

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.

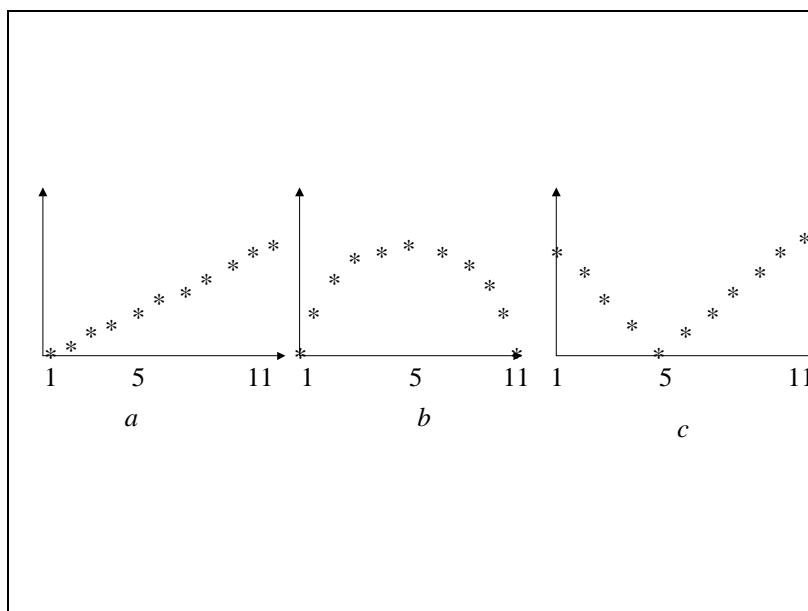


Figure 1 : Les trois formes a , b et c sont construites à partir de 11 variables

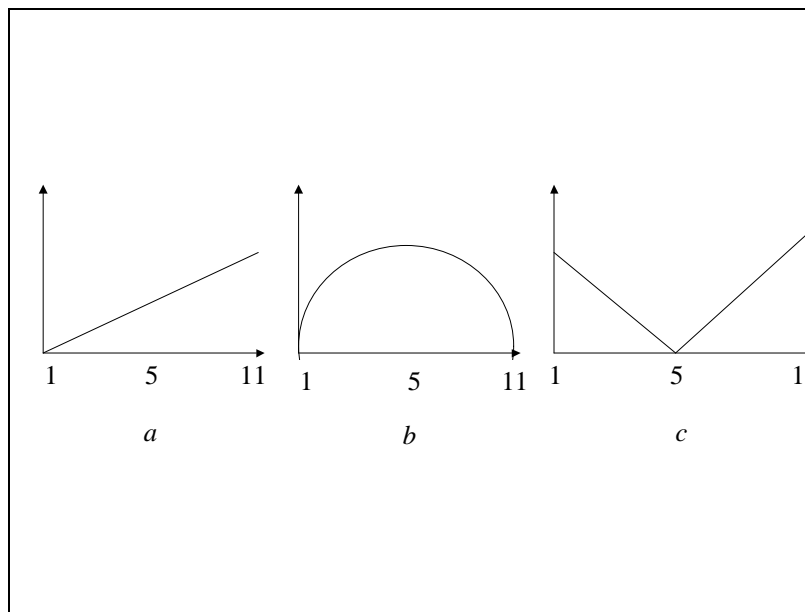


Figure 2 : Les 11 points sont joints pour mettre en évidence la forme

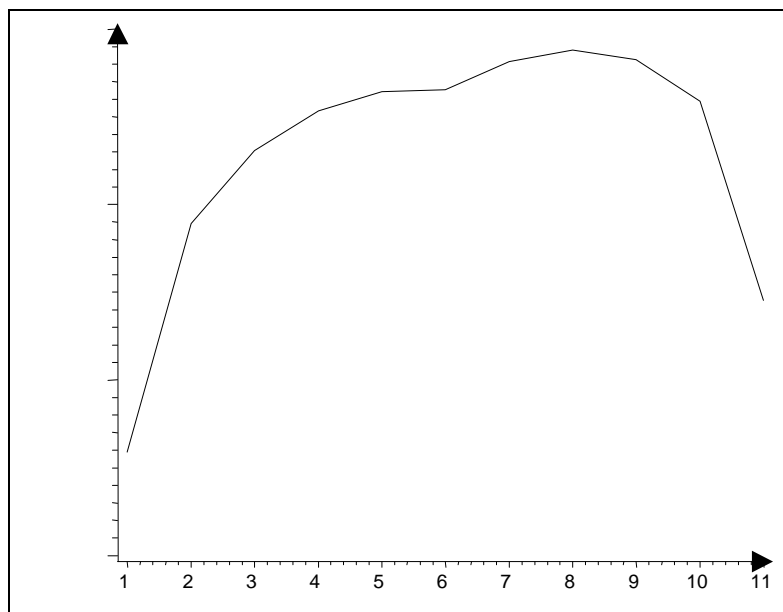


Figure 3 : Exemple d'une forme construite à partir du mélange de a , b et c avec $a=0.44$, $b=0.17$ et $g=0.12$

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.

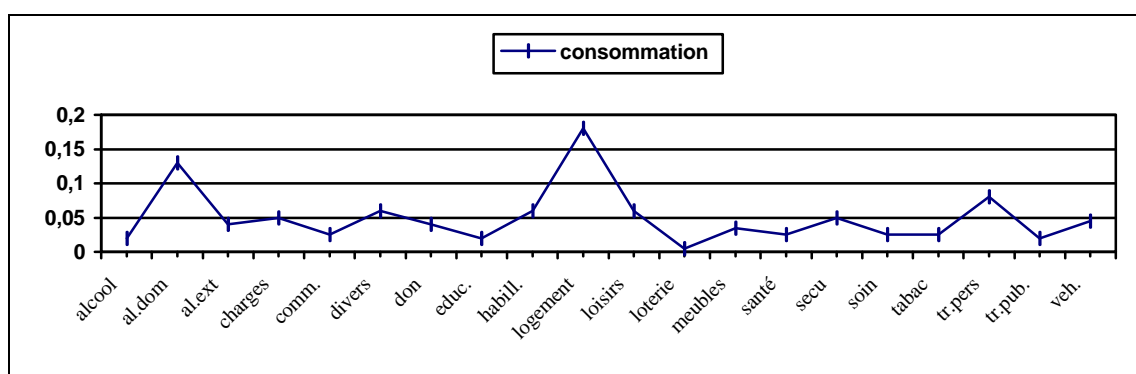


Figure 4 : Profil de consommation de l'individu i

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.

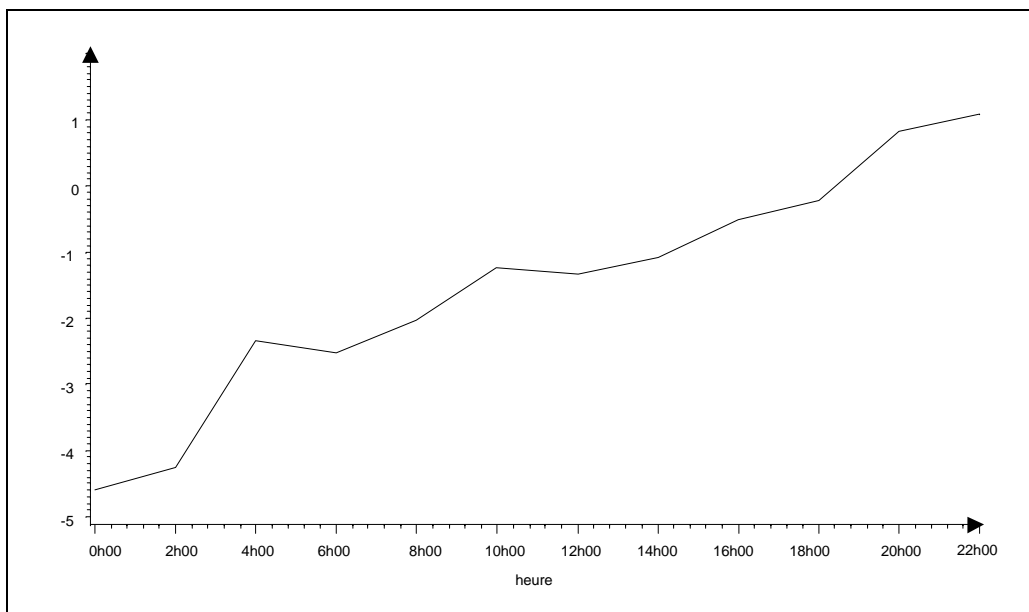


Figure 5 :Relevé bi-horaire du jour i_1

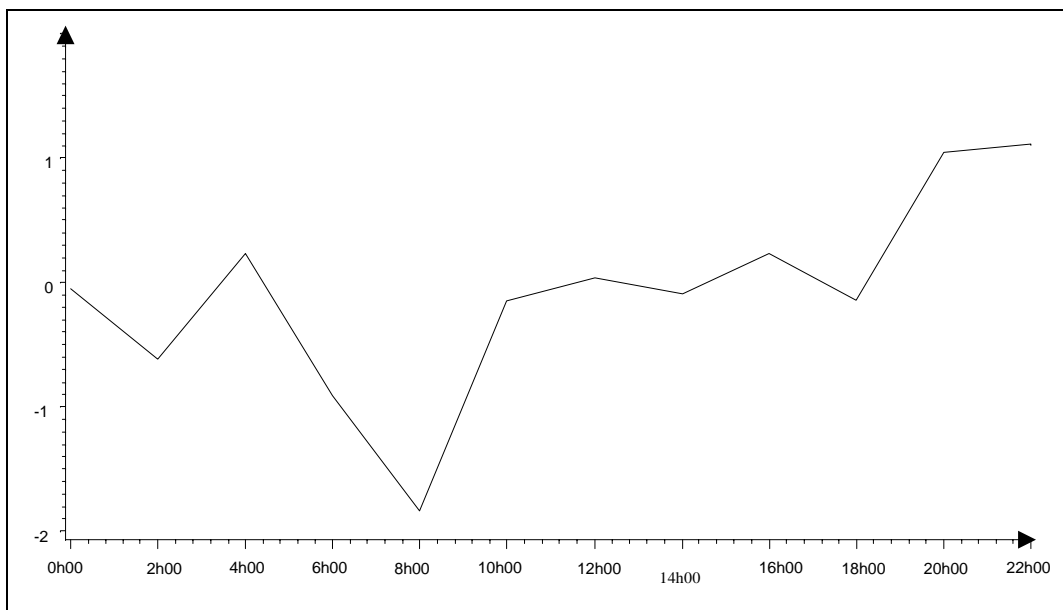


Figure 6 :Relevé bi-horaire du jour i_2

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

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 $\{G_k^{m+1}\}_{k=1, \dots, U}$

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

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

Conclusion

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

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.

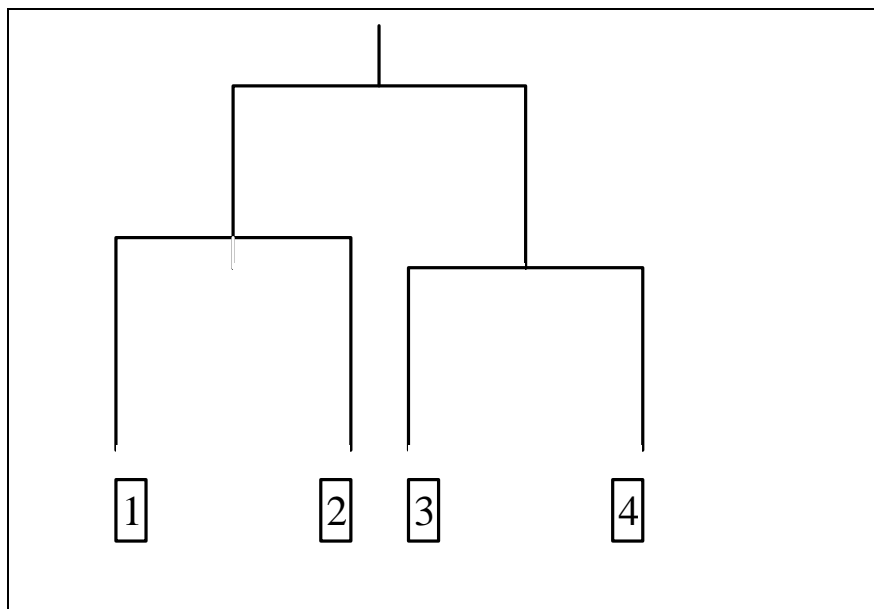


Figure 7 : Dendrogramme 1, la classe2 est représentée entre la classe1 et de la classe3

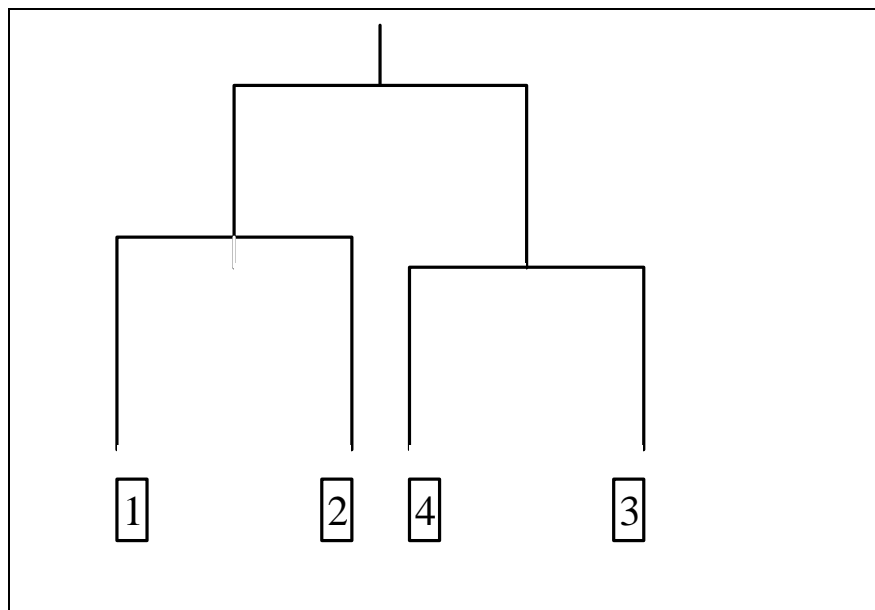


Figure 8: Dendrogramme 2, la classe2 est représentée entre la classe1 et de la classe4

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 :

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

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

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

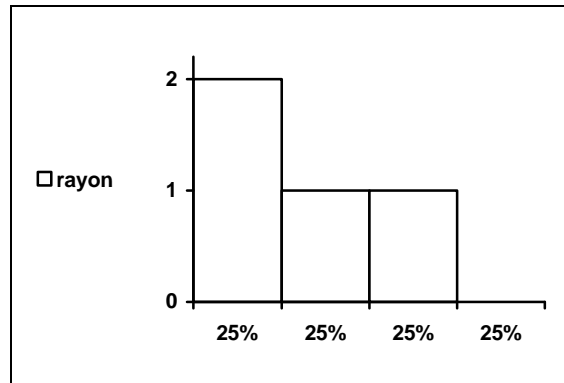


Figure 11 : Répartition en pourcentage du nombre d'itérations utilisées pour chaque rayon

- 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.

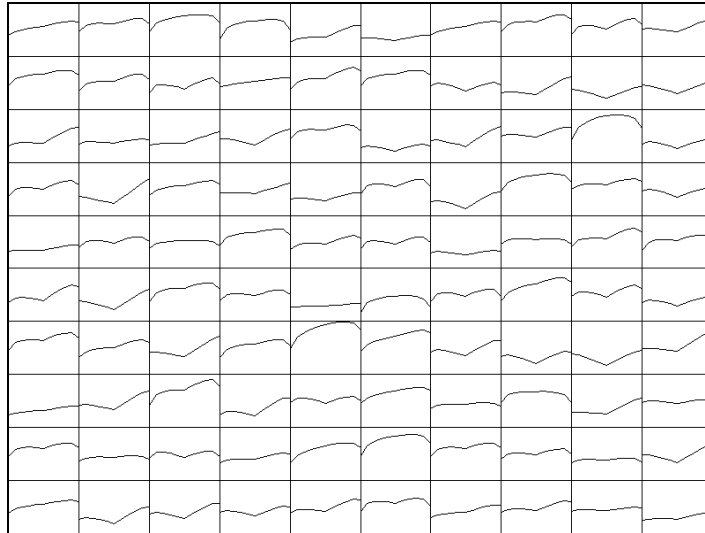


Figure 12 : L'initialisation



Figure 13 : Premier individu présenté

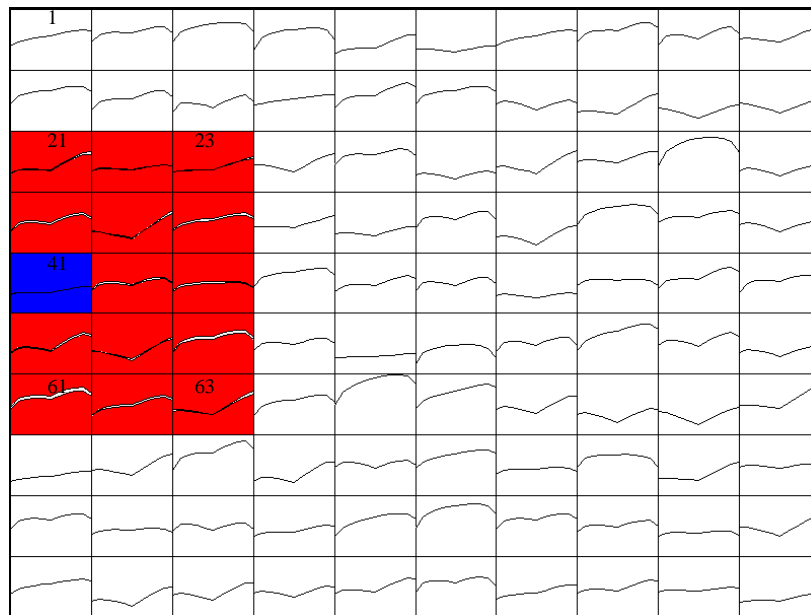


Figure 14 : Première modification des vecteurs codes au rayon2

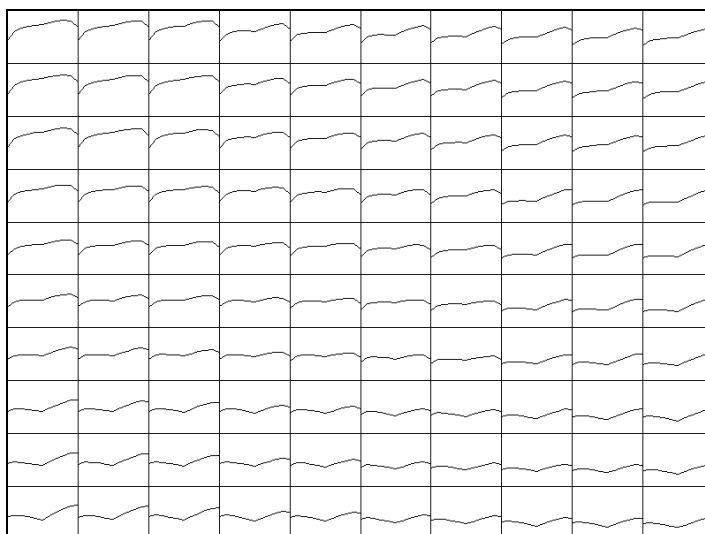


Figure 15 : Etat de la carte à la fin de l'apprentissage au rayon 2

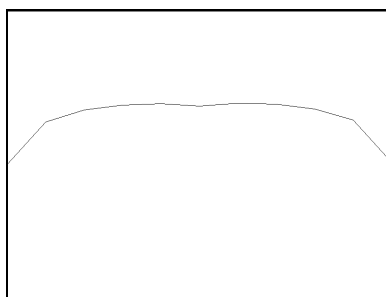


Figure 16 : Premier individu présenté pour l'apprentissage au rayon 1

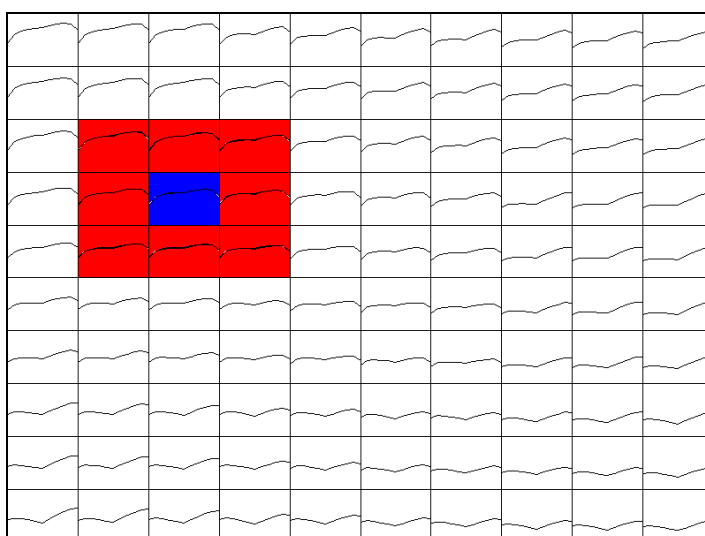


Figure 17 : Première modification des vecteurs codes au rayon 1

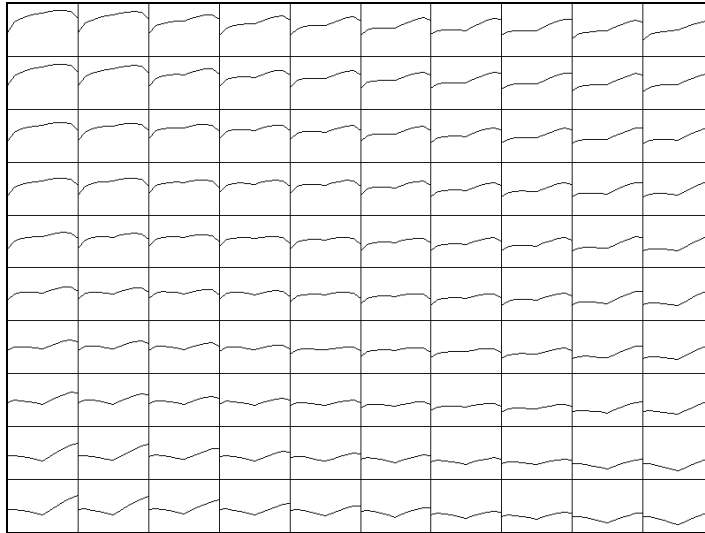


Figure 18 : Etat de la carte à la fin de l'apprentissage au rayon 1



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

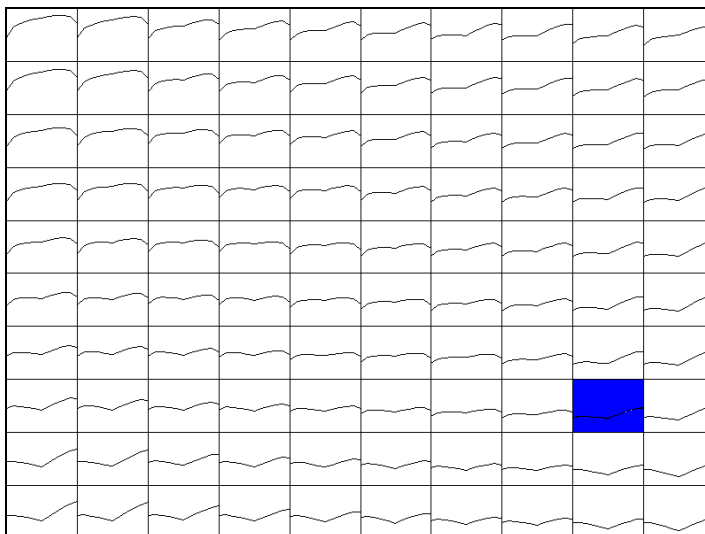


Figure 20 : Première modification des vecteurs codes au rayon 0

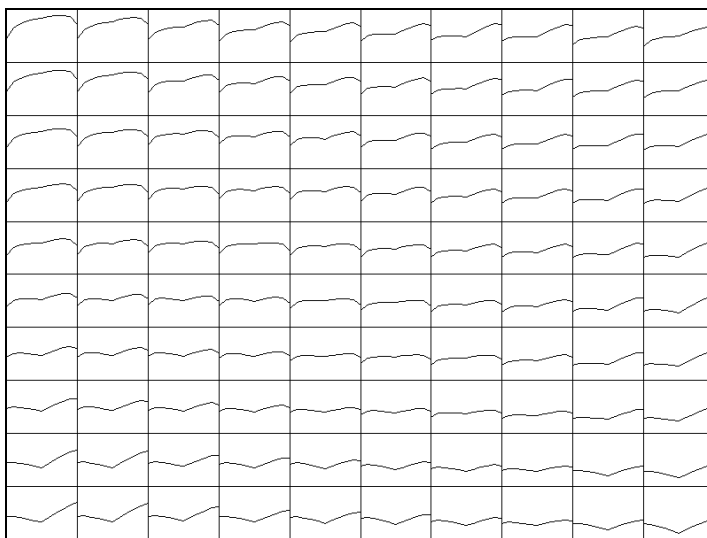


Figure 21 : Etat final de la carte

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.

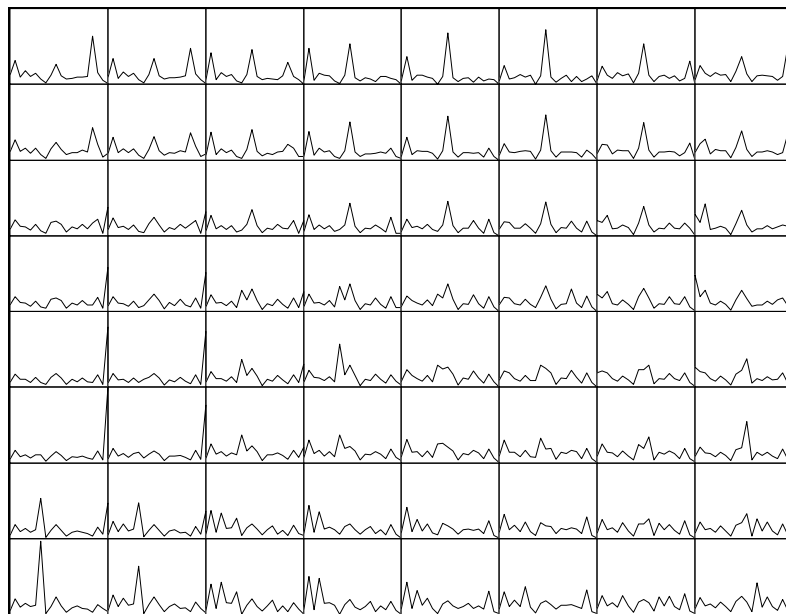


Figure 22 : Les vecteurs codes sont représentés dans leur unité d'affectation

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.

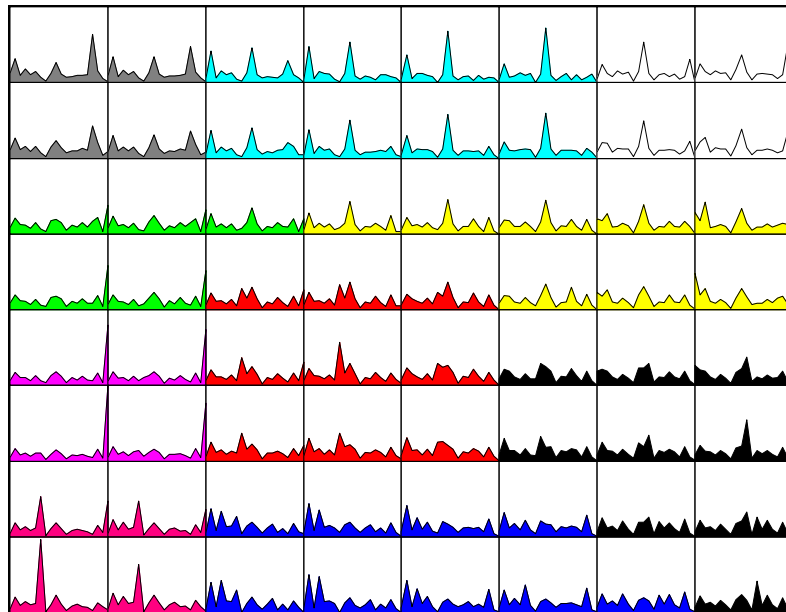


Figure 23 : Au graphique précédent a été rajouté une couleur du remplissage qui correspond à la macro-classe associée à l'unité

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 $G_{u'}$.

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

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 $u+1$.

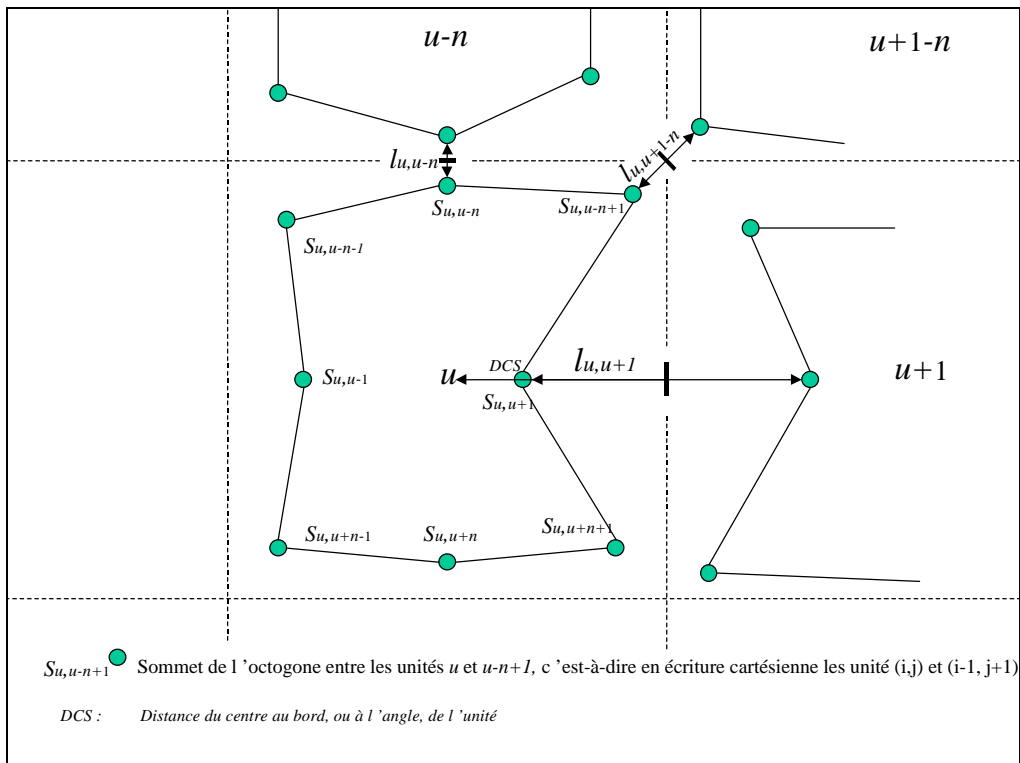


Figure 24 : Chaque sommet de l'octogone associé à l'unité u se rétracte d'une longueur proportionnelle à la distance entre les vecteurs codes des unités qui partagent le bord ou l'angle correspondant

- 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.

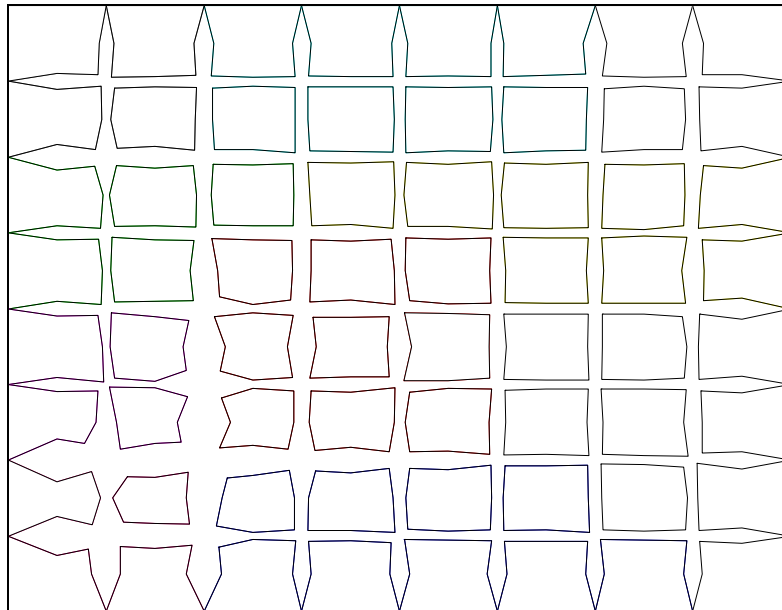


Figure 25 : Carte des distances

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.

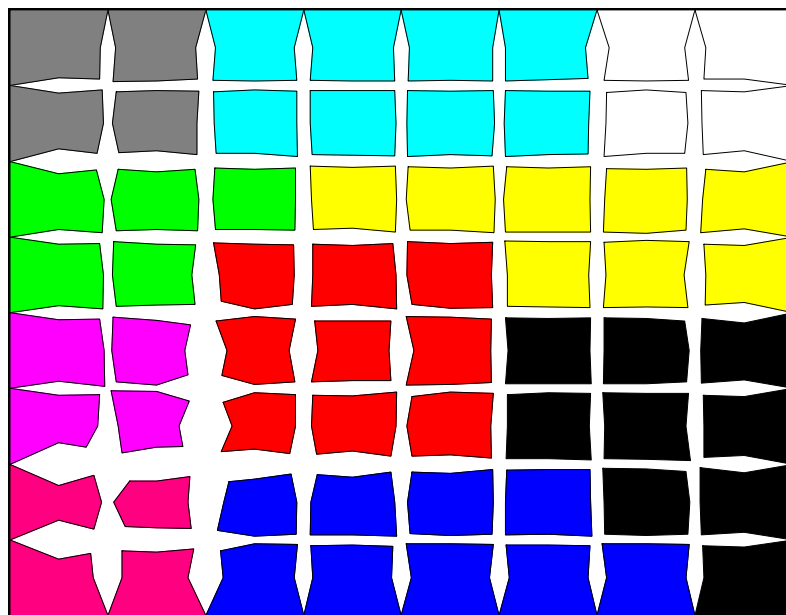


Figure 26 : Les octogones de la carte des distances sont remplis de la couleur qui correspond à la macro-classe de l'unité

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.

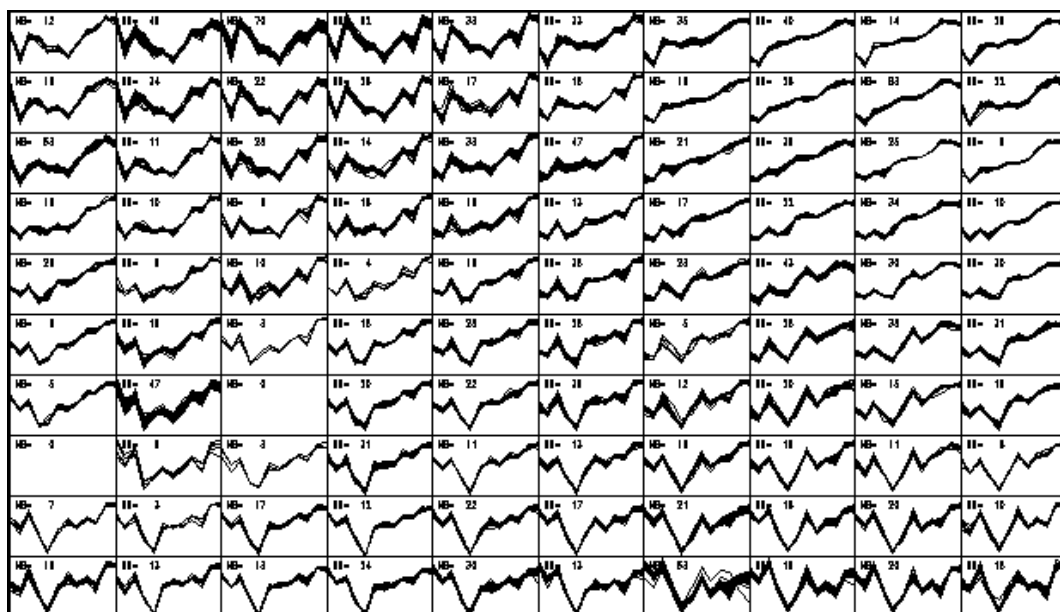


Figure 27 : Tous les individus et l'effectif de chaque classe sont représentés dans leur unité d'affectation

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 :

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

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 ? ».

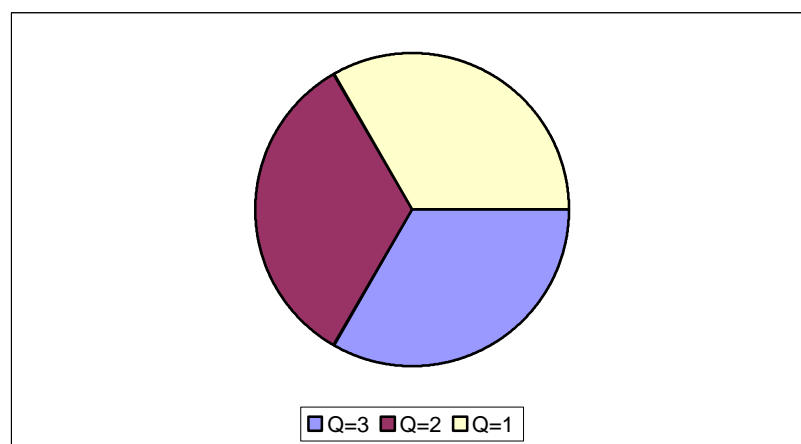


Figure 28 : Camembert qui représente la répartition des modalités 1,2 et 3 de la variable Q

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.

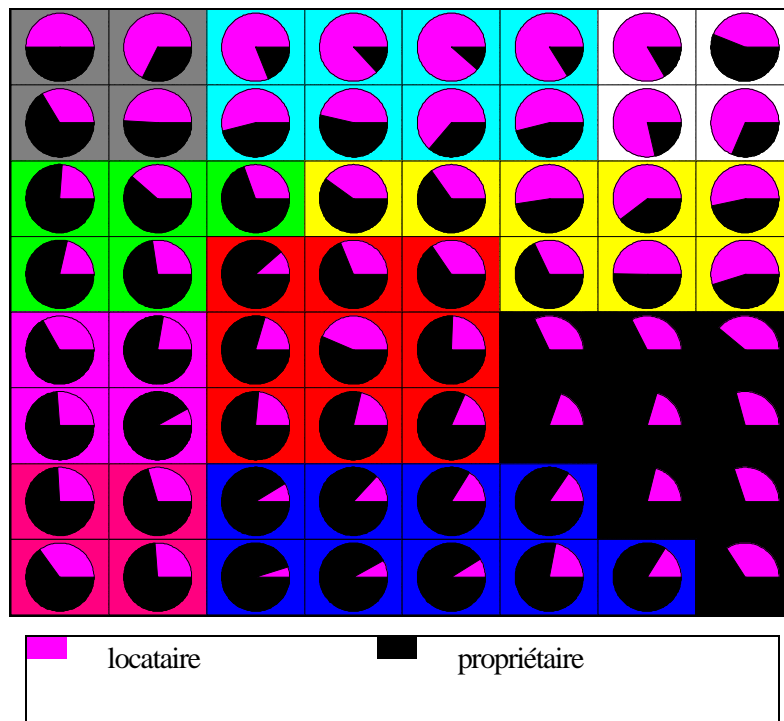


Figure 29 : Cartographie des propriétaires-locataires dans l'exemple 2

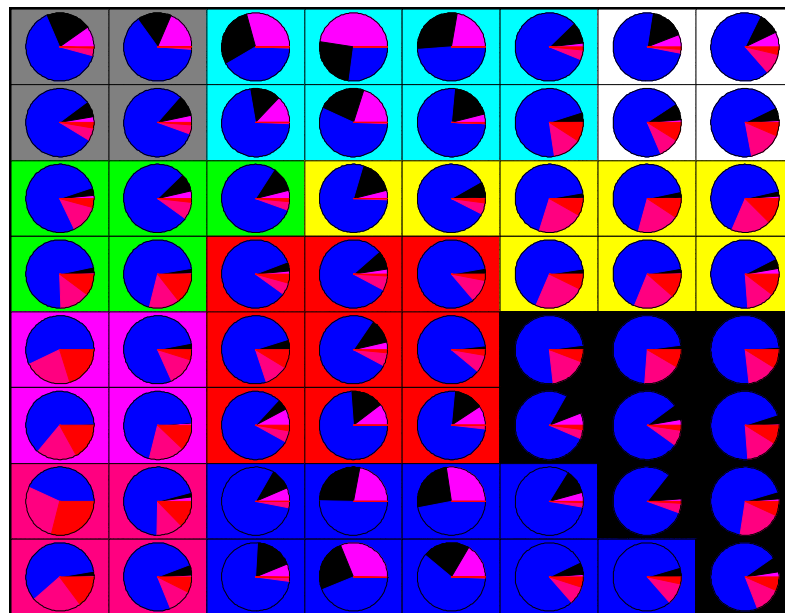


Figure 30 : Cartographie de la richesse-pauvreté dans l'exemple 2

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.

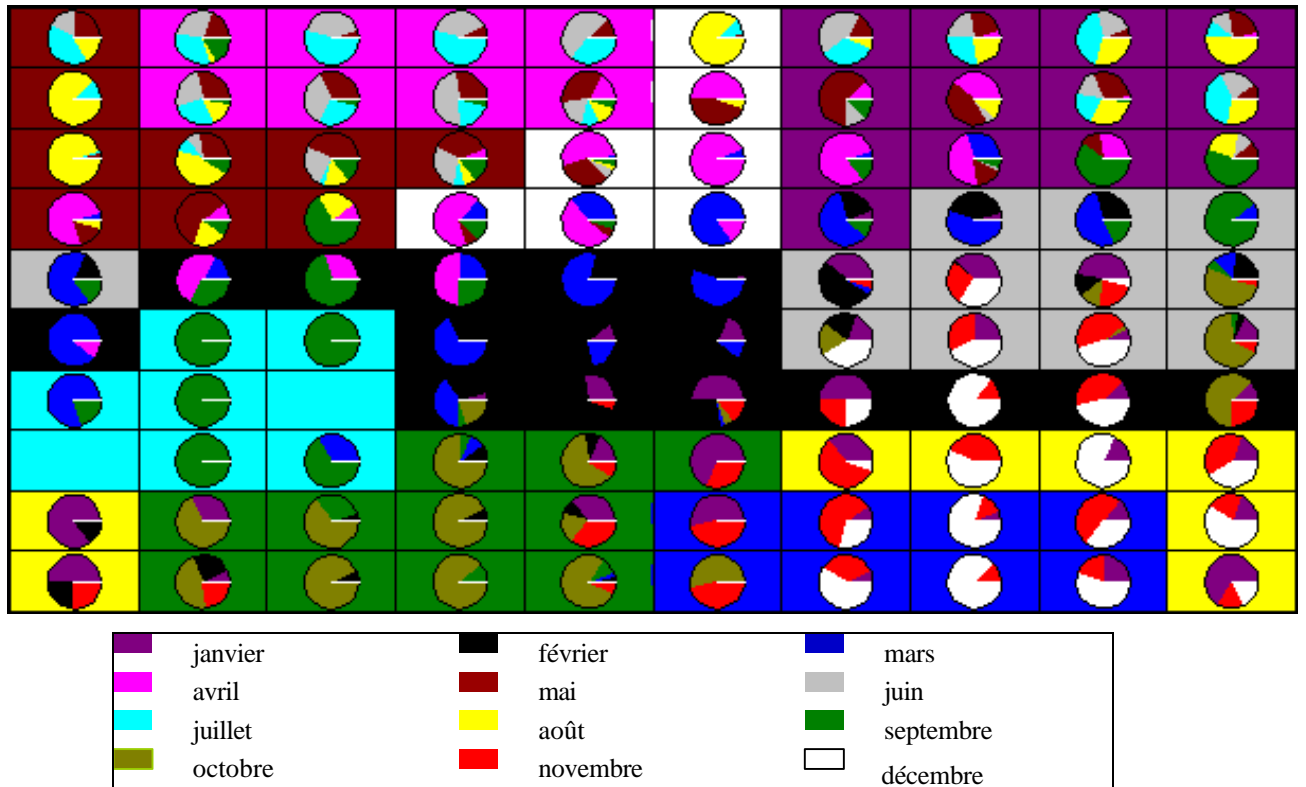


Figure 31 : Cartographie des mois pour l'exemple 3

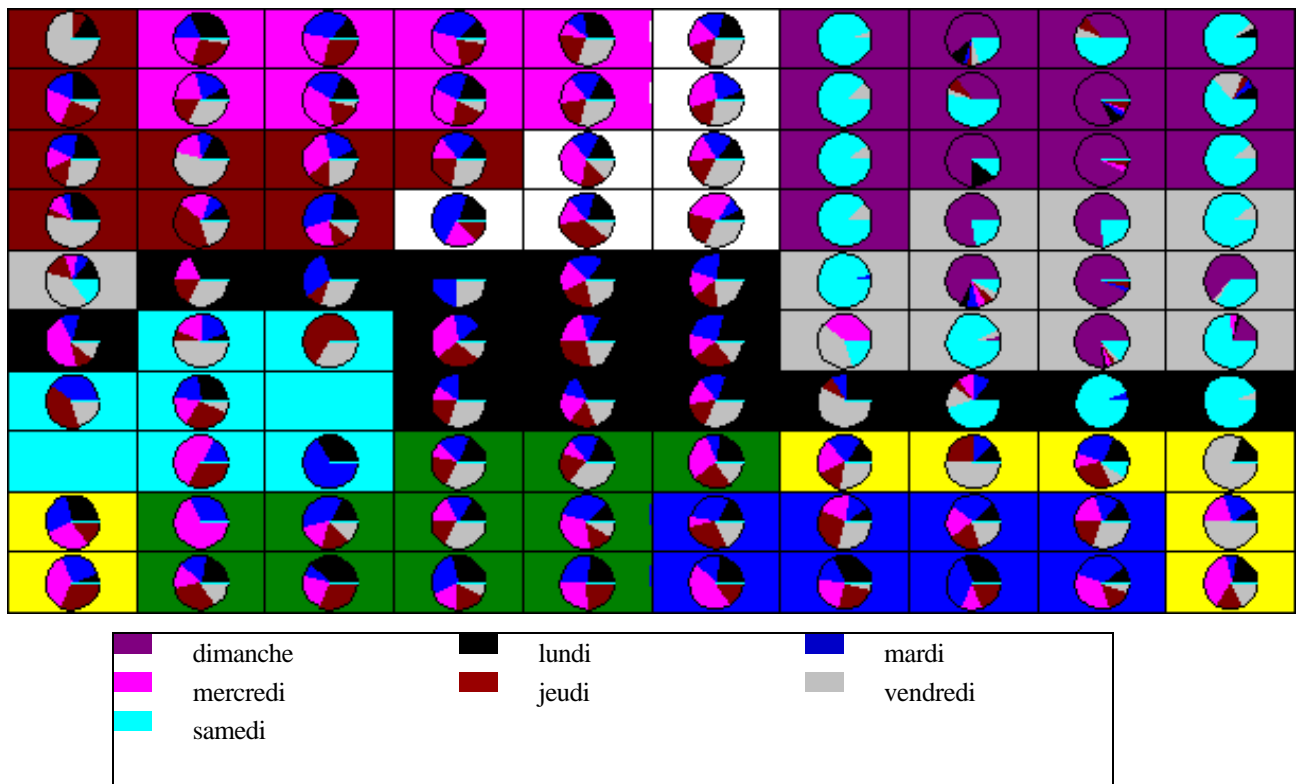


Figure 32 : Cartographie des jours de semaine pour l'exemple 3

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.

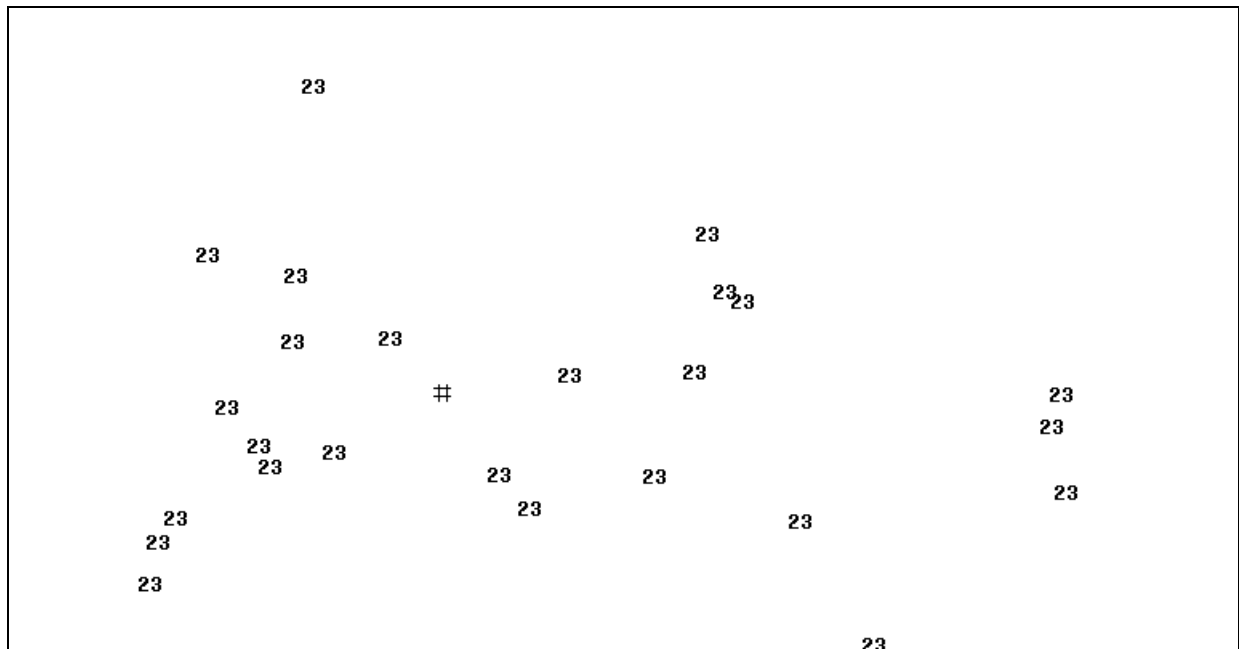


Figure 33 : Plan principal restreint à la classe associée à l'unité 26 ou plan tangent à la surface de la carte au point qui coïncide avec le vecteur code de cette même unité

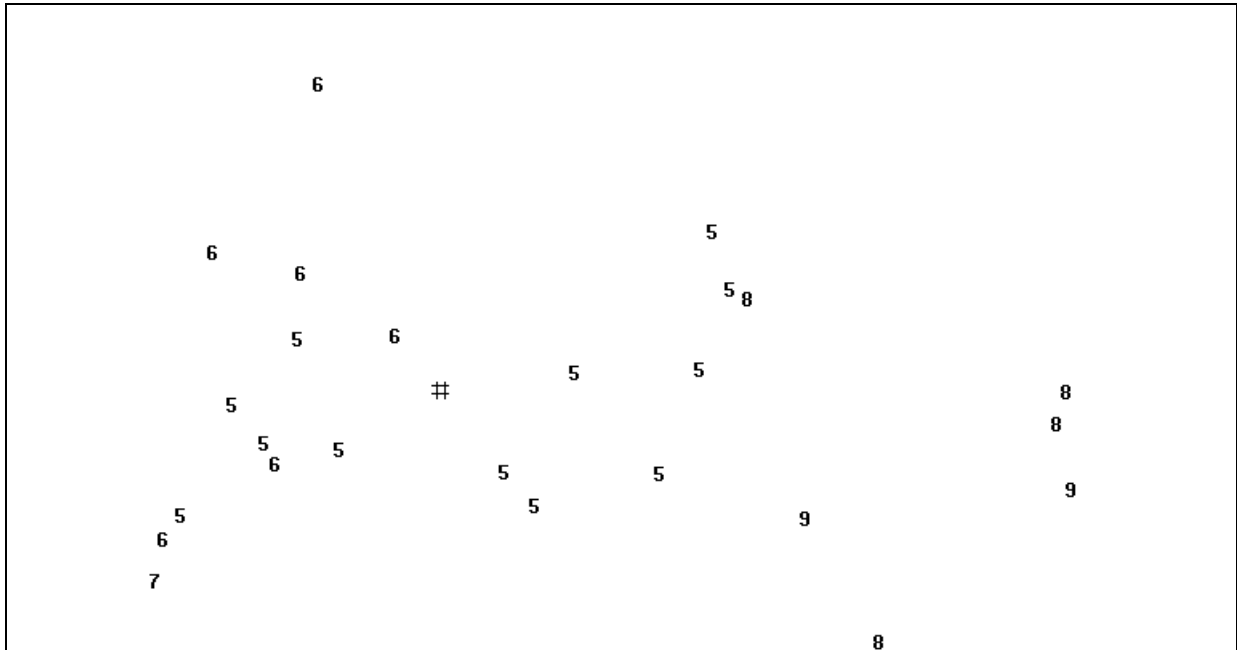


Figure 34 : Les points sont repérés par le mois. Le mois de juin est plutôt à gauche, mai au centre et août à droite

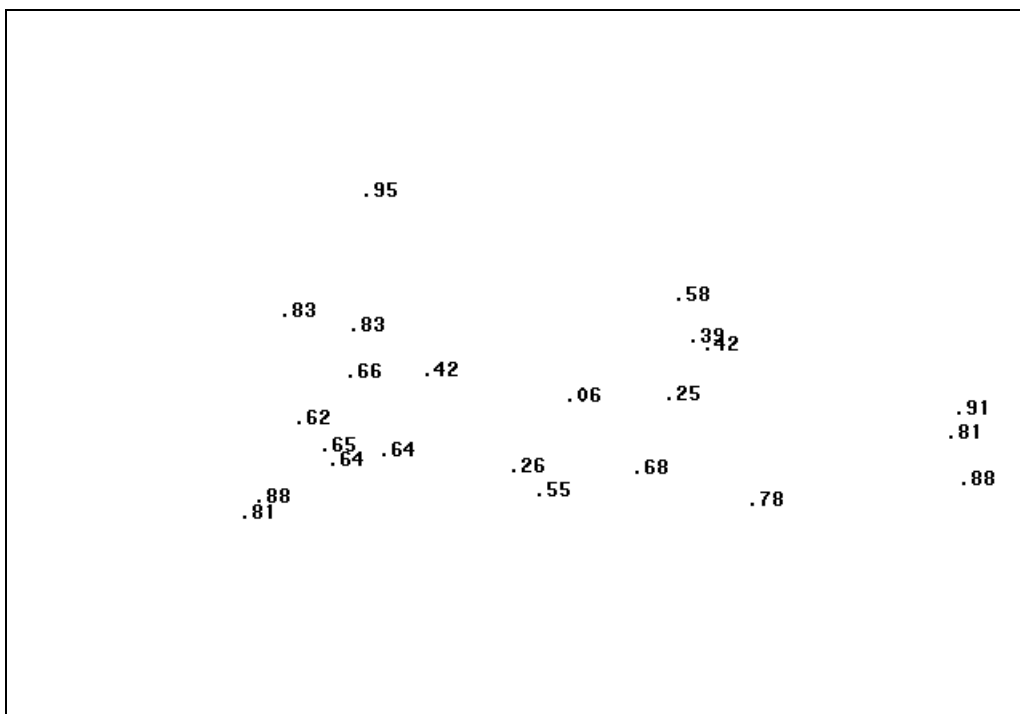


Figure 35 : Les points sont repérés par la qualité de leur projection. Les jours de mai sont mal représentés par ce plan.

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.

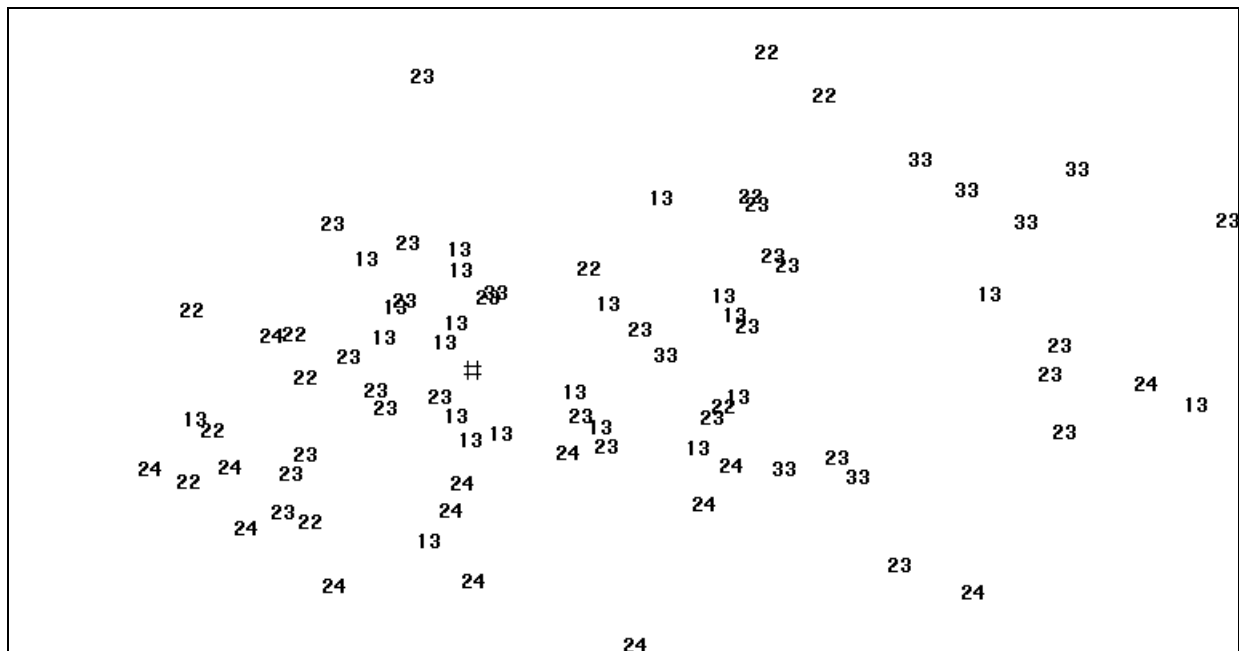


Figure 36 : Plan principal restreint à la classe associée à l'unité 23 sur lequel sont projetés les individus affectés aux unités voisines 13, 12, 24, 33

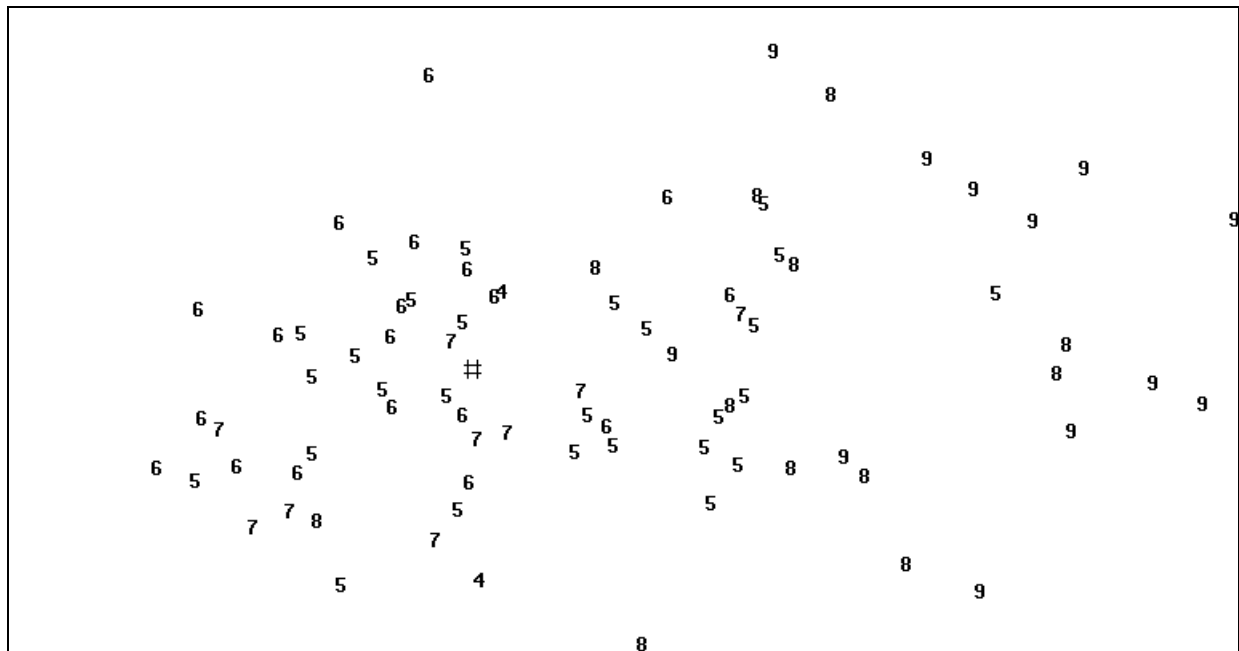


Figure 37: Les points sont repérés par le mois. Les individus de juin-juillet sont à gauche, de mai au centre et de août-septembre à droite

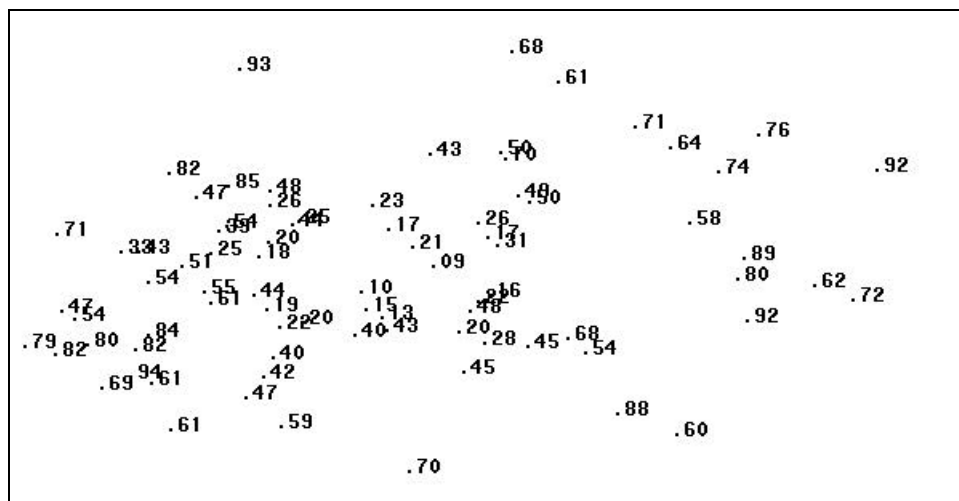


Figure 38: : Les points sont repérés par la qualité de leur projection. Les jours de mai sont mal représentés par ce plan, les autres le sont très bien.

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.

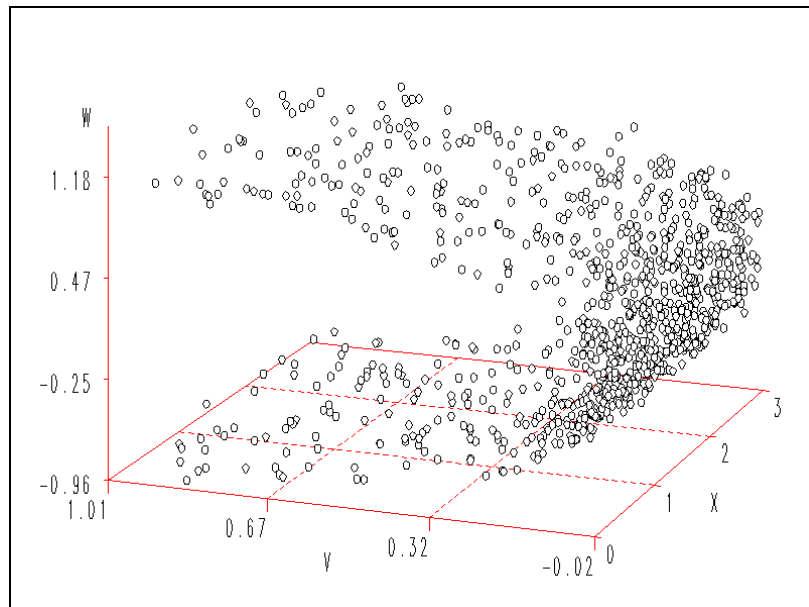


Figure 39 : Représentation du fer à cheval

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).

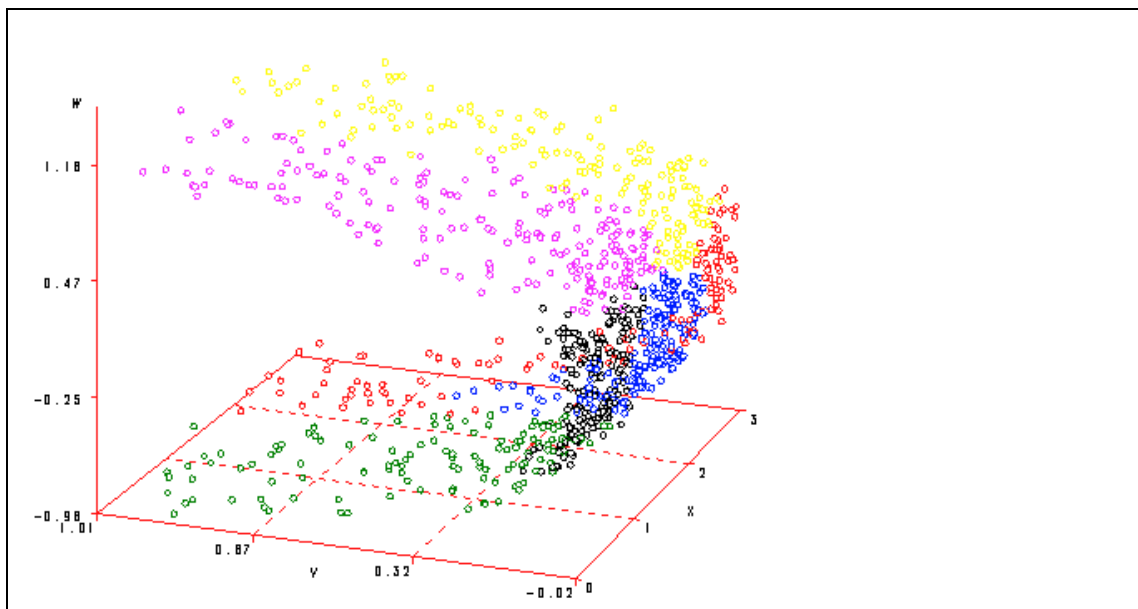


Figure 40 : Les couleurs qui représentent la classe de l'individu améliorent la lisibilité du graphique

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.

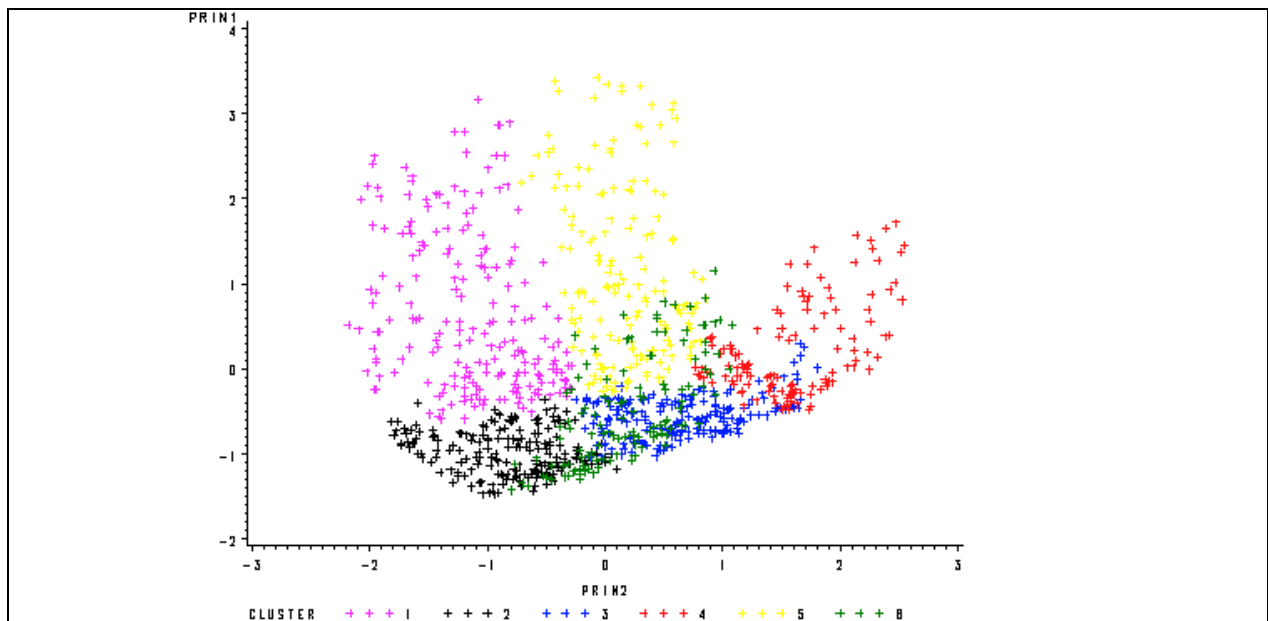


Figure 41: Plan principal (dim1, dim2)

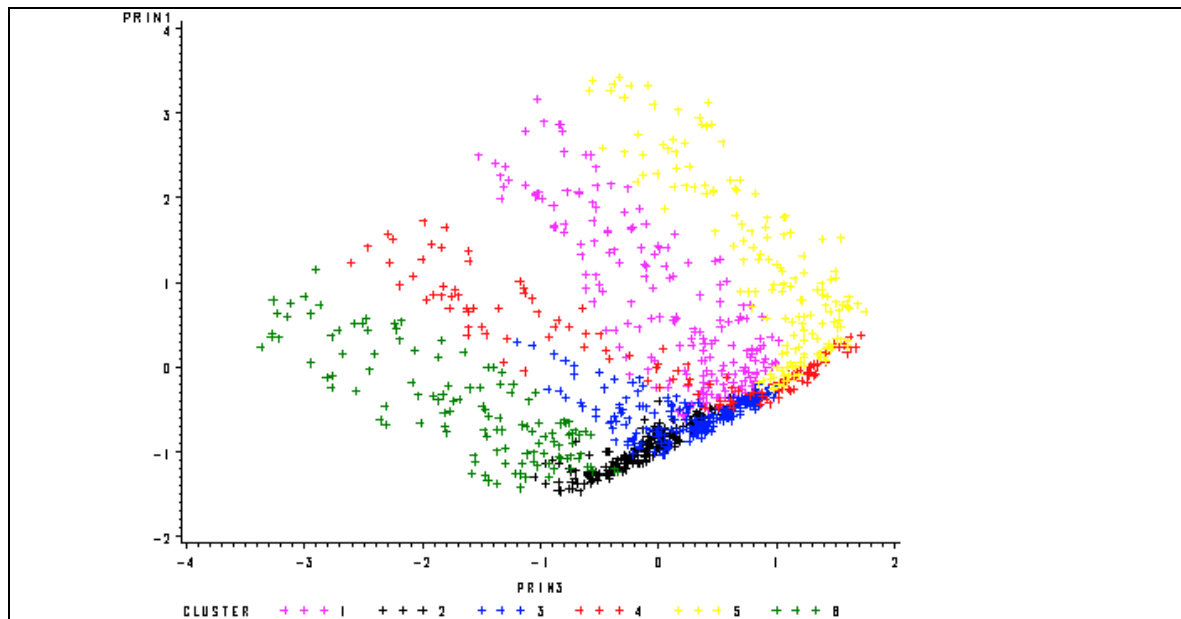


Figure 42: Plan secondaire (dim1, dim3)

- 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.

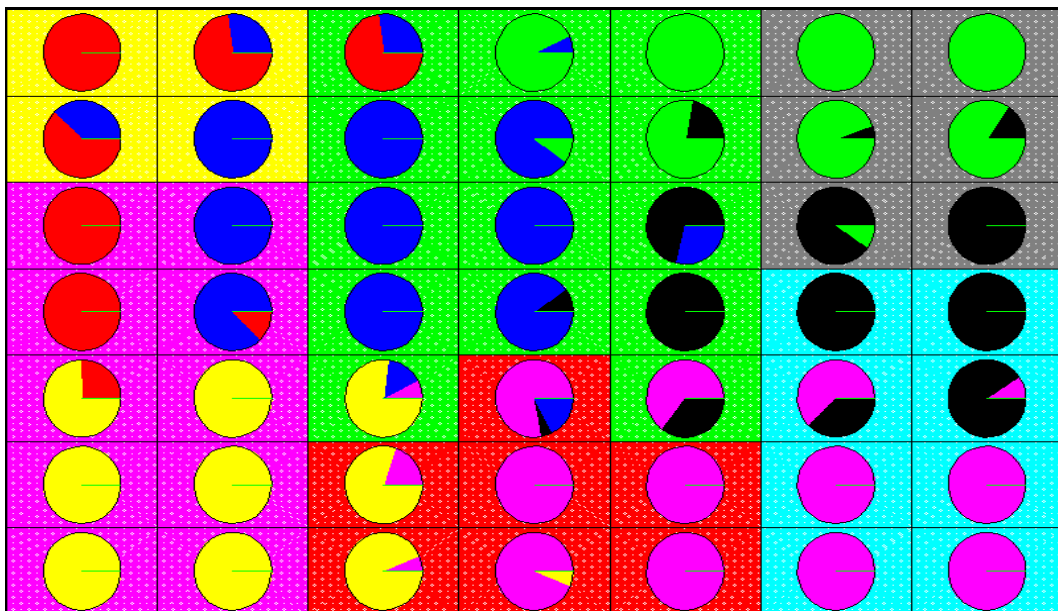


Figure 43: La carte de Kohonen qui ajuste le fer à cheval

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.

2^{ème} PARTIE

PREVISION A L'AIDE DES CARTES DE KOHONEN

Chap.4 **Problématique**

Chap.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.**

Chap.6 **Application de la MEP à l'exemple de *la prévision de la consommation électrique nationale polonaise***

Chap.7 **Quelques réflexions sur la méthode**

4 Préviation de courbes à l'aide des cartes de Kohonen : Problématique.

4.1 Le problème

Définition:

Supposons que l'on mesure le temps grâce à deux unités hiérarchiquement emboîtées (par exemple *le jour et l'heure* ou *la semaine et le jour* ou *l'heure et la seconde* etc.), l'unité lente – le jour, la semaine ou l'heure – étant multiple de l'unité rapide – respectivement l'heure, le jour ou la seconde. On qualifie de « basse fréquence, période, séquence, échelle lente » ce qui est relatif à l'unité lente et de « haute fréquence, instant, unité de temps, échelle rapide » ce qui est relatif à l'unité rapide.

Le problème consiste à élaborer une méthode de prévision qui prenne en compte la structure temporelle intrinsèque d'une série chronologique qui cumule deux

évolutions ayant chacune sa cadence propre. On se limite au cas où ces « cadences propres » correspondent à deux échelles de temps hiérarchiquement emboîtées, c'est-à-dire au cadre de la définition précédente.

De plus, on désire effectuer une prévision à long terme qui ait une qualité identique pour tout le futur. On ne se satisfait donc pas des méthodes qui ajustent bien les premières valeurs, mais qui perdent leur qualité d'estimation sur la partie à prévoir la plus lointaine. En particulier on écarte celles qui sont performantes au sens d'une erreur moyenne sur la durée à prévoir mais qui perdent leur fiabilité quand on considère le maximum de l'erreur.

4.2 Domaine d'application

La prévision des courbes de charge électrique – c'est-à-dire des 24 valeurs de puissance électrique utilisée pendant une journée et mesurée chaque heure – à partir d'un relevé de consommation électrique $\{Y_t\}$ est un problème important pour les compagnies d'électricité. En effet, sous évaluer le besoin d'énergie a pour conséquence l'obligation d'avoir recours à des centrales thermiques plus rapides à mettre en marche, mais également plus coûteuses que les centrales nucléaires. Au contraire, utiliser inutilement une centrale nucléaire représente une dépense inutile très importante. Pour pouvoir contrôler la production d'électricité en satisfaisant à la fois la demande intérieure et les contrats de vente, il est essentiel de pouvoir prévoir ces courbes de charge avec la même qualité tout au long de la journée. Ce problème a été, jusqu'à maintenant, abordé à partir de modèles récurrents qui ont pour entrées le passé $\{Y_t, t \leq t\}$ et pour sortie l'estimation de la première heure à venir $\{\tilde{Y}_{t+1}\}$ qui est, elle-même, présentée en entrée pour la prévision de l'heure suivante et ainsi de suite.

Ces courbes de charge dépendent du type de jour (dimanche, samedi, jour de semaine, jour férié, etc.), comme le montrent les études [107, 82, 84]). Ce type de jour peut être décrit à partir d'une série A_t de q variables qualitatives nominales $\{A_t^1, \dots, A_t^q\}$ qui accompagnent la

série chronologique $\{Y_t\}$, par exemple le jour de semaine (lundi, mardi, etc.) ou le mois. Dans la pratique, le plus souvent, le modèle utilisé est de type Arimax (présenté au chapitre 4.4.1). Il a l'inconvénient d'avoir la même forme, donc la même formule de prédiction tout au long de l'année. Pour ce genre de problème, on constate également le phénomène classique d'aplatissement des prévisions dans le cadre de la prédiction à long terme.

4.3 Exemple

Les données :

L'étude qui, à l'origine, a motivé notre travail est *la prédiction de la consommation nationale française d'électricité*. L'objectif était d'estimer la consommation demi horaire du lendemain, soit les 48 valeurs **du jour** suivant (et non les 48 valeurs suivantes). La consommation horaire a la propriété d'avoir une structure variable avec **le type** de jour (*jours fériés, samedis, dimanches, jours ouvrés, etc.*) qui se cumule avec les évolutions **du niveau** et **de l'amplitude** dues à la *température*. Les résultats de cette étude ont fait l'objet d'une publication présentée en annexe.

Pour des questions de confidentialité, il n'est pas possible de se servir de cette étude pour illustrer la présentation de la méthode. Néanmoins, nous disposons d'une base de données, concernant *la consommation électrique polonaise*, qui nous permet de faire une démonstration de l'utilisation de cette technique dans le cadre de données réelles. Malheureusement, nous ne possédons pas certaines connaissances telles que le relevé des températures qui permettraient de rendre cette prédiction plus performante. Nous invitons donc le lecteur à suivre la démarche sur des données de consommation électrique polonaise et de se référer à la prédiction de la consommation française d'électricité pour apprécier la qualité de l'estimation par cette méthode.

Premières remarques :

La base de données de *la consommation d'électricité nationale polonaise* est constituée d'un relevé horaire (Y_t) entre le 01 janvier 1986 et le 31 décembre 1993, une unité valant 20Mwh. Un premier aperçu général confirme ce que le bon sens suggère : on trouve une certaine répétition à 24 heures ou d'une semaine à l'autre comme l'illustre la figure 44 où est tracée *la consommation électrique horaire entre le dimanche 22 février 1987 et le 07 mars 1987*. En particulier la poussée de consommation domestique autour de 20 heures se retrouve quotidiennement ainsi qu'une petite pointe entre 11h00 et 13h00 qui coïncide avec la préparation du déjeuner. Sur les figures 45, 46 et 47, on constate aussi des variations de courbe de charge entre *un lundi de mars et un dimanche du même mois ou un lundi d'août*. En particulier il existe un changement de structure entre *un jour travaillé et un jour chômé* qui s'explique par le fait que la montée de consommation entre 5h00 et 8h00 – qui correspond à la mise en route des usines, du chauffage des lieux de travail et le début des activités professionnelles – est spécifique des jours ouvrés (ce phénomène est par exemple visible sur la figure 46). Outre des comportements de consommation différents à certaines heures, il existe des variations du niveau journalier. Un exemple entre des lundis d'hiver et d'été est fourni par les figures 46 et 47. Le chauffage est la première cause de ce phénomène auquel on peut ajouter parfois la baisse de l'activité professionnelle pendant les périodes de vacances.

Dans *la consommation nationale électrique française*, les jours à tarifs spéciaux (appelés *jours ejp* par l'EDF) et les jours fériés devaient aussi être abordés de façon particulière. En Pologne certains samedis sont ouvrés, la consommation d'un tel samedi peut donc se rapprocher plus du vendredi précédent que du samedi chômé de la semaine précédente, etc.

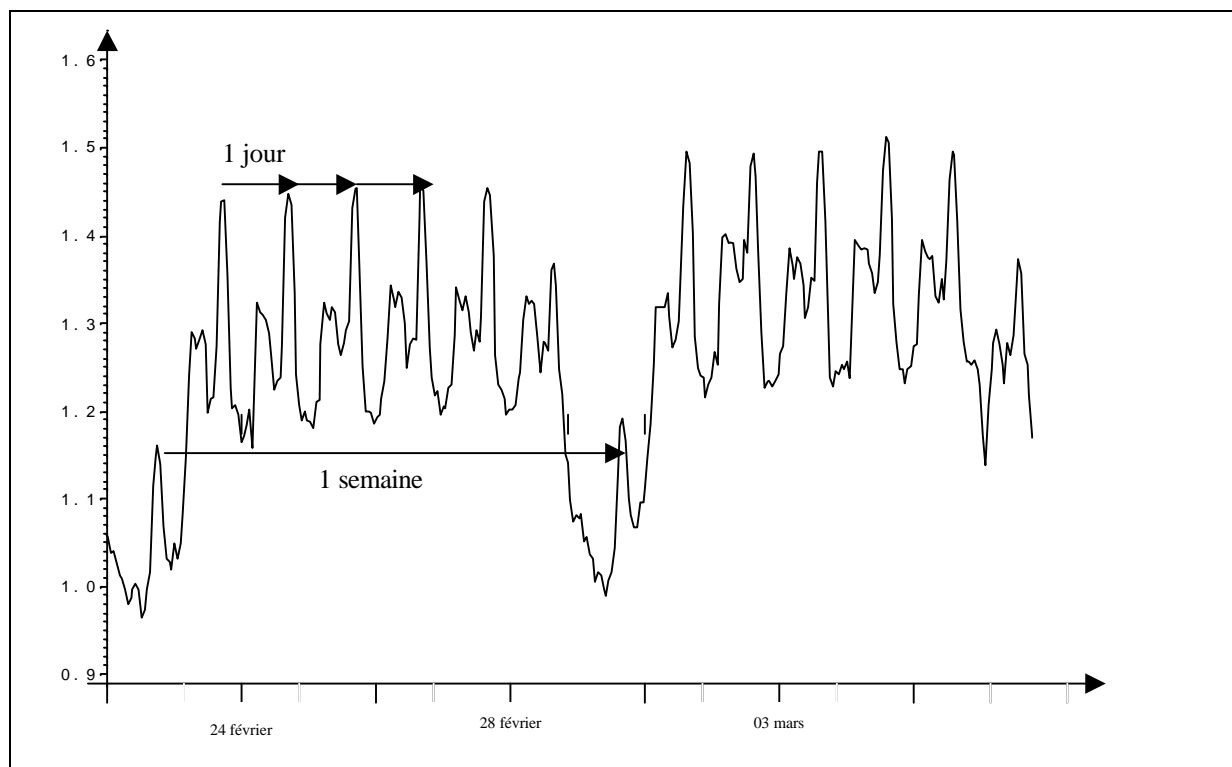


Figure 44: Deux semaines successives du dimanche 22 février 1987 au 07 mars 1987

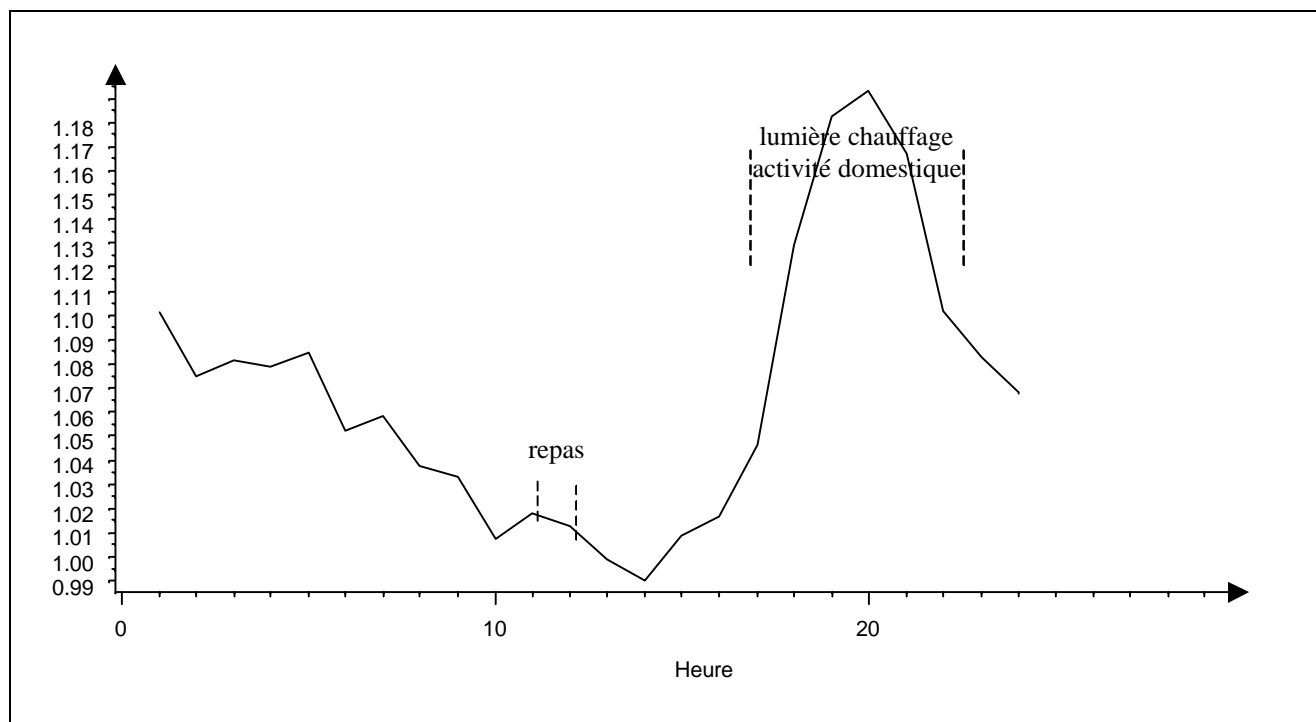


Figure 45: Courbe de consommation du dimanche 01 mars 1987

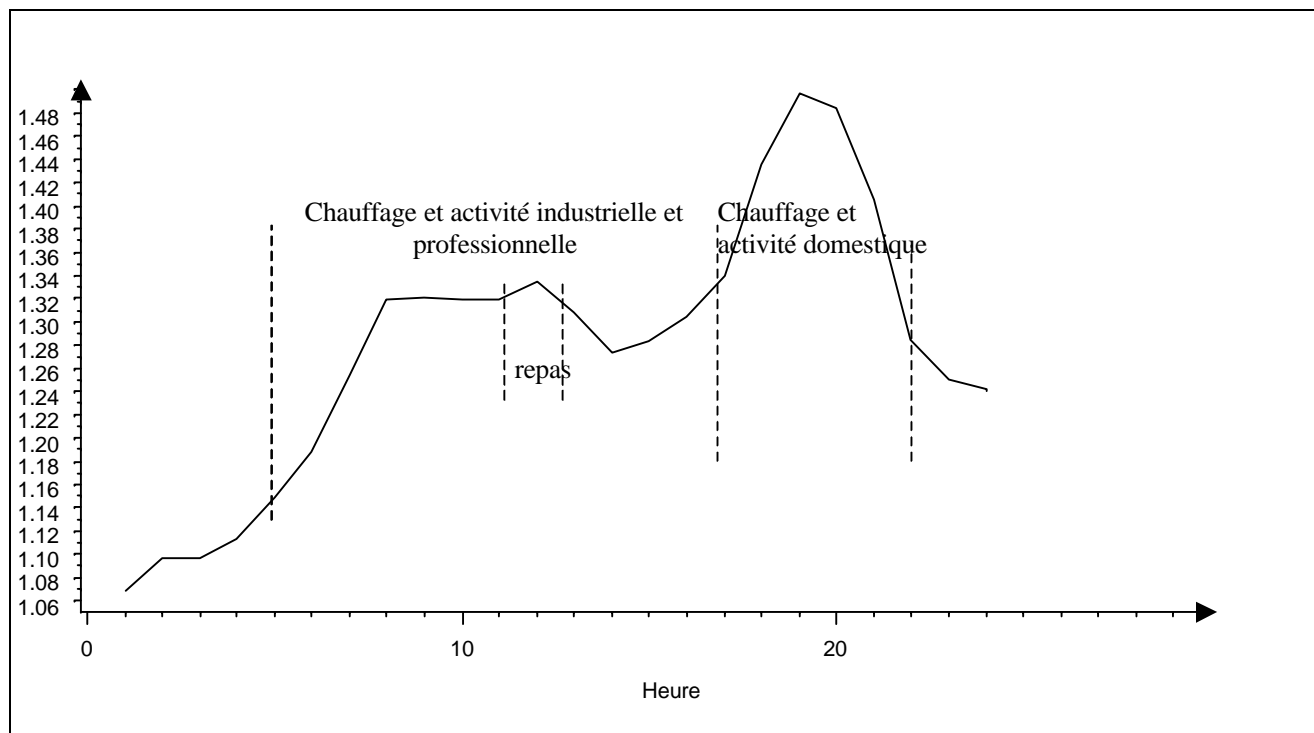


Figure 46: Courbe de consommation du lundi 02 mars 1987

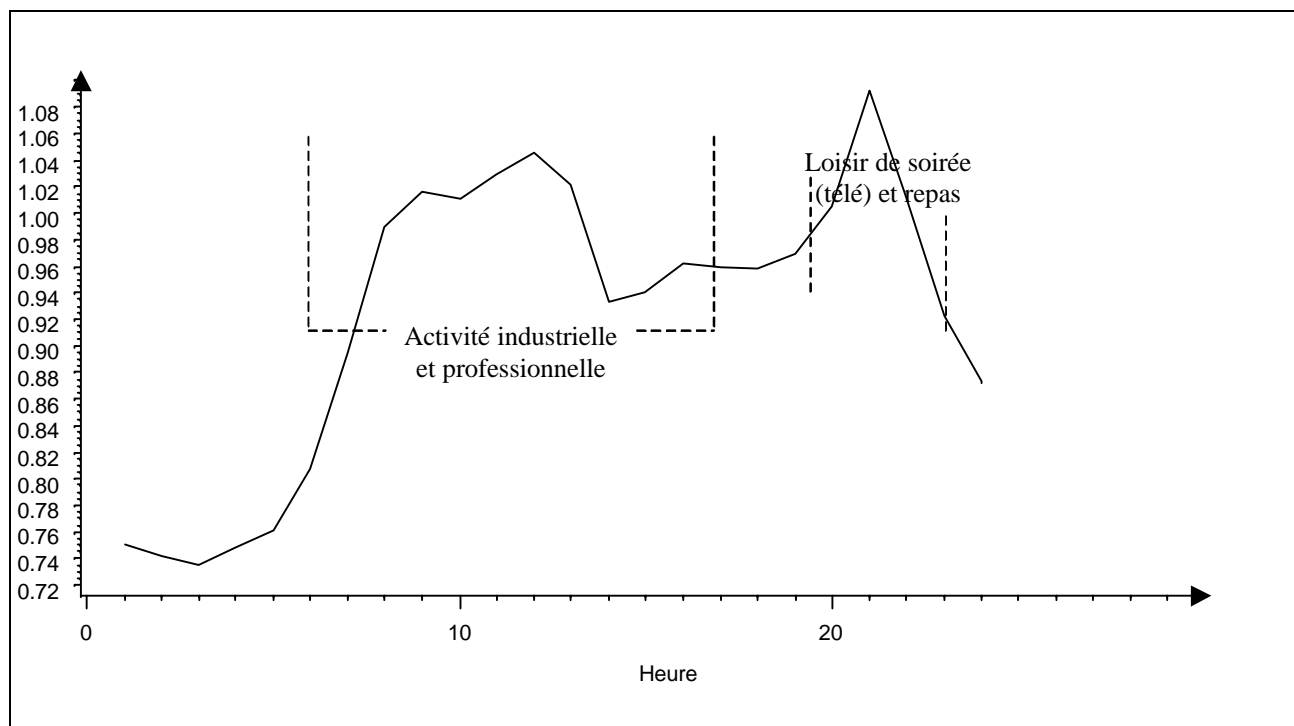


Figure 47 : Courbe de consommation du lundi 08 août 1988. Le niveau moyen est plus faible que celui de la figure 3

4.4 Les modèles classiques

La prédiction de la consommation électrique est le sujet de nombreuses études dans le monde. Parmi les techniques linéaires qui ont été appliquées ([60 et 96], on peut distinguer la famille des modèles Arma (ils sont au centre du modèle PRECOCE [4] appliqué par l'EDF et des études [11, 83, 48, 98]). De plus, depuis la publication de l'étude de Lapedes et Farber (1987)[67] et les résultats théoriques sur les qualités de prédiction à court terme des réseaux de neurones ([59, 54, 20, 21, 58, 87, 105, 102, 72]), de nombreuses expériences d'approximation des séries chronologiques par ces modèles non linéaires ont été faites avec des résultats satisfaisants. Weigend et Gerchenfeld (1994) [101] en fournit plusieurs références, par exemple [16,39,50,92,97] et dans le cadre plus précis des séries électriques [34, 18, 36, 91, 55, 84, 17]. Une comparaison des techniques neuronales est proposée dans diverses études telles que [53]. Le réseau le plus souvent utilisé est toutefois le perceptron multicouches [82, 32, 56, 103].

Dans ce paragraphe deux méthodes classiques seront présentées. L'Armax et le perceptron multicouches représenteront respectivement la famille linéaire et neuronale. (on peut citer aussi l'étude [2, 40] qui compare les modèles linéaires et non linéaires). Ils seront de plus confrontés au problème de la prédiction de la consommation électrique polonaise. On s'intéressera par exemple à la façon dont elles prennent en compte les particularités des données constatées précédemment, c'est-à-dire l'existence de structures journalières de consommation sensibles au type de jour.

4.4.1 Un modèle linéaire: L'Armax

L'Armax est un modèle de type linéaire et paramétrique qui est parmi les plus utilisés dans la prédiction des séries temporelles. Rappelons quelques définitions et notations de base pour présenter cette méthode de prédiction dont les propriétés sont détaillées dans de nombreux

ouvrages. Parmi les classiques, on peut citer celui de Brockwell et Davis (1991) [15], ainsi que ceux de Box et de Jenkins (Box et Jenkins 1976 [14]) et Ansley (Ansley, (1979) [1]) pour les techniques de prévisions.

Définition de la stationnarité:

Un processus stochastique $\{Y_t\}_{t \geq 0}$ est :

- *stationnaire au sens strict* si les vecteurs (Y_1, \dots, Y_p) et $(Y_{1+k}, \dots, Y_{p+k})$ ont la même distribution pour tout entier p et k .

- *stationnaire au sens large* (ou *stationnaire du second ordre*) si $E(Y_t^2) < \infty$, $E(Y_t)$ est constant et si $Cov(Y_t, Y_{t+k})$ ne dépend que k . Dans la suite, *stationnaire* désignera un processus stationnaire au sens large.

- La *différentiation d'ordre d* d'un processus $\{Y_t\}_{t \geq 0}$ est l'opération qui produit le nouveau processus $\{(Y_t - Y_{t-d})\}_{t \geq d}$, noté $\{\Delta^d(Y_t)\}$. On l'utilise en particulier, lorsque ce dernier est stationnaire, on dit alors qu'on a stationnarisé $\{Y_t\}$. Dans le cadre de la prévision des données polonaises, les répétitions d'un jour à l'autre ou à une semaine d'intervalle décrites à la section 4.3.2 laissent penser qu'une différenciation d'ordre 24 et 168 (une semaine) peuvent être nécessaires pour obtenir un processus stationnaire.

Définition d'un modèle Armax(p,q):

Un processus $\{Y_t\}$ stationnaire au sens large suit un **modèle Armax(p,q)** où $\{A_i\}_{1 \leq i \leq m}$ sont m variables exogènes, s'il existe un bruit blanc (e_t), deux polynômes f et q qui s'écrivent sous la forme $f(u) = 1 + f_1u + \dots + f_pu^p$ et $q(u) = 1 + q_1u + \dots + q_qu^q$ avec des coefficients a_1, \dots, a_m tels que pour tout t :

$$Y_t + f_1Y_{t-1} + f_2Y_{t-2} + \dots + f_pY_{t-p} = e_t + q_1e_{t-1} + \dots + q_qe_{t-q} + \dots + a_1A_t^1 + \dots + a_mA_t^m$$

où a_1, \dots, a_m sont les paramètres d'ajustement des variables exogènes A^1, \dots, A^m .

On appelle *opérateur retard d'ordre d* l'application B telle que $B(Y_t) = Y_{t-d}$.

Un processus $\{Y_t\}$ suit un modèle **Armax(p,q)** de moyenne m si $\{Y_t - m\}$ suit un modèle **Armax(p,q)**.

Un processus $\{Y_t\}$ suit un modèle **Arimax(p,d,q)** si $\{\Delta^d(Y_t)\}$ suit un modèle **Armax(p,q)**.

Les modèles Armax sont très utilisés pour les problèmes de prévision tels que celui des données polonaises de consommation électrique. Dans cet exemple, une différenciation d'ordre 168 (une semaine) ou un opérateur retard du même ordre (Y_{t-168}) peuvent permettre de traduire l'écart entre un jour ouvré et un dimanche. Par contre, la différence entre un jour ouvré et un jour férié n'est pas un phénomène saisonnier, elle ne peut être prise en compte dans un modèle Armax ni par une différenciation ni par un opérateur retard. Dans un modèle Armax, une indicatrice de jour férié introduite en variable exogène pourrait compenser un changement de niveau commun aux 24 heures. Mais ici, on est en présence d'un changement de structure, soit 24 changements de niveaux différents, ce qui fait un nombre de paramètres très important à estimer. Mal prendre en compte un jour parce qu'il est férié est d'autant plus gênant que l'on connaît sa caractéristique à l'avance. De plus cette remarque s'étend aux jours de changement d'horaires (qui décalent une partie de la consommation d'une heure) ou à l'alternance entre les samedis ouvrés et travaillés (dans ce cas, la différenciation d'ordre 168 ne s'applique plus). Elle concerne aussi les jours *ejp*, à tarification particulière réservée aux entreprises, qui ne vont pas avoir le même impact sur la consommation pour les heures travaillées et la soirée ou la nuit.

On constate donc que l'on est confronté à de nombreux changements de structure qui pénalisent la performance d'un modèle Armax. Il est de plus à souligner que ces changements correspondent à des phénomènes journaliers. Une approche différente de l'Armax consisterait donc à prendre en compte la notion de type de jour.

4.4.2 Un modèle neuronal: Le perceptron multicouches

Dans la famille des réseaux de neurones, le perceptron multicouches fournit un modèle non linéaire de prévision. Il peut en particulier généraliser les modèles Armax au cadre non linéaire et on note alors *Narmax* le modèle correspondant. On va voir dans cette section comment utiliser un tel modèle pour l'adapter au problème des données électriques polonaises.

- Introduction:

Les modèles connexionistes qui sont inspirés des connaissances biologiques sur le fonctionnement du cerveau sont apparus dès les années 40. W. Mc Culloch et W.Pitts (Mc

Culloch et Pitts, (1943) [78]) ont introduit le neurone formel défini à partir de son potentiel et d'une fonction d'activation. Le premier est une somme pondérée des entrées, la seconde est une fonction y des entrées de type sigmoïde comme par exemple $y(x) = \frac{e^x - 1}{e^x + 1}$.

Rosenblatt a associé un algorithme d'apprentissage au perceptron simple qui est le nom donné à un neurone formel (Rosenblatt, 1962 [86]). Néanmoins, Minsky et Paper montrèrent rapidement (Minsky et Paper, (1969) [80]) les limites de ce perceptron simple incapable de résoudre des problèmes de base – en particulier séparer des ensembles non linéairement séparables. Le perceptron multicouches qui désigne une architecture en réseau constitué de plusieurs couches de perceptrons simples est tout de suite apparu comme une solution naturelle. Mais l'algorithme d'apprentissage de Rosenblatt ne s'étendait pas au perceptron multicouches.

Ce n'est que dans les années 80 que le perceptron multicouches est devenu populaire, notamment grâce aux travaux parallèles de l'équipe de Lecun et al. (Lecun et al., 1985 [70]) en France et celle de Rumelhart et al. (Rumelhart et al. 1986 [87]) aux Etats-Unis sur l'algorithme de rétropropagation du gradient (des centres de gradient du premier ordre) dont est issu un algorithme d'apprentissage du perceptron multicouches par minimisation d'une fonction d'erreur.

L'utilisation du perceptron pour prévoir les séries stationnaires a été justifiée par ses propriétés d'approximation démontrées par Cybenko (1989) [31] ou Hornik, Stinchcombe and White (1989) [57], Attali et Pages [3], Barron [7]. En effet, un perceptron multicouches avec une couche cachée pourvue d'un nombre suffisant de neurones peut approximer n'importe quelle fonction continue sur un compact de \mathbf{R}^p , donc en particulier les séries stationnaires.

1) Modèle autorégressif NAR (Neuronal autorégressive) avec un perceptron multicouches à 1 couche cachée constituée de n unités et pour fonction de transfert la sigmoïde y :

Pour réaliser un tel modèle, on utilise une architecture de réseau du type de la figure 48. Le problème du choix du nombre d'unités et de couches a fait l'objet de nombreux travaux, on peut citer Weigend et al. (1990)[100] , Le Cun et al. (1990) [69] et Cottrell et al. (1995) [23 et 24].

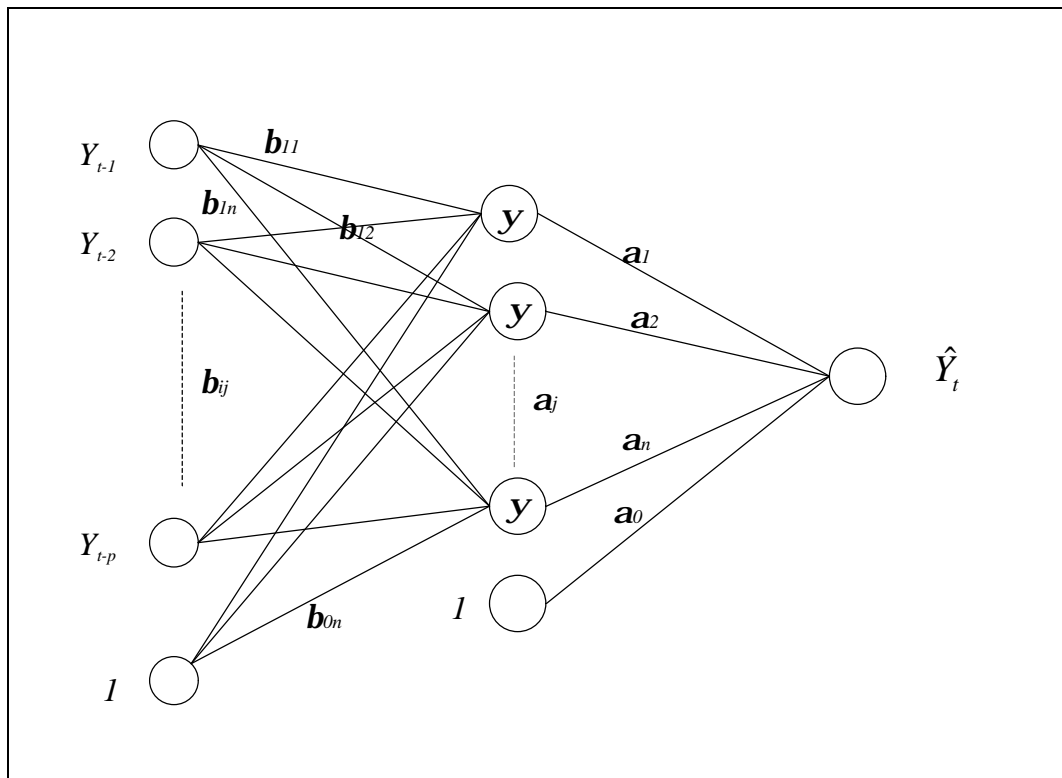


Figure 48 : Perceptron multicouches pour un NAR

$$\text{On a } Y_t = \sum_{j=1}^n a_j y \left(\sum_{i=1}^p b_{ij} Y_{t-i} + b_{0i} \right) + a_0 + e_t \quad (1)$$

où les e_t sont des variables indépendantes équidistribuées de loi $N(0, \mathbf{s}^2)$, $(a_k)_{0 \leq k \leq n}$ et $(b_{ij})_{0 \leq i \leq p, 1 \leq j \leq n}$ sont les paramètres du modèle (et les poids des connexions).

Estimer un tel modèle nécessite d'estimer $n(p+1)+1$ paramètres.

Remarque:

Dans la pratique, des méthodes mixtes ont été essayées. Elles consistent à réaliser une pré-classification et à construire un perceptron pour chaque classe trouvée. De telles méthodes sont présentées par exemple dans Garcia-Tejedor et al. (1994) [49], Muller et al. (1994) [81], Peng T.M. et al. (1992) [84], Park D.C. et al. (1991) [82], Lee K.Y. et al. (1992) [71]. Leurs limites sont développées dans la suite.

2) Modèle NARMAX(p,q) avec variables exogènes :

Le modèle de type perceptron qui généralise l'Armax(p,q) au cadre non linéaire peut s'écrire sous la forme (2),

$$Y_t = \sum_{j=1}^n a_j y \left(\sum_{i=1}^p b_{ij} Y_{t-i} + \sum_{l=1}^m b'_{lj} A_t^l + \sum_{k=1}^q b''_{kj} e_{t-k} + b_{0j} \right) + a_0 + e_t \quad (2)$$

où $(a_k)_{0 \leq k \leq n}$, $(b_{ij})_{0 \leq i \leq p \text{ et } 1 \leq j \leq n}$, $(b'_{lj})_{1 \leq l \leq m \text{ et } 0 \leq j \leq n}$ et $(b''_{kj})_{1 \leq k \leq q \text{ et } 1 \leq j \leq n}$, sont les paramètres du modèle et $A=(A^l)_{1 \leq l \leq m}$ le vecteur des variables exogènes et e_t des variables indépendantes équidistribuées de loi $N(0, \sigma^2)$.

Estimer un tel modèle (schématisé par la figure 49) nécessite ainsi d'estimer $n(p+m+q+1)+1$ paramètres et l'architecture devient dans ce cas celle de la figure suivante :

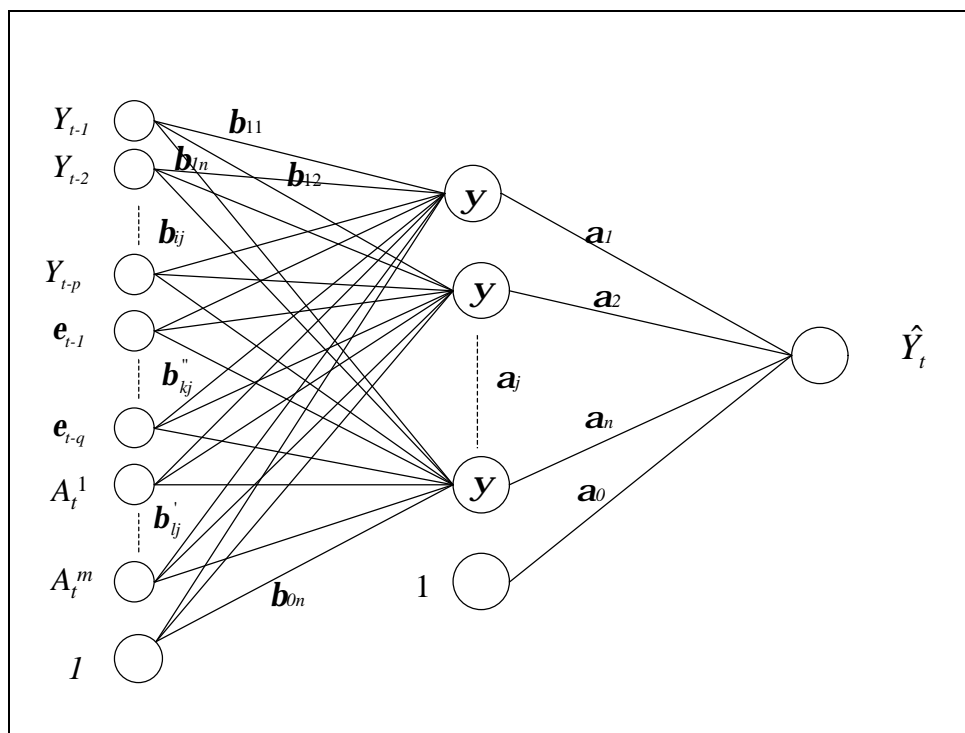


Figure 49 : Perceptron multicouches pour un NARMAX

3) Estimation à plusieurs pas avec un perceptron multicouches :

Pour estimer Y_{t+1} , l'idée qui vient naturellement à l'esprit est de remplacer Y_t par \hat{Y}_t dans l'une des formules (1) et (2) pour respectivement un NAR et un NARMAX comme le suggère la figure suivante. Néanmoins, cette technique régulièrement utilisée dans le cadre linéaire est ici impossible pour des raisons développées ci-après, liées à la fonction de récurrence (décrits également dans [106]).

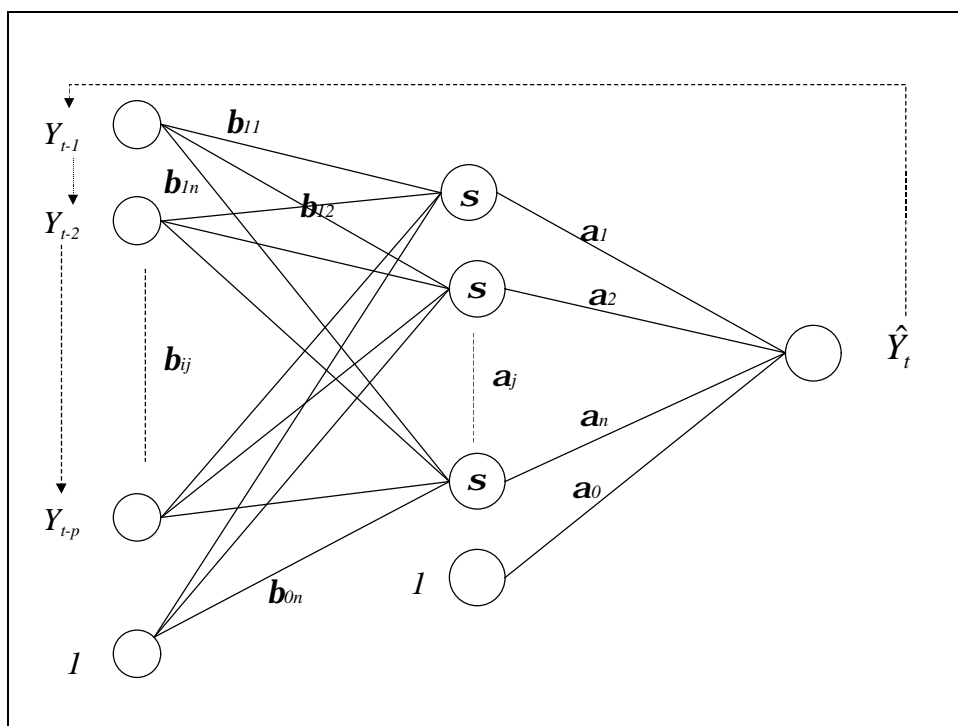


Figure 50 : Perceptron multicouches susceptible de modéliser un modèle récurrent avec un NAR

- Considérons l'exemple simple d'un perceptron à 2 unités cachées qui modélise un NAR pour prévoir deux unités de temps successives :

On se place dans le cas où l'on cherche à prévoir la série temporelle $\{Y_t\}$ par le modèle fourni par le perceptron de la figure 51 qui ne prend en compte que le retard d'ordre 1. on a donc:

$$Y_t = f(Y_{t-1}) + e_t$$

$$\hat{Y}_t = f(Y_{t-1})$$

$$\text{avec } f(x) = 9y\left(\frac{3}{2}x\right) - 5y\left(\frac{3}{2}x\right)$$

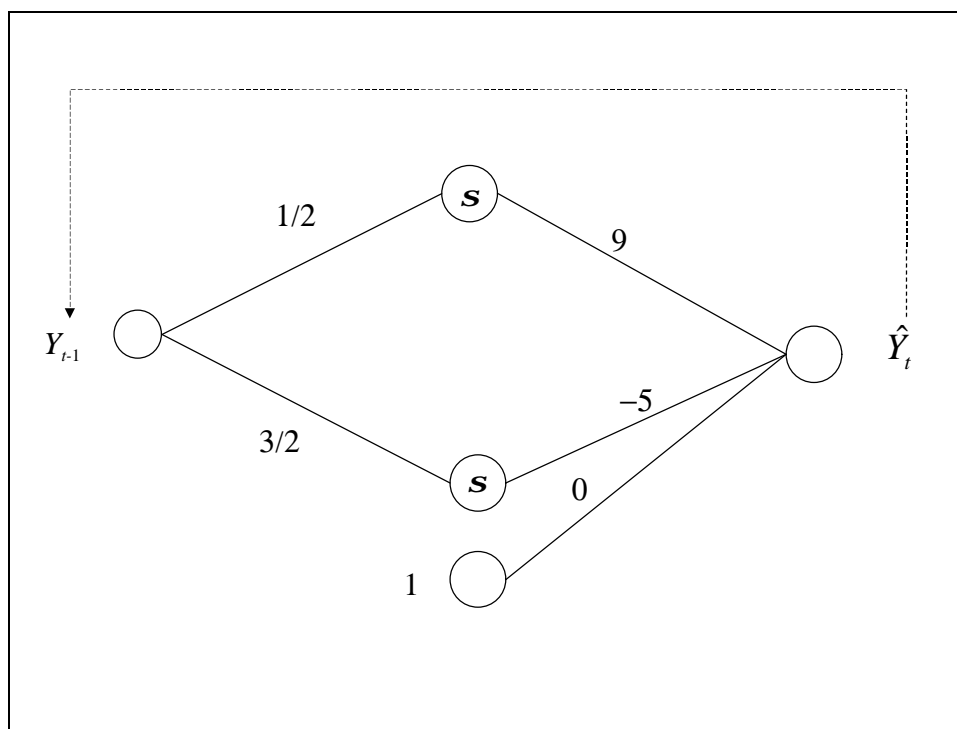


Figure 51 : Exemple d'un perceptron multicouches utilisé pour faire une récurrence

Etudions la suite (x_t) suivante:

$$x_{t+1} = f(x_t)$$

$$\text{où } f(x) = 9y\left(\frac{3}{2}x\right) - 5y\left(\frac{3}{2}x\right)$$

La fonction f a 3 points fixes répulsifs en 0, 2.1, -2.1 et deux points fixes attractifs en 3,44 et -3,44 (figure 52).

Le comportement de la suite suivant la valeur de l'initialisation x_1 est:

$$\text{Si } x_1 \geq 2.1 \quad \lim_{n \rightarrow \infty} x_n = 3,44 \text{ (figure 55).}$$

$$\text{Si } x_1 \leq -2.1 \quad \lim_{n \rightarrow \infty} x_n = -3,44 \text{ (figure 56).}$$

$$\text{Si } -2.1 < x_1 < 2.1 \quad \forall t \hat{I} N, x_t \in]-2.1 ; 2.1[\text{ (figures 53 et 54).}$$

Les valeurs oscillent. On peut noter par exemple que $f(-x) = -f(x)$ et donc pour les valeurs de x telles que $f(x) = -x$ (c'est-à-dire -0.8 et 0.8) vérifient $f(f(x)) = f(-x) = -f(x) = -(-x) = x$ et sont donc des valeurs d'adhérence de la suite.

Cas particulier:

Pour $x_1=0.001$ $x_6>0.6$
 Pour $x_1=-0.001$ $x_6<-0.6$

Quand on confronte les deux suites $(x_t)_{0.001}$ et $(x_t)_{-0.001}$ pour des initialisations respectives de 0.001 et -0.001 , on constate qu'un écart de 0.002 à l'initialisation peut générer en 6 itérations un écart 600 fois plus grand. Cette suite est donc très sensible à l'initialisation.

Une solution pour prévoir I valeurs successives avec un Narmax est de construire I perceptrons multicouches du type de celui de la figure 50. Mais ceci oblige à estimer $I(n(p+n+q+1)+1)$ paramètres ce qui devient rapidement très important. Dans le cadre particulier des données polonaises, I vaut 24 ce qui rend le nombre de paramètres à estimer beaucoup trop grand pour faire du perceptron multicouches un modèle de prévision à plusieurs pas adéquat.

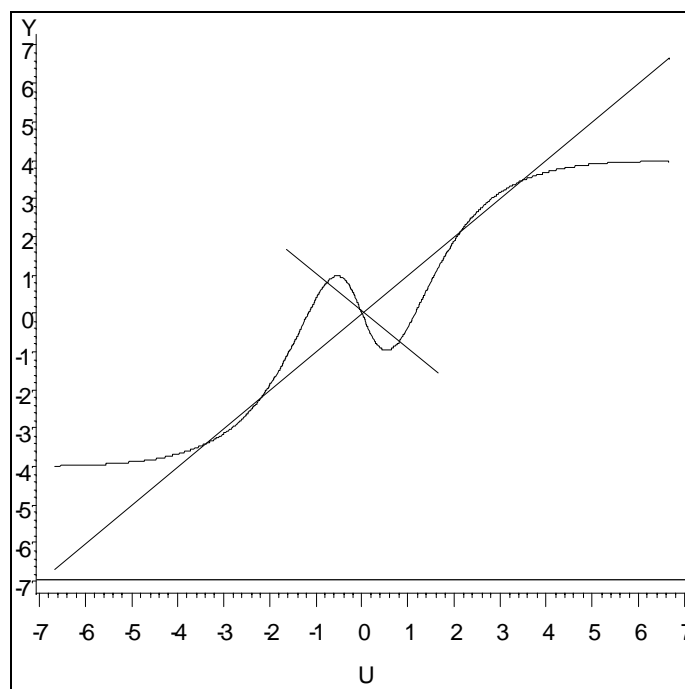


Figure 52 : Superposition de la fonction d'activation et des droites $y=x$ et $y=-x$

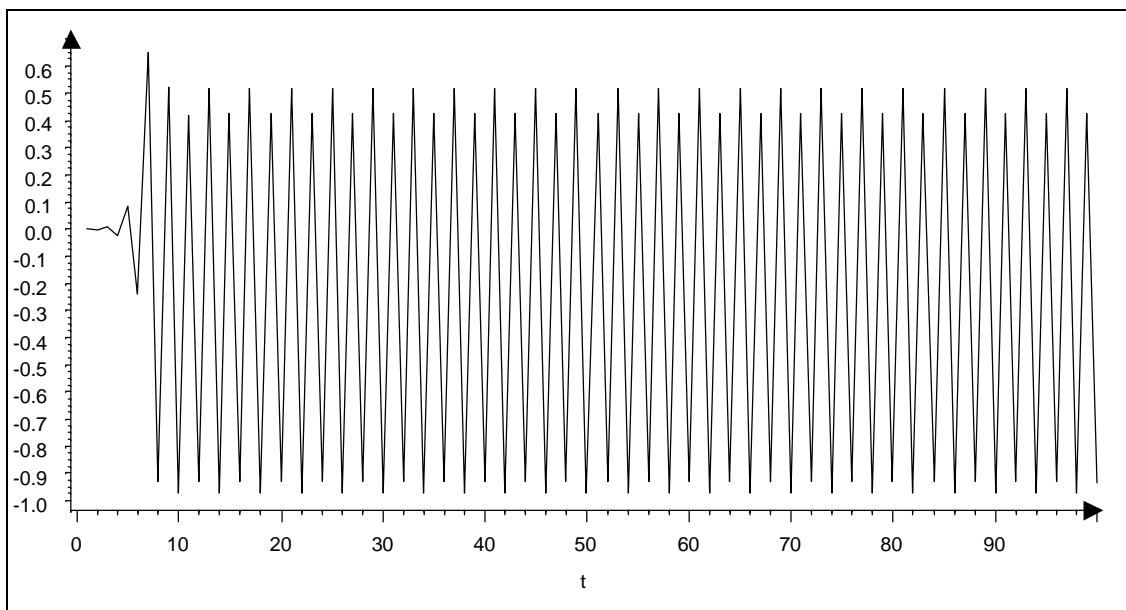


Figure 53 : La suite (x_t) pour une initialisation $x_1=0.001$

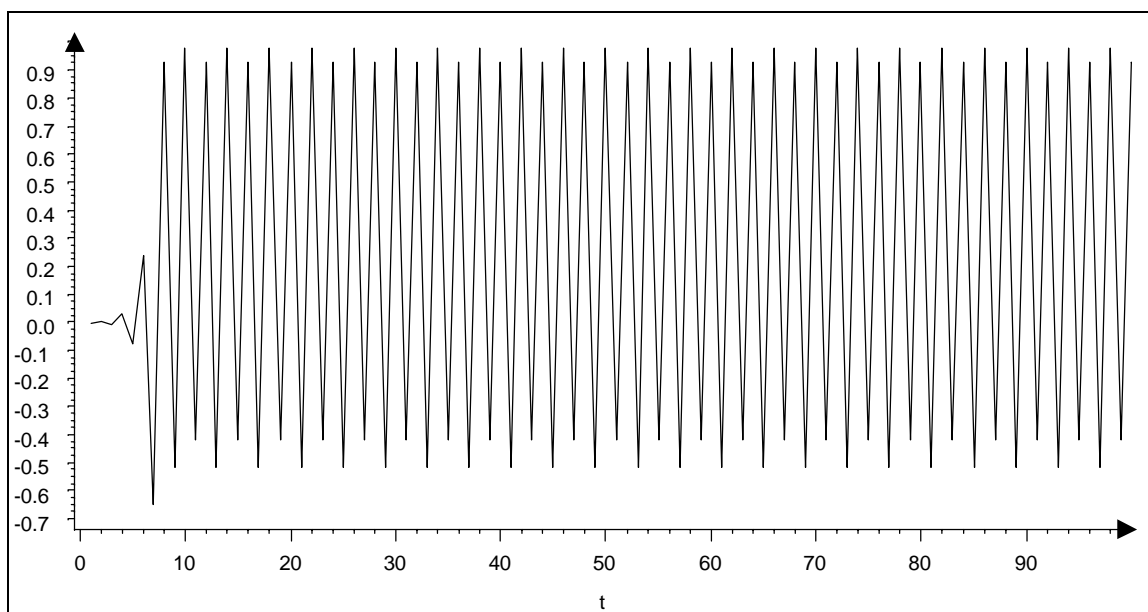


Figure 54 : La suite (x_t) pour une initialisation $x_1=-0.001$

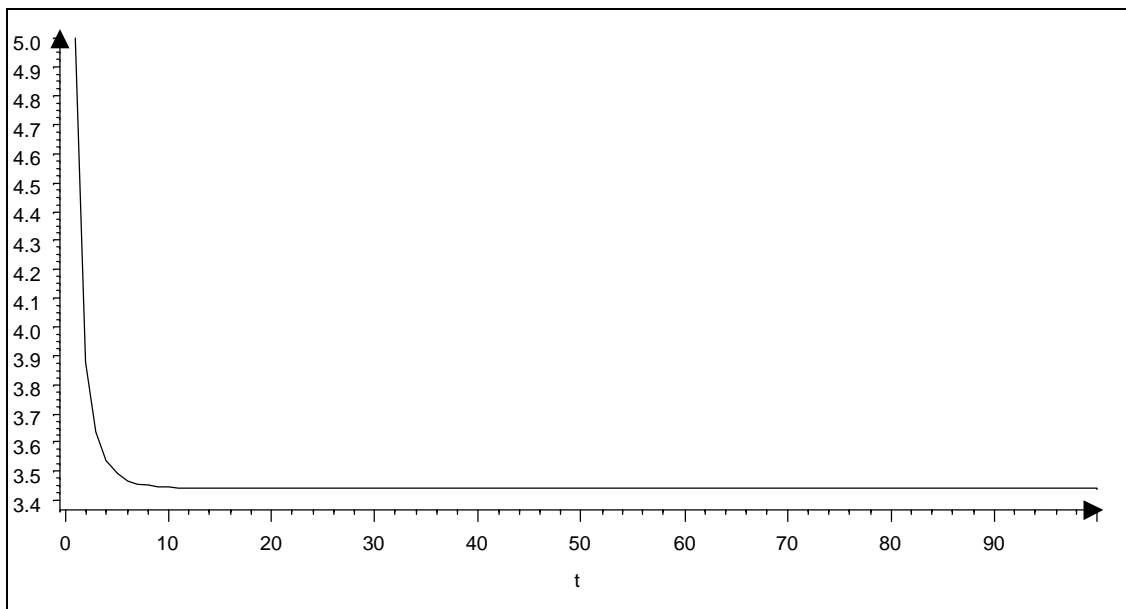


Figure 55 : La suite (x_t) pour une initialisation $x_1=5$

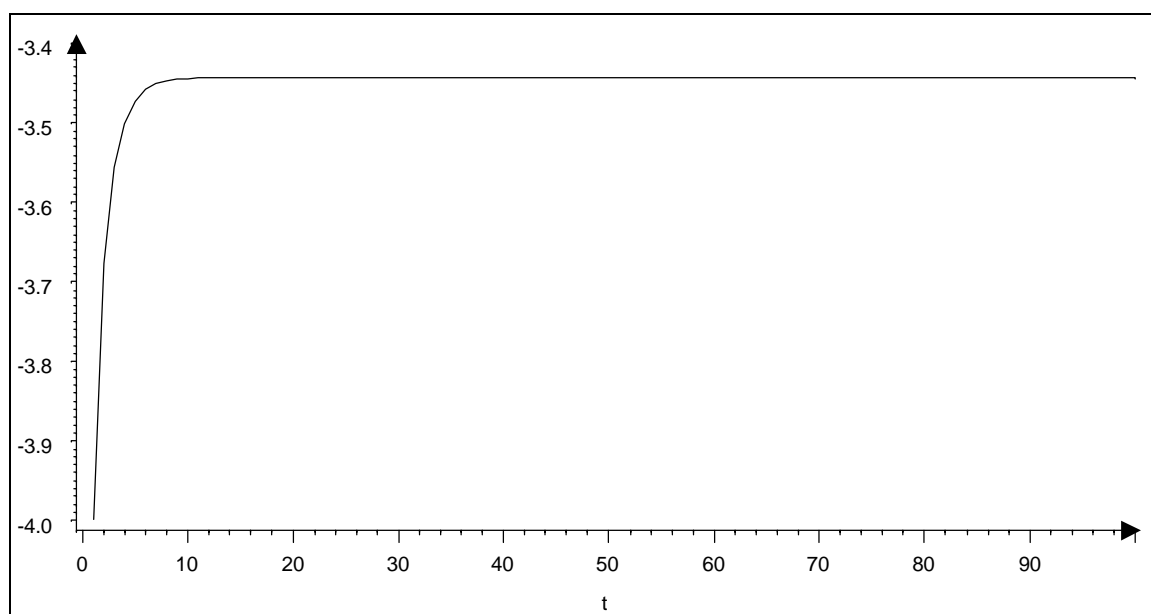


Figure 56 : La suite (x_t) pour une initialisation $x_1=-4$

4.4.3 Conclusion

Les difficultés d'adaptation de l'Armax et du perceptron multicouches au problème de la prévision de la consommation électrique proviennent de deux raisons distinctes. La première est due à la récurrence sur 24 itérations. On a montré en effet qu'elle est à exclure pour le perceptron multicouches. L'Armax, quand à lui, commettra plus d'erreur en fin de journée qu'en début et ne satisfait donc pas à la condition de prévoir aussi bien tout au long de la journée. Par contre ces modèles sont tout à fait adaptés à la prévision d'un paramètre journalier comme par exemple la moyenne, le minimum, le maximum, etc. Dans ce cas, on estime en effet une seule valeur à une échéance. L'autre raison est la prise en compte des jours particuliers (des exemples tels que les jours fériés ont été cités dans la section 4.3), ce que fait mal l'Armax. De nombreuses études s'appuient sur une classification des jours comme par exemple [61] ou les études mixtes classification-perceptron précitées.

A ce stade, pour s'adapter aux données, il apparaît naturel de diviser l'information pour isoler la partie qui concerne le type de jour. Les dates des jours fériés, samedis ouvrés etc. étant connues, seule la consommation électrique qui en résulte est à évaluer. On est donc dans un cadre décisionnel où des modèles basés sur des classifications (par exemple bayésiens) paraissent justifiés, la classification de Kohonen présentée dans la première partie a d'ailleurs donné satisfaction lorsqu'elle a été utilisée([107], [19]). Par contre, l'information qui ne dépend pas du type de jour peut-être traitée par des modèles récurrents comme ceux que l'on vient de voir. En particulier, si elle se réduit à un ou deux paramètres journaliers le perceptron multicouches devient tout à fait adapté. On va proposer dans le paragraphe suivant un modèle de prévision basé sur le découpage de l'information et la réduction du nombre de paramètres. On pourra constater que cette méthode résout en même temps le problème de la récurrence sur 24 heures et la prise en compte du type de jour en introduisant une notion de profil de consommation.

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

5.1 Introduction

On a vu précédemment que le modèle Armax ne prend pas en compte les changements de structures non périodiques et que le perceptron multicouches ne permet pas de faire de la prévision à plusieurs pas. On a donc cherché un modèle adapté, par exemple, aux spécificités des données de consommation électrique polonaises dont des exemples ont été donnés à la section 4.3. En particulier, on peut rappeler la présence d'un changement de structure

journalière entre *un jour travaillé* et *un jour férié* ou un *dimanche*. Cela nous invite à prendre en compte la notion de jour et ainsi **superposer deux échelles**, *le jour* et *l'heure*. On s'est ainsi intéressé à une approche différente de celle des modèles récurrents, qui consiste notamment à **étudier les journées prises dans leur ensemble**. Les unités de temps sont *le jour* et *l'heure* et le vecteur à estimer est la courbe de charge journalière qui est dans notre exemple un vecteur de dimension 24, $X_k = (X_k^1, \dots, X_k^{24}) = (X_k^h)_{1 \leq h \leq 24}$, où X_k^h est la consommation électrique à l'heure h du jour k . Ce choix devrait éviter les inconvénients décrits dans la présentation des méthodes classiques (chapitre 4.4).

Les données de courbes de charge s'intègrent donc au problème plus général défini précédemment au chapitre 4.1 et la méthode que nous allons présenter constitue donc une tentative de réponse à ce problème. *L'heure* devient ainsi l'unité "*haute fréquence*" et *le jour* l'unité "*basse fréquence*". On peut remarquer que l'unité de l'échelle rapide (l'heure) est imposée par la mesure des données alors que l'échelle lente, ici le jour, est déduite des observations préalables et des propriétés de la variable à prévoir.

Pour une courbe de charge de dimension 24, on distingue alors *sa forme (le profil)* de dimension 24, son *niveau de consommation globale (la moyenne)* de dimension 1 et *sa variabilité* représentée par *l'écart type* de dimension 1. On définit le profil P_k du jour k comme la courbe de charge de la journée centrée par la moyenne des 24 valeurs de consommation du jour et réduite par l'écart type de ces 24 valeurs. Ce profil est ainsi un vecteur de dimension 24.

La moyenne et l'écart type de chaque jour peuvent être estimés par une méthode quelconque de type Armax ou neuronale, qui est ici adaptée, car on ne prévoit qu'une seule valeur pour le jour suivant (il s'agit d'estimer 1 paramètre à 1 pas). Le profil, quant à lui, est déterminé à partir d'une classification qui permet à la fois la construction de classes et leur regroupement par un critère de voisinage. Une classification de type Kohonen qui diminue la somme des carrés intra-classes étendue aux classes voisines s'est révélée être un excellent outil pour réaliser cet objectif, comme on va le voir dans la suite.

Nous allons présenter dans ce chapitre la méthode que nous avons utilisée pour prévoir la consommation électrique polonaise dans le cadre général défini au chapitre 4.1.

5.2 Définitions et notations:

Données:

Soit une série chronologique $(Y_t)_{0 \leq t \leq T}$ de T valeurs mesurant un événement temporel qui varie en suivant deux évolutions de temps ayant chacune sa cadence propre. La première a une cadence qui correspond à 1 unité de temps, on dit que la première évolution est *instantanée* et on appelle *instant* la durée de sa réalisation. La seconde, quant à elle, se réalise au cours de p unités, p étant constant et on appelle *période* la séquence de temps correspondant à l'événement lent. On a ainsi N périodes de p observations ($N \times p = T$).

A chaque *période* (séquence basse fréquence) d'indice k , on affecte éventuellement q variables qualitatives A_k^1, \dots, A_k^q qui constituent le vecteur A_k de dimension q (A_k^1, \dots, A_k^q) $_{0 \leq k \leq N}$. Elles permettent de définir le *type* de la période d'indice k .

Notations:

On note par le double indice X_k^h la valeur de la série chronologique Y à l'instant h de la période k (par exemple l'heure h du jour k), $X_k^h = Y_{k \times p + h}$.

On note p le nombre d'événements *haute fréquence* qui se produisent au cours d'une période *basse fréquence*, p est supposé constant.

On note X_k la suite de valeurs successives communes à la période k , appelée *courbe de charge de la période k* .

$$X_k = (X_k^1, \dots, X_k^p)$$

On note t la variable – résultante de la conjonction des variables A^1, \dots, A^q – qui renseigne le type de la période et $t(k)$ le type de la période k . Si m_1, \dots, m_q sont les nombres de modalités respectifs des variables A^1, \dots, A^q , alors t a un nombre de modalités inférieur ou égal au produit $m_1 \times \dots \times m_q$. Par exemple, dans le cas de la consommation électrique polonaise, on dispose des variables qualitatives telles que le jour de semaine, le mois et l'indicatrice de jour férié. On peut définir le type du 31/12/98 comme un *mercredi non férié de décembre* et le 01/01/99 comme un *jeudi férié de janvier*.

Définitions:

La valeur moyenne de X pour la période k , notée M_k est donnée par :

$$M_k = \frac{1}{p} \sum_{h=1}^p X_k^h$$

L'écart type pour la période k , notée s_k , est donné par la formule:

$$s_k = \sqrt{\frac{1}{p} \sum_{h=1}^p (X_k^h - M_k)^2} \text{ et vérifie la propriété } \|X_k^h - M_k\|^2 = p s_k^2$$

Le profil pour la période k , noté P_k , représente la courbe de charge centrée réduite. C'est un vecteur de dimension p et de norme 1, $P_k = (P_k^1, \dots, P_k^p)$, donné par la formule

$$P_k^h = \frac{X_k^h - M_k}{\|X_k - M_k\|} = \frac{X_k^h - M_k}{\sqrt{p} \times s_k}$$

Le profil est un vecteur de la sphère de rayon 1 dans l'espace de dimension p .

On a donc **la décomposition suivante de la courbe X_k de la période k en profil, moyenne et écart type :**

$$X_k = M_k \mathbf{1}_p + \sqrt{p} s_k P_k$$

où $\mathbf{1}_p$ est le vecteur à p composantes toutes égales à 1

Cette formule est centrale dans la méthode de prévision.

5.3 Méthode de prévision – la MEP

On appelle *MEP* (moyenne – écart type – profil) la méthode de prévision proposée. Elle s'effectue en plusieurs étapes qui seront appliquées plus loin dans le cadre de l'analyse de la *consommation électrique polonaise*. Elle peut se résumer comme suit :

- 1) On classe les N profils P_k en U classes (C_1, \dots, C_U). Le choix de la classification est libre, on utilisera celle de Kohonen qui offre des possibilités de visualisation qui seront présentées ultérieurement. On note G_u le représentant (ou le vecteur code) de la classe C_u .
- 2) A chaque période d'indice k , on associe sa classe d'affectation $C_{(k)}=C_{u(k)}$, soit celle qui vérifie: $\|G_{u(k)} - P_k\| = \min_{u=1\dots U} \|G_u - P_k\|$
- 3) On définit **le type $t(k)$ de chaque période k** à partir des variables qualitatives (A^1, \dots, A^q) qu'on enregistre dans la variable t .
- 4) A chaque classe C_u , on associe $\mathcal{P}(u / t_0)$ qui est la probabilité qu'une courbe X_k appartienne à la classe C_u sachant qu'elle a le type t_0 , elle est estimée par la fréquence empirique

$$\tilde{\mathcal{P}}(u / t_0) = \frac{\sum_{k=1}^N \mathbf{1}_{\{C(k)=C_u\}} \times \mathbf{1}_{\{t(k)=t_0\}}}{\sum_{k=1}^N \mathbf{1}_{\{t(k)=t_0\}}}$$

où $\mathbf{1}_{\{condition\}}$ est la fonction qui prend la valeur 1 si la condition est respectée, 0 sinon.

5) On définit **la règle d'affectation** qui à la période d'indice k associe le profil $\tilde{P}_k = \tilde{P}_{t(k)}$ commun à tous les jours k' qui ont le même type $t(k)$.

\tilde{P}_k est donné par la formule:

$$\tilde{P}_k = \frac{\sum_{u=1}^U \tilde{P}(u/t(k)) G_u}{\left\| \sum_{u=1}^U \tilde{P}(u/t(k)) G_u \right\|}$$

\tilde{P}_k est ainsi de norme 1 et conserve la propriété d'être un profil.

6) On prévoit séparément M_k et s_k par une méthode de prévision récursive de type Armax ou de type neuronale telle que le perceptron multicouches. Comme on prévoit à la période $k-1$ la valeur à la période k , on est dans le cadre où ces techniques donnent de bons résultats. On note \tilde{M}_k et \tilde{s}_k les estimés respectifs de M_k et s_k .

7) On prévoit X_k par \tilde{X}_k , donné par la formule

$\tilde{X}_k = \tilde{M}_k + \sqrt{p} \tilde{s}_k \times \tilde{P}_k$

6 Application de la MEP à l'exemple de la *prévision de la consommation électrique nationale polonaise*

Données :

Nous allons prévoir la courbe de charge à partir *du relevé de la consommation électrique horaire nationale polonaise du 01 janvier 1986 au 31 décembre 1993 effectué par la compagnie nationale polonaise d'électricité*, ce qui constitue 2922 jours avec 24 valeurs horaires.

Définitions:

Conformément à la notation du chapitre 5, on note X_k^h la valeur de *la consommation électrique* du jour k à l'heure h .

L'unité de temps étant le jour, un individu X_k représentant *la courbe de charge du jour* k est donc déterminé par la donnée de 24 valeurs successives. On note:

$$X_k = (X_k^1, \dots, X_k^{24})$$

La consommation moyenne du jour k se calcule comme suit:

$$M_k = \frac{1}{24} \sum_{h=1}^{24} X_k^h$$

L'écart type est donné par la formule:

$$s_k = \sqrt{\frac{1}{24} \sum_{h=1}^{24} (X_k^h - M_k)^2}$$

Le profil du jour k , noté P_k est donné par la formule:

$$P_k^h = \frac{X_k^h - M_k}{\|X_k - M_k \mathbf{1}_{24}\|}$$

On a donc $X_k = M_k + \sqrt{p} s_k P_k$

Intérêt particulier de la méthode :

On a supposé d'une part que le profil évoluait indépendamment de la température et d'autre part que la météorologie était entièrement prise en compte dans la moyenne et l'écart type, ce qui est très intéressant dans le cadre d'une prévision. Les résultats de la classification et l'étude détaillée des erreurs effectuée à posteriori ont confirmé cette hypothèse.

6.1 Classification des profils

N'importe quel type de classification peut être appliqué. Néanmoins, dans le chapitre 7.3 sont développés des éléments qui argumentent plutôt vers le choix d'une classification réalisée à l'aide de l'algorithme de Kohonen. Les sections suivantes présentent la classification hiérarchique, puis celle qui utilise les cartes de Kohonen.

6.1.1 Utilisation d'une classification hiérarchique

Rappelons que la classification hiérarchique de type Ward a été présentée au chapitre 2.2. On utilise la méthode cluster-ward du logiciel SAS pour faire une première classification grossière des différents profils.

Si on s'intéresse aux regroupements correspondants à un changement de niveau d'inertie important, on constate que le niveau à 8 classes est trop simpliste. Ceux au delà de 10 classes étant trop difficiles à interpréter, on a donc choisi un découpage en 10 classes.

En considérant le tableau 1 qui décrit chacune des 10 classes à partir du croisement des jours et des mois, **on propose la description qualitative des classes suivante:**

- Classe 1 : jours de semaine d'avril août et septembre
- 2 : dimanches d'avril à août
- 3 : jours de semaine et samedis de mai à août
- 4 : jours de semaine d'octobre à février et lundis de mars ainsi que les samedis de novembre à janvier
- 5 : dimanches février mars et octobre
- 6 : samedis et dimanches d'avril et septembre
- 7 : samedis de mai à août
- 8 : samedis d'octobre à février
- 9 : dimanches d'octobre à janvier
- 10: jours de semaine de février mars et octobre sauf le lundi

Tableau 1 : Table de contingence qui croise les jours et les mois pour chacune des 10 classes de la classification hiérarchique.

CLASSE	MOIS	LUNDI	MARDI	MERCREDI	JEUDI	VENDREDI	SAMEDI	DIMANCHE
1	3	5	6	4	4	2	.	.
1	4	22	32	33	32	32	9	.
1	5	.	1	8	6	1	.	.
1	6	.	1	.	1	.	.	.
1	7	.	2	.	2	.	.	.
1	8	15	27	22	22	18	1	.
1	9	26	29	30	31	32	7	.
2	3	1	5
2	4	7	23
2	5	1	1	1	5	3	8	36
2	6	.	.	.	5	.	2	32
2	7	.	1	1	.	1	6	34
2	8	.	1	1	1	.	7	34
2	9	1
3	4	6	2	2	2	1	3	.
3	5	33	33	26	23	31	10	.
3	6	35	34	34	28	34	7	.
3	7	34	32	35	34	35	6	.
3	8	20	7	12	12	17	9	.
4	1	33	34	34	35	35	11	.
4	2	33	18	22	19	19	8	.
4	3	23	2	3	.	2	2	.
4	4	1	.	.
4	8	1	.
4	9	7	.	2	.	1	.	.
4	10	34	17	21	18	19	5	1
4	11	33	33	32	32	33	8	.
4	12	31	34	34	34	31	14	.
5	1	7
5	2	5	25
5	3	1	22	28
5	4	1	.
5	9	4	9
5	10	6	19
5	11	1	1	1	1	1	1	5
6	3	2	3
6	4	6	.	.	1	.	11	10
6	8	3	2
6	9	22	24
7	4	3	1
7	5	1	.	.	1	1	18	.
7	6	25	2
7	7	1	23	1
7	8	1	15	.
7	9	1
8	1	22	.
8	2	20	.
8	3	6	.
8	9	1	.
8	10	1	1	.	1	2	24	.
8	11	.	.	1	.	.	24	.
8	12	2	.	.	.	1	14	2
9	1	1	1	2	1	2	2	28
9	2	7
9	10	1	15
9	11	1	.	.	1	.	1	30
9	12	2	2	2	2	3	7	33
10	2	.	14	10	13	13	.	.
10	3	6	28	28	31	30	4	.
10	9	.	6	3	3	1	.	.
10	10	.	17	14	17	15	.	.

Remarques:

- La première constatation concerne la remarquable correspondance entre les caractéristiques calendaires (*type de jour et mois*) et les classes qui se dégagent des résultats au point qu'ils semblent être les critères qui permettent de caractériser les classes. A titre d'exemple, la classe 1 est essentiellement constituée *des jours travaillés*¹ des mois de *mi-saison avril et septembre* auxquels il faut ajouter certains *jours travaillés d'août* et les jours de *mars* postérieurs au changement d'heure.

- La deuxième constatation porte sur le fait que **la classification ne fait pas ressortir d'effet de température, mais au contraire, des effets saisonniers**. En effet, il est clair que certains jours de mars ont la même température que ceux de novembre par exemple, et pourtant, ils ne sont pas regroupés dans la classification. Une conclusion similaire s'impose lorsque l'on suit les jours et leur classement par ordre chronologique et que l'on constate une régularité qui ne correspond pas aux aléas de température (spécialement en hiver). **Ce résultat confirme l'hypothèse d'indépendance du profil et de la météorologie.**

- Refaire la classification avec un plus grand nombre de classes n'apporte pas de précisions, les ruptures entre les classes étant moins nettes du point de vue calendaire.

- On constate que certaines catégories se répartissent sur plusieurs classes (par exemple les mardis de mars se retrouvent dans 3 classes 1, 4 et 10). La classification hiérarchique ne fournit pas de moyen simple et rapide pour savoir si ces classes sont voisines ou non. Cela pose un problème qui sera détaillé dans le chapitre 7.3.

Limites de la classification hiérarchique:

Les limites de la classification hiérarchique concernant l'application de notre méthode seront décrites plus loin. Néanmoins, on peut déjà se rendre compte que si les mardis de mars se retrouvent sur 3 classes, notre méthode leur affectera un représentant qui est construit comme le barycentre de 3 vecteurs dont il n'est pas immédiat de savoir si ils sont semblables. Ce problème augmente avec le nombre de classes. On peut voir aussi ce représentant comme un barycentre de tous les éléments de ces 3 classes qui représentent presque la moitié des individus, et donc on n'aura pas la précision souhaitée.

¹ Nous n'avons pas la possibilité de vérifier si les samedis de la classe CL10 sont des samedis travaillés.

Conclusion:

Cette classification hiérarchique montre que les profils semblent indépendants de la température. De plus, on a remarqué un suivi temporel qu'il serait très intéressant de conserver lors d'une classification plus fine comprenant d'avantage de classes.

Une classification de type Kohonen qui permet de mieux visualiser les ressemblances entre individus et qui ordonne les représentants des classes (le cas échéant, en deux dimensions), favorise l'interprétation des résultats, la mise en valeur de la saisonnalité et permet d'augmenter considérablement le nombre de classes tout en gardant un résultat exploitable.

6.1.2 Utilisation de la classification de Kohonen (présentée au chapitre 2)

On effectue, dans ce chapitre, une classification au moyen de **l'algorithme de Kohonen**. Cette méthode possède la particularité, outre de classer, d'organiser les représentants des classes suivant une notion de voisinage qui est visualisé dans une carte (ici à deux dimensions). Cette classification sert à faire ressortir les profils des situations stables et à les ordonner, ce qui est intéressant dans le cadre de la prévision.

Les choix techniques qui ont été faits sont les suivants. Chaque jour est présenté 12 fois lors de l'apprentissage, ce qui fait *35064 itérations*. *Le voisinage* passe de 49 voisins à 25 puis 9 et enfin 1 voisin à respectivement $5/12^{\circ}$, $10/12^{\circ}$ et $11/12^{\circ}$ du temps d'apprentissage. *Le choix de la carte* s'est porté sur 100 unités disposées sur une carte 10x10. L'initialisation de la carte est faite avec 100 profils pris au hasard dans la base de données.

6.1.2.1 Variantes de l'algorithme de Kohonen susceptible d'améliorer la classification des courbes de charge :

Dans le chapitre 2.3² qui présente l'algorithme de Kohonen, il est indiqué que certaines modifications de l'algorithme peuvent être intéressantes dans certains cas. Dans le paragraphe suivant, sont présentées certaines variantes sur la structure de la carte et sur l'apprentissage.

- Variantes sur la structure :

On peut modifier la structure d'un réseau en grille en considérant un réseau en cylindre ou tore.

- Le cylindre:

On appelle cylindre un réseau de type grille où on impose *aux unités de deux bords opposés d'être voisines* comme l'illustre la suite de figures 57, 58, 59.

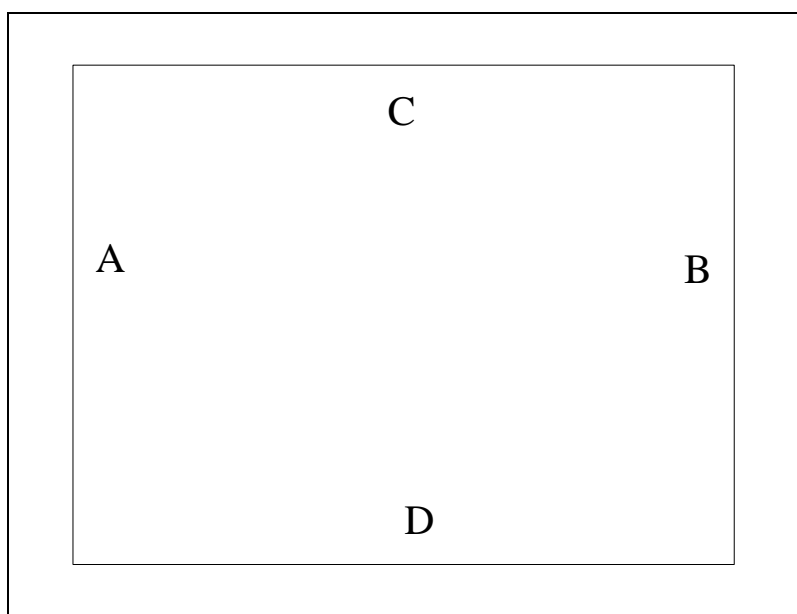


Figure 57 : Structure de carte rectangulaire classique

² Chapitre 2.3 paragraphe Variantes de l'algorithme de Kohonen

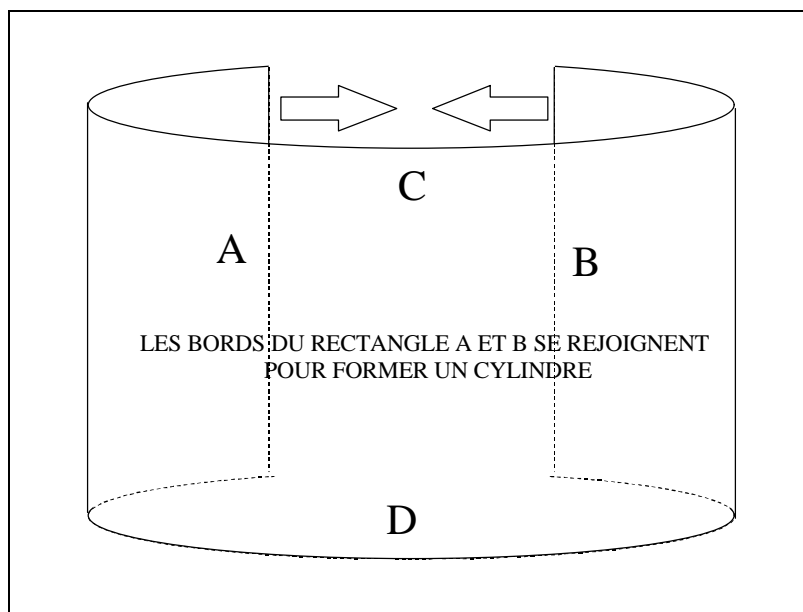


Figure 58 : Les bords A et B se rejoignent

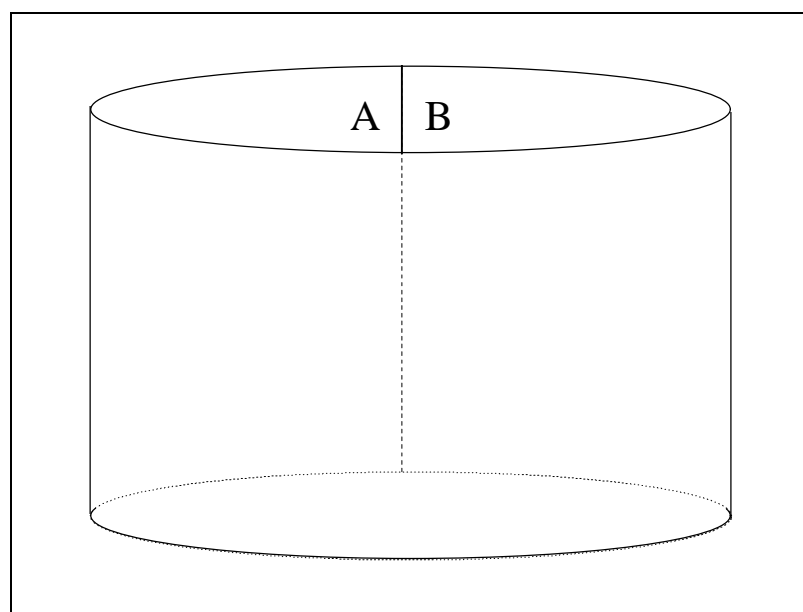


Figure 59 : Structure de carte cylindrique

- Le tore :

Le tore est une grille où les quatre bords sont voisins de leurs bords opposés respectifs comme sur la figure 60 (il peut aussi être perçu comme un cylindre où les 2 bases se rejoignent).

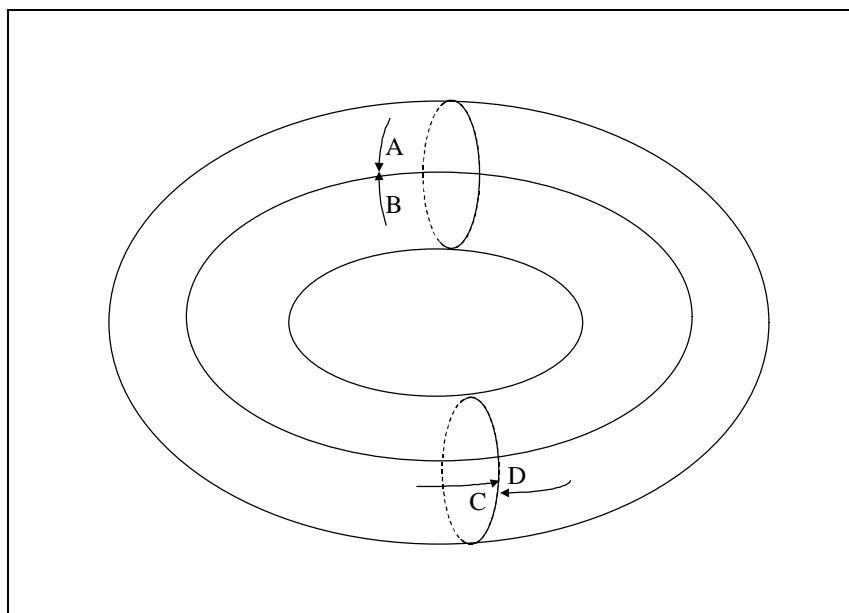


Figure 60 : Structure en tore

- Choix de la structure dans l'étude de la consommation électrique nationale polonaise :

Le cylindre favorise la mise en évidence d'une saisonnalité et en particulier du mouvement annuel hiver → printemps → été → automne → hiver. Dans le cas de la consommation d'électricité, l'hiver et l'été sont fortement opposés alors que les saisons intermédiaires sont plus proches. La grille rectangulaire a donc tendance à privilégier la représentation de l'opposition entre l'hiver et l'été en les plaçant sur des unités extrêmes de la carte et à centrer les saisons intermédiaires de sorte que cela provoque un regroupement du printemps et de l'automne. Une grille plus grande aurait peut-être résolu le problème, mais le cylindre a fourni de bons résultats en forçant l'étalement de la saisonnalité à la fois dans l'étude de la consommation d'électricité nationale polonaise et française³.

³ Cf. annexe

Il est très difficile de définir une règle pour savoir dans quel cas le cylindre apporte un plus ou non. Seul un bilan à posteriori comme dans cette étude peut montrer l'intérêt de cette structure.

- Variante sur l'apprentissage : La renormalisation des poids à chaque étape.

Les vecteurs codes G_u sont initialisés dans l'enveloppe convexe E de l'ensemble des éléments à classer X . Ils évoluent suivant la formule présentée au chapitre 2.3 $G_u(t+1) = G_u(t) + \mathbf{h}(t)(X(t+1) - G_u(t))$ et sont donc, à chaque étape t , un barycentre de deux vecteurs de E . Ils sont donc tout au long de l'apprentissage dans l'enveloppe convexe E .

Par exemple, si les éléments à classer possèdent une propriété linéaire ($\sum_{j=1}^p a_j X^j = cste$), les vecteurs G_u auront aussi cette propriété qui se transmet aux barycentres.

Par contre, si les éléments X possèdent une propriété non linéaire et en particulier, si les vecteurs X sont sur la sphère unité, comme c'est le cas des profils journaliers de consommation électrique, les vecteurs codes seront à l'intérieur et non sur la boule unité comme l'illustre la figure 61. Et dès que 2 vecteurs X distincts sont affectés à l'unité u ,

$$\sum_{j=1}^p G_u^{j^2} < 1 .$$

Il est possible d'aménager l'algorithme pour que les vecteurs codes restent à chaque étape sur la sphère unité comme les vecteurs à classer en normalisant le vecteur code $G_u(t)$ à chaque étape de l'apprentissage par la formule suivante⁴:

$$G_u(t+1) = \frac{G_u(t) + \mathbf{h}(t)(X(t+1) - G_u(t))}{\|G_u(t) + \mathbf{h}(t)(X(t+1) - G_u(t))\|}, \text{ pour } u \in V_r(u_0)$$

$G_u(t+1) = G_u(t)$, pour les autres.

En résumé, dans cette étude on choisit donc une *carte cylindrique* pour illustrer la saisonnalité des données et *une renormalisation des poids à chaque itération* afin de les laisser sur la sphère unité et ainsi garder cette propriété des profils.

⁴ Cette formule remplace son équivalent du chapitre 2.4

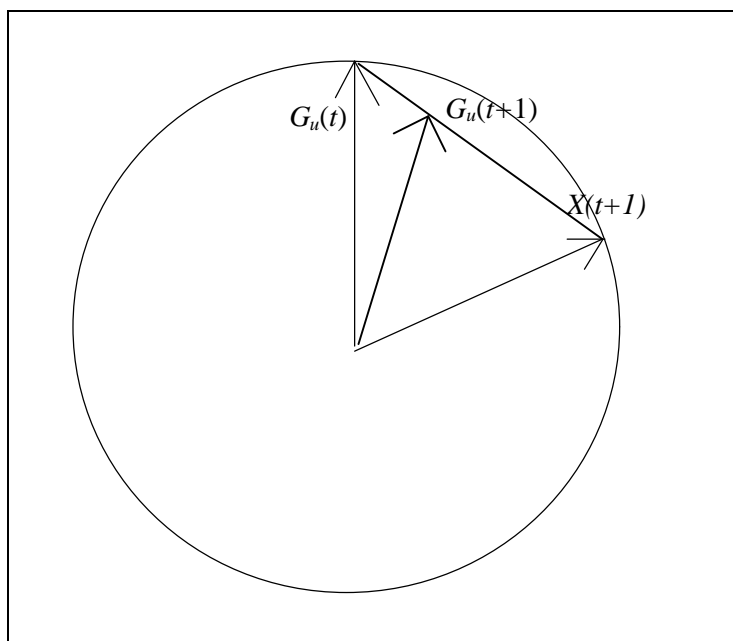


Figure 61 : Le vecteur code $G_u(t+1)$ étant un barycentre de $G_u(t)$ et de X se trouve à l'intérieur du cercle unité

6.1.2.2 Classification et prévision des profils

A la suite de l'apprentissage de la carte de Kohonen modifié comme expliqué ci-dessus, on obtient **100 classes** dont les représentants sont les vecteurs codes de l'algorithme de Kohonen (vecteurs de 24 composantes). Ils sont ordonnés sur une carte où il faut considérer que les bords verticaux sont voisins (le réseau de voisinage est un cylindre).

Représentation des centres de classes :

On peut aisément vérifier sur le graphique de la figure 62, où est tracé pour chaque unité u le vecteur code final correspondant G_u , que **les vecteurs codes des unités voisines sur la carte sont des profils voisins** ainsi que ceux des bords verticaux.

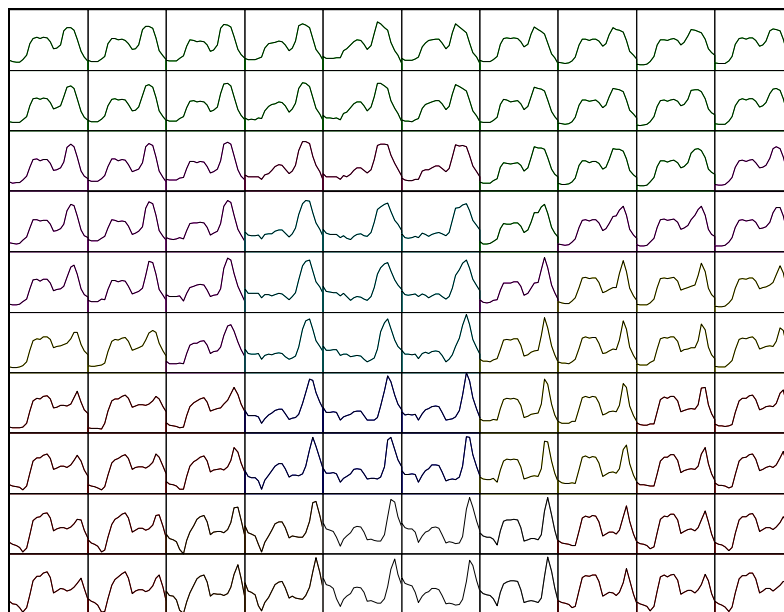


Figure 62 : Les vecteurs codes sont représentés sur la carte dans leur unité d'affectation

Représentation de tous les jours sur la carte :

On affecte ensuite à chacun des 2922 jours, l'unité dont le vecteur code est le plus proche en distance de son profil. Le graphique suivant superpose tous les profils dans l'unité qui leur est affectée et permet de vérifier que la dispersion à l'intérieur de chaque classe est très faible.

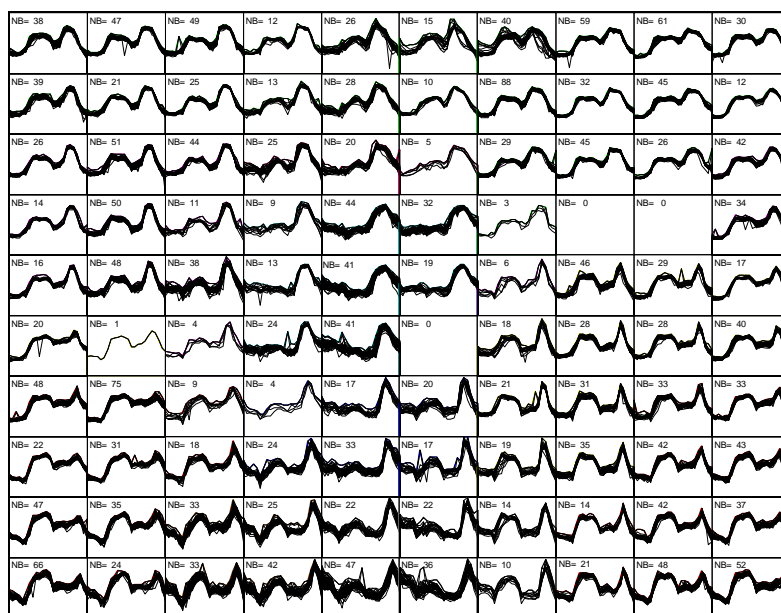


Figure 63 : Tous les profils sont représentés dans leur unité d'affectation

Carte des distances :

On cherche maintenant à positionner le vecteur code G_u par rapport à l'ensemble des individus extérieurs à la classe u , pour avoir un indicateur de l'éloignement de la classe u par rapport au reste du nuage de points. Pour cela, on pourrait mesurer la distance entre G_u et les $U-1$ autres vecteurs codes. En fait, la carte de Kohonen ayant l'avantage d'ordonner les classes selon un critère de proximité, il suffit de mesurer les distances à ses seuls voisins. Le graphique suivant donne pour chaque unité la distance entre le vecteur code associé et celui des 8 unités voisines. Pour plus d'information sur ce graphique, on peut se référer au chapitre 3. Cette carte indique les fossés intra-classes (c'est-à-dire les frontières correspondantes aux grandes distances entre les classes). Elle invite ainsi à regrouper les classes dont les vecteurs codes sont proches et séparer les autres.

Pour réunir les vecteurs codes proches et visualiser les regroupements de classes correspondants, on effectue une classification hiérarchique de ces 100 classes en macro-classes. La carte des distances sert alors à vérifier la cohérence de ces dernières avec les distances entre les vecteurs codes.

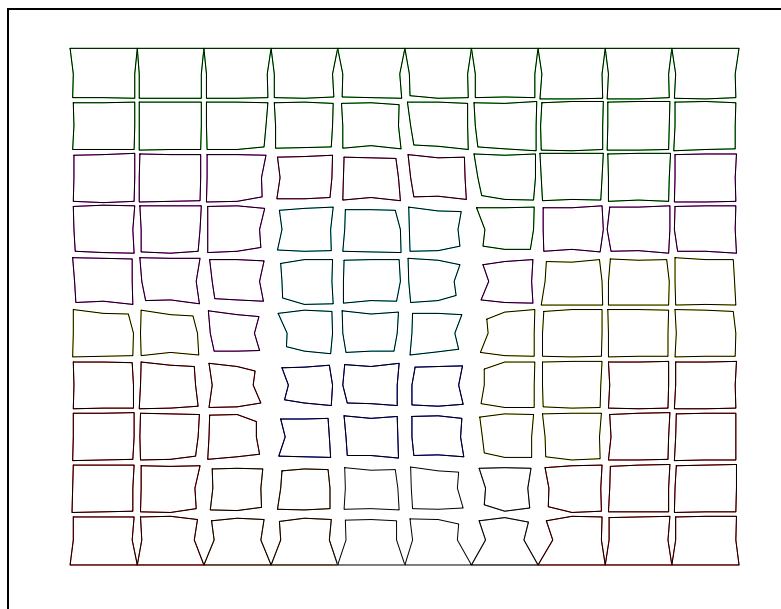


Figure 64 : Carte des distances

Visualisation des macro-classes :

Afin de rendre plus lisible la carte résultat de l'apprentissage (et non dans l'optique de la prévision), **on effectue une classification hiérarchique (méthode `cluster_ward` de SAS) à 10 classes à partir des 100 vecteurs codes précédemment obtenus** (le choix de 10 classes correspondant à un palier d'inertie et un stade où l'interprétation reste aisée). Conformément à la notation du chapitre 3, on appellera ces 10 classes les macro-classes pour les distinguer des 100 micro-classes issues de la classification de Kohonen. En associant à chaque macro-classe une couleur, on peut représenter sur la même carte, les deux classifications comme sur la figure 65. Les cartes *des distances* et de *la dispersion intra-classes* deviennent celles des figures 66 et 67:

On constate sur la carte des distances que les changements de classes correspondent à de grandes distances. On peut de plus distinguer des changements de classes voisines par continuité et d'autres par des ruptures.

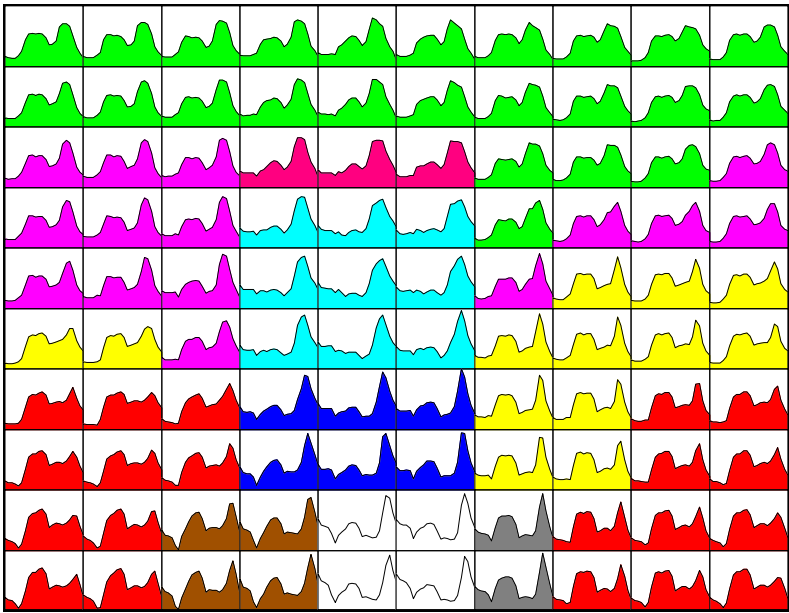


Figure 65 : Superposition des macro-classes et des micro-classes

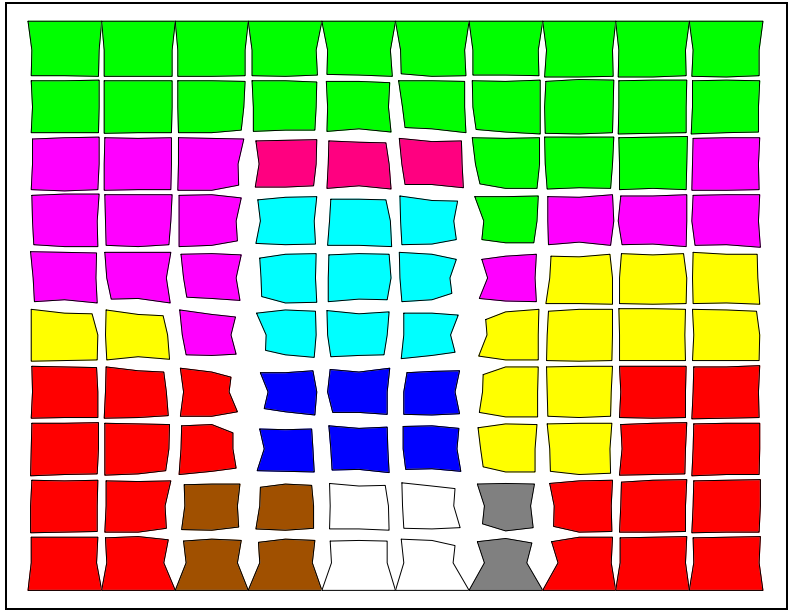


Figure 66 : Nouvelle carte des distances

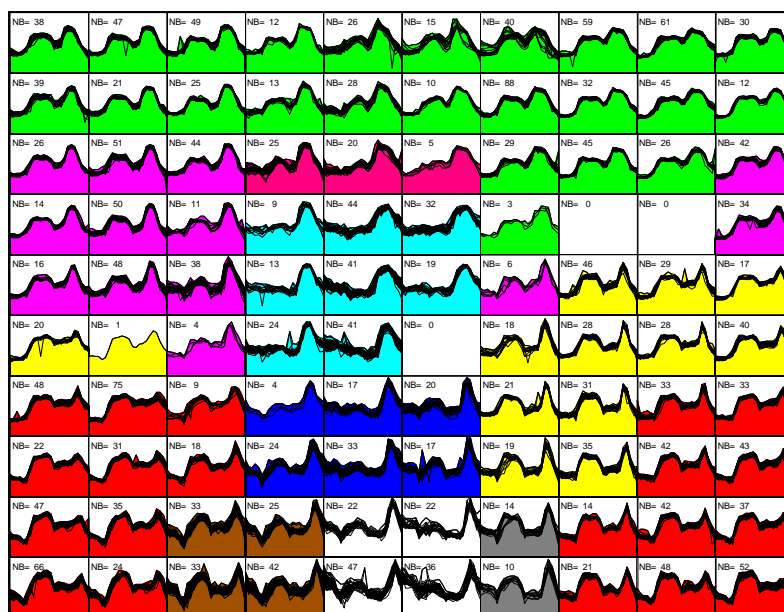


Figure 67 : Les individus sont représentés simultanément dans leur micro-classe et leur macro-classe

Description et typologie des classes :

Pour chaque niveau de classification, on peut, par la suite, essayer de **caractériser chaque classe par des variables exogènes** calendaires (donc déterministes), telles que le type de jour, le mois, et la spécificité de certains jours (jours fériés et ponts).

Niveau macro-classes:

Pour qualifier les différentes macro-classes en terme de "type de jour" et de "mois", on utilise le tableau 2 de contingence. qui croise les classes, les jours et les mois.

Tableau 2 : Table de contingence qui croise les jours et les mois pour chaque macro-classe

MACRO-CLASSE	MOIS	LUNDI	MAR->VEN	SAMEDI	DIMANCHE
1	1	33	136	26	.
1	2	17	72	8	.
1	3	.	6	.	.
1	9	1	2	.	.
1	10	13	82	10	1
1	11	33	130	30	.
1	12	33	133	27	.
2	1	.	2	.	.
2	2	16	56	11	.
2	3	29	117	27	.
2	4	.	3	2	.
2	9	6	14	4	1
2	10	21	60	15	.
3	4	10	23	5	.
3	5	33	126	12	.
3	6	34	131	11	.
3	7	35	139	9	.
3	8	24	132	10	.
3	9	1	4	1	.
4	3	5	18	.	.
4	4	18	103	4	.
4	5	.	4	.	.
4	6	1	1	.	.
4	7	.	1	.	.
4	8	12	5	2	.
4	9	26	118	15	.
5	1	1	6	.	34
5	2	.	.	1	32
5	3	.	.	1	16
5	9	.	.	.	7
5	10	.	.	4	34
5	11	1	4	1	34
5	12	2	9	3	33
6	4	.	9	9	.
6	5	.	.	2	.
6	8	.	.	3	1
7	3	2	.	3	15
7	4	5	.	10	15
7	5	2	2	2	4
7	7	.	.	.	1
7	8	.	.	4	12
7	9	.	.	13	25
8	5	.	.	20	5
8	6	.	.	23	13
8	7	.	.	25	25
8	8	.	.	17	5
9	1	.	.	9	1
9	2	.	.	13	.
9	3	.	.	3	.
9	9	.	.	1	.
9	10	1	.	7	.
9	11	1	2	3	1
9	12	.	1	5	2
10	3	.	.	1	5
10	4	1	.	4	19
10	5	.	9	.	27
10	6	.	5	.	21
10	7	.	3	1	9
10	8	.	3	.	18
10	9	.	.	.	1

Le résultat obtenu est illustré par la figure 68 et par le tableau 2 dans lesquels l'on peut vérifier les propriétés suivantes:

On constate sur la figure 65 que les macro-classes regroupent des micro-classes voisines sur la carte pour créer des surfaces connexes et l'association de ces macro-classes à une notion qualitative de type "saisonnier" ou "jour de semaine" est aisée.

De plus les séparations entre les 10 classes coïncident avec les grandes distances entre les poids comme le montrent la figure 66.

On constate un regroupement en fonction d'une part de la notion de jour (jours travaillés, samedis et dimanches), d'autre part de celle de la saison.

Jusqu'ici, ces classifications fournissent une typologie des jours selon les profils de charge. On pourrait se contenter de ce niveau de classification pour définir la *variable type de jour* à partir des 8 modalités "jour de semaine d'octobre", "samedi d'hiver", etc. En fait, a posteriori cette définition du jour est très large et induit une erreur de prévision trop importante. Comme on dispose d'un niveau micro-classes avec une notion de voisinage, il est possible de définir un type de jour plus fin, c'est-à-dire qui respecte la distinction entre ces 8 types. Il suffit pour cela de créer, comme au paragraphe suivant, des catégories qui soient emboîtées dans ces 8 types.

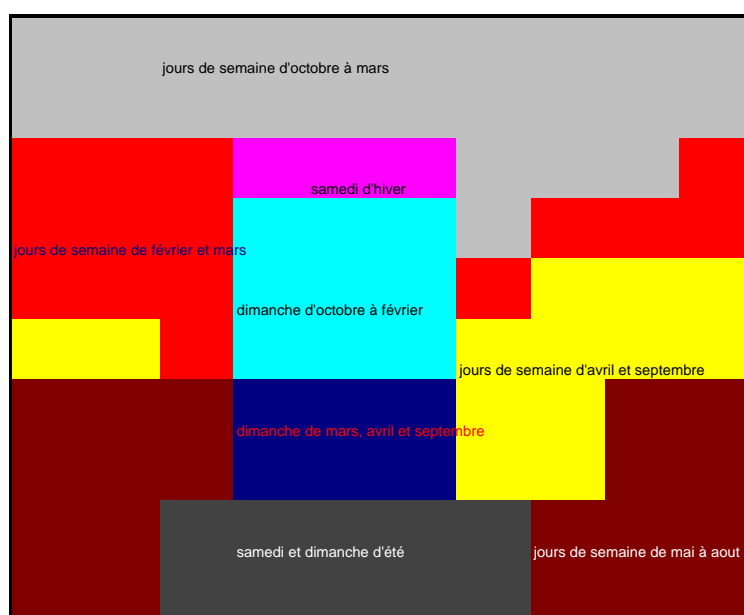


Figure 68 : Caractérisation qualitative des macro-classes

Niveau micro-classes :

On cherche à élaborer une variable type de jour construite à partir des variables calendaires dans le but de trouver une règle qui affecte à chaque jour les micro-classes qui correspondent à son type. Pour cela, on utilise **les tableaux de contingence** qui indiquent d'une part **comment est reparti chaque couple** (jour de semaine, mois) dans les 100 classes (dont le tableau 3 constitue un extrait) et d'autre part **quels sont, pour une classe donnée, les couples** (jour de semaine, mois) **associés** (le tableau 4 en donne un extrait), la première approche est nécessaire pour la prévision alors que la deuxième sert à qualifier les unités.

Exemple d'interprétation du tableau 3:

Les dimanches de février se répartissent sur 5 classes voisines sur la carte (44, 45, 46, 54, 55), les deux tiers étant affectés à la classe 55.

			44	45	46				
			54	55					

Figure 69: Positionnement des unités 44, 45, 46, 54, 55 où sont contenus les dimanches de février

Exemple d'interprétation du tableau 4:

La classe 1 est essentiellement constituée des jours de semaine (mardi à vendredi) des mois de janvier et février.

Tableau 3 : Extrait du tableau de contingence des couples (jour, mois) dans chaque micro-classe

MOIS	JOUR	MICRO-CLASSE	NOMBRE	POURCENTAGE
2	dimanche	44	4	13
2	dimanche	45	1	3.1
2	dimanche	46	4	13
2	dimanche	54	3	9.4
2	dimanche	55	20	63
2	lundi	19	3	9.1
2	lundi	20	7	21
2	lundi	29	7	21
2	lundi	30	8	24
2	lundi	31	1	3.0
2	lundi	40	7	21
2	mar->vend	1	11	8.6
2	mar->vend	2	5	3.9
2	mar->vend	10	7	5.5
2	mar->vend	11	24	19
2	mar->vend	12	19	15
2	mar->vend	13	3	2.3
2	mar->vend	14	1	.78
2	mar->vend	20	2	1.6
2	mar->vend	21	10	7.8
2	mar->vend	22	22	17
2	mar->vend	23	3	2.3
2	mar->vend	30	7	5.5
2	mar->vend	31	1	.78
2	mar->vend	32	13	10
2	samedi	1	1	3.0
2	samedi	11	3	9.1
2	samedi	14	3	9.1
2	samedi	15	1	3.0
2	samedi	22	3	9.1
2	samedi	24	12	36
2	samedi	25	1	3.0
2	samedi	32	1	3.0
2	samedi	33	1	3.0
2	samedi	34	1	3.0

Tableau 4 : Extrait du tableau de contingence qui croise les jours et les mois pour chaque micro-classe

MICRO-CLASSE	MOIS	LUNDI	MAR->VEN	SAMEDI	DIMANCHE
1	1	.	19	2	.
1	2	.	11	1	.
1	10	.	.	4	.
1	11	.	.	1	.
2	1	.	29	.	.
2	2	.	5	.	.
2	10	.	8	1	.
2	11	.	4	.	.
3	1	.	8	.	.
3	10	.	29	.	.
3	11	.	12	.	.
4	1	.	.	1	.
4	10	.	3	.	.
4	11	.	7	.	.
4	12	.	1	.	.
5	1	.	.	1	.
5	11	.	.	12	.
5	12	1	2	10	.
6	1	.	.	2	.
6	10	.	1	.	.
6	11	.	2	2	.
6	12	.	5	3	.
7	1	.	3	5	.
7	11	.	7	4	.
7	12	.	14	7	.
8	1	.	6	1	.
8	11	.	20	.	.
8	12	.	32	.	.
9	1	10	3	.	.
9	10	1	.	.	.
9	11	19	.	.	.
9	12	23	5	.	.
10	1	2	18	.	.
10	2	.	7	.	.
10	10	.	1	.	.
10	11	.	2	.	.

Cartographie des mois et jours :

Les tableaux précédents ne tiennent pas compte des voisinages. il est donc difficile de savoir, lorsqu'un type de jour se répartit dans deux classes différentes, si ces classes sont voisines et donc assimilables ou réellement disjointes. Pour visualiser ce problème, on utilise les cartes des jours et des mois (cf. section 3.6).

La première carte – qu'on appelle *cartographie des jours* (figure 70) – projette la répartition des jours dans les micro-classes. On constate que les mardis, mercredi, jeudi et vendredi ont la même répartition, on les regroupe donc sous une même modalité pour réaliser une nouvelle cartographie des jours plus simple représentée dans la figure 71. Les nouvelles modalités calendaires sont donc:

- 1: dimanche
- 2: lundi
- 3: mardi mercredi jeudi et vendredi
- 4: samedi

La figure 72 représente, quant à elle, la cartographie des mois.

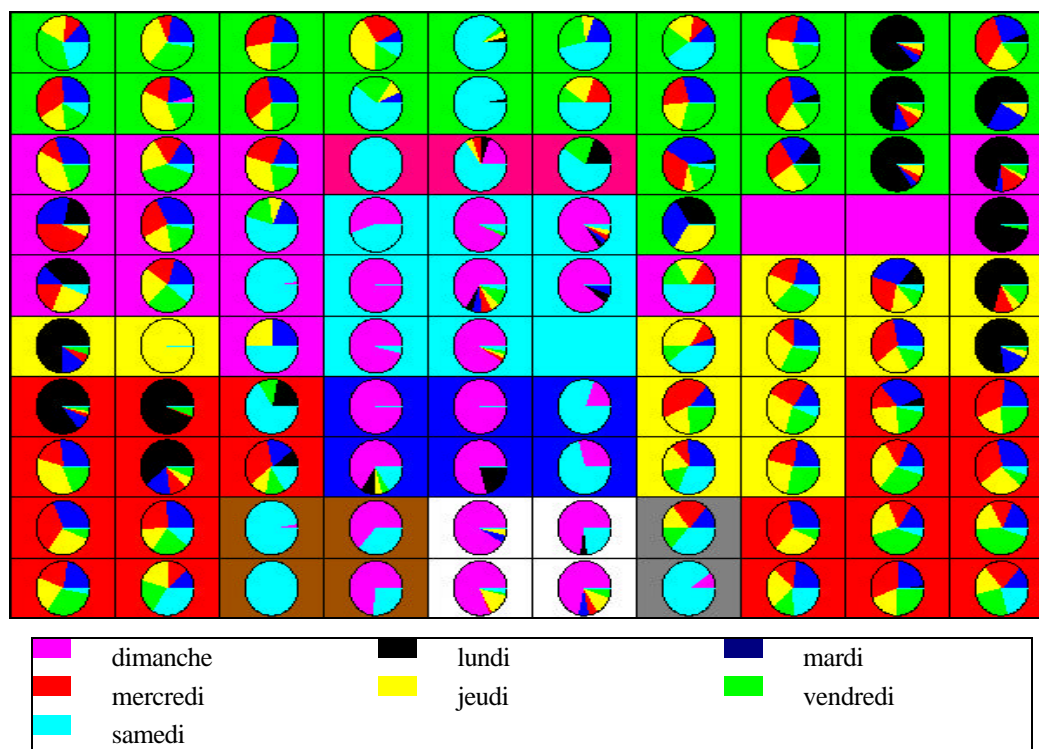


Figure 70 : les jours

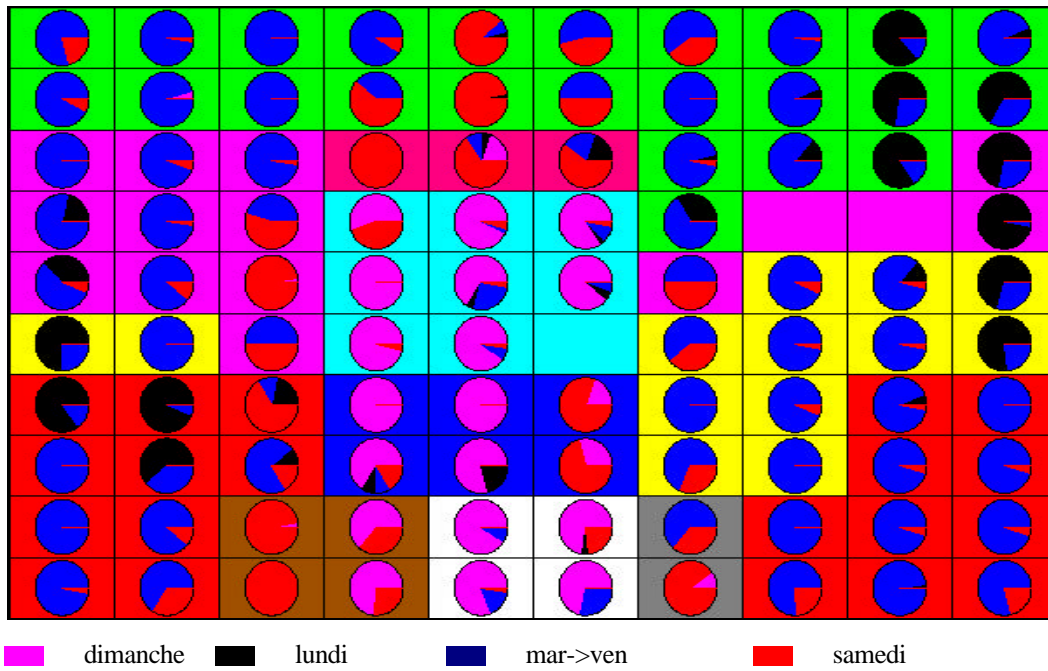


Figure 71 : Les jours avec regroupement des jours de mardi à vendredi

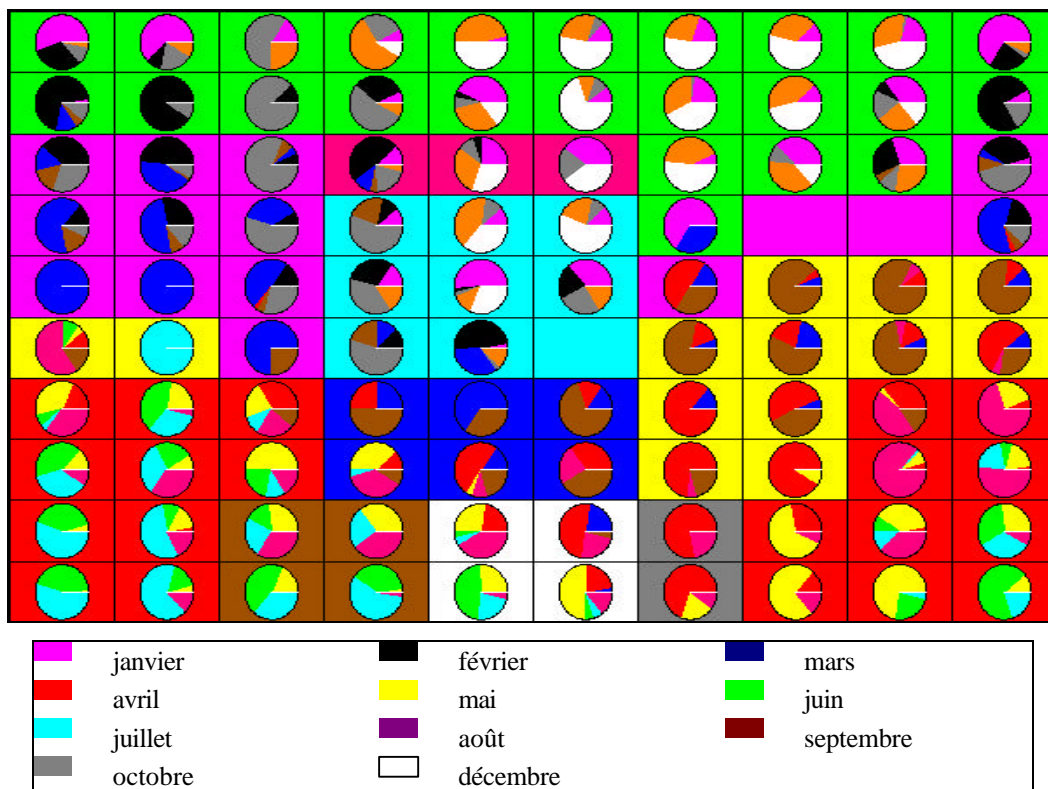


Figure 72 : Les mois

Affinage de la description de la carte:

On constate que le mois de janvier ne couvre pas toute la surface « jours de semaine d'octobre à mars » de la carte mais seulement une région. De plus, les jours de décembre et de février ne se mélangent pas. En fait, chaque mois est réparti dans une région de la carte beaucoup plus réduite que la zone réservée à sa super-classe. De plus, on peut suivre sur la cartographie des mois le passage d'un mois au suivant. Les unités concernées par deux mois successifs sont en effet communes ou contiguës la plupart du temps. Ce phénomène est illustré par les figures 73 et 74. La première indique les grandes lignes pour les « dimanches d'hiver » et les « jours de semaine d'hiver ». La seconde, établie à partir d'une rotation du cylindre autour de son axe, montre comment sont répartis les jours de semaine des mois d'été. Le passage d'une unité à l'autre ne peut cependant pas être daté mais cette répartition des mois sur la carte est facile à exploiter dans le cadre d'une prévision.

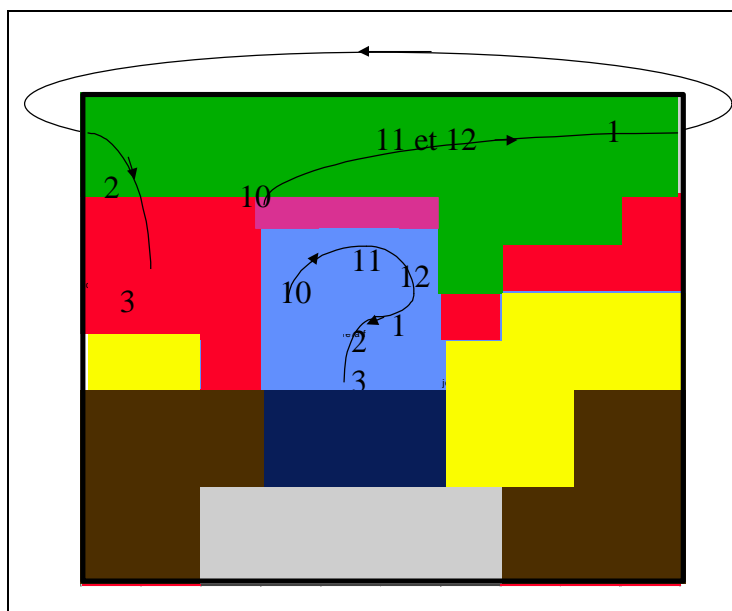


Figure 73 : Disposition sur la carte des mois d'hiver pour les deux catégories - jours de semaine et dimanches

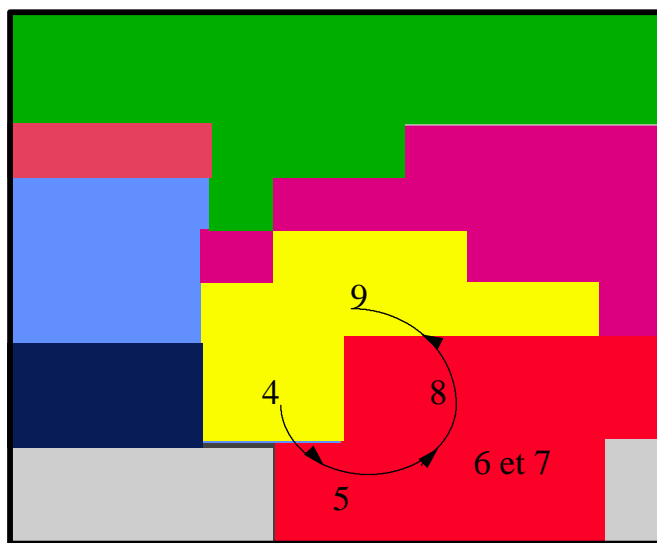


Figure 74 : Disposition sur la carte des mois d'été pour les jours de semaine

Le cas particulier des jours fériés:

Le tableau suivant, qui donne l'unité d'affectation de chaque jour férié, montre que les jours fériés sont classés comme des dimanches du même mois.

Tableau 5 : Numéro de l'unité d'affectation de chacun des jours fériés

ANNEE	1986	1987	1988	1989	1990	1991	1992	1993
1er janvier	45	45	45		45	45	45	45
Pâques	75	75	75		75	75	75	86
1er mai	96	96		74	96	96	96	
Constitution					96	74		74
Fête-Dieu	96	95	95	95	95	96	95	95
22 juillet	85	95	95					
15 août	96			96	96	85		
1er novemb	46		46	55	55	55		46
11 novemb	3					25	25	25
25 décemb	45	45		45	45	45	45	
26 décemb	35		36	36	36	36		

Conclusion :

La classification en macro-classes permet de séparer les jours de semaine, les samedis et les dimanches. Les samedis sont situés dans des classes intermédiaires entre celles des dimanches et celles des jours de semaine. Certains samedis se trouvent dans la zone des classes de jours de semaine – à la frontière de cette zone, comme par exemple les classes 5 14 15 16 – (on sait que certains samedis sont travaillés en Pologne). Malheureusement, faute de calendrier précis, on n'a pas pu vérifier l'hypothèse que les samedis ouvrés soient précisément les samedis classés comme des « *jours de mardi à vendredi* » qui formeraient ainsi la continuité entre les week-end et les jours de travail.

La fracture entre l'hiver et le printemps et celle entre l'été et l'automne se font au changement horaire. Par contre, les ruptures entre avril et mai ou entre août et septembre ne répondent pas à une règle calendaire évidente. Une étude sur une période plus longue pourrait peut-être identifier la date de rupture de façon plus précise. De plus, les jours fériés sont classés comme des dimanches du même mois.

La classification micro-classes permet d'affirmer qu'on trouve un mouvement au sein des macro-classes qui correspond au cycle des mois. De plus, on peut distinguer les lundis des autres jours de semaine.

Construction de la variable type de jour :

L'analyse de la classification par l'algorithme de Kohonen qui vient d'être effectuée montre que les profils se regroupent suivant certaines variables calendaires telles que *le mois*, *le jour de semaine*, *l'indicatrice de jour férié*, *l'indicatrice d'horaire d'été*, *l'indicatrice d'horaire d'hiver* etc.

On appellera *mars* (respectivement *septembre*) les jours de mars (respectivement septembre) jusqu'au changement horaire. Les jours suivants le changement horaire seront rattachés au mois suivant (*avril*, respectivement *octobre*). Par exemple, le 31 mars 1986 est un jour d'avril pour notre notation. De plus, on appelle dimanche *un dimanche* ou *un jour férié*.

On appelle alors **type du jour** k noté $t(k)$ la conjonction (jour de semaine, mois) où la notion de jour de semaine et de mois correspondent à la remarque précédente (*un jour férié* est un *dimanche* et les jours de mars ou de septembre qui suivent le changement horaire sont du type jours d'avril, respectivement d'octobre).

On distingue alors 48 modalités calendaires qui correspondent aux 12 mois de la nouvelle notation (voir ci-dessus) et aux 4 types de jours (lundis, dimanches, samedi, autres).

Prévision du profil :

Considérons un type $t=(\text{jour de semaine, mois})$ parmi les 48 possibles. A partir du tableau 2, on connaît la probabilité empirique $\tilde{P}(u/t)$ pour un tel élément d'être classé dans l'unité u dont le représentant est G_u .

Le profil prévu associé à t vaut alors:

$$\tilde{P}(t) = \frac{\sum_{u=1}^{100} \tilde{P}(u/t) \times G_u}{\left\| \sum_{u=1}^{100} \tilde{P}(u/t) \times G_u \right\|}$$

Par exemple, considérons un dimanche de février. A partir du tableau 2, on définit:

$$\tilde{P}(\text{dimanche de février}) = \frac{4}{32} \times G_{44} + \frac{1}{32} \times G_{45} + \frac{4}{32} \times G_{46} + \frac{3}{32} \times G_{54} + \frac{20}{32} \times G_{55}$$

Le graphique suivant superpose les 5 profils G_{44} , G_{45} , G_{46} , G_{54} , G_{55} tracés en bleu avec $\tilde{P}(\text{dimanche de février})$ tracé en rouge.

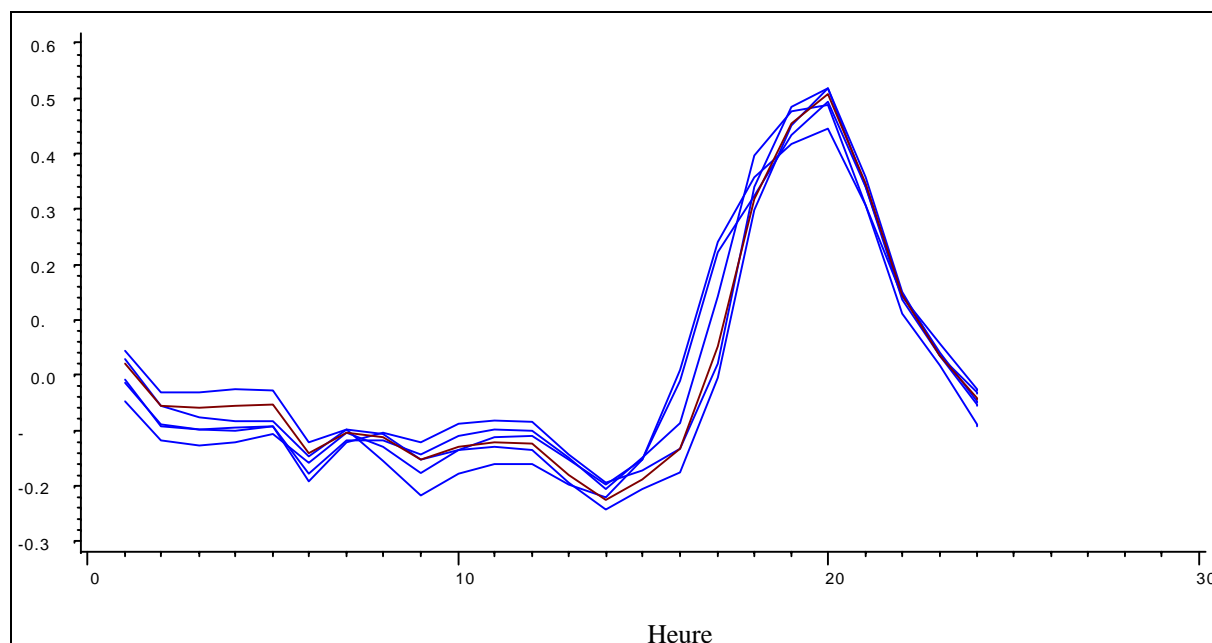


Figure 75 : Superposition des représentants des 5 classes qui contiennent des dimanches de février (courbes bleues) et du profil estimé de ce type de jour (courbe rouge)

6.2 Estimation des courbes de charges électrique et performances

Après avoir défini la règle de prévision des profils, on va maintenant estimer l'ensemble de la courbe de charge électrique polonaise. On validera la méthode MPE en la comparant avec un modèle armax classique qui prend en compte l'information des données par demi-heure et qui prévoit l'heure h du lendemain comme la prévision à l'échéance h du modèle. On présentera dans ce chapitre un bilan de l'erreur globale qui compare les deux modèles mais aussi indique la part d'erreur apportée par chacune des trois composantes de la MEP.

6.2.1 Estimation des courbes de charge électrique par le méthode MPE.

Pour prévoir la courbe de charge électrique polonaise, il suffit de compléter l'estimation du profil présenté dans la section 6.1 par celle de la moyenne et de l'écart type.

Estimation de la moyenne et de l'écart type :

La moyenne est estimée par un modèle armax(7,1 7,(1 2 3)(7)) avec une différenciation préalable d'ordre 1 et 7. On utilise, de plus, l'apport de la variable supplémentaire qui vaut 1 si le jour est férié et 0 sinon.

Le modèle utilisé est le suivant :

$$(1-B)(1-B^7)(1-0.04 \times B^7)M_t = (1-0.93 \times B^7)(1-0.23 \times B - 0.10 \times B^2 - 0.10 \times B^3)e_t - 0.17 \times \mathbf{1}_{\{\text{jour férié}\}}$$

On utilise le même type de modèle pour prévoir l'écart type :

$$(1-B)(1-B^7)(1-0.06 \times B^7)s_t = (1-0.85 \times B^7)(1-0.76 \times B - 0.07 \times B^2 - 0.037 \times B^3)e_t - 0.037 \times \mathbf{1}_{\{\text{jour férié}\}}$$

Estimation de la courbe de charge X_k du jour k :

L'estimation \tilde{X}_k de X_k se fait naturellement par la formule suivante:

$$\tilde{X}_k = \tilde{M}_k \mathbf{1}_p + \sqrt{p} \tilde{\mathbf{S}}_k \tilde{P}_k$$

Les figures suivantes superposent les courbes X_k et \tilde{X}_k (respectivement bleue et rouge). L'estimation de la courbe de charge résultant de trois prévisions (la moyenne, l'écart type et le profil), il est intéressant de connaître la contribution de chacune à l'erreur globale c'est-à-dire évaluer la performance d'une prévision où M est connu (ce qui revient à estimer $\mathbf{s}P$) et celle où M et \mathbf{s} sont connus (seul P est alors estimé). Dans ce dernier cas, il convient de considérer la qualité d'estimation de P mais aussi de $\mathbf{s}P$. Dans un premier temps on illustrera ces différents cas sur deux échantillons de la base de données. La semaine *du lundi 02 mars 1987 au dimanche 08 mars 87* est une semaine d'hiver à priori classique, celle *du jeudi 30 avril 87 au mercredi 07 mai 87* est une semaine qui contient un jour férié (*01 mai*) qui est un vendredi et fait partie des jours qui créent de grandes erreurs. Par la suite, on comparera la méthode *MEP* à un modèle armax de référence.

Ajustement des profils (quand \mathbf{s} et M sont connus) par la méthode *MEP* :

Avec cette méthode, on ajuste plutôt bien les formes (graphiques 76 et 77) quel que soit le type de jour, y compris le jour férié. Les ponts ne sont pas pris en compte dans notre modèle (comme pour l'armax) et il ne faut donc pas trop tenir compte du *samedi 02 mai 87*. Par contre, un décalage dans le temps des profils, comme celui du *dimanche 08 mars 87* (on perçoit sur la figure 76 que la prévision est en retard d'une heure dans la soirée), est une source d'erreur typique de la méthode. Pour ce jour précis, on peut voir que cette erreur⁵ n'a pas d'effet sur le maximum ou le minimum et le décalage entre la forme estimée et la réelle n'est que d'une heure. Ceci est intéressant car notre estimation permet d'être plus efficace que ne le mesure l'erreur dans une problématique du type « Quelle charge électrique maximale la société d'électricité nationale polonaise devra elle fournir demain ? » et « A quelle heure? ». De plus, « l'erreur par décalage horaire » possède sûrement une propriété temporelle qu'une étude spécifique complémentaire permettrait de réduire ou de contrôler.

Représentation de P_k :

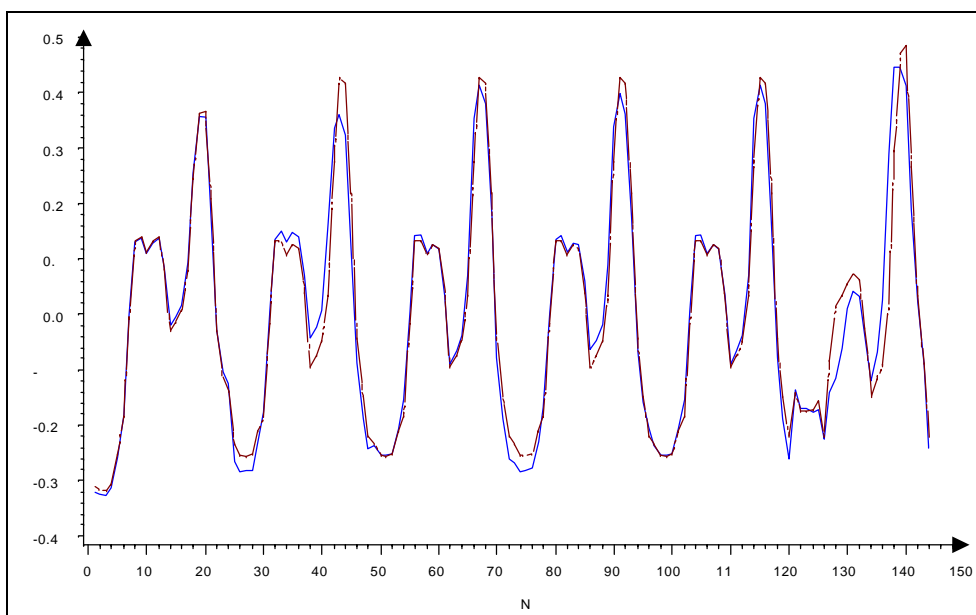


Figure 76 : Représentation du profil P entre le lundi 02 mars et le dimanche 08 mars 97

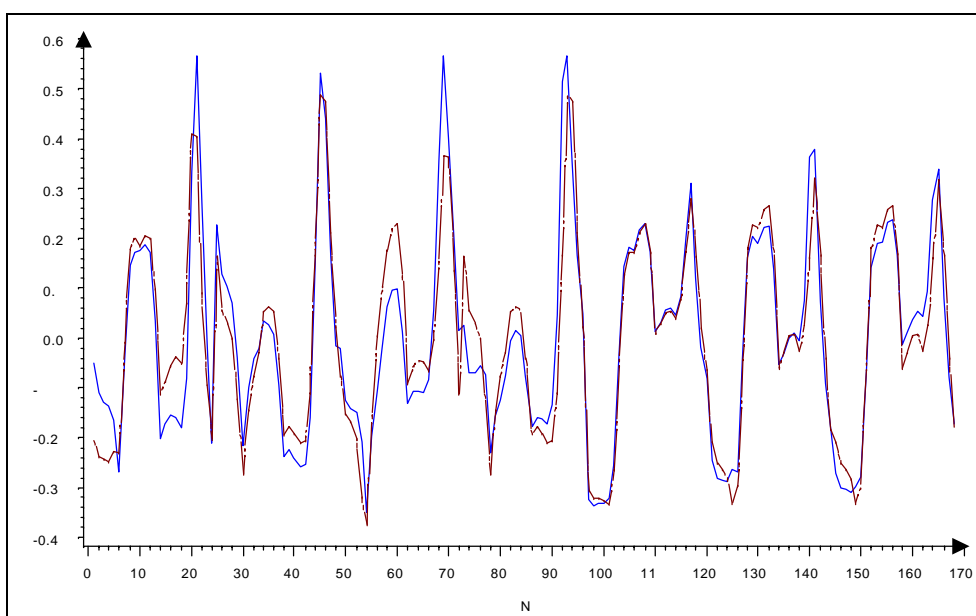


Figure 77 : Représentation du profil P entre le jeudi 30 avril et le mercredi 06 mai 87

⁵ On isole ici l'erreur en niveau pour se concentrer sur les propriétés de l'erreur en forme

Représentation de $s \hat{P}_k$, seul P étant estimé :

On reste dans le cas où M et s sont connus, mais on représente la courbe $s P$ au lieu du profil P .

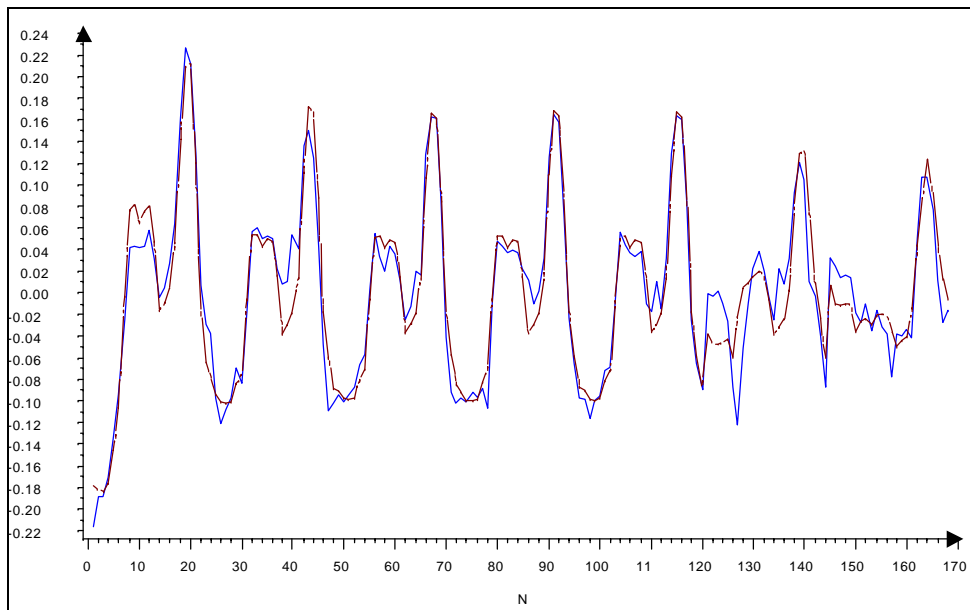


Figure 78 : Représentation de sP entre le lundi 02 mars 87 et le dimanche 08 mars 97

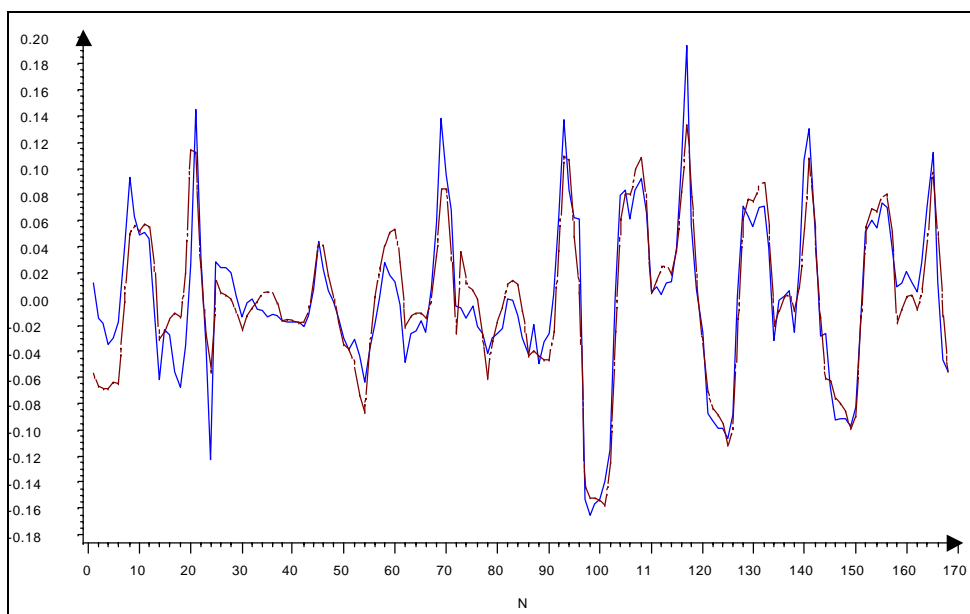
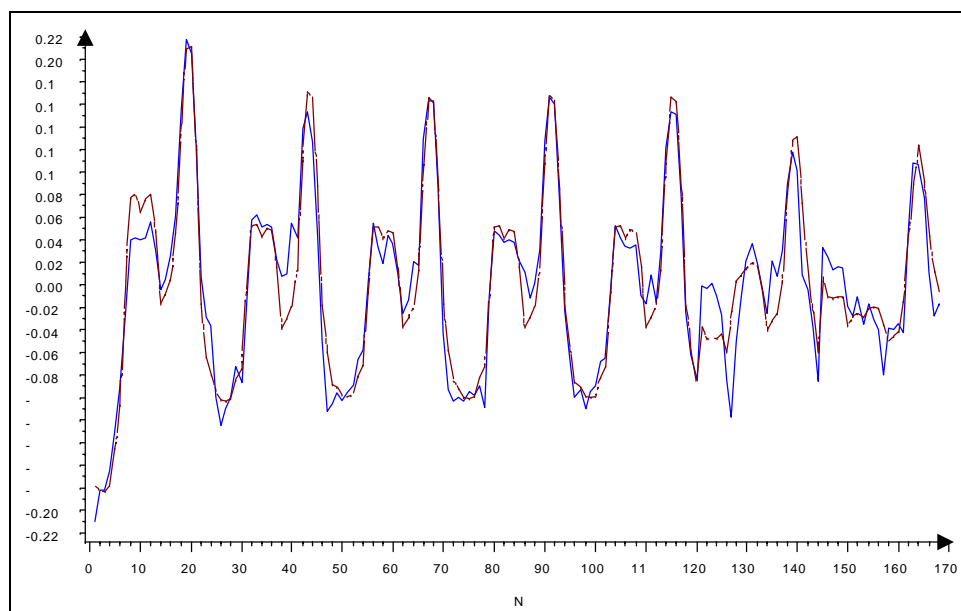


Figure 79 : Représentation de sP entre le jeudi 30 avril 87 et le mercredi 06 mai 87

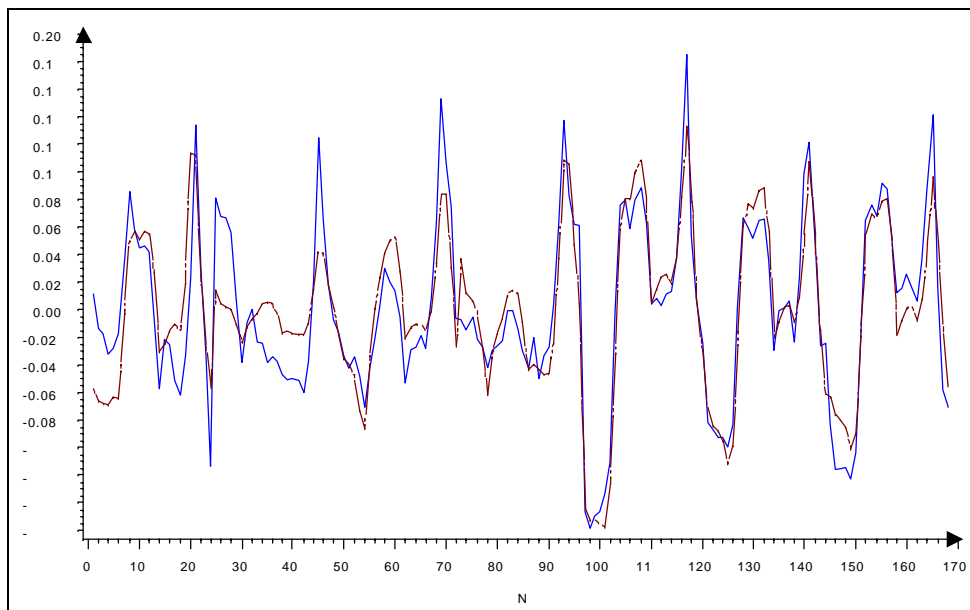
La méthode MEP pour la courbe de charge :

Dans la prévision de la semaine du 30 avril 87 au 06 mai 87, on commet une erreur importante en niveau le 1er mai due à l'armax qui prévoit la moyenne. On constate qu'il ne récupère le « bon niveau » que le lundi 04 mai. Ceci provient du fait que l'on prévoit la moyenne à partir des moyennes des jours précédents et il faut donc plusieurs jours pour restabiliser la prévision. On peut choisir une autre méthode de prévision de la moyenne si l'armax est mal adapté sans que cela ne change réellement la méthode qui repose sur une division de l'information en trois paramètres, le profil étant estimé à partir d'une règle de prévision. On a donc choisi de représenter à la fois la courbe sP et la courbe totale pour isoler la part d'erreur issue de l'estimation de sP de celle de M .

Représentation de $s \hat{P}_k$:

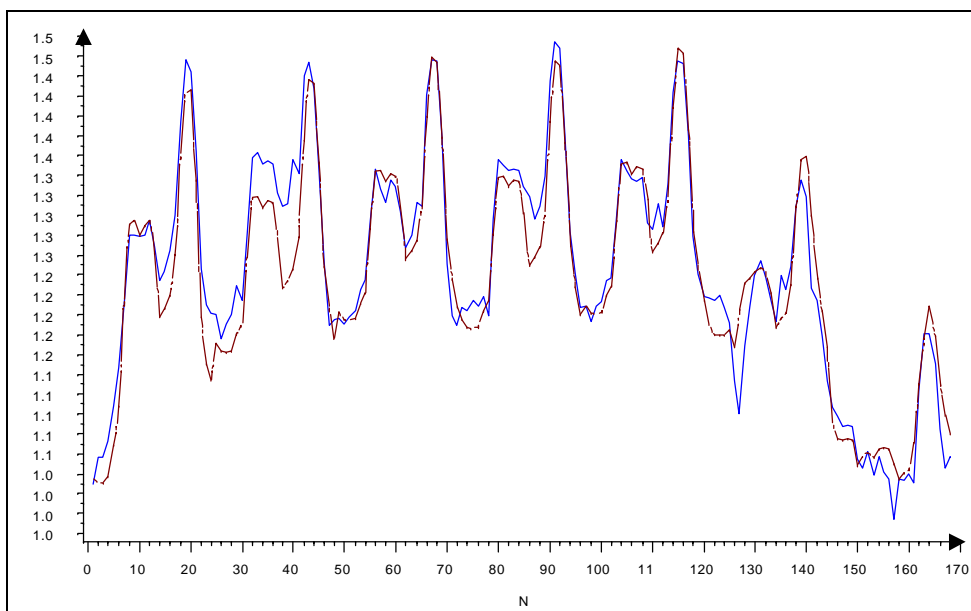


**Figure 80 : Prévision avec la méthode moyenne- écart type- profil quand M est connu et S et P sont estimés:
Représentation de $M+s \hat{P}_k$ entre le lundi 02 mars 87 et le dimanche 08 mars 97**

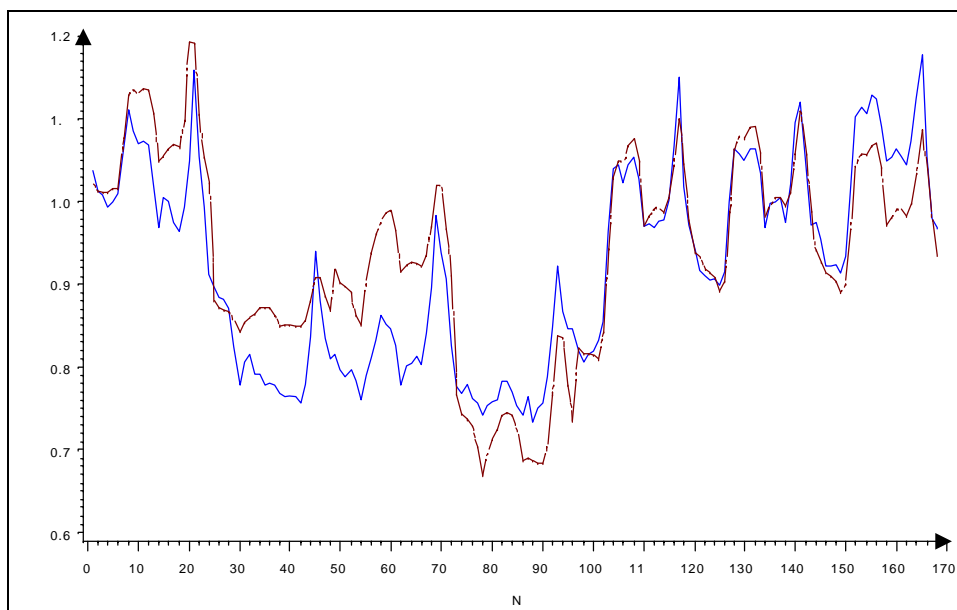


**Figure 81 : Prédiction avec la méthode moyenne- écart type- profil quand M est connu et S et P sont estimés:
Représentation de $M+S \hat{P}_k$ entre le jeudi 30 avril 87 et le mercredi 06 mai 87**

Représentation de $M_k+S_kP_k$ c'est-à-dire la série X_t elle-même :



**Figure 82 : Prédiction avec la méthode moyenne- écart type- profil quand les trois paramètres sont estimés:
Représentation de $M+S \hat{P}_k$ entre le lundi 02 mars 87 et le dimanche 08 mars 97**



**Figure 83 : Prévision avec la méthode moyenne- écart type- profil quand les trois paramètres sont estimés:
Représentation de $M+s \hat{P}_k$ entre le jeudi 30 avril 87 et le mercredi 06 mai 87**

6.2.2 Un modèle de référence pour valider la méthode par comparaison

Le modèle que l'on a choisi pour valider la MEP est un modèle armax qui considère les données brutes, sans considération de notion de courbe de charge. Ses paramètres sont estimés en utilisant l'intégralité de l'information. Ensuite, le jour $k+1$ est prévu à partir des heures antérieures à la 24^{ème} heure du jour k .

Le modèle armax de référence :

Nous avons prévu la consommation électrique horaire du lendemain sur l'échantillon allant du 01 mars 87 au 28 février 88 soit 365 jours. Le modèle est estimé sur la base complète et le passé du 01 mars 87 sert à l'estimation. L'estimation de la première demi-heure est une estimation à un pas, celle de la deuxième à deux pas, l'estimée de la première servant à estimer la deuxième, etc.

On utilise un modèle armax (24,(24) (1 2 3)) qui procède à une différentiation d'ordre 24 et 168 (le jour et la semaine) et de la correction par l'indicatrice du jour férié.

On obtient le modèle:

$$(1 - B^{24})(1 - B^{168})(1 - 0.36 \times B^{24})X_t = 0 + (1 - 0.96 \times B^{24})(1 - 0.90 \times B - 0.69 \times B^2 - 0.35 \times B^3)Z_t - 0.12 \times 1_{\{\text{jour férié}\}}$$

Dans le graphique 84 qui représente la prévision par l'armax de la semaine de mars, on voit la prise en compte du passé et de l'héritage de l'erreur. L'armax s'adapte au changement de forme des samedis et dimanches par différenciation. Le modèle est assez performant pour rendre compte de la forme.

Par contre, le graphique 85 qui représente la prévision par le même modèle armax de la semaine de mai illustre le problème du jour férié. Dans un tel modèle, l'information « le 01 mai est un jour férié » n'est prise en compte que par une variable exogène qui vaut 1 au lieu de 0 quand le jour est férié. C'est donc une compensation en niveau et non en forme qui est appliquée. Le 01 mai 87 est ainsi prévu avec une forme de vendredi. L'erreur en niveau du samedi provient du fait que nous n'avons pas pris en compte les ponts. Par contre, l'erreur en forme provient à la fois du phénomène précédent (le 02 mai 87 est prévu avec une forme de samedi), mais aussi du fait que le modèle tient compte de l'erreur de la veille pour ajuster ce jour. Or dans le cas d'un jour férié ce système adaptatif est parasite et génère une erreur supplémentaire. On constate que le dimanche permet de recadrer le modèle car le dimanche qui suit un jour férié se comporte encore comme un dimanche. Ce modèle armax se restabilise donc plus vite que celui qui estime la moyenne car il tient compte des heures précédentes et non des jours. On pourrait concevoir d'ailleurs un modèle basé sur les heures pour prévoir la moyenne.

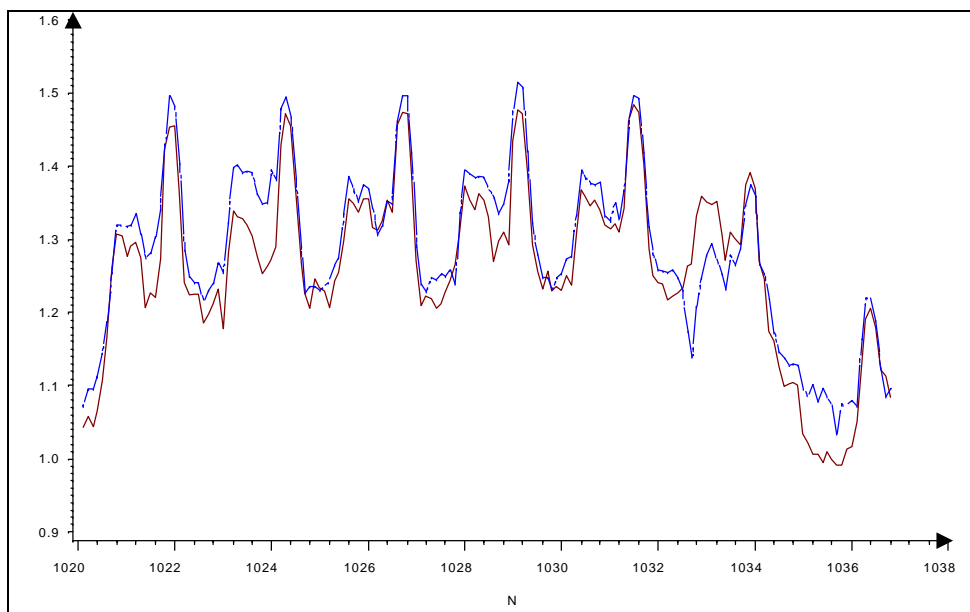


Figure 84 : Prévision avec un armax : lundi 02 mars 87-> dimanche 08mars 97

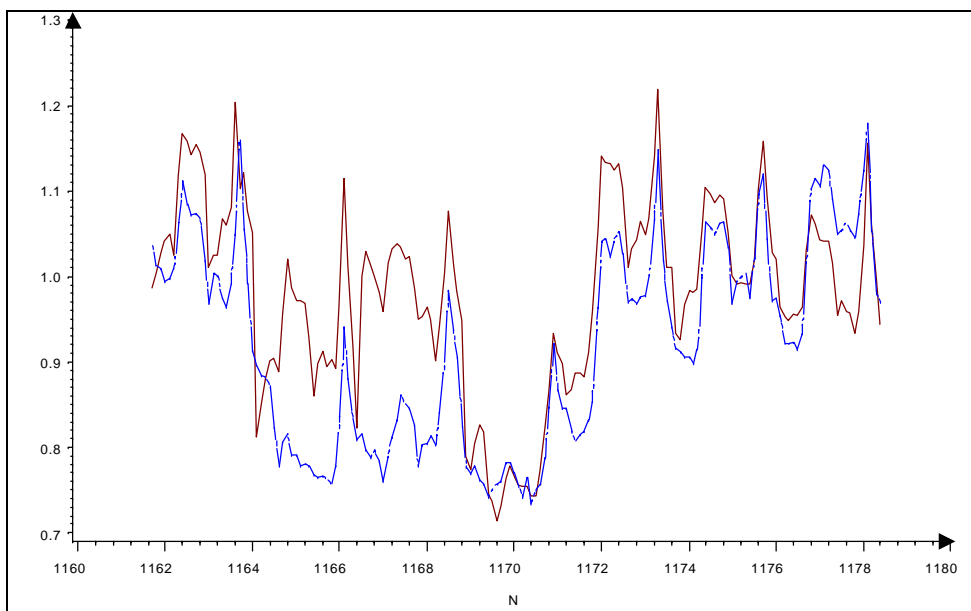


Figure 85 : Prévision avec un armax : du jeudi 30 avril 87 -> mercredi 06 mai 87

6.2.3 Performances de la MEP

Mesure de l'erreur :

L'erreur commise en estimant un vecteur X de dimension 24 par le vecteur \tilde{X} est mesurée par l'écart quadratique classique calculé par la formule suivante:

$$E_k = E(k) = E(X, \tilde{X}) = \frac{1}{24} \times \sum_{h=1}^{24} (X_k^h - \tilde{X}_k^h)^2$$

et l'erreur moyenne de prévision sur la base de données complète (sur N jours) vaut:

$$E = \frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (X_k^h - \tilde{X}_k^h)^2$$

L'écart type de l'erreur journalière $(E_k)_{k=1, \dots, N}$ sur la base complète est noté $s(E_k)$, le maximum de celle-ci est noté $\max(E_k)$.

Le tableau suivant fait le bilan comparatif de l'erreur commise avec le modèle *MEP* respectivement lorsque M et s sont connus (seul P est estimé), M est connu (s et P sont estimés), les trois sont estimés et le modèle armax. Il montre que le modèle *MEP* est globalement bien meilleur que l'armax, la section suivante présente une étude comparative plus fine de cette comparaison. Il met également en évidence les difficultés que l'on a à estimer la moyenne, c'est-à-dire le niveau de consommation. Ce dernier dépend fortement de la température qui n'est pas connue dans notre étude. Néanmoins, on retrouve ce phénomène dans l'étude de la consommation française présente en annexe qui peut provenir du fait que l'effet de la température n'est pas forcément linéaire (par exemple elle n'influence plus le chauffage à partir d'un seuil voisin de 19 degrés). Néanmoins, ce problème ne remet pas en cause la méthode car, comme on l'a expliqué précédemment, un modèle qui prendrait bien en compte la moyenne pourrait être utilisé dans la méthode que l'on présente.

Tableau 6 : Bilan de l'erreur globale

Modèle	Erreur globale
M et S connus	0,00044
M connu	0,00052
modèle <i>moyenne-écart type-profil global</i>	0,00114
Modèle armax	0,0030

6.3 Analyse de l'erreur :

Estimation de la courbe de charge :

On rappelle que l'on a décidé d'estimer la courbe de charge X par $\tilde{X} = \tilde{M} + \tilde{S} \times \tilde{P}$ où \tilde{M} et \tilde{P} est l'estimé du profil à partir d'une règle d'affectation.

Remarque:

$$X - \tilde{X} = (M - \tilde{M}) + S \times P - \tilde{S} \times \tilde{P}$$

$$X - \tilde{X} = (M - \tilde{M}) + S \times (P - \tilde{P}) + (S - \tilde{S}) \times (\tilde{P} - P) + (S - \tilde{S}) \times P$$

On constate que dans l'écart (X, \tilde{X}) , l'erreur sur la moyenne s'additionne à celle combinée du profil et de l'écart type. Par contre, l'influence de l'erreur sur le profil $(P - \tilde{P})$ sur l'écart (X, \tilde{X}) est proportionnel à $(S + (S - \tilde{S}))$ qui est de l'ordre de S .

La particularité de notre méthode étant l'introduction du profil, la remarque précédente nous invite à étudier séparément les erreurs sur P , sP et $X=M+sP$ pour connaître la contribution de chacune des erreurs à l'erreur globale.

La moyenne de consommation électrique horaire et l'écart type associé valent respectivement: $\bar{M}=1,017$ et $\bar{S}=0,076$.

6.3.1- Analyse de l'erreur de prévision du profil

Selon, le choix de mesure de l'erreur précédente, on a:

$$E_k = \frac{1}{24} \times \sum_{h=1}^{24} (P_k^h - \tilde{P}_k^h)^2$$

$$E = \frac{1}{N} \sum_{j=1}^N E(k) = \frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (P_k^h - \tilde{P}_k^h)^2 = 0.0018$$

$$s(E_k) = 0.0033$$

l'écart type de cette erreur valant $s(E_k) = 0.0033$ et l'erreur maximale étant $\max(E_k) = 0.06$.

L'erreur sur le profil est la même quelle que soit l'échéance de prévision (un jour ou plus). On fait une erreur globale de 0,0018 qui se cumulera avec les erreurs sur la moyenne et le profil.

Etude comparative par type de jour :

On constate sur le tableau 7 qu'en moyenne, on fait deux fois plus d'erreurs le dimanche que les jours de semaine et quatre fois plus le samedi. L'erreur sur le dimanche peut s'expliquer par le fait que les jours fériés sont mélangés aux dimanches dans la règle d'affectation des profils. Le lundi est le jour travaillé pour lequel on fait le moins d'erreur, ceci provenant du fait que la règle lui affecte un profil particulier.

La répartition de l'erreur illustrée par le graphique 86 aboutit aux mêmes conclusions. En effet, la concentration sur les erreurs faibles est plus dense pour *les lundis et les jours de semaine* que *les samedis et les dimanches*. La fonction de répartition de l'erreur des dimanches présente la particularité de posséder deux pointes en 0,0005 et 0,0025 qui laisse présager la présence de deux populations. Les graphiques 88 et 89 montrent que l'on prévoit très bien les profils des jours fériés et moins bien les dimanches. Etablir une nouvelle règle

d'affectation qui distingue les deux populations peut améliorer l'estimation des dimanches, mais risque de diminuer la robustesse. Les samedis sont les seuls jours où l'on fait une très grande erreur de prévision du profil. L'existence en Pologne de samedis travaillés peut constituer une explication, mais nous n'avons pas les moyens de vérifier cette hypothèse, puisque nous ne disposons pas du calendrier des samedis ouvrés.

Les moyennes des erreurs sur les mois sont dans un rapport de 1 à 3. Le mois d'avril, qui constitue un mois de transition entre l'hiver et l'été est le moins bien estimé. Les mois de mai et août contribuent également plus à l'erreur que les autres. La répartition de l'erreur en terme de fréquences est plutôt équilibrée sur les mois comme l'illustre le graphique 87.

On pouvait s'attendre à ce que la répartition de l'erreur sur les années répercute la tourmente politique polonaise sur la période 86-93. En fait, seule l'année 86 se démarque par une erreur moyenne plus importante.

On peut mesurer la contribution de l'heure h en définissant $E(h)$ comme suit:

$$E(h) = \frac{1}{N} \times \sum_{k=1}^N (P_k^h - \tilde{P}_k^h)^2$$

On constate que l'on fait moins d'erreurs en moyenne aux heures de travail (la plage horaire de 10h00 à 16h00) comme le montre le graphique 90 qui représente $E(h)$ en fonction de h . On fait, par contre, une plus grande erreur au moment de la reprise et de la sortie du travail, ainsi que la soirée (l'erreur à 18h00 et 21h00 est deux fois plus grande que la moyenne).

Conclusions:

On fait d'avantage d'erreurs à des moments au cours desquels la consommation électrique est plus faible comme *les samedis, les dimanches* et les mois *avril, mai* et *août* en raison respectivement d'une activité industrielle plus faible et de la suppression du chauffage. Ces catégories coïncident avec un écart type s plus faible. On a déjà vu que dans l'écart entre la courbe de charge et son estimation, l'erreur sur les profils est pondérée par s . On peut donc espérer un rééquilibrage des erreurs.

Tableau 7 : Erreur moyenne par type de jour

JOUR	ERREUR
dimanche	0.0022
lundi	0.00084
mardi	0.0012
mercredi	0.0010
jeudi	0.0012
vendredi	0.0012
samedi	0.0048

Tableau 8 : Erreur moyenne par mois

MOIS	ERREUR
1 ⁶	0.0010
2	0.0012
3	0.0015
4	0.0033
5	0.0030
6	0.0014
7	0.0015
8	0.0023
9	0.0015
10	0.0017
11	0.0017
12	0.0010

Tableau 9 : Erreur moyenne par année

ANNEE	ERREUR
1986	0.0025
1987	0.0018
1988	0.0017
1989	0.0016
1990	0.0017
1991	0.0016
1992	0.0015
1993	0.0017

⁶ Le mois 1 correspond au mois de janvier

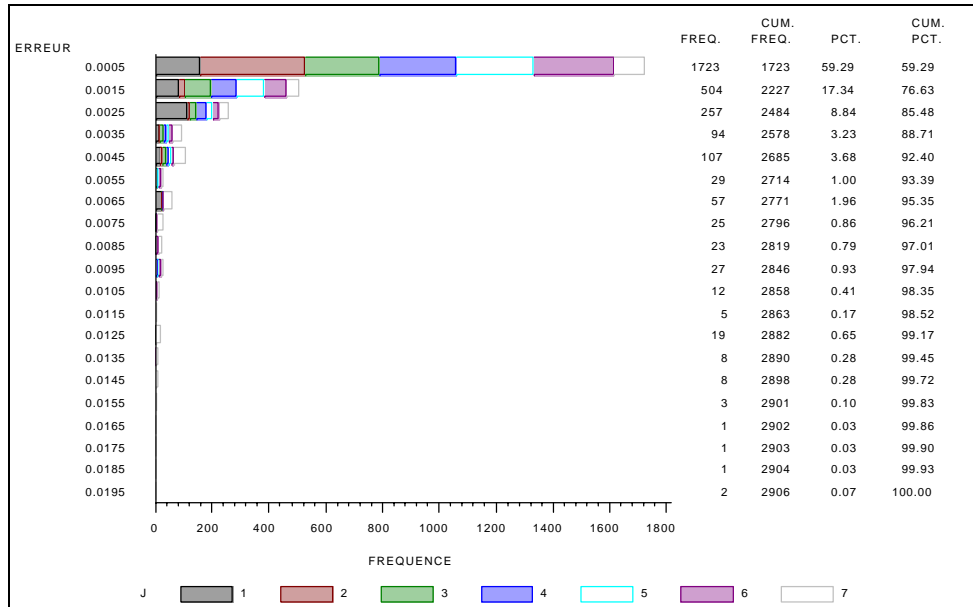


Figure 86 : Erreur pour chaque catégorie de type de jour

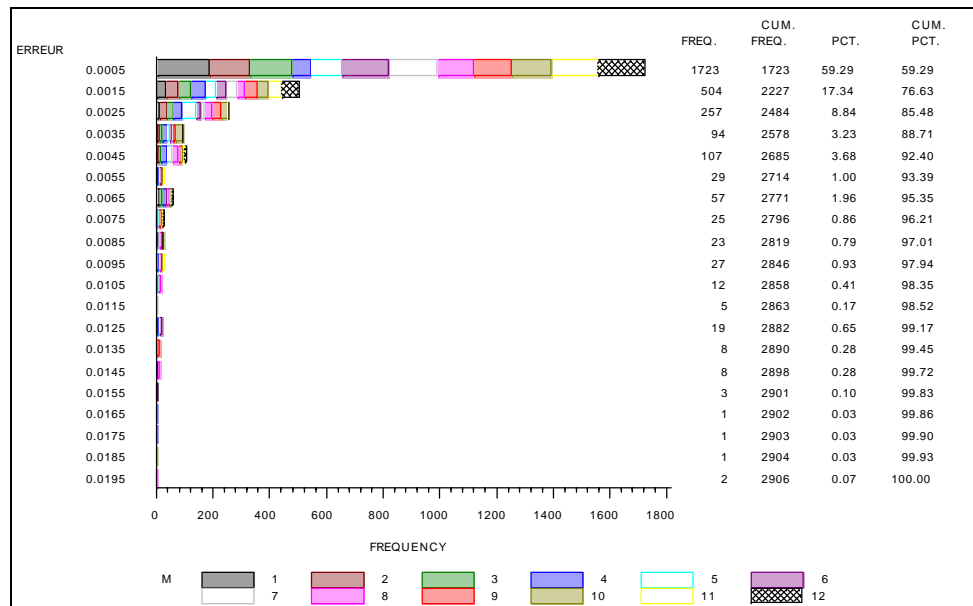


Figure 87 : Erreur pour chaque catégorie de mois

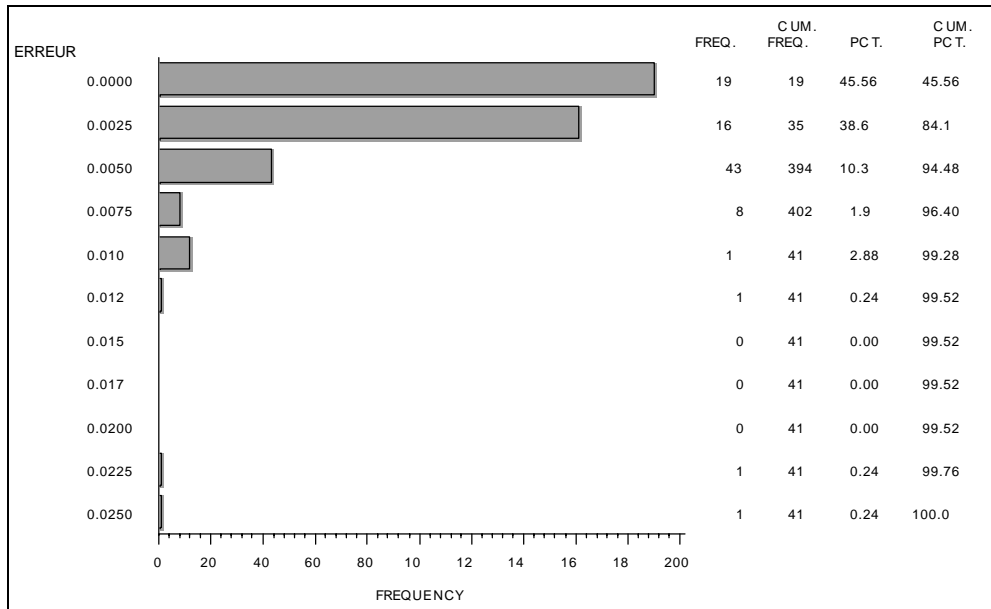


Figure 88 : Répartition des dimanches

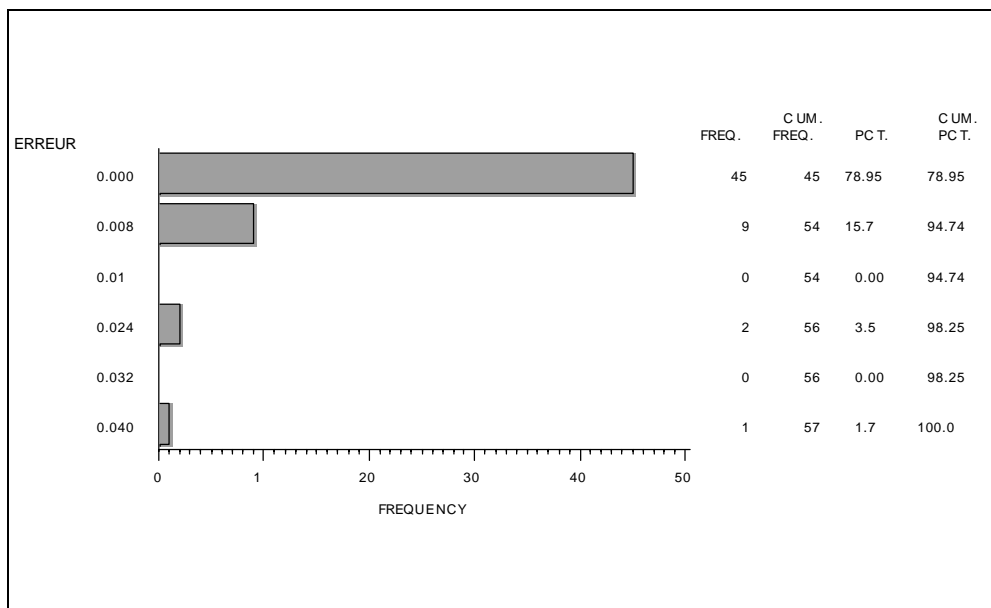


Figure 89 : Répartition de l'erreur des jours fériés

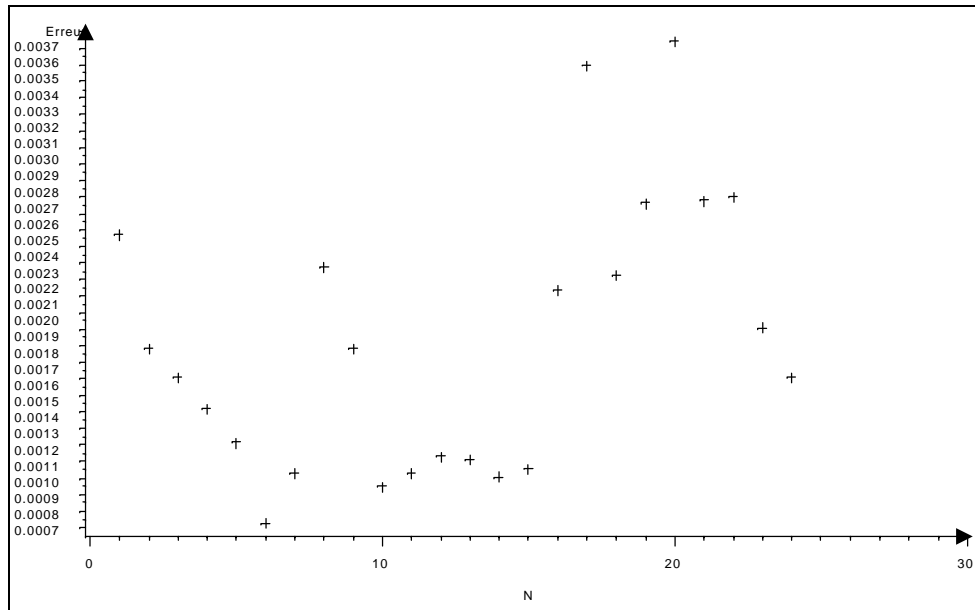


Figure 90 : Erreur moyenne pour chaque heure

Localisation de l'erreur sur la carte :

On code l'erreur que l'on fait sur l'estimation du profil par la variable qualitative Q . Les 500 plus fortes erreurs sont codées magenta, les autres noires. On effectue ensuite la cartographie de la variable Q selon la méthode présentée au chapitre I.3.6 pour obtenir la carte de la figure 92. Si la répartition des erreurs sur la carte est uniforme, le camembert de chaque unité doit avoir le même découpage (« angle magenta » = $\pi \times 500 / 2900$, « angle noir » = $\pi \times 2400 / 2900$), soit celui du graphique 91 qui constitue ainsi le camembert témoin.

Une unité qui produit plus d'erreurs que la normale voit sa part magenta s'accroître par rapport au modèle précédent. Parmi ces unités, on retrouve beaucoup d'unités qui mélangent les samedis aux jours ouvrés, ce qui ajoute à notre conviction que la connaissance du calendrier des samedis travaillés en Pologne permettrait d'affiner notre prévision.

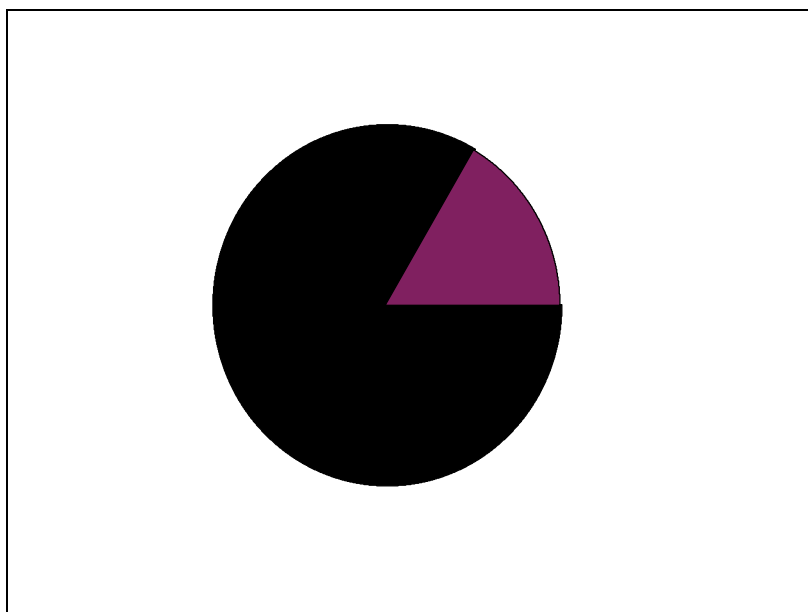


Figure 91 : Camembert d'une classe dont la proportion d'éléments appartenants aux 500 plus grandes erreurs est identique à celle de la population globale

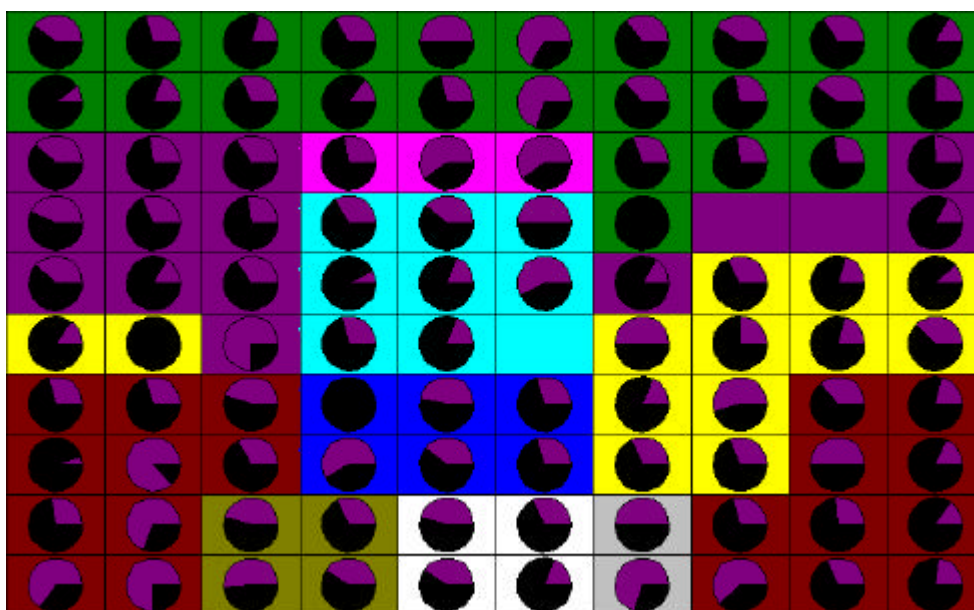


Figure 92 : Répartition des 500 plus fortes erreurs dans les 100 classes

6.3.2 Analyse de l'erreur de prévision de la courbe de charge

Erreur sur l'estimation de s P à s connu

Quand on se place dans le cadre où s est connu, les quatre types d'erreurs présentés au chapitre 4.1 appliqués à la prévision de s P deviennent

$$E_k = \frac{1}{24} \times \sum_{h=1}^{24} (\mathbf{s}_k \times P_k^h - \mathbf{s}_k \times \tilde{P}_k^h)^2$$

$$\text{et } E = \frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (\mathbf{s}_k \times P_k^h - \mathbf{s}_k \times \tilde{P}_k^h)^2 = 0.00044$$

$$s(E_k) = 0.00073$$

$$\text{et } \max(E_k) = 0.0159$$

Commentaires :

L'erreur moyenne est divisée par 4.

L'erreur de la moyenne de l'erreur par type de jour et de mois confirme le rééquilibrage entre les différentes catégories annoncé dans la section 6.3.1. En effet, les dimanches, mois d'août, de mai et d'avril rentrent dans le rang. Les samedis restent plus sujets à l'erreur mais dans une proportion ramenée de 2,6 à 1. Le mois de décembre devient celui qui produit le plus d'erreur (1,8 fois la moyenne).

Les répartitions des erreurs sont identiques pour toutes les catégories sauf les samedis qui sont décalés vers les grandes valeurs.

Conclusion:

Comme attendu, on constate un nivellement entre les catégories.

Tableau 10 : Erreur moyenne par type de jour

JOUR	ERREUR
Dimanche	0.00039
lundi	0.00047
mardi	0.00036
Mercredi	0.00037
jeudi	0.00040
Vendredi	0.00041
samedi	0.00067

Tableau 11 : Erreur moyenne par mois

MOIS	ERREUR
1	0.00041
2	0.00039
3	0.00047
4	0.00052
5	0.00038
6	0.00028
7	0.00021
8	0.00030
9	0.00040
10	0.00057
11	0.00052
12	0.00079

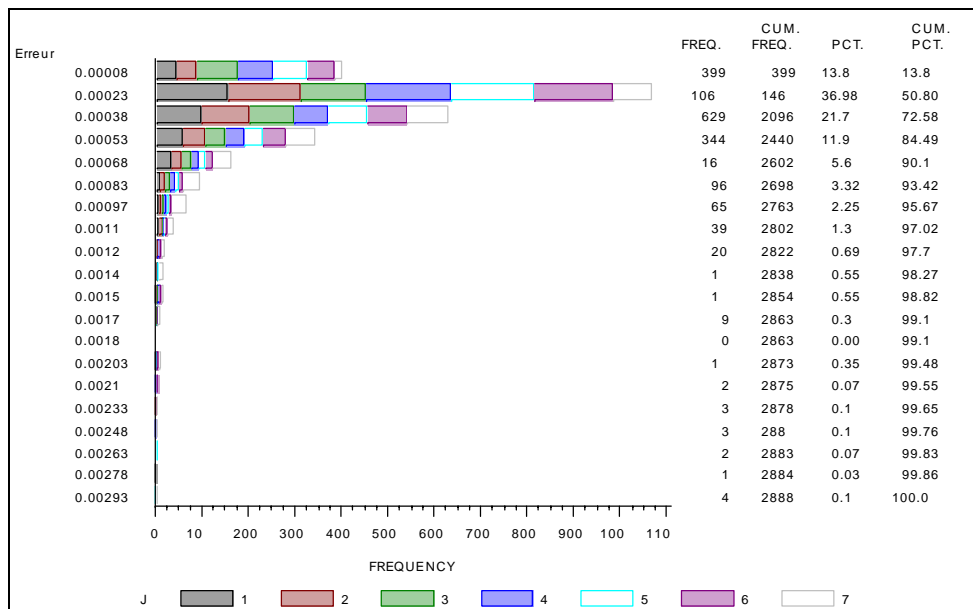


Figure 93 : Erreur par type de jour

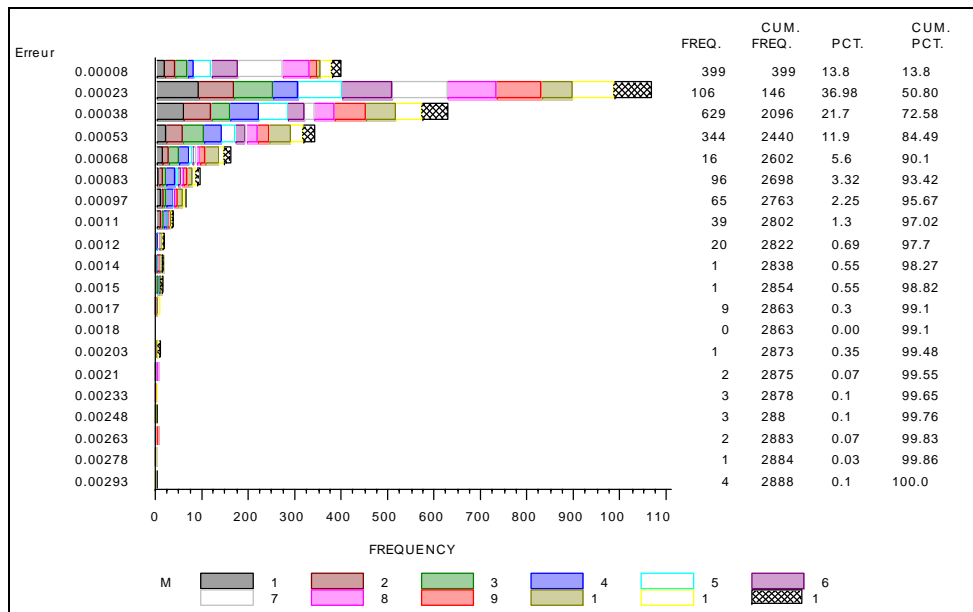


Figure 94 : Erreur par mois

Erreur d'estimation de s P à s estimé :

Dans le cas où s est estimé par un arma séparément de P , les erreurs deviennent :

$$E_k = \frac{1}{24} \times \sum_{h=1}^{24} (s_k P_k^h - \tilde{s}_k \tilde{P}_k^h)^2$$

$$\text{et } E = \frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (s_k P_k^h - \tilde{s}_k \tilde{P}_k^h)^2 = 0.00052$$

$$s(E_k) = 0.00073$$

$$\text{et } \max(E_k) = 0.011$$

L'estimation de σ apporte une augmentation de l'erreur E de 18% mais l'écart type de l'erreur n'a pas varié. C'est une augmentation de l'erreur E_k en niveau. La fonction de répartition de l'erreur E_k est très proche de celle avec s connu. L'étude par catégories amène aux mêmes conclusions que celles de la section 6.3.2. Seul les lundis produisent relativement plus d'erreur. Ceci est dû à une perte de la stationnarité entre les jours ouvrés, les samedis et les dimanches qui n'est pas annulé par la différentiation d'ordre 7 (une semaine).

Conclusion:

L'estimation de S par un arma modifie peu l'erreur et sa fonction de répartition. A ce stade de l'estimation, c'est donc davantage le profil qui crée l'erreur. Mais comme on l'a déjà remarqué, l'estimation de S est une estimation à 1 jour alors que la méthode d'estimation du profil fournit la même erreur quelle que soit l'échéance de prévision.

Tableau 12 : Erreur moyenne par type de jour

JOUR	ERREUR
dimanche	0.00045
lundi	0.00060
mardi	0.00046
mercredi	0.00044
jeudi	0.00046
vendredi	0.00048
samedi	0.00077

Tableau 13 : Erreur moyenne par mois

MOIS	ERREUR
1	0.00055
2	0.00044
3	0.00055
4	0.00063
5	0.00053
6	0.00035
7	0.00027
8	0.00041
9	0.00047
10	0.00061
11	0.00063
12	0.00084

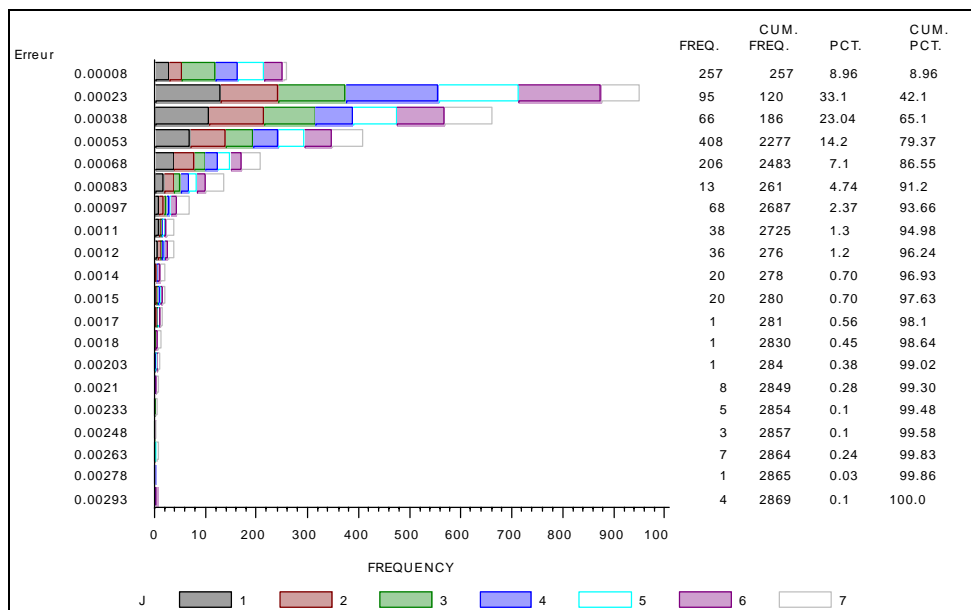


Figure 95 : Erreur par type de jour

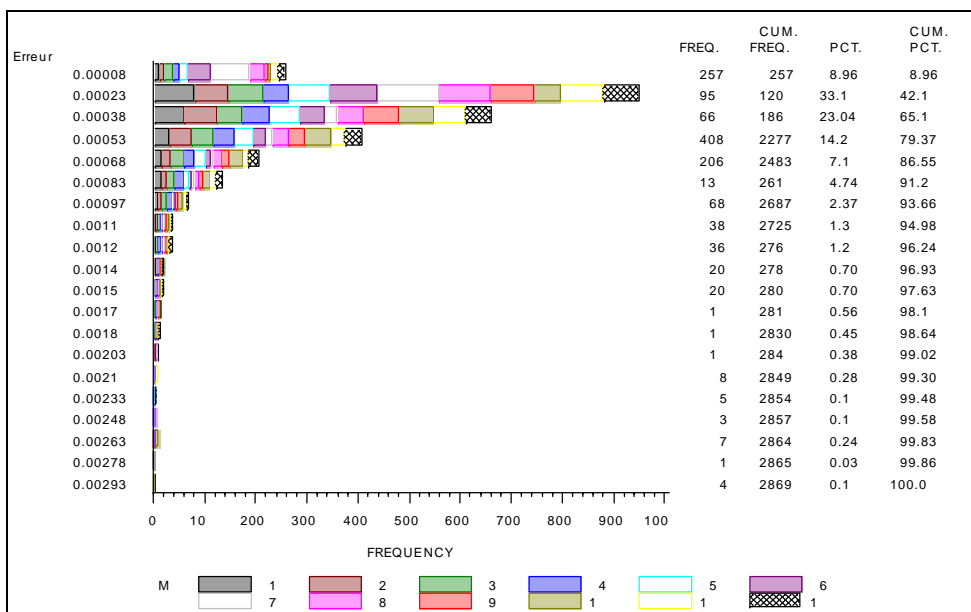


Figure 96 : Erreur par mois

Erreur d'estimation de $M + \mathbf{s} P$ à \tilde{M} et $\tilde{\mathbf{s}}$ estimés :

Dans ce chapitre, on estime la courbe de consommation complète, M et \mathbf{s} étant estimés par un modèle arma. Les erreurs deviennent :

$$E_k = \frac{1}{24} \times \sum_{h=1}^{24} (M + \mathbf{s}_k P_k^h - \tilde{M} + \tilde{\mathbf{s}}_k \tilde{P}_k^h)^2$$

$$\text{et } E = \frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (M + \mathbf{s}_k P_k^h - (\tilde{M} + \tilde{\mathbf{s}}_k \tilde{P}_k^h))^2 = 0.00114$$

$$\mathbf{s}(E_k) = 0.003$$

$$\text{et } \max(E_k) = 0.087$$

Commentaire :

L'erreur moyenne augmente beaucoup, son écart type aussi, la fonction de répartition se décale vers les grandes valeurs et la queue de distribution s'allonge. Ceci peut s'expliquer par le manque d'information sur la température qui nous pénalise lourdement dans l'évaluation du niveau de consommation électrique. Ce qui pourrait expliquer que l'erreur est deux fois plus grande en hiver.

Remarque :

Toute méthode linéaire qui prévoirait bien la consommation électrique serait performante pour estimer la moyenne et pourrait donc être adoptée par notre méthode pour remplacer le modèle arma. En effet, la particularité de la *méthode MEP* réside dans l'estimation de la journée dans son ensemble et le résumé de 24 variables en trois paramètres ainsi que l'utilisation de l'algorithme de Kohonen pour construire la règle d'affectation qui permet d'estimer le profil. Il est évident que notre méthode permet de remplacer le modèle arma par tout autre modèle, en particulier un modèle issu des réseaux de neurones comme le perceptron multicouches tout à fait adapté dans le cadre de la prévision à un pas d'un seul paramètre.

Dans la suite, nous allons comparer notre méthode à celle d'un arma qui sera naturellement confronté au même problème du manque d'information sur la température.

Tableau 14 : Erreur moyenne par type de jour

JOUR	ERREUR
dimanche	0.0012
lundi	0.0011
mardi	0.00076
mercredi	0.00079
jeudi	0.00086
vendredi	0.0010
samedi	0.0022

Tableau 15 : Erreur moyenne par mois

MOIS	ERREUR
1	0.0015
2	0.0010
3	0.0012
4	0.0015
5	0.0012
6	0.00084
7	0.00064
8	0.00084
9	0.00084
10	0.0011
11	0.0014
12	0.0017

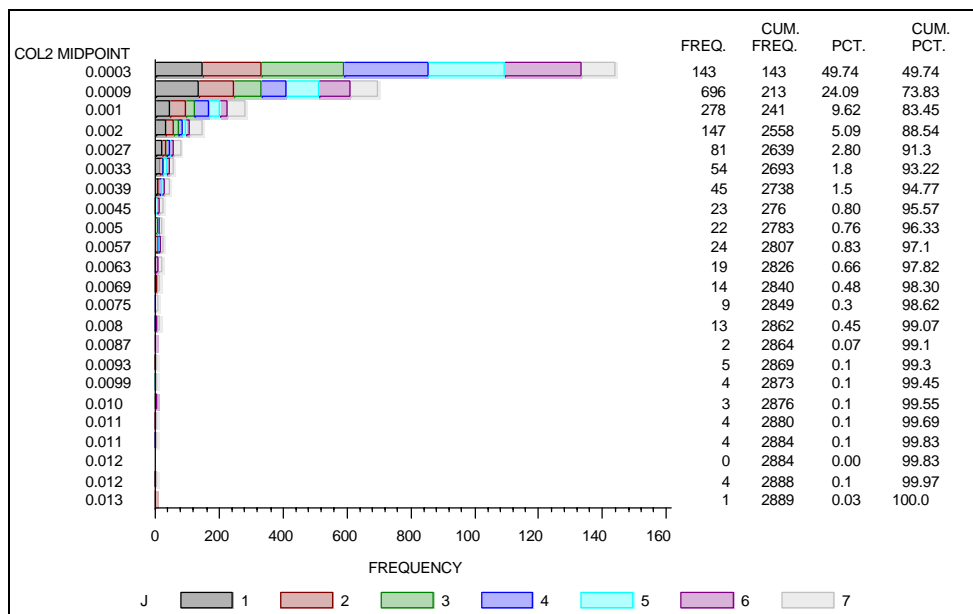


Figure 97 : Erreur par type de jour

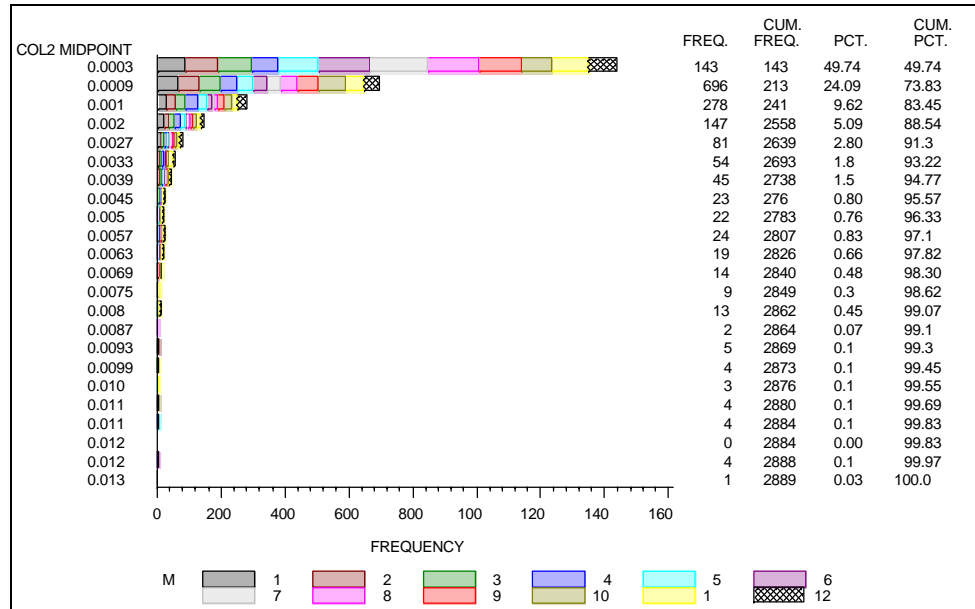


Figure 98 : Erreur par mois

Erreur commise avec le modèle armax :

On rappelle que nous avons prévu la consommation électrique horaire du lendemain sur l'échantillon allant du 01 mars 87 au 28 février 88 soit 365 jours à l'aide d'un modèle armax (24,(24) (1 2 3)) qui procède à une différentiation d'ordre 24 et 168 et corrige les jours fériés par une fonction indicatrice. Les paramètres sont estimés sur la base complète et le passé du 01 mars 87 sert à l'estimation. Le modèle s'écrit :

$$\begin{aligned} (1 - B^{24})(1 - B^{168})(1 - 0.36 \times B^{24})X_t = \\ 0 + (1 - 0.96 \times B^{24})(1 - 0.90 \times B - 0.69 \times B^2 - 0.35 \times B^3)Z_t - 0.12 \times 1_{\{\text{jour férié}\}} \end{aligned}$$

Il fournit l'erreur suivante où on a regroupé l'erreur par jour.

$$\frac{1}{N} \times \frac{1}{24} \times \sum_{k=1}^N \sum_{h=1}^{24} (X_k^h - \tilde{X}_k^h)^2 = 0.0030$$

On réalise une erreur moyenne nettement supérieure (presque trois fois) à celle obtenue avec la méthode qui nous intéresse, essentiellement due aux grandes erreurs. En effet, la fonction de répartition de la figure 99 montre une concentration de l'erreur vers 0,0008 comparable à celle de notre méthode mais la queue de distribution est bien plus longue. Sur cette même figure, on vérifie que l'erreur se répartit de la même façon sur les différents jours, hormis les samedis, alors que la moyenne de l'erreur par type de jour présentée dans le tableau 16 montre aussi un écart pour les jeudis et les vendredis, ce qui confirme que ce sont les grandes valeurs qui rendent l'erreur moyenne importante.

Tableau 16 : Erreur moyenne par type de jour

JOUR	ERREUR
dimanche	0.0015
lundi	0.0023
mardi	0.0015
mercredi	0.0015
jeudi	0.0035
vendredi	0.0040
samedi	0.0066

Tableau 17 : Erreur moyenne par mois

MOIS	ERREUR
1	0.0048
2	0.0012
3	0.0027
4	0.0051
5	0.0045
6	0.0031
7	0.0016
8	0.0016
9	0.0016
10	0.0023
11	0.0012
12	0.0059

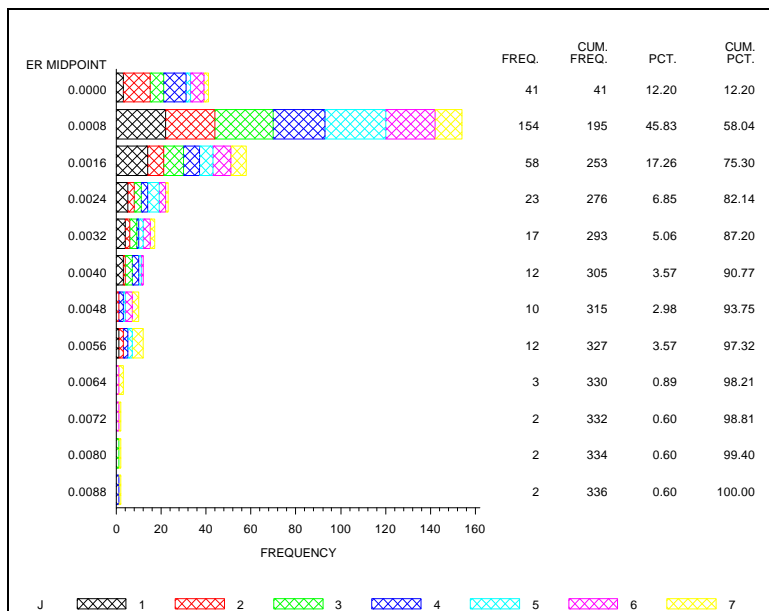


Figure 99 : Erreur par type de jour

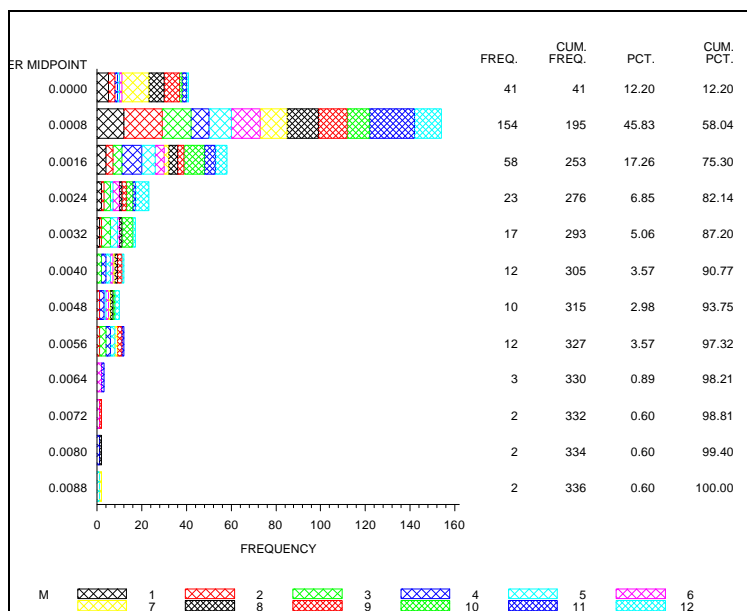


Figure 100 : Erreur par mois

Conclusions :

La méthode MEP est meilleure que le modèle armax lorsque l'on considère l'erreur globale, mais elle fait aussi moins de grandes erreurs et est plus performante pour chaque type de jour. Dans l'étude de la consommation électrique française fournie en annexe, on a aboutit aux mêmes conclusions. De plus, on y a prévu l'année qui suivait la période qui a permis d'établir la règle (en particulier qui a servi à construire la carte de Kohonen) avec un résultat très satisfaisant puisque si l'erreur était supérieure à la moyenne elle était meilleure que certaines années de la base d'étalonnage.

7. Quelques réflexions sur la méthode.

Certains aspects ont été éludés pendant la présentation de la méthode pour l'alléger et la simplifier. On va maintenant les exposer dans ce chapitre. En particulier, on reviendra sur le choix d'estimer le profil par un vecteur normé et on indiquera certaines options qui ont été prises en compte sans être retenues dans l'étude de la consommation électrique polonaise. On présentera ensuite les outils de contrôle visuels fournis par les cartes de Kohonen qui préviennent certains risques d'erreur issus de la construction de la règle de prévision. On indiquera également une limite de cette méthode et plus particulièrement du choix de classer les profils.

7.1 Normalisation de l'estimé du profil et renormalisation des poids à chaque itération de l'apprentissage.

Normalisation de l'estimé du profil :

Dans la prévision de la consommation électrique, on a choisi une règle d'affectation qui associe à chaque type de jour un profil constitué d'un barycentre de U vecteurs¹ renormalisé (voir à la section 5.3 le paragraphe 5)). En effet, le barycentre est de norme inférieure à 1. Si on se place dans le cadre de l'exemple, on veut estimer X par $\tilde{M} + \tilde{S}\tilde{P}$ où \tilde{M} et \tilde{S} sont des estimateurs de M et de \mathbf{s} . Si on ne renormalise pas \tilde{P} , on a davantage de risque de sous estimer l'amplitude ce qui n'est pas souhaité. En particulier, dans la consommation d'électricité, il n'est pas équivalent de mal estimer le maximum, le minimum ou une valeur moyenne car le coût de l'erreur n'est pas la même (par exemple le maximum peut nécessiter l'apport de centrales thermiques, etc.).

Description du phénomène :

Si on prévoit le profil P_t du type de jour t par le barycentre \hat{P}_t des vecteurs codes (qui sont de norme 1), il sera de norme strictement inférieure à 1 et n'appartiendra donc pas à la sphère unité, ce phénomène a déjà été décrit au paragraphe 6.1.2.1. Il est donc à une distance \mathbf{e}_t de la sphère unité. Le nombre de types de jours étant fini, on note \mathbf{e} le minimum des \mathbf{e}_t .

Pour chaque jour k , P_k est sur la sphère unité, d'où :

Si $\hat{P}_{t(k)}$ est l'estimé du type du jour de k ,

$$\forall k \quad \|P_k - \hat{P}_{t(k)}\| > \mathbf{e}_t \geq \mathbf{e}.$$

¹ Les U vecteurs codes G_u issus de l'apprentissage de l'algorithme de Kohonen

Si on ne normalise pas les vecteurs \hat{P}_t , on commet donc une erreur systématique sur le profil, supérieure à \mathbf{e} . De plus, quand on regarde sa conséquence sur l'erreur de prévision de la courbe de charge à M et \mathbf{s} connus, on constate une erreur systématique supérieure à \mathbf{se} . En effet, à M et \mathbf{s} connus, on a :

$$\|X - \tilde{X}\| = \|M + \mathbf{s}P - M - \mathbf{s}\hat{P}\| = \mathbf{s} \|P - \hat{P}\|$$

et $\|X - \tilde{X}\| > \mathbf{se}$

C'est pourquoi, dans les études de la consommation électrique polonaise et française, on a choisi $\tilde{P}_t = \frac{\hat{P}_t}{\|\hat{P}_t\|}$.

On constate de plus que $\forall h \quad \left| \frac{\hat{P}_k^h}{\|\hat{P}_k\|} \right| > |\hat{P}_k^h|$ et par conséquent, $\forall h \quad |\tilde{P}_k^h| > |\hat{P}_k^h|$

En estimant le profil P par \tilde{P} au lieu de \hat{P} , on diminue le risque de sous évaluer la consommation électrique et par conséquent de ne pouvoir fournir les usagers ou de devoir mettre en marche des centrales d'appoint plus coûteuses.

7.2 Options techniques :

Dans l'étude de la consommation électrique, la méthode *MEP* impose de découper la série chronologique en séquences de même durée. L'objectif de prévoir les 24 heures suivantes justifie de choisir des périodes de 24 heures. Néanmoins, le fait de prendre les valeurs de 1h00 à 24h00 peut être remis en question. C'est le problème du choix de l'origine du jour. La *MEP* propose aussi un découpage de l'information journalière en 3 paramètres (moyenne, écart type et profil) dont des variantes sont présentées dans ce chapitre.

Choix de l'origine du jour :

On a défini la *basse fréquence* comme la séquence constituée d'un nombre fixe p ($p=24$) d'unités de temps. On appelle *événement* ce qui se passe au cours de cette séquence. On attend de la classification qu'elle en fasse ressortir les propriétés de répétition et de variation. Il arrive aussi, de façon intuitive, que l'on veuille recouvrir sous le même terme, *événement*, quelque chose qui se produit entre deux instants où "il ne se passe rien". Ceci correspond à la volonté de regrouper dans sa continuité tout ce qui est dû au même phénomène. Par exemple, si les vendredis et les samedis partagent l'activité nocturne du vendredi soir au samedi matin, les variations de cette période provoqueront en même temps des changements de classes pour les vendredis et les samedis. On peut préférer que seul un type de jour (les vendredis) soit concerné par le comportement *de la sortie du vendredi soir*. Pour faire coïncider notre définition de l'événement avec cette notion intuitive, il suffit de choisir comme origine du jour, si elle existe, une heure où "il ne se passe rien", c'est à dire une heure où la variable X est constante. De façon plus générale, on peut choisir l'heure où la variance de la variable est minimale. C'est cette option qui a été choisie dans l'étude de la consommation nationale d'électricité où les jours commencent à 4h00.

Autre définition du profil :

Dans la prévision de la courbe de charge, on a décidé de découper l'information selon 3 paramètres à savoir la moyenne, l'écart type et le profil, comme défini au chapitre 5.2. On peut imaginer un découpage avec d'autres caractéristiques, qui peut être motivé par les propriétés de ces derniers.

Par exemple, la bonne stabilité du minimum de la courbe de charge électrique journalière aurait pu nous faire préférer le découpage suivant :

$$X_k = \min_k + (\max_k - \min_k) \times F_k$$

où $\min_k = \min_{h=1, \dots, 24} (X_k^h)$

et $\max_k = \max_{h=1, \dots, 24} (X_k^h)$

Si on appelle *forme* (au lieu de profil) la courbe corrigée du niveau et de l'amplitude, F_k est la forme associée à une correction en niveau par le minimum et en amplitude par l'écart entre le minimum et le maximum.

Lors des études de la consommation électrique, une correction par le minimum et l'écart min-max a été envisagée, mais l'écart min-max ne présente pas de propriété aussi intéressante que la stabilité du minimum. Dans notre étude, l'indépendance du profil et de la température est une propriété essentielle qui justifie à posteriori notre découpage.

7.3 Nécessité d'un outil de contrôle visuel des effets nuisibles dus aux regroupements de profils effectués dans la règle de prévision

Dans la méthode *MEP*, la classification sert à établir une règle de prévision du profil à partir d'une typologie. On attend donc de celle-ci, outre un bon regroupement des jours en classes, qu'elle se prête à développer des moyens de sélectionner les critères de définition des types de jours. La qualité de cette prévision dépend en effet de cette sélection. En particulier, on préfère que le profil \tilde{P}_t associé au type de jour t soit classé dans une classe où apparaissent des formes de type t . L'exemple suivant illustre un cas où une mauvaise typologie (ici due à un manque de critère dans la définition du type) engendre une erreur de prévision importante.

On montrera ensuite comment la représentation en cartes de la notion de voisinage entre les représentants des classes inhérente à la classification de Kohonen constitue un outil visuel qui permet d'évaluer si les critères qui définissent les types de jours sont suffisants, et ce dès l'établissement de la règle et non au moment de l'analyse de l'erreur.

Exemple théorique où une classification aboutit à un échec :

Supposons que l'on utilise une autre classification que celle de l'algorithme de Kohonen dans un cas concret de \mathbf{R}^2 (2 valeurs par jour). Une variable qualitative permet de distinguer 3 types (les dimanches, samedis et jours de semaine ouvrés) qui seront ceux retenus. La classification répartit les profils des jours en 3 classes avec pour représentants associés $P_1=(1,0)$, $P_2=(0,1)$ et $P_3=\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$. La répartition des jours passés dans les 3 classes se traduit par le tableau de contingence suivant dans lequel les pourcentages sont en ligne :

classe	1	2	3
t			
jour ouvré	50%	50%	0%
samedi	0%	0%	100%
dimanche	0%	0%	100%

Dans ce cas, la règle définie au chapitre précédent associe aux *jours ouvrés* k le profil \tilde{P}_k suivant :

$$\tilde{P}_k = \tilde{P}(\text{jour ouvré}) = \frac{\frac{1}{2}P_1 + \frac{1}{2}P_2}{\left\| \frac{1}{2}P_1 + \frac{1}{2}P_2 \right\|} = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right).$$

Maintenant reclassons \tilde{P}_k dans les 3 classes. On rappelle que k , en tant que jour ouvré, a 50% de chance d'être classé dans la classe 1, 50% dans la 2 et 0% dans la 3. Pourtant $\tilde{P}_k = P_3 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right)$ et donc \tilde{P}_k est reclassé dans la classe 3.

En résumé la répartition des profils des jours ouvrés dans la classification est la suivante :

classe	1	2	3
t			
Jour ouvré	50%	50%	0%

En revanche, celle des estimés des profils des jours ouvrés est :

Classe	1	2	3
t			
jour ouvré	0%	0%	100%

On constate que ces deux répartitions sont à une distance maximale au sens du χ^2 . On a en fait classé tous les jours ouvrés comme les samedis et les dimanches, ce qui est bien sûr à l'opposé de l'objectif.

Remarques :

- Le problème de la méthode à ce stade de développement provient de ce que la distinction *jours ouvrés, samedis et dimanches* n'explique pas suffisamment la classification au sens où elle n'explique pas la différence entre P_1 et P_2 .

- Si on avait eu $P_1 = \frac{1}{\sqrt{0.9^2 + 0.1^2}}(0.9, 0.1)$, $P_2 = (1, 0)$ et $P_3 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, il n'y aurait pas

eu le problème, car l'estimé du profil des jours ouvrés \tilde{P}_k aurait été plus proche de P_1 et de P_2 que de P_3 . Il faut donc savoir si les profils qu'on associe pour estimer le profil d'une famille, par exemple *jours ouvrés*, sont proches ou non.

- Dans \mathbf{R}^2 et pour 3 types il est aisé de réaliser que la typologie est insuffisante, par exemple avec le tableau de contingence précédent. Mais cela se complique vite lorsque augmente le nombre des données ou de types. Il est donc nécessaire d'avoir un outil qui puisse aider à savoir dès l'établissement de la règle s'il manque un critère dans la définition des types.

Conclusion :

Il est indispensable d'établir une bonne typologie et donc disposer d'un outil pour contrôler la qualité des critères et leur éventuelle insuffisance (comme le souligne la première remarque). Un outil de visualisation qui permette de sélectionner les critères judicieux et voir si les profils qu'ils regroupent sont voisins constituerait une solution. Il aurait, de plus, l'avantage d'intervenir au moment du choix de la typologie et non au moment de l'étude de l'erreur. Les cartes de Kohonen appelées *cartographie du critère* (figure 70, 71 et 72 du chapitre 6.1.2.2) fournissent une solution en montrant à la fois les profils qui sont sélectionnés par le critère et si ces derniers sont voisins. La suite détaille comment les cartes de Kohonen peuvent déceler les situations délicates.

Comment les cartes de Kohonen apportent un outil visuel de contrôle de ce phénomène :

Le problème évoqué précédemment se visualise très bien. Le graphique suivant illustre le cas précédent où les données sont dans \mathbf{R}^2 . La classification fait ressortir 3 groupes, les jours ouvrés sont répartis équitablement dans les classes C_1 et C_2 de représentants P_1 et P_2 et les samedis et dimanches sont regroupés dans la classe C_3 de représentant P_3 . Le barycentre B des profils P_1 et P_2 est reclassé dans C_3 . Dans \mathbf{R}^2 , le cas redouté est typiquement celui où les classes dont les profils apportent une contribution au barycentre encerclent une autre classe. C'est cela qu'on cherche à détecter par les cartes de Kohonen ou tout autre outil visuel.

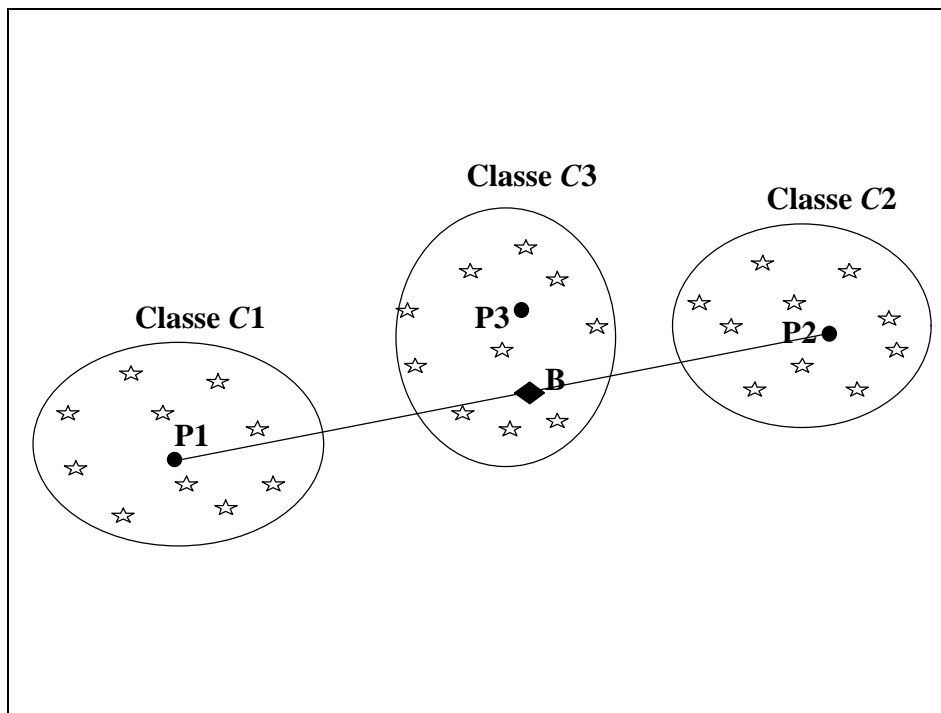


Figure 101 : Le barycentre B des profils moyens P_1 et P_2 respectivement des classes $C1$ et $C2$ est reclassé dans la classe $C3$

Gestion de ce problème dans le cadre de la classification de Kohonen :

Cas où les unités associées à un type ne sont pas dans une zone connexe :

D'après ce qu'on vient de voir, on s'attend à ce que le profil associé à un type soit reclassé dans l'une des classes dont le représentant a fourni une contribution au barycentre correspondant. Il est donc important de contrôler les regroupements des unités en classes connexes de façon à avoir l'assurance d'un bon comportement.

Or la carte de Kohonen, par sa notion de voisinage, fournit immédiatement une réponse. Faire le barycentre entre les vecteurs codes d'unités qui, lorsqu'elles sont regroupées, forment une zone connexe de la carte garantit que le barycentre B est proche de ces classes comme le montre la figure 102. Au contraire, si on calcule le barycentre des classes disposées à l'opposé sur la carte, on s'expose à un risque. Si la carte se replie, comme sur la figure 103, le cas reste valable, par contre sur la figure 104 la méthode sera mal utilisée.

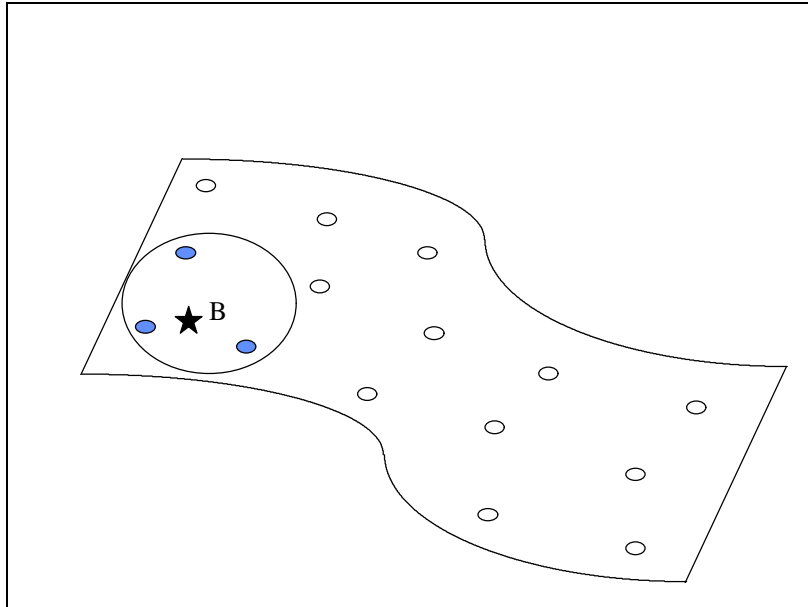


Figure 102 : Le barycentre des centres de classes bleues est reclassé dans une classe bleue uniquement si ces classes sont voisines sur la carte.

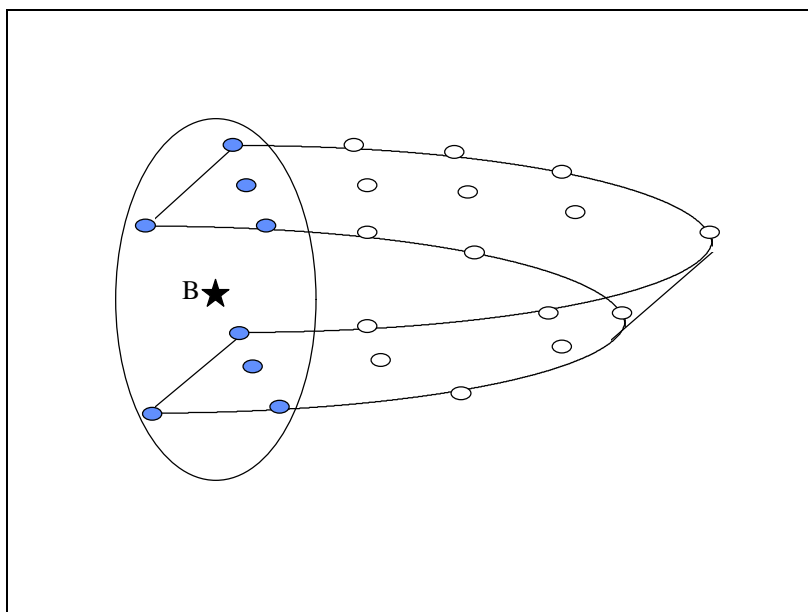


Figure 103 : Le repliement de la carte permet au barycentre des centres des classes bleues non voisines sur la carte d'être reclassé dans une classe bleue

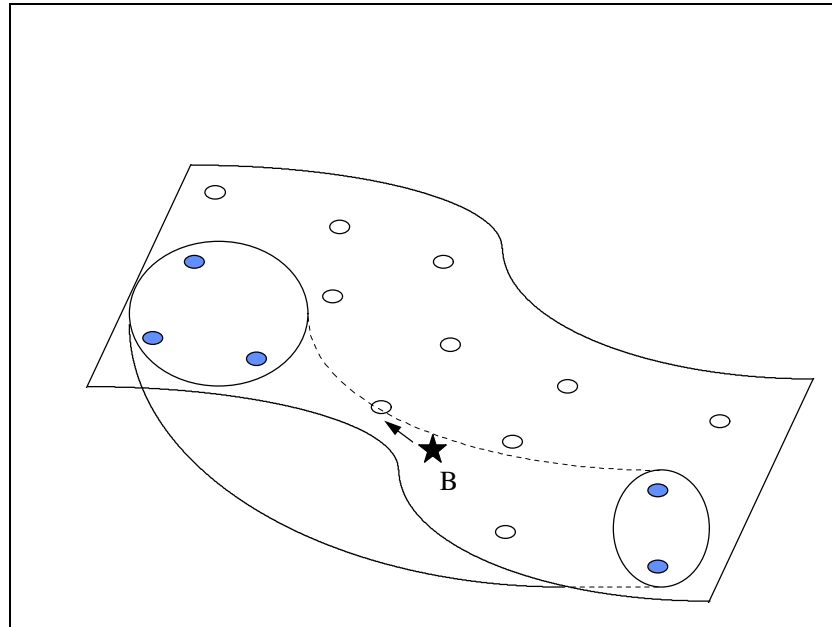


Figure 104 : Carte où le barycentre B des centres d'une sélection de classes éloignées sur la carte est reclassé dans une classe externe à cette sélection.

Remarque:

Supposons que le profil estimé \tilde{P}_k soit calculé avec un vecteur code G_{u_0} qui ne soit pas dans la région connexe des autres codes utilisés, ce dernier interviendrait avec le poids $P_{\text{roba}}(\{k' \in u_0, k' \in t(k)\})$. Les jours particuliers ne remettent donc pas en cause la méthode, car cette probabilité est négligeable. Seuls deux groupes de classes fortement représentés posent un problème. Mais cette configuration assurera qu'il y a un phénomène réel à découvrir et à prendre en compte dans la règle, la carte de Kohonen aboutira dans ce cas à la conclusion d'un manque de critère. Le problème ne doit donc pas être perçu comme pouvant restreindre le champ d'utilisation de la méthode, mais au contraire comme des situations que cette technique est la plus apte à mettre en évidence et à traiter.

D'autres outils de visualisation sont possibles, en particulier le couplage classification hiérarchique – analyse factorielle (il faut dans ce cas se méfier des effets des projections). La comparaison de la classification de Kohonen avec ces méthodes fait l'objet du chapitre 3.8. Les liens très forts entre représentation et classification dans le cas de Kohonen constitue un avantage dans la détection de la prévention du problème lié au mauvais choix des critères de définition des types.

7.4 Limites

Limites de la réduction :

La réduction a pour effet de mettre au même niveau toutes les variations. Elle n'est pas adaptée à tous les cas. En particulier, considérons une base de données qui contienne à la fois des individus remarquables et des individus qui ne représentent aucune forme typique (bruit blanc). Dans ce cas, la normalisation du bruit blanc créera toutes les formes possibles et la classification ne révélera pas de formes particulières.

Pour illustrer ce phénomène, considérons la figure 105. Elle représente deux formes typiques, une forme plate et une forme à forte variation. La base de données de notre exemple est construite à partir de ces deux formes étalons en rajoutant un bruit. Des exemples d'individus de la base ainsi constitués sont tracés dans la figure 106. Cependant, l'écart type de la forme plate est très faible par rapport à celui de l'autre forme et quand on normalise un individu sans forme, on donne de l'importance au bruit. De sorte que la classe de la forme à forte amplitude est retrouvée par la classification alors que la forme plate génère de nombreuses classes où on ne retrouve d'ailleurs pas la forme plate, ce phénomène est illustré par la figure 107.

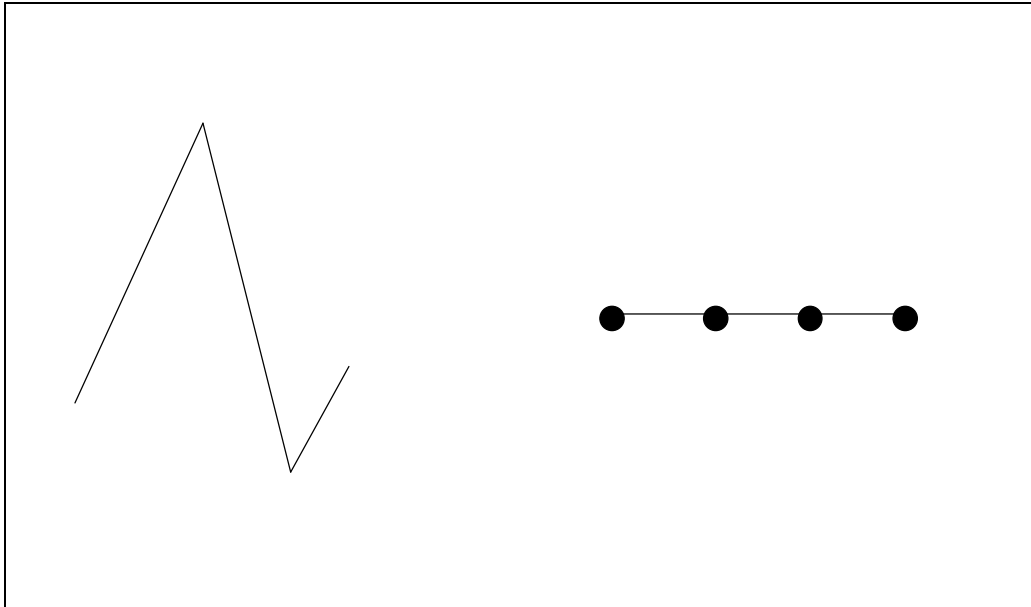


Figure 105 : Deux formes marquées

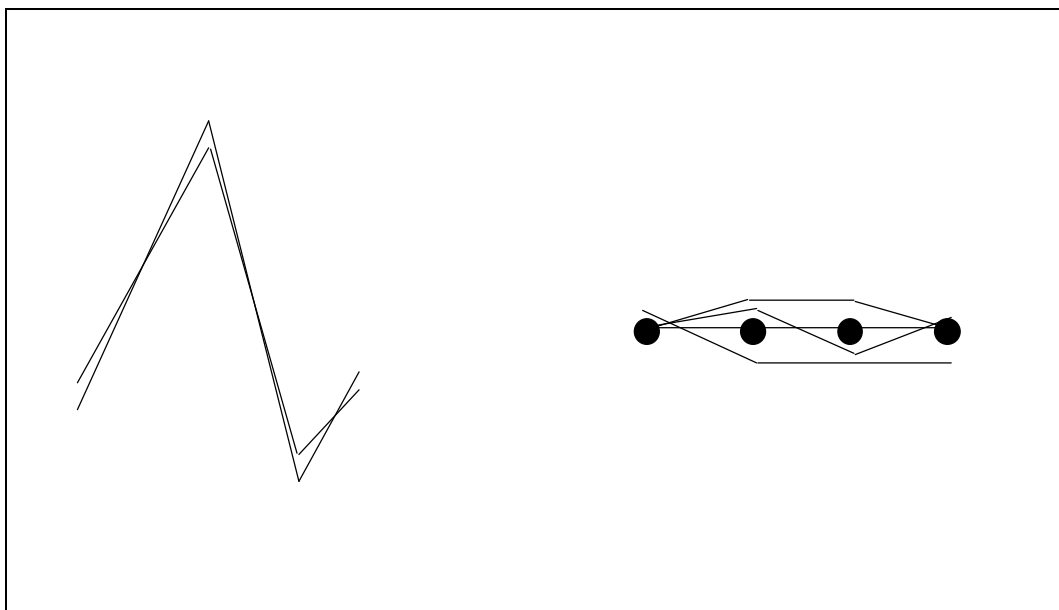


Figure 106 : Formes constituées par l'ajout d'un bruit aux deux formes marquées précédentes

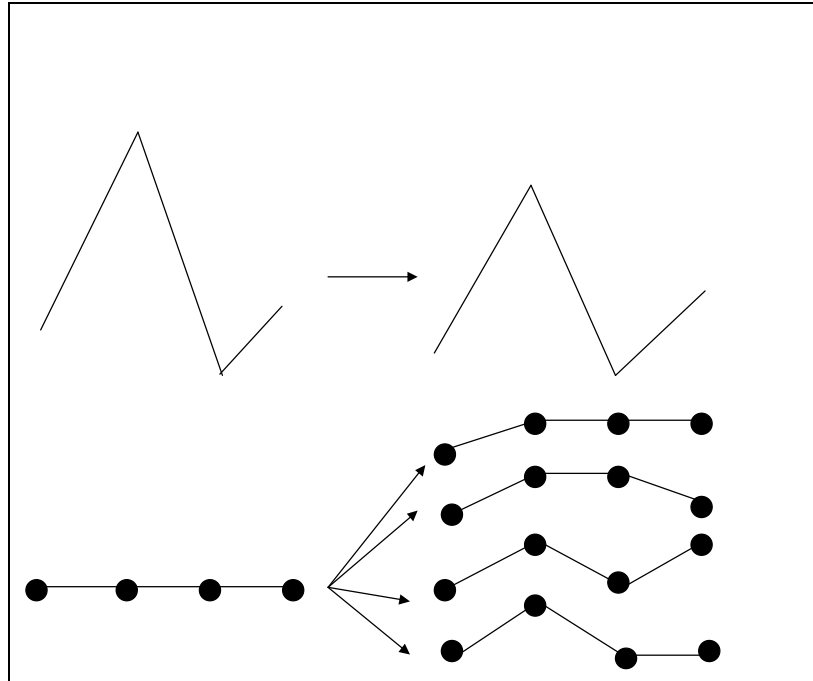


Figure 107 : Exemple de sorties de la classification des formes créées par le bruitage et la normalisation des courbes de gauche. La courbe plate génère un grand nombre de types

Conséquence :

Ce phénomène fait qu'il est difficile d'utiliser cette méthode pour prévoir des données corrigées par une autre méthode de prévision, et peut-être réduites à un bruit.

Couplage de la MEP avec une méthode linéaire :

On sait qu'un modèle linéaire de type armax est utilisé dans la pratique pour prévoir la consommation électrique. De plus, on vient d'expliquer pourquoi on ne peut utiliser la méthode *MEP* pour prévoir l'erreur du modèle précédent. De la même façon, l'erreur issue de la *MEP* a très peu de chance d'avoir des propriétés qui lui permettent d'être estimée par un modèle linéaire. Il faudrait pour cela, au minimum, qu'elle soit indépendante du type de jour ce qui serait très surprenant.

En conclusion, cette méthode ne se prête pas facilement à un couplage avec un modèle linéaire.

8 Conclusion et perspectives

On a exposé le modèle de prévision *MEP* dans un cadre volontairement restrictif, en s'appuyant sur une étude pratique dans le cadre de données réelles de consommation électrique. Ce choix s'est justifié par une volonté de simplifier la présentation et de mettre l'accent sur les aspects spécifiques de la méthode. Il est en effet possible d'étendre les domaines de données et d'applications de ce modèle de prévision.

Extension du domaine de données par une extension du critère de la règle d'affectation :

Dans l'étude de la consommation électrique, la règle d'affectation s'appuie sur des variables calendaires. On peut en fait établir cette règle à partir de critères de qualité répertoriés dans des variables à modalités. Dans la prévision des courbes de charge les jours fériés sont un peu dans ce cas, mais on peut ajouter d'autres facteurs tels que les jours de grève, les jours de modification tarifaire ou les jours de forte production industrielle qui ne sont pas fonction de la date.

L'introduction de tels critères élargit le domaine de données, car il offre une méthode de prévision dans le cadre de séries temporelles où l'on constate la répétition de formes d'évolution d'intervalles fixes, mais à des dates qui ne répondent pas à une loi calendaire. Par exemple, une série sur la pollution peut présenter des formes journalières de pollution dues à des conditions climatiques réunies au cours d'une seule journée. Un tel phénomène constitue une rupture dans la série chronologique et la courbe de jour n 'a de lien qu'avec un jour du passé où sont réunies les mêmes conditions atmosphériques. Notre méthode reste adaptée à ce type de séries.

Remarque :

Un modèle armax pour ce type de série se heurte au même problème que la prise en compte d'un jour férié dans la prévision des courbes de charge électrique (il arrive à prendre en compte des sauts de niveau mais pas des changements de forme).

Extension de l'espace d'application :

Le découpage des données en séquences de 24h00 dans le cadre de la consommation électrique correspond à l'objectif. En effet, on cherche à prévoir le lendemain avec la même précision pour toutes les heures, connaissant la consommation jusqu'à 24h00. Dans la pratique, on trouve des *contraintes techniques ou stratégiques* qui peuvent influencer le choix de la méthode. Parmi les *contraintes techniques*, on peut citer la fréquence de renouvellement des données. En effet, si le renouvellement des données s'effectue toutes les 5 heures ou toutes les 48 heures, on peut choisir de prévoir respectivement des séquences de 5 heures ou de 48 heures. Notre méthode répond à cette contrainte par le choix de la basse fréquence et devient plus pertinente que les méthodes récursives lorsque la fréquence de renouvellement des données est lente.

En effet, nous prévoyons les profils avec la même qualité quelle que soit l'échéance de prévision. Dans le cadre des *contraintes stratégiques*, on peut ajouter à l'exemple particulier de la consommation électrique *l'ensemble des projets qui établissent des scénarios*. Par exemple, on peut vouloir prévoir la consommation électrique d'un jour futur j selon différentes hypothèses telles que *j est un jour férié, un jour ouvré, un jour de grève, un jour de*

tarification particulière, dans une période de forte productivité industrielle, dans une conjoncture économique morose ou autre etc. Notre méthode reste adaptée à ce type d'étude au contraire des méthodes récursives. Cette fois, c'est la règle qui permet de prendre en compte la contrainte de *faire des scénarios*.

Ainsi, la méthode répond à des contraintes très diversifiées et peut donc s'appliquer à un domaine très étendu.

Perspectives :

La partie centrale de la méthode *moyenne – écart type – profil* est la construction d'une règle de prévision d'une partie de l'information à partir de variables calendaires ou qualitatives. De plus, telle qu'elle a été présentée, cette technique ne donne pas plus d'importance au passé proche qu'au lointain. **Il est donc naturel de vouloir faire progresser ce modèle en intégrant dans la règle la prise en compte de la succession temporelle et des variables quantitatives.**

Pour des raisons que l'on a déjà expliquées, **il est nécessaire de contrôler l'effet de ces notions sur les classes étendues à leurs voisinages.** C'est une fois de plus, en développant des outils d'analyse et de visualisation des cartes de Kohonen que l'on pourra le réaliser. En particulier, il serait intéressant de pouvoir **visualiser le suivi temporel sur la carte.** Ceci **aurait d'ailleurs des applications dans d'autres domaines que la prévision, par exemple l'analyse temporelle des panels.**

BIBLIOGRAPHIE

- 1 Ansley C.F.
An algorithm for the exact likelihood of a mixed autoregressive-moving average models,
Biometrika, 66, pp. 59-65, (1979).
- 2 Atlas L., Connor J. and Damborg M.
Comparisons of conventional techniques and neural networks in power system load
forecasting and stability
In Proceedings of the American Power Conference (1991).
- 3 Attali J. et Pagès G.
Approximation of functions by perceptrons: a new approach.
Neural Processing Letters, 22, 5, 19 (1995).
- 4 Auriault J., Lesne H.
Le modèle précoce (PRÉvision à COURT terme de la Consommation d'Énergie électrique),
Note HR/25-1780, EDF
- 5 Azencott R., Girard Y., astier R., Baudin M., Girard B., Jakubowicz P., et Martin M.
MANDRAKE: un progiciel expert en analyse de séries temporelles.
Harcourt Brace and World, Paris. (1991).
- 6 Ball G.H, Hall D.J
A clustering technique for summarizing multivariate data.
Behavioral Sciences , 12, pp. 153-155, (1967).
- 7 Barron A.
Universal approximation bounds for superpositions of a sigmoidal function
I.E.E.E. Transactions on Information Theory, 39, pp. 390-945. (1993).
- 8 Benaïm M., Fort J.C, Pages. G.,
Convergence of the one-dimensional Kohonen algorithm,
advanced appl. Prob. 30, pp. 850-869
- 9 Bishop C.M., Svensén M., Williams C.K.I.
GTM: a principled Alternative to the self-organizing map
In advances in Neural Information
Proceedings Systems, Editor : MC Mozer and M.I Jordan and T. Pitche, pp. 354-360
(1997).
- 10 Bishop C. M., Tipping M.E., Jordan M. I.
Hierarchical latent variable model for a data visualization

- 11 Bolzern, P. and Fronza, G.
Role of Weather Inputs in Short-Term Forecasting of Electric Load,
Electric Power and Energy Systems, 8. 1, pp. 42-46. (1986).
- 12 Bouton C., Pagès G.
Self-organization of the one-dimensional Kohonen algorithm with non uniformly
distributed stimuli,
Stochastic Process. Appl. 47, pp. 249-274 (1993).
- 13 Bouton C., Pagès G.
Convergence in distribution of the one-dimensional Kohonen algorithm when the stimuli
are not uniform.
Adv. Appl. Probab., 26, pp. 80-103 (1994).
- 14 Box G.E.P., Jenkins G.M.,
Time Series Analysis Forecasting and Control,
Holden-Day. San Francisco, (1976).
- 15 Brockwell, P. et Davis, R.
Time Series: Theory and Methods.
Springer-Verlag, Berlin (1991).
- 16 Caire P., Hatabian G. and Muller C.
Progress in Forecasting by Neural Networks
IJCNN, 2, pp. 540-545, Baltimore (1992).
- 17 Capone R., Kimbrough S.,
Using a neural network to predict electricity generation,
Proceedings WCNN 94, San Diego, 1, pp. 324-329, (1994).
- 18 Chaudary S.D., Kabra P.K, Srivastava S.C. and Vinod D.M.
Short Term Electric Load Forecasting Using Artificial Neural Network
In Proceedings of 4th International Conference on Expert System Applications to Power
Systems (1993).
- 19 Cosculluela M.J., Dominguez M.J., Montes R. and Garcia-tejedor AJ.
Day type Identification for electric hourly load demand forecasting using self-organized
maps
In Proceedings of NeuroNimes. (1993)
- 20 Cottrell M., Girard B. and Girard Y.
Réseaux de neurones et séries temporelles
XXIV Journées de Statistiques, ASU, Bruxelles, (1992).
- 21 Cottrell M., Girard B. and Girard Y., Mangeas M.
Time Series and Neural Network : a Statistical Method for Weight Elimination

- Proceedings of ESANN'93, pp. 157-165, M. Verleysen Ed. Quorum, Bruxelles. (ISBN 2-9600049-06) (1993).
- 22 Cottrell M., Fort J.C.,
Etude d'un algorithme d'auto-organisation,
Ann. Inst. Henri Poincaré , 23, 1, pp. 1-20 (1987).
 - 23 Cottrell M., Fort J.C., Pagès G.,
Comments about analysis of the convergence properties of topology preserving neural networks,
IEEE Trans. Neural Networks 6 (3), pp. 797-799 (1995).
 - 24 Cottrell M., Girard B., Girard Y., Mangeas M. and Muller C.
Neural modeling for time series: a statistical stepwise method for weight elimination
IEEE Transactions on Neural Networks, 6, 6, pp. 1355-1364 (1995).
 - 25 Cottrell M., Letremy P., Roy E.,
Analysing a contingency table with Kohonen maps: a factorial correspondence analysis
in: J. Cabestany, J. Mary, A. Prieto (Eds.), Proceedings IWANN'93, Lecture Notes in
Computer Science, Springer, Berlin, pp. 305-311 (1993).
 - 26 Cottrell M., Letremy P.,
Classification et analyse des correspondances au moyen de l'algorithme de Kohonen:
application à l'étude de données socio-économiques
Proceedings Neuro-Nimes, pp. 74-83 (1994).
 - 27 Cottrell M., de Bodt, E.,
A Kohonen Map Representation to avoid Misleading Interpretations ,
Proceedings of ESANN'96, Editions D Facto, Bruxelles, pp. 103-110 (avril 1996).
 - 28 Cottrell M., de Bodt, E., Gregoire, Ph.
Simulation de l'évolution de la structure à terme des taux d'intérêt : une approche non
paramétrique,
Banque & Marchés, 36, pp. 21-28
 - 29 Cottrell M., de Bodt, E., Gregoire, Ph., Henrion E.F.
Simulating Interest Rate Structure Evolution on a Long Term Horizon : A Kohonen
Map Application
Proceedings of the Conference Neuronal Networks in the Capital Markets, Californian
Institute of Technology, Pasadena, World Scientific (1996)
 - 30 Cottrell M., Fort J.C., Pages G.
Theoretical aspects of the SOM algorithm
Neuro Computing, 21, pp. 119-138 (1998).
 - 31 Cybenko G.,
Approximation by superposition of a sigmoidal function
Math. Control Signal Systems, 2, pp. 303-314 (1989).

- 32 Damborg. D. J. et al.
Potential of Artificial Neural Network to Power System Operation
Proceedings of IEEE International Symposium on Circuits and Systems, New Orleans
Louisiana, May 1-3, pp. 2933-2937 (1990).
- 33 De Bodt E., Cottrell M., Verleysen
Using the Kohonen Algorithm for Quick Initialisation of Simple Competitive Learning
Algorithm
(1999)
- 34 Dipti S, Lies A.C. and Chen J.C..
A novel approach to electric load forecasting based on a neural network
IEEE International Joint Conference on Neural Networks IJCNN 91, (Nov. 1991).
- 35 Duflo M.
Méthodes récursives aléatoires
Masson, Paris 1990
- 36 El-Sharkawi M.A., Oh S., Marks R.J, Damborg M. J., Brace C.M.
Short Term Electric Load Forecasting Using an Adaptively Trained Layered Perceptron
Forum on Applications of Neural Networks to Power Systems, Seattle, WA, 1991.
- 37 Erwin E., Obermayer K., Shulten K.,
Self-organizing maps: stationary states, metastability and convergence rate
Biol. Cybernet. 67, pp. 35-45 (1992).
- 38 Erwin E., Obermayer K., Shulten K,
Self-organizing maps: ordering, convergence properties and energy functions,
Biol. Cybernet. 67, pp. 47-55 (1992).
- 39 Finnoff W., Zimmiermann H.G,
Detecting Structure in Small Data sets by Network Fitting under Complexity Constraints
Proceedings of the International Work-shop PASE 1991, Zurich, (1991).
- 40 Fishwick P.
Neural network models in simulation: A comparison with traditional modeling
approaches
Working Paper, University of Florida, Gainesville, FL, (1989).
- 41 Flanagan J.A.,
Self-organizing neural networks,
Phd. Thesis, Ecole Polytechnique Fédérale de Lausanne, (1994).
- 42 Flanagan J.A.,
Self-organisation in Kohonen's SOM,
Neural Networks 6 (7) , pp. 1185-1197 (1996).
- 43 Forgy E. W.
Cluster analysis of multivariate data : efficiency versus interpretability of classifications.

- Biometric Society Meetings, Riverside, California (1965).
- 44 Fort J.C., Pagès G.,
About the convergence of the generalized Kohonen algorithm
in: M. Marinero, P.G. Morasso (Eds.), Proceedings ICANN'94, Springer, Berlin, pp.
318-321 (1994).
 - 45 Fort J.C., Pagès G,
On the a.s. convergence of the Kohonen algorithm with a general neighborhood function
Ann. Appl. Probab., 5, 4, pp. 1177-1216 (1995).
 - 46 Fort J.C., Pagès G,
About the Kohonen algorithm: strong or weak self-organization
Neural Networks 9, 5, pp.773-785 (1995)
 - 47 Fort J.C., Pagès G,
Convergence of stochastic algorithms: from the Kushner & Clark theorem to the
Lyapunov functional,
Adv. Appl. Probab. 28, 4, pp. 1072-1094 (1996).
 - 48 Galiana F., Handschin. E. and Fiechter A.
Identification of Stochastic Electric Load Models for Physical Data
IEEE Transactions on Automatic Control, AC-19, 6, pp. 887-893 (1974).
 - 49 Garcia Tejedor A., Cosculluela M., Bermejo C. and Montes R.,
A neural system for short-term load forecasting based on day-type classification
Proceedings ISAP 94, (1994).
 - 50 Ginzberg I. and Horn D.
Learning the rule of a time series
Int. J. of Neural Systems, 3, 2, pp 167-177 (1992).
 - 51 Gorodkin J., Hanses L.K, Krogh A., Svarer C. and Winther O.
A quantitative Study of Pruning by Optimal Brain Damage
int. J of Neural Systems, 4, 2, pp. 159-169 (1993).
 - 52 Gordon A.D.
A review of hierarchical classification
J.R. Statit. Soc., A, 150, 2, pp. 119-137 (1987).
 - 53 Groot C. and Wurtz D.
Analysis of Univariate Time Series with Connectionist Nets : A Case Study of Two
Classical Examples
Proceedings of the Munotec workshop, Dublin, Dec 1990
 - 54 Hertz, J. Krogh A., Palmer R.
Introduction to the theory of neural computation
Santa Fe Institute (1991).

- 55 Hinkley, D.
Jackknife Methods
Encyclopedia of Statistical Sciences, 4, pp. 280-287 (1983).
- 56 Ho K.L., Hsu Y.Y., Yang C.C.
Short term forecasting using a multilayer neural network with an adaptative learning algorithm
Transactions on Power Systems, 7, 1, pp. 141-149 (1992).
- 57 Hornik K., Stinchcombe M. and White H.
Multilayer feedforward networks are universal approximators
Neural Networks, 2, pp. 359-366 (1989).
- 58 Hornik. K.
Approximation capabilities of multilayer feedforward networks
Neural Networks. 4. (1991).
- 59 Hripesak G.
Problem Solving Using Neural Networks
M.D. Computing, 5, 3, pp. 25-37 (1988).
- 60 Hubele N. F. and Cheng C. S.
Identification of Seasonal Short Term Forecasting Models Using Statistical Decision Functions
IEEE Transactions on Power Systems, 5, 1. pp. 40-45 (1990).
- 61 Kaufman L and Rousseeuw, P.
Finding Groups in Data
Wiley, New York (1990).
- 62 Kohonen T.,
Self-organized formation of topologically correct feature maps
Biol. Cybernet, 43, pp. 59-69 (1982).
- 63 Kohonen T.,
Analysis of a simple self-organizing process
Biol. Cybernet, 44, pp. 135-140 (1982).
- 64 Kohonen T.,
Self-organization and Associative Memory
3rd ed., Springer, New York (1989).
- 65 Kohonen T.,
Self-Organizing Maps
Springer, New York, 30 (1995).
- 66 Lance G. N., Williams W. T.
A general theory of classification sorting strategies
Computer J., 9, pp. 373-380 (1967).

- 67 Lapedes, A. S. and Farber, R.,
Nonlinear signal processing using neural networks: prediction and system modeling Los
Technical Report, Alamos National Laboratory (1987).
- 68 Lebart, L., Morineau A., Piron M.
Statistique exploratoire multidimensionnelle.
Dunod, Paris (1995).
- 69 Le Cun Y., Denker J. S. and Solla S. A.
Optimal brain damage
in Touretzky, E. D. (ed.), Advances in Neural Information Processing Syst. II, San
Mateo, CA: Morgan Kaufman, pp. 598-605 (1990).
- 70 Le Cun Y
Une procédure d'apprentissage pour réseau à seuil assymétrique.
Cognitiva 85, pp. 599-604 (1985).
- 71 Lee K. Y., Cha Y. T., Park J. H.
Short-term load forecasting using an artificial neural network
IEEE, Trans. Power Systems, 7, 1, pp.124-132 (Feb. 1992).
- 72 Lee K.Y., Cha Y.T., Park J.H.
Artificial neural network methodology for short-term load forecasting
NSF Workshop on Artificial Neural Network Methodology in Power System
Engineering Clemson University, SC, Apr., pp. 9-10 (1990).
- 73 Leshno M., Ya Lin V., Pinkus A. and Schocken S.
multilayer Feedforward Networks with a Nonpolynomial Activation Function can
Approximate any Function
Neural Networks, 6, pp. 861-867 (1993).
- 74 Linde Y., Buzo A., Gray R.M.,
An algorithm for vector quantizer design
IEEE Transactions on Communications, COM-28, 1, (January 1980).
- 75 Lloyd S.P.,
least squares quantization in PCM
IEEE Transactions on Information Theory, IT-28, 2, pp. 129-149 (March 1982).
- 76 Mangeas M.,
Propriétés Statistiques des Modèles Paramétriques Non-linéaires de Prévion de Séries
Temporelles
Th : Mathématiques : Paris
- 77 Mangeas M., Muller C., Weigend A.S
Forecasting electricity demand using a mixture of nonlinear experts
In World congress on Neural Networks WCNN'95, 2, pp. 48-53 (1995).

- 78 Mac Culloch W. et Pitts W
A Logical Calculus of the Ideas Immanent in Nervous Activity
Bulletin of Mathematical Biophysics, 5 (1943).
- 79 MacQueen J.B
Some methods for classification and analysis of multivariate observations.
Proceedings Symp. Math. Statist. And probability (5th), Berkley, 1, pp. 281-297,
University of Californie. Press, Berkley (1967).
- 80 Minsky M. et Papert S.
Perceptrons. MIT Press, Cambridge (E.U) (1969).
- 81 Muller, M., Cottrell, M., Girard, B., Girard, Y. and Mangeas, M.
A neural network tool for forecasting french electricity consumption
Proceedings of WCNN 94, San Diego, 1, pp. 360-365 (1994).
- 82 Park D.C., El-Sharkawi M.A, Marks R.J., Atlas L.E., and Damborg M.J.
Electric load forecasting using an artificial neural network
IEEE Transactions on Power Systems, 6 (1991).
- 83 Park, J. H., Park, Y. M., and Lee K. Y.
Composite Modeling for Adaptive Short Term Load Forecasting
IEEE Power Systems Engineering Summer Meeting. (1990).
- 84 Peng. T. M., Hubele. N. F., Karady. G. G.
Conceptual Approach to the Application of Neural Network for Short Term Load
Forecasting
Proceedings of IEEE International Symposium on Circuits and System, New Orleans,
Louisiana May 1-3. pp. 2942-2945. (1990)
- 85 Peng T.M., Hubele N.F. and Karady G.G.
Advancement in the application of neural network for short-term load forecasting
IEEE Transactions on Power Systems, 7 (1992).
- 86 Rosenblatt F.
Principles of Neurodynamics
Spartan, New York (1962).
- 87 Rumelhart D. E., Hinton G. E. et Williams R. J.
Learning internal representations by error propagation
In Rumelhart D. E. et McClelland J. L. editors, Parallel Distributed Processing:
Explorations in the Microstructure of Cognition, 1: Foundations, pp. 318-362. MIT
Press/Bradford Books, Cambridge, MA. (1986).
- 88 Sadeghi A.
Asymptotic behaviour of self-organizing maps with non-uniform stimuli distribution Ann.
Appl. Probab., 8, 1, pp.281-299 (1998).

- 89 Sadegui A.
Self-organization property of Kohonen's map with general type of stimuli distribution
Neural Networks (1997).
- 90 Saporta, S.
Probabilités Analyse de données et Statistiques.
Editions Technip (1990).
- 91 SAS/ETS SAS/IML SAS/STAT
User's Guide. Version 6. First Edition. SAS Institute Inc.
- 92 Sobajic D. J. and Pao Y. H.
Artificial Neural Net Based Dynamic Security Assessment for Electric Power System
IEEE Transactions on Power Systems, 4, 1, pp. 220-228 (1989).
- 93 Sobral R. and Canu S.
Application des Réseaux de Neurones Artificiels à la Prévision : la Consommation d'Eau
Actes du Congrès Neuro-.Nimes. (1990).
- 94 Sokal R.R., Sneath P.H.A
Principles of numerical Taxonomy
Freeman and co., San-Francisco (1963).
- 95 Thorndike R.L
Who belongs in the family
Psychometrika, 18, pp. 267-276 (1953).
- 96 Van Meeteren H. and Van Son P.
Short Term Load Prediction with a Combination of Different Models
IEEE Power Industry Computer Applications Conference, pp. 192-197 (1979).
- 97 Varfis A. and Versino. C.
Univariate Econoinic Time Series Forecasting by Connectionist Methods
INNC. Paris (1990).
- 98 Vemuri S., Huang W. L. and Nelson D. J.
On-line algorithms for forecasting hourly loads of an electric utility
IEEE Trans. on Power Appr. and Sys., PAS-100, pp. 3775-3784 (1981).
- 99 Ward J.H
Hierarchical grouping to optimize an objective function
J. of Amer. Statist. Assoc., 58, pp.236-244 (1963).
- 100 Weigend A. S., Huberman B. A. and Rumelhart D. E.
Predicting the future: a connectionist approach
Int. J. Neural Systems, 1, 3, pp.193-209 (1990).
- 101 Weigend, A. S. and Gershenfeld N. A.
Times Series Prediction

- Proceedings Volume XV, Santa Fe Institute, Reading, MA: Addison-Wesley (1994).
- 102 Weigend A.S., Mangeas M., Srivastava A. N.
Nonlinear gated experts for time series: discovering regimes and avoiding overfitting
International Journal of Neural Systems 6, pp. 373-399 (1995).
- 103 Werbos P.
Generalization of backpropagation with application to recurrent gas market model
Neural Networks, 1, pp. 339-356 (1988).
- 104 Wong M.A.
A hybrid clustering method for identifying high density clusters
J. of Amer. Statist. Assoc., 77, p 841-847 (1982).
- 105 Yao J.F, Mangeas M.
Sur l'estimation des moindres carrés d'un model autoregressif fonctionnel
CRAS de Paris, 324-I, PP. 471-474
- 106 Yao Q., Tong H.
Quantifying the influence of initial values on nonlinear prediction
Technical Report, UKC/IMS/S92/5c, University of Kent, U.K. (1992).
- 107 Yuan-Yih Hsu and Chien-Chuen Yang.
Desing of artificial neural networks for short-term load forecasting. part 1: self-
organising feature maps for day type identification
IEEE Proceedings-C, 138 (1991).