Tertiary building stock modeling: Area determination by fusion of different datasets Tertiary building stock modeling: Area determination by fusion of different datasets

Huu Tam Nguyen, Fabrice Decellas, Bruno Duplessis, Philippe Rivière, Dominique Marchio

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Tertiary building stock modeling: Area determination by fusion of different datasets

To cite this article: Huu Tam Nguyen *et al* 2019 *J. Phys.: Conf. Ser.* **1343** 012017

Tertiary building stock modeling: Area determination by fusion of different datasets

Huu Tam Nguyen*[1,2], Fabrice Decellas[1], Bruno Duplessis[2], Philippe Rivière[2], Dominique Marchio[2]

[1] EDF R&D Lab les Renardières, Avenue des Renardières, 77250 Écuelles, France

[2]Mines Paristech – Paris Sciences et Lettres (PSL), 60 Boulevard Saint-Michel, 75006 Paris, France

*huu-tam.nguyen@mines-paristech.fr

Abstract. The building stock accounts for 43% of the total French final energy consumption in 2017. The French building stock mainly includes residential and tertiary buildings, which have very different consumption patterns. Today, building stock models are largely used in many countries for energy demand forecast and decision making. In this context, good knowledge of the building stock is essential for modeling. However, this type of model is lacking for the tertiary sector, as it is difficult to categorize buildings that have a wide variety of area and energy uses, different systems and complex dynamic trends. Existing stock models are mostly based on an aggregated representation of the total heated floor area of the building stock. Such a representation limits explicit studies of the factors affecting the energy demand and fails at taking into account heterogeneity. This paper exposes and discusses new methods aiming at modeling the tertiary building stock at the city scale in France. While mainly relying on open source public data, methods are designed to be applicable at any scale (from small cities to large region or even national scale).

1. Introduction

For many years, building energy modeling and simulation have been largely used in advising decision makers in the domain of energy policy [1]. The building stock accounts for 43% of the total French final energy consumption in 2017 [2]. The tertiary sector is responsible for one third of the total building stock's consumption. Whereas considerable research has been devoted to the residential building stock, less attention has been paid to the tertiary building stock. Heterogeneity is the main reason that makes it difficult to study the tertiary sector. The representation of the tertiary building stock is usually based on the total heated floor area of economic sub-sectors, aggregated at national scale, omitting heterogeneity. Through this work, we develop a building stock model which allows a detailed description of this sector's building stock. The purpose of this model is to study the energy demand as well as long-term projections of the tertiary stock. Recently, detailed individual building and building stock models have begun to merge into hybrid methods that might help evaluating the energy demand of a group of buildings, which may contain from dozens to thousands individuals [3]. Our ambition is to represent the tertiary building stock as a group of individual buildings. The first important step is to determine the area of each building.

Many building stock models use the area as a key variable for estimating the energy demand. For example in the econometric model NEMS - National Energy Modeling System, the US tertiary building stock is described in its Commercial Demand Module [4]. In NEMS, the floor area is used in many of the model equations. The CBECS (Commercial Buildings Energy Consumption Survey [5]) is the main source providing statistics of buildings area and its characteristics such as geographic division, building type, construction period, etc.

In France, the main statistics source is the CEREN[6], which provides estimates of the *heated floor area* of the national building stock. In consequence, most of French tertiary building stock models are based on CEREN's statistics available under conditions. For models which concern smaller scales than national, the current method consists in disaggregating the heated floor area of the CEREN by establishing an area per job ratio for each sub-sector, and using national and local employment statistics. It is deployed in tools such as ENERTER [7] developed by Energies Demain SARL. This method will be detailed in the following section.

Errors in evaluating the floor area may have important impacts on the results. This variable comes often from surveys which make it difficult to evaluate uncertainties. Indeed, inadequate responses to surveys might results from: lack of knowledge in the issue, bad understanding of questions, lack of definition for technical vocabularies, and so on. Geographic information system (GIS) is a mean to overcome the lack of information on buildings [8], as it provides the geometric data of each one.

In this paper, we propose datasets fusion methods that evaluate the floor area at building's scale in two different ways, each of them having its own pros and cons which will be discussed further. Those methods are designed to be applicable at any scale (from small cities to large region or even national scale). They have a good potential for improvement, because more and more datasets are becoming available in open source.

2. Method

2.1. Datasets selection

The final goal is to create a new dataset that will be the foundation of energy modeling of group of buildings. To do so, original databases must have information at individual scale, or can be disaggregated at that scale by merging with others. Databases are chosen in order to harvest most important information of the stock such as building shape, construction period, and economic activity. The main databases employed in this study are:

- SIRENE: Register system of companies and public organizations [9]
- BDTOPO: Buildings GIS data [10]
- SITADEL: Construction permit chronological data (year by year, from 1975 to 2015), aggregated at city scale [11]
- CEREN: Statistics of building stock aggregated data at national scale [6]

These databases are designed for different goals with various formats and units. The particularity of our approach is to integrate these data to represent the tertiary building stock as a collection of individuals with their own characteristics, which is a necessary input to design a detailed level bottom-up model of the tertiary building stock.

The data fusion step allows obtaining a description of each individual with its area. Before that, it requires pre-analysis of datasets in order to reconcile definitions and units of building area. With available data, this work introduces two ways to obtain the area. The first method consists in identifying areas by crossing SIRENE and CEREN information to calculate the heated area of each company in the city by economic sector. The second method consists in crossing BDTOPO and SIRENE information to identify the buildings in which tertiary companies are located. Finally, we compare the area calculated by each method to SITADEL data.

2.2. Datasets fusion methods

Each data source below provides a great quantity of information, but none is designed specifically for building stock modeling. To avoid being overwhelmed by too much information, important data must be filtered and then extracted. From data sources shown above, SIRENE and BDTOPO contain individual information of companies and buildings, respectively. In order to join national statistics of the tertiary building stock from CEREN to other datasets, they must be disaggregated to smaller scales. We deploy the same disaggregation method as ENERTER, which consists in establishing an area per job ratio (m²/job) for each economic sub-sector by dividing its total area (from CEREN) by its number of jobs from national employment statistics. Afterwards, by merging SIRENE and BDTOPO, jobs of companies and organizations are allocated to buildings. Unknown occupancy is a frequent problem when we use GIS data. In the absence of a perfect solution, the allocation is realized by the matching of addresses of companies, organizations and buildings. The final step consists in evaluating the floor area of each building in 2 different ways, either by applying the area per job ratios to the allocated jobs, or by estimating the floor area from the footprint and the height. The results are compared to the statistics of construction (SITADEL) at city scale. **Table 1** lists data we have selected.

Table 1. Extraction from datasets

Dataset	Extracted data	Geographic scale
SIRENE	Company/organization ID, address, economic activity code, employment statistics	Multi-scale*
CEREN	Economic sector & sub-sector, ratio heated floor area per job	Aggregated national scale
BDTOPO	Building footprint, height, parcel, topography data, building-address links	Multi-scale
SITADEL	Building type, total built floor area per area range, construction period (1975-2015)	Aggregated city scale

*A dataset is considered multi-scale when it contains individual information that can be aggregated at any scale

In detail, two main ways to evaluate the tertiary building stock's area are:

- The first one by the area per job ratio (*area per job method*):

$$S_{f\_heated} = R_{S/j} \times N_j$$

S_{f_heated}: Heated floor area $R_{S/j}$: Ratio heated floor area per job (CEREN) N_j: Number of jobs (SIRENE)

- The second one by the GIS data of each tertiary building (*GIS method*):

$$Nf = \frac{Hb}{Hc}$$

$$S_f = Fp \times Nf$$

Nf: Number of floors
Hb: Building height (BDTOPO)
Hc: Ceiling height (Assumption)
Fp: Building footprint (BDTOPO)

Datasets fusion and results comparison are not possible without the standardization of definitions, units and classifications. For example, the floor area in SITADEL is defined in the official French construction regulation, but the heated floor area in CEREN is not. In French urban planning law, we have several specific building's area definitions. In general, there are 4 major measurements shown in the following table.

Table 2. Floor area measurements

French area measurement	Parallel international area measurement	Definition
SHOB	GFA - gross floor area	The total floor area contained within the building measured to the external face of the external walls
SHON*	/	The SHON is the SHOB less the floor areas taken up by unenclosed areas, parking, basement and attic with a ceiling height less than 1,8

(area definition in SITADEL)		meters, enclosed machinery rooms on the roof, non-convertible basement, and cellars
SDP	GIA - Gross internal area	The floor area contained within the building measured to the internal face of the external walls of enclosed levels; less floor areas the taken up by floor areas with a ceiling height less than 1,8 metres, stairs and escalators, parking, non-convertible attic, machinery rooms, mechanical and electrical services, and cellars
SU	UFA - usable floor area	The SU is the SDP less the floor areas taken up by internal insulation, internal walls, fixed partitions, convertible basement and attic, circulation areas, and mezzanines

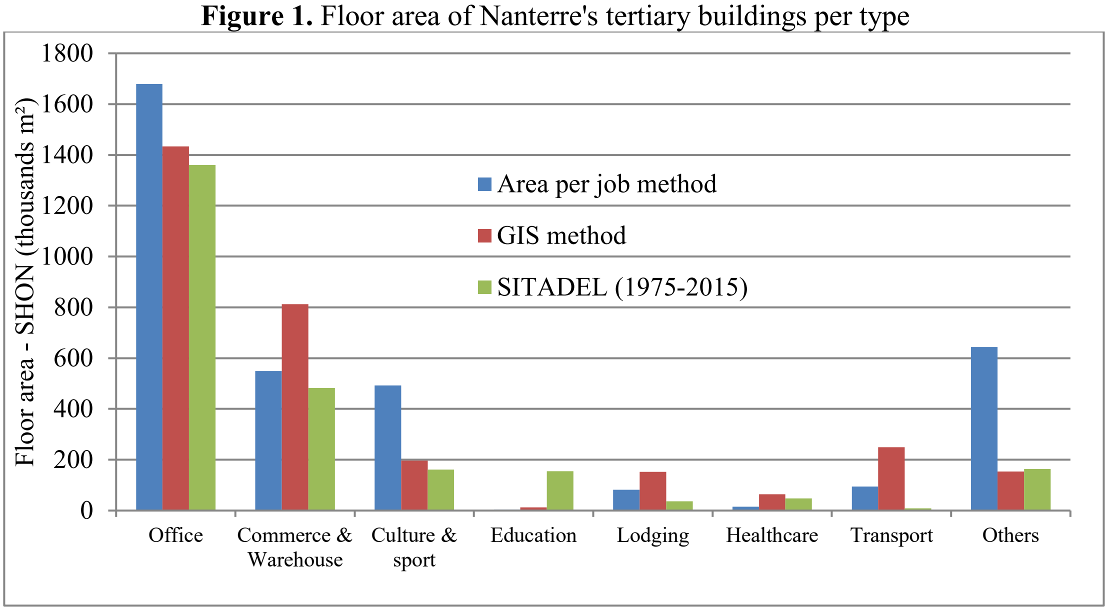
Considering that SHON is the area definition used in SITADEL, all floor area must be converted to SHON for final comparison. In this case, assumptions are made to link the heated floor area to the closest metrics in urban regulation, which is the SU. Afterwards, we establish a conversion ratio between SHON and SU by referencing to real building designs. Given that the floor area calculated by the GIS method corresponds to the SHOB's definition, a conversion ratio between SHOB and SHON must also be established.

Not only the floor area definition, but classifications of building area must also be standardized. In SITADEL, the floor area is classified by building type. Buildings in BDTOPO also have a classification by type, but a large part of buildings is classified as "indifferent", and all tertiary buildings are aggregated under a single "tertiary" class. In order to align categories, we establish the links between activity codes (SIRENE's classification) and building types (SITADEL's classification) using expert rules, and then classify each building according its activity code with highest job number.

3. Results

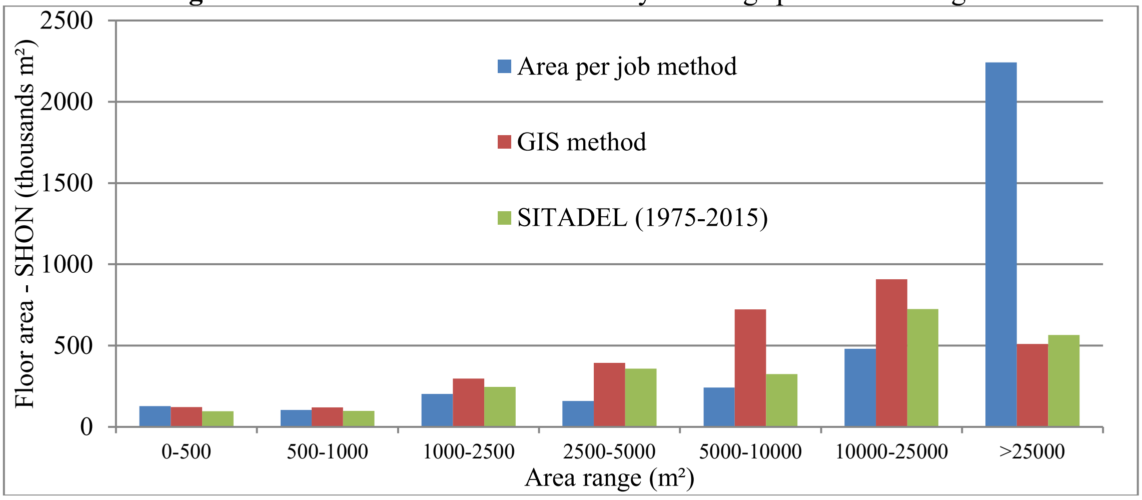
Methods above are applied to the case study of a city. In this paper, we present the tertiary stock of Nanterre, a city in the northwest sector of the Paris region. By construction of the methodology, any city of Metropolitan France could be used. The floor area evaluation is compared to SITADEL's data. At city scale, the comparison is carried out according to the building types and to the floor area ranges respectively in **Figure 1** and **Figure 2**. We observe that the total floor area in SITADEL has the lowest value, we assume that the gap is due to the lack of old buildings (built before 1975).

Figure 1. Floor area of Nanterre's tertiary buildings per type



As we can see, the distribution by building type is relatively even. The floor area gaps between the three sources in **Figure 1** differ according to the building type, while in **Figure 2** the value given by the area per job method is outranged compared to the two others for the range > 25000 m². The reason can be found in the datasets specificities.

Figure 2. Floor area of Nanterre's tertiary buildings per surface range

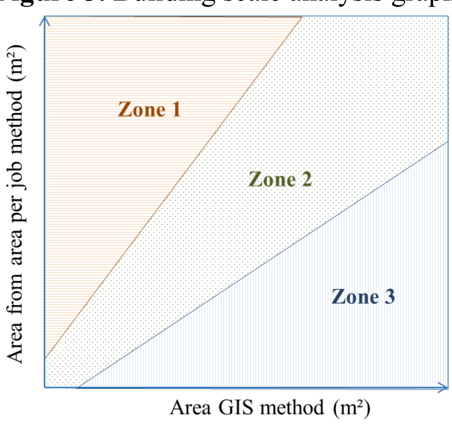


Concerning the area per job method, the important value for the > 25000 m² range is due to the employment statistics in SIRENE. There are many headquarters of big companies and organizations in Nanterre, who also have other sites all over the country. For centralized management, many companies report a significant number of jobs at headquarter, but many workers are located in other sites. In this case, the area per job method overvalues headquarters area. At small scale, the area per job method is also limited by the aggregation which cannot account for the local specificities. For example, the aggregated ratio area per job for culture & sport buildings doesn't differentiate a football stadium and a cinema.

The GIS method has difficulties to evaluate the real area used for tertiary activities within buildings. Especially in big cities, buildings are shared between residential and tertiary activities. Topography data also has bugs such as lack of buildings or bad building shapes compared to reality. The lack of data also regards address matching, as sometimes the links address-buildings are not complete. It appears that many campus (university & research site) have this problem, and in this case jobs cannot be allocated to all concerned buildings.

As SITADEL is not available at building scale, the comparison is only possible between the two other evaluation methods: area estimates are compared in a graph (shown in **Figure 3**) with the area from the GIS method on the X-axis and the area per job method in Y-axis.

Figure 3. Building scale analysis graph



The position of each building in **Figure 3** allows verifying the coherence of estimation methods. In case of perfect coherency the results should be in zone 2 around the diagonal. In fact, results already obtained are very spread in the figure. We propose an assumption/verification method to explain the differences and conclude on the best assessment of the surface. It is a case-by-case study of buildings. To illustrate our method, we give here after examples of assumptions that can be formulated and then empirically tested by survey and statistical analysis:

- A building in zone 1 could be a headquarter or located in a campus
- A building in zone 3 could be shared between tertiary and residential

Specific rules based on this analysis could be included to our method of building stock modeling. For example: In the case of shared buildings, the area per job method could be more relevant than the GIS method. In consequence, we could make a rule that whenever a shared building is detected according to formulated assumptions, we account the value given by the area per job method for the floor area occupied by tertiary activities. However, these rules must probably be adapted at least by building type (offices, schools, hospitals,...).

4. Discussion

The methods presented in this work allow describing the tertiary stock as a group of individuals. By merging datasets and statistics, buildings geometry and occupancy characteristics can be attributed. The heterogeneity of this sector is reflected through the different characteristics of each individual. As open source data availability increases, this representation can be improved in the future. Assumptions concerning the thermal characteristics and energy-consuming systems could be made via topography data and occupants activities of each building. Those characteristics are necessary to make the modeling of the stock more realistic, which improves the study of energy demand as well as long-term projections.The datasets fusion also allows to compare the quality of different sources and to use best area estimate for each building.

References

[1] Qudrat-Ullah H 2015 Modelling and Simulation in Service of Energy Policy *Energy Procedia* **75** 2819–2825
[2] ADEME - French Environment & Energy Management Agency, "ADEME - Buildings." [Online]. Available: https://www.ademe.fr/en/expertise/buildings. [Accessed: 30-Apr-2019].
[3] Reinhart C F and Cerezo Davila C 2016 Urban building energy modeling – A review of a nascent field *Building and Environment* **97** 196–202
[4] U.S. Energy Information Administration (EIA) 2017 Commercial Demand Module of the National Energy Modeling System: Model Documentation
[5] U.S. Energy Information Administration (EIA) "Commercial Buildings Energy Consumption Survey (CBECS)." [Online] Available: https://www.eia.gov/consumption/commercial/about.php.
[6] Centre d'étude et de recherches économique sur l'énergie - France, "CEREN." [Online] Available: https://www.ceren.fr/. [Accessed: 14-May-2018].
[7] Energies Demain SARL 2011 Suivi de l'impact énergie-climat des secteurs résidentiel et tertiaire en région Centre 87-109
[8] Buffat R, Froemelt A, Heeren N, Raubal M, and Hellweg S 2017 Big data GIS analysis for novel approaches in building stock modelling *Applied Energy* **208** 277–290
[9] Institut national de la statistique et des études économiques (INSEE) - France, "SIRENE." [Online] Available: https://www.sirene.fr/sirene/public/accueil. [Accessed: 14-May-2018].
[10] Institut national de l'information géographique et forestière (IGN) - France, "BDTOPO®." [Online] Available: http://professionnels.ign.fr/bdtopo. [Accessed: 14-May-2018].
[11] Ministère de la transition écologique et solidaire - France, "SITADEL." [Online]. Available: http://www.statistiques.developpement-durable.gouv.fr/donnees-ligne/th/logement-construction-2.html. [Accessed: 14-May-2018].