

1

UNIVERSITE PARIS I – PANTHEON SORBONNE
U.F.R. DE MATHEMATIQUES et INFORMATIQUE

Année 1999

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS I

Discipline : Mathématiques

Présentée et soutenue publiquement

Par

Patrick ROUSSET

Titre :

Applications des algorithmes d'auto-organisation
à la classification et à la prévision

Directeur de Thèse : Marie COTTRELL

JURY

Michel Verleysen

Jeanny Herault

Marie Cottrell

Jean-Pierre Fénelon

François Gardes

Stephane Canu

Remerciements

Si pour beaucoup la thèse est le moyen de trouver un emploi dans la recherche ou l'enseignement, c'est mon travail à l'université et une expérience préalable d'enseignement qui m'ont donné l'envie de faire celle-ci. Elle m'a permis de continuer à découvrir et progresser, de concrétiser des idées nées de rencontres et de lectures, ainsi que de mieux me connaître.

Au moment de conclure ce travail, mes premières pensées vont à mes parents et mes proches. Etre le témoin privilégié de la destinée de mes élèves a élargi l'angle de vue sur mon propre parcours, c'est pourquoi je suis heureux d'avoir ici l'opportunité de remercier tous ceux qui m'ont aidé, encouragé ou soutenu tout au long de mes études.

En ce qui concerne la thèse, elle-même, je dois son aboutissement aux professeurs Marie Cottrell et Bernard Girard. Décrire leur contribution en réduirait l'importance. Je les remercie donc succinctement d'avoir ajouté à leur grande compétence la gentillesse et la disponibilité. Ces qualités se trouvent d'ailleurs facilement au SAMOS (ainsi que la réponse à la plupart de mes questions). Merci aussi à Yvonne Girard et au SCIPRE de m'avoir fourni matériel et assistance, à Chaiya pour la qualité des impressions couleurs et à ma cousine Dominique pour sa relecture. J'ai eu aussi plaisir à travailler en collaboration avec Patrice Gaubert, Christiane Guinot, François Gardes, Yvonne Girard, Bertrand Maillet, Christian Derquenne et Morgan Mangeas, à avoir une rencontre fructueuse avec Michael Jordan.

Je voudrais remercier aussi Michel Verleysen et Jeanny Herault pour avoir accepté la charge de rapporter cette thèse, pour leurs remarques et leurs encouragements ainsi que les membres du jury Stéphane Canu, Jean-Pierre Fénelon et François Gardes.

Un clin d'œil à Krystyna, Thierry, Arnaud, Olivier, Salah, ... et aux habitants de Sari d'Orcino. Mes encouragements enfin pour Elena, Maxence, Alexandre, Mika, Clara, Laura, Hugo et les autres pour qui le chemin est encore long.

TABLE DES MATIERES

Introduction	1
1^{ère} PARTIE : ANALYSE DE DONNEES A L'AIDE DES CARTES DE KOHONEN . 5	
1 Analyse de données – Problème et exemples	7
1.1 Le problème	7
1.2 Présentation de trois exemples de bases de données	8
2 La classification de Kohonen	15
2.1 Introduction de l’algorithme de Kohonen.	15
2.2 Quelques algorithmes de classification.	16
2.3 Algorithme de Kohonen	23
2.4 La relation entre la méthode basée sur la classification de Kohonen et les autres méthodes	34
3 Analyse de données en utilisant les cartes de Kohonen	37
3.1 Analyse de données appliquées en pratique	39
3.1.1 Analyse linéaire des données	39
3.1.2 Analyse de données à l’aide des réseaux de neurones	41
3.2 Analyse de données à l’aide des cartes de Kohonen.	41
3.3 2 niveaux de classification	43
3.4 Représentation des distances inter classes.	45
3.5 Contenu des classes et homogénéité	49
3.6 Croisement de la classification avec une variable qualitative.	50
3.7 Analyse restreinte à une région de la carte.	55
3.7.1 Projection des observations de A dans le plan $P(A)$	56
3.7.2 Représentation des classes voisines.	59
3.8 Comparaison avec les méthodes classiques.	62
3.8.1 L'exemple du fer à cheval – Les représentations issues de l'ACP	63
3.8.2 Utilisation de la classification à l'aide de l'algorithme de Kohonen	65
3.8.3 Comparaison entre la classification de Kohonen et l'analyse factorielle	66
2^{ème} PARTIE : PREVISION A L'AIDE DES CARTES DE KOHONEN	69
4 Problématique.	71
4.1 Problème	71
4.2 Domaine d'application.	72
4.3 Exemple	73

4.4	Les modèles classiques.	77
4.4.1	Un modèle linéaire: L'Armax.	77
4.4.2	Un modèle neuronal: Le perceptron multicouches	79
4.4.3	Conclusion.	88
5	La MEP : une méthode de prévision dans le cas d'une série chronologique qui cumule deux évolutions de temps ayant chacune sa cadence propre.	89
5.1	Définition et notations	89
5.2	Méthode de prévision.	91
6	Application de la MEP à l'exemple de la prévision de la consommation électrique nationale polonaise.	95
6.1	Classification des profils.	96
6.1.1	Utilisation d'une classification hiérarchique.	97
6.1.2	Utilisation de la classification de Kohonen.	100
6.1.2.1	Variantes de l'algorithme de Kohonen susceptible d'améliorer la classification des courbes de charge	101
6.1.2.2	Classification et prévision des profils.	105
6.2	Estimation des courbes de charges électriques et performances	122
6.2.1	Estimation des courbes de charges électriques par la méthode MPE.	122
6.2.2	Un modèle de référence pour valider la méthode par comparaison.	128
6.2.3	Performances de la MEP.	131
6.3	Analyse de l'erreur.	132
6.3.1	Analyse de l'erreur de prévision du profil.	133
6.3.2	Analyse de l'erreur de prévision de la courbe de charge.	140
7	Quelques réflexions sur la méthode.	151
7.1	Normalisation de l'estimé du profil et renormalisation des poids à chaque itération de l'apprentissage.	152
7.2	Options techniques.	153
7.3	Choix de la classification	155
7.4	Limites.	162
8	Conclusion et perspectives.	165

INTRODUCTION

Les travaux scientifiques nécessitent souvent une confrontation entre les nouvelles techniques et celles existantes. On aura également cette exigence dans la suite, c'est pourquoi il peut être intéressant de faire un point sommaire sur l'évolution des outils statistiques et de se demander ce que l'on attend d'une nouveauté. En particulier, quelles caractéristiques elle doit posséder pour répondre à l'évolution des moyens comme par exemple le matériel et pour servir de relais aux méthodes existantes pour résoudre les problèmes.

Evolution de la statistique :

Les statistiques ont rapidement évolué au cours des dernières décennies. Les techniques se sont en effet beaucoup développées et de nouvelles sont apparues. De plus, son domaine d'application s'est élargi et la population des utilisateurs s'est diversifiée. Cette évolution provient des progrès de l'informatique qui a d'abord révolutionné le monde de la statistique à son arrivée et l'a, depuis, fait évoluer continuellement. La capacité grandissante des ordinateurs a imposé un renouvellement très fréquent des techniques. Parmi les grands axes d'influence de l'informatique sur les techniques, on peut citer trois exemples.

La capacité de stockage des ordinateurs a permis un accroissement de la taille des bases de données dans des proportions telles qu'il est aujourd'hui courant de disposer de recueil de données dont le nombre d'individus s'exprime en millions et celui des variables en centaines. Ce phénomène réclame naturellement une approche et des techniques nouvelles.

Les progrès des outils informatiques pour réaliser des représentations graphiques ont orienté les statistiques vers des techniques qui aboutissent à des résultats qui se traduisent par des représentations graphiques sous forme de tableaux ou de cartes. Par exemple, les analyses factorielles, dont le but est de représenter "au mieux" l'information sur des plans, ont pris une grande importance dans l'analyse de données multidimensionnelles.

D'autre part, la vitesse d'exécution grandissante des ordinateurs a permis d'envisager l'utilisation d'algorithmes qui ont un coût élevé en terme de temps de calcul. En particulier, cela a provoqué une diversification des techniques de classification.

Outre l'évolution des techniques, l'informatique a aussi modifié le cadre d'application des statistiques. Les domaines d'étude se sont diversifiés et les techniques sont de plus en plus utilisées par des non statisticiens, spécialistes d'autres sciences comme l'économie, la finance, l'assurance, ou de façon plus générale la plupart des ingénieurs. Cette évolution provient de la mise à disposition des logiciels qui proposent une gamme de modèles statistiques qu'il suffit d'exécuter sur son ordinateur personnel. Il est à noter que les sorties sous forme de graphiques et de tableaux ont largement contribué à la vulgarisation des techniques statistiques.

En même temps de l'informatique, le développement récent des réseaux de neurones apporte aussi une source de renouvellement pour les statistiques. De nouveaux algorithmes répondent ainsi au besoin d'approches nouvelles déjà évoqué. En particulier, ils permettent une étude non linéaire mieux adaptée à certaines structures.

"Quelles caractéristiques pourrait avoir un nouvel algorithme qui s'inscrirait dans l'évolution des statistiques?" :

D'après le bilan historique fait précédemment, il apparaît que c'est plus au besoin d'analyse qu'il doit répondre qu'au problème des moyens (les problèmes de calcul étant moindres). Il doit fournir un moyen de traiter des bases de données de plus en plus importantes et complexes. Un algorithme issu des réseaux de neurones peut par exemple répondre à cette demande. De plus, pour être appliqué par des non spécialistes et suivre ainsi l'évolution de l'analyse classique, il doit être relativement simple à utiliser et doit fournir des sorties sous une forme qui facilite leur interprétation. C'est-à-dire sous forme de tableaux ou

de graphiques. Il est d'ailleurs possible d'envisager des représentations graphiques autres que les projections linéaires. Avant de présenter les avantages de celles proposées dans la suite, nous ne résistons pas au plaisir de contempler une de celles produites par un analyste plus doué :



Picasso : L'acrobate

Contexte et objectif de la thèse :

L'esprit dans lequel nous avons essayé de concevoir un modèle de prévision des courbes de consommation nationale électrique française est à rapprocher d'une citation de Jean-Paul Benzécri présente dans l'introduction de l'ouvrage "Statistiques exploratoire multidimensionnelle" [68], "le modèle doit suivre les données et non l'inverse". On a, ainsi, mis au point une méthode de prévision qui s'appuie sur une analyse de données, car cela paraissait mieux adapté à cette série chronologique. On a choisi de le faire à partir de la classification de Kohonen car elle offrait plus de possibilités. On a développé les outils

d'interprétation de cette dernière pour améliorer notre potentiel d'analyse et par là même la qualité de notre prévision. En parallèle, d'autres études se sont présentées, elles appartenaient à des domaines variés tels que l'analyse de la *consommation des canadiens* ou *du chômage* qui sont fournies en annexe. Les outils précités se sont avérés performants pour les analyser, mais on a souvent dû les faire évoluer ou en ajouter d'autres. Cet échange entre les problèmes et les solutions a abouti à un ensemble de moyens d'analyse complémentaires et adaptatifs.

C'est pourquoi, dans la suite, l'analyse de données sera présentée indépendamment de la prévision. Elle constituera la première partie, la prévision étant le sujet de la seconde. Pour montrer l'adaptabilité des outils développés, on illustrera leur utilisation à l'aide d'exemples de bases de données variées, souvent issues d'études réelles. Par contre, leurs études complètes fournies en annexe ne sont pas traitées.

A la suite de la présentation de l'analyse de données et de sa comparaison à l'association *classification – analyse factorielle*, on montrera dans la seconde partie comment s'appuyer sur celle-ci pour mettre au point une méthode de prévision qui s'écarte des méthodes récursives et s'adapte à des contextes nouveaux. En particulier, on l'appliquera aux données de consommation électrique polonaise, mis gentiment à notre disposition par le professeur Osowski de l'université Warsaw Technical, pour lesquelles les techniques précitées ont montré à la fois leurs qualités et leurs limites. L'étude de la consommation française est présentée en annexe.

1^{ère} PARTIE

ANALYSE DE DONNEES A L'AIDE DES CARTES DE KOHONEN

Chap.1 **Analyse de données – Problème et exemples**

Chap.2 **La classification de Kohonen**

Chap.3 **Analyse de données en utilisant les cartes de Kohonen**

1 Analyse de données - Problème et exemples.

1.1 Le problème

L'apparition puis le développement des ordinateurs ont permis d'envisager de nouvelles techniques d'analyse de données, notamment en exécutant rapidement certains calculs tels que l'inversion de matrice ou la diagonalisation. C'est le cas des analyses factorielles – comme par exemple l'analyse en composantes principales ou ACP – qui ont répondu au besoin de visualiser les données. Ces méthodes réalisent en effet, des projections sur les plans dits "principaux" (contenant le maximum d'inertie) où l'on voit le mieux le nuage de points.

Mais il existe un besoin de compléter ces méthodes par de nouvelles qui seraient capables de traiter des bases de données de grande taille, ayant une structure non linéaire, et par conséquent mal expliquées par les outils traditionnels. Certains réseaux de neurones, comme le perceptron multicouche, permettent de remplacer les analyses classiques. Mais ils n'offrent pas toujours de représentations graphiques et les interprétations sont parfois

difficiles. La classification de Kohonen a, quant à elle, le double avantage de faire une analyse non linéaire de la base de données et de se prêter à des représentations graphiques faciles à utiliser.

Dans ce chapitre vont être développées différentes façons d'exploiter cette dernière méthode pour répondre à plusieurs problèmes de l'analyse de données tels que la classification, l'analyse de la dispersion, la caractérisation des individus par des variables explicatives qualitatives, etc. Les techniques qui vont être présentées seront appliquées à des exemples simulés pour l'illustration ou issus d'études pratiques dont le lecteur pourra trouver les analyses complètes en annexe.

1.2 Présentation de trois exemples de bases de données.

Soit une base de données X de N observations dans laquelle chaque individu i est décrit par p variables quantitatives et q variables qualitatives. On note $x_i=(x_i^1, \dots, x_i^j, \dots, x_i^p)$ le vecteur de \mathbf{R}^p constitué des p variables quantitatives.

Trois exemples de bases de données vont servir à illustrer les techniques que nous proposons. L'objectif n'est pas d'en faire l'étude, mais d'exploiter la spécificité de chacune pour illustrer les différentes parties développées par la suite.

Le premier exemple, d'effectif réduit, est construit artificiellement pour visualiser le mécanisme d'apprentissage de l'algorithme de Kohonen. Le deuxième – issu d'une étude pratique – permet d'effectuer des analyses qui combinent des variables quantitatives et qualitatives, il servira de plus à donner un exemple de classification avec la distance du χ^2 . Enfin, le troisième utilise une notion temporelle qui facilitera l'interprétation des rapprochements des classes à l'intérieur des cartes de Kohonen.

1^{er} exemple :

Le premier exemple est constitué d'une base de données simulées contenant 300 individus et 11 variables. Chaque individu est construit à partir du mélange aléatoire des 3 formes a, b, c de \mathbf{R}^{11} de la figure 1 – mieux visualisées par le type de représentation de la figure 2, adopté pour la suite, qui joint les points indiquant les 11 valeurs. Ceci revient à prendre une combinaison linéaire des trois vecteurs de base a, b et c fixés – du type $\mathbf{a}a+\mathbf{b}b+\mathbf{g}c$ où \mathbf{a}, \mathbf{b} et \mathbf{g} sont des scalaires aléatoires compris entre 0 et 1 – afin d'obtenir des formes variées, comme par exemple celles de la figure 3.

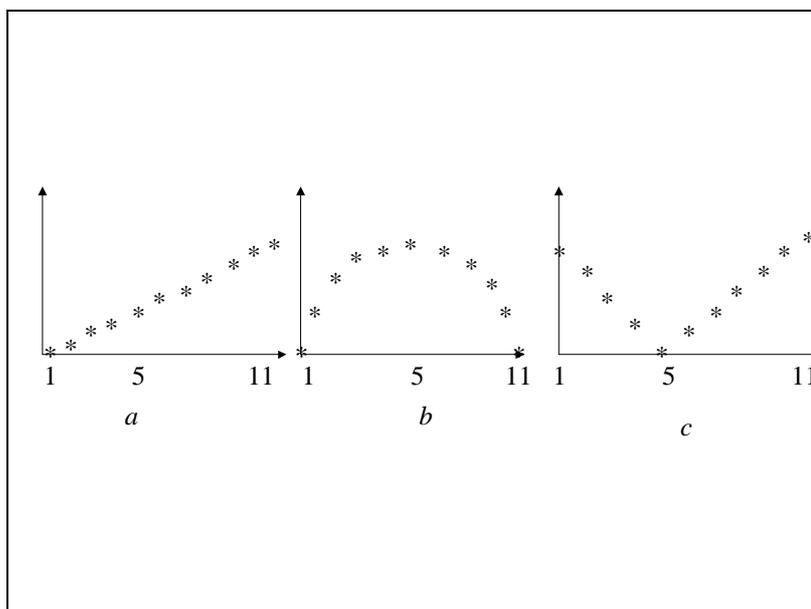


Figure 1 : Les trois formes a, b et c sont construites à partir de 11 variables

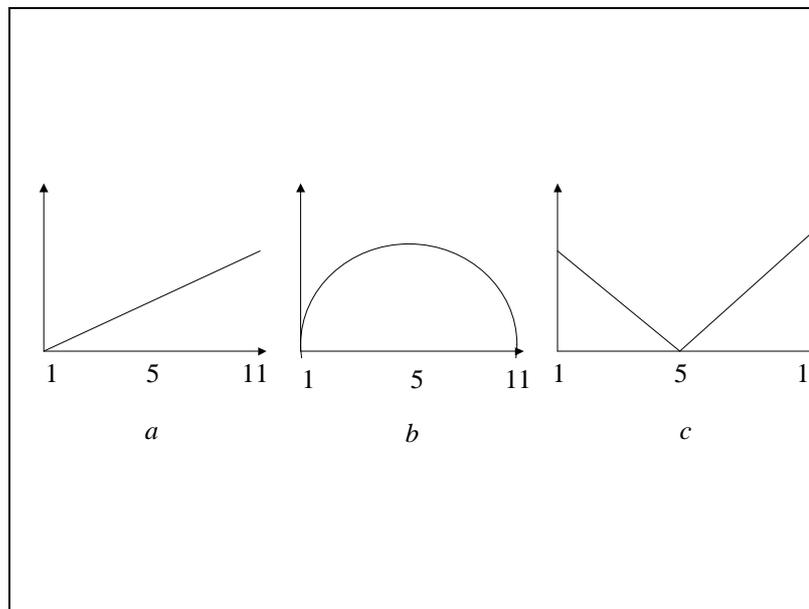


Figure 2 : Les 11 points sont joints pour mettre en évidence la forme

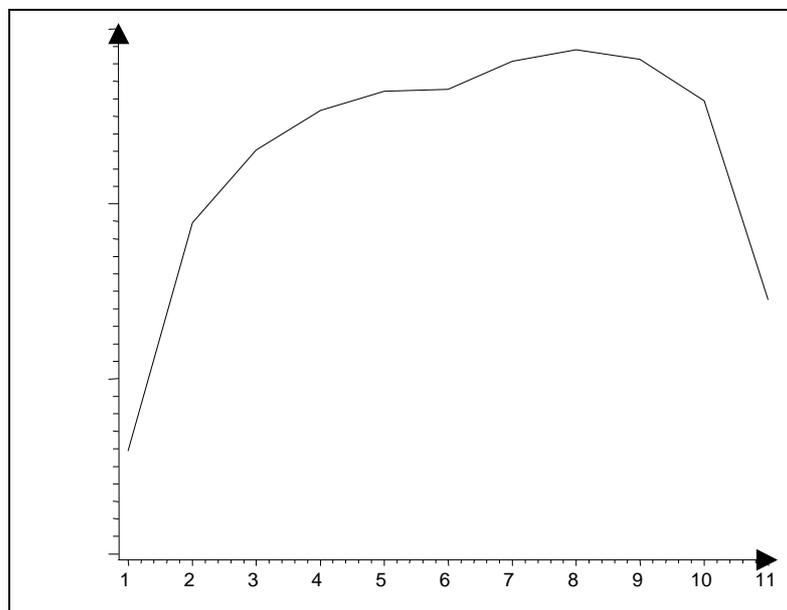


Figure 3 : Exemple d'une forme construite à partir du mélange de a , b et c avec $a=0.44$, $b=0.17$ et $g=0.12$

2^{ème} exemple :

Cet exemple est issu de deux enquêtes de consommation réalisées par Statistiques Canada entre 1986 et 1992 auprès d'environ 10000 ménages. Un individu est défini par son profil de consommation, c'est-à-dire par la répartition de sa consommation dans 20 produits (Logement, achat de véhicule etc.¹). Pour un individu, la valeur associée à une variable représente le ratio de la consommation de ce produit rapporté à la consommation totale (par exemple, consommation dans le logement / consommation totale). Les variables sont donc des pourcentages, elles prennent des valeurs positives et leur somme vaut 1.

$$j=1, \dots, 20 \quad x^j \geq 0$$

et
$$\sum_{j=1}^{20} x^j = 1$$

Comparer deux individus revient à comparer deux profils. Lorsque les observations sont des pourcentages, il est classique de choisir la distance du χ^2 , c'est donc celle qui sera utilisée dans cet exemple. La figure 4 représente la consommation d'un individu i . Sur l'axe des abscisses sont indiqués les produits de consommation, l'ordre est alphabétique et n'a donc aucun sens (ce n'est pas un axe temporel). De même, les points sont reliés entre eux dans le seul but de rendre le profil plus lisible.

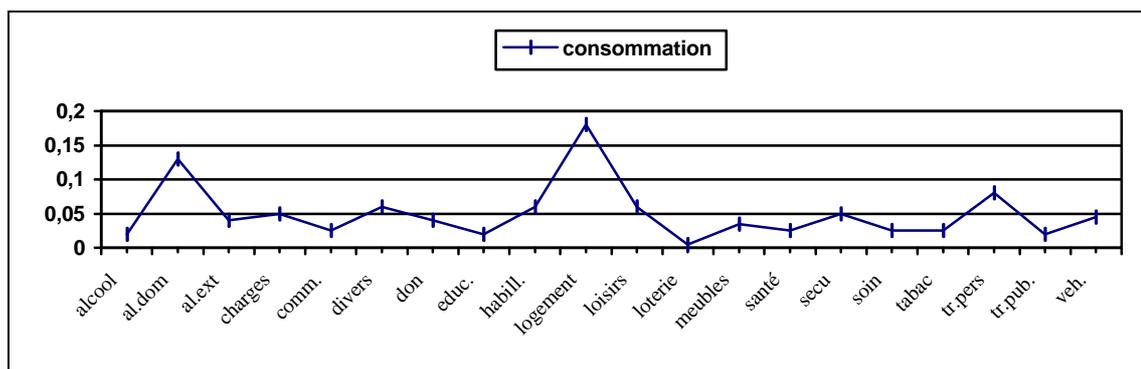


Figure 4 : Profil de consommation de l'individu i

¹ La liste exhaustive des variables de consommation est la suivante : alcool, alimentation au domicile, alimentation à l'extérieur, charges, communication (dépense de téléphone), divers (intérêts des prêts personnels, produits d'entretien), dons, éducation, habillement, logement, loisirs, loterie, meubles, santé, dépenses de sécurité (assurance chômage, vie, retraite), soins, tabac, transport personnel, transport public, achat de véhicule

Cette enquête fournit aussi des caractéristiques qualitatives complémentaires sur les individus : le niveau d'éducation, la tranche d'âge, la détention de logement et le niveau de pauvreté². Cet exemple permettra donc, dans la suite, d'illustrer la façon dont on peut utiliser les cartes de Kohonen – qui réalisent une classification des individus par rapport des variables quantitatives (ici la consommation) – afin de réaliser la cartographie d'un autre critère qui n'a pas servi à la classification (par exemple la détention de logement).

3^{ème} exemple :

Cet exemple a été construit artificiellement à partir d'une base de données réelles. Il contient 2000 observations qui représentent les relevés journaliers d'une mesure faite toute les deux heures (la première valeur est le relevé à 0h00, le deuxième à 2h00, le troisième à 4h00, etc.). Sur les figures 5 et 6 sont représentées les 12 mesures d'une journée pour les individus i_1 et i_2 de la base de données.

On exploitera cet exemple pour illustrer les études sur la dispersion et les relations entre les unités de la carte de Kohonen qui s'expliquent dans ce cas particulier par une notion familière aux statisticiens, l'évolution temporelle. Cet exemple permettra ainsi de faire le lien avec *l'exploitation des cartes de Kohonen dans le cadre de la prévision* présentée au chapitre 2.

² Le niveau de pauvreté a été évalué par des économistes à partir de critères dont l'un est la consommation.

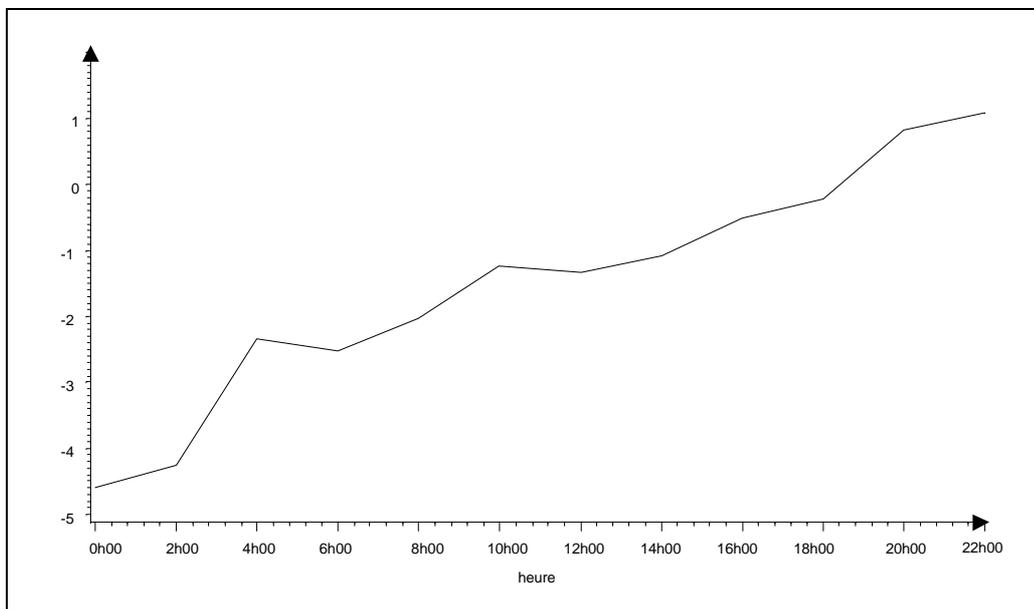


Figure 5 :Relevé bi-horaire du jour i_1

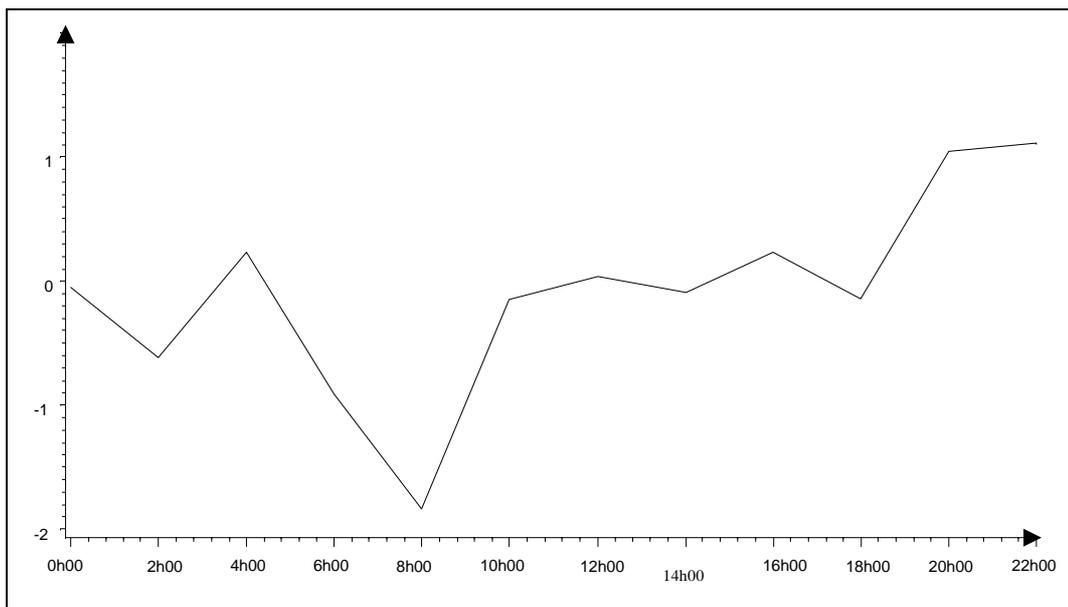


Figure 6 :Relevé bi-horaire du jour i_2

2 La classification de Kohonen

2.1 Introduction de l'algorithme de Kohonen.

Teuvo Kohonen a proposé dès 1982 (Kohonen, 1982 [62, 63]) un algorithme qui produit une carte d'organisation topologique ([64, 65]). Le processus ne dépend que des entrées et ne nécessite pas l'intervention d'un superviseur, on parle dans ce cas d'auto-organisation. Sa fonction principale est de faire correspondre les éléments de l'espace d'entrée avec des unités ordonnées sur une carte – qui est une représentation graphique où chaque unité est entourée de ses voisines, les voisinages ayant été définis à priori. Le résultat est une fonction de l'espace des entrées vers l'ensemble des unités, telle que les images de deux éléments voisins au sens d'une certaine distance dans l'espace des entrées sont la même unité ou des unités voisines sur la carte.

Les propriétés de l'algorithme et de ses cartes de sortie permettent conjointement d'analyser des données qui ont une structure non linéaire et de réaliser des représentations graphiques qui mettent en évidence certaines de leurs caractéristiques. L'application la plus courante est la classification de l'espace d'entrée, où l'on définit une notion de voisinage entre les classes qui n'est pas prise en compte par les méthodes de classification classiques.

2.2 Quelques algorithmes de classification.

Introduction :

On distingue deux grandes catégories d'algorithmes de classification. La première consiste à effectuer une partition de l'espace en un nombre de classes fixé, c'est le cas de la méthode *des centres mobiles* et ses variantes issues des travaux de Forgy (1965) [43], Thorndike (1953) [95], Mac Queen (1967) [79], Ball et Hall (1967) [6]. La deuxième fournit une classification pour chaque niveau de regroupement, chacune se déduisant de l'agrégation ou de la dissociation des classes du niveau précédent. *La classification hiérarchique de type WARD* procède par exemple par agrégation (Ward J.H. (1963) [99]). Ces méthodes sont présentées dans les ouvrages de statistiques classiques tels que *Statistique Exploratoire Multidimensionnelle* (Lebart, Morineau, Piron (1995) [68]) ou *Probabilités Analyse de Données et Statistiques* (Saporta, S. (1990) [94]).

1^{er} type de classification : les classifications à nombre de classes fixé.**a) La méthode des centres mobiles :**

Cette méthode, dont la paternité est attribuée à Forgy, est particulièrement adaptée aux données de grande taille. En effet, dans sa mise en place informatique, le tableau de données conservé en mémoire auxiliaire est lu plusieurs fois mais de façon séquentielle, et ainsi n'occupe jamais de zone mémoire importante. Son efficacité est en grande partie attestée par les seuls résultats expérimentaux. Ce type de classification est souvent utilisé comme technique de description et de réduction en association avec des analyses factorielles. Pour plus de précisions, on peut se rapporter aux ouvrages de Sokal et Sneath (1963) [93], Lance et Williams (1967) [66] ou plus récemment Gordon (1987) [52].

- L'algorithme :

Considérons l'ensemble X de N individus décrits par p variables à partitionner en U classes à l'aide d'une distance d (souvent la distance euclidienne ou celle du χ^2). Les étapes de l'algorithme sont les suivantes :

Etape 0:

On détermine U centres provisoires de classes, les centres $\{G_1^0, \dots, G_k^0, \dots, G_U^0\}$ (tirés selon un tirage pseudo aléatoire sans remise de U individus dans la population comme le préconise Mac Queen (Mac Queen J. B. (1967) [79]).

Les U centres induisent une première partition $\{C_1^0, \dots, C_k^0, \dots, C_U^0\}$. L'individu i appartient à la classe C_k^0 s'il est plus proche de G_k^0 que de tous les autres centres.

Etape 1:

On définit U nouveaux centres de classes $\{G_1^1, \dots, G_k^1, \dots, G_U^1\}$ en déterminant les centres de gravité des classes $C_1^0, \dots, C_k^0, \dots, C_U^0$. Ces nouveaux centres induisent une nouvelle partition formée des classes notées $\{C_1^1, \dots, C_k^1, \dots, C_U^1\}$.

Etape m :

On détermine U nouveaux centres de classe $\{G_1^m, \dots, G_k^m, \dots, G_U^m\}$ en prenant les centres de gravité des classes $C_1^{m-1}, \dots, C_k^{m-1}, \dots, C_U^{m-1}$, ce qui induit la nouvelle partition $\{C_1^m, \dots, C_k^m, \dots, C_U^m\}$.

- Montrons que le processus converge :

On note $v(m)$ la mesure de la dispersion autour des centres de classes à l'étape m :

$$v(m) = \sum_{k=1}^U \sum_{i \in C_k^m} \|x_i - G_k^m\|^2$$

Considérons la somme des carrés des écarts à la moyenne intra classe à l'étape m , $V(m)$

$$V(m) = \sum_{k=1}^U \sum_{i \in C_k^m} \|x_i - G_k^{m+1}\|^2 \text{ où } G_k^{m+1} \text{ étant le centre de gravité de la classe } C_k^m.$$

On va montrer que $v(m) \geq V(m) \geq v(m+1)$:

Ceci montrera la convergence vers le minimum à la fois de la dispersion autour des centres de classes et de la somme des carrés des écarts à la moyenne intra classe.

D'après le théorème de Huygens,

$$v(m) = V(m) + \sum_{k=1}^U n_k \|G_k^{m+1} - G_k^m\|^2 \text{ où } n_k \text{ est l'effectif de } C_k$$

$$d'où \quad v(m) \geq V(m)$$

$$et \quad v(m) = V(m) \iff \forall k=1, \dots, U \quad G_k^{m+1} = G_k^m$$

Montrons que $V(m) \geq v(m+1)$:

A l'étape $m+1$

$$v(m+1) = \sum_{k=1}^U \sum_{i \in C_k^{m+1}} \|x_i - G_k^{m+1}\|^2$$

L'expression de $V(m)$ est identique à celle de $v(m+1)$ si on remplace C_k^m par C_k^{m+1} .

Or par définition C_k^{m+1} est celui qui minimise la dispersion autour de $\{G_k^{m+1}\}_{k=1, \dots, U}$

$$d'où \quad \sum_{i \in C_k^{m+1}} \|x_i - G_k^{m+1}\|^2 \leq \sum_{i \in C_k^m} \|x_i - G_k^{m+1}\|^2$$

$$donc \quad V(m) \geq v(m+1).$$

Conclusion

$$v(m) \geq V(m) \geq v(m+1).$$

$$avec \quad (1) \quad v(m) = V(m) \iff \forall k=1, \dots, U \quad G_k^{m+1} = G_k^m$$

Une suite positive strictement décroissante converge, donc l'algorithme converge vers un minimum local. De plus, si deux étapes conduisent à la même partition, cette dernière est une solution de l'algorithme.

b) Simple competitive learning ou algorithme de Lloyd : une version stochastique en ligne des centres mobiles :

L'algorithme *SCL* est une version stochastique de l'algorithme de Forgy présenté au paragraphe précédent, qui a la particularité de déplacer un seul centre de classe à la fois lors de la présentation d'un individu, ce qui le rend mieux adapté aux grandes bases de données. Son principe, dont le détail peut être trouvé dans les ouvrages Hertz J. et al. (1991) [54] Linde Y. et al. (1980) [74], Lloyd S.P. (1982) [75] et Mac Queen J. (1967) [79], est le suivant :

Les centres de classes G_1, \dots, G_U sont initialisés de façon aléatoire.

A l'étape t , on présente l'observation $x = x(t+1)$ extraite au hasard. Puis, on définit l'indice gagnant u_0 comme celui du centre de classe qui minimise la distance euclidienne au vecteur présenté:

$$\|x - G_{u_0}(t)\| = \min_u \|x - G_u\|$$

G_{u_0} est alors modifié selon la formule:

$$G_{u_0}(t+1) = G_{u_0}(t) + \mathbf{e}(t)(x - G_{u_0}(t))$$

où le paramètre d'adaptation $\mathbf{e}(t)$ décroît vers 0 en satisfaisant la condition de Robbins-Monro ($\sum \mathbf{e}(t) = \infty$ et $\sum \mathbf{e}^2(t) < \infty$).

On peut ajouter une autre version stochastique des centres mobiles appelée K-means introduite par Mac Queen (1967) et analogue à celle de Forgy. On montre que ces trois algorithmes convergent vers un minimum local de la somme des carrés des écarts aux centres de classes. Celui de Lloyd est très proche de l'algorithme de Kohonen présenté au paragraphe 2.3.

2^{ème} type de classification : parmi les classifications qui fournissent un regroupement pour chaque niveau, on utilise en particulier les classifications hiérarchiques, avec la distance de WARD :

Le principe de la classification ascendante hiérarchique consiste à créer, à chaque étape, une nouvelle partition de l'ensemble des données en agrégeant les deux éléments les plus proches selon une distance choisie à priori.

- *Principe de l'algorithme (commun à toutes les distances) :*

L'algorithme consiste à calculer à chaque étape s la matrice des distances entre les n_s représentants des classes¹ et à déterminer les deux classes dont les représentants sont les plus proches afin de les agréger et d'obtenir $(n_s - 1)$ classes à l'étape suivante. A l'étape 0, il y a autant de classes que d'individus dans la base de données. A l'étape s , on en trouve $n-s$. A la fin, il ne reste plus qu'une classe.

- *Définitions:*

Pour tout k , on note g_k le barycentre de la classe C_k et m_k son effectif et g désigne le barycentre de toutes les données. On appelle somme des carrés des écarts aux centres de classes, notée SC_W (pour somme des carrés intra ou within), la quantité.

$$SC_W = \sum_{k=1}^s \sum_{i \in C_k} \|x_i - g_k\|^2$$

La somme des carrés des écarts entre les centres de classes et le centre de gravité est notée SC_B (pour somme des carrés inter ou between) et vaut

$$SC_B = \sum_{k=1}^s m_k \|g_k - g\|^2$$

La somme des carrés des écarts au centre de gravité, notée SC_T (pour somme des carrés totale) s'écrit

$$SC_T = \sum_{k=1}^s \sum_{i \in C_k} \|x_i - g\|^2$$

La relation de Huygens induit la propriété suivante :

$$SC_T = SC_B + SC_W$$

- *La classification hiérarchique qui utilise la distance de WARD :*

La stratégie consiste, ici, à choisir à chaque étape l'agrégation entre deux classes qui fait le moins varier la somme des carrés des écarts aux centres de classes. C'est-à-dire qu'on cherche à obtenir à chaque pas un minimum local de la quantité SC_W ou de façon équivalente un maximum de SC_B . Considérons le passage de la partition P_s à s classes à P_{s-1} formée de $s-1$ classes par l'agrégation des 2 classes $C_{k_1}(g_{k_1}, m_{k_1})$ et $C_{k_2}(g_{k_2}, m_{k_2})$ qui donne la classe C . On peut calculer son effectif m_t et son barycentre g_t .

$$\text{On a } m_t = m_{k_1} + m_{k_2}$$

$$\text{et } g_t = \frac{m_{k_1} g_{k_1} + m_{k_2} g_{k_2}}{m_{k_1} + m_{k_2}}$$

Avant l'agrégation des classes, la somme des carrés des écarts entre les centres de classes et le centre de gravité s'écrit:

$$SC_B(s) = \sum_{k=1}^s m_k \|g_k - g\|^2 = m_{k_1} \|g_{k_1} - g\|^2 + m_{k_2} \|g_{k_2} - g\|^2 + \sum_{\substack{k=1 \\ k \neq k_1 \& k \neq k_2}}^s m_k \|g_k - g\|^2$$

Après l'agrégation des classes, elle s'écrit:

$$SC_B(s-1) = m_t \|g_t - g\|^2 + \sum_{\substack{k=1 \\ k \neq k_1 \& k \neq k_2}}^s m_k \|g_k - g\|^2$$

La diminution de la quantité SC_B , notée D_s , vaut:

$$\Delta_s = m_{k_1} \|g_{k_1} - g\|^2 + m_{k_2} \|g_{k_2} - g\|^2 - m_t \|g_t - g\|^2$$

Le théorème de Huygens s'écrit :

$$m_{k_1} \|g_{k_1} - g\|^2 + m_{k_2} \|g_{k_2} - g\|^2 = m_{k_1} \|g_{k_1} - g_t\|^2 + m_{k_2} \|g_{k_2} - g_t\|^2 + m_t \|g_t - g\|^2$$

et la diminution de la quantité SC_B , D_s due au passage de s à $s-1$ classes devient :

$$\Delta_s = m_{k_1} \|g_{k_1} - g_t\|^2 + m_{k_2} \|g_{k_2} - g_t\|^2.$$

$$\text{De plus } g_t = \frac{m_{k_1} g_{k_1} + m_{k_2} g_{k_2}}{m_{k_1} + m_{k_2}}$$

$$\text{d'où } \Delta_s = \frac{m_{k_1} m_{k_2}}{m_{k_1} + m_{k_2}} \|g_{k_1} - g_{k_2}\|^2 = \frac{m_{k_1} m_{k_2}}{m_{k_1} + m_{k_2}} d^2(x_{k_1}, x_{k_2})$$

¹ Les barycentres

A chaque étape, on cherche les classes C_{k_1} et C_{k_2} qui correspondent au plus petit D_s . D_s est appelé *indice de niveau* et on vérifie que la somme des indices de niveaux est l'inertie totale des n éléments.

- *Résultat:*

Les classifications hiérarchiques déterminent une notion de paternité qui se traduit par un dendrogramme indiquant pour chaque niveau, le regroupement effectué. Mais il ne fournit pas d'organisation des classes qui permette d'associer facilement à l'une d'entre elles celles qui lui sont proches. En effet, les figures 7 et 8 montrent deux représentations possibles de la même classification hiérarchique. Il n'est donc pas possible, à partir de l'arbre, de donner les classes les plus proches de la classe 2 autres que la classe 1 issue du même père. Il faudrait une notion de voisinage entre les classes pour pouvoir les disposer sur le graphique selon un ordre qui reflète leur proximité. C'est précisément cette notion – centrale dans l'algorithme de Kohonen – qui va fournir des représentations graphiques qui identifient les classes voisines dans le cas de la classification présentée au chapitre suivant.

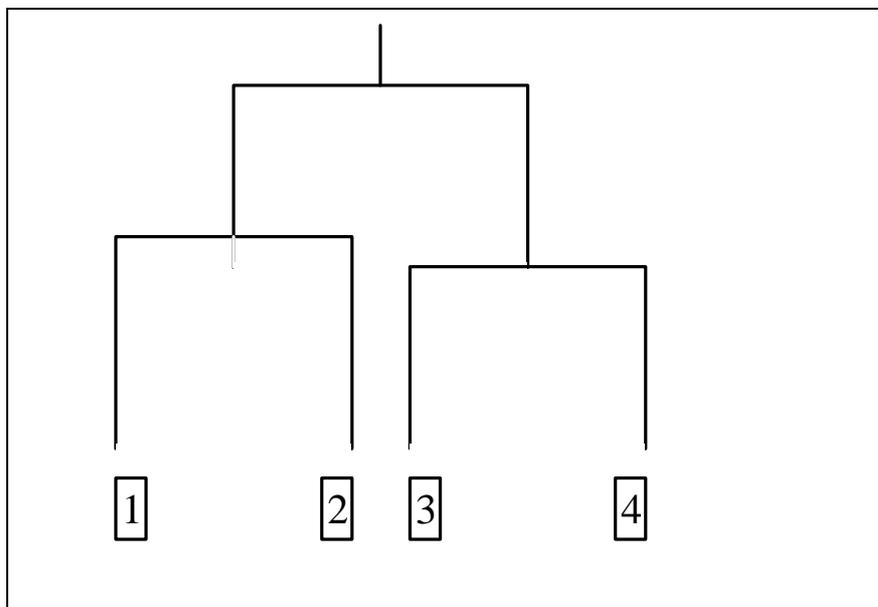


Figure 7 : Dendrogramme 1, la classe2 est représentée entre la classe1 et de la classe3

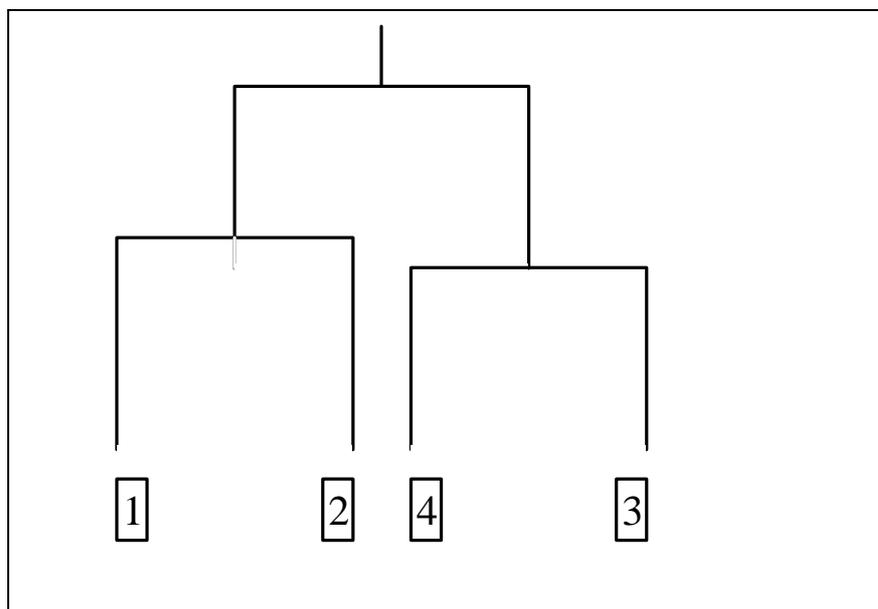


Figure 8: Dendrogramme 2, la classe2 est représentée entre la classe1 et de la classe4

2.3 Algorithme de Kohonen

L'algorithme de Kohonen est un processus d'auto-organisation très puissant. Analogue à l'algorithme des Lloyd, il contient en plus une notion de voisinage entre les U classes – que l'on peut symboliser par des liens entre U unités disposées en réseau. Dans la plupart des cas, cette structure est de dimension une ou deux², on parle alors respectivement de ficelle ou de grille. Ce type de représentation – illustré par les figures 9 et 10 dans lesquelles les unités proches correspondent à des classes voisines – définit naturellement le voisinage des classes.

L'algorithme permet ainsi d'effectuer une classification selon une distance quelconque – dans les exemples, on utilisera la distance euclidienne et celle du χ^2 – et de la traduire par une

carte où l'on peut repérer très facilement les classes voisines. C'est cette propriété qui rend attractif cet outil – par la suite détaillé et utilisé dans les applications. On trouve d'autres algorithmes de ce type réunis sous la dénomination *self-organizing map* (SOM), on peut citer par exemple celui proposé par Bishop et al. (1997) [9]

- **Structures et définitions :**

a) Matérialisation du Voisinage :

On utilisera deux façons de repérer une unité sur une grille. La première consiste à numéroter les unités de 1 à U comme sur les figures 9 et 10. Dans la seconde, on affecte à l'unité u ses coordonnées cartésiennes sur la carte (i_u, j_u) comme indiqué sur la figure 9. Pour la ficelle, ces deux notations sont identiques.

On définit *le voisinage* de rayon r d'une unité u_0 , noté $V_r(u_0)$, comme l'ensemble des unités u situées sur le réseau à une distance inférieure ou égale à r . En utilisant les coordonnées cartésiennes, on peut définir la distance d par:

$$d(u, u_0) = \max(|i_u - i_{u_0}|, |j_u - j_{u_0}|) \text{ pour une grille}$$

$$d(u, u_0) = \max(|i_u - i_{u_0}|) = \max(|u - u_0|) \text{ pour une ficelle}$$

$$\text{Dans les deux cas, } V_r(u_0) = \{u \in \{1, \dots, U\} / d(u, u_0) \leq r\}$$

Exemples:

- Pour une grille :

Dans le cas d'une grille 10x10, les unités qui constituent le voisinage de rayon 2 de l'unité 34 – repérée par le bipoint (4,4) – vérifient la relation:

$$\max(|i_u - i_{34}|, |j_u - j_{34}|) = \max(|i_u - 4|, |j_u - 4|) \leq 2$$

La figure 9 représente une telle grille où l'unité 34 correspond à la case rouge et ses voisines aux cases bleues, c'est-à-dire les unités 12, 13, 14, 15, 16, 21 à 26, 31 à 36, 41 à 46, 51 à 56.

² La dimension 3 ne fournit pas des cartes suffisamment lisibles, mais elle est techniquement facile à réaliser

1 (1,1)	2 (1,2)	3							10
11 (2,1)	12				16				20
			34 (4,4)						
	52				56				
91	92								100

Figure 9 : Voisinage de rayon 2 pour une grille

- Pour une ficelle :

Le voisinage $V_1(5)$ de l'unité 5 – l'unité rouge – de l'exemple de la figure 10 est constitué des unités 4, 5, 6 – les unités bleues et rouges.

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

Figure 10 : Voisinage de rayon 1 pour une ficelle

b) Fonction de voisinage

Le rayon peut varier au cours du temps on note alors $V_{r(t)}(u)$ le voisinage de l'unité u à l'instant t , on a : $V_{r(t)}(u_0) = \{u \in \{1, \dots, U\} / d(u, u_0) \leq r(t)\}$

Exemple de fonction de rayon :

A titre d'exemple, on propose la règle suivante en pourcentage du nombre d'itérations de l'algorithme T résumée par le graphique de la figure 11 :

$$\text{Pour } t \leq \frac{25}{100} T, \quad r=2$$

$$\text{Pour } \frac{25}{100} T < t \leq \frac{75}{100} T, \quad r=1$$

$$\text{Pour } \frac{75}{100} T < t \quad r=0$$

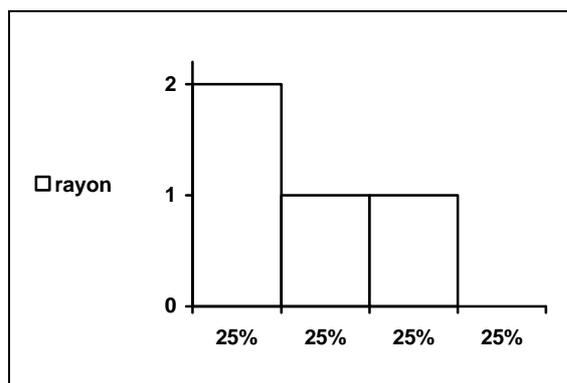


Figure 11 : Répartition en pourcentage du nombre d'itérations utilisées pour chaque rayon

- L'algorithme :

Les observations sont mesurées par p variables quantitatives et forment un nuage de points de \mathbf{R}^p . A chaque itération t , l'unité u est caractérisée par un vecteur code $G_u(t)$ également de dimension p . L'apprentissage de l'algorithme se fait comme suit.

- L'initialisation des vecteurs codes dans l'espace \mathbf{R}^p .

On initialise les vecteurs codes par un tirage aléatoire dans l'enveloppe convexe des observations ou dans l'ensemble des observations ou encore dans le premier plan fourni par une préalable analyse en composantes principales. Nous utilisons le plus souvent la deuxième solution.

- A chaque étape t ,

- On présente l'observation $x=x(t+1)$ extraite au hasard de la base de données.
- On détermine l'unité gagnante u_0 comme celle dont le vecteur code G_{u_0} est le plus près

de x au sens de la distance choisie dans \mathbf{R}^p , c'est-à-dire, celui qui réalise :

$$\|x - G_{u_0}\| = \min_u \|x - G_u\|$$

- On détermine les voisins de u_0 suivant la règle de voisinage à l'instant t .
- On modifie les vecteurs codes de u_0 et des unités voisines de u_0 par les transformations suivantes :

$$G_u(t+1) = G_u(t) + \mathbf{h}(t)(x - G_u(t)), \quad \text{" } u \hat{\mathbf{I}} V_{r(t)}(u_0) \text{"}^3$$

$$G_u(t+1) = G_u(t), \text{ pour les autres unités}$$

où $\mathbf{h}(t)$ est un paramètre d'adaptation qui décroît vers 0 selon les critères de Robbins-Monro ($\sum \mathbf{h}(t) = \infty$ et $\sum \mathbf{h}^2(t) < \infty$).

Le vecteur code $G_u(t)$ est modifié pour devenir $G_u(t+1)$ qui est le barycentre de x et de $G_u(t)$ de coefficients respectifs $\mathbf{h}(t)$ et $1-\mathbf{h}(t)$. Cette loi renforce donc d'une part la similarité entre la nouvelle entrée x et les vecteurs codes de l'unité u_0 et de ses voisines et d'autre part la similarité entre les vecteurs codes des unités de $V_{r(t)}(u)$.

Après l'apprentissage de la carte de Kohonen, chaque unité u est représentée par son vecteur code G_u .

La classification de Kohonen de la base de données consiste alors à associer à chacun de ses vecteurs la classe définie par le numéro de l'unité qui gagne lorsque l'on présente ce vecteur.

L'algorithme de Kohonen nous permet de classer les observations dans U classes, en ajoutant à la propriété usuelle *les classes regroupent des observations semblables* la nuance suivante : *2 vecteurs similaires appartiennent à la même classe ou éventuellement à des classes voisines*. Le grand intérêt d'une telle classification est que le voisinage entre les classes est porteur d'informations, cette propriété sera exploitée dans les applications suivantes.

Un point sur les avancées théoriques concernant l'algorithme de Kohonen.

Si la simplicité de l'algorithme se traduit par une facilité à le programmer, il offre encore une résistance aux mathématiciens qui cherchent à établir ses propriétés et sa convergence vers un état stable. Le seul cas où l'étude de ces propriétés est achevée est le cas où la base de données est de dimension 1 et la structure de la carte une ficelle, un bilan complet de l'état des recherches est par exemple présenté dans Cottrell et al. (1998) [30].

La première démonstration, dans le cas où la base de données suit une distribution uniforme et la fonction de voisinage affecte le même poids aux deux voisins de l'unité

³ C'est-à-dire les vecteurs codes des unités voisines de u_0

gagnante (Cottrell et Fort [22]), a été généralisée d'une part à une grande classe de distributions (Bouton et Pagès [12, 13]) et à une structure de voisinage plus générale par Erwin et Al. [37, 38]. L'auto-organisation a fait l'objet d'une récente étude de Sadeghi [88, 89]. Finalement Fort et Pagès [44, 45] en collaboration avec Benaïm [8] ont achevé la démonstration de la convergence presque sûre vers un état unique, après auto-organisation, pour un ensemble de fonctions de voisinage très large et une grande classe de distributions des entrées.

Dans le cas multidimensionnel, les résultats sont moins avancés. On peut citer les travaux de Fort et Pagès [44, 46] et Flanagan [41, 42] qui aboutissent pour le moment à des résultats incomplets.

Illustration pas à pas de l'apprentissage :

L'exemple 1 – cas de 300 formes simulées comme il l'est indiqué au paragraphe 1.2 – va servir de support à l'illustration de l'apprentissage réalisée en représentant l'état de la carte de Kohonen à différentes étapes de l'algorithme.

L'initialisation :

A chacune des 100 unités, on affecte un vecteur tiré au hasard dans la base de données. La carte de la figure 12 représente ces unités dans lesquelles sont tracées les formes de leurs représentants. A ce stade, l'ordre des unités ne correspond pas à une organisation des formes.

Présentation du premier élément :

Le premier individu présenté est tiré au hasard dans la base de données, il est représenté dans la figure 13. Le vecteur code le plus proche est celui de l'unité 41 (l'unité gagnante est celle qui est repérée, dans la figure 14 par un fond bleu).

Première modification au rayon 2 :

Les unités modifiées sont la 41 et ses 14 voisines elles sont repérées par un fond rouge sur la carte de la figure 14 (pour le rayon 2, ce sont généralement 25 vecteurs codes qui sont transformés mais l'unité 41 est sur un bord). Dans ces unités sont représentées les formes des

vecteurs codes avant la modification et après celle-ci. L'espace blanc entre ces deux courbes permet de visualiser l'ampleur de cette modification.

Fin de l'apprentissage au rayon 2 :

L'état de la carte à la fin de l'apprentissage au rayon 2 est représenté par la carte de la figure 15. Dès ce stade, l'organisation des unités sur la carte correspond à un ordonnancement des formes. Les formes des vecteurs codes des unités voisines sont semblables.

Première modification au rayon 1 :

L'apprentissage au rayon 1 sert à affiner l'organisation déjà en place. Lors de la présentation de l'individu i de la base de donnée (figure 16), l'unité 33 est la gagnante et les 9 unités 22,23,24,32,33,34,42,43,44 sont modifiées (figure 17). On constate que l'ampleur de la modification est, à ce stade, plus faible qu'au début de l'apprentissage puisque l'espace blanc entre les vecteurs codes avant et après leur modification est devenu trop fin pour être visible.

Fin de l'apprentissage au rayon 1 :

La carte de la figure 18 montre que la carte a gardé la structure de la fin de l'apprentissage au rayon 2. Elle s'est en fait affinée.

Apprentissage au rayon 0 :

Pour ce rayon, seule l'unité gagnante est modifiée, ici l'unité 79 (figure 21), la modification à chaque pas est très faible. Cet apprentissage sert à rapprocher les vecteurs codes des barycentres des classes.

Fin de l'apprentissage :

Dans l'état final, les formes sont ordonnées sur la carte de la figure 21.

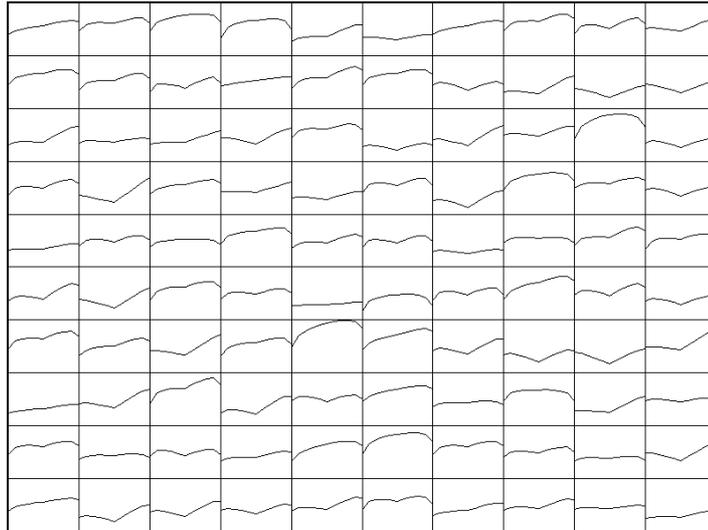


Figure 12 : L'initialisation

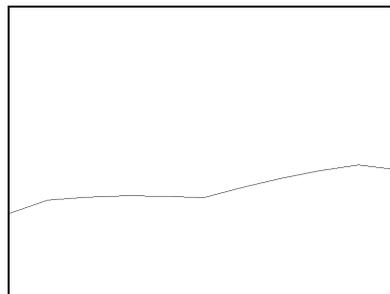


Figure 13 : Premier individu présenté

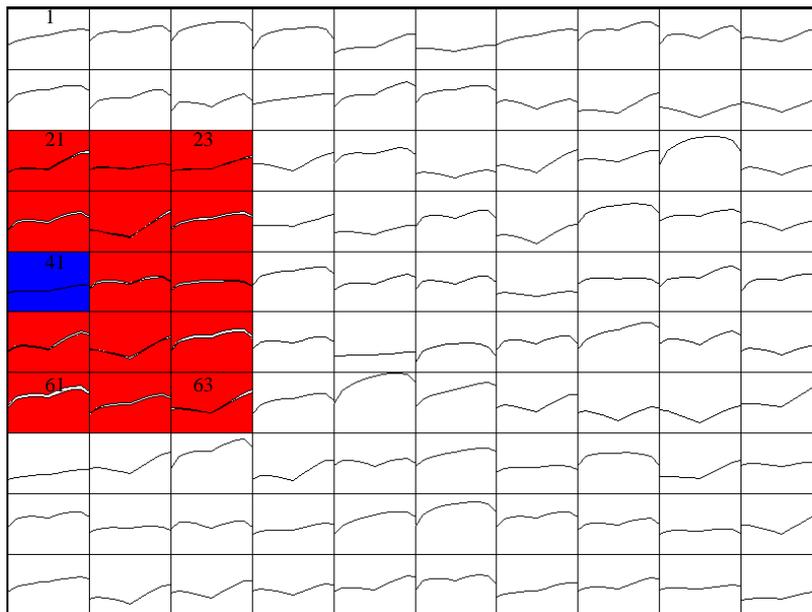


Figure 14 : Première modification des vecteurs codes au rayon2

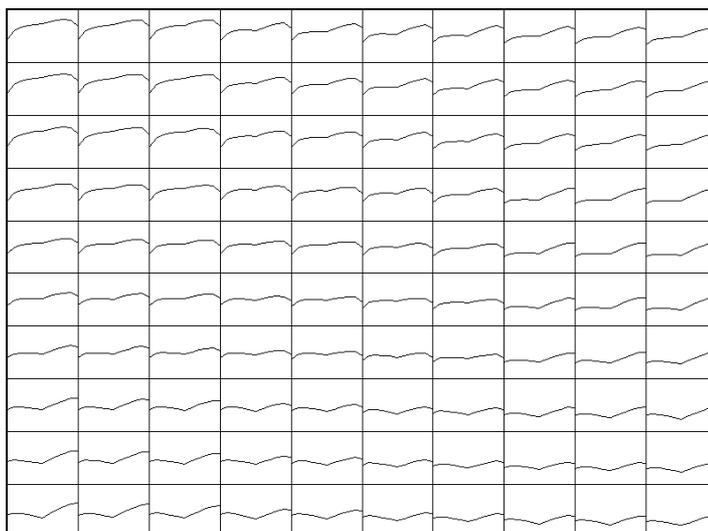


Figure 15 : Etat de la carte à la fin de l'apprentissage au rayon 2

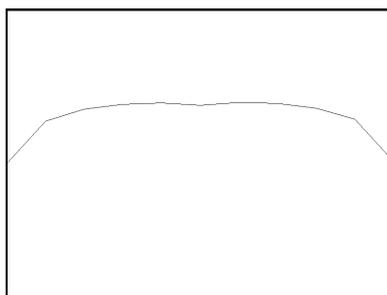


Figure 16 : Premier individu présenté pour l'apprentissage au rayon 1

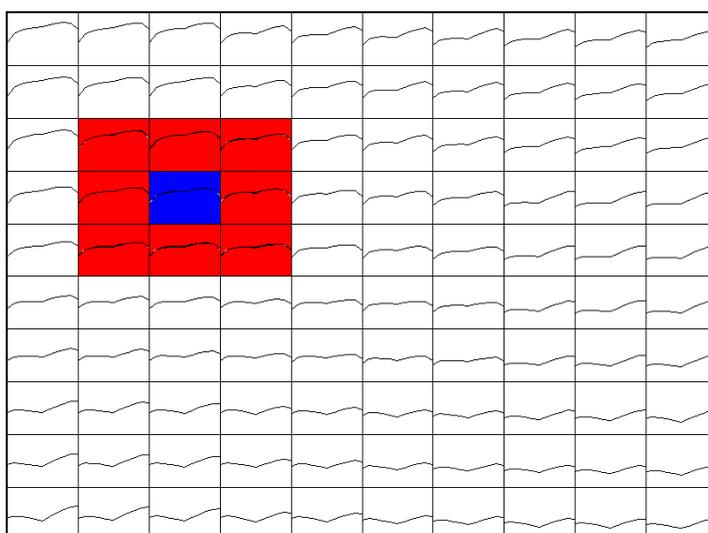


Figure 17 : Première modification des vecteurs codes au rayon 1

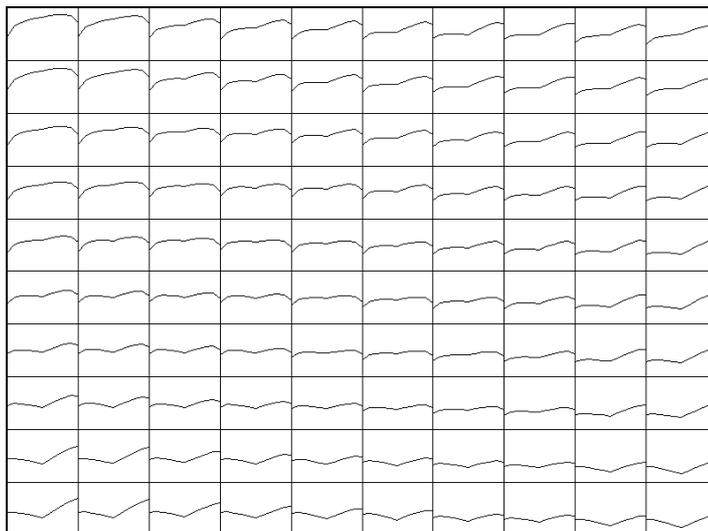


Figure 18 : Etat de la carte à la fin de l'apprentissage au rayon 1

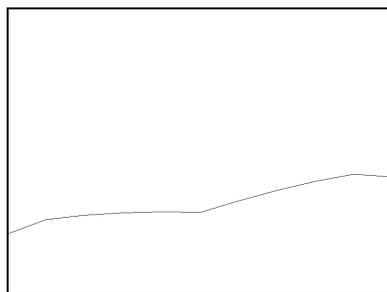


Figure 19 : Premier individu présenté pour l'apprentissage au rayon 0

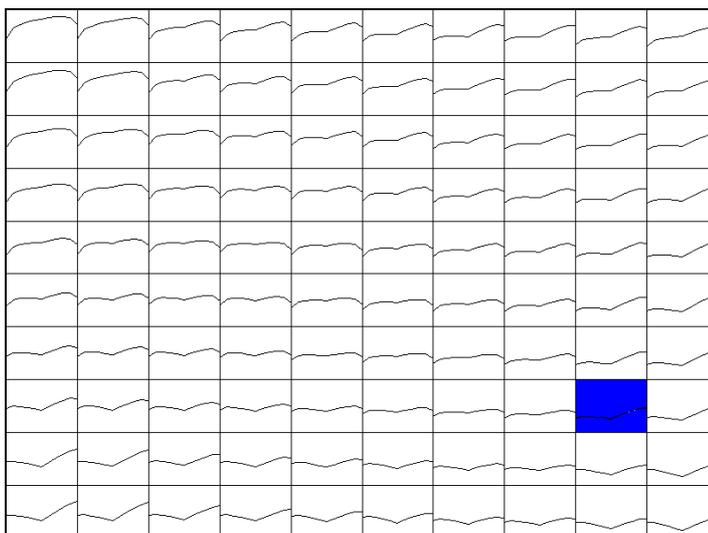


Figure 20 : Première modification des vecteurs codes au rayon 0

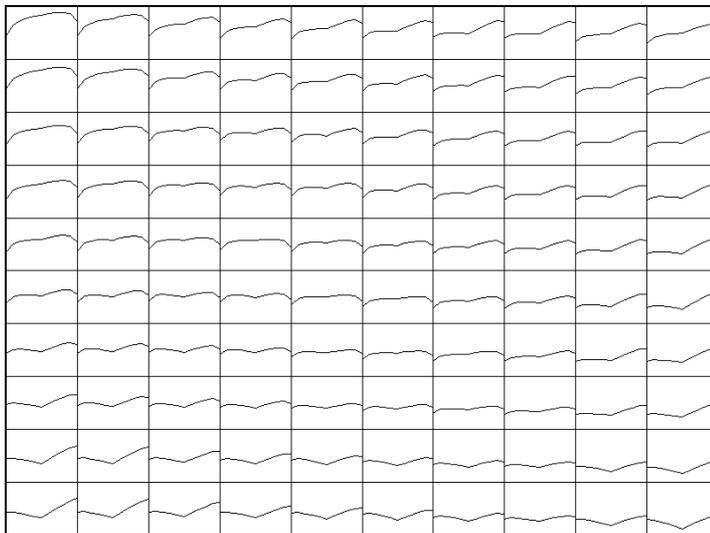


Figure 21 : Etat final de la carte

Variantes de l'algorithme de Kohonen :

Il arrive que le nuage de points possède une caractéristique particulière et on souhaite, bien entendu, que la classification la traduise soit par la structure de la carte, soit par les représentants des classes. Dans l'exemple 2, la consommation des canadiens, est exprimée en pourcentages, et les individus vérifient donc la contrainte $\forall i, \sum_{j=1}^p x_i^j = 1$. Les vecteurs codes résultants de la classification de Kohonen étant des barycentres de ces individus, ils conservent cette relation – on a en effet $\forall t, \forall u, \sum_{j=1}^p G_u^j(t) = 1$. Et pour la même raison, toute propriété linéaire de la sorte sera transmise aux centres de classes. En revanche si les composantes des x_i suivent une relation non linéaire, il n'y a aucune raison qu'elle se conserve. Par exemple, dans l'étude de la consommation d'électricité présentée au chapitre 6, les individus – qui sont des profils – satisfont à la propriété suivante: $\forall i, \sum_{j=1}^p x_i^{j^2} = 1$. Or cette propriété ne se conserve pas si on n'aménage pas l'algorithme.

Il est en effet possible d'utiliser des variantes de l'algorithme mieux adaptées aux particularités des données. Dans l'étude de la consommation d'électricité, chapitre 6, on a

défini une variante permettant de conserver la propriété commune aux individus ($\forall i, \sum_{j=1}^p x_j^2 = 1$). On y présente également un exemple de modification de la structure de la carte, le rectangle y a été remplacé par un cylindre, qui permet de prendre en compte la saisonnalité interne à la série temporelle. Ces variantes sont présentées au chapitre 2.

2.4 La relation entre la méthode basée sur la classification de Kohonen et les autres méthodes.

La classification de Kohonen par rapport aux autres méthodes :

Si la notion de voisinage lui est particulière, la classification de Kohonen a des analogies avec certaines méthodes présentées au chapitre 2.2 et des propriétés qui permettent d'envisager d'éventuels couplages avec elles. Parmi elles, l'algorithme de Lloyd (ou le Simple Competitive Learning) est exactement celui de Kohonen à 0 voisin. Ces deux techniques de classification sont **mieux adaptées aux grandes bases de données** que les algorithmes comme celui de Forgy – qui obligent à classer tous les individus à chaque itération – ou celui de la classification hiérarchique avec la distance de Ward – dans lequel il faut calculer toutes les distances entre les centres de classes – même s'il existe dans chaque cas des aménagements qui réduisent les coûts informatiques.

La classification de Kohonen est **robuste** – au sens où le résultat ne peut être grandement modifié par l'ajout d'un nouvel élément à la base de donnée si celui-ci n'est pas trop extravagant (valeur erronée ou aberrante). Cette propriété est aussi vérifiée par la méthode des centres mobiles mais n'est pas partagée par la classification ascendante hiérarchique dont le résultat peut être remis en cause par l'apport d'un individu

supplémentaire. Par contre, cette dernière est la seule à fournir exactement le même résultat quand on relance l'algorithme car les autres – qui aboutissent à un minimum local de la somme des carrés des écarts aux centres de classes – dépendent de l'ordre de présentation des individus et de l'initialisation.

Ces méthodes peuvent être complémentaires et donner naissance à des **combinaisons hybrides** du type *centres mobiles – classification hiérarchique* (dont on peut trouver une présentation dans Wong (1982) [104]), *carte de Kohonen – Lloyd* (expliquée dans Debodt, Cottrell, Verleysen (1999) [33]) ou *carte de Kohonen – classification hiérarchique*.

Il n'y a pas plus de règles pour déterminer le nombre de classes U dans les méthodes de type Kohonen que dans les classifications classiques. Seuls les critères habituels tels que les tests de Fisher et Wilks peuvent être utilisés. De plus, augmenter le nombre de classes signifie passer de U classes à $U+1$ dans le cas d'une ficelle et n^2 classes à $(n+1)^2$ dans le cas d'une grille ($U=n \times n$). Dans le cas de la classification de type Ward, on préfère U classes plutôt que $U+1$ si le saut d'inertie entre ces deux regroupements est faible (par rapport au saut entre les niveaux $U-1$ et U). Mais cet outil ne permet ni de choisir le meilleur saut d'inertie ni celui du meilleur niveau de classification.

3 Analyse de données en utilisant les cartes de Kohonen

L'algorithme de Kohonen – qui a été présenté et comparé à d'autres méthodes de classification au paragraphe 2 – va maintenant être exploité dans le cadre de l'analyse de données. On peut se demander, avant de développer ses propriétés, en quoi ses caractéristiques permettent d'envisager qu'il soit mieux adapté à certaines données ou à certaines analyses que les méthodes classiques. On a déjà signalé précédemment (voir au paragraphe 1) que les projections sur des plans produits par les analyses factorielles ne sont pas adaptées à des bases de données qui nécessitent une représentation dans des espaces de dimension trop grande (supérieure à 4). Cela pose des problèmes dans la pratique de visualisation et de synthèse. De plus, on souhaite utiliser des bases de données définies éventuellement par un plus grand nombre de variables. Aujourd'hui, on espère améliorer les méthodes existantes en les complétant.

Il est en effet frustrant qu'une base de données plus grande ne fournisse pas une information exploitable significativement plus importante. La raison en est que les variables ajoutées sont secondaires en pouvoir explicatif des individus et qu'elles les influencent sous deux formes : le cumul et la spécificité. Le cumul des variables secondaires – qui peut prendre une importance significative par rapport aux variables primaires – est rarement linéaire et l'addition de nombreux facteurs ne constitue pas un facteur facile à interpréter. D'autre part, les variables secondaires ont généralement un effet, non pas sur l'ensemble des individus mais sur un groupe, c'est un effet de spécificité que traduit mal la linéarité. Dans *les couplages entre analyses factorielles et classifications* – présentés dans la suite et que l'on peut trouver dans des manuels tels que celui de Lebart, Morineau, Piron (1995) [68] ou dans Wong (1982) [104] – cet effet peut être pris en compte par la classification, mais perdu par le système de projection inhérent à l'analyse factorielle.

Ceci vient d'ailleurs d'un problème plus général, *le couplage analyse factorielle – classification n'optimise pas la représentation de la classification* et inversement *la classification n'optimise pas l'explication de la représentation* pour des raisons développées dans la section suivante. Or, c'est précisément ce que fait l'algorithme de Kohonen – où la représentation et la classification sont jumelées – et c'est pourquoi il peut prendre en compte une augmentation de l'information ignorée de l'analyse classique et apporter ainsi un plus par rapport aux autres méthodes.

L'autre apport est celui d'une méthode offrant à la fois une grande facilité d'interprétation, une grande souplesse d'utilisation – on peut choisir la distance et donc l'adapter à un cadre d'étude assez large – et permettant des représentations graphiques de nombreuses sortes de bases de données en respectant leur topologie. Ces représentations peuvent rendre cette méthode aussi attractive et souvent mieux adaptée que les méthodes linéaires dont un exemple simple en trois dimensions sera étudié dans ce chapitre. Dans la suite, seront développés des outils pour exploiter au mieux les possibilités de cet algorithme – en particulier, la représentation de la dispersion et de la discrimination des classes, l'analyse des variables quantitatives au moyen de variables qualitatives sous forme de cartographies, l'analyse locale.

3.1 Analyse de données appliquées en pratique.

Dans cette section sont présentés d'une part le couplage analyse factorielle – classification dont les représentations sont dans le même esprit que celles des cartes de Kohonen telles qu'elles seront développées dans la suite, et d'autre part l'analyse de données avec le perceptron multicouches qui constitue une approche par les réseaux de neurones. Par la suite, on va montrer comment l'algorithme de Kohonen répond à la fois au souci de produire des sorties graphiques explicites de la première technique et d'échapper à la contrainte linéaire de la seconde. D'autres études ont fait l'objet de comparaisons entre les possibilités qu'offrent respectivement l'algorithme de Kohonen et les méthodes classiques (comme par exemple l'analyse en correspondance [25 et 26]).

3.1.1 Analyse linéaire des données.

Parmi les méthodes linéaires, l'analyse factorielle est très souvent utilisée pour explorer de grands tableaux de données. Elle consiste à rechercher la meilleure structure euclidienne approchant un nuage de points, d'abord un plan, puis un espace à 3 dimensions, etc... Ces méthodes sont très appréciées, car elles produisent des représentations graphiques par projection de la base de données.

En effet, elles fournissent les axes principaux d'inertie dans l'ordre de leur importance. Ceci permet de choisir le plan sur lequel projeter pour voir au mieux le nuage de points – c'est-à-dire celui qui contient le maximum d'inertie, appelé plan principal et obtenu à partir des deux premiers axes. La contribution du troisième axe permet de construire l'espace à trois

dimensions qui épouse le mieux le nuage de points, le quatrième axe complète ce dernier pour obtenir l'espace de dimension quatre etc..

Néanmoins, l'espace de dimension 3 se déduisant des 2 premiers plans est difficile à imaginer sans expérience et cela devient rapidement impossible dans les dimensions supérieures. C'est ce qui explique que l'on ait des difficultés pour analyser les bases de données qui ont une structure de dimension intrinsèque assez grande (supérieure à 4). Parmi les autres inconvénients, on peut indiquer que les axes principaux sont sensibles aux points extrêmes ou aberrants ce qui entraîne des problèmes de robustesse. Et il est nécessaire de contrôler cette influence pour avoir une représentation graphique significative de la base de données. De plus, la projection sur un plan d'un nombre de points très important ($\gg 100$) peut charger le graphique au point de le rendre illisible. Enfin, elle peut entraîner des distorsions fâcheuses et superposer des points qui peuvent être éloignés dans l'espace. Pour pallier ces lacunes, on trouve dans l'analyse classique le couplage *classification – analyse factorielle* qui apporte une réelle amélioration. On verra, par la suite que l'algorithme de Kohonen qui conserve l'esprit de ce couplage va plus loin dans les liens entre classification et représentation.

Couplage de l'analyse factorielle avec une classification :

Il consiste à effectuer une classification sur l'espace entier, les classes sont ainsi constituées à partir de l'ensemble de l'information et pas seulement à partir du sous espace à 2 ou 3 dimensions qui sert à l'interprétation. La représentation des classes sur les plans augmente l'information et allège les graphiques de l'analyse factorielle. De plus, l'éventuelle possibilité fournie par la classification de déterminer les individus marginaux améliore la robustesse. Néanmoins, même si le plan principal est le plan qui contient le maximum d'inertie, l'analyse linéaire n'optimise pas nécessairement la séparation des classes et de même la classification n'est pas conçue pour expliquer au mieux les projections qui servent de représentation graphique dans l'analyse factorielle. Cette remarque, déjà faite en introduction du paragraphe 3 est développée dans la section 3.7.3 où une base de données est analysée à la fois par cette méthode et par l'algorithme de Kohonen en vue d'une comparaison.

3.1.2 Analyse des données avec les réseaux de neurones.

Comme méthode neuronale d'analyse des données pour remplacer les méthodes linéaires classiques quand celles-ci ne s'appliquent plus, le perceptron multicouches qui admet en entrée à la fois des variables qualitatives et quantitatives peut être une solution. Cependant, comme pour tout apprentissage supervisé, il est nécessaire de connaître les classes des observations utilisées pour l'apprentissage, ce qui est une situation particulière. De plus, même si beaucoup de progrès ont été réalisés depuis l'époque de la « boîte noire » – pour mieux choisir l'architecture du réseau, pour supprimer les connexion non significatives, entre autres – il n'est pas facile d'interpréter le modèle, d'en extraire les variables principales, ni de déduire du résultat une représentation des observations respectant la topologie de l'espace d'entrée.

Ce genre de modèle est, en fait, plus approprié pour la régression, pour la prévision à court terme et pour la classification sans visualisation ni interprétation. La prévision à court terme sera développée dans le chapitre 4.4.2 en même temps qu'une présentation du perceptron multicouches

3.2 Analyse de données à l'aide des cartes de Kohonen

Comme nous l'avons décrit dans la section 2.3, le premier résultat obtenu à la suite de l'apprentissage est une classification des N observations en U classes (une par unité, certaines étant éventuellement vides), représentées par le vecteur code associé et liées entre elles par une notion de voisinage.

Pour exploiter au mieux la carte de Kohonen, il est naturel d'utiliser les cases qui symbolisent les unités pour y représenter certaines caractéristiques des individus. En

particulier, on peut y tracer les représentants des classes. Si on s'appuie sur l'exemple 2 de *la consommation des Canadiens*, après avoir réalisé l'apprentissage de l'algorithme de Kohonen sur cette base de données, on peut représenter la carte de sortie de la figure 22, ici une grille, où dans chaque unité est représenté le vecteur code associé. Pour comprendre la signification de ces courbes, on peut se référer à la section 1.2 qui présente les exemples. On constate que les formes des vecteurs codes de deux unités voisines se ressemblent (relativement aux autres).

De plus, dans certaines unités, le tracé du vecteur code fait apparaître clairement le type de consommation dominant chez les individus de la classe – par exemple la consommation pour le logement domine en effet dans les unités 5, 6, 7, 13, 14 et 15.

Une fois que les classes ont été définies, tous les critères de la statistique classique peuvent être utilisés pour mesurer les variances intra classes et inter classes. En complément, nous proposons de les visualiser par des méthodes graphiques qui superposent ces résultats avec les propriétés topologiques des cartes de Kohonen.

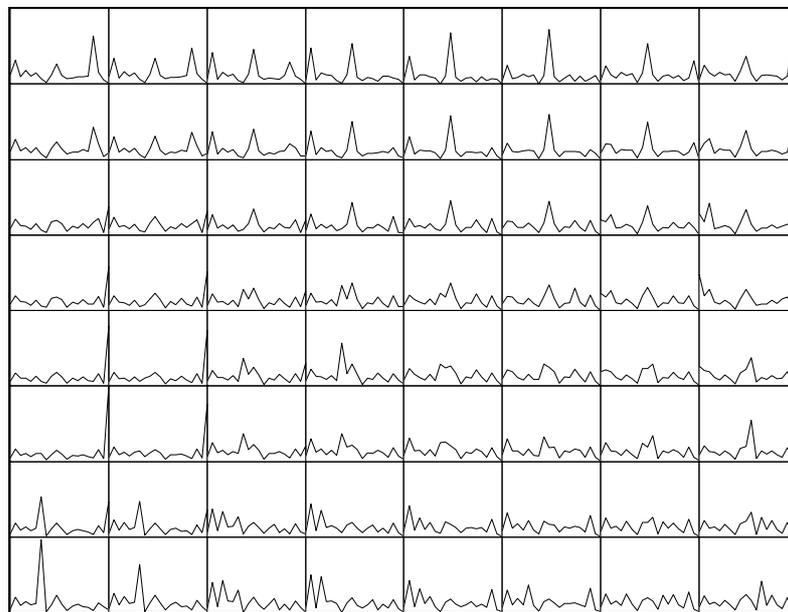


Figure 22 : Les vecteurs codes sont représentés dans leur unité d'affectation

3.3 2 niveaux de classification

Le nombre de vecteurs codes U étant grand (pour une grille, dans nos exemples $U=100$), un niveau de classification plus grossier peut être également révélateur. Un niveau élevé est intéressant, car il fournit une qualité d'analyse plus fine et comprime moins l'information que si l'on résume l'ensemble des individus par les représentants d'un petit nombre de classes. Par contre, une classification qui cumule 2 niveaux de regroupement permet d'avoir l'analyse fine de celle qui a un grand nombre de classes avec le recul et les propriétés de synthèses de celle qui en a peu. Elle est de fait plus efficace qu'une classification à un seul niveau de regroupement ou que deux qui ne s'emboîtent pas.

Pour passer du regroupement en U classes de la carte de Kohonen à un niveau plus grossier, une classification hiérarchique des U vecteurs codes est suffisante. Il est à noter que U est en général très petit devant le nombre de données N et cette deuxième classification a donc un coût de calcul très faible. Une représentation des deux niveaux de classification sur la carte permettra de prendre en compte la notion de voisinage et constituera ainsi un enrichissement par rapport au couplage classique *centres mobiles - classification hiérarchique*.

Dans la suite, on appellera *micro-classes* les U classes de Kohonen et *macro-classes* celles qui résultent de la classification hiérarchique qui les regroupe. Pour visualiser conjointement ces deux classifications, chaque unité de la carte est remplie avec la couleur de sa macro-classe. Quand on applique cette méthode à l'exemple 2, on obtient la figure 23 qui représente ainsi le regroupement des 100 micro classes en 10 macro-classes. L'avantage de la double classification est la possibilité d'analyser au niveau macro d'où émergent les conclusions générales et au niveau micro qui révèle les phénomènes plus précis et plus particuliers. Le niveau micro informe de plus sur la façon dont se fait le passage d'une macro classe à une autre.

Dans les applications où nous avons utilisé cette méthode (certaines sont présentées dans ce document en annexe), les macro classes regroupent des micro classes voisines. Cette remarque est très intéressante, car elle confirme la propriété de respect de la topologie des cartes de Kohonen. Néanmoins, on peut concevoir des cas où une macro classe est constituée de 2 groupes de micro classes. Par exemple des données repliées en fer à cheval peuvent provoquer ce phénomène. La section suivante présente un nouvel outil qui complète l'information des macro classes en aidant à comprendre la structure de la carte elle même.

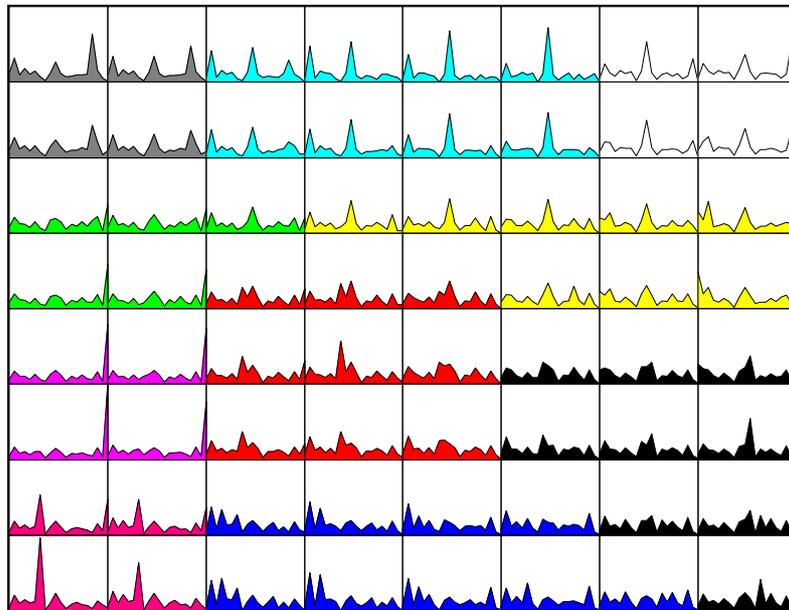


Figure 23 : Au graphique précédent a été rajouté une couleur du remplissage qui correspond à la macro-classe associée à l'unité

3.4 Représentation des distances inter classes

Comme il a déjà été suggéré précédemment, il est très important d'avoir une bonne représentation de la géométrie de la carte. La technique qui va être présentée maintenant a pour but d'augmenter la connaissance de la structure de la carte elle-même en ne s'intéressant qu'aux vecteurs codes.

La sortie de l'algorithme est constituée de U classes liées par une règle de voisinage et représentées par les vecteurs codes. Les distances entre les classes sont irrégulières et les U vecteurs codes engendrent une surface dans l'espace de dimension p qui est le nombre de variables quantitatives. La représentation de la sortie par une grille plane où toutes les unités sont disposées à égales distances omet naturellement ces propriétés. Pour éviter que cela entraîne des confusions lors de l'interprétation, nous proposons de représenter les distances entre les classes à partir de la méthode proposée par Cottrell, De Bodt [27], dont le principe est le

suivant:

La représentation des distances respectives entre les vecteurs codes des unités voisines u et u' se fait à l'aide d'un octogone dont les sommets $S_{u,u'}$ se rétractent du bord vers le centre de l'unité u d'une longueur $l_{u,u'}$ proportionnelle à la distance entre G_u et $G_{u'}$.

$$\text{On a donc } l_{u,u'} = C \times \|G_u - G_{u'}\|_{\mathbb{R}^p}$$

La figure 24 illustre la déformation des bords de l'unité u dont le vecteur code associé est plus proche, dans l'ordre, de celui de l'unité $u-n$, puis $u+1-n$, et enfin, de l'unité $u+1$.

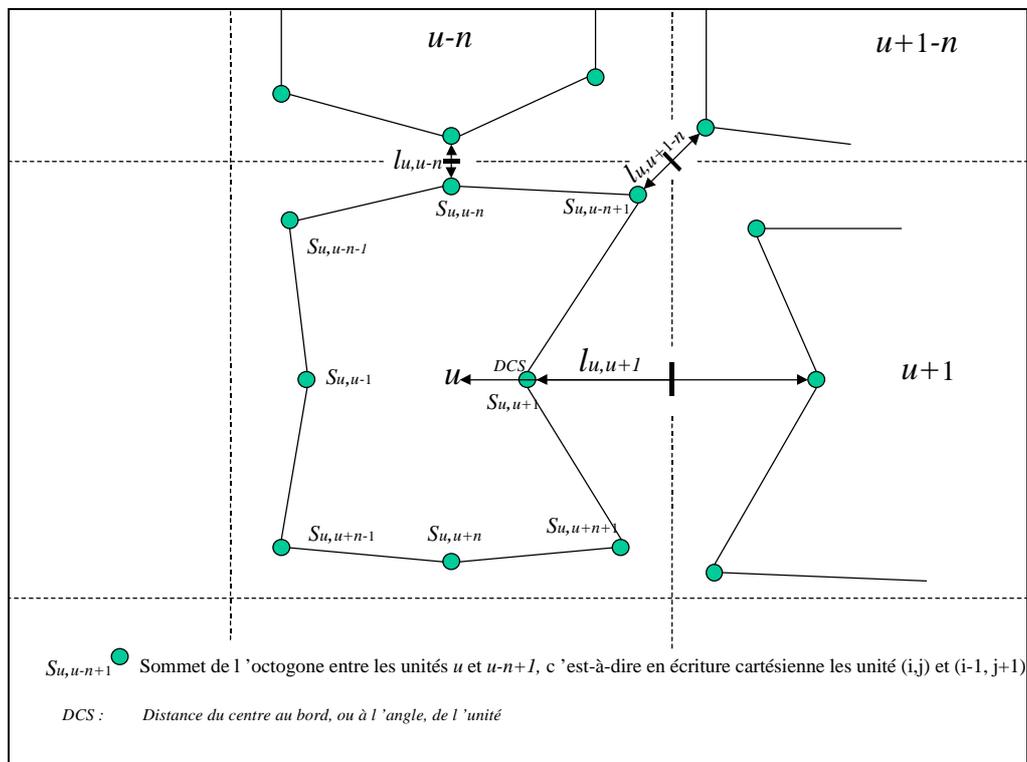


Figure 24 : Chaque sommet de l'octogone associé à l'unité u se rétracte d'une longueur proportionnelle à la distance entre les vecteurs codes des unités qui partagent le bord ou l'angle correspondant

- Choix de la constante C :

On ne peut pas rétracter le sommet d'une longueur supérieure à la distance DCS du centre de l'octogone au sommet : $l_{u,u'} < DCS$ (1)

Considérons M la distance maximum entre 2 vecteurs codes voisins sur toute la carte,

$M = \sup_{u,u'} \{ \|G_u - G_{u'}\|, u' \text{ voisin de } u \}$. Une constante C qui vérifie $C = m \times \frac{DCS}{M}$, avec $0 < m < 1$,

permet de vérifier la relation (1). Dans la pratique, le choix $C = \frac{1}{3} \times \frac{DCS}{M}$ s'est révélé

satisfaisant dans les études pratiques que nous avons accomplies. Il permet en effet de bien visualiser la distance entre les classes et laisse les octogones suffisamment grands pour superposer les 2 classifications du chapitre 3.3 comme on le verra dans la suite.

La carte obtenue est du type de celle de la figure 25 réalisée pour la base de l'exemple 2 *des consommations canadiennes*. On constate que la grande dispersion a lieu dans le quart de carte “bas-gauche” de la carte.

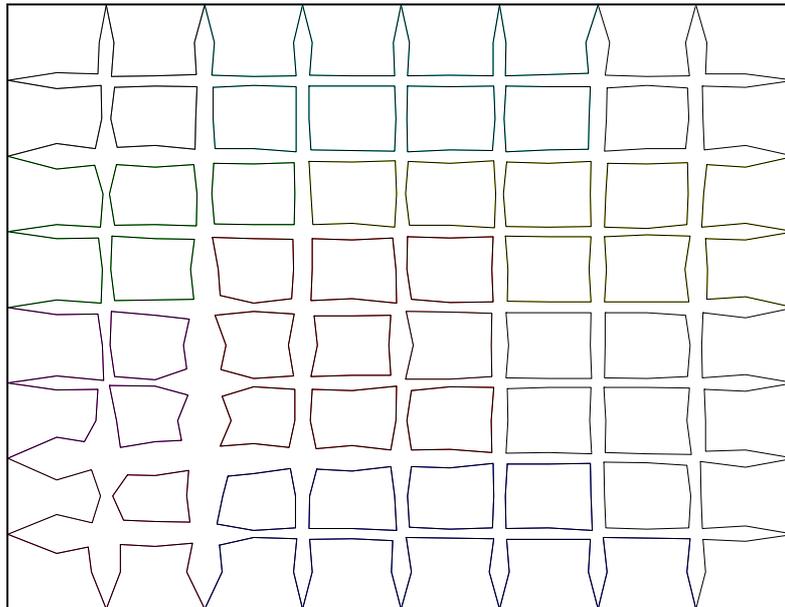


Figure 25 : Carte des distances

Superposition des représentations des distances entre les vecteurs codes et les macro-classes:

Pour superposer la représentation de la dispersion à celle des macro classes, il suffit d’associer à chaque unité u la couleur correspondant à la macro classe de G_u comme précédemment. Il s’ensuit la nouvelle carte de la figure 26 qui suscite quelques remarques.

La superposition des macro-classes avec les distances entre les classes, fait apparaître en général que les frontières des macro-classes coïncident localement avec les plus grandes distances entre les micro-classes. La notion de *local* est ici primordiale, car il est faux de dire que les changements de classes coïncident avec les grandes distances absolues. Cependant, quand on ne considère que les unités de deux macro-classes (ce qu'on appelle local), la séparation de celles-ci se situe à l'endroit des grandes distances entre les micro-classes de la zone. Ceci confirme la pertinence du second niveau de classification. D'autre part, un

changement de macro-classes qui se fait avec une petite distance signifie qu'une population homogène de taille importante a été divisée en deux et que le passage d'un groupe à l'autre se fait avec une continuité.

Enfin, la notion de grande distance est locale. En effet, on trouve à l'intérieur de certaines macro-classes des distances qui sont parmi les plus grandes de la carte. Par contre, dans les macro-classes qui regroupent un grand nombre de micro-classes, les distances sont plus petites. Ce phénomène provient du fait que l'algorithme affecte plus d'unités à une population qui possède un gros effectif. Il convient donc de ne comparer les distances que localement. Pour expliquer ce phénomène, considérons la base de *la consommation des canadiens* qui a fourni la classification de la figure 26. Rajoutons à cette base les individus de la 1ère macro-classe (de sorte qu'ils apparaissent 2 fois dans la base). L'algorithme de Kohonen affecte plus de classes à la première macro-classe et de fait les distances entre ses micro-classes deviennent plus petites.

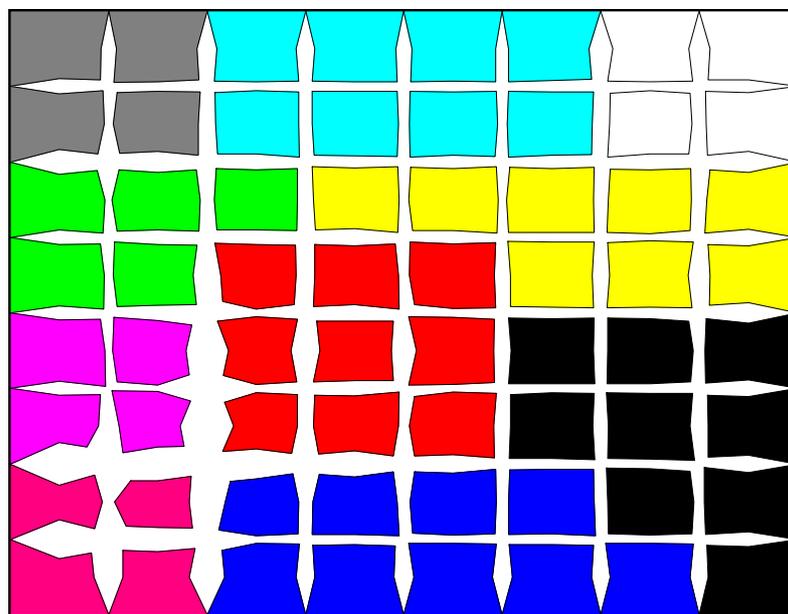


Figure 26 : Les octogones de la carte des distances sont remplis de la couleur qui correspond à la macro-classe de l'unité

3.5 Contenu des classes et homogénéité.

La dispersion intra-classes est également une question qui mérite d'être étudiée. Il est en effet important d'avoir une idée de l'homogénéité des classes. De plus, dans le cadre des cartes de Kohonen, il est très utile de savoir si la qualité de cette variabilité est liée à la topologie. Il est donc naturel de traduire cette dispersion sur une carte. Pour cela, le choix qui a été fait consiste à représenter simultanément chaque individu dans son unité d'affectation comme sur la figure 27 commentée ci-après. Ce n'est donc pas une mesure de la dispersion qui est renseignée sur la carte mais plutôt la qualité de celle-ci. On peut en effet y vérifier si l'homogénéité de la classe est bonne, mais également si on y trouve plusieurs sous-classes ou encore y déceler les valeurs marginales ou aberrantes.

Appliqué à l'exemple 3 des données bi-horaires, ce type de représentation fournit la carte de la figure 27 qui traduit une homogénéité satisfaisante. Les individus de chaque classe ont en effet tendance à se superposer pour créer un profil commun très net, l'épaisseur du trait reflétant la légère variabilité. Dans l'unité 97, on distingue deux individus qui ne sont pas des valeurs aberrantes mais plutôt des valeurs marginales. Dans l'unité 72, on constate que la perte d'homogénéité se situe sur les premières heures. Dans la classification des consommations électriques de la deuxième partie, on trouve des unités où cohabitent deux groupes – les jours fériés et les dimanches – ce qui se traduit par des profils qui se regroupent autour de deux formes voisines mais distinctes. C'est donc bien la qualité de l'homogénéité qui est représentée sur ces cartes, elle sert avant tout de guide pour orienter d'éventuelles études.

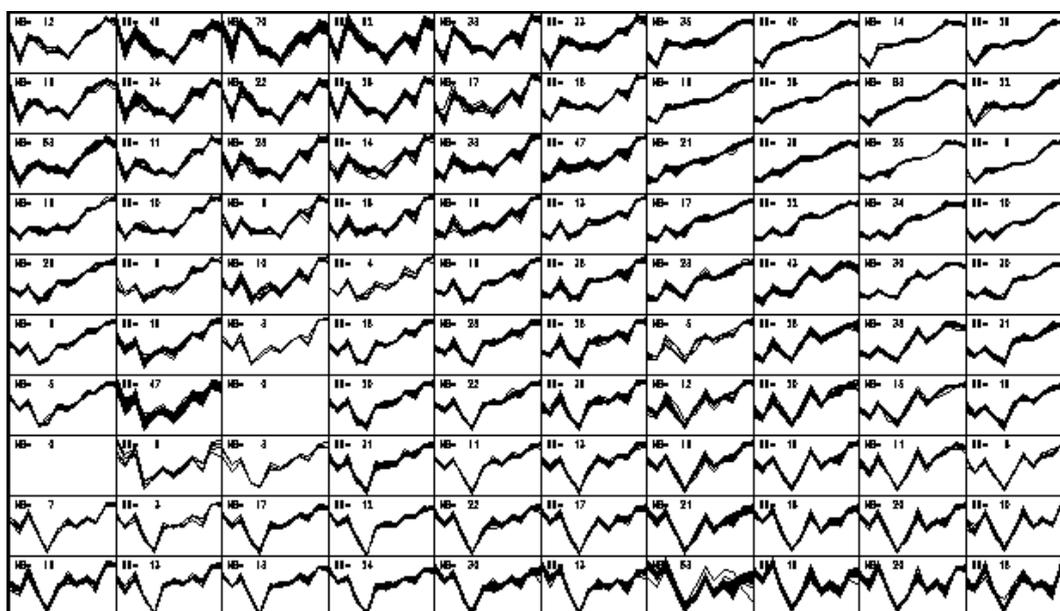


Figure 27 : Tous les individus et l'effectif de chaque classe sont représentés dans leur unité d'affectation

3.6 Croisement de la classification avec une variable qualitative.

On a vu précédemment comment répartir les individus d'une base de données en U classes – qui respectent une relation de voisinage – à partir de p variables quantitatives. On va maintenant chercher à caractériser ces groupes par une variable Q parmi les q variables qualitatives qui n'ont pas participé à la classification. La méthode classique consiste à croiser la variable qui affecte chaque individu à sa classe avec chacune de ces q variables. Elle aboutit à un tableau de contingence de taille U fois le nombre de modalités de la variable. Dans nos exemples, U prend des valeurs importantes – entre 64 et 100 – ce qui rend la représentation en tableau inadaptée. L'exploitation d'un aussi grand nombre de classes ne peut se faire qu'avec le

support des cartes et la notion de voisinage. On procède donc de la même façon que dans les sections précédentes en traçant dans chaque unité u l'information associée – ici la fréquence avec laquelle apparaît chaque modalité dans la population de la classe C_u . On appelle cette représentation la *cartographie* de la variable Q qui se construit comme suit.

Pour chaque unité u , on extrait de la base de données les N_u individus de la classe C_u . On calcul la fréquence $F_{u,k}$ de la modalité k de la variable qualitative Q dans la classe C_u par :

$$F_{u,k} = \sum_{x \in C_u} 1_{\{Q(x)=k\}} / N_u$$

Pour résumer les fréquences $F_{u,k}$ de la classe C_u , on peut représenter un camembert comme sur la figure 28 où la portion de la modalité k vaut $F_{u,k} \times 2\pi$. Chaque modalité k est alors représentée par une couleur de remplissage. La figure 28 représente le cas où les éléments de l'unité prennent les modalités 1, 2 et 3 de Q avec la même fréquence de $1/3$. En dessinant dans chaque unité u le camembert associé à F_u , on obtient la cartographie de la variable Q sur l'ensemble de la base de données comme sur les figures 29, 30, 31 et 32. Elle répond aux questions du type « Qui sont les individus de telle classe? » « Existe-t-il une caractéristique commune aux voisins d'une classe ? » « Pouvons nous qualifier un regroupement de classes par un critère qualitatif ? ».

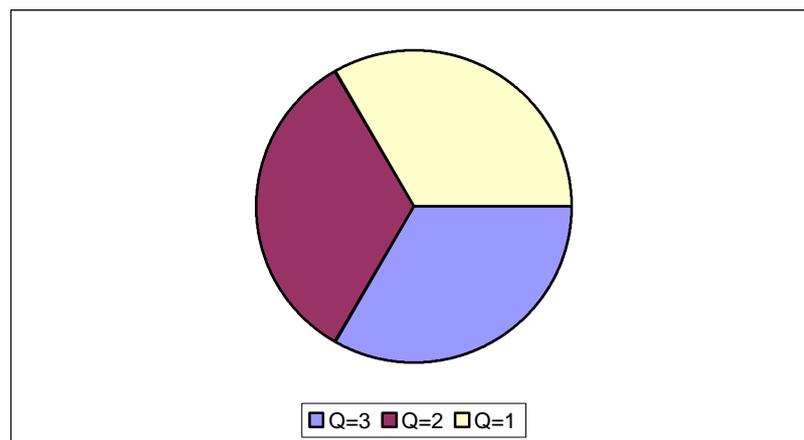


Figure 28 : Camembert qui représente la répartition des modalités 1,2 et 3 de la variable Q

Exemple 2 de la consommation des canadiens:

Dans le cadre de l'exemple 2 de la *consommation des canadiens*, la cartographie du critère *pauvreté- richesse* (figure 30) indique que les unités à forte proportion de riches et quasi- riches sont placés plutôt *aux extrémités de la diagonale angle supérieur droit - angle inférieur gauche*, les *pauvres et quasi - pauvres* occupent, pour leur part, *la zone centrale*.

De son côté, la cartographie des *propriétaires- locataires*, figure 29, oppose sur cette diagonale les *propriétaires* placés autour *de l'angle inférieur gauche* et les *locataires* à l'opposé. La combinaison des deux cartographies fournit ainsi une première indication sur la répartition des canadiens selon leur consommation. Par exemple, la forte dépense dans le logement pour la population des unités 5, 6, 7, 13, 14, et 15 constatée dans la section 3.2 correspond à la consommation des locataires pauvres et quasi pauvres conformément à la logique. Les autres critères apportent également leur part d'explication, l'étude complète est fournie en annexe avec l'étude discriminante qui a confirmé l'impression laissée par la carte.

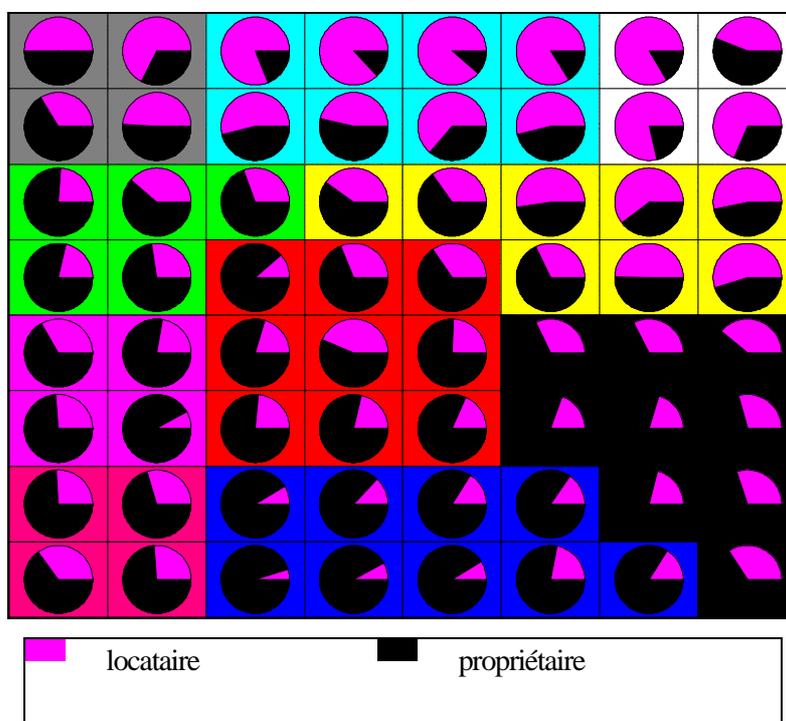


Figure 29 : Cartographie des propriétaires-locataires dans l'exemple 2

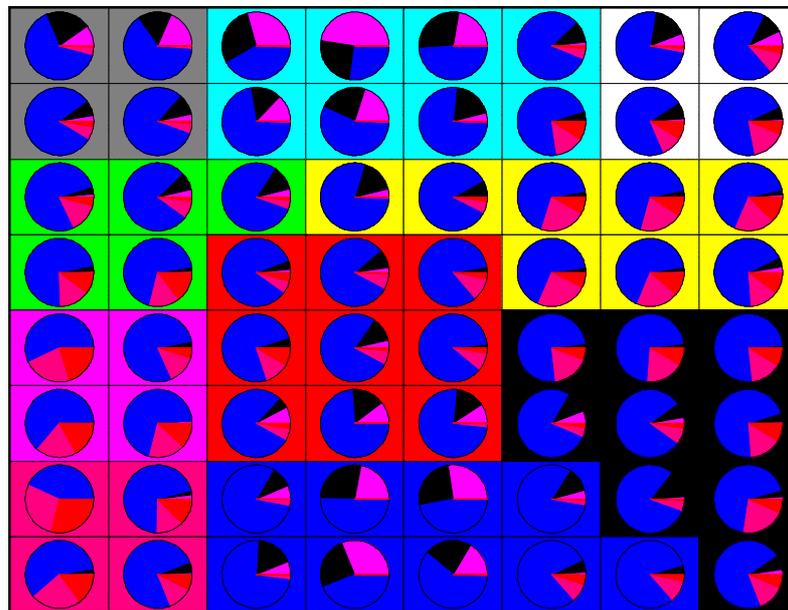


Figure 30 : Cartographie de la richesse-pauvreté dans l'exemple 2

Exemple 3 des données constituées des relevés journaliers d'une mesure faite toute les deux heures:

La cartographie des jours de semaine de la figure 32 définit 3 grandes zones: Celle des dimanches – les unités 8, 9, 18, 19, 28, 29, 38, 39, 48, 49, 50 et 59 – celle des samedis – les unités 7, 10, 17, 20, 27, 30, 37, 40, 47, 50, 58, 60, 69, et 70 – et celle des jours de semaine. La figure 31 permet de décomposer ces trois régions en fonction des mois. Par exemple, l'unité 23 est constituée des jours de semaine des mois de mai, juin, août et septembre, ses unités voisines 13, 22, 24 et 33 sont constituées des jours de semaine des mois de mai à septembre. Il serait intéressant de savoir comment sont répartis ces mois dans la région de ces 5 unités ou à l'intérieur de chacune d'entre elles. C'est ce que propose l'outil présenté dans la section suivante.

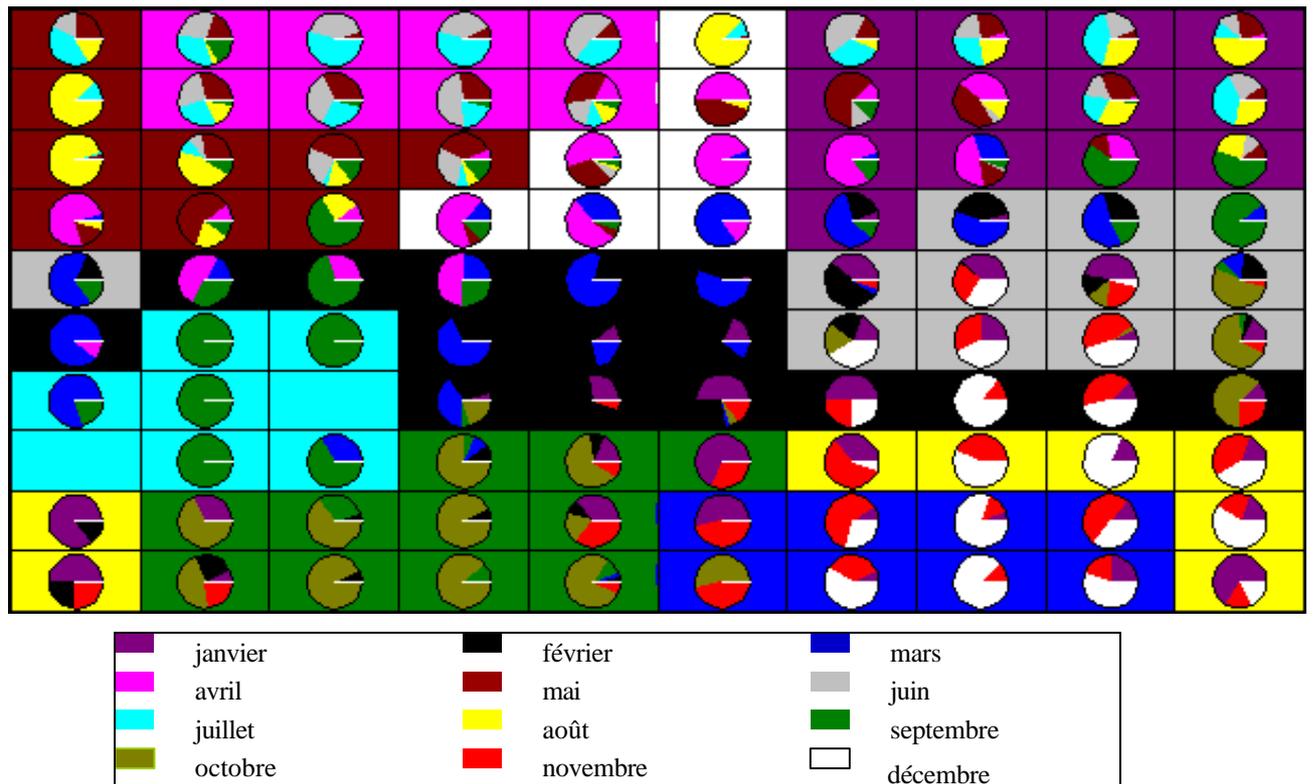


Figure 31 : Cartographie des mois pour l'exemple 3

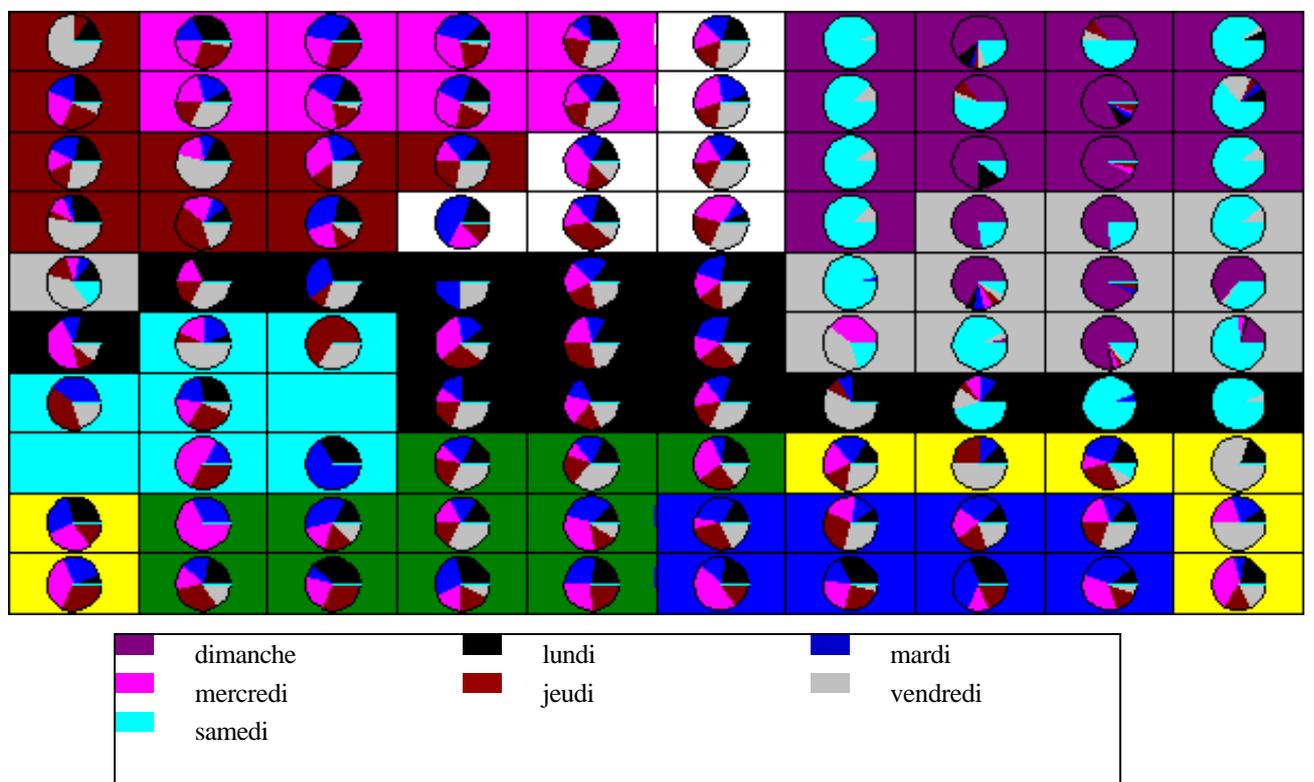


Figure 32 : Cartographie des jours de semaine pour l'exemple 3

3.7 Analyse restreinte à une région de la carte

On a vu que l'algorithme de Kohonen pouvait remplacer les analyses classiques lorsque celles-ci ne sont pas applicables. Cependant, la topologie fournie par la carte de Kohonen définit naturellement une notion de "localité" à partir de laquelle on peut réaliser des analyses linéaires restreintes, parfois judicieuses. Il suffit, en effet, d'extraire de la base de données les individus d'un ensemble de classes voisines et de les analyser avec une technique classique. On parlera dans ce cas d'analyse locale. On note que la notion de localité provient directement de la topologie fournie par la carte de Kohonen. Le principe a été inspiré du papier de M. Bishop, M. Jordan, M. Tipping [10] qui propose une classification descendante hiérarchique où sont représentés les plans principaux de tous les ensembles d'individus constitués par toutes les classes de tous les niveaux de dissociation.

Faire une étude des régions de la carte revient à multiplier les études. Ces dernières doivent donc être simples et rapides. C'est pourquoi il est raisonnable de se contenter de déterminer le plan qui ajuste le mieux, au sens des moindres carrés, la base de données extraite – c'est-à-dire le plan principal que définirait une analyse en composantes principales de cette région. Il peut également être perçu comme une approximation du plan tangent de la surface de la carte, en un point particulier de celle-ci (par exemple le vecteur code d'une unité) dont la projection associée est celle qui visualise le mieux le voisinage de ce point. L'exemple 3 choisi pour illustrer les phénomènes de dispersion va servir à présenter cette approche.

- *Notations:*

On choisit sur la carte les unités voisines qui nous intéressent. L'ensemble des individus affectés à ces unités constitue la base A des données extraites et on appelle $P(A)$ le plan qui ajuste ce nouveau nuage de points. La projection de l'individu i de coordonnées x_i dans \mathbf{R}^p sur $P(A)$ est, elle, notée $P(A)(x_i)$.

- *Technique d'ajustement du plan:*

Comme nous sommes supposés effectuer de nombreuses analyses locales, il est important que l'ajustement du plan s'effectue rapidement. Nous avons choisi d'utiliser l'algorithme EM (§...) pour construire le plan tangent afin d'avoir un temps de calcul faible.

Représentation:

En projetant les individus sur le plan tangent en un point (unité) particulier de la carte, on peut réaliser une représentation graphique qui décrit la zone de voisinage de ce point grâce au plan le mieux adapté à ce sous-ensemble de la base de données (celui qui contient maximum d'inertie).

3.7.1 Projections des observations de A dans le plan $P(A)$.

On peut marquer la projection des individus sur le plan par différents symboles. En choisissant un symbole qui contient des renseignements sur l'individu, on peut cumuler de l'information. Le symbole peut ainsi être le nom, mais aussi le numéro de la classe, la modalité correspondant à l'observation pour les q variables qualitatives, la qualité de la projection ou les valeurs des variables quantitatives ayant participé ou non à l'apprentissage.

La qualité de la projection *qual* de l'individu i pour le plan $P(A)$ est mesurée par:

$$qual(x_i) = \frac{\|x_i - P_{(A)}(x_i)\|}{\|x_i - \bar{x}_{(A)}\|} \text{ où } \bar{x}_{(A)} \text{ représente la moyenne de } x \text{ pour la population } A.$$

Ce mode de représentation facilite l'étude de la dispersion sur le plan et les relations entre une classe et ses voisines (quand la population A contient plus d'une unité).

Application à l'exemple 3 des données bi-horaires:

Les figures 33, 34, et 35 sont 3 façons de représenter ces projections. On a choisi d'ajuster la classe 23 (A est constitué de l'ensemble des observations classées dans l'unité 23). Dans la figure 34, les individus sont représentés par la variable qualitative qui renseigne sur le mois. Dans la figure 35, on a caractérisé l'observation par sa *qualité de projection*. On

constate dans cet exemple que *les jours du mois de mai* sont mal représentés dans le plan au contraire *des mois de juin, août, septembre*. On a de plus un mouvement de *juin à août-septembre* très clairement visible. Cette constatation invite à vérifier si ce mouvement reste vrai lorsque l'on projette sur ce plan les classes voisines comme le propose le chapitre suivant.

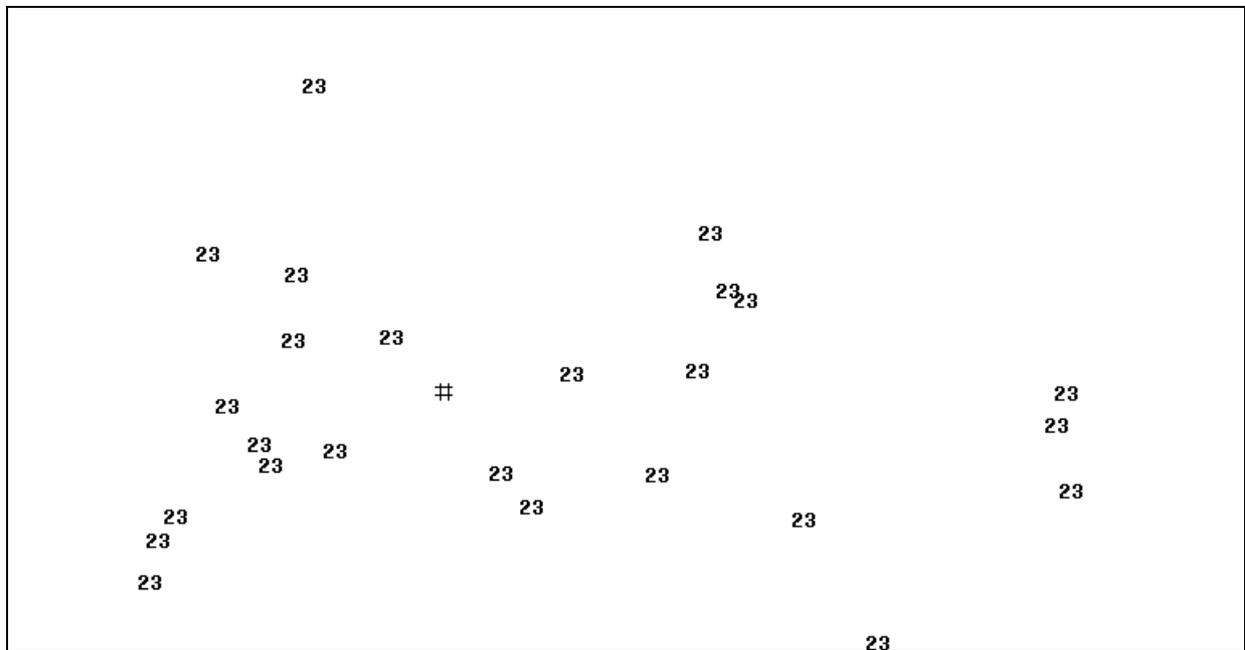


Figure 33 : Plan principal restreint à la classe associée à l'unité 26 ou plan tangent à la surface de la carte au point qui coïncide avec le vecteur code de cette même unité

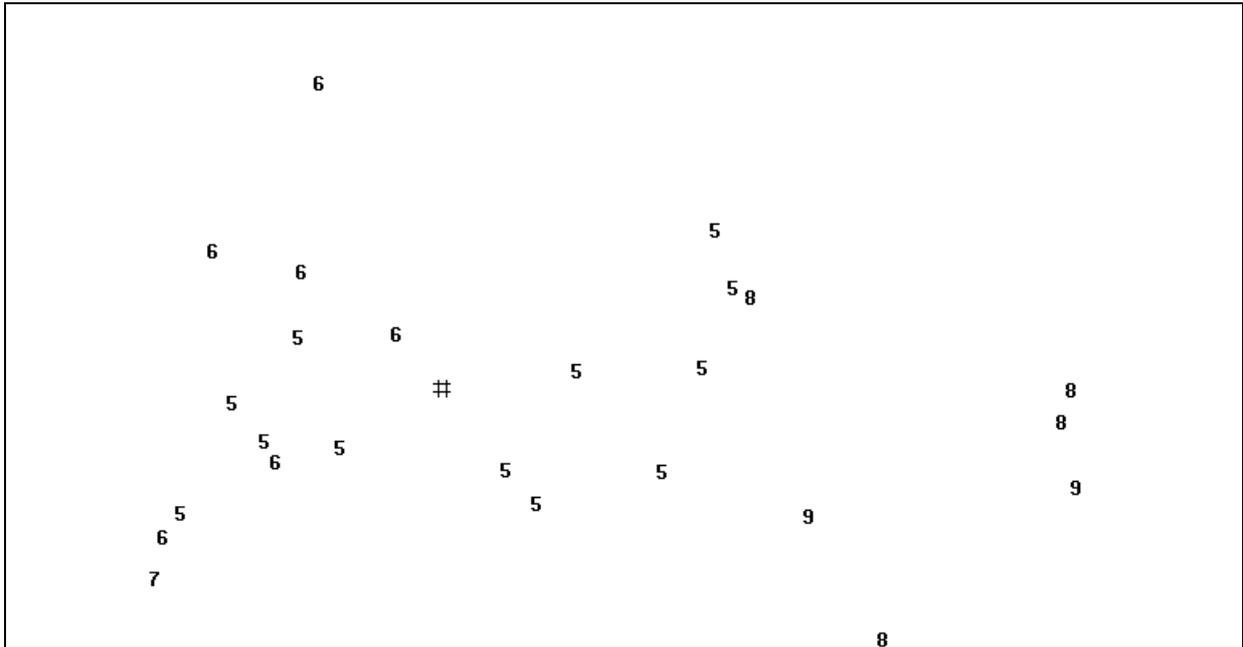


Figure 34 : Les points sont repérés par le mois. Le mois de juin est plutôt à gauche, mai au centre et août à droite

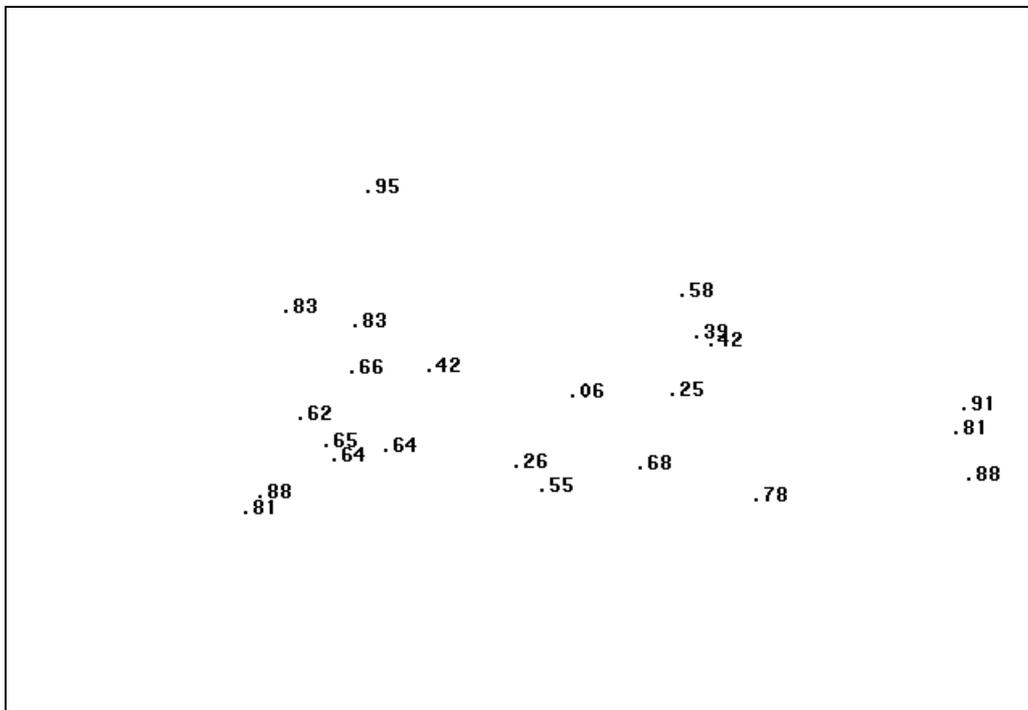


Figure 35 : Les points sont repérés par la qualité de leur projection. Les jours de mai sont mal représentés par ce plan.

3.7.2 Représentation des classes voisines

Comme nous l'avons déjà souligné dans la section 3.4, l'algorithme de Kohonen affecte davantage de classes aux populations qui ont un gros effectif. Il est donc possible que les propriétés constatées sur une unité soient celles d'un ensemble élargi à ses voisines. En particulier il est intéressant de vérifier si le plan qui permet de visualiser la typologie d'une unité explique également celle de ses voisines. Pour cela, deux approches sont possibles:

La première approche consiste à reprendre la même représentation que précédemment en remplaçant A par A et ses voisins, on note cet ensemble $V(A)$. Ceci change $P(A)$ en $P(A \cup V(A))$ et peut modifier les conclusions (en particulier les axes principaux). Pour éviter que cette méthode ne génère trop d'études, il vaut mieux posséder un critère de sélection de $V(A)$, la deuxième approche peut en fournir un.

La deuxième approche s'applique dans le cas où on veut privilégier l'étude de A et en même temps découvrir si les conclusions sont généralisables, il faut dans un premier temps ajuster A par $P(A)$ et projeter sur ce plan les individus de $A \cup V(A)$. Une classe voisine dont les individus ont un bon critère de projection et amènent aux mêmes conclusions que ceux de A peut lui être associée. C'est aussi un moyen de choisir les classes qui méritent d'être ajoutées à A dans une étude de type « première approche ».

Exemple:

On a choisi d'ajuster la classe 23 par le plan $P(23)$ et d'y projeter les classes voisines de 23 soit 13, 22, 24, 33. Les individus de l'unité 23 sont estimés de la même façon qu'à la section 3.7.1 et la projection des autres confirme les mêmes conclusions (*un mouvement de juin -juillet vers août-septembre*).

Remarques:

Sur la carte de la figure 38, on peut analyser la projection des individus des autres classes et distinguer ceux qui ont une qualité de projection satisfaisante. La carte de la figure 36 ne permet pas de distinguer par ce plan les différentes classes. Les axes principaux ne sont pas discriminants des 5 classes, ce qui montre que l'analyse de données avec la carte de

Kohonen a apporté un plus par rapport à une analyse linéaire à 2 dimensions. La carte de la figure 37 montre que l'on peut généraliser le mouvement *juin-juillet* -> *août-septembre* aux classes voisines.

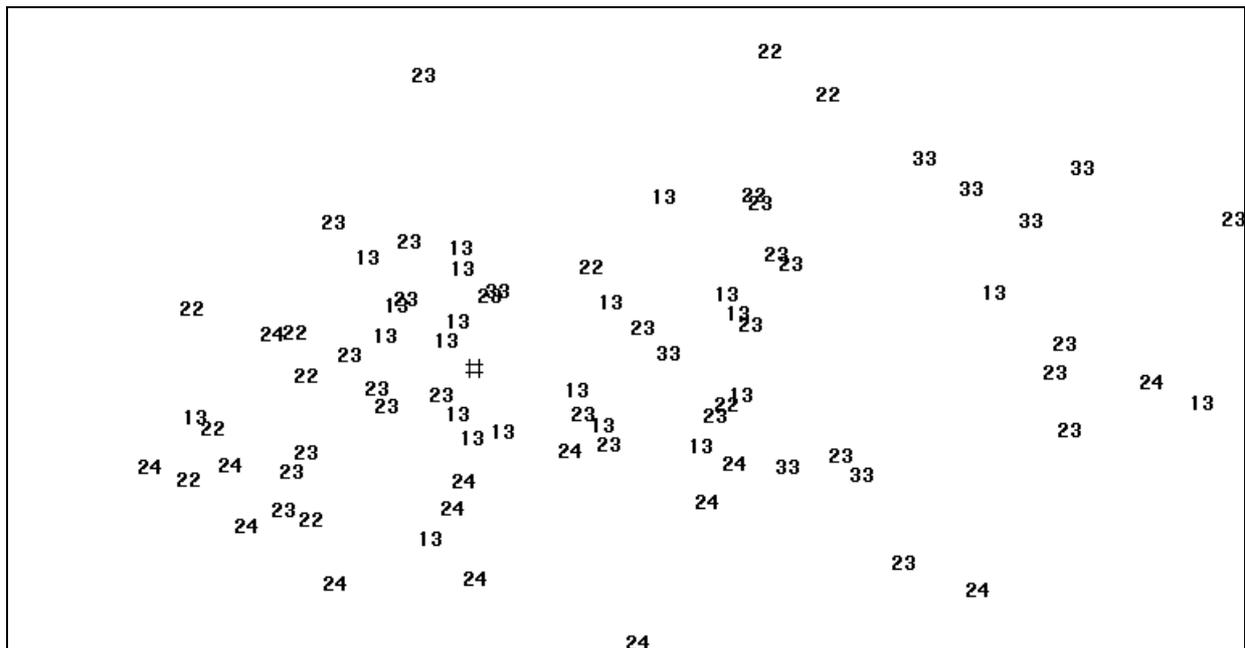


Figure 36 : Plan principal restreint à la classe associée à l'unité 23 sur lequel sont projetés les individus affectés aux unités voisines 13, 12, 24, 33

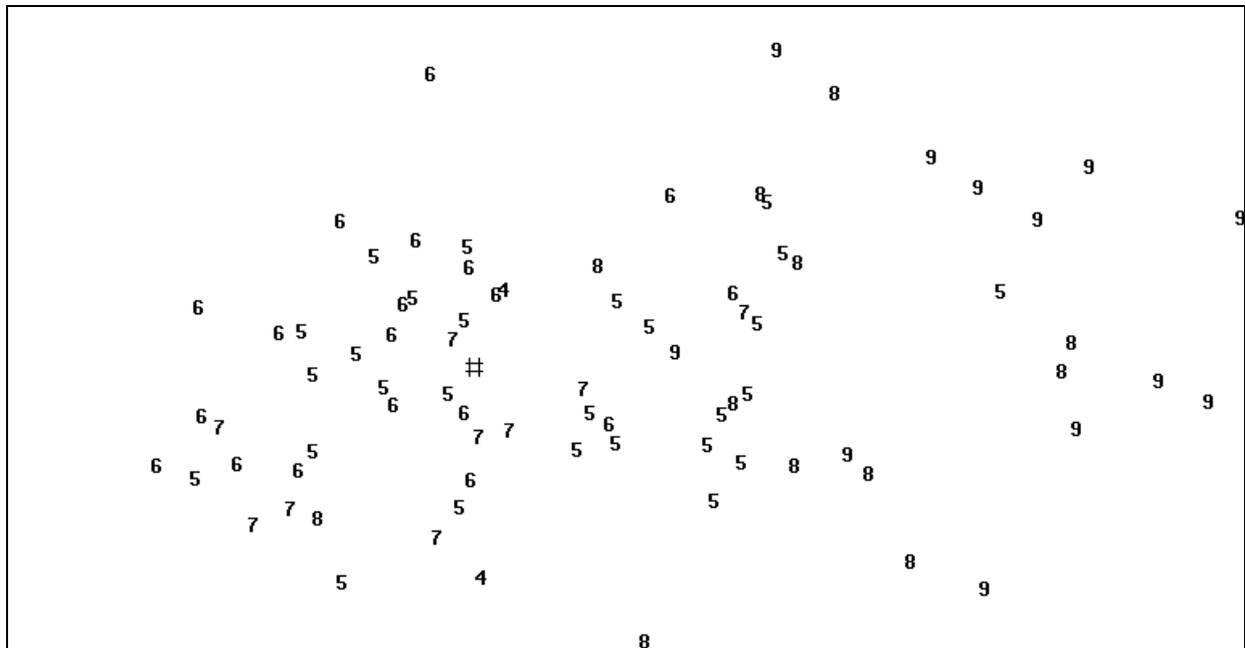


Figure 37: Les points sont repérés par le mois. Les individus de juin-juillet sont à gauche, de mai au centre et de août-septembre à droite

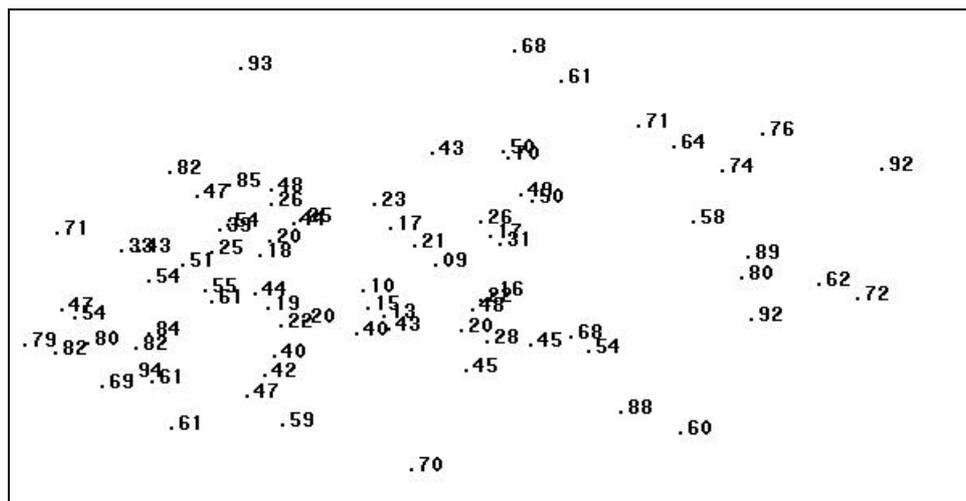


Figure 38: Les points sont repérés par la qualité de leur projection. Les jours de mai sont mal représentés par ce plan, les autres le sont très bien.

3.8 Comparaison avec les méthodes classiques.

Dans ce chapitre, la classification à l'aide de l'algorithme de Kohonen va être comparée à la méthode présentée au chapitre 3.1.1 qui couple l'analyse factorielle avec une classification et en particulier l'exemple de l'analyse en composantes principales couplée avec une classification hiérarchique qui utilise la distance de Ward. La base de données qui sert de test, issue d'une simulation, est celle représentée par la figure 39 où le noyau de points a une forme de fer à cheval. Elle a été élaborée pour illustrer certaines limites de l'analyse en composantes principales et comment la classification de Kohonen peut remplacer une analyse factorielle quand celle-ci est mal adaptée.

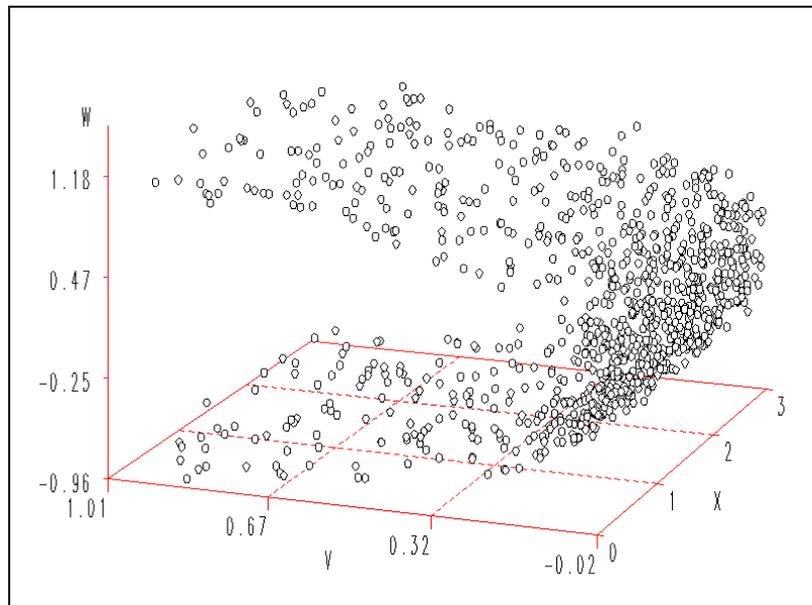


Figure 39 : Représentation du fer à cheval

3.8.1 L'exemple du fer à cheval – Les représentations issues de l'ACP.

Première étape: la classification de type Ward.

La première étape consiste à effectuer une classification de type Ward. Suivant la méthode présentée au chapitre I.2.2, on a choisi 6 classes au regard des niveaux d'inertie. La couleur permet de repérer la classe des points sur la figure 40. Dans la suite, tous les graphiques respecteront le même codage des classes (une même couleur correspond à la même classe dans toutes les figures).

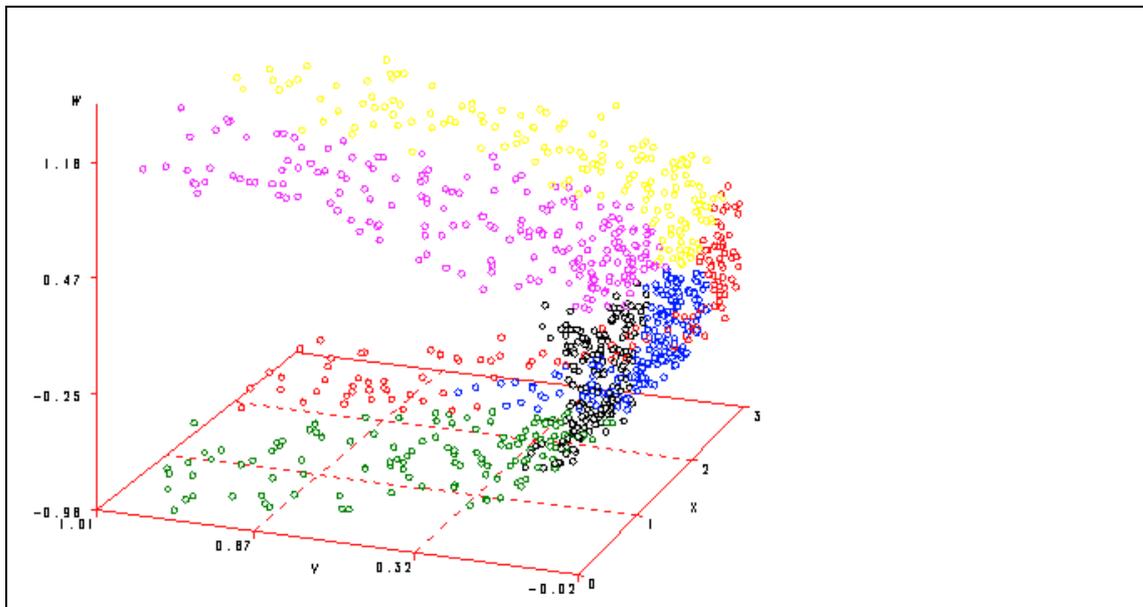


Figure 40 : Les couleurs qui représentent la classe de l'individu améliorent la lisibilité du graphique

Deuxième étape: l'analyse en composantes principales

La base de données étant de dimension trois, on trouve trois axes principaux dim1, dim2, dim3 qui contiennent 100% de l'inertie. Sur la figure 41 est représentée la projection orthogonale du nuage de points sur le plan défini par les axes dim1 et dim2. On constate d'abord que la représentation n'est pas spécialement meilleure que celle de la figure 40. De plus, la projection écrase la troisième dimension ce qui nous laisse penser que les classes noire,

verte, et bleue sont mélangées. Il est à noter que la classe jaune est bien isolée, au contraire de la verte, alors que la symétrie du nuage de points devrait aboutir à une même qualité de représentation pour ces deux classes. Ceci amène deux remarques : la classification n'explique pas complètement la représentation sur le plan principal et la réduction de l'information induit en erreur par un phénomène d'écrasement classique dû à la projection.

Sur la figure 42 supplémentaire de la précédente est représentée le plan secondaire (dim1, dim3). Les deux plans représentent ainsi trois dimensions c'est-à-dire, dans notre exemple, l'ensemble de l'information (la base étant de dimension 3). Si la représentation n'est pas excellente, c'est donc uniquement dû au système de projection.

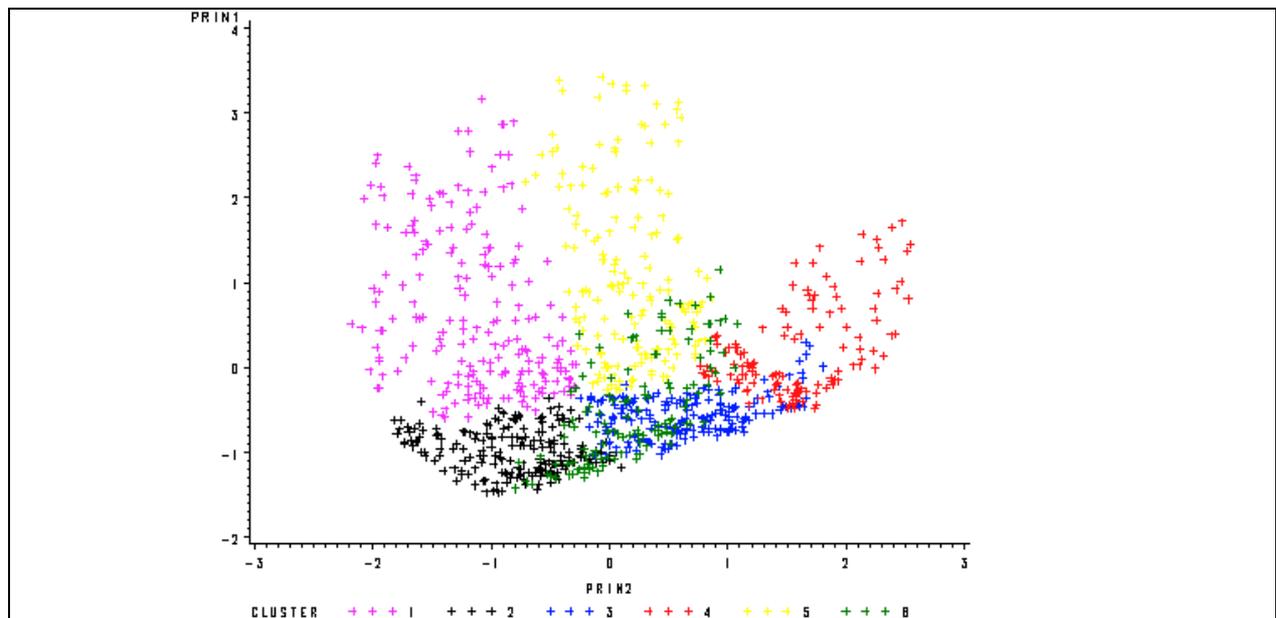


Figure 41: Plan principal (dim1, dim2)

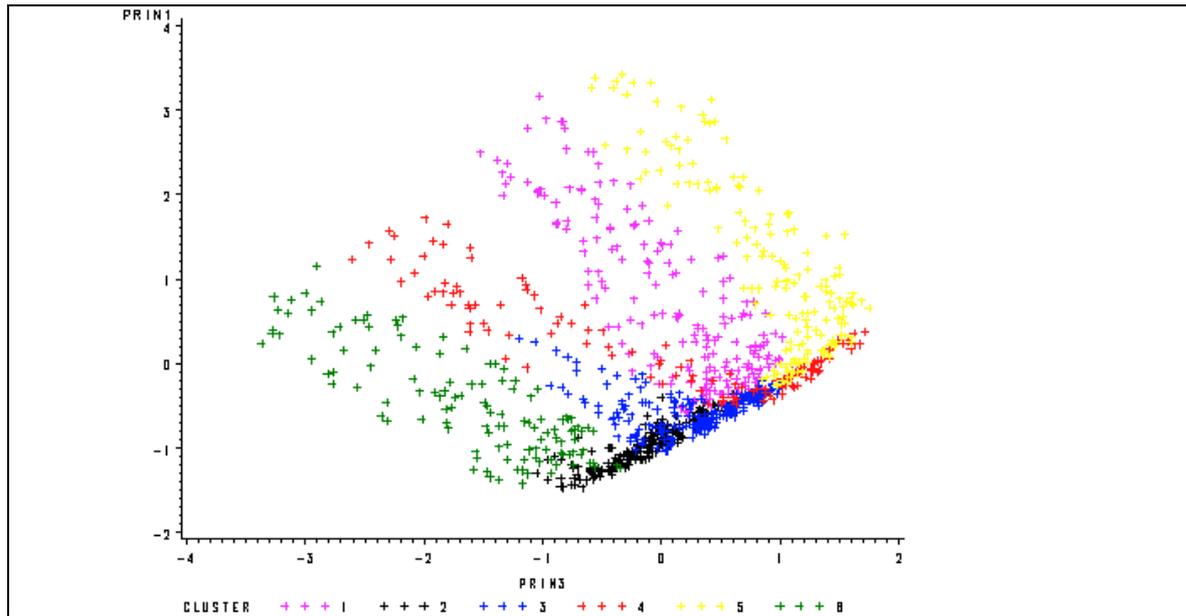


Figure 42: Plan secondaire (dim1, dim3)

- Remarques:

On se rend compte dans le cas du fer à cheval, que même sans réduire l'information, la représentation n'est pas complètement satisfaisante. En particulier les classes des verts, rouges, magentas et jaunes sont mieux représentées que les classes des noirs et bleus.

Cet exemple aussi met en évidence le fait que l'analyse en composante principale n'optimise pas nécessairement la représentation d'une classification avec laquelle elle serait couplée.

3.8.2 Utilisation de la classification à l'aide de l'algorithme de Kohonen.

La figure 43 représente le résultat de la classification de Kohonen. C'est un résumé de l'information qui est représenté, ici, au contraire de l'analyse en composantes principales du chapitre précédent. Dans chaque camembert, la part de couleur c est la proportion d'éléments de la classe de couleur c issue de la classification hiérarchique faite dans la section 3.7.3.1. Il est noter que la représentation est conçue pour bien représenter la classification de Kohonen et pas forcément la classification hiérarchique.

Remarques:

Les classes issues de la classification hiérarchique sont très bien différenciées et ordonnées d'une façon qui montre la continuité et le mouvement interne de la structure du noyau.

On a perdu la structure en fer à cheval, c'est-à-dire que les extrémités du noyau de points ont tendance à se rapprocher. Deux surfaces, l'une courbe et l'autre plane, fournissent le même type de carte. Cependant, le fait de ne pas retrouver la forme fer à cheval n'est absolument pas frustrant dans la mesure où la notion de forme n'existe plus dès que la dimension excède 3.

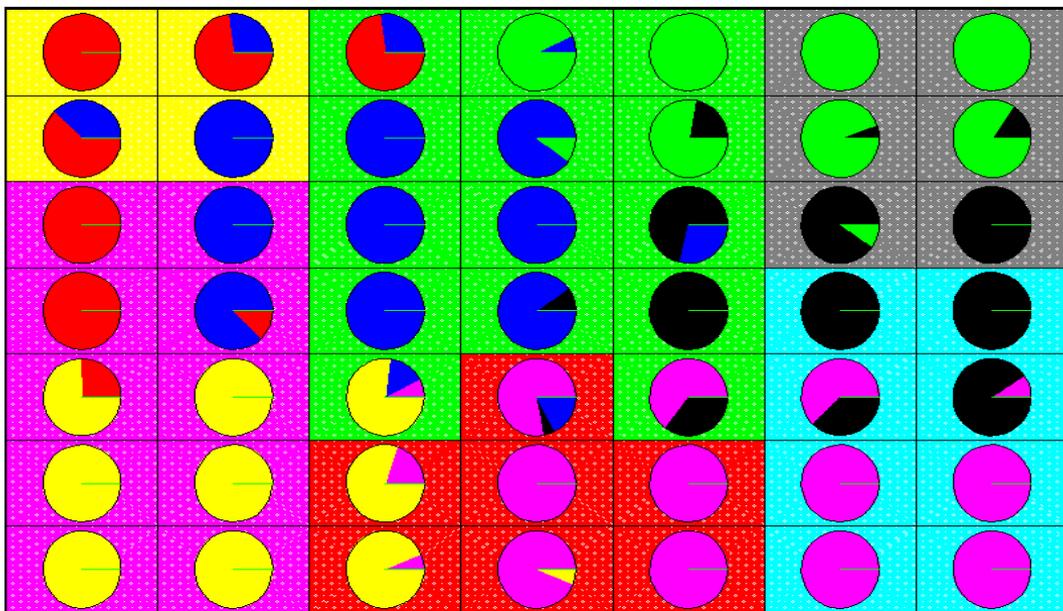


Figure 43: La carte de Kohonen qui ajuste le fer à cheval

3.8.3 Comparaison entre la classification de Kohonen et l'analyse factorielle

Différences d'approche entre la classification de Kohonen et le couplage « classification – analyse de données » issue de la statistique classique:

Dans le cadre classique, l'analyse de données est faite indépendamment de la classification. Chacune d'entre elles n'est donc pas conçue pour optimiser l'explication de l'autre. En effet, c'est l'ensemble des variables (de dimension supérieur à 3) qui a déterminé la

classification et seuls les deux premiers plans de l'analyse factorielle (les 3 dimensions principales) peuvent être exploités le plus souvent. Il n'y a donc aucune garantie d'avoir une bonne représentation de la classification, si on la projette sur ces axes. On ne percevra pas, entre autres, les classes issues de l'inertie ignorée dans ces trois dimensions. En résumé, la classification est indépendante de l'analyse factorielle; si elle aide à son interprétation, elle n'est pas conçue pour la compléter.

Au contraire, dans le cas des cartes de Kohonen, la classification et la représentation sont liées. On cherche, en effet, une surface qui épouse au mieux le nuage de points dont l'information est réduite à un nombre de classes fixes. Le souci de la représentation est, ici, de relater au mieux la classification. De plus, les contraintes imposées à la classification (contraintes de voisinage) ont pour but de faire émerger des propriétés qui sont traduites par la carte.

Notre méthode prend donc en compte simultanément les problèmes de classification et de représentation. De son côté, l'analyse factorielle recherche une structure linéaire, car c'est la plus facile à représenter et la méthode de classification est choisie en fonction de caractéristiques de la base de données.

Différences d'exploitation:

Dans le cadre de la statistique classique, au delà du plan principal, c'est la valeur ajoutée qui est représentée. La dimension est généralement réduite à 3 en raison de notre incapacité à interpréter les dimensions supérieures. Dans l'algorithme de Kohonen, le support de représentation (ficelle, grille) est généralement d'ordre inférieur à 2 (les représentations en 3 dimensions étant jusqu'à présent peu lisibles) mais on ne fixe pas la contrainte de linéarité ce qui permet à la surface de mieux épouser le nuage de points.

La contrainte de linéarité est bien plus forte que celle du simple voisinage. Elle rend la lecture du plan principal plus facile, mais peut compromettre la qualité de l'ajustement. Or, en pratique, c'est surtout la continuité et la tendance qui sont exploitées, car la combinaison linéaire des vecteurs propres n'a pas toujours de réalité pour l'utilisateur. Par exemple, considérons le cas suivant : $dim1$ et $dim2$ sont les deux premiers axes d'inertie, P la projection sur le plan principal et les individus i et i' vérifient $P(x_i) = 3 \cdot dim1 + 2 \cdot dim2$ et

$P(x_i) = \dim1 + 0.66 \cdot \dim2$. Si M est la moyenne du nuage, comme le modèle est linéaire, x_i et $x_{i'}$ vérifient $\overrightarrow{MP}(x_i) = 3 \times \overrightarrow{MP}(x_{i'})$. Mais l'interprétation devient souvent x_i et $x_{i'}$ vont une même direction dans le sens « $\dim1$ et $\dim2$ positifs ». En effet, $\dim1$ et $\dim2$ étant déjà difficiles à interpréter par rapport aux notions connues (les p variables) leurs combinaisons linéaires telles que $3 \cdot \dim1 + 2 \cdot \dim2$ le sont tout autant.

Par le critère de voisinage, la carte de Kohonen traduit les qualités de continuité et permet de visualiser les tendances et les sens sans imposer la contrainte d'ajustement linéaire. Par conséquent, la classification par la carte de Kohonen a moins d'exigence pour presque autant de possibilités d'exploitation.

Bilan de la comparaison de l'analyse factorielle et des cartes de Kohonen :

La contrainte de linéarité de l'analyse factorielle est beaucoup plus forte que celle de voisinage des cartes de Kohonen, elle peut donc contrarier un bon ajustement des données. Elle permet, néanmoins, de faire des projections sur des plans qui sont des représentations plus familières aux utilisateurs que les cartes de Kohonen. Par contre, pour ces dernières, la représentation est conjuguée avec la classification, ce qui la rend plus fiable que la combinaison classification hiérarchique – analyse factorielle. De plus, les cartes de Kohonen ont des propriétés de robustesse que ne possède pas l'analyse factorielle dont les axes principaux sont sensibles aux valeurs extrêmes. Enfin, l'ensemble des bases de données où la carte de Kohonen peut être appliquée contient le domaine d'application de l'analyse factorielle.