ENSG Géomatique

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES

CSTB le futur en construction

Rapport de Thèse professionnelle

Cycle : Mastère Spécialisé DESIGEO

Study of Multilevel Modelling Approach of Household Energy Consumption in France

&

Data Visualisation with R-Shiny

insérer ici une image (8 x 10 cm minimum et 10 x 15 cm maximum) et effacer ce texte

Rémy Zumbiehl

le 29/05/2017

☐ Non confidentiel ☐ Confidentiel IGN ☐ Confidentiel Industrie ☐ jusqu'au ……

ECOLE NATIONALE DES SCIENCES GEOGRAPHIQUES
6 et 8 avenue Blaise Pascal - Cité Descartes - Champs sur Marne - 77455 MARNE-LA-VALLEE CEDEX 2
Téléphone 01 64 15 31 00 Télécopie 01 64 15 31 07

Jury

Président(e) du jury

Commanditaire :

Prénom PATRONYME,

Développé de l'organisme (SIGLE)

adresse

Encadrement de stage :

Fateh BELAÏD, CSTB, maître de stage

Emmanuel FRITSCH, unité, ENSG/IGN, rapporteur principal

Prénom PATRONYME, organisme (facultatif)

Prénom PATRONYME, unité, ENSG/IGN (facutatif)

Responsable pédagogique du cycle :

Anna CRISTOFOL , ENSG Cycle Mastère Specialisé DESIGEO

Stage du 12/09/16 au 15/05/17

Diffusion Web : □ Internet □ Intranet ENSG

Situation du document :

rapport de stage présenté en fin du cycle du Mastère Spécialisé DESIGEO

Nombre de pages : xx dont xx d'annexes

Système hote : Word

Remerciements

Insérer ici votre texte de remerciements

Abstract

As many other countries, France is facing the challenge of modelling its residential energy consumption. Modelling residential energy consumption is essential to be able to understand national energy problematic, to predict future trends, thus to be prepared to adapt future French policies and legislation in order to meet energy efficiency requirements at global and European levels. Most of residential energy consumption models created in the past, and based on datasets taken from various surveys, were built under standard multiple regression frameworks with usual variables taken in account for explaining the residential energy consumption. Those standard models are generally not capable of capturing more than 55% of the residential energy consumption variance. A multilevel regression model (MRM) offers an interesting approach in the comprehension and modelling of residential energy consumption (REC) based on the dataset results of the “Phebus” national survey. This dataset is consisting of 2090 unique cases distributed within 81 geographical administrative divisions called “départements” (DEP). MRM offers the possibility of extracting area effects from total variation of REC and explaining the remaining variation with relevant explanatory variables and their interactions. Multilevel Regression Models can answer the following question: Is the geographical context influencing the residential household energy consumption? The study showed the ability and effectiveness of the MRM to quantify 12% of area effects (aggregated level) and 55% of household effects (individual level).

Table des matières

Table des matières

Pour générer la table des matières, veuillez suivre la procédure suivante :

1. Ouvrir le menu "Références" – "Table des matières"
2. Choisir un type prédéfini ou choisir "Insérer une table des matières…"
3. Choisissez le nombre de niveau à afficher, le format utilisé, les caractères de suite,…Pour faire une mise à jour, cliquez sur le raccourci "Mettre à jour la table"

N'oubliez pas d'effacer ce texte quand vous n'en aurez plus besoin.

Liste des tableaux

Pour générer la liste des tableaux, veuillez suivre la procédure suivante :
1. Ouvrir le menu "Références" – "Insérer une table des illustrations"
2. Choisir la légende "tableau" et la mise en forme désirée, puis faire OK
3. Pour faire une mise à jour, cliquez dans la table et appuyez sur la touche "F9"

N’oubliez pas d’effacer ce texte quand vous n’en aurez plus besoin.

Liste des figures

Pour générer la liste des figures, veuillez suivre la procédure suivante :
1. Ouvrir le menu "Références" – "Insérer une table des illustrations"
2. Choisir la légende "figure" et la mise en forme désirée, puis faire OK
3. Pour faire une mise à jour, cliquez dans la table et appuyez sur la touche "F9"

N’oubliez pas d’effacer ce texte quand vous n’en aurez plus besoin.

Liste des équations

Pour générer la liste des figures, veuillez suivre la procédure suivante :
1. Ouvrir le menu "Références" – "Insérer une table des illustrations"
2. Choisir la légende "équation" et les caractères de suite désirés, puis faire OK
3. Pour faire une mise à jour, cliquer dans la table et appuyer sur la touche "F9", choisir l'option désirée et faire OK

N'oubliez pas d'effacer ce texte quand vous n'en aurez plus besoin.

Liste des annexes

Pour générer la liste des annexes, veuillez suivre la procédure suivante :
1. Ouvrir le menu "Références" – "Insérer une table des illustrations"
2. Choisir "Options" et le style "3|Ann_titre1" pour construire la table, puis faire OK
3. Pour faire une mise à jour, cliquez dans la table et appuyez sur la touche "F9"

N’oubliez pas d’effacer ce texte quand vous n’en aurez plus besoin.

Glossary

CSTB	Centre Scientifique et Technique du bâtiment
INSEE	Institut National de Statistiques et Etudes Economiques
DEP	Département
REG	Région
HDD	Heating Degree Day
MRM	Multi level Regression model
REC	Residential Household Energy Consumption
GeS	Greenhouse Gas Emission

N’oubliez pas d’effacer ce texte quand vous n’en aurez plus besoin.

INTRODUCTION

Accounting for more than 30% of the total energy consumption in France, the residential sector is consuming more energy than the industrial sector and almost a similar amount than the transportation sector. Moreover, residential sector is contributing for more than 16% of national CO_2 emissions, hence representing a high potential for energy efficiency incentive measures in order to mitigate greenhouse gas emission (GeS). As housing units built prior 1975 represent 61% of the housing stock across the country, this particular group of housing units constitutes the primary target for housing refurbishment programs.

Electricity and gas constitute the two main sources of energy consumed by households in France, and electricity used for space heating represents more than 60% of household energy consumption. Noticeable improvements were made on space heating technology in recent years with the use of more efficient space heating systems, such as condensing boilers and heat pumps, along with the quality enhancement of housing insulation materials. Furthermore housing refurbishment programs reduced significantly the yearly household energy consumption to approximately 180 kWh/m^2 in 2012.

In recent years, energy consumed for space heating represents the type of energy consumption that has reduced the most significantly across France. The amount of energy consumed for space heating has decreased by 33% since 1990, but at the same time the amount of energy consumed specifically for electrical appliances has increased by 40% due to the use of numerous new home electronics devices (smartphones, electronic tooth brushes, …). In order to pursue the global reduction effort on household energy consumption and adopt incentive measures in this sense, a recent policy framework has set up the objective of reducing by 40% the amount of residential energy consumption by 2030 [1].

In this context, and with the aim of adapting framework policies to enhance energy efficiency within the residential sector, conducting and analysing a national energy consumption survey that would provide detailed information on households characteristics and their impacts on the energy consumption has proven to be particularly judicious.

---

[1] Loi Relative à la Transition Energétique (2015)

Such a study would allow for better understanding of the households energy consumption, and for modelling and predicting energy consumption according relevant households characteristics.

Structure of the paper

Part 1 is introducing a brief description of the CSTB, while Part 2 is providing a detailed description of Phebus dataset. It is reminded here that the results of the national Phebus survey constitutes the base of all modelling approaches that will be built for the purpose of explaining variations of residential energy consumption among households in France. Part 2 will deliberate on the motivation for applying multilevel regression analysis and will particularly focus on data preparation. In Part 3, various step-by-step models will be carried out, and a first standard multiple regression model approach, using only household features at micro-level (individual) will be proposed. The advantage of starting with a standard multiple regression model is that, following a selection of relevant predictors that are supposed the explain residential household energy consumption (REC), it will also point out the limitation of explanation power of such model and will therefore motivate the use of a new modelling approach like multilevel regression analysis. The step-by-step work will be pursued, starting with building “null models” in Part 3, aiming to select which grouping variable would represent the best aggregated level (level-2) to be considered for a multilevel regression model (DEP or REG). Once a level-2 grouping variable is selected, a third model introducing level-2 predictors will be elaborated, again in part 3. An estimation of the explanation power of the third model compared to the null model will give a batch of results that will confirm or not the selection of the level-2 explanatory variables. Again in Part 3, a fourth model will be built, first incorporating some of level-1 predictors used with the standard regression model, then together with level-2 predictors, prior estimating the significance of such models. Building a model with all level-1 and level-2 explanatory variables will then form an “almost full” multilevel model capable, hopefully, to demonstrate the effects of environment or contextual indicators and specific household characteristics on residential energy consumption. All models elaborated in Part 3 will contain methodological discussions and empirical specifications. Following a cautious conclusion on the results obtained after the use of a multilevel regression model, some data visualization maps built using R-Shiny framework will be proposed on a few appendixes.

PART 1 : CSTB – A KEY ACTOR

1.1 Presentation

CSTB stands for Technical and Scientific Center dedicated for Building.

The organisation is an EPIC ("Etablissement Public à caractère Industriel et Commercial"), which is a category of public undertaking in France. It includes state control entities of an industrial or commercial nature, including research institutes and infrastructure operators such as RATP, IFREMER, ONERA, BRGM.

The CSTB was founded in 1947 following World War II, aiming to support the reconstruction effort. The main mission of the CSTB is to ensure quality and safety of the buildings, and support innovation from the idea to the market. There are approximately 900 people working within four CSTB sites in France (Champs-sur-Marne, Sofia-Antipolis, Nantes and Grenoble).

1.2 Four keys activities

The CSTB is focusing its effort on four key activities: research and consulting, assessment, certification, and diffusion of information.

Diffusion of Information

The CSTB is rending scientific and regulatory information accessible and directly usable trough edited products and information services, product softwares and adapted training sessions to companies. Hence it contributes to the sharing of knowledge of professionals in relation with performance stakes of a building, evolutions of regulations, and progression of innovation.

Certification

Certification is a process allowing the characteristics of an offer to comply within a reference framework. This quality label is essential as it provides the buyers and users with confidence

when comparing and selecting a new offer available in the market. Moreover this certification process provides actors (promoters, constructors, independent tradesmen such as plumbers, electricians, painters, carpenters,) with a medium that differentiate their offer from others offers in competition. Certifying organization accredited, the CSTB is a key actor in the certification of products and building services.

Assessment

Innovation assessment by CSTB provides building actors with some crucial information regarding level of performance of processes used, materials, or any elements or equipment involved in the building contraction process. CSTB delivers guidelines to building actors, thus privileging emergence of innovations and access to the building market, while securing and re-assuring them. Moreover, the CSTB offers assessment services to construction companies wishing to develop innovations on the market. On a European scale, CSTB is a technical assessment certified organism, which is guiding certification processes and delivering CE label.

Research and expertise

The CSTB is focusing its research efforts in priorities domains. It mobilises its expertise to support framework policies and assist building professional actors. Also it develop a systemic approach including overall socio-economic stakes regarding safety, health and comfort, environment and energy. Research work in the CSTB is sometimes carried with the cooperation of the “Ministère de l’Enseignement Supérieur et de la Recherche”, and is generally financed through European Union partnerships, national programs and various socio-economic actors.

Based on the knowledge acquired from past research studies and perpetual innovation assessment, activity of CSTB expertise is also leaning on a deep knowledge of professional buildings actors.

At a national, European, and international level, the CSTB is participating to the normalization and technical regulations related to building construction.

PART 2 : DATA & MODELLING APPROACH

2.1 Phebus Dataset

2.1.1 Description of dataset

Phebus is a new punctual national survey, implemented by INSEE (National Institute of Statistic and Economical Studies) and including two sections realised separately: a face to face interview with housing occupants randomly selected, with questions regarding their house equipment energy- consuming, global energy consumption and attitude vis-à-vis energy, and another section which comprises a diagnosis of energy efficiency of the housing unit. The objective of Phebus national survey is to deliver a clear photography of households energy use within French metropolitan housing stock in 2012. The 2012 Phebus dataset consists of observations taken from 2356 housing units selected to represent the 27.6 million housing units that are occupied as a primary residence. Only housing units corresponding to an individual house, housing units located inside building, independent rooms inside buildings with private entrance, and home dedicated to elderly people are taken in consideration during data collection and survey. Phebus survey has been conducted from April to October 2013, across 81 “départements” (DEP) and 12 “regions” (REG) within French metropolitan territory. No data is available for Corsica. The data was collected using an area-probability sampling scheme, coming from national census data collection in France, and is representative of housing units across regions, climatic zones, housing type (insulated house or multi-unit housing) and year of housing construction.

2.1.2 Geographical divisions in France

Since 2015, French territory inside Europe is divided in 13 administrative regions (12 metropolitan regions plus Corsica region). The scope of intervention of French administrative regions is quite wide as it concerns among other things school and transport administrations, economic development, tax system, sustainable development policy, and biodiversity protection. French “départements” constitutes another essential subdivision of France metropolitan territory. There are 96 “départements” located in French territory inside Europe, in both metropolitan regions and Corsica. The subdivision “département” represents a level comprised in-between levels “Régions” and level “Arrondissement”, last one being another smaller subdivision of level ”département”. From now on, let’s consider level “”département”

as level DEP and level “region” as level REG. From a general point of view, one REG contains various DEP and one DEP is subdivided in various “arrondissements”. The two levels REG and DEP will form the two possible aggregated levels that could be considered for a multilevel approach.

Let’s investigate how the subdivisions DEP are represented among the households interviewed during Phebus survey. The below figure is a “bar plot” type of graphic summarizing the number of households interviewed within each “department” in France where the survey took place. Each bar corresponds to a household interrogated during Phebus survey. The amount of energy consumed is corresponding to a yearly energy consumption scale at the bottom of each bar plot. The black vertical line is related to the national mean amount of energy consumed per household (18000 kWh).

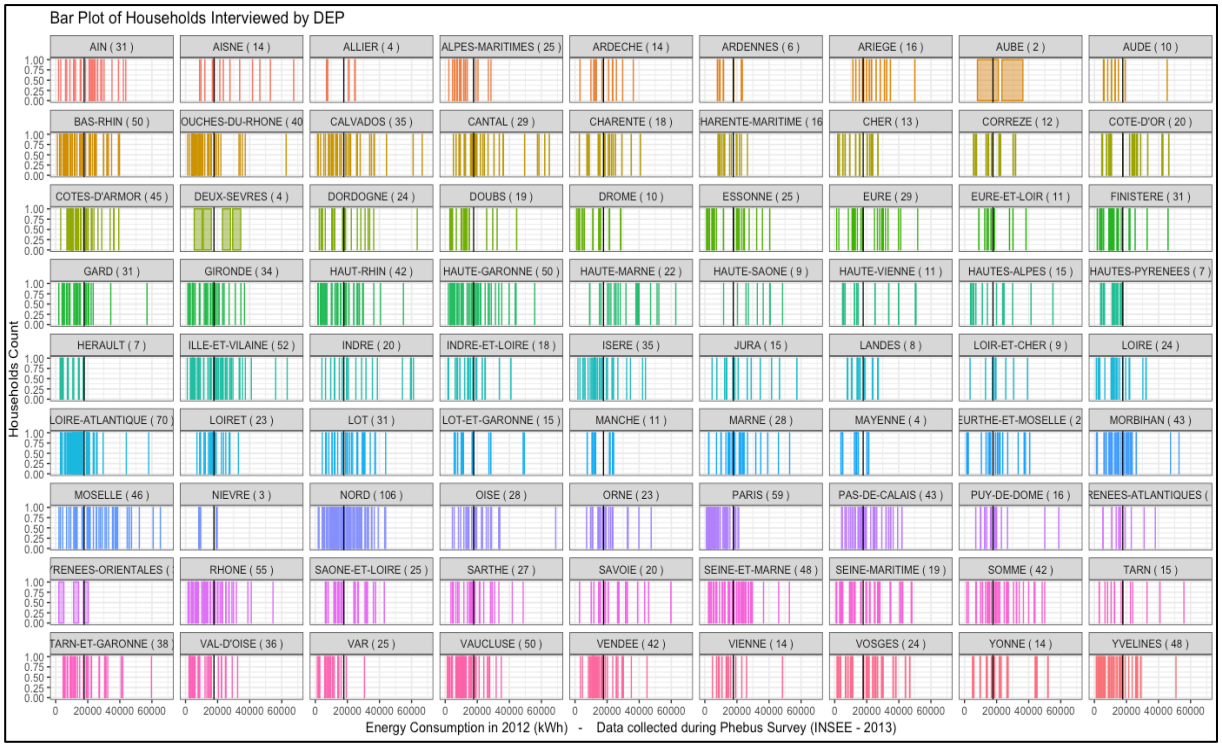


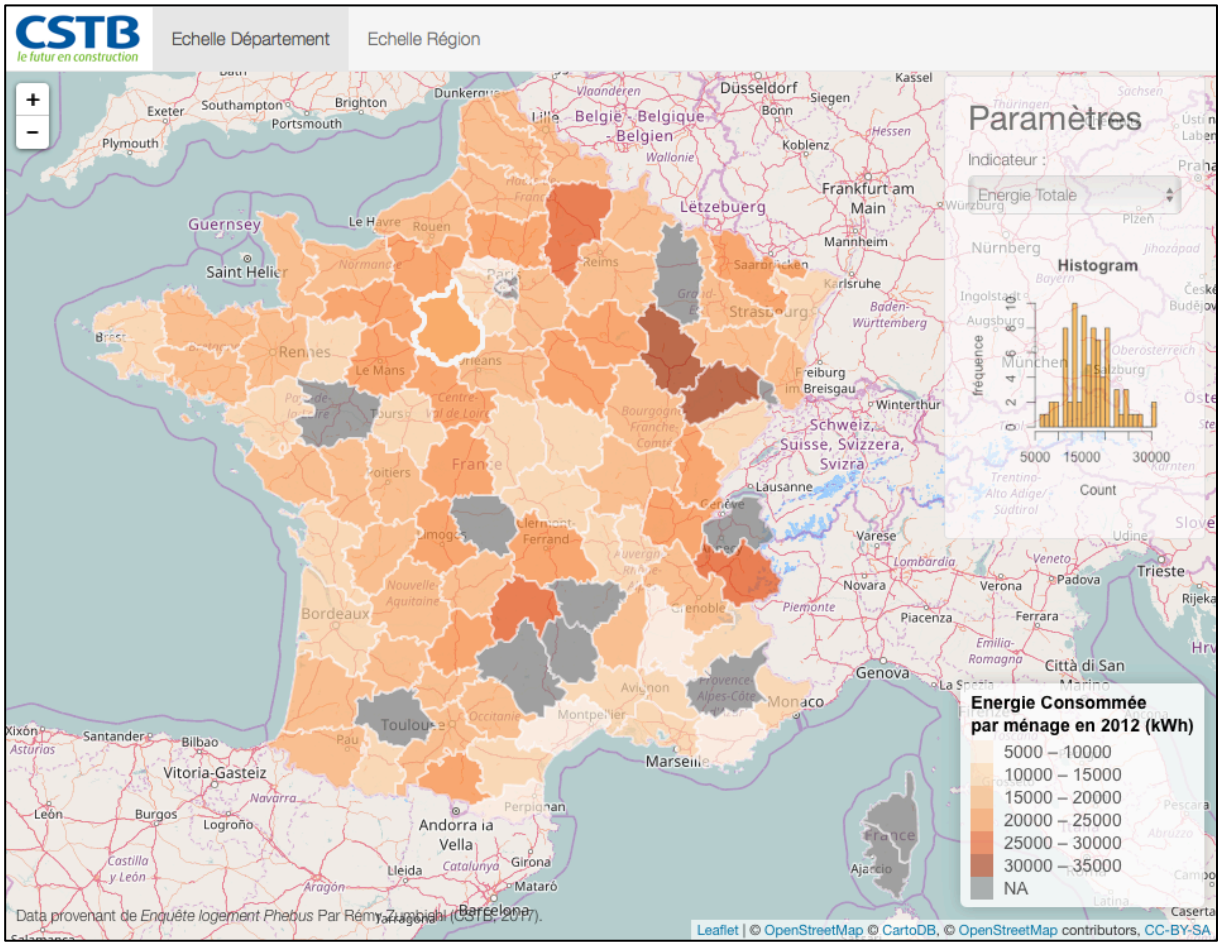
Figure 1 : Bar Plot representing the amount of households interviewed, along with their 2012 Energy Consumption and organized by DEP

At first glance, one can be noticed that some DEPs comprise a number of households interviewed much higher than others DEPs. For instance, there are only four households interviewed in the geographical division “Mayenne” while there are 23 households

interviewed in “Orne”, although both DEPs are pretty much similar in terms of number of habitants.

Also from the same graphic, one can observe that the average household energy consumption in Paris is significantly lower than the average of household energy consumption in France.

Similar analysis of household energy consumption can be carried by observing the below map of France where .



2.2 Modelling approach

The national Phebus survey provides detailed information of household characteristics, that could explain variations of the total amount of energy yearly consumed. Nevertheless, previous modelling approaches, including various approaches studied in the USA, enabled to point the fact that variations of energy consumed by households can be explained not only at a household level by the inclusion of individual household's characteristics, but also at a more aggregated level in which are considered the impact of regional effects. Indeed very few literature, such as Daniel Courgeau study, author of the article "Du groupe à l'individu" in 2004, who introduced the concept of scale or aggregation level into regressions models explaining variations of a dependant variable.

The key concept here is to understand that the group in which is belonging individuals can be considered as a relevant context to the extent that it comprises a game of influences having effects on the behaviour of individuals. Other interesting approaches were made by Wang & Wang study (spatial interaction models for biomass consumption in the United States - 2011) or Tso & Guan study (Effects of environment indicators and households features on residential energy consumption), and indicate that area variations or regional effects have a significant impact on energy consumption. The results of both studies are showing that spatial interaction among households energy consumption in the United States becomes weaker with the farther neighbour states, thus confirming that within a same group, households can interact between themselves and may influence each other in their way of consuming energy.

In this document it is proposed a multilevel modelling approach of household energy consumption in France by using predictors at both household-level data (level-1) and aggregated-level data (level-2) which considers regional effects within a national sample.

2.2.1 Multilevel approach adapted to Phebus dataset

Various simplified methods exist for estimating residential energy consumption. One of the simplest methods would be the heating degree day (HDD), which can be resumed as a tool to estimate thermic energy consumptions in function of winter temperatures. It is a typical indicator of household energy consumption for space heating. However this method is far

from being sufficient to explain variations of a complex variable such as the yearly household energy consumption, which depends on numerous independent variables. Researchers have worked on much more sophisticated methods such as neural networks for predicting energy consumption, or principal components analysis followed by regression analysis for explaining energy consumption. Each method has its advantages and disadvantages, and results are found to be good, but very few methods are taking in account the contextual effects on household energy consumption.

Multilevel models are particularly appropriated for analysing data presenting complex structures involving stratified characteristic levels. Those levels are found to be a forming a combination of micro-unity and macro-unity, for instance households and their environment, which could hereby be defined as DEP and REG. In the study of relations between households and their environment, the environment characterizes a context in which households are included and that is assessed to be relevant in order to understand the household energy consumption. In social sciences, the group of affiliations of individuals is often considered as an extremely relevant context to be studied, to the extent that in the same context are practised a game of influences having substantial effects on individuals behaviours.

The two questions that could be asked at this point are: “Does the households energy consumption is influenced by the geographic localization of the same households within a DEP or REG? – “How and according which process is the environment inflecting household energy consumption? “.
One of the benefits of multilevel regression models, compared to a more traditional regression model, is that regional effects are extracted from the variance of residential energy consumption, thus allowing explaining the remaining variance with usual explanatory variables. In a statistical model, as for instance multiple regression models, there is always an unobserved part, in other words a part of reality that is not explained by the model. In a multilevel model, dissociating different levels of observation allows to finely detecting this unobserved heterogeneity and provides a measure of variance per level. (in MRM the heterogeneity is expressed in terms of random intercepts and slopes).

2.2.2 Aggregation bias and various assumptions discarded

An important benefit of using multilevel models is that, by differentiating levels of analysis, it helps avoid the aggregation bias effect (also known as the Robinson effect). Basing on results given by aggregated data models and inferring conclusions on individual behaviours (hence household energy consumption) may well turn out to be false, and lead to what can be called the aggregation bias effect.

One of the base-assumption of classical regression models is the independency of errors, thus excluding grouping effect involving that members of the same group would tend to look alike than members not belonging to the same group. Yet it is precisely the environmental and grouping effect that is studied with multilevel regression models. When using MRM, the assumption of independency of errors is discarded and replaced by intra-group (intra-class) errors, which corresponds to the fact that households within the same group, or geographical division, tend to look alike. Moreover, classical regression models are founded on the homoscedasticity of residuals assumption, i.e. the stability of residual variance. MRM replace the homoscedasticity assumption by a weaker assumption stipulating that residual variance can vary as a linear or non-linear function of explanatory variables.

Multilevel models are mixed models adapted to stratified data analysis. It contains various error terms, at least one error term at each level of the structure model. More than one population type is then considered of being part in a multilevel analysis: one population inside each level. According Snijders & Bosker (1999), it is important to note that the dependant variable in a MRM shall be of a level-1 type. In other words, a MRM is a model with the aim of explaining something happening at the lowest, i.e. the most detailed, level possible.

2.2.3 Fixed effects and random effects

When modelling effects of some variables, it is convenient to understand which type of effect is modelled. It is therefore important to operate a distinction between fixed effects and random effects. Fixed effects are non-stochastic effects that are falling within a limited subset of modalities of a factor. When studying fixed effects, only effects of this or that specific modality of a factor on a dependant variable is assessed. With Phebus dataset, evaluating fixed effects would be limited for instance to assess the effects of the 3 housing area

modalities on household energy consumption. It is the precise effect of the affiliation of the household to one of the three housing area modalities that is studied.

At the contrary, when extrapolating the results of a study beyond the modalities strictly observed of a factor, random effects are in this case the objects of consideration. Random effects are falling within a wider subset, in fact within infinity of modalities of a factor, in which only a sample is studied. When taking an interest in random effects, one is trying to extrapolate results given by a subset of a dataset beyond the modalities strictly observed of a factor. In many modelling issues, experimental conditions are not allowing to dispose and study all potential modalities of a variable. Only a few modalities can be considered, starting from which the results will be extrapolated from other modalities that are presenting an interest for the study. In the present case, the goal is to understand to what extent the geographical context has a control over the household energy consumption. The geographical context can here be resumed by two level-2 grouping variables indicating where households are geographically located: DEP and REG. It is not a particular effect of one DEP or one REG on the REC that we are concerned for, but the global effect of DEP and REG as an overall group.

2.3 Data Preparation

As previously said, Phebus is a big dataset consisting of 768 variables for each one of the 2356 individuals interviewed, including quantitative and qualitative variables. All variables included in the dataset can be easily decoded using a dedicated dictionary provided by INSEE and associating each one of variable code naming to a detailed description of what the code stands for.
Following the appropriation of the dataset and variable code dictionary associated, one of the first main tasks can be resumed in filtering and reducing the amount of variable response to a number of variables explaining the most the yearly household energy consumption which will remain our variable response during the study.

Referring to the existing literature on the subject, a first selection-base of variables explaining the household energy consumption in France can be made, before starting with cleaning data process such as dealing with NAs, checking for any irrelevant data and studying data distribution.

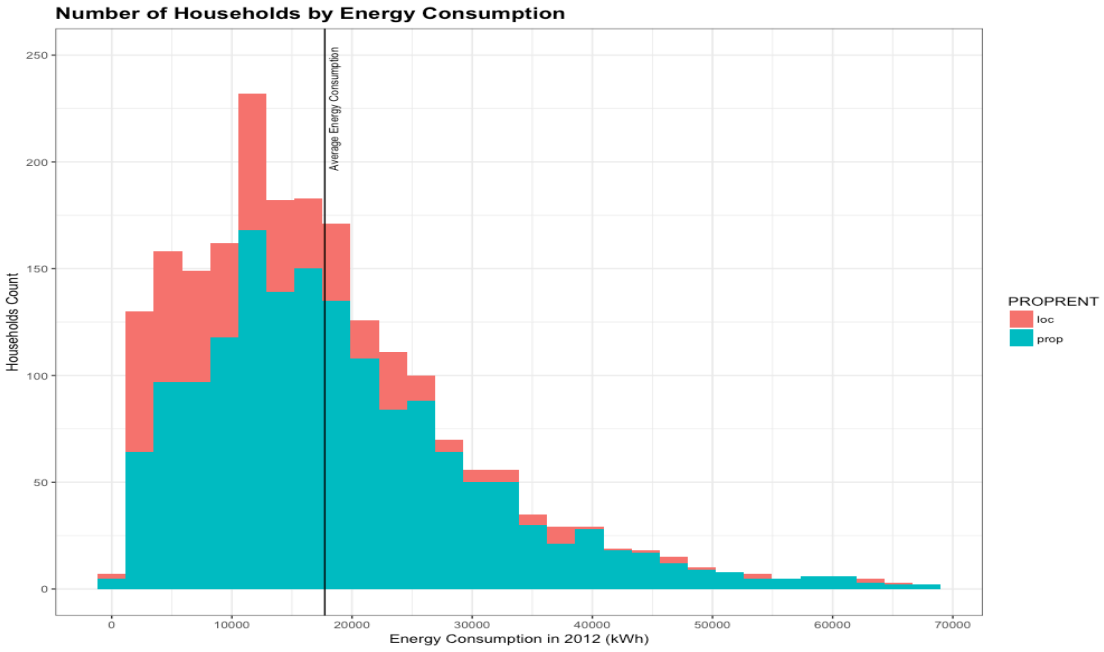
PART 3 : EMPIRICAL SPECIFICATION AND MODELS

3.1 Variables description

Prior starting to model household energy consumption using a Multilevel Regression approach, an interesting step would be to conduct a classical regression model using relevant explanatory variables at household level only (level-1- Households Individuals).
Let's introduce the variable response and some level-1 predictors.

3.1.1 Variable response

In all models that will be conducted in this paper, the response variable is 2012 energy consumption per household, measured in kWh. The household energy consumption corresponds to the quantity of final energy consumed by a household for space heating, production of hot water and electrical appliance.
The histogram below is showing the number of households interviewed during Phebus enquiry in function of their energy consumption. The graphic shows an evident skewed distribution, with most of households consuming between 0 and 30000kWh of energy.



Graphic 2 : Number of households in function of their energy consumption

On the above graphic, the red part is representing the household renting their house/flat while the blue part represents the household owners of their living space. At first glance it can be noted that, with the amount of energy consumed increasing, the owners are slightly more represented than renters.

In order to reduce the skewed distribution of the response variable, and to increase the impact of the differences between values on the left side of the distribution, where the majorities of the observations are taking place, a log transformation is implemented.
The transformed response variable is named *LOGCONSTOT*.

3.1.2 Level – 1 Explanatory variables

LOGREV is a numerical variable indicating the gross household income disposed in 2012, with a log transformation to reduce a right skewed distribution.

AREA3G is a categorical variable indicating the area of the housing, divided in three groups: 0 - 40 m^2 ; 40 m^2 - 100 m^2 ; 100m^2 and above.

INSULHOUS is a binary indicator measuring whether the housing is insulated (=1) or attached to another housing (=0).

YEARCONST is a numerical variable recording the year when the house was constructed.

ROOMNBR is a numerical variable indicating the number of bedrooms in the housing.

HEATSYST is a categorical variable indicating whether the space heating system is dedicated only for heating the housing, for heating a cluster of housing (collective space heating system), or a mixed system (individual and collective).

HEATSOURCE indicates the type of energy used for the space heating system. Three categories are defined: electricity ; gas ; other.

RURAL is a binary indicator showing if the housing is located in a rural area (i.e. less than 2000 hab). 1= Yes, 0 = No.

HEATTEMP is another binary variable indicating the heating temperature selected by the households to heat their housing is above 21°C (included). 1= Yes, 0= No.

ECS is a categorical variable indicating how is produced hot water. Three categories are defined: Electricity, boiler (using gas, fuel, or wood as energy source), and others.

UNOCCWEEK is a binary variable indicating whether the housing is unoccupied less than four hours during weekdays. 1= Yes, 0= No.

PCS is a categorical variable indicating household employment status. Three categories are defined: Executive status, Middle-level status, and other status.

NBRPERS is a numerical variable indicating the number of persons living in housing.

3.1.2 Level – 2 Grouping variables

DEP is a categorical variable indicating the number of the "département" where is located the household.

REG is a categorical variable indicating the number of the "Région" where is located the household.

3.2 Model 1 – Multiple Regression Model

Let's introduce a first multiple regression - Model 1, explaining residential households energy consumption using only level-1 explanatory variables.
Model 1 can be contextually written with equation (1) as :

$$Y_i = \gamma_0 + \sum_k \beta_j . X_j + e_i$$

In Eq. (1), Y_i is the annual energy consumption of household i.

X_j is the matrix of level-1 explanatory variables.

Other parameters in the equation need to be estimated.

γ_0 is the intercept and β_j is the slope of level-1 explanatory variables X_j.

e_i is the error term which represents the variability non explicated by the model.

The results of the multiple regression model can be found on the table below.

Model 1 (multiple regression)	
Parameters	Estimate (Std Error) – p value codes
Intercept	9.113417 (0.508465) - ***
LOGREV	0.086045(0.025743) - ***
AREA3G	
100-inf	0.398126(0.109430) - ***
40-100	0.223487(0.105132) - *
INSULHOUS	0.220906(0.028592) - ***
YEARCONST	-0.001316(0.000220) - ***
ROOMNBR	0.099743(0.014281) - ***
HEATSYST	
Indiv	1.238481(0.049766) - ***
Mixt	1.501356(0.059138) - ***
HEATSOURCE	
Electricity	-0.471268(0.035131) - ***
Gas	-0.076311(0.034192) - *
RURAL	0.106759(0.029169) - ***
HEATTEMP (>21°C)	0.106637(0.026012) - ***
ECS *(ref.mod. others)*	
gas,fuel,wood	0.239602(0.066698) - ***
electricity	0.158986(0.066228) - *
UNOCCWEEK	0.096351(0.025179) - ***
PCS*(ref.mod. others)*	
Executive	-0.066157(0.040036) - .
Middle-level	-0.085812(0.036835) - *
NBRPERS	0.043580(0.010263) - ***

Table 1 : Results of Model 1 obtained using lm command in R

All coefficients (slopes) estimated with Model-1 appear to be significant (p-value < 0.1).

Analysis of F-Statistic for the 13 level-1 explanatory variables and 2071 DOF as well as the associated probability indicates a correct global significance of Model1.

Various Model 1 diagnostics such as residuals normality diagnosis, homoscedasticity and co-linearity evaluation, consolidate the acceptation of the model.

Calculation of the R^2 determination coefficient, which indicates the part of the variance explained by the model (i.e. ratio of variance explicated by the model divided by total variance) gives a value of 0.5382.

Hence, approximately 54 % of variance of the yearly household energy consumption in France is explained with a classical regression model, using the 13 above explanatory variables.

At this point it is interesting to consider the limitation of a traditional regression model, as the results are showing that only half of the variance of REC can be explained. All variables incorporated in the model have a significant influence on the energy consumption but an important part of its variance remains unexplained. The question that could be asked at this point is: "Does the geographic context has an influence on the household energy consumption? ". In such case, multilevel regression approach are found to be one the effective tools in order to demonstrate the regional effects on a dependant variable.

Multilevel regression modelling is found to be well adapted to Phebus dataset in the way that the modelling method can establish a link between geographical context and household energy consumption. MRM can easily be implemented with dedicated statistical packages in R.

3.3 Model 2 – Null Models

3.3.1 Introduction to Null Models

One of the steps that must be taken when analysing stratified data consists of estimating how the variance of the studied phenomena is shared among the different levels that are supposed to structure the dataset. To this purpose, null models, which are the simplest possible multilevel models, are proving to be useful. A null model is equivalent to a variance

analysis with random effects (ANOVA) and is completely unconditional, as no explanatory variables are introduced in the null model.

A Null Model can be contextually written with the following equations:

$$Y_{ij} = \beta_{0j} + e_{ij} \qquad \text{(at level -1)}$$
$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad \text{(at level -2)}$$

Integrating both equations in the equation (2):

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad \text{(both levels 1 \& 2)}$$

In Eq. (2), Y_{ij} is the annual energy consumption of household i in reportable geographic group j. β_{0j} is the estimated intercept for each reportable geographical group j.

γ_{00} represents the average annual energy consumption Y.
u_{0j} is the random error associated to each geographical group j, and supposedly having normal distribution with mean value of zero and variance σ_{u0}^2.
e_{ij} represents a random error associated to each household i, supposedly having normal distribution, mean value of zero, and variance σ_e^2.

A null model, or intercept-only model, comprises a fixed part (γ_{00}) and a random part with the two error terms u_{0j} & e_{ij}. At this point, the calculation of an intra-class coefficient (ICC) turns out to be very useful in order to assess what would represent the share of intra-class variance compared to the global variance. The ICC coefficient can be contextually written with equation (3) as:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

ICC coefficient can hereby be interpreted as a degree of similarity of households within the same geographical cluster (DEP or REG). Moreover, it can be resumed as a simple variance decomposition of the dependant variable. The key point to understand, when fitting null models, is that disparities u_{0j} between geographical groups are considered as random.

Indeed, not taking in account the random composition of geographical clusters leads to neglect the sampling variance affecting estimation calculations, and subsequently to lean (biaiser) information due to an overestimation the share of inter-class variance.

Null models are providing crucial information regarding the variance shared among levels that are considered to stratify data. The evolution of the part of residual variance over the subsequent modelling work will always be related to the variance shared among levels and provided by null models.

3.3.2 Fitting two Null Models

Let's start by fitting two null models with Phebus dataset, each one with a different level-2 grouping variable: DEP and REG.

In R this procedure can easily be done with lme4 package using the following line commands:

```
NullModel1 <- lmer(LOGCONSTOT ~ 1 + (1 | DEP), data = Phebus)
NullModel2 <- lmer(LOGCONSTOT ~ 1 + (1 | REG), data = Phebus)
```

The intercept denoted by 1 immediately following the tilde sign, is the intercept for the fixed effects. Within the parentheses, 1 denotes the random effects intercept, and the variables DEP or REG are specified as the level-2 grouping variable.

Results of the null models can be seen in the table below.

Fixed Effects	**NullModel 1** Est (StdErr)	**NullModel 2** Est (StdErr)
Intercept γ_{00}	*9.56344(0.03428)*	*9.54377(0.05815)*

Random Effects	**NullModel 1** Var(StdDev)	**NullModel 2** Var(StdDev)
Level 1 - Intercept σ_{u0}^2	*0.06652(0.2579)*	*0.03709(0.1926)*
Level 2 - Residual σ_e^2	*0.52176(0.7223)* *Groups DEP : 81*	*0.55489(0.7449)* *Groups REG : 12*

Number of obs : 2090		

Quality criteria	**NullModel 1**	**NullModel 2**
AIC	*4685.0*	*4736.1*
BIC	*4702.9*	*4753.1*

ICC	**NullModel 1**	**NullModel 2**
$\sigma_{u0}^2 / (\sigma_{u0}^2 + \sigma_e^2)$	*0.11307*	*0.06265*

Table 2 : Results of Null models obtain using lme4 command in R

The table above represents the parameters estimates and standard errors for both null models. NullModel 1 estimates the intercept as 9.56 which is simply the logarithm of the average household energy consumption across all DEP geographical divisions and households (i.e. 14186kWh). On the other hand, NullModel 2 estimates the intercept as 9.54, thus average household energy consumption across all REG and households (i.e. 13904kWh).

Households features and their environment (geographical context) are two distinct sources of the variance of the energy consumption in the housing, and both sources of variance have to be modelled as random effects. A null model does not explain specifically the variance of the dependent variable. It only decomposes the variance into two independent components: σ_e^2 which is the variance of the lowest-level errors e_{ij}, and σ_{u0}^2 which is the variance of the highest-level error u_{0j}.

The two sources of the variance of household energy consumption can be resumed as an inter-class variance in the level of households (level -1) and an intra-class variance, or residual variance, in the level of the grouping variable (*level -2 DEP or REG*). The “residual” term is used to denote part of variance that cannot be explained or modelled with the other terms. It is the variation in the observed data that is “left over” after are determined the estimates of the parameters in the other parts of the model.

Considering the models described above, the variance component corresponding to the random intercept is 0.06652. The two-variance components can be used to partition the variance across levels. The Intra-class correlation coefficient ICC for NullModel1 is equal to *0.06652/ (0.06652+0.52176)* = 0.113, meaning that roughly 11.3% of the variance of the

yearly energy consumption per household is attributable to the DEP-level. The calculated ICC implies that clustering effect is greater than 10% and a multilevel regression analysis is needed to control for this clustering effect. A similar process for the calculation of the ICC for model 2 indicates that roughly 6.2% of the global variance of the energy consumption per household is attributable to the REG-level. ICC calculated for Nullmodel 1 has a higher value than ICC calculated for Nullmodel2, and this difference would indicate that households are more similar in their consumption of energy within “départements” than within “regions”. Thus, in order to build a multilevel model, a level-2 grouping variable related to “département” could be more appropriated than a level-2 grouping variable related to “regions”.

We shall therefore at this point consider working on a multilevel regression using DEP as the level-2 grouping variable.

Since the null models described above do not contain explanatory variables, the residual variances represent unexplained error variance. The deviance term reported in the same table is a measure of model misfit; when adding explanatory variables to the model, the deviance is expected to go down.

3.4 Step-by-step – MRM Models

3.4.1 Level – 2 Explanatory variables

Prior getting to a complete multilevel model, including all relevant predictors, let’s first introduce significant level-2 predictors that may explain variation of household energy consumption. All candidate variables described below aim to capture regional effects.

LOGREVDEP is a numerical variable indicating the logarithm of the average per capita disposable household income, per DEP in 2012 (source of data INSEE).

HDDDEP is a numerical variable indicating the heating degree days in 2012 (source of data ADEME).

In France, three level-2 explanatory variables could explain the clustering effects within DEP geographical divisions : DEP-average per capita yearly income in 2012 (*LOGREVDEP*), and heating degree-day per department in 2012 (*HDDDEP*). To be noted that due to homogeneous energy policies among French administrative divisions (DEP and REG), average prices of energy remain identical across whole French territory and therefore can not be taken in account in our models.

3.4.2 Two Models introducing Level – 2 Explanatory variables

The three following R command lines will fit multilevel mixed models starting the inclusion of a first level-2 explanatory variable for the first model, inclusion of a second level-2 explanatory variable for the second model, and finally all three level-2 predictors incorporated into the last model.

```
Model3_1 <- lmer(LOGCONSTOT ~ 1 +
                                    HDDDEP  + (1 | DEP), data = phebus, REML = FALSE)

Model3_2 <- lmer(LOGCONSTOT ~ 1 +
                        HDDDEP +
                        LOGREVDEP + (1 | DEP), data = phebus, REML = FALSE)
```

Results of the two models using level-2 explanatory variables can be seen in the table below.

Fixed Effects	**Model 3_1** Est (StdErr)	**Model 3_2** Est (StdErr)
Intercept γ_{00}	*2,6169(1,2602) - **	*14,8102(2,4373) - ***
LOGHDDDEP	*0,8897(0,1614) - ***	*0,9686(0,1357) - ***
LOGREVDEP		*-1,2503(0,2258) - ***
LOGDENSPOPDEP		

Random Effects	**Model 3_1** Var(StdDev)	**Model 3_2** Var(StdDev)
Level 2- Intercept σ_{u0}^2	*0,04322(0,2079)*	*0,02334(0,1528)*

Level 1 - Residual σ_e^2 *Number of obs : 2090*	*0,52093(0,7218)* *Groups DEP : 81*	*0,52179(0,7223)* *Groups DEP : 81*

Quality criteria	Model 3_1	Model 3_2
AIC	*4685.0*	*4639*
BIC	*4702.9*	*4667,2*

Table 3 : Results obtained in R when fitting models with level-2 explanatory variables only

AIC and BIC are information criterions that can be taken in consideration when checking for the suitability of above models. AIC is taking in account the number of parameters to estimate while BIC is taking in account the number of parameters and the size of the sample. The difference of value of AIC or BIC, when passing from a model to another, indicates the suitability of the models. Level-2 variable candidates are selected if it decreases AIC (or BIC) by 10 or more. To be noted that information criterions can't be considered as objectives indicators when they are considered alone. Those criterions represent an interesting tool when their value can be compared from a model to another.

The table above indicates that, when comparing from the null model to model 3_1, the variance components corresponding to the random intercept has decreased from 0.0665 to 0.0432. Thus, the inclusion of two level-2 predictors has accounted for some of the unexplained variance in the household's energy consumption. Nevertheless, the estimate is still more than twice of the size of it's standard error, suggesting that there remains unexplained variance.

In Model 3_2, the aggregated predictors represent the contextual effect and somehow the inter-class effect on the households energy consumption. On one hand, a high value of heating degree day implies a high yearly energy consumption by households, which could be easily explained by the fact that it is necessary to use more energy to maintain a comfortable living temperature in housings during winter time. On another hand, the higher is the value of the average per capita disposable household income in a "département" division, the lower will becomes the household energy consumption in the same division, thus meaning that households living in geographical divisions where the mean income is low are consuming more energy than those who are living in divisions where the mean level of income is high. Although this last remark seems unusual, as usually a high level of income is correlated with high-energy expenses, it can be explained by what one can call "the Parisian effect" on the study of contextual effect. In Paris for instance, very few housings are insulated from each

other and less energy is demanded for heating housings compared to other “départements” with a lower mean level of income but at the same time with housing more insulated.

3.4.3 Model 3 - Level – 1 & 2 Explanatory variables

After having incorporated suitable level-2 variables, let’s introduce level-1 explanatory variables described in chapter 3.1.2 in order to build a multilevel regression model.

Model 3 can be contextually written with the following equation (3):

$$Y_{ij} = \gamma_{00} + \sum_{k} \beta_{k0} . X_{kij} + \sum_{l} \beta_{0q} . Z_{lj} + \sum_{k} u_{kj} . X_{kij} + u_{0j} + e_{ij}$$

In Eq. (3), Y_{ij} is the annual energy consumption of household i in geographical division j (DEP).
X_{ij} is the matrix of level-1 explanatory variables and Z_j is the matrix of level-2 explanatory variables. Other parameters in the equation need to be estimated. For the fixed effects, β_{00} is the intercept for fixed effects, β_{k0} is slope for level-1 explanatory variables, and β_{0q} is slope for level-2 explanatory variables. Regarding random effects, e_{ij} are errors at level 1 (households), u_{0j} and u_{kj} are residuals terms at level 2 (DEP). The variance of the residual error u_{0j} is the variance of the intercepts between DEPs and the variances of the residual term u_{kj} are the variances of slopes between groups.

The table below is resuming results given by NullModel1 fitted in chapter 3.3.2 as well as the results obtained after fitting a multilevel model including level-1 and level-2 predictors with no interactions between variables.

Fixed Effects	**NullModel1 (DEP)** Est (StdErr)	**Model 3** Est (StdErr)
Intercept **γ_{00}**	*9.56344(0.03428)*	*5.292210(1.581710)* ***
LOGHDDDEP		*0.594353(0.077933)* ***

LOGREVDEP		*-0.069042(0.155860)*
LOGDENSPOPDEP		*-0.024734(0.0163385)*
LOGREV		*0.099270(0.025610) ****
AREA3G		
100-Inf		*0.329653(0.107710) ****
40-100		*0.160861(0.103419)*
INSULHOUS		*0.201693(0.028163) ****
YEARCONST		*-0.01324(0.000215) ****
ROOMNBR		*0.098093(0.014047) ****
HEATSYST		
Indiv		*1.222301(0.049564) ****
Mixt		*1.489877(0.058777) ****
HEATSOURCE		
Electricity		*-0.444014(0.034445) ****
Gas		*-0.065390(0.033781) **
RURAL		*0.059387(0.030740) **
HEATTEMP		*0.104282(0.025533) ****
ECS		
Gas,Fioul,Wood		*0.232124(0.065039) ****
Other		*0.162513(0.064558) ***
UNOCCWEEK		*0.111200(0.024596) - ****
PCS		
Executive		
Middle level		*-0.050374(0.039155) -*
NBRPERS		*-0.083686(0.035862) - ***
		*0.042803(0.010016) - ****

Random Effects	**NullModel1 (DEP)** Var(StdDev)	**Model 3** Var(StdDev)
Level 2 - Intercept σ_{u0}^2	*0.06652(0.2579)*	*0.002826(0.05316)*
Level 1 - Residual σ_e^2	*0.52176(0.7223)*	*0.260016(0.50992)*
Number of obs : 2090	*Groups DEP : 81*	*Groups DEP : 81*

Quality criteria	**NullModel1 (DEP)**	**Model 3**
AIC	*4661*	*3183.1*
BIC	*4683.5*	*3318.6*

Table 4 : Results obtained in R when fitting a MRM model with all level 1 & 2 predictors

The evolution of the estimations of the random effects between both models is quite interesting. Passing from Nullmodel1 to multilevel model3 brought an explanatory profit. Let's examine this explanatory profit level by level. At level-1 (households), residual variance is 0.52176 for Nullmodel1 and is 0.260016 for model 3. Thus, the explanatory profit of model 3 is about (0.52176-0.260016)/0.52176 = 0.502. Model 3 is therefore explaining 50.2% of the level-1 variance of households energy consumption. At level-2 (DEP), residual variance is about 0.06652 for Nullmodel1 and 0.002826 for model 3. Thus, the explanatory profit of model 3 is about (0.06652-0.002826)/0.06652 = 0.957, hence model 3 is explaining roughly 95% of the level-2 variance of households energy consumption.

CONCLUSION

Scientists are usually describing a society with hierarchical structure and multilevel models were developed to appropriately represent such data structures and incorporate hierarchical levels inside the model. Countries like France are showing various hierarchical levels such as cities, department, counties, or region. Households leaving in different geographical divisions might differ in the way they consume energy, even if the households characteristics are similar (income, age, etc..). This difference could be due to numerous environmental indicators including cultural, economic, politic, or historic reasons. Households sharing the same environment might also some similarity in the way they are consuming energy, even if their individual characteristics are different. When using multilevel modelling, one is assuming that within a same group, or geographical division, individuals interact each other and therefore can mutually influence their way of consuming energy. In other terms “like attracts like” and “birds of a feather flock together”.

Modelling the household energy consumption using a multilevel model allows for better understanding and increases the explanatory profit when compared to classical multiple regression models that are considering only a single hierarchical level in the data. The research studied in previous chapters seems to suggest that, while individual or level-1 information explains a larger part of energy consumption variation (88 % of global variance), there is some statistical evidence for contextual effects in the household energy consumption variation within French departments (12% of global variance).

Using multilevel regression modelling brings numerous advantages. For instance, and unlike classical multiple regression model, the independence of the residuals assumption can be violated as it is precisely the grouping effects that is studied when working with MRM. Another assumption can be violated: the homoscedasticity of the residuals (consistency of residual variance). Multilevel model replace the homoscedasticity assumption by a weaker assumption whereby the variance of residuals can be represented by a linear (or non linear) function of explicative variables. Another significant advantage of multilevel models is that working simultaneously with two hierarchical levels on a stratified dataset, mitigates the risk of having a bias aggregation error (or atomist error) which consists in inferring on an individual level what has been observed on a aggregated level.

Multilevel regression modelling also has its limitation. Firstly our sample size taken from Phebus dataset might un-sufficiently large to draw inference for a population of households and a population of departments. Secondly, the results of the ad hoc geographical clustering confirmed the significant regional effect on households energy consumption, however the interpretation of the results was proved to be quite complex. In effect, although a multilevel model is increasing the explanatory profit from 55% to 67% of the global variance of the household energy consumption, it is quite difficult to precisely analyse and quantify the geographical effect of one specific department compared to another. Our model tested

Multilevel models were developed in order to capture the spatial variation of the effect of alcohol enforcement, which was found to be highly significant. It is noted that no other variables were found to add explanatory effect in the reduction of road accidents in Greece. This was not surprising, as no other parameter (e.g. vehicle ownership, road network length, etc.) presented a significant variation, comparable to the increase of enforcement, in the examined period. Consequently, as in other studies (Goldenbeld and van Schagen, 2005), the intensi- fication of enforcement is considered to be the main cause of the improvement of road safety in Greece. Although additional explanatory variables (for which data was not available) could contribute to the description of this trend, the present models are efficient in describing the regional variation of this trend.

Our study is not without limitations. Firstly, our study was cross sectional, limiting our ability to draw causal inferences, in particular the direction of the association between house- hold income and self rated health that could be potentially bi-directional. Secondly, while our sample size was suffi- ciently large to draw inferences for the population of individuals and the population of communities, at the region level, as there are only 13 regions in Chile, we need to be cautious about drawing extensive inferences on the between region variations observed in the analysis. While an alternative strategy would have been to model the 13 regions of Chile as a fixed effect, as our primary interest was not in making region specific inferences we did not specify them as ‘‘fixed effects’’. At the same time, we recognise that community variation should be estimated after taking into account the regions to which they belong and hence the treatment of regions as a ‘‘random effect’’, which is compatible with our conceptual framework.

the research reviewed above seems to suggest that, while individual or micro-level information explains a larger part of health varia- tion, there is some statistical evidence for contextual effects in health varia- tion in the

British population, which can be expressed in terms of information on geographic setting. These contextual effects may operate at more than one geographic scale.

Our study is not without limitations. Firstly, our study was cross sectional, limiting our ability to draw causal inferences, in particular the direction of the association between house- hold income and self rated health that could be potentially bi-directional. Secondly, while our sample size was suffi- ciently large to draw inferences for the population of individuals and the population of communities, at the region level, as there are only 13 regions in Chile, we need to be cautious about drawing extensive inferences on the between region variations observed in the analysis. While an alternative strategy would have been to model the 13 regions of Chile as a fixed effect, as our primary interest was not in making region specific inferences we did not specify them as ''fixed effects''. At the same time, we recognise that community variation should be estimated after taking into account the regions to which they belong and hence the treatment of regions as a ''random effect'', which is compatible with our conceptual framework.

The results of the ad hoc geographical clustering confirmed the significant regional variation of the effect of enforcement, however the interpretation of results was proved to be quite com- plex. It might be reasonable to assume that the regional variation of the effect is mainly a result of different regional practices in the implementation of enforcement.

Nested data structures—observations organized in non-overlapping groups—arise naturally from numerous data collection schemes. Examples resulting in such a data structure include situations when individuals are observed over time (repeated measures); when a field is sub- divided into smaller plots on which a treatment is applied (split plots); or when a stratified sampling scheme is used, such as when sampling students within schools within districts (mul- tilevel data). When data are organized in this manner observations are no longer independent. Any statistical model used must accommodate for these dependencies by allowing for a more general covariance structure, in which observations from the same group or individual can be correlated.

Mixed models were developed to appropriately represent such data structures. They incor- porate parameters that govern the dependence structure—the random effects—and parameters that govern the global trend—fixed effects. With the additional flexibility provided by the random effects comes additional complexity at each stage of statistical modeling. The prob- lem of parameter estimation (i.e., model fitting) has been widely addressed in the literature, with the most commonly used approaches being maximum likelihood (ML) estimation and re- stricted maximum likelihood (REML) estimation (see Christensen, 2002; Demidenko, 2004, for detailed overviews). We discuss existing approaches for exploratory and confirmatory analysis in chapter 2.

Hierarchical structures are omnipresent in today's society—this is reflected in the data that we collect on all aspects of this society. Hierarchical linear models allow a representation of structural levels in a statistical modeling framework. Diagnostic tools are used to assess the quality of model estimation and explore features of the data not well described by the model. Residual and influence diagnostics are familiar tools for the classical regression model with independent observations. For hierarchical linear models, these diagnostic tools must be adjusted to reflect the dependence introduced by the nested data structure. Residual analysis now includes the assessment of distributional assumptions at each level of the model. This requires the use of level-dependent residual quantities. Similarly, the parameter estimates may be influenced at each level of the model, requiring influence diagnostics that can pinpoint specific levels of the model, as well as specific aspects of the model. We present an overview of the diagnostic tools available for hierarchical linear models that are familiar from linear models. Additionally, we discuss the utility of the lineup protocol for residual analysis with complex models.

BIBLIOGRAPHIE

P. Bressoux. *Modélisation Statistique Appliquée aux Sciences Sociales*. De Boeck, 2010, 462p.

D. Courgeau. *Du groupe à l'individu – Synthèse Multiniveau*. Institut National d'Etudes Démographiques, 2004.

Humphreys K, Carr-Hill R. *Area variations in health outcomes: artefact or ecology*. Epidemiol Community Heath, 1991, 20(1), 251-8.

Tso & Guan study. *A multilevel approach to understand effects of environment indicators and households features on residential energy consumption*. Science Direct (Volume 66). March 2014. p722-731. (https://doi.org/10.1016/j.energy.2014.01.056)

Wang & Wang. Spatial Interaction models for biomass consumption in the United States. Science Direct(Volume36).2011.P6555-6558. (https://doi.org/10.1016/j.energy.2011.09.009).

Snijders & Bosker. *Multilevel Analysis : an introduction to basic and advanced multilevel modelling.* SAGE Publication. 2011. 368p.

Gelman & Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. 2006. 656p.

Luca Campanelli. *Introduction to mixed-effects modeling using the lme4 package.* 2017. https://www.lcampanelli.org/mixed-effects-modeling-lme4/

K.Magnusson. Using R and lme/lmer to fit different two – and three level - longitudinal models. 2016. http://rpsychologist.com/r-guide-longitudinal-lme-lmer

Bressoux, Coustere, Leroy-Audouin. *Les modèles multiniveau dans l'analyse écologique : le cas de la recherche en éducation.* Revue Française de Sociologie. 1997. Volume 38. P67-97.

Golaz & Bringé. *Enjeux et limites de l'analyse multiniveau en démographie*. INED – CEPED. 2009. http://jms.insee.fr/files/documents/2009/53_4-JMS2009_S03-CS_GOLAZ-ACTE.PDF

Attention n'oubliez pas d'effacer le texte suivant quand vous n'en aurez plus besoin.

Une synthèse introductive d'une demi-page fait le point sur l'état de l'art percu au travers de la bibliographie et précède les références présentées par thème. Chaque référence bibliographique est suivie d'un commentaire qui en souligne l'intérêt (théorie, méthodologie, applications, originalité du procédé, etc…) en style TEXTE.

Pour la structuration de votre bibliographie, vous avez à votre disposition, dans le menu déroulant des Pages postliminaires, un niveau de titre : 'Subdivision de biblio' (style 3|Bibli_tit2) et dans la liste des styles, si nécessaire, un autre niveau de subdivision (style 3|Bibli_tit3). Toutes les références sont stylées en style « 3|Bibli_item ».

Pour l'écriture de vos entrées bibliographiques, vous avez 2 possibilités :

- Soit, écrire vos références selon les instructions (du cours de recherche documentaire) en choisissant dans le menu déroulant des Pages postliminaires 'Référence biblio' (style '3|Bibli_item)
- Soit, utiliser les champs prédéfinis ci-dessous en prenant soin, avant de les remplir, de les recopier autant de fois que nécessaire ou tout du moins en prenant la précaution d'en garder un vierge. **Attention, votre présentation ne doit pas suivre les types de document, mais les thèmes de votre rapport**.

Ouvrages imprimés

NOM, Prénom ou Initiales. *Titre en italique : sous titre en italique*. Lieu d'édition : Editeur, Date de publication, nombre de pages p.

Ouvrages électroniques

NOM, Prénom ou Initiales. *Titre en italique : sous titre en italique*. [en ligne ou cédérom ou bande magnétique ou disquette], Lieu d'édition : Editeur, Date de publication recommandée, [référence du JJ mois AAAA (ou visité le JJ mois AAAA)]. renseignements nécessaires pour localiser le document cité ex disponible sur Internet <http://www.xxxxx>

Chapitre dans un ouvrage imprimé

NOM, Prénom ou Initiales. Titre du chapitre. In : NOM, Prénom ou Initiales (éd. sc.), *Titre de l'ouvrage en italique : sous titre en italique*. Lieu d'édition : Editeur, Date de publication, nombre de pages p.

Rapports imprimés

NOM, Prénom ou Initiales. *Titre en italique : sous titre en italique*. Lieu de publication, Date de publication, nombre de pages p.

Travaux universitaires

NOM, Prénom ou Initiales. *Titre du mémoire ou de la thèse en italique : sous titre en italique*. Nature de la thèse ou du mémoire, Université ou Ecole de soutenance, Date de soutenance, Nombre de pages p.

Articles de périodiques imprimés

NOM, Prénom ou Initiales. Titre de l'article. *Titre du périodique en italique*, (pays facultatif), Année, volume et/ou numéro, pp xx - xx

Articles de périodiques électroniques

NOM, Prénom ou Initiales. Titre de l'article. *Titre du périodique en italique*, [en ligne ou cédérom ou bande magnétique ou disquette], (pays facultatif), Année, volume et/ou numéro, [référence du JJ mois AAAA (ou visité le JJ mois AAAA)]. renseignements nécessaires pour localiser le document cité ex disponible sur Internet <http://www.xxxxx>

Communication dans un congrès

NOM, Prénom ou Initiales. Titre de la communication. In : NOM, Prénom ou Initiales (éd. sc.), *Titre du congrès*, *Lieu du congrès*, *Date du congrès*. Lieu d'édition : Editeur, Date de publication, pp xx - xx

Sites web consultés

Nom du site. [référence du JJ mois AAAA (ou visité le JJ mois AAAA)], URL du site <http://xxxxxxxxxx>

Bases de données

Organisme auteur, sigle. Nom développé de la base, sigle. Date de création

ANNEXES

Annexe 1 Titre de votre annexe

Le contenu des annexes doit être stylé de la même façon que le reste du document , en style « Texte » pour le texte et en style « 3|Ann_Titre3 » (Partie d'annexe) ou « 3|Ann_Titre4 » (sous-partie d'annexe) si des subdivisions sont nécessaires à l'intérieur de l'annexe.

Partie d'annexe

Sous-partie d'annexe