🏆 *reviewed paper*

A Spatial Microsimulation Model for the Estimation of Heat Demand in Hamburg

M. Esteban Muñoz H.

(Marcelo Esteban Muñoz Hidalgo, HafenCity University Hamburg, Germany, marcelo.hidalgo@hcu-hamburg.de)

1 ABSTRACT

Most spatial bottom up domestic heat models are based on an existing building stock data set, this can be the official digital cadastre (ALKIS in Germany), community based data sources (e.g: open-street) or collected data stored on a GIS system. On this paper we present an innovative method for the creation of spatial bottom up domestic heat models that do not need an existing building stock data-set as input. The advantage of this type of models are its transferability, speed and flexibility.

The presented model uses national standardized statistical data, making it possible to apply it for any region in the country without having to modify the model. Because the input data for the model is more compact the speed of the model increases significantly, the type of analysis possible with a high speed model allow us to perform a microsimulation of the building stock for the entire country, see (Muñoz H., Seller, & Peters, 2015). The presented model structure allows us to quickly develop dynamic simulation scenarios.

We present a spatial microsimulation model developed for the simulation of domestic heat demand. The presented model use the GREGWT R package to create a synthetic building stock benchmarked to aggregated small area statistics from the 2011 German census. We create this synthetic building stock from the 2010 microsensus. The heat estimation if performed on the microcensus with help of the heat R package.

The simulation results present a monthly heat demand at a microlevel for the entire city of Hamburg. The generated data for the estimation of heat demand can be use as input data for other Agent Based Models. By enriching the microcensus with time use data, we are able to generate the basis data for the construction of activity based urban models. We see the development of this type of urban models as an essential development of urban planning, specially for a smart urban development. The generation of microdata is a fundamental part of the smart city development.

2 INTRODUCTION

This paper present a spatial microsimulation model used for the estimation of residential heat demand for the city of Hamburg. In this paper we first reweight a population sample to small geographical areas and create a synthetic population based on this reweighted sample. The reweighting algorithm is performed with an implementation of the GREGWT method and the creation of synthetic population with a Fitness Based (fbs) method. The presented results show a good internal validation of the spatial microsimulation model.

Microsimulation, introduced by (Orcutt, 1957), is a simulation method used by many scientific disciplines. The method aims to simulate many types of social phenomena at a micro-level. The main idea of this type of models is to create or sample a representative synthetic population and design simulation on this population. There is not a defined scope of the model application because this type of models can be apply to all kinds of phenomena. Commonly the subjects of the synthetic population are individuals of families, but this micro-units can be any type of agents. We can develop a model simulating firms, animals or in this case buildings.

The spatial on spatial microsimulation was introduces by social geographers (Clarke & Holm, 1987). Spatial microsimulation add the spatial constrain to these models. Instead of having a single representative population we generate synthetic populations representative of small geographical areas. The spatial "granularity" at which this models can be internally validated are these geographical areas. For overview of spatial microsimulation models see (Tanton, 2014, O'Donoghue, Morrissey, & Lennon, 2014). Theoretically we can generate synthetic populations at any geographical level, nonetheless the performance of the model will reduce as does the geographical aggregation. Trying to find a representative population for a very small area with just a few residents is harder that for larger populations. Also, on smaller geographical areas we encounter more zeros on the benchmarks, many algorithms can't deal with the zeros percent on the benchmarks, GREGWT cat. An issue that we encounter in the city of Hamburg is that at a smaller spatial aggregation we see that available data is harder to get or has been scrambled in order to protect anonymity of the residents.

On this paper we make use of a generalised regression for the weighting of sample survey results, an implementation of this process is known as the GREGWT algorithm. We use the GREGWT R library (Muñoz H., Vidyattama, & Tanton, 2015), the library is an implementation of the GREGWT algorithm. The original GREGWT algorithm was developed by the Australian Bureau of Statistics (Bell, 2000). The algorithm is also used at the National Center for Social and Economic Modeling (NATSEM) on their spatial microsimulation model spatialMSM (Tanton, 2007, Tanton, Williamson, & Harding, 2014).

The GREGWT algorithm only reweights a sample to known benchmarks of small geographical areas. The results from the GREGWT algorithm are non integer weights. We do not consider the non integer weights as a synthetic population, because we can not use this population for other urban simulations, like the population of digital building stocks or as input to agent based models. The GREGWT R library implements a slightly modified version of the algorithm proposed my Ma and Srinivasan (2015) for the creation of synthetic populations.

For the computation of domestic heat demand we make use of the HEAT R library (Muñoz H., 2015). This library implements a monthly quasi steady state model for the estimation of domestic heat demand.

3 METHOD

3.1 Reweighting the Survey Sample with GREGWT

For the reweighting of the survey to generate the representative weights for each geographical area we need two data-sets: (1) a survey to reweight with design weights; and (2) the benchmarks for each geographical area. We use the German Mikrozensus (Statistische Ämter des Bundes und der Länder, 2010), this is a representative sample of the German population of individual records. The Scientific-File contains 1% of the total population. The geographical aggregation of this sample is the federal state. Hamburg is both, a state and a city. For the reweighting of the survey we take only records from Hamburg into account. A reduction of the initial survey considerably reduces the computational time of the algorithm, specially for the later part of the analysis on which we implement the fbs method. On special cases, the results can be better if we use a bigger sample size. In this case the internal validation of the model is performing extremely well, and therefor we do not need a bigger sample size.

The parameters used on the survey sample and the corresponding benchmarks are listed on Table 1. It is important to notice that not all these parameters have the same units. While age, marital status and sex count individuals, that is f.ex: number of people on a geographical area that are male, floor area refers to the total number of dwelling units on a given floor area category. The last two parameters: year of construction and number of dwelling units count the number of buildings on these categories. The R implementation of GREGWT can deal with this variation of units and benchmark the sample accordingly.

The reweighting of the survey sample and the latter creation of a synthetic population and building stock is benchmarked not only to demographic characteristics of the small geographical areas but to characteristics of the building stock. For the computation of heat demand we need to estimate the heat transmission coefficients of the building components. The defining parameter for the estimation of these coefficients is the construction year of the building. We use this parameter, and the other two parameters describing the building stock for the classification of the individual records on the sample into building types. Attach to each building type is a heat transmission coefficient and other relevant parameters needed for the estimation of heat demand.

The GREGWT algorithm is an implementation of Sigh & Mohl (A. Singh & Mohl, 1996). For a description of the algorithm and its applications see Tanton, Vidyattama, Nepal, and McNamara (2011). A mathematical explanation and comparison to other smaptial microsimulation method can be found an Rahman, Harding, Tanton, and Shuangzhe (2010). For a comparison to the well establish IPF method and a discussion of the roll of design/initial weights see (Muñoz H., Tanton, & Vidattama, 2015).

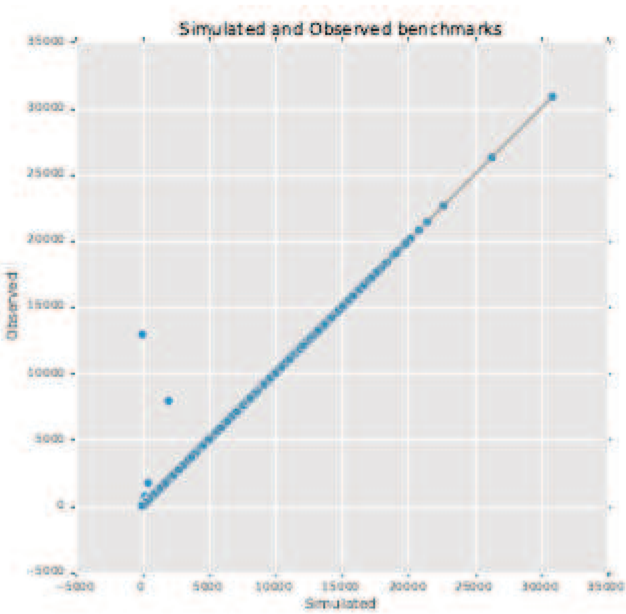

Fig. 1: Plot of simulated (x-axis) and observed (y-axis) sum of marginal totals.

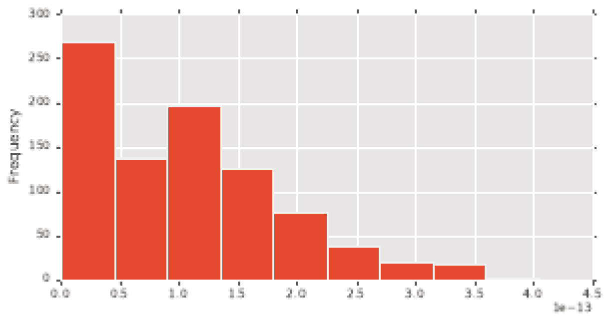

Fig. 2: Distribution of PSAE

MC Code	Census Code	Description
EF1	/	Federal State (NUTS 2)
EF952	/	Weight
EF44	ALTER KURZ	Age
EF49	FAMSTND AUSF	Marital status
EF46	GESCHLECHT	Sex
EF492	WOHNFLAECHE 20S	Floor area of the dwelling
EF494	BAUJAHR MZ	Year of construction
EF635	ZAHLWOHNGN HHG	Number of dwellings in a building

Table 1: Used benchmarks from the 2011 Census and corresponding micro census attributes. Source: Microcensus (Statistische Ämter des Bundes und der Länder, 2010) & Census (Statistische Ämter des Bundes und der Länder, 2011).

Aim of the GREGWT algorithm is to find a set of weights w that can be used to match the microdata survey X to a set of small geographical areas benchmarks T so that $T=\sum w_j X_j$. The Algorithm aims to reduce the difference between the initial d and the estimated weights w. The algorithm maintains the distance D between design and estimated weights low. The GREGWT algorithm makes use of the truncated Chi-Squared distance function, represented in Equation 1 for the computation of the distance between weights.

$$D = 1/2 \sum_j (w_j - d_j)^2/d_j \qquad (1)$$

The distance minimization equation is expressed as the Lagrangian function of the Chi-Squared function. With this equation we formulate an equation for the computation of new weights. Where $X'_j=\sum \lambda_k X_{j,k}$.

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

$$w_j=d_j+d_j X'_j \tag{2}$$

The resulting weights from this computation can be negative. For most simulations having a negative weight presents a problem. In this case having negative weight would result in negative heat demands. In order to cope with the negative weights the GREGWT algorithm introduced an upper and lower bound constrains. In this case we set the lower bound constrain to 0. If the algorithm computes weights outside the defined bounds, the weights are truncated to the bounds and the algorithm will loop until a predefine convergence threshold is reached or there is no improvement in the iteration.

$$TAE_i=\sum_j |T_{i,j}-\hat{T}_{i,j}| \tag{3}$$

$$PTAE_i=TAE_i \div pop_i \times 100 \tag{4}$$

The internal validation of the simulation is quantified by means of: (a) the Total Absolute Error (TAE); and (b) the Percentage Total Absolute Error (PTAE). The TAE measures the absolute distance between the benchmarks and the marginal sums of the sample survey with the computed new weights. A TAE of zero would represent a perfect match. Equation 3 represents the TAE mathematically and Figure 1 plots the sum of benchmarks (y-axis) and the marginal sums of the sample survey (x-axis). The performance of the algorithm shows very good results, almost all small geographical areas are aligned to the 45 degree line. There are some small areas with a bigger error for which the GREGWT can not find a representative set of weight. The distribution of the PTAE error is plotted on Figure 2. This error measure is equivalent to the TAE but normalized by the total population of the small area. Equation 4 shows this normalization. The results from the reweighting algorithm show an extremely low error.

3.2 Creating a Synthetic Population with the fbs Algorithm

The reweighting of the sample survey computes non integer weights. On the next step we convert these weights to integer values. In order to create a synthetic population we implement a modified version of the fitness-based synthesis introduced by Ma and Srinivasan (2015).

In the following section we describe the implemented method for the construction of synthetic populations. The construction of synthetic populations consists of using the reweighted survey and using the estimated new weights for the construction of a synthetic population. Ma and Srinivasan (2015) proposed the computation of two fitness values expressing: (FI) the subtracting and (FII) the adding probability of individuals from the random selected population out of the initial sample survey.

The algorithm iterates until no record in the input data has positive values for either type FI (Equation 5) or type FII (Equation 6) fitness measure. R is the difference between the small area totals T and estimated totals $\hat{T}$ (with an integer weight w) for benchmark category k. Both fitness measures are computed for each individual j of the initial survey sample.

$$FIj=\sum_k R_k^2-(R_k-x_{j,k})^2 \tag{5}$$

$$FIIj=\sum_k R_k^2-(R_k+x_{j,k})^2 \tag{6}$$

Where: $R_k=T_k-\sum_j x_{j,k}\times w_j$.

On the algorithm implementation we introduced an extra constrain, the total absolute error TAE. Performing a spatial microsimulation at a low level of aggregation with just a few people on each area (10 individuals) is difficult. By using the fbs algorithm we are able to reduce the TAE achieved by GREGWT on areas with a small population. The extra constrain introduced makes sure that changing individuals results in a reduction of TAE (i.e. the change is only accepted if the achieved TAE is lower than the previous TAE value).

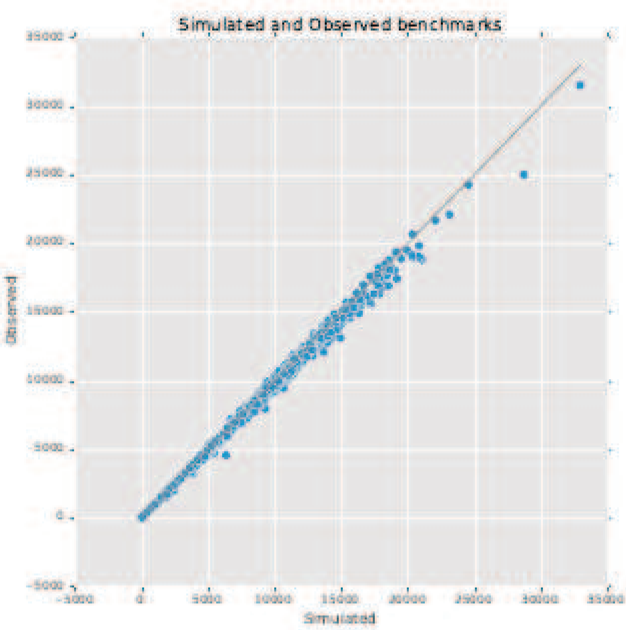

Fig. 3: Plot of simulated (x-axis) and observed (y-axis) sum of marginal totals for reweighted population with the fbs algorithm

We implement this algorithm as an addition to the GREGWT algorithm. This combination allows us to speed up the creation of synthetic families. Normally with the fbs method alone we start with a random sample of the survey. In this case, Instead of starting with a random sample of records, we start with a sample selected with the selection probability vector defined by the GREGWT computed weights. With this implementation the fbs method only needs a couple of iteration to find the best population instead of performing twice the sample size number of iterations, as reported by Ma and Srinivasan (2015).

The results from the combination of the GREGWT and the fbs method output very good results. Figure 3 shows the comparison between the small area benchmarks (y-axis) and the marginal sums of the synthetic population (x-axis). As expected, the overall performance of the model decreases. The achieved results are still good, the mean PSAE value for the synthetic population is 5.14e−1% compared to the mean PSAE value achieved by GREGWT alone of 9.93e−14%.

3.3 Computing Heat Demand

The computation of heat demand is performed for the entire input sample survey, i.e. we compute the heat demand for each individual on the sample. For the computation of heat demand we need to define the heat transmission coefficients of the building components. In order to define them we classify the survey into building types. We make use of the IWU typology (Diefenbach, Cischinsky, Rodenfels, & Clausnitzer, 2010, Loga et al., 2011) for this classification. The classification is based on the construction year, number of dwelling units and floor area. Table 2 shows the structure of the building typology.

The use of building typologies for the estimation of top down national energy models of bottom up building stock models is common practise (Caputo, Costa, & Ferrari, 2013, Hrabovszky-Horváth, Pálvölgyi, Csoknyai, & Talamon, 2013, Kragh & Wittchen, 2013, M. K. Singh, Mahapatra, & Teller, 2013, Dall'O', Galante, & Torri, 2012, Dascalaki, Droutsa, Balaras, & Kontoyiannidis, 2011, Balaras et al., 2007) among the scientific community. We make use of the R library HEAT (Muñoz H., 2015) for the computation of monthly and yearly heat demands. This library implements a quasi steady state model for the estimation of heat demand.

	<1859	1860–1918	1919–1948	1949–1957	1958–1968	1969–1978	1979–1983	1984–1994	1995–2001	2002–2009
EFHa	183	180	164	181	146	155	118	132	110	88

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

RH		153	137	156	106	127	127	98	78	86
KMH	190	143	168	156	129	134	118	122	92	79
GMH		127	144	142	131	117				
HH					114	113				

Table 2: IWU-de building typology matrix for Germany. Source: (Loga, Diefenbach, & Born, 2011) Specific Heat demand (spez. Wärmebedarfskennzahl) [kWh/m2a]. (EFH) Single family house “Einfamilienhaus”; (RH) Terrace house “Reihenhaus”; (KMH) Apartment house “Mehrfamilienhaus”; (GMH) Large apartment house “Großes Mehrfamilienhaus”; (HH) High-rise “Hochhaus”;

The library used for the computation of heat demand needs a building geometry definition for the computation of heat demand. The created synthetic building stock does not represent the buildings geometrically. In order to estimate heat demand we construct a virtual geometry based on the floor area and number of stories of the building. The computation of heat demand is performed for each individual in the survey sample. Because of this, we need to divide the estimated heat demand by the individual household size. By doing this we avoid counting the same heat demand twice. The computation of energy demand at an individual level (per capita) rather that the estimation of heat demand per building can present itself as an opportunity for the development of activity based urban models. Instead of computing the heat demand of a residential unit and of a building office we compute the energy demand at home and at the office of the same individual.

4 RESULTS & CONCLUSIONS

The presented method makes use of a spatial microsimulation model for the estimation of residential heat demand at a low level of aggregation. In order to achieve this, we compute the heat demand for each record in the national population survey sample and (a) reweight and latter (b) create a synthetic population for each geographical small area of the city of Hamburg.

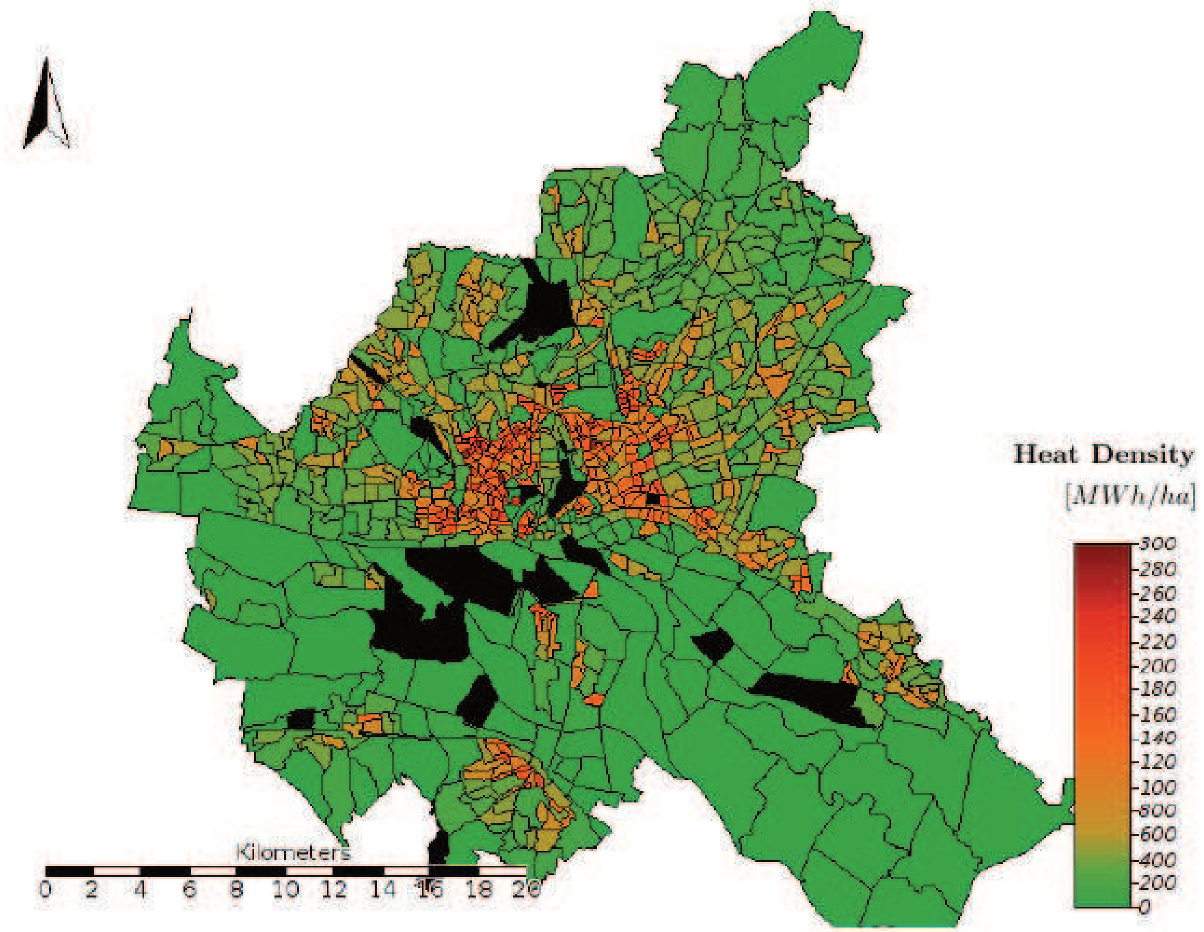

Fig. 4: Heat density for the residential sector in Hamburg

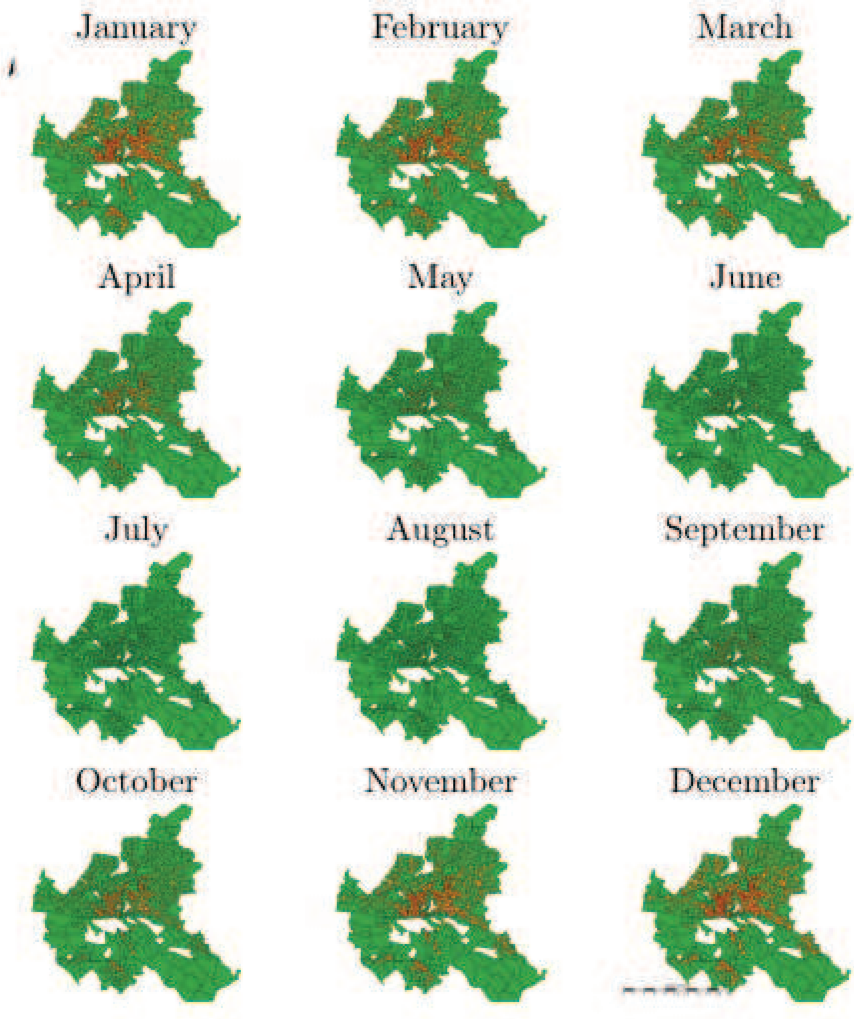

Fig. 5: Monthly heat density for the city of Hamburg

The computed heat demand is presented as heat densities of the small geographical areas. We performed the estimation of heat demand two times: (1) computing yearly heat demand and (2) a monthly heat demand. The results are presented in form of two density maps. Figure 4 shows the yearly heat density for each small geographical area in the city of Hamburg. It is clear that areas with a bigger population and with an older building stock with have a higher energy density. This model only accounts for the residential sector. A large part of the city center is covered my mix used buildings or office buildings. The heat used on this type of buildings is not taken into account on the model. The city of Hamburg also has a large industrial area, the harbor. Energy used on the industry is also not taken into account on this simulation exercise. Figure 5 shows the estimated monthly heat demand for each geographical area.

The presented method for the estimation of heat demand at a low level of aggregation shows a quick and robust performance. We have also shown the combination of two methods for the creation of synthetic populations, we take advantage of the speed provided by GREGWT to quickly compute representative weights of each small area. We use this weights to take a sample of the survey as initial population. Because this population is already close to the small area benchmarks, the fbs only needs a couple of iterations to achieve convergence. The model can be internally validated, this ensure us that the synthetic population created for each small area is a good representative of the area. Unfortunately we are still not able to make an external validation of the model.

We aim to extend the method by benchmarking the survey to aggregated energy consumption values. By benchmarking the estimated heat demands of each individual to know aggregated we can ensure that the synthetic population is not only representative by means of demographics but also by means on energy demand.

ISBN 978-3-9504173-0-2 (CD), 978-3-9504173-1-9 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER

5 REFERENCES

Balaras, C. A., Gaglia, A. G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., & Lalas, D. P. (2007). European residential buildings and empirical assessment of the hellenic building stock, energy consumption, emissions and potential energy savings. Buildingand Environment, 42 (3), 1298–1314. doi: 10.1016/j.buildenv.2005.11.001

Bell, P. (2000). GREGWT and TABLE macros — User guide [Computer software manual]. Australian Bureau of Statistics (ABS), Canberra.

Caputo, P., Costa, G., & Ferrari, S. (2013). A supporting method for defining energy strategies in the building sector at urban scale. Energy Policy, 55 , 261–270. doi:10.1016/j.enpol.2012.12.006

Clarke, M., & Holm, E. (1987). Microsimulation methods in spatial analysis and planning. Geografiska Annaler. Series B, Human Geography, 69 (2), pp. 145-164. Retrieved from http://www.jstor.org/stable/490448

Dall'O', G., Galante, A., & Torri, M. (2012). A methodology for the energy performance classification of residential building stock on an urban scale. Energy and Buildings, 48, 211–219. doi: 10.1016/j.enbuild.2012.01.034

Dascalaki, E. G., Droutsa, K. G., Balaras, C. A., & Kontoyiannidis, S. (2011). Building typologies as a tool for assessing the energy performance of residential buildings – a case study for the hellenic building stock. Energy and Buildings, 43 (12), 3400–3409.doi: 10.1016/j.enbuild.2011.09.002

Diefenbach, N., Cischinsky, H., Rodenfels, M., & Clausnitzer, K.-D. (2010). Daten-basis gebäudebestand: Datenerhebung zur energetischen qualität und zu den modernisierungstrends im deutschen wohngebäudebestand (1. ed.). Darmstadt: Institut Wohnen und Umwelt (IWU) and Bremer Energie Institut (BEI).

Hrabovszky-Horváth, S., Pálvölgyi, T., Csoknyai, T., & Talamon, A. (2013). Generalized residential building typology for urban climate change mitigation and adaptation strategies: The case of Hungary. Energy and Buildings, 62 , 475–485. doi: 10.1016/j.enbuild.2013.03.011

Kragh, J., & Wittchen, K. (2013). Development of two danish building typologies for residential buildings. Energy and Buildings, 68, Part A, 79 - 86. doi: 10.1016/j.enbuild.2013.04.028

Loga, T., Diefenbach, N., & Born, R. (2011). Deutsche Gebäudetypologie: Beispielhafte Maßnahmen zur Verbesserung der Energieeffizienz von typischen Wohngebäuden.

Ma, L., & Srinivasan, S. (2015). Synthetic population generation with multilevel controls: A fitness-based synthesis approach and validations. Computer-Aided Civil and Infrastructure Engineering, 30 (2), 135–150. doi: 10.1111/mice.12085

Muñoz H., M. E. (2015). heat: R package to estimate heat demand of residential buildings. github.com/emunozh/heat. Retrieved from github.com/emunozh/heat

Muñoz H., M. E., Seller, H., & Peters, I. (2015). Constructing a synthetic city for estimating spatially disaggregated heat demand. In 5th world congress of the international microsimulation association (IMA).

Muñoz H., M. E., Tanton, R., & Vidattama, Y. (2015). A comparison of the GREGWT and IPF methods for the re-weighting of surveys. In 5th world congress of the international microsimulation association (IMA).

Muñoz H., M. E., Vidyattama, Y., & Tanton, R. (2015). GREGWT: an implementation of the GREGWT algorithm in R. github.com/emunozh/gregwt. Retrieved from github.com/emunozh/GREGWT

O'Donoghue, C., Morrissey, K., & Lennon, J. (2014). Spatial microsimulation modelling: a review of applications and methodological choices. International Journal of Microsimulation, 7 (1), 26-75.

Orcutt, G. H. (1957). A new type of socio-economic system. Review of Economics and Statistics, 39 (2), 116–123.

Rahman, A., Harding, A., Tanton, R., & Shuangzhe, L. (2010). Methodological issues in spatial microsimulation modelling for small area estimation. International Journal of Microsimulation, 3 (2), 3–22.

Singh, A., & Mohl, C. (1996). Understanding calibration estimators in survey sampling. Survey Methodology, 22 , 107–115.

Singh, M. K., Mahapatra, S., & Teller, J. (2013). An analysis on energy efficiency initiatives in the building stock of Liege, Belgium. Energy Policy, 62 , 729–741. doi:10.1016/j.enpol.2013.07.138

Statistische Ämter des Bundes und der Länder. (2010). Forschungsdatenzentren der statistischen Ämter des bundes und der länder — Mikrozensus (EVAS 12211) — 2010 Scientific-Use-File. Retrieved 20.11.2015, from http://www.forschungsdatenzentrum.de/bestand/mikrozensus/