

Relations entre les caractéristiques socio-économiques de la population et l'exposition au risque d'inondation en Wallonie

Auteur : Lecron, Séléna

Promoteur(s) : Cornet, Yves; Dewals, Benjamin; Donnay, Jean-Paul

Faculté : Faculté des Sciences

Diplôme : Master en sciences géographiques, orientation géomatique et géométrologie, à finalité spécialisée

Année académique : 2018-2019

URI/URL : <http://hdl.handle.net/2268.2/7298>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Relations entre les caractéristiques socio-économiques de la population et l'exposition au risque d'inondation en Wallonie

Mémoire présenté par : Séléna LECRON
pour l'obtention du titre de
*Master en sciences géographiques,
orientation géomatique et géométrologie*

Année académique : **2018-2019**
Date de défense : **septembre 2019**

Président de jury : **Jean-Paul DONNAY**
Promoteurs : **Yves CORNET**
Benjamin DEWALS
Jean-Paul DONNAY
Jury de lecture : **Pierre ARCHAMBEAU**
Geoffrey HOUBRECHTS

Relations entre les caractéristiques socio-économiques de la population et l'exposition au risque d'inondation en Wallonie

Mémoire présenté par : Sélène LECRON
pour l'obtention du titre de
*Master en sciences géographiques,
orientation géomatique et géométrologie*

Année académique : **2018-2019**
Date de défense : **septembre 2019**

Président de jury : **Jean-Paul DONNAY**
Promoteurs : **Yves CORNET**
Benjamin DEWALS
Jean-Paul DONNAY
Jury de lecture : **Pierre ARCHAMBEAU**
Geoffrey HOUBRECHTS

Remerciements

Je souhaite tout d'abord remercier M. Donnay pour son suivi et sa connaissance sans lesquels je n'aurais pas pu réaliser mes analyses.

Je tiens également à remercier M. Cornet pour son aide et ses remarques. Son dynamisme tout au long de l'année a été d'un grand soutien.

Je remercie M. Dewals pour la proposition de ce sujet de mémoire aussi intéressant que pertinent. Je veux également dire merci à l'HECE pour les données hydrographiques nécessaire à la réalisation de ce mémoire.

Je remercie M. Archambeau pour les données hydrologiques ainsi que pour ses explications concernant les données.

Je remercie M. Binard pour les données fournies sur l'occupation du sol et les secteurs statistiques.

Enfin, je souhaite remercier ma famille pour le soutien tout au long de l'année. Je remercie également mes amis pour l'aide et la bonne humeur qu'ils m'ont apportées.

Résumé

Le phénomène naturel d'inondation est un sujet de préoccupation des politiques de gestion de l'environnement. C'est dans ce contexte que le plan de Prévention et Lutte contre les Inondations et leurs Effets sur les Sinistrés (PLUIE) a été élaboré. Ce plan est à l'origine de la cartographie de l'aléa d'inondation. L'objectif de notre recherche est de déterminer si la population présente dans ces zones à risque d'inondation présente des caractéristiques socio-économiques dites faibles. Pour ce faire, deux approches sont envisagées. D'une part, la population des secteurs statistiques, échelle à laquelle notre recherche est effectuée, est caractérisée par le revenu médian et par l'écart relatif entre le revenu médian du secteur et le revenu médian de sa commune. Des régressions linéaires multiples sont réalisées pour expliquer les deux variables en tenant compte de 3 variables indépendantes, du statut d'inondation du secteur et de l'autocorrélation spatiale de la variable dépendante. Le statut d'inondation d'un secteur est défini par une variable binaire. Elle qualifie un secteur statistique « d'inondé » ou de « non inondé » en fonction de la part de sa surface habitée en zone à risque d'inondation. Ensuite, une auto-régression spatiale LAG est effectuée avec les 3 mêmes variables explicatives. D'autre part, un indice synthétique, créé à partir d'une analyse en composantes principales et décrivant le niveau de vie des secteurs statistiques, est exploité pour caractériser la population. Un test d'indépendance du χ^2 et une ANOVA à un et deux facteurs sont utilisés afin de déterminer s'il existe une relation entre le statut d'inondation d'un secteur et son niveau de vie. Ensuite, nous avons exploité une régression Logit ayant pour but de prédire l'occurrence des événements « inondés » et « non inondés » en fonction du niveau de vie. Les résultats de ces deux approches sont mitigés quant à l'influence du statut d'inondation d'un secteur statistique sur les caractéristiques socio-économiques de la population de celui-ci. Certains résultats maintiennent que la population située dans un secteur inondé possède un revenu médian et un niveau de vie faible ainsi qu'un écart relatif des revenus négatif. Cependant la tendance générale des résultats infirme l'hypothèse de départ, c'est-à-dire que la population définie par des caractéristiques socio-économiques faibles ne vit pas incontestablement dans les secteurs statistiques à risque d'inondation élevé.

Abstract

The natural phenomenon of flooding is a concern of environmental management policies. Therefore the plan of « Prévention et Lutte contre les Inondations et leurs Effets sur les Sinistrés » (PLUIE) was developed. This plan is at the origin of flood hazard mapping. The objective of our research is to determine whether the population living in these flood risk areas has so-called low socio-economic characteristics. For this purpose, two approaches are considered. On the one hand, the population of statistical sectors, at which scale our research is conducted, is characterized by the median income and by the relative difference between the median income of the sector and the median income of its municipality. Multiple linear regressions are performed to explain the two variables taking into account 3 independent variables, the flooding status of the area and the spatial autocorrelation of the dependent variable. The flood status of an area is defined by a binary variable. It qualifies a statistical sector as "flooded" or "unflooded" according to the proportion of its inhabited area in the flood risk zone. Then, a spatial LAG auto-regression is performed with the same 3 explicative variables. On the other hand, a synthetic index, created from a principal components analysis and describing the quality of life of statistical sectors, is used to characterize the population. A χ^2 independence test and one- and two-way ANOVA are used to determine whether there is a relationship between the flood status of an area and its quality of life. Then, we applied a Logit regression aiming to predict the occurrence of "flooded" and "unflooded" events in function of the quality of life. The results of these two approaches are mitigated regarding the influence of the flood status of a statistical sector on the socio-economic characteristics of its population. Some results maintain that the population living in a flooded area has a low median income and a low quality of life as well as a negative relative income difference. However, the general trend of the results reverses the initial hypothesis, i.e. that the population defined by low socio-economic characteristics does not undoubtedly live in statistical sectors at high risk of flooding.

Table des matières

Remerciements.....	1
Résumé.....	2
Abstract	3
Liste des figures	7
Liste des tableaux	8
CHAPITRE 1: INTRODUCTION	1
CHAPITRE 2: ÉTAT DE L'ART.....	3
I.Description des caractéristiques socio-économiques de la population à risque	3
II.Risque d'inondation en Wallonie: l'aléa d'inondation	7
III.Cartographie densimétrique.....	8
IV.mMéthodes et résultats de recherches précédentes	9
CHAPITRE 3: HYPOTHÈSE.....	11
CHAPITRE 4: DÉVELOPPEMENT THÉORIQUE	12
I.Présentation des données	12
I. 1.Recensement socio-économique de 2011 et données statistiques complémentaires	12
I. 2.Les limites des secteurs statistiques	13
I. 3.L'occupation du sol.....	13
I. 4.L'aléa d'inondation.....	13
II.Prétraitements et analyse descriptive des données.....	14
II. 1.Prétraitements et analyse des indices	14
II.1. i.Sélection et construction des indices.....	14
II.1. ii.Analyse descriptive des valeurs des indices	14
II. 2.Prétraitement de l'aléa d'inondation	18
II. 3.Prétraitement des secteurs statistiques	20
III.Présentation des traitements	22
III. 1Analyse en composantes principales	22
III.1. i.Choix des composantes à conserver	23
III.1. ii.Score factoriel	23
III. 2.Analyse de contingence par χ^2 (χ^2) ou test d'indépendance	24
III. 3.ANOVA	25
III.3. i.ANOVA à un facteur de classification : Une seule variable explicative X	25
III.3. ii.ANOVA à deux facteurs de classification : deux variables explicatives X1 et X2..	27
III.3. iii.Comparaison multiple des moyennes	28

III.3. iv.Non-respect des conditions d'utilisation de l'ANOVA	29
III. 4.Régressions classiques	30
III.4. i.Régression linéaire simple	30
III.4. ii.Régression linéaire multiple	31
III.4. iii.Ajout d'une variable auxiliaire	32
III.4. iv.Introduction du concept d'interaction des variables explicatives	33
III. 5.Régression Logit	34
III.5. i.Rapport de cote (Odd ratio)	34
III.5. ii.Modèle de régression Logit	34
III.5. iii.Interprétation des coefficients de la régression	34
III.5. iv.Estimation des coefficients par maximum de vraisemblance	35
III. 6.Mesure de l'autocorrélation spatiale	37
III.6. i.Autocorrélation spatiale	37
III.6. ii.Matrice de voisinage	37
III.6. iii.Indice de Moran	38
III.6. iv.Auto-régression spatiale	40
CHAPITRE 5: APPLICATION ET VALIDATION	42
I.Première approche: Population définie selon le revenu médian	42
I. 1.Variation spatiale du revenu	43
I. 2.Autocorrélation spatiale	43
I. 3.Régressions sur le revenu médian des secteurs statistiques	46
I.3. i.Régression multiple sur tous les secteurs	46
I.3. ii. Régressions multiples pour les 2 jeux de données, secteurs inondés / secteurs non inondé	47
I.3. iii. Régression multiple avec une variable auxiliaire binaire (inondé / non inondé)	48
I.3. iv.Régression multiple avec deux variables auxiliaires	50
I.3. v.Régression LAG	51
I.3. vi.Conclusions sur les régressions multiples du revenu	54
I. 4.Régressions sur l'écart relatif entre le revenu des secteurs et le revenu de leur commune	54
I.4. i.Analyse et distribution spatiale de l'écart relatif des revenus	55
I.4. ii.Régression multiple sur tous les secteurs	56
I.4. iii. . Régressions multiples pour les 2 jeux de données, secteur inondés / non inondés	56

I.4. iv. Régression multiple avec une variable auxiliaire binaire (inondé / non inondé)	58
I.4. v. Régression multiple avec deux variables auxiliaires	58
I.4. vi. Régression LAG	60
I.4. vii. Conclusion générale de la première approche sur le revenu médian	63
II... Deuxième approche : Population définie selon la composante synthétique du niveau de vie	64
II. 1. Sélection des indices sur base des corrélations et de la variance	64
II. 2. Construction de la composante « niveau de vie »	65
II. 3. Analyse du score factoriel	66
II. 4. Mesure de l'autocorrélation du score factoriel	69
III. Analyse combinée du statut d'inondation et du niveau de vie d'un secteur statistique	70
III. 1. χ^2	70
III. 2. ANOVA	72
III.2. i. Anova à un facteur	72
III.2. ii. Anova à deux facteurs	76
III. 3. Régression Logit du score factoriel	84
CHAPITRE 6: CONCLUSION	86
Bibliographie	89
Liste des annexes	94

Liste des figures

Figure 1: Modèle de vulnérabilité à l'aléa naturel (Cutter et al., 2003).....	4
Figure 2: Schéma de détermination de l'aléa inondation par débordement (Beckers et al., 2013; modifiée).....	7
Figure 3: Partitionnement d'un secteur statistique en fonction de deux critères (la zones habitée et l'aléa d'inondation)	19
Figure 4: Diagramme des trois proportions de surface inondée en zone habitée avec un seuil de 80% de surface non inondée en zone habitée permettant de classer les secteurs en 2 modalités : inondés et non inondés	19
Figure 5 : Observations réparties en 3 groupes avec la moyenne des groupes (en jaune) et la moyenne générale (en vert). L'écart total (en bleu), l'écart résiduel (en rouge) et l'écart factoriel (en mauve)(Timmermans, 2017)26	
Figure 6 : Fonction logit, proche d'une fonction sigmoïde exemple de pour différente valeur du coefficient de la régression (Donnay, 2019b).....	35
Figure 7: Exemple de la définition de contiguïté.....	38
Figure 8 : Diagramme de Moran : exemple obtenu avec 7.512 secteurs statistiques. Variable : première composante de l'ACP, niveau de vie. En abscisse : le niveau de vie des secteurs statistiques, En ordonnée : moyenne du niveau de vie des secteurs voisins au secteur	39
Figure 9: Test de permutation aléatoire du revenu médian	44
Figure 10 :Diagramme de l'indice de Moran du revenu médian par secteur statistique.	44
Figure 11: Test de permutation aléatoire de l'écart relatif des revenus médians	45
Figure 12 : Diagramme de l'indice de Moran de l'écart relatif des revenus médians par secteur statistique.	45
Figure 13: Test de permutation de l'indice de Moran des résidus de la régression sur le revenu	51
Figure 14: Diagramme de Moran des résidus de la régression classique du revenu	51
Figure 15: Diagramme de Moran des résidus de la régression classique sur l'écart relatif des revenus	60
Figure 16 : Test de permutation de l'indice de Moran des résidus de la régression sur l'écart relatif des revenus	60
Figure 17 : Résultat de la régression LAG sur l'écart relatif des revenus médian où les variables indépendantes sont $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage}$	61
Figure 18: Corrélation entre les 24 indices socio-économiques calculés à partir du recensement de 2011	65
Figure 19: Histogramme du score factoriel du niveau de vie pour 7.525 secteurs	67
Figure 20: Test de permutation de l'indice de Moran du score factoriel	69
Figure 21: Indice de Moran calculé sur le score factoriel du niveau de vie avec le nombre de secteurs par quadrant.....	69
Figure 22:Dispersion du score factoriel des 7.512 secteurs statistiques dans les deux classes d'inondation	73
Figure 23: Histogramme des résidus.....	74
Figure 24: Dispersion du score factoriel des 7510 secteurs en fonction de leur position dans le diagramme de Moran	76
Figure 25: Dispersion du score factoriel des 7510 secteurs en fonction à leur statut d'inondation et de leur position dans le diagramme de Moran.....	77
Figure 26: Représentation graphique de la moyenne du score factoriel des secteurs pour chaque couple de facteurs déterminant les groupes de l'ANOVA (statut d'inondation et classe de Moran).....	78
Figure 27: Histogramme des résidus de l'ANOVA	81
Figure 28: Probabilités prédites pour $y=1$ (inondé) en fonction du score factoriel suite à la régression logit	85

Liste des tableaux

Tableau 1 : Concept de la vulnérabilité sociale (Cutter et al., 2003)	5
Tableau 2: Vingt-quatre indices socio-économiques calculés à partir des données du recensement de 2011 (1/3 pages)	15
Tableau 3: Explication de la variance totale, intergroupe, intragroupe et de la statistique de Fisher (Timmermans, 2017)	26
Tableau 4: Somme des carrés des écarts, variance intergroupe, intragroupe, totale et statistique de Fisher d'une ANOVA à 2 facteurs où le nombre d'individus par catégorie est le même (Donnay, 2019a).....	28
Tableau 5: Nombre de secteurs statistiques dans les 4 quadrants du diagramme de Moran obtenu à partir du revenu médian et de l'écart relatif des revenus.....	45
Tableau 6: Régression sur le revenu médian des secteurs statistiques (Y) (7.512) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	46
Tableau 7: Régression sur le revenu médian par secteur statistique inondé (1.391) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	47
Tableau 8 : régression multiple du revenu effectuée sur les 6121 secteurs non inondés avec $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	47
Tableau 9 : Comparaison des moyennes du revenu des secteurs inondés et non inondés	48
Tableau 10 : Régression sur le revenu médian des secteurs statistiques (7.512) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage} + X4 = \text{ClasseInondation}$	48
Tableau 11 : Régression sur le revenu médian des secteurs statistiques (7.512) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage} + X4 = \text{ClasseInondation}$ avec interactions.....	49
Tableau 12 : Régression sur le revenu médian des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage} + X4 = \text{ClasseMoran} + X5 = \text{ClasseInondation}$	50
Tableau 13 : Résultat de la régression LAG sur le revenu médian où les variables indépendantes sont $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	52
Tableau 14 : Statistiques descriptives de l'écart relatif des revenus.....	55
Tableau 15 : Régression sur l'écart relatif des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	56
Tableau 16 : Régression sur l'écart relatif des revenus des secteurs statistiques INONDES (1.391) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	56
Tableau 17 : Régression sur l'écart relatif des revenus des secteurs statistiques NON INONDES (6.121) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	57
Tableau 18 : Comparaison des moyennes de l'écart des revenus des secteurs inondés et non inondés	57
Tableau 19 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$	58
Tableau 20 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage} + X4 = \text{ClasseMoran}$	59
Tableau 21 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage} + X4 = \text{ClasseMoran} + X5 = \text{ClasseInondation}$	59
Tableau 22: Indices redondants avec leur variance et leur corrélation pour justifier la sélection	64
Tableau 23: Résultat de l'ACP sur la composante 1, celle correspondant au niveau de vie	66
Tableau 24 : Résumé des statistiques descriptives des 7.525 secteurs (avant suppression des cas erratiques) ..	67
Tableau 25: Classement des secteurs statistiques en fonction du score factoriel du niveau de vie (5 classes)	70
Tableau 26: Table de contingence des fréquences observées entre le score factoriel mesurant le niveau de vie et le statut d'inondation	71

Tableau 27: Table de contingence des fréquences attendues entre le score factoriel mesurant le niveau de vie et le statut d'inondation	71
Tableau 28: Ecart entre les fréquences observées et les fréquences attendues entre le score factoriel et le statut d'inondation.....	71
Tableau 29 : Statistique descriptive du score factoriel du niveau de vie des 7.512 secteurs en fonction de leur statut d'inondation	72
Tableau 30: Résultat de l'ANOVA à 1 facteur (statut d'inondation) sur le niveau de vie.....	73
Tableau 31: Comparaison couple de la moyenne des groupes formés par le statut d'inondation	74
Tableau 32: Test d'homogénéité de la variance intragroupe de l'ANOVA à 1 facteur (le statut d'inondation) ..	74
Tableau 33: Test de normalité des résidus de l'ANOVA à 1 facteur (le statut d'inondation).....	74
Tableau 34: Statistique descriptive du score factoriel des secteurs en fonction du quadrant où ils se trouvent dans le diagramme de Moran	76
Tableau 35: ANOVA à 2 facteurs (statut d'inondation et classe de Moran) sans interaction	77
Tableau 36: ANOVA à 2 facteurs (le statut d'inondation et la classe Moran) avec interaction.....	79
Tableau 37: Comparaison par paire des moyennes des groupes avec interaction dans le cadre d'une ANOVA à 2 facteurs (statut d'inondation et classe de Moran).....	80
Tableau 38: Test d'homogénéité de la variance intragroupe de l'ANOVA à 2 facteurs (statut d'inondation et classe de Moran).....	81
Tableau 39: Test de normalité des résidus de l'ANOVA à 2 facteurs (statut d'inondation et classe de Moran) .	81
Tableau 40: ANOVA 2 facteurs type III : Comparaison de la moyenne des 8 groupes à la moyenne générale sur 7.510 secteurs statistiques	82
Tableau 41: ANOVA 2 facteurs type III : Comparaison de la moyenne des 8 groupes à la moyenne générale sur 7510 secteurs statistiques avec interaction.....	82
Tableau 42: Résultat de la régression Logit sur le statut d'inondation Y.....	84
Tableau 43: Test de significativité des odds ratio du modèle Logit où Y= le statut d'inondation et X= le score factoriel.....	84
Tableau 44: Matrice de confusion issue du test de prédiction des valeurs supérieures à 0,5.....	85

CHAPITRE I: INTRODUCTION

« Depuis 1996 en Wallonie, 450 nouveaux bâtiments, représentant 500.000 m² de terrain, sont ainsi construits en moyenne chaque année. » (CEEW, 2007). Cependant, cette urbanisation croissante a un impact non négligeable sur l'environnement. Les infrastructures érigées provoquent l'imperméabilisation du sol engendrant ainsi un changement local des conditions hydrologiques. En effet, cette artificialisation des terres réduit considérablement l'infiltration de l'eau dans le sol. Bien que l'exemple puisse paraître anecdotique, on a pu mesurer par exemple dans la bande de Gaza, qu'une augmentation de 1% des surfaces urbanisées réduit de 41% le total d'infiltration (McGrane, 2016). Par ailleurs, les précipitations, qui précédemment s'infiltraient, se transforment en grande partie en ruissellement favorisant ainsi les inondations.

Ces dix dernières années, 98 % des communes wallonnes ont été touchées par des inondations (CEEW, 2007). Les eaux pluviales s'accumulent plus rapidement et en plus grande quantité dans les rivières avoisinant les sols imperméabilisés. Des études démontrent également que l'artificialisation des terres augmente le volume de ruissellement, diminue le temps de concentration du ruissellement, augmente la fréquence de retour des inondations et élève le débit de pointe des événements orageux (McGrane, 2016). Le risque et l'amplitude des inondations augmentant a pour conséquence d'engendrer des dommages économiques et des pertes sociales plus importants. En 1995, l'inondation survenue dans le bassin de la Meuse a causé 25 millions d'euros de coûts (CEEW, 2007). Néanmoins, afin d'atténuer ou de compenser l'imperméabilisation des sols, des démarches préventives sont mises en œuvre tant à l'échelle européenne que wallonne. De plus, il existe des recherches qui tentent de quantifier et prédire les dommages que provoqueraient les inondations (Becker *et al.*, 2013 ; Mustafa *et al.*, 2018 ; Ernst *et al.*, 2010 etc.). « L'estimation des coûts des projets locaux ou généraux et des études pour le district hydrographique de la Meuse est de l'ordre de 65 millions d'euros. » (GTI, 2015).

Des scénarios, basés sur l'évolution de l'occupation du sol, sont estimés jusqu'en 2100 afin de prédire la future urbanisation. Ces scénarios mettent en œuvre plusieurs cas de figure. Dans un premier cas, l'évolution de l'occupation du sol suit la dynamique actuelle, c'est-à-dire un étalement urbain. Dans un second cas, il y a une densification du bâti existant. Afin d'estimer au mieux les évolutions, les scénarios prennent en compte différents paramètres tels que la croissance de la population permettant d'estimer le nombre de ménages futurs (Becker *et al.*, 2013 ; Mustafa *et al.*, 2018).

Suivant chaque scénario d'urbanisation, le risque d'inondation est calculé. Afin de quantifier l'évolution du risque d'inondation, un indicateur, basé sur les 3 périodes de retour (25, 50 et 100 ans), est estimé permettant ainsi d'évaluer les dommages futurs. Pour un taux d'urbanisation moyen et un désir de densifier le bâti existant, l'indicateur du risque d'inondation augmente de 37% s'il n'y a aucune restriction (c'est-à-dire que les nouvelles constructions peuvent s'implanter dans les plaines d'inondation). Par contre, l'indicateur n'augmente que de 2% si les politiques d'aménagement du territoire interdisent la construction dans les plaines d'inondation fortement et moyennement à risque (Mustafa *et al.*, 2018). Suivant les scénarios d'urbanisation, les dommages provoqués par l'inondation, dont la période de retour est de 100 ans, le long de la Meuse, peuvent varier de 334 à 462 millions d'euros, c'est-à-dire augmenter de 1 à 40%. (Becker *et al.*, 2013)

Cependant, les inondations ne provoquent pas que des pertes économiques, mais également sociales. En effet, le risque d'inondation peut être décrit par 3 composantes : l'aléa d'inondation, l'exposition et la vulnérabilité sociale et économique (Ernst *et al.*, 2010 ; Dewals *et al.*, 2011).

Selon la théorie de Coninx et Bachus (2007), la vulnérabilité sociale dépend de l'exposition au risque d'inondation et des caractéristiques socio-économiques de la population. Cela rejoint la théorie plus générale de Cutter *et al.* (2003) qui aborde la vulnérabilité sociale face à l'aléa naturel. Il n'est donc pas sans intérêt de s'interroger sur les individus habitant dans les zones à risque d'inondation.

Dès lors, la question de recherche se pose comme étant : « **Quelles sont les caractéristiques socio-économiques de la population exposée au risque d'inondation ?** »

Dans le chapitre 2, constitué de l'état de l'art, nous nous intéressons aux études précédentes traitant de la manière de caractériser la population sujette à l'aléa naturel ; au risque d'inondation en Wallonie ; et aux méthodes ainsi qu'aux résultats de la combinaison des deux phénomènes.

Le chapitre 3 énonce l'hypothèse de travail, justifie son choix et évoque les limitations des études précédentes.

Le chapitre 4 consiste à rassembler l'essentiel des éléments théoriques relatifs aux méthodes qui seront appliquées dans notre cas d'étude.

Le cinquième chapitre est dédié à l'application et la validation des résultats obtenus par le croisement des caractéristiques socio-économiques de la population et du risque d'inondation au moyen de différentes méthodes statistiques.

Le dernier chapitre regroupe les conclusions partielles émises dans le chapitre 4 et propose une conclusion générale.

CHAPITRE 2: ÉTAT DE L'ART

I. Description des caractéristiques socio-économiques de la population à risque

Dans la littérature, de nombreux termes et concepts apparaissent fréquemment pour essayer de qualifier le statut socio-économique d'un individu. Comme nous le verrons dans un instant, beaucoup de sociologues, notamment, s'étendent sur des termes comme l'impact social, l'inégalité environnementale, la vulnérabilité sociale, etc. De ces notions, provient la confection d'indices qui tâchent de quantifier le bien-être humain, les conditions de vie...

D'une part, le terme d'inégalité au sens de **l'inégalité environnementale** revient de façon récurrente dans la littérature. Ce terme sociologique postule que les inégalités socio-économiques sont renforcées par des facteurs environnementaux comme l'offre et l'accessibilité à des espaces verts ; la qualité des conditions de vie et d'hébergement ; et l'inégalité d'exposition aux sources de pollution et de nuisance (Lejeune *et al.*, 2016). Pour utiliser le terme d'inégalité environnementale, il faut montrer qu'un aspect de l'environnement est distribué de manière inhabituelle à travers différents groupes sociaux. Ces groupes peuvent être différenciés par la classe sociale, l'ethnie, le genre, l'âge, etc. (Walker *et al.*, 2006). À l'origine, le concept de justice environnementale dénonçait la localisation excessive des déchets toxiques dans les communautés pauvres et les minorités ethniques aux États-Unis. Ensuite, le terme d'inégalité environnementale s'est étendu pour qualifier la distribution disproportionnée de la population à l'exposition et l'expérience à l'aléa naturel (Fielding & Burningham, 2005 ; Walker *et al.*, 2006). Cependant, ce concept sociologique est imagé et n'est pas basé sur des données quantifiables.

D'autre part, le terme de **vulnérabilité sociale** est corrélé au terme d'inégalité environnementale. « La vulnérabilité d'un groupe ou d'une personne est déterminée par sa capacité à récupérer, à faire face et à sa sensibilisation au risque naturel » (Walker *et al.*, 2006 ; Cutter *et al.*, 2000). Les personnes plus vulnérables sont inégalement plus exposées au risque naturel et manquent d'accès à un support social, financier ou politique. Ces individus souffrent plus de la perte relative et prennent plus de temps à recouvrer d'un désastre (Maantay *et al.*, 2009). Plusieurs études citent la vulnérabilité comme participant à la détermination du risque naturel. La figure 1 schématise l'approche de Cutter *et al.* (2003) pour décrire les composantes du risque. Le risque (une mesure objective de la probabilité d'un événement d'aléa naturel) interagit avec les limitations (des actions pour atténuer ou réduire le risque) pour produire l'aléa potentiel. L'aléa ou le danger potentiel est modéré ou au contraire est décuplé par la situation géographique du lieu ainsi que par le tissu social. Ce dernier inclut l'expérience de la communauté au danger et son habilité à répondre, faire face, récupérer et s'adapter à ce danger. La vulnérabilité biophysique est mesurée par la fréquence de l'événement et la délimitation de la zone de danger. Quant à la vulnérabilité sociale, elle est mesurée par les caractéristiques socio-économiques (Cutter *et al.*, 2003).

D'autres recherches abordent le terme d'impact social comme étant lié à la vulnérabilité (Coninx & Bachus, 2007 ; Wakler *et al.*, 2006 ; Ersnt *et al.*, 2010 ; Dewals *et al.*, 2011). Contrairement à l'estimation des dommages, l'estimation de l'impact social se concentre sur les personnes comme éléments à risque. L'impact social est évalué suivant trois aspects : (1-2) deux aspects reliés aux personnes et (3) un aspect lié aux caractéristiques de l'événement naturel. Dans le cas des inondations, ce dernier aspect (3) est décrit par la hauteur d'eau, la vitesse d'écoulement et la vitesse de montée des eaux.

Quant aux aspects liés aux personnes, ils correspondent à la vulnérabilité sociale et à la vulnérabilité biophysique des personnes. Ce dernier est déterminé par leur exposition au risque d'inondation (Coninx & Bachus, 2007). La vulnérabilité sociale est, comme on l'a dit plus haut, l'habilité à répondre, faire face, récupérer et s'adapter à l'aléa naturel, ce qui est directement influencé par les caractéristiques socio-économiques des individus et les caractéristiques de leur logement (Cutter *et al.*, 2003).

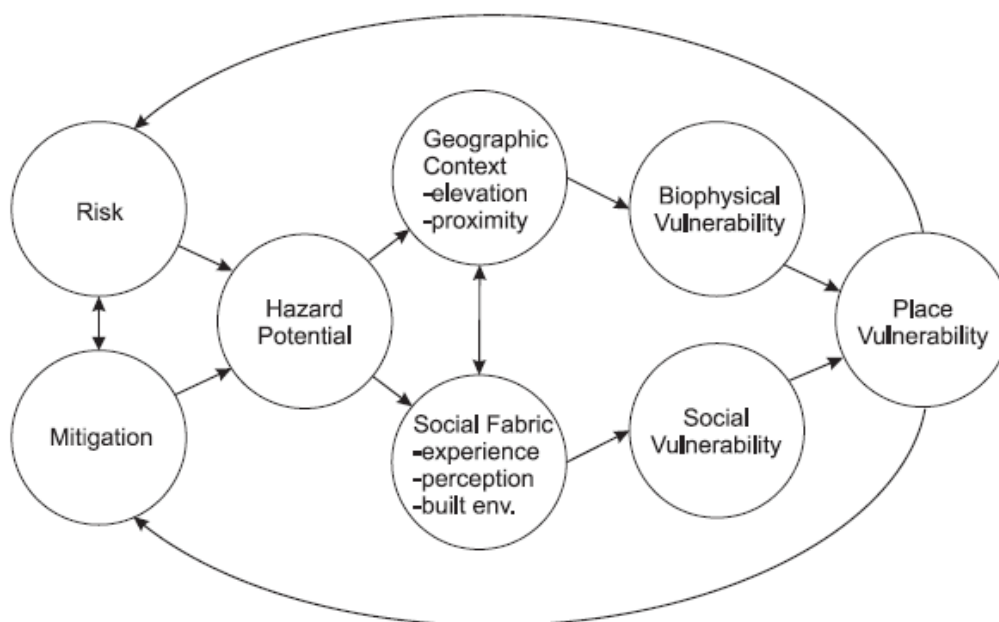


Figure 1: Modèle de vulnérabilité à l'aléa naturel (Cutter *et al.*, 2003)

De nombreux auteurs tentent de définir cette vulnérabilité sociale par des facteurs tels que : le manque d'accès aux ressources (information, savoir et technologie), la limitation de pouvoir politique et de représentation, le capital social (incluant réseaux et connexion sociale), la croyance et les coutumes, l'âge et la disponibilité du bâti, les individus fragiles et physiquement limités et le type et la densité d'infrastructure (Walker *et al.*, 2006 ; Cutter *et al.* 2003). Les chercheurs semblent s'accorder sur ces facteurs, cependant il subsiste un désaccord concernant les variables et les concepts permettant de les décrire. Les concepts utilisés pour définir ces facteurs comme l'âge, le genre, la race et l'ethnie, et le statut socio-économique sont généralement acceptés. Dans le tableau 1 sont illustrés les concepts proposés par Cutter *et al.* (2003) ainsi que la correspondance de ces concepts avec les autres recherches citées. Par ailleurs, selon une autre étude menée par le Centre de recherche de l'aléa d'inondation en Angleterre, les caractéristiques déterminant la vulnérabilité sociale aux inondations sont l'âge, l'état de santé, les revenus et la composition familiale (Coninx & Bachus, 2007).

Pour résumer, la vulnérabilité sociale est un paradigme social qui est défini comme étant l'habilité de la population à récupérer, à faire face et à s'adapter à l'aléa naturel. Elle est décrite par des facteurs aboutissant à des concepts (tableau 1) qui varient selon les considérations des auteurs. Elle n'a pas pour objectif de déterminer quelles sont les populations vulnérables lors d'aléas naturels. Au contraire, elle **postule** que c'est la population décrite par les concepts utilisés qui est la plus susceptible de subir un aléa naturel.

Tableau 1 : Concept de la vulnérabilité sociale (Cutter et al., 2003)

Concept	Articles abordant les concepts
Statut socio-économique (revenu, pouvoir politique, prestige)	Coninx & Bachus, 2007 ; Maantay et al., 2009 ; Fielding, 2012 ; Rasch, 2017
Genre	
Race et ethnité	Maantay et al., 2009; Fielding, 2012
Age	Coninx & Bachus, 2007 ; Maantay et al., 2009 ; Fielding, 2012 ; Rasch, 2017
Développement commercial et industriel	
Perte d'emploi	Maantay et al., 2009 ; Rasch, 2017
Population rurale/urbaine	
Type de construction résidentielle	Rasch, 2017
Infrastructure et subsistance	Maantay et al., 2009 ; Rasch, 2017
Locataire	Rasch, 2017
Profession	
Composition familiale	Coninx & Bachus, 2007 ; Maantay et al., 2009
Education	Maantay et al., 2009 ; Rasch, 2017
Croissance de la population	
Dépendance aux services sociaux	
Service médical (hôpitaux, médecins ...)	Maantay et al., 2009
Population aux besoins spéciaux (infirme, sans-abris...)	Coninx & Bachus, 2007 ; Maantay et al., 2009 ; Fielding, 2012

À partir des concepts présentés dans le tableau 1, des variables jugées pertinentes sont sélectionnées. Ces variables, et par conséquent les indices obtenus, diffèrent suivant les études. Dans le but d'établir un indice de vulnérabilité sociale, Cutter et al. (2003) teste la multi-colinéarité de 250 variables et en conserve 85 variables. À partir de ce résultat, des indices sont calculés et ensuite normalisés obtenant ainsi 42 indices. Par la suite, une analyse en composantes principales est effectuée. 11 composantes sont conservées et décrites. Enfin, l'indice composite de vulnérabilité sociale (SoVI) est créé en combinant le score factoriel de ces 11 composantes.

Coninx & Bachus (2007) développent également un indice composite de vulnérabilité sociale basé sur l'indice élaboré par le Centre de recherche de l'aléa d'inondation en Angleterre, auquel ils ajoutent deux variables. L'indice final se compose de 6 variables : la privation financière, les malades à long terme, les parents isolés, les personnes âgées, les étrangers et les logements non permanents.

L'indice de vulnérabilité à l'aléa d'inondation de la ville de New York (NYCHVI) est dérivé de « l'évaluation de la vulnérabilité humaine » (HVA). Afin de calculer le HVA, quinze indices ont été utilisés et classés selon 4 groupes d'indicateurs : (1) socio-économiques (pourcentage d'individus pauvres, pourcentage d'individus sans emploi, pourcentage de personnes sans diplôme de secondaire...), (2) structure des ménages et invalidité (pourcentage de personnes âgées de 65 ans ou plus...), (3) race et ethnie (pourcentage de minorités...), (4) logement et transport (pourcentage de ménages sans véhicule...). Chaque indice est choisi sur base d'études identifiant les personnes les plus « à risque » lors d'une catastrophe naturelle. Le NYCHVI modifie et ajoute des indices afin de mieux correspondre à la réalité de la ville de New York (Maantay *et al.*, 2009).

Dès lors, il est possible de constituer toute sorte d'indices de vulnérabilité sociale élaborés à partir de variables diverses (Rasch, 2017 ; Fielding, 2012). Il n'existe donc pas un indice unique de la vulnérabilité sociale.

En outre, la population peut être définie autrement que par sa vulnérabilité sociale. Certains auteurs tentent de caractériser la population « à risque » par des indices de **bien-être**, de **pauvreté**, de la **condition de vie**, etc.

L'indice de développement humain (HDI) se calcule par la moyenne géométrique de trois indices : l'espérance de vie, l'éducation et les revenus (UNDP, 2018).

L'indice de privation multiple (IMD) combine des indicateurs appartenant (1) au capital social, (2) au capital humain (pourcentage de personnes instruites, pourcentage de travailleurs...), (3) au capital financier (pourcentage de ménages utilisant des services bancaires...) et (4) capital physique (pourcentage des ménages sans électricité...). Ces indicateurs permettent d'appréhender le concept de pauvreté comme un élément multidimensionnel, une combinaison de plusieurs privations (Baud *et al.*, 2009).

Dans les régions urbaines belges, l'indice synthétique de difficulté comprend 23 indices pouvant être regroupés en quatre dimensions principales : « l'origine », « le revenu », « la précarité sur le marché de l'emploi » et « les ménages précaires et de revenu de transfert » (Grippa *et al.*, 2015).

En conclusion, il existe une multitude de termes, de concepts et de variables se rejoignant parfois, pour tenter de décrire les caractéristiques socio-économiques des individus impactés par le risque naturel en général et le risque d'inondation en particulier.

II. Risque d'inondation en Wallonie: l'aléa d'inondation

Le risque d'inondation en Wallonie est un sujet largement abordé dans la littérature. Afin d'anticiper et de prédire les dommages causés par ce risque, des scénarios sont réalisés (Ernst *et al.*, 2010 ; Dewals *et al.*, 2015 ; Beckers *et al.*, 2015, Mustafa *et al.*, 2018). Ces prédictions nécessitent de disposer de l'étendue de la future plaine d'inondation. Par conséquent, un modèle hydraulique simule, à l'aide d'un modèle numérique de terrain, les débits de crues associés aux périodes de retour afin d'obtenir l'étendue d'inondation, la hauteur d'eau et la vitesse d'écoulement sur des mailles de 5x5m. Le modèle d'écoulements à surface libre WOLF 2D a été développé par l'Université de Liège (Detrembleur *et al.*, 2008 ; Erpicum *et al.*, 2010, etc.). Le modèle a notamment été employé sur 800 km de cours d'eau principaux en Région wallonne afin de calculer la carte de l'aléa d'inondation.

Les cartes relatives aux inondations en Wallonie ont été réalisées pour répondre aux exigences du plan PLUIES (Plan de LUTte contre les Inondations et leurs Effets sur les Sinistrés) adopté par le gouvernement wallon en 2003. Ce plan a légèrement été modifié en 2007 pour correspondre à la directive européenne 2007/60/CE. A la suite d'une enquête publique, les cartes relatives aux inondations ont été revues et approuvées par le gouvernement wallon le 10 mars 2016.

Dès lors, il existe 3 cartographies différentes relatives aux inondations : la carte des risques d'inondation, la carte des zones inondables et la carte de l'aléa d'inondation (GTI, 2016). La carte d'aléa d'inondation est une combinaison des périodes de retour, il n'y a donc qu'une seule carte alors que les autres cartographies possèdent une carte pour chaque occurrence (25, 50, 100 ans et supérieure à 100 ans). La carte d'aléa d'inondation est confectionnée à partir d'un modèle hydraulique, de la méthode hydropédologique, de cartes géologiques et d'observations de terrain. Le modèle hydraulique est obtenu à partir des statistiques hydrologiques disponibles sur la plupart des rivières. Afin de valider et de compléter le modèle, des enquêtes de terrain ont été effectuées (Peeters *et al.*, 2006). L'aléa d'inondation par débordement (figure 2) est déterminé en fonction de la période de retour de l'inondation (25 ans, 50 ans, 100 ans et supérieure à 100 ans) et de la profondeur de submersion. L'aléa d'inondation est une variable ordonnée qui peut prendre la valeur : très faible, faible, moyenne et élevée. L'aléa d'inondation regroupe les inondations par débordement naturel d'un cours d'eau et par le ruissellement des eaux de pluie (GTI, 2016).

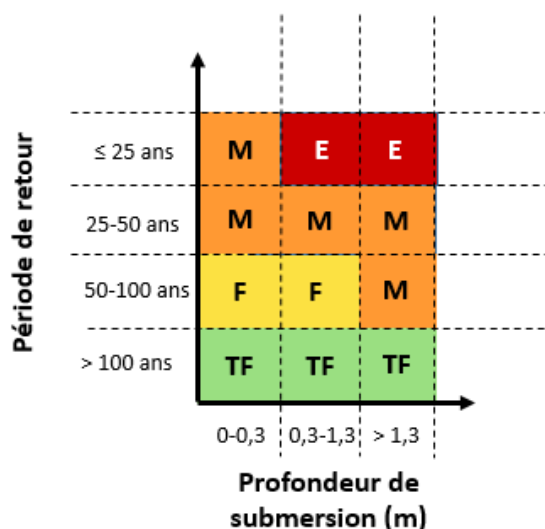


Figure 2: Schéma de détermination de l'aléa inondation par débordement (Beckers *et al.*, 2013; modifiée)

III. Cartographie densimétrique

Les unités administratives délimitées artificiellement, tels que les secteurs statistiques et les communes, représentent l'échelle à laquelle il est possible de disposer des données socio-économiques agrégées. Le concept de « Modifiable Areal Unit Problem » (MAUP) fait référence à un biais dans les traitements statistiques dû à l'effet d'échelle et l'effet de zone des données agrégées. L'effet d'échelle est relié à la taille de l'unité de surface utilisée. L'analyse se fait donc sur les mêmes données, mais à échelle d'agrégation différente (ex. : secteurs statistiques ou communes). Quant à l'effet de zone, il est relié à la forme de l'unité de surface pour une échelle fixe (ex. : grille ou polygone). Il est donc possible d'obtenir des résultats différents suivant l'agrégation des données (Openshaw, 1984). De plus, les données sont définies comme homogènes à l'intérieur du niveau d'agrégation, or les unités administratives sont des entités abstraites surimposées aux distributions réelles des phénomènes géographiques. Pour empêcher ce type de problèmes, les données recensées sont désagrégées à des niveaux géographiques plus pertinents, plus représentatifs de la réalité (l'occupation du sol par exemple). Cette technique est également appelée cartographie densimétrique. Par exemple, la population est redistribuée dans les zones résidentielles. Pour cela, des données auxiliaires sont utilisées afin d'affiner et redistribuer le phénomène étudié. Plusieurs méthodes pour effectuer cette désagrégation sont possibles :

- Wright (1936) propose une désagrégation appelée la « méthode binaire ». Celle-ci consiste à appliquer un masque binaire (habité/non habité) aux données statistiques. Cela permet d'éliminer les zones de densité nulle et donc de redistribuer la population à l'intérieur des surfaces habitées. Présentement, les données géographiques auxiliaires correspondent généralement à des cartes d'occupation du sol obtenues par télédétection.
- Il est aussi possible de désagréger la population suivant plusieurs classes résidentielles avec l'aide d'images satellitaires à de haute résolution et d'un algorithme de classification sophistiqué. Chaque classe résidentielle dispose ainsi de sa propre densité de population (Donnay & Unwin, 2001).
- Plus récemment, des recherches ont permis de désagréger la population du secteur au niveau des bâtiments pris individuellement. La population est redistribuée en fonction de la surface ou du volume du bâtiment. La donnée auxiliaire correspond donc à l'emprise du bâtiment ou à son volume, ainsi qu'à son affectation fonctionnelle (ce qui permet de conserver les bâtiments résidentiels). En Région wallonne, ces informations peuvent être obtenues grâce au PICC (Projet Informatique de Cartographie Continue) (Cornet *et al.*, 2012 ; Ledant, 2009).

Le but de la cartographie densimétrique est également de faire correspondre au mieux deux phénomènes dont les limites ne coïncident pas. En effet, il est rare d'observer des limites identiques de deux phénomènes distincts. Les cartes densimétriques redistribuent la population à une échelle plus fine. Cela permet d'identifier plus aisément les cooccurrences de deux phénomènes. Dans la littérature, il existe des exemples de désagrégation spatiale réalisée afin de comparer la distribution de la population et le risque d'inondation avec une meilleure précision. C'est le cas du « New York Cadastral based Expert Dasymeric System » (CEDS) qui correspond au nombre de ménages ou à la superficie de logements redistribués à l'échelle parcellaire (Maantay *et al.*, 2009). Fielding & Burningham (2005) redistribuent quant à eux la population suivant un quadrillage de 200 x 200m grâce un modèle de population de surface utilisant une méthode d'estimation à noyaux de densité (KDE).

IV. Méthodes et résultats de recherches précédentes

La littérature témoigne de nombreuses recherches portant sur les caractéristiques des personnes « à risque » dans les plaines d'inondation. Comme évoqué précédemment, ces individus sont souvent définis par la vulnérabilité sociale. Celle-ci est d'habitude quantifiée par un indice composé de plusieurs variables socio-économiques qui varient d'une région et d'une recherche à l'autre. Dans les exemples abordés ci-dessous, les régions étudiées sont situées dans des pays en voie de développement tel que Sri Lanka, mais également dans des pays développés tels que l'Angleterre, le Pays de Galles ou encore les États-Unis. Dans ces quelques exemples, nous soulignerons les méthodes de comparaison utilisées (régression, χ^2 , simple superposition de couches...).

Une étude faite en Angleterre et au Pays de Galles n'utilise pas d'indices de vulnérabilité mais met en évidence l'inégalité des distributions de classes sociales dans les zones d'inondation (Fielding & Burningham, 2005). Ces classes sociales sont définies par la profession d'un membre du ménage et sont connues au niveau de « l'Enumeration District » (ED). Les EDs représentent la plus fine échelle de recensement en Angleterre et au Pays de Galles. Un ED correspond à environ 200 ménages. Cette classification est utilisée, car la profession est jugée comme étant liée au revenu, aux perspectives économiques et à la sécurité de l'emploi. Le risque d'inondation est cartographié selon la zone correspondant à une période de retour 100 ans. Deux approches méthodologiques sont envisagées pour étudier les classes sociales à l'intérieur de la plaine d'inondation. La première utilise la carte représentant les plaines d'inondation qu'elle découpe selon les EDs afin d'estimer la distribution des classes sociales par rapport au nombre d'habitants à l'extérieur et à l'intérieur de la zone. La deuxième est une redistribution de la population des EDs selon des cellules de 200 x 200 m à l'aide d'un modèle de populations de surface utilisant une méthode KDE. Pour comparer avec les classes sociales, celles-ci sont redistribuées dans des grilles de même résolution. Ensuite, les auteurs étudient la proportion de populations à l'intérieur et à l'extérieur des plaines d'inondation préalablement rastérisées. Un test de χ^2 est effectué pour déterminer la significativité de la distribution des deux approches. En conclusion, la méthode de quadrillage démontre que les classes sociales les plus précaires sont les plus exposées aux inondations alors que la méthode utilisant les « Enumeration District » indique le contraire. Néanmoins, les auteurs de la recherche affirment que la méthode de quadrillage est plus pertinente, notamment parce que l'approche avec les EDs est une estimation grossière. En effet, il y a une surestimation de la population, car les plaines d'inondation ne coïncident pas avec les limites des EDs et donc, par facilité, tous les EDs intersectés sont repris entièrement comme étant à l'intérieur d'une zone d'inondation. De plus, la méthode par quadrillage prend en compte des carrés non peuplés. Cela permet de ne pas estimer de populations en zone inondée dans une cellule où il n'y pas d'habitants (Fielding & Burningham, 2005). En outre, des recherches plus récentes sur la même région et par le même auteur concluent que les classes ouvrières sont plus exposées au risque d'inondation que les classes moyennes (Fielding, 2012).

Aux États-Unis, un indice de vulnérabilité de la ville de New York (NYCHVI), composé de dix-neuf indices et obtenu au niveau des secteurs de recensement, est utilisé pour étudier la localisation et la proportion populations vulnérables au risque d'inondation. Afin d'estimer la proportion populations à l'intérieur des plaines d'inondation, les chercheurs travaillent en complémentarité avec une carte du nombre de ménages (ou la superficie des logements) désagrégés à l'échelle de la parcelle (CEDS) (Maantay *et al.*, 2009). Le but de cet article est de prouver que l'utilisation du CEDS permet une meilleure estimation de la population mais également de déterminer quel type de variables (le pourcentage de minorités, le pourcentage de personnes de plus de 65 ans etc.) est le plus représenté à l'intérieur des zones d'inondation. Cependant, aucune comparaison n'est faite avec l'extérieur. Les auteurs n'ont pas pour objectif de déterminer si la population dans les plaines d'inondation est significativement différente de la population à l'extérieur (Maantay *et al.*, 2009).

Au Sri Lanka, des enquêtes ont été réalisées auprès de 517 ménages questionnés sur leur condition socio-économique et leur moyen de subsistance ainsi que sur leur exposition à l'inondation de 2016. À partir de ces données, une régression multiple a été effectuée afin d'observer les relations entre la pauvreté, l'exposition à l'inondation et les données démographiques. Il conclut que les ménages souffrant le plus de l'inondation sont les ménages pauvres dépendant de l'agriculture pour subvenir, ayant une éducation limitée et possédant de petites portions de terre. Alors que les ménages les moins affectés sont financièrement les plus stables et représentent des propriétaires de grandes superficies de terre, souvent instruits et ne dépendant pas de l'agriculture pour vivre (Silva & Kawasaki, 2018).

Malgré les spécificités des régions abordées par ces cas d'étude, les conclusions restent cependant semblables : les ménages « pauvres » et les classes sociales faibles sont préférentiellement distribués dans les plaines d'inondation.

CHAPITRE 3: HYPOTHÈSE

Pour rappel, la question de recherche est formulée ainsi : « **Quelles sont les caractéristiques socio-économiques de la population exposée au risque d'inondation ?** »

Au regard de l'état de l'art, on s'aperçoit qu'il existe une multitude d'indices et de variables possibles pour décrire la population. Comme l'a souligné Cutter *et al.* (2003), à partir des concepts du tableau 1, différentes variables sont élaborées suivant l'appréciation de l'auteur. Par exemple, à partir de « l'âge », certains auteurs vont préférer exploiter le pourcentage de la population âgée de plus 65 ans tandis que d'autres vont préférer utiliser le pourcentage de la population âgée de moins de 10 ans. En outre, tous les concepts ne sont pas abordés par chaque indice. Par exemple, certaines recherches prennent en compte des variables témoignant de la « santé » alors que d'autres ne les évoquent guère. Conséquemment, plusieurs indices de vulnérabilité sont déterminés. De surcroît, il est possible de qualifier la population moyennant des indices distincts de la vulnérabilité et selon d'autres classifications (Fielding *et al.*, 2005 ; UNDP, 2018 ; Baud *et al.*, 2009 ; Grippa *et al.*, 2015, etc.). Un des objectifs de cette recherche est donc de simplifier ces différentes approches en utilisant un indice unique.

Si seuls les résultats des recherches antérieures étaient pris en compte, on aboutirait à une conclusion selon laquelle les personnes « vulnérables » sont localisées dans les zones d'inondation. Or cette conclusion rapide n'est pas tout à fait correcte. Premièrement, la Région wallonne n'est pas culturellement ni spatialement analogue aux étendues précédemment étudiées. En effet, le peuplement des vallées en Wallonie est régi par des faits historiques tels que l'industrialisation. La concentration de la population autour des rivières et dans les plaines d'inondation n'est pas un effet du hasard mais est dû, tout du moins en partie, à l'histoire du peuplement (Vandermotten & Vandewattyne, 1985 ; Sporck *et al.*, 1985 ; NCGB, 2012 ; Grimmeau *et al.*, 2015 ; Donnay *et al.*, 1986, etc.) Deuxièmement, les études, comme celle de Maantay *et al.* (2009), postulent que les personnes caractérisées par les variables formées à partir des concepts du tableau 1 se localisent dans les plaines d'inondation. Les auteurs ne cherchent pas à vérifier si les individus dans les zones d'inondation sont plus vulnérables que les individus en dehors.

Par conséquent, l'intérêt de la recherche proposée réside dans le fait de prouver que ce sont effectivement les personnes « vulnérables » qui se localisent en majorité dans les plaines d'inondation. Ce mémoire tentera de démontrer, moyennant des analyses statistiques, que les individus dont les caractéristiques socio-économiques sont jugées « vulnérables » se situent dans les zones d'inondation.

Dès lors, l'hypothèse de ce travail est : « **La population la plus exposée au risque d'inondation en Wallonie possède des caractéristiques socio-économiques faibles.** »

CHAPITRE 4: DÉVELOPPEMENT THÉORIQUE

I. Présentation des données

I. 1. Recensement socio-économique de 2011 et données statistiques complémentaires

Les données statistiques proviennent du recensement effectué par Statbel en 2011 à l'échelle des secteurs statistiques. « Le census [recensement] est une photographie au 1er janvier 2011 de la population belge c'est-à-dire de l'ensemble des habitants du territoire belge, quelle que soit leur nationalité. Le census 2011 fournit un large éventail de chiffres au niveau démographique, socio-économique, de l'enseignement ainsi que par rapport au logement [...] La population prise en compte pour le census 2011 correspond à l'ensemble des personnes domiciliées en Belgique, c'est-à-dire enregistrées au Registre National » (Statbel, 2014).

Le recensement de 2011 a été réalisé au moyen de bases de données administratives. C'est le premier recensement à être établi de cette manière. Les recensements précédents interrogeaient directement la population. Le recensement fournit un grand nombre de variables socio-économiques à l'échelle des secteurs statistiques, mais elles datent de 2011 et le nombre de secteurs statistiques varie selon la feuille de données

Cinq fichiers Excel contenant les données brutes sont accessibles sur le site du Census 2011. Chaque fichier Excel regroupe plusieurs feuilles de calcul dont chacune aborde une thématique différente. Les titres de ces feuilles sont repris à l'annexe 1. Après quelques modifications préalables, telles que la suppression de colonnes redondantes non nécessaires à l'analyse, les feuilles de calcul jugées utiles ont été importées dans une base de données au format Access. Les feuilles conservées sont indiquées dans l'annexe 1.

Outre les données du recensement, les informations purement économiques ont été obtenues sur le site de Stabel et de l'IWEPS. Parmi les fichiers fournis par l'office, la donnée la plus proche temporellement du recensement de 2011 est la statistique fiscale de 2012. Plusieurs variables, telles que les revenus par déclaration, sont proposées dans le fichier Excel de la statistique fiscale (Statbel, 2017c). Il a été décidé de conserver le revenu médian par déclaration fiscale disponible à l'échelle des secteurs statistiques. Cette variable apporte une information sur le statut économique des ménages par secteur statistique, moins sensible aux valeurs extrêmes que le revenu moyen. La même variable, mais calculée par commune et non plus par secteur, a également été reprise à des fins de comparaison. Cette variable est accessible sur le site de l'IWEPS pour l'année 2011 (IWEPS, 2019).

Pour permettre le calcul du taux d'accroissement entre 2011 et 2017, le chiffre de population par secteur statistique en 2017 a été ajouté à notre base de données (Statbel, 2017a).

I. 2. Les limites des secteurs statistiques

La géométrie des secteurs statistiques provient de Statbel au format numérique vectoriel. Le système de référence est le Lambert belge 1972. Le découpage des secteurs statistiques correspond aux actualisations de 2001 et 2011 qui ont permis d'adapter les secteurs à l'accroissement des noyaux d'habitat, aux grands travaux d'infrastructure et à la modification de limites communales. Ce découpage administratif représente l'échelle la plus fine à laquelle il est possible de disposer de données statistiques en Belgique. En 2011, il existe 19.781 secteurs statistiques en Belgique, dont les dimensions varient de 1,124 ha à 5.848,999 ha et dont population moyenne est de 539 habitants (Statbel, 2017c).

I. 3. L'occupation du sol

La carte d'occupation du sol de Wallonie COSW V2.07 est accessible via le Géoportail de Wallonie (Géoportail de Wallonie, 2019b). La couche du niveau 1 représente les terrains artificialisés. Parmi ceux-ci sont distingués les terrains résidentiels, qui sont répertoriés dans la couche du niveau 11. Un exemple de l'emprise des terrains résidentiels est disponible à l'annexe 2.

I. 4. L'aléa d'inondation

La version 2016 de la carte d'aléa d'inondation est disponible sur le Géoportail de Wallonie (Géoportail de Wallonie, 2019a). Ces données sont fournies au format vectoriel, alors que le résultat original de la modélisation de l'aléa d'inondation est obtenu en mode raster. Le format vectoriel est obtenu par vectorisation des pixels de résolution de 5x5 m. Chaque pixel est transformé en un polygone en format vectoriel. Les données contenues dans chaque polygone correspondent aux valeurs de l'aléa et aux types d'inondation, toutes deux qualitatives. Les valeurs de l'aléa sont sur une échelle ordonnée (très faible, faible, moyen et élevée), tandis que les types d'inondation sont sur une échelle nominale (ruissellement, débordement, ruissellement et débordement). Le fichier (format Shapefile) regroupe 622.641 polygones sur l'ensemble de la Région wallonne. À titre d'exemple, l'aléa d'inondation dans le centre de Liège est représenté à l'annexe 3.

II. Prétraitements et analyse descriptive des données

II. 1. Prétraitements et analyse des indices

II.1. i. Sélection et construction des indices

Les indices utilisés (tableau 2) sont des plus classiques et inspirés par la littérature. Néanmoins, les données disponibles à l'échelle des secteurs statistiques ne sont pas optimales. Certaines informations indispensables à l'élaboration d'indices particuliers sont manquantes. En effet, les indices suivants auraient pu être pertinents dans notre analyse, mais les données nécessaires à leur construction ne sont pas disponibles à l'échelle des secteurs statistiques : l'espérance de vie, le taux de fécondité, le taux de mortalité, la part des maisons individuelles, le taux de rénovation, le taux d'allocation GRAPA, le taux de personnes dépendantes du CPAS, le taux d'incapacité de travail...

Les 24 indices utilisés sont présentés dans le tableau 2 avec leur équation. Ils ont été élaborés au départ des données importées dans notre base de données Access. Le passage par une base de données permet de bénéficier des opérateurs SQL (requêtes, transactions, jointures, etc.) particulièrement utiles face à des données hétérogènes (nombre variable de secteurs par feuille Excel originale, etc.) et pour la construction des indices.

II.1. ii. Analyse descriptive des valeurs des indices

Les indices constituent des proportions, ils ne possèdent donc pas d'unité, à l'exception des variables de revenu et de la densité d'habitants. Le tableau en annexe 5 présentent les statistiques descriptives de chacun des indices élaborés, tandis que les histogrammes de ces indices sont disponibles en annexe 4. Suite à leur analyse, la densité de population et la part de noyaux familiaux de plus de 5 personnes, qui présentaient une distribution sensiblement asymétrique à gauche, ont subi une transformation logarithmique.

Tableau 2: Vingt-quatre indices socio-économiques calculés à partir des données du recensement de 2011 (1/3 pages)

Indice créé (Nom simplifié)	Méthode de calcul	Commentaire
Démographie		
Densité d'habitants (DensiteHab) Logarithme de la densité d'habitants (LogDensHab)	$\frac{\text{Population totale}}{\text{Superficie du secteur}}$	Unité : Hab/km ² Unité : log (Hab/km ²)
Part de la population célibataire, veuve et divorcée (PartCelibDiv)	$\frac{P_c + P_v + P_d + P_{pdi} + P_{pde}}{P_T - \text{Non indiqué}}$	P _c = Célibataire, jamais mariée et jamais en partenariat P _v = Veuve, pas remariée, ni nouveau partenariat P _d = Divorcée, pas remariée, ni nouveau partenariat P _{pdi} = Partenariat dissous P _{pde} = Partenariat terminé par le décès d'un des partenaires P _T = Population totale
Part de la population mariée, en couple et en partenariat (PartMarieCoup)	$\frac{P_m + P_{cms} + P_{cso} + P_{pe}}{P_T}$	P _m = Mariée P _{cms} = Couple de même sexe P _{cso} = Couple de sexe opposé P _{pe} = Partenariat enregistré P _T = Population totale
Part de la population étrangère totale (PartEtrangTot)	$\frac{\text{Population étrangère totale}}{\text{Population totale}}$	Etranger = N'étant pas de nationalité belge
Part d'étrangers de l'UE parmi la population étrangère (PartEtrangUE)	$\frac{\text{Population étrangère de l'UE}}{\text{Population étrangère totale}}$	UE = Union Européen
Taux d'accroissement de la population entre 2011 et 2017 (TauxAccr)	$\frac{P_{17} - P_{11}}{P_{11}} \times 100$	P ₁₇ = Population totale en 2017 P ₁₁ = Population totale en 2011
Éducation		
Part de la population disposant d'un diplôme d'enseignement supérieur au sein de la population de plus de 25 ans (PartDipSup)	$\frac{P_{es} + P_d}{\text{Population plus de 25ans}}$	P _{es} = Population disposant d'un diplôme de l'enseignement supérieur P _d = Population disposant un doctorat
Part de la population sans diplôme de secondaire au sein de la population majeure (PartSSDipSec)	$\frac{P_{sd} + P_{dp} + P_{di}}{P_{+18ans} - \text{Non indiqué}}$	P _{ds} = Population sans diplôme P _{dp} = Population avec un diplôme de l'enseignement primaire P _{di} = Population avec un diplôme de l'enseignement secondaire inférieur P _{+18ans} = Population âgée de plus de 18ans

Tableau 2: suite (2/3 pages)

Composition des ménages		
Part des ménages d'une personne (PartMenage1pers)	$\frac{\text{Ménage privé* d'une personne}}{\text{Nombre de ménages total}}$	
Part de la population monoparentale (PartMonoparental)	$\frac{\text{Population monoparentale}}{\text{Population totale}}$	La population monoparentale correspond à un parent seul avec au moins 1 enfant de moins de 25 ans
Part de noyaux familiaux de plus de 5 personnes (PartNoyauFam5pers) Logarithme de la part de noyaux familiaux de plus de 5 personnes (LogPartNoyauFam5pers)	$\frac{N_{f+5}}{\text{Nombre de noyaux familiaux}}$	N_{f+5} = Noyaux familiaux** de plus de 5 personnes Cela correspond à une famille nombreuse
Emploi		
Taux d'activité (TauxActivite)	$\frac{\text{Population active totale}}{\text{Population totale}}$	Population active, correspondant aux personnes susceptibles de travailler, soit les personnes âgées de 15 à 65 ans
Taux de dépendance vieillesse (TauxDepViel)	$\frac{\text{Population âgée de plus de 65ans}}{\text{Population de 20 à 65ans}}$	
Taux de dépendance (TauxDependance)	$\frac{P_{0-20ans} + P_{+65ans}}{\text{Population de 20 à 65ans}}$	$P_{0-20ans}$ = Population entre 0 et 20 ans P_{+65ans} = Population de plus de 65 ans
Part d'indépendants (PartIndependance)	$\frac{\text{Nombre d'indépendants}}{\text{Population active totale}}$	Le nombre d'indépendants correspond au statut professionnel de la population (les autres statuts étant salariés, employeurs et autres)
Taux de chômage (TauxChomage)	$\frac{\text{Nombre de chômeurs}}{\text{Population active totale}}$	Correspondant à la situation sur le marché de l'emploi
Logement		
Part de logements antérieur à 1945 (PartLogAnt45)	$\frac{\text{Logements antérieur à 1945}}{\text{Logement total – Non indiqué}}$	La date correspondant à la période de construction
Part de logements occupés par des locataires (PartLogOccLocataire)	$\frac{\text{Logements occupés par le locataire}}{\text{Logement total}}$	
Part de logements occupés par leur propriétaire (PartLogOccPropri)	$\frac{\text{Logements occupé par le propriétaire}}{\text{Logement total}}$	
Nombre d'habitants par logement (PartHabParLog)	$\frac{\text{Population totale}}{\text{Logement total}}$	
Part de logements ayant plus de 6 pièces (PartLogplus6piece)	$\frac{\text{Logements de plus de 6 pièces}}{\text{Logement total – Non indiqué}}$	Indication sur le confort des logements

Tableau 2: suite (3/3 pages)

Fiscalité		
Revenu médian par déclaration fiscale en 2012 par secteur statistique (RevenuMed)	<i>Revenu médian par secteur statistique</i>	Revenu annuel en € par déclaration fiscale***
Revenu médian par déclaration fiscale en 2011 par commune (RevenuMedCom)	<i>Revenu médian par commune</i>	
Écart relatif entre le revenu médian par secteur statistique et le revenu médian de sa commune (EcartRevenu)	$\frac{RM_{sec} - RM_{com}}{RM_{com}} \times 100$	RM_{sec} = Revenu médian par déclaration fiscale en 2012 par secteur statistique RM_{com} = Revenu médian par déclaration fiscale en 2011 par commune
<p><i>Quelques définitions:</i></p> <p>*Un noyau familial est composé de deux personnes ou plus, qui vivent dans le même ménage et dont les liens sont ceux de mari et d'épouse, de partenaire en partenariat enregistré, de partenaire vivant en union consensuelle, ou de parent et d'enfant (Statbel, 2014).</p> <p>**Un ménage privé regroupe les personnes vivant seules dans un logement et les groupes de plusieurs personnes vivant dans un même logement et pourvoyant ensemble aux besoins essentiels de l'existence (Statbel, 2014).</p> <p>***Type de déclaration fiscale : Une déclaration peut être remplie par une (déclaration individuelle) ou deux personnes (déclaration commune). La règle générale veut que chacun remplisse une déclaration individuelle, à l'exception des personnes mariées, des cohabitants légaux et des personnes à charge. (Statbel, 2017b ; modifié)</p>		

II. 2. Prétraitement de l'aléa d'inondation

Afin d'obtenir une variable binaire à deux modalités (inondé et non inondé), indispensable dans la suite de l'analyse, l'aléa d'inondation et les terrains résidentiels de la carte d'occupation du sol ont été combinés comme expliqué ci-dessous.

Pour rappel, l'aléa d'inondation compte quatre modalités : très faible, faible, moyen et élevé. L'aléa est utilisé pour calculer la proportion de surface inondée dans la seule zone habitée. Selon la valeur de cette proportion, les secteurs statistiques seront définis comme étant inondés ou non inondés. La variable binaire inondé / non inondé sera nommée « statut d'inondation » ou « classe d'inondation » dans la suite de la recherche.

Les explications simplifiées suivantes décrivent la manière dont nous avons obtenu le statut d'inondation (des explications plus détaillées figurent en annexe 6). L'étape 1 décrit comment nous avons obtenu la proportion de surface inondée en zone habitée et l'étape 2 décrit comment nous avons classé ces proportions en deux modalités : inondé et non inondé.

Étape 1 : Calcul de la proportion de surface inondée en zone habitée par secteur statistique

La seule commune non affectée par l'aléa d'inondation est préalablement supprimée (annexe 7).

- i. Sélection des secteurs statistiques contenant une zone habitée. L'annexe 8 reprend l'ensemble des secteurs supprimés par cette sélection et l'annexe 11 est un récapitulatif de la suppression des secteurs durant l'ensemble de l'étude.
- ii. Calcul de la surface habitée par secteur statistique.
- iii. Sélection en parallèle de l'aléa de type « débordement ».
- iv. Reclassification de l'aléa en 2 groupes :



- v. Intersection de la zone habitée des secteurs avec les zones d'inondation des 2 reclassifications de l'aléa. On obtient une zone habitée faiblement inondée et une zone habitée fortement inondée.
- vi. Calcul des superficies habitées en zones fortement inondées, en zones faiblement inondées et en zones non inondées. On obtient, pour chaque secteur, la superficie faiblement inondée en zone habitée, la superficie fortement inondée en zone habitée et la superficie non inondée en zone habitée.
- vii. Calcul de la proportion des 3 caractéristiques d'inondation en normalisant par rapport à la surface habitée. On obtient, pour chaque secteur, la proportion de surface faiblement inondée en zone habitée, la proportion de surface fortement inondée en zone habitée et la proportion de surface non inondée en zone habitée.

Cette première étape est reprise schématiquement dans la figure 3.

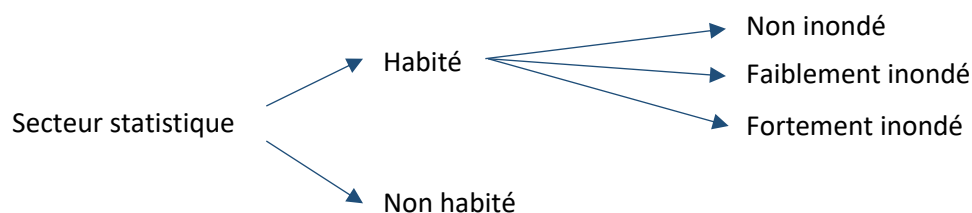


Figure 3: Partitionnement d'un secteur statistique en fonction de deux critères (la zones habitée et l'aléa d'inondation)

Étape 2 : Classement des secteurs en deux types : secteurs inondés et secteurs non inondés

On construit alors un diagramme triangulaire dont les sommets correspondent à 100% de surface fortement inondée en zone habitée, 100% de surface faiblement inondée en zone habitée, 100% de surface non inondée en zone habitée (figure 3 et annexe 9). La position de chaque secteur dans le diagramme est calculée en fonction du pourcentage de sa surface habitée dans chacune des trois catégories.

Dans un premier temps, 4 classes ont été créées (Inondée fortement, Inondée faiblement, Non inondée et Neutre) selon 3 seuils, soient, toujours dans la seule zone habitée : 50% de surface fortement inondée, 50% de surface faiblement inondée et 50% de surface non inondée (cf. annexe 9). Cependant, les secteurs statistiques contenus dans la classe non inondée sont largement majoritaires (7.078 sur 7.512 secteurs), de sorte qu'un seuil unique défini par 80% de la surface non inondée en zone habitée a finalement été retenu. Au-dessus de ce seuil, les secteurs sont associés à la classe non inondée et en-dessous, ils sont associés à la classe inondée. Malgré cette nouvelle classification, le nombre de secteurs dans la classe non inondée reste relativement important (6.121 secteurs). La figure 4 représente la distribution des secteurs dans les deux classes.

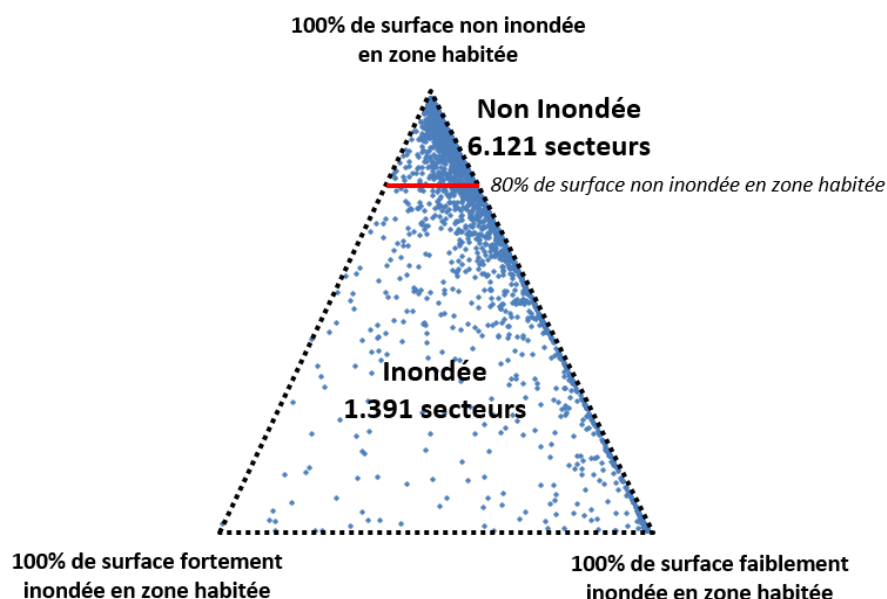


Figure 4: Diagramme des trois proportions de surface inondée en zone habitée avec un seuil de 80% de surface non inondée en zone habitée permettant de classer les secteurs en 2 modalités : inondés et non inondés

II. 3. Prétraitement des secteurs statistiques

Influence de la construction des indices sur le nombre de secteurs statistiques

L'élaboration des indices réduit considérablement le nombre de secteurs statistiques exploitables. En 2011, la Belgique compte **19.781** secteurs statistiques. Dans la feuille comprenant la densité d'habitants par secteur statistique fournie par Statbel, il existe 20.125 secteurs. Mais celle-ci comporte des secteurs dit « non-localisés » dont la densité d'habitants est absente (aucune donnée).

Après avoir construit les 24 indices dans la base de donnée Access, une jointure des tables comportant ces indices dans une table unique a été effectuée. Suite à cette requête, le nombre de secteurs statistiques chute à **16.940**. Cette diminution est due à la suppression des secteurs composés d'au moins une valeur sans donnée et possédant moins de 50 habitants. Premièrement, les valeurs sans données de certains indices proviennent de la différence du nombre de secteurs par feuille (cf. annexe 1). La part de ménages d'une personne, la part de noyaux familiaux et la part de logements de plus de 6 pièces possèdent moins de secteurs dans leurs feuilles Excel donc moins d'entités que les autres tables. Deuxièmement, il existe 2.073 secteurs pour lesquels le revenu médian n'est pas publié car ils possèdent moins de 20 déclarations fiscales. Enfin, des données manquantes correspondent aux secteurs pour lesquels le dénominateur des indices calculés est nul. Le taux de chômage et la part d'indépendants sont indéterminés lorsqu'il n'y a pas de population active. La part d'étrangers de l'UE parmi les étrangers est indéterminée lorsqu'il n'y a pas de population de nationalité étrangère dans le secteur.

Conservation des secteurs en Wallonie

Puisque la région étudiée est la Wallonie, seuls les secteurs statistiques de cette région sont sélectionnés (**7.895** secteurs).

Analyse des valeurs aberrantes de l'ensemble des données

L'histogramme (annexe 4) de chaque variable est construit et analysé enfin d'éliminer les éventuelles valeurs erratiques. L'annexe 10 reprend les quelques valeurs aberrantes détectées.

Suppression préalable des secteurs pour rendre possible la comparaison entre les traitements

Dans la suite de la recherche, nous calculons l'autocorrélation spatiale des secteurs statistiques. Pour ce calcul, une matrice de voisinage est nécessaire (cf. infra III.6.ii). Lors de la création de cette matrice, des « îles », qui correspondent à des secteurs sans voisins, sont supprimées du calcul. Étant donné la quantité de secteurs statistiques précédemment éliminés, de nombreuses îles ont été produites (350 îles lors du calcul de la matrice de voisinage des 7.895 secteurs). Autrement dit, un grand nombre de secteurs sont entourés de secteurs sans données. Dès lors, afin de pouvoir comparer les traitements qui ne font pas intervenir l'autocorrélation spatiale avec les traitements qui prennent en compte l'autocorrélation spatiale, il est nécessaire de supprimer ces 350 secteurs.

Suppression des secteurs dont la commune n'est pas intersectée par l'aléa d'inondation

Lors du calcul des proportions de surface inondée en zone habitée (cf. § II. 2.), la seule commune qui n'est pas intersectée par l'aléa d'inondation est supprimée au début du traitement. Il s'agit de la commune de Faimés qui comptent 11 secteurs statistiques (annexe 7).

Suppression des secteurs en zone non habitée

Lors du calcul des proportions de surface inondée en zone habitée (cf. § II. 2.), les secteurs qui ne contiennent pas de terrains résidentiels sont supprimés de l'analyse (annexe 8).

Un tableau récapitulatif de l'ensemble des prétraitements et du nombre de secteurs éliminés par ceux-ci est disponible à l'annexe 11.

III. Présentation des traitements

Les parties théoriques des traitements effectués dans ce mémoire sont présentées dans de nombreux ouvrages de statistique (Dagnelie, 1973 ; Dagnelie, 1977 ; Snedecor & Cochran, 1980 ; Dixon & Massey, 1983, etc.) et d'analyse spatiale (De Smith *et al.*, 2018 ; Caloz & Collet, 2011 ; la collection CATMOG [QMRG, 2011] ; Hammond & Mccullagh, 1978 ; Griffith & Chun, 2013 ; Yeates, 1974 ; Zaninetti, 2005 ; Bailey & Gatrell, 1995, etc.), ainsi que dans des notes de cours (Donnay, 2015, 2019a, 2019b ; Timmermans, 2015 ; Reboul, 2015, etc.) et les manuels de logiciels spécialisés (Anselin & Rey, 2014 ; Bivand *et al.*, 2013 ; Dorman, 2014 ; Pebesma, E. & Bivand, 2019, etc.). Les notations pouvant changer d'un auteur à l'autre, nous reprenons ici le minimum requis pour permettre la compréhension de l'application qui suivra.

III. 1. Analyse en composantes principales

Afin de considérer un indice unique et synthétique à partir des indices préalablement sélectionnés, une analyse en composantes principales (ACP) a été effectuée. Les indices calculés précédemment correspondent aux variables originelles que l'on introduit dans l'ACP.

L'ACP est une analyse factorielle, c'est-à-dire une analyse multivariée qui crée de nouvelles variables à partir de variables corrélées afin de diminuer le nombre de variables originelles, d'obtenir des variables non corrélées et synthétiques, tout en conservant un maximum l'information contenue dans les données initiales (Rummel, 1970). L'analyse en composantes principales traite des variables quantitatives. Les nouvelles variables correspondent à une combinaison linéaire des variables originelles, et elles sont généralement appelées facteurs, axes, dimensions ou composantes.

Dans une matrice des variances-covariances, l'information totale ou variance totale correspond à la somme des variances de chacune des variables initiales donc à la trace principale de la matrice. Dans une **matrice de corrélation**, les variables sont standardisées, et le poids de chacune est donc identique (variances égales à l'unité). La somme des variances, donc l'information totale, correspond au nombre de variables initiales.

Les facteurs créés suite à l'ACP sont caractérisés par leur corrélation avec les variables originelles. Ces corrélations sont représentées dans la **matrice des saturations**, formée de l'ensemble des vecteurs propres issus de la matrice d'inertie.

La **part de variance expliquée** par chaque facteur correspond à la valeur propre du vecteur propre correspondant, soit la somme en colonne des carrés de la matrice des saturations. Si toutes les composantes sont extraites, c'est-à-dire si l'on conserve le même nombre de composantes que de variables originelles, le total de variance expliquée est de 100%.

La valeur de la **communauté** d'une variable pour un nombre de composantes extraites permet de déterminer dans quelle mesure cette variable est bien représentée dans l'espace des composantes. La communauté correspond à la somme en ligne des carrés de la matrice des saturations. Si toutes les composantes sont extraites la communauté de chaque variable vaut 1.

La corrélation de Bravais-Pearson, entre une variable originelle i et une variable originelle j est déterminée par l'équation suivante :

$$r_{v_i v_j} = \frac{\text{cov}(v_i, v_j)}{s_{v_i} s_{v_j}} = \frac{\sum (v_i - \bar{v}_i)(v_j - \bar{v}_j)}{n s_{v_i} s_{v_j}}$$

Où s = l'écart type

La corrélation entre les composantes et les variables originelles dans la matrice des saturations utilise la même formule en remplaçant la variable j (v_j) par la composante j (c_j).

III.1. i. Choix des composantes à conserver

La sélection du nombre de facteurs à conserver (composantes « principales ») est basée sur la part de variance expliquée, les communautés et éventuellement l'interprétation de la signification des composantes extraites de l'ACP. Les seuils retenus en pratique sont empiriques. À titre d'exemple, les composantes dont les valeurs propres sont supérieures à 1 sont souvent conservées (car elles « expliquent » au moins autant qu'une variable originale). Face à un grand nombre de variables originelles, il est également courant de conserver les composantes dont la variance expliquée relative atteint un seuil minimal (par exemple 5 %).

III.1. ii. Score factoriel

Pour chaque individu (dans notre analyse, chaque secteur statistique), il est possible de calculer un score (ou une cote ou une note) factoriel standardisé sur chaque composante selon la formule suivante :

$$\text{Score factoriel sur la composante } j = \sum_{i=1}^n a_{i,j} x_i$$

n = Nombre de variables originelles ; j = Composante principale ; i = Variable originelle

$a_{i,j} = S_{j,i}/vp_j$ (vecteur propre) ; $S_{j,i}$ = Saturation entre la composante j et la variable i

vp_j = Valeur propre de la composante j ; x_i = Valeur standardisée du secteur sur la variable i

III. 2. Analyse de contingence par χ^2 (χ^2) ou test d'indépendance

Le test du χ^2 s'applique à des variables qualitatives catégorielles, nominales ou ordonnées. L'analyse de contingence consiste à tester statistiquement l'association entre deux variables qualitatives. La table de contingence construite pour ce test croise les modalités des deux variables (ou plus) pour obtenir les fréquences observées dans chaque cellule de la table. Parallèlement, une table des fréquences attendues (ou théoriques) est construite en multipliant la somme en ligne i des fréquences observées avec la somme en colonne j , et en divisant ce produit par le nombre total d'observations, soit :

$$\text{La fréquence attendue : } f_{ij} = \frac{(\Sigma_i \times \Sigma_j)}{n}$$

Où Σ_i = Somme de la ligne i ; Σ_j = Somme de la colonne j ; n = nombre total d'observations

La valeur de la statistique χ^2 est obtenue en sommant les rapports du carré des écarts entre fréquence observée et fréquence attendue, à la fréquence attendue, sur toutes les cellules de la table. Elle est comparée à la valeur théorique de la distribution de la variable χ^2 au degré de liberté correspondant (nombre de ligne – 1, nombre de colonnes – 1) et au seuil de signification α accepté. L'hypothèse d'une association entre les deux variables comparées (H_1) est acceptée si la valeur de la statistique χ^2 calculée dépasse la valeur théorique.

$$\text{La statistique } \chi^2 : \chi^2 = \sum \sum \frac{(O_{i,j} - f_{i,j})^2}{f_{i,j}}$$

Où $O_{i,j}$ = Fréquence observée de la cellule i, j ; $f_{i,j}$ = Fréquence attendue de la cellule i, j

Pour un seuil de signification α , si la différence entre le χ^2 théorique et le χ^2 observé est en dessous du seuil, l'écart des distributions n'est pas du hasard dès lors l'hypothèse nulle est rejetée. Les deux variables sont associées. La différence entre les 2 χ^2 est appelée la p-valeur.

Les conditions nécessaires à l'utilisation du test :

- Les observations sont sous forme de fréquences absolues par catégorie.
- Le nombre total d'observations (n) doit être supérieur à 20.
- Le nombre d'observations attendues par catégorie doit être supérieur à 5.

III. 3. ANOVA

L'ANOVA permet de comparer les moyennes d'une variable Y entre plusieurs groupes. La ou les variable(s) explicative(s) X est/sont qualitative(s) et présente(nt) plusieurs modalités, tandis que la variable dépendante ou « réponse » Y est une variable quantitative. L'objectif du test est de détecter si un groupe de la variable X a un effet significatif sur la valeur de la moyenne de la réponse (Y).

Les conditions d'utilisation qui doivent être vérifiées sont les suivantes :

- Homogénéité des variances :
 - Dans chaque groupe, on effectue une analyse de variance sur les écarts absolus entre les valeurs observées de la variable dépendante Y et la moyenne de ces valeurs dans le groupe construit sur la variable X. Le test vérifie que la variance de chacun des groupes est égale. Si la p-valeur est $> 0,05$, on accepte l'hypothèse d'homogénéité.
- Normalité des résidus :
 - Teste si les variables présentent une distribution normale.
 - Pour ce faire, le test de Kolmogorov-Smirnov compare la distribution des résidus avec une distribution selon une loi normale. Si la p-valeur est $> 0,05$, on accepte l'hypothèse de normalité.

III.3. i. ANOVA à un facteur de classification : Une seule variable explicative X

L'ANOVA à un facteur prend en compte 1 variable dépendante quantitative (Y) et 1 variable indépendante nominale (X) caractérisée par plusieurs modalités.

Le test consiste à vérifier si les moyennes de Y des différentes modalités de X s'écartent significativement de la moyenne générale.

Les hypothèses sont les suivantes :

- H_0 : L'écart entre les moyennes de Y pour chaque modalité de X et la moyenne générale de Y est égal à 0.
- H_1 : Il existe au moins une moyenne de Y d'une modalité α du X qui diffère significativement de la moyenne générale.

Le test consiste à comparer la variance intergroupe à la variance intragroupe à l'aide d'un test de Fisher-Snedecor. Les définitions de la variance intergroupe, de la variance intragroupe et de la statistique de Fisher sont disponibles dans le tableau 3 pour un nombre total d'observations n et un nombre de groupes k.

Tableau 3: Explication de la variance totale, intergroupe, intragroupe et de la statistique de Fisher (Timmermans, 2017)

Variance totale, intragroupe, intergroupe et statistique de Fisher		
Totale	$\frac{SCT}{n - 1}$	Estimateur de la variance globale (variabilité autour de la moyenne globale)
Intragroupe (erreur)	$\frac{SCR}{n - k}$	Estimateur de la variance résiduelle (variabilité individuelle inexpliquée autour de la moyenne de groupe)
Intergroupe	$\frac{SCF}{k - 1}$	Estimateur de la variance factorielle (variabilité autour de la moyenne expliquée par l'appartenance au groupe X)
Statistique de Fischer-Snedecor	$F_{obs} = \frac{SCF/k - 1}{SCR/n - k}$	Le rapport de la variance expliquée (venant de la variation au niveau de la variable explicative X) à la variance inexpliquée (variance résiduelle à l'intérieur des groupes)

La variance intragroupe provient de la somme des carrés des écarts entre la valeur de Y pour un individu i et la moyenne de Y du groupe X (SCR). Cet écart est appelé écart résiduel. La variance intergroupe provient de la somme des carrés des écarts entre la moyenne de Y du groupe X et la moyenne générale de Y (SCF). Cet écart est appelé écart factoriel. La variance totale provient de la somme des carrés des écarts entre la valeur de Y pour un individu i et la moyenne générale de Y (SCT). Cet écart est appelé écart totale. La relation entre ces sommes des carrés représente la formule classique de décomposition de la variance, reprise ci-dessous et dans la figure 5.

L'équation de la décomposition de la variance correspond à :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \bar{Y}_X)^2 + \sum_{i=1}^n (\bar{Y}_X - \bar{Y})^2$$

← SCT →
← SCR →
← SCF →

Y_i = Valeur de Y pour l'individu i (ex: score factoriel) ; \bar{Y} = Moyenne générale de Y
 \bar{Y}_X = Moyenne du groupe X ; n = Nombre total d'observations ; i = Individu (ex: secteur statistique)

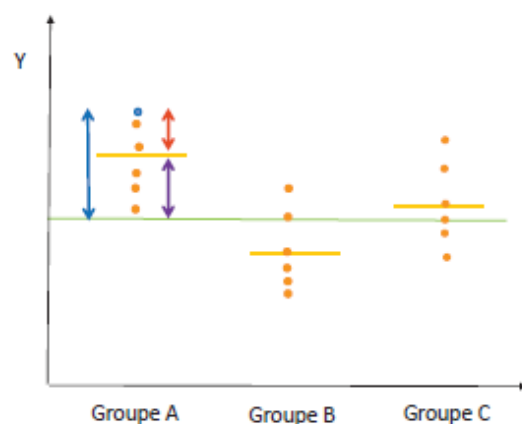


Figure 5 : Observations réparties en 3 groupes avec la moyenne des groupes (en jaune) et la moyenne générale (en vert). L'écart total (en bleu), l'écart résiduel (en rouge) et l'écart factoriel (en mauve)(Timmermans, 2017)

III.3. ii. ANOVA à deux facteurs de classification : deux variables explicatives X1 et X2

L'ANOVA à deux facteurs prend en compte 1 variable dépendante quantitative (Y) et 2 variables indépendantes nominales (X1 et X2) caractérisées chacune par plusieurs modalités. Le but est d'étudier si le comportement de la variable dépendante Y est influencé par les variables X1, X2 et éventuellement par leur interaction.

Étant donné que nous ne testons plus un mais deux facteurs, les hypothèses sont dédoublées, voir triplées si on prend en compte l'interaction des deux facteurs. Le tableau 4 reprend les équations de la statistique de Fisher, de la somme des carrés des écarts et de la variance intragroupe, intergroupe et totale. Ce tableau ne prend pas en compte l'interaction et il fait l'hypothèse que le nombre d'observations par modalité est égal.

Dès lors, les hypothèses à tester sont déterminées comme suit :

- Hypothèse en ligne (pour le facteur X1) :
 - H0 : L'écart entre les moyennes de Y pour chaque modalité en ligne et la moyenne générale (\bar{y}) est égal à 0
 - H1 : Il existe au moins une moyenne en ligne d'une modalité « l » qui diffère significativement de la moyenne générale
- Hypothèse en colonne (pour le facteur X2):
 - H0 : L'écart entre les moyennes de Y pour chaque modalité en colonne et la moyenne générale est égal à 0
 - H1 : Il existe une moyenne en colonne d'une modalité « c » qui diffère significativement de la moyenne générale

Dans le cas où l'interaction des facteurs X1 et X2 doit être prise en compte, une hypothèse complémentaire s'ajoute :

- Hypothèse d'interaction (Combinaison des facteurs X1 et X2):
 - H0 : L'écart entre les moyennes de Y pour chaque paire de modalité de X1 et X2 et la moyenne générale est égal à 0
 - H1 : Il existe au moins une paire de modalités X1 et X2 pour laquelle la moyenne diffère significativement de la moyenne générale

La décomposition de la variance s'écrit cette fois :

$$SCT = SCR + SCF_l + SCF_c [+SCF_{lc}]$$

SCF_l = Somme des carré des écarts des intergroupes en ligne (Facteur 1)

SCF_c = Somme des carré des écarts des intergroupes en colonne (Facteur 2)

SCF_{lc} = Somme des carré des écarts des intergroupes (Interaction des deux facteurs)

III.3. iv. Non-respect des conditions d'utilisation de l'ANOVA

Dans le cas où les hypothèses ne sont pas respectées, il est possible d'utiliser des tests sensiblement équivalents mais non paramétriques (Siegel, 1956).

Le premier test est équivalent à l'ANOVA à 1 facteur excepté qu'au lieu de jouer avec des moyennes de Y, il s'intéresse aux rangs (numéros des valeurs classées) et à la médiane de Y. Il s'agit du test de Kruskal-Wallis qui vérifie si les k groupes de X considérés comme des échantillons font partie de la même population Y. Pour ce faire, il teste l'hypothèse nulle définie comme suit : la médiane des groupes de X est égale à la médiane de la population Y. Si au moins une médiane est différente, l'hypothèse nulle est rejetée. La p-valeur est obtenue par une loi χ^2 dès que le nombre d'individus est supérieur à 5.

Les tests de Wilcoxon et de Mann-Whitney sont équivalents au test de Student. Ils réalisent une comparaison des sommes ou des différences des numéros d'ordre de Y selon les groupes de X et test la significativité de la différence de leur médiane.

En général, les tests paramétriques comme l'ANOVA sont plus fiables car, si les tests non-paramétriques sont plus robustes, ils ont tendance à rejeter à tort l'hypothèse nulle.

III. 4. Régressions classiques

III.4. i. Régression linéaire simple

La régression linéaire simple tente d'expliquer la variance de Y par la variable X. L'écart entre la valeur prédite par le modèle linéaire et la valeur observée correspond à l'erreur d'ajustement (le résidu).

$$Y = a + bX + \varepsilon$$

Y = Variable aléatoire à expliquer, réponse ou variable dépendante

X = Variable indépendante, explicative ou régresseur

a = Ordonnée à l'origine ou terme indépendant

b = Pente de régression, coefficient angulaire de la droite ou coefficient de régression

ε = Résidu

La variance totale du modèle est donc décomposée comme suit :

$$SCT = SCR + SCF$$

SCT = Variance totale, la variation de la variable dépendante (Y)

SCF = Variance expliquée par le modèle, donc la variance expliquée par la régression

SCR = Variance des résidus, donc la variation résiduelle

Afin de déterminer la qualité de l'ajustement, un **coefficient de détermination** R^2 est utilisé. Celui-ci est calculé par le rapport entre la variance expliquée par le modèle et la variance totale du modèle :

$$R^2 = \frac{SCF}{SCT}$$

R^2 varie entre 0 et 1. Il exprime, en %, la part de la variance de Y expliquée par le modèle.

Plusieurs conditions sont nécessaires à l'utilisation d'une régression simple (ou multiple). Celles-ci comportent des hypothèses à vérifier *a priori* telles que la corrélation entre les variables X et Y, l'explication de la variance Y par la variance de X doit être plausible et le coefficient de régression doit être une constante afin de déterminer si la relation entre Y et X est linéaire. Ensuite, ces conditions comportent des hypothèses à vérifier *a posteriori* après l'estimation des paramètres du modèle. Celles-ci sont la normalité et l'homoscédasticité des résidus et l'absence de corrélation des résidus (Ferguson, 1977 ; Donnay, 2019a).

La signification statistique du coefficient de détermination R^2 est déterminée par un test de Fisher.

La statistique de Fisher, dont l'équation est présentée ci-dessous, teste si la variance expliquée par le modèle est significativement supérieure à la variance résiduelle du modèle. Si la p-valeur est inférieure au seuil de confiance α , l'hypothèse nulle est rejetée. La variance expliquée sera donc significativement supérieure à la variance résiduelle du modèle.

$$F = (n - 2) \frac{SCF}{SCR}$$

n = Nombre total d'observation

$(n - 2)$ = Degré de liberté

La signification statistique du coefficient de régression et de l'ordonnée à l'origine est donnée par un test de conformité de Student. Dans le cas du coefficient de régression, les hypothèses testées sont :

- H_0 : b est égal à 0
- H_1 : b est significativement différent de 0

Dans le cas de l'ordonnée à l'origine, les hypothèses testées sont:

- H_0 : a est égal à 0
- H_1 : a est significativement différent de 0

Dès lors, dans les deux cas, si la p-valeur résultante du test est supérieure au seuil de confiance de 0,05, l'hypothèse nulle est rejetée, le coefficient de régression et/ou l'ordonnée à l'origine sont significativement différents de 0.

III.4. ii. Régression linéaire multiple

$$Y = a + b_1X_1 + \dots + b_kX_k + \varepsilon$$

n = Nombre d'observations ; k = Nombre de paramètres ; a = Ordonnée à l'origine

b_i = Coefficient de régression partielle de la variable explicative i sur Y

ε = Résidu ; $(n - 1 - k)$ = Degré de liberté

La régression linéaire multiple tente d'expliquer la variance de Y par les variables X_i . L'écart entre la valeur prédite par le modèle linéaire et la valeur observée correspond à l'erreur d'ajustement (résidu).

La qualité de l'ajustement de la régression est à présent déterminée par le **coefficient de détermination multiple**. $R^2 = \frac{SCF}{SCT}$ exprime la part de variance de la variable dépendante Y expliquée par l'ensemble des régresseurs (X_k).

On utilise le plus souvent le **R^2 ajusté**. Il peut diminuer avec l'ajout d'une nouvelle variable explicative. Le R^2 ajusté permet de prendre en considération le critère de parcimonie dans l'élaboration du modèle. On essaie de maximiser le R^2 pour un minimum de variables explicatives. Son équation est donnée par :

$$R'^2 = \frac{n-1}{(n-k-1)} (1 - R^2).$$

Les conditions nécessaires à l'utilisation de la régression multiple sont semblables aux conditions de la régression linéaire simple.

La signification statistique du coefficient de détermination multiple est déterminée par un test de Fisher dont l'équation est rappelée ci-dessous. Ce dernier correspond à un test dit **global** pour $(k ; n-k-1)$ degré de liberté. Dès lors, avec ce test de Fisher, on teste globalement la significativité du R^2 en comparant la part de variance expliquée par tous les régresseurs à la variance résiduelle, chacune rapportée à son degré de liberté. La statistique de Fisher est donnée par :

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

Il est possible de tester l'**utilité** de chaque régresseur (X) individuellement grâce à la corrélation partielle de chacun de ceux-ci entretient avec la variable dépendante (Y). Sa signification statistique est mesurée par un test de Student à (n-k-1) degrés de liberté. La statistique de Student est donnée par :

$$t = r \sqrt{\frac{n - k - 1}{1 - r^2}}$$

Si, par exemple, on ne dispose que de deux régresseurs x_1 et x_2 , la corrélation partielle de x_1 par rapport à y est donnée par :

$$r_{yx_1.x_2} = \frac{r_{yx_1} - (r_{yx_2} r_{x_1x_2})}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}$$

Le test de signification statistique de l'ordonnée à l'origine est la même que celui dans la régression linéaire simple.

III.4. iii. Ajout d'une variable auxiliaire

Dans une régression simple ou multiple, il est possible d'ajouter une variable qualitative souvent binaire, par exemple X_2 . Il en résulte deux droites de régression avec la même pente, mais dont l'ordonnée à l'origine diffère.

$$Y = a + b_1X_1 + b_2X_2 + \varepsilon$$

Si $X_2 = 0$ on obtient : $E(Y) = a + b_1X_1$. L'ordonnée à l'origine prend donc seulement la valeur de la constante de base.

Si $X_2 = 1$ on obtient : $E(Y) = (a + b_2) + b_1X_1$. L'ordonnée à l'origine de la droite correspond à l'addition de la valeur de la constante (a) et de la valeur du coefficient de la variable auxiliaire (b_2).

Dans le cas d'une variable auxiliaire catégorielle à plusieurs modalités, le principe est le même. Chaque individu possède p modalités d'une variable X_2 . La variable qualitative est introduite dans le modèle de régression avec p-1 modalités. La modalité non prise en considération correspond à la modalité de référence. Si la valeur de la variable X_2 ne correspond pas à une des p-1 modalités, cela signifie qu'elle correspond à la modalité de référence. Souvent chaque modalité p-1 a pour valeur 1 ou 0. Par conséquent, si une des modalités est égale à 1 le reste des modalités valent 0. Lorsque toutes les modalités sont à 0, cela correspond à la modalité de référence.

Un test conformité de Student permet de mesurer la signification statistique de l'écart entre les 2 ordonnées à l'origine.

III.4. iv. Introduction du concept d'interaction des variables explicatives

Il existe un effet d'interaction lorsque l'effet d'un facteur dépend de la modalité de l'autre facteur. Dans une régression linéaire simple et multiple, l'interaction signifie que la ou les variables explicatives (X) ne sont pas totalement indépendantes les unes des autres. Il est possible de prendre en compte l'interaction entre ces variables explicatives dans les régressions. Prenons l'exemple suivant d'une régression multiple, où X_1 et X_2 sont deux variables explicatives. Leur interaction est mise en évidence par le terme dX_1X_2 .

$$Y = a + bX_1 + cX_2 + dX_1X_2 + \varepsilon$$

Le test de conformité de Student fonctionne également pour tester la significativité de la valeur du coefficient de régression d qui tient compte de l'interaction des variables X_1 et X_2 .

III. 5. Régression Logit

III.5. i. Rapport de cote (Odd ratio)

Le rapport de cote de y est égal à 1 correspond au rapport de la probabilité que y soit égal à 1 sur la probabilité que y soit différent de 1. L'équation étant formulée comme suit :

$$Cote(y = 1) = \frac{P(y = 1)}{[1 - P(y = 1)]}$$

Le ratio de cote variant de 0 à l'infini positif, le logarithme est appliqué afin d'obtenir un intervalle variant de l'infini négatif à l'infini positif.

On obtient ainsi le **Logit de y** : $\ln\left(\frac{P(y=1)}{1-P(y=1)}\right)$

III.5. ii. Modèle de régression Logit

Les variables explicatives (x) peuvent être quantitatives ou qualitatives, avec ou sans interactions. Le **Logit de y** peut être estimé par l'équation linéaire suivante :

$$\ln\left(\frac{P(y = 1)}{1 - P(y = 1)}\right) = b_0 + b_1x_1 + \dots + b_kx_k$$

Par transformation, on obtient :

$$P(y = 1) = \frac{e^{b_0 + b_1x_1 + \dots + b_kx_k}}{1 + e^{b_0 + b_1x_1 + \dots + b_kx_k}}$$

La fonction Logit correspond à la réciproque d'une fonction sigmoïde, représentante des fonctions logistiques.

III.5. iii. Interprétation des coefficients de la régression

L'interprétation des coefficients permet de mesurer l'association entre X et Y . Rappelons que les odds ratio correspondent à l'exponentielle des coefficients. Dans un cas simplifié, où l'on ne dispose que d'une seule variable explicative (x_1) et pas d'ordonnée à l'origine, l'équation de régression se ramène à :

$$P(y = 1) = \frac{e^{b_1x_1}}{1 + e^{b_1x_1}}$$

Comme le montre la figure 6 :

- Si $b_1 = 0$: le coefficient est non significatif
- Si $b_1 < 0.5$: donc s'il est petit, une large plage de valeurs de x_1 est ambigu   (mod  le peu discriminant)
- Si $b_1 > 1$: le coefficient devient significatif

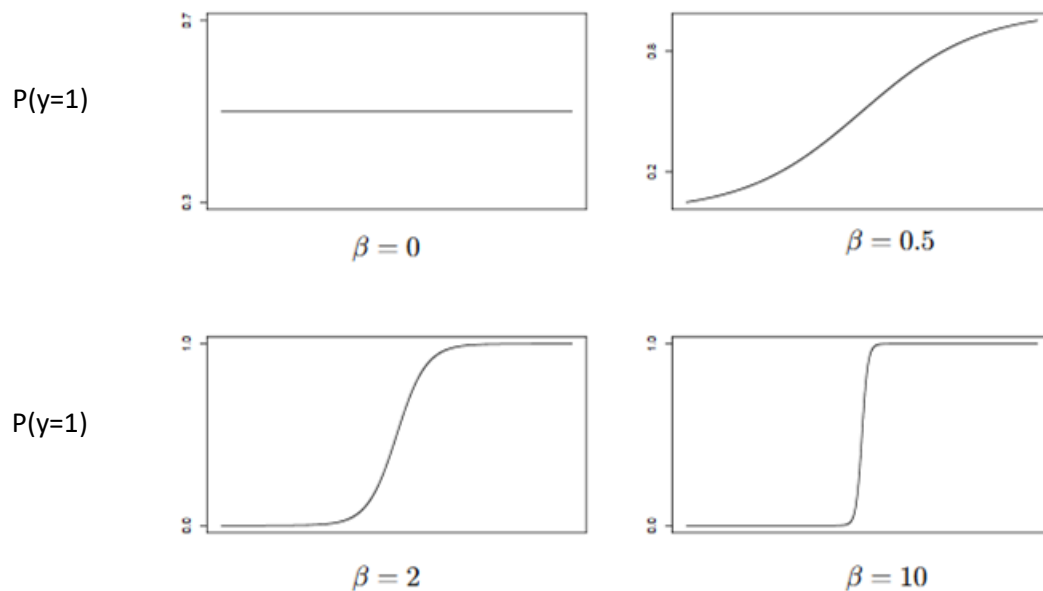


Figure 6 : Fonction logit, proche d'une fonction sigmo  de exemple de pour diff  rente valeur du coefficient de la r  gression (Donnay, 2019b)

III.5. iv. Estimation des coefficients par maximum de vraisemblance

La probabilit   qu'un individu i prenne la valeur $y_i = 1$ est :

$$P_i = \frac{n_1}{N}$$

n_1 = Nombre d'occurrences pour lesquelles y   gal 1 ; N = Nombre total d'observations binaires

La **fonction de vraisemblance L** pour l'ensemble des N individus est donn  e par :

$$L = \prod_i P_i^{y_i} (1 - P_i)^{(1-y_i)}$$

Ou sous forme logarithmique :

$$\ln L = \sum_{i=1}^N [y_i(b_0 + b_1 x_{i1} + \dots + b_k x_{ik}) - \ln(1 - e^{(b_0 + b_1 x_{i1} + \dots + b_k x_{ik})})]$$

La fonction suit une loi de Bernouilli o   y_i est la valeur observ  e 1 ou 0 et les coefficients sont extraits par maximum de vraisemblance (Wrigley, 1976 ; Donnay, 2019b).

Le test de conformité des coefficients de la régression n'est plus un test de Student mais un test de χ^2 .

Les hypothèses à tester sont données par :

- $H_0 : b_1 = b_2 = \dots = 0$
- $H_1 : b_1 \neq 0 \text{ ou } b_2 \neq 0 \dots$

Il existe en fait 3 tests possibles qui se basent sur la statistique χ^2 afin de vérifier la qualité de l'ajustement par maximum de vraisemblance. Ces trois tests se nomment :

- Le rapport de vraisemblance ou de déviance : il calcule la différence entre 2 valeurs de vraisemblance et est surtout utilisé pour tester la signification globale du modèle.
- Test de Wald : il utilise la pente de la fonction de vraisemblance à proximité de sa valeur maximale et est utilisé pour tester le rôle de chaque paramètre.
- Le score test (ou test du multiplicateur de Lagrange) : il utilise la fonction de vraisemblance en un point quelconque (pas nécessairement son maximum) pour tester qu'un paramètre d'intérêt présente une valeur particulière.

III. 6. Mesure de l'autocorrélation spatiale

III.6. i. Autocorrélation spatiale

La première loi géographique énoncée par Tobler (1970) est "everything is related to everything else, but near things are more related than distant things." L'autocorrélation spatiale est définie par l'existence « d'une relation entre la proximité spatiale des lieux et leur degré de ressemblance ou de dissemblance. » (Donnay, 2015). On parle également d'effet de voisinage.

« L'autocorrélation » correspond à la corrélation du phénomène avec lui-même tandis que le terme « spatial(e) » fait référence à la corrélation mesurée entre les valeurs d'un même phénomène présentées par 2 entités spatiales. Les observations géographiques sont rarement indépendantes, en particulier si elles relèvent de phénomène spatialement continus. L'autocorrélation spatiale peut être calculée au moyen d'indices tels que celui de Moran, celui de Geary ou encore celui de Ord (Goodchild, 1986). L'autocorrélation spatiale est calculée sur l'ensemble d'un territoire (mesure globale) mais peut aussi être décomposée en contributions locales. Dans cette recherche, nous nous intéressons à l'autocorrélation globale d'une variable sur une région donnée.

III.6. ii. Matrice de voisinage

Afin de mesurer l'autocorrélation spatiale, il est nécessaire de connaître la relation spatiale (topologique ou proximale) entre les entités étudiées. Pour cela, une matrice de voisinage (W) est utilisée. Cette matrice nous informe si une entité i est voisine d'une autre entité j (Pebesma & Bivand, 2019 ; Anselin & Rey, 2014). La matrice booléenne de voisinage est une matrice carrée d'ordre n (nombre d'individus spatiaux) dont les éléments prennent les valeurs de 1 ou 0. Pour la colonne i et la ligne j , si la valeur est égale à 1, cela signifie que les entités i et j sont voisines. Dans le cas contraire, l'élément est égal à 0, cela signifie que les deux entités ne sont pas voisines. Le terme de voisinage au sens topologique prend en compte la contiguïté et l'ordre de voisinage. Dès lors, si deux entités possèdent une limite commune, si elles sont contigües, elles sont voisines au niveau d'ordre 1. La contiguïté dépend cependant de la définition du terme de frontière. En effet, il existe une contiguïté de type « Queen » et une contiguïté de type « Rooks ». En mode vectoriel, le type « Rooks » est plus restreint que le type « Queen », car il ne prend pas en compte les voisins dont la frontière est définie par un seul point. Dans la figure 7, l'individu 1 n'est pas contigu à l'individu 2 pour le cas du type « Rooks », mais bien pour le cas du type « Queen ». Les entités ne présentant aucune contiguïté correspondent à des îles (des éléments sans aucun voisin). Elles sont retirées du calcul de l'autocorrélation spatiale.

L'ordre de voisinage 1 prend en considération le(s) voisin(s) d'une entité i partageant sa frontière. L'ordre 2 étend la considération du voisinage à la 2^{ème} couronne de voisins, et ainsi de suite pour les ordres supérieurs de voisinage. L'ordre de voisinage $k > 1$ permet de limiter l'effet des enclaves (1 seul voisin topologique) et des îles qui risquent d'apporter un biais dans le calcul de l'autocorrélation.

La matrice de voisinage est de la forme suivante :

$$\begin{pmatrix} w_{1,1} & \cdots & w_{1,j} & \cdots & w_{1,n} \\ \vdots & & \vdots & & \vdots \\ w_{i,1} & & w_{i,j} & & w_{i,n} \\ \vdots & & \vdots & & \vdots \\ w_{n,1} & \cdots & w_{n,j} & \cdots & w_{n,n} \end{pmatrix}$$

Le poids total de la matrice correspond    :

$$W = \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n w_{i,j}$$

*i et j sont deux individus spatiaux distincts
(ex: secteurs statistiques)*

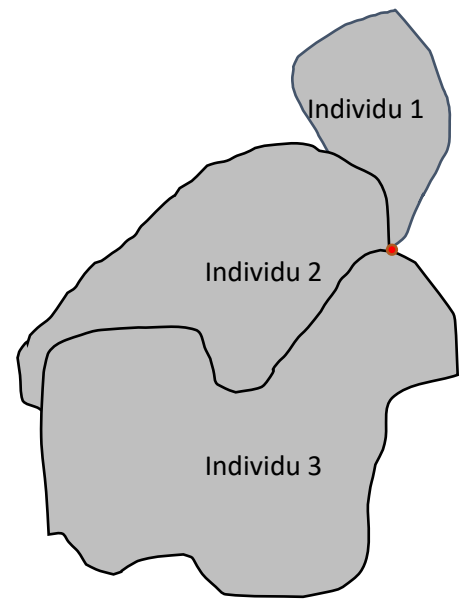


Figure 7: Exemple de la d  finition de contiguit  

La matrice de voisinage est unique, quelle que soit la variable quantitative dont on souhaite   valuer l'autocorr  lation spatiale, pourvu que les entit  s spatiales restent identiques.

III.6. iii. Indice de Moran

L'indice de Moran correspond    une mesure de l'autocorr  lation spatiale de la variable x :

$$I_{Moran} = \frac{n \sum_i^n \sum_j^n w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_i^n (x_i - \bar{x})^2}$$

Auto-covariance

i et j sont deux individus spatiaux (ex: secteurs statistiques)

W = Poids total de la matrice de voisinage

$w_{i,j}$ = valeur d'une cellule de la matrice de voisinage

n = Nombre total d'individus

L'indice de Moran compare une « auto-covariance spatiale » à la variance de la variable x . Afin de tester la signification statistique de l'indice, un test de permutation aléatoire est utilisé. En comparant la valeur calculée de l'indice, avec celles d'un très grand nombre de simulations obtenues par permutations aléatoires des valeurs entre les individus, une pseudo-probabilité est fournie. Elle peut être comparée à un seuil de confiance α , de sorte que si la valeur calculée est très inférieure au seuil α , la présence d'autocorrélation spatiale entre les individus peut être validée.

À partir de cet indice, il est possible de créer un graphique appelé diagramme de Moran. L'axe des abscisses correspond aux valeurs standardisées de la variable x , et l'axe des ordonnées correspond à la moyenne des valeurs standardisées des voisins de chaque individu. Le graphique est divisé en 4 quadrants (figure 8). Le quadrant supérieur gauche LH correspond aux entités de valeur faible au milieu de valeurs élevées du voisinage, le quadrant supérieur droit HH correspond aux entités de valeur élevée au milieu de valeurs élevées du voisinage, le quadrant inférieur gauche LL correspond aux entités de valeur faible au milieu de valeurs faibles du voisinage et le quadrant inférieur droit HL correspond aux entités de valeur élevée au milieu de valeurs faibles du voisinage.

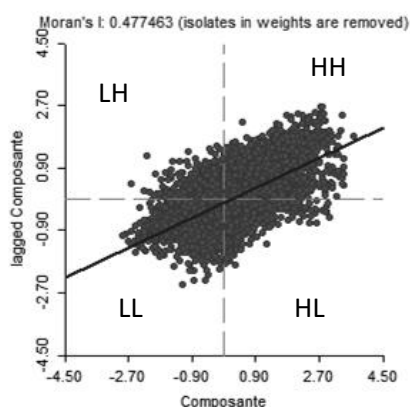


Figure 8 : Diagramme de Moran : exemple obtenu avec 7.512 secteurs statistiques. Variable : première composante de l'ACP, niveau de vie. En abscisse : le niveau de vie des secteurs statistiques, En ordonnée : moyenne du niveau de vie des secteurs voisins au secteur

III.6. iv. Auto-régression spatiale

La présence d'autocorrélation spatiale est causée par un effet de mimétisme entre les entités voisines, susceptible de perturber les résultats d'une analyse de régression. Afin d'évaluer la dépendance spatiale, l'indice de Moran est calculé sur les résidus d'une régression classique. Si l'indice de Moran est significatif, les conditions de la régression ne sont plus strictement respectées et un modèle d'auto-régression spatial peut être testé pour tâcher de prendre en compte l'effet de mimétisme (Anselin & Ray, 2014; Le Gallo, 2002).

Le **modèle général d'auto-régression** spatiale est donné par :

$$Y = f(WY, X, WX, \varepsilon, W\varepsilon)$$

W = Matrice de voisinage standardisée en ligne ; Y = Variable réponse
 X = Matrice des variables explicatives ; ε = Résidus

Le modèle prend en compte la présence de l'autocorrélation spatiale dans la variable dépendante (Y), les variables indépendantes (X) et les résidus par l'intermédiaire des variables spatialement décalées WY, WX et Wε.

Modèle auto-régressif LAG

Dans le cas de la régression LAG, le plus utilisé, seule la dépendance spatiale de la variable dépendante (Y) est prise en compte dans la forme élémentaire du modèle. Ceci se traduit par la présence de la variable spatialement décalée WY. La valeur de Y dans une entité spatiale dépend des valeurs des variables explicatives X dans cette entité ET des valeurs de Y dans les entités voisines.

$$Y = \rho WY + X\beta + \varepsilon$$

Y = Vecteur des valeurs observées de la variable dépendante
 ρ = Coefficient d'autocorrélation spatiale
 W = Matrice des valeurs observées du voisinage, standardisée en ligne
 X = Matrice des variables indépendantes
 ε = Résidu

Le **coefficient** d'autocorrélation spatiale ρ mesure l'intensité d'interaction spatiale entre les individus qualifiés par la valeur de la variable dépendante (Y). Si $\rho = 0$, on retombe sur un modèle sans dépendance spatiale. En d'autres termes, ce coefficient quantifie l'effet de mimétisme entre les voisins sur la variable réponse. Le coefficient ρ est obtenu par maximum de vraisemblance.

Sa signification statistique est vérifiée par un test de Wald, un rapport de vraisemblance (LR) et le multiplicateur de Lagrange (LM). Les hypothèses testées sont classiquement :

- H_0 : ρ est égal à 0
- H_1 : ρ est significativement différent de 0

Les conditions d'utilisation de la régression à vérifier sont l'indépendance des variables explicatives et la normalité, l'homoscédasticité et l'absence de corrélation des résidus.

On notera que la forme réduite du modèle LAG permet d'identifier un effet multiplicatif et un effet de diffusion :

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$$

Le premier terme représente l'effet multiplicatif, soit le fait que dans une entité donnée, la valeur de Y n'est pas uniquement expliquée par les valeurs de X dans cette entité, mais aussi par celles présentées dans les autres régions, voisines ou non (l'effet diminue avec l'éloignement). Le second terme mesure l'effet de diffusion, témoignant que le résidu aléatoire d'une entité affecte non seulement la valeur Y de l'entité, mais aussi les autres entités (même diminution avec l'éloignement).

Il est possible de comparer la qualité d'un modèle de régression par rapport à un autre. Il est évidemment intéressant de comparer le modèle auto-régressif LAG au modèle classique de la régression entre la même variable dépendante et les mêmes variables indépendantes, la seule différence étant la prise en compte de la dépendance spatiale dans le modèle LAG. Cette comparaison permet de déterminer si l'utilisation d'un modèle auto-régressif LAG a été pertinente. Cependant, le coefficient de détermination (R^2) n'est pas toujours comparable. On compare alors les modèles sur le critère « Akaike information » (AIC). Plus l'AIC est petit, meilleur sera le modèle.

CHAPITRE 5: APPLICATION ET VALIDATION

Le chapitre précédent a présenté une série d'outils statistiques d'analyse spatiale que nous comptons utiliser selon deux protocoles distincts, selon la définition choisie du statut socio-économique de la population résidente. Dans un premier temps, nous limiterons ce concept à la seule notion de revenu des ménages dont la définition est non ambiguë et stable dans le temps. Ensuite, nous l'étendrons à une construction synthétique qualifiée de « niveau de vie », via une ACP, exploitant les différentes variables et indices préparés lors de prétraitements. Bien qu'*a priori* plus riche, cette dernière approche est fragilisée par une part d'arbitraire (choix des variables, interprétation des facteurs...) et par une instabilité spatiale et temporelle (une même ACP menée sur les mêmes variables ne garantit pas des résultats identiques si les individus sont différents et/ou en nombres distincts et si les variables sont levées à des dates différentes). Des analyses statistiques similaires sont menées en parallèle selon ces deux approches, revenus et niveau de vie, afin de comparer les résultats et de conforter les conclusions.

I. Première approche: Population définie selon le revenu médian

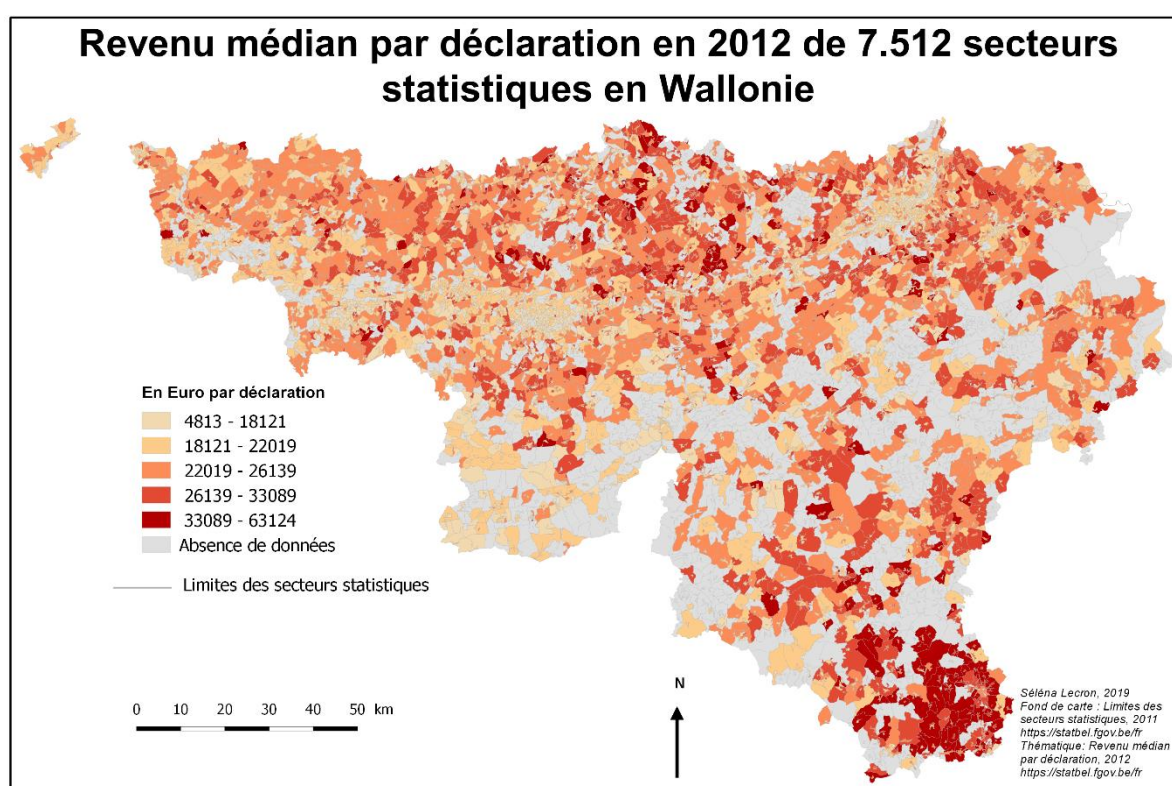
Dans un premier temps, le revenu a été analysé au niveau des secteurs statistiques. Pour rappel, il s'agit du revenu médian par déclaration fiscale qui est utilisé. Ensuite, l'écart relatif entre le revenu médian des secteurs statistiques et le revenu médian de leur commune a été étudié. Ce dernier indice est qualifié par la suite d'écart relatif des revenus.

Parmi les multiples variables et indices présélectionnés (cf. Développement théorique II.1), nous avons identifié les variables susceptibles d'expliquer le revenu. Pour cela, les corrélations du revenu avec les différents indices ont été examinées (annexe 12). Les indices disposant d'une corrélation supérieure à 0,5 avec le revenu médian par secteur statistique sont multiples et largement redondants (part de la population célibataire, part de la population mariée, part de la population disposant d'un diplôme de l'enseignement supérieur, part de la population sans diplôme de secondaire, part de la population composée de ménages d'une personne, taux de chômage, part de logements occupés par les locataires et part de logements occupés par les propriétaires). Parmi ceux-ci, les indices susceptibles d'expliquer le montant des revenus sont : la part de la population disposant d'un diplôme de l'enseignement supérieur, la part de ménages d'une personne et le taux de chômage. Les autres indices sont considérés comme résultant du revenu ou bien présentent une redondance d'information avec les critères déjà retenus.

À partir de cette sélection, plusieurs modèles de régressions multiples sont appliqués : certains avec l'ajout de variables auxiliaires (binaire ou/et catégorielle), d'autres en tenant compte de l'interaction entre les variables explicatives. En outre, une auto-régression spatiale LAG a été exécutée pour tenter de corriger l'autocorrélation spatiale des résidus des régressions multiples classiques.

I. 1. Variation spatiale du revenu

Les prétraitements ont retenu 7.512 secteurs statistiques en Wallonie, susceptibles d'être incorporés dans l'analyse. La distribution spatiale des revenus médians est illustrée à la carte 1 et présente des variations largement discutées dans la littérature (NCGB, 2012 ; Grimmeau *et al.*, 2015 ; Donnay *et al.*, 1986 ; Vandermotten & Vandewattyne, 1985 ; Sporck *et al.*, 1985, etc.). À titre exemple, la région frontalière grand-ducale se distingue par ses revenus médians élevés dans le courant de la seconde moitié du 20^e s., tandis que le sillon Sambre-et-Meuse se détache également, mais par ces revenus plus faibles et cela dès le 19^e s. Des variations plus subtiles apparaissent notamment dans les fonds de vallées, et trouvent leurs origines dans des périodes plus anciennes encore. Ces multiples disparités témoignent d'une autocorrélation spatiale du phénomène et justifient son étude préalable.



Carte 1 : Représentation du revenu médian au sein des secteurs statistiques en 2012 sur l'ensemble de la Wallonie

I. 2. Autocorrélation spatiale

Parmi les variables auxiliaires employées, une variable catégorielle a été traitée pour apporter une information sur l'autocorrélation spatiale de la variable dépendante. Pour ce faire, l'autocorrélation spatiale du revenu ainsi que celle de l'écart relatif des revenus ont été calculées. Les étapes décrites ci-dessous sont : (1) la construction d'une matrice de voisinage et (2) le calcul de l'indice de Moran. Elles suffisent pour réaliser la régression LAG, mais pour la régression multiple, il est en plus nécessaire de classer l'indice de Moran pour construire une variable auxiliaire catégorielle, ce qui correspond à la 3^e étape discutée ci-dessous.

Étape 1 : Création d'une matrice de voisinage (annexe 13)

Une première matrice de voisinage d'ordre 1 a été réalisée (voisinage de type Queen). Elle engendre 865 enclaves, c'est-à-dire 865 secteurs statistiques n'ayant qu'un voisin, et 350 îles. Les îles sont supprimées pour la suite du traitement, et pour éviter le biais engendré par ces enclaves dans le calcul de l'indice de Moran, une matrice de voisinage d'ordre 2 est construite. Le nombre de secteurs ne possédant qu'un voisin diminue ainsi de 865 à 72, mais 2 îles sont isolées. Les secteurs correspondant aux îles sont également supprimés pour la suite du traitement. Cette suppression réduit le nombre de secteurs à 7.510. L'annexe 13 reprend cette construction et l'annexe 14 montre l'effet que l'ordre de la matrice de voisinage a sur l'indice de Moran avec le revenu médian comme exemple.

Étape 2 : Calcul de l'indice de Moran

L'indice de Moran est calculé successivement sur chacune des deux variables : le revenu médian et ensuite l'écart relatif des revenus. Les résultats de ces calculs sont présentés ci-dessous.

L'indice d'autocorrélation du revenu médian présente une valeur de 0,35 dont la significativité est évaluée par une pseudo-probabilité inférieure à 0,0002 (figures 9 et 10).

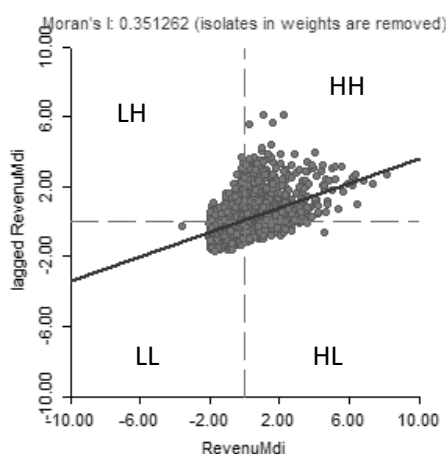


Figure 10 : Diagramme de l'indice de Moran du revenu médian par secteur statistique.

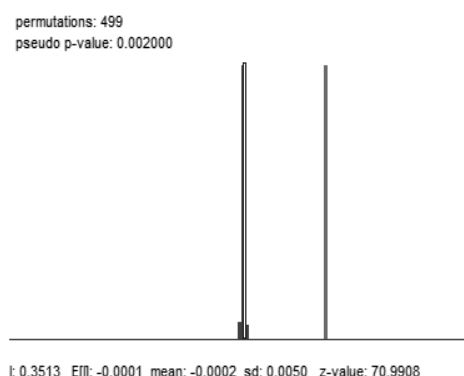


Figure 9: Test de permutation aléatoire du revenu médian

Le calcul de l'indice de Moran sur l'écart relatif des revenus permet de déceler une faible autocorrélation spatiale (0,14) dont la significativité est du même ordre que celle de l'indice précédent (p-valeur = 0.002) (figures 11 et 12)

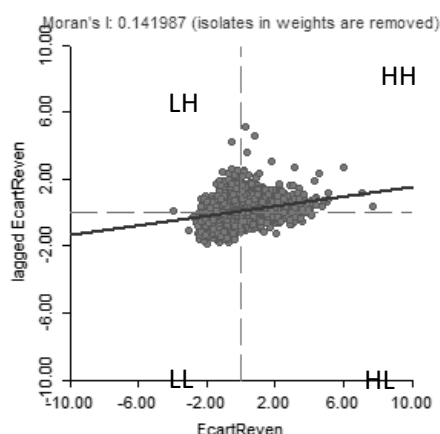


Figure 12 : Diagramme de l'indice de Moran de l'écart relatif des revenus médians par secteur statistique.

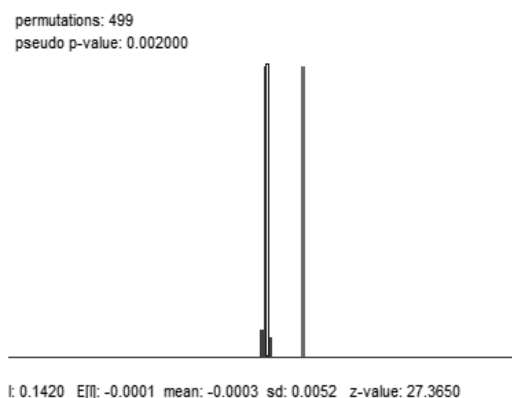


Figure 11: Test de permutation aléatoire de l'écart relatif des revenus médians

Etape 3 : Création de la variable auxiliaire catégorielle

Afin d'obtenir une variable auxiliaire catégorielle, les secteurs statistiques sont classés en fonction de leur contexte spatiale illustré par leur position dans le diagramme de Moran (figure 9 et 11). La répartition des secteurs statistiques dans chacun des quadrants de Moran est repris dans le tableau 5.

Revenu médian des secteurs statistiques

Le quadrant LH correspond aux secteurs statistiques de bas revenus entourés de secteurs avec de hauts revenus. Le quadrant HH correspond aux secteurs statistiques de hauts revenus entourés de secteurs avec de hauts revenus. Le quadrant LL correspond aux secteurs statistiques de bas revenus entourés de secteurs avec de bas revenus. Enfin, le quadrant HL correspond aux secteurs statistiques de hauts revenus entourés de secteurs avec de bas revenus.

Écart relatif des revenus

Les 4 quadrants ont la même signification en termes de voisinage. Ceci dit, un écart relatif élevé ne signifie pas forcément que l'écart est positif. Cela signifie que la valeur standardisée de l'écart relatif des revenus compte parmi les plus grandes. Parmi les valeurs des écarts aux revenus présents dans la classe HH, il existe de nombreux écarts négatifs, ils sont juste « moins négatifs » que les autres.

Tableau 5: Nombre de secteurs statistiques dans les 4 quadrants du diagramme de Moran obtenu à partir du revenu médian et de l'écart relatif des revenus

Nombre de secteurs statistiques dans les 4 quadrants du diagramme de Moran			
Indice de Moran créé à partir du revenu médian des 7.510 secteurs statistiques			
LH	HH	LL	HL
1589	2624	2447	850
Indice de Moran créé à partir de l'écart relatif des revenus des 7.510 secteurs statistiques			
1987	2277	2157	1089

I. 3. Régressions sur le revenu médian des secteurs statistiques

Une régression multiple sur les revenus est réalisée à partir des 3 variables jugées explicatives identifiées précédemment (part de la population disposant d'un diplôme de l'enseignement supérieur, part de ménages d'une personne, taux de chômage).

Ensuite, deux jeux de données sont créés à partir de l'ensemble des 7.512 secteurs statistiques. Ceux-ci sont distingués selon le statut d'inondation défini lors des prétraitements. Par conséquent, une régression multiple est appliquée aux 6.121 secteurs non inondés, et une autre régression est appliquée aux 1.391 secteurs inondés. En outre, cette distinction est traduite par une variable auxiliaire binaire, introduite dans deux nouvelles versions du modèle de régression multiple, l'une sans et l'autre avec interaction de la variable auxiliaire et les autres variables explicatives.

La dernière régression multiple sur le revenu ne concerne que 7.510 secteurs. Les variables dépendantes sont cette fois-ci les 3 indices, les deux classes d'inondation (première variable auxiliaire) et une variable auxiliaire catégorielle décrit ci-dessus.

I.3. i. Régression multiple sur tous les secteurs

Tableau 6: Régression sur le revenu médian des secteurs statistiques (Y) (7.512) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage}$

Régression sur le revenu médian des secteurs statistiques (Y)				
Nombre de secteurs : 7.512				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	23181,3	180,1	128,73	<2e-16
Part de diplômés du supérieur	18316,3	373,3	49,06	<2e-16
Part de ménages 1 personne	-13659,3	406,5	-33,60	<2e-16
Taux de chômage	-15934,4	691,1	-23,06	<2e-16
Écart-type des résidus	3137	Degré de liberté		7508
R²	0,6028	R² ajusté		0,6026
Statistique F	3798	Degrés de liberté		(3 ; 7508)
p-valeur	2,2e-16			

Au regard du coefficient de détermination (tableau 6), la part de variation du revenu expliquée par les 3 variables est de 60%. Tous les coefficients sont significatifs et ils participent donc tous à l'explication de la variance du revenu. Les coefficients de la part de ménages comportant une personne et du taux de chômage sont négatifs (quand ils augmentent, le revenu diminue, ce qui est plausible). On peut donc constater que les 3 indices sélectionnés sont effectivement pertinents à la description du revenu cependant il faut constater que 40% de la variance du revenu n'est pas expliquée par la régression.

I.3. ii. Régressions multiples pour les 2 jeux de données, secteurs inondés / secteurs non inondés

Afin d'analyser l'effet des 3 indices sur le revenu varie selon le statut d'inondation du secteur, les secteurs inondés et non inondés sont étudiés séparément. Une première régression multiple est effectuée sur les 1.391 secteurs inondés.

Tableau 7: Régression sur le revenu médian par secteur statistique inondé (1.391) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage}$

Régression sur le revenu médian des secteurs statistiques INONDÉS				
Nombre de secteurs inondés : 1.391				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	22383,1	351,4	63,70	<2e-16
Part de diplômés du supérieur	18536,6	863,3	21,47	<2e-16
Part de ménages1 personne	-12216,8	729,0	-16,76	<2e-16
Taux de chômage	-14004,0	1331,3	-10,52	<2e-16
Erreur standard des résidus	2789	Degré de liberté		1387
R²	0,6364	R² ajusté		0,6356
Statistique F	809,1	Degrés de liberté		(3 ; 1387)
p-valeur	2,2e-16			

Une deuxième régression multiple est effectuée sur les 6.121 secteurs non inondés.

Tableau 8: régression multiple du revenu effectuée sur les 6121 secteurs non inondés avec $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage}$

Régression sur le revenu médian des secteurs statistiques NON INONDÉS (Y)				
Nombre de secteurs inondés : 6.121				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	23514,1	211,1	111,38	<2e-16
Part de diplômés du supérieur	18055,6	417,3	43,27	<2e-16
Part de ménages 1 personne	-14291,5	489,0	-29,22	<2e-16
Taux de chômage	-16688.2	799.6	-20.87	<2e-16
Erreur standard des résidus	3207	Degré de liberté		6117
R²	0,589	R² ajusté		0,5888
Statistique F	2922	Degrés de liberté		(3 ; 6117)
p-valeur	2,2e-16			

Les coefficients de régression des 3 variables explicatives sont significatives, elles expliquent donc une partie de la variance du revenu. Le coefficient de détermination des secteurs inondés (64%) est plus grand que celui obtenu pour la régression multiple portant sur l'ensemble des secteurs (60%) et que celui de la régression multiple traitant des secteurs non inondés (59%). La part de variance du revenu est donc légèrement mieux expliquée par les 3 indices dans les secteurs inondés. Cependant, le R² reste du même ordre de grandeur (tableaux 7 et 8). Afin de déterminer, si les deux jeux de données sont réellement distincts, une comparaison de leur moyenne est effectuée.

Vérification de la significativité de la différence des moyennes du revenu des secteurs inondés et non inondés

Tableau 9 : Comparaison des moyennes du revenu des secteurs inondés et non inondés

Comparaison de la moyenne du revenu des secteurs inondés avec la moyenne des secteurs non inondés	
Test de Student	
Moyenne du revenu	
Inondé	Non inondé
21.255,30	22.940,41
Statistique t	11,501
Degré de liberté	7510
p-valeur	< 2,2e-16
L'intervalle de confiance de la différence des moyennes à 95%	[1.397,882 ; 1.972,334]

Les moyennes du revenu des secteurs inondés et non inondés sont significativement différentes (tableau 9). De plus, on constate que le revenu des secteurs inondés est plus faible que le revenu des secteurs non inondés. Le revenu paraît donc statistiquement associé au statut d'inondation du secteur.

I.3. iii. Régression multiple avec une variable auxiliaire binaire (inondé / non inondé)

Une autre approche visant à intégrer le statut d'inondation dans l'étude de la variation du revenu consiste à ajouter une variable auxiliaire binaire (inondé/non inondé) à la régression générale des 3 indices.

Sans interaction

Tableau 10 : Régression sur le revenu médian des secteurs statistiques (7.512) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage} + X4 = \text{ClasseInondation}$
Où $X4$ =variable auxiliaire binaire

Régression sans interaction sur le revenu médian des secteurs statistiques (Y)					
Nombre de secteurs inondés : 7.512					
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4= ClasseInondation					
Où X4=variable auxiliaire binaire					
Coefficient	Estimation du coefficient		Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	23183,56		180,14	128,701	<2e-16
Part de diplômés du supérieur	18312,17		373,59	49,017	<2e-16
Part de ménages 1 personne	-13638,27		411,18	-33,169	<2e-16
Taux de chômage	-15939,62		691,21	-23,061	<2e-16
Classe d'inondation	-33,95		95,03	-0,357	0,721
Erreur standard des résidus	3137	Degré de liberté		7505	
R²	0,6028	R² ajusté		0,6026	
Statistique F	2847	Degrés de liberté		(4 ; 7505)	
p-valeur	2.2e-16				

Lorsque $X_4 = 0$ (non inondé) l'ordonnée à l'origine correspond à celle du modèle de base. Lorsque $X_4 = 1$, l'ordonnée à l'origine devient (23183 - 33,95). Au vu de la p-valeur de la variable binaire (en grisé dans le tableau), la différence entre les 2 ordonnées à l'origine de la classe d'inondation n'est pas significative. Selon cette démarche, et contrairement à la précédente, le statut d'inondation ne semble pas participer à l'explication de la variance du revenu des secteurs statistiques (tableau 10).

Avec interactions

Tableau 11 : Régression sur le revenu médian des secteurs statistiques (7.512) avec pour variables indépendantes $X_1 = \text{PartDipSup} + X_2 = \text{PartMénage1pers} + X_3 = \text{TauxCh\^omage} + X_4 = \text{ClasseInondation}$ avec interactions

Régression avec interactions sur le revenu médian des secteurs statistiques (Y)				
Nombre de secteurs : 7.512				
X1 = PartDipSup * X2= PartMénage1pers * X3=TauxChômage * X4=ClasseInondation				
Où X4=variable auxiliaire binaire				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	21257,9	508,6	41,793	< 2e-16
Part de diplômés du supérieur	30784,1	1444,0	21,319	< 2e-16
Part de ménages 1 personne	-5412,2	1703,5	-3,177	0.00149
Taux de chômage	-20567,8	3248,7	-6,331	2.58e-10
Classe d'inondation	-1902,7	1192,2	-1,596	0,11053
PartMénage1pers:PartDipSup	-43384,8	5184,5	-8,368	< 2e-16
PartMénage1pers:TauxChomage	13237,6	8603,9	1,539	0,12395
PartDipSup:TauxChomage	-53191,9	13154,6	-4,044	5.32e-05
PartMénage1pers:PartDipSup:TauxChomage	91117,0	37410,8	2,436	0,01489
ClasseInondation:PartMénage1pers	2447,8	3668,7	0,667	0,50466
ClasseInondation:PartDipSup	2770,0	3568,4	0,776	0,43762
ClasseInondation:TauxChomage	8791,0	6762,2	1,300	0,19363
ClasseInondation:PartDipSup:PartMénage1pers	3393,0	11634,8	0,292	0,77058
ClasseInondation:PartDipSup:TauxChomage	-14356,9	28599,0	-0,502	0,61568
PartDipSup:PartMénage1pers:TauxChomage	91117,0	37410,8	2,436	0,01489
ClasseInondation:PartDipSup:PartMénage1pers:TauxChomage	-11451,9	66804,3	-0,171	0,86389
Erreur standard des résidus	3054	Degré de liberté		7494
R²	0,6241	R² ajusté		0,6234
Statistique F	829,6	Degrés de liberté		(15 ; 7494)
p-valeur	2,2e-16			

Dans le tableau 11, le coefficient de détermination de la régression multiple intégrant les interactions entre les variables explicatives ($R^2 = 62\%$) est mieux expliqué que celui de la régression générale ($R^2 = 60\%$). Néanmoins, cette valeur supérieure provient des seules interactions entre les 3 indices initiaux. À titre de confirmation (annexe 15), nous avons réalisé une régression multiple du revenu par les 3 indices et leurs interactions mais sans tenir compte du statut d'inondation, et elle conduit également à R^2 de 62%. De plus, l'ordre de grandeur des deux coefficients de détermination avec et sans interactions reste semblable. Concernant la significativité des coefficients de régression, on constate que le statut d'inondation n'est toujours pas significatif, pas plus que l'interaction de cette variable binaire explicative avec les autres variables explicatives. De nouveau, le revenu n'est donc pas expliqué par le statut d'inondation du secteur selon ce modèle.

I.3. iv. Régression multiple avec deux variables auxiliaires

Dans la régression qui suit, une deuxième variable auxiliaire est prise en compte. Il s'agit de la variable auxiliaire catégorielle résultant de la classification des secteurs statistiques en fonction de leur position dans le du diagramme de Moran.

Tableau 12 : Régression sur le revenu médian des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage} + X4 = \text{ClasseMoran} + X5 = \text{ClasseInondation}$

Régression sur le revenu médian des secteurs statistiques (Y)				
Nombre de secteurs : 7.510				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4=ClasseMoran + X5=ClasseInondation				
Où X4 = variable auxiliaire de 4 modalités et X5 = variable auxiliaire binaire				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	25184,07	171,74	146,638	< 2e-16
Part de diplômés du supérieur	11342,69	353,60	32,078	< 2e-16
Part de ménages 1 personne	-8681,04	365,96	-23,721	< 2e-16
Taux de chômage	-12073,34	621,38	-19,430	< 2e-16
LH	-4114,09	117,03	-35,154	< 2e-16
HH	212,22	53,99	3,931	8,54e-05
LL	-1238,12	40,58	-30,514	< 2e-16
HL	/	/	/	/
Classe d'inondation	59,16	81,41	0,727	0,467
Erreur standard des résidus	2686	Degré de liberté		7502
R²	0,7089	R² ajusté		0,7087
Statistique F	2610	Degrés de liberté		(6 ; 7502)
p-valeur	2,2e-16			

La variable témoignant de l'autocorrélation spatiale présente donc à 4 modalités (LH, HH, LL, HL). La valeur du coefficient de X4 dépendra de l'appartenance des secteurs à l'une de ces 4 catégories et l'ordonnée à l'origine varie selon 4 possibilités. HL n'a pas de valeur, car il correspond à la modalité de référence. Si la valeur de « la classe de Moran » pour un secteur n'est ni LH, ni HH, ni LL alors il s'agit de HL pour laquelle X correspond à 0. Il y a donc initialement 8 ordonnées à l'origine possibles. Cependant, comme la classe d'inondation n'est toujours pas significative, seules les 4 ordonnées à l'origine correspondant aux 4 contextes spatiaux identifiés grâce au diagramme de Moran sont pertinentes (tableau 12).

Le coefficient de détermination ($R^2 = 71\%$) est supérieur au coefficient de détermination de toutes les régressions précédentes (régression générale $R^2 = 60\%$; régression multiple avec la variable auxiliaire d'inondation $R^2 = 60\%$; régression multiple des secteurs seulement inondés $R^2 = 64\%$; et régression multiple des secteurs seulement non inondés $R^2 = 59\%$). Cette augmentation de qualité de régression est uniquement due à l'apport de la variable auxiliaire de l'autocorrélation spatiale. En effet, comme nous le montrons en annexe, le coefficient de détermination obtenu sans la variable du statut d'inondation reste égal à 71% (annexe 16). Contrairement la classe d'inondation qui est la seule variable n'étant pas significative, l'ajout de la variable décrivant l'appartenance à un quadrant du diagramme de Moran permet de mieux expliquer le revenu.

I.3. v. Régression LAG

Le but de cette régression, dans le contexte de cette étude, est de percevoir si la part de variance non expliquée par le revenu pourrait être due, en partie, à un phénomène d'inondation.

Les étapes suivies pour la réalisation d'une régression LAG sont les suivantes. Premièrement, l'autocorrélation des variables indépendantes est étudiée. Deuxièmement, on examine l'autocorrélation des résidus provenant de la première régression multiple sur les 3 indices. Si les résidus sont autocorrélés spatialement, une régression LAG est justifiée. Si c'est le cas, elle est exécutée et les résultats de la régression sont interprétés. Enfin, on vérifie si l'autocorrélation spatiale a effectivement été éliminée dans les résidus de la régression LAG.

Étape 1 : Analyse de l'autocorrélation spatiale des variables indépendantes et dépendantes

Il y a présence d'autocorrélation spatiale forte dans le cas des 3 variables (0,35 ; 0,49 ; 0,43). Celles-ci apparaissent clairement dans les diagrammes de Moran et les tests de permutation montrent que les indices sont significatifs (annexe 17). Pour appel, l'autocorrélation du revenu médian (figure 9) est également importante (0,35) et jugée significative.

Étape 2 : Analyse de l'autocorrélation spatiale des résidus de la première régression multiple

L'examen des résidus découlant de la régression multiple générale avec 3 variables explicatives (3 indices) montre qu'une autocorrélation significative subsiste dans ces résidus (0,19), qui justifie l'exécution d'une auto-régression spatiale. On doit cependant constater qu'elle est inférieure à l'autocorrélation spatiale des variables initiales, tant indépendantes que dépendante. On peut dès lors présumer que la majeure partie de l'autocorrélation spatiale du revenu est liée à celles des 3 variables de base.

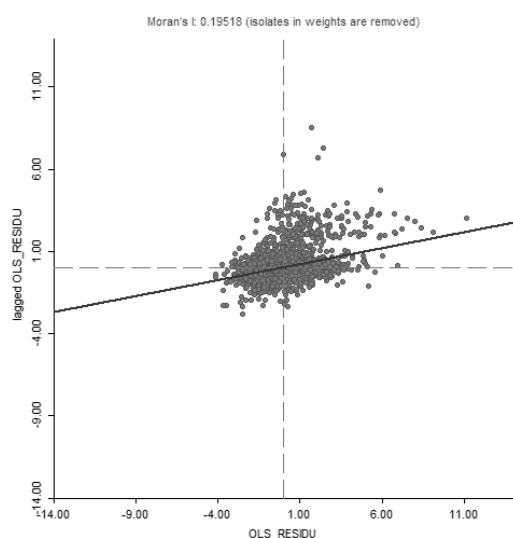


Figure 14: Diagramme de Moran des résidus de la régression classique du revenu

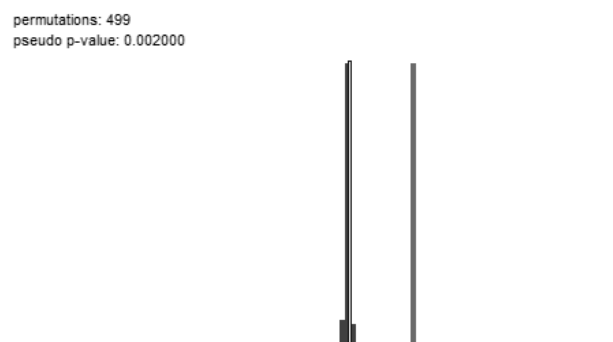


Figure 13: Test de permutation de l'indice de Moran des résidus de la régression sur le revenu

  tape 3 : Analyse des r  sultats de la r  gression LAG

Tableau 13 : R  sultat de la r  gression LAG sur le revenu m  dian o   les variables ind  pendantes sont $X_1 = \text{PartDipSup}$ + $X_2 = \text{PartM  nage1pers}$ + $X_3 = \text{TauxCh  mage}$

REGRESSION

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : RegClassiqueIndice
Spatial Weight : RegClassiqueIndice
Dependent Variable : RevenuM?di Number of Observations: 7512
Mean dependent var : 22628.4 Number of Variables : 5
S.D. dependent var : 4975.49 Degrees of Freedom : 7507
Lag coeff. (Rho) : 0.116795

R-squared : 0.612426 Log likelihood : -71048.6
Sq. Correlation : - Akaike info criterion : 142107
Sigma-square : 9.59457e+006 Schwarz criterion : 142142
S.E of regression : 3097.51

Variable	Coefficient	Std.Error	z-value	Probability
W RevenuM?di	0.116795	0.00970045	12.0401	0.00000
CONSTANT	20487.9	279.883	73.2017	0.00000
TauxChomag	-14455.4	690.887	-20.923	0.00000
PartDipSup	17508.8	378.988	46.1988	0.00000
PartM?nage	-13296.6	402.777	-33.0122	0.00000

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	3	2499.4828	0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : RegClassiqueIndice

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	173.7526	0.00000

Test l'autocorr  lation
spatiale dans les r  sidus
de la r  gression LAG

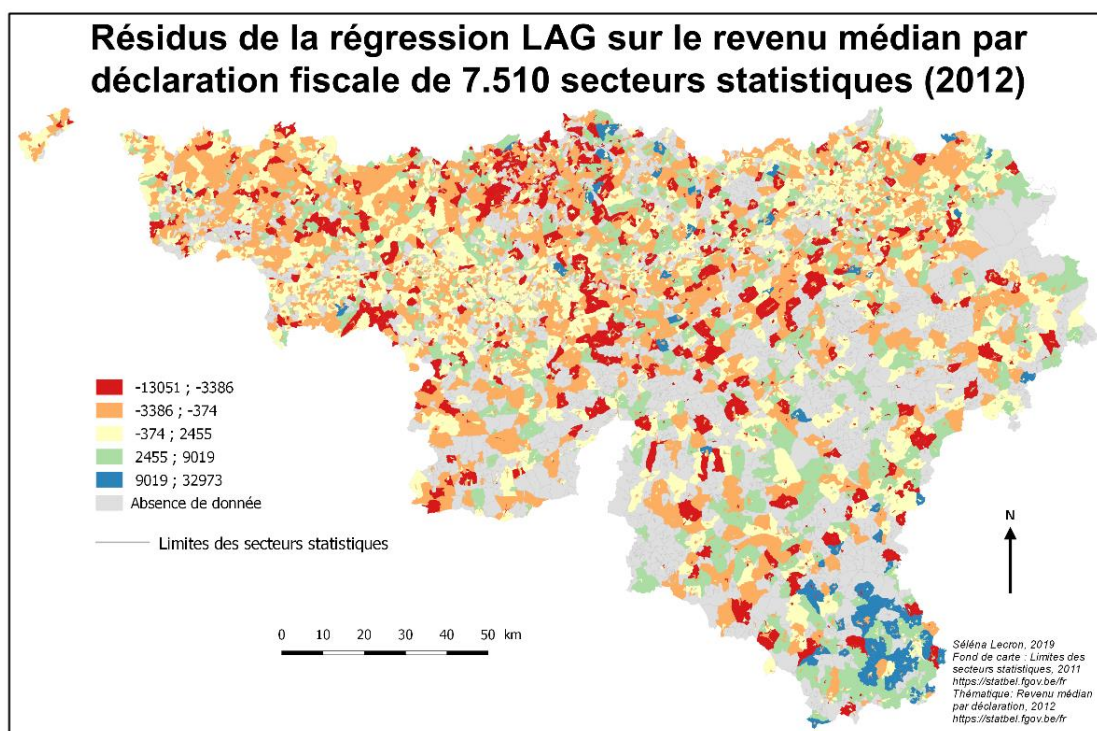
La figure 13, reprend les r  sultats bruts de la r  gression LAG. Suite    la r  gression LAG, le coefficient de d  termination a tr  s l  g  rement augment   (de 0,60, le R^2 passe    0,61). L'AIC est l  g  rement inf  rieur    celui de la r  gression classique (142,107 contre 142,279), ce qui tend    montrer que la r  gression LAG est l  g  rement meilleure que la r  gression classique.

La variable endog  ne d  cal  e ρ , t  moignant de l'effet de mim  tisme entre voisins sur la variable du revenu des secteurs statistiques, est significative (0,18) mais relativement faible. Les effets multiplicatifs et de diffusion ne modifient pas sensiblement les r  sultats de la r  gression g  n  rale, comme le confirme l'analyse des r  sidus.

Étape 4 : Analyse de l'autocorrélation spatiale des résidus de la régression LAG

L'indice de Moran calculé sur les résidus de la régression LAG est faible mais reste significatif (0,13). L'effet de mimétisme n'est pas suffisant pour annuler totalement l'autocorrélation des résidus (Annexe 20) comme le laissait sous-entendre la faible valeur du coefficient ρ . Reste à savoir si l'autocorrélation spatiale résiduelle est susceptible d'être interprétée comme la conséquence d'une cause locale, telle que la situation dans une vallée, ou d'une cause jouant à plus petite échelle ?

La cartographie des résidus de la régression LAG permet de répondre à cette question (carte 2). Elle montre que la plus grande part d'erreur est localisée dans la région d'Arlon et dans le Brabant wallon. Cela pourrait se traduire par « l'ensemble de la population dans la région au sud du Luxembourg belge (ou dans le Brabant wallon) est caractérisée par un revenu élevé ». Il n'y a donc pas vraiment d'effet d'interaction spatiale au niveau des secteurs, mais un effet « régional » traduisant la richesse de l'ensemble de la région. Les résidus non expliqués proviennent donc vraisemblablement des différences de richesse entre les régions, plutôt que d'une cause locale, quelle qu'elle soit. Nous ne pouvons pas donc affirmer au terme de cette régression que le caractère inondable pourrait être un élément caché de la part non expliquée du revenu.



Carte 2 : Représentation spatiale des résidus de la régression LAG sur le revenu médian

I.3. vi. Conclusions sur les régressions multiples du revenu

Les régressions sur les deux jeux de données (inondé, non inondé) pourraient mener à la conclusion que les secteurs inondés expliquent mieux le revenu, les secteurs inondés présentant des revenus plus faibles que les secteurs non inondés. Cependant, outre le fait que les coefficients restent très proches, les autres modèles de régression vont dans le sens inverse.

Ainsi, l'ajout d'une variable auxiliaire témoignant du statut de l'inondation n'explique guère le revenu médian par secteur. En outre, comme évoquée ci-dessus, l'autocorrélation des résidus de la régression LAG implique un effet régional plutôt qu'un effet local. Nous ne pouvons donc pas à ce stade valider l'hypothèse qu'un phénomène d'inondation pourrait être caché dans l'autocorrélation de ses résidus.

I. 4. Régressions sur l'écart relatif entre le revenu des secteurs et le revenu de leur commune

Pour rappel, l'écart relatif des revenus est calculé comme suit :

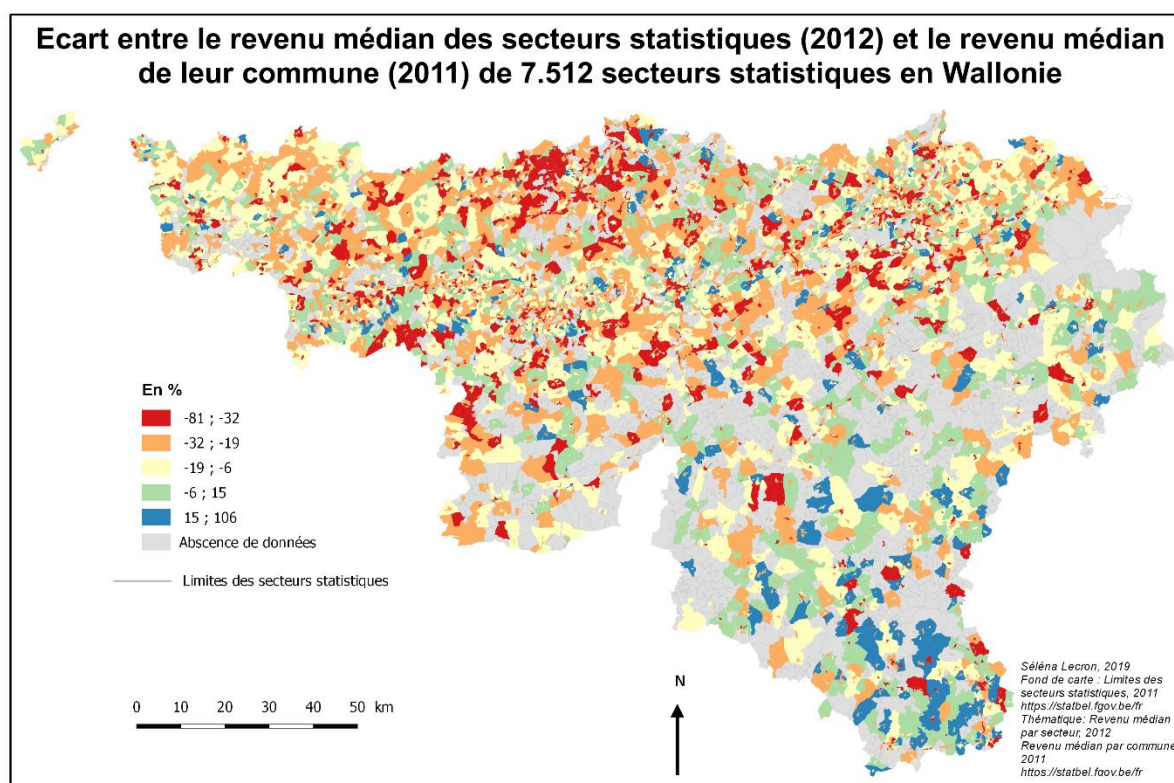
$$\text{Ecart relatif des revenus} = \frac{RM_{sec} - RM_{com}}{RM_{com}}$$

où RM_{sec} = Revenu médian par déclaration en 2012 par secteur statistique
et RM_{com} = Revenu médian par déclaration en 2011 par commune

L'écart relatif entre le revenu médian par secteur statistique et du revenu médian par commune a été élaboré afin d'étudier plus localement la variation du revenu. En effet, au vu des conclusions précédentes de la régression LAG, on constate que le revenu est fortement influencé par un effet régional. L'objectif du calcul de l'écart relatif des revenus est donc de tenter de diminuer l'effet régional observé précédemment (carte 1).

Dans un premier temps, nous analysons la variable correspondant à l'écart relatif entre le revenu médian des secteurs et le revenu médian de leur commune. Pour cela, nous réalisons une analyse statistique et spatiale de ces écarts. Ensuite, une analyse de variance à 2 facteurs, quatre régressions multiples et une autorégression sont appliquées aux données. Ces régressions sont semblables à celles effectuées dans la partie précédente sur le revenu médian.

I.4. i. Analyse et distribution spatiale de l'écart relatif des revenus



Carte 3 : Écart relatif des revenus sur l'ensemble de la Wallonie

Le secteur ayant la plus grande valeur d'écart relatif des revenus (105,6%) se situe dans la commune de Seraing. Cependant, celle-ci correspond à une valeur extrême. Le secteur ayant la valeur la plus basse (-80,6%) se situe dans la commune de Durbuy. Les secteurs présentant des écarts de revenus négatifs se localisent en fait un peu partout en Wallonie. La médiane de l'écart relatif des revenus de l'ensemble des secteurs est négative, ce qui signifie qu'il y a plus de secteurs s'écartant négativement de la valeur du revenu de la commune.

Tableau 14 : Statistiques descriptives de l'écart relatif des revenus

Statistiques descriptives de l'écart relatif des revenus (en %)						
Minimum	1 ^{er} quartile	Médiane	Moyenne	3 ^{ème} quartile	Maximum	Ecart type
-80,578	-27,770	-19,385	-17,632	-9,412	105,667	15,917

Au vu de la carte 3, la région d'Arlon se distingue légèrement par des écarts relatifs des revenus fortement positif (en bleu). Le revenu des secteurs est généralement supérieur au revenu de la commune. La région ouest du Brabant Wallon se distingue elle aussi, mais par des écarts fortement négatifs. Les secteurs présentent donc des revenus inférieurs au revenu de leur commune. L'effet régional n'est donc pas entièrement effacé.

I.4. ii. Régression multiple sur tous les secteurs

Nous avons effectué une régression multiple sur les 3 indices (part de population disposant d'un diplôme de l'enseignement supérieur, part de ménages d'une personne, taux de chômage) où la variable dépendante est cette fois-ci l'écart relatif des revenus.

Tableau 15 : Régression sur l'écart relatif des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$

Régression sur l'écart relatif des revenus (Y)					
Nombre de secteurs : 7.510					
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage					
Coefficient	Estimation du coefficient		Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-6,3284		0,7909	-8,002	1,41e-15
Part de diplômés du supérieur	19,7533		1,6395	12,048	<2e-16
Part de ménages 1 personne	-43,6118		1,7852	-24,430	<2e-16
Taux de chômage	-38,9434		3,0352	-12,831	<2e-16
Erreur standard des résidus	13,78	Degré de liberté		7506	
R²	0,252	R² ajusté		0,2517	
Statistique F	842,8	Degrés de liberté		(3 ; 7506)	
p-valeur	2.2e-16				

La première constatation est que le coefficient de détermination est très faible : les 3 indices n'expliquent que 25% de la variation de l'écart relatif des revenus. Cependant, la variance expliquée est supérieure à la variance résiduelle (test de Fisher) et la régression est donc significative. De plus, tous les coefficients sont significatifs, et donc tous utiles pour expliquer la variance de l'écart relatif des revenus.

I.4. iii. Régressions multiples pour les 2 jeux de données, secteur inondés / non inondés

Les secteurs inondés ont été séparés des secteurs non inondés. À partir de ces deux sets de données, deux régressions multiples sur l'écart relatif des revenus ont été exécutées avec trois variables indépendantes les 3 indices.

La première régression multiple est effectuée sur les 1.391 secteurs inondés.

Tableau 16 : Régression sur l'écart relatif des revenus des secteurs statistiques INONDES (1.391) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxChômage}$

Régression sur l'écart relatif des revenus (Y)				
Nombre de secteurs inondés : 1.391				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-13,522	1,539	-8,789	< 2e-16
Part de diplômés du supérieur	26,556	3,780	7,026	3,32e-12
Part de ménages 1 personne	-36,448	3,192	-11,420	< 2e-16
Taux de chômage	-27,128	5.829	-4.654	3.57e-06
Erreur standard des résidus	12,21	Degré de liberté		1387
R²	0,2885	R² ajusté		0,287
Statistique F	187,5	Degrés de liberté		(3 ; 1397)
p-valeur	2,2e-16			

La deuxième régression multiple est effectuée sur les 6.121 secteurs non inondés.

Tableau 17 : Régression sur l'écart relatif des revenus des secteurs statistiques NON INONDES (6.121) avec pour variables indépendantes $X1 = \text{PartDipSup}$ + $X2 = \text{PartMénage1pers}$ + $X3 = \text{TauxCh\^omage}$

Régression sur l'écart relatif des revenus (Y)				
Nombre de secteurs non inondés : 6.121				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-4,4974	0,9251	-4,862	1,19e-06
Part de diplômés du supérieur	17,3248	1,8284	9,476	< 2e-16
Part de ménages 1 personne	-44,9712	2,1429	-20,986	< 2e-16
Taux de chômage	-43,1796	3,5037	-12,324	< 2e-16
Erreur standard des résidus	14,05	Degré de liberté		6117
R²	0,2328	R² ajusté		0,2324
Statistique F	618,7	Degrés de liberté		(3 ; 6117)
p-valeur	2,2e-16			

On arrive aux mêmes conclusions que pour le revenu, c'est-à-dire que le coefficient de détermination des secteurs inondés ($R^2 = 29\%$) est légèrement supérieur au coefficient de détermination de la régression de l'ensemble des secteurs ($R^2 = 25\%$) et au coefficient de détermination de la régression des secteurs non inondés ($R^2 = 23\%$). Cependant, comme précédemment, les valeurs des R^2 restent du même ordre de grandeur (tableau 16 et 17).

Vérification de la significativité de la différence des moyennes des deux groupes

Tableau 18 : Comparaison des moyennes de l'écart des revenus des secteurs inondés et non inondés

Comparaison de la moyenne de l'écart relatif des revenus des secteurs inondés avec la moyenne de l'écart relatif des revenus des secteurs non inondés	
Test de Student	
Moyenne de l'écart relatif des revenus	
Inondés	Non inondés
-22,333	-16,530
Statistique t	12,391
Degré de liberté	7508
p-valeur	< 2,2e-16
L'intervalle de confiance de la différence des moyennes à 95%	[4,885 ; 6,721]

Dans le tableau 18, on constate que les moyennes de l'écart relatif des revenus des secteurs inondés et non inondés sont significativement différentes. Les secteurs inondés possèdent un écart relatif moyen inférieur à l'écart relatif moyen des secteurs non inondés. Dans les deux cas, les écarts sont négatifs : le revenu des secteurs est généralement inférieur au revenu de la commune et les secteurs dans les zones inondées ont un écart négatif plus prononcé vis-à-vis du le revenu de la commune. Les secteurs inondés sont donc plus « pauvres » (en termes purement financiers) que le reste de la commune. Malgré la significativité du test, l'écart n'est que de 6%, alors que l'écart relatif des revenus de l'ensemble des secteurs varie de -81 à 105%.

I.4. iv. Régression multiple avec une variable auxiliaire binaire (inondé / non inondé)

À partir de la régression générale sur l'écart relatif des revenus utilisant les 3 indices, une variable auxiliaire binaire (inondé / non inondé) est ajoutée.

Tableau 19 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X1 = \text{PartDipSup} + X2 = \text{PartMénage1pers} + X3 = \text{TauxCh\^omage}$

Régression sur l'écart relatif des revenus (Y) avec classes d'inondation				
Nombre de secteurs : 7.510				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4=Classelnondation				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-6,2574	0,7897	-7,924	2,64e-15
Part de diplômés du supérieur	19,4689	1,6378	11,887	< 2e-16
Part de ménages 1 personne	-42,2494	1,8026	-23,438	< 2e-16
Taux de chômage	38,9525	3,0303	-12,854	< 2e-16
Classe d'inondation	-2,1028	0,4166	-5,048	4,58e-07
Erreur standard des résidus	13,75	Degré de liberté		7505
R²	0,2545	R² ajusté		0,2541
Statistique F	640,5	Degré de liberté		7505
p-valeur	2e-16			

Contrairement à la régression des revenus avec la variable auxiliaire de l'inondation menée dans la section précédente, le coefficient d'inondation est cette fois significatif (tableau 19). Dès lors, l'ordonnée à l'origine dans le cas où les secteurs sont non inondés ($X4 = 0$) et dans le cas où les secteurs sont inondés ($X4 = 1$) est significativement différente. L'écart relatif de revenus varie donc en fonction du statut d'inondation.

De plus, cette variation est négative, c'est-à-dire que si le secteur est inondé ($X4 = 1$), l'ordonnée à l'origine est réduite de 2,10. Le revenu des secteurs inondés est donc généralement plus faible que le revenu de sa commune. Cependant, le coefficient de détermination reste faible et n'augmente que très peu par rapport à la régression générale (de 25,2 à 25,4%).

I.4. v. Régression multiple avec deux variables auxiliaires

À partir de la régression de base utilisant les 3 indices, deux variables auxiliaires sont ajoutées. La première témoigne du statut d'inondation comporte 2 valeurs : inondé et non inondé. La seconde prend en compte l'autocorrélation spatiale de l'écart relatif des revenus, et comporte 4 valeurs : HL, HH, LL, LH. Pour rappel, la signification de ces 4 classes est expliquée dans la section § I.2. Pour comparer les effets de ces deux variables auxiliaires, deux régressions sont exécutées, la première sans la classe d'inondation (tableau 20) et la seconde avec la classe d'inondation (tableau 21).

Tableau 20 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X_1 = \text{PartDipSup} + X_2 = \text{PartMénage1pers} + X_3 = \text{TauxCh\^omage} + X_4 = \text{ClasseMoran}$

Régression sur l'écart relatif des revenus (Y) avec modalités de l'indice de Moran				
Nombre de secteurs : 7.510				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4=ClasseMoran				
Où X4 = variable auxiliaire de 4 modalités				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-22,6253	0,5883	-38,460	< 2e-16
Part de diplômés du supérieur	13,1575	1,1832	11,120	< 2e-16
Part de ménages 1 personne	-18,7395	1,2866	-14,566	< 2e-16
Taux de chômage	-14,0041	2,1633	-6,473	1,02e-10
LH	/	/	/	/
HH	21,2124	0,3102	68,375	< 2e-16
LL	-2,4998	0,3079	-8,119	5,43e-16
HL	17,7721	0,3805	46,702	< 2e-16
Erreur standard des résidus	9,665	Degré de liberté		7503
R²	0,6319	R² ajusté		0,6316
Statistique F	2147	Degrés de liberté		(6 ; 7503)
p-valeur	2,2e-16			

On voit que l'apport de la simple variable auxiliaire témoignant de l'autocorrélation spatiale augmente grandement la qualité de l'ajustement ($R^2 = 63\%$). L'écart relatif des revenus du secteur et de son voisinage influence fortement la part de variabilité expliquée de l'écart relatif des revenus. Les 4 valeurs de la variable auxiliaire sont significatives, c'est-à-dire que les ordonnées à l'origine des 4 droites diffèrent, et que les 4 modalités sont nécessaires à la compréhension de la variabilité de l'écart relatif des revenus. (LH est la modalité de référence, correspondant à la valeur de $X=0$, soit à l'ordonnée à l'origine de base) (tableau 20).

Tableau 21 : Régression sur l'écart des revenus des secteurs statistiques (7.510) avec pour variables indépendantes $X_1 = \text{PartDipSup} + X_2 = \text{PartMénage1pers} + X_3 = \text{TauxCh\^omage} + X_4 = \text{ClasseMoran} + X_5 = \text{ClasseInondation}$

Régression sur l'écart relatif des revenus (Y) avec deux variables auxiliaires				
Nombre de secteurs : 7.510				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4=ClasseMoran + X5=ClasseInondation				
Où X4 = variable auxiliaire de 4 modalités et X5 = variable auxiliaire binaire				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Ordonnée à l'origine	-22,5859	0,5885	-38,377	< 2e-16
Part de diplômés du supérieur	13,0706	1,1838	11,041	< 2e-16
Part de ménages 1 personne	-18,3994	1,2977	-14,179	< 2e-16
Taux de chômage	-14,0555	2,1631	-6,498	8,66e-11
LH	/	/	/	< 2e-16
HH	21,1896	0,3104	68,267	< 2e-16
LL	-2,4808	0,3080	-8,055	9,15e-16
HL	17,7681	0,3805	46,700	< 2e-16
Classe d'inondation	-0,5816	0,2934	-1,982	0,0475
Erreur standard des résidus	9,664	Degré de liberté		7502
R²	0,6321	R² ajusté		0,6317
Statistique F	1841	Degrés de liberté		(7 ; 7502)
p-valeur	2,2e-16			

Lorsque l'on injecte le caractère inondable comme seconde variable auxiliaire, le coefficient de détermination multiple ne change quasiment pas ($R^2 = 63\%$), mais le coefficient de la seconde variable auxiliaire est nettement moins significatif que dans le cas où elle apparaît sans celle de l'autocorrélation spatiale (tableau 21). Le statut d'inondation n'explique donc que très peu la variation de l'écart relatif du revenu. L'appartenance à un groupe du diagramme de Moran, par contre, l'influence sensiblement.

I.4. vi. Régression LAG

Étape 1 : Analyse de l'autocorrélation des variables indépendantes et de la variable dépendante

Comme nous l'avons déjà évoqué précédemment, il y a présence d'autocorrélation spatiale forte dans les 3 variables indépendantes (0,35 ; 0,49 ; 0,43) (annexe 17), tandis que l'autocorrélation de l'écart relatif des revenus est sensiblement plus faible (0,14) mais toujours jugée significative (figure 11).

Étape 2 : Analyse de l'autocorrélation dans les résidus de la première régression multiple

Afin de déterminer s'il existe une autocorrélation spatiale de la variable dépendante non prise en compte par les variables indépendantes, nous calculons l'autocorrélation des résidus.

Comme nous l'indique les figures 15 et 16, l'indice de Moran est significatif (0,27). L'interaction spatiale des résidus est même nettement plus importante que l'interaction spatiale de la variable dépendante. Cela pourrait justifier pourquoi la variable auxiliaire catégorielle explique aussi bien ($R^2 = 63\%$) l'écart relatif des revenus.

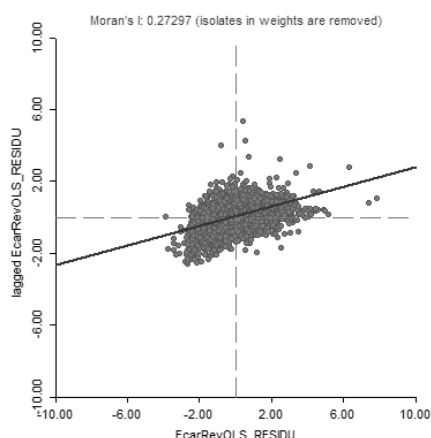


Figure 15: Diagramme de Moran des résidus de la régression classique sur l'écart relatif des revenus

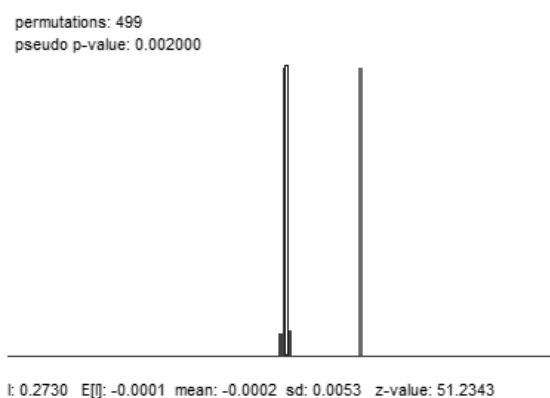


Figure 16 : Test de permutation de l'indice de Moran des résidus de la régression sur l'écart relatif des revenus

Étant donné la présence l'autocorrélation dans les résidus de la régression de base, nous effectuons une autorégression spatiale LAG avec l'écart relatif des revenus comme variable dépendante et les 3 variables indépendantes qui sont la part de diplômés de l'enseignement supérieur, la part de ménages d'une personne et le taux de chômage.

Étape 3 : Analyse des résultats de la régression LAG

La variable endogène décalée ρ témoigne de l'intensité de l'interaction spatiale des variables dépendantes. Elle vaut ici 0,37 et son apport est significatif dans la régression. L'AIC, qui était de 60.712,3 dans la régression classique, vaut à présent 60.277,2. Il a donc diminué légèrement ce qui indique que la régression LAG est meilleure. Le coefficient de détermination multiple a également augmenté. Il était de 25% dans la régression classique et il est de 30% pour la régression LAG (figure 17).

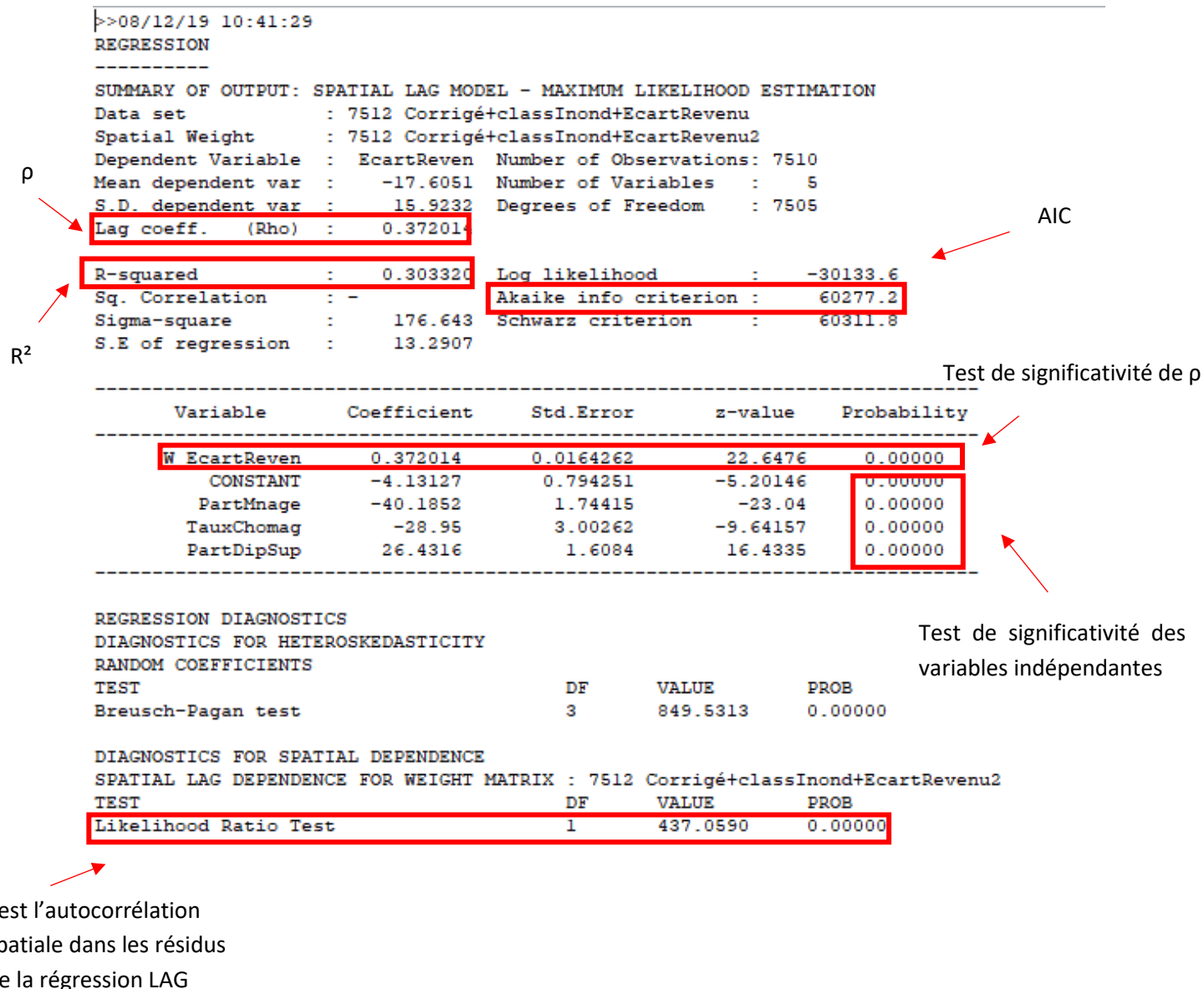
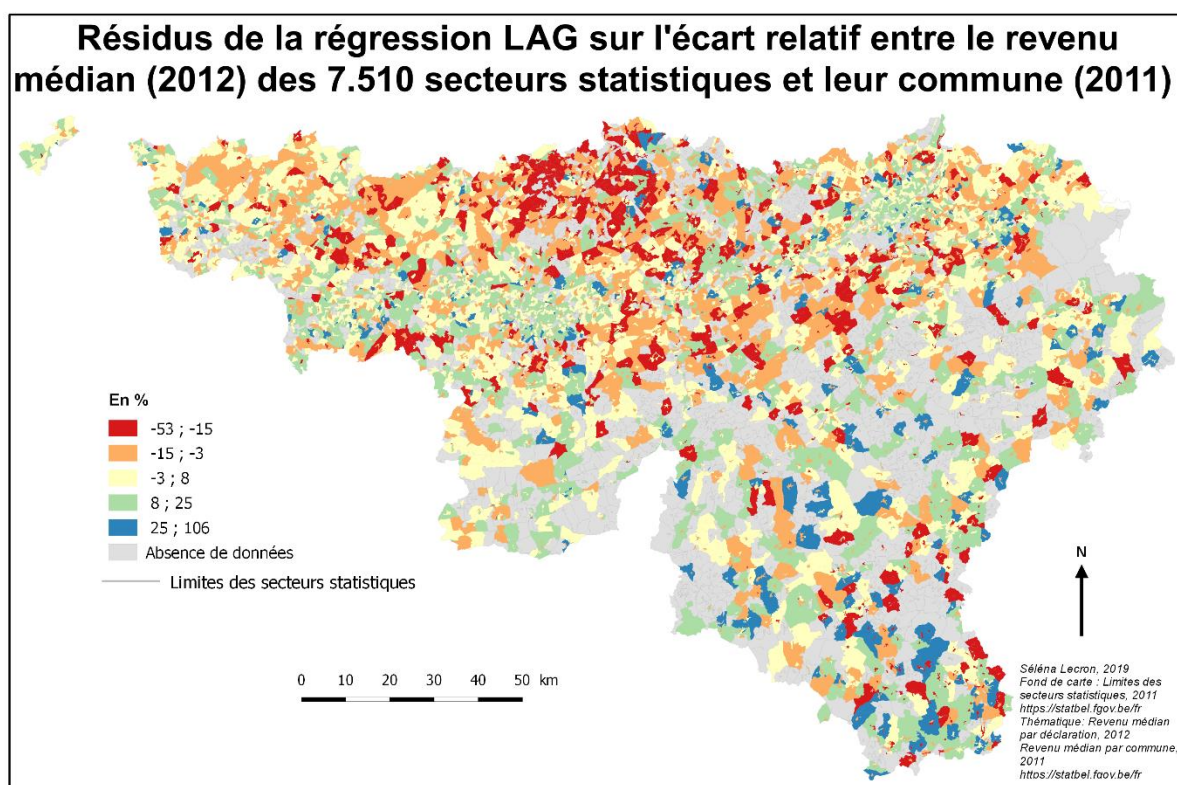


Figure 17 : Résultat de la régression LAG sur l'écart relatif des revenus médian où les variables indépendantes sont $X1 = \text{PartDipSup}$ + $X2 = \text{PartMénage1pers}$ + $X3 = \text{TauxCh\^omage}$

Étape 4 : Analyse de l'autocorrélation des résidus de la régression LAG

Le test de déviance de la régression LAG indique que la régression LAG est significative. Cependant, comme l'indique l'annexe 18, l'indice de Moran des résidus (0.13) reste significatif bien qu'ayant diminué de moitié vis-à-vis de la régression de base.

La carte des résidus montre que le sillon Sambre-et-Meuse se distingue légèrement par des résidus élevés, tout comme une bonne partie de la province du Luxembourg au sein de laquelle la région d'Arlon ne se démarque pas particulièrement. La région du Brabant Wallon et la bordure méridionale du sillon sambro-mosan (tout particulièrement au sud de Namur) se distinguent, elles, par des résidus négatifs. Un phénomène régional reste donc présent, mais apparemment distinct de celui évoqué pour la distribution des résidus des revenus. Il s'agirait plutôt ici d'un phénomène lié à la mise en place de banlieues résidentielles relativement riches autour des agglomérations urbaines, mais cela reste à démontrer d'autre part et ne semble pas impliquer le phénomène d'inondation potentielle.



Carte 4 : Représentation spatiale des résidus issus de la régression LAG sur l'écart relatif des revenus

I.4. vii. Conclusion générale de la première approche sur le revenu médian

L'objectif en calculant l'écart relatif entre le revenu médian des secteurs et le revenu médian de leur commune était de diminuer voire d'effacer un effet régional sensible dans l'analyse des seuls revenus médians par secteur. Nous constatons, via la carte de l'écart relatif des revenus et la carte des résidus du LAG, que cette approche n'a pas été concluante (carte 3 et 4).

Le statut d'inondation est significatif dans l'ensemble des régressions effectuée dans cette section, cependant il n'explique quasiment rien de la variance de l'écart relatif des revenus. Seule les régressions multiples séparant les secteurs inondés et non inondés semblent convenir que l'écart relatif des revenus dépend du statut d'inondation. Néanmoins le coefficient de détermination reste fort proche du coefficient de la régression multiple générale.

Dès lors, au terme de ces démarches, il faut admettre que le statut d'inondation n'influence pas de façon incontestable le revenu et l'écart relatif des revenus. La notion de « revenus » est peut-être trop restrictive ? D'où l'intérêt de créer une variable synthétique traduisant le « niveau de vie », peut-être plus encline à être influencée par le caractère inondé du secteur ?

II. Deuxième approche : Population définie selon la composante synthétique du niveau de vie

La composante du niveau de vie est censée synthétiser une série de variables socio-économiques disponibles par secteur statistique en Wallonie. La démarche relève d'une analyse multivariée (ACP) et elle a été utilisée à de nombreuses reprises dans les études de sociologie et de géographie humaine et économique, à des échelles spatiales diverses (« écologie factorielle urbaine » par exemple ; Berry, 1971 ; Racine & Raymond, 1973 ; Mather & Openshaw, 1974 ; Donnay et al., 1986). Elle pêche cependant par le fait qu'elle est foncièrement exploratoire et non strictement reproductible dans le temps et dans l'espace, et que l'interprétation des résultats ne peut s'affranchir d'un certain arbitraire.

Les pages précédentes, traitant du seul revenu médian par secteur statistique, ne s'étant pas avérées concluantes vis-à-vis de l'hypothèse de départ, nous tentons ici de vérifier si le caractère inondable des entités spatiales peut influencer le niveau de vie *sensu lato* de la population résidente. L'hypothèse devient donc que le caractère inondable d'un secteur statistique a tendance à faire diminuer le niveau de vie de la population.

II. 1. Sélection des indices sur base des corrélations et de la variance

Parmi les variables et indices identifiés dès les prétraitements, il existe des redondances plus ou moins fortes dont l'ACP tirera parti pour construire les nouvelles variables synthétiques (composantes ou facteurs). Pour éviter toute interférence des unités dans lesquelles sont exprimés les variables et indices, ils sont standardisés. En d'autres termes, la matrice d'inertie dont on souhaite extraire les vecteurs propres est la matrice des corrélations entre les variables et indices (figure 18 et annexe 12).

Si la corrélation entre variables initiales est indispensable à la démarche, il est clair qu'une redondance totale, ou presque, entre deux variables ou indices est inutile. Par conséquent, une sélection parmi les variables et indices est effectuée sur base de cette matrice des corrélations. Dans le tableau 22, figurent les indices se ressemblant dans leur définition, parmi lesquels une sélection devrait pouvoir être effectuée, par exemple sur la base de la plus grande variance (avant standardisation). En première analyse, sont ainsi conservés : la part de la population mariée, la part de la population disposant d'un diplôme de l'enseignement supérieur et le taux de dépendance.

Tableau 22: Indices redondants avec leur variance et leur corrélation pour justifier la sélection

Indices redondants		
Indices	Corrélation	Variance
Part de la population célibataire / divorcée	-0.988	0.0044
Part de la population mariée		0.0167
Part la population disposant d'un diplôme de l'enseignement supérieur	-0.903	0.0139
Part de la population sans diplôme de l'enseignement secondaire		0.0137
Taux de dépendance vieillesse	0.785	0.0223
Taux de dépendance		0.0272

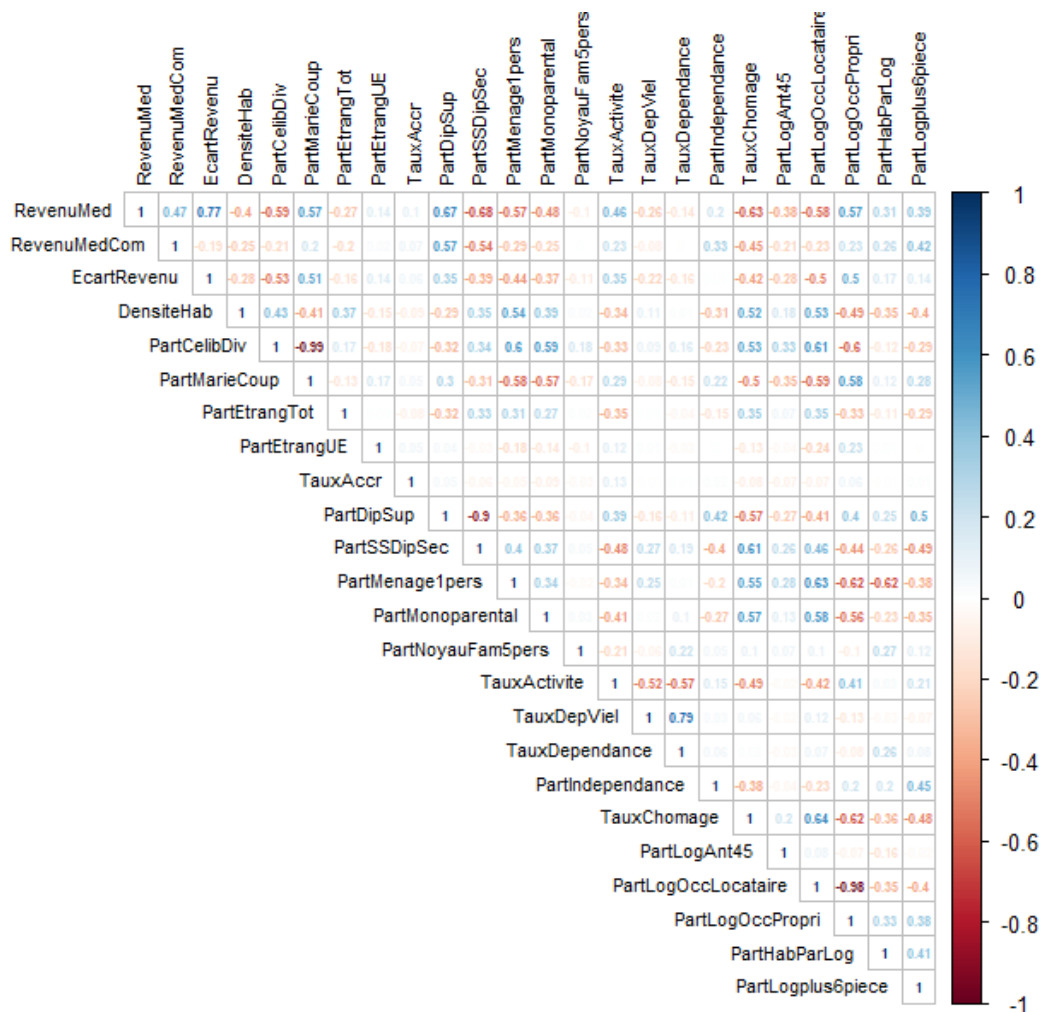


Figure 18: Corrélation entre les 24 indices socio-économiques calculés à partir du recensement de 2011

II. 2. Construction de la composante « niveau de vie »

Comme dans toute démarche exploratoire, la recherche d'une composante traduisant le niveau de vie s'est faite par essais et erreurs. L'objectif est d'obtenir une composante unique, reprenant le maximum d'information et capable d'une interprétation non ambiguë si possible sans rotation complémentaire des axes factoriels.

Afin d'identifier les variables et indices complémentaires aux revenus, incontournables dans la définition générale du niveau de vie, une première ACP a été réalisée sans tenir compte de la variable du revenu médian (annexe 19). Quelques indices se regroupent rapidement, en l'occurrence ceux utilisés dans les analyses de régressions sur les revenus menés dans la section précédente (part de la population disposant d'un diplôme de l'enseignement supérieur, par de ménages d'une personne et taux de chômage). À l'opposé, plusieurs variables et indices paraissent inconciliables avec les variables précédentes. Une nouvelle sélection et des substitutions entre variables largement redondantes ont ainsi été effectuées, en se basant sur les valeurs des saturations et des communautés, ainsi que sur les analyses descriptives détaillées de chacun des indices et variables candidats (annexes 4, 5 et 19). Enfin, le revenu médian a été introduit pour lancer une deuxième ACP sur un total de 14 variables et indices standardisés (annexe 20).

La première composante extraite rassemble 41% de l'information totale, alors que la deuxième n'en explique que de 9%. Les saturations observées avec les variables et indices originaux (tableau 23) permettent d'interpréter ce premier facteur comme la composante du niveau de vie attendue. Ainsi, le taux de chômage est l'indice le plus corrélé avec le niveau de vie (82%), suivi du revenu médian (-80%), de la part de logements occupés par le locataire (79%) et de la part de ménages d'une personne (76%). Le signe de ces saturations (notamment positif pour le taux de chômage et négatif pour le revenu) permet de déterminer que cette composante est une mesure synthétique de la précarité de la population.

Tableau 23: Résultat de l'ACP sur la composante 1, celle correspondant au niveau de vie

Saturations, valeur propre et variance expliquée de la composante du « niveau de vie »	
	Saturations
Revenu médian	-0.801
Log de la densité d'habitants	0.585
Part de population mariée	-0.681
Part de ménages d'une personne	0.758
Part de la population monoparentale	0.683
Part totale d'étrangers	0.465
Part de population disposant d'un diplôme de supérieur	-0.692
Part d'indépendants	-0.462
Taux d'activité	-0.567
Taux de chômage	0.828
Part de logements antérieurs à 1945	0.317
Part de logements occupés par le locataire	0.787
Part de logements de plus de 6 pièces	-0.628
Nombre d'habitants par logement	-0.503
Valeur Propre	5.77
Proportion de variance expliquée	41%

II. 3. Analyse du score factoriel

Le score factoriel correspond à une cote du niveau de vie (précarité) pour chaque secteur. Dès lors, un score élevé signifie que la population du secteur est précaire et, à l'inverse, un score faible signifie que la population présente un bon niveau de vie.

L'histogramme des fréquences des scores factoriels obtenus pour tous les secteurs est représenté à la figure 19. Les scores de valeurs absolues importantes ($> |3|$) sont interpellant et ils ont fait l'objet d'une analyse individuelle statistique et cartographique en utilisant les ressources Q-GIS et Google Maps. Les commentaires, justifiant dans quelques cas l'élimination du secteur statistique pour la suite des analyses, sont fournis à l'annexe 21. L'histogramme des scores, après suppression des secteurs jugés erratiques, est disponible à l'annexe 22.

Tableau 24 : Résumé des statistiques descriptives des 7.525 secteurs (avant suppression des cas erratiques)

Résumé des statistiques descriptives des 7.525 secteurs (avant suppression)					
Minimum Q_0	1 ^{er} quartile Q_{25}	Médiane Q_{50}	Moyenne	3 ^{ème} quartile Q_{75}	Maximum Q_{100}
-2,849	-0,693	-0,145	0,000577	0,603	3,891
Résumé des statistiques descriptives des 7.512 secteurs (après suppression)					
-2,849	-0,694	-0,147	-0,00521	0,598	3,646

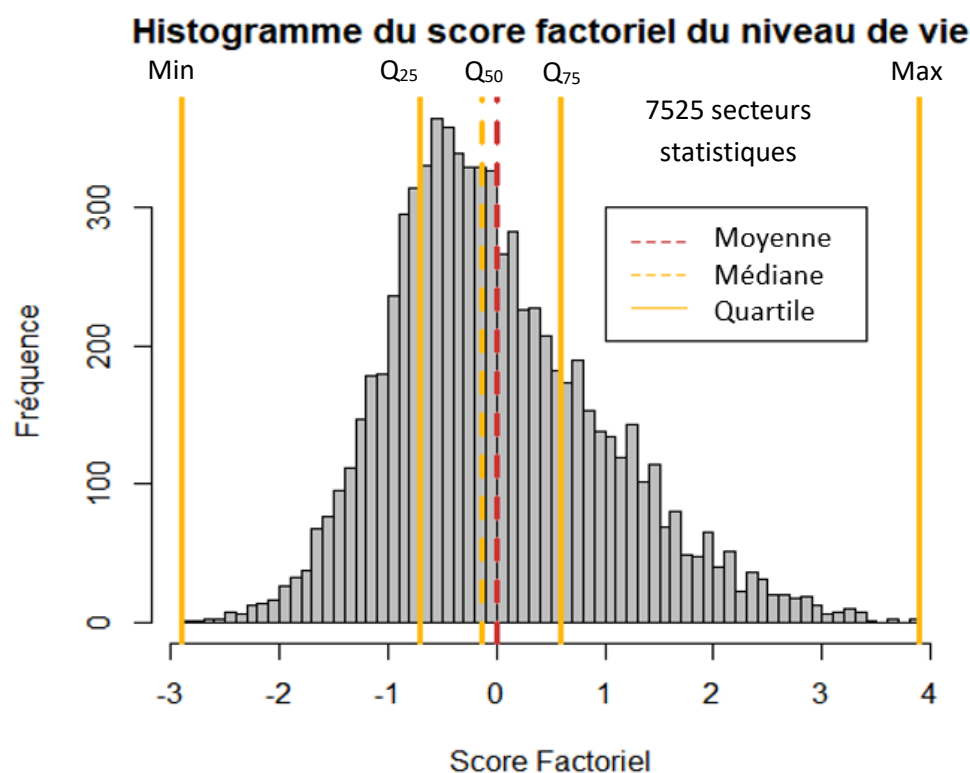
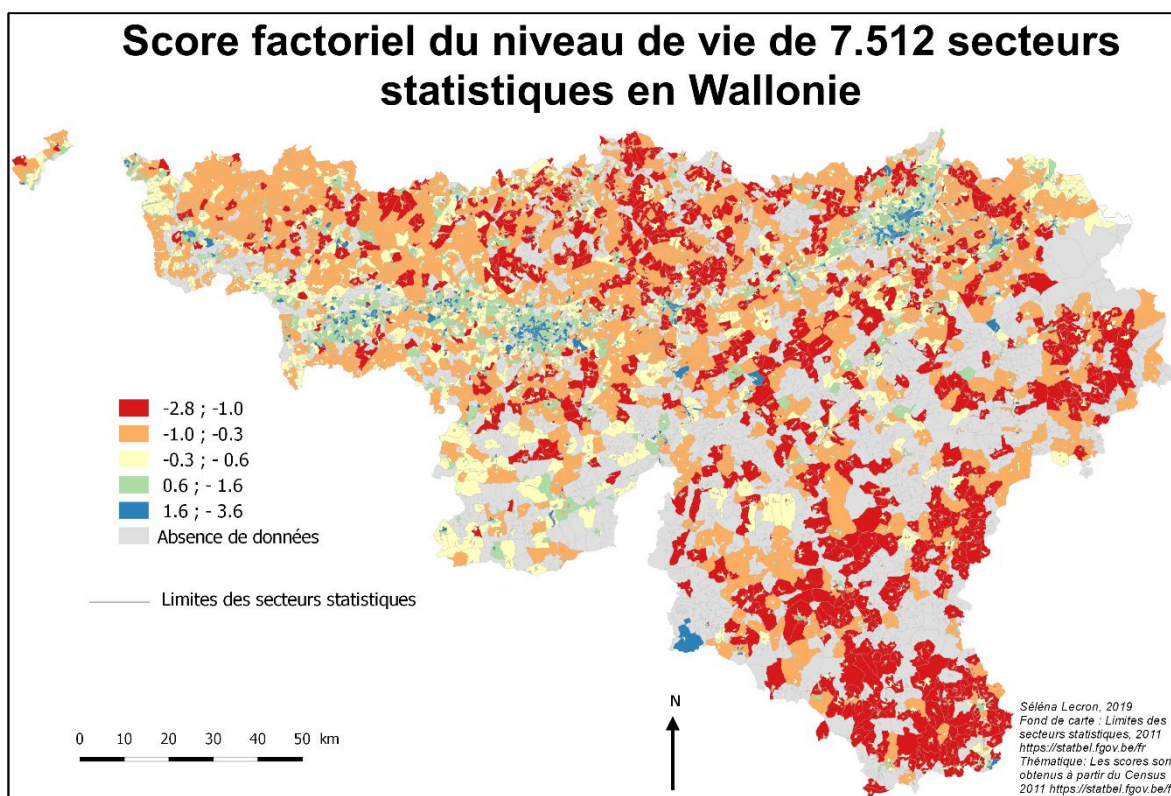


Figure 19: Histogramme du score factoriel du niveau de vie pour 7.525 secteurs

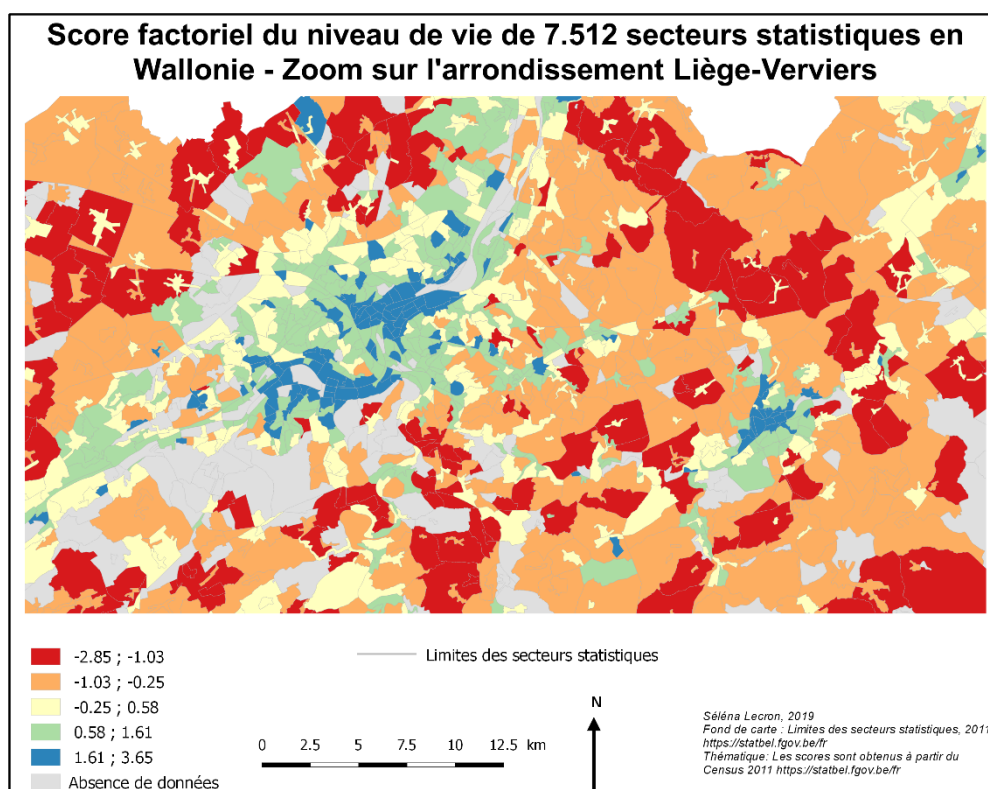
Analyse cartographique de la distribution spatiale des secteurs

Le sillon Sambre-et-Meuse ressort clairement sur la carte 5 figurant les scores factoriels de la composante du niveau de vie. Cette région est caractérisée par les scores les plus élevés donc par la population présentant les niveaux de vie les plus bas. À l'inverse, la province du Luxembourg possède des scores très négatifs, c'est-à-dire une population dont le niveau de vie est élevé.

La carte 6 reprend un extrait à plus grande échelle de la carte précédente, centré sur l'arrondissement de Liège-Verviers. Les scores élevés des centres-villes sont bien marqués, à l'inverse des scores plus faibles de la périphérie, témoignant d'une dualité spatiale désormais bien connue dans toutes les agglomérations urbaines.



Carte 5 : Score factoriel du niveau de vie de 7512 secteurs en Wallonie



Carte 6 : Score factoriel du niveau de vie de 7.512 secteurs en Wallonie - Zoom sur l'arrondissement de Liège-Verviers

II. 4. Mesure de l'autocorrélation du score factoriel

Les observations précédentes laissent entendre que le score factoriel du niveau de vie présente une autocorrélation spatiale importante. Pour le vérifier, l'indice de Moran est calculé en utilisant la même matrice de voisinage que dans les analyses précédentes, puisque la décomposition spatiale de la variable étudiée est identique (matrice binaire d'ordre 2, voisinage de type Queen sur 7.510 secteurs statistiques).

La valeur de l'indice de Moran est désormais de 0,48 et jugée largement significative (figure 21 et 20). Le voisinage spatial ayant montré son importance dans la section précédente traitant du revenu, nous conservons l'idée de classer les secteurs statistiques en fonction de leur appartenance à 4 quadrants du digramme de Moran. Dans la figure 21, on observe la fréquence des secteurs par quadrant. L'interprétation de l'appartenance à un quadrant est évidemment identique aux précédentes, tout en se souvenant que les scores positifs correspondent aux secteurs de faible niveau de vie, et les scores négatifs aux secteurs de niveau de vie élevé. À titre d'exemple, le quadrant LH rassemble des secteurs statistiques d'un niveau de vie élevé (faibles scores) entourés de secteurs de niveau de vie faible (scores élevés). Chaque secteur statistique se retrouve ainsi qualifié par une des 4 « classes de Moran » comme dans la section précédente réservée à l'analyse des revenus.

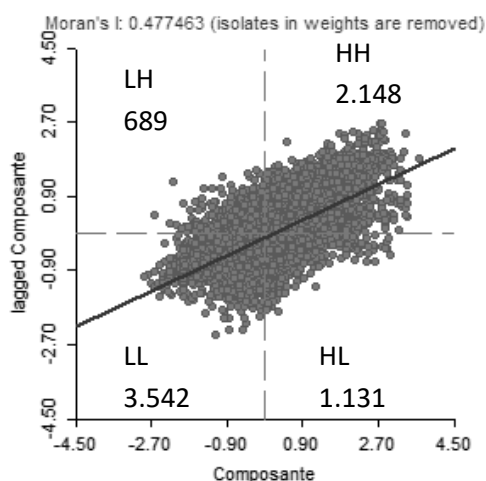


Figure 21: Indice de Moran calculé sur le score factoriel du niveau de vie avec le nombre de secteurs par quadrant

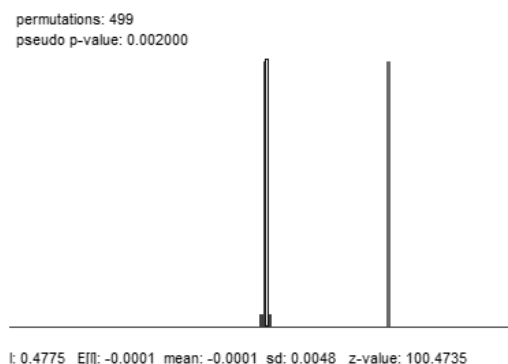


Figure 20: Test de permutation de l'indice de Moran du score factoriel

III. Analyse combinée du statut d'inondation et du niveau de vie d'un secteur statistique

Puisque désormais toutes les indices socio-économiques pris en compte dans notre recherche sont synthétisés par une seule composante, il n'est plus question de tenter une régression des scores factoriels sur des variables « explicatives » agrémentées de quelques variables auxiliaires. Il s'agit de confronter directement les valeurs des scores aux catégories d'inondation et/ou aux classes de Moran, afin de voir s'il existe une relation entre les phénomènes.

Les méthodes classiques relèvent dès lors des statistiques non paramétriques (test du χ^2) et des analyses de variances à une ou plusieurs entrées. On complètera ce protocole par une régression logistique afin de tester si le niveau de vie de la population peut expliquer que celle-ci réside dans une zone inondable ou non (variable réponse binaire).

III. 1. χ^2

Le test de χ^2 permet de tester l'association entre deux variables qualitatives, nominales ou ordonnées. Pour cela, il faut préalablement classer les scores factoriels du niveau de vie (transformation en une variable ordonnée). Les valeurs des scores étant standardisées, une classification par écarts-types était naturelle. Elle utilise 5 classes d'une amplitude d'un écart-type, avec une classe centrale centrée sur 0, et deux classes ouvertes aux extrémités (tableau 25). Le niveau de vie ainsi classé est croisé avec une matrice de contingence est définie afin de croiser les 5 classes du score factoriel du niveau de vie avec les 2 modalités du statut d'inondation.

Tableau 25: Classement des secteurs statistiques en fonction du score factoriel du niveau de vie (5 classes)

Classement des secteurs en fonction de leur score factoriel				
Classe 1	Classe 2	Classe 3	Classe 4	Classe 5
< -1,5	[-1,5 ; -0,5[[-0,5 ; 0,5[[0,5 ; 1,5[$\geq 1,5$
305	2.255	2.895	1.449	608

Les tableaux des fréquences observées, des fréquences attendues et de l'écart entre ces deux fréquences sont représentés ci-dessous (tableaux 26, 27, 28). S'il y a indépendance entre les deux variables (le niveau de vie et le statut d'inondation), on s'attend à ce que les fréquences attendues, sous cette hypothèse d'indépendance, soient approximativement égales aux fréquences observées.

Tableau 26: Table de contingence des fréquences observées entre le score factoriel mesurant le niveau de vie et le statut d'inondation

Fréquences observées							
Statut d'inondation		Niveau de vie					Somme
		< -1,5	[-1,5 ; -0,5[[-0,5 ; 0,5[[0,5 ; 1,5[≥ 1,5	
	Inondé	33	262	530	374	192	1.391
	Non-inondé	272	1.993	2.365	1.075	416	6.121
Somme		305	2.255	2.895	1.449	608	7.512

Tableau 27: Table de contingence des fréquences attendues entre le score factoriel mesurant le niveau de vie et le statut d'inondation

Fréquences attendues							
Classe d'inondation		Niveau de vie					Somme
		< -1,5	[-1,5 ; -0,5[[-0,5 ; 0,5[[0,5 ; 1,5[≥ 1,5	
	Inondé	56	418	536	268	113	1.391
	Non inondé	249	1.837	2.359	1.181	495	6.121
Somme		305	2.255	2.895	1.449	608	7.512

Tableau 28: Ecart entre les fréquences observées et les fréquences attendues entre le score factoriel et le statut d'inondation

Ecart entre les fréquences observées et attendues							
Classe d'inondation		Niveau de vie					
		< -1,5	[-1,5 ; -0,5[[-0,5 ; 0,5[[0,5 ; 1,5[≥ 1,5	
	Inondé	-23	-156	-6	106	79	
	Non inondé	23	156	6	-106	-79	

Le tableau 28 des écarts entre les fréquences observées et les fréquences attendues montre que, pour les scores négatifs (les classes 1 et 2) correspondant à un niveau de vie plus élevé, la fréquence observée des secteurs inondés est toujours inférieure à la fréquence attendue. Il y a donc moins de secteurs dans la classe « inondé » qu'attendu. Au contraire, pour les scores positifs (les classes 4 et 5), correspondant à un niveau de vie faible, la fréquence observée des secteurs inondés est toujours supérieure à la fréquence attendue. La constatation inverse peut être faite pour les secteurs non inondés. Au vu de ces résultats, on peut en déduire que, d'un point de vue relatif, il y a relativement plus de secteurs inondés pour un niveau de vie faible.

Le test de χ^2 permet de tester cette dépendance statistiquement et globalement. Il aboutit à une p-valeur inférieure au seuil de confiance 0,05. (**Test de χ^2** : $\chi^2 = 203.2$ pour un degré de liberté = 4 et une p-valeur < 2.2×10^{-16}). L'hypothèse nulle peut donc être rejetée : le niveau de vie et le statut d'inondation des secteurs statistiques sont fortement associés (dépendants). On modérera cependant quelque peu ces propos en se souvenant que le test du χ^2 , s'il est robuste, est aussi particulièrement laxiste en rejetant volontiers l'hypothèse nulle dès que le nombre d'individus est élevé...

III. 2. ANOVA

L'analyse de variance (ANOVA) consiste à examiner la part de la variance du niveau de vie expliquée par le statut d'inondation d'un secteur. Deux types d'ANOVA sont exploités : à 1 facteur et à 2 facteurs de classification.

La première ANOVA est composée d'un seul facteur qualitatif de classification : le statut d'inondation dont les modalités prennent les valeurs suivantes : inondé, non inondé. La seconde ANOVA est déterminée par deux facteurs qualitatifs : le statut d'inondation et la « classe de Moran ». Cette dernière est, comme explicité plus haut, un indicateur du contexte d'autocorrélation spatiale des secteurs composé de 4 valeurs : HH, HL, LL, HL. Ces 4 classes sont définies par la valeur du score factoriel du secteur et la moyenne des valeurs de score factoriel des secteurs voisins. Elles correspondent aux 4 quadrants dans lesquels un secteur peut se trouver dans le diagramme de Moran. Dans les deux cas, la variable à expliquer correspond au score factoriel du niveau de vie. Si l'hypothèse d'homogénéité des variances et l'hypothèse de normalité des résidus ne sont pas vérifiées, un test de Kruskal-Wallis et un test de Wilcoxon peuvent être appliqués en lieu et place de l'ANOVA.

III.2. i. Anova à un facteur

Tableau 29 : Statistique descriptive du score factoriel du niveau de vie des 7.512 secteurs en fonction de leur statut d'inondation

Statistiques descriptives du score factoriel des 7.512 secteurs selon leur statut d'inondation								
Statut d'inondation	Total	Moyenne	Écart-type	Q25	Q75	Médiane	Max	Min
Inondé	1.391	0,337	1,04	-0,382	1,04	0,221	3,36	-2,85
Non inondé	6.121	-0,0826	0,964	-0,741	0,473	-0,228	3,65	-2,77

La classe des secteurs non inondés possède une moyenne du score factoriel à peine négative et inférieure à celle de la classe des secteurs inondés (tableau 29). Étant donné qu'un score factoriel négatif témoigne d'un niveau de vie plus élevé, ces résultats sont cohérents avec notre hypothèse de départ. Cependant, comme l'illustre la figure 22, les deux boxplots construites sur les intervalles interquartiles se chevauchent largement. Les valeurs extrêmes visibles dans la même figure correspondent aux éléments supérieurs à 1,5 fois l'intervalle interquartile. Le nombre, l'identifiant et le statut d'inondation des secteurs statistiques de ces valeurs extrêmes sont repris à l'annexe 23.

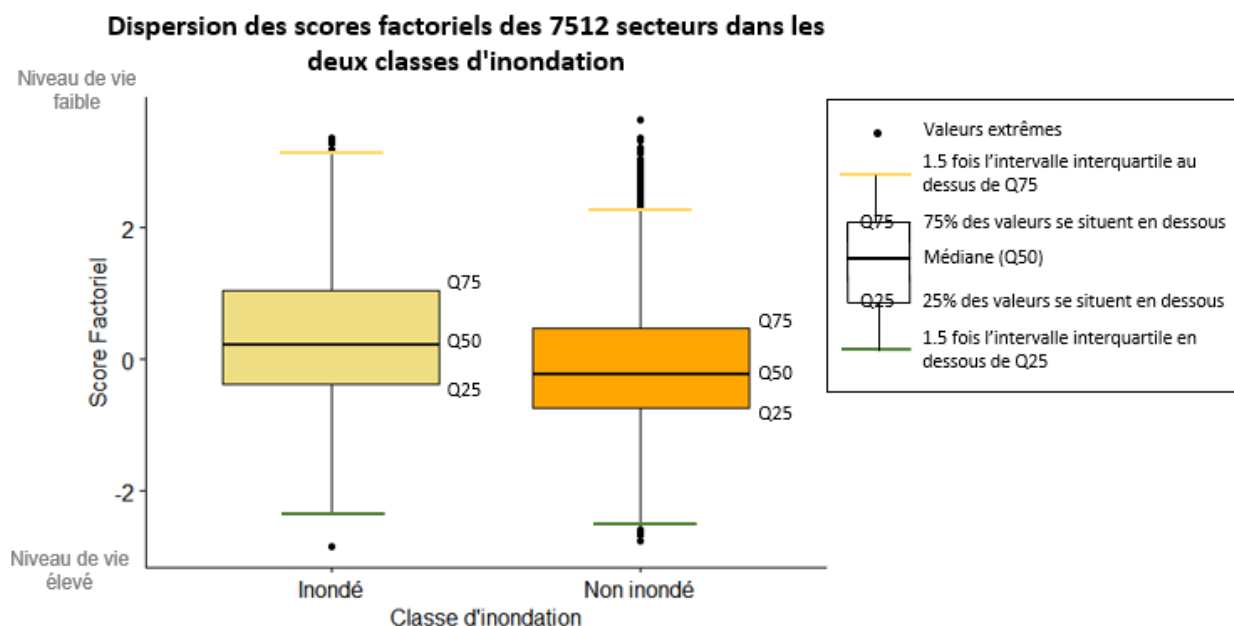


Figure 22: Dispersion du score factoriel des 7.512 secteurs statistiques dans les deux classes d'inondation

Comparaison des moyennes du score factoriel des deux groupes selon le statut d'inondation

Tableau 30: Résultat de l'ANOVA à 1 facteur (statut d'inondation) sur le niveau de vie

ANOVA à 1 facteur					
Comparaison des moyennes du score des 2 groupes formés par les secteurs inondés et non inondés à la moyenne générale					
	Degré de liberté	Somme des carrés	Variance	Statistique de Fisher	p-valeur
Statut d'inondation	1	199	199,45	208,8	< 2e-16
Résidus	7.510	7.174	0,96		

Le résultat de l'ANOVA (tableau 30) nous indique que la variance de la régression est supérieure à la variance résiduelle (intragroupe). La p-valeur est donc inférieure au seuil de confiance 0,05. Par conséquent, l'hypothèse H0 est rejetée : la moyenne d'au moins un des groupes (inondé- non inondé) est significativement différente de la moyenne générale.

Comparaison par couple de la moyenne des groupes formés par le statut d'inondation

Tableau 31: Comparaison couple de la moyenne des groupes formés par le statut d'inondation

Comparaison par paire de la moyenne des groupes formés par le statut d'inondation				
Test de Tukey				
	Écart des moyennes	IC _{inf}	IC _{sup}	p-valeur
Inonde - Non inondé	-0,419	-0,476	-0,362	< 2e-16

Étant donné que la p-valeur est inférieure au seuil de confiance 0,05, l'hypothèse nulle est rejetée. Le test de Tukey confirme que la moyenne des scores des secteurs inondés et celle des secteurs non inondés sont significativement différentes. Le niveau de vie serait donc influencé par son appartenance à un secteur inondé ou non (tableau 31).

Vérification des conditions d'application de l'ANOVA

Tableau 32: Test d'homogénéité de la variance intragroupe de l'ANOVA à 1 facteur (le statut d'inondation)

Test d'homogénéité de la variance intragroupe			
7.512 secteurs statistiques 1 facteur			
	Degré de liberté	Statistique de Fisher	p-valeur
Classe d'inondation	1	19,104	1,254e-05
Résidus	7.510		

La p-valeur étant inférieure à 0,05, le test d'homogénéité des variances rejette l'hypothèse nulle. Cela signifie que les variances intragroupes sont significativement différentes, rejetant ainsi l'hypothèse d'égale variance (tableau 32).

Tableau 33: Test de normalité des résidus de l'ANOVA à 1 facteur (le statut d'inondation)

Test de normalité sur les résidus		
Test de Kolmogorv-Smirnov sur l'ANOVA à 1 facteur		
	Statistique D	p-valeur
	0.0619	2.2e-16

La p-valeur étant inférieur à 0,05, l'hypothèse nulle est donc rejetée. Les résidus de l'ANOVA ne suivent pas une distribution normale (tableau 33 et figure 23).

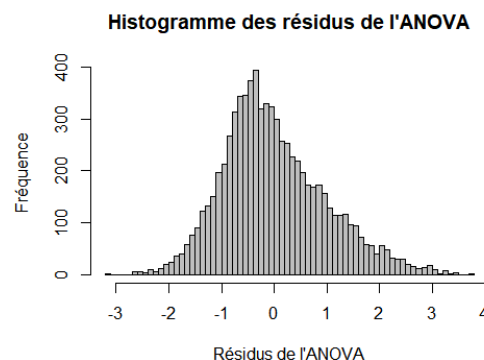


Figure 23: Histogramme des résidus de l'ANOVA à 1 facteur : le statut d'inondation

Test non paramétrique au vu du non-respect des conditions d'application de l'ANOVA à 1 facteur

Comme les deux conditions nécessaires à l'application de l'ANOVA, le test d'homogénéité des variances et le test de normalité des résidus, ne sont pas vérifiées, les tests de Kruskal-Wallis et de Wilcoxon ont été effectués en substituant la moyenne par la médiane. Le résultat de ces deux tests sont disponibles à l'annexe 24. Le premier compare la médiane du score des groupes et la médiane générale du score. Le deuxième compare les médianes des groupes. Les conclusions sont similaires à l'étude sur la moyenne. En effet, les hypothèses nulles sont rejetées : la médiane du score des groupes est significativement différente de la médiane générale du score et les médianes du score des groupes sont significativement différentes entre elles.

Conclusion de l'ANOVA à 1 facteur

L'ANOVA à un facteur nous indique que l'appartenance d'un secteur à un statut d'inondation, influence la valeur de son score factoriel. Le score factoriel et donc le niveau de vie des secteurs varient suivant leur statut « inondé » ou « non inondé ».

Il faut cependant souligner que les conditions nécessaires à l'application d'une ANOVA ne sont pas respectées. Nous avons tenté de remédier à ce problème en exploitant des tests comparables non paramétriques. Ces derniers étant plus laxistes quant au rejet des hypothèses nulles, il convient d'être prudent quant aux conclusions que l'on peut en tirer ;

De plus, il est important de remarquer que les boxplots de la figure 22 coïncident largement. Par conséquent, un grand nombre de secteurs détiennent un score factoriel similaire tout en étant situés dans des groupes de statut d'inondation différent.

III.2. ii. Anova à deux facteurs

Préambule : analyse à 1 facteur de la classe de Moran

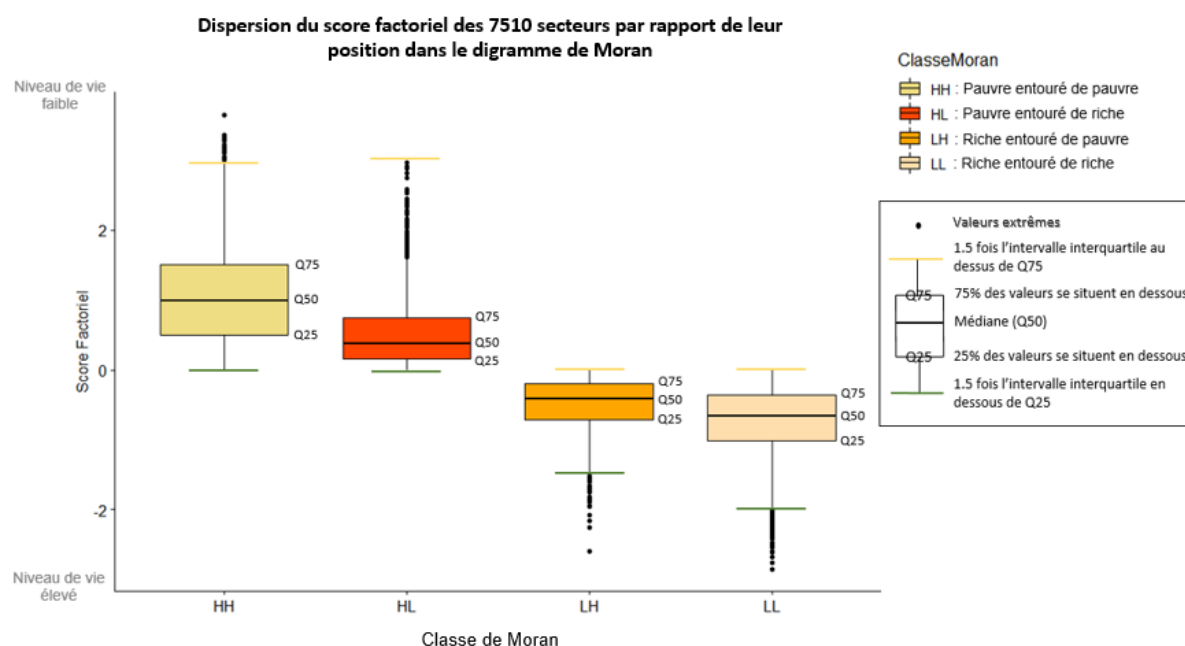


Figure 24: Dispersion du score factoriel des 7510 secteurs en fonction de leur position dans le diagramme de Moran

Tableau 34: Statistique descriptive du score factoriel des secteurs en fonction du quadrant où ils se trouvent dans le diagramme de Moran

Statistique descriptive du score factoriel des secteurs en fonction du quadrant où ils se trouvent dans le diagramme de Moran								
Classe Moran	Total	Moyenne	Écart-type	Q25	Médiane	Q75	Min	Max
LH	689	-0,515	0,413	-0,720	-0,415	-0,198	-0,00222	3,65
HH	2.148	1,09	0,73	0,501	0,985	1,50	-0,00369	2,97
LL	3.542	-0,738	0,490	-1,02	-0,665	-0,365	-2,59	-0,00780
HL	1.131	0,53	0,515	0,160	0,379	0,741	-2,85	-0,00512

Tout d'abord, la figure 24 et le tableau 34, s'intéressent uniquement à la dispersion dans les classes de Moran. La valeur du score est séparée en 2 sous-groupes dont les boxplots ne coïncident presque pas. Les secteurs dont les scores sont élevés (HH et HL) sont logiquement séparés des scores dont les scores sont faibles (LH et LL). Dans ces 2 sous-groupes, si le score du voisinage est pris en considération, on constate que la classe HH (0,985) possèdent une médiane du score supérieure à la classe HL (0,379) et que la classe LL (-0,665) possèdent une médiane du score inférieure à la classe LH (-0,415) (Figure 24).

Analyse à 2 facteurs : le statut d'inondation et la classe de Moran

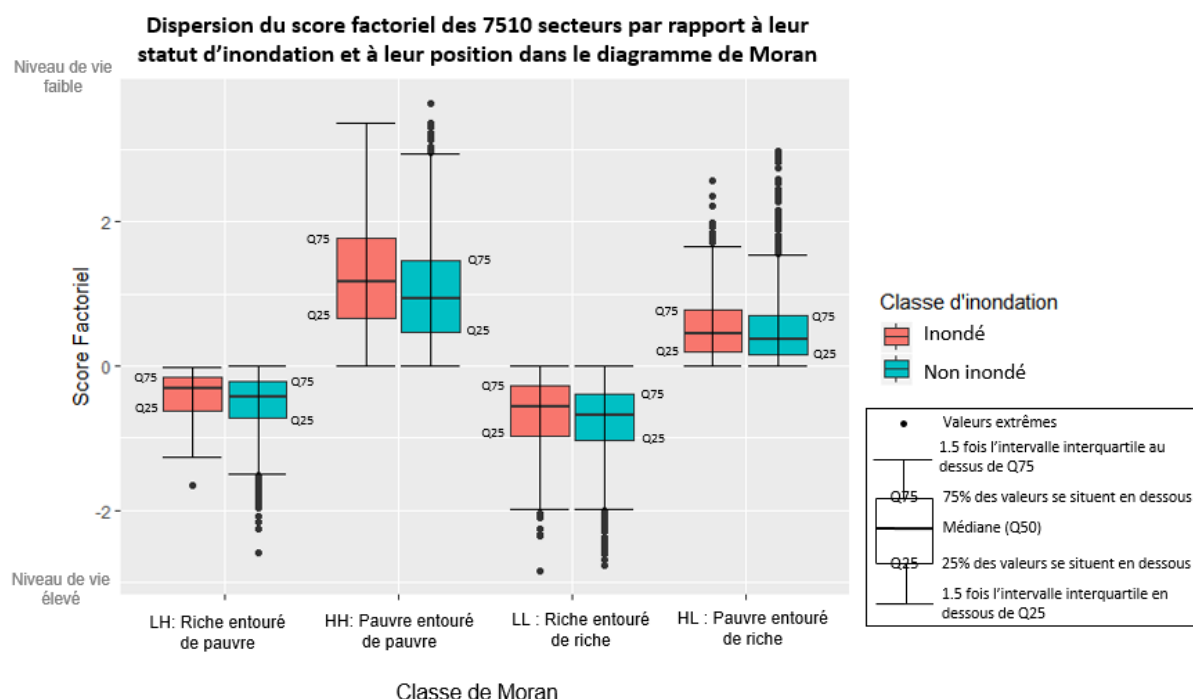


Figure 25: Dispersion du score factoriel des 7510 secteurs en fonction de leur statut d'inondation et de leur position dans le diagramme de Moran

La combinaison des deux facteurs du statut d'inondation et de la classe de Moran permet d'obtenir la figure 25 et le tableau des statistiques descriptives de l'annexe 25. Afin de simplifier la compréhension de la figure 25, les termes de « pauvres » et de « riches » ont été employés dans la légende. Néanmoins, il est évident que nous traitons toujours le score factoriel du niveau de vie et non le revenu. Vis-à-vis de la figure 24, des éléments supplémentaires sont apportés par le statut d'inondation. Dans chaque classe de Moran, les secteurs inondés ont toujours une médiane du score supérieure à la médiane du score des secteurs non inondés. Dès lors, dans chaque groupe, le niveau de vie est plus faible pour les secteurs inondés quelle que soit la classe de Moran. Par exemple, dans les secteurs dont le niveau de vie est faible entourés de secteurs dont le niveau de vie est également faible (HH), les secteurs inondés présentent une médiane des scores supérieure (1,77) aux secteurs non inondés (1,45). Rappelons que les scores élevés décrivent un niveau de vie faible.

Comparaison de la moyenne générale aux moyennes des groupes de secteurs établis selon leur statut d'inondation, d'un part, et leur position dans le diagramme de Moran d'autre part

Tableau 35: ANOVA à 2 facteurs (statut d'inondation et classe de Moran) sans interaction

ANOVA 2 facteurs sans interaction					
Comparaison de la moyenne générale aux moyennes des groupes de secteurs établis selon leur statut d'inondation, d'un part, et leur position dans le diagramme de Moran d'autre part					
	Degré de liberté	Somme des carrés	Variance	Statistique de Fisher	p-valeur
Classe de Moran	3	4.960	1.653,3	5.182,01	< 2e-16
Statut d'inondation	1	17	17,5	54,82	1,47e-13
Résidus	7.505	2.394	0,3		

Au terme de l'ANOVA, on constate que les p-valeurs des deux facteurs sont inférieures à 0,05. La variance intergroupe est donc supérieure à la variance intragroupe. L'hypothèse nulle est rejetée : la moyenne générale du score est significativement différente de la moyenne du score d'au moins une des classes de Moran et de la moyenne du score d'au moins une des classes du statut d'inondation (tableau 35).

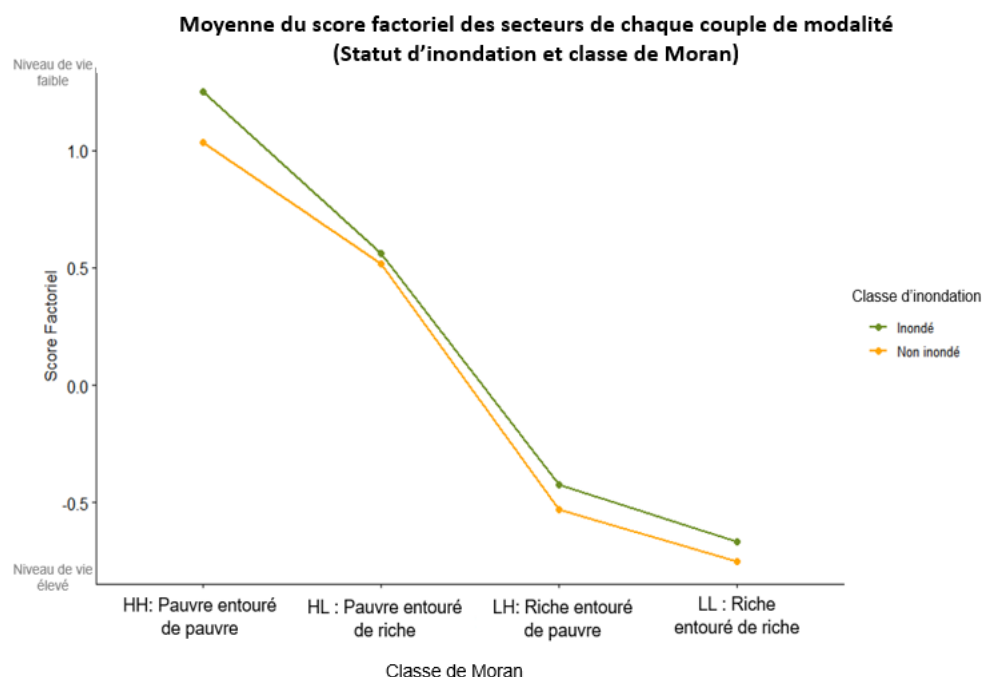


Figure 26: Représentation graphique de la moyenne du score factoriel des secteurs pour chaque couple de facteurs déterminant les groupes de l'ANOVA (statut d'inondation et classe de Moran)

La figure 26 permet par une première approche visuelle de déterminer s'il y a présence d'une interaction entre les deux facteurs. Les segments de droites n'étant pas parallèles, nous concluons qu'il existe une interaction entre la classe de Moran et le statut d'inondation. Pour vérifier cette hypothèse, une ANOVA prenant en compte l'interaction des deux facteurs a été réalisée (tableau 36).

On constate au tableau 36 que les deux facteurs et leur interaction sont significatifs. La variance résiduelle de chacun des termes est inférieure à la variance intergroupe. Dès lors, la moyenne générale est différente d'au moins une des moyennes de la classe de Moran ; la moyenne générale est différente d'au moins une des moyennes du statut d'inondation ; et la moyenne générale est différente d'au moins une des moyennes de la combinaison du statut d'inondation et de la classe de Moran.

L'effet d'interaction signifie que l'effet des deux facteurs de la classification de l'ANOVA (le statut d'inondation) dépend de l'autre (classe de Moran). En d'autres termes, le score factoriel est influencé conjointement par son appartenance à une catégorie de la classe de Moran et à son statut d'inondation.

Tableau 36: ANOVA à 2 facteurs (le statut d'inondation et la classe Moran) avec interaction

ANOVA 2 facteurs avec interaction					
Comparaison de la moyenne du statut d'inondation et des 4 quadrants de Moran et de leur interaction à la moyenne générale					
	Degré de liberté	Somme des carrés	Variance	Statistique de Fisher	p-valeur
Classe de Moran	3	4.960	1.653,3	5.192,21	< 2e-16
Classe d'inondation	1	17	17,5	54,92	1,39e-13
Classe de Moran : Classe d'inondation	3	6	1,9	5,92	0,000497
Résidus	7.502	2.389	0,3		

Comparaison par couple de la moyenne de score du statut d'inondation, de la moyenne du score des 4 quadrants de Moran et de la moyenne du score des groupes formés par l'interaction du statut d'inondation et des 4 quadrants de Moran

Le test de Tukey permet de réaliser une comparaison multiple des moyennes des modalités de chacun des facteurs. Le tableau 37 reprend la comparaison des moyennes de chaque groupe entre elles. La comparaison des moyennes sans interaction correspond aux 7 premières lignes (de HH à Non inondé-inondé). Ensuite, viennent les comparaisons tenant compte de l'interaction.

Les moyennes des classes de Moran sont toujours toutes significativement différentes entre elles. La différence de moyennes des secteurs inondés et non inondés est également significative. Parmi les interactions, seules 3 combinaisons de moyennes ne sont pas significatives. La moyenne du score des secteurs riches inondés entourés de secteurs pauvres (LH : inondé) n'est pas significativement différente de la moyenne des secteurs riches non inondés entourés de secteurs pauvres (LH : non inondé). La moyenne du score des secteurs pauvres inondés entourés des secteurs riches (HL : inondé) n'est pas significativement différente de la moyenne du score des secteurs pauvres non inondés entourés de secteurs riches (LH : non inondé). La moyenne du score des secteurs pauvres inondés entourés des secteurs pauvres (HH : inondé) n'est pas significativement différente de la moyenne du score des secteurs pauvres non inondés entourés de secteurs pauvres (HH : non inondé).

En d'autres termes, la moyenne du score factoriel des classes du diagramme de Moran LL, LH, HL n'est pas influencée par le statut d'inondation. Plus particulièrement, la moyenne du score des secteurs riches entourés de secteurs riches (LL), n'est pas influencée par le statut d'inondation du secteur. À l'inverse, les secteurs pauvres entourés de pauvres (HH) détiennent une moyenne du score factoriel significativement différente lorsque le secteur est inondé ou non inondé. Cette dernière constatation va dans le sens de notre hypothèse de départ.

Tableau 37: Comparaison par paire des moyennes des groupes avec interaction dans le cadre d'une ANOVA    2 facteurs
(statut d'inondation et classe de Moran)

Comparaison par paire des moyennes des scores des 2 facteurs le statut d'inondation et la classe de Moran avec interaction Test de Tukey				
Paire de groupes	��cart des moyennes	IC _{inf}	IC _{sup}	p-valeur
HH - LH	1,601	1,537	1,664	0
LL- LH	-0,223	-0,283	-0,162	0
HL-LH	1,045	0,975	1,115	0
LL-HH	-1,823	-1,863	-1,784	0
HL-HH	-0,556	-0,609	-0,503	0
HL -LL	1,267	1,218	1,317	0
Non inond��-Inond��	-0,123	-0,156	-0,090	0
HL : Inond�� – HH : Inond��	-0,688	-0,813	-0,564	0
LH : Inond�� – HH : Inond��	-1,673	-1,864	-1,483	0
LL : Inond�� – HH : Inond��	-1,916	-2,024	-1,808	0
HH : Non inond�� – HH : Inond��	-0,218	-0,304	-0,132	0
HL : Non inond�� – HH : Inond��	-0,732	-0,827	-0,637	0
LH : Non inond�� – HH : Inond��	-1,780	-1,882	-1,678	0
LL : Non inond�� – HH : Inond��	-1,999	-2,080	-1,919	0
LH : Inond�� – HL : Inond��	-0,985	-1,187	-0,783	0
LL : Inond�� – HL : Inond��	-1,228	-1,355	-1,101	0
HH : Non inond�� – HL : Inond��	0,470	0,362	0,579	0
HL : Non inond�� – HL : Inond��	-0,043	-0,159	0,073	0,950
LH : Non inond�� – HL : Inond��	-1,092	-1,214	-0,970	0
LL : Non inond�� – HL : Inond��	-1,311	-1,415	-1,206	0
LL : Inond�� – LH : Inond��	-0,243	-0,435	-0,050	0,003
HH : Non inond�� – LH : Inond��	1,455	1,275	1,636	0
HL : Non inond�� – LH : Inond��	0,942	0,757	1,127	0
LH : Non inond�� – LH : Inond��	-0,107	-0,296	0,082	0,680
LL : Non inond�� – LH : Inond��	-0,326	-0,504	-0,147	8,68E-07
HH : Non inond�� – LL : Inond��	1,698	1,609	1,787	0
HL : Non inond�� – LL : Inond��	1,184	1,086	1,283	0
LH : Non inond�� – LL : Inond��	0,136	0,031	0,241	0,00235212
LL : Non inond�� – LL : Inond��	-0,083	-0,168	0,001	0,05649998
HL : Non inond�� – HH : Non inond��	-0,513	-0,586	-0,441	0
LH : Non inond�� – HH : Non inond��	-1,562	-1,644	-1,480	0
LL : Non inond�� – HH : Non inond��	-1,781	-1,834	-1,728	0
LH : Non inond�� – HL : Non inond��	-1,048	-1,140	-0,957	0
LL : Non inond�� – HL : Non inond��	-1,267	-1,334	-1,201	0
LL : Non inond�� – LH : Non inond��	-0,219	-0,296	-0,142	0

Vérification des conditions d'utilisation de l'ANOVA à 2 facteurs

Tableau 38: Test d'homogénéité de la variance intragroupe de l'ANOVA à 2 facteurs (statut d'inondation et classe de Moran)

Test d'homogénéité de la variance intragroupe 7.512 secteurs statistiques			
	Degré de liberté	Statistique de Fisher	p-valeur
	7	77.669	< 2.2e-16
	7.502		

La p-valeur étant inférieure à 0,05, le test d'homogénéité des variances rejette l'hypothèse nulle. Cela signifie que les variances intragroupes sont significativement différentes, rejetant ainsi l'hypothèse d'égale variance (tableau 38)

Tableau 39: Test de normalité des résidus de l'ANOVA à 2 facteurs (statut d'inondation et classe de Moran)

Test de normalité sur les résidus Test de Kolmogorov-Smirnov sur l'ANOVA à 2 facteurs		
	Statistique D	p-valeur
	0,0345	2,2e-16

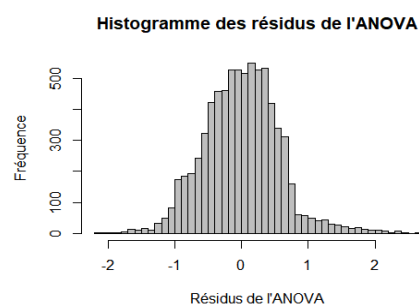


Figure 27: Histogramme des résidus de l'ANOVA à deux facteurs (Statut d'inondation et Classe de Moran)

La p-valeur étant inférieure à 0,05, H_0 est donc rejetée. Les résidus de l'ANOVA ne suivent pas une distribution normale (tableau 39 et figure 27).

Test non paramétrique au vu du non-respect des conditions d'application de l'ANOVA à 2 facteurs

Les conditions nécessaires à l'application de l'ANOVA n'étant pas réunies, nous devrions effectuer un test sur la médiane du score de chaque groupe. Cependant, le test n'est pas réalisable conjointement sur deux facteurs. Pour y remédier, avec un test de Kruskal-Wallis, nous comparons la médiane des groupes d'un facteur à la fois à la médiane générale de ce facteur. Ensuite, avec un test de Wilcoxon, nous comparons par paire les médianes du score des groupes des deux facteurs pris séparément. Ces tests ont déjà été exécutés sur le statut d'inondation à l'annexe 24. Les deux tests sur la médiane sont réalisés sur les 4 quadrants de Moran. Les résultats sont disponibles à l'annexe 26. Pour les deux facteurs, le résultat du test de Kruskal-Wallis confirme qu'au moins une médiane du score des groupes est significativement différente de la médiane générale. Pour les 2 facteurs, le test de Wilcoxon rejette l'hypothèse nulle pour chaque paire de médiane du score des groupes : l'ensemble des médianes sont différentes entre elles. Nous remarquons donc que nous ne savons pas prendre en compte l'interaction des deux facteurs dans aucun des deux tests.

ANOVA à 2 facteurs sur un nombre d'observations distinct par groupe

Comme le nombre de secteurs statistiques contenu dans chaque modalité diffère (annexe 25), un test particulier, dit « de type III » dans la littérature, est utilisé. Celui-ci réalise une ANOVA à 2 facteurs, il compare donc la moyenne des 8 groupes avec la moyenne générale. Dans le premier test du tableau 40, sans interaction, le statut d'inondation et le contexte spatial (diagramme de Moran) présentent des p-valeurs inférieures au seuil de significativité 0,05. La moyenne du score de chaque groupe est différente de la moyenne générale du score. En ajoutant l'interaction dans l'équation (tableau 41), le statut d'inondation perd sa significativité, sa p-valeur devient supérieure à 0,05. Donc le calcul de l'ANOVA a été influencé par le nombre d'individus différents dans les groupes. La moyenne du score des secteurs inondés n'est pas différente de la moyenne générale du score et la moyenne du score des secteurs non inondés n'est pas différente de la moyenne générale du score.

Cependant, l'interaction de deux facteurs est significative. Cela se traduit par le fait qu'au moins une des moyennes du score des groupes formés par l'interaction entre le statut d'inondation et la classe de Moran est significativement différente de la moyenne générale. Malheureusement, avec ce test, il n'existe pas de comparaisons multiples des moyennes prenant en compte le nombre de secteurs distincts par groupe. Dès lors, il nous est impossible de déterminer les couples de groupes présentant des moyennes différentes. Ceci aurait pourtant été très utile.

Tableau 40: ANOVA 2 facteurs type III : Comparaison de la moyenne des 8 groupes à la moyenne générale sur 7.510 secteurs statistiques

ANOVA 2 facteurs type III sans interaction Comparaison de la moyenne des 8 groupes à la moyenne générale Sur 7.510 secteurs statistiques				
	Degré de liberté	Somme des carrés	Statistique de Fisher	p-valeur
Ordonnée à l'origine	1	78,1	244,649	<2e-16
Classe de Moran	3	4778,1	4992,159	<2e-16
Classe d'inondation	1	17,5	54,816	1,468e-13
Résidus	7.505	2.394,4		

Tableau 41: ANOVA 2 facteurs type III : Comparaison de la moyenne des 8 groupes à la moyenne générale sur 7510 secteurs statistiques avec interaction

ANOVA 2 facteurs type III avec interaction Comparaison de la moyenne des groupes à la moyenne générale Sur 7.510 secteurs statistiques				
	Degré de liberté	Somme des carrés	Statistique de Fisher	p-valeur
Ordonnée à l'origine	1	17,02	53,4600	<2e-16
Classe de Moran	3	987,68	1033,9502	<2e-16
Classe d'inondation	1	0,93	2,9258	0,0872141
Classe de Moran : Classe d'inondation	3	5,66	5,9202	0,0004974
Résidus	7.502	2.388,75		

Conclusion de l'ANOVA à 2 facteurs

Sans interaction, peu importe le test effectué, les moyennes de chaque groupe des deux facteurs sont significativement différentes de la moyenne générale. Le même constat est à formuler pour les comparaisons par paire : sans interaction les moyennes de chaque groupe des deux facteurs sont significativement différentes entre elles.

Avec interaction, les résultats sont plus nuancés. Le test « de type III » qui prend en compte le nombre distinct d'observations par groupe, démontrent que ni la moyenne du score du groupe inondé ni celle du groupe non inondé ne sont significativement différentes de la moyenne générale. Par conséquent, le facteur de statut d'inondation pris individuellement n'influence pas le score factoriel du niveau de vie. Cependant, le même test excepté qu'il ne prend pas en compte le nombre distinct d'observations, prétend que les groupes formés par l'interaction entre le statut d'inondation et le contexte spatiale déterminé par la position dans le diagramme de Moran influencent la valeur du score factoriel du niveau de vie. Donc l'appartenance d'un secteur à un statut d'inondation particulier et à un contexte spatial (diagramme de Moran) particulier influence son score factoriel. Un autre test, celui de la comparaison par paire des moyennes, révèle que l'interaction du statut d'inondation dans un même contexte spatial (diagramme de Moran) n'influence pas le score du secteur excepté pour le contexte spatiale HH. En effet, les secteurs appartenant au quadrant HH (pauvre entouré de pauvres) ont un score plus élevé, si leur statut d'inondation est inondé et ils ont un score plus faible si leur statut d'inondation est non inondé.

Dès lors la tendance générale des résultats de l'ANOVA à 2 facteurs va dans le sens de l'hypothèse de l'étude, c'est-à-dire que les secteurs inondés ont un niveau de vie plus faible que les secteurs non inondés, avec la contrainte que ces secteurs soient pauvres et entourés d'autres secteurs pauvres.

Conclusion générale sur les résultats des deux ANOVA

Tout d'abord, il ne faut pas oublier que les conditions d'application de l'ANOVA ne sont pas respectées. Des tests complémentaires sur la médiane ont été effectués mais ils ne remplacent pas tous les tests sur la moyenne effectués par l'ANOVA. En outre, les figures 22, 24 et 25 montrent clairement que les boxplots coïncident c'est-à-dire que de nombreux secteurs statistiques ont la même valeur de score factoriel mais appartiennent à des groupes distincts. Il ne faut pas oublier que l'ANOVA reste une comparaison de moyennes de groupes.

Si l'on se permet d'ignorer ces précautions, les résultats des ANOVAs tendent à affirmer que le statut d'inondation – inondé ou non inondé – influence le score factoriel d'un secteur. Cette affirmation est d'autant plus vraie pour les secteurs de niveau de vie faible entourés d'autres secteurs de niveau de vie faible.

III. 3. Régression Logit du score factoriel

La régression Logit est effectuée sur une variable dépendante binaire, le statut d'inondation dont les modalités sont « inondé » (1) ou « non inondé » (0). La variable indépendante est la variable quantitative du niveau de vie définie par le score factoriel. Il ne s'agit évidemment pas de prédire le statut d'inondation par le niveau de vie d'un secteur ! Le but est de prédire l'occurrence des événements « inondés » et « non inondés » en fonction du niveau de vie et d'estimer l'influence du niveau de vie. En d'autres termes, est-ce que le fait que la population soit dans un secteur inondé est expliqué par son relativement faible niveau de vie ?

Tableau 42: Résultat de la régression Logit sur le statut d'inondation Y

Régression Logit sur le statut d'inondation (Y) Nombre de secteurs : 7.512 X= Score factoriel du niveau de vie				
Coefficient	Estimation des coefficients	Erreur standard	Valeur de z	p-valeur de z
Intercepte	-1,5307	0,03092	-49,50	< 2e-16
Score factorielle	0,4104	0,02929	14,01	< 2e-16
Déviance	7.199	Degré de liberté	7.511	
Déviance résiduelle	7.002	Degré de liberté	7.510	
AIC	7.006			

Tableau 43: Test de significativité des odds ratio du modèle Logit où Y= le statut d'inondation et X= le score factoriel

Test de la significativité des odd ratio				
Coefficient	Odd ratio	2,5 %	97,5%	p-valeur de z
Intercepte	0,216	0,03092	-49,50	< 2e-16
Score factoriel	1,507	0,02929	14,01	< 2e-16

Les résultats du test indiquent que leur p-valeur est inférieure à 0,05 ; que l'odd ratio (1,5) diffère significativement de 1 et que le coefficient du score factoriel (0,41) diffère significativement de 0,5.

Il est possible d'interpréter la signification de l'odd ratio afin de déterminer l'influence du score factoriel du niveau de vie dans la prédiction du statut d'inondation. L'odd ratio étant égal à 1,5 (> 1), cela peut se traduire par « quand le score factoriel augmente (faible niveau de vie), la prédiction que y = 1, donc que le statut d'inondation soit inondé, augmente également ».

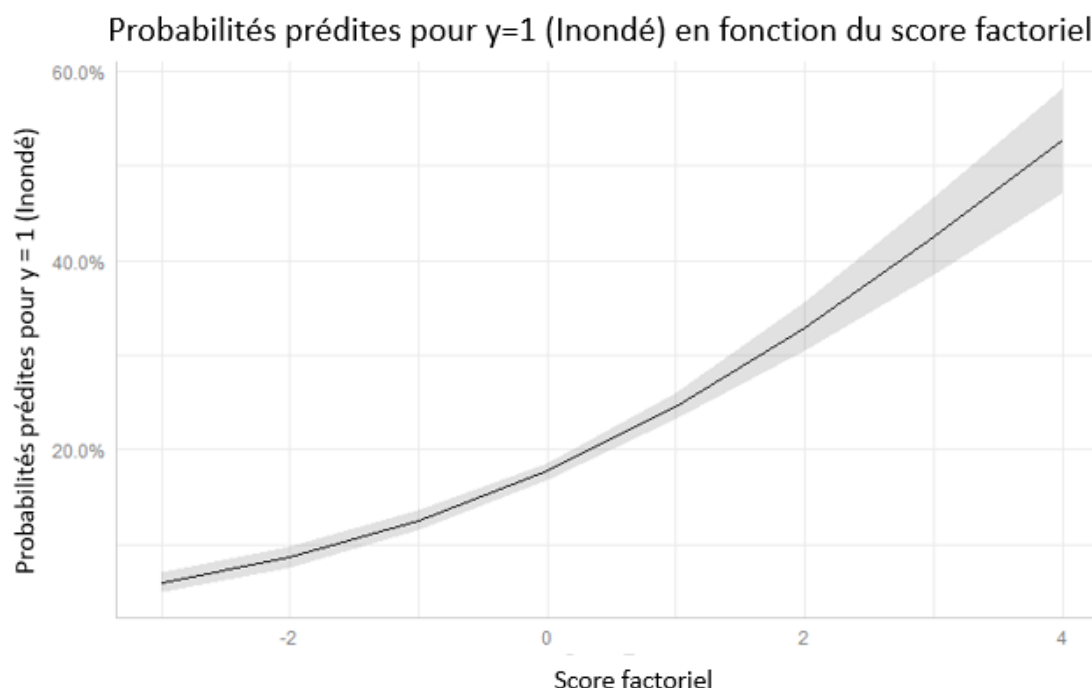


Figure 28: Probabilités prédites pour $y=1$ (inondé) en fonction du score factoriel suite à la régression logit

Tableau 44: Matrice de confusion issue du test de prédiction des valeurs supérieures à 0,5

Matrice de confusion		
	Non inondé (0)	Inondé (1)
Faux	6.121	1.391
Vrai	0	0

La matrice de confusion (tableau 44) correspond au tableau croisé des valeurs observées et celles des valeurs prédites en appliquant le modèle aux données d'origine. Si les valeurs prédites pour $y = 1$ sont supérieures à 0,5, le secteur est classé en « vrai ». Or comme nous pouvons le constater, les effectifs positifs classés en négatif sont au nombre de 1.391. Cela signifie qu'aucune probabilité de $y = 1$ n'est supérieure à 0,5. En effet, comme on peut l'observer sur la figure 28 les valeurs prédites pour $y = 1$ supérieures à 0,5, correspondent à un score factoriel supérieur à 3. Or le score factoriel supérieur à 3 correspond à des valeurs extrêmes de la valeur du niveau de vie.

Conclusion de la Logit

Malgré la significativité des coefficients et des odd ratio de la régression, le modèle n'est pas prédictif. En d'autres termes, un secteur dont le niveau de vie est faible n'est pas localisé dans les seules zones inondables. Le fait d'être « pauvre » au sens large du terme (pas seulement économique) n'est pas discriminant. Donc les secteurs statistiques dont le niveau de vie est faible ne sont pas systématiquement inondés. Plus généralement la notion de faible niveau de vie ne signifie pas / n'est pas strictement liée aux zones d'inondation.

CHAPITRE 6: CONCLUSION

Pour rappel, la question à laquelle tente de répondre de cette recherche était : **« Quelles sont les caractéristiques socio-économiques de la population exposée au risque d'inondation ? »**. L'hypothèse sous-jacente est que les zones inondables sont principalement occupées par des populations caractérisées par un profil socio-économique précaire.

Dans un premier temps, il est nécessaire de définir la notion de caractéristiques socio-économiques de la population et de fixer l'emprise spatiale de la ou des variables capables de traduire ce concept. Du point de vue spatial, il nous a été demandé de traiter la totalité du territoire wallon. Dès lors, la plus petite entité spatiale pour laquelle des données socio-économiques diverses étaient disponibles, est le secteur statistique. Ensuite, en ce qui concerne la notion de caractéristiques socio-économiques, deux points de vue ont été traités en parallèle. D'une part, nous avons réduit la notion au seul revenu des ménages (revenu médian par déclaration par secteur statistique). Le critère est quantitatif, non ambigu, stable dans le temps et dans l'espace à différents niveaux spatiaux. D'autre part, nous avons travaillé sur un indice synthétique de niveau de vie constitué par la première composante principale d'une matrice de données socio-économiques disponibles par secteur statistique. Le critère est également quantitatif (les scores factoriels sont standardisés) et, conceptuellement, couvre plus que le seul aspect financier. Cependant, il est construit de manière exploratoire, avec un certain arbitraire, et rien ne garantit sa stabilité spatiale et temporelle.

Ensuite, il est nécessaire de définir le risque d'inondation. Nous disposons de plusieurs données rassemblées par le laboratoire HECE de l'Université de Liège, parmi lesquelles une carte des aléas classés d'inondation en Wallonie, et les hauteurs d'eau à trois périodes de retour, rasterisées à la résolution de 5 m, individuellement pour tous les grands cours d'eau de Wallonie. Ces dernières données quantitatives n'ayant pu être fusionnées de façon satisfaisante sur tout le territoire, seule la carte des aléas d'inondation a été utilisée. Une rapide analyse statistique et cartographique des aléas nous a amené à reclasser les secteurs statistiques en deux catégories : « inondés » et « non inondés ».

Afin de tester l'hypothèse de recherche, deux protocoles de traitements ont été imaginés, un pour les revenus, l'autre pour l'indice synthétique fourni par le score factoriel.

Dans le cas des revenus, l'hypothèse de recherche consiste à prétendre que les ménages aux revenus les plus bas résident prioritairement dans les secteurs statistiques qualifiés d'inondés. Le protocole dédié aux revenus repose sur des analyses de régression. Le revenu médian par secteur statistique est susceptible d'être largement expliqué par trois variables socio-économiques qui sont : la part de la population disposant d'un diplôme de l'enseignement supérieur, à la part de ménages d'une seule personne et le taux de chômage. Il existe certes un résidu non négligeable et toutes les variables, ainsi que le résidu, sont entachés d'une autocorrélation spatiale sensible. L'idée est donc de tenter de réduire significativement les résidus en introduisant parmi les variables explicatives, le caractère inondable et le phénomène d'autocorrélation spatiale. Plusieurs modèles de régressions multiples ont ainsi été exécutés, en ajoutant aux trois variables explicatives de départ, des variables auxiliaires traduisant le statut d'inondation des secteurs statistiques et leurs modalités de voisinage (quadrants de Moran), ainsi que toutes les interactions entre les variables indépendantes et auxiliaires. Les résultats n'ont pas été jugés significatifs : le caractère inondé ou non d'un secteur statistique n'influence pas la variance des revenus médians des ménages qui y résident. Afin de réduire l'autocorrélation spatiale des résidus, une

auto-régression spatiale de type LAG a également été tentée. La variable endogène décalée est cependant peu efficace et la répartition spatiale des résidus semble indiquer que la distribution des revenus est soumise à un phénomène régional, plutôt que local comme le serait l'influence des zones inondables. Pour tenter de limiter cet effet régional, la totalité du protocole des analyses de régression a été de nouveau appliqué sur une nouvelle variable réponse constituée par l'écart relatif des revenus des secteurs statistiques par rapport au revenu de leur commune. Cette approche fait apparaître un rôle statistiquement significatif au statut d'inondation des secteurs statistiques mais sa contribution à la variance de la variable réponse reste excessivement limitée. De plus, les résidus de l'auto-régression LAG menée dans cette partie, indiquent l'influence d'un nouveau phénomène supra-local, vraisemblablement lié à la distribution particulière des zones résidentielles aux alentours des grandes agglomérations.

Nous nous sommes alors penché sur l'analyse de la variable synthétique du niveau de vie. L'hypothèse de recherche consiste à prétendre que les secteurs statistiques inondés possèdent des populations dont le niveau de vie est jugé plus précaire. Le protocole d'analyse se doit d'être différent du précédent, puisque, dans ce cas, toutes les variables socio-économiques utiles ont été rassemblées par l'ACP. Il faut donc comparer directement le score factoriel du niveau de vie avec le caractère inondable des secteurs de résidence. Comme on constate également une très forte autocorrélation spatiale dans la répartition du score factoriel, cet aspect est aussi intégré à l'analyse. Les méthodes statistiques sont évidemment moins puissantes dans ce second protocole puisqu'une seule variable quantitative est disponible. Tout d'abord, un simple test du χ^2 est effectué, après avoir classé le niveau de vie en 5 classes. Ensuite des analyses de variance à un facteur (caractère inondable) et à deux facteurs (caractère inondable et contexte spatial d'autocorrélation du secteur) sont effectuées. Des tests équivalents non paramétriques sont réalisés lorsque les conditions d'application de l'analyse de variance ne sont pas remplies. Dans la majorité des cas, les tests tendent à affirmer que le statut d'inondation – inondé ou non inondé – influence le score factoriel d'un secteur. Cette affirmation est d'autant plus vraie pour les secteurs de niveau de vie faible entourés d'autres secteurs de faible niveau de vie. L'hypothèse de départ semblerait donc validée selon ce protocole. Il reste que plusieurs cas d'ambiguïté persistent, que l'indice de niveau de vie est, par construction, une variable relativement instable, et que les tests effectués, qui n'ont pas la puissance des méthodes paramétriques, ne sont pas unanimes. Au terme de cette seconde partie, une dernière méthode a été appliquée, pour tenter d'exploiter au mieux le caractère quantitatif du score factoriel. Il s'agit une régression Logit dont l'objectif est de déterminer si le fait que la population soit dans un secteur inondé est expliqué par son relativement faible niveau de vie. Les résultats de cette dernière analyse sont parfaitement cohérents, mais le modèle – comme on s'y attendait un peu... - ne peut prétendre être prédictif : la notion de faible niveau de vie ne signifie pas / n'est pas strictement liée aux zones d'inondation.

Les résultats de l'étude sont donc mitigés. Même si les tests les plus simples ont tendance à valider l'hypothèse, il est délicat d'affirmer à ce stade que la population dont les caractéristiques socio-économiques sont faibles, se localise préférentiellement dans les zones présentant un risque sensible d'inondation.

Plusieurs aspects de l'étude pourraient être revisités pour tenter de valider l'hypothèse de façon plus nette.

Tout d'abord, d'un point de vue spatial, le fait de vouloir traiter toute la Wallonie en une fois s'est avéré peu efficace, les secteurs statistiques présentant une beaucoup trop grande hétérogénéité, et les

variables de hauteur d'eau n'ayant pu être exploitées. Une approche utilisant un plan d'échantillonnage par grappes (vallées, communes ou parties de communes...), et l'emprise des bâtiments résidentiels, et non la superficie des secteurs statistiques, constituerait sans doute, du point de vue spatial, une alternative intéressante. Cependant, elle se heurterait vraisemblablement à la difficulté d'obtention des critères socio-économiques, sinon peut-être par enquêtes (les ressources de l'administration cadastrale étant peu sujettes à être diffusées).

Ensuite, du point de vue thématique – celui du risque d'inondation – il est clair que les données ont été sous-exploitées. Pour des raisons techniques liées au mode d'acquisition et de préparation des données, il n'a pas été possible de traiter les hauteurs d'eau sur la totalité du territoire imposé. Ces variables quantitatives ont fait défaut dans notre étude. D'autre part, la classification des zones inondables, réduite *in fine* à deux catégories, laisse une distribution totalement déséquilibrée des secteurs statistiques. Le trop grand nombre de secteurs qualifiés d'inondés renforce d'ailleurs la critique précédente portant sur l'emprise spatiale de l'analyse.

Enfin, on peut revenir sur la question de recherche et la définition de l'hypothèse qui, bien qu'alléchantes a priori, s'avèrent somme toute assez naïves. La répartition de la population ne résulte pas d'un processus aléatoire ! Question de recherche et hypothèse excluent, en effet, toute réflexion préalable sur les modalités d'installation, en particulier historiques, de la population résidentielle. La régression Logit réalisée dans cette étude démontre bien que la population caractérisée par un niveau de vie faible n'est pas systématiquement localisée dans les secteurs inondés, tandis que les auto-régressions spatiales laissent suspecter divers effets régionaux ou, en tout cas, supra-locaux, pour expliquer la répartition des revenus. Or les analyses menées sur la totalité des secteurs statistiques de Wallonie n'auraient été légitimement interprétables que « toutes autres choses étant égales »...

Bibliographie

Anselin, L. & Rey, S.J. (2014). *Modern Spatial Econometrics in Practice*. Chicago : GeoDa Press LLC, 394 p.

Bailey, T.C. & Gatrell, A.C. (1995). *Interactive Spatial Data Analysis*. Harlow : Longman Scientific and Technical, 432 p.

Baud, I.S.A., Pfeffer, K., Sridharan, N. & Nainan, N. (2009). Matching Deprivation Mapping To Urban Governance In Three Indian Mega-Cities. *Habitat International*, 33 (4), 365-377.
<https://doi.org/10.1016/j.habitatint.2008.10.024>

Beckers, A., Dewals, B., Erpicum, S., Dujardin, S., Detrembleur, S., Teller, J., Pirotton, M. & Archambeau, P. (2013). Contribution Of Land Use Changes To Future Flood Damage Along The River Meuse In The Walloon Region. *Natural Hazards And Earth System Sciences*, 13, 2301-2318. <https://doi.org/10.5194/nhess-13-2301-2013>

Berry, B. (1971). Introduction: The Logic and Limitations of Comparative Factorial Ecology. *Economic Geography*, 47(2), 209 p. <https://doi.org/10.2307/143204>

Bivand, R.S., Pebesma, E. & Gomez-Rubio, V., (2013). *Applied Data Analysis with R*. 2 ed. New York : Springer, 405 p.

Caloz, R. & Collet, C. (2011). *Analyse spatiale de l'information géographique*. Lausanne : Presses Polytechniques et Universitaires Romandes, Science et ingénierie de l'environnement, 384 p.

Cellule Etat de l'Environnement Wallon (CEEW). (2007). *Rapport analytique sur l'état de l'environnement wallon 2006-2007*. Namur : Ministère de la Région wallonne - Direction générale des Ressources naturelles et de l'Environnement, 736 p.
[http://etat.environnement.wallonie.be/files/Publications/Rapport%20analytique%202006-2007/Rapport_analytique\(1\).pdf](http://etat.environnement.wallonie.be/files/Publications/Rapport%20analytique%202006-2007/Rapport_analytique(1).pdf). Consulté le 11 janvier 2019.

Coninx, I. & Bachus, K. (2007). *Integrating Social Vulnerability To Floods In A Climate Change Context*. In CAIWA 2007 International Conference On Adaptive & Integrated Water Management. Bâle : Osnabrück University, 30 p.
http://dev.ulb.ac.be/ceese/ADAPT/public_section/Doc/Doc/Coninx_and_Bachus.pdf. Consulté le 20 juillet 2019.

Cornet, Y., Binard, M., Ledant, M., Van Der Kwast, J. & Van de Voorde, T. (2012). Predicting The Spatial Distribution Of Population Based On Impervious Surface Maps And Modeled Land Use Change. In Perakis, K.G. & Moysiadis, A.K. (eds), *European Association of Remote Sensing Laboratories 32nd EARSeL Symposium Proceedings Advances, Mykonos, 2012*. Volos : Université de Thessalie, 575-586.
https://orbi.uliege.be/bitstream/2268/144921/1/Cornet_et_al_2012.pdf. Consulté le 20 juillet 2019.

Cutter, S.L., Boruff, B.J. & Shirley, W.L. (2003). Social Vulnerability To Environmental Hazards. *Social Science Quarterly*, 84 (2), 242–261. <https://doi.org/10.1111/1540-6237.8402002>

Cutter, S.L., Mitchell, J.T. & Scott, M.S. (2000). Revealing The Vulnerability Of People And Places: A Case Study Of Georgetown County, South Carolina. *Annals Of The Association Of American Geographers*, 90 (4), 713-737.
<https://doi.org/10.1111/0004-5608.00219>

Dagnelie, P. (1973). *Théorie et méthodes statistiques, 2 tomes*. Gembloux : Les Presses agronomiques de Gembloux, 378 p.

Dagnelie, P. (1977). *Analyse statistique à plusieurs variables*. Gembloux : Les Presses agronomiques de Gembloux, 362 p.

De Silva, M.M.G.T. & Kawasaki, A. (2018). Socioeconomic Vulnerability To Disaster Risk: A Case Study Of Flood And Drought Impact In A Rural Sri Lankan Community. *Ecological Economics*, 152, 131-140.
<https://doi.org/10.1016/j.ecolecon.2018.05.010>

De Smith, M., Goodchild, M.F. & Longley, P.A. (2018). *Geospatial Analysis*. 6 ed. Londres : Winchelsea Press, 602 p.

Detrembleur, S., Dewals, B., Archambeau, P., Erpicum, S. & Piroton, M. (2008). *Modélisation numérique et gestion du risque inondation en milieu urbain sur base de modèles topographiques à haute résolution*. In 3e journées doctorales en hydrologie urbaine, Neustrie, 2018. Bouguenais Cedex : Laboratoire Eau et Environnement de l'IFSTTAR. <http://hdl.handle.net/2268/35259>

Dewals, B., Archambeau, P., Erpicum, S. & Piroton, M. (2015). Modélisation hydraulique détaillée d'inondations extrêmes sur un tronçon transnational de la Meuse. *La Houille Blanche*, 2, 78-84.
<https://doi.org/10.1051/lhb/20150022>

Dewals, B., Detrembleur, S., Archambeau, P., Erpicum, S., Ernst, J. & Piroton, M. (2011). Caractérisation micro-échelle du risque d'inondation : modélisation hydraulique détaillée et quantification des impacts socio-économiques. *La Houille Blanche*, 2, 28-34. <https://doi.org/10.1051/lhb/2011015>

Dixon, W.J. & Massey, F.J. (1983). *Introduction To Statistical Analysis*. 4 ed. Singapour : McGraw-Hill International Student Edition, 672 p.

Donnay, J-P. (2015). *Mode Zonal*. Analyse spatiale, Liège, Université de Liège, inédit, 40 p.

Donnay, J-P. (2019a). *Analyse variance et régressions simples*. Compléments d'analyse spatiale, Liège, Université de Liège, inédit, 29 p.

Donnay, J-P. (2019b). *Régressions Logit-Probit, Poisson-binomiale négative*. Compléments d'analyse spatiale, Liège, Université de Liège, inédit, 29 p.

Donnay, J-P., Laloux, J-P. & Nadasdi, I. (1986). Analyse typologique du logement par secteur statistique dans l'agglomération liégeoise. In Sporck, J.A. (ed.) *Liège aujourd'hui et demain*. Liège : Eugène WHALE ÉDITEUR, 29-38.

Donnay, J-P. & Unwin, D. (2001). Modelling Geographical Distributions In Urban Areas. In Donnay, J-P., Barnsley, M., & Longley, P. (eds), *Remote Sensing And Urban Analysis*. London : Taylor & Francis, 205-224.

Dorman, M. (2014). *Learning R for Geospatial Analysis*. Birmingham : Packt Publishing, 364 p.

Ernst, J., Dewals, B., Detrembleur, S., Archambeau, P., Erpicum, S. & Piroton, M. (2010). Micro-Scale Flood Risk Analysis Based On Detailed 2D Hydraulic Modelling And High Resolution Geographic Data. *Natural Hazards*, 55 (2), 181-209. <https://doi.org/10.1007/s11069-010-9520-y>

Erpicum, S., Dewals, B., Archambeau, P., Detrembleur, S. & Piroton, M. (2010). Detailed Inundation Modelling Using High Resolution Dams. *Engineering Applications Of Computational Fluid Mechanics*, 4 (2), 196-208.
<https://doi.org/10.1080/19942060.2010.1101531>

Ferguson, R. (1977). *Linear Regression in Geography*. Norwich : Geo Abstracts University of East Anglia, CATMOG, 15, 24 p.

Fielding, J. & Burningham, K. (2005). Environmental Inequality And Flood Hazard. *Local Environment*, 10 (4), 379-395. <https://doi.org/10.1080/13549830500160875>

Fielding, J. (2012). Inequalities In Exposure And Awareness Of Flood Risk In England And Wales. *Disasters*, 36 (3), 477-494. <https://doi.org/10.1111/j.1467-7717.2011.01270.x>

Géoportail de Wallonie. (2019a). *Aléa d'inondation (Version 2016) – Série*. Géoportail de Wallonie. <http://geoportail.wallonie.be/catalogue/04c3ee3d-4036-4aea-a80c-9da59dad3b87.html>. Consulté le 20 août 2019.

Géoportail de Wallonie. (2019b). *Carte d'Occupation du Sol de Wallonie [COSW] (Version 2_07) – Série*. Géoportail de Wallonie. <http://geoportail.wallonie.be/catalogue/290e1fe8-0d99-410e-967b-a02f389b931a.html>. Consulté le 20 août 2019.

Goodchild, M.F. (1986). *Spatial Autocorrelation*. Norwich : Geo Abstracts University of East Anglia, CATMOG, 47, 56 p.

Griffith, D.A. & Chun, Y. (2013). *Spatial Statistics and Geostatistics*. Thousand Oaks (Cal) : SAGE Publications, 330 p.

Grimmeau, J-P., Deboosere, P., Eggerickx, T., Gadeyne, S., Hermia, J-P., Marissal, P., Romainville, A., Van Hecke, E. & Willaert, D. (2015). *Atlas de Belgique - Population*. Gand : Acamedia Press, Atlas de Belgique, 6, 104 p.

Grippa, T., Marissal, P., May, X., Wertz, I. & Loopmans, M. (2015). *Etude dynamiques des quartiers en difficulté dans les régions urbaines belges*. Bruxelles : Service fédéral de programmation intégration social, 87 p. <https://www.mi-is.be/fr/etudes-publications-statistiques/analyse-dynamique-des-quartiers-en-difficulte-dans-les-regions>. Consulté le 12 janvier 2019.

Groupe Transversal Inondations (GTI). (2015). *District hydrographique international de la Meuse : Projet de plan de gestion des risques d'inondation en Wallonie 2015-2021*. Jambes : DGO3, 302 p. http://environnement.wallonie.be/inondations/files/pgri/PGRI_MEUSE_20150424.pdf. Consulté le 20 juillet 2019.

Groupe Transversal Inondations (GTI). (2016). Notice méthodologique d'élaboration des cartographies des zones soumises à l'aléa d'inondation et du risque de dommages dus aux inondations. SPW, 51 p. http://environnement.wallonie.be/inondations/files/2016_carto/Methodo_GW20160310_final.pdf. Consulté le 20 juillet 2019.

Hammond, R. & McCullagh, P.S. (1978). *Quantitative Techniques in Geography. An Introduction*. 2 ed. Oxford : Clarendon Press, 384 p.

IWEPS. (2019). *Niveaux et conditions de vie - Revenus nets imposables médian et moyen*. Institut wallon de l'évaluation, de la prospective et de la statistique (IWEPS). https://walstat.iweps.be/walstat-catalogue.php?indicateur_id=831101. Consulté le 20 août 2019.

Le Gallo, J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire, *Économie et prévision*, 155, 139-157.

Ledant, M. (2009). *Évaluation du potentiel de la cartographie densimétrique appliquée à des données de très grande échelle spatiale 2D et 3D. Le cas de la région liégeoise*. Mémoire de master en sciences géographiques, Liège, Université de Liège, inédit, 107 p.

Lejeune, Z., Xhignesse, G., Kryvobokov, M. & Teller, J. (2016). Housing Quality As Environmental Inequality: The Case Of Wallonia, Belgium. *Journal Of Housing And The Built Environment*, 31 (3), 495-512.
<https://doi.org/10.1007/s10901-015-9470-5>

Maantay, J.A., Marako, A. & Culp, G. (2009). Using Geographic Information Science To Estimate Vulnerable Urban Populations For Flood Hazard And Risk Assessment In New York City. In Showalter P. & Lu Y. (eds), *Geospatial Techniques In Urban Hazard And Disaster Analysis*. Dordrecht : Springer, Geotechnologies and the environment, 2, 71-97. https://doi.org/10.1007/978-90-481-2238-7_5

Mather, P. & Openshaw, S. (1974). Multivariate Methods and Geographical Data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 23, 283-308.

McGrane, S. (2016). Impacts Of Urbanisation On Hydrological And Waterquality Dynamics, And Urban Water Management: A Review. *Hydrological Sciences Journal*, 61 (13), 2295-2311.
<https://doi.org/10.1080/02626667.2015.1128084>

Mustafa, A., Bruwier, M., Archambeau, P., Erpicum, S., Piroton, M., Dewals, B. & Teller, J. (2018). Effects Of Spatial Planning On Future Flood Risks In Urban Environments. *Journal Of Environmental Management*, 225, 193-204. <https://doi.org/10.1016/j.jenvman.2018.07.090>

NCGB. (2012). *A concise geography of Belgium*. Merelbeke : National Committe of Geography, 46 p.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich : Geo abstracts University of East Anglia, CATMOG, 38, 41 p. <https://www.uio.no/studier/emner/sv/iss/SGO9010/openshaw1983.pdf>. Consulté le 20 juillet 2019.

Pebesma, E. & Bivand, R., (2019). *Spatial Data Science*. <https://keen-swartz-3146c4.netlify.com/>. Consulté le 20 août 2019.

Peeters, A., Van Campenhout, J., Donnay, F., Mols, J., Snijders, J.-P. & Petit, F. (2006). La cartographie des zones inondées: approche géomorphologique et enquêtes de terrain. In *Les risques majeurs en Région wallonne: prévenir en aménageant*. Namur : Direction générale de l'Aménagement du territoire du Logement et du Patrimoine, Ministère de la Région wallonne, 44– 59.

QMRG. (2011). CATMOG. QMRG. <http://www.qmrg.org.uk/catmog/index.html>. Consulté le 20 août 2019.

Racine, J-B & Reymong H. (1973). *L'analyse quantitative en géographie*. Paris : PUF, Collection SUP.

Rasch, R. (2017). Income Inequality And Urban Vulnerability To Flood Hazard In Brazil. *Social Science Quarterly*, 98 (1), 299-325. <https://doi.org/10.1111/ssqu.12274>

Reboul, L. (2015). *Statistique spatiale ; cours 3*. I2M, Aix-Marseille, Université d'Aix-Marseille, 47 p. <http://iml.univ-mrs.fr/~reboul/SP3.pptx.pdf>. Consulté le 20 août 2019.

- Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston : Northwestern University Press, 617 p.
- Siegel, S. (1956). *Nonparametrics Statistics for the Behavioral Sciences*. New York : McGraw Hill Book Company, 312 p.
- Snedecor, G.W. & Cochran, W.G. (1980). *Statistical methods*. 7 ed. Ames (Iowa) : The Iowa State University Press, 476 p.
- Sporck, J., Van Der Haegen, H. & Pattyn, M. (1985). L'organisation spatiale de l'espace urbain. In Preudhomme, C. & Viaene-Awouters, L. (eds), *La cité belge d'aujourd'hui : quel devenir ?*. Bruxelles : Crédit communal, Bulletin trimestriel du Crédit Communal de Belgique, 154, 153-164.
- Statbel. (2014). *Census 2011 Belgique*. Direction générale Statistique - Statistics Belgium (Statbel). http://www.census2011.be/index_fr.html. Consulté le 20 août 2019.
- Statbel. (2017a). *Population par secteur statistique - 2017*. Direction générale Statistique - Statistics Belgium (Statbel). <https://statbel.fgov.be/fr/open-data/population-par-secteur-statistique-5>. Consulté le 20 août 2019.
- Statbel. (2017b). *Revenus fiscaux*. Direction générale Statistique - Statistics Belgium (Statbel). <https://statbel.fgov.be/fr/themes/menages/revenus-fiscaux#documents>. Consulté le 20 août 2019.
- Statbel. (2017c). *Statistique fiscale des revenus par secteur statistique 2005-2016*. Direction générale Statistique - Statistics Belgium (Statbel). <https://statbel.fgov.be/fr/open-data/statistique-fiscale-des-revenus-par-secteur-statistique>. Consulté le 20 août 2019.
- Timmermans, C. (2015). *STAT0076 – Cours 5. Statistiques générales*, Liège, Université de Liège, inédit, 40 p.
- UNDP. (2018). *Human Development Index (HDI)*. United Nations Development Programme (UNDP). <http://www.hdr.undp.org/en/content/human-development-index-hdi>. Consulté le 12 janvier 2019.
- Vandermotten, C. & Vandewattyne, P. (1985). Les étapes de la croissance et de la formation des armatures urbaines en Belgique. In Preudhomme, C. & Viaene-Awouters, L. (eds), *La cité belge d'aujourd'hui : quel devenir ?*. Bruxelles : Crédit communal, Bulletin trimestriel du Crédit Communal de Belgique, 154, 41-62.
- Walker, G., Burningham, K., Fielding, J., Smith, G., Thrush, D. & Fay, H. (2006). *Addressing Environmental Inequalities: Flood Risk*. Rotherham : Environment agency, 120 p. https://www.staffs.ac.uk/assets/SC020061_SR1%20report%20-%20inequalities%20flood%20risk_tcm44-21951.pdf. Consulté le 20 juillet 2019.
- Wright, J. K. (1936). A method of mapping densities of population : With Cape Cod as an Example. *The geographical review*, 26 (1), 103-110. https://www.jstor.org/stable/209467?seq=1#metadata_info_tab_contents. Consulté le 20 juillet 2019.
- Wrigley, N. (1976). *Introduction to the use of logit models in geography*. Norwich : Geo Abstracts University of East Anglia, CATMOG, 10, 18 p.
- Yeates, M. (1974). *An Introduction to Quantitative Analysis in Human Geography*. New York : McGraw-Hill, 300 p.
- Zaninetti, J-M. (2005). *Statistique spatiale. Méthodes et applications géomatiques*. Paris : Hermès-Lavoisier, 300 p.

Liste des annexes

<i>Annexe 1 : Ensemble des fichiers Excel réalisés à la suite du recensement de 2011, disponible sur Statbel.</i>	<i>95</i>
<i>Annexe 2 : Partie de la légende de la carte d'occupation du sol de Wallonie – COSW V2.07 et cartographie des terrains résidentiels à Lambermont.....</i>	<i>97</i>
<i>Annexe 3 : Cartographie de l'aléa d'inondation dans le centre de Liège.....</i>	<i>98</i>
<i>Annexe 4 : Histogrammes des indices socio-économiques et tableau d'analyse des distributions</i>	<i>99</i>
<i>Annexe 5: Statistiques descriptives des 24 indices socio-économiques élaborés à partir du recensement de 2011</i>	<i>104</i>
<i>Annexe 6 : Compte-rendu des opérations réalisées sur Arcgis de la classification de l'aléa d'inondation en une variable binaire.....</i>	<i>105</i>
<i>Annexe 7 : Secteurs statistiques dans la commune de Faimies qui n'est pas intersectée par une zone d'inondation</i>	<i>110</i>
<i>Annexe 8 : Secteurs statistiques supprimés lors de la sélection des zones habitées</i>	<i>110</i>
<i>Annexe 9 : Diagramme triangulaire des proportions de surface inondée en zone habitée réparties entre 4 classes</i>	<i>110</i>
<i>Annexe 10 : Valeurs aberrantes décelées dans la part d'habitants par logement et le taux d'accroissement</i>	<i>111</i>
<i>Annexe 11 : Tableau récapitulatif de la suppression des secteurs statistiques dans l'ensemble de l'étude</i>	<i>111</i>
<i>Annexe 12: Matrice de corrélation des 24 indices socio-économiques obtenus à partir du recensement de 2011</i>	<i>112</i>
<i>Annexe 13 : Construction de la matrice de voisinage des secteurs statistiques</i>	<i>113</i>
<i>Annexe 14 : Calcul de l'indice de Moran pour le revenu médian par secteur statistique en fonction de l'ordre de la matrice de voisinage.....</i>	<i>114</i>
<i>Annexe 15 : Régression avec interaction sur le revenu médian des secteurs statistiques (7.512) avec 3 variables indépendantes</i>	<i>115</i>
<i>Annexe 16 : Régression sur le revenu médian des secteurs statistiques (7.510) avec 4 variables indépendantes</i>	<i>116</i>
<i>Annexe 17 : Autocorrélation spatiale des 3 indices (Part de la population ayant un diplôme d'enseignement supérieur, part de ménages d'une personne, taux de chômage)</i>	<i>117</i>
<i>Annexe 18 : Autocorrélation des résidus suite à l'auto-régression LAG de l'écart relatif des revenus</i>	<i>118</i>
<i>Annexe 19 : Résultat de la première ACP comprenant 15 variables.....</i>	<i>119</i>
<i>Annexe 20: Résultat final de l'ACP sur 14 indices</i>	<i>120</i>
<i>Annexe 21 : Tableau des secteurs dont le score factoriel est supérieur à 3.....</i>	<i>121</i>
<i>Annexe 22 : Histogramme des fréquences du score factoriel des 7512 secteurs obtenus après suppression des valeurs aberrantes</i>	<i>123</i>
<i>Annexe 23 : Tableau des valeurs extrêmes de l'ANOVA à un facteur sur le niveau de vie et le statut d'inondation</i>	<i>124</i>
<i>Annexe 24 : Tests non paramétriques à 1 facteur</i>	<i>126</i>
<i>Annexe 25 : Statistiques descriptives du score factoriel des groupes formés par leur appartenance à un quadrant du diagramme de Moran et leur statut d'inondation.....</i>	<i>127</i>
<i>Annexe 26 : Tests non paramétriques à 2 facteurs.....</i>	<i>127</i>

Annexe 1 : Ensemble des fichiers Excel réalisés à la suite du recensement de 2011, disponible sur Statbel.

L'annexe comprend le titre des feuilles, une indication sur l'utilisation de la feuille dans l'étude, ainsi que le nombre de secteurs par fichier Excel

Census_Demo_Sector_Pieter_FR_10FEB15	
20 125/ 19 748 secteurs statistiques	
Titre de la feuille	Utilisation
Population de nationalité belge et étrangère selon le sexe	Non
Population de nationalité belge et étrangère selon le sexe - Secteurs statistiques	Oui
Population selon le sexe et l'état civil	Non
Population selon le sexe et l'état civil - Secteurs statistiques	Oui
Population de nationalité belge et étrangère selon le sexe et le groupe d'âge quinquennal	Non
Population de nationalité belge et étrangère selon le sexe et le groupe d'âge quinquennal - Secteurs statistiques	Oui
Densité de population	Non
Densité de population par secteur statistique - Secteurs statistiques	Oui

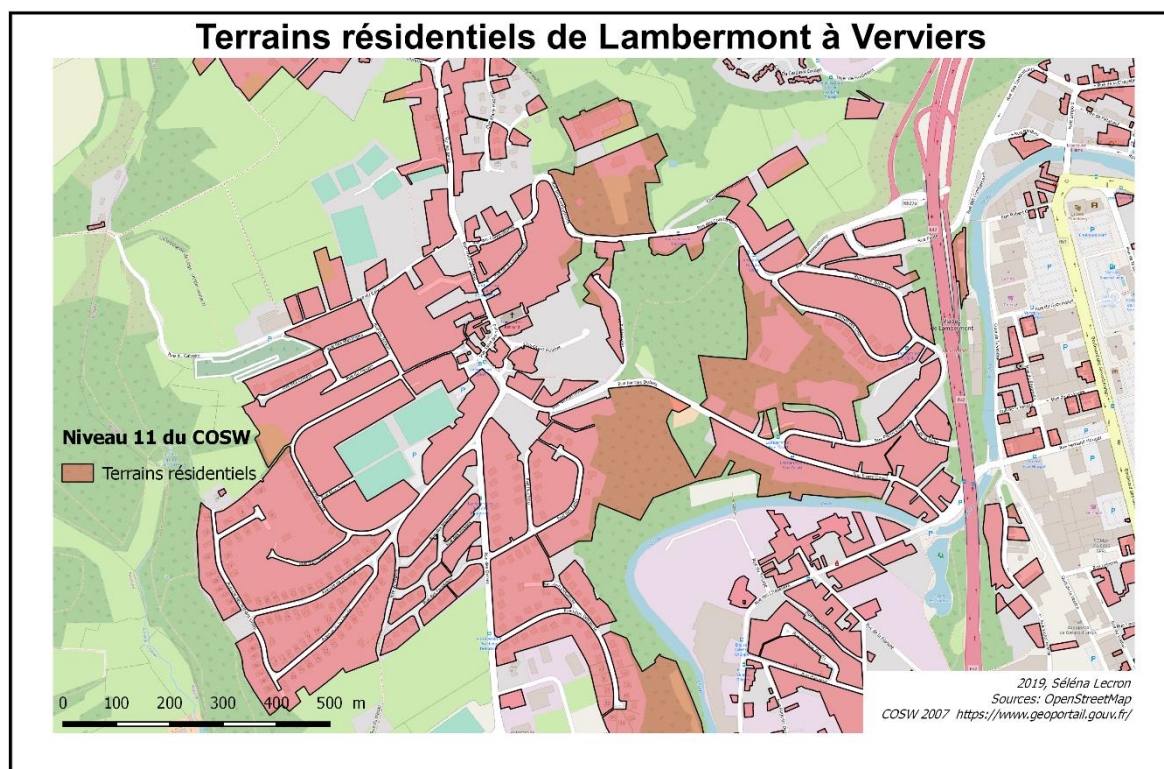
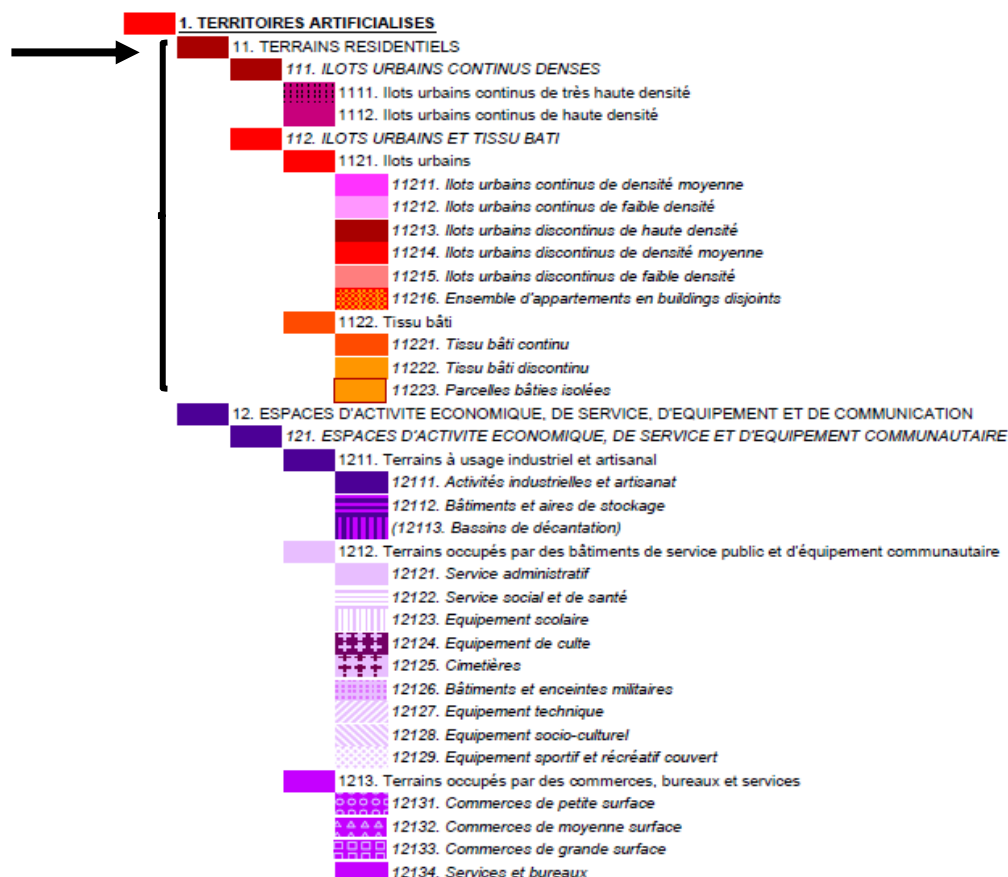
Census_Educ_Sector_FR_10FEB15	
19 738 secteurs statistiques	
Titre de la feuille	Utilisation
Population belge et étrangère suivant ou non un enseignement selon le lieu d'habitation et le sexe	Non
Population belge et étrangère suivant ou non un enseignement selon le lieu d'habitation et le sexe - Secteurs statistiques	Non
Population de 18 ans et plus, ne suivant plus d'enseignement selon le lieu d'habitation, le sexe et le niveau d'instruction	Non
Population de 18 ans et plus, ne suivant plus d'enseignement selon le lieu d'habitation, le sexe et le niveau d'instruction - Secteurs statistiques	Oui

Census_Logements_Sector_FR_23MAR15	
19 693 / 20 088 / 19 705 secteurs statistiques	
Titre de la feuille	Utilisation
Logements occupés selon le régime de propriété	Non
Logements occupés selon le régime de propriété - Secteurs statistiques	Oui
Population selon le régime de propriété de son logement	Non
Population selon le régime de propriété de son logement - Secteurs statistiques	Non
Logements classiques selon la période de construction	Non
Logements classiques selon la période de construction - Secteurs statistiques	Oui
Population selon la période de construction de son logement	Non
Population selon la période de construction de son logement - Secteurs statistiques	Non
Logements classiques occupés selon le nombre de pièces	Non
Logements classiques occupés selon le nombre de pièces - Secteurs statistiques	Oui

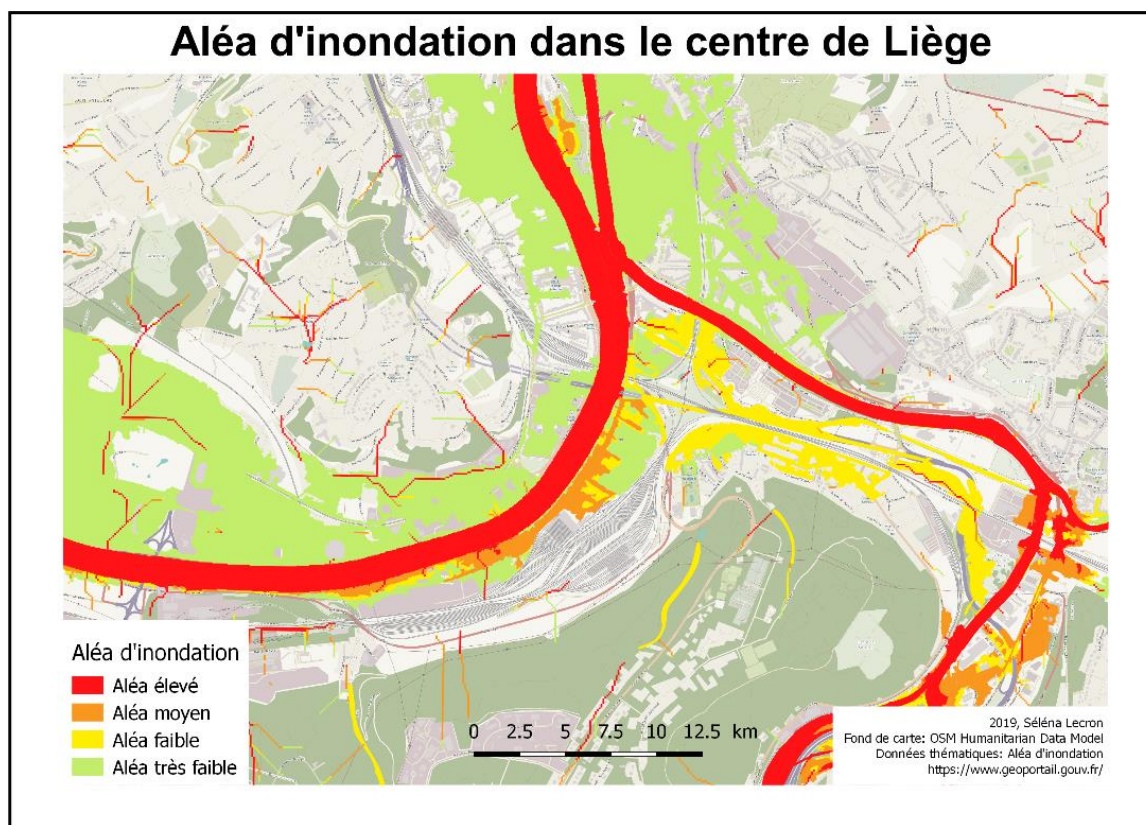
Census_emploi_Sector_FR_02MAR15	
19 748/ 19 555 secteurs statistiques	
Titre de la feuille	Utilisation
Population de nationalité belge et étrangère selon le sexe et la situation sur le marché de l'emploi	Non
Population de nationalité belge et étrangère selon le sexe et la situation sur le marché de l'emploi - Secteurs statistiques	Oui
Population active occupée selon le sexe, le lieu de travail et le statut professionnel - Population totale	Non
Population active occupée selon le sexe, le lieu de travail et le statut professionnel - Population totale - Secteurs statistiques	Oui
Population active occupée selon le sexe et le secteur d'activité économique - Population totale	Non
Population active occupée selon le sexe et le secteur d'activité économique - Population totale - Secteurs statistiques	Non
Population active occupée selon le sexe et le groupe d'âge quinquennal - Population totale	Non
Population active occupée selon le sexe et le groupe d'âge quinquennal - Population totale - Secteurs statistiques	Non
Demandeurs d'emploi selon le sexe et le groupe d'âge quinquennal - Population totale	Non
Demandeurs d'emploi selon le sexe et le groupe d'âge quinquennal - Population totale - Secteurs statistiques	Non

Census_Menages_Noyaux_Sector_FR_17APR15	
19 487 / 19 738 / 19 718 secteurs statistiques	
Titre de la feuille	Utilisation
Population vivant en ménage privé selon le sexe et le type de ménage	Non
Population vivant en ménage privé selon le sexe et le type de ménage - Secteurs statistiques	Oui
Ménages privés selon la taille du ménage	Non
Ménages privés selon la taille du ménage - Secteurs statistiques	Oui
Types de noyaux familiaux	Non
Types de noyaux familiaux - Secteurs statistiques	Non
Taille des noyaux familiaux	Non
Taille des noyaux familiaux - Secteurs statistiques	Oui

Annexe 2 : Partie de la légende de la carte d'occupation du sol de Wallonie – COSW V2.07 et cartographie des terrains résidentiels à Lambermont

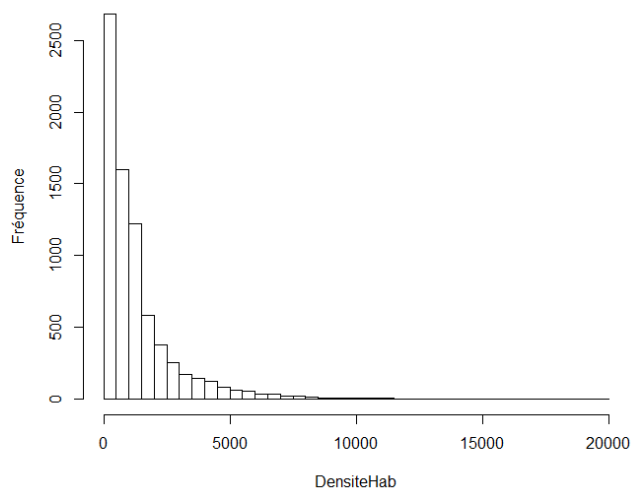


Annexe 3 : Cartographie de l'aléa d'inondation dans le centre de Liège

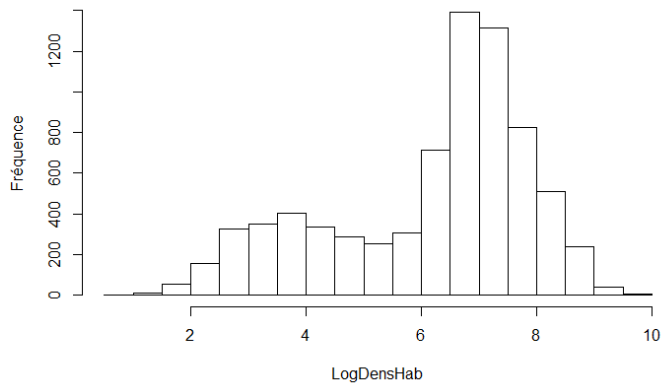


Annexe 4 : Histogrammes des indices socio-économiques et tableau d'analyse des distributions

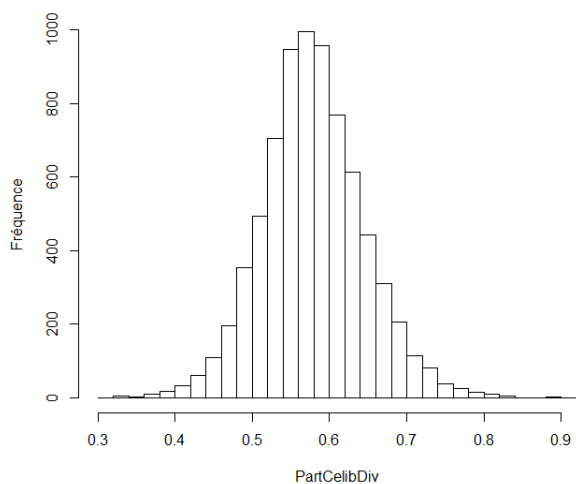
Histogramme de DensiteHab



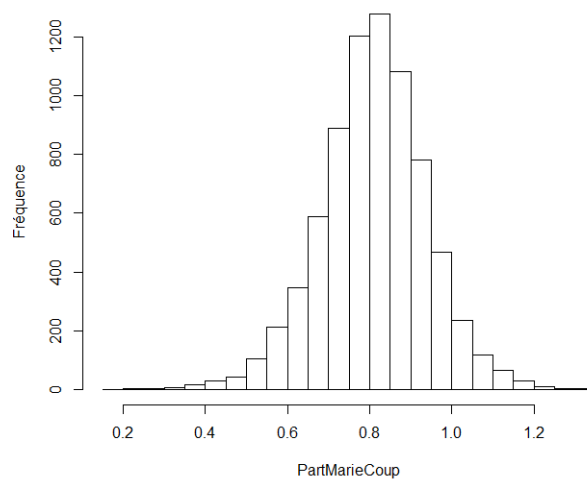
Histogramme de LogDensHab



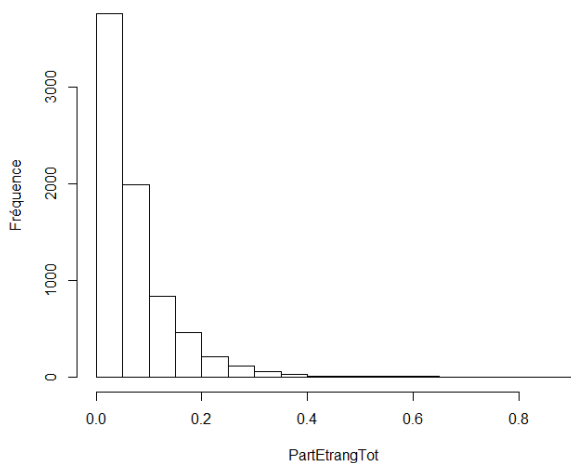
Histogramme de PartCelibDiv



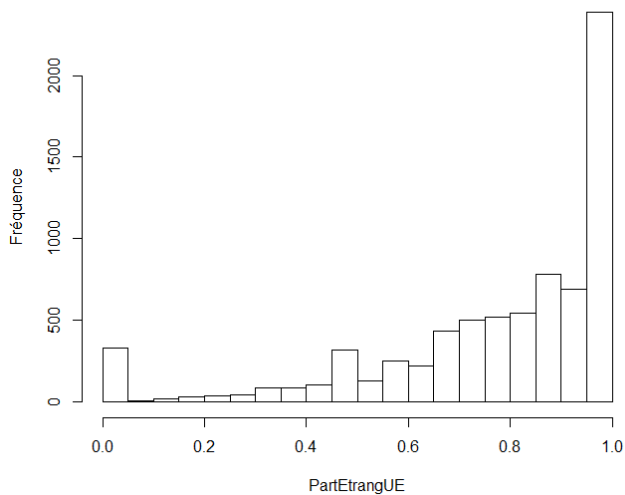
Histogramme de PartMarieCoup



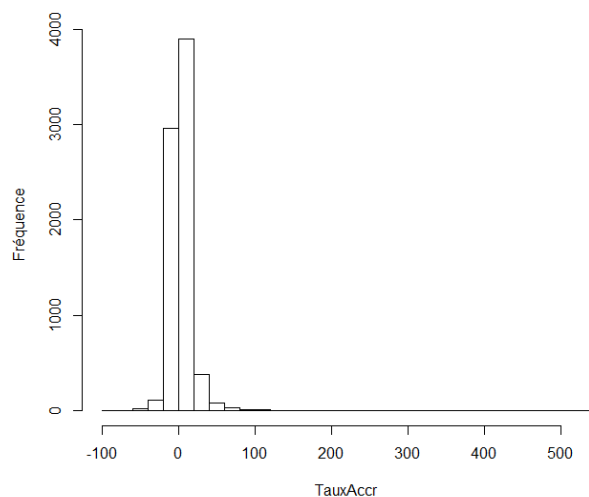
Histogramme de PartEtrangTot



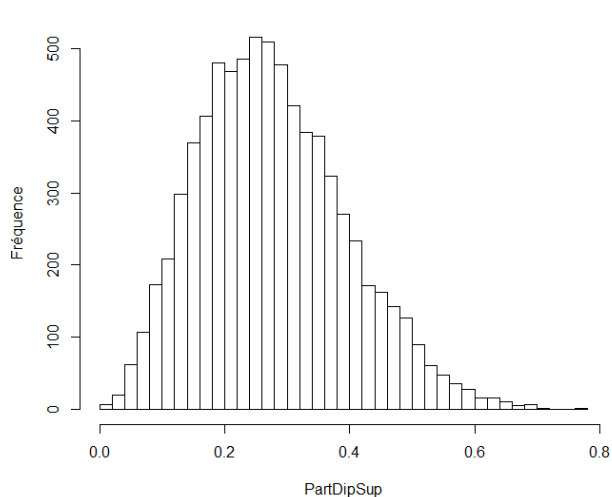
Histogramme de PartEtrangUE



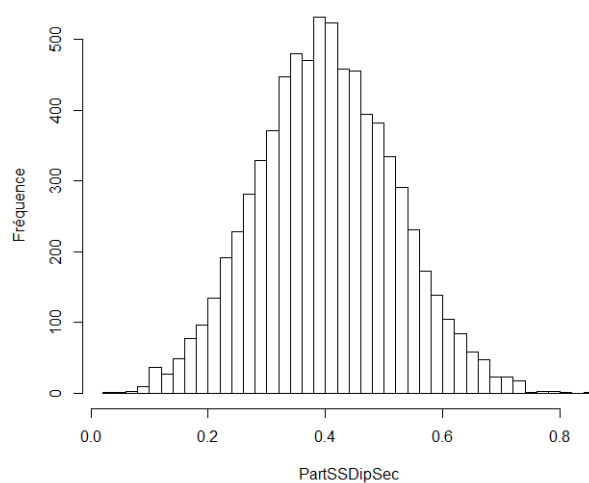
Histogramme de TauxAccr



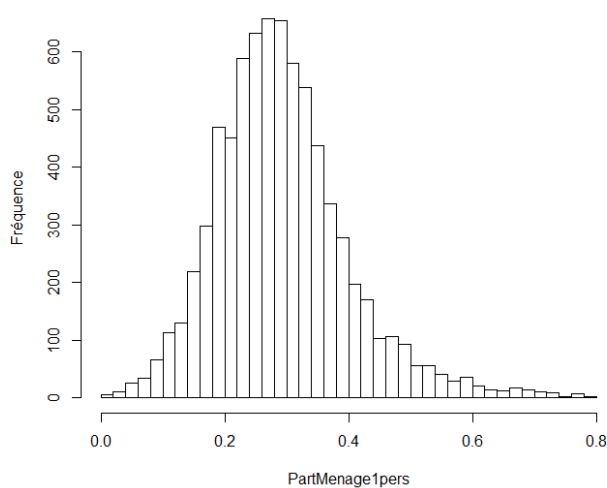
Histogramme de PartDipSup



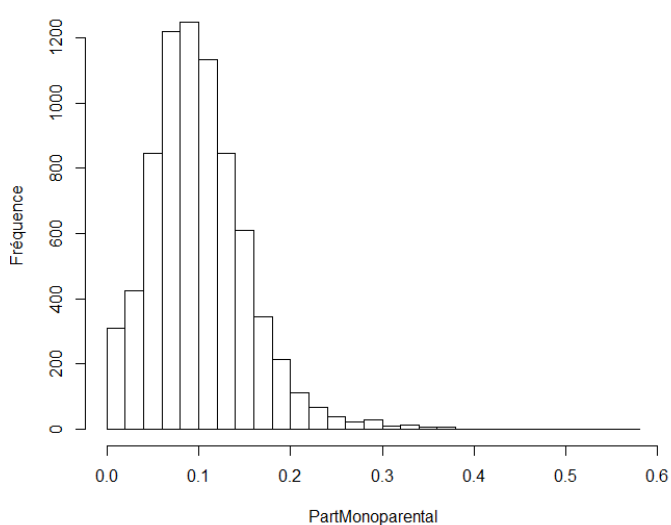
Histogramme de PartSSDipSec



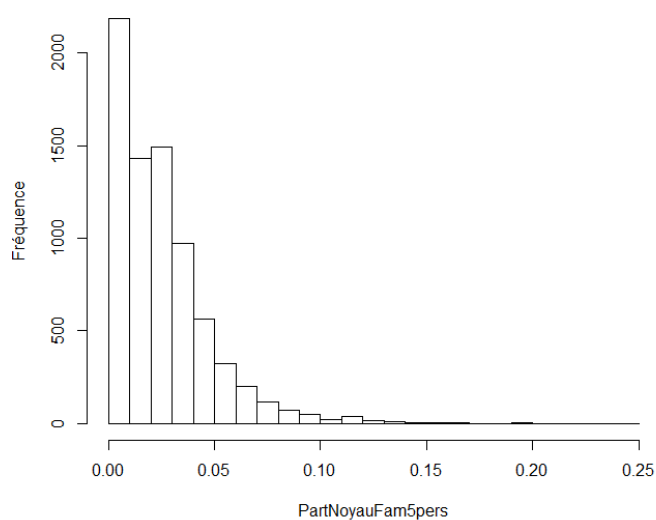
Histogramme de PartMenage1pers



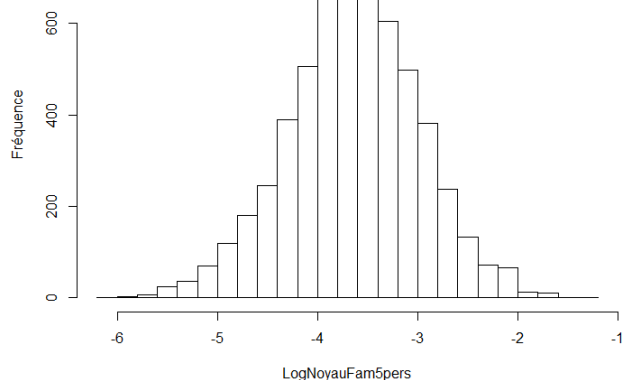
Histogramme de PartMonoparental



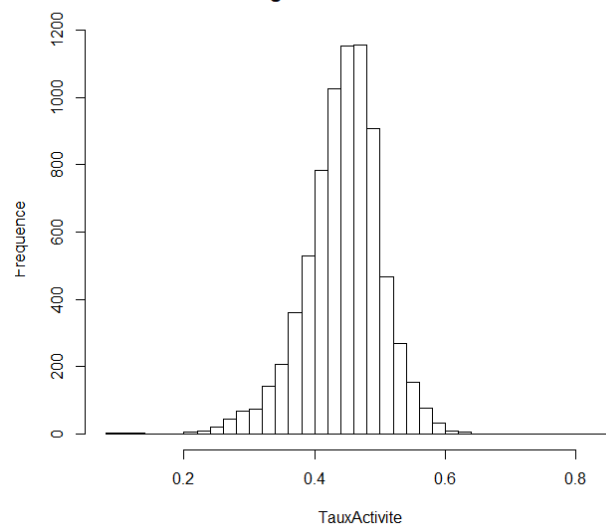
Histogramme de PartNoyauFam5pers



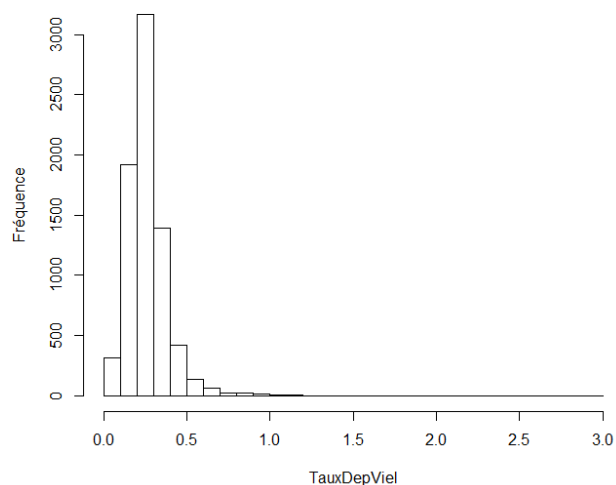
Histogramme de LogNoyauFam5pers



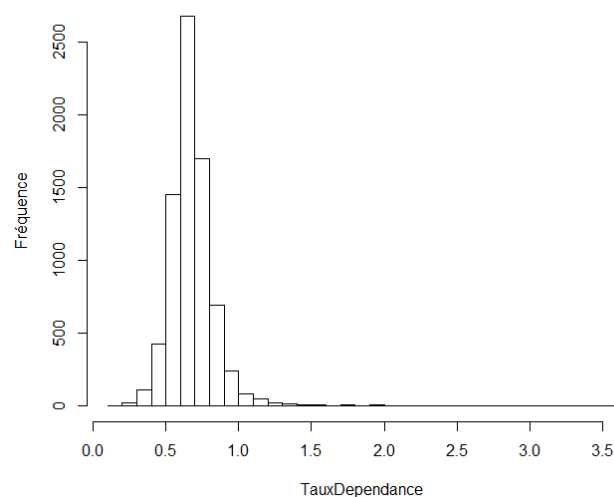
Histogramme de TauxActivite



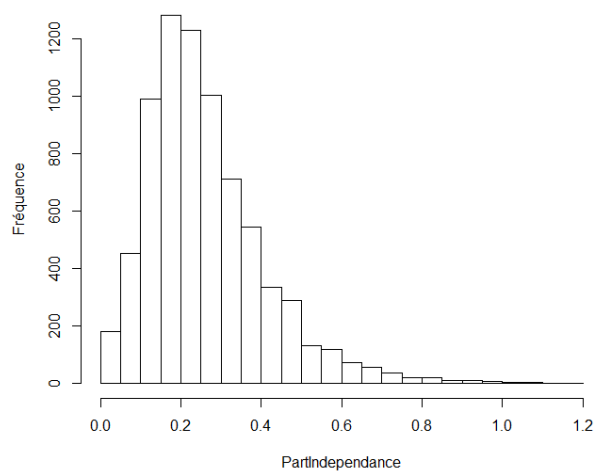
Histogramme de TauxDepViel



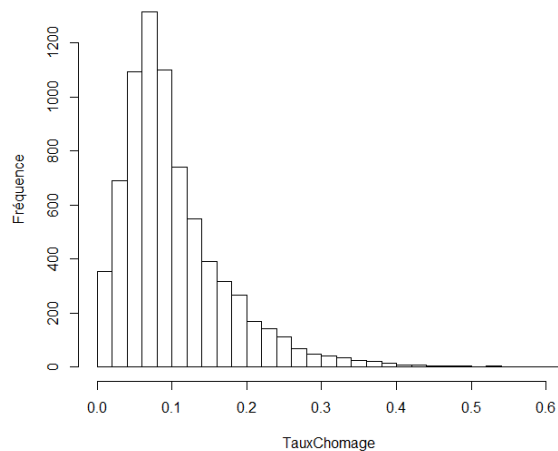
Histogramme de TauxDependance



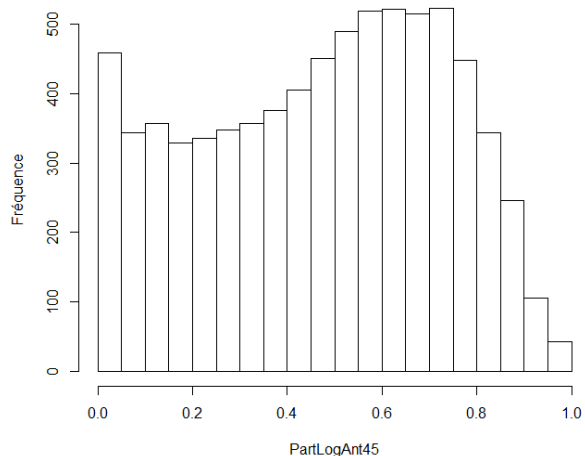
Histogramme de PartIndependance



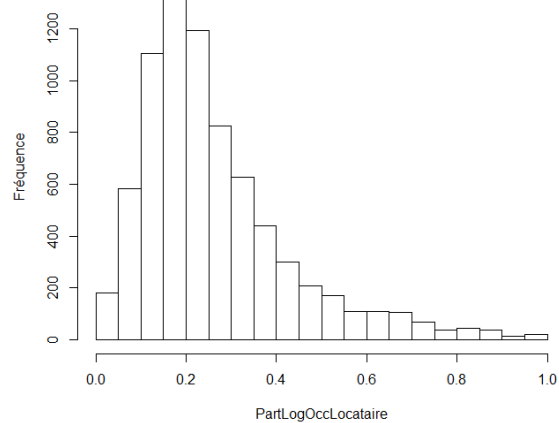
Histogramme de TauxChomage



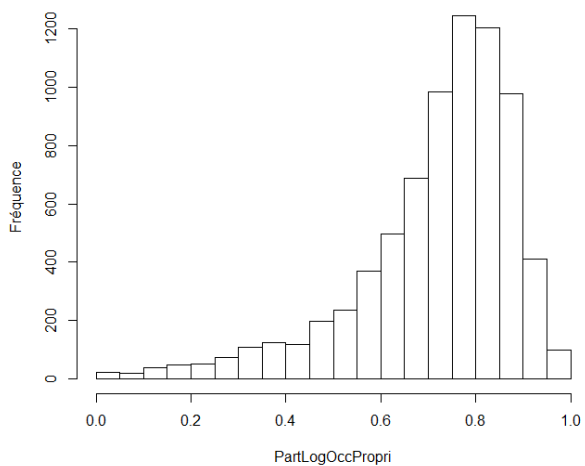
Histogramme de PartLogAnt45



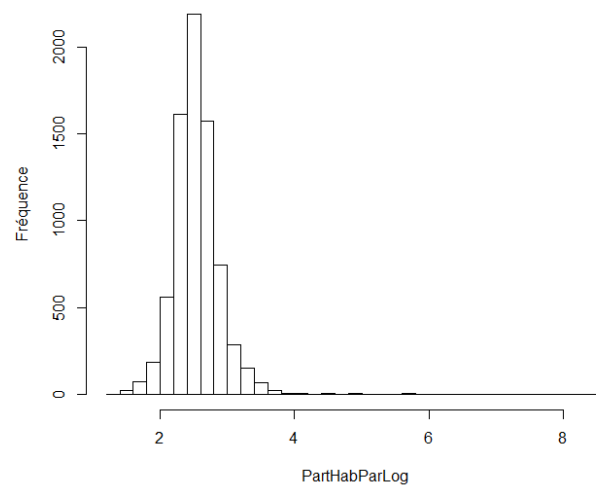
Histogramme de PartLogOccLocataire



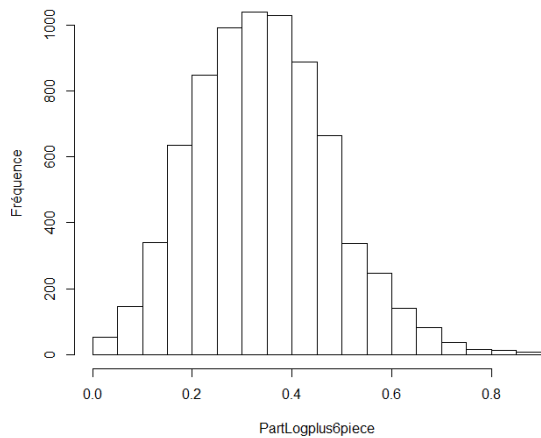
Histogramme de PartLogOccPropri  



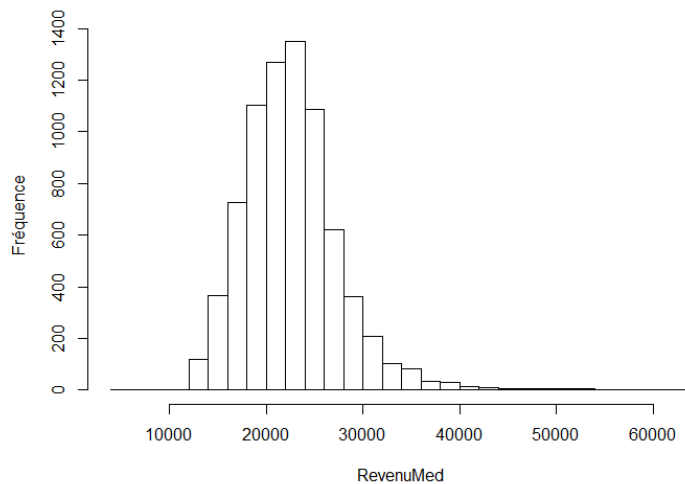
Histogramme de PartHabParLog

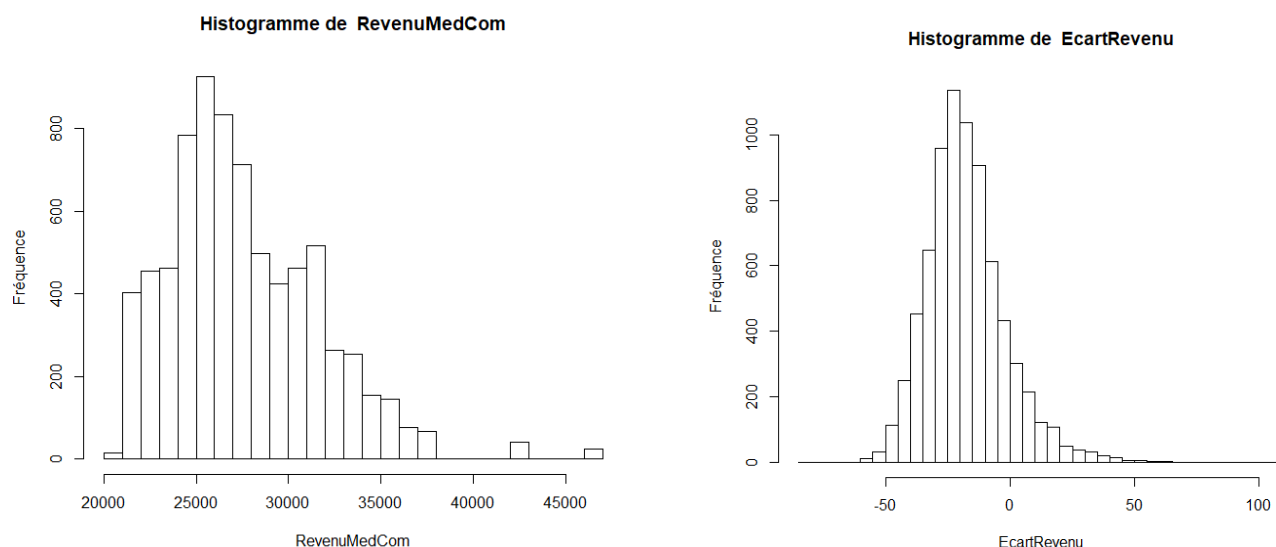


Histogramme de PartLogplus6piece



Histogramme de RevenuMed





Nom simplifié de l'indice	Analyse de la distribution des histogrammes des 7512 secteurs statistiques Construction de 30 classes par histogramme
Démographie	
DensiteHab	Distribution asymétrique vers la gauche
LogDensiteHab	Distribution bimodale
PartCelibDiv	Distribution normale
PartMarieCoup	Distribution normale
PartEtrangTot	Distribution asymétrique vers la gauche
PartEtrangUE	Distribution anormale
TauxAccr	Distribution normale avec des valeurs extrêmes
Education	
PartDipSup	Distribution légèrement asymétrique vers la gauche
PartSSDipSec	Distribution normale
Composition des ménages	
PartMenage1pers	Distribution légèrement asymétrique vers la gauche
PartMonoparental	Distribution légèrement asymétrique vers la gauche
PartNoyauFam5pers	Distribution asymétrique vers la gauche
LogPartNoyauFam5pers	Distribution normale
Emploi	
TauxActivite	Distribution normale
TauxDepViel	Distribution légèrement asymétrique vers la gauche
TauxDependance	Distribution normale
PartIndependance	Distribution légèrement asymétrique vers la gauche
TauxChomage	Distribution légèrement asymétrique vers la gauche
Logement	
PartLogAnt45	Distribution uniforme
PartLogOccLocataire	Distribution légèrement asymétrique vers la gauche
PartLogOccPropri	Distribution légèrement asymétrique vers la droite
PartHabParLog	Distribution normale avec des valeurs extrêmes
PartLogplus6piece	Distribution normale
Fiscalité	
RevenuMed	Distribution normale
RevenuMedCom	Distribution anormale
EcartRevenu	Distribution légèrement asymétrique vers la gauche

Annexe 5: Statistiques descriptives des 24 indices socio-économiques élaborés à partir du recensement de 2011

Nom simplifié de l'indice	Statistiques descriptives des 7512 secteurs statistiques					
	Minimum	Maximum	Moyenne	Médiane	Variance	Ecart type
Démographie						
DensiteHab	2,605	19.656,47	1.259,83	831,40	2.463.274,9	1.569,48
PartCelibDiv	0,3011	0,9186	0,57942	0,5766	0,0044	0,0661
PartMarieCoup	0,1628	1	0,8104	0,8132	0,0167	0,1293
PartEtrangTot	0	0,8549	0,0748	0,05	0,0061	0,0781
PartEtrangUE	0	1	0,7806	0,8571	0,0636	0,2521
TauxAccr	-97,5	526,1628	3,3734	1,7476	240,2696	15,5006
Education						
PartDipSup	0	0,7662	0,2772	0,2660	0,0139	0,1180
PartSSDipSec	0,0294	0,8505	0,4020	0,4	0,0137	0,1169
Composition des ménages						
PartMenage1pers	0	0,7859	0,2923	0,2825	0,0114	0,1066
PartMonoparental	0	0,5681	0,1008	0,0957	0,0028	0,0533
PartNoyauFam5pers	0	0,25	0,0248	0,0210	0,0006	0,0241
Emploi						
TauxActivite	0,0944	0,8533	0,4436	0,4485	0,0035	0,0590
TauxDepViel	0	2,9697	0,2660	0,2464	0,0223	0,1494
TauxDependance	0,1719	3,5625	0,6821	0,6667	0,0272	0,1651
PartIndependance	0	1,1875	0,2628	0,2349	0,0221	0,1486
TauxChomage	0	0,6087	0,1028	0,0854	0,0050	0,0710
Logement						
PartLogAnt45	0	1	0,4724	0,5	0,0660	0,2569
PartLogOccLocataire	0	1	0,2644	0,2223	0,0281	0,1676
PartLogOccPropri	0	1	0,7591	0,7183	0,0283	0,1682
PartHabParLog	1,317	8,6	2,5504	2,5217	0,1381	0,3716
PartLogplus6piece	0	0,893	0,3422	0,3333	0,0191	0,1381
Fiscalité						
RevenuMed	4.812,85	63.124,30	22.628,38	22.246,87	24.758.769	4975,82
RevenuMedCom	20.717	46.659	27.619,02	26.815	16.162.672	4020,28
EcartRevenu	-80,58	105,67	-17,63	-19,38	253,36	15,92

Annexe 6 : Compte-rendu des opérations réalisées sur Arcgis de la classification de l'aléa d'inondation en une variable binaire

Prétraitement : Les secteurs statistiques d'une même commune ont été fusionnés pour obtenir une couche vectorielle de l'ensemble des communes wallonnes.

Sélection des communes intersectées par les inondations

1) Fonction : Dissolve

Fichier INPUT :

- scbel01012011_gen13.shp : Couche des secteurs statistiques

Fichier OUTPUT :

- Commune.shp

Paramètre de la fonction :

- DissolveField : Nis_012011 = Code INS des communes

3) Fonction : Select by attributes

Fichier INPUT :

- ALEA_INOND_2016__ALEA.shp : Couche de l'aléa d'inondation

Paramètre de la fonction :

- "TYPEALEA" ='Débordement'

2) Fonction : Select by location

Fichiers INPUT :

- Commune.shp (*Target layer*)
- ALEA_INOND_2016__ALEA.shp : Couche des aléas d'inondation (*Source layer/ Selected features*)

Fichier OUTPUT :

- Sélection des communes intersectées par les débordements

Paramètre de la fonction :

- Spatial selection : intersect the source layer feature
- ⇒ A partir de la sélection des débordements, les communes intersectées sont sélectionnées. Par une sélection inverse, on remarque qu'il n'y a qu'une commune non intersectée par les zones d'inondation correspondant à 11 secteurs statistiques.

3) Fonction : Select by attribute

Fichiers INPUT :

- Commune.shp (*Selected features*)
- scbel01012011_gen13.shp : Couche des secteurs statistiques

Fichiers OUTPUT :

- Sélection des secteurs dans les communes intersectées

Paramètre de la fonction :

- Select from current selection

Classification des aléas d'inondation

1) Fonction : Clip

Fichiers INPUT :

- scbel01012011_gen13.shp : Couche des secteurs statistiques (*Selected features*)
- Niv2_11.shp : Couche des zones résidentielles (*Clip Features*)

Fichier OUTPUT :

- DécoupZHSS .shp : couche des zones habitées par secteur statistique

2) Fonction : Calculate geometry

Fichier INPUT :

- DécoupZHSS.shp

Fichier OUTPUT :

- Nouveau champ : AreaZH = contient la superficie habitée pour chaque secteur en m²

Paramètres de la fonction :

- Property : Area
- Use coordinate system of the data source : Belge Lambert 1972
- Units : Square meters

3) Fonction : Select by attributes

Fichier INPUT :

- ALEA_INOND_2016__ALEA.shp : Couche des aléas d'inondation

Fichier OUTPUT :

- ALEA_INOND_2016__ALEA.shp (*Selected features*)

Paramètre de la fonction :

"TYPEALEA" ='Débordement' AND ("VALEUR"= 'Aléa faible' OR "VALEUR"= 'Aléa très faible')

⇒ Sélection des valeurs d'aléa faible et d'aléa très faible

4) Fonction : Export Data

Fichier INPUT :

- ALEA_INOND_2016__ALEA.shp (*Selected features*)

Fichier OUTPUT :

- AléaFT.shp

Paramètre(s) de la fonction :

- Export : Selected features
- Same coordinate system as : this layer's source data

⇒ Enregistre dans un shapefile la nouvelle valeur « d'aléa faible » comprenant l'aléa faible et très faible

5) Fonction : Select by attributes

Fichier INPUT :

- ALEA_INOND_2016__ALEA.shp

Fichier OUTPUT :

- ALEA_INOND_2016__ALEA.shp (*Selected features*)

Paramètre de la fonction :

"TYPEALEA"='Débordement' AND ("VALEUR"= 'Aléa moyen' OR "VALEUR"= 'Aléa élevé')

⇒ Sélection des deux valeurs d'aléa élevé qu'on regroupe en une seule

6) Fonction : Export Data

Fichier INPUT :

- ALEA_INOND_2016__ALEA.shp (*Selected features*)

Fichier OUTPUT :

- AléaEM.shp

Paramètre(s) de la fonction :

- Export : Selected features
- Same coordinate system as : this layer's source data

⇒ Enregistre dans un shapefile la nouvelle valeur « d'aléa élevé » comprenant l'aléa moyen et élevé

7) Fonction : Clip

Fichiers INPUT :

- DécoupZHSS .shp : Couche des zones résidentielles par secteur statistique
- AléaFT.shp : Couche d'aléa d'inondation faible (*Clip Features*)

Fichier OUTPUT :

- DécoupInonFT.shp : Couche des zones habitées par secteur intersectées par un aléa d'inondation faible

8) Fonction : Clip

Fichiers INPUT :

- DécoupZHSS .shp : Couche des zones résidentielles par secteur statistique
- AléaEM.shp : Couche d'aléa d'inondation élevé (*Clip Features*)

Fichier OUTPUT :

- DécoupInonEM.shp : Couche des zones habitées par secteur intersectées par un aléa d'inondation élevé

9) Fonction : Calculate geometry

Fichier INPUT :

- DécoupInonFT.shp

Fichier OUTPUT :

- Nouveau champ : AireInoFT = contient la superficie habitée faiblement inondée pour chaque secteur en m²

Paramètres de la fonction :

- Property : Area
- Use coordinate system of the data source : Belge Lambert 1972
- Units : Square meters

10) Fonction : Calculate geometry

Fichier INPUT :

- DécoupInonFT.shp

Fichier OUTPUT :

- Nouveau champ : AireInoEM = contient la superficie habitée fortement inondée pour chaque secteur en m²

Paramètres de la fonction :

- Property : Area
- Use coordinate system of the data source : Belge Lambert 1972
- Units : Square meters

11) Fonction : Join Data

Fichiers INPUT :

- DécoupZHSS .shp : Couche des zones résidentielles par secteur statistique
- DécoupInonFT.shp : Couche contenant la superficie des zones habitées faiblement inondées (*Table to join to the layer*)

Fichier OUTPUT :

- Nouveaux champs dans la couche DécoupZHSS.shp provenant de la couche DécoupInonFT.shp

Paramètres de la fonction :

- Field in this layer that the join will be based on : Cs012011 = l'identifiant des secteurs statistiques
- Choose the table to join to this layer : DécoupInonFT.shp
- Choose the field in the table to base the join on : Cs012011
- Join Option : Keep all record (cela permet de conserver les secteurs statistiques qui ne sont pas intersectés par l'aléa d'inondation faible)

12) Fonction : Join Data

Fichiers INPUT :

- DécoupZHSS .shp : Couche des zones résidentielles par secteur statistique
- DécoupInonEM.shp : Couche contenant la superficie des zones habitées fortement inondées (*Table to join to the layer*)

Fichiers OUTPUT :

- Nouveaux champs dans la couche DécoupZHSS.shp provenant de la couche DécoupInonEM.shp

Paramètres de la fonction :

- Field in this layer that the join will be based on : Cs012011 = l'identifiant des secteurs statistiques
- Choose the table to join to this layer : DécoupInonEM.shp
- Choose the field in the table to base the join on : Cs012011
- Join Option : Keep all record

13) Fonction : Export Data

Fichier INPUT :

- DécoupZHSS .shp : Couche des zones résidentielles par secteur statistique contenant les superficies des zones inondées

Fichier OUTPUT :

- DécoupZHSSAireIno.shp : Couche des zones résidentielles par secteur statistique contenant des zones inondées

Paramètres de la fonction :

- Export : All features
- Same coordinate system as : this layer's source data

- ⇒ Enregistre dans un shapefile de l'ensemble des superficies (habitée, inondé faiblement et inondé fortement)

14) Fonction : Calculate Field

Fichier INPUT :

- DécoupZHSSAireIno.shp : Couche des zones résidentielles par secteur statistique contenant les superficies des zones inondées

Fichier OUTPUT :

- Nouveau champ : AireSansIn = la superficie des zones non inondées par secteur

Paramètre de la fonction :

- Calcul : $\text{AireSansIn} = [\text{AreaZH}] - [\text{AireInoEM}] - [\text{AireInoFT}]$

15) Fonction : Calculate Field

Fichier INPUT :

- DécoupZHSSAireIno.shp : Couche des zones résidentielles par secteur statistique contenant les superficies des zones inondées

Fichier OUTPUT :

- Nouveau champ : PourcInoEM = Pourcentage de la superficie fortement inondée par rapport à la zone habitée pour chaque secteur

Paramètre de la fonction :

- Calcul : $\text{PourcInoEM} = ([\text{AireInoEM}] / [\text{AreaZH}]) * 100$

16) Fonction : Calculate Field

Fichier INPUT :

- DécoupZHSSAireIno.shp : Couche des zones résidentielles par secteur statistique contenant les superficies des zones inondées

Fichier OUTPUT :

- Nouveau champ : PourcInoFT = Pourcentage de la superficie faiblement inondée par rapport à la zone habitée pour chaque secteur

Paramètre de la fonction :

- Calcul : $\text{PourcInoFT} = ([\text{AireInoFT}] / [\text{AreaZH}]) * 100$

17) Fonction : Calculate Field

Fichier INPUT :

- DécoupZHSSAireIno.shp : Couche des zones résidentielles par secteur statistique contenant les superficies des zones inondées

Fichier OUTPUT :

- Nouveau champ : PourcSANSi = Pourcentage de la superficie non inondée par rapport à la zone habitée pour chaque secteur

Paramètre de la fonction :

- Calcul : $\text{PourcSANSi} = ([\text{AireSANSiT}] / [\text{AreaZH}]) * 100$

- ⇒ Résultat : une couche vectorielle (DécoupZHSSAireIno.shp) comprenant le pourcentage de superficie fortement, faiblement et non inondée en zone habitée

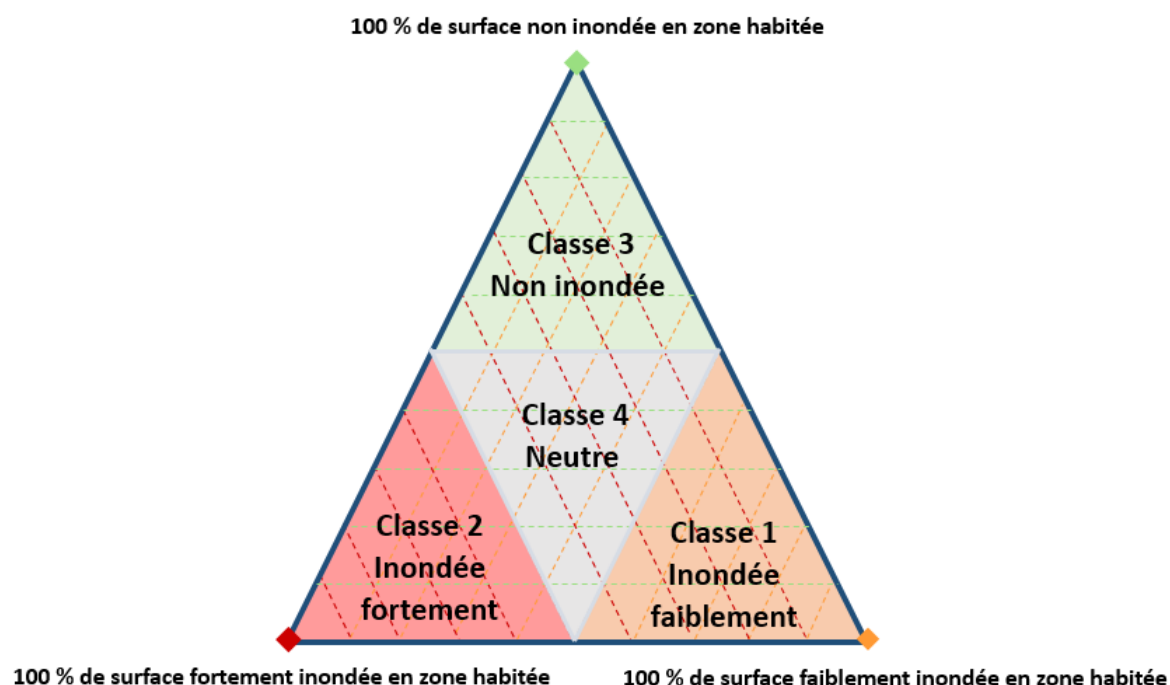
Annexe 7 : Secteurs statistiques dans la commune de Faimés qui n'est pas intersectée par une zone d'inondation

Identifiant du secteur	Nom du secteur	Identifiant du secteur	Nom du secteur
64076A00-	CELLES-CENTRE	64076C00-	LES WALEFFES-CENTRE
64076A01-	TERMOGNE	64076C099	WALEFFE - BORLEZ - CAMPAGNE
64076A02-	SAIVE	64076D00-	VIEMME-CENTRE
64076A09-	CELLES-HABITATIONS DISPERSEES	64076E000	AINEFFE-CENTRE
64076A100	FAIMES	64076E09-	AINEFFE-HABITATIONS DISPERSEES
64076B000	BORLEZ-CENTRE		

Annexe 8 : Secteurs statistiques supprimés lors de la sélection des zones habitées

Identifiant du secteur	Nom du secteur
93056M164	BOIS DE ROLY
93056N062	GUEULE DU LOUP
93056R062	LES VALISETTES
93056R163	NEUVILLE-LA-FORET

Annexe 9 : Diagramme triangulaire des proportions de surface inondée en zone habitée réparties entre 4 classes



Classe 1 Inondée faiblement	Classe 2 Inondée fortement	Classe 3 Non inondée	Classe 4 Neutre
31	308	7078	108

Annexe 10 : Valeurs aberrantes décelées dans la part d'habitants par logement et le taux d'accroissement

Identifiant du secteur	Nom du secteur	Part d'habitants par logement	Taux d'accroissement
93022A39-	AERODROME	13.21	-80.61
92094J13-	BRICGNIOT	11.69	-11.15
62063H223	SART-LOHAY	10.22	51.73
25048A09-	LES CAILLOUX	9.43	-70.830
62009B112	SEDOZ	9.14	-66.56

Annexe 11 : Tableau récapitulatif de la suppression des secteurs statistiques dans l'ensemble de l'étude

Traitement	Nombre de secteurs	Commentaire
Nombre de lignes maximum dans les feuilles de densité d'habitants de Stabel	20 125	Prend en compte les secteurs non-localisés
Suppression des secteurs sans revenu médian, de moins de 50 habitants et sans donnée pour certains indices	16 940	Les indices comportant des secteurs sans données sont : -DensiteHab -PartEtrangUE -PartNoyauFam5pers -PartMenage1pers -PartMonoparental -PartIndependance -TauxChomage -PartLogplus6piece
Conservation des secteurs seulement dans la région wallonne	7895	
Elimination des îles	7545	350 îles
Suppression de valeurs aberrantes	7540	Voir Annexe 4
Suppression des secteurs non intersectés par la zone habitée	7536	Voir Annexe 8
Suppression des secteurs dont la commune n'est pas intersectée par l'aléa d'inondation	7525	Supprime 11 secteurs statistiques Voir Annexe 7
Suppression des secteurs obtenant un score factoriel aberrant lors de l'ACP	7512	Voir Annexe 21

Annexe 12: Matrice de corrélation des 24 indices socio-économiques obtenus à partir du recensement de 2011

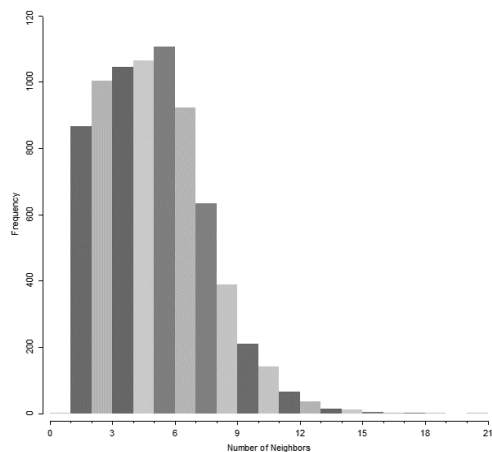
	RevenuMed	RevenuMedCom	EcartRevenu	DensiteHab	PartCelibDiv	PartMarieCoup	PartEtrangTot	PartEtrangUE	TauxAccr	PartDipSup	PartSSDipSec	PartMenage1pers
RevenuMed	1.000	0.467	0.769	-0.402	-0.592	0.572	-0.274	0.135	0.104	0.669	-0.682	-0.574
RevenuMedCom	0.467	1.000	-0.189	-0.254	-0.209	0.197	-0.204	0.024	0.073	0.573	-0.538	-0.292
EcartRevenu	0.769	-0.189	1.000	-0.281	-0.528	0.514	-0.162	0.144	0.062	0.351	-0.394	-0.440
DensiteHab	-0.402	-0.254	-0.281	1.000	0.430	-0.407	0.370	-0.151	-0.092	-0.287	0.347	0.545
PartCelibDiv	-0.592	-0.209	-0.528	0.430	1.000	-0.988	0.167	-0.176	-0.065	-0.324	0.339	0.599
PartMarieCoup	0.572	0.197	0.514	-0.407	-0.988	1.000	-0.132	0.175	0.050	0.304	-0.312	-0.582
PartEtrangTot	-0.274	-0.204	-0.162	0.370	0.167	-0.132	1.000	0.009	-0.078	-0.321	0.326	0.311
PartEtrangUE	0.135	0.024	0.144	-0.151	-0.176	0.175	0.009	1.000	0.053	0.040	-0.032	-0.180
TauxAccr	0.104	0.073	0.062	-0.092	-0.065	0.050	-0.078	0.053	1.000	0.048	-0.059	-0.053
PartDipSup	0.669	0.573	0.351	-0.287	-0.324	0.304	-0.321	0.040	0.048	1.000	-0.903	-0.360
PartSSDipSec	-0.682	-0.538	-0.394	0.347	0.339	-0.312	0.326	-0.032	-0.059	-0.903	1.000	0.398
PartMenage1pers	-0.574	-0.292	-0.440	0.545	0.599	-0.582	0.311	-0.180	-0.053	-0.360	0.398	1.000
PartMonoparental	-0.480	-0.248	-0.372	0.391	0.586	-0.565	0.272	-0.137	-0.085	-0.355	0.368	0.337
PartNouveauFam5pers	-0.097	0.001	-0.115	0.022	0.181	-0.175	0.016	-0.096	-0.030	-0.041	0.047	-0.023
TauxActivite	0.457	0.232	0.354	-0.335	-0.327	0.286	-0.352	0.120	0.126	0.388	-0.482	-0.337
TauxDepViel	-0.256	-0.081	-0.216	0.110	0.093	-0.076	0.008	0.014	-0.008	-0.163	0.268	0.253
TauxDependance	-0.142	-0.002	-0.157	0.013	0.159	-0.149	-0.044	-0.026	0.007	-0.110	0.188	0.013
PartIndependance	0.197	0.328	0.006	-0.313	-0.225	0.221	-0.148	0.004	0.016	0.416	-0.402	-0.197
TauxChomage	-0.634	-0.448	-0.417	0.521	0.530	-0.500	0.351	-0.125	-0.083	-0.567	0.610	0.547
PartLogAnt45	-0.382	-0.212	-0.281	0.180	0.334	-0.345	0.065	-0.037	-0.070	-0.268	0.259	0.282
PartLogOccLocataire	-0.584	-0.233	-0.503	0.530	0.609	-0.586	0.346	-0.238	-0.072	-0.408	0.455	0.631
PartLogOccPropri	0.575	0.225	0.498	-0.493	-0.602	0.581	-0.326	0.234	0.065	0.396	-0.441	-0.621
PartHabParLog	0.315	0.256	0.166	-0.353	-0.125	0.119	-0.114	0.009	-0.005	0.249	-0.261	-0.622
PartLogplus6piece	0.395	0.415	0.145	-0.396	-0.294	0.282	-0.287	-0.003	0.013	0.500	-0.489	-0.377

	PartMonoparental	PartNouveauFam5pers	TauxActivite	TauxDepViel	TauxDependance	PartIndependance	TauxChomage	PartLogAnt45	PartLogOccLocataire	PartLogOccPropri	PartHabParLog	PartLogplus6piece
RevenuMed	-0.480	-0.097	0.457	-0.256	-0.142	0.197	-0.634	-0.382	-0.584	0.575	0.315	0.395
RevenuMedCom	-0.248	0.001	0.232	-0.081	-0.002	0.328	-0.448	-0.212	-0.233	0.225	0.256	0.415
EcartRevenu	-0.372	-0.115	0.354	-0.216	-0.157	0.006	-0.417	-0.281	-0.503	0.498	0.166	0.145
DensiteHab	0.391	0.022	-0.335	0.110	0.013	-0.313	0.521	0.180	0.530	-0.493	-0.353	-0.396
PartCelibDiv	0.586	0.181	-0.327	0.093	0.159	-0.225	0.530	0.334	0.609	-0.602	-0.125	-0.294
PartMarieCoup	-0.565	-0.175	0.286	-0.076	-0.149	0.221	-0.500	-0.345	-0.586	0.581	0.119	0.282
PartEtrangTot	0.272	0.016	-0.352	0.008	-0.044	-0.148	0.351	0.065	0.346	-0.326	-0.114	-0.287
PartEtrangUE	-0.137	-0.096	0.120	0.014	-0.026	0.004	-0.125	-0.037	-0.238	0.234	0.009	-0.003
TauxAccr	-0.085	-0.030	0.126	-0.008	0.007	0.016	-0.083	-0.070	-0.072	0.065	-0.005	0.013
PartDipSup	-0.355	-0.041	0.388	-0.163	-0.110	0.416	-0.567	-0.268	-0.408	0.396	0.249	0.500
PartSSDipSec	0.368	0.047	-0.482	0.268	0.188	-0.402	0.610	0.259	0.455	-0.441	-0.261	-0.489
PartMenage1pers	0.337	-0.023	-0.337	0.253	0.013	-0.197	0.547	0.282	0.631	-0.621	-0.622	-0.377
PartMonoparental	1.000	0.030	-0.409	0.015	0.098	-0.270	0.571	0.126	0.581	-0.558	-0.233	-0.353
PartNouveauFam5pe	0.030	1.000	-0.208	-0.063	0.219	0.051	-0.500	-0.345	-0.586	-0.098	0.267	0.124
TauxActivite	-0.409	-0.208	1.000	-0.523	-0.575	0.152	-0.493	-0.024	-0.418	0.413	0.033	0.214
TauxDepViel	0.015	-0.063	-0.523	1.000	0.785	0.032	0.610	0.259	0.455	-0.441	-0.261	-0.489
TauxDependance	0.098	0.219	-0.575	0.785	1.000	0.057	0.547	0.282	0.631	-0.621	-0.622	-0.377
PartIndependance	-0.270	0.051	0.152	0.032	0.057	1.000	-0.385	-0.044	-0.230	0.082	0.256	0.076
TauxChomage	0.571	0.098	-0.493	0.059	0.016	-0.385	1.000	-0.044	0.640	-0.624	-0.360	-0.481
PartLogAnt45	0.126	0.070	-0.024	-0.021	-0.030	-0.402	-0.385	0.200	0.640	-0.624	-0.360	-0.481
PartLogOccLocataire	0.581	0.098	-0.418	0.119	0.073	-0.230	0.640	0.077	1.000	-0.982	-0.353	-0.397
PartLogOccPropri	-0.558	-0.098	0.413	-0.129	-0.082	0.413	-0.624	-0.067	-0.982	1.000	0.327	0.375
PartHabParLog	-0.233	0.267	0.033	-0.035	0.256	0.196	-0.360	-0.159	-0.353	0.327	1.000	0.408
PartLogplus6piece	-0.353	0.124	0.214	-0.067	0.076	0.451	-0.481	-0.022	-0.397	0.375	0.408	1.000

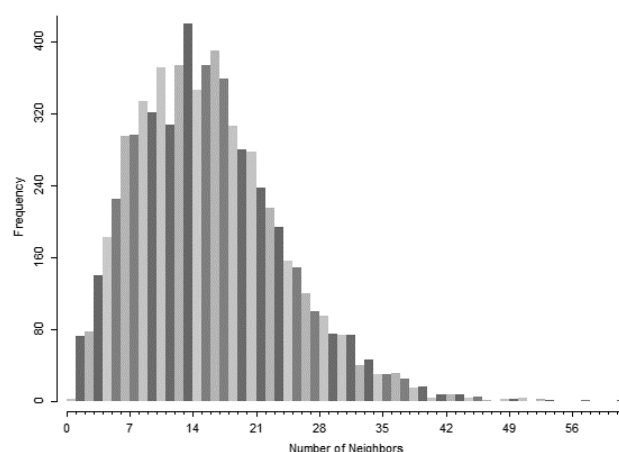
Annexe 13 : Construction de la matrice de voisinage des secteurs statistiques

La fréquence observée du nombre de voisins par secteur statistique est donnée par :

Pour un voisinage d'ordre 1 et une contiguïté de type « Queen »



Pour un voisinage d'ordre 2 et une contiguïté de type « Queen »

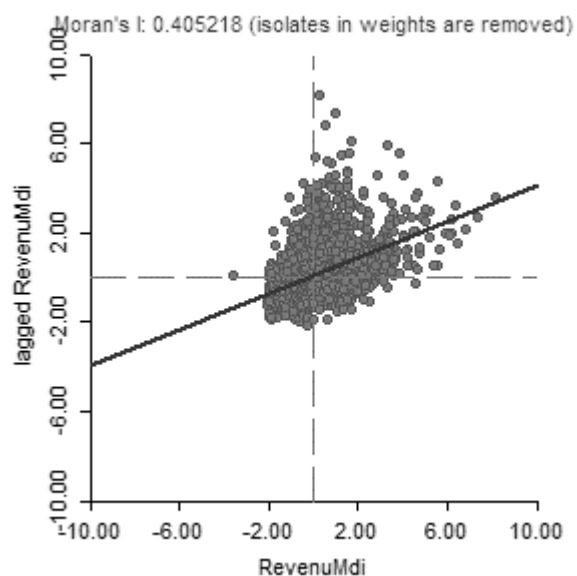


Comparaison des matrices de voisinage		
	Ordre 1	Ordre2
Enclaves	865	72
Îles	0	2
Nombre de voisins minimum	0	0
Nombre de voisins maximum	20	61
Moyenne du nombre de voisins	4.5	15.29
Médiane	4	14

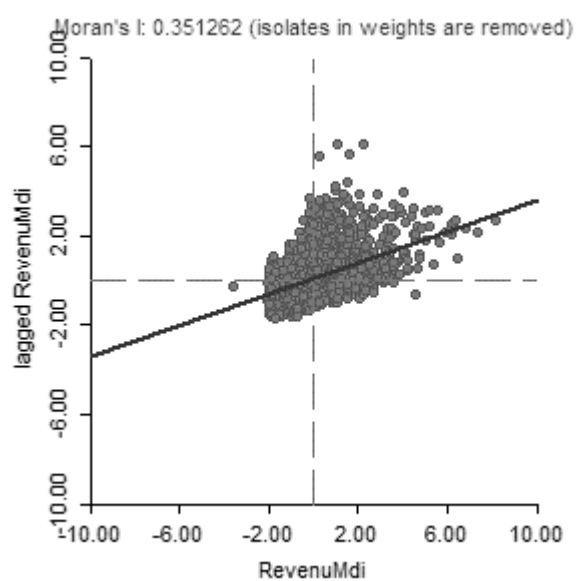
Îles supprimées dans la matrice de voisinage d'ordre 2	
Identifiant du secteur	Nom du secteur
61068D10-	BORSET-CENTRE
25048A19-	MOLEMBAIS SAINT-JOSSE

Annexe 14 : Calcul de l'indice de Moran pour le revenu médian par secteur statistique en fonction de l'ordre de la matrice de voisinage

Indice de Moran pour une matrice de voisinage d'ordre 1



Indice de Moran pour une matrice de voisinage d'ordre 2



Annexe 15 : Régression avec interaction sur le revenu médian des secteurs statistiques (7.512)
avec 3 variables indépendantes

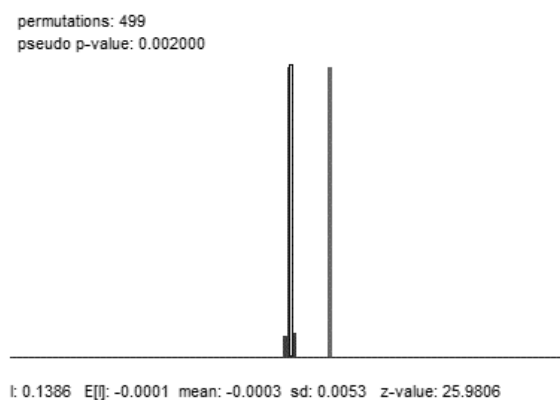
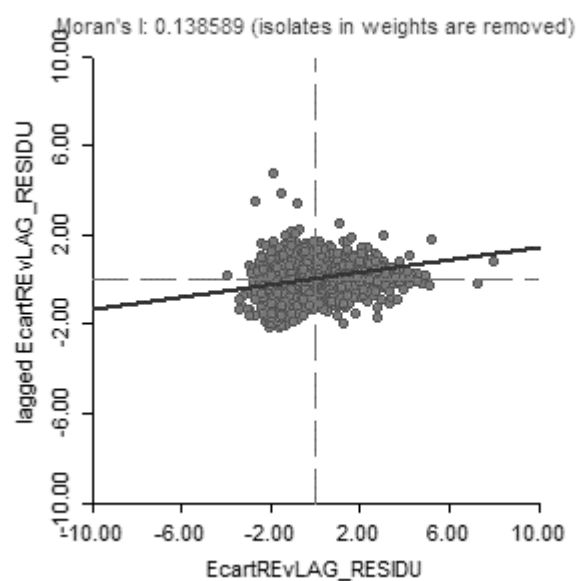
Régression avec interaction sur le revenu médian des secteurs statistiques (Y)				
Nombre de secteurs : 7.512				
X1 = PartDipSup * X2= PartMénage1pers * X3=TauxChômage				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Intercepte	20968,5	448,9	46,715	< 2e-16
Part de diplôme supérieur	-5061,8	1476,6	-3,428	0.000611
Part de ménage 1 personne	31095,0	1284,8	24,202	< 2e-16
Taux de chômage	-19131,4	2736,0	-6,993	2.93e-12
PartMénage1pers:PartDipSup	-42465,9	4527,0	-9,381	< 2e-16
PartMénage1pers:TauxChomage	11253,7	6916,7	1,627	0.103770
PartDipSup:TauxChomage	54860,2	11185,7	-4,905	9.56e-07
PartMénage1pers:PartDipSup:TauxChomage	87909,9	29411,0	2,989	0.002808
Ecart type des résidus	3054	Degré de liberté		7502
R²	0,6237	R² ajusté		0,6233
Statistique F	1776	Degré de liberté		(7, 7502)
p-valeur	2,2e-16			

Annexe 16 : Régression sur le revenu médian des secteurs statistiques (7.510) avec 4 variables indépendantes

Régression sur le revenu médian des secteurs statistiques (Y)				
Nombre de secteurs : 7.510				
X1 = PartDipSup + X2= PartMénage1pers + X3=TauxChômage + X4=ClasseMoran				
Où X4 = variable auxiliaire de 4 modalités				
Coefficient	Estimation du coefficient	Erreur standard	Statistique t	p-valeur
Intercepte	25184,91	171,73	146,651	< 2e-16
Part de diplôme supérieur	11335,58	353,45	32,071	< 2e-16
Part de ménage 1 personne	-8643,67	362,32	-23,856	< 2e-16
Taux de chômage	-12071,96	621,36	-19,428	< 2e-16
LH	-4111,63	116,98	-35,149	< 2e-16
HH	212,62	53,98	3,939	8,27e-05
LL	-1237,88	40,57	-30,510	< 2e-16
HL	/	/	/	/
Ecart type des résidus	2686		Degré de liberté	7503
R²	0,7089		R² ajusté	0,7087
Statistique F	3046		Degré de liberté	(6,7503)
p-valeur	2,2e-16			



Annexe 18 : Autocorrélation des résidus suite à l'auto-régression LAG de l'écart relatif des revenus



Annexe 19 : Résultat de la première ACP comprenant 15 variables

	PC1	PC2	PC3	PC4	PC5	PC6
Log de la densitéHabitant	-0.594	-0.136	0.013	0.273	-0.170	-0.016
Taux d'accroissement	0.130	-0.109	-0.221	-0.181	0.907	-0.013
Part de marié	0.683	-0.196	-0.362	0.031	-0.223	-0.046
Part popu monoparentale	-0.727	0.090	0.087	-0.207	0.020	0.099
Log de la part noyaux fam de plus de 5pers	-0.165	0.497	0.196	0.302	0.116	-0.570
Part d'étranger de l'UE	0.190	-0.201	-0.417	0.544	0.094	0.295
Part de diplôme supérieur	0.686	0.083	0.199	-0.209	-0.087	0.008
Part d'indépendant	0.516	0.293	0.309	-0.209	0.008	0.465
Taux de dépendance	-0.160	0.749	-0.369	0.096	0.046	0.270
Taux d'activité	0.597	-0.543	0.321	0.014	0.130	-0.180
Taux de chômage	-0.838	-0.028	-0.009	-0.051	0.020	-0.009
Part de logement ant 45	-0.283	-0.044	0.587	0.567	0.210	0.268
Part de logement occupé par le propriétaire	0.757	-0.079	-0.123	0.364	-0.031	-0.058
Part de logement de plus de 6pièces	0.637	0.389	0.307	0.030	0.011	0.118
Part d'habitant par logement	0.441	0.602	-0.070	0.094	0.060	-0.258
Valeur Propre	4.502	1.835	1.233	1.102	0.999	0.902
Proportion de variance	30%	12%	8%	7%	7%	6%
Proportion cumulée de la variance	30%	42%	50%	58%	64%	70%

PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
Log de la densitéHabitant	0.310	0.505	0.256	-0.001	0.073	0.294	-0.111	0.049	0.014
Taux d'accroissement	0.052	0.229	0.018	0.074	0.045	0.053	0.015	-0.021	-0.017
Part de marié	-0.105	0.242	-0.277	0.248	0.146	0.013	-0.032	-0.196	0.206
Part popu monoparentale	0.187	-0.232	0.223	0.220	0.392	-0.197	-0.133	-0.038	0.067
Log de la part noyaux fam de plus de 5pers	0.319	0.051	-0.362	-0.045	0.051	-0.127	-0.038	-0.003	0.010
Part d'étranger de l'UE	0.445	-0.387	-0.044	0.029	-0.076	0.019	-0.006	-0.032	-0.010
Part de diplôme supérieur	0.438	0.137	0.276	-0.058	-0.036	-0.120	0.329	-0.116	0.070
Part d'indépendant	0.173	0.007	-0.375	-0.151	0.223	0.202	-0.030	0.085	0.001
Taux de dépendance	-0.103	0.186	0.121	-0.195	-0.093	-0.178	-0.068	0.082	0.207
Taux d'activité	0.038	-0.161	0.117	-0.153	-0.044	-0.004	-0.186	0.143	0.274
Taux de chômage	0.003	-0.106	-0.177	0.232	-0.065	0.133	0.308	0.206	0.193
Part de logement ant 45	-0.285	0.071	0.014	0.011	0.040	-0.045	0.101	-0.177	0.065
Part de logement occupé par le propriétaire	-0.163	0.110	0.115	0.050	0.301	-0.137	0.144	0.274	-0.099
Part de logement de plus de 6pièces	0.068	0.061	0.081	0.474	-0.244	-0.016	-0.154	0.093	-0.038
Part d'habitant par logement	-0.166	-0.325	0.277	0.006	0.116	0.353	0.043	-0.085	0.051
Valeur Propre	0.819	0.790	0.699	0.487	0.422	0.395	0.333	0.256	0.226
Proportion de variance	5%	5%	5%	3%	3%	3%	2%	2%	2%
Proportion cumulée de la variance	76%	81%	86%	89%	92%	95%	97%	98%	100%

Annexe 20: Résultat final de l'ACP sur 14 indices

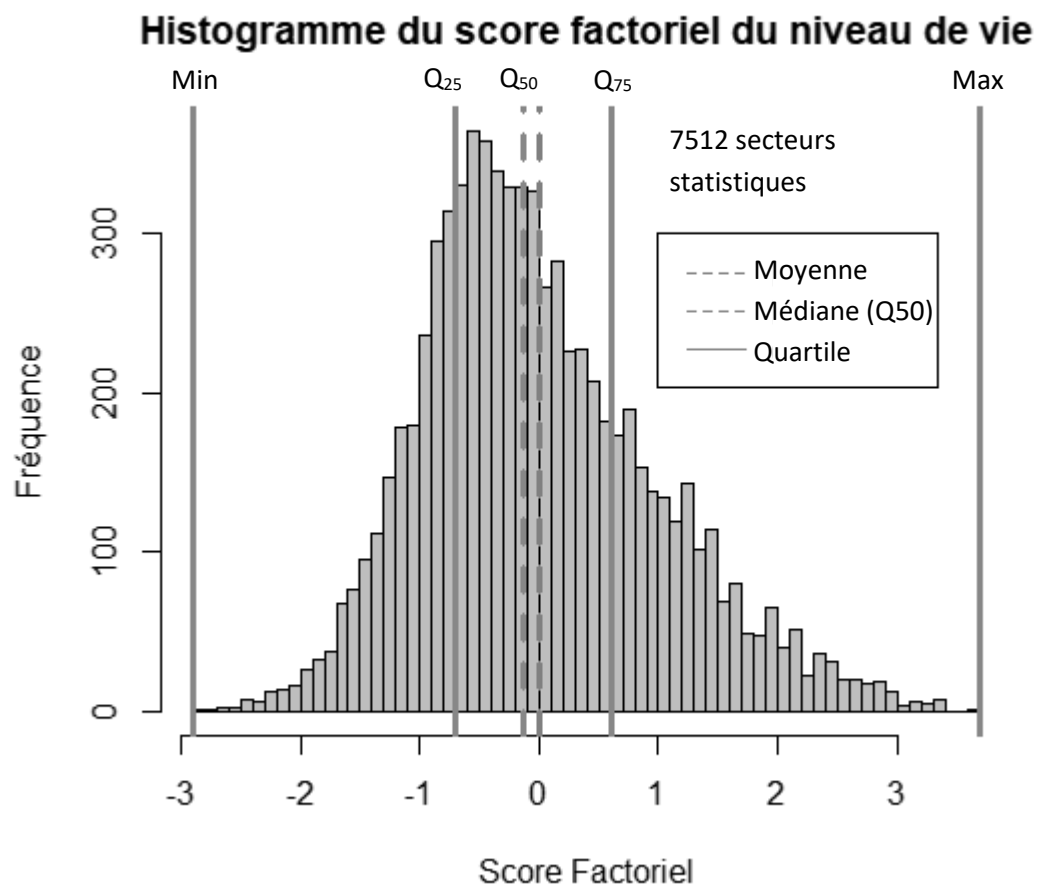
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Revenu médian	-0.801	-0.262	-0.055	0.168	-0.123	0.224	0.135	-0.017	0.124	-0.188	0.093	0.126	0.314	-0.021
Log de la densité d'habitant	0.585	-0.131	-0.238	0.044	0.017	0.641	-0.234	0.078	-0.242	0.216	0.032	-0.033	0.057	0.001
Part de marié	-0.681	-0.440	-0.088	0.022	0.394	-0.055	-0.199	0.216	0.145	0.086	0.077	0.137	-0.084	0.174
Part de ménage 1 personne	0.758	0.203	-0.291	0.337	0.096	-0.047	-0.016	-0.310	0.039	-0.100	-0.079	-0.072	0.087	0.217
Part de la population monoparentale	0.683	0.053	0.277	0.077	-0.350	0.093	0.149	0.486	0.138	-0.048	-0.152	-0.040	0.024	0.107
Part d'étranger totale	0.465	-0.232	0.302	0.160	0.568	0.212	0.481	-0.040	0.055	0.021	-0.029	-0.071	-0.033	-0.044
Part de diplôme supérieur	-0.692	0.151	-0.058	0.456	-0.223	0.321	0.081	-0.031	-0.006	-0.194	0.139	-0.008	-0.265	0.014
Part d'indépendant	-0.462	0.533	0.153	0.416	0.301	-0.200	-0.047	0.260	-0.287	0.030	0.093	-0.028	0.114	-0.009
Taux d'activité	-0.567	0.039	-0.551	-0.069	-0.192	-0.110	0.452	0.053	-0.029	0.310	0.080	-0.088	0.010	0.052
Taux de chômage	0.828	-0.047	0.111	-0.023	-0.051	-0.097	-0.079	0.033	0.186	0.004	0.469	-0.159	0.030	-0.005
Part de logement ant 45	0.317	0.601	-0.322	-0.479	0.252	0.205	0.119	0.109	0.078	-0.169	0.085	0.177	-0.016	0.014
Part de logement occupé par le locataire	0.787	0.079	0.139	0.320	-0.158	-0.094	0.085	-0.095	0.029	0.230	0.069	0.378	-0.026	-0.024
Part de logement de plus de 6 pièces	-0.628	0.474	0.191	0.082	0.059	0.207	-0.158	-0.089	0.422	0.257	-0.083	-0.078	0.033	-0.028
Part d'habitant par logement	-0.503	0.115	0.665	-0.357	-0.123	0.143	0.104	-0.191	-0.201	0.068	0.100	0.023	0.020	0.159
Valeur Propre	5.772	1.293	1.264	1.028	0.912	0.798	0.642	0.525	0.461	0.389	0.329	0.262	0.205	0.121
Proportion de variance	41%	9%	9%	7%	7%	6%	5%	4%	3%	3%	2%	2%	1%	1%
Proportion cumulée de la variance	41%	50%	59%	67%	73%	79%	84%	87%	91%	93%	96%	98%	99%	100%

Annexe 21 : Tableau des secteurs dont le score factoriel est supérieur à 3

Secteur dont le score est supérieur à 3				
INS	Secteur statistique	Score factoriel du niveau de vie	Suppression	Justification de la suppression
52011A374	ABATTOIR	3.891	Supprimé	Abattoir, 528 habitants, Commune de Charleroi
52011G0PR	PONT DE PHILIPPEVILLE	3.843	Supprimé	Entreprise Alstom et parking. 124 habitants, Commune de Charleroi
62096A172	L'ESPERANCE	3.649	Supprimé	Accelor Mital, 50 habitants, Commune de Liège
52011E123	YERNAUX	3.646	Conservé	Terril, d'immeubles et maisons. 745 habitants, Commune de Charleroi
63079A172	QUARTIER EST	3.415	Supprimé	Essentiellement des entreprises. 271 habitants, Commune de Verviers
52011J122	LE SPIGNAT	3.368	Conservé	Immeubles et maisons à proximité d'une carrière non comprise dans le secteur. 748 habitants, Commune de Charleroi
53070A031	NORD-HAINE	3.360	Conservé	Lotissement, lycée, prairie. 439 habitants, Commune de Saint-Ghislain
62096A242	LA BERGERIE BAS	3.349	Conservé	Nombreux d'immeubles, garage BMW et autres magasins 1355 habitants, Commune de Seraing
62120A822	LA FONTAINE	3.343	Conservé	Ensemble de maisons. 654 habitants, Commune de Flémalle
52011A00-	CHARLEROI - VILLE HAUTE	3.325	Conservé	Centre de Charleroi. 3260 habitants, Commune de Charleroi
56016J162	CAMARGUE - LONGUE TAILLE	3.317	Conservé	Ensemble de maisons. 194 habitants, Commune de Chimay
62120A402	LES TRIXHES-CENTRE	3.315	Conservé	Immeubles, école primaire, terrain vague. 913 habitants, Commune de Flémalle
92094B312	SEDEN	3.298	Supprimé	Lycée, Liddle, terrain de foot, club de rugby, immeubles. 1410 habitants, Commune de Namur
25121A503	BICREAU-CENTRE	3.280	Supprimé	Centre de Louvain-La-Neuve. 459 habitants, Commune d'Ottignies-Louvain-La-Neuve
52012B512	LA CHAMBRE	3.277	Conservé	Ilôt d'immeuble, parking, terrain vague. 68 habitants, Commune de Chatelet
52011E0MH	FRICHE INDUSTRIELLE	3.257	Supprimé	Friche industrielle. 99 habitants, commune de Charleroi
52011G0WR	PARC D'ACTIVITES ECONOMIQUES	3.246	Supprimé	Parc d'activité économique. 56 habitants, Commune de Charleroi
52022A20-	CITE DES OISEAUX	3.224	Conservé	Maisons et immeubles, 679 habitants, Commune Fontaine-L'éveque
25121A724	CENTRE SPORTIF	3.224	Supprimé	Centre sportif de Louvain-la-Neuve. 263 habitants, Commune d'Ottignies-Louvain-La-Neuve
63079A61-	RUE DE DISON	3.214	Conservé	Ensemble de maisons. 2866 habitants, Commune de Verviers
63079A00-	VERVIERS-CENTRE	3.206	Conservé	Centre de Verviers. 2459 habitants, Commune de Verviers

62118B112	CITE "LOGIS-MILITAIRES"	3.204	Conservé	Ensemble de maisons, 338 habitants, Commune de Grace-Hollogne
53053D243	QUARTIER AUX FLEURS	3.195	Conservé	Ensemble d'immeubles et maisons. 633 habitants, Commune de Mons
63079A600	HODIMONT	3.194	Conservé	Ensemble de maisons. 1850 habitants, Commune de Verviers
53053U412	CITE DE LA FRANCOPHONIE	3.173	Conservé	Ensemble de maisons. 616 habitants, Commune de Mons
62118D11-	FONTAINE-CITE	3.166	Conservé	Présence d'une écurie mais le reste est composé de maisons. 121 habitants, Commune de Grace-Hollogne
62096A07-	COCKERILL	3.163	Supprimé	Ancien haut-fourneau. 85 habitants, Commune de Seraing
52011G221	CITE PARC	3.131	Conservé	Ensemble d'immeubles, maisons. 1811 habitants, Commune de Charleroi
53053G113	SHAPE - MAISIERES	3.127	Supprimé	Petite zone de maisons et le reste est privé car c'est une zone militaire 817 habitants, Commune Mons
53014B032	BAS CORTILS - BOUVEAUX	3.109	Conservé	Moitié maisons, moitié retail park. 773 habitants, Commune de Boussu
57081A973	TOURNAI - ZONE INDUSTRIELLE	3.094	Supprimé	Chemin de fer, Intermarché, entrepôt et autres entreprises. 100 habitants Commune de Tournai
63040B063	GRUNSTRASSES BACH ERHOLUNGSGEB.	3.047	Conservé	Appartements, maisons. 163 habitants, Commune de Kelmis.
56087A31-	HOME FAMILIAL	3.033	Conservé	Ensemble de maisons. 430 habitants, Commune de Morlanwelz
52011A433	VIADUC	3.026	Supprimé	Garage, chemin de fer, industrie. 404 habitants, Commune de Charleroi
53053A04-	MONS STATION	3.020	Conservé	Gare de Mons et habitations à proximité. 681 habitants, Commune de Mons
52015A102	COURCELLES-CENTRE - GARE	3.014	Conservé	Maisons, appartements, Trafic, magasins. 220 habitants, Commune de Courcelles

Annexe 22 : Histogramme des fréquences du score factoriel des 7512 secteurs obtenus après suppression des valeurs aberrantes



Annexe 23 : Tableau des valeurs extrêmes de l'ANOVA à un facteur sur le niveau de vie et le statut d'inondation

Valeurs extrêmes des secteurs inondés (4)			
ID du secteur statistique	Nom du secteur statistique	Score factoriel	Classe d'inondation
52012B512	LA CHAMBRE	3.277	Inondé
53053D243	QUARTIER AUX FLEURS	3.195	Inondé
53070A031	NORD-HAINE	3.360	Inondé
56016J162	CAMARGUE - LONGUE TAILLE	3.317	Inondé
81015C090	HABERGY-HABITATIONS DISPERSEES	-2.849	Inondé

Valeurs extrêmes des secteurs non inondés (123)			
ID du secteur statistique	Nom du secteur statistique	Score factoriel	Classe d'inondation
82003C390	LUTREBOIS - HABITATIONS DISP.	-2.766	Non inondé
83034B222	LA FAMENNOISE	2.537	Non inondé
85009C090	CHANTEMELLE-HABITATIONS DISP.	-2.684	Non inondé
85045A00-	VIRTON-CENTRE	2.411	Non inondé
85046A090	HABAY-LA-NEUVE-HAB. DISP.	-2.593	Non inondé
52010A41-	RUCHE CHAPELLOISE	2.955	Non inondé
52011A00-	CHARLEROI - VILLE HAUTE	3.325	Non inondé
52011A02-	PARC REINE ASTRID	2.396	Non inondé
52011A031	BEAUX ARTS	2.659	Non inondé
52011A04-	UNIVERSITE DU TRAVAIL	2.772	Non inondé
52011A0AR	STADE	2.840	Non inondé
52011A0BR	SPINOIS	2.515	Non inondé
52011A222	MAMBOURG	2.684	Non inondé
52011A301	LE FAUBOURG	2.296	Non inondé
52011A31-	BOIS DEL BOL	2.338	Non inondé
52011A322	CITES "RESIDENCES"	2.296	Non inondé
52011A333	L'ALOUETTE	2.788	Non inondé
52011B11-	CARABIN	2.535	Non inondé
52011C0BH	WARCHAT	2.438	Non inondé
52011C2MR	LODELINSART-GARE	2.467	Non inondé
52011D010	QUARTRE BRAS	2.320	Non inondé
52011D40-	LES HAIES	2.305	Non inondé
52011E123	YERNAUX	3.646	Non inondé
52011E201	CITE SELESTAT	2.637	Non inondé
52011F032	BOIS DES CLOCHES	2.775	Non inondé
52011G221	CITE PARC	3.131	Non inondé
91034A062	VAL DE CHESSION	2.996	Non inondé
52011H432	MON LOGIS	2.907	Non inondé
52011J122	LE SPIGNAT	3.368	Non inondé
91034A800	GEMECHENNE	2.443	Non inondé
52011J42-	BOUGNOU	2.362	Non inondé
52011J452	TROISIEME ECLUSE	2.974	Non inondé
52011K212	CITE MALGHEM	2.327	Non inondé
52011M04-	CITE DE LA LACHE	2.635	Non inondé
52011N132	BAYEMONT	2.825	Non inondé
52011P00-	GOSSELIES-CENTRE	2.469	Non inondé
52011P422	CITE HUBINON ET SPINOIS	2.425	Non inondé
52012A022	LA CALVAIRE	2.351	Non inondé
52012B4AJ	LE LONG PRE	2.367	Non inondé

52015A102	COURCELLES-CENTRE - GARE	3.014	Non inondé
52015A220	MIAUCOURT	2.681	Non inondé
91103A062	DOMAINE MAIEUR FRANCOIS	2.302	Non inondé
52022A20-	CITE DES OISEAUX	3.224	Non inondé
91141A083	TRICOINTE	2.887	Non inondé
52043E012	PARC DE BELLECOURT	2.515	Non inondé
91142A464	BOIS DE LENS	2.531	Non inondé
52048A042	MALFALIM	-2.593	Non inondé
52074B09-	LES VIEUX-SARTS	2.749	Non inondé
52074C042	LE BATI	2.384	Non inondé
53014A401	CITE DU FOYER MODERNE	2.760	Non inondé
53020A312	CHEVALIERES	2.532	Non inondé
53020D01-	FERRAND	2.615	Non inondé
53028A121	CITE PIERRARD	2.830	Non inondé
53028C021	PLANTINCHOU	2.424	Non inondé
53053A00-	MONS-CENTRE	2.892	Non inondé
53053C312	ESPINETTE	2.344	Non inondé
92094A02-	LE GROGNON	2.454	Non inondé
53053G214	BRUYERE DE JURBISE	2.539	Non inondé
53053T412	CITE DU CAMPIAU	2.858	Non inondé
53053U412	CITE DE LA FRANCOPHONIE	3.173	Non inondé
92094J212	HASTEDON	2.883	Non inondé
53065A301	CITE ASTRID	2.385	Non inondé
53065A312	LA FLACHE	2.587	Non inondé
53082B34-	CITE GERMINAL-RES. COLFONTAINE	2.311	Non inondé
54007A323	OUTRE-MEUSE	2.899	Non inondé
92140E052	CITE	2.971	Non inondé
55022A11-	CITE DU BOCAGE	2.598	Non inondé
55022A31-	CITE ASTRID	2.896	Non inondé
55022C14-	CITE BEAU SITE	2.439	Non inondé
55022F030	CITE LEBURTON (MAURAGE)	2.845	Non inondé
55022J032	CITE DE GENIVAL	2.606	Non inondé
93022H212	MINIERES	2.582	Non inondé
56029B163	CUL DE CHEVAL	2.338	Non inondé
56087A31-	HOME FAMILIAL	3.033	Non inondé
57081A052	SAINT-JACQUES	2.321	Non inondé
57081A522	RESIDENCE CARBONELLE	2.439	Non inondé
57081V043	VAL DE LA RESIDENCE	2.449	Non inondé
62060H09-	PAIFVE-HABITATIONS DISPERSEES	2.823	Non inondé
62063A101	SAINT-CHRISTOPHE	2.492	Non inondé
62063A50-	SAINTE-MARGUERITE	2.693	Non inondé
62118B112	CITE "LOGIS-MILITAIRES"	3.204	Non inondé
62118D11-	FONTAINE-CITE	3.166	Non inondé
61080A100	LES FAGNES	2.415	Non inondé
61080D012	CITE VANDEWEGHE	2.694	Non inondé
62003A211	MOLINVAUX-LA-TONNE	2.664	Non inondé
62051A302	LAIXHAUT	2.755	Non inondé
62051B0AN	QUARTIER DE LA CHAPELLE	2.691	Non inondé
25117C012	LE BOSQUET	2.390	Non inondé
62063F20-	BONNE-FEMME	2.344	Non inondé
25121A322	BAULOY	2.304	Non inondé
25121A608	BRUYERE-CENTRE	2.355	Non inondé
62093B10-	HAUT-DE-TILLEUR	2.335	Non inondé
62096A210	VAL-POTET	2.757	Non inondé
62096A221	LA BERGERIE-BAS	2.315	Non inondé
62096A242	LA BERGERIE BAS	3.349	Non inondé

62096A442	PRE-SORAY	2.753	Non inondé
62096B201	CHANT DES OISEAUX	2.563	Non inondé
25124C19-	LERINNES-HABITATIONS DISP.	-2.610	Non inondé
62096C000	OUGREE-BAS	2.736	Non inondé
62096C042	OUGREE NICOLAY	2.889	Non inondé
62096C101	BIENS-COMMUNAUX	2.888	Non inondé
62100B122	SAPINIERE -FORESTIER	2.421	Non inondé
62120A221	LES TRIXHES SUD	2.683	Non inondé
62120A402	LES TRIXHES-CENTRE	3.315	Non inondé
62120A822	LA FONTAINE	3.343	Non inondé
62120A833	LES GRANDS PRES	2.304	Non inondé
62120B312	RAMIOUL EST	2.485	Non inondé
63020A00-	DISON-CENTRE	2.810	Non inondé
63020A20-	PISSEROULE	2.591	Non inondé
63040B063	GRUNSTRASSESBAACH ERHOLUNGSGEB.	3.047	Non inondé
63075B190	BOIS JACQUES MATHIEU	2.919	Non inondé
63079A00-	VERVIERS-CENTRE	3.206	Non inondé
63079A01-	PLACE VERTE	2.993	Non inondé
63079A02-	NOTRE DAME DE RECOLLETS	2.783	Non inondé
63079A03-	GARES CENTRALE ET OUEST	2.775	Non inondé
63079A101	SAINT-REMACLE	2.903	Non inondé
63079A11-	PRES JAVAIS	2.405	Non inondé
63079A30-	MANGOMBROUX	2.935	Non inondé
63079A40-	RUE DU PALAIS	2.323	Non inondé
63079A50-	GERARD-CHAMPS	2.387	Non inondé
63079A600	HODIMONT	3.194	Non inondé
63079A61-	RUE DE DISON	3.214	Non inondé
63079F120	CLARISSSES	2.353	Non inondé

Annexe 24 : Tests non paramétriques à 1 facteur

Comparaison de la médiane des groupes inondés et non inondés à la médiane générale Test de Kruskal-Wallis (Chi ²)			
	Degré de liberté	Chi ²	p-valeur
	1	0.0619	2.2e-16

La médiane générale est significativement différente de la médiane des groupes.

Comparaison par paire des médianes des groupes Test de Wilcoxon	
	p-valeur
Inondé – Non inondé	2e-16

Les médianes des groupes sont significativement différentes.

Annexe 25 : Statistiques descriptives du score factoriel des groupes formés par leur appartenance à un quadrant du diagramme de Moran et leur statut d'inondation

Statistiques descriptives du score factoriel selon chaque quadrant du diagramme de Moran et le statut d'inondation					
Classe Moran	Classe d'inondation	Total	Moyenne	Erreur standard	Q25
LH	Inondé	95	-0.423	0.343	-0.633
	Non-inondé	594	-0.530	0.422	-0.725
HH	Inondé	528	1.25	0.758	0.663
	Non-inondé	1620	1.03	0.713	0.460
LL	Inondé	475	-0.666	0.503	-0.967
	Non-inondé	3067	-0.749	0.487	-1.03
HL	Inondé	293	0.562	0.487	0.190
	Non-inondé	838	0.518	0.524	0.150
Classe Moran	Classe d'inondation	Médiane	Q75	Min	Max
LH	Inondé	-0.158	-0.308	-0.0182	-1.65
	Non-inondé	-0.208	-0.430	-0.00780	-2.59
HH	Inondé	1.77	1.17	3.36	0.00824
	Non-inondé	1.45	0.925	3.65	-0.00222
LL	Inondé	-0.282	-0.570	-0.00527	-2.85
	Non-inondé	-0.385	-0.682	-0.00512	-2.77
HL	Inondé	0.783	0.440	2.58	-0.00320
	Non-inondé	0.710	0.365	2.97	-0.00369

Annexe 26 : Tests non paramétriques à 2 facteurs

Comparaison de la médiane des classes de Moran à la médiane générale Test de Kruskal-Wallis (Chi ²)			
	Degré de liberté	Chi ²	p-valeur
	3	5686.7	2.2e-16

La médiane générale est significativement différente de la médiane des groupes.

Comparaison par paire des médianes des groupes Test de Wilcoxon (Sans l'hypothèse d'égale variance)	
	p-valeur
HH - LH	2e-16
LL- LH	2e-16
HL-LH	2e-16
LL-HH	2e-16
HL-HH	2e-16
HL-LL	2e-16

Les médianes des groupes sont significativement différentes.