

**Direction de la Diffusion et de l'Action Régionale**

**H 2013/01**

**Concentration et spécialisation des  
activités économiques : des outils pour  
analyser les tissus productifs locaux**

Claire Kubrak

**Document de travail**



**Institut National de la Statistique et des Études Économiques**



Institut National de la Statistique et des Études Économiques

*Série des documents de travail  
de la Direction de la Diffusion et de l'Action Régionale*

**E 2013/01**

**Concentration et spécialisation des activités  
économiques :  
des outils pour analyser les tissus productifs locaux**

***Claire Kubrak (Insee, PSAR Études économiques régionales)***

*Mars 2013*

*Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.  
Working papers do not reflect the position of INSEE but only their author's views.*



# Table des matières

Résumé.....	6
Abstract .....	6
<b>I. Introduction .....</b>	<b>7</b>
<b>II. Les indicateurs de concentration et de spécialisation .....</b>	<b>8</b>
<b>II.1. Comment analyser la concentration et la spécialisation économiques des territoires ?</b>	<b>8</b>
<b>II.2. La courbe de Lorenz et l'indice de Gini.....</b>	<b>9</b>
<b>II.3. Les mesures de concentration .....</b>	<b>11</b>
II.3.1 La part des m plus grand établissements.....	11
II.3.2 L'indice de Herfindhal .....	12
II.3.3 L'indice d'entropie de Theil .....	12
II.3.4 L'indice de concentration de Gini.....	13
II.3.5 Le coefficient de répartition.....	13
II.3.6 La densité économique.....	14
<b>II.4. Les mesures de spécialisation et spécificité .....</b>	<b>14</b>
II.4.1 Le coefficient de spécialisation d'une zone .....	14
II.4.2 Les indices de spécificité .....	14
<b>II.5. La mesure d'agglomération .....</b>	<b>17</b>
<b>II.6. Quels indicateurs pour quel type de concentration ?.....</b>	<b>20</b>
<b>II.7. Limites des mesures de répartition spatiale et d'agglomération .....</b>	<b>20</b>
<b>III. Analyse de la structure productive d'une zone d'un territoire.....</b>	<b>21</b>
<b>III.1. Les données « Connaissance locale de l'appareil productif » (CLAP) .....</b>	<b>21</b>
<b>III.2. Les indices de spécialisation .....</b>	<b>21</b>
<b>III.3. Les indices de spécificité.....</b>	<b>24</b>
<b>III.4. Les indices de concentration .....</b>	<b>28</b>
<b>III.5. Le lien entre les indices de concentration et de spécialisation .....</b>	<b>30</b>
<b>III.6. L'indice de densité économique des zones .....</b>	<b>32</b>
<b>IV. Répartition spatiale des activités d'un secteur de l'économie .....</b>	<b>34</b>
<b>IV.1. La concentration productive des établissements .....</b>	<b>34</b>
<b>IV.2. La répartition spatiale des activités .....</b>	<b>36</b>
<b>IV.3. Le lien entre la concentration productive et la répartition spatiale des activités .....</b>	<b>36</b>
<b>IV.4. La mesure d'agglomération .....</b>	<b>38</b>
<b>Bibliographie.....</b>	<b>43</b>

## Résumé

Ce document présente plusieurs indicateurs permettant d'étudier la structure des activités économiques locales en termes de concentration et de spécialisation.

Après une définition des différentes notions de concentration sous-jacentes - concentration géographique, concentration productive, répartition spatiale, spécialisation - il donne une description théorique des indicateurs proposés.

Les deux parties suivantes présentent une illustration concrète de ces indicateurs et de leur interprétation en utilisant l'effectif salarié et les rémunérations des établissements français, disponibles dans le fichier « Connaissance de l'appareil productif » (Clap) de l'année 2010. Ces résultats sont interprétés selon deux approches distinctes : une approche géographique en analysant la structure du tissu productif des zones géographiques (régions, départements, zones d'emploi), et une approche sectorielle avec l'étude de la répartition spatiale des activités d'un secteur donné.

**Mots clefs** : concentration, spécialisation, spécificité, agglomération, densité économique, indice de Gini, région, département, zone d'emploi, secteur d'activité, emploi, effectif salarié, rémunération, CLAP.

## Abstract

This paper presents several indicators for studying the structure of local economic activities in terms of concentration and specialisation. Using a definition of the different underlying measures of concentration : geographic, productive, spatial distribution and specialisation, it gives a theoretical description of the proposed indicators. The two following sections present a worked example of these indicators and their interpretation in using the paid workforce and payments by the French establishments which are available in the file « Connaissance de l'appareil productif » (Clap) from 2010. The results are interpreted using two distinct approaches : a geographical approach by analysing the productive structures in geographical zones (regions, departments, employment zones), and a sectorial approach with a study of the spatial distribution of activities in a given sector.

**Keywords** : concentration, specialisation, clustering, economic density, Gini coefficient, region, department, employment zone, sector, workforce, payments, CLAP.

## I. Introduction

La connaissance des atouts et des faiblesses du système productif est un enjeu important pour les pouvoirs publics locaux. Ils doivent pouvoir évaluer les risques pour leur économie d'un choc sur l'activité économique locale : choc sectoriel ou fermeture d'un gros établissement par exemple. La concentration et la spécialisation des activités d'une zone sont des éléments essentiels pour le diagnostic de la structure du tissu productif. La concentration des emplois dans les établissements permet de mesurer la dépendance de l'emploi local à quelques gros établissements. La spécialisation de l'activité économique dans quelques secteurs renseigne sur les risques d'un choc sectoriel dans le domaine considéré. Toutefois, une concentration ou une spécialisation forte ne sont pas nécessairement signes de vulnérabilité de l'appareil productif. Elles peuvent être bénéfiques tant que les établissements ou les secteurs concernés sont en bonne santé.

L'étude de la concentration et de la spécialisation des activités d'une zone permet de répondre à des questions du type : l'activité de la zone repose-t-elle sur quelques gros établissements ou est-elle bien répartie dans de nombreux établissements ? Quel est le poids des très gros établissements sur l'économie de la zone ? Les activités de la zone sont-elles semblables à celles du reste du territoire ou la zone est-elle spécialisée dans un secteur particulier ?

La question des spécialisations locales, qui montre que les activités économiques ne se répartissent pas de façon uniforme sur le territoire, amène à s'intéresser également aux choix de localisation des entreprises. Plusieurs considérations peuvent être avancées pour expliquer ces choix : les avantages naturels (disponibilité d'une ressource comme le fer en Lorraine ou d'une situation naturelle comme la proximité de la mer pour les zones côtières), les externalités de localisation des entreprises (effets positifs de l'agglomération d'entreprises comme les échanges de connaissances ou de savoir-faire) [voir Ellison et Glaeser].

Cette problématique peut être abordée par la mesure de la concentration des activités d'un secteur dans un nombre plus ou moins important de zones géographiques. On tente ainsi de répondre aux interrogations suivantes : comment les activités économiques sont-elles distribuées géographiquement ? Les établissements d'un secteur sont-ils répartis au hasard sur le territoire ?

Ce document présente différents indicateurs de concentration permettant d'une part d'analyser la structure productive d'une zone en la comparant à celle d'un territoire de référence, d'autre part d'étudier la répartition géographique des activités d'un secteur dans les différentes zones d'un territoire. Ces indicateurs sont tous décrits dans la littérature économique. Ils ont été sélectionnés parmi d'autres indicateurs, pour leur facilité d'interprétation ou les propriétés particulières qu'ils présentent. Au-delà de la description technique des indicateurs exposée en partie II, le document produit en parties III et IV des éléments d'interprétation des différentes mesures de concentration en les mettant en perspective avec des exemples d'utilisation à partir du fichier « Connaissance de l'appareil productif » (Clap) de l'année 2010.

Les indicateurs présentés sont ceux utilisés dans les travaux du PSAR Études économiques régionales de l'Insee pour l'analyse des territoires en termes de concentration et de spécialisation économiques. A ce titre, ce document est le produit d'un travail collectif, initié il y a dix ans par Benoît Mulkay alors responsable du PSAR. Il doit également beaucoup aux contributions de Claire Poisson, Bérénice Costes et Lionel Doisneau. Toutefois, certaines des interprétations sont personnelles et les erreurs qui apparaîtraient ne sont imputables qu'à l'auteur du document.

## II. Les indicateurs de concentration et de spécialisation

### II.1. Comment analyser la concentration et la spécialisation économiques des territoires ?

Les mesures de caractérisation de zones ou de secteurs d'activité qui sont présentées dans ce travail sont purement descriptives. Elles ne visent pas à expliquer une structure d'activité particulière qui peut reposer sur des comportements économiques tout aussi bien que sur l'héritage historique ou géographique de la zone étudiée. Donc il faut être prudent dans une interprétation trop explicative des caractéristiques qui sont mises en évidence par les statistiques présentées.

De même, il faut se garder d'un regard normatif sur les indicateurs produits. Certains observateurs peuvent interpréter une forte concentration ou spécialisation comme des facteurs de vulnérabilité de l'appareil productif. Or, différentes études empiriques ont montré que les effets associés peuvent être très différents suivant le contexte ou la zone étudiée.

Dans tous les cas, ces indicateurs permettent d'établir un diagnostic de la structure productive de la zone étudiée en termes de spécificité par rapport à d'autres zones ou à un territoire englobant.

L'analyse du tissu productif est menée principalement à partir de l'emploi salarié des établissements. Elle est également possible en utilisant les rémunérations, qui sont aussi un indicateur intéressant du poids économique local. Ainsi, à nombre de salariés équivalent, un établissement avec une rémunération globale importante aura sans doute un impact plus fort sur l'économie locale qu'un autre avec une rémunération plus faible. D'autres données économiques pourraient également être pertinentes, comme les emplois y compris non salariés, la valeur ajoutée, les investissements, les exportations, etc. Ces informations ne sont pas disponibles à un niveau finement localisé dans les sources actuellement utilisées.

La notion de concentration repose ici sur la répartition inégale des activités mesurées par les emplois salariés. On peut étudier la concentration des emplois d'une zone géographique ou d'un secteur d'activité. Par ailleurs, on peut considérer la répartition des emplois dans les différents établissements, dans les différentes zones ou encore dans les différents secteurs. On définit alors plusieurs types de concentration :

- la **concentration productive** : répartition des emplois d'un secteur dans un nombre plus ou moins grand d'établissements
- la **concentration géographique** : répartition des emplois d'une zone géographique dans un nombre plus ou moins grand d'établissements
- la **répartition spatiale** : répartition des emplois d'un secteur dans un nombre plus ou moins grand de zones géographiques
- la **spécialisation** : répartition des emplois d'une zone géographique dans un nombre plus ou moins grand de secteurs d'activités.

Les indicateurs peuvent être mis en œuvre sur de nombreux zonages : régions, départements, zones d'emploi, aires urbaines, espaces urbains, cantons, arrondissements, communes, ... Ils sont calculés ici pour les régions, les départements et les zones d'emploi.

De la même manière, la prise en compte des secteurs d'activités peut être réalisée selon de nombreux niveaux d'agrégation de la nomenclature d'activités française : 10 secteurs, 17, 88, etc. Le choix de tel ou tel niveau a évidemment un impact sur la valeur des indicateurs produits. L'analyse sera aussi différente selon les secteurs sélectionnés dans le champ de l'étude : ensemble de l'économie, ou seulement la partie marchande par exemple. Ainsi, les hôpitaux et les collectivités territoriales sont souvent les plus gros employeurs dans des zones relativement petites. La prise en compte du secteur public aura un impact fort sur ces zones.



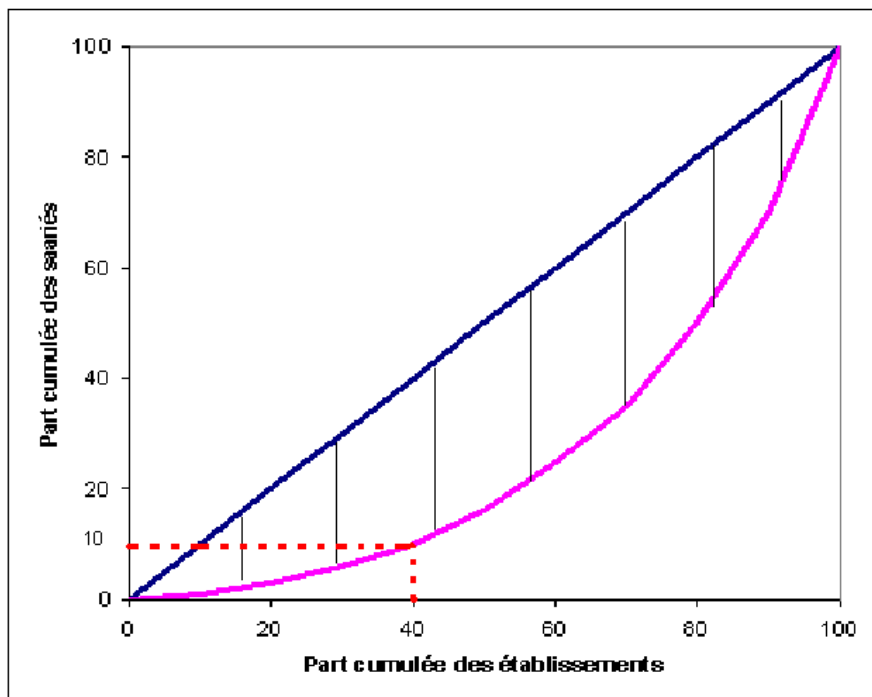
## II.2. La courbe de Lorenz et l'indice de Gini

La courbe de Lorenz est un instrument graphique très utilisé pour décrire la répartition inégale d'une variable entre plusieurs individus. Elle permet de comparer - en abscisse - la distribution des individus à la distribution d'une variable d'intérêt parmi les différents individus - en ordonnée.

Exemple : construction de la courbe de Lorenz pour mesurer la répartition des salariés d'une zone dans les établissements.

On porte en abscisse les n % d'établissements les plus petits ( 10 %, 20 %, ... 100 %) de la zone et en ordonnée la part des effectifs concentrés dans ces n % d'établissements.

*Graphique 1 : La courbe de Lorenz*



Dans la zone étudiée, les 40 % d'établissements les plus petits regroupent seulement 10 % des salariés. Les derniers 20 % concentrent à eux seuls la moitié des salariés.

Cette courbe de Lorenz est inscrite dans un carré (0,1) représentant les proportions d'individus de la population et la proportion de la variable totale. Elle va du point (0,0) au point (1,1) par définition, et aucun point ne se situe au-dessus de la diagonale principale. Cette courbe est croissante, ainsi que sa dérivée première<sup>1</sup>. Si la répartition de la variable est parfaitement uniforme entre les individus, cette courbe de Lorenz se confondra avec la diagonale principale. En revanche, elle s'éloignera de cette diagonale principale pour une distribution inégalitaire de la variable. Si toute la variable considérée est attribuée à un seul individu, la courbe de Lorenz épousera alors à la limite les bords inférieurs et à droite du carré (0,1).

Jusqu'à présent on a présenté la courbe de Lorenz concernant une variable observée pour chaque individu de la population, sous l'hypothèse que ces individus sont identiques. On peut cependant calculer cette courbe dans des situations où les individus n'ont pas les mêmes poids. C'est le cas lorsque l'on ne dispose pas de la variable au niveau individuel, mais à un niveau d'agrégation plus élevé, que celle-ci provienne d'une agrégation sectorielle (données regroupées par secteurs) ou spatiale (données sur un territoire). Dans ce cas, on va donner à chaque classe sectorielle ou géographique (département, zone d'emploi, ...) un poids égal à son importance relative.

<sup>1</sup> Pour des détails concernant cette courbe de Lorenz, on peut se référer à Saporta (1990) ou à Jayet (1993).

Prenons l'exemple de données regroupées au niveau des régions. On donnera à chaque région  $r$  une importance égale à un poids relatif qui est donné par une variable de référence  $z_r/z$ . Dans la courbe de Lorenz, la région  $r$  représente la fraction  $x_r/x$  de la variable d'intérêt. En conséquence, la pente de la courbe sera donnée par l'indice de spécificité de la zone :  $s_r = \frac{x_r/x}{z_r/z} = \frac{x_r/z_r}{x/z}$

Comme la courbe de Lorenz a une pente croissante, les zones seront ordonnées par ordre croissant de spécificité. Les différents points de la courbe seront donnés en abscisse par le cumul des poids relatifs des zones ( $z_r/z$ ), et en ordonnée par le cumul de la part relative de la variable considérée ( $x_r/x$ ).

Pour mesurer la répartition spatiale (dans les différentes zones) de l'emploi d'un secteur (voir § II.3.5) : on aura la part de la zone dans l'emploi total du territoire de référence comme poids relatif de la zone ( $z_r/z$ ) (en abscisse), et la part de la zone dans l'emploi du secteur comme part relative de la variable d'intérêt ( $x_r/x$ ) (en ordonnée).

Inversement, si on cherche à mesurer la concentration de l'emploi d'une zone dans les secteurs d'activité, on pourra avoir en abscisse la part de chaque secteur dans l'emploi total du territoire de référence, et en ordonnée la part de chaque secteur dans l'emploi de la zone considérée (voir § II.4.1).

L'indice de Gini permet de résumer l'information lue sur la courbe de Lorenz. C'est le double de l'aire comprise entre la courbe et la première bissectrice. Ce coefficient a des valeurs comprises entre zéro, lorsqu'il y a une distribution uniforme de la variable entre les différents individus, et l'unité, quand un seul individu dispose de l'intégralité de la variable considérée. Entre ces deux extrêmes, on a une valeur positive qui augmente lorsque la courbe de Lorenz se déplace vers le bas et qu'il y a un accroissement des inégalités.

D'un point de vue pratique, une première manière de calculer est d'ordonner les  $N$  observations par ordre croissant et d'utiliser la formule :

$$Gini = \frac{1}{N(N-1)\bar{x}} \sum_{i=1}^N \sum_{j=i+1}^N (x_j - x_i)$$

où  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  est la moyenne de la variable considérée. Cependant ce calcul est difficile du fait de

la double sommation qui implique un nombre de calculs important : ces calculs et donc le temps sont fonction du carré du nombre d'observations.

Une seconde méthode moins coûteuse en temps, mais plus approximative, consiste à calculer la surface en dessous de la courbe de Lorenz et d'en déduire le coefficient de Gini. L'approximation est d'autant meilleure que le nombre d'individus est important. Elle revient à calculer la surface d'un grand nombre de trapèzes :

$$Gini = 1 - 2 \left[ \sum_{i=1}^N \frac{1}{2N} \left( \frac{x_{i-1}^C}{x} + \frac{x_i^C}{x} \right) \right] = 1 - \frac{1}{N} \left[ \sum_{i=1}^N \left( \frac{x_{i-1}^C}{x} + \frac{x_i^C}{x} \right) \right]$$

$$= 1 + \frac{1}{N} - \frac{2}{N} \sum_{i=1}^N \sum_{j=0}^i \frac{x_j}{x}$$

avec  $x_i^C = \sum_{j=0}^i x_j$ ,  $x_{i-1}^C = \sum_{j=0}^{i-1} x_j = \sum_{j=0}^i x_j - x_i = x_i^C - x_i$ ,  $x = \sum_{j=0}^N x_j = x_N^C$  et  $x_0 = 0$

On peut également adapter cette méthode de calcul pour le cas où les données individuelles auraient été regroupées par régions ou par secteurs, en utilisant le poids relatif de la région ou du secteur.

$$Gini = 1 - 2 \left[ \sum_{r=1}^R \left( \frac{z_r^C}{z} - \frac{z_{r-1}^C}{z} \right) \frac{1}{2} \left( \frac{x_{r-1}^C}{x} + \frac{x_r^C}{x} \right) \right] = 1 - \left[ \sum_{r=1}^R \frac{z_r}{z} \left( \frac{x_{r-1}^C}{x} + \frac{x_r^C}{x} \right) \right]$$

$$= 1 + \sum_{r=1}^R \frac{z_r}{z} \frac{x_r}{x} - 2 \sum_{r=1}^R \frac{z_r}{z} \sum_{s=1}^r \frac{x_s}{x}$$

Dans ce document, cet outil est utilisé pour mesurer 4 indicateurs :

- le coefficient de concentration productive, mesurant la répartition des salariés d'un secteur d'activité dans les établissements
- le coefficient de concentration géographique, mesurant la répartition des salariés d'une zone dans les établissements
- le coefficient de répartition, caractérisant la répartition des salariés d'un secteur d'activité dans un nombre plus ou moins important de zones
- le coefficient de spécialisation, évaluant la répartition des salariés d'une zone dans un nombre plus ou moins important de secteurs d'activité.

### II.3. Les mesures de concentration

Elles visent à résumer la distribution de la taille des établissements en matière d'emploi ou de revenu dans un secteur ou dans une zone.

Les indicateurs qui suivent sont présentés dans la littérature économique. Ils satisfont les conditions de Lorenz : premièrement s'il y a une augmentation de la dispersion de la distribution avec la même moyenne, cela doit accroître l'indice de concentration. La seconde condition établit que si tous les individus sont de même taille, une augmentation du nombre d'individus doit entraîner une diminution de l'indice de concentration.

#### II.3.1 La part des m plus grand établissements

On classe les N établissements par ordre décroissant selon la variable d'intérêt, par exemple l'effectif salarié de l'établissement, de série  $(x_1, \dots, x_N)$ . On note ainsi la taille relative de l'établissement i

mesurée par l'emploi  $\xi_i$  où  $\xi_i = xi / \sum_{i=1}^n xi$  telle que  $\xi_1 \geq \xi_2 \geq \dots \geq \xi_m \geq \dots \geq \xi_N$ .

On définit alors la part des m plus grands établissements par la formule :  $C_m = \sum_{i=1}^m \xi_i$ ,

On s'intéresse ici en particulier à la part des 4 (C4) ou des 10 (C10) plus grands établissements de la distribution. Ces indicateurs permettent de calculer la concentration de la variable d'intérêt dans les 4 ou 10 plus grands établissements de la distribution.

	0 %	Pourcentage faible	Pourcentage élevé	100 %
Pour chaque zone ou chaque secteur	Aucun salarié dans la zone (ou le secteur)			Tous les salariés sont employés dans les m plus grands établissements
	Cas limite			Cas limite
		L'emploi est peu concentré dans la zone (ou le secteur)	L'emploi est concentré dans la zone (ou le secteur)	Concentration maximale dans la zone (ou le secteur).

### II.3.2 L'indice de Herfindhal

La deuxième mesure de concentration, proposée par Herfindhal, représente la somme des carrés des

parts de tous les individus :  $H = \sum_{i=1}^N \xi_i^2$ .

Cet indice de Herfindhal est compris entre  $1/N$  si tous les  $N$  établissements sont de même taille, et 1 lorsqu'un seul établissement regroupe l'ensemble des salariés. On peut interpréter l'inverse de cet indice de Herfindhal comme étant le nombre équivalent d'établissements de taille égale qui se partageraient les emplois de manière uniforme.

Contrairement à C4 ou C10 cet indicateur prend en compte l'ensemble des établissements pour mesurer la concentration. Il est toutefois très sensible aux valeurs élevées des  $\xi_i$ .

	1/N			1
Pour chaque zone ou chaque secteur	Distribution uniforme : les N établissements sont de même taille			Tous les salariés sont employés dans un seul établissement
	Cas limite			Cas limite
	L'emploi n'est pas du tout concentré dans la zone (ou le secteur)	L'emploi est peu concentré dans la zone (ou le secteur)	L'emploi est concentré dans la zone (ou le secteur)	Concentration maximale dans la zone (ou le secteur).

### II.3.3 L'indice d'entropie de Theil

On peut également utiliser l'indice d'entropie, proposé par Theil (1967), qui est égal à la somme des

parts pondérées par leur logarithme :  $E = \sum_{i=1}^N \xi_i \ln \xi_i$ .

Comme pour l'indice de Herfindhal, on peut calculer une valeur minimale et maximale de cet indice d'entropie. Ainsi si tous les établissements sont de taille égale avec la même part de marché  $1/N$ , l'indice d'entropie sera égal à  $-\ln(N)$ , en revanche la valeur maximale (0) est obtenue dans le cas où une seule firme détient tout l'emploi.

	$-\ln(N)$			0
Pour chaque zone ou chaque secteur	Distribution uniforme : les N établissements sont de même taille			Tous les salariés sont employés dans un seul établissement
	Cas limite			Cas limite
	L'emploi n'est pas du tout concentré dans la zone (ou le secteur)	L'emploi est peu concentré dans la zone (ou le secteur)	L'emploi est concentré dans la zone (ou le secteur)	Concentration maximale dans la zone (ou le secteur).

Cet indice d'entropie possède également une propriété remarquable de décomposition. Lorsque la population est divisée en plusieurs groupes ( $g=1..G$ ), l'indice d'entropie général se décompose en 2 éléments : l'entropie intra-groupe, mesurant les inégalités internes à chaque groupe et l'entropie inter-groupe, mesurant les inégalités entre les différents groupes.

$$E_{\text{intra}} = \sum_{g=1}^G \frac{x_g}{x} \left( \sum_{i \in g} \frac{x_i}{x_g} \ln \frac{x_i}{x_g} \right) = \sum_{g=1}^G \frac{x_g}{x} E_g \quad E_{\text{inter}} = \sum_{g=1}^G \frac{x_g}{x} \ln \frac{x_g}{x}$$

$$E = E_{\text{intra}} + E_{\text{inter}}$$

Ainsi, on peut effectuer des décompositions de plus en plus fines de la concentration des salariés dans un secteur ou dans une zone d'un territoire.

Par exemple, si on considère une région REG à D départements :

$$E_{REG} = \sum_{dep=1}^D \frac{x_{dep}}{x} \ln \frac{x_{dep}}{x} + \sum_{dep=1}^D \frac{x_{dep}}{x} E_{dep}$$

### II.3.4 L'indice de concentration de Gini

Comme les indicateurs C4 et C10, l'indice de Herfindhal ou celui de Theil, cet indice mesure la concentration de l'emploi ou de la rémunération dans les établissements. Il peut s'agir de l'emploi d'une zone (concentration géographique) ou de l'emploi d'un secteur (concentration productive).

	0			1
Pour chaque zone ou chaque secteur	Distribution uniforme : les N établissements sont de même taille			Tous les salariés sont employés dans un seul établissement
	Cas limite			Cas limite
	L'emploi n'est pas du tout concentré dans la zone (ou le secteur)	L'emploi est peu concentré dans la zone (ou le secteur)	L'emploi est concentré dans la zone (ou le secteur)	Concentration maximale dans la zone (ou le secteur).

### II.3.5 Le coefficient de répartition

Le coefficient de répartition mesure la concentration des secteurs dans les zones géographiques c'est-à-dire la répartition de l'emploi d'un secteur parmi les différentes zones du territoire. C'est un indice de Gini relatif, chaque zone géographique est pondérée par sa part dans l'emploi total du territoire. Il tient compte du poids des zones dans l'emploi total. Sur la courbe de Lorenz associée, on a donc en abscisse le poids cumulé des zones dans l'emploi de l'ensemble de l'économie et en ordonnée le poids cumulé des zones dans l'emploi du secteur considéré.

Les différentes zones sont classées par ordre croissant de leur spécificité sectorielle, ce qui correspond à la pente de la courbe de Lorenz :

$$r_i^k = \frac{X_i^k / X^k}{X_i / X} = \frac{X_i^k / X_i}{X^k / X}$$

	0			1
Pour chaque secteur k	Le poids des zones dans l'emploi du secteur k est identique à leur poids dans l'ensemble de l'économie.			Une seule zone géographique emploie la totalité du secteur k.
	Cas limite			Cas limite
	L'emploi du secteur k n'est pas du tout concentré	L'emploi du secteur k est peu concentré	L'emploi du secteur k est concentré	Le secteur k est totalement concentré dans une seule zone géographique

### II.3.6 La densité économique

La densité économique permet de caractériser l'importance de l'activité économique d'une zone. La densité économique  $D_j^E$  d'une zone  $j$  est calculée de manière analogue à la densité de population  $D_j^P$  comme le nombre d'emplois salariés par km<sup>2</sup>.

$$D_j^E = \frac{\text{Emploi}_j}{\text{Surface}_j} \quad \text{et} \quad D_j^P = \frac{\text{Population}_j}{\text{Surface}_j}$$

Ces deux densités sont naturellement très corrélées.

On peut également construire un taux d'emploi de la région : c'est-à-dire le rapport de l'emploi salarié à la population. Il s'agit aussi du rapport entre la densité économique et la densité de population.

$$TE_j = \frac{\text{Emploi}_j}{\text{Population}_j} = \frac{D_j^E}{D_j^P}$$

## II.4. Les mesures de spécialisation et spécificité

### II.4.1 Le coefficient de spécialisation d'une zone

Il mesure la concentration de l'emploi d'une zone dans un nombre plus ou moins grand de secteurs. C'est un indice de Gini relatif, c'est-à-dire que chaque secteur est pondéré par son poids dans l'emploi total. Ainsi ce coefficient tient compte de l'importance de chacun des secteurs au niveau agrégé du territoire de référence. On mesure donc la spécialisation par rapport aux activités du territoire de référence.

Sur la courbe de Lorenz associée, on cumule en abscisse la part de l'emploi des secteurs  $k$  dans l'emploi total du territoire de référence et en ordonnée la part de l'emploi des secteurs  $k$  dans l'emploi de la zone  $i$ . Les différents secteurs sont classés par ordre croissant de spécificité sectorielle de la zone, ce qui correspond à la pente de la courbe de Lorenz :

$$r_i^k = \frac{X_i^k / X_i}{X^k / X} = \frac{X_i^k / X^k}{X_i / X}$$

	0			1
Pour chaque zone $i$	Les poids des secteurs sont identiques dans la zone et dans le territoire de référence			Toute l'activité de la zone est concentrée dans un seul secteur
	Cas limite			Cas limite
	La zone $i$ n'est pas du tout spécialisée	La zone $i$ est peu spécialisée / très diversifiée	La zone $i$ est très spécialisée / peu diversifiée	La zone $i$ est complètement spécialisée

### II.4.2 Les indices de spécificité

Ces indicateurs mesurent la spécificité de chacune des zones d'étude à l'intérieur du territoire : d'abord les spécificités sectorielles des zones puis leur spécificité globale. La spécificité compare la structure sectorielle d'une zone d'intérêt à celle du reste du territoire. Comme pour le coefficient de

spécialisation, l'idée de spécificité globale développée ici est relative et revient à considérer si la zone d'étude s'est spécialisée dans des activités que l'on ne retrouve pas ailleurs.

#### II.4.2.1 Les indices de spécificité sectoriels

L'indice de spécificité sectoriel permet de comparer l'importance d'un secteur d'activités dans la zone

d'étude et dans l'ensemble du territoire. Il est généralement défini par :  $r_i^k = \frac{X_i^k / X_i}{X^k / X}$

Il est égal à 1 si le secteur k a la même importance dans la zone i et dans l'ensemble du territoire ; il vaut 0 si aucune activité du secteur k n'est localisée dans la zone i. Lorsque tout l'emploi du secteur k est localisé dans la zone i, il est égal à  $X/X_i$ . Sa valeur maximale dépend donc de la taille de la zone, ce qui peut être gênant quand on travaille sur une zone de grande taille par rapport au territoire de référence. Pour pallier ce problème, on choisit de comparer la zone non plus à l'ensemble du territoire mais au reste du territoire (territoire de référence hors zone d'étude). On obtient alors un indice de

spécificité modifié :  $\tilde{r}_i^k = \frac{X_i^k / X_i}{(X^k - X_i^k) / (X - X_i)}$

C'est cet indice modifié qui est privilégié par la suite.

	0			infini
Pour chaque secteur k	Aucun emploi du secteur k dans la zone i.			Tout l'emploi du secteur k est localisé dans la zone i
	Cas limite			Cas limite
	Le secteur k n'est pas du tout représenté	Le secteur k est peu représenté	Le secteur k est très représenté	Le secteur k n'est représenté que dans la zone i

#### II.4.2.2 L'indice de spécificité global de Krugman

L'inconvénient de ces indices de spécificité sectoriels provient du fait qu'il y en a autant que de secteurs considérés pour chaque zone du territoire. On peut les analyser pour découvrir quelles sont les spécialités sectorielles de la zone, mais cela ne fournit pas d'indication sur la spécificité globale de la zone. Afin de remédier à cet inconvénient et de mesurer globalement la spécificité de la zone, on va utiliser l'indice de spécificité global proposé par Paul Krugman en 1991 dans son livre « Geography and Trade ».

Cet indice calcule la différence entre la structure industrielle de la zone et celle des autres zones du territoire de référence. Ainsi en évitant de comparer la zone à l'ensemble du territoire, on évite de biaiser la mesure pour de grandes zones qui sont toujours plus proches de la structure industrielle moyenne. L'indice de Krugman est donc la somme des écarts en valeur absolue entre la structure industrielle de la zone et celle du reste du territoire de référence :

$$\begin{aligned}
 K_i &= \frac{1}{2} \sum_k \left| \frac{X_i^k}{X_i} - \frac{X^k - X_i^k}{X - X_i} \right| = \frac{1}{2} \sum_k \left| \frac{X_i^k}{X_i} - \frac{\tilde{X}_i^k}{\tilde{X}_i} \right| \\
 &= \frac{1}{2} \sum_k \left| \omega_i^k - \tilde{\omega}_i^k \right| = \frac{1}{2} \sum_k \tilde{\omega}_i^k \left| \frac{\omega_i^k}{\tilde{\omega}_i^k} - 1 \right| \\
 &= \frac{1}{2} \sum_k \tilde{\omega}_i^k \left| \tilde{r}_i^k - 1 \right|
 \end{aligned}$$

Cet indice de Krugman s'interprète également comme la somme des divergences, prises en valeur absolue, de l'indice de spécificité de la zone par rapport à l'unité, c'est-à-dire une situation où la zone possède la même part d'emploi que l'ensemble du territoire. Cette somme est pondérée par

l'importance du secteur  $k$ , mesurée par la part relative de l'emploi du secteur  $k$  dans le reste du territoire.

En conséquence, on obtient un indice synthétique de la spécificité de la zone par rapport au reste du territoire. On peut montrer que cet indice de Krugman est compris entre zéro et un<sup>2</sup>. Si la zone ressemble pour tous les secteurs au reste du territoire, on aura pour tout  $k$  :  $\tilde{r}_i^k = 1$  parce que  $\omega_i^k = \tilde{\omega}_i^k$  ou encore  $\omega_i^k = X_i^k / X_i = X^k / X = s^k$ . En conséquence, l'indice de Krugman sera nul et on pourra dire que la zone ressemble parfaitement au reste du territoire : elle n'a aucune spécificité du fait qu'elle reflète parfaitement la structure sectorielle de ce territoire. En revanche, si la zone est entièrement spécialisée dans des activités que l'on ne retrouve pas ailleurs, cet indice de Krugman sera égal à l'unité et on aura une zone parfaitement spécifique.

On peut d'ailleurs interpréter cet indice comme le pourcentage de réallocation sectorielle qu'il devrait y avoir dans la zone pour qu'elle ressemble parfaitement au reste du territoire en termes de structure productive. En effet, cet indice de Krugman est la somme des différences en pourcentage entre la structure productive de la zone et celle du reste du territoire. Donc pour ressembler parfaitement au reste du territoire et avoir alors une spécificité nulle, il faut que tous les indices de spécificité soient unitaires ( $\tilde{r}_i^k = 1$ ), ce qui se produit par une redistribution des activités productives entre les secteurs.

	0			1
Pour chaque zone $i$	La structure sectorielle de la zone $i$ est exactement la même que celle du territoire			La zone est entièrement spécialisée dans des activités que l'on ne retrouve pas ailleurs
	Cas limite			Cas limite
	Aucune spécificité de la zone $i$	La zone $i$ est peu spécifique	La zone $i$ est très spécifique	Parfaite spécificité de la zone $i$

### II.4.2.3 L'indice global de spécificité bilatérale

On peut également proposer un indice de spécificité bilatérale entre deux zones pour rechercher les zones qui ressemblent le plus à la zone étudiée ou les zones qui en sont les plus éloignées en termes de structure industrielle. Pour cela l'indice de Krugman est amendé de telle manière que la comparaison des structures industrielles ne se fasse plus avec le reste du territoire, mais avec une autre zone  $j$  de ce territoire :

$$B_{i,j} = \frac{1}{2} \sum_k \left| \frac{X_i^k}{X_i} - \frac{X_j^k}{X_j} \right| = \frac{1}{2} \sum_k |\omega_i^k - \omega_j^k|$$

Ainsi on peut construire un tableau carré (symétrique) de ces indices bilatéraux qui s'interprètent de la même manière que l'indice de Krugman : si l'indice est proche de zéro, les deux zones seront très semblables dans leur structure industrielle, sinon l'indice indiquera le pourcentage de réallocation des activités qu'il faudrait effectuer dans la zone  $i$  pour que cette zone ait la même structure industrielle que la zone  $j$ .

<sup>2</sup> Dans le livre de Krugman, l'indice est présenté sans le facteur 1/2, ainsi sa valeur est comprise entre zéro et deux. Pour faciliter son interprétation, nous avons divisé cet indice par deux pour que sa valeur maximale soit l'unité.



	0			1
Pour chaque couple de zones (i ; j)	La structure sectorielle de la zone i est exactement la même que celle de la zone j			La zone i est totalement différente de la zone j
	Cas limite			Cas limite
	Aucune spécificité de la zone i par rapport à la zone j	La zone i est peu spécifique par rapport à la zone j	La zone i est très spécifique par rapport à la zone j	Parfaite spécificité de la zone i par rapport à la zone j

## II.5. La mesure d'agglomération

L'objectif est de mesurer l'agglomération des établissements dans quelques zones du territoire. Pour cela, on étudie la corrélation entre les choix de localisation (zone i) de deux établissements appartenant à un même secteur  $k$ , cette corrélation représentant soit les avantages naturels à la localisation, soit les externalités liées à l'agglomération des établissements.

$$\text{Corr}(u_{j,i}^k, u_{h,i}^k) = \gamma^k \quad \text{pour } j \neq h \quad \text{avec } -1 \leq \gamma^k \leq 1. \quad (1)$$

Les deux variables  $u_{j,i}^k$  et  $u_{h,i}^k$  sont deux variables binaires qui valent 1 si l'établissement  $j$  (ou  $h$ ) sont localisées dans la zone  $i$  et 0 autrement. Ces deux variables de localisation peuvent être considérées comme des variables de Bernoulli avec une probabilité égale à la part de la zone  $i$  dans l'emploi total<sup>3</sup>  $s_i = X_i / X$  quel que soit le secteur d'activité. On a donc :

$$\Pr(u_{j,i}^k = 1) = \frac{X_i}{X} = s_i \quad \text{et} \quad E[u_{j,i}^k] = s_i \quad \text{et} \quad V[u_{j,i}^k] = s_i(1 - s_i)$$

En conséquence, on peut calculer facilement la probabilité que deux établissements d'un même secteur appartiennent à la même zone, en utilisant la définition de la covariance :

$$\begin{aligned} \Pr(u_{j,i}^k = 1, u_{h,i}^k = 1) &= E[u_{j,i}^k \cdot u_{h,i}^k] = \text{Cov}[u_{j,i}^k, u_{h,i}^k] + E[u_{j,i}^k]E[u_{h,i}^k] \\ &= \gamma^k s_i(1 - s_i) + s_i^2 \end{aligned}$$

La dernière égalité s'obtient avec la définition du coefficient de corrélation de localisation (1). En faisant la somme sur toutes les zones, on obtient alors la probabilité totale que deux établissements soient localisés dans une même zone quelconque du territoire :

$$p^k = \sum_i \Pr(u_{j,i}^k = 1, u_{h,i}^k = 1) = \gamma^k \sum_i s_i(1 - s_i) + \sum_i s_i^2 = \gamma^k \left(1 - \sum_i s_i^2\right) + \sum_i s_i^2.$$

ce qui peut se réécrire sous la forme :

$$\gamma^k = \frac{p^k - \sum_i s_i^2}{1 - \sum_i s_i^2}$$

Maintenant on veut estimer la valeur du paramètre d'intérêt  $\gamma^k$  : la corrélation pour que deux établissements soient localisés dans la même zone. Pour cela, il faut tout d'abord obtenir un estimateur de la probabilité  $p^k$  que deux établissements soient localisés dans la même zone.

<sup>3</sup> On considère ici que la variable de calcul est l'emploi de l'établissement.

Maurel et Sédillot (*MS*) ont proposé d'utiliser l'estimateur suivant :

$$\hat{p}^k = \sum_i \frac{\sum_{\substack{j,h \in i \\ j \neq h}} z_j^k z_h^k}{\sum_{\substack{j,h \\ j \neq h}} z_j^k z_h^k} \quad \text{avec } z_j^k = X_j^k / X^k \text{ la part de l'établissement } j \text{ dans l'emploi total du secteur } k$$

$$\text{Or } \sum_{\substack{j,h \in i \\ j \neq h}} z_j^k z_h^k = \left( \sum_{j \in i} z_j^k \right)^2 - \sum_{j \in i} z_j^{k2}$$

On note  $\omega_i^k = \sum_{j \in i} z_j^k$  la part de l'emploi du secteur  $k$  localisée dans la zone  $i$ .

$$\text{et } \sum_{\substack{j,h \\ j \neq h}} z_j^k z_h^k = \left( \sum_j z_j^k \right)^2 - \sum_j z_j^{k2} = 1 - H^k$$

où  $H^k$  est l'indice de Herfindhal de concentration horizontale (productive) du secteur  $k$  :  
 $H^k = \sum_j z_j^{k2}$

$$\text{On a alors } \hat{p}^k = \frac{\sum_i \omega_i^{k2} - H^k}{1 - H^k} \quad \text{et} \quad \hat{\gamma}_{MS}^k = \frac{\hat{p}^k - \sum_i s_i^2}{1 - \sum_i s_i^2} = \frac{G_{MS}^k - H^k}{1 - H^k} \quad \text{où}$$

$$G_{MS}^k = \frac{\sum_i \omega_i^{k2} - \sum_i s_i^2}{1 - \sum_i s_i^2} \quad \text{avec } \omega_i^k = X_i^k / X^k \text{ la part de l'emploi du secteur } k \text{ localisée dans la zone } i$$

Ellison et Glaeser (*EG*) proposent un estimateur de  $\gamma^k$  légèrement différent, dont la définition ne découle pas directement du modèle probabiliste comme celui de Maurel-Sédillot. Il est construit à partir de la définition a priori d'une « mesure de concentration géographique brute »  $G_{EG}^k$  définie intuitivement en comparant la répartition géographique de l'emploi dans une industrie  $k$  (définie par les  $\omega_i^k$ ) à celle de l'emploi total (définie par les  $s_i$ ) :

$$G_{EG}^k = \frac{\sum_i (\omega_i^k - s_i)^2}{1 - \sum_i s_i^2}$$

Les deux indicateurs  $G_{MS}^k$  et  $G_{EG}^k$  comparent la part de l'emploi dans la zone pour le secteur  $k$  et pour l'emploi total.

On peut montrer que l'espérance de ces deux mesures est identique :

$$E[G_{EG}^k] = E[G_{MS}^k] = H^k + \gamma^k (1 - H^k) \quad (2)$$

Si la corrélation de localisation est nulle ( $\gamma^k = 0$ ) : il n'y a pas d'avantages naturels à la localisation ou d'externalités d'agglomération entre les établissements du secteur. En conséquence, cette

espérance sera égale à l'indice de Herfindhal de concentration horizontale du secteur. Ce qui permet de faire un test de l'absence d'externalités d'agglomération conditionnellement à la structure productive du secteur, c'est-à-dire que l'hypothèse nulle sera  $H_0 : \gamma^k = 0$  :

$$\text{Test } t_{H_0: \gamma^k=0} = \frac{G_a^k - H^k}{s(G_a^k)} \quad \text{avec } a = EG \text{ ou } MS$$

avec  $s(G_a^k)$  l'écart-type estimé de la mesure de concentration géographique brute. On peut estimer ces écarts-type sous l'hypothèse nulle avec les formules données dans les articles respectifs de Ellison-Glaeser et Maurel-Sédillot. Cependant le calcul de ces écarts-type en dehors de l'hypothèse nulle d'absence de corrélation de localisation se révèle très complexe et nécessite des hypothèses supplémentaires sur le modèle de localisation des établissements.

En comparant les deux estimateurs proposés, on peut toutefois remarquer que l'estimateur de Maurel et Sédillot possède souvent une variance plus élevée que l'estimateur de Ellison et Glaeser, surtout avec des niveaux de concentration productive (indice de Herfindhal) faibles. Ce dernier cas étant le plus fréquent, on préférera utiliser l'estimateur original de Ellison et Glaeser qui possède en général une meilleure précision que l'estimateur de Maurel et Sédillot, bien qu'ils soient tous les deux sans biais.

Au-delà de ce test, on peut obtenir un estimateur de la corrélation de localisation sous la forme :

$$\hat{\gamma}_a^k = \frac{G_a^k - H^k}{1 - H^k} \quad \text{avec } a = EG \text{ ou } MS.$$

Cet estimateur est évidemment sans biais par la propriété d'espérance précédente (2). On rappelle que l'on prend la structure productive sectorielle comme donnée et donc que l'indice de Herfindhal est un paramètre fixe exogène déterminé par les caractéristiques productives du secteur étudié.

## II.6. Quels indicateurs pour quel type de concentration ?

En résumé, les indicateurs proposés sont, pour chaque type de concentration, les suivants :

*Tableau 1 : Liste des différents indicateurs proposés pour mesurer chaque type de concentration*

Type de concentration	Indicateurs proposés
Concentration productive	C4, C10 Indice de Herfindhal Indice d'entropie de Theil Indice de Gini de concentration
Concentration géographique	C4, C10 Indice de Herfindhal Indice d'entropie de Theil Indice de Gini de concentration
Répartition spatiale	Indice de Gini de répartition Mesure de concentration géographique brute d'Ellison et Glaeser Mesure de concentration géographique brute de Maurel et Sédillot
Spécialisation	Indice de Gini de spécialisation Spécificités sectorielles Spécificités bilatérales Spécificité globale de Krugman

## II.7. Limites des mesures de répartition spatiale et d'agglomération

Les mesures de répartition « spatiale » des activités présentées précédemment ont des limites. Le problème majeur repose sur le fait que l'on classe la localisation des établissements dans des zones prédéterminées de manière administrative (régions, départements) ou économique-sociale (zone d'emploi, aires urbaines). Cela revient à faire abstraction de la localisation de l'établissement en un point du territoire. Comme le soulignent Duranton et Overmann [2005], on transforme ainsi des « points sur une carte en éléments dans une boîte », en perdant ainsi de l'information.

D'un point de vue pratique, cela entraîne deux difficultés. La première est un *effet de frontière* qui peut séparer un pôle d'activité dans deux zones différentes. Selon que la frontière traverse le pôle d'activité ou non, on peut obtenir des mesures d'agglomération différentes. Soit toute l'activité se situe dans une zone et on aura une mesure d'agglomération très élevée, soit la frontière sépare le pôle d'activité, et on aura alors une mesure d'agglomération plus faible, alors que la réalité économique reste identique.

La seconde difficulté provient du fait que le découpage prédéterminé supprime l'effet de distance. Si on considère une activité sur deux zones du territoire, on obtiendra la même mesure d'agglomération pour des cas très différents : soit les deux zones sont contiguës et forment un seul pôle d'activité, soit les deux zones sont très éloignées et forment alors deux pôles d'activité sur le territoire. Dès lors l'interprétation des mesures d'agglomération doit se faire avec une certaine prudence et nécessite que l'on s'intéresse à différents niveaux de zonage pour tenter d'évaluer des agglomérations au niveau des aires urbaines, des zones d'emploi ou des régions par exemple.

Pour remédier à ces difficultés, Arbia et Espa [1996], Marcon et Puech [2003] et [2012] et Duranton et Overmann [2005] considèrent non plus un espace divisé en zones figées mais un espace continu et introduisent des mesures de concentration spatiale des activités basées sur la distance entre les unités considérées (les établissements dans notre cas). Cependant la mise en œuvre de ces méthodes nécessite des logiciels et des puissances de calcul peu souvent disponibles. Le *Document de travail* « Détection des disparités socio-économiques, l'apport de la statistique spatiale » de Jean-Michel Floch [2012] expose quelques méthodes d'analyse spatiale et fournit des exemples de leur utilisation à l'Insee.

### III. Analyse de la structure productive d'une zone d'un territoire

On présente dans cette partie des pistes d'utilisation et d'interprétation des indicateurs de spécialisation et de concentration qui permettent de décrire la structure productive d'une zone d'un territoire.

#### III.1. Les données « Connaissance locale de l'appareil productif » (CLAP)

Les exemples présentés dans cette partie et dans la suivante sont réalisés à partir des données « Connaissance locale de l'appareil productif » (CLAP) de l'Insee.

CLAP est un outil conçu à partir de différents fichiers administratifs : SIRENE, URSSAF, DADS et SIASP. Il fournit chaque année une image de l'activité économique, en mettant à disposition des données d'emplois salariés et de rémunérations finement localisées tant au niveau géographique qu'à celui de l'activité. Il couvre la France métropolitaine et les départements d'outre-mer, et l'ensemble des activités économiques à l'exception du secteur de la défense et des ménages employeurs domestiques.

A ce titre, CLAP permet de répondre aux besoins locaux de connaissance de l'appareil productif et aux besoins nationaux sur la localisation de celui-ci. Ce sont les données privilégiées de l'Insee pour la description du tissu productif local.

Plusieurs outils mobilisant les données CLAP ont été développés à l'Insee pour l'analyse de la structure productive des territoires. En particulier, l'investissement « Concentration et spécialisation » du PSAR Études économiques régionales permet de produire les différents indicateurs de concentration présentés dans ce document sur divers zonages et pour différents niveaux sectoriels. A titre d'exemple, les études suivantes ont été réalisées à partir de cet investissement : « L'industrie en Bourgogne : diversifiée avec des spécialisations locales », Insee Bourgogne Dimensions n°179 - 2012, « Des territoires inégalement dotés pour faire face aux mutations économiques », SUD INSEE L'essentiel n°131 - 2009.

Dans la suite du document, les exemples sont tous produits à partir des données CLAP de 2010. Les effectifs correspondent aux effectifs salariés au 31/12/2010 hors intérimaires et les rémunérations au cumul des rémunérations brutes sur l'année 2010. Le territoire étudié est la France métropolitaine et les départements d'Outre-mer (Guadeloupe, Martinique, Guyane, La Réunion), détaillé par région, département et zone d'emploi. Pour certains exemples, il a été réduit à la France métropolitaine. En termes d'activités, le champ privilégié est celui de l'ensemble de l'économie, à l'exception du secteur de la défense et des ménages employeurs domestiques. Il est parfois restreint au champ marchand non agricole pour ne pas prendre en compte dans l'analyse les établissements du secteur public et les exploitations agricoles qui présentent des concentrations atypiques. Les indicateurs sont produits en agrégeant les activités en 38 secteurs (niveau A38), 88 secteurs (niveau A88) ou sans les détailler, c'est-à-dire en 732 secteurs (niveau NAF732).

#### III.2. Les indices de spécialisation

*La zone d'étude est-elle spécialisée dans un faible nombre de secteurs ou offre-t-elle un grand éventail d'activités ?*

Pour répondre à cette question, on propose l'utilisation du coefficient de Gini (relatif) pour mesurer la concentration sectorielle des activités dans la zone par rapport à celle du territoire de référence (voir § II.4.1). On appellera ce coefficient de Gini le **coefficient de spécialisation**. On pourra également utiliser l'indice de spécificité global de Krugman (voir § III.3)

Le tableau 2 donne la moyenne et l'écart-type de ce coefficient de spécialisation suivant les niveaux de désagrégation spatial et sectoriel. On a calculé les coefficients de spécialisation des régions, des départements et des zones d'emploi, pour les secteurs des niveaux A38, A88 et NAF732. On remarque immédiatement que plus on travaille à un niveau fin de désagrégation sectorielle, plus le coefficient de spécialisation est élevé car la répartition des activités par secteur peut devenir plus

diversifiée entre secteurs fins d'activité. De même, plus on considère des zones petites, plus leur spécialisation est forte.

*Tableau 2 - Coefficient de spécialisation, moyenne et écart-type (entre parenthèses)*

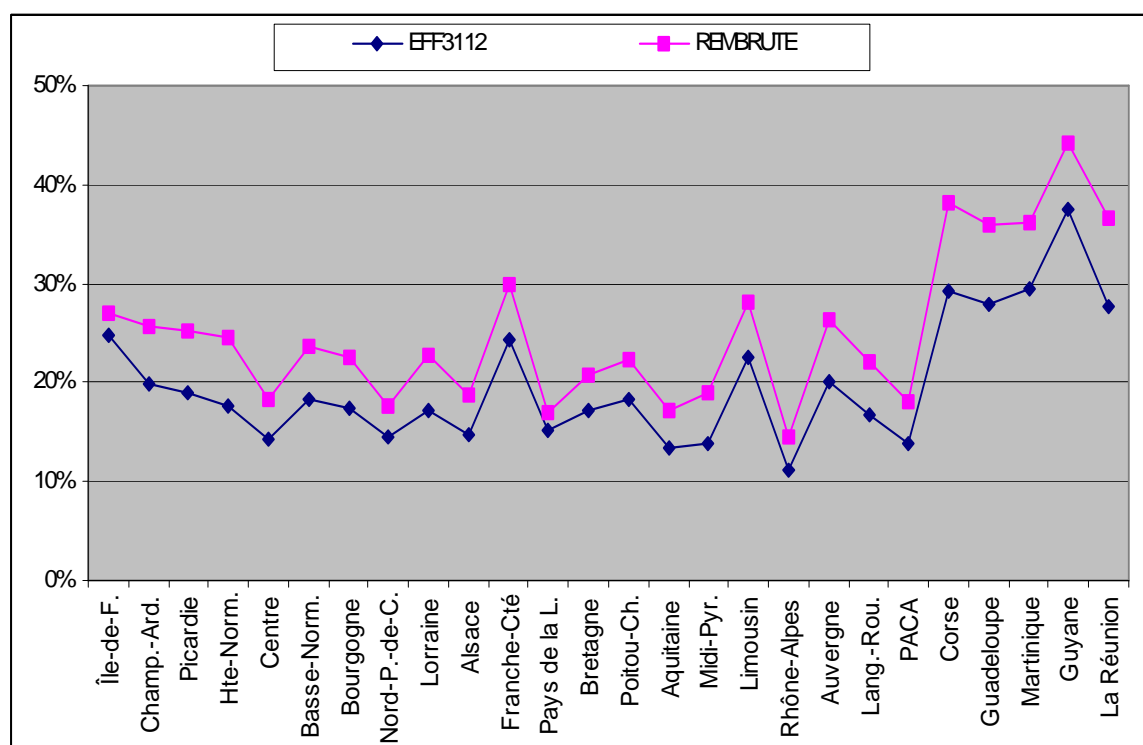
	Région	Département	Zone d'emploi
A38	0.166 (0.055)	0.208 (0.057)	0.296 (0.091)
A88	0.199 (0.064)	0.249 (0.061)	0.351 (0.098)
NAF732	0.297 (0.081)	0.385 (0.072)	0.529 (0.117)

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

26 régions - 100 départements - 322 zones d'emploi

*Graphique 2 : Coefficients de spécialisation régionaux*



Le graphique 2 montre les divergences observées lorsqu'on s'intéresse aux rémunérations plutôt qu'aux effectifs. Dans l'ensemble la différence est relativement faible au niveau régional, où les deux variables suivent le même profil.

On observe bien également la particularité des DOM et de la Corse qui ont une structure sectorielle bien différente de la référence France.

On remarque que la spécialisation est systématiquement plus marquée pour les salaires que pour les effectifs. Cela peut s'expliquer par une distribution des salaires plus inégalitaire que celle des emplois.

Le tableau 3 présente les valeurs du coefficient de spécialisation par région pour les effectifs et les rémunérations, au niveau sectoriel A88. Comme pour le tableau et le graphique précédents, la référence est le niveau France entière.

Même si l'étude des rémunérations donne en moyenne un coefficient de spécialisation plus élevé, le classement des régions n'est que légèrement modifié. Le coefficient de corrélation entre la spécialisation selon les effectifs et la spécialisation selon la rémunération est de 0,98.

*Tableau 3 : Coefficient de spécialisation par région (en pourcentage)  
(classement par ordre décroissant pour les effectifs)*

#	REGION	Effectifs salariés	Rémunérations
1	03 GUYANE	37,6	44,1
2	94 CORSE	29,3	38,1
3	02 MARTINIQUE	29,5	36,2
4	01 GUADELOUPE	28,0	35,9
5	04 LA RÉUNION	27,7	36,5
6	11 ÎLE-DE-FRANCE	24,8	27,0
7	43 FRANCHE-COMTÉ	24,3	29,8
8	74 LIMOUSIN	22,5	28,2
9	83 AUVERGNE	20,2	26,3
10	21 CHAMPAGNE-ARDENNE	19,9	25,7
11	22 PICARDIE	19,0	25,1
12	54 POITOU-CHARENTES	18,4	22,3
13	25 BASSE-NORMANDIE	18,3	23,6
14	23 HAUTE-NORMANDIE	17,7	24,5
15	26 BOURGOGNE	17,4	22,4
16	53 BRETAGNE	17,3	20,8
17	41 LORRAINE	17,2	22,7
18	91 LANGUEDOC-ROUSSILLON	16,7	22,0
19	52 PAYS DE LA LOIRE	15,1	17,0
20	42 ALSACE	14,8	18,8
21	31 NORD-PAS-DE-CALAIS	14,5	17,6
22	24 CENTRE	14,3	18,4
23	73 MIDI-PYRÉNÉES	13,9	19,0
24	93 PACA	13,9	18,0
25	72 AQUITAINE	13,3	17,3
26	82 RHÔNE-ALPES	11,1	14,4
Statistiques descriptives sur les 26 régions			
	Moyenne	19,9	25,1
	Écart-type	6,4	7,6
	Minimum	11,1	14,4
	Maximum	37,6	44,1
Corrélations			
	Effectifs	1,000	0,983
	Rémunérations	0,983	1,000

Source : CLAP 2010  
Niveau sectoriel : a88  
Territoire de référence : France entière

On peut conclure de cette première analyse que les petites régions, à caractère plus rural, sont plutôt spécialisées par rapport à la référence France, alors que les grandes régions sont assez diversifiées, à l'exception de l'Île-de-France dont le caractère spécialisé peut s'expliquer par la prédominance d'activités de service. Mais cette spécialisation au niveau régional peut cacher de grandes variations à l'intérieur de chacune des régions. En effet, lorsqu'on calcule les coefficients de spécialisation de chaque zone d'emploi, la variabilité est beaucoup plus importante que pour le niveau régional : l'écart-type passe de 6,4 % (niveau région) à 9,8 % (niveau zone d'emploi) pour les effectifs salariés.

### III.3. Les indices de spécificité

La notion de spécificité globale est assez proche de celle de spécialisation. Il s'agit de comparer la structure sectorielle d'une zone d'étude à celle des autres zones. La principale différence entre les deux tient au fait que l'indice de Gini de spécialisation compare la zone d'étude à l'ensemble du territoire de référence, alors que l'indice de Krugman limite la comparaison aux autres zones du territoire, c'est-à-dire le territoire de référence moins la zone étudiée.

L'idée de spécificité globale revient à considérer si la zone d'étude s'est spécialisée dans des activités que l'on ne retrouve pas ailleurs.

#### *Quelle est la spécificité des activités économiques d'une zone ?*

On peut calculer un **indice de spécificité sectoriel** pour chaque secteur de la zone. Il permet de comparer le poids du secteur  $k$  dans l'activité de la zone et dans celle du territoire de référence (voir § II.4.2.1).

Par exemple, en Midi-Pyrénées, la spécificité de la région pour le secteur de la fabrication de matériels de transport est égale à 2,41. Cela signifie que ce secteur est 2,41 fois plus important dans cette région qu'en moyenne dans le reste de la France. Cela s'explique par la représentation forte de la construction aéronautique dans la région.

#### *Est-on proche ou éloigné de la structure industrielle des autres zones ?*

L'**indice de Krugman** permet de synthétiser au niveau de la zone l'information contenue dans les différents indices sectoriels (voir § II.4.2.3). Il compare la structure industrielle de la zone à celle du reste du territoire (territoire de référence hors zone étudiée).

Le tableau 4 donne les moyennes et écarts-type de cet indice de Krugman pour différents zonages (région, département, zone d'emploi) aux niveaux sectoriels A38, A88 et NAF732, sur l'effectif salarié au 31/12/2010. On remarque encore clairement que plus la classification sectorielle est détaillée, plus grand est l'indice de Krugman en moyenne. Ce qui s'explique par le fait que la diversification sectorielle est plus importante pour des secteurs définis de manière plus fine car la répartition des emplois dans les sous-secteurs n'est certainement pas uniforme. Pour la même raison, plus le détail géographique augmente, plus l'indice de spécificité augmente en moyenne. En effet, en considérant un zonage plus fin et une classification sectorielle plus détaillée, il y a davantage de risques de trouver des activités plus agglomérées dans des zones plus spécialisées. Ainsi il est normal que cet indice de spécificité, comme le coefficient de spécialisation, augmente en moyenne avec des classifications plus détaillées. Dit autrement, l'hypothèse de non-uniformité sectorielle et géographique du tissu productif français semble être vérifiée sur ces données.

La dispersion de ces indices s'accroît également avec la finesse du zonage géographique ou de la classification sectorielle. Cependant si on rapporte l'écart-type à la moyenne pour obtenir le coefficient de variation, celui-ci diminue légèrement avec le détail de la classification sectorielle, alors qu'il reste relativement stable entre les différents niveaux géographiques. Il semble donc qu'il y ait une stabilité de la dispersion relative de l'indice de spécificité.

*Tableau 4 - Indice de Krugman, moyenne et écart-type (entre parenthèses)*

	Région	Département	Zone d'emploi
A38	0.116 (0.043)	0.143 (0.044)	0.206 (0.067)
A88	0.140 (0.050)	0.168 (0.046)	0.242 (0.072)
NAF732	0.208 (0.064)	0.261 (0.056)	0.370 (0.093)

Source : CLAP 2010

Variable : effectifs salariés au 31/12

26 régions - 100 départements - 322 zones d'emplois

Le tableau 5 présente les valeurs de l'indice de spécificité de Krugman par région suivant les 3 nomenclatures sectorielles utilisées plus haut. Même s'il existe de faibles différences suivant la classification sectorielle adoptée, on peut cependant noter une bonne cohérence entre les



classements relatifs que l'on obtient avec les différents niveaux sectoriels (corrélation de rang > 0,9 entre les niveaux A88 et NAF732).

Excepté en Île-de-France qui concentre un grand nombre d'activités très spécifiques surtout dans le tertiaire, les régions les plus spécifiques sont les plus petites et les plus agricoles, alors que les grandes régions peuvent offrir une palette d'activités plus diversifiées et plus proches de la moyenne nationale.

*Tableau 5 : Indice de Krugman par région (en pourcentage), classement par ordre décroissant pour le niveau A732*

#	REGION	A38	A88	A732
1	03 GUYANE	24,2	28,0	38,1
2	01 GUADELOUPE	14,4	20,4	30,8
3	02 MARTINIQUE	16,6	21,4	30,5
4	04 LA RÉUNION	16,1	20,0	29,1
5	11 ÎLE-DE-FRANCE	20,4	22,6	28,8
6	94 CORSE	16,3	20,3	28,4
7	43 FRANCHE-COMTÉ	14,1	15,5	23,8
8	74 LIMOUSIN	12,2	15,3	23,3
9	22 PICARDIE	10,6	12,4	20,2
10	21 CHAMPAGNE-ARDENNE	10,9	12,5	19,8
11	83 AUVERGNE	12,1	13,6	19,5
12	25 BASSE-NORMANDIE	10,8	12,5	19,0
13	53 BRETAGNE	10,6	12,0	18,7
14	54 POITOU-CHARENTES	9,3	11,9	18,4
15	41 LORRAINE	9,4	12,0	18,2
16	23 HAUTE-NORMANDIE	10,2	11,4	17,9
17	26 BOURGOGNE	10,5	11,6	17,7
18	42 ALSACE	9,0	10,2	17,5
19	91 LANGUEDOC-ROUSSILLON	11,8	12,9	17,4
20	31 NORD-PAS-DE-CALAIS	7,3	10,6	16,5
21	52 PAYS DE LA LOIRE	9,0	10,6	16,0
22	24 CENTRE	7,7	9,5	15,6
23	93 PACA	8,3	10,3	15,5
24	73 MIDI-PYRÉNÉES	7,5	9,5	15,1
25	72 AQUITAINE	7,0	8,7	13,4
26	82 RHÔNE-ALPES	6,6	8,1	12,5

Statistiques descriptives sur les 26 régions

Moyenne	11,6	14,0	20,8
Écart-type	4,3	5,0	6,4
Minimum	6,6	8,1	12,5
Maximum	24,2	28,0	38,1

Corrélations

A38	1,000	0,976	0,948
A88	0,976	1,000	0,987
NAF732	0,948	0,987	1,000

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

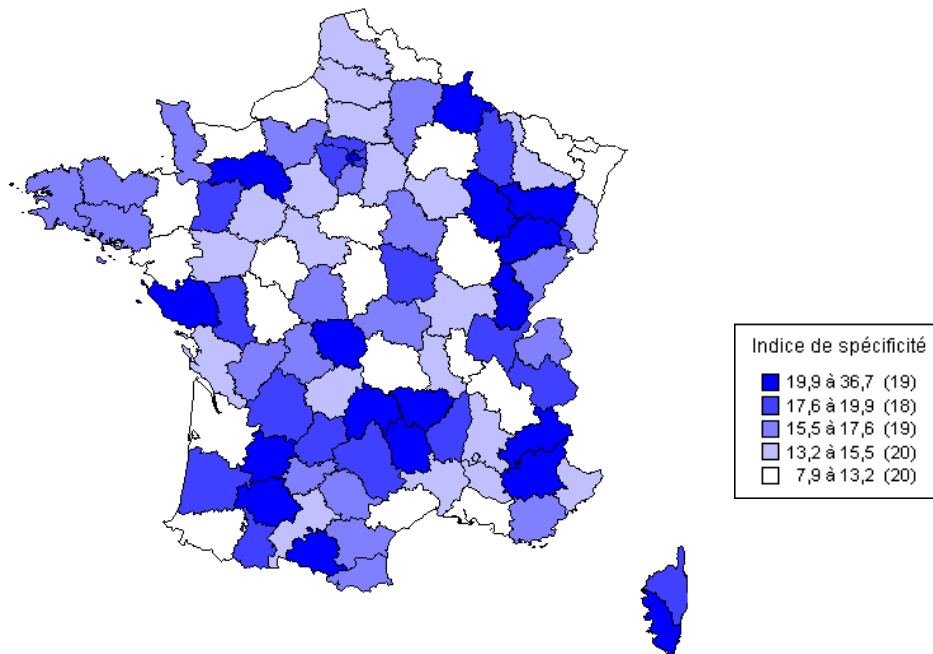
Territoire de référence : France entière

On peut également représenter cet indice de Krugman sur une carte comme celle qui est présentée ci-dessous (Carte 1) au niveau départemental en A88. On remarque sur cette carte qu'il n'y pas de structure particulièrement marquée pour la spécificité des départements. Néanmoins il semble que les

départements les plus spécifiques sont les départements assez isolés avec peu d'activités industrielles, à l'exception de la région parisienne.

D'un autre côté, les départements abritant une grande métropole régionale connaissent en général un faible degré de spécificité car ces métropoles régionales regroupent un vaste éventail d'activités. Cela conforte l'idée d'une polarisation régionale autour de la métropole principale de chaque région.

Carte 1 : Indice de Spécificité de Krugman par département (en A38)



### *De quelle autre zone est-on proche en termes d'activité économique ?*

**L'indice de spécificité bilatéral** compare les spécificités globales des zones deux à deux (voir § II.4.2.2). On présente dans le tableau 6 les indices de spécificité bilatérale calculés à partir des effectifs salariés au 31 décembre de CLAP 2010 qui comparent les structures industrielles au niveau A38 des différentes régions. La lecture de ce tableau peut se faire en ligne ou en colonne, nous privilégierons ici la lecture en ligne. Dans ce tableau, on a indiqué en rouge la région la plus éloignée en termes de structure sectorielle, et en bleu, la région sectoriellement la plus semblable.

Les deux régions qui semblent immédiatement les plus atypiques dans leur structure industrielle sont la Corse et l'Île-de-France qui apparaissent comme les plus éloignées des autres régions. L'Île-de-France a une structure industrielle très éloignée de la Franche-Comté (27,0 %) et de la Corse (24,7 %), mais aussi des régions voisines telles que la Picardie (25,1 %) ou Champagne-Ardenne (25,3 %). En revanche, elle semble être plus proche de grandes régions du sud comme PACA (16,2 %), Rhône-Alpes (18,0 %), Midi-Pyrénées (18,7 %), Aquitaine (20,1 %) ou Languedoc-Roussillon (21,9 %), tout en ayant une valeur assez élevée de l'indice de spécificité bilatérale.

En enlevant de l'analyse l'Île-de-France et la Corse qui ont une structure sectorielle très particulière, et en se concentrant sur les 20 autres régions, on remarque que les deux régions du sud PACA et Languedoc-Roussillon sont assez différentes, dans leur structure sectorielle, des régions de la moitié Nord de la France. De même, la Franche-Comté est très spécifique par rapport aux régions de l'Ouest et du Sud.

Si on s'intéresse à une région particulière, par exemple à la Picardie, on voit que cette région est assez proche dans sa structure industrielle de la Lorraine (6,5 %), de la Bourgogne (6,6 %), de la Champagne-Ardenne (6,8 %) et du Centre (7,5 %) c'est-à-dire de régions qui sont géographiquement assez proches. En revanche, elle est très différente de l'Île-de-France (25,1 %) et de la Corse (17,6 %) et dans une moindre mesure des régions du sud PACA (14,4 %) et Languedoc-Roussillon (13,7 %).

Cet indicateur permet ainsi d'aborder une analyse plus détaillée, pour caractériser les proximités entre les régions, ou au contraire leur éloignement en termes de structure du tissu productif.

*Tableau 6 - Indice de spécificité bilatéral par région*

Région	Île-de-France	Champagne-Ardenne	Picardie	Haute-Normandie	Centre	Basse-Normandie	Bourgogne	Nord-Pas-de-Calais	Lorraine	Alsace	Franche-Comté	Pays de la Loire	Bretagne	Poitou-Charentes	Aquitaine	Midi-Pyrénées	Limousin	Rhône-Alpes	Auvergne	Languedoc-Roussillon	Provence-Alpes-Côte d'Azur	Corse
Île-de-France		25.3	25.1	22.3	21.1	24.7	24.4	21.7	24.1	21.3	27.0	22.1	23.0	23.2	20.1	18.7	26.6	18.0	25.3	21.9	16.2	24.7
Champagne-Ardenne	25.3		6.8	11.2	8.4	8.7	6.2	8.8	6.0	11.4	9.3	8.6	10.6	9.1	9.8	12.3	10.2	11.0	8.7	13.8	14.7	17.5
Picardie	25.1	6.8		8.7	7.5	9.7	6.6	8.3	6.5	11.8	10.6	10.1	11.3	10.3	11.0	11.7	9.8	10.8	7.5	13.7	14.4	17.6
Haute-Normandie	22.3	11.2	8.7		6.3	11.5	10.6	8.4	9.5	11.6	12.3	9.6	12.8	11.3	11.1	11.5	13.4	9.4	11.7	14.7	14.4	18.3
Centre	21.1	8.4	7.5	6.3		9.3	6.9	8.2	7.0	8.4	11.8	6.4	9.8	8.3	8.1	9.9	11.4	6.3	9.3	12.8	12.5	16.7
Basse-Normandie	24.7	8.7	9.7	11.5	9.3		8.2	9.2	8.5	10.4	10.4	7.7	7.7	7.5	8.8	7.7	10.1	12.1	7.9	11.2	15.0	18.1
Bourgogne	24.4	6.2	6.6	10.6	6.9	8.2		8.4	6.2	9.9	9.3	8.9	9.8	6.8	8.1	11.2	8.2	10.0	7.3	12.0	13.7	16.1
Nord-Pas-de-Calais	21.7	8.8	8.3	8.4	8.2	9.2	8.4		6.5	9.8	10.8	8.4	9.6	9.2	7.7	7.4	10.1	8.9	9.1	11.5	10.8	17.9
Lorraine	24.1	6.0	6.5	9.5	7.0	8.5	6.2	6.5		9.9	8.7	9.2	10.7	8.8	9.8	9.8	7.7	9.8	7.7	12.1	12.6	16.5
Alsace	21.3	11.4	11.8	11.6	8.4	10.4	9.9	9.8	9.9		12.5	8.3	10.3	10.5	9.3	10.6	13.7	7.4	12.7	13.6	11.8	18.4
Franche-Comté	27.0	9.3	10.6	12.3	11.8	10.4	9.3	10.8	8.7	12.5		11.7	14.3	12.3	14.4	12.3	13.3	13.9	10.8	16.9	17.9	22.1
Pays de la Loire	22.1	8.6	10.1	9.6	6.4	7.7	8.9	8.4	9.2	8.3	11.7		8.5	8.4	9.4	10.5	13.2	8.6	11.6	14.3	14.2	18.0
Bretagne	23.0	10.6	11.3	12.8	9.8	7.7	9.8	9.6	10.7	10.3	14.3	8.5		9.5	7.5	8.8	12.4	11.1	10.6	10.1	12.4	17.6
Poitou-Charentes	23.2	9.1	10.3	11.3	8.3	7.5	6.8	9.2	8.8	10.5	12.3	8.4	9.5		6.7	10.0	8.9	11.9	10.0	11.3	14.1	16.8
Aquitaine	20.1	9.8	11.0	11.1	8.1	8.8	8.1	7.7	9.8	9.3	14.4	9.4	7.5	6.7		7.7	11.1	10.2	10.5	7.8	8.7	14.4
Midi-Pyrénées	18.7	12.3	11.7	11.5	9.9	7.7	11.2	7.4	9.8	10.6	12.3	10.5	8.8	10.0	7.7		12.2	9.8	9.9	9.4	10.9	17.5
Limousin	26.6	10.2	9.8	13.4	11.4	10.1	8.2	10.1	7.7	13.7	13.3	13.2	12.4	8.9	11.1	12.2		13.9	9.2	11.9	13.7	16.2
Rhône-Alpes	18.0	11.0	10.8	9.4	6.3	12.1	10.0	8.9	9.8	7.4	13.9	8.6	11.1	11.9	10.2	9.8	13.9		12.3	11.9	10.3	18.3
Auvergne	25.3	8.7	7.5	11.7	9.3	7.9	7.3	9.1	7.7	12.7	10.8	11.6	10.6	10.0	10.5	9.9	9.2	12.3		11.0	14.5	18.8
Languedoc-Roussillon	21.9	13.8	13.7	14.7	12.8	11.2	12.0	11.5	12.1	13.6	16.9	14.3	10.1	11.3	7.8	9.4	11.9	13.5	11.0		7.6	10.6
Provence-Alpes-Côte d'Azur	16.2	14.7	14.4	14.4	12.5	15.0	13.7	10.8	12.6	11.8	17.9	14.2	12.4	14.1	8.7	10.9	13.7	10.3	14.5	7.6		12.2
Corse	24.7	17.5	17.6	18.3	16.7	18.1	16.1	17.9	16.5	18.4	22.1	18.0	17.6	16.8	14.4	17.5	16.2	18.3	18.8	10.6	12.2	

### III.4. Les indices de concentration

*Quelle est la distribution de l'emploi dans la zone ?*

**Les indicateurs C4 et C10** permettent de déterminer la part dans l'emploi des 4 ou des 10 plus grands établissements à l'intérieur d'une zone déterminée (voir § II.3.1).

Ces indicateurs permettent de savoir si, dans un zone géographique donnée, les salariés travaillent dans peu d'établissements ou s'ils sont répartis dans un grand nombre d'établissements.

Étudié en termes de zone géographique, **l'indice de Herfindhal** permet de savoir comment les salariés d'une zone sont répartis dans les établissements de la zone. C'est-à-dire si l'emploi de la zone est peu ou au contraire très concentré dans les établissements (voir § II.3.2).

Contrairement aux indicateurs C4 ou C10, il prend en considération l'ensemble des établissements de la zone.

Comme l'indice de Herfindhal, **l'indice d'entropie** mesure la concentration des emplois de la zone dans les établissements (voir § II.3.3). Il possède une propriété intéressante, permettant de décomposer la concentration des salariés dans les établissements selon les différents niveaux d'agrégation du zonage géographique ou du secteur d'activité.

Exemple : indice d'entropie de la région Midi-Pyrénées

DEP	LIBDEP	Effectif	Entropie
09	Ariège	40 537	-7,15
12	Aveyron	83 564	-7,89
31	Haute-Garonne	506 555	-8,65
32	Gers	50 943	-7,60
46	Lot	47 866	-7,34
65	Hautes-Pyrénées	71 161	-7,52
81	Tarn	103 884	-8,07
82	Tarn-et-Garonne	63 948	-7,63

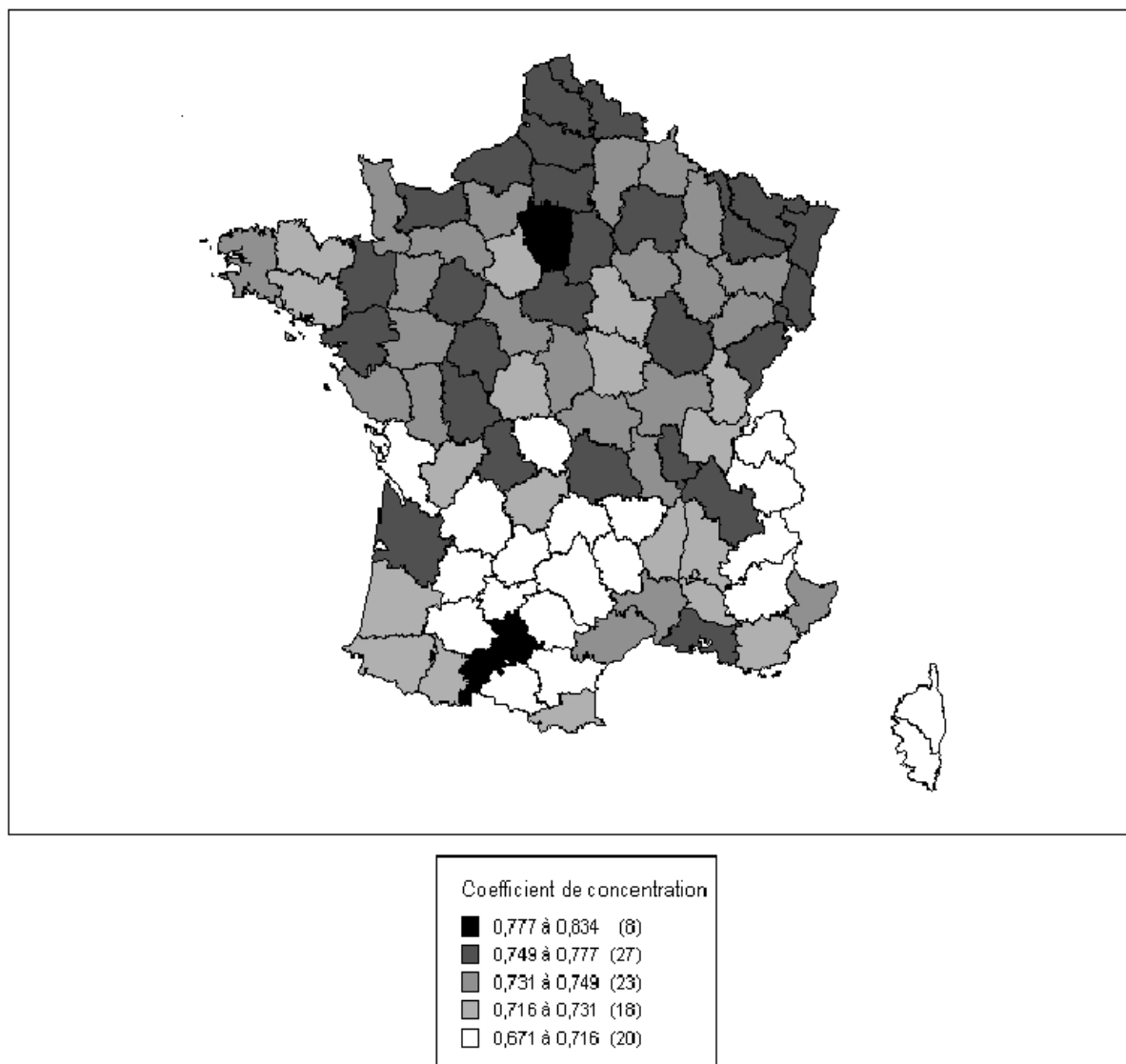
REG	LIBREG	Effectif	Entropie
73	Midi-Pyrénées	968 458	-9,79

$$\begin{aligned}
 E_{reg73} &= \underbrace{\sum_{dep} \frac{effectif_{dep}}{effectif_{reg73}} \ln \left( \frac{effectif_{dep}}{effectif_{reg73}} \right)}_{\text{entropie inter départements}} + \underbrace{\sum_{dep} \frac{effectif_{dep}}{effectif_{reg73}} E_{dep}}_{\text{entropie intra départements}} \\
 &= -1,60 - 8,19 = -9,79
 \end{aligned}$$

L'entropie intra-départements est une moyenne pondérée des indices d'entropie des départements. L'entropie inter-départements sera d'autant plus élevée que les salariés seront inégalement répartis entre les départements. Dans cet exemple, la concentration des salariés en Midi-Pyrénées s'explique en quasi-totalité par leur inégale répartition dans les établissements au sein des différents départements. Les poids inégaux des départements dans l'emploi salarié de la région n'interviennent qu'à hauteur de 16 %.

**Le coefficient de concentration** est un indice de Gini (voir § II.3.4). Comme les indices précédents, il mesure la concentration des salariés de la zone dans les établissements.

*Carte 2 : Coefficients de concentration départementaux*



La carte 2 représente les coefficients de concentration départementaux, calculés pour les effectifs salariés au 31/12/2010. La France semble coupée en deux par une ligne allant de la Vendée à la Savoie, avec une moitié nord où la concentration des salariés dans les établissements est plus importante que dans la moitié sud. Les départements présentant les plus fortes concentrations sont les départements franciliens hormis la Seine-et-Marne, et la Haute-Garonne. En revanche la concentration des salariés dans les établissements est très faible en Corse, dans des départements ruraux tels que la Lozère, la Haute-Loire, le Tarn, la Creuse, le Cantal et dans les départements montagneux du Sud-Est.

Le tableau 7 présente les statistiques descriptives pour ces mesures de concentration sur l'emploi salarié. Le niveau géographique choisi est celui de la zone d'emploi, ainsi les statistiques sont calculées sur les 322 zones d'emploi.

*Tableau 7 - Statistiques descriptives sur les mesures de concentration géographique*

	Moyenne	Écart-type	Minimum	1er Quartile	Médiane	3ième Quartile	Maximum
Établissements	5 875	16 057	242	1 450	2 874	5 718	267 968
Postes	47 144	151 523	804	9 982	18 444	43 470	2 534 884
C4	10,7%	6,3%	1,4%	6,2%	9,0%	13,9%	37,7%
C10	17,6%	8,6%	2,5%	11,1%	15,5%	22,9%	50,5%
Indice Herfindhal	0,7%	0,8%	0,0%	0,3%	0,5%	0,9%	6,2%
Indice Entropie	-6,6	0,9	-10,4	-7,2	-6,6	-5,9	-4,4
Coeff. Concentration	74,6%	3,9%	54,3%	72,3%	75,1%	77,1%	84,3%

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

Champ marchand non agricole hors intérim sur les 322 zones d'emploi

On constate que les 10 plus gros établissements d'une zone d'emploi représentent entre 3 % et 50 % de l'emploi salarié du champ marchand non agricole (et hors établissements d'intérim) des zones. En moyenne ces 10 établissements concentrent environ un sixième de l'emploi de la zone. D'ailleurs les coefficients de concentration de Gini sont assez élevés avec des valeurs comprises entre 54 % et 84 % et une moyenne de 75 %.

### III.5. Le lien entre les indices de concentration et de spécialisation

Le tableau 8 présente les corrélations entre les mesures de concentration (C10, indice de Herfindhal et coefficient de concentration) et les mesures de spécialisation (indice de Krugman et coefficient de spécialisation) auxquelles le logarithme de l'effectif salarié a été ajouté pour tenir compte de la taille des établissements.

*Tableau 8 : Corrélations entre les mesures de spécialisation et de concentration géographique.*

	log(effectif)	Indice de Krugman	Coefficient de spécialisation	C10	Indice de Herfindhal	Coefficient de concentration
log(effectif)	1,000					
Indice de Krugman	-0,738	1,000				
Coefficient de spécialisation	-0,740	0,988	1,000			
C10	-0,629	0,722	0,761	1,000		
Indice de Herfindhal	-0,435	0,591	0,620	0,884	1,000	
Coefficient de concentration	0,552	-0,177	-0,130	0,195	0,257	1,000

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

Champ marchand non agricole (hors intérim) sur les 322 zones d'emploi

Le coefficient de spécialisation et l'indice de spécificité global de Krugman sont naturellement très fortement corrélés. Rappelons qu'ils comparent tous deux la structure de la zone à celle du territoire de référence en termes de concentration des salariés dans les secteurs d'activité.

Parmi les indices de concentration, les indices C10 et de Herfindhal sont très corrélés entre eux, alors qu'ils apparaissent indépendants du coefficient de Gini de concentration. Cela s'explique par le fait que l'indicateur C10 ne tient compte que des 10 plus gros établissements pour mesurer la concentration alors que le coefficient de concentration considère l'ensemble des établissements. Ainsi l'indicateur C10 est fortement lié au nombre d'établissements présents dans la zone : il sera mécaniquement plus élevé dans les zones comptant peu d'établissements et inversement. Par exemple, c'est la zone d'emploi de Paris qui enregistre le C10 le plus faible avec seulement 2,5 % des salariés (contre 17,6 % en moyenne) dans les 10 plus gros établissements, alors que le coefficient de concentration place cette zone en 5<sup>ème</sup> position. Le problème existe aussi dans une moindre mesure avec l'indice de Herfindhal, qui est très sensible aux valeurs élevées du poids des établissements (le poids de l'établissement dans l'emploi ou la rémunération de la zone est élevé au carré). Or ces valeurs élevées sont peu nombreuses dans les grosses zones puisque les salariés sont répartis sur un nombre très important d'établissements.

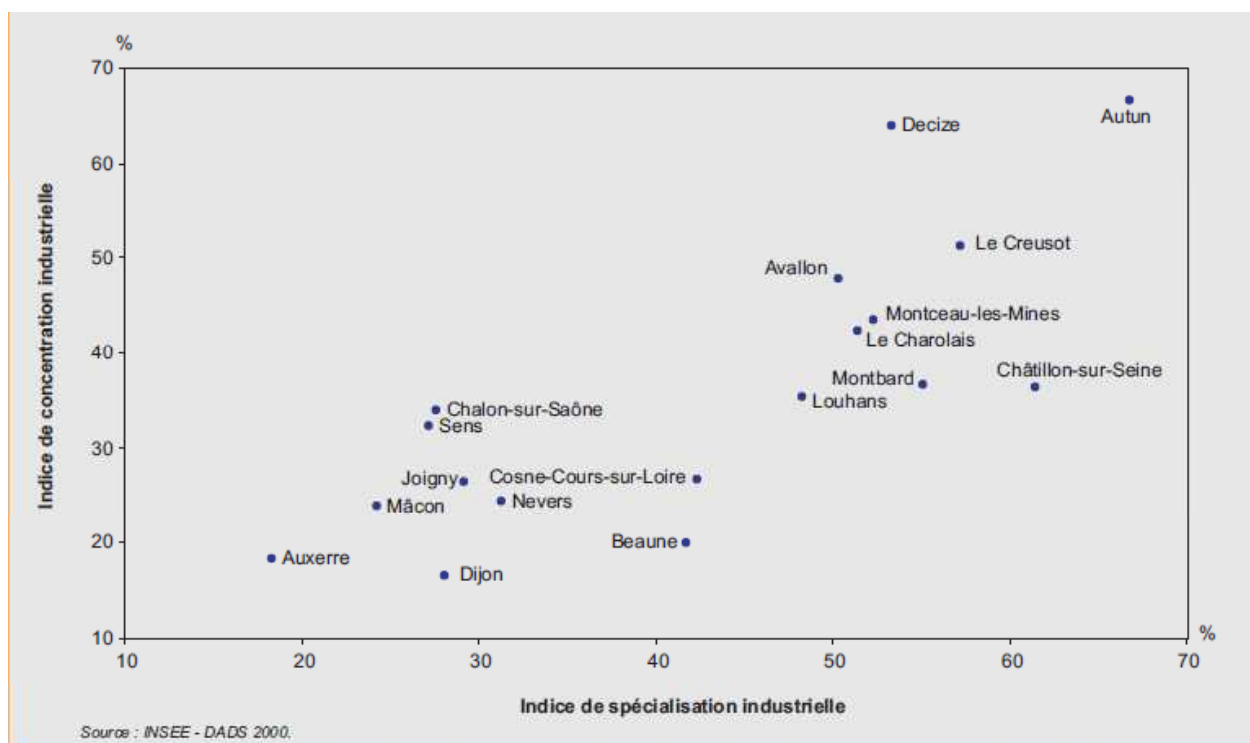
La corrélation positive entre l'indice de concentration et l'effectif salarié de la zone s'explique par le fait que les gros établissements, qui contribuent à augmenter l'indice de concentration, sont souvent situés dans des zones à forte agglomération d'activités comme celles de Paris, Saclay, Roissy-Sud Picardie, Lille, Toulouse. Les autres indicateurs de concentration sont corrélés négativement avec l'effectif salarié de la zone. Cela tient au fait qu'ils sont mécaniquement plus faibles dans les zones regroupant beaucoup d'établissements et donc beaucoup de salariés (cf. ci-dessus).

Les indicateurs de spécialisation ont logiquement un lien négatif avec les effectifs salariés. En effet, plus une zone est grande, plus elle compte de salariés et plus elle a de chances d'avoir une activité diversifiée.

La forte corrélation positive entre C10 et Herfindhal et les indicateurs de spécialisation est liée au fait que lorsque l'activité d'une zone est fortement concentrée dans quelques établissements, elle a de grandes chances d'être spécialisée dans un petit nombre de secteurs. On ne retrouve pas le même lien entre concentration et spécialisation lorsqu'on mesure la concentration avec le coefficient de Gini. Les zones très concentrées au sens du coefficient de Gini peuvent être des zones importantes comportant un grand nombre de gros établissements et donc potentiellement un grand éventail d'activités.

Le graphique ci-dessous extrait d'une publication de la Direction régionale de Bourgogne de Novembre 2004, « Spécialisation, concentration et dépendance de l'emploi industriel en Bourgogne » illustre la corrélation forte entre la spécialisation et la concentration mesurée par l'indicateur C4 dans les zones d'emploi de Bourgogne. Ainsi, les zones d'emploi spécialisées sont également concentrées et celles non spécialisées ne sont pas non plus concentrées (selon C4).

Graphique 3 : Lien entre l'indice de spécialisation et de concentration (région Bourgogne sur les 18 zones d'emploi)



### III.6. L'indice de densité économique des zones

*Quel est le lien entre la densité de population et la densité d'emploi dans la zone ?*

Une autre manière d'aborder les caractéristiques de la structure productive locale est de mesurer la **densité économique** d'une zone. Cette densité économique est fortement corrélée avec la densité de population car une forte activité économique locale va de pair avec la présence d'une population active importante et donc de nombreux habitants. Il faut cependant noter que le concept de densité économique repose sur le lieu de travail d'une personne (salariée), alors que la densité de population tient compte du lieu de résidence. La différence entre les deux concepts provient des migrations alternantes qui peuvent alimenter en force de travail une zone à faible densité de population.

Cette forte corrélation conduit à s'intéresser plutôt au rapport entre les deux densités, c'est-à-dire le taux d'emploi de la zone. Celui-ci montre des différences sensibles entre les zones.

$$TE = \frac{\text{Emploi}}{\text{Population}} = \frac{D^E}{D^P}$$

Il faut interpréter cet indicateur avec prudence pour plusieurs raisons. Tout d'abord, le numérateur ne tient compte que des salariés et pas des indépendants, des intérimaires et des actifs non-salariés. Par ailleurs, la structure par âge de la population peut expliquer une partie des différences observées entre les zones : par exemple, une zone habitée par une population âgée aura mécaniquement un taux d'emploi plus faible. Les interprétations semblent plus aisées à des niveaux géographiques plus fins que la région du type zone d'emploi, le niveau régional pouvant masquer de fortes disparités. Enfin l'utilisation de cet indicateur au niveau des catégories socioprofessionnelles peut révéler des différences entre les zones plus importantes que celles observées pour l'ensemble des salariés.

En moyenne, la densité économique de la France métropolitaine est de 40,8 salariés par km<sup>2</sup>, alors que la densité de population s'élève à 116,8 habitants par km<sup>2</sup>. Cet écart s'explique par le taux d'activité global en France. On observe des disparités très fortes entre les régions : de 11 salariés par km<sup>2</sup> en Corse à 451 en Île-de-France. Cependant s'il y a un facteur de 28 entre la densité de



population la plus élevée (Île-de-France) et la moins élevée (Corse), ce facteur passe à 40 en termes de densité économique. Cet écart montre que certaines régions ont une part plus importante d'actifs salariés dans leur population ou encore qu'elles attirent des travailleurs résidant dans les autres régions.

*Tableau 9 : Densité de population et densité économique par région*

REGION	Densité de population	Densité économique	Taux d'emploi salarié
11 ILE-DE-FRANCE	982,6	451,5	45,9%
21 CHAMPAGNE-ARDENNE	53,7	17,0	31,6%
22 PICARDIE	100,7	29,3	29,1%
23 HAUTE-NORMANDIE	151,6	49,4	32,6%
24 CENTRE	66,5	20,9	31,5%
25 BASSE-NORMANDIE	85,9	26,8	31,1%
26 BOURGOGNE	53,5	17,0	31,8%
31 NORD-PAS-DE-CALAIS	329,8	106,3	32,2%
41 LORRAINE	102,0	30,0	29,4%
42 ALSACE	225,8	79,0	35,0%
43 FRANCHE-COMTE	74,1	22,8	30,7%
52 PAYS DE LA LOIRE	112,6	38,7	34,3%
53 BRETAGNE	119,4	37,9	31,8%
54 POITOU-CHARENTES	70,0	21,5	30,7%
72 AQUITAINE	79,1	25,6	32,4%
73 MIDI-PYRENEES	64,4	21,4	33,2%
74 LIMOUSIN	45,0	13,6	30,3%
82 RHONE-ALPES	143,4	51,6	36,0%
83 AUVERGNE	53,2	16,6	31,2%
91 LANGUEDOC-ROUSSILLON	96,3	28,3	29,3%
93 PROVENCE-ALPES-COTE D'AZUR	158,1	51,8	32,8%
94 CORSE	35,5	11,3	31,9%
<b>France métropolitaine</b>	<b>116,8</b>	<b>40,8</b>	<b>34,9%</b>
Moyenne	145,6	53,1	32,5%
Écart-Type	198,6	91,9	3,5%
Minimum	35,5	11,3	29,1%
Maximum	982,6	451,5	45,9%

## IV. Répartition spatiale des activités d'un secteur de l'économie

Dans cette section, la perspective est renversée car, au lieu de porter l'analyse sur les différentes zones d'un territoire, on caractérise la répartition de l'emploi des différents secteurs sur le territoire. On mesure tout d'abord le niveau de concentration de l'activité d'un secteur afin de déterminer si le secteur d'activité est dominé par quelques gros établissements : on parle alors de concentration productive. On étudie ensuite comment ces emplois se répartissent dans les différentes zones.

### IV.1. La concentration productive des établissements

*Quel est le degré de concentration productive d'un secteur sur un territoire ?  
Est-ce qu'un secteur est dominé par quelques grands établissements ?*

Plusieurs mesures peuvent être utilisées afin d'évaluer le degré de concentration dans chacun des secteurs d'activité. La concentration productive est très souvent utilisée en matière de description des secteurs économiques du fait qu'elle révèle la structure de marché de ce secteur et qu'elle influence la politique de la concurrence. On se pose la question de savoir comment est distribué l'emploi salarié dans les établissements du secteur et quelle est l'influence des plus gros établissements dans ce secteur. Ces mesures de concentration productive sont présentées dans le § II.3. Il s'agit de la part des 4 ou des 10 plus grands établissements (C4 ou C10), de l'indice de Herfindhal, de l'indice d'entropie ou encore du coefficient de Gini de concentration.

A la différence des indices de concentration géographique qui mesurent la dépendance d'une zone au nombre d'établissements et à la répartition des emplois dans ces établissements quel que soit leur secteur d'activité, on considère ici uniquement la concentration productive des emplois dans les établissements d'un même secteur (à un niveau plus ou moins fin).

Ces indicateurs de concentration productive sont indépendants du niveau géographique retenu.

*Tableau 10 : Mesures de concentration productive - statistiques descriptives*

	Nombre d'Établis.	C 10 (en %)	Indice de Herfindhal (en %)	Indice d'Entropie (en %)	Coefficient Concentrat. (en %)
<b>A38</b>					
Moyenne	63 551	8,9	0,4	-8,3	76,8
Écart-type	91 880	11,8	1,1	1,9	6,7
Minimum	95	0,6	0,0	-11,8	59,2
Médiane	23 043	5,5	0,1	-8,4	76,6
Maximum	456 320	64,8	6,0	-3,3	93,2
<b>A88</b>					
Moyenne	27 342	16,3	2,1	-7,3	74,4
Écart-type	48 055	22,3	7,6	2,2	9,5
Minimum	4	0,7	0,0	-11,3	41,7
Médiane	8 028	8,1	0,2	-7,5	76,4
Maximum	285 746	100,0	52,3	-1,0	93,2
<b>NAF732</b>					
Moyenne	3 243	33,3	4,7	-5,3	69,0
Écart-type	7 667	28,5	10,8	2,1	11,3
Minimum	1	0,4	0,0	-10,7	0,0
Médiane	681	24,1	1,1	-5,3	70,3
Maximum	86 681	100,0	100,0	0,0	97,1

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

Champ marchand non agricole (hors intérim)

Le Tableau 10 donne quelques statistiques descriptives de certaines mesures de concentration productive pour 3 classifications sectorielles : A38, A88 et NAF 732.

On remarque que plus la classification sectorielle est détaillée, plus les mesures de concentration productive sont en moyenne importantes, sauf pour le coefficient de Gini de concentration. En effet, le nombre d'établissements par secteur d'activité est sensiblement plus faible pour des classifications sectorielles détaillées et on aura dès lors davantage de concentration dans ces secteurs avec un effectif d'établissements faible. De plus, la présence d'un établissement dominant dans le secteur est plus probable avec ce type de classification détaillée. Ce qui explique la progression des mesures de concentration lorsque l'on distingue des secteurs à un échelon plus fin.

La médiane de la part des 10 plus grands établissements de chacun des secteurs est de 5,5 % en au niveau A38, 8,1 % en A88 et atteint 24 % en NAF 732. Ce qui veut dire que près d'un quart des emplois sont regroupés dans les 10 plus gros établissements pour plus de la moitié des secteurs de la NAF 732.

Le coefficient de Gini de concentration productive a un comportement un peu différent car il est moins sensible au nombre d'établissements du secteur concerné. Il considère davantage la répartition par taille de ces établissements, mettant plus de poids sur la présence d'un petit nombre de gros établissements. C'est pourquoi ce coefficient de concentration est assez peu sensible en moyenne à la classification sectorielle utilisée. Son écart-type varie également peu en fonction de la classification utilisée.

Ceci est confirmé par le tableau 11 de corrélation entre ces différentes mesures de concentration productive. Il y a une très forte corrélation entre l'indicateur C10 et les indices de Herfindhal et d'Entropie qui indiquent le même phénomène de concentration. Cependant la corrélation, même si elle est encore positive, est nettement plus faible entre ces mesures et le coefficient de Gini de concentration, indiquant deux manières de mesurer la concentration productive des établissements du secteur. Encore une fois, ce coefficient de concentration tient compte de l'ensemble de la répartition des établissements, alors que C10, de même que les indices de Herfindhal ou d'Entropie, donnent davantage de poids aux gros établissements du secteur. Finalement toutes ces mesures de concentration productive sont négativement corrélées avec le nombre d'établissements du secteur. Ainsi un secteur avec peu d'établissements sera automatiquement plus concentré qu'un secteur où il y a davantage d'établissements.

*Tableau 11 : Corrélations entre les mesures de concentration productive*

	Nombre d'Établ.	C 10	Indice de Herfindhal	Indice d'Entropie	Coefficient Concentr.
Nombre d'Établissements	1,000				
C 10	-0,387	1,000			
Indice de Herfindhal	-0,173	0,701	1,000		
Indice d'Entropie	-0,607	0,932	0,636	1,000	
Coef. Concentration	-0,156	0,227	-0,027	0,257	1,000

Source : CLAP2010

Variable : effectifs salariés au 31/12

Niveau sectoriel : NAF732

## IV.2. La répartition spatiale des activités

*Y-a-t-il une concentration géographique de certaines activités économiques ?  
Quelle est la répartition spatiale des établissements d'un secteur ?*

Il s'agit de la répartition des emplois d'un secteur dans un nombre plus ou moins grand de zones (voir § II.3.5). Le tableau 12 présente les moyennes et les écarts-type pour ce coefficient de répartition calculé pour la France métropolitaine suivant le type de zonage (région, département ou zone d'emploi) et le niveau de classification sectoriel (A38, A88 et NAF732).

On retrouve encore le résultat habituel que plus la classification géographique ou sectorielle est détaillée, plus le coefficient de répartition est en moyenne élevé. On peut toutefois noter que la localisation des activités au niveau le plus fin (NAF 732) est relativement concentrée dans quelques zones d'emploi avec un coefficient moyen assez élevé (0,646) ce qui montre de nouveau la spécialisation relativement importante au niveau des zones d'emploi. Cependant il existe des larges différences entre les secteurs d'activités comme l'indiquent les écarts-type de la distribution de ces coefficients de répartition spatiale.

*Tableau 12 : Coefficient de répartition, moyenne et écart-type (entre parenthèses)*

	REGION	DEPARTEMENT	ZONE EMPLOI
A38	0,246 (0,151)	0,344 (0,192)	0,393 (0,203)
A88	0,309 (0,213)	0,419 (0,235)	0,471 (0,243)
NAF 732	0,430 (0,234)	0,584 (0,249)	0,646 (0,247)

Source : CLAP 2010

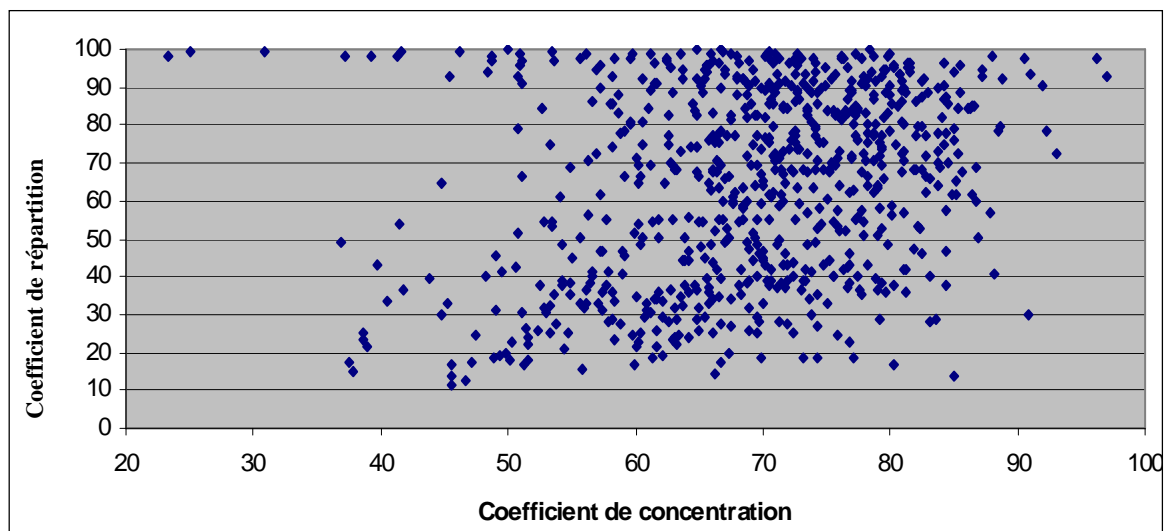
Variable : Effectifs salariés au 31/12

22 Régions - 96 Départements - 322 Zones d'Emploi

## IV.3. Le lien entre la concentration productive et la répartition spatiale des activités

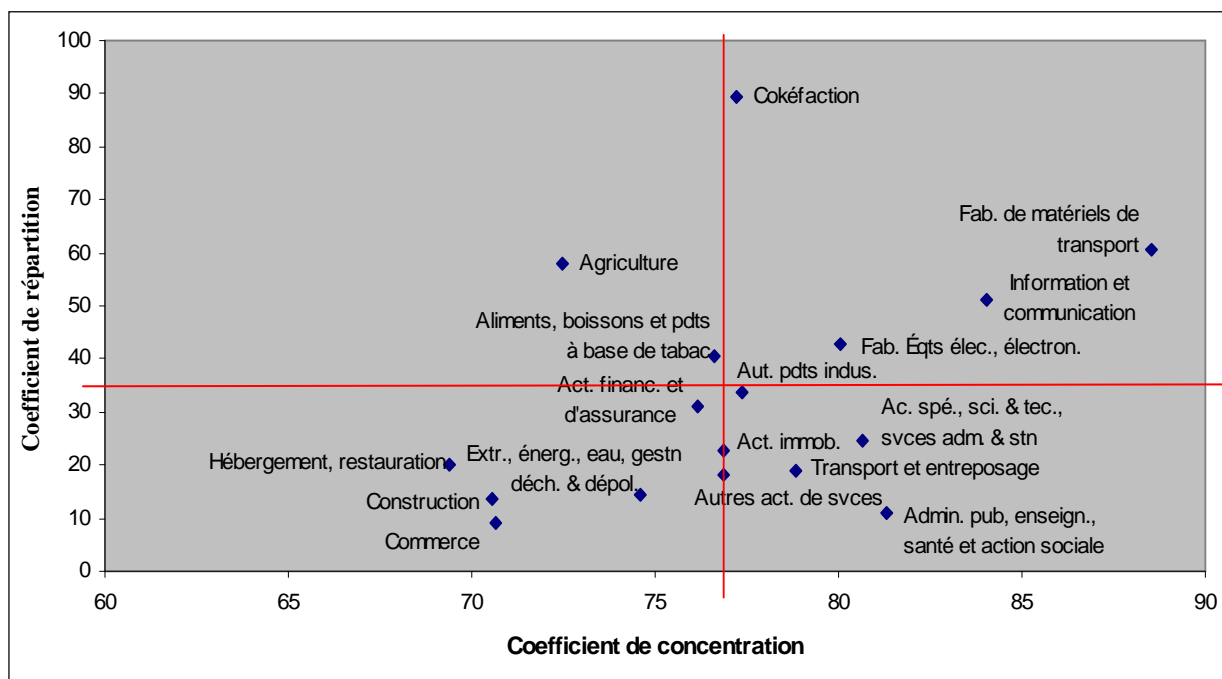
On présente dans les graphiques 4 et 5 la relation entre la répartition géographique et la concentration productive des établissements d'un secteur. Ces graphiques sont réalisés en utilisant les effectifs salariés au 31/12, avec un découpage en zones d'emploi pour le calcul de la répartition géographique des activités. Le graphique 4 montre le nuage de points des secteurs de la NAF732 pour le coefficient de concentration productive et le coefficient de répartition spatiale. En abscisse on mesure la concentration de l'emploi des différents secteurs dans les établissements, en ordonnée on a une mesure de la concentration de l'emploi des différents secteurs dans les zones d'emploi (répartition spatiale des activités).

*Graphique 4 : Répartition spatiale des activités et concentration productive*



Ce nuage de points, représentant chacun un secteur, ne possède pas de structure particulièrement marquée avec une faible corrélation de 25 % entre le coefficient de répartition et le coefficient de concentration productive de l'emploi. Il y a une légère liaison positive entre la concentration géographique et la concentration productive des secteurs. On remarque également de nombreux secteurs avec une répartition géographique assez inégale (coefficient de répartition supérieur à 75 % par exemple), pour un ensemble de valeurs de coefficient de concentration assez large. Ainsi ce n'est pas parce que la concentration productive est élevée, que la concentration géographique est nécessairement forte. Pour étudier en détail ce phénomène, on présentera dans le § IV.4 une mesure de l'agglomération des établissements qui tient compte de ces effets de taille.

*Graphique 5 : Répartition des activités et concentration productive*



Le graphique 5 répète l'exercice pour une classification sectorielle plus agrégée au niveau A 17. Ici la corrélation entre les deux coefficients devient plus importante avec 38%. On remarque premièrement que le coefficient de répartition géographique est nettement plus faible que précédemment, avec moins de valeurs élevées, la moyenne se situant autour de 33 % (contre 65 % en NAF732), alors que les valeurs pour le coefficient de concentration productive conservent une moyenne proche (environ 77 % contre 70 % en NAF732).

En considérant les moyennes de ces coefficients, on peut découper le graphique en 4 parties avec en bas à gauche des secteurs qui sont peu concentrés et bien répartis sur le territoire. On y retrouve la construction, le commerce, l'hébergement et la restauration. Ce sont les activités proches des consommateurs. A l'opposé, en haut à droite, on a les secteurs de la fabrication de matériels de transport et d'information et de communication qui sont fortement concentrés productivement dans quelques établissements, et géographiquement dans quelques zones.

Un secteur très concentré horizontalement mais très étalé géographiquement est le secteur de l'administration publique, l'enseignement, la santé et l'action sociale. Au contraire le secteur agricole est bien réparti dans les établissements mais concentré autour de quelques zones.

#### IV.4. La mesure d'agglomération

*Comment mesurer le degré d'agglomération d'un secteur sur un territoire ?  
Comment tester si les établissements sont répartis au hasard sur le territoire ?*

Après avoir mesuré la répartition géographique et la concentration productive des établissements d'un secteur d'activité, on voudrait prolonger l'analyse en déterminant si un secteur d'activité est plus ou moins fortement localisé ou aggloméré<sup>4</sup>. En effet, l'analyse de la répartition spatiale peut être trompeuse car elle est fortement conditionnée par la concentration productive et par le nombre d'établissements du secteur. Si le secteur est horizontalement concentré, cela veut dire qu'il y est dominé par quelques gros établissements. Dans ce cas, il est probable que ce secteur sera aussi fortement concentré géographiquement dans quelques zones. A la limite, si on considère un secteur avec uniquement deux établissements, le degré de concentration géographique de ce secteur sera aussi élevé car les deux établissements se localiseront dans une ou deux zones du territoire, alors que les autres zones n'auront aucun établissement de ce secteur.

Afin d'éviter que la mesure de concentration géographique ne dépende trop de la structure productive du secteur, deux auteurs américains, Glenn Ellison et Edward Glaeser ont développé dans les années 90 une mesure d'agglomération des établissements sur un territoire. La méthode proposée cherche à mesurer l'agglomération des activités sur un territoire en tenant compte de la répartition par taille des établissements dans un secteur. En effet, si la technologie de production requiert seulement quelques établissements de grande taille dans un secteur, celui-ci sera souvent considéré comme plus localisé et plus aggloméré qu'un secteur avec une multitude de petits établissements, si on prend les mesures habituelles de répartition géographique sur le territoire. Afin de neutraliser cet effet de la concentration horizontale dans un secteur, Ellison et Glaeser ont proposé un indicateur basé sur deux modèles alternatifs de choix de localisation. Le premier formalise la notion d'avantages naturels (disponibilité d'une ressource ou d'une situation naturelle), alors que le second modèle tient compte des externalités de localisation des entreprises, c'est-à-dire des avantages non marchands à la constitution de pôles d'entreprises. Cependant ces deux explications théoriques impliquent les mêmes mesures de l'agglomération des établissements - voir § II.5.

On produit la mesure de l'agglomération des activités proposées par Ellison et Glaeser pour la France, en utilisant la variable « effectifs salariés au 31/12 » de CLAP 2010 par établissement. On donne dans le tableau 13 les statistiques descriptives suivant les classifications sectorielles et géographiques habituelles, pour la mesure de concentration géographique brute de Ellison et Glaeser  $G_{EG}^k$  qui s'interprète comme la probabilité que deux établissements du même secteur se localisent dans la même zone, ainsi que sur la corrélation de localisation  $\hat{\gamma}_{EG}^k$  qui mesure la corrélation entre les choix de localisation de deux établissements du même secteur.

---

<sup>4</sup> Ces deux termes seront utilisés indifféremment ici. Un secteur avec une agglomération d'établissements importante sera également un secteur où les établissements seront fortement localisés dans certaines zones. On pourra parfois parler de polarisation sectorielle pour certaines zones.

*Tableau 13 : Mesures d'agglomération, moyenne et écart-type (entre parenthèses)*

	REGION	DEPARTEMENT	ZONE EMPLOI
Concentration géographique brute $G_{EG}^k$			
A38	0,037 (0,055)	0,027 (0,064)	0,032 (0,060)
A88	0,075 (0,155)	0,053 (0,118)	0,056 (0,101)
NAF 732	0,104 (0,142)	0,086 (0,133)	0,084 (0,130)
Corrélation de localisation $\hat{\gamma}_{EG}^k$			
A38	0,034 (0,051)	0,023 (0,060)	0,028 (0,056)
A88	0,060 (0,035)	0,032 (0,093)	0,036 (0,067)
NAF 732	0,058 (0,130)	0,035 (0,086)	0,039 (0,079)

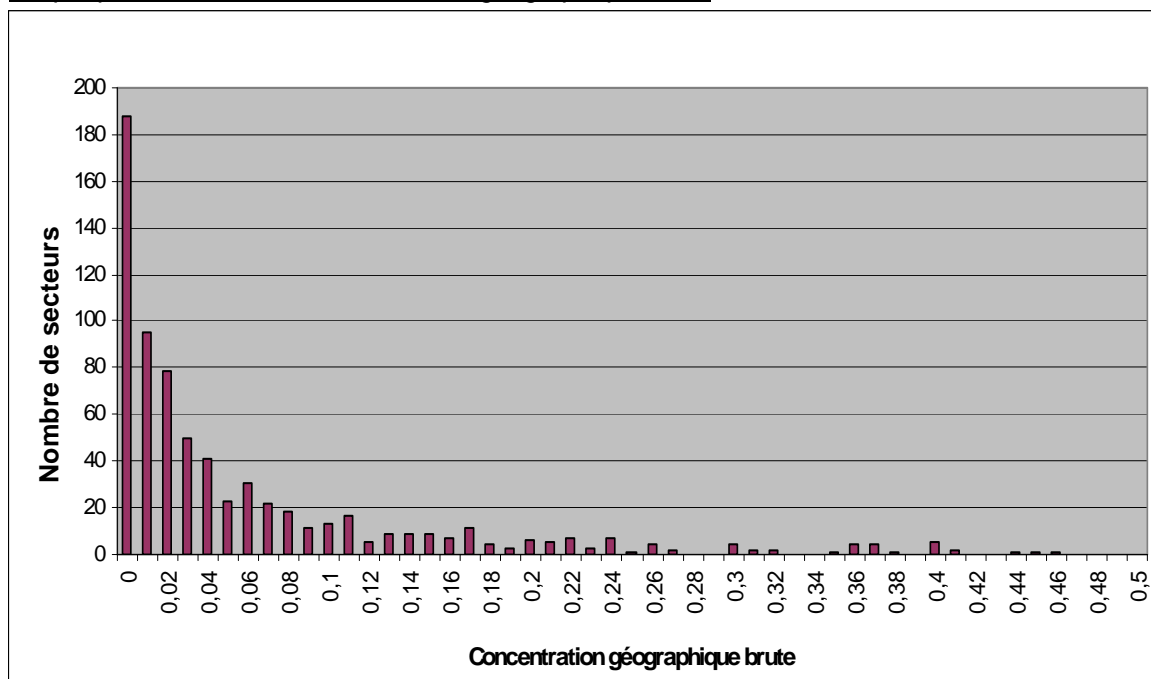
Source : CLAP 2010

Variable : Effectifs salariés au 31/12

26 Régions - 100 Départements - 322 Zones d'Emploi

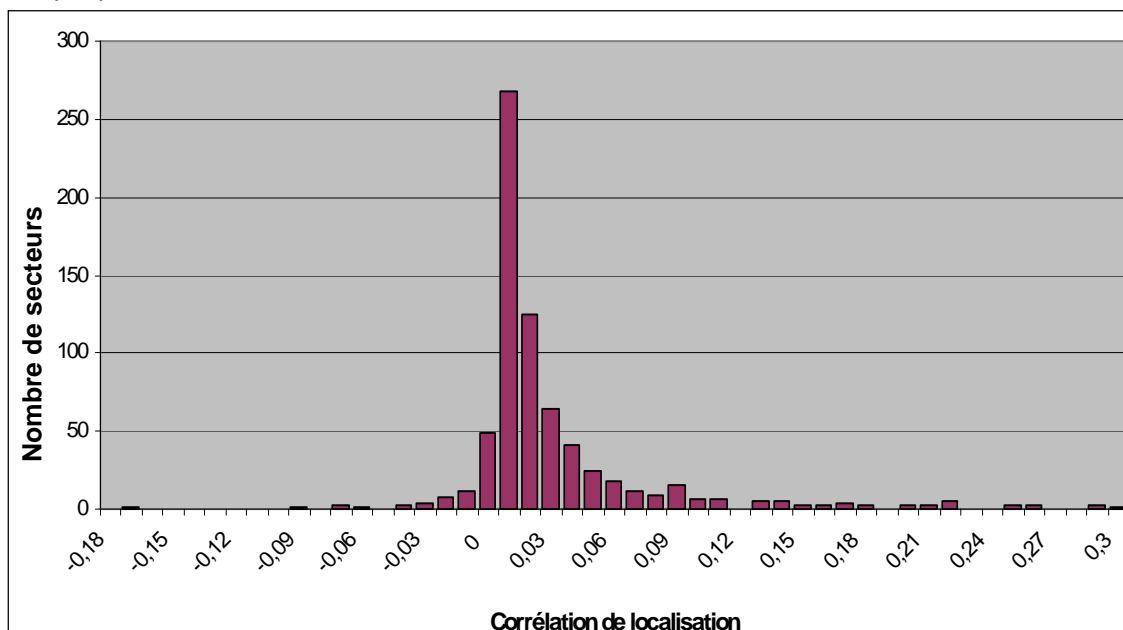
La probabilité pour que deux établissements d'un même secteur se localisent dans la même région, le même département ou la même zone d'emploi reste faible avec moins de 8 % en moyenne au niveau A88 pour les régions. Cependant, vu les écarts-type de la mesure de concentration géographique brute, on peut supposer que la distribution de ces probabilités est très asymétrique, ce qui est confirmé par le graphique 6, réalisé au niveau départemental en NAF732. Cette probabilité moyenne augmente avec la finesse du niveau d'agrégation sectorielle. En effet, un niveau sectoriel très fin permet d'identifier des activités spécifiques souvent localisées, par exemple la construction aéronautique et spatiale (NAF732) dans la fabrication de matériel de transport (A38). On voit aussi une augmentation de la probabilité de co-localisation lorsque l'on passe à un niveau régional indiquant une agglomération plus importante pour des zones de taille plus importante.

*Graphique 6 : Mesure de concentration géographique brute*



En ce qui concerne la corrélation de localisation  $\hat{\gamma}_{EG}^k$ , on trouve des valeurs relativement faibles n'excédant pas 6 % en moyenne pour les plus élevées (voir tableau 13). On a également une diminution de cette corrélation lorsque l'on va vers des découpages territoriaux plus fins. Cela indique que l'agglomération des activités est un peu plus importante au niveau régional, qu'au niveau des zones d'emploi : les activités s'agglomèrent davantage au niveau régional avec une répartition dans les différentes zones d'emploi. La distribution de ces corrélations par département (graphique 7) est également fortement asymétrique, même si on obtient ici des valeurs négatives pouvant indiquer que les établissements tendent à se disperser davantage que la répartition uniforme sur le territoire le voudrait. On remarque aussi une forte concentration de ces corrélations sur des valeurs faibles de l'ordre de 1 à 3 %.

*Graphique 7 : Mesure de corrélation de localisation*



Ellison-Glaeser propose, sans aucune justification théorique, de déterminer qu'un secteur est fortement localisé lorsque sa corrélation de localisation est supérieure à 5 %, alors que si cette corrélation est comprise entre 2 % et 5 %, le secteur sera dit peu aggloméré. Une valeur inférieure à 2 % indiquerait un secteur dont les établissements ne sont pas agglomérés (ou faiblement concentrés sur le territoire).

En appliquant cette règle, on obtient dans le tableau 14 un nombre relativement important de secteurs fortement localisés dans l'Industrie ou l'Agriculture, alors que la Construction ou le Commerce sont plutôt des secteurs faiblement concentrés géographiquement. Toutefois cette dernière catégorie est la plus importante quelle que soit le grand secteur considéré. Ces résultats sont proches de ceux de Maurel et Sédillot qui trouvent que 23 % des secteurs de l'Industrie sont assez localisés et 27 % fortement localisés.

Cependant à la suite des travaux utilisant cette méthodologie, on conseille souvent de comparer les résultats obtenus suivant différents niveaux de zonage et suivant différentes classifications d'activité afin de choisir les conclusions les plus robustes sur la localisation et l'agglomération des activités.



**Tableau 14 : Corrélation de Localisation, calculée par zone d'emploi en NAF 732**  
(Nombre et pourcentage du secteur dans chaque type de localisation)

SECTEUR	Dispersé	Faible Concentration	Assez Localisé	Fortement Localisé	Ensemble
	$\hat{\gamma}_{EG}^k < 0$	$0 \leq \hat{\gamma}_{EG}^k < 2\%$	$2\% \leq \hat{\gamma}_{EG}^k < 5\%$	$\hat{\gamma}_{EG}^k \geq 5\%$	
Agriculture	6 (16%)	6 (16%)	17 (46%)	8 (22%)	37
Industrie	27 (9%)	88 (30%)	121 (42%)	53 (18%)	289
Construction	2 (5%)	31 (82%)	2 (5%)	3 (8%)	38
Commerce	9 (8%)	80 (69%)	20 (17%)	7 (6%)	116
Services	22 (9%)	119 (51%)	30 (13%)	63 (27%)	234
Administration	5 (45%)	5 (45%)	(0%)	1 (9%)	11
Ensemble	71 (10%)	329 (45%)	190 (26%)	135 (19%)	725

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

Finalement on présente dans le tableau 15 la liste des 20 secteurs de la NAF 732 qui sont les plus agglomérés au niveau zone d'emploi en utilisant le coefficient de corrélation de localisation de Ellison et Glaeser. Toutes ces corrélations dépassent 29 %, ce qui est relativement élevé. On remarque que la majorité de ces activités appartiennent au secteur des services (16 sur 20). Ces activités sont fortement localisées du fait de leur caractéristique productive très centralisée dans le domaine financier ou culturel. Toutefois, plus de la moitié des activités des services sont faiblement concentrées. Il s'agit en particulier de tous les services de la sphère présenteielle, directement destinés à la population, et donc bien répartis sur le territoire.

**Tableau 15 : Liste des 20 secteurs les plus agglomérés selon Ellison-Glaeser**

NAF732	$G_{EG}^k$	$\hat{\gamma}_{EG}^k$	Etabl.	Effectifs
2013A Enrichissement & retrait. mat. nucléaire	0,7619	0,6652	7	2 950
0123Z Culture d'agrumes	0,5993	0,5726	47	269
5911C Production de films pour le cinéma	0,5372	0,5340	2 028	7 525
5121Z Transports aériens de fret	0,6251	0,4596	35	3 352
6612Z Courtage valeur mobilière & marchandise	0,4891	0,4591	271	3 632
6391Z Activités des agences de presse	0,4526	0,4241	500	5 629
0112Z Culture du riz	0,4371	0,4179	54	124
6520Z Réassurance	0,6307	0,4128	42	5 810
5814Z Édition de revues et périodiques	0,4074	0,4013	2 348	34 178
5913A Distribution de films cinématographiques	0,4082	0,3944	181	1 264
5912Z Post-production film & prog. télévision	0,3796	0,3715	884	9 929
5920Z Enregistrement sonore & édition musicale	0,3851	0,3674	1 550	3 801
6530Z Caisses de retraite	0,5054	0,3574	16	1 425
5911A Prod. film & prog. pour la télévision	0,3586	0,3521	2 296	12 992
6630Z Gestion de fonds	0,3460	0,3376	1 798	17 846
5811Z Édition de livres	0,3403	0,3312	1 373	13 694
2391Z Fabrication de produits abrasifs	0,4934	0,3300	37	1 812
6020B Edition de chaînes thématiques	0,3521	0,3243	152	6 147
9900Z Act. organisations extraterritoriales	0,3149	0,2967	398	4 261
5110Z Transports aériens de passagers	0,3430	0,2935	566	70 619

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

En ce qui concerne l'industrie, les secteurs les plus agglomérés sont ceux qui ont un avantage naturel évident à se localiser géographiquement dans la même zone : Enrichissement & retraitement de matières nucléaires (2013A), Fabrication de vins effervescents (1102A), Extraction de houille (0510Z). On retrouve également les secteurs de vieille tradition industrielle comme la Fabrication de coutellerie (2571Z), les activités textiles (1411Z, 1420Z ou 1396Z), l'horlogerie (2652Z).

Le tableau 16 montre qu'on obtient sensiblement les mêmes résultats en utilisant les estimateurs de Maurel et Sédillot. 17 des 20 secteurs les plus agglomérés sont les mêmes avec les deux méthodes. Les estimateurs des corrélations de localisation  $\hat{\gamma}_{EG}$  et  $\hat{\gamma}_{MS}$  sont très corrélés ( $\rho=0,90$ ).

*Tableau 16 : Liste des 20 secteurs les plus agglomérés selon Maurel-Sédillot*

	<b>NAF732</b>	<b>G_MS</b>	<b>Gamma_MS</b>	<b>Etabl.</b>	<b>Effectifs</b>
6520Z	Réassurance	0,8681	0,7902	42	5 810
5911C	Production de films pour le cinéma	0,7462	0,7444	2 028	7 525
6612Z	Courtage valeur mobilière & marchandise	0,6908	0,6727	271	3 632
6530Z	Caisses de retraite	0,7188	0,6345	16	1 425
6391Z	Activités des agences de presse	0,6467	0,6282	500	5 629
2013A	Enrichissement & retrait. mat. nucléaire	0,7241	0,6121	7	2 950
5913A	Distribution de films cinématographiques	0,5949	0,5855	181	1 264
5814Z	Édition de revues et périodiques	0,5889	0,5847	2 348	34 178
0123Z	Culture d'agrumes	0,6003	0,5736	47	269
5920Z	Enregistrement sonore & édition musicale	0,5626	0,5499	1 550	3 801
5912Z	Post-production film & prog. télévision	0,5544	0,5486	884	9 929
5911A	Prod. film & prog. pour la télévision	0,5280	0,5232	2 296	12 992
6020B	Edition de chaînes thématiques	0,5291	0,5089	152	6 147
6630Z	Gestion de fonds	0,5136	0,5073	1 798	17 846
5811Z	Édition de livres	0,5063	0,4995	1 373	13 694
5121Z	Transports aériens de fret	0,6504	0,4961	35	3 352
9900Z	Act. organisations extraterritoriales	0,4833	0,4696	398	4 261
9102Z	Gestion des musées	0,4881	0,4604	393	5 933
6130Z	Télécommunications par satellite	0,5100	0,4422	64	2 221
6611Z	Administration de marchés financiers	0,5170	0,4266	40	802

Source : CLAP 2010

Variable : Effectifs salariés au 31/12

On peut également étudier l'agglomération des activités d'un secteur sur le territoire régional, en utilisant les fichiers CLAP régionaux. On mesure alors l'agglomération dans les différentes zones (zones infra-régionales : départements, zones d'emploi....) des activités régionales d'un secteur. Pour que la corrélation de localisation ait un sens, on recommande d'utiliser un découpage assez fin du niveau régional. Il serait par exemple peu pertinent d'étudier l'agglomération des activités dans les départements d'une petite région (deux ou trois départements seulement).

## Bibliographie

Arbia, G. et Espa, G. (1996) *Statistica Economica Territoriale*. Padua : Cedam.

Duranton, Gilles et Henry Overman (2001) : "Localisation in UK Manufacturing Industries: Assessing Non-Randomness Using Micro-Geographic Data", mimeo, London School of Economics.

Duranton, Gilles et Henry Overman (2005) : "Testing for localization using micro-geographic data", *Review of Economic Studies*, 72(4) pp. 1077-1106.

Ellison, Glenn et Edward Glaeser (1994) : "Geographic Concentration in U.S. Manufacturing Industries : A Dartboard Approach", NBER Working Paper N°4840.

Ellison, Glenn et Edward Glaeser (1997) : "Geographic Concentration in U.S. Manufacturing Industries : A Dartboard Approach », *Journal of Political Economy*, Vol. 105, N°5, pp. 889-927.

Jayet, Hubert (1993) : *Analyse Spatiale Quantitative : une introduction*, Paris : Economica.

Krugman, Paul (1991) : *Geography and Trade*, Leuven: Leuven University Press et Cambridge, MA : The MIT Press.

Marcon, Eric, Florence Puech (2010) : « Measures of the geographic concentration of industries : improving distance-based methods », *Journal of Economic Geography* 10(5) pp. 745-762.

Marcon, Eric, Florence Puech (2012) : « A typology of distance-based measures of spatial concentration », <http://halshs.archives-ouvertes.fr/halshs-00679993>

Maurel, Françoise et Béatrice Sédillot (1997) : « La concentration géographique des industries françaises », *Économie et Prévision*, 1997-5, N°131, pp. 25-45.

Maurel, Françoise et Béatrice Sédillot (1997) : « A measure of the Geographic Concentration in French Manufacturing Industries », *Regional Science and Urban Economics*, 1999, Vol. 29, pp. 575-604.

Saporta, Gilbert (1990) « Probabilités, Analyse des Données et Statistique », Paris : Editions Technip.

Theil, Henri (1967) : *Economics and Information Theory*, Amsterdam.

### Publications nationales de l'INSEE

Floch, Jean-Michel (2012) : « [Détection des disparités socio-économiques - L'apport de la statistique spatiale](#) », *Document de travail de la DDAR*, H2012/04.

Houdebine, Michel (1999) : « [Concentration géographique des activités et spécialisation des départements français](#) », *Economie et Statistique*, 1999-6/7, N°326-327, pp. 189-204.

Lainé, Frédéric et Carole Rieu (1999a) : « [Le tissu productif régional : diversité et concentration](#) », *INSEE Première*, N°630 (Janvier).

Lainé, Frédéric et Carole Rieu (1999b) : « [La diversité industrielle des territoires](#) », *INSEE Première*, N°650 (Juin).

Utilisations des mesures de concentration et spécialisation dans les publications régionales de l'INSEE, quelques exemples

« [L'industrie en Bourgogne : diversifiée avec des spécialisations locales](#) », Insee Bourgogne Dimensions n°179, 2012.

« [Les grandes zones d'emploi sont les mieux armées face aux mutations économiques](#) », Insee Bourgogne Dimensions n°157, 2010.

« [Des territoires inégalement dotés pour faire face aux mutations économiques](#) », SUD INSEE L'essentiel n°131, 2009.

« [Économie de la Moselle-Est : un appareil productif en mutation accélérée](#) », Économie lorraine n°131, 2008.

« [Concentration et spécialisation de l'économie bas-normande, des activités exposées aux risques économiques](#) », Cent pour cent Basse-Normandie n°144, 2005.