Synthèse : From energy behaviors to lifestyles: Contribution of behavioral archetypes to the description of energy consumption patterns in the residential sector

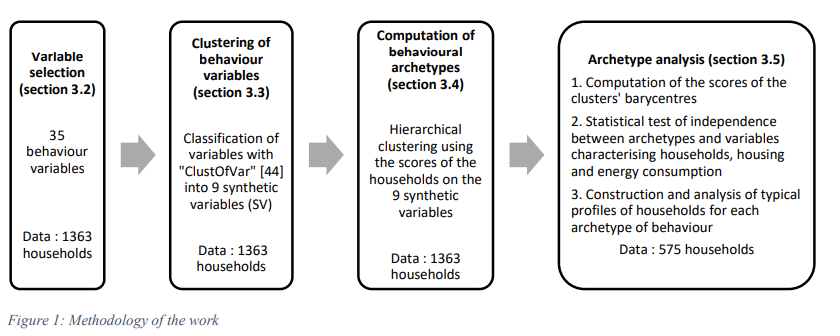
Article de décembre 2022

Construction d’archétypes comportementaux dans le but de mieux comprendre le comportement énergétique des ménages.

Algorithme de clustering

Dataset de 35 variables couvrant l'hygiène, l'alimentation, le chauffage, l'éclairage, les pratiques de loisirs et l'occupation du logement avec 1363 ménages d’Ile de France

7 archétypes distincts de comportements domestiques ont été identifiés

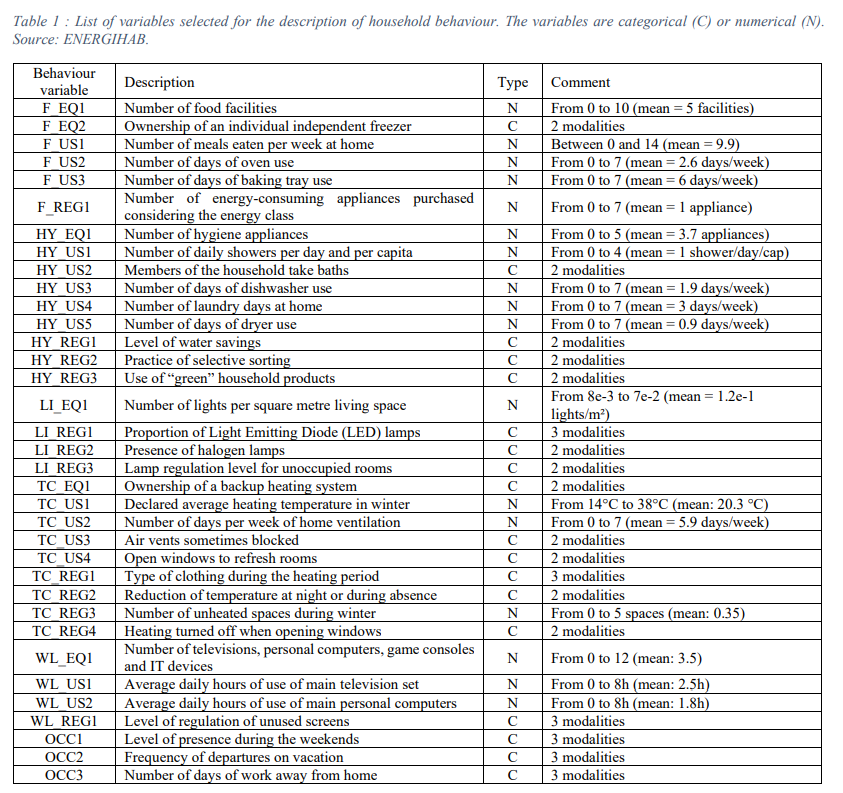


## Données

Source : projet ENERGIHAB de l’ANR (Agence Nationale de la Recherche française)

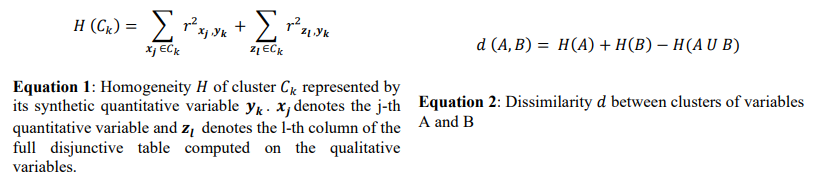
Enquête téléphonique menée en 2010 auprès de 1950 ménages de la région Île-de-France. Le questionnaire comprenait 362 questions portant sur les comportements, la consommation d'énergie, ainsi que les caractéristiques des ménages et des logements.

35 variables comportementales ont été sélectionnées, comprenant 19 variables qualitatives et 16 variables quantitatives, 1363 ménages.



## Clustering

* Clustering par la méthode de Chavent qui calcule des coefficients de corrélation pour les variables quantitatives et des ratios de corrélation pour les variables qualitatives. Ensuite, une variable synthétique est calculée pour chaque groupe, représentant une sorte de moyenne des variables du groupe.



Pour choisir combien de groupes (clusters) nous voulons créer à partir de nos variables, nous avons suivi trois étapes. Tout d'abord, nous avons utilisé une méthode appelée Agglomerative Hierarchical (AHC) Clustering pour organiser nos variables dans une sorte d'arbre en fonction de leur similarité. Ensuite, nous avons examiné cet arbre pour voir comment sa hauteur changeait en fonction du nombre de groupes que nous voulions former. Nous avons choisi le nombre idéal de groupes en regardant où la hauteur de l'arbre diminuait le plus.

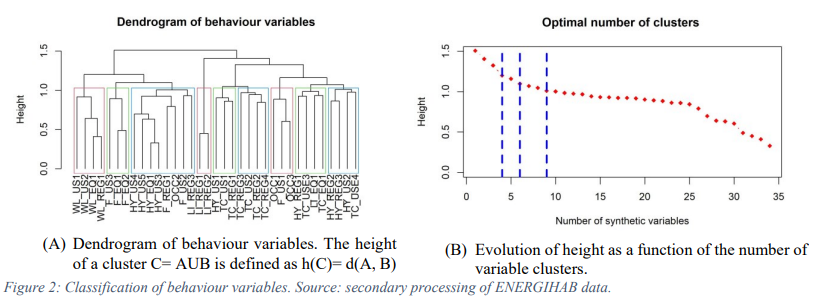
* Regroupement des ménages similaires avec la distance euclidienne
* Liaison des clusters avec les variables synthétiques avec test de Student. Ce test compare les moyennes des coordonnées des barycentres pour chaque variable synthétique. Il détermine si les moyennes des coordonnées sont significativement différentes entre les clusters. En d'autres termes, il vérifie s'il y a une différence significative dans la manière dont chaque cluster est représenté par les variables synthétiques.

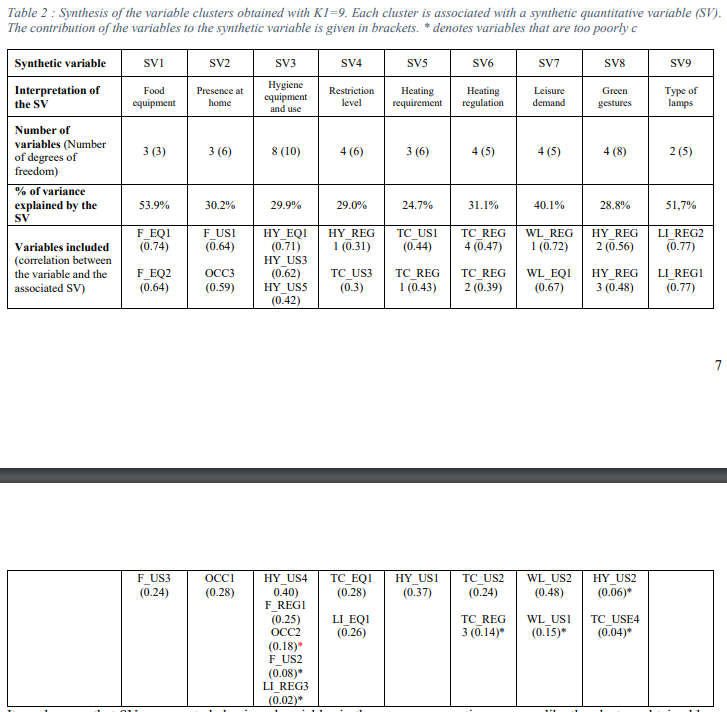
Puis vérification avec un test d’indépendance Chi2 pour vérifier l’indépendance des clusters

## Evaluation des performances :

Bootstrapping : opération de clustering avec 100 sous-datasets et mesure de la persistance des associations de variable

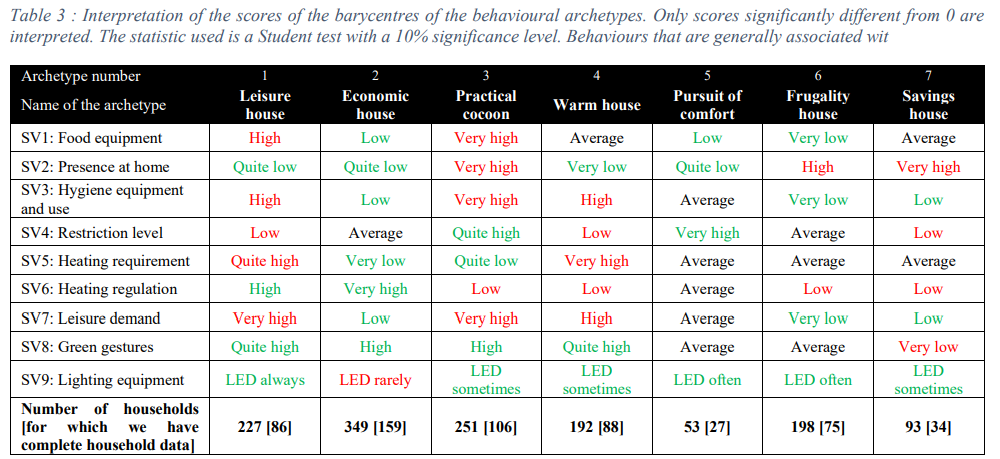
## Résultats



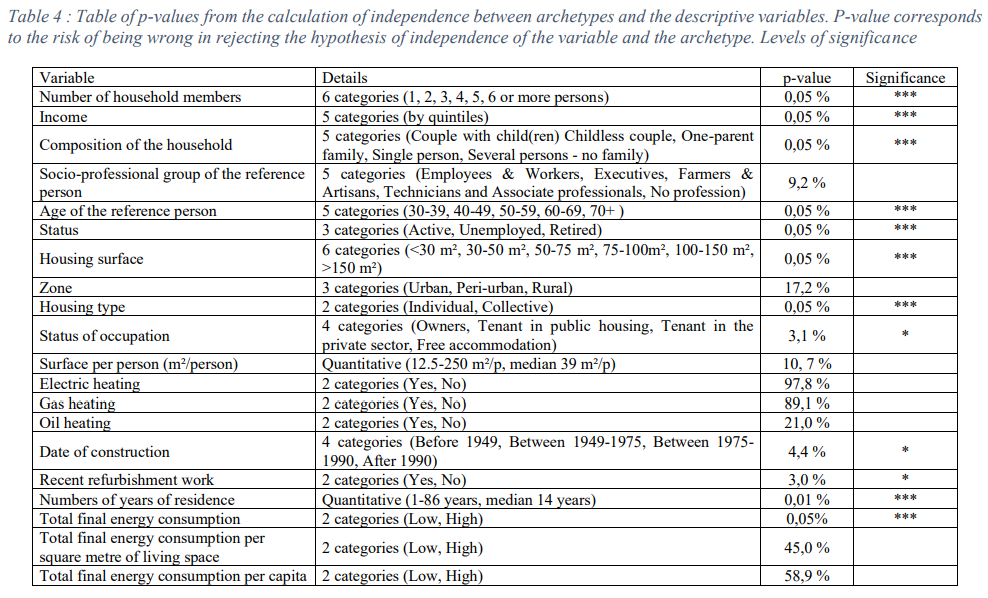


9 variables synthétiques

Ménages réparties en 7 clusters

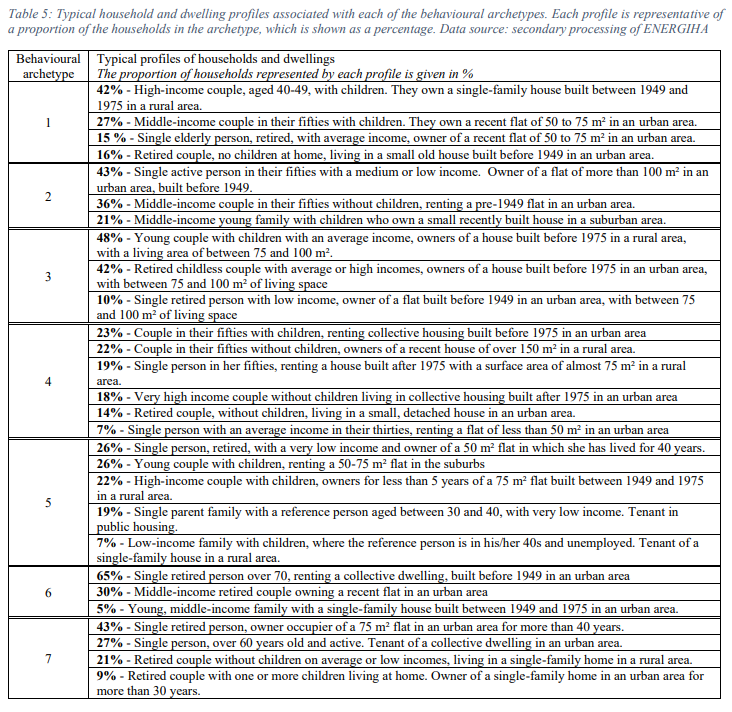


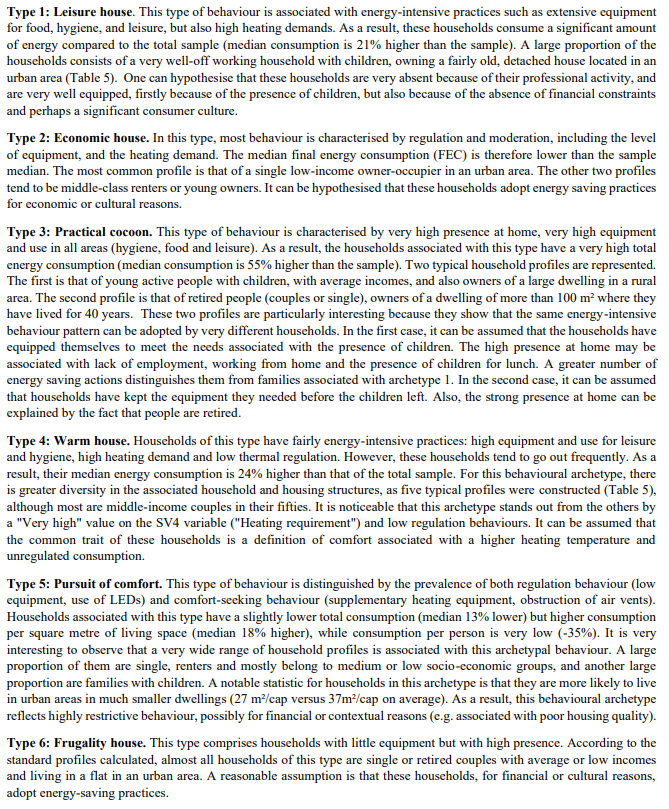
Etude de la corrélation entre les variables et leur SV

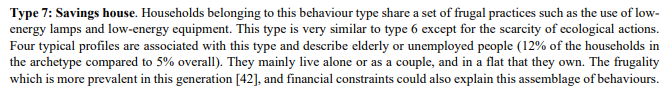


Le test d’indépendance montre des similarités entre certains archétypes comportementaux.









## Limites

Données anciennes de 2010, collectées par téléphone (possibles biais, erreurs ou manque d’informations), n’incluent pas les développements technologiques (appareils disponibles, temps et fréquence d’utilisation, télétravail, mobiles et numériques), limitées géographiquement