Cartes auto-organisatrices pour l'analyse de données

Vincent Lemaire

Plan

1. Généralités et algorithme
2. Visualisations des données et interprétation
3. Analyse exploratoire en 'grande' dimension
4. Sélection de variable, données manquantes…
5. Utilisation de distances adaptées

Généralités

- Carte auto-organisatrice ou carte de Kohonen

  Algorithme de classification développé par Teuvo Kohonen dès 1982

- Propriétés
  - Apprentissage non supervisé
  - Méthode de quantification vectorielle
  - Regroupement des informations en classes tout en respectant la topologie de l'espace des observations
    - définition a priori d'une notion de voisinage entre classes
    - des observations voisines dans l'espace des données appartiennent après classement à la même classe ou à des classes voisines
    - compression de données multidimensionnelles tout en préservant leurs caractéristiques

Description du modèle

- Deux espaces indépendants

  l'espace des données généralement de grande dimension

  l'espace des représentations (la carte) de dimension réduite

  les nœuds de la carte sont disposés géométriquement selon une topologie fixée a priori

- Trouver la projection entre les deux espaces

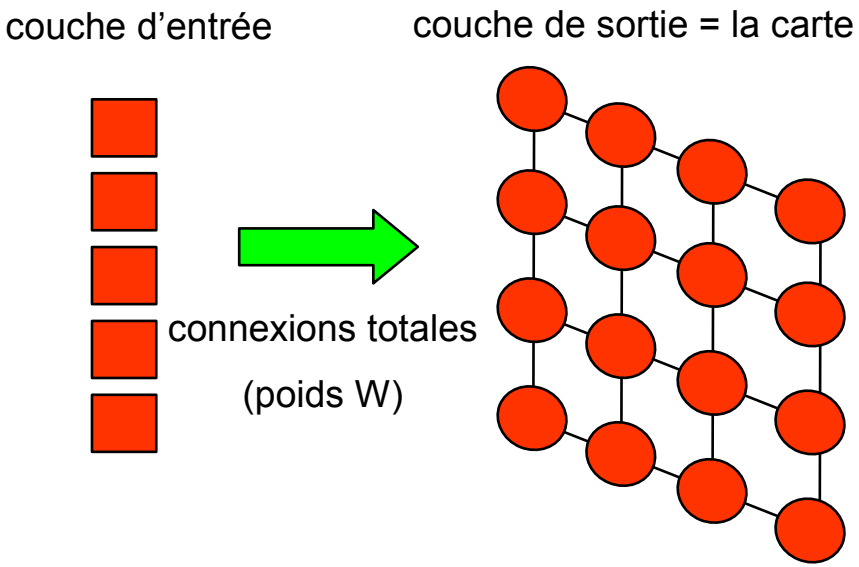
  la projection doit conserver la topologie des données

- Un nœud dans la carte possède des

  coordonnées ***fixes*** sur la carte

  coordonnées ***adaptables*** *W*

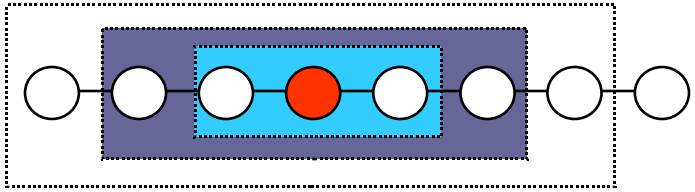
  dans l'espace d'entrée original



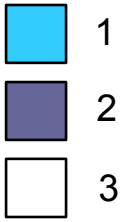
Différentes topologies et voisinages

- La structure topologique de la carte introduit la notion de voisinage et de distance entre les neurones de la carte

- Carte unidimensionnelle ou fil

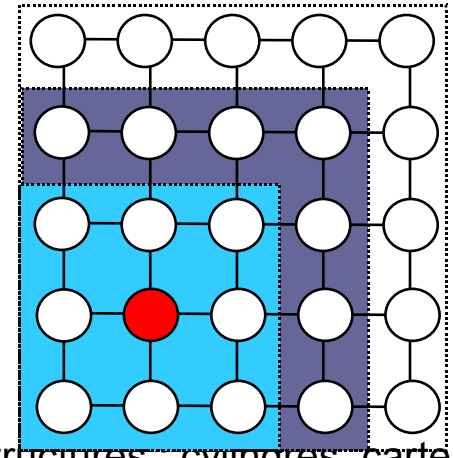


rayon de voisinage

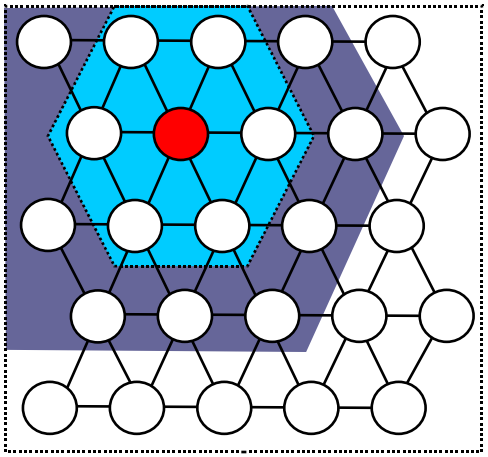


- Carte bi-dimensionnelle (carrée, rectangle)

Voisinage rectangulaire



Voisinage hexagonal



- Autres structures : cylindres, carte 3D

Apprentissage de la carte

- L'apprentissage met en correspondance l'espace des entrées et la carte

  Adaptation des poids *W* de telle manière que des exemples proches dans l'espace d'entrée sont associés au même neurone ou à des neurones proches dans la carte

- Algorithme

  Initialisation aléatoire des poids *W*

  A chaque itération t

  Présentation d'un exemple d'apprentissage *X(t),* choisi au hasard, à l'entrée de la carte

  Comparaison de l'exemple à tous les vecteurs poids, le neurone gagnant *j** est celui dont le vecteur poids $W_{j^*}(t)$ est le plus proche de l'entrée *X(t)* *(phase de compétition)*

$$d_N(X(t), W_{j^*}(t)) = \min_j d_N(X(t), W_j(t))$$

  Évaluation du voisinage du neurone gagnant dans la carte

$$h_{j*}(j,t) = h(d(j,j^*),t)$$

  d_N distance dans l'espace d'entrée

  d distance dans la carte

  ε pas d'apprentissage

  h fonction voisinage

  Mise à jour des poids pour tous les neurones de la carte, l'adaptation est d'autant plus forte que les neurones sont voisins de *j** *(phase de coopération)*

$$W_j(t+1) = W_j(t) + \Delta W_j(t)$$
$$\Delta W_j(t) = \varepsilon(t) \cdot h_{j*}(j,t) \cdot (X(t) - W_j(t))$$

Paramètres d'apprentissage

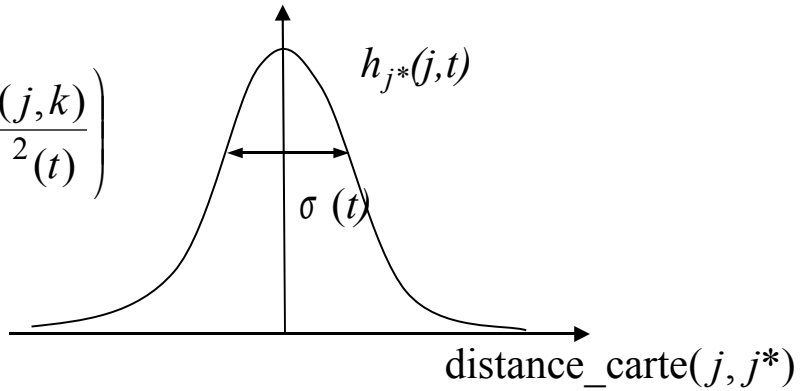
- La fonction de voisinage dans la carte est une fonction continue de forme gaussienne

  La décroissance de la taille du voisinage s'obtient par diminution de l'écart type σ

  σ grand, beaucoup de neurones se rapprochent de *X(t)*, σ petit, l'adaptation reste très localisée

$$h_j(k,t) = \exp\left(-\frac{d^2(j,k)}{2\sigma^2(t)}\right)$$

$$\sigma(t) = \sigma_i\left(\frac{\sigma_f}{\sigma_i}\right)^{\frac{t}{t\max}} \qquad \varepsilon(t) = \varepsilon_i\left(\frac{\varepsilon_f}{\varepsilon_i}\right)^{\frac{t}{t\max}}$$



- Le pas de l'apprentissage contrôle la vitesse d'apprentissage

  ε trop petit, le modèle ne s'adapte pas assez aux données, ε trop grand, risque d'instabilité

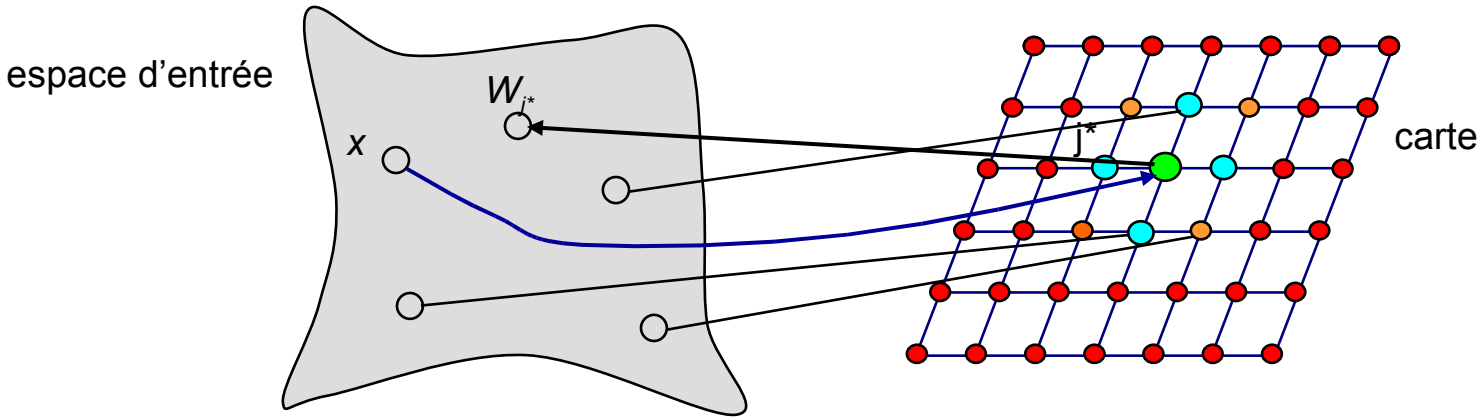
- Condition de convergence

  Choisir les valeurs σ et ε grandes au départ et les réduire graduellement (on choisira des paramètres à décroissance exponentielle)

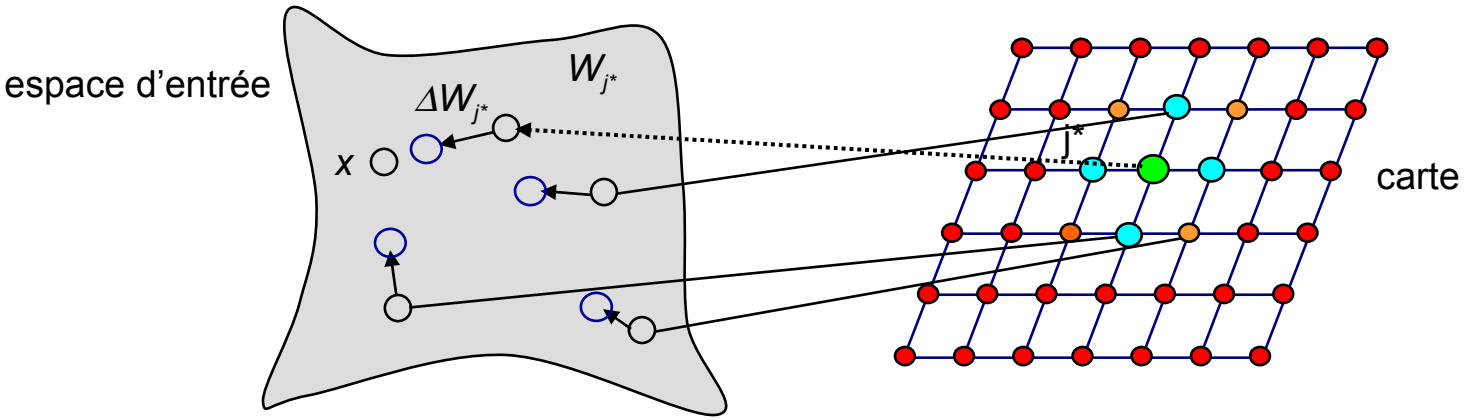
  ordonnancement rapide au départ qui s'affine ensuite

Interprétation de l'algorithme

- Sélection du neurone le plus proche de l'exemple présenté *X*



- Adaptation des poids

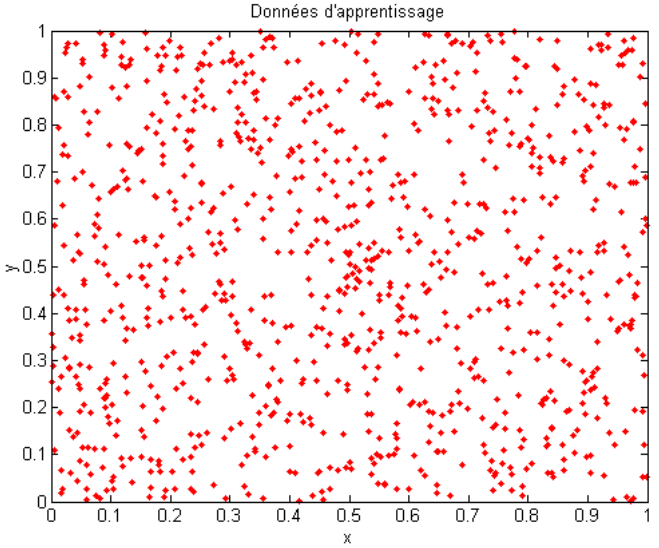


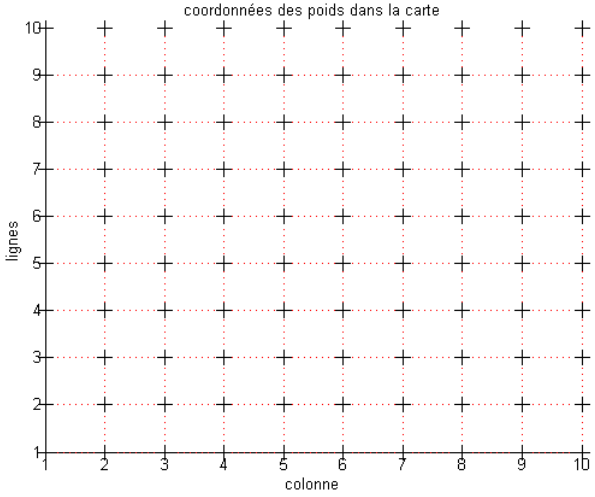
Les poids du neurone gagnant *j** et de ses voisins sont rapprochés plus fortement du vecteur d'entrée *X*

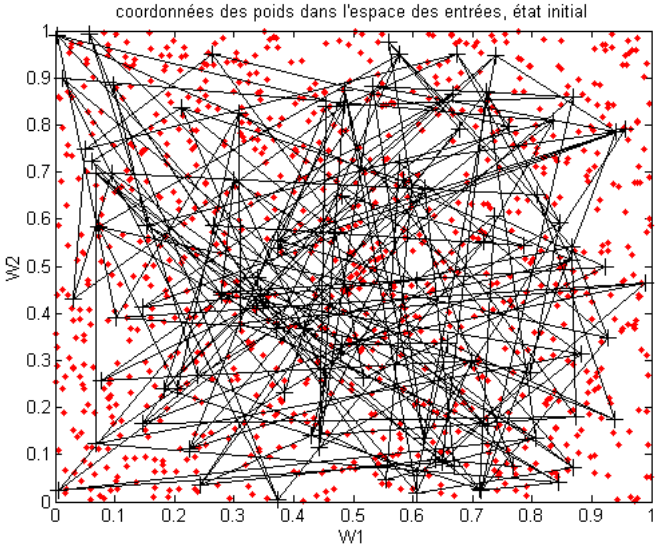
L'amplitude de l'adaptation est fonction de la distance du neurone au neurone gagnant (flèches de longueurs différentes)

Démonstration en dimension 2

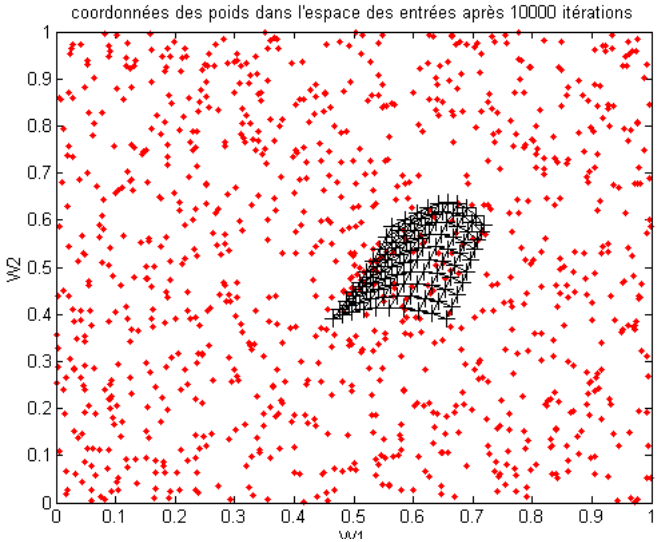
- 1000 exemples d'apprentissage, les données sont uniformément distribuées dans un carré
- Grille 10x10 avec un voisinage carré dans la carte
- A partir d'une carte totalement désordonnée (initialisation aléatoire des poids), l'algorithme doit converger vers une carte organisée qui conserve la topologie de l'espace d'entrée

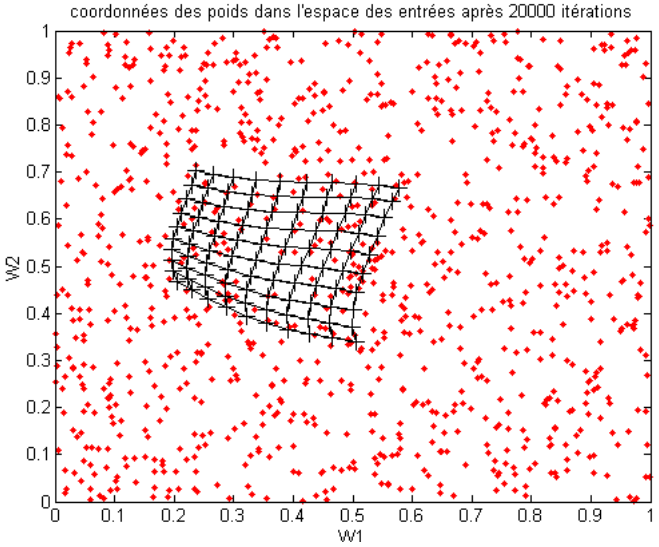


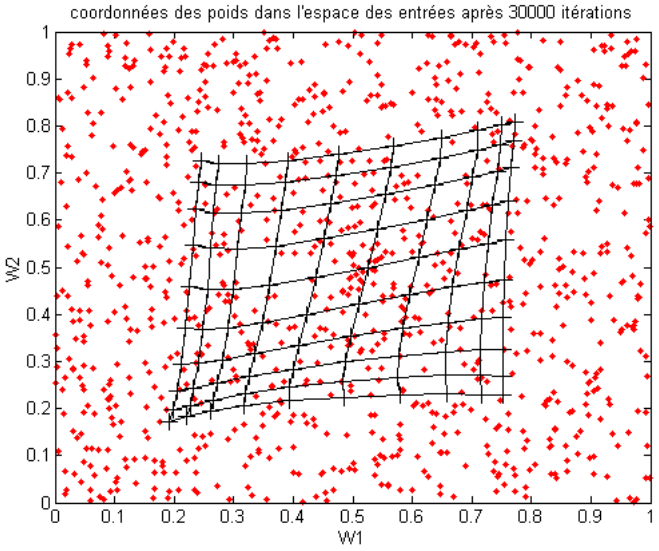


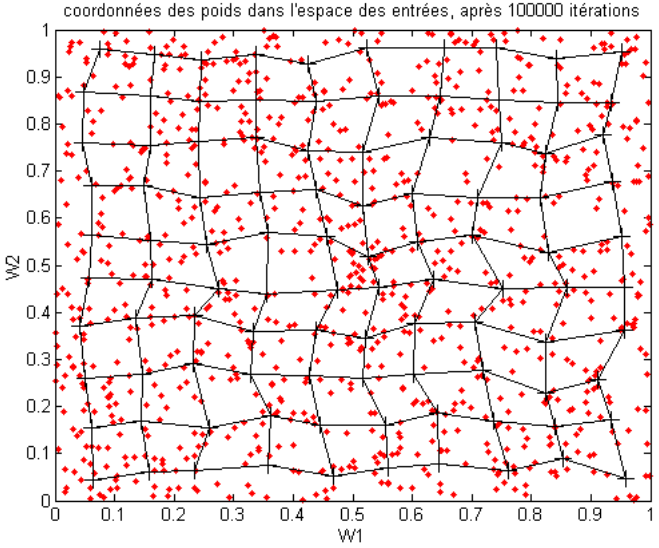


Démonstration en dimension 2









Démonstration en dimension 2

- Auto-organisation

  la répartition des neurones devient uniforme au fur et à mesure de l'apprentissage

  la répartition des neurones est calquée sur la densité de probabilité des vecteurs d'apprentissage

  après apprentissage, un exemple d'entrée sera représenté par le neurone dont il se rapproche le plus

  une fois l'apprentissage achevé, les valeurs des connexions définissent un pavage de Voronoi de l'espace des entrées qui reflète au mieux la distribution de probabilité des motifs d'entrée

- La carte s'organise en vérifiant les propriétés suivantes

  chaque neurone se spécialise dans une portion de l'espace d'entrée

  deux entrées qui sont proches dans l'espace d'entrée provoquent la réponse du même neurone ou de deux neurones voisins dans la carte

  chaque neurone a la même probabilité d'être activé : on trouvera plus de neurones spécialisés dans des zones de l'espace d'entrée où la probabilité est plus grande que dans celles où les observations sont rares

Plan

1. Généralités et algorithme
2. Visualisations des données et interprétation
3. Analyse exploratoire en 'grande' dimension
4. Sélection de variable, données manquantes…
5. Utilisation de distances adaptées

Analyse exploratoire de données

- Analyse exploratoire

  visualiser « intelligemment » les données

- Beaucoup d'individus (N) X beaucoup de variables (p)

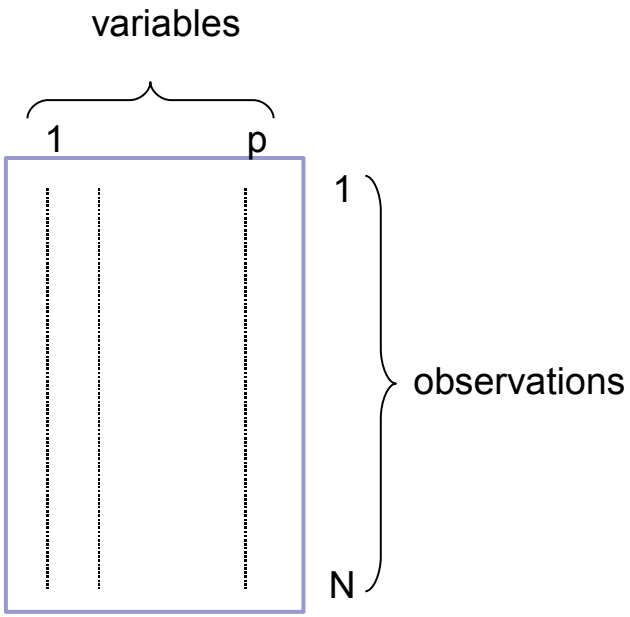
  effectuer des regroupements qui respectent la structure des données

  rendre visibles les similarités entre données

- Les cartes de Kohonen qui permettent de réaliser une segmentation et de la visualiser

  projection non-linéaire sur un plan (donc en dimension 2)

  qui respecte en dimension 2 les relations de similarité existant en dimension V



Exemple d'analyse exploratoire

- Description des données

  53 pays décrits par 6 indicateurs sociaux et économiques (1991) : croissance annuelle (%), mortalité infantile (‰), taux d'analphabétisme (%), taux de scolarisation (%), PNB par habitant, augmentation annuelle du PNB (%)

  Représentation des données sous la forme d'une table de 53 lignes et 6 colonnes

pays	croissance annuelle	taux de mortalité infantile	taux d'analphabétisme	taux de scolarisation	PNB par habitant	augmentation du PNB
Afrique sud	2,9	89	50	19	2680	-2,9
Algérie	2,9	114	58,5	47,9	2266	0,1
Argentine	1,2	44	5,3	69,5	2264	2
Australie	1,3	10,4	0	86	9938	-1,2
Bahreïn	3,8	57	20,9	76,3	8960	-10,1
Brésil	2,2	75	23,9	62,3	1853	-3,9

- Construction de la carte auto-organisatrice à partir des données

  un vecteur d'entrée de la carte = une ligne de la table (6 dimensions d'entrée)

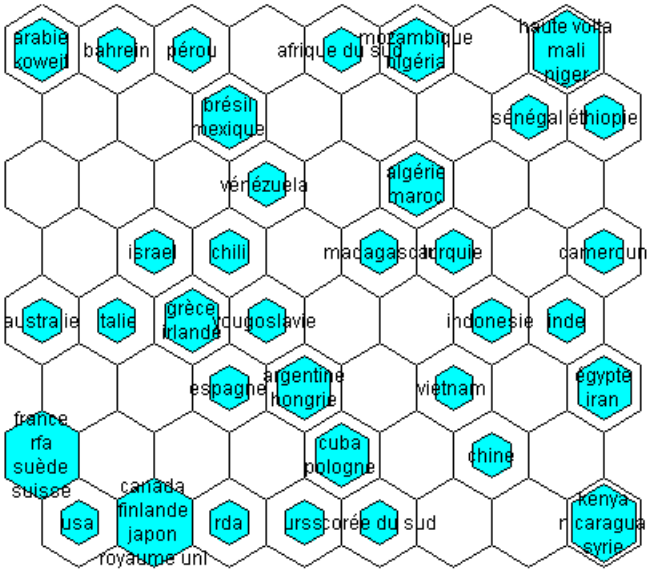
  paramètres de la carte : carte carrée 8 lignes, 8 colonnes (64 neurones), voisinage hexagonal

- Après apprentissage, chaque individu (un vecteur d'entrée) est associé à un neurone de la carte (le neurone le plus proche de l'individu)

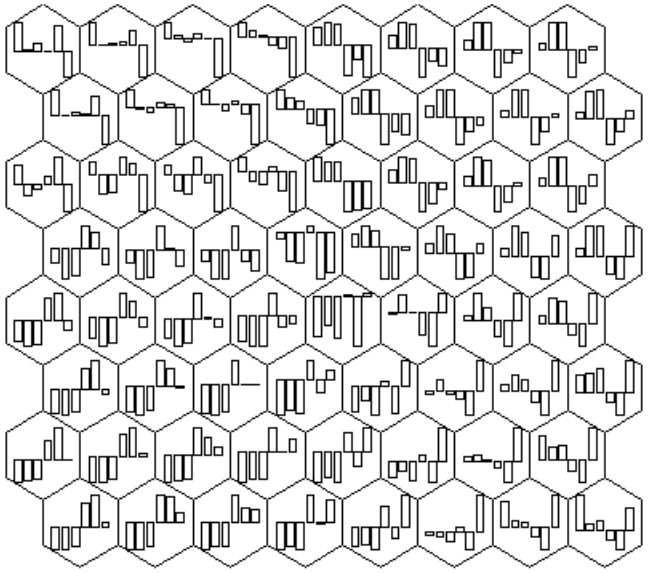
Représentations (1)

- Différents outils de visualisation

carte des individus



histogramme des vecteurs de poids



- Carte des individus

  on liste sur la carte les observations classées par le neurone (figuré par un hexagone)

  les observations semblables ont la même projection

  la taille de l'hexagone est proportionnelle au nombre d'individus classés par le neurone

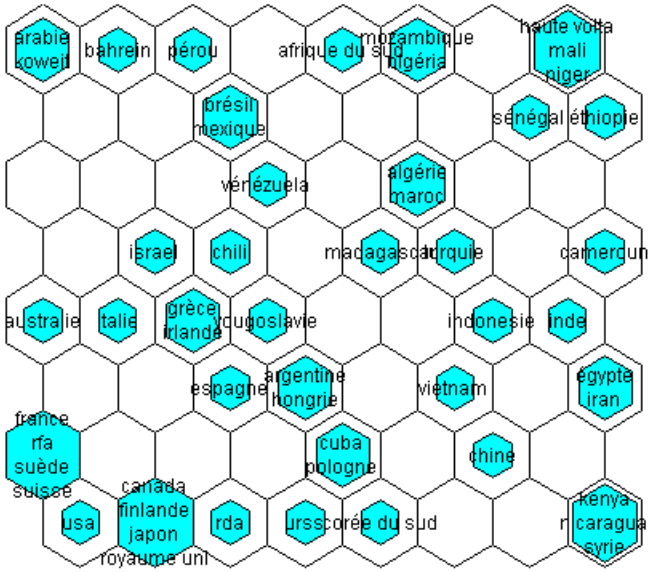
- Histogramme des poids

  on représente, pour chaque neurone, les 6 composantes du vecteur de poids avec un histogramme

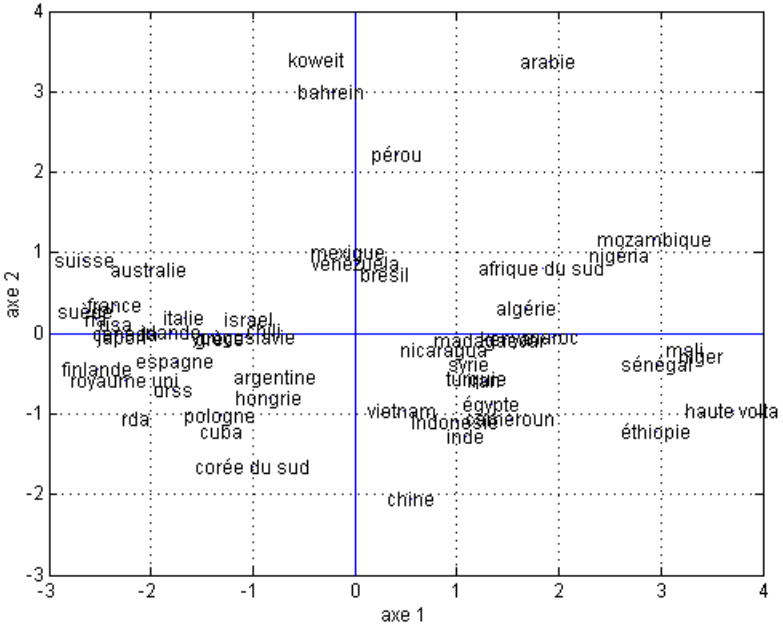
Représentations (2)

- Représentation plane analogue à l'ACP (mais représentation sur une seule carte)

carte des individus



ACP : projection sur les 2 premiers axes



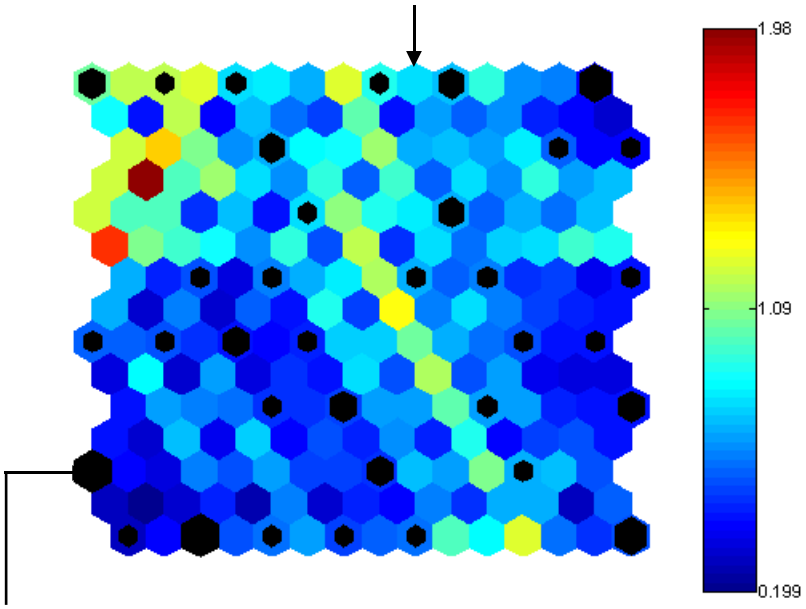
- La carte peut être vue comme une extension non linéaire de l'ACP

- L'ACP est une méthode linéaire qui réalise une réduction de dimension par projections orthogonales sur un plan, une représentation linéaire n'est pas toujours adaptée au nuage des points
- Pour l'exemple étudié on constate une répartition plus uniforme des pays sur la carte de Kohonen

Représentations (3)

- Carte des distances (ou U-matrice)

Distance euclidienne entre deux neurones voisins dans la carte



population du neurone

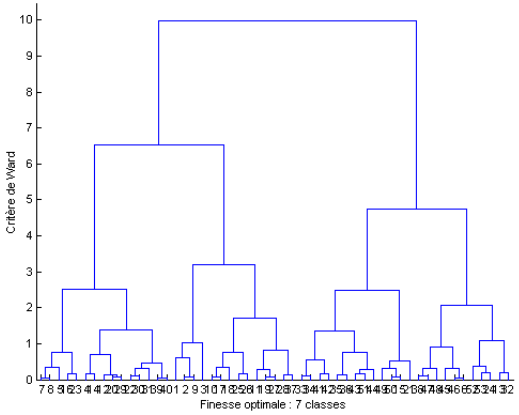
- Une couleur rouge indique une zone frontière

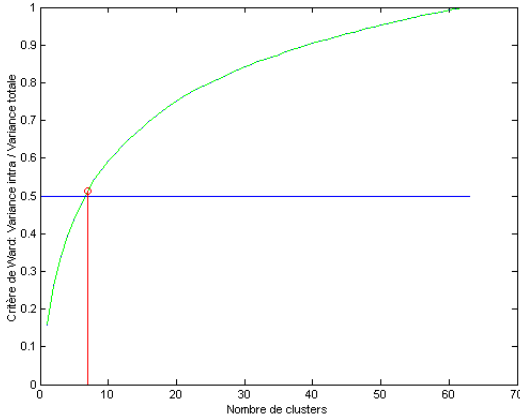
- Une couleur bleue indique que les distances entre neurones sont faibles et caractérise une zone homogène dans la carte (ensemble de neurones peu différents)

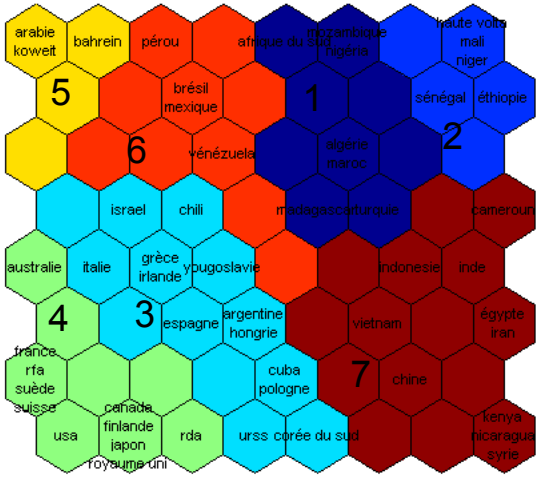
Les zones frontières coïncident souvent avec les unités vides de la carte

Exemple de clustering à partir de la carte

- Les neurones de la carte sont regroupés à l'aide d'une procédure de classification hiérarchique ascendante
- Le critère de regroupement est le critère de Ward
- On considère que le nombre optimal de clusters est obtenu pour un rapport inertie intra classe /inertie totale de 0,5



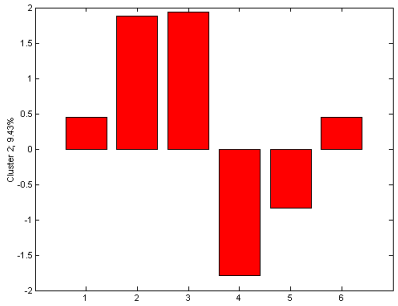




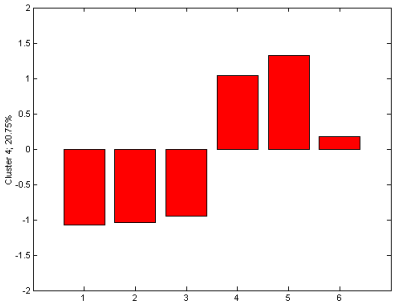
Les clusters coïncident avec les zones homogènes de la carte des distances

Caractéristiques des clusters

2 : pays pauvres d'Afrique, 4 : pays riches, 5 : pays producteur de pétrole, etc …



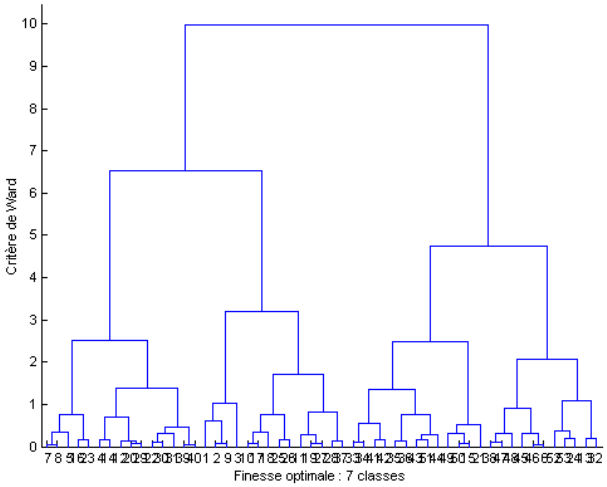
Profil du cluster 2

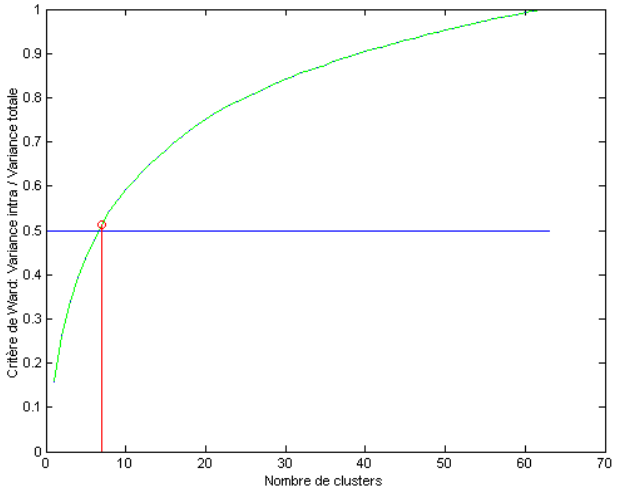


Profil du cluster 4

Classification ascendante hiérarchique

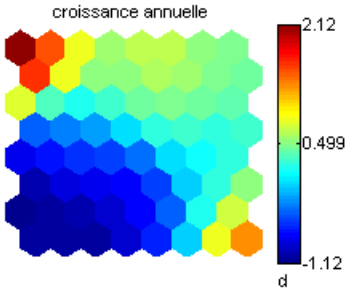
- La méthode de classification hiérarchique appliquée à la carte consiste à regrouper *au mieux* les neurones de manière à donner une vision plus globale de la carte
- On construit une suite de classifications emboîtées en regroupant les neurones les plus proches, puis les groupes les plus proches au sens d'une distance *convenable*
- On choisit en général la distance de Ward qui favorise des groupements les plus compacts possibles dans l'espace des données (qui font le moins varier l'inertie intraclasse)
- Le résultat est un dendrogramme
- Au fur et à mesure des regroupements la somme des carrés intra classe augmente de 0 jusqu'à la somme des carrés totale
- On considère que le nombre optimal de clusters est obtenu pour un rapport inertie intra classe /inertie totale de 0,5

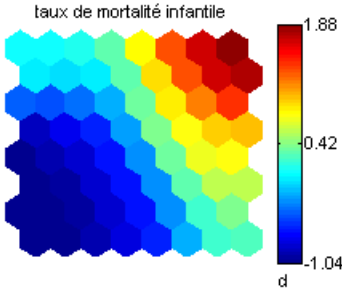


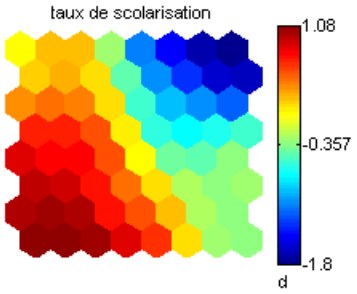


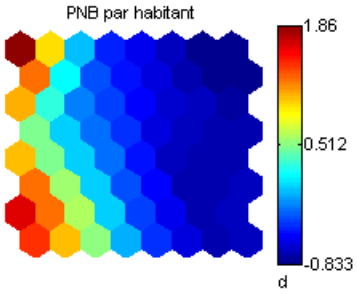
Représentations (4)

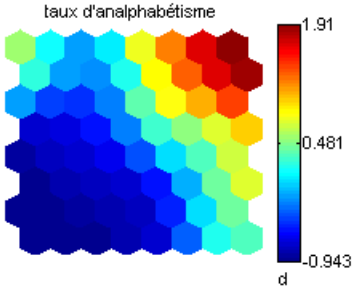
- Une carte de poids par dimension d'entrée

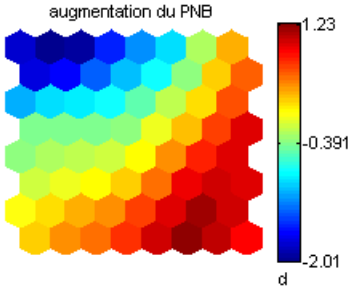


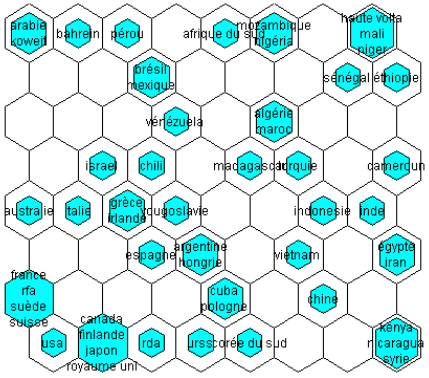












- Interprétation des cartes

  taux de mortalité infantile élevé pour les pays d'Afrique centrale (forte valeur de la variable dans le coin NE de la carte)

  taux d'analphabétisme faible pour les pays développés (faible valeur de la variable dans le coin SO de la carte)

- Les variables semblables ont la même projection

  les taux de mortalité infantile et d'analphabétisme sont très corrélés et anti-corrélés au taux de scolarisation, etc …

  étude de la corrélation des variables

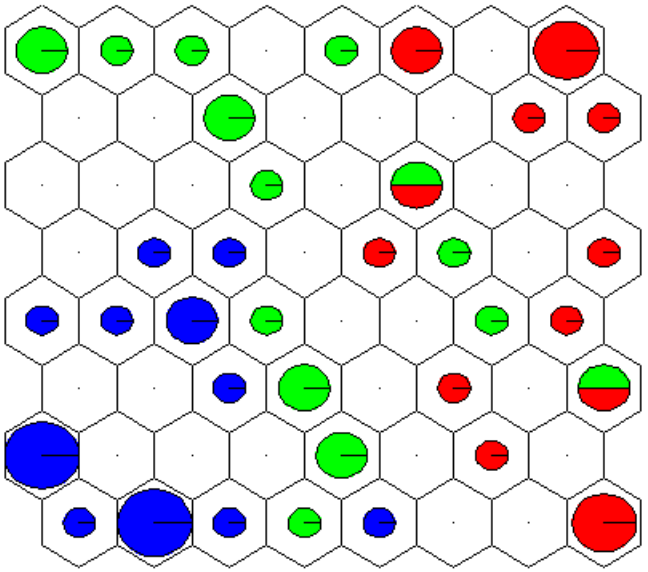
Projection d'informations sur la carte

- On souhaite étudier la relation entre une variable qui n'a pas été utilisée pour l'apprentissage de la carte et les variables qui ont servi à la construire

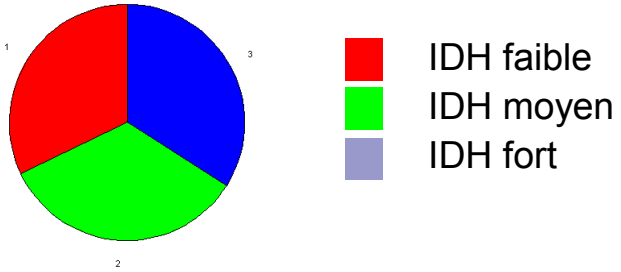
  projection de l'information supplémentaire sur la carte

  on trace pour chaque neurone un histogramme qui indique la répartition de la variable pour les observations classée par le neurone

- Application : projection de l'indice de développement humain



IDH codé sur trois valeurs



- On constate une forte inhomogénéité de la distribution de l'IDH dans la carte

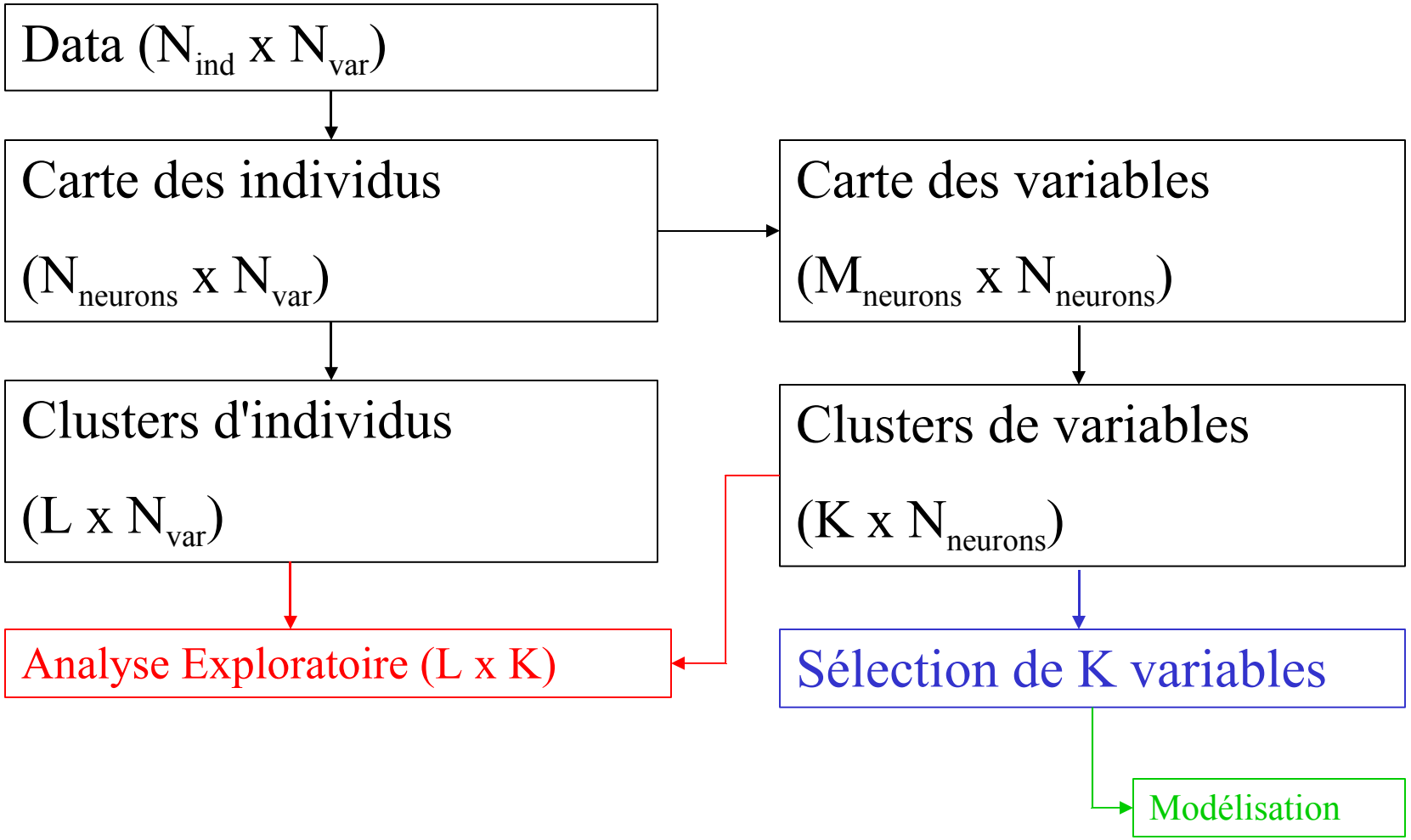
  Un IDH élevé correspond aux pays riche

  Un IDH faible correspond aux pays pauvres d'Afrique

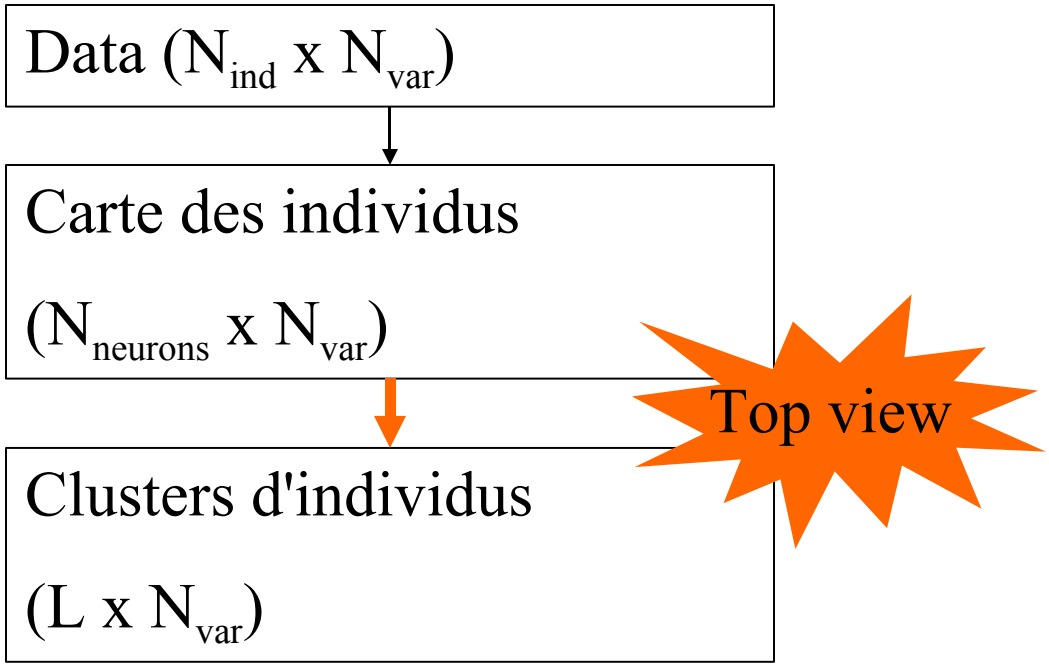
Plan

1. Généralités et algorithme
2. Visualisations des données et interprétation
3. Analyse exploratoire en 'grande' dimension
4. Sélection de variable, données manquantes…
5. Utilisation de distances adaptées

Analyse conjointe Individus x Variables

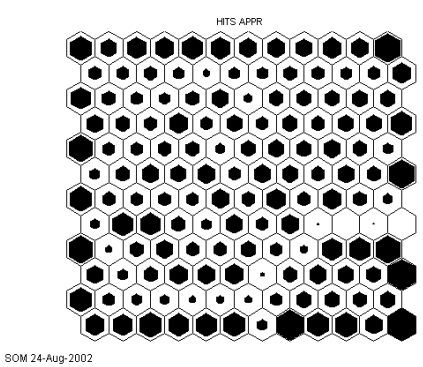


Analyse des individus

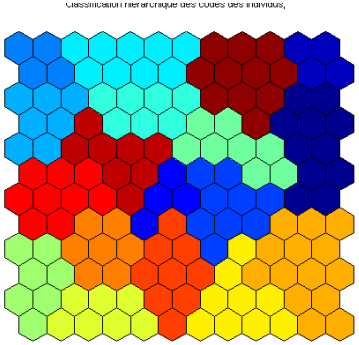


Les nombreux visages d'une carte

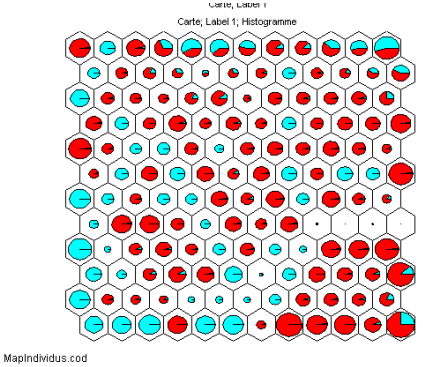
« Top view » : projection des individus



Populations



Clustering

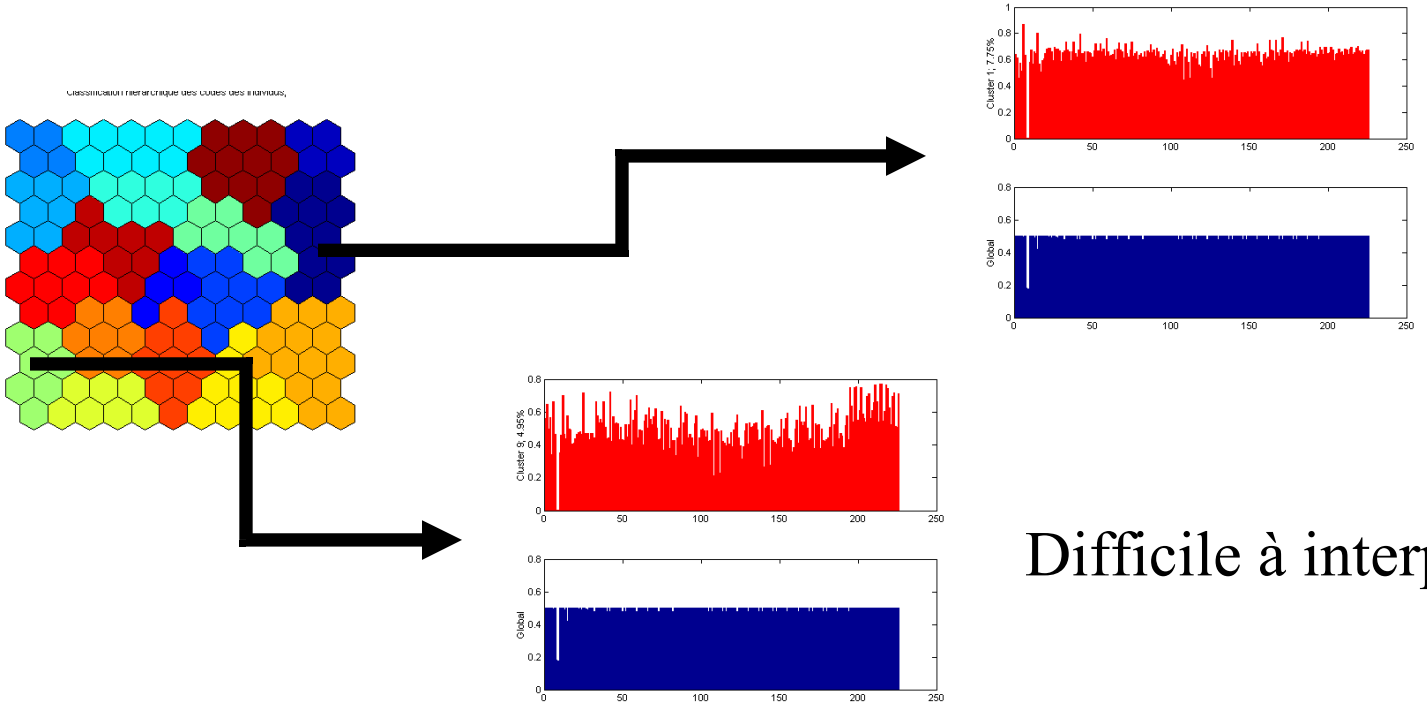


Projection

Les nombreux visages d'une carte

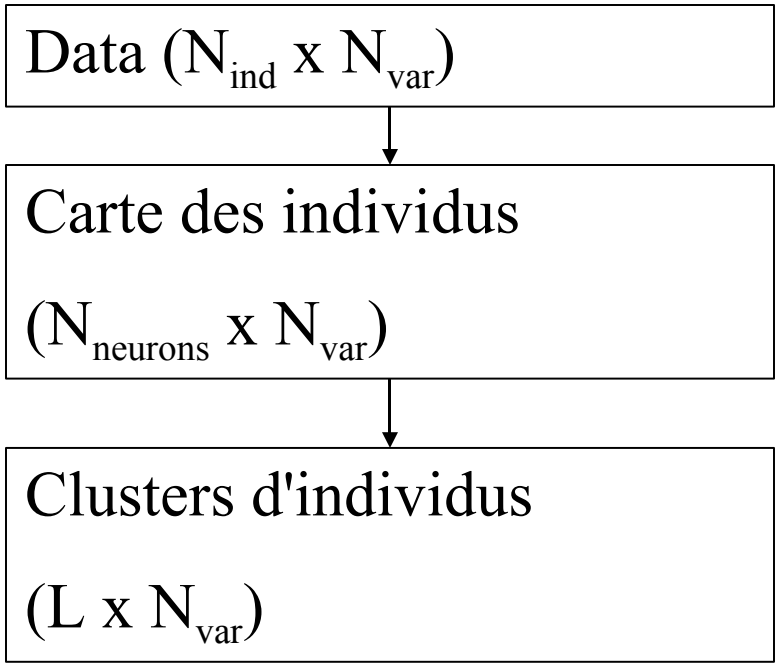
« Top view » : $\{N_p(v)\}_{p=1\ldots Nneurons}$

prototypes de la carte, vecteurs en N_{var} dimensions :

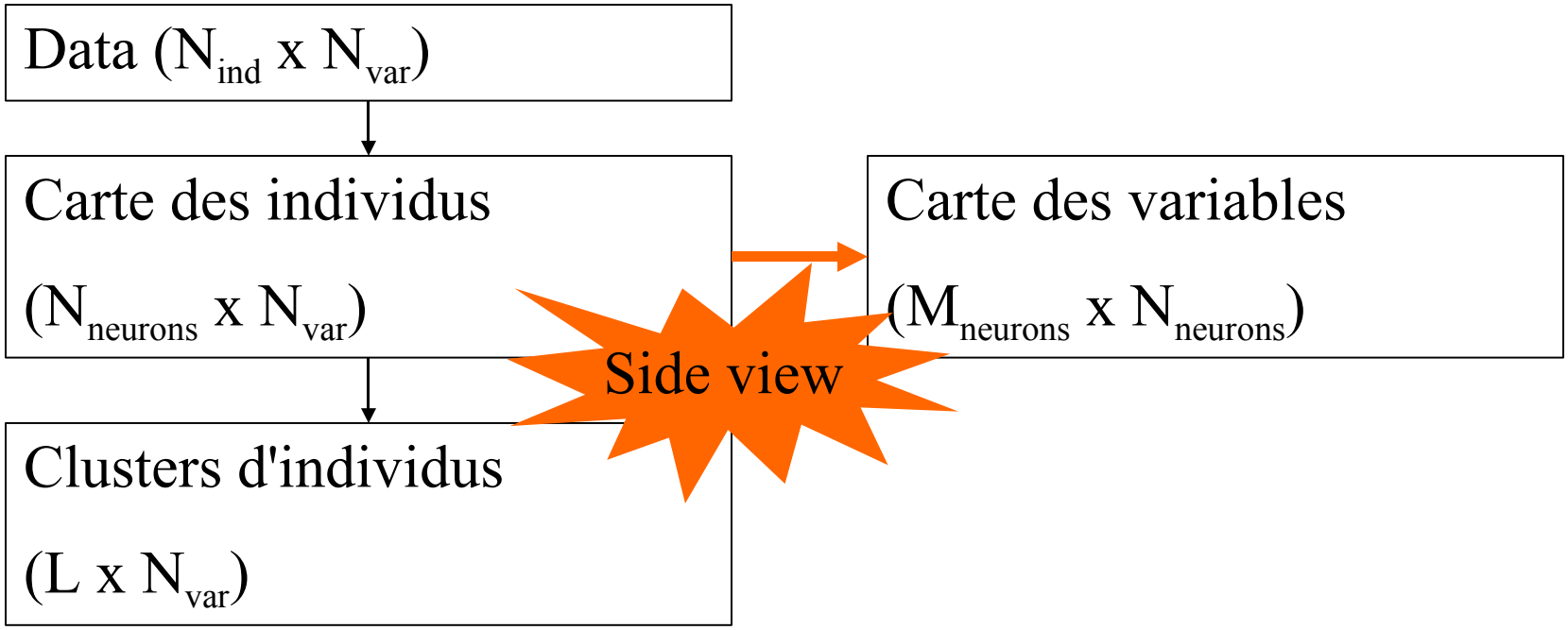


Difficile à interpréter

Les nombreux visages d'une carte : l'intérêt de la carte des variables



Les nombreux visages d'une carte : l'intérêt de la carte des variables

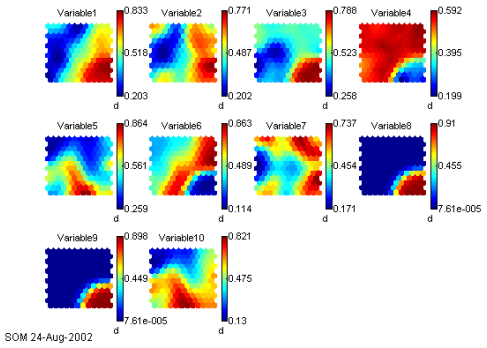


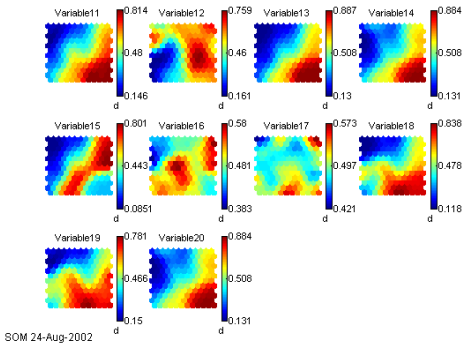
Les nombreux visages d'une carte : l'intérêt de la carte des variables

« Side view » : projection des variables

$\{M_v(p)\}_{v=1\ldots Nvar}$

Projections sur la carte des composantes des neurones
Une 'tranche' de la carte pour chaque variable :



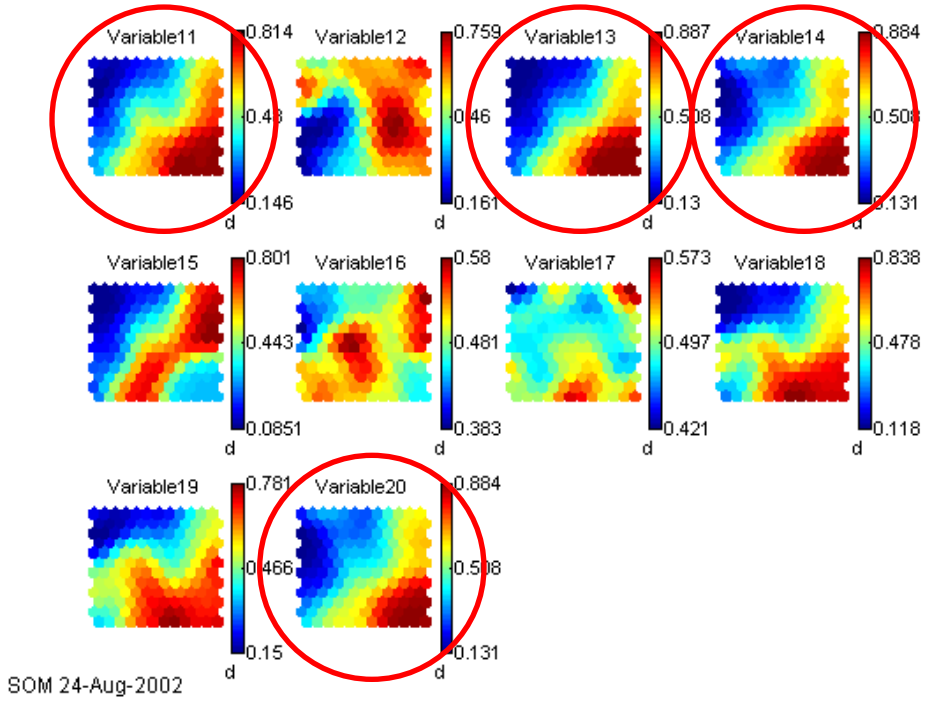


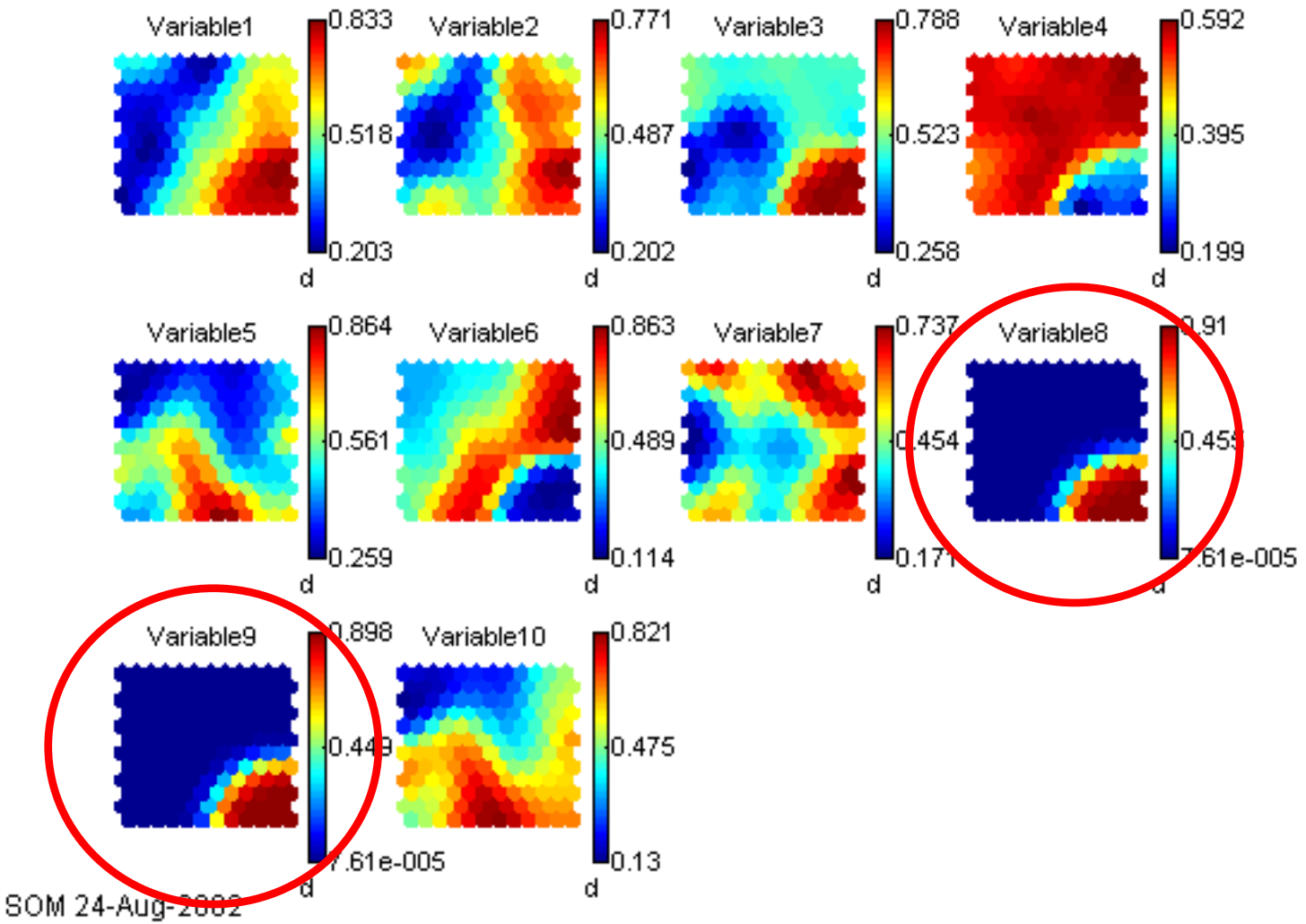
• • •

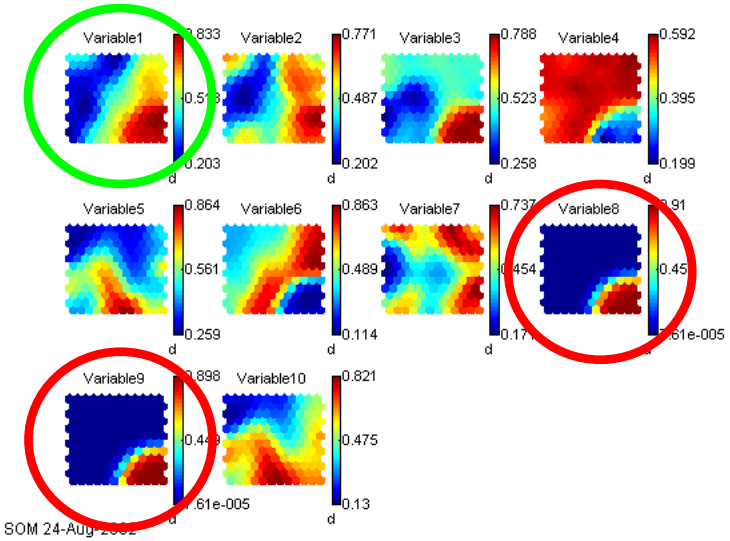
Les nombreux visages d'une carte : l'intérêt de la carte des variables

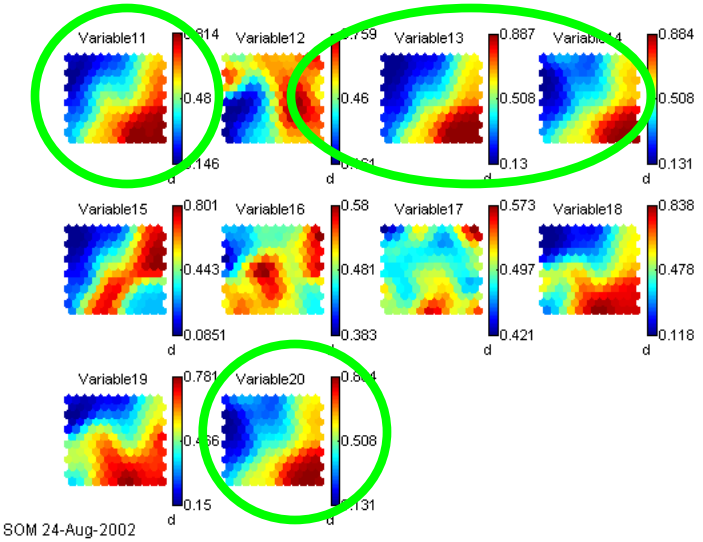
« Side view »

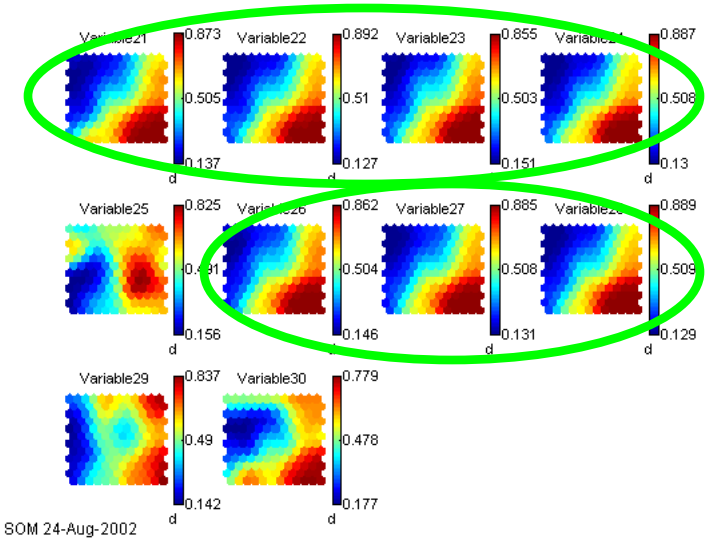
Chaque carte est représentative de la variable correspondante :
cela permet d'étudier les corrélations entre les variables
'au sens de la carte des individus'

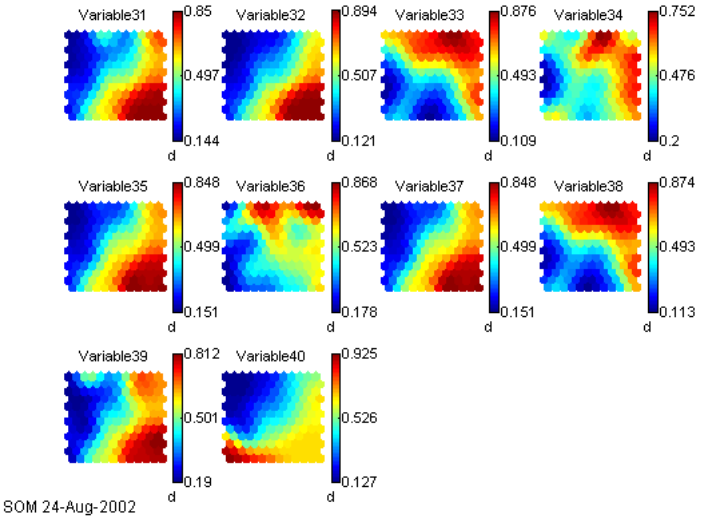


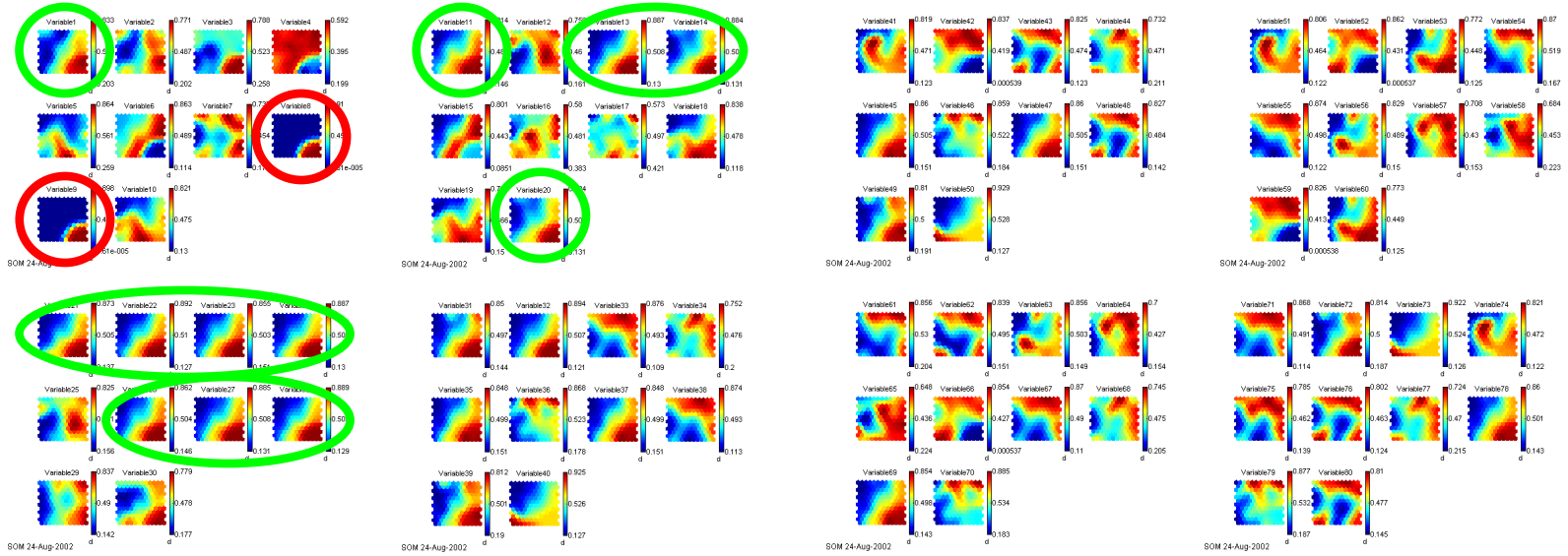

Variable1
0.833
0.518
0.203
d
Variable2
0.771
0.487
0.202
d
Variable3
0.788
0.523
0.258
d
Variable4
0.592
0.395
0.199
d
Variable5
0.864
0.561
0.259
d
Variable6
0.863
0.489
0.114
d
Variable7
0.737
0.454
0.171
d
Variable8
0.91
0.455
7.61e-005
d
Variable9
0.898
0.443
7.61e-005
d
Variable10
0.821
0.475
0.13
d
SOM 24-Aug-2002


Variable1
Variable2
Variable3
Variable4
Variable5
Variable6
Variable7
Variable8
Variable9
Variable10
SOM 24-Aug-2002


Variable11
Variable12
Variable13
Variable14
Variable15
Variable16
Variable17
Variable18
Variable19
Variable20
SOM 24-Aug-2002


Variable21
Variable22
Variable23
Variable24
Variable25
Variable26
Variable27
Variable28
Variable29
Variable30
SOM 24-Aug-2002


Variable31
Variable32
Variable33
Variable34
Variable35
Variable36
Variable37
Variable38
Variable39
Variable40
SOM 24-Aug-2002



• • •

Trop "d'inspections" visuelles à réaliser…

•Transformation de chaque carte en un vecteur, représentatif de la variable au sens de la carte des individus

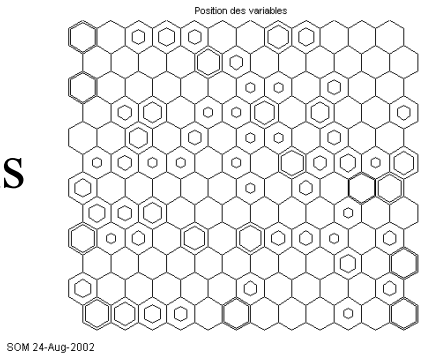
•Création d'une nouvelle carte : la carte des variables

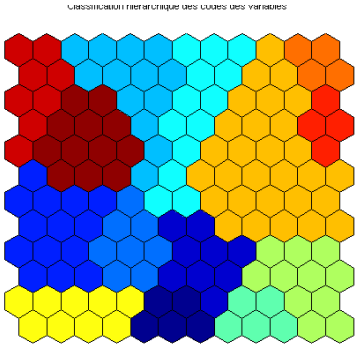
La carte des variables

Données : Projection des variables sur la carte des individus

Carte des variables….

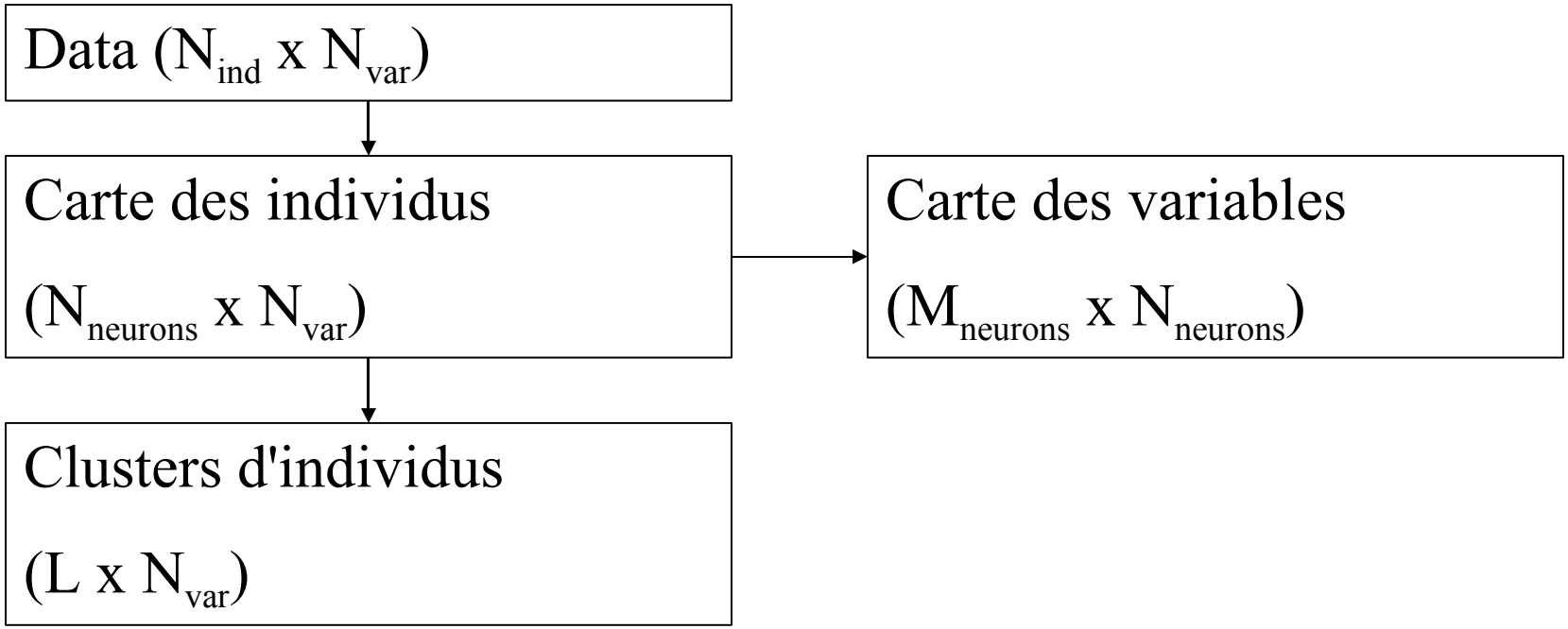
Populations



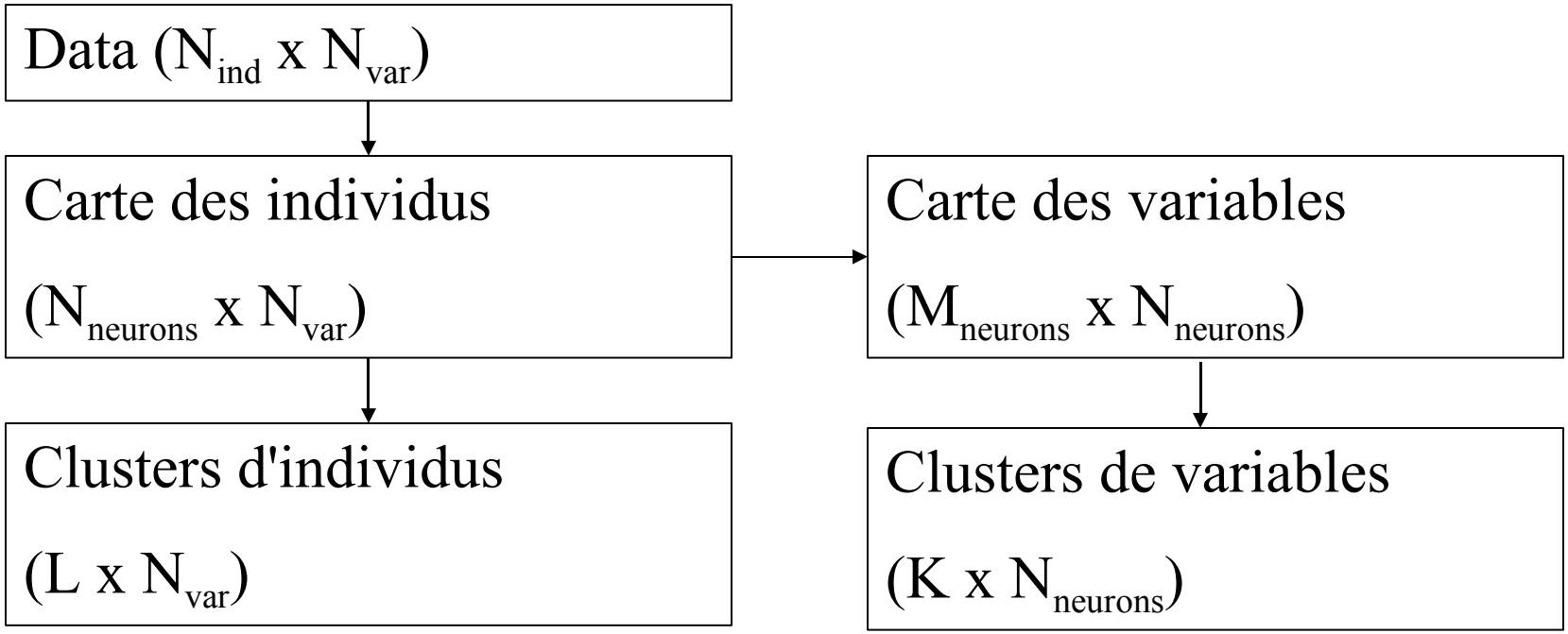


Clustering

Synopsis

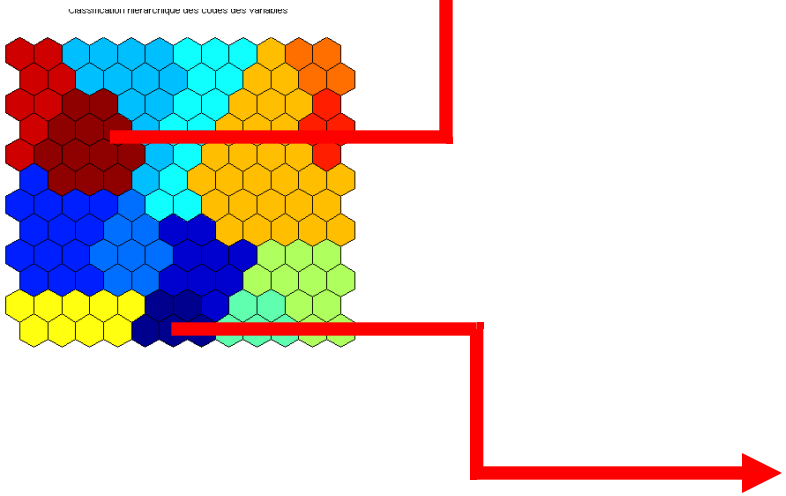

Data (N_{ind} x N_{var})
Carte des individus
($N_{neurons}$ x N_{var})
Carte des variables
($M_{neurons}$ x $N_{neurons}$)
Clusters d'individus
(L x N_{var})

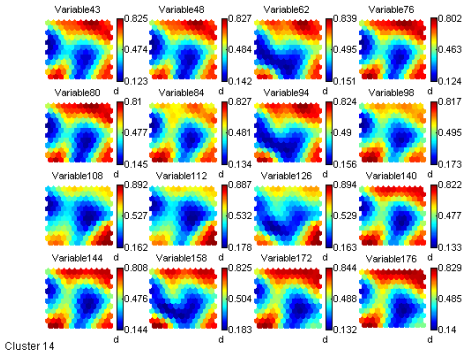
Synopsis

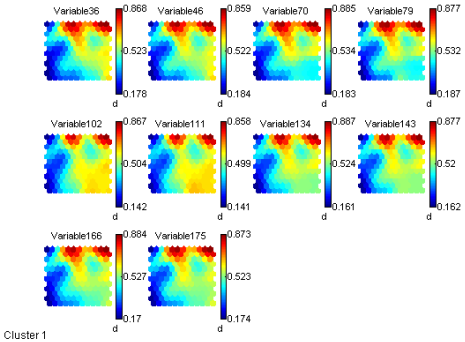

Data (N_{ind} x N_{var})
Carte des individus
($N_{neurons}$ x N_{var})
Clusters d'individus
(L x N_{var})
Carte des variables
($M_{neurons}$ x $N_{neurons}$)
Clusters de variables
(K x $N_{neurons}$)

Carte des variables

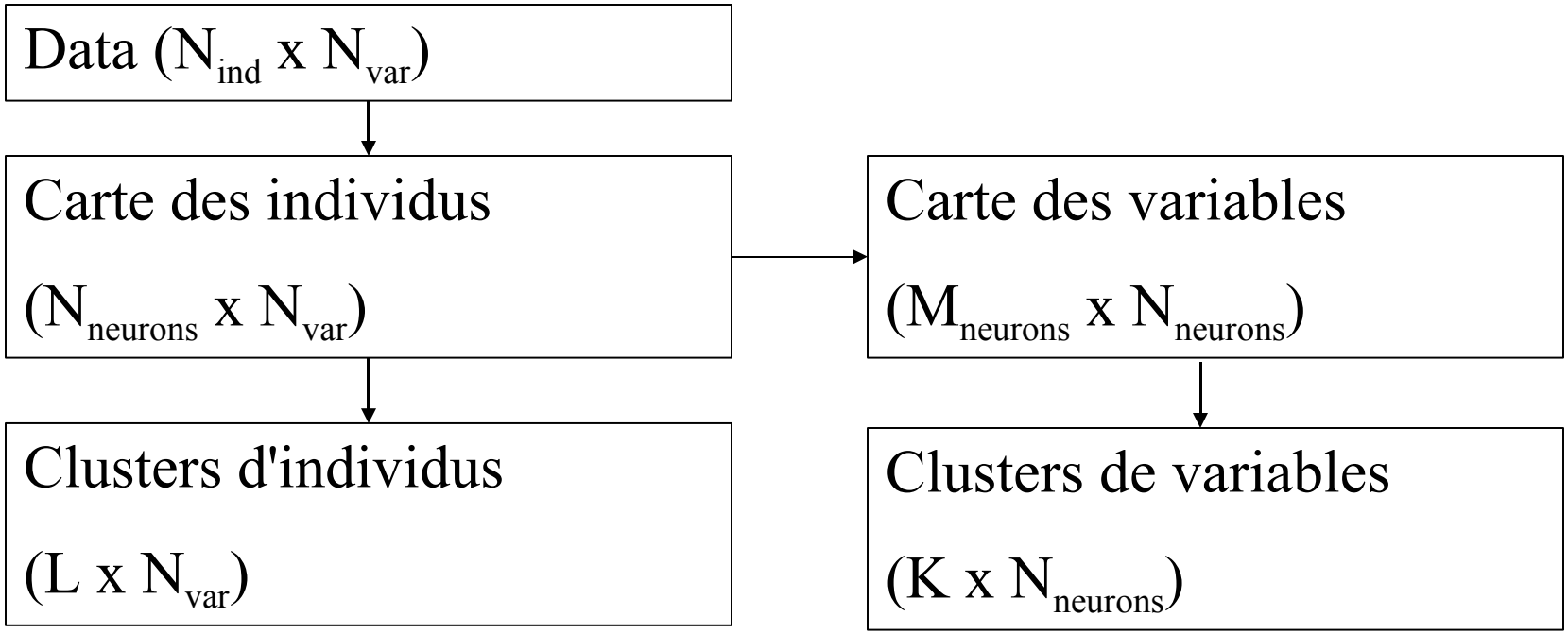
Clustering



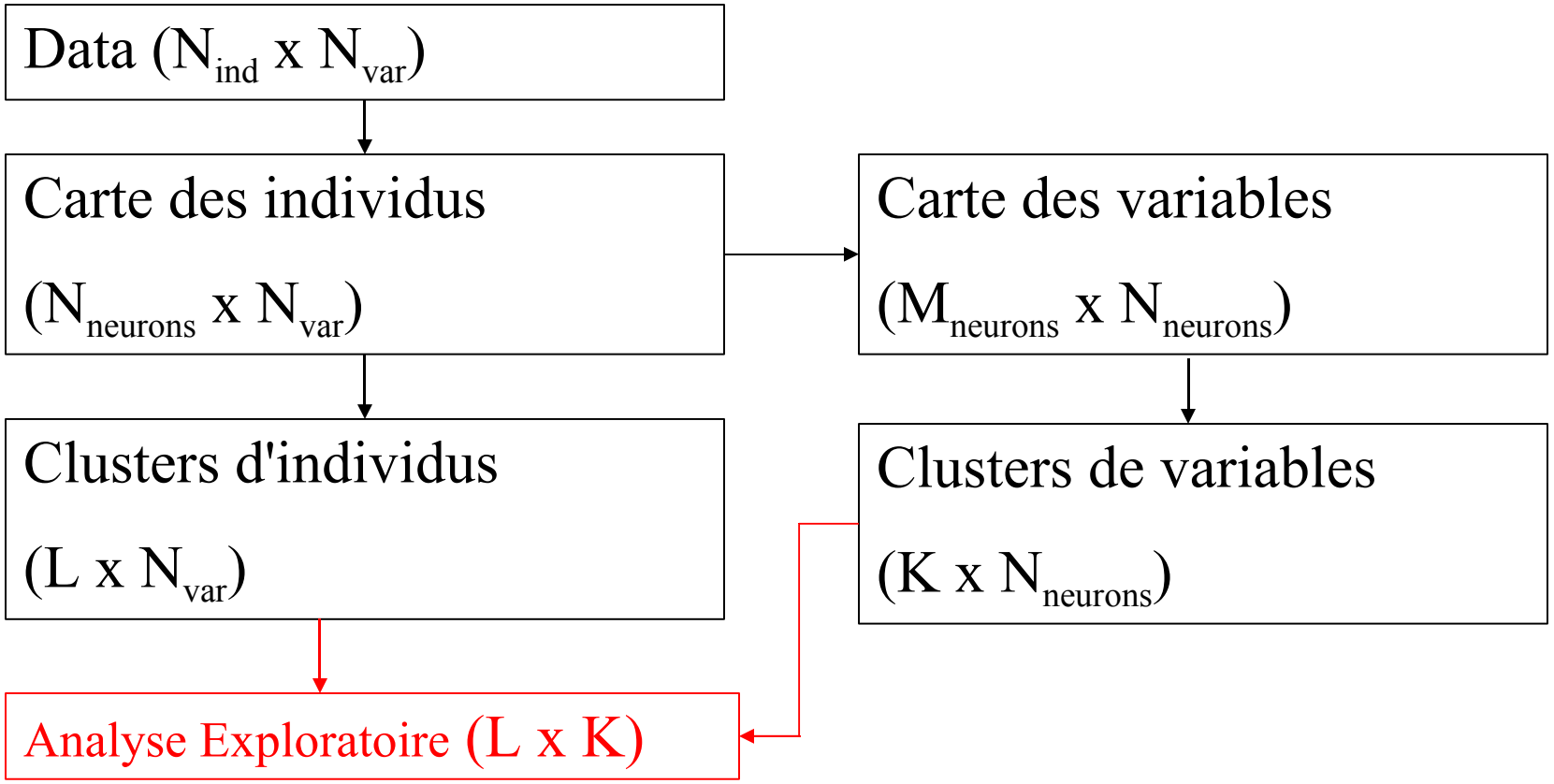




Synopsis

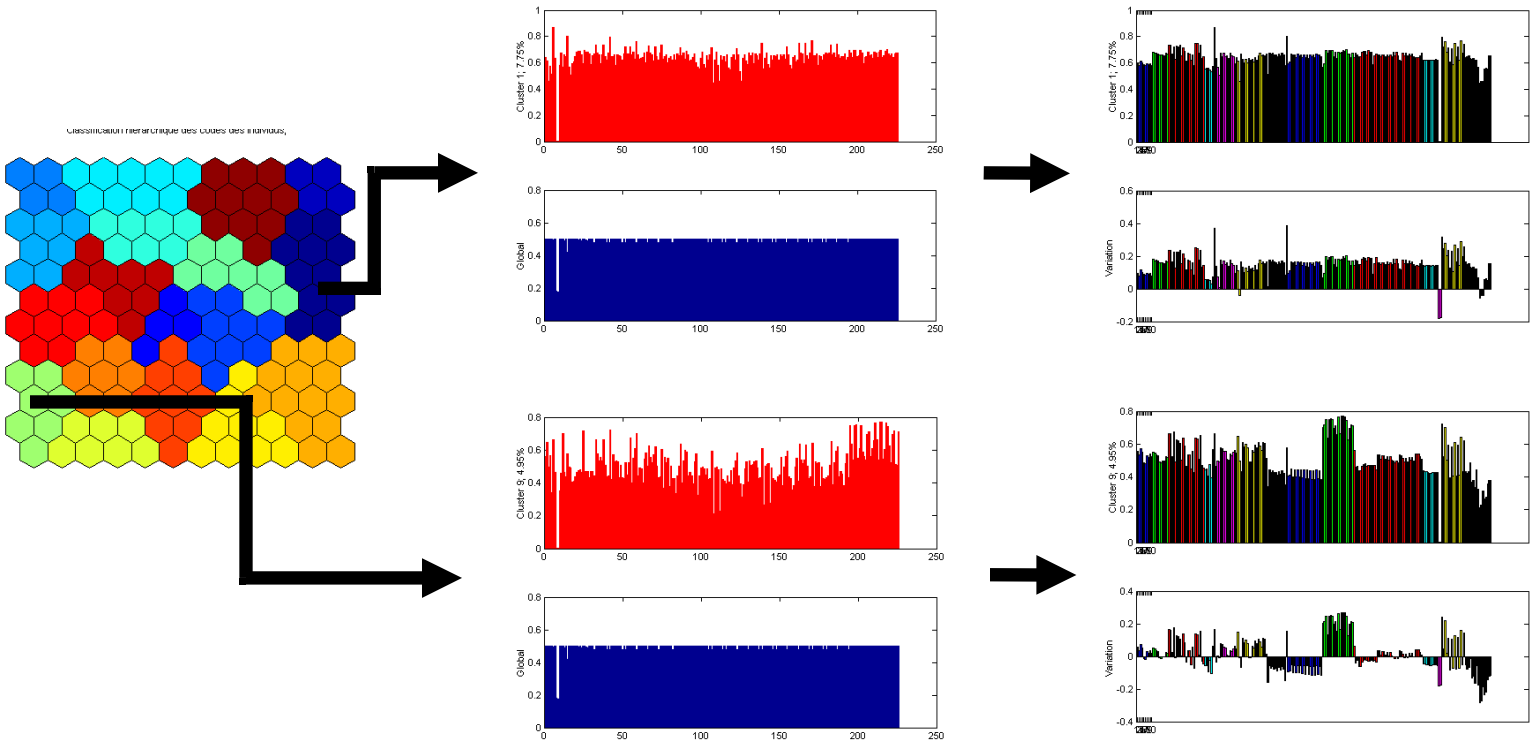

Data (N_{ind} x N_{var})
Carte des individus
($N_{neurons}$ x N_{var})
Clusters d'individus
(L x N_{var})
Carte des variables
($M_{neurons}$ x $N_{neurons}$)
Clusters de variables
(K x $N_{neurons}$)

Synopsis


Data (N_{ind} x N_{var})
Carte des individus
($N_{neurons}$ x N_{var})
Carte des variables
($M_{neurons}$ x $N_{neurons}$)
Clusters d'individus
(L x N_{var})
Clusters de variables
(K x $N_{neurons}$)
Analyse Exploratoire (L x K)

Utilisation de la carte des variables

Interprétation plus facile des clusters d'individus en réordonnant les variables selon leur cluster



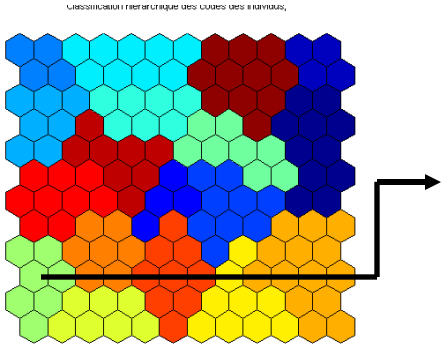
Analyse exploratoire

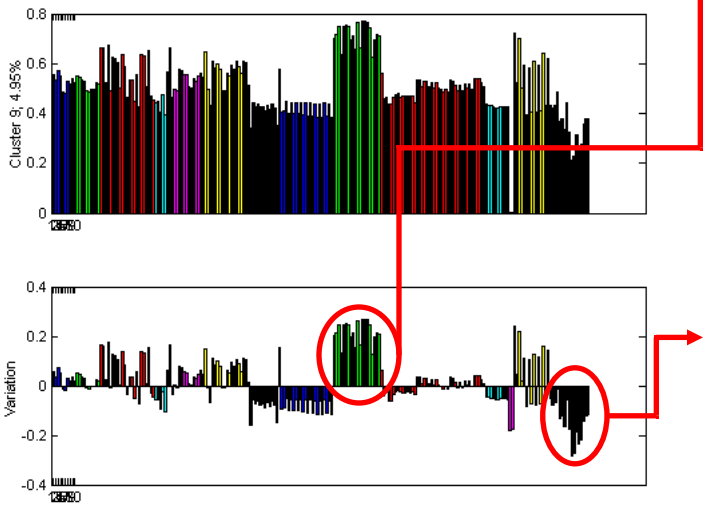
Double clustering :

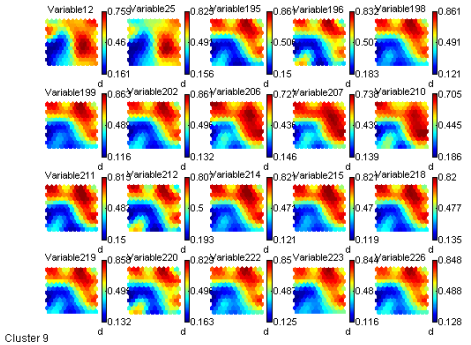
•des individus L clusters

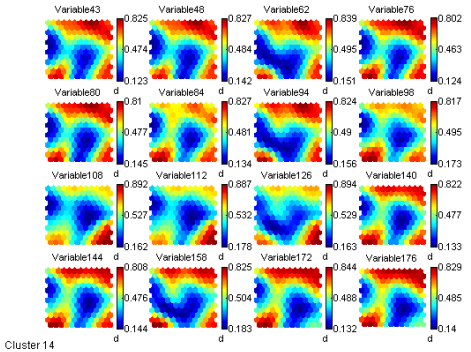
•des variables K clusters

les deux clustering sont cohérents simultanément





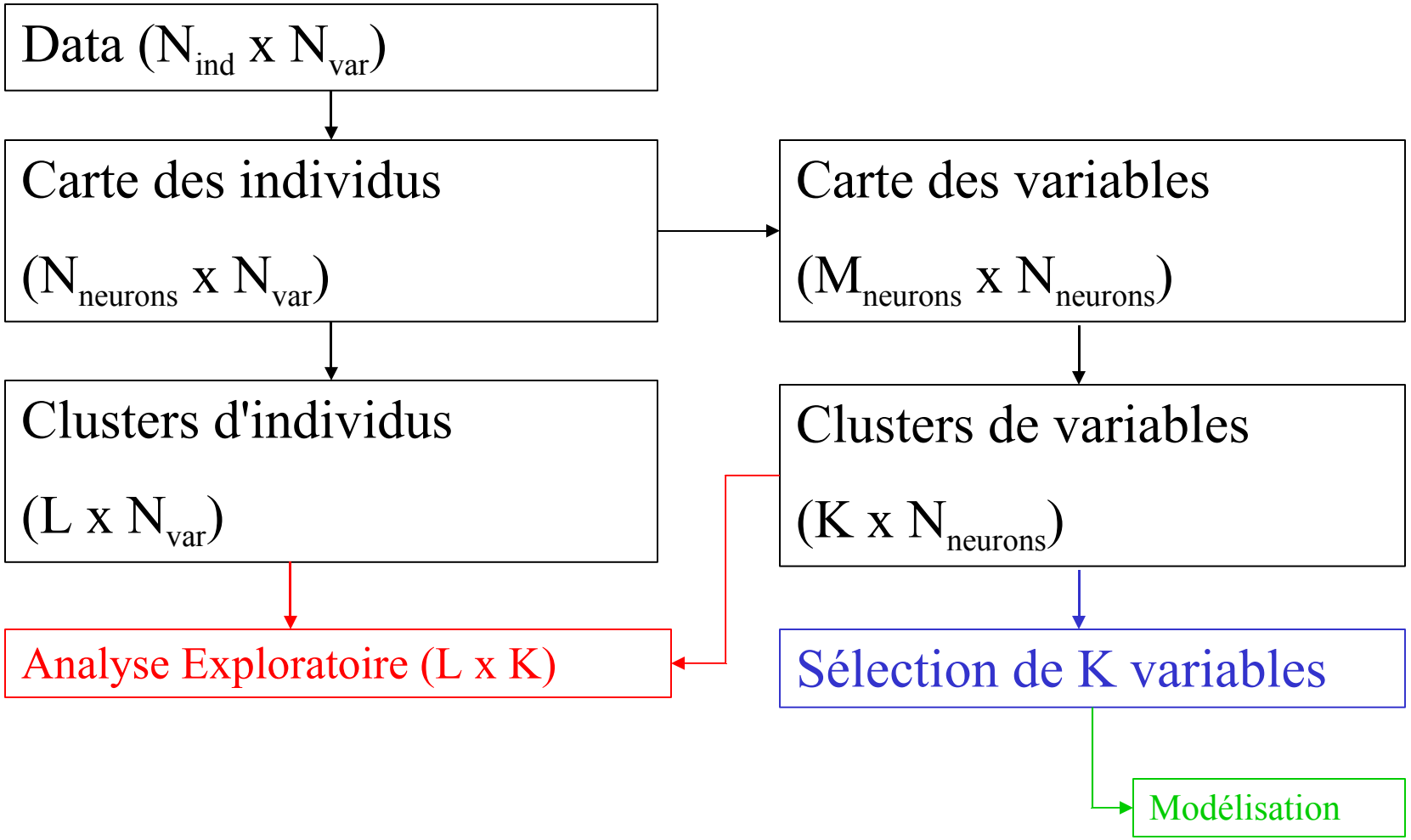




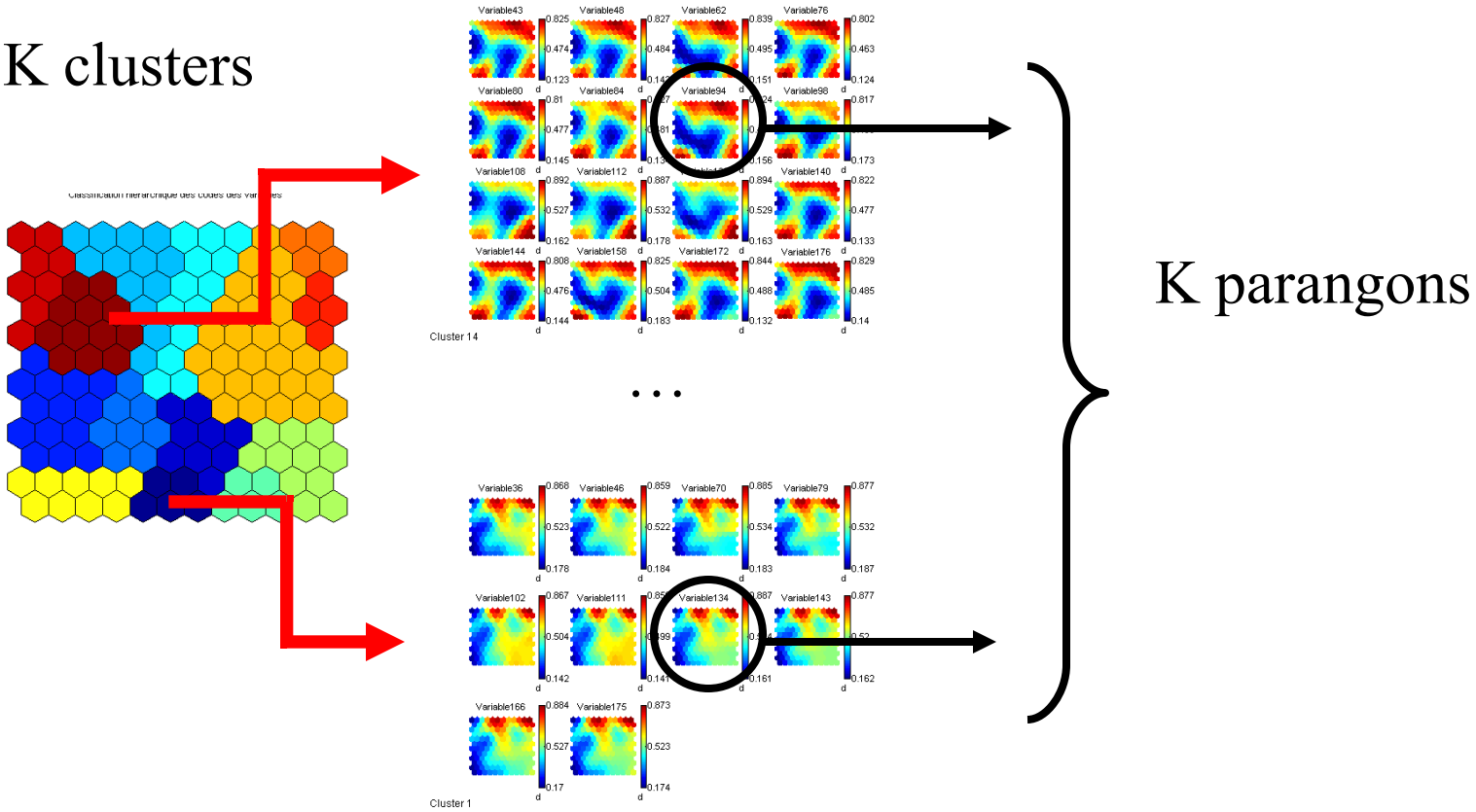
Plan

1. Généralités et algorithme
2. Visualisations des données et interprétation
3. Analyse exploratoire en 'grande' dimension
4. Sélection de variable, données manquantes…
5. Utilisation de distances adaptées

Sélection de variables à l'aide de Kohonen



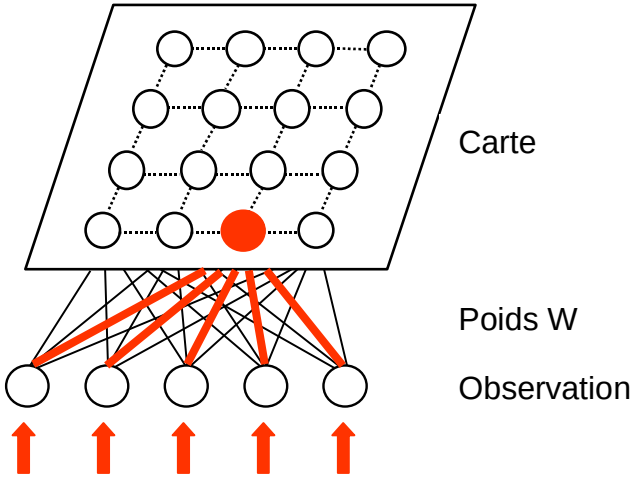
Sélection de variables



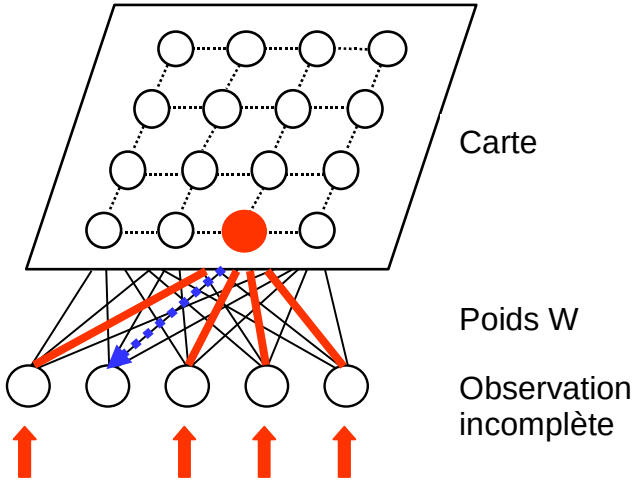
Traitement de données manquantes

- Les composantes correspondant à la variable manquante sont ignorées lors de la sélection du neurone gagnant (phases d'apprentissage et d'exploitation)
- En phase d'exploitation, la valeur associée à la variable manquante est la valeur du poids du neurone gagnant dans la dimension correspondante

- Classification traditionnelle



- Estimation d'une valeur manquante



Détection des données aberrantes

- Principe

  La carte auto-organisatrice est une méthode de quantification vectorielle

  L'erreur de quantification est un indicateur du caractère atypique d'une observation

- Mise en oeuvre

  Après apprentissage on calcule pour chaque neurone la moyenne des distances observation-neurone des observations classées par le neurone

  Puis, pour chaque observation que l'on veut vérifier, on compare la distance entre l'observation et le neurone le plus proche, avec la distance moyenne

  calcul du coefficient de représentativité *CR(X)*

$$CR(X) = \exp \frac{-d(X, W_{j^*})}{2 dist\_mean(j^*)}$$

  si *CR(x)* < *seuil*, alors l'observation est considérée comme erronée

Plan

1. Généralités et algorithme
2. Visualisations des données et interprétation
3. Analyse exploratoire en 'grande' dimension
4. Sélection de variable, données manquantes…
5. Utilisation de distances adaptées

Rappel : Apprentissage de la carte

- L'apprentissage met en correspondance l'espace des entrées et la carte

  Adaptation des poids *W* de telle manière que des exemples proches dans l'espace d'entrée sont associés au même neurone ou à des neurones proches dans la carte

- Algorithme

  Initialisation aléatoire des poids *W*

  A chaque itération t

  Présentation d'un exemple d'apprentissage *X(t),* choisi au hasard, à l'entrée de la carte

  Comparaison de l'exemple à tous les vecteurs poids, le neurone gagnant *j** est celui dont le vecteur poids $W_{j^*}(t)$ est le plus proche de l'entrée *X(t)* *(phase de compétition)*

$$d_N(X(t), W_{j^*}(t)) = \min_j d_N(X(t), W_j(t))$$

  Évaluation du voisinage du neurone gagnant dans la carte

$$h_{j*}(j,t) = h(d(j,j*),t)$$

  Mise à jour des poids pour tous les neurones de la carte, l'adaptation est d'autant plus forte que les neurones sont voisins de *j** *(phase de coopération)*

$$W_j(t+1) = W_j(t) + \Delta W_j(t)$$
$$\Delta W_j(t) = \varepsilon(t) \cdot h_{j*}(j,t) \cdot (X(t) - W_j(t))$$

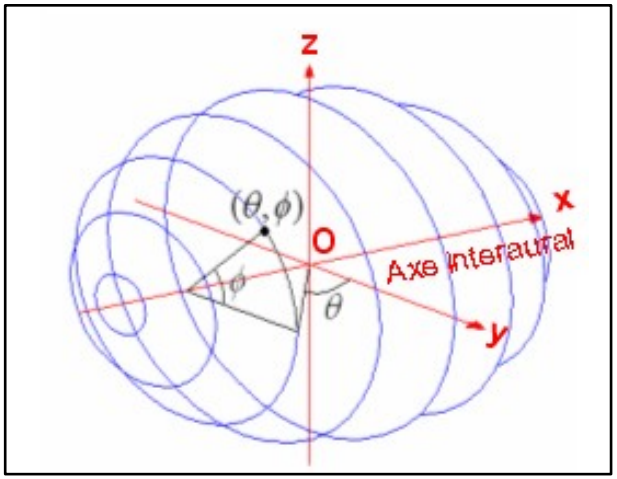
d_N distance dans l'espace d'entrée

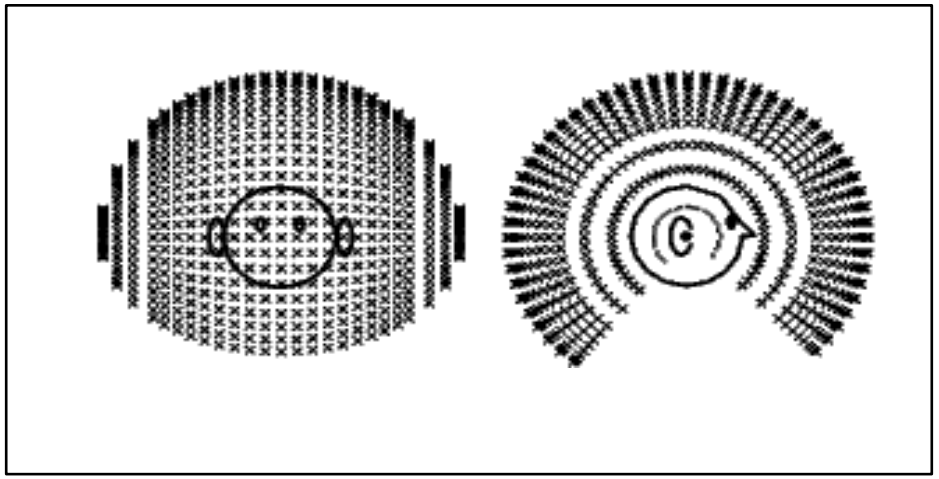
d distance dans la carte

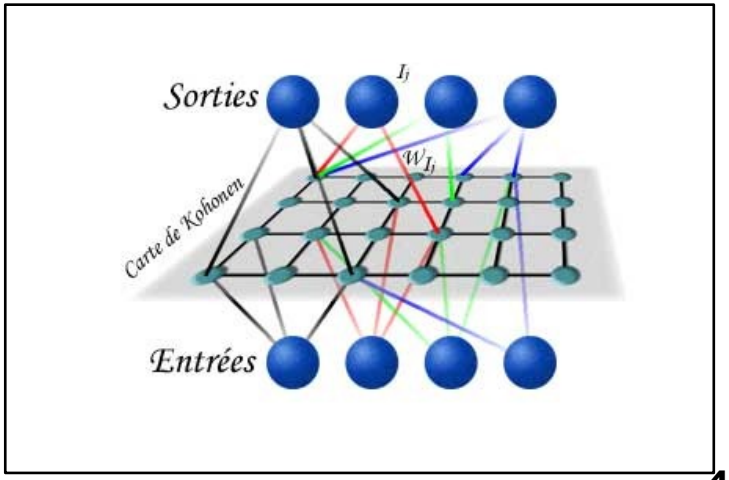
ε pas d'apprentissage

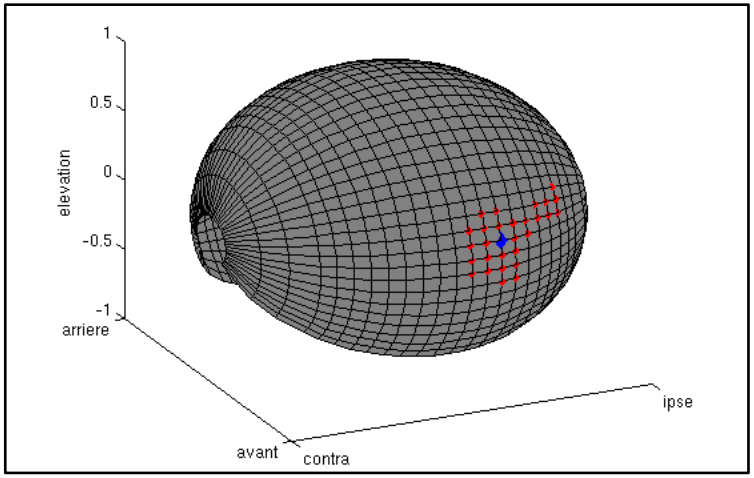
h fonction voisinage

Exemple : Cluster d'HRTF



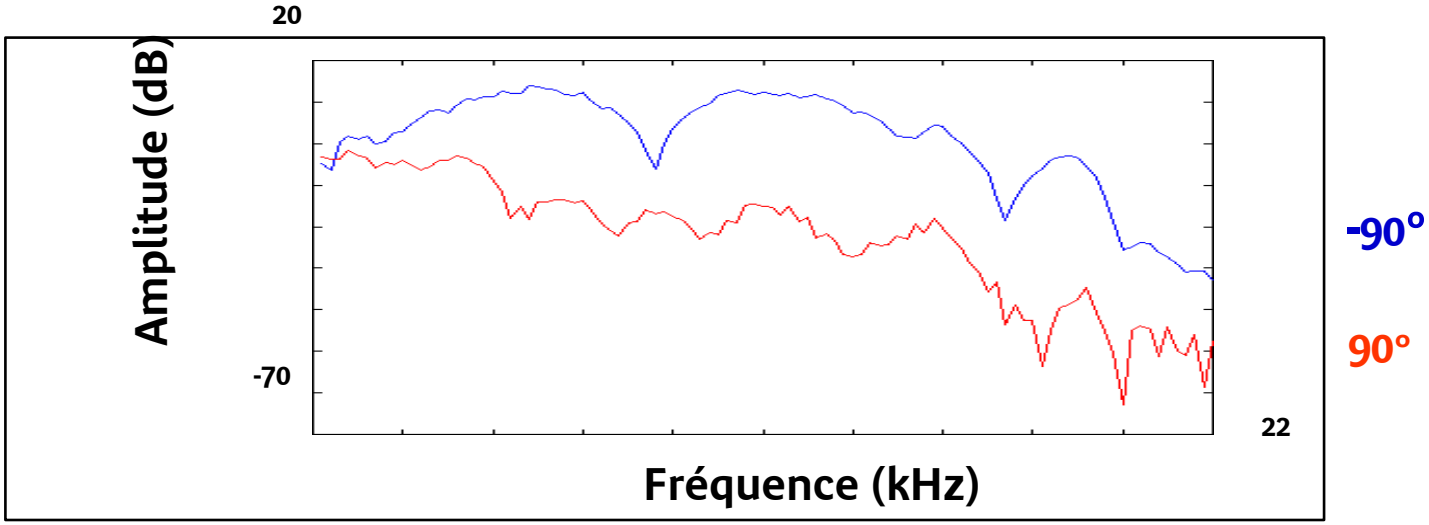




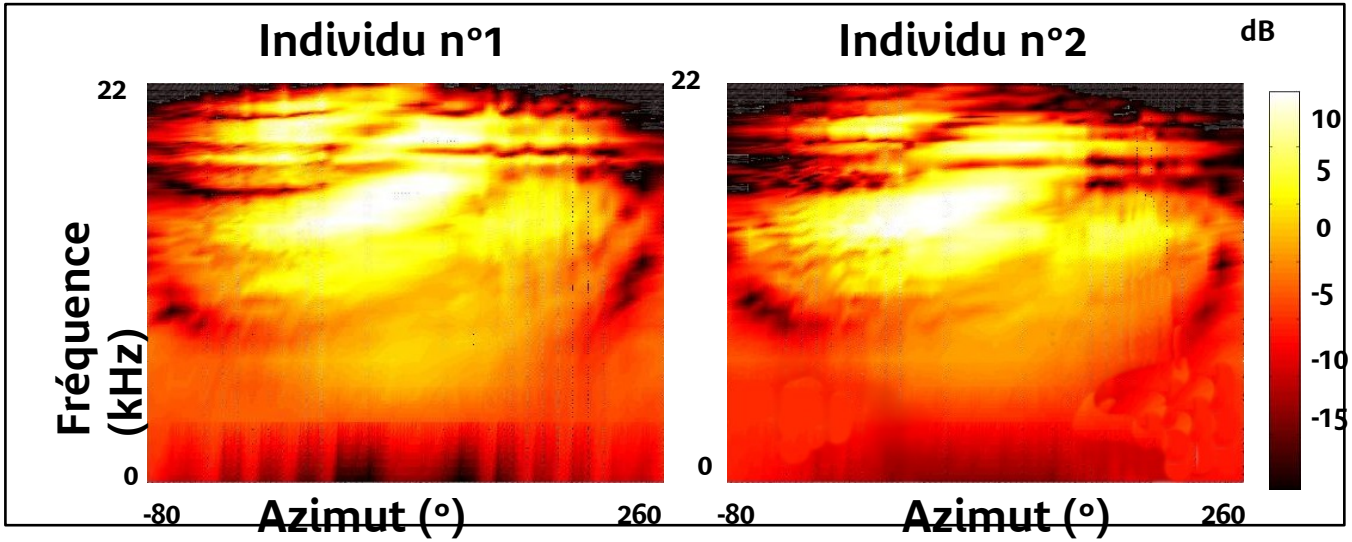


Exemple : Cluster d'HRTF

Même individu à différentes positions



Même positions pour deux individus



Distances adaptées potentielles

a. Critère Mean Squared Error (MSE) :

$$Crit_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left[H_1(i) - H_2(i)\right]^2$$

b. Critère Mean Squared Error (MSE) avec des poids sur les fréquences basés sur des filtres auditif (Bark ou ERB) :

$$Crit_{MSE\ Bark} = \frac{1}{N}\sum_{i=1}^{N}\left\{\alpha(i)\left[H_1(i) - H_2(i)\right]\right\}^2$$

c. Critère Fahn

$$Crit_{Fahn} = \frac{\sum_{i=1}^{N}\left[H_1(i) - H_2(i)\right]^2}{\sum_{i=1}^{N}\left[H_1(i)\right]^2}$$

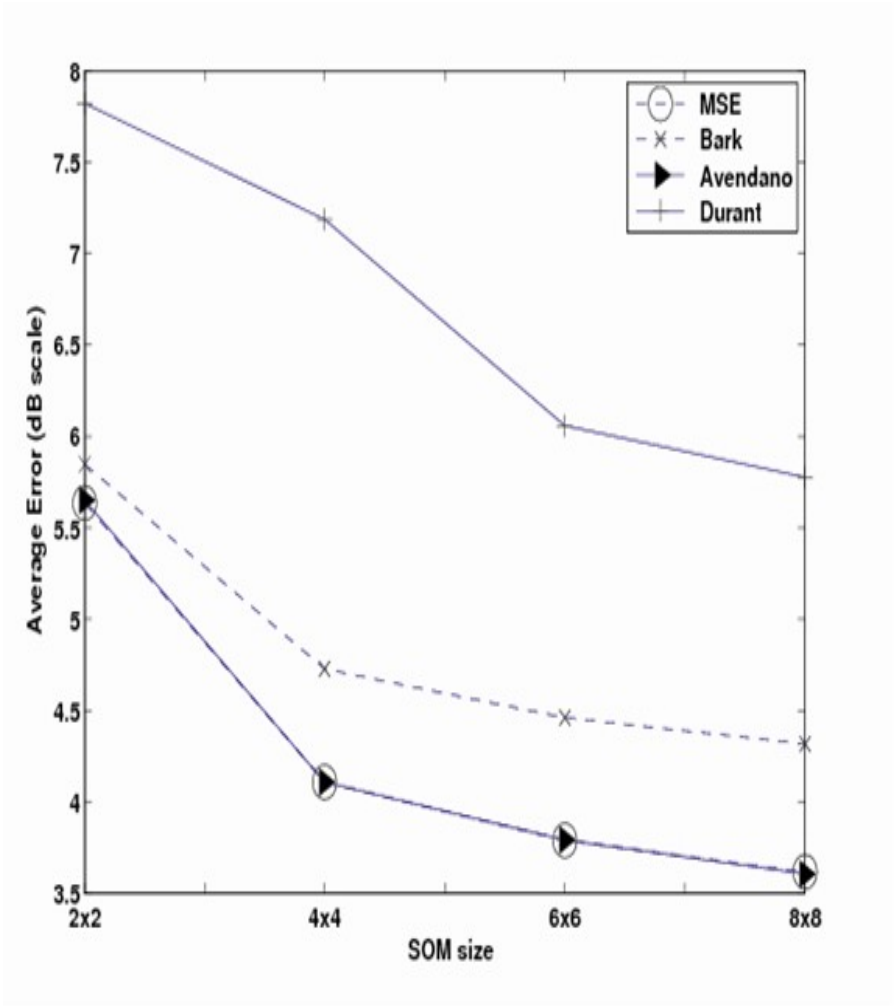
d. Critère Algazi

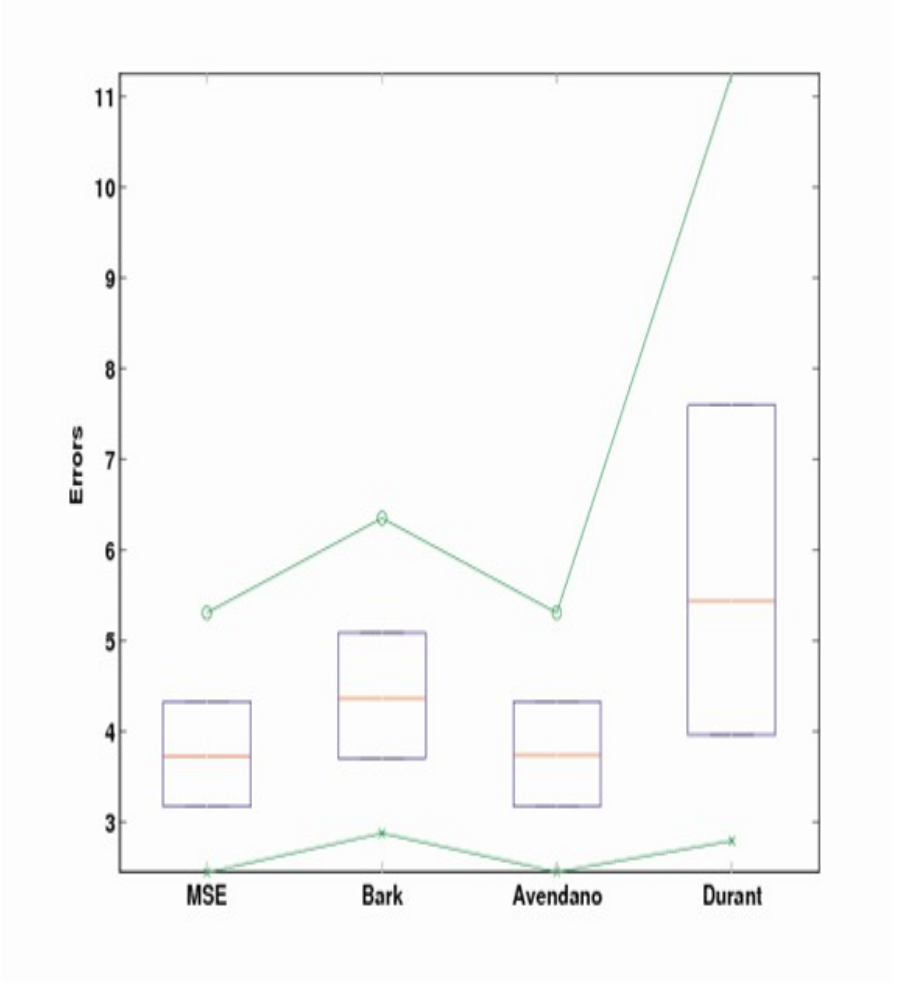
$$Crit_{Algazi} = 10\log 10\left\{\frac{\sum_{i=1}^{N}\left[H_1(i) - H_2(i)\right]^2}{\sum_{i=1}^{N}\left[H_1(i)\right]^2} + 1\right\}$$

e. Critère Durant

$$Crit_{Durant} = \frac{1}{N}\sum_{i=1}^{N}\left\{20\log 10\left[\frac{H_2(i)}{H_1(i)}\right] - \bar{d}\right\}^2, \quad \text{with :} \quad \bar{d} = \frac{1}{N}\sum_{i=1}^{N}20\log 10\left[\frac{H_2(i)}{H_1(i)}\right]$$

Distances : Résultats





Références

- Site de l'université d'Helsinki http://www.cis.hut.fr/research/som-research
- Site de l'université Paris 1 (SAMOS) http://www.univ-paris1.fr/
- T. Kohonen, Self-Organizing Maps, Springer, Heidelberg, 1995
- S. Midenet and A. Grumbach, Learning Associations by Self-Organization : the LASSO model, NeuroComputing, vol. 6, pp. 343-361, Elsevier, 1994
- S. Ibbou, Classification, analyse des correspondances et méthodes neuronales, Thèse de doctorat de l'université Paris 1, 1996
- G. Saporta, Probabilités, Analyse des Données et Statistique, Technip, 1990
- F. Blayo, M. Verleysen. Les réseaux de neurones artificiels, que sais je, PUF, 1996
- S. Thiria et al. Statistiques et méthodes neuronales, Dunod, 1997
- Ritter, H., T. Martinetz and K. Schulten Topology conserving maps for learning visuo motor coordination, Neural Networks, vol. 2, 159-168, 1989
- E. Oja and S. Kaski, Kohonen Maps, Elsevier, Amsterdam, 1999
- T. Samad and S. Harp, Self organization with partial data, Network, vol. 3, 205-212, 1992
- J. Vesanto, Neural network tool for data mining : SOM toolbox, TOOLNET, 184-196, 2000
- S. Kaski, Data exploration using self organizing maps, Acta Polytechnica Scandinavica, n°82, 1997

Des applications nombreuses, des domaines différents

- Plus de 4000 papiers de recherche recensés, de nombreuses applications industrielles
- Applications

  Analyse exploratoire de données, clustering, classification

  Quantification, détection de données erronées, sélection de variable

  Données manquantes, prédiction, diagnostic

- Domaines

  Télécommunication : analyse et détection de la fraude sur les cartes FT (Lemaire, FTR&D), diagnostic de l'état du réseau à partir de mesures de sondes (Fessant, Clérot, FTRD)

  Socio Economie : analyse du marché immobilier de Paris (Ibbou, U Paris I), segmentation du marché du travail (Gaubert, U Paris I), dépenses de formation en entreprise (Perraudin, U Paris I)

  TextMining : organisation d'une grande base de documents, le système WEBSOM (Kohonen, UTH), recherche de mots clés dans de grands textes (Kohonen, UTH)

  Processus industriels : optimisation du dosage d'un coagulant pour un problème de traitement de l'eau (Valentin, Suez)

  Energie : prédiction de la consommation électrique (Rousset, U Paris I)

  Télédétection : analyse de la couleur de l'océan à partir d'images satellites (Thiria, U de Versailles)