Bottom-up statistical analysis of the energy consumption of French single-family dwellings

Maxime Raynaud, Bernard Bourges, Dominique Osso, Bruno Duplessis, Jérôme Adnot

WORKING PAPER finalized in July 2014

Bottom-up statistical analysis of the energy consumption of French single-family dwellings

M. Raynaud[a, b, *], B. Bourges[c], D. Osso[b], B. Duplessis[a], J. Adnot[a]

[a] *MINES ParisTech, CES, 60 boulevard Saint-Michel, 75272 Paris Cedex 06, France*

[b] *EDF R&D, Site des Renardières, Avenue des Renardières - Ecuelles, 77818 Moret-sur-Loing Cedex, France*

[c] *Ecole des Mines de Nantes, UMR GEPEA – FR IRSTV, 4 rue Alfred Kastler, 44300 Nantes, France*

[*] Corresponding author. Tel.: +33 (0)1 60 73 67 93. Fax: +33 (0)1 60 73 65 70 E-mail address: maxime.raynaud@mines-paristech.fr

ABSTRACT

Implementing effective energy policies in the residential sector requires better understanding of the sources for the dispersion of energy consumption amongst households. Bottom-up statistical models have been identified as one major modelling technique, particularly accounting for the diversity of inhabitants behaviours. In the various statistical models of the literature, behaviour characteristics are seldom incorporated in the data set but socio-economic data are most often used as "proxy" of occupant behaviour. The present study is based on a detailed survey, combining energy billing data with technical, geographical, socio-economic and behavioural variables. The corresponding sample, despite of its limited size (420 individuals), is representative of French households. A statistical model relating energy consumption to the other variables has been applied, enabling simultaneous use of quantitative and qualitative explanatory factors (ANCOVA, analysis of covariance). The main determinants found for energy consumption in the sector of single-family dwellings in France are, by decreasing order of weights, surface area, type of main heating system, age of the household head and climate zone. We detected that the most influential behaviour variable is night temperature setting reduction. Explanation and prediction capacities of the model as the accuracy of the model coefficients are studied and some possibilities of improvement are proposed.

Keywords: Total energy consumption, single-family dwellings, France, main determinants, occupant behaviour

1. Introduction

Faced with climate and energy-related issues, energy efficiency in the building sector (1) is one of the areas for action in European (2) and French (3) policies. In France the residential sector accounted for 30% of final energy consumption in 2009 (4). The implementation of effective energy policies requires, *inter alia*, proper understanding of the factors influencing energy consumption in this sector and of the variability sources amongst households.

Quantitative modelling is a key issue to select the main targets and the relevant solutions and to evaluate the impacts of implemented measures. Bottom-up methods have been identified as one major family of modelling techniques in this respect, based on disaggregated data by households or groups of households (5).

A first set of bottom-up methods is classified as "**engineering models**". Those are knowledge based models (based on thermodynamics or heat transfer, e.g.) using data on the building, its energy systems and the way they are operated. These models generally use as inputs statistical information, either through samples of representative individuals or through archetypes (5). Another engineering technique employs distributions of appliance ownership and use in the overall population with common appliance ratings to calculate energy consumption of end-use (i.e the product of appliance ownership, appliance use, appliance rating and the inverse of appliance efficiency). Some studies (6; 7) have shown the importance of occupants behaviour in determining the energy consumption of a dwelling and the limitations of engineering models, which are based on assumptions concerning behaviours (5). These assumptions in the engineering models are one of the main causes cited for explaining the deviations between predicted and observed consumptions when energy-efficiency measures have been implemented (8; 9).

Bottom-up "**statistical models**" are alternate methods, particularly well suited for estimating actual energy consumptions, accounting for the diversity of inhabitants behaviours (5). Bottom-up statistical models are generally based on a sample of representative households, using data of energy billing for one or more years, complemented by a questionnaire collecting data on building, energy systems and household characteristics. Statistical techniques such as linear regression or neural networks are used to derive a predictive model of energy consumption via explanatory variables from the questionnaire (5) and to identify the most important predictors.

The present paper concentrates on bottom-up statistical models while until now in many countries, more attention has been paid to engineering models, due to the important data cost for statistical models.

The study is based on a detailed survey, combining energy billing data with technical, geographical, socio-economic and behavioural variables, enabling to develop a statistical model. After a literature review, the sample data set is presented and the statistical analysis method is introduced. Results of the statistical model are reported and discussed. In particular, explanation and prediction capacities of the model are analysed, as well as the accuracy of model coefficients. Possible improvements for statistical bottom-up models and related input data and surveys are proposed as a conclusion.

2. Bottom-up statistical models of energy consumption in residential sector: a short literature review

A selection of studies based on bottom-up statistical models in various countries is reviewed below. Most of them make use of regression type analysis (6-7; 10-20). Neural networks or similar pure black-box approaches have not been considered here, due to the lack of physical significance of the coefficients of these models, making difficult comparison between different studies.

Table 1 presents the selected studies as regards data sets, response variable considered and main significant explanatory variables by categories (building characteristics, energy systems, refurbishment, geography and climate, socio-economic characteristics of households and behavioural data). Explanation capacities of each model are summarized by the determination coefficient of the regression (R^2 or adjusted R^2), which represents the fraction of response variance explained by the model.

Table 1 emphasizes the diversity of situations and results. Sample size ranges from limited data sets (less than 100) to a few thousand households. Response variables may cover:

- different end-uses: space heating alone (7; 15; 16; 18; 20), space heating and hot water production (6; 14), all end-uses (11-13);
- or specific energy carriers: natural gas (17), electricity (19), wood (10), oil (16).

Often the yearly energy consumption (in kWh, MJ or specific units) is directly the response variable (6), but sometimes a transformed response is used with the energy consumption per unit area (18), or per room (15) or the logarithmic function of the energy (7; 20) or both transforms (18). In some cases, annual expenditure for energy (in € or £) is analysed, combined or not with former transforms (11; 13; 14; 15).

Depending on data sets used, explanatory variables incorporated in the model are very diverse, building characteristics being always significant. Behaviour characteristics are seldom incorporated in the data set (6; 7; 10; 16; 19) but socio-economic data are most often used as “proxy” of occupant behaviour. A distinction can be made between econometric studies mainly oriented towards estimate of price and income elasticities (generally using multi-year periods with varying energy prices) and more physical analyses trying to identify and to quantify all major influences on energy consumption, with a shorter term perspective and a predictive aim.

Most of the statistical models presented by these studies present a low level of explanation capacity (i.e. a coefficient of determination less than 0.5).

Moreover, noticing that most of these papers show the expected effects of some variables (technical, climate and socio-economic) on energy consumption (or bills), consensus is rarely reached and it is common to encounter studies that find, if not an opposite effect, at least a non-significant effect (e.g., Cavailhes et al. (11) as well as Guerra-Santin et al. (6) find that the presence of double glazing generates a less energy consumption but Risch and Salmon (12) determine the effect as non-significant). Non-significant effect does not necessarily mean that the corresponding variable has no effect. The effect can remain undetected by the statistical model because it is hidden by noise, lack of one or more major influences amongst predictors or by multicollinearity amongst predictors, particularly with small size samples.

Former comments clearly show that comparison between different bottom-up statistical models is difficult and must be made carefully. Furthermore, differences between countries -even with similar development levels- cannot be neglected, whatever their origin (building and energy regulation, building industry practices, climate, way of life, available census or energy billing information), even if partially taken into account by some predictors in the model. Comparisons are likely to be relevant only within one particular country or between very similar countries (the common framework given by European Union Directives for European countries is far from being sufficient in this respect). Representative open data bases should be developed in every country with such bottom-up modelling in order to support decision making, implementing and evaluation of energy efficiency policies.

Like in many other countries, the literature on bottom-up statistical modelling of French residential energy consumption is rather limited and more often published as grey literature. Only four recent studies (10-13) addressing the question of energy demand in the French residential sector, according to data panels on the household scale were identified.

The public database used by Cavailhes et al. (11) and Risch and Salmon (12), following the example of their European counterparts (14; 15), provide data that relate to the features of the dwellings (type of dwelling, surface area, insulation level, etc.), the features of the energy systems (energy used, collective or individual heating, etc.), as well as geographical information (climate, urban density, etc.) and finally information relating to the households characteristics (number of people, income, occupation status, etc.). This database (Housing Survey

2006 or “Enquête logement 2006”, by the National Institute of Statistics and Economic Studies, INSEE) is fully representative of the national building stock and of the population (21). However, these panels do not provide information on the behaviour of the households. Furthermore, energy data are not directly collected by the survey but they are estimated from energy expenditure data (in €), based on additional assumptions. Anyway, rich information is available on buildings and occupants, so these studies deserve careful analysis. More detailed results from references and their comparison with the results of this study can be found in the Discussion section below.

Calvet and Marical work (13) is also based on expenditure data, from a 2006 national survey on family budgets, but very limited information on dwellings and occupants is available (nothing on behaviours) and reporting of the statistical model is not complete.

Finally, Couture et al. (10) developed an econometric model tackling a very specific problem related to fuelwood consumption in a particular region (Midi-Pyrénées) whereas only 16.4% of the sample use wood as primary heating source.

This review clearly emphasizes the need to improve statistical models, particularly by taking into account behaviour characteristics and not only building technical data. Within one specific country, like France, collecting and processing new data sets from enriched surveys is necessary to develop more extended knowledge about building stock, main determining factors of energy demand and ways to reduce it effectively. The present work aims to complement former studies for the French residential sector, using a new representative panel incorporating actual energy consumptions for all energy carriers and all end-uses. The data source on which this paper is based helps to fill the lack of behaviour information with practices reported by the occupants, while presenting information on all the elements mentioned above (technical, geographical and socio-economic) as well as on recent undertaken retrofit work—a situation that very few databases include.

Thus, this study aims to provide information that answer to the following questions:

- What share of the variation in energy consumption can be explained by a model combining technical (building, energy systems, recent work), geographical, socio-economic and behavioural variables? What is the prediction capacity of such a model?
- What are the main determinants for energy consumption in the sector of single-family dwellings in France? What behavioural elements do these factors include?
- What are the relative weights of these various determinants in the model?

The next section presents the data and the statistical analysis method used in this respect.

Table 1
Synthesis of a set of bottom-up statistical models on residential energy consumptions and expenditures in several countries.

Reference	Information about data (sample size, country, years of observation, source of data)	Dependent variable (in unit used)	Main explanatory variables (significance at least 10% level)						Explanatory capacities
			Bui	En. Sys.	Re	Geo	Soc. Eco	Beh.	
Guerra-Santin et al. (6)	15,000 observations – the Netherlands – 2000 (3 years of data) - interviews	Annual energy consumption for space and water heating (in MJ)	14	1	-	-	5	6	R^2=0.459 (total model = steps 1+2+3)
Haas et al. (7)	400 households – Austria - 1993-1996 - accounting, monitoring and questionnaires	Log. of annual space heating energy consumption (in kWh)	3	1	-	-	1	1	Adj. R^2=0.795 (unrestricted model)
Kjaerbye et al. (17)	150,000 observations (34,700 single-family owner-occupied houses) - Denmark - 1998–2003 - different administrative data bases	Log. of annual natural gas consumption (in kWh)	3	1	1	2	3	-	R^2=0.243
Leth-Petersen and Togeby (18)	36,000 observations (4820 apartment blocks > 1500 m² with a central heating system) – Denmark - 1984–1995 -different data bases	Log. of annual space heating energy consumption per square meter (in kWh/m²)	2	2	-	-	1	-	R^2=0.322
Meier and Rehdanz (15)	64,000 observations -Great Britain - 1991-2005 excepted 1996 - interviews	Log. of annual space heating expenditure per room (in £/room)	3	1	-	2	7	-	R^2=0.274 (global model)
Ndiaye and Gabriel (19)	62 houses - Oshawa (Canada) – 2007-2008 - energy audits, phone surveys and smart meter.	Annual electricity consumption per square foot (in kWh/ft²)	1	5	-	-	2	1	Adj. R^2=0.754
Rehdanz (14)	12,000 observations - Germany - 1998 and 2003 - interviews	Log. of monthly expenditure for space and water heating per m² (in cent/m²)	4	2	1	1	7	-	R^2=0.194 (global model)
Sardianou (16)	500 households – Greece - 2003 - interviews	Log. of annual fuel quantity consumed for space heating (in l)	1	-	-	-	4	3	Adj. R^2=0.37 (final model II)
Schuler et al (20)	15,000 observations – West Germany - 1998 - interviews	Utilization intensity (space heating utilization)	5	2	-	-	4	-	R^2=0.008 (household char. only); R^2=0.144 (building char. only)
Couture et al. (10)	2254 households – France (the Midi-Pyrénées region) - 2004 – 2005 - phone survey	Annual wood quantity consumed (in m^3)	1	2	-	-	1	1	R^2 unknown
Cavailhes et al. (11)	47,000 observations – France – 2002 and 2006 - interviews	Log. of annual total energy expenditure per square meter (€/m²)	6	1	-	1	5	-	R^2 between 0.12 and 0.20
		Log. of annual total energy consumption per m² (in kWh/m²)							R^2 unknown
Risch et Salmon (12)	37,000 observations – France – 2002 and 2006 - interviews	Log. of annual total energy consumption per m² (in kWh/m²)	9	2	-	1	4	-	R^2 =0.36 (single-family dwelling)
Calvet and Marical (13)	10,000 observations – France – 2006 - interviews	Log. of annual total energy expenditure per square meter (in €/m²)	1	1	-	-	1	-.	R^2 unknown

* = Bui.: Building; En. Sys.: Energy Systems; Re.: Refurbishment; Geo.: Geography; Soc.-eco.: Socio-economic; Beh.: Behaviour.

3. A representative sample with a large set of explanatory variables and an adapted method

Results of a survey conducted by TNS Sofres on behalf of EDF in June 2009 were used as an input data set for the present study. This survey was conducted amongst 2,012 households selected from a panel of 20,000 households (“Métascope” panel (22)), to be representative of the population of households living in France. The questions related to energy consumption by energy type (electricity, gas, heating oil, LPG and wood) of households in 2008 and its determinants (characteristics of the dwelling, heating system, socio-economic characteristics of the household, household habits, etc.). The questionnaire used was drawn up by Cayla, who used the results to study the effect of behaviour on the long-term estimation of heating energy consumption in the French residential sector, mainly based on an engineering model (23; 24).

Before exploiting these data, we proceeded to clean up the panel. More than two third of data (1,392 dwellings) were removed. Most of removed data concerned the lack of energy data; either some energy consumption was not declared (e.g., no electricity consumption) or it was unknown (housing with collective heating but without individual metering of consumption). Cases, with inconsistencies between energy systems and declared consumption or with extreme total consumption levels (higher than 1000 kWh/(m².year) or lower than 40 kWh/(m².year)[1]) were also discarded. Finally, the size of the sample was reduced to 620 households or main residences (420 single-family dwellings and 200 apartments).

Since only 30% of the source sample was kept, we needed to verify that it remains nationally representative. Four criteria are usually applied to ensure the representativeness of the "Métascope" panel; they were applied here to the reduced sample. Figure 1 compares the distribution, according to type of dwelling (1-a), the socio-professional group of the head of the household (1-b), the age of this person (1-c) and the number of people living in the household (1-d), between the data from the national institute of statistics INSEE (25) (national benchmark), the original sample (2,012 dwellings) and the final sample of 620 dwellings.

In figure 1-a, we can see that apartments are under-represented (-10%) in the final sample, mainly because we eliminated apartments with collective heating.

In the diagram 1-b presenting the socio-professional group of the head of the household, we can see that the representativeness is quite good, with a maximum deviation of 4% for "executives".

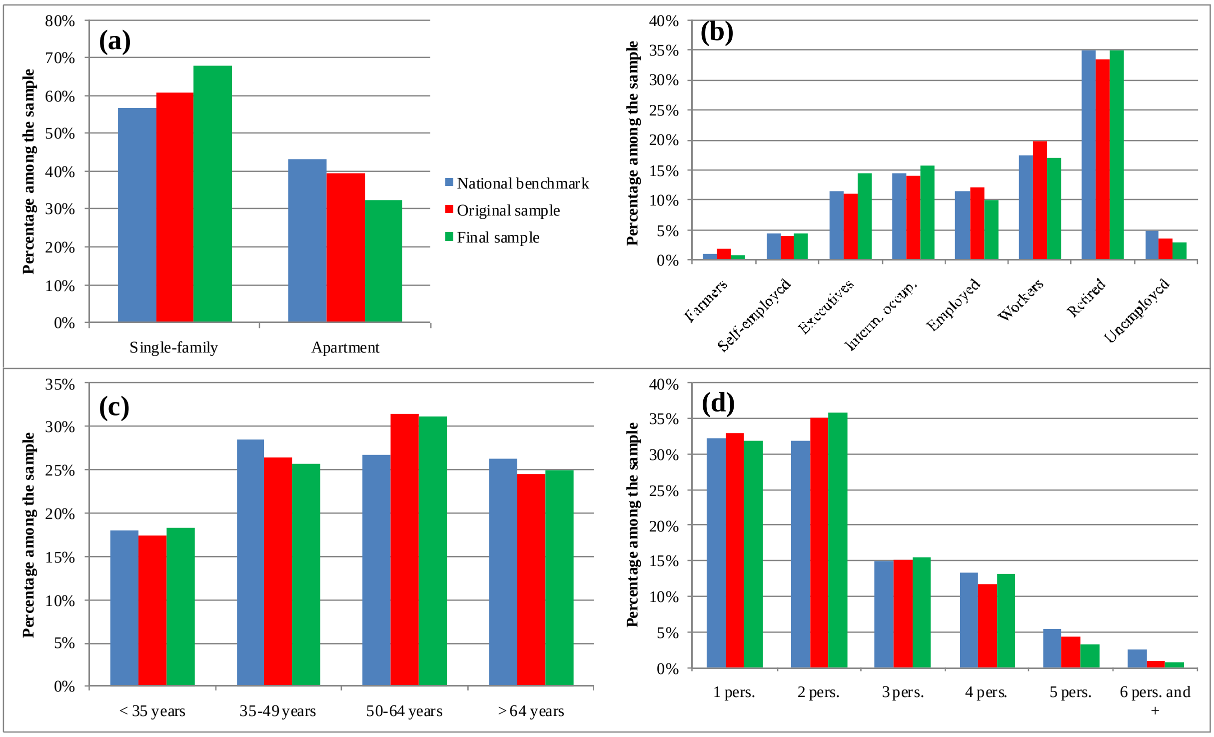

Fig. 1. Study of the representativeness of the final sample. (a) Type of dwelling, (b) Socio-professional category of the head of the household, (c) Age of the head of the household, (d) Number of people in the household

The graphic relating to the age of the head of the household shows maximum deviations of 4% in the two intermediate classes (35-49 years and 50-64 years). Figure 1-d shows the greatest deviation for the category "two people in the household", with a difference of 5% between the data from INSEE and the final sample.

In conclusion, compared with the national benchmark, the sample deviations are around 10% for one criterion (type of housing: single-family houses vs apartments) and less than 5% for the other three. We therefore postulate that the 620 households of the final sample are quite representative of French households.

However, due to the bias present (no cases with collective heating) amongst the 200 apartment dwellings, we chose not to study this sub-sample and, thus, we concentrated only on the 420 single-family dwellings.

[1] Total consumption level per surface area, in final energy, slightly less than that of the most efficient dwellings currently being built.

In the context of this study, we sought to explain the natural logarithm (Ln) of consumptions in final energy for all end-uses of the dwellings for the year 2008. This total energy consumption was obtained from adding up five energy consumptions (electricity, gas, heating oil, LPG and wood), converted into kWh_{LCV}[2], supplied by the questionnaire[3]. The use of the Ln is justified from an analytic point of view. Indeed, the main share of energy consumption in a dwelling corresponds to space heating load, it is described mathematically by multiplying effects of building thermal characteristics, HVAC system efficiency (precisely inverse of efficiency) and degree-days (26). Thus, using the Ln makes it possible to transform these effects into additive relationships as required for a linear regression model.

In addition, the use of the Ln for the dependent variable is particularly interesting in terms of information given by the model coefficients. For a quantitative explanatory variable used with a Ln transform, the coefficient corresponds to the elasticity of the dependent variable in accordance with the explanatory variable. In the case of a quantitative explanatory variable not expressed in Ln, the coefficient is approximately the same as the percentage change in the dependent variable for an increase of one unit in the explanatory variable. In the last case of a qualitative explanatory variable, the coefficient of one category is also approximately equal to the percentage change of the dependent variable between the reference category and the considered category, everything else being equal.

Regarding the set of explanatory variables, we have obtained a set of twelve variables relating to the building, eight relating to energy systems, one relating to refurbishment, three relating to geography, four relating to socio-economic characteristics and ten relating to behaviour (Table 2). We can notice that very few statistical models in the literature include such a number of behaviour variables.

In order to check that multicollinearity of explanatory variables remains in a reasonable range, we made sure that none of the quantitative variables or qualitative variable categories has a Variance Inflation Factor (VIF) of more than 3.

---

[2] LCV: Lower Calorific Value. HCV: Higher Calorific Value. Conversion factor of gas: 0.9 kWh_{LCV}/kWh_{HCV}. Energy content of heating oil: 9.96 kWh_{LCV}/litre. Energy content of LPG: 12,769 kWh_{LCV}/tonne. Energy content of wood: 1,480 kWh_{LCV}/stere.

[3] A certain number of LPG consumption data for cooking use are missing. They are filled in with the mean national consumption for the year 2008 supplied by the Center of Studies and Economic Researches on Energy (CEREN) (34) segmented by type of dwelling. 700 kWh_{LCV} for cooking only with LPG added for 28% of the 420 dwellings. 465 kWh_{LCV} for mixed cooking with LPG and electricity added for 21% of the sample.

Table 2
Set of explanatory variables used for ANCOVA.

Variable	Definition (modality percentage of qualitative variable) [min; max of quantitative variable]
Building	
LnSurface	Ln of the surface area in m², [3.40; 5.94]
Ceiling_H	1- ceiling height < 2.49 m (82.6% of the sample); 2- 2.5 m to 2.99 m (12.9%); 3- >3 m (4.5%)
Num_storeys	1- one-storey (54.5% of the sample); 2- two storeys (42.1%); 3- three or more storeys (15.5%)
House Shape	1- compact shape (77.4% of the sample); 2- L-shaped or elongated (17.6%); 3-complex (5,0%)
Party_walls	1- detached house (57.1% of the sample); 2- semi-detached (27.4%); 3- terraced (15.5%)
Veranda	1- yes, the house has a veranda (14.8% of the sample); 2- no (85.2%)
Floor	1- on a platform or ground (35.0% of the sample); 2- above a crawl space (20.2%); 3- above a basement, cellar or other unheated room (35.5%); 4- combination of the above (9.3%)
Roof	1- with virgin loft (63.3% of the sample); 2- heated habitable roof space (17.4%); 3- unheated habitable roof space (11.7%); 4- flat roof or combination of the above (7.6%)
Building_vintage	1- dwelling built before 1975 (45.5% of the sample); 2- between 1975 and 1988 (28.6%); 3- between 1989 and 2000 (16.2%); 4- after 2000 (9.7%)
Double_glazing	1- no double glazing (DG) (8.8% of the sample); 2- partial old DG (before 1990) (3.1%); 3- partial recent DG (after 1990) (12.1%); 4- all old DG (20.5%); 5- all recent DG (55.5%)
Outer_wall_insulation	1- the outer walls are not insulated (31.0% of the sample); 2- 5 cm or less (21.0% of the sample); 3- around 10 cm (33.5%); 4- around 15 cm (10.0%); 5- around 20 cm or more (4.5%)
Roof_insulation	1- roof not insulated (22.1% of the sample); 2- less than 10 cm (49.3%); 3- >10 cm (28.6%)
Energy systems	
Main_heating	1- heat pump (4%); 2- elec. heated floor (3%); 3- elec. heated panels (6%); 4.0- elec. convectors <5 years old (6%); 4.1- elec. conv., 5 to 10 years (6%); 4.2- conv., 11 to 15 years (4%); 4.3- conv., 16 to 25 years (6%); 4.4- conv., > 25 years (3%); 5- efficient gas boiler (6%); 6- efficient boiler, other energy (4%); 7.0- standard gas boiler, < 5 years old (5% of the sample); 7.1- stand. gas boiler, 5 to 10 years (9%); 7.2- stand. gas boiler, 11 to 15 years (4%); 7.3- stand. gas boiler, >16 years (4%); 8.0- standard oil boiler, <10 years (6%); 8.1- stand. oil boiler, 11 to 15 years (4%); 8.2- stand. oil boiler, 16 to 25 years (3%); 8.3 stand. oil boiler, > 25 years (3%); 9- stand. boiler, other energy (4%); 10- wood-burning stove (3%); 11- enclosed wood roomheater (6%); 12- mobile electric radiator or kerdane stove, or heating cooker or others (3%).
Occasional_supplement	1-occasional supplementary heating (more than 15 days per year) (39.3%); 2- no (60.7%)
Exceptional_supplement	1- exceptional supplementary heating (15 days or less per year) (19.3%); 2- no (80.7%)
Regulator_system	1- yes, the dwelling has a heating temperature regulator or programmer (71.4%); 2- no (28.6%)
DHW_prod_system	1- DHW produced by a boiler (38.1% of the sample); 2- electric storage tank (48.3%); 3- electric instant water heater (11.2%); 4- gas instant water heater (2.4%)
Cooking_energy	1- cooking with gas only (15.2% of the sample); 2- mixed gas and electricity (11.7%); 3- LPG only (30.2%); 4- mixed LPG and electricity (21.2%); 5- electricity only (21.7%)
Appliance_ratio	1- at least two different types of electrical appliance (6.7%); 2- three different types of electrical appliance (12.1%)…8- eight or more different types of electrical appliance (7.1%)
Energy_saving_lights	1- dwelling mostly fitted with energy-saving bulbs (31.4%); 2- partly (47.6%); 3- no (21.0%)
Refurbishment	
Works_last 12 months	0- no energy-saving or improvement works conducted in the last year (77.6% of the sample); 1- improvement works conducted (6.2%); 2- energy-saving works conducted (11.4%); 3- energy-saving and improvement works conducted (4.8%)
Geography	
Climate_zone	1- dwelling located in thermal regulation zone H1a (north and Paris) (27.4% of the sample); 2- H1b (east) (17.1%); 3- H1c (northern part of the southeast) (12.4%); 4- H2a (Brittany) (6.2%); 5- H2b (west and centre) (13.3%); 6- H2c (southwest) (10.0%); 7- H2d (southern part of the southeast) (4.8%); 8- H3 (Mediterranean) (8.8%)
Urban_density	1- rural area (63% of the sample); 2- suburban area (26%); 3- urban area (11%)
Altitude_km	mean altitude of the town in km, [0; 1.48]
Socio-economics	
Age_head_household	1- head of the household less than 35 years old (13.1% of the sample); 2- 35 to 49 years old (28.6%); 3- 50 to 64 years old (31.1%); 4- at least 65 years old (26.2%)
Monthly_income	1- monthly income less than €1200 (8.8% of the sample); 2- €1201 to €1900 (23.8%); 3- €1901 to €3000 (42.4%); 4 - €3001 to €5300 (21.2%); 5- at least €5301 (3.8%)
Occupation_status	1- owner occupied dwelling (76.4% of the sample); 2- council tenant (4.8%); 3- other including leased for free (5.7%); 4- private tenant (13.1%)
LnNum_people	Ln of the people number in the household, [0; 1.79]
Behaviour	
Bath_usage	1- less than one bath per week per person (81.2%); 2- one bath per week per person (6.4%); 3- more than one bath per week per person (12.4%)
Shower_usage	1- less than seven showers per week per person (69.3%); 2- seven showers per week per person (24.1%); 3- more than seven showers per week per person (6.7%)
Living_room_temp	1- mean winter temperature in living room 19 °C or lower (42.6% of the sample); 2- 20 °C to 22

	°C (54.5%); 3- 23 °C or higher (2.9%)
Less_heated_room	1- some rooms not heated or less heated than living room (75% of the sample); 2- no (25%)
Half-day_temp_reduction	1- heating temperature setting always reduced when absent for half a day (29% of the sample); 2- yes, sometimes (23%); 3- no, never (48%)
Night_temp_reduction	1- temperature setting always reduced at night (49%); 2- sometimes (21%); 3- never (30%)
Ventilation_frequency	1- house ventilated every day (69%); 2- several times a week (22%); 3-once a week or less (9%)
Ventilation_time	1- ventilation <10 mn (30%); 2- 10 to 29 mn (54%); 3- 30 to 59 min (10%); 4- >1 hour (6%)
Lights_empty_rooms	1- when entering an empty room, light is found on regularly (68%); 2- never (32%)
Lights_during_day	1- lights on during the day regularly to occasionally (86.7% of the sample); 2- never (13.3%)

We conducted a covariance analysis (ANCOVA, generalisation of a multiple linear regression with both quantitative and qualitative explanatory variables), based on the ordinary least squares estimator. The natural logarithm of the total energy consumption of 420 single-family dwellings was taken as response variable (dependent variable). Since we were seeking the main determinants of the latter amongst a very large number of explanatory variables (38 variables), we started by selecting variables according to their contribution to the model. For this purpose, we recursively removed the variable that made the smallest contribution, by analyzing the probability associated with the F-test of Type III SS (null hypothesis test: "The variance of the model with the variable is not significantly different from that of the model without the variable (null variable coefficient(s))."), until only variables with a probability of no more than 0.15 remained. The table of the Type III SS (Sum of Squares) shows how—regardless of the order of selection of the variables in the model—removing one explanatory variable affects the adjustment of the model, all other variables being kept. Since this first selection was conducted only according to the global contribution of the variables, we applied the method of backward selection to the model as second selection.

This backward selection is applied with a critical probability (Pr) to the Student's t-test of 0.05 (null hypothesis test: "The coefficient is equal to zero."), in order only to keep in the model the variables or categories (for qualitative variables) that have at least this level of significance. In terms of reference for the qualitative variables, we chose to work with the sum of the coefficients of the categories equal to zero. This choice had the advantage of making the model independent from the encoding of the categories, unlike the references (coefficient zero) for the first or last most arbitrary category. For non significant categories, a zero effect is applied.

This statistical work was carried out using XLSTAT software, version 2010.5.07.

4. A valid model with significant effects

4.1. *Validity of the model*

An ANCOVA implies a set of statistical assumptions which must be verified before interpreting the results both from the statistical and from the physical points of views. The key assumptions to be checked concern the model errors which should have no auto-correlation and should be normally distributed with zero expectation and constant variance (homoscedasticity), whatever the values of explanatory variables. Most of these checkings are based on the residuals of the regression model, defined as differences between observed values and values predicted by the model (estimations of model errors).

First of all, we used an F-test (null hypothesis test: "The variance of the model is not significantly different from that of the model with all coefficients equal to zero.") to verify the overall significance of the model. Since the critical probability associated with the F value of 30.04 for 33 and 386 degrees of freedom is lower than 10^{-4}, the null hypothesis can be rejected and the model can be considered as highly significant, concluding that the explanatory variables contribute a significant amount of information to the model.

The auto-correlation amongst the residuals was tested and nothing of significance was found.

Figure 2-a represents the standardized residuals (residuals divided by the estimate of the standard deviation on the error of the model) versus the values predicted by the model for the 420 dwellings of the sample. We observed a symmetric random distribution of the residuals around zero (by construction with the least squares, the residuals have a mean value of zero). Levene and Bartlett tests were conducted to compare variances of residuals between the levels of the qualitative variables and between different categories of the quantitative variables ("LnSurface" and "Altitude_km" were split into four classes while "LnNum_people" was split into six classes). All these tests confirmed the equality of the variances of the various categories, with the exception of the tests conducted on the "Building_vintage" (Pr = 0.020), "Occasional_supplement" (Pr = 0.020) and "Age_head_household" (Pr = 0.036) variables. The oldest buildings (build before 1975) have a bigger variance of residuals than the others (0.109 versus around 0.077), certainly due to a great heterogeneity in thermal characteristics amongst these dwellings (e.g., refurbished or not). Dwellings with occasional supplementary heating system show more scattered energy consumption (variances of residuals: 0.107 with versus 0.075

without); supplementary system may be used on very different ways (e.g., all the days or sometime only). Heads of the household between 35 to 49 years old have a less variance than the others (0.062 versus around 0.098); more homogeneous occupation scenarios might be observed for this category of "young active" people if compared to retired persons. The tests are negative for only these three variables and the differences of variances remain small, thus we consider the hypothesis of homoscedasticity validated.

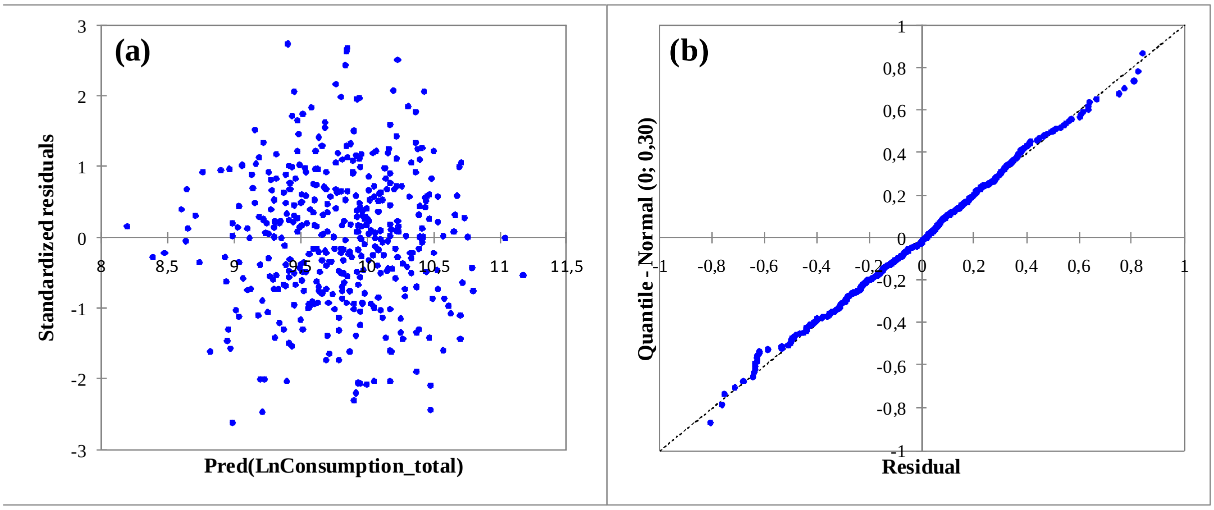

Fig. 2. Summary analysis of residuals of the model. (a) Standardized residuals according to values predicted by the model, (b) Normal Q-Q plot of residuals

Figure 2-b presents the normal probability plot of residuals (Q-Q plot); it makes it possible to compare the probability distribution of the residuals of the sample (on the abscissa, with increasing order) with that of a normal distribution with the same mean (zero) and the same variance would have (on the ordinates). The alignment of the points on the first bisector indicates the normality of the residuals of the model. The critical probability associated with the Jarque-Bera test (null hypothesis test: "The sample follows a normal distribution.") of 0.962 guarantees the hypothesis of normality. In parallel to this verification, we calculated that 93.8% of the standardized residuals are in an interval [-1.96; 1.96] which is totally consistent with the Gaussian assumption (about 95% are expected in this interval).

In conclusion, the main hypotheses are verified and the results of the model can be considered to be reliable. It is important to note that verifying usual assumptions of linear regression models (normality and homoscedasticity) here clearly justifies *a posteriori* the logarithmic transform of the energy consumption as a response variable in the model. Moreover, checking usual assumptions of ANCOVA is a prerequisite before analysing the results of the model and interpreting them both from a statistical point of view and from a "physical" (or causality) perspective.

4.2. *The statistical model*

Table 3 provides the variables eliminated during the first selection of the variables according to their contribution to the model (probability associated with the F-Fisher test of the Type III SS higher than 0.15).

Table 3
List of variables discarded due to minimal contribution to the model.

Order of removal	Variable	Pr > F (Type III SS)
1st	Shape	0.917
2nd	Double_glazing	0.841
3rd	Num_storeys	0.788
4th	DHW_prod_system	0.747
5th	Veranda	0.753
6th	Roof_insulation	0.748
7th	Occupation_status	0.706
8th	Regulator_system	0.610
9th	Less_heated_room	0.509
10th	Lights_empty_rooms	0.491
11th	Half-day_temp_reduction	0.450
12th	Ventilation_frequency	0.351
13th	Monthly_income	0.289
14th	Urban_density	0.323
15th	Appliance_ratio	0.241
16th	Works_last 12 months	0.283
17th	Ceiling_H	0.163

The model (significant terms only) is presented in Table 4. This one presents an adjusted R^2 (coefficient of determination taking into account the number of explanatory variables used by the model) of around 0.7.

$$\text{adjusted } R^2 = 1 - (1 - R^2) * \frac{n-1}{n - DF_{model} - 1} \quad (1)$$

where R^2 is coefficient of determination, n is the number of observations in the sample and DF_{model} is the model degrees of freedom.

The RMSE (Root Mean Square Error)—estimator of the standard deviation in the error of the model—is 0.309.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_i')^2}{n - DF_{model} - 1}} \quad (2)$$

where y_i is the actual dependent variable and y_i' is the predicted dependent variable for the i^{th} observation in the sample.

Assuming a normal distribution of model errors, we can derive than 95% of the values are within the interval [-1.96 * 0.309 ; + 1.96 * 0.309], i.e. [-0.606 ; + 0.606]. Keeping in mind that the model is based on the logarithm of energy consumption as response variable, the ratio between the observed energy consumptions and predicted values (exponential of the residual) may be in a range between exp(-0.606) = 0.55 and exp(+0.606) = 1.83, leading to an uncertainty range [-45% ; + 83%] for the predicted values.

The coefficient of each category or each variable indicates to what extent the latter affects the outcome of the model, with all other effects being maintained constant. This model has 50 coefficients distributed over 19 variables. Their effects are analysed in the Discussion section.

The last column of Table 4 provides the standardized coefficients of this model, representing ratio between the change in response variable for a change of one standard deviation in explanatory variable or category and the standard deviation in observed response variable. The higher the absolute value of a coefficient, the greater the relative weight of this variable or this category. On the other hand, if the confidence interval around the standardized coefficient comprises zero, its weight is non-significant. By design, this is the case for all the first categories of the qualitative variables of the model. Since the sum of the standardized coefficients of the categories of a qualitative variable should be null also, by our own choice, this necessarily leads to a non-significant relative weight for one of the categories (if we know the other categories, this is automatically deduced and thus knowing it does not provide any additional information).

Table 4
Results of the model. The response variable is the natural logarithm of the yearly total final energy consumption (in kWh).

Variable	Coef.	Standard error	Lower bound (95%)	Upper bound (95%)	Pr > |t|	Standar-dized coef.
LnSurface	0.526	0.055	0.417	0.635	< 0.0001	0.302
Party_walls- detached	0.086	0.023	0.040	0.132	0.0003	0.114*
Party_walls- terraced	-0.086	0.023	-0.132	-0.040	0.0003	-0.114
Floor- on a platform	-0.061	0.019	-0.099	-0.022	0.002	-0.091*
Floor- above unheated room	0.061	0.019	0.022	0.099	0.002	0.091
Building_vintage- before 1975	0.109	0.029	0.052	0.166	0.0002	0.127*
Building_vintage- after 2000	-0.109	0.029	-0.166	-0.052	0.0002	-0.127
Outer_wall_insulation- not insulated	0.078	0.021	0.037	0.119	0.0002	0.112*
Outer_wall_insulationt- 10 cm	-0.078	0.021	-0.119	-0.037	0.0002	-0.112
Main_heating- heat pump	-0.362	0.078	-0.516	-0.208	< 0.0001	-0.163*
Main_heating- electrical heated floor	-0.213	0.091	-0.392	-0.035	0.019	-0.095
Main_heating- electrical heated panels	-0.147	0.063	-0.271	-0.024	0.019	-0.081
Main_heating- elec. convectors (<5 years)	-0.318	0.064	-0.443	-0.193	< 0.0001	-0.175
Main_heating- elec. convectors (5-10 years)	-0.151	0.062	-0.274	-0.029	0.016	-0.084
Main_heating- elec. convectors (11-15 years)	-0.211	0.076	-0.360	-0.061	0.006	-0.102
Main_heating- elec. convectors (16-25 years)	-0.195	0.062	-0.318	-0.072	0.002	-0.107
Main_heating- elec. convectors (> 25 years)	-0.192	0.082	-0.353	-0.031	0.020	-0.090
Main_heating- standard oil boiler (<10 years)	0.200	0.064	0.075	0.326	0.002	0.110
Main_heating- stand. oil boiler(11-15 years)	0.340	0.076	0.191	0.489	< 0.0001	0.165
Main_heating- stand. oil boiler (16-25 years)	0.403	0.082	0.240	0.565	< 0.0001	0.189
Main_heating- standard boiler other energy	0.408	0.079	0.253	0.563	< 0.0001	0.195
Main_heating- enclosed wood roomheater	0.438	0.065	0.310	0.565	< 0.0001	0.238
Occasionnal_supplement- yes	0.094	0.017	0.060	0.128	< 0.0001	0.165*
Occasionnal_supplement- no	-0.094	0.017	-0.128	-0.060	< 0.0001	-0.165
Exceptional_supplement- yes	0.069	0.020	0.029	0.108	0.001	0.097*
Exceptional_supplement- no	-0.069	0.020	-0.108	-0.029	0.001	-0.097
Cooking_energy- gas only	0.096	0.027	0.043	0.149	0.0004	0.104*
Cooking_energy- electricity only	-0.096	0.027	-0.149	-0.043	0.0004	-0.104
Energy_saving_lights- yes, mostly	0.067	0.023	0.023	0.111	0.003	0.086*
Energy_saving_lights- no	-0.067	0.023	-0.111	-0.023	0.003	-0.086
Climate_zone- north and Paris	0.080	0.033	0.016	0.144	0.015	0.045*
Climate_zone- east	0.174	0.036	0.102	0.246	< 0.0001	0.205
Climate_zone- southwest	-0.122	0.045	-0.211	-0.034	0.007	-0.128
Climate_ southern part of the southeast	-0.132	0.056	-0.241	-0.022	0.018	-0.122
Altitude_km	0.191	0.081	0.032	0.351	0.019	0.071
LnNum_people	0.199	0.039	0.122	0.276	< 0.0001	0.175
Age_head_household- less than 35 years old	-0.140	0.034	-0.207	-0.073	< 0.0001	-0.152*
Age_head_household- 35 à 49 years old	-0.059	0.027	-0.113	-0.005	0.033	-0.066
Age_head_household- at least 65 years old	0.199	0.032	0.135	0.262	< 0.0001	0.218
Bath_usage- less than 1/week/person	-0.051	0.024	-0.098	-0.005	0.031	-0.062*
Bath_usage- more than 1/week/person	0.051	0.024	0.005	0.098	0.031	0.062
Shower_usage-less than 7/week/person	-0.074	0.019	-0.111	-0.038	< 0.0001	-0.114*
Shower_usage- 7/week/person	0.074	0.019	0.038	0.111	< 0.0001	0.114
Living_room_temp-19 °C or lower	-0.118	0.030	-0.177	-0.060	< 0.0001	-0.115*
Living_room_temp- 23 °C or higher	0.118	0.030	0.060	0.177	< 0.0001	0.115
Night_temp_reduction- yes, always	-0.091	0.018	-0.126	-0.056	< 0.0001	-0.142*
Night_temp_reduction- no, never	0.091	0.018	0.056	0.126	< 0.0001	0.142
Ventilation_time-less than 10 min	-0.051	0.018	-0.087	-0.016	0.005	-0.081*
Ventilation_time- 10 to 29 min	0.051	0.018	0.016	0.087	0.005	0.081
Intercept	7.231	0.253	6.733	7.729	< 0.0001	
Number of observations	420					
DF of model	33					
Adjusted R^2	0.696					
RMSE	0.309					

* = the confidence interval of the standardized coefficient comprises 0 (confidence level 95%).

5. Comparison with the literature and lessons to draw

5.1. *Explanation and prediction capacities*

First, a comparison of explanation and prediction capacities of the present model with literature data is proposed here.

The established model is based on 38 explanatory variables (of which 3 are quantitative), grouped in different sets: technical (building and energy systems), geographical, socio-economic and behavioural; its adjusted R^2 is around 0.7. Most of reviewed papers presenting similar models in different countries (11; 12; 14; 15; 16-18; 20), have much lower adjusted R^2, typically in the range 0.1-0.4, so with a very limited explanation capacity. Although based on similar data sets, they generally include much less detailed information either on the building (size, shape, insulation, etc.) or on energy systems installed. Furthermore information regarding behaviour is not available. Corresponding data sets have not been collected in a specific way for developing energy bottom-up statistical model, but these models are developed *a posteriori* on existing data sets.

By introducing inhabitant practices in their data, Guerra-Santin et al. (6) found a model that helps explaining 46% of the variations in annual consumption for heating and DHW production in Dutch single-family dwellings. Only two reported models give comparable results with the present model, in terms of prediction capacity with adjusted R^2 of 0.75-0.78, using data from Austria (7) and Canada (19), including preliminary building energy rating and monitoring of consumption (7) or complete energy audit (19) (with ventilation rate measurement, e.g.). Sample size is comparable to ours (7) or much smaller (19) and larger samples of such detailed data would be very difficult to gather. Representativeness at a national scale had not been analysed and is probably questionable for the Canadian case with only 62 dwellings in the sample (19).

The former comparison with models from literature is based only on the coefficient of determination. Even with its adjusted form, R^2 is only a rough criterion for comparing different models based on totally different input data sets or different forms of the response variable. The RMSE or estimator of standard deviation of model error is a better criterion for models using the same response variable; it characterizes the statistical distribution of errors caused by non controlled phenomena not included amongst explanatory variables. But this important characteristic of a statistical model is seldom reported in literature. Actually, none of the papers analysed above provide an estimation of the errors in their model.

Anyway, the RMSE deserves some additional discussion. The RMSE value of 0.309 characterizes obviously an inaccurate prediction model: as stated above, the average 95% confidence interval for the predicted energy consumptions of each house is the interval of [-45%; +83%]. Consequently, the model is not accurate enough to predict energy consumption on the scale of one particular dwelling or of a small group of dwellings.

5.2. *Building characteristics*

One of the main determinants is the surface area, as in almost all of the literature studied (7; 14; 16). The positive sign of the coefficient (0.526) is in line with the thermal rating of the building and the cited literature (consumption increases with surface area). The number of party walls is also significant (a terraced house consumes 17.2% less than a detached house), as in the model of Guerra-Santin et al. (6). In accordance with the rules for calculating current French thermal regulations (27), a house with its floor above an unheated room has higher consumption (+12.2%) than a similar house with its floor on a platform.

Regarding building vintage, as expected, the model found that the more recent dwellings have lower consumption (houses built before 1975 consume 21.8% more than houses built after 2000). This result is identical to those provided by Cavailhes et al. (11) and Kjaerbye et al. (17). However, only the difference between dwellings built before 1975 (date of the first French thermal regulation) and the most recent dwellings (since 2000) is identified as being significant. Moreover, in a similar fashion to Kjaerbye et al. (17) for Danish data, we noted that the recorded difference in consumption (-21% between before 1975 and after 2000) is lower than expected by the theoretical assessments of the thermal regulations[4].

On the subject of dwelling insulation, as in the study by Guerra-Santin et al. (6), the model shows that 10 cm of insulation on the outer walls leads to a consumption that is 15.6% lower than that of dwellings with no insulation. Nevertheless, surprisingly, larger insulation thickness (15cm or >20cm) does not seem to lead to lower energy consumption. A comeback to the definition of what is statistically significant or not enables to remind that a non-significant coefficient does not mean "absence of effect". It means that the effect remains below some detection threshold which depends partially on sub-sample size. Here the two categories concerned

[4] -20% corresponds, for new residential buildings, to the theoretical difference in the maximum authorised consumption between the thermal regulations of 1988 and 2000. Two thermal regulations were passed between 1975 and 1988 (respectively RT 1974 and RT 1981).

represent only a small part of sample size (respectively 10% and 5%) and this fact probably explains the non-significance encountered.

Roof insulation is considered to be non-significant as in Cavailhes et al. (11) but unlike in the papers by Risch and Salmon (12). Amongst the other variables relating to the thermal rating of the dwelling, the model considers the presence of double glazing to be non-significant, this time contradicting Cavailhes et al. (11) as well as Guerra-Santin et al. (6) but similar to Risch and Salmon (12).

5.3. *Characteristics of the energy systems*

The type of main heating system is a very important variable for explaining consumption. According to the model, a wood-burning stove or standard oil boilers or boilers using other energies result in higher consumption, in terms of final energy, than electric heating systems (convectors, heating panels, heating floors and heat pumps). According to the theoretical efficiency of the various systems[5], these observations appear to be logical, in particular the fact that wood-burning stoves and heat pumps are the systems that result, respectively, in the highest (43.8% higher than average) and lowest (36.2% lower than average) consumption levels. Part of the differences, as far as electrical heating systems are concerned, may be also explained by higher insulation levels required by thermal regulation when these systems are initially installed, leading to lower final energy consumption if compared with fuel heating systems.

Furthermore, this variable confirms that, overall, the older the heating system, the higher its consumption (e.g., a standard oil boiler 16 to 25 years old consumes 20.3% more than an equivalent boiler less than 10 years old). This observation is identical to the results of Kjaerbye et al. (17). A similar "age" effect is observed for electrical convectors (e.g., equipments older than 25 years old consume 12.6% more than ones less than 5 years old), while the calculation models used to diagnose energy efficiency in France (28) do not consider this effect. This age effect is likely to be linked to technological developments regarding convector control and comfort.

The presence of occasional (more than 15 days per year) or exceptional (up to 15 days per year) supplementary heating is considered to be a source of increased consumption (+18.8% and +13.8%, respectively). These results match those of Cavailhes et al. (11) with a supplementary fireplace but not those of Guerra-Santin et al. (6), which have a non-significant "supplementary heating" variable. Also contradicting Guerra-Santin et al. (6), we found that the presence of a temperature control system is not a determinant for consumption.

As regards energy systems for other end-uses, the type of system used for producing **DHW** is not significant, in a similar fashion to Ndiaye and Gabriel (19).

Concerning end-uses linked to **specific electricity**, it was found that a house without energy-saving light bulbs has lower energy consumption (-13.4%) than the same house using mostly this type of bulbs. This effect is not confirmed by Ndiaye and Gabriel (19), who found the number of energy-saving bulbs to be non-significant. In other respects, Wallenborn and Dozzi (29) highlight that amongst the households having a high environmental consciousness, so likely to have energy-savings light bulbs, there are predominantly people having the greatest incomes, the highest education level and also the highest energy consumptions notably due to great electrical appliance ratio. Thus, the variable "Energy_saving_lights" could hide an "income" effect and/or an effect of electrical appliance ratio (not selected in the model).

Finally, the energy used for **cooking** is significant in the model (cooking using only electricity leads to consumption that is 19.2% lower than when cooking using only gas). This value seems too high for a secondary end-use such as cooking, which represented in 2008, less than 10% of the final consumption of the residential sector in France (30). This variable hides probably others effects than cooking equipments or cooking practices. Anyway, this explanatory variable is the only one for which unexpected effects were found and no reason was identified.

5.4. *Characteristics of refurbishment*

According to the model obtained, the variable relating to energy-saving or improvement works conducted during the previous 12 months is not significant for explaining the consumption of a dwelling. Ndiaye and Gabriel (19) obtained an identical result, while Kjaerbye et al. (17) and Rehdanz (14), for a new heating system, showed reduced consumption due to the retrofit performed. Thus, due to the scarcity of information at our disposal regarding the refurbishment performed, we could not draw any conclusions regarding the influence of this type of works.

---

[5] According to (28), the generation performance of electric convectors or electric heating panels or of an electric heating floor = 1, generation performance (on LHV) of a standard oil or LPG boiler between 0.67 (installed before 1988) and 0.87 (installed after 2000), generation performance of a heat pump between 2.2 (air/air heat pump) and 4 (air/water heat pump), overall efficiency of a wood-burning stove = 0.5.

Otherwise, through an *ex-post* billing analysis of an energy efficiency scheme, Scheer et al. (31) show that after refurbishment, the energy demand of the retrofitted houses is approximately equal to the rest of the stock because the before refurbishment energy efficiency of these houses is much lower than this of the others. This result could explain the non effect of energy-saving or improvement works when the statistical analysis is done on energy consumption of one year.

This shows us the usefulness of working with panels of dwellings that have undergone energy-saving works and provide precise information from before and after the works, if we want to study the effects of energy-efficiency actions (32).

5.5. *Geographical characteristics*

With the "Climate_zone" variable in the model, we found results similar to those of Risch and Salmon (12). Namely, as an expected result, the coldest zones in winter (H1b – east and H1a – north and Paris) have higher consumption than milder zones (H2c – southwest and H2d – southern part of the southeast)[6]. While this matches the findings of Meier and Rehdanz (15), the recorded effects are far from being weak as they are for the latter. Due to large diversity of climates in France compared to most of European countries, regional sensitivity of energy consumption is probably less in other countries.

Anyway, the effect of climatic zones is more limited than expected from heating degree-days values; insulation requirements from thermal regulation are higher in colder climates and this fact reduces the effect of climate zones in the model.

Furthermore, the model logically estimates that the higher the altitude at which the dwelling is located, the higher its consumption (+19.1% per 1000 m).

As regards urban density, it is not significant in the model, while Kjaerbye et al. (17) found that it had a significant effect. Its effect may be partially hidden by the variable regarding party walls, which was not included in the model of Kjaerbye et al. (17).

5.6. *Socio-economic characteristics of households*

Socio-economic variables are also amongst the main determinants for the consumption of a single-family dwelling. As in the vast majority of the literature (13; 16; 20), the model estimated that the consumption of a household increases with the age of the head of the household, with a particularly marked effect for people aged 65 and older (33.9% higher than the group of under 35 and 25.8% higher than the group aged 35 to 49). It was also found that the greater the number of people living in the household, the greater the consumption (+19.9% for each additional person). This is in line with Rehdanz (14), Risch and Salmon (12) and Schuler et al. (20).

Unlike these latter studies and Sardianou (16), occupation status is not significant in the model.

The models of Risch and Salmon (12) as well as Cavailhes et al. (11) show that household income has a significant effect on consumption, while we found a non-significant effect. This lack of significance for household income can also be found in the works of Ndiaye and Gabriel (19). It appears to us that it may, to a large extent, be explained by the fact that income is, *inter alia*, a “proxy” of the household practices and thus, as soon as a certain number of behaviour variables are found in the models, this variable loses a large amount of the information that it contains.

5.7. *Behavioural characteristics*

Indeed, the model contains several behaviour variables, not included in most other studies reported, like other French studies (10-13). As regards behaviour relating to space heating, it has been verified that the higher the temperature setting in the main living room, the higher the consumption (consumption with a temperature of 19°C or less is 23.6% lower than with a temperature of 23°C or more). This observation matches the observations of Haas et al. (7) and Guerra-Santin et al. (6). Like the latter reference, never reducing the heating temperature during the night leads to a higher consumption (+18.2%) compared with reducing the temperature every night.

As for ventilation practices, in a more or less similar manner to Ndiaye and Gabriel (19) and Iwashita and Akasaka (33), ventilation time seems to be a determinant. Thus, ventilation periods of 10 to 29 min on average lead to higher consumption (+10.2%) than ventilation periods of less than 10 min. Larger ventilation periods (30-59mn or >60mn) do no lead to significant higher energy consumption. Again (as for insulation thickness), small

[6] Heating degree-days (base 18°C) for the different zones: H1b – east: 2852 HDD (Nancy), H1a – north and Paris: 2688 HDD (Trappes), H2c – southwest: 2192 HDD (Agen) and H2d – southern part of the southeast: 2054 HDD (Carpentras).

sub-sample size for each category (10% and 6% respectively) is likely to explain that they cannot be found significantly different from the zero-average of the coefficients.

Furthermore, it was found—as expected—that the higher the number of showers taken, the higher the consumption (a household in which less than seven showers are taken per person per week consumes 14.8% less than a household in which seven showers are taken per week per person). The same influence is determined for the number of baths taken (a household in which less than one bath is taken per person per week consumes 10.2% less than a household in which more than one bath is taken per week per person).

Finally, according to the model, leaving the light on in empty rooms is not significant. This is in line with the lack of significance found for leaving the lights on when leaving a room for a short time in the model of Ndiaye and Gabriel (19). Using the lights during the day is also considered to be non-significant.

5.8. *Weight of variables*

After the identification of the main determinants for energy consumption, we will now look at the most influential factors amongst them. For this purpose, absolute values of the standardized coefficients of the model, introduced in section 4.2, are used as criteria. For qualitative variables, only the maximum standardized coefficient value amongst the various modalities of the variable is considered.

In this respect, the natural logarithm of the surface area is the variable with the greatest weight (standardized coefficient 0.30), then, by decreasing order (the absolute value of the standardized coefficient, for the most important category for qualitative variables):

- main heating system (enclosed wood roomheater, 0.24; but four other categories have standardized coefficient higher than 0.15);
- age of the household head, (at least 65 years old, 0.22);
- climate zone (east, 0.21);
- Logarithm of the inhabitants number (0.18);
- occasional supplementary heating (0.17);
- night temperature setting reduction (0.14);
- building vintage (0.13).

Six other variables have absolute values of standardized coefficient in the range [0.10; 0.12], two of which being behavioural characteristics (living room temperature and shower usage).

Amongst literature cited in the present paper, only Guerra-Santin et al. (6) presented the standardized coefficients of their model. Comparison with Guerra-Santin et al (6) shows rather different results. The surface area is found one of the most influential factors as well, but the type of heating system is not part of their set of explanatory variables (not statistically significant during the screening of variables) and the age of the respondent is not observed by them as a major determinant. Moreover, amongst the household characteristics and behaviour variables found significant by Guerra-Santin et al. (6), the most influential is the number of heated bedrooms while temperature during the night is considered to have little influence.

5.9. *Accuracy of the model coefficients*

Accuracy of model coefficients is another item to be discussed, in order to determine how the model could be relevant for sensitivity analysis of average energy consumption with respect to the different parameters. In Table 4, uncertainties of the coefficients are characterized by standard errors and 95% confidence intervals (coefficient ± 1.96 * standard error). The three more accurately estimated coefficients (highest absolute value of t-values, ratios between the estimated coefficients and theirs standard errors) are indicated below with their 95% confidence intervals, estimated values (from Table 4) and relative uncertainties (ratio between half of 95% confidence interval and estimated value):

- LnSurface (Ln of surface area) [0.42; 0.64] 0.53 ±21% (t-value 9.5)
- Main_heating- wood roomheater [0.31; 0.57] 0.44 ±29% (t-value 6.7)
- Age_head_household- at least 65 years old [0.14; 0.26] 0.20 ±32% (t-value 6.1)

All other coefficients have t-values lower than 5.5 and relative uncertainties equal or larger than ±36%. With such uncertainties, it is only possible to check orders of magnitude and physical meaning of coefficients. Use for precise quantitative assessment (of the impact of such category of space heating system or type of behaviour, e.g.) is not relevant. Decreasing the relative uncertainties of the main parameters to ±10% - ±20% would be necessary before planning quantitative application of the model, typically reducing standard errors (and increasing t-values) by a factor 2 or more. With the usual statistical rule of thumb of standard errors being proportional to the inverse of square root of sample size, a complete data set of 1600 households or more (instead of 420) would be necessary to reach this accuracy target.

Furthermore, larger samples could make possible to identify models accounting for interactions between the main explanatory variables. Sensitivity analysis of outputs from energy engineering models (26) clearly showed the importance of interactions between the main explanatory variables when they are defined in a rather extended range: size parameters, insulation level, ventilation rate and set-point temperature for space heating. This result emphasizes the limits of simple linear models for statistical prediction. Incorporating major interactions could improve the accuracy of the prediction model, leading to smaller RMSE and, due to positive feedback, further reduced standard errors for the coefficients.

These results should be kept in mind when specifying future surveys to improve bottom up statistical models. However, based on the most significant coefficients of the model and their orders of magnitude, the present results enable to draw lessons regarding the main determinants for energy consumption in the French residential sector thanks to the high overall significance of the model.

6. Conclusion

This study is based on a sample of 420 single-family dwellings occupied by households, representative of the French population. One of the goals was to determine the explanation and prediction capacities of a bottom-up statistical model regarding the total energy consumption of these dwellings, starting from an extended set of explanatory variables: technical (building, energy systems and refurbishment), geographical, socio-economic and behavioural variables. A set of explanatory variables with such an information quantity is present in very few literature studies. The others goals were to identify the main determinants of energy consumption in the French residential sector and possibly to rank them.

The created model has an explanation capacity (adjusted R^2 of around 0.7), close to the best found in the literature, keeping in mind that no energy audit or monitoring was available (only questionnaire data provided by occupants). Moreover, the explanation capacity obtained is higher than in studies conducted using similar sets of explanatory variables but without information on behaviour. Nevertheless, by analysing the estimations of errors made by the model, seldom reported in literature, we determined that the latter was not suitable for predicting consumption on the scale of one dwelling. Indeed, the 95% confidence interval for the predicted energy consumption of one house is too large.

Despite of its low prediction capacity but thanks to its high overall significance, such statistical model may be used as rough benchmark before energy audits for a first selection of houses to be handled in priority: houses with high residuals (whatever positive or negative) are likely to be associated to missing or erroneous information.

The main determinants for energy consumption in the single-family dwelling sector in France have been examined. Most of effects identified are consistent with effects either predicted by engineering models or found in the bottom-up statistical literature. However, the difference between theory and practice for certain results is highlighted (e.g. efficiency of thermal regulations). It also stressed the lack of significance of household income when the models have a certain amount of behaviour information. Thanks to the database to which we had access, we were also able to analyse certain practices of households, which is something rarely found in the literature.

Also infrequently done in literature, we have looked at the most influential variables amongst the main determinants identified. The variables having the greatest weights in the model are surface area, type of main heating system, age of the household head and climate zone. We detected that the most influential behaviour variable is night temperature setting reduction. Under the classification obtained from the consequent set of explanatory variables studied, the most influential variables are technical. However, behaviour variables and variables of socio-economic characteristics household, which are for a great part “proxy” of household behaviours, are preponderant determinants too.

Finally, the study confirmed that in order to study the effects of energy efficiency actions it made sense to use dedicated panels of dwellings that have undergone energy-saving measure, with precise information before and after refurbishment, allowing us to draw unambiguous conclusions.

In terms of outlooks, an analysis of the results seen from the point of view of energy-savings potentials and energy policy could be proposed. To verify the robustness of results obtained, a study with behaviours measured or monitoring instead declared (less uncertainty) could be interesting. For identifying more robust statistical model in terms of prediction capacity, we would need of a bigger data sample which could enable both to decrease standard errors of coefficient (with an objective of reduction by a factor 2, quantification of suitable data set size has been done) and to account for the major interactions between explanatory variables. At last, a similar study should be realized on sample of apartments, representative of the French households living in collective dwelling, for knowing all the determinants of the French dwelling stock.

References

[1] The Parliament and the Council, Directive 2010/31/EU on the energy performance of buildings, The European Parliament and the Council of the European Union, 19 May 2010.

[2] Europa, Energy efficiency for the 2020, 29 March 2009. http://europa.eu/legislation_summaries/energy/energy_efficiency/en0002_en.htm (accessed on June 05, 2012).

[3] Ministère de l'écologie, du développement durable et de l'énergie, Efficacité énergétique : la France soutient activement la proposition de directive européenne, 16 April 2012, http://www.developpement-durable.gouv.fr/Efficacite-energetique-la-France.html (accessed on June 05, 2012)

[4] Service de l'Observation et des Statistiques, Bilan énergétique de la France pour 2010, Commissariat général au développement durable, Références, 2011.

[5] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (2009) 1819-1835.

[6] O. Guerra-Santin, L. Itard, H. Visscher, The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock, Energy and Buildings 41 (2009) 1223-1232.

[7] R. Haas, H. Auer, P. Biermayr, The impact of consumer behavior on residential energy demand for space heating, Energy and Buildings 27 (1998) 195-205.

[8] A. Audenaert, K. Briffaerts, L. Engels, Pratical versus theoretical domestic energy consumption for space heating, Energy Policy 39 (2011) 5219-5227.

[9] G. Branco, B. Lachal, P. Gallinelli, W. Weber, Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data, Energy and Buildings 36 (2004) 543-555.

[10] S. Couture, S. Garcia, A. Reynaud, Household energy choices and fuelwood: An econometric approach using French data, Energy Economics 34 (2012) 1972-1981.

[11] J. Cavailhes, D. Joly, T. Brossard, H. Cardot, M. Hilal, P. Wavresky, La consommation d'énergie des ménages en France, Convention MEEDM - INRA no. 22 0009 3249, 2011.

[12] A. Risch, C. Salmon, French residential energy demand: Micro-econometric analysis of household multi-fuel energy consumption, Note de recherche (http://www.irege.univ-savoie.fr) (Also published in: 5th Annual International Symposiums on Economic Theory, Policy and Applications, Athens, Greece, 2010), 2010.

[13] L. Calvet, F. Marical, Le budget "énergie du logement" : les déterminants des écarts entre les ménages, Commissariat Général au Développement Durable (Eds), Le point sur 56 (2010).

[14] K. Rehdanz, Determinants of residential space heating expenditures in Germany, Energy Economics 29 (2007) 167-182.

[15] H. Meier, K. Rehdanz, Determinants of residential space heating expenditures in Great Britain, Energy Economics 32 (2010) 949-959.

[16] E. Sardianou, Estimating space heating determinants: An analysis of Greek households, Energy and Buildings 40 (2008) 1084-1093.

[17] V.H. Kjaerbye, M. Togeby, A.E. Larsen, Do changes in regulatory requirements for energy efficiency in single-family houses result in the expected energy savings?, Proceedings of the ECEEE Summer Study 2011, Presqu'île de Gien, France, 2011.

[18] S. Leth-Petersen, M. Togeby, Demand for space heating in apartment blocks: measuring effects of policy measures aiming at reducing energy consumption, Energy Economics 23 (2001) 387-403.

[19] D. Ndiaye, K. Gabriel, Principal component analysis of the electricity consumption in residential dwellings, Energy and Buildings 43 (2011) 446-453.

[20] A. Schuler, C. Weber, U. Fahl, Energy consumption for space heating of West-German households: empirical evidence, scenario projections and policy implications, Energy Policy 28 (2000) 877-894.

[21] Institut National de la Statistique et des Etudes Economiques, Statistical operation: Housing survey in 2006, http://www.insee.fr/en/methodes/default.asp?page=sources/ope-enq-logement.htm (accessed on January 24 2014).

[22] TNS SOFRES, Le panel postal: Métascope, http://www.tns-sofres.com/notre-offre/outils/25096226A30148689BD36FC625260267.aspx (accessed on June 05, 2012).

[23] J-M. Cayla, N. Maizi, C. Marchand, The role of income in energy consumption behaviour: Evidence from French households data, Energy Policy 39 (2011) 7874-7883.

[24] J-M. Cayla, B. Allibe, M-H. Laurent, From pratices to behaviors: Estimating the impact of household behavior on space heating energy consumption, Proceedings of the ACEEE Summer Study 2010, Pacific Grove, United States, 2010.

[25] Institut National de la Statistique et des Etudes Economiques, Bases de données, http://www.insee.fr/fr/bases-de-donnees/ (accessed on June 01, 2012).

[26] D. Garcia Sanchez, B. Lacarrière, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, Energy and Buildings 68 (2014) 741–750.

[27] Réglementation thermique 2005, Règles Th-U - Fasciule 4, 2006.

[28] Ministère de l'emploi, de la cohésion sociale et du logement, Arrêté portant approbation de diverses méthodes de calcul pour le diagnostic de performance, 9 Novembre 2006.

[29] G. Wallenborn, J. Dozzi, Du point de vue environnemental, ne vaut-il pas mieux être pauvre et mal informé que riche et conscientisé ?, in: P. Cornut, T. Bauler, E. Zaccaï, Université de Bruxelles (Eds.), Environnement et inégalités sociales, Bruxelles, 2007, pp. 47-59

[30] Agence de l'Environnement et de la Maîtrise de l'Energie, Chiffres clés – Energie et climat – édition 2009, 2009

[31] J. Scheer, M. Clancy, S. N. Hógáin, Quantification of energy savings from Ireland's Home Energy Saving scheme: an ex post billing analysis, Energy Efficiency (2012)

[32] M. Raynaud, D. Osso, J. Adnot, B. Bourges, B. Duplessis, An energy efficiency program analysis to understand the gaps between ex-ante and ex-post evaluations, Proceeding of the 2012 International Energy Program Evaluation Conference, Rome, Italy, 2012.
[33] G. Iwashita, H. Akasaka, The effects of human behavior on natural ventilation rate and indoor air environment in summer - a field study on southern Japan, Energy and Buildings 11 (1997) 267-275.
[34] Centre d'Etudes et de Recherches Economiques sur l'Energie, Secteur résidentiel - Suivi du parc et des consommations d'énergie - Consommations en 2008, Etude 9102, 2009.