

Quelle répartition de la consommation d'eau potable entre les différents usagers sur le bassin Seine-Normandie ?

Analyse statistique

AESN. Auteurs : Jules Barbazanges, Théophile Clerc

2024

Table des matières

RESUME	1
CONTEXTE DE L'ETUDE	3
RESULTATS ET DISCUSSION	4
1. QUATRE MODELES RETENUS	4
a. Structure des modèles	4
b. Performances	6
2. INFLUENCE DU TOURISME	8
a. Analyse des données observées	9
b. Projections sur les communes touristiques	11
3. RESULTATS DES PROJECTIONS SUR TOUTES LES COMMUNES	12
a. Consommation projetée par échantillon	12
b. Consommation projetée totale	17
c. Spatialisation	19
4. INTERPRETATION ET VALIDITE DES MODELES	21
a. Interprétation et cohérence des coefficients des modèles	22
b. Comparaison avec les données fournies et intégrées aux modèles	30
c. Comparaison avec les données fournies non intégrées aux modèles	34
MATERIELS ET METHODES	36
1. RECOLTE ET TRAITEMENT DES DONNEES	37
a. Eau potable	37
b. Données communales	40
2. CONSTRUCTION DES MODELES	42
a. La régression linéaire	43
b. Construction des échantillons	44
c. Variables utilisées	45
d. Autres variables potentiellement intéressantes	47
e. Méthodes de recherche des modèles	47
3. PROJECTIONS	49
LIMITES ET AXES D'AMELIORATION	50
1. BIAIS DES DONNEES	50
2. QUALITE DES MODELES	51
a. Précision des modèles	51
b. Hypothèses de la régression linéaire	51
c. Variabilité et fonctionnalité des modèles	52
d. Omissions de variables pertinentes	53
e. Temporalité	53
f. Viabilité des projections	54
ANNEXES	55
ANNEXE I : TYPOLOGIES FOURNIES PAR LES DIFFERENTS SERVICES D'EAU	55
ANNEXE II : MODELES OBTENUS PAR LA FONCTION « REGBEST » ET LA METHODE ITERATIVE POUR CHAQUE TYPOLOGIE DE COMMUNE	56
ANNEXE III : GRAPHIQUES DES RESIDUS, RESULTATS DES TESTS DE BREUSCH-PAGAN ET VIF CALCULES POUR LES MODELES RETENUS	61
ANNEXE IV : SPATIALISATION DES CONSOMMATIONS PROJETEES POUR CHAQUE SEGMENT D'USAGER	67
ANNEXE V : RESULTAT DES VOLUMES PRELEVES PREDITS POUR CHAQUE SEGMENT D'USAGER	77
ANNEXE VI : BOITE A MOUSTACHE DES CONSOMMATIONS TOTALES ET TOTALES PAR HABITANT OBSERVEES SUR LES COMMUNES TOURISTIQUES ET NON TOURISTIQUES	78

Table des illustrations

FIGURE 1 : COMMUNES CONSIDEREES COMME TOURISTIQUES SUR LE BASSIN	8
FIGURE 2 : BOITE A MOUSTACHE DE LA CONSOMMATION TOTALE OBSERVEE SUR LES COMMUNES TOURISTIQUES ET NON TOURISTIQUES, POUR LA TYPOLOGIE DE COMMUNES DE 1 000 A 5 000 HABITANTS	10
FIGURE 3 : BOITE A MOUSTACHE DE LA CONSOMMATION TOTALE PAR HABITANT OBSERVEE SUR LES COMMUNES TOURISTIQUES ET NON TOURISTIQUES, POUR LA TYPOLOGIE DE COMMUNES DE 1 000 A 5 000 HABITANTS.....	10
FIGURE 4 : CONSOMMATION D'EAU TOTALE PROJETEE, PAR COMMUNE ET A L'ECHELLE DU BASSIN VERSANT.....	19
FIGURE 5 : CONSOMMATION D'EAU TOTALE PROJETEE RAPPORTEE AU NOMBRE D'HABITANTS, PAR COMMUNE ET A L'ECHELLE DU BASSIN VERSANT	20
FIGURE 6 : BOITE A MOUSTACHE DES CONSOMMATIONS LIEES AUX RESIDENCES PRINCIPALES RAPPORTEES AU NOMBRE D'HABITANTS, PAR TYPOLOGIE DE COMMUNE	22
FIGURE 7 : COMMUNES DU BASSIN POUR LESQUELLES ONT ETE RECUEILLIES DES DONNEES.....	38
FIGURE 8 : DISTRIBUTION SPATIALE DES TYPOLOGIES DE COMMUNES PAR LE NOMBRE D'HABITANTS.....	44

Liste des tableaux

TABLEAU 1 : MODELE EXPLICATIF RETENU POUR LES COMMUNES DE PLUS DE 10 000 HABITANTS.....	4
TABLEAU 2 : MODELE EXPLICATIF RETENU POUR LES COMMUNES DE 5 000 A 10 000 HABITANTS.....	4
TABLEAU 3 : MODELE EXPLICATIF RETENU POUR LES COMMUNES DE 1 000 A 5 000 HABITANTS.....	5
TABLEAU 4 : MODELE EXPLICATIF RETENU POUR LES COMMUNES DE MOINS DE 1 000 HABITANTS.....	5
TABLEAU 5 : INDICATIONS SUR LA PERFORMANCE DES MODELES	6
TABLEAU 6 : ESTIMATION DE LA PRECISION DES MODELES A L'AIDE DE L'ERREUR STANDARD RESIDUELLE	6
TABLEAU 7 : ANALYSE DES DONNEES FOURNIES DE CONSOMMATION D'EAU TOTALE ET PAR HABITANT, ENTRE COMMUNES TOURISTIQUES ET NON TOURISTIQUES	9
TABLEAU 8 : CONSOMMATIONS TOTALES DES COMMUNES TOURISTIQUES.....	11
TABLEAU 9 : RESULTAT DES PROJECTIONS POUR LES COMMUNES DE PLUS DE 10 000 HABITANTS.....	12
TABLEAU 10 : RESULTAT DES PROJECTIONS POUR LES COMMUNES DE 5 000 A 10 000 HABITANTS.....	13
TABLEAU 11 : RESULTAT DES PROJECTIONS POUR LES COMMUNES DE 1 000 A 5 000 HABITANTS.....	14
TABLEAU 12 : RESULTAT DES PROJECTIONS POUR LES COMMUNES DE MOINS DE 1 000 HABITANTS.....	14
TABLEAU 13 : RESULTAT DES PROJECTIONS POUR PARIS INTRA-MUROS.....	15
TABLEAU 14 : CONSOMMATIONS PROJETEES TOTALE ET DOMESTIQUE, RAPPORTEES AU NOMBRE D'HABITANTS	16
TABLEAU 15 : RESULTAT DES PROJECTIONS A L'ECHELLE DU BASSIN ET PAR SEGMENTS D'USAGER	18
TABLEAU 16 : RENDEMENTS DES RESEAUX PAR TYPOLOGIE DE COMMUNE ET RESULTATS DES VOLUMES PRELEVES ESTIMES A L'AIDE DES RENDEMENTS.....	19
TABLEAU 17 : COEFFICIENTS CALCULES POUR LE SECTEUR DES MENAGES, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE.....	21
TABLEAU 18 : COEFFICIENTS CALCULES POUR LE SECTEUR TERTIAIRE ET DES COMMERCE, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE	23
TABLEAU 19 : COEFFICIENTS CALCULES POUR LE SECTEUR DE L'ENSEIGNEMENT, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE.....	26
TABLEAU 20 : COEFFICIENTS CALCULES POUR LE SECTEUR DE LA SANTE, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE.....	27
TABLEAU 21 : COEFFICIENTS CALCULES POUR LES EQUIPEMENTS SPORTIFS, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE.....	28
TABLEAU 22 : COEFFICIENTS CALCULES POUR LE SECTEUR DE L'AGRICULTURE, ESTIMATION DE LA COHERENCE ET VALEURS DE LA LITTERATURE ..	29
TABLEAU 23 : RESULTATS DES PRELEVEMENTS PREDITS PAR LES PROJECTIONS AJUSTEES ET LES HYPOTHESES	35
TABLEAU 24 : CARACTERISTIQUES DES COMMUNES POUR LESQUELLES DES DONNEES ONT ETE RECUEILLIES, PAR RAPPORT AUX COMMUNES DU BASSIN.....	39
TABLEAU 25 : CARACTERISTIQUES DES ECHANTILLONS DE COMMUNES	43
TABLEAU 26 : LISTE DES VARIABLES EXPLICATIVES UTILISEES DANS LES MODELES, SELON LA TYPOLOGIE D'USAGER.....	45

Résumé

L'objectif de cette étude est de fournir une image de la consommation de l'eau potable sur le bassin Seine-Normandie, en différenciant les usages principaux : la **consommation domestique, des commerces, du secteur tertiaire, des collectivités, des industries, et de l'agriculture**. Pour cela, il a été choisi une méthode basée sur la construction de modèles de régressions linéaires, permettant l'explication de la consommation d'eau potable par des caractéristiques communales (type d'habitats, équipements...). L'élaboration des modèles a été possible grâce à la richesse des données disponibles, socio-économiques et d'eau potable pour l'année 2022, avec la contribution des principaux fournisseurs sur le bassin : Veolia, SUEZ, SENEQ, SEDIF, VEDIF, Eau de Paris et la Communauté d'Agglomération de Reims. Une fois que les modèles les plus pertinents ont été construits puis choisis, des projections sur les communes du bassin pour lesquelles il n'y avait pas de données d'eau potable ont pu être réalisées.

A l'échelle du bassin, **la consommation d'eau potable moyenne estimée par les modèles est largement dominée par la consommation liée aux ménages, avec une proportion d'environ 74,1 %**, finalement plus faible que les 80 % estimés initialement dans une ancienne étude. Ensuite, les **commerces et le secteur tertiaire représentent 10,8 %** de la consommation, suivis de la consommation attribuée aux **établissements publics et liés à la collectivité représentant 9,7 % (3,1 % pour la santé, 0,7 % pour les équipements sportifs, 3,6 % pour l'enseignement, 2,2 % pour l'administration publique)**. Le **secteur industriel représente 2,3 %** et le **secteur agricole 0,9 %**. Enfin, **la consommation du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts représente environ 2,2 %**, et est pratiquement exclusivement liée à du nettoyage de bâtiments, sans qu'il ne soit possible de savoir à quel type de bâtiment elle se rapporte (bâtiments publiques, bâtiments du secteur tertiaire, universités...).

Les résultats indiquent également une **consommation uniquement domestique par habitant égale à près de 41 m³ par an, ce qui revient à une consommation journalière d'environ 112 litres**. Ce résultat est plus faible que les moyennes nationales d'environ 150 litres, mais peut s'expliquer par une prédominance d'appartements, potentiellement moins consommateurs d'eau que les maisons, et par un climat plus froid et humide qui réduit les besoins en eau pour des usages tels que l'arrosage ou la climatisation, au contraire de territoires soumis à un climat plus sec.

Un autre résultat important de cette étude, qui mériterait d'être approfondi, est lié au tourisme. **Les données fournies par les services d'eau potable montrent que les consommations sont plus élevées de près de 30 % dans les communes de moins de 10 000 habitants qui ont été classées comme touristiques.**

Une première analyse des modèles montre qu'ils expliquent entre 82 % et 97 % de la consommation, ce qui est très satisfaisant. Les pourcentages d'erreur, calculés après avoir appliqué les modèles aux communes pour lesquelles des données étaient disponibles, montrent des résultats convaincants pour les communes fortement peuplées. Cependant, ils ont rencontré plus de difficultés dans les communes moins peuplées. Ces dernières présentent en effet des caractéristiques très variées, ce qui rend la construction d'un modèle unique capable de capter les variations de la consommation complexe. Localement, cela se traduit

donc par des marges d'erreur pouvant être assez élevées dans ces communes, mais pour des volumes relativement faibles.

Après analyse des coefficients obtenus et comparaison aux données fournies, il est conclu que nos modèles et les projections sont cohérents et fournissent des résultats satisfaisants. Par exemple, en tenant compte des rendements des réseaux, **les modèles prédisent des prélèvements totaux pour la production d'eau potable sur le bassin environ 8 % inférieurs aux chiffres réels de 2021. En comparant également les résultats avec les chiffres globaux fournis par les distributeurs d'eau potable, à la fois au total et par typologie d'usager, il a été également constaté des marges d'erreur assez faibles, ne dépassant pas 2 % avec le VEDIF, par exemple. Le volume total consommé par la ville de Paris est également prédit avec une précision de 0,4 %, par rapport aux données fournies par Eau de Paris.**

Cependant, il semble qu'il y ait une surestimation de la consommation dans les secteurs de la santé et de l'enseignement, ainsi qu'une sous-estimation, importante pour les équipements sportifs, et l'industrie. Les segments des ménages, des commerces et du secteur tertiaire, qui représentent les plus gros volumes de consommation, semblent quant à eux bien prédits. **En prenant en compte ces observations, une hypothèse qui réduit les consommations des secteurs surestimés (santé et enseignement) et augmente celles des secteurs sous-estimés (équipements sportifs et industries) et des ménages, est proposée.** Cette hypothèse fournit des résultats finalement un peu plus proches des chiffres de prélèvement total sur le bassin, par rapport aux projections ajustées.

Malgré ces bons résultats, il est important de souligner qu'en tant que méthode exploratoire, elle présente naturellement plusieurs limites, notamment en raison de la variabilité des types de communes et des biais dans les données. De plus, disposer de plusieurs années de données aurait permis de réaliser une analyse rétrospective afin de vérifier la stabilité de nos modèles dans le temps. Certaines variables, climatiques pour tenir compte des variations météorologiques par exemple, auraient également mérité d'être intégrées à l'étude. En prenant en considération ces observations, cette méthode pourrait servir de base pour le développement de méthodes similaires sur d'autres territoires.

Cette étude permet donc de fournir une idée des principaux postes de consommation d'eau potable sur le bassin, et d'identifier les secteurs prioritaires sur lesquels agir et les leviers mobilisables dans le cadre du Plan Eau. Il convient maintenant de se rapprocher des structures locales pour valider ces résultats. En effet, un objectif de réduction de 14 % des prélèvements pour la production d'eau potable a été adopté par le Comité de bassin, ce qui représente près de 192 millions de m³ sur les 1,37 milliard prélevés en 2021. L'un des premiers leviers mobilisables, souvent perçu comme une priorité, est l'amélioration du rendement des réseaux, en moyenne plus faible sur les communes faiblement peuplées (Tableau 16). Un objectif visant à atteindre un rendement moyen de 80 % pour ces communes permettrait une économie de seulement un peu plus de 8 millions de m³, soit une réduction de 0,6 %, bien loin des objectifs du Plan Eau et à des coûts potentiellement élevés pour les collectivités et l'Agence. L'importance relative de ce type de mesure souligne l'intérêt d'obtenir une meilleure connaissance des consommateurs d'eau potable, afin de cibler plus précisément les stratégies et trajectoires de sobriété à mettre en place en identifiant les plus gros consommateurs. Les ménages et les secteurs des commerces et du tertiaire, étant les postes de consommation les plus importants, des actions de sensibilisation et d'incitations à l'installation d'équipements hydro-économiques pourraient être un levier pour réduire les consommations, de la même manière que la mise en place d'une tarification progressive de l'eau. Les postes de

consommation directement liés aux collectivités représentent aussi une part importante et un accompagnement auprès de celles-ci sera donc également nécessaire.

Contexte de l'étude

En France, la gestion de l'eau suit deux principes fondamentaux : d'une part, la notion de récupération des coûts implique que les usagers des services publics d'eau et d'assainissement doivent payer pour les coûts générés par ces services, et d'autre part le principe « pollueur payeur » implique que « les frais résultants des mesures de prévention, de réduction de la pollution et de lutte contre celle-ci doivent être supportés par le pollueur ». Afin que ces principes soient applicables, il est nécessaire de pouvoir suivre les prélèvements et pollutions des différents usagers afin de pouvoir ensuite leur faire payer le coût correspondant.

Pourtant, les agences de l'eau n'ont pas une vision précise des types d'usager qui se cachent derrière les prélèvements pour la production d'eau potable. La redevance étant payée par le service d'eau et d'assainissement, l'information s'arrête à ce niveau. Les services d'eau et d'assainissement n'ont d'ailleurs que rarement des informations plus précises sur le type d'usagers que sont leurs clients.

La question de la structuration de ces prélèvements pour l'eau potable revient régulièrement. Elle peut par exemple éclairer les taux des redevances dans la mesure où des organismes professionnels se retrouvent à payer des redevances prélèvements au même taux que les ménages. Elle est également essentielle lors du rapportage à la Commission Européenne du bon respect du principe de récupération des coûts, afin d'identifier quelle partie est payée par les ménages et laquelle ne l'est pas. Récemment, le Plan Eau a aussi redonné de l'importance à cette question. En effet, ce plan prévoit une réduction des prélèvements de 10 % entre 2019 et 2030. Sur le bassin, l'objectif de réduction décliné par le comité de bassin, se traduit par un objectif de 14 % pour l'eau potable. Afin que l'Agence puisse accompagner au mieux les acteurs du territoire, il est d'abord nécessaire d'identifier les usagers finaux, en particulier les principaux types de consommateurs de ces prélèvements, ainsi que leur localisation.

Jusqu'à-là, l'Agence de l'eau se basait sur des enquêtes de consommateurs très localisées, des études redevances datant d'il y a plus de 10 ans ou bien sur un calcul par ordre de grandeur en multipliant la consommation d'un ménage type par le nombre de ménages du bassin. Ces méthodes permettent d'obtenir un ordre de grandeur des prélèvements liés aux ménages, de l'ordre de 80 % sur le bassin Seine-Normandie. Néanmoins, elles présentent quelques limites :

- Des enquêtes trop localisées qui rendent difficiles la généralisation ;
- Aucune de ces méthodes ne permet de préciser la partie hors ménages, différencier l'administration publique des bureaux privés du tertiaire par exemple ;
- Les résultats sont difficiles à spatialiser.

La présente étude cherche à répondre à ces différentes limites. Elle surpasse l'échelle locale en prenant appui sur un large échantillon de données réparties sur tout le bassin. De plus, ses résultats permettent de remonter à des usagers plus fins que les études précédentes et peuvent être projetés sur différentes échelles du bassin. Le modèle fournit des résultats

satisfaisants, avec tout de même quelques limites détaillées dans le rapport. **Les résultats et la discussion sont détaillés en premier lieu, avant la partie sur la méthodologie.**

Résultats et discussion

1. Quatre modèles retenus

Quatre modèles sont présentés ici. Ils ont été retenus car ils semblent les plus pertinents, couvrant efficacement les différents segments d'utilisateurs à représenter.

Ces modèles ont été élaborés à l'aide de régressions linéaires, en suivant la formule suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Les données intégrées aux modèles sont des données d'eau potable issues de plusieurs services présents sur le bassin : SUEZ, SENEQ, Veolia, VEDIF, Eau de Paris, Agglomération du Havre et la Communauté urbaine du Grand Reims. Elles représentent la variable à expliquer, ceci à l'aide de déterminants de la consommation qui seront les variables explicatives : type d'habitats, équipements, caractéristiques de la population et des ménages etc.

L'échelle de travail adoptée dans cette étude est celle de la commune. Étant donné la grande diversité de leurs caractéristiques, des échantillons de communes ont été construits en fonction du nombre d'habitants. Ainsi, chaque modèle correspond à un échantillon.

a. Structure des modèles

Les tableaux suivants présentent les modèles pour chaque échantillon, avec les variables utilisées, les coefficients associés et leur significativité¹.

Tableau 1 : modèle explicatif retenu pour les communes de plus de 10 000 habitants

Communes de plus de 10 000 habitants	
Variables explicatives	Coefficients
Constante	-45940,875
Nombre d'emplois dans le secteur industriel	36,877***
Nombre de lits en secteur hospitalier	313,837***
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	115,477***
Nombre d'emplois dans le secteur de l'enseignement	51,222*
Nombre de résidences principales étant des maisons	112,412***
Nombre de résidences principales étant des appartements	92,489***
Nombre d'emplois dans le secteur de l'administration publique	30,238***
Nombre d'établissements du secteur tertiaire marchand	30,608**

¹ La significativité, basée sur le résultat du test de Student, est estimée à l'aide des étoiles dans les tableaux. Plus il y a d'étoiles, plus le signal statistique est fort indiquant que le coefficient est non nul, alors qu'une absence d'étoile indique un résultat non significatif.

Tableau 2 : modèle explicatif retenu pour les communes de 5 000 à 10 000 habitants

Communes de 5 000 à 10 000 habitants	
Variables explicatives	Coefficients
Constante	38314, 24
Nombre d'élèves de l'enseignement supérieur	33,94 ***
Nombre de résidences principales étant des maisons	36,58 *
Nombre de résidences principales étant des appartements	86,41 ***
Nombre de ménages possédant 2 voitures ou plus	87,11 **
Nombre d'emplois dans le secteur tertiaire marchand	21,86 ***
Nombre de bassins de natation	13389,7 *

Tableau 3 : modèle explicatif retenu pour les communes de 1 000 à 5 000 habitants

Communes de 1 000 à 5 000 habitants	
Variables explicatives	Coefficients
Constante	4278, 51
Nombre d'emplois dans le secteur industriel	29,20 ***
Nombre d'hôpitaux ou d'EHPAD	3017,79 ***
Nombre d'emplois dans le secteur de l'enseignement	74,29 *
Nombre de résidences principales étant des appartements	58,48 ***
Nombre de ménages possédant 2 voitures ou plus	157,48 ***
Nombre d'emplois dans le secteur tertiaire marchand	20,91 ***
Nombre d'UGB porcins	17,03 ***
Nombre d'équipements sportifs avec douches et sanitaires	371,11 *
Nombre de bassins de natation	5097,31 *
Surface totale des terrains de grands jeux	1,08E-04 ***

Tableau 4 : modèle explicatif retenu pour les communes de moins de 1 000 habitants

Communes de moins de 1 000 habitants	
Variables explicatives	Coefficients
Constante	1366,57 ***
Nombre de lits en secteur hospitalier	93,68 ***
Nombre d'emplois dans le secteur de la santé et de l'action sociale	69,01 ***
Nombre d'emplois dans le secteur de l'enseignement	129,28 ***
Nombre de résidences principales étant des maisons	34,60 ***
Nombre de résidences secondaires	55,12 ***
Nombre de résidences principales étant des appartements	45,58 **
Nombre de ménages possédant 2 voitures ou plus	93,21 ***
Nombre d'emplois dans le secteur tertiaire marchand	20,32 ***
Nombre d'UGB bovins	2,51 ***
Nombre d'emplois du secteur de l'agriculture et de la sylviculture	21,31 ***

b. Performances

Le tableau suivant présente certaines caractéristiques des modèles permettant d'évaluer leur performance et leur précision. Les coefficients de tous les modèles sont significatifs, et les modèles sont globalement très significatifs d'après le F-test. En présence d'hétéroscédasticité cependant, les tests de significativité ne sont normalement pas complètement fiables, mais elle paraît finalement suffisamment faible pour qu'ils soient interprétables. De plus, les résultats du R² sont particulièrement élevés et indiquent que les indicateurs captent entre 82 % et 97 % de la variabilité de la consommation de l'eau.

En revanche les erreurs moyennes et médianes, calculées à partir des différences entre les valeurs théoriques calculées par commune et les valeurs réelles, sont plutôt élevées pour les communes de moins de 5 000 habitants. Le calcul pour évaluer les différences entre les valeurs obtenues par les modèles et les valeurs réelles provenant des données, se base sur la formule suivante :

$$\left| \frac{\text{Valeur calculée par les modèles} - \text{Valeur observée}}{\text{Valeur observée}} \right| \times 100$$

Tableau 5 : indications sur la performance des modèles

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants
Taille de l'échantillon	1 095	323	56	181
R ² ajusté	0,87	0,88	0,82	0,97
F-test (p-value)	< 2,2204E-16	< 2,2204E-16	< 2,2204E-16	< 2,2204E-16
Erreur moyenne (%)	20	15,2	8,7	10
Erreur médiane (%)	15	12,4	6,2	7,2
Erreur standard résiduelle	4 695	19 266	42 304	212 601
Résidus supprimés	7	4	0	3

L'erreur standard résiduelle affichée dans le tableau est l'erreur moyenne du modèle en valeur absolue. Elle diminue avec la taille des communes. Cependant, elle est à mettre en relation avec les valeurs réelles observées, à l'aide du tableau suivant :

Tableau 6 : estimation de la précision des modèles à l'aide de l'erreur standard résiduelle

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants
Erreur standard résiduelle	4 695	19 266	42 304	212 601
Plage des valeurs réelles	1 918 – 158 724	34 061 – 325 509	206 025 - 693 834	426 785 - 7 131 200
Valeurs moyennes réelles	19 421	98 334	350 849	1 778 662
Proportion de l'erreur par rapport aux valeurs moyennes réelles (%)	24,2	19,6	12,1	12

La comparaison des erreurs résiduelles avec les valeurs réelles confirme que les communes de moins de 5 000 habitants sont moins bien prédites par les modèles. En effet, l'erreur résiduelle pour ces communes, même si elle est plus faible en valeur absolue, représente au moins 19,6 % des valeurs moyennes réelles, ce qui est assez élevés lorsqu'on compare avec les communes de plus de 5 000 habitants.

Plusieurs mécanismes peuvent expliquer cela. Tout d'abord, les choix faits pour la construction des modèles, tels que la suppression des indicateurs avec des coefficients négatifs par exemple, expliquent cette précision parfois plus faible. Les échantillons sont également plus grands par rapport à ceux pour les communes de plus de 5 000 habitants, donc il est plus probable d'avoir des valeurs aberrantes, et une plus grande disparité dans les données peut être observée qui impactera la qualité des modèles. De plus, les consommations d'eau étant plus faibles par rapport aux communes de plus de 5 000 habitants, la présence d'infrastructures importantes ou une concentration d'habitats plus élevées va rapidement conduire à de fortes variations de la consommation et à des valeurs extrêmes. Et enfin, dans les échantillons de communes de moins de 5 000 habitants, nous pouvons notamment mettre en évidence que la majorité des communes sont rurales selon la classification de l'INSEE (à habitat dispersé et très dispersé, ou bourgs ruraux), mais tout de même quelques communes urbaines sont présentes. En effet, ces échantillons sont constitués de 1 276 communes rurales et 153 communes urbaines, soit plus de 10 % de communes urbaines. Les échantillons sont donc assez hétérogènes sur cet aspect au contraire des échantillons de communes de plus de 5 000 habitants, pour lesquels les communes sont pratiquement exclusivement urbaines.

Enfin, un autre facteur, indépendant des modèles, est que les usagers domestiques notamment en milieu rural ont la possibilité de se tourner vers des ressources alternatives, tels que l'eau de pluie ou les forages individuels. Ces volumes ne se retrouveront donc pas dans les données d'eau potable et peuvent fausser les résultats.

En fin de compte, les petites communes ont des spécificités propres qu'il est difficile de capter à l'aide d'un modèle unique.

2. Influence du tourisme

Comme expliqué dans la partie présentée dans la méthodologie, portant sur la construction de nos échantillons, les communes touristiques sont celles labellisées « Station classée de tourisme » ou « Commune touristique » par l'INSEE, ou bien celles avec plus de 10 places pour 100 habitants dans au moins l'une des catégories suivantes : places en hôtels, places de camping, places en autres hébergements collectifs de tourisme. Cette typologie n'est donc pas exhaustive, Paris n'est par exemple pas considérée comme touristique. En effet, le nombre de places pour 100 habitants est dans ce cas amoindri par la forte population de Paris, d'un peu plus de deux millions d'habitants, et beaucoup de personnes venant à Paris passent par la plateforme de locations de logement Airbnb. Or ces logements n'apparaissent pas dans les données communales mobilisées. Il peut cependant être supposé que le tourisme n'aura que peu d'influence sur la consommation, étant donné que de nombreux parisiens quittent la ville en période estivale, ce qui pourrait maintenir une population assez stable.

Les 298 communes classées comme touristique sont visualisées sur la figure suivante.

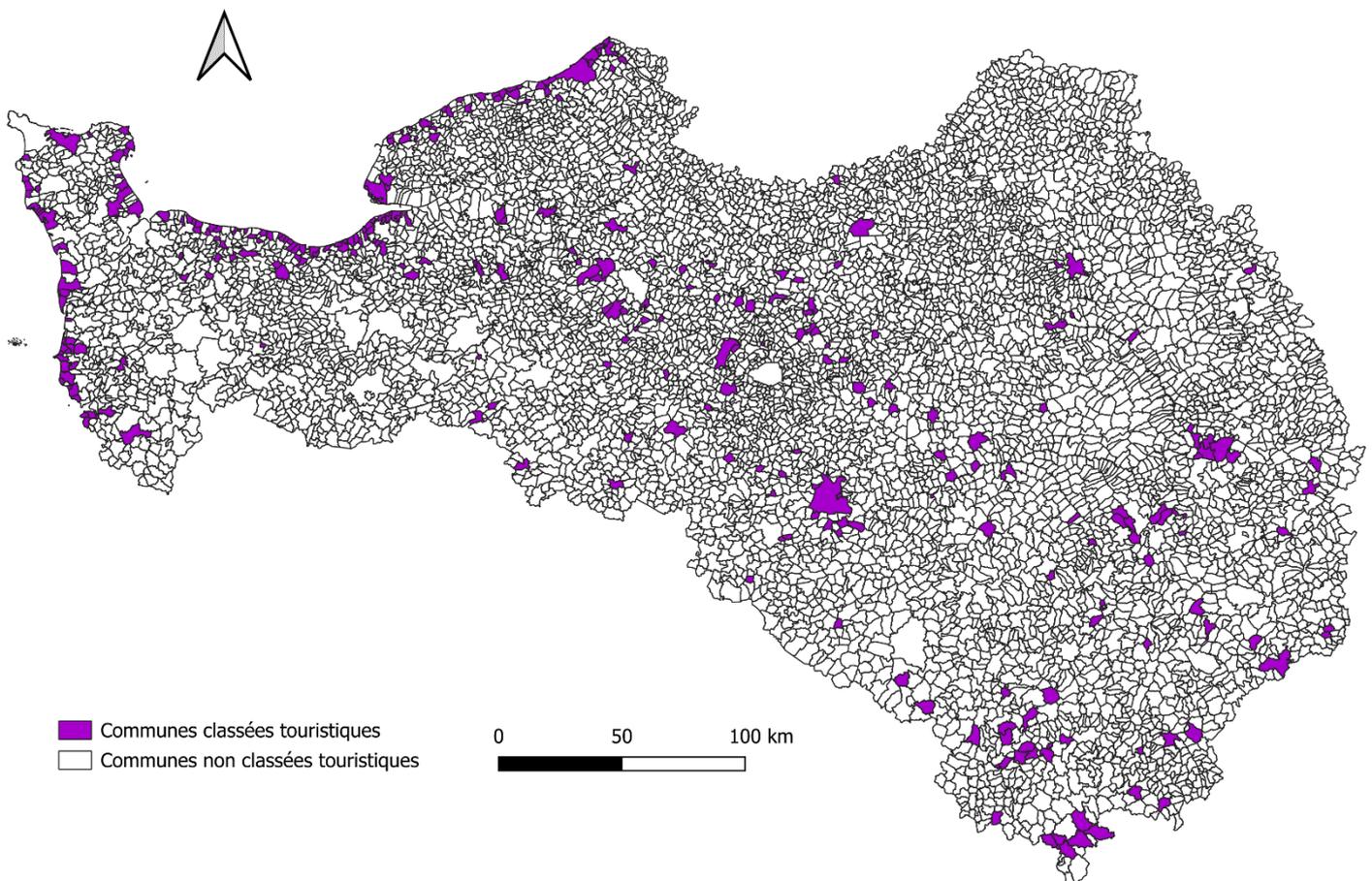


Figure 1 : communes considérées comme touristiques sur le bassin

Les communes du littoral, Reims, Versailles, Fontainebleau, ainsi que des communes situées au début du Massif du Morvan sur le Parc naturel sont mises en évidence.

a. Analyse des données observées

Des données de consommation d'eau potable ont été recueillies pour 82 communes classées touristiques. Ces données ont permis de faire une première comparaison entre les communes touristiques et celles non touristiques. L'objectif est de voir si les communes considérées comme touristiques ont tendance à avoir une consommation totale annuelle plus élevée que les communes non touristiques.

Les communes sont classées selon la typologie utilisée pour construire les modèles. Le tableau suivant présente les consommations moyennes totales et les consommations moyennes totales par habitant, pour les communes touristiques et non touristiques par échantillon.

Tableau 7 : analyse des données fournies de consommation d'eau totale et par habitant, entre communes touristiques et non touristiques

	Moins de 1 000 habitants		1 000 – 5 000 habitants		5 000 – 10 000 habitants		Plus de 10 000 habitants	
	Oui	Non	Oui	Non	Oui	Non	Oui	Non
Commune touristique								
Nombre de communes	40	1102	30	327	3	56	9	184
Volume consommé total moyen (Millions de m ³ /an)	0,03	0,02	0,17	0,1	0,47	0,35	1,98	1,82
Ecart (%)	+ 33,7		+ 40		+ 26,1		+ 8,1	
Volume consommé total moyen par habitant (m ³ /an)	74,2	53,3	76,8	49,1	74	50	57,5	51,7
Ecart (%)	+ 28,2		+ 36		+ 32,4		+ 10,2	

Les résultats montrent que sur les données fournies, la consommation moyenne et la consommation par habitant moyenne sont systématiquement plus élevées chez les communes classées comme touristiques. Le fait que la consommation moyenne par habitant soit également plus élevée confirme bien que la consommation moyenne totale n'est pas expliquée uniquement par un nombre d'habitants plus élevé. L'échantillon de communes touristiques pour les communes de plus de 5 000 habitants est petit, donc la question de la robustesse de la comparaison peut se poser pour ce cas-là.

Les figures suivantes présentent la distribution des consommations totales et par habitant pour les communes de 1 000 à 5 000 habitants. A noter que les points rouges indiquent la moyenne. Elles confirment bien que les communes touristiques ont des consommations totales et des consommations par habitant plus élevées, par rapport aux communes non touristiques. Elles montrent également une variabilité plus forte. Les graphiques pour les autres typologies de commune sont présentés en Annexe VI : boîte à moustache des consommations totales et totales par habitant observées sur les communes touristiques et non touristiques.

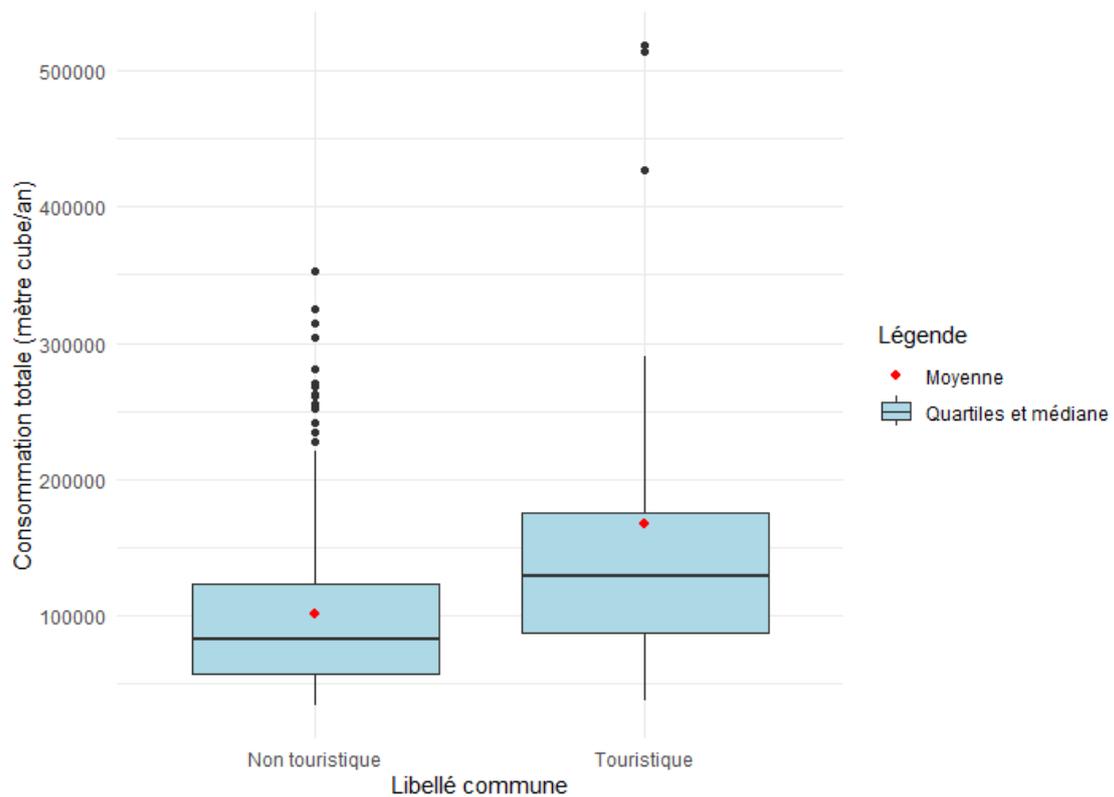


Figure 2 : boîte à moustache de la consommation totale observée sur les communes touristiques et non touristiques, pour la typologie de communes de 1 000 à 5 000 habitants

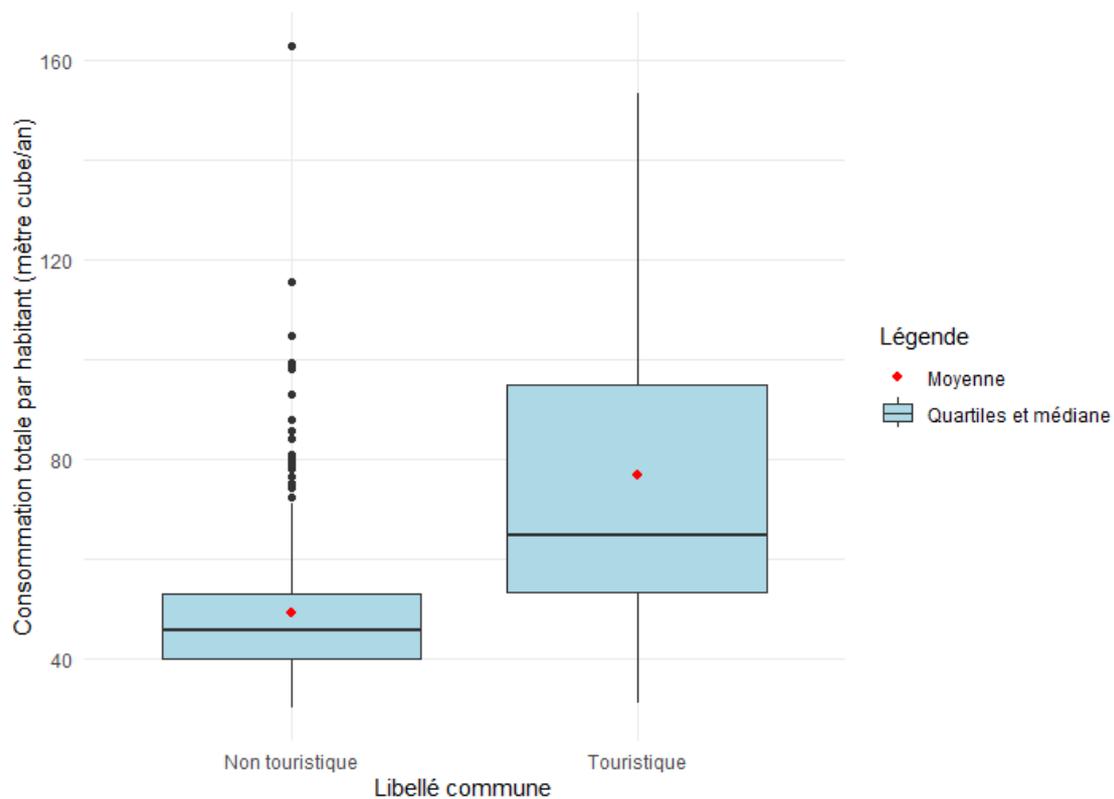


Figure 3 : boîte à moustache de la consommation totale par habitant observée sur les communes touristiques et non touristiques, pour la typologie de communes de 1 000 à 5 000 habitants

Ces diagrammes et le tableau précédent suggèrent que le tourisme a un impact significatif sur la consommation d'eau, augmentant à la fois la consommation moyenne et la variabilité de cette consommation.

b. Projections sur les communes touristiques

Les projections peuvent être appliqués sur les communes touristiques, puis l'écart dû au tourisme mesuré précédemment est rajouté hormis pour les communes touristiques de plus de 10 000 habitants. En effet, les modèles ont tendance à déjà surestimer la consommation sur ces communes. Par conséquent, il ne semble pas cohérent d'ajouter les 8 % de surplus mesuré après analyse des données pour cet échantillon. En revanche, sur les autres échantillons où les modèles sous-estiment bien la consommation des communes touristiques, il paraît plus justifié d'ajouter ce surplus, attribué à l'influence du tourisme. Les résultats sont présentés dans le tableau suivant :

Tableau 8 : consommations totales des communes touristiques

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants
Volume consommé total prédit (Millions de m ³ /an)	4,6	10,5	3,5	71,5
Ecart mesuré (%)	+ 33,7	+ 40	+ 26,1	
Volume consommé total prédit (Millions de m ³ /an), avec surplus du tourisme	6,5	14,7	4,4	71,5

L'étude de l'influence du tourisme pourrait faire l'objet d'une étude plus approfondie. D'une part en mobilisant des données portant sur la fréquentation touristique (INSEE), une nouvelle méthode pour considérer qu'une commune est touristique ou non pourrait être définie. Les communes classées comme telles sont en effet principalement des communes présentant une forte capacité d'accueil, sans témoigner du fait qu'elles accueillent réellement une forte population touristique. Ensuite, il serait également possible à l'avenir de mobiliser des approches plus poussées pour expliquer la consommation d'eau potable liée au tourisme. Par exemple avec des méthodes d'appariement, afin de regrouper des communes aux caractéristiques similaires (population, nombre d'infrastructures...) et de comparer, parmi ces communes, la consommation d'eau des communes touristiques et non touristiques.

La construction de modèles uniquement sur les communes touristiques a été testée, en intégrant des variables sur le nombre d'hébergements touristiques ou le nombre d'emplois du secteur de l'hébergement touristique. Cependant, les régressions linéaires fonctionnent mal pour des échantillons aussi petits, donc nous ne sommes pas allés plus loin pour le moment, en attendant de nouvelles données. Une autre possibilité serait d'abaisser le seuil de 10 places en hébergement touristique pour 100 habitants qui a été utilisé pour la classification, afin d'intégrer plus de communes.

3. Résultats des projections sur toutes les communes

a. Consommation projetée par échantillon

Les tableaux suivants présentent les résultats des projections ajustées pour les quatre échantillons, en prenant en compte les communes touristiques. La projection ajustée est calculée simplement en multipliant les coefficients, présentés précédemment, par les données communales.

Il est nécessaire de préciser que la constante semble importante pour l'ajustement des modèles. Comme nous ne savons pas comment elle se répartit entre les segments d'utilisateur, elle n'a pas été prise en compte pour le calcul des proportions et la consommation projetée par segment. En revanche, le total la prend en compte, bien que la constante soit significative uniquement dans le dernier modèle.

De la même manière, le volume d'eau consommé considéré comme le surplus touristique est présenté à part, hormis pour le modèle sur les communes de plus de 10 000 habitants, pour les raisons expliquées dans la section précédente. En effet, il n'est pas possible de savoir comment ce volume se répartit parmi les types d'utilisateurs, même s'il est probable que ça soit principalement une consommation domestique.

En ce qui concerne la question du secteur du nettoyage de bâtiment et d'entretien des espaces verts, il est conclu, après des tests, que la consommation expliquée par ce segment d'utilisateur correspond exclusivement au nettoyage de bâtiments. Cependant, il n'est pas non plus possible de savoir s'il est lié à du nettoyage d'établissements publics ou à des bâtiments tertiaires, donc ce segment est laissé à part dans la suite de l'étude.

Tableau 9 : résultat des projections pour les communes de plus de 10 000 habitants

Communes de plus de 10 000 habitants			
Variables explicatives	Segment d'utilisateur	Part	Consommation projetée ajustée (Millions de m ³ /an)
<i>Constante</i>			-15,8
Nombre de résidences principales étant des maisons	Ménages	72,5	532,6
Nombre de résidences principales étant des appartements			
Nombre d'établissements du secteur tertiaire marchand	Commerces et tertiaire	11,2	82,1
Nombre d'emplois dans le secteur de l'enseignement	Enseignement	4,1	30,2
Nombre de lits en secteur hospitalier	Santé	3,5	25,8
Nombre d'établissements de l'administration publique	Administration publique	3,2	23,4
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	Nettoyage de bâtiment	3,1	22,8
Nombre d'emplois dans le secteur industriel	Industries	2,4	17,9
Total			719,2

La consommation pour cet échantillon de commune est comme attendu, dominée par la consommation des ménages, à hauteur de 72,5 %, plus faible que le chiffre de 80 % initialement avancé. Mais cette proportion relativement plus faible s'explique par la présence importante d'établissement publics et liés à la collectivité (administration publique, santé,

enseignement) dans les grands centres urbains, qui caractérisent presque exclusivement cet échantillon, et qui représente ici près de 11 % de la consommation. Ensuite, la forte présence de commerces et d'établissements du secteur tertiaire dans les grands centres urbains explique également la part importante de ce secteur dans la consommation. Aucune consommation liée aux équipements sportifs n'est modélisée pour ces communes, du fait de l'absence d'indicateur associé dans ce modèle. Cela met en évidence l'une des limites des modèles car ils n'expliquent pas toutes les consommations, ce qui se répercute sur les projections.

Tableau 10 : résultat des projections pour les communes de 5 000 à 10 000 habitants

Communes de 5 000 à 10 000 habitants			
Variables explicatives	Segment d'utilisateur	Part	Consommation projetée ajustée (Millions de m ³ /an)
<i>Constante</i>			7,3
<i>Surplus tourisme</i>			0,9
Nombre de résidences principales étant des maisons	Ménages	82,6	50
Nombre de résidences principales étant des appartements			
Nombre de ménages possédant 2 voitures ou plus			
Nombre d'emplois dans le secteur tertiaire marchand	Commerces et tertiaire	13,3	8
Nombre de bassins de natation	Equipements sportifs	3,4	2
Nombre d'élèves de l'enseignement supérieur	Enseignement	0,7	0,4
Total			68,7

La proportion de la consommation expliquée par la consommation des ménages est plus forte dans cet échantillon. Une des raisons pourrait être que ces communes correspondent à des communes urbaines périphériques de grandes villes (en Ile-de-France, Caen, Rouen, Le Havre), et donc potentiellement majoritairement résidentielles du fait de loyers plus faibles par exemple, comme le montre la Figure 8 présentées dans la méthodologie. En effet, sur les 191 communes de l'échantillon, 94 sont des communes de densité intermédiaires (petites villes, centres urbains intermédiaires et ceintures urbaines), dont 36 sont par exemple considérées comme étant des ceintures urbaines selon la classification de l'INSEE.

Ce modèle, à l'inverse de ce que l'on pourrait attendre, ne montre rien sur la consommation des établissements de santé, de l'administration publique et semble sous-estimer la consommation liée à l'enseignement du fait qu'elle ne montre que celle des élèves du supérieur.

Tableau 11 : résultat des projections pour les communes de 1 000 à 5 000 habitants

Communes de 1 000 à 5 000 habitants			
Variables explicatives	Segment d'utilisateur	Part	Consommation projetée ajustée (Millions de m³/an)
<i>Constante</i>			5,9
<i>Surplus tourisme</i>			4,2
Nombre de résidences principales étant des appartements	Ménages	73,1	100
Nombre de ménages possédant 2 voitures ou plus			
Nombre d'emplois dans le secteur tertiaire marchand	Commerces et tertiaire	10,7	14,7
Nombre d'emplois dans le secteur industriel	Industries	4	5,5
Nombre d'équipements sportifs avec douches et sanitaires	Equipements sportifs	3,5	4,8
Nombre de bassins de natation			
Surface totale des terrains de grands jeux			
Nombre d'hôpitaux ou d'EHPAD	Santé	3,5	4,7
Nombre d'emplois dans le secteur de l'enseignement	Enseignement	3,3	4,5
Nombre d'UGB porcins	Agriculture - élevage	1,8	2,5
Total			146,9

Les proportions de consommation sont assez similaires pour cet échantillon à celles pour l'échantillon de communes de plus de 10 000 habitants. Cet échantillon est caractérisé par une forte présence de bourgs ruraux, au nombre de 644 sur les 1379 communes, c'est-à-dire des communes rurales plus denses que les communes à habitat rural dispersé et très dispersé. Un bourg est considéré comme une petite agglomération, et accueille donc de nombreux services de proximité. L'échantillon compte également de nombreuses communes urbaines (44 grands centres urbains, 71 petites villes, 316 ceintures urbaines). Cela peut donc expliquer les similarités entre cet échantillon et celui des communes de plus de 10 000 habitants, malgré des consommations projetées bien plus élevées dans ce dernier. La proportion de la consommation des industries est assez élevée, égale à 4 %, ce qui peut s'expliquer par la présence d'industries agro-alimentaires dans les communes rurales.

Tableau 12 : résultat des projections pour les communes de moins de 1 000 habitants

Communes de moins de 1 000 habitants			
Variables explicatives	Segment d'utilisateur	Part	Consommation projetée ajustée (Millions de m ³ /an)
<i>Constante</i>			8,3
<i>Surplus tourisme</i>			1,6
Nombre de résidences principales étant des maisons	Ménages	82,6	81,3
Nombre de résidences secondaires			
Nombre de résidences principales étant des appartements			
Nombre de ménages possédant 2 voitures ou plus			
Nombre d'UGB bovins	Agriculture - élevage	7	6,8
Nombre d'employé du secteur de l'agriculture et de la sylviculture			
Nombre d'emplois dans le secteur tertiaire marchand	Commerces et tertiaire	6,3	6,2
Nombre d'emplois dans le secteur de l'enseignement	Enseignement	2,6	2,6
Nombre de lits en secteur hospitalier	Santé	1,4	1,4
Nombre d'emplois dans le secteur de la santé et de l'action sociale			
Total			108,3

La consommation pour les communes de moins de 1 000 habitants est logiquement expliquée principalement par la consommation des ménages. Le secteur de l'agriculture et de l'élevage explique également une proportion importante de la consommation, plus encore que le secteur des commerces et du tertiaire.

Il est également intéressant de regarder les consommations et les proportions calculées pour la ville de Paris, montrées par le tableau suivant. Il sera montré par la suite que la consommation totale à Paris est très bien prédite par le modèle sur les communes de plus de 10 000 habitants.

Tableau 13 : résultat des projections pour Paris intra-muros

Paris			
Variables explicatives	Segment d'utilisateur	Part	Consommation projetée ajustée (Millions de m ³ /an)
<i>Constante</i>			-0,05
Nombre de résidences principales étant des maisons	Ménages	64,3	102,4
Nombre de résidences principales étant des appartements			
Nombre d'établissements du secteur tertiaire marchand	Commerces et tertiaire	19,9	31,7
Nombre d'emplois dans le secteur de l'enseignement	Enseignement	5,4	8,6
Nombre d'établissements de l'administration publique	Administration publique	3,2	5,1
Nombre de lits en secteur hospitalier	Santé	3,2	5,1
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	Nettoyage de bâtiment	2,5	3,9
Nombre d'emplois dans le secteur industriel	Industries	1,5	2,4
Total			159,1

Ce qui est surtout remarquable à Paris est la forte proportion de la consommation liées aux commerces et au secteur tertiaire, et la proportion plus faible liée aux ménages. La consommation des commerces et du secteur tertiaire à Paris représente pratiquement la moitié de toute la consommation du secteur prédite sur l'échantillon de communes de plus de 10 000 habitants. C'est assez cohérent car il y a une forte densité de ce type d'établissements à Paris, et de nombreux commerces sont situés en bas d'immeubles d'habitation, avec un compteur commun pour tout l'immeuble. Une partie de la consommation liée aux ménages peut donc être comptabilisée dans celle liée aux commerces.

Ces résultats peuvent ensuite être mis en relation avec le nombre d'habitants. Les résultats sont fournis dans le tableau suivant. La consommation domestique par habitant est calculée en divisant la consommation projetée du segment des ménages par le nombre d'habitants.

Tableau 14 : consommations projetées totale et domestique, rapportées au nombre d'habitants

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants	Total
Population	2 074 715	2 843 780	1 350 745	12 513 835	18 783 075
Consommation domestique / habitant (m ³ /an)	39,2	35,2	37	42,6	40,7
Consommation totale / habitant (m ³ /an)	52,2	51,7	52,2	57,5	56

Au total sur le bassin la **consommation domestique obtenue par nos projections par an et par habitant est égale environ à 40,7 m³, soit environ 111 litres/jour/habitant**. Elle est assez faible si l'on se réfère au chiffre moyen revenant régulièrement à l'échelle de la France d'environ 150 litres/jour/habitant, qui équivaut à 55 m³ par an. En revanche cette moyenne à l'échelle de la France n'est pas homogène sur l'ensemble du territoire, avec de fortes disparités (R.Barbier & M.Montginoul, 2013)². Par exemple, une étude de l'INSEE portant sur le périmètre du SEDIF indique que la consommation domestique y est de 36 m³/an/habitant, soit près de 99 litres/jour/habitant, soit un chiffre cohérent par rapport aux résultats présentés ici. Ce chiffre plus faible est notamment expliqué dans l'étude par une proportion plus importante d'appartement sur le périmètre du SEDIF, égale à 74 % par rapport à la moyenne française de 43 %, du fait d'une consommation moindre de ces habitats (INSEE, 2023)³. Sur le périmètre du bassin, la proportion d'appartements est d'environ 58 %, ce qui la situe entre la moyenne française et celle du SEDIF. Ainsi, la consommation domestique par habitant calculée, située entre ces deux valeurs, est cohérente. Une autre explication à ces disparités dans la consommation d'eau domestique par habitant serait liée au climat. Les régions du Sud pourraient avoir une moyenne de consommation plus élevée du fait de précipitations plus faibles, ce qui nécessite davantage d'eau pour l'arrosage des jardins, pour la consommation ou pour la climatisation.

² Barbier & Montginoul, « Avant-propos : le consommateur d'eau en question », 2013.

³ INSEE, « La consommation d'eau potable illustrée par la situation dans 133 communes franciliennes : 100 litres en moyenne par jour et par habitant », 2023.

La consommation domestique par habitant est plus élevée sur les communes de moins de 1 000 habitants, par rapport aux communes de 1 000 à 10 000 habitants. Cela peut s'expliquer par la présence de l'indicateur « Nombre de résidences secondaires » dans ce modèle, qui indique de fortes proportions de résidences secondaires sur ces communes. La présence des résidences secondaires va, de la même manière que le tourisme, augmenter la consommation d'eau par habitant, notamment en raison d'une occupation saisonnière que ne reflète pas le nombre d'habitants permanents.

En ce qui concerne la **consommation d'eau totale par habitant, prenant en compte tous les autres usages, un résultat de 56 m³/an/habitant est obtenu, soit 153 litres/habitant/jour**. L'étude du Crédoc propose un chiffre similaire en Ile-de-France, avec une majorité de valeurs par commune comprises entre 30 et 70 m³, et une valeur de 55 m³ qui est la plus fréquente (Crédoc, 2016).

Il est également à noter que les consommations par habitant augmentent en fonction du nombre d'habitants par commune, hormis pour les communes de moins de 1 000 habitants à cause des résidences secondaires. C'est logique pour la consommation totale par habitant, car les communes plus peuplées proposent plus d'équipements ou d'infrastructures ayant de fortes consommations d'eau, et sont le lieu d'activités économiques plus nombreuses. En revanche, les consommations uniquement domestiques par habitant sont relativement proches pour les communes de moins de 10 000 habitants, mais celles de plus de 10 000 habitants ont un résultat plus élevé, ce qui est plus surprenant. Cela peut être dû en premier lieu à un effet du tourisme que les modèles et l'analyse sur les communes touristiques n'ont pas réussi à mettre en évidence. D'autre part, il est aussi possible de l'expliquer par le fait que les communes fortement peuplées ont des niveaux de vie plus élevés, et comme évoqué précédemment, un niveau de vie plus élevé impacte positivement la consommation. Dans les villes notamment, l'accès à l'eau potable peut également paraître plus sécurisé et constant par rapport aux zones rurales, et les ménages peuvent donc être moins sensibles à la question de préservation de la ressource en eau.

b. Consommation projetée totale

Les résultats des projections à l'échelle du bassin versant sont présentés dans le tableau suivant. Les estimations hautes et basses de consommation et les proportions hautes et basses, pour chaque segment d'utilisateur, sont également présentées. Il n'est pas possible de calculer de borne pour le surplus touristique, donc le même volume est considéré pour le calcul de la consommation totale pour chaque borne.

Les bornes basses et hautes sont estimées à l'aide de la fonction « predict » sur Rstudio. En effet, elle permet de calculer des intervalles pour la valeur de Y , représentant la consommation d'eau totale, par commune. La fonction ne permettant pas de décomposer le calcul pour chaque variable explicative, le pourcentage d'écart entre la borne basse et la valeur ajustée, ainsi qu'entre la borne haute et la valeur ajustée de la consommation totale, est calculé, puis appliqué uniformément à chaque segment d'utilisateur. Il est donc considéré que l'intervalle global calculé reste constant pour chaque variable explicative et, par conséquent, pour chaque segment d'utilisateur.

Tableau 15 : résultat des projections à l'échelle du bassin et par segments d'utilisateur

Bassin Seine-Normandie						
Segment d'utilisateur	Consommation projetée (Millions de m ³ /an)			Part (%)		
	Basse	Ajustée	Haute	Basse	Ajustée	Haute
Ménages	618,8	763,9	909,1	66,2	74,1	80,7
Commerces et tertiaire	90,6	111,1	131,6	7,7	10,8	15
Etablissements publics	82,2	100	117,8	6,8	9,7	14
<i>Enseignement</i>	30,9	37,8	44,6	2,5	3,7	5,2
<i>Santé</i>	26,1	32	37,8	2,2	3,1	4,4
<i>Administration publique</i>	20,2	23,4	26,6	1,7	2,3	3,2
<i>Equipements sportifs</i>	4,9	6,9	8,8	0,4	0,7	1
Industries	19,4	23,4	27,4	1,6	2,3	3,2
Services de nettoyage de bâtiments et d'entretien des espaces verts	20,1	22,8	25,5	1,6	2,2	3
Agriculture - élevage	5,4	9,3	13,3	0,4	0,9	1,6
<i>Surplus tourisme</i>	6,8	6,8	6,8			
<i>Constante</i>	0,6	5,7	10,8			
Total	838,7	1 043,1	1 230,7			

La consommation projetée ajustée totale des ménages représente 74,1 % de la consommation totale. Ce qui représente finalement un peu moins que le chiffre initial estimé de 80 % dans l'étude sur les APAD. **Le secteur des commerces et du tertiaire représente 10,8 %.** **Les secteurs de l'enseignement, la santé, l'administration publique et les équipements sportifs peuvent être regroupés en une seule catégorie, représentant les infrastructures publiques et liées à la collectivité. En regroupant ces segments, la proportion est de 9,7 %.**

L'écart entre la borne moyenne et la borne basse est d'environ 24 % et d'environ 18 % entre la borne moyenne et la borne haute, ce qui représente des volumes assez élevés de près de 250 millions de m³.

Néanmoins, ces chiffres peuvent être comparés avec les chiffres des prélèvements pour la production d'eau potable sur le bassin pour l'année 2021. Lorsqu'on corrige les projections de volumes facturés avec le rendement moyen des réseaux présentés dans le tableau suivant, pour obtenir les volumes prélevés prédits par les modèles, le résultat est d'environ 1,26 milliard de m³. Les données de prélèvement sur l'année 2021 indiquent 1,37 milliard de m³ prélevés. Ce qui représente donc une différence relativement faible de 8,5 %, qui permet d'estimer que la réalité est plus proche des projections ajustées que des bornes hautes et basses. En suivant le même raisonnement avec les résultats de la borne haute, le volume prélevé prédit est supérieur de 9,4 % aux données réelles.

Tableau 16 : rendements des réseaux par typologie de commune et résultats des volumes prélevés estimés à l'aide des rendements

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants	Total
Rendement des réseaux (%)	75,6	79,9	83	84,4	
Volume consommé total prédit (Millions de m ³ /an)	108,3	146,9	68,7	719,2	1 043,1
Volume prélevé total prédit (Millions de m ³ /an)	143,3	183,9	82,8	852,1	1 262,1

L'Annexe V : résultat des volumes prélevés prédits pour chaque segment d'utilisateur, présente la répartition des prélèvements prédits par usager (c'est-à-dire en tenant compte des rendements des réseaux présentés dans le tableau précédent). Les résultats sont finalement très similaires.

c. Spatialisation

La carte suivante montre, comme attendu, une consommation totale projetée qui est cohérente avec la carte de la population (Figure 8) présentée dans la méthodologie. L'Île-de-France est la zone avec les consommations les plus élevées, c'est également là qu'il y a le plus d'habitants, ainsi que le Cotentin et les Métropoles de Rouen et du Havre. Les consommations les plus faibles sont situées à l'est du bassin, au niveau des communes les moins peuplées. Les cartes sur les autres segments d'utilisateur sont présentées en Annexe IV : spatialisation des consommations projetées pour chaque segment d'utilisateur.

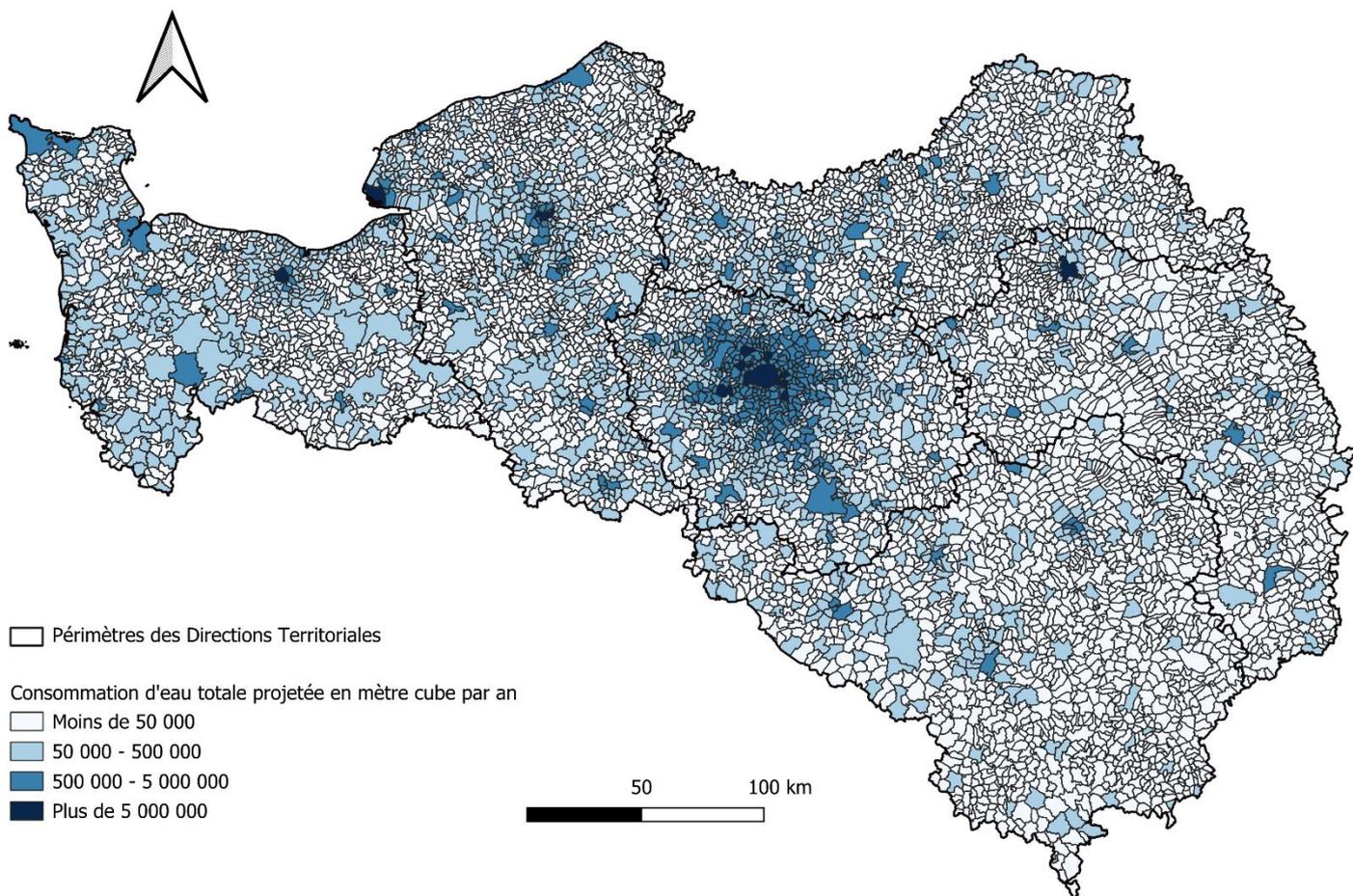


Figure 4 : consommation d'eau totale projetée, par commune et à l'échelle du bassin versant, en prenant en compte les communes touristiques et le surplus associé

La tendance en regardant la consommation totale par habitant est complètement différente (Figure 5). En effet, les communes pour lesquelles les consommations totales par habitant sont les plus élevées sont situées au niveau des zones les moins peuplées, à l'est du bassin particulièrement, et de la Basse-Normandie. Certaines de ces communes sont notamment celles ayant des proportions importantes de résidences secondaires, les communes touristiques du littoral et au sud du bassin (Figure 1) et certaines communes ayant une consommation liée à l'agriculture et l'élevage importante (Annexe IV : spatialisation des consommations projetées pour chaque segment d'utilisateur).

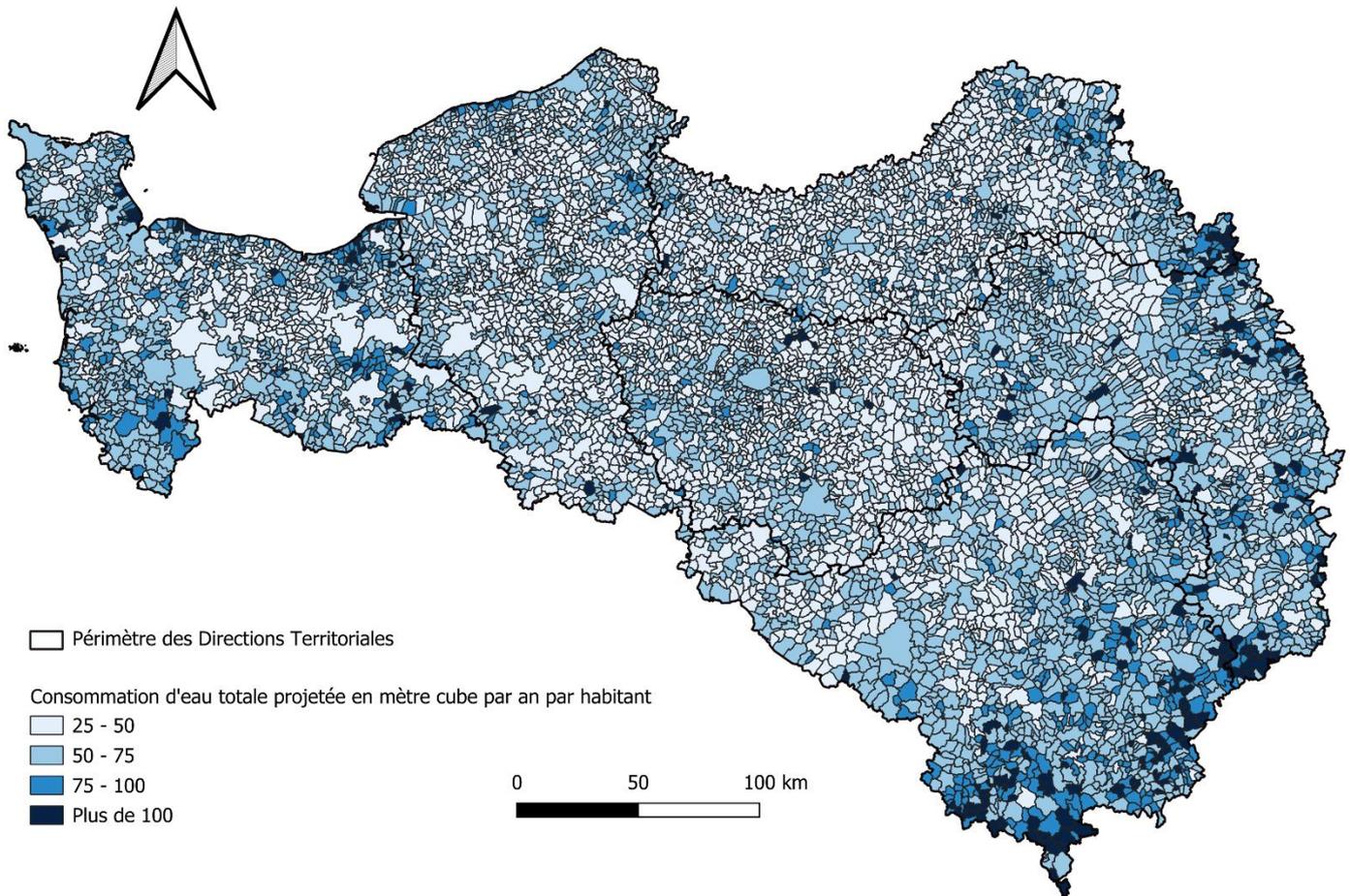


Figure 5 : consommation d'eau totale projetée rapportée au nombre d'habitants, par commune et à l'échelle du bassin versant

4. Interprétation et validité des modèles

Dans cette partie, la validité des coefficients obtenus pour les modèles sera discutée en les comparant à travers les différents modèles et à des valeurs issues de la littérature. Les volumes d'eau pour les indicateurs que les modèles ne montrent pas, par exemple les équipements sportifs, seront également estimés.

La validité des modèles sera ensuite évaluée en comparant les résultats des projections aux données fournies par les services d'eau potable, puis les limites de l'étude seront évoquées.

a. Interprétation et cohérence des coefficients des modèles

L'analyse des coefficients est présentée par grand type d'usager. L'étude principale utilisée pour comparer les valeurs des coefficients provient du Syndicat mixte d'étude et de gestion de la ressource en eau du département de la Gironde (SMEGREG).⁴

Chaque tableau de comparaison entre les modèles comprend une colonne estimant la cohérence entre les quatre modèles choisis. Elle est estimée par un code couleur : vert pour une bonne cohérence, orange pour une cohérence moyenne et rouge pour une mauvaise cohérence. Le numéro dans chaque case correspond au modèle comparé : 1 pour celui sur les communes de plus de 10 000 habitants, 2 pour celles de 5 000 à 10 000 habitants, 3 pour celles de 1 000 à 5 000 habitants et 4 pour celles de moins de 1 000 habitants.

Ménages

Le tableau suivant compare les coefficients entre les modèles pour le segment des ménages. Le bouquet de variables explicatif de la consommation des ménages est différent entre les modèles. Les coefficients associés à chaque variable ne sont donc pas systématiquement identiques, et peuvent être influencés par la présence de l'indicateur « Nombre de ménages possédant 2 voitures ou plus ».

Tableau 17 : coefficients calculés pour le secteur des ménages, estimation de la cohérence et valeurs de la littérature

Echantillon	Variables	Coefficients	Cohérence inter-modèles		Valeurs littéraires
Plus de 10 000 habitants	Nombre de résidences principales étant des maisons	112,412		2	115 m ³ /an par logement et logement de fonction par extension
	Nombre de résidences principales étant des appartements	92,489	3	4	
5 000 à 10 000 habitants	Nombre de résidences principales étant des maisons	36,58	1		
	Nombre de résidences principales étant des appartements	86,41	3	4	
	Nombre de ménages possédant 2 voitures ou plus	87,11			
1 000 à 5 000 habitants	Nombre de résidences principales étant des appartements	58,48	1	2	
	Nombre de ménages possédant 2 voitures ou plus	157,48		4	
Moins de 1 000 habitants	Nombre de résidences principales étant des maisons	34,6	1	2	
	Nombre de résidences principales étant des appartements	45,58			
	Nombre de ménages possédant 2 voitures ou plus	93,21	3		
	Nombre de résidences secondaires	55,12			

⁴ Commission Locale de l'Eau du SAGE Nappes profondes de Gironde, « Liste de ratios de consommation d'eau potable et leur domaine d'emploi pour se repérer, poser un premier diagnostic ou prévoir une consommation », 2007

Cependant, pour estimer la cohérence des modèles entre eux pour cette partie, il peut être observé que les consommations domestiques par habitant, présentées précédemment dans le Tableau 14 sont assez similaires hormis pour les communes de plus de 10 000 habitants. De plus, pour l'échantillon de moins de 1 000 habitants, la consommation domestique est expliquée comme évoquée précédemment, par l'indicateur sur les résidences secondaires qui tire cette consommation vers le haut. La figure suivante, montrant la distribution des résultats par commune, et le résultat global obtenu de 35,9 m³/habitant en enlevant la consommation liée aux résidences secondaires pour les communes de moins de 1 000 habitants, montrent des résultats plus homogènes entre les échantillons qu'en gardant la consommation liée aux résidences secondaires.

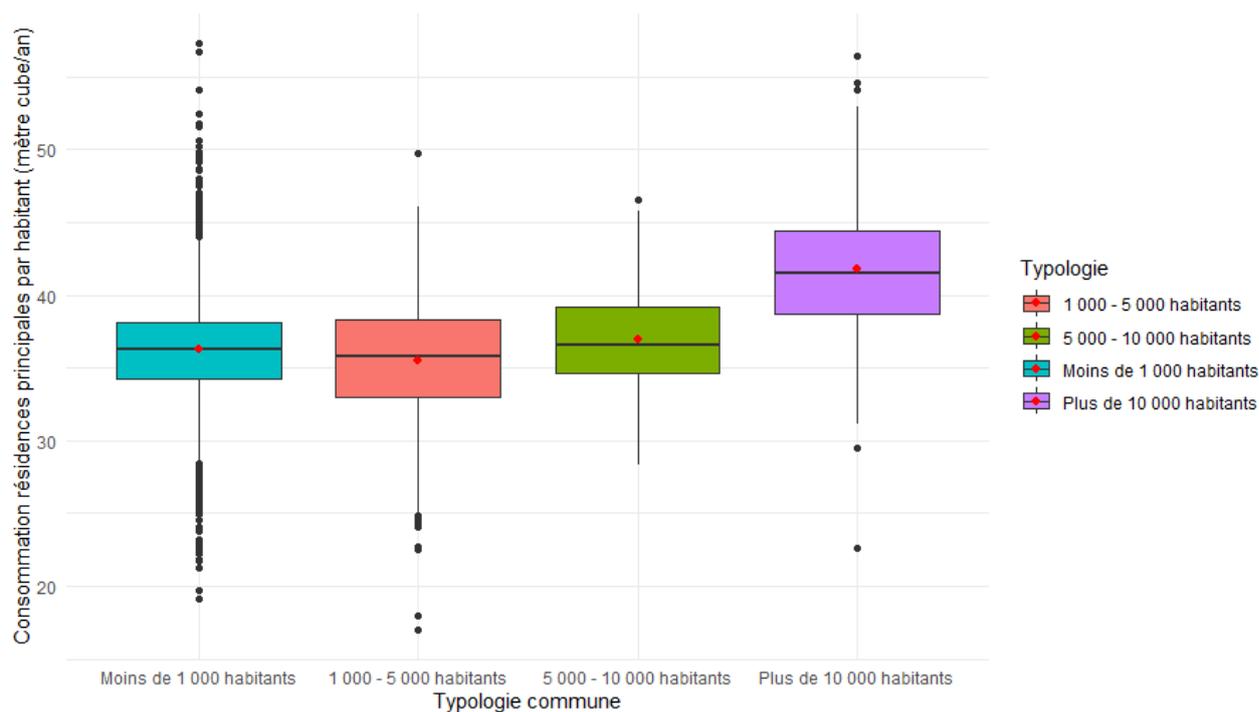


Figure 6 : boîte à moustache des consommations liées aux résidences principales rapportées au nombre d'habitants, par typologie de commune

Dans le premier modèle, où n'apparaît pas la variable « Nombre de ménages possédant 2 voitures ou plus », le coefficient pour l'indicateur « Nombre de résidences principales étant des maisons » égal à 112,412, est plus élevé que pour l'indicateur « Nombre de résidences principales étant des appartements », égal à 92,489. Cela indique que la consommation unitaire des maisons est plus élevée, en cohérence avec ce qui a été évoqué plus haut (plus grandes surfaces, possibles piscines et voitures supplémentaires...). Ce chiffre est également assez cohérent à la fois avec les chiffres d'une consommation par ménage d'environ 120m³ et ceux du SMEGREG, en considérant environ 2,5 personnes par ménage.

Pour comparer plus précisément avec la littérature, il aurait été intéressant d'avoir des données sur le nombre d'habitants en maisons et le nombre d'habitants en appartements afin d'avoir une moyenne d'habitants pour chaque, et d'expliquer grâce à ça la différence entre le coefficient portant sur les maisons et celui portant sur les appartements. En effet, la différence de consommation entre les deux peut aussi s'expliquer par le fait que les ménages vivant en appartements peuvent être en moyenne plus petits que ceux vivant en maisons. Pendant l'étude, il a par exemple été mis en évidence que la variable sur le nombre de résidences

principales étant des appartements était fortement corrélé à l'indicateur sur le nombre de ménage ne comptant qu'une seule personne.

Enfin, la variable sur les résidences secondaires n'apparaît que dans le dernier modèle, et aucune consommation liée n'est donc visible dans les autres. En appliquant le résultat du coefficient obtenu (55,12) sur les autres échantillons, un total d'environ 21,1 millions de m³ ne serait pas expliqué par nos modèles. En valeurs relatives cependant cela reste faible et explique peut-être pourquoi cet indicateur ne ressort pas. Par exemple, pour les communes de plus de 10 000 habitants, la consommation liée aux résidences secondaires ne représenterait seulement que 2,5 % de la consommation domestique. Et ce, si tant est qu'il soit judicieux de transposer tel quel le coefficient trouvé pour les résidences secondaires rurales. Dans les grands centres urbains, les résidences secondaires consomment sans doute bien moins.

Tertiaire marchand

Le tableau suivant montre une bonne cohérence entre les modèles. Pour le modèle sur les communes de plus de 10 000 habitants, la variable « Nombre d'établissements du secteur tertiaire marchand » n'est pas directement comparable avec la variable « Nombre d'emplois dans le secteur tertiaire marchand » utilisée dans les autres modèles. La consommation obtenue pour ce secteur, d'environ 82,1 millions de m³, peut cependant être rapportée sur le nombre d'emplois du secteur tertiaire marchand observé sur ces communes pour obtenir un chiffre de 14,1 m³/employé, plus faible par rapport aux autres modèles. En rajoutant le secteur du service de nettoyage des bâtiments et d'entretien des espaces verts ici, le résultat est de 17,4 m³/employé. Cette attribution est confortée par le fait que la consommation modélisée associée à ces services se concentre dans des communes avec beaucoup de tertiaire, où les emplois associés sont plutôt liés à l'entretien, probablement de bureaux.

Tableau 18 : coefficients calculés pour le secteur tertiaire et des commerces, estimation de la cohérence et valeurs de la littérature

Echantillon	Variables	Coefficients	Cohérence inter-modèles		Valeurs littéraires
Plus de 10 000 habitants	Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	115,477		1	36,5m ³ à 54,75 m ³ /emploi dans l'administration, pour un grand ensemble de bureaux (>30 emplois)
	Nombre d'établissements du secteur tertiaire marchand	30,608	3	4	
5 000 à 10 000 habitants	Nombre d'emplois dans le secteur tertiaire marchand	21,86	1		
			3	4	
1 000 à 5 000 habitants	Nombre d'emplois dans le secteur tertiaire marchand	20,91	1	2	
				4	
Moins de 1 000 habitants	Nombre d'emplois dans le secteur tertiaire marchand	20,32	1	2	
			3		
					10,95 à 18,25 pour un bâtiment de 5 à 30 employés

Ces résultats se situent entre les valeurs de la littérature, ces dernières concernant uniquement les bureaux du secteur tertiaire. Alors que la variable présentée ici correspond au secteur tertiaire, avec d'une part des grands ensembles de bureaux multiservices (cantine, climatisation etc...) et des petits bureaux, mais aussi à tous les types de commerces. Comme indiqué dans l'étude du SMEGREG, la consommation des petits bureaux (5 à 30 employés) peut correspondre à la consommation stricte des sanitaires d'un employé, ce qui peut également correspondre au cas de nombreux commerces. Ces valeurs, plutôt proches de la consommation pour les petits bureaux, sont donc finalement assez logiques.

Administration publique

La variable portant sur l'administration publique n'apparaît que dans le modèle sur les communes de plus de 10 000 habitants, avec un coefficient égal à 30,238. La comparaison peut se faire avec les mêmes valeurs de la littérature présentées dans le tableau précédent. Le résultat obtenu est plus proche de celle pour les grands ensembles de bureaux, ce qui est logique du fait de la plus forte concentration de ce type d'établissements dans les grandes villes. En revanche elle paraît un peu faible, car pour cet échantillon les données indiquent environ 100 emplois dans l'administration par établissement, donc le coefficient aurait pu être plus élevé si l'on se base sur les valeurs de la littérature.

Aucun indicateur sur ce segment n'apparaît dans les autres modèles. A partir des valeurs de la littérature et du nombre d'employés par établissement calculé à partir des données, il est possible d'estimer que les volumes correspondant à ce que les modèles n'expliquent pas sont d'un peu plus de 2 millions de m³ environ, soit un résultat assez faible.

Enseignement

Le tableau suivant montre une bonne homogénéité entre les modèles, hormis avec l'échantillon des communes de 5 000 à 10 000 habitants qui ne montre que la consommation liée à l'enseignement supérieur, mais rien concernant l'enseignement du primaire et du secondaire. Les coefficients varient beaucoup entre l'échantillon de plus de 10 000 habitants et celui pour les communes de moins de 1 000 habitants, avec un rapport d'environ 2,5 entre les deux. Mais en rapportant la consommation modélisée totale liée à l'enseignement, au nombre d'élèves du primaire, du secondaire et de l'enseignement supérieur, les résultats obtenus sont plus proches malgré l'incohérence de l'échantillon pour les communes de 5 000 à 10 000 habitants.

Le résultat pour les communes de moins de 1 000 habitants reste tout de même plus élevé. D'une part le modèle surestime peut-être la consommation dans ces communes, mais la proportion du segment de l'enseignement de 2,6 % modélisée pour ces communes semble assez cohérent. D'autre part cette différence peut être plutôt expliquée par le fait qu'il y a moins d'enseignants en milieu rural donc la consommation par enseignant, montrée par le coefficient, est plus élevée. En effet, il est observé dans les données communales près de 6 élèves par enseignant pour les communes de plus de 10 000 habitants, contre 8,5 élèves par enseignant dans celles de moins de 1 000 habitants. Il est également possible qu'il y ait des types d'enseignement plus consommateurs en eau dans les communes rurales, par exemple avec des formations liées à l'agriculture.

Tableau 19 : coefficients calculés pour le secteur de l'enseignement, estimation de la cohérence et valeurs de la littérature

Echantillon	Variabes	Coefficients	Consommation (m ³) par élève/étudiant/an	Cohérence inter-modèles		Valeurs littéraires
Plus de 10 000 habitants	Nombre d'emplois dans le secteur de l'enseignement	51,222	8,6	3	2, 4	3 à 4 m ³ /an/élève, jusqu'à 6 m ³ /an/élève pour les établissements techniques ou les laboratoires universitaires 20 litres/élève/jour (Centre d'Information sur l'Eau
5 000 à 10 000 habitants	Nombre d'élèves de l'enseignement supérieur	33,94	1,4	1, 3	4	
1 000 à 5 000 habitants	Nombre d'emplois dans le secteur de l'enseignement	74,29	9,2	1	2, 4	
Moins de 1 000 habitants	Nombre d'emplois dans le secteur de l'enseignement	129,28	15,8	1, 3	2	

Par rapport aux chiffres du SMEGREG, les résultats sont assez élevés. Sur le total, hormis les communes de 5 000 à 10 000 habitants, près de 9 m³/élève/an est obtenu soit environ deux à trois fois les valeurs du SMEGREG. Le Centre d'Information sur l'Eau donne un chiffre équivalent à 7,3 m³/élève/an, plus cohérent avec ces résultats.

Globalement, hormis pour l'échantillon de communes de 5 000 à 10 000 habitants pour lequel le modèle n'est pas assez bon sur ce thème, les autres modèles semblent assez justes malgré une probable surestimation.

Santé

La comparaison pour ce secteur est délicate car les indicateurs ne montrent pas les mêmes choses. En revanche, comme le nombre de lits en secteur hospitalier prend également en compte les lits en EHPAD, la consommation modélisée par le modèle sur les communes de 1 000 à 5 000 habitants peut être rapportée sur le nombre de lits dans cet échantillon. Le résultat obtenu est d'environ 529 m³/lit, plus élevé que les 313,8 m³ calculés par le modèle des communes de plus de 10 000 habitants. Les deux chiffres sont toutefois dans le même ordre de grandeur.

La consommation modélisée pour ces deux modèles peut également être rapportée au nombre d'emplois dans les hôpitaux et les EHPAD, pour obtenir une valeur assez proche. Pour le dernier modèle, le secteur de la santé et de l'action sociale représente une chose différente. Cependant, en effectuant le même calcul avec la consommation expliquée uniquement par le nombre de lits en secteur hospitalier, le résultat obtenu est bien plus faible.

Tableau 20 : coefficients calculés pour le secteur de la santé, estimation de la cohérence et valeurs de la littérature

Echantillon	Variables	Coefficients	Consommation (m ³) par emploi hôpitaux/EHPAD/an	Cohérence inter-modèles		Valeurs littéraires
Plus de 10 000 habitants	Nombre de lits en secteur hospitalier	313,837	51,7	3	4	150 m ³ /lit d'hôpital/an, 60-90 m ³ /lit d'EHPAD/an
5 000 à 10 000 habitants						
1 000 à 5 000 habitants	Nombre d'hôpitaux ou d'EHPAD	3017,79	67,8	1	4	47,45-474,5 m ³ /lit/an (AHP)
Moins de 1 000 habitants	Nombre de lits en secteur hospitalier	93,68	9,9	1		
	Nombre d'emplois dans le secteur de la santé et de l'action sociale	69,01			3	

Les valeurs rapportées dans la littérature concernant la consommation par lit sont très variables, d'un facteur 10 selon une étude sur l'eau dans le domaine de la santé en Picardie⁵. Cela indique que les situations peuvent être très dépendantes des services de soin de l'hôpital, de sa taille etc, d'un hôpital à l'autre. Une approche statistique aussi large que celle présentée dans cette étude, et consistant finalement à calculer un coefficient moyennant différentes valeurs, peut avoir du mal à capter ces fluctuations.

Ces chiffres semblent être surestimés par les modèles si l'on se base sur les données du SMEGREG, et il peut également être conclu que les résultats des modèles ne sont peut-être pas représentatifs de la réalité au cas par cas.

Enfin, en appliquant le résultat obtenu par lit pour les communes de plus de 10 000 habitants et par emplois dans le secteur de la santé et de l'action sociale pour les communes de moins de 1 000 habitants, il est estimé que les modèles n'expliquent pas une consommation d'environ 4,2 millions de m³, du fait de l'absence d'indicateurs pour les communes de 5 000 à 10 000 habitants. Ce chiffre représenterait environ 11 % de la consommation des équipements de santé à l'échelle du bassin, et 5,9 % de la consommation totale au niveau de l'échantillon. Cela paraît assez élevé, mais cette estimation est cependant à relativiser car les coefficients peuvent être surestimés et l'utilisation du coefficient associé aux emplois du secteur de la santé et de l'action sociale tel quel peut être questionnée, étant donné qu'il n'apparaît que dans un échantillon.

Equipements sportifs

Les équipements sportifs n'apparaissent que dans deux échantillons : celui des communes de 5 000 à 10 000 habitants et celui des communes de 1 000 à 5 000 habitants. Le seul élément comparable entre les deux modèles concerne les bassins de natation. Il est observé une différence importante entre les deux modèles sur la valeur du coefficient, mais la littérature indique que la consommation des piscines et des équipements sportifs en général varie

⁵ Association des Hygiénistes de Picardie (AHP), « L'eau à l'hôpital », 2015.

grandement en fonction de la taille et de la fréquentation des équipements. Un coefficient par bassin de natation plus élevé pour les communes plus grande, qui verront donc une fréquentation plus forte, est donc logique.

Tableau 21 : coefficients calculés pour les équipements sportifs, estimation de la cohérence et valeurs de la littérature

Echantillon	Variabes	Coefficients	Cohérence inter-modèles	Valeurs littéraires
Plus de 10 000 habitants				
5 000 à 10 000 habitants	Nombre de bassins de natation	13389,7	3	300 à 500 m ³ /an pour les salles de sport
1 000 à 5 000 habitants	Nombre de bassins de natation	5097,31	2	
	Nombre d'équipements sportifs avec douches et sanitaires	371,11		
	Surface totale des terrains de grands jeux	1,08E-04		
Moins de 1 000 habitants				

Le coefficient obtenu pour les équipements sportifs avec douches et sanitaires correspond au chiffre de l'étude du SMEGREG.

En revanche, l'absence d'indicateurs sur les autres échantillons est problématique, et il peut être estimé que ce manque correspond à des volumes d'eau potable importants qui ne sont pas expliqués par les modèles. Pour les communes de plus de 10 000 habitants par exemple, en appliquant le coefficient de 13389,7 obtenu pour les bassins de natation, le volume est d'environ 10 millions de m³. Cela représenterait une proportion au niveau de l'échantillon de 1,4 % de la consommation d'eau potable totale, et sur tous les échantillons, ce volume représenterait la grande majorité de la consommation des équipements sportifs sur le bassin (environ 59 %). Pour les communes de moins de 1 000 habitants le problème est surtout lié à l'absence d'indicateurs sur la surface des terrains de sport, ce qui pourrait représenter, en comptant aussi les bassins de natation et les autres équipements sportifs, un volume non expliqué d'un peu plus de 3 millions de m³. Ce volume serait équivalent à environ 3 % de la consommation de l'échantillon.

Finalement, les coefficients obtenus sont assez cohérents. En revanche, il est estimé que les modèles n'expliquent pas une consommation d'un peu plus de 13 millions de m³, assez importante.

Industries

Pour ce secteur, la comparaison des coefficients entre eux et avec la littérature n'est pas forcément pertinente étant donné que ce sont des industries pouvant être de nature très différente, et dont les process peuvent consommer plus ou moins d'eau. Les deux coefficients obtenus sont cependant relativement proches (36,877 et 29,2).

Pour les modèles sur les communes de moins de 1 000 habitants et de 5 000 à 10 000 habitants, pour lesquels il n'apparaît pas d'indicateurs sur les industries, il est possible d'estimer qu'en appliquant les coefficients obtenus pour les autres échantillons, environ 5 millions de m³ ne sont pas expliqués par les modèles.

Pour ce secteur, il est également possible de se baser sur la base de données de l'IREEP qui estime un volume prélevé par les industries du bassin entre 35 et 40 millions de m³, plus élevé que les 28 millions calculés par les modèles (Annexe V : résultat des volumes prélevés prédits pour chaque segment d'usager).

Agriculture

Le tableau suivant présente la cohérence des modèles. Pour les deux modèles où les indicateurs apparaissent, la cohérence est bonne. En effet, la comparaison directe des coefficients des variables est compliquée car ce sont trois variables différentes, mais en ramenant la consommation modélisée pour le secteur de l'agriculture au nombre d'emplois du secteur, les chiffres obtenus sont similaires.

Tableau 22 : coefficients calculés pour le secteur de l'agriculture, estimation de la cohérence et valeurs de la littérature

Echantillon	Variables	Coefficients	Consommation (m ³) par emploi agriculture/sylviculture/an	Cohérence inter-modèles		Valeurs littéraires
Plus de 10 000 habitants						
5 000 à 10 000 habitants						
1 000 à 5 000 habitants	Nombre d'UGB porcins	17,03	57		4	
Moins de 1 000 habitants	Nombre d'UGB bovins Nombre d'emplois du secteur de l'agriculture et de la sylviculture	2,51 21,31	68	3		

Il est logique qu'aucun indicateur ne ressorte dans les communes de plus de 5 000 habitants. En appliquant les coefficients obtenus pour les autres modèles, la consommation projetée est très faible. Sur l'échantillon de communes de moins de 1 000 habitants en revanche, l'application du coefficient de 17,03 calculé sur l'autre modèle pour les UGB porcins modélise une consommation d'environ 7,6 millions m³, ce qui est très important pour cet échantillon.

Cependant, la pertinence d'appliquer ce coefficient sur le dernier modèle peut être questionnée. En effet, comme le montre l'échantillon de communes de moins de 1 000 habitants, le secteur de l'agriculture est expliqué par deux coefficients : un sur les UGB bovins et l'autre sur le nombre d'emplois du secteur qui est bien plus élevé. Alors que dans le modèle sur les communes de 1 000 à 5 000 habitants il n'y a qu'une variable sur les UGB porcins qui explique la consommation du secteur, avec un coefficient bien plus élevé que pour les UGB bovins, qui peut s'expliquer par le fait qu'il prend également en compte les emplois du secteur. Donc, de la même manière que sur le dernier modèle, le coefficient sur les UGB porcins pourrait en réalité être plus faible s'il y avait un indicateur sur le nombre d'emplois.

b. Comparaison avec les données fournies et intégrées aux modèles

Les résultats des projections peuvent être comparés aux données qui ont été envoyées par les différents services d'eau, même si leur typologie par segment d'utilisateur est différente. La catégorie « Professionnels », présente chez tous les fournisseurs, est plus compliquée à rattacher aux segments d'utilisateur définis dans l'étude par exemple, mais certaines typologies comme les « Syndicats Gestionnaires d'immeubles » sont facilement rattachables au segment sur la consommation domestique. Pour la comparaison avec les typologies « Collectivités », les quatre segments suivants sont regroupés : administration publique, santé, éducation et équipements sportifs.

En ce qui concerne la constante, il est choisi ici de l'intégrer au calcul en la répartissant proportionnellement aux différents segments d'utilisateurs, car elle semble permettre un meilleur ajustement des modèles et donc d'obtenir des résultats plus proches des données. Par exemple, si sur un périmètre la constante est de 4 millions de m³, et que sur ce périmètre la part projetée pour la consommation domestique est égale à 75 %, alors l'équivalent de 75 % de 4 millions de m³ est rajouté à la consommation domestique projetée.

Les comparaisons sont présentées sans inclure les communes touristiques.

SUEZ/SENEO

Les modèles surestiment la consommation totale sur le périmètre de SUEZ de 10,6 %, avec un volume prédit de près de 22,6 millions de m³ et réel d'environ 20,2 millions de m³.

En ce qui concerne la consommation des ménages, les typologies « `Syndic (gestionnaire d'immeuble) ` » et « Particuliers » peuvent être utilisées pour l'estimer. Les projections indiquent une proportion de la consommation liée aux ménages égale à 75,7 % sur les communes de SUEZ, contre une proportion assez proche de 76,6 % pour la consommation des deux typologies citées. En valeur absolue, les modèles prédisent une consommation d'environ 16,9 millions de m³, contre une consommation observée de 15,5 millions de m³, soit une surestimation d'environ 8,7 %.

Les typologies « Administration » et « Collectivité » sont comparables avec le regroupement des quatre segments sur l'administration publique, la santé, l'éducation et les équipements sportifs. Les proportions sont très similaires : 8 % pour les données observées et 8,4 % pour les projections. En valeurs absolues, les modèles prédisent une consommation d'environ 1,9 millions de m³, contre une consommation observée de 1,6 millions de m³, soit une surestimation d'environ 15,1 %.

La typologie des professionnels pourrait être comparée aux segments des commerces et du secteur tertiaire, aux industries et au segment du nettoyage des bâtiments et de l'entretien des espaces verts. Des proportions très proches sont obtenues : 14,9 % sur les projections et 15,1 % dans les données. En valeurs absolues, les modèles prédisent une consommation d'environ 3,6 millions de m³, contre une consommation observée de 3 millions de m³, soit une surestimation d'environ 15,1 %.

Ensuite, les 10 communes de SENE0 étant également gérées par SUEZ, les typologies des bases de données sont les mêmes. Ces 10 communes représentent une consommation de près de 33,6 millions de m³ sur l'année 2022 qui se répartit de cette manière :

- Administration = 4,1%
- Collectivité = 5,1 %
- Particulier = 13,4 %
- Syndicat gestionnaire d'immeuble = 56,7 %
- Professionnel = 20,7 %

Les projections indiquent une consommation d'environ 34,5 millions de m³, soit un différentiel de 2,6 % qui est bien plus faible que sur les données de SUEZ.

Pour la consommation domestique, SENE0 indique une proportion de 70,1 % alors que les projections fournissent un résultat de 71,4 % de la consommation totale, ce qui est donc très proche. En valeurs absolues, les projections indiquent un résultat d'environ 24,6 millions de m³, plus élevée d'environ 4,4 % par rapport aux 23,6 millions de m³ des données. La différence est encore une fois plus faible par rapport à la comparaison faite sur SUEZ.

Pour la consommation liée au secteur public et aux collectivités, la proportion de 9,2 % indiquée dans les données de SENE0 est très proche des 9,3 % obtenus avec les projections. En valeurs absolues, les modèles prédisent une consommation un peu plus élevée de 4,1 %.

Pour la typologie « Professionnels », par rapport aux trois segments correspondants, des proportions similaires sont obtenues : 19,2 % sur les projections et 20,7 % dans les données. En valeurs absolues, les modèles prédisent une consommation d'environ 6,6 millions de m³, contre une consommation observée de 6,9 millions de m³, soit une sous-estimation d'environ 4,5 %.

Enfin, étant donné que les communes de SENE0 sont des communes de plus de 10 000 habitants, il est également intéressant de comparer les données fournies avec le résultat des proportions issus du modèle général pour les communes de plus de 10 000 habitants (Tableau 9). Les proportions associées au secteur domestique sont légèrement plus faibles dans les données (70 % contre 72,5 %), de même que pour les établissements publics et liés à la collectivité (9,2 % contre 10,8 %).

Veolia

En ce qui concerne le volume total, les projections estiment un volume consommé de 132,9 millions de m³ contre une consommation observée de 117,8 millions de m³, soit une surestimation de près de 11,4 %.

La typologie des données fournies par Veolia n'est pas évidente à rapprocher aux segments d'utilisateur définis. Il n'y a par exemple aucune catégorie correspondant clairement et exclusivement aux établissements de santé, aux équipements sportifs, ou encore au secteur tertiaire. Les contacts à Veolia ont par exemple indiqué que les établissements de santé pouvaient être retrouvés dans plusieurs catégories. Il est tout de même possible de considérer que les typologies « Collectifs », « Individuels » et « Individuels (SRU) » sont représentatives de la consommation liée aux ménages. Une proportion de 86 % est obtenue, bien plus élevée que la proportion des ménages projetée, égale à 77 %. En revanche la différence en valeurs absolues, et en prenant en compte la constante, est de moins de 1 million de m³ soit une

surestimation des modèles d'environ uniquement 1 %. Les autres segments d'utilisateur sont donc probablement surestimés par les modèles.

Les typologies de Veolia sur les bâtiments communaux et les appareils publics peuvent être regroupées dans une catégorie portant sur la consommation liée aux collectivités. Le problème est que comme expliqué juste avant, il n'est pas possible de savoir exactement ce qu'il y a dans ces catégories. Les établissements de santé par exemple ne sont pas forcément compris dedans. En regroupant les mêmes typologies issues des projections que précédemment pour estimer la consommation liée aux collectivités, les modèles prédisent une consommation plus élevée de près de 39,9 %. Cette comparaison est peut-être peu fiable, mais rejoint finalement ce qui a été observé sur l'analyse des coefficients, c'est-à-dire une surestimation probable des chiffres liés à ces établissements.

La typologie « Irrigations et agricoles » peut être comparée avec le segment sur l'agriculture et l'élevage. Cette typologie représente environ 2,3 % de la consommation totale, alors que les projections indiquent une proportion de près de 1,5 % pour ce segment d'utilisateur. Les modèles prédisent une consommation de 2,1 millions de m³, contre une consommation observée de 2,7 millions de m³, soit une sous-estimation des modèles d'environ 22,5 %.

VEDIF

Les projections prédisent un volume consommé total sur le VEDIF 242,5 millions de m³ contre une consommation observée de 241,4 millions de m³, soit une très faible différence d'environ 0,4 %. Cette différence est la plus faible observée.

Le VEDIF a ensuite fourni deux types de segmentation par utilisateur : une en quatre postes et une en 22 postes. La typologie en 22 postes est bien sûr plus détaillée mais n'est pas forcément évidente à rattacher aux segments d'utilisateur définis dans les modèles. La typologie en quatre postes est plus simple pour la comparaison, et propose un segment « Particulier » et un segment « Gestionnaire d'immeubles », qui correspondent théoriquement principalement à de l'habitat. La consommation des ménages projetée représente 75,4 %, alors que les consommations réelles associées à ces segments représentent 74,5 % et sont donc très proches des résultats. En valeurs absolues, cela représente une consommation observée égale à environ 179,9 millions de m³, contre une consommation projetée de 182,9 millions de m³, ce qui équivaut à un delta de 1,6 %.

La segmentation en quatre postes propose également une typologie « Collectivités ». La consommation observée pour ce segment représente 8,7 % de la consommation totale. Alors qu'en agrégeant les quatre segments d'utilisateurs correspondant aux établissements publics, une proportion de 8,5 % est obtenue dans nos projections, soit un résultat très similaire. En valeur absolue, cela donne une consommation de 20,9 millions de m³ dans les données observées contre 20,6 millions de m³ projetée, soit une différence faible de 1,6 %.

En comparant la typologie des Professionnels avec les trois segments équivalents évoqués précédemment, des résultats similaires sont également obtenus : une proportion de 16,1 % estimée par les projections et de 16,8 % dans les données. En valeurs absolues, les modèles prédisent une consommation d'environ 39 millions de m³, contre une consommation observée de 40,6 millions de m³ soit une sous-estimation d'environ 4 %.

En ce qui concerne la segmentation en 22 postes, une typologie concerne les établissements de santé, comparable avec les projections pour le segment de la santé. Les proportions sont similaires, représentant 2,1 % de la consommation dans les données observées et 2,6 % dans

les projections. En valeurs absolues, la consommation des établissements de santé est égale à environ 5,1 millions de m³ d'après les données et à 6,3 millions de m³ dans les projections, soit un delta plus élevé d'environ 19,4 %.

En ce qui concerne les équipements sportifs, un problème du modèle pour les communes de plus de 10 000 habitants est qu'il ne montre pas les consommations liées aux équipements sportifs. Cependant, les données du VEDIF indiquent une consommation de 1,7 millions de m³ consommées, soit une part très faible du volume total sur le VEDIF (environ 0,6 %). Ces faibles volumes expliquent peut-être pourquoi les indicateurs ne sont pas significatifs. Les projections sur le VEDIF pour les équipements sportifs estiment un volume consommé de 0,4 million de m³, avec la présence de 33 communes ayant moins de 10 000 habitants.

Conclusion

En conclusion, ces comparaisons permettent de mettre en évidence et de confirmer que les modèles fonctionnent très bien sur les communes du VEDIF et de SENEQ, qui concernent principalement des communes de plus de 5 000 habitants, même si localement les marges d'erreur peuvent être plus élevées. Néanmoins, les résultats sont un peu moins bons pour les communes de Veolia. En effet, ces communes concernent plutôt les communes moins peuplées pour lesquelles les modèles fonctionnent moins bien, d'où les plus grands écarts entre les valeurs observées et théoriques. De plus, ces données sont les plus hétérogènes avec de nombreux reports de facturation d'une année à l'autre et la présence récurrente de valeurs aberrantes. De la même manière que pour Veolia, les résultats sont aussi un peu moins satisfaisants sur les communes de SUEZ. C'est plus surprenant car ce sont majoritairement des communes de plus de 5 000 habitants, pour lesquelles les modèles fonctionnent le mieux. L'explication peut également être due aux biais introduits par la gestion des volumes apparaissant comme négatifs dans la base de données, à cause des reports de facturation. En effet, il a été choisi de laisser les volumes négatifs dans la base de SUEZ ce qui a pour effet de tirer les consommations observées vers le bas, et explique donc pourquoi les modèles surestiment les consommations sur ce périmètre.

Les consommations domestiques sont les mieux prédites, ce sont en effet pour ces indicateurs que les coefficients dans les modèles étaient les plus significatifs. Cependant, les modèles semblent surestimer plutôt les acteurs de la santé (VEDIF) et liés aux établissements publics en général (Veolia, SENEQ, SUEZ), ce qui confirme l'analyse des coefficients faites précédemment. Sur le VEDIF en revanche, la catégorie des établissements publics est légèrement sous-estimée. C'est logique car finalement la consommation des établissements de santé est ici surestimée et probablement aussi celle pour l'enseignement. Mais cette surestimation est en fait compensée par le fait que la consommation des équipements sportifs est grandement sous-estimée, car ils n'apparaissent pas pour les communes de plus de 10 000 habitants qui caractérisent majoritairement les communes du VEDIF. Les modèles ont tendance à sous-estimer la catégorie des Professionnels (VEDIF, SENEQ), ce qui est une nouvelle fois logique par rapport à ce qui est observé dans la section précédente. Le secteur des commerces et du tertiaire semble bien expliqué par les modèles, mais il y a en effet probablement des volumes d'eau liés aux secteurs de l'industrie et de l'agriculture qui ne sont pas modélisés.

Il peut finalement être souligné qu'il est assez cohérent, pour les communes ayant permis de construire les modèles, d'obtenir des volumes totaux projetés similaires. En revanche, étant donné qu'uniquement les volumes totaux par commune ont été utilisés, retrouver des valeurs

aussi proches par segment d'utilisateur est plus surprenant et montre que les modèles fonctionnent bien, en tout cas sur les communes fortement peuplées.

c. Comparaison avec les données fournies non intégrées aux modèles

Pour répondre au possible biais introduit par le fait de comparer les projections aux données même ayant permis de construire les modèles, il est également possible de comparer les résultats des projections aux données d'Eau de Paris et de l'Agglomération de Reims, qui elles n'ont pas été intégrées aux modèles.

Eau de Paris

Les données d'Eau de Paris ne proposent pas de typologies d'utilisateurs, il n'y a donc pour l'instant pas d'éléments de comparaison par segment. Il est tout de même possible de comparer le résultat projeté pour Paris issu du modèle pour les communes de plus de 10 000 habitants, et la consommation totale observée, et voir qu'il prédit parfaitement bien la consommation avec une surestimation d'à peu près seulement 0,4 %.

Agglomération de Reims

Sur les 49 communes de l'Agglomération pour lesquelles des données ont été obtenues, la consommation totale est d'environ 14,6 millions de m³. Les projections estiment un volume d'environ 15,5 millions de m³, soit une surestimation plutôt faible d'environ 5,9 %.

Les typologies de l'Agglomération sont très détaillées et les consommations indiquées sont rattachées directement à un site. Il a donc été réalisé un traitement pour essayer de rapprocher les typologies fournies aux segments d'utilisateur.

L'Agglomération indique des typologies correspondant aux ménages, aux ménages + commerces lorsque des commerces sont situés en bas d'immeuble, et au tertiaire. En ce qui concerne uniquement la consommation des ménages, les modèles prédisent une consommation d'environ 11,6 millions de m³ contre une consommation de 11,4 millions de m³ réelle, soit une légère surestimation d'environ 2 %. En revanche, le secteur des commerces et du tertiaire est fortement surévalué, d'environ 3 fois la valeur réelles des données. Cependant, il est possible que certaines consommations dans les données soient attribuées aux ménages alors qu'elles correspondent en fait à la catégorie ménages + commerces et donc qu'elles soient liées à des commerces. Une autre possibilité est que le modèle sur les communes de plus de 10 000 habitants, qui explique principalement la différence importante sur ce segment, présente un indicateur sur le nombre de commerces et d'établissements du tertiaire qui est très corrélé à l'indicateur sur le nombre d'appartements. Cette corrélation peut donc impliquer, à l'inverse, que les consommations attribuées aux commerces en bas d'immeuble sont en fait des consommations domestiques, comme nous l'avons déjà évoqué.

Les modèles prédisent ensuite très bien ce qui se rattache aux industries, à la santé et à l'administration publique, avec des différences inférieures à 2 %. En revanche, comme prévu l'enseignement est grandement surestimé, environ 2 à 3 fois plus que ce qu'indique les données. Les équipements sportifs sont également sous-estimés, environ 2 fois moins.

Il est donc finalement démontré qu'il n'y a pas de biais car les modèles fonctionnent aussi pour les communes qui n'ont pas été intégrées dans leur conception, tout en présentant les mêmes limites déjà évoquées précédemment soit : la surestimation de l'enseignement, une

surestimation moins importante pour la santé et une sous-estimation des équipements sportifs du fait de leur absence dans certains modèles.

Afin de proposer un test de sensibilité qualitatif et au vu des éléments soulevés dans les parties précédentes, il est possible d'essayer de comparer les résultats avec une situation où les bornes basses de consommation pour l'enseignement et la santé sont attribuées (postes surestimés), que les consommations liées aux équipements sportifs sont augmentées (10 millions pour les communes de plus de 10 000 habitants et 3 millions pour les communes de moins de 1 000 habitants), et que la valeur de prélèvement pour les industries de 35 millions de m³, indiquée dans les données de l'IREP et que nous considérons comme fiables, est attribuée pour ce secteur. Les prélèvements liés aux ménages sont également augmentés en leur attribuant la valeur de la borne haute, malgré le fait que les comparaisons effectuées précédemment indiquent que ce segment semble correctement prédit. La consommation des ménages étant le plus gros facteur d'influence sur la variabilité des résultats, ce choix permet de se rapprocher des chiffres de prélèvement totaux sur le bassin pour la production d'eau potable, et donne plus de poids à la consommation des ménages, au détriment de celle des collectivités et du secteur des commerces et tertiaire. Les résultats des prélèvements associés sont présentés dans le tableau suivant, accompagnés des résultats initiaux qui avaient été calculés :

Tableau 23 : résultats des prélèvements prédits par les projections ajustées et les hypothèses

Segments d'utilisateur	Projections ajustées		Hypothèses	
	Prélèvements projetés (Millions de m ³ /an)	Part	Prélèvements projetés (Millions de m ³ /an)	Part
Ménages	924	74,2	1 103,1 (Borne haute)	77,1
Commerces et tertiaire	133,7	10,7	133,7	9,3
Etablissements publics	120	9,6	120,4	8,4
<i>Enseignement</i>	45,4	3,6	37,1 (Borne basse)	2,6
<i>Santé</i>	38,4	3,1	31,3 (Borne basse)	2,2
<i>Administration publique</i>	27,7	2,2	27,7	1,9
<i>Equipements sportifs</i>	8,5	0,7	24,3 (Rajout volumes manquants estimés)	1,7
Industries	28,1	2,3	35 (Données IREP)	2,4
Services de nettoyage de bâtiments et d'entretien des espaces verts	27,1	2,2	27,1	1,9
Agriculture - élevage	12,2	1	12,2	0,8
<i>Surplus tourisme</i>	8,5		8,5	
<i>Constante</i>	8,4		8,4	
Total	1 262		1 448,4	

Avec cette hypothèse, le total des prélèvements pour la production d'eau potable est cette fois surestimé d'un peu moins de 6 % par rapport au chiffre de 1,37 milliard de m³. Ce résultat est tout de même plus proche qu'en utilisant les projections ajustées. **Les projections ajustées et cette hypothèse permettent finalement d'établir un intervalle relativement fiable des prélèvements par segment d'utilisateur pour la production d'eau potable.**

Pour finir, malgré ces observations positives sur la fonctionnalité des modèles, il est crucial de reconnaître certaines limites de l'étude. Bien que les modèles utilisés offrent des projections ajustées qui apparaissent globalement cohérentes, la présence d'hétéroscédasticité, la largeur

des intervalles de confiance ou encore le fait que certaines variables n'apparaissent pas dans certains modèles, soulèvent des questions quant à la précision de ces estimations. De plus, d'autres facteurs non modélisés pourraient influencer les résultats, ce qui invite à une interprétation prudente des projections. Ces éléments ouvrent la discussion sur les contraintes méthodologiques et les possibles biais qui pourraient affecter la robustesse des conclusions.

Matériels et méthodes

L'étude présente est un travail exploratoire, basé sur l'élaboration de modèles statistiques multi-variés pour prédire et expliquer la consommation en eau potable. Cette approche n'était encore récemment que peu utilisée. La majorité des études de ce type, que l'on peut par exemple retrouver dans les documents de planification, consistaient uniquement à utiliser un ratio de consommation par abonné considéré comme stable dans le temps, et à projeter la population future pour faire des prévisions de consommation d'eau potable (J.D Rinaudo, 2013).⁶

Certaines méthodes basées sur les modèles statistiques multi-variés ont vu le jour. C'est par exemple le cas de l'étude du Crédoc de 2016⁷, qui cherche à construire des modèles explicatifs de la consommation en eau potable en Ile-de-France, dans l'objectif de créer un outil permettant de simuler les consommations en modifiant certains paramètres communaux. Une autre étude, plus ancienne, réalisée par N. Neverre & Rinaudo en 2010⁸ dans le cadre d'un projet de service public, propose une étude économétrique de la demande en eau potable dans le département de l'Hérault. Enfin, une étude de M. Favre & Montginoul en 2020⁹ décrit la création d'un observatoire des consommations en eau potable sur un territoire pilote, Nantes Métropole, basée sur des fonctions de demande en eau.

Ces études ont pour objectif de construire des modèles explicatifs de la consommation afin ensuite de simuler l'évolution de la consommation totale dans le temps, si les paramètres explicatifs sont modifiés. Leur objectif est de visualiser des tendances sur des périmètres d'étude plus petits et similaires, pour estimer l'élasticité de la consommation de l'eau par rapport aux variables utilisées¹⁰. Elles privilégient pour cela l'utilisation de données relatives, comme la proportion de maisons par commune par exemple.

L'étude présente a plutôt pour objectif de **dresser une photographie de l'état actuel de la consommation en eau potable sur le bassin et d'identifier les différents types d'usagers (ménages, commerces, industries, établissements de santé...)**, en construisant des modèles explicatifs puis en faisant des projections sur les communes où n'ont pas été obtenues de données. Une approche quantitative est donc privilégiée, avec l'utilisation de variables basées sur des données brutes, par exemple le nombre de maisons pour notre cas, témoignant du fait que plus la population d'une commune et le nombre d'équipements augmente, plus la consommation d'eau potable devrait également augmenter. Au contraire,

⁶ Rinaudo, « Prévoir la demande en eau potable », 2013.

⁷ CREDOC, « Prévoir la consommation d'eau en Ile-de-France », 2016.

⁸ Neverre et Rinaudo, « Conseil Général de l'Hérault et Brgm », 2010.

⁹ Favre et Montginoul, « Comment observer de façon optimale la consommation en eau potable ? », 2020

¹⁰ L'élasticité de la demande en eau est la mesure qui témoigne de la variation de la demande en eau d'un territoire lorsqu'on fait varier l'un de ses paramètres (comme le prix de l'eau par exemple).

cette méthodologie permet plutôt de dégager des ratios de consommation unitaire, par maison, par appartement ou par lit d'hôpital par exemple, chose encore relativement peu étudiée dans la littérature.

1. Récolte et traitement des données

La première phase de l'étude a été consacrée à la récolte et aux traitements des données à intégrer dans nos modèles. D'une part des données de consommation d'eau potable, de l'autre des données communales qui serviront de variables explicatives.

a. Eau potable

L'une des idées de l'étude est d'approcher la consommation d'eau potable par des variables spatiales (nombre d'habitants, etc), Or, l'agence de l'eau ne dispose que du lieu de prélèvement de l'eau potable et pas du lieu où l'eau est effectivement consommée par l'utilisateur final. Ces deux lieux peuvent être très éloignés. Il a donc été décidé de récupérer des données de consommation auprès de fournisseurs d'eau potable du bassin. Plusieurs gestionnaires ont ainsi été contactés, avec la volonté de couvrir une grande partie du territoire et d'avoir des organismes variés (délégations, régies...) :

- Veolia Eau d'Ile-de-France (VEDIF)
- SUEZ
- Veolia
- SENEIO
- Communauté urbaine du Grand Reims
- Agglomération de la région de Compiègne
- Le Havre Seine Métropole
- Eau de Paris
- Syndicat des eaux d'Ile-de-France (SEDIF)

Parmi cette liste, les données des fournisseurs suivants ont été récupérées :

Fournisseur	Nombre de communes	Période
VEDIF	150	2018-2022
Veolia	2495	2022
SUEZ	51	2018-2023
SENEIO	10	2019-2022
Eau de Paris	1	2018-2022
Agglomération du Havre	46	2022
Communauté urbaine du Grand Reims	49	2023

Les données d'eau potable fournies correspondent aux volumes facturés, c'est-à-dire les volumes d'eau vendus et arrivant aux compteurs des abonnés. Ces volumes sont considérés représentatifs des volumes consommés. Notamment, ils ne sont pas influencés par l'état du réseau d'eau potable. L'objectif de l'étude étant de comprendre comment la consommation d'eau potable se distribue entre les différents types d'utilisateurs, il a été demandé aux opérateurs

de nous fournir une typologie des consommations par poste et par commune la plus précise possible.

Veolia, le VEDIF et SENE0 nous fournissent des données par commune et selon différentes typologies, qui sont présentées en Annexe I : typologies fournies par les différents services d'eau. Les différentes bases de données présentent une ligne par commune et des valeurs de consommation correspondante par typologie. En revanche, SUEZ a fourni une base de données bien plus détaillée, par site de consommation, lui-même rattaché à une commune.

Nous avons reçu les données d'Eau de Paris et de l'agglomération du Grand Reims mais elles étaient trop difficiles à homogénéiser avec les autres données pour les utiliser au sein des modèles. Au vu de la taille de la capitale, Eau de Paris aurait mérité un traitement à part, que la durée de l'étude ne nous a pas permis de réaliser. Comme on le verra plus tard, ces données ont tout de même pu être utilisées pour tester la pertinence du modèle.

Enfin, remarquons également que certaines communes étaient présentes dans plusieurs bases de données : celles de SENE0 dans celles de SUEZ, mais aussi celles de l'Agglomération du Havre dans celles de Veolia.

Les données ont ensuite été traitées et homogénéisées : addition des volumes facturés lorsque plusieurs distributeurs opèrent sur une même commune (c'est le cas pour quatre communes de Suez, également desservies par Veolia et le VEDIF), travail sur les dénominations des communes qui n'étaient pas toujours fournies avec leur code INSEE afin de les identifier, etc. Une limite importante de plusieurs des bases de données était la présence de volumes négatifs. Ces volumes négatifs sont liés aux traitements effectués pour la facturation, et des reports d'une année sur l'autre. Pour Veolia, il a été nécessaire de supprimer les communes correspondantes, soit 156 communes. Pour SUEZ, les volumes négatifs ont été laissés tels quels dans la base de données. D'une part parce que nous avons les données par site de consommation, et qu'en agréant par commune il n'y avait pas de volumes négatifs par commune, à l'inverse de Veolia. Et comme les données couvraient plusieurs années, les valeurs ont pu être comparées pour chaque commune d'une année sur l'autre. Cela a permis de conclure que les données étaient relativement stables, et que conserver les volumes négatifs n'introduisait pas de biais trop importants. Remarquons qu'en tout état de cause, les reports de facturation n'apparaissent pas forcément sous la forme d'un volume négatif. Il suffit que le remboursement soit inférieur à la facture de l'année. De plus, tout report entraîne aussi une année de surévaluation des volumes consommés. Il n'est pas possible ici de discerner ces reports, bien qu'ils créent des biais sur les données de consommation.

Après tous les traitements détaillés ci-dessus, un total de 2024 communes pour lesquelles des données ont été recueillies est obtenu, et qu'il est possible de localiser spatialement à l'aide de la carte suivante :

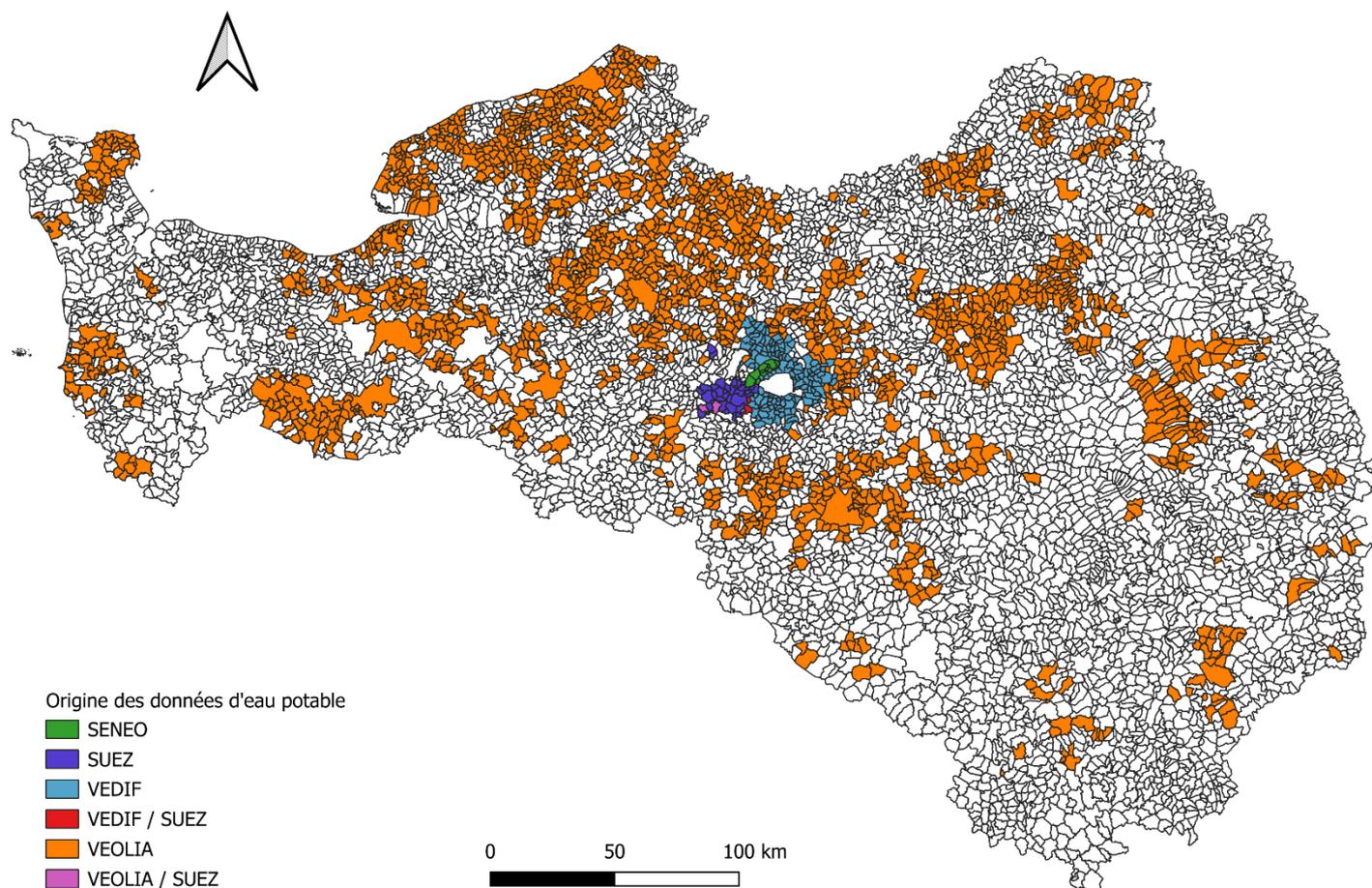


Figure 7 : communes du bassin pour lesquelles ont été recueillies des données

Les données du VEDIF, de SENEQ et de SUEZ concernent des communes d'Ile-de-France et donc principalement des communes urbaines avec une forte population. Quant aux données de Veolia, elles couvrent de nombreuses communes sur le reste du bassin, principalement rurales. Il est donc considéré que les données collectées sont représentatives du bassin ainsi que de la diversité des types de communes. Le tableau suivant présente le nombre de communes avec des données par rapport au nombre de communes totales, en fonction de leur population. Il montre que même s'il manque environ les $\frac{3}{4}$ des communes, rurales notamment, près de la moitié de la population totale du bassin est représentée.

Tableau 24 : caractéristiques des communes pour lesquelles des données ont été recueillies, par rapport aux communes du bassin

Taille des communes	Plus de 10 000 habitants	5 000 – 10 000 habitants	1 000 – 5 000 habitants	Moins de 1 000 habitants	Total
Nombre de communes avec données	208	68	433	1 315	2024
Population sur les communes avec données	6 925 298	478 136	894 281	519 192	8 816 907
Nombre de communes total	345	191	1 379	6 100	8 015
Population totale	12 513 835	1 350 745	2 843 780	2 074 715	18 783 075
Part communes avec données (%)	60,3	35,6	31,4	21,6	25,3
Part population (%)	55,3	35,4	31,4	25	46,9

Comme évoqué, parmi les 2024 communes présentées il peut donc y avoir des biais dans les données entraînant des valeurs aberrantes, que les premiers traitements n'ont pas décelées. Afin de repérer et traiter ces anomalies, pouvant être induites par les reports de facturation, par des erreurs dans la saisie des données ou par le fait que certaines communes puissent être desservies par plusieurs distributeurs non identifiés, une solution a été de calculer la consommation totale par habitant. En effet, une commune de 22 150 habitants par exemple est indiquée comme consommant un volume de 5271 m³, correspondant à une consommation totale par habitant de 0,23 m³ ce qui est bien trop peu, et indique un problème dans la donnée. Les communes où les valeurs sont inférieures à 30 m³/habitant et supérieures à 200 m³/habitant sont donc supprimées. En sachant que la consommation moyenne par habitant et par an est de 36 m³ d'après une étude de l'INSEE en Ile-de-France, en prenant en compte uniquement les volumes directement livrés aux ménages, une borne basse de 30 m³/habitant d'eau totale (donc l'eau arrivant aux ménages + le reste) est considérée comme cohérente, tout en restant assez large. La borne haute est fixée à 200 m³/habitant, une consommation totale élevée par habitant pouvant être expliquée par la présence d'infrastructures importantes sur la commune (hôpitaux, industries, nombreux bureaux) qui ferait gonfler la consommation par habitant. Pour le reste, impossible à prétraiter, il sera décidé plus tard de retirer les communes les plus aberrantes par rapport aux modèles de régression.

b. Données communales

Afin d'expliquer la consommation d'eau potable selon différents usagers, des données communales permettant de les caractériser au mieux ont été collectées. Les catégories principales d'usagers qu'il est nécessaire de visualiser sont : les ménages, le secteur industriel, les commerces et le secteur tertiaire (correspondant au secteur tertiaire marchand d'après l'INSEE), l'agriculture, les établissements de santé, l'éducation, l'administration publique et les équipements sportifs.

Afin de caractériser la consommation associée aux ménages, des données sur les caractéristiques des logements ont d'une part été recueillies (nombre de résidences principales qui sont des maisons ou des appartements, nombre de résidences secondaires, nombre moyen de pièces par résidences principales...). En effet, une maison pourrait

consommer plus qu'un appartement, du fait d'une surface plus importante à entretenir ou de la présence éventuelle de jardin et de piscine. De plus, il est possible de penser qu'une résidence secondaire aura tendance à consommer moins que les résidences principales, étant donné qu'elle ne sera occupée qu'une partie de l'année. Enfin, des données sur l'ancienneté des logements ont été collectées, des logements anciens pouvant potentiellement consommer plus que des logements récents à cause d'éventuelles fuites.

D'autre part, des données sur les caractéristiques des ménages ont également été collectées (nombre de ménages, taille moyenne des ménages, nombre de ménages possédant 2 voitures ou plus, nombre de ménages d'une seule personne...) et de la population (nombre d'habitants, nombre d'habitants non scolarisés et non diplômés, nombre de chômeurs). En effet, la population est évidemment le facteur qui va jouer le plus sur la consommation en eau potable. La taille des ménages devrait en théorie jouer un rôle positif dans la consommation, mais certaines études montrent que ça n'est pas le cas et que plus le ménage est grand, plus il réalise des économies d'échelles. Certains facteurs socio-économiques peuvent également jouer un rôle. C'est par exemple le cas du revenu, les ménages plus aisés étant moins sensibles à leur facture d'eau ils pourraient avoir tendance à consommer plus. Les personnes ayant un plus haut niveau d'étude pourraient également être plus sensibles aux sujets environnementaux, et donc chercher à être plus attentives à leur consommation d'eau et à éviter les gaspillages.

Ces différentes données proviennent de diverses bases de données de l'INSEE : le recensement de la population et le dispositif Fichier Localisé Social et Fiscal (Filosofi) en particulier.

Pour les autres usages, des données issues d'une étude socio-économique menée sur le bassin, qui détermine par commune le nombre d'établissements et d'emplois par code NAF, a été mobilisée. Cette base a permis de caractériser les usages suivants :

- **Industries** : les industries sont définies à l'aide des codes NAF 35.11Z, 35.21Z et 35.30Z. Ces codes NAF correspondent aux industries manufacturières et extractives, ainsi que les industries de production d'électricité, de gaz, de vapeur et d'air conditionné.
- **Commerces et secteurs tertiaires** : les nombreux codes NAF suivants ont été utilisés pour définir cette catégorie : ceux de la section 35, hormis ceux utilisés pour l'industrie, les codes NAF des sections 36 à 39, tous les codes NAF des sections 45, 46, 47, de 49 à 53, 56, de 58 à 66, de 68 à 75, de 77 à 82, de 90 à 92 et de 94 à 99.
- **Agriculture** : cet usage est caractérisé par les codes NAF présents dans les sections 1 et 2, correspondant aux établissements de l'agriculture et de la sylviculture.
- **Santé** : les établissements considérés comme Hôpitaux et EHPAD correspondant aux codes NAF 86.10Z (activités hospitalières) et ceux de la section 87 (hébergement médico-social et social), sont séparés des établissements considérés comme relevant de la santé et de l'action social. Ces derniers correspondent aux autres codes NAF de la section 86 et ceux de la section 88, et représentent en théorie les activités liées à la santé sans capacité d'hébergement.
- **Education** : les codes NAF correspondant à la section 85 sont utilisés pour caractériser l'éducation. A noter que cette section regroupe les emplois de l'enseignement du pré-primaire au supérieur.

- **Administration publique** : ce segment est caractérisé par les codes NAF de la section 84.
- **Services de nettoyage de bâtiments et d'entretien des espaces verts** : les sections 81.2 et 81.3 sont utilisées.

D'autres sources de données ont permis de compléter l'étude socio-économique pour certains usages. Pour l'agriculture, des données de l'Agreste fournissant le nombre d'Unité de Gros Bétail (UGB) bovins et porcins ont aussi été mobilisées. En effet, c'est un indicateur de l'intensité de l'élevage sur la commune, secteur pour lequel il est potentiellement utilisé de l'eau potable pour abreuver les bêtes.

En ce qui concerne le domaine de la santé, la Statistique Annuelle des Equipements de santé (SAE), afin d'obtenir la capacité d'accueil des établissements (nombre de lits, nombre de séjours en hospitalisation complète, nombre de séances d'hémodialyse...) a aussi été utilisée. Le fichier « capact21 » a été mobilisé pour cela, qui indique ces infos par unité (médecine, chirurgie...) et par centre hospitalier (CHU, cliniques...) et EHPAD.

Pour l'enseignement, des données issues de l'Observatoire des Territoires ont été collectées pour obtenir le nombre d'écoles et d'élèves du primaire et du secondaire, et le nombre d'étudiants du supérieur.

Pour finir, le recensement des équipements sportifs de l'INSEE a été utilisé, qui indique le type d'équipement (terrains de grands jeux, bassins de natation...) ainsi que leur surface, et si l'équipement possède des douches et des sanitaires. Un traitement a été effectué pour avoir la surface totale de « Terrains de grands jeux », afin d'essayer d'avoir un indicateur témoignant de la quantité d'eau utilisée pour l'arrosage des terrains de foot par exemple, et la surface totale des bassins de natation, par commune. En revanche, certains terrains et bassin n'avaient pas de données de surface. La solution a été de leur attribuer la surface moyenne calculée sur tout le jeu de données. Cette méthode peut introduire des biais statistiques, mais il est considéré que cela en entraînerait moins que de supprimer les équipements sportifs sans surface.

2. Construction des modèles

Dans cette étude, des modèles explicatifs de la consommation en eau potable sont donc construits. Pour cela, des modèles de régression linéaire multiple sont élaborés, c'est-à-dire combinant plusieurs variables, qui permettent, en plus d'expliquer la consommation en eau potable de réaliser ensuite des projections.

Les typologies de consommation d'eau potable fournies par les distributeurs ne sont pas homogènes entre eux, à cause notamment du fait que chaque opérateur n'a pas forcément la même vision sur qui consomme l'eau. Il a donc été choisi de construire les modèles en se basant sur un seul objet d'étude commun, à savoir la consommation d'eau totale par commune, et sur l'année 2022 uniquement. Veolia, qui représente la grande majorité des communes, n'a en effet fourni que cette année.

a. La régression linéaire

Les régressions linéaires sont des outils statistiques permettant de modéliser une relation entre une variable dépendante Y , ici la consommation d'eau totale par commune, et une ou plusieurs variables explicatives X , qui sont les données communales.

Formule générale de la régression linéaire multiple

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Dans cette formule, β_0 est la constante, c'est-à-dire la valeur de Y lorsque les variables explicatives sont égales à 0, X_1, X_2, \dots, X_n sont les variables explicatives, $\beta_1, \beta_2, \dots, \beta_n$ sont les coefficients des variables explicatives, et ϵ l'erreur résiduelle du modèle. Les coefficients associés aux variables explicatives sont estimés par la méthode des moindres carrés ordinaires. Cette méthode minimise la somme des carrés des différences entre les valeurs observées et les valeurs prédites du modèle.

Hypothèses de la régression

La régression linéaire doit répondre à certaines hypothèses sous-jacentes. Parmi celles-ci :

- *Linéarité* : la relation entre la variable dépendante et les variables indépendantes doit être linéaire.
- *Homoscédasticité* : cette hypothèse énonce que la variance des erreurs de prédiction du modèle, que l'on appelle les résidus, doit être constante à travers les valeurs des variables indépendantes et donc ne pas présenter d'hétéroscédasticité.
- *Non colinéarité* : les variables indépendantes ne doivent pas être colinéaires, c'est-à-dire qu'elles ne doivent pas évoluer de la même façon partout.

Evaluation des modèles

Enfin, les modèles de régression linéaire peuvent être évalués par quelques méthodes :

- *Le coefficient de détermination R^2* : le R^2 mesure la proportion de la variance de la variable dépendante qui est expliquée par les variables explicatives du modèle. Il varie de 0 à 1 : plus il est proche de 1, mieux le modèle explique la variable Y .
- *Le test de Fischer (F-test)* : le test de Fischer évalue la significativité globale du modèle et si les variables explicatives expliquent une proportion importante de la variabilité de Y . La p-value associée à ce test indique la probabilité que les coefficients des variables explicatives soient tous égaux à zéro.
- *Le test de Student (T-test)* : il permet de tester la significativité individuelle de chaque coefficient, à l'inverse du F-test qui teste la significativité globale. La p-value associée à ce test indique la probabilité que chaque coefficient des variables explicatives soient égaux à zéro.
- *L'erreur standard résiduelle* : c'est une mesure de la qualité de l'ajustement du modèle de régression linéaire. Cela représente l'écart moyen entre les valeurs théoriques calculées par le modèle et les valeurs réelles.
- *Test de Breusch-Pagan et graphique Residuals vs Fitted* : ces deux méthodes permettent d'estimer la présence d'hétéroscédasticité dans les modèles. Le graphique montre la distribution des résidus, qui doit être aléatoire autour de 0 et ne pas présenter

de structure particulière pour répondre à l'hypothèse d'homoscédasticité. Alors que le test statistique fournit un résultat interprétable comme suit : si la p-value est supérieure au seuil de 5 % (0,05) alors les résidus sont homoscédastiques, dans le cas contraire le test démontre la présence d'hétéroscédasticité.

- *VIF (Variance Inflation Factor)* : le VIF est calculé pour chaque variable du modèle de régression, et évalue si les variables sont corrélées les unes aux autres (multicolinéarité). Un VIF supérieur à 10 est considéré comme élevé, et indique une multicolinéarité importante.

b. Construction des échantillons

Des échantillons de commune ont été élaborés afin de réaliser les régressions linéaires et construire des modèles robustes statistiquement. En effet, les communes avec peu d'habitants ou peu denses peuvent se comporter d'une manière très différente des communes densément peuplées, et un seul modèle ne pourrait expliquer toutes les variations de la consommation. Après quelques tests de construction de modèles sur tout le jeu de données, il a été conclu qu'il n'était pas possible de ne construire qu'un seul modèle. Le choix entre deux critères communaux était possible pour faire les échantillons : la densité communale faite par l'INSEE et le nombre d'habitants. Après plusieurs essais avec la densité communale, il a été décidé de se baser sur une classification définie par le nombre d'habitants.

Le tourisme a également été traité à part. En effet, il est possible que les communes touristiques présentent des consommations annuelles plus élevées qui ne seraient pas expliquées par les modèles pour les communes classiques, ce qui introduirait des biais. A partir des données sur les capacités d'accueil en établissements touristiques de l'INSEE, une commune est considérée comme touristique si elle a un nombre de place d'hôtels supérieur à 10 pour 100 habitants, ou un nombre de place de camping supérieur à 10 pour 100 habitants, ou un nombre de place en autre hébergement collectif de tourisme de 10 pour 100 habitants, ou enfin si elle est labellisée « Station classée de tourisme » ou « Commune touristique » par l'INSEE. Les communes touristiques sont ensuite enlevées des échantillons construits pour tester les modèles. Voici les caractéristiques finales des échantillons, en prenant en compte toutes les communes du bassin :

Tableau 25 : caractéristiques des échantillons de communes

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4
Nombre d'habitants	Plus de 10 000	5 000 – 10 000	1 000 – 5 000	Moins de 1 000
Nombre de communes, hors communes touristiques	322	182	1299	5913
Nombre de communes touristiques	23	9	81	187

La carte suivante (Figure 8) montre que les communes de plus de 10 000 habitants sont concentrées en Ile-de-France, ainsi qu'au niveau du Cotentin (Cherbourg...) et des Métropoles de Rouen, du Havre et de Reims. De manière générale, les communes avec une forte population semblent plutôt situées à l'Ouest de Paris.

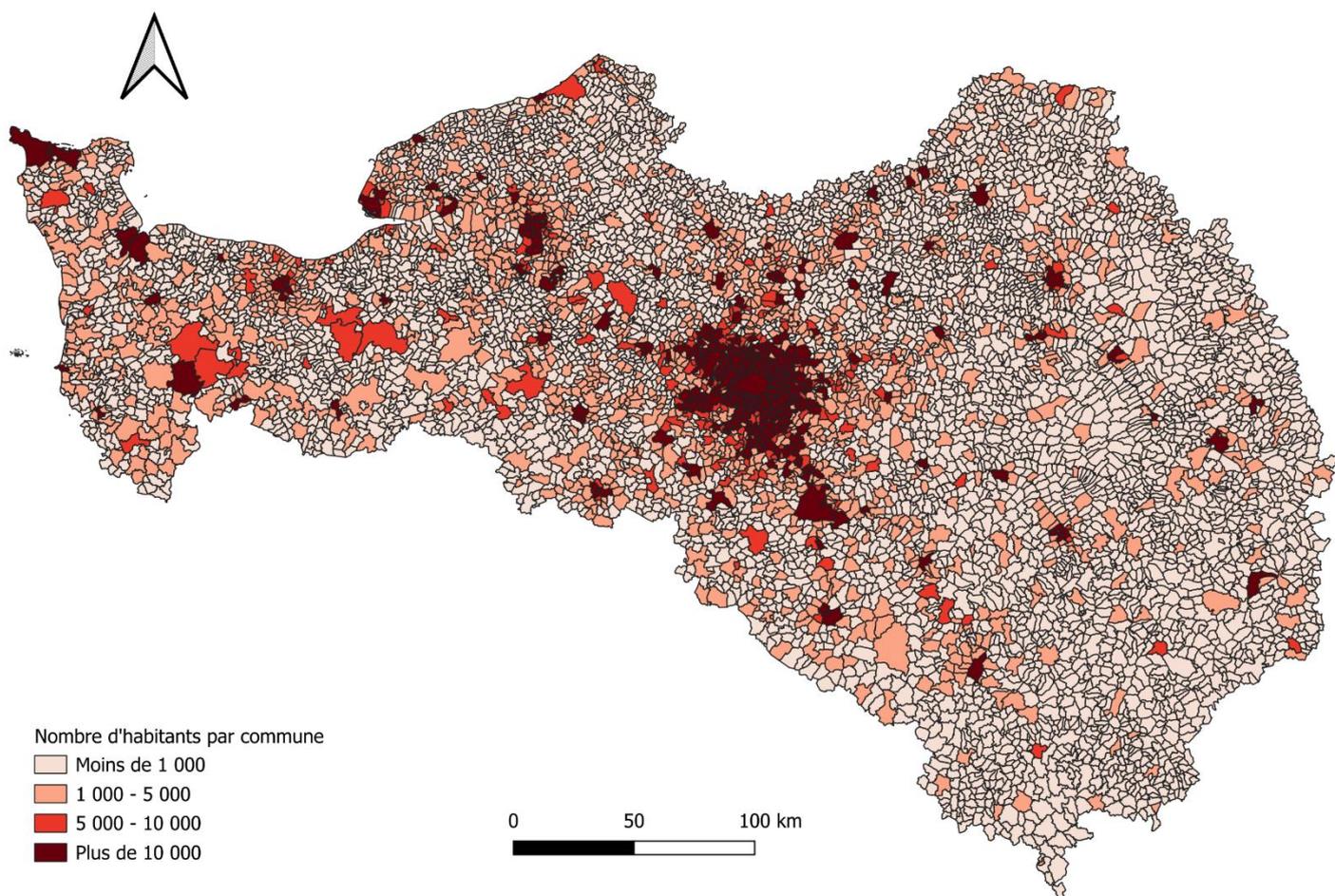


Figure 8 : distribution spatiale des typologies de communes par le nombre d'habitants

c. Variables utilisées

Parmi toutes les données récupérées, après différents tests et dans l'objectif de représenter au maximum les différents secteurs liés à la consommation d'eau potable, un pool de variables explicatives a été sélectionné puis testé sur chaque échantillon. Ces variables sont présentées dans le tableau suivant :

Tableau 26 : liste des variables explicatives utilisées dans les modèles, selon la typologie d'usager

Catégories d'usagers	Variables explicatives
Ménages	Nombre de résidences principales étant des maisons Nombre de résidences principales étant des appartements Nombre de résidences secondaires Nombre de ménages possédant 2 voitures ou plus Taille moyenne des ménages
Tertiaire marchand	Nombre d'établissement du secteur tertiaire marchand Nombre d'emplois du secteur tertiaire marchand
Administration publique	Nombre d'établissement du secteur de l'administration publique Nombre d'emplois du secteur de l'administration publique
Enseignement	Nombre d'établissement du secteur de l'enseignement Nombre d'emplois du secteur de l'enseignement Nombre d'élèves du primaire et du secondaire Nombre d'élèves de l'enseignement supérieur
Santé	Nombre d'hôpitaux ou d'EHPAD Nombre d'emplois des hôpitaux ou des EHPAD Nombre d'établissement du secteur de la santé et de l'action sociale Nombre d'emplois du secteur de la santé et de l'action sociale Nombre de lits en secteur hospitalier
Equipements sportifs	Nombre d'équipements sportifs avec douches et sanitaires Surface totale des terrains de grands jeux Surface totale des bassins de natation Nombre de bassins de natation
Industrie	Nombre d'établissement du secteur de l'industrie Nombre d'emplois du secteur de l'industrie
Agriculture - élevage	Nombre d'établissement du secteur de l'agriculture et de la sylviculture Nombre d'emplois du secteur de l'agriculture et de la sylviculture Nombre d'UGB bovins Nombre d'UGB porcins SAU totale
Services de nettoyage de bâtiments et d'entretien des espaces verts	Nombre d'établissement du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts Nombre d'emplois du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts

Certaines variables n'ont pas été utilisées, soit parce qu'elles ne fonctionnaient pas dans les modèles, soit parce qu'elles expliquaient moins bien les variations de consommation ou alors parce qu'elles n'étaient pas adaptées.

C'est par exemple le cas du revenu médian pour lequel Favre & Montginoul (2020) ont montré, au niveau des Iris, une élasticité de 0,3 à 0,4. C'est-à-dire qu'une augmentation des revenus de 10 % conduit à une augmentation de la consommation de 3 à 4 % (Favre & Montginoul, 2020). Le revenu médian fonctionne dans certains cas dans les modèles. Cependant, de la même manière qu'en ajoutant des variables du type moyenne ou médiane, comme la taille moyenne des ménages par exemple, une augmentation très significative de la constante est observée que nous ne savons pas comment l'interpréter.

Les variables sur le nombre de ménages d'une seule personne, le nombre de résidences principales occupées par des propriétaires et celles liées à l'ancienneté des logements n'ont pas été utilisées non plus. En effet, elles étaient extrêmement corrélées aux variables « Nombre de résidences principales étant des maisons » et « Nombre de résidences principales étant des appartements », qu'il semblait plus pertinent d'utiliser.

d. Autres variables potentiellement intéressantes

D'autres variables auraient pu être intégrées aux modèles. Par exemple, le climat peut influencer la consommation de l'eau. D'une part, car plus la pluviométrie est élevée plus la consommation d'eau aurait tendance à diminuer, du fait notamment des usages extérieurs comme l'arrosage des jardins. Et des températures élevées auraient également tendance à entraîner une augmentation de la consommation de l'eau, avec une augmentation de la fréquence des douches ou de la consommation d'eau de boisson par exemple (Neverre & Rinaudo, 2010). Les données climatiques n'ont finalement pas été utilisées dans les modèles car il aurait été plus adapté d'avoir une chronique journalière des consommations d'eau potable, pour montrer que la consommation serait plus élevée lors des jours les plus chauds par exemple. En effet, les températures moyennes annuelles varient peu d'une année à l'autre, encore plus sur des communes spatialement proches donc nous pensons qu'au pas de temps annuel disponibles dans les données d'eau potable, l'impact des températures devrait être peu visible.

Le prix de l'eau potable est également un facteur explicatif qui ressort régulièrement dans la littérature. L'étude de Neverre & Rinaudo (2010) par exemple montre une élasticité de la consommation par rapport au prix de près de -0,2. En analysant les données fournies par SISPEA sur le prix de l'eau par commune, il a été constaté qu'elles n'étaient pas complètes à l'échelle de tout le bassin. Le prix de l'eau n'est donc pas intégré dans les modèles.

Enfin, il aurait également été intéressant d'avoir des données sur l'occupation des sols notamment pour estimer la surface d'espaces verts ou de jardin, afin de modéliser l'arrosage des espaces verts. Il existe des données cartographiques de ce type mais uniquement sur l'Île-de-France, et les données de Corine Land Cover n'ont pas semblé assez précises. De plus, c'est notamment le cas pour Paris, certaines communes n'utilisent pas d'eau potable pour l'entretien des espaces verts.

e. Méthodes de recherche des modèles

Avec toutes les variables à notre disposition, le nombre de combinaisons possibles est trop grand pour toutes les tester. Deux méthodes principales ont été utilisées pour trouver les modèles les plus pertinents, sans avoir à tester toutes les combinaisons possibles.

- Méthode itérative

Des modèles ont d'abord été construits à l'aide d'une méthode itérative. C'est-à-dire que l'on a d'abord construit un modèle avec toutes les variables, puis la variable qui aura sa p-value la plus élevée, et qui est donc la moins significative, est enlevée. Le modèle est ensuite relancé pour enlever une nouvelle fois la variable la moins significative etc. Ceci jusqu'à obtenir un modèle avec uniquement des variables significatives.

- Fonction « regbest »

Une autre méthode est d'utiliser la fonction « regbest » du package « leaps » sur Rstudio. En rentrant la variable de consommation d'eau et les variables explicatives dans la fonction, elle

va construire le meilleur modèle linéaire possible en fonction du nombre de variables explicatives, puis proposera le meilleur modèle final. C'est-à-dire qu'avec les 31 variables présentées dans le

Tableau 26, la fonction propose le meilleur modèle avec deux variables, le meilleur avec trois, jusqu'au meilleur modèle avec 31 variables. Pour définir ce meilleur modèle, la méthode se base sur le coefficient de détermination en choisissant celui avec le R^2 le plus élevé à chaque fois. Enfin, le meilleur modèle final sera celui qui, parmi les modèles choisis à chaque nombre de variables, aura sa probabilité critique du test du R^2 la plus faible. C'est-à-dire le modèle qui aura sa probabilité d'avoir un R^2 égal à zéro la plus faible. Le meilleur modèle final n'est donc pas forcément celui qui a le R^2 le plus élevé ou un nombre de variables élevés. En effet, il est possible de constater qu'en ajoutant de plus en plus de variables, le coefficient de détermination avait tendance à augmenter, car ce que l'on appelle du « bruit de mesure » est modélisé. L'objectif a donc été de chercher à être parcimonieux dans le nombre de variables à intégrer aux modèles.

Pour chaque échantillon, les modèles construits à l'aide de la méthode itérative et de la fonction « regbest » sont présentés en Annexe II : modèles obtenus par la fonction « regbest » et la méthode itérative pour chaque typologie de commune.

Les modèles sont finalement choisis en se basant sur ces deux méthodes, tout en modifiant les choix afin de respecter les critères suivants :

- Le modèle doit permettre de différencier les différents usagers. Chaque variable doit être rattachable à un type d'utilisateur uniquement et sa consommation associée ainsi modélisée ne peut pas être négative. Dans certains cas, si les modèles présentent des coefficients négatifs, alors l'indicateur associé est exclu car la consommation projetée pour cette variable sera négative, même si la consommation d'eau totale Y est correctement prédite. En revanche le modèle peut potentiellement perdre en précision en faisant cela.
- Les résultats ne doivent pas être incohérents avec ce qui est connu du monde de l'eau. Par exemple, pour les communes de plus de 10 000 habitants la fonction RegBest a proposé un modèle avec la variable « Nombre d'élèves du primaire et du secondaire ». En réalisant les projections avec ce modèle, la consommation d'eau associée à cette variable donne un résultat très élevé, qui représente entre 10 et 20 % de la consommation d'eau totale. Que le secteur de l'enseignement représente une aussi grosse part de la consommation d'eau potable ne paraît pas cohérent. En remplaçant cette variable par la variable « Nombre d'employé du secteur de l'enseignement », le résultat obtenu semble bien plus cohérent bien que le modèle soit un peu moins précis.
- Les modèles doivent respecter au mieux les hypothèses nécessaires à la régression, notamment la non-colinéarité des variables et l'homoscédasticité. En ce qui concerne cette dernière hypothèse par exemple, la distribution des résidus, c'est-à-dire les erreurs de prédiction des modèles, peut être observée à l'aide d'un graphique Residuals vs Fitted. Des résidus avec des valeurs trop élevées par rapport aux autres ont cependant parfois été constatés. Comme la régression linéaire cherche à minimiser la somme des carrés des erreurs du modèle, les observations ayant des résidus très

élevés contribuent donc fortement aux modèles. Dans certains cas, les communes correspondant à ces résidus ont donc été supprimées pour reconstruire un modèle ensuite. La suppression des résidus est tout de même limitée pour ne pas manquer certains phénomènes, ce qui fait que les critères statistiques d'évaluation de l'homoscédasticité ne sont pas parfaitement respectés dans les modèles. Cependant, la littérature indique que cette condition n'est pas forcément rédhibitoire pour la validité de ceux-ci.

3. Projections

L'objectif des projections est de calculer la valeur Y , la consommation d'eau potable, sur les communes où cette donnée n'est pas disponible mais où des données communales, c'est-à-dire les variables explicatives, le sont. L'autre objectif est de décomposer le calcul. C'est-à-dire que l'objectif n'est pas uniquement d'obtenir la valeur projetée de Y , mais également la valeur projetée associée à chaque variable explicative, et donc à chaque segment d'utilisateurs. Les projections doivent également prendre en compte les erreurs associées à nos modèles, ce qui aboutit au calcul de marges d'erreur, dans laquelle la valeur Y projetée est a priori réellement comprise. Pour faire les projections, il est nécessaire de construire un nouveau jeu de données avec les données liées aux variables explicatives pour ces communes, et appliquer les valeurs des modèles pour calculer Y .

- Calcul de la projection ajustée

Le calcul de la projection ajustée est simple, il suffit de multiplier les coefficients obtenus pour chaque modèle avec les données communales. Cette méthode permet également d'estimer la consommation associée à chaque variable explicative. Par exemple, pour la variable « Nombre d'emploi dans le secteur industriel », si le coefficient est égal à 30 et la valeur de l'indicateur est égale à 100, alors la consommation associée à cet indicateur est égale à 3 000 m³. Les projections sont effectuées sur toutes les communes de chaque échantillon, même celles pour lesquelles des données d'eau potable avaient été fournies.

- Calcul des bornes basses et hautes de la projection

Pour calculer les bornes basses et hautes entre lesquelles la valeur de Y sera comprise, la fonction « predict » du package de base « stats » sur Rstudio a été utilisée. En rentrant les nouvelles données liées aux variables explicatives et le modèle, la fonction calcule les valeurs basses, ajustées et hautes de Y par commune. Une subtilité de la fonction permet également de préciser si le calcul doit être effectué en utilisant un intervalle de confiance ou un intervalle de prédiction. L'intervalle de prédiction sera bien plus large car il prend en compte non seulement l'incertitude liée au modèle et à l'échantillonnage, mais également la variabilité des nouvelles observations, pour une observation individuelle.

L'intervalle de prédiction a été utilisé pour les communes où il n'y avait pas de données d'eau potable, car il est plus adapté pour faire des projections. En revanche, les consommations théoriques ont également été calculées par les modèles pour les communes pour lesquelles des données avaient été recueillies, en utilisant l'intervalle de confiance.

Le problème lié à la fonction « predict » est qu'elle ne permet pas de décomposer le calcul pour chaque variable explicative. Une valeur de Y basse et haute est calculée, mais on ne

peut pas remonter aux valeurs basses et hautes pour chaque variable, car les intervalles de prédictions et de confiance sont calculés à partir du modèle global et non indicateur par indicateur. La solution a été de considérer que ces intervalles sont constants pour chaque usager : si la consommation d'eau potable totale est connue avec une erreur de 10 % entre les marges, alors on estime que le sous-total attribué aux ménages est connu à 10 %, de même pour l'industrie etc. Cette méthode a des limites car chaque variable ne participe pas de manière égale à l'erreur du modèle.

Pour chaque indicateur, une consommation ajustée, une consommation basse et une consommation haute ont donc été calculées. Pour calculer la part que représente la consommation avec l'hypothèse basse pour chaque indicateur, cette valeur a été rapportée sur la somme des consommations avec l'hypothèse haute pour les autres indicateurs. Et inversement pour calculer la part que représente l'hypothèse haute.

Limites et axes d'amélioration

Comme évoqué, cette étude est une étude exploratoire, ayant l'objectif de tester une nouvelle approche de l'utilisation de modèles statistiques pour expliquer et prédire la consommation en eau potable. Cette partie a donc pour objectif de recenser les limites de notre étude, certaines étant déjà évoquées au long de ce rapport, et les points d'améliorations possibles.

1. Biais des données

Tout d'abord, les données d'eau potable fournies peuvent induire de nombreux biais. Il y a d'une part la question de savoir si les volumes comptabilisés au niveau des compteurs d'eau sont fiables. En effet, les compteurs d'eau peuvent avoir des erreurs de mesure, menant à des données inexactes, et notamment les compteurs anciens ou mal entretenus. Ensuite, les données provenant de volumes facturés, le report d'une année sur l'autre des facturations que l'on a pu observer dans les données de Veolia et SUEZ peut également poser problème. Pour Veolia par exemple, il a été choisi de supprimer les communes pour lesquelles il y avait un volume négatif d'indiqué dans les données. Mais pour les autres communes il y a peut-être des reports que l'on n'a pas pu voir, si les volumes sont à la fin toujours positifs.

Enfin, il a été mis en évidence que certaines communes sont alimentées en eau potable par plusieurs fournisseurs dans certains cas. Mais il peut y en avoir d'autres, comme le recensement des fournisseurs n'a pas été exhaustif, notamment sur des communes rurales. Dans ce cas-là, s'il n'y qu'un fournisseur recensé le volume indiqué sera donc sous-estimé. Malgré tout, pour un périmètre aussi large que celui d'un bassin versant recenser tous les fournisseurs d'eau potable est compliqué et chronophage.

En ce qui concerne les données communales, elles sont en principe assez fiables, et la richesse des données provenant de l'INSEE, de l'étude socio-économique et des recensements des établissements de santé et sportifs, a permis d'obtenir de nombreuses variables différentes pour construire nos modèles. En revanche, alors que les données d'eau potable utilisées datent de 2022, étant donné que Veolia n'a fourni qu'une année, les données communales ne datent pas forcément de la même année. Les données de population de

l'INSEE par exemple datent de 2020. Il serait plus rigoureux d'harmoniser l'année des données même si les données de l'INSEE ne varient pas forcément beaucoup d'une année à l'autre.

2. Qualité des modèles

Bien que nos modèles soient assez robustes, ils présentent tout de même certains défauts. En effet, ils sont assez variables, les hypothèses de la régression sont difficilement respectées et certains secteurs consommateurs d'eau ne sont pas expliqués.

a. Précision des modèles

Comme cela a déjà été expliqué, les modèles sont moins précis sur les communes faiblement peuplées. Les communes densément peuplées peuvent globalement présenter des infrastructures similaires et donc un comportement général de consommation de l'eau également similaire. Mais dans les communes rurales, la variabilité des consommations peut être plus élevée à cause de la présence d'infrastructures consommatrices d'eau ou d'une concentration élevée d'habitats sur une commune, qui ne sera pas forcément représentatif des autres communes. Cela questionne donc la pertinence d'un modèle unique pour ces communes. Une solution aurait pu être de construire des sous-échantillons pour ces communes, par exemple en différenciant les communes de moins de 5 000 habitants urbaines et rurales, afin d'améliorer les modèles.

b. Hypothèses de la régression linéaire

En lien avec leur précision, les modèles ne respectent pas tout à fait les conditions d'homoscédasticité. Il y a deux moyens d'évaluer cette condition : une méthode d'analyse graphique des résidus, pas forcément très précise, mais qui a été conseillé par un statisticien, et une méthode statistique à l'aide du test de Breusch-Pagan. Bien que les graphiques sur la distribution des résidus soient satisfaisants (Annexe III : graphiques des résidus, résultats des tests de Breusch-Pagan et VIF calculés pour les modèles retenus), le test montre une hétéroscédasticité dans nos modèles. Elle est faible dans celui pour les communes de plus de 10 000 habitants, mais plus élevée pour ceux sur les communes de moins de 5 000 habitants d'après les tests, et même très élevée pour les communes de moins de 1 000 habitants. Le modèle sur les communes de 5 000 à 10 000 habitants ne présente pas d'hétéroscédasticité. Cependant, l'hétéroscédasticité n'est pas forcément fatale pour l'interprétation des modèles, mais les tests de Student et de Fischer permettant d'estimer la significativité des coefficients dans les modèles peuvent être faussés. La question est de savoir à quel niveau il est possible de considérer qu'elle n'est pas trop critique, et si les deux derniers modèles pourraient être dans ce cas.

Pour pallier ces problèmes, une solution a été de supprimer des résidus aberrants, ce qui a pour effet de réduire l'hétéroscédasticité. Mais pour ne pas supprimer trop d'observations cette solution a été limitée à seulement quelques résidus, et l'hétéroscédasticité a donc été seulement réduite. Cette limite est liée aux caractéristiques très variables des communes faiblement peuplées et est donc en lien avec la précision des modèles, et de la même manière, des sous-échantillons auraient pu être construits pour ces communes. Sinon, tester des méthodes de transformation des données pourrait aussi être nécessaire. L'étude de N. Neverre & Rinaudo (2010) propose quatre différentes méthodes de transformation de données : logarithme-linéaire, logarithme-logarithme, semi-logarithme et mixte.

Les résultats des VIF (Annexe III : graphiques des résidus, résultats des tests de Breusch-Pagan et VIF calculés pour les modèles retenus), estimant la multi-colinéarité des variables

des modèles sont bons, hormis dans deux cas. Le premier cas est dans le modèle des communes de plus de 10 000 habitants, pour lequel les variables « Nombre d'établissements du secteur tertiaire marchand » et « Nombre de résidences principales étant des appartements » ont des valeurs de VIF égaux à 11 environ, ce qui est élevé. Ce VIF élevé indique donc une multi-colinéarité principalement entre ces deux variables, qui implique une augmentation de la variance des coefficients et qui diminue donc la fiabilité des valeurs des coefficients calculés. Le fait que ces deux variables soient fortement corrélées peut s'expliquer par exemple par le fait que de nombreux commerces sont situés en bas d'immeubles d'habitation dans les grandes villes. Cette colinéarité avait été identifiée mais il a été choisi de finalement garder ces variables dans les modèles, car elles sont essentielles pour ces communes. En effet, ne pas montrer de consommation liée au secteur des commerces et du tertiaire, ou au nombre de résidences principales étant des appartements dans les communes de plus de 10 000 habitants aurait été problématique, et aurait diminué l'intérêt de cette étude

L'autre cas problématique concerne le modèle pour les communes de moins de 1 000 habitants. Les variables « Nombre de ménages possédant 2 voitures ou plus » et « Nombre de résidences principales étant des maisons » sont fortement corrélées, et leurs VIF sont d'environ 21. Dans ce cas-là, la valeur du VIF est trop élevée en théorie mais le choix a également été fait de conserver les deux variables. Bien que la suppression de l'un des deux indicateurs n'entraîne pas une perte significative d'informations, car les deux variables sont liées au secteur domestique, elle réduit toutefois la précision du modèle sur des communes où la précision des modèles n'est déjà pas optimale.

Ces observations indiquent donc que les coefficients associés aux variables sont parfois à interpréter avec précaution.

c. Variabilité et fonctionnalité des modèles

Comme il est possible de le voir avec les modèles présentés en Annexe II : modèles obtenus par la fonction « regbest » et la méthode itérative pour chaque typologie de commune, obtenus par une méthode itérative et à l'aide de la fonction « regbest », ils peuvent être très variables au niveau des indicateurs. Les modèles peuvent donc grandement varier en termes d'indicateurs et de coefficients associés, sur les consommations projetées par secteur associées également, même s'ils sont finalement assez peu sensibles sur les consommations totales projetées.

De plus, les modèles finalement choisis semblent surestimer certains secteurs (santé et enseignement notamment) mais sous-estimer la consommation associée aux équipements sportifs. D'une part parce que certaines variables ne sont pas intégrées à l'analyse, mais également aussi à cause de la variabilité des modèles. Par exemple, le modèle pour les communes de plus de 10 000 habitants obtenus par la méthode itérative, propose l'indicateur « Surface totale des bassins de natation ». Alors que le modèle qui a été choisi, ne propose rien sur les équipements sportifs, ce qui se traduit par une consommation prédite nulle sur ces communes pour les équipements sportifs. Ce qui est surprenant, c'est que si cet indicateur est rajouté dans le modèle choisi pour les communes de plus de 10 000 habitants, il ne sera pas significatif et aura en plus un coefficient négatif. Pour les équipements sportifs, des indicateurs sur le nombre total d'équipements sportifs ou le nombre d'emplois liés aux activités sportives auraient pu être utilisés. Mais ces indicateurs n'ont pas semblé pertinents car certaines activités sportives ne consomment pas d'eau en tant que telles, mais ce sont les équipements

liés à certaines activités qui consomment (arrosage des terrains de grands jeux, bassins de natation, douches et sanitaires).

Il serait nécessaire, même si cela prend plus de temps, d'être plus exhaustif sur la diversité de modèles possibles avec les indicateurs disponibles et de mieux comprendre les relations entre les indicateurs, en lien avec leur colinéarité notamment.

Enfin, les indicateurs qui ont été utilisés sur le nombre d'emplois par secteur sont pertinents pour prendre en compte la taille des établissements. Un nombre plus important d'emplois signifie des établissements plus importants et qui peuvent donc consommer plus d'eau. Mais la relation n'est pas forcément parfaite. Pour le secteur tertiaire par exemple, la consommation d'eau potable est expliquée principalement par la consommation des employés (sanitaire, alimentation...), donc dans ce cas-là la relation avec le nombre d'emplois du secteur est pertinente. En revanche, dans le cas de l'industrie, la consommation est faiblement liée à la consommation des employés, mais plutôt par le type de processus industriel et va donc grandement dépendre de l'activité de l'industrie. Dans ce cas-là, l'utilisation du nombre d'emplois n'est pas très fonctionnelle et présente des limites, alors que l'utilisation du nombre d'industries ne prend pas en compte la taille de l'industrie, et fonctionne rarement dans les modèles. Cette affirmation est également valable pour les secteurs de la santé, de l'agriculture ou des équipements sportifs.

Pour le secteur industriel, l'utilisation de la base de données de l'IREP, qui recense les prélèvements industriels sur le réseau d'alimentation en eau potable, est pertinente pour obtenir des résultats plus fins. En effet, après analyse, il a été conclu que cette base de données était fiable et pouvait donc servir de base pour ce secteur.

d. Omissions de variables pertinentes

De nombreux facteurs déjà évoqués précédemment influençant la consommation de l'eau ne sont pas utilisés dans les modèles, soit par manque de données adéquates (climat, prix de l'eau, arrosage des espaces verts), soit car elles ne fonctionnent pas ou ne sont pas forcément pertinentes pour l'étude (revenu médian, nombre de personnes non scolarisées et non diplômées...). Les modèles ne montrent rien en ce qui concerne l'arrosage des espaces verts, alors que SUEZ indique une consommation pour cet usage d'environ 400 000 m³ pour l'année 2022. Cependant, certaines communes d'Ile-de-France, dont Paris, n'utilisent pas d'eau potable pour cet usage.

e. Temporalité

Une autre limite de l'étude est qu'elle se concentre uniquement sur une année. Avoir des données d'eau potable sur plusieurs années aurait permis de prendre du recul sur les modèles, en construisant des modèles explicatifs sur plusieurs années afin de voir s'ils sont cohérents d'une année à l'autre, ainsi que d'observer des tendances d'augmentation ou de réduction des consommations par usage.

Enfin, il serait intéressant d'obtenir des données de consommation d'eau potable à un pas de temps plus fin que le pas de temps annuel, afin de capter les variations de consommation saisonnières du fait d'une activité touristique importante par exemple. Techniquement c'est difficilement réalisable actuellement, mais la tendance est au développement de l'appareil statistique public et de l'Open data, ce qui permettra d'obtenir des données plus précises (Crédoc, 2016).

f. Viabilité des projections

La première limite à ce sujet, évidente, repose sur le postulat des projections qui partent du principe que les relations identifiées dans les zones avec données sont applicables ailleurs. Or cette hypothèse peut ne pas tenir si les conditions locales diffèrent significativement.

La méthode pour calculer les bornes hautes et basses par segment d'utilisateur a également des limites, car l'écart entre la projection ajustée et les bornes hautes et basses est appliqué sur chaque segment de manière équivalente. Alors qu'il est fortement probable que les segments d'utilisateur ne participent pas de manière équivalente à la variabilité expliquant ces bornes.

Annexes

Annexe I : typologies fournies par les différents services d'eau

VEDIF 4 postes	Collectivités	Gestionnaires immeubles	Particuliers	Professionnels
-----------------------	---------------	-------------------------	--------------	----------------

VEDIF 22 postes
Ambassade
Appareil public
Appartement (AI-SRU)
Bâtiment industriel
BI-PI public
Borne fontaine
Bureaux
Chantier
Cimetière
Commerce
Enseignement
Equipement incendie prive
Equipement sportif
Espace vert
Etablissement de sante
Groupe d'immeubles
Immeuble
Mairie
Pavillon
Piscine
Terrain agricole
Nettoyage voirie

SENEO	Administration 2022	Collectivité 2022	Particulier 2022	Professionnel 2022	Syndic (gestionnaire d'immeuble) 2022
--------------	---------------------	-------------------	------------------	--------------------	---------------------------------------

SUEZ
Administration
Agriculteur
Client de passage (forain, navire, ...)
Collectivité

Intra groupe
Particulier
Professionnel
Syndic (gestionnaire d'immeuble)

Veolia
Appareils publics
Bâtiments communaux
Chasse sur réseau assainissement
Collectifs
Espaces verts
Essais PI/BI
Fournitures temporaires
Indéterminé
Individuels
Individuels (SRU)
Industriels
Irrigations et agricoles
Lavage de voirie

Annexe II : modèles obtenus par la fonction « regbest » et la méthode itérative pour chaque typologie de commune

Tableau 1 : modèle explicatif obtenu par la fonction « regbest » pour les communes de plus de 10 000 habitants et caractéristiques de base du modèle

Communes de plus de 10 000 habitants	
Variables explicatives	Coefficients
Constante	-938899,693***
Nombre d'emplois dans le secteur industriel	34,551***
Nombre de lits en secteur hospitalier	300,207***
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	95,149***
Nombre d'élèves du primaire et du secondaire	55,346***
Nombre de résidences principales étant des maisons	69,522***
Nombre de résidences principales étant des appartements	87,585***
Nombre d'établissement du secteur de l'administration publique	5536,924***
Nombre d'établissements du secteur tertiaire marchand	34,657***
Taille moyenne des ménages	378058,387***
Nombre d'employé du secteur de la santé et de l'action sociale	-173,668**

	Modèle Regbest
Taille de l'échantillon	184
R ² ajusté	0,98
F-test (p-value)	< 2,2204E-16

Erreur standard résiduelle	186 372,2
Résidus enlevés	0

Tableau 2 : modèle explicatif obtenu par la méthode itérative pour les communes de plus de 10 000 habitants et caractéristiques de base du modèle

Communes de plus de 10 000 habitants	
Variables explicatives	Coefficients
Constante	-755552,742***
Nombre d'emplois dans le secteur industriel	41,549***
Nombre de lits en secteur hospitalier	312,9***
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	80,306***
Nombre d'élèves du primaire et du secondaire	52,022***
Nombre de résidences principales étant des maisons	75,188***
Nombre de résidences principales étant des appartements	82,344***
Nombre d'établissement du secteur de l'administration publique	3938,502***
Nombre d'établissements du secteur tertiaire marchand	51,674***
Taille moyenne des ménages	312506,389***
Nombre d'établissements du secteur de la santé et de l'action sociale	-587,167***
Nombre d'employé du secteur de l'agriculture et de la sylviculture	-1770,909*
Surface totale des bassins de natation	73,788**

	Méthode itérative
Taille de l'échantillon	184
R ² ajusté	0,982
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	182 261,8
Résidus enlevés	0

Tableau 3 : modèle explicatif obtenu par la fonction « regbest » pour les communes de 5 000 à 10 000 habitants et caractéristiques de base du modèle

Communes de 5 000 à 10 000 habitants	
Variables explicatives	Coefficients
Constante	58836,632*
Nombre d'élèves de l'enseignement supérieur	32,589 ***
Nombre de résidences principales étant des maisons	60,383***
Nombre de résidences principales étant des appartements	67,632***
Nombre d'établissement du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts	-1128,563*

Nombre d'emplois dans le secteur tertiaire marchand	15,682 ***
Nombre d'établissements du secteur tertiaire marchand	109,834***

	Modèle Regbest
Taille de l'échantillon	56
R ² ajusté	0,85
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	39 105
Résidus enlevés	0

Tableau 4 : modèle explicatif obtenu par la méthode itérative pour les communes de 5 000 à 10 000 habitants et caractéristiques de base du modèle

Communes de 5 000 à 10 000 habitants	
Variabiles explicatives	Coefficients
Constante	58836,632*
Nombre d'élèves de l'enseignement supérieur	32,589 ***
Nombre de résidences principales étant des maisons	60,383***
Nombre de résidences principales étant des appartements	67,632***
Nombre d'établissement du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts	-1128,563*
Nombre d'emplois dans le secteur tertiaire marchand	15,682 ***
Nombre d'établissements du secteur tertiaire marchand	109,834***

	Méthode itérative
Taille de l'échantillon	56
R ² ajusté	0,85
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	39 105
Résidus enlevés	0

Tableau 5 : modèle explicatif obtenu par la fonction « regbest » pour les communes de 1 000 à 5 000 habitants et caractéristiques de base du modèle

Communes de 1 000 à 5 000 habitants	
Variabiles explicatives	Coefficients
Constante	4113,306
Nombre d'emplois dans le secteur industriel	72,462***
Nombre d'hôpitaux ou d'EHPAD	2144,472**
Nombre d'élèves du primaire et du secondaire	18,332**
Nombre de résidences principales étant des appartements	37,443**
Nombre de ménages possédant 2 voitures ou plus	124,148***

Nombre d'établissements du secteur tertiaire marchand	98,957***
Nombre d'UGB porcins	13,129**
Surface totale des terrains de grands jeux	1,03E-04 ***

	Modèle Regbest
Taille de l'échantillon	327
R ² ajusté	0,808
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	25 931,47
Résidus enlevés	0

Tableau 6 : modèle explicatif obtenu par la méthode itérative pour les communes de 1 000 à 5 000 habitants et caractéristiques de base du modèle

Communes de 1 000 à 5 000 habitants	
Variables explicatives	Coefficients
Constante	3236,851
Nombre d'emplois dans le secteur industriel	76,095***
Nombre d'hôpitaux ou d'EHPAD	2374,604**
Nombre d'élèves du primaire et du secondaire	12,961*
Nombre de résidences principales étant des maisons	31,288*
Nombre de résidences principales étant des appartements	50,224***
Nombre de ménages possédant 2 voitures ou plus	103,829***
Nombre d'UGB bovins	-4,309*
Nombre d'UGB porcins	14,922**
Nombre de résidences secondaires	68,95*
Nombre d'établissement du secteur de l'industrie	-597,006*
Nombre d'emplois dans le secteur tertiaire marchand	21,763***
Nombre d'établissement du secteur de l'enseignement	782,94**
Nombre d'élèves de l'enseignement supérieur	25,585**
Surface totale des terrains de grands jeux	1,06E-04 ***

	Méthode itérative
Taille de l'échantillon	327
R ² ajusté	0,817
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	25 312,11
Résidus enlevés	0

Tableau 7 : modèle explicatif obtenu par la fonction « regbest » pour les communes de moins de 1 000 habitants et caractéristiques de base du modèle

Communes de moins de 1 000 habitants	
Variables explicatives	Coefficients
Constante	1807,165***
Nombre de résidences principales étant des maisons	64,125***
Nombre d'établissements du secteur tertiaire marchand	196,008***

	Modèle Regbest
Taille de l'échantillon	1 102
R ² ajusté	0,764
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	6 654,72
Résidus enlevés	0

Tableau 8 : modèle explicatif obtenu par la méthode itérative pour les communes de moins de 1 000 habitants et caractéristiques de base du modèle

Communes de moins de 1 000 habitants	
Variables explicatives	Coefficients
Constante	1289,465***
Nombre de résidences principales étant des maisons	24,472**
Nombre d'établissements du secteur tertiaire marchand	51,423***
Nombre d'emplois du secteur de l'enseignement	101,072***
Nombre d'emplois des hôpitaux ou des EHPAD	-15,568***
Nombre de lits en secteur hospitalier	185,084***
Nombre d'emplois du secteur de la santé et de l'action sociale	68,764***
Nombre d'emplois du secteur de l'administration publique	-58,115***
Nombre d'emplois du secteur de l'agriculture et de la sylviculture	13,079*
Nombre d'emplois du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts	13,245**
Nombre d'emplois du secteur tertiaire marchand	15,591***
Nombre de résidences secondaires	48,351***
Nombre de résidences principales étant des appartements	45,897**
Nombre de ménages possédant 2 voitures ou plus	92,096***
Surface totale des bassins de natation	-98,32*
Nombre de bassins de natation	12229,201***
Nombre d'emplois du secteur de l'industrie	-5,93**
Nombre d'UGB bovins	2,945***
Nombre d'établissement du secteur de l'enseignement	245,778*
Nombre d'hôpitaux ou d'EHPAD	1729,459***

	Méthode itérative

Taille de l'échantillon	1 102
R ² ajusté	0,852
F-test (p-value)	< 2,2204E-16
Erreur standard résiduelle	5 268, 593
Résidus enlevés	0

Annexe III : graphiques des résidus, résultats des tests de Breusch-Pagan et VIF calculés pour les modèles retenus

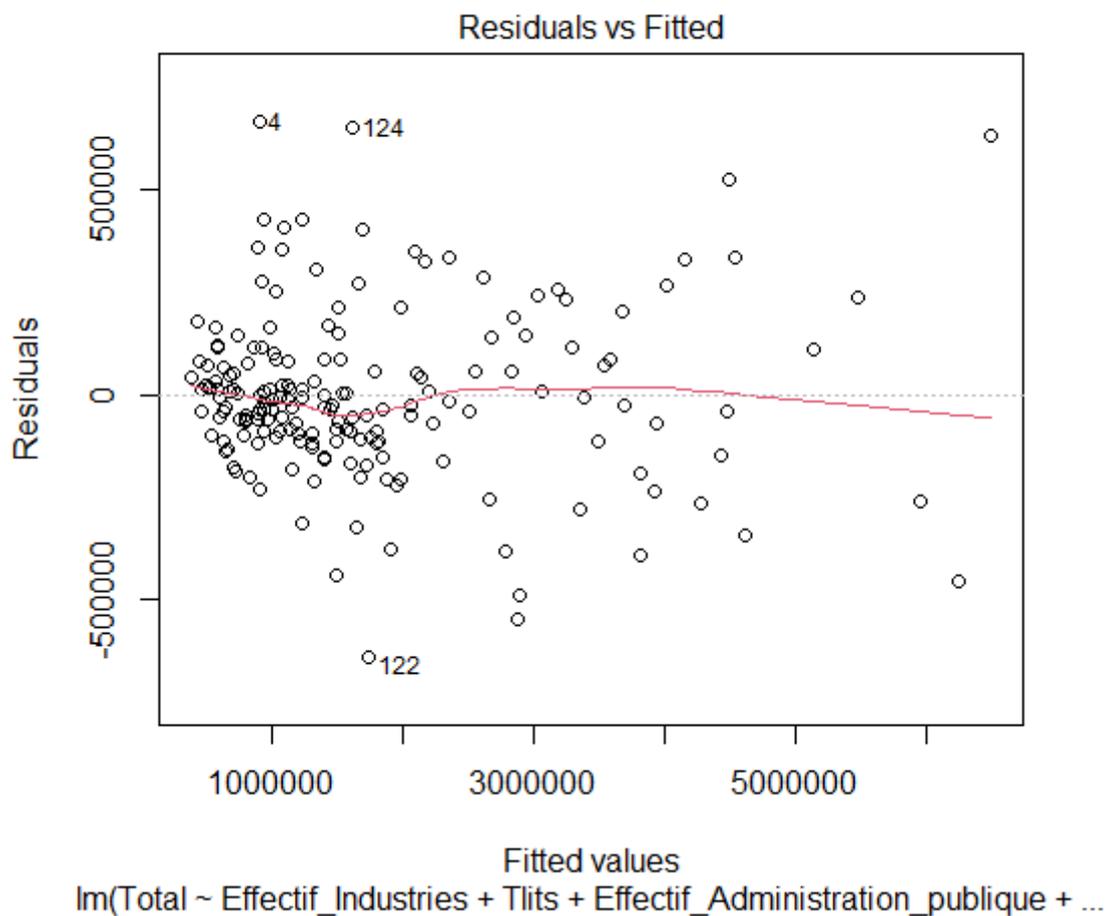
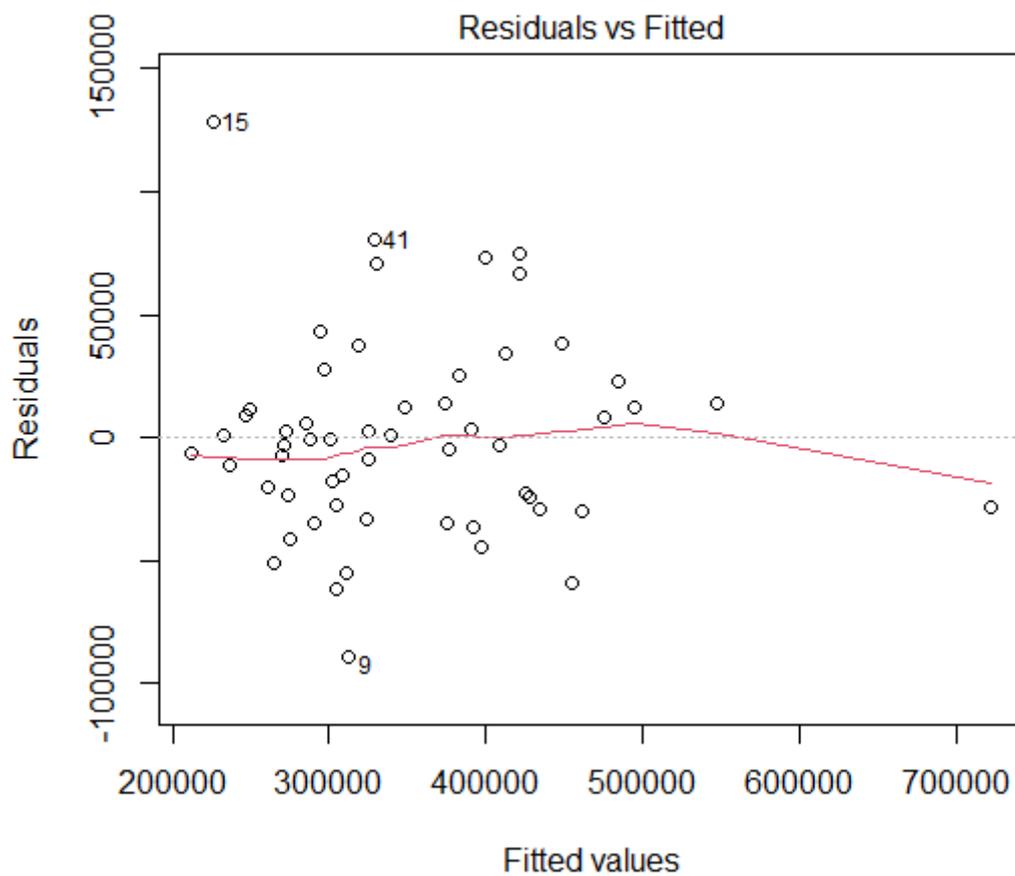
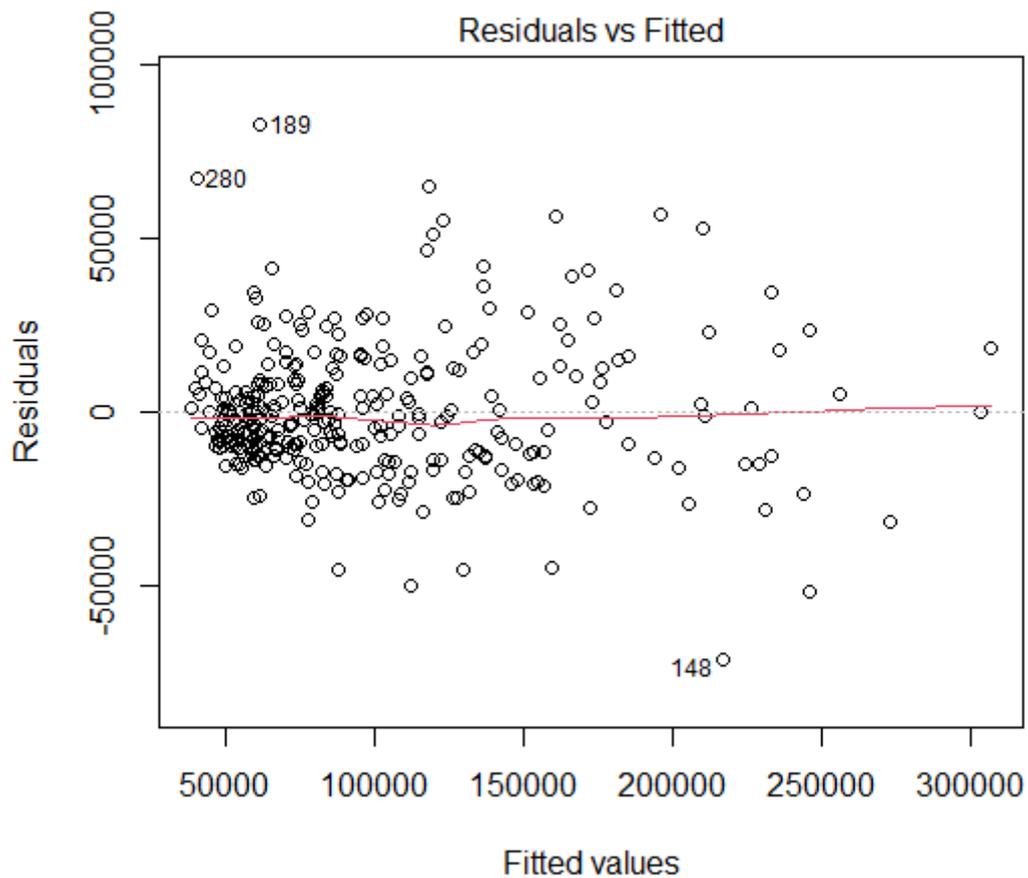


Figure 1 : graphique Residuals vs Fitted pour la typologie de communes de plus de 10 000 habitants



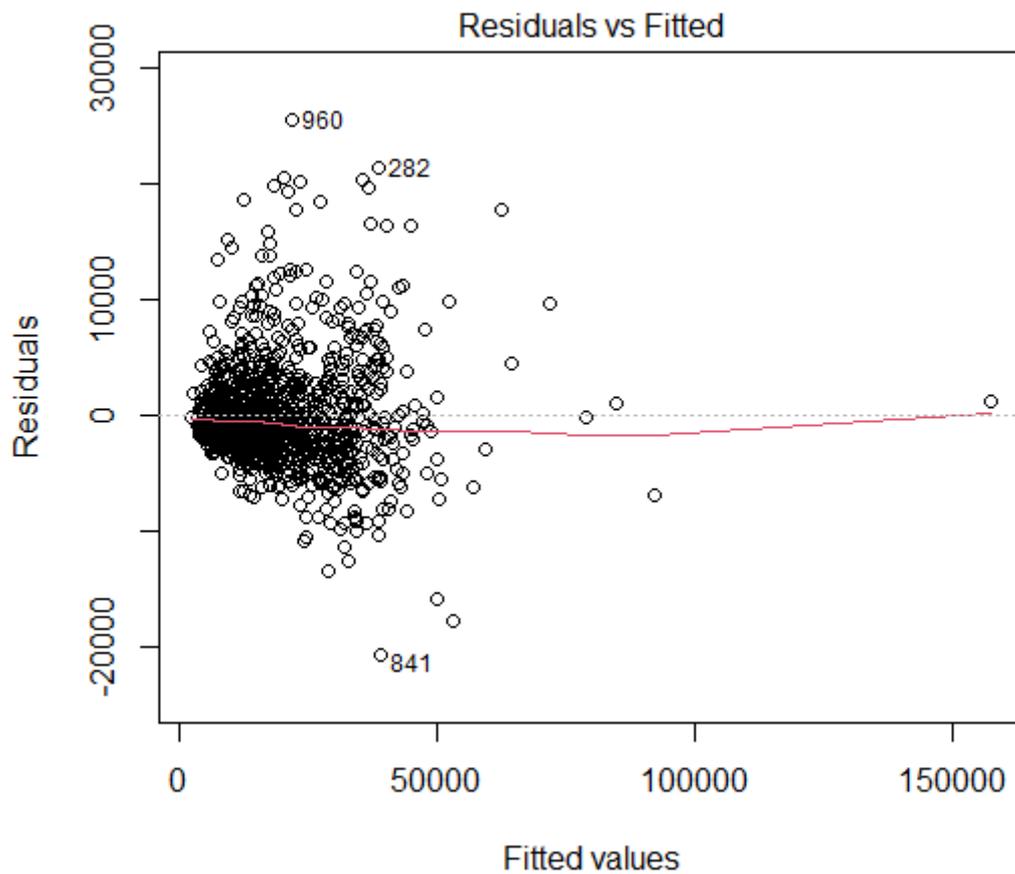
$\text{lm}(\text{Total} \sim \text{Total_effectif_Esuperieur} + \text{Effectif_Tertiaire_marchand} + \text{Nres_p}$

Figure 2 : graphique Residuals vs Fitted pour la typologie de communes de 5 000 à 10 000 habitants



$\text{lm}(\text{Total} \sim \text{Effectif_Enseignement} + \text{Effectif_Tertiaire_marchand} + \text{Nres_princ})$

Figure 3 : graphique Residuals vs Fitted pour la typologie de communes de 1 000 à 5 000 habitants



$\text{lm}(\text{Total} \sim \text{UGB_BOVINS} + \text{Effectif_Enseignement} + \text{Tiits} + \text{Effectif_Sante_ac})$

Figure 4 : graphique Residuals vs Fitted pour la typologie de communes de moins de 1 000 habitants

Tableau 27 : p-value associées au test de Breusch-Pagan, réalisé pour chaque échantillon

	Moins de 1 000 habitants	1 000 – 5 000 habitants	5 000 – 10 000 habitants	Plus de 10 000 habitants
P-value	0,01299583	0,7848986	0,0000007005354	< 2,2204E-16

Plus de 10 000 habitants	VIF
Nombre d'emplois dans le secteur industriel	1,527659158
Nombre de lits en secteur hospitalier	1,357802631
Nombre d'emplois dans le secteur du nettoyage de bâtiment et d'entretien des espaces verts	1,471619170
Nombre d'emplois dans le secteur de l'enseignement	2,823280603
Nombre de résidences principales étant des maisons	1,241399590
Nombre de résidences principales étant des appartements	11,717489546
Nombre d'emplois dans le secteur de l'administration publique	1,930296508
Nombre d'établissements du secteur tertiaire marchand	10,936353358

5 000 – 10 000 habitants	VIF
Nombre d'élèves de l'enseignement supérieur	1,059371434
Nombre de résidences principales étant des maisons	2,658277563
Nombre de résidences principales étant des appartements	2,210618546
Nombre de ménages possédant 2 voitures ou plus	1,751072511
Nombre d'emplois dans le secteur tertiaire marchand	1,058859932
Nombre de bassins de natation	1,499923973

1 000 – 5 000 habitants	VIF
Nombre d'emplois dans le secteur industriel	1,250500973
Nombre d'hôpitaux ou d'EHPAD	1,162770782
Nombre d'emplois dans le secteur de l'enseignement	1,769686262
Nombre de résidences principales étant des appartements	2,007634711
Nombre de ménages possédant 2 voitures ou plus	1,851318667

Nombre d'emplois dans le secteur tertiaire marchand	1,711182494
Nombre d'UGB porcins	1,065552082
Nombre d'équipements sportifs avec douches et sanitaires	1,210878763
Nombre de bassins de natation	1,166652843
Surface totale des terrains de grands jeux	1,081672739

Moins de 1 000 habitants	VIF
Nombre de lits en secteur hospitalier	1,063577907
Nombre d'emplois dans le secteur de la santé et de l'action sociale	1,166454180
Nombre d'emplois dans le secteur de l'enseignement	1,307250591
Nombre de résidences principales étant des maisons	21,507746649
Nombre de résidences secondaires	1,452756881
Nombre de résidences principales étant des appartements	1,852177367
Nombre de ménages possédant 2 voitures ou plus	21,049813019
Nombre d'emplois dans le secteur tertiaire marchand	1,176119093
Nombre d'UGB bovins	1,079269450
Nombre d'employé du secteur de l'agriculture et de la sylviculture	1,058249782

Annexe IV : spatialisation des consommations projetées pour chaque segment d'utilisateur

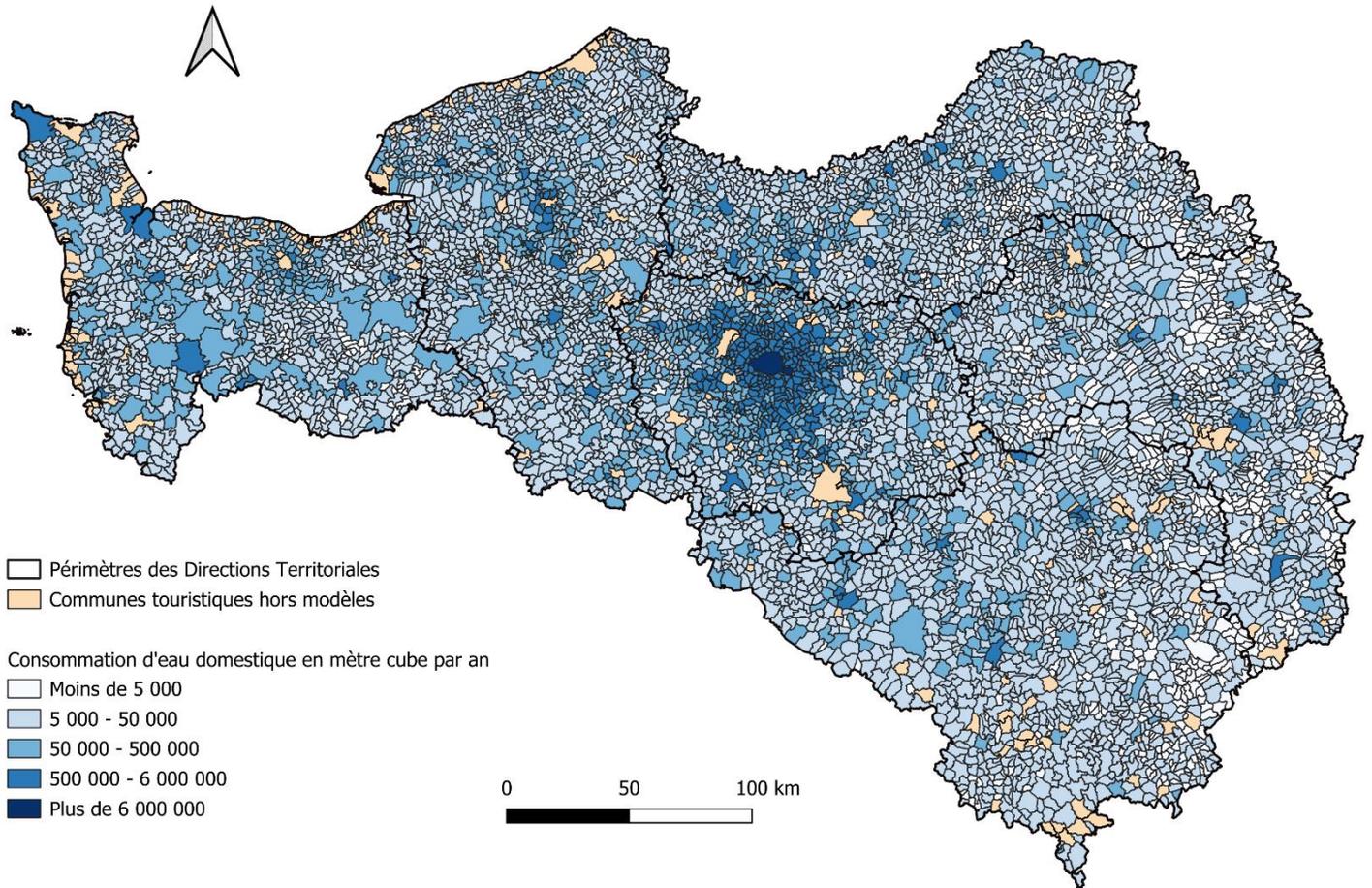


Figure 1 : consommation d'eau domestique projetée, par commune et à l'échelle du bassin versant

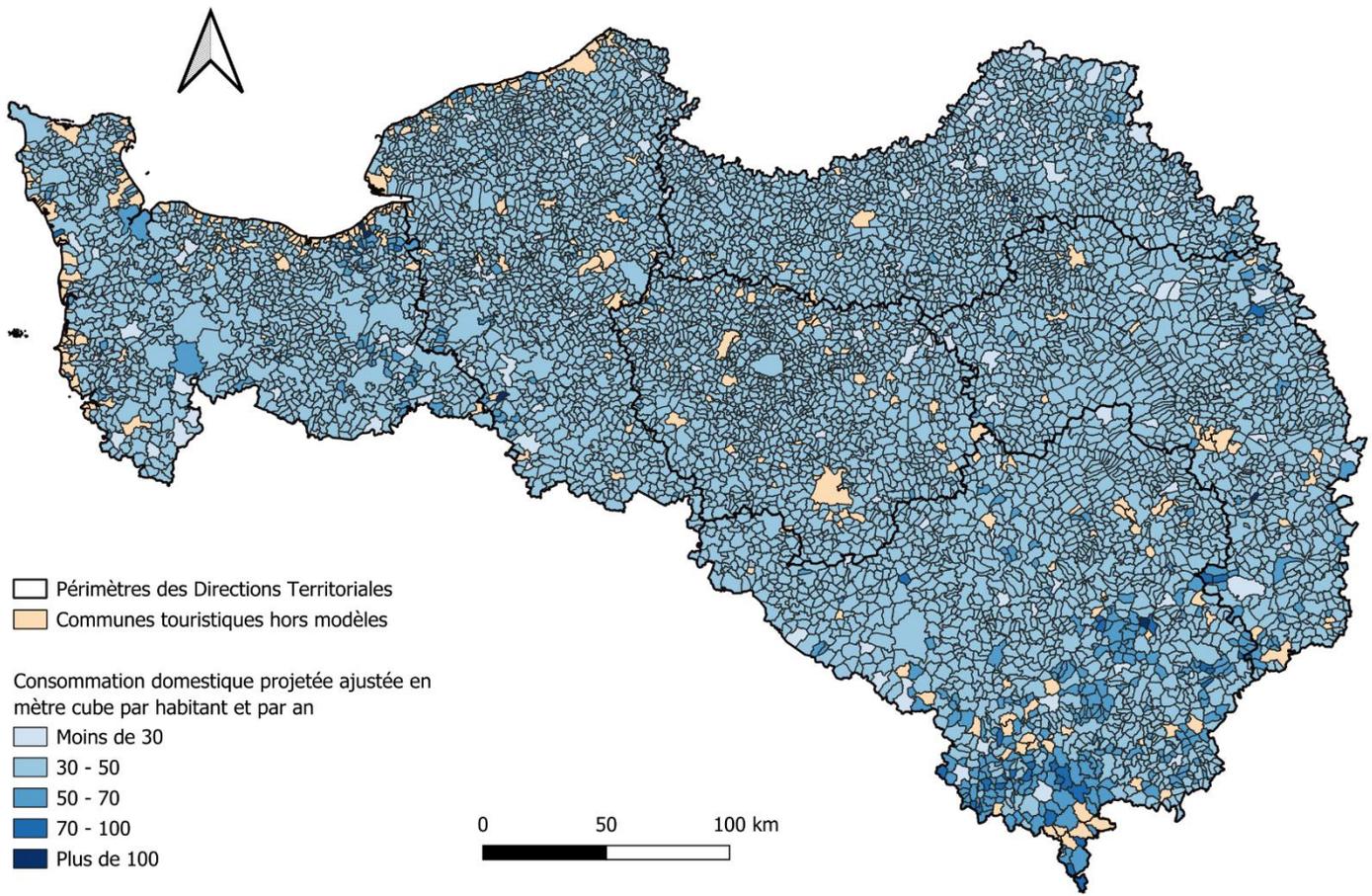


Figure 2 : consommation d'eau domestique rapportée au nombre d'habitants, par commune et à l'échelle du bassin versant

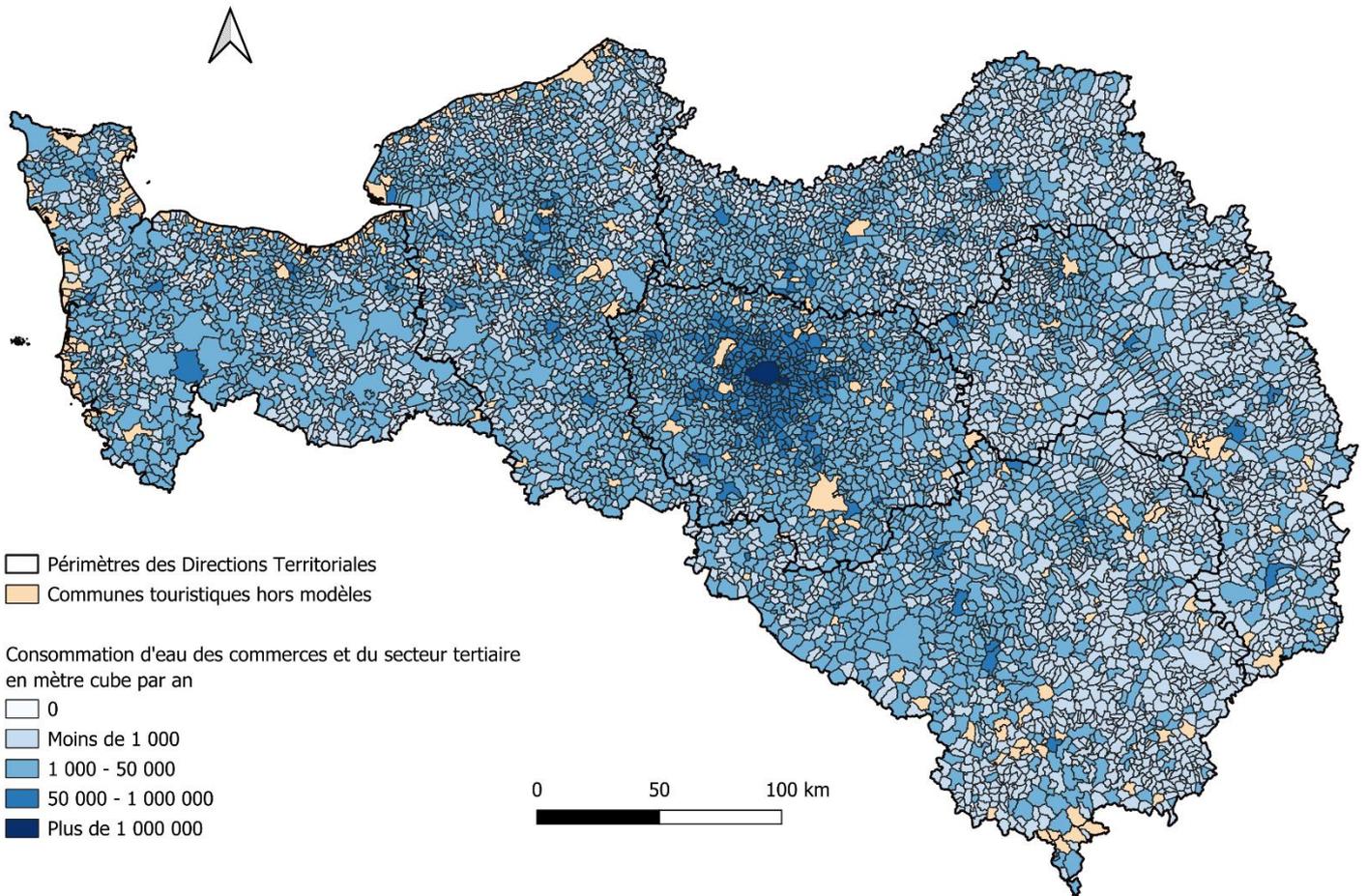


Figure 3 : consommation d'eau des commerces et du secteur tertiaire projetée, par commune et à l'échelle du bassin versant

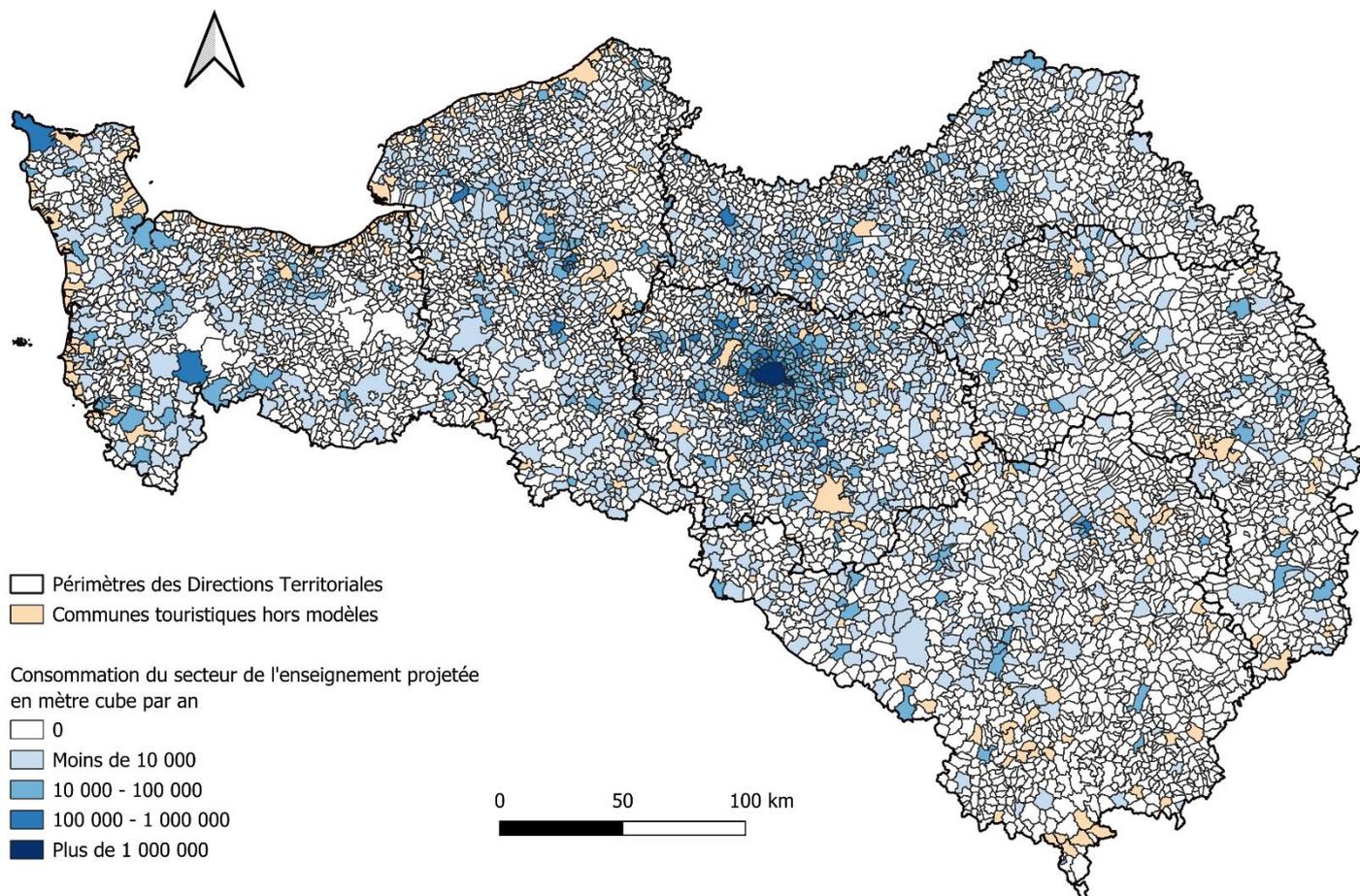


Figure 4 : consommation d'eau du secteur de l'enseignement projetée, par commune et à l'échelle du bassin versant

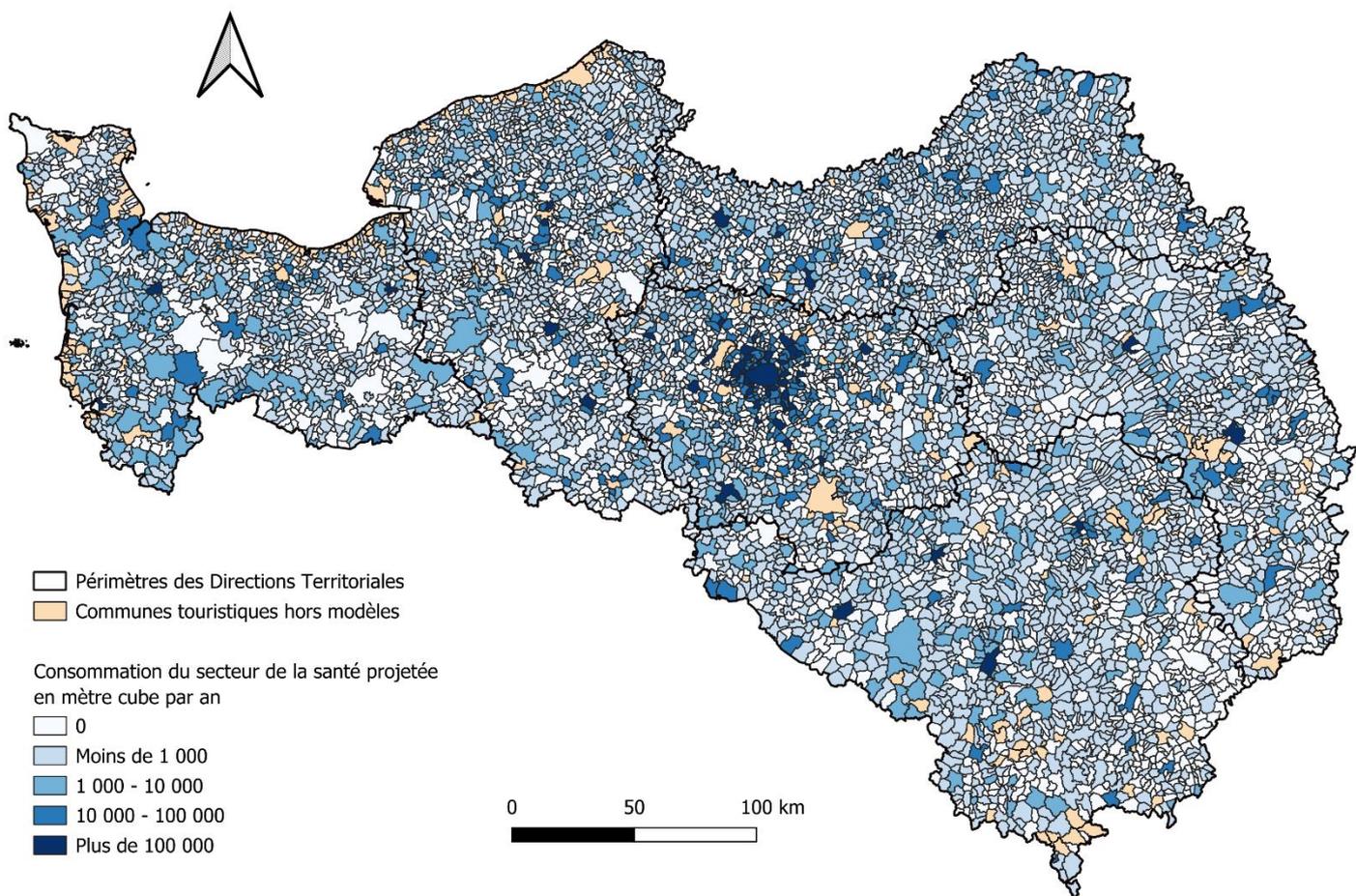


Figure 5 : consommation d'eau du secteur de la santé projetée, par commune et à l'échelle du bassin versant

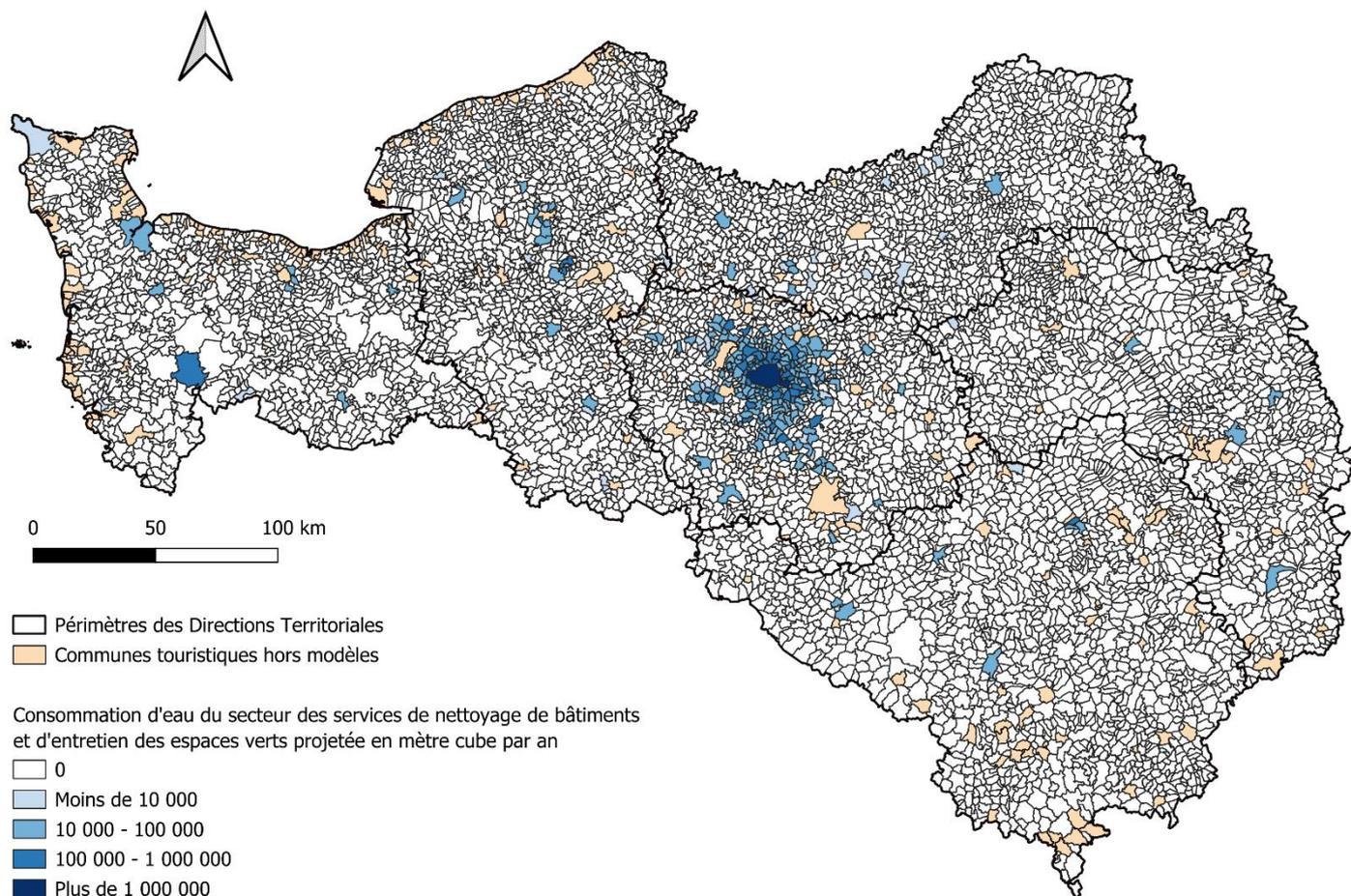


Figure 6 : consommation d'eau du secteur des services de nettoyage de bâtiments et d'entretien des espaces verts projetée, par commune et à l'échelle du bassin versant

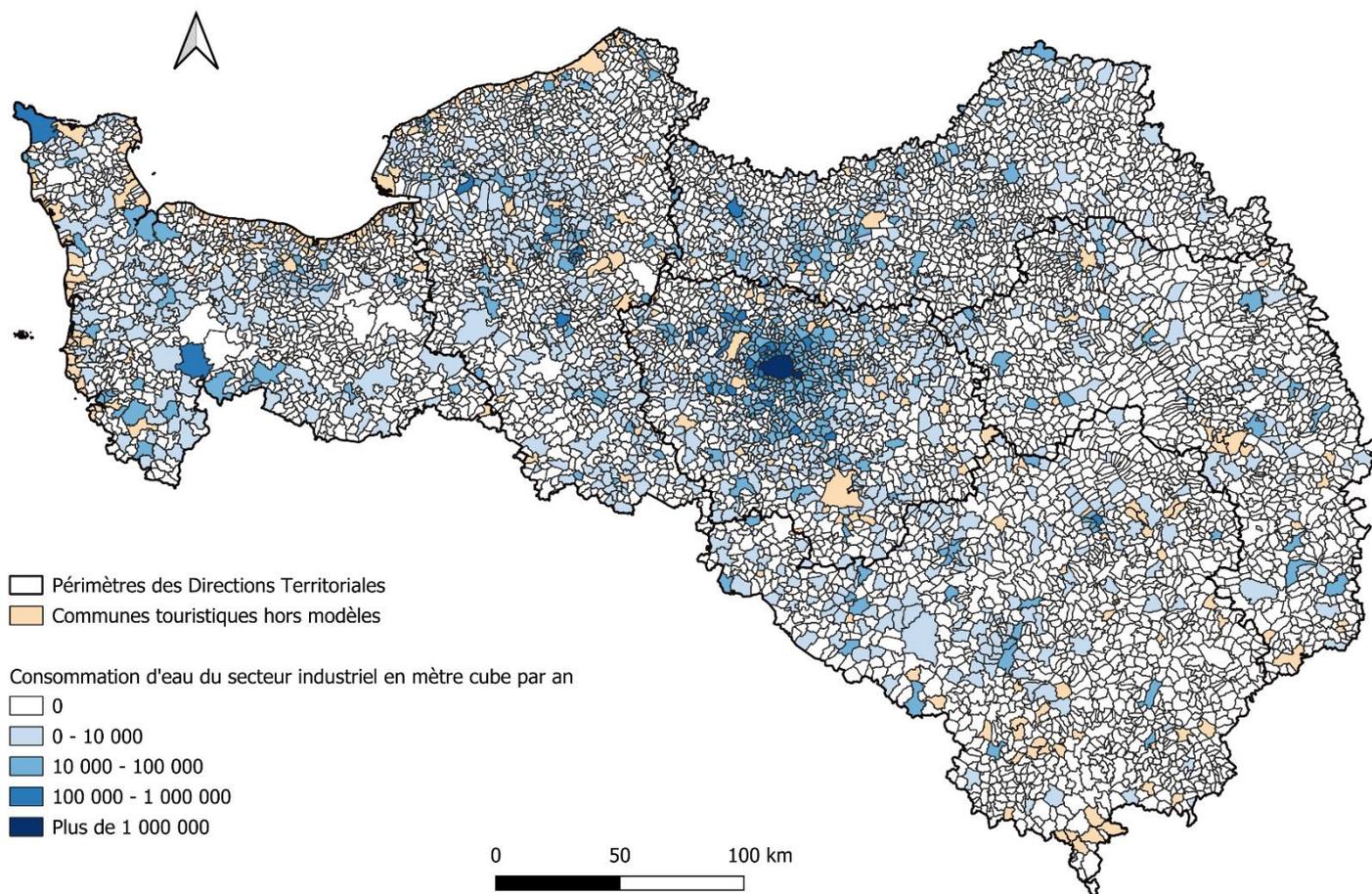


Figure 7 : consommation d'eau du secteur industriel projetée, par commune et à l'échelle du bassin versant

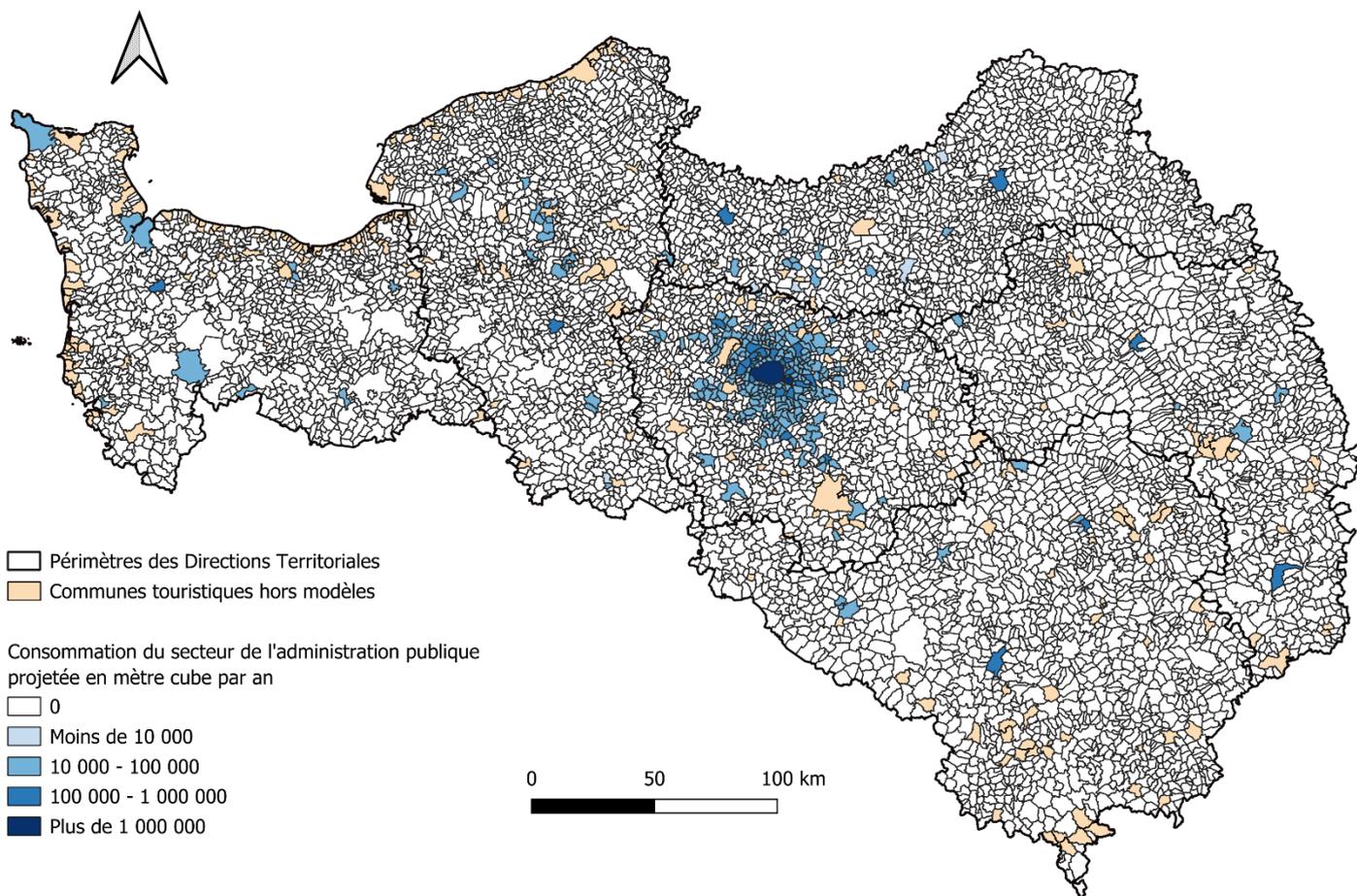


Figure 8 : consommation d'eau du secteur de l'administration publique projetée, par commune et à l'échelle du bassin versant

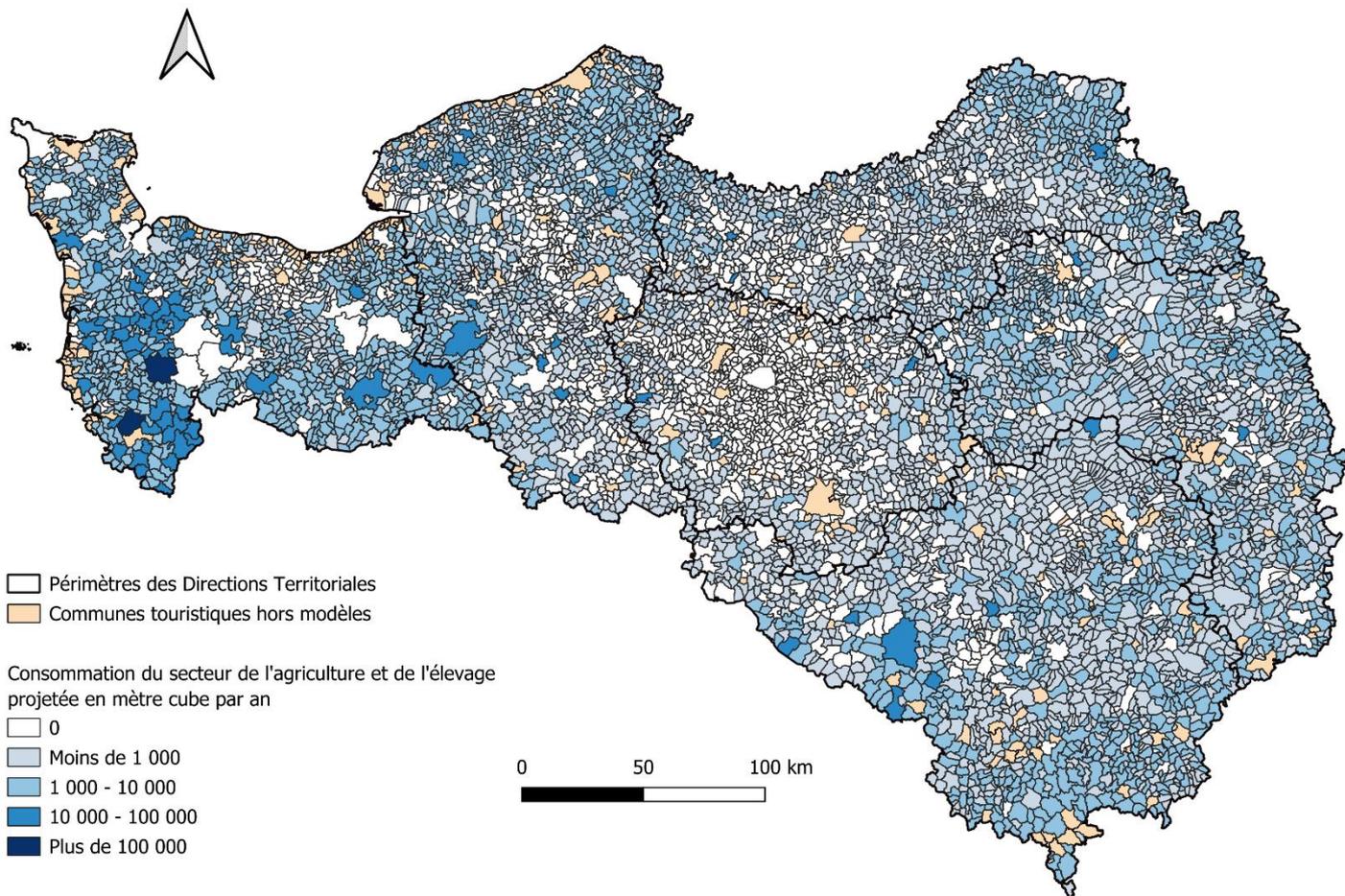


Figure 9 : consommation d'eau du secteur de l'agriculture et de l'élevage projetée, par commune et à l'échelle du bassin versant

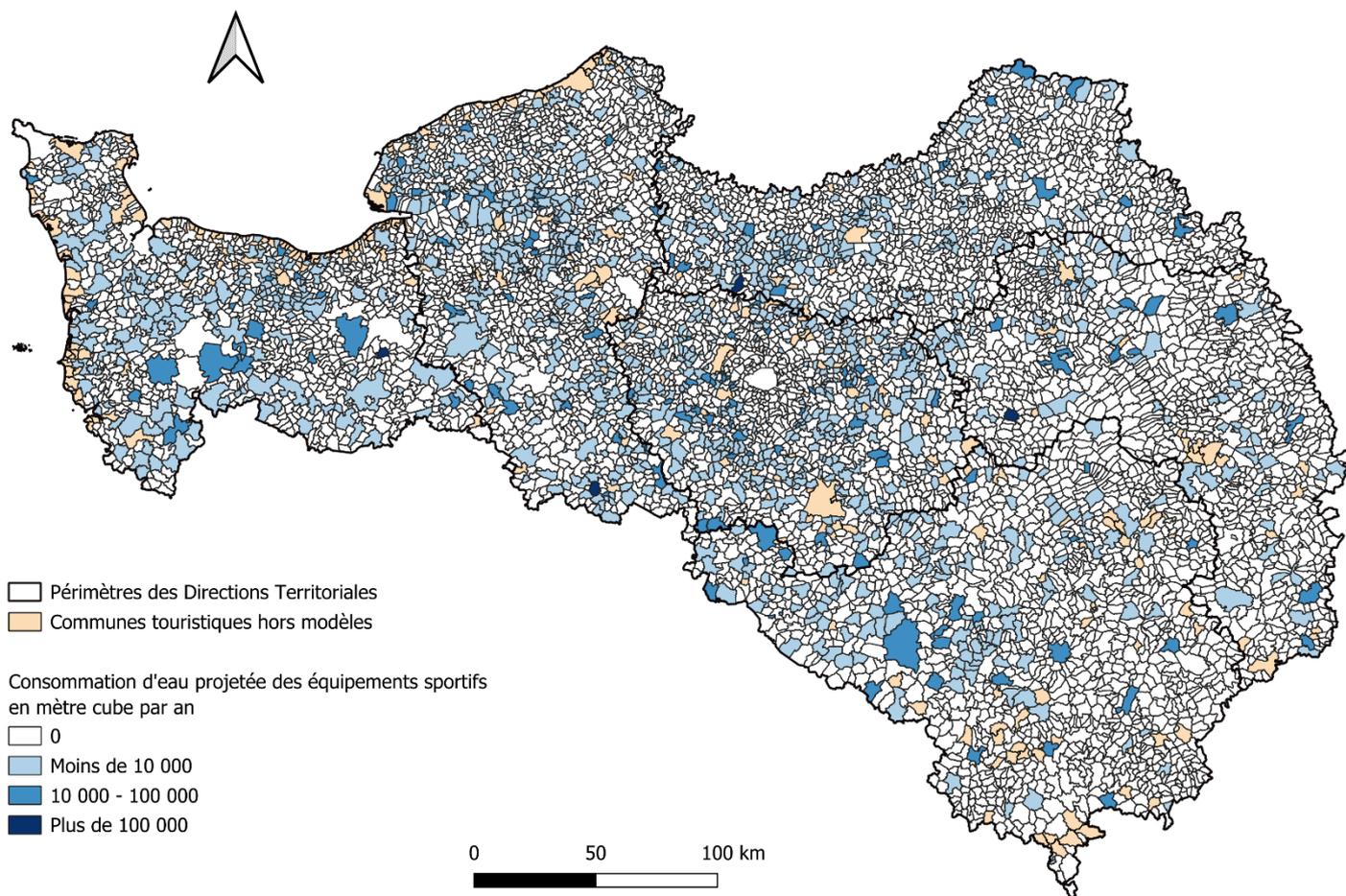


Figure 10 : consommation d'eau des équipements sportifs projetée, par commune et à l'échelle du bassin versant

Annexe V : résultat des volumes prélevés prédits pour chaque segment d'usager

	Prélèvements projetées (m3/an)	Part
Ménages	924030464	74,21
Commerces et tertiaire	133655480,3	10,73
Enseignement	45400144,2	3,65
Santé	38389571,87	3,08
Services de nettoyage de bâtiments et d'entretien des espaces verts	27056751,73	2,17
Industries	28120301,13	2,26
Administration publique	27744404,5	2,23
Agriculture - élevage	12188926,84	0,98
Equipements sportifs	8489765,848	0,68
<i>Surplus tourisme</i>	8546831,874	
<i>Constante</i>	8448642,557	
Total	1262071285	

Annexe VI : boîte à moustache des consommations totales et totales par habitant observées sur les communes touristiques et non touristiques

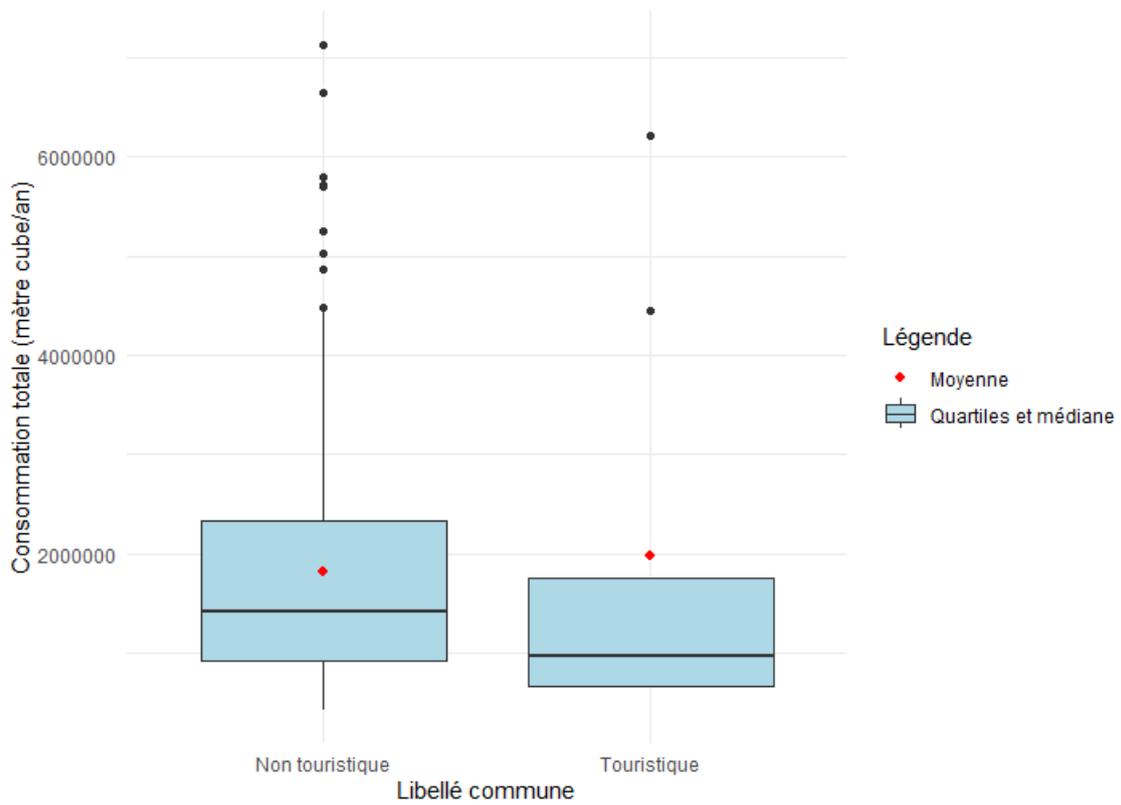


Figure 1 : boîte à moustache de la consommation totale observée sur les communes touristiques et non touristiques, pour la typologie de communes de plus de 10 000 habitants

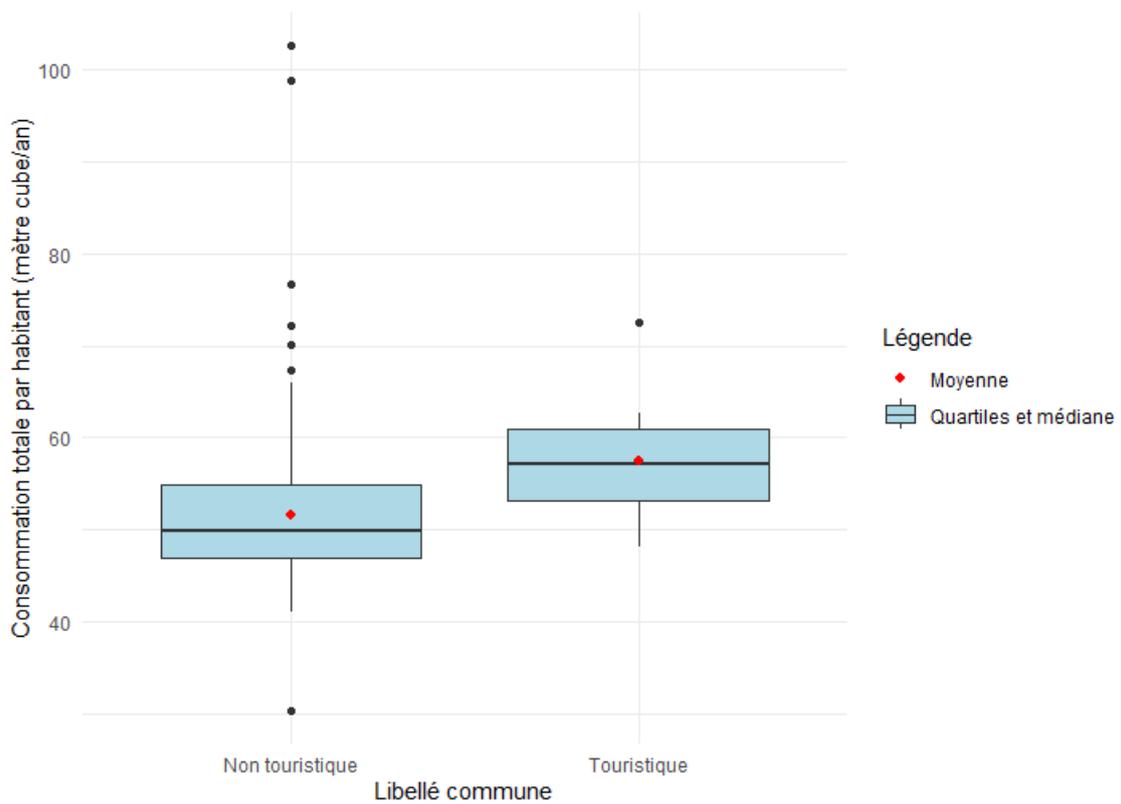


Figure 2 : boîte à moustache de la consommation totale par habitant observée sur les communes touristiques et non touristiques, pour la typologie de communes de plus de 10 000 habitants

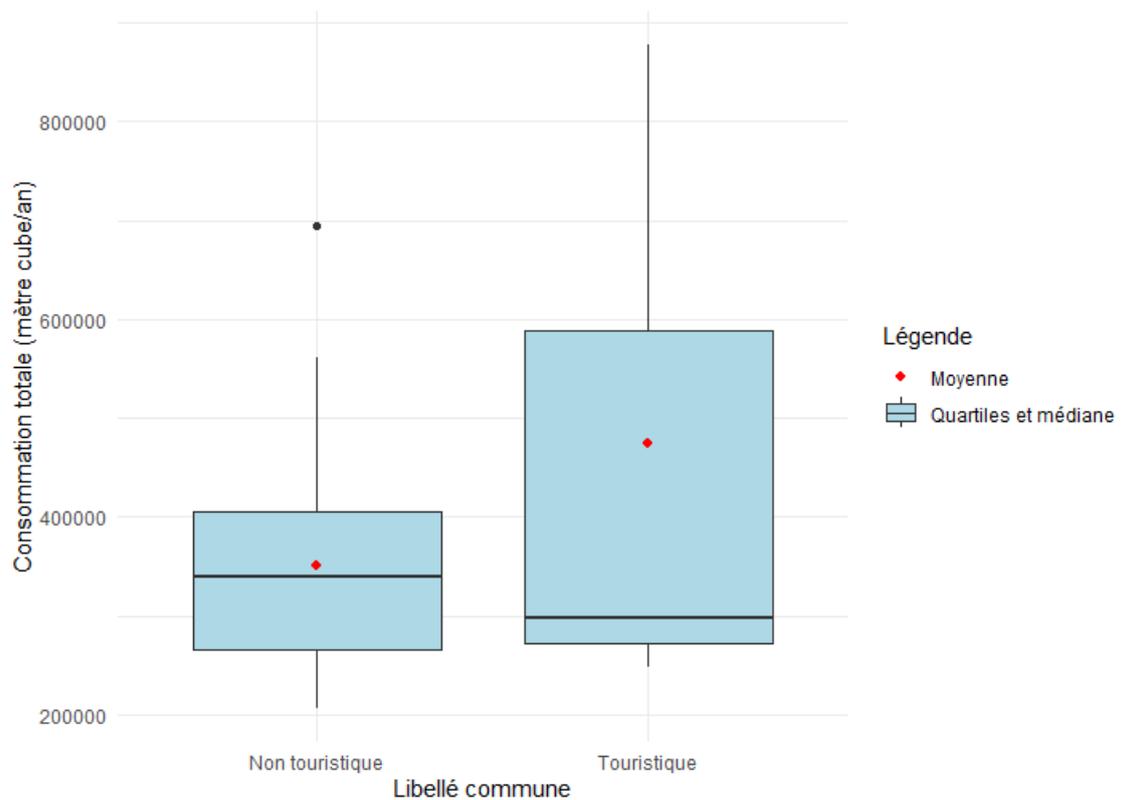


Figure 3 : boîte à moustache de la consommation totale observée sur les communes touristiques et non touristiques, pour la typologie de communes de 5 000 à 10 000 habitants

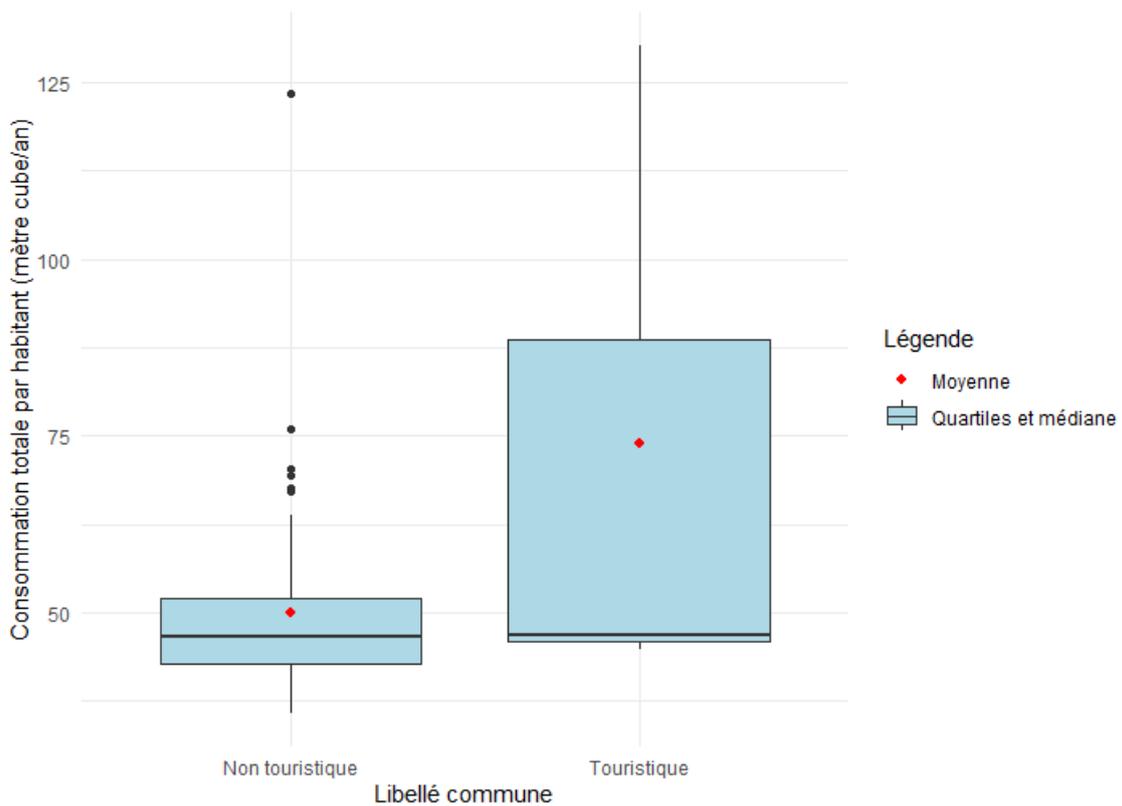


Figure 4 : boîte à moustache de la consommation totale par habitant observée sur les communes touristiques et non touristiques, pour la typologie de communes de 5 000 à 10 000 habitants

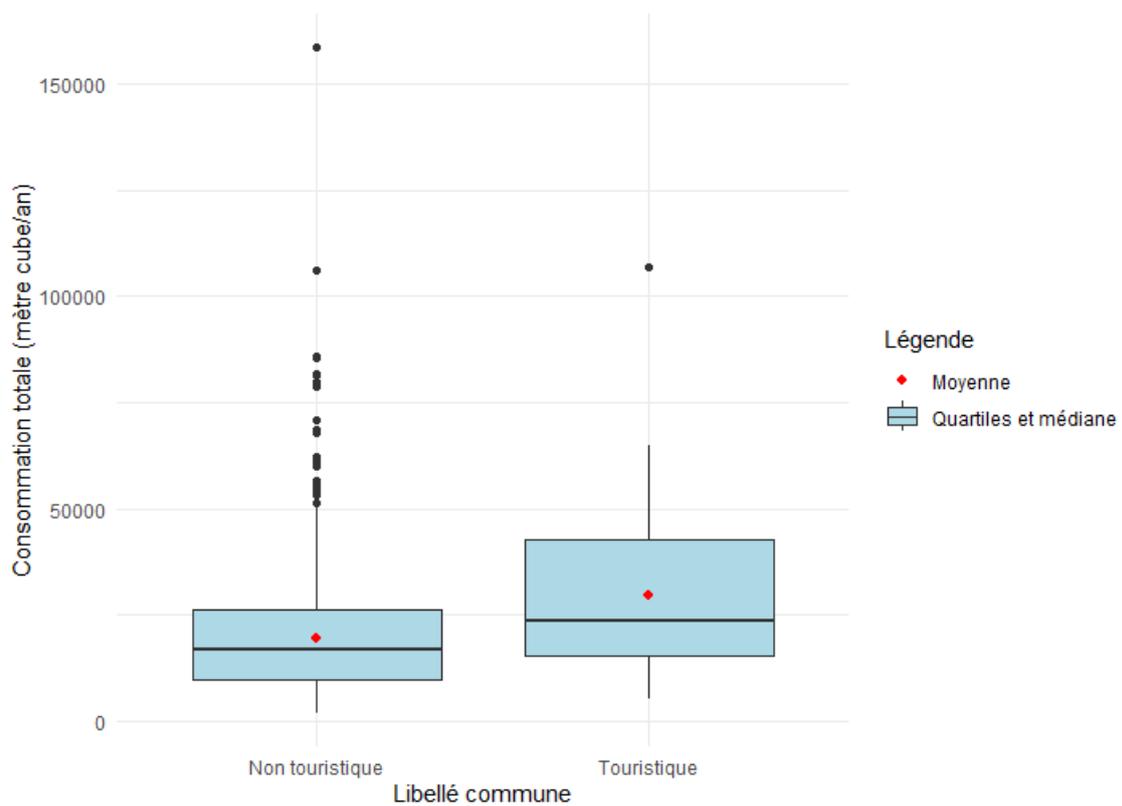


Figure 5 : boîte à moustache de la consommation totale observée sur les communes touristiques et non touristiques, pour la typologie de communes de moins de 1 000 habitants

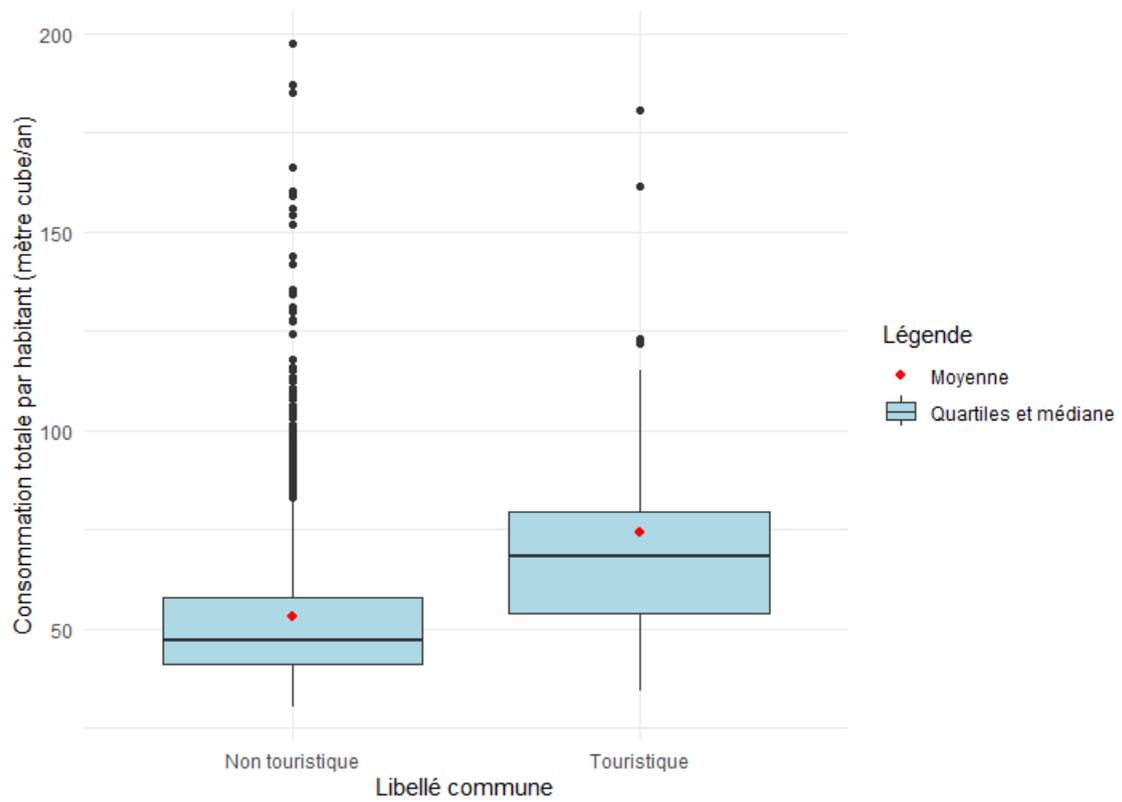


Figure 6 : boîte à moustache de la consommation totale par habitant observée sur les communes touristiques et non touristiques, pour la typologie de communes de moins de 1 000 habitants