



HAL
open science

An address-matching algorithm for household-scale databases to enhance electricity demand characterization

Antoine Rogeau, Robin Girard, Georges Kariniotakis, Nicolas Kong

► To cite this version:

Antoine Rogeau, Robin Girard, Georges Kariniotakis, Nicolas Kong. An address-matching algorithm for household-scale databases to enhance electricity demand characterization. 11th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2018), Nov 2018, Dubrovnik (Cavtat), Croatia. hal-01948608

HAL Id: hal-01948608

<https://hal.science/hal-01948608>

Submitted on 7 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An address-matching algorithm for household-scale databases to enhance electricity demand characterization

Antoine Rogeau, Robin Girard, Georges Kariniotakis
Centre for Processes, Renewable Energy and Energy Systems (PERSEE)
MINES ParisTech, PSL University
Sophia Antipolis, France
Email: antoine.rogeau@mines-paristech.fr

Nicolas Kong
ENEDIS
Paris, France
Email: nicolas.kong@enedis.fr

Abstract—This paper participates in the challenging data science opportunity offered by the growing number of databases made available to public institutions. It presents an innovative method to match household-scale databases using address information. The developed algorithm authorizes different matching qualities, depending on the reliability of the link between the paired elements. This work was carried out in collaboration with the French DSO Enedis, which provided valuable customer information that was matched with a national database describing dwellings. The matching algorithm performances are analyzed, and adjustments are proposed to improve the matching quality in urban, suburban and rural contexts. Lastly, two basic characterization analyses were made to highlight the potential of these consolidated databases.

Index Terms—database matching, address matching, building, electricity consumption, characterization, disaggregation

I. INTRODUCTION

In response to growing environmental concerns, public organizations are developing action plans to tackle current trends and initiate a global energy transition. The latter covers several fields, including renewable energy integration, migration from thermal to cleaner transportation systems, power systems operation optimization using smart-grid technologies, and reduction of energy consumption by improving energy efficiency. In order to encourage and facilitate initiatives fostering a cleaner and more sustainable future, a number of organizations, encouraged by legislation, have built energy-related databases, made available to different actors. These databases contain valuable information of interest for carrying out studies and broadening collective knowledge of the energy field.

In addition, combining these databases can often lead to deeper and more complex analyses. To emphasize this point, this paper considers one of the main fields involved in energy transition, namely energy efficiency, through a reduction in energy demand.

With the aim of broadening knowledge of the factors affecting energy consumption, governments and academics carry out characterization studies to understand end-user behavior. Jones et al. [1] reviews studies on this topic, and identifies factors recognized as having an impact on electricity consumption. Huebner et al. [2] quantifies the correlations between

household characteristics and annual electricity demand. The majority of existing works use data harvested from surveys, and so are carried out using a restricted amount of data (less than 1,000 households). This kind of data source, while highly reliable, is difficult to obtain due to the cumbersome collection procedure. To work on larger samples requires taking advantage of the existence of large energy-related databases.

Consequently, this paper introduces the basics of an electricity consumption characterization method based on an individual annual consumption database provided by the French DSO Enedis and a building database covering the whole country. One of the main challenges is to identify the most appropriate way to combine heterogeneous sources of data to obtain a reliable and unified database for characterizing consumption as finely as possible, i.e. at household scale.

The first part of this paper describes the matching algorithm developed to combine two dwelling-scale databases containing address information. Then, the second part presents the results achieved by the matching algorithm on different test datasets. Finally, the third part of this article illustrates the work that can be carried out using data obtained from the matching algorithm, presenting two electricity demand characterization analyses.

II. DESCRIPTION OF THE MATCHING ALGORITHM

Combining information from heterogeneous sources has been a major technical concern for decades. To do so, equivalent information must be identified from both sources, so that the best combination can be found. This task is called the field matching problem and basic approaches are presented by Monge and Elkan [3]. These authors use different matching algorithms relying on basic strings correspondence ratios, i.e. a proportion of the words matched in the two compared strings, or total string distances. Distances between entire address labels have also been used by Dengyue Li [4] while some approaches have split address labels into relevant components (number, street, etc.) to establish a matching ranking, such as presented by Brummet [5]. The three methods mentioned above aim at matching a database to a unique reference field set, e.g. for geocoding purposes, and do not handle multiple matches. To work with consumption and dwelling data at

a disaggregated level, as intended in this paper, specifically requires matching the correct items, even when they share a common address. The developed algorithm considers this issue.

The matching algorithm presented in this paper comprises two steps: the preliminary preparation of the data, and the so-called matching.

A. Preliminary preparation

As expressed in several articles, to achieve reliable address matching, street names must first be cleaned and standardized [3, 4, 5, 6]. Stop words, extra spaces and symbols are removed [3]. Also, recurrent words with different possible orthography are replaced by a standardized equivalent in both data sets using a dictionary (e.g. "AVENUE", "AVE", etc. to "AV").

This matching should be carried out on large data sets (several thousands of elements). For efficient and fast operation, a loop structure for an iterative process should be avoided; the present method proposes to pre-process addresses to generate factor matrices. Address information is split into three categories, corresponding to the city code (CITY), street number (NUMBER) and street name. The latter is sub-split in two, namely the last word of the street name (LAST), and the rest of the street name (STREET). An example of a split is presented in Table I.

TABLE I
EXAMPLE OF SPLIT FOR THE ADDRESS: "1 RUE CLAUDE DAUNESSE, 06152 SOPHIA ANTIPOLIS".

Description	Name	Value
Code of the city	CITY	06152
Formatted first part of the street name	STREET	R CLAUDE
Last word of the street name	LAST	DAUNESSE
Street number	NUMBER	1

This approach, and particularly the split of the street name into two parts, was inspired by the availability of the FANTOIR database in France¹. This file describes all street names per city using a key called FANTOIR, which corresponds to the first eight letters of the last word of the street names. Unlike standard matching methods, where one threshold defining the matching quality has to be set for the whole process, this structure allows to work at different quality levels of matching within the same procedure, from high quality (all factors matched) to low quality (e.g. only CITY and LAST) and with no deterioration from one to the other. Indeed, splitting information and working with factors allows to rapidly build a tree structure (see Fig. 1), achieving different levels of appropriate matching depending on the factors selected. If no database such as FANTOIR is available or can be built, it is also possible to work on whole street names.

The distance between LAST words is calculated using the Optimal String Alignment (OSA), also known as the Damerau-Levenshtein distance. A high similarity is needed between LAST words, allowing a maximum distance of 1 character.

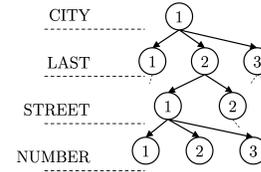


Fig. 1. Tree structure of the data used for the matching - selecting all factors

This is one of the advantages of working with the last word instead of the entire street label. To validate the STREET matching criteria, one common word has to be found in the rest of the street label (if there is only one word in the label, then this criteria is validated). NUMBER and CITY have to match perfectly.

B. Matching of addresses

After this preliminary preparation is performed, the next step is the application of the proposed matching algorithm. The structure of the algorithm is presented in Fig. 2.

- ① Perfect Matching (PM): the first step is to attribute the uniqueness of the CITY, LAST, STREET and NUMBER match, before processing data in which uniqueness or number match are not achieved
- ② Multiple matching at the same NUMBER (NM): a sampling allows to perform pairings; any remaining elements are kept for the next step. The same factors are considered as for the previous step, but uniqueness is no longer required.
- ③ Matching the same STREET (SM): use of the same sampling process.
- ④ Matching the same LAST word (LM): use of the same sampling process.
- ⑤ Some data cannot be matched (NoM).

Both SM and LM present a strong deterioration of the information: the numbers do not correspond between the two paired elements, and the only certainty is that the elements pertain to the same geographical element (street, avenue etc.). Consequently, these matches are only valuable for working at an aggregated level.

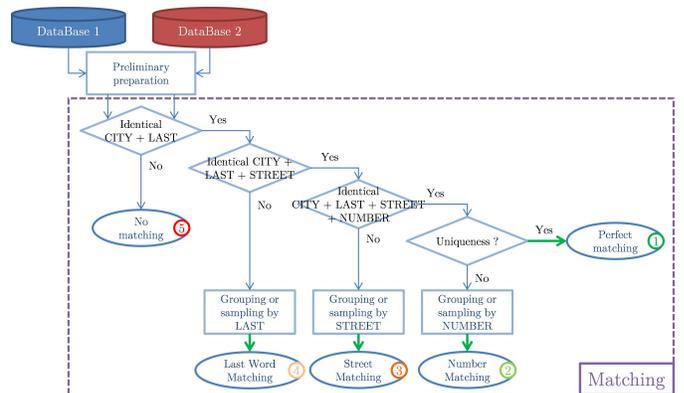


Fig. 2. Structure of the matching algorithm

¹<https://www.data.gouv.fr/fr/datasets/fichier-fantoir-des-voies-et-lieux-dits/>

III. PERFORMANCES OF THE MATCHING ALGORITHM AND IMPROVEMENTS

The above algorithm was employed to match two databases: On the one hand, customer information, containing mainly the contract power, yearly consumption and geographical address, was provided by the main French DSO, Enedis. On the other hand, structural information such as the dwelling area, year of construction, number of rooms and building material, was gathered from national dwelling descriptions, provided by the French Public Finances Directorate. Areas of different sizes were tested to validate the matching algorithm, but also in perspective to the electricity consumption characterization.

A. Performance in an urban area. Study of a city: Brest

The matching algorithm was run first for the city of Brest, France, considering 96,691 clients for a process time of about 400 seconds. Results are presented in Table II.

TABLE II
OCCURRENCES OF QUALITY OF MATCHING, APPLICATION TO BREST.

Step	Quality	Occurrences
1	PM	17,770
2	NM	59,733
3	SM	4,924
4	LM	2,222
5	NoM	12,142

Note that only about 20% of the data match perfectly. Although this figure may seem low at first sight, it is perfectly understandable when considering the structure of Brest, an urban area in which many dwellings are part of collective buildings. Consequently, several elements share the same NUMBER label and cannot be uniquely matched. This observation motivated a deeper analysis of the influence of the study area structure on the quality of matching. For this reason, the next section presents the results of the matching algorithm when applied to a whole county.

B. Performances on different city types. Study of a region: Rhone

The matching process was carried out on the entire French region of Rhone, which represents more than one million French DSO Enedis clients. The results of the matching are presented in Table III.

TABLE III
OCCURRENCES OF MATCHING QUALITY, APPLICATION TO RHONE REGION.

Step	Precision	Occurrences
1	PM	110,723
2	NM	555,663
3	SM	134,487
4	LM	57,799
5	NoM	204,994

While the PM proportion now drops to about 10%, an observation of matching type distributions in the different cities within this region (291 in total) reveals some of the limits of the developed algorithm. Fig. 3 compares the proportions

of MP, NM and SM depending on the number of dwellings in the city.

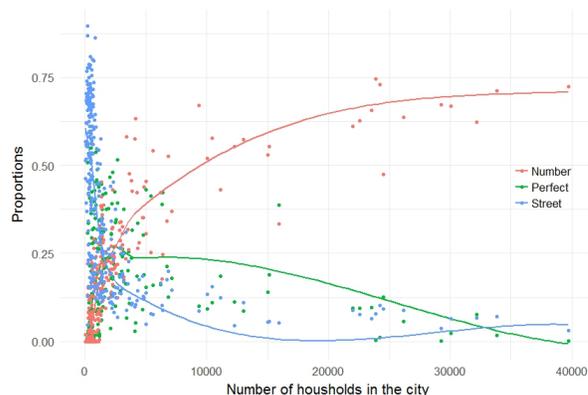


Fig. 3. Evolution of matching proportions depending on the number of dwellings in the city

One first observation is that very small cities present a majority of SMs. This can be explained by the fact that in very rural areas, addresses have no number but only a place name (called *lieu-dit* in France), and several dwellings often share the same address. Fig. 4 considers this *lieu-dit* specificity. Cities with a significant proportion of each type of matching (> 30% of the maximum observed proportion) are plotted against the proportion of *lieu-dit* on the x-axis and the proportion Houses vs. Apartment on the y-axis.

Three clear groups appear, from which the following archetypes can be drawn:

- **Urban context:** few houses and *lieu-dit*, present a high proportion of NMs.
- **Suburban context:** majority of houses and few *lieu-dit*, present the highest proportion of PMs.
- **Rural context:** majority of houses and *lieu-dit*, the SMs predominate.

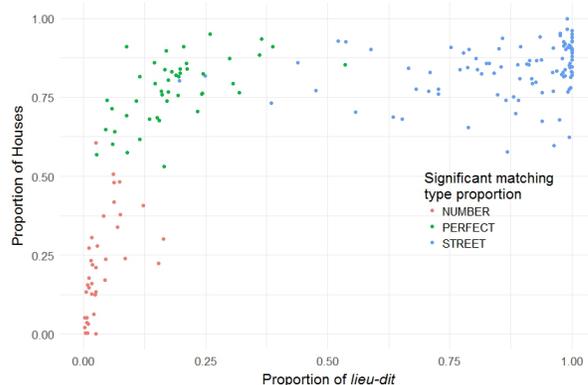


Fig. 4. Representation of the most significant matching type, i.e. proportion in city of over 30% of the maximum observed proportion, depending on the house proportion and the *lieu-dit* proportion in each city

To carry out an individual energy consumption analysis requires PMs, whereby the correspondence between the building

characteristics and the customer information is certain. The matching algorithm thus has to be improved for both rural and urban contexts. The problems confronted are radically different for both contexts and require a specific solution:

- **Urban context:** For areas where apartments are dominant and thus NMs are common, a grouping of dwellings can be applied at building level. This approach is specifically relevant when working on energy consumption characterization, as building characteristics are the same for all dwellings in a building. This leads to an equivalent client with a new equivalent dwelling corresponding to the sum of all dwellings (i.e. in terms of area, number of rooms, etc.).
- **Rural context:** For areas where several dwellings are attached to the same address, an aggregation is insufficient to carry out a reliable energy analysis, as the building characteristics of each dwelling may differ. Using geographic information such as coordinates could help to identify the right match, by running a least-distance geographic matching within each group.

The results of the aggregation at building level for the city of Brest are presented in Table IV. Since no elements are rejected by a sampling during the aggregation, the number of low-quality matches occurrences decreases.

TABLE IV
OCCURRENCES OF QUALITY OF MATCHING WITH BUILDING
AGGREGATION, APPLICATION TO BREST.

Step	Quality	Aggregation	Occurrences
1	PM	No	17,770
2	NM	Yes	67,953 (8,265 buildings)
2	NM	No	343
3	SM	No	2,288
4	LM	No	2,071
5	NoM	No	6,266

IV. STUDY CASE: USAGE OF MATCHED DATA TO CHARACTERIZE ENERGY DEMAND

This section presents the studies that can be carried out thanks to database consolidation. Over recent decades, the academic field has developed a growing interest in understanding residential energy demand with the aim of reducing end-use demand. Indeed, energy efficiency is a key to tackle global warming and the environmental crisis. Bottom-up approaches are particularly well-suited to identify the factors impacting residential consumption, and two main categories of models are developed: (1) Engineering models use appliance and building physics-based models to describe the consumer's behaviour and predict the consumption while (2) statistical models rely on the analysis of observation data using Machine Learning methods to describe consumer's energy demand [7, 8]. Two main approaches can be observed: first, qualitative approaches attempt to identify the most impacting factors on energy consumption, without trying to identify the actual link between cause and effect [9, 10]. On the other hand, quantitative approaches do try to measure this link using linear

regression models [11] or even by training Machine Learning algorithms (Neural Network, Random Forest) [12].

In this article, both basic quantitative and qualitative approaches are drawn, showing how database matching can be used to extract relevant information, which is not visible when the databases are split. An Odds-Ratio analysis is presented to identify the factors impacting electricity demand, then carry out a linear regression fit to quantify the impact of the dwelling area on energy consumption.

A. Odds-ratio analysis

This analysis is based on the work presented by Jones and Lomas [9]. It consists in splitting the consolidated consumer-dwelling database into two subsets depending on energy consumption. The first one is constituted by low-energy consumers, while high-energy consumers are stored in another subset. In this electricity consumption context, the Odds-Ratio approach estimates the odds that a client will be a high consumer (resp. a low consumer) given a certain characteristic of his dwelling, compared to the odds of being a high consumer (resp. a low consumer) if his dwelling features the reference characteristics. Reference refers to the most frequent characteristic in the sample. This analysis was applied to the city of Brest: Among the 17,779 PMs, the work was led on the 5,476 clients with annual energy consumption information provided. The resulting table is presented in Table V. Each line represents a category for a given parameter (e.g. floor area), the number of elements falling into this category, and the number of elements pertaining to the high and low energy consumer subsets. The Odds-Ratio is calculated as follows:

$$OR = \frac{S_H^{100+}/S_L^{100+}}{S_H^{REF}/S_L^{REF}} = \frac{976/1553}{858/2089} = 1.53 \quad (1)$$

where: OR = odds ratio; S_H^{100+} = number of dwellings with area greater than $100m^2$ and high electricity consumption; S_L^{100+} = number of dwellings with area greater than $100m^2$ and low electricity consumption; S_H^{REF} = number of dwellings with reference area and high electricity consumption; S_L^{REF} = number of dwellings with reference area and low electricity consumption.

As expected, the probability of being a high-energy consumer appears to grow with the size of the dwelling, but also with the number of bedrooms and rooms. It also emerges that owners tend to consume more than tenants. One last observation concerns the building material, whereby stone and gritstone seem to give a better energy performance than brick or concrete for example. Nevertheless, it has to be considered that construction methods have evolved over the years, along with heating systems. Consequently, heating fuels and thermal standards are strongly correlated, and no major conclusion can be made on the latter category.

B. Characterization of individual electricity demand

The objective of the characterization of individual electricity demand is to establish a link between dwelling characteristics

TABLE V
TABLE OF ODDS-RATIOS FOR PMs IN THE CITY OF BREST, FRANCE

	Category	Occurrences	Number of dwellings		Odds Ratio
			High consumers	Low consumers	
Floor Area	>100m ²	2529	976	1553	1.53
	>50m ²	2947	858	2089	REF
	<50m ²	58	12	46	0.64
Insulation Standard	None	3065	674	2391	REF
	1974	1179	657	522	4.46
	1988	617	241	376	2.27
	2000	352	100	252	1.41
	2005	321	174	147	4.20
Number of Bedrooms	5+	709	288	421	1.31
	4	2287	784	1503	REF
	3	1938	614	1324	0.89
	2	498	134	364	0.71
	1	88	24	64	0.72
	0	11	1	10	0.19
Number of Rooms	5+	5463	1826	3637	REF
	4	54	16	38	0.84
	3	10	2	8	0.50
	2	3	1	2	1.00
Occupation	Owner	4739	1636	3103	REF
	Tenant	654	142	512	0.53
Walls Materials	Stone	742	199	543	0.84
	Gritstone	1656	379	1277	0.68
	Concrete	852	404	448	2.07
	Brick	948	359	589	1.40
	Chilboard	2502	760	1742	REF
	Wood	64	32	32	2.29
	Other	117	51	66	1.77

and electricity consumption through simple models (e.g. linear regressions).

PMs and NMs aggregated at building level can be considered as highly reliable and be used for single building consumption characterization. Nevertheless, this work focuses on PMs with consumption information provided, for the city of Brest. In order to guarantee robust interpretability, linear regression models are used to quantify the impact of dwelling characteristics on electricity consumption.

A reference cross-validated regression model, applied to match the consumption of a household with its dwelling surface, shows a very low r-squared ($R^2 = 0.055$) and cannot be considered as relevant. Data is then clustered into Data Groups (DGs), built according to the presence of electric Domestic Hot Water (DHW), Electric Heating (EH), age of building, etc.

Information on the presence of electric appliances is not available in the databases and so an algorithm has been implemented to infer their presence in dwellings, based on the households floor area and the energy contract. Table VI transcripts this simple algorithm. DHW and EH are dummy variables, where 1 represents the presence of DHW and EH, and 0 that the fuel is inferred to be non-electric (gas, thermal, etc.), depending on the dwelling type, floor area, energy contract and type of contract.

TABLE VI
REPRESENTATION OF THE DHW AND EH ATTRIBUTION ALGORITHM.

	Floor area	Contract power	Domestic Hot Water		Electric Heating
			Contract type		
			Peak/Off-Peak	Base	
House	<70m ²	3kVA	1	0	0
		>3kVA	1	1	1
	>70m ²	6kVA	1	0	0
		>6kVA	1	1	1
Apartment	<100m ²	6kVA	1	0	0
		>6kVA	1	1	1
	>100m ²	9kVA	1	0	0
		>9kVA	1	1	1

The resulting model, when considering two parameters DG_1 and DG_2 , is described by Equation 2.

$$C_p(\text{Surf} | DG_1 \times DG_2) = \sum_i^{DG_1} \sum_j^{DG_2} (\alpha_{ij} \times A_{ij,p} + \beta_{ij}) \quad (2)$$

where C_p is the consumption of the customer p , i (resp. j) is the group index built on the parameter DG_1 (resp. DG_2), $A_{ij,p}$ is the area of the dwelling matched to the client p for i and j corresponding to this dwelling, and zero otherwise.

Building DGs from the presence or absence of DHW and EH, the r-squared is higher ($R^2 = 0.478$) and the accuracy of the consumption estimates increases by 44.56% compared to the reference. Equations (3) to (5) describe the link detected between electricity consumption C and dwelling Area A by regression models, which are plotted on Fig. 5.

$$C_{EH,DHW} = 173,62 \times A_{EH,DHW} + 8.553,23 \quad (3)$$

$$C_{\overline{EH},DHW} = 133,49 \times A_{\overline{EH},DHW} + 5.354,86 \quad (4)$$

$$C_{\overline{EH},\overline{DHW}} = 49,89 \times A_{\overline{EH},\overline{DHW}} + 3.820,47 \quad (5)$$

In France, regulations define the insulation standards to be respected when constructing a new building. This legislation dates from 1974 and standards were upgraded in 1988, 2000, 2005 and 2012. When constructing smaller DGs depending upon both the insulation standard and installed fuel, it is possible to improve the electricity consumption estimation, reaching an r-squared slightly higher than 0.5 (0.525). In the knowledge that the presence of Electric Heating and electric Domestic Hot Water has a strong impact on electricity consumption, note that the attribution of such devices was made using a very basic algorithm which prevents from making any definitive conclusions. It is also possible to observe a strong variability around the linear fit, which can probably be explained by other factors, e.g. structural or socio-economic. The lack of reliable data and other relevant databases in this domain makes it impossible to take this analysis further.

V. CONCLUSIONS AND PERSPECTIVES

This work presents an innovative method aiming to take advantage of new trends in energy, i.e. Open Data and Big

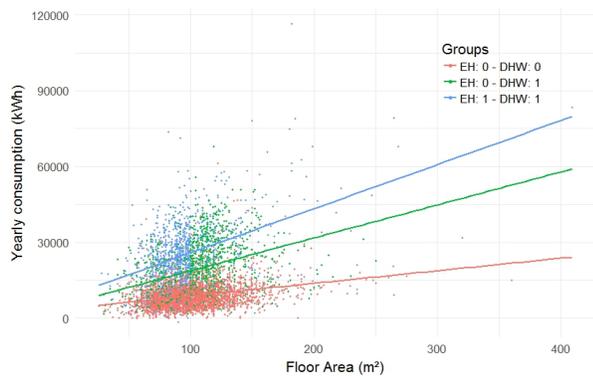


Fig. 5. Fitting of electricity consumption when Electric Heating and Sanitary Hot Water are used for groups

Data. A generic algorithm is presented for matching databases comprising address information, with the aim of working at dwelling, building or higher aggregation levels. The results of two algorithm runs were presented and allowed to identify shortcomings depending on the area. Improvements were thus implemented to permit a better matching in urban and suburban areas.

Two studies were carried out regarding the best quality matches obtained for the city of Brest to characterize end-users electricity demand. A first approach used the Odds-Ratio analysis to identify the key factor affecting electricity consumption, while a second study presented a linear regression model to estimate the impact of the dwelling area on annual energy consumption via simple linear regressions with groups. Both studies show the potential of such matching techniques in the energy domain. In addition, they highlight the fact that the reliability of data is another crucial issue to deal with.

Matching with other energetic vectors (i.e. gas) is easily achievable, and some, such as district heating, could facilitate understanding given the absence of other usages to perturb the analysis. This work can also be extended to aggregated levels, as LV or MV feeders where load curves are available, to characterize peak demand and thermo-sensitivity. At such level, structural information could also be combined with non-address based matching algorithms, e.g. optimization algorithm for adding socio-economic information [13].

This approach is expected to be a useful tool for territorial authorities and actors developing policies and actions to facilitate the energy transition at the level of a territory, as it will enable them to estimate the consumption of their territory, identify the most energy-consuming dwellings and households, and encourage them to take action to reduce their energy consumption (i.e. through thermal renovation).

ACKNOWLEDGEMENT

The authors would like to thank the French DSO Enedis for providing some very valuable customer data that allowed them to develop an efficient matching algorithm and undertake the characterization work presented in this paper.

REFERENCES

- [1] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 901–917, mar 2015.
- [2] G. Huebner, D. Shipworth, I. Hamilton, Z. Chalabi, and T. Oreszczyn, "Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes," *Applied Energy*, vol. 177, pp. 692–702, sep 2016.
- [3] A. E. Monge and C. P. Elkan, "The field matching problem: Algorithms and applications," in *Second International Conference on Knowledge Discovery & Data Mining*, 1996.
- [4] Z. M. Dengyue Li, Shengrui Wang, "Approximate address matching," in *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2010.
- [5] Q. Brummet, "Matching addresses between household surveys and commercial data," *CARRA Working Paper Series*, 2015.
- [6] A. V. S. Bianca Salas and E. Shuster, "Fun with address matching: Use of the compged function and the sql procedure," in *SAS Global Forum 2018*, 2018.
- [7] N. Fumo, "A review on the basics of building energy estimation," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 53–60, mar 2014.
- [8] M. Brøgger and K. B. Wittchen, "Estimating the energy-saving potential in national building stocks – a methodology review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1489–1496, feb 2018.
- [9] R. V. Jones and K. J. Lomas, "Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics," *Energy and Buildings*, vol. 101, pp. 24–34, aug 2015.
- [10] C. Bartusch, M. Odlare, F. Wallin, and L. Wester, "Exploring variance in residential electricity consumption: Household features and building properties," *Applied Energy*, vol. 92, pp. 637–643, apr 2012.
- [11] I. G. Hamilton, P. J. Steadman, H. Bruhns, A. J. Summerfield, and R. Lowe, "Energy efficiency in the british housing stock: Energy demand and the homes energy efficiency database," *Energy Policy*, vol. 60, pp. 462 – 480, 2013.
- [12] D. Brounen, N. Kok, and J. M. Quigley, "Residential energy use and conservation: Economics and demographics," *European Economic Review*, vol. 56, no. 5, pp. 931 – 945, 2012, green Building, the Economy, and Public Policy.
- [13] T. Barbier, R. Girard, N. Kong, F.-P. Neirac, G. Kariniotakis, and E. Magliaro, "An optimisation algorithm for matching large scale databases on customers for improved characterisation of electricity consumption," *IET Conference Proceedings*, January 2016.