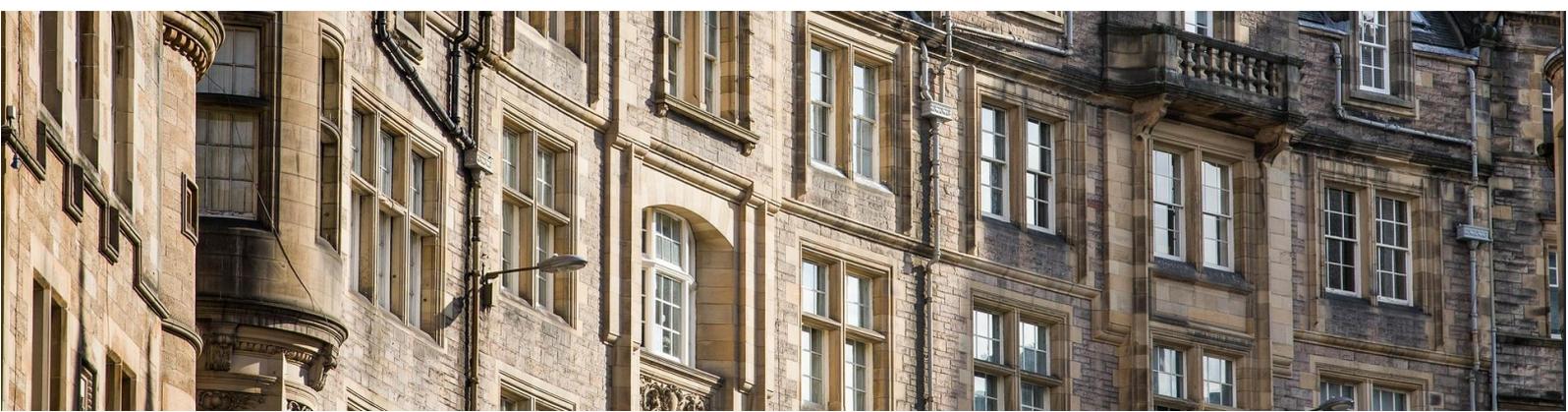


TAXINOMIE

NOTE METHODOLOGIQUE CALCUL DES INDICATEURS TOP15 ET TOP30

Décembre 2022

NOTE METHODOLOGIQUE



Sommaire

Sommaire.....	2
1. Contexte.....	3
2. Etude préliminaire.....	4
2.1 Définition des sous-typologies.....	4
2.2 Taille minimale des échantillons pour chaque sous-typologie.....	5
3. Création de la base de données utilisée en 2022.....	7
3.1 Traitement préliminaire de la base de données.....	7
3.2 Repérage et élimination des valeurs aberrantes de consommation d'énergie primaire.....	7
3.3 Occurrences uniques.....	8
4. Calcul des indicateurs.....	9
4.1 Correction géographique.....	9
4.2 Calcul des indicateurs taxinomiques.....	10
4.2.1 Indicateurs généraux.....	10
4.2.2 Indicateurs par sous-typologies.....	10
ANNEXE A : évaluation de la proportion p et calcul du nombre de Cochran n_c	12
A.1 Evaluation de p	12
A.2 Calcul de n_c	12
ANNEXE B : Choix et calcul des bornes.....	14
ANNEXE C : Biais géographique et pondération.....	16

1. Contexte

Le présent document s'inscrit dans le cadre de l'application de critères relatifs à l'enjeu d'atténuation de la Taxinomie européenne. Parmi ces critères d'atténuation du changement climatique, l'appartenance à un Top15 et Top30 de son marché régional ou national permet - si l'ensemble des autres critères est respecté - de démontrer respectivement une contribution substantielle à l'atténuation, ou une validation du *Do No Significant Harm* correspondant.

En 2020, l'OID a publié une première série d'indicateurs permettant aux acteurs d'évaluer le comportement de leurs patrimoines vis-à-vis de ces critères. Ces indicateurs ont été mis à jour régulièrement depuis cette date.

Actifs résidentiels

En octobre 2022, des indicateurs taxinomiques de top 15% et top 30% pour les bâtiments résidentiels ont été produits par le Ministère de la transition écologique et de la cohésion des territoires (MTE). Les indicateurs s'appliquent désormais officiellement, et ils n'ont donc pas été recalculés par l'OID dans sa mise à jour de janvier 2023.

Actifs non résidentiels

Les indicateurs tertiaires en revanche n'ont pas pour le moment été communiqués par le MTE qui attend de pouvoir exploiter les premiers résultats de remontée de données de consommations sur la plateforme OPERAT, plateforme mise en place dans le cadre du Dispositif Eco-énergie Tertiaire. Aussi, l'OID propose des indicateurs calculés sur les données collectées lors de la collecte du BPE 2022 pour les bâtiments tertiaires.

C'est la raison pour laquelle l'OID propose une nouvelle mise à jour des indicateurs taxinomiques associés en janvier 2023, permettant notamment :

- De calculer des indicateurs sur des typologies plus fines que la simple décomposition tertiaire/résidentiel précisée dans le référentiel de la Taxinomie ;
- D'améliorer la granularité des indicateurs en établissant également des indicateurs par sous-typologies pour avoir un niveau de compréhension et d'évaluation plus fin et plus juste du parc de bâtiments tertiaires français ;
- De proposer pour la première fois des indicateurs concernant certaines typologies d'actifs.

La décomposition des bâtiments tertiaires en typologies plus fines permet de refléter les différences évidentes entre des catégories de bâtiments très hétérogènes (bâtiments de bureaux, bâtiments de santé, etc).

Le présent document se veut être une explication détaillée des méthodes appliquées pour définir les sous-typologies, pour calculer les indicateurs et pour s'assurer de la cohérence des résultats. Un guide d'utilisation des indicateurs taxinomiques qui détaille également comment choisir entre une approche à l'échelle de la typologie ou de la sous-typologie est disponible dans un document explicatif séparé.

2. Etude préliminaire

2.1 DEFINITION DES SOUS-TYPOLOGIES

Pour certaines typologies, il semble pertinent de définir des sous-typologies plus précises. En effet, au sein d'une même typologie, les profils de consommation énergétique des bâtiments peuvent diverger significativement. Par exemple, pour la typologie « santé », les consommations d'énergie d'un EHPAD semblent instinctivement différentes de celles d'un CHU.

Aussi, les typologies « bureaux », « hôtels » et « santé » ont été subdivisées en sous-typologies. La définition des sous-typologies a été faite en concertation avec des experts et des acteurs du marché pour parvenir à une division qui soit pertinente au sens des consommations potentielles, et qui soit effectivement utile aux acteurs du marché.

Ainsi, les bureaux ont été divisés en :

- IGH (voir encadré ci-dessous),
- Bureaux haussmanniens,
- Bureaux haut de gamme ou à haute fonctionnalité,
- Bureaux standards RT2012,
- Bureaux standards RT1978 à 2005
- Autres bureaux

Dans le cas des bureaux, les immeubles de grande hauteur ont été traités à part. En effet, l'univers des IGH de bureaux est très limité : seulement une cinquantaine d'immeubles en France. Il est impossible de faire une étude statistique sur un univers si petit car évaluer un indicateur sur un échantillon d'une trentaine de bâtiments dans l'espoir qu'il représente un univers à peine deux fois plus grand laisse trop d'incertitude. Aussi, pour les IGH, on se contentera de calculer un top 15 et un top 30 qui représentent uniquement les bâtiments pour lesquels nous possédons des informations. Ils ne se veulent en aucun cas être une évaluation du véritable top 15 ou top 30 des IGH de bureaux pour l'ensemble du pays, au contraire de tous les autres indicateurs de top 15 et top 30 pour les autres sous-typologies qui sont construits comme une évaluation statistique de la véritable valeur du top 15 et du top 30 de leur sous-typologie pour la France entière.

Les sous-typologies des bâtiments de santé ont été définies en fonction du niveau de service et du niveau d'équipement qui est nécessaire pour assurer les services voulus. Ces catégories sont aussi en accord avec les classifications classiques des bâtiments de santé. On a donc les sous-typologies suivantes :

- Les établissements hospitaliers (médecine, chirurgie, obstétrique),
- Les établissements médico-sociaux,
- Les EHPAD

Quant aux hôtels, c'est leur classement qui détermine les sous-typologies, puisque leur classement détermine les tailles des chambres et les niveaux de services attendus. On a donc trois sous-typologies :

- Hôtels 1 et 2 étoiles,
- Hôtels 3 et 4 étoiles,
- Autres hôtels

2.2 TAILLE MINIMALE DES ECHANTILLONS POUR CHAQUE SOUS-TYPOLOGIE

Pour être certain que les indicateurs taxinomiques de consommation d'énergie primaire que nous produisons ont un sens, il faut s'assurer que l'échantillon pour lequel une contribution a été faite lors de la collecte 2022 est :

- d'une part représentatif de l'ensemble de la population totale ;
- d'autre part d'une taille suffisante pour qu'on puisse en extrapoler des indicateurs applicables à l'ensemble de la population.

L'échantillon est considéré **représentatif** : les données de consommation d'énergie ne sont apparemment pas biaisées. On peut le vérifier empiriquement en comparant la répartition des consommations de nos contributeurs à la répartition des consommations de la base des DPE (cf figure 1. pour exemple).

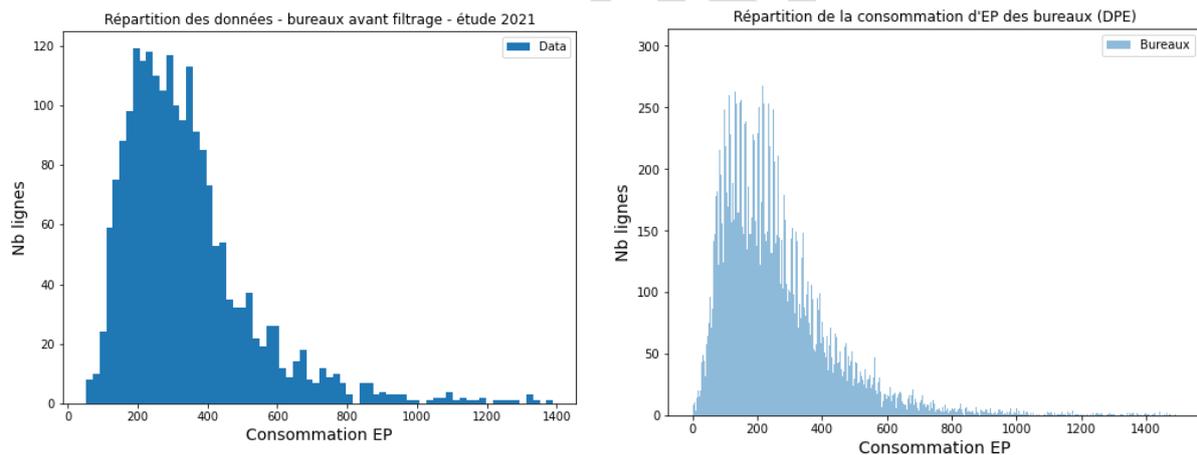


Figure 1. Gauche : Histogramme de la consommation d'énergie primaire des bureaux dans la collecte OID. Droite : Histogramme de la consommation d'énergie primaire des bureaux dans la base ADEME des DPE.

On considérera qu'un échantillon a une **taille suffisante** s'il respecte le critère de Cochran. C'est-à-dire qu'on considérera qu'on peut calculer des indicateurs qui ont du sens pour une sous typologie si la taille de l'échantillon est au moins égale au nombre de Cochran n_c . Ce nombre se calcule à partir de la part que représente la sous typologie étudiée dans la population totale ; et il dépend du niveau de confiance qu'on veut avoir dans l'indicateur qu'on calcule.

Pour être plus précis, le nombre de Cochran n_c est tel que $n_c = t^2 \frac{p(1-p)}{m^2}$, où :

- t est le quantile du niveau de confiance qu'on veut avoir sur l'estimation produite ;
- p est la part estimée de la population totale que représente la sous-typologie étudiée ;
- m est la marge d'erreur qu'on s'autorise sur la valeur même de l'indicateur calculé.

Dans notre cas, on fixe le niveau de confiance à 90% et la marge d'erreur à 5%. Ainsi t vaut 1,64 et m vaut 0,05 (voir annexe A pour le calcul de t). Quant à p , on estime la proportion de chaque sous-typologie dans le patrimoine français en étudiant des jeux de données fournis par l'ADEME et l'INSEE. Les valeurs de p et donc de n_c obtenues pour chaque sous typologie sont résumées dans le tableau 1.

	p	n_c
Bureaux		
Bureaux Haut de Gamme ou Haute Fonctionnalité	0,01	11
Bureaux standards RT 2012	0,05	51
Bureaux standards avant RT 2012	0,22	185
Hausmannien	0,02	21
Hotels		
1 & 2 étoiles	0,25	201
3 & 4 étoiles	0,48	269
Santé		
Centres hospitaliers	0,07	67
EHPAD	0,14	129
Etablissements médico-sociaux	0,77	191

Tableau 1. Répartition estimée par sous-typologie des actifs français et nombres de Cochran associés.

Ainsi, s'il y a au moins n_c bâtiments pour chaque sous-typologie, on est certain à 90% de donner un indicateur précis à 5% près.

Pour plus de précision sur l'évaluation de la proportion p pour chaque sous typologie et sur le calcul du nombre de Cochran n_c , voir l'annexe A.

Remarque : les sous-typologies présentées dans le tableau 1 ne sont pas définitives : les prochaines collectes de données plus larges et plus précises permettront une granularité plus fine dans l'étude que l'OID mène. D'autre part, les typologies qui ne sont pas encore sous-divisées feront l'objet de travaux ultérieurs pour aboutir à une distinction en sous-typologies.

3. Création de la base de données utilisée en 2022

3.1 TRAITEMENT PRELIMINAIRE DE LA BASE DE DONNEES

A partir de la base de données complète issue de la collecte du BPE, on sélectionne toutes les données rapportées en 2019, 2020, 2021.

Application d'un filtre pour d'une part conserver les valeurs exploitables, et d'autre part supprimer les valeurs visiblement erronées. On conserve ainsi les lignes pour lesquelles :

- La surface de reporting est positive non nulle ;
- Le taux de vacance est inférieur à 10%
- La valeur de consommation d'énergie primaire est positive non nulle
- La valeur de consommation d'énergie finale est supérieure à 50kWh/m²/an

La base de données obtenue est appelée *base de données préliminaire* dans la suite de ce document.

3.2 REPERAGE ET ELIMINATION DES VALEURS ABERRANTES DE CONSOMMATION D'ENERGIE PRIMAIRE

La *base de données préliminaire* est répartie selon les sous typologies. Pour chaque sous typologie, on va chercher à éliminer les éventuelles données pour lesquelles la consommation d'énergie primaire serait aberrante et qui risqueraient donc de fausser le calcul des indicateurs.

Pour ce faire, on détermine :

- le premier quartile Q_1 de consommation d'énergie primaire de l'échantillon étudié, c'est-à-dire, la consommation minimale telle que 25% des consommations lui soient inférieures ; et
- le troisième quartile Q_3 de consommation d'énergie primaire de l'échantillon étudié, c'est-à-dire, la consommation maximale telle que 25% des consommations lui soient supérieures.

Ainsi, 50% des valeurs de consommation d'énergie primaires sont comprises entre Q_1 et Q_3 . Elles se situent également « au milieu » de notre échantillon (voir figure 2).

Il est donc désormais possible de déterminer l'écart interquartile $IQR = Q_3 - Q_1$ qui servira de référence pour repérer les valeurs qu'on considérera comme aberrantes. Au vu de la répartition des données collectées (voir figure 1), on considérera que les données sont aberrantes si la consommation est :

- inférieure à la borne inférieure $Q_1 - 1 \times IQR^1$; ou
- supérieure à la borne supérieure $Q_3 + 2 \times IQR^1$.

Seront exclues de l'étude toutes les données qui ne respectent pas ce critère. La répartition des données conservées est illustrée par la figure 3.

¹ Pour plus de détails concernant le choix de bornes, se référer à l'annexe B

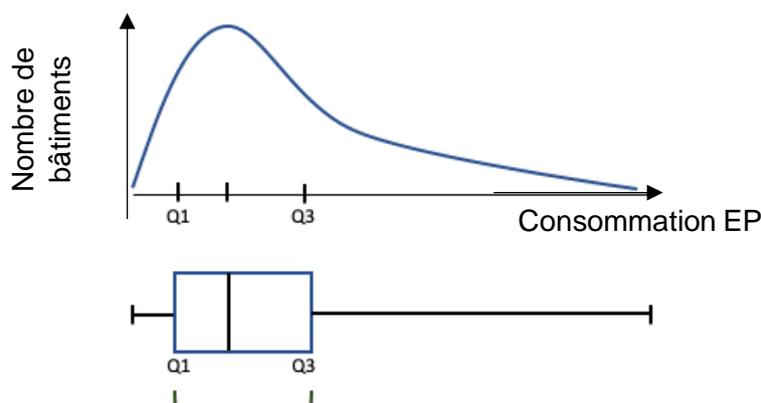


Figure 2. Illustration du principe de recherche des valeurs aberrantes.

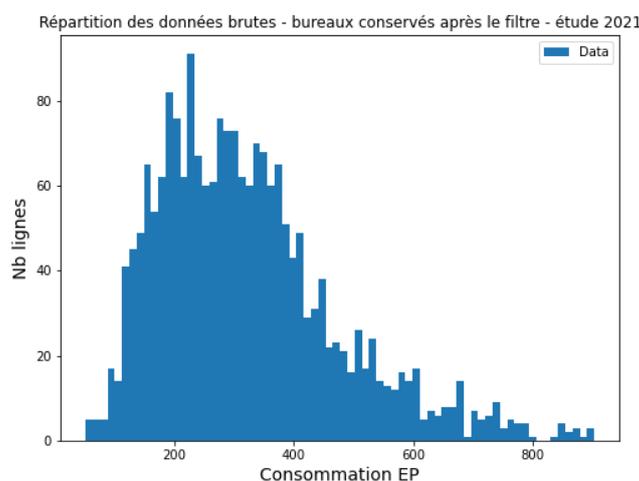


Figure 3. Histogramme de la consommation d'énergie primaire dans la base de données préliminaire.

3.3 OCCURRENCES UNIQUES

La *base de données préliminaire* comprend à ce stade des données recueillies sur 3 ans. Pour obtenir la base de données finale, on ne conserve que l'occurrence la plus récente de chaque bâtiment dans le cas où les données auraient été transmises plusieurs fois sur les trois dernières années. Cette sélection effectuée, on obtient la base de travail définitive qu'on appellera la *base de données finale* dans le reste de ce document. L'influence de chaque année dans la *base de données finale* est résumée dans le tableau 2. Le tableau 3 résume le nombre de valeurs par sous typologie dans la *base de données finale*.

Année	bureaux	santé	hébergement	commerce
2019	17%	1%	0%	14%
2020	23%	15%	29%	10%
2021	60%	84%	71%	76%

Tableau 2. Poids des années de reporting dans la base de données finale.

	Etat des données
Bureaux (hors IGH)	Suffisant
Hausmannien	Suffisant
Bureaux Haut de Gamme ou Haute Fonctionnalité	Suffisant
Bureaux standards avant RT 2012	Suffisant
Bureaux standards après RT 2012	Suffisant
IGH	Suffisant
Santé	Suffisant

Centres hospitaliers publics ou privés	Suffisant
Etablissements médico-sociaux	Suffisant
EHPAD	Suffisant
Hébergement	Suffisant
Hôtels 1 et 2 étoiles	Suffisant
Hôtels 3 et 4 étoiles	Suffisant
Logistique/stockage²	Insuffisant
Centre Commercial	Suffisant

Tableau 3. Nombre de données par sous-typologie dans la base de données finale.

Remarque : pour chaque sous-typologie retenue dans l'étude préliminaire, le nombre de données conservées dépasse bien le nombre minimal requis calculé précédemment. On peut considérer que l'évaluation des indicateurs taxinomiques sur chaque sous-typologie à partir de la *base de données finale* a du sens.

4. Calcul des indicateurs

4.1 CORRECTION GEOGRAPHIQUE

La collecte du BPE 2022 est fortement biaisée en faveur de la métropole parisienne. Une comparaison avec la base de données ADEME des DPE donne une estimation de la véritable répartition du patrimoine bâti français. La figure 4 illustre cette différence de répartition.

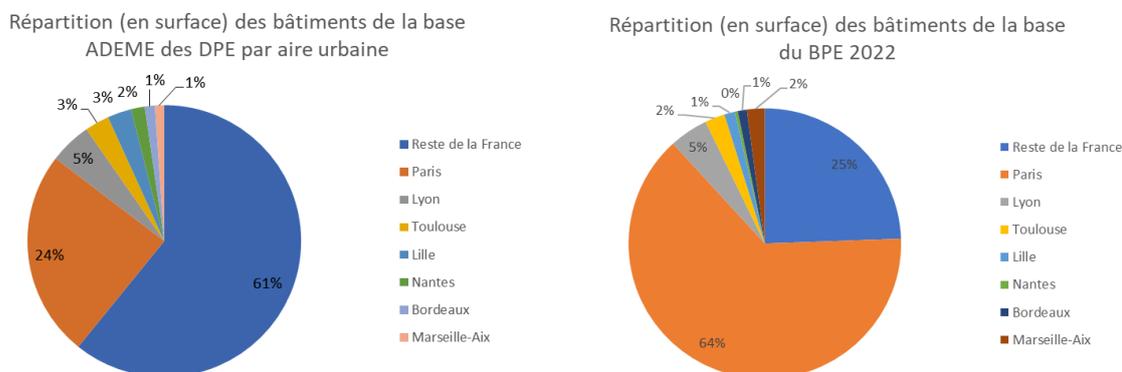


Figure 4. Gauche : Répartition géographique de la base de données ADEME des DPE ; Droite : Répartition géographique de la base du BPE 2022

Pour compenser ce biais géographique et donner des indicateurs taxinomiques pertinents sur l'ensemble du territoire, nous appliquons un coefficient de correction géographique pour le calcul de certains indicateurs. Pour tenir compte de la différence répartition de la *base de données finale* et de celle des DPE, nous appliquons le découpage géographique suivant :

- Unité urbaine de Paris,
- Unité urbaine de Lyon,
- Unité urbaine de Toulouse,
- Unité urbaine de Lille,
- Unité urbaine de Nantes,
- Unité urbaine de Bordeaux,

² La catégorie logistique a été étudiée mais ne fait pas l'objet d'un calcul d'indicateurs taxinomiques car le nombre de données collecté n'est pas suffisant pour produire des indicateurs qui soient suffisamment fiables.

- Unité urbaine de Marseille-Aix,
- Reste de la France.

La définition des unités urbaines est donnée par la [base des unités urbaines 2020](#) de l'INSEE.

Pour chacune de ces unités (y compris le reste de la France), on calcule le coefficient de correction géographique $c_{\text{unité}}$ défini par :

$$c_{\text{unité}} = \frac{\text{Surface}_{\text{unité}}^{\text{DPE}}}{\text{Surface}_{\text{totale}}^{\text{DPE}}} \times \frac{\text{Surface}_{\text{totale}}^{\text{BPE 2022}}}{\text{Surface}_{\text{unité}}^{\text{BPE 2022}}}$$

La correction du biais géographique s'effectue de la façon suivante :

- Détermination de l'unité géographique à laquelle appartient chaque bâtiment recueilli ;
- Calcul d'une surface fictive pondérée : $S_{\text{pondérée}} = S_{\text{initiale}} \times c_{\text{unité}}$

Pour plus de détails sur l'évaluation du biais géographique et sur sa correction, se référer à l'annexe C.

4.2 CALCUL DES INDICATEURS TAXINOMIQUES

4.2.1 INDICATEURS GÉNÉRAUX

La *base de données finale* est classée par ordre croissant de consommation d'énergie primaire. On effectue ensuite la somme cumulée des surfaces pondérées.

On calcule alors 15% et 30% de la surface totale.

La valeur du top 15 en consommation d'énergie primaire est donc celle associée à la ligne de la première surface cumulée dépassant le seuil de 15% de la surface totale.

La figure 3 illustre avec un exemple fictif, le processus de recherche du top 15.

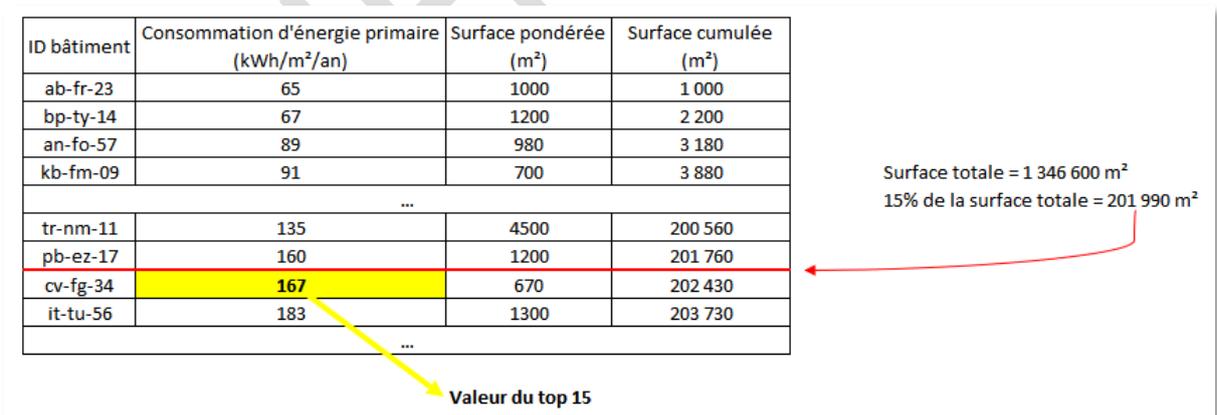


Figure 3. Illustration du processus de recherche du top 15.

Le processus de recherche du top 30 est entièrement similaire : il correspond à la consommation associée à la ligne de la première surface cumulée dépassant le seuil de 30% de la surface totale.

4.2.2 INDICATEURS PAR SOUS-TYPOLOGIES

La démarche de recherche des indicateurs top 15 et top 30 par sous-typologie est en tout point similaire.

On filtre la *base de données finale* pour ne garder que les lignes correspondant à la sous typologie étudiée. On ordonne la base filtrée par ordre croissant d'énergie primaire consommée. On effectue ensuite la somme cumulée des surfaces pondérées et on applique exactement la même méthodologie de recherche que pour les indicateurs généraux.

CONFIDENTIAL

ANNEXE A : évaluation de la proportion p et calcul du nombre de Cochran n_c

A.1 EVALUATION DE p

On extrait de la base de données ADEME des DPE³ toutes les entrées dont l'attribut « secteur_activite » contient la chaîne de caractères « bureau ».

La base de données ADEME des DPE ne contient pas d'information spécifique pour classer ses entrées dans les sous-typologies que nous avons déterminées. On évalue les proportions de chaque catégorie par d'autres moyens, en s'appuyant sur d'autres critères renseignés :

- On considèrera que les bureaux construits entre 1870 et 1920 sont haussmanniens ;
- On considèrera que les bureaux construits en 2014 et après sont RT 2012 ;
- On considèrera que les bureaux construits entre 1975 et 2013 sont RT 1974 à 2005 ;
- On considèrera que 1% des bureaux sont hauts de gamme ou à haute fonctionnalité ;
- On classera tous les autres bureaux dans « autres ».

La répartition des hôtels du parc français par classement est donnée directement par l'INSEE⁴.

Le nombre d'établissements de santé⁵ et le nombre d'EHPAD⁶ sont aussi fournis par l'INSEE, et le ministère des solidarités donne le nombre d'établissements médico-sociaux⁷.

On peut alors calculer pour chaque typologie la proportion p qu'elle représente au sein de la typologie mère. Les valeurs de p obtenues sont rassemblées dans le tableau A.1.

A.2 CALCUL DE n_c

Pour chaque sous typologie on calcule le nombre de Cochran n_c tel que $n_c = t^2 \frac{p(1-p)}{m^2}$, où t est le quantile associé au niveau de confiance, p est la proportion estimée de la population qui présente la caractéristique étudiée, et m est la marge d'erreur.

On choisit un niveau de confiance à 90%. Une [table de la loi de Student](#) à l'infini nous donne la valeur de $t = 1,64$ pour un tel niveau de confiance. On choisit m comme on le souhaite. Dans notre cas, on choisit $m = 0,005$ pour une marge d'erreur de 5%.

Avec les proportions p calculées précédemment, on obtient facilement n_c (voir tableau A.1.).

³ <https://www.data.gouv.fr/fr/datasets/dpe-tertiaire-avant-juillet-2021/> et <https://www.data.gouv.fr/fr/datasets/dpe-tertiaire-depuis-juillet-2021/>

⁴ https://www.insee.fr/fr/statistiques/2015412#tableau-figure1_radio1

⁵

<https://www.insee.fr/fr/statistiques/4277748?sommaire=4318291#:~:text=Parmi%20les%201%20356%20%C3%A9tablissements,des%20%C3%A9tablissements%20de%20soins%20de>

⁶ <https://www.insee.fr/fr/statistiques/3676717?sommaire=3696937>

⁷ <https://solidarites->

sante.gouv.fr/IMG/pdf/guide_activites_et_metiers_du_secteur_social_medico_social_et_petite_enfance_sept_2010.pdf

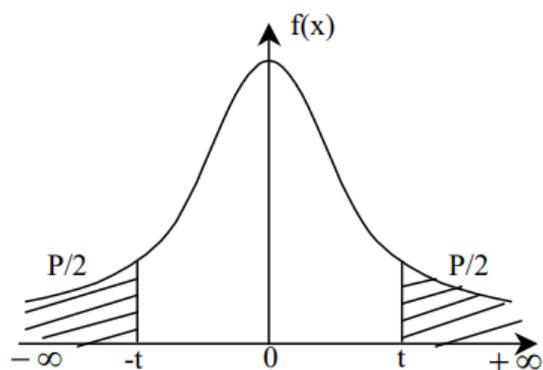


Figure A.1. Loi de Student.

	p	n_c
Bureaux		
Bureaux Haut de Gamme ou Haute Fonctionnalité	0,01	11
Bureaux standards RT 2012	0,05	51
Bureaux standards avant RT 2012	0,22	185
Hausmannien	0,02	21
Hotels		
1&2 étoiles	0,25	201
3&4 étoiles	0,48	269
Santé		
Etablissement de santé	0,07	67
EHPAD	0,14	129
Médico-sociaux	0,77	191

Tableau A.1. Taille minimale des échantillons à recueillir par sous-typologie dans la collecte du BPE 2022.

ANNEXE B : Choix et calcul des bornes

La façon usuelle de traiter les valeurs aberrantes dans une étude statistique est de considérer une plage d'étude de 4 écarts interquartiles autour de la médiane. En effet si la distribution de nos données récoltées suivait une répartition gaussienne, on pourrait s'assurer que plus de 99% des données se trouvent dans l'intervalle $[Q_1 - 1,5 \times IQR ; Q_3 + 1,5 \times IQR]$, où Q_1 et Q_3 sont le premier et troisième quartile, et où $IQR = Q_3 - Q_1$ est l'écart interquartile (voir figure B.1).

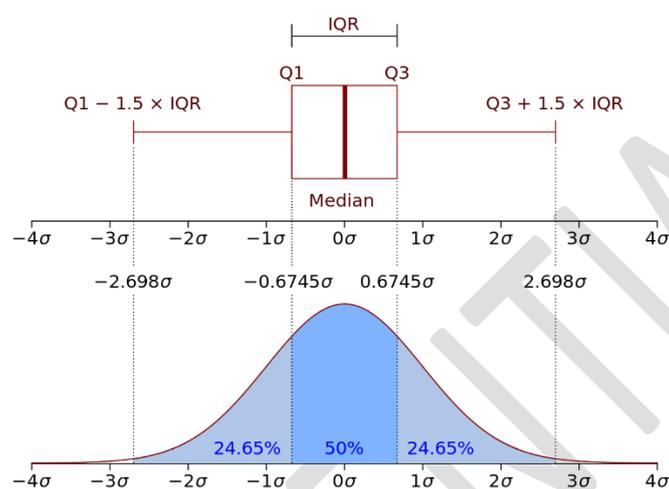


Figure B.1. Densité de probabilité d'une loi normale et répartition.

Cependant, les données que nous étudions dans la base sont décalées à droite (cf. fig. 1 gauche pour exemple et fig. B.2 pour un schéma). En effet, les actifs sont rassemblés autour de la médiane, mais sont coincés à gauche puisqu'une consommation d'énergie est évidemment positive. La distribution des actifs n'est en revanche pas limitée à droite et des actifs peu efficaces existent encore dans le patrimoine.

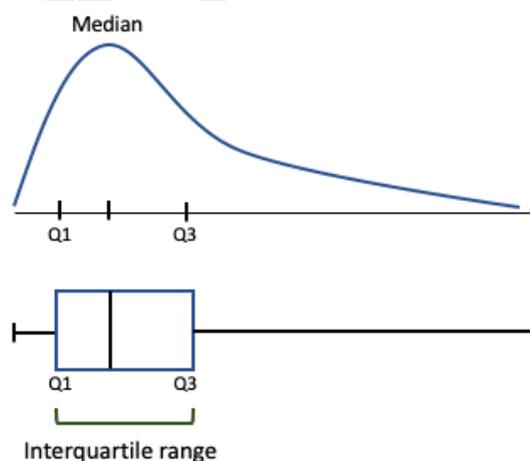


Figure B.2. Illustration de la recherche du biais de la distribution de nos données.

Aussi, dans les cas d'une distribution décalée à droite, on ne peut pas considérer comme avec une distribution centrée qu'on pourra se débarrasser des valeurs aberrantes avec un filtre $[Q_1 - 1,5 \times IQR ; Q_3 + 1,5 \times IQR]$. Notre choix s'est porté sur un filtre $[Q_1 - 1 \times IQR ; Q_3 + 2 \times IQR]$, ce qui prend en compte le décalage à droite de la distribution et donne des bornes acceptables pour un intervalle de contrôle, comme le montre le tableau B.1.

Typologie	Q1	Q3	IQR	Borne inférieure	Borne supérieure
Hotels					
Hôtels 1/2 étoiles	295	377	82	213	541
Hôtels 3/4 étoiles	253	361	108	145	577
Santé					
Centres hospitaliers publics ou privés	317,75	513	195,25	122,5	903,5
Etablissements médico-sociaux	183	254,75	71,75	111,25	398,25
EHPAD	265	319	54	211	427
IGH	273,5	487,5	214	59,5	915,5
Bureaux (hors IGH)					
Hausmannien	227	401	174	53	749
Bureaux Haut de Gamme ou Haute Fonctionnalité	237,25	501	263,75	-26,5	1028,5
Bureaux standards après RT 2012	179,5	331,5	152	27,5	635,5
Bureaux standards avant RT 2012	224,25	455	230,75	-6,5	916,5
Logistique/stockage	134,5	286,5	152	-17,5	590,5

Tableau B.1. Premiers et troisièmes quartiles par sous-typologie et limites de recherche des valeurs aberrantes.

ANNEXE C : Biais géographique et pondération

La première étape pour repérer le biais géographique de la typologie bureau, est d'avoir une répartition des bureaux en France de référence. On crée cette répartition de référence grâce à la base de données des DPE qui recense l'essentiel des bâtiments français. Dans la base des DPE tertiaire, on extrait tous les DPE qui décrivent un bureau ou assimilé.

A partir de la [base de données des aires urbaines de l'INSEE](#), on associe chaque commune à son aire urbaine. On peut donc répartir chaque bureau dans la métropole qui le concerne. Les 6 métropoles les plus représentées dans nos collectes sont également les plus représentées dans la base de données de DPE. Aussi, on cantonne notre étude à ces 6 villes.

On fait alors la somme de toutes les surfaces recensées dans notre collecte BPE et dans la base des DPE par aire urbaine. Le tableau C.1 donne la répartition que nous obtenons.

	Surface des bureaux (DPE)	Surface des bureaux (BPE)	Coefficient de l'unité urbaine
Reste de la France	20166986	3242870	2,4898
Paris	8152370	8470690	0,3853
Lyon	1647178	611888	1,0778
Toulouse	959654	313848	1,2242
Lille	936634	167822	2,2345
Nantes	541823	46834	4,6318
Bordeaux	391916	141062	1,1123
Marseille-Aix	365609	281827	0,5194
Total	33162173	13276842	1

Tableau C.1. Répartitions des surfaces et coefficients multiplicateurs.

Ainsi, pour l'étude des bureaux, afin de rendre compte du biais géographique en faveur des bureaux parisiens dans notre collecte, on pondère pour le calcul des top 15 et top 30 les surfaces de chaque bureau en fonction de l'unité urbaine à laquelle il appartient. Le coefficient d'une unité urbaine précise est défini ainsi :

$$c_{\text{unité}} = \frac{\text{Surface}_{\text{unité}}^{\text{DPE}}}{\text{Surface}_{\text{totale}}^{\text{DPE}}} \times \frac{\text{Surface}_{\text{totale}}^{\text{BPE 2022}}}{\text{Surface}_{\text{unité}}^{\text{BPE 2022}}}$$

Par exemple : un bureau parisien de 100m² comptera comme un bureau de 38m² alors qu'un bureau situé hors des six villes étudiées d'une surface de 100m² comptera comme un bureau de 248m².