Validation of building stock heat demand estimation using consumption data under GDPR

Aktive Endkunden-/Prosumerpartizipation & Gebäudesektor

Mostafa Fallahnejad, Aadit Malla, Lukas Kranzl
Energy Economics Group – TU Wien

Motivation and Research Question

Bottom-up methods use various data on building use, construction period, renovation status of the building, heating system, footprint, and heated gross floor area to estimate the heat demand of buildings. Once the estimation is done, the plausibility and validity of the results should be checked. For the plausibility check, often it is checked if the obtained indicators are in a reasonable range and if there is any strange behaviour in the output data. The values are also compared with other sources and energy balances. In the case of the validity checks, however, the obtained results should be compared to consumption data to have a robust basis.
Directive 2007/2/EC elaborates on establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). The INSPIRE directive aims to create a spatial data infrastructure within the European Union (EU). This should facilitate the exchange of spatially relevant information between organizations in the public sector and public access to spatial data across Europe. Non-personal geodata falls into this category. In contrast, combining those data with personal data or civil register falls under General Data Protection Regulation (GDPR). For example, accessing energy consumption data of residents of buildings for third parties is forbidden.
The problem is, however, that the consumption data cannot be shared due to the GDPR restrictions. As a result, the validation of the building stock heat demand cannot be easily done and requires cooperation between DSOs and third parties. This paper proposes an approach for performing validity studies.

Method

The heat demand estimation used in this work originates from Spatial Energy Plan (SEP) project[1], funded by Green Energy Lab, and covers buildings in Vienna. We focus on buildings with single-use (no mixed-use) and single energy carrier usage (e.g. Natural gas). In the first step, the goal is to select a subset of data sets representing all buildings in Vienna. Table 1 shows the parameters used to categorize the buildings along with their unique cases.

Table 1 Parameters for categorization of samples

PARAMETER	UNIQUE CASES
CONSTRUCTION PERIODS	10
END-USE	14
RENOVATION STATUS	3
FOOTPRINT SIZE	5
TOTAL COMBINATIONS	2100

Although heated gross floor area is also a relevant parameter for the categorization of samples, including it could lead to having a high number of categories. Subsequently, the analysis of the categories would be more difficult. It is also not a good practice to substitute footprint with the heated gross floor area, as a lower footprint area with a higher number of floors could be equivalent to a higher footprint area and a lower number of floors. Whereas footprint can categorize buildings better.

Out of the possible 2100 combinations provided in the table above:

- Only 318 combinations have more than 5 samples;
- Only 219 combinations have more than 10 samples;
- Only 76 combinations have more than 100 samples.

Certain combinations do not exist among the buildings in Vienna. Additionally, the suggested granularity is high for a few other categories. These categories account for only 71 entries in total and due to the scarcity of the buildings in these categories and the insignificance of their heat demand,

[1] http://www.waermeplanung.at/

they were excluded from the remaining steps. Two sample subsets of five buildings were built for each remaining category. The sample subsets should be representative of their corresponding category.

The categorization of samples does not include the heated gross floor area. However, this parameter affects the overall heat demand of a building. To classify the buildings within a sample set, the Jenks natural break algorithm is applied to the heated gross floor area. The Jenks natural break classification algorithm assigns samples within a category to a pre-defined number of groups so that each group's average deviation from the class mean is minimized while each group's deviation from the means of the other groups is maximized. Accordingly, the creation of a sample sub-set is done as follows:

- Sort all samples in the category based on the "heated gross floor area" descendingly.
- Use Jenks Natural Break to allocate the samples into two to four different gross floor area classes (depending on the overall number of samples in a category and their distribution) and label them with: GFA1, GFA2, GFA3 and GFA4
- Sample subset creation:
    - Subset "A":
        - Find the label with the largest size.
        - Select the class median and its 4 immediate neighbours
            - If this class has less than five samples, the remaining samples will be selected from the next largest label class.
    - Subset "B":
        - Select the median of each class.
        - From the largest class, two samples will be selected.

The procedure is illustrated in Fig. 1.

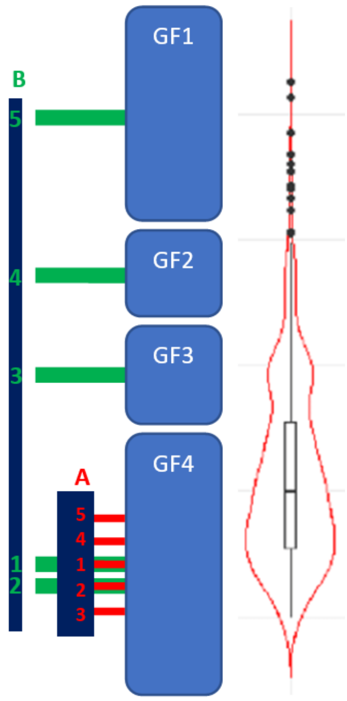

Figure 1 Selection of subset

Results and conclusions

Two tables are prepared and sent to the DSO for further evaluation. The first table shows the estimated demand on the building level. If the deviation of the estimated demand from the consumption value is greater than 50%, the sample should be replaced with a new one. This is because some buildings are not occupied the whole year or may end their contract with the DSO. In the second table, the aggregated demand values as well as consumption data (climate corrected) from different years, are provided. Considering the aggregation, the GDPR is respected. Furthermore, since the buildings in each category have similar characteristics, the deviations can be used to correct specific demands in the category. The sample subset A and B help to estimate the level of mismatches. While sample A is more representative of each category, sample B demonstrates the impact of outliers within each category.

Table 2 Sample IDs and their estimated demand

Sample ID	Subset	ID-SEP	ID-DSO	Demand [MWh]
1	A	SEP_ID0	DSO_ID0	D0
1	A	SEP_ID1	DSO_ID1	D1
1	A	SEP_ID2	DSO_ID2	D2
…	…	…	…	…

Table 3 Aggregated demand and consumption data for each sample subset

Sample ID	Subset	Aggregated demand (SEP) [MWh]	Consumption [MWh]						
			2015	2016	2017	2018	2019	2020	2021
1	A	X	X1	X2	X3	X4	X5	X6	X7
1	B	Y	Y1	Y2	Y3	Y4	Y5	Y6	Y7
…	…	…	…	…	…	…	…	…	…