

REVUE DE STATISTIQUE APPLIQUÉE

J.-J. DENIMAL

M. ALPHA DIALLO

**Réalisation d'un zonage du département du
Nord à partir de l'inventaire communal (INSEE
Nord-Pas-de-Calais, 1988)**

Revue de statistique appliquée, tome 48, n° 1 (2000), p. 85-102

http://www.numdam.org/item?id=RSA_2000__48_1_85_0

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

RÉALISATION D'UN ZONAGE DU DÉPARTEMENT DU NORD À PARTIR DE L'INVENTAIRE COMMUNAL (INSEE-Nord Pas-de-Calais, 1988)

J.-J. Denimal⁽¹⁾ et M. Alpha Diallo⁽²⁾

(1) *Université des Sciences et Technologies de Lille (USTL).*

(2) *Université des Sciences de Conakry (République de Guinée).*

RÉSUMÉ

Dans le cadre de l'aménagement du territoire, on se propose, à partir de données provenant de l'inventaire communal réalisé par l'INSEE Nord Pas de Calais en 1988, de regrouper les communes du département du Nord en classes, chacune d'elles étant autonome du point de vue des équipements qu'elles possèdent. Après comparaison avec d'autres techniques de classification, une technique de classification hiérarchique basée sur des tableaux de flux (Domengès 1982) sera utilisée pour la construction de la hiérarchie, puis complétée par une étude plus détaillée des classes obtenues. Dans chacune d'elles, la détermination des villes-pôles, attirant par ses équipements les villes satellites voisines, sera ensuite réalisée. Le graphe représentant ces flux de populations sera décrit et visualisé sous forme de graphiques et de cartes. Enfin, des propositions de traitements statistiques pour étendre cette étude à l'ensemble des villes de France seront faites en conclusion de cet article.

Mots-clés : *Tableaux de Flux, Classification Ascendante Hiérarchique, Analyse des Graphes.*

ABSTRACT

Within the framework of the town and country planning, a classification of the towns of the North of France based on a survey realized by the french Institute INSEE Nord Pas de Calais in 1988, is proposed in order to create clusters of towns, defining independent zones from the viewpoint of their equipments. After comparisons with other methods of classification, a technique of hierarchical clustering based on input/output tables (Domengès 1982) will be applied for the construction of these clusters of towns. In each of them, the determination of main cities whose equipments attract satellite towns, will be realized. The graph representing these flows of populations will be then described and displayed into graphical representations or maps. Finally, a proposition of a statistical methodology is made, at the end of this paper, in order to extend the current work to the totality of the french towns.

Keywords : *Input/output Tables, Hierarchical Clustering, Graph Analysis.*

1. Introduction

La présente communication consiste, dans le cadre de l'aménagement du territoire, en le traitement de données provenant de l'inventaire communal réalisé par l'INSEE Nord Pas de Calais en 1988. Ce traitement mettra, en particulier, en œuvre une technique de classification ascendante hiérarchique basée sur des tableaux de flux (Domengès 1982). Cette technique sera, ensuite, complétée par une étude plus fine des classes de villes obtenues. De manière plus précise, l'objet de cette enquête est d'une part de mieux connaître le niveau d'équipement de chaque commune du département du Nord et d'autre part de quantifier et de caractériser les flux de populations provenant des communes voisines d'une commune donnée et engendrés par l'utilisation d'un équipement particulier de cette dernière. Le problème ainsi posé se traduit, d'un point de vue statistique, par la construction d'une classification de l'ensemble des communes du département du Nord en classes, chacune d'elles constituant une zone autonome du point de vue de ses équipements. Dans chacune de ces zones (encore appelées « bassins de vie »), les flux de populations seront caractérisés par un certain nombre de villes- pôles attirant par leurs équipements les communes environnantes. Le travail à réaliser consiste donc, en un premier temps en la construction de cette classification et en un second temps en la détermination et la caractérisation de ces villes-pôles.

Concernant l'enquête réalisée par l'INSEE pour élaborer cet inventaire communal, un questionnaire regroupant plus de 500 questions a été proposé à chaque mairie des 652 communes du Nord afin de déterminer et de mesurer la fréquentation de leurs équipements. Après dépouillement des premiers résultats, l'INSEE Nord Pas de Calais a sélectionné 39 équipements comme étant les plus caractéristiques. La présente étude a été conduite sur la base de ce choix. Le détail de ces 39 équipements retenus (hypermarché, garagiste, perception, collègue, médecin,) pourra être trouvé au paragraphe 4, lors de l'examen des graphiques 7 et suivants.

L'article est structuré en 5 chapitres : Après l'introduction (§ I), le paragraphe II présentera la technique de classification retenue (Domengès 1982) et sa comparaison avec d'autres méthodes, en particulier celle utilisée par l'INSEE (la méthode appelée Mirabel). La classification que nous avons appliquée étant hiérarchique, une coupure de l'arbre obtenu sera réalisée après examen de la représentation graphique de l'indice d'agrégation (graphique 1) et donnera naissance à une partition de l'ensemble des communes, composée des classes terminales dites associées à la coupure réalisée. Une cartographie de l'ensemble de ces classes de communes sera proposée et permettra de visualiser les résultats obtenus (graphiques 2 et 3). Le paragraphe III sera consacré à la détermination des villes-pôles dans chacune des classes de communes ainsi construites. Différents types de villes-pôles pourront alors être observés. Nous donnerons au paragraphe IV, à titre d'exemple, une étude plus fine de l'une des classes de communes obtenues. Les équipements utilisés par ces communes seront d'abord décrits de manière globale, puis en distinguant les flux de populations associés à chacun des pôles de cette classe. Cette étude permettra d'obtenir, d'une part une cartographie des flux de populations, et d'autre part des représentations graphiques caractérisant ces derniers en terme de types d'équipements utilisés et ceci pour chaque pôle de la classe.

2. Classification des 652 communes du Nord

2.1. Les données, que l'INSEE Nord Pas de Calais nous a communiquées, se présentent sous la forme d'un tableau croisant l'ensemble des 652 communes du Nord avec celui des 39 équipements retenus, chaque case (i, j) contenant le code postal de la commune possédant l'équipement j et attirant les habitants de la ville i . (Si plusieurs communes possèdent cet équipement j , seule est retenue la ville attirant le plus grand nombre d'habitants de i). Une colonne supplémentaire est adjointe à ce tableau contenant les nombres d'habitants des 652 communes.

Un tableau de flux k_{II} , croisant l'ensemble I des 652 communes avec lui-même, sera alors constitué. Chaque case (i, i') croisant les communes ligne et colonne i et i' , contient la quantification $k(i, i')$ du flux de population allant de la commune i vers celle notée i' : $k(i, i')$ est égale au nombre d'équipements de i' attirant les habitants de i multiplié par le nombre d'habitants de la ville i .

Dans l'ensemble de cette étude, seuls seront considérés les flux internes au département du Nord. Autrement dit, Les flux de populations générés par des équipements extérieurs au département du Nord ne seront pas pris en compte.

2.2. Il¹ est clair que le tableau de flux k_{II} comporte un très grand nombre de zéros. Aussi, son stockage sur disque a été facilité (gain de place mémoire) par l'implantation de trois vecteurs notés ci-dessous COL , VL , ADL construits grâce au passage « ligne-colonne ». Ainsi, par ce procédé, une matrice M possédant n lignes, p colonnes et k éléments non nuls peut se résumer en 3 vecteurs :

- Un vecteur COL de taille k où $COL(\ell)$ est l'indice de la colonne incidente pour la ℓ^e valeur non nulle, $1 \leq \ell \leq k$ (la lecture de la matrice se faisant de gauche à droite et de haut en bas).

- Un vecteur VL de taille k où $VL(\ell)$ est la valeur du ℓ^e élément non nul de la matrice.

- Un vecteur ADL de taille $n + 1$ où $ADL(i)$ vaut :

Pour $i = 1$, $ADL(1) = 0$

Pour $i > 1$, $ADL(i) =$ Nombre d'éléments non nuls dans la ligne $i - 1$ additionné des nombres d'éléments non nuls des lignes précédentes.

(Ainsi, $ADL(i + 1) - ADL(i) =$ Nombre d'éléments non nuls de la ligne i)

En conséquence, I désignant l'ensemble des n lignes de la matrice M , on peut alors écrire : $\forall i \in I, \forall \ell \in [ADL(i), ADL(i + 1)], M(i, COL(\ell)) = VL(\ell)$.

2.3. Diverses méthodes de construction peuvent être envisagées pour le classement des 652 communes en zones autonomes du point de vue des 39 équipements retenus.

¹ La rédaction d'un programme en langage IML de SAS permettant de codifier un tableau sous la forme de ces trois vecteurs a été réalisée par Mesdemoiselles Dauphinot et Julien étudiantes en maîtrise de Mathématiques Appliquées de Lille1 dans le cadre d'un stage effectué dans le service de Monsieur Houssin, Directeur Adjoint de l'INSEE NORD PAS DE CALAIS.

2.3.1. L'INSEE utilise une technique appelée « Mirabel » qui s'apparente à une classification hiérarchique ascendante particulière. De manière plus précise, I désignant l'ensemble des villes du département du Nord et k_{II} le tableau de flux défini précédemment, posons :

$$k(i, \cdot) = \sum_{i' \in I} k(i, i'), \forall i \in I$$

c et c' désignant deux sous-ensembles disjoints de I ,

$$k(c, \cdot) = \sum_{i \in c} \sum_{i' \in I} k(i, i') \quad k(c, c') = \sum_{i \in c} \sum_{i' \in c'} k(i, i')$$

L'indice d'association, noté $\text{lien}(c, c')$, utilisé par Mirabel vaut :

$$\text{lien}(c, c') = \frac{k(c, c')}{k(c, \cdot)}$$

La technique consiste à agréger à chaque niveau les deux classes admettant l'indice d'association maximum.

L'indice d'association $\text{lien}(c, c')$ n'est pas symétrique. Cependant, on peut remarquer que l'indice d'association $\text{Max}\{\text{lien}(c, c'); \text{lien}(c', c)\}$, symétrique par définition, construira le même arbre.

La principale difficulté liée à l'application de cette technique est qu'elle engendre des inversions. Autrement dit, deux classes c et c' peuvent s'agréger à un niveau d'association $\text{lien}(c, c')$ qui soit supérieur à celui $\text{lien}(d, d')$ deux classes d et d' agrégées antérieurement.

2.3.2. Une seconde technique, utilisée par l'INSEE, et appelée Zonage, se propose, par permutations des lignes et des colonnes du tableau de flux k_{II} , de former une suite de sous- matrices diagonales chargées des flux les plus importants. Malheureusement, la mise en œuvre précise de cette technique ne nous a pas été communiquée.

2.3.3. Une procédure de zonage du type de celle de l'INSEE serait possible après avoir transformé le tableau de flux k_{II} initial en un tableau binaire composé de 0 et de 1 en ne retenant que les flux les plus importants. En effet, à partir du choix d'un seuil α , un tableau binaire kb_{II} pourrait être ainsi construit : $\forall i \in I, \forall i' \in I$

$$kb(i, i') = \begin{cases} 1 & \text{si } \frac{k(i, i')}{k(i)} > \alpha \\ 0 & \text{sinon} \end{cases}$$

La technique de classification croisée de Govaert (1984) pourrait alors être appliquée et reviendrait à réordonner les lignes et les colonnes du tableau kb suivant deux partitions de façon à obtenir des blocs homogènes de 1 ou de 0. Cependant, Govaert utilise des techniques du type des nuées dynamiques de Diday (1971) pour

lesquelles il est nécessaire de connaître à priori les nombres de classes des partitions que l'on cherche.

Ainsi, le choix du seuil α et la connaissance à priori des nombres de classes des partitions ligne et colonne cherchées constituent les difficultés posées par cette approche.

2.3.4. La technique que nous avons retenue a été proposée par Domengès (1982) dans les Cahiers de l'Analyse des Données, sous le titre classification hiérarchique ascendante d'après un critère adapté aux tableaux de flux. Nous compléterons cette méthode aux paragraphes suivants (§III et IV) par une procédure de détermination des villes-pôles dans chacune des classes de communes obtenues.

I désignant l'ensemble des communes du département du Nord, le principe général de la méthode de Domengès consiste, tout d'abord à se ramener à un tableau symétrique ks_{II} en posant : $\forall i \in I, \forall i' \in I$

$$ks_{II}(i, i') = \begin{cases} k(i, i') + k(i', i) & \text{si } i \neq i' \\ k(i, i) & \text{sinon} \end{cases}$$

Dans le cadre de cette étude, il est à noter que les flux de populations s'effectuent généralement dans un seul sens (soit de i vers i' , soit de i' vers i).

On pose alors :

$$\begin{aligned} \forall i \in I, \quad ks(i) &= \sum_{i' \in I} ks(i, i') \\ ks &= \sum_{i \in I} \sum_{i' \in I} ks(i, i') \end{aligned}$$

c et c' étant deux classes disjointes de communes, on pose :

$$ks(c, c') = \sum_{i \in c} \sum_{i' \in c'} ks(i, i') \quad ks(c) = \sum_{i \in c} \sum_{i' \in I} ks(i, i')$$

Le taux d'association entre deux classes c et c' de communes est défini par :

$$\text{lien}(c, c') = \frac{ks(c, c')}{ks(c).ks(c')}$$

L'inverse de ce taux d'association sera considéré comme une « pseudo distance » entre les deux classes c et c' et sera retenu comme indice d'agrégation dans la classification ascendante hiérarchique construite sur I . Autrement dit, un nœud $n = (c, c')$ regroupant deux classes c et c' aura pour indice d'agrégation (ou niveau d'agrégation) :

$$\nu(n) = \frac{1}{\text{lien}(c, c')}$$

A chaque niveau de la hiérarchie, seront agrégées les deux classes c et c' pour lesquelles le taux d'association $\text{lien}(c, c')$ est maximum. Ainsi, les nœuds n

les plus hauts dans la hiérarchie ainsi construite, qui sont ceux pour lesquels le niveau d'agrégation $\nu(n)$ est le plus élevé, sont donc également ceux pour lesquels le taux d'association lien(c, c') entre les deux classes c et c' qui les définissent est le plus faible.

La hiérarchie obtenue sera coupée de façon à ne retenir que les nœuds n dont les indices d'agrégation $\nu(n)$ sont les plus importants ce qui générera une partition de l'ensemble des communes en classes (les classes dites terminales associées à la coupure) autonomes du point des équipements puisque, par construction, deux quelconques c et c' d'entre elles auront un taux d'association lien(c, c') faible.

L'avantage de la méthode de Domengès sur la méthode Mirabel de l'INSEE provient de l'impossibilité de générer des inversions (voir §2.3.1). Domengès démontre, en effet, dans son article que son indice d'agrégation vérifie l'axiome de la médiane (pour plus de détails sur cet axiome, voir Jambu 1998, ou Diday *et al.* 1982). Ce résultat assure la non-existence d'inversions ainsi que la possible application de techniques de construction accélérée de la hiérarchie (voir les articles de Bruynhooge 1978 ou de De Rahm 1980).

2.4. Représentation de la hiérarchie restreinte à ses nœuds supérieurs

La classification ascendante hiérarchique (Domengès 1982) construite sur l'ensemble I des 652 communes, à partir du tableau symétrique kS , génère une suite d'indices d'agrégation $\{\nu(n)/n \in N\}$ où N est l'ensemble des 651 nœuds formés. Chaque indice d'agrégation mesure la « distance » entre les deux classes de communes qui se sont agrégées au nœud n .

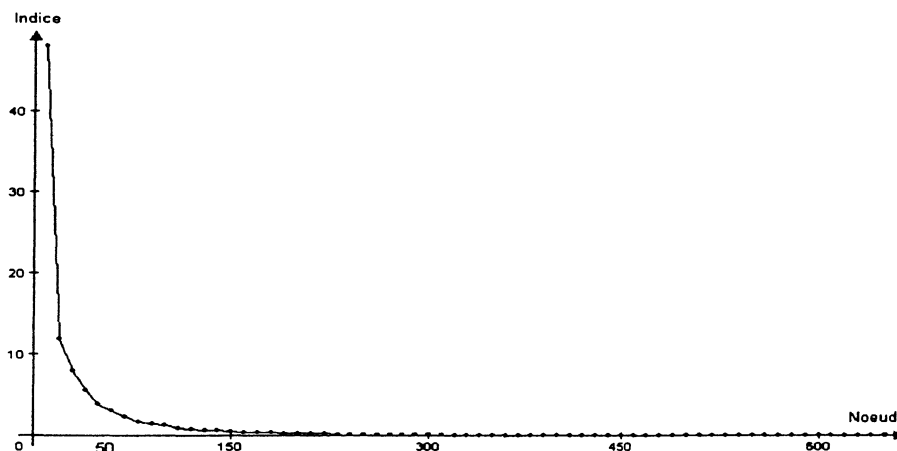


FIGURE 1
Représentation graphique des indices d'agrégation

On trouvera, sur la figure 1 ci-dessus, la représentation graphique de ces indices, les nœuds étant numérotés suivant une lecture descendante de l'arbre (le sommet de

la hiérarchie correspondant au nœud 1 et le nœud le plus bas – c'est-à-dire en fait le premier formé – ayant pour numéro 651).

Après examen de la courbe des indices donnée ci-dessus, nous avons retenu les 50 nœuds supérieurs, ce qui donne naissance à une partition de l'ensemble I des communes en 51 classes terminales. On peut montrer, par ailleurs, que le total des flux internes de ces 51 classes représente environ 90 % du flux total de populations observé dans le département du Nord.

On trouvera, ci-dessous, la représentation graphique (figure 2) de la hiérarchie restreinte aux 50 nœuds supérieurs ainsi que la cartographie des 51 classes de la partition des 652 communes.

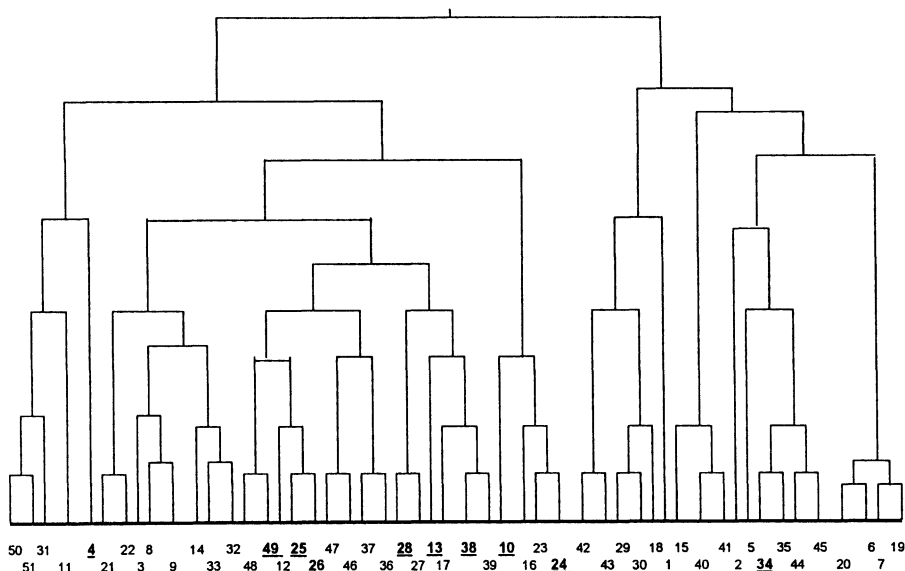


FIGURE 2

Représentation de la hiérarchie restreinte à ses 50 nœuds supérieurs

Il apparaît ainsi que chacune de ces 51 classes forme un ensemble connexe de villes, certaines de ces classes étant des singletons (indiqués en caractères gras sur la figure 2). Il s'agit de :

4 = Coudekerque-Branche	49 = Villeneuve d'Ascq	25 = Mons-en-Baroeul
26 = Lille	28 = Tourcoing	13 = La Madeleine
38 = Marcq-en-Baroeul	10 = Wattrelos	34 = Bruay-sur-Escaut
24 = Roubaix		

Ces 10 singletons correspondent à des villes importantes du Nord, se suffisant à elles-mêmes. Hormis les deux villes Coudekerque-Branche et Bruay-sur-Escaut qui sont des « îlots » au milieu d'autres classes, les 8 autres singletons appartiennent à l'agglomération lilloise. En consultant les données, on constate que ces villes-singletons sont caractérisées par un nombre important d'équipements et par le

fait qu'elles n'utilisent que très peu d'équipements extérieurs et aucun équipement « hors-Nord ». Il convient d'ajouter que les villes Mons-en-baroeul, Bruay-sur-Escaut, Wattrelos, La Madeleine, Marcq-en-Baroeul, Coudekerque-Branches sont non seulement autonomes du point de vue des 39 équipements retenus, mais que très peu de citoyens extérieurs viennent utiliser leurs équipements. Ce qui s'explique par le fait que ces villes sont entourées par des villes elles-mêmes importantes.

3. Détermination des villes-pôles pour chacune des 51 classes obtenues

L'objectif est de déterminer, au sein des classes qui ne sont pas des singletons, les villes-pôles (encore appelées bourgs-centres par l'INSEE). Par ville-pôle, il faut comprendre une commune qui possède un nombre suffisant d'équipements sollicités par des communes-satellites voisines. Les villes-pôles sont donc des communes bien équipées attirant les populations des communes voisines moins nanties.

Pour la détermination des villes-pôles d'une classe donnée c , on revient au tableau k_{CC} croisant l'ensemble des villes de c avec lui-même tel que : $\forall i \in c \forall i' \in c$
 $k_{CC}(i, i') = \text{population}(i) \cdot [\text{nombre d'équipements que } i \text{ va chercher dans } i']$

On considère alors le tableau des fréquences associées f_{CC} :

$$k_{CC} = \sum_{i \in c} \sum_{i' \in c} k_{CC}(i, i')$$

$$f_{CC}(i, i') = \frac{k_{CC}(i, i')}{k_{CC}}$$

A partir d'un seuil donné à priori α donné à priori, une commune i' est déclarée ville-pôle s'il existe au moins une ville i différente de i' pour laquelle $f_{CC}(i, i') > \alpha$.

Autrement dit, un tableau binaire kb_{CC} est ainsi associé au tableau f_{CC} : $\forall i \in c \forall i' \in c$

$$kb_{CC}(i, i') = \begin{cases} 1 & \text{si } f_{CC}(i, i') > \alpha \\ 0 & \text{sinon} \end{cases}$$

Un graphe peut alors représenter la matrice kb_{CC} en considérant chaque ville comme un sommet et en traduisant chaque valeur $kb_{CC}(i, i')$ égale à 1 par une arête du graphe allant de i vers i' . On retrouve alors l'analyse d'un tableau de flux proposé par Camiz (1996) utilisant des graphes ainsi définis à partir de seuils α donnés à priori.

La difficulté consiste en le choix de α . Deux exigences sont à remplir :

D'une part, il faut que le flux de population conservée, en se limitant aux flux de populations qui se déplacent vers les pôles retenus soit maximum et ceci pour perdre le moins d'information possible, et d'autre part il est souhaitable que le nombre de pôles retenus ne soit pas trop grand, ceci pour obtenir une vue synthétique claire des flux de populations dans la classe considérée.

Le flux de population total k_{CC} (voir ci-dessus) d'une classe c correspond aux déplacements de populations d'une ville vers une autre, augmentés des flux internes

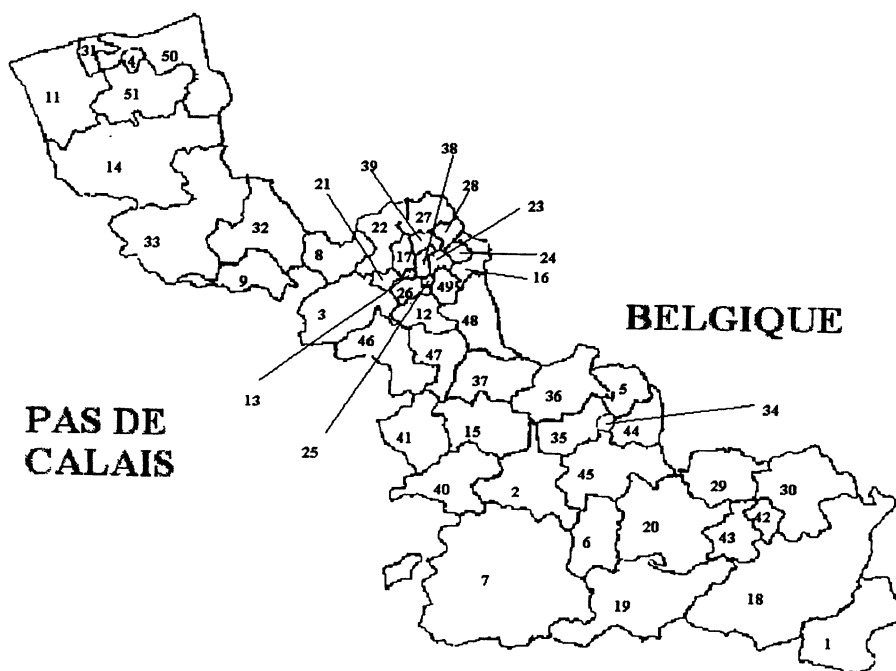


FIGURE 3
Cartographie des 51 classes

à chacune des villes de la classe. Autrement dit,

$$k_{CC} = \sum_{i \in c} k_{CC}(i, i) + \sum_{i \in c} \sum_{\substack{i' \in c \\ i \neq i'}} k_{CC}(i, i')$$

Pour chaque classe c de communes obtenue, un choix de seuil α a été réalisé à partir de l'examen du graphique visualisant les différentes valeurs de α en plaçant en abscisse le nombre de pôles retenus et en ordonnée le flux de population conservé (ce flux correspond au flux total diminué des flux de populations entre deux villes qui ne sont pas des pôles) :

Ainsi, pour la classe 14 (voir figure 3) en faisant varier le seuil α de 1 % à 10 %, nous avons obtenu le graphique ci-après (figure 4).

Il apparaît ainsi que la valeur $\alpha = 5\%$ correspond aux deux exigences citées précédemment : on conserve 76 % du flux de la classe avec un nombre de pôles égal à 5.

A titre indicatif, nous donnons ci-dessous le nombre de pôles et le pourcentage de flux conservé pour chacune des classes qui ne sont pas des singletons (ceci pour $\alpha = 5\%$). On rappelle également que le flux conservé correspond au flux total diminué des flux de population entre deux villes qui ne sont pas des pôles :

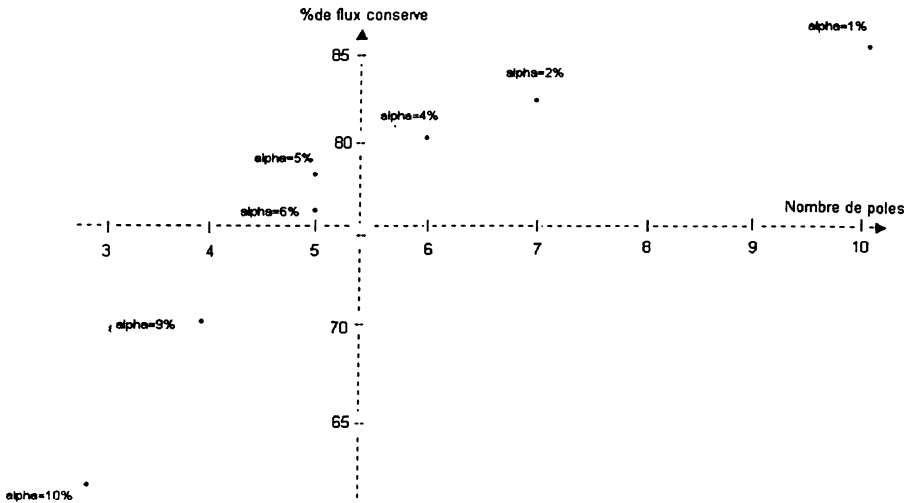


FIGURE 4
Représentation des diverses valeurs du seuil α pour la classe 14

N°c	1	2	3	5	6	7	8	9	11	12	14	15	16	17	18	19	20	21	22	23
N. P	4	4	4	4	3	2	4	3	4	4	5	7	4	3	5	3	4	2	4	2
%C	93	73	70	93	83	59	90	94	93	89	76	87	92	100	73	73	90	100	92	100

27	29	30	31	32	33	35	36	37	39	40	41	42	43	44	45	46	47	48	50	51
5	2	5	3	3	2	6	3	4	2	5	5	1	4	5	4	6	5	5	1	4
100	88	85	100	87	78	91	85	85	100	75	80	93	94	94	82	75	85	76	81	84

On s’aperçoit ainsi, que pour $\alpha = 5\%$ le pourcentage de flux conservé est globalement bon (en moyenne 83 %).

Certaines particularités sont à noter :

– Tout d’abord, la classe 7 qui regroupe un nombre important de communes (87 communes) présente un nombre de pôles égal à 6 et un pourcentage de flux conservé de 70 % pour $\alpha = 3\%$. Il est à noter que cette classe est caractérisée par un grand pôle « Cambrai » qui absorbe à lui seul 50 % du flux total de la classe, le flux restant étant réparti sur un ensemble de communes beaucoup plus petites.

– On note, d’autre part, l’existence de classes où toute commune est un pôle et ceci pour toute valeur du seuil α variant entre 1 % et 10 %. Ce sont des classes comportant un petit nombre de communes, généralement relativement importantes.

Classe 17 = Marquette-lez-Lille, Wambrechies, Saint André-lez-Lille

Classe 21 = Lambersart, Lomme

Classe 23 = Wasquehal, Croix

Classe 31 = Fort-Mardyck, Saint Paul-sur-mer, Grande Synthe.

Classe 39 = Bondues, Mouvaux.

On peut aussi ajouter la classe 27 qui possède 5 éléments et pour laquelle toute commune est un pôle pour $\alpha \leq 5\%$.

Enfin, on peut noter que certaines classes conservent un pourcentage de flux constant au-delà d'un seuil donné α (α restant inférieur à 10 %) :

Classe 9 ($\alpha \geq 3\%$), classe 11 ($\alpha \geq 2\%$), classe 12 ($\alpha \geq 4\%$), classe 22 ($\alpha \geq 3\%$), classe 33 ($\alpha \geq 4\%$), classe 35 ($\alpha \geq 5\%$), classe 42 ($\alpha \geq 6\%$), classe 44 ($\alpha \geq 6\%$), classe 50 ($\alpha \geq 5\%$). Cela signifie encore que, dans les conditions citées précédemment, le nombre de pôles de chacune de ces classes reste constant.

– Nous donnons ci-dessous la liste des 155 pôles retenus pour un choix de seuil $\alpha = 5\%$ pour l'ensemble des classes de communes étudiées (les classes dont toute commune est un pôle apparaissant en gris)

Classes	Pôles ($\alpha=5\%$)
1	Anor Trelon Wignehies Fourmies
2	Bouchain Escaudain Douchy les mines Denain
3	La Bassée Wavrin Haubourdin Loos
5	Escaupont Fresnes sur Escaut Vieux Condé Condé sur Escaut
6	Viesly Haussy Solesmes
7	Cambrai Caudry
8	Chapelle d'Armentières Nieppe Houplines Armentières
9	La Gorge Estaires Merville
11	Loon-plage Grand Fort Philippe Bourbourg Gravelines
12	Lesquin Wattignies Ronchin Faches-Thumesnil
14	Bollezele Esquelbecq Cassel Watten Wormhout
15	Marchiennes Montigny en Ostrevant Flines les Raches Fenain Lallaing Requencourt Somain
16	Leers Lys les Lannoy Hem
17	Wasmuël Sains-aux-Bois
18	Consoire Soire le Château Avesnelles

	Sains du Nord Avesnes sur Helpe
19	Busigny Leers Cateau-Cambresis
20	Englefontaine Gommegnies Poix du Nord Quesnoy
21	
22	Werwicq Sud Quesnoy sur Deule Perenchies Comines
23	
27	
29	Longueville Bavay
30	Ferrières la Grande Feignies Louvroil Jeumont Maubeuge
31	
32	Meterin Steenwerk Bailleul
33	Steenvoorde Hazebrouck
35	Hérin Petite Forêt Wallers Beuvrage Anzin Raimes
36	Lecelles Hasnon Saint Amand les eaux
37	Nomain Landas Beuvry la forêt Orchies
39	
40	Fechain Monchecourt Masny Auberchicourt Aniche

41	Cuincy Auby Waziers Pradelles Douai	45	Trith Saint leger Aulnoy lez Valenciennes Marly Valenciennes
42	Hautmont	46	Thumeries Gondecourt Bauvin Semeries Annoeullin Seclin
43	Pont sur Sambre Leval Berlaimont Aulnoye Aymeries		
44	Quarouble Crespin Quievrechain Onnaing Saint Saulve		
47	Avelin Bersée Cappel en Pevele Pont à Marcq Templeuve	51	Armbouts Cappel Hoynullle Bergues Cappelle la Grande
48	Sainghin en Mélantois Genech Willems Baisieux Cysoing		
50	Dunkerque		

4. Caractérisation des flux de populations internes aux classes

Chacune des classes retenues a été, ensuite, étudiée de manière plus précise d'une part en cartographiant les flux de populations vers les pôles retenus et d'autre part en caractérisant pour chacun de ces pôles les flux de populations observés par les types d'équipements correspondants.

On se limitera, dans cet article, à la présentation de la classe 14 (voir figure 3). Cette dernière est une classe d'effectif moyen comportant 36 communes et 5 pôles : Wormhout (5135 habitants), Cassel (2351 habitants), Watten (3235 habitants), Bollezeele (1500 habitants), Esquelbecq (1907 habitants). Cette classe 14 est donc composée de petites communes situées plutôt en zone rurale.

4.1. Description des pôles de la classe 14

La figure 5 ci-dessous représente les cinq pôles de cette classe caractérisée chacune en abscisse par le nombre d'équipements qu'elle possède (au maximum 36) et en ordonnée par le flux de population engendré (Dans ce flux, chaque habitant i est compté ki fois, ki étant le nombre d'équipements du pôle concerné utilisés par i). Il apparaît, en particulier, que la ville- pôle Bollezeele dont le nombre d'habitants est le plus faible parmi les cinq pôles retenus possède néanmoins un nombre important d'équipements qui engendrent un flux d'environ 50 000 habitants.

On trouvera, sur la figure 6 ci-dessous, la cartographie des flux de populations engendrés par chacun de ces 5 pôles :

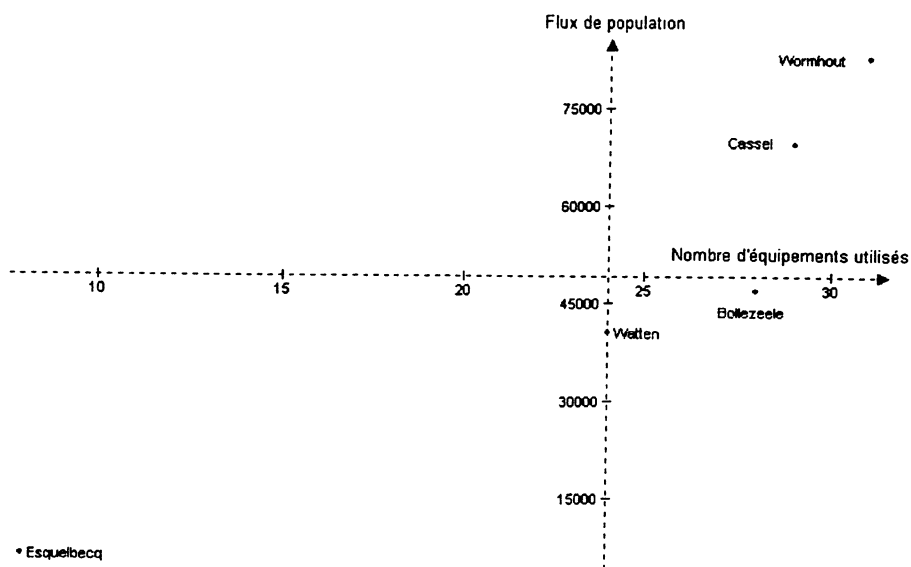


FIGURE 5
Description des pôles de la classe 14

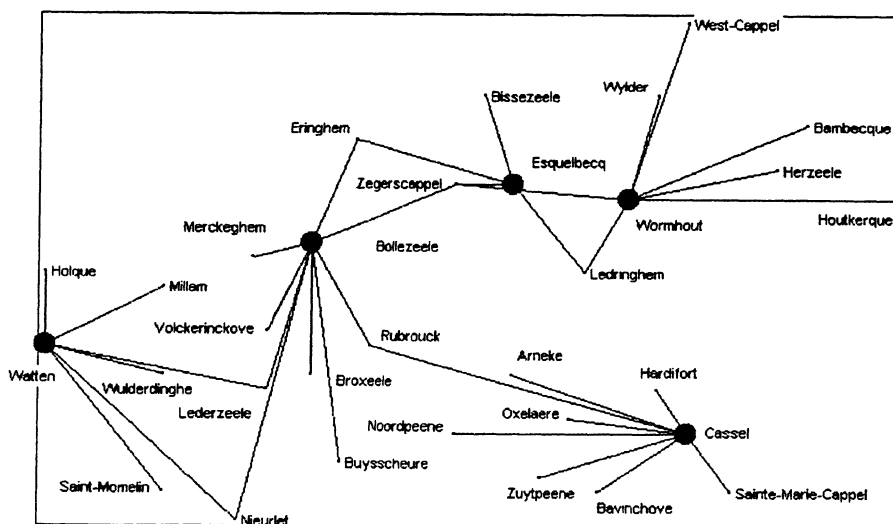


FIGURE 6
Cartographie des flux de la classe 14

4.2. Description des équipements utilisés par les communes de la classe 14

En général, les équipements sont implantés pour desservir sur place une population importante. Ils sont donc rares dans les communes faiblement peuplées et plus nombreux dans celles ayant une taille démographique plus élevée (certaines exceptions peuvent cependant survenir, ainsi la ville Bollezele de la classe 14). Pour illustrer ce phénomène, on peut construire un graphique (graphique 7 ci-dessous) où chaque point représente un équipement, caractérisé en abscisse par le nombre de communes de la classe 14 (au maximum 36) qui le possèdent et en ordonnée par le nombre d'habitants desservis sur « place » (c'est-à-dire dans sa commune). Deux courbes d'effectifs cumulés croissants ont été tracées. La courbe supérieure a été obtenue en classant les 36 communes par taille décroissante et en cumulant leurs nombres et leurs populations. La courbe inférieure a été obtenue de la même façon mais en rangeant cette fois les communes par taille croissante. Il apparaît, ainsi, à la figure 7 que la totalité des équipements est située à proximité de la courbe supérieure, un équipement se trouvant exactement sur cette courbe serait présent dans toutes les communes au-delà d'une certaine taille et absent dans toutes celles de taille inférieure. C'est quasiment le cas par exemple pour l'équipement « Médecin Généraliste ». Une interprétation similaire pourrait être faite pour un équipement situé à proximité de la courbe inférieure, ce que l'on n'observe pas.

Il apparaît, d'autre part, que les équipements comme Cinéma, Maternité, Etablissement hospitalier, Hypermarché ne sont pas dans la classe 14.

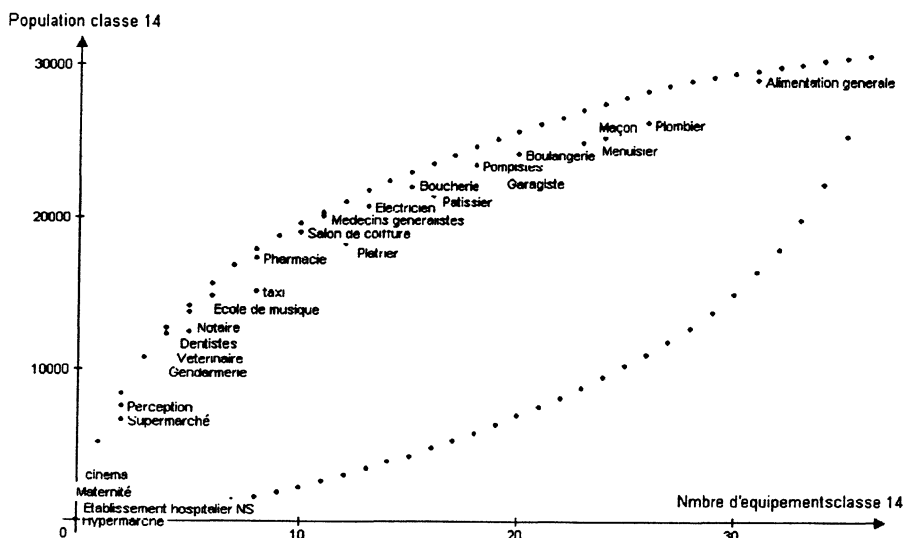


FIGURE 7
Présentation des équipements de la classe 14

4.3. Description des équipements de chacun des cinq pôles de la classe 14

On trouvera sur les graphiques suivants 8,9,10,11,12 la représentation graphique des équipements utilisés pour chacun des 5 pôles retenus, chaque équipement du pôle considéré étant caractérisé en abscisse par le nombre de villes-satellites qui l'utilisent et en ordonnée par le flux de population engendré.

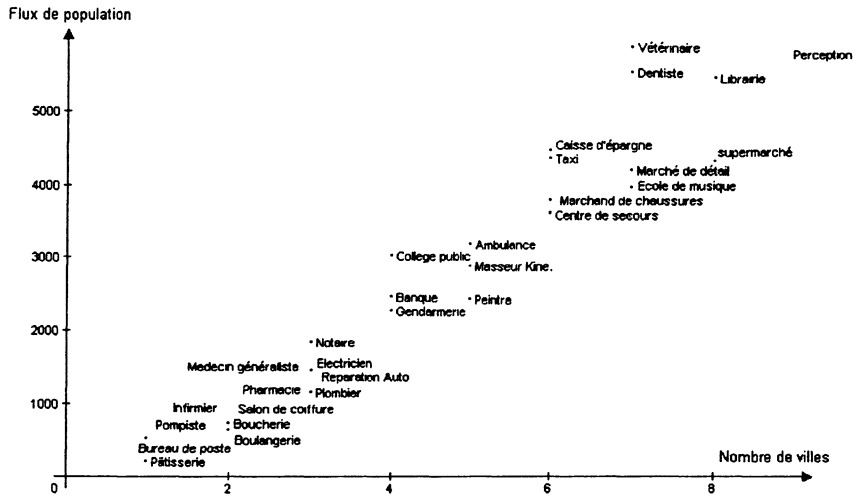


FIGURE 8
Etude du pôle Wormhout

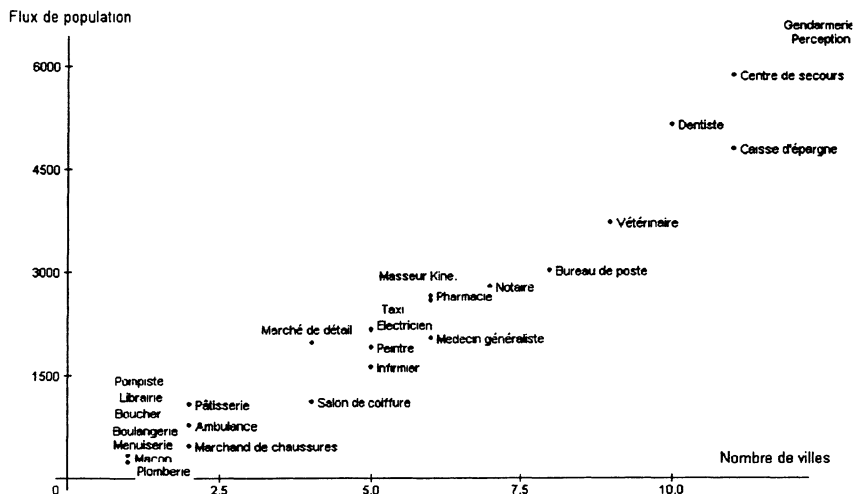


FIGURE 9
Etude du pôle Cassel

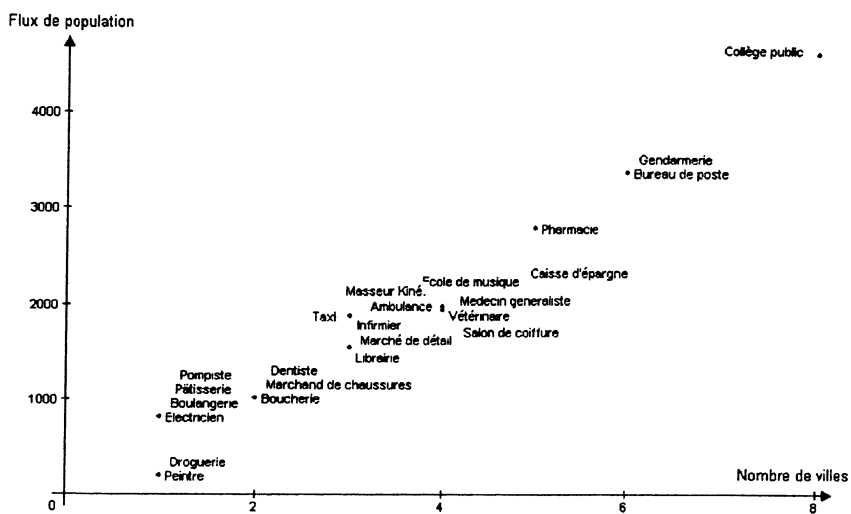


FIGURE 10
Etude du pôle Watten

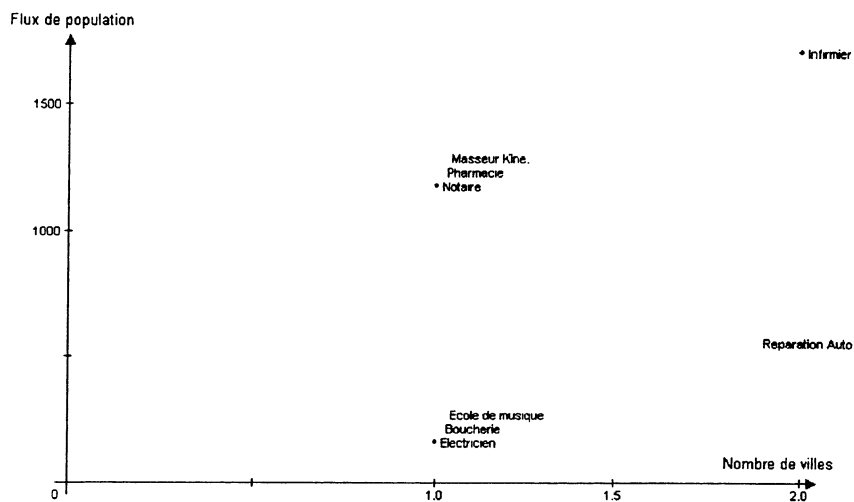


FIGURE 11
Etude du pôle Esquelbecq

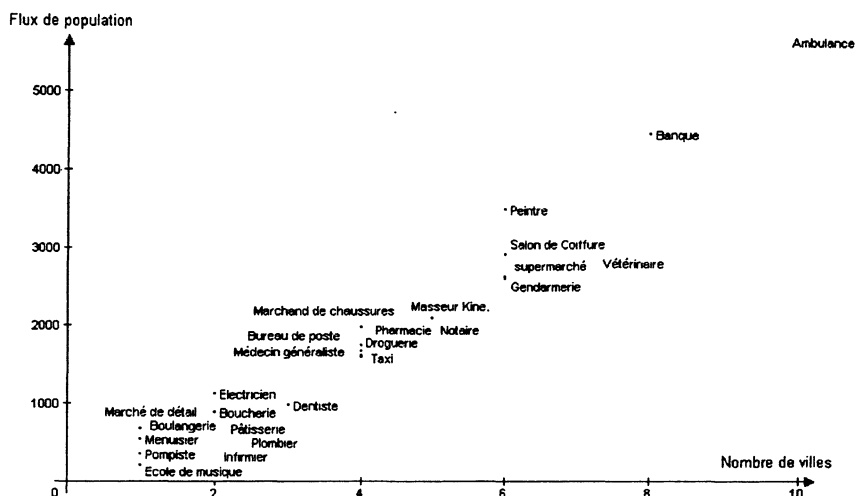


FIGURE 12
Etude du pôle Bollezelee

5. Conclusion

La présente étude réalisée dans le cadre du département du Nord pourrait être étendue à la France entière. La difficulté proviendrait de la taille élevée des tableaux à analyser. Nous proposons, en conclusion, trois grandes étapes d'un possible traitement de ces données :

- Stocker le tableau de flux croisant l'ensemble des villes avec elle-même sous la forme « ligne-colonne » (voir §2.2 autrement dit sous la forme de trois vecteurs)
- Lancer une classification de l'ensemble des communes en une centaine de classes par la technique de Govaert après avoir transformé le tableau de flux en un tableau binaire (voir §2.3.3), cette technique étant très rapide.
- Appliquer ensuite sur chacune des classes obtenues la présente approche développée ici dans le cadre du département du Nord.

Bibliographie

- BENZECRI, J.P. (1973), *L'analyse des données*, tomes 1 et 2. Dunod (Paris).
- BOCK, H.H. (1974), *Automatische Klassifikation*. Vandenhoeck und Ruprecht in Gottingen.
- BRUYNHOOGE, M. (1978), *Classification ascendante hiérarchique des grands ensembles de données, un algorithme fondé sur les voisinages réductibles*. Les Cahiers de l'analyse des données, Vol 3, n° 1, pp. 7-33.

- BRUYNHOOGHE, M. (1978), *Large data set clustering methods using the concept of space contraction*, COMPSTAT 3, Phisika Verlag, Vienna, pp. 239, 245.
- CAMIZ, S. (1987), *The analysis of graph structure as a method for the analysis of the economy Input/output tables*. In Heiberger R.M. (ed), Computer Science and Statistics.
- Proceeding of the 13th Symposium on the interface. American Statistical Association, pp. 169-178.
- CAMIZ, S. (1996), *I/O Analysis : Old and New Analysis Techniques*. In S. Camiz and S. Stefani (eds), Matrices and Graphs theory and Applications. Singapore World Scientific Publishing, pp. 92-119.
- DIDAY, E. (1971), *Une nouvelle méthode en classification automatique et reconnaissance des formes*. Revue de Statistique Appliquée. Vol. 19, n° 2.
- DIDAY, E., LEMAIRE, J., POUGET, J., TESTU, F. (1982), *Eléments d'analyse des données*. Dunod.
- DOMENGÈS, D. (1982), *Classification Ascendante hiérarchique d'après un critère adapté aux tableaux de flux*. Les cahiers de l'analyse des données. Vol. VII, n° 2, pp. 169-172.
- GOVAERT, G. (1984), *Classification croisée*. Thèse d'état. Université Paris VI.
- GOVAERT, G. (1984), *Classification simultanée de tableaux binaires*. Data Analysis and Informatics 3, eds. Diday E., Jambu M., Lebart L.
- JAMBU, M. (1978), *Classification Automatique pour l'analyse des données* (Tomes 1 et 2), Dunod, Paris.
- DE RAHM, C. (1980), *La CAH selon la méthode des voisins réciproques*. Cahiers de l'analyse des données, Vol V, n° 2, pp. 135-144.