

**ANNEXE**  
**L'outil de construction des zonages : ANABEL<sup>1</sup>**

<sup>1</sup> Extraits des manuels d'utilisation des logiciels Mirabel et Anabel.



La construction des principaux zonages d'études, notamment le zonage en aires urbaines, les zones d'emploi et les bassins de vie, consiste à rassembler des territoires sur la base des flux échangés. Des méthodes de zonage basées sur l'analyse des flux ont été développées à l'Insee. Il n'y a pas qu'une seule méthode pour définir des « bassins », des « aires d'influence » ou des « espaces de mobilité ». Le point de départ commun à la plupart d'entre elles est la matrice donnant, pour chaque paire de communes, le flux correspondant. A partir de cette matrice, plusieurs algorithmes permettent de passer aux « bassins ». De même qu'avec une même méthode, le choix du lien peut créer des zonages différents.

Historiquement, la première méthode conçue a été le logiciel **ZONAGE**<sup>2</sup> (Loeiz Laurent), avec le principe d'obtenir un zonage optimisé par application d'un certain nombre de règles. Puis les logiciels **MIRABEL/ANABEL** ont été développés. Plus récemment, d'autres méthodes ont été explorées, notamment des méthodes issues de la **théorie des graphes et de l'analyse de réseaux**.

L'objectif ici est de présenter la méthode utilisée dans nos zonages actuels ZAU2010, ZE2010 et BV2012, à savoir les outils **MIRABEL/ANABEL**.

Le logiciel **MIRABEL**<sup>3</sup> a été élaboré au début des années 80 (Christophe Terrier, Bernard Sinou). Cette méthode a été beaucoup utilisée dans la construction des zonages d'étude de l'Insee. A l'origine, cet outil a été conçu pour étudier les déplacements entre le lieu de domicile et le lieu de travail afin d'effectuer un zonage du territoire en bassins d'emploi. En principe, deux points du territoire sont d'autant plus liés que les échanges entre eux sont nombreux. Telle est l'idée de départ de Mirabel. Pour mesurer cette liaison, plusieurs types de liens étaient disponibles. Mirabel permettait d'obtenir, par agrégation, un zonage. Par extension, Mirabel a pu traiter tous types de données bilocalisées.

Cet outil a été rénové en 2010 sous le nom d'**ANABEL**<sup>4</sup>. Les zonages d'études les plus récents ont été construits avec cette méthode. Il reprend l'algorithme mis au point pour **MIRABEL**, dont l'applicatif n'était plus opérationnel. **ANABEL** est développé sous R (logiciel libre d'analyse statistique). Le principe est identique à celui de Mirabel, les différences étant principalement au niveau des choix pour les liens, des possibilités et des sorties proposées.

### **Principe de la méthode**

La méthode utilisée est un procédé d'agrégation qui fonctionne par itération.

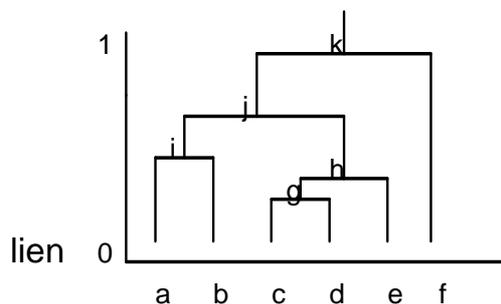
On dispose d'un ensemble d'observations. On sait calculer le "lien" caractérisant deux observations quelconques. En règle générale, les observations correspondent aux communes à regrouper, et les "liens" sont calculés à partir des flux observés entre les observations. Dès que le nombre d'observations dépasse la vingtaine, il devient très vite difficile d'analyser cette matrice. L'idée consiste donc à résumer l'information contenue dans cette matrice sous la forme d'un arbre. Les racines de cet arbre, qui représentent chacune des observations à classer, s'assemblent progressivement, par ordre de "liens" décroissants, regroupant un nombre de plus en plus grand d'observations pour aboutir finalement à un sommet unique (figure 1).

<sup>2</sup> Pour en savoir plus sur ce logiciel, cf. « *Les zonages : enjeux et méthodes* » Insee-Méthodes n°83, décembre 1998.

<sup>3</sup> Méthode Informatisée de Recherche et d'Analyse de Bassins par l'Etude de Liaisons

<sup>4</sup> ANALyse Bilocalisée pour les Etudes Locales

**Figure 1 - L'arbre de classification**



Plus le lien entre deux communes est élevé, plus ces communes seront 'proches'. Dans la figure 1, **c** et **d** sont les observations les plus proches : elles s'agrègent pour former le "noeud" **g**. Puis **e** vient s'agréger pour former le noeud **h**, etc.

### Mise en œuvre de la méthode

Pour faciliter la compréhension du processus d'agrégation, l'exemple de la figure 2 illustre son déroulement.

Chaque commune est identifiée par une lettre. On trouve :

- en ligne la commune de résidence X ;
- en colonne la commune de travail Y ;
- à droite du tableau la population active de la commune de résidence.

Chaque case (X,Y) du tableau contient deux valeurs :

- en haut le nombre de personnes habitant en X et travaillant en Y ;

- en bas le lien défini comme précédemment  $\left[ \text{lien}(A, B) = \frac{\text{Flux}(A, B)}{\text{PopAct}(A)} \right]$

Flux (A, B) = flux de A vers B (sorties de A vers B)

PopAct(A) = population active de A

Flux (A, A) = stables de A (personnes qui résident et travaillent dans A)

La somme des effectifs en ligne est égale à la population totale, sauf si la commune envoie des actifs en dehors du champ de l'étude (ce qui est le cas ici).

**Figure 2 - Exemple de flux entre communes**

Lieu de travail	A	B	C	F	G	I	J	K	M	Autres Destinations	Population active
Lieu de résidence											
A ARRENTES	20 57.1					10 28.6				5 14.3	35
B AUMONTZEY		130 53.1			10 4.1	35 14.3	5 2.0			65 26.5	245
C BARBEY-SEROUX			15 20.0			50 66.6				10 13.3	75
F LA FORGE				55 29.7	25 13.5				50 27.0	55 29.7	185
G GERARDMER				5 0.1	3585 92.9	15 0.4			15 0.4	235 6.1	3855
I GRANGES/VOLOGNE		20 1.8			65 5.8	970 85.8			5 0.4	70 6.2	1130
J JUSSARUPT		5 4.0			5 4.0	15 12.0	85 68.0			15 12	125
K LIEZEY					50 50.0			40 40.0		10 10	100
M LE THOLY					110 18.8				415 70.9	60 10.3	585

Processus :

- On cherche **les deux communes les plus liées** et on **les agrège**<sup>5</sup>.

Dans l'exemple, il s'agit de C et de I, C envoyant 50 personnes sur 75 soit 66.6% de ses actifs vers I. Il faut donc agréger C à I pour **former une nouvelle unité géographique** ou **classe**. En pratique, on appellera I cette nouvelle classe du nom de la commune attirante. Dorénavant les communes C et I n'existeront plus mais seront représentées toutes les deux par la classe I. On dira aussi que I est le pôle<sup>6</sup> de la classe et C le premier élément de cette classe.

- Il convient alors d'additionner les effectifs de la ligne Y à ceux de la ligne X, y compris les populations totales (suppression de la ligne Y), et d'additionner les effectifs de la colonne Y à ceux de la colonne X (suppression la colonne Y).

- On recalcule le lien entre la classe ainsi créée et chacune des autres communes ou classes déjà constituées.

- On recherche les deux communes ou classes les plus liées, compte tenu des liens nouvellement calculés.

- On recommence le processus : les classes prennent de l'importance puis s'agrègent entre elles jusqu'à former un seul ensemble représentant la région.

Une nouvelle lecture montre que K et G sont maintenant les plus liées, puis A et I. Après l'agrégation de K à G puis de A à I, on obtient le tableau suivant, les pôles étant en majuscules et les satellites en minuscules :

<sup>5</sup> On exclut de cette recherche la diagonale de la matrice qui donne la population stable de la commune ainsi que le taux de stabilité correspondant. On rappelle que le taux de stabilité pour une zone donnée est le rapport du nombre d'actifs résidant et travaillant dans cette zone à sa population active totale.

<sup>6</sup> Lorsque le lien est symétrique, le pôle sera par convention la commune la plus peuplée.

**Figure 3 - La figure 2 après trois agrégations**

Lieu de travail	B	F	G, k	I, c, a	J	M	Autres destinations	POP
Lieu de résidence								
B	130		10	35	5		65	245
AUMONTZEY	53.1		4.1	14.3	2.0		26.5	
F		55	25			50	55	185
LA FORGE		29.7	13.5			27.0	29.7	
G, k		5	3675	15		15	245	3955
GERARDMER		0.1	92.9	0.4		0.4	6.2	
I, c, a	20		65	1065		5	85	1240
GRANGES/VOLOGNE	1.6		5.2	85.9		0.4	6.9	
J	5		5	15	85		15	125
JUSSARUPT	4.0		4.0	12.0	68.0		12	
M			110			415	60	585
LE THOLY			18.8			70.9	10.3	

Le lecteur est invité à continuer jusqu'à la fin et il découvrira que G demeure pôle de la classe qui regroupe les 8 autres communes. Ce travail, possible manuellement pour 9 communes, nécessite obligatoirement l'emploi de l'ordinateur lorsqu'il s'agit d'en étudier plusieurs milliers.

## Résultats

Les résultats peuvent se résumer sous la forme d'une liste, les communes sont en minuscules et les classes désignées par le nom de leur pôle en majuscules :

### **Figure 4 - Le détail des agrégations successives**

Etape 1 : c s'agrège à i au taux de 66.6 %.  
i devient pôle de la zone I tandis que c est appelé satellite.

Etape 2 : k s'agrège à g au taux de 50.0 %.  
g devient pôle de la zone G et k son satellite.

Etape 3 : a s'agrège à I au taux de 28.6 %.  
a s'agrège à la zone I et en devient satellite.

Etape 4 : f s'agrège à m au taux de 27.0 %.  
m devient pôle de la zone M et f son satellite.

Etape 5 : M s'agrège à G au taux de 17.5 %.  
M devient sous-pôle<sup>7</sup> de la zone G. La classe (M, f) devient une sous-classe de G et ces 2 communes deviennent deux satellites de G.

Etape 6 : b s'agrège à I au taux de 14.3 %.  
b devient satellite de I.

Etape 7 : j s'agrège à I au taux de 16.0 %.  
j devient satellite de I : *Phénomène de rebond (voir ci-dessous).*

Etape 8 : I s'agrège à G au taux de 5.2 %.

Les 9 communes sont finalement regroupées au sein d'une seule classe dont le pôle est G. Cette classe contient deux sous classes (I, c, a, b, j) et (M, f). La hiérarchie de la classe finale est donc (G, k, (M, f), (I, c, a, b, j)).

### **Remarque :**

Si une classe Y s'agrège à une classe X pour former la classe résultante  $Z = X+Y$  alors :

- les liens des autres communes vers la classe Z sont plus élevés que pour X ou Y. C'est cette propriété qui explique le "rebond". En effet, si T est une commune quelconque différente de X et Y

$$\text{lien}(T \rightarrow Z) = \text{lien}(T \rightarrow X) + \text{lien}(T \rightarrow Y)$$

- de plus si  $\text{lien}(X \rightarrow T) \geq \text{lien}(Y \rightarrow T)$ ,  $\text{lien}(X \rightarrow T) \geq \text{lien}(Z \rightarrow T) \geq \text{lien}(Y \rightarrow T)$

<sup>7</sup> Sous-pôle : on appellera sous-pôle ou pôle rattaché, une commune qui, s'étant en premier lieu comportée comme un pôle, s'est rattachée avec l'ensemble de sa zone à une autre classe.

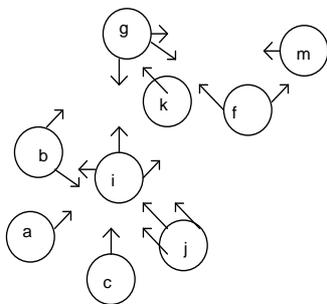
Puisque le processus consiste à toujours traiter le lien le plus fort, on pourrait s'attendre à ce que les liens soient toujours décroissants dans la liste des résultats. Ce n'est pas toujours le cas. Ainsi à l'étape 7, le lien de j à I (16%) est plus fort que le lien de b à I (14.3%) qui a pourtant été traité le premier. C'est ce que l'on appelle le phénomène de "rebond".

L'explication en est simple. Après l'étape 6, la classe I composée des communes I, c et a attire la commune b à 14% et la commune j à 12%. C'est donc b qui s'agrège à I à l'étape 10. La commune j était en outre attirée par b à 4%. Après l'étape 6, j est donc liée à 12+4 soit 16% à la classe I, formée maintenant des communes I, c, a et b. Ce lien, établi avec une classe plus grande, et par l'intermédiaire de B, relate une "polarisation" moins forte.

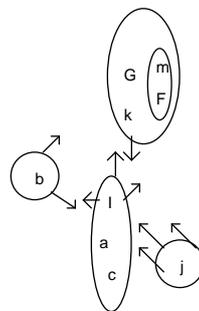
Les résultats peuvent être représentés de différentes façons. Sous la forme d'une liste relatant le détail des agrégations comme ci-dessus. Cette retranscription est insuffisante pour rendre compte de toute la richesse de l'information. La représentation graphique sous forme d'arbre est au contraire bien adaptée pour saisir la hiérarchie des rattachements. Enfin la représentation cartographique est également utile pour finaliser le zonage.

**Figure 5 - Représentation schématique des agrégations<sup>8</sup>**

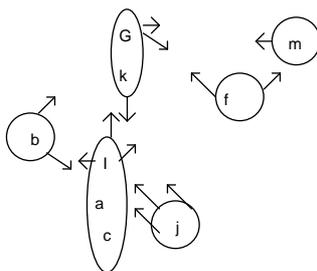
1- Au départ, chaque commune constitue sa propre zone.



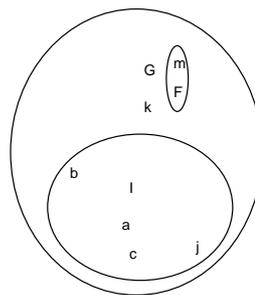
3- Etape 5 : F est devenu sous-pôle (ou satellite) de la zone G.



2- Trois étapes plus loin : c s'est agrégé à I, k à G et a à I.



4- Fin du processus : l'ensemble des communes ne forme plus qu'une seule zone dont le pôle est G.



<sup>8</sup> Les pôles des zones sont en majuscules (sauf les communes isolées qui sont en minuscule).

## Les critères d'arrêt

Comme le démontre le principe explicité précédemment, sans critère d'arrêt le résultat final de l'agrégation correspondra à notre région étudiée dans son ensemble. Il faut donc déterminer quand stopper le processus.

Pour cela 3 paramètres sont possibles, et peuvent être combinés :

### a) Lien minimal

Il est possible de fixer une valeur du lien en-deçà de laquelle aucune agrégation n'est possible. C'est-à-dire que si  $L$  est le lien maximal entre entités trouvé, et que  $L < L_{\min}$ , alors on stoppe le processus d'agrégation.

### b) Isolation partielle

Il est possible de déterminer une valeur de  $PopAct(A)$  au-delà de laquelle une entité ne peut plus être déclarée sous-pôle (ou satellite) d'une autre entité.

Ainsi, si on reprend la Figure 5, lors de l'étape 5 (dessin 3) nous avons l'entité F (formée de f & m) qui devenait satellite de G. Or si  $PopAct(F) > Seuil\_Isolation\_Partielle$  alors F ne pourra être déclaré satellite et donc l'agrégation ne pourra se faire. En revanche rien n'empêche une autre entité de venir s'agréger avec F, devenant ainsi son satellite et F le pôle.

Classe partiellement isolée : si elle se comporte comme un pôle, ne peut plus s'agréger à une autre classe mais d'autres classes pourront s'y agréger.

### c) Isolation totale

Ce paramètre correspond à un seuil de  $PopAct(A)$  au-delà duquel la zone constituée ne peut plus évoluer. Elle ne pourra donc plus être ni satellite d'une autre zone, ni accueillir par agrégation une quelconque autre entité.

Ainsi, si on reprend la Figure 5, lors de l'étape 5 (dessin 3) nous avons l'entité F (formée de f & m) qui devenait satellite de G. Or si  $PopAct(G) > Seuil\_Isolation\_Totale$  alors la zone G ne pourra plus évoluer et donc l'agrégation ne pourra se faire.

Classe totalement isolée : lorsqu'elle ne pourra plus s'agréger à une autre et qu'aucune classe ne pourra s'y agréger.

**L'algorithme prend fin lorsqu'aucune agrégation n'est possible en fonction des critères définis.**

Quelques définitions :

- Classe : ensemble d'unités géographiques agrégées lors du processus d'agrégation.
- Pôle : lors d'une agrégation, on parle de pôle pour désigner l'unité géographique attirante.
- Pôle rattaché : pôle dont la zone a été agrégée dans la suite du processus à un autre pôle. On parle aussi de sous-pôle.
- Rebond : une unité géographique peut se trouver attirée à un pôle avec un taux plus fort que l'unité géographique précédemment agrégée à ce pôle, tout simplement parce qu'elle se trouve attirée à la fois par le pôle et à la fois par cette unité déjà agrégée. Le rebond n'intervient que lors du recalcul de la table de données.
- Satellite : lorsque le lien ne fait intervenir qu'un flux orienté, on dira qu'une unité géographique devient satellite si c'est elle qui est attirée. La commune attirante est la commune pôle.

## Les paramètres

L'application permet de fixer un ensemble de paramètres, comme les critères d'arrêt (cf. supra), mais aussi d'autres critères permettant d'affiner l'étude par rapport au zonage attendu :

### Type de liaison

ANABEL propose 2 types de fonctions de liaison :

$$\begin{aligned} \text{Attractivité (AA)} : & \left[ \text{lien}(A, B) = \frac{\text{Flux}(A, B)}{\text{PopAct}(A)} \right] \\ \text{Echange (EN)} : & \left[ \text{lien}(A, B) = \frac{\text{Flux}(A, B) + \text{Flux}(B, A)}{\text{PopAct}(A)} \right] \end{aligned}$$

Le choix entre ces 2 fonctions de lien se fera selon les flux étudiés :

- Ainsi pour des flux du type polarisés (par ex. navettes domicile/travail), la fonction à utiliser sera plutôt la AA.
- Pour les flux sans polarisation apparente (par ex. migration résidentielles), la fonction à utiliser sera plutôt la EN.

Le lien AA prend une valeur comprise entre 0 et 1 et correspond à la part des individus de A se rendant dans B.

Le lien EE est quasi-symétrique. Il prend en compte les flux dans les 2 sens ( $A \rightarrow B$  et  $B \rightarrow A$ ). Il prend des valeurs minorées par 0 mais pas majorées.

ANABEL ne reprend pas l'ensemble des fonctions de lien proposées par Mirabel, mais uniquement celle utilisées régulièrement.

### Remarques :

- En attraction, le calcul ne fait intervenir qu'un flux orienté de A vers B. En échange, le calcul fait intervenir les flux dans les deux sens, A vers B et B vers A. En échange, les liens sont dits symétriques lorsque les deux unités géographiques élémentaires sont prises en compte de manière égalitaire.
- Les liens en attraction sont conseillés lorsqu'on cherche à délimiter l'aire d'attraction d'une commune ou d'une zone. Les liens en échange privilégient la réciprocité des flux et donnent souvent plus de poids à la commune la plus peuplée. Par exemple, une zone industrielle implantée en périphérie d'une ville pourra être considérée comme pôle en attraction dans la mesure où elle attire les habitants des communes voisines qui viennent y travailler. Mais en échange, cette zone ne s'individualisera probablement pas et sera comprise dans la zone d'échange de la ville avec laquelle elle fonctionne.

### Contiguïté

Il est possible d'ajouter une contrainte de contiguïté, afin de ne rendre possible l'agrégation entre entités que si elles sont contigües l'une à l'autre.

### Prise en compte des stables

Ce paramètre permet de ne pas prendre en compte les individus ayant des flux stables (i.e.  $A \rightarrow A$ ).

En effet, si ces flux ne sont pas pris en compte dans le calcul du lien maximal (une entité ne sera jamais agrégée avec elle-même), ils interfèrent dans le calcul de  $\text{PopAct}(A)$  dans la fonction de lien. Ne pas prendre en compte les stables focalise donc les calculs uniquement sur les individus mobiles.

### Pôles a priori

Il est possible de fixer au départ un ensemble de communes qui joueront le rôle de pôles a priori. C'est-à-dire que les agrégations ne se feront qu'autour de ces pôles, les pôles ne pouvant être satellites d'autres pôles.

Le résultat final sera donc une partition contenant autant de zones que de pôles a priori, avec un pôle par zone.

### L'effet « boule de neige »

Dans le cas d'une ville par exemple, les communes de la périphérie vont être fortement attirées. De fil en aiguille, des communes de plus en plus éloignées vont s'agréger à cette zone car elles sont elles-mêmes attirées par cette zone en constitution, plus faiblement certes, mais l'agrégation sera tout de même effective. C'est ce que l'on appelle l'effet « boule de neige ». Cet effet rend plus difficile l'émergence de petites zones à la périphérie de gros pôles d'emploi. C'est également ce processus qui peut expliquer qu'au final, une commune peut se trouver plus fortement attirée par une autre zone que la sienne (à la limite de deux grosses zones par exemple). En effet, après son agrégation, Anabel continue à faire grossir toutes les zones, aussi bien la sienne que ses voisines. L'une de celle-ci peut très bien, en atteignant une configuration plus avancée, devenir la plus attirante.

Comme on l'a vu, il peut arriver qu'une commune affectée à une zone se trouve plus attirée au final par une autre zone. Ce cas est a priori rare mais il résulte du fait que l'outil affecte une fois pour toute une commune à une zone, et que l'ordre d'agrégation n'est pas anodin.