

Analyse de données spatiales du marché du travail. Application aux 260 communes wallonnes

Christophe Croux, Etienne Wasmer

► **To cite this version:**

Christophe Croux, Etienne Wasmer. Analyse de données spatiales du marché du travail. Application aux 260 communes wallonnes. 14ème Congrès des Economistes Belges de Langue Française, Jul 2000, Liège, Belgique. pp.1-19. hal-01053490

HAL Id: hal-01053490

<https://hal-sciencespo.archives-ouvertes.fr/hal-01053490>

Submitted on 31 Jul 2014

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

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

Analyse de Données Spatiales du Marché du Travail

Application aux 260 communes wallonnes

Christophe Croux (ECARES, ULB)

Etienne Wasmer (ECARES, ULB)

14ème Congrès des Economistes Belges

de Langue Française

26/07/2000

Liège

Résumé

L'analyse de données spatiales permet d'appréhender des phénomènes en général ignorés de l'analyse économique standard. Dans le cadre du marché du travail, en particulier, l'approche spatiale peut révéler l'importance de la concentration spatiale du chômage, les corrélations dans l'espace de la qualification de la force de travail, la propagation dans l'espace des chocs d'emplois, enfin les raisons possibles de la persistance d'un niveau de chômage élevé dans certains secteurs géographiques. Ce papier tente d'adresser ces questions en appliquant des techniques d'analyse spatiale à des données des 260 communes de la Wallonie. L'objectif est d'abord méthodologique et exploratoire : après une première partie descriptive utilisant les techniques d'analyse de données spatiales, on procède à l'estimation d'un modèle de chômage et d'offre de travail avec ou sans termes de corrélations spatiales. Les conclusions d'étape permettent d'éclairer le rôle de la mobilité du travail et des politiques intra-régionales.

Les auteurs tiennent à remercier très chaleureusement Michel Oris, de l'Université Liège, pour avoir mis à notre disposition la base de données des communes de Wallonie que nous avons utilisée dans ce texte. Nous remercions également Catherine Dehon et Arnout Van Messem pour leur assistance précieuse dans la préparation des données, en particulier la localisation et le regroupement des communes. Nous remercions enfin Muriel Dejemeppe pour ses commentaires.

« *Theoreticians of today are chiefly preoccupied with introducing the time element in full into their analysis, and the literature abounds with models of a dynamic nature. Yet who can deny that the spatial aspects of economic development : that all economic process exist in space, as well as over time ? Realistically, both time and space must be vital considerations in any theory of economy.* » Walter Isard, "The General Theory of Location and Space-Economy", *Quarterly Journal of Economics*, 1949.

1. Introduction

En dépit d'un regain d'intérêt pour les aspects géographiques et spatiaux dans le cadre de la Nouvelle Economie Géographique (voir en particulier Fujita et al. 1999), et malgré le grand nombre de travaux en économie régionale, l'analyse économique du marché du travail tend à ignorer les aspects spatiaux à un niveau fin de désagrégation. Le commentaire de Walter Isard a donc conservé toute son actualité un demi-siècle plus tard. Pourtant, un très grand nombre de mécanismes propres au marché du travail ont des aspects locaux de premier ordre : par exemple, la prospection d'emploi, en particulier pour les travailleurs les moins qualifiés, a un caractère local très marqué. Les décisions d'accepter ou de refuser une offre d'emploi dépendent énormément de la longueur des trajets quotidiens entre le domicile et le travail et des coûts financiers de transports. Toutes les études réalisées pour les collectivités locales indiquent que les décisions de créer des emplois ou d'implanter des industries dépendent de la qualité de la force de travail dans la zone prospectée, de la réserve de main d'œuvre locale, de la connectivité de la région aux autres zones d'activité. Les situations de monopsonne¹ des entreprises sont inhérentes aux aspects spatiaux et au manque de mobilité de la main d'œuvre. Enfin, les décisions de mobilité des travailleurs dépendent de façon cruciale des conditions locales du marché du travail dans la zone de départ et d'arrivée.

Malgré le caractère non-controversé de l'importance de ces mécanismes, les études empiriques (et même théoriques) essayant de capturer les éléments géographiques et spatiaux du marché du travail sont encore assez rares. Cela tient essentiellement à deux raisons : les données à un niveau fin de désagrégation, typiquement le bassin d'emploi (*travel-to-work areas*) ou la commune, voire le quartier d'une commune, ne sont pas facilement disponibles. Les enquêtes du marché du travail, basée sur un échantillonnage de la population, ne permettent pas d'avoir des données statistiquement fiables sur des unités géographiques trop désagrégées, et il faut avoir recours aux données des recensements ou des données exhaustives des administrations pour obtenir une vision précise des phénomènes locaux. D'autre part, la mise en œuvre d'études empiriques à ces niveaux de désagrégation fins implique une complexité accrue, en raison de l'importance de phénomènes dits de « spillovers » ou d'interaction entre les unités considérées : ainsi, si on peut raisonnablement négliger au premier ordre l'interaction entre les marchés du travail des grandes régions d'un pays, il est en revanche plus difficile de faire ce type d'hypothèses dès lors qu'on augmente la précision géographique de l'analyse.² Dès lors, des techniques économétriques habituelles peuvent être inadaptées dans la mesure où elles conduisent à des inefficacités voire des biais systématiques dans l'estimation des paramètres d'intérêt économique. Case (1991) justifie ainsi l'analyse spatiale des données lorsque des variables inobservées présentent une structure de corrélation spatiale (telles que le climat) : si des données de panel sont disponibles, le traitement par effet fixe peut être un substitut. Mais dans le cas de coupes transversales, la perte d'efficacité des estimateurs classiques peut être très grande.

¹ C'est à dire de pouvoir de marché des entreprises sur leurs salariés

² Même au niveau régional, par exemple les 22 régions françaises ou les 10 régions anglaises, on peut identifier des termes d'interaction spatiale, par exemple dans l'estimation de fonctions d'appariement (voir Petrongolo et Wasmer 1999 sur ce point).

Dans cette étude, nous avons résolu cette double difficulté (non-disponibilité des données et mise en œuvre de techniques plus complexes d'estimation). En effet, nous avons obtenu un accès à une base de données d'une très grande richesse, qui reprend, commune par commune de Wallonie, soit 260 unités, l'ensemble des informations statistiques connues des services régionaux ou fédéraux. Ainsi, les données du recensement, des données de l'INS, voire celles de la police et de la gendarmerie, qui ont été réunies par une équipe de chercheurs de l'Université de Liège, nous ont été communiquées dans le cadre exclusif du 14ème Congrès des Economistes Belges de Langue Française. La description des données que nous utiliserons est placée en annexe de ce document. D'autre part, la technique d'analyse des données que nous avons retenue (analyse descriptive puis économétrique) tient compte des interactions spatiales. Nous nous sommes basés sur Cliff et Ord (1981) et Cressie (1993) pour la théorie statistique.

Pourquoi cet intérêt pour les données spatiales ? La réponse est simple : elles révèlent d'intéressantes régularités. Notamment, en anticipant un peu sur la description des données, il apparaît que le taux de chômage des communes a une très forte variabilité, et que cette variabilité définit des gradients³ de taux de chômage importants, avec la coexistence de zones à chômage élevé et des zones à chômage faible. Cette dispersion, plus forte dans les communes défavorisées, a alimenté le thème de la composante géographique de la *fracture sociale*, c'est à dire l'existence de zones ou de quartiers en difficulté économique ou sociale. Pourtant, comme le notait tout récemment le géographe Christophe Guilhy (*Libération*, 16 Août 2000), '*S'il n'est pas contestable que certains quartiers concentrent plus que d'autres les difficultés, force est de constater que ceux-ci se définissent aussi par leur diversité sociale et urbaine*'. Notre analyse a précisément pour but d'intégrer ce type de variables explicatives, variables d'éducation pour la composante sociale et économique, taille de la population ainsi que la structure du chômage dans le voisinage des communes pour la composante urbaine.

Dans ce travail, on s'intéressera à la concentration spatiale du chômage, aux corrélations dans l'espace de la qualification de la force de travail, à la propagation dans l'espace des chocs d'emplois, enfin aux raisons possibles de la persistance d'un niveau de chômage élevé dans certains secteurs géographiques. Etant donné l'absence (à notre connaissance) de travaux économiques sur le chômage au niveau des communes de Wallonie, nous commencerons par une partie descriptive utilisant les techniques d'analyse de données spatiales. On étudiera notamment la décroissance des corrélations entre les taux de chômage des communes dans l'espace, ainsi que la concentration des qualifications de la population active dans certaines zones de la Wallonie. Ensuite, on procédera à l'estimation d'un modèle de chômage et d'offre de travail minimaliste pour vérifier que l'inclusion de termes de corrélations spatiales a un impact sur les coefficients estimés. Etant donné le caractère exploratoire de notre approche, nous resterons prudents quant aux implications de politique économique, nous contentant de souligner telles ou telles d'entre elles conditionnellement à l'interprétation des résultats du modèle économétrique. Ainsi, on évoquera brièvement la nécessité de la mobilité du travail et des politiques intra-régionales.

2. Définition des données.

Bien que la base de données soit très vaste, nous en extrayons un sous-ensemble limité de variables : le taux de chômage des communes ; le niveau d'éducation, mesuré comme la fraction des individus de la population active dans les différentes catégories de diplôme ;

³ Ce terme technique emprunté à la physique signifie la variation dans l'espace de la grandeur considérée, ce qu'on peut voir comme une déclivité si le taux de chômage représentait une altitude.

enfin la taille de la population active.

Notre définition du taux de chômage par rapport aux données a nécessité un petit ajustement détaillé ci-après. Pour comprendre le problème, rappelons qu'il y a potentiellement deux définitions du chômage. La première, de nature administrative, est la suivante: il s'agit des *chômeurs complets indemnisés*, c'est à dire des chômeurs inscrits comme demandeurs d'emploi temps plein, non lié par un contrat de travail, qui bénéficie d'allocations de chômage ou d'attente. Cette catégorie regroupe à la fois des travailleurs dont le contrat de travail a pris fin et des jeunes qui ont terminé leurs études ou leur contrat d'apprentissage et effectuent leur stage d'attente.

La seconde correspond aux critères retenus par l'Office Statistique des Communautés Européennes (OSCE). Le demandeur d'emploi doit remplir les trois conditions suivantes : être sans emploi, à la recherche d'un emploi et être directement disponible pour un emploi. Malheureusement, cette définition théorique n'a pas été respectée et les données issues du recensement que nous utilisons ont pris l'ensemble des demandeurs d'emploi en compte, qu'ils soient employés ou non. De plus, le taux de chômage de chaque commune est défini comme le ratio des demandeurs d'emploi à la population active occupée qui est l'ensemble de la population exerçant une profession, une fonction ou une activité lucrative, y compris les aidants – les chômeurs étant exclus). Cette distribution (commune par commune) a les propriétés statistiques suivantes :

Statistiques pour la variable 'demandeur d'emplois' rapportée à l'emploi:

Min.	1 ^{er} Q.	Méd.	Moy.	3 ^{ème} Q.	Max.
5.2	11.1	14.0	15.9	19.4	41.7

On a donc par rapport à la mesure traditionnelle du taux de chômage, un double biais à la hausse : le numérateur comprend tous les demandeurs d'emplois (employés et chômeurs) ; et le dénominateur ne comporte que la population active occupée. On peut traiter le deuxième biais, de façon à obtenir le taux de chômage conventionnel, défini comme le ratio du nombre de chômeurs sur la population active total, incluant les chômeurs, en appliquant la transformation $x/(1+x)$. On obtient ainsi :

Statistiques pour le taux de chômage:

Min.	1 ^{er} Q.	Méd.	Moy.	3 ^{ème} Q.	Max.
4.9	10.0	12.2	13.4	16.2	29.4

On peut constater que le taux de chômage moyen non pondéré pour la taille de la commune est de 13.4%, ce qui est comparable aux chiffres officiels pour la Wallonie (d'après l'INS, le taux de chômage administratif était de 13.1% en Wallonie en 1991). Il est à noter toutefois que ce chiffre diverge du taux de chômage au sens du Bureau International du Travail qui était de 10.1%. Cette définition est la plus proche des critères de l'OSCE, qui était de 10.1%). Dans l'analyse qui suit, on utilisera le taux de chômage ou sa transformée logistique basée sur le taux corrigé de la seconde tabulation.

On peut également constater dans le deuxième tableau ci-dessus, ainsi qu'il était annoncé en introduction, la forte dispersion des taux de chômage des communes, puisque l'écart inter-quartile est de 5.1% de taux de chômage entre le minimum et le quart inférieur, et de 13.2% entre le quart supérieur et maximum !

La population active est classée selon le plus haut niveau de diplôme obtenu. Il y a en tout 12 classes de diplômes que nous avons regroupés dans trois classes (A, B et C). La première

correspond aux individus ‘sans diplôme ou secondaire inférieur’, la seconde aux ‘secondaire supérieur’, la troisième aux individus ayant un diplôme du supérieur (universitaire ou de type court). On a exclu les 3.6% de la population pour lesquels le diplôme était déclaré ‘non connu’ dans l’enquête.

3. Définition des concepts spatiaux.

Nous sommes intéressés par l’étude d’un échantillon $\{y_s \mid s \in D\}$, où D est une partie finie de \mathfrak{R}^2 . Chaque indice s correspond aux coordonnées géographiques du centre d’une commune. L’ensemble D forme une lattice (une grille) irrégulière en \mathfrak{R}^2 . Dans la Figure 1, nous voyons l’ensemble D pour l’échantillon de base, à savoir la position des 260 communes et les villes principales. Nous utiliserons donc des techniques de la statistique spatiale pour l’analyse des données sur une lattice. La Figure 2 donne une idée de la dispersion du taux de chômage : on peut voir des zones de concentration du chômage autour de Mons d’une part, et de Charleroi-Namur d’autre part. La région de Liège présente également une zone de chômage quoi que moins marquée. Les zones du Nord proches de Bruxelles et de l’Est proches de l’Allemagne et du Grand-Duché du Luxembourg sont en revanche nettement plus favorisées. On peut noter ainsi l’existence de gradients de chômage et de corrélations spatiales assez fortes qu’il nous faudra analyser.

Afin d’avoir une première idée des dépendances spatiales dans l’échantillon, nous allons utiliser deux mesures classiques en analyse spatiale : le variogramme et le corrélogramme. Ce sont en fait des analogues du covariogramme et corrélogramme pour l’analyse d’une série chronologique.

3.1 Le variogramme. Le variogramme théorique est la fonction

$$\gamma(h) = \frac{1}{2} E[(y_s - y_r)^2]$$

pour $\|s-r\| = h$ et $h > 0$.⁴ La quantité $\gamma(h)$ mesure la grandeur de la différence entre des observations y_s à distance h . En l’absence de corrélation spatiale, nous trouverons $\gamma(h) = \text{constante} = \gamma(\infty) = \text{Var}(y_s)$ pour chaque $h > 0$. En présence de corrélation spatiale (positive), la fonction $\gamma(h)$ augmentera plus lentement vers sa valeur limite. La version empirique de $\gamma(h)$ se calcule sur une grille des distances h_j pour $j = 1, 2, \dots, M$ où $h_j = j\delta - \frac{\delta}{2}$ et $\delta > 0$. On considère maintenant les ensembles

$$N(h_j) = \{(s, r) \in D \times D \mid \|s - r\| \in [h_j - \frac{\delta}{2}, h_j + \frac{\delta}{2}]\}$$

et on calcule

$$\hat{\gamma}(h_j) = \frac{1}{2|N(h_j)|} \sum_{(s,r) \in N(h_j)} (y_s - y_r)^2,$$

⁴ $\|\cdot\|$ représente la norme euclidienne. Le variogramme théorique est bien défini si le processus est stationnaire (dans l’espace).

où $|N(h_j)|$ est le nombre de points (s, r) appartenant à $N(h_j)$.⁵ Le paramètre δ est un paramètre de lissage qui doit être suffisamment grand de façon à ce que $N(h_j)$ contienne suffisamment d'observations pour pouvoir calculer une moyenne précise. D'un autre côté, δ doit rester suffisamment petit pour s'assurer que seules des différences $\|s-r\|$ proches de h_j soient prises en compte.⁶ La Figure 3a donne le variogramme empirique pour nos données de chômage.⁷ Nous remarquons immédiatement la présence d'une corrélation spatiale, très nette dans un rayon de 30 kilomètres, dans la mesure où le variogramme augmente nettement de 0 à 30 km. Au delà de 30 km, le variogramme continue à croître, mais plus lentement, ce qui suggère une atténuation des phénomènes spatiaux entre 30 et 100 km.

3.2 Le corrélogramme. Pour tenter de confirmer ces constats, nous introduisons maintenant le corrélogramme, qui est souvent considéré comme plus facile à interpréter. Le corrélogramme est défini par

$$\rho(h) = \text{Corr}(y_s, y_r) \text{ pour } \|s-r\| = h.$$

La version empirique de la statistique est représentée sur la Figure 3b. On peut à nouveau constater que la distance de 30 kilomètres joue un rôle pivot: la statistique décroît de façon importante de zéro à 30 km, puis le corrélogramme est proche de zéro au-delà. On en tire la conclusion que l'analyse des données communales nécessite l'utilisation de techniques spécifiques, dans la mesure où toutes les communes ont des voisins à une distance de moins de cette limite des 30 kilomètres (aucune commune n'est isolée).

4. Estimation d'un modèle de chômage basé sur l'offre de travail

4.1. Le modèle retenu. Dans ce travail, on souhaite être relativement simple par rapport au modèle utilisé de façon à souligner l'intérêt de l'approche spatiale. En particulier, on se limite dans le choix des explicatives à un nombre restreint de variables décrivant l'offre de travail et le type de commune. En ce sens, notre modèle n'est pas un vrai modèle structurel, mais plutôt une forme réduite dont on souhaite filtrer les résidus (les termes non-expliqués par le modèle) pour leurs aspects spatiaux. De plus, nous faisons le choix, dans cette première étape, de ne prendre que les variables d'offre de travail en considération: de ce fait, les résidus de ce modèle seront interprétables, le cas échéant, comme la somme de variables de demande de travail. On aura donc dans la structure spatiale des résidus l'impact de la structure spatiale des termes de demande de travail. Nous reviendrons sur ce point ultérieurement. Dans l'immédiat, les variables explicatives du modèle sont donc des termes d'offre: d'une part, les variables EduA, EduB et EduC (les parts des individus des différents niveaux d'étude dans la population active) et d'autre part une variable de taille de la commune, le logarithme de la population active de la commune. Cette variable doit capturer les problèmes spécifiques aux grandes agglomérations. Il est important avant de choisir la spécification du modèle d'analyser la variation des régresseurs dans l'espace, représentée sur la Figure 4. Sur cette figure, on peut remarquer, en blanc sur le graphique a), la faible densité des communes de la

⁵ Si le processus y_s est stationnaire, $\hat{\gamma}(h)$ est un estimateur convergent pour $\gamma(h)$ pour $\delta \rightarrow 0$ et $N(h) \rightarrow \infty$. Dans notre papier, $\hat{\gamma}(h)$ est plutôt utilisé comme outil descriptif de la structure spatiale que comme un estimateur.

⁶ Dans la littérature, on conseille de prendre δ tel que chaque $N(h_j)$ contienne au moins 30 paires $(s, r) \in D \times D$ avec un M tel que h_M est la moitié de la distance maximale entre 2 points de D .

⁷ Tous les calculs sont réalisés avec le logiciel « S⁺, Spatial stats » et en prenant les valeurs par défaut des routines (Kaluzny et al. 1998).

province du Luxembourg et des Ardennes belges. On peut aussi noter que la concentration de qualifications en Wallonie est très marquée dans la région du Brabant Wallon, ainsi qu'autour de Mons et de Liège, mais que le reste de la région est relativement pauvre en qualification, ainsi que l'atteste la prévalence de couleur blanche dans la graphique d) représentant la densité de la part des diplômés de la catégorie C. Une représentation graphique de type 'boxplot' non reportée ici nous indique que les fractions EduA et EduC sont relativement dispersées entre communes, alors que la dispersion de EduB est beaucoup plus faible. On choisira de prendre EduC comme référence du modèle. Par ailleurs, après de nombreux choix de spécification, il s'avère que la présence d'un terme non-linéaire en EduA est nécessaire : on expliquera et interprétera ce choix plus tard, en commentant le signe du coefficient quadratique.

On aurait également pu inclure des variables d'âge ou de sexe de la commune. Les variables d'âge présentent la caractéristique d'être très homogène entre communes, ainsi qu'une analyse 'boxplot' non reportée nous l'a révélé. De plus, et sans doute comme conséquence de ce fait, le modèle ainsi estimé présentait des coefficients pour les variables d'âge qui étaient peu significatifs. Quant à la composition hommes-femmes de l'emploi des communes, nous avons préféré ne pas l'inclure étant donné son endogénéité possible et avons préféré estimer le modèle simple séparément pour les hommes et pour les femmes. Étant donné que les résultats ont été assez proches du modèle agrégé hommes+femmes, nous ne reportons pas ces résultats. Au total, le modèle que nous présentons ici nous paraît être un bon choix parmi l'ensemble des modèles possibles.

La variable endogène sera la transformée logistiquée du taux de chômage qui est représentée sur la Figure 5. Cette figure est très proche de la Figure 2 qui représentait le chômage en taux, et on retrouve en fait la même structure spatiale avec la transformée.

4.2 Résultats du modèle linéaire. D'après la littérature économique, on s'attend à ce que la variable EduA ait un signe positif : une fraction de non-diplômés plus élevée, par un simple effet de composition, contribuera à augmenter le chômage d'équilibre de la commune⁸. La variable EduB devrait également avoir un coefficient positif, pour les mêmes raisons. Le coefficient du logarithme de la population active devrait être positif, si on en croit Harris et Todaro (1970) : le taux de chômage élevé des grandes agglomérations est un fait connu, qui s'explique par l'existence d'emplois mieux payés dans les villes, avec néanmoins un nombre rationné. Dès lors, les migrations vers les villes augmentent le chômage comme un phénomène de files d'attente.

La Table 1 en Annexe donne les résultats du modèle linéaire. La proportion de variance expliquée est de 40%, ce qui est élevé pour notre simple modèle de chômage en coupe transversale. On s'aperçoit effectivement que les coefficients d'EduA, EduB et le logarithme de la taille sont tous trois positifs, et significatifs au seuil de 1%. En revanche, le terme quadratique en EduA est négatif, et également significatif à 1%. L'interprétation statistique pour ce signe est que l'impact marginal d'un nombre croissant de non-qualifiés sur le taux de chômage de la population de la commune est négatif. Ce résultat est à priori surprenant : on aurait pu croire à un coefficient positif si la concentration de non-qualifiés génère une externalité négative sur le niveau de qualification de la commune, dans l'esprit de la littérature sur les ghettos (voir par exemple Benabou 1993). En réalité, le coefficient

⁸ L'explication théorique habituelle du chômage plus élevé des non-qualifiés est que leur productivité est plus faible que celles des qualifiés, mais que leur salaire n'est pas plus bas en proportion, ce qui réduit d'autant la demande de travail pour cette catégorie.

négatif du terme quadratique semble indiquer que les concentrations très élevées de non-qualifiés dans une commune vont de pair avec l'existence d'industrie ou d'emplois pour ces faibles niveaux de qualification : ainsi, des industries de main d'œuvre s'implanteront plus aisément lorsqu'elles pourront facilement recruter dans la population locale. Une autre interprétation est un mécanisme d'ajustement économique par les salaires : les concentrations très élevées de non-qualifiés ne peuvent indéfiniment faire croître le chômage, car les salaires doivent s'ajuster à la baisse. Enfin, les mécanismes d'aide économique aux régions défavorisées peuvent également justifier la concavité de l'effet de la variable EduA.

4.3 Statistique de Moran. Nous avons maintenant besoin de tester la pertinence du modèle non-spatial précédent et notamment l'absence de corrélation spatiale de ses résidus. Il faut définir un test basé sur une statistique dite de Moran.

En premier lieu, nous devons spécifier une matrice W , de dimension $N \times N$ (où N est le nombre de communes), qui sera appelée la matrice de voisinage (neighborhood matrix). La matrice W contient l'information sur la proximité des communes. L'élément w_{rs} est une fonction de la distance entre deux communes : nous pourrions imaginer que w_{rs} est déterminé par d'autres facteurs tels que les infrastructures de transports, le temps de transport entre les deux communes ou encore le coût de transport d'une unité de bien : c'est en fait une distance économique que nous cherchons à approcher. Le choix d'une spécification particulière pour les éléments de cette matrice dépend de la théorie économique que l'on utilise. Ayant également expérimenté les éventuelles asymétries axiales (les axes Nord-Sud et Est-Ouest semblent ne pas jouer de rôle très distincts), nous nous en sommes tenus à la spécification la plus simple et travaillons ici avec la spécification suivante des poids : $w_{rs} = 1$ si $\|s-r\| \leq 30$ et $w_{rs} = 0$ sinon. Cela implique que les communes situées à une distance inférieure à 30 kilomètres sont considérées comme des voisins.⁹

Une mesure de corrélation spatiale est la statistique de Moran, appelée encore l'indice de Moran qui est noté I

$$I = \frac{\sum_{r=1}^N \sum_{s=1}^N w_{rs} (y_s - \bar{y})(y_r - \bar{y})}{\sum_{r=1}^N \sum_{s=1}^N w_{rs}} \frac{N}{\sum_{r=1}^N (y_s - \bar{y})^2}$$

où $\bar{y} = \frac{1}{S} \sum_{s=1}^N y_s$. Cette mesure doit être vue comme une sorte de coefficient d'autocorrélation pour une analyse spatiale. La loi asymptotique de l'indice de Moran I a été obtenue sous l'hypothèse nulle d'absence de corrélation spatiale et peut être utilisée pour un test statistique.¹⁰

Nous appliquons effectivement cette statistique pour tester l'absence de corrélation dans les résidus de l'ajustement linéaire. On peut également calculer l'indice de Moran pour les données de chômage brute (en transformation logistique). On trouve une valeur de 0.2902,

⁹ La matrice W contient donc beaucoup de zéros. Il est conseillé de prendre des voisinages plutôt petits de façon à limiter la complexité des calculs sur la base de cette matrice W et de ne pas perdre trop de degrés de liberté. Le choix de 30 km correspond à l'atténuation du corrélogramme, mais les résultats sont robustes dans un intervalle de 20 à 40 km.

¹⁰ Une comparaison de différents tests de corrélation spatiale est donnée par Kelejian et Robinson (1998).

avec une statistique normale de 22.3. La ‘normal p-value (2-sided)’ étant de 3.389e-110, on doit rejeter – sans surprise - l’hypothèse d’absence de corrélation pour le chômage. La statistique de Moran pour les résidus de ce modèle est de 0.1805 ce qui est plus faible que la corrélation spatiale des données brutes de chômage: les variables explicatives ont elles-mêmes une structure spatiale qui capture une partie des corrélations dans l’espace. Néanmoins, on rejette avec un *p-value* inférieur à 10^{-3} l’absence de structure spatiale des résidus de ce modèle. Cela peut également se voir sur la Figure 6 qui présente la structure des résidus : une corrélation spatiale assez forte peut être observée. On reviendra plus loin sur l’interprétation économique de ces résidus qui est très importante. Mais avant, il apparaît nécessaire du fait de ces corrélations résiduelles d’essayer d’améliorer la méthode d’estimation. Cela justifie le choix d’un modèle spatial, dit le modèle SAR d’ordre 1 qui tient compte de ces interactions spatiales. La méthode et les résultats sont présentés dans les deux sections suivantes.

4.4 Modèle Spatial. Nous utilisons un modèle autorégressif simultané (SAR ; Simultaneous Autoregressive Model, selon Cressie (1993) et Wittle (1954). Soit y_s la variable d’intérêt, nous écrivons

$$y_s = f(x_s) + \delta_s \quad \delta_s = 1, \dots, N.$$

Nous nous limitons à un modèle linéaire, donc $f(x_s) = \beta' x_s$ et voulons estimer le paramètre β . Nous appelons $f(x_s)$ le terme structurel. Pour le terme d’erreur, nous utilisons une forme autorégressive

$$\delta_s = \sum_{r=1}^N B_{sr} \delta_r + \varepsilon_s.$$

Le terme ε_s représente un bruit sans corrélation spatiale. La matrice B a comme propriété que $B_{ss} = 0$ pour $s = 1, \dots, N$. Pour diminuer le nombre de paramètres inconnus, il faut spécifier la forme de B . Pour un modèle SAR (d’ordre 1) nous prenons

$$B = \rho W,$$

où W est la matrice des poids (mais avec $w_{ii}=0$) et ρ est un paramètre à estimer. Si $\rho = 0$, on a absence de corrélation spatiale. En notation matricielle, avec y , δ et $\varepsilon \in \mathfrak{R}^N$, le modèle s’écrit comme $y = X\beta + \delta$ et $\delta = B\delta + \varepsilon$, donc

$$y = X\beta + B(y - X\beta) + \varepsilon \tag{1}$$

Nous voyons que y se décompose en 3 termes, un terme structurel, un signal spatial $B\delta = B(y - X\beta)$ et un bruit ε . Supposons que $\varepsilon \sim N(0, A)$, où A est une matrice diagonale.¹¹ Comme $y = X\beta + (I-B)^{-1} \varepsilon$ on obtient

$$Y \sim N(X\beta, (I-B)^{-1} A (I-B)^{-1}).$$

¹¹ Nous avons supposé l’absence d’hétéroscédasticité et donc choisi $A = \sigma^2 I$.

Nous estimerons β et les éléments inconnus de B et A par maximum de vraisemblance.

4.5 Résultats du modèle linéaire spatial SAR1. La Table 2 donne les résultats du modèle autorégressif spatial. L'erreur standard des résidus est de 0.24, ce qui est plus faible que l'erreur standard des résidus du modèle linéaire simple, qui était de 0.31.

Les coefficients estimés changent quelque peu : celui d'EduA passe de 9.71 à 9.67, le terme quadratique reste à -7.94 . Le coefficient de la variable de taille diminue de façon plus significative, de 0.19 à 0.14. Enfin, celui d'EduB par contre est divisé par deux, et passe de 7.14 à 3.80.

La statistique de Moran pour les résidus est maintenant de -0.013 et la p-value correspondante est de 0.48, ce qui veut dire qu'on ne peut plus rejeter l'hypothèse d'absence de corrélation spatiale des résidus.

4.6 Interprétation des résultats des deux modèles. Les Figures 5 à 9 permettent de comparer respectivement la structure spatiale du chômage en Wallonie (5) et l'écart par rapport aux prédictions du modèle linéaire (6) et du modèle spatial (7 à 9).

On peut faire plusieurs constats. D'abord, les deux modèles ne prédisent pas partout le taux de chômage correct : les zones sombres des Figures 6 et 8 correspondent à un taux de chômage supérieur à ce que le modèle prédit, les zones claires correspondent à un taux inférieur. Commençons par les résidus du modèle linéaire simple (Figure 6). On peut constater que le sud du Hainaut ainsi que la région autour de Liège sont des régions où le taux de chômage est 30 à 50% plus élevé que ce que les taux d'éducation laisseraient présager. En revanche, le Nord du Brabant et l'Est de la Wallonie (la région frontalière) ont un niveau de chômage de 40% plus faible environ par rapport aux prédictions du modèle. Il est intéressant en particulier de constater que ces régions sont frontalières: mais, si la proximité de la Flandre, de Bruxelles ou de l'Allemagne semble avoir un effet bénéfique, tel n'est pas le cas de la proximité de la France. Une interprétation possible est que l'éloignement des centres d'emplois de la Wallonie peut être compensé par le travail transfrontalier vers la Flandre et l'Allemagne, mais pas vers la France. On a même, de façon inattendue, une zone de chômage élevée par rapport à ce que le modèle prédit, au sud de Namur, qui correspond aux deux communes de Philippeville et Viroinval, sans doute une partie enclavée des Ardennes.

Les résidus non-corrélés spatialement du modèle spatial (Figure 9) sont en général un peu plus faibles que pour le modèle linéaire simple (les écarts sont de -40 à $+40\%$ ainsi que l'atteste l'échelle) ce qui indique une amélioration de l'adéquation du modèle. On peut retrouver dans une moindre mesure le même phénomène des régions au chômage plus élevé (Hainaut) et plus faible (régions Est) que ce qui est prédit par le modèle.

Les Figures 7 à 9 permettent d'étudier plus en détail les résultats du modèle spatial. Ces figures regroupent les trois termes définis par l'équation (1) : le terme structurel (c'est à dire la valeur prédite par le modèle), le signal spatial et le bruit, c'est à dire la partie des résidus débarrassée des corrélations spatiales. On peut voir que la densité du terme structurel est assez uniforme, à l'exception de deux zones : la zone blanche au Nord, à proximité de Bruxelles : le chômage y est plus faible car la main d'œuvre y est mieux formée. En revanche, les zones urbaines de Charleroi Namur ont un taux de chômage prédit très élevé en raison de la variable de taille (configuration urbaine) et peut-être de la moindre qualification de la main d'œuvre.

Cette deuxième explication est cependant moins convaincante au vu de l'absence de zone marquée de non-qualification (Figures 4b,c et d).

Le signal spatial (Figure 8) est également intéressant à commenter: il traduit l'impact d'une commune sur l'autre en terme de chômage, c'est à dire la propagation d'une commune à l'autre de déséquilibres du marché de l'emploi. On peut ainsi voir que la zone autour des centres urbains de Charleroi et de Namur contient des termes de spillovers spatiaux positifs très forts: ces termes se propagent en fait à toute la région environnante, sur une zone qui comprend presque la moitié de la Wallonie. Tout se passe comme si le déclin des industries traditionnelles entraînait un sous-emploi structurel qui s'atténue avec l'éloignement. On retrouve le même phénomène, quoique fort atténué, aux environs de Liège. En revanche, la proximité de l'Allemagne (zone claire) et dans une moindre mesure, du Luxembourg, entraîne des spillovers positifs qui réduisent le chômage de près de 40%.

Pour conclure sur l'intérêt de la méthode spatiale, on peut considérer la décomposition de la variance du modèle SAR1. Le terme structurel représente 23% de la variance du chômage. Le terme spatial représente 28.5% de la variance du chômage. Enfin, le bruit représente 37%. La somme est inférieure à 1 car la corrélation entre le terme structurel et le signal spatial est positive, de +0.30. Cette décomposition de la variance montre que d'une part, les termes structurels jouent un rôle important, mais qu'il existe des interactions spatiales qui jouent un rôle comparable, voire supérieur avec notre modèle volontairement simple.

Deux questions connexes peuvent également être soulevées par notre analyse: la première consiste à corrélérer les résidus à la structure sectorielle de la commune, ou en d'autres termes, à résoudre un problème de variable manquante. Ainsi, la forte composante industrielle des régions du Hainaut associée à un déclin structurel pourrait expliquer le chômage supplémentaire non prédit par le modèle. L'autre question, reliée, est celle du manque de diversité de la composition industrielle, qui amplifie les chocs (positifs ou négatifs) d'emplois dans une localité donnée. Une étude sur les villes américaines dans la période d'après-guerre¹² avait ainsi montré que le chômage est positivement corrélé à l'indice de Herfindhal de la composition industrielle. Dans nos recherches futures, nous nous attacherons à relier le taux de chômage à la composition industrielle et à sa diversité, de façon à corroborer ou infirmer cette analyse.

5. Conclusion.

Notre approche a souligné l'intérêt de l'examen de données désagrégées de chômage au niveau des communes. Elle révèle l'existence de fortes corrélations spatiales entre les variables du marché du travail. De plus, il existe des zones touchées par le chômage, à la fois structurellement (de par l'existence de faibles niveaux de qualification) et spatialement (par l'éloignement des centres d'emplois de la Wallonie ou des régions environnantes). Il existe par ailleurs des zones favorisées en terme de taux de chômage, ce sont les zones frontalières, du Luxembourg, de la Flandres et surtout de l'Allemagne.

En tout état de cause, l'existence des gradients de chômage révélés par notre travail souligne la nécessité de la mobilité des travailleurs pour deux raisons: d'une part, lorsque les interactions spatiales sont fortes, la proximité de zones de sous-emplois, notamment des zones industrielles en reconversion nécessiterait une mobilité plus forte des travailleurs. D'autre

¹² De Simon, C.J. et Nardinelli C., *Economic Inquiry*, Avril 1992, citée par Margo (1993).

part, le chômage est plus faible là où les opportunités de mobilité transfrontalière sont plus importantes.

Enfin, l'existence d'écarts de chômage par rapport à ce que prédit le modèle basé sur les qualifications de la force de travail indique qu'il existe divers moyens de traiter le chômage des zones élevées : soit d'augmenter le niveau local de ces qualifications en investissant dans la formation des populations ; soit de traiter les composantes spatiales du chômage, c'est à dire en aidant à la reconversion des zones en difficulté. L'importance relative des composantes spatiales dans la décomposition de la variance du chômage indiquerait que ce type de mesure peut être efficace au moins dans le court-terme.

Cela nous permet de conclure sur le thème de la fracture sociale évoqué en introduction. Nous avons constaté dans notre analyse que l'éloignement des centres d'emplois a un rôle négatif sur le taux de chômage local, mais que cet éloignement peut être compensé par la proximité des frontières allemandes et luxembourgeoises. En revanche, la proximité des zones de Charleroi-Namur a un effet particulièrement important et négatif. Il n'est pas possible dans le cadre de notre contribution d'en approfondir les causes. On laissera au lecteur le soin d'interpréter ces corrélations spatiales comme le résultat d'un manque de mobilité spécifique des travailleurs de cette région, ou comme le manque d'attractivité de la région pour les créateurs d'emplois, dû à un déficit d'image ou à des taux d'imposition trop lourds par rapport au reste de la Belgique.

Références

- Benabou, R. (1993), « Workings of a City: Location, Education and Production », *Quarterly Journal of Economics*, 108, 619-52
- Case, A. C. (1991), « Spatial Patterns in Household Demand », *Econometrica*, 59, 953—965.
- Cliff, A.D., and Ord, J.K. (1981), *Spatial Processes: Models and Applications*. Pion Limited, Londen.(moran)
- Cressie, N.A.C. (1993), *Statistics for Spatial Data*, Wiley, New York.
- Fujita, M., Krugman, P. and Venables, A.-J. (1999). *The spatial economy: Cities, regions, and international trade*, Cambridge and London: MIT Press, 1999, pages xiii, 367.
- Harris J.R. et Todaro, M.P. (1970), « Migration, Unemployment and Development: A Two-Sector Analysis », *The American Economic Review*, Vol. 60, No. 1. (1970), pp. 126-142.
- Kaluzny, S.P., Vega, S.C., Cardoso, T.P., et Shelly, A.A. (1998), *S+Spatial Stats, User's Manuel*, Springer.
- Kelejian H. H., Robinson D. R. (1998), “A Suggested Test for Spatial Autocorrelation and/or Heteroskedasticity and Corresponding Monte Carlo Results”, *Regional Science and Urban Economics*, 28, 389—417.
- Whittle P, 1954, “On stationary processes in the plane”, *Biometrika* 41, 434-449
- Margo R.A. (1993). “Employment and Unemployment in the 1930's” , *The Journal of Economic Perspectives*, Vol 7, Issue 2, pp. 41-59.
- Petrongolo, B. et Wasmer, E. (1999) “Matching et Spillovers Régionaux en Grande-Bretagne et en France”. Dans *Développements Récents en Economie Spatiale: Emploi, Concurrence Spatiale et Dynamiques Régionales*, 1999, Economica, Catin M., Lesueur J.-Y. et Zénou Y., eds.

Annexe A1

Description des variables

Sources des données

Bade de données de Michel Oris, reprenant les données suivantes : Recensement de la Population et des Logements au 1er mars 1981 et 1991, Institut National de Statistique, Base de données statistiques I.N.S., séries chronologiques par commune, Banque de données DB-GEO (DB2), I.N.S. , Service général d'appui policier, Division Appui en matière de politique policière, Statistique de criminalité, 1994 : Statistique criminelle interpolicière intégrée CSII, 1er mai 1995, Kluwer Editorial-SGAP.APSD, Diegem-Bruxelles, s.d. (Manuel des Services de Police, Dossier 1995). Registre National

Catégories de diplôme:

avec DB : ens. Primaire	Edu1
avec DB : ens. Secondaire inférieur	Edu2
avec DB : ens. Secondaire supérieur	Edu3

avec DB : ens. Pédagogique de type court (normal)	Edu4	
avec DB : autre ens. non universitaire (type court)	Edu5	
avec DB : ens. Universitaire et assimilé	Edu6	
avec DB : niveau d'ens. Indéterminé	Edu7	
avec DB : total	Edu8	=1+2+,,,+7
sans DB : avec diplôme ou certif. Étranger	Edu9	
sans DB : sans diplôme ou certif. Étranger	Edu10	
niveau d'instruction inconnu	Edu11	
total général	Edu tot	=8+9+10

Regroupement des diplômes:

EduA = Edu1+Edu2+Edu10

EduB= Edu3

EduC= Edu4+Edu5+Edu6

Annexe A2

Tables

Table 1 : Résultats du modèle linéaire. Variable endogène: log du chômage rapporté à la population active occupée.

	Value	Std. Error
(Intercept)	-8.31	(0.55)
eduB	7.14	(1.03)
eduA	9.71	(2.40)
eduA^2	-7.94	(2.64)
log(pop. active occ.)	0.187	(0.024)

Residual standard error: 0.3119 on 255 degrees of freedom
Multiple R-Squared: 0.4094
F-statistic: 44.2 on 4 and 255 degrees of freedom, the p-value is 0

La statistique de Moran pour les résidus donne:
Correlation = 0.1805
Normal statistic = 13.98
Normal p-value (2-sided) = 1.925e-44

Table 2 : Résultats du modèle spatial SAR1. Variable endogène: log du chômage rapporté à la population active occupée.

	Value	Std. Error
(Intercept)	-6.84	(0.46)
eduB	3.81	(0.91)
eduA	9.67	(1.86)
eduA^2	-7.94	(2.04)
log(pop. active occ.)	0.138	(0.020)

Residual standard error: 0.249075 on 254 degrees of freedom
rho = 0.04975

La statistique de Moran pour les résidus donne:
Correlation = -0.01327
Normal statistic = -0.7139
Normal p-value (2-sided) = 0.4753

Figure 1

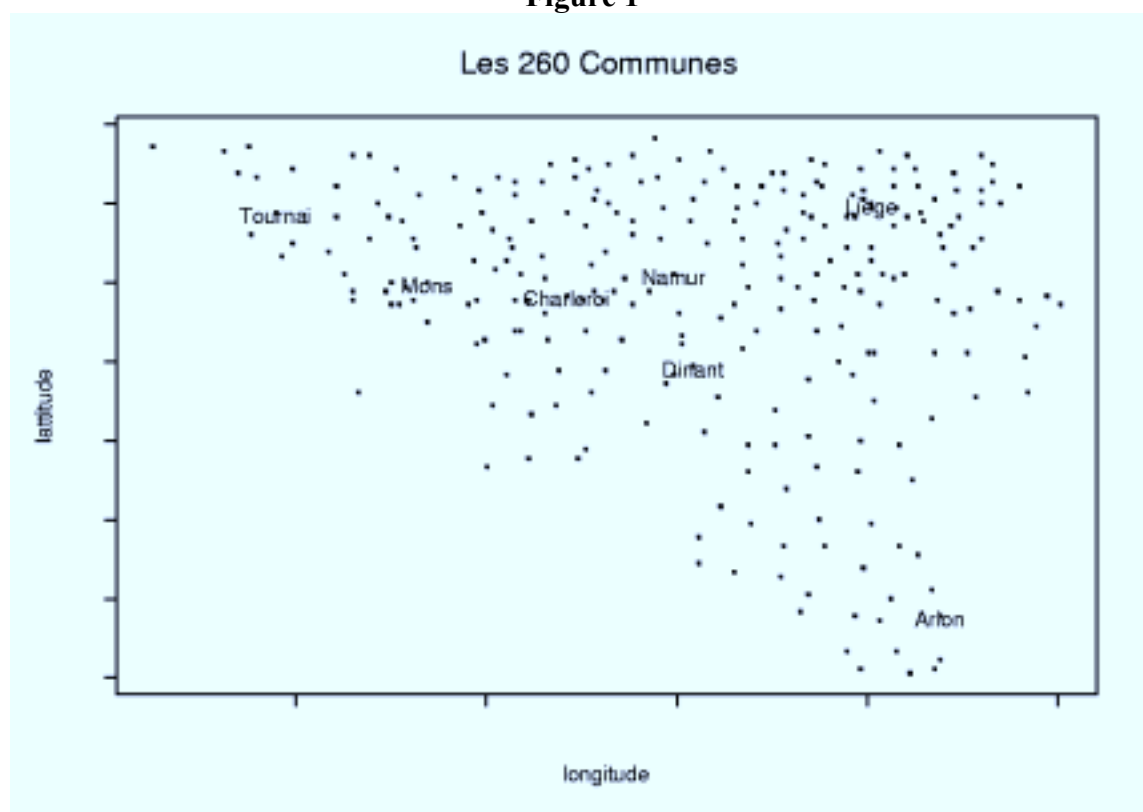


Figure 2

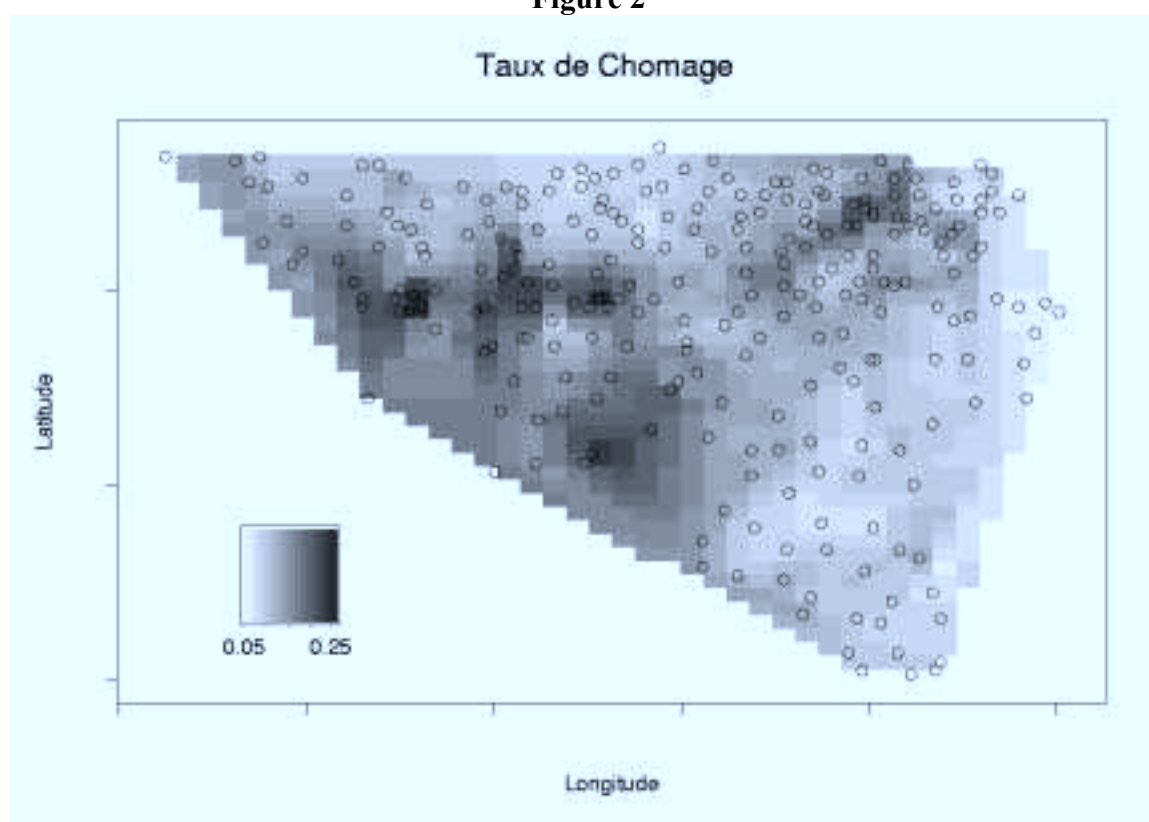


Figure 3

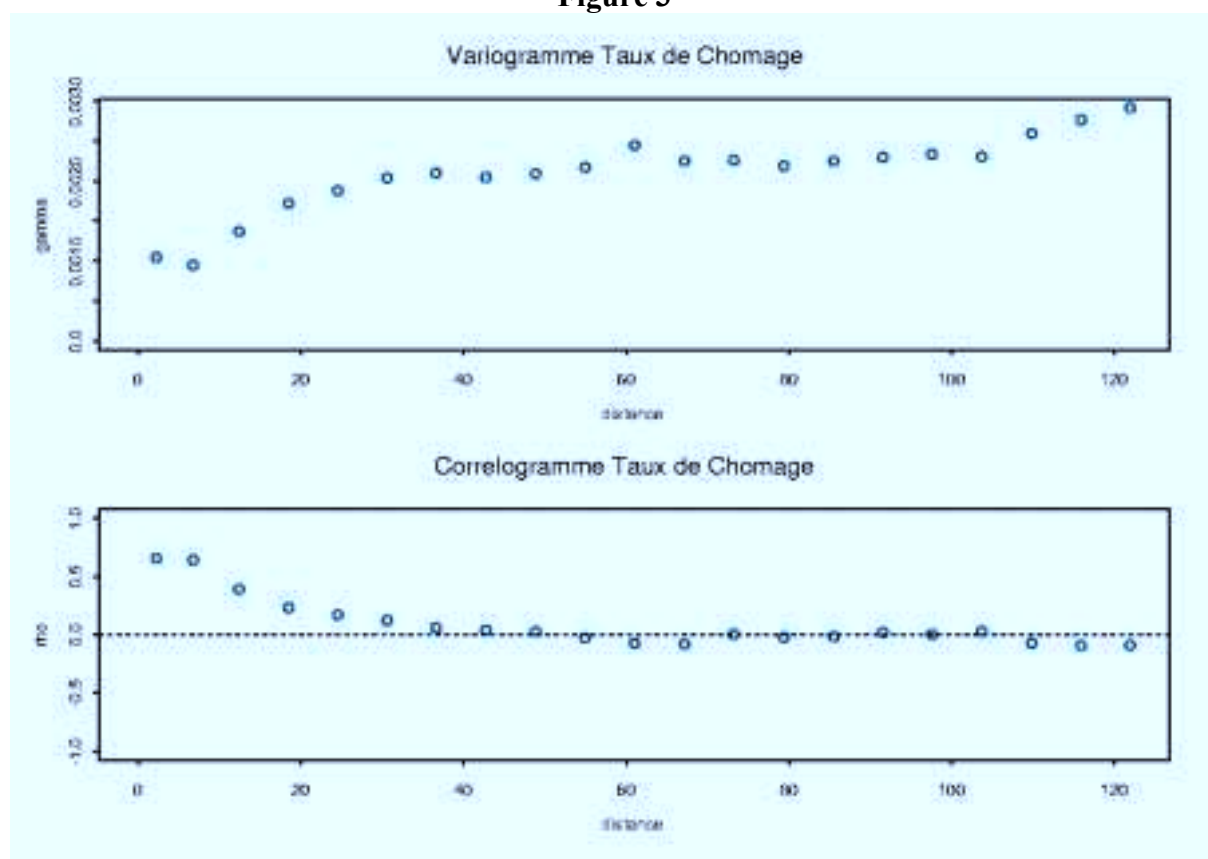


Figure 4

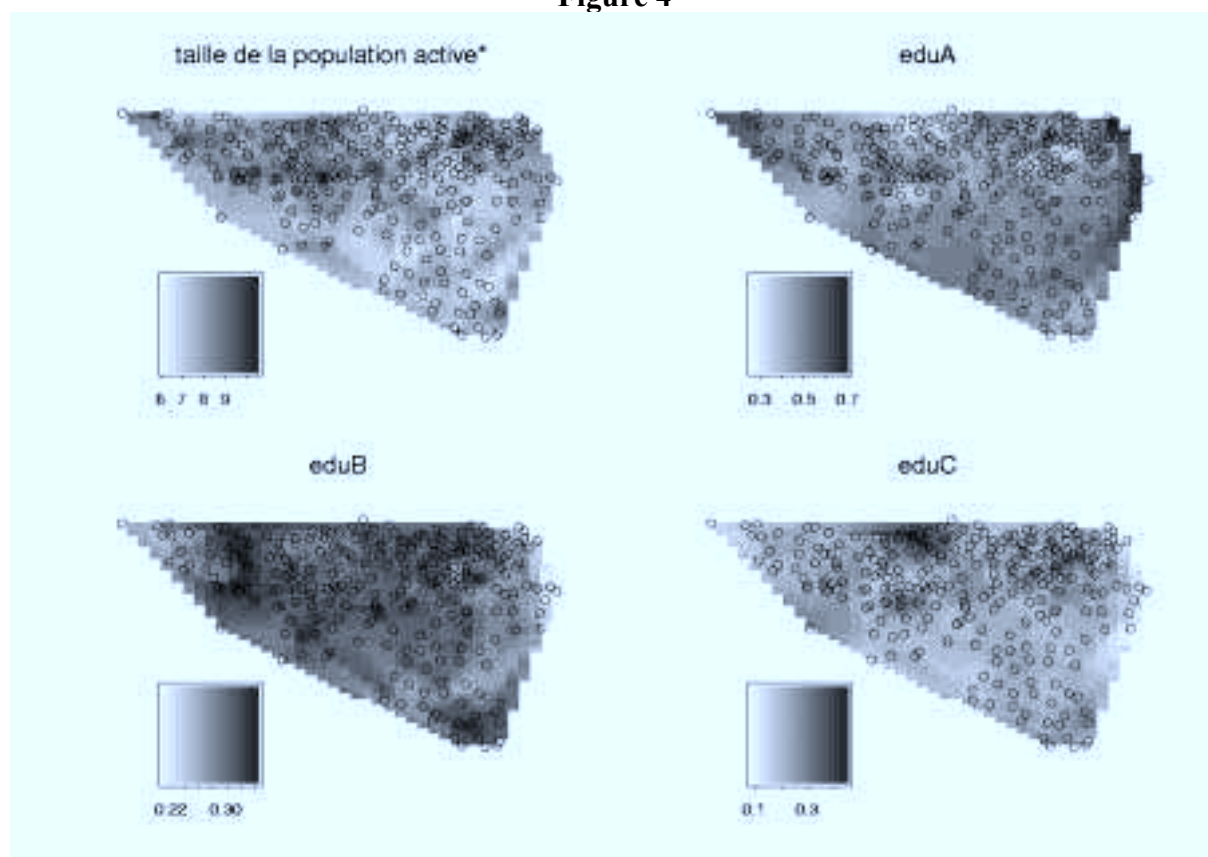


Figure 5

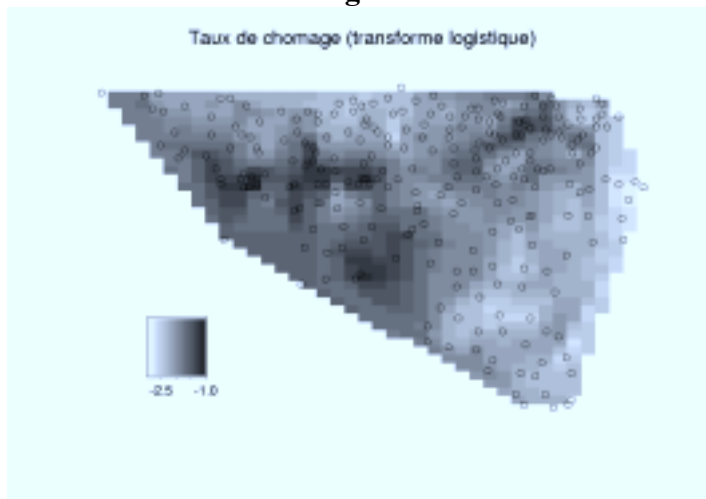


Figure 6

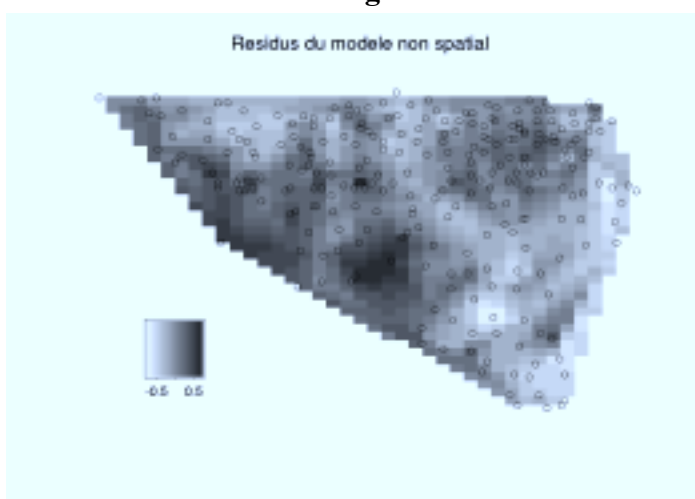


Figure 7

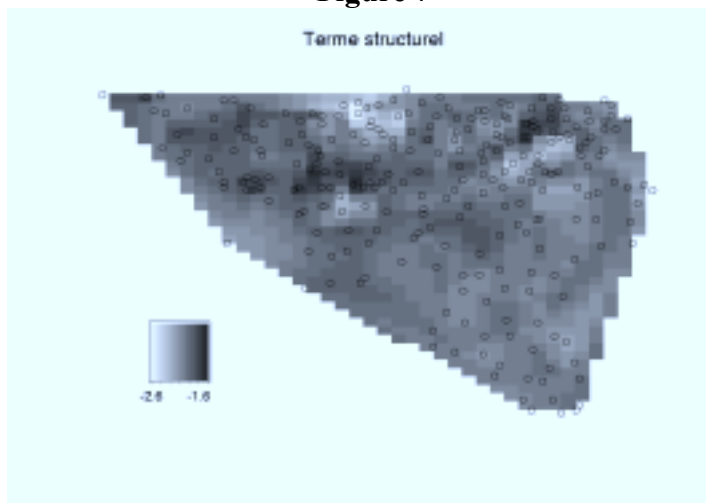


Figure 8

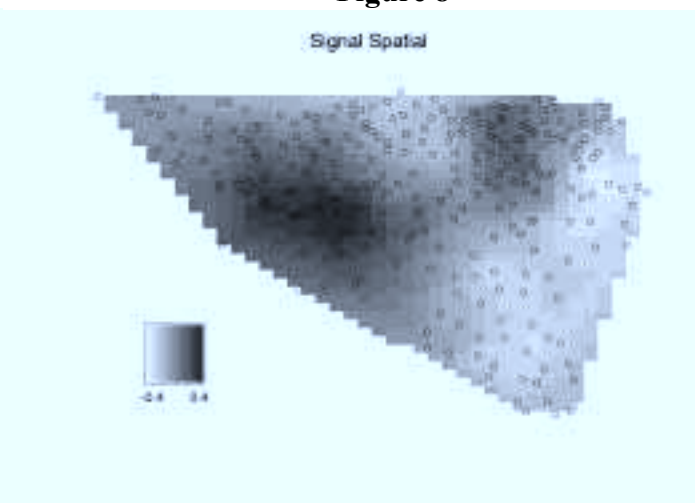
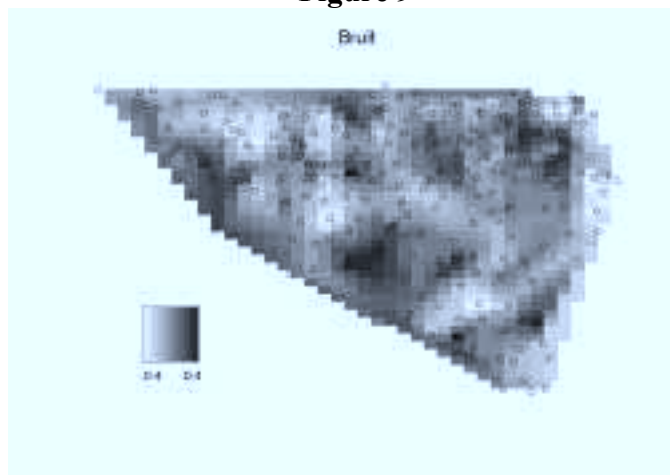


Figure 9



□