JMS Journées de méthodologie statistique de l'Insee 2018

LES INDICES DE CONCENTRATION GÉOGRAPHIQUE A L'ÉPREUVE DE L'AGRÉGATION DES DONNÉES

Emmanuel AUVRAY (*), Salima BOUAYAD AGHA (**)

(*) Gains (Tepp), Le Mans Université
(**) Gains (Tepp), Le Mans Université et Crest

Emmanuel.Auvray@univ-lemans.fr

Mots-clés : Concentration, Agglomération, Statistiques spatiales, Maup

Résumé

Pour caractériser la concentration spatiale des activités économiques il convient de disposer de mesures statistiques fiables afin d'évaluer les disparités existantes et de pouvoir comparer les niveaux de concentration par secteurs dans le temps et dans l'espace. L'espace est continu mais sa discrétisation du fait du regroupement spatial d'observations à des échelles géographiques différentes (communes, départements, régions) peut induire une erreur de mesure (Briant *et al.*, 2010). Comme il n'est pas toujours possible de mobiliser la position exacte des entités, ce travail se propose d'étudier, à partir de données simulées, jusqu'à quel point les indices de concentration géographique des activités peuvent être biaisés par l'agrégation géographique. Nous montrons que les valeurs des indices sont sensibles à l'échelle géographique sur la base desquels ils sont calculés et que certains indices sont plus robustes que d'autres à l'agrégation géographique.

Abstract

To characterize the spatial concentration of economic activities, reliable statistical measures are needed to assess existing disparities and to be able to compare concentration levels by sector in time and space. By nature, the space is continuous but its discretization due to the spatial grouping of observations at different geographical scales (municipalities, departments, regions) can induce a measurement error (Briant *et al.*, 2010), thus affecting the representation of the concentration. Since it is not always possible to mobilize the exact position of the entities, this work proposes to study, from simulated data, the extent to which the indices of geographic concentration of activities the most commonly used can be biased by geographic aggregation. We show that index values are sensitive to the geographic scale on which they are calculated and that some indices are more robust than others to geographic aggregation.

Introduction

Les avantages liés à la concentration géographique des entreprises ont été mis en évidence il y a plus d'un siècle par Marshall (1890). Pour lui, ces bénéfices sont favorisés par la proximité géographique des entreprises d'un même secteur et permettent *i)* de réduire les coûts de transport entre clients et fournisseurs, *ii)* de pouvoir disposer d'une main d'œuvre spécialisée et stable et *iii)* de bénéficier d'un phénomène de diffusion des connaissances (externalités technologiques). Plus tard, les travaux de Porter (1990, 1998) et l'étude des *learning regions* (Florida et Smith, 1995) se focalisent sur les effets de l'innovation et de la coopération active en R&D sur l'agglomération. Ils soulignent l'importance de la formation de clusters, tant en termes d'innovation technologique qu'organisationnelle. Ainsi, la formation de clusters à forte intensité en R&D participe à la spécialisation économique de certains territoires[1].

Les bénéfices directs ou indirects de la proximité géographique des firmes sont appelées économies d'agglomération. Elles peuvent désigner les économies d'échelle internes (propres à l'entreprise), les économies sectorielles (qui bénéficient aux entreprises du secteur même si elles ne sont pas sur le même territoire) et les économies d'urbanisation (qui profitent aux entreprises d'une zone géographique mêmes si elles ne sont pas du même secteur). L'ensemble de ces économies peuvent donc conduire à une spécialisation du territoire.

Que ce soit pour bénéficier des avantages de la proximité, éviter les inconvénients de la dispersion (ou l'inverse selon la logique de localisation d'un secteur), le choix d'implantation est souvent stratégique pour les entreprises et de ce fait, le choix d'une mesure de concentration doit faire l'objet d'une attention particulière pour éviter les erreurs d'interprétation (Marcon et Puech, 2014). D'autant plus que ces indices servent aussi à évaluer la spécialisation des territoires dans le temps (Houdebine, 1999). Si sur le plan théorique, l'explication des phénomènes d'agglomération est relativement bien décrite (Fujita *et al*.,1999 ; Fujita et Thisse, 2002), les travaux empiriques sont moins avancés (Ellison *et al.*, 2010). Pour étudier la localisation des activités et l'intensité de l'agglomération des activités d'un secteur, les recherches en économie spatiale se sont intéressées à la définition de mesures de concentration géographique généralement présentées sous la forme d'indices. Ceux-ci sont d'autant plus importants qu'ils peuvent apporter des informations sur les forces et les faiblesses de l'activité économique locale ce qui représente un enjeu important pour les pouvoirs publics. Les travaux de Briant *et al.* (2010) montrent que ces indices qui reposent sur une discrétisation de l'espace peuvent être biaisés et sont sensibles aux découpages et aux niveaux d'agrégation considérés. Par ailleurs, des travaux récents (Billings et Johnson, 2015 ; Lafourcade et Mion, 2007 ; Barlet *et al.,* 2013) mettent en évidence un biais lié à la taille de l'échantillon (et donc de l'industrie). En effet, lorsqu'un secteur comporte peu d'établissements l'indice calculé peut amener à conclure que le secteur concerné est concentré.

Alors que la décentralisation est au cœur des politiques publiques d'aménagement du territoire, il est donc important de pouvoir déterminer l'échelle géographique la mieux adaptée à l'étude de la concentration ou tout du moins, de vérifier quels sont les indices les plus robustes aux variations d'échelle. On notera que pour neutraliser cet effet, Marcon et Puech (2003) et Duranton et Overman (2005) proposent des indices qui prennent en compte la distance entre les établissements et non plus l'appartenance à une entité géographique. Ces indices qui font l'hypothèse d'un espace continu ne sont pas sensibles, par construction, aux découpages géographiques. Ils sont cependant bien plus coûteux en temps de calcul lorsqu'on mobilise des données à une échelle désagrégée et continuent de présenter un biais négatif lié au nombre d'établissements (Barlet *et al.*, 2013).

Les valeurs des indices calculés à partir de données agrégées à une échelle géographique spécifique dépendent étroitement des découpages considérés. Sous l'acronyme Maup[2] on désigne l'influence du découpage spatial sur les résultats de traitements statistiques ou de modélisation. Plus précisément, les formes irrégulières et les limites des maillages administratifs qui ne reflètent pas nécessairement la réalité des distributions spatiales étudiées sont un obstacle à la comparabilité des unités spatiales inégalement subdivisées.

Selon Openshaw (1984), le Maup est une combinaison de deux problèmes distincts mais proches :

- Le problème de l'échelle (*scale effect*) est lié à une variation de l'information engendrée lorsqu'un jeu d'unités spatiales est agrégé afin de former des unités moins nombreuses et plus grandes pour les besoins d'une

[1] Nous désignerons par concentration l'agglomération d'entreprises d'un même secteur sur quelques territoires. La spécialisation signifie que l'activité d'un territoire est très marquée par un secteur. La concentration d'un secteur sur un territoire ne s'accompagne pas nécessairement de la spécialisation du territoire.

[2] Modifiable Areal Unit Problem (Openshaw et Taylor, 1979 ; Openshaw et Taylor, 1981).

analyse ou pour des questions de disponibilité des données. Dans ce cas, si le nombre de zones est trop faible celles-ci seront trop homogènes tandis qu'un nombre de zones trop élevé augmente le risque de ne pas disposer, sur celles-ci, d'observations sur les variables d'intérêt. Cependant, augmenter ou diminuer le nombre de régions qui découpent le territoire traduit l'effet d'échelle uniquement si les zones font parties d'une logique d'imbrication les unes dans les autres. Si ce n'est pas le cas, cela ne relève pas de l'effet d'échelle du Maup.

- Le problème de l'agrégation (ou de zonage) (*zone effect* ou *shape effect*) est lié à un changement dans la diversité de l'information engendré par les différents schémas possibles d'agrégation à une même échelle. Cet effet est caractéristique des découpages administratifs (particulièrement électoraux) et vient s'ajouter à l'effet d'échelle.

Dans le contexte de l'étude de la concentration géographique, la comparaison des résultats de travaux issus de découpages géographiques différents se heurte au Maup. Les indices de concentration utilisés dans la littérature ne font pas exception aux mesures dépendantes de l'échelle géographique adoptée. En effet, pour un même secteur, la concentration mesurée à l'échelle communale et départementale peut donner lieu à des conclusions différentes.

Afin de souligner l'importance qu'il convient d'accorder à cette question, ce travail se propose d'étudier la sensibilité des indices de concentration les plus couramment utilisés aux deux effets du Maup. L'originalité de cette contribution tient à la distinction et à l'évaluation des deux effets. Tout d'abord, à partir des premiers résultats d'une étude empirique, nous mettons les indices à l'épreuve de l'effet d'échelle. Puis, pour mettre en lumière la sensibilité des indices à l'effet de zonage, nous simulons différentes configurations de localisation en faisant varier systématiquement des frontières de manière à disposer de différents découpages d'un même espace. Ensuite, nous étudions la capacité des mesures standards de concentration à restituer ces schémas préalablement définis. Nous montrons que les résultats sont divergents et que certains indices sont plus robustes que d'autres à la manière dont l'espace considéré est divisé.

Mises bout à bout, ces deux approches nous permettent d'évaluer la sensibilité des indices de concentration au Maup. De plus, nous pouvons également évaluer la sensibilité de l'effet de zonage au nombre de régions qui découpent le territoire.

La seconde section présente une revue des indices de concentration les plus souvent mobilisés pour quantifier les niveaux d'agglomération des activités. La troisième partie analyse le problème d'échelle lié au Maup sur les indices de concentration. Après la présentation de résultats sur données françaises, nous démontrons analytiquement l'effet de l'agrégation géographique sur les indices de concentration. Dans une quatrième section, afin d'étudier le problème de zonage lié au Maup, nous présentons les différents schémas de localisation retenus et le protocole général de simulation des données avant de présenter les résultats de l'analyse de sensibilité. Puis dans une dernière partie nous conclurons sur l'importance du choix de découpage et l'utilisation des indices de concentration.

1. Mesurer la concentration spatiale : une revue des principaux indices

On juge de la qualité d'un indice de concentration selon que celui-ci respecte ou non certaines propriétés (Combes et Overman, 2004 ; Duranton et Overman, 2005). Un indice doit être comparable entre secteurs, être insensible au découpage géographique et à la classification sectorielle, permettre de mesurer significativement les écarts (entre zones, périodes ou secteurs) et prendre en compte l'agglomération géographique de l'activité. La plupart des indices couramment utilisés ne vérifie pas simultanément l'ensemble de ces propriétés.

La littérature sur les indices de concentration considère deux générations d'indice. Ceux de première génération (Gini, 1912 ; Herfindahl, 1950) ne sont pas spécifiques à l'étude de la concentration des activités et sont utilisés en économie géographique alors qu'ils ont initialement été élaborés pour mesurer les inégalités de revenu et la concentration (non géographique) des entreprises sur un marché. Les indices de deuxième génération (Ellison et Glaeser, 1997 ; Maurel et Sédillot, 1999), prennent en compte l'agglomération de l'ensemble des activités pour mesurer la concentration d'un secteur. Ils permettent de vérifier si la concentration des établissements est due à la répartition géographique de l'ensemble des activités sur l'ensemble du territoire. Un secteur est dit concentré si la répartition du secteur sur chaque sous-territoire ne correspond pas au poids des sous-territoires. Ces indices reposent sur une discrétisation prédéfinie du territoire : l'espace analysé est divisé en plusieurs sous-territoires distincts (comme le découpage régional ou départemental). Les indices de concentrations présentés

précédemment ont l'inconvénient d'êtres aspatiaux : ils sont invariants par permutation des zones. Comme ces indices reposent sur le découpage discret du territoire, le positionnement des zones et les interactions que cela peut engendrer ne sont pas pris en compte.

1.1. Indices discrets de première génération

1.1.1. Indice de Gini

L'indice de Gini compare une statistique calculée à partir des données observées sur N régions à sa valeur dans le cas d'une répartition égalitaire parfaite (équirépartition) entre celles-ci. Pour un secteur k donné, on désigne respectivement par $S_k$et S^i la part des effectifs du secteur k et de la région i au niveau agrégé. On note également $s_k^i = \frac{Effectif\ du\ secteur\ k\ dans\ la\ zone\ i}{Effectif\ total\ du\ secteur\ k}$ la part de l'effectif du secteur k dans une région i. On définit r_k^i comme l'écart entre l'effectif constaté (s_k^i) et l'effectif théorique ($S^i \times S_k$), soit :

$$r_k^i = s_k^i - S^i \times S_k$$

Dans le cas où cet écart est nul ($s_k^i = S^i \times S_k$) pour tout $i \in \{1, 2, \dots, N-1, N\}$, la répartition du secteur k correspond à la localisation de l'ensemble des activités au niveau agrégé. Dans le cas contraire ($s_k^i \neq S^i \times S_k$ pour au moins une valeur de $i \in [1, 2, \dots, N-1, N]$), la répartition du secteur k ne correspond pas à la localisation de l'ensemble des activités au niveau agrégé. La courbe de Lorenz associée, notée R_k^i , n'est autre que la fonction de répartition des r_k^i. L'indice de Gini est calculé comme la somme des écarts entre la répartition homogène et la courbe de Lorenz :

$$G_k = 1 - 2 \times \sum_{i=1}^{N} R_k^i$$

Si la répartition observée correspond à une équirépartition, on a alors $G_k = 0$ ($r_k^i = 0\ \forall\ i$). Plus G_k augmente et plus on s'éloigne d'une équirépartition (avec un maximum de $G_k = 1$[3]).

Notons qu'il n'y a pas de valeur précise pour une localisation d'établissement ni concentrée, ni aléatoire. Des valeurs similaires de cet indice peuvent correspondre à des répartitions dans l'espace distinctes[4]. Une variante de cet indice consiste à pondérer la part d'un secteur sur un territoire, par le poids du territoire total (par exemple, on pondère la part d'un secteur sur un département, en tenant compte du poids du département sur le territoire national). On parle alors d'indice de Gini relatif (WGini). Cette variante permet, pour ce travail, de tenir compte dans le calcul de l'importance relative d'une zone donnée.

1.1.2. Indice de Herfindahl

L'indice de Herfindahl permet de comparer la répartition des effectifs dans chacun des secteurs en considérant un découpage selon *N* régions. Si on reprend les mêmes notations que précédemment celui-ci est égal à :

$$H_k = \sum_{i=1}^{N} \left(s_k^i\right)^2$$

Pour $H_k = 1$, l'intégralité des effectifs d'un secteur se situe dans une seule région : il y a donc une concentration parfaite du secteur et tous les établissements se localiseront dans cette région. Lorsque $H_k = 1/N$, les effectifs du secteur k sont répartis de manière homogène entre les N régions, ce qui correspond à une dispersion parfaite des établissements[5]. Dans ce cas, à l'inverse de ce qui se passe dans le cas d'un secteur concentré, la présence d'un établissement du même secteur à un endroit donné réduit les chances qu'un autre établissement du même secteur se localise à un endroit proche. Lorsque la présence d'un établissement d'un secteur donné à un endroit donné n'a pas d'impact sur le choix de localisation d'un établissement du même secteur on admet que la répartition des

[3] On considère parfois que $G_k = 0.5 - \sum_{i=1}^{N} R_k^i$; G_k est alors compris entre 0 et 0.5. Afin de faciliter les comparaisons, nous normaliserons les valeurs pour qu'elles soient bornées entre 0 et 1.

[5] Afin de faciliter les comparaisons, nous normaliserons également les valeurs pour qu'elles soient bornées entre 0 et 1.

établissements sur le territoire est aléatoire. Comme pour l'indice de Gini, on ne peut pas associer une valeur particulière de l'indice à cette configuration spécifique.

1.2. Indices discrets de deuxième génération

Les indices de Herfindahl et de Gini ne sont pas des mesures spécifiquement dédiées à la concentration géographique. Ils ne prennent pas en compte l'agglomération globale de l'activité sur les différents territoires. Les indices de seconde génération sont plus pertinents dans la mesure où ils prennent en compte de manière explicite la localisation des établissements.

1.2.1. Indice d'Ellison et Glaeser

L'indice d'Ellison et Glaeser (1997) repose sur une approche probabiliste. Ils définissent un indice de concentration spatiale à partir d'une distribution aléatoire des établissements, selon une probabilité proportionnelle à la taille des zones géographiques. Les deux valeurs extrêmes de cet indice sont, comme pour l'indice de Herfindahl, l'équirépartition et la concentration totale. Cet indice repose sur un modèle théorique de choix de localisation dans lequel l'agglomération est expliquée par la présence d'avantages naturels (typiquement un meilleur accès au réseau de communication ou la présence d'une matière première à proximité) et/ou d'externalités du fait de la présence d'autres établissements. Le cadre aléatoire proposé par les deux auteurs permet également de tester la significativité des résultats obtenus.

Dans ce modèle, les *M* établissements choisissent séquentiellement leur localisation parmi les sites *N*. Un établissement décide de faire le même choix que l'entreprise précédente ou fait le choix d'une localisation au hasard. À partir de ce modèle, on obtient l'indice d'Ellison et Glaeser qui s'écrit :

$$\gamma_k^{EG} = \frac{\sum_{i=1}^{N}(s_i - x_i)^2 - (1 - \sum_{i=1}^{N} x_i^2)(\sum_{i=1}^{M} z_j^2)}{(1 - \sum_{i=1}^{N} x_i^2)(1 - \sum_{i=1}^{M} z_j^2)}$$

où s_i est la part de l'emploi du secteur k dans la zone i, x_i est la part relative de l'emploi total dans la zone i et les z_j mesurent les tailles des établissements j du secteur k.

Un secteur est dit fortement concentré si la valeur est supérieure à 0.05 et faiblement concentrée si la valeur est inférieure à 0.02. Ces bornes sont néanmoins déterminées de manière *ad hoc* à partir des valeurs des secteurs d'activités qui sont reconnus comme concentrés.

1.2.2. Indice de Maurel et Sédillot

Dans le prolongement de ces travaux, Maurel et Sédillot (1999) proposent un indice modifié à partir d'un modèle de choix séquentiel de localisation. Cet indice s'écrit sous la forme suivante :

$$\gamma_k^{MS} = \frac{\frac{\sum_{i=1}^{N} s_i^2 - \sum_{i=1}^{N} x_i^2}{1 - \sum_{i=1}^{N} x_i^2} - H_{Etab_k}}{1 - H_{Etab_k}}$$

Il repose sur une pondération des externalités de chaque établissement dans chaque zone *i* par la taille des établissements de cette zone.

De manière générale, les indices d'Ellison et Glaeser (EG) et Maurel et Sédillot (MS) vérifient :

$$\gamma_m = \frac{C_m - H}{1 - H} \; m \in \{EG, MS\}$$

où :

$$C_{MS} = \frac{\sum_{i=1}^{N} s_i^2 - \sum_{i=1}^{N} x_i^2}{1 - \sum_{i=1}^{N} x_i^2}$$

On peut aussi réécrire C_m de la manière suivante :

$$C_m = H + \gamma_m(1 - H)\ m \in \{EG, MS\}$$

Si $\gamma_m = 0$, alors la répartition des établissements du secteur correspond à la répartition globale des activités ; γ est « un excès » de concentration pure (G) par rapport à la concentration de la production (*H*). Avec :

$$C_{EG} = \frac{\sum_{i=1}^{N}(s_i - x_i)^2}{1 - \sum_{i=1}^{N} x_i^2}$$

Les deux mesures prennent en compte l'agglomération des firmes, pour corriger la mesure de concentration d'un secteur (suivant le principe qu'il est normal de trouver plus d'entreprises correspondant à un certain critère, dans une zone où il y a plus d'entreprises). Devereux *et al.*, (2004) soulignent qu'il y a une nuance entre le s_i d'EG et de MS. EG corrige de l'agglomération du secteur industriel dans son ensemble, alors que MS corrige de l'agglomération totale de l'emploi. La différence la plus notable entre les deux indices réside dans le calcul de l'écart entre le poids du secteur dans la zone par rapport au poids de la zone. Alors que, EG prend en compte la somme des écarts pour chaque zone géographique i, MS considère ces écarts sur l'ensemble du territoire. De ce fait, l'indice MS ne prend pas en compte les spécificités pour chaque zone. Les bornes qui servent à identifier les secteurs concentrés sont les mêmes que pour l'indice EG, tout en admettant que ces valeurs sont choisies de manière arbitraire.

2. Sensibilité des indices à l'effet d'échelle

2.1. Application sur données françaises

Pour illustrer notre propos nous avons calculé les indices de concentration présentés dans la section précédente afin d'analyser la répartition des activités en France. Les données sont issues du fichier des stocks d'établissements en France au 31 décembre 2014 accessible sur le site de l'Insee[6]. Nous restreignons notre étude à la France métropolitaine (hors Corse) pour garantir la continuité du territoire et permettre ainsi de garder des logiques de localisation et de concentration similaires. Les activités sont localisées à la commune, ce qui nous permet également d'agréger directement les données par arrondissement, département, ancienne région[7], nouvelle région[8]. Nous excluons volontairement les découpages géographiques qui ne recouvrent pas totalement le territoire. Les découpages « Bassin de vie » et « Zone d'emploi » sont également exclus car deux communes d'un même bassin de vie/d'une même zone d'emploi peuvent ne pas appartenir à un même département et ne s'inscrivent pas dans l'imbrication des découpages français. De ce fait, l'effet de l'agrégation pour les découpages « Zones d'emploi » et « Bassin de vie » ne peut se faire qu'en comparaison du découpage communal. Les résultats de l'agrégation géographique sont présentés dans le Tableau 1[9].

Les résultats indiquent que, pour un même indice, l'effet de l'agrégation géographique des données est identique quel que soit le passage du niveau désagrégé au niveau agrégé immédiatement supérieur (« Commune » à « Arrondissement », « Arrondissement » à « Département », etc). On retrouve des résultats identiques quelle que soit la classification NAF (nomenclature d'activités française) considérée. Comme cela est synthétisé dans la Table 1, l'agrégation géographique des données se traduit par une baisse systématique de la valeur des indices de Gini et de Gini pondéré : pour un secteur donné, plus le découpage géographique est à une échelle fine, plus la valeur de ces deux indices augmente. Le résultat obtenu pour l'indice pondéré est tel que la pondération des zones ne corrige pas le biais de l'indice de Gini. À l'inverse, l'agrégation géographique augmente la concentration mesurée par l'indice de Herfindahl et les indices de seconde génération : pour un secteur donné, plus le découpage géographique est à une échelle agrégée et plus le secteur semble concentré.

Les variations dues au découpage géographique étaient attendues car la proximité géographique n'est pas la même selon les secteurs. En revanche, l'effet apparemment systématique de l'agrégation des données ainsi qu'un sens de variation différent selon l'indice n'étaient pas prévus.

[6] Institut National de la Statistique et des Études Économiques.
[7] Régions existantes entre 1970 et 2015.
[8] Régions effectives depuis 2016.
[9] Le détail des moyennes est en annexes.

Pour s'assurer que ce résultat n'est pas exclusivement lié à la nature particulière des données mobilisées, nous étudions dans ce qui suit les propriétés théoriques de l'effet de l'agrégation géographique des données sur chacun des indices de concentrations considérés.

Tableau 1 Effets de l'agrégation géographique sur la moyenne des indices de concentration

Indice	**Agrégation géographique (Commune à Nouvelle région)**
Gini	↘
WGini	↘
Herfindahl	↗
EG	↗
MS	↗

2.2. Approche analytique

Pour chacune des cinq mesures précédentes nous étudions le biais potentiel engendré par l'agrégation géographique des données et ce que cela peut avoir comme conséquence sur la validité des comparaisons selon les découpages géographiques considérés. Pour cela nous comparons les expressions des mesures pour un découpage géographique en N régions et en K secteurs d'activité avec un découpage géographique en $N-1$ régions

2.2.1. Indice de Gini

La difficulté pour cet indice vient du fait que l'agrégation peut potentiellement modifier l'ordre dans lequel les régions sont classées selon leur taille croissante. Si l'on considère trois régions, l'agrégation des deux plus petites d'entre elles peut être telle que la région nouvellement formée soit ou non celle qui a la plus grande taille. Nous avons testé les différentes configurations d'agrégation envisageables en considérant trois, puis quatre régions. L'indice de Gini du secteur k est :

$$G_k(N,K) = 1 - 2 \times \sum_{i=1}^{N} (R_k^i)$$

que l'on peut également écrire :

$$G_k(N,K) = \frac{2\sum_{i=1}^{N} iE_i^k}{N\sum_{i=1}^{N} E_i^k} - \frac{N+1}{N}$$

avec E_i^k le nombre d'emplois du secteur k dans la région i.

L'objectif est de savoir si $G_k(N,K) > G_k(N-1,K)$. Ce qui revient à vérifier si :

$$\sum_{i=1}^{N} (2(N-1)i+1)E_i^k > 2N \sum_{i'=1}^{N-1} E_{i'}^k$$

avec $E_{i'}^k$ le nombre d'emplois du secteur k dans la région i' issue du découpage géographique agrégé[10].

[10] De manière générale, en posant Z le nombre de régions supprimées suite à l'agrégation géographique, la condition s'écrit :

$$\sum_{i=1}^{N} (2(N-Z)i+1)E_i^k > 2N \sum_{i'=1}^{N-Z} E_{i'}^k$$

En procédant aux différentes agrégations géographique possibles avec deux, trois puis quatre régions, on montre que l'indice de Gini diminue suite à l'agrégation géographique si :

- On agrège l'ensemble des régions.
- Le processus d'agrégation implique un nombre important de régions.
- L'agrégation n'implique pas une homogénéisation de l'activité dans les différentes régions.
- La différence de taille, avant la fusion, entre la région la plus grande et la région la petite est élevée.

2.2.2. Indice de Herfindahl

Pour $N = 3$ régions et $K = 3$ secteurs d'activité, l'indice de Herfindahl du secteur k s'écrit :

$$H_k(N,K) = \sum_{i=1}^{N} {s_k^i}^2 = \sum_{i=1}^{N} \left(\frac{E_i^k}{E^k}\right)^2 = \frac{1}{(E^k)^2}\sum_{i=1}^{N}\left(E_i^k\right)^2$$

$$H_k(3,3) = \frac{\left(E_1^k\right)^2 + \left(E_2^k\right)^2 + \left(E_3^k\right)^2}{(E^k)^2}$$

On agrège géographiquement les zones 2 et 3 pour former la nouvelle région 2.

$$H_k(2,3) = \frac{\left(E_1^k\right)^2}{(E^k)^2} + \frac{\left(E_2^k + E_3^k\right)^2}{(E^k)^2} = H_k(3,3) + \frac{2E_2^k E_3^k}{(E^k)^2}$$

$$H_k(N',K) = \frac{1}{(E^k)^2}\sum_{i=1}^{N}\left(E_i^k\right)^2 + \frac{1}{(E^k)^2}\sum_{i,j=N'}^{N} E_i^k E_j^k$$

avec $i \neq j$.

$$H_k(N',K) = H_k(N,K) + \frac{1}{(E^k)^2}\sum_{i,j=N'}^{N} E_i^k E_j^k$$

avec $i \neq j$.

Comme les effectifs sont supérieurs ou égaux à 0 :

$$H_k(N',K) \geq H_k(N,K)$$

L'agrégation géographique ne peut donc pas diminuer la valeur de l'indice de Herfindahl. Plus précisément, l'agrégation géographique augmente la valeur de l'indice à l'exception des cas pour lesquels les effectifs du secteur dans au moins $N-1$ des N régions agrégées sont égaux à 0. Lorsque l'agrégation géographique ne modifie pas l'indice de Herfindahl cela signifie qu'au plus une des régions agrégées comporte le secteur considéré. De fait, plus le secteur est présent dans chacune des régions avant l'agrégation, plus l'indice de Herfindahl augmente du fait du regroupement.

2.2.3. Indice de Maurel et Sédillot

On part de l'expression simplifiée :

$$MS_k(N,K) = \frac{C_{MS_k(N,K)} - H_{Etab_k}}{1 - H_{Etab_k}}$$

avec

$$C_{MS_k(N,K)} = \frac{\sum_{i=1}^{N} s_i^2 - \sum_{i=1}^{N} x_1^2}{1 - \sum_{i=1}^{N} x_1^2} = \frac{H_k(N) - H_T(N)}{1 - H_T(N)}$$

où

$$H_T(N) = \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} \left(\frac{E_k}{E}\right)^2$$

L'agrégation géographique diminue la valeur de l'indice si $MS_k(N', K) < MS_k(N, K)$. Comme H_{Etab_k} n'est pas modifié par l'agrégation géographique des données, on a donc $C_{MS_k(N',K)} < C_{MS_k(N,K)}$ si :

$$\frac{\sum_{i,j=N'}^{N} E_i^k \, E_j^k}{\sum_{i=1}^{N'-1} E_i^k \, E_j^k} < \frac{\sum_{i,j=N'}^{N} E_i \, E_j}{\sum_{i=1}^{N'-1} E_i E_j}$$

Cela revient à conclure que dans le cas où l'agrégation géographique augmente la concentration de l'ensemble des activités proportionnellement plus que la concentration du secteur k, alors la valeur de l'indice diminue. Inversement, si proportionnellement l'agrégation géographique augmente la concentration de l'ensemble de l'activité moins que la concentration du secteur *k*, alors la valeur de l'indice augmente. On ne peut donc pas conclure de manière systématique à l'impact de l'agrégation géographique sur le sens de variation de l'indice MS.

2.2.4. Indice d'Ellison et Glaeser

$$EG_k(N, K) = \frac{\frac{\sum_{i=1}^{N}(s_i^2 - x_i^2)}{1 - \sum_{i=1}^{N} x_i^2} - H_{Etab_k}}{1 - H_{Etab_k}}$$

$$EG_k(N, K) = \frac{\frac{H_k(N) - H_T(N) - \frac{2}{E^k E}\sum_{i=1}^{N} E_i^k E_i}{1 - H_T(N)} - H_{Etab_k}}{1 - H_{Etab_k}}$$

EG augmente suite à l'agrégation géographique des données si :

$$\frac{\sum_{i,j=N'}^{N} E_i^k E_j^k}{(E^k)^2} + \frac{\sum_{i,j=N'}^{N} E_i E_j}{(E)^2} - \frac{2\sum_{i,j=N'}^{N} E_i^k E_j}{E^k E} > 0$$

avec $i \neq j$. Or, si $N = 3$, $N' = 2$ et $K = 3$, EG augmente si :

$$(E)^2 2E_2^k E_3^k + 2\left(E^k\right)^2 E_2 E_3 - 2E_2^k E_3 - 2E_3^k E_2 > 0$$

Comme $E^k = \sum_{i,j=1}^{N} E_i^k$ et $E = \sum_{i,j=1}^{N} E_i$ alors $(E)^2 2E_2^k E_3^k > 2E_2^k E_3 + 2E_3^k E_2$ et l'agrégation géographique augmente nécessairement la valeur de l'indice EG pour un secteur donné.

En conclusion, les résultats peuvent être synthétisés ci-dessous (*cf*. Tableau 2).

Tableau 2 Effets de l'agrégation géographique sur les indices de concentration

	Effet empirique	Effet théorique
Gini	↘	↘ sous conditions
Herfindahl	↗	↗
MS	↗	↗ si la concentration du secteur augmente plus que la concentration de l'ensemble des activités
EG	↗	↗

3. Sensibilité des indices à l'effet de zonage

3.1. Protocole général

Afin d'étudier l'impact de l'effet de zonage du Maup sur les indices de concentration, nous simulons une distribution de points représentant les coordonnées des localisations d'établissements de plusieurs secteurs, puis à partir de cette répartition, nous simulons une division du territoire pour pouvoir calculer les valeurs des indices de concentration.

Plus précisément, nous prenons en compte 2 000 établissements (localisations) se répartissant en cinq secteurs désignés par A, B, C, D et E. Cette répartition se fait en deux temps :

Dans un premier temps, les 1 000 établissements des secteurs A, B, C et D sont répartis selon deux schémas, suffisants pour mettre en lumière la sensibilité des indices, et tels que le nombre d'établissement présents dans chaque secteur soit identique dans les deux cas (*cf*. Tableau 3[11]) :

1. Répartition contrôlée : répartition homogène des établissements et affectation sectorielle arbitraire.
2. Répartitions semi-contrôlées : répartitions homogène ou concentrée de chaque secteur A, B, C et D.

Dans un second temps, 1 000 autres établissements du secteur E sont répartis de manière homogène sur l'ensemble du territoire et ce quel que soit le schéma de localisation retenu pour les quatre autres secteurs. Dans le cas de la répartition contrôlée, la prise en compte de ce secteur E permet d'éviter qu'il n'y ait qu'un seul type de secteur dans une zone géographique et d'obtenir ainsi une répartition des établissements plus réaliste. Pour les répartitions semi-contrôlées, les indices calculés pour le secteur E servent de *benchmark* pour étudier comment les logiques de localisation des autres secteurs peuvent avoir une influence et quelle en est la nature.

Tableau 3 Nombre d'établissements par secteur

Secteur A	Secteur B	Secteur C	Secteur D	Secteur E
271	231	492	6	1 000

Afin d'en étudier la sensibilité à l'effet de zonage du Maup, les indices de concentration sont calculés à partir d'une variation systématique d'un territoire donné (par application de frontières différentes) sur lequel la localisation ou la logique de localisation d'un nombre fixe d'établissements est contrôlée. La définition *a priori* de la localisation des secteurs selon des schémas préétablis nous permet de maitriser les valeurs attendues des indices présentés plus haut tandis que, l'application de frontières variables nous permet d'étudier l'impact de ces découpages sur la variation des valeurs de ces mêmes indices.

Nous utiliserons les deux stratégies suivantes pour diviser le territoire :

[11] Ces effectifs sont issus de la simulation de la répartition contrôlée.

1. Une division en deux zones pour laquelle nous créons 17 820 découpages différents en faisant varier la frontière selon son inclinaison et les aires des zones.
2. Une division allant de deux à cinquante zones en appliquant un diagramme de Voronoï[12]. Pour une même répartition et à nombre de régions donné, nous effectuons 100 découpages géographiques. Ainsi nos résultats ne dépendent pas d'un découpage unique. Cette méthode conduit à découper le territoire 4 900 fois pour une même répartition d'établissements.

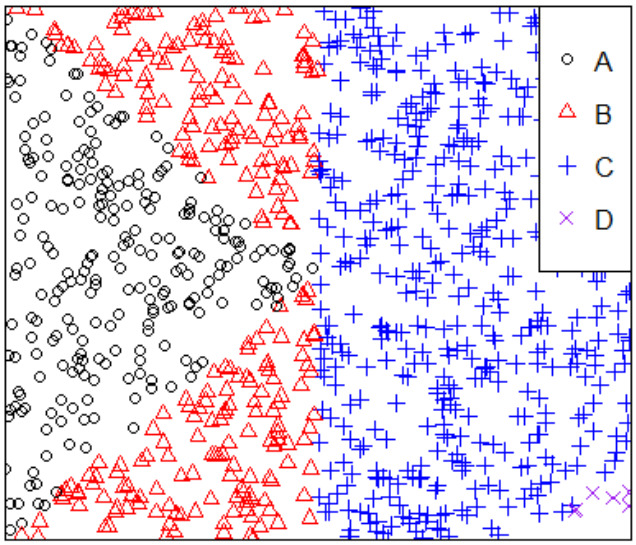
En procédant ainsi, nous pouvons évaluer la sensibilité de l'effet de zonage sur les indices de concentration mais également étudier la sensibilité de cet effet au nombre de régions qui découpent le territoire.

3.1.1. Répartition contrôlée

Nous répartissons 1 000 établissements de manière homogène sur le territoire qui sont affectés à chacun des secteurs A, B, C et D à partir de leurs coordonnées géographiques. Le secteur D se caractérise par une concentration d'établissements sur une petite aire. Les établissements du secteur B sont localisés de chaque côté de la zone où sont situés les établissements du secteur A, tandis que le secteur C est réparti sur le territoire restant. Ce schéma est représenté par la Figure 1.

Avec cette répartition des activités, les établissements du secteur E sont répartis de manière homogène sur un territoire où l'ensemble des autres activités est réparti de manière homogène. La variation d'un indice de concentration pour ce secteur doit donc être faible et dans le cas où la valeur de l'indice est sensible à la frontière, on dira qu'il est sensible au Maup.

Figure 1. Répartition contrôlée[13] des 1000 établissements des secteurs A, B, C et D.



3.1.2. Répartitions semi-contrôlées

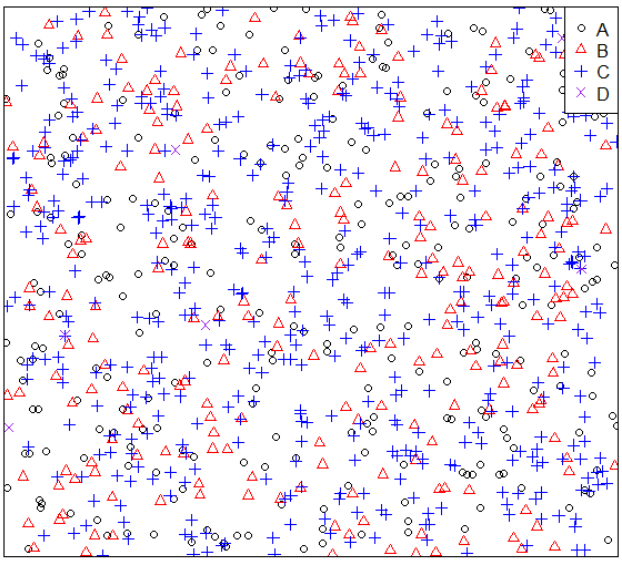
Pour que nos résultats ne dépendent pas du schéma contrôlé, nous modifions la répartition des établissements. En conservant les effectifs de chaque secteur (*cf*. Table 3) et la localisation des établissements du secteur E, nous modifions la répartition des établissements des secteurs A, B, C et D. Ceux-ci peuvent être localisés de manière homogène ou concentrée[14]. Cela revient à considérer seize combinaisons possibles de logique de localisation pour lesquelles nous simulons une répartition d'établissements.

[12] Représente une décomposition particulière d'un espace métrique déterminée par les distances à un ensemble discret d'objets de l'espace, en général un ensemble discret de points

[13] Répartition homogène et appartenance sectorielle arbitraire.

[14] Pour simuler une répartition concentrée, nous prenons un processus de Matèrn conduisant à une localisation des établissements en un ou plusieurs pôles.

Figure 2. Répartition semi-contrôlée des 1 000 établissements des secteurs A, B, C et D.



3.1.3. Découpages géographiques en deux zones

Nous privilégions une analyse graphique à une analyse de corrélation en raison des problèmes liés aux corrélations de Pearson et de Spearman (Hauke et Kossowski, 2011). Nous indiquons également la moyenne et l'écart type de chaque indice, par secteur et par répartition.

Le territoire considéré est un carré de côté égal à 1 de sorte que les localisations des établissements varient entre 0 et 1. Cet espace est divisé en deux zones par une frontière linéaire, ce qui permet en modifiant systématiquement son inclinaison de faire varier les aires des zones géographiques prises en compte. Plus précisément, on considère un ensemble d'inclinaisons de la frontière allant de 0° à 179° avec un pas de variation de 1°. On obtient un ensemble d'aires allant de 0.01 à 0.99 avec un pas de variation de 0.01. Cela revient à considérer 17 820 frontières différentes délimitant à chaque fois deux surfaces complémentaires sur l'ensemble de l'espace considéré. La construction de cet ensemble de couples de surfaces nous permet de vérifier, à aire égale, la conséquence du choix de la frontière sur les variations des valeurs des indices de concentration, pour chacune des répartitions décrites précédemment. L'effet de la taille et de la forme des zones géographiques est synthétisé par un ensemble de figures en trois dimensions retraçant la forme de la distribution des valeurs des indices selon le degré d'inclinaison de la frontière et la surface des zones. L'effet « forme » du Maup sera mis en évidence si la valeur de l'indice est sensible au degré d'inclinaison de la frontière pour des tailles de zones données.

En guise d'exemple de lecture de ces graphiques, la Figure 3 présente différentes frontières associées à trois surfaces (0.01, 0.5 et 0.99) et trois inclinaisons (0, 90 et 179 degrés). Ces différents points sont représentés dans la Figure 4 où l'axe « Aire » correspond à la surface entre la courbe et la droite d'équation x=1 (soit la frontière Est du cadre) et l'axe « Degré » correspond à l'inclinaison de la frontière. Le Tableau 4 synthétise ce à quoi correspondent les neufs points représentés dans la Figure 3[15].

[15] La valeur de l'indice de concentration au point 1 de la Figure 4 correspond à la frontière pour laquelle l'aire sous la courbe est de 0.01 et le degré d'inclinaison est de 0° (*cf.* Table 4). La frontière correspondante est la droite — représentée dans la Figure 3.

Tableau 4 Configurations type de frontière et d'inclinaison

	Degré = 0	Degré = 90	Degré = 179
Aire = 0.01	1 ————	2 ————	3 ————
Aire = 0.5	4 — — —	5 — — —	6 — — —
Aire = 0.99	7 ·····	8 ·····	9 ·····

Figure 3. Exemples de frontières

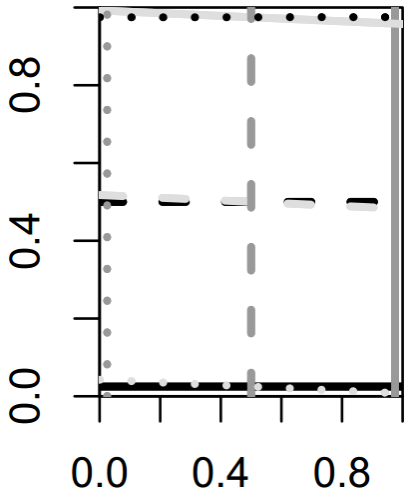
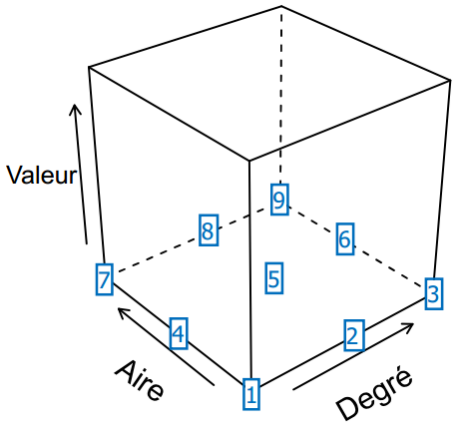


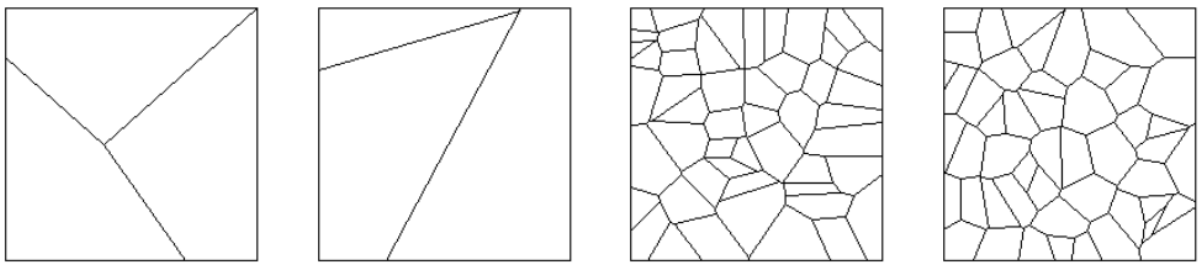
Figure 4. Lecture



3.1.4. Découpages de Voronoï

Afin de vérifier si les résultats précédents ne sont pas spécifiquement liés à un découpage en deux zones, nous menons le même travail pour des découpages géographiques allant jusqu'à cinquante zones en appliquant un diagramme de Voronoï (*cf*. Figure 5) et en calculant les indices de concentration pour chaque secteur.

Figure 5. Exemples de diagrammes de Voronoï en trois et cinquante zones



3.2. Résultats

Le Tableau 5 présente la moyenne et l'écart-type de la distribution de chacun des indices par secteur. Comme attendu, le secteur D (par construction) est le plus concentré quel que soit l'indice pris en compte. On constate également que la dispersion de la distribution des indices de seconde génération est la plus importante pour ce secteur. Ces indices semblent donc plus sensibles mais ce résultat est sans doute lié au faible nombre d'établissements qui compose ce secteur.

Par ailleurs, pour l'ensemble des indices, le secteur B (après le secteur E) est toujours celui pour lequel la valeur de la concentration est la plus faible, ce qui est conforme à ce qui est attendu, les établissements de ce secteur étant les seuls à être répartis en deux pôles.

Les établissements du secteur E étant répartis de manière homogène, on s'attend à ce que pour celui-ci les moyennes et les écarts-types de la distribution des indices soient faibles. Cela n'est vérifié que pour l'indice EG (moyenne proche de 0.000 ; écart-type de 0.001). Les indices de seconde génération semblent donc être ceux qui sont les plus conformes aux résultats attendus au vu du schéma de répartition des établissements considéré. De plus, l'indice EG semble plus robuste au Maup car les écarts-types sont plus faibles qu'ils ne le sont pour l'indice MS.

Tableau 5 Répartition contrôlée : Moyenne et écart-type par indice et par secteur.

Indice	Secteur A	Secteur B	Secteur C	Secteur D	Secteur E
Gini	0.727 (0.320)	0.547 (0.361)	0.620 (0.326)	0.955 (0.169)	0.507 (0.288)
WGini	0.200 (0.156)	0.115 (0.010)	0.148 (0.113)	0.385 (0.270)	0.099 (0.047)
Herfindahl	0.631 (0.387)	0.430 (0.400)	0.491 (0.376)	0.940 (0.213)	0.340 (0.299)
EG	0.402 (0.422)	0.125 (0.167)	0.197 (0.204)	2.491 (8.000)	0.000 (0.001)
MS	0.345 (0.779)	0.138 (0.624)	0.190 (0.647)	0.746 (1.861)	0.025 (0.078)

La lecture des graphiques en trois dimensions suggère que pour le secteur E (*cf*. Figure 6) les indices de première génération sont plus sensibles aux surfaces considérées qu'aux formes des zones définies. Quel que soit le degré d'inclinaison de la frontière, la valeur de l'indice de concentration est déterminée à partir de la surface géographique la plus élevée. De plus, pour les indices de Gini et de Herfindahl, les valeurs sont élevées dès lors qu'il y a plus d'établissements du secteur dans une zone géographique. Comme les indices de seconde génération prennent en compte le poids de chaque région pour calculer la concentration, ces indices varient peu selon les différents découpages : ils sont donc les moins sensibles au Maup.

Figure 6 Indices de concentration du Secteur E. Répartition contrôlée

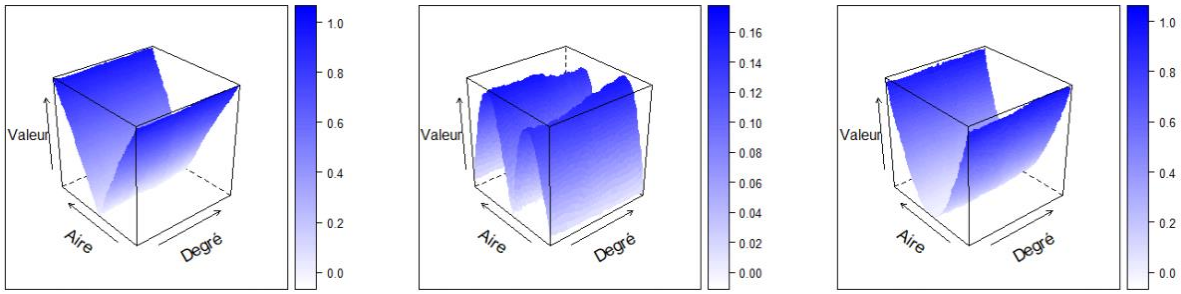


Figure 6-A – Gini Figure 6-B – WGini Figure 6-C – Herfindahl

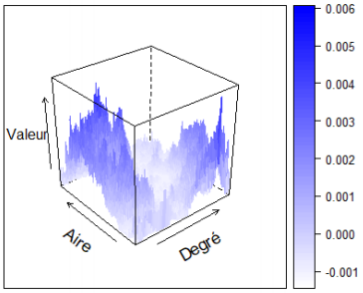


Figure 6-D – Ellison et Glaeser

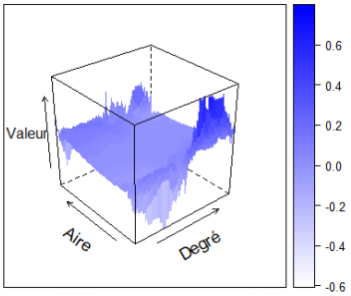


Figure 6-E – Maurel et Sédillot

Pour le secteur D (*cf*. Figure 7), on constate une rupture marquée pour les indices de Gini, Herfindahl et MS. Celle-ci, s'observe à partir d'un seuil pour lequel le dessin de la frontière est tel qu'il y a des établissements du secteur dans les deux zones délimitées par cette frontière. Pour tous les autres découpages pour lesquels les établissements du secteur D ne se localisent que dans une seule zone géographique, la concentration est maximale. L'indice EG prend en compte, l'écart entre la répartition du secteur et la répartition totale, et de ce fait, même si les établissements du secteur D ne se localisent que dans une seule région, la valeur de l'indice de concentration est faible lorsque la part relative des établissements du secteur D est faible. Le fait que les écarts-types de l'indice MS soient relativement plus élevés que ceux de l'indice EG (à l'exception du secteur D) est essentiellement lié aux valeurs associées aux découpages géographiques les plus spécifiques pour lesquels l'aire de la zone la plus petite est inférieure à 5% du territoire.

Figure 7 Indices de concentration : Secteur D. Répartition contrôlée

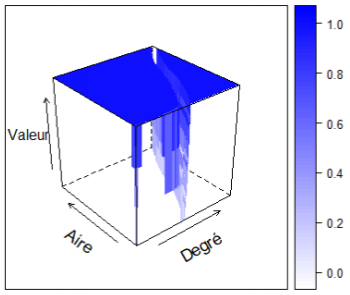


Figure 7-A – Gini

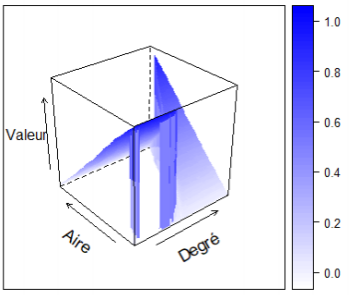


Figure 7-B – WGini

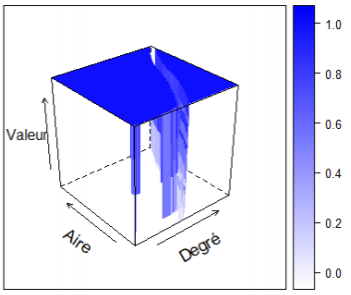


Figure 7-C – Herfindahl

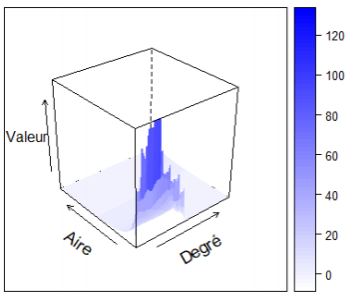


Figure 7-D – Ellison et Glaeser

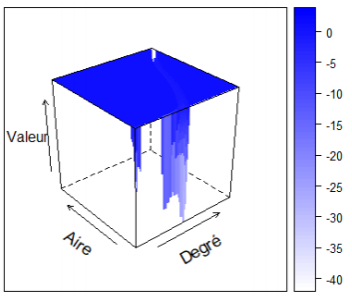


Figure 7-E – Maurel et Sédillot

3.2.1. Variation des découpages géographiques en N zones

Un test non paramétrique de Kolmogorov-Smirnov permet de vérifier si les distributions des valeurs des indices de concentration selon un découpage en N régions et N+1 régions sont significativement proches. Cela représente 48 tests à indice et secteur donnés. Le Tableau 6 présente le pourcentage des cas où l'hypothèse de similarité des deux distributions est rejetée au seuil de 5%.

Tableau 6 Pourcentage de tests refusés de Kolmogorov-Smirnov

Indice	Secteur A	Secteur B	Secteur C	Secteur D	Secteur E	Moyenne
Gini	4	13	4	44	2	13
WGini	19	15	17	8	6	13
Herfindahl	13	13	8	58	10	20
EG	19	13	40	8	13	18
MS	10	6	8	4	4	7
Moyenne	13	12	15	25	7	14

En moyenne, nous constatons que dans 14% des cas il y a une différence significative entre les distributions des valeurs d'un indice de concentration à N et N+1 régions. L'indice MS est celui qui est le moins influencé par le passage de N à N+1 régions (7% de dissimilarité). La représentation des distributions des valeurs des indices selon le nombre de régions prises en compte (Figures 8 et 9) permet de vérifier si les variations de chacun des indices sont proches pour un même secteur. Quel que soit le secteur, c'est pour l'indice MS que la proximité des deux distributions (N et N+1 régions) est la plus forte. On constate, à partir de ce que l'on obtient pour le secteur D, la forte sensibilité de l'indice de Gini et de Herfindahl lorsque le secteur pris en compte comporte un faible nombre d'établissements concentrés en un coin du territoire (respectivement 44% et 58% de dissimilarités). Pour le secteur C qui présente des localisations d'établissements moins spécifiques, on observe pourtant des différences notables entre les indices, la dissimilarité allant de 7% (Gini) à 40% (EG) contre 15% en moyenne pour ce secteur.

Enfin, si l'on admet que le secteur E peut servir de référence, puisque ces établissements sont répartis de manière homogène sur un territoire où l'ensemble des activités est réparti de manière homogène, on peut conclure que, selon ce critère, les indices de Gini, Gini pondéré et MS sont les plus adéquats.

L'indice MS semble donc le plus approprié pour comparer des résultats issus de différents découpages géographiques car il est le moins sensible à la logique de localisation du secteur, et montre le moins de dissimilarité.

Figure 8 Évolution des valeurs de l'indice de Gini et de Herfindahl pour le secteur D

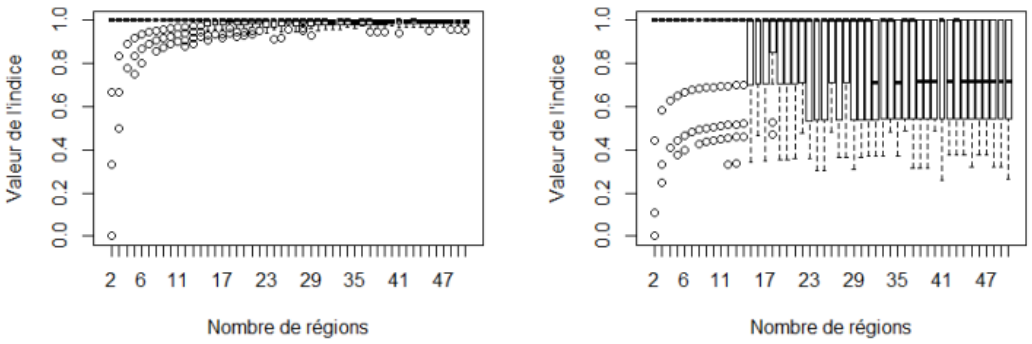
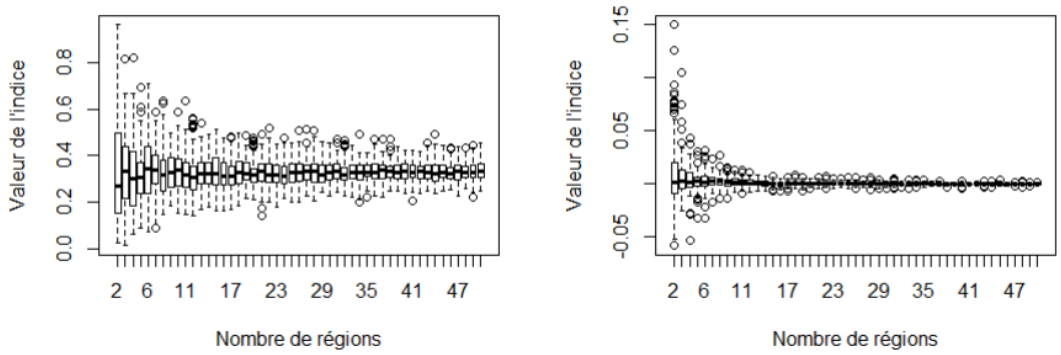


Figure 9 Évolution des valeurs de l'indice de Gini et de Maurel et Sédillot pour le secteur E



3.2.2. Variation de la logique de localisation des établissements

Les résultats précédents peuvent être sensibles à la répartition contrôlée des établissements. Pour tester leur robustesse, nous considérons plusieurs autres répartitions d'établissements (dites semi-contrôlées) pour lesquelles seule la logique de localisation de chaque secteur est contrôlée.

Le Tableau 7, qui synthétise les résultats obtenus quand tous les secteurs sont répartis de manière homogène, montre que les valeurs sont proches pour tous les secteurs à l'exception de D, secteur pour lequel les écarts-types sont plus élevés. Par ailleurs, les valeurs de concentration sont en moyenne plus élevées pour les indices de première génération et plus faibles pour les indices de seconde génération.

Tableau 7 La valeur des indices de concentration si les localisations sont toutes aléatoires

	Secteur A	**Secteur B**	**Secteur C**	**Secteur D**	**Secteur E**
	Aléa.	Aléa.	Aléa.	Aléa.	Aléa.
Gini	0.483	0.493	0.494	0.501	0.499
	(0.287)	(0.293)	(0.286)	(0.317)	(0.289)
WGini	0.094	0.096	0.097	0.108	0.098
	(0.045)	(0.046)	(0.046)	(0.077)	(0.046)
Herfindahl	0.316	0.329	0.326	0.351	0.333
	(0.291)	(0.302)	(0.294)	(0.349)	(0.297)
EG	-0.002	-0.001	-0.001	-0.043	-0.001
	(0.003)	(0.003)	(0.002)	(0.237)	(<0.001)
MS	-0.030	0.010	-0.008	-0.248	0.009
	(0.132)	(0.143)	(0.122)	(1.260)	(0.045)

Le Tableau 8 montre l'impact de la concentration d'un secteur lorsque tous les autres secteurs se répartissent de manière aléatoire. Cela ne modifie pas les conclusions précédentes : la concentration d'un secteur augmente la valeur de l'indice ainsi que son écart-type ; moins il y a d'établissements, plus cette augmentation est importante. On peut donc tester si les distributions des valeurs d'un indice pour un secteur lorsque tous les autres se répartissent de manière aléatoire sont significativement différentes lorsque ce secteur est lui-même réparti de

manière aléatoire ou est concentré. Les résultats du test de Kolmogorov-Smirnov montrent que pour tous les indices les distributions sont significativement différentes au seuil de 1%.

Tableau 8 Impact de la concentration du secteur

Indice	**Secteur A**	**Secteur B**	**Secteur C**	**Secteur D**
	Conc.	Conc.	Conc.	Conc.
Gini	0.719***	0.775***	0.513***	0.790***
	(0.298)	(0.310)	(0.405)	(0.337)
WGini	0.191***	0.176***	0.110***	0.210***
	(0.137)	(0.121)	(0.113)	(0.161)
Herfindahl	0.606***	0.697***	0.427***	0.737***
	(0.367)	(0.380)	(0.422)	(0.389)
EG	0.317***	0.231***	0.184***	0.320***
	(0.325)	(0.209)	(0.174)	(0.523)
MS	0.368***	0.608***	0.180***	0.555***
	(0.668)	(0.494)	(0.661)	(0.722)

*** : Significatif au seuil de 1%

Comme évoqué précédemment, la concentration géographique des établissements d'un secteur donné modifie la répartition globale de l'activité et par conséquent doit modifier les valeurs des indices de concentration de ce secteur mais également celles des autres secteurs. Pour mesurer cet effet, nous considérons le secteur E, réparti de manière homogène quelle que soit la logique de localisation des autres secteurs. Le Tableau 9 nous permet de comparer l'effet de la concentration d'un seul des quatre secteurs sur les mesures de concentration du secteur E. Comme attendu, il n'y a aucun impact sur les valeurs de l'indice de Gini et de Herfindahl puisque ces indices ne tiennent pas compte de la localisation des autres secteurs. Pour les indices de Gini pondéré et EG, la concentration d'un secteur augmente celle du secteur E ; cette augmentation est d'autant plus importante que le nombre d'établissements du secteur qui se concentre est élevé. Il n'y a en revanche pas d'effet systématique sur les valeurs de l'indice MS.

Tableau 9 La valeur des indices de concentration du secteur E selon la logique de localisation des autres secteurs

Indice	Si A, B, C, D aléa.	Si A conc.	Si B conc.	Si C conc.	Si D conc.
Gini	0.499	0.499	0.499	0.499	0.499
	(0.289)	(0.289)	(0.289)	(0.289)	(0.289)
WGini	0.098	0.100***	0.094***	0.105***	0.097***
	(0.046)	(0.050)	(0.046)	(0.055)	(0.045)
Herfindahl	0.333	0.333	0.333	0.333	0.333
	(0.297)	(0.297)	(0.297)	(0.297)	(0.297)
EG	-0.001	0.008***	0.005***	0.018***	-0.001***
	(<0.001)	(0.010)	(0.006)	(0.018)	(0.001)
MS	0.009	0.008***	-0.042***	0.019***	0.015***
	(0.045)	(0.138)	(0.085)	(0.217)	(0.060)

*** : Significatif au seuil de 1%

Conclusion

Le travail précédent illustre l'importance du choix de l'indice adéquat pour rendre compte de la concentration des activités dans l'espace. Nous avons mis en évidence, aussi bien empiriquement qu'analytiquement que l'agrégation géographique entraîne systématiquement une hausse (indice de Herfindhal) ou une baisse (indice de Gini) de la concentration mesurée des indices de première génération. Les indices de seconde génération plus spécifiquement destinés à mesurer la concentration géographique des établissements sont plus adaptés car ils prennent en compte l'agglomération globale de l'activité sur les différents territoires. Nous avons également constaté la pertinence des indices de seconde génération pour prendre en compte les interactions des logiques de localisation de différents secteurs. En faisant varier de manière systématique des frontières partitionnant l'espace sur une même répartition d'établissements, puis en faisant varier le nombre de régions, on constate que l'indice MS est l'indice discret le plus robuste pour restituer les valeurs attendues et celui qui semble être le moins sensible au Maup malgré variation plus importante que l'indice EG lorsque les découpages géographiques sont atypiques.

Du fait de cette sensibilité au MAUP, le choix d'un indice selon le découpage géographique considéré n'est pas anodin lorsqu'il s'agit de comparer des secteurs ou lorsqu'il s'agit de les prendre en compte dans des modèles de choix de localisation. De plus, les indices discrets ne permettent pas de savoir si les secteurs concentrés le sont en un ou plusieurs pôles. Il serait intéressant de prolonger cette étude en utilisant des indices continus.

Annexes

Concentration moyenne des secteurs par indice et par découpage

Indice	Commune	Arrondissement	Département	Région	Nouvelle région	Bassin de vie	Zone d'emploi
Gini	0.96890	0.69676	0.57698	0.51780	0.42725	0.87655	0.73647
WGini	0.85510	0.69544	0.55111	0.46351	0.36826	0.76159	0.71277
Herfindahl	0.01527	0.03722	0.04260	0.09226	0.08451	0.07701	0.05589
EG	0.00590	0.02056	0.02547	0.04507	0.04892	0.02882	0.02699
MS	0.00581	0.02001	0.02513	0.04505	0.04841	0.02951	0.02642

Bibliographie

Barlet M., Briant A., Crusson L., « Location patterns of service industries in France: A distance-based approach », *Regional Science and Urban Economics,* vol 43, n°2, pp 338-351, 2013.

Billings S., Johnson E., « Measuring Agglomeration: Which estimator should we use? », 2015.

Briant A., Combes P.P., Lafourcade M., « Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? », *Journal of Urban Economics,* vol 67, n°3, pp 287-302, 2010.

Combes P.P., Overman, H.G., « The spatial distribution of economic activities in the European Union », *Handbook of regional and urban economics,* vol 4, pp 2845-2909, 2004.

Devereux M.P., Griffith R., Simpson H., « The geographic distribution of production activity in the UK », *Regional science and urban economics* , vol 34, n°5, pp 533-564, 2004.

Di Salvo M., Gadais M,. Roche-Woillez G., « L'estimation de la densité par la méthode du noyau: méthode et outils », note méthodologique et technique. *CERTU,* 2005.

Duranton G., Overman H.G., « Testing for localization using micro-geographic data », *The Review of Economic Studies*, vol 72, n°4, pp 1077-1106, 2005.

Ellison G., Glaeser E.L., « Geographic concentration in US manufacturing industries: a dartboard approach », *Journal of political economy,* vol 105, n°5, pp 889-927, 1997.

Ellison G., Glaeser E.L., Kerr W.R., « What causes industry agglomeration? Evidence from coagglomeration patterns », *The American Economic Review,* vol 100, n°3, pp 1195-1213, 2010.

Florida R., Smith D., « Toward the learning region », *Futures*, vol 27, n°5, pp 527-536, 1995.

Fujita M., Krugman P., Venables A.J., *The spatial economy: cities, regions and international trade*, MIT Press, 1999.

Fujita M., Thisse, J.F., *Economics of agglomeration: cities, industrial location, and regional growth,* Cambridge University Press, 2002.

Gini C., *Variabilità e mutabilità*. Reprinted in Memorie di metodologica statistica, Rome: Libreria Eredi Virgilio Veschi. Pizetti E, Salvemini, T., 1992.

Hauke J., Kossowski T., « Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data », *Quaestiones geographicae*, vol 30, n°2, pp 87-93, 2011.

Herfindahl O.C., *Concentration in the steel industry*, Doctoral dissertation, Columbia University, 1950.

Houdebine M., « Concentration géographique des activités et spécialisation des départements français », *Économie et statistique*, vol 326, n°1, pp 189-204, 1999.

Lafourcade M., Mion G., « Concentration, agglomeration and the size of plants », *Regional Science and Urban Economics*, vol 37, n°1, pp 46-68, 2007.

Marcon E., Puech F., « Evaluating the geographic concentration of industries using distance-based methods », *Journal of Economic Geography,* vol 3, n°4, pp 409-428, 2003.

Marcon E., Puech F., « Measures of the geographic concentration of industries: improving distance-based methods », *Journal of Economic Geography,* vol 10, n°5, pp 745-762, 2010.

Marcon E., Puech F., « Mesures de la concentration spatiale en espace continu: théorie et applications », *Économie et statistique,* vol 474, n°1, pp 105-131, 2014.

Marshall A., *Principles of political economy*, Maxmillan, New York, 1890.

Maurel F., Sédillot B., « A measure of the geographic concentration in French manufacturing industries », *Regional Science and Urban Economics,* vol 29, n°5, pp 575-604, 1999.

Openshaw S., « Ecological fallacies and the analysis of areal census data », *Environment and planning A,* vol 16, n°1, pp 17-31, 1984.

Openshaw S., Taylor P.J., « A million or so correlation coefficients », *Statistical methods in the spatial sciences*, pp 127-144, 1979.

Openshaw S., Taylor P.J., « The modifiable areal unit problem », *In*: N. Wrigley, R. Bennett (Ed.), *Quantitative Geography: a British View.* London: Routledge et Kegan Paul, 1979.

Porter M.E., « The competitive advantage of nations », *Harvard Business Review,* vol 68, n°2, pp 73-93, 1990.

Porter M.E., « Clusters and the new economics of competition », *Harvard Business Review,* vol 76, n°6, pp 77-90, 1998.