

XXVI IUSSP International Population Conference

Poster Session 3

Migration, environment and spatial demography

Wednesday, September 30
12:00 PM - 1:30 PM

**Poster Room
Royal room wings**

**Un outil pour analyser la distribution
spatiale d'un événement ou d'une structure
démographique rare**

Michel Poulain, *Université Catholique de Louvain (Louvain-la-Neuve, Belgique)*

Luc Dal, *Université Catholique de Louvain (Louvain-la-Neuve, Belgique)*

Claude Grasland, *Centre National de la Recherche Scientifique (CNRS, Paris, France)*

Un outil pour analyser la distribution spatiale d'un événement ou d'une structure démographique rare

Michel Poulain, *Université Catholique de Louvain (Louvain-la-Neuve, Belgique)*
Luc Dal, *Université Catholique de Louvain (Louvain-la-Neuve, Belgique)*
Claude Grasland, *Centre National de la Recherche Scientifique (CNRS, Paris, France)*

Contact : michel.poulain@uclouvain.be

Introduction

Le démographe est souvent confronté au problème des petits nombres par le fait que certains effectifs de populations sont trop ténus ou que certains phénomènes démographiques correspondent à des probabilités d'occurrence trop faible. Concrètement, il se trouve en présence d'événements démographiques ou de groupes de population que l'on peut qualifier de « rares ». C'est le cas, par exemple, pour la mortalité par certaines causes de décès ou la probabilité de survie entre certains âges, mais aussi pour des groupes de population comme celui composé par les centenaires.

Soit X l'occurrence d'un événement E qualifié de « rare », au sein d'une population de taille P . E peut se réaliser avec une probabilité inconnue θ . Dire que l'événement E est « rare » revient à dire que les estimations de θ seront entachées de fortes fluctuations aléatoires. Ces dernières empêchent toute analyse de la distribution spatiale du phénomène. Notre objectif est de rendre possible une telle analyse en optimisant la visualisation de cette distribution par le recours au lissage spatial et à une cartographie simplifiée.

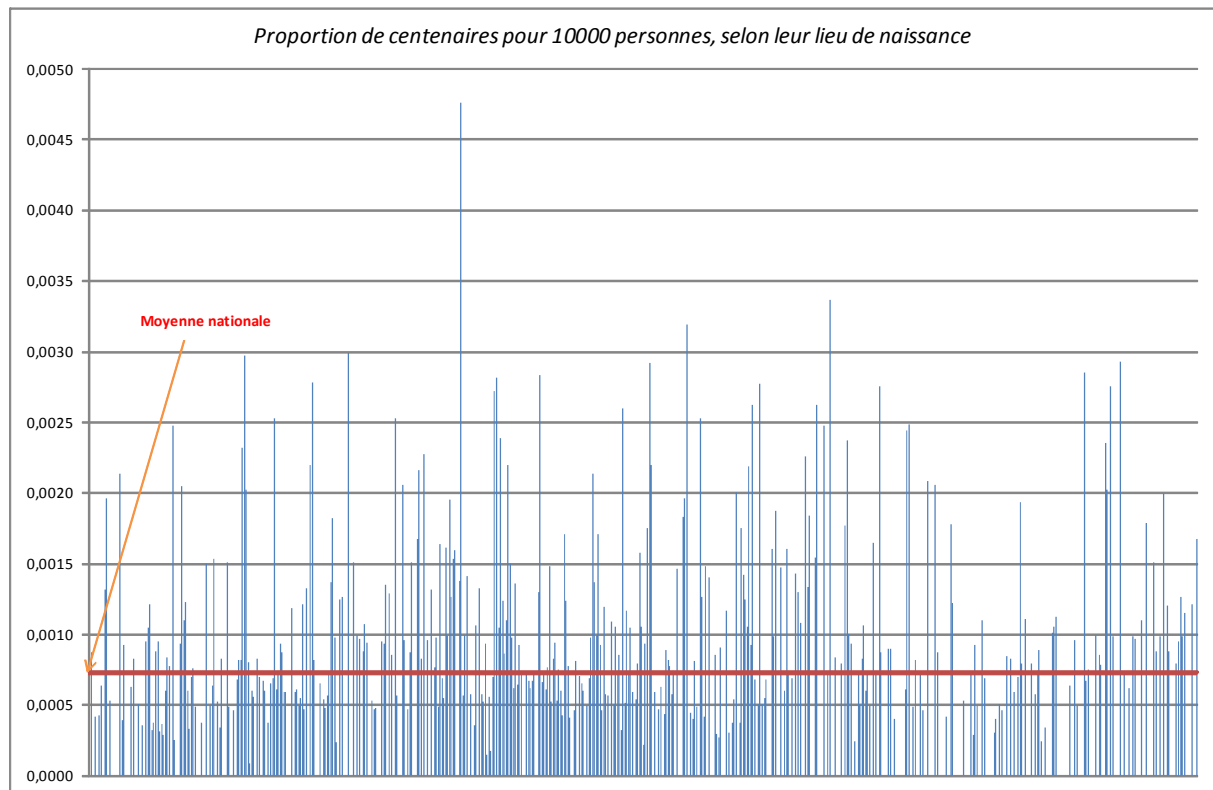
La prévalence des centenaires à l'échelle des 589 communes belges

L'exemple retenu dans cette présentation concerne la distribution des centenaires vivant en Belgique soit en fonction de leur lieu de naissance. Leur nombre sera rapporté à la population des personnes âgées de 65 ans et plus également distribuée par lieu de naissance¹. Pour l'ensemble de la Belgique, le nombre de centenaires au 1^{er} janvier 2006 était de 1174² ($=N$) et la taille de la population des 65+ était de 1.615.844 ($=P$). En moyenne, donc, à l'échelle nationale, l'estimation de θ est donc de $0,727 \cdot 10^{-3}$ ($=N/P$). Si on estime cette même valeur de θ sur base des observations désagrégées au niveau des 589 communes belges, on observera une forte variabilité des estimations comme le montre la distribution de ces estimations sur la figure 1.

¹ Le choix de la population des 65+ est guidé par le souci de rattacher chaque individu à son lieu de résidence à la naissance et non au lieu de naissance de fait qui pour les générations plus récentes est souvent la maternité de la ville voisine.

² Le nombre de centenaires à cette date est de 1298 mais 124 d'entre eux ont été écartés de l'analyse parce qu'ils sont nés à l'étranger.

Figure 1. Distribution des estimations de la proportion de centenaires par commune, au 01/01/2006 en Belgique



L'objectif cette analyse est double : tout d'abord on souhaite mettre en évidence les communes qui présentent une proportion de centenaires significativement plus élevée, ou plus faible, par rapport à la moyenne nationale. Ensuite, on cherche à identifier les zones ou régions plus larges où cette proportion est significativement plus élevée ou plus faible. Cette seconde approche devrait contribuer à mettre en évidence certains déterminants de la longévité ayant une empreinte spatiale. Ce faisant, on fait l'hypothèse d'un déterminisme spatial qui se traduit par le fait que des personnes vivant dans un même lieu présentent une longévité plus semblable par le fait qu'elles partagent certaines caractéristiques communes liées à l'espace (alimentation, environnement, gènes ou caractéristiques socioculturelles, par exemple).

Pour la première démarche visant à mettre en évidence des communes où la proportion de centenaires s'écarte significativement de la moyenne nationale, on procède classiquement à un test de chi-carré; on sait qu'un minimum 'théorique' de 5 observations par cellule est indispensable pour pouvoir tirer des conclusions statistiquement pertinentes. Dans le cas présent, compte tenu de la valeur moyenne calculée à l'échelle nationale, le seuil de taille minimale pour la population soumise au risque (65+) devrait être de 6.882 individus (5/0,000727).

Le test du χ^2

Si la condition du seuil de taille minimale est satisfaite, on peut appliquer le test du χ^2 et tirer des conclusions statistiquement pertinentes.

Pour une commune « i », on calcule donc $\chi^2(i) = \frac{(N(i) - A(i))^2}{A(i)}$ avec $N(i)$ = nombre de centenaires observés et $A(i)$ le nombre de centenaires attendus (si la répartition de ces derniers était uniforme dans l'ensemble du pays) : $A(i) = P(i) * \frac{\sum_j N(j)}{\sum_j P(j)}$.

En retenant un seuil de signification de 5%, ce qui correspond à une valeur de χ^2 est > à 3,841 on pourra admettre pour la commune concernée que la proportion de centenaires y est significativement différente de celle du pays (avec un risque d'erreur de 5%). Inversement, si on observe une valeur de χ^2 inférieure à 3,841, alors on peut dire que la proportion de centenaires dans cette commune est la même que celle du pays, avec un risque d'erreur de 5%.

Seules quelques communes (32 sur 589 dans le cas des lieux de naissance) ont une population des 65+ suffisante ($N > 6.882$ individus) pour pouvoir appliquer valablement le test du chi-carré. On ne peut donc tirer aucune conclusion pour les autres communes.

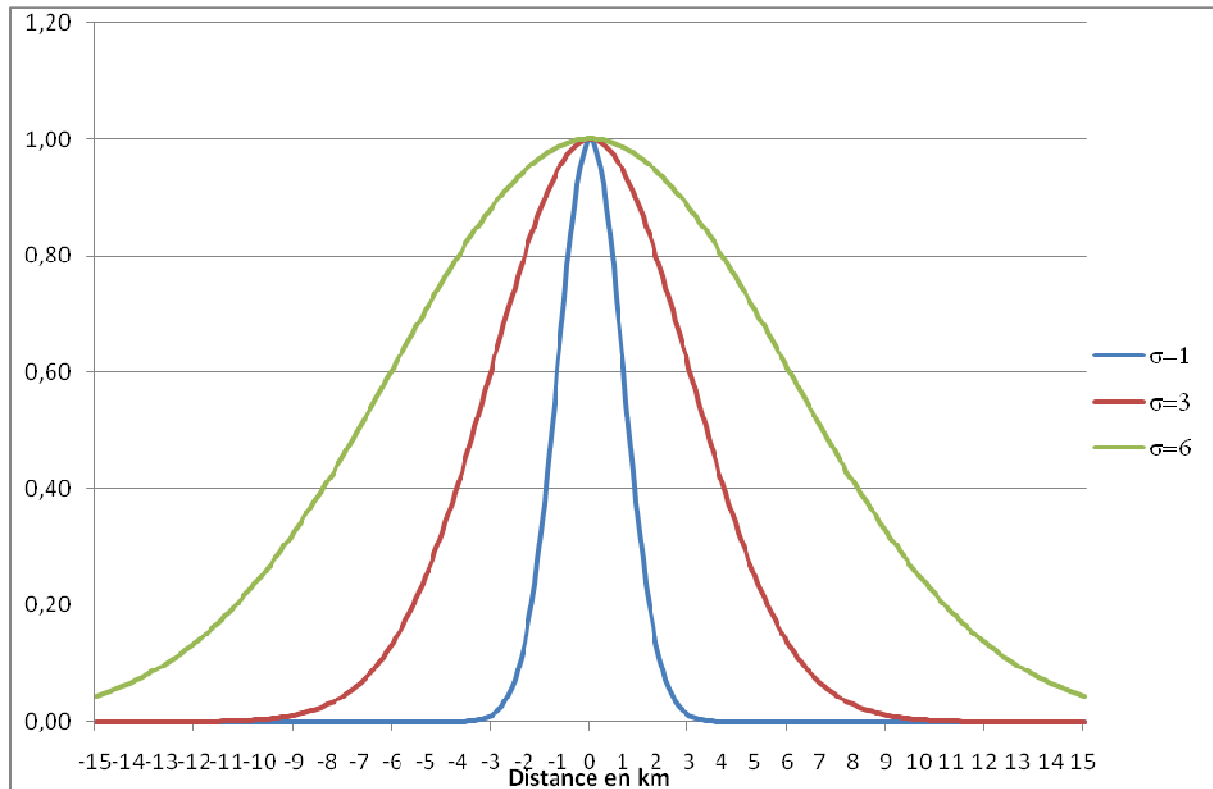
Le recours au lissage gaussien

Pour contourner ce problème, on adopte la méthode de lissage gaussien. Il s'agit d'une méthode de lissage par potentiel basée sur des voisinages spatiaux de portée variable. Brièvement, celle-ci remplace les mesures de comptages par commune par des comptages dans un voisinage centré sur chaque commune en

recourant à la fonction de **Gauss centrée non normalisée** $f(x) = e^{-\frac{1}{2} * \left(\frac{x}{\sigma}\right)^2}$.

Comme le montre la figure 2, cette fonction propose une distribution symétrique autour de la valeur 0 où elle prend une valeur unitaire et décroît avec la distance x selon le paramètre de variance σ^2 .

Figure 2. Fonction de Gauss centrée non normalisée avec trois valeurs croissantes de la variance σ^2



A l'aide de cette fonction on calcule le **potentiel gaussien** pour une commune donnée en ajoutant à la valeur observée pour cette commune celles de toutes les autres communes affectées d'un poids qui décroît avec leur éloignement par rapport à la commune considérée. Cette décroissance suit la distribution gaussienne dont on peut faire varier pas par pas la variance σ^2 de 0 à l'infini. Lorsque la variance tend vers 0, seule la valeur observée pour la commune considérée est prise en compte pour le calcul du potentiel gaussien. Par contre, lorsque l'on incrémente progressivement la valeur de cette variance σ^2 , les valeurs des autres communes s'ajoutent avec un poids qui augmente tout en diminuant toujours avec la distance. A la limite, lorsque la variance tend vers l'infini, le potentiel gaussien correspond à la somme des valeurs de toutes les communes avec un poids unitaire.

Cette méthode de calcul du potentiel gaussien est appliquée aussi bien au nombre de centenaires répartis par leur lieu de naissance qu'à la population des 65+ soumise au risque³. Le potentiel gaussien de population augmente à chaque pas avec la variance et la procédure sera arrêtée lorsqu'il dépasse le seuil critique minimum de 6.882 individus. Sur cette base, on détermine la proportion gaussienne de centenaires par lieu de résidence ou lieu de naissance en rapportant le potentiel gaussien de centenaires au potentiel gaussien de population des 65+ en retenant le niveau de variance critique ainsi atteint afin de pouvoir procéder valablement à un test de χ^2 .

³ Ce qui permet de ne pas considérer les facteurs constants habituels de la distribution gaussienne, puisqu'ils se simplifient par division.

En d'autres termes, la méthode proposée vise à estimer une densité de probabilités (dans l'espace) par la méthode (classique en statistique non paramétrique) des noyaux (« kernel density estimation »). (W. Härdle (1990): *Applied Nonparametric Regression*, B. Silverman (1986): *Density estimation for Statistics and Data Analysis*, M. Wand, R. Jones (1995): *Kernel smoothing*) et présente un intérêt dans le cas d'estimation de densités de probabilités d'événement rares spatialement localisés⁴.

Il s'agit donc de calculer donc pour un voisinage gaussien de portée σ et pour toute commune i , les fonctions suivantes, les sommes sur j se faisant sur toutes les communes et D_{ij} étant la distance entre les communes i et j :

- Potentiel gaussien de centenaires en i : $N^*(i, \sigma) = \sum_j N_j e^{-\frac{\ln(\sigma^2)}{\sigma^2} + D_{ij}^2}$
- Potentiel gaussien de population en i : $P^*(i, \sigma) = \sum_j P_j e^{-\frac{\ln(\sigma^2)}{\sigma^2} + D_{ij}^2}$
- Proportion gaussienne de centenaires en i : $p(i, \sigma) = \frac{N^*(i, \sigma)}{P^*(i, \sigma)}$
- Chi2 de déviation pour i : $\chi^2(i, \sigma) = \frac{(N^*(i, \sigma) - P^*(i, \sigma))^2}{P^*(i, \sigma)}$

Résultats

La proportion de centenaires parmi la population âgée de 65 ans et plus, classée par commune de naissance, est présentée par classe de valeur sur la figure 3. Les proportions y sont exprimées pour mille personnes et la première classe regroupe exclusivement les communes auxquelles aucun centenaire n'a pu être attribué, soit 203 communes sur un total de 589. De toute évidence, la lecture de cette carte est malaisée et une simplification est nécessaire afin de faciliter la visualisation de la répartition spatiale de cette proportion.

Le lissage gaussien présenté dans cette contribution permet de calculer les proportions gaussiennes correspondantes qui sont reprises sur la figure 4 avec les mêmes classes de valeur. Sur celle-ci, des plages plus uniformes ressortent avec notamment la Campine au Nord-est, où la proportion est moindre, et le Hainaut Occidental et les deux Flandres au Nord-Ouest avec une proportion plus élevée.

Toutefois il convient d'appliquer le test du χ^2 afin de vérifier le niveau de signification des écarts observés par rapport à la moyenne nationale. Ce sont ces écarts qui sont présentés sur la figure 5 de la façon suivante : les valeurs sont positives si la proportion gaussienne est supérieure à la moyenne nationale et négatives dans le cas inverse. Hormis le signe, les valeurs proposées sont le complément à l'unité de la p-valeur résultant du test du χ^2 . Cette transformation a été rendue nécessaire afin de faciliter la visualisation de cette

⁴ Le noyau gaussien est non borné et permet de s'étendre à l'infini pour atteindre le niveau de significativité requis (ce qui ne serait pas le cas avec les autres noyaux traditionnels : triangulaire, quadratique, Epanechnikov, ...).

distribution spatiale. A l'examen de cette carte, on peut en déduire qu'une grande partie de la Campine, au Nord-est de la Belgique, présente une proportion de centenaires significativement moindre. A l'inverse, parmi les régions présentant une proportion significativement supérieure, on note principalement la Flandre Occidentale, au Nord-Ouest du pays, et le Hainaut Occidental, au Centre-Ouest.

Enfin, la figure 7 montre le résultat du calcul des proportions gaussiennes lorsque l'on porte le nombre minimum de cas théoriques de 5, comme précédemment, à 20. Ce faisant, on force l'effet du lissage gaussien et l'on perd évidemment en précision tout en facilitant la visualisation. Il ressort de cette dernière analyse que la Campine et plus précisément la province du Limbourg est marquée par une longévité moindre alors qu'à l'ouest du pays, le Westland, entre Veurne, Poperinge et Oostende ainsi que le Hainaut Occidental sont les deux principales aires de longévité de la Belgique. Ces résultats devraient permettre de focaliser les recherches futures sur ces régions afin de déceler ce qui les oppose et qui pourrait être considéré comme déterminant de la longévité.

Figure 3. Nombre de centenaires pour 1000 personnes âgées de 65 ans et plus par commune de naissance (Belgique, 2006, Source : Statistics Belgium).

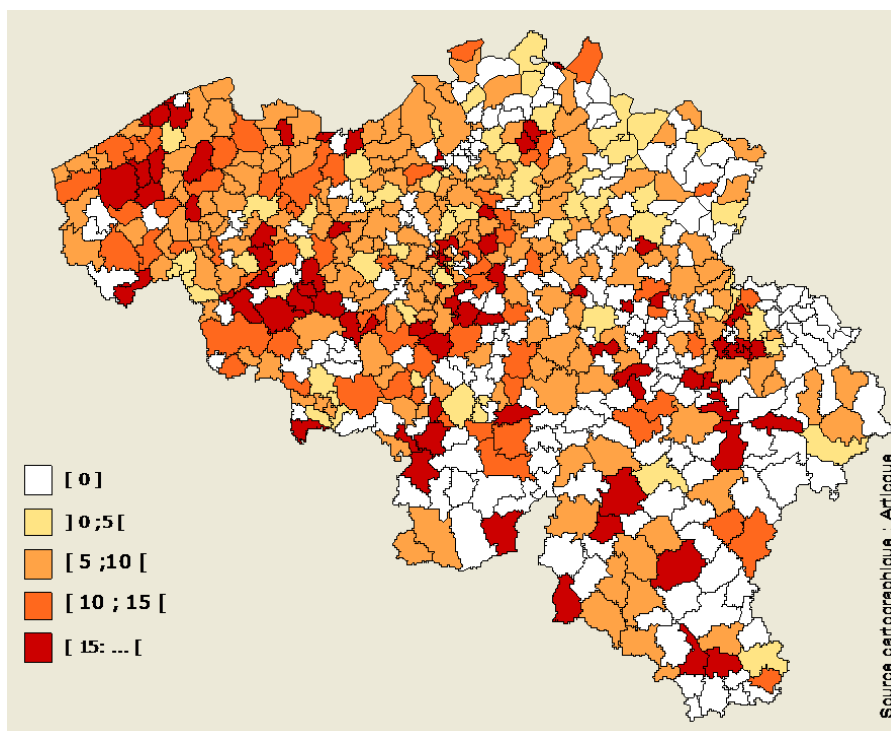


Figure 4. Proportion gaussienne de centenaires par commune de naissance (pour 1000 personnes âgées de 65 ans et plus : Belgique, 2006, Source : Statistics Belgium).

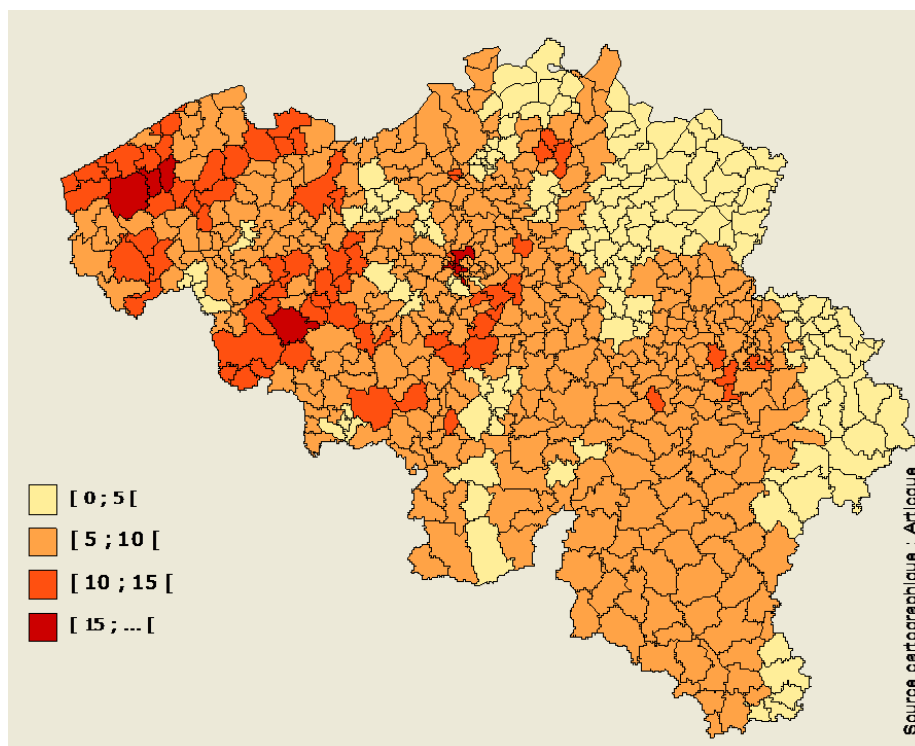


Figure 5. Complément à l'unité des p-valeurs du test du χ^2 , valeurs affectées d'un signe positif lorsque la proportion gaussienne est supérieure à la moyenne nationale et d'un signe négatif, dans le cas inverse.

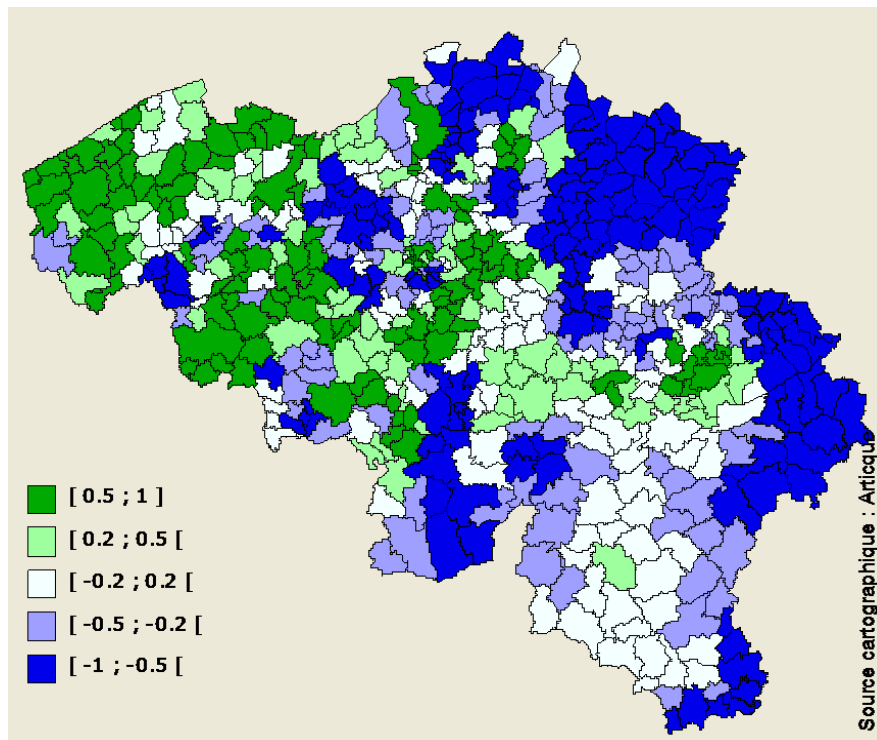


Figure 7. Proportion gaussienne de centenaires par commune de naissance lorsque le seuil pour le lissage gaussien est porté de 5 à 20 cas théoriques.

