

# *Tableaux multiples Et Données Evolutives 1*

## **Méthode STATIS**

**N. Niang Keita**  
niang@cnam.fr

## SOMMAIRE

**Introduction**

**La méthode STATIS**

**Exemple: 1er tour scrutins élections présidentielles**

**La méthode STATIS DUALE**

**Application au contrôle des procédés à temps variable**

## Introduction

- Cadre général: méthodes d'analyse simultanée de plusieurs tableaux de données
- Diverses situations:
  - $p_t$  variables mesurées sur  $n_t$  individus à différentes dates: données évolutives
  - $p_t$  variables mesurées sur  $n_t$  individus à différentes occasions (non temporel)
  - Plusieurs tableaux de contingence ou de proximités

## Introduction

- Plusieurs méthodes:
  - STATIS, STATIS DUALE
  - AFM
  - DACP
  - + AUTRES ANGLO SAXONS

## Méthodes STATIS

- Structuration de Tableaux A Trois Indices de la Statistique (Escoufier et L'Hermier des plantes (1976) + C.Lavit)
- permet l'exploration simultanée de plusieurs tableaux de données quantitatives:
- $n$  individus et  $p_t$  var différentes → STATIS
  - Privilégie la position des individus
- $p$  var et  $n_t$  individus différents → STATIS DUALE
  - Privilégie les relations entre variables

## L'approche de la méthode:

- Etude classique d'un tableau: l'analyse factorielle associe au tableau un ensemble de représentation graphiques (cercle de corrélations, plans factoriels)
- Plusieurs tableaux: études séparées → trop de représentations indépendantes, pas pertinent; d'où:
- Recherche d'un ensemble unique de représentation, un résumé global, un compromis des tableaux

## Méthode STATIS

- Idée essentielle de la méthode:
- recherche d'une réponse à la question: **les distances entre individus sont elles stables d'un tableaux à un autre ?**
- Nécessité de trouver une structure commune aux études appelée **intrastructure**

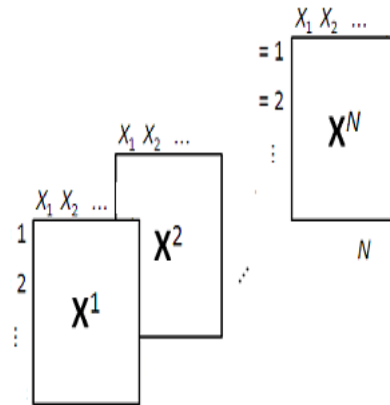
### 4 étapes:

- **Interstructure** : étude globale des différences entre tableaux
- **Compromis** : résumer les tableaux en un seul représentatif selon certains critères
- **Intrastructure** : étude fine des différences entre tableaux
- **Trajectoires** : évolution des individus ou variables suivant les tableaux

## Notations

- On dispose de N tableaux  $\mathbf{X}_t$  à **p** variables quantitatives décrivant les **mêmes n** individus

$$\mathbf{X}_t = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{pt1} \\ 2 & X_{12} & X_{22} & \dots & X_{pt2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & X_{1n} & X_{2n} & \dots & X_{ptn} \end{bmatrix}$$



## Rappels ACP

- ACP classique : triplet  $(\mathbf{X}, \mathbf{M}, \mathbf{D})$

– facteurs principaux :  $\mathbf{M}\mathbf{V}\mathbf{u} = \lambda\mathbf{u} = \textcolor{red}{(1)}$

– composantes principales :  $\mathbf{c} = \textcolor{blue}{\mathbf{X}\mathbf{u}}$

$$\mathbf{M}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}\textcolor{blue}{\mathbf{X}\mathbf{u}} = \lambda\textcolor{blue}{\mathbf{X}\mathbf{u}} \quad \mathbf{X}^* \textcolor{red}{(1)}$$

$$\mathbf{W}\mathbf{D}\mathbf{c} = \lambda\mathbf{c}$$

$\mathbf{W}$  =matrice des produits scalaires entre individus

**Composantes principales**  $\mathbf{W}\mathbf{D}\mathbf{c} = \lambda\mathbf{c}$

## Etude de l'interstructure : analyse globale

- ACP particulière: 3 phases
- Définir un objet représentatif
- Définir une métrique pour distances entre objets
- Trouver une image des objets représentatifs

## Interstructure : objet représentatif $W_t$

- Individu de l'ACP = objet représentatif d'un tableau  $X_t$
- $W_t = X_t M_t X_t'$  caractérise  $(X_t, M_t, D)$

$W_t$  contient les produits scalaires entre individus = tous les liens inter individus

## Interstructure objet représentatif $W_t$

- On représente chaque étude  $(X_t, M_t, D)$   $t=1, \dots, N$  par  $W_t$

- Même nombre d'individus

$D$  = Matrice de poids des individus

$$X_t = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 2 & X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \quad \text{par} \quad W_t = X_t M_t X_t'$$

- On associe un poids  $\pi_t$  à chaque  $W_t$

$$\Delta = \begin{bmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_N \end{bmatrix}$$

## Interstructure : métrique pour distances entre objets

- Metrique de l'ACP:

définir une distance entre objets

Produit scalaire de Hilbert Schmidt:

$$S_{tt'} = \langle W_t | W_{t'} \rangle_{HS} = \text{Tr} (DW_t DW_{t'})$$

Si les objets  $W_t$  ont des normes très différentes:

$$\langle W_t | W_{t'} \rangle_{HS} = \text{Tr} (DW_t DW_{t'}) / \sqrt{\text{Tr}[(DW_t)^2] \text{Tr}[(DW_{t'})^2]}$$

## Interstructure **métrique pour distances entre objets**

On reconnaît le **coefficient RV** (cf cours de Gonzalez) :

On définit la matrice S qui contient les coefficients RV

On associe à chaque tableau un poids:  $\Delta = \begin{bmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_N \end{bmatrix}$

**Comme en ACP on peut donc construire une image représentative des objets**

**Rque: S a tous ses termes positifs, on aura un facteur taille**

## Interstructure **image représentative des objets**

- Rappel : STATIS = ACP particulière :
  - **Individu** =  $W_i$  = objet représentatif d'un tableau
  - **Tableau de données** = S = Matrice des coefficients RV (produits scalaires entre objets  $W_i$ ) et poids
  - **Composante principale** = vecteur propre de  $S\Delta$



## Interstructure image représentative des objets

- **ACP de  $\mathbf{S}\Delta$**

Les vecteurs propres de  $\mathbf{S}\Delta$  associés aux deux plus grandes valeurs propres permettent la représentation des objets  $\mathbf{W}_1, \dots, \mathbf{W}_N$  sur le 1er plan factoriel :

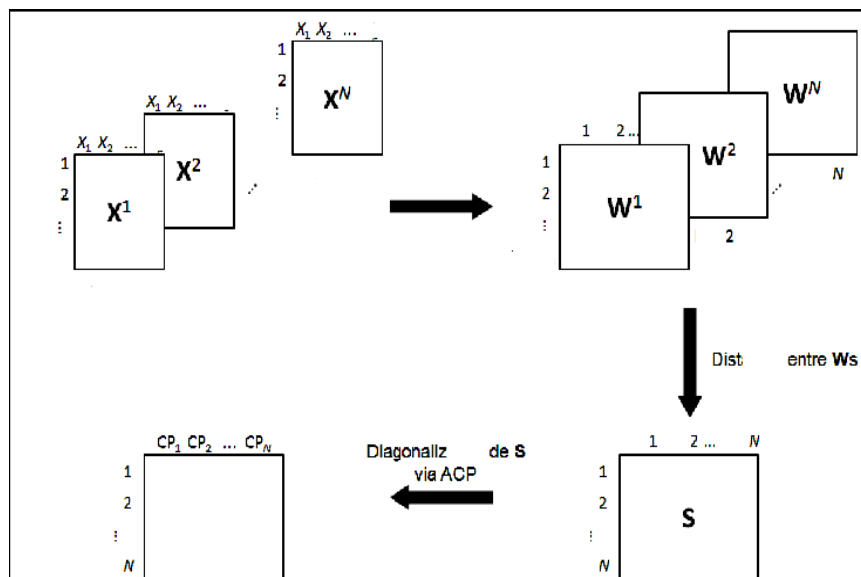
- Les coordonnées des  $\mathbf{W}_t$  sur l'axe  $i$  sont contenues dans  $\mathbf{c}_{t,i}$ :

$$\mathbf{c}_{t,i} = \sqrt{\lambda_i} \mathbf{u}_i$$

$\lambda_i$  :  $i^{\text{ème}}$  valeur propre de  $\mathbf{S}\Delta$

$\mathbf{u}_i$  :  $i^{\text{ème}}$  vecteur propre de  $\mathbf{S}\Delta$

## STATIS - Résumé Interstructure



Extrait L'analyse des données multivariées  
méthodes et applications. GEA I. Techni

Exemple 1:  
1er tour  
présidentielle

Scrutin 1969

dep	GI1	DI1	GP1	AT1	NP1	DI2	CN2	NP2
01								
95								

Scrutin 1974

dep	GI1	DI1	GP1	DP1	AT1	NP1	DI2	GI2	NP2
01									
95									

Scrutin 1981

dep	GI1	DI1	GP1	AT1	NP1	DI2	GI2	NP2
01								
95								

Scrutin 1988

dep	GI1	DI1	GP1	DP1	AT1	NP1	DI2	GI2	NP2
01									
95									

Fig. 3.2. Les tableaux de données étudiés.

En 1969 :

- GI1 : Vote Defferre
- DI1 : Vote Poher + Vote Pompidou
- GP1 : Vote Rocard + Vote Duclos + Vote Krivine
- AT1 : Vote Ducatel
- NP1 : Votes blancs et nuls + Abstention
- CN2 : Vote Poher
- DI2 : Vote Pompidou
- NP2 : Votes blancs et nuls + Abstention

(gauche institutionnelle)  
(droite institutionnelle)  
(gauche protestataire)  
(centre)  
(non participation réelle)  
(centre)

En 1974 :

- GI1 : Vote Mitterrand
- DI1 : Vote Giscard + Vote Muller + Vote Royer + Vote Chaban
- GP1 : Vote Laguille + Vote Krivine
- DP1 : Vote Le Pen + Vote Renouvin
- AT1 : Vote Héraut + Vote Sebag + Vote Dumont
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- DI2 : Vote Giscard d'Estaing
- NP2 : Votes blancs et nuls + Abstention

(droite protestataire)

En 1981 :

- GI1 : Vote Mitterrand + Vote Crépeau
- DI1 : Vote Giscard + Vote Chirac + Vote Debré + Vote Giscard
- GP1 : Vote Laguille + Vote Marchais + Vote Bouchardau
- AT1 : Vote Lalonde
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- DI2 : Vote Giscard d'Estaing
- NP2 : Votes blancs et nuls + Abstention

En 1988 :

- GI1 : Vote Mitterrand
- DI1 : Vote Barre + Vote Chirac
- GP1 : Vote Lajoie + Vote Juquin + Vote Laguille + Vote Bousset
- DP1 : Vote Le Pen
- AT1 : Vote Waechter
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- DI2 : Vote Chirac
- NP2 : Votes blancs et nuls + Abstention

mêmes individus = départements.  
différentes variables : différents ans de données.  
Statistiques élémentaires sur les tableaux de données

Scrutin 1969

(en %)	GI1	DI1	GP1	AT1	NP1	DI2	CN2	NP2
Moyenne	3,74	52,77	19,48	0,98	23,04	37,79	28,33	33,88
Ecart-type	1,07	5,82	3,18	0,22	1,68	5,35	4,37	5,36

Scrutin 1974

(en %)	GI1	DI1	GP1	DP1	AT1	NP1	DI2	GI2	NP2
Moyenne	36,33	43,41	2,46	0,78	0,98	15,81	44,19	42,86	12,95
Ecart-type	5,55	5,58	0,49	0,18	0,36	1,68	5,83	6,13	1,42

Scrutin 1981

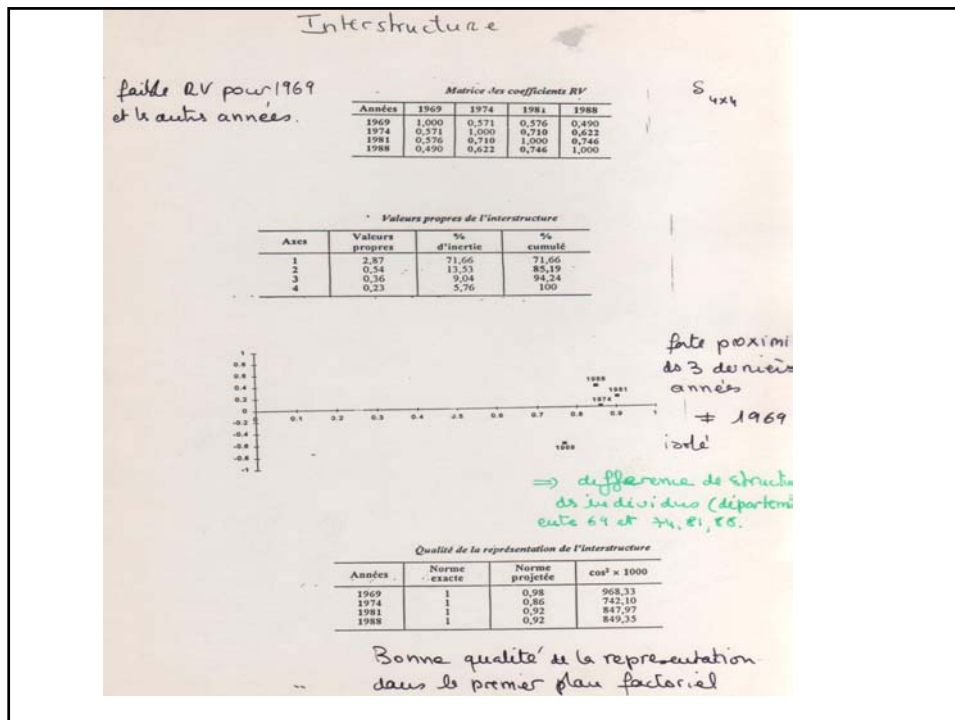
(en %)	GI1	DI1	GP1	AT1	NP1	DI2	GI2	NP2
Moyenne	23,12	39,61	14,84	1,21	19,36	40,35	44,10	15,54
Ecart-type	2,92	4,75	3,94	0,55	2,01	4,84	4,94	1,61

Scrutin 1988

(en %)	GI1	DI1	GP1	DP1	AT1	NP1	DI2	GI2	NP2
Moyenne	27,70	29,66	9,16	11,09	3,12	19,26	37,78	44,51	17,71
Ecart-type	3,42	4,05	2,61	3,30	0,82	1,65	3,94	4,10	1,96

$D = \frac{1}{95} I$

Importante différence de dispersion  $\Rightarrow$  on centre et réduit chaque tableau.  $\Pi_t = \frac{1}{4} \quad \forall t = 1, 4$



## Compromis-Intrastructure

**Interstructure = analyse globale**

mise en évidence de ressemblances ou différences globales entre tableaux sans les expliquer.

Objectif de la suite = analyse plus fine pour expliquer

Deux étapes:

- \* recherche d'un point de repère = **compromis**
- \* étude de **l'intrastructure**

## Recherche d'un compromis

Compromis = bon résumé, de même nature que les objets

solution:  $W_{co}$  moyenne pondérée des  $W_t$

Les coefficients étant tels que  $W_{co}$  soit le plus corrélé avec les  $W_t$

## Recherche d'un compromis

Matrice Compromis  $W_{CO}$

$$W_{CO} = \sum_{t=1}^N \alpha_t W_t$$

$\alpha_t = \frac{1}{\sqrt{\lambda_1}} \pi_t u_1^{(t)}$

$W_t = \mathbf{X}_t \mathbf{M} \mathbf{X}_t'$

$t^{\text{ème}}$  élément du 1<sup>er</sup> vecteur propre de  $\mathbf{S}\Delta$

1<sup>ère</sup> valeur propre de  $\mathbf{S}\Delta$

Matrice de données initiales

## Recherche d'un compromis

### Représentation et Interprétation

- 4 cas: \*
- \*  $W_t$  ont des normes voisines et des grands RV: existence d'une structure commune bien décrite par le compromis
  - \* un  $W_t$  différent des autres: il intervient peu dans la construction du compromis, robuste
  - \*  $W_t$  normes trop différentes: il faut normer
  - \*  $W_t$  très différents, RV faibles: pas de structure commune

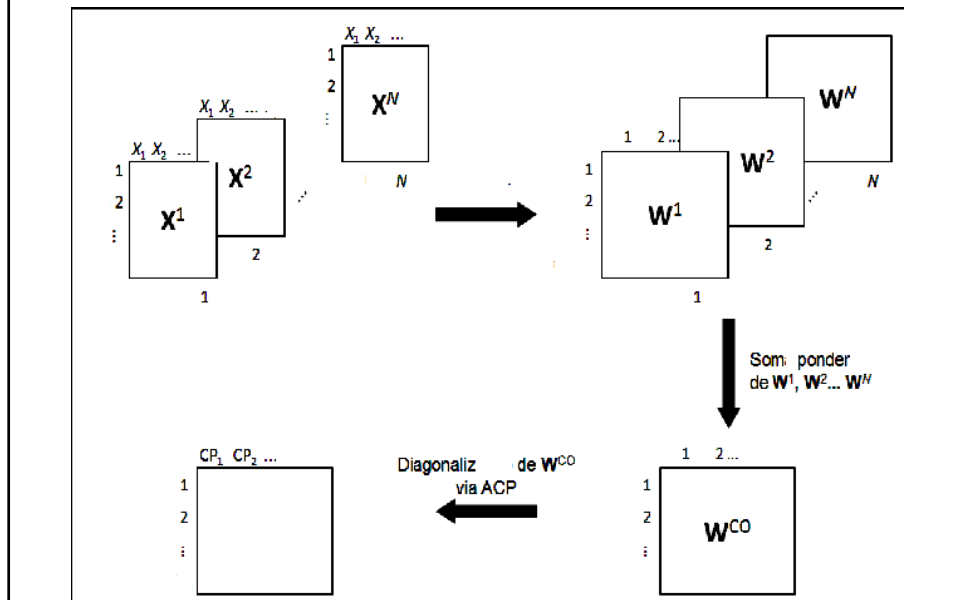
## Intrastructure image représentative des points compromis

- ACP de  $W_{co}D$
- Les vecteurs propres de  $W_{co}D$  associés aux deux plus grandes valeurs propres fournissent une image euclidienne des points compromis
  - Les coordonnées compromises sur l'axe  $i$  sont contenues dans le vecteur  $c_{co,i}$ :

$$c_{co,i} = \sqrt{\delta_i} v_i = \frac{1}{\sqrt{\delta_i}} W_{co} D v_i$$

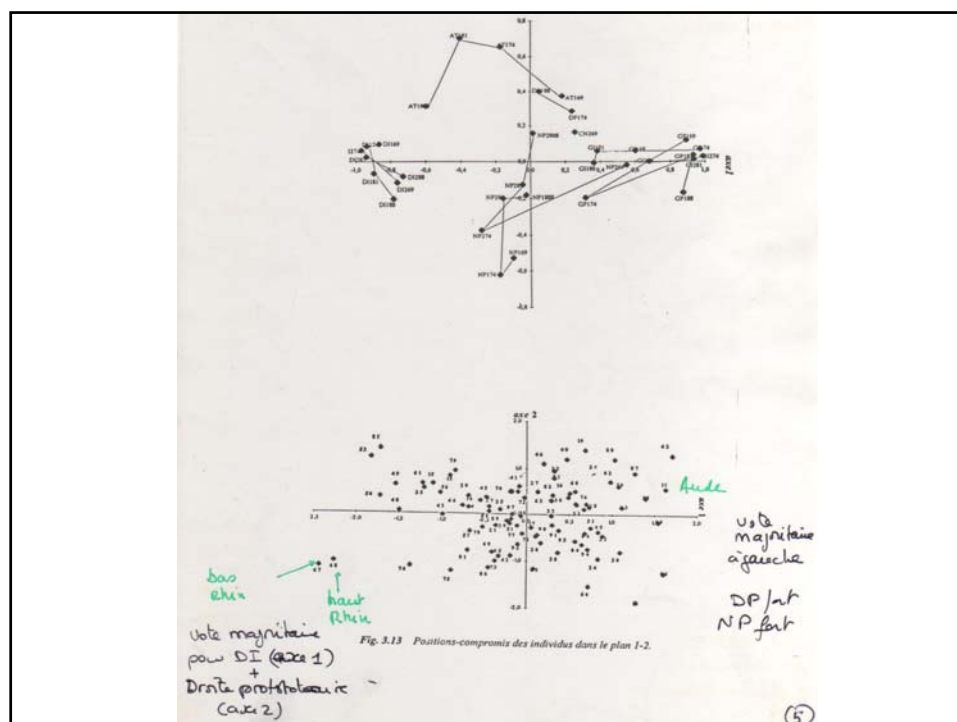
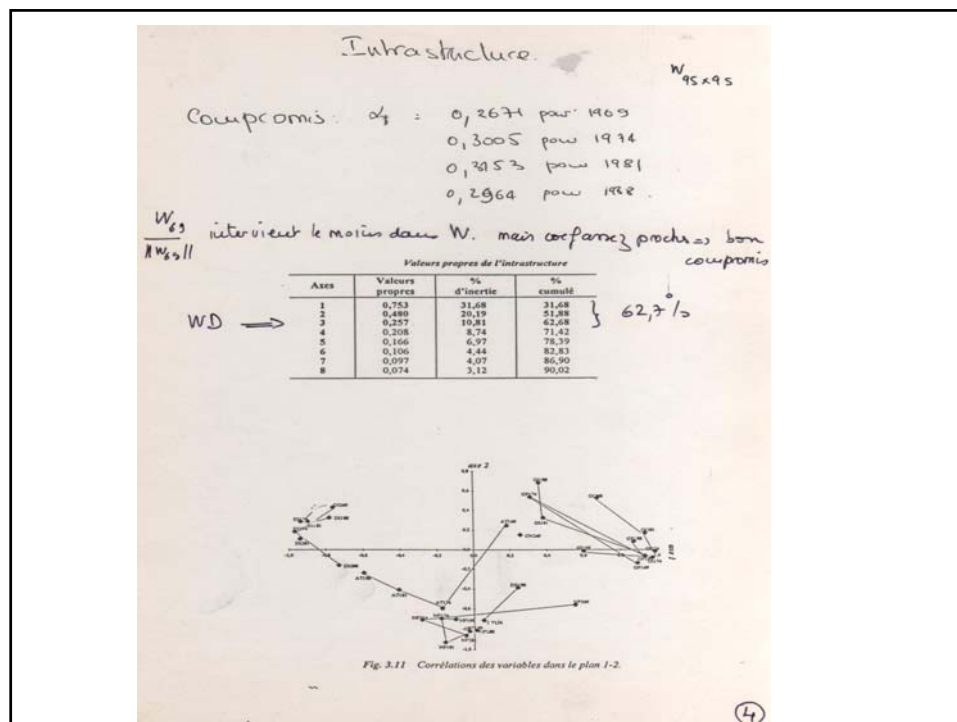
$\delta_i$  1<sup>ère</sup> valeur propre de  $W_{co}D$        $v_i$  1<sup>ère</sup> vecteur propre de  $W_{co}D$

## STATIS - Intrastructure



### Intrastructure : corrélation variables-compromis

- Chaque composante principale du compromis est un **vecteur à n dimension**, n étant le nombre d'individus des tableaux initiaux.
- Calcul de leurs **corrélations avec les variables initiales** des tableaux  $X_t$
- Représentation des **cercles de corrélation**
- **Interprétation des axes pour expliquer les positions compromis des individus**



## STATIS - Trajectoires des individus

- Méthode: technique des points supplémentaires:

Représentation sur l'image euclidienne compromis des  $N$  nuages des individus

$$\mathbf{c}_{CO\ t,i} = \frac{1}{\sqrt{\delta_i}} \mathbf{W}_t \mathbf{D} \mathbf{v}_i$$

- Remarque :  $\mathbf{c}_{CO,i}$  est le “centre de gravité” des points  $\mathbf{c}_{CO\ t,i}$

## STATIS - Trajectoires des individus

- Objectif :

- Mise en évidence des écarts entre les  $\mathbf{W}_t$  et avec le compromis au niveau individuel
- Détection des individus responsables des écarts entre tableaux

- Définition :

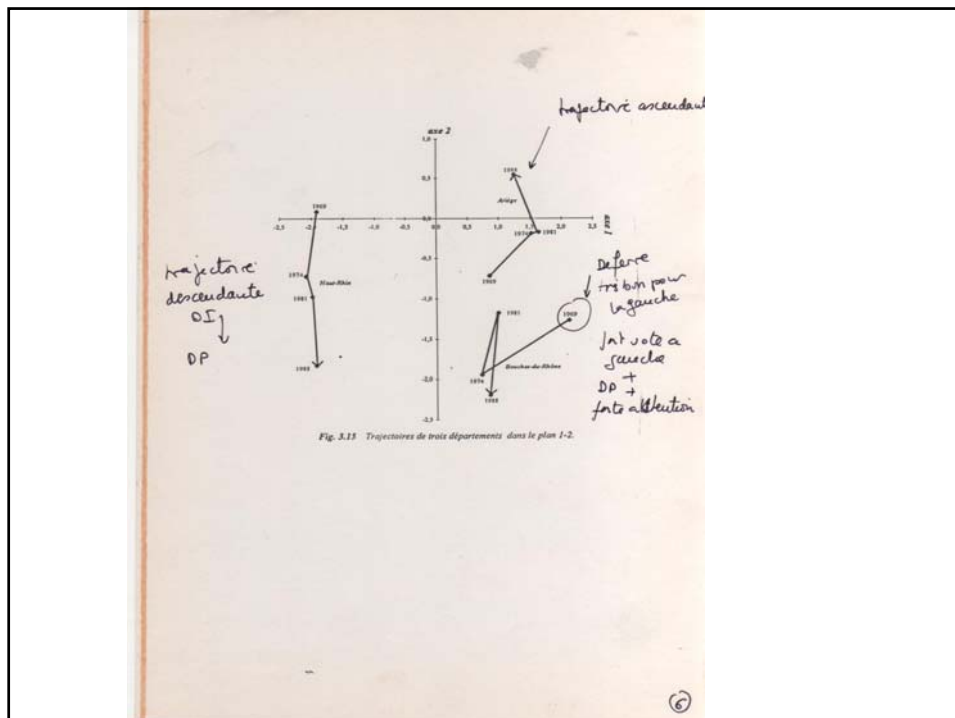
- On place les différentes positions d'un individu tel qu'il est décrit par chaque étude. Ces différentes positions définissent sa trajectoire



## Interprétation des trajectoires

- Deux classes de trajectoires des individus :
  - **Peu étendue autour de sa position compromis** = individus dont l'évolution suit l'évolution moyenne, écart par rapport à la moyenne régulier d'un tableau à l'autre
  - **Trajectoire de grande amplitude** = changement de structure suivant les tableaux, différence avec l'évolution moyenne

Rq: si les axes du compromis sont bien corrélés avec les variables on pourra expliquer les axes par les variables et on interprètera de façon détaillée les trajectoires des individus



## STATIS DUALE

- Même démarche que STATIS en prenant comme matrice représentant un tableau,  $V_t$  la matrice de covariance des variables à la place des  $W_t$ .
- On privilégie les variables au lieu des individus.
  - **Interstructure**: étude globale des variables
  - **Compromis**: moyenne pondérée des  $V_t$
  - **Intrastructure**: compromis des variables
  - **Trajectoires** de variables

## STATIS DUALE

Exemple d'application:

Contrôle de procédés par lots à temps variable