



**HAL**  
open science

# Etude des déterminants et modélisation de la consistance du patrimoine des réseaux d'eau potable

K. Han Mui

► **To cite this version:**

K. Han Mui. Etude des déterminants et modélisation de la consistance du patrimoine des réseaux d'eau potable. Sciences de l'environnement. 2016. hal-02605536

**HAL Id: hal-02605536**

**<https://hal.inrae.fr/hal-02605536>**

Submitted on 16 May 2020

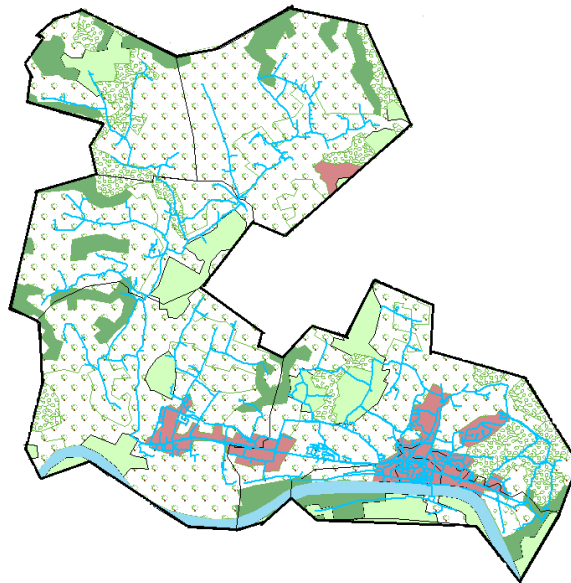
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



*Promotion Stuttgart 2013/2016*

**Travail de Fin d'Etudes présenté pour l'obtention du Diplôme  
d'Ingénieur ENGEES –spécialité GEDEMU**



**Etude des déterminants et modélisation de la consistance du patrimoine des réseaux  
d'eau potable**

*Kimberley Han Mui*

Maître de stage : Alain Husson

*4 janvier au 24 juin 2016*

*Irstea de Bordeaux,  
50 Avenue de Verdun  
33612 Cestas*

---

## REMERCIEMENTS

---

Je voudrais d'abord remercier Alain Husson, ingénieur au sein de l'équipe GPIE de l'Irstea de Bordeaux, qui m'a suivi en tant que maître de stage et avec qui j'ai beaucoup travaillé durant tout ce stage. Il a su me guider et me conseiller à divers étapes de mon travail en gardant toujours un esprit posé et pédagogue. Il m'a beaucoup transmis au cours de ces six mois et a contribué à faire de la recherche un domaine intéressant et agréable pour moi.

Je remercie également Eddy Renaud, chef de l'équipe GPIE, qui a été un pilier important durant toute l'étude et qui a su, à travers conseils et critiques, me pousser à être rigoureuse et précise dans mes démarches.

Je voudrais aussi remercier Camille Rodicq, également ingénieur au sein de l'équipe et collègue directe durant mon stage, pour ses conseils et sa sympathie durant tout le stage. Nous avons beaucoup réfléchi ensemble et ses relectures approfondies ont beaucoup contribué à la qualité de mon rendu final.

Je remercie aussi Yves Legat, l'expert en statistique de l'équipe, pour ses lumières lors de nos recherches de traitement de données, en particulier pour son aide lors des codages sur R.

Je voudrais enfin saluer Constant Gourlaouen, mon collègue de bureau, ainsi que tout le reste de l'équipe GPIE pour leur bonne humeur (surtout aux pauses café et au déjeuner) et leur disponibilité, sans parler de leur réception positive de ma passion pour la Corée...

---

RESUME

---

**Etude des déterminants et modélisation de la consistance du patrimoine des réseaux d'eau potable**

L'étude menée par **Irstea** pour le compte du ministère chargé de l'écologie et de l'Office National des Milieux Aquatiques (**ONEMA**) vise à construire un **dispositif permanent d'évaluation du patrimoine des réseaux d'eau potable aux échelles nationale et de bassin**. Après analyse de travaux précédents sur le sujet, les liens avec les **réseaux de voirie** ont été privilégiés pour **l'évaluation du linéaire** des réseaux d'eau potable. Les routes ont notamment été traitées en prenant en compte la présence de **bâtis**, la **démographie** et la **position géographique** des services. Pour le calage des modèles construits, une base de services a été créée à partir des données de la **base SISPEA**, de l'Observatoire national des services publics d'eau potable et d'assainissement. La validation de ces modèles s'est effectuée en vérifiant leur **applicabilité directe à l'échelle communale**, puis sur une deuxième base de services, et finalement par la méthode de **Monte-Carlo**. Le modèle retenu au terme de toutes ces étapes permet d'évaluer le linéaire de réseau d'eau potable à l'échelle nationale de manière satisfaisante ; cela est moins évident à l'échelle du service. Le travail effectué dans le cadre de ce stage permet une évaluation du linéaire global de canalisations à différentes échelles et ouvre la voie pour la modélisation des linéaires en fonction des diamètres, des matériaux et des âges des canalisations dans la continuité de l'étude.

---

ABSTRACT

---

**Study of the determining factors and modelling of the consistency of drinkable water networks**

The goal of the study led by Irstea and commissioned by the ONEMA and the Ministry of ecology seeks to **model** as closely as possible the **different features** of drinking water networks in metropolitan France and such on a **national scale**. After going through past works on the same theme, the **road networks** appeared to have the most potential for the modelling and were thus the main focus throughout this internship. These were especially processed using **GIS** (Geographical information system) programs and taking any **infrastructure present** into account. The models were then constructed by combining the effects of the **population** and/or **population density** in each studied service. In order to calibrate the created models, sets of data extracted from the **SISPEA database** of the National Observatory of Public Water Services were used. To validate the models, the variation between the results when applied on the services and on the associated municipalities were analysed, a second set of data was tested and the **Monte-Carlo method** was used. The final model obtained is believed to offer reliable results on the national scale, though it may produce more significant errors on the service level. The work produced throughout this internship concludes the modelling for the length evaluation of drinkable water networks in France for the ONEMA's study and shall be followed by the modelling of more characteristics such as the diameters, materials and age of the networks.

## SOMMAIRE

Table des figures et tableaux.....	5
Glossaire.....	7
Introduction.....	8
1. Contexte de l'étude.....	10
2. Etudes précédentes.....	11
2.1 Etudes notables.....	11
2.2 Etude Cador.....	11
3. Evaluation du linéaire de réseau d'eau potable : modèles linéaires simples.....	15
3.1 Données disponibles.....	15
3.2 Outils utilisés.....	16
3.3 Traitements primaires.....	16
3.3.1 Service de la Réole.....	16
3.3.2 Service de Lacanau.....	20
3.3.3 Syndicat de Guîtres.....	21
3.4 Combinaison avec les bases SISPEA.....	23
3.4.1 Première base : Communes mono-services.....	23
3.4.2 Deuxième base : Méga-services.....	24
3.4.3 Troisième base : Communes multi-services.....	24
3.5 Construction des premiers modèles.....	26
3.6 Affinage du modèle linéaire.....	31
3.6.1 Méthode de classification par taille de population.....	31
3.6.2 Méthode de classification par densité.....	36
3.7 Application sur les communes des services.....	38
4. Evaluation du linéaire de réseau d'eau potable : modèles multilinéaires.....	42
4.1 Construction des modèles.....	43
4.1.1 Corrélation entre paramètres deux à deux.....	43
4.1.2 Calcul du VIF (Variance Inflation Factor).....	43
4.1.3 Construction pas à pas du modèle : Critère d'Information d'Akaike.....	45
4.1.4 Modèles.....	45
4.2 Validation des modèles multilinéaires.....	48
Conclusion.....	52
Bibliographie.....	54
Annexes.....	56

TABLE DES FIGURES ET TABLEAUX

Figure 1: Départements étudiés dans l'étude de Cador, tirée de l'Etude Cador.....	12
Figure 2: Relation entre densité de population et longueurs de réseau à l'habitant, tirée de l'Etude Cador.....	13
Figure 3: Relation entre dispersion de la population et longueurs de réseau à l'habitant, tirée de l'Etude Cador.....	14
Figure 4: Comparaison entre le linéaire routier et le linéaire de réseau d'eau potable sur la Réole .....	17
Figure 5: Comparaison des réseaux sur la Réole après traitement par tampon 50 m .....	18
Figure 6: Effets des zones tampons sur les bâtis isolés.....	19
Figure 7: Comparaison des réseaux sur le service de Lacanau.....	21
Figure 8: Comparaison des réseaux sur le service de Guîtres.....	22
Figure 9 : Schéma de regroupement des méga-services.....	25
Figure 10 : Schéma des regroupements de données de la base SISPEA.....	26
Figure 11: Comparaison routes-réseaux eau potable sur la Gironde (1) .....	26
Figure 12: Comparaison routes-réseaux eau potable sur la Gironde (2) .....	27
Figure 13: Corrélations entre Routes brutes et linéaires de réseaux.....	28
Figure 14: Corrélations entre Routes filtrées et linéaires de réseaux.....	28
Figure 15: Corrélations entre Routes traitées par tampon 100 m et linéaires de réseaux .....	29
Figure 16: Corrélations entre Routes traitées par tampon 250 m et linéaires de réseaux .....	29
Figure 17: Comparaison des linéaires modélisés par rapport aux linéaires réels du premier modèle selon le linéaire réel de chaque service.....	31
Figure 18: Pente tous les 1000 points des linéaires de routes traitées par tampon par rapport au linéaire réel.....	32
Figure 19: R <sup>2</sup> des valeurs de pentes obtenues tous les 1000 points .....	32
Figure 20: Courbe des pentes à 1000 points après correction .....	33
Figure 21: Courbe des R <sup>2</sup> pour les pentes à 1000 points après correction .....	33
Figure 22: Pentés à appliquer pour les classes de populations brutes.....	34
Figure 23: Pentés tous les 1000 points en prenant en compte la densité des services.....	37
Figure 24: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la première base .....	38
Figure 25: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la deuxième base .....	39
Figure 26: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la troisième base (limite supérieure) .....	40
Figure 27: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de densités sur la première base.....	41
Figure 28: Exemple de distribution des densités de bâtis sur un réseau routier .....	42
Figure 29: Corrélations deux à deux pour les huit premiers paramètres .....	43
Figure 30: Analyse des résidus par rapport à la population des services .....	47
Figure 31: Résultats de la méthode Monte-Carlo sur le modèle K1.....	49
Figure 32: Résultats de la méthode Monte-Carlo sur le modèle K2.....	49
Figure 33: Schéma boîte à moustaches.....	50

Tableau 1: Caractéristiques de chaque département étudié, tirées de l'Etude Cador.....	12
Tableau 2: Données sur chaque département étudié, tirées de l'Etude Cador.....	13
Tableau 3: Comparaison numérique des réseaux sur la Réole .....	17
Tableau 4: Comparaison des parties non cohérentes des réseaux.....	17
Tableau 5: Comparaisons numériques entre le réseau routier et le réseau d'eau potable de la Réole.....	18
Tableau 6: Comparaisons numériques sur le réseau de la Réole après traitement par zone tampon 50 m.....	19
Tableau 7: Résultats des traitements primaires sur le service de Lacanau.....	20
Tableau 8: Résultats des traitements primaires sur le service de Guîtres.....	21
Tableau 9: Coefficients obtenus après traitement par zones tampon sur chaque région .....	30
Tableau 10: Résultats du premier modèle.....	30
Tableau 11: Erreurs corrigées dans la première base.....	33
Tableau 12: Classes de populations brutes obtenues .....	34
Tableau 13: Classes de populations retenues .....	34
Tableau 14: Pentés associées aux classes retenues.....	35
Tableau 15: Pentés obtenues sur les nouveaux indicateurs .....	36
Tableau 16: Modélisation par classe de densité.....	37
Tableau 17: Résultats obtenus avec les modèles simples linéaires .....	41
Tableau 18: Valeurs de VIF pour la recherche de nouveaux paramètres (1) .....	44
Tableau 19: Valeurs de VIF pour la recherche de nouveaux paramètres (2) .....	44
Tableau 20: Valeurs de VIF pour la recherche de nouveaux paramètres (3) .....	44
Tableau 21: Résultats des modèles multilinéaires.....	48
Tableau 22: Coefficients du modèle K2 retenu.....	50
Tableau 23 : Résumé des résultats trouvés sur les différents modèles .....	52
Tableau 24: Résultats des premières comparaisons sur la Réole.....	60

---

## GLOSSAIRE

---

AITF = Association des Ingénieurs Territoriaux de France

ASTEE = Association Scientifique et Technique pour l'Eau et l'Environnement

CGCT = Code Général des Collectivités Territoriales

DDT = Direction Départementale des Territoires

IFEN = Institut Français de l'Environnement

IGN = Institut National Information Géographique Forestière

INSEE = L'Institut National de la Statistique et des Etudes Economiques

Irstea = Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture

Laboratoire GESTE = Gestion Territoriale de l'Eau et de l'Environnement

GPIE = Gestion Patrimoniale des Infrastructures liées à l'Eau

MEEM = Ministère de l'Environnement, de l'Energie et de la Mer

ONEMA = Office National de l'Eau et des Milieux Aquatiques

SOeS = Service de l'Observation et des Statistiques

SISPEA = Système d'Information sur les Services Publics d'Eau et d'Assainissement

VIF = Variance Inflation Factor



## INTRODUCTION

---

Le **décret n° 2012-97 du 27 janvier 2012** de la loi « Grenelle II » impose désormais à tous les services publics de l'eau le rendu d'un **descriptif détaillé** de leurs réseaux. Cette obligation constitue une opportunité pour construire un dispositif permanent d'évaluation du patrimoine des réseaux d'eau potable aux échelles nationale et de bassin qui pourra servir la construction des politiques publiques en matière de gestion patrimoniale des réseaux d'eau. Aujourd'hui, seule une petite proportion des services d'eau (environ **4000 sur 14 000 services**) renseignent effectivement des informations patrimoniales dans la base de données de l'observatoire national. Pour une meilleure gestion du patrimoine, une planification plus efficace et une implémentation durable des travaux à effectuer par rapport aux besoins des collectivités en France, l'ONEMA recherche donc un moyen d'évaluer le patrimoine du réseau d'eau potable national malgré ces lacunes.

Cette étude, intitulé **Construction d'un dispositif permanent d'évaluation du patrimoine des réseaux d'eau potable aux échelles nationale et de bassin**, fera intervenir 3 équipes d'Irstea dont l'équipe GPIE à Bordeaux, pour une durée de 3 ans (2016-2018). Elle se décompose en 3 axes d'études :

- Une **approche technique** et statistique de l'évaluation du patrimoine (GPIE Bordeaux)
- L'identification des **besoins et validation du dispositif** (G-EAU Montpellier)
- **L'évaluation financière et économique** du patrimoine (GESTE Strasbourg et GPIE Bordeaux)

L'équipe GPIE se focalise dans un premier temps sur **l'élaboration, le calage et la validation de l'outil d'évaluation**. Le but de ce stage de TFE est de contribuer au développement initial de l'outil et d'aider à son amélioration au fil de rencontres avec l'ONEMA et le MEEM afin de s'assurer que l'outil convient à leurs besoins. Il faudra dégager les premiers résultats, en se focalisant sur le premier axe, notamment en tentant de modéliser les premières caractéristiques liées aux patrimoines régionaux (et de bassin) avant de pouvoir l'étendre à l'échelle nationale.

Dans un premier temps, un travail de bibliographie a été effectué, principalement pour prendre connaissance des études passées sur des sujets similaires. Puis, une comparaison réseau eau potable – environnement du réseau a été faite sur une commune pour obtenir un premier aperçu des possibilités et des limites des ressources d'informations disponibles. Un traitement centré sur les linéaires de voirie est ensuite appliqué sur l'ensemble des régions de France afin de pouvoir construire un premier modèle évaluant les linéaires de réseaux d'eau potable. Des ajustements de modèles par classes et des méthodes statistiques ont ensuite été testés, et leurs effets évalués sur les résultats obtenus. Enfin, un modèle a été retenu et appliqué sur l'ensemble de la France et son intervalle de confiance a été évalué.



## 1. CONTEXTE DE L'ETUDE

Depuis le **décret n° 2012-97 du 27 janvier 2012** de la loi « Grenelle II », le Code Général des Collectivités Territoriales (CGCT) et le Code de l'Environnement imposent le rendu d'un descriptif détaillé des réseaux des services publics de l'eau, en sus de réglementer aussi les aspects techniques de la gestion du patrimoine des systèmes d'eau potable en France. La rédaction de ces descriptifs vise notamment à aider les collectivités dans leur gestion de leurs réseaux, dans le but principal de réduire leurs pertes en eau.

L'ONEMA est l'organisme technique français de référence sur la connaissance et la surveillance de l'état des eaux et sur le fonctionnement écologique des milieux aquatiques. Il assure le pilotage fonctionnel national du Système d'Information sur l'Eau (SIE) et, pour répondre aux besoins des acteurs de la politique de l'eau, il soutient des projets de recherche-développement-innovation dont les résultats sont transférés sur le terrain. Dans le cadre du décret mentionné plus haut, l'ONEMA s'est associé à l'ASTEE et l'AITF afin de créer un guide pour aider les collectivités à élaborer leurs descriptifs détaillés et à le mettre à jour tous les ans [1]. Ce guide indique notamment les informations à renseigner :

- Plan d'ensemble (constamment mis à jour) avec au minimum les canalisations, leur diamètre et les dispositifs généraux
- La ressource en eau mobilisée (pour le calcul du seuil de rendement)
- Inventaire
  - o Linéaire de canalisations sans branchements
  - o Nombre de branchements particuliers (non obligatoire)
  - o Diamètres des canalisations
  - o Matériaux des canalisations
  - o Conduites (date de pose, pressions nominales...)
  - o Catégorie d'ouvrage : sensible ou non sensible (en général non sensible pour les ouvrages d'eau potable)
  - o Classe d'incertitude pour la précision cartographique (A: incertitude =0,4m, B : 0,4m <incertitude < 1,5m, C : incertitude >1,5m)

Devant la charge de travail que représente la collecte de données à partir d'archives et le besoin de systèmes de gestion pour acquérir toutes ces données, les services de l'eau sont souvent dans l'impossibilité de remettre un descriptif complet, ce qui rend la programmation efficace et économique des travaux nécessaires plus laborieuse. En effet, selon le dernier rapport de l'Observatoire des services publics d'eau et d'assainissement, [2] près de **41% des services** d'eau potable en France n'ont jamais renseigné la base de données depuis 2008 (base SISPEA).

Ce stage s'insère donc dans une collaboration entre l'ONEMA, Irstea et d'autres partenaires (dont le laboratoire GESTE et les différentes agences de l'eau) pour construire un dispositif permanent d'évaluation du patrimoine des réseaux d'eau potable aux échelles nationale et de bassin. Compte tenu de la lourdeur et de la difficulté de constituer une base de données nationale exhaustive (cf. les taux de remplissage de la base de données SISPEA évoqués ci-dessus), il apparaît qu'une approche par modélisation statistique est à privilégier. Le modèle attendu devra permettre de connaître le linéaire de canalisation d'eau potable à l'échelle nationale mais aussi de caractériser les réseaux en termes de matériaux, de diamètres moyens et de date de pose.

## 2. ETUDES PRECEDENTES

---

Peu d'études existent à ce jour sur le sujet. En effet, plusieurs estimations ont été faites dans le passé afin d'obtenir des ordres de grandeur pour illustrer un rapport environnemental ou patrimonial. Cependant, les études spécifiques sur ce sujet sont rares, ce qui a conduit à l'usage de chiffres variés d'un article à un autre et d'année en année.

Quelques chiffres que nous pouvons retrouver dans la littérature générale vont de 800 000 km [3] à 1 050 000 km [4] en passant par 906 000 km [5]. La problématique aujourd'hui est donc de trouver un linéaire cohérent qui permettrait d'unifier l'évaluation du patrimoine partout en France et faciliter la coordination des politiques de gestion de ce patrimoine.

### 2.1 ETUDES NOTABLES

---

Parmi les études qui ont été faites précédemment, nous pouvons citer deux exemples le plus souvent mentionnés :

- **L'estimation de l'ONEMA**, consistant à utiliser des regroupements de services selon des caractéristiques urbaines et démographiques avant de calculer des coefficients selon la taille de population desservie par le service. Les coefficients sont ensuite utilisés directement pour l'extrapolation. (Etude interne, 2010)
- L'estimation **du SOeS** (anciennement IFEN), qui a été calée sur un échantillon créé en vue de calculer le prix de l'eau. Cette étude n'est donc pas adaptée pour l'extrapolation du réseau dans le cadre de la gestion patrimoniale. [6]
- **L'Etude Cadour** [7] effectuée en 2002 est détaillée ci-après.

### 2.2 ETUDE CADOR [7]

---

L'étude la plus poussée a été effectuée en 2002 par Jean Michel Cadour, sur huit départements de France, les résultats étant extrapolés à la France entière. Les données utilisées étant le résultat d'enquêtes sur les différents territoires, cette étude a notamment permis de faire un bilan sur les moyens et méthodes employés pour de telles enquêtes.

Les départements étudiés par Cadour sont :

- Allier (03)
- Aveyron (12)
- Doubs (25)
- Hérault (34)
- Indre-et-Loire (37)
- La Manche (50)
- La Somme (80)
- Bas-Rhin (67)

Les départements sont répartis géographiquement sur toute la France comme le montre la figure suivante :

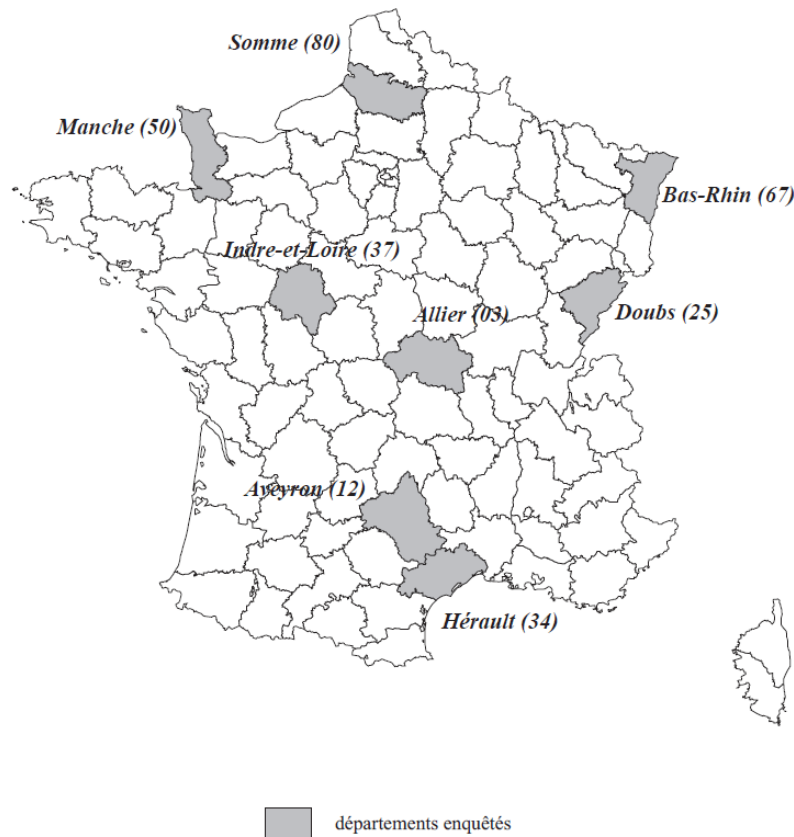


Figure 1: Départements étudiés dans l'étude de cador, tirée de l'Etude Cador

Nous pouvons aussi observer que les caractéristiques démographiques sont plutôt diverses selon le département. ([Tableau 1](#))

Tableau 1: Caractéristiques de chaque département étudié, tirées de l'Etude Cador

Département enquêté	Agence de l'eau	Nombre de communes	Taille de la population	Surface (km <sup>2</sup> )
Allier 03	Loire-Bretagne	320	345 000	7300
Aveyron 12	Adour-Garonne	304	263 800	8735
Doubs 25	Rhône Méditerranée Corse	592	499 000	5230
Hérault 34	Rhône Méditerranée Corse	343	896 500	6100
Indre-et-Loire 37	Loire-Bretagne	277	554 000	6126
La Manche 50	Seine-Normandie	516	481 470	5983
Somme 80	Artois-Picardie	782	555 550	6170
Bas Rhin 67	Rhin-Meuse	527	1 026 120	4755

L'occupation du sol est assez hétérogène d'un département à un autre ce qui a permis d'obtenir une représentativité du sol français a priori, à l'exception de la présence d'un département à dominante urbaine comme l'Île-de-France.

Les données récoltées et utilisées sous forme de paramètres pour la modélisation du linéaire sont définies dans le Tableau 2 suivant :

Tableau 2: Données sur chaque département étudié, tirées de l'Etude Cadot

	longueur du réseau (km)	densité de pop. hbt/km <sup>2</sup>	taux de pop. éparsé %	densité du réseau km/km <sup>2</sup>	habitant par km de canal.	mètres de canal. par habitant	diamètre moyen (mm)
<b>Allier</b>	9 000	47	27	1,2	38	26	117
<b>Aveyron</b>	6450	30	39	0,7	41	25	100
<b>Doubs</b>	6 800	95	8	1,3	73	14	105
<b>Hérault</b>	7 300	147	8	1,2	123	8	145
<b>Indre-et-Loire</b>	9 900	90	22	1,6	56	18	103
<b>Manche</b>	12 110	81	40	2,0	40	25	89
<b>Somme</b>	5 950	90	5	1,0	93	11	116
<b>Bas-Rhin</b>	7 300	216	3	1,5	141	7	126

Une première approche graphique simple a permis à Cadot d'obtenir les tendances suivantes :

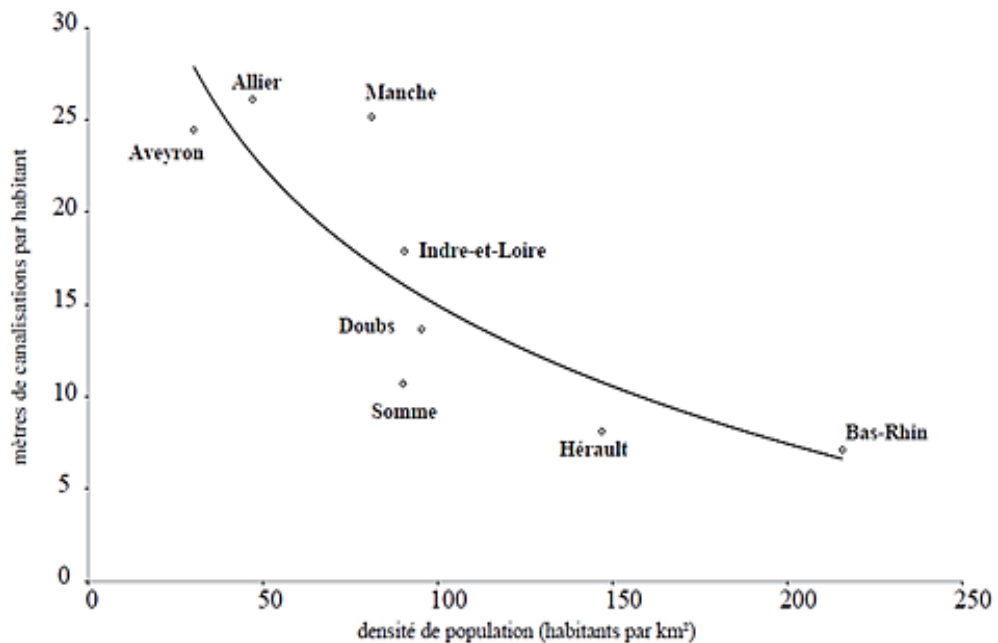


Figure 2: Relation entre densité de population et longueurs de réseau à l'habitant, tirée de l'Etude Cadot

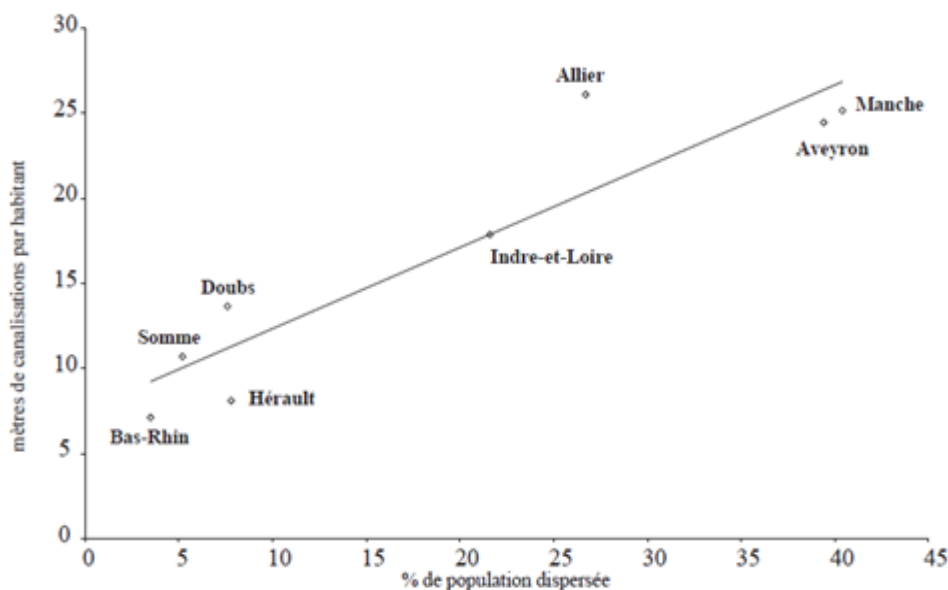


Figure 3: Relation entre dispersion de la population et longueurs de réseau à l'habitant, tirée l'Etude Cador

L'accent est mis sur le type de tendance suivi (linéaire, courbe, ascendant, descendant), cependant il existe un doute considérable quant à l'exploitabilité et la précision des coefficients obtenus de telles courbes de tendance. Cador spécifie d'ailleurs que les corrélations ne peuvent être utilisées qu'à l'échelle du **département** et pas à l'échelle du **service**.

Ces différents paramètres ont été ensuite évalués par une **Analyse en Composantes Principales** (ACP) afin de retenir ceux qui sont le mieux corrélés au linéaire d'eau potable. Une régression linéaire multiple sur les paramètres issus de l'ACP donne alors des modèles simples tels :

$$L = 35,14 * pop. dispersée + 5430000$$

$$L = -0,323 * pop. agglomérées + 35,07 * pop. dispersée - 178,87 * surface + 6672931$$

Le calage de ces modèles a été effectué sur la base d'une enquête du SOeS avant de l'appliquer sur l'ensemble de la France pour une estimation de linéaire total d'environ 850 000 km.

De la même manière, lors de son étude, Cador a évalué l'âge des canalisations en France, les diamètres et les matériaux selon l'emplacement du réseau dans les différents départements.

Les modèles de Cador ont donc permis d'obtenir un **ordre de grandeur des caractéristiques** de réseau à l'échelle nationale en 2002 en se basant sur un échantillon assez restreint. Nous constatons cependant à partir des figures 2 et 3 que les différents départements ne suivent tout de même pas de manière significatives les courbes de tendances produites. De plus, les démarches de Cador n'ayant pas été précisément détaillés dans son mémoire, il est difficile de **vérifier** la pertinence de ses résultats. Par exemple, la population dispersée utilisée qui est considérée par Cador comme étant un paramètre majeur de la modélisation, n'est pas définie clairement.

### 3. EVALUATION DU LINEAIRE DE RESEAU D'EAU POTABLE : MODELES LINEAIRES SIMPLES

---

---

#### 3.1 DONNEES DISPONIBLES

---

Un des objectifs principaux de l'étude est de pouvoir construire des modélisations à partir de bases de données **publiques** et donc **accessibles** à toute structure souhaitant réaliser une évaluation de patrimoine. Une première étape a donc consisté à faire un bilan des données disponibles.

La liste suivante résume la totalité des bases répertoriées :

- La **base de données SISPEA** pour les informations relatives aux services et aux communes lorsque celles-ci sont renseignées
- La base de données de **bâtis** sur tout le territoire français obtenue sous forme de couche SIG à partir de l'**IGN** (Ces couches ne sont pas accessibles directement aux particuliers, mais sont gratuites pour tout organisme de recherche ou de service public)
- La base de **réseau routier** sur le territoire français obtenue sous forme de couche SIG à partir de la base Geofabrik, qui source cette information **d'Open Street Map**
- La base **Corine Land Cover**, sous forme de couche SIG, qui permet de connaître l'occupation du sol sur le territoire français
- Le **réseau de contour IRIS** de l'IGN, en couche SIG, regroupant la population par « îlots » de 2000 habitants selon le recensement de l'INSEE.
- La couche des **communes de France** sous SIG, qui regroupe les informations disponibles par communes et permettra d'appliquer les traitements à cette échelle pour l'estimation nationale.
- **Les bases de données de l'INSEE**, notamment en ce qui concerne les données de types de logement exploités selon les communes, pour calculer l'indicateur « Maisons par bâti », le nombre de maisons par rapport aux nombre de bâtis sur une commune.

Afin de créer les premiers modèles, des **données réelles de réseaux d'eau** potable sur certaines communes/services ont aussi été récupérées à partir des services concernés sous forme de couche SIG. Ces données recueillies étaient relatives à des réseaux d'eau situés principalement en **Aquitaine**.



## 3.2 OUTILS UTILISÉS

Afin de traiter ces données, les outils à notre disposition sont aussi des outils **accessibles** à la majorité des structures (Certains sont payants mais généralement déjà en place dans la plupart des organismes publics).

Pour traiter les données d'occupation du sol ainsi que les routes et les bâtis, **Arcgis** a été utilisé dans cette étude (le même travail peut aussi être effectué sur **Qgis** qui est un logiciel libre). Le logiciel de Système d'Information Géographique va permettre d'effectuer le croisement des informations spatiales sur l'ensemble du territoire français (routes, bâtis,..) afin d'obtenir des paramètres qu'il sera ensuite possible de corréliser avec les linéaires des réseaux d'eau potable.

**Access** et **Excel** sont les principaux outils utilisés pour le regroupement de données ainsi que la réorganisation des différentes bases de données accessibles afin de produire une unique base avec les paramètres nécessaires à la construction du modèle.

Le **logiciel R** a été utilisé car il est gratuitement accessible sur le web. Ses points forts sont **la traçabilité et la reproductibilité** des traitements. Le logiciel permet d'effectuer une grande quantité de calculs, spécialement ceux qui ont trait à la **statistique**. Les jointures complexes, les études sur les régressions linéaires et multilinéaires lors de la construction des modèles ont été réalisées sous R. Le logiciel permet notamment d'observer les données résiduelles lors des modélisations et d'évaluer les erreurs liées à nos modèles.

## 3.3 TRAITEMENTS PRIMAIRES

### 3.3.1 SERVICE DE LA REOLE

Dans l'étude de Cador [7], un modèle sur le **linéaire routier** a aussi été testé suivant l'observation d'une cohérence intéressante trouvée durant l'enquête sur le département du Bas-Rhin. Le modèle issu donne une estimation linéaire de 870 000 km mais **n'a pu être calé** de manière satisfaisante sur les bases SOeS de par une différence conséquente entre les résultats obtenus lors de l'application sur les zones rurales et urbaines. Ce modèle tentait de lier directement le linéaire de routes au linéaire de réseau d'eau potable ; il était donc logique que les résultats ne soient pas concluant d'une zone urbaine à une zone en campagne.

Cependant, intuitivement, la piste du linéaire routier reste une piste intéressante, puisque la majeure partie du réseau d'eau potable est généralement posée sous voirie (82% d'après Cador [7]). Dans ce stage, j'ai donc choisi d'approfondir cette hypothèse, en apportant un traitement au linéaire du réseau routier avant de tenter de le relier au réseau d'eau potable.

Il s'agit de dégager dans un premier temps les limites de chaque outil et de chaque base de données qui nous est accessible dans le cadre de la modélisation du linéaire de réseau. La proximité entre les deux réseaux (routier et eau potable) a d'abord été confirmée en superposant les deux réseaux sur des services de la Gironde, en commençant par celui de la Réole.

Le service considéré contient à la fois un centre-ville très urbanisé et des secteurs de types agricoles ou sylvestres, ce qui permet d'observer les éventuels effets de traitements sur les zones urbaines ou rurales.

Après récupération des linéaires routiers sur la France et un recoupage par rapport aux limites du service, l'affichage sur ArcGis permet vite de se rendre compte qu'en effet, le réseau d'eau potable se calque correctement sur le réseau routier sur le groupement de communes. (Figure 4)

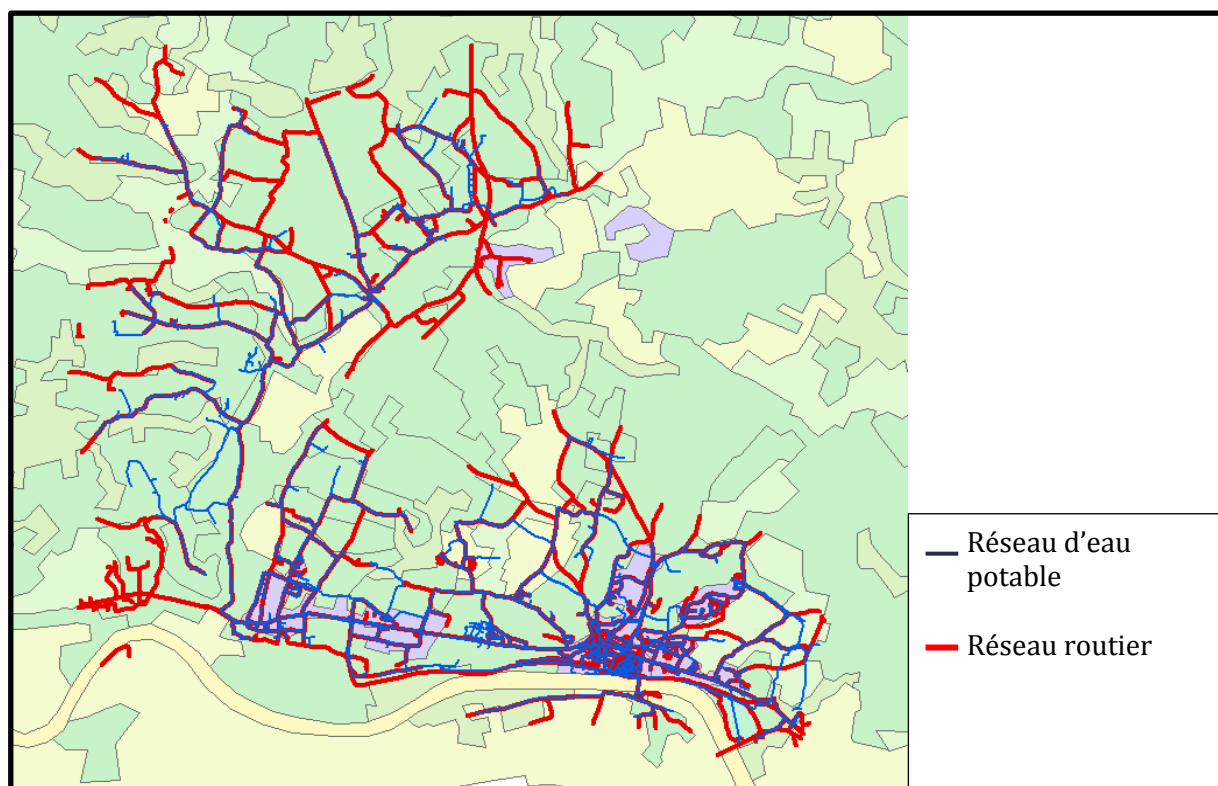


Figure 4: Comparaison entre le linéaire routier et le linéaire de réseau d'eau potable sur la Réole

L'annexe A1 présente les différentes démarches d'abord utilisées sur la commune de La Réole pour trouver la meilleure méthode d'évaluation de la cohérence avec le linéaire de réseau routier (calculée en pourcentage par rapport au linéaire de réseau d'AEP).

ArcGis permet de calculer le linéaire de route ainsi découpé pour le comparer au linéaire de canalisation AEP actuel du service (donnée renseignée par le service).

Tableau 3: Comparaison numérique des réseaux sur la Réole

Longueur de réseau (donnée service)	143,723 km
Longueur de réseau routier sur les secteurs	138,625 km (~96%)

Il ne faut évidemment pas tirer de conclusion hâtive de la proximité entre ces deux valeurs. En effet, lorsque nous calculons les linéaires de chaque réseau ne se recoupant pas avec l'autre (calculs explicités dans l'annexe A1), nous obtenons les chiffres suivants :

Tableau 4: Comparaison des parties non cohérentes des réseaux

Longueur de canalisations ne suivant pas le réseau routier	42,9 km (~29,9% de longueur de réseau)
Longueur de routes sans réseau eau potable	51,0 km (~35,5% de longueur de réseau)

C'est donc une certaine compensation entre ces deux valeurs qui peut expliquer la proximité entre les valeurs des réseaux de route et d'eau potable vues dans le Tableau 3.

Des calculs similaires en différenciant des secteurs du service par rapport aux caractéristiques d'occupation de sol permettent d'obtenir le Tableau 5 suivant :

Tableau 5: Comparaisons numériques entre le réseau routier et le réseau d'eau potable de la Réole

Secteur	Longueur de réseau AEP sur le secteur (km)	Longueur de réseau routier sur le secteur (km)	Longueur de réseau AEP hors réseau routier (km)	Longueur de réseau routier hors réseau AEP (km)
La Réole	143,723	138,625 (96,5%)*	42,95 (29,9%)	51 (35,5%)
Secteur péri-urbain + urbain	73,652	71,863 (97,6%)	18,8 (25,5%)	25,2 (34,2%)
Secteur rural	20,371	28,48 (139%)	6,77 (33,2%)	15,9 (78%)
Secteur urbain	42,659	34,191 (80%)	11,2 (26,2%)	7,07 (16,6%)

\*les pourcentages sont calculés par rapport à la longueur de réseau AEP associé (c'est-à-dire la deuxième colonne)

Nous pouvons donc observer que les secteurs ruraux provoqueraient un **sur-comptage** des réseaux d'eau potable en ne considérant que les linéaires de routes, et que à l'inverse, les secteurs plus urbanisés sont plus proches de la réalité en termes de cohérence mais **sous estiment** légèrement le linéaire de réseau d'eau potable. Cela est vraisemblablement dû aux doubles canalisations présents généralement dans les secteurs où les habitations sont concentrées.

Pour tenter de réduire l'erreur, il a fallu réfléchir à un moyen de prendre en compte les routes **seulement** lorsque celles-ci permettent effectivement d'atteindre un bâtiment, ce qui induit la nécessité d'un branchement au réseau d'eau potable. La couche des bâtis de l'IGN a donc été utilisée afin de pouvoir créer des **zones tampons** (rond d'un diamètre défini avec pour centre le bâti) autour de chaque construction présente sur le territoire. Sont alors prises en compte seulement les portions de routes passant par ces zones tampons.

La Figure 5 suivante donne une idée de ce qu'on obtient grâce au traitement par tampon 50 m :

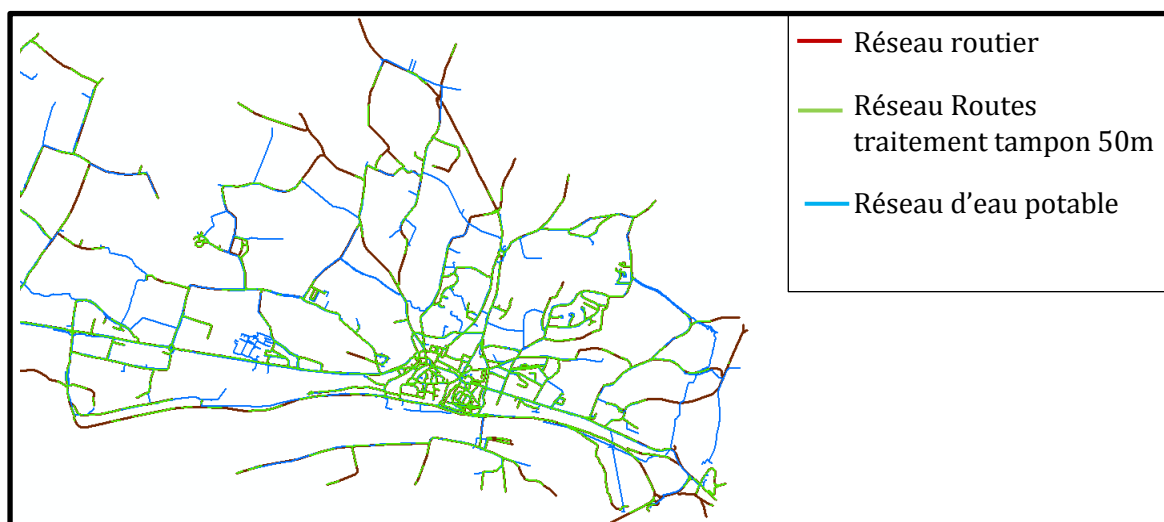


Figure 5: Comparaison des réseaux sur la Réole après traitement par tampon 50 m

Les résultats des linéaires après traitement sont présentés dans le tableau suivant :

Tableau 6: Comparaisons numériques sur le réseau de la Réole après traitement par zone tampon 50 m

Secteur	Longueur de réseau AEP sur le secteur (km)	Longueur de réseau routier sur le secteur (km)	Longueur de réseau AEP hors réseau routier	Longueur de réseau routier hors réseau AEP
La Réole traité	143,723	105,876 (87%)	53,5 (37,2%)	25,7 (17,9%)
Secteur péri-urbain +urbain traité	73,652	61,488 (83,5%)	35 (47,6%)	18,2 (24,7%)
Secteur rural traité	20,371	18,525 <b>(90,9%)</b>	12,2 (60%)	7,86 (38,6%)

Les valeurs obtenues montrent que le traitement **réduit** le taux de réseau routier compté en surplus, optimisant alors le rapport réseau routier/linéaire eau potable dans les secteurs ruraux. Le traitement induit cependant des erreurs dues au fractionnement et découpage effectués, enlevant des tronçons de routes où les canalisations doivent pourtant passer afin d'atteindre certains abonnés isolés. (Figure 6) Il serait donc intéressant de tester **plusieurs distances** de zones tampons qui pourraient prendre en compte, notamment en zone de regroupements d'habitations où ce traitement est vraiment néfaste, la présence de voisins directs. Il faudrait aussi considérer une distance qui permettrait de **traiter effectivement** la plupart des types de communes en France, dont La Réole ne peut être représentative.

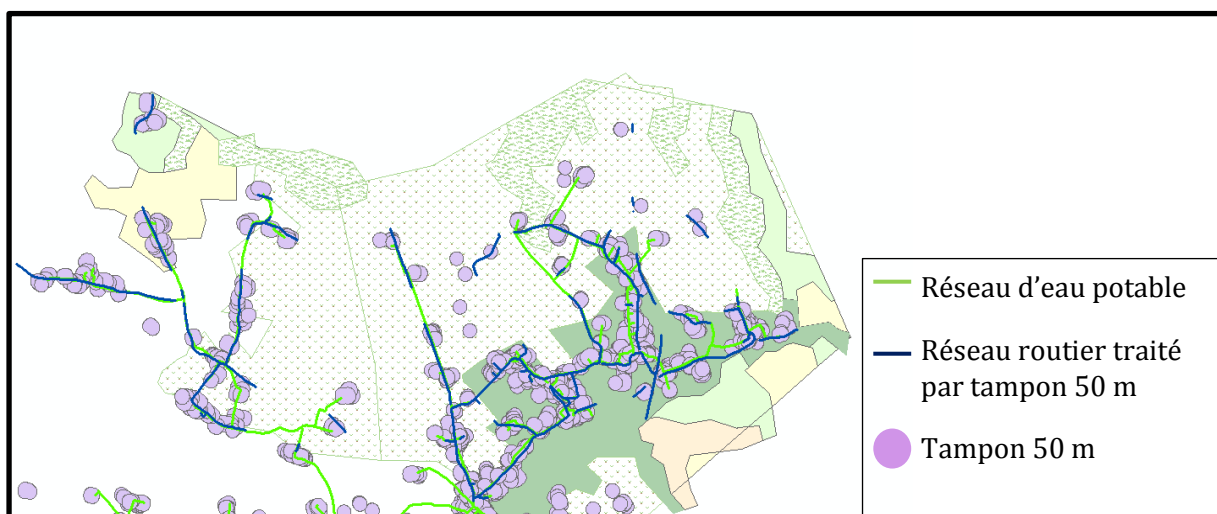


Figure 6: Effets des zones tampons sur les bâtis isolés

Les traitements primaires permettent de dégager les limites suivantes:

- La base **Corine Land Cover** contenant plus d'une centaine de catégories, un regroupement de ces dernières en moins d'une dizaine permettra de mieux observer l'occupation du sol à l'échelle nationale.
- Les routes sont aussi classées dans un nombre important de catégories dont toutes ne sont pas forcément **pertinentes** pour notre étude (Il existe par exemple beaucoup de via ferrata et de sentiers en dessous desquels il est clair qu'un réseau d'eau potable ne peut être placé). Il faudra donc effectuer un « **filtrage** » des routes avant le traitement par zones tampons.
- La **couche des bâtis** est à l'inverse peu détaillée ; certains objets telles les constructions ponctuelles ont pu être triés selon leur pertinence par rapport à un réseau d'eau potable, mais une bonne majorité des objets se trouvent dans la catégorie « Indifférencié ». Nous ne pourrions donc exploiter que la position des objets de cette couche.
- Le **traitement considéré** prenant en compte les couches précédentes incorporeront forcément une **erreur** pour les bâtis isolés de par l'enlèvement de tronçons importants.

Malgré ces limites, notamment concernant l'erreur sur les bâtis isolés, il sera intéressant de voir si l'erreur résiduelle est importante ainsi que la possibilité qu'elle soit compensée à l'échelle nationale.

Nous avons donc pu voir que le linéaire routier peut apporter une information essentielle pour l'évaluation du linéaire de conduites d'AEP, a priori assez satisfaisante sur le service de la Réole. Afin de vérifier un peu plus ces résultats, le traitement a aussi été effectué sur **deux autres services** de la Gironde (Annexe A2). Cette fois-ci, le réseau routier en zone urbaine a été compté en totalité en faisant l'hypothèse que la concentration des habitations est assez grande pour que toutes les routes soient équipées au minimum d'une canalisation. Les zones tampons n'ont été créées qu'en **zones non-urbaines** selon les données de la base Corine Land Cover.

### 3.3.2 SERVICE DE LACANAU

Le service de Lacanau a la particularité d'être situé au **bord de la plage**, et donc de présenter plus de sentiers pédestres et d'avoir un réseau plus adapté à des consommateurs saisonniers, et une certaine dispersion dans la présence des logements. Nous observons d'ailleurs que plus de la moitié des voies présentes sur le service est enlevée dès le filtrage des routes. (Figure 7, Tableau 7) La zone de tampon utilisée a aussi dû être augmentée à 200 m afin de rester cohérent avec la **dispersion** des bâtiments, la taille du service et les routes desservant la commune.

Tableau 7: Résultats des traitements primaires sur le service de Lacanau

Longueur de réseau	155,083 km
Longueur réseau routier avant traitement	556,206 km (359%)*
Longueur de routes filtrées	241,013 km (155%)
Longueur de routes après traitement (zone tampon)	186,129 km (120%)

\*Les pourcentages sont calculés par rapport à la longueur de réseau AEP

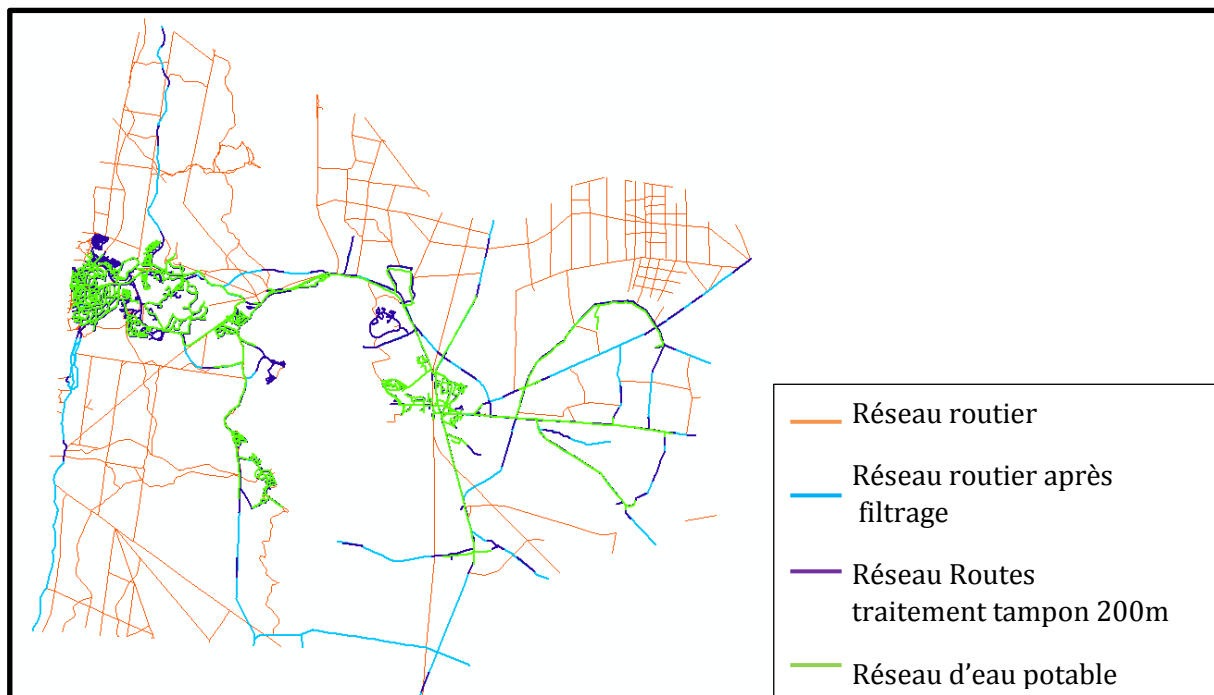


Figure 7: Comparaison des réseaux sur le service de Lacanau

Le traitement participe donc bien à l'enlèvement de tronçons clairement non-cohérents avec un réseau d'eau potable. Le traitement par zone tampon réduit l'écart entre les deux linéaires, mais à un taux encore pas assez satisfaisant (120%) et nous observons notamment par la Figure 7 que certains tronçons routiers constituant les seuls chemins d'accès sont enlevés par le traitement alors que cela n'a pas lieu d'être.

### 3.3.3 SYNDICAT DE GUITRES

Le Syndicat de Guîtres contient 13 communes et est donc caractérisé par une occupation du **sol hétérogène** (Bonzac, Bayas, Guîtres, Lagorce, Sablons, Saint-Martin-de-Laye, Saint-Martin-du-Bois, Maransin, Saint-Ciers-d'Abzac, Tizac-de-Lapouyade, Savignac-de-l'Isle, Lapouyade, Saint-Denis-de-Pile). Là encore, nous observons un écart considérable entre le linéaire routier et le linéaire de réseau d'eau potable qui diminue déjà fortement grâce au filtrage des voies qui ne peuvent pas être desservies. Le traitement par zone tampon de 200 m contribue une nouvelle fois à la réduction de l'écart entre les deux linéaires, mais ne permet pas une correspondance totale entre les deux. (Figure 8, Tableau 8)

Tableau 8: Résultats des traitements primaires sur le service de Guîtres

Longueur de réseau	401,735 km
Longueur réseau routier avant traitement	613,797 km (153%)
Longueur des routes filtrées	502,895 km (125%)
Longueur de routes après traitement (zone tampon)	449,578 km (112%)

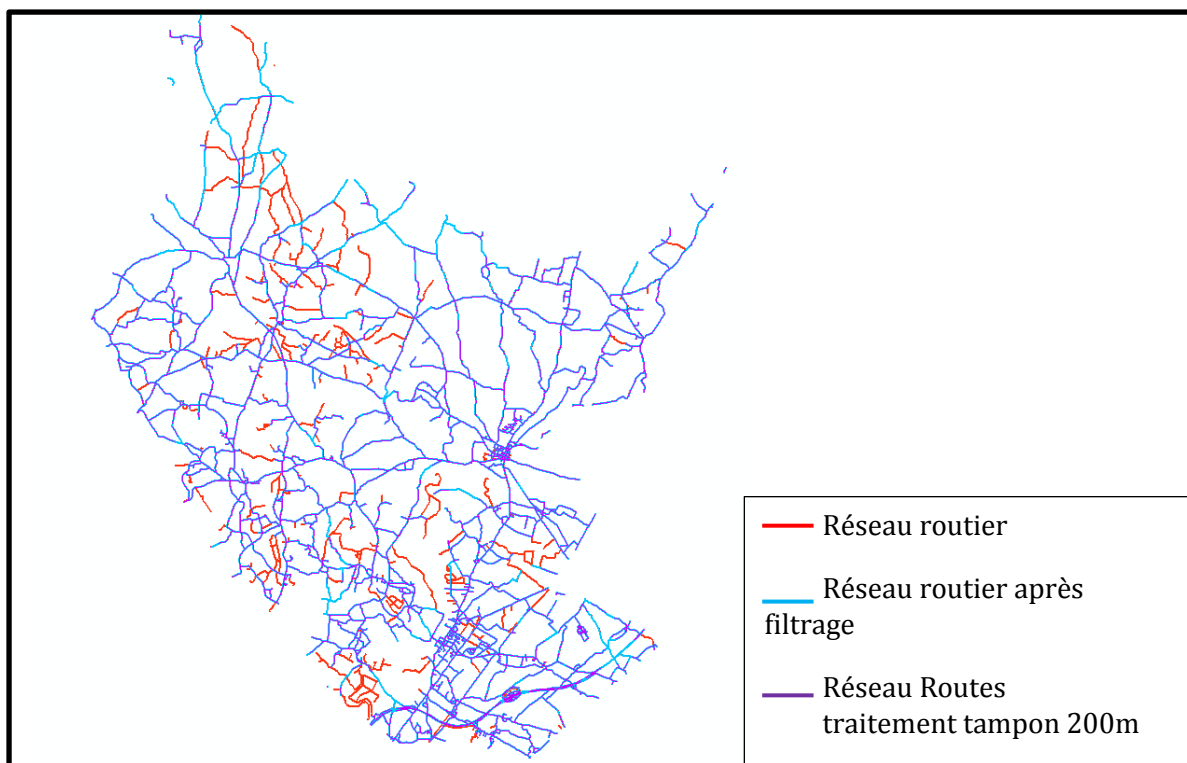


Figure 8: Comparaison des réseaux sur le service de Guîtres

Le traitement en général est donc plutôt **favorable** du point de vue de la modélisation ; en effet bien que la correspondance entre la longueur des deux réseaux ne soit pas optimale, les résultats obtenus permettent de se **rapprocher** de manière intéressante du linéaire réel présent sur chaque **service**.

Avant d'appliquer cette méthode à un plus grand nombre de services connus, il convient de se demander quel est **le rayon** de la zone tampon à appliquer pour obtenir une relation optimale entre les deux réseaux. En effet, si un traitement par zone tampon de 200 m sur un territoire de petite taille telle que la Réole n'apporte finalement pas de différence par rapport au réseau routier filtré, l'erreur créée n'est pas importante et sera probablement compensée sur d'autres services. Ainsi, il nous faut donc chercher la taille optimale de zone tampon qui nous permettra d'appliquer le traitement efficacement sur toute la France malgré les **différences démographiques** et **géographiques** des services. Nous rappelons donc que l'échelle recherchée des résultats se situe au niveau du bassin ou à l'échelle nationale ; la cohérence au niveau de la commune est donc appréciée mais pas nécessaire. Nous nous attarderons surtout sur les cohérences de linéaires par service et par département.

Des zones tampons de rayon de **50 à 300 m** ont été testées (par pallier de 50 m) afin de sélectionner le traitement permettant de se rapprocher au mieux de la réalité et en faisant l'hypothèse qu'au-delà de 600 m (distance entre deux zones tampons), le raccordement au bâti le plus éloigné aura été desservi par un autre chemin (passant sous un champ par exemple).

### 3.4 COMBINAISON AVEC LES BASES SISPEA

Il s'agit ensuite de **recouper** les résultats de traitement de réseau routier avec les autres caractéristiques des communes sur un grand échantillon afin d'obtenir les plus d'informations possible pour caractériser chaque service.

Les données des bases ont été extraites à partir de la **base de données SISPEA** sur les années 2009 à 2013. Cette base de données, renseignée tous les ans par seulement environ 25% des services d'eau, nous permet en effet de récupérer les **caractéristiques à l'échelle des services** qui serviront à caler nos modèles. Un service est constitué de communes ou de parties de communes, ainsi son **périmètre n'est pas défini** et exploitable sous forme de couche SIG. L'échelle retenue pour l'agrégation des couches SIG (routes, bâtis...) est donc la commune. Les modèles créés se baseront cependant sur des paramètres à **l'échelle communale** extraits des SIG qu'il faudra ensuite regrouper suivant la composition communale des services afin de permettre la comparaison aux linéaires réels extrait de la base SISPEA. L'**extrapolation** à l'échelle nationale se fera donc elle aussi à **l'échelle communale**, ce qui requerra que les modèles calés sur les services des bases créées soient **vérifiés** sur les communes associées.

Afin de **caler les modèles** que nous désirons mettre en place, nous avons créé **3 bases** qui regroupent des services où nous pourrions accéder au linéaire de réseau d'eau potable.

#### 3.4.1 PREMIERE BASE : COMMUNES MONO-SERVICES

Nous désirons une première base de calage dont les données sont connues parfaitement afin de ne permettre aucun biais initial dans les résultats. Pour ce faire, nous avons décidé de ne retenir pour cette base que les communes appartenant **entièrement** à un seul service, lorsque ce service est effectivement **renseigné en linéaire réel et en nombre d'abonnés**. Les données de routes traitées des communes de cette base pourront donc être comptées en totalité pour constituer le linéaire de réseau modélisé sur le service auquel elles appartiennent.

Afin de gagner un maximum de données sur les données de 2009 à 2013, nous avons utilisé une méthode de **retour en arrière** permettant d'enlever les services contenant des communes ayant changé d'adhésion lors d'une année plus récente. Ainsi les caractéristiques des communes et services retenus seront les plus récents renseignés dans notre base.

Le détail des étapes suivies afin de constituer cette base sous le logiciel Access (Microsoft) se trouve en annexe A3. Le nombre de services retenus dans cette base s'élève alors à **5043 services** regroupant **13 442 communes (37% des communes de France)**.



### 3.4.2 DEUXIEME BASE : MEGA-SERVICES

La deuxième base a été construite en complément de la première base et a par la suite été utilisée afin de valider les modèles calés sur la première base. Lorsqu'une commune est sur **plus d'un service à la fois**, il est possible de regrouper certains services entre eux si :

- Toutes les autres communes appartenant aux mêmes services font **intégralement** partie du groupement de services
- Tous les services du groupement sont **renseignés** en linéaire de réseau réel et d'abonnés

Le grand pseudo-service résultant sera appelé **méga-services**. Pour faire partie de cette base, il est donc indispensable que tout le **périmètre** de chaque regroupement de services soit entièrement connu.

Les étapes suivies afin de créer cette base sont résumées dans l'annexe A4. Le nombre de méga-services ainsi obtenus s'élève à 104 méga-services regroupant **285 services** et **1776 communes (5% des communes de France)**.

### 3.4.3 TROISIEME BASE : COMMUNES MULTI-SERVICES

La dernière base a été créée à partir des tous les services renseignés qui ne satisfont pas les conditions pour intégrer les deux premières bases. Les services de cette base sont donc composés de communes qui appartiennent aussi à au moins un autre service non-renseigné. De ce fait, il est impossible de connaître **exactement** le périmètre des communes contenues dans chaque service ; cette base servira seulement à confirmer les **ordres de grandeur** des linéaires modélisés par rapport aux linéaires réels des services.

La base retenue contient **1482 services** regroupant **10 340 communes**. La base représente environ **28% des communes en France**.

Le schéma suivant résume la logique selon laquelle les trois bases ont été construites :

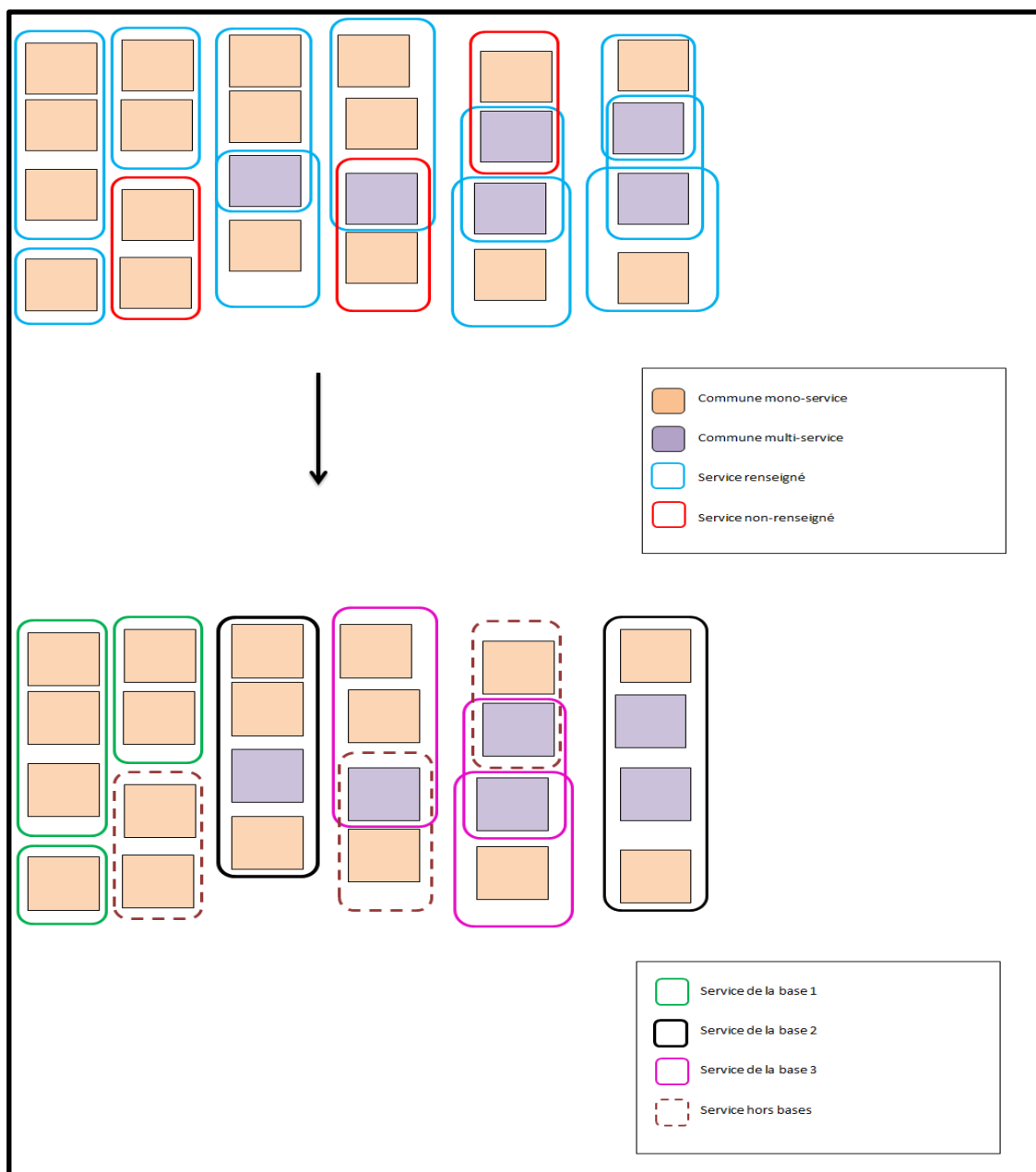


Figure 9 : Schéma de regroupement des méga-services

### 3.5 CONSTRUCTION DES PREMIERS MODELES

Les données sur chaque base sont ensuite **combinées** avec la somme des linéaires de routes traitées des communes appartenant à chaque base. (Figure 10)

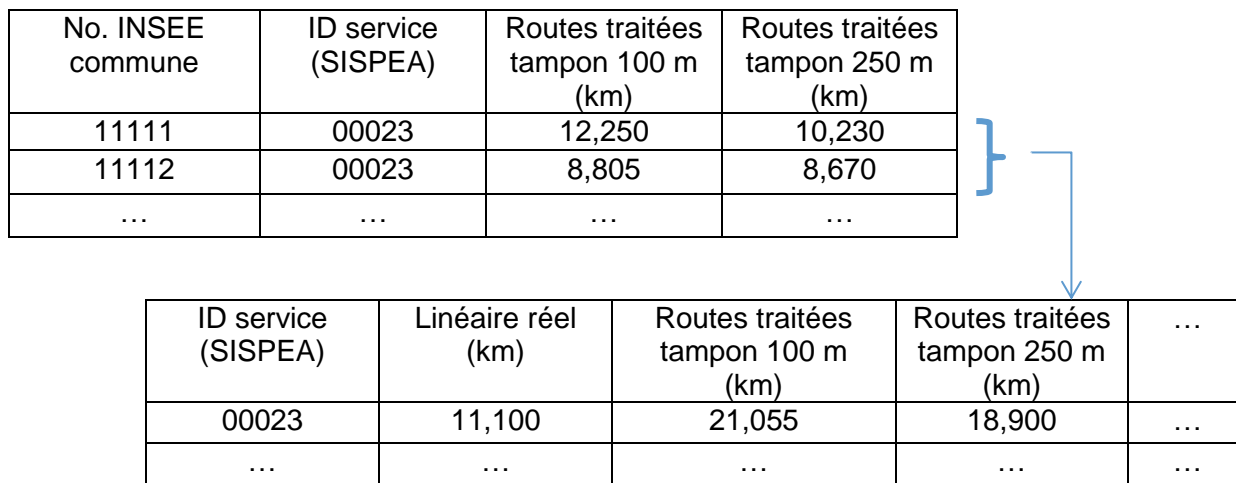


Figure 10 : Schéma des regroupements de données de la base SISPEA

Les routes ainsi traitées sont alors **comparées aux linéaires réels** sur les services retenus afin d'observer l'effet des traitements. Les graphiques suivants donnent les premières tendances obtenues avec les zones tampons de différentes tailles lorsque celles-ci ont été appliquées aux données obtenues sur le département de **la Gironde** contenant 92 services de la première base. En effet, les traitements étant longs, les résultats ont d'abord été observé sur ce département afin d'éviter des calculs inutiles sur le reste de la France.

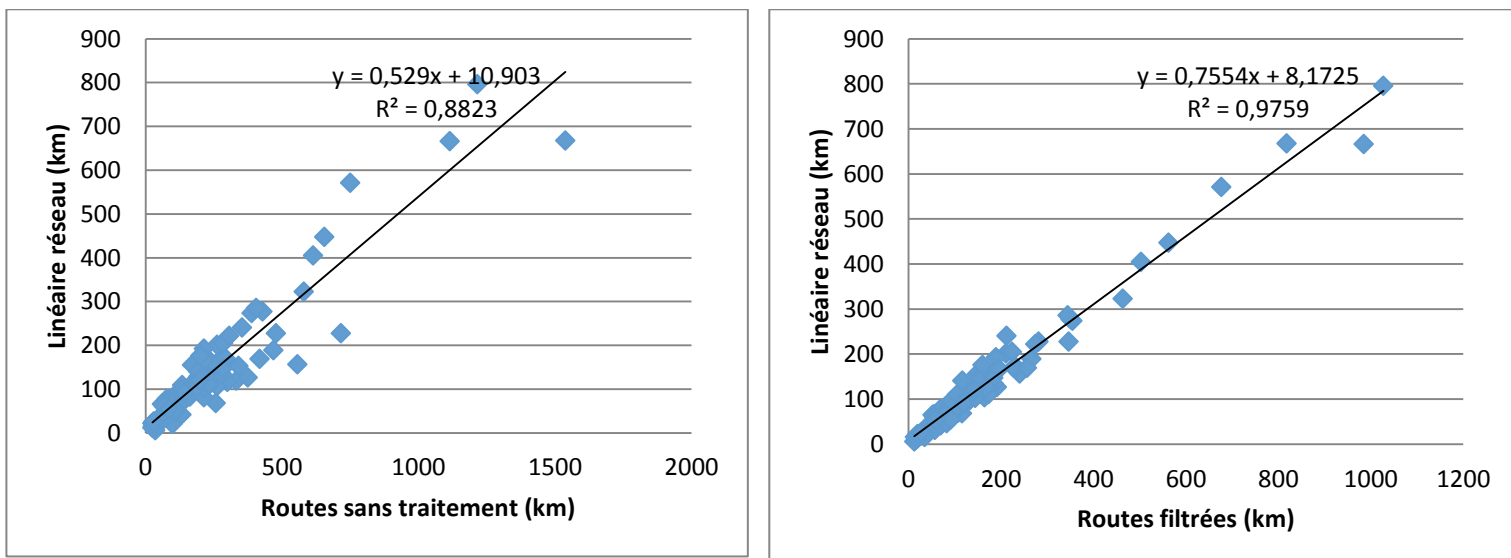


Figure 11: Comparaison routes-réseaux eau potable sur la Gironde (1)

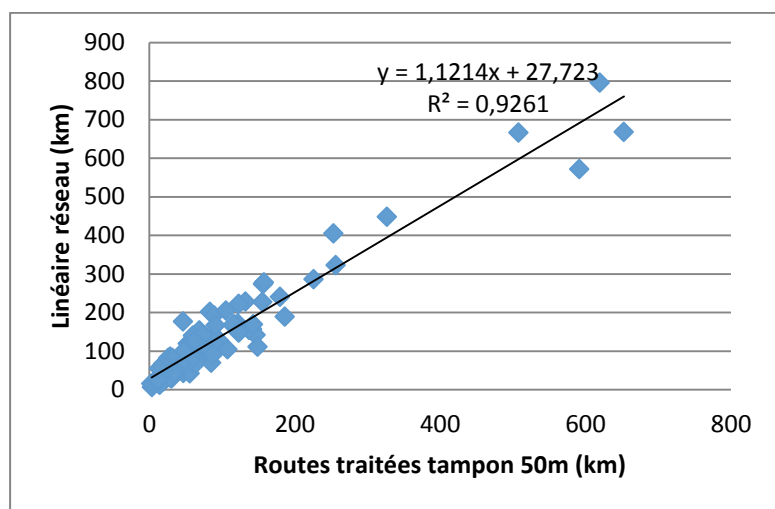
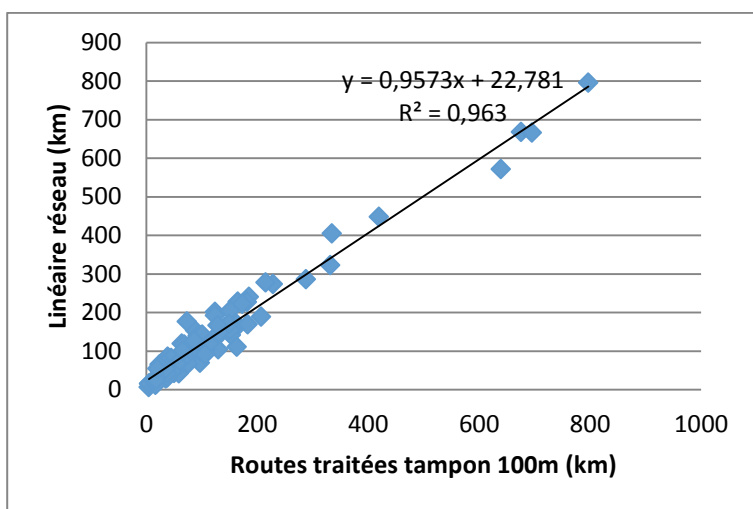
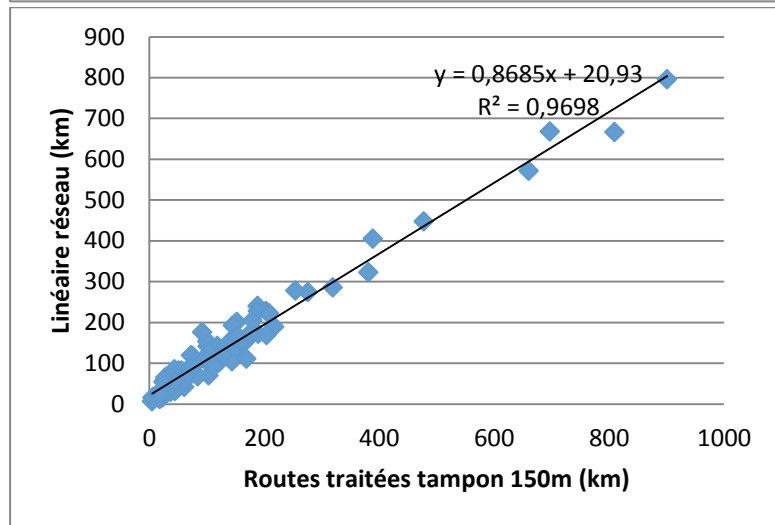
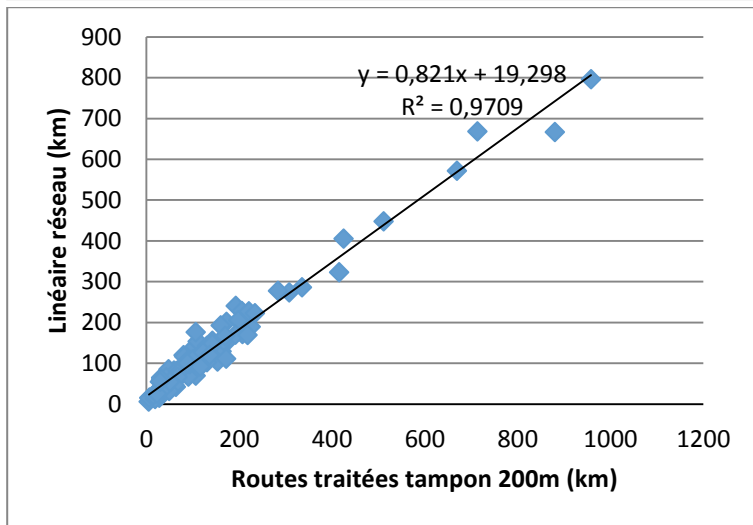
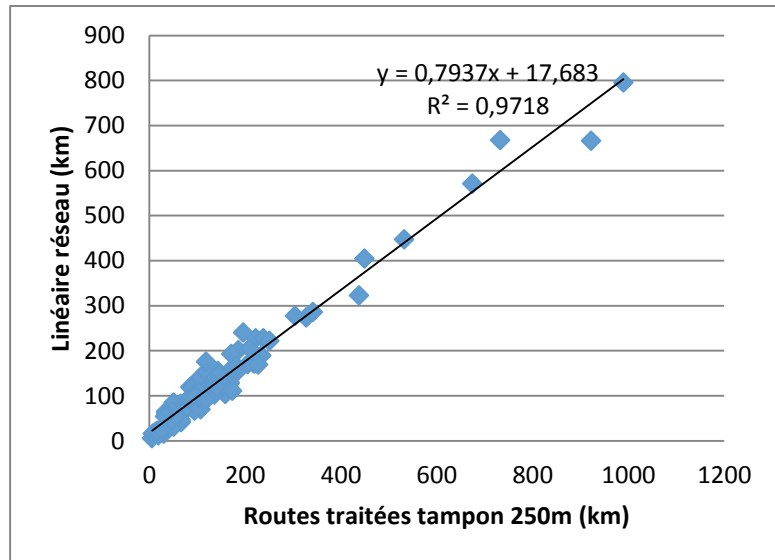
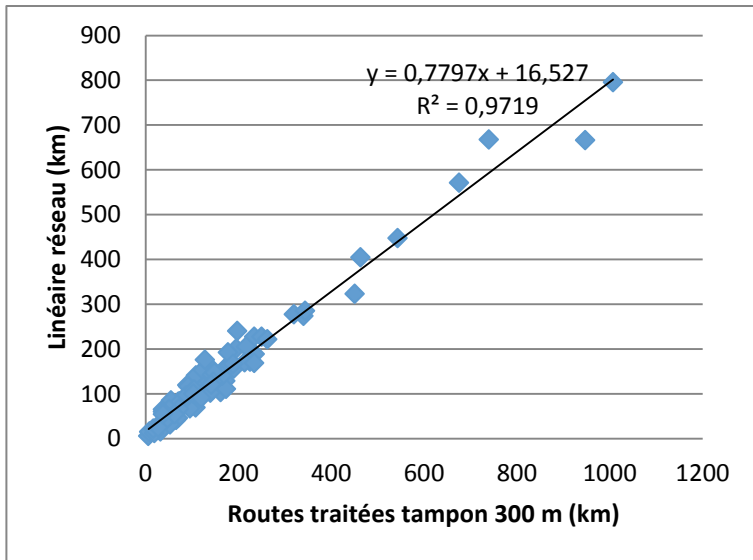


Figure 12: Comparaison routes-réseaux eau potable sur la Gironde (2)

Nous observons dans un premier temps que les **traitements** restent intéressants à l'échelle du département par rapport aux linéaires de routes brutes lorsque comparés aux linéaires de réseau d'eau potable réels. Nous pouvons aussi voir qu'à l'échelle de la Gironde, les tailles de zones tampons entre **100 et 300 m** permettent toutes d'obtenir des résultats similaires.

Les traitements ont donc ensuite été appliqués à **9 régions contenant 2335 services** de la première base. Les 9 régions choisies (Alsace, Aquitaine, Auvergne, Bourgogne, Bretagne, Corse, Languedoc-Roussillon, Poitou-Charentes et Rhône-Alpes) ont des **caractéristiques hétérogènes** en termes d'occupation du sol, de position géographique et de démographie.

Les figures 13 à 16 suivants résument les corrélations obtenues d'abord sur les routes brutes, puis après le filtrage, et enfin sur les routes traitées par tampon 100m et 250m, les zones tampons offrant les **meilleurs coefficients** de détermination R<sup>2</sup>.

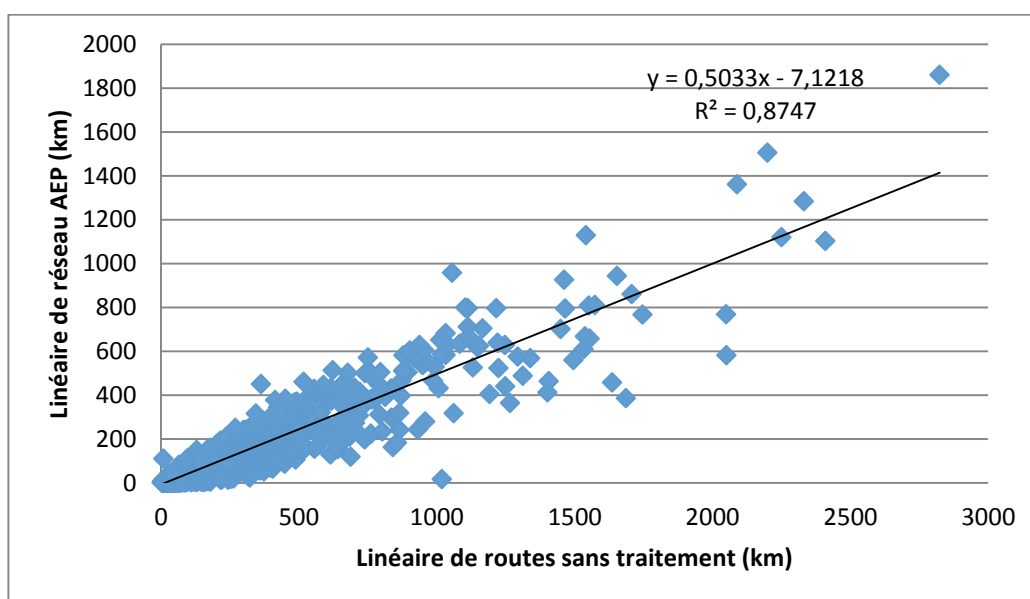


Figure 13: Corrélation entre Routes brutes et linéaires de réseaux

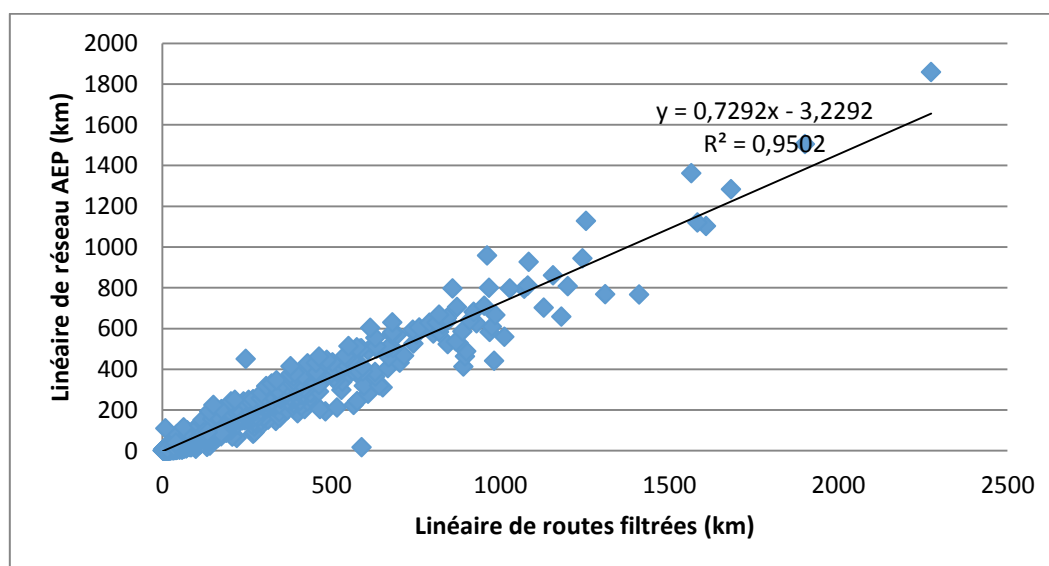


Figure 14: Corrélation entre Routes filtrées et linéaires de réseaux

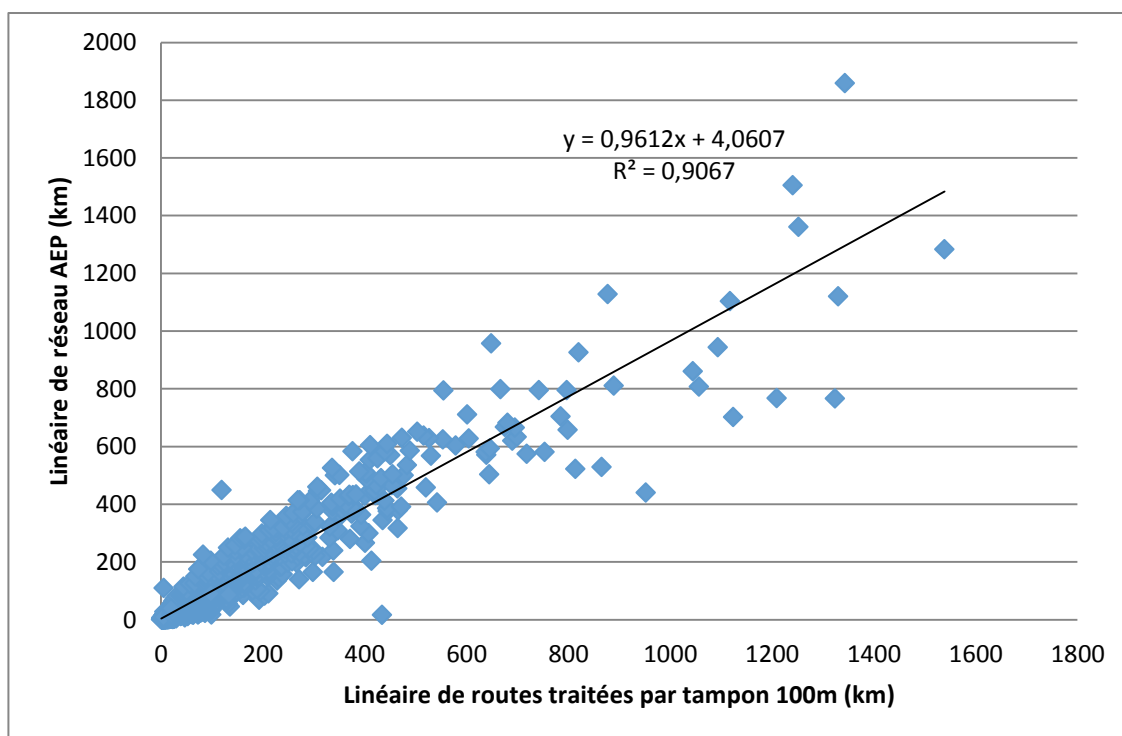


Figure 15: Corrélation entre Routes traitées par tampon 100 m et linéaires de réseaux

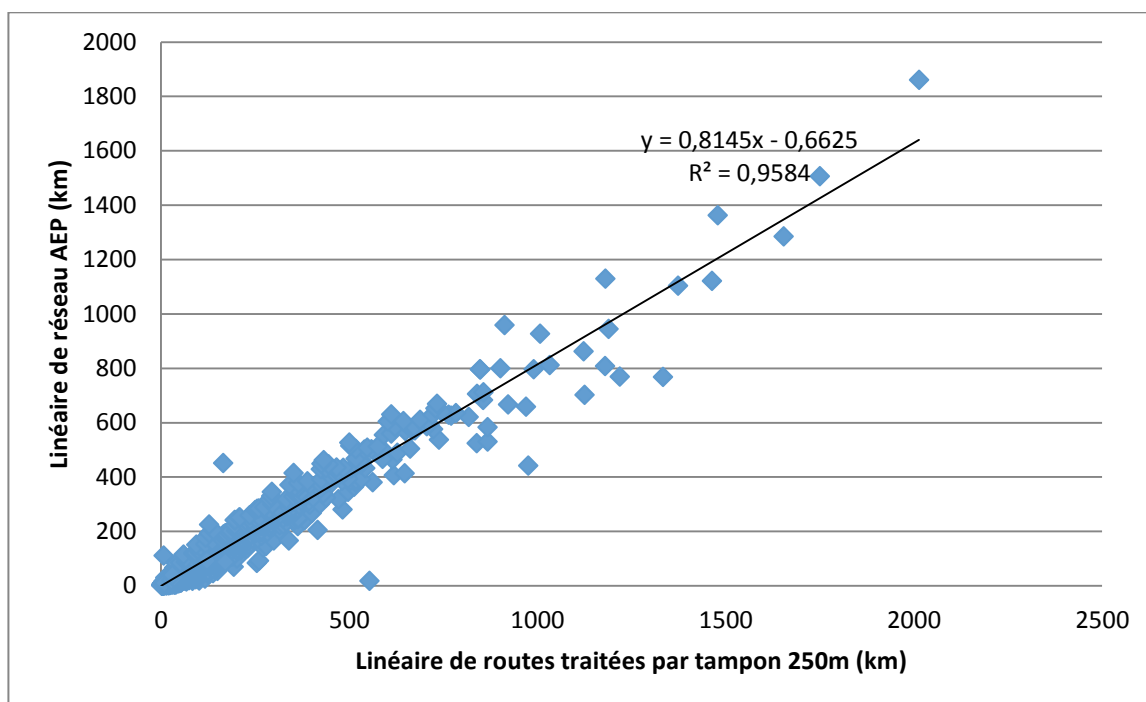


Figure 16: Corrélation entre Routes traitées par tampon 250 m et linéaires de réseaux

Les résultats observés par région ainsi que sur la totalité des 2335 services nous permettent de choisir le **tampon 250 m** comme meilleure taille pour notre traitement. Dans la suite, seule cette taille a été appliquée sur les régions restantes pour une question de gain de temps de traitement.

Le Tableau 9 suivant donne alors la répartition des coefficients de corrélations ainsi que les coefficients multiplicateurs associés pour les graphiques obtenus par région sur les **5043 services de la première base** (France entière) uniquement. Il a été choisi pour les courbes de tendance de **forcer l'interception** à l'origine afin de simplifier l'application sur les communes lors de l'estimation nationale. Les courbes restent cependant cohérentes et les coefficients de détermination varient très peu par rapport à ce changement (variation inférieur à 1% dans les valeurs de R<sup>2</sup>).

Tableau 9: Coefficients obtenus après traitement par zones tampon sur chaque région

	Tampon 250	
	a	R <sup>2</sup>
Région Alsace	0,7781	0,9394
Région Aquitaine	0,8794	0,9609
Région Rhône Alpes	0,7653	0,9314
Région Poitou Charentes	0,8071	0,959
Région Auvergne	0,9324	0,9733
Région Bourgogne	0,8372	0,9702
Région Bretagne	0,8399	0,9567
Région Corse	0,7537	0,9802
Région Languedoc Roussillon	0,6354	0,9678
Région Basse Normandie	0,9001	0,9811
Région Limousin	0,9632	0,915
Région Champagne-Ardenne	0,8776	0,9619
Région Centre	1,0152	0,9219
Région Picardie	0,8218	0,9694
Région Pays de la Loire	0,8162	0,9218
Région Nord Pas de Calais	0,8497	0,984
Région PACA	0,5999	0,717
Région Franche Comte	0,6748	0,9758
Région Midi-Pyrénées	0,935	0,9169
Région Lorraine	0,8867	0,9503
Région Ile-de-France	0,8036	0,8457
Région Haute Normandie	0,8623	0,9857
France entière	0,8337	0,9562

La **variation importante** des coefficients de détermination et les résultats moins consistants sur certaines régions nous indiquent que les coefficients par région ne constituent pas les meilleurs modèles à considérer pour estimer le linéaire national. Les sommes se compensent cependant quelque peu sur les services considérés sur la base entière avec un R<sup>2</sup> supérieur à 0,95. Pour notre premier modèle, nous nous contenterons donc d'utiliser le coefficient calé de la table précédente sur la France entière.

Tableau 10: Résultats du premier modèle

a	R <sup>2</sup>	Estimation nationale (km)
0,8337	0,9562	883985,23

Ce modèle présente cependant plus d'un biais sur ses résultats. Non seulement il ne corrige pas les **déviations** d'une partie des résultats surlignés dans le Tableau 9, mais il les masque aussi. Nous observons aussi une certaine **surestimation** sur les petits services qui rendra forcément l'intervalle de confiance du modèle trop important.

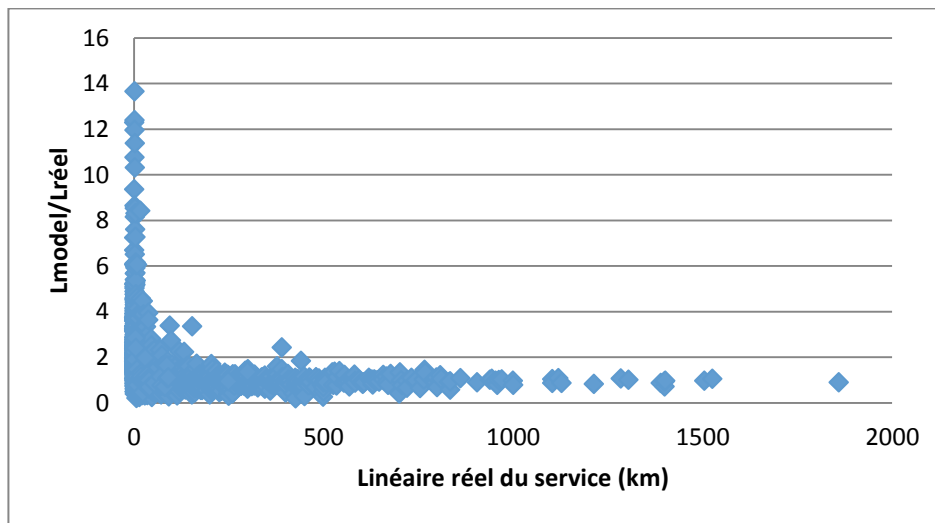


Figure 17: Comparaison des linéaires modélisés par rapport aux linéaires réels du premier modèle selon le linéaire réel de chaque service

Sur la Figure 17, nous observons bien une très grande amplitude relative des erreurs sur les plus petits services qui s'estompent quand la longueur du réseau augmente. Pour que l'erreur relative soit inférieure à 40% entre le linéaire routier et le linéaire de réseau d'eau potable réel, il ne faudrait considérer que les réseaux de longueur supérieure à 15 km.

### 3.6 AFFINAGE DU MODELE LINEAIRE

Pour remédier aux lacunes du premier modèle, nous avons tenté de créer des **classes de services** et ainsi appliquer un deuxième traitement au réseau linéaire routier pris en considération.

#### 3.6.1 METHODE DE CLASSIFICATION PAR TAILLE DE POPULATION

Nous avons premièrement vérifié si la taille du service termes de en population joue un rôle conséquent dans la corrélation entre les linéaires de réseaux et les caractéristiques calculées précédemment.

Un premier regroupement rapide par classes de population permet d'observer que la majeure partie des services (80%) ont une population de moins de 5000 habitants et que les taux de remplissage (proportion de services utilisés dans la base par rapport aux services présents sur le territoire) sont en général plus élevés sur les classes intermédiaires (taux de 40% à 60% pour les populations entre 5000 et 100 000 habitants, pour environ 20% sur les petits services de moins de 5000 habitants).

Nous avons ensuite observé la variation de l'effet de nos traitements sur les routes par rapport au linéaire réel de réseau d'eau potable selon la taille de la population. Nous avons calculé la **penne glissante** des linéaires de réseaux d'eau potable par rapport à la longueur de routes traitées par **tampon 250 m** sur une base de 1000 points par penne. (Figure 18)



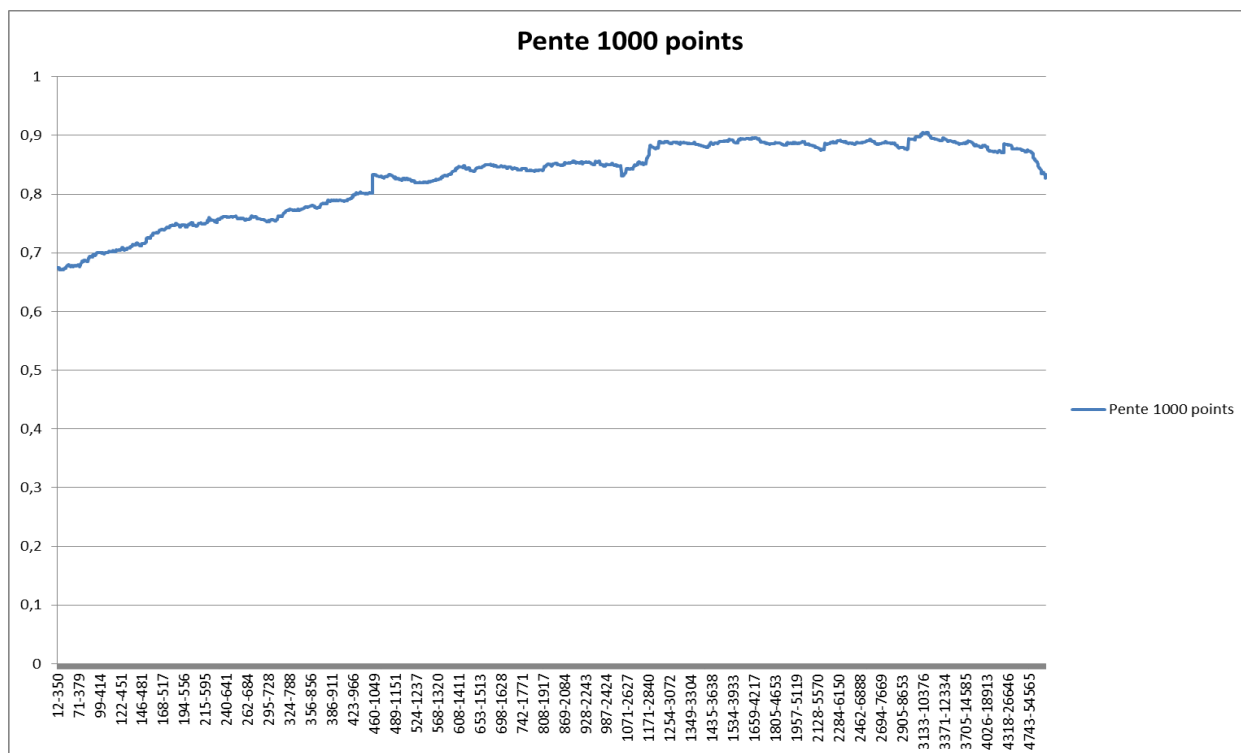


Figure 18: Pente tous les 1000 points des linéaires de routes traitées par tampon par rapport au linéaire réel

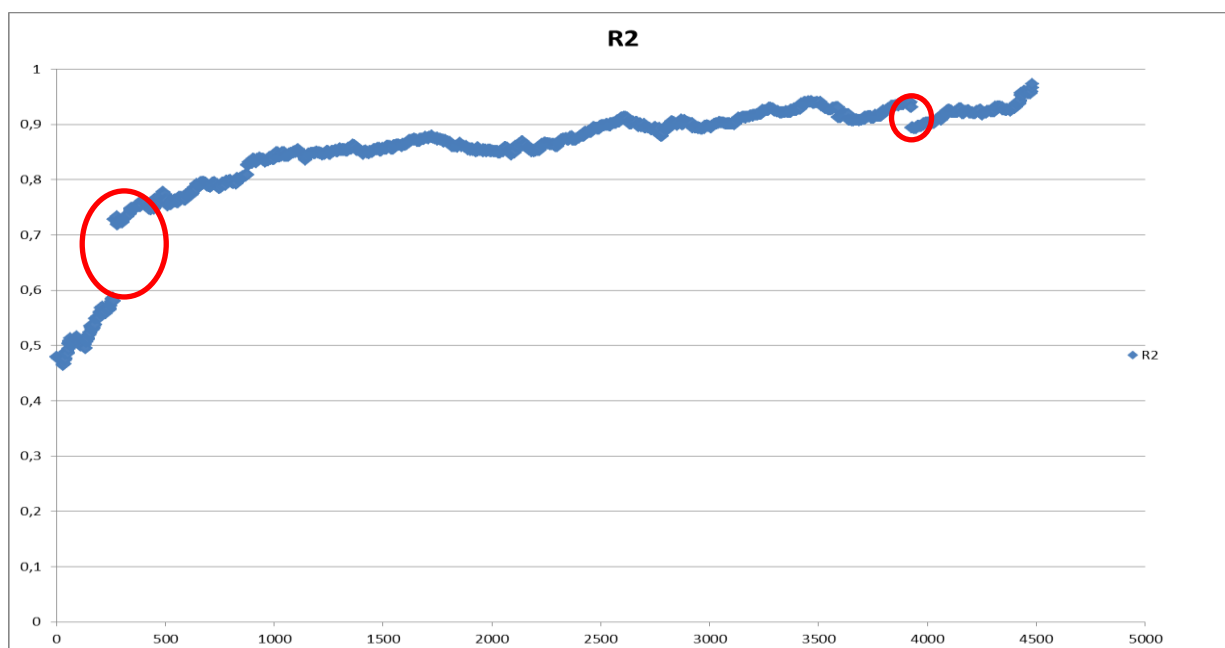


Figure 19: R<sup>2</sup> des valeurs de pentes obtenues tous les 1000 points

Les zones en rouge sur la [Figure 19](#) désignent des sauts dans la continuité des résultats trouvés, et montrent que certaines erreurs subsistent dans la base de données. Les points concernés ont été recherchés dans la base et corrigés (puisque l'on observe que ce sont des données qui ont très probablement des fautes de frappe) selon le [Tableau 11](#) suivant.

Tableau 11: Erreurs corrigées dans la première base

ID SISPEA du service	Population du service	L <sub>aep</sub> base SISPEA	L <sub>aep</sub> Rectifié
78167	126	110	11
54172	1045	500	50
116305	9122	1300	300

Nous obtenons alors les figures suivantes :

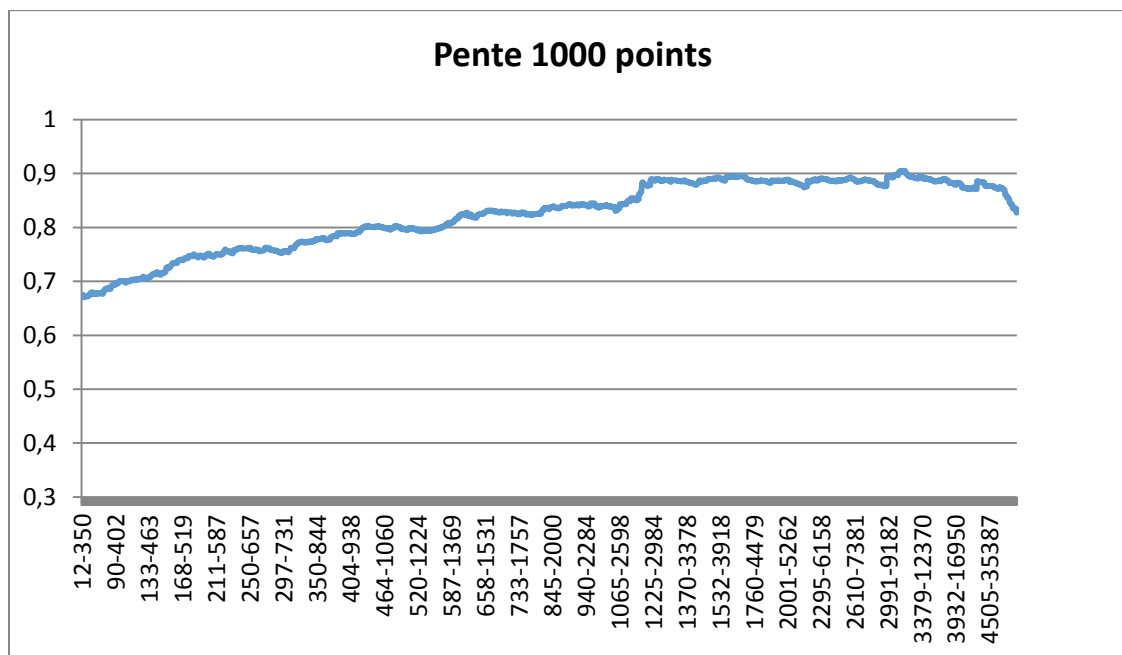


Figure 20: Courbe des pentes à 1000 points après correction

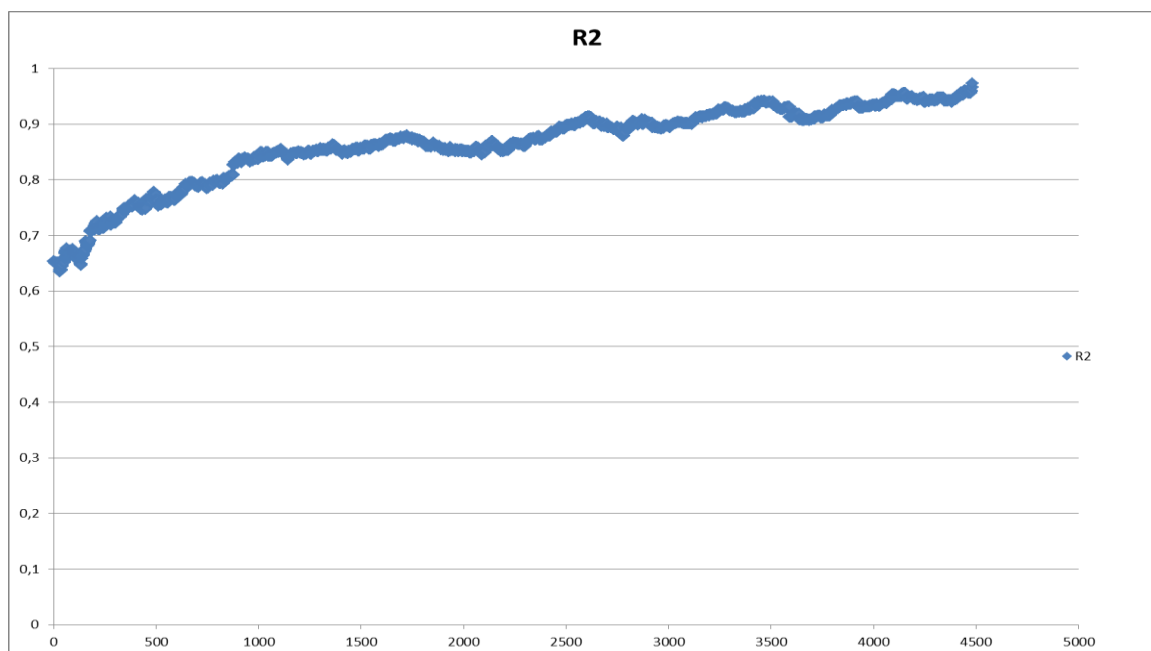


Figure 21: Courbe des R<sup>2</sup> pour les pentes à 1000 points après correction

Nous observons que les pentes ont tendances à augmenter dans les services de petite taille et à se stabiliser sur les plus grands services.

Une répartition de la **population en classes** est effectuée afin que chaque classe contienne **un effectif de 500 services**. Nous obtenons alors pour les services renseignés de notre **première base**, les limites suivantes :

Tableau 12: Classes de populations brutes obtenues

Classe 1 : 12- 193 hab	Classe 2 : 194-350 hab	Classe 3 : 351-551 hab
Classe 4 : 552-843 hab	Classe 5 : 844-1290 hab	Classe 6 : 1291-1995 hab
Classe 7 : 1996-3159 hab	Classe 8 : 3160-5255 hab	Classe 9 : 5256-10230 hab
Classe 10 : 10230+ hab		

Les pentes sont alors calculées sur chaque classe pour obtenir les valeurs suivantes :

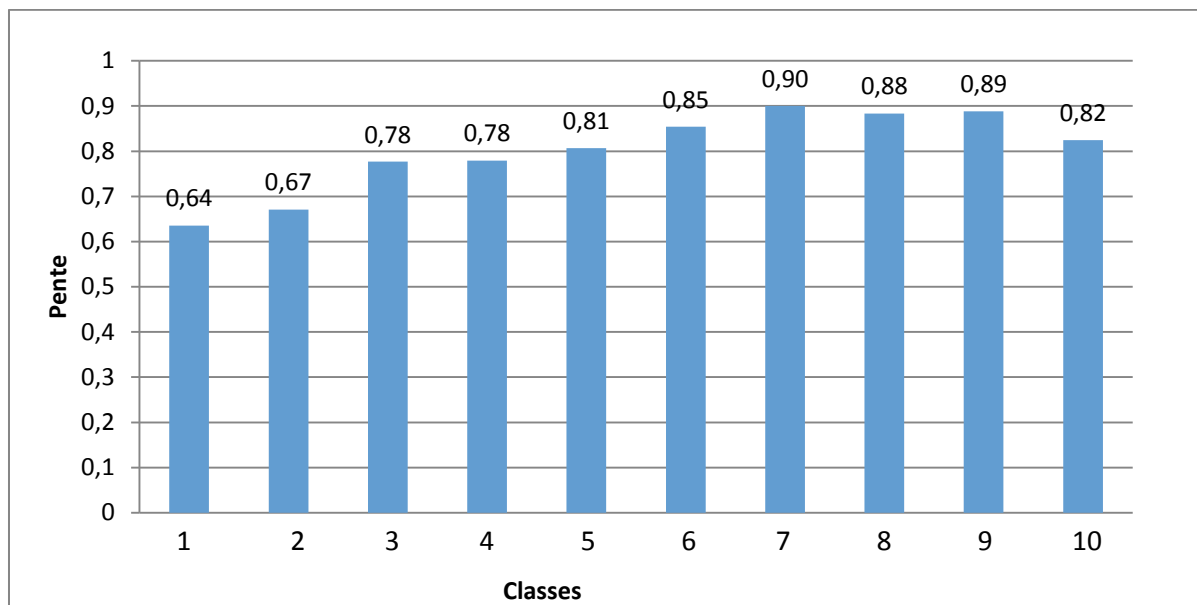


Figure 22: Pentes à appliquer pour les classes de populations brutes

Nous observons alors qu'il est possible de **combiner les classes** construites deux à deux afin d'obtenir des classes contenant 1000 services.

Tableau 13: Classes de populations retenues

Classe 1 : 12-350 hab	Classe 2 : 351-1000 hab	Classe 3 : 1001-2000 hab
Classe 4 : 2001-5000 hab	Classe 5 : 5001+ hab	

Les pentes retenues sont celles calculées sur chaque nouvelle classe :

Tableau 14: Pentés associées aux classes retenues

Classe	Habitants	Pente	R <sup>2</sup>
1	12-350	0,6746	0,7452
2	351-1000	0,7960	0,8940
3	1002-2000	0,8377	0,9130
4	2001-5002	0,8854	0,9280
5	5002-1077200	0,8337	0,9599

Nous appliquons alors les pentes aux **communes** (les routes par communes généralement la seule information accessible sur les réseaux non renseignés) selon la classe de population comme calculée précédemment. Ceci permet d'avoir une première estimation du linéaire à l'échelle nationale.

Le linéaire français des réseaux d'eau potable serait alors (par cette modélisation) de **852 692 km**.

Il faut cependant se demander si les pentes calées sur les services renseignés peuvent être appliquées directement sur les communes, car celles-ci **n'appartiennent pas forcément à la même catégorie** que celle où se trouve le/les services qui leur sont associés. Par exemple, un service de 10 000 habitants composé de 10 communes de 1000 habitants chacune se verra attribuer une pente différente que celles appliquées sur les communes qui lui sont associées séparément.

Nous voudrions donc réaliser un autre modèle avec un indicateur différent et qui permettrait des corrélations plus satisfaisantes.

### 3.6.2 METHODE DE CLASSIFICATION PAR DENSITE

Les différents indicateurs testés sont :

- Indic1 :  $Densité_{pop} = \frac{Population}{Superficie\ du\ territoire}$
- Indic2:  $\frac{Surface\ urbaine}{Superficie\ du\ territoire}$
- Indic3:  $\frac{Routes\ filtrées}{Superficie\ du\ territoire}$
- Indic4:  $\frac{Population}{Routes\ filtrées}$
- Indic5:  $\frac{Population}{Routes\ traitées\ tampon\ 250m}$

De même que pour les populations, ces indicateurs ont été calculés sur les services de la première base, puis les graphiques de pentes glissantes à 1000 points ont été tracés. (Figure 23) Les pentes obtenues et R<sup>2</sup> associés sont présentés pour chaque classe dans le tableau suivant par indicateur :

Tableau 15: Pentes obtenues sur les nouveaux indicateurs

	Indic1		Indic2		Indic3		Indic4		Indic5	
	Pente 1000 pts	R <sup>2</sup>	Pente 1000 pts	R <sup>2</sup>	Pente 1000 pts	R <sup>2</sup>	Pente 1000 pts	R <sup>2</sup>	Pente 1000 pts	R <sup>2</sup>
Classe 1	0,953	0,964	1,034	0,934	1,024	0,937	0,896	0,930	0,948	0,957
Classe 2	0,923	0,957	0,924	0,959	0,963	0,932	0,910	0,950	0,873	0,982
Classe 3	0,869	0,977	0,860	0,957	0,898	0,974	0,851	0,967	0,829	0,930
Classe 4	0,791	0,953	0,773	0,942	0,830	0,954	0,803	0,942	0,786	0,936
Classe 5	0,793	0,969	0,799	0,969	0,790	0,965	0,797	0,973	0,799	0,974

L'indicateur de **densité de population** (Indic 1) présente les meilleures corrélations sur la majorité de l'échantillon avec des coefficients multiplicateurs variant moins d'une taille de densité à l'autre. Nous nous sommes donc concentrés sur cette variable pour la suite du deuxième modèle.

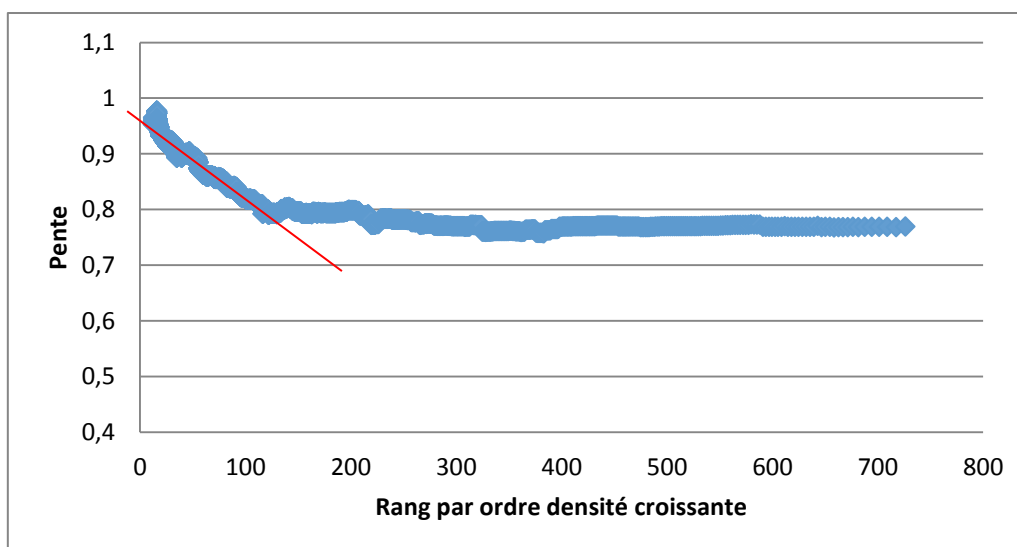


Figure 23: Pentés tous les 1000 points en prenant en compte la densité des services

Nous remarquons une tendance **décroissante** au début de la distribution, traduisant une dépendance du linéaire par rapport à la densité de population. Cette dépendance décroissante est suivie d'un **premier plateau** entre les rangs 150 et 250, puis d'un autre **plateau** à partir des pentes suivantes, traduisant des variations à coefficients constants sur les services concernés. Une recherche des points de changement de tendance conduit à obtenir les classes de densités suivantes : **[0-120 hab/km<sup>2</sup>, 120-400 hab/km<sup>2</sup>, 400+ hab/km<sup>2</sup>]**

À partir des graphiques ainsi construits, nous obtenons alors des tendances pour chaque classe, selon le tableau suivant :

Tableau 16: Modélisation par classe de densité

Classe	Modélisation
Classe 1 : 0 à 120 hab/km <sup>2</sup>	$L_{aep} = (1 - \text{Densité} * 0.0017) * L_{R250}$
Classe 2 : 120 à 400 hab/km <sup>2</sup>	$L_{aep} = 0,796 * L_{R250}$
Classe 3 : > à 400 hab/km <sup>2</sup>	$L_{aep} = 0,761 * L_{R250}$

Note :  $L_{R250}$  = Longueur des routes traitées par tampon 250 m

Le linéaire français des réseaux d'eau potable serait alors (par cette modélisation) de **925 559 km**.

### 3.7 APPLICATION SUR LES COMMUNES DES SERVICES

#### Classes de population : Première base

Il convient de vérifier que les modèles créés soient applicables directement sur les communes, bien que calés sur les services, sans introduire de biais important. En effet, le modèle final sera appliqué sur la base nationale de communes puisque les données de services ne sont pas complètes sur la France entière.

Sur le modèle par classes de populations par exemple, les pentes sont d'abord appliquées sur l'ensemble des services de la première base par classes de populations, et la somme des linéaires obtenue est comparée à la somme de linéaire réel des réseaux d'eau potable. Nous obtenons alors une somme de linéaire modélisée égale à 100,5% du réseau linéaire réel sur la **première base**; le modèle surestime donc légèrement le linéaire réel.

Les pentes sont alors à nouveau appliquées cette fois-ci sur les communes de chaque service par rapport à la population de chaque commune. Les linéaires obtenus sont alors sommés par service. Ces linéaires sont ensuite comparés aux linéaires calculés directement sur les services ainsi qu'aux linéaires réels sur la **première base**.

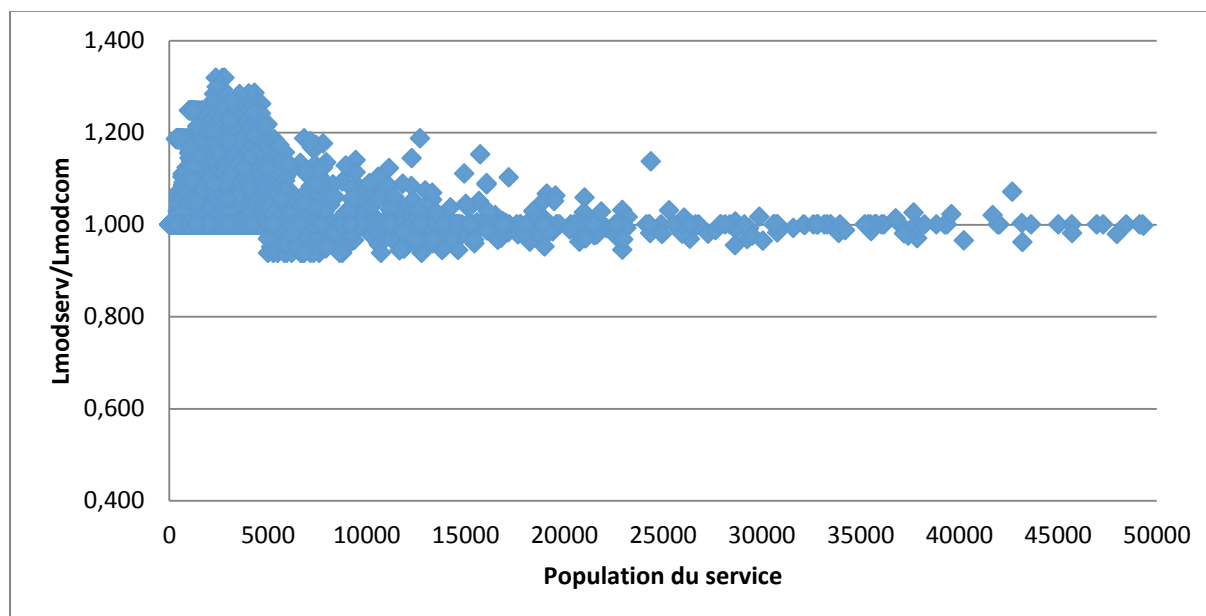


Figure 24: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la première base

Le Figure 24 ci-dessus montre la tendance des rapports de linéaires avec application des pentes sur les services, et les linéaires avec application des pentes sur les communes. Nous observons que pour les **petits services** (moins de 5000 habitants), il y a **surestimation** sur une majeure partie des valeurs de linéaires entre le calcul sur le service et celui sur les communes associées, avec une surestimation maximale de 32%. Pour les plus grands services, le rapport des linéaires oscille en général autour de 1.

En faisant la somme des linéaires avec les pentes appliquées sur les communes, nous obtenons un linéaire total égal à 97,4% de la somme de linéaire réel, soit un écart de 3,1% avec le calcul effectué sur les services.

### Classes de population : Deuxième base

Afin de vérifier si le modèle reste efficace en dehors de notre base de données de calage (première base), les pentes sont testées sur les **méga-services** (deuxième base) ainsi que sur les services renseignés (troisième base) mais ne contenant pas de communes mono-services ou ne pouvant pas former de méga-services. Sur les méga-services (deuxième base), la variation de  $L_{\text{mod.serv}}/L_{\text{mod.com}}$  suit généralement la même **tendance** que sur les services de notre base de calage. La somme des linéaires modélisés sur les services puis sur les communes de la première base sont respectivement de 96,1% et 93,9% par rapport à la somme des linéaires réels sur les services. (Figure 25)

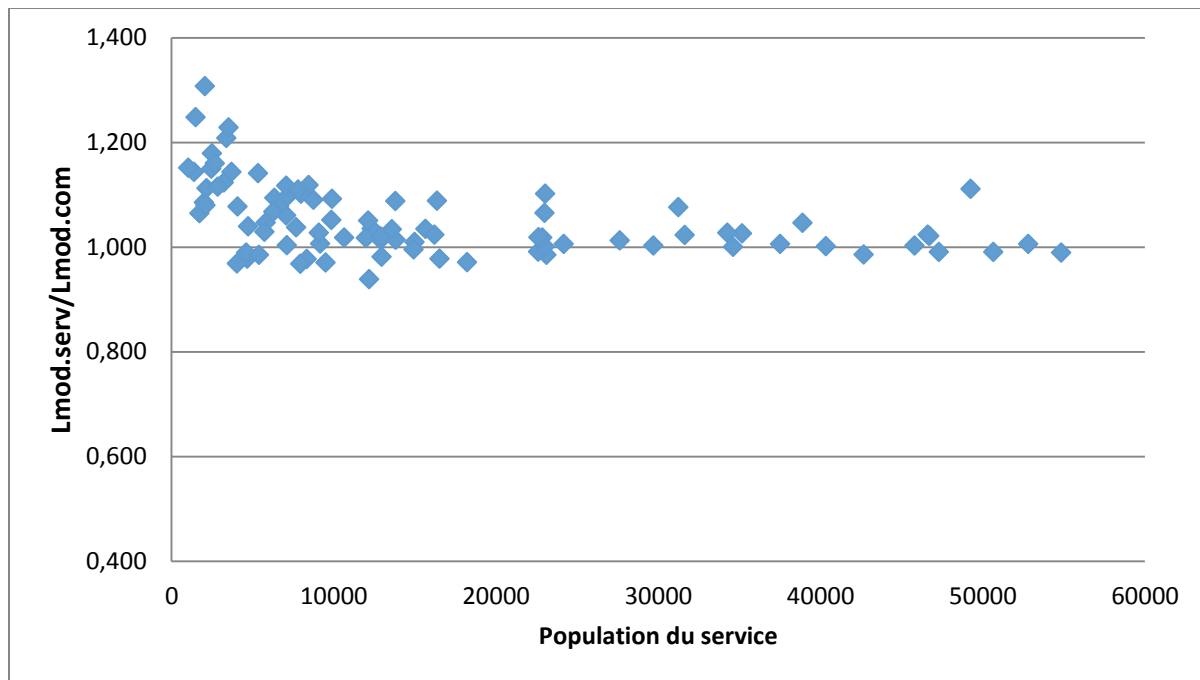


Figure 25: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la deuxième base

### Classes de population : Troisième base

Sur les **services renseignés mais non utilisés** (troisième base), nous avons donc des services qui :

- Contiennent au moins une commune appartenant à un autre service
- Si nous tentons d'agréger ce service avec d'autres services à la manière d'un méga-service, alors il existe au moins un service à agréger qui n'est pas renseigné dans le méga-service.

Ces caractéristiques font que sur ces services renseignés, il sera impossible d'obtenir un linéaire exactement comparable en faisant une simple somme sur les communes du service. Il est donc possible de raisonner par **prorata** ou par **limites**. Le prorata s'effectue en répartissant les routes de chaque commune de manière égale entre le nombre de services auxquels elles appartiennent. La limite inférieure sera créée à partir des communes mono-services de chaque service (donc les communes appartenant en totalité uniquement à ce service), et la limite supérieure sera créée en sommant les linéaires de toutes les communes appartenant au service ; cela créera donc une surestimation puisqu'il y a au moins un partie



du linéaire réel de la commune qui appartiendrait à un service différent. Les tailles de populations sur les différents services sont calculées de la même manière avant de pouvoir appliquer les pentes.

Les résultats obtenus alors sont donc à prendre avec précaution puisque ceux-ci appartiendront à une fourchette de valeurs. Les pourcentages et rapports  $L_{\text{mod.serv}}/L_{\text{mod.com}}$  calculés doivent être considérés comme ayant des marges d'erreurs.

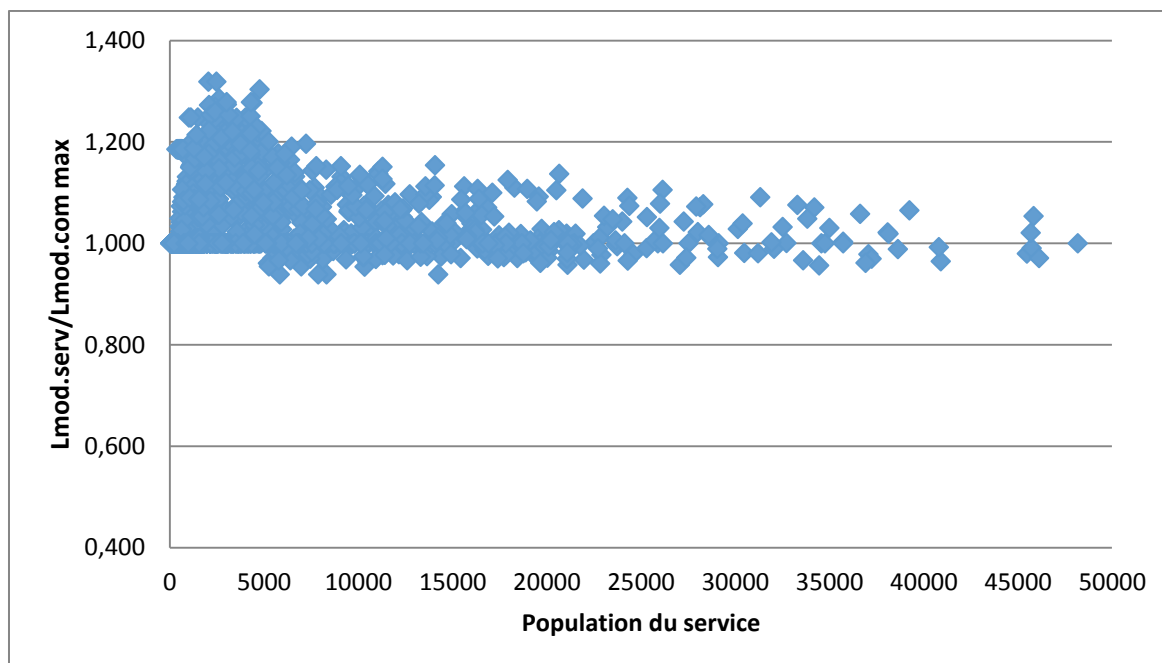


Figure 26: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de populations sur la troisième base (limite supérieure)

La Figure 26 précédente montre la tendance lorsque les linéaires sont comparés aux linéaires modélisés avec la limite supérieure. Les calculs avec les prorata de communes par service varient de manière similaire alors que ceux avec la limite inférieure présentent une marge beaucoup trop importante pour présenter une réelle représentativité des résultats du modèle. La tendance est similaire aux résultats sur la base de calage.

### Classes de densité

Il est donc difficile de dire que le modèle par classes de population soit optimal pour l'estimation du linéaire national. De plus la marge de différence entre les résultats lorsque les modèles sont appliqués sur les services et sur les communes sont inquiétants puisque l'estimation se fera en effet sur les bases de communes françaises.

Nous pouvons cependant observer qu'avec le modèle des densités, cette tendance se réduit fortement (Figure 27). En effet, utiliser la densité permet de normaliser la population, il est donc normal que la transition du modèle des services aux communes s'effectue en minimisant l'erreur. Ce modèle serait donc plus adapté à ce que nous recherchons.

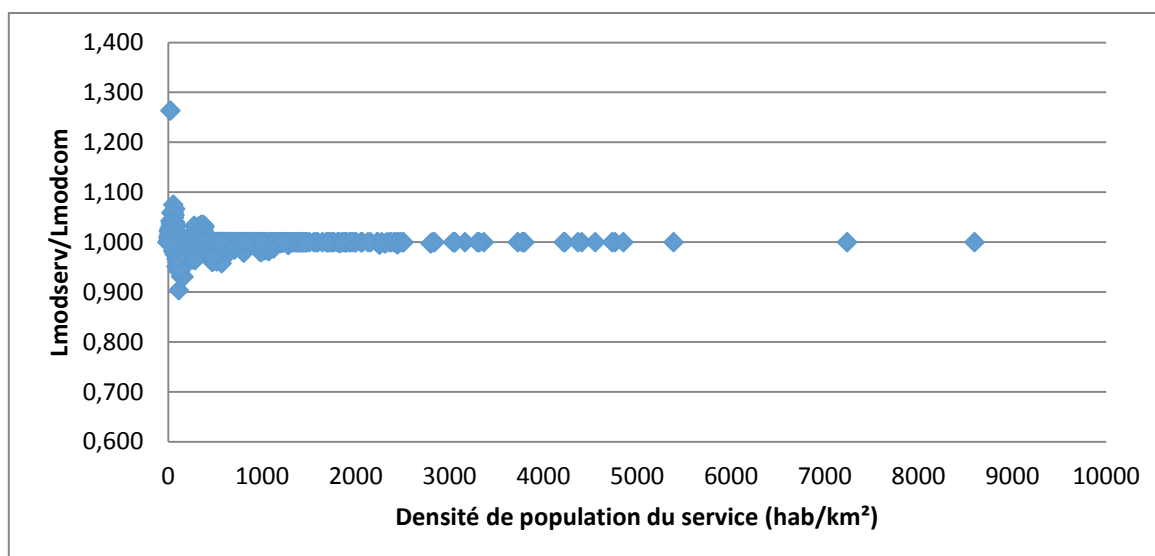


Figure 27: Comparaison des linéaires modélisés sur les services et sur les communes avec les classes de densités sur la première base

## BILAN

Le tableau suivant résume les linéaires modélisés obtenus lorsque les trois modèles linéaires simples sont appliqués sur la base de communes nationale :

Tableau 17: Résultats obtenus avec les modèles simples linéaires

	Laep réel	Modèle 1 (Tampon 250 m)		Modèle 2 (Classes population)		Modèle 3 (Classes Densité)	
		Lmod.serv	Lmod.com	Lmod.serv	Lmod.com	Lmod.serv	Lmod.com
Base 1	350920,8	100,5%		100,5%	97,4%	103,3%	103,2%
Base 2	55392,6	96,1%		96,1%	93,9%	98,3%	99,5%
Base 3	203696,3	140%		139,8%	129%	145,5%	140%
France entière			883985,2		852692,2		925559,2

Le **modèle 3** (par classes de densité) présente la meilleure capacité d'application sur les communes, cependant, il présente aussi une plus grande surestimation du linéaire sur la première base de calage.

Nous observons aussi que l'application des modèles sur la **base 3** induit une grande marge d'erreur de par le manque de précision dans sa composition. Cette base ne peut donc pas servir à mieux ajuster nos modèles et ne servira pas par la suite de l'étude.

#### 4. EVALUATION DU LINEAIRE DE RESEAU D'EAU POTABLE : MODELES MULTILINEAIRES

Les modèles précédents ont permis de prendre en compte **le réseau routier, les bâtis ainsi que les populations** des différentes communes pour arriver à un coefficient de corrélation ( $R^2$ ) supérieur à 0,90 et une somme totale du linéaire modélisée pratiquement égale au linéaire réel sur la base de calage.

Il sera difficile d'obtenir une précision plus importante, cependant, il est encore possible de prendre en compte plus de paramètres afin d'affiner l'estimation du linéaire de réseau à l'échelle du **service**.

Quelques nouveaux paramètres pris en compte sont notamment **la partie des routes traitées hors zones urbaines, la proportion des maisons par rapport aux logements présents sur la commune** ainsi que divers autres paramètres faisant intervenir **les surfaces de tampon** ont été créés durant le traitement sur SIG. Afin de prendre en compte une éventuelle influence régionale, l'altitude et l'agence de l'eau associée à chaque service seront aussi par exemple pris en compte.

**Deux principaux modèles multilinéaires** ont été recherchés en prenant en compte deux paramètres centraux différents : le premier restant les routes traitées par tampon de rayon de 250 m, le second étant des **classes de routes filtrées** construites en suivant des classes de **densité de bâtis**.

Ce dernier paramètre a été défini afin d'obtenir une autre manière d'observer l'effet de la répartition du bâti par rapport au réseau routier. La densité de bâtis dans chaque maille (de taille 500 x 500 m) a été calculée en **faisant le rapport de la surface de bâti au sol par rapport à la surface de la maille**.

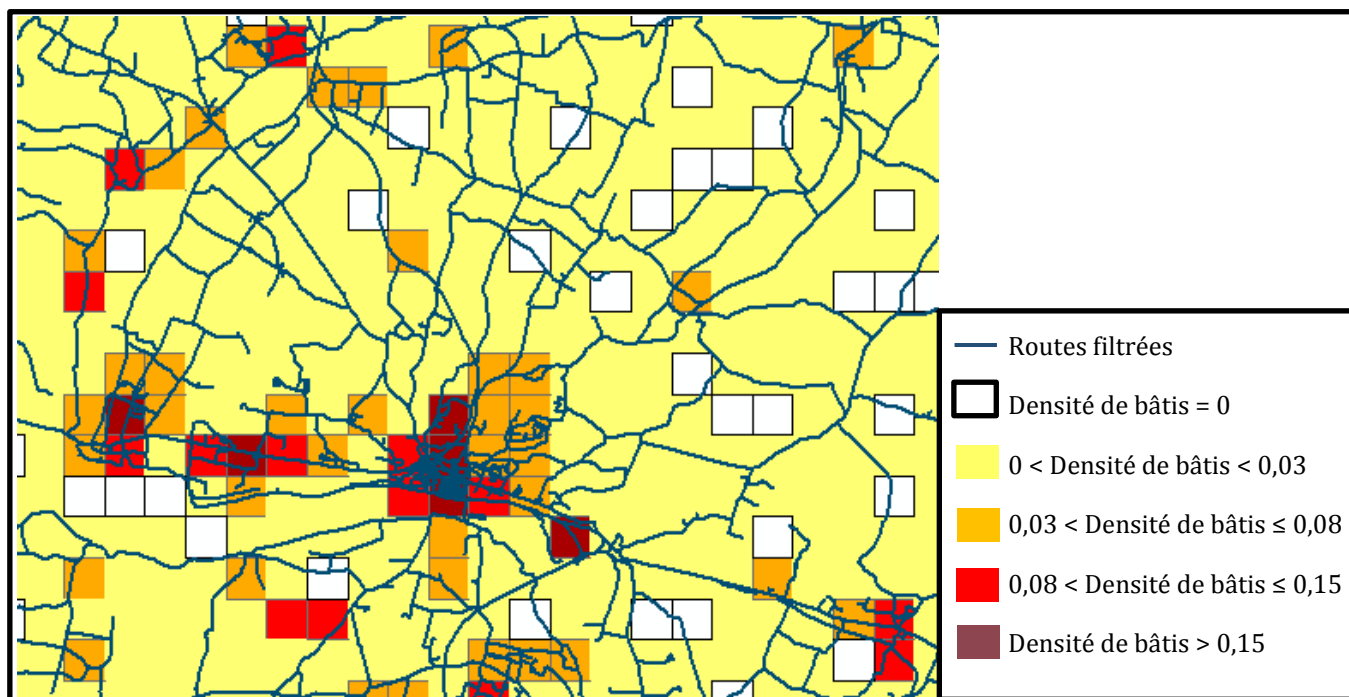


Figure 28: Exemple de distribution des densités de bâtis sur un réseau routier

Dans la Figure 28 ci-dessus, les différentes couleurs de maille représentent **5 différentes classes de densités de bâtis** comme définies dans la légende. La longueur totale des routes filtrées présentes dans chacune des classes est ensuite calculée par commune afin d'être exploitée dans les modèles.

#### 4.1 CONSTRUCTION DES MODELES

Cette partie décrit les étapes suivies pour tester la pertinence de tous les paramètres par rapport au linéaire réel.

##### 4.1.1 CORRELATION ENTRE PARAMETRES DEUX A DEUX

Nous avons d'abord observé les corrélations des paramètres deux à deux afin de déterminer si certains paramètres ne sont pas **redondants**. Nous pouvons observer, par exemple sur la Figure 29 que les routes sans traitement, les routes filtrées et les routes traitées sont toutes directement corrélées les unes avec les autres (**graphique linéaire et coefficient de corrélation forte**). Il faudra donc faire attention à ne retenir qu'un de ces paramètres pour la modélisation. De même entre le nombre d'abonnés par commune et la population de la commune.

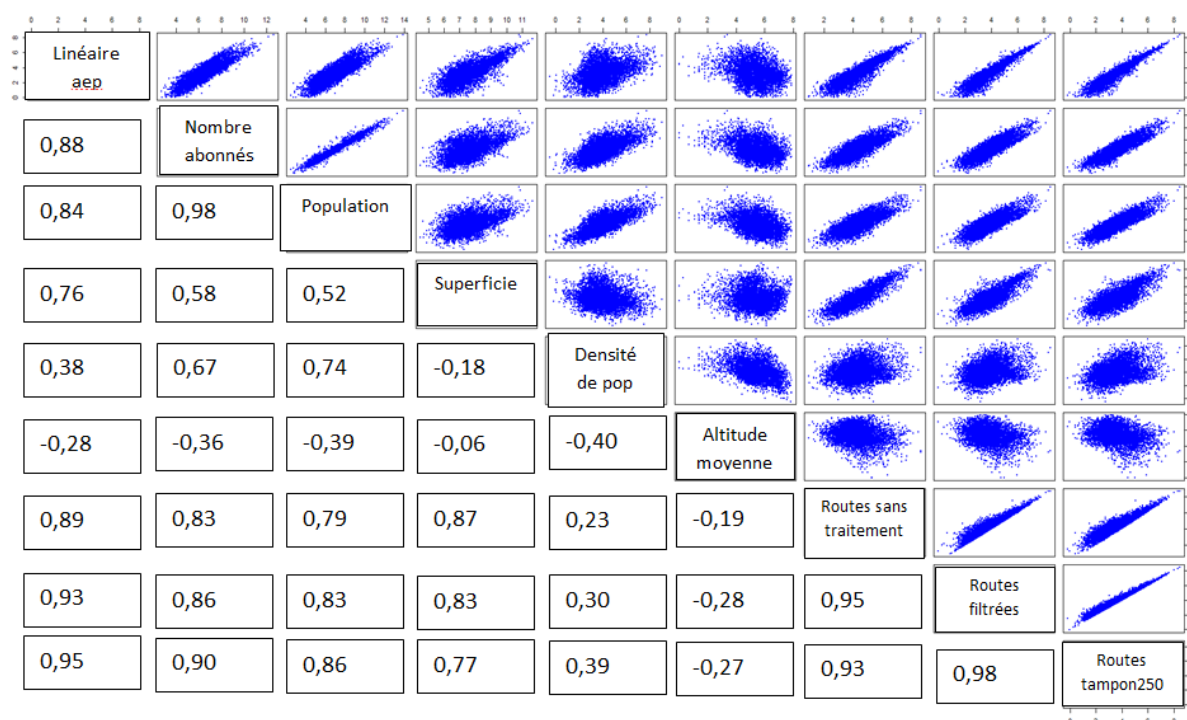


Figure 29: Corrélations deux à deux pour les huit premiers paramètres

##### 4.1.2 CALCUL DU VIF (VARIANCE INFLATION FACTOR)

Pour compléter l'étude des corrélations entre les paramètres, il est possible de calculer le VIF. Ce calcul, que permet le logiciel R, donne une idée de la **multi-colinéarité** des paramètres entre eux. Cependant, il ne permet pas de dire spécifiquement quels sont les paramètres directement corrélés. Il faut donc calculer ce facteur sur différentes combinaisons jusqu'à obtenir celles dont les coefficients finaux sont **inférieurs à 10**.

Les tableaux suivants donnent un exemple du processus suivi en utilisant le calcul du VIF pour la construction du modèle de routes traitées par tampon 250 m :

Tableau 18: Valeurs de VIF pour la recherche de nouveaux paramètres (1)

Variables	VIF
Populations	13.937072
Superficie du territoire	22.342733
Altitude moyenne	7.751593
Routes sans traitement	29.471144
Routes filtrées	217.157783
Routes urbaines	92.731849
Surface totale (urbaine + tampon)	47.849805
Routes traitées tampon 250	110.505880
Population/Surface zone tampon	4.039569
Maisons par bâti	3.007527

Nous pouvons observer ici au moins 5 paramètres dont le VIF dépasse 10 ; il y a donc une forte multi-colinéarité au sein de ce groupe. C'est un résultat logique puisque nous savons que des corrélations directes existent notamment entre tous les paramètres routiers.

Tableau 19: Valeurs de VIF pour la recherche de nouveaux paramètres (2)

Variables	VIF
Populations	10.124146
Superficie du territoire	8.509076
Altitude moyenne	1.269675
Routes urbaines	36.823705
Surface totale (urbaine + tampon)	35.276658
Routes traitées tampon 250	31.995463
Population/Surface zone tampon	3.820122
Maisons par bâti	1.836675

De la visualisation des corrélations directes, nous pouvons choisir de ne retenir qu'un seul paramètre « Routes » parmi celles non traitées, celles filtrées et les traitées. Les routes filtrées sont retenues par pertinence. Les valeurs de VIF diminuent considérablement mais sont toujours supérieures à 10 pour la moitié des paramètres.

Tableau 20: Valeurs de VIF pour la recherche de nouveaux paramètres (3)

Variables	VIF
Populations	1.448777
Superficie du territoire	5.066449
Altitude moyenne	1.274376
Routes traitées tampon 250	5.225446
Population/Surface zone tampon	3.890807
Maisons par bâti	1.783172

La surface totale pouvant être redondante avec la superficie du territoire, la première est enlevée du calcul du VIF. Après plusieurs essais en échangeant les variables une à une, il apparaît que les valeurs du VIF sont le plus intéressantes lorsque le paramètre « Routes urbaines » est aussi enlevé.

Les résultats suggèrent qu'avec **les routes traitées par tampon 250 m, l'altitude et le nombre de maisons rapportés au nombre de bâtiments** par commune peuvent constituer de nouvelles variables intéressantes pour notre modèle.

#### 4.1.3 CONSTRUCTION PAS A PAS DU MODELE : CRITERE D'INFORMATION D'AKAIKE

Dans cette méthode de construction d'un modèle multilinéaire, il s'agit de sélectionner d'abord la variable explicative la plus corrélée à la variable à expliquer, puis de sélectionner, parmi celles qui restent, la variable explicative dont la corrélation partielle est la plus élevée (en gardant constantes la ou les variables déjà retenues). A chaque fois, **la pertinence des variables** déjà sélectionnées est vérifiée : si l'une d'elles n'est plus significative, elle est retirée. Le critère de sélection retenu est le critère AIC (Akaike Information Criterion) qui permet de comparer deux modèles entre eux (plus il est faible, meilleur est le modèle). Il est calculé après chaque ajout ou suppression de variables, la solution qui le minimise est retenue avant de passer à l'étape suivante. Ceci est répété tant qu'il reste des variables candidates dont le coefficient de corrélation partiel est significatif. Chaque paramètre est cumulé un à un par ordre décroissant de pertinence et lorsque ce coefficient semble se stabiliser (varie de moins en moins), nous pouvons considérer que les paramètres pris en compte sont de moins en moins significatifs par rapport à notre modélisation voulue.

En y incorporant les paramètres retenus des étapes précédentes, il est donc possible de connaître les **paramètres les plus importants** pour bâtir notre modélisation.

Il faut cependant être prudent puisque **ce résultat n'est qu'indicatif** et ne fournit pas directement un modèle complet avec des coefficients associés à chaque paramètre. Il est donc intéressant de prendre en compte prioritairement les paramètres indiqués par le critère sans pour autant s'y enfermer lors de la construction finale du modèle.

Dans le cas du modèle par tampon 250 m, le critère considère par exemple que l'apport Maisons/Nombre de bâtiments n'est pas très significatif pour notre modèle, mais que **l'altitude** présente possiblement une pertinence.

#### 4.1.4 MODELES

##### Premier modèle multilinéaire K1

Il s'agit ensuite de construire plusieurs modèles différents en échangeant à chaque fois les variables afin d'observer leurs effets réels retenus pour la modélisation. Le choix des variables retenues est validé en fonction de la significativité de leur lien avec la variable à expliquer. Les termes d'interaction non significatifs sont éliminés. En l'occurrence, les variables dont la p-value calculée est supérieure à 5 % n'ont statistiquement pas de relation avec le linéaire des réseaux, et ne sont donc pas retenues dans le modèle final. L'apport des variables au modèle est évalué par l'évolution du coefficient de corrélation. Par exemple, il a été facilement observé qu'une interaction entre les routes retenues et l'agence de l'eau à laquelle appartient le service permet d'améliorer les caractéristiques de la modélisation (**meilleur R<sup>2</sup> et p-values très inférieures à 5%**).

Nous avons alors cumulé d'autres paramètres à la suite de cette interaction afin d'obtenir un modèle de type :

$$\text{Linéaire aep} = \beta_a * \text{Routes}_{\text{traitéestampon250}} + \beta_b * \text{Pop}_{\text{service}} + \beta_c * \text{Altitude}_{\text{moyenne}} + \beta_d * \frac{\text{Maisons}}{\text{Logements}}$$

Avec  $\beta_a$  un coefficient dépendant de l'agence à laquelle appartient le service

$\beta_b, \beta_c, \beta_d$  des coefficients indépendants associés à chaque paramètre

Cependant, les caractéristiques des modèles varient de façon **minimale** lorsque les paramètres sont ainsi cumulés, avec notamment un gain en  $R^2$  d'environ 0,4% pour l'altitude et le ratio Maisons/Logements. Il apparaît donc que leurs effets ne soient pas assez significatifs pour notre modélisation par rapport à la complexification qu'ils apportent. La **population du service** apporte cependant un peu plus de poids pour la modélisation (gain de  $R^2$  d'environ 1,3%) ; il en va de même pour la **superficie du territoire**. Le modèle qui devrait alors être retenu serait de la forme :

$$\text{Linéaire aep} = \beta_a * \text{Routes}_{\text{traitéestampon250}} + \beta_b * \text{Pop}_{\text{service}} + \beta_e * \text{Superficie du territoire}$$

Avec  $\beta_e$  un coefficient indépendant associé à la superficie du territoire

En se rappelant les **modèles simples linéaires** construits en partie 3.6, il a alors été jugé intéressant d'intégrer la **densité** à notre modèle au lieu de la population et de la superficie. En combinant les **classes de densité de population** créées dans la partie 3.6.2, nous obtenons le modèle linéaire simple suivant :

#### **Modèle K1 :**

$$\text{Linéaire aep} = \beta_f * \text{Routes}_{\text{traitéestampon250}}$$

Avec  $\beta_f$  un coefficient dépendant de l'agence de l'eau et de la classe de densité du service parmi les 3 classes (<120 hab/km<sup>2</sup>, 120-400 hab/km<sup>2</sup>, >400 hab/km<sup>2</sup>)

Nous obtenons alors un  **$R^2$  d'environ 0,977** avec des **p-values toutes très inférieures à 5%** pour chaque paramètre considéré (Annexe A5).

#### Deuxième modèle multilinéaire K2

Dans l'optique d'améliorer le lien avec la densité qui s'exprime uniquement au niveau de la commune, mais qui ne rend pas compte de la dispersion de l'habitat, **les classes de routes filtrées dépendant des densités de bâtis** définies en début de ce chapitre ont été créées et exploitées dans un deuxième modèle. La variable introduite rend compte de la densité de l'habitat à une maille de 500 m<sup>2</sup>.

En suivant les mêmes étapes avec comme paramètre central ces classes de routes, le meilleur modèle obtenu avec un  $R^2$  de **0,973** (Annexe A6) est le suivant :

**Modèle K2 :**

$$\text{Linéaire aep} = \beta_{a1} * \text{Routes}_{\text{classe1}} + \beta_{a2} * \text{Routes}_{\text{classe2}} + \beta_{a3} * \text{Routes}_{\text{classe3}} + \beta_{a4} * \text{Routes}_{\text{classe4}} + \beta_{a5} * \text{Routes}_{\text{classe5}}$$

Avec les  $\beta_{ai}$  des coefficients dépendant de l'agence de l'eau pour classe de densité de bâtis  $i$

Cumuler à cette équation le paramètre de population des services augmenterait la valeur du  $R^2$  à 0,98. Cependant le tracé des résidus de ce modèle par rapport à la population (Figure 30) montre que ce paramètre n'a pas d'effet direct sur les linéaires calculés. En effet les résidus étant centrés en zéro dans la Figure 30, nous pouvons a priori conclure que les résidus ne sont pas une fonction de la population.

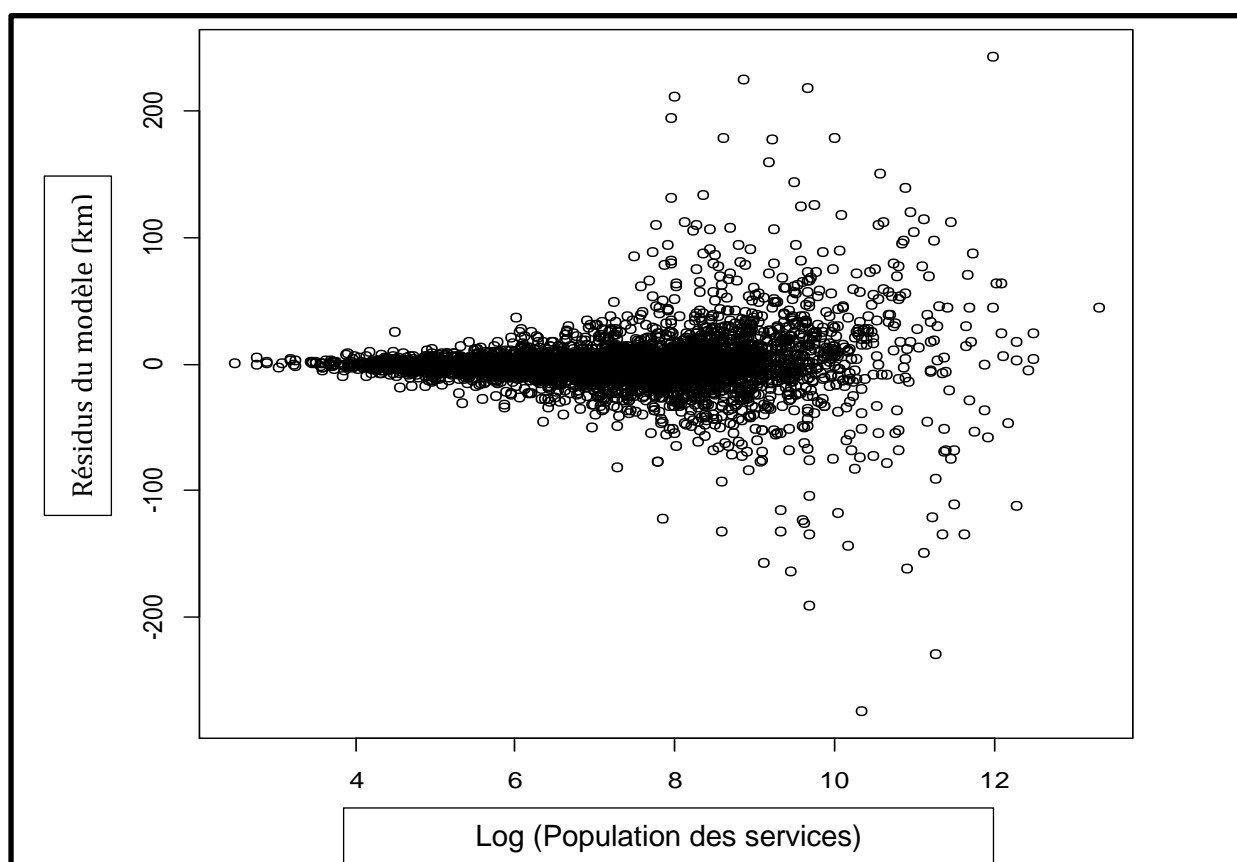


Figure 30: Analyse des résidus par rapport à la population des services

Le gain en  $R^2$  pourrait donc découler d'une correction sur quelques grands services seulement, mais en général ne rien apporter au modèle en lui-même.

Il s'agit ensuite de valider les modèles K1 et K2 retenus.



## 4.2 VALIDATION DES MODELES MULTILINEAIRES

Pour cela, il serait intéressant d'abord de vérifier que les résidus suivent une **distribution normale**. Un calcul simple sur le logiciel R (Annexe A7) nous permet de tracer les résidus des modèles calés de manière à observer leur normalité. Dans notre cas, les deux modèles présentent des graphiques similaires et mais peu compatibles avec une distribution normale. Cependant, même si les résidus **ne suivent pas tout à fait une loi normale**, les **p-values très faible de l'ordre de  $10^{-16}$**  confortent le choix du modèle et la significativité du lien entre les variables explicatives et le linéaire de réseau.

**Dix points extrêmes** (sur un total de 5043 services) de la base ont cependant été écartés car ils présentaient une trop grande déviation par rapport à une distribution normale. Les services concernés contiennent soit une erreur de donnée ou sont dotés d'une caractéristique qui ne permet pas de les modéliser correctement.

Après visualisation des valeurs prédites par le modèle (Annexe A8), nous vérifions aussi, comme pour les modèles linéaires simples, que les résultats obtenus en appliquant les modèles sur **les communes** de la base de calage sont similaires aux résultats obtenus sur les services.

Le tableau suivant résume les valeurs obtenues sur chaque modèle :

Tableau 21: Résultats des modèles multilinéaires

	Laep réel	Modèle K1		Modèle K2	
		Lmod.serv	Lmod.com	Lmod.serv	Lmod.com
Base 1	350920,8	100,60%	100,57%	100,47%	100,52%
Base 2	55392,6	97,1%	96,3%	96,1%	96,1%

Nous validons ensuite les modèles par la **méthode de Monte-Carlo**. Cette méthode permet, à partir d'un nombre important de **tirages aléatoires**, d'évaluer une valeur numérique (en l'occurrence la stabilité du linéaire modélisé). Elle consiste à **caler les modèles sur un échantillon** extrait de la base, et à appliquer le modèle résultant sur le reste de la base utilisée. Pour un même modèle et une même taille d'échantillon de calage, 2000 tirages aléatoires de la base sont effectués afin d'observer **le résultat de la prédiction** sur la totalité de la **base 1**. Si le modèle est pertinent, on doit en toute logique observer **la convergence des valeurs de linéaires** extrapolées avec celle observée.

Les mêmes étapes sont ensuite répétées en augmentant la taille de l'échantillon de calage (**25%, 50% puis 75%**) afin d'observer la variation des résultats produits. Les figures 31 et 32 suivants illustrent les résultats obtenus sur les deux modèles avec la méthode Monte-Carlo. Chaque boîte à moustache (explicités en Figure 33) représente graphiquement les résultats des 2 000 évaluations du linéaire de la base 1 entière réalisées pour un taux d'échantillonnage.

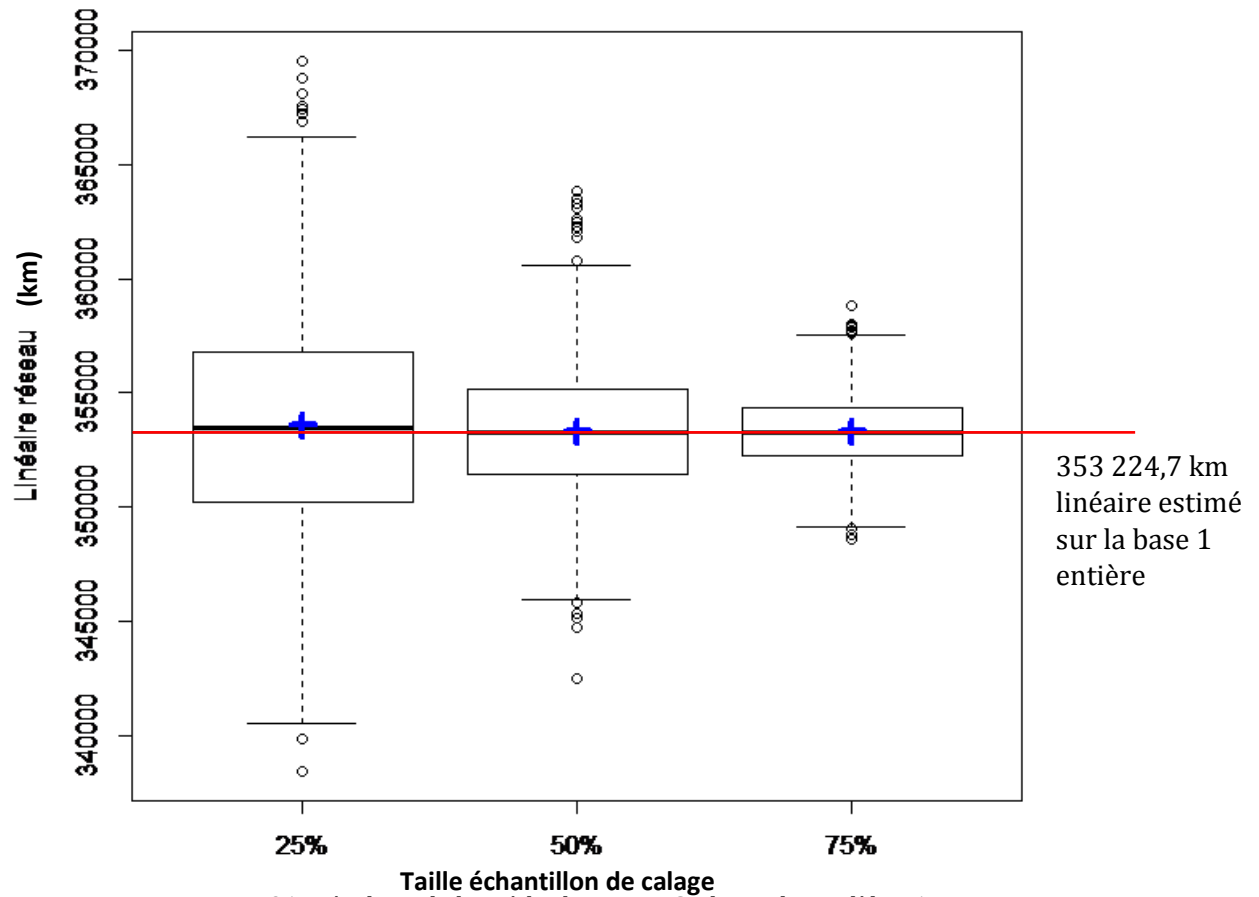


Figure 31: Résultats de la méthode Monte-Carlo sur le modèle K1

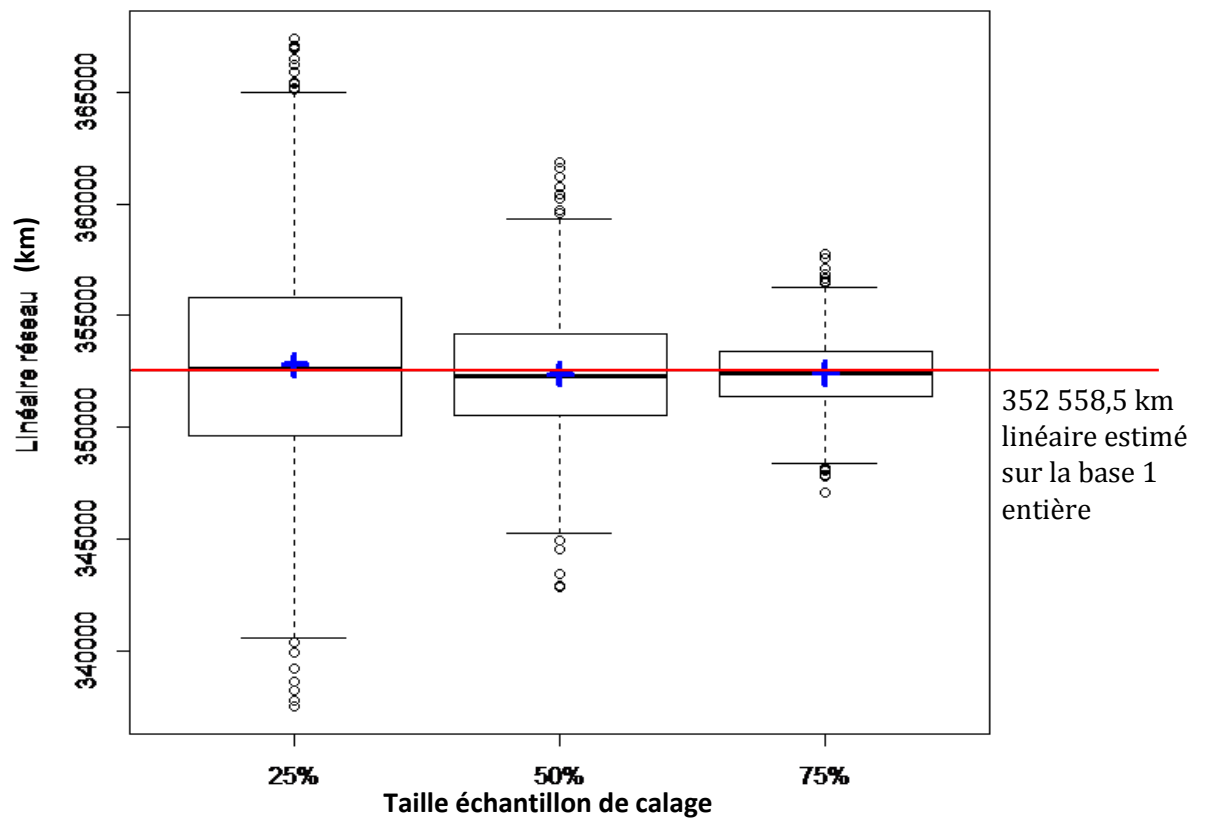


Figure 32: Résultats de la méthode Monte-Carlo sur le modèle K2

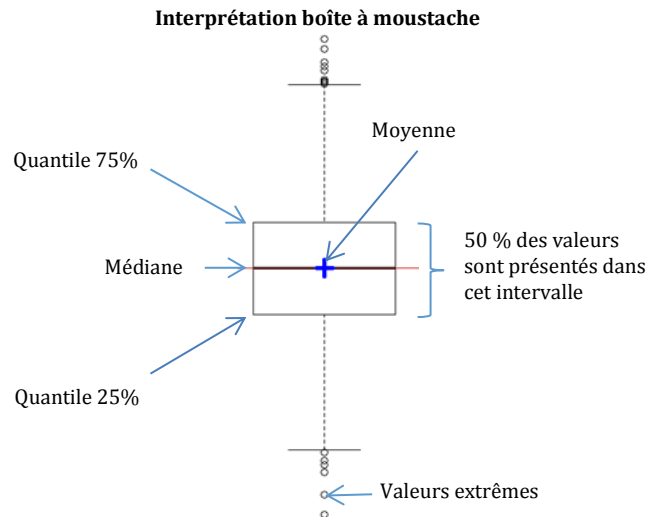


Figure 33: Schéma boîte à moustaches

Dans les deux cas, les prédictions des modèles calés sur des échantillons de taille croissante convergent vers la valeur estimée lorsque le calage est effectué sur la base entière. Le choix des variables des modèles est donc **robuste** : il ne dépend pas de manière significative de la taille de l'échantillon ou des individus spécifiques de la base.

## BILAN

Il y a très peu de différences entre nos **modèles K1 et K2** pour les linéaires prédits prédites, à la fois sur les services et les communes. Ils peuvent tous deux être considérés comme **robustes** suite à l'utilisation de la méthode Monte Carlo. Le **modèle K2** est cependant **plus facile à mettre en œuvre**, les étapes nécessaires à la création des variables densités de bâtis étant plus simples sur ArcGis. Ce modèle est donc retenu pour l'estimation nationale du réseau linéaire d'eau potable.

Tableau 22: Coefficients du modèle K2 retenu

Agence	Adour-Garonne	Artois-Picardie	Loire Bretagne	Rhin-Meuse	Rhône-Méditerranée-Corse	Seine Normandie
$\beta_{a1}$	-8.264e-03	-0,1168	0,1752	0,402	-5.878e-02	8.900e-02
$\beta_{a2}$	0,9483	0,5363	0,929	0,3035	0,7215	0,914
$\beta_{a3}$	0,1252	1,207	0,649	1,782	0,825	0,806
$\beta_{a4}$	1,242	0,859	0,659	0,467	0,719	0,6062
$\beta_{a5}$	0,7861	0,8104	0,781	0,853	0,53	0,881

L'intervalle de confiance du modèle est calculé grâce à l'écart-type des valeurs prédites selon l'équation suivante :

$$\text{Intervalle de confiance}(95\% \pm) = \text{Linéaire estimé} \pm 1.96 * \sqrt{\sum (\text{écarttype}_i)^2}$$

Nous obtenons alors **896 403,5 km** en estimation nationale de linéaire avec le modèle K2 retenu et un intervalle de confiance de plus ou moins **309,8 km**.



## CONCLUSION

La première partie de l'étude des déterminants et modélisation de la consistance du patrimoine d'eau potable en France commandée par l'ONEMA consiste à **modéliser de manière explicite les linéaires des réseaux à l'échelle nationale** pour compenser le manque de renseignements sur la majeure partie des services d'eaux. Le principe de la modélisation est de **relier le linéaire de réseau d'eau potable** à des caractéristiques pour lesquelles **l'information est connue** sur l'ensemble du **territoire national**. Le modèle calé sur les services pour lesquels les linéaires de canalisations sont connus pouvant ainsi être appliqué à l'ensemble des communes.

Après une étude bibliographique et des investigations exploratoires, la piste du lien entre **réseau routier** et **réseau de canalisations** a été exploitée au travers de **traitements SIG** pour dégager les routes desservant uniquement des bâtis et prendre en compte la densité de l'habitat à une échelle infra-communale. Un **prétraitement** par enlèvement des voies non-cohérentes a été préconisé avant un traitement du réseau par zone tampon pour ne retenir le linéaire routier que lorsque celui-ci est proche de zones construites.

Afin de caler et d'évaluer les modèles construits, des **bases de services** ont été créés à partir des données de la **base SISPEA**. Une première base composée de services simples et entièrement connus en matière de géographie et de démographie permet de caler les modèles. Une deuxième base permet de vérifier l'application des modèles sur des services plus complexes mais toujours complètement définis. Les **communes** appartenant à chaque service ont aussi été utilisées afin de pouvoir comparer l'effet des modèles appliqués parallèlement aux services et aux communes. En effet les données extraites de la base SISPEA sont disponibles à **l'échelle du service** et il a fallu les expliciter grâce à des variables à **l'échelle communale**.

**Trois premiers modèles** ont été créés à partir des réseaux routiers traités par zones tampons, les deux derniers prenant notamment en compte respectivement **la population et la densité de population** des services.

**Deux modèles multilinéaires** ont ensuite été créés sous le logiciel de calcul R afin de prendre en compte l'appartenance à une **agence de l'eau** ainsi que le réseau routier associé à certaines **classes de densité de bâtis**. Les résultats en détail des modèles sont disponibles en Annexe A9. Le Tableau 23 suivant résume les résultats en pourcentage ainsi que les estimations nationales pour chaque modèle.

Tableau 23 : Résumé des résultats trouvés sur les différents modèles

Modèles linéaires		Modèle 1 (tampon250)		Modèle 2 (Classe population)		Modèle 3 (Classe Densité)	
		Lmod.serv	Lmod.com	Lmod.serv	Lmod.com	Lmod.serv	Lmod.com
	Laep réel						
Base 1	350920,8	100,5%		100,5%	97,4%	103,3%	103,2%
Base 2	55392,6	96,1%		96,1%	93,9%	98,3%	99,5%
France entière			883985,2		852692,23		925559,19
Modèles multilinéaires		Modèle K1		Modèle K2			
	Laep réel	Lmod.serv	Lmod.com	Lmod.serv	Lmod.com		
Base 1	350920,8	100,60%	100,57%	100,47%	100,52%		
Base 2	55392,6	97,10%	96,3%	96,10%	96,10%		
France entière			894212,68		<b>896403,53</b>		

Il sera difficile d'aller plus loin quant à l'estimation nationale du linéaire de par la **grande proximité** déjà obtenue entre les résultats modélisés sur la base de calage et les linéaires réels accessibles. L'obtention des modèles vus dans ce rapport demandent d'ailleurs **un temps considérable passé au tri et à la correction de données**. A l'échelle **du service**, il est certes encore possible d'obtenir une meilleure précision, notamment sur les petits services, cependant cela requerra un **plus grand nombre de données** dont la précision n'est pas assurée.

Les résultats obtenus des modélisations sur le linéaire étant satisfaisants, **la suite** de l'étude commandée par l'ONEMA consistera à modéliser de **nouvelles caractéristiques du patrimoine de réseaux d'eau potable** en France. Ces caractéristiques incluent notamment **les diamètres, les matériaux et les dates de pose des canalisations**, afin de permettre une meilleure programmation du renouvellement des réseaux. La tâche s'avère cependant **moins aisée** que sur les linéaires puisqu'elle nécessite de connaître en détail le patrimoine de plusieurs réseaux pour caler les modèles et nécessite donc la **coopération de nombreux exploitants**.

---

## BIBLIOGRAPHIE

---

1. Office National de l'eau et des milieux aquatiques (ONEMA), Association scientifique et technique pour l'eau et l'environnement (ASTEE), Association des Ingénieurs territoriaux de France (AITF). **Gestion patrimoniale des réseaux d'eau potable : Elaboration du descriptif détaillé des ouvrages de transport et de distribution d'eau (Décret 2012-97 du 27 janvier 2012)**. ONEMA éditions, 2013. 48p. (Guides et protocoles)
2. Observatoire national des services d'eau et assainissement. **Panorama des services et de leur performance en 2012**. ONEMA éditions, 2015. 88 p.
3. REBEIX Gisèle. **800 000 km de conduites pour distribuer l'eau potable**. *Les données de l'environnement*, 2001, n°71. p. 1-4.
4. Canaliseurs de France. **Communiqué de presse pour l'Assemblée Générale de Canaliseurs de France du 16 Juin 2015** [en ligne], [Consulté le 25/02/2016] <<http://www.campagne-presse.com/wp-content/uploads/2015/10/CP-AG-15.pdf>>
5. LE JEANNIC Françoise, GICQUIAUX Cyril, GREGOIRE Patrice. **Services d'eau et d'assainissement : une inflexion des tendances ?** *Le Point Sur (SOeS-SSP, Enquête Eau 2008)*, 2010, n°67. p.1-4.
6. Service de l'Observation et des Statistiques. **Les services publics de l'eau en 2004 – Volet eau potable (Octobre 2007)** [en ligne], [Consulté le 25/02/2016] <<http://www.statistiques.developpement-durable.gouv.fr/publications/p/145/1168/services-publics-leau-2004-volet-eau-potable.html>>
7. CADOR Jean-Michel. **Le patrimoine en canalisation d'AEP en France, Bilan des huit enquêtes départementales et estimation nationale**. Rapport au Ministère de l'Aménagement du Territoire et de l'Environnement et à l'Assemblée des Départements de France, 2002. 193 p.





# ANNEXES

---

---

## A1 : Premiers traitements sous ArcGis sur la Réole

### I. Inventaire des couches SIG

Les différentes couches utilisées proviennent des sources suivantes :

- Pour l'occupation du sol :
  - o **Corine Land Cover** afin de repérer les types d'activités sur le secteur étudié
  - o La couche de bâtis de l'**IGN** afin d'évaluer les routes pertinentes à corrélérer avec le réseau d'eau potable
- Le réseau routier sur le territoire obtenu du site **Geofabrik**
- Le **réseau d'AEP** sur la collectivité et le découpage du secteur du service de l'eau de la Réole

Les couches récupérées ont été transformées en projection **Lambert 93** (si elles ne l'étaient pas déjà) afin que les calculs soient applicables et plus précis.

Une première organisation des catégories d'occupation du sol sur Corine Land Cover a été effectuée, les catégories d'origine étant trop précises pour être pratique à ce stade.

Les catégories obtenues sont les suivantes :

- Les mines et les décharges
- Les plages et sable
- Les surfaces en eau
- Les zones humides
- Les zones commerciales, aéroports, ports et équipements de loisir
- Le tissu urbain (continu et discontinu)
- Les espaces verts artificialisés
- Les espaces agricoles (dont vignobles et cultures parcellaires)
- Les forêts et milieux verts semi-naturels
- Le réseau routier, ferroviaire et espaces associés

Sur notre secteur particulier, seules 5 de ces catégories sont présentes, une nouvelle division a donc plus tard été réalisée afin de mieux visualiser l'occupation du sol sur la Réole :

- Tissu urbain continu et discontinu
- Zones industrielles, commerciales et équipements de loisirs
- Terres arables
- Vignobles
- Prairies
- Systèmes cultureux et parcellaires
- Surface essentiellement agricoles
- Forêts de feuillus
- Forêts et végétation arbustive
- Cours et voies d'eau

Cette nouvelle catégorisation sera utilisée plus tard lors de la correction du réseau routier sur le secteur.

## II. Premiers calculs de longueur

Un premier état de la longueur des réseaux sur la Réole est effectué (Par mesure de la longueur des polygones sur Arcgis) :

Longueur du réseau d'**AEP** = 143,723 km

Longueur du **réseau routier** = 138, 576 km (~=96%)

Cette forte proximité entre les deux valeurs pourrait laisser à penser que le réseau routier suffisant pour l'évaluation de longueur. Cependant, la cohérence réelle entre le parcours des deux réseaux doit être calculée avant de pouvoir confirmer cette hypothèse.

Ce calcul a été effectué de 3 manières différentes :

### 1. En utilisant une transformation en raster

Les polygones des deux réseaux sont converties en pixels rasters de taille 37 x 37 m en utilisant les outils **Polygone vers Raster** puis **Calculatrice Raster** de l'ArcToolbox, ces pixels sont affectés de la valeur 1 sur leurs réseaux respectifs et de 0 partout ailleurs.

Les pixels sont ensuite recoupés par un calcul simple de soustraction (**Réseau AEP - 2\*Réseau routier**) afin de déterminer les endroits où les réseaux se superposent ou non. A chaque pixel du secteur nous obtenons donc une valeur qui détermine sa position sur le réseau :

- Les routes où il n'y a pas de canalisations présentes = -2
- Les routes et les canalisations coïncident = -1
- Les canas ne passent pas en dessous des routes répertoriées = 1
- 0 partout ailleurs

Le nombre de pixels dans chaque catégorie et une proportion simple permet alors de retrouver leurs longueurs associées.

**Inconvénient de cette méthode :** Le réseau d'AEP passant souvent dans une certaine marge par rapport à la route, la pixellisation crée des régions où les réseaux qui coïncident ne sont pas pris en compte. (Figure 34)

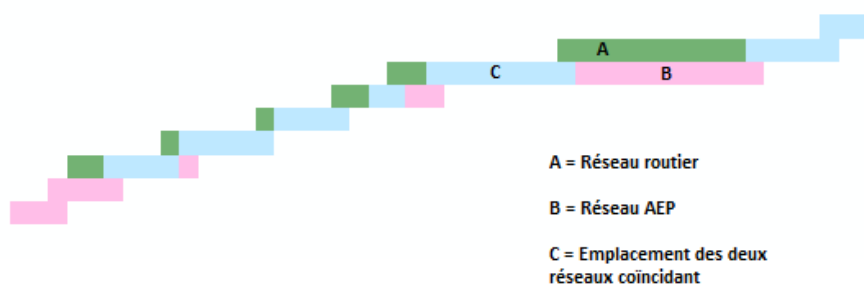


Figure 34: Exemple de canalisation résultant du calcul par raster

Sur la [Figure 34](#) précédente, en A et B, les deux réseaux devraient clairement être considérés comme coïncidents, mais les pixels ne se superposant pas exactement, ces derniers produiront une erreur dans le nombre réel de longueur des catégories en résultat.

## 2. Par ligne d'intersection

La fonction **Intersecter** d'Arcgis permet de créer une nouvelle polyligne suivant l'axe de l'intersection entre les deux réseaux considérés, et cela à une certaine distance près. Pour être certaine de prendre en compte les réseaux coïncidant sans pour autant créer une polyligne d'intersection trop grossière, j'ai considéré une distance de 25m. Le calcul de géométrie permet ensuite de calculer la longueur de l'intersection créée.

**Inconvénient de cette méthode** : La marge incorporée crée une polyligne qui donne une indication de la longueur de l'intersection, mais n'est pas plus précise de par la création de cette ligne entre les deux réseaux.



Figure 35: Exemple de lignes obtenues par intersection

## 3. Par découpage à partir d'une zone tampon

Afin d'être plus précis en terme de longueur d'intersection, la dernière méthode consiste à créer une **zone tampon** (de 15 m) autour du réseau d'eau potable sur la collectivité, et de **découper** le réseau routier par rapport à cette zone tampon. La polyligne de route découpée est alors mesurée par calcul de géométrie sur Arcgis. Cette méthode permet de prendre en compte le décalage spatiale entre les deux réseaux ainsi que d'obtenir une longueur précise de l'intersection entre les deux réseaux. Cette valeur de longueur permet alors aussi de calculer par soustraction simple les longueurs de chaque réseau non cohérent avec l'autre.

Longueur routes hors réseau AEP = Longueur du réseau routier – longueur d'intersection

Longueur AEP hors réseau routier = Longueur du réseau AEP – longueur d'intersection

### III. Calculs avec un réseau routier modifié

Il est apparent des résultats précédents qu'il sera difficile de lier le linéaire de routes au linéaire de réseau AEP du aux divergences importantes entre les longueurs de réseaux non cohérents. Dans cette partie, une méthode a été proposée afin de modifier le réseau routier pour ne prendre en compte que celles qui desservent des zones réelles d'activités.

Pour cela, une couche de tronçons de route a été créée en combinant les parties de secteurs dont l'occupation du sol était **d'origine agricole ou naturelle**, et en supprimant celles qui sont en contact avec des **bâties à moins de 50m** (grâce à des zones tampons). Afin de pouvoir effectuer ce traitement, le réseau routier doit d'abord être **fractionné en tronçons** de longueur inférieurs à leur état initial afin que seules des parties spécifiques du réseau soient affectés par les caractéristiques de la modification. Afin de limiter par ailleurs l'enlèvement de certains tronçons ne desservant a priori pas d'habitations proches mais étant nécessaires tout de même à la connexion logique du réseau d'eau potable, le fractionnement a été implémenté à une distance d'environ 277 m (500 parties égales sur toute la Réole).

La couche de tronçons obtenue est ensuite retranchée au réseau routier initial sur la collectivité et les calculs précédents sont de nouveau effectués sur tous les secteurs. Sur toute la collectivité, 23 tronçons sont alors enlevés.

Les tableaux suivants résument les résultats obtenus avec chaque méthode de mesure et sur chaque secteur pour le réseau routier modifié.

Tableau 24: Résultats des premières comparaisons sur la Réole

Secteur en Réole entier	( Avec correction de réseau routier )			
Longueur du réseau AEP	<b>143,723</b>	km		
Longueur du réseau routier		118,556	km	
% route/AEP		82,49%		
		Méthode raster	Méthode intersection	Méthode tampons
Longueur de réseau AEP hors réseau routier		53,5		54,085
	%	37,22%		37,63%
Longueur de réseau routier hors réseau AEP		25,7		28,918
	%	21,68%		24,39%
Longueur d'intersection		90,223	102,757	89,638
	%	62,78%	71,50%	62,37%

Sur le secteur entier, cette modification est à double tranchant : elle diminue le taux de réseau routier hors intersection mais diminue aussi le taux d'intersection moyen à environ 65%.

Le réseau routier ne suffit pas, mais il peut avoir une participation importante dans le calcul du linéaire de réseau d'eau potable. L'intersection est particulièrement importante sur les secteurs urbains, et une correction du réseau routier peut y améliorer la cohérence, mais cette correction n'est pas à généraliser trop hâtivement sur les secteurs plus ruraux. En termes de méthode, le calcul effectué avec la zone tampon autour du réseau AEP (II.3) semble être plus précis pour le calcul des indicateurs d'écart ; cette méthode sera donc retenue si ces calculs doivent être effectués sur d'autres collectivités dans le futur.

## A2 : Traitement sous ArcGis généralisé

Liste des couches utiles sous ArcGis :

Nom de la couche	Source
Routes par région (à convertir en lambert 93)	Geofabrik (openstreetmaps)
Bâti par département	IGN (BD topo)
Zones construites sur le département	Données EEA
Communes et ses limites	Géofla (IGN)

- 1- Conversion en Lambert-93 de la couche « routes »  
Couche Résultante : « *routes93\_[secteur étudié]* »
- 2- Sélection par soustraction des routes négligeables en termes de réseau d'eau potable, notamment dans les catégories :

bridleway	cycleway
footway	track
path	elevator
steps	via ferrata

- 3- Fusionner les routes de la région ensemble  
Couche résultante : un seul grand tronçon de route « *routes93\_all\_[secteur étudié]* »
- 4- Regrouper le bâti en deux couches :
  - « *bati\_ind\_[secteur étudié]* » regroupe les couches : bâti industriel et bâti indifférencié



Figure 36: Regroupement des couches bati industriel et indifférencié

- « *bati\_nonind\_[secteur étudié]* » regroupe les couches : toutes les autres couches de bâti de l'IGN hors les constructions linéaires et ponctuelles, qui ont été jugés négligeables de par leurs caractéristiques par rapport au réseau d'eau potable. (Définitions - Constructions linéaires : barrages, ponts, murs remarquables etc,

- Constructions ponctuelles : bâti d'emprise moins de 20m<sup>2</sup> et hauteur plus de 50m comme des éoliennes, antennes, cheminées, phares etc)
- 5- Création des couches de bâtis considérant uniquement les zones non urbanisés en croisant avec la couche de zone construites de EEA. Couches résultantes :
  - « *bati\_ind[secteur étudié] campagne* » et « *bati\_nonind[secteur étudié]\_campagne* »
- 6- Création d'une zone tampon (taille selon la distance testée, par exemple 250 m) autour de chacune des couches de bâti (hors zones urbaines), Dans les couches résultantes, un « rond » est présent autour de chaque bâti (beaucoup de ronds se chevauchent)
  - « *tampon250\_bati\_nonind[secteur étudié]\_campagne* »
  - « *tampon250\_bati\_ind[secteur étudié]\_campagne* »

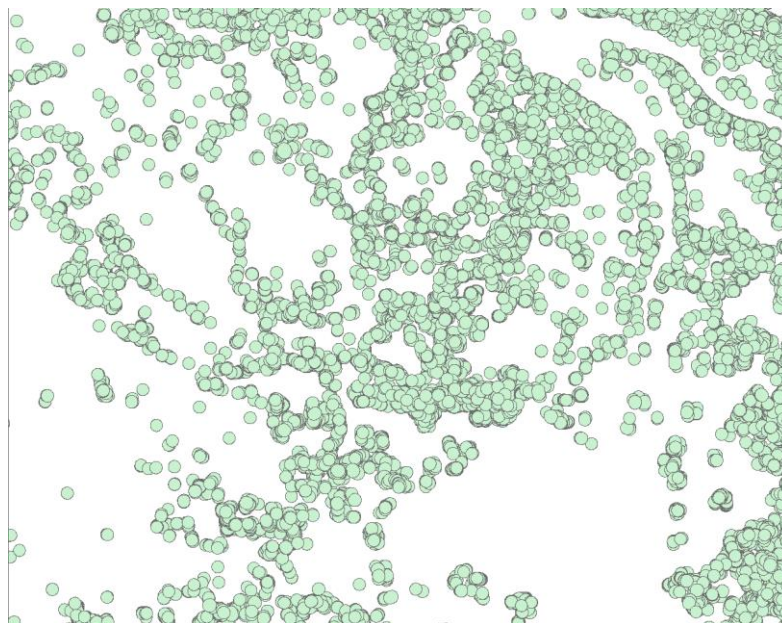


Figure 37: Couche de zone tampon pour la couche bâti indus/indiff

- 7- Fusion de chacune des couches avec les zones tampon autour du bâti (attention lors de la fusion il faut décocher « créer des entités multiparties »)
  - Les « ronds » formés en 4 fusionnent s'ils se touchent
  - « *tampon250\_bati\_nonind[secteur étudié]\_campagne\_fusion* »
  - « *tampon250\_bati\_ind[secteur étudié]\_campagne\_fusion* »

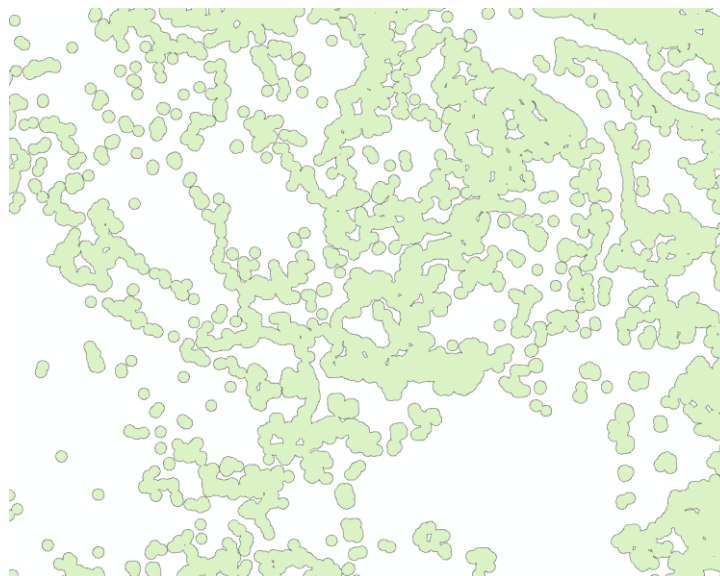


Figure 38: Couche de zones tampon fusionnées



Figure 39: couche des zones tampons fusionnées, vue avec réseau routier

- 8- Découper la couche « *routes93\_all\_[secteur étudié]* » avec chacune des couches « *tampon250\_bati\_nonind[secteur étudié]\_campagne\_fusion* », « *tampon250\_bati\_ind [secteur étudié]\_campagne\_fusion* » et zones urbaines « *UMZ006* »

Cette découpe permet de conserver uniquement les routes comprises à moins de 250m du bâti (soit 500m entre deux bâtis). Couches résultantes : « *routes\_non ind\_[secteur étudié]* » et « *routes\_ind\_[secteur étudié]* » et « *route\_zoneconstruite\_[secteur étudié]* »



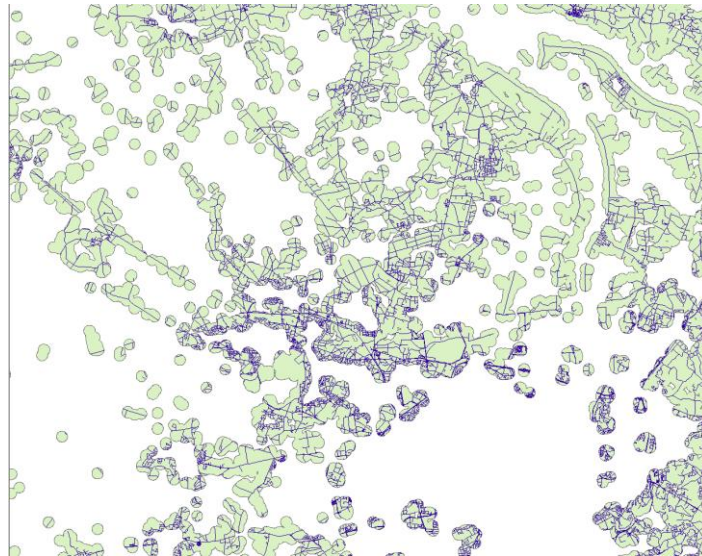


Figure 40: Réseau routier après la découpe par zones tampons fusionnées

- 9- Combiner les routes découpées dans les couches précédemment pour obtenir la couche « route\_traitee\_[secteur étudié] »
- 10- Fusionner les lignes de cette couche en une seule polygone « route\_traitee\_fusion\_[secteur étudié] »
- 11- Intersecter la couche « route\_traitee\_fusion\_[secteur étudié] » avec la couche « commune » (Données Geofla). On obtient pour chaque commune le linéaire de routes retenues « route\_traitee\_parcommune\_[secteur étudié] ».
- 12- Les résultats sont ensuite calculés à plusieurs échelles :
 

Le réseau routier sur chaque commune sans aucun traitement est obtenu à partir de l'intersection entre la couche « Commune » et « Route [secteur étudié]\_ensemble »

Le réseau routier sur chaque commune avant le recoupement par les zones tampons est obtenu à partir de l'intersection entre la couche « Commune » et « routes93\_all\_[secteur étudié] »

Le réseau routier après traitement complet est obtenu à partir de la couche « route\_traitee\_parcommune\_[secteur étudié] ».
- 13- Comparer le réseau routier sur chaque commune au réseau d'eau potable

### A3 : Construction de la base 1 (Services simples et complets)

Données utilisées : Base SISPEA pour la composition communale et les données des services d'eau sur toutes la France de 2009 à 2013.

Objectif : Obtenir une base de données sur les communes en France comportant des informations pouvant être corrélés au linéaire d'eau potable. Pour que cette base puisse être exploitée, nous voudrions que les communes étudiées ne soient pas sur plusieurs services à la fois (pour éviter les erreurs suite à leur double comptage). Les services répertoriées en 2013 dans ce cas ne sont qu'environ 3000 (échantillon non suffisant), il faudra donc prendre en considération les données des années précédentes, en faisant attention à ne pas avoir de doublons et en ne prenant pas en compte les services qui contiennent des communes appartenant à d'autres services dans les années futures.

Démarche :

- 1- Combiner les deux bases en liant leurs N° INSEE Commune sur Access (par requête Sélection). Garder dans la sélection les champs qui peuvent nous intéresser :
  - a. N° du département de la commune
  - b. N°INSEE de la commune
  - c. Le nom de la commune
  - d. Population de la commune
  - e. Distribution
  - f. Production
  - g. Année
  - h. Agence de l'eau
  - i. ID SISPEA du service
  - j. Communes adhérentes au service (nombre)
  - k. Population des communes adhérentes
  - l. Nombre d'abonnés (VP 056)
  - m. Linéaire de réseau AEP hors branchements (VP 077)

Cette liste peut être modifiée si besoin en faisant des nouvelles requêtes de Sélection après la mise en place de la base de données.

Le tableau correspondant à la requête est ensuite exporté en format Excel :  
« **Services\_Année\_filtre** »

- 2- Les tableaux précédemment créés sont importer dans la base Access sous le nom de « **Services\_Année\_filtre** »
- 3- Sur chaque tableau **Services\_Année\_filtre**, effectuer une requête Sélection permettant par regroupement de N°INSEE Commune de compter le nombre d'ID SISPEA. Le tableau donne alors le nombre de services par communes pour l'année considérée :  
« **services\_par\_communes\_Année** »
- 4- Une nouvelle requête Sélection permet ensuite en liant le N°INSEE commune de « **Services\_Année\_filtre** » et « **services\_par\_communes\_Année** » (jointure affichant

tous les enregistrements de « Services\_Année\_filtre »), d'obtenir les enregistrements des communes n'appartenant qu'à un seul service. (CompteIDsispeaservice =1) :  
« Services\_Année\_sansservicesmultiples »

- 5- De la table « **Services\_Année\_sansservicesmultiples** », une nouvelle Sélection permet avec le critère « Est Pas Null » sur les champs chiffrés de trier les services retenues qui ont été renseignés pour l'année considérée : « **Communes\_renseignes\_Année** »
- 6- La table « **Communes\_renseignes\_2013** » est exportée puis réimportée en table afin de permettre un cumul des données sous le nom « **Services\_2009-2013\_traités** »
- 7- Afin de ne retenir que les ID SISPEA de services non-récurrents dans les années futures, une sélection est effectuée en liant les tables « **Services\_2009-2013\_traités** » et « **Communes\_renseignes\_2012** » (jointure affichant tous les enregistrements de « Communes\_renseignes\_2012 ») par leur ID SISPEA service et en spécifiant un ID SISPEA service nul pour la table « Services\_2009-2013\_traités » : « **R\_ID2012** »
- 8- Nous devons aussi éliminer toutes les communes qui se trouvent dans un service différent dans les années futures, et les services associées correspondant en l'année considérée (2012). Cela se fait en deux étapes :
  - a. Une sélection est effectuée en liant les N°INSEE commune de « **Services\_2009-2013\_traités** » et « **R\_ID2012** » (jointure affichant tous les enregistrements de « R\_ID2012 »), et en utilisant le critère N°INSEE commune de la table « Services\_2009-2013\_traités » = Est Pas Null. Ainsi, nous obtenons les communes qui sont dans des différents services en 2012 et les informations associées à ces communes (dont l'ID SISPEA service) : « **IDserviceàsupprimer\_2012** »
  - b. Nous devons ensuite ne retenir que les communes de « **R\_ID2012** » qui n'ont pas les ID SISPEA service de « **IDserviceàsupprimer\_2012** ». Cela se fait en liant ces deux tables par l'ID SISPEA service (jointure affichant tous les enregistrements de « R\_ID2012 »), et en utilisant le critère ID SISPEA service de la table « IDserviceàsupprimer\_2012 » = Est Null : « **Communes\_àajouter\_2012** »
- 9- Les informations des communes de la requête « **Communes\_àajouter\_2012** » sont alors ajoutés à la table « **Services\_2009-2013\_traités** » par requête ajout 'Ajout\_2012-2013'.
- 10- Les étapes 7 à 9 sont répétées pour les années 2011, puis 2010 et enfin 2009 afin d'obtenir une base de données dont les informations sont les plus récentes possibles.
- 11- Le tableau final « **Services\_2009-2013\_traités** » est alors exporté en format Excel afin de pouvoir l'exploiter avec les données SIG.

## A4 : Constructions de la base 2 (Méga-services)

Etapas sous Access (Microsoft)

- 1- *R00\_vider T Tempo* : Table temporaire des services à supprimer initialement vide
- 2- *R01\_met 1 dans service retenu* : Requête Mise à jour qui permet de mettre à jour une colonne « Retenu » avec valeur 1 à chaque ligne de la table de données brutes des services (*Services1\_2013*)
- 3- *R02a\_select serv non renseigne* et *R02a\_select serv renseigne* : Requête Sélection pour obtenir en deux tables tous les services renseignés et non renseignés respectivement
- 4- *R02\_MAJ service non renseigne* : Requête Mise à jour qui met le champ retenu à 0 quand service non renseignés (Vp056, Vp077). Cette requête s'appuie sur la requête *R02a\_select serv non renseigne*
- 5- *R03\_service non retenu* : Alimente la table *T\_tempo* avec ID service sélectionné comme suit : recherche de toutes les communes appartenant à un service non renseignés, puis recherche des services associés à ces communes pour les supprimer (requêtes associées, *R03a\_selection service non retenu*, *R03b\_select commune\_2*, *R03c\_select serv renseigne*, *R03d\_services non renseignes*)
- 6- *R04\_MAJ service à partir T Tempo* : Le champ retenu de tous les services de la table *T Tempo* est mis à 0 (Requête Mise à jour)
- 7- *R05\_Ajout service a T Tempo* et *R06\_MAJ service à partir T Tempo* : Les procédures 5 et 6 sont répétées autant de fois que le nombre de services ajouté à la table *T Tempo* n'est pas nul en relançant successivement ces deux requêtes.
- 8- *R07\_ajout service incohérence nombre comm* : Vérification de la cohérence commune desservie + commune adhérente avec somme des communes du service (requête associées, *R07a\_recherche des communes multiservices*, *R07b\_services contenant des comm multiserv*, *R07c\_recherche différence nb commune du service*)
- 9- *R06\_MAJ service à partir T Tempo* : Le champ retenu de tous les services de la table *T Tempo* est mis à 0
- 10- *R05\_Ajout service a T Tempo* et *R06\_MAJ service à partir T Tempo* : Les procédures 5 et 6 sont répétées autant de fois que le nombre de services ajouté à la table *T Tempo* n'est pas nul en relançant successivement ces deux requêtes.
- 11- *R08\_Bilan* : sélection des communes restantes pour export et traitement sous R.
- 12- *R09\_export Services1\_2013* : Export *services1\_2013*

Ces étapes étant laborieuses et ne pouvant pas être cumulées sur les années précédentes comme pour la base 1, la recherche de méga-services n'a été effectuée que sur l'année 2013.

La prochaine étape consiste à déterminer quels services doivent être fusionnés entre eux.

Cette étape, et le regroupement des données par méta-services obtenus étant complexe, elle a été réalisée préférablement sur le logiciel de calcul R.

Script R :

*#somme des valeurs de la colonne Retenu en regroupant numéro de communes*

```
base_com_uniq=aggregate(base$Retenu,subset(base, select=c("N_INSEE_commune")),sum)
```

#sélection des communes seulement pour lesquelles la somme précédente est >1, ce sont les communes multiservices et appartenant à des services renseignés)

```
base_com_uniq=subset(base_com_uniq,(x>1),select=c("N_INSEE_commune","x"))
```

#récupération des infos associés à la sélection précédente

```
service=merge(base,base_com_uniq,by="N_INSEE_commune")
```

#dataframe précédent en ordre croissant d'ID service

```
service=service[order(service$Id_SISPEA_du_service),]
```

```
service$N_INSEE_commune=as.character(service$N_INSEE_commune) #conversion de la  
colonne insee commune en texte
```

```
#####
```

#Initialisation des variables de la boucle

```
e=data.frame()
```

```
a=NULL;b=NULL;d=NULL
```

```
j=1
```

#Boucle sur le numéro de service

```
for (i in (1:nrow(service))) {
```

```
  a=service$Nv_Id[i]
```

```
  ee=as.matrix(e)
```

#Est ce que l'Id service est déjà dans la matrice, si oui, il a déjà été traite

```
  if (!a %in% ee) {
```

```
    b=service$N_INSEE_commune[which (service$Nv_Id==a)]
```

```
    d=unique(service$Nv_Id[which (service$N_INSEE_commune %in% b)])
```

```
    r=0
```

```
    while (r==0) {
```

```
      b=service$N_INSEE_commune[which (service$Nv_Id %in% d)]
```

```
      d2=unique(service$Nv_Id[which (service$N_INSEE_commune %in% b)])
```

```
      if (length(d2)>length(d)) {d=d2}
```

```
      else {r=1;d=d2}
```

```
    }
```

#Sinon, creation d'une nouvelle ligne dans la matrice

```
  e[j,1]=a
```

```
  service$Nv_Id[which (service$Nv_Id %in% d)]=a
```

```
  d=d[which(d!=a)]
```

```
  for (l in 1:length(d)) {e[j,l+1]=d[l] }
```

```

j=j+1
}
}
#####
#Remplacement service dans base complète
ee=as.matrix(e)
for (i in 1:nrow(ee)) {
  le=length(e[i,][!is.na(e[i,])])
  dd=e[i,2:le]
  base$Nv_Id[which (base$Nv_Id %in% dd)]=e[i,1]
}
#Mise a jour service
service2013<-read.csv2("Data/Multiserv/Services1_2013.csv",as.is=TRUE) #chargement de la
base
ee=as.matrix(e)
for (i in 1:nrow(ee)) {
  le=length(e[i,][!is.na(e[i,])])
  dd=e[i,2:le]
  service2013$Nv_Id[which (service2013$Nv_Id %in% dd)]=e[i,1]
}
#regroupement des informations par méga-service
communes_routes1<-read_excel("csv/Résultats_communes_tout_1902.xlsx")
serv=subset(base,base$Nv_Id %in% e[,1])
serv2=serv[,c(2:4,10:12,16:17)]
serv3=aggregate(serv2[,c(2:8)], subset(serv2, select=c("N_INSEE_commune")), max)
routes_communes_fusion<-merge(serv3,communes_routes1,by="N_INSEE_commune")
routes_communes_fusion$Routes_urbaines_km[which(is.na(routes_communes_fusion$Routes_u
rbaines_km))]=0
routes_communes_fusion$Surface_urbaine[which(is.na(routes_communes_fusion$Surface_urbai
ne))]=0
routes_communes_fusion$Surface_urbaine[which(is.na(routes_communes_fusion$Surface_tamp
on250))]=0
comm.routes.somme=aggregate(routes_communes_fusion[,c(3,10,13:18)],subset(routes_commu
nes_fusion, select=c("Nv_Id")),sum) #agrégation des caractéristiques de communes selon méga-
service

```

```
services_à_fusionner=subset(service2013,service2013$Nv_Id %in% ee)
```

```
#services_à_fusionner
```

```
mega_services1=aggregate(services_à_fusionner[14:15],subset(services_à_fusionner,  
select=c("Nv_Id")),sum) #agrégation des caractéristiques services selon méga-service
```

```
mega_services2=merge(comm.routes.somme,mega_services1,by="Nv_Id")
```

```
mega_services3=merge(meg.alt,mega_services2,by="Nv_Id")
```

```
mega_services3$surface_totale_ha=mega_services3$Surface_urbaine+(mega_services3$Surface_t  
ampon250/10000) #calcul de la surface total pour les méga services
```

```
#agences de l'eau
```

```
library(plyr)
```

```
pl<-function(x) x[which.max(nchar(x))]
```

```
ag<-ddply(serv,.(Nv_Id),summarize,pl(Agence.s._de_leau))
```

```
mega_services4=merge(mega_services3,ag,by="Nv_Id")
```

```
mega_services4$Nv_Id=mega_services4$Nv_Id+9000000 #changement des noms des méga-  
services créés pour facilement les retrouver
```

## A5 : Résultats Modèle K1

Lineaire\_aep ~ 0+Route\_filtrees\_tampon250\_km:Agence:densite

Coefficients	Estimate	Std.Error	t-value	Pr(> t )
Route_filtrees_tampon250_km:AgenceAEAG:densite>400	0.796684	0.014463	55.08	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAEAP:densite>400	0.846861	0.008350	101.43	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAELB:densite>400	0.773140	0.004330	178.57	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERM:densite>400	0.849682	0.013184	64.45	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERMC:densite>400	0.641362	0.006792	94.43	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAESN:densite>400	0.820648	0.011302	72.61	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAEAG:densite120	0.875977	0.004836	181.15	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAEAP:densite120	0.762524	0.035948	21.21	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAELB:densite120	0.937698	0.003711	252.70	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERM:densite120	0.946900	0.017790	53.23	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERMC:densite120	0.745097	0.006751	110.37	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAESN:densite120	0.922052	0.011330	81.38	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAEAG:densite400	0.790268	0.012646	62.49	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAEAP:densite400	0.828335	0.031399	26.38	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAELB:densite400	0.838130	0.006431	130.33	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERM:densite400	0.840841	0.017976	46.78	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAERMC:densite400	0.741350	0.006729	110.18	<2e-16 ***
Route_filtrees_tampon250_km:AgenceAESN:densite400	0.869863	0.019102	45.54	<2e-16 ***

Residual standard error: 23.27 on 5015 degrees of freedom  
Multiple R-squared: 0.9777, Adjusted R-squared: 0.9777  
F-statistic: 1.224e+04 on 18 and 5015 DF, p-value: < 2.2e-16



## A6 : Résultats Modèle K2

Lineaire\_aep ~ 0+Agence:RouteF\_Class1 + Agence:RouteF\_Class2+ Agence:RouteF\_Class3+  
Agence:RouteF\_Class4+ Agence:RouteF\_Class5

Coefficients	Estimate	Std.Error	t-value	Pr(> t )
AgenceAEAG:RouteF_Class1	-8.264e-06	4.667e-05	-0.177	0.859453
AgenceAEAP:RouteF_Class1	-1.168e-04	7.708e-05	-1.515	0.129777
AgenceAELB:RouteF_Class1	1.752e-04	2.670e-05	6.564	5.75e-11 ***
AgenceAERM:RouteF_Class1	4.020e-04	8.237e-05	4.880	1.09e-06 ***
AgenceAERMC:RouteF_Class1	-5.878e-05	3.663e-05	-1.605	0.108589
AgenceAESN:RouteF_Class1	8.900e-05	4.323e-05	2.059	0.039566 *
AgenceAEAG:RouteF_Class2	9.483e-04	1.120e-05	84.643	<2e-16 ***
AgenceAEAP:RouteF_Class2	5.363e-04	1.051e-04	5.104	3.44e-07 ***
AgenceAELB:RouteF_Class2	9.290e-04	9.015e-06	103.049	<2e-16 ***
AgenceAERM:RouteF_Class2	3.035e-04	7.584e-05	4.002	6.38e-05 ***
AgenceAERMC:RouteF_Class2	7.215e-04	1.543e-05	46.751	<2e-16 ***
AgenceAESN:RouteF_Class2	9.139e-04	2.044e-05	44.716	<2e-16 ***
AgenceAEAG:RouteF_Class3	1.252e-04	6.718e-05	1.864	0.062418
AgenceAEAP:RouteF_Class3	1.207e-03	2.192e-04	5.505	3.88e-08 ***
AgenceAELB:RouteF_Class3	6.491e-04	5.404e-05	12.012	<2e-16 ***
AgenceAERM:RouteF_Class3	1.782e-03	1.476e-04	12.073	<2e-16 ***
AgenceAERMC:RouteF_Class3	8.250e-04	5.050e-05	16.336	<2e-16 ***
AgenceAESN:RouteF_Class3	8.061e-04	1.080e-04	7.467	9.60e-14 ***
AgenceAEAG:RouteF_Class4	1.242e-03	8.325e-05	14.924	<2e-16 ***
AgenceAEAP:RouteF_Class4	8.592e-04	1.496e-04	5.742	9.92e-09 ***
AgenceAELB:RouteF_Class4	6.590e-04	6.294e-05	10.470	<2e-16 ***
AgenceAERM:RouteF_Class4	4.673e-04	1.333e-04	3.507	0.000458 ***
AgenceAERMC:RouteF_Class4	7.192e-04	5.781e-05	12.440	<2e-16 ***
AgenceAESN:RouteF_Class4	6.062e-04	1.001e-04	6.056	1.50e-09 ***
AgenceAEAG:RouteF_Class5	7.861e-04	5.617e-05	13.997	<2e-16 ***
AgenceAEAP:RouteF_Class5	8.104e-04	7.426e-05	10.914	<2e-16 ***
AgenceAELB:RouteF_Class5	7.808e-04	2.401e-05	32.524	<2e-16 ***
AgenceAERM:RouteF_Class5	8.525e-04	6.244e-05	13.654	<2e-16 ***
AgenceAERMC:RouteF_Class5	5.301e-04	2.250e-05	23.561	<2e-16 ***
AgenceAESN:RouteF_Class5	8.808e-04	3.867e-05	22.777	<2e-16 ***

Residual standard error: 22.9 on 5003 degrees of freedom  
Multiple R-squared: 0.9785, Adjusted R-squared: 0.9784  
F-statistic: 7589 on 30 and 5003 DF, p-value: < 2.2e-16

### A7 : Scripts R – Résidus graphiques

```
windows()
```

```
par(mfrow=c(2,2))
```

```
plot(ModK1,labels.id = base_net$Id_SISPEA_service, id.n = 6,cex.id=0.8)
```

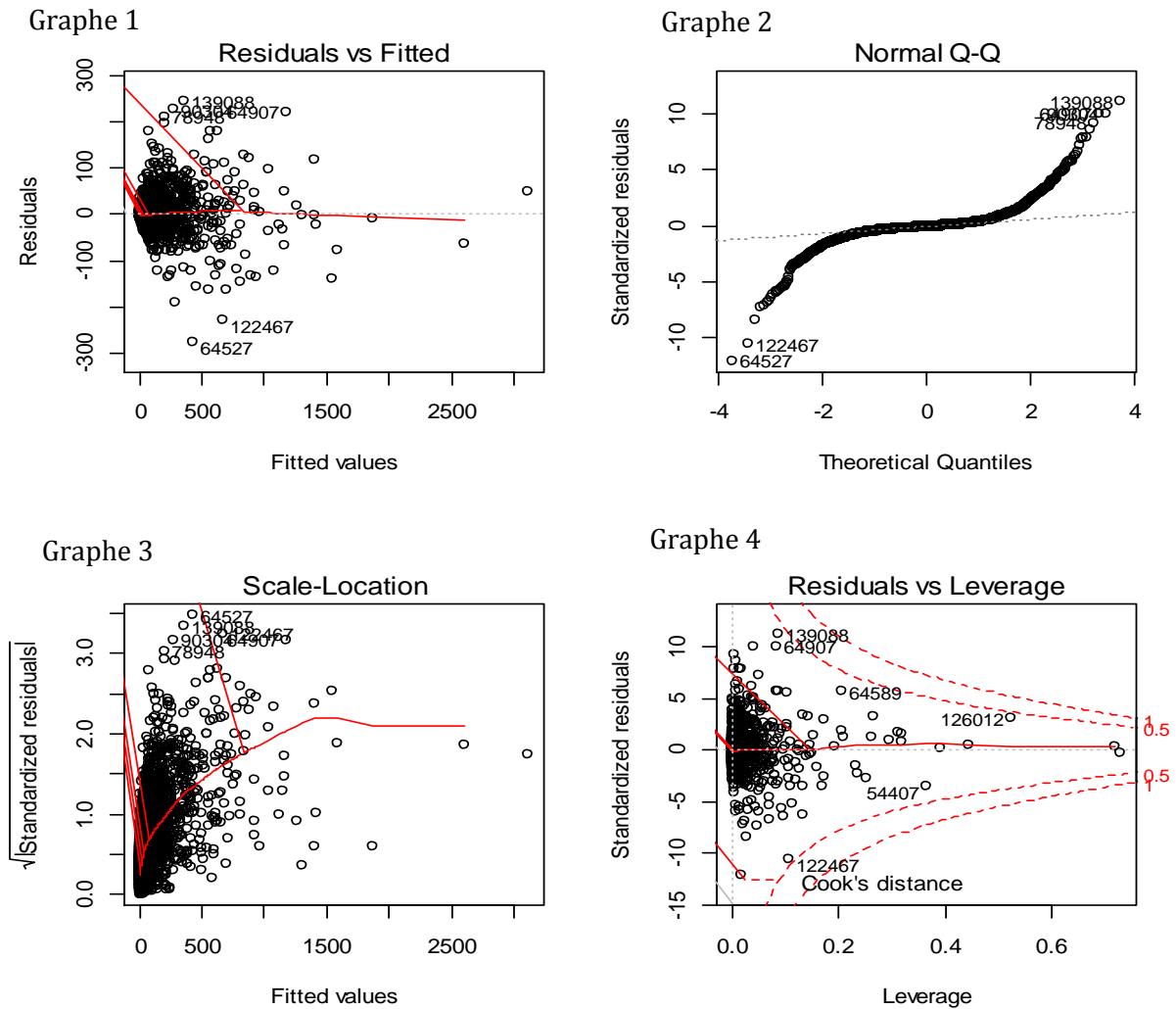


Figure 41: Résidus du modèle K1

Graphe 1 :

Ce graphique permet de voir si les résidus suivent une tendance linéaire ou non. Que les résidus soient dispersés aléatoirement d'une part et d'autre de la ligne rouge horizontale, alors ils ne suivent a priori pas de tendance linéaire ; aucun facteur important n'a donc été omis lors de notre modélisation.

Graphe 2 :

Le graphique 2 compare la distribution des résidus à une distribution normale. Le graphique obtenu suit en partie la ligne en pointillé mais diverge pour le début et la fin de la distribution. La distribution de nos résidus n'est donc pas parfaitement normale, il conviendra d'enlever les services les plus éloignés de la ligne de distribution normale.

Graphe 3 :

Ce graphique présente la distribution des résidus par rapport aux valeurs prédites de manière à observer la distribution des variances au sein des résidus. Une modélisation parfaite devrait permettre d'observer une ligne rouge horizontale avec les points dispersés de manière égale et sans structure a priori de part et d'autre de la ligne. Nous observons ici que la plupart des points sont en général dispersés de manière similaire des deux côtés de la ligne, cette dernière n'est cependant pas horizontale dû à la grande concentration de résidus en début du graphique. Nous savons que notre base contient majoritairement des petits services et que ces derniers induisent quelques erreurs.

Graphe 4 :

Le dernier graphique présente les résidus par rapport aux courbes de Cook et permet de déceler si certains points influencent le reste de la modélisation de manière trop importante. De manière générale, les points à l'extérieur au-dessus de la courbe supérieure et en-dessous de la courbe inférieure sont à exclure. Nous voyons ici que tous les services considérés sont bons par rapport à ce critère.

## A8 : Scripts R – Exemple de modélisation et évaluation de l'estimation

```

ModK1<- lm((Lineaire_aep)~0+Route_filtrees_tampon250_km:Agence:densite
, data=base_net_reduite1)
summary(ModK1) #visualisation des coefficients de la modélisation
plot(ModK1$residuals~log(base_net_reduite1$Total_populations_communes))
#visualisation de la Figure 30
prediction=(predict(ModK1,newdata=base_net,type="response",se.fit=TRUE))
#extrapolation sur la base1
base_net$predict=(prediction$fit)
base_net$se1=prediction$se.fit #écart-type
base_net$var1=prediction$se.fit^2 #variance
#visualisation des résultats
windows()
plot(base_net$predict~base_net$Lineaire_aep)

```

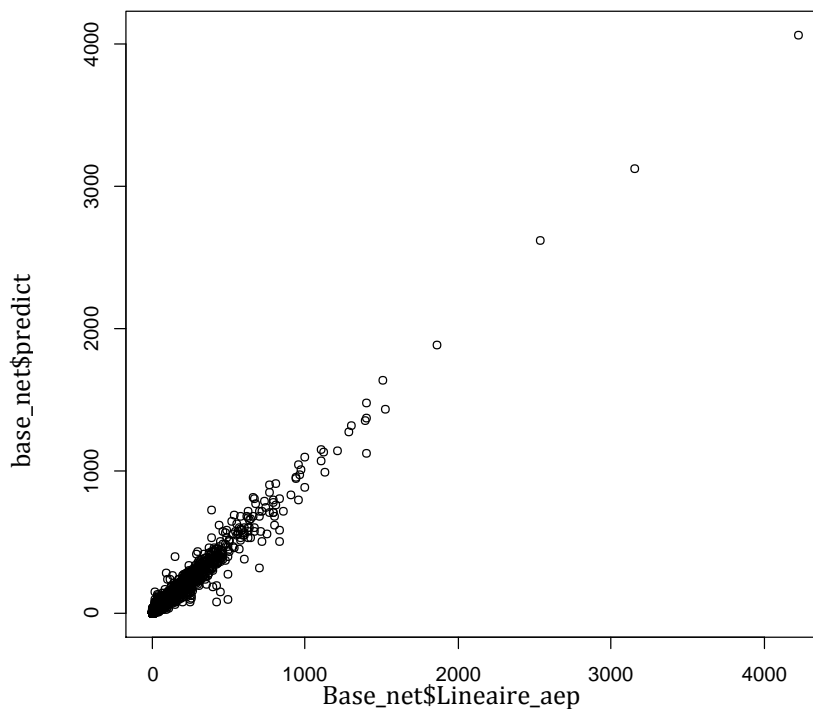


Figure 42: Valeurs prédites par le modèle par rapport aux valeurs réelles

### #calcul des sommes de linéaires estimées

```
(Lineaire_serv=sum(base_net$Lineaire_aep))
```

```
(Lineaire_serv_pred=sum(base_net$predict))
```

```
(Lineaire_serv_95p=Lineaire_serv_pred+1.96*sqrt(sum(base_net$var1))) #borne inférieure de l'intervalle à 95%
```

```
(Lineaire_serv_95m=Lineaire_serv_pred-1.96*sqrt(sum(base_net$var1))) #borne supérieure de l'intervalle à 95%
```

## A9 : Tableaux récapitulatifs des résultats de tous les modèles créés

Base	Laep réel	Modèle 1 (tampon250)			Modèle 2 (Class population)				Modèle 3 (Class Densité)			
		Lmod.serv	%	Lmod.com	Lmod.serv	%	Lmod.com	%	Lmod.serv	%	Lmod.com	%
Base 1	350920,8	352547,6	100,5%		352687,95	100,5%	341914,6	97,4%	362617,7	103,3%	362324,5	103,2%
Base 2	55392,6	53220,14	96,1%		53246,4	96,1%	52009,7	93,9%	54475,98	98,3%	55133,92	99,5%
Base 3 (calcul au prorata)	203696,3	284176,684	140%		284807,497	139,8%	262787,2	129%	296372,397	145,5%	285717,6	140%
France entière				883985,2			852692,23				925559,19	

Base	Laep réel	Modèle K1				Modèle K2			
		Lmod.serv	%	Lmod.com	%	Lmod.serv	%	Lmod.com	%
Base 1	350920,8	353224,7	100,60%	352915,6	100,57%	352558,5	100,47%	352733,4	100,52%
Base 2	55392,6	53796,1	97,10%	53338,65	96,3%	53226,8	96,10%	53244,97	96,10%
Base 3 (calcul au prorata)	203696,3								
France entière					894213,0				896403,5

## MEMOIRE DE FIN D'ETUDES

**Diplôme(s)** Diplôme d'Ingénieur de l'ENGEES

**Spécialité** GEDEMU (Gestion Durable de l'Eau en Milieux Urbains)

**Auteur :** Kimberley Han Mui

**Année**  
2016

### Titre

Etude des déterminants et modélisation de la consistance des réseaux d'eau potable

**Nombre de pages**      texte 45                      annexes 20

**Nombre de références bibliographiques** 7

**Structure d'accueil** Equipe GPIE, IRSTEA Bordeaux, France

**Maître de stage :** M. Alain Husson

### Résumé

L'étude vise à construire un dispositif permanent d'évaluation du patrimoine des réseaux d'eau potable aux échelles nationale et de bassin. Les réseaux routiers ont été privilégiés pour l'évaluation du linéaire. Les modèles construits ont pris en compte la présence de bâtis, la démographie et la position géographique des services. Pour le calage des modèles, les données de la base SISPEA ont été utilisées. La validation des modèles s'est effectuée en plusieurs étapes, notamment par la méthode de Monte-Carlo. Le modèle final permet d'évaluer le linéaire de réseau d'eau potable à l'échelle nationale; cela est moins évident à l'échelle du service.

### Mots-clés

Gestion patrimoniale, modélisation linéaire, modélisation multilinéaire, base SISPEA, Linéaire de réseau, Eau potable, territoire France métropolitaine, Système d'informations géographiques, statistiques analytiques