

ANNEXE L'ANALYSE EN COMPOSANTES PRINCIPALES DES DIFFERENCIATIONS LOCALES

On considère un ensemble de six régions caractérisées par la distribution de trois secteurs d'activités et on se propose de dégager successivement des **facteurs de différenciation globaux** (concernant l'ensemble des paires de régions) et des **facteurs de différenciation locaux** (caractérisant les paires de régions contiguës).

I. Passage des lieux aux couples de lieux

La première étape de l'analyse consiste à passer d'un *Tableau A.1* décrivant les attributs des n lieux à un *Tableau A.2* décrivant les différences d'attributs entre les lieux. Chaque variable V est transformée en une variable DV définie par $DV_{ij} = (V_i - V_j)$. On adjoint à ce *Tableau A.2* une variable C_{ij} définissant la contiguïté territoriale des régions : $C_{ij}=1$ si les régions i et j sont contiguës et $C_{ij}=0$ dans le cas contraire. Enfin, on élimine les paires reliant une région à elle-même ($i=j$) ce qui conduit à un tableau contenant $n(n-1)$ lignes.

On extrait enfin du *Tableau A.2* les lignes correspondant aux paires de régions contiguës ($C_{ij}=1$), ce qui conduit à un *Tableau A.3* contenant $2.k$ lignes, k étant le nombre de limites entre deux régions.

II. Calcul des variations, des covariations et des corrélations locales

Soit E un sous-ensemble quelconque de l'ensemble des couples de lieux et contenant $2.l$ paires symétriques. Soit V une variable numérique définie sur l'ensemble des n lieux ayant servi à constituer les paires. On appelle *variation locale de V dans E* la quantité $\sigma^2(V,E)$ définie par :

$$\sigma^2(V,E) = \frac{1}{2.l} \sum_{(i,j) \in E} (V_i - V_j)^2 \quad (1)$$

Si on note T l'ensemble des paires de lieux vérifiant ($i \neq j$) on retrouve la relation habituelle avec la variance de la variable V :

$$\sigma^2(V,T) = 2.\sigma^2(V) \quad (2)$$

Et si on note C l'ensemble des paires contiguës, l'indice d'autocorrélation spatiale de Geary peut s'écrire à l'aide de la formule :

$$I_{Geary} = \frac{\sigma^2(V,C)}{\sigma^2(V,T)} \quad (3)$$

De la même manière on peut définir la covariation locale de deux variables V et W dans un sous-ensemble de couples de lieux E par la relation :

$$Cov(V, W, E) = \frac{1}{2.l} \sum_{(i,j) \in E} (V_i - V_j) \cdot (W_i - W_j) \quad (4)$$

Appliqué à l'ensemble de couples lieux T , la formule (4) montre l'existence d'une relation simple avec la formule habituelle de la covariance.

$$Cov(V, W) = 2 \cdot Cov(V, W, T) \quad (5)$$

Enfin, on peut définir le coefficient de corrélation des variations locales de deux indicateurs V et W dans un sous-ensemble de couples de lieux E par la relation :

$$r(V, W, E) = \frac{Cov(V, W, E)}{\sigma(V, E) \cdot \sigma(W, E)} \quad (6)$$

On déduit des formules (3), (5) et (6) que le coefficient de corrélation des variations locales appliqué à l'ensemble de couples de lieux T est égal au coefficient de corrélation linéaire entre les variables V et W .

$$r(V, W, T) = \frac{2 \cdot Cov(V, W)}{\sqrt{2} \cdot \sigma(V) \sqrt{2} \cdot \sigma(W)} = r(V, W) \quad (7)$$

III. Signification de l'analyse en composantes principales des différenciations locales

On déduit des équations précédentes qu'une analyse en composante principale effectuée sur les variables (DX, DY, DZ) du *Tableau A.2* (ensemble des couples de lieux) doit conduire à l'apparition des mêmes facteurs qu'une analyse en composante principale classique effectuée sur les variables (X, Y, Z) du *Tableau A.1*. Etudier les facteurs de différenciation des attributs des lieux revient en effet au même qu'étudier les différences entre les attributs de l'ensemble des couples de lieux. L'application d'un logiciel classique d'analyse des données aux *Tableaux A.1 et A.2* a permis de confirmer le résultat théorique : que l'on utilise une ACP normée (matrice des corrélations) ou une ACP non normée (matrice de covariance), les résultats de l'analyse sur les lieux et de l'analyse sur l'échantillon exhaustif des couples de lieux sont effectivement les mêmes.

Bien qu'évident sur le plan statistique, ce résultat est intéressant sur le plan théorique car il permet de préciser les différences entre une analyse globale et une analyse locale des facteurs de différenciations territoriales. En effet, si l'on effectue une ACP sur le *Tableau A.3*, les facteurs qui vont apparaître n'exprimeront plus les combinaisons de différences entre deux lieux quelconques mais entre deux lieux proches (en l'occurrence, deux lieux contigus). Or, il n'y a aucune raison que les facteurs de différenciation intervenant au niveau local soient les mêmes que ceux observés au niveau global et ce, pour deux raisons au moins :

- Si l'on effectue une analyse en composante principale non normée, le poids des variables dépend de leur variance. Transposé à l'étude des couples de lieux, ce poids est proportionnel à la variation locale de chaque indicateur. Du coup, les indicateurs qui présentent une distribution organisée (autocorrélation spatiale positive) vont générer des différences beaucoup plus faible au niveau local qu'au niveau global. Ils vont donc beaucoup moins contribuer à la constitution des facteurs si l'ACP est réduite à

l'échantillon des couples de lieux contigus. Inversement, les variables présentant une autocorrélation spatiale négative vont créer des différences locales proportionnellement plus fortes que celles qu'elles génèrent au niveau global et leur rôle dans la constitution des facteurs va être majoré.

- Si l'on effectue une analyse en composante principale normée sur les couples de lieux, l'effet précédent disparaît (chaque variable se voit attribuer le même poids), mais les facteurs obtenus ne seront pas pour autant les mêmes au niveau local et au niveau global car les corrélations entre les différences peuvent se modifier voire s'inverser quant on ne considère qu'un échantillon particulier de couples de lieux. Dans le cas d'un processus de ségrégation multiscalaire, on peut imaginer une distribution de deux populations A et B qui se localisent préférentiellement dans les mêmes quartiers mais tendent à se dissocier lorsqu'ils sont majoritaires sur un ensemble de quartiers. Une corrélation des différences établie sur l'ensemble des paires de quartiers montrera alors une corrélation positive (si un quartier i a plus de membres du groupe A qu'un quartier j , alors il a également en moyenne plus de membres du groupe B), tandis qu'une corrélation réduite aux paires de quartiers contigus pourra afficher une corrélation négative (si un quartier i a plus de membres du groupe A qu'un quartier j voisin, alors il a probablement moins de membres du groupe B que celui-ci).

IV. Exemple d'application

Dans l'exemple proposé, les variables qui contribuent en moyenne le plus aux *variations globales* de niveaux entre les régions sont les variables X et Y , la variable Z générant des variations environ trois fois plus faibles (*Tableau A.4*). Mais si l'on examine les *variations locales*, c'est-à-dire les différences entre les régions contiguës, la hiérarchie est différente car les variables X et Y présentent une autocorrélation spatiale positive (en particulier Y) tandis que la variable Z présente une autocorrélation spatiale négative. Les variables X et Z génèrent des variations de même intensité entre les régions contiguës tandis que les différences liées à la variable Y apparaissent négligeables.

Tableau A.4 : Variations globales et variations locales

	X	Y	Z
Variation globale	0.070	0.070	0.024
Variation locale	0.042	0.010	0.040
$1 - I_{\text{Geary}}$	0.400	0.857	-0.667

En ce qui concerne les corrélations de différences entre les régions, des différences tout aussi spectaculaires apparaissent quant on passe du niveau global au niveau local (*Tableau A.5*). Bien que le signe des corrélations soit conservé, leur intensité est fortement bouleversée. Au niveau global, les plus fortes associations des différences s'observent entre les variables X et Y (-0.83) tandis qu'au niveau local c'est entre les variables X et Z (-0.88).

Tableau A.5 : Corrélations globales et corrélations locales

	corrélation globale ($E=T$)	corrélation locale ($E=C$)
$r(X, Y, E)$	-0.83	-0.29
$r(X, Z, E)$	-0.29	-0.88
$r(Y, X, E)$	-0.29	-0.20

Ces différentes se répercutent logiquement sur la constitution des axes factoriels. Dans le cas d'une ACP normée (*Tableau A.6*) comme dans le cas d'une ACP non normée (*Tableau A.7*), des différences profondes apparaissent dans les contributions des variables à la constitution des axes factoriels et dans l'inertie de ceux-ci. Si on se limite à la première composante des différenciations, on voit qu'au niveau global elle oppose les régions agricoles (*X*) aux régions industrielles (*Y*) tandis qu'au niveau local, les plus fortes discontinuités entre régions contiguës sont liées à l'opposition entre régions rurales (*X*) et régions tertiaires (*Z*).

Tableau A.6 : Résultats des ACP locales et globale (normées)

	<i>ACP globale</i>		<i>ACP locales</i>	
	F1 (61%)	F2 (39%)	F1 (63%)	F2 (37%)
X	+ 0.96	+ 0.29	- 0.98	+ 0.19
Y	- 0.96	+ 0.29	+ 0.11	- 0.99
Z	+ 0.00	- 1.00	+ 0.95	+ 0.31

Tableau A.7 : Résultats des ACP locales et globale (non normées)

	<i>ACP globale</i>		<i>ACP locales</i>	
	F1 (78%)	F2 (22%)	F1 (84%)	F2 (16%)
X	+ 0.179	+ 0.055	- 0.210	+ 0.050
Y	- 0.179	+ 0.055	+ 0.007	- 0.105
Z	0.000	- 0.110	+ 0.203	+ 0.055

On pourrait proposer une interprétation imaginaire de ces résultats faisant intervenir deux processus historiques de transformation. L'économie agricole (*X*) domine initialement sur l'ensemble du territoire et joue le rôle de proie face à deux secteurs en expansion qui jouent le rôle de prédateur, mais selon des logiques spatiales différentes. L'industrialisation (*Y*) est apparue dans la partie occidentale du pays et s'est propagée selon un processus de diffusion continu vers la partie orientale du pays sans rencontrer d'obstacles. On peut imaginer par exemple qu'une grande manufacture en expansion, localisée dans la région 1, sous-traite sa production à des travailleurs à domicile qu'elle recrute dans des campagnes de plus en plus éloignées au fur et à mesure que les réservoirs de main-d'œuvre s'épuisent. Le résultat est un gradient régulier qui engendre une forte opposition globale entre l'ensemble des régions mais ne se traduit pas par l'apparition de discontinuités entre celles-ci au niveau local. L'urbanisation et le développement du secteur tertiaire (*Z*) ont obéi quant à eux à une logique concurrentielle, l'apparition d'une grande ville dans une région limitant le développement de villes dans les régions contiguës. Le processus est donc fondamentalement discontinu et il engendre au niveau local de fortes discontinuités entre les régions contiguës.

Les deux processus de prédation génèrent donc des formes d'opposition territoriale à des niveaux d'organisation différents : les effets de l'industrialisation sont perceptibles au niveau global (opposition entre *X* et *Y* sur le premier facteur de l'ACP utilisant tous les couples de régions) tandis que ceux de l'urbanisation sont plus localisés et ne génèrent des différenciations importantes qu'au niveau local (opposition entre *X* et *Z* sur le premier facteur de l'ACP n'utilisant que les couples de lieux contigus).

Figure A.1 : Information disponible

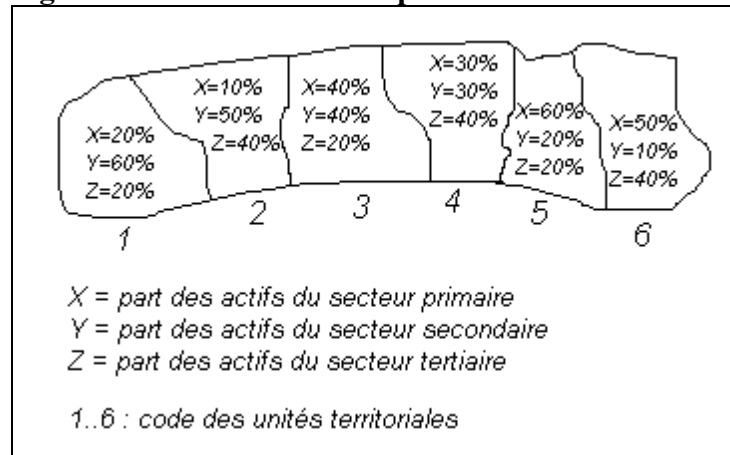


Tableau A.1

i	X_i	Y_i	Z_i
1	0.2	0.6	0.2
2	0.1	0.5	0.4
3	0.4	0.4	0.2
4	0.3	0.3	0.4
5	0.6	0.2	0.2
6	0.5	0.1	0.4

Tableau A.2

i	j	DX_{ij}	DY_{ij}	DZ_{ij}	C_{ij}
1	2	0.1	0.1	-0.2	1
1	3	-0.1	0.2	0	0
1	4	-0.1	0.3	-0.2	0
1	5	-0.4	0.4	0	0
1	6	-0.3	0.5	-0.2	0
2	1	-0.1	-0.1	0.2	1
2	3	-0.2	0.1	0.2	1
2	4	-0.2	0.2	0	0
2	5	-0.5	0.3	0.2	0
2	6	-0.4	0.4	0	0
3	1	0.1	-0.2	0	0
3	2	0.2	-0.1	-0.2	1
3	4	0	0.1	-0.2	1
3	5	-0.3	0.2	0	0
3	6	-0.2	0.3	-0.2	0
4	1	0.1	-0.3	0.2	0
4	2	0.2	-0.2	0	0
4	3	0	-0.1	0.2	1
4	5	-0.3	0.1	0.2	1
4	6	-0.2	0.2	0	0
5	1	0.4	-0.4	0	0
5	2	0.5	-0.3	-0.2	0
5	3	0.3	-0.2	0	0
5	4	0.3	-0.1	-0.2	1
5	6	0.1	0.1	-0.2	1
6	1	0.3	-0.5	0.2	0
6	2	0.4	-0.4	0	0
6	3	0.2	-0.3	0.2	0
6	4	0.2	-0.2	0	0
6	5	-0.1	-0.1	0.2	1

Tableau A.3

i	j	DX_{ij}	DY_{ij}	DZ_{ij}	C_{ij}
1	2	0.1	0.1	-0.2	1
2	1	-0.1	-0.1	0.2	1
2	3	-0.3	0.1	0.2	1
3	2	0.3	-0.1	-0.2	1
3	4	0.1	0.1	-0.2	1
4	3	-0.1	-0.1	0.2	1
4	5	-0.3	0.1	0.2	1
5	4	0.3	-0.1	-0.2	1
5	6	0.1	0.1	-0.2	1
6	5	-0.1	-0.1	0.2	1