

UNE MÉTHODE DE SOFT SUBSPACE CLUSTERING BASÉE SUR LES CARTES  
TOPOLOGIQUES POUR DONNÉES MULTIBLOCS ET UNE MÉTHODE DE  
RECHERCHE DE CONSENSUS DE PARTITIONS

Ouattara  
N Niang & F Badran & C Mandin

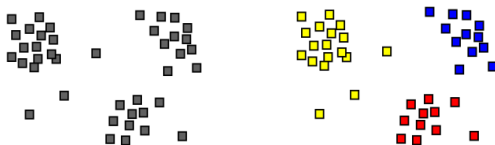
22-01-2014



- 1 Introduction
- 2 Classification de données multi-vues basée l'algorithme SOM (2S-SOM)
- 3 Méthode fusion de SOM basée sur la matrice des corrélations vectorielles ( $R_V$ -CSOM)
- 5 Conclusions & Perspectives

# La classification

La **Classification** est une tâche d'apprentissage non supervisé visant à regrouper les d'objets d'une base de données en classes **homogènes** et **contrastées**.



Plusieurs méthodes de partitionnement fournissent les classes (K-moyennes (Forgy (1965) ; MacQueen (1967))), ou les méthodes hiérarchiques (CAH) (Gordon (1987)).

# Représentation d'une partition

- $\pi$  : Une variable qualitative dont les modalités sont les labels des classes
- $H$  : La matrice disjonctive associée à  $\pi$
- $Co_{\pi}$  : La matrice adjacence ( $N \times N$ ) associée à  $\pi$

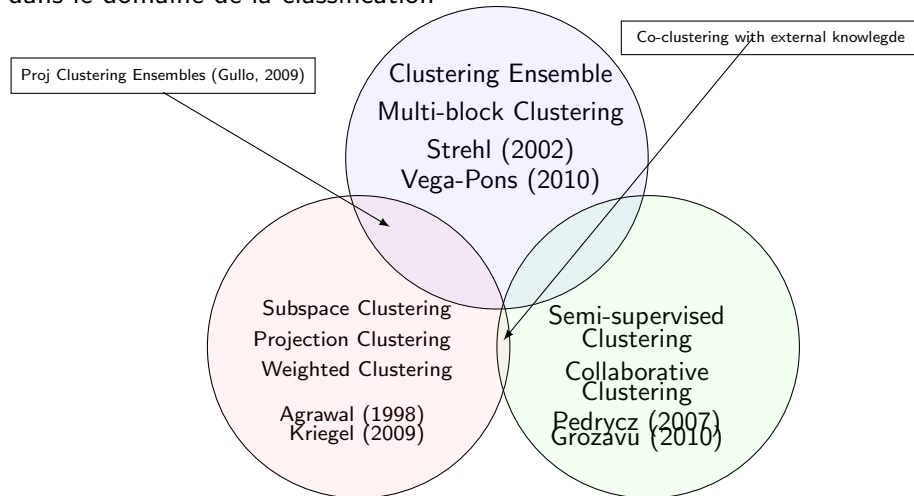
	$\pi$
$z_1$	1
$z_2$	2
$z_3$	2
$z_4$	1
$z_5$	3
$z_6$	3

	$H$		
	$\pi_{11}$	$\pi_{12}$	$\pi_{13}$
$z_1$	1	0	0
$z_2$	0	1	0
$z_3$	0	1	0
$z_4$	1	0	0
$z_5$	0	0	1
$z_6$	0	0	1

	$Co_{\pi}$					
	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$
$z_1$	1	0	0	1	0	0
$z_2$	0	1	1	0	0	0
$z_3$	0	1	1	0	0	0
$z_4$	1	0	0	1	0	0
$z_5$	0	0	0	0	1	1
$z_6$	0	0	0	0	1	1

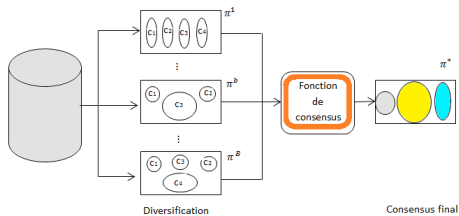
# Domaine très vaste

La complexité croissante des données a engendré de nombreux progrès dans le domaine de la classification



# Ensemble cluster (1)

- 1 **Diversification.** Un ensemble  $\Pi = \{\pi^1, \dots, \pi^B\}$  de partitions



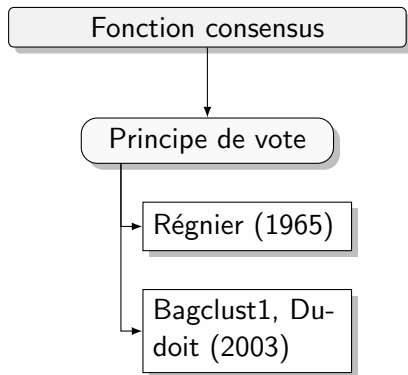
- 2 **Agrégation.** Recherche du consensus des partitions  $\pi^b$  à l'aide d'une fonction de consensus.

# Ensemble cluster (2)

## Propriétés souhaitées

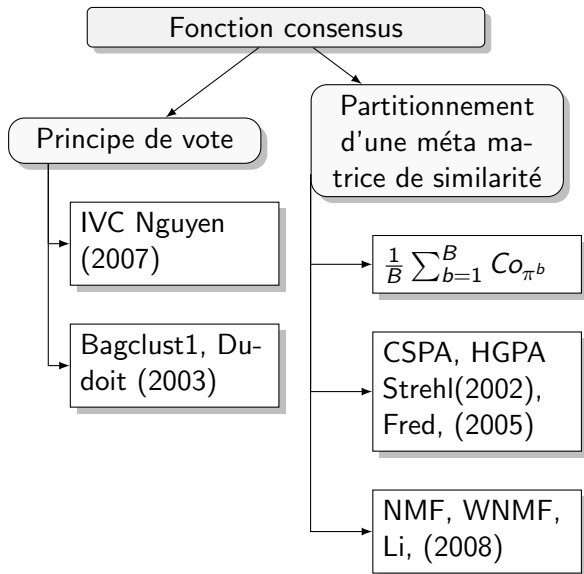
- 1 **Consistance**, le consensus est un compromis des partitions
- 2 **Robustesse**, le consensus est en moyenne meilleure que les partitions.
- 3 **Stabilité**, le consensus est insensible à la présence d'outliers et de variables de bruit
- 4 **Nouveauté**, la partition consensus doit mettre en évidence des propriétés des données qui sont difficilement atteignables par chaque partition de l'ensemble.

## Ensemble cluster (3)

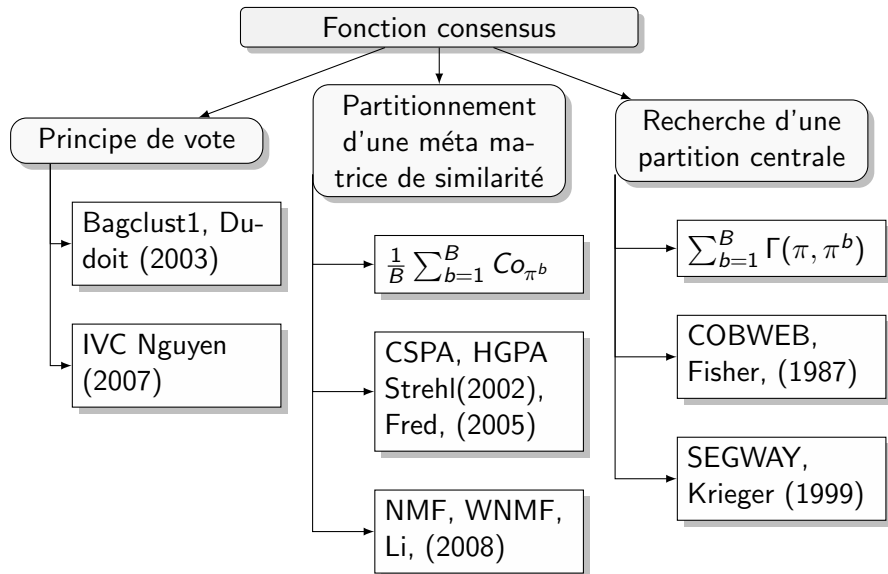




## Ensemble cluster (3)



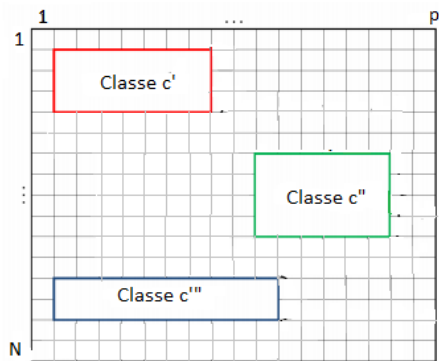
# Ensemble cluster (3)



# Subspace clustering

**Découvrir des classes d'objets dans des sous-espaces de l'espace initiale** Un ensemble  $\mathcal{Z}$  d'objets définis dans  $\mathbb{R}^p$

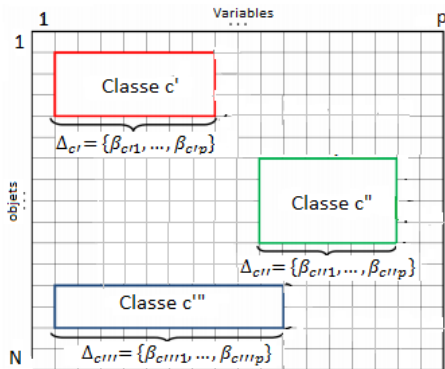
- une partition  $\pi$  des observations et un ensemble de sous-espaces de variables caractéristiques de chaque classes de  $\pi$

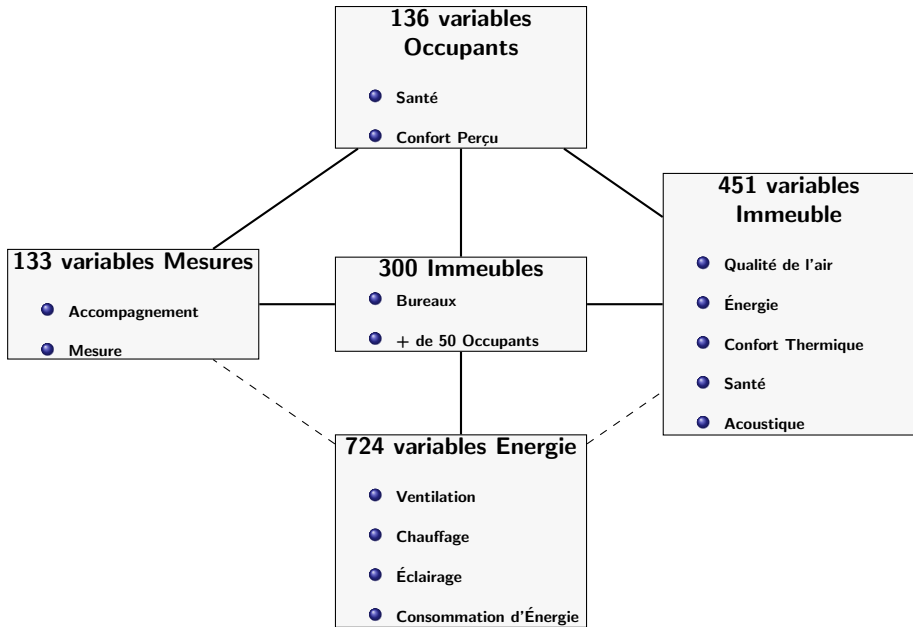


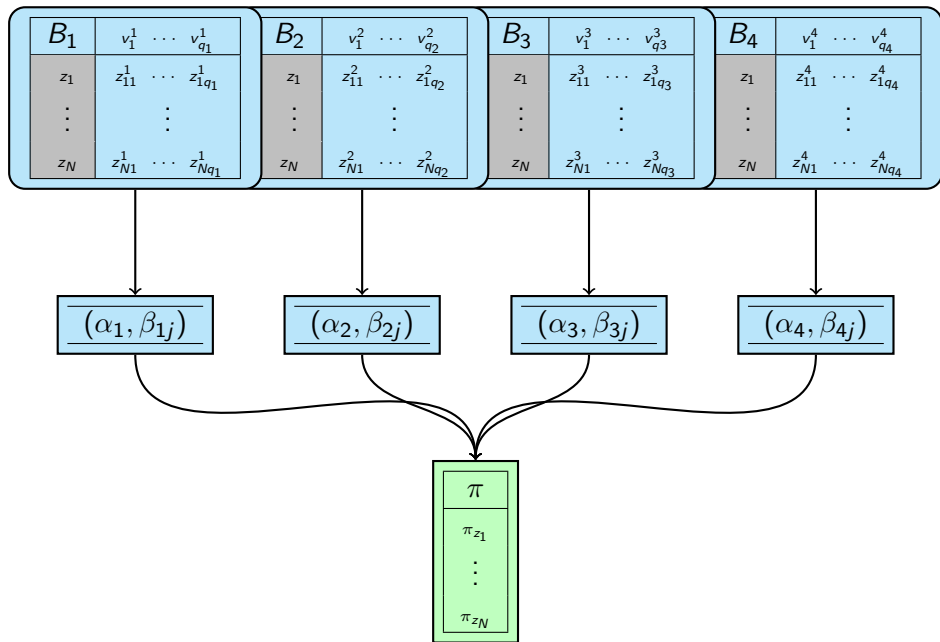
# Subspace clustering

**Découvrir des classes d'objets dans des sous-espaces de l'espace initiale** Un ensemble  $\mathcal{Z}$  d'objets définis dans  $\mathbb{R}^p$

- une partition  $\pi$  des observations et un ensemble de sous-espaces de variables caractéristiques de chaque classes de  $\pi$
- A chaque classe  $c \in \pi$  est associée un vecteur  $\beta_c$   
 $\beta_c = \{\beta_{c1}, \dots, \beta_{cp}\}$  définissant la contribution des variables pour cette classe.







# 2S-SOM

- C'est une extension des Cartes Topologiques Auto-organisées pour gérer les données multi-bloc : **Soft Subspace Self Organizing Map** (2S-SOM)

2S-SOM est Basée sur les principes de pondération des variables proposés par Liping (2007) et Chen(2012) pour prendre en compte les propriétés de visualisation et de réduction des dimensions qu'offre SOM.

# SOM, Kohonen (1992)

## Fonction Objectif

$$G_{som}^T(\mathcal{X}, \mathcal{W}) = \sum_{z_i \in \mathcal{Z}} \sum_{c \in \mathcal{C}} \mathcal{K}^T(\sigma(c, \mathcal{X}(z_i))) \|z_i - w_c\|^2$$

**Optimisation** : Processus itération en deux phases

$$\text{Affectation : } \mathcal{X}^T(z_i) = \underset{c \in \mathcal{C}}{\operatorname{argmin}} \sum_{c \in \mathcal{C}} \mathcal{K}^T(\sigma(c, \mathcal{X}(z_i))) \|z_i - w_c\|^2$$

$$\text{Actualisation : } w_c^T = \frac{\sum_{r \in \mathcal{C}} \sum_{z_i \in r} \mathcal{K}^T(\sigma(c, r)) z_i}{\sum_{r \in \mathcal{C}} \mathcal{K}^T(\sigma(c, r)) n_r}$$

- $\mathcal{C}$  désigne la carte topologique et  $c$  une cellule de la carte.
- $\mathcal{X}$  une fonction d'affectation des observations  $z_i$  aux cellules  $c$
- $\mathcal{K}^T$  est la fonction évaluant le voisinage entre deux cellules  $c$  et  $r$  de la carte  $\mathcal{C}$  et  $\sigma$  un fonction évaluant le voisinage entre deux cellules  $c$  et  $r$ .



# Critère objectif de 2S-SOM

$$\mathcal{J}_{2S-SOM}^T(\mathcal{X}, \mathcal{W}, \alpha, \beta) = \sum_{c \in \mathcal{C}} \left( \sum_{b=1}^B \left( \sum_{z_i \in \mathcal{Z}} \alpha_{cb} \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c)) d_{\beta_{cb}}(i) + J_{cb} \right) + I_c \right)$$

avec  $d_{\beta_{cb}}(i) = \sum_{j=1}^{p_b} \beta_{cbj} (z_{ib}^j - \omega_{cb}^j)^2$  et  $I_c = \lambda \sum_{b=1}^B \alpha_{cb} \log(\alpha_{cb})$  et  $J_{cb} = \eta \sum_{j=1}^{p_b} \beta_{cbj} \log(\beta_{cbj})$  Son optimisation se fait sous les contraintes :

$$\left\{ \begin{array}{l} \sum_{j=1}^{p_b} \beta_{cbj} = 1, \beta_{cbj} \in [0, 1], \forall c \in \mathcal{C}, \forall b \\ \sum_{b=1}^B \alpha_{cb} = 1, \alpha_{cb} \in [0, 1], \forall c \in \mathcal{C} \end{array} \right.$$

# Affectation et Actualisation des centres de classes

**Phase d'Affectation**, pour  $T$  donné, les paramètres de pondération  $\alpha_{ck}$  sont connus ainsi que les prototypes. Les individus sont affectés à leur cellule gagnante  $c_g$  grâce à la fonction  $\mathcal{X}$  :

$$c_g(z_i) = \mathcal{X}(z_i) = \underset{c \in \mathcal{C}}{\operatorname{argmin}} \left( \sum_{r \in \mathcal{C}} \mathcal{K}^T(\sigma(r, c)) \left( \sum_{b=1}^B \alpha_{cb} d_{\beta_{cb}}(i) \right) \right)$$

**Phase d'Actualisation** du vecteur référent  $\omega_{c_g}^T$  associé à la cellule  $c_g$  :

$$\omega_{c_g}^T = \frac{\sum_{z_i \in \mathcal{Z}} \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c_g)) z_i}{\sum_{z_i \in \mathcal{Z}} \mathcal{K}^T(\sigma(\mathcal{X}(z_i), c_g))}$$

## Calcul des $\beta$

supposons que les paramètres  $\mathcal{X} = \hat{\mathcal{X}}$ ,  $\omega = \hat{\omega}$  et  $\alpha = \hat{\alpha}$  sont connus et fixés à leurs valeurs courantes alors  $\forall \theta > 0$ ,  $\mathcal{J}_{2S-SOM}^T(\hat{\mathcal{X}}, \hat{\omega}, \hat{\alpha}, \beta)$  atteint son minimum pour une cellule  $c$  et pour un bloc  $b_k$  en

$$\beta_{cbj} = \frac{\exp\left(\frac{-\Phi_{cbj}}{\eta}\right)}{\sum_{j=1}^{p_b} \exp\left(\frac{-\Phi_{cbj}}{\eta}\right)}$$

$$\Phi_{cbj} = \sum_{z_i \in r, r \neq c} \alpha_{cb} \mathcal{K}^T(r, c) (z_{ib}^j - \omega_{cb}^j)^2 + \mathcal{K}^T(c, c) \sum_{z_i \in c} \alpha_{cb} (z_{ib}^j - \omega_{cb}^j)^2$$

## Calcul des $\alpha$

supposons que les paramètres  $\mathcal{X} = \hat{\mathcal{X}}$ ,  $\omega = \hat{\omega}$  et  $\beta = \hat{\beta}$  sont connus et fixés à leurs valeurs courantes alors  $\forall \lambda > 0$ ,  $\mathcal{J}_{2S-SOM}^T(\hat{\mathcal{X}}, \hat{\omega}, \alpha, \hat{\beta})$  atteint son minimum pour une cellule  $c$  et pour un bloc  $k$  en

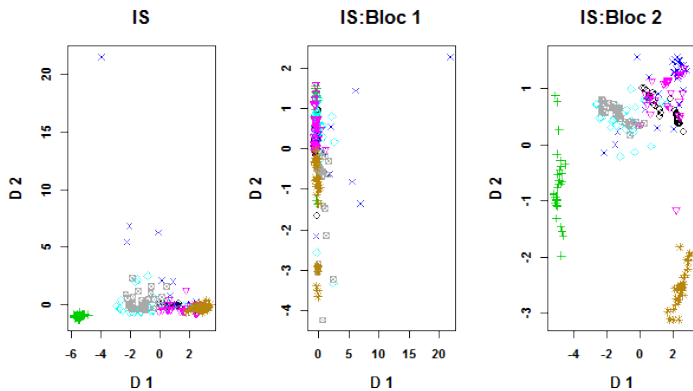
$$\alpha_{cb} = \frac{\exp\left(\frac{-\Psi_{cb}}{\lambda}\right)}{\sum_{b=1}^B \exp\left(\frac{-\Psi_{cb}}{\lambda}\right)}$$

$$\Psi_{cb} = \sum_{z_i \in r, r \neq c} \mathcal{K}^T(r, c) d_{\beta_{cb}}(i) + \mathcal{K}^T(c, c) \sum_{z_i \in c} d_{\beta_{cb}}(i)$$

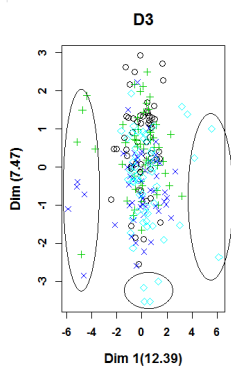
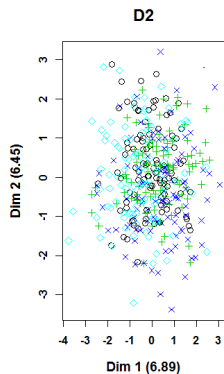
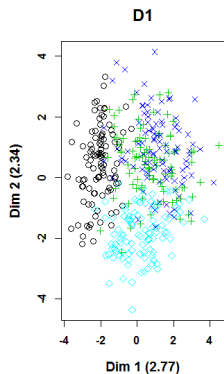
## Données

Données	Structure en blocs		Structure de la carte			
	$\#blocs$	$\#VB$	Niter	Dim	$T_i \times T_f (\lambda, \eta)$	
IS	9-10		150	$9 \times 9$	$2 \times 0.82$	(3,31)
CT	7-4-10		150	$10 \times 10$	$2 \times 0.1$	(7,11)
DMU	76-216-64-240-47-6		150	$10 \times 7$	$3 \times 0.2$	(10, 20)
D1	25-25-25-25	9-18-10-7	150	$10 \times 9$	$2 \times 0.1$	(2, 3)
D2	5-5-5-5	2-2-4-4	150	$10 \times 9$	$3 \times 0.2$	(1, 5)
D3	5-5-5-5-5	2-2-4-4-5	150	$10 \times 10$	$3 \times 0.2$	(1, 5)

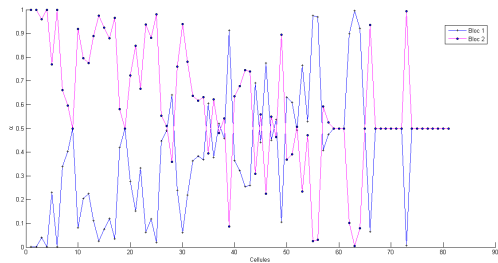
IS : 2310 obs, 19 variables, 2 blocs et 7 classes



# D1, D2 et D3



# Données IS



(a)





	Données	Indices	kM	EWKM	FGKM	SOM	2S-SOM
IS		Précision	0.38	0.66	0.60	0.63	<b>0.71</b>
		F-mesure	0.50	0.64	0.59	0.59	<b>0.69</b>
		Pureté	0.41	0.59	0.63	0.61	<b>0.63</b>
Réelles	CT	Precision	<b>0.50</b>	0.45	0.40	0.44	0.47
		F-mesure	<b>0.48</b>	0.45	0.27	0.44	0.45
		Pureté	<b>0.47</b>	0.43	0.38	0.45	0.45
DMU		Precision	0.59	<b>0.81</b>	0.60	0.75	0.80
		F-mesure	0.59	<b>0.80</b>	0.62	0.74	<b>0.80</b>
		Pureté	0.61	<b>0.77</b>	0.40	0.72	<b>0.77</b>
D1		Precision	0.37	0.98	0.90	0.31	<b>0.99</b>
		F-mesure	0.36	0.77	0.77	0.29	<b>0.78</b>
		Pureté	0.47	0.72	0.72	0.38	<b>0.74</b>
Simulées	D2	Précision	0.46	0.37	0.70	0.28	<b>0.87</b>
		F-mesure	0.45	0.36	0.60	0.27	<b>0.70</b>
		Pureté	0.58	0.45	0.61	0.33	<b>0.71</b>
D3		Précision	0.33	0.35	0.75	0.35	<b>0.90</b>
		F-mesure	0.31	0.36	0.61	0.29	<b>0.62</b>
		Pureté	0.37	0.47	0.49	0.35	<b>0.51</b>

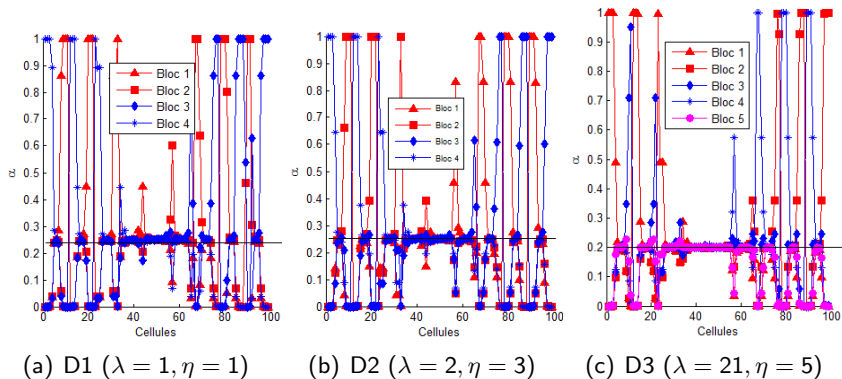
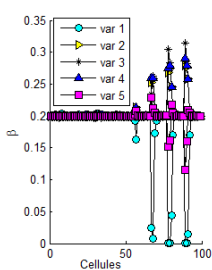
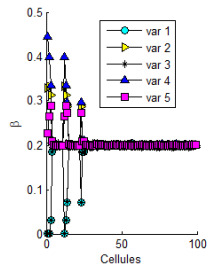


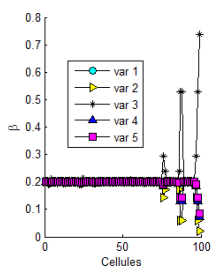
FIGURE : Les poids  $\alpha_{cb}$  associés aux blocs par rapport aux cellules des cartes associées aux tables D1, D2 et D3



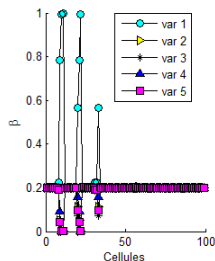
(a) Bloc 1



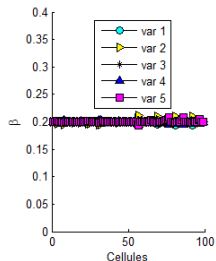
(b) Bloc 2



(c) Bloc 3



(d) Bloc 4

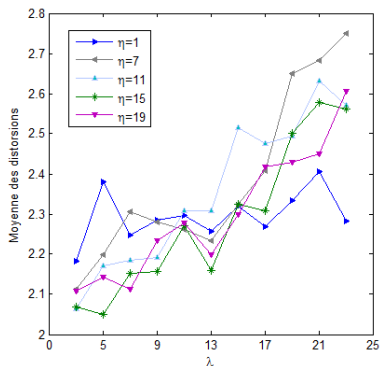


(e) Bloc 5

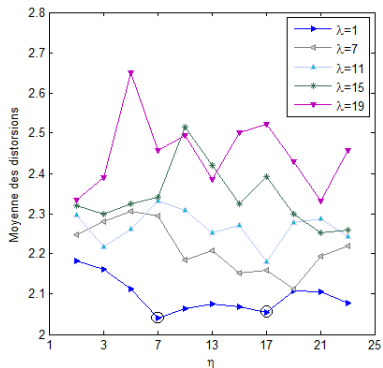
# Campagne Nationale "Logements" (CNL) 2003-2005

Logements		Ménage	
quantitatives	qualitatives	quantitatives	qualitatives
32	40	5	6
Habitudes		Polluants	
quantitatives	qualitatives	quantitatives	qualitatives
22	20	15	

TABLE : Les blocs de variables



(a)



(b)

FIGURE : Évolution de la moyenne de la mesure de distorsion pour les couples de paramètres  $\lambda$  et  $\eta$

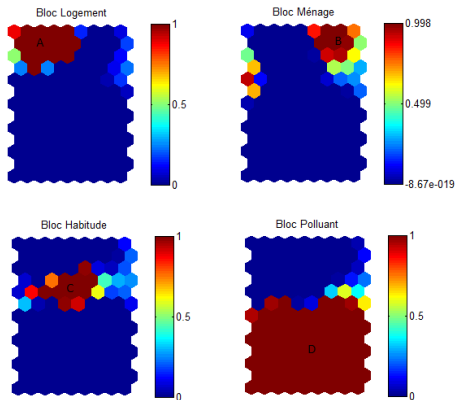


FIGURE : Les poids des blocs sur chaque cellule de la carte ; Les codes couleurs correspondent aux poids des blocs dans les cellules

Classe 5 Surface 57m<sup>2</sup> vs 110m<sup>2</sup>, fort taux de meubles en pvc et en bois aggloméré, Jeunes (38 ans vs 48 ans) en activité (65%), faibles revenus (1593 vs 2522), seuls, très faible activité d'entretien, faible utilisation des déodorants et désodorisant, forte concentration en acétaldéhyde (11 vs 8), en acroléine (0.89 vs 0.59), faibles concentrations en toluène (4.29 vs 8)

Classe 3 Surface 73m<sup>2</sup> vs 110m<sup>2</sup>, personnes en activité (0.80%), 84% employés, ouvriers ou profession, fort taux de meubles en pvc et en bois aggloméré, faibles revenus (2097 vs 2522), couples.

Classe 4 Surface 113m<sup>2</sup> vs 110m<sup>2</sup>, Jeunes (36 ans vs 48 ans), couples avec 2 enfants de moins de 2 ans, en activité (0.83), forte concentration de formaldéhyde (22 vs 18)

Classe 6 Surface 140m<sup>2</sup> vs 110m<sup>2</sup>, fort taux de meubles en bois massif, cadres supérieurs ou artisan (60%) en couple, Agé (57 ans vs 48 ans), fort revenus (4144 vs 2522), faible utilisation des déodorants, faibles concentrations en acétaldéhyde (6 vs 8), en toluène (3.93 vs 5.87).

Classe 2 : Surface 157m<sup>2</sup> vs 110m<sup>2</sup>, fort taux de menuiseries en bois, revenus élevés (4771 vs 2522), artisans ou cadre supérieurs (64%) personnes en activité (89%), couples avec 2 enfants de + de 10 ans, très forte activité d'entretien, forte utilisation des Déodorants opposée à une faible utilisation des désodorisants. Faibles concentrations en acétaldéhyde (6 vs 8), en benzène (0.64 vs 0.83), en toluène (3.95 vs 5.87).

Classe 1 : Surface 122m<sup>2</sup> vs 110m<sup>2</sup>, personnes en activité (87%), Employés ou ouvriers (64%), faibles revenus (2212 vs 2522), couples avec 2 enfants de + de 10 ans, très forte activité d'entretien, forte utilisation des Déodorants.

Classe 7 Surface 115m<sup>2</sup> vs 110m<sup>2</sup>, fort taux de menuiseries et de meubles en bois massif, retraité (67 ans vs 48 ans), faibles revenus (1761 vs 2522), couples, très faible activité d'entretien, très faible utilisation des Déodorants, concentrations faibles en formaldéhyde (15 vs 18), en acroléine (0.43 vs 0.59), en hexaldéhyde (3.42 vs 4.82).



# Méthodes de fusion de SOM

**Objectif** : Favoriser les cartes ou les partitions les meilleures au sens d'un certain critère

## CSOM

Le consensus finale tient compte de l'importance relative de chaque carte topologique de l'ensemble  $\Pi$ .

## $R_V$ -CSOM

La prise en compte de l'information commune à toutes les partitions. La recherche d'une matrice compromis moyenne des matrices issues des partitions représentant au mieux la similarité entre les cartes topologiques.

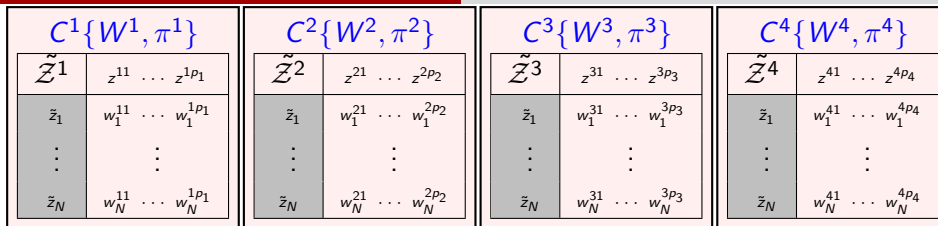


# Notations

Soit  $\Pi = \{\mathcal{C}^1, \dots, \mathcal{C}^b, \dots, \mathcal{C}^B\}$  et  $\mathcal{C}^b = (W^b, \pi^b)$

où  $\pi^b$  est une partition des observations et  $W^b = \{w_1^b, \dots, w_{N_b}^b\}$   
l'ensemble des vecteurs référents de la carte  $b$ .

Problème : définir la partition  $\pi^*$  et la matrice des vecteurs référents  
 $W^* = \{w_1^*, \dots, w_{N_{cell}}^*\}$  de la carte fusion des cartes  $\mathcal{C}^b$ .



## Détermination des poids $\mu$

**Évaluation** de la qualité des cartes avec la mesure de distorsion :

$$Dist_b = \sum_{i=1}^N \sum_c K^T(X(z_i), c) \|z_i - w_c\|^2 \quad (\text{Vesanto, 2003})$$

$$\mu_b = \frac{1}{Dist_b Dist_f} \quad \text{Avec} \quad Dist_f = \sum_{b=1}^B \frac{1}{Dist_b}$$

En posant :

$$d_f(\tilde{z}_i, \tilde{w}_c) = \sum_{b=1}^B \mu_b \|\tilde{z}_i^b - \tilde{w}_c^b\|^2$$

On redéfinit la fonction objectif de SOM comme suit :

$$\mathcal{J}_{CSOM} = \sum_{\tilde{z}_i \in \tilde{\mathcal{Z}}} \sum_{c \in \mathcal{C}^*} K^T(\sigma(\mathcal{X}(\tilde{z}_i), c)) \left( \sum_{b=1}^B \mu_b \|\tilde{z}_i^b - \tilde{w}_c^b\|^2 \right)$$

# Optimisation de $\mathcal{J}_{CSOM}$

Processus itératif en deux phases identique celle de SOM

- Affectation des observations :

$$\mathcal{X}(\tilde{z}_i) = \arg \min_{c \in \mathcal{C}^*} \sum_{c \in \mathcal{C}^*} \mathcal{K}^T(\sigma(\mathcal{X}(\tilde{z}_i), c)) \sum_{b=1}^B \mu_b \|\tilde{z}_i^b - \tilde{w}_c^b\|^2$$

- Actualisation des centres de classe

$$\tilde{w}_c^* = \frac{\sum_{\tilde{z}_i \in \tilde{\mathcal{Z}}} \mathcal{K}^T(\sigma(\mathcal{X}(\tilde{z}_i), c)) \tilde{y}_i}{\sum_{\tilde{z}_i \in \tilde{\mathcal{Z}}} \mathcal{K}^T(\sigma(\mathcal{X}(\tilde{z}_i), c))}$$

où  $y_i = [\sqrt{\mu_1} \tilde{z}_i^1, \dots, \sqrt{\mu_B} \tilde{z}_i^B]$  et  $\tilde{z}_i^b$  la portion de  $\tilde{z}_i$  restreinte aux dimensions du bloc  $b$ .

## Détermination des poids $\mu$

$$X^b = \tilde{Z}^b \tilde{Z}^{b'} \quad (N \times N) \text{ et } HS(X^a, X^b) = \text{trace}(DX^a DX^b)$$

$$R_V(X^a, X^b) = HS \left( \frac{X^a}{\|X^a\|_{HS}}, \frac{X^b}{\|X^b\|_{HS}} \right) = \frac{\text{trace}(DX^a DX^b)}{\sqrt{\text{trace}((DX^a)^2) \text{trace}((DX^b)^2)}}$$

$$\tilde{C}o = \max_C \sum_{b=1}^B R_V^2(C, X^b) \text{ avec } \tilde{C}o = \sum_{b=1}^B \sigma_b X^b$$

Ici on retrouve le critère d'optimisation de la méthode STATIS (Lavit, 1994)

Les solutions, les poids  $\sigma_b$  se définissent comme les coordonnées du premier vecteur propre de la matrice des corrélations vectorielles  $R_V$ .

# La partition consensus

- La carte Fusion s'obtient par application de l'algorithme SOM sur la matrice :

$$[\sigma_1 \tilde{Z}^1, \dots, \sigma_B \tilde{Z}^B]$$

- Il est aussi possible d'appliquer un algorithme de partitionnement directement sur la matrice  $\tilde{C}_o$

## CSOM

D	Indice	SOM	KM	NMF	WNMF	CSPA	CSOM	CSOM <sub>s</sub>
IS	Pur	0.61	0.61	0.45	0.58	0.37	<b>0.65</b>	0.60
	NMI	0.59	0.62	0.36	0.51	0.27	<b>0.63</b>	0.58
	NVI	0.58	0.56	0.78	0.65	0.84	<b>0.54</b>	0.58
CT3	Pur	0.78	0.78	0.78	0.78	0.78	0.78	0.78
	NMI	0.13	0.15	0.10	0.10	0.15	<b>0.16</b>	0.12
	NVI	0.93	0.92	0.95	0.96	0.92	<b>0.91</b>	0.94
DMU	Pur	0.78	0.76	0.84	0.86	<b>0.88</b>	0.84	0.75
	NMI	0.75	0.76	0.79	0.81	0.83	<b>0.85</b>	0.76
	NVI	0.39	0.38	0.34	0.31	0.28	<b>0.26</b>	0.39
FGKM	Pur	0.55	<b>0.61</b>	0.53	0.52	0.51	0.60	0.59
	NMI	0.18	<b>0.27</b>	0.12	0.11	0.12	0.22	0.22
	NVI	0.89	<b>0.84</b>	0.94	0.94	0.93	0.87	<b>0.84</b>

TABLE : Performances des algorithmes SOM, K-moyennes (KM), NMF, WNMF, CSPA et Fusion de SOM (CSOM et CSOM<sub>s</sub>) sur les bases IS, CT3, DMU et FGKM. Les valeurs entre parenthèses sont les écarts types des 25 apprentissages.

D	Indice	SOM	KM	NMF	WNMF	CSPA	$R_V$ -CSOM	$R_V$ -CSOM <sub>1</sub>
IS	Pur	0.61	0.61	0.45	<b>0.58</b>	0.37	0.55	0.53
	NMI	0.59	0.62	0.36	0.51	0.27	<b>0.57</b>	0.44
	NVI	0.58	0.56	0.78	0.65	0.84	<b>0.61</b>	0.71
CT3	Pur	0.78	0.78	0.78	0.78	0.78	<b>0.78</b>	0.78
	NMI	0.13	0.15	0.10	0.10	0.15	<b>0.20</b>	0.11
	NVI	0.93	0.92	0.95	0.96	0.92	<b>0.89</b>	0.94
DMU	Pur	0.76	0.78	0.85	0.85	<b>0.86</b>	0.54	0.82
	NMI	0.76	0.77	0.79	0.81	<b>0.82</b>	0.57	0.79
	NVI	0.39	0.37	0.34	0.31	<b>0.30</b>	0.60	0.35
FGKM	Pur	0.55	<b>0.61</b>	0.53	0.52	0.51	<b>0.59</b>	0.53
	NMI	0.18	<b>0.27</b>	0.12	0.11	0.12	<b>0.23</b>	0.14
	NVI	0.89	<b>0.84</b>	0.94	0.94	0.93	<b>0.87</b>	0.93

**TABLE :** Performances des algorithmes SOM, K-moyennes (KM), NMF, WNMF, CSPA,  $R_V$ -CSOM et  $R_V$ -CSOM<sub>1</sub> sur les bases IS, CT et DMU ;



# Conclusions

La méthode de classification présentée permet de réduire l'effet des blocs comportant des variables distribuées uniformément dans la classification.

- Le système de poids introduit permet de définir les blocs importants au sens de la variance d'une classe.
- Les approches proposées permettent d'apporter les premiers éléments de réponses à de nombreux problèmes de détection de groupes homogènes d'observations dans des données collectées au cours d'enquête de grande ampleur.

# Perspectives

- Redéfinir les blocs à l'aide d'une méthode classification des variables puis d'appliquer la méthode de partitionnement multi-blocs 2S-SOM sur ces blocs.
- la définition des blocs en même temps que la recherche des classes.

## Merci pour votre attention

- Pacifico, L.D.S. et de. Carvalho (2011). A batch self organizing maps algorithm based on adaptive distances. Neural Networks (IJCNN), The 2011 International Joint Conference on 3, 2297-2304.
- Chen, X., Ye, Y., al., 2012. A feature group weighting method for subspace clustering of high-dimensional data. Pattern Recogn 45(1), 434-446.
- Kohonen, T. (1998). The self-organizing map. Neurocomputing 21, 1-3.
- Strehl, A. et J. Ghosh (2002). Cluster ensembles - a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research 3, 583-617.