

# Estimation de la performance énergétique en phase amont

Travaux de stage

## Contexte et objectifs initiaux

Dans le cadre de leurs missions, les ingénieurs CVC d’AREP réalisent des simulations thermiques sous ClimaWin sur gares et technicentres, mais la saisie manuelle des paramètres repousse ces études en phase avancée.

AREP souhaite estimer dès la phase préliminaire la consommation énergétique à partir de grandeurs macroscopiques pour dégager rapidement des tendances. Ce stage vise à analyser la quantité, la qualité et la variabilité du corpus ClimaWin existant afin de définir un prototype d’outil statistique d’estimation précoce.

## Méthodologie

Etape 1 – Sélection des variables

- Identification, dans les rapports XML de ClimaWin et en lien avec les besoins métiers (ingénieurs CVC, architectes), des données d’entrée (géographiques, géométriques, scénarios) et de sortie (consommation, besoins énergétiques).

Etape 2 – Calcul des indicateurs de performance énergétique

- Confort thermique unifié selon la réglementation projet (RT2012 & RE2020)
- Consommation spécifique (besoins : chauffage + froid + éclairage) rapportée à la surface du bâtiment

Bâtiment	Géographique	Utilisation - équipements	Caractéristiques géométriques
Propriétés thermo physiques	Fenêtres - vitrages	Eclairage	Consommation - besoins énergétiques

Regroupement par type des variables du jeu de données

Etape 3 – Collecte des études

- Regroupement de 90 fichiers ClimaWin (RT2012 & RE2020), l’intégralité des simulations disponibles

Etape 4 – Bot d’extraction des XML

- Développement d’un petit robot Python pour récupérer automatiquement tous les fichiers XML associés aux projets ClimaWin collectés

Etape 5 – Structuration des données

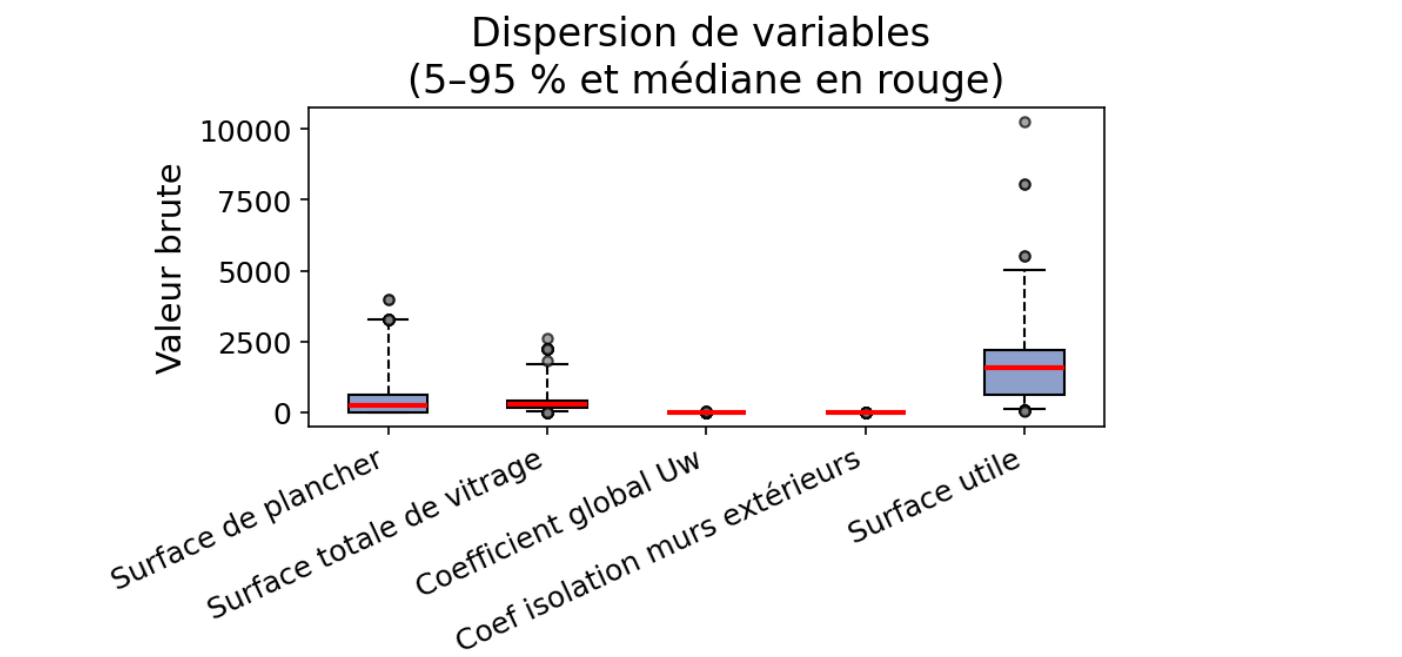
- Écriture d’un parseur Python qui scanne chaque XML et en extrait les balises identifiées en amont
- Stockage dans un tableau (DataFrame pandas) où chaque ligne = un bâtiment et chaque colonne = une variable métier
- Jeu de données ainsi prêt pour l’analyse statistique

Réglementation	Batiment	vraie surf utile	Uw_total	Conso_spec	Ratio_Global
RE2020	['BV', 'BAT 2.5']	1586.71	7.01264	30.1919	2.36715
RE2020	['BV', 'BAT 2.5']	3228.51	3.25354	14.8384	2.36715
RT2012	['Bâtiment A1']	10261.1	26.5015	22.1	0.610127
RE2020	['DALLE 01']	2119.3	7.10758	29.7	0.686171
RE2020	['Bâtiment n°1']	2217.66	7.62082	33.2025	1.38866
RE2020	['Bâtiment n°1']	2217.66	7.51004	45.1176	1.93786

Organisation du jeu de données : extrait du DataFrame final

## Analyse du jeu de données

Observation de la dispersion des variables. Les boîtes à moustache permettent de repérer les écarts et biais éventuels et visualiser la distribution des données.



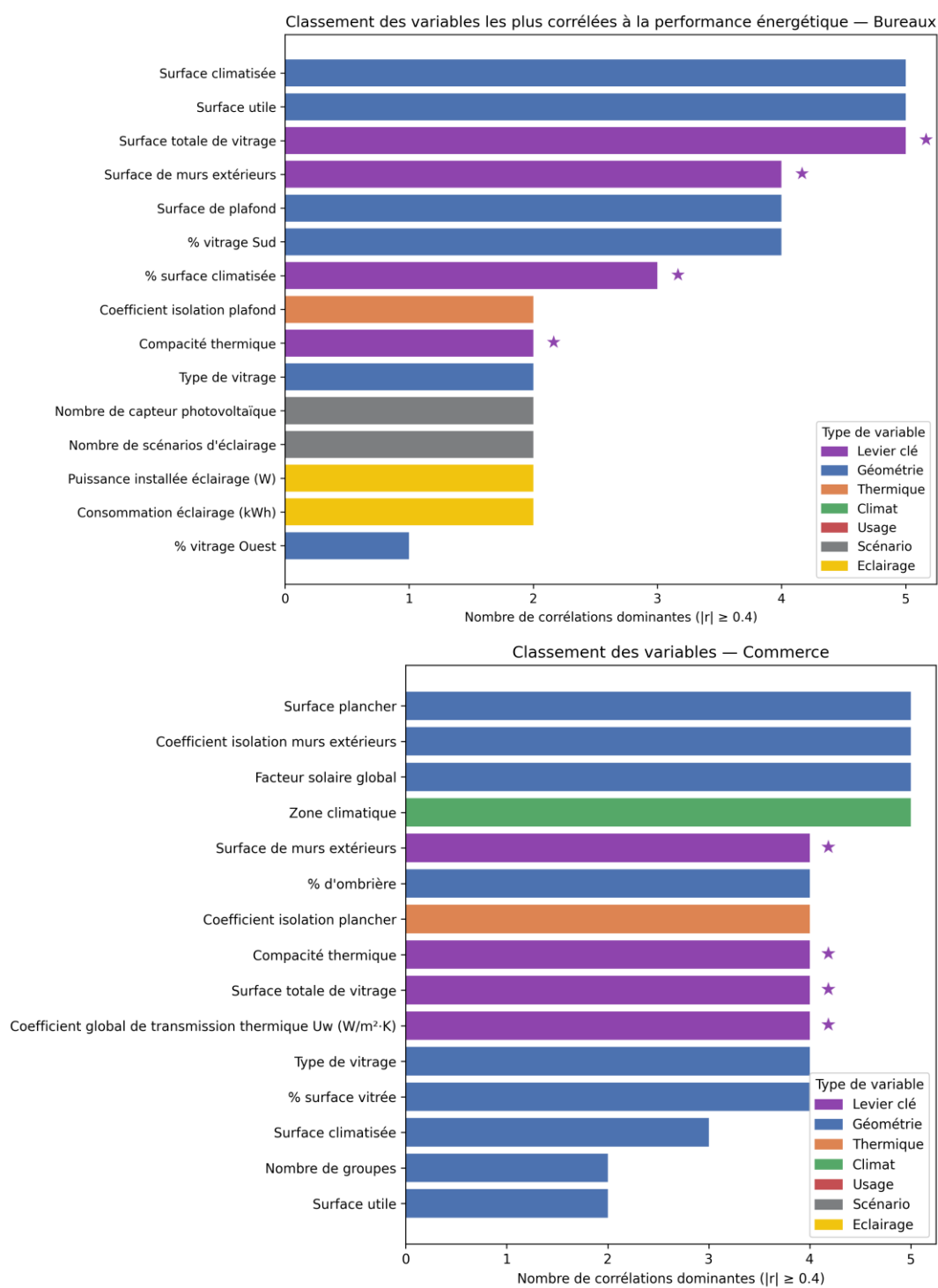
- La plupart des valeurs se situent dans un intervalle étroit, avec très peu de points en dehors des moustaches.
- Ce niveau d’homogénéité traduit une faible variabilité entre les projets recueillis, rendant tout apprentissage automatique impossible.

Deux axes d’approfondissement :

- Identifier les leviers clés de performance énergétique,
- Segmenter les bâtiments en familles homogènes en découpant chacun des 5 leviers clés en quintiles (5 groupes de 20 % classés du plus faible au plus élevé), puis comparer la répartition consommation / confort au sein de chaque groupe.

1 – Identification des leviers clés

J’ai calculé les coefficients de corrélation entre chaque variable d’entrée et nos deux indicateurs de sortie (conso spécifique & confort thermique). Les deux classements des 15 variables les plus fortement corrélées (coef de corrélation  $|r| \geq 0,4$ ) – « Bureaux » en haut et « Commerce » en bas – font ressortir cinq leviers décisifs (★) pour l’éco-conception :



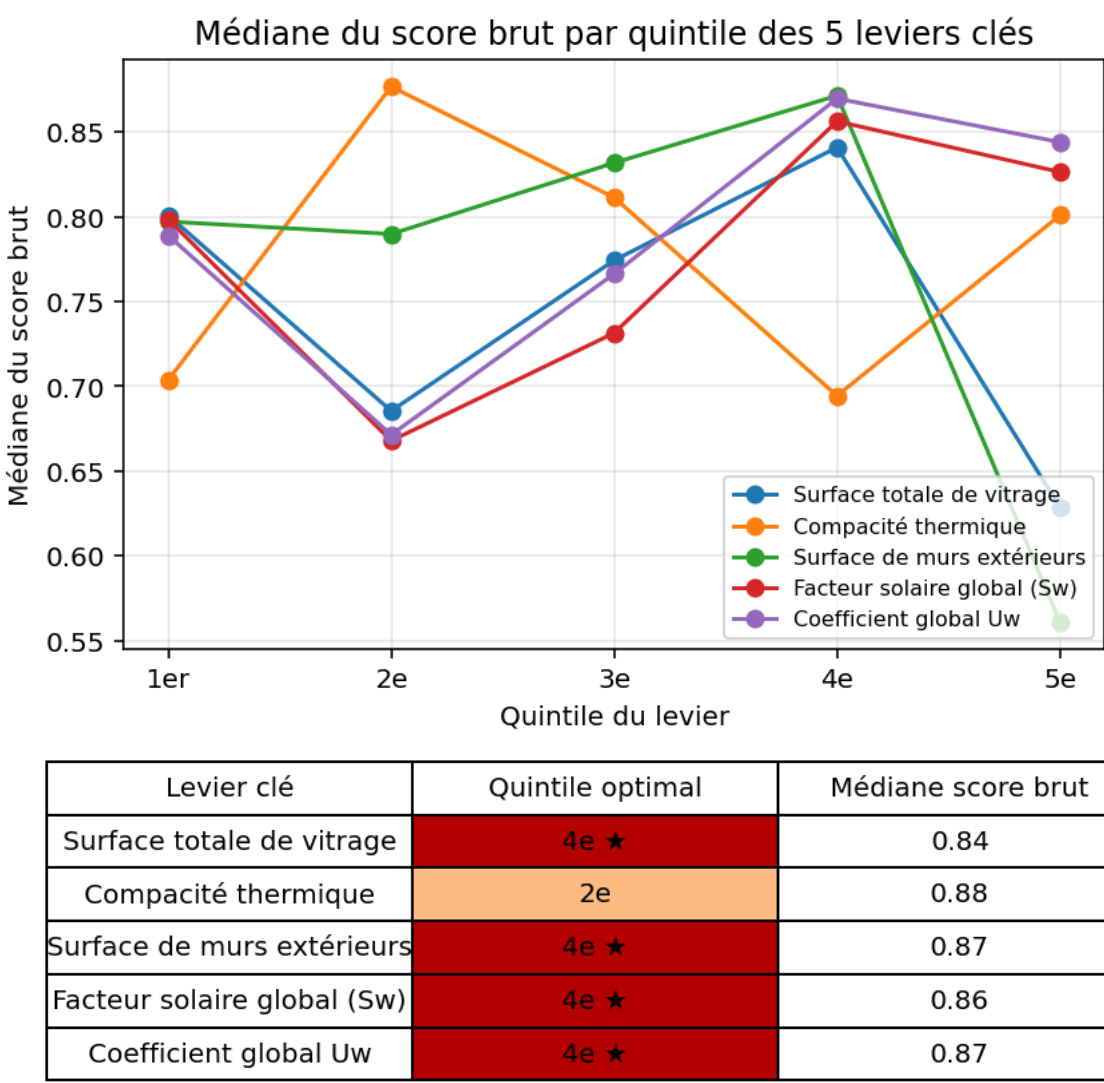
- Surface totale de vitrage (optimisation de l’éclairément naturel et apports solaires)
- Surface de murs extérieurs (masse et inertie thermique)
- Facteur solaire global (Sw) (maîtrise des gains solaires)
- Compacité thermique (rapport volume/surface)
- Coefficient global Uw (qualité d’isolation de l’enveloppe)

## Résultats finaux

2 – Segmentation par quintiles

Pour chacun des cinq leviers, les bâtiments ont été répartis en cinq quintiles.

J’ai tracé, pour chaque quintile, la relation conso spécifique vs confort thermique. J’ai défini un score brut (70 % conso spécifique / 30 % confort) et calculé sa médiane par quintile, pour repérer le groupe de bâtiments dont la performance énergétique moyenne est la plus élevée (score brut proche de 1).



AREP



IMT Nord Europe  
École Mines-Télécom  
IMT-Université de Lille

Étudiant.e Raphaël Launay  
Niveau d’études M1  
Spécialité Energie Environnement

Tuteur.ice école Bernard Baudoin

Durée 17 semaines  
Période Du 7 avril au 1<sup>er</sup> août 2025

Entité AREP Hypercube  
Tuteur.ice AREP Alexis Sauvageon

## EMC2B



Énergie Matière Carbone Climat Biodiversité

Cet outil d’estimation précoce de la performance énergétique, fondé sur cinq leviers clés, peut fournir des repères concrets pour :

- Réduire la consommation dès les phases préliminaires
- Optimiser les matériaux et l’isolation pour limiter les déperditions
- Garantir le confort thermique tout en maîtrisant l’empreinte carbone

Principaux enseignements :

Compacité thermique : 2<sup>e</sup> quintile  
→ meilleur compromis volume / inertie / surface

Les 4 autres leviers : médiane la plus haute au 4<sup>e</sup> quintile (★)  
→ dimensionnement modéré-élevé favorise isolation, inertie et apports contrôlés

Ces résultats confirment que notre segmentation met en lumière les zones de configuration les plus performantes, offrant le meilleur compromis consommation/confort, à partir de notre jeu de donnée.