Working Paper Sustainability and Innovation
No. S 10/2019

Charlotte Huang
Rainer Elsland

A survey-based approach to estimate residential electricity consumption at municipal level in Germany

Summary

In the context of the German *Energiewende* (energy transition), energy system modelling is used to investigate possible future scenarios of the national energy system. These models depend on regionally disaggregated input data to adequately capture interdependencies in the energy system at high resolution. In Germany, official energy consumption statistics are only published at the national (AGEB, 2019a) and federal state (LAK, 2019) levels – there are no official statistics on residential electricity consumption with higher regional resolution.

So far, energy system modelling has typically relied on specific consumption values or constant per-capita estimators (see Beer (2012) and Hartel et al. (2017)) to approximate residential electricity consumption with a regional resolution beyond that of federal state. The use of primary data on electricity consumption at household level has so far been limited. Yet, primary data, e.g. from the German Residential Energy Consumption Survey (GRECS), displays heterogeneity in household-level electricity consumption, which cannot be captured by constant per-capita distribution keys. This study aims to investigate whether integrating primary data into quantitative modelling of energy systems contributes to a more realistic representation of regional electricity consumption by accounting for heterogeneity in household electricity consumption.

A synthetic population is generated via the Iterative Proportional Fitting (IPF) algorithm based on primary data on household electricity consumption taken from the German Residential Energy Consumption Survey (GRECS), as well as region-specific data at municipal (*Gemeindeebene*) level taken from the 2011 German census. Total residential electricity consumption at municipal level is then inferred from the synthetic population. Estimates of total residential electricity consumption were derived for 2011 for all municipalities of the German state Rhineland-Palatinate and evaluated against benchmark values from the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017) of the same year, which is available at the regional resolution of German urban and rural districts, referred to here as “counties” (*Stadt- und Landkreise*).

The derived estimates achieved an R^2 of around 0.99 with respect to benchmark values. Overall, the estimates were 6.8% below the benchmark value for Rhineland-Palatinate. It can be concluded that Iterative Proportional Fitting (IPF) constitutes a viable approach to integrate primary data into deriving regional estimates of residential electricity consumption at municipal level.

Key words: Regional analysis, techno-economic modelling, survey-based approach, residential sector, Iterative Proportional Fitting (IPF), synthetic population, electricity consumption

Table of Contents Page

1 Background and objective

The Climate Action Plan 2050, adopted by the Federal Republic of Germany in 2016, outlines Germany's strategy to lower its long-term greenhouse gas emissions as required under the Paris Climate Agreement (BMUB, 2016). In addition to the overall climate target of reducing total greenhouse gas emissions by 55 percent by 2030 compared to 1990 levels, sector-specific targets have been set (BMUB, 2016, p. 8). These plan emission reductions ranging from 31% for industry to 67% for the building sector (BMUB, 2016, p. 8).

The targets specified above are to be realized through a transformation process known as the German *Energiewende* (energy transition), which consists of a gradual decarbonisation of the national energy system, combined with efforts to increase energy efficiency across sectors. Further, energy demand is to be met through an increased share of renewable energy sources. Specifically, renewable energies should make up at least 80 percent of gross final electricity consumption by 2050 (BMWi, 2017, p. 4). This requires significant restructuring of the national energy system and its infrastructures.

Energy system models are employed to analyse alternative transition pathways to reach these political targets. These models can incorporate technological and economic constraints to assess energy system dynamics at high resolution. In Germany, the grid development plan commissioned by the four German Transmission Network Operators (*50Hertz Transmission*, *Amprion*, *TransnetBW,* and *TenneT TSO*) called the "*Netzentwicklungsplan 2030*" (Grid Development Plan Electricity 2030), provides such an analysis by identifying necessary transmission network requirements to ensure future grid stability under different possible future demand scenarios (Fraunhofer ISI, 2017).

A similar assessment of possible consequences of the energy transition on the energy system is done by Beer (2012), who uses a regional energy model to assess the effects of flexible cogeneration plant operation in the context of increased regional expansion of renewable energy. Similarly, a recent study published by Fraunhofer ISI together with the Karlsruhe Institute of Technology (KIT) analyses the possible future effects of an increased share of renewable energies in industry, transport, and the heating sector, as well as the impacts on energy security and grid stability in southern Germany (Hartel, et al., 2017).

To calibrate such techno-economic models, data with high regional resolution are required as input. Whereas official statistics are freely accessible on the location of energy generating power plants with a minimum net nominal capacity of 10

MW (Bundesnetzagentur, 2019), official data on energy consumption is severely restricted in its regional scope. At the national level, annual official statistics on energy consumption by sector are published by the *Arbeitsgemeinschaft Energiebilanzen e.V. (AGEB),* the German Working Group on Energy Balances (AGEB, 2019a). The *Anwendungsbilanz* (energy end-use balance) further breaks down national total energy consumption by sector, energy carrier, and energy end-use (AGEB, 2019b). Similarly, at the federal state level (*Länderebene*), annual consumption data differentiated by sector are provided by the Federal State Statistical Offices (LAK, 2019). However, there are no statistics on energy end-use at the federal state level comparable to those given by the *Anwendungsbilanz* at national level.

In the framework of the federal states' energy atlases (*Energieatlanten)*, the regional offices of several German federal states have countered the lack of spatially disaggregated data by deriving per-capita estimates at municipal level (*Gemeinden*). For this purpose, homogenous consumption patterns are assumed and total electricity consumption is distributed regionally based on simplified drivers, such as population size (see Bayerische Staatsregierung (2018) and Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (2019)). Faced with the lack of regionally disaggregated electricity consumption data, energy system modelling research has also used auxiliary data available at the desired regional resolution to approximate electricity consumption. Beer (2012), and Fraunhofer ISI (2017), for instance, used data on sectoral drivers of energy consumption, such as production figures in industry and specific consumption values per dwelling size to derive regional estimates of electricity consumption for subsequent analysis. Whereas electricity consumption data at a higher regional aggregation level are commonly used in energy system modelling together with distribution keys based on auxiliary data, to our knowledge, integrating primary data on electricity consumption into energy system modelling in Germany has been limited.

Contrary to the commonly used specific consumption values (Beer, 2012) or the constant per-capita estimators (Bayerisches Landesamt für Umwelt, 2018), primary data on electricity consumption reflect the heterogeneity of actual electricity consumption at household level. Findings from the German Residential Energy Consumption Survey (GRECS), for instance, indicate significant differences in mean electricity consumption between households in eastern and western federal states, as well as between different household types (RWI and forsa, 2015, p. 44).

The aim of this study is to explore whether integrating primary data into quantitative modelling of energy systems contributes to a more realistic representation of residential electricity consumption by accounting for possible heterogeneity at household-level. The methodology uses primary data as an alternative to the commonly used official statistics on annual total electricity consumption published at national or federal state level, and auxiliary data of electricity consumption drivers. The modelling approach is applied to a case study of municipalities (*Gemeinden*) in the federal state of Rhineland-Palatinate and evaluated using the *Netzentwicklungsplan 2030* (Grid Development Plan Electricity 2030) as a benchmark.

The study is structured as follows: First, an overview of methodologies to derive residential electricity consumption at high regional resolution will be given (Chapter 2). This is followed by a description of the method to be applied (Chapter 3). This method is used to estimate residential electricity consumption for the federal state of Rhineland-Palatinate at municipal level, and the derived estimates are evaluated (Chapter 4). Following concluding remarks (Chapter 5), a critical appraisal of the methodology, as well as an outlook to future research are given (Chapter 6).

2 Literature review

Overall, the methodologies to derive regionally disaggregated electricity consumption estimates are largely determined by the availability of official statistics on electricity consumption, which differ by country. The subsequent literature review will therefore be structured by country to gain insights into the regional resolution, the methodologies used and the corresponding data requirements. A detailed overview of the methodologies used to derive residential energy consumption data in Europe has been published by the European Commission as part of a *Manual for Statistics on Energy Consumption in Households* (Eurostat, 2013). A brief overview of international practices can be found in Diekmann et al. (2000). In addition to the efforts made by national and local authorities to compile residential electricity consumption data, research has used the available official statistics and enhanced the spatial resolution of these data by applying quantitative modelling approaches.

The subsequent section gives a brief overview of selected international practices relying on primary data to derive residential electricity consumption. Primary data on residential electricity consumption are generally derived either from obligatory business surveys directed at electricity suppliers, or from voluntary household surveys, which often only cover a nationally representative fraction of the respective country. In addition to the overview of international methodologies, an overview of data availability in Germany and selected German studies with a resolution beyond the German federal state level is given in Section 2.3.

2.1 Obligatory business surveys

Business surveys often achieve nationwide coverage at high regional resolution but depend on a corresponding legal framework being in place. Both the Netherlands and the United Kingdom have made data collection on residential electricity consumption possible at a regional resolution of household addresses by introducing obligatory business surveys directed at electricity suppliers. By surveying electricity companies, Statistics Netherlands (*Centraal Bureau voor de Statistiek, CBS*) derives the annual electricity consumption of private households based on dwelling address (Statistics Netherlands, 2019). The business registry is used to differentiate between residential electricity consumption and electricity used by small businesses (Eurostat, 2013, p. 113). A similar approach is used by the Department for Business, Energy and Industrial Strategy (BEIS) in the United Kingdom. Annual electricity consumption data from meter point readings are gathered from data aggregator companies in obligatory business surveys (BEIS, 2018, p.

25). Meter point readings are matched to dwelling addresses using meter post-code address files (BEIS, 2018, p. 27) to obtain address-level consumption data.

2.2 Voluntary household surveys

Voluntary household surveys are also used to estimate residential electricity consumption. Unlike business surveys, voluntary household surveys generally do not achieve national coverage, making the additional use of modelling approaches necessary. Statistik Austria (2018) has extrapolated data from household surveys on electricity end-use to the level of NUTS-2 administrative regions based on household and dwelling characteristics (Statistik Austria, 2018, pp. 4, 13). In the United States of America (USA), the U.S. Energy Information Administration (EIA) periodically administers the Residential Energy Consumption Survey (RECS) to a nationally representative sample of around 5,600 households (EIA, n.d.). To allow inference at regional resolution beyond the national level, Zhang et al. (2018) have enhanced the pool of household and dwelling characteristics contained in the RECS sample by statistically matching observations with observations from the U.S. Public Use Microdata Samples (PUMS). The resultant dataset contains an enlarged set of household variables as well as energy consumption data from the RECS. The information contained was used to impute energy consumption estimates for the remaining unmatched household observations of the PUMS. The pool of household observations together with the matched or imputed energy consumption data was then used to generate synthetic populations matching actual populations based on selected household variables up to U.S. ZIP code level. Inferences on the region’s total residential electricity consumption were then drawn from the synthetic population.

Table 1 below provides a brief overview of the methodologies employed to generate residential electricity consumption statistics by country, the achieved regional resolution, as well as the data used.

Table 1: Overview of methodologies using primary data to derive residential electricity consumption at high regional resolution.

Study	Regional resolution	Approach	Data sources
Austria, Statistik Austria (2018)	NUTS-2	Extrapolation based on household and dwelling characteristics	Voluntary household survey on electricity consumption; Auxiliary data from the Austrian Microcensus
The Netherlands, Statistics Netherlands (CBS) (2019)	Dwelling address	Data matching based on addresses and business registry	Obligatory business surveys on electricity supply; Addresses from business registry
United Kingdom, Department for Business, Energy and Industrial Strategy (BEIS, 2018)	Dwelling address	Data matching based on meter postcode address files	Obligatory business survey on electricity consumption; Meter postcode address files
United States of America, Zhang et al. (2018)	ZIP-code level	Statistical matching to generate large sample of household types with corresponding electricity consumption Generation of a synthetic population to infer energy consumption at ZIP-code level	Household electricity consumption survey (RECS); Household characteristics survey (Public Use Microdata Samples (PUMS)); Regional distribution of household characteristics at ZIP-code level (American Community Survey (ACS))

2.3 Survey data in Germany

In Germany, both obligatory business surveys and voluntary household surveys have been used to estimate electricity consumption – albeit mostly restricted to regional areas above the municipal level. Official statistics on national and federal state residential electricity consumption are published by the *Arbeitsgemeinschaft Energiebilanzen e.V. (AGEB),* the German Working Group on Energy Balances, and the statistical offices of the federal states, respectively. The underlying data are from obligatory business surveys directed at electricity utility companies (Statistisches Bundesamt Deutschland, 2018). Consumption data contained in the survey are not differentiated regionally below the federal state level (Ibid.). In the federal state energy atlases, however, these official statistics are distributed based on municipal population size to approximate corresponding electricity consumption for the region (see Bayerische Staatsregierung (2018) and Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (2019)).

A voluntary household survey, the German Residential Energy Consumption Survey (GRECS), commissioned by the German Federal Ministry for Economic Affairs and Energy (RWI and forsa, 2005), has been conducted in addition to the annual official statistics on energy consumption described above. The GRECS has been used to assess residential energy end-use (RWI and forsa, 2005), as well as the response of households to changing electricity prices (Frondel & Kussel, 2019). Although the GRECS can be seen as the German counterpart to the U.S. Residential Energy Consumption Survey (RECS), to the best of our knowledge, no attempts have been made so far to use the information contained in it to estimate regionally disaggregated consumption data based on a synthetic population, as done by Zhang et al. (2018) for the U.S.

Table 2 provides an overview of selected German studies with a regional resolution beyond federal state level.

Table 2: Overview of methodologies used in Germany to derive residential electricity consumption at a regional resolution higher than the federal state level.

Study	**Regional resolution**	**Approach**	**Data sources**
Fraunhofer ISI (2017), Germany	German counties (*Stadt- und Landkreise*)	Distribution of national residential electricity consumption through distribution keys	Total national residential electricity consumption by end-use (AGEB, 2019b); Household and dwelling characteristics differentiated by German counties
Bayerisches Landesamt für Umwelt (2018), Germany	Municipalities (*Gemeinden*) in the German state of Bavaria (*Bayern*)	Distribution of federal state residential electricity consumption by population size	Total residential electricity consumption of the German state of Bavaria; Population size of municipalities in Bavaria
Beer (2012), Germany	German municipalities (*Gemeinden*)	Specific electricity consumption values per household based on dwelling size are extrapolated to the region of interest	Specific consumption values published by the Association of German Engineers (*Verein Deutscher Ingenieure e.V.*) (VDI, 1994); Dwelling area by municipality (DESTATIS, 2008)
Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (2019), Germany	Municipalities (*Gemeinden*) in the German state of North Rhine-Westphalia (*Nordrhein-Westfalen*)	Distribution of federal state electricity consumption by population size	Total electricity consumption of the German state of North Rhine-Westphalia; Population size of municipalities in North Rhine-Westphalia

3 Methodology

3.1 Methodological concept

Primary data from household surveys are often limited to nationally representative, anonymized samples. Collecting detailed micro-level data at high spatial resolution is subject to high costs and data privacy concerns. Detailed micro-level data for smaller (geographical) areas of interest are therefore less common (see for example Moeckel, Spiekermann, & Wegener (2003, p. 4)). This is reflected in the availability of residential electricity consumption data in Germany. Whereas primary data on household electricity consumption representative at the national level are given in the German Residential Energy Consumption Survey (GRECS) (forsa; RWI, 2016), higher-resolution data are lacking.

In order to deal with this problem, spatial microsimulation research has resorted to generating synthetic populations (see O'Donoghue, Morrissey, & Lennon (2014) for a comprehensive overview of spatial microsimulations and their applications). This is one way to generate robust micro-level data for smaller (geographical) areas of interest for which official data are either lacking or severely restricted due to data privacy issues.

Synthetic populations generally rely on combining sample data[1] representative for a higher level of aggregation and auxiliary data characterizing the region of interest (see Zhu & Ferreira Jr. (2014) and Zhang et al. (2018) for example). In the context of energy system modelling, synthetic populations offer a way to combine nationally representative primary data on electricity consumption, such as information contained in the GRECS, with region-specific data on electricity consumption drivers (such as the number of households or the distribution of population and dwelling characteristics) of the region of interest. This allows an additional source of information, namely primary survey data, to be incorporated into regional estimates of residential electricity consumption.

Figure 1 illustrates the methodological concept used for the subsequent analysis. Based on survey data from a nationally representative sample and the distribution of selected characteristics in the region of interest (referred to as "aggregate totals" or "control totals"), a synthetic population is generated to represent the actual population of interest. The synthesized population can then be used to draw an

1 It should be noted that sample-free approaches also exist (see Gargiulo, Ternes, Huet, & Deffuant (2010) for an example).

inference on a variable of interest, in this case, total residential electricity consumption.

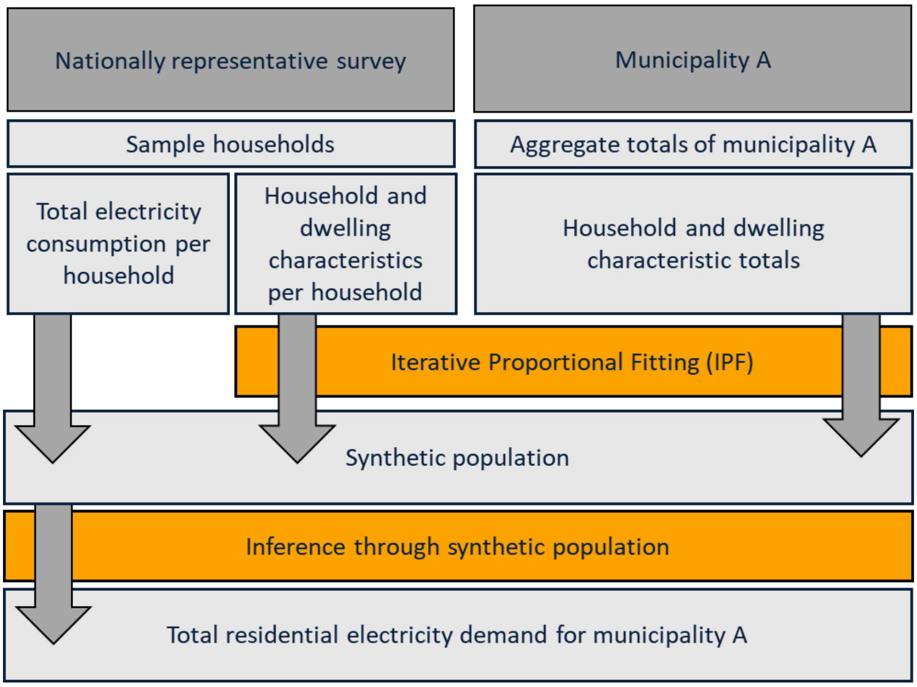

Figure 1: Schematic overview of methodological approach.

3.2 Synthetic population

A synthetic population refers to a synthesized pool of individual agents for which a statistical distribution of selected features approximates that of the real population (Moeckel, Spiekermann, & Wegener, 2003).

Whereas one-dimensional statistical distributions are often available at the desired regional resolution, e.g. from a national census, the joint distribution at that level of resolution is often lacking or inaccessible due to confidentiality concerns (Ryan, Maoh, & Kanaroglou, 2009). Synthetic populations combine the information from several one-dimensional distributions to approximate a joint distribution of the selected features. Whereas the distribution of gender and age may be given at high regional resolution, for example, the joint distribution, *age x gender,* may be lacking. This joint distribution can then be approximated through a synthetic population.

Synthetic populations have largely been generated as input for subsequent agent-based modelling due to the need for detailed micro-level data in this field of research (see for example Zhu & Ferreira Jr. (2014), or the Transportation Analysis and Simulation System (TRANSIMS) by Smith, Beckman, & Baggerly (1995)). However, the derived statistical characteristics of a synthetic population can also be used for inference purposes.

In the context of energy consumption inference, Zhang et al. (2018) used synthetic populations to derive residential energy consumption at ZIP-code level for the United States of America (USA). Based on an enlarged dataset of household observations containing both household characteristics and energy consumption values, Zhang et al. (2018) generated synthetic populations to fit one-dimensional distributions of selected characteristics of the regions of interest. The average percentage difference between the derived electricity consumption estimates and observed consumption data was 13.6% (Zhang, et al., 2018, p. 170).

3.3 Iterative Proportional Fitting

Synthetic populations can be derived via Iterative Proportional Fitting (IPF). The method is described in detail by Birkin & Clarke (1988) and was implemented in their SYNTHESIS model - a sample database of synthetically IPF-generated populations for the Leeds Metropolitan district. More recent examples include Moreno & Moeckel (2018), who adapt the IPF algorithm to synthesize a population for the greater Munich metropolitan area using the German census.

By adjusting multi-dimensional micro-level data, referred to as a “seed matrix”, to fit one-dimensional distributions of aggregate totals of the region of interest, the IPF approximates the unobserved joint distribution of the target area by preserving the internal (correlation) structure of the seed matrix.

The underlying concept of this approach is illustrated in Figure 2 below using two exemplary one-dimensional distributions of features (*Feature 1* and *Feature 2*) reflecting the population characteristics from a fictive area of interest. It is not possible to derive the joint distribution based only on the one-dimensional distributions. However, sample data including both features contains information on their correlation structure and can be used to approximate the joint distribution.

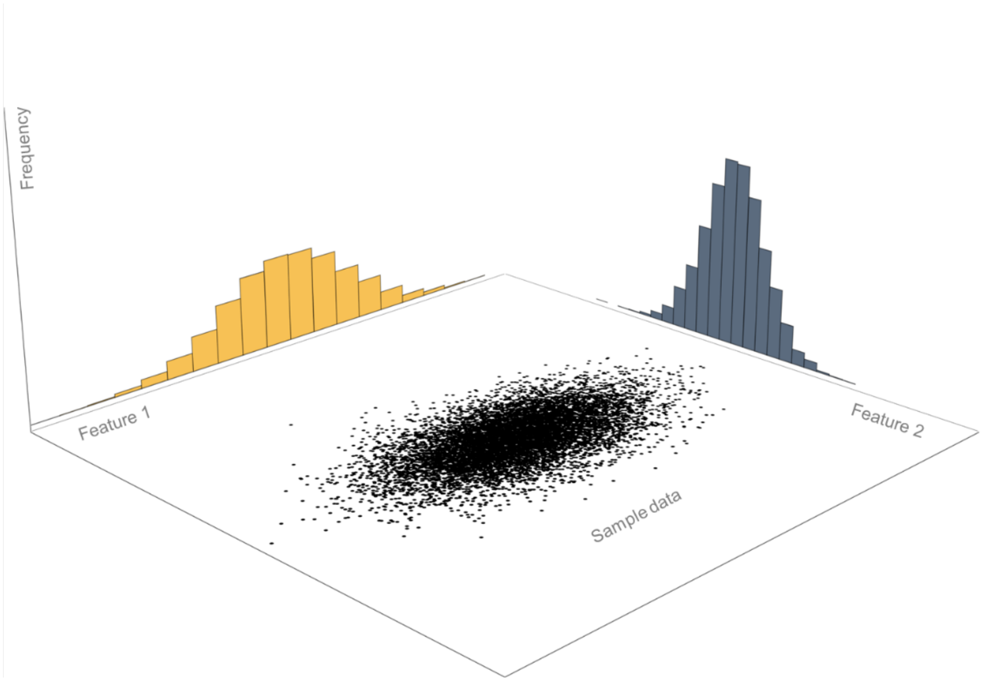

Figure 2: Visualization of the relationship between one-dimensional distributions at the region of interest and the joint distribution from sample data.

How the IPF algorithm works is best understood through the framework of contingency table analysis, where observations from the seed matrix are iteratively assigned new weights to fit row and column totals ("aggregate totals") from the target population. The survey weight calibration is based on pre-selected variables. These need to be present in both the seed matrix and the aggregate totals of the population of interest. This illustrates one main restriction of the IPF algorithm, namely that data needs to be categorical or discrete for the algorithm to reach convergence.

The IPF algorithm deterministically assigns and adjusts weights to each observation until the distance between the total of cross-tabulation values (individual records) and specified aggregate totals is minimized[2]. Because weights are assigned to individual records of the seed matrix, the correlation structure between the characteristics of each observation remains untouched. Under the assumption that the sample observations share the same correlation structures as the population of interest for selected features, the derived joint distribution matches

2 For a proof of convergence of the algorithm, see Fienberg (1970).

that of the target population. Assuming statistical equivalence between the synthesized joint distribution and the unobserved real distribution, third variable inferences can then be drawn based on the synthetic population.

The IPF-generated synthetic population adapts the sample to fit known one-dimensional distributions of selected characteristics of the region of interest while preserving the correlation structure. This is illustrated using a simplified example in Figure 3: The IPF algorithm adjusts the underlying sample (*a.*) to fit known distributions of household types. Populations *b.* and *c.* represent populations of two distinct regions of interest, each with a different number of total households, as well as different composition of household types; the IPF selects and weights each sample household from *a.* to conform to these constraints ("aggregate totals"). The weights assigned by the IPF algorithm therefore reflect the frequency with which a specific household is represented in the synthetic population.

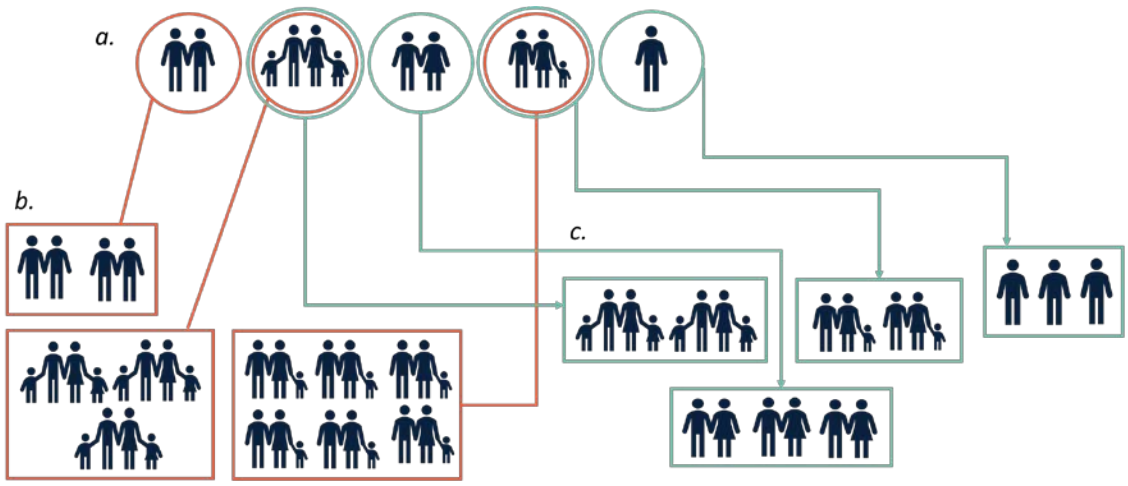

Figure 3: Schematic representation of the IPF algorithm.

To allow inference, the IPF assumes that differences between regions are fully captured by differences in their composition of household types, population and dwelling characteristics. It needs to be assumed that each household type of the region of interest is represented at least once in the underlying sample. The IPF algorithm then identifies the relevant household and assigns weights accordingly. The quality of inference is therefore dependent on three elements: the quality of the underlying sample that the correlation structure is adequately captured, and that variables are measured correctly. Additionally, the validity of the derived synthetic population is also dependent on the accuracy of the aggregate totals for the region of interest.

Further, inference validity is also linked to variable selection. Inference validity improves with increasing explanatory power of the selected variables over the variable of interest. It is important to note that the influence does not need to be direct, but can be through third unobserved variables not explicitly accounted for in the IPF procedure. For example: Dwelling size can affect electricity consumption directly, but also indirectly through third variables, such as income and the corresponding ownership of electrical appliances. Inference therefore relies on the assumption that the relationship of selected characteristics and a variable of interest in the target population is reflected in the underlying micro-level sample data without needing to explicitly identify or quantify the causal relationship of those variables.

Variable selection is subject to a trade-off between the number and categories of selected variables and the speed and ease of convergence of the IPF algorithm. The speed and likelihood of convergence decreases with the number of variables and categories used, and it is therefore recommended to collapse categories where possible to reduce computation time (Battaglia, Izrael, Hoaglin, & Frankel, 2009). Furthermore, it has been recommended that each category should be represented with around 5 percent in the seed matrix (sample) to further aid convergence of the IPF algorithm (Ibid.).

The IPF algorithm has been implemented in various software packages (see *mipfp* from Barthélemy & Suesse (2018) for an example in R). In this paper, the STATA module *ipfraking* by Kolenikov (2014) was used.

The algorithm of the *ipfraking* package generates and adjusts weights w_j for each observation j of the sample S. After each iteration k, the weights are adjusted with respect to the v^{th} control variable. $T[X_v]$ represents the aggregate total of the target population for control variable v. Convergence is declared and the algorithm stopped if the weighted totals $\sum_{j \in S} w_j^{k,p} x_v$ match the specified aggregate totals $T[X_v]$ within a default tolerance of $\delta_D = 10^{-6}$ for all $v = 1, \dots, p$ control variables.

$$w_j^{k,v} = \begin{cases} w_j^{k,v-1} \dfrac{T[X_v]}{\sum_{l \in S} w_l^{k,v-1} x_{vl}}, & x_{vj} \neq 0 \\ w_j^{k,v-1}, & x_{vj} = 0 \end{cases}$$

The iterations are repeated until convergence is reached, defined as $D_k < \delta_D$, (Kolenikov, 2014, p. 12).

$$D_k = \max_{j \in S} \frac{|w_j^{k,p} - w_j^{k-1,p}|}{w_j^{k-1,p}}$$

Further, the algorithm stops when divergence, defined as $D_k > D_{k-1}$, is detected, meaning that the distance between the weighted observations and the aggregate controls increases with subsequent iterations, or a maximum number of iterations has been reached.

3.4 Assessment of estimator variability

The IPF procedure described above was used to generate a synthetic population, from which point estimates of total annual residential electricity consumption at municipal level were derived. Based on the obtained synthetic populations, the variability of these point estimates can be assessed via bootstrapping (Efron, 1979). Under the assumption that the distribution of household electricity consumption in the region of interest is matched, resampling with replacement from the synthesized population can be used to obtain confidence intervals for the statistic of interest – namely, total residential electricity consumption.

The weights assigned by the IPF algorithm reflect the frequency with which a particular household is represented in the synthesized population. By taking the IPF weights as probability weights and iteratively resampling synthetic populations with replacement from the pool of household observations, several estimates of the statistic of interest can be derived and a distribution of estimated total residential electricity consumption for a selected municipality can be obtained. Confidence intervals can be computed from this distribution to assess the variability.

Formally, this translates to the following:

Let $Y_A = \sum_i y_i$ be the variable of interest defined as the sum of electricity consumption y of all households $i = 1, \dots, N$ living in municipality A, and let $\hat{Y}_A = \sum_j x_j w_j$ be the estimate of the variable of interest, defined as the sum of frequency weights w as derived by the IPF algorithm, and the corresponding electricity consumption y of all households $j = 1, \dots, S$ from the GRECS sample.

By taking $b = 1, \dots, B$ samples of N households each, using the weights w_j as probability weights, bootstrap samples $\hat{Y}_{A1}, \dots, \hat{Y}_{AB}$ can be derived. With B approaching ∞, the sampling distribution of the estimate $\hat{Y}_A$ can be approximated.

It is important to note that the confidence interval generated for a region's total residential electricity consumption depends on the assumption that the synthetic population reflects the actual population of interest. If this is the case, the confidence intervals indicate the variability of the point estimate. However, the derived confidence intervals cannot be used to evaluate the accuracy of the IPF procedure in approximating the actual population of interest.

4 Case study

4.1 Data basis

The IPF algorithm was used to derive regional estimates of residential electricity consumption based on primary data of household electricity consumption. For the subsequent analysis, data from a nationally representative sample in the German Residential Energy Consumption Survey (GRECS) was taken as micro-level sample data, and selected variables at municipal level from the German 2011 census used as aggregate totals.

Table 3: Overview of data sources used to estimate residential electricity at municipal level, as well as benchmark data.

	Iterative Proportional Fitting		**Benchmark**
	Micro-level sample data	**Aggregate controls**	
Data source	German Residential Energy Consumption Survey (GRECS) (forsa; RWI, 2016)	German 2011 census (Statistisches Landesamt Rheinland-Pfalz, n.d.)	*Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017)
Regional resolution	Nationally representative household sample	Aggregate statistics at municipal level in Rhineland-Palatinate	Counties of Rhineland-Palatinate
Variables used	Per household: • Total annual electricity consumption (kWh) • Tenure type • Dwelling type • Household size • Degree of urban density • East/West German federal state	Per municipality: • Number of households by tenure type • Number of households by dwelling type • Number of households by household size • Degree of urban density of municipality	Per county: Total residential electricity consumption (GWh)

Residential electricity consumption was estimated at municipal level for the German state of Rhineland-Palatinate for the year 2011 and validated at county level

(*Stadt- und Landkreise*) using data from the *Netzentwicklungsplan 2030* of the same year.

4.1.1 Data for electricity demand assessment

The German Residential Energy Consumption Survey (GRECS) is a nationwide survey dedicated to collecting residential energy consumption data in Germany. The survey was commissioned by the German Federal Ministry of Economic Affairs and Energy and conducted by *RWI Leibniz-Institut für Wirtschaftsforschung e.V.* (RWI) and *forsa Gesellschaft für Sozialforschung und statistische Analysen mbH* (forsa). Its aim is to gain insights into residential energy consumption and end-use, and to supplement the national energy consumption statistics derived by the *AG Energiebilanzen* (RWI and forsa, 2005, p. 7). The survey was conducted in 2003 and 2005, as well as for the periods 2006-2008, 2008-2011, and 2011-2013. The survey was administered to a nationally representative sample of 15,000 households, with the latest survey period (2011-2013) achieving a net sample size of 8,561 German households (RWI and forsa, 2015, p. 5). In addition to detailed information on energy consumption by energy carrier, the survey collects information on socio-economic household characteristics, dwelling characteristics, and housing conditions.

In the period from 2011 to 2013, estimates of total national residential electricity consumption derived from the GRECS were within three petajoules, i.e. less than 1%, of the official national residential electricity consumption derived by the *AG Energiebilanzen* (RWI and forsa, 2015, p. 9). This is remarkable considering the contrasting data sources used: whereas the *AG Energiebilanzen* (AGEB) uses business surveys to derive figures of national electricity consumption (Statistisches Bundesamt Deutschland, 2018), estimates using the GRECS are based on household observations. The GRECS constitutes the most comprehensive source of primary data on residential energy consumption in Germany and is therefore employed to assess the use of such primary data to derive regionally disaggregated consumption estimates.

Aggregate controls at municipal level are taken from the German 2011 census[3]. The German 2011 census contains population and housing statistics at municipal level and provides a detailed overview of household characteristics as well as dwelling conditions. In the framework of IPF analysis, the German 2011 census

3 Whereas data are available with a resolution below that of municipalities, they are not freely accessible and are subject to strict data protection measures (FDZ, n.d.).

was used by Moreno & Moeckel (2018) to synthesize a population for the greater Munich metropolitan area. The analysis in this paper is done for the year 2011 to match the year of the German census. The analysis is restricted to municipalities in the German state of Rhineland-Palatinate.

For the IPF algorithm, the selection of aggregate controls has to be limited to those variables present in both the sample data, in this case the GRECS, and the regionally specific aggregate controls from the German 2011 census. This restricts the possible variables to the following household, dwelling and regional characteristics:

Table 4: Overview of variables used in the IPF procedure.

Household characteristics	Dwelling characteristics	Regional characteristics
Household size	Tenure type	Eastern/Western federal states
	Dwelling type	Degree of urbanization of municipalities

The selected variables are described in detail below.

Eastern/Western federal states

The GRECS shows a significant difference in annual per-capita electricity consumption between West (≈ 1,538 kWh) and East German federal states (≈ 1,368 kWh). The lower per-capita electricity consumption has been linked to higher electricity costs in East German federal states (RWI and forsa, 2015, p. 44) and persists at the household level (see Figure 4)

Figure 4 below depicts the sample distribution of per-capita electricity consumption at household level based on the household's location in either East or West German federal states.

This difference persists across household types, as illustrated exemplarily for a single household in Figure 5 below.

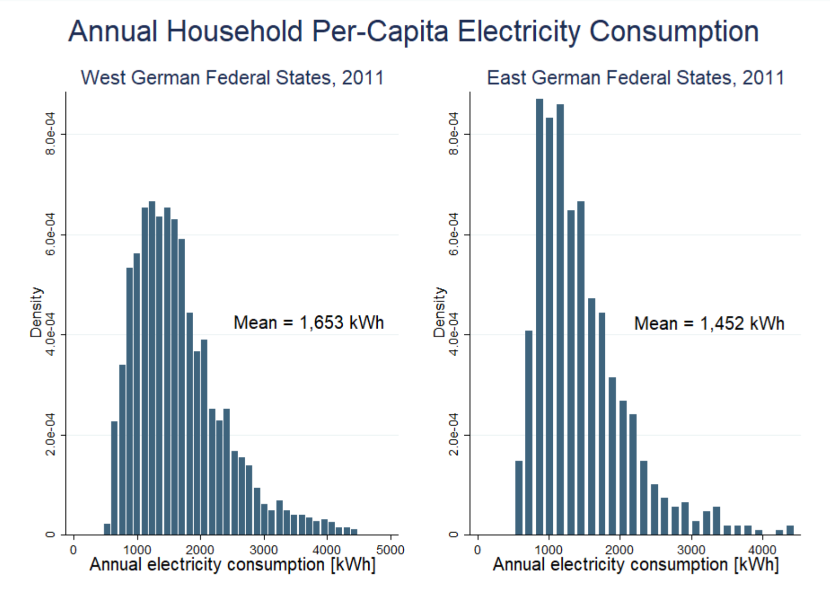

Figure 4: Annual per-capita electricity consumption of households from West and East German federal states based on data from forsa and RWI (2016).

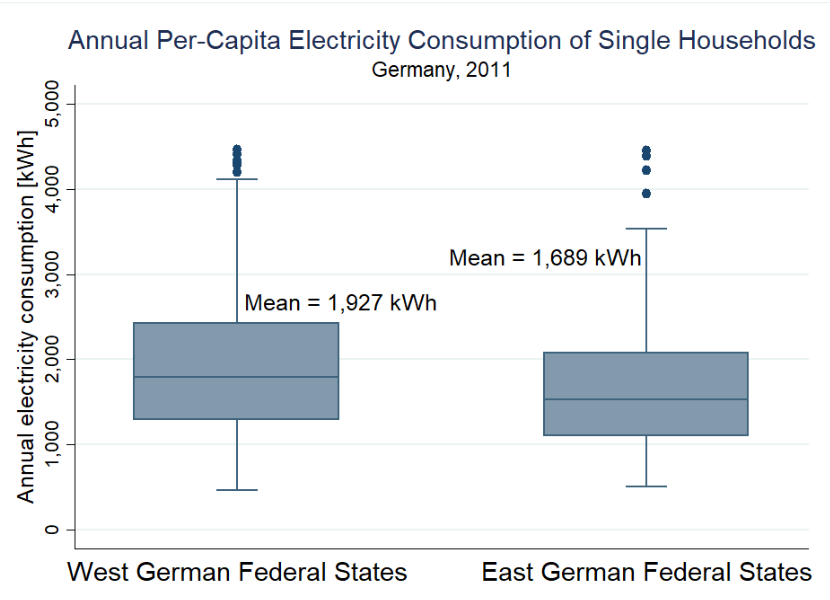

Figure 5: Annual per-capita electricity consumption by federal state for single households based on data from forsa and RWI (2016)[4].

4 Lower and upper adjacent lines for all box plots have been calculated as defined by Tukey (1977) (see StataCorp (2013, p. 93) for more information).

In line with the estimation of national residential electricity consumption by RWI and forsa (2015) based on the GRECS, the subsequent IPF analysis also differentiates between household observations from Eastern and Western states. Given that the analysis was done for municipalities of the federal state of Rhineland-Palatinate, which is located in Germany's West, only observations from West German households were considered in the subsequent IPF analysis, so as to prevent structural differences in electricity consumption between East and West German households from distorting the estimates. A total sample of 2,995 West German households were therefore considered in the subsequent estimation.

Degree of urbanization

The degree of urbanization categorizes administrative units based on their population agglomeration using minimum population thresholds per 1 km². The indicator is given for each municipality as part of the German municipal directory (*Gemeindeverzeichnis*) (Statistisches Bundesamt, 2012). Cities are defined as densely populated areas (code 1), towns and suburbs are assigned intermediate density (code 2), and rural areas are defined as thinly populated areas (code 3) (Eurostat, 2018). Similar to the differentiation by East and West German states, the IPF analysis only considers subsets of the GRECS categorized by the same degree of urbanization as the municipality of interest.

Household size

Electricity consumption generally increases with household size due to increased usage and ownership of electrical appliances. However, changes in electricity consumption do not increase linearly with household size (Schlomann, et al., 2004, p. 25). This is reflected in the GRECS, which shows decreasing per-capita electricity consumption values with increasing household size. This emphasizes the importance of accounting for the heterogeneity in per-household electricity consumption, which cannot be captured using constant per-capita estimators.

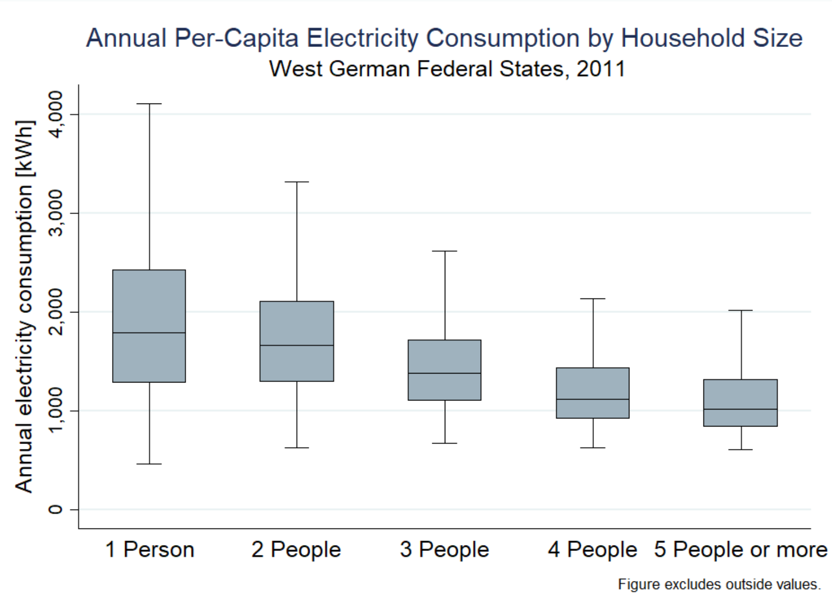

Figure 6: Annual per-capita electricity consumption by household size based on data from forsa and RWI (2016).

Tenure type

A dwelling is classified as "rented" when it is occupied or used in return for regular payments (which can include free rent) by the tenant. The term "tenants" therefore refers to occupants that pay rent, as well as those living rent-free in the dwelling, but who are not considered owners of the dwelling (Statistische Ämter des Bundes und der Länder, 2015, p. 63). The differences in household per-capita electricity consumption by tenure type are notable when controlling for household size.

Figure 7 below indicates a significantly lower electricity consumption for tenants. This might be due to the effects of unobserved third variables, such as wealth, on electricity consumption.

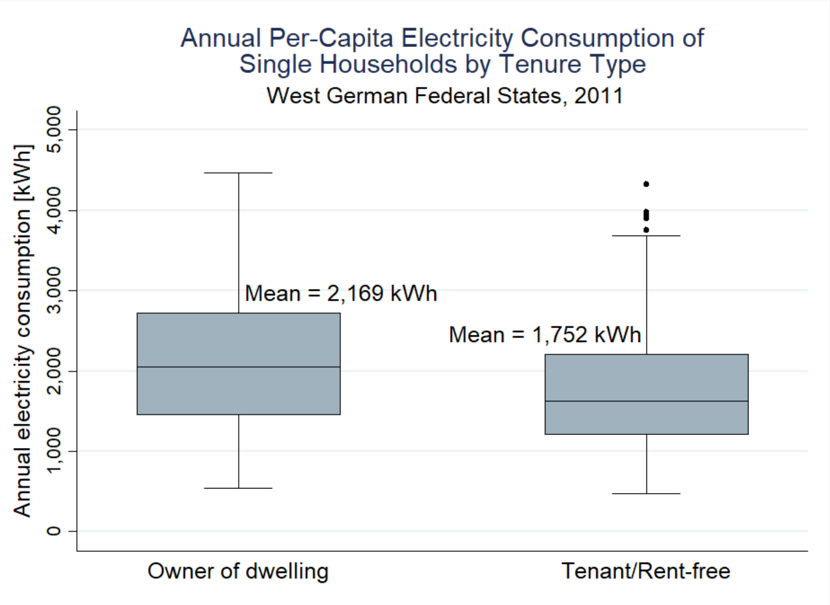

Figure 7: Annual per-capita electricity consumption by tenure type for single households based on data from forsa and RWI (2016).

Dwelling type

The number of dwelling units in the building of residence have been collapsed in both the data source for aggregate totals from the German 2011 census and the GRECS to form the following categories of dwelling type: “single-family house” (*Einfamilienhaus*), “two-family house” (*Zweifamilienhaus*) and “apartment building” (*Mehrfamilienhaus*).

Figure 8 below illustrates the differences in per-capita electricity consumption for a single household by dwelling type.

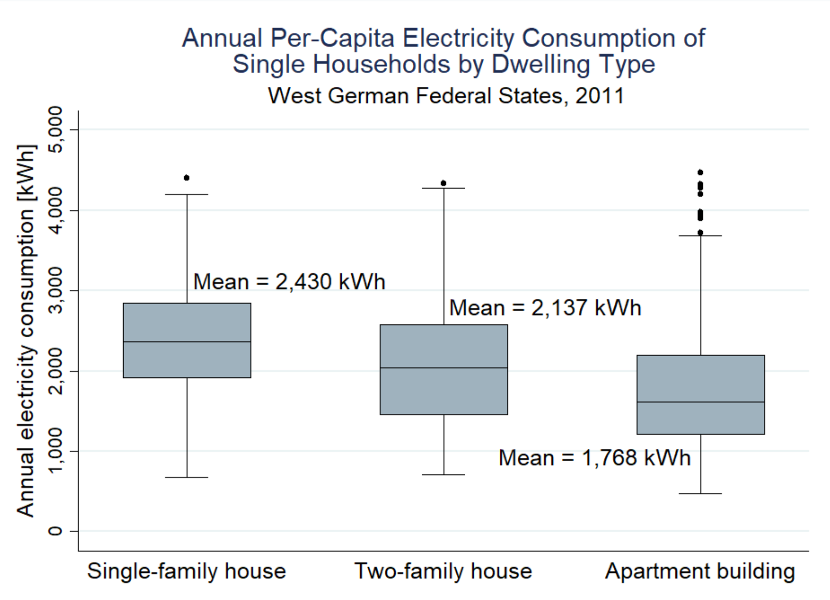

Figure 8: Annual per-capita electricity consumption by dwelling type for single households based on data from forsa and RWI (2016).

4.1.2 Benchmark data

The *Netzentwicklungsplan 2030* (Grid Development Plan Electricity 2030), commissioned by the four German Transmission Network Operators and developed by the Fraunhofer ISI (Fraunhofer ISI, 2017), was used to evaluate the estimations derived using the IPF algorithm. The *Netzentwicklungsplan 2030* assesses the grid expansion plans of the transmission network operators in the context of regional electricity demand in Germany until 2030. In this context, three different alternative scenarios are analysed that consider different possible future network usage situations for the next ten to fifteen/fifteen to twenty years (Fraunhofer ISI, 2017).

The *Netzentwicklungsplan 2030* used FORECAST (FORecasting Energy Consumption Analysis and Simulation Tool) for the analysis. FORECAST is a bottom-up model, which integrates socio-economic and technological drivers to model long-term scenarios for future energy demand at the national and regional levels (Fraunhofer ISI, 2017).

The *Netzentwicklungsplan 2030* regionalizes electricity consumption retroactively until 2008 to the county (*Stadt- und Landkreise*) level. Estimates have been validated at the level of German states using official statistics on electricity consumption by sector published by the Federal State Statistical Offices (LAK, 2019). In addition to validation at the level of federal state, estimates have been validated at county level using historical data provided by the Transmission Network Operators. Estimates were in 90% accordance with the data provided by the Transmission Network Operators (Fraunhofer ISI, 2017, p. 16). The *Netzentwicklungsplan 2030* therefore contains validated data with a regional resolution beyond federal state level, making it a viable benchmark for subsequent IPF estimates. IPF-derived estimates are evaluated at county level by aggregating municipal estimates to match the resolution of the benchmark data provided.

4.2 IPF specifications

The IPF algorithm was investigated using the following six specifications summarized below.

Table 5: Overview of IPF specifications.

	Household size	**Tenure type**	**Dwelling type**	**Degree of urbanization**
IPF 1	X	X		
IPF 2	X	X	X	
IPF 3	X		X	
IPF 4	X	X		X
IPF 5	X	X	X	X
IPF 6	X		X	X

IPF 1 iteratively adjusts household observations of the GRECS sample to fit the distribution of household size and tenure type at municipal level. This information is given as a cross table as part of the German 2011 census (Statistisches Landesamt Rheinland-Pfalz, n.d.).

IPF 2 uses the information contained in the cross table and further includes aggregate totals of the distribution of dwelling types[5].

IPF 3 uses household size and dwelling type aggregates.

IPF 4, IPF 5, IPF 6 include a region's degree of urbanization. To do so, subsamples of the GRECS were formed based on the degree of urbanization of the households' regions of residence. The IPF only considers a subsample matching the urban density of the target region. An overview of the distribution of households across different degrees of urban density is given in Table 6 below.

Table 6: Distribution of GRECS household observations by degree of urban density.

	Degree of urban density 1	**Degree of urban density 2**	**Degree of urban density 3**
Household observations	915 households	1,452 households	628 households

Based on the regional identification number included in the GRECS for each household observation, households were assigned a degree of urban density using information from the German municipal directory (*Gemeindeverzeichnis*). IPF 4, 5 and 6 operate on a subsample of observations that match the region of interest's degree of urbanization.

4.3 Results

4.3.1 IPF estimation results

Estimates of total residential electricity consumption for 2011 were derived using IPF for each municipality of Rhineland-Palatinate. The estimations derived using IPF specification 1 are illustrated in Figure 9 below.

5 The census gives data on tenure type by household size, but does not include households in dwellings categorized as holiday and leisure dwellings (Statistische Ämter des Bundes und der Länder, 2015, p. 64). As a result, the aggregate totals of household size and tenure type are 10% lower on average than the aggregate totals of dwelling type. This discrepancy is countered by applying a constant factor to all three categories of dwelling type to ensure matching aggregate controls. The underlying assumption is that the discrepancy is equally distributed across dwelling types. Similarly, RWI and forsa (2015) have scaled dwelling types by applying the average vacancy rate across dwelling type categories in their estimation of national energy consumption using the GRECS (RWI and forsa, 2015, p. 74).

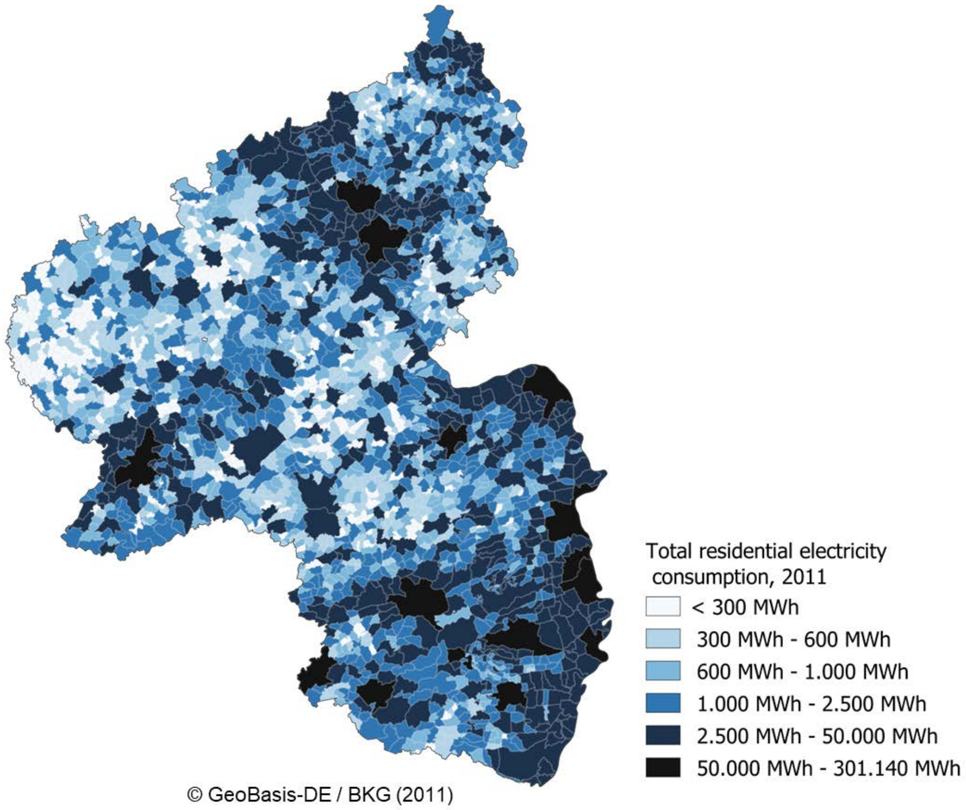

Figure 9: IPF estimates of total residential electricity consumption for municipalities of Rhineland-Palatinate, 2011.

In the absence of benchmark data at the level of German municipalities, data at the resolution of German counties (*Stadt-und Landkreise*) taken from the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017) were used to evaluate the IPF-derived estimates. The total residential electricity consumption of each municipality in Rhineland-Palatinate obtained via IPF was therefore aggregated to the level of German county to allow a comparison with the benchmark.

Table 7 below provides an overview of the mean absolute deviations across all municipalities in Rhineland-Palatinate aggregated to the county level as well as the statistical measure of R^2 for all six IPF specifications. The last column displays the total percentual deviation at state level.

Table 7: Overview of IPF estimates.

	Mean absolute deviation [GWh]	Mean absolute percentage deviation [%]	R^2	Total percentual deviation [%], Rhineland-Palatinate
IPF 1	11.33 GWh	7.64%	0.989	- 6.84%
IPF 2	13.73 GWh	8.97%	0.981	- 7.81%
IPF 3	15.01 GWh	9.90%	0.982	- 9.10%
IPF 4	14.05 GWh	9.49%	0.988	- 8.73%
IPF 5	14.08 GWh	9.23%	0.981	- 8.10%
IPF 6	15.89 GWh	10.60%	0.983	- 9.92%

Overall, there is relatively low variability between the estimates derived from the six different IPF specifications. A high R^2 value was achieved across all IPF specifications with IPF 1, which included household size and tenure type as aggregate totals, displaying the highest R^2 value, as well as the lowest mean absolute deviation with respect to the benchmark. The degree of urbanization does not seem to have high explanatory power for the variable of interest, since the IPF specifications not accounting for the degree of urbanization outperform their respective counterparts including the degree of urbanization[6].

The estimates derived by IPF 1 and aggregated to the county level are depicted together with the benchmark data in Figure 10 below.

6 See Table 5: Overview of IPF specifications.

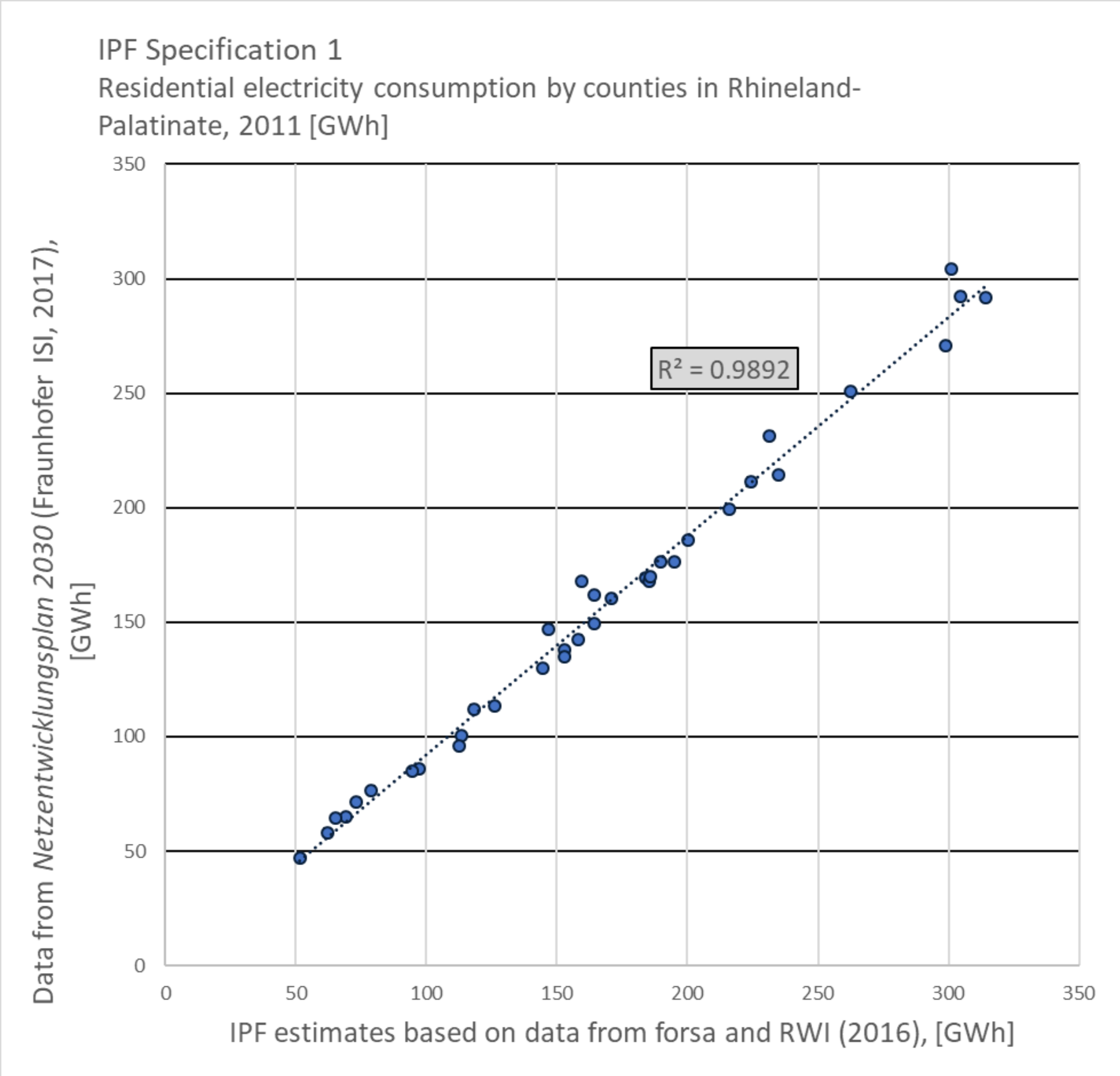

Figure 10: Illustration of IPF estimates vs. benchmark data.

Figure 11 below illustrates the estimates derived using IPF 1 as the percentual absolute deviation from the benchmark data of the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017). Percentage absolute deviation is generally lower for metropolitan areas (Stadtkreise).

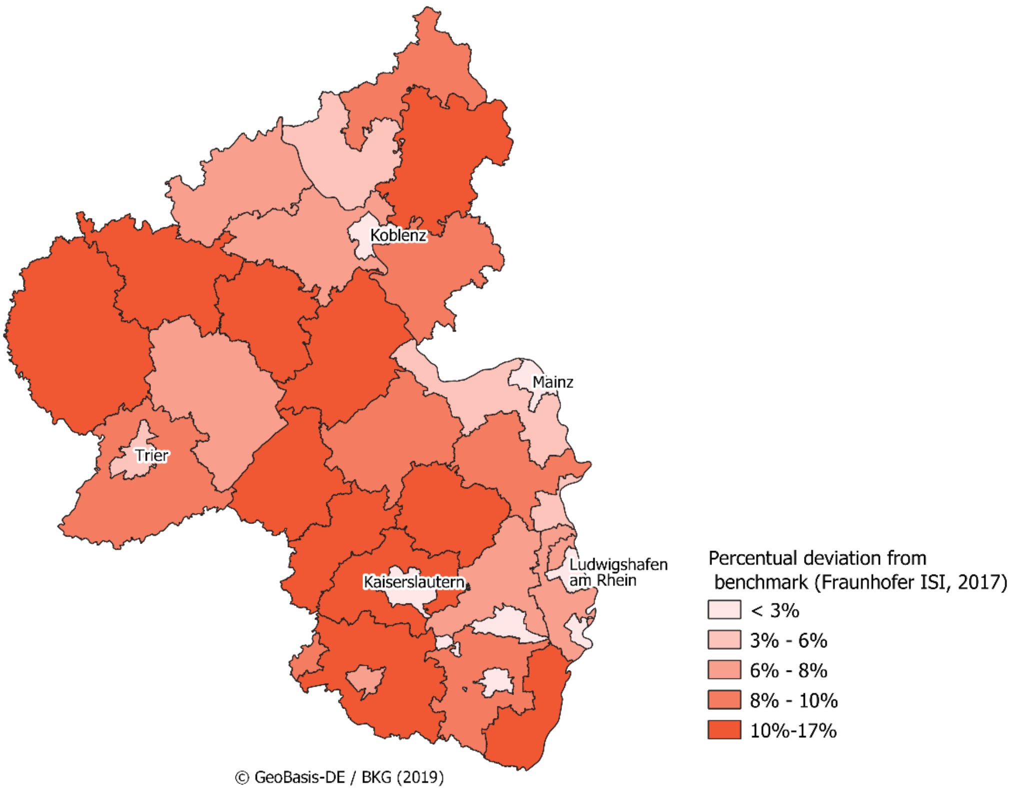

Figure 11: Absolute percentage deviation of IPF estimates at county level based on benchmark data from the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017).

4.3.2 Uncertainty of estimation results

Bootstrapping was applied to assess the confidence of the derived point estimate. Based on 50,000 Monte-Carlo sampling iterations, a distribution of the point estimate of total residential electricity consumption for 2011 was derived for each municipality in Rhineland-Palatinate. Based on the derived distribution, 95%-confidence intervals are obtained as the 2.5 and 97.5 percentiles of the bootstrap sample distribution.

This is illustrated for the example of Mainz in Figure 12 below.

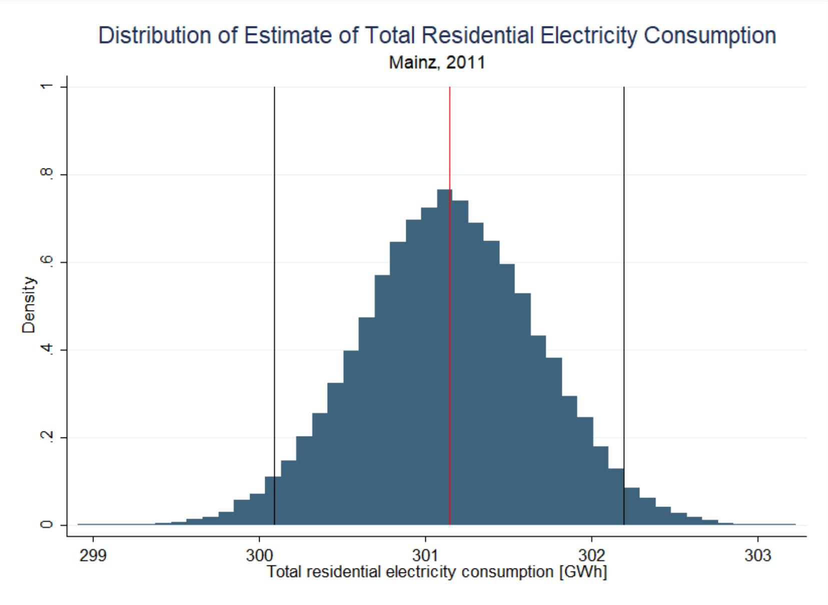

Figure 12: Bootstrap sample distribution of total residential electricity consumption of Mainz, 2011.

It is important to note that the derived confidence intervals do not assess the validity of the synthetic population derived via IPF. Rather, the bootstrap analysis assesses the variability of the derived estimator based on the given data and under the assumption that the derived synthetic population matches the actual population of interest.

As benchmark data are only available at the county level, point estimates of the total residential electricity consumption at municipal level were aggregated to the level of German counties and confidence intervals were obtained for this level.

Figure 13 depicts the relationship between the IPF-derived point estimates of total residential electricity consumption at county level together with their respective confidence intervals, and the benchmark from the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017).

The variability of many estimates is quite low. Even when including the upper confidence intervals, total residential electricity consumption is underestimated compared to data from the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017). There is discrepancy between the benchmark values and the distribution of residential electricity consumption derived via IPF. Changes in the IPF specifications with respect to variable selection and data used might therefore be required to move closer to the benchmark values of the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017).

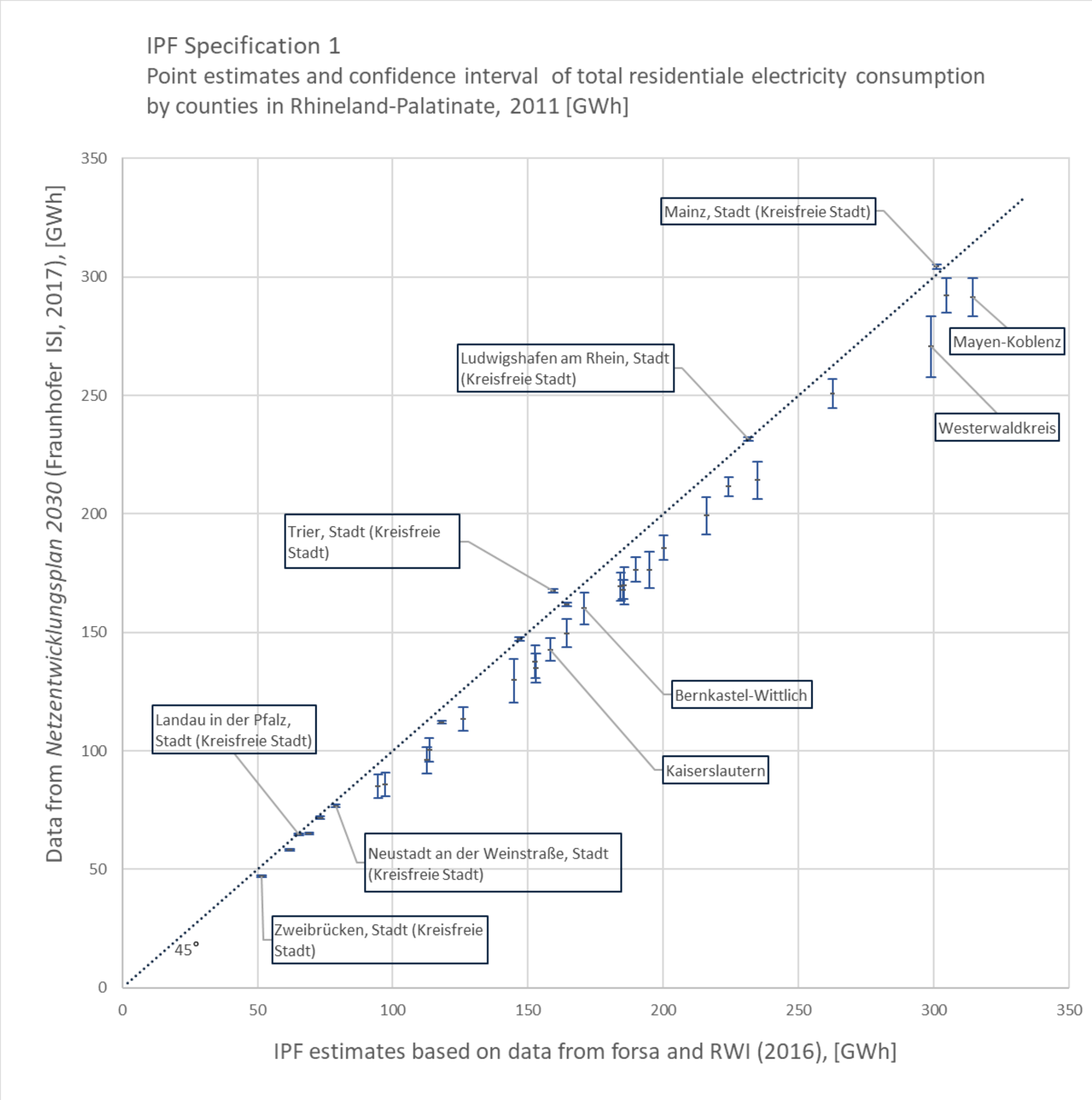

Figure 13: IPF point estimates of total residential electricity consumption including confidence intervals vs. benchmark data.

4.3.3 Discussion

The smallest error measurement was achieved for IPF 1, which used the fewest variables, namely household size and tenure type, as aggregate totals. Although this might initially suggest that the additional variables investigated, namely dwelling type and degree of urbanization, have low explanatory power for the variable of interest, several points should be considered.

Subsamples of varying size were used to include the degree of urbanization in IPF analyses (see Chapter 4.2). For instance, there were only 628 household observations in the subsample for regions with a degree of urbanization 3. This sample may be too small, and may have resulted in higher error measurements for estimations including the degree of urbanization.

Possible improvements could be achieved by including population level aggregate totals in addition to the household level aggregate totals investigated (see Müller & Axhausen (2012) for multi-level fitting algorithms). Underestimating the total residential electricity consumption may have been in part due to the collapsed category of households with 5 or more household members. For example, a region with many larger households may have been represented in the synthetic population by sample households of lower size, resulting in matching numbers of total households, but discrepancies regarding the overall population size. However, such variability should have been captured by bootstrapping. Yet, the IPF estimates including their respective upper confidence intervals were still below the benchmark values.

With respect to dwelling type, aggregate totals at municipal level did not include information about the ratio of holiday/leisure dwellings per dwelling type. The total number of dwellings was therefore often higher than the total number of households. To ensure convergence, adjustments in the number of dwellings per dwelling type were necessary (see Chapter 14.2) and a constant was applied across all dwelling types until aggregate totals of households were matched. This may have led to misspecifications of aggregate totals and therefore resulted in discrepancies between the synthesized population and the actual population.

5 Conclusions

By generating a synthetic population via Iterative Proportional Fitting (IPF), nationally representative primary data on household electricity consumption can be used to derive estimates of total residential electricity consumption at the level of German municipalities. The derived IPF estimates have a relatively high R^2 overall with respect to benchmark data of the *Netzentwicklungsplan 2030* (Fraunhofer ISI, 2017), but are generally lower than benchmark values at county (*Stadt- und Landkreise*) level.

Due to the lack of benchmark data on residential electricity consumption with a resolution below that of German counties, the estimates' validity at municipal (*Gemeinde*) level could not be adequately assessed. Unlike benchmark data at a higher aggregation level, benchmark data at municipal level would allow the derived estimates to be evaluated based on their ability to reflect distinct regional differences in residential electricity consumption between municipalities.

Even though the method described is able to combine data from various sources into the estimation, a major challenge when using IPF to derive regional estimates is the limited number of variables that can currently be selected as aggregate totals. Even though the German 2011 census constitutes the most detailed data source on population, household and dwelling characteristics at municipal level, there were a very limited number of variables available for subsequent IPF estimation based on the GRECS. Although data at higher resolution do exist, they are subject to confidentiality concerns (FDZ, n.d.), and were therefore not used in this assessment.

6 Critical appraisal and outlook

Possible applications of the method described are by no means restricted to the estimation of regional residential electricity consumption. The case study at the resolution of municipalities in Rhineland-Palatinate illustrates just one possibility of how information from survey data can be included into energy system modelling. The scope of the analysis could easily be extended to estimate energy consumption by end-use or for other energy carriers at municipal level or beyond, providing primary data containing such information is available.

Given the importance of high-resolution regional estimates for energy system modelling, and the relative scarcity of electricity consumption data with a resolution higher than federal state level, it seems sensible to combine as much information as possible to make the most of the available data. In this respect, Iterative Proportional Fitting (IPF) constitutes a viable approach, as it is able to combine nationally representative survey data with data on region-specific characteristics. Adaptations of the IPF, such as Iterative Proportional Updating (IPU), are able to account for aggregate totals at varying geographical levels (see Moreno & Moeckel (2018)), and through this, allow the integration of further additional information.

Whereas the German Residential Energy Consumption Survey (GRECS) specifically targets residential energy consumption, the use of other surveys for energy system modelling is yet to be assessed. Information contained in regularly conducted official surveys, such as the German time-use survey (*Zeitverwendungserhebungen*), or the German Income and Consumption Survey (*Einkommens- und Verbrauchsstichprobe*), represent possible sources of primary data on the energy consumption behaviour of households. In this context, IPF estimations may be combined with a preceding variable analysis to identify possible aggregate totals based on their explanatory power of household energy consumption. Using cluster analysis to typify households based on their energy consumption characteristics, for instance, may prove useful when combined with IPF estimation to derive high-resolution regional estimates of residential energy consumption.

Whereas benchmark data from the *Netzentwicklungsplan 2030* were validated at county level and achieved 90% accordance with historical data from the Transmission Network Operators (Fraunhofer ISI, 2017, p. 16), data at a higher regional resolution could provide additional information on how to further improve estimates of regional electricity consumption. The lack of adequate benchmark data at municipal level therefore remains the main challenge when assessing methodologies to derive estimates at high regional resolution.

7 References

AGEB. (2019a). *Energiebilanz der Bundesrepublik 2017.* Berlin: Arbeitsgemeinschaft Energiebilanzen e.V.

AGEB. (2019b). *Anwendungsbilanzen für die Endenergiesektoren in Deutschland in den Jahren 2013 bis 2017.* Berlin: AG Energiebilanzen e.V.

Barthélemy, J., & Suesse, T. (2018). mipfp: An R Package for Multidimensional Array Fitting and Simulating Multivariate Bernoulli Distributions. *Journal of Statistical Software, 86*(1), pp. 1-20.

Battaglia, M. P., Izrael, D., Hoaglin, D. C., & Frankel, M. R. (2009). Practical Considerations in Raking Survey Data. *Survey Practice, 2*(5), pp. 1-10.

Bayerisches Landesamt für Umwelt. (2018, Mai). *Informationen zu den Rahmendaten: Stromverbrauch.* Retrieved September 10, 2019, from Energie-Atlas Bayern: https://www.energieatlas.bayern.de/file/pdf/1231/Rahmendaten_Stromverbrauch.pdf

Beer, M. G. (2012). *Regionalisiertes Energiemodell zur Analyse der flexiblen Betriebsweise von Kraft-Wärme-Kopplungsanlagen.* München: Dissertation Technische Universität München.

BEIS. (2018). *Sub-National Consumption Statistics: Methodology and Guidance Booklet.* London: Department for Business, Energy & Industrial Strategy (BEIS).

Birkin, M., & Clarke, M. (1988). Synthesis—A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples. *Environment and Planning A: Economy and Space, 20*(12), pp. 1645-1671. doi:https://doi.org/10.1068/a201645

BKG. (2019). *Verwaltungsgebiete 1:2 500 000, Stand 01.01. (VG2500)*. Retrieved September 25, 2019, from Bundesamt für Kartographie und Geodäsie (BKG): https://gdz.bkg.bund.de/index.php/default/open-data/verwaltungsgebiete-1-2-500-000-stand-01-01-vg2500.html

BMUB. (2016). *Climate Action Plan 2050 - Principles and goals of the German government's climate policy.* Berlin: Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB).

BMWi. (2017). *Report of the Federal Government on Energy Research 2017.* Berlin: Federal Ministry for Economic Affairs and Energy (BMWi).

Bundesnetzagentur. (2019, March 8). *Power plant list*. Retrieved September 16, 2019, from Bundesnetzagentur: https://www.bundesnetzagentur.de/EN/Areas/Energy/Companies/SecuritYOfSupply/GeneratingCapacity/PowerPlantList/PubliPowerPlantList_node.html;jsessionid=56FB0CEC870459DFF200232CE037CE33

Capasso, A., Grattieri, W., Lamedica, R., & Prudenzi, A. (1994). A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems, 9*(2), pp. 957-964.

DESTATIS. (2008). *Statistik lokal 2008 - Daten für die Gemeinden, kreisfreien Städte und Kreise Deutschlands.* Düsseldorf: Statistische Ämter des Bundes und der Länder.

Diekmann, J., Wittke, F., Ziesing, H.-J., Eichhammer, W., Gruber, E., Kling, N., . . . Geiger, B. (2000). *Erarbeitung kostengünstiger Erhebungsformen zur Erfassung des effektiven Energieverbrauchs.* Berlin, Karlsruhe, München: Zusammenarbeit mit dem DIW Berlin, Fraunhofer ISI, TU München, im Auftrag des Bundesministeriums für Wirtschaft und Technologie.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics, 7*(1), pp. 1-26.

EIA. (n.d.). *Residential Energy Consumption Survey (RECS)*. Retrieved September 21, 2019, from U.S. Energy Information Administration: https://www.eia.gov/consumption/residential/about.php

Eurostat. (2013). *Manual for statistics on energy consumption in households.* Luxembourg: Publications Office of the European Union.

Eurostat. (2018, December 6). *eurostat - Statistics Explained*. Retrieved October 3, 2019, from Glossary: Degree of urbanisation: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Degree_of_urbanisation

Eymann, L., Rohrer, J., & Stucki, M. (2014). *Energieverbrauch der Schweizer Kantone: Endenergieverbrauch und Mittelabfluss durch den Energie-Import.* Wädenswil: ZHAW Zürcher Hochschule für Angewandte Wissenschaften im Auftrag der Schweizerischen Energie-Stiftung (SES).

FDZ. (n.d.). *Forschungsdatenzentrum der Statistischen Ämter der Länder*. Retrieved September 20, 2019, from Zensus 2011: https://www.forschungsdatenzentrum.de/de/haushalte/zensus-2011

Fienberg, S. E. (1970). An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics, 41*(3), pp. 907-917.

forsa; RWI. (2016). Energieverbrauch privater Haushalte in Deutschland, 2011-2013. RWI-GRECS. Version:1. RWI - Leibniz-Institut für Wirtschaftsforschung. Dataset. doi:http://doi.org/10.7807/EVpH:11-13:V1

Fraunhofer ISI. (2017). *Netzentwicklungsplan Strom: Entwicklung der regionalen Stromnachfrage und Lastprofile: Begleitgutachten: Studie im Auftrag der deutschen Netzbetreiber.* Karlsruhe: Fraunhofer-Institut für System-und Innovationsforschung ISI.

Frondel, M., & Kussel, G. (2019). Switching on Electricity Demand Response: Evidence for German Households. *Energy Journal, 40*(5), pp. 1-16. doi:10.5547/01956574.40.5.mfro

Gargiulo, F., Ternes, S., Huet, S., & Deffuant, G. (2010, January 22). An Iterative Approach for Generating Statistically Realistic Populations of Households. *PLoS ONE, 5*(1). doi:https://doi.org/10.1371/journal.pone.0008828

Ge, Y., Dai, J., Qian, K., Hepburn, D., & Zhou, C. (2015). *Simulation of Domestic Electricity Load Profile by Multiple Gaussian Distribution.* Lyon: 23rd International Conference on Electricity Distribution.

Hartel, R., Slednev, V., Yilma, H. Ü., Keles, D., Fichtner, W., Eßner, A., . . . Wietschel, M. (2017). *Dekarbonisierung des Energiesystems durch verstärkten Einsatz erneuerbaren Stroms im Wärme-, Verkehrs- und Industriesektor bei gleichzeitigen Stilllegungen von Kraftwerken – Auswirkungen auf die Versorgungssicherheit in.* Karlsruhe: Karlsruhe Institute of Technology (KIT), Fraunhofer Institute for Systems and Innovation Research ISI.

Kolenikov, S. (2014). Calibrating Survey Data using Iterative Proportional Fitting (Raking). *The Stata Journal: Promoting communications on statistics and Stata, 14*(1), pp. 22–59. doi:10.1177/1536867x1401400104

LAK. (2019). *Energiebilanzen der Länder.* Bremen: Statistische Landesamt Bremen.

Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen. (2019). *Daten- und Berechnungsgrundlagen Strom*. Retrieved August 6, 2019, from Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen: https://www.energieatlas.nrw.de/site/strom/daten-und-berechnungsgrundlagen

Moeckel, R., Spiekermann, K., & Wegener, M. (2003). Creating a synthetic population. *In Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM)* (pp. 1-18). Sendai, Japan: Centre for Northeast Asian Studies.

Moreno, A., & Moeckel, R. (2018). Population synthesis handling three geographical resolutions. *ISPRS International Journal of Geo-Information, 7*(5), p. 174.

Müller, K., & Axhausen, K. W. (2012). *Multi-level fitting algorithms for population synthesis.* Zurich: IVT, ETH Zurich.

O'Donoghue, C., Morrissey, K., & Lennon, J. (2014). Spatial Microsimulation Modelling: a Review of Applications. *International Journal of Microsimulation, 7*(1), pp. 26-75.

Richardson, I., Thomson, M., Infield, D., & Clifford, C. (2010). Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings, 42*(10), pp. 1878-1887. doi:https://doi.org/10.1016/j.enbuild.2010.05.023

RWI and forsa. (2005). *Erhebung des Energieverbrauchs der privaten Haushalte für das Jahr 2003, Forschungsprojekt Nr. 61/04 des Bundesministeriums für Wirtschaft und Technologie.* Essen, Berlin: Rheinisch-Westfälisches Institu für Wirtschaftsforschung, forsa Gesellschaft für Sozialforschung und statistische Analysen mbH.

RWI and forsa. (2015). *Erhebung des Energieverbrauchs der privaten Haushalte für die Jahre 2011-2013: Forschungsprojekt Nr. 54/09 des Bundesministeriums für Wirtschaft und Technologie.* RWI Projektberichte.

Ryan, J., Maoh, H., & Kanaroglou, P. (2009). Population Synthesis: Comparing the Major Techniques Using a Small, Complete Population of Firms. *Geographical Analysis, 41*(2), pp. 181-203. doi:https://doi.org/10.1111/j.1538-4632.2009.00750.x

Schlomann, B., Gruber, E., Eichhammer, W., Kling , N., Diekmann, J., Ziesing, H.-J., . . . Eckl , R. (2004). *Energieverbrauch der privaten Haushalte und des Sektors Gewerbe, Handel, Dienstleistungen (GHD).* Karlsruhe, Berlin, Nürnberg, Leipzig, München: Zusammenarbeit mit dem DIW Berlin, TU München, GfK Nürnberg, IE Leipzig, im Auftrag des Bundesministeriums für Wirtschaft und Arbeit.

Smith, L., Beckman, R., & Baggerly, K. (1995). TRANSIMS: Transportation analysis and simulation system. Seattle: Proceedings of the 5th National Transportation Planning Methods Applications Conference.

StataCorp. (2013). *Stata 13 Stata Graphics Reference Manual.* College Station, TX: Stata Press.

Statistics Netherlands. (2019, April 4). *Supply of electricity and natural gas via the national grid.* Retrieved October 1, 2019, from CBS:

https://www.cbs.nl/en-gb/our-services/methods/surveys/korte-onderzoeksbeschrijvingen/supply-of-electricity-and-natural-gas-via-the-national-grid

Statistik Austria. (2018). *Strom- und Gastagebuch.* Wien: Statistik Austria.

Statistische Ämter des Bundes und der Länder. (2015). *Zensus Kompakt, Ausgabe 2014 .* Statistische Ämter des Bundes und der Länder.

Statistisches Bundesamt. (2012). *Gemeindeverzeichnis - Gebietsstand: 31.12.2011.* Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2019). *Fortschreibung des Bevölkerungsstandes – Bevölkerung: Bundesländer, Stichtag*. Retrieved September 25, 2019, from Genesis-Online Datenbank: https://www-genesis.destatis.de/genesis/online/data;sid=BDC6F543E41CC55672DE3F98A0DEC1CA.GO_1_1?operation=abruftabelleBearbeiten&levelindex=1&levelid=1569406409209&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&ausw

Statistisches Bundesamt Deutschland. (2018). *Jahreserhebung über Stromabsatz und Erlöse in der Elektrizitätsversorgung: Qualitätsbericht.* Wiesbaden: Statistisches Bundesamt Deutschland.

Statistisches Landesamt Rheinland-Pfalz. (n.d.). *Statistisches Landesamt Rheinland-Pfalz*. Retrieved October 3, 2019, from Zensus: Kreise, Städte, Gemeinden: https://www.statistik.rlp.de/no_cache/de/gesellschaft-staat/zensus/kreise-staedte-gemeinden/

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

Tukey, J. W. (1977). *Exploratory data analysis.* Massachusetts: Addison-Wesley.

U.S. Energy Information Administration. (2018). *Residential Demand Module of the National Energy Modeling System: Model Documentation.* Washington, DC: U.S. Department of Energy.

VDI. (1994). *VDI-02 94 Energieverbrauchskennwerte für Gebäude - VDI 3807.* Düsseldorf: Verein Deutscher Ingenieure (VDI).

Velte, D., & Kuittinen, H. (2018). *Case Study Report - Energiewende.* Luxembourg: Publications Office of the European Union.

Wiesmann, D., Azevedo, I. L., Ferrão, P., & Fernández, J. E. (2011). Residential electricity consumption in Portugal: Findings from top-down and bottom-up models. *Energy Policy, 39*(5), pp. 2772-2779.

Zhang, W., Caleb, R., Subhrajit, G., Garikap, V. M., Dilkina, B., Brown, M. A., & Pendyalag, R. M. (2018). Estimating residential energy consumption in metropolitan areas: A microsimulation approach. *Energy, 155*, pp. 162-173. doi:https://doi.org/10.1016/j.energy.2018.04.161

Zhu, Y., & Ferreira Jr., J. (2014). Synthetic population generation at disaggregated spatial scales for land use and transportation microsimulation. *Transportation Research Record: Journal of the Transportation Research Board No.2429*, pp. 168–177.

Authors’ affiliations

Prof. Dr. Rainer Elsland
Fraunhofer Institute for Systems and Research (Fraunhofer ISI)
Competence Center Energy Technology and Energy Systems
Karlsruhe, Germany

Charlotte Huang
Fraunhofer Institute for Systems and Research (Fraunhofer ISI)
Competence Center Energy Technology and Energy Systems
Karlsruhe, Germany

Contact: Prof. Dr. Rainer Elsland

Fraunhofer Institute for Systems
and Innovation Research (Fraunhofer ISI)
Breslauer Straße 48
76139 Karlsruhe
Germany
Phone: +49 721 6809-438
E-Mail: rainer.elsland@isi.fraunhofer.de
www.isi.fraunhofer.de

Karlsruhe 2019