

Perspectives de population au niveau des communes wallonnes : rapport méthodologique

Jean-Paul Sanderson¹, Luc Dal²,
Thierry Eggerickx³, Michel Poulain

¹ Logisticien de recherche (FNRS) au Centre de recherche en démographie de l'UCL (jean-paul.sanderson@uclouvain.be)

² Informaticien au Centre de recherche en démographie de l'UCL

³ Professeur et Maître de recherche (FNRS) au Centre de recherche en démographie de l'UCL

1. INTRODUCTION

Aujourd'hui, les administrations locales sont de plus en plus confrontées aux questions d'aménagement de leur territoire. En Belgique, si ce sont les Régions qui fixent les règles en matière de logement et d'aménagement du territoire (Wooncode en Flandre⁴, Code wallon du logement en Région wallonne⁵ et Code bruxellois du logement⁶), ce sont, en partie, les communes qui veillent à leur application tout en disposant d'une certaine autonomie communale pour faire face à ces problèmes.

Or, les communes ne sont pas forcément outillées pour faire face à ces questions. Un premier outil a été mis au point pour les communes wallonnes, CYTISE qui fournit des données statistiques à l'échelle des communes et des quartiers. Toutefois, pour résoudre les questions liées à l'aménagement du territoire, il est essentiel de pouvoir se projeter dans un avenir à court ou moyen termes et de disposer d'estimations de chiffre de la population, de répartitions par âge et par type et/ou taille des ménages.

Parallèlement, les Régions doivent pouvoir estimer l'évolution de leur population et de sa répartition dans l'espace régional afin de pouvoir déterminer les besoins et les moyens à mettre en œuvre. En 2010-2011, le Centre de recherche en démographie de l'Université catholique de Louvain mettait au point, à la demande de l'Institut Wallon de l'Évaluation, de la Prospective et de la Statistique (IWEPS). L'enjeu était de taille puisqu'en Belgique, un tel exercice n'a jamais été réalisé ni pour la Wallonie ni pour Bruxelles.

Cette demande a abouti en la réalisation d'un jeu de projections avec scénarios alternatifs. L'objectif du présent projet est d'actualiser le jeu de projection et de transférer la méthodologie à l'IWEPS.

La complexité de l'exercice réside principalement dans le défi posé par les petits nombres. De fait, la plupart des auteurs de perspectives locales ne projettent pas les populations en-deçà d'un certain seuil de population (Sanderson et al., 2011). La principale raison de ce choix est liée à la problématique des petits nombres. Les décès, les naissances et les migrations peuvent être qualifiés «d'événements rares» et cette caractéristique s'amplifie au fur et à mesure que le découpage spatial s'affine (Eggerickx, Poulain, 1996). Les solutions adoptées consistent le plus souvent à ne pas descendre en dessous d'un seuil critique. En France, dans le cas du modèle OMPHALE (Outil Méthodologique de Projection d'Habitants d'Actifs de Logements et d'Elèves) utilisé actuellement (Bergouignan, 2008 ; Rudant, 2011), celui-ci est fixé à 50.000 habitants. Si ce choix permet de contourner la difficulté liée aux petits nombres, il ne permet pas de prendre en compte la grande diversité des situations rencontrées au niveau local. Pour les 262 communes wallonnes, seules 9 (3%) comptent au moins 50.000 habitants en 2013 et 223 (85%) en recensent moins de 20.000⁷. D'autres méthodes ont été développées, par exemple :

⁴ Vlaamse Wooncode, décret du Gouvernement flamand du 15 juillet 1997 (modifié en 1999, 2004, 2005, 2006, 2008, 2009 et 2010)

⁵ Code wallon du Logement, décret du Gouvernement wallon du 29 octobre 1998 (modifié en 2000, 2001, 2002, 2003, 2005, 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014)

⁶ Code bruxellois du Logement, ordonnance du Gouvernement de la région de Bruxelles-capitale du 17 juillet 2003 (modifié en 2008 et 2009)

⁷ Source : www.cytise.be

- La méthode des tendances partielles (Banens, 1994) préconise de réaliser les perspectives à un niveau agrégé selon la méthode des composantes et de redistribuer ensuite le résultat obtenu proportionnellement au poids de chaque localité. Concrètement, dans le cas de la Belgique, il s'agirait de se baser sur les projections de population réalisées par le Bureau fédéral du Plan à l'échelle des arrondissements et de répartir le chiffre de population de chaque arrondissement entre les communes qui le composent. Le poids de chaque commune dans l'arrondissement pouvant être estimé en analysant les tendances observées pour les années antérieures. Les ménages sont déduits a posteriori. Cette méthode présente l'avantage de ne pas être affectée par le problème des petits nombres, mais elle a l'inconvénient de ne pas tenir compte des spécificités communales et de l'information disponible à l'échelle des communes.
- La méthode des projections par profil socio-économique des individus appliquée à la Suède (Fransen et Karlson, 2010) permet de réaliser des projections à un niveau territorial très fin. Elle consiste à appliquer la méthode des composantes aux populations communales à ceci près que lorsque le chiffre de population est insuffisant, les paramètres de fécondité, de mortalité et de migration vont être déterminés au niveau national. Toutefois, afin de tenir compte des spécificités de la composition sociale des espaces locaux, on va calculer pour l'ensemble du pays, les paramètres du mouvement de la population par catégorie sociale, celle-ci étant déterminée par le niveau d'instruction, l'âge, le type de logement occupé... On construit ainsi des comportements types par catégorie sociale et on applique ces comportements types aux communes proportionnellement à leur composition sociale. Cette méthode présente l'avantage de trouver une solution au problème des petits nombres et de tenir compte de certaines disparités locales (dans la composition socio-économique), mais présente l'inconvénient de considérer que les comportements individuels sont uniquement tributaires de la catégorie sociale et ne sont nullement affectés par le milieu d'habitat. Enfin, le nombre de ménage est déduit a posteriori.

Enfin, dans le cadre de certaines entités, par exemple Montréal (Maurois et Bélanger, 2014), des outils de micro-simulation ont été élaborés pour répondre à des demandes spécifiques. Toutefois, ce type d'outil n'a pas été conçu pour réaliser des projections à l'échelle d'un grand nombre de communes.

Ce rapport présentera d'abord les problèmes liés à la démographie locale, et ensuite la méthode originale employée dans le cadre de cet exercice de projection de la population et des ménages des communes de Wallonie.

2. PROBLÈMES ET SPÉCIFICITÉS DE LA DÉMOGRAPHIE LOCALE

Les principaux problèmes liés à la démographie locale sont les petits nombres et la disponibilité des données.

2.1. Le problème des petits nombres

En démographie, la méthode la plus classique de projection de population est la méthode des composantes. Elle consiste à faire évoluer une population à partir d'hypothèses portant sur l'évolution de chaque composante du mouvement démographique : fécondité, mortalité et migration. On part d'une population distribuée par âge et sexe pour une année donnée et on la projette en lui appliquant les quotients de mortalité, les taux de fécondité et les bilans migratoires estimés par sexe et par âge (ou groupe d'âges quinquennal). Or, sous un certain seuil

de population, le risque est grand de voir ces paramètres varier de façon très aléatoire suite à un « accident statistique ». Ainsi, si on prend l'exemple de la mortalité pour 3 communes (Daverdisse, Dinant et Namur) de tailles différentes, on le voit sur le graphique 1 et le tableau 1 que le risque est grand d'avoir un nombre de décès nul du fait de la faiblesse des effectifs. Même dans une commune de plus de 13.000 habitants, comme Dinant, les effectifs sont faibles compte tenu de la rareté de l'événement et empêche tout calcul fiable.

En d'autres termes, les quotients de mortalité, les taux de fécondité et dans une moindre mesure les bilans migratoires sont très sensibles au problème des petits nombres qui, sans un minimum de précautions méthodologiques et d'esprit critique s'appuyant sur une connaissance approfondies des tendances démographiques locales, – risque d'invalidier nombre de résultats de projection. La dimension locale est donc un terrain miné où les spécialistes de projections ont rarement osé s'aventurer.

Figure 1. La distribution des décès à Daverdisse (1426 hab.), Dinant (13.668) et Namur (110.500) entre 2008 et 2012 par groupe quinquennal d'âges (source : Cytise-RN)

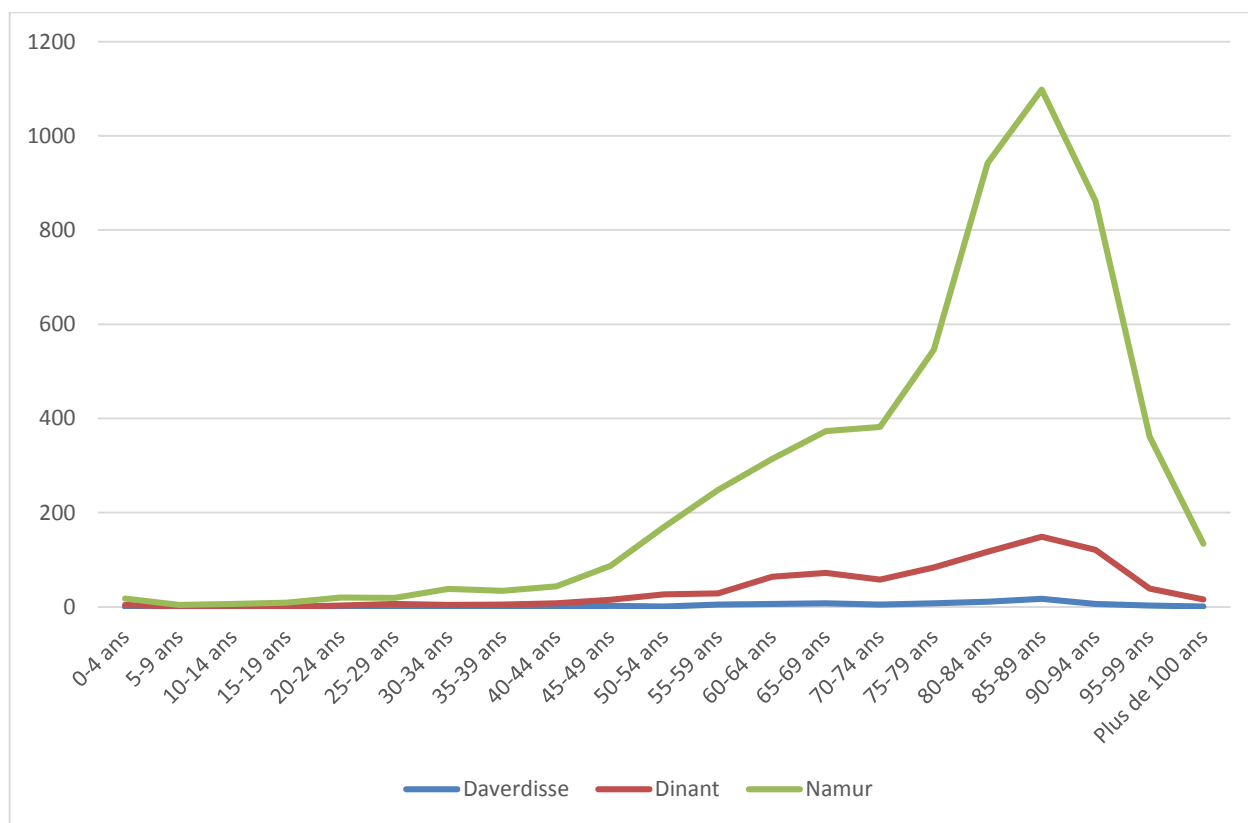


Tableau 1. Nombre de décès à Daverdisse (1426 hab.), Dinant (13.668) et Namur (110.500) entre 2008 et 2012 par groupe quinquennal d'âges (Source : Cytise-RN)

Groupes d'âges	Daverdisse	Dinant	Namur
0-4 ans	0	4	18
5-9 ans	0	1	4
10-14 ans	0	2	6
15-19 ans	0	0	9
20-24 ans	0	3	20
25-29 ans	1	6	19
30-34 ans	0	4	38
35-39 ans	1	5	34
40-44 ans	0	8	44
45-49 ans	2	15	87
50-54 ans	1	27	170
55-59 ans	5	29	248
60-64 ans	6	64	314
65-69 ans	8	72	373
70-74 ans	5	58	382
75-79 ans	8	84	546
80-84 ans	11	117	942
85-89 ans	17	149	1098
90-94 ans	6	121	862
95-99 ans	3	39	362
Plus de 100 ans	1	16	134
Total	75	824	5710

2.1. La disponibilité des données

Etablir des projections démographiques suppose une excellente connaissance des situations démographiques actuelles et passées. Celle-ci repose inévitablement sur la disponibilité et la qualité des données au niveau local. En Belgique, nous avons la chance de disposer du Registre national qui permet une observation continue et fiable des populations communales, au moins depuis 1991. Néanmoins, l'exploitation de ces données nécessite une solide expertise scientifique dans la production des informations nécessaires aux projections démographiques, informations qui ne sont pas disponibles auprès de la DGSIE à l'échelle locale, telles que les tables de mortalité, les taux de fécondité, la ventilation de l'immigration et de l'émigration par âge et sexe, les types et taille des ménages.

En d'autres termes, une connaissance fine de la diversité démographique locale, des évolutions différenciées et des mécanismes qui les sous-tendent constituent des conditions nécessaires à l'élaboration des hypothèses et des scénarii les plus probants. Il en est de même en particulier des mouvements migratoires et des mécanismes qui déterminent leur intensité, leur orientation et leurs composantes sociodémographiques.

2.2. Les spécificités de la démographie locale : la variabilité

Dans le contexte occidental, à l'échelle des pays, des régions et des sous-régions (exemple des arrondissements en Belgique ou des départements en France), on observe une tendance générale à l'uniformisation des comportements démographiques (Monnier, 2004). En d'autres termes, les écarts entre les niveaux de fécondité et de mortalité ont sensiblement diminué... ce

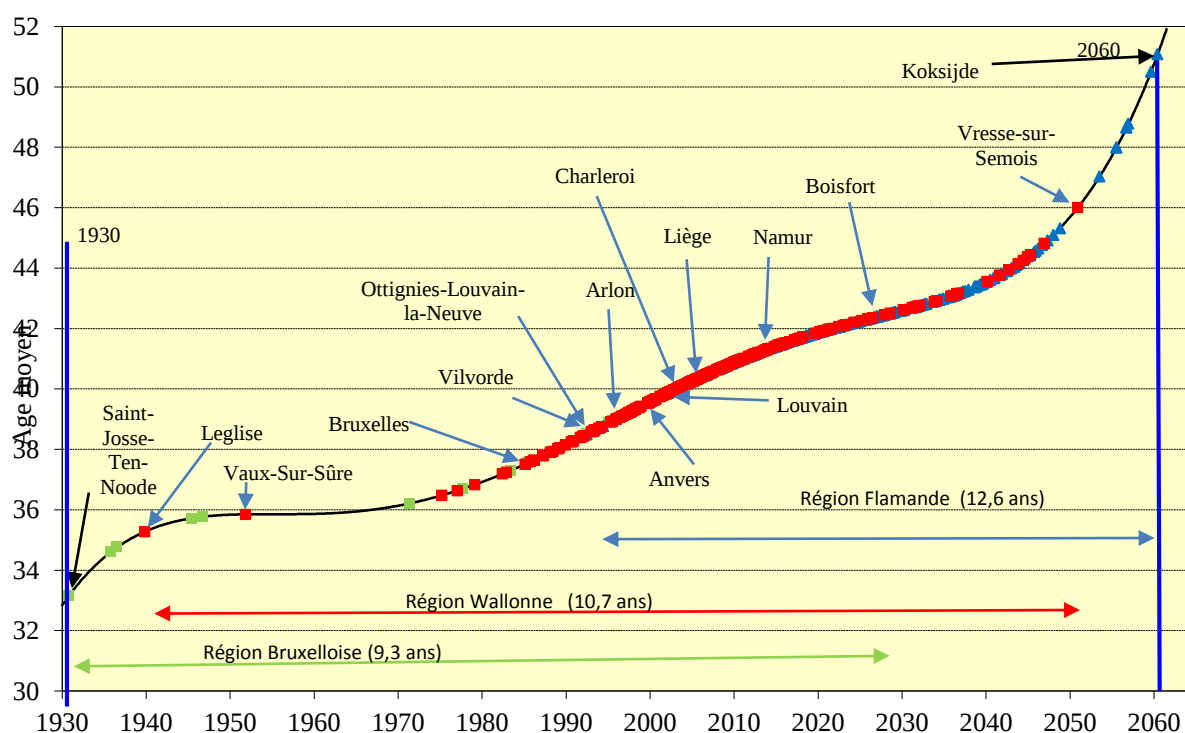
qui facilite les hypothèses d'évolutions différenciées de ces comportements (convergence) alimentant les projections régionales ou par arrondissement.

Or, à l'échelle des communes, les écarts demeurent très importants. D'une manière générale, plus le maillage spatial considéré est fin, plus le nombre d'unités spatiales est important et plus l'éventail des différences est grand (Eggerickx, Poulain, 1996). En outre, l'hypothèse de convergence des comportements démographiques ne se vérifie pas forcément à l'échelle de « sous-régions » ou des communes.

Ces aspects seront illustrés par un exemple concret, portant sur l'âge moyen de la population.

Il y a aujourd'hui moins d'une année d'écart d'âge moyen entre la Flandre et la Wallonie, 7 années entre l'arrondissement de Bastogne et celui de Furnes, mais 18 années entre la commune la plus jeune de Belgique et la plus vieillie. Si nous plaçons la valeur de l'âge moyen de 2013 de chaque commune sur une échelle du temps qui suit l'évolution de l'âge moyen de l'ensemble de la population belge, on constate que la commune la plus jeune, Saint-Josse-ten-Noode, à l'âge moyen qu'avait la population de la Belgique en 1930, alors que les communes les plus âgées – par exemple Koksijde – ont l'âge moyen qu'aurait la Belgique en 2061, selon les dernières projections démographiques du Bureau du Plan. En termes de processus de vieillissement, il y a donc un écart de plus d'un siècle entre ces cas extrêmes (figure 2).

Figure 2. L'âge moyen de la population au 1er janvier 2013 et la position de chaque commune sur l'échelle du temps correspondant à l'évolution de la moyenne nationale (source : DGSIE-RN)



En outre, les dynamiques de changement à l'échelle locale sont rapides. La cartographie de l'âge moyen s'est radicalement modifiée entre 1970 et 2013 (figures 3 et 4). Non seulement, les espaces communaux de sur-vieillessement et de sous-vieillessement ne sont plus les mêmes, mais certaines communes se caractérisent même par une évolution contraire aux tendances généralement observées. Ainsi, Bruxelles et un certain nombre de communes hesbignonnes

figuraient parmi les plus vieilles en 1970 ; aujourd’hui, ces communes se distinguent par un rajeunissement démographique inédit et par un âge moyen de leur population parmi les plus faibles. Ces évolutions particulières se justifient, dans le cas de Bruxelles, par l’intensification des migrations internationales, et dans le cas des communes hesbignonnes par le caractère sélectif selon l’âge des migrations internes (Eggerickx et al., 2010).

Figure 3. L’âge moyen de la population des communes en 1970 (source : DGSIE)

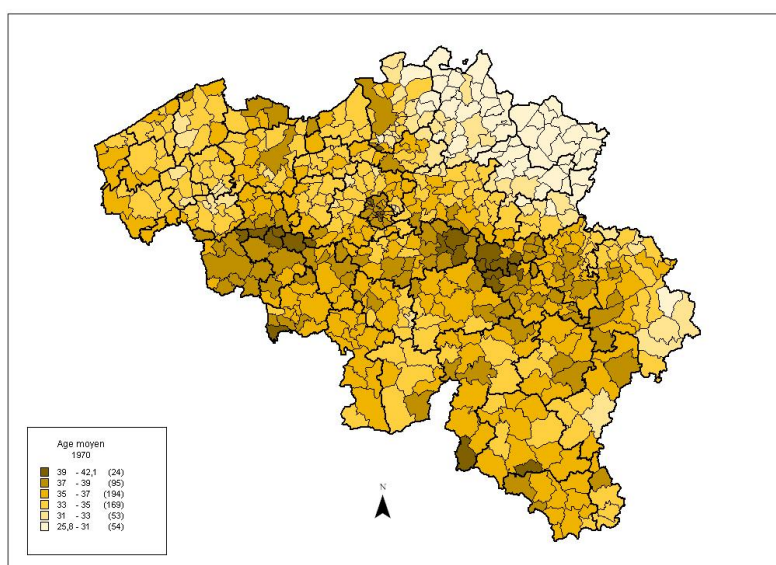
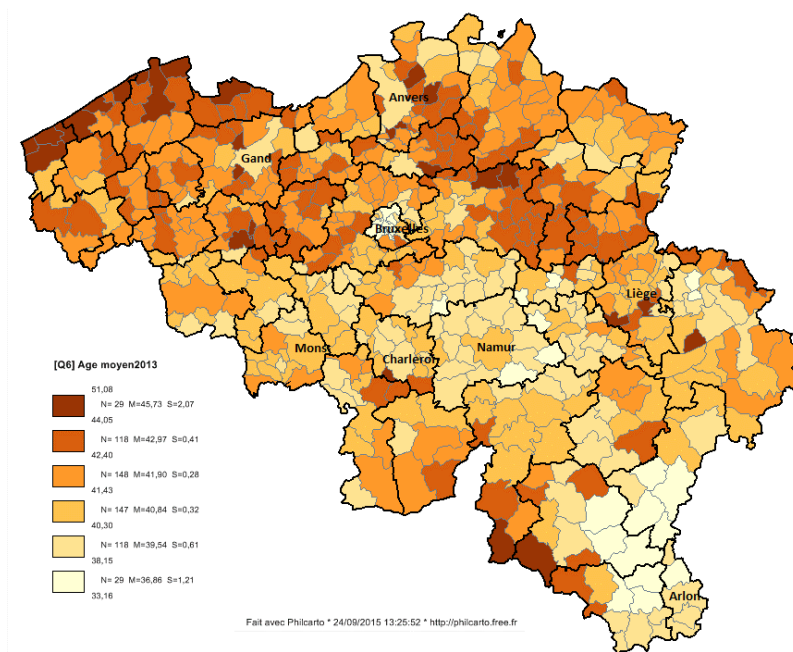


Figure 4. L’âge moyen de la population des communes en 2013 (source : DGSIE)



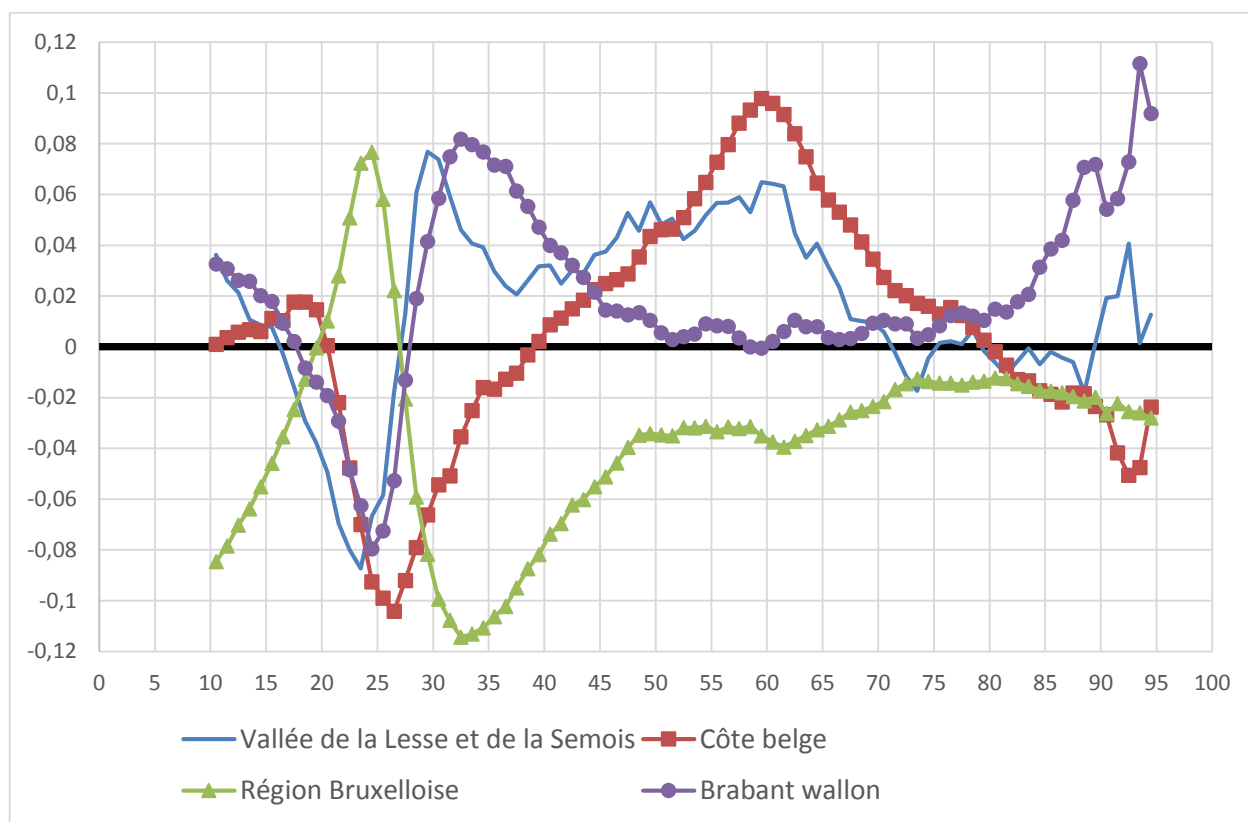
Un constat similaire a pu être posé pour les phénomènes démographiques que sont la fécondité (Costa et al., 2010 ; Brée et al., 2014) ; la mortalité (Eggerickx, Sanderson, 2010 ; Deboosere, Fizman, 2009). Il est donc indispensable de tirer au mieux parti des informations disponibles afin de pouvoir prendre en compte la variabilité des phénomènes démographiques au niveau local.

2.3. Le rôle déterminant des migrations à l'échelle des communes

A l'échelle des communes, les mouvements migratoires se caractérisent par leur fréquence (en moyenne, par commune, 5 fois plus d'entrées et de sorties que de naissances et de décès), leur poids dans la croissance démographique et leur sensibilité conjoncturelle.

La fréquence des mouvements migratoires à l'échelle des communes en fait des éléments déterminants de leur dynamique démographique. Les migrations sont le principal moteur de l'évolution du chiffre de la population des communes et exercent aussi une influence déterminante sur l'évolution des structures par âge (Sanderson et al., 2009) de la distribution des ménages et de la composition socioéconomique des populations.

Figure 5. Les bilans migratoires par âge selon certains types de communes (2006-2010) (source : DGSIE)



L'intensité et la composition sociodémographique des migrations internes varient d'un type de commune à l'autre. Deux éléments guident les choix résidentiels :

- la position des individus et des ménages dans leur cycle de vie : l'offre d'emploi et de logements locatifs ainsi que les atouts socioculturels des villes rencontrent les aspirations de nombreux jeunes adultes au moment de leur émancipation ; les espaces périurbains rencontrent les attentes des 30-45 ans et de leurs enfants en proposant des logements plus vastes dans un cadre champêtre à proximité des axes de pénétration en ville et les personnes âgées migrant au moment de la retraite, etc. Ainsi, en fonction du cycle de vie des individus et des ménages et des choix résidentiels qui y sont associés, les bilans migratoires selon l'âge varient sensiblement selon le type de communes (Courgeau, Lelièvre, 2003). Des courants migratoires plus importants et plus spécifiques à certains

groupes d'âges sont donc observés vers des lieux bien déterminés en réponse à des attraits bien précis (figure 5).

- Le second point focal autour duquel s'organisent les migrations internes concerne l'offre de logement (importance, coût, type et taille des logements...).

3. MÉTHODE DE PROJECTION

Outre l'échelle locale, l'autre spécificité de notre méthode de projection est de produire simultanément des projections de population et de ménage. Une telle simultanéité ne peut s'obtenir qu'à partir d'un modèle multi-états de type LIPRO ou de celui produit dans le cadre du projet « Mobidic » (Gusbin et al., 2007). Pour ces deux modèles, les individus, à un temps donné, sont caractérisés par leur appartenance, non seulement, à un groupe d'âges et à un sexe, mais également à un type de ménage et à leur position au sein de celui-ci. Au fil du temps, les individus évoluent d'un état initial vers un état final, avec une certaine probabilité. A partir de la connaissance de ces probabilités et de l'état de départ, le processus de projection génère les individus dans leur état final. En itérant ce processus, on calculera de la sorte les effectifs projetés par âge, sexe et type de ménage.

La présentation de la méthode, ci-dessous, développera le propos en prenant étape par étape les problèmes rencontrés et les solutions adoptées. On commencera par la définition des états ; ensuite, on abordera la solution trouvée pour résoudre les problèmes de variations aléatoires dû aux petits nombres : le lissage gaussien ; suivra la technique d'estimation et de projection des transitions d'un état à l'autre pour enfin terminer par la description de la méthode de projection (en ce compris les calculs de la mortalité, de la fécondité et de la migration).

3.1. La définition des états

Compte tenu des objectifs, quatre variables vont permettre de définir les états, auxquels il faudra ajouter l'état de né, de décédé, d'immigrant et d'émigrant. Ces quatre variables sont :

- le sexe : 2 possibilités : homme – femme ;
- l'âge : 18 possibilités (le dernier groupe est celui des personnes âgées de 85 ans et plus) définies par les âges quinquennaux ;
- la commune de résidence : 589 communes belges ;
- le ménage : 42 possibilités en combinant la taille et le type de ménage.

Tableau 2. Types de ménages et position des individus dans le ménage

Type de ménage	Position			
	1	2	3	4
1. Isolés	Personne de référence			
2. Couples mariés sans enfant	Personne de référence	Conjoint		
3. Cohabitants sans enfant	Personne de référence	Conjoint		
4. Couples mariés avec 1 enfant	Personne de référence	Conjoint	Enfant	
5. Couples mariés avec 2 enfants	Personne de référence	Conjoint	Enfant	
6. Couples mariés avec plus de 2 enfants	Personne de référence	Conjoint	Enfant	
7. Cohabitants avec 1 enfant	Personne de référence	Conjoint	Enfant	
8. Cohabitants avec 2 enfants	Personne de référence	Conjoint	Enfant	
9. Cohabitants avec plus de 2 enfants	Personne de référence	Conjoint	Enfant	
10. Familles monoparentales avec 1 enfant	Personne de référence		Enfant	
11. Familles monoparentales avec 2 enfants	Personne de référence		Enfant	
12. Familles monoparentales avec plus de 2 enfants	Personne de référence		Enfant	
13. Autres ménages privés de deux personnes	Personne de référence	Conjoint	Enfant	Autre personne
14. Autres ménages privés de 3 personnes	Personne de référence	Conjoint	Enfant	Autre personne
15. Autres ménages privés de plus de 3 personnes	Personne de référence	Conjoint	Enfant	Autre personne
16. Ménages collectifs				

En combinant l'ensemble des possibilités présentées, on a :

589 communes *1.428 possibilités (âge, sexe et ménage) + naissances + décédés + immigrants + émigrants.

L'examen de la distribution des individus selon le groupe d'âge et le type de ménage auxquels ils appartiennent (par exemple, la commune de Gerpinnes en 2013) montre qu'un grand nombre de catégories compte un effectif réduit. On peut aisément imaginer ce qu'il en serait en y ajoutant le sexe des individus et la taille des ménages.

Tableau 3. Distribution des individus selon leur âge et le type de ménages – Gerpinnes, 1/1/2013

Classes d'âges	Isolés	Couples sans enfant	Couples avec enfants	Cohabitants sans enfant	Cohabitants avec enfants	Monoparentaux	Autres	Total
0 à 4 ans	0	0	205	0	189	70	53	517
5 à 9 ans	0	0	289	0	162	116	38	605
10 à 14 ans	0	0	367	0	121	179	43	710
15 à 19 ans	3	0	398	1	76	171	81	730
20 à 24 ans	36	1	346	60	61	178	91	773
25 à 29 ans	64	35	211	96	76	75	58	615
30 à 34 ans	64	26	190	55	168	51	49	603
35 à 39 ans	79	16	253	52	132	87	62	681
40 à 44 ans	94	25	417	41	162	107	54	900
45 à 49 ans	121	56	407	29	96	120	65	894
50 à 54 ans	130	157	388	58	45	102	93	973
55 à 59 ans	136	312	290	52	23	68	82	963
60 à 64 ans	145	472	113	53	10	42	67	902
65 à 69 ans	140	485	70	46	4	20	44	809
70 à 74 ans	118	278	37	18	4	13	40	508
75 à 79 ans	115	220	11	11	0	15	45	417
80 à 84 ans	121	143	12	8	1	18	47	350
85 à 89 ans	75	68	4	2	1	14	26	190
90 à 94 ans	26	10	1	0	0	6	11	54
95 à 99 ans	5	0	0	0	0	0	2	7
100 ans et plus	0	0	0	0	0	0	0	0
Total	1472	2304	4009	582	1331	1452	1051	12201

Il est donc essentiel de limiter le nombre d'états potentiels. Compte tenu des variables retenues pour définir les états, seule la variable ménage offre des possibilités de regroupement. Une analyse menée sur les comportements migratoires et sur le type de logement occupé par les différents types de ménage a permis de réduire le nombre de catégories liées au ménage à 9 :

1. isolés,
2. couples mariés sans enfants,
3. couples mariés avec un seul enfant,
4. couples mariés avec 2 enfants,
5. autres ménages de taille 2,
6. autres ménages de taille 3,
7. chefs de ménage de couples mariés avec 3 enfants ou plus et chefs de ménages des autres ménages de taille supérieure ou égale à 4,
8. autres membres de couples mariés avec 3 enfants ou plus et autres membres des autres ménages de taille supérieure ou égale à 4,
9. collectifs.

A ces 9 états, on en ajoute 3 qui correspondent aux entrées et sorties du système⁸. Ce sont les

10. naissances / décès
11. immigrants internes / émigrants internes
12. immigrants externes / émigrants externes

Dans les lignes qui suivent, ces situations de ménages seront désignées par $m_1, \dots, m_9$ et m_{10}, m_{11}, m_{12} . Selon que l'on considère la situation en début ou en fin de période, il n'y a aucune ambiguïté sur les notations m_{10}, m_{11} et m_{12} .

3.2. Technique d'estimation des transitions entre les états

L'objectif de l'exercice est de projeter les populations et les ménages à l'horizon 2035. La méthodologie est basée sur (a) des données détaillées sur la population au niveau local par âge, sexe et situation de ménages et de tous les événements démographiques pour la période 2000-2015 (b) Les chiffres de toutes les transitions possibles entre les Etats observées au terme de chaque périodes de cinq ans 2000-2005, 2005-2010 et 2010-2015 et (c) une estimation robuste de chaque probabilité de transition qui permet l'extrapolation à l'avenir.

La méthode étant une méthode multi-états, tout le monde est caractérisé à un moment donné selon quatre dimensions: la commune, le sexe, 5 ans groupe d'âge et mode de vie. En appliquant les probabilités de transition extrapolées à la population initiale dans le temps "t", l'algorithme calcule la population dans le temps "t + 5", avec les mêmes caractéristiques permettant la projection simultanée du nombre de personnes et du nombre de ménages avec leurs caractéristiques.

Les données utilisées proviennent du système d'enregistrement continu de la population de droit belge appelé "Registre National". Nous disposons d'une « photo » de la situation au 1er janvier de chacune des années 2000, 2005, 2010 et 2015. Pour chaque individu, on dispose de la commune de résidence, l'année de naissance, le sexe et le type de ménage. Il est donc possible de calculer les probabilités de transition entre deux dates. Ces données sont exhaustives et couvrent l'ensemble de la population résidant en Belgique. La fiabilité de ces données est

⁸ Au sens strict, il faudrait parler d'états « non encore né », « non encore immigré », « non encore décédé » et « non encore émigré » puisque les migrations, naissances et décès ne sont pas des états mais des événements.

considérée comme exceptionnelle, une situation similaire avec celles des pays nordiques, des Pays-Bas et de l'Autriche.

Les entrées et sorties du système, entre deux dates, sont les naissances, les immigrations internes au pays, les immigrations externes (en provenance de l'étranger), les décès, les émigrations internes au pays et les émigrations externes (à destination de l'étranger).

A un instant donné, chaque individu est caractérisé par son sexe, sa commune de résidence, sa situation de ménage et sa classe d'âge (révolu) quinquennal.

Dans une commune donnée, au cours d'une période de 5 ans, il effectue une transition d'une situation de ménage à une autre situation de ménage (éventuellement la même) et il vieillit de 5 ans. De plus, il peut entrer dans la commune, dans une situation de ménage (par naissance ou immigration interne ou en provenance de l'étranger) ou en sortir (par décès ou émigration interne ou à destination de l'étranger).

Cela donne lieu à une famille de matrices de transitions de dimensions 12 par 12 avec donc 144 possibilités de transition dont 9 ne seront pas considérées (les transitions entre les 3 états complémentaires : par exemple, les nés décédés, les nés émigrés internes, les nés émigrés internationaux, les immigrés internes décédés, ...) car elles concernent des personnes qui entrent et sortent pendant la période et ne modifient en rien les effectifs de population finals par rapport à la situation initiale.

Un individu de sexe s_j , se trouvant dans une commune c_i , dans une situation de ménage m_l , et appartenant à la classe d'âges a_k au temps « t » appartient à un état noté e_{c_i, s_j, a_k, m_l} .

A la fin de la période de 5 ans, il est nécessairement dans un des 4 états suivants :

- $e_{c_i, s_j, a_{k+1}, m_r}$ s'il est survivant dans la même commune dans un ménage de type « r »,
- $e_{c_i, s_j, a_{k+1}, m_{10}}$ s'il est décédé,
- $e_{c_i, s_j, a_{k+1}, m_{11}}$ s'il a émigré en Belgique,
- $e_{c_i, s_j, a_{k+1}, m_{12}}$ s'il a émigré vers l'étranger.

Inversement, un individu se trouvant dans un état $e_{c_i, s_j, a_{k+1}, m_l}$ en « t + 5 » était dans un des 4 états suivants 5 ans plus tôt :

- e_{c_i, s_j, a_k, m_r} s'il était vivant dans la même commune dans un ménage de type « r »,
- $e_{c_i, s_j, a_1, m_{10}}$ s'il est né pendant l'année,
- $e_{c_i, s_j, a_k, m_{11}}$ s'il a immigré d'une autre commune,
- $e_{c_i, s_j, a_k, m_{12}}$ s'il a immigré de l'étranger.

Pour chaque sexe, commune et groupe d'âge, les matrices de transitions se présentent sous la forme :

Tableau 4 : structure générale d'une matrice de transition

				$e_{c_i, s_j, a_{k+1}, m_r}$				$e_{c_i, s_j, a_{k+1}, m_{10}}$	$e_{c_i, s_j, a_{k+1}, m_{11}}$	$e_{c_i, s_j, a_{k+1}, m_{12}}$	Situation initiale
...	...	...	...	...	...	...	...	...	...	...	
e_{c_i, s_j, a_k, m_l}	...	...	...	N_{m_r, m_l}	...	...	...	...	...	...	$\sum_{l \leq 10} N_{m_r, m_l}$
...	...	...	...	...	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...	...	...	
$e_{c_i, s_j, a_1, m_{10}}$	...	...	...	...	...	...	...				
$e_{c_i, s_j, a_k, m_{11}}$	...	...	...	...	...	...	...				
$e_{c_i, s_j, a_k, m_{12}}$	...	...	...	...	...	...	...				
Situation finale				$\sum_{r \leq 10} N_{m_r, m_l}$							

L'analyse des observations indique que de nombreuses transitions sont extrêmement rares et pour assurer une robustesse aux projections, les probabilités de transition sont estimées, non pas à l'aide des fréquences observées, mais en faisant appel à un estimateur plus élaboré.

Par exemple, dans le cas de la commune de Gerpinnes, la matrice des comptages des transitions entre le 1/1/2000 et le 1/1/2005, pour les hommes âgés de 20 à 24 ans au 1/1/2000 se présente comme suit :

Tableau 5 : transitions des hommes de 20-24 ans à Gerpinnes entre le 1/1/2000 et le 1/1/2005.

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	1/01/2000
m1	1	0	0	0	1	0	0	0	0	0	5	1	8
m2	0	0	0	0	0	0	0	0	0	0	2	0	2
m3	3	2	29	1	7	2	0	0	0	0	27	7	78
m4	1	4	14	13	8	4	0	3	0	1	34	7	89
m5	3	2	0	1	14	3	0	0	0	0	13	2	38
m6	1	0	0	0	8	1	0	2	0	2	13	2	29
m7	1	0	0	0	0	0	0	0	0	0	0	1	2
m8	8	2	5	7	2	3	0	28	0	0	40	8	103
m9	0	0	0	0	0	0	0	0	0	0	0	1	1
m10										Transitions impossibles			
m11	20	11	14	4	26	7	3	7	0				80
m12	1	1	1	2	0	1	0	2	0				14
1/01/2005	39	22	63	28	66	21	3	42	0	3	134	29	

Au cours de la période pour cette sous-population, on compte un total de 3 décès, de 134 émigrants internes (vers une autre commune) et de 29 émigrations externes. Inversement, on observe 80 immigrants internes (en provenance d'une autre commune), 14 immigrants externes (et bien entendu pour cette classe d'âges, aucune naissance).

Ce type de table est élaboré pour chaque sexe, chaque groupe d'âges, chaque commune et chacune des périodes d'observation, soit un ensemble de $2 * 19 * 262 * 3 = 29\,868$ tables

A partir d'une telle table, on peut immédiatement estimer les probabilités de transition par les fréquences relatives.

Tableau 6 : fréquences relatives des transitions des hommes de 20-24 ans à Gerpinnes entre le 1/1/2000 et le 1/1/2005.

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
m1	12,5%	0,0%	0,0%	0,0%	12,5%	0,0%	0,0%	0,0%	0,0%	0,0%	62,5%	12,5%
m2	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%
m3	3,8%	2,6%	37,2%	1,3%	9,0%	2,6%	0,0%	0,0%	0,0%	0,0%	34,6%	9,0%
m4	1,1%	4,5%	15,7%	14,6%	9,0%	4,5%	0,0%	3,4%	0,0%	1,1%	38,2%	7,9%
m5	7,9%	5,3%	0,0%	2,6%	36,8%	7,9%	0,0%	0,0%	0,0%	0,0%	34,2%	5,3%
m6	3,4%	0,0%	0,0%	0,0%	27,6%	3,4%	0,0%	6,9%	0,0%	6,9%	44,8%	6,9%
m7	50,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	50,0%
m8	7,8%	1,9%	4,9%	6,8%	1,9%	2,9%	0,0%	27,2%	0,0%	0,0%	38,8%	7,8%
m9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%

Remarque : dans ce tableau, on ne considère pas les lignes relatives à m₁₀ m₁₁ et m₁₂ car ces lignes concernent les naissances et les immigrations qui, dans l'algorithme de projection, seront introduites en valeur absolue : cela sera explicité ultérieurement.

On peut y lire, par exemple, que la probabilité de transition de m₈ vers m₃ est estimée par 5/103 = 4,9%. Les probabilités de transition de m₉ vers toute autre situation sont nulles, exception faite de celle vers l'émigration externe (100 %), mais ce cas ne concerne qu'une seule personne ! On peut raisonnablement suspecter que ces estimations ne sont pas les meilleures et risquent de présenter une très forte variabilité.

Si on considère par exemple, dans la même commune, les transitions pour les hommes âgés de 60 à 64 ans en début de période de la situation « autres ménages de taille 2 » (m₅) vers celle de « couple marié avec un seul enfant » (m₃), les observations successives sont les suivantes :

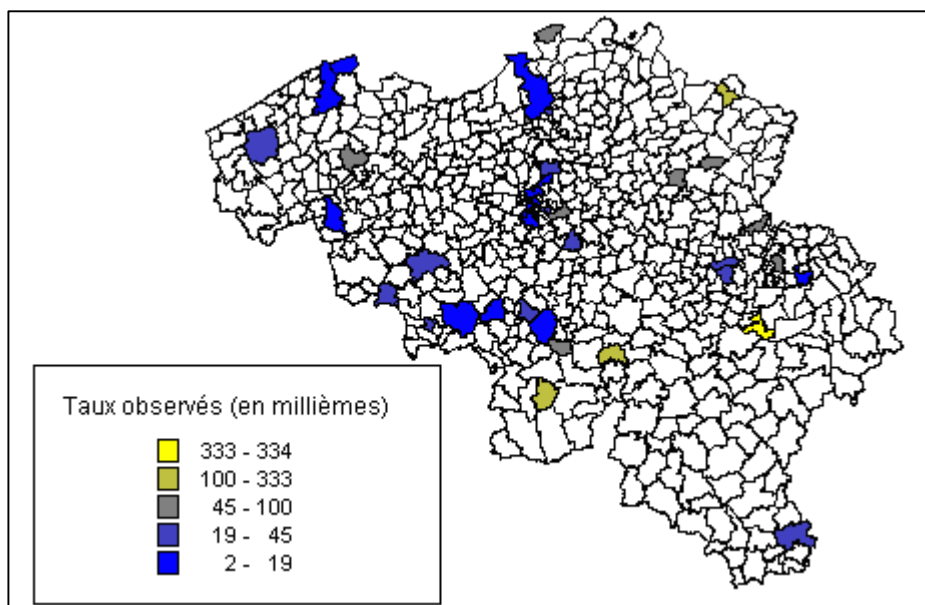
Tableau 7 : comptages des transitions m₅ vers m₃ des hommes de 60-64 ans à Gerpinnes entre le 1/1/2000 et le 1/1/2005.

	2000-2005	2005-2010	2010-2015
Numérateur	1	0	0
Dénominateur	20	20	42
$\hat{\theta}$	5%	0%	0%

Les numérateurs et les dénominateurs sont faibles. La probabilité ainsi calculée présente une forte variabilité aléatoire au fil du temps. Il y a un risque des sous(sur)-évaluation du phénomène. Il est donc nécessaire de recourir à une estimation plus robuste de manière à pouvoir extrapoler les probabilités calculées. La solution envisagée ici consiste à procéder à un lissage gaussien. Pratiquement, il s'agit de procéder à un lissage spatial en complétant l'information disponible à l'échelle d'une commune en regardant ce qui se passe dans les communes voisines et en pondérant l'information par la distance. L'hypothèse sous-jacente étant que des communes proches géographiquement ont des comportements démographiques similaires.

Toutefois, certaines probabilités présentent une forte variabilité spatiale comme le montre la carte dessous.

Figure 6. Probabilités de transition des hommes âgés de 60 à 64 ans le 1/1/2000 d'une situation de ménage de type « autre ménage de taille 3 » à celle de couple marié avec un seul enfant.



Cette absence de structure spatiale montre que la contiguïté spatiale n'implique pas nécessairement la similitude des comportements démographiques et ce d'autant plus que des communes contiguës peuvent avoir des populations socioéconomiquement totalement différentes.

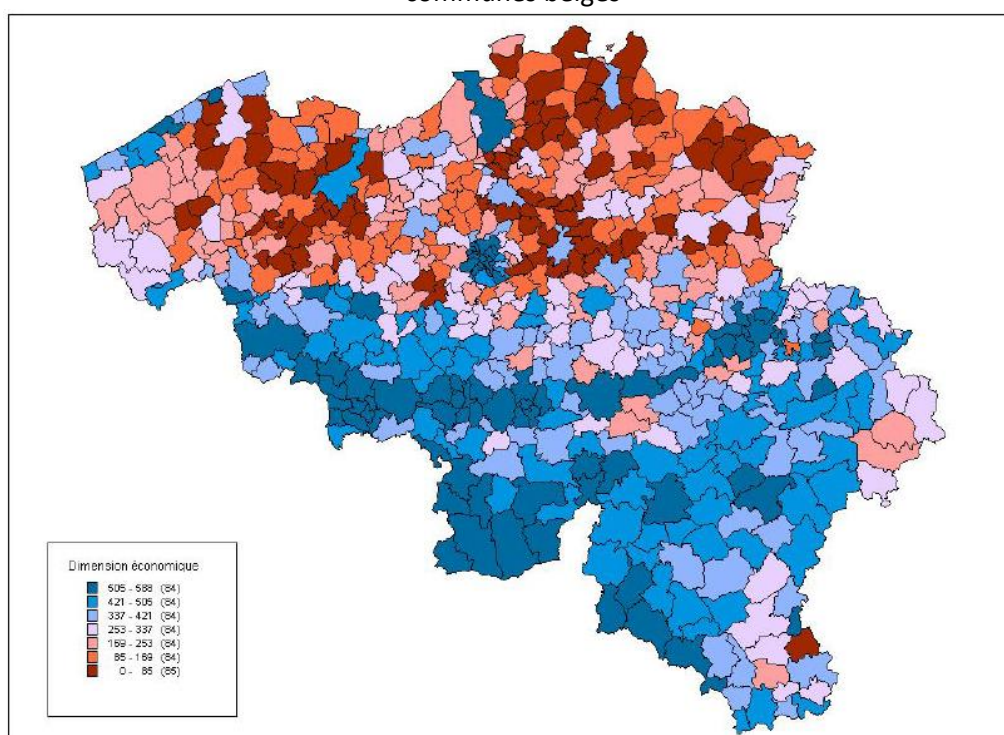
Aussi, afin d'éviter d'intégrer les données de communes ayant des profils totalement différents, nous avons « complexifié » le lissage gaussien en introduisant une « distance généralisée » combinant à poids égaux une distance géographique et une distance socio-économique. Celle-ci est calculée à partir de plusieurs variables traduisant la situation socio-économique de chaque commune. Le choix des indicateurs se base sur une étude antérieure (Eggerickx et al., 2007) qui avait abouti en la construction d'un baromètre des conditions de vie dans les communes belges qui comprenait une dimension socioéconomique. Toutefois, à la différence de l'indicateur calculé dans cette étude, nous avons, ici enlevé les variables démographiques qui contribuaient à la définition de la dimension socio-économique car, on se serait trouvé dans le cas d'une réflexion tautologique où on chercherait à expliquer les variations d'un phénomène démographique par des variables démographiques.

Tableau 8. Liste des indicateurs retenus pour construire la dimension socio-économique

Dimension	Intitulé des variables	Source de l'information
Socio-économique	Revenu médian (1994 et 2008)	(DGSIE-DEMO)
	Taux de chômage (2001)	(DGSIE-ESE-IWEPS)
	Taux de chômage (1991)	(DGSIE-ESE-IWEPS)
	Proportion de personnes se déclarant en mauvaise santé	(DGSIE-ESE-DEMO)
	Part des logements disposant du double vitrage	(DGSIE-ESE-DEMO)
	Part des personnes considérant leur maison en bon état	(DGSIE-ESE-DEMO)
	Part des bénéficiaires du revenu d'intégration sociale dans la population en âge d'activité (2000)	(SPP Intégration sociale IWEPS)
	Part de travailleurs saisonniers ou à temps partiel	(ONSS DMFA, IWEPS)
	% de diplômés disposant d'un diplôme de l'enseignement supérieur (2001)	(DGSIE-ESE-DEMO)
	Taux d'emploi (1991)	(DGSIE-ESE-IWEPS)
	Taux d'emploi (2001)	(DGSIE-ESE-IWEPS)
	Prix moyen par transaction immobilière (1985 et 2008)	(DGSIE-DEMO)
	Part des logements disposant d'une connection internet (2001)	(DGSIE-ESE-DEMO)

La figure 7 présente une cartographie de cette dimension socio-économique. En bleu apparaissent les communes connaissant la situation la moins favorable et en rouge, celles occupant les meilleures positions sur l'indicateur. Si cette carte met en évidence la dualisation Flandre-Wallonie, elle met également en évidence l'hétérogénéité des deux régions et la persistance tant au nord qu'au sud du pays de communes connaissant des situations très différentes de leur voisine. Dès lors, en combinant distance socio-économique et distance géographique, on évite de donner un poids important à une commune certes voisine mais dont le profil socio-économique ne correspondrait pas à celui de la commune pour laquelle il faut appliquer le lissage gaussien. Ainsi, dans le cas de Attert (au-dessus de Arlon), le poids d'Arlon dans le calcul des taux gaussiens, sera plus faible que celui de Habay. Toutes deux jouxtent de Attert mais Arlon a un profil socio-économique plus éloigné de Attert que Habay.

Figure 7. Distribution spatiale de la dimension socio-économique du baromètre des conditions de vie des communes belges



Les variables retenues sont combinées à partir d'une analyse en composante principale dont les deux premiers facteurs permettent de construire un espace bidimensionnel où une distance socio-économique peut être calculée. Les variables retenues pour la construction de cette analyse en composante principale sont celles utilisées pour la construction de la dimension socio-économique du baromètre des conditions de vie en Région wallonne.

Les résultats de l'ACP nous permettent de nous limiter aux deux premiers facteurs puisque ceux-ci rendent compte de 75 % de la variance. Au-delà, les gains successifs en augmentant le nombre de dimensions prises en compte sont nettement plus réduits.

Tableau 9. Variance totale expliquée de l'ACP

Composante	Valeurs propres initiales		
	Total	% de la variance	% cumulé
1	8,727	58,183	58,183
2	2,507	16,713	74,896
3	,954	6,360	81,255
4	,820	5,467	86,722
5	,517	3,445	90,167
6	,324	2,160	92,327
7	,286	1,906	94,233
8	,275	1,836	96,069
9	,155	1,036	97,105
10	,126	,843	97,948
11	,095	,631	98,579
12	,077	,513	99,092
13	,064	,429	99,522
14	,051	,339	99,861
15	,021	,139	100,000

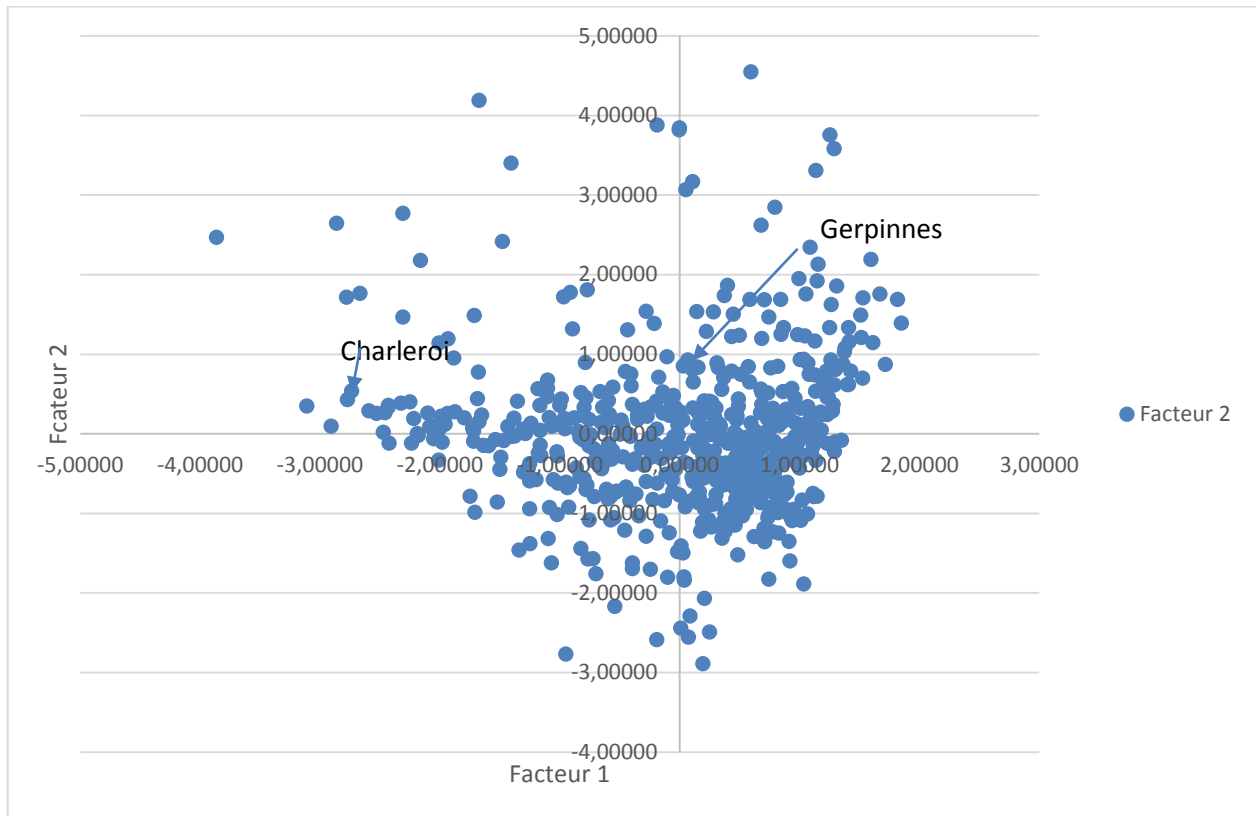
La première dimension reflète les conditions de vie des individus avec des variables liées à l'état du logement, à leur état de santé et à leur statut sur le marché du travail tandis que la deuxième reflètent leur situation socio-économique avec le niveau d'instruction, le coût de l'immobilier et la part des connections internet.

Tableau 10. Matrice des composantes

	Composante	
	1	2
Proportion de personnes en mauvaise santé	-,806	-,370
Revenu médian 1994	,429	,731
Revenu médian 2008	,757	,463
Taux de chômage 1991	-,831	-,246
Taux de chômage 2001	-,949	-,177
Part des bénéficiaires du revenu d'intégration sociale dans la population en âge d'activité	-,842	-,044
Taux d'emploi 1991	,863	,137
Taux d'emploi 2001	,939	,096
Part des logements disposant d'une connection internet	,610	,718
Part de travailleurs saisonniers ou à temps partiel	-,866	,026
% de diplômés disposant d'un diplôme de l'enseignement supérieur (2001)	-,013	,837
Part des personnes considérant leur maison en bon état	,855	,129
Part des logements disposant du double vitrage	,635	,248
Prix moyen par transaction immobilière 1985	,150	,787
Prix moyen par transaction immobilière 2008	,018	,837

A titre illustratif, on a identifié deux communes (Charleroi et Gerpinnes) sur le plan factoriel. Celles-ci bien que voisines géographiquement sont très éloignées socioéconomiquement parlant. Ce serait donc clairement une erreur de ne pas en tenir compte dans le calcul de la dimension généralisée.

Figure 8. Projection des communes sur les deux premiers axes factoriels



Le nouvel estimateur se fonde sur l'hypothèse que des populations proches au sens géographique et socio-économique ont des comportements démographiques semblables et la construction de cet estimateur permet d'exploiter de manière optimale les données disponibles afin de conserver les spécificités locales, pour autant qu'elles soient représentatives. Dans le cadre de cet exercice de projection, c'est la conjonction de deux éléments qui entraîne une perte de signification statistique des probabilités calculées :

- d'une part, des événements peuvent être qualifiés de « rares » pour certains âges, car leur probabilité d'occurrence est proche de 0 ;
- d'autre part, dans certaines communes, les effectifs soumis au risque sont également très faibles. Dès lors, le risque de sous (sur)-évaluation du phénomène est important.

L'estimateur que nous utilisons par la suite est un estimateur à noyaux⁹ défini comme étant le rapport des potentiels du numérateur et du dénominateur : dans le calcul des fréquences, on remplace les mesures de comptages (numérateurs et dénominateurs) par leurs potentiels respectifs. En d'autres termes, on procède à un lissage spatial. Les lignes qui suivent décrivent la méthode.

Soit $\delta_{i,j}$ la distance (généralisée) entre les communes « i » et « j ».

Pour une transition donnée, on note N_j le numérateur observé dans la commune « j » et D_j le dénominateur observé dans la commune « j ». Les potentiels sont deux fonctions définies par

⁹ Voir par exemple Silverman. B.W. Density Estimation. London: Chapman and Hall, 1986

$N(i, \sigma) = \sum_j N_j e^{-\frac{\delta_{i,j}}{\sqrt{2}\sigma^2}}$ et $D(i, \sigma) = \sum_j D_j e^{-\frac{\delta_{i,j}}{\sqrt{2}\sigma^2}}$. Le rapport de ces deux fonctions est l'estimateur retenu et il est calculé en chaque point.

L'estimateur fréquentiel $\hat{\theta}_i = \frac{N_i}{D_i}$ est ainsi remplacé par $\hat{\theta}_i = \frac{N(i, \sigma)}{D(i, \sigma)}$, expression qui est une fonction de $\delta_{i,j}$ de σ . Elle contient l'information de tous les voisinages de la commune « i » et prend une infinité de valeurs dont la valeur (unique) minimale à retenir sera atteinte en accroissant σ jusqu'à un certain seuil. Ce dernier est déterminé par un critère sur le potentiel théorique du dénominateur : il doit être suffisamment grand pour pouvoir effectuer un test de chi-carré valide, à savoir 5.

Remarque :

*Dans l'approche basée sur les fréquences, pour une transition donnée, on calcule la probabilité de transition au niveau national, et on en déduit les numérateurs communaux attendus, sous l'hypothèse que la transition communale soit la même que la transition nationale (hypothèse d'indépendance) : si N_c = numérateur observé de la commune « C », et D_c = dénominateur observé de la commune « C », l'estimateur élémentaire de la probabilité de transition est $\theta_c = \frac{N_c}{D_c}$ et on calcule $A_c = \frac{\sum_c N_c}{\sum_c D_c} * D_c$, le numérateur attendu sous l'hypothèse d'indépendance. Pour tester cette hypothèse, on procède classiquement à un test de chi-carré et on compare $\left(\frac{N_c - A_c}{\sqrt{A_c}}\right)^2$ à une distribution de chi carré à un degré de liberté afin d'accepter ou de rejeter l'hypothèse. Pour pouvoir tester valablement, il faut un minimum de 5 observations attendues par cellule¹⁰.*

Le potentiel théorique du dénominateur est alors $D^*(i, \sigma) = N(i, \sigma) * \frac{D_{nat}}{N_{nat}}$, et le chi-carré est $\chi^2(i, \sigma) = \frac{(D(i, \sigma) - D^*(i, \sigma))^2}{D^*(i, \sigma)}$.

Dans cette méthode d'estimation, on accepte donc de perdre de l'information locale, mais on gagne en robustesse et cette approche permettra de « lisser » les variabilités temporelle et spatiale. Les numérateur et dénominateur résultants sont dénommés « numérateur gaussien » et « dénominateur gaussien » puisque la fonction qui intervient dans la définition des potentiels est une fonction de densité gaussienne et le nouvel estimateur sera dénommé « estimateur gaussien ».

En reprenant l'exemple précédent (transitions pour les hommes âgés de 60 à 64 ans en début de période de la situation « autres ménages de taille 2 » (m_5) vers celle de « couple marié avec un seul enfant » (m_3)), avec l'approche élémentaire, on calcule :

Tableau 11 : Comptage des transitions m_5 vers m_3 des hommes de 60-64 ans à Gerpinnes et au niveau national.

	Gerpinnes			Niveau national		
	2000-2005	2005-2010	2010-2015	2000-2005	2005-2010	2010-2015
Numérateurs	1	0	0	43	57	151
Dénominateur	20	20	42	15233	16522	25274
	5%	0%	0%	0,28%	0,34%	0,60%

¹⁰ Voir par exemple Dagnelie P. *Théorie et méthodes statistiques*, vol. 2, 1975 p. 63

Sous l'hypothèse que la probabilité de transition de Gerpinnes soit celle observée au niveau national, on devrait observer les numérateurs suivants :

Tableau 12 : Numérateurs attendus pour les transitions m_5 vers m_3 (hommes de 60-64 ans à Gerpinnes).

	Gerpinnes		
	2000-2005	2005-2010	2010-2015
Numérateurs attendus	$20 \times 0,28\% = 0,056$	$20 \times 0,34\% = 0,069$	$42 \times 0,60\% = 0,0251$

Or il est impossible, avec ces données, de procéder au test, puisqu'il faut atteindre au minimum la valeur de 5 et on est loin du compte ! Par exemple pour la première période, le numérateur attendu devrait être au moins égal à $5 / 0,28\% = 1773$.

L'estimation par lissage spatial permet quant à elle d'obtenir un estimateur satisfaisant. A partir des données précédentes, le tableau dessous présente les étapes successives du calcul effectué pour atteindre cette valeur :

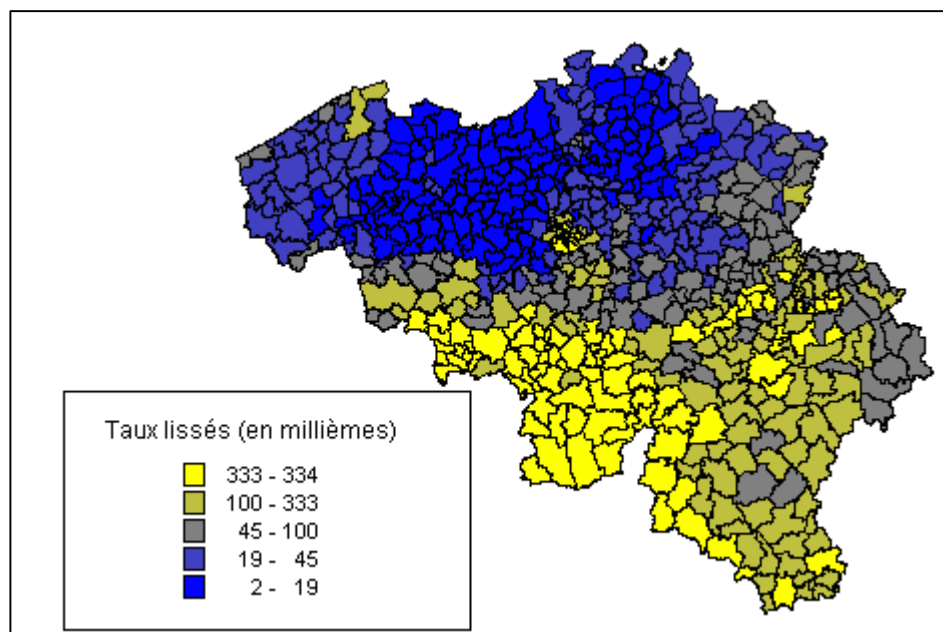
Tableau 13 : Etapes de l'estimation gaussienne des probabilités de transition m_5 vers m_3 (hommes de 60-64 ans à Gerpinnes entre le 1/1/2000 et le 1/1/2005).

	Dénominateur estimé	Numérateur estimé	Probabilité estimée	Communes intervenant dans le processus	Distance généralisée
0	20	1,000	5,0%	Gerpinnes	0
1	20,006	1,000	5,0%	Ham-sur-Heure Nalinnes	3,442
2	24,232	1,000	4,1%	Montigny-le-Tilleul	4,114
3	57,867	1,025	1,8%	Profondeville	4,916
4	238,676	1,407	0,6%	Gembloux	5,494
5	782,742	2,844	0,4%	Villers-la-Ville	5,594
6	1741,052	5,508	0,3%	Genappe	5,786
7	2959,320	8,930	0,3%	Les Bons Villers	6,169
8	4252,501	12,562	0,3%	Floreffe	6,239
	...	...	...	...	...

La valeur 1773 est obtenue après la 6^{ème} étape et il faut donc prendre les observations de 6 communes (autres que Gerpinnes) pour atteindre le seuil recherché.

Ce nouvel estimateur présente à l'évidence une forte structure spatiale ainsi que le montre la carte ci-dessous et les effets aléatoires, résultant de la rareté des observations et de la taille réduite des populations en jeu, ont été éliminés :

Figure 9. Probabilités (résultant du lissage gaussien) de transition des hommes âgés de 60 à 64 ans le 1/1/2000 d'une situation de ménage de type « autre ménage de taille 3 » à celle de couple marié avec un seul enfant.



Les trois estimations précédentes sont remplacées par leurs valeurs « lissées » et ce sont ces dernières qui seront projetées selon la méthodologie décrite dans les lignes qui suivent.

Tableau 14 : Comparaison des estimations élémentaires et gaussiennes des probabilités de transition m_5 vers m_3 (hommes de 60-64 ans à Gerpinnes).

	Estimateur fréquentiel			Estimateur gaussien		
	2000-2005	2005-2010	2010-2015	2000-2005	2005-2010	2010-2015
N	1	0	0	8,93	7,47	8,92
D	20	20	42	2959,32	1884,32	1354,91
θ	5,0%	0%	0 %	0,30%	0,40%	0,66%

Remarque :

*Dans le calcul de l'estimation gaussienne, si au lieu de considérer l'ensemble des communes du pays, on limite l'espace de référence à la seule région wallonne, les estimateurs résultants sont légèrement modifiés puisque la matrice des distances généralisées qui intervient dans le processus de calcul est de dimension 262*262 au lieu de 589*589. Les différences entre les deux projections, tant absolues que relatives, sont faibles (moins de ½ % pour Namur et moins de 3 % dans le cas de Beauvechain, commune bordant la frontière linguistique où l'effet du choix de la distance est plus important). Toutefois, on retiendra les 589 communes. D'une part, il s'agit d'exploiter de manière optimale toutes les informations disponibles et, d'autre part, cela permet d'inscrire le projet dans la continuité de l'exercice précédent.*

Tableau 15: Comparaison des projections (non calibrées) de population à Beauvechain et à Namur.

	2000	2005	2010	2015	2020	2025	2030	2035
Beauvechain								
(réf = Wallonie)	6199	6454	6714	7023	7226	7370	7463	7515
(réf = Pays)	6199	6454	6714	7023	7168	7259	7304	7317
Namur								
(réf = Wallonie)	105419	106767	108950	110646	111192	111914	112625	113066
(réf = Pays)	105419	106767	108950	110646	111185	111901	112603	113037

3.3. Projection des transitions entre les états

Le lissage gaussien terminé, on dispose d'estimateurs robustes des probabilités de transition pour chacune des trois périodes d'observations. Ce sont ces dernières qui doivent être projetées à l'horizon retenu. Dans cet exercice, cet horizon est choisi, la date du 1/1/2035, ce qui correspond à 4 bonds quinquennaux.

Le principe général de la méthode de projection consiste à prolonger les tendances observées au cours des trois périodes précédentes. L'hypothèse sous-jacente à cette démarche est que les tendances observées entre 2000 et 2015, en termes de transition d'un état vers un autre, se poursuivent. Cette hypothèse rejoint le scénario tendanciel identifié le plus souvent comme scénario central notamment dans les projections du Bureau fédéral du Plan. Il se justifie d'autant plus aisément qu'il est demandé qu'en fin de projection, les résultats soient calibrés sur ceux du Bureau Fédéral du Plan, opération qui serait quasiment impossible si les hypothèses de projection étaient différentes.

Le principe général du modèle de projection est que si une tendance à la hausse ou à la baisse se dessine, on prolonge celle-ci et dans le cas contraire, on opte pour la moyenne des trois observations.

La transcription mathématique de ce principe a cependant mis en lumière certaines particularités : en effet, dans le cas de certaines transitions liées à l'émergence ou à l'augmentation rapide de certains comportements, on risque, en maintenant la tendance, d'aboutir à des résultats, a priori, peu plausibles. Ainsi une régression linéaire pourrait produire des probabilités projetées supérieures à l'unité ou négatives. Il est donc impératif de borner les probabilités projetées et la régression logistique (croissante ou décroissante) est une solution élégante à cet égard.

Trois modèles sont retenus en définitive :

Si θ_1 , θ_2 , θ_3 désignent les probabilités estimées et lissées au cours des 3 périodes d'observation

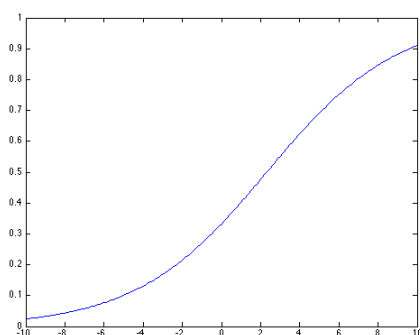
- Modèle 1 : si la probabilité gaussienne moyenne est nulle, ou indéterminée, ou si la probabilité moyenne calculée au niveau national est inférieure à 0,5 % ou si la population gaussienne soumise au risque est inférieure à 30 unités, alors dans ces cas, le modèle mis en œuvre est une constante pondérée (à poids croissants dans le temps) égale à $\sum_{i=1:3} w_i \theta_i$ où $\sum_{i=1:3} w_i = 1$. Afin de donner plus de poids aux valeurs récentes, on retiendra pour poids le vecteur (1/6, 1/3, 1/2).
- Modèle 2 : si les probabilités gaussiennes successives sont croissantes et présentent une concavité positive ($\theta_3 > \theta_1$ et $|\theta_2 - \frac{\theta_1 + \theta_3}{2}| \leq |\theta_3 - \theta_1|$), alors le modèle à appliquer est une régression logistique pondérée croissante dont les asymptotes sont 0 et A. $A = \min\left(1, \frac{\theta_1 + \theta_2 + \theta_3}{3} - (\theta_1 + \theta_2 - 2\theta_3)\right)$. Si $\theta_2 \geq A$, alors le modèle sera le modèle 1.
- Modèle 3 : si les probabilités successives sont décroissantes et présentent une concavité positive ($\theta_3 < \theta_1$ et $|\theta_2 - \frac{\theta_1 + \theta_3}{2}| \leq |\theta_3 - \theta_1|$) alors le modèle sera celui d'une régression logistique pondérée décroissante dont les asymptotes sont 1 et A. $A = \max\left(1, \frac{\theta_1 + \theta_2 + \theta_3}{3} - (\theta_1 + \theta_2 - 2\theta_3)\right)$. Si $\theta_2 \leq A$, alors le modèle sera le modèle 1.

La régression logistique peut se ramener à une régression linéaire de la forme en effectuant un changement de variable obtenu en posant $\ln\left(\frac{\theta_t}{1-\theta_t}\right) = \alpha + \beta t \Leftrightarrow \theta_t = \frac{1}{1+\gamma e^{\delta t}}$ avec $\delta = -\beta$ et $\gamma = e^{-\alpha}$. Si $\delta < 0$, la fonction est croissante avec des asymptotes horizontales en 1 et en 0. Inversement, si $\delta > 0$, alors, la fonction est décroissante avec des asymptotes horizontales en 0 et 1.

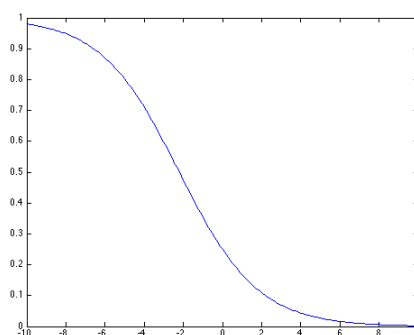
On peut fixer les asymptotes gauche et droite, λ_G et λ_D . Par transformation linéaire en posant $\phi_t = \frac{\theta_t - \lambda_D}{\lambda_R - \lambda_D}$ et en écrivant le modèle $\ln\left(\frac{\phi_t}{1-\phi_t}\right) = \alpha + \beta t$, on se ramène au cas précédent.

Exemples de fonctions logistiques

$$y = \frac{1}{1 + 2\exp(-0,3x)}$$



$$y = \frac{1}{1 + 3\exp(-0,5x)}$$



3.4. Cas particulier des transitions de mortalité

Pour ce qui est de la mortalité, dans le cas des ménages privés, les probabilités de transition par type de ménage sont projetées sur la base d'une famille de fonctions logistiques pondérées de pente commune mais dont l'intercept dépend uniquement du groupe d'âges considéré. Cela signifie que la pente est la même pour tous les âges et sexe. Elle est estimée en regroupant les sexes et les classes d'âges. Si une des probabilités est nulle, alors le(s) groupe(s) d'âges concernés est (sont) exclu(s) du calcul et pour ce(s) groupe(s) d'âges particulier, la projection sera une constante pondérée. Les probabilités projetées commencent donc à un niveau dépendant de l'âge et évoluent en synchronie avec les autres âges et le sexe opposé. L'hypothèse sous-jacente est que la mortalité d'un groupe d'âges et d'un sexe donné n'est pas indépendante de la mortalité de l'autre sexe et des autres groupes d'âges. Pour le dernier groupe d'âges, qui est un groupe d'âges ouvert, le modèle choisi est modèle 1 ou 2 ou 3 en fonction des critères définis précédemment.

En ce qui concerne les ménages collectifs, les probabilités de transition par sexe et groupe d'âges sont projetées à l'aide d'un modèle de régression logistique pondérée croissante analogue au modèle 2. Si le critère n'est pas satisfait alors on retient le modèle constant.

3.5. Normalisation des probabilités de transition Projetées

Au terme du processus de projection, les probabilités projetées, par sexe, groupes d'âges et situation de ménage, sont normalisées : pour un état donné, la somme des probabilités de transition pourrait dépasser, ou être inférieure à l'unité vu que les projections sont réalisées sans

contrainte. Or, il ne peut y avoir ni génération spontanée, ni « évaporation » des individus. Chaque probabilité de transition, autre que celle vers le décès est donc affectée d'un facteur correctif tel que la somme de toutes les probabilités de transition soit bien unitaire. Le décès ne faisant pas l'objet d'une décision personnelle, contrairement aux autres transitions, il n'est pas affecté par cette correction.

Ce sont en définitive les valeurs normalisées qui interviendront dans l'algorithme de la projection.

Dans le cas de Gerpinnes, le tableau ci-dessous reprend les probabilités de transition de la première période projetée des hommes âgés de 30 à 34 ans qui sont initialement membres d'un couple marié avec exactement 1 enfant (ménage de type m₃).

Tableau 16 : normalisation des probabilités de transition projetées.

Situation finale	Probabilité projetée non normalisée	Probabilité projetée normalisée
m1	3,32%	3,40%
m2	0,47%	0,48%
m3	35,24%	36,06%
m4	20,81%	21,30%
m5	2,58%	2,64%
m6	1,60%	1,64%
m7	7,10%	7,26%
m8	3,23%	3,30%
m9	0,02%	0,02%
m10	0,46%	0,46%
m11	20,00%	20,47%
m12	2,90%	2,97%
Somme	97,73%	100%

On constate bien que les probabilités ont dû être corrigées afin que leur somme soit unitaire.

3.6. Les entrées dans le système

Les probabilités de transition étant estimées, projetées et normalisées, pour pouvoir calculer les populations et les ménages en fin de période, il y a lieu de préciser les individus qui entrent dans le système. Ces entrées se font soit par naissance (pour le premier groupe d'âges), soit par immigration en provenance d'une autre commune belge ou en provenance de l'étranger (pour tous les groupes d'âges).

3.6.1. La première classe d'âges

L'effectif des 0-4 ans en fin de période par sexe et situation de ménage est calculé par addition des nouveau-nés survivants dans la population concernée et des nouveau-nés immigrés (internes et externes) entre leur naissance et la fin de la période. La projection de cet effectif se fait en estimant un pseudo-taux de fécondité spécifique à chaque type de ménage et chaque groupe d'âges des femmes âgées de 20-24 à 45-49 ans en fin de période. Les taux observés sont estimés par lissage spatial et ensuite, ils sont projetés, au moyen d'une régression logistique pondérée croissante ou décroissante (ou d'une constante). Les taux éventuellement indéterminés sont fixés à 0.

Si $\phi_i, i = 1..3$ désignent les pseudo-taux de fécondité (sexes regroupés) observés et lissés, par type de ménage, on calcule la valeur numérique de l'asymptote : $A = \left(\frac{\phi_1 + \phi_2 + \phi_3}{3} \right) - (\phi_1 + \phi_2 - 2\phi_3)$.

- Si $A \leq 0$ alors le modèle de projection est une régression logistique pondérée croissante de 0 vers 1.
- Dans le cas contraire,
 - si $\phi_1 > A$ et $\phi_2 > A$ et $\phi_3 > A$ alors la méthode sera une régression logistique pondérée décroissante de 1 vers A,
 - si $\phi_1 < A$ et $\phi_2 < A$ et $\phi_3 < A$ alors la méthode sera une régression logistique pondérée croissante de 0 vers A.
- dans toutes les autres situations, on retiendra comme modèle une moyenne pondérée des observations (éventuellement lissées).

Ces taux extrapolés sont multipliés par le nombre de femmes âgées de 20 à 49 ans en fin de période, afin de déterminer le nombre total de nouveau-nés survivants et de nouveau-nés immigrés. Le calcul se fait en distinguant les types de ménage en fin de période. De plus, on regroupe toutes les femmes des couples mariés avec 3 enfants ou plus et des autres ménages de taille supérieure ou égale à 4, indépendamment de leur statut de chef de ménage ou non (ménages m_7 et m_8).

Le coefficient de masculinité calculé au niveau national (sur base des observations 2000-2015) est appliqué aux taux projetés, afin d'obtenir un taux de fécondité selon le sexe et ce pour chaque commune.

3.6.2. L'introduction des immigrants

A la fin de chaque pas de la projection, les immigrants internes sont introduits en chiffres absolus, selon le sexe, groupe d'âges et type de ménage auxquels ils appartiennent. Les chiffres retenus sont les moyennes pondérées des trois observations, un poids plus important étant accordé aux observations plus récentes. Le même principe est appliqué aux immigrants externes.

3.7. Algorithme complet

3.7.1. Notations

- Communes : $c_j, j = 1..262$
- Sexes : $s_j, j = 1..2$
- Situations de ménages : $m, j = 1..9$
- Groupes d'âges en début de période : $a_j, j = 1..18$

Pour rappel, un état à un instant « t » est caractérisé par le croisement des modalités d'une commune, d'un sexe, d'une classe d'âges, et d'un type de ménage et est noté e_{c_i, s_j, a_k, m_l} .

On note encore

- $P_t(c_i, s_j, a_k, h_l)$: la population dans l'état en début de période,
- $II_{t \rightarrow t+5}(c_i, s_j, a_k, m_l)$: le nombre d'immigrants internes de sexe s_j , âgés de a_k ans (au temps « t ») dans les ménages de type m_l , entrés dans la commune c_i pendant la période $[t: t + 5]$,
- $IE_{t \rightarrow t+5}(c_i, s_j, a_k, m_l)$: le nombre d'immigrants externes de sexe s_j , âgés de a_k ans (au temps « t ») dans les ménages de type m_l , entrés dans la commune c_i pendant la période $[t: t + 5]$,

- $El_{t \rightarrow t+5}(c_i, s_j, a_k, m_l)$: le nombre d'émigrants internes de sexe s_j , âgés de a_k ans (au temps « t ») en provenance des ménages de type m_l , sortis de la commune c_i pendant la période $[t: t + 5]$,
- $EE_{t \rightarrow t+5}(c_i, s_j, a_k, m_l)$: le nombre d'émigrants externes de sexe s_j , âgés de a_k ans (au temps « t ») en provenance des ménages de type m_l , sortis de la commune c_i pendant la période $[t: t + 5]$,
- $\Theta_{t \rightarrow t+5}(c_i, a_k, s_l, m_i, m_j)$: la probabilité de transition du type de ménage m_l vers le type de ménage m_j , des individus de la commune c_i , de sexe s_j , âgés de a_k ans (au temps « t ») au cours de la période $[t: t + 5]$.

L'algorithme mis en œuvre requiert en entrée les données suivantes :

- la population initiale décomposée selon le sexe, la classe d'âges et la situation de ménage auquel elle appartient en début période,
- les naissances et immigrations du premier groupe d'âges, selon le sexe et type de ménage auquel elle appartient en fin de période,
- les probabilités de transitions entre les différents états,
- les effectifs d'immigrants distingués selon le sexe, la classe d'âges et la situation de ménage finale.

Il produit pour chaque point de la projection

- les chiffres détaillés de population dont les sommes marginales permettent d'obtenir facilement les effectifs selon les groupes d'âges, les sexes, les types de ménages, les communes ainsi que toute combinaison de ces facteurs.
- le nombre des ménages privés qui s'obtient en considérant les ménages de type « isolé », la moitié des individus vivant couples mariés sans enfants et autres ménages de taille 2, le tiers des individus vivant en couples mariés avec un seul enfant et autres ménages de taille 3, le quart des couples mariés avec 2 enfants et les chefs de ménage de couples mariés avec 3 enfants ou plus et chefs de ménages des autres ménages de taille supérieure ou égale à 4.

3.7.2. Etape 1

Les populations de la première classe d'âges sont calculées. Pour ce faire, la population féminine soumise au risque de la natalité est calculée par groupes d'âges et types de ménages.

On calcule :

$$\begin{aligned}
 \forall i = 1..262, \forall k = 4..9 : P_{t+5}(c_i, F, a_{k+1}, m_i) \\
 = \sum_{j=1}^9 P_t(c_i, F, a_{k+1}, m_i) * \Theta_{t \rightarrow t+5}(m_i, m_j, a_k, F) + EE_{t \rightarrow t+5}(c_i, F, a_k, m_l) \\
 + IE_{t \rightarrow t+5}(c_i, F, a_k, m_l)
 \end{aligned}$$

Pour assurer la cohérence de ces chiffres avec ceux calculés par le Bureau Fédéral du Plan, on procède à un calibrage : si $BFP_{t+5}(W, a_{k+1}, F)$ désigne la population féminine wallonne appartenant au groupe d'âges « k+1 » et projetée au temps t+5 par le Bureau Fédéral du Plan, alors, on impose : $\forall k = 4..9 : \sum_{i=1}^{262} P_{t+5}(c_i, F, a_{k+1}, m_i) = BFP_{t+5}(W, a_{k+1}, F)$

Le calibrage consiste à calculer :

$$\forall i = 1..262, \forall k = 4..9: P_{t+5}(c_i, F, a_{k+1}, m_i) \leftarrow P_{t+5}(c_i, F, a_{k+1}, m_i) * \frac{\sum_{i=1}^{262} P_{t+5}(c_i, F, a_{k+1}, m_i)}{BFP_{t+5}(W, a_{k+1}, F)}.$$

On en déduit le nombre d'enfants de 0-4 ans (sexes confondus) dans les ménages de type m_i en fin de la période $[t: t + 5]$: $N(m_i, [t: t + 5]) = \sum_{k=4}^9 P_{t+5}(c_i, F, a_{k+1}, m_i) * \Phi_{t \rightarrow t+5}(m_i, a_k)$, où $\Phi_{t \rightarrow t+5}(m_i, a_k)$ est pseudo taux de fécondité des femmes appartenant aux ménages de type « i » d et à la classe d'âges « k +1 ». En appliquant à ces chiffres le coefficient de masculinité national α , on en déduit le nombre de garçons âgés de 0-4 ans = $P_{t+5}(c_q, s_1, a_0, m_j) = N(m_i, [t: t + 5]) * \alpha$ et le nombre de filles âgées de 0-4 ans = $P_{t+5}(c_q, s_2, a_0, m_j) = N(m_i, [t: t + 5]) * (1 - \alpha)$.

3.7.3. Etape 2

On projette les populations des classes d'âges supérieures par sexe et situation de ménage. On introduit ensuite les immigrants internés et externes.

$$\begin{aligned} \forall q = 1..262, \forall k = 1..19, \forall l = 1..2, \forall i = 1..9 : P_{t+5}(c_q, s_l, a_{k+1}, m_i) \\ = \sum_{j=1}^9 P_t(c_q, s_l, a_k, m_j) * \Theta_{t \rightarrow t+5}(c_q, a_k, s_l, m_i, m_j) + \Pi_{t \rightarrow t+5}(c_q, s_l, a_k, m_i) \\ + IE_{t \rightarrow t+5}(c_q, s_l, a_k, m_i) \end{aligned}$$

Au terme de cette étape, un facteur correctif est introduit afin s'assurer l'égalité des effectifs masculins et féminins appartenant au type de ménage « couple sans enfants ». On redistribue la population totale vivant dans ce type de ménage selon les sexes et les classes d'âges et ce pour chaque commune :

$$\begin{aligned} \forall q = 1..262 \sum_{k=4}^{19} P_{t+5}(c_q, s_1, a_{k+1}, m_2) = \sum_{k=4}^{19} P_{t+5}(c_q, s_2, a_{k+1}, m_2) \\ \text{et } \forall q = 1..262 \sum_{i=1}^9 \sum_{k=4}^{19} P_{t+5}(c_q, s_1, a_{k+1}, m_i) = \sum_{i=1}^9 \sum_{k=4}^{19} P_{t+5}(c_q, s_2, a_{k+1}, m_i) \end{aligned}$$

$$\begin{aligned} \forall q = 1..262, \forall k = 1..19, \forall l = 1..2 : P_{t+5}(c_q, s_l, a_{k+1}, m_2) \\ \leftarrow P_{t+5}(c_q, s_l, a_{k+1}, m_2) \\ * \frac{\sum_{k=4}^{19} P_{t+5}(c_q, s_1, a_{k+1}, m_2) + \sum_{k=4}^{19} P_{t+5}(c_q, s_2, a_{k+1}, m_2)}{2} \end{aligned}$$

$$\begin{aligned} \forall q = 1..262, \forall k = 1..19, \forall l = 1..2, \forall i <> 2 : P_{t+5}(c_q, s_l, a_{k+1}, m_i) \\ \leftarrow P_{t+5}(c_q, s_l, a_{k+1}, m_i) \\ * \frac{\sum_{i=1}^9 \sum_{k=4}^{19} P_{t+5}(c_q, s_1, a_{k+1}, m_i) + \sum_{i=1}^9 \sum_{k=4}^{19} P_{t+5}(c_q, s_2, a_{k+1}, m_i)}{2} \end{aligned}$$

3.7.4. Etape 3

Afin d'assurer la cohérence des projections avec celles produites par le Bureau Fédéral du Plan pour la région wallonne, un facteur correctif est appliqué à chaque commune, sexe, type de ménage et groupes d'âges, de telle sorte que les totaux des âges coïncident avec les chiffres du Bureau Fédéral du Plan. On impose donc : $\forall k = 1..19 : \sum_{i=1}^{262} P_{t+5}(c_q, s_l, a_{k+1}, m_i) = \text{BFP}_{t+5}(W, a_{k+1}, s_l)$. Dès lors $\forall q = 1..262, \forall k = 1..19, \forall l = 1..2, \forall i = 1..9$

$$P_{t+5}(c_q, s_l, a_{k+1}, m_i) \leftarrow P_{t+5}(c_q, s_l, a_{k+1}, m_i) * \frac{\sum_{i=1}^{262} P_{t+5}(c_q, s_l, a_{k+1}, m_i)}{\text{BFP}_{t+5}(W, a_{k+1}, s_l)}$$

Cet algorithme est itéré jusqu'à l'horizon de la projection, la population finale au sortir d'un pas d'itération devenant la population initiale pour le pas suivant.

Il est en cours de développement en Matlab.

4. BIBLIOGRAPHIE INDICATIVE

- Bahri A., Eggerickx T., Sanderson J.-P., 2010, « Transformation des ménages et ségrégation sociale de l'espace. Une analyse de la précarisation des parcs résidentiels en Belgique (1991-2005) », *Dynamiques de pauvretés et vulnérabilités en démographie et en sciences sociales*, Chaire Quetelet 2007, B. Masquelier, T. Eggerickx, ed(s), Presses Universitaires de Louvain, Louvain-la-Neuve, pp. 187-221.
- Banens M., 1994, « La prévision de sous-populations par la méthode des tendances partielles », *Population*, n°4-5, pp. 1130-1137.
- Bergouignan C., 2008, « Projeter les populations soumises à une forte mobilité résidentielle I », *Cahiers de démographie locale*, pp. 19-53.
- Bergouignan C., 2008, « Projeter les populations soumises à une forte mobilité résidentielle II », *Cahiers de démographie locale*, pp. 55-97.
- Costa R., Eggerickx T., Rizzi E., Sanderson J.-P., 2010, « Analyse spatiale et temporelle de la fécondité en Belgique : une approche communale », *La fécondité : représentation, causalité, prospective*, CUDEP, (à paraître).
- Costa R., Eggerickx T., Sanderson J.-P., 2011, « Les territoires de la fécondité en Belgique au 20^e siècle Une approche longitudinale et communale », *Espaces, Populations, Sociétés*, 3, 2011.
- Courgeau D., Lelièvre E., 2003, « Les motifs individuels et sociaux des migrations », *Démographie : analyse et synthèse. IV Les déterminants de la Migration*, sous la direction de G. Cazelli, J. Vallin et G. Wunsch, Editions de l'Institut National d'Etudes Démographiques, Paris, pp. 147-169.
- Dagnelie P. Théorie et méthodes statistiques, vol. 2, 1975
- Deboosere P., Fizman P., 2009, « De la persistance des inégalités socio-spatiales en santé : le cas belge », *Espace, Populations et Sociétés*, Différences et inégalités socio-démographiques : l'approche par le local, n°, pp. 149-158.
- Eggerickx T., Hermia J.-P., Surkijn J., Willart D., 2010, (avec la collaboration de L. Dal, M. Poulain et J.-P. Sanderson) *Les migrations internes en Belgique*, Monographie 2 de l'Enquête socioéconomique de 2001, DGSIE, Bruxelles, 200 p., (à paraître).
- Eggerickx T., Oris M., Sanderson J.-P., Vilpert S., 2010, *L'évolution de la population, l'âge et le sexe*, Monographie 1 de l'Enquête socio-économique de 2001, DGSIE, Bruxelles, 230 p. (à paraître).
- Eggerickx T., Sanderson J.-P., 2010, « Les inégalités spatiales de mortalité en Belgique : 1980-2005 », *Démographie et santé*, CUDEP, Bordeaux, pp. 145-161.
- Eggerickx T., Debuissou M., Dal L., 2008, « Les tendances récentes des migrations entre les arrondissements » Perspectives de population 2007-2060, *Planning Paper*, Bureau du Plan-DGSIE, Bruxelles, 105, pp. 35-38.

- Eggerickx T., Poulain M., 1996, « De la variabilité des paramètres démographiques pour les petites populations », *Espace, Populations et Sociétés*, Mélanges D. Noin, n°1, pp. 93-102.
- Eggerickx T., Sanderson J.P., Bahri A. et Hermia J.P., 2005, *Les espaces « ruraux » en Wallonie (Belgique). La diversité des trajectoires démographiques et le rôle des migrations dans la structuration socio-démographique des espaces*, communication présentée au colloque « FAIRE CAMPAGNE » Pratiques et projets des espaces ruraux aujourd'hui, Rennes, 17 et 18 mars 2005, 20 p.
- Franzén M. et Karlsson T., 2010, *Using national data to obtain small area estimators for population projections on sub-national level*, Communication présentée à la Conference of European statisticians, 28-30 avril 2010 à Lisbonne, 10 p.
- Grasland C., 1991, « Potentiel de population, interaction spatiale et frontière: des deux Allemagne à l'unification », *L'Espace Géographique*, 3, pp. 243-254
- Grasland C., 2001, *Lissage cartographique et animation spatio-temporelle : quelques réflexions méthodologiques*, Travaux de l'Institut de Géographie de Reims, 12 p
- Gusbin D., Toint P., Cornelis E., Poulain M., Eggerickx T., 2007, *Démographie, géographie et mobilité. Perspectives à long terme et politiques pour un développement durable*, Bruxelles, Politique Scientifique Fédérale, 2007, 144 p., Plan d'appui scientifique à une politique de développement durable, PADD II
- Maurois G. et Bélanger A., 2014, "Microsimulation Model Projecting Small Area Populations Using Contextual Variables: An Application to the Montreal Metropolitan Area, 2006-2031", *International Journal of Microsimulation*, 7, pp. 158-193
- Monnier A., 2004, *Démographie contemporaine de l'Europe. Evolutions, tendances, défis*, Collection U, Paris, 415 p.
- Reher D., 1998, « Family ties in Western Europe: persistent contrasts », *Population and Development Review*, vol. 24, n°2, pp. 203-234.
- Roussel L., 1992, « La famille en Europe occidentale : divergences et convergences », *Population*, 1, pp. 133-152.
- Rudant J.B., 2011, « Le modèle de projections localisées Omphale 2010 », *Cahiers de démographie locale 2010*, 3, pp.83-110
- Sanderson J.-P., Eggerickx T., Poulain M., 2009, « Impact des migrations internes sur le vieillissement des populations locales : les communes belges de 1992 à 2002 », *Reproduction et renouvellement des populations*, CUDEP, Bordeaux, pp. 193-212
- Sanderson J.-P., Dal L., Debuissson M., Eggerickx T., Poulain M., 2011, « Perspectives de population et de ménages des communes belges. Confrontation des résultats à la réalité des chiffres », *Cahiers de démographie locale 2010*, 3.
- Sanderson J.-P., Eggerickx T., 2010, Des outils d'aide à la décision : les indicateurs de condition de vie et de mixité socio-démographique des communes et des quartiers en Belgique, *Cahiers de démographie locale 2009*, 2, p. 115-158.
- Silverman B.W., 1986, *Density Estimation*. Londres: Chapman and Hall.

- Sobotka T., 2008, "The diverse faces of the Second Demographic Transition in Europe", *Demographic Research*, 19, 8, pp. 171-224.
- Van Hecke E., 1998, "Actualisering van de stedelijke hiërarchie in België" in *Tijdschrift van het Gemeentekrediet van België*, année 52, n°205, pp. 45-76 ;
- Wattelar C., 2004, "Perspectives démographiques : histoire de la méthode et méthodes actuelles" in Caselli G., Vallin J. et Wunsch G., *Démographie : analyse et synthèse. V. Histoire du peuplement et prévisions*, INED, Paris, pp. 253-276 ;
- Willems P., 2006, *Bevolkingsprojecties 2004-2025 voor de 308 gemeenten van het Vlaamse Gewest*, Bruxelles, SVR – Technisch rapport 2006/2, 52 p.;
- Willems P., 2007, *Projecties van aantallen huishoudens naar huishoudgrootte voor de 308 gemeenten van het Vlaamse Gewest. Twee scenario's voor de periode 2005-2025*, Bruxelles, SVR – Technisch rapport 2007/1, 32 p.;
- Willems P., 2008, "Perspectives de population pour la Région flamande, 2004-2025" in Bureau fédéral du Plan, *Perspectives de population 2007-2060*, Bruxelles, Planning paper 105, pp. 19-22 ;

Rapport de l'exercice précédent

- Sanderson J.P., Dal L., Eggerickx T. et Poulain M., 2010, *Perspectives de population locale en Belgique. Rapport I: Etat de l'art et test de quelques méthodes*, Louvain-la-Neuve, 24 p.
- Sanderson J.P., Dal L., Eggerickx T. et Poulain M., 2010, *Perspectives de population locale en Belgique. Rapport II: Lissage gaussien et typologie des communes*, Louvain-la-Neuve, 23 p.
- Sanderson J.P., Dal L., Eggerickx T. et Poulain M., 2010, *Perspectives de population locale en Belgique. Rapport III : Définition des états*, Louvain-la-Neuve, 15 p.
- Sanderson J.P., Dal L., Eggerickx T. et Poulain M., 2011, *Perspectives de population locale en Belgique. Rapport IV : Lissage gaussien*, Louvain-la-Neuve, 11 p.
- Dal L., Eggerickx T., Poulain M. et Sanderson J.P., 2011, *Perspectives de population locale en Belgique. Rapport V : Méthodologie de la projection*, Louvain-la-Neuve, Décembre 2011, 13 p.

5. PROGRAMME SAS

Luc Dal DEMO – UCL septembre 2015

Le programme fourni en annexe prépare les données extraites du Registre National et effectue les opérations suivantes en vue du traitement ultérieur en Matlab.

1. Lecture des données des stocks aux 4 dates retenues
2. Attribution de la typologie des ménages
3. Couplage individuel pour les 3 périodes
4. Lecture des flux utiles au cours des 3 périodes
5. Elaboration des structures communales, régionales et nationale de la population par sexe, âge et type de ménage sur base des observations quinquennales
6. Elaboration des matrices de transitions communales de la population par sexe, âge et type de ménage sur base des observations quinquennales dans le cadre régional.
7. Exportation des données de structure et des matrices de transition au format csv. Ce sont ces données qui sont importées en Matlab pour produire les projections de population et de ménages.

Remarques

Les flux sont considérés dans leur seule région comme espace de référence (Bruxelles, Wallonie, Flandre); dans ce cas les migrations de/vers une autre région du pays sont considérées comme étant des migrations extérieures.

Parallèlement, lorsque l'espace de référence est la pays dans son intégralité, les migrations extérieures reprennent uniquement les migrations de/vers l'étranger.

```

/*****/
/*Definition des chemins*/
/*Luc Dal DEMO - UCL 2015*/
/*****/
%let stocks=XXXXXX;/*chemin des stocks*/
%let flux=YYYYYY;/*chemin des flux*/
%let donnees=ZZZZZ\datas;/*chemin des donnees
prepees*/
%let txt=UUUUUU\txt;/*chemin des donnees txt -
-> import Matlab*/
options obs=max;
%let debut=2000;
%let fin=%eval(&debut+5);
libname donnees "&donnees";
libname stocks "&stocks";
libname flux "&flux";
libname txt "&txt";
/*****/
/*1. Selection des donnees utiles en t*/
/*****/
%macro etats(t);
data donnees.e&t;
    set stocks.stockmen&t;
    rename    n3lien=lienCM    n3tymen=typomen
nouvelletaille=taillemenage;
    if com='21099' then delete;/*Suppression des
registres d attente*/
    keep codern sexe n3lien n3tymen nouvelletaille
age anaiss reg com;
run;
/*Typologie des menages utile pour la projection*/
data donnees.e&t;
    set donnees.e&t;
    select;
    when(typomen='Col') men='09';/*collectifs*/
    when(typomen in('IH ','IF ') or taillemenage=1)
men='01';/*isoles*/
    when(typomen          in('CSE','HSE'))
men='02';/*couples sans enfants*/
    when(typomen          in('CAE','HAE')          and
taillemenage=3)  men='03';/*couples avec 1
enfant*/
    when(typomen          in('CAE','HAE')          and
taillemenage=4)  men='04';/*couples avec 2
enfants*/
    when(taillemenage eq 2) men='05';/*autres
menages de taille 2*/
    when(taillemenage eq 3) men='06';/*autres
menages de taille 3*/
    when(typomen in('CAE','HAE') and taillemenage
ge 5 and lienCM eq '01') men='07';/*CM couples
avec 3 enfants ou +*/
    when(taillemenage ge 4 and lienCM eq '01')
men='07';/*CM autres menages de taille 4+*/

```

```

    when(typomen in('CAE','HAE') and taillemenage
ge 5 and lienCM ne '01') men='08';/*Autre
membres couples avec 3 enfants ou +*/
    when(taillemenage ge 4 and lienCM ne '01')
men='08';/*Autre membres autres menages de
taille 4+*/
end;
    keep codern lienCM reg age anaiss sexe com
men typomen taillemenage;
run;
/* Correctifs*/
data donnees.e&t;set donnees.e&t;
    if typomen='Aut' and taillemenage=1 and
sexe='1' then men='01';
    if typomen='Aut' and taillemenage=1 and
sexe='2' then men='01';
    if men='07' and age lt 18 and sexe='1' then
men='08';/*prise en compte des CM de moins de
15/18 ans*/
    if men='07' and age lt 15 and sexe='2' then
men='08';
    if men='01' and age lt 18 and sexe='1' then
men='08';/*isoles de moins de 18 ans*/
    if men='01' and age lt 15 and sexe='2' then
men='08';/*isolees de moins de 15 ans*/
run;
proc datasets lib=donnees memtype=data nolist
noprnt;
    modify e&t;
    attrib _all_ label=";
    attrib _all_ format=;
run;
quit;
proc sort data=donnees.e&t;
    by codern;
run;
%mend;
/*Boucle de selection des donnees utiles aux 4
dates*/
%macro boucle1();
    %do t=&debut %to &fin %by 5;
    %etats(&t);
    %end;
%mend;
%boucle1();
/*****/
/*2. Couplage des individus a deux dates
consecutives*/
/*****/
%macro fus(t);
data null;
    %let t1=%eval(&t);
    %let t2=%eval(&t+5);
    %put &t1 &t2;
run;

```

```

data e&t1;set donnees.e&t1;
    drop typomen liencm;
    rename com=com&t1 reg=reg&t1 com=com&t1
men=men&t1;
run;
data e&t2;set donnees.e&t2;
    drop typomen liencm;
    rename com=com&t2 reg=reg&t2 com=com&t2
men=men&t2;
run;
data donnees.couplage&t1&t2;
    merge e&t1(in=a) e&t2(in=b);
    by codern;
    select;
    when(a and b) do;p='PP';end;
    when(a and not b) do;p='PA';end;
    when(not a and b) do;p='AP';end;
    end;
run;
data donnees.couplage&t1&t2;
    set donnees.couplage&t1&t2;
    if p='PP' then do;/* presents en Belgique aux
deux moments */
        com1=com&t1;
        com2=com&t2;
        reg1=reg&t1;
        reg2=reg&t2;
        men1=men&t1;
        men2=men&t2;
        end;
        if p='PA' then do;/* sortis de Belgique */
            com1=com&t1;
            reg1=reg&t1;
            men1=men&t1;
            end;
        if p='AP' then do;/* entres en Belgique */
            com2=com&t2;
            reg2=reg&t2;
            men2=men&t2;
            end;
keep p anaiss codern com1 com2 sexe men1 men2
reg1 reg2;
run;
%mend;
/*Boucle de couplage des donnees */
%macro boucle2();
    %do t=&debut %to %eval(&fin-5) %by 5;
    %fus(&t);
    %end;
%mend;
%boucle2();
/***** */
/*3. Entrees et sorties du systeme*/
/***** */
/*3.1. Extraction des mouvements utiles*/

```

```

/*naissances, deces, immigrations interne
emigrations internes*/
data nai dec imiint emiint;
set flux.flux19912014(keep=codern com comD
amig mmig jmig tmig anaiss rename=(anaiss=anam
com=comc1 comD=comc2) where=(tmig
in('11','21','12','22')));
*age au mouvement;
    age=amig-1-anam;
    select;
    when (tmig='11') output nai;/*naissances*/
    when (tmig='21') output dec;/*deces*/
    when (tmig='12') output imiint;/*immigrations
internes*/
    when (tmig='22') output emiint;/*emigrations
internes*/
    end;
run;
/***** */
/*3.2 Nettoyage des naissances*/
/*on s assure que les naissances ont pour annee de
mouvement l annee de naissance*/
/*Attribution de la region de naissance*/
data donnees.nai;length reg$1;
    set nai(where=(amig=anam));
    rename comc1=commvt;
    select;
    when(substr(comc1,1,1) in('1','3','4','7') or
substr(comc1,1,2) in('23','24')) reg='V';
    when(substr(comc1,1,2)='21') reg='B';
    otherwise reg='W';
    end;
    keep codern comc1 amig reg age;
run;
proc sort data=donnees.nai nodupkey;
    by codern;
run;
/***** */
/*3.3 Nettoyage des deces*/
/*on s assure que chaque deces est enregistre une
seule fois*/
/*Si certain deces sont enregistres plusieurs fois, on
conserve la premiere occurrence*/
/*Attribution de la region de deces*/
data donnees.dec;length reg$1;
    set dec(where=(amig ge anam));
    rename comc1=commvt;
    select;
    when(substr(comc1,1,1) in('1','3','4','7') or
substr(comc1,1,2) in('23','24')) reg='V';
    when(substr(comc1,1,2)='21') reg='B';
    otherwise reg='W';
    end;
    keep codern comc1 amig reg age;
run;

```

```

proc sort data=donnees.dec;
  by codern amig;
run;
data donnees.dec;set donnees.dec;
  by codern amig;
  if first.codern;
run;
/*****/
/*3.4 Nettoyage des immigrations internes*/
/*Attribution de la region d immigration*/
data donnees.imi;length reg$1;
  set imiint(where=(amig ge anam));
  rename comc1=commvt;
  select;
  when(substr(comc1,1,1) in('1','3','4','7') or
substr(comc1,1,2) in('23','24')) reg='V';
  when(substr(comc1,1,2)='21') reg='B';
  otherwise reg='W';
  end;
  keep codern comc1 amig mmig jmig reg
age;/*commune arrivee*/
run;
proc sort data=donnees.imi;
  by codern amig mmig jmig;
run;
/*****/
/*3.5 Nettoyage des emigrations internes*/
/*Attribution de la region d emigration*/
data donnees.emi;length reg$1;
  set emiint(where=(amig ge anam));
  rename comc1=commvt;
  select;
  when(substr(comc1,1,1) in('1','3','4','7') or
substr(comc1,1,2) in('23','24')) reg='V';
  when(substr(comc1,1,2)='21') reg='B';
  otherwise reg='W';
  end;
  keep codern comc1 amig mmig jmig reg
age;/*commune depart*/
run;
proc sort data=donnees.emi;
  by codern amig mmig jmig;
run;
/*****/
/*4. Structures*/
/*****/
data structures;
run;
%macro struct(t);
data tmp;
  set donnees.e&t;
  select;
  when(substr(com,1,1) in('1','3','4','7') or
substr(com,1,2) in('23','24')) reg='V';
  when(substr(com,1,2)='21') reg='B';

```

```

otherwise reg='W';
end;
select;
when(age lt 0) grage='00000';
when(0 le age lt 5) grage='00_04';
when(5 le age lt 10) grage='05_09';
when(10 le age lt 15) grage='10_14';
when(15 le age lt 20) grage='15_19';
when(20 le age lt 25) grage='20_24';
when(25 le age lt 30) grage='25_29';
when(30 le age lt 35) grage='30_34';
when(35 le age lt 40) grage='35_39';
when(40 le age lt 45) grage='40_44';
when(45 le age lt 50) grage='45_49';
when(50 le age lt 55) grage='50_54';
when(55 le age lt 60) grage='55_59';
when(60 le age lt 65) grage='60_64';
when(65 le age lt 70) grage='65_69';
when(70 le age lt 75) grage='70_74';
when(75 le age lt 80) grage='75_79';
when(80 le age lt 85) grage='80-85';
when(85 le age lt 90) grage='85-90';
when(90 le age) grage='90+ ';
end;
run;
/*Communes*/
proc freq data=tmp noprint;tables
com*men*sexe*grage/list sparse
out=com(drop=percent rename=count=eff);
run;
/*Pays*/
proc freq data=tmp noprint;tables
men*sexe*grage/list sparse out=nat(drop=percent
rename=count=eff);
run;
data nat;length com $5;set nat;
  com='00100';
run;
/*Bruxelles*/
proc freq data=tmp(where=(reg='B')) noprint;tables
men*sexe*grage/list sparse out=b(drop=percent
rename=count=eff);
run;
data b;length com $5;set b;
  com='02000';
run;
/*Flandre*/
proc freq data=tmp(where=(reg='V')) noprint;tables
men*sexe*grage/list sparse out=v(drop=percent
rename=count=eff);
run;
data v;length com $5;set v;
  com='01000';
run;

```

```

proc      freq      data=tmp(where=(reg='W'))
noprnt;tables      men*sexe*grage/list      sparse
out=w(drop=percent rename=count=eff);
run;
/*Wallonie*/
data w;length com $5;set w;
      com='03000';
run;
data tmp;length annee $4;
      set com nat b v w;
      annee=&t;
run;
data structures;
      set structures tmp;
run;
proc sort data=structures;
      by com annee;
%mend;
%macro boucle3();
      %do t=&debut %to %eval(&fin) %by 5;
      %struct(&t);
      %end;
%mend;
%boucle3();
data donnees.structures;
      set structures;
      if _N_ >1;
run;
proc datasets lib=donnees memtype=data nolist
noprnt;
      modify structures;
      attrib _all_ label=";
      attrib _all_ format=;
run;
quit;
/*****/
/*5. Transitions au niveau regional*/
/*****/
data transitions;
run;
%macro mouvementsreg(t,r);
%let t1=&t;
%let t2=%eval(&t1+5);
%put &t1 &t2 &r;
data couplage&t1&t2;
      set donnees.couplage&t1&t2;
      age=%eval(&t1)-anaiss-1;/*les      ages      sont
negatifs pour ceux qui ne sont pas encore nes
puisque l age considere est celui du debut de
periode */
      select;
      when(age lt 0) grage='00000';
      when(0 le age lt 5) grage='00_04';
      when(5 le age lt 10) grage='05_09';
      when(10 le age lt 15) grage='10_14';

```

```

      when(15 le age lt 20) grage='15_19';
      when(20 le age lt 25) grage='20_24';
      when(25 le age lt 30) grage='25_29';
      when(30 le age lt 35) grage='30_34';
      when(35 le age lt 40) grage='35_39';
      when(40 le age lt 45) grage='40_44';
      when(45 le age lt 50) grage='45_49';
      when(50 le age lt 55) grage='50_54';
      when(55 le age lt 60) grage='55_59';
      when(60 le age lt 65) grage='60_64';
      when(65 le age lt 70) grage='65_69';
      when(70 le age lt 75) grage='70_74';
      when(75 le age lt 80) grage='75_79';
      when(80 le age lt 85) grage='80-85';
      when(85 le age lt 90) grage='85-90';
      when(90 le age)      grage='90+ ' ;
end;
run;
data pa ap pp;set couplage&t1&t2;
      select;
      when((reg1 eq symget('r') and reg2 eq
symget('r'))) output pp;/*ils restent dans la region*/
      when((reg1 eq symget('r') and reg2 ne
symget('r'))) output pa;/*ils sortent de la region*/
      when((reg1 ne symget('r') and reg2 eq
symget('r'))) output ap;/*ils entrent dans la
region*/
      otherwise;
      end;
run;
/*Cas des PA : ils sortent de la region par deces ou
emi exterieure (etranger ou autre region)*/
data pa;merge pa(in=a) donnees.dec(in=b
where=(put(&t1,4.) le amig lt put(&t2,4.)));
      by codern;
      if (a and not b) then mvt='PAExt ';/*Emi
exterieur*/
      if (a and b) then mvt='PADec ';/*Deces*/
      if a;
      rename com1=com reg1=reg;
      keep codern sexe grage age com1 reg1 men1
mvt;
run;
/*Cas des AP : ils entrent de la region par naissance
ou immi exterieure (etranger ou autre region)*/
data ap;merge ap(in=a) donnees.nai(in=b
where=(put(&t1,4.) le amig lt put(&t2,4.)));
      by codern;
      if (a and not b) then mvt='APExt ';/*Immi
exterieur*/
      if (a and b and com2 eq commvt) then
mvt='APNais';/*Naissance*/
      if (a and b and com2 ne commvt) then
mvt='APInt';/*Immi interne nne*/
      if a;

```

```

        rename com2=com reg2=reg;
        keep codern sexe grage age com2 reg2 men2
mvt;
run;
/*Cas des PP : niveau regional*/
data ppe;set pp;/*emi internes des pp*/
    com=com1;
    reg=reg1;
    if com ne com2 then do;mvt='PPEint';output
ppe;end;/*Emigration interne*/
    keep codern sexe grage age com reg men1 men2
mvt;
run;
data ppi;set pp;/*immi internes des pp*/
    com=com2;
    reg=reg2;
    if com ne com1 then do;mvt='PPlint';output
ppi;end;/*Emigration interne*/
    keep codern sexe grage age com reg men1 men2
mvt;
run;
data ppp;set pp;/*pas de mouvements internes des
pp*/
    com=com2;
    reg=reg2;
    if com eq com1 then do;output ppp;end;
    keep codern sexe grage age com reg men1
men2;
run;
data final;
    set ppp ppi ppe ap pa;
    if mvt='APNais' then
do;men1='10';grage='00000';end;/*Naissances*/
    if mvt='APExt ' then men1='12';/*Immigrations
en provenance exterieur*/
    if mvt='PADec ' then men2='10';/*Deces*/
    if mvt='PAExt ' then men2='12';/*Emigrations a
destination exterieur*/
    if mvt='APlint' then men1='11';/*Immigrations
en provenance d une autre commune*/
    if mvt='PPlint' then men1='11';/*Immigrations
en provenance d une autre commune de la meme
region*/
    if mvt='PPEint' then men2='11';/*Emigrations
vers une autre commune de la meme region*/
run;
/*Numerateurs*/
proc freq data=final noprint;tables
com*men1*men2*grage*sexe/list sparse
out=num(drop=percent rename=count=numreg);
data num;set num;cle=sexe||grage||com||men1;
proc sort data=num;by cle;
/*Denominateurs*/

```

```

proc freq data=final noprint;tables
com*men1*grage*sexe/list sparse
out=den(drop=percent rename=count=denreg);
data den;set den;cle=sexe||grage||com||men1;
proc sort data=den;by cle;
/*Regroupement*/
data transcom;merge num den;by cle;
drop cle;
run;
/*Ensemble de la region */
/*Numerateurs*/
proc freq data=final noprint;tables
men1*men2*grage*sexe/list sparse
out=num(drop=percent rename=count=numreg);
data num;set num;cle=sexe||grage||men1;

proc sort data=num;by cle;
/*Denominateurs*/
proc freq data=final noprint;tables
men1*grage*sexe/list sparse
out=den(drop=percent rename=count=denreg);
data den;set den;cle=sexe||grage||men1;

proc sort data=den;by cle;
/*Regroupement*/
data transnat;length com$5;merge num den;by
cle;
if symget('r')='B' then com='02000';
if symget('r')='V' then com='01000';
if symget('r')='W' then com='03000';
drop cle;
run;
data trans;
length periode $ 9;
set transnat transcom;
periode=compress(&t1||'-'||&t2-1);
run;
proc sort data=trans;
by com men1 men2 grage sexe;
run;
data transitions;
set transitions trans;
run;
%mend;
%macro boucle4();
/*Boucle de couplage des donnees */
%do t=&debut %to %eval(&fin-5) %by 5;
%mouvementsreg(&t,B);
%mouvementsreg(&t,V);
%mouvementsreg(&t,W);
%end;
%mend;
%boucle4();
data donnees.transitionsreg;

```

```

set transitions;
if _N_ > 1;
run;
proc datasets lib=donnees memtype=data nolist
noprint;
modify transitionsreg;
attrib _all_ label="";
attrib _all_ format=;
run;
quit;
/*****
/*6. Transitions au niveau national*/
*****/
data transitions;
run;
%macro mouvements(t);
%let t1=&t;
%let t2=%eval(&t1+5);
%put &t1 &t2;
data couplage&t1&t2;
set donnees.couplage&t1&t2;
age=%eval(&t1)-anaiss-1;
select;
when(age lt 0) grage='00000';
when(0 le age lt 5) grage='00_04';
when(5 le age lt 10) grage='05_09';
when(10 le age lt 15) grage='10_14';
when(15 le age lt 20) grage='15_19';
when(20 le age lt 25) grage='20_24';
when(25 le age lt 30) grage='25_29';
when(30 le age lt 35) grage='30_34';
when(35 le age lt 40) grage='35_39';
when(40 le age lt 45) grage='40_44';
when(45 le age lt 50) grage='45_49';
when(50 le age lt 55) grage='50_54';
when(55 le age lt 60) grage='55_59';
when(60 le age lt 65) grage='60_64';
when(65 le age lt 70) grage='65_69';
when(70 le age lt 75) grage='70_74';
when(75 le age lt 80) grage='75_79';
when(80 le age lt 85) grage='80-85';
when(85 le age lt 90) grage='85-90';
when(90 le age) grage='90+ ';
end;
run;
data pa ap pp;set couplage&t1&t2;
select;
when(p='PP') output pp;
when(p='PA') output pa;
when(p='AP') output ap;
end;
run;
/*Cas des PA : ils sortent du pays par deces ou emi
internationale*/

```

```

data pa;merge pa(in=a) donnees.dec(in=b
where=(put(&t1,4.) le amig lt put(&t2,4.)));
by codern;
if (a and not b) then mvt='PAExt ';/*Emi
etranger*/
if (a and b) then mvt='PADec ';/*Deces dans la
commune parmi ceux qui sont presents en debut de
periode*/
if a;
rename com1=com;
keep p codern sexe grage age com1 men1 mvt;
run;
/*Cas des AP : ils entrent dans le pays par naissance
ou immi internationale*/
data ap;merge ap(in=a) donnees.nai(in=b
where=(put(&t1,4.) le amig lt put(&t2,4.)));
by codern;
if (a and not b) then mvt='APExt ';/*Immi
etranger*/
if (a and b and com2 eq commvt) then
mvt='APNais';/*Naissance*/
if (a and b and com2 ne commvt) then
mvt='APInt';/*Immi interne nne*/
if a;
rename com2=com;
keep p codern sexe grage age com2 men2 mvt;
run;
/*Cas des PP : niveau national*/
data ppe;set pp;/*emi internes des pp*/
com=com1;
if com ne com2 then do;mvt='PPEint';output
ppe;end;/*Emigration interne*/
keep p codern sexe grage age com men1 men2
mvt;
run;
data ppi;set pp;/*Immi internes des pp*/
com=com2;
if com ne com1 then do;mvt='PPlint';output
ppi;end;/*Immigration interne*/
keep p codern sexe grage age com men1 men2
mvt;
run;
data ppp;set pp;/*pas de mouvements internes des
pp*/
com=com2;
if com eq com1 then do;output ppp;end;
keep p codern sexe grage age com men1 men2;
run;
data final;
set ppp ppi ppe ap pa;
if mvt='APNais' then
do;men1='10';grage='00000';end;/*Naissances*/
if mvt='APExt ' then men1='12';/*Immigrations
en provenance etranger*/
if mvt='PADec ' then men2='10';/*Deces*/

```

```

        if mvt='PAExt ' then men2='12';/*Emigrations
vers etranger*/
        if mvt='APIint' then men1='11';/*Immigrations
en provenance d une autre commune*/
        if mvt='PPIint' then men1='11';/*Immigrations
en provenance d une autre commune*/
        if mvt='PPEint' then men2='11';/*Emigrations
vers une autre commune*/
run;
/*Numerateurs*/
proc freq data=final noprint;tables
com*men1*men2*grage*sexe/list sparse
out=num(drop=percent rename=count=numnat);
data num;set
num;cle=sexe| |grage| |com| |men1;
proc sort data=num;by cle;
/*Denominateurs*/
proc freq data=final noprint;tables
com*men1*grage*sexe/list sparse
out=den(drop=percent rename=count=dennat);
data den;set
den;cle=sexe| |grage| |com| |men1;
proc sort data=den;by cle;
/*Regroupement*/
data transcom;merge num den;by cle;
drop cle;
run;
/*Ensemble du pays */
/*Numerateurs*/
proc freq data=final noprint;tables
men1*men2*grage*sexe/list sparse
out=num(drop=percent rename=count=numnat);
data num;set num;cle=sexe| |grage| |men1;

proc sort data=num;by cle;
/*Denominateurs*/
proc freq data=final noprint;tables
men1*grage*sexe/list sparse
out=den(drop=percent rename=count=dennat);
data den;set den;cle=sexe| |grage| |men1;

proc sort data=den;by cle;
/*Regroupement*/
data transnat;length com$5;merge num den;by
cle;
com='00100';
drop cle;
run;
data trans;
length periode $ 9;
set transnat transcom;
periode=compress(&t1|'-'||&t2-1);
run;
proc sort data=trans;
by com men1 men2 grage sexe;

```

```

run;
data transitions;
set transitions trans;
run;
%mend;
%macro boucle();
/*Boucle de couplage des donnees */
%do t=&debut %to %eval(&fin-5) %by 5;
%mouvements(&t);
%end;
%mend;
%boucle();
data donnees.transitionsnat;
set transitions;
if _N_ >1;
run;
proc datasets lib=donnees memtype=data nolist
noprint;
modify transitionsnat;
attrib _all_ label=";
attrib _all_ format=;
run;
quit;

/*****/
/*recodification des fichiers de donnees -->
Matlab*/
/*limitation de la taille des fichiers --> tout en
numerique*/
/*****/
/*7. Exportation structures de la population */
data structures;
retain an com men sexe age eff;
set donnees.structures;
select;
when(grage='00000') age=0;
when(grage='00_04') age=1;
when(grage='05_09') age=2;
when(grage='10_14') age=3;
when(grage='15_19') age=4;
when(grage='20_24') age=5;
when(grage='25_29') age=6;
when(grage='30_34') age=7;
when(grage='35_39') age=8;
when(grage='40_44') age=9;
when(grage='45_49') age=10;
when(grage='50_54') age=11;
when(grage='55_59') age=12;
when(grage='60_64') age=13;
when(grage='65_69') age=14;
when(grage='70_74') age=15;
when(grage='75_79') age=16;
when(grage='80-85') age=17;
when(grage='85-90') age=18;
when(grage='90+ ') age=19;

```

```

end;
select;
when(annee='2000') an=1;
when(annee='2005') an=2;
when(annee='2010') an=3;
when(annee='2015') an=4;
end;

drop grage annee;
run;
proc sort data=structures;
    by an com;
run;
proc export data=structures
    outfile ="&txt\structuresML.csv"
    dbms=csv
    replace;
run;
/*****
/*8. Exportation transits nationaux */
data transitionsnat;
retain p com men1 men2 sexe age num den;
    set
donnees.transitionsnat(rename=(numnat=num
dennat=den));

    select;
    when(grage='00000') age=0;
    when(grage='00_04') age=1;
    when(grage='05_09') age=2;
    when(grage='10_14') age=3;
    when(grage='15_19') age=4;
    when(grage='20_24') age=5;
    when(grage='25_29') age=6;
    when(grage='30_34') age=7;
    when(grage='35_39') age=8;
    when(grage='40_44') age=9;
    when(grage='45_49') age=10;
    when(grage='50_54') age=11;
    when(grage='55_59') age=12;
    when(grage='60_64') age=13;
    when(grage='65_69') age=14;
    when(grage='70_74') age=15;
    when(grage='75_79') age=16;
    when(grage='80-85') age=17;
    when(grage='85-90') age=18;
    when(grage='90+ ') age=19;
end;
select;
when(periode='2000-2004') p=1;
when(periode='2005-2009') p=2;
when(periode='2010-2014') p=3;
end;
drop grage periode;
run;

```

```

proc sort data=transitionsnat;
    by p com men1 men2 sexe age;
run;
proc export data=transitionsnat
    outfile="&txt\transitionsMLp.csv"
    dbms=csv
    replace;
run;
/*****
/*9. Exportation transits regionales */
data transitionsreg;
retain p com men1 men2 sexe age num den;
    set
donnees.transitionsreg(rename=(numreg=num
denreg=den));
    select;
    when(grage='00000') age=0;
    when(grage='00_04') age=1;
    when(grage='05_09') age=2;
    when(grage='10_14') age=3;
    when(grage='15_19') age=4;
    when(grage='20_24') age=5;
    when(grage='25_29') age=6;
    when(grage='30_34') age=7;
    when(grage='35_39') age=8;
    when(grage='40_44') age=9;
    when(grage='45_49') age=10;
    when(grage='50_54') age=11;
    when(grage='55_59') age=12;
    when(grage='60_64') age=13;
    when(grage='65_69') age=14;
    when(grage='70_74') age=15;
    when(grage='75_79') age=16;
    when(grage='80-85') age=17;
    when(grage='85-90') age=18;
    when(grage='90+ ') age=19;
end;
select;
when(periode='2000-2004') p=1;
when(periode='2005-2009') p=2;
when(periode='2010-2014') p=3;
end;
drop grage periode;
run;
proc sort data=transitionsreg;
    by p com men1 men2 sexe age;
run;
data transitionsreg;length reg $1;
    set transitionsreg;
    select;
    when(substr(com,1,2)='21' or com='02000')
reg='b';
    when(substr(com,1,2)='25' or substr(com,1,1)
in('5','6','8','9') or com='03000') reg='w';
    otherwise reg='v';

```

```

    end;
run;
data tmp;set transitionsreg(where=(reg='w'));
    drop reg;
proc export data=tmp
    outfile="%txt%\transitionsMLw.csv"
    dbms=csv
    replace;
run;
data tmp;set transitionsreg(where=(reg='v'));
    drop reg;
proc export data=tmp
    outfile="%txt%\transitionsMLv.csv"
    dbms=csv
    replace;
run;
data tmp;set transitionsreg(where=(reg='b'));
    drop reg;
proc export data=tmp
    outfile="%txt%\transitionsMLb.csv"
    dbms=csv
    replace;
run;
proc export data=transitionsnat
    outfile="%txt%\transitionsMLp.csv"
    dbms=csv
    replace;
run;

```