



# Modélisation des liens entre la morphologie urbaine, la génération de polluants et leur concentration dans l'air par l'approche PLS path modelling

Samba Diop

## ► To cite this version:

Samba Diop. Modélisation des liens entre la morphologie urbaine, la génération de polluants et leur concentration dans l'air par l'approche PLS path modelling. Modélisation et simulation. 2018. hal-01880194

HAL Id: hal-01880194

<https://hal.inria.fr/hal-01880194>

Submitted on 25 Sep 2018

**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.

**Modélisation des liens entre la  
morphologie urbaine, la génération de  
polluants et leur concentration dans l'air  
par l'approche PLS path modelling**



Samba Fall DIOP

Master 2 MIASHS C2ES

Tuteurs professionnels:

Peter STURM et Serge FENET

Directeur de mémoire: Sandrine MATHY

Université Grenoble-Alpes

17 Juillet, 2018

# Remerciements

Je tiens à remercier tout spécialement mes deux tuteurs de stage, Serge FENET et Peter STURM, pour leur confiance, tous leurs précieux temps qu'ils m'ont accordé et les conseils avisés qu'ils ont prodigué tout le long de ma mission afin que celle ci se déroule dans les meilleures conditions. Mes remerciements s'adressent également Sandrine Mathy, ma directrice de mémoire, également pour ses conseils utiles pour la rédaction de celui-ci.

Je remercie enfin l'ensemble des membres de l'équipe STEEP, pour leur disponibilité et ouverture d'esprit ; qui m'a permis tout de suite de m'y intégrer de façon immédiate.

# Table des matières

<b>Table des figures</b>	<b>5</b>
<b>Liste des tableaux</b>	<b>5</b>
<b>I</b>	<b>7</b>
<b>1 Déroulement du stage</b>	<b>7</b>
<b>2 Présentation de l'Institut nationale de recherche en informatique et automatique</b>	<b>7</b>
<b>3 Présentation de l'équipe STEEP</b>	<b>7</b>
<b>4 Les missions effectuées</b>	<b>8</b>

<b>5</b>	<b>Les apports du stage</b>	<b>8</b>
5.1	Les compétences acquises et développées . . . . .	8
5.2	Les difficultés rencontrées et les solutions apportées . . . . .	9
<b>6</b>	<b>Conclusion</b>	<b>9</b>
<b>II</b>		<b>10</b>
<b>1</b>	<b>Revue de la littérature et contexte de travail</b>	<b>11</b>
1.1	Qu'est-ce l'étalement urbain . . . . .	11
1.2	Présentation générale des mesure de l'étalement urbain . . . . .	12
1.3	Présentation d'indicateurs d'étalement urbain calculés sur données ouvertes . . . . .	14
1.3.1	Mixité d'usage du sol . . . . .	15
1.3.2	Accessibilité des activités . . . . .	17
1.3.3	Dispersion du bâti . . . . .	18
1.4	Causes de l'étalement urbain . . . . .	18
1.4.1	Sources démographiques . . . . .	19
1.4.2	Préférences individuelles . . . . .	19
1.4.3	Baisse des coûts de transports . . . . .	20
1.4.4	Imperfections du marché . . . . .	21
1.4.5	Politiques publiques . . . . .	21
1.4.6	La topographie . . . . .	22
1.5	Conséquences de l'étalement urbain . . . . .	23
1.5.1	La consommation énergétique . . . . .	23
1.5.1.1	Le secteur des transports . . . . .	23

1.5.1.2	Le secteur résidentiel . . . . .	26
1.5.2	L'émission de polluants et dégradation de la qualité de l'air .	27
1.5.3	Les effets de l'étalement urbain sur la santé . . . . .	28
<b>2</b>	<b>Méthodologie</b>	<b>29</b>
2.1	Choix du modèle . . . . .	30
2.2	Quelques mots sur les modèles d'équations structurelles . . . . .	30
2.3	Bref historique de la méthode PLS path modeling . . . . .	31
2.4	Principes de l'approche PLS path modeling . . . . .	31
2.4.1	Le modèle externe ou de mesure . . . . .	31
2.4.2	Le modèle interne ou structurel . . . . .	34
2.5	Fonctionnement de l'algorithme PLS-PM . . . . .	34
2.5.1	Estimation des variables latentes . . . . .	35
2.5.2	Estimation externe . . . . .	35
2.5.3	Estimation interne . . . . .	36
2.5.4	Mise à jour des poids dans l'estimation externe . . . . .	36
2.5.5	Vérification de la convergence . . . . .	37
2.6	Évaluation du modèle . . . . .	37
2.6.1	Évaluation du modèle externe . . . . .	38
2.6.2	Évaluation du modèle structurel . . . . .	40
2.6.3	La qualité de l'ajustement global du modèle structurel . . . .	40
2.7	Présentation du modèle a priori . . . . .	41
2.7.1	Modèle structurel . . . . .	42
2.7.2	Modèle de mesure . . . . .	43
2.8	Présentation de l'échantillon de données . . . . .	44
2.9	Description de l'échantillon . . . . .	45

<b>3</b>	<b>Résultats et discussion</b>	<b>48</b>
3.1	Modèle initial . . . . .	48
3.2	Modèle final . . . . .	50
<b>4</b>	<b>Conclusion</b>	<b>56</b>
<b>5</b>	<b>Limites</b>	<b>57</b>
<b>6</b>	<b>Références</b>	<b>58</b>
	<b>Annexes</b>	<b>64</b>

## Table des figures

1.1	(a) Image satellite de Grenoble, (b) mixité d'usage du sol . . . . .	16
1.2	(a) Image satellite de Grenoble, (b) accessibilité aux activités . . . . .	17
1.3	(a) Image satellite de Grenoble, (b) dispersion du bâti . . . . .	18
1.4	Diverses causes de l'étalement urbain (Source : auteur) . . . . .	24
1.5	Relation entre la densité de population (en abcisse) et la consommation en pétrole par habitant(en ordonnée). Source : Newman et Kentworthy (1989) . . . . .	26
1.6	Conséquences de l'étalement urbain (Source : auteur) . . . . .	29
2.1	Trois blocs de variables sous différents schémas (Source : auteur) . . . . .	32
2.2	Structure du modèle PLS path modelling (Source : auteur) . . . . .	34
2.3	Algorithme PLS path modelling (Esposito Vinzi, 2005 : 5th symposium on PLS and related methods) . . . . .	35
2.4	Modèle structurel a priori. La couleur bleue designe une neutralité de la variable, le vert designe un effet attendu positif sur le plan environnemental, le rouge designe l'effet inverse. Une flèche de a à b signifie que a influence b (Source : auteur) . . . . .	43
3.1	Modèle 1 (Source : auteur) . . . . .	48
3.2	Modèle 2 (Source : auteur) . . . . .	51
.1	(a) Image satellite de Chambéry, (b) mixité d'usage du sol . . . . .	65
.2	(a) Image satellite de Chambéry, (b) accessibilité aux activités . . . . .	65
.3	(a) Image satellite de Chambéry, (b) dispersion du bâti . . . . .	66

## Liste des tableaux

1	Evaluation PLS path modelling (Source : auteur) . . . . .	41
---	---	----

2	Statistiques descriptives (Source : auteur) . . . . .	47
3	Poids externes modèle 1 (Source : auteur) . . . . .	49
4	Inidimensionnalité :modèle 2 (Source : auteur) . . . . .	51
5	Poids externes du modèle 2 (Source : auteur) . . . . .	53
6	Résumé statistique du modele interne 2 (Source : auteur) . . . . .	54
7	Effets direct, indirects et totaux (Source : auteur) . . . . .	55
8	Modèle externe 2 (Source : auteur) . . . . .	67
9	Crossloadings du modèle 2 (Source : auteur) . . . . .	68
10	Modele interne 1 (Source : auteur) . . . . .	69
11	Dictionnaire des variables . . . . .	70

# Première partie

## 1 Dérroulement du stage

J'ai effectué un stage au sein de l'équipe STEEP de l'Inria sur le site de Monbonnot-Saint-Martin. Au cours de ce stage j'ai pu m'intéresser à la programmation, au traitement de données massives et à la modélisation statistique. Plus généralement, ce stage a été l'occasion pour moi de découvrir le monde de la recherche, de travailler dans un environnement où la rigueur scientifique et l'ouverture d'esprit se conjuguent parfaitement, ce qui m'a permis de m'y épanouir professionnellement et sur le plan personnel. Au-delà de m'apporter une myriade de nouvelles connaissances, ce stage m'a permis de saisir l'importance de la communication au sein d'une équipe et son grand potentiel dans l'apport de solutions, même dans les situations les plus complexes.

## 2 Présentation de l'Institut nationale de recherche en informatique et automatique

Créée en 1967, l'institut Nationale de Recherche en Informatique et Automatique (Inria) est une organisation publique un à caractère scientifique et technologique spécialisé dans les domaines de l'informatique et des mathématiques. Il est sous la tutelle du ministère de l'Enseignement supérieur et du ministère de l'Économie et des Finances. Il est organisé en équipes-projets au sein desquelles des chercheurs aux compétences complémentaires travaillent autour d'un projet scientifique déterminé.

## 3 Présentation de l'équipe STEEP

L'équipe STEEP ou Soutenabilité, Territoires, Environnement, Economie et Politique est une équipe orientée vers la modélisation des systèmes mettant en relation

les acteurs sociaux, économiques et environnementaux dans le cadre d'une transition vers le développement durable à l'échelle des régions. Elle a débuté en 2010 en tant qu'« action exploratoire » avant de devenir une « Équipe projet Inria ». Elle est également affiliée au laboratoire Jean Kuntzmann (LJK). Sa mission est de mettre à disposition des outils d'aide à la décision qui s'appuient sur des logiciels de simulation, de visualisation et d'optimisation numérique.

## **4 Les missions effectuées**

Mon stage au sein de l'équipe projet STEEP a consisté dans un premier temps en la réalisation d'une revue de la littérature sur la question de l'étalement urbain. Dans un deuxième temps, il a été question d'étudier et de s'approprier l'outil de calcul des indices de forme urbaine et de réaliser quelques modifications afin de corriger un biais relatif au calcul des indices. Ce fut un vrai challenge pour moi, n'étant pas issu d'une formation spécialisée en informatique. Mais après plusieurs essais, recherches j'ai pu acquérir progressivement une compréhension du fonctionnement des programmes informatiques et ainsi réussi à communiquer plus ou moins bien avec la machine dans son langage.

## **5 Les apports du stage**

### **5.1 Les compétences acquises et développées**

Ce stage m'a permis d'approfondir mes connaissances en informatique et de développer des connaissances très variées des divers outils informatiques nécessaires au traitement et à l'analyse des données. En effet j'ai eu l'opportunité de suivre une formation python à Lyon à l'occasion de laquelle j'ai acquis des notions essentielles en programmation orientée objet, de parallélisation de calculs et de tests unitaires de code. Ces nouvelles compétences ont eu la qualité d'être spontanément utiles dans la pratique ; en m'ayant servi par la suite à manipuler aisément les données, en automatisant par exemple diverses opérations sur des volumes importants d'information.

Ayant l'opportunité de travailler dans une équipe pluridisciplinaire, les différents échanges que j'ai pu avoir sur les questions du développement durable, des moyens de l'appliquer au niveau régional m'ont permis de faire du recul dans ma réflexion sur la résolution de la problématique de mon mémoire, qui a été d'étudier l'étalement urbain dans une approche systémique.

## **5.2 Les difficultés rencontrées et les solutions apportées**

Durant les premiers mois du stage, j'ai eu des difficultés à choisir une méthode pour traiter le problème posé par la mission qui m'a été confiée. En effet les pistes de réflexions me semblaient infinies, me poussant à expérimenter plusieurs modèles sans pour autant être satisfait des résultats. En effet les méthodes courantes utilisées en statistiques et en économétrie consistent à expliquer une variable  $Y$  par une ou plusieurs autres variables  $X$ . Or l'étalement urbain est un concept complexe mettant en jeu plusieurs facteurs qu'il fallait donc d'une part expliquer et d'autre part utiliser pour prédire la consommation d'énergie et l'émission de polluants, qui sont à leur tour des concepts mettant en jeu plusieurs variables. C'est alors en multipliant les discussions avec mes tuteurs de stage et en poussant plus loin les recherches que je suis parvenu progressivement à comprendre les particularités de la question de recherche et à obtenir une piste de résolution satisfaisante.

## **6 Conclusion**

Ce stage fut enrichissant pour moi car j'ai pu développer des compétences indispensables à ma formation de chargé d'étude économiques et statistiques. Il m'a aussi permis de comprendre ce qu'est le travail de recherche et à prendre du recul devant des questions complexes en les projetant sur la réalité pour pouvoir y apporter des réponses pratiques et utiles. J'ai également pu y développer un grand intérêt pour le travail en équipe et la réflexion en groupe. Cette expérience a confirmé mon intérêt de poursuivre ma carrière professionnelle dans le domaine de la recherche.

## Deuxième partie

### Introduction

En 2014, dans son rapport sur les perspectives de l'urbanisation, l'organisation des Nations Unies a fait état d'une proportion de la population mondiale vivant en milieu urbain qui est passée de 30% en 1950 à 54% en 2014 et qui devrait atteindre 66% en 2050 [Wup, 2014]. Dans ce contexte, l'étalement urbain en particulier, est une forme d'urbanisation dont les effets en termes de développement durable soulèvent plusieurs inquiétudes. Selon l'Agence européenne pour l'environnement, l'étalement urbain est à l'origine en Europe de conflits d'usage du sol et représente une grande menace quant à l'utilisation durable de l'espace [Hennig et al., 2016].

La préservation des espaces naturels et agricoles, la maîtrise de la consommation d'énergie, de la pollution et le bien être des populations sont au cœur des préoccupations actuelles avec l'urbanisation croissante des territoires. En France, les schémas de cohérence territoriale (SCoT), apparus avec la loi Solidarité et renouvellement urbain du 13 décembre 2000, ont été conçus comme outils de planification stratégique à différentes échelles territoriales avec comme objectif l'application des principes du développement durable, ceci notamment en veillant à la cohésion des différents programmes sectoriels. Or atteindre ces objectifs suppose une connaissance suffisante des interactions entre ces paramètres que l'on cherche à maîtriser. Comme nous le verrons plus loin ; l'étalement urbain était largement qualifié de non durable en tant que modèle d'urbanisation mais les mécanismes qui y conduisent et ceux inhérents à ses effets néfastes ne sont pas simultanément pris en compte dans la littérature. Or cette approche peut se révéler trompeuse car ces mécanismes sont en général interdépendants et ainsi les résultats obtenus dans une grande majorité des travaux sont potentiellement très biaisés. Dans ce contexte, nous nous sommes posés la question de la possibilité d'analyser empiriquement l'étalement urbain en amont en déterminant ses causes, et en aval en évaluant ses effets simultanément sur plusieurs facteurs liés à la qualité de l'environnement et ceci dans un même modèle. Pour répondre à cette problématique nous avons tenté de construire un modèle guidé par la littérature autour du concept d'étalement urbain et par les

données sélectionnées avec le plus grand souci de représentativité des phénomènes que nous souhaitons mesurer. Notre objectif a été de tester, qualifier et quantifier les mécanismes allant de la genèse de la morphologie urbaine aux effets de celle-ci sur la l'émission de pollution et la dégradation de la qualité de l'air.

En préambule, nous allons présenter les travaux sur la mesure, les causes et les conséquences de l'étalement. Ensuite nous introduirons dans un deuxième temps la méthodologie adoptée dans la conduite de cette étude. Dans un troisième temps les résultats de l'analyse seront présentés avant d'exposer les principales conclusions de ce travail de recherche et ses limites dans une dernière étape.

## 1 Revue de la littérature et contexte de travail

### 1.1 Qu'est-ce l'étalement urbain

Bien qu'il n'y ait pas de consensus sur une définition de la notion d'étalement urbain au sein de la littérature, il existe un ensemble de concepts souvent utilisés pour la définir. Sa définition varie selon le fait qu'on cherche à mettre en évidence les processus qui l'accompagnent ou que l'on vise une description des caractéristiques physiques du territoire où le phénomène est observé. Dans le premier cas de figure, [Balestrat et al., 2011] par exemple, le définissent comme une dynamique accompagnée d'une augmentation accélérée et non planifiée des surfaces artificialisées au détriment des surfaces agricoles et des espaces naturels. [Sainteny, 2008] donne une description similaire en le définissant comme un découplage entre la croissance démographique et l'artificialisation des sols. Dans la seconde approche, on peut citer [Ewing, 1997], [Burchell et al., 1998], [Galster et al., 2001], parmi un grand nombre d'auteurs, qui le caractérisent entre autres par une faible densité de population et du bâti, un aménagement discontinu et fragmenté de l'espace, une ségrégation spatiale des différentes catégories d'usage du sol et une mauvaise accessibilité à ces dernières [Ewing and Rong, 2008].

## 1.2 Présentation générale des mesure de l'étalement urbain

Il existe une diversité d'approches dans la littérature pour mesurer l'étalement urbain qui aboutissent parfois à des interprétations contradictoires, l'objectif étant de passer d'un concept subjectif et qualitatif à une mesure objective et quantitative de l'étalement urbain [Hamidi et al., 2015]. La densité de population et celle de la surface bâtie sont fréquemment utilisées pour quantifier ce phénomène dans la littérature [Enault, 2004].

Dans ce registre, on peut citer [Lopez and Hynes, 2003], qui ont défini un score d'étalement comme étant la différence entre la proportion de la population urbaine vivant dans un secteur de recensement à faible densité (entre 3500 et 200 habitants par km<sup>2</sup>) et celle vivant dans un secteur à haute densité (supérieure à 3500 habitants par km<sup>2</sup>). On y retrouve également les travaux [Anthony, 2004] qui en étudiant l'efficacité de la législation pour la gestion de l'expansion urbaine à travers 49 états des USA, mesure l'étalement urbain par le changement en densité des zones correspondantes sur une période de quinze ans. Les travaux de [Yue et al., 2013] pour analyser l'étalement urbain et ses causes pour la ville chinoise de Hangzhou utilisent eux aussi la densité de population comme indicateur, ainsi que l'expansion de la surface urbanisée. Enfin, [Arti et al., 2013] ont examiné les liens entre le nombre de décharges dans la ville indienne de Chhattisgarh et l'étalement urbain, en adoptant la densité de population urbaine comme mesure pour ce dernier.

D'autres auteurs ont choisi un nombre plus important de variables pour définir l'étalement urbain. C'est le cas du travail de [Galster et al., 2001] qui l'ont défini à travers huit dimensions : la densité par miles carrés des unités résidentielles, le degré de concentration spatiale des aménagements dans une zone urbaine définie, la centralité ; qui est le degré de proximité des constructions au quartier d'affaire d'une aire urbaine, le degrés de regroupement et la continuité du bâti, le caractère nucléaire du bâti, l'usage mixte de l'espace entre activités et résidences et le niveau de proximité entre les différents types d'usages. L'étude comparative entre aires urbaines compactes et étalées sur données simulées réalisée par [Tsai, 2005] a utilisé également un lot de quatre variables pour caractériser la morphologie urbaine :

la superficie de l'aire urbaine, sa densité en activités, l'inégalité de la distribution spatiale et le degré de regroupement de la population ou des emplois. On peut citer également [Arribas-Bel and Schmidt, 2013] qui ont conduit une analyse exploratoire de la structure urbaine de 209 villes européennes. Ils ont pour cela défini six dimensions pour caractériser l'étalement urbain, qui sont le temps moyen du trajet vers le lieu de travail, la proportion de la population vivant en dehors du centre ville, la densité de population, le niveau de fragmentation du bâti, l'accessibilité aux espaces libres et la mixité d'usage du sol.

Une troisième catégorie d'auteurs ont développé des indices synthétiques par sous catégories constituant des dimensions de l'étalement urbain par l'application de techniques statistiques de réduction dimensionnelle de l'information. Par exemple [Burton, 2002] ; en étudiant la morphologie urbaine des villes du Royaume-Uni a calculé cinq indices d'étalement urbain (ou compacité urbaine) qui sont la somme de scores  $z$  (ou scores standards) des variables qui composent chaque catégorie. Au total 41 variables ont été utilisées pour définir les indices de densité, de mixité d'usage, d'intensification en population et en bâti. La moyenne de ces scores a été utilisée ensuite pour comparer les différentes villes.

Une approche plus complexe a été adoptée par [Ewing et al., 2003] pour mesurer l'étalement urbain dans les 101 plus grandes régions statistiques métropolitaines des États-Unis. Des lots de 7, 6, 6 et 3 variables sont sélectionnées pour représenter respectivement la densité, la mixité, la centralité et l'accessibilité. Un score par lot de variables a été calculé en retenant après une analyse en composantes principales (ACP) ; l'axe factoriel avec la plus grande part de variance expliquée. Ensuite les composantes sont standardisées, normalisées par la taille de la population puis sommées pour obtenir un indicateur unique d'étalement urbain. L'ACP permet d'effectuer une synthèse de l'information lorsqu'on est en présence d'une quantité importante de variables à analyser ([Guerrien, 2003]). Une technique assez similaire a été employée par [Frenkel and Ashkenazi, 2008] pour mesurer l'étalement urbain dans 78 zones urbaines en Israël sur une période de quinze ans. Ils définissent un indicateur intégré en utilisant l'analyse factorielle qui permet de quantifier une variable inobservable (ou variable latente) comme étant la combinaison linéaire de 13 variables représentant la densité de population, la fragmentation du bâti et la

mixité d’usage du sol. Une toute autre technique de réduction dimensionnelle des facteurs a été utilisée par [Tian et al., 2017] pour étudier l’expansion urbaine au sein de la ville de Shanghai qu’ils ont divisée en cinq zones. Les indicateurs d’étalement choisis sont l’expansion de la ville, la disponibilité d’infrastructures urbaines et la forme urbaine. Ils sont mesurés au total par 13 variables. La méthode d’analyse hiérarchique des procédés, consistant en une hiérarchisation, puis comparaison puis à une détermination de poids pour chaque indicateur [Cherrared et al., 2011] a été utilisée avant de calculer un score unique qui est une combinaison linéaire des indicateurs pondérés par ces poids.

Cependant, les différents indicateurs d’étalement urbain présents dans la littérature ne sont ni construits sur les mêmes données, ni aux mêmes échelles géographiques et territoriales. Ces différences font qu’il est difficile de comparer les différents travaux sur la question, sans potentiellement induire un nombre considérable de biais [Gervasoni et al., 2017]. Le caractère commercial des données et logiciels employés dans beaucoup de travaux est également un frein important à la comparabilité et la reproductibilité des différents résultats.

### **1.3 Présentation d’indicateurs d’étalement urbain calculés sur données ouvertes**

Dans le but d’apporter une réponse à cette problématique, [Gervasoni et al., 2017] ont développé une infrastructure logicielle de production d’indicateurs d’étalement urbain en libre accès écrit en Python (langage de programmation multiparadigme), utilisant en entrée les données issues de la plateforme OSM (OpenStreetMap). Ces données ont l’avantage d’être disponibles partout dans le monde et bien que toutes les zones géographiques ne soient pas renseignées au même niveau d’exhaustivité et de précision, de nombreuses études sur leur qualité ont souligné leurs multiples potentiels à servir dans de nombreux domaines d’application ([Arsanjani et al., 2015], [Zielstra et al., 2014]). Les trois dimensions d’étalement urbain mesurées par ces indices sont la mixité d’usage du sol, le niveau d’accessibilité des activités et le degré de dispersion du bâti dans la zone d’intérêt considérée.

Pour illustrer chaque indice, nous avons choisi l’exemple de la commune de Gre-

noble (et Chambéry en annexe).

### 1.3.1 Mixité d'usage du sol

Nous avons précédemment vu que l'étalement urbain a été défini comme étant caractérisé par une ségrégation territoriale des fonctions urbaines entraînant un mauvais équilibre entre les différents types d'activités et les lieux résidentiels en particulier [Ewing and Rong, 2008]. L'objectif de cet indicateur a été de visualiser sur un territoire donné le niveau de co-occurrence des usages résidentiels et ceux pour des activités (commerces, activités de loisir etc) sur un rayon prédéfini. Un prétraitement sur les données relatives aux différents usages est réalisé en estimant une fonction de densité de probabilité pour les deux types d'usage. Ceci a été effectué par une estimation à noyaux (Gaussien) de la densité (Kernel Density Estimation ou KDE) qui permet de lisser la distribution spatiale des différents usages et d'interpoler les données existantes sur les bâtiments non renseignés. L'hypothèse de normalité de la distribution spatiale des deux types d'usage a été faite au préalable. La fonction utilisée KDE est définie ainsi :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\|x - x_i\|}{h}\right) \quad (1.1)$$

Avec  $f$  comme la fonction de densité de probabilité estimée par type d'usage,  $x_i$  correspondant aux coordonnées latitude longitude par point d'une grille régulière avec un pas de 100 m dans le cadre de ce travail,  $h$  étant le paramètre de lissage appelé aussi fenêtre, choisi en fonction de la distance à pied dans un certain voisinage (ici distance moyenne parcourue à pied pour accéder à un voisinage urbain),  $n$  représentant le nombre d'observations et le noyau  $K$  est ici la fonction de densité de la loi normale à espérance nulle et d'écart-type égal à 1. Ainsi le noyau s'écrit :

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (1.2)$$

Après ce processus, l'entropie de Shannon a été utilisée pour définir l'indice de mixité d'usage du sol. En théorie de l'information, l'entropie de Shannon introduite

en 1948 par Claude Shannon, est une mesure moyenne de la quantité d'information (en Bit) nécessaire à la transmission d'un message contenant des symboles identiquement et indépendamment distribués. Elle est définie dans le cas discret comme suit :

$$H(x) = - \sum_{i=1}^n p_i \log_2 p_i \quad (1.3)$$

où  $n$ , est le nombre de type d'usage du sol possible et  $p_i$  la probabilité à un point  $i$  d'occurrence d'un type d'usage.  $H$  s'approche de 0 si la mixité d'usage est faible, et de  $\log_2(n = 2)$  si l'occurrence conjointe des deux types d'usages se rapproche d'une distribution uniforme dans l'espace. La visualisation de la mixité d'usage du sol est ensuite effectuée sous forme de carte de chaleur, avec en abscisse et en ordonnée les coordonnées géographiques des différents points et une variation de la couleur par point du maillage régulier en fonction du niveau de l'entropie obtenu dans un rayon donné.

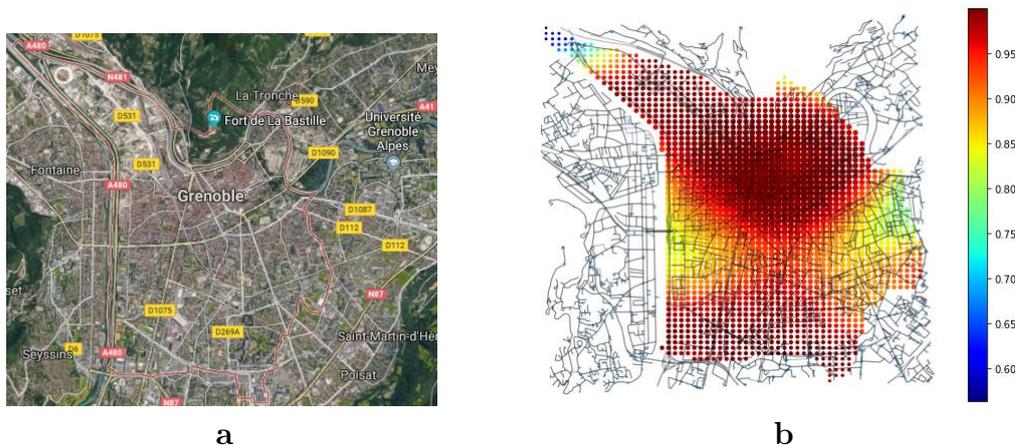


FIGURE 1.1 – (a) Image satellite de Grenoble, (b) mixité d'usage du sol

Le niveau minimal et maximal de mixité d'usage du sol sont respectivement représentés par les couleurs bleu foncé et rouge foncé. Nous pouvons voir que le niveau de mixité est très élevé au niveau du quartier hypercentre de la ville et plus faible sur la presqu'île de Grenoble appelée également polygone scientifique et qui concentre une faible partie de la population et un nombre important de centres de

recherche scientifique.

### 1.3.2 Accessibilité des activités

L'étalement urbain est souvent associé à une dépendance accrue à l'automobile due à une moindre accessibilité des opportunités professionnelles dans les zones où il est observé comparées aux zones urbaines plus compactes [Ewing and Hamidi, 2015]. La question à laquelle l'indice d'accessibilité construit par [Gervasoni et al., 2017] tente de répondre est : combien d'activités relatives aux services nécessaires à la vie quotidienne des habitants sont disponibles dans un rayon donné ? Autrement dit, outre la possibilité offerte aux habitants par la mixité d'usage du sol d'accéder à différents services urbains à pied, jusqu'où devront-ils aller en moyenne pour pouvoir bénéficier des différents services urbains ?

Dans un premier temps, le réseau routier de la zone d'intérêt est extrait d'OSM, puis divisé en tronçons dont la longueur maximale est de 150 mètres. Ensuite chaque activité est associée au plus proche nœud d'un tronçon en fixant une distance maximale de liaison à 250 m pour éviter l'inclusion de surfaces naturellement inaccessibles dans le calcul (forêts, parcs, zones montagneuses etc.). La dernière étape consiste ensuite à cumuler pour chaque point d'une grille régulière le nombre d'activités associées aux nœuds situés dans un rayon de 1 km. La visualisation se fait selon le même principe que la mixité d'usage du sol.

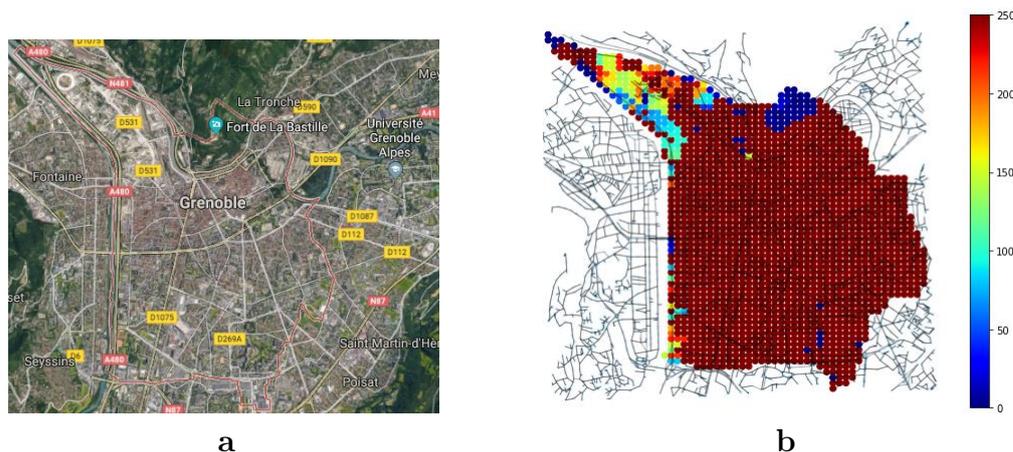


FIGURE 1.2 – (a) Image satellite de Grenoble, (b) accessibilité aux activités

Nous constatons que le niveau d'accessibilité par le réseau routier est au maximum dans une majeure partie du territoire à l'exception d'une partie de la presqu'île qui est coupée partiellement du reste du territoire par l'Isère et en allant vers le fort de la Bastille. Cette zone est caractérisée une forte pente du terrain et une faible présence d'activités.

### 1.3.3 Dispersion du bâti

v

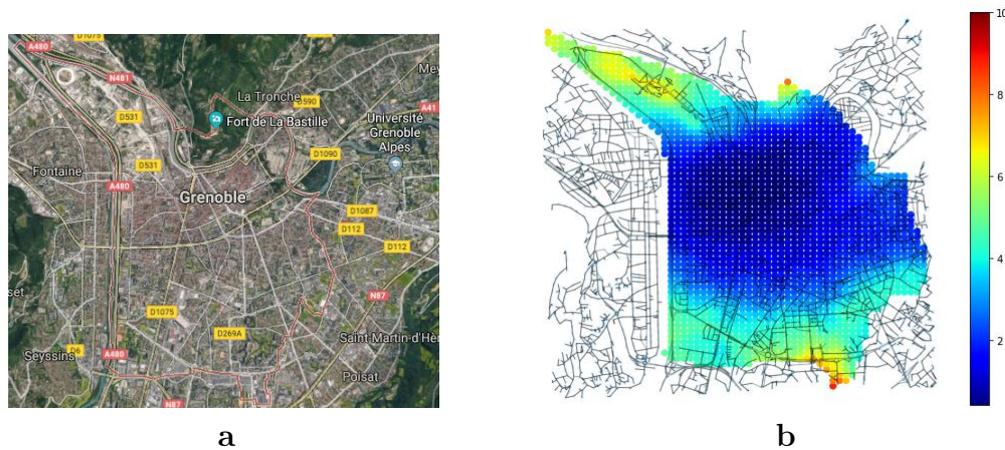


FIGURE 1.3 – (a) Image satellite de Grenoble, (b) dispersion du bâti

Nous observons une forte concentration du bâti au niveau du centre ville et une dispersion plus importante au nord ouest et au sud de la commune.

Les causes du phénomène de l'étalement urbain ont fait l'objet d'une littérature abondante comme on peut le voir dans la section suivante.

## 1.4 Causes de l'étalement urbain

L'étalement urbain peut être causé par une multitude de facteurs. Selon l'Agence européenne de l'environnement, les sources de l'étalement urbain peuvent être classées en cinq facteurs : démographiques, socioéconomiques, politiques, géomorphologiques et technologiques [Hennig et al., 2016]. En général, l'étalement urbain

se détermine par la combinaison de forces attractives issues de la périphérie des aires urbaines et de forces répulsives présentes au centre de ces aires. Nous présentons ci-après une typologie de sources correspondant aux travaux de littérature étudiés dans le cadre de ce travail.

#### **1.4.1 Sources démographiques**

Selon [Gottmann and Harper, 1990] l'explosion en surface des métropoles modernes a été accompagnée par un accroissement de leur population. Cette augmentation de la population peut être de source naturelle ou due à la migration (interne ou internationale) vers les zones urbaines. En exerçant une pression sur le marché immobilier, l'augmentation de la population entraîne une hausse des prix dans le secteur résidentiel [Christiansen and Loftsgarden, 2011]. [Combes et al., 2012] trouvent également une relation positive entre les coûts urbains (part du logement dans les dépenses des habitants et élasticité des prix des logements en centre ville) et la taille de la population. Ce qui oblige les ménages les plus modestes à se localiser à l'extérieur des centres urbains [Chin, 2002], où la majeure partie de population est souvent concentrée. De même lorsque cette augmentation de population s'accompagne de celle des nuisances urbaines telles que les embouteillages, une mauvaise qualité de l'air, les nuisances sonores et l'insécurité, elle peut amener une partie des habitants à se localiser à l'extérieur des zones urbaines. Cependant, d'après l'Agence européenne pour l'environnement, cet effet devrait être faible en Europe car il y est prévu un déclin de la population d'ici la fin du siècle [Hennig et al., 2016]. De même, les choix résidentiels des individus sont influencés par leurs situations professionnelles et les structures des ménages auxquels ils appartiennent [Da Cunha, 2005].

#### **1.4.2 Préférences individuelles**

Selon [Pumain, 2003], un rejet des habitats collectifs dans les années 1960 par les classes moyennes et le changement structurel des familles ont entraîné les premières vagues de construction de lotissements. Dans une grande partie des pays développés, les habitants des centres urbains observent une insuffisance d'espace

habitable ([Bhatta, 2010]). On observe une tendance des jeunes couples avec enfant à déménager des centres urbains vers des zones plus rurales pour accéder à des logements plus vastes (par [Chin, 2002] ; [Christiansen and Loftsgarden, 2011]) etc. Cette demande en supplément d'espace indique un étalement urbain uniquement lorsqu'il en résulte un aménagement moins dense dans la marge de la zone urbaine ([Bhatta, 2010]). Sur le plan quantitatif, [Kiakou, 2017] trouve en estimant la demande en logements situés à la périphérie de la ville de Larissa en Grèce une influence positive et significative des espaces ouverts sur celle-ci. Les attributs sociaux et physiques du lieu résidentiel (image, environnement naturel, voisins, etc.) qui lorsqu'ils favorisent les zones périphériques, peuvent ainsi entraîner un mouvement de la population vers ces dernières ([Da Cunha, 2005]) suivie par une partie des équipements et enseignes commerciales ([Bochet, 2005]) qui à leur tour rendent ces zones encore plus attractives. Les développements en infrastructures routières et les progrès technologiques ont permis une baisse des coûts de transport qui à été en faveur de la dispersion spatiale des fonctions urbaines.

### 1.4.3 Baisse des coûts de transports

Les transports ont toujours joué un rôle majeur dans la forme des villes ([Glaeser and Kahn, 2004]). Les avancées technologiques et les baisses de coûts de transport ont permis aux industries, commerces et ménages de gagner en flexibilité en matière de localisation géographique et une possibilité d'intensifier les échanges ([Bochet, 2005]) grâce la possibilité d'effectuer des distances plus longues avec le même temps de trajet ([Christiansen and Loftsgarden, 2011]). Selon [Frumkin, 2002], il existe un lien étroit entre la faible densité du bâti et l'utilisation de l'automobile comme moyen de transport. Ceci est soutenu par le fait que contrairement aux transports publics qui nécessitent une densité minimale pour pouvoir être déployés, la voiture individuelle n'impose aucune contrainte aux habitants en ce sens ([Glaeser and Kahn, 2004]).

#### 1.4.4 Imperfections du marché

Les travaux de [Brueckner, 2000] reviennent très souvent dans la littérature pour exprimer la partie de l'étalement urbain imputable à l'imperfection du marché quant à la détermination du prix des terres en périphérie des aires urbaines. Le modèle postule que les résidents d'une localité se déplacent vers le centre des affaires avec un coût  $t \cdot x$  ( $t$  : temps,  $x$  : distance) où ils obtiennent un revenu  $y$ . Le revenu net,  $y - t \cdot x$ , est dépensé en logement ( $q$ ) et en autres biens ( $c$ ). Pour compenser la distance du trajet au travail, le prix foncier ( $r$ ) baisse à mesure que l'on s'éloigne du quartier d'affaires. Cette relation pousse les résidents à se localiser de plus en plus loin des centres urbains pour habiter dans des logements plus larges à un prix plus faible. Trois catégories de défaillances du marché ont été soulevées par l'auteur. Celles-ci sont la non prise en compte de la valeur des espaces ouverts lors de la conversion en surface urbanisée, la non intégration par les résidents des coûts liés à la congestion engendrée par leur usage des routes et l'absence de comptabilisation par les entreprises immobilières dans leur prix de vente de l'ensemble des coûts en infrastructures publics générés par leur offre. Ainsi, [Sainteny, 2008] qualifie cet effet du coût foncier sur l'étalement urbain comme étant pervers car son niveau global est parfois supérieur en périphérie par rapport à un logement situé au centre d'une agglomération urbaine. Ce dernier soulève une autre imperfection qui est le fait que l'accès aux prêts immobiliers ne soit soumis qu'à des conditions de ressources et non aux choix de localisation. Une tarification sous optimale au sein du marché immobilier est donc une cause essentielle du problème de l'étalement. Bien que l'intervention des autorités publiques soit souvent justifiées par la défaillance du marché, elle peut aussi être à la source de l'étalement urbain comme on peut le voir dans la section suivante.

#### 1.4.5 Politiques publiques

Les politiques publiques ont une influence importante sur l'aménagement urbain à travers leurs impacts, entre autres, sur les coûts de transports, les prix immobiliers, le crédit immobilier, les taxes et subventions ([Hennig et al., 2016]). Cette influence se ramène soit à leur non utilisation pour corriger les imperfections du marché ou à

leur inefficacité. Selon certains auteurs, il existe en France une coordination insuffisante des instruments financiers et fiscaux, d'une part, et les politiques en matière d'urbanisme, d'autre part. Dans beaucoup de cas, les taxes sont indépendantes de la localisation et sont même souvent plus faibles en périphérie (Bhatta, 2010), ce qui encourage les ménages à s'y installer. Les politiques de limitation des surfaces urbanisées appliquées partiellement sur un territoire peuvent conduire à un aménagement discontinu de l'espace à cause de la volonté de les contourner ([Ehrlich et al., 2017]). Elles peuvent également conduire à une hausse des coûts fonciers ([Zaninetti, 2014]) avec l'exemple des politiques des périmètres d'urbanisation initiées dans les années 1970 à Portland, aux États-Unis ([O'Toole, 2001]). Cette hausse de coûts peut entraîner un étalement urbain à l'extérieur de ces périmètres comme pour le cas du comté de Clark appartenant à la région métropolitaine d'Oregon-Washington dans les années 1990 ([Jun, 2004]). En Espagne, entre 1960 et 2011, on constate un effet positif de la construction d'autoroutes sur la périurbanisation d'après les travaux de [Garcia-López and Muñiz, 2013]. La même conclusion a été faite par [Baum-Snow et al., 2016] dans le cas de la Chine avec une distinction des effets décentralisateurs des différents types d'infrastructures de transport sur les localisations des activités industrielles et de service.

#### 1.4.6 La topographie

Plusieurs aspects relatifs à la topographie ont été désignés comme responsables de la structure de l'aménagement du territoire ([Hennig et al., 2016]). En analysant les déterminants de la demande de logement, [Saiz, 2010] trouve que dans la plupart des lieux où celle-ci était inélastique la structure topographique définie par la raideur de la pente du terrain était contraignante. [Burchfield et al., 2006] concluent également une relation positive entre leur mesure de l'étalement urbain et la rugosité du terrain et l'absence de montagnes hautes. La présence de gisements de ressources naturelles exploitables peut également créer de l'étalement urbain ([Hennig et al., 2016]). A titre d'exemple, le processus d'urbanisation de la région Xinjiang dans l'ouest de la Chine est fortement lié à l'exploitation des ressources des sous-sols ([Dong and Yang, 2014]). Ainsi la composition du sol ou la topographie à travers leur effets sur l'attractivité des territoires peuvent être des

sources non négligeables de l'étalement urbain.

Par le biais de mécanismes complexes, l'étalement urbain a conséquences socioéconomiques et environnementales qui ont largement été couvertes par la littérature.

## **1.5 Conséquences de l'étalement urbain**

La littérature sur les conséquences de l'étalement est très abondante et couvre ses impacts économiques, sociaux et environnementaux. Bien que ces trois impacts ne soient pas dissociables, nous allons nous focaliser sur les impacts environnementaux et en particulier sur la consommation énergétique, sur les émissions de polluants et sur la qualité de l'air. Nous allons également évoquer brièvement les impacts sur la santé. Ce choix est guidé par les contraintes de la mission dans le cadre duquel ce travail s'inscrit.

### **1.5.1 La consommation énergétique**

Beaucoup d'études empiriques durant les trois dernières décennies ont tenté d'analyser les relations entre la forme urbaine et les modes de déplacement en analysant les comportements des ménages ou les caractéristiques globales de la population dans un certain voisinage. [Shammin et al., 2010] ont dans ce sens estimé les intensités en énergies des dépenses de 2982 ménages aux États-Unis répartis dans des villes à degrés de compacité variables. Ainsi, ils ont obtenu que les ménages des villes étalées présentent une intensité énergétique de 17% à 19% supérieure par rapport aux ménages de villes plus compactes. Pour avoir une vision plus détaillée de cette relation, nous présentons quelques travaux sur la question en distinguant la consommation d'énergie due au transport et celle liée au secteur résidentiel.

#### **1.5.1.1 Le secteur des transports**

L'étude des liens entre la forme urbaine et la consommation énergétique dans le secteur des transports par [Newman and Kenworthy, 1989] comme on peut le voir sur la figure 1.5 suggère une relation négative entre le nombre d'habitants par

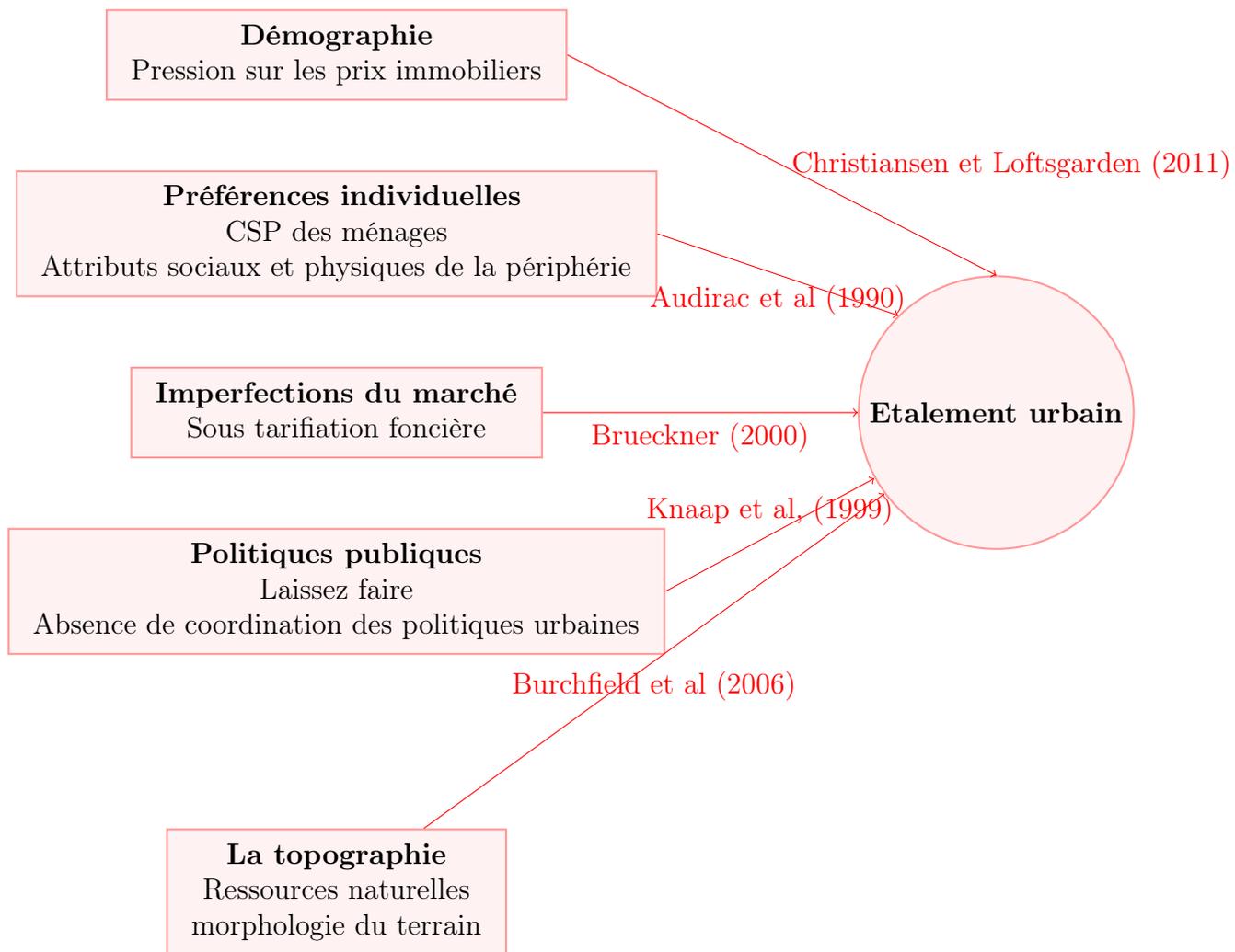


FIGURE 1.4 – Diverses causes de l'étalement urbain (Source : auteur)

hectare et la consommation de pétrole par tête. En analysant la performance énergétique en matière de transport de personnes dans trois villes périurbaines de la Wallonie en Belgique, [Marique and Reiter, 2012] concluent que la distance entre les unités résidentielles et les autres fonctions urbaines jouent un rôle majeur dans la consommation énergétique. Plus la mixité d'usage du sol est faible et plus la distance entre le domicile et les activités est grande, plus la consommation énergétique en transport sera élevée. [da Silva et al., 2007] aboutissent à des résultats similaires en étudiant la relation entre la consommation d'énergie du secteur des transports et la morphologie urbaine tout en contrôlant leurs caractéristiques socioéconomiques. La densité par kilomètre carré et le rapport entre les distances Nord-Sud et Est-Ouest les plus longues ont montré respectivement des effets négatifs et positifs sur la quantité de carburant consommée à des fins de transport. Selon [Bhatta, 2010], bien que la consommation automobile d'énergie ne soit pas aussi efficiente en zone dense qu'en zone étalée à cause des embouteillages dans la première, sa quantité par habitant y est plus faible car on y conduit moins. Contrairement aux résultats précédents, une relation positive entre le degré de concentration de la population et la consommation annuelle d'essence par habitant a été trouvée par [Shim et al., 2006] dans leur étude de l'efficacité énergétique des transports sur 71 villes coréennes.

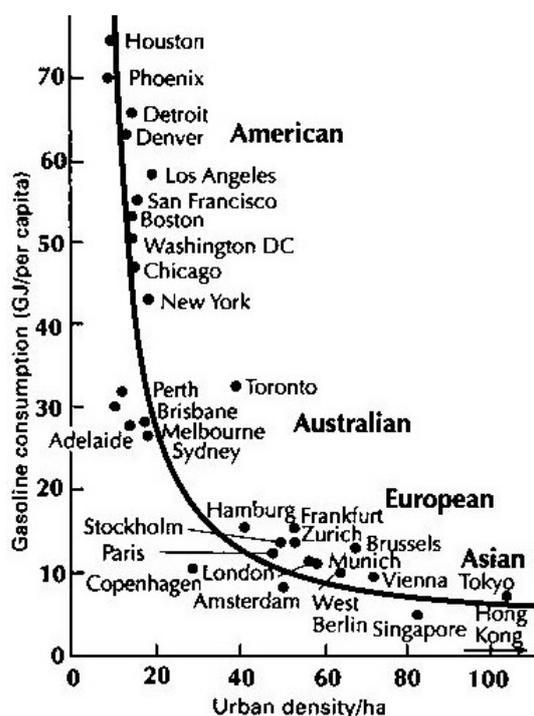


FIGURE 1.5 – Relation entre la densité de population (en abscisse) et la consommation en pétrole par habitant(en ordonnée). Source : Newman et Kentworthy (1989)

### 1.5.1.2 Le secteur résidentiel

En 2015, le secteur résidentiel représente 30,2% de la consommation totale d'énergie en France ([Cavaud et al., 2016]). [Ewing and Rong, 2008] lient la consommation énergétique résidentielle à la forme urbaine par trois canaux : le transport et la distribution d'électricité, le parc de logement et via la formation d'îlots de chaleur urbains. Pour des raisons de disponibilité de données et de faible contribution de la forme urbaine sur la consommation d'énergie due au transport et à la distribution d'électricité (7%), les auteurs ne s'intéressent qu'aux deux autres champs. Le premier effet est qu'indirectement, les types de choix résidentiels généralement associés à l'étalement urbain tels que des logements de tailles plus grandes qu'en milieu urbain dense entraînent une augmentation de la consommation d'énergie nécessaire pour contrôler la température des pièces plus larges. Le second effet de

la forme urbaine est relatif aux îlots de chaleur dus à une absorption des radiations solaires associées aux surfaces noires (goudrons, constructions vitrées etc) et aux activités humaines. D'après [Tabeaud, 2010], les températures minimales sont plus élevées dans les centres urbains qu'en zone rurale. Cet écart périphérie-centre est en moyenne entre 1 et 2°C et peut aller jusqu'à 10°C lorsque le temps est ensoleillé et qu'il y a absence du vent [Tabeaud, 2010]. Ainsi, la demande en climatisation est plus élevée en été dans les milieux urbains plus denses, et celle en chauffage plus faible en hiver comparé aux zones périphériques moins denses [Ewing and Rong, 2008].

Ces deux types de consommations d'énergie ont un impact direct sur l'émission de divers polluants dans l'air dont l'importance dépend de plusieurs paramètres.

### 1.5.2 L'émission de polluants et dégradation de la qualité de l'air

Les polluants liés à la morphologie urbaine considérés dans notre analyse sont uniquement ceux d'origine énergétique c'est-à-dire émis lors de la combustion d'énergie fossile. Leur quantité est souvent directement proportionnelle aux consommations d'énergie (DGEC, 2012) et donc indirectement à l'étalement urbain comme nous l'avons vu précédemment. Bien que l'industrie automobile ait observé d'importants progrès en matière d'émission de polluants, le nombre important de kilomètres parcourus en voiture entraîne des rejets importants de monoxyde de carbone (CO), de dioxyde de carbone (CO<sub>2</sub>), d'oxydes d'azote (NO), de particules fines [Frumkin, 2002], de composés organiques volatiles (VOC pour Volatile Organic Compound) etc. Ces polluants dits primaires ou précurseurs, par le biais de réactions chimiques complexes avec les radiations solaires et la chaleur, conduisent à la formation d'ozone et de particules ainsi appelés polluants secondaires [Catherine Vakil, 2005]. Par leur capacité de se loger dans les alvéoles pulmonaires, les particules ultrafines ( $<2.5\mu m$ ) sont sources de hausse de la mortalité liée aux maladies cardiovasculaires [Brunekreef and Holgate, 2002].

De nombreuses études ont tenté d'évaluer les effets de la morphologie urbaine sur la qualité de l'air à différents niveaux géographiques et ont abouti à des résultats contrastés. [Stone Jr, 2008], utilise un ensemble d'indicateurs synthétiques d'étalement urbain développé par [Ewing et al., 2003] comme variables explicatives de la

qualité de l'air qui est mesurée par le nombre de jours de dépassement de la valeur limite de l'ozone sur 8 heures. Pour un échantillon composé des 45 plus grandes zones statistiques métropolitaines de États-Unis, il conclut à un effet négatif de l'étalement urbain sur la qualité de l'air. De façon contradictoire, [Chen et al., 2008] trouvent pour les 45 plus grandes villes de la Chine une relation positive entre la compacité urbaine mesurée par la densité de population et le niveau de concentration de dioxyde soufre par mètre carré de surface bâtie. Encore plus surprenante, l'étude de la relation forme urbaine et qualité de l'air menée par [Cho and Choi, 2014] sur des données en panel (combinaison de données individuelles et temporelles) de 17 villes coréennes a montré des effets nuancés selon les différents types de polluants considérés. Notons que le seul effet à y avoir été significatif est celui de la densité nette de surface bâtie sur la concentration d'oxydes de carbone avec un signe positif. [Mansfield et al., 2015] ont développé un modèle prédictif des concentrations de particules fines de diamètres inférieures à  $2.5\mu\text{m}$  (PM : Particulate Matter en anglais) au niveau régional et local en simulant deux scénarios de développement urbain (compact et étalé) par une redistribution de niveaux différents de densité aux villes de leur zone d'étude situées dans la région de Caroline du Nord aux États-Unis. Ils obtiennent ainsi pour le scénario étalé une baisse des concentrations de ces particules au niveau des stations locales mais une augmentation en terme de niveau moyen régional. Un résultat inverse est obtenu pour le scénario compact avec une hausse des concentrations locales de PM 2.5 mais un niveau régional inchangé comparé au scénario de base. Ce résultat laisse croire que la relation entre forme urbaine et qualité de l'air est fortement liée au niveau géographique auquel elle est étudiée et que l'effet de la dépendance automobile sur la qualité de l'air s'observe plutôt au niveau régional.

### 1.5.3 Les effets de l'étalement urbain sur la santé

Par l'intermédiaire de la génération de pollution due à une forte dépendance à l'automobile, l'étalement urbain a également été désigné comme l'un des responsables de l'accroissement des cas de maladies respiratoires. En effet, par leur capacité de se loger dans les alvéoles pulmonaires, les particules ultrafines ( $<2.5\mu\text{m}$ ) sont sources de hausse de la mortalité liée aux maladies cardiovasculaires [Brunekreef

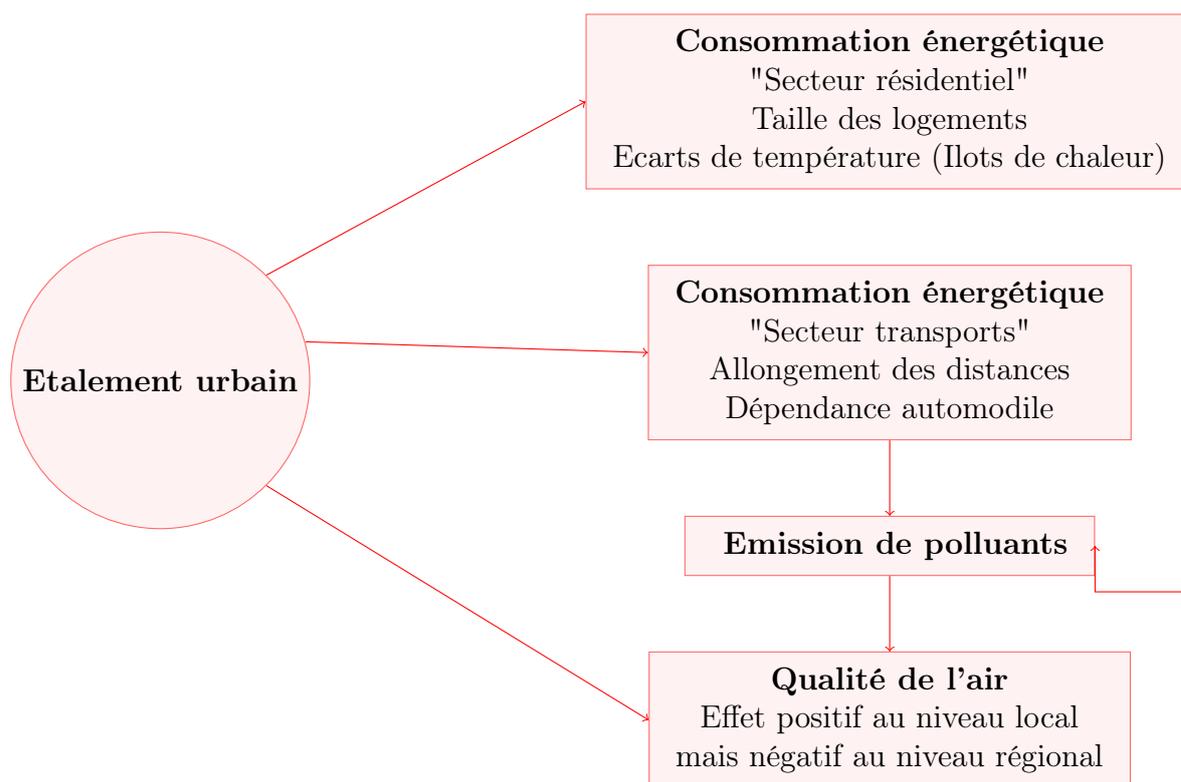


FIGURE 1.6 – Conséquences de l'étalement urbain (Source : auteur)

and Holgate, 2002]. L'étalement urbain étant associé à une faible densité des activités, la faible accessibilité de ces dernières liée à une mauvaise connectivité des réseaux de transports urbains et à une forte dépendance à l'automobile, elle induit potentiellement à une réduction importante de l'activité physique des habitants des zones urbaines concernées et donc favorise la multiplication des cas d'obésité chez ces derniers ([Frumkin, 2002] ; [Garfinkel-Castro et al., 2017]). [Lopez, 2004] à travers une analyse multiniveau (individuel et métropolitain) conclut également à un lien positif entre l'étalement urbain et l'obésité.

## 2 Méthodologie

On a pu observer que l'étalement urbain est un concept complexe et multidimensionnel, tant au niveau de sa mesure, de ses déterminants qu'à celui de ses

conséquences. L'objectif de notre travail consiste à analyser l'étalement urbain par une approche intégrée, en tentant d'estimer d'une part les relations entre la morphologie urbaine et ses déterminants, et d'autre part les liens de cette dernière avec l'émission de polluants et la qualité de l'air. Du fait qu'aucun de ces concepts ne soit directement observable, un ensemble de variables est choisi pour représenter chacun d'entre eux. Notre ambition initiale était d'élargir l'étude sur tout le territoire national, mais la non disponibilité de données homogènes entre régions nous a contraint à nous limiter à la région Auvergne-Rhône-Alpes. L'autre raison est que la quantité de données ouvertes est largement supérieure en comparaison aux autres régions françaises.

## 2.1 Choix du modèle

## 2.2 Quelques mots sur les modèles d'équations structurelles

L'approche PLS path modeling (PLS-PM) pouvant être traduite par modélisation de piste causale par moindres carrés partiels a été introduite par [Wold, 1982] en tant que modèle d'équation structurelle (Structural equation modeling ou SEM en anglais). Comme son nom l'indique, il est basé sur les moindres carrés partiels en opposition au modèle de relations structurelles linéaires (Linear structural relationships ou LISREL) développé par [Jöreskog, 1970], qui lui est basé sur l'estimation des covariances par maximum de vraisemblance [Jakobowicz, 2007]. Les modèles d'équations structurelles ont pour but d'estimer le réseau de relations causales défini selon la théorie, en mettant en relation un plusieurs concepts latents (phénomènes non observables directement) censés représenter la complexité au sein d'un système et mesurés chacun par un ensemble de facteurs observables appelés communément variables manifestes. Ces modèles se situent entre l'analyse de piste causale et l'analyse factorielle confirmatoire initiées respectivement par Tukey (1964) et [Thurstone, 1931] d'après [Vinzi et al., 2010b]. Selon H. Wold, l'approche PLS-PM peut-être qualifiée de modélisation douce car elle nécessite peu d'observations et d'hypothèses sur les distributions et l'approche par maximum de vraisemblance de modélisation dure du fait de ses hypothèses lourdes sur la distribution des données et qu'elle requiert des centaines d'observations [Tenenhaus

et al., 2005]. Nous ne nous focalisons dans notre étude que sur l'approche PLS-PM pour cette raison principalement et du fait que les données que nous utiliserons sont de sources très variées et sont régies très probablement par des lois statistiques diverses. Elle peut être utilisée pour confirmer la théorie ou tester des hypothèses de relations non explorées ([Chin, 1998]).

## **2.3 Bref historique de la méthode PLS path modeling**

La méthode PLS est le fruit des recherches de [Wold, 1966] sur l'analyse en composantes principales avec plus d'un ensemble de variables ou bloc, qui ont débouché sur l'algorithme NILES (Nonlinear Iterative Least Squares). Avec son extension à trois blocs ou plus, celui-ci est rebaptisé NIPALS (Nonlinear Iterative Partial Least Squares) dans les articles de Wold de 1973 et 1975 ([Tenenhaus et al., 2005], [Jakobowicz, 2007]). L'application de l'approche PLS à la modélisation de piste causale avec variables latentes est apparue dans les travaux de Wold à la fin des années 1970 ([Tenenhaus et al., 2005]). Cette méthode est présente dans plusieurs domaines d'application, notamment dans plusieurs sous disciplines des sciences sociales et en chimiométrie.

## **2.4 Principes de l'approche PLS path modeling**

Cette approche est décrite par deux sous modèles : le modèle externe et le modèle interne.

### **2.4.1 Le modèle externe ou de mesure**

Celui-ci relie une variable latente aux variables manifestes qui lui sont associées. Trois schémas d'association sont possibles à savoir réflexif, formatif et MIMIC (Multiple effect Indicators for Multiple Causes) qui combine les deux premiers.

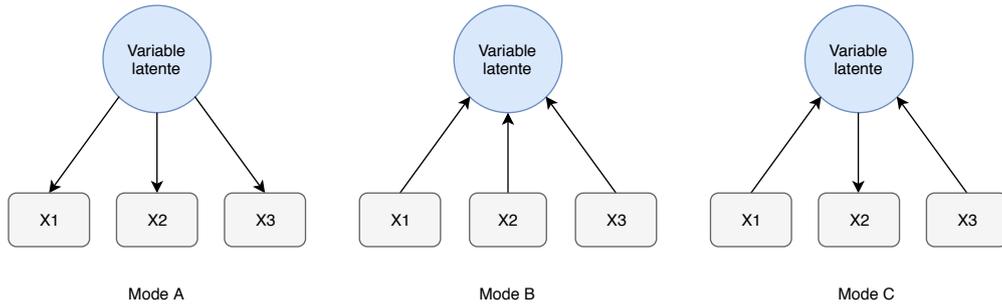


FIGURE 2.1 – Trois blocs de variables sous différents schémas (Source : auteur)

Dans le schéma réflectif ou mode A chaque variable manifeste d'un même bloc est expliquée par une seule variable latente. Ceci est modélisé par l'équation de régression simple suivante :

$$x_{ik} = \lambda_{i0} + \lambda_{ik}\xi_k + \epsilon_{ik} \quad (2.1)$$

avec  $\lambda_{i0}$  la constante du modèle,  $x_{ik}$  une variable manifeste  $i$  d'un bloc  $k$ ,  $\lambda_{ik}$  le coefficient de régression associé à la variable latente  $\xi_k$  standardisée et  $\epsilon_{ik}$  le terme d'erreur. L'hypothèse derrière ce modèle est que  $\epsilon_{ik}$  est de moyenne nulle et non corrélé avec la variable latente associée à son bloc d'appartenance et l'équation précédente devient donc :

$$E(x_{ik}|\xi_k) = \lambda_{i0} + \lambda_{ik}\xi_k \quad (2.2)$$

Puisque les variables manifestes sont standardisées donc de variance unitaire cela revient à estimer la covariance entre  $\xi_k$  et  $x_{ik}$ .

Inversement, dans le schéma formatif ou mode B chaque variable latente est expliquée par les variables manifestes du bloc qu'elle représente. La relation s'écrit dans ce cas :

$$\xi_k = \sum_i w_{ik}x_{ik} + \delta_{ik} \quad (2.3)$$

avec  $w_{ik}$  les coefficients associant chaque variable manifestes  $i$  à la variable latente de son bloc  $k$  et  $\delta_{ik}$  le terme d'erreur de l'équation qui doit ici également satisfaire

comme contraintes une espérance et une corrélation avec les variables manifeste toutes deux égales à 0. Cependant la non linéarité de la relation entre la variable latente et les variables manifestes peut être prise en compte en utilisant la régression PLS ([Vinzi et al., 2010b]).

Le schéma MIMIC ou mode C, est un mélange entre les deux schémas précédents dans un même bloc de variables manifestes. Si on considère que les  $i_1$  variables manifestes d'un bloc sont définies selon un schéma réflectif et les  $i - i_1$  autres selon un schéma formatif on obtient alors :

$$x_{ik} = \lambda_{i0} + \lambda_{ik}\xi_k + \epsilon_{ik} \quad (2.4)$$

pour  $i=1$  à  $i_1$ , où la variable latente est définie par :

$$\xi_k = \sum_i w_{ik}x_{ik} + \delta_{ik} \quad (2.5)$$

Les mêmes contraintes pour les deux schémas précédents s'appliquent ici également.

En 2010 [Vinzi et al., 2010a] ont présenté les modes PLScore et PLScow qui sont orientés respectivement vers une maximisation des corrélations entre variables latentes et celle de la covariance entre elles. Le mode PLScore est un compromis entre le mode B et le mode A. Lorsque les variables manifestes sont standardisées, la première composante du bloc de variables  $X$  est obtenue de la manière suivante :

$$t_{1i} = \frac{1}{\sqrt{\sum_k cor^2(z_k, x_{ki})}} \sum_k cor(z_k, x_{ki}) \quad (2.6)$$

Le nombre de composantes dépend du schéma duquel on veut se rapprocher entre le mode A et B. Dans le premier, une seule composante est calculée alors que dans le second autant de composantes que de variables sont utilisées. Ces composantes ont l'avantage de ne pas être corrélées entre elles et donc adaptées en cas de multicolinéarité dans cadre de la régression multiple (mode B). Le mode PLScow suit le même principe avec comme différence une non standardisation des variables par bloc. La corrélation dans la formule de la composante est remplacée ainsi par

la covariance.

### 2.4.2 Le modèle interne ou structurel

Il relie les variables latentes entre elles en régressant celles qui sont endogènes sur les variables exogènes qui leur sont adjacentes :

$$\xi_k = \beta_{k0} + \sum_i \beta_{ki} \xi_i + \zeta_k \quad (2.7)$$

Les contraintes sont :  $cor(\zeta_k, \epsilon_k) = 0$  et  $cor(\zeta_k, \xi_j) = 0, \forall j \neq k$

Le schéma 2.2 illustre la relation entre les deux modèles .

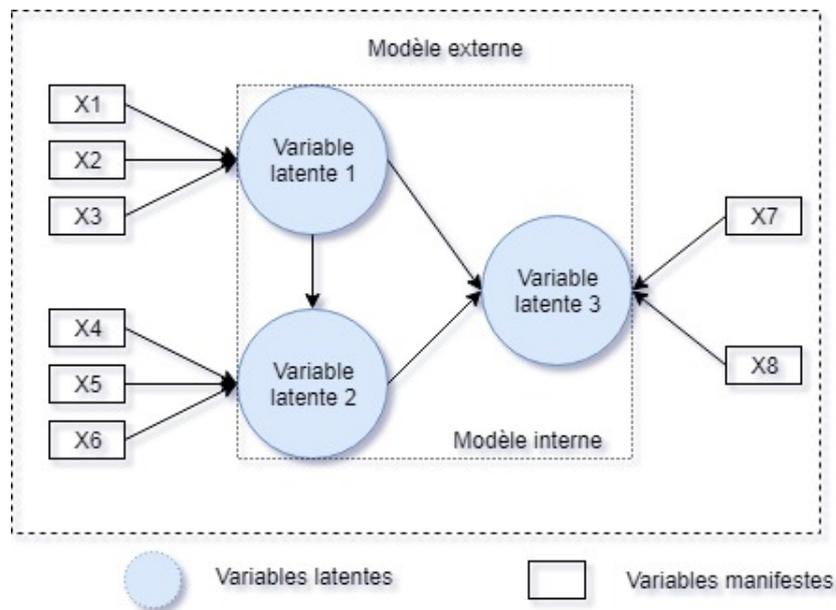


FIGURE 2.2 – Structure du modèle PLS path modelling (Source : auteur)

## 2.5 Fonctionnement de l’algorithme PLS-PM

L’algorithme PLS-PM est un procédé itératif d’estimation des variables latentes qui s’effectue en alternant des estimations par le modèle externe et par le modèle interne.

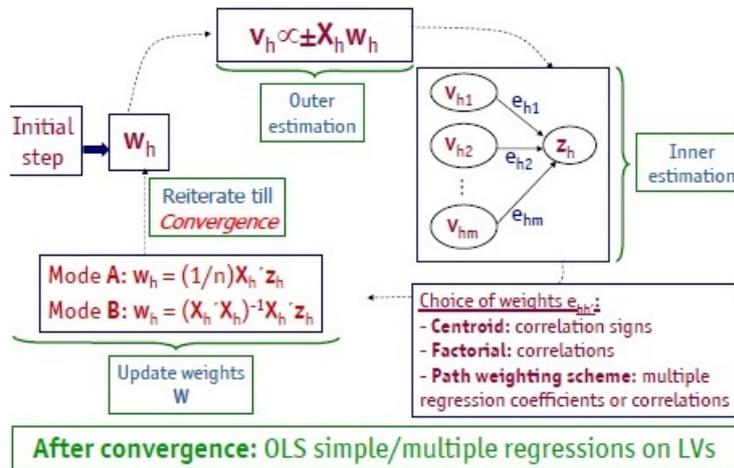


FIGURE 2.3 – Algorithm PLS path modelling (Esposito Vinzi, 2005 : 5th symposium on PLS and related methods)

### 2.5.1 Estimation des variables latentes

### 2.5.2 Estimation externe

D'abord un vecteur de poids  $w_{ki}$  des variables manifestes  $i$  d'un bloc  $k$  est choisi arbitrairement comme valeur d'initialisation (ex. 0,0,1...). Ils sont standardisés pour obtenir des variables latentes à variance unitaire ([Tenenhaus et al., 2005]).

Ensuite les variables latentes sont estimées en tant que combinaisons linéaires de leurs variables manifestes centrées (et réduites dans certains cas, eg. [Vinzi et al., 2010b]) :

$$y_k \propto \pm \left[ \sum w_{ki}(x_{ki} - \bar{x}_{ki}) \right] \quad (2.8)$$

Le terme  $\propto$  signifie que  $y_k$  est obtenu par la somme pondérée par les poids  $w_i$  des variables manifestes  $x_i$  standardisées du bloc  $k$ .

### 2.5.3 Estimation interne

- Le schéma centroïde

Il découle des travaux originaux de Wold sur l'approche PLS-PM. Il retient le signe de la corrélation entre  $y_k$  et la variable latente  $y_{k'}$  qui est reliée à elle :

$$e_{kk'} = \text{sign}[\text{cor}(y_k y_{k'})] \quad (2.9)$$

- Le schéma factoriel

Cette méthode est issue des travaux de Lohmöller et prend directement la valeur de la corrélation entre deux variables latentes adjacentes  $y_k$  et  $y_{k'}$  estimées dans le modèle externe comme poids interne :

$$e_{kk'} = \text{cor}(y_k, y_{k'}) \quad (2.10)$$

- Le schéma structurel ou pondération de piste

Il prend le coefficient de régression de  $y_k$  sur  $y_{k'}$  si la structure de la relation va de  $k'$  vers  $k$ . Si celle-ci va de  $k$  vers  $k'$  alors c'est le coefficient de corrélation qui est retenu :

$$e_{kk'} = \begin{cases} (Y_{k'} Y_{k'}')^{-1} Y_{k'} Y_k, & \text{si } \xi' \rightarrow \xi \\ \text{cor}(y_k, y_{k'}), & \text{dans le cas inverse} \end{cases}$$

### 2.5.4 Mise à jour des poids dans l'estimation externe

Les valeurs  $z_k$  obtenues par le modèle interne sont ensuite utilisées dans le modèle externe :

- Pour le mode A

Le poids  $w_{ki}$  est mis à jour en tant que coefficient de la régression simple suivante :

$$w_{ki} = \text{cov}(x_{ik}, z_k) \quad (2.11)$$

Si les variables manifestes sont standardisées alors l'équation se transforme en corrélation (Esposito Vinzi et al, 2009).

- Pour le mode B

Les poids  $w_k$  sont recalculés comme étant les coefficients de la régression multiple suivante :

$$w_k = \frac{1}{N} [(X'_k X_k)^{-1} X'_k Z_k] \quad (2.12)$$

- Pour le mode C

Le poids  $w_{ki}$  est mis à jour en tant que signe de la corrélation entre les variables manifestes est le poids issu du modèle interne :

$$w_{ki} = \text{sign}[\text{cor}(x_{ik}, z_k)] \quad (2.13)$$

### 2.5.5 Vérification de la convergence

Tant que  $w_{ki}$  varie, le processus reprend à partir de l'estimation externe. Lorsque la valeur de  $w_{ki}$  ne varie plus ou varie uniquement en dessous d'un seuil de tolérance prédéfini par l'utilisateur alors il y a convergence. Dans ce cas l'algorithme procède à la dernière phase.

#### 9.5.6 Estimation des relations dans les modèles internes et externes

Les valeurs des coefficients affectés aux relations entre variables latentes sont obtenues par moindres carrés ordinaires en régressant celles qui sont adjacentes entre elles. Les poids externes (loadings) des variables manifestes sont obtenus en calculant leurs corrélations avec le score final de la variable latente issue de leur bloc. (voir section sur le modèle a priori).

L'estimation du modèle PLS PM étant faite de manière récursive, les interactions bidirectionnelles entre variables latentes ne sont pas permises.

## 2.6 Évaluation du modèle

Le modèle PLS-PM est basé sur la variance et est donc orienté vers une prédiction des relations et ne dispose pas ainsi de critère d'ajustement global tel que le  $\chi^2$  pour le modèle LISREL mais plutôt d'un indice de qualité globale de la prédiction du modèle. Son évaluation se fait en trois étapes, d'abord celle du modèle externe

et ensuite celle du modèle structurel ([Henseler et al., 2009]; [Sanchez, 2013]) et enfin par l'indice de qualité de prédiction globale.

### 2.6.1 Évaluation du modèle externe

Pour le modèle externe, différents critères sont utilisés pour la validation en fonction du mode utilisé (A ou B). Pour le mode A, trois critères issus de la théorie des tests développés en psychométrie ([Sanchez, 2013]) sont utilisés : l'unidimensionnalité, la communauté et la validité discriminante.

L'unidimensionnalité

Du fait que dans le schéma réflectif comme nous l'avons montré dans le graphique présentant les différents modes, les variables manifestes sont le reflet d'une même variable latente, elles doivent être corrélées positivement entre elles et varier dans un même espace dimensionnel. Pour mesurer cela, trois indicateurs sont employés dans la littérature : l'alpha de Cronbach, le rho de Dillon-Goldstein et l'analyse en composante principale (critère de Kaiser).

- Pour le premier, il s'agit du rapport entre la somme des corrélations entre les variables manifestes  $i$  d'un bloc  $h$  et la variance totale de ce bloc obtenue en y rajoutant le nombre de variables manifestes dans ce bloc lorsqu'elles sont standardisées. Ce ratio est divisé par sa valeur maximale  $\frac{i-1}{i}$  lorsque  $Cor(x_h, x'_h)$  est égal à 1 ([Tenenhaus et al., 2005]). Il est alors exprimé ainsi :

$$\alpha = \frac{\sum_{h \neq h'} Cor(x_h, x'_h)}{i + \sum_{h \neq h'} Cor(x_h, x'_h)} * \frac{i}{i-1} \quad (2.14)$$

Un bloc est considéré comme unidimensionnel si l'alpha de Cronbach est supérieur à 0,7. Cependant selon [Jakobowicz, 2007], plusieurs travaux ont montré que cet indice ne mesure pas l'unidimensionnalité mais la consistance interne au sein d'un bloc de variables manifestes, et est donc à employer avec précaution.

- Le rho de Dillon-Goldstein est quand à lui calculé ainsi lorsque les variables manifestes sont standardisées :

$$\hat{\rho} = \frac{[\sum^i Cor(x_i, t1)]^2}{[\sum^i Cor(x_i, t1)]^2 + \sum^i [1 - Cor^2(x_i, t1)]} \quad (2.15)$$

Il mesure le poids de la corrélation entre les variables manifestes et la première composante principale obtenue à partir du bloc correspondant. Là également, une valeur supérieure à 0.7 désigne l'unidimensionnalité d'un bloc. Il est considéré plus fiable que l'alpha de Cronbach ([Tenenhaus et al., 2005]), notamment car il est lié directement à l'analyse en composante principale [Jakobowicz, 2007].

- Le critère de Kaiser est le troisième indicateur d'unidimensionnalité et consiste à vérifier que la première valeur propre de la matrice des corrélations d'un bloc de variables est supérieure à 1 et la seconde inférieure à 1. [Tenenhaus et al., 2005] suggèrent que lorsqu'une variable manifeste est négativement corrélée à la première composante principale d'un bloc, celle-ci doit en être exclue.

La communauté

Elle permet d'évaluer la qualité du modèle externe pour chaque variable en y mesurant la part de sa variabilité expliquée par la variable latente de son bloc d'appartenance. Il s'agit du carré de la corrélation entre une variable manifeste et la variable latente appelée loading. Une valeur supérieure à 0,7 de cet indicateur indique que l'indicatrice est suffisamment reproductible à partir de la variable latente. La communauté se calcule ainsi par :

$$Com_i = Cor^2(x_i, \xi_k) \quad (2.16)$$

La moyenne des communautés au sein d'un bloc fournit la communauté moyenne nommée encore variance moyenne extraite ou AVE (Average Variance Extracted). Celle-ci doit être supérieure à 0,5 ([Sanchez, 2013], [Ravand and Baghaei, 2016]).

La validité discriminante

Deux critères essentiels pour mesurer la validité discriminante au sein d'un bloc sous schéma réflexif sont les corrélations croisées (crossloadings) et le critère de

Fornell-Larcker. Le premier consiste à comparer les corrélations entre les variables manifestes d'un bloc et leur variable latente à celles avec les variables latentes des autres blocs. Les premières doivent être supérieures aux deuxièmes ou autrement on dit qu'il y a trahison de l'indicateur. Il peut être formalisé ainsi :  $Cor(x_{ik}, \xi_k) > Cor(x_{ik}, \xi_j), \forall k \neq j$ .

Le critère de Fornell-Larcker quant à lui impose que les AVEs soient supérieures au carré des corrélations entre variable latente d'un bloc et n'importe quelle autre variable latente. Cela peut se formuler ainsi :  $AVE_k > Cor(\xi_k, \xi_j), \forall k \neq j$ .

Pour le mode B, l'évaluation impose simplement que les coefficients de régression associés à chaque variable manifeste soient significatifs et qu'il n'y ait pas de multicollinéarité entre les variables au sein d'un bloc ([Fricker et al., 2012]).

### 2.6.2 Évaluation du modèle structurel

Pour le modèle structurel, les coefficients de régression entre variables latentes endogènes et exogènes doivent être significatifs. La proportion de la variance expliquée ( $R^2$ ) obtenue doit être au moins supérieure à 0,1 ([Fricker et al., 2012]). Deux autres critères sont vérifiés : la redondance et la qualité de l'ajustement.

La redondance

Elle mesure la part de variabilité d'une variable au sein d'un bloc qui est expliquée par une variable latente qui est directement connectée à ce bloc ([Vinzi et al., 2010b]). C'est le produit entre la communauté moyenne par bloc ou AVE et le coefficient de détermination linéaire entre variables latentes endogènes et exogènes. Elle se calcule de la manière suivante pour chaque bloc k :

$$Red_k = Com_k \times R^2(\xi_k, \xi_{k'}) \quad (2.17)$$

### 2.6.3 La qualité de l'ajustement global du modèle structurel

[Tenenhaus et al., 2005] ont proposé un indice de qualité globale du modèle interne noté communément GoF (Goodness of Fit). Il combine l'indice de communauté

moyen issu du modèle externe et la moyenne des  $R^2$  calculés à partir du modèle interne pour calculer cet indice de qualité globale qui est leur moyenne géométrique :

$$GoF = \sqrt{Com_k \times \bar{R}^2} \quad (2.18)$$

<b>Critères de validation du modèle PLS-PM</b>	
Modèle externe	Modèle interne
<b>Mode A</b>	
<p><b>1. Unidimensionnalité</b></p> <ul style="list-style-type: none"> <li>• <math>\alpha</math> de Cronbach (<math>\geq 0,7</math>)</li> <li>• <math>\rho</math> de Dillon-Goldstein (<math>\geq 0,7</math>)</li> <li>• Valeurs propres : <math>\lambda_1 &gt; 1</math> et <math>\lambda_2 &lt; 1</math></li> </ul> <p><b>2. Communauté</b></p> <ul style="list-style-type: none"> <li>• <math>Cor^2(x_{ik}, \xi_k) &gt; 0,7</math></li> <li>• <math>AVE &gt; 0,5</math></li> </ul> <p><b>3. Validité discriminante</b></p> <ul style="list-style-type: none"> <li>• Crossloadings : <math>Cor(x_i, \xi_i) &gt; Cor(x_i, \xi_j), i \neq j</math></li> <li>• Fornell-Lacker : <math>AVE &gt; Cor(\xi_k, \xi_{k'}), \forall k \neq k'</math></li> </ul>	<p><b>1. Coefficients de régression</b></p> <ul style="list-style-type: none"> <li>• Coefficients des <math>\xi_k</math> significatifs</li> </ul> <p><b>2. Coefficient de détermination</b></p> <ul style="list-style-type: none"> <li>• <math>R^2 &gt; 0.1</math></li> </ul> <p><b>3. Redondance</b></p> <ul style="list-style-type: none"> <li>• Valeur à maximiser</li> </ul>
<b>Mode B</b>	
<ul style="list-style-type: none"> <li>• Coefficients des <math>x_i</math> significatifs</li> <li>• Non multicolinéarité</li> </ul>	
<b>Qualité prédictive globale</b>	
<ul style="list-style-type: none"> <li>• <math>GoF \geq 0,7</math></li> </ul>	

TABLE 1 – Evaluation PLS path modelling (Source : auteur)

## 2.7 Présentation du modèle a priori

La construction de notre modèle est basée sur la revue de littérature effectuée plus haut. Le modèle structurel qui est celui liant les différents concepts latents permet de modéliser d'une part les relations entre la forme urbaine et ses précurseurs po-

tentiels, et d'autres part celle de cette morphologie urbaine avec les conséquences qui lui sont communément attribuées. Ces relations s'enchaînent de manière séquentielle de façon à prendre en compte les effets directs et indirects de chaque variable latente sur celle qu'elle est censée affecter. Le modèle de mesure ; quand à lui comme sont nom l'indique ; essaye de mesurer les concepts latents soit par l'utilisation de variables conformément aux travaux qui ont été abordés dans la revue littérature ; soit par celle de données jugées pertinentes par leur nature et leur relation étroite avec ces concepts. Les variables latentes que nous choisissons de mesurer sont donc contraintes par la disponibilité de données permettant de les mesurer de manière satisfaisante.

### **2.7.1 Modèle structurel**

Le schéma ci-dessous permet de lier les différentes relations en amont et en aval de la morphologie urbaine. Comme nous l'avons vu précédemment, l'étalement urbain est générée par la croissance démographique par l'intermédiaire d'une pression sur les prix immobilier qui pousse les ménages les plus modestes à se localiser en périphérie des villes. Il a été souligné par ailleurs que les couples avec enfants avaient tendance à préférer des logements plus spacieux, qui à leur tour étaient plus accessibles hors des centres-villes. On a également cités différents travaux mettant en avant une relation positive entre la pente du terrain et l'étalement urbain en forçant une discontinuité du bâti. Les autres causes de l'étalement urbain telles que la fiscalité publique, les imperfections du marché et les évolutions des couts de transport n'ont pas été intégrés à cause de la non disponibilité de données en libre accès ou de sources fiables.

En ce qui concerne les conséquences de la forme urbaine, nous avons retenu de plusieurs sources une influence positive de l'étalement urbain sur la consommation énergétique du secteur des transports par le biais de l'augmentation de la dépendance automobile qu'elle produit. L'augmentation des tailles de logement en périphérie entraîne également une hausse de consommation en énergie dans le secteur résidentiel. En résultat de ces deux effets, l'étalement urbain conduit indirectement à une augmentation du niveau des émissions de différents polluants qui

à leur tour se retrouvent dans l'air en en dégradent ainsi la qualité. La morphologie urbaine a également une influence sur la concentration de ces polluants par l'intermédiaire des effets de canyon urbain (Hauteur des bâtiments/Largeur de la rue).

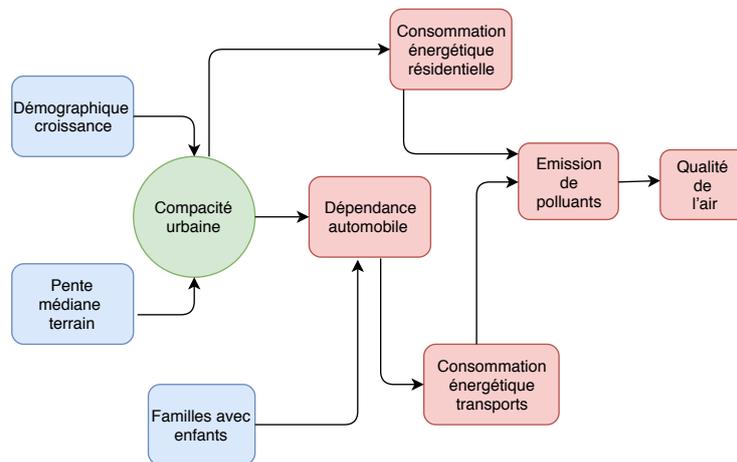


FIGURE 2.4 – Modèle structurel a priori. La couleur bleue designe une neutralité de la variable, le vert designe un effet attendu positif sur le plan environnemental, le rouge designe l'effet inverse. Une flèche de a à b signifie que a influence b (Source : auteur)

Nous avons ensuite intégré les données météorologiques et les informations concernant le parc du logement dans le modèle afin de contrôler leurs effets respectifs sur la concentration de polluants et le niveau de consommation d'énergie dans le secteur résidentiel. Les effets de la pente du terrain sur le climat et la concentration de polluants ont également été pris en considération. Les effets de la forme urbaine sur la santé n'ont pas été modélisés du fait de l'absence d'information sur ce dernier aspect.

### 2.7.2 Modèle de mesure

Il représente les liens entre les variables latentes et les variables permettant de les mesurer. La compacité urbaine est représentée par deux des indices développés par [Gervasoni et al., 2017], à savoir l'accessibilité des activités, la mixité d'usage du

sol et la densité de population par hectare. Le schéma de construction ici étant réflexif (les variables reflètent la compacité urbaine), la dispersion n'a pas été incluse dans le modèle car elle est en contradiction avec le reste des indices. Toutes les autres variables latentes du modèle sont définies par le schéma formatif c'est à dire qu'elles sont formées par les variables qui les mesurent et non l'inverse. Le concept de croissance démographique est mesuré par les taux d'évolution de la population par tranche d'âge . Les familles avec enfants comprennent le nombre de ménages avec enfants et la population de ces ménages. La variable latente représentant la dépendance automobile comprend les déplacements domicile-travail de la population active effectués en voiture par sexe. Les consommations énergétiques des secteurs transport et résidentiel sont mesurées respectivement par la consommation énergétique en transport routier et en transport de personnes d'une part, et d'autre part de la consommation en chauffage et des autres types de consommation. Pour les caractéristiques des logements, les nombre de résidences principales par nombre de pièces qui vont de 1 à 5, les années de construction des logement par tranches comprises entre 1919 et 2012 et le nombre de logement par type de chauffage (collectif et individuel). La variable émission de polluants est mesurée par les quantités de particules PM2.5 et PM10 totales émises en 2015. La concentration de polluants comprend les niveaux médians de concentration des particules PM2.5 et PM10 par commune. La pente du terrain est définie par le niveau médian d'élévation obtenu à partir du modèle numérique d'élévation des communes de la région. Et enfin, la variable climat qui est représentée par les niveau de températures et de pression vapeur saturante de l'eau annuelles médianes. Pour les variable latentes définies en schéma formatif, le mode PLScore a été choisi plutôt que la régression linéaire multivariée pour éviter les problèmes de multicollinéarité au sein des blocs de variables et maximiser la corrélation entre les variables latentes.

## 2.8 Présentation de l'échantillon de données

L'échantillon initial est constitué de 3507 communes sur les 4095 de la région Auvergne-Rhône-Alpes. Après omission des observations avec données manquantes contrainte par le logiciel utilisé pour l'estimation de notre modèle (Plspm, par [Sanchez et al., 2013]), cette taille est réduite à 3456, puis 1146 lorsque uniquement

les communes de plus de 1000 habitants sont sélectionnées, cela en application des préconisations de l'INSEE (Institut national de la statistique et des études économiques) à ne pas comparer les territoires de petites tailles (INSEE : Conseils d'utilisation, 2018).

Les indices de forme urbaine ont été obtenus à partir de la plateforme OpenStreetMap et un programme écrit en Python développé par [Gervasoni et al., 2017]. Les variables liées à la démographie, le parc du logement et les modes de déplacement des populations actives des communes sont issues de l'INSEE. Les données sur les différents types de polluants et leur niveau de concentration moyenne annuelle ont été obtenues du site internet du CRAIG (Centre régional Auvergne-Rhône-Alpes de l'information géographique) et la consommation énergétique par secteur à partir de celui de l'OREGES de la région Auvergne-Rhône-Alpes (Observatoire de l'énergie et des gaz à effet de serre). Les données sur la météo et sur les aspects topographiques des communes ont été respectivement tirées des sites web WorldClim et USGS. Le nettoyage et la fusion de l'ensemble des données ont été fait en développant le code des traitements nécessaires dans les langages Python et R. Les données météorologiques, topographiques et de concentration de polluant initialement sous format raster ont été extraites via le logiciel QGIS.

## 2.9 Description de l'échantillon

En matière de démographie, entre 2010 et 2014 le taux d'évolution de la population varie entre -13% et 38% environ. On observe en moyenne une baisse de la population des 30-44 ans de 1.1% dans la région, toutes les autres tranches étant en hausse. Le nombre moyen de couples avec enfant est de 445 par commune dans l'échantillon final, avec une forte disparité entre elles comme le montre l'écart-type qui est de 759879 couples en moyenne. Le nombre d'hommes à se rendre au travail en voiture est légèrement supérieur à celui des femmes avec des valeurs respectives de 159 et de 141 personnes approximativement. Là également les données sont très dispersées autour de la moyenne. En ce qui concerne l'énergie, le niveau de consommation lié au chauffage résidentiel est très supérieur à celui des autres types de consommation. En effet sa valeur moyenne est de 22,993 Gwh (gi-

gawattheure ) contre 9,936 Gwh pour les autres types de consommation d'énergie résidentiels, 13,898 Gwh pour le transport routier et 18,924 Gwh pour le transport de personnes. Les quantités de PM2.5 et PM10 sont là aussi très écartées entre les commune, avec des étendues des valeurs en tonnes s'élevant respectivement à 175.24 et à 248,19 pour des niveaux moyens régionales de 10 et 13 tonnes environs sur l'année 2015. Le tableau 2.9 présente quelques indicateurs statistiques par variable de notre échantillon.

Statistic	N	Mean	St. Dev.	Min	Max
Population	1,145	3,551.344	6,467.472	1,002	141,463
evo0014	1,146	0.038	0.185	-0.271	2.213
evo1529	1,146	0.035	0.201	-0.332	3.844
evo3044	1,146	-0.011	0.170	-0.328	2.019
evo4559	1,146	0.070	0.182	-0.199	2.822
evo6074	1,146	0.206	0.197	-0.126	2.765
men_coup_enf	1,146	445.121	759.879	75.927	15,940.760
pmen_coup_enf	1,146	1,724.595	2,999.758	283.462	63,756.270
elev_median	1,146	461.311	313.921	49.000	2,615.000
P14_RP_1P	1,146	62.895	418.731	0.000	11,067.010
P14_RP_2P	1,146	177.650	744.714	2.000	15,710.460
P14_RP_3P	1,146	340.451	1,187.419	17.586	25,816.550
P14_RP_4P	1,146	448.901	1,030.635	66.243	21,118.650
P14_RP_5PP	1,146	592.872	783.989	56.673	15,058.640
P14_RP_ACH19	1,146	203.473	351.505	18.216	7,290.923
P14_RP_ACH45	1,146	114.363	431.996	1.012	8,919.234
P14_RP_ACH70	1,146	340.436	1,377.003	8.090	30,421.940
P14_RP_ACH90	1,146	499.756	1,224.192	34.773	25,677.620
P14_RP_ACH05	1,146	275.419	493.117	13.028	9,620.361
P14_RP_ACH11	1,146	143.546	280.684	3.068	5,855.812
P14_RP_CCCOLL	1,146	313.184	1,637.966	0.000	37,808.630
P14_RP_CCIND	1,146	732.499	1,756.078	52.625	34,320.110
mean_temp	1,146	10.585	1.647	-1.042	13.871
mean_press_vap	1,146	0.971	0.092	0.433	1.125
landusemix_m2	1,146	0.034	0.026	0.0001	0.197
access_ha	1,146	6.187	9.442	0.017	69.672
pop_density	1,146	3.025	5.211	0.083	69.354
Nb_homm_voit	1,146	158.516	540.540	0.000	11,445.490
Nb_femm_voit	1,146	140.680	476.691	0.000	9,117.204
transport_routier	1,146	13.898	24.290	0.056	310.946
transport_pers	1,146	18.924	34.078	0.189	569.216
autre_conso_resid	1,146	9.936	23.579	1.741	385.418
chauff_resid	1,146	22.994 <sup>7</sup>	51.837	3.666	1,081.897
pm25_emi	1,146	10.165	11.193	0.760	176.000
pm10_emi	1,146	13.167	14.609	0.810	249.000
pm10_median	1,146	15.809	2.793	8.100	22.600
pm25_median	1,146	9.172	2.432	2.600	15.400

TABLE 2 – Statistiques descriptives (Source : auteur)

## 3 Résultats et discussion

### 3.1 Modèle initial

Après l'exécution du modèle, le schéma structurel suivant a été obtenu :

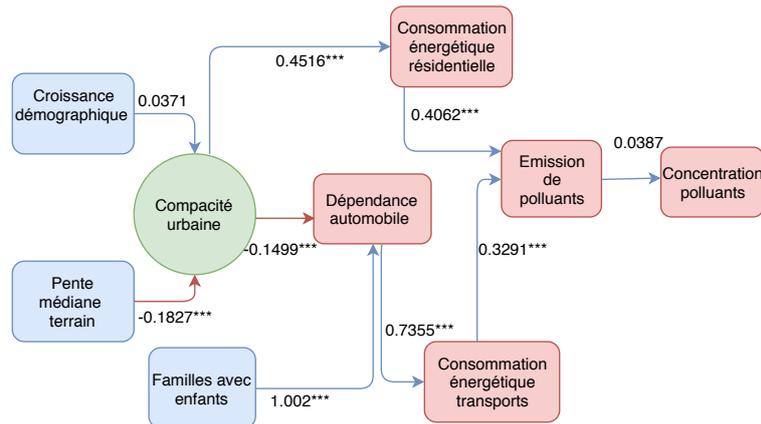


FIGURE 3.1 – Modèle 1 (Source : auteur)

Avant d'interpréter les résultats sur l'ampleur et le sens des relations entre les différentes variables de notre modèle théorique, il convient d'abord d'évaluer sa qualité d'abord au niveau du modèle de mesure, ensuite à celui du modèle structurel.

Comme le montre le tableau 1 de synthèse des critères d'évaluation du modèle PLS path modelling, les indicateurs de qualité du modèle externe dépendent du schéma du bloc qui est examiné. Sur les neuf variables, seule la compacité urbaine est définie en schéma réflectif (mode A) et les autres étant sous le schéma PLScore comme nous l'avons évoqué plus haut (sauf l'élévation médiane qui est mesurée par une seule variable B). Ce mode d'estimation des variables latentes est équivalent au mode B lorsque autant de composantes PLS que de variables manifeste par bloc sont utilisées dans le processus d'estimation (ce qui est le cas dans notre étude). Pour la compacité urbaine, les valeurs du l'alpha de Cronbach et du Rho de Dillon-Goldstein sont de 0.87 et 0.924 et donc largement au dessus du seuil de 0.7 suggéré dans la littérature pour vérifier l'unidimensionnalité au sein d'un

bloc. Le critère de Kaiser est également satisfait avec une première valeur propre à 2.4 et une deuxième à 0.2. En analysant la communauté, les valeurs des loadings (corrélations des variables avec le score de la variable latente) de la densité de population, de l'accessibilité et de la mixité d'usage sont respectivement de 0.94, 0.95 et de 0.77. Elles sont bien au dessus du seuil de 0.7. Pour les autres blocs, nous avons vu que la non multicolinéarité et la significativité des coefficients des variables manifestes régressées sur la variable latente sont utilisées pour évaluer le modèle externe. Pour la multicolinéarité, l'utilisation du mode PLScore qui remplace les variables par des composantes orthogonales permet d'emblée d'éviter ce problème. Pour le deuxième critère, la table 3 montre que tous les coefficients sont significatifs même lorsque les tests sont réalisés par Bootstrap (Valeur moyenne des coefficients calculée sur plusieurs tirages de sous échantillons avec remise à partir de l'échantillon initial). Cela se traduit par la non inclusion de la valeur zéro dans les intervalles de confiance estimés à partir de la distribution Bootstrap (distribution des valeurs obtenues pour toutes les échantillons).

	Original	Mean.Boot	Std.Error	perc.025	perc.975
croiss_demo-evo0014	0.282	0.284	0.008	0.272	0.303
croiss_demo-evo1529	0.282	0.284	0.008	0.272	0.303
croiss_demo-evo3044	0.282	0.284	0.008	0.272	0.303
croiss_demo-evo4559	0.282	0.284	0.008	0.272	0.303
struc_men-men_coup_enf	0.500	0.500	0.00002	0.500	0.500
struc_men-pmen_coup_enf	0.500	0.500	0.00002	0.500	0.500
topogr-elev_median	1	1	0	1	1
urban_compacity-landusemix_m2	0.211	0.209	0.027	0.143	0.252
urban_compacity-access_ha	0.410	0.413	0.014	0.387	0.441
urban_compactness-pop_density	0.476	0.474	0.023	0.428	0.525
auto_dep-nb_homm_voit	0.502	0.501	0.0005	0.501	0.502
auto_dep-nb_femm_voit	0.502	0.501	0.0005	0.501	0.502
transp_cons-transport_routier	0.509	0.509	0.002	0.506	0.513
transp_cons-transport_pers	0.509	0.509	0.002	0.506	0.513
resid_cons-autre_conso_resid	0.510	0.510	0.006	0.502	0.523
resid_cons-chauff_resid	0.510	0.510	0.006	0.502	0.523
polluant_emiss-pm10_emi	0.503	0.503	0.001	0.502	0.506
polluant_emiss-pm25_emi	0.503	0.503	0.001	0.502	0.506
air_pollution-pm10_median	0.516	0.516	0.001	0.515	0.518
air_pollution-pm25_median	0.516	0.516	0.001	0.515	0.518

TABLE 3 – Poids externes modèle 1 (Source : auteur)

Pour ce qui est du modèle interne, le degré de significativité des coefficients obtenu par régression des variables latentes structurellement liées est illustré sur le schéma précédent par le nombre d'astérisque affiché (\* : Pvalue < 0.1, \*\* : Pvalue < 0.05, \*\*\* : Pvalue < 0.01). Nous pouvons constater que seules les relations croissance démographique - compacité urbaine et émission de polluants - concentration de polluants ne sont pas significativement différents de 0. Les niveaux de  $R^2$  des modèles explicatifs de la compacité urbaine et de l'émission de polluants sont extrêmement bas (0.035 et 0.014 respectivement). Cela traduit une mauvaise qualité d'ajustement au niveau structurel, qui se répercute à son tour sur la valeur de l'indice de mesure de la qualité globale du modèle (GoF), qui s'élève à 0.5686. Rappelons que celui-ci est calculé à partir de la moyenne géométrique entre la moyenne des  $R^2$  et celle des AVE ou communautés moyennes par bloc.

Afin de contrôler les effets respectifs du climat et du parc de logement sur la pollution de l'air et la consommation énergétique du secteur résidentiel, nous avons donc conçu un second modèle en incluant les variables mesurant ces concepts.

## 3.2 Modèle final

Trois nouvelles interactions sont également introduites :

- Celle de la pente du terrain avec le climat et la concentration de la pollution
- Celle du climat avec le niveau de concentration de polluants
- Celle de la compacité urbaine sur la consommation énergétique du secteur des transports

L'estimation de ce modèle nous donne le nouveau schéma ci-dessous :

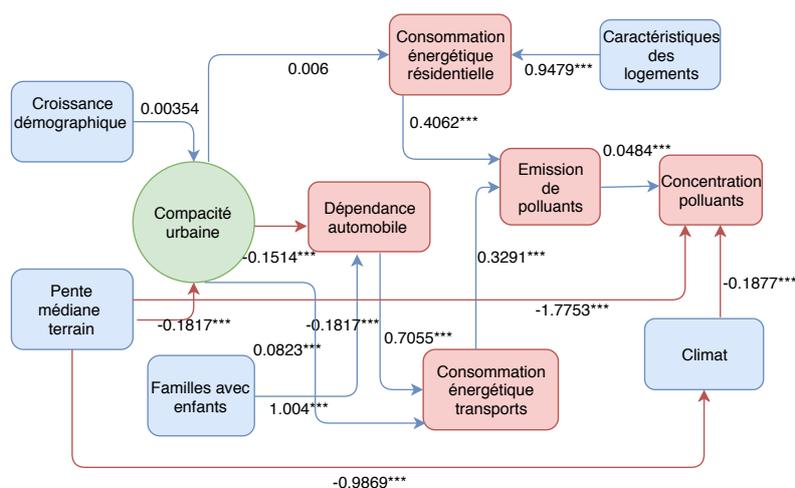


FIGURE 3.2 – Modèle 2 (Source : auteur)

Comme avec le modèle précédent, nous avons successivement évalué le modèle externe puis le modèle interne. Les critères d'unidimensionnalité sont tous satisfaits pour le seul bloc en mode réflexif de notre modèle, comme nous pouvons le voir dans le tableau 4 ci-dessous.

	Mode	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
croiss_demo	PLSCORE	4	0	0	3.155	0.530
struc_men	PLSCORE	2	0	0	2.000	0.0004
topogr	B	1	1	1	1	0
house_chars	PLSCORE	13	0	0	11.863	0.418
climate	PLSCORE	2	0	0	1.975	0.025
urban_compacity	A	3	0.876	0.924	2.410	0.428
auto_dep	PLSCORE	2	0	0	1.988	0.012
transp_cons	PLSCORE	2	0	0	1.929	0.071
resid_cons	PLSCORE	2	0	0	1.922	0.078
polluant_emiss	PLSCORE	2	0	0	1.974	0.026
air_pollution	PLSCORE	2	0	0	1.877	0.123

TABLE 4 – Inidimensionnalité : modèle 2 (Source : auteur)

Pour le critère de validité discriminante, les corrélations des variables mesurant la compacité urbaine avec le score de ce dernier sont toutes supérieures à leurs

corrélations avec les autres scores des variables latentes. Pour les autres blocs, tous les coefficients sont également significatifs, comme on peut le voir sur la table 5.

Dans le modèle interne, après contrôle de l'effet des caractéristiques des logements, l'effet de la compacité urbaine sur la consommation résidentielle baisse considérablement et devient non significatif. L'effet de la croissance démographique reste toujours non significative. Tous les autres coefficients sont significatifs à un niveau de risque de 1%. Pour un résumé des informations sur la qualité du modèle interne, voir la table 6.

Tous les  $R^2$  sont satisfaisants, sauf celui concernant le modèle explicatif de la forme urbaine. La part de la variance expliquée n'y est que de 3.5%. Par conséquent le niveau de la redondance qui lui est associé est très faible. En revanche, la forme urbaine a un effet significatif sur la dépendance automobile et la consommation d'énergie en transports. Son effet sur les consommation d'énergie du secteur résidentiel n'est pas significatif. Pour les autres blocs, toutes les liaisons sont significatives comme on a pu le voir sur le schéma du modèle structurel précédant. Le GoF connaît une nette amélioration pour ce second modèle en passant à 0.7416.

Nous pouvons ainsi passer à une analyse du graphique obtenu après exécution du calcul des paramètres du modèle. La compacité urbaine et l'étalement urbain étant étymologiquement deux concepts opposés, il est attendu que les effets du premier correspondent à l'opposé de ceux du deuxième. Ainsi par exemple, si l'étalement urbain entraîne une dégradation de la qualité de l'air, cela signifie qu'à contrario une forme urbaine caractérisée par une plus grande densité en habitants et en activités auront l'effet inverse. Nous avons ainsi obtenu conformément aux travaux cités dans la littérature un effet négatif et significatif de la pente du terrain sur la compacité urbaine (Corollaire des résultats obtenus par [Saiz, 2010]). Nous pouvons lire sur le graphique du modèle 2, la compacité urbaine a une influence négative significative sur la dépendance automobile, qui a sont tour affecte positivement et significativement la consommation en énergie du secteur des transports. Sans surprise, la consommation énergétique des deux secteurs inclus dans notre modèle ont un impact positif et significatif sur l'émission de particules fines dans l'air qui à leur tour impactent positivement le niveau de concentration de particules.

	Original	Mean.Boot	Std.Error	perc.025	perc.975
croiss_demo-evo0014	0.282	0.283	0.007	0.271	0.297
croiss_demo-evo1529	0.282	0.283	0.007	0.271	0.297
croiss_demo-evo3044	0.282	0.283	0.007	0.271	0.297
croiss_demo-evo4559	0.282	0.283	0.007	0.271	0.297
struc_men-men_coup_enf	0.500	0.500	0.00002	0.500	0.500
struc_men-pmen_coup_enf	0.500	0.500	0.00002	0.500	0.500
topogr-elev_median	1	1	0	1	1
house_chars-P14_RP_1P	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_2P	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_3P	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_4P	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_5PP	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH19	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH45	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH70	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH90	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH05	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_ACH11	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_CCCOLL	0.081	0.081	0.002	0.079	0.085
house_chars-P14_RP_CCIND	0.081	0.081	0.002	0.079	0.085
climate-mean_temp	0.503	0.503	0.0003	0.503	0.504
climate-mean_press_vap	0.503	0.503	0.0003	0.503	0.504
urban_compacity-landusemix_m2	0.183	0.183	0.024	0.130	0.232
urban_compacity-access_ha	0.425	0.429	0.015	0.399	0.463
urban_compacity-pop_density	0.483	0.479	0.027	0.428	0.541
auto_dep-nb_homm_voit	0.502	0.501	0.0004	0.501	0.502
auto_dep-nb_femm_voit	0.502	0.501	0.0004	0.501	0.502
transp_cons-transport_routier	0.509	0.509	0.002	0.506	0.512
transp_cons-transport_pers	0.509	0.509	0.002	0.506	0.512
resid_cons-autre_conso_resid	0.510	0.511	0.007	0.503	0.527
resid_cons-chauff_resid	0.510	0.511	0.007	0.503	0.527
polluant_emiss-pm10_emi	0.503	0.503	0.001	0.502	0.505
polluant_emiss-pm25_emi	0.503	0.503	0.001	0.502	0.505
air_pollution-pm10_median	0.516	0.516	0.001	0.515	0.517
air_pollution-pm25_median	0.516	0.516	0.001	0.515	0.517

TABLE 5 – Poids externes du modèle 2 (Source : auteur)

	Type	R2	Block_Community	Mean_Redundancy	AVE
croiss_demo	Exogenous	0	0.789	0	0
struc_men	Exogenous	0	1.000	0	0
topogr	Exogenous	0	1	0	0
house_chars	Exogenous	0	0.913	0	0
climate	Endogenous	0.974	0.988	0.962	0
urban_compacity	Endogenous	0.035	0.792	0.028	0.792
auto_dep	Endogenous	0.874	0.994	0.869	0
transp_cons	Endogenous	0.547	0.964	0.527	0
resid_cons	Endogenous	0.904	0.961	0.869	0
polluant_emiss	Endogenous	0.484	0.987	0.478	0
air_pollution	Endogenous	0.398	0.939	0.373	0

TABLE 6 – Résumé statistique du modele interne 2 (Source : auteur)

Cependant, bien que le graphique ne montre que des effets directs entre les diverses composantes du modèle, des effets indirects sont aussi calculables dans le modèle PLS path modelling. Ils sont calculés en effectuant le produit des coefficients des chemins qui vont indirectement d’une variable latente à une autre. Ainsi l’effet indirect de la pente du terrain sur la qualité de l’air est obtenu en effectuant le produit entre son effet sur le climat et l’influence de celui-ci sur la concentration de particules. Un effet total est ainsi obtenu en effectuant la somme de ces deux effets. Il est important de tenir compte de cela pour une approche globale des interactions de la forme urbaine avec les différents facteurs du modèle. La table 7 présente ces différents effets. Nous pouvons y voir que bien que l’effet direct de la compacité urbaine sur la consommation énergétique dans le secteur des transports soit positif, celui-ci est négatif au niveau global. De manière indirecte, nous observons également que la compacité urbaine affecte négativement l’émission de polluants et leur concentration dans l’air.

	Relations	directe	indirecte	total
1	croiss_demo ->urban_compacity	0.035	0	0.035
2	croiss_demo ->auto_dep	0	-0.005	-0.005
3	croiss_demo ->transp_cons	0	-0.001	-0.001
4	croiss_demo ->resid_cons	0	0.0002	0.0002
5	croiss_demo ->polluant_emiss	0	-0.0002	-0.0002
6	croiss_demo ->air_pollution	0	-0.00001	-0.00001
7	struc_men ->auto_dep	1.004	0	1.004
8	struc_men ->transp_cons	0	0.708	0.708
9	struc_men ->polluant_emiss	0	0.233	0.233
10	struc_men ->air_pollution	0	0.011	0.011
11	topogr ->climate	-0.987	0	-0.987
12	topogr ->urban_compacity	-0.182	0	-0.182
13	topogr ->auto_dep	0	0.028	0.028
14	topogr ->transp_cons	0	0.004	0.004
15	topogr ->resid_cons	0	-0.001	-0.001
16	topogr ->polluant_emiss	0	0.001	0.001
17	topogr ->air_pollution	-1.775	1.176	-0.600
18	house_chars ->resid_cons	0.948	0	0.948
19	house_chars ->polluant_emiss	0	0.385	0.385
20	house_chars ->air_pollution	0	0.019	0.019
21	climate ->air_pollution	-1.191	0	-1.191
22	urban_compacity ->auto_dep	-0.151	0	-0.151
23	urban_compacity ->transp_cons	0.082	-0.107	-0.024
24	urban_compacity ->resid_cons	0.006	0	0.006
25	urban_compacity ->polluant_emiss	0	-0.006	-0.006
26	urban_compacity ->air_pollution	0	-0.0003	-0.0003
27	auto_dep ->transp_cons	0.705	0	0.705
28	auto_dep ->polluant_emiss	0	0.232	0.232
29	auto_dep ->air_pollution	0	0.011	0.011
30	transp_cons ->polluant_emiss	0.329	0	0.329
31	transp_cons ->air_pollution	0	0.016	0.016
32	resid_cons ->polluant_emiss	0.406	0	0.406
33	resid_cons ->air_pollution	0	0.020	0.020
34	polluant_emiss ->air_pollution	0.048	0	0.048

TABLE 7 – Effets direct, indirects et totaux (Source : auteur)

## 4 Conclusion

L'objectif de ce travail a été de modéliser simultanément les facteurs qui contribuent à façonner l'environnement urbain en matière de densité de population et en activités et du niveau de mixité d'usages des sols d'un côté, et les mécanismes à travers lesquelles cette morphologie urbaine est liée à l'émission de polluants et leur concentration dans l'air au niveau des communes de la région Auvergne-Rhône-Alpes. Pour cela nous avons utilisé l'approche PLS path modelling, qui est une méthode d'estimation d'équations structurelles permettant de modéliser les relations entre des variables latentes obtenues par extraction de l'information contenue dans les données multidimensionnelles qui servent à les mesurer. La revue de littérature nous a permis de construire un modèle basé sur les connaissances a priori sur les relations entre les différents concepts étudiés. Au niveau de la mesure des variables latentes, les critères de validation ont été satisfaits pour toutes les variables. Ce qui traduit une bonne adéquation entre les données choisies et les concepts qu'elles représentent (ou l'inverse). Sur le plan structurel, nous avons obtenu une qualité de prédiction satisfaisante pour toutes les variables latentes à l'exception de la compacité urbaine dont le  $R^2$  est en dessous du seuil de 10% suggéré dans la littérature. Cela peut-être du au fait que la période entre 2010 et 2015 pour laquelle l'évolution de la population a été mesurée soit extrêmement courte pour contribuer à la formation de la structure urbaine des communes. Les nombre de facteurs explicatifs de la forme urbaine dans notre modèle est également très faible compte tenu de la complexité du phénomène. Comme nous l'avons vu dans la figure 1.4, l'étalement urbain est expliqué par plusieurs facteurs tels que l'imperfection du marché immobilier, les politiques publiques ou encore les préférences individuelles. Or nous ne disposons pas des données nécessaires pour intégrer ces éléments dans notre modèle. Les résultats de notre analyse montrent une association négative et significative entre la compacité urbaine et la dépendance automobile. Bien que cette compacité urbaine influe positivement sur la consommation énergétique du secteur des transports, nous avons pu voir que l'effet total sur ce dernier est négatif. Cela conforte une grande partie des travaux dans la littérature ([Newman and Kenworthy, 1989], [?] etc.). En revanche l'effet de la

compacité urbaine sur la consommation énergétique du secteur résidentiel n'est pas significatif. Soulignons par ailleurs que la littérature portant sur cette liaison est beaucoup moins abondante que la précédente. Ce que nous pouvons retenir de notre analyse est que la compacité urbaine a un effet négatif sur la dépendance automobile et la consommation énergétique en transports des personnes et pour le transport routier. En revanche, nous n'avons pas trouvé de lien significatif entre la forme urbaine et la consommation en énergie dans le secteur résidentiel. Une densité urbaine plus compacte est également liée négativement à la pollution émise et concentrée dans l'air mais ceci de façon indirecte.

## 5 Limites

Bien que notre modèle ait produit des résultats significatifs, l'impossibilité de spécifier des relations à double sens dans le modèle PLSPM est un handicap majeur pour étudier des relations de causalité. En effet, il est plausible d'observer dans la réalité des interactions dans les deux sens entre les différents concepts étudiés dans notre modèle. Par ailleurs, si une série chronologique des données est disponible, il est possible de spécifier des relations bidirectionnelles en alternant leurs sens sur des périodes distinctes. Le fait également que les indices de forme urbaine aient été calculés sur les données OpenStreetMap de l'année 2018 rend les enchaînements logiques des relations de notre modèle anachroniques. Cependant, nous avons fait l'hypothèse implicite que la forme urbaine d'aujourd'hui n'est pas significativement différente de celle d'il y a trois ans. Enfin, bien que nous utilisions le terme effet, le PLS path modelling est plutôt orienté prédiction (les scores obtenus par bloc de variables sont calculés de sorte à maximiser la variance contenue dans un bloc et non la covariance entre eux), nous interprétons néanmoins les résultats comme des effets en rapport avec les connaissances a priori sur l'interaction entre les différentes composantes guidée par la littérature.

## 6 Références

- [Wup, 2014] (2014). World urbanization prospects the 2014 revision. Technical report, United Nations.
- [Anthony, 2004] Anthony, J. (2004). Do state growth management regulations reduce sprawl? *Urban Affairs Review*, 39(3) :376–397.
- [Arribas-Bel and Schmidt, 2013] Arribas-Bel, D. and Schmidt, C. R. (2013). Self-organizing maps and the us urban spatial structure. *Environment and Planning B : Planning and Design*, 40(2) :362–371.
- [Arsanjani et al., 2015] Arsanjani, J. J., Mooney, P., Zipf, A., and Schauss, A. (2015). *Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets*. Springer.
- [Arti et al., 2013] Arti, S., Ayushman, M., and Bhawana, M. (2013). Urban sprawl development and need assessment of landfills for waste disposal : A case study of bilaspur municipal corporation of chhattisgarh, india. *Journal of Environmental Research and Development*, 7(4A) :1718.
- [Balestrat et al., 2011] Balestrat, M., Barbe, É., Chery, J.-P., Lagacherie, P., and Tonneau, J.-P. (2011). Reconnaissance du patrimoine agronomique des sols : une démarche novatrice en languedoc-roussillon. *Norois. Environnement, aménagement, société*, (221) :83–96.
- [Baum-Snow et al., 2016] Baum-Snow, N., Henderson, V., Turner, M. A., Zhang, Q., and Brandt, L. (2016). Highways, market access, and urban growth in china. *SERC Discussion Papers*, 200.
- [Bhatta, 2010] Bhatta, B. (2010). Causes and consequences of urban growth and sprawl. pages 17–36.
- [Bochet, 2005] Bochet, B. (2005). Morphologie urbaine et développement durable : transformations urbaines et régulation de l'étalement. A. DA CUNHA, JP. LERESCHE, P. KNOEPFEL, S. NAHRATH, *Enjeux du développement urbain durable. Transformations urbaines, gestion des ressources et régulation urbaine*, Presses Polytechniques et Universitaires Romandes, Lausanne, pages 55–77.

- [Brueckner, 2000] Brueckner, J. K. (2000). Urban sprawl : diagnosis and remedies. *International regional science review*, 23(2) :160–171.
- [Brunekreef and Holgate, 2002] Brunekreef, B. and Holgate, S. T. (2002). Air pollution and health. *The lancet*, 360(9341) :1233–1242.
- [Burchell et al., 1998] Burchell, R. W., Shad, N. A., Listokin, D., Phillips, H., Downs, A., Seskin, S., Davis, J. S., Moore, T., Helton, D., and Gall, M. (1998). *The costs of sprawl-revisited*. Number Project H-10 FY'95.
- [Burchfield et al., 2006] Burchfield, M., Overman, H. G., Puga, D., and Turner, M. A. (2006). Causes of sprawl : A portrait from space. *The Quarterly Journal of Economics*, 121(2) :587–633.
- [Burton, 2002] Burton, E. (2002). Measuring urban compactness in uk towns and cities. *Environment and Planning B : Planning and Design*, 29(2) :219–250.
- [Catherine Vakil, 2005] Catherine Vakil, M. (2005). Report on public health and urban sprawl in ontario.
- [Cavaud et al., 2016] Cavaud, D., Coléou, Z., Coltier, Y., Dussud, F.-X., Ecoiffier, M., Gong, Z., Misak, E., Mombel, D., and Reynaud, D. (2016). Chiffres clés de l'énergie edition 2016. Technical report, Service de l'Observation et des Statistiques.
- [Chen et al., 2008] Chen, H., Jia, B., and Lau, S. (2008). Sustainable urban form for chinese compact cities : Challenges of a rapid urbanized economy. *Habitat international*, 32(1) :28–40.
- [Cherrared et al., 2011] Cherrared, M., Zekiouk, T., and Chocat, B. (2011). Application de la méthode ahp pour l'évaluation de la performance des systèmes d'assainissement urbains. *Journal of Decision Systems*, 20(1) :103–127.
- [Chin, 2002] Chin, N. (2002). Unearthing the roots of urban sprawl : a critical analysis of form, function and methodology.
- [Chin, 1998] Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2) :295–336.
- [Cho and Choi, 2014] Cho, H.-S. and Choi, M. J. (2014). Effects of compact urban development on air pollution : Empirical evidence from korea. *Sustainability*, 6(9) :5968–5982.

- [Christiansen and Loftsgarden, 2011] Christiansen, P. and Loftsgarden, T. (2011). Drivers behind urban sprawl in europe. *TØI Report*, 1136 :2011.
- [Combes et al., 2012] Combes, P.-P., Duranton, G., and Gobillon, L. (2012). The costs of agglomeration : Land prices in french cities.
- [Da Cunha, 2005] Da Cunha, A. (2005). *Enjeux du développement urbain durable : transformations urbaines, gestion des ressources et gouvernance*. PPUR presses polytechniques.
- [da Silva et al., 2007] da Silva, A. N. R., Costa, G. C. F., and Brondino, N. C. M. (2007). Urban sprawl and energy use for transportation in the largest brazilian cities. *Energy for Sustainable Development*, 11(3) :44–50.
- [Dong and Yang, 2014] Dong, W. and Yang, Y. (2014). Exploitation of mineral resource and its influence on regional development and urban evolution in xinjiang, china. *Journal of Geographical Sciences*, 24(6) :1131–1146.
- [Ehrlich et al., 2017] Ehrlich, M. V., Hilber, C. A., and Schöni, O. (2017). Institutional settings and urban sprawl : Evidence from europe. *Journal of Housing Economics*.
- [Enault, 2004] Enault, C. (2004). La dilution : note méthodologique pour l’analyse de l’étalement urbain. *L’Espace géographique*, 33(3) :241–255.
- [Ewing, 1997] Ewing, R. (1997). Is los angeles-style sprawl desirable? *Journal of the American planning association*, 63(1) :107–126.
- [Ewing and Hamidi, 2015] Ewing, R. and Hamidi, S. (2015). Compactness versus sprawl : A review of recent evidence from the united states. *Journal of Planning Literature*, 30(4) :413–432.
- [Ewing et al., 2003] Ewing, R., Pendall, R., and Chen, D. (2003). Measuring sprawl and its transportation impacts. *Transportation Research Record : Journal of the Transportation Research Board*, (1831) :175–183.
- [Ewing and Rong, 2008] Ewing, R. and Rong, F. (2008). The impact of urban form on us residential energy use. *Housing policy debate*, 19(1) :1–30.
- [Frenkel and Ashkenazi, 2008] Frenkel, A. and Ashkenazi, M. (2008). Measuring urban sprawl : how can we deal with it? *Environment and Planning B : Planning and Design*, 35(1) :56–79.

- [Fricker et al., 2012] Fricker, S., Kreisler, C., and Tan, L. (2012). *An exploration of the application of PLS path modeling approach to creating a summary index of respondent burden*, pages 4141–4155.
- [Frumkin, 2002] Frumkin, H. (2002). Urban sprawl and public health. *Public health reports*.
- [Galster et al., 2001] Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., and Freihage, J. (2001). Wrestling sprawl to the ground : defining and measuring an elusive concept. *Housing policy debate*, 12(4) :681–717.
- [Garcia-López and Muñiz, 2013] Garcia-López, M.-À. and Muñiz, I. (2013). Urban spatial structure, agglomeration economies, and economic growth in barcelona : An intra-metropolitan perspective. *Papers in Regional Science*, 92(3) :515–534.
- [Garfinkel-Castro et al., 2017] Garfinkel-Castro, A., Kim, K., Hamidi, S., and Ewing, R. (2017). Obesity and the built environment at different urban scales : examining the literature. *Nutrition reviews*, 75(suppl\_1) :51–61.
- [Gervasoni et al., 2017] Gervasoni, L., Bosch, M., Fenet, S., and Sturm, P. (2017). Calculating spatial urban sprawl indices using open data. In *15th International Conference on Computers in Urban Planning and Urban Management*.
- [Glaeser and Kahn, 2004] Glaeser, E. L. and Kahn, M. E. (2004). Sprawl and urban growth. 4 :2481–2527.
- [Gottmann and Harper, 1990] Gottmann, J. and Harper, R. A. (1990). *Since Megalopolis : the urban writings of Jean Gottmann*. Johns Hopkins Univ Pr.
- [Guerrien, 2003] Guerrien, M. (2003). L’intérêt de l’analyse en composantes principales (acp) pour la recherche en sciences sociales. présentation à partir d’une étude sur le mexique. *Cahiers des Amériques latines*, (43) :181–192.
- [Hamidi et al., 2015] Hamidi, S., Ewing, R., Preuss, I., and Dodds, A. (2015). Measuring sprawl and its impacts : An update. *Journal of Planning Education and Research*, 35(1) :35–50.
- [Hennig et al., 2016] Hennig, Soukup, Orlitová, Schwick, and Kienast (2016). Urban sprawl in europe. Technical report, European Environment Agency–Federal Office for the Environment.

- [Henseler et al., 2009] Henseler, J., Ringle, C. M., and Sinkovics, R. R. (2009). *The use of partial least squares path modeling in international marketing*. Emerald Group Publishing Limited.
- [Jakobowicz, 2007] Jakobowicz, E. (2007). *Contributions aux modèles d'équations structurelles à variables latentes*. PhD thesis, Conservatoire national des arts et metiers-CNAM.
- [Jöreskog, 1970] Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. *ETS Research Bulletin Series*, 1970(2) :i–41.
- [Jun, 2004] Jun, M.-J. (2004). The effects of portland's urban growth boundary on urban development patterns and commuting. *Urban Studies*, 41(7) :1333–1348.
- [Kiakou, 2017] Kiakou (2017). The “urban sprawl” effect on out-of-town real estate market. *Openscience journal*.
- [Lopez, 2004] Lopez, R. (2004). Urban sprawl and risk for being overweight or obese. *American journal of public health*, 94(9) :1574–1579.
- [Lopez and Hynes, 2003] Lopez, R. and Hynes, H. P. (2003). Sprawl in the 1990s : measurement, distribution, and trends. *Urban Affairs Review*, 38(3) :325–355.
- [Mansfield et al., 2015] Mansfield, T. J., Rodriguez, D. A., Huegy, J., and MacDonald Gibson, J. (2015). The effects of urban form on ambient air pollution and public health risk : a case study in raleigh, north carolina. *Risk Analysis*, 35(5) :901–918.
- [Marique and Reiter, 2012] Marique, A.-F. and Reiter, S. (2012). A method for evaluating transport energy consumption in suburban areas. *Environmental Impact Assessment Review*, 33(1) :1–6.
- [Newman and Kenworthy, 1989] Newman, P. W. and Kenworthy, J. R. (1989). Gasoline consumption and cities : a comparison of us cities with a global survey. *Journal of the American planning association*, 55(1) :24–37.
- [O'Toole, 2001] O'Toole, R. (2001). The folly of smart growth. *Regulation*, 24 :20.
- [Pumain, 2003] Pumain, D. (2003). Une approche de la complexité en géographie. *Géocarrefour*, 78(1) :25–31.

- [Ravand and Baghaei, 2016] Ravand, H. and Baghaei, P. (2016). Partial least squares structural equation modeling with r. *Practical Assessment, Research & Evaluation*, 21(11) :1–16.
- [Sainteny, 2008] Sainteny, G. (2008). L'étalement urbain. (1) :7–15.
- [Saiz, 2010] Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3) :1253–1296.
- [Sanchez, 2013] Sanchez, G. (2013). Pls path modeling with r. *Berkeley : Trowchez Editions*, 383.
- [Sanchez et al., 2013] Sanchez, G., Trinchera, L., Sanchez, M. G., and FactoMineR, S. (2013). Package 'plsmpm'.
- [Shammin et al., 2010] Shammin, M. R., Herendeen, R. A., Hanson, M. J., and Wilson, E. J. (2010). A multivariate analysis of the energy intensity of sprawl versus compact living in the us for 2003. *Ecological Economics*, 69(12) :2363–2373.
- [Shim et al., 2006] Shim, G.-E., Rhee, S.-M., Ahn, K.-H., and Chung, S.-B. (2006). The relationship between the characteristics of transportation energy consumption and urban form. *The Annals of Regional Science*, 40(2) :351–367.
- [Stone Jr, 2008] Stone Jr, B. (2008). Urban sprawl and air quality in large us cities. *Journal of Environmental Management*, 86(4) :688–698.
- [Tabeaud, 2010] Tabeaud, M. (2010). Climats urbains. savoirs experts et pratiques sociales. *Ethnologie française*.
- [Tenenhaus et al., 2005] Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., and Lauro, C. (2005). Pls path modeling. *Computational statistics & data analysis*, 48(1) :159–205.
- [Thurstone, 1931] Thurstone, L. L. (1931). The measurement of social attitudes. *The journal of abnormal and social psychology*, 26(3) :249.
- [Tian et al., 2017] Tian, L., Li, Y., Yan, Y., and Wang, B. (2017). Measuring urban sprawl and exploring the role planning plays : A shanghai case study. *Land Use Policy*, 67 :426–435.
- [Tsai, 2005] Tsai, Y.-H. (2005). Quantifying urban form : compactness versus 'sprawl'. *Urban studies*, 42(1) :141–161.

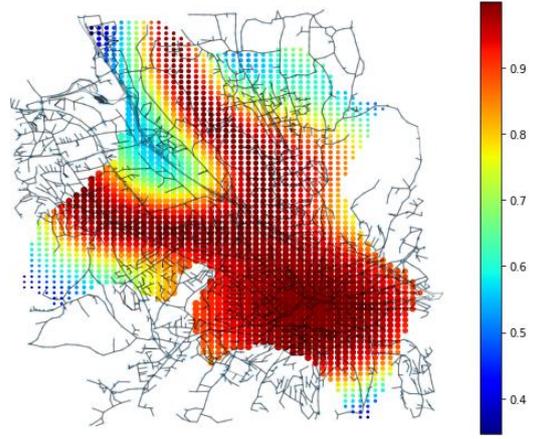
- [Vinzi et al., 2010a] Vinzi, V. E., Russolillo, G., and Trinchera, L. (2010a). An integrated pls regression-based approach for multidimensional blocks in pls path modeling. In *42èmes Journées de Statistique*.
- [Vinzi et al., 2010b] Vinzi, V. E., Trinchera, L., and Amato, S. (2010b). *PLS path modeling : from foundations to recent developments and open issues for model assessment and improvement*. Springer.
- [Wold, 1982] Wold, H. (1982). *Models for knowledge*. Springer.
- [Wold, 1966] Wold, H. O. (1966). A fix-point theorem with econometric background. *Arkiv för Matematik*, 6(3) :221–240.
- [Yue et al., 2013] Yue, W., Liu, Y., and Fan, P. (2013). Measuring urban sprawl and its drivers in large chinese cities : The case of hangzhou. *Land use policy*, 31 :358–370.
- [Zaninetti, 2014] Zaninetti, J.-M. (2014). La politique des périmètres d’urbanisation est-elle pertinente ? le cas de la région métropolitaine de portland, oregon. *Cybergeog : European Journal of Geography*.
- [Zielstra et al., 2014] Zielstra, D., Hochmair, H., Neis, P., and Tonini, F. (2014). Areal delineation of home regions from contribution and editing patterns in openstreetmap. *ISPRS International Journal of Geo-Information*, 3(4) :1211–1233.

## Annexes

Indices d’étalement urbain pour Chambéry

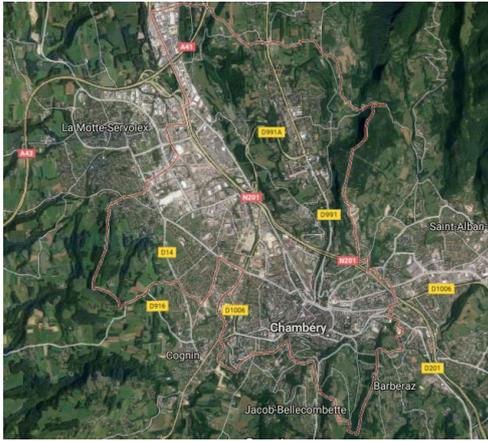


a

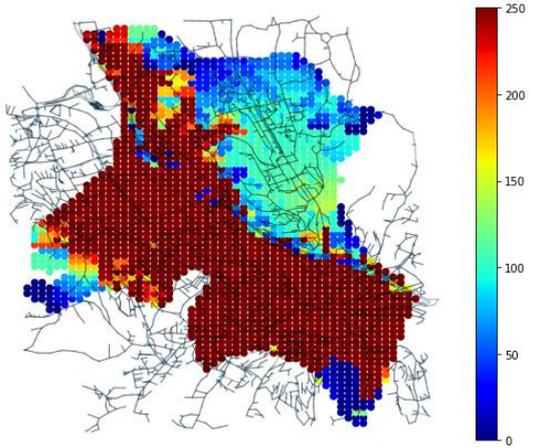


b

FIGURE .1 – (a) Image satellite de Chambéry, (b) mixité d'usage du sol



a

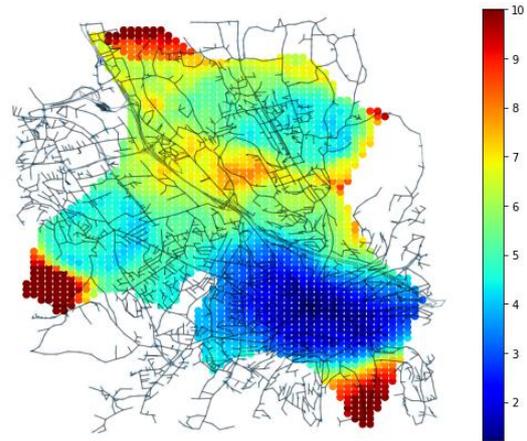


b

FIGURE .2 – (a) Image satellite de Chambéry, (b) accessibilité aux activités



a



b

FIGURE .3 – (a) Image satellite de Chambéry, (b) dispersion du bâti

Modèle externe 2

Crossloadings du modèle 2

Modèle interne (1)

Dictionnaire des variables

	name	block	weight	loading	communality	redundancy
1	evo0014	croiss_demo	0.282	0.913	0.833	0
2	evo1529	croiss_demo	0.282	0.853	0.728	0
3	evo3044	croiss_demo	0.282	0.906	0.820	0
4	evo4559	croiss_demo	0.282	0.879	0.773	0
5	men_coup_enf	struc_men	0.500	1.000	1.000	0
6	pmen_coup_enf	struc_men	0.500	1.000	1.000	0
7	elev_median	topogr	1	1	1	0
8	P14_RP_1P	house_chars	0.081	0.922	0.850	0
9	P14_RP_2P	house_chars	0.081	0.987	0.974	0
10	P14_RP_3P	house_chars	0.081	0.992	0.984	0
11	P14_RP_4P	house_chars	0.081	0.989	0.978	0
12	P14_RP_5PP	house_chars	0.081	0.947	0.898	0
13	P14_RP_ACH19	house_chars	0.081	0.914	0.835	0
14	P14_RP_ACH45	house_chars	0.081	0.940	0.883	0
15	P14_RP_ACH70	house_chars	0.081	0.978	0.956	0
16	P14_RP_ACH90	house_chars	0.081	0.981	0.962	0
17	P14_RP_ACH05	house_chars	0.081	0.946	0.895	0
18	P14_RP_ACH11	house_chars	0.081	0.882	0.777	0
19	P14_RP_CCCOLL	house_chars	0.081	0.954	0.911	0
20	P14_RP_CCIND	house_chars	0.081	0.979	0.959	0
21	mean_temp	climate	0.503	0.994	0.988	0.962
22	mean_press_vap	climate	0.503	0.994	0.988	0.962
23	landusemix_m2	urban_compacity	0.183	0.762	0.580	0.020
24	access_ha	urban_compacity	0.425	0.950	0.902	0.031
25	pop_density	urban_compacity	0.483	0.946	0.894	0.031
26	nb_homm_voit	auto_dep	0.502	0.997	0.994	0.869
27	nb_femm_voit	auto_dep	0.502	0.997	0.994	0.869
28	transport_routier	transp_cons	0.509	0.982	0.964	0.527
29	transport_pers	transp_cons	0.509	0.982	0.964	0.527
30	autre_conso_resid	resid_cons	0.510	0.980	0.961	0.869
31	chauff_resid	resid_cons	0.510	0.980	0.961	0.869
32	pm10_emi	polluant_emiss	0.503	0.994	0.987	0.478
33	pm25_emi	polluant_emiss	0.503	0.994	0.987	0.478
34	pm10_median	air_pollution	0.516	0.969	0.939	0.373
35	pm25_median	air_pollution	0.516	0.969	0.939	0.373

TABLE 8 – Modèle externe 2 (Source : auteur)

name	block	croiss_demo	struc_men	topogr	house_chars	climate	urban_compacity	auto_dep	transp_cons	resid_cons	polluant_emiss	air_pollution
1	evo0014	croiss_demo	0.913	-0.059	0.073	0.049	0.094	0.035	0.172	0.132	0.037	0.051
2	evo1529	croiss_demo	0.853	-0.007	-0.013	0.042	-0.021	-0.028	0.064	0.024	-0.029	0.060
3	evo3044	croiss_demo	0.906	0.088	0.070	0.036	0.109	0.033	0.159	0.128	0.021	0.070
4	evo4559	croiss_demo	0.879	-0.003	-0.015	-0.005	-0.028	-0.027	0.092	0.038	-0.034	0.049
5	men_coup_enf	struc_men	0.047	1.000	0.965	0.064	0.515	0.927	0.773	0.928	0.687	0.151
6	pmen_coup_enf	struc_men	0.046	1.000	0.967	0.066	0.513	0.924	0.769	0.923	0.682	0.154
7	elev_median	topogr	-0.044	1	-0.046	-0.987	-0.183	-0.001	-0.122	0.007	0.017	-0.399
8	P14_RP_1P	house_chars	0.045	0.819	-0.014	0.922	0.014	0.892	0.709	0.858	0.595	0.042
9	P14_RP_2P	house_chars	0.043	0.934	0.014	0.987	0.013	0.957	0.742	0.947	0.641	0.052
10	P14_RP_3P	house_chars	0.035	0.956	0.028	0.992	0.029	0.961	0.752	0.953	0.657	0.065
11	P14_RP_4P	house_chars	0.018	0.978	0.058	0.989	0.061	0.962	0.770	0.940	0.698	0.099
12	P14_RP_5PP	house_chars	0.018	0.965	0.087	0.947	0.089	0.930	0.728	0.902	0.728	0.151
13	P14_RP_ACH19	house_chars	0.017	0.852	0.065	0.914	0.312	0.912	0.671	0.859	0.645	0.024
14	P14_RP_ACH45	house_chars	0.006	0.854	0.028	0.940	0.383	0.919	0.671	0.885	0.632	0.031
15	P14_RP_ACH70	house_chars	0.031	0.942	0.028	0.978	0.031	0.952	0.753	0.941	0.644	0.068
16	P14_RP_ACH90	house_chars	0.010	0.973	0.041	0.981	0.042	0.952	0.760	0.931	0.688	0.099
17	P14_RP_ACH05	house_chars	0.057	0.946	0.059	0.946	0.058	0.908	0.773	0.911	0.680	0.140
18	P14_RP_ACH11	house_chars	0.068	0.868	0.072	0.882	0.068	0.844	0.686	0.830	0.649	0.133
19	P14_RP_CCCOLL	house_chars	0.057	0.933	0.006	0.954	0.005	0.915	0.744	0.928	0.601	0.057
20	P14_RP_CGIND	house_chars	-0.0002	0.941	-0.070	0.979	0.073	0.959	0.742	0.917	0.697	0.103
21	mean_temp	climate	0.038	0.070	-0.970	0.052	0.994	0.009	0.124	-0.005	-0.015	0.528
22	mean_press_vap	climate	0.031	0.060	-0.991	0.042	0.994	-0.0003	0.118	-0.012	-0.019	0.585
23	landusemix_m2	urban_compacity	0.079	-0.166	0.169	0.158	0.762	0.103	0.068	0.151	0.035	0.210
24	access_ha	urban_compacity	0.014	-0.186	0.436	0.179	0.950	0.348	0.349	0.423	0.321	0.302
25	pop_density	urban_compacity	0.048	-0.153	0.538	0.145	0.946	0.410	0.371	0.518	0.291	0.248
26	nb_homm_voit	auto_dep	0.006	0.927	0.005	0.973	0.353	0.997	0.736	0.918	0.694	0.040
27	nb_femm_voit	auto_dep	0.002	0.919	-0.008	0.966	0.375	0.997	0.730	0.917	0.699	0.041
28	transport_router	transp_cons	0.110	0.698	-0.142	0.686	0.290	0.662	0.982	0.699	0.614	0.218
29	transport_pers	transp_cons	0.159	0.816	-0.097	0.822	0.377	0.782	0.982	0.849	0.662	0.169
30	autre_conso_resid	resid_cons	0.086	0.890	-0.027	0.912	0.451	0.884	0.776	0.980	0.647	0.083
31	chauff_resid	resid_cons	0.091	0.925	0.040	0.952	0.446	0.920	0.769	0.980	0.658	0.029
32	pm10_emi	polluant_emiss	0.0003	0.650	-0.008	0.654	0.260	0.620	0.630	0.630	0.994	0.051
33	pm25_emi	polluant_emiss	-0.004	0.710	0.042	0.716	0.304	0.724	0.660	0.693	0.994	0.026
34	pm10_median	air_pollution	0.053	0.147	-0.579	0.080	0.249	0.044	0.194	0.053	0.033	0.969
35	pm25_median	air_pollution	0.071	0.148	-0.582	0.086	0.306	0.035	0.188	0.059	0.042	0.969

TABLE 9 – Crossloadings du modèle 2 (Source : auteur)

climat	Estimate	Std. Error	t value	Pr(> t )
Intercept	-0	0.005	-0	1
topogr	-0.987	0.005	-206.928	0
urban compacity				
Intercept	0	0.029	0	1
croiss_demo	0.035	0.029	1.219	0.223
topogr	-0.182	0.029	-6.246	0
Automobile dependance				
Intercept	-0	0.011	-0	1
struc_men	1.004	0.012	81.951	0
urban_compacity	-0.151	0.012	-12.361	0
transport energy cons				
Intercept	0	0.020	0	1
urban_compacity	0.082	0.021	3.850	0.0001
auto_dep	0.705	0.021	32.986	0
Residential energy cons				
Intercept	-0	0.009	-0	1
house_chars	0.948	0.010	90.970	0
urban_compacity	0.006	0.010	0.577	0.564
Polluant emission				
Intercept	0	0.021	0	1
transp_cons	0.329	0.035	9.530	0
resid_cons	0.406	0.035	11.765	0
Air quality				
Intercept	-0	0.023	-0	1
topogr	-1.775	0.142	-12.472	0
climate	-1.191	0.142	-8.369	0
polluant_emiss	0.048	0.023	2.108	0.035

TABLE 10 – Modele interne 1 (Source : auteur)

TABLE 11 – Dictionnaire des variables

Variable	Signification	Source
POPULATION	Niveau de population	IGN BD Carto
evo0014	Évolution du nombre de personnes de 0 à 14 ans	Insee
evo1529	Évolution du nombre de personnes de 15 à 29 ans	Insee
evo3044	Évolution du nombre de personnes de 30 à 44 ans	Insee
evo4559	Évolution du nombre de personnes de 45 à 49 ans	Insee
evo6074	Évolution du nombre de personnes de 60 à 74 ans	Insee
men_coup_enf	Nb de ménages formées d'un couple avec enfant(s)	Insee, RP.
pmen_coup_enf	Nb de personnes des ménages formées d'un couple avec enfant(s)	Insee, RP.
elev_median	Niveau median de la pente du terrain	USGS
P14_RP_1P	Rés princ 1 pièce en 2014	Insee
P14_RP_2P	Rés princ 2 pièce en 2014	Insee
P14_RP_3P	Rés princ 3 pièce en 2014	Insee
P14_RP_4P	Rés princ 4 pièce en 2014	Insee
P14_RP_5PP	Rés princ 5 pièce en 2014	Insee
P14_RP_ACH19	Rés princ avt 1919 en 2014	Insee
P14_RP_ACH45	Rés princ avt 1945 en 2014	Insee
P14_RP_ACH70	Rés princ avt 1970 en 2014	Insee
P14_RP_ACH90	Rés princ avt 1990 en 2014	Insee
P14_RP_ACH05	Rés princ avt 2005 en 2014	Insee
P14_RP_ACH11	Rés princ avt 2011 en 2014	Insee
P14_RP_CCCOLL	Rés princ Chauffage Central Collectif en 2014	Insee
P14_RP_CCIND	Rés princ Chauffage Central individuel en 2014	Insee
landusemix_m2	Indice de mixité d'usage du sol	Gervasoni et al (2017)
access_ha	Indice de d'accessibilité des activités	Gervasoni et al (2017)
pop_density	Densité de population	IGN BD Carto
nb_homm_voit	Nb d'hommes qui vont au travail ds la commune en voiture	Insee, exploitation complémentaire
nb_femm_voit	Nb de femmes qui vont au travail ds la commune en voiture	Insee, exploitation complémentaire
transport_routier	consommation en énergie des transports routiers	OREGES
transport_pers	consommation en énergie en transports de personnes	OREGES
autre_conso_resid	autres consommation en énergie du sect résidentiel	OREGES
chauff_resid	consommation en résidentielle en chauffage	OREGES
pm10_emi	PM10 émis en tonnes	CRAIG
pm25_emi	PM25 émis en tonnes	CRAIG
pm10_median	concentration médiane de PM10	CRAIG
pm25_median	concentration médiane de PM25	CRAIG