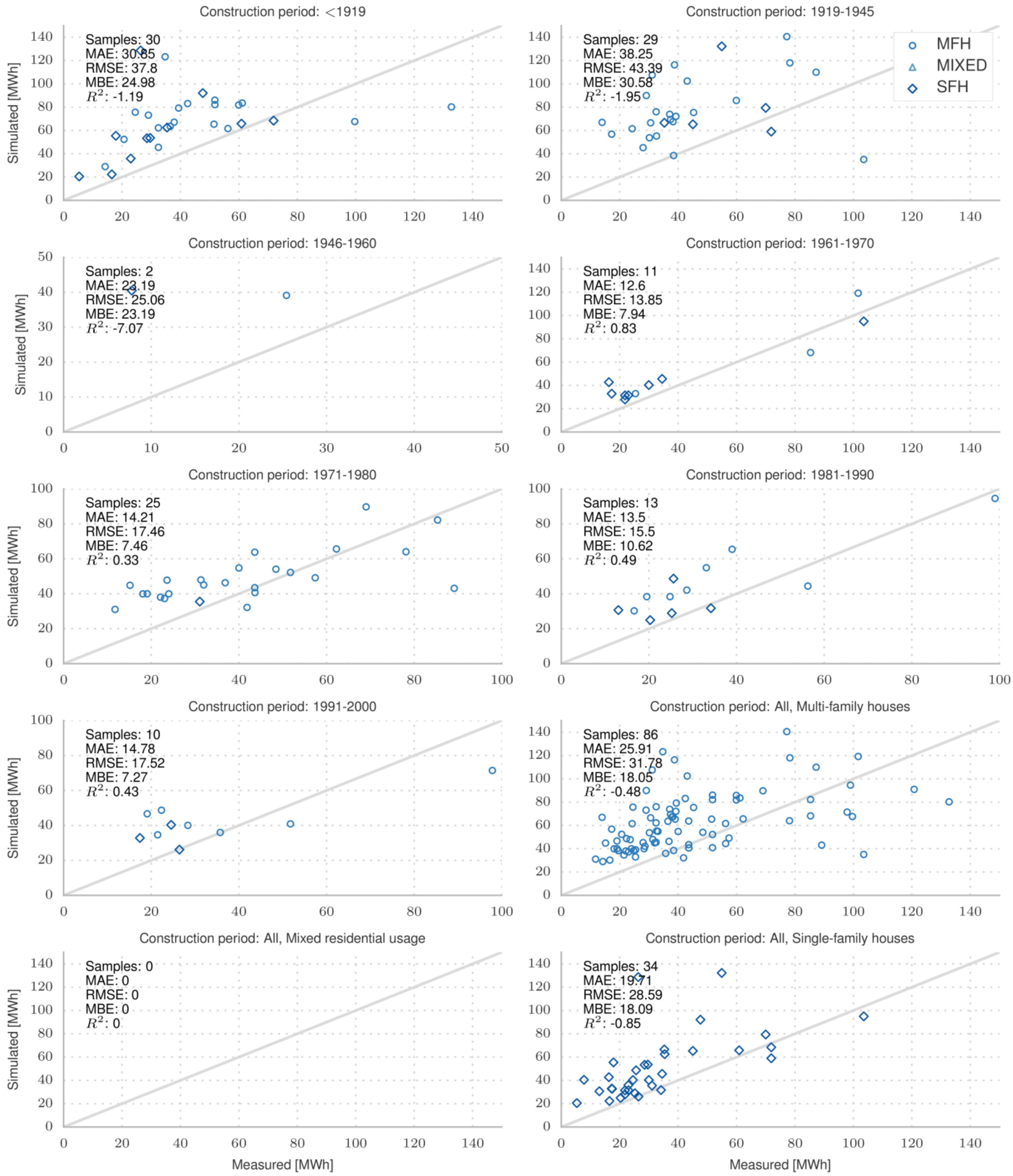


Figure C.5: Comparison of simulated and measured space heating demand for different construction periods as well as the different building types single-family houses (SFH), multi-family houses (MFH), and mixed residential usage (MIXED).

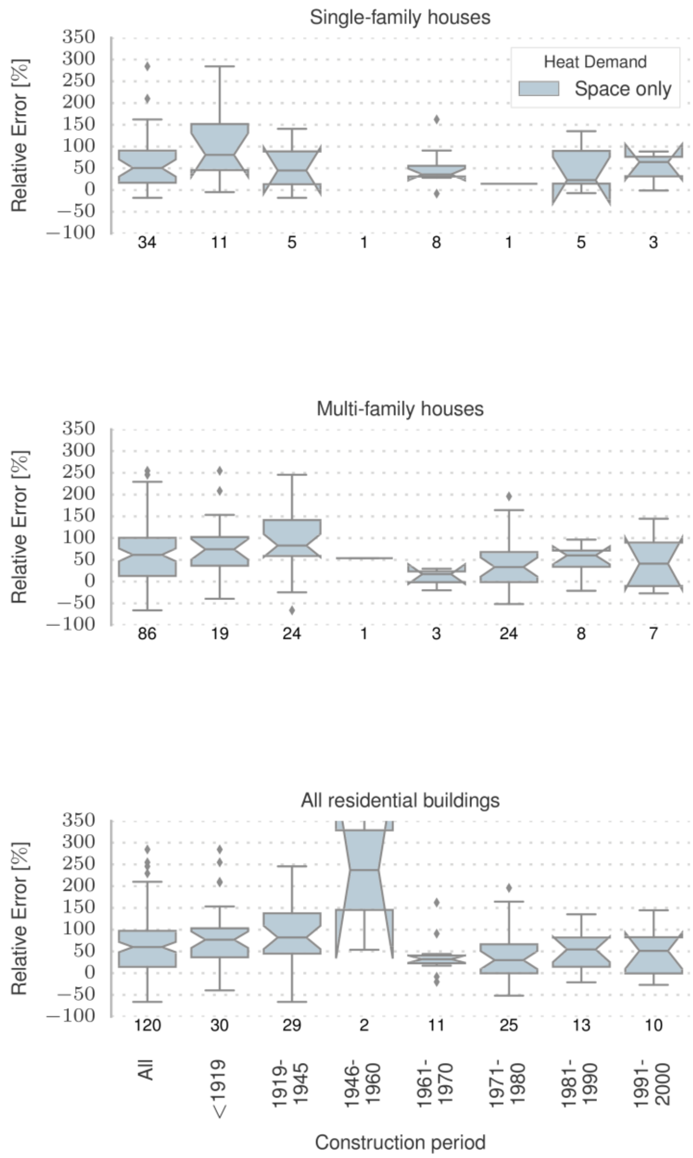


Figure C.6: **Relative error of simulated heat demand (only space heating and space heating + warm water) to the measured consumed energy demand for different building types and construction periods. Sample size *n* is given below the plots.**

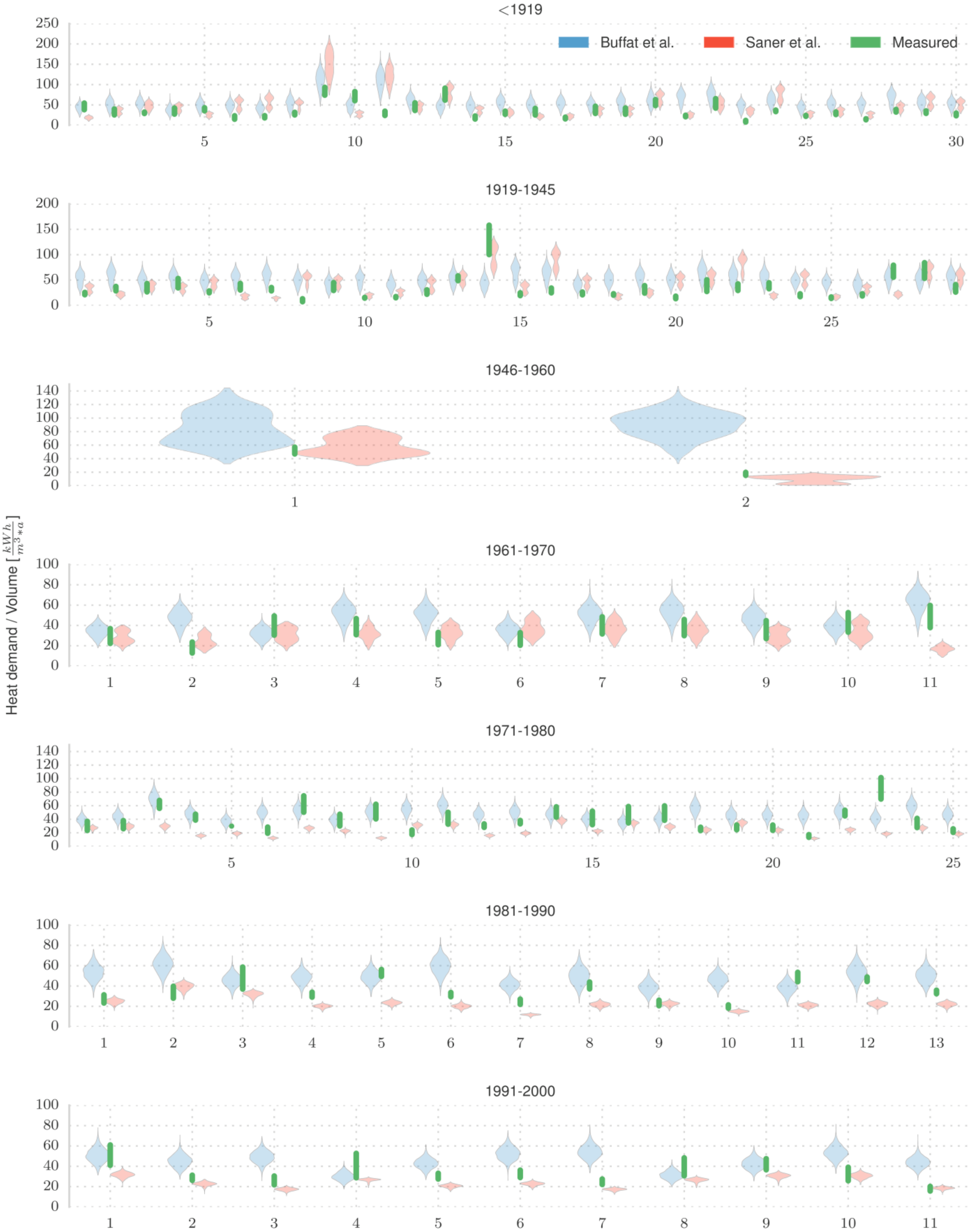


Figure C.7: Comparison of simulated space heat demand of this work, simulated demand from Saner et al. [4] validated in Froemelt & Hellweg [3] (Chapter 3) and measured heat demand for the village of Zernez.

C.5 Building Matching

This section describes the implemented process to match FRBD building with building foot-prints.

C.5.1 Data Sets

C.5.1.1 Swissboundaries3D

The Swissboundaries3D [6] dataset contains the area geometries of each municipality in Switzer-land. Each municipality has an assigned unique id (FOSNR).

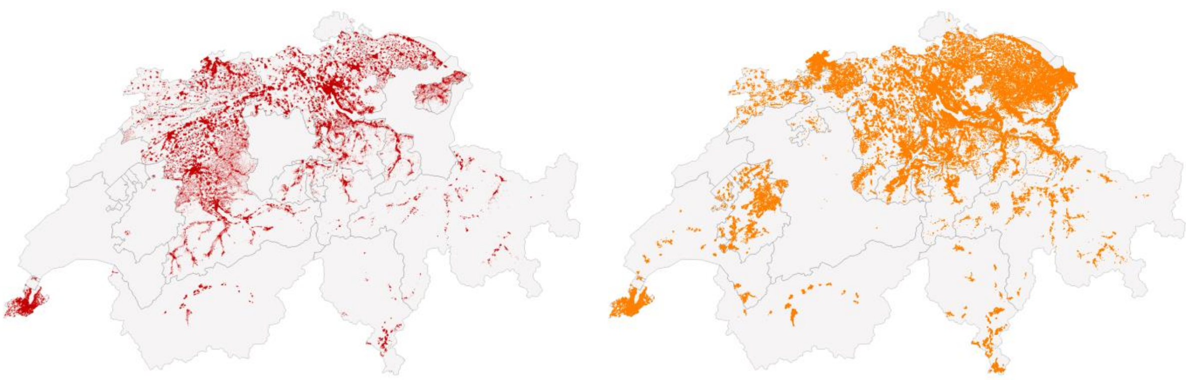
C.5.1.2 Historic Municipality Mutations

The Swiss Statistical Office provides a regularly updated dataset in XML format with all historic municipality mutations since 1960. As municipalities can merge or split, the municipality numbers between datasets created at a different time need not be the same. Therefore, we were required to update the municipality numbers in all datasets to the current municipality number according to the historic municipality mutations. This allows grouping all datasets by the current municipality numbers and thus parallelizing the matching process.

C.5.1.3 Building Footprints

The cadastral survey building footprints data set contains polygons for every recorded building as well as the attributes EGID, FOSNR and date of change. The EGID key can be used to directly combine the building with FRBD data points if available. However, as seen in Figure C.8a, the EGID is not used in large parts of Switzerland. From 2,990,415 building footprints 35.1% have a value for the EGID.

We also include building footprints from OpenStreetMap and the SwissTLM dataset. These footprints do not have an EGID attribute.



(a) Buildings (b) Building entrances

Figure C.8: Presence of EGID in the building footprint and building entrances dataset.

C.5.1.4 Lots

For each lot the polygon geometry as well as the attributes LOTNR, FOSNR and date of change are available. From 3,426,153 lots 98.4% have a value for the LOTNR.

C.5.1.5 Building Entrances

For each entrance of a building a dataset with point coordinates of the entrances and the attributes EGID, FOSNR and date of change are available. Multiple building entrance points can be located on the same building polygon, as a building can have multiple entrances. The EGID attribute can be used as link to the FRBD data points to the building footprint the entrance point is located in. As seen in Figure C.8b the EGID value is not present in every region of Switzerland. From 2,134,984 building 48.29% have a value for the EGID.

C.5.2 Attributes

This section discusses the relevant attributes of the used datasets shown in Figure C.9.

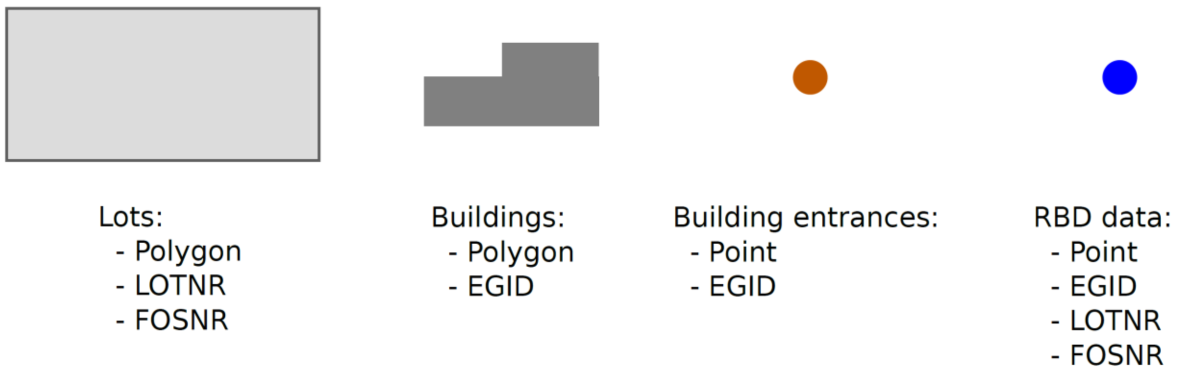


Figure C.9: Used data sets and relevant attributes.

C.5.2.1 FOSNR

Each municipality in Switzerland is represented with a unique identification number (FOSNR). When municipalities split or merge the number will be updated. However, no merge of municipalities occurred recently. The resulting municipality can either receive a completely new FOSNR assigned or reuse one of the numbers assigned to one of the original municipalities.

C.5.2.2 LOTNR

Each real estate in Switzerland has an assigned number (LOTNR). A municipality can be divided into several zones. This can have historic reasons, such as when municipalities merged in the past, the old LOTNR are kept and each zone represents an old municipality.

C.5.2.3 EGID

Every building in Switzerland used for habitation is assigned a unique identification number (EGID) by the Federal Statistical Office.

C.5.3 Handling Data Inconsistencies

Errors or inconsistencies in the data can occur for multiple reasons. Reasons can be that the datasets are managed independently. Thus when a building changes, e.g. when a new building is created or an existing extended, this change can be reflected already in one but not necessary all datasets. Further the definition of a building in the cadastral survey is not exactly the same as in the FRBD as the FRBD consists mainly from buildings with habitation use. As the cadastral survey is not completely covered digitally in all regions cadastral buildings might be missing. Depending on the organization of the municipality the data is collected by another person than entered in the system. The data might even be first collected handwritten before entered in the system. Thus errors in the process of collection and input can occur.

We assume that persons not familiar with the data are less able to identify errors (both in the original data or the detection of typographic errors while inputting the data into the system). E.g. if a person needs to enter every time the same FOSNR, he will more likely detect a typographic error in the FOSNR he entered. Similar for postal codes or street names a person familiar with the region he lives is more likely to recognize if the street name is wrong or not existent inside the postal code area. However, for humans it is difficult to check coordinates thus we assume it is more likely that typographic errors occur in coordinates.

Errors can be detected by the system through automatically validating the input data. For some attributes this can be done easily. For example, the municipality number can be checked against the municipality of the person is working for entering the data. Other attributes are much harder to validate. For example, it is impossible to validate that FRBD coordinates are within the corresponding building footprint without access to a digital version of this building footprint.

In order to handle data inconsistencies and errors the FRBD points and buildings are matched with different matching strategies. Each of the strategy tries to find if it can detect path between a FRBD point and a building by matching attributes or spatial relationships.

Typographic errors can be detected by string matching. Various algorithms exist today for matching strings [7]. For our cases we decided to measure the edit distance between two strings using the Damerau-Levenshtein [8] distance. The Damerau-Levenshtein distance measures the minimum number of operations needed to transform a string to another. Supported operations are the insertion of a character at any position, the deletion of a character, the substitution of a character with another and the transposition of two adjacent characters. An edit distance of 1 allowed operation or for coordinates an absolute distance of at least 10 meter was defined as typographic error.

C.5.4 Method

Figure C.10 shows an overview of the implemented architecture to match the data. First, the FOSNRs of all input datasets are updated to the current FOSNR. Then the data is partitioned by the FOSNRs to allow parallelized computation. The workflow was due to the large amount of data (around 9 GB) that needed to be processed.

C.5.4.1 Pre-Processing

Using the swissboundaries3D dataset each building footprint and building entrance is assigned the current FOSNR. If building footprints overlap multiple municipalities they are assigned to the municipality they overlap with the most. Then the FOSNRs are updated with the historic FOSNR dataset to the FOSNR of the current municipality. The FOSNRs of the FRBD dataset were only updated with the historic FOSNR dataset, because it was assumed that the coordinates of the cadastral datasets can be trusted more than the coordinates of the FRBD dataset.

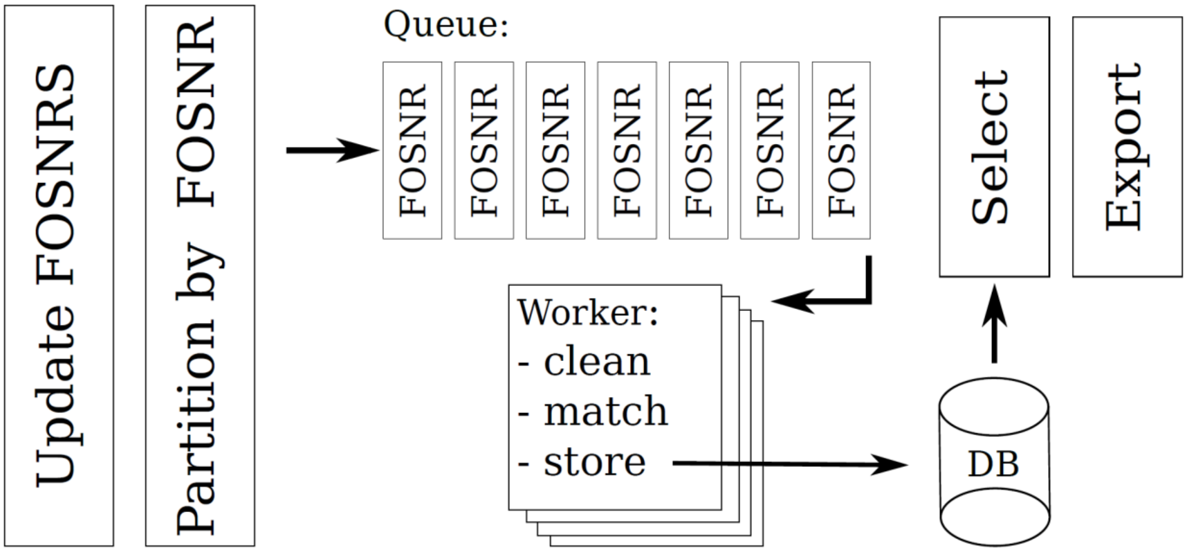


Figure C.10: **Workflow of the FRBD matching process.**

C.5.4.2 Matching Strategies

The data model allows different possibilities to match FRBD data points with buildings. Errors in the data prevent successful matches or lead to erroneous matches. Different matching strategies were implemented using different error correction methods. These strategies are presented in this section.

Building Footprint + EGID

As seen in Figure C.11 in this strategy all FRBD points are matched to buildings with a matching EGID attribute. A FRBD point needs be inside a matching building.

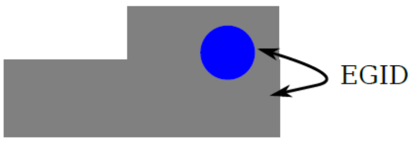


Figure C.11: **Matching strategy for buildings with matching EGID.**

Building Footprint + Building Entrance + EGID

In this strategy FRBD data points are linked to buildings with the help of the building entrances as seen in Figure C.12. The EGID of the FRBD point and the building entrance point need to match and both the FRBD point and the building entrance point need to be inside the same building.

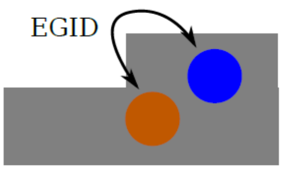


Figure C.12: **Matching strategy over building entrances with matching EGID.**

Mincost Distance

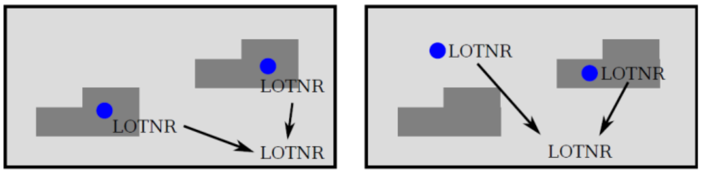


Figure C.13: **Matching strategy using a min cost assignment.**

This strategy does not rely on the EGID to match buildings. It is applied if the number of buildings on a lot match the number of buildings assigned to this lot. If multiple lots with the same LOTNR exist in the same municipality, a FRBD building is assigned its nearest lot with the same LOTNR. Min cost matching, also known as assignment problem, is used to match the buildings. Min cost matching can be solved with the Hungarian method in polynomial time [9]. The costs are in our case the distance between the buildings and the FRBD points. Figure C.14 shows a real case where the min cost matching is superior to a nearest neighbor matching with a spatial join as in the case of a spatial join two buildings would be matched to the same building.

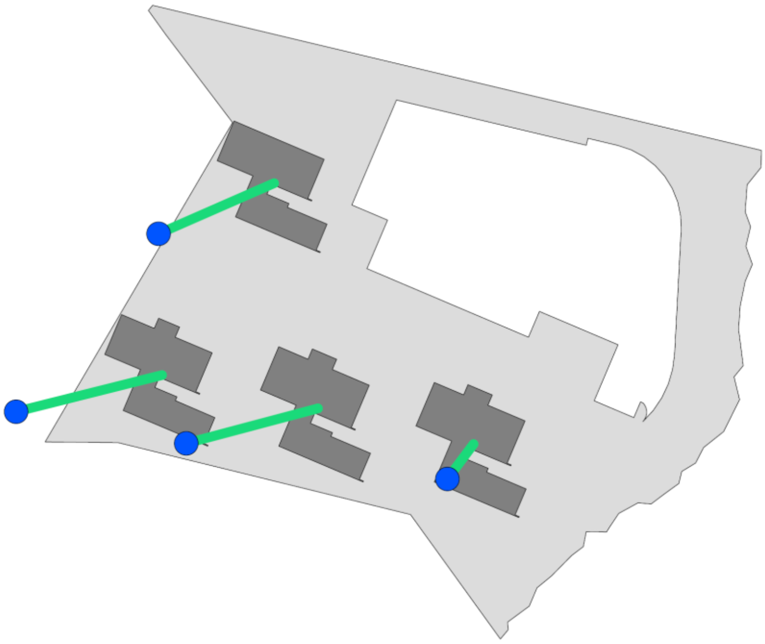


Figure C.14: **Mincost matching recovering faulty coordinates.**

Error Aware Building Footprint + EGID

With the building EGID matching strategy two sources of errors can occur as shown in Figure C.15. The coordinates of a FRBD point can lie outside a building with a matching EGID or a FRBD point can be on a building with a different EGID. In this strategy, for each FRBD point outside a building with a matching EGID it is checked if a typographic error in the coordinates can be detected so that the FRBD point would land within the building. A typographic error is allowed to occur in either the X or Y coordinate but not both. If a FRBD point is inside a building with a different EGID, then it is checked if the difference can be explained by a typographic error.

Figure C.15: Matching strategy with typographic error corrections for buildings with matching EGID. Red arrows indicate that EGIDs do not match. This relation is only valid if the mismatch of the EGIDs can be explained by a typographic error.

Error Aware Building Footprint + Building Entrance + EGID

Two kinds of errors are detectable when FRBD points are matched to building footprints over a common EGID and building entrance points. The FRBD point can be inside a building but none of the FRBD points of the building have the same EGID as the FRBD point or a FRBD point can be outside a building with a building entrance point with the same EGID. This situation is shown in Figure C.16. For the first case, a match is established if a typographic error in the EGID can be found. In the second case, a match is established if a typographic error in the coordinates of the FRBD point would move the point within the building.

Figure C.16: Matching strategy with typographic error corrections for building entrances with matching EGID. Red arrows indicate that EGIDs do not match. This relation is only valid if the mismatch of the EGIDs can be explained by a typographic error.

Same LOTNR and Typographic Error in Coordinates

Figure C.17 shows the case that a FRBD point is outside a lot with a matching LOTNR and an unequal number of FRBD points and building matching to this lot. In this case, it is checked if a typographic error in the coordinates can be found that would move the point within a building on the lot. As only an edit distance of 1 is allowed for both coordinates buildings need to be either on the same X or Y axis.

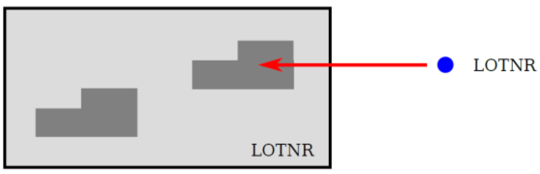


Figure C.17: Matching strategy for FRBD points with erroneous coordinates.

Spatial

In this strategy FRBD points are matched to buildings according to their spatial relationship. The matching is divided in different quality levels according to if the LOTNR of the FRBD points equals the LOTNR of real estate the building is on. If a FRBD point is inside a building and the LOTNR are equal the quality is 1. If a FRBD point is outside a building it is matched to the nearest building within 5 meters. If the LOTNR matches, the match quality is 2. For FRBD points without a matching LOTNR, the quality is 3 for FRBD points inside a building and 4 for an FRBD point within 5 meters to a building.

C.5.5 Best Matching Selection

For each FRBD points all detected matchings are ranked according to our assumptions which errors are more likely. For each FRBD point the matching with the highest ranked strategy is selected. The ranking of the different strategies from best to worst is as follows:

1. Building Footprint + Building Entrance + EGID
2. Building Footprint + EGID
3. Mincost Distance
4. Error Aware Building Footprint + Building Entrance + EGID
5. Error Aware Building Footprint + EGID
6. Same LOTNR and Typographic Error in Coordinates
7. Spatial

C.6 References

[1] SIA, *SIA 380/1: Thermische Energie im Hochbau*, Zurich, Switzerland: Swiss Society of Engineers and Architects, 2009.

[2] Kanton Luzern (Umwelt und Energie), *Energiespiegel - Methodik und Diskussion*, 2015.

[3] A. Froemelt and S. Hellweg, *Assessing Space Heating Demand on a Regional Level: Evaluation of a Bottom-Up Model in the Scope of a Case Study*, *Journal of Industrial Ecology*, vol. 21, no. 2, pp. 332–343, 2017.

[4] D. Saner, N. Heeren, B. Jäggi, R. A. Waraich, and S. Hellweg, *Housing and Mobility Demands of Individual Households and their Life Cycle Assessment*, *Environ. Sci. Technol.*, vol. 47, no. 11, pp. 5988–5997, 2013.

[5] H. Wallbaum, N. Heeren, M. Jakob, G. Martius, and N. Gross, *Gebäudeparkmodell*, Stadt Zürich, Zurich, Switzerland, 2010.

[6] Bundesamt für Landestopografie (swisstopo), *swissBOUNDARIES3D*, 2014. [Online]. Available: http:%0A//www.swisstopo.admin.ch/internet/swisstopo/en/home/products/landscape/%0AswissBOUNDARIES3D.html. [Accessed: 01-Jan-2014].

[7] G. Navarro, *A Guided Tour to Approximate String Matching*, *ACM Computing Surveys*, vol. 33, no. 1, pp. 31–88, 2001.

[8] F. J. Damerau, *A technique for computer detection and correction of spelling errors*, *Communications of the ACM*, vol. 7, no. 3, pp. 171–176, 1964.

[9] J. Munkres, *Algorithms for the Assignment and Transportation Problems*, *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–8, 1957.

Appendix D

Using Data Mining To Assess Environmental Impacts of Household Consumption Behaviors

D.1 General Remarks

D.1.1 Open-Source Software

All computations for this study were performed with the following open-source software:

PostgreSQL 9.4 [1]

Python 3.6.1 [2]

Python packages:

Brightway2 (2, 1, 1) [3]

SciKit-learn 0.18.1 [4]

NumPy 1.12.1 [5]

SciPy 0.19.0 [6]

Pandas 0.19.2 [7]

psycopg2 2.7.1 [8]

SOMPY [9]

Matplotlib 2.0.0 [10]

Seaborn 0.8.0 [11]

D.1.2 Used Categorization

The applied categorization and especially the subdivisions, subcategories and different aggregation levels are based on the structure of the Swiss Household Budget Survey (HBS) data [12] which itself is based on the United Nation's "Classification of Individual Consumption according to Purpose" (COICOP) [13]. The HBS-structure can be found in the supplemental EXCEL-file[1]. Throughout the study, we mainly refer to the following 11 main categories: Food, Restaurants&Hotels, Clothing, Housing, Furnishings, Health, Transport, Communication, Recreation, Education and Others. These main categories correspond to the top level of the HBS-structure and COICOP, except for some additional merges for Food and Others:

- Food comprises the two categories "Food and non-alcoholic beverages" as well as "Alcoholic beverages, tobacco and narcotics".
- Others encompasses "Miscellaneous goods and services", "Insurance" (in contrast to COICOP, this category is separated in HBS from "Miscellaneous goods and services") as well as "Fees".

For the other top categories, the official description of COICOP is as follows:

- Restaurants&Hotels: "Restaurants and hotels"

[1] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

- Clothing: "Clothing and footwear"
- Housing: "Housing, water, electricity, gas and other fuels"
- Furnishings: "Furnishings, household equipment and routine household maintenance"
- Health: "Health"
- Transport: "Transport"
- Communication: "Communication"
- Recreation: "Recreation and culture"
- Education: "Education"

For the subcategories in Figure 5.4 in Chapter 5, we used the second lowest level according to HBS-structure.

D.2 Pre-Processing of Consumption Data

This section explains in detail how new variables that are needed for the clustering as well as for the life cycle assessment (LCA) were created based on the HBS-data [12]. The procedure is briefly outlined in section 5.2.2 in Chapter 5.

D.2.1 Creation of Dummy Variables and Count-Statistics

The Python-implementations of regression and clustering techniques [4] that were used require continuous input values. Therefore, categorical variables that feature distinct, but unordered categories need to be transformed to so called "dummy variables", also known as "one-hot encoding" or "one-of-K encoding" [4]. This means that for each categorical value, a new binary variable that takes 1 if the sample features the respective categorical value and 0 otherwise is introduced. For instance, a household living in the canton of Zurich gets a 1 for the dummy variable "Zurich" and zeros for all other canton dummy variables.

The following categorical household variables were converted to dummy variables (categorical values are in parentheses):

- Major region (Lake Geneva, Espace Mittelland, Northwestern Switzerland, Zurich, Eastern Switzerland, Central Switzerland, and Ticino)
- Language region (German- / Rhaeto-Romanic-speaking, Italian-speaking, French-speaking)
- Canton (Aargau, Bern, Geneva, Lucerne, St. Gallen, Ticino, Vaud, Zurich, other canton)

Based on the data that were available for the members of a specific household, the following count-statistics were computed:

- Females aged between 0 and 4
- Females aged between 5 and 14
- Females aged between 15 and 24
- Females aged between 25 and 34
- Females aged between 35 and 44
- Females aged between 45 and 54
- Females aged between 55 and 64
- Females aged between 65 and 74
- Females at the age of >75
- Males aged between 0 and 4
- Males aged between 5 and 14
- Males aged between 15 and 24
- Males aged between 25 and 34
- Males aged between 35 and 44
- Males aged between 45 and 54
- Males aged between 55 and 64
- Males aged between 65 and 74
- Males at the age of >75
- Number of foreigners in the household
- Number of Swiss persons in the household
- Number of divorced persons in the household
- Number of married persons in the household
- Number of unmarried persons in the household
- Number of widowed persons in the household

Please note that the chosen age cohorts correspond to the age groups used by the Federal Statistical Office for their statistical analyses based on the HBS [14].

D.2.2 Estimating Full Detail of Utility Costs

As mentioned in section 5.2.2 in Chapter 5, an in-depth environmental assessment requires full detail of utility costs for each household. Full details of utility costs comprise expenditures for water supply, wastewater collection, refuse collection, electricity, and heating fuels. However, this is not available for all households taking part in the HBS. The following subsections will describe the applied approach to impute missing information on dwelling extra costs. Subsection D.2.2.1 (*Description of the Procedure*) will start with a textual description on the applied procedure, while subsection D.2.2.2 (*Details of the Modeling Framework*) will provide information on technical details.

D.2.2.1 Description of the Procedure

For imputing missing information, the following steps were completed one after the other (details will be given further below):

1. Missing water supply costs were estimated based on wastewater information if latter was available and vice versa
2. Modeling of total utility costs
3. Modeling heating costs as shares in total utility costs
4. Modeling wastewater costs as shares in total utility costs
5. Modeling water costs as shares in total utility costs
6. Modeling electricity in kilowatt-hours
7. Modeling refuse collection as amounts of waste bags
8. Computing quantities or expenditures based on the modeled estimates

Step 1: In case information on wastewater expenditures was available, but not on water supply, then water supply costs were estimated by assuming that the amount of consumed water equals the amount of discharged wastewater. This assumption is plausible given the price formation [15] for water and wastewater in Switzerland which indeed assumes the same amounts for water and wastewater. However, for applying this assumption, wastewater expenditures needed to be first converted to quantities (see section D.2.3 for details). Once the amounts were set equal, the water supply costs were then computed by means of the prices given in section D.2.3. Understandably, the same procedure was applied vice versa if water supply data was available, but no information given on wastewater.

Step 2: It is important to know that in Switzerland, many tenants (and depending on the municipality also some homeowners) might not have full insight into the breakdown of their utility costs, but very often they have a specification of a part of these costs. All other utilities are paid in a lump sum (in German called "Nebenkosten pauschal", and hereafter called "utility lump sum"). This means that the missing details of utility costs are "hidden" in the aggregated form of that utility lump sum (note that the utility lump sum comprises different costs for different households). This is at least some piece of information that can be used to model missing data in the following. Indeed, 85% of all households in the HBS have either full details available or have at least information on a part of the utility costs as well as on the utility lump sum. However, in order to use the information given in the utility lump sum for further modeling, this utility lump sum needed to be estimated for the remaining 15% of HBS-households in a first step. For these households, we thus modeled the total dwelling extra costs and then subtracted the known utility costs (see subsection D.2.2.2 (*Details of the Modeling Framework*) for the details of the modeling).

Steps 3-5: In order to use the hidden information in the utility lump sum to impute missing water supply, wastewater collection and heating costs, these three housing categories were modeled one after the other as respective shares in the total of utility costs (= sum of utility lump sum, water supply, wastewater collection, electricity, heating, and maintenance costs).

Step 6: Even though electricity demand could have been estimated also as shares in total utility costs, we modeled electricity directly as amounts of bought kilowatt-hours for two reasons: First, most households have separate electricity bills available which renders the situation easier than for other housing categories. Only 3% of all HBS-households have no information on power consumption. This large amount of training data encourages for a direct modeling of electricity. Second, electricity was modeled in kilowatt-hours instead of expenditure in Swiss Francs since detailed price data was available (see section D.2.3) which allows for de-coupling the dependency of expenditures on local electric power companies.

Step 7: Expenditures for refuse collection are not part of the utility lump sum and needed thus to be estimated separately. Furthermore, there are some municipalities in which no fees are demanded for waste treatment, which of course does not mean that these households do not produce waste. Therefore, waste production was modeled as amount of waste bags and not as expenditure in Swiss Francs (see section D.2.3 for the conversion of expenditures to waste bags).

Step 8: Finally, the life cycle assessment step will need the conversion of expenditures into the respective functional units. Therefore, water and wastewater costs (estimated in steps 1, 4 or 5) were translated into cubic meters and heating costs (step 3) into mega-joules. In order to provide consistent data for subsequent computations, electricity demand (step 6) and amount of waste bags (step 7) produced were back-calculated to expenditures (refer to section D.2.3 for more details regarding the conversion from expenditures to quantities).

Please note that only the results of steps 1 and 2 entered the modeling in the subsequent steps 3-8, while the results of steps 3-8 did not interfere with each other.

D.2.2.2 Details of the Modeling Framework

Steps 2 to 7 in subsection D.2.2.1 (*Description of the Procedure*) refer to "modeling data". In the present case, this means that missing data was predicted based on given data in the HBS. For all these modeling steps, the scheme presented in Figure D.1 was applied which is described in more detail below.

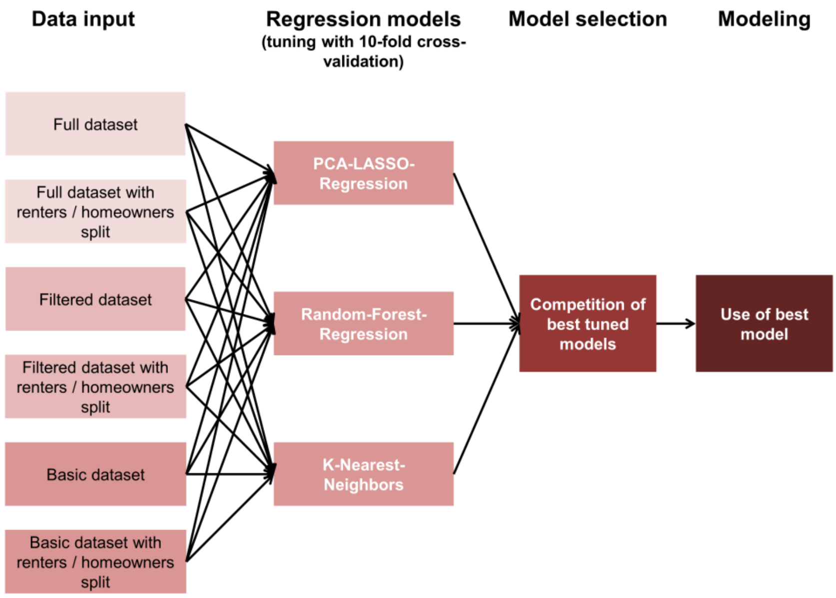


Figure D.1: **Modeling scheme that was applied for imputing missing details on utility costs.**

Data input: The full HBS-dataset with all data on the lowest level as well as on aggregated levels of course features many multicollinearities. However, not every regression model is able to handle these multicollinearities [16]. Therefore, three different datasets were created based on the HBS: the full HBS-dataset, a filtered dataset which corresponds to the dataset used for training the self-organizing map (see filtering step in section D.3.1) as well as a basic dataset which only consists of household characteristics and housing-relevant expenditures. Each of these three datasets were then further subdivided into the respective dataset as a whole and a dataset in which renter-households are distinguished from homeowners. This resulted in a total of six datasets which entered the model selection process of the regression models.

Regression models: Three different regression model types were then trained based on each of the different datasets. The hyperparameters of these models were tuned by means of 10-fold cross-validation [16]. In general, standardized data was used, except for Random-Forest-Regression [17] for which also the use of non-standardized data was tested.

- PCA-LASSO-Regression: LASSO [18] (least absolute shrinkage and selection operator) belongs to the generalized linear models. By introducing a regularization/penalty term (called alpha) to the least squares model, LASSO can turn non-informative features to zero. The value of alpha was determined by evaluating 1000 different alphas in the 10-fold cross-validation step. However, just as other linear models, also LASSO can become unstable if between-predictor relationships are present [16] (collinearities or multicollinearities). We therefore decided to apply principal component analysis (PCA) [19, 20] to the dataset in a preparatory step in order to transform the data into principal components which are de-correlated. Based on the variable selection ability of LASSO, we then let the LASSO choose the principal components which are important for the respective prediction.

- Random-Forest-Regression: Random-Forest-Regressors [17] are robust, do not need extensive preparation and can also capture non-linearities [16]. It shall be mentioned that the Python-Implementation of Random-Forests in the SciKit-Learn-Package [4] that was used – just as many other software programs – does not exactly correspond to the regression model presented in Breiman 2001 [17] (which uses a sort of regression model tree [16]), but applies as a base learner an improved version of a CART (Classification and Regression Tree) that was proposed by Breiman 1984 [21]. The hyperparameters (number of trees and maximum number of features considered in a split step) of the Random-Forest were tuned in a 10-fold cross-validation [16] procedure and based on the recommendations of [22–24]. Instead of using the "best" number of trees, we applied the "one-standard-deviation"-rule [16] and chose the number of trees of the model whose mean squared error was within one standard deviation from the model with the best performance. Furthermore, the scoring function for the splits in the Random Forest was set to minimizing the mean squared error.
- K-Nearest-Neighbor-Regression: A K-Nearest-Neighbor-Regression [25] using Euclidean distances was also applied to the datasets. The number of neighbors was determined again in 10-fold cross-validation [16]. Just as for the Random-Forest-Regression, the "one-standard-deviation"-rule was also applied for choosing the number of neighbors in this regression model.

Model selection: After determining the best hyperparameters for all regression models, these best-tuned models entered a model selection competition. The criteria for selecting the best model for a certain housing cost category were the following:

1. Mean squared error was used as the most important criterion.
2. As a second criterion the coefficient of determination (R^2) was also considered.
3. A visual plausibility check based on basic diagnostic plots (see Figures D.2-8) was also taken into account. It needs to be mentioned that models predicting negative values were excluded regardless of their other performance scores.

Modeling: In the following, the basic diagnostic plots as well as the performance metrics of the eventually chosen models will be presented.

- Modeling total of utility costs:
  - Chosen regression model: PCA-LASSO-Regression
  - Chosen dataset: Full dataset with split renters/homeowners
  - MSE renters: 0.0031
  - MSE homeowners: 0.0077
  - R^2 renters: 0.997
  - R^2 homeowners: 0.992

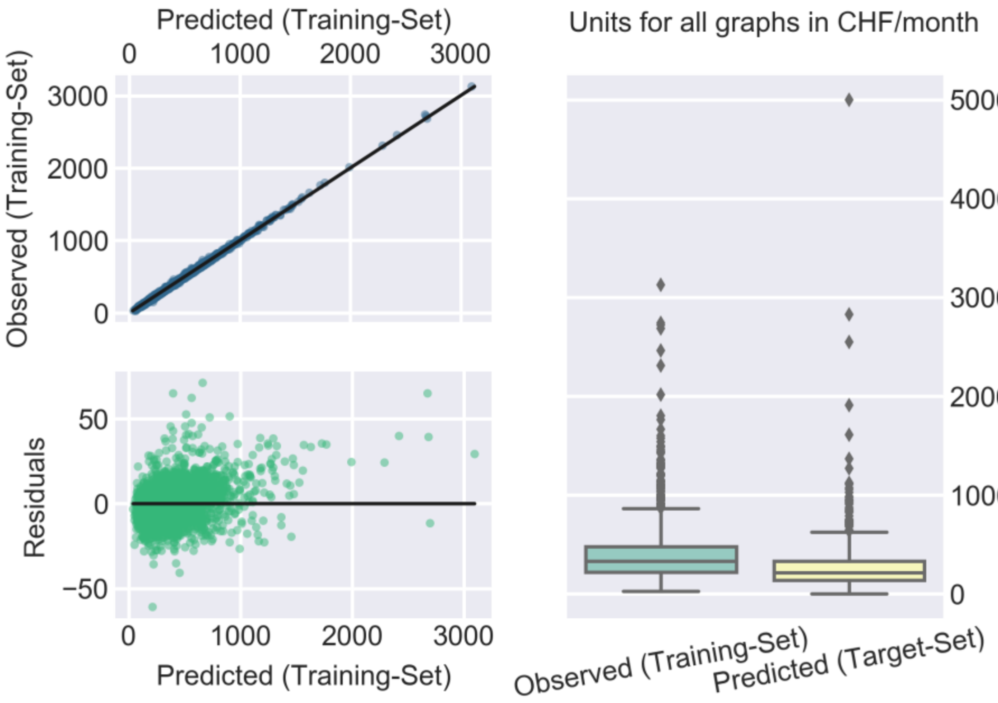


Figure D.2: Basic diagnostic plot for modeling total of utility costs; Chosen model: PCA-LASSO-Regression (full dataset, here only homeowners). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).

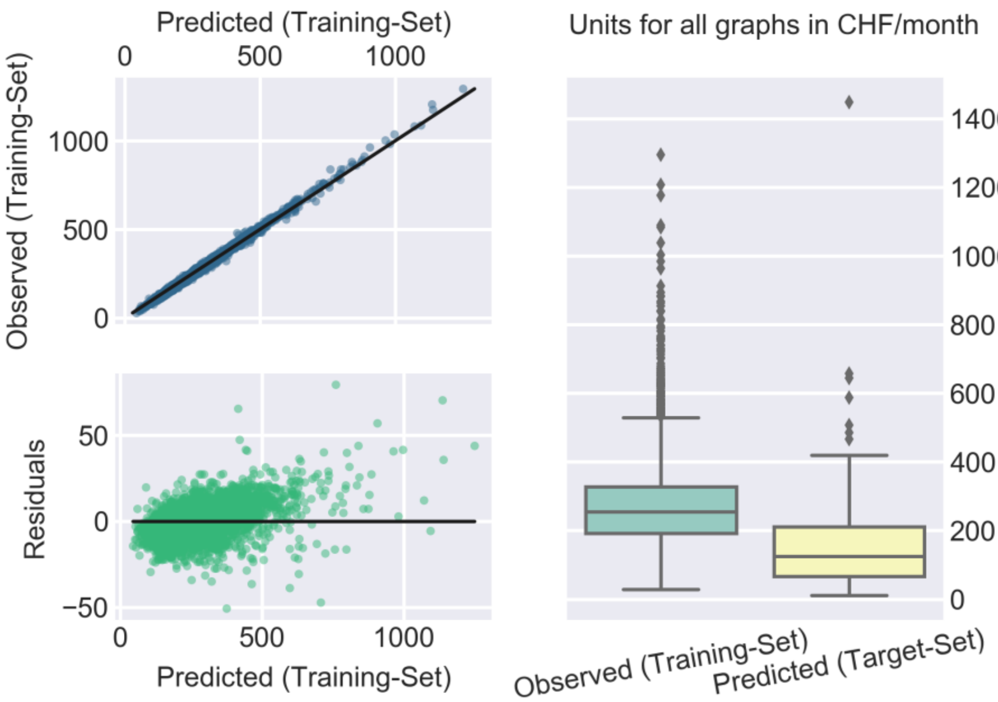


Figure D.3: Basic diagnostic plot for modeling total of utility costs; Chosen model: PCA-LASSO-Regression (full dataset, here only renters). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).

- Modeling heating costs (shares in total utility costs):
    - Chosen regression model: Random-Forest-Regression
    - Chosen dataset: non-standardized filtered dataset
    - MSE: 0.0349
    - R^2: 0.330

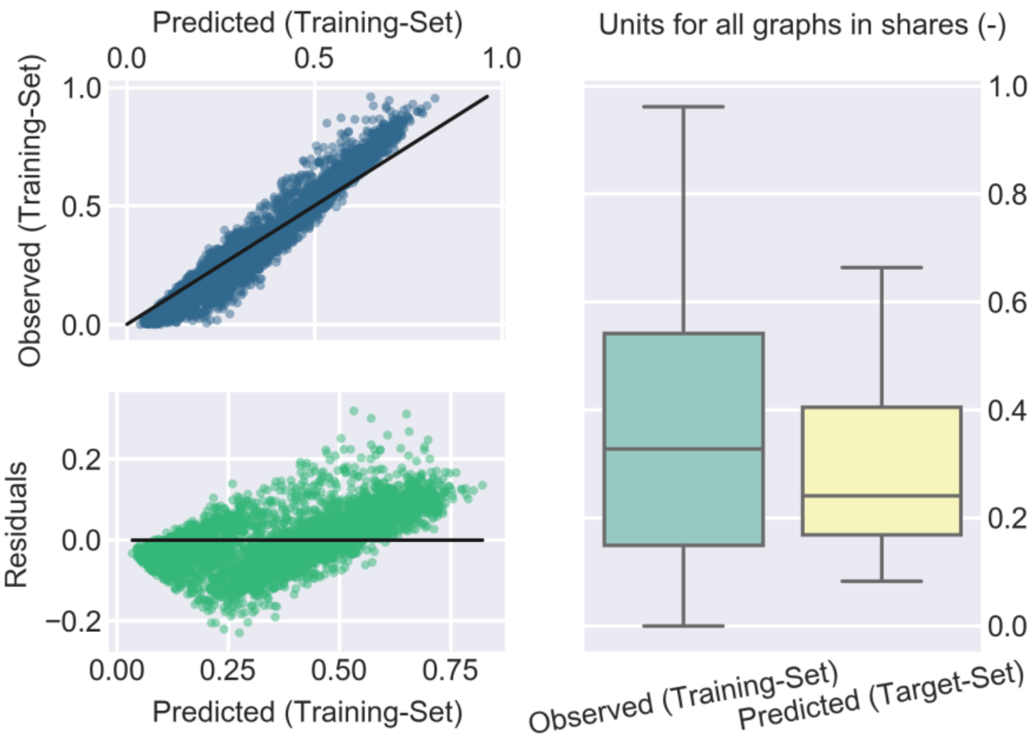


Figure D.4: **Basic diagnostic plot for modeling heating costs; Chosen model: Random-Forest-Regression (filtered dataset). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).**

- Modeling wastewater costs (shares in total utility costs):
  - Chosen regression model: Random-Forest-Regression
  - Chosen dataset: non-standardized basic dataset
  - MSE: 0.0043
  - R^2: 0.099

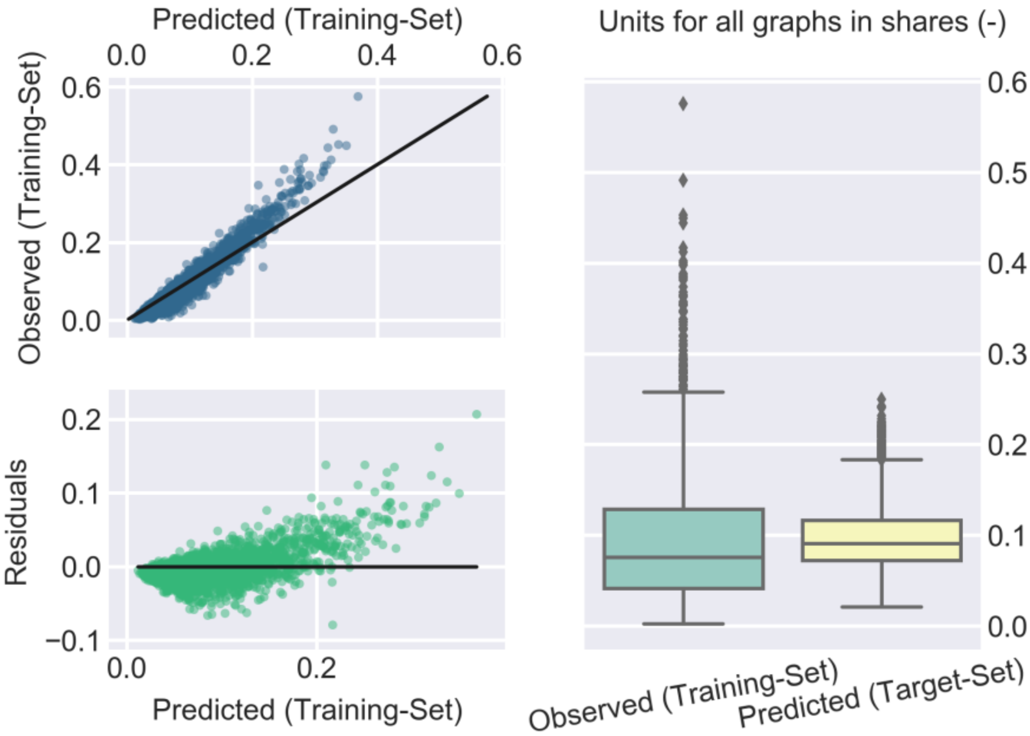


Figure D.5: **Basic diagnostic plot for modeling wastewater costs; Chosen model: Random-Forest-Regression (basic dataset). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).**

- Modeling water supply costs (shares in total utility costs):
    - Chosen regression model: Random-Forest-Regression
    - Chosen dataset: non-standardized basic dataset
    - MSE: 0.0041
    - R^2: 0.056

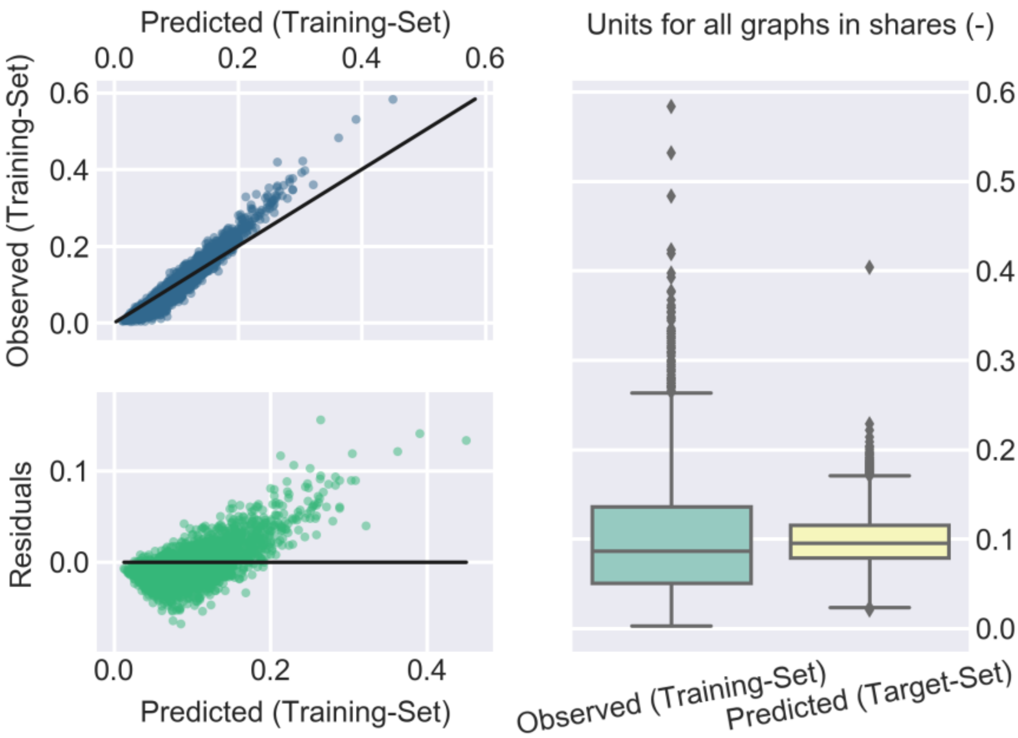


Figure D.6: **Basic diagnostic plot for modeling water supply costs; Chosen model: Random-Forest-Regression (basic dataset). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).**

- Modeling electricity demand:
  - Chosen regression model: Random-Forest-Regression
  - Chosen dataset: non-standardized full dataset
  - MSE: 26244.3
  - R^2: 0.869

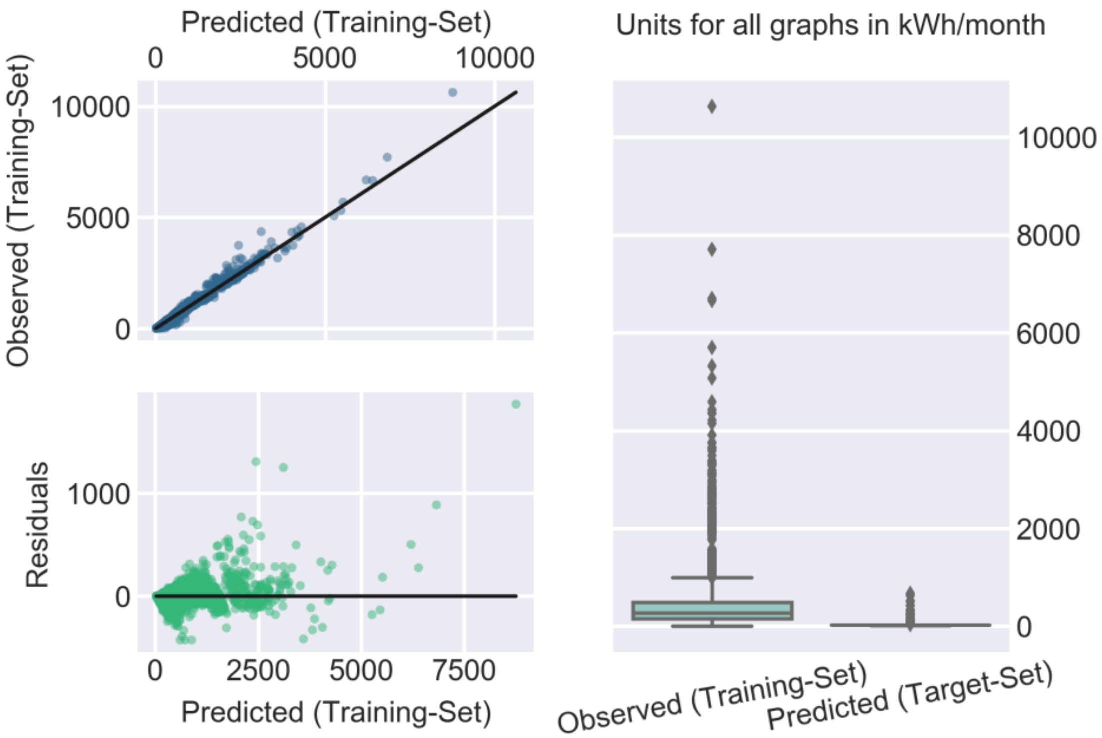


Figure D.7: **Basic diagnostic plot for modeling electricity demand; Chosen model: Random-Forest-Regression (full dataset). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).**

- Modeling waste production:
  - Chosen regression model: Random-Forest-Regression
  - Chosen dataset: non-standardized full dataset
  - MSE: 87.56
  - R^2: 0.030

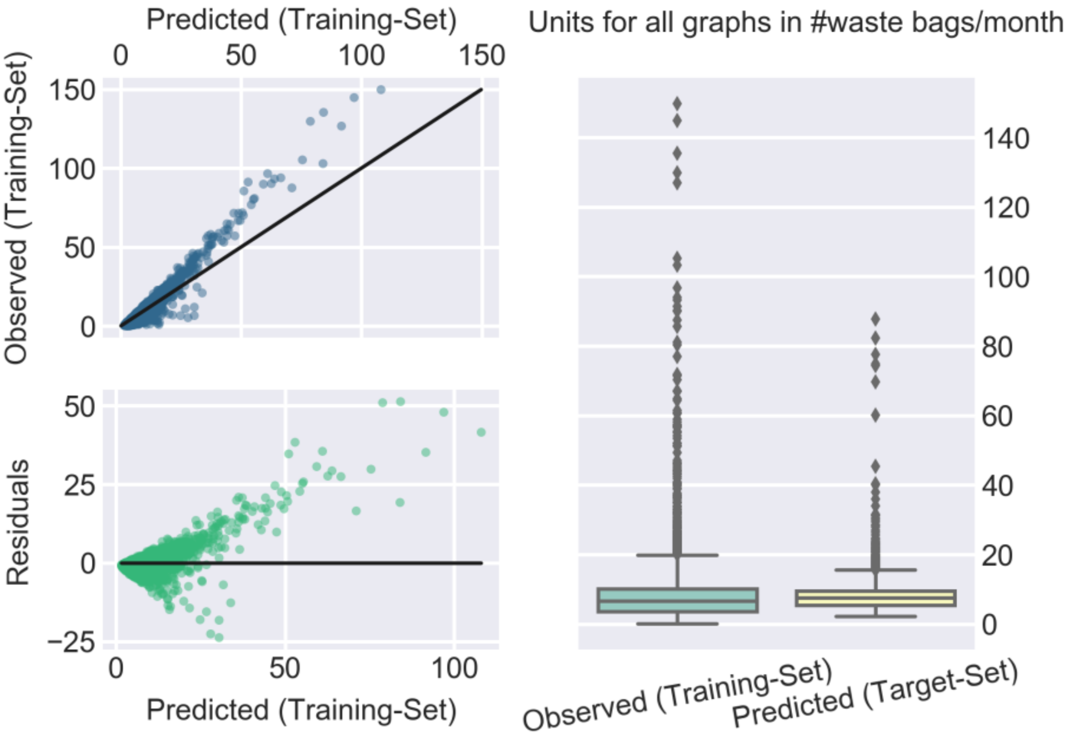


Figure D.8: **Basic diagnostic plot for modeling waste production; Chosen model: Random-Forest-Regression (full dataset). Upper left: Observed versus predicted values; Lower left: Residual plot; Right: Comparison of the distributions in the observed training dataset and in the predicted target dataset (please note that this is only meant for a visual plausibility check since these distributions need not to be equal).**

D.2.3 Conversion of Utility Costs to Quantities

The use of the process-based life cycle inventory database ecoinvent [26] to assess housing-related environmental impacts requires functional units in terms of quantities (cubic meters, kilowatt-hours, etc.) instead of expenditures in Swiss Francs. Although more information on LCA-modeling is given in section D.4, the conversion of utility costs to quantities was implicitly needed in section D.2.2 and shall thus be discussed below.

Electricity: Detailed price information was available for electricity [27]. Expenditures on electricity could thus be converted to kilowatt-hours based on year, canton and household type. It needs mentioning that Swiss power supply companies set their prices depending on the total volume of purchased electricity and on the time of the purchase. Therefore, the electricity prices strongly depend on the electricity use behavior of households. In order to be able to perform reasonable price comparisons and thus come up with representative prices, the Federal Electricity Commission (ElCom) [27] defines seven "household types". Thereby, a household type not only exhibits a total annual demand, but features also a typical diurnal load profile. However, the HBS-data only provides expenditure data and hence, does not allow for drawing conclusions about diurnal profiles. Since also the total annual electricity demand is not known initially and dependent of the price data, the corresponding household price type for a certain HBS-household was determined iteratively. Table D.1 shows the prices used for the conversion.

Table D.1: Electricity prices retrieved from [27] in centimes per kilowatt-hour (Rp./kWh). The household types are based on different characteristics and show the following annual power consumption: H1: 1600 kWh; H2: 2500 kWh; H3: 4500 kWh; H4: 4500 kWh; H5: 7500 kWh; H6: 25000 kWh; H7: 13000 kWh.

	Canton	H1	H2	H3	H4	H5	H6	H7
2009	Canton Zurich	20.12	17.46	13.50	15.61	12.81	9.72	12.18
	Canton Bern	31.46	27.62	20.94	24.68	19.94	14.88	18.26
	Canton Lucerne	32.86	29.07	22.42	25.96	21.31	14.35	18.73
	Canton St. Gallen	26.38	22.50	17.69	20.13	16.70	12.72	14.69
	Canton Aargau	25.09	21.42	16.86	18.73	15.71	12.14	14.62
	Canton Ticino	25.12	22.67	18.33	20.79	17.42	15.45	17.25
	Canton Vaud	28.64	26.71	21.48	24.26	20.58	16.83	19.86
	Canton Geneva	23.43	23.00	19.54	22.64	19.74	16.58	20.24
	Swiss Average (other cantons)	25.17	22.76	17.76	20.90	17.10	13.31	16.53
2010	Canton Zurich	21.25	18.45	14.25	16.47	13.51	10.24	12.84
	Canton Bern	31.46	27.44	20.91	24.57	19.86	14.88	18.26
	Canton Lucerne	32.86	29.07	22.42	25.96	21.31	14.35	18.73
	Canton St. Gallen	27.17	23.36	18.32	20.23	16.96	13.40	15.17
	Canton Aargau	25.10	21.41	16.82	18.67	15.65	12.31	14.57
	Canton Ticino	25.12	22.64	19.46	21.86	18.50	16.55	18.33
	Canton Vaud	28.75	26.71	21.58	24.38	20.58	16.92	19.93
	Canton Geneva	22.89	22.46	19.00	22.10	19.20	16.04	19.70
	Swiss Average (other cantons)	26.10	23.63	18.77	21.69	18.14	14.31	17.47
2011	Canton Zurich	23.60	20.69	16.50	18.60	15.74	12.45	15.01
	Canton Bern	31.69	27.89	21.93	25.17	21.12	16.25	19.49
	Canton Lucerne	31.08	27.45	21.19	24.57	20.16	13.88	17.90
	Canton St. Gallen	27.68	23.63	19.34	21.12	18.09	15.14	16.19
	Canton Aargau	25.05	21.85	18.12	19.46	17.09	13.54	16.26
	Canton Ticino	26.22	23.04	19.11	21.24	18.00	16.12	17.82
	Canton Vaud	28.24	26.76	22.36	24.76	21.84	18.09	21.30
	Canton Geneva	22.26	21.86	18.53	21.49	18.74	15.69	18.92
	Swiss Average (other cantons)	26.98	24.32	19.76	22.26	18.98	15.24	18.19

Heating energy: Expenditures on heating fuels were converted to mega joules of final energy. However, the energy carrier used by the household's heating system is unknown in the HBS [12]. Therefore, an average Swiss energy mix needed to be assumed and the conversion implied several steps. Note that the conversions took into account the year in which a particular household was surveyed.

First, we assumed fuel oil as the household's energy carrier and computed liters of final energy based on the prices retrieved from [28] (see Table D.2). The lower calorific value was then calculated by means of figures from [29] (0.86 kg/l x 42.6 MJ/kg).

Table D.2: Fuel oil prices according to [28] in Swiss Francs per 100 liters (CHF/100l).

	Purchased amount						
	800 - 1500 l	1501 - 3000 l	3001 - 6000 l	6001 - 9000 l	9001 - 14000 l	14001 - 20000 l	> 20000 l
2009	80.86	72.83	68.90	67.45	66.39	65.18	64.60
2010	97.08	89.14	85.41	84.00	82.97	81.94	81.39
2011	109.32	101.71	98.03	96.74	95.71	94.66	94.13

Second, natural gas was assumed as the household's heating energy carrier. Final energy was computed based on gas prices presented in Table D.3. As the prices refer to the upper calorific value, lower calorific values were computed based on data from [29] (50.4 MJ/kg for upper calorific value and 45.5 MJ/kg for lower calorific value).

Table D.3: Prices for natural gas according to [28] in Swiss Francs per kilowatt-hour (CHF/kWh).

	Purchased amount			
	20,000 kWh	50,000 kWh	100,000 kWh	500,000 kWh
2009	0.0959	0.0908	0.0889	0.0863
2010	0.0911	0.0866	0.0848	0.0828
2011	0.0953	0.0911	0.0896	0.0875

Third, analogous computations were done assuming wood pellets as heating energy carrier. For this, the prices given in Table D.4 were used and the energy density of 18 MJ/kg was extracted from [30].

Table D.4: Prices for wood pellets according to [28] in Swiss Francs per 6000 kg (CHF/6000kg).

	Wood pellets 6000 kg
2009	2309.00
2010	2378.35
2011	2344.28

Finally, these three estimates of final energy in MJ of lower calorific values were weighted based on the nationwide final energy consumption statistics for households [30] given in Table D.5 and then averaged.

Table D.5: Statistics on final energy consumption of households in Switzerland [30]. Note that the shares were adjusted for the present purpose. The original data also contains information on coal, district heating and other renewable energy sources, but these shares were negligibly low.

	Shares in final energy consumption of households		
	Fuel oil	Natural gas	Wood energy
2009	0.63	0.26	0.11
2010	0.62	0.27	0.11
2011	0.60	0.28	0.12

Waste, wastewater collection and water supply: The expenditures on refuse collection, wastewater treatment and water supply were converted to number of waste bags, cubic meters of wastewater and cubic meters of water, respectively. The prices were retrieved from [15] and are presented in Tables D.6-8. Note that [15] only provides data for individual municipalities. Since the residential municipality of the HBS-households was unknown, we assumed the largest municipality within a canton (usually its capital) to be representative for the whole canton. Furthermore, the household types used by [15] to compute the prices are based on more living conditions than just number of household members. However, this was the only information provided by the HBS in this regard.

Table D.6: **Prices for waste bags according to** [15] **in CHF per 35-liters-bag. HH1: 1-person household; HH2: 2-persons household; HH3: 3-persons household; HH4: 4-persons household. Note that prices for HH2 could not be retrieved from** [15] **but were computed as the average of HH1 and HH3. Refer to subsection D.2.2.2 (*Details of the Modeling Framework*) for a description on how municipalities without waste prices (e.g. Geneva) were dealt with.**

Canton	HH1	HH2	HH3	HH4
Canton Zurich	3.65	2.86	2.39	2.22
Canton Bern	3.51	2.97	2.74	2.89
Canton Lucerne	2.49	2.22	2.19	2.19
Canton St. Gallen	2.60	2.29	2.22	2.18
Canton Aargau	2.92	2.52	2.40	2.34
Canton Ticino	2.82	2.30	2.03	1.93
Canton Vaud	3.64	3.17	3.05	3.24
Canton Geneva	0.00	0.00	0.00	0.00
Swiss Average (other cantons)	3.37	2.80	2.54	2.44

Table D.7: **Prices for wastewater treatment according to** [15] **in CHF per m^3. HH1: 1-person household; HH2: 2-persons household; HH3: 3-persons household; HH4: 4-persons household. Note that prices for HH2 could not be retrieved from** [15] **but were computed as the average of HH1 and HH3.**

Canton	HH1	HH2	HH3	HH4
Canton Zurich	3.36	3.05	2.73	3.57
Canton Bern	2.44	2.41	2.38	3.09
Canton Lucerne	1.73	1.73	1.73	1.73
Canton St. Gallen	2.14	2.12	2.11	2.46
Canton Aargau	1.56	1.40	1.25	1.35
Canton Ticino	0.78	0.78	0.78	0.78
Canton Vaud	1.30	1.30	1.30	1.30
Canton Geneva	2.49	2.59	2.69	3.07
Swiss Average (other cantons)	2.21	2.12	2.02	2.19

Table D.8: **Prices for water supply according to** [15] **in CHF per m^3. HH1: 1-person household; HH2: 2-persons household; HH3: 3-persons household; HH4: 4-persons household. Note that prices for HH2 could not be retrieved from** [15] **but were computed as the average of HH1 and HH3.**

Canton	HH1	HH2	HH3	HH4
Canton Zurich	1.93	1.80	1.68	2.04
Canton Bern	2.36	2.36	2.37	2.80
Canton Lucerne	1.96	1.87	1.78	2.31
Canton St. Gallen	3.11	3.05	3.00	3.92
Canton Aargau	1.35	1.35	1.35	1.49
Canton Ticino	1.64	1.59	1.54	2.36
Canton Vaud	2.39	2.35	2.32	3.26
Canton Geneva	2.18	2.19	2.19	2.57
Swiss Average (other cantons)	2.04	1.88	1.72	2.01

D.2.4 Evaluation of the Imputed Housing Data

In view of the many assumptions behind the computations in sections D.2.2 and D.2.3, plausibility checks with nationwide statistics were conducted as explained below. The modeled and converted final energy demand for heating totals at 21,222 MJ per person per year. This corresponds well with the national statistics [30], which indicates an average final energy consumption for heating of 22,890 MJ per person per year for the years 2009 to 2011. Similarly, the average final energy consumption of electricity based on the HBS was computed to be 2,101 kWh per person per year, while the national statistics [30] reveals an average value of 2,308 kWh per person per year in the years 2009 to 2011.

According to [31], the current average direct daily water consumption amounts to 142 liters per person or 52 m^3 per person and year. Our estimated average of 68 m^3 per person and year is slightly higher. It still might be considered a reasonable estimate since the statistics itself is an approximation rather than a measured value and it also does not specifically refer to the years 2009 to 2011.

Finally, waste statistics [32–34] amount to 344 kg per person per year in the years 2009 to 2011. The computed average of 187 kg per person and year based on the models in D.2.2 and D.2.3 is definitely too low, but still in the same order of magnitude. This discrepancy can partly be attributed to the fact that the waste statistics comprises all municipal solid waste and this also includes commercial waste apart from household waste. However, this probably does not explain the entire difference and the possibility exists that the modeling approach underestimates real waste production of households. Nonetheless, a rough sensitivity analysis with the final LCA results showed negligible effects even in the case of doubling the waste production. Note that the estimated amount of waste bags was converted to kilograms according to [35] (4.44 kg/bag).

D.2.5 Estimating Public Transport and Bicycle Demand

As mentioned in Chapter 5, the HBS-data does not allow for estimating kilometers driven by public transport and bicycles. In order to be able to assess environmental impacts on account of public transport and bicycle use nevertheless, person-kilometers of these traffic modes were estimated based on data from the Swiss Mobility and Transport Microcensus [36]. The data retrieved from this survey is compiled in Table D.9.

Based on the data provided in Table D.9, four different estimates for public transport (train, coach, urban vehicles) and bicycle demand were performed by taking into account the household's circumstances with regard to size, monthly income and age and gender of household members. These four estimates were finally averaged.

Table D.9: Daily distances per person in kilometers according to [36].

		Daily distance train (km)	Daily distance coach (km)	Daily distance tram and bus (km)	Daily distance bicycle (km)
		Average	Average	Average	Average
Household size	1 person	8.0	0.1	1.5	0.6
	2 persons	7.0	0.1	1.1	0.7
	3 persons	7.1	0.1	1.5	0.7
	4 persons	6.1	0.2	1.5	0.9
	5 persons and more	7.9	0.2	1.8	1.0
Monthly income	<= 2000 CHF	4.6	0.1	1.4	0.3
	2001 - 6000 CHF	5.3	0.2	1.3	0.5
	6001 - 10000 CHF	6.8	0.1	1.4	0.9
	10001 - 14000 CHF	9.6	0.1	1.5	1.1
	> 14000 CHF	10.9	0.1	1.5	1.1
Age	age: 6-17	4.6	0.2	2.2	1.0
	age: 18-24	15.6	0.3	3.2	0.8
	age: 25-44	8.0	0.1	1.2	0.8
	age: 45-64	6.1	0.1	0.9	0.9
	age: >= 65	4.6	0.1	1.0	0.4
Gender	Males	7.4	0.1	1.3	1.0
	Females	6.7	0.1	1.5	0.6

D.3 Pattern Recognition and Clustering of Households

D.3.1 Preparation for Pattern Recognition

This section shall provide more details on the filtering approach described in subsection 5.2.3.1 (*Preparation for Pattern Recognition*) in Chapter 5. The overall goal of this preparatory step was to find attributes of the HBS-data that make similarly behaving households identifiable independent of the month in which they were surveyed. While the general procedure was presented in Chapter 5, Figure D.9 illustrates it visually. Furthermore, the present section will focus on some additional details. For instance, the fourth diamond shape ("Is it representative for a month?") and its associated processing step ("Consider using an aggregated level") did not lead to the simultaneous use of an aggregated level together with its own sub-levels. This means for instance, if "fruits" (aggregated level) and simultaneously some of its sub-categories (e.g. "apples" or "pears") had fulfilled the requirements for inclusion, then either the aggregated level or the sub-levels were considered for the next steps, but not both (please note that this was just a clarifying example and does not correspond to an effective case). Furthermore, there seems to be some subjectivity attached to the fifth diamond shape ("Is this attribute very specific and consumed on an irregular basis?") in Figure D.9. This is true to some extent, but it also needs to be mentioned, that these decisions were informed by statistics for the respective attribute including histograms, boxplots, different percentiles, mean and coefficient of variation. It should also be noticed that only 10 attributes were excluded because of this decision point.

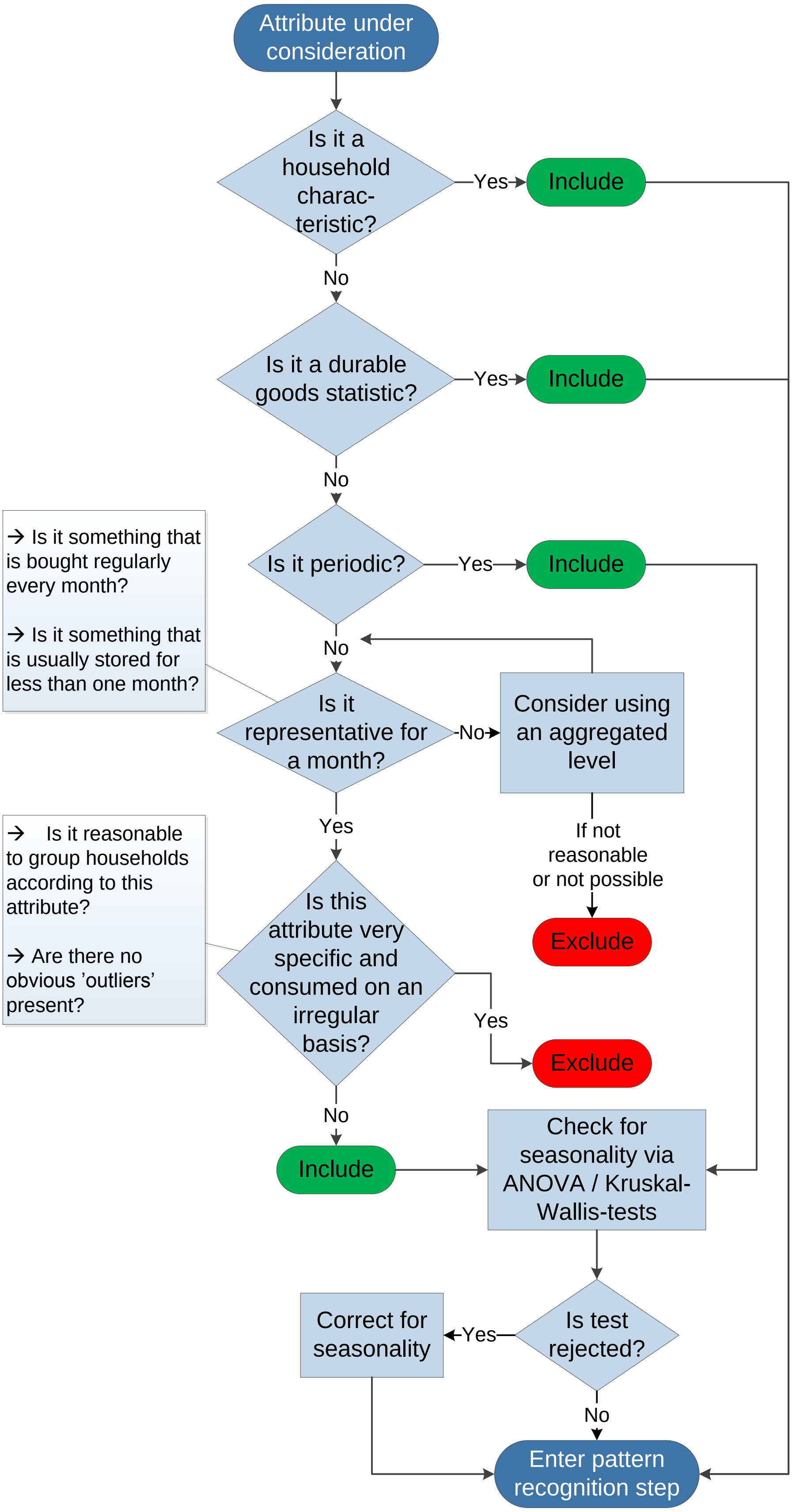


Figure D.9: **Flow scheme outlining the general procedure for preparing data for the sub-sequent pattern recognition.**

Finally, we also would like to mention that housing-related and public transport-related data that were imputed in section D.2 were completely excluded for the clustering, to avoid introducing uncertainties from the modeling procedure.

As mentioned in Chapter 5, the final dataset for pattern recognition comprises 157 attributes in total, thereof 85 consumption and income categories covering about 80% of all expenditures and 95% of total income. The 20% of excluded expenditures pertain mainly to infrequent expenses on durable goods (e.g. purchase of a new car). However, since the finally bought durable goods entered the clustering via the durable goods statistics (e.g. number of new cars), this piece of information is not lost but just considered in a different attribute.

The full list of attributes which entered the following computations can be found in the supplemental EXCEL-file[2] (column "Pattern recognition filter"). Furthermore, attributes that were corrected for seasonality were marked accordingly in the EXCEL-file (column "Correction for seasonality"). The seasonality correction is also illustrated in Figure D.10 for the example of "fruits". The figure shows the original data on the left and the corrected, "de-seasonalized" data on the right-hand side.

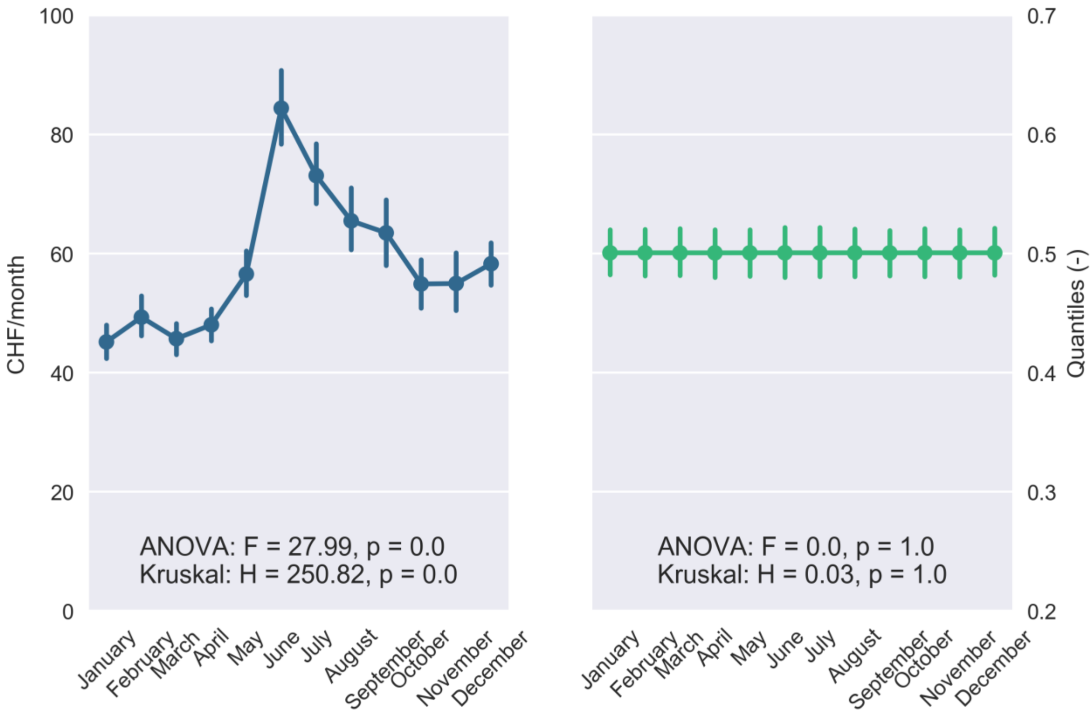


Figure D.10: **Example of correcting for seasonality for "fruits" including test statistics of ANOVA and Kruskal-Wallis-test. Left: original data; right: corrected data; Error bars are the 95%-confidence interval of the monthly mean.**

The correction procedure as such can be described as follows: The dataset is partitioned into monthly subsets (e.g. "expenditures on fruits in January", "expenditures on fruits in February", etc.). In a next step, the original values are then replaced by its monthly quantile ranks. This

[2] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

means, if e.g. a household buys fruits for 7 Swiss Francs in January, this corresponds to a quantile rank of 0.42 in this month; or in other words: 42% of the expenditures for fruits of all households in January were below 7 Swiss Francs. Finally, the original value of 7 Swiss Francs is replaced by 0.42. Please note that this is a made-up example.

As mentioned in Chapter 5, the need for seasonality correction was judged based on ANOVA [37] and Kruskal-Wallis-test [38]. With these tests, we tested if the survey month shows a statistically significant influence. While the null hypothesis of ANOVA assumes the means of different data groups (here: monthly groups) are the same, Kruskal-Wallis focuses on the groups' medians. If the tests suggest rejection of the null hypothesis, this means that monthly samples do not originate from the same distribution and thus that seasonality is present (so the survey month shows a statistically significant impact). The α-levels for both tests were set at 0.05 and we decided to correct for seasonality if at least one of the tests indicated rejection of the null hypothesis (however, in almost all cases, there was no contradiction between the two tests).

D.3.2 Self-Organizing Map (SOM)

Even though there is plenty of literature available describing how self-organizing maps (SOM) work (e.g. [39–43]), we would like to support the reader with a brief overview (note that the subsequent description is closely based on the very helpful explanations of Vesanto [43] and that the description is simplified to a certain degree):

A SOM consists of so-called neurons which are arranged on a regular low-dimensional grid. A so-called prototype vector is attached to each neuron which features the same dimension as the input vectors (here, the input vector is a HBS-household with all of its attributes after the preparation step in section D.3.1). Moreover, each neuron is linked to neurons which are adjacent on the map by a neighborhood function. This neighborhood relation then dictates the topology of the map. There are two different possibilities for training the SOM: sequential and batch mode. Even though batch mode was used for the present study, we start explaining the sequential algorithm because it is easier to understand from our point of view.

In the sequential algorithm, the SOM is trained iteratively: in each step, an input vector is chosen randomly and the distance between it and all the prototype vectors is computed using Euclidean distances. The closest neuron is called the "Best-Matching Unit" (BMU). After identifying the BMU, its prototype vector as well as the prototype vectors of its neighboring neurons are updated, such that they move closer to the input vector. Depending on the neighborhood function, the BMU might move most, while the neighbors are adjusted the less the further away they are located from the BMU. Such a behavior can for instance be achieved by a Gaussian neighborhood function whose kernel is at the BMU. A possible update rule can look like equation (D.1) [44]:

$$\underline{m_i}(t+1) = \underline{m_i}(t) + \alpha(t) \cdot h_{ci}(t) \cdot [\underline{x}(t) - \underline{m_i}(t)] \tag{D.1}$$

$\underline{m_i}$:	prototype vector of neuron i
$\underline{x}$:	vector from input dataset
α:	learning rate
h_{ci}:	neighborhood kernel around BMU c
t:	epoch

The first so-called epoch is accomplished if all input vectors were exposed to the SOM. The next epoch works in exactly the same way, except for the fact that the neighborhood radius (this determines which neurons on the map are considered neighbors) and the learning rate decrease. This process is then repeated over many epochs until the SOM "converges". However, it is to be noted that "convergence" is not really provable [45], but according to Tan and George [45], good maps can be produced if the SOM is trained with "enough" epochs. It is generally recommended to subdivide the training into a rough training and a fine-tuning phase. The first starts with relatively large initial learning rates and neighborhood radius. In the second, both, learning rate and neighborhood radius, are small from the beginning. In simplified terms, one could say that "cooperative learning" usually prevails in the first phase, while "competitive learning" is predominant in the second phase because the neighborhood radius is small and in the end usually comprises only one neuron.

The batch training algorithm works basically along the same lines as explained above. Instead of using a single input vector at a time, the whole input dataset is exposed to the SOM before any updates are conducted. In each epoch, the dataset is partitioned into Voronoi-regions to determine which input vector is closest to which neuron. The updated prototype vector is then a weighted average of the data samples. Thereby, the weights are based on the neighborhood function. In this batch-mode, no learning rate is needed anymore [46].

According to literature [39, 40, 42–45] and above description, the following parameters need to be tuned to obtain the "best" SOM: number of neurons, arrangement of neurons, initial and final neighborhood radius of the rough training phase, initial and final neighborhood radius of the fine-tuning phase, neighborhood function, and number of epochs for both: the rough and fine-tuning phases. However, a tuning procedure also implicitly needs an evaluation measure to finally determine the "best" choice. Thereby, we followed Tan and George [45] and used the topographic error in a first instance and looked at the quantization error in a second instance. The quantization error thus only played a role to decide among maps with similar topographic errors. Unfortunately, the computational burden of training a SOM did prevent from an exhaustive search of the whole parameter space in the sense of computing all parameter value combinations. Therefore, we decided to take literature recommendations [39, 43–45] as a starting point and tried then to find the way towards the best map by applying the following iterative approach (in the end we computed more than 35 different maps):

1. Radii: We started tuning the radii of the rough and fine-tuning training phases since these are apparently the most important parameters [43]. Besides the recommendations in [39, 43–45], we also experimented with more extreme radii. All other parameters were set to the recommendations of [43].

2. Epochs: For tuning the number of epochs (for rough and fine-tuning phase), we set all other parameters according to [43] and then applied again different literature recommendations [39, 43–45] as well as own trials.

3. Number of neurons: While in principle, the procedure was exactly the same as for steps 1 and 2, it needs to be mentioned that radii and number of epochs were set dynamically since the recommendations in [43] are dependent on the number of neurons.

4. In a fourth step, the settings of different (less important) parameters were tested: neighborhood function (bubble vs. Gaussian), initialization (PCA vs. random), as well as different map ratios. Please note that all sources generally agree that the ratio of the side lengths shall correspond to the ratio between the two greatest eigenvalues of the covariance matrix of the training data.

5. Based on all findings above, we constructed a "best" combination of parameters. Taking this "best" map as a basis, we experimented again with the radii.

6. And finally we also tested different numbers of epochs on top of this "best" combination. However, we realized that at this point, the U-Matrices obviously hardly change indicating that convergence is probably reached already in step 5.

Table D.10 shows the finally selected SOM-parameters. To give an impression of the achieved improvements by the above tuning procedure, the largest topographic error of all maps was 0.400 and the largest quantization error 10.050, while the final map reached 0.063 and 9.266 for these performance metrics. The following figures deliver full insight into the so-called component maps of the final SOM. These component maps provide a wealth of information since they visually depict correlations among attributes. For instance, in Figure D.11 we see that higher numbers of persons per households can be found on the left-hand side of the map. At the same time, the number of pensioners per household is larger on the right-hand side of the map, which in turn also correlates strongly with the income from "pensions and social benefits". Overlaying these maps with the clustering maps presented in section 5.3.3 and in Chapter 5 thus also gives insights into the archetypes' characteristics and consumption behavior. Further maps (e.g. a hits-map) are presented in section D.3.3. Note that pixels in these maps correspond to the map positions of neurons.

Table D.10: Parameters of the final SOM.

Parameter	Value
Normalization	Standardized data
Initialization	PCA
Neighborhood	Gaussian
Map-Ratio	21:47
No. of neurons	987
No. of epochs (rough training phase)	7896
No. of epochs (fine-tuning phase)	31584
Initial radius (rough training phase)	35
Final radius (rough training phase)	9
Initial radius (fine-tuning phase)	9
Final radius (fine-tuning phase)	1
Topographic error	0.063
Quantization error	9.266

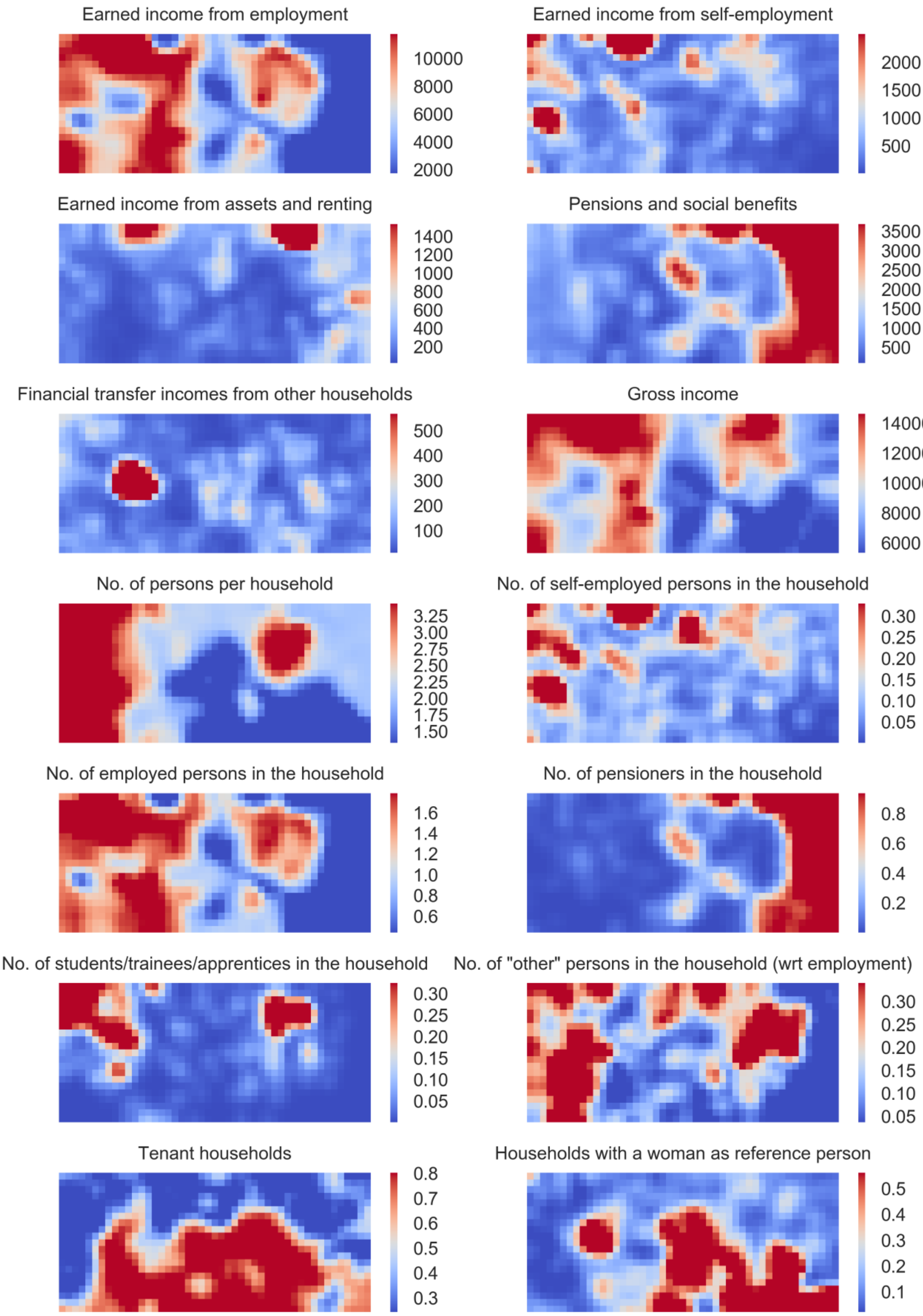


Figure D.11: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

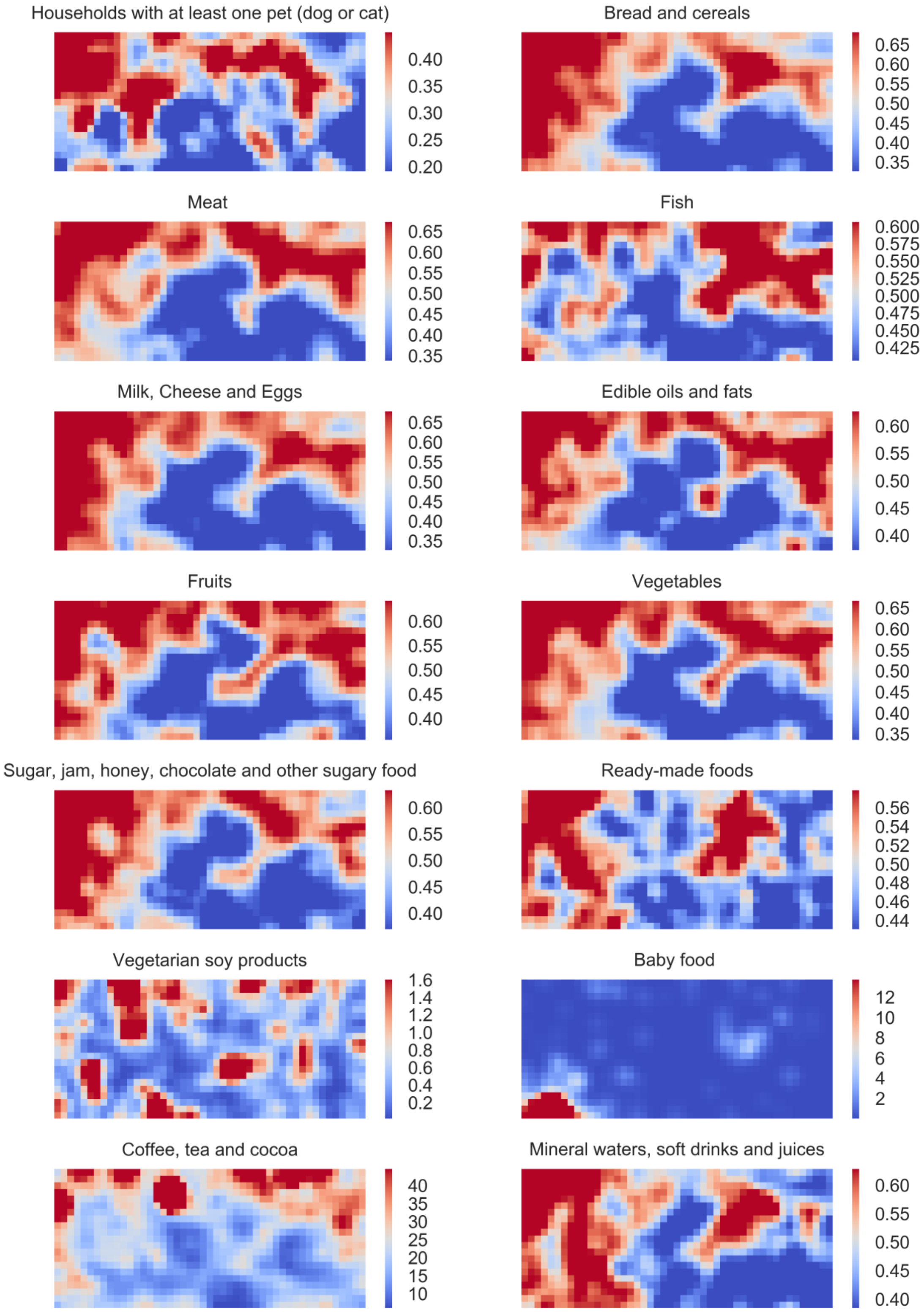


Figure D.12: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

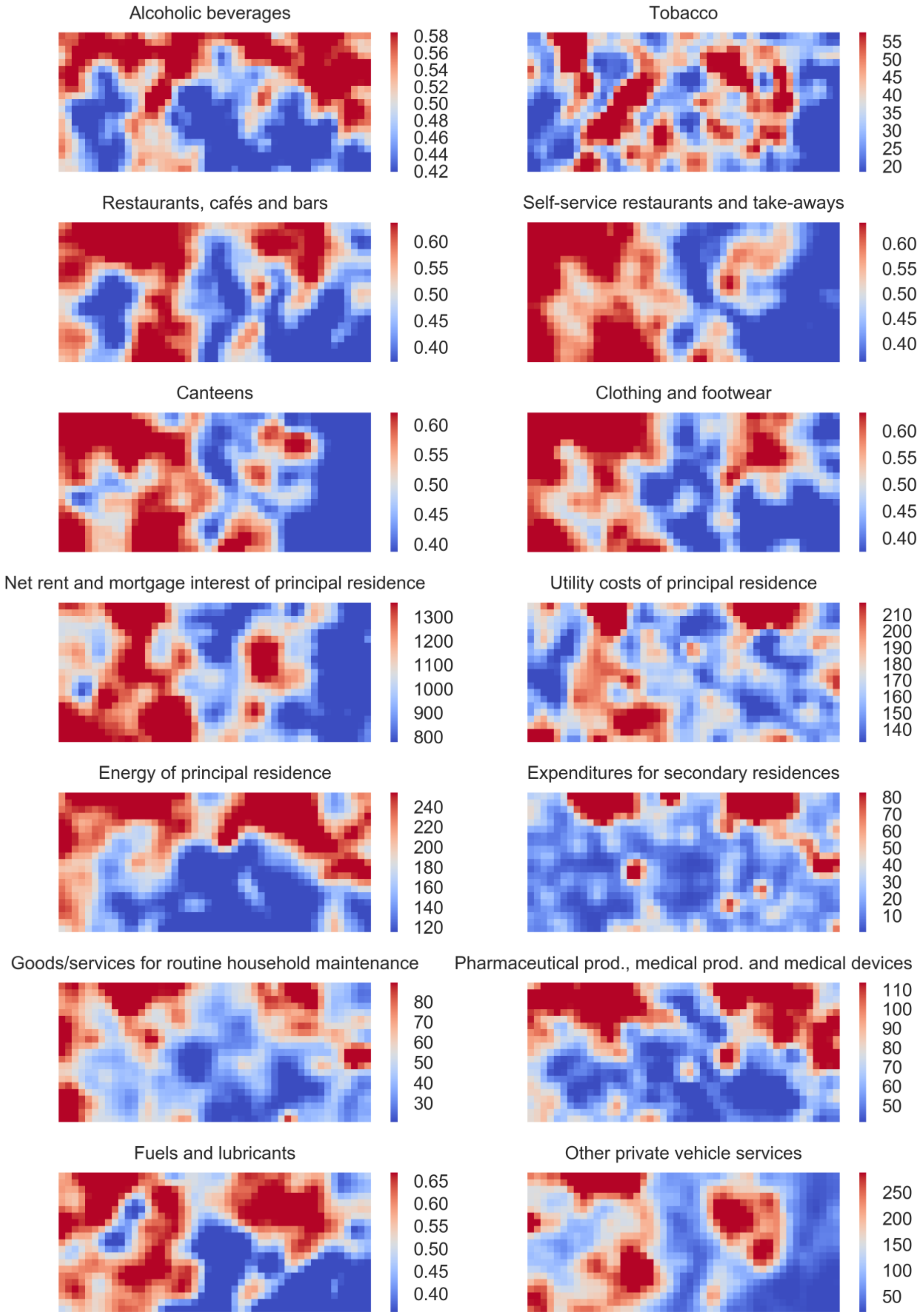


Figure D.13: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

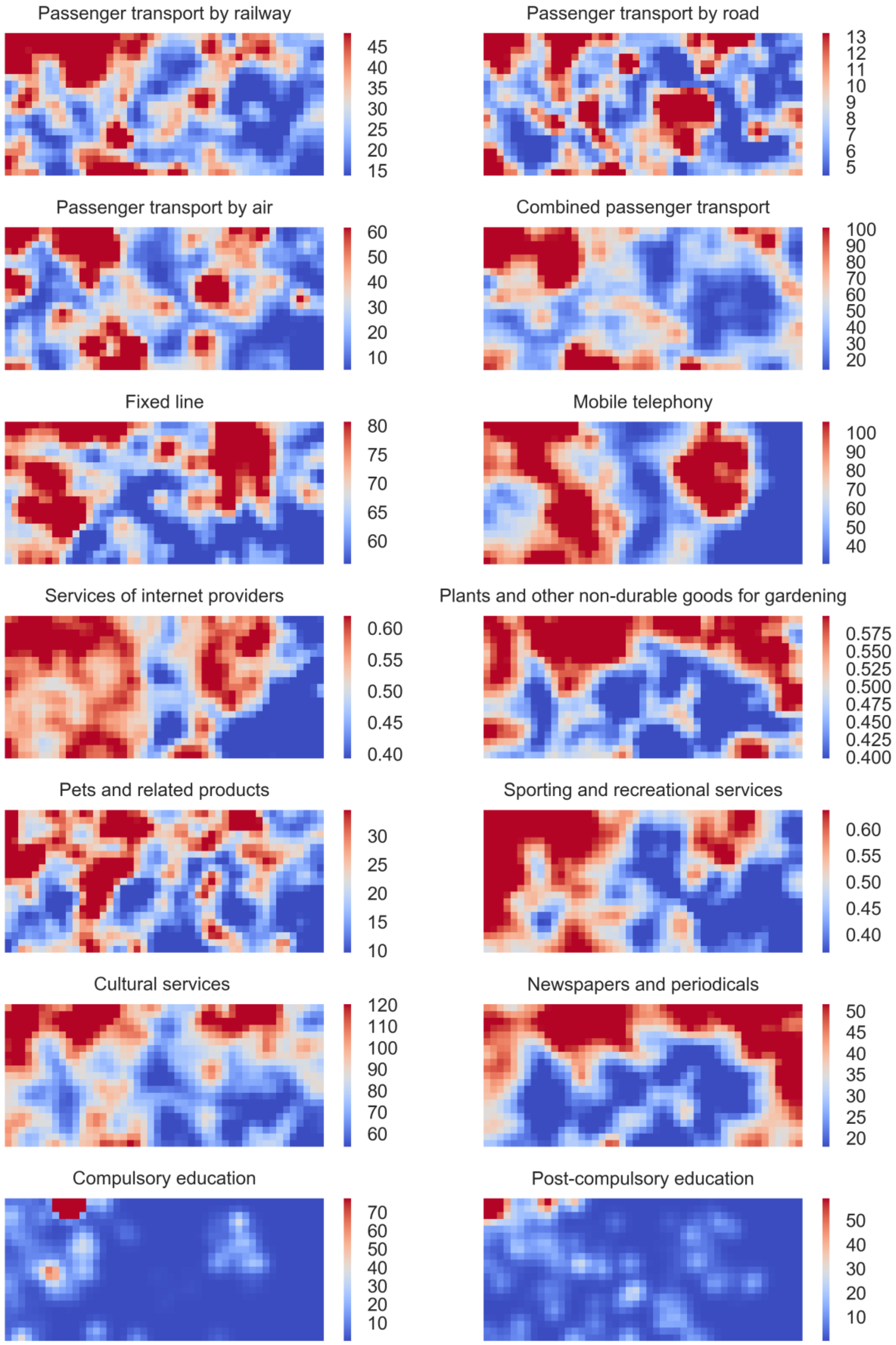


Figure D.14: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

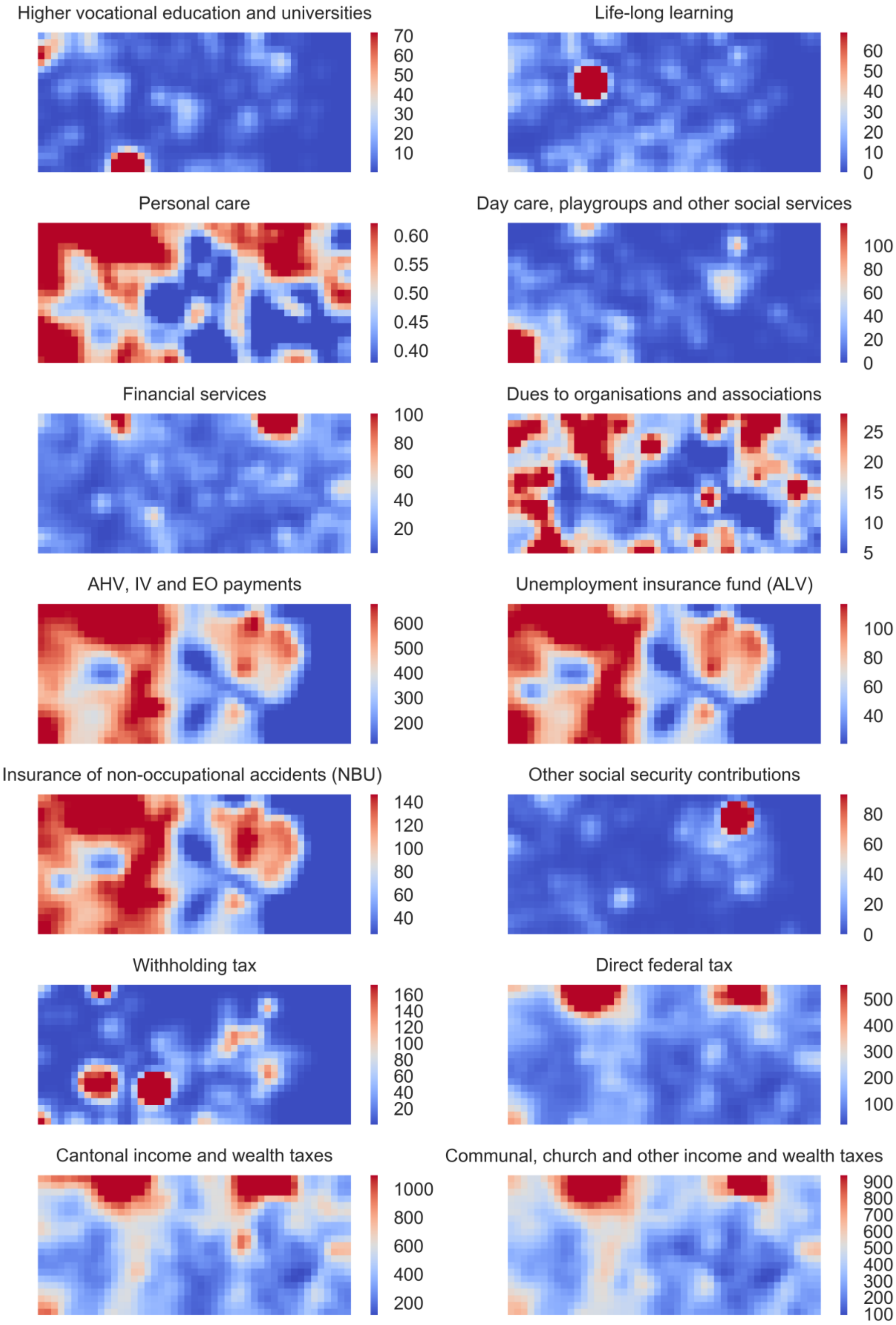


Figure D.15: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

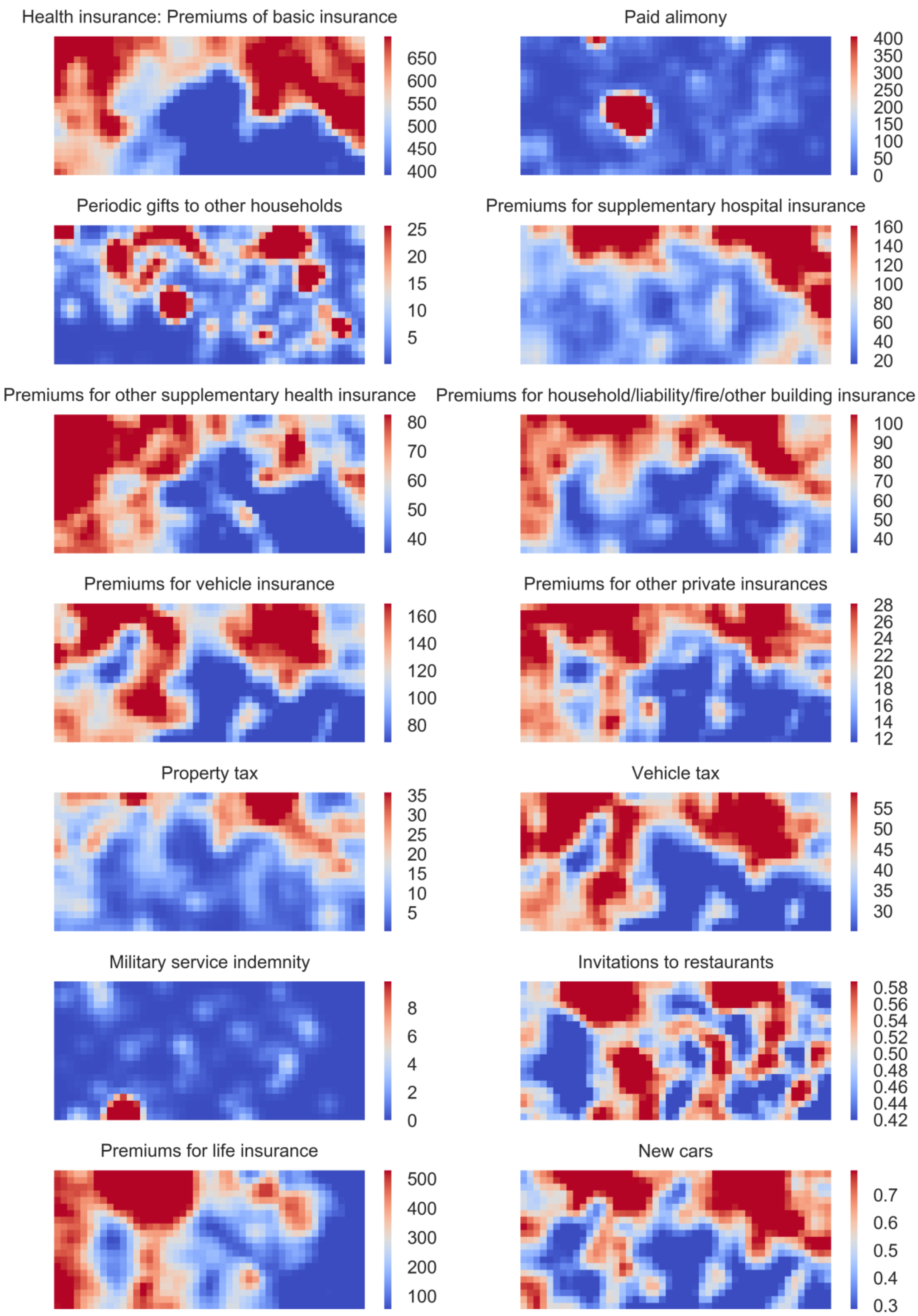


Figure D.16: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

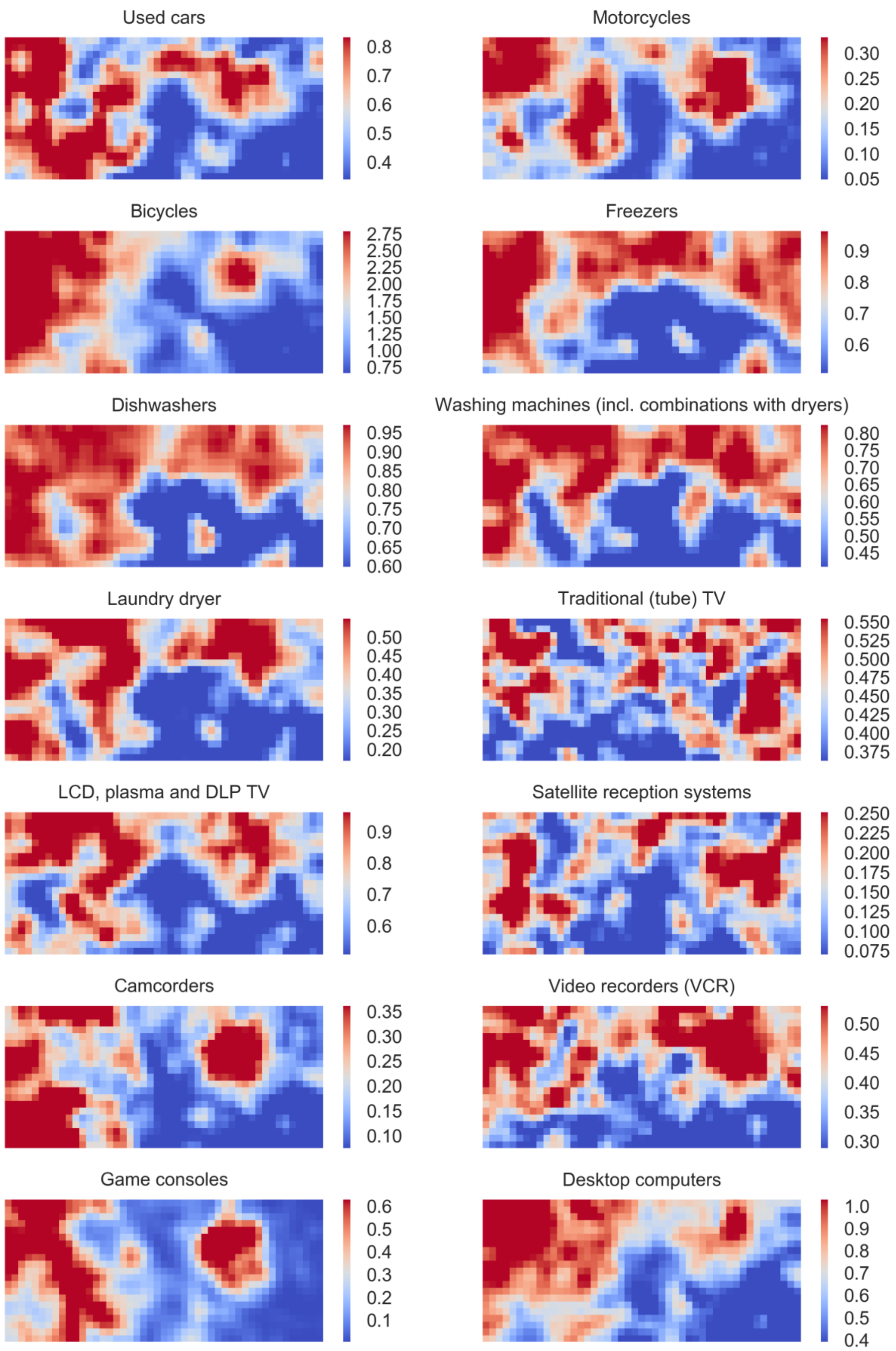


Figure D.17: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

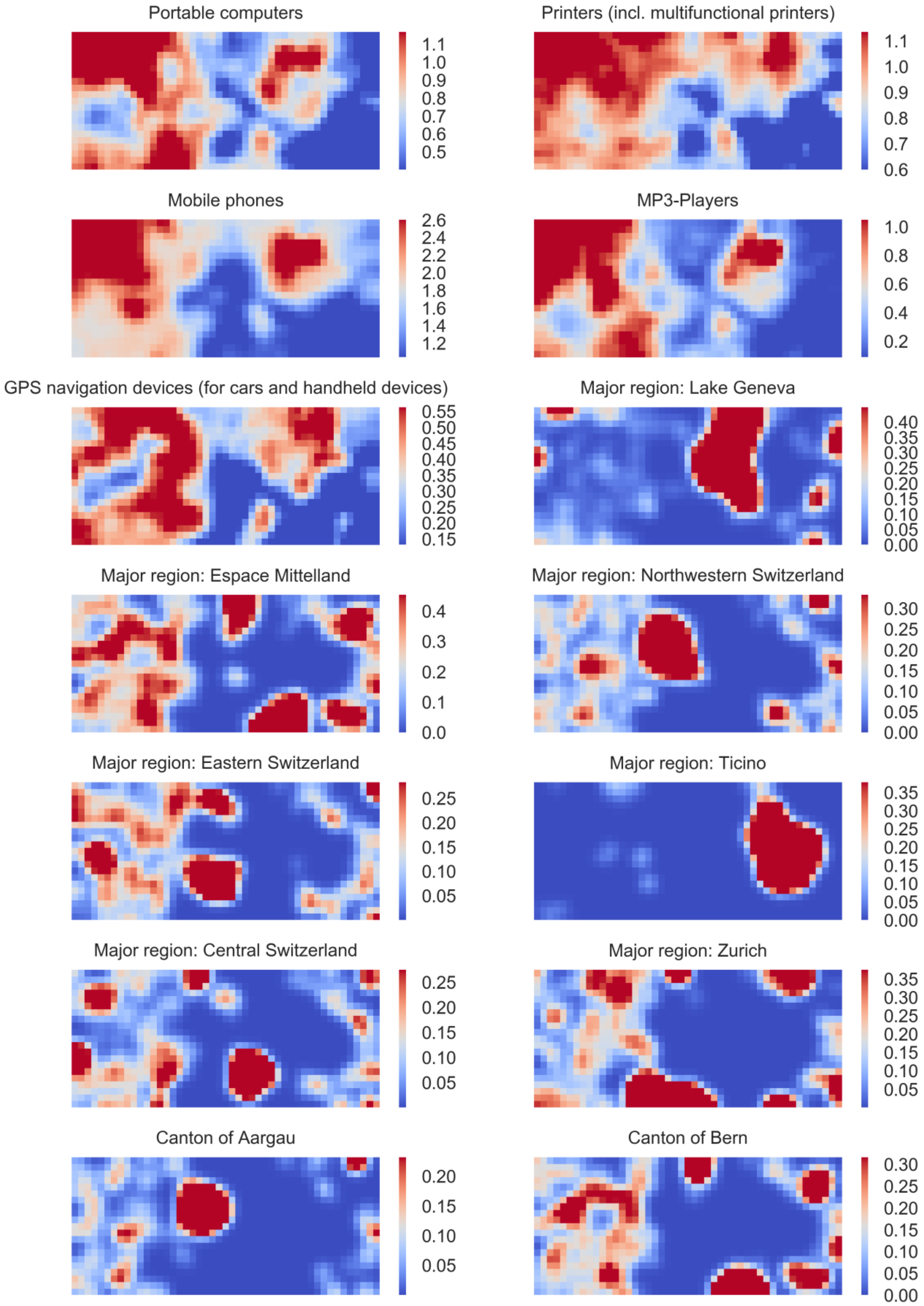


Figure D.18: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

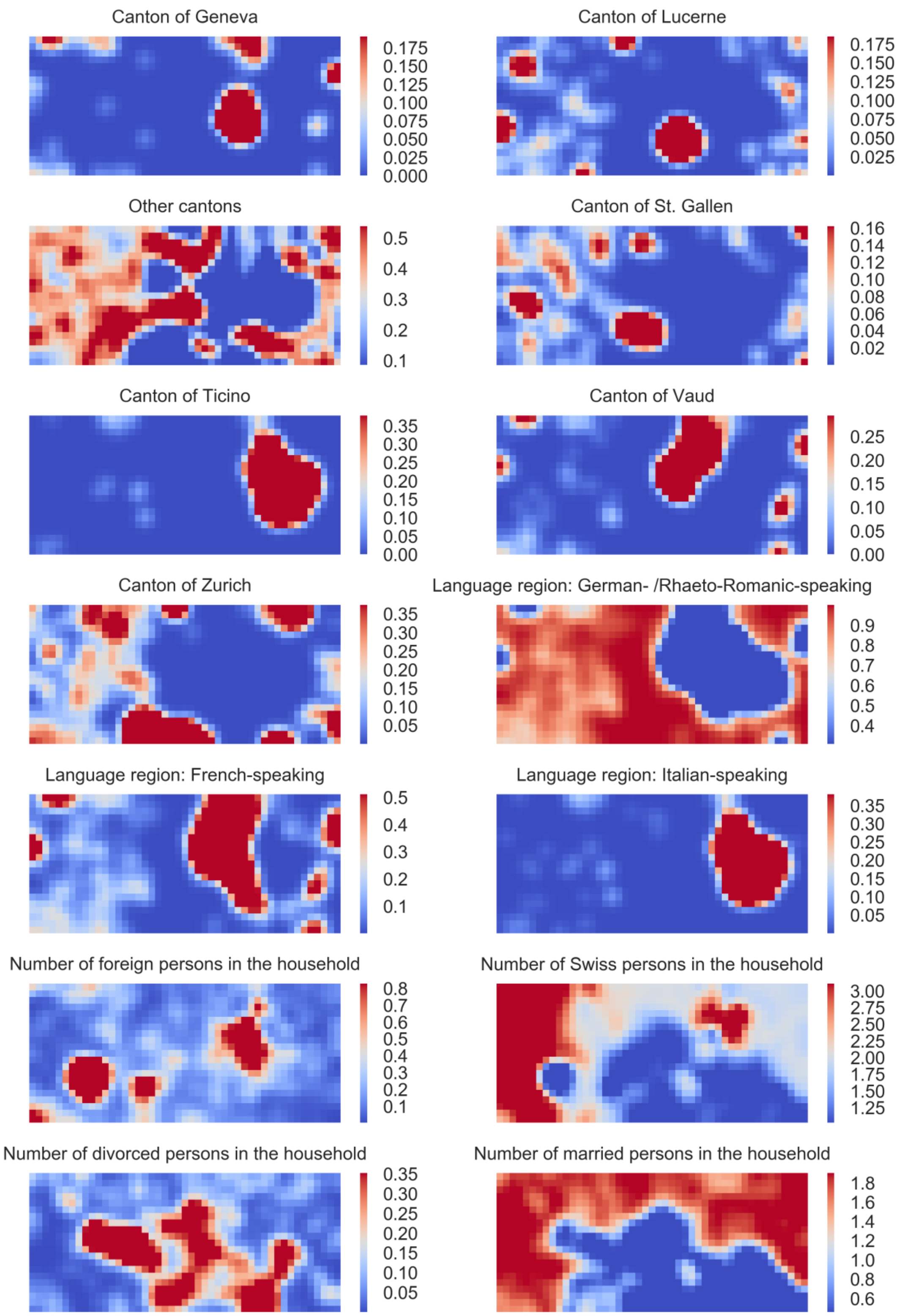


Figure D.19: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

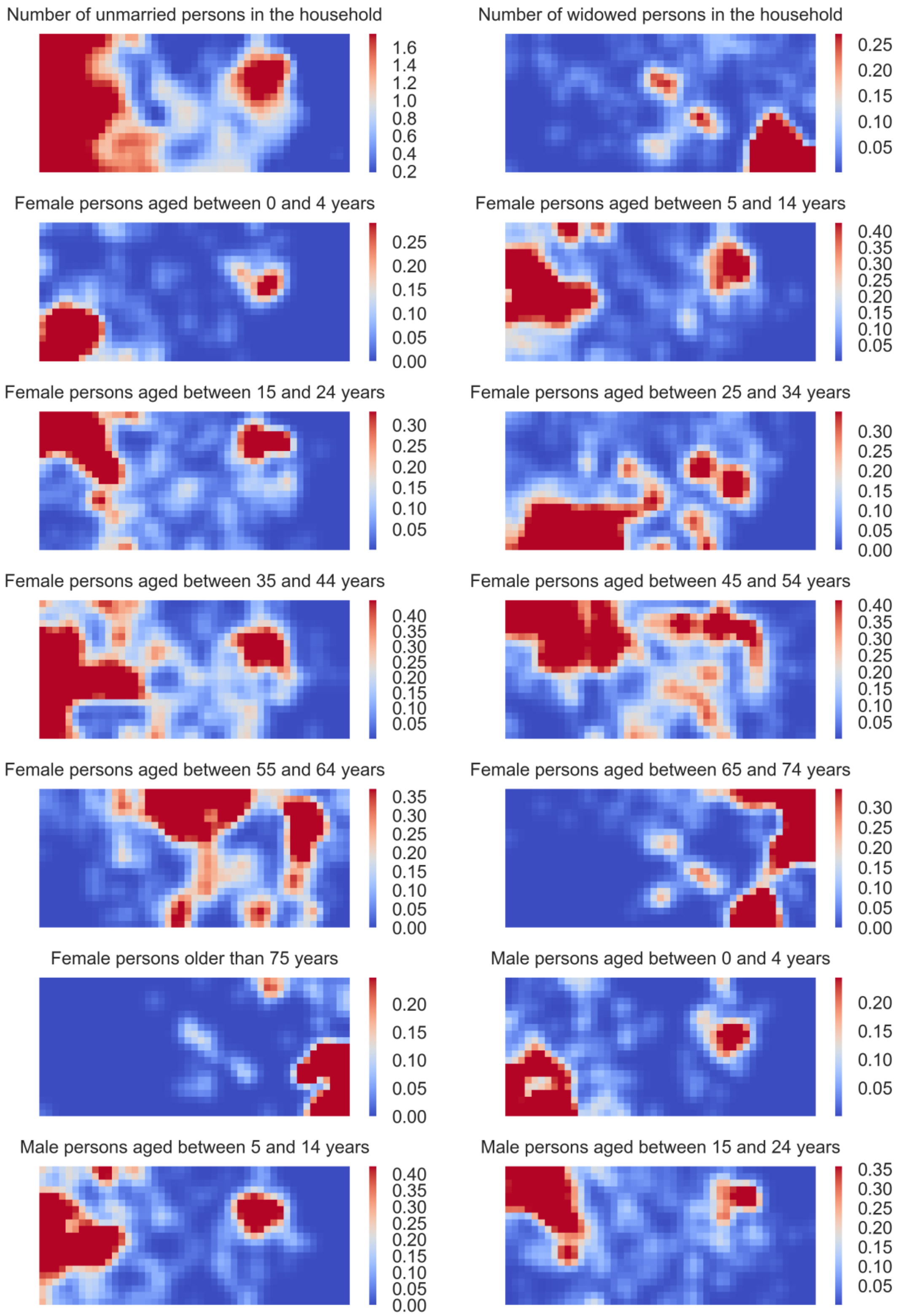


Figure D.20: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

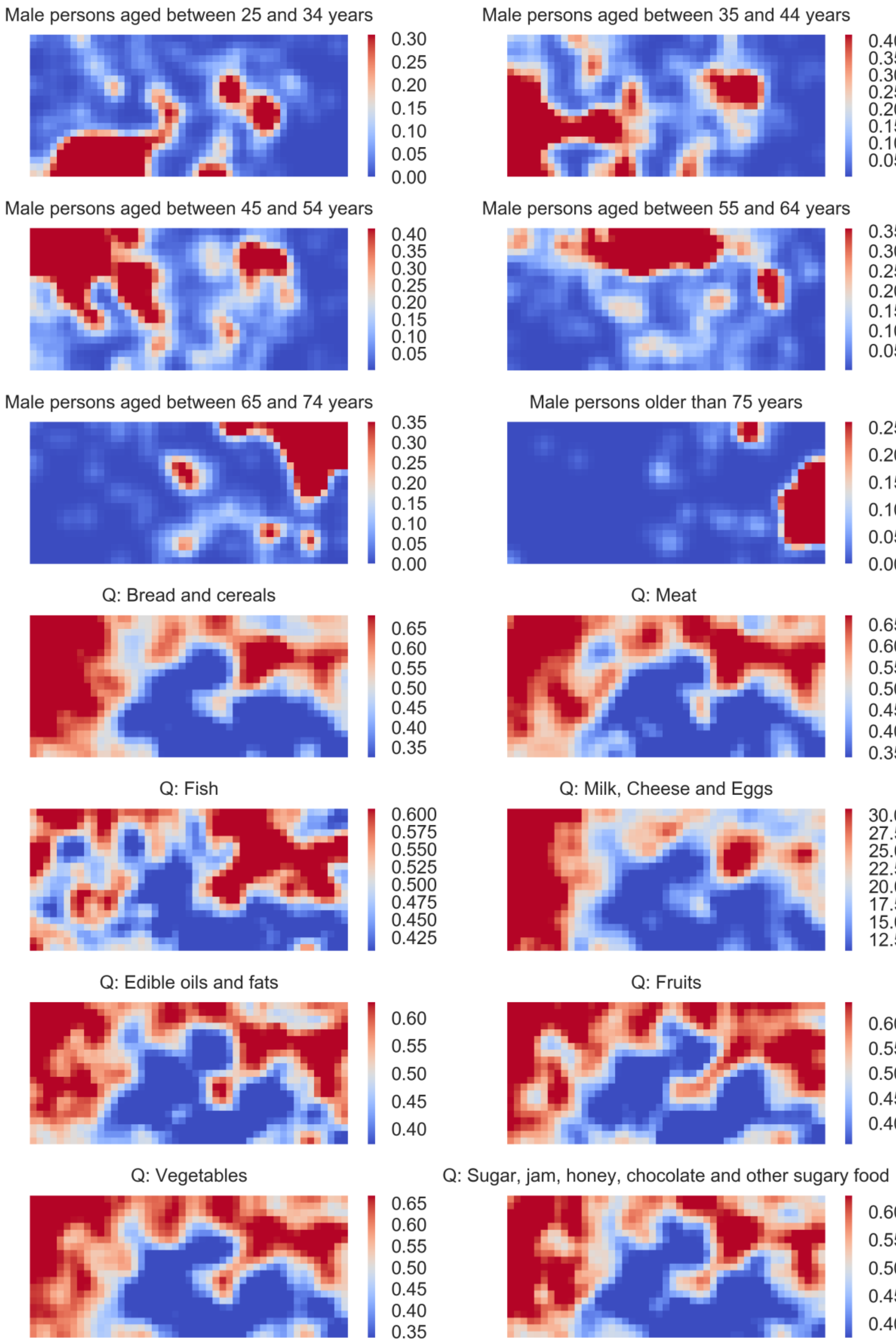


Figure D.21: Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).

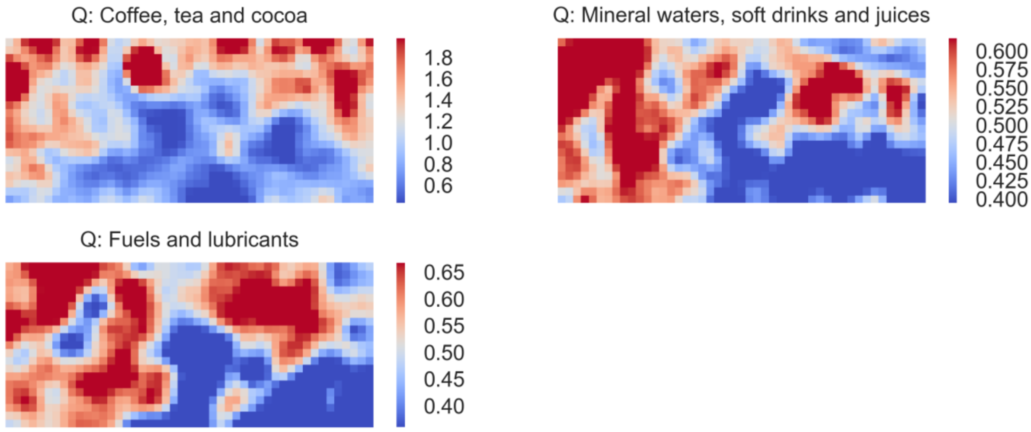


Figure D.22: **Component maps of the final SOM. The units are as follows: income and consumption expenditures in CHF/month; quantities (marked with a "Q") are in liters or kg/month; durable goods statistics, count-statistics and dummy variables do not have a unit (-); Note that variables corrected for seasonality are in monthly quantiles (-).**

D.3.3 Clustering

The present section shall provide more technical details on the two applied clustering approaches: K-Means [47–49] and agglomerative clustering [50, 51] as well as on the procedure to determine the final clustering.

K-Means separates data in a predefined number of *k* groups by minimizing the "within-sum-of-squares", resulting in clusters of similar size and variance. Agglomerative clustering uses a bottom-up approach and begins with each prototype vector as its own cluster. The clusters are then successively merged starting with the two closest clusters. Various agglomerative clustering techniques differ in the use of affinity metrics ("which distance") and in the linkage criteria ("how this distance is computed").

D.3.3.1 Technical Details of the Clustering Algorithms

For the present study, the implementation of the clustering approaches in [4] was used. The settings for the computations were as follows:

- K-Means:
    - K-Means++ [49] was applied for seeding. This ensures that the initialized centroids are generally distant from each other leading to faster convergence and more robust results than random initialization [49].
    - Even though K-Means++ already helps to prevent convergence to local minima, we still applied 100 independent runs.
    - An absolute number of maximum iterations was not set. The iterations stopped due to a tolerance criterion with regard to inertia.
    - The only tuning parameter: number of clusters.

- Agglomerative clustering:
    - Affinity metrics investigated: Euclidean distance, L1-norm (Manhattan distance) and cosine distance.
    - Linkage criteria investigated: Ward [52] and average. Similar to the objective function of K-Means (minimizing the "within-sum-of-squares"/inertia), Ward-linkage minimizes variances, but applied in the scope of an agglomerative hierarchical approach. In contrast, average-linkage minimizes the average of the distances between all member vectors of pairs of clusters. The applied implementation [4] allows Ward-linkage only in combination with Euclidean distance, while for average-linkage, also the other affinity metrics were analyzed.
    - Connectivity: A very interesting feature of the agglomerative clustering implementation [4] is the possibility to take connectivity constraints into account. This means that the hierarchical algorithm is allowed to merge only clusters which are adjacent to each other on the SOM. All combinations above are therefore computed in both modes: once with connectivity constraints and once without.

D.3.3.2 Evaluation of the Best Clustering

As mentioned in Chapter 5, a two-step evaluation procedure was applied to determine the "best" clustering: A pre-selection which only focused on the parameters within each approach followed by a final comparison between the two clustering methods.

Pre-Selection

In this pre-evaluation, Silhouette-Coefficients (S) [53] constituted the main evaluation metric. However, according to Liu et al. [54], S might face problems in the presence of sub-clusters. Therefore, we also took the Calinski-Harabasz-Coefficient (CH) [55] into consideration which evaluates scatter between groups and within groups and which succeeded in the cases where S failed. However, CH might fail in the case of noise and skewed distributions, but S fortunately performed well in the respective experiments of [54].

The pre-selection applied to the two approaches:

- K-Means:
    - The pre-selection basically narrowed down the possible range of the number of clusters (k). Note that we refrained from just taking the k with the largest S-value since especially in the case of sub-clusters [54], it seems to be reasonable to determine a range of k with similar (large) S-values.
    - The k's for further investigations, were determined based on a range of high S-values, a range of high CH-values as well as on the "elbow" of the inertia-plot: 25 to 93 clusters.

- Agglomerative Clustering:
  - Based on the plots of S (overall average) and CH for different numbers of clusters, the best combination of parameters was determined: Euclidean distance with Ward-linkage [52] including connectivity constraints.
  - Just as for K-Means, also the range of numbers of clusters to be further investigated was selected based on a range of high S-values, a range of high CH-values as well as the largest gap in the dendrogram: 20 to 153 clusters.

Final Evaluation

In the second evaluation step, the pre-selected clustering algorithms were judged according to their ability to reproduce the structure of visible groups of neurons in the U-Matrix. Furthermore, two other criteria entered the decision process:

- ANOVA-tests: ANOVA was performed for each attribute separately. Rejection of the null hypothesis (see also section D.3.1) shows that there is at least one cluster which statistically significantly differs from the others. This test shall thus give an impression of how reasonable the clustering might be.
- Number of HBS-households per cluster to get an indication of representativeness of the clustering. Besides statistics for each cluster, also overall statistics such as mean, minimum, maximum and median were computed. Furthermore, it was also analyzed if all survey months are represented in each cluster and if not, which are missing.

Note that these two aspects were meant to broaden the information basis rather than being the main decision criteria. Moreover, at this advanced stage of evaluating the clustering, the null hypothesis of ANOVA was rejected for all attributes in all of the investigated clusterings.

For the agglomerative clustering, comparing the U-Matrices of different numbers of clusters was interesting since the use of connectivity constraints enabled the tracking of merging clusters. Therefore, we could directly judge how reasonable the merge of two clusters (breaking the borders) is.

The two best versions of each clustering algorithm are shown in Figure D.23. After careful consideration of all criteria, we decided for the agglomerative clustering approach with Ward-linkage and connectivity constraints. Apart from all criteria being either in favor of this approach or show reasonable results, the preservation of connectivity was seen as an advantage over K-Means which distributed some clusters across the whole U-Matrix. In addition, the use of these connectivity constraints is also in line with the suggestions of Vesanto and Sulkava [56] who propose a similar approach to perform clustering on top of a SOM.

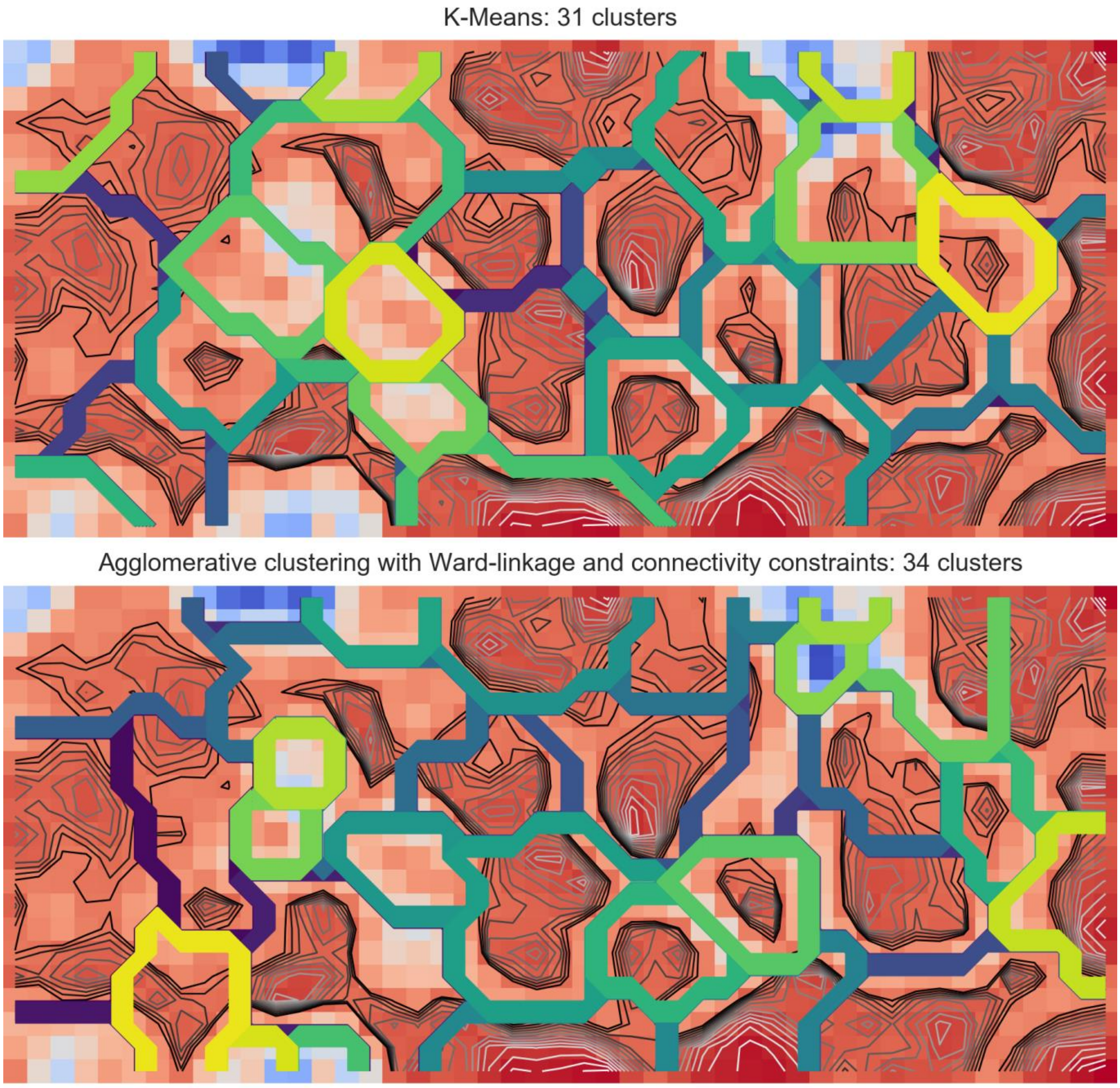


Figure D.23: **Comparison of the best K-Means (31 clusters, top) and the best Ward-clustering (34 clusters, bottom). The borders of the found clusters were projected onto the U-Matrix.**

D.3.3.3 Final Clustering and Deriving Archetypes

As explained in Chapter 5, some manual merging of clusters was applied in a post-processing step. This was primarily necessary because not all clusters reached the minimum count of households (130) that was regarded as being representative for performing statistics [14]. Additionally, this manual post-processing is also supported by the dendrogram in Figure D.24 which shows some "Fuzziness" between 34 and 24 clusters indicating that there are several clusters which are close to each other. This might also be a consequence of using connectivity constraints. Also Vesanto and Alhoniemi [46] point out that in case of sub-clusters, it might be reasonable to cut a dendrogram at different positions. Therefore, we decided to merge all clusters with less than 130 households with their closest neighboring clusters if these merges would also happen in the dendrogram between 34 and 24 clusters. This procedure resulted in 26 clusters that were considered representative (more than 130 households as members) and 2 clusters for which the minimum count of households was not reached. These two clusters will be part of all subsequent analyses, but they will be marked as “outlier A” (OA) and “outlier B” (OB). To name all other clusters, a random letter was assigned. Note that in all clusters all survey months are present which indicates another aspect of representativeness and also shows that the applied correction for seasonality worked.

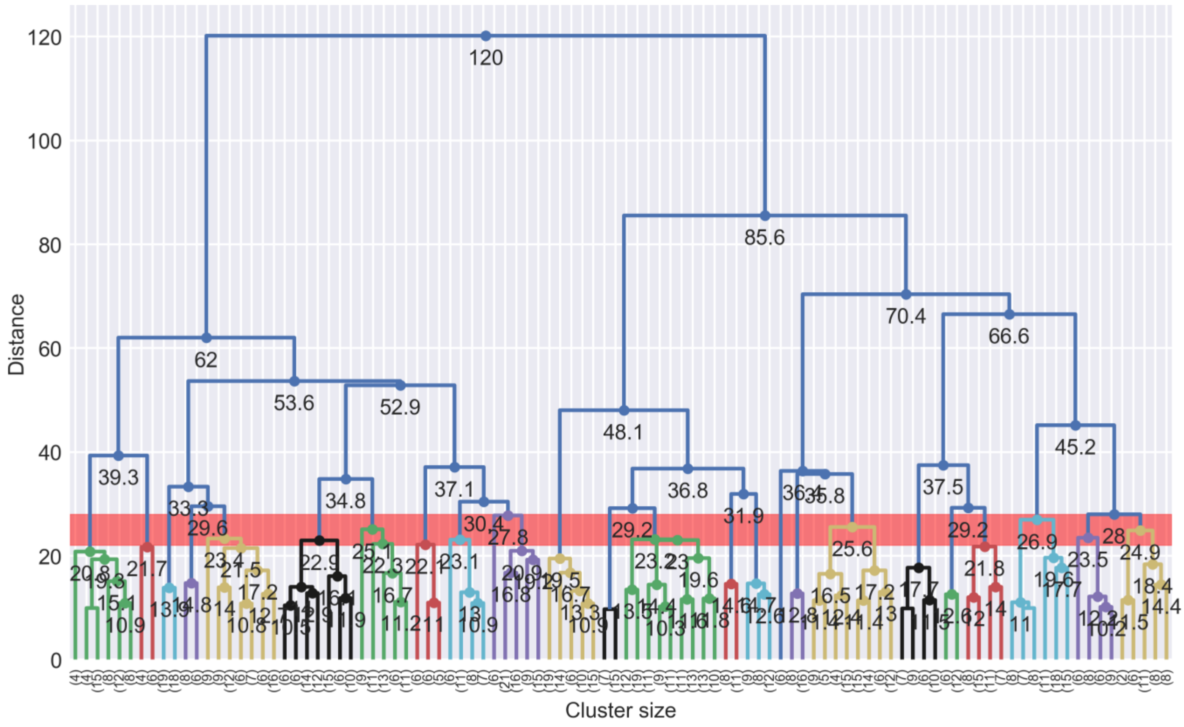


Figure D.24: **Simplified and truncated dendrogram for the agglomerative clustering approach with Ward-linkage and connectivity constraints. The red zone approximates the blur between 34 and 24 clusters.**

However, the question might arise why not directly working with 24 clusters. There are several reasons against this: First of all, the version with 34 clusters outperformed 24 clusters in the evaluation procedure above. Second, the two “outlier” clusters would not change and still persist. Third, Vesanto and Alhoniemi [46] suggest that for clustering the SOM, also the so-called hits-map shall be consulted. This map depicts the hits per neuron, which means that it shows how many times the neuron became a BMU (see Figure D.25). Or in other words: it indicates the number of HBS-households for which a certain neuron is closest. Vesanto and Alhoniemi [46] argue that interpolating neurons (neurons with few or even zero hits) might obscure cluster borders or the other way around: the localization of zero-hit-neurons can be utilized to identify cluster borders. The hits-map in Figure D.25 shows indeed some zero-hit-neurons along the borders of clusters which would be merged on the way from 34 to 24 clusters. We took these zero-hit-neurons as indication for not breaking these borders. To a certain degree, this is also in line with other literature recommendations suggesting that for clustering on top of a SOM, also density information should be used [57]. Taking the hits-map as a rough approximation of a P-Matrix [58] (similar to the U-Matrix, but based on density instead of distances), this would result in the same conclusion not to conduct all possible merges between 34 and 24 clusters.

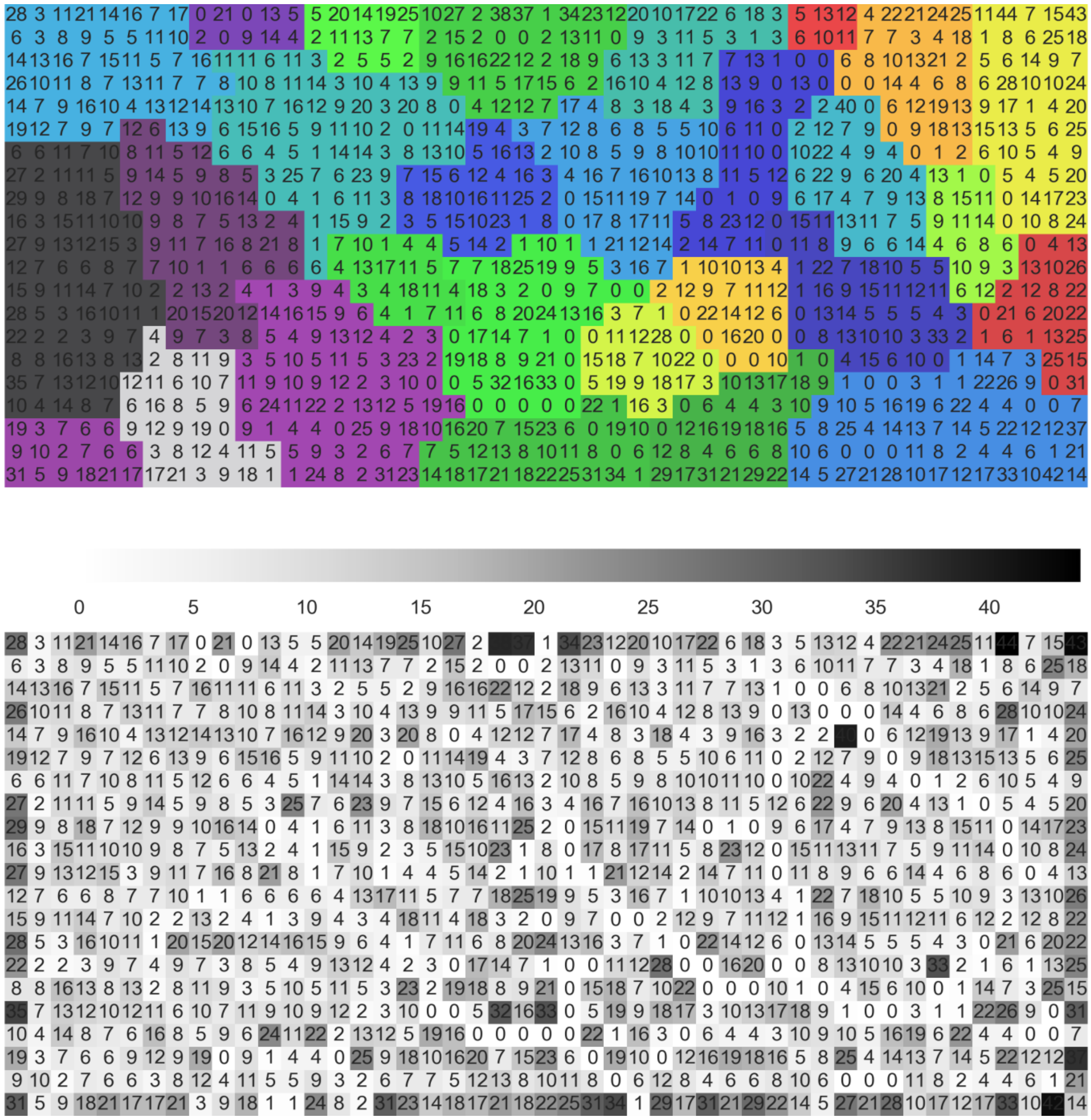


Figure D.25: Hits-maps showing the count of how many times a neuron became BMU. Top: Hits-map with cluster borders; bottom: hits-map as heatmap.

The following figures (Figures D.26-40) show some more evaluation results and quality checks of the final clustering and Table D.11 gives information on the number of households per cluster.

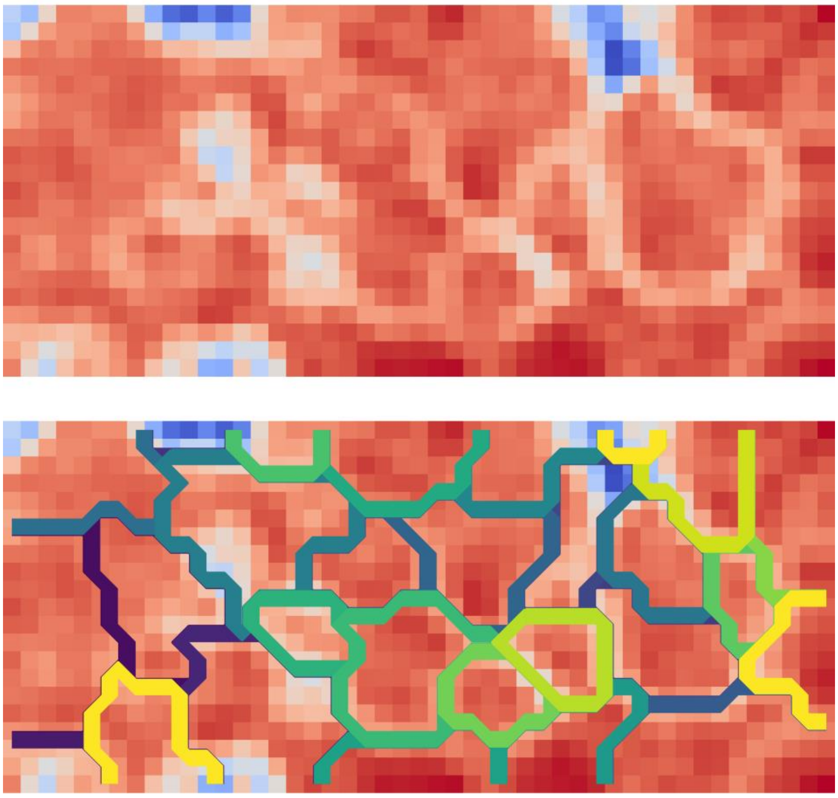


Figure D.26: Top: U-Matrix of the final SOM. Bottom: cluster borders projected onto the U-Matrix.

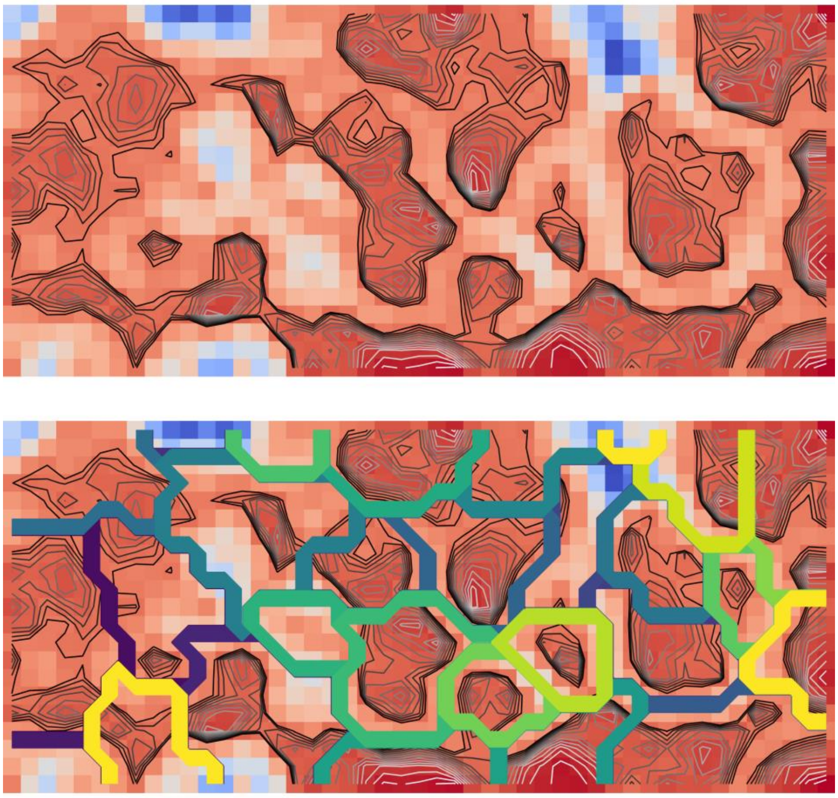


Figure D.27: Similar to Figure D.26, but with contour lines to improve visibility of suggested groupings of neurons by the U-matrix. Top: U-Matrix of the final SOM. Bottom: cluster borders projected onto the U-Matrix.

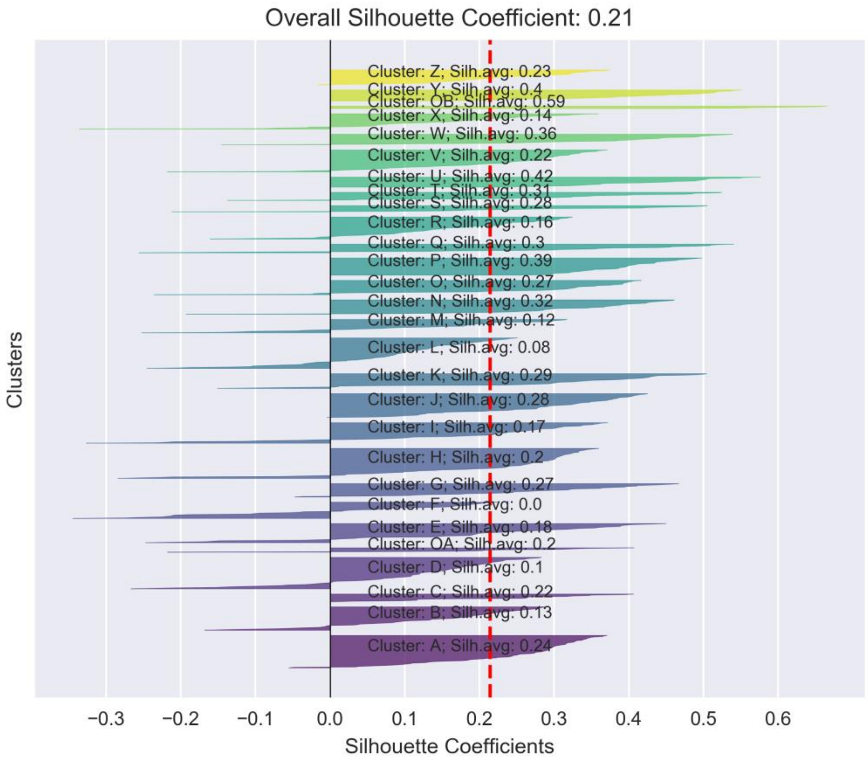


Figure D.28: Silhouette plot [53] of the final clustering, showing the silhouette scores for all samples. The sample scores are grouped by their cluster membership and ranked in decreasing order within each cluster. The red vertical line indicates the overall average Silhouette coefficient.

Table D.11: Statistics of number of households per cluster in the final clustering.

Cluster name	No. of households
A	730
B	450
C	182
D	597
E	351
F	244
G	296
H	691
I	414
J	561
K	315
L	557
M	257
N	369
O	420
P	433
Q	139
R	428
S	137
T	146
U	224
V	576
W	191
X	320
Y	296
Z	285
OA	68
OB	57

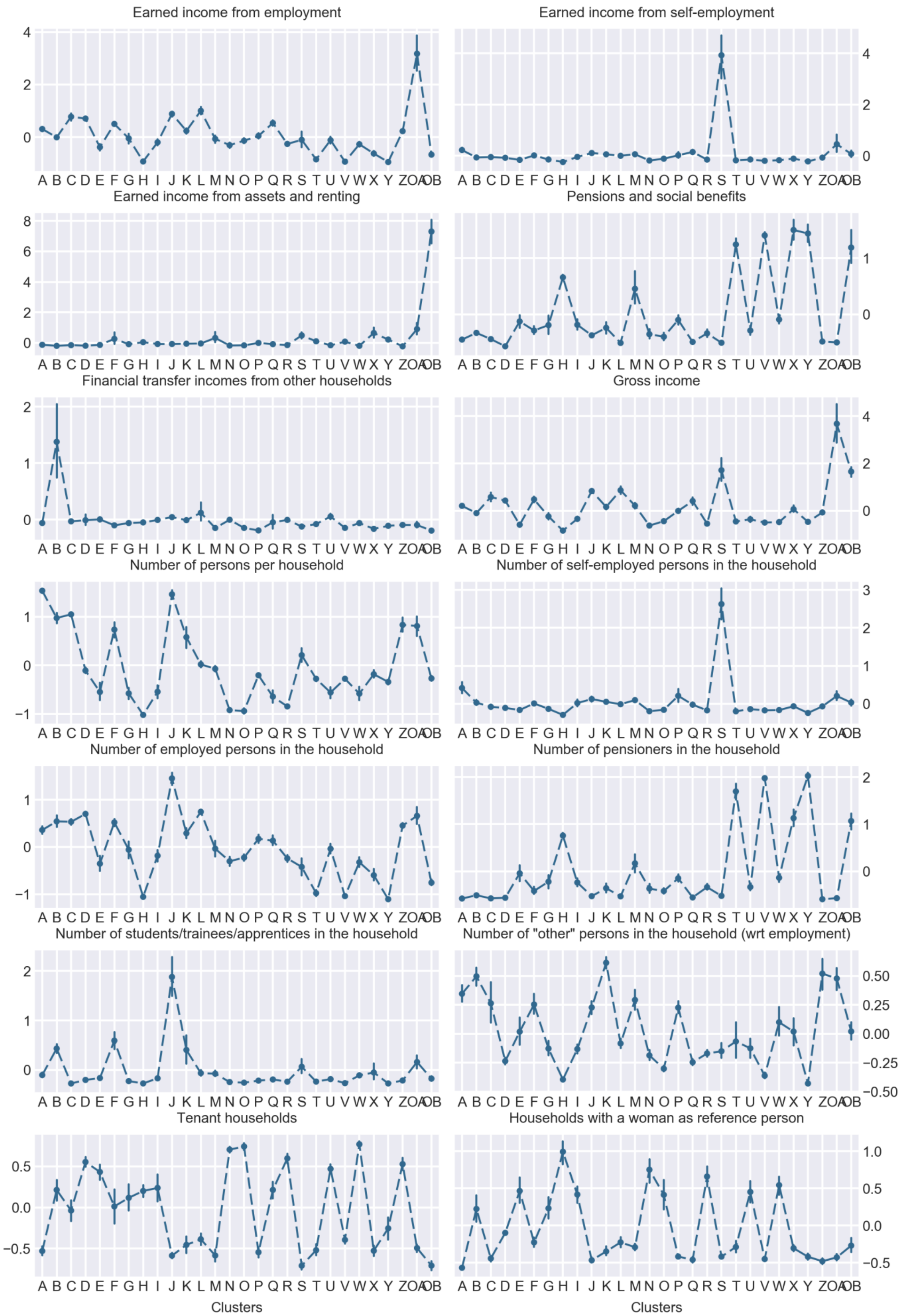


Figure D.29: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

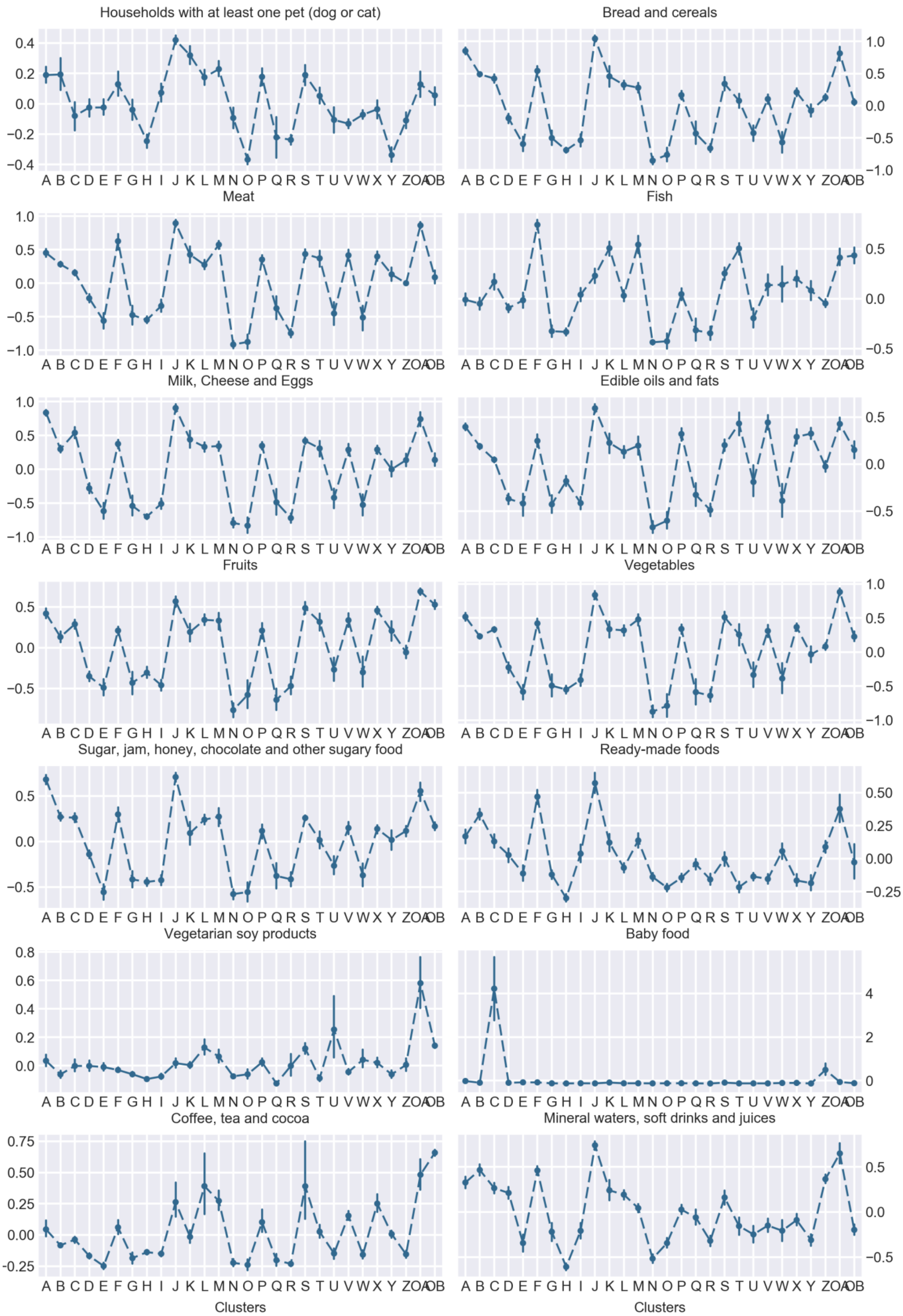


Figure D.30: Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".

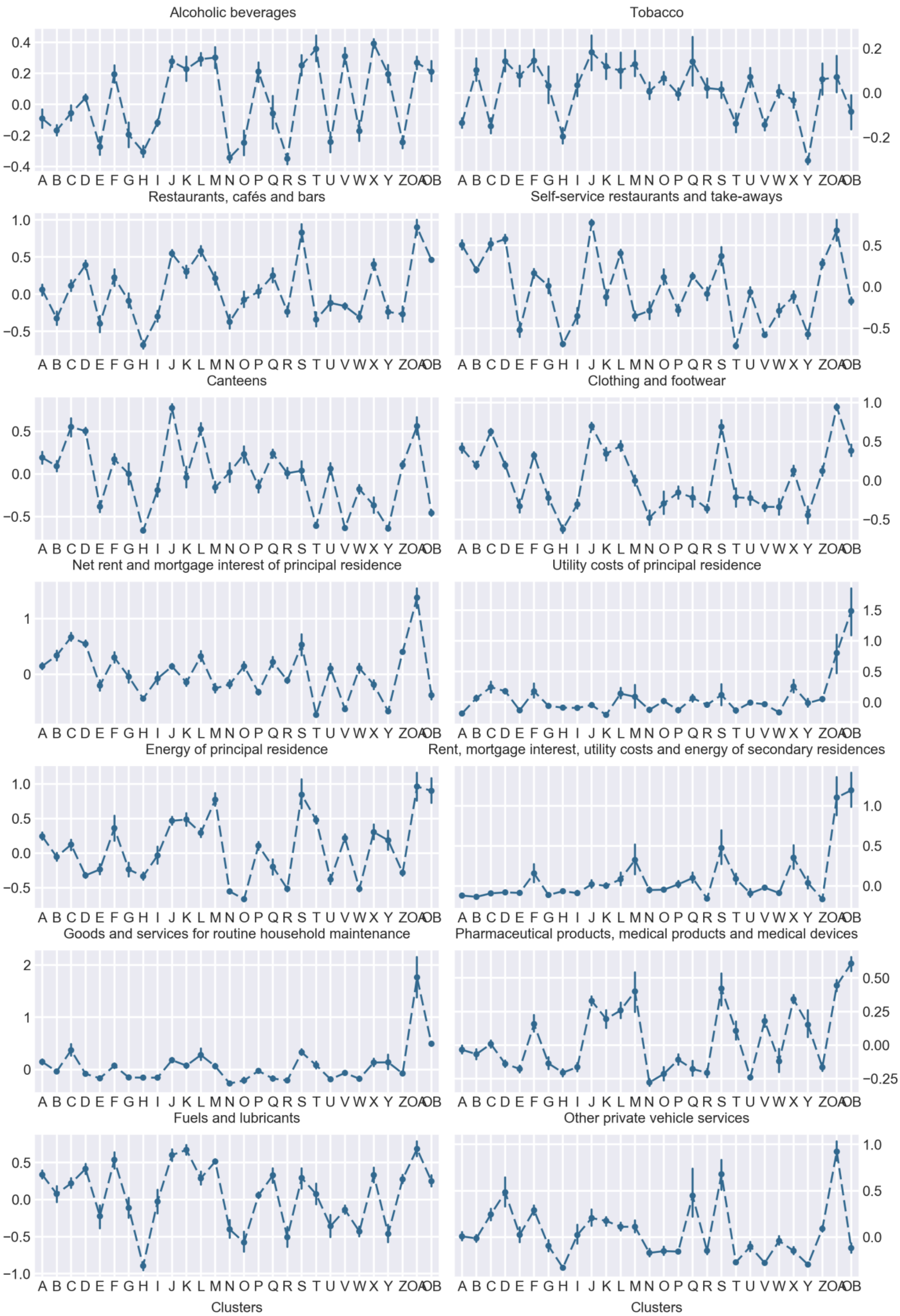


Figure D.31: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

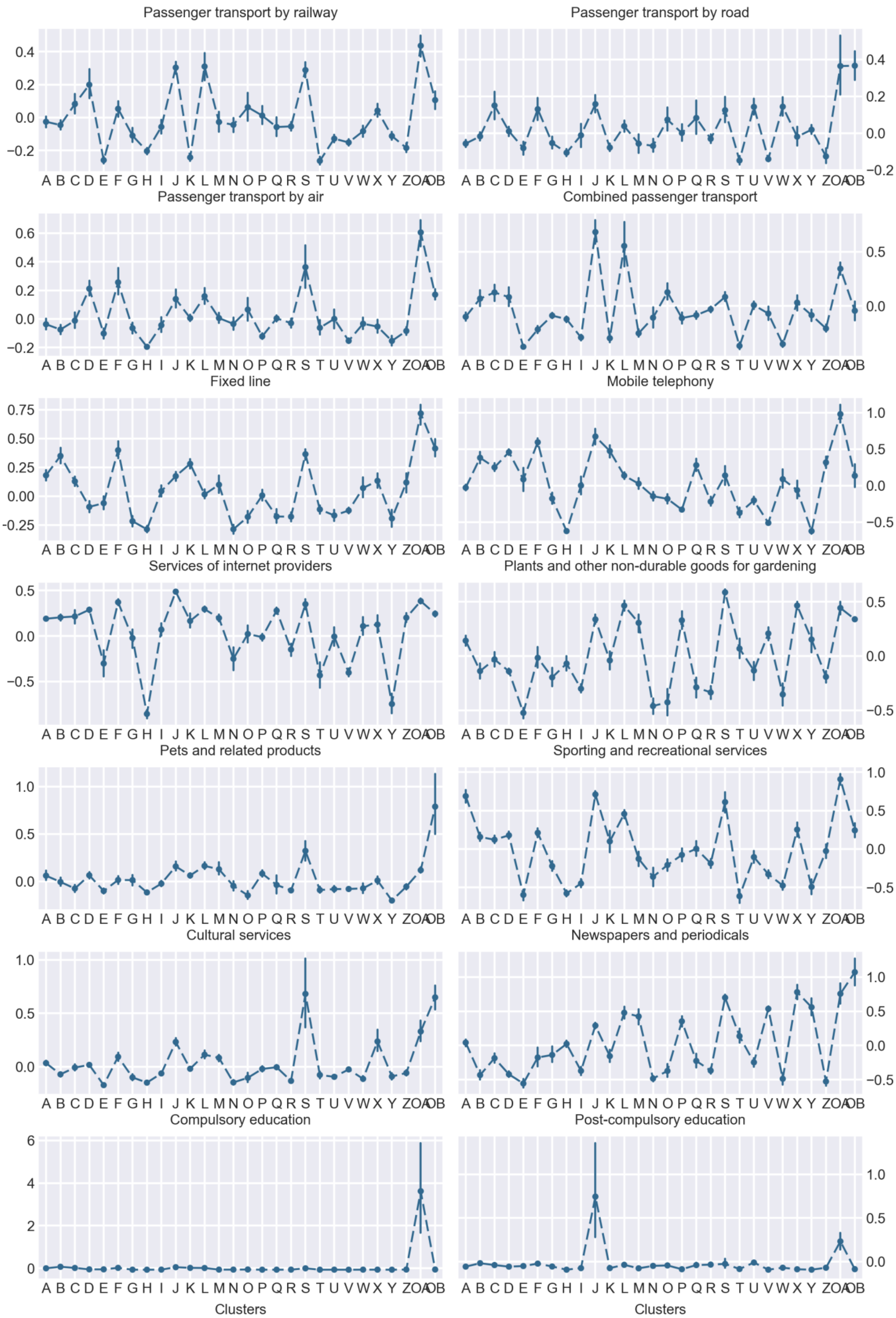


Figure D.32: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with “Q:”.**

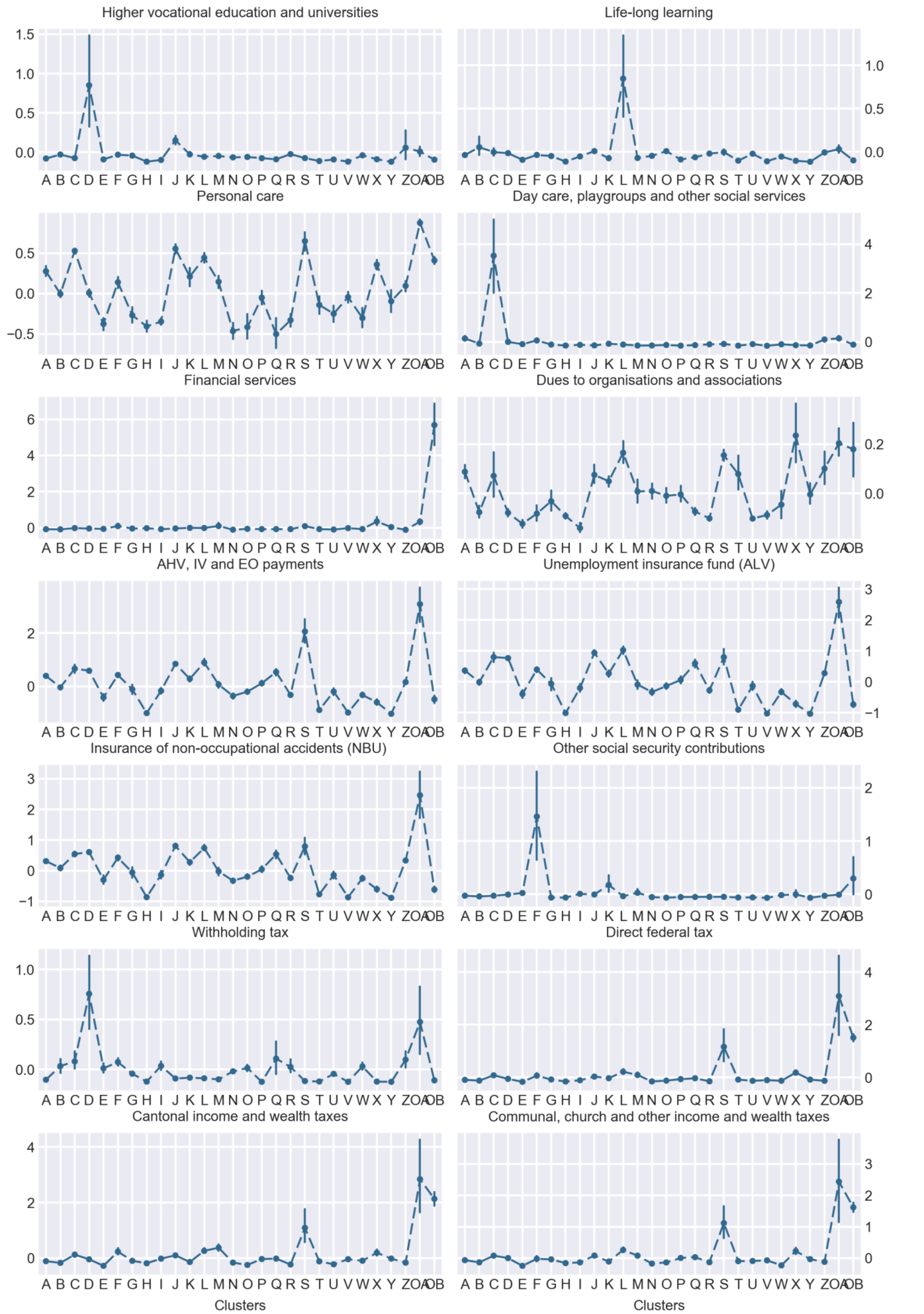


Figure D.33: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

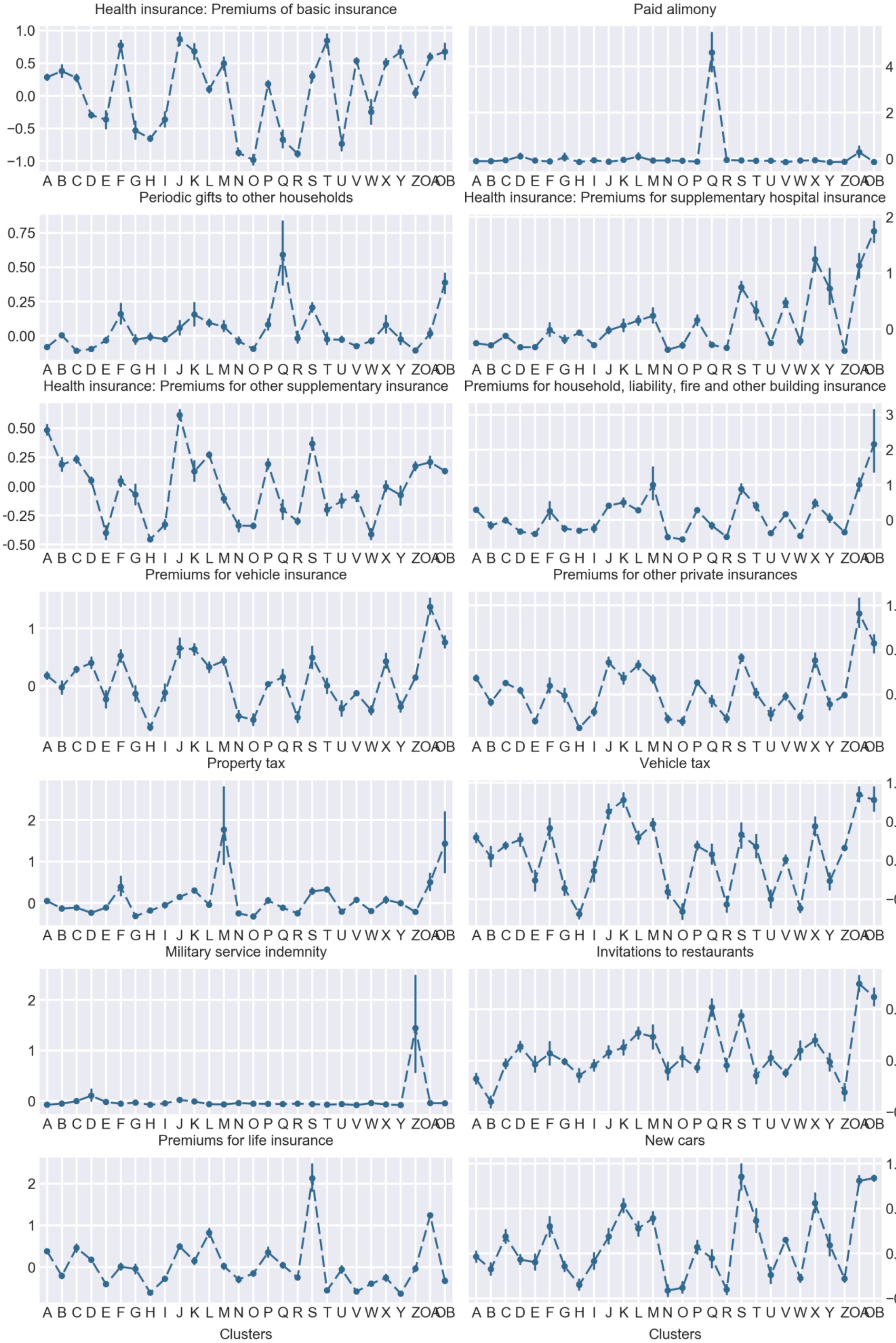


Figure D.34: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

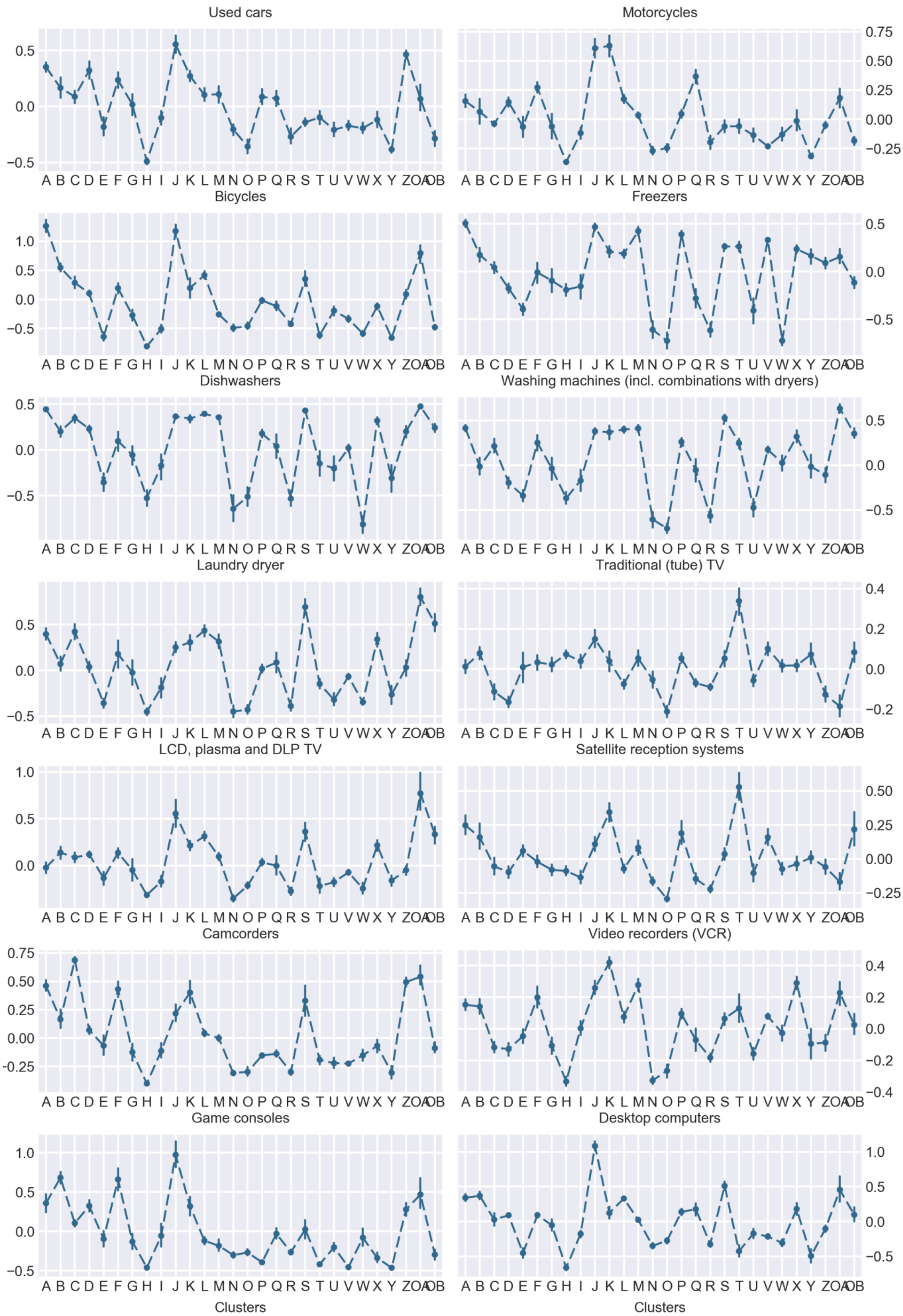


Figure D.35: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

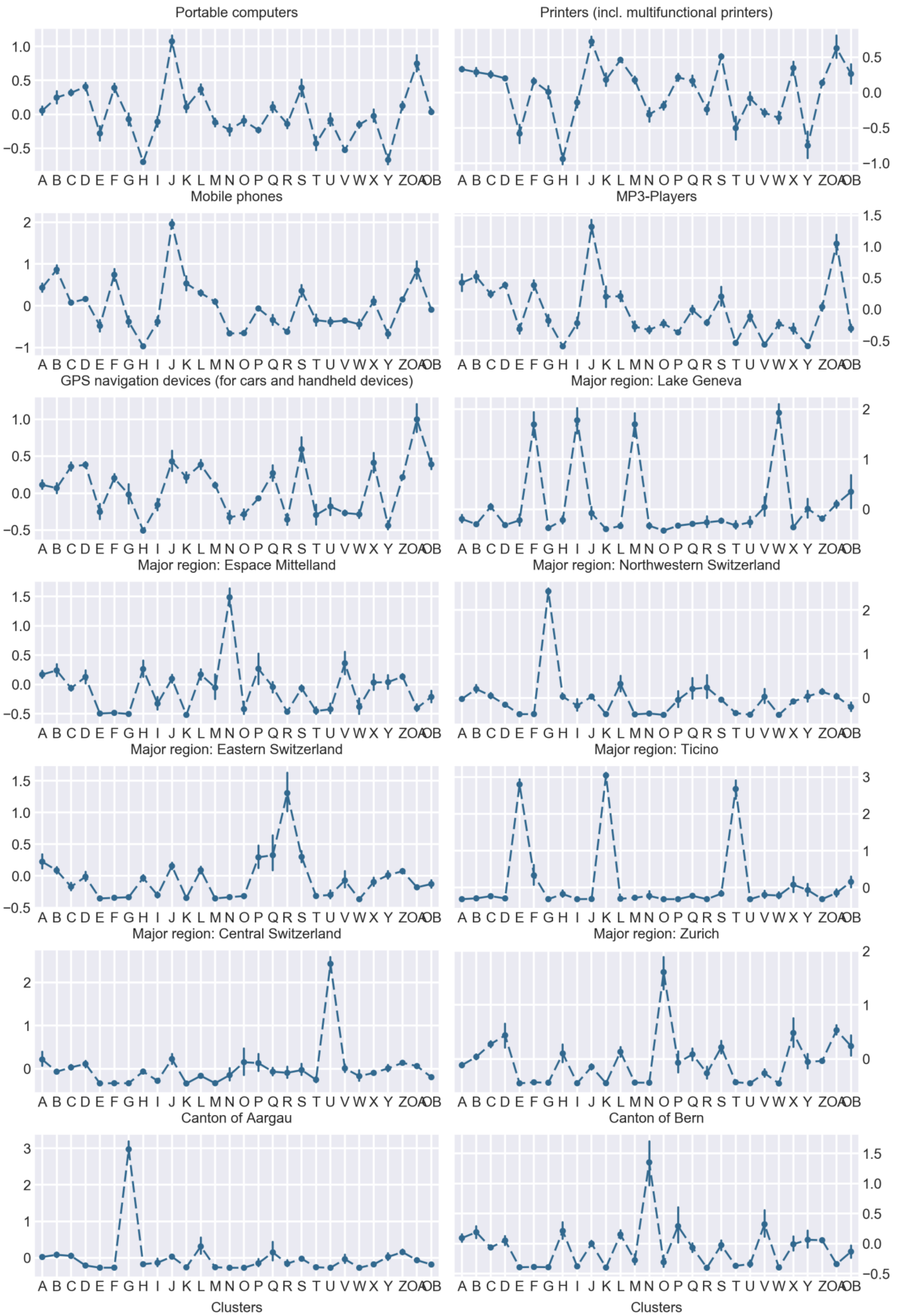


Figure D.36: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

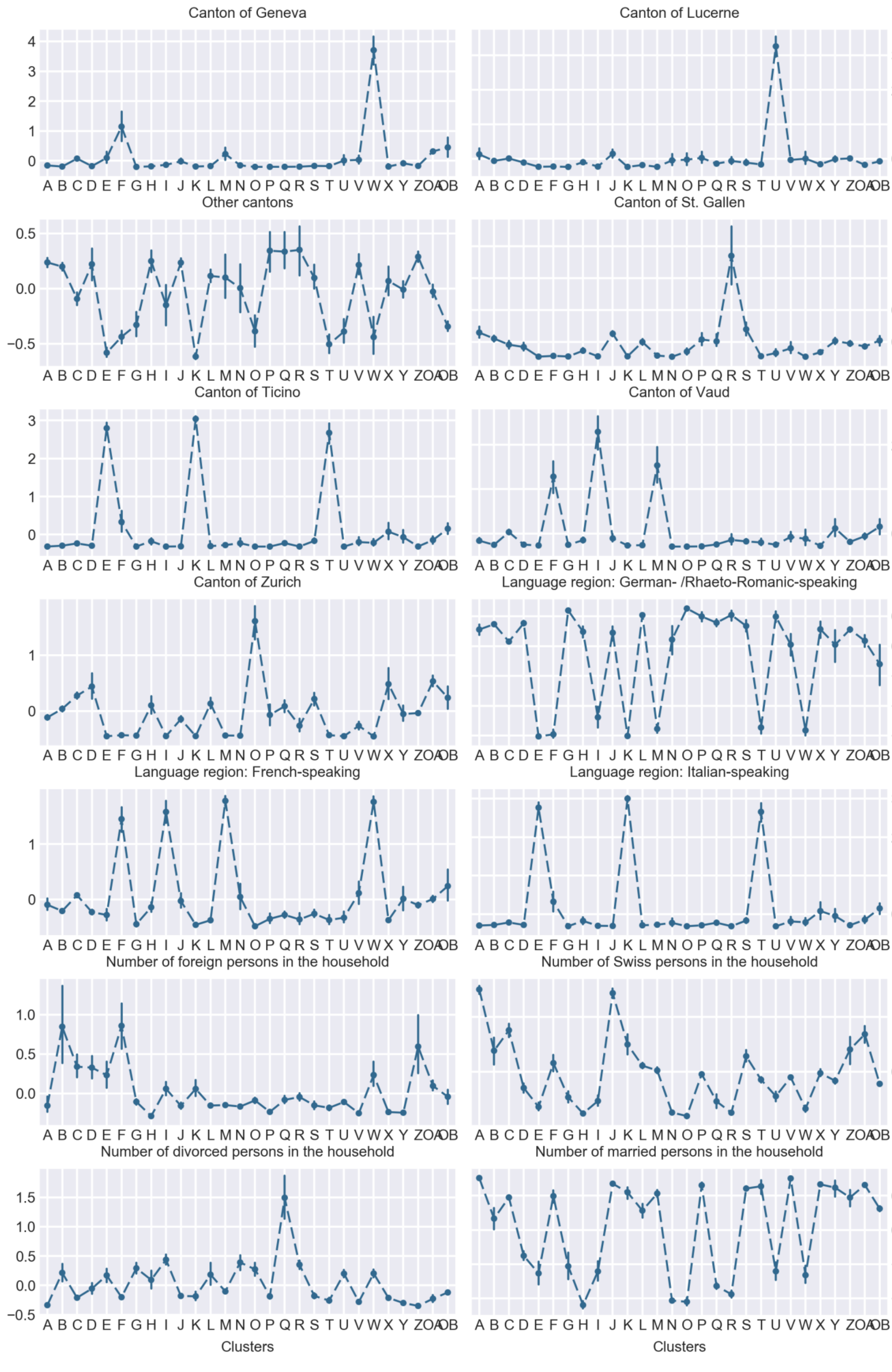


Figure D.37: Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with “Q:”.

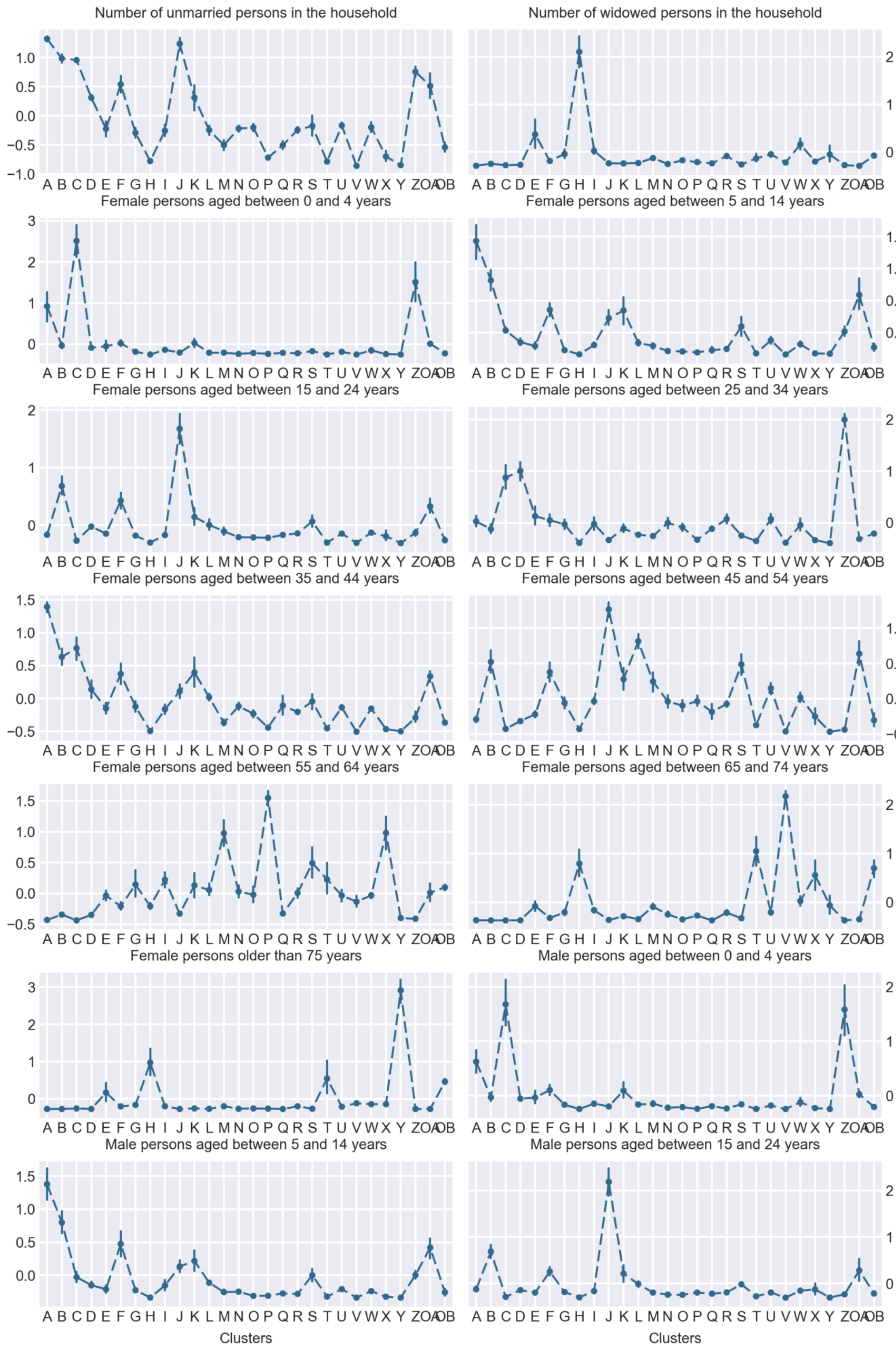


Figure D.38: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

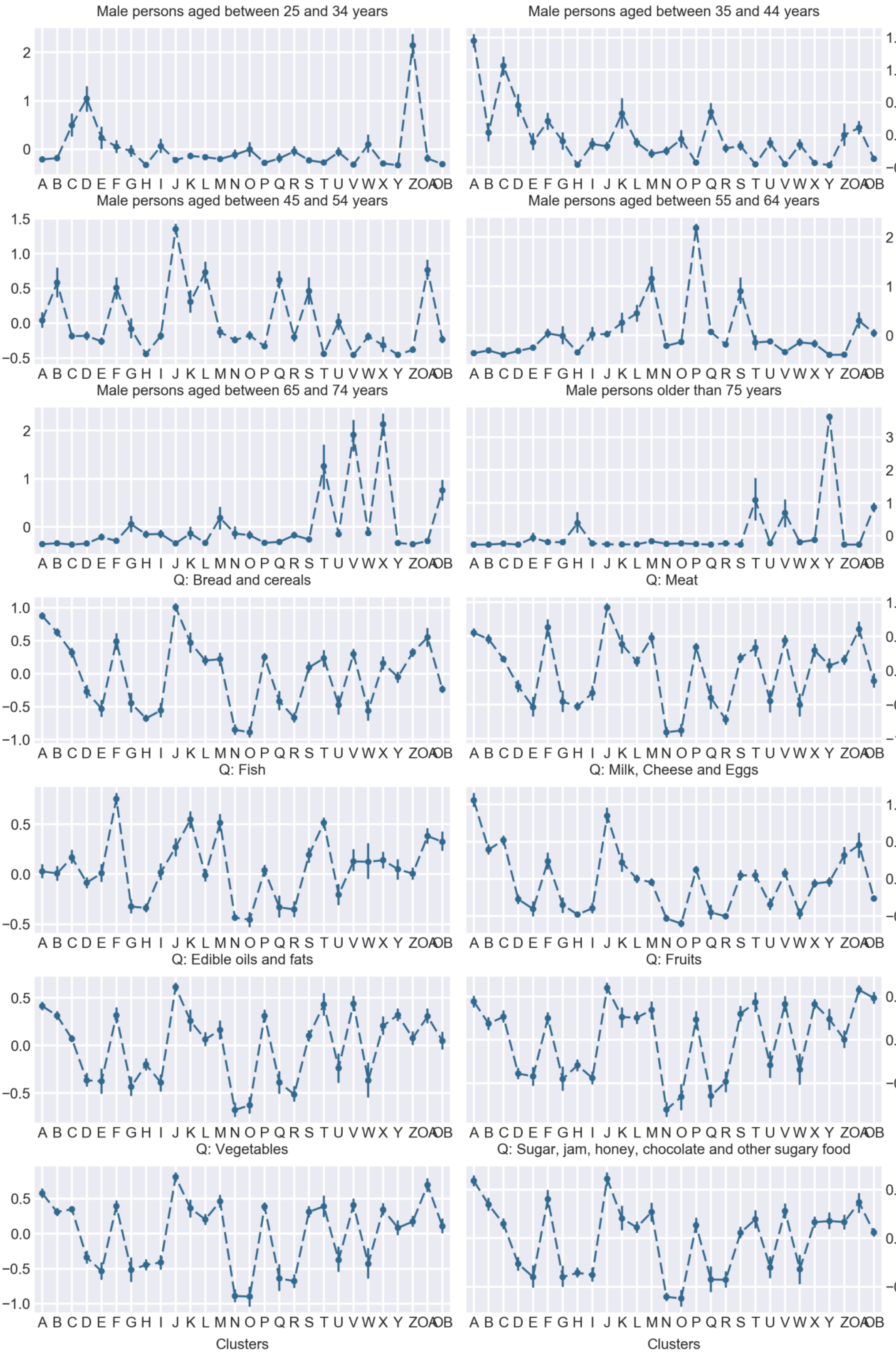


Figure D.39: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

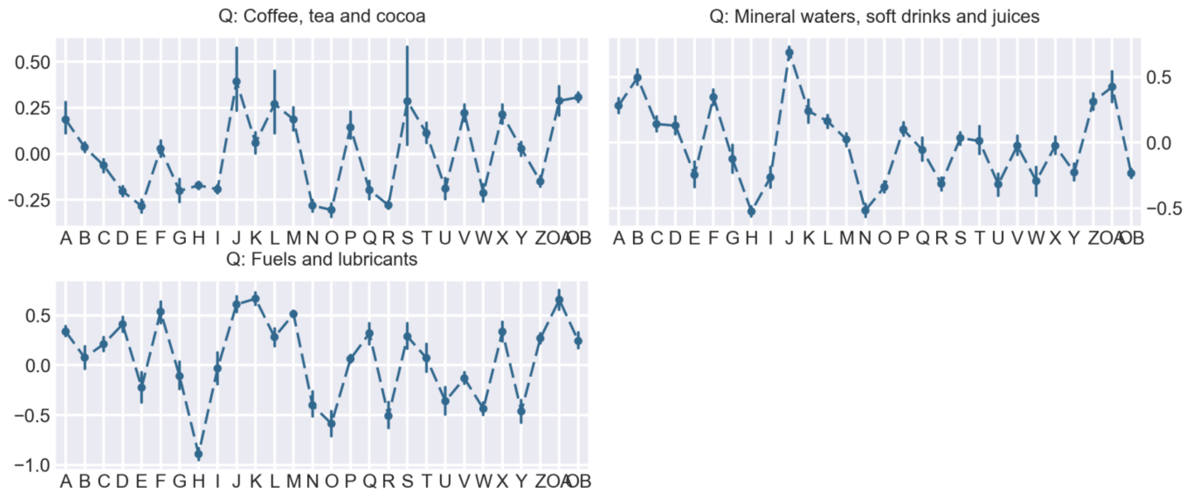


Figure D.40: **Plots to check the quality of the clustering. The points illustrate the centroids of the archetypes, while the error bars visualize the 95%-confidence interval of the centroids. The data is unit-less since standardized and partly corrected for seasonality. Quantity-attributes are marked with "Q:".**

D.4 LCA-Modeling

The whole LCA-modeling applied in this study is disclosed in the accompanying EXCEL-file[3]. This section shall explain how this EXCEL-file needs to be read and provide more insights into some peculiarities.

First of all, the EXCEL-file should be gone through line-by-line. All unit processes in a row form the process model which approximates the consumption category. Each unit process comprises 5 columns providing more information on how to include the unit process in the process model: "On" determines if the unit process is active (whether it shall be included or not); "Activity" holds the key to find the activity in the respective database via brightway2 [3]; this also shows if the unit process originates from ecoinvent [26], Agribalyse [59] or EXIOBASE [60, 61]; "DB Act" shows a human readable name of the unit process; "CFL Act" indicates a conversion factor for individual unit processes. Different uses of the "CFL Act"-field will be explained below. In the case of food categories, this factor corresponds to a conversion factor from Scherer and Pfister [62] which translates the unit process as it is given in the respective database into the actually needed product for the process model. The idea to use the factor from [62] is based on the LCA-modeling of food by Walker et al. [63]. "Amount Act" finally shows the amount needed for the functional unit of the respective consumption category. Thereby, the "Amount Act"-entry needs to be multiplied with the "CFL Act"-entry when the unit process enters the process model. Last but not least, the finally computed LCA-factor (characterization factor) for the functional unit of the process model is multiplied by the entry in the column "ConversionDem2FU" in order to convert the functional unit of the process model into the units of the demand. In the case of EXIOBASE-activities, this means for instance that the functional unit of the process model in

[3] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

millions of € of basic-prices are translated into purchaser-prices in Swiss Francs (see below). While for the very most of all cases, this conversion is self-explanatory since it is either the aforementioned translation of EXIOBASE-activities or simply the factor of 1 (in the case the HBS already provides the demands in kilograms or liters), some very few conversions need more explanations:

- For housing-related activities please refer to section D.2.3.
- The following categories needed to be converted from Swiss Francs to kilograms mainly based on price and weights data from the largest retailers in Switzerland [64, 65]:
  - Bakery products [64, 65]
  - Sandwich [64, 66]
  - Fresh eggs, processed eggs [67, 68]
  - Culinary herbs [64]
  - Confectionery [64, 65]
  - Other sugar or cocoa based foods [64, 65]
  - Sauces, seasonings and spices [64, 65]
  - Soups and bouillons [64, 65]
  - Ready-to-cook meals [64, 65]
  - Vegetarian soy products [64]
  - Newspapers and periodicals [64, 68]
  - Body wash and bath additive [69, 70]

Table D.12: Data to convert bought liters of petrol to vehicle-kilometers according to [26] **and based on the density of 0.75 kg/l from** [29].

ecoinvent-activity	Petrol use kg/km	Conversion km/l
'transport, passenger car, small size, petrol, EURO 3' (kilometer, RER, None)	0.058	12.974
'transport, passenger car, small size, petrol, EURO 4' (kilometer, RER, None)	0.054	14.006
'transport, passenger car, small size, petrol, EURO 5' (kilometer, RER, None)	0.050	14.897
'transport, passenger car, medium size, petrol, EURO 3' (kilometer, RER, None)	0.070	10.789
'transport, passenger car, medium size, petrol, EURO 4' (kilometer, RER, None)	0.065	11.475
'transport, passenger car, medium size, petrol, EURO 5' (kilometer, RER, None)	0.062	12.084
'transport, passenger car, large size, petrol, EURO 3' (kilometer, RER, None)	0.081	9.232
'transport, passenger car, large size, petrol, EURO 4' (kilometer, RER, None)	0.077	9.720
'transport, passenger car, large size, petrol, EURO 5' (kilometer, RER, None)	0.074	10.166

Furthermore, bought liters of petrol and diesel were converted to vehicle-kilometers. This was done based on the density information given in [29] (0.75 kg/l for petrol and 0.84 kg/l for diesel) and the amounts of kilograms of diesel and petrol used to drive 1 kilometer according to ecoinvent [26]. For convenience, this data is also made available in Table D.12 and Table D.13 and entered accordingly the “CFL Act”-field in the EXCEL-file. By the way, the “Amount Act” in the EXCEL-file for the use of diesel and petrol corresponds to the shares of the different petrol and diesel cars in the total Swiss petrol and diesel car fleet [71]. Similarly, the “Amount Act”-field for trains and urban vehicles contains the shares based on Swiss public transportation statistics [72].

Table D.13: Data to convert bought liters of diesel to vehicle-kilometers according to [26] and based on the density of 0.84 kg/l from [29].

ecoinvent-activity	Diesel use kg/km	Conversion km/l
'transport, passenger car, small size, diesel, EURO 3' (kilometer, RER, None)	0.045	18.688
'transport, passenger car, small size, diesel, EURO 4' (kilometer, RER, None)	0.043	19.333
'transport, passenger car, small size, diesel, EURO 5' (kilometer, RER, None)	0.042	20.169
'transport, passenger car, medium size, diesel, EURO 3' (kilometer, RER, None)	0.061	13.850
'transport, passenger car, medium size, diesel, EURO 4' (kilometer, RER, None)	0.057	14.622
'transport, passenger car, medium size, diesel, EURO 5' (kilometer, RER, None)	0.056	15.095
'transport, passenger car, large size, diesel, EURO 3' (kilometer, RER, None)	0.076	11.017
'transport, passenger car, large size, diesel, EURO 4' (kilometer, RER, None)	0.072	11.740
'transport, passenger car, large size, diesel, EURO 5' (kilometer, RER, None)	0.070	12.061

As mentioned in Chapter 5, we generally attempted to adjust the process models as close as possible to the situation of Switzerland. With regard to ecoinvent-activities, this means, we tried to construct "Swiss Markets" in the case there are no Swiss activities directly available. Thereby, the different production activities were weighted according to the import shares provided by Scherer and Pfister [62]. Due to a lack of information, associated transport activities could not be adjusted and correspond thus to the transport modeling provided by ecoinvent [26] for the respective global markets. In the EXCEL-file, these global market transport processes are indicated to be located in a database called "heia", which only contains manually adjusted ecoinvent-activities. Of course, if neither import shares are given by Scherer and Pfister [62] nor ecoinvent [26] distinguishes different production processes, the global-market-activities were taken to approximate the respective part of a process model.

Additionally, the previously mentioned "heia"-database also contains a fruit- and vegetable-mix (called "fruitnes" and "vegetablenes"). These mixes were also created based on the import data from Scherer and Pfister [62] and their composition is presented in Table D.14. Again, whenever possible, Swiss markets were also generated for activities being part of these mixes.

Table D.14: Swiss fruit- and vegetable-mixes to cover process models which need a general unit activity for fruits or vegetables. The shares are retrieved from [62].

Fruitnes		Vegetablenes	
Fruit	Share	Vegetable	Share
Lemon	0.027	Spinach	0.0611
Orange	0.171	Fennel	0.0170
Mandarin	0.021	Celery	0.0170
Banana	0.112	Onion	0.0862
Apple	0.381	Cabbage red	0.0604
Pear	0.081	Cabbage white	0.0604
Peach	0.047	Cauliflower	0.0331
Apricot	0.032	Tomato	0.3567
Melon	0.040	Fava bean	0.0018
Pineapple	0.066	Aubergine	0.0134
Palm date	0.003	Zucchini	0.0182
Kiwi	0.017	Green bell pepper	0.0182
Papaya	0.002	Cucumber	0.0893
		Carrot	0.1468
		Asparagus green	0.0102
		Asparagus white	0.0102

Similar to the modeling of the car-driven kilometers, also the heating technologies need some more explanations. The conversion of expenditures to final energy was already explained in section D.2.3. The “Amount Act”-field in the EXCEL-file carries the averaged shares given in Table D.5 while the “CFL Act” contains an average efficiency to translate final energy to useful energy based on data given by [73]. To split the shares among technologies with the same energy carrier, either additional data could be used from [73] or equal shares needed to be assumed. However, a quick sensitivity analysis revealed that this assumption is of minor importance.

Another issue with regard to housing energy are expenditures for secondary residences. Unfortunately, the HBS does not provide a detailed breakdown of expenditures for secondary homes. Environmental impacts induced by secondary residences are thus approximated as follows: The total of secondary home costs are – just virtual – proportionally distributed to the primary home categories. The increase in primary home expenditures is then directly transferred to an analogue increase in amounts of energy, electricity, water, wastewater, and waste bags. The difference between the environmental impacts caused by this “virtual” increase and the environmental impacts before this increase are determined as the impacts originating from secondary homes.

As mentioned in Chapter 5, the coupling of HBS-consumption categories with EXIOBASE-activities followed the recommendations of Steen-Olsen et al. [74]. In the following, we will give a short overview of the linking-process and highlight where our approach deviates from [74]:

1. Back-calculating from HBS-prices (2009, 2010, 2011) to prices of 2007 (EXIOBASE-basis) by means of data from [70].
2. Conversion from million € to Swiss Francs based on [75].
3. We then matched HBS-consumption categories with EXIOBASE-activities based on the ISIC rev. 3.1 [76] descriptions. However, EXIOBASE obviously merged some industry sectors and for some parts uses the ISIC rev. 4. We thus tried to follow the EXIOBASE-documentation and the respective ISIC-descriptions, but also double-checked with the Federal Statistical Office in order to match to the correct consumption categories. In the case of one-to-many-matches, Steen-Olsen et al. [74] applied an optimization process to determine the shares of different EXIOBASE-sectors for a certain consumption category. Because of the different modeling structure in our hybrid LCA, we were not able to employ the same optimization approach, but decided to determine the shares of different sectors by the Swiss final demand vector provided by EXIOBASE (these shares entered the “CFL Act”-fields). Similarly, the contributions from different countries and regions to a certain industry sector were also weighted according to the household consumption vector of Swiss households (these shares entered the “Amount Act”-fields).
4. Finally, the purchaser-prices (HBS-data) needed to be converted to basic-prices (EXIOBASE-basis). The “ConversionDem2FU”-field in the EXCEL-file thus not only considers the conversions in steps 1 and 2, but also the subtraction of taxes as well as trade and transport margins. The trade and transport margins were then re-distributed to the

trade- and margins-sectors which also become part of the different EXIOBASE-process models. For these sectors, the "Amount Act"-field contains the share of the respective margin-sector according to EXIOBASE, while the "CFL Act"-entry corresponds to the share of a purchase going to the margins-sectors.

Finally, we need to point out that EXIOBASE provides a large, but – compared to e.g. ecoinvent – limited number of biosphere flows. Thereby, greenhouse gases are well covered. This limitation is thus not an issue for the results according to IPCC 2013 (100a). However, the comprehensive impact assessment method of ReCiPe is slightly affected. While climate change, acidification, eutrophication, particulate matter formation, photochemical oxidant formation and impacts of land use are well considered, toxicity can fairly be accounted for and resource depletion at least in parts. However, ozone depletion and ionizing radiation – to our knowledge – cannot be assessed by the environmental data provided by EXIOBASE. But we also would like to highlight that the most important biosphere flows for ReCiPe [77] are included in EXIOBASE and that mobility, food and housing, which are the three most important consumption areas from an environmental perspective (see e.g. [74, 78–82]), are anyway modeled by ecoinvent-activities. Moreover, the presented ReCiPe-results can also be regarded as plausible since the prevalence-weighted Swiss average of this study amounts to 0.98 kPts. per person per year, which is very close to the 1.0 kPts. per person per year for the average citizen in the European Union (please note that this average of 1000 points per person per year are the result of the normalization and weighting step of ReCiPe and thus set by definition for the average footprint of EU-citizens in the year 2000 [77]).

Last but not least, system boundaries with regard to invitations and donations were set as follows: if a household invites another household to a restaurant or hotel, the environmental impacts are allocated to the invited household, except for the case, the invitation takes place at the inviting household's home. Even though these assumptions might be questionable or seem to be contradictory to some extent, it is the only way to avoid double-counting and to make full use of the HBS-data. The food bought for an invitation to one's own home cannot be distinguished from other food purchases. At the same time, the expenditures for restaurants and hotels also include costs which were not paid by the surveyed household.

D.5 Additional Results

D.5.1 Results in the Excel-File

The supplemental EXCEL-file[4] was already mentioned and partly described in sections D.1.2 (overview of HBS-structure), D.3.1 (indication of which attributes passed the filtering process and which of them were corrected for seasonality), and D.4 (full description of LCA-modeling). Besides the aforementioned information, the EXCEL-file also contains the following additional results:

[4] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

- Results of ANOVA applied to all clustering attributes as well as results of visual quality check.
- The centroid vector of all clusters containing the average demands, amounts, expenditures, household characteristics, income and durable goods statistics for all attributes provided by the HBS.
- The LCA-results for IPCC 2013 (100a) [83].
- The LCA-results for ReCiPe 2008 (total endpoints, H,A) [84].

D.5.2 Heatmaps of Household Characteristics

In addition to the full results of the centroid vector provided in the supplementary EXCEL-file[5], Figure D.41 and Figure D.42 visually support the comparison of the archetypes' household characteristics. These heatmaps are normalized along the attribute's axis on a minimum-maximum-scale. For instance, if gross income shall be compared, the cluster with the largest gross income is assigned a value of 1, while the minimum income receives a 0. All others are scaled proportionally between 0 and 1.

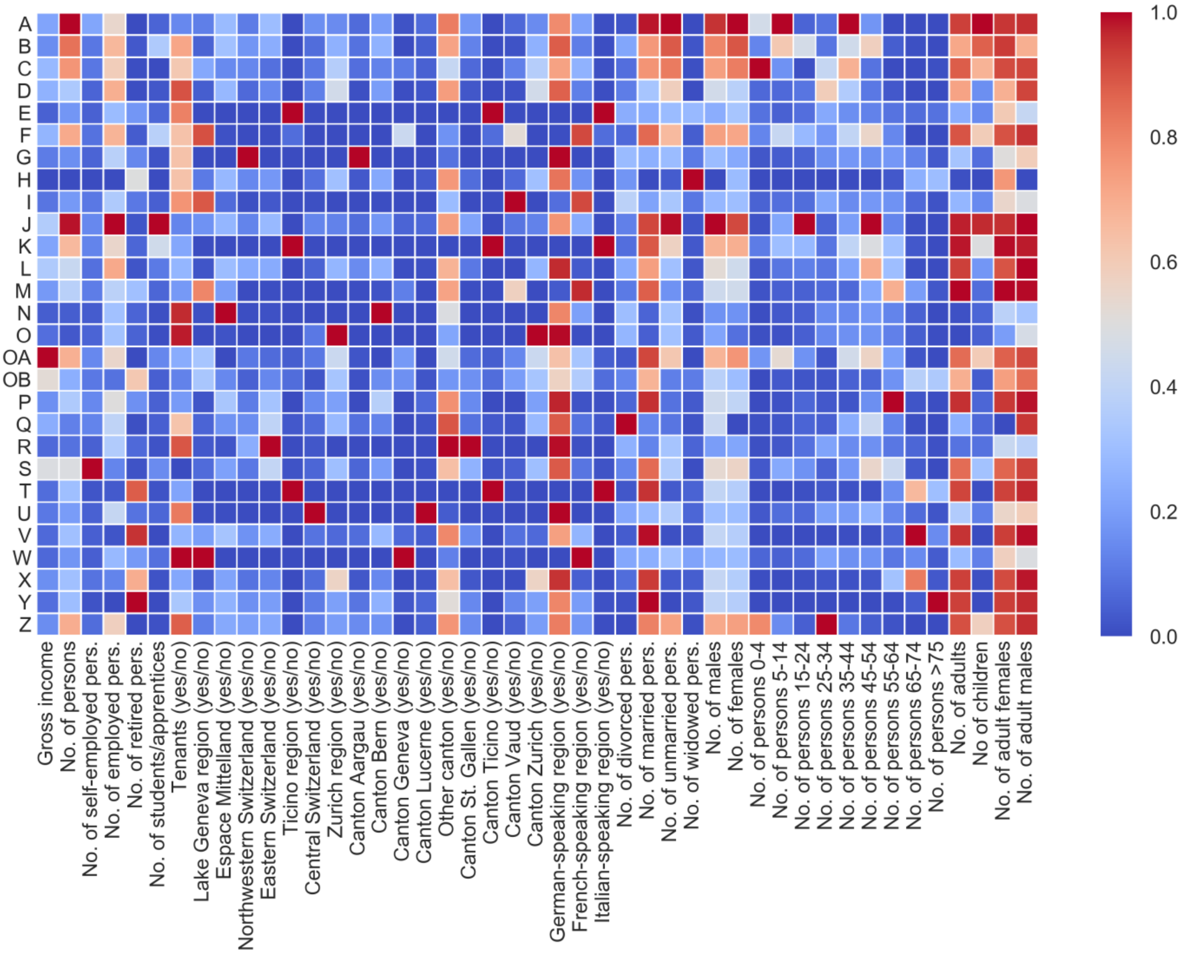


Figure D.41: **Heatmap of most important household characteristics (total view, not per-capita).**

[5] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

Some characteristics are better compared on a per-capita basis. Therefore, this view is also provided in Figure D.42. Please note that in both figures, not all household characteristics are displayed, but only the most important ones and some characteristics are provided in an aggregated form.

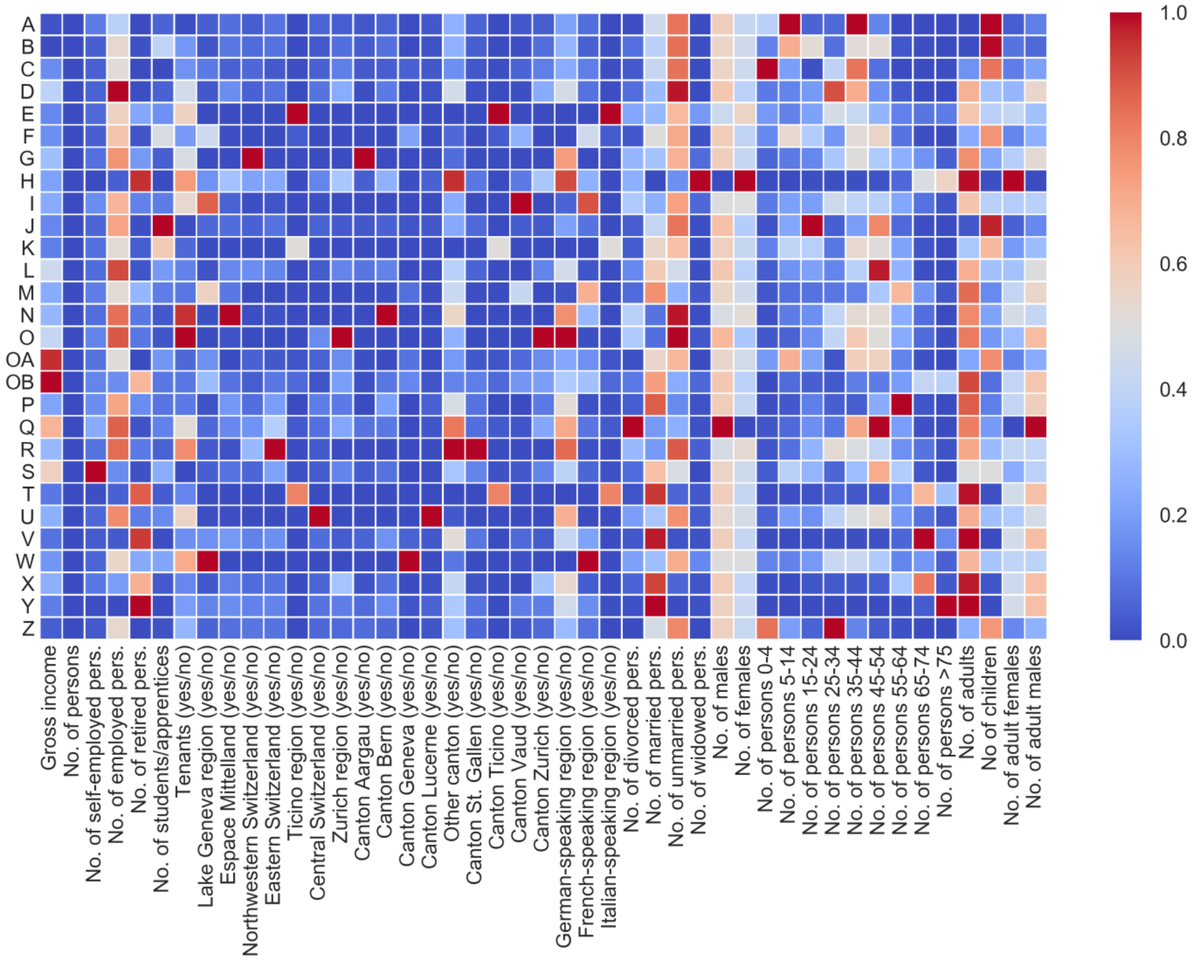


Figure D.42: **Heatmap of most important household characteristics (per-capita).**

D.5.3 Heatmaps of Household Demands

Similar to the previous section, Figure D.43 and Figure D.44 help to compare the archetypes' centroid vectors with regard to expenditures. Again, a minimum-maximum-scaling was applied and Figure D.43 shows the totals-perspective, while Figure D.44 enables a per-capita comparison. It was not possible to show all expenditures in a single heatmap. Besides excluding durable goods statistics, the second and sometimes the third lowest level of expenditures according to the HBS-structure is shown. Furthermore, some categories which were considered "less important" were also excluded for constructing the heatmaps. However, we would like to point out that all results are provided in the supplementary EXCEL-file[6].

[6] The EXCEL-file can be downloaded at https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b01452 or it can be requested via froemelt@ifu.baug.ethz.ch.

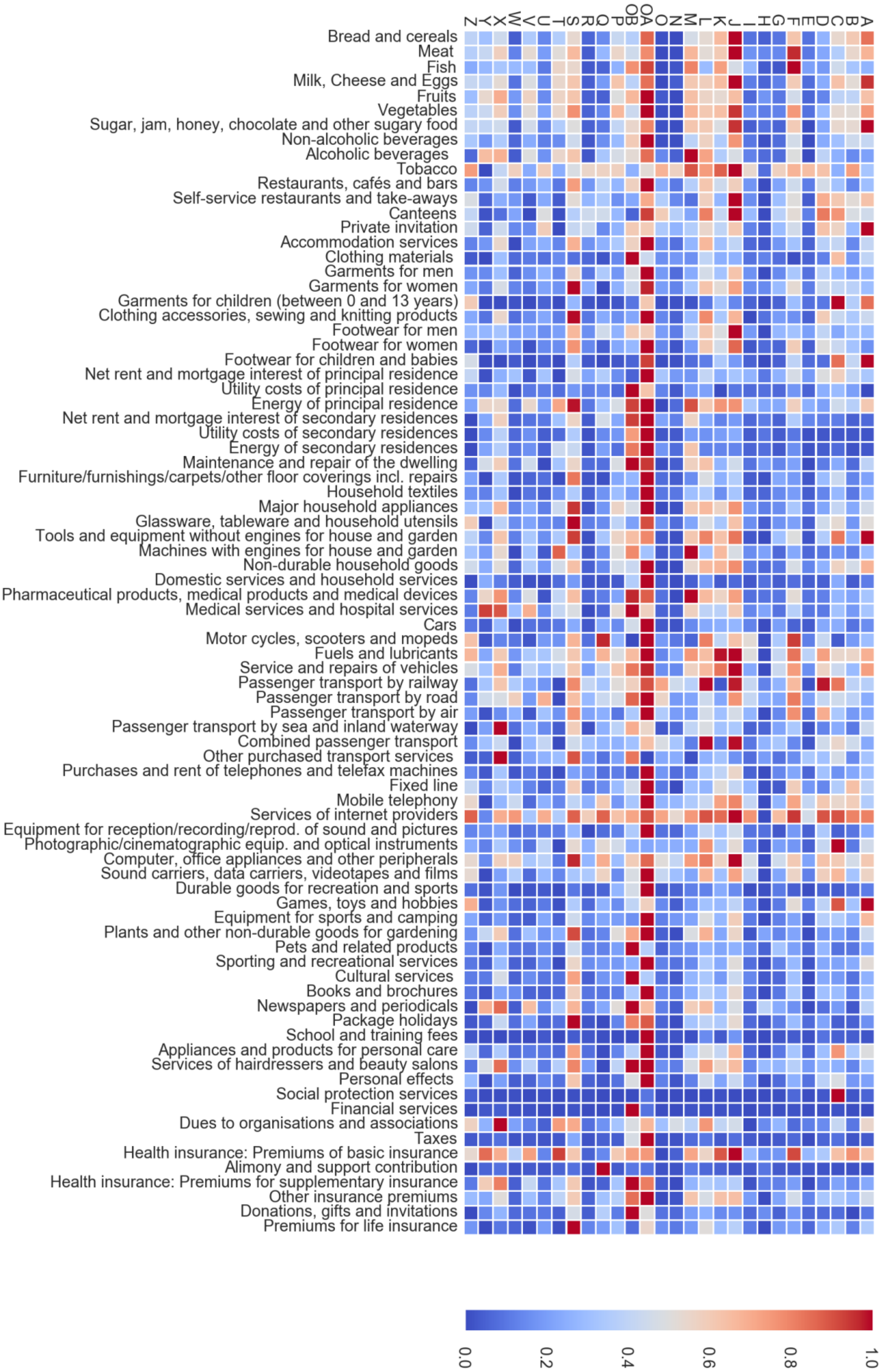


Figure D.43: **Heatmap of selected expenditures (total view, not per-capita).**

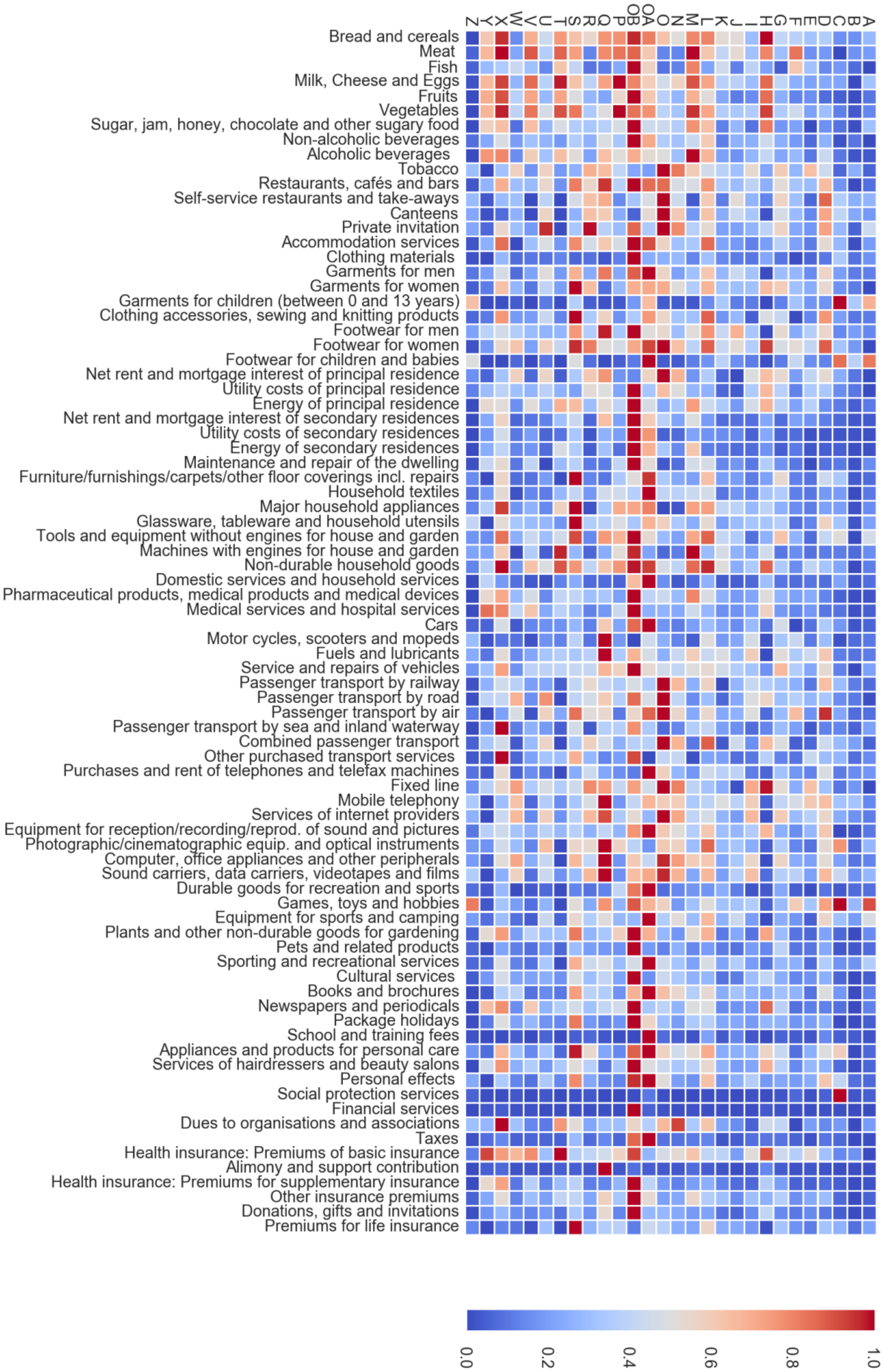


Figure D.44: **Heatmap of selected expenditures (per-capita).**

D.5.4 Heatmaps of Environmental Impacts

This section shall finally also provide a visually enhanced way to compare the environmental impacts of the archetypes. Figures D.45-48 visualize the results of greenhouse gas emissions (IPCC 2013, 100a), while Figures D.49-52 show the results of ReCiPe-endpoints (H,A). Additionally, different views are presented:

- Figure D.45 (greenhouse gas emissions) and Figure D.49 (ReCiPe-endpoints) take a "per-cluster-view". This means, the minimum-maximum-scaling is applied along the archetype's axis and allows comparing the main consumption categories within each archetype.
- Figure D.46 (greenhouse gas emissions) and Figure D.50 (ReCiPe-endpoints) provide a "per-category-view" and allows for comparisons among archetypes but within a certain consumption category.
- Figure D.47 (greenhouse gas emissions) and Figure D.51 (ReCiPe-endpoints) are prevalence-weighted and are not scaled enabling comparisons among different archetypes and different consumption areas at the same time.
- Finally, Figure D.48 (greenhouse gas emissions) and Figure D.52 (ReCiPe-endpoints) enable a prevalence-weighted "per-category-view".

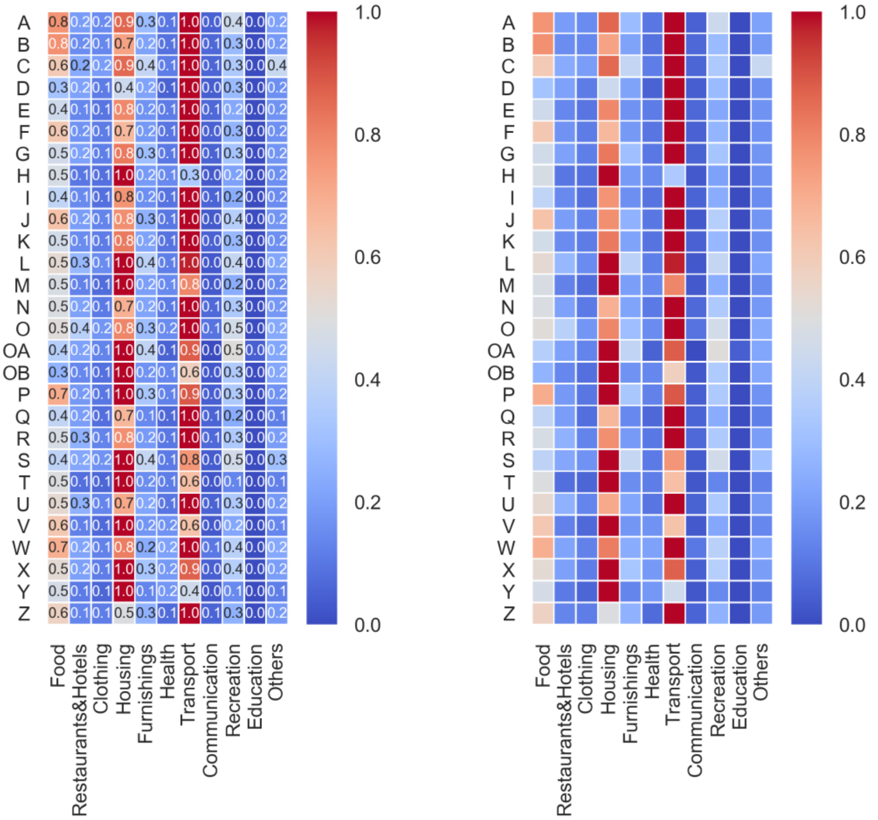


Figure D.45: **Greenhouse gas emissions (IPCC 2013, 100a), "per-cluster-view": Minimum-maximum-scaling along the archetype's axis to compare the consumption categories for each archetype separately. Left and right figures show the same, but the left-hand side also provides annotations of the applied scaling.**

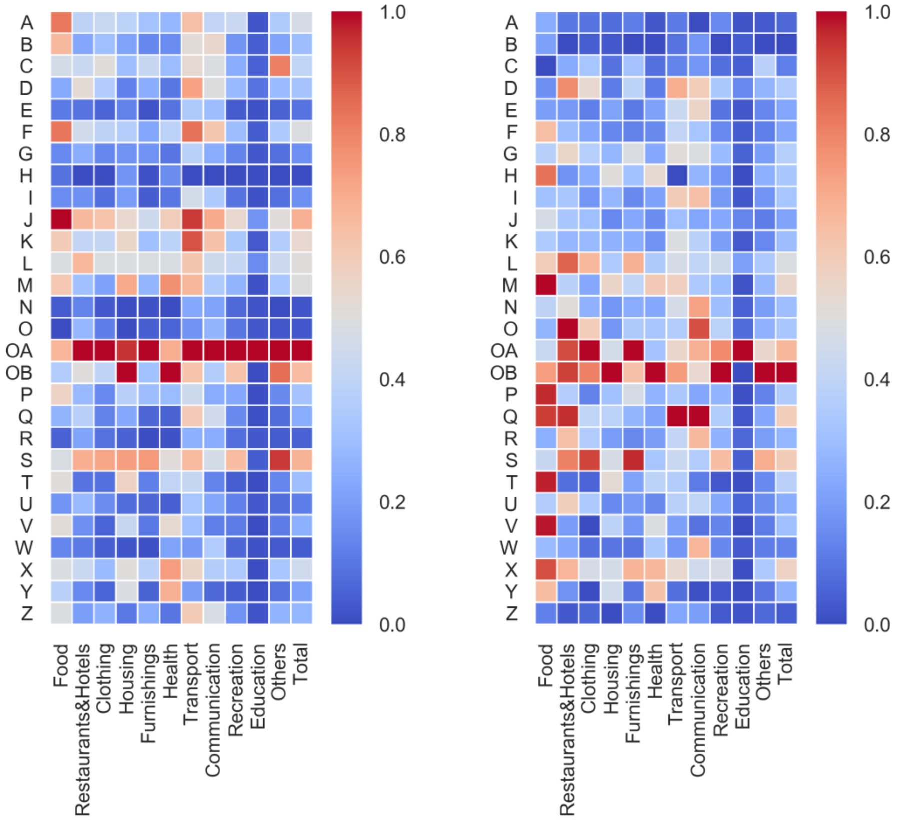


Figure D.46: Greenhouse gas emissions (IPCC 2013, 100a), "per-category-view": Minimum-maximum-scaling along the consumption area axis to compare different archetypes within a consumption area. Left: total-view; Right: per-capita view.

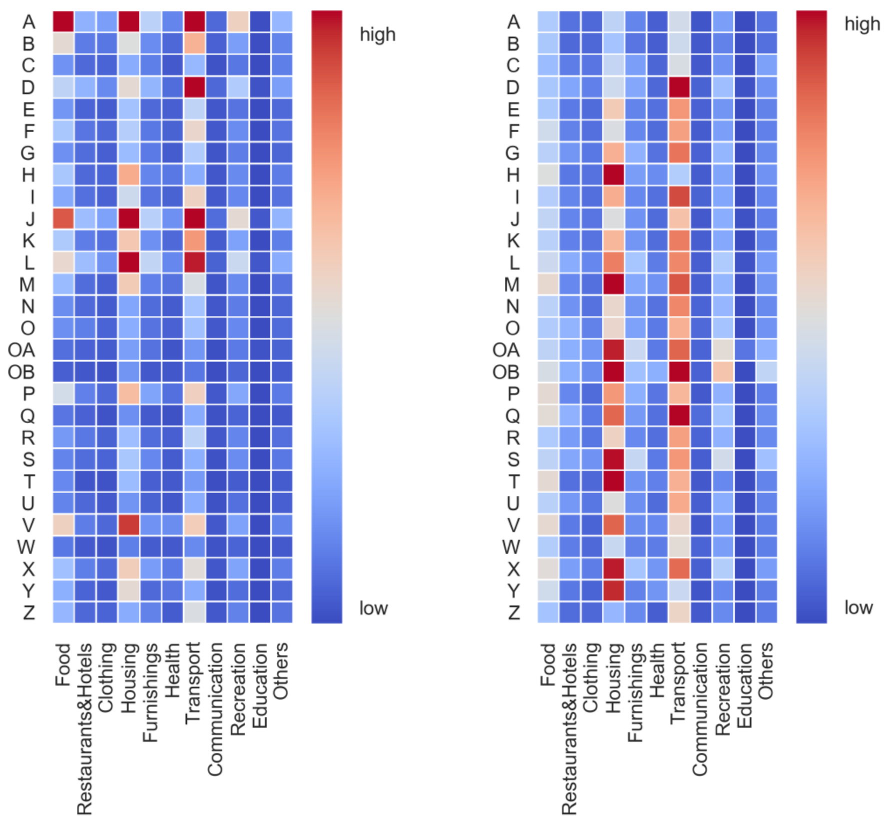


Figure D.47: Greenhouse gas emissions (IPCC 2013, 100a), prevalence-weighted results to compare archetypes and consumption categories at the same time. Left: total-view; Right: per-capita view.

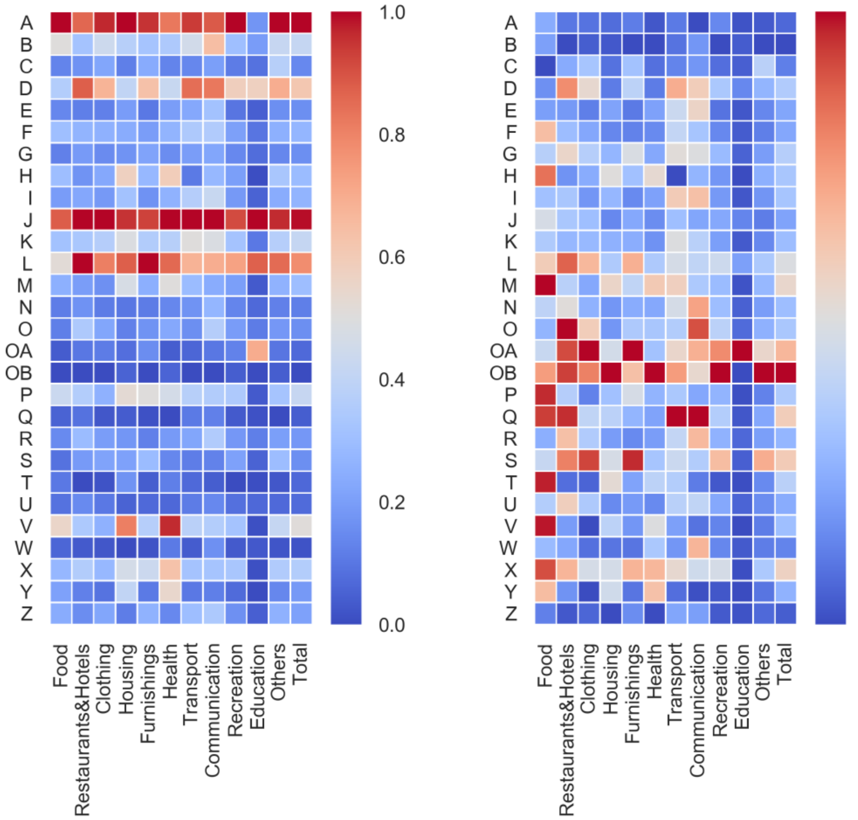


Figure D.48: **Greenhouse gas emissions (IPCC 2013, 100a), prevalence-weighted "per-category-view": Minimum-maximum-scaling along the consumption area axis to compare different archetypes within a consumption area. Left: total-view; Right: per-capita view.**

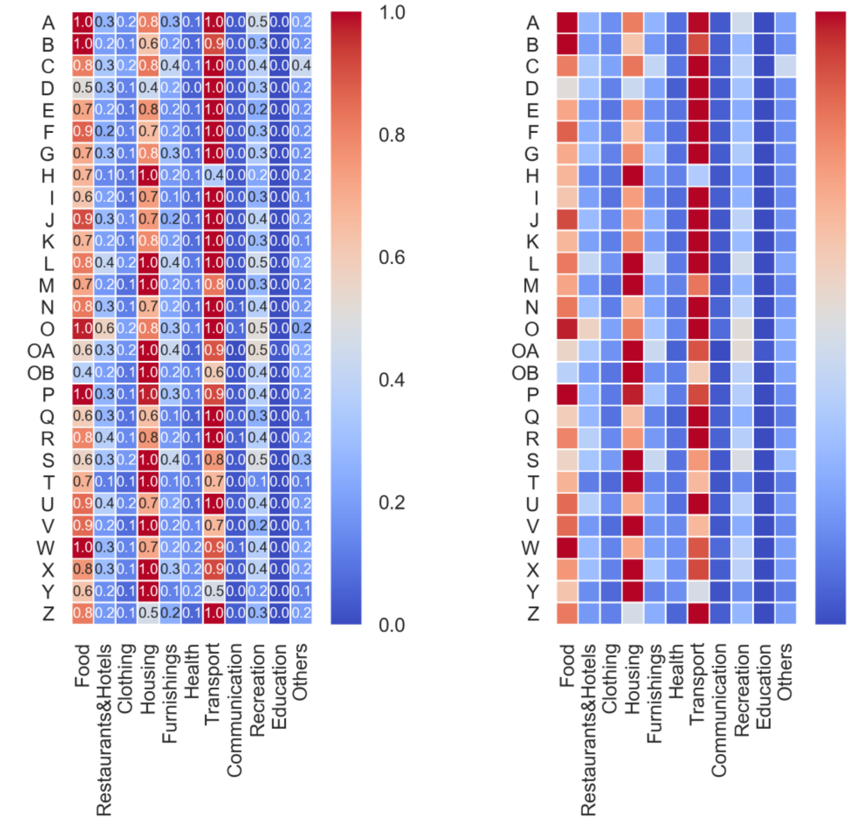


Figure D.49: **ReCiPe-endpoints (H, A), "per-cluster-view": Minimum-maximum-scaling along the archetype's axis to compare the consumption categories for each archetype separately. Left and right figures show the same, but the left-hand side also provides annotations of the applied scaling.**

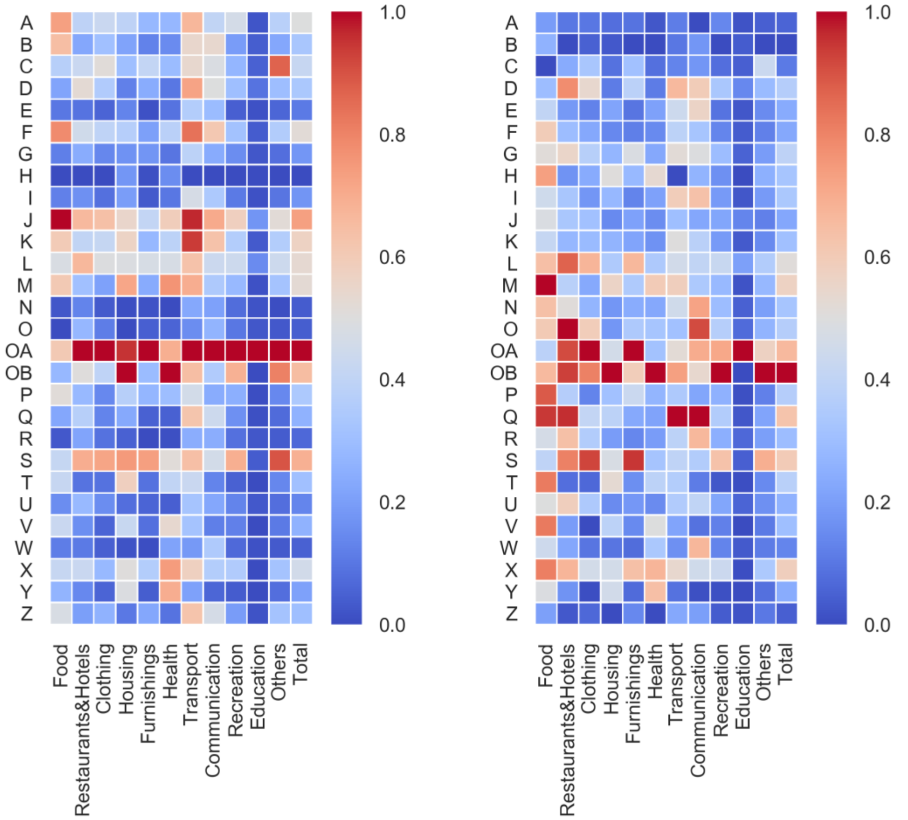


Figure D.50: **ReCiPe-endpoints (H,A), "per-category-view": Minimum-maximum-scaling along the consumption area axis to compare different archetypes within a consumption area. Left: total-view; Right: per-capita view.**

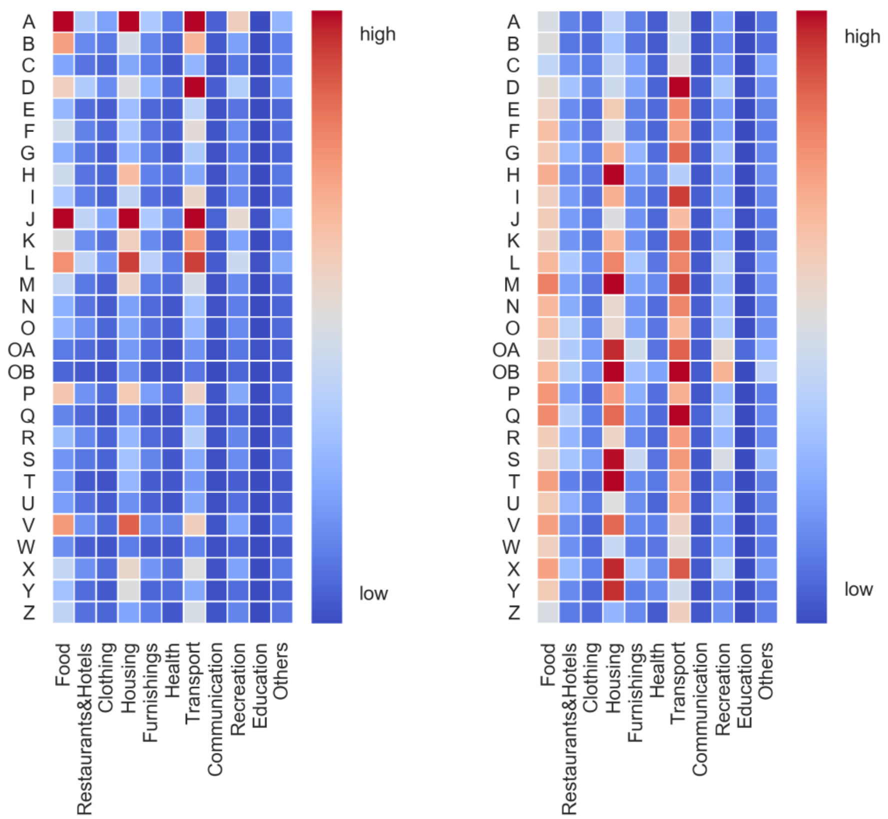


Figure D.51: **ReCiPe-endpoints (H,A), prevalence-weighted results to compare archetypes and consumption categories at the same time. Left: total-view; Right: per-capita view.**

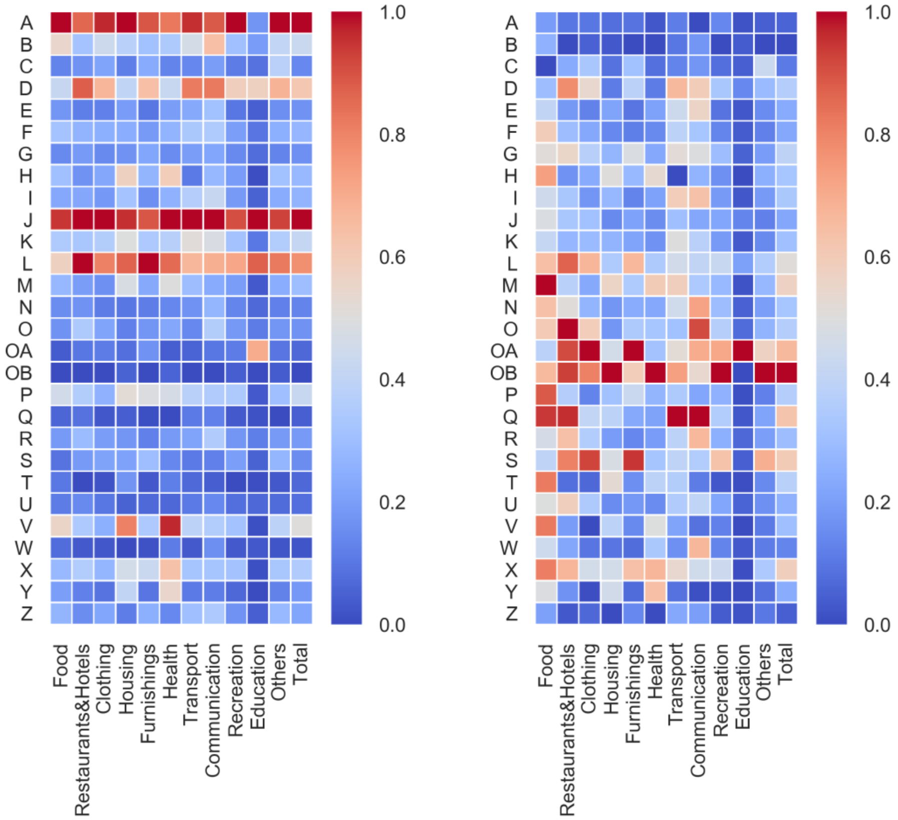


Figure D.52: **ReCiPe-endpoints (H,A), prevalence-weighted "per-category-view": Minimum-maximum-scaling along the consumption area axis to compare different archetypes within a consumption area. Left: total-view; Right: per-capita view.**

D.5.5 Correlation Analysis

As mentioned in Chapter 5, a correlation analysis was performed with the LCA-results of the archetypes' demands and their respective household characteristics. Please note that this analysis is just a rudimentary univariate correlation analysis and its only goal is to demonstrate that the archetype-approach can also be used for such investigations. This is also the reason for only considering correlations between different consumption areas and income, age and household size which are often cited as the most important household characteristics affecting the environmental footprints of households [82, 85]. The results of this analysis are depicted in Figures D.53-56. Note that the Spearman-Correlation-Coefficients as well as the linear regressions shown in these figures were computed on a prevalence-weighted basis. Figure D.53 and Figure D.55 show the results for greenhouse gas emissions (IPCC 2013, 100a), while Figure D.54 and Figure D.56 provide the findings of the correlation analysis of ReCiPe-endpoints (H,A). Furthermore, Figure D.53 and Figure D.54 are based on the total environmental impacts per archetype and Figure D.55 and Figure D.56 illustrate the correlation with per-capita footprints.

The Spearman-Correlation-Coefficients between total emissions and income (0.91 for both impact methods) as well as for household size (0.86) confirm the overall picture in Figure 5.3 in Chapter 5. Additionally, while a strong correlation can be found between income and recreation

(0.93), income correlates less strongly with food impacts (0.73 for ReCiPe and 0.67 for greenhouse gas emissions) and exhibits even a negative correlation with per-capita food impacts. This is also in line with the findings of Girod and De Haan [86] and with e.g. the footprints of OA and OB. Both archetypes show high impacts in the field of leisure activities (especially due to "package holidays", but also because of "major durable goods for recreation", and "pets services"), but impacts induced by food do not differ from other archetypes.

Total footprints often negatively correlate with age. This might be surprising since the analyses in Chapter 5 demonstrate that elder archetypes, particularly the "retired-couple"-cluster V, can be responsible for considerable contributions. However, from a per-capita perspective, the tendencies found in Chapter 5 for elder clusters can be confirmed. For instance, H ("old, widowed females"), V ("low-income, retired couple") and Y ("low-income, very old couple") show larger housing and health per-capita footprints than other archetypes. Corresponding to this, Figure D.55 and Figure D.56 reveal large Spearman-Correlation-Coefficients between age and housing (0.88 for both impact methods) and between age and health (0.93 for both impact methods).

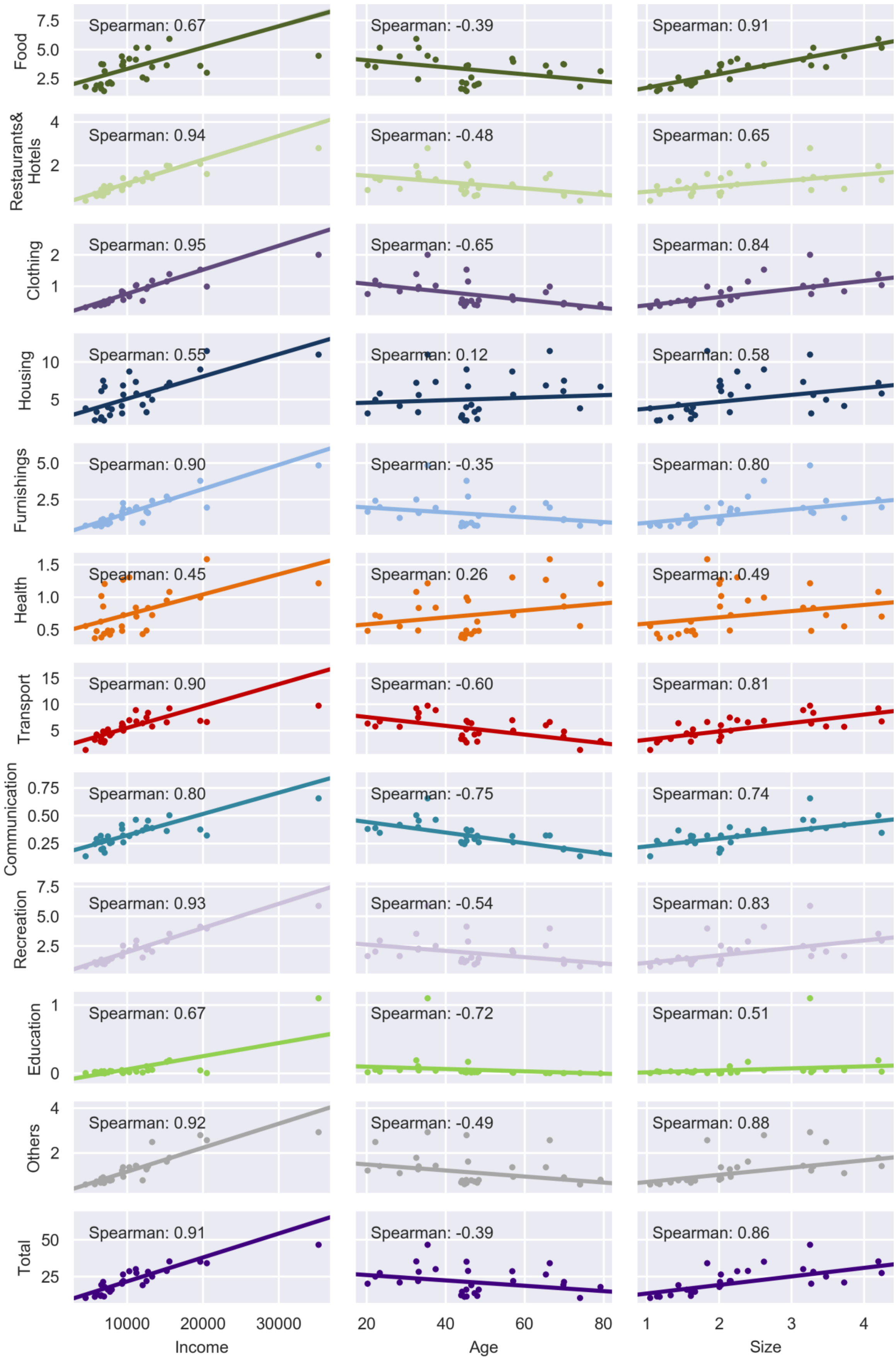


Figure D.53: Correlation analysis of prevalence-weighted total greenhouse gas emissions (IPCC 2013, 100a).

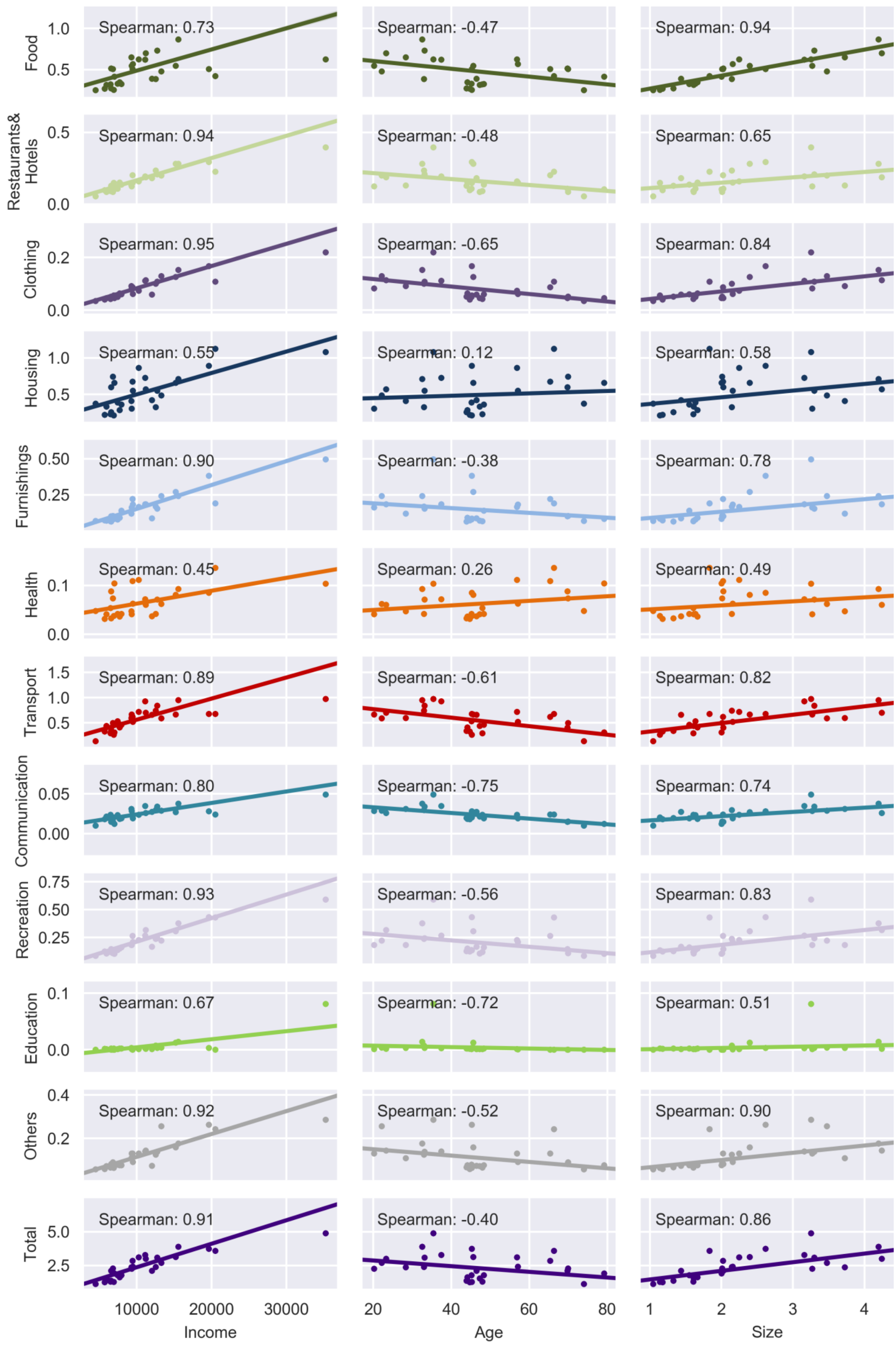


Figure D.54: Correlation analysis of prevalence-weighted total ReCiPe-endpoints (H,A).

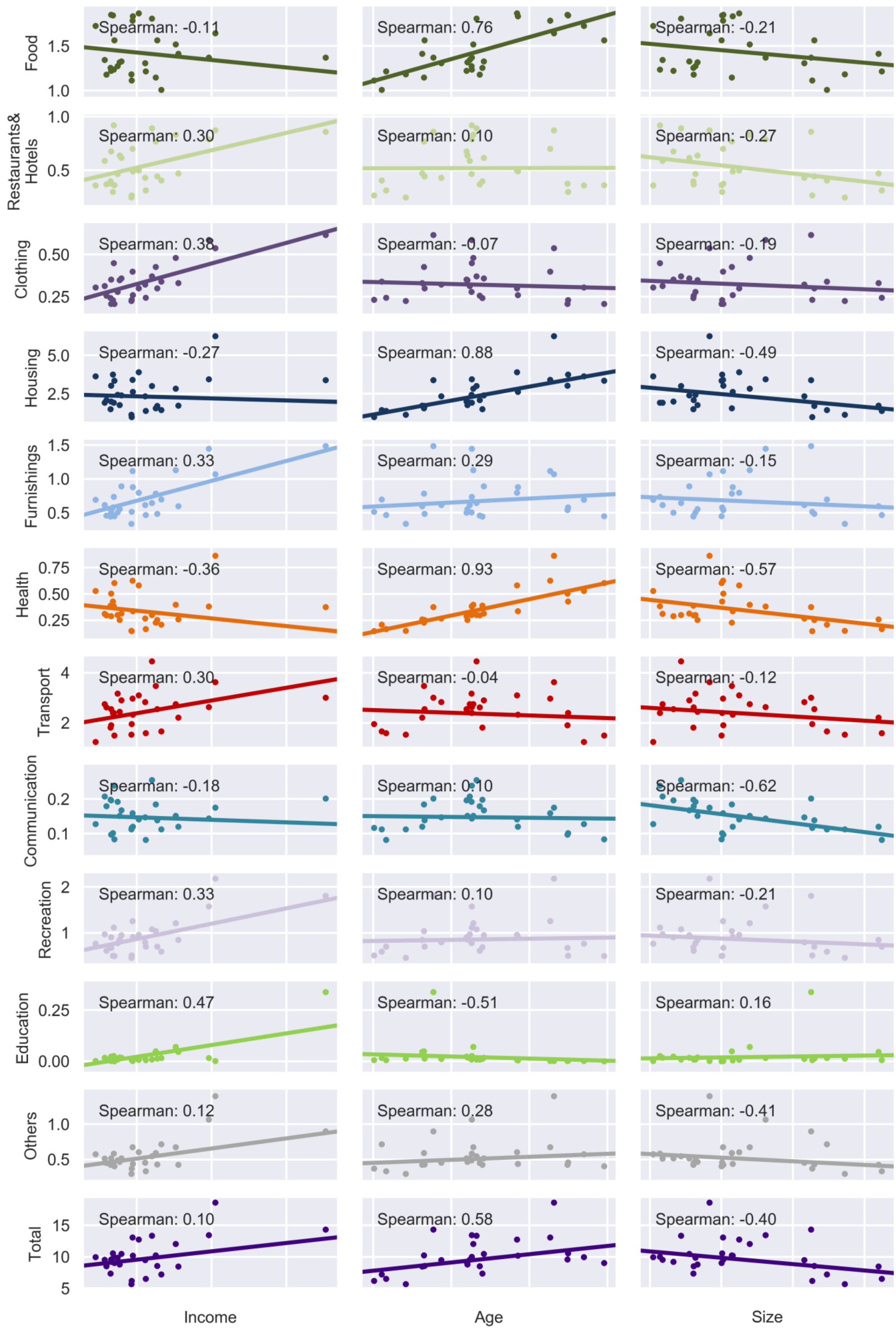


Figure D.55: **Correlation analysis of prevalence-weighted per-capita greenhouse gas emissions (IPCC 2013, 100a).**

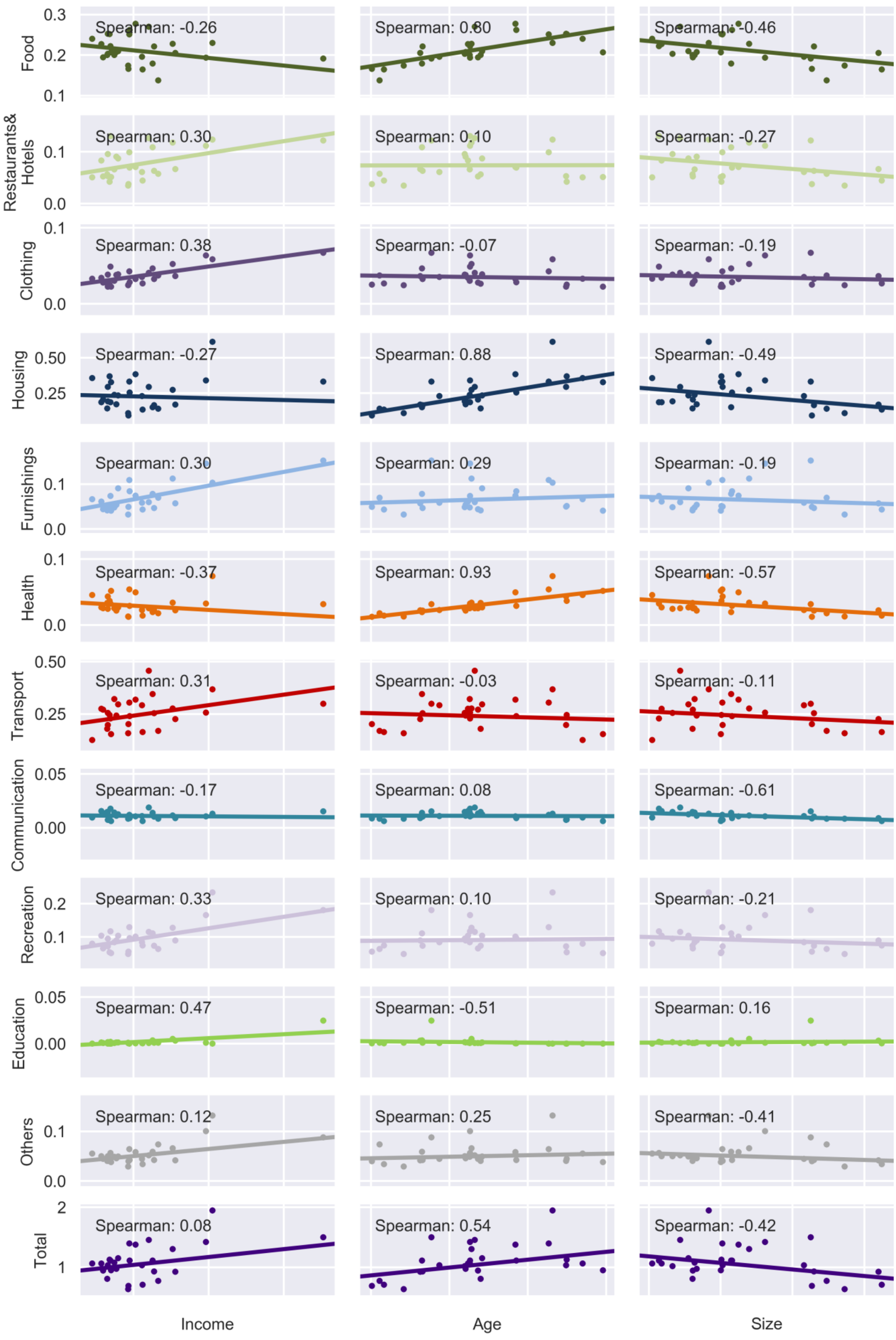


Figure D.56: **Correlation analysis of prevalence-weighted per-capita ReCiPe-endpoints (H,A).**

D.5.6 Results of Ecological Scarcity

Finally, we also present the results of the archetypes for the Ecological Scarcity (2013) [87] method. Since this is a Switzerland-tailored impact assessment method and our study is also focused on Switzerland, the following results might be interesting for Swiss policymakers. The overall, prevalence-weighted footprint is given in Figure D.57. Similar to the comparisons based on greenhouse gas emissions conducted in Chapter 5, these results again prove agreement with the study of Jungbluth et al. [88] although the prevalence-weighted per-capita footprint of 13.8 Mio. UBP (Umweltbelastungspunkte="environmental impact points") deviates by about 20% from the 17.3 Mio. UBP found in Jungbluth et al. [88]. Please note that the original value of Jungbluth et al. [88] (20 Mio. UBP) was again adjusted for better comparison. We also would like to point out that Jungbluth et al.'s study refers to another time period as the present study and that they used Ecological Scarcity 2006 [89] instead of 2013. Finally, also the footprint compositions on a total and a per-capita basis are provided in Figure D.58.

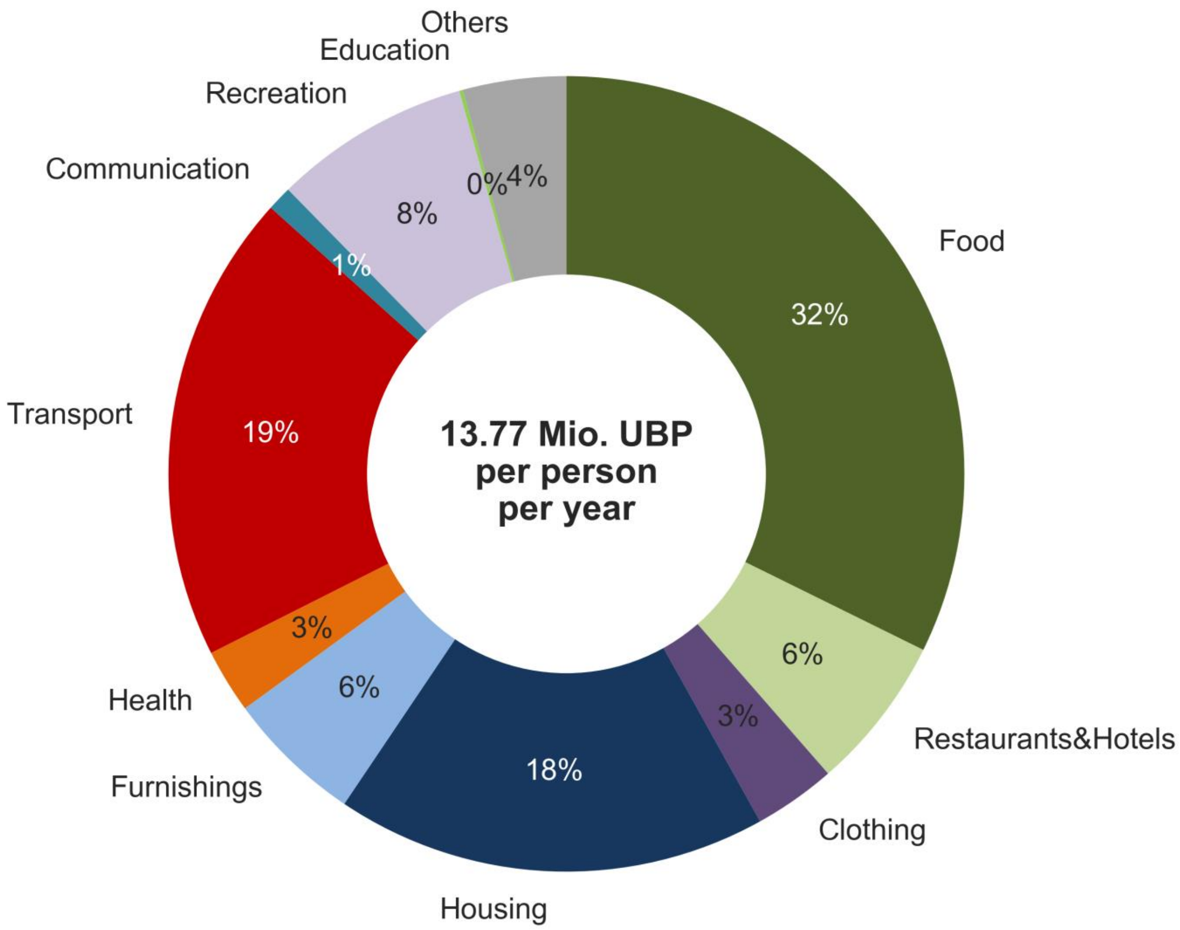


Figure D.57: Prevalence-weighted Swiss average footprint based on the Ecological Scarcity (2013) [87] **method.**

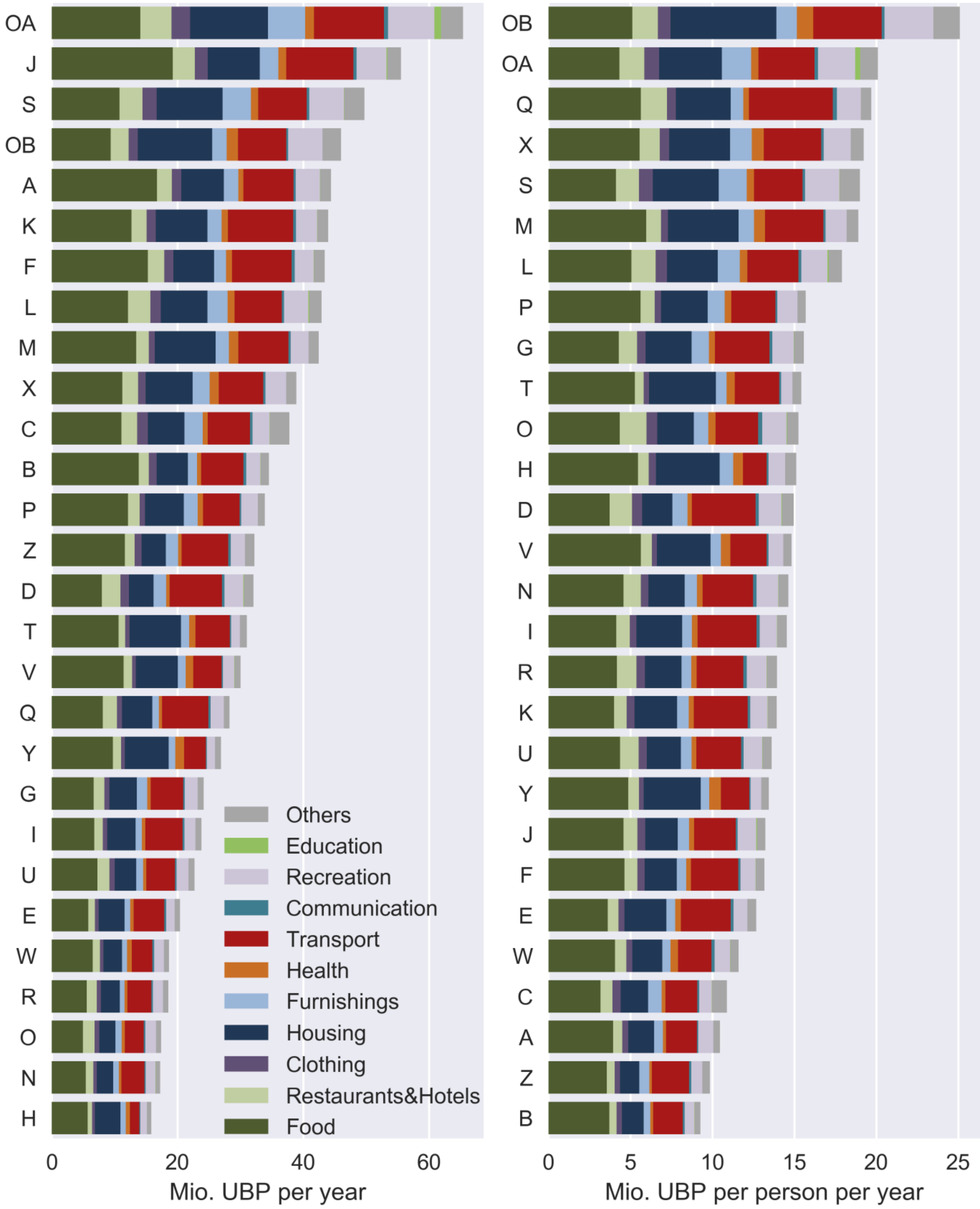


Figure D.58: Results of Ecological Scarcity (2013) [87]. The archetypes are ranked according to their footprints in both subfigures. Left: total footprint composition; Right: Per-capita footprints.

D.6 REFERENCES

[1] The PostgreSQL Global Development Group, *PostgreSQL*. [Online]. Available: https://www.postgresql.org. [Accessed: 19-Jan-2017].

[2] Python Software Foundation, *Python*, 2017. [Online]. Available: https://www.python.org/. [Accessed: 01-Apr-2017].

[3] C. Mutel, *Brightway: An open source framework for Life Cycle Assessment*, *Journal of Open Source Software*, vol. 12, p. 2, 2017.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2012.

[5] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, *The NumPy Array: A Structure for Efficient Numerical Computation*, *Computing in Science and Engineering*, vol. 13, pp. 22–30, 2011.

[6] E. Jones, T. Oliphant, P. Peterson, and et al., *SciPy: Open Source Scientific Tools for Python*. [Online]. Available: http://www.scipy.org/. [Accessed: 01-Apr-2017].

[7] W. McKinney, *Data Structures for Statistical Computing in Python*, in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 51–56.

[8] D. Varrazzo, *psycopg*, 2017. [Online]. Available: http://initd.org/psycopg/articles/. [Accessed: 01-Apr-2017].

[9] V. Moosavi, S. Packmann, and I. Vallés, *SOMPY*, 2017. [Online]. Available: https://github.com/sevamoo/SOMPY. [Accessed: 01-Apr-2017].

[10] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Computing in Science and Engineering*, vol. 9, pp. 90–95, 2007.

[11] M. Waskom, O. Botvinnik, D. O'Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, Brian, C. Fonnesbeck, A. Lee, and A. Qualieh, *Seaborn: Statistical Data Visualization*, 2017. [Online]. Available: https://seaborn.pydata.org/.

[12] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[13] United Nations Statistics Division, *COICOP (Classification of Individual Consumption According to Purpose)*, 2017. [Online]. Available: https://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5. [Accessed: 22-Nov-2017].

[14] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung 2011: Kommentierte Ergebnisse und Tabellen*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2013.

[15] Federal Department of Economic Affairs Education and Research, *Preisüberwachung*, 2017. [Online]. Available: http://www.preisvergleiche.preisueberwacher.admin.ch. [Accessed: 12-Apr-2017].

[16] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer Dordrecht, Heidelberg / New York / London, 2013.

[17] L. Breiman, *Random Forests*, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, *Journal of the Royal Statistical*

Society. Series B: Statistical (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[19] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, *Journal of Educational Psychology1*, vol. 24, no. 7, pp. 498–520, 1933.

[20] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[21] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, CRC Press, 1984.

[22] L. Breiman and A. Cutler, *Manual on Setting Up, Using, and Understanding Random Forests V4.0*, 2003.

[23] A. Liaw and M. Wiener, *Classification and Regression by randomForest*, *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[24] R. Genuer, J.-M. Poggi, and C. Tuleau, *Random Forests: some methodological insights*, Institut National de Recherche en Informatique et en Automatique: Le Chesnay Cedex, France, 2008.

[25] N. S. Altman, *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*, *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[26] G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, and B. Weidema, *The ecoinvent database version 3 (part I): overview and methodology*, *The International Journal of Life Cycle Assessment*, vol. 21, no. 9, pp. 1218–1230, 2016.

[27] Federal Electricity Commission (ElCom), *Strompreis-Webseite*, 2017. [Online]. Available: https://www.strompreis.elcom.admin.ch. [Accessed: 20-Jan-2017].

[28] Federal Statistical Office (BFS), *Landesindex der Konsumentenpreise. Durchschnittspreise Energie*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2017.

[29] B. P. Weidema, C. Bauer, R. Hischier, C. Mutel, T. Nemecek, J. Reinhard, C. O. Vadenbo, and G. Wernet, *Overview and methodology - Data quality guideline for the ecoinvent database version 3*, ecoinvent report No. 1(v3); ecoinvent: St. Gallen, Switzerland, 2013.

[30] Bundesamt für Energie (BFE), *Schweizerische Gesamtenergiestatistik 2015*, Bundesamt für Energie (BFE): Bern-Ittigen, Switzerland, 2016.

[31] Schweizerischer Verein des Gas- und Wasserfaches (SVGW), *Trinkwasser*, 2017. [Online]. Available: http://wasserqualitaet.svgw.ch/index.php?id=874. [Accessed: 28-Nov-2017].

[32] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2009 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2010.

[33] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2010 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2011.

[34] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2011 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2012.

[35] Bundesamt für Umwelt (BAFU), *Erhebung der Kehrichtzusammensetzung 2012*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2014.

[36] Bundesamt für Statistik (BFS) and Bundesamt für Raumentwicklung (ARE), *Mobilität in der Schweiz. Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2012.

[37] R. A. Fisher, *Statistical methods for research workers*, 14th ed., New York, NY, USA: Hafner, 1973.

[38] W. H. Kruskal and W. A. Wallis, *Use of ranks in one-criterion variance analysis*, *J. Am. Stat.*

Assoc., vol. 47, no. 260, pp. 583–621, 1952.

[39] T. Kohonen, *The Self-Organizing Map*, *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[40] T. K. Kohonen, *Self-Organizing Maps*, vol. 30, Springer Berlin Heidelberg, 2001.

[41] T. Kohonen, *Exploration of Very Large Databases by Self-Organizing Maps*, *Proceedings of IEEE International Conference on Neural Networks. v1*, pp. 1–6, 1997.

[42] S. M. Guthikonda, *Kohonen Self-Organizing Maps*, Wittenberg University, 2005.

[43] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, *SOM Toolbox for Matlab 5*, Technical Report A57; Helsinki University of Technology: Helsinki, Finland, 2000.

[44] J. Vesanto, *Neural network tool for data mining: SOM toolbox*, in *Proceedings of symposium on tool environments and development methods for intelligent systems (TOOLMET 2000)*, 2000, pp. 184–196.

[45] H. Sen Tan and S. E. George, *Investigating learning parameters in a standard 2-D SOM model to select good maps and avoid poor ones*, in *Australasian Joint Conference on Artificial Intelligence*, 2004, pp. 425–437.

[46] J. Vesanto and E. Alhoniemi, *Clustering of the Self-Organizing Map*, *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.

[47] J. MacQueen, *Some methods for classification and analysis of multivariate observations*, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281–297, 1967.

[48] S. Lloyd, *Least squares quantization in PCM*, *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[49] D. Arthur and S. Vassilvitskii, *K-Means++: the Advantages of Careful Seeding*, in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1035.

[50] M. R. Anderberg, *Cluster analysis for applications*, New York, NY, USA: Academic Press, 1973.

[51] J. A. Hartigan, *Clustering algorithms*, no. Wiley series in probability and mathematical statistics. Applied probability and statistics, New York, NY, USA: Wiley, 1975.

[52] J. H. Ward, *Hierarchical Grouping to Optimize an Objective Function*, *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[53] P. J. Rousseeuw, *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*, *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[54] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, *Understanding and enhancement of internal clustering validation measures*, *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 982–994, 2013.

[55] T. Calinski and J. Harabasz, *A dendrite method for cluster analysis*, *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974.

[56] J. Vesanto and M. Sulkava, *Distance matrix based clustering of the Self-Organizing Map*, in *Artificial Neural Networks - ICANN 2002*, 2002.

[57] J. Lötsch and A. Ultsch, *Exploiting the Structures of the U-Matrix*, in *Advances in Self-Organizing Maps and Learning Vector Quantization*, Cham: Springer, 2014, pp. 249–257.

[58] A. Ultsch, *U*-Matrix: a Tool to visualize Clusters in high dimensional Data*, University of Marburg: Marburg, Germany, 2003.

[59] P. Koch and T. Salou, *AGRIBALYSE: Methodology Report - Version 1.2*, ADEME: Angers,

France, 2015.

[60] A. Tukker, A. de Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, J. M. Rueda Cantuche, M. Bouwmeester, J. Oosterhaven, T. Drosdowski, and J. Kuenen, *Exiopol - Development and Illustrative Analyses of a Detailed Global MR EE SUT/IOT*, *Economic Systems Research*, vol. 25, no. 1, pp. 50–70, 2013.

[61] R. Wood, K. Stadler, T. Bulavskaya, S. Lutter, S. Giljum, A. de Koning, J. Kuenen, H. Schütz, J. Acosta-Fernández, A. Usubiaga, M. Simas, O. Ivanova, J. Weinzettel, J. H. Schmidt, S. Merciai, and A. Tukker, *Global Sustainability Accounting-Developing EXIOBASE for Multi-Regional Footprint Analysis*, *Sustainability (Switzerland)*, vol. 7, no. 1, pp. 138–163, 2015.

[62] L. Scherer and S. Pfister, *Global Biodiversity Loss by Freshwater Consumption and Eutrophication from Swiss Food Consumption*, *Environ. Sci. Technol.*, vol. 50, no. 13, pp. 7019–7028, 2016.

[63] C. Walker, E. R. Gibney, and S. Hellweg, *Comparison of Environmental Impact and Nutritional Quality among a European Sample Population – findings from the Food4me study*, *Scientific Reports*, vol. 8, no. 1, 2018.

[64] Migros Genossenschafts Bund (CH), *Migros*, 2017. [Online]. Available: https://www.migros.ch/de.html. [Accessed: 17-Mar-2017].

[65] Coop Genossenschaft, *Coop*, 2017. [Online]. Available: http://www.coop.ch/de.html. [Accessed: 17-Mar-2017].

[66] Steiner-Beck AG, *Steiner*, 2017. [Online]. Available: https://www.steiner-beck.ch/. [Accessed: 17-Mar-2017].

[67] GalloSuisse - Vereinigung der Schweizer Eierproduzenten, *Das Schweizer Ei*, 2017. [Online]. Available: http://www.gallosuisse.ch/. [Accessed: 17-Mar-2017].

[68] Federal Statistical Office (BFS), *Landesindex der Konsumentenpreise. Durchschnittspreise ausgewählter Produkte*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2017.

[69] Forschungsstelle für Sozial- und Wirtschaftsgeschichte (UZH), *Historical Statistics of Switzerland Online*, 2017. [Online]. Available: http://www.fsw.uzh.ch/hstat/nls_rev/ls_files.php?chapter_var=./h. [Accessed: 17-Mar-2017].

[70] Federal Statistical Office (BFS), *Swiss Consumer Price Index. LIK66, LIK77, 2015*, Neuchâtel, Switzerland: Federal Statistical Office (BFS), 2017.

[71] Federal Statistical Office (BFS), *STAT-TAB: Die interaktive Statistikdatenbank*, Bundesamt für Statistik (BFS): Neuchâtel, Switzerland, 2016.

[72] Bundesamt für Statistik (BFS) and Bundesamt für Verkehr (BAV), *Alpen- und grenzquerender Personenverkehr (A+GQPV), Statistik des öffentlichen Verkehrs (OeV). Kennzahlen RPV*, Bundesamt für Statistik (BFS), Bundesamt für Verkehr (BAV): Neuchâtel, Switzerland, 2016.

[73] ecoinvent Center, *Ecoinvent data, version 2.2*, http://www.ecoinvent.org/database/, 1998-2013, 2013.

[74] K. Steen-Olsen, R. Wood, and E. G. Hertwich, *The Carbon Footprint of Norwegian Household Consumption 1999-2012*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[75] Oanda Corporation, *Foreign Exchange Rates*, 2017. [Online]. Available: https://www.oanda.com/. [Accessed: 01-Dec-2017].

[76] United Nations Statistics Division, *ISIC (International Standard Industrial Classification of All Economic Activities) rev. 3.1*, 2017. [Online]. Available:

https://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17; [Accessed: 22-Nov-2017].

[77] A. W. Sleeswijk, L. F. C. M. van Oers, J. B. Guinée, J. Struijs, and M. A. J. Huijbregts, *Normalisation in product life cycle assessment: An LCA of the global and European economic systems in the year 2000*, *Science of the Total Environment*, vol. 390, no. 1, pp. 227–240, 2008.

[78] D. Ivanova, K. Stadler, K. Steen-Olsen, R. Wood, G. Vita, A. Tukker, and E. G. Hertwich, *Environmental Impact Assessment of Household Consumption*, *Journal of Industrial Ecology*, vol. 20, no. 3, 2016.

[79] E. G. Hertwich and G. P. Peters, *Carbon Footprint of Nations: A Global, Trade-Linked Analysis*, *Environ. Sci. Technol.*, vol. 43, no. 16, pp. 6414–6420, 2009.

[80] M. Di Donato, P. L. Lomas, and Ó. Carpintero, *Metabolism and Environmental Impacts of Household Consumption: A Review on the Assessment, Methodology, and Drivers*, *Journal of Industrial Ecology*, vol. 19, no. 5, pp. 904–916, 2015.

[81] A. Tukker and B. Jansen, *Environmental Impacts of Products: A Detailed Review of Studies*, *Journal of Industrial Ecology*, vol. 10, no. 3, pp. 159–182, 2006.

[82] E. G. Hertwich, *Life Cycle Approaches to Sustainable Consumption: A Critical Review*, *Environ. Sci. Technol.*, vol. 39, no. 13, pp. 4673–4684, 2005.

[83] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013.

[84] M. Goedkoop, R. Heijungs, M. Huijbregts, A. De Schryver, J. Struijs, and R. van Zelm, *ReCiPe 2008*, Ministrie van VROM: Den Haag, Netherlands, 2012.

[85] A. Tukker, M. J. Cohen, K. Hubacek, and O. Mont, *The Impacts of Household Consumption and Options for Change*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 13–30, 2010.

[86] B. Girod and P. De Haan, *More or Better? A Model for Changes in Household Greenhouse Gas Emissions due to Higher Income*, *Journal of Industrial Ecology*, vol. 14, no. 1, pp. 31–49, 2010.

[87] R. Frischknecht and S. Büsser Knöpfel, *Swiss Eco-Factors 2013 according to the Ecological Scarcity Method. Methodological fundamentals and their application in Switzerland*, Environmental studies no. 1330; Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2013.

[88] N. Jungbluth, C. Nathani, M. Stucki, and M. Leuenberger, *Environmental Impacts of Swiss Consumption and Production. A Combination of Input-Output Analysis with Life Cycle Assessment*, Federal Office for the Environment (FOEN): Bern, Switzerland, 2011.

[89] R. Frischknecht, R. Steiner, and N. Jungbluth, *The Ecological Scarcity Method – Eco-Factors 2006. A method for impact assessment in LCA*, Environmental Studies No. 0906; Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2009.

Appendix E

Outlook: Model Evaluation

E.1 OUTLOOK: MODEL EVALUATION

This appendix aims at giving an impression of the plausibility of the model results by presenting a selection of comparisons of national statistics with the results of the mobility sub-model, the consumption sub-model as well as the overall model (see Figure 1.1 in Chapter 1).

E.1.1 Mobility Sub-Model

Table E.1 compares the mean daily distance (kilometers per person per day) according to the model in Chapter 6 with the Swiss Mobility and Transport Microcensus 2010 [1] in different aggregations, whereas Table E.2 shows a similar comparison but with regard to modal splits. Note that in respect of cities, Geneva, Basel and Lugano were not considered because of their proximity to the national borders. Since routine cross-border trips were not simulated in the used MATSim-applications [2], we anticipated that the comparisons with these cities might not yield reasonable results. Furthermore, we would like to mention that the modal split relate only car trips and public transportation trips to each other and do not take bike and walk trips into account. Note that the modal splits are computed based on kilometers driven.

The used "PT-classes" in Tables E.1 and E.2 need some more explanation: The Federal Office for Spatial Development subdivides the whole area of Switzerland into five classes according to the goodness of public transport (PT) access. While a PT-class A indicates a very good public transport service, a not-classified area means no or very infrequent public transport access [3].

Table E.1: Comparison of the mean daily distance of the mobility sub-model results (Chapter 6) with the Swiss Mobility and Transport Microcensus 2010 [1]. [rel. Diff. = relative difference; abs. Diff. = absolute difference; PT = Public transport]

	Mean daily distance (km/pers/day)	Microcensus	Model	rel. Diff.	abs. Diff.
Total	Total	36.7	37.1	0.011	0.4
Spatial Structures	Agglomeration (cores and isolated cities)	32.1	30.9	-0.037	-1.2
	Agglomeration (other municipalities)	37.0	37.2	0.005	0.2
	Rural municipalities	41.3	43.8	0.061	2.5
Household Types	1-Person HH	33.2	33.9	0.021	0.7
	2-Persons HH	37.2	36.3	-0.024	-0.9
	3-Persons HH	38.7	38.1	-0.016	-0.6
	4-Persons HH	36.6	38.5	0.052	1.9
	5+-Persons HH	37.0	38.8	0.049	1.8
Cities	Winterthur	37.8	39.4	0.042	1.6
	Zurich	35.6	35.2	-0.011	-0.4
	Bern	38.9	30.2	-0.224	-8.7
	Lucerne	37.2	33.2	-0.108	-4.0
	Lausanne	32.7	34.2	0.046	1.5
PT-classes	PT-class A	29.3	27.6	-0.058	-1.7
	PT-class B	31.3	32.2	0.029	0.9
	PT-class C	37.1	37.1	0.000	0.0
	PT-class D	39.1	41.3	0.056	2.2
	not classified	40.5	45.1	0.114	4.6

Table E.2: Comparison of the modal splits in the mobility sub-model (Chapter 6) with the Swiss Mobility and Transport Microcensus 2010 [1]. [rel. Diff. = relative difference; abs. Diff. = absolute difference]

	Modal splits (%)	Microcensus	Model	rel. Diff.	abs. Diff.
Total	Total: Car	73.9	72.4	-0.021	-1.6
	Total: PT	26.1	27.6	0.060	1.6
Household Types	1-Person HH: Car	67.0	64.0	-0.045	-3.0
	1-Person HH: PT	33.0	36.0	0.091	3.0
	2-Persons HH: Car	75.2	70.6	-0.062	-4.6
	2-Persons HH: PT	24.8	29.4	0.187	4.6
	3-Persons HH: Car	75.0	71.7	-0.045	-3.3
	3-Persons HH: PT	25.0	28.3	0.134	3.3
	4-Persons HH: Car	75.8	72.8	-0.039	-3.0
	4-Persons HH: PT	24.2	27.2	0.124	3.0
	5+-Persons HH: Car	69.8	72.5	0.038	2.7
	5+-Persons HH: PT	30.2	27.5	-0.088	-2.7
Cities	Winterthur: Car	56.0	67.0	0.198	11.1
	Winterthur: PT	44.0	33.0	-0.251	-11.1
	Zurich: Car	65.2	67.0	0.028	1.8
	Zurich: PT	34.8	33.0	-0.053	-1.8
	Bern: Car	64.9	69.9	0.077	5.0
	Bern: PT	35.1	30.1	-0.143	-5.0
	Lucerne: Car	67.7	78.3	0.157	10.6
	Lucerne: PT	32.3	21.7	-0.329	-10.6
	Lausanne: Car	79.1	74.9	-0.053	-4.2
	Lausanne: PT	20.9	25.1	0.201	4.2

E.1.2 Consumption Sub-Model

Tables E.3-9 compare a selection of modeled revenues and expenditures with the original data from the Swiss Household Budget Survey (HBS) [4]. Deviations of more than 10% are indicated in red.

Table E.3: Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, all households are included. [rel. Diff. = relative difference; abs. Diff. = absolute difference]

(Swiss Francs per month)	All households			
	HBS	Model	rel. Diff.	abs. Diff.
Earned income	7227	7498	0.04	270
Primary income	7600	7908	0.04	307
Pensions and social benefits	1805	1762	-0.02	-43
Gross income	9530	9797	0.03	267
Compulsory transfer expenditure	2600	2720	0.05	120
Disposable income	6741	6888	0.02	147
Other insurances, fees and transfers	568	579	0.02	11
Total expenditures	5417	5564	0.03	147
Food and non-alcoholic beverages	654	663	0.01	9
Alcoholic beverages and tobacco	108	109	0.02	2
Restaurants and hotels	543	557	0.03	14
Clothing and footwear	234	240	0.03	6
Housing, water, electricity, gas and other fuels	1489	1542	0.04	52
Furnishings/household equip. and routine maintenance	277	286	0.03	9
Health	266	269	0.01	4
Transport	750	768	0.02	19
Communication	178	180	0.01	3
Recreation and culture	624	640	0.03	16
Miscellaneous goods and services	295	309	0.05	14

Table E.4: Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, the households are differentiated according to major regions (will be continued in Table E.5). [rel. Diff. = relative difference; abs. Diff. = absolute difference]

(Swiss Francs per month)	Major regions											
	Lake Geneva				Espace Mittelland				Northwestern Switzerland			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	7162	7308	0.02	146	6760	7348	0.09	588	7177	7664	0.07	487
Primary income	7557	7714	0.02	157	7045	7756	***0.10***	711	7511	8088	0.08	577
Pensions and social benefits	1975	1797	-0.09	-178	1832	1835	0.00	4	1917	1766	-0.08	-151
Gross income	9632	9627	0.00	-5	8992	9722	0.08	730	9588	9985	0.04	397
Compulsory transfer expenditure	2844	2748	-0.03	-96	2548	2694	0.06	145	2610	2771	0.06	161
Disposable income	6629	6712	0.01	83	6208	6831	***0.10***	623	6793	6993	0.03	200
Other insurances, fees and transfers	594	577	-0.03	-17	571	581	0.02	10	536	584	0.09	48
Total expenditures	5451	5479	0.01	28	5132	5533	0.08	401	5348	5617	0.05	269
Food and non-alcoholic beverages	701	664	-0.05	-37	668	670	0.00	1	625	664	0.06	38
Alcoholic beverages and tobacco	117	110	-0.06	-7	102	109	0.07	7	106	108	0.02	2
Restaurants and hotels	501	518	0.03	17	508	551	0.08	43	533	566	0.06	33
Clothing and footwear	215	227	0.06	13	217	237	0.09	21	240	244	0.02	5
Housing, water, electricity, gas and other fuels	1538	1554	0.01	16	1361	1520	***0.12***	158	1471	1541	0.05	70
Furnishings/household equip. and routine maintenance	250	266	0.06	16	273	290	0.06	17	306	300	-0.02	-6
Health	291	277	-0.05	-14	261	273	0.04	12	248	267	0.08	19
Transport	735	760	0.03	25	736	755	0.03	19	730	786	0.08	56
Communication	199	189	-0.05	-10	171	175	0.02	4	166	177	0.06	11
Recreation and culture	626	616	-0.02	-10	583	645	***0.11***	62	626	650	0.04	24
Miscellaneous goods and services	278	297	0.07	19	250	308	***0.23***	57	296	315	0.06	18

Table E.5: **Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data** [4]**. Here, the households are differentiated according to major regions (continued from Table E.4). [rel. Diff. = relative difference; abs. Diff. = absolute difference]**

(Swiss Francs per month)	Major regions															
	Zurich				Eastern Switzerland				Central Switzerland				Ticino			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	7953	7917	0.00	-36	7132	7527	0.06	395	7640	7691	0.01	51	6240	6331	0.01	91
Primary income	8407	8337	-0.01	-70	7471	7935	0.06	465	8095	8099	0.00	3	6638	6678	0.01	40
Pensions and social benefits	1860	1675	-0.10	-185	1485	1692	***0.14***	206	1571	1754	***0.12***	183	1813	1828	0.01	15
Gross income	10380	10132	-0.02	-248	9086	9764	0.07	678	9814	9983	0.02	170	8594	8635	0.00	41
Compulsory transfer expenditure	2732	2790	0.02	58	2409	2698	***0.12***	289	2368	2742	***0.16***	374	2253	2313	0.03	60
Disposable income	7461	7144	-0.04	-316	6501	6884	0.06	383	7292	7071	-0.03	-220	6129	6171	0.01	42
Other insurances, fees and transfers	560	579	0.03	19	565	571	0.01	6	560	586	0.05	26	598	564	-0.06	-35
Total expenditures	5972	5736	-0.04	-236	5113	5533	0.08	420	5594	5690	0.02	96	5016	5046	0.01	30
Food and non-alcoholic beverages	621	646	0.04	26	644	664	0.03	21	659	689	0.05	30	636	627	-0.01	-9
Alcoholic beverages and tobacco	112	112	0.01	1	96	107	***0.11***	11	114	111	-0.03	-3	103	104	0.01	1
Restaurants and hotels	641	598	-0.07	-43	524	568	0.08	44	599	584	-0.03	-15	431	459	0.06	27
Clothing and footwear	262	251	-0.04	-11	234	242	0.03	8	248	249	0.00	1	229	219	-0.04	-10
Housing, water, electricity, gas and other fuels	1710	1598	-0.07	-112	1342	1519	***0.13***	177	1533	1557	0.02	24	1358	1416	0.04	58
Furnishings/household equip. and routine maintenance	296	299	0.01	3	266	283	0.07	18	294	292	-0.01	-2	248	252	0.01	3
Health	287	265	-0.08	-22	240	265	***0.10***	25	250	275	***0.10***	25	251	251	0.00	-1
Transport	782	794	0.02	12	724	754	0.04	29	810	781	-0.04	-29	752	727	-0.03	-25
Communication	179	182	0.02	3	165	177	0.07	12	169	178	0.05	9	201	188	-0.06	-13
Recreation and culture	693	662	-0.04	-30	609	646	0.06	37	625	660	0.06	35	555	540	-0.03	-15
Miscellaneous goods and services	389	328	***-0.16***	-62	269	308	***0.15***	39	293	314	0.07	21	251	265	0.05	13

Table E.6: **Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, the households are differentiated according to income classes (will be continued in Table E.7). [rel. Diff. = relative difference; abs. Diff. = absolute difference]**

(Swiss Francs per month)	Income classes (Swiss Francs per month)											
	< 4880				4880 - 7173				7174 - 9702			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	904	212	***-0.77***	-692	3540	3828	0.08	288	6147	6363	0.04	216
Primary income	1072	685	***-0.36***	-386	3767	4084	0.08	316	6367	6632	0.04	266
Pensions and social benefits	2270	3736	***0.65***	1466	2153	2346	0.09	192	1941	1764	-0.09	-177
Gross income	3475	4527	***0.30***	1052	6053	6524	0.08	471	8414	8607	0.02	193
Compulsory transfer expenditure	865	1024	***0.18***	159	1542	1669	0.08	127	2188	2267	0.04	79
Disposable income	2538	3212	***0.27***	674	4299	4752	***0.11***	453	6074	6225	0.02	151
Other insurances, fees and transfers	315	423	***0.34***	107	446	437	-0.02	-9	552	528	-0.04	-24
Total expenditures	3038	3067	0.01	28	4089	3994	-0.02	-94	5105	5147	0.01	43
Food and non-alcoholic beverages	444	410	-0.08	-34	530	471	***-0.11***	-59	638	635	0.00	-2
Alcoholic beverages and tobacco	68	52	***-0.24***	-16	81	91	***0.12***	10	100	101	0.01	1
Restaurants and hotels	226	191	***-0.16***	-35	358	393	0.10	35	495	475	-0.04	-20
Clothing and footwear	96	106	0.10	9	150	145	-0.03	-5	197	212	0.08	15
Housing, water, electricity, gas and other fuels	1061	1156	0.09	95	1266	1280	0.01	14	1448	1519	0.05	71
Furnishings/household equip. and routine maintenance	120	150	***0.25***	30	185	159	***-0.14***	-26	253	262	0.04	9
Health	194	213	0.10	19	234	246	0.05	11	266	247	-0.07	-19
Transport	298	232	***-0.22***	-66	528	470	***-0.11***	-58	701	728	0.04	28
Communication	109	78	***-0.29***	-31	152	141	-0.07	-10	181	185	0.02	4
Recreation and culture	294	318	0.08	24	418	426	0.02	8	574	559	-0.03	-15
Miscellaneous goods and services	127	161	***0.26***	33	186	172	-0.08	-14	253	224	***-0.11***	-28

Table E.7: Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, the households are differentiated according to income classes (continued from Table E.6). [rel. Diff. = relative difference; abs. Diff. = absolute difference]

(Swiss Francs per month)	Income classes (Swiss Francs per month)							
	9703 - 13170				> 13171			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	9443	10615	*0.12*	1172	16099	14716	-0.09	-1383
Primary income	9718	10840	*0.12*	1122	17074	15755	-0.08	-1319
Pensions and social benefits	1441	935	*-0.35*	-507	1221	905	*-0.26*	-316
Gross income	11255	11854	0.05	599	18448	16772	-0.09	-1675
Compulsory transfer expenditure	2957	3111	0.05	154	5447	5357	-0.02	-91
Disposable income	8108	8351	0.03	243	12682	11242	*-0.11*	-1441
Other insurances, fees and transfers	630	646	0.03	16	897	874	-0.03	-23
Total expenditures	6246	6521	0.04	274	8604	8703	0.01	99
Food and non-alcoholic beverages	768	778	0.01	10	891	992	*0.11*	101
Alcoholic beverages and tobacco	129	127	-0.01	-2	159	155	-0.02	-4
Restaurants and hotels	659	704	0.07	46	975	933	-0.04	-42
Clothing and footwear	290	295	0.02	5	436	430	-0.01	-6
Housing, water, electricity, gas and other fuels	1611	1728	0.07	117	2061	1955	-0.05	-105
Furnishings/household equip. and routine maintenance	307	322	0.05	15	522	544	0.04	22
Health	290	257	*-0.11*	-33	344	378	0.10	34
Transport	942	1004	0.07	63	1279	1274	0.00	-6
Communication	210	217	0.03	7	237	238	0.01	1
Recreation and culture	722	764	0.06	42	1113	1106	-0.01	-7
Miscellaneous goods and services	319	325	0.02	6	588	699	*0.19*	110

Table E.8: Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, the households are differentiated according to household types (will be continued in Table E.9). [rel. Diff. = relative difference; abs. Diff. = absolute difference]

(Swiss Francs per month)	All 1-person households				All couples with children			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	3899	5769	*0.48*	1870	10770	10111	-0.06	-659
Primary income	4254	6122	*0.44*	1868	11004	10479	-0.05	-525
Pensions and social benefits	1815	1660	-0.09	-155	780	1158	*0.49*	379
Gross income	6156	7896	*0.28*	1740	11890	11803	-0.01	-87
Compulsory transfer expenditure	1620	2151	*0.33*	531	3291	3290	0.00	0
Disposable income	4273	5519	*0.29*	1246	8510	8374	-0.02	-136
Other insurances, fees and transfers	419	473	*0.13*	54	622	648	0.04	26
Total expenditures	3696	4534	*0.23*	839	6868	6641	-0.03	-227
Food and non-alcoholic beverages	378	491	*0.30*	113	909	852	-0.06	-56
Alcoholic beverages and tobacco	75	89	*0.18*	14	101	118	*0.17*	17
Restaurants and hotels	365	447	*0.22*	82	654	670	0.02	16
Clothing and footwear	136	182	*0.34*	46	338	316	-0.07	-22
Housing, water, electricity, gas and other fuels	1249	1414	*0.13*	166	1772	1686	-0.05	-86
Furnishings/household equip. and routine maintenance	155	209	*0.34*	53	362	360	-0.01	-3
Health	181	210	*0.16*	29	279	293	0.05	14
Transport	458	600	*0.31*	142	968	944	-0.02	-24
Communication	124	156	*0.26*	32	226	216	-0.04	-10
Recreation and culture	400	502	*0.25*	102	815	790	-0.03	-25
Miscellaneous goods and services	174	234	*0.35*	60	444	395	*-0.11*	-49

Table E.9: Selection of modeled revenues and expenditures in the overall model (Chapter 6) compared with the original HBS-data [4]. Here, the households are differentiated according to household types (continued from Table E.8). [rel. Diff. = relative difference; abs. Diff. = absolute difference]

(Swiss Francs per month)	Household types											
	Couples with 1 child				Couples with 2 children				Couples with 3+ children			
	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.	HBS	Model	rel. Diff.	abs. Diff.
Earned income	10250	9436	-0.08	-815	11124	10282	-0.08	-842	11019	10534	-0.04	-486
Primary income	10448	9788	-0.06	-660	11359	10635	-0.06	-724	11337	10908	-0.04	-429
Pensions and social benefits	693	1167	***0.69***	475	752	1123	***0.49***	371	1057	1119	0.06	61
Gross income	11215	11106	-0.01	-109	12238	11936	-0.02	-302	12522	12210	-0.02	-312
Compulsory transfer expenditure	3213	3122	-0.03	-91	3323	3304	-0.01	-19	3383	3365	-0.01	-17
Disposable income	7893	7840	-0.01	-53	8823	8502	-0.04	-321	9099	8715	-0.04	-384
Other insurances, fees and transfers	577	602	0.04	25	660	654	-0.01	-6	622	669	0.07	46
Total expenditures	6440	6206	-0.04	-234	7131	6755	-0.05	-376	7153	6915	-0.03	-239
Food and non-alcoholic beverages	779	745	-0.04	-34	953	890	-0.07	-63	1089	923	***-0.15***	-167
Alcoholic beverages and tobacco	109	113	0.04	4	105	117	***0.12***	12	74	119	***0.60***	45
Restaurants and hotels	618	615	0.00	-3	706	680	-0.04	-26	599	702	***0.17***	104
Clothing and footwear	297	285	-0.04	-12	365	326	***-0.11***	-39	359	337	-0.06	-22
Housing, water, electricity, gas and other fuels	1738	1681	-0.03	-57	1810	1688	-0.07	-122	1748	1685	-0.04	-62
Furnishings/household equip. and routine maintenance	336	334	-0.01	-3	361	364	0.01	3	426	375	***-0.12***	-51
Health	249	262	0.05	13	298	299	0.00	1	296	310	0.05	14
Transport	1011	883	***-0.13***	-127	950	959	0.01	8	917	986	0.08	69
Communication	228	207	-0.09	-21	223	220	-0.01	-3	227	223	-0.02	-4
Recreation and culture	619	695	***0.12***	77	934	814	***-0.13***	-119	951	851	***-0.10***	-100
Miscellaneous goods and services	456	385	***-0.16***	-71	426	398	-0.07	-28	468	404	***-0.14***	-63

The comparisons in Tables E.3-9 show that the model is able to satisfactorily reproduce the statistics according to the HBS in different aggregations, be it grouped by geographical major regions, income classes or household types. Only in two tables, larger deviations can be observed: for the income class "< 4880 Swiss Francs per month" in Table E.6 and for "1-person households" in Table E.8. Even though these deviations are still in an acceptable range, they are also explainable. In fact, only one archetype shows an income lower than 4880 Swiss Francs per month, which is H – the "old, widowed females"-cluster. This means that only one specific archetype was considered in this income class and deviations were clearly to be expected. In other words: The subdivision of income classes according to the official HBS-statistics is not reasonable for comparing with our model. Though, this does not apply for single-person households. Here, the reasons for the deviations originate from the averaging procedure to derive the archetypes and the classification approach to assign the archetypes to households. Since there is no archetype which is a "pure" single-person household (average number of persons equals exactly 1), it is very likely that larger households in the same cluster tend to increase the average values of "almost" single-person household archetypes. Indeed, Table E.8 shows overestimations by trend.

E.1.3 Comparison with National Statistics

Table E.10 provides a similar evaluation of the model results as was already presented in Appendix D.2.4 for the archetypes. Final energy demand for heating, electricity as well as total energy demand correspond well with national statistics [5]. The modeled water consumption is higher than the current direct water consumption according to [6]. As already mentioned in D.2.4, the model results can still be regarded reasonable since the statistics itself is not a measured value but an estimation and does not specifically refer to a particular year. Also in respect of waste production, the same reasoning as presented in Appendix D.2.4 applies: The underestimation of the model can partly be explained by the fact that the waste statistics [7] takes not only household waste but also commercial waste into account. However, a rough sensitivity analysis revealed that the final life cycle assessment results are only negligibly affected even in the case of doubling the waste production.

Table E.10: Comparison of national statistics with model results. Final energy demand statistics originate from [5] and refer to the year 2013, while water consumption was retrieved from [6] and waste statistics from [7].

	Statistics	Model
Total final energy for heating (GJ/pers/yr)	23.5	19.6
Final energy for heating (only fuel oil) (GJ/pers/yr)	12.2	10.8
Final energy for electricity (kWh/pers/yr)	2306	2389
Water consumption (m^3/pers/yr)	52	69
Waste production (kg/pers/yr)	344	193

E.2 REFERENCES

[1] Bundesamt für Statistik (BFS) and Bundesamt für Raumentwicklung (ARE), *Mobilität in der Schweiz. Ergebnisse des Mikrozensus Mobilität und Verkehr 2010*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, 2012.

[2] S. Hirschberg (ed.), C. Bauer, B. Cox, T. Heck, J. Hofer, W. Schenler, A. Simons, A. Del Duce, H.-J. Althaus, G. Georges, T. Krause, M. Vayá González, F. Ciari, R. A. Waraich, B. Jäggi, A. Stahel, A. Froemelt, and D. Saner, *Opportunities and challenges for electric mobility: An interdisciplinary assessment of passenger vehicles*, Villigen, Dübendorf and Zurich, Switzerland, 2016.

[3] ARE (Bundesamt für Raumentwicklung), *ÖV-Güteklassen. Berechnungsmethodik ARE. Grundlagenbericht für die Beurteilung der Agglomerationsprogramme Verkehr und Siedlung*, 2017.

[4] Bundesamt für Statistik (BFS), *Haushaltsbudgeterhebung (HABE) 2009 bis 2011*, Neuchâtel, Switzerland: Bundesamt für Statistik (BFS), 2013.

[5] Bundesamt für Energie (BFE), *Schweizerische Gesamtenergiestatistik 2015*, Bundesamt für Energie (BFE): Bern-Ittigen, Switzerland, 2016.

[6] Schweizerischer Verein des Gas- und Wasserfaches (SVGW), *Trinkwasser*, 2017. [Online]. Available: http://wasserqualitaet.svgw.ch/index.php?id=874. [Accessed: 28-Nov-2017].

[7] Bundesamt für Umwelt (BAFU), *Abfallmengen und Recycling 2013 im Überblick*, Bundesamt für Umwelt (BAFU): Bern, Switzerland, 2014.

CURRICULUM VITAE

Name	Andreas Daniel Frömelt
Date of birth	20 April 1983
Citizen of	Stettfurt (TG)
Nationality	Switzerland

EDUCATION

2013 – present	Doctoral studies, ETH Zurich, Institute of Environmental Engineering, Chair of Ecological Systems Design, Switzerland
2014 – 2015	Certificate of Advanced Studies ETH in Spatial Information Systems
2007 – 2009	Master of Science in Environmental Engineering (major courses: Urban Water Management and Water Resources Management), ETH Zurich, Switzerland
2003 – 2006	Bachelor of Science in Environmental Engineering, ETH Zurich, Switzerland
1996 – 2002	Swiss Upper Secondary School, Kantonsschule Frauenfeld, Switzerland

PRACTICAL EXPERIENCE

Jan 2013 – present	Scientific Assistant, ETH Zurich, Institute of Environmental Engineering, Chair of Ecological Systems Design, Switzerland
Oct 2011 – Dec 2016	Scientific Assistant, ETH Zurich, Institute of Environmental Engineering, Laboratory for Environmental Engineering, Switzerland (Oct 2011 – Dec 2012: full-time position / Jan 2013 – Dec 2014: 20%-position as lecturer / Jan 2015 – Dec 2016: 10%-position as lecturer)
Jul 2010 – Dec 2013	Sub-Project Manager / Project Engineer, Pöyry Infra AG, Division of Water and Wastewater Treatment, Switzerland (Jul 2010 – Sep 2011: full-time position / Oct 2011 – Dec 2013: as hourly-paid employee)
Sep 2009 – Jun 2010	Project Engineer, CSD Ingenieurs, Division of Water and Wastewater Treatment (former Morgenthaler Ingenieure AG), Switzerland
Nov 2008 – Jan 2009	Internship, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department of Water and Sanitation in Developing Countries (Sandec), Dodoma, Tanzania

Apr 2007 – Jul 2007	Internship, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department of Water and Sanitation in Developing Countries (Sandec), Kathmandu, Nepal
Nov 2006 – Mar 2007	Internship, CSD Ingenieurs, Division of Water and Wastewater Treatment (former Morgenthaler Ingenieure AG), Switzerland

TEACHING EXPERIENCE

Jan 2017 – present	Lecturer for "Development and Assessment of Transport Scenarios", ETH Zurich, Master of Advanced Studies in Future Transport Systems and Certificate of Advanced Studies in Systemic Aspects of Future Transport, SCCER Mobility, Switzerland
Jan 2013 – Dec 2016	Lecturer for "Labor für Umweltingenieurwissenschaften I" and "Labor für Umweltingenieurwissenschaften II", ETH Zurich, Institute of Environmental Engineering, Laboratory for Environmental Engineering, Switzerland

LANGUAGES

German	Mother tongue
English	Fluent in spoken and written
French	Conversant
Russian	Basic knowledge

PUBLICATION LIST

ARTICLES INCLUDED IN THIS DISSERTATION

A. Froemelt, M. Mauchle, B. Steubing, and S. Hellweg, *Greenhouse Gas Emissions Quantification and Reduction Efforts in a Rural Municipality*, *Journal of Industrial Ecology*, vol. 22, no. 1, pp. 92–105, 2018.

A. Froemelt and S. Hellweg, *Assessing Space Heating Demand on a Regional Level: Evaluation of a Bottom-Up Model in the Scope of a Case Study*, *Journal of Industrial Ecology*, vol. 21, no. 2, pp. 332–343, 2017.

R. Buffat, **A. Froemelt**, N. Heeren, M. Raubal, and S. Hellweg, *Big data GIS analysis for novel approaches in building stock modelling*, *Applied Energy*, vol. 208, pp. 277–290, 2017.

A. Froemelt, D. J. Dürrenmatt, and S. Hellweg, *Using Data Mining To Assess Environmental Impacts of Household Consumption Behaviors*, *Environmental Science & Technology*, 2018.

A. Froemelt, R. Buffat, N. Heeren, and S. Hellweg, *Assessing environmental impacts of individual households: A large-scale bottom-up LCA-model for Switzerland*, in *SETAC Europe 28th Annual Meeting*, 13-17 May, Rome, Italy, 2018.

ADDITIONAL PUBLICATIONS

D. Bucher, R. Buffat, **A. Froemelt**, and M. Raubal, *Can the Use of Electric Bicycles Reduce Commuter Energy Demand? A Case Study in Switzerland*, in preparation.

S. Cucurachi, S. Schiess, **A. Froemelt**, and S. Hellweg, *Noise-footprint from Personal Land-Based Mobility*, accepted for publication in the *Journal of Industrial Ecology*.

R. Frischknecht, C. Bauer, **A. Froemelt**, S. Hellweg, K. Biemann, T. Buetler, B. Cox, P. De Haan, S. Hoerl, R. Itten, N. Jungbluth, Y. Ligen, N. A. Mathys, S. Schiess, S. Schori, P. van Loon, J. Wang, and S. Wettstein, *LCA of mobility solutions: approaches and findings — 66th LCA forum, Swiss Federal Institute of Technology, Zurich, 30 August, 2017*, *The International Journal of Life Cycle Assessment*, vol. 23, no. 2, pp. 381–386, 2018.

R. Buffat, L. Schmid, N. Heeren, **A. Froemelt**, M. Raubal, and S. Hellweg, *GIS-based Decision Support System for Building Retrofit*, *Energy Procedia*, vol. 122, pp. 403–408, 2017.

S. Hirschberg (ed.), C. Bauer, B. Cox, T. Heck, J. Hofer, W. Schenler, A. Simons, A. Del Duce, H.-J. Althaus, G. Georges, T. Krause, M. Vayá González, F. Ciari, R. A. Waraich, B. Jäggi, A. Stahel, **A. Froemelt**, and D. Saner, *Opportunities and challenges for electric mobility:*

An interdisciplinary assessment of passenger vehicles, Villigen, Dübendorf and Zurich, Switzerland, 2016.

M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, **A. Frömelt**, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Leitfaden*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010577816, 2015.

M. Wagner, C. Weyell, K. Christiaanse, A. Mikoleit, S. Hellweg, **A. Frömelt**, B. Steubing, A. Schlüter, P. Geyer, S. Cisar, J. Carmeliet, K. Orehounig, G. Mavromatidis, M. Schlegel, H. Zaugg, and S. Gruber, *Zernez Energia 2020 - Aktionsplan*, Eidgenössische Technische Hochschule Zürich (ETHZ): Zurich, Switzerland; DOI: http://dx.doi.org/10.3929/ethz-a-010525998, 2015.

P. Meier, **A. Frömelt**, and W. Kinzelbach, *Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data*, *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 999–1008, 2011.

A. Frömelt, *Marketing Compost in Nepal – Field Testing of Sandec's Compost Marketing Handbook, Eawag-Publication,* 2007