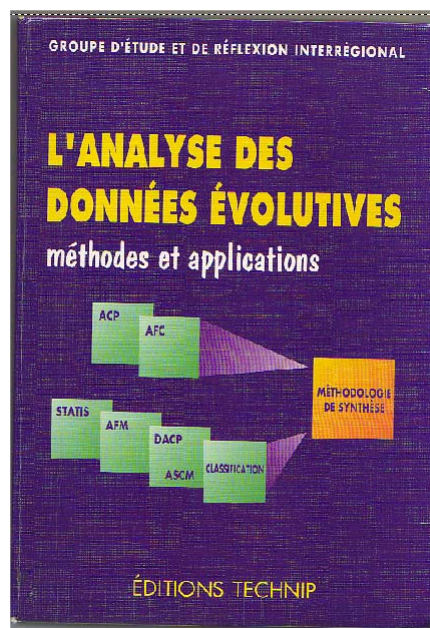


# Tableaux multiples et données évolutives

**N. Niang Keita** [niang@cnam.fr](mailto:niang@cnam.fr)  
**Gilbert Saporta** [gilbert.saporta@cnam.fr](mailto:gilbert.saporta@cnam.fr)

novembre 2011

1



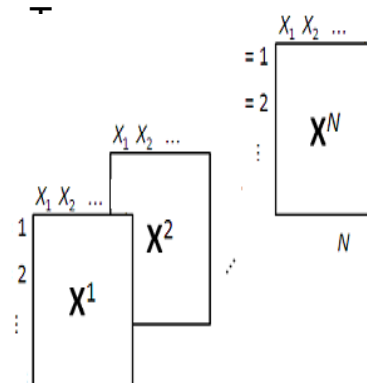
2

# Plan

1. Généralités
2. Double ACP
3. STATIS
4. AFM

3

- 1 Généralités
- Suite de tableaux individusxvariablesxtemps
  - temps discret  $1, 2, \dots, t$ ,
  - « données de panels »
  - données « cubiques »



4

- 1 Généralités
- **Etude classique d'un tableau**: l'ACP 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

5

## 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

6

## 2. La double ACP (Bouroche, J.M., 1975)

- ACP n°1: nuage des centres de gravité  $g_t$  comme étude de l'**interstructure**
- Recherche de l'**intrastructure**
  - ACP de chaque tableau: T systèmes d'axes
  - Système d'axes **compromis** maximise l'inertie de la somme des projections de chaque nuage = somme des inerties après centrage

7

## double ACP (suite)

- Axes **compromis** : vecteurs propres de  $\sum_{t=1}^T \mathbf{V}_t$
- ACP n°2: ACP de la concaténation verticale (superposition) des  $X_t$  centrés

$X_1$
$X_2$
$X_T$

8

## double ACP: exemple

- Crimes et délits en 9 catégories pour l'ensemble des départements métropolitains, par année de 1974 à 1993
  - VO : vols et recels
  - FX : faux et escroqueries
  - DF: délits financiers
  - CH: chèques sans provisions
  - CR: coups, règlements de comptes, traumatismes
  - ST: stupéfiants
  - DD: destructions et dégradations
  - ET : délits à la police des étrangers
  - DV : divers

9

Tableau 5.1 Centres de gravités des tableaux analysés<sup>1</sup>

Années	VO	FX	DF	CH	CR	ST	DD	ET
1974	15,46	1,31	0,27	3,31	1,41	0,040	1,06	0,088
1975	16,84	1,51	0,24	2,64	1,45	0,056	1,27	0,092
1976	16,44	1,63	0,27	2,11	1,44	0,068	1,25	0,076
1977	18,14	1,98	0,31	3,39	1,56	0,090	1,55	0,093
1978	18,26	2,13	0,37	3,88	1,54	0,123	1,70	0,113
1979	19,90	2,35	0,40	4,34	1,64	0,185	1,98	0,134
1980	21,89	2,67	0,44	5,90	1,74	0,201	2,41	0,179
1981	23,56	3,08	0,45	7,00	1,81	0,275	2,68	0,203
1982	27,20	3,88	0,73	8,33	1,86	0,388	3,18	0,158
1983	28,41	4,32	0,70	8,41	1,93	0,371	3,24	0,346
1984	30,18	4,53	0,95	7,64	1,95	0,401	3,39	0,341
1985	30,65	4,68	0,84	6,46	1,97	0,499	3,26	0,369
1986	28,72	4,74	0,66	4,89	1,78	0,727	2,98	0,375
1987	27,64	4,80	0,45	4,42	1,72	0,699	3,02	0,364
1988	27,18	4,54	0,50	3,96	1,88	0,666	2,98	0,365
1989	28,34	4,84	0,35	3,90	1,99	0,682	3,26	0,414
1990	30,99	5,17	0,34	3,58	2,03	0,741	3,77	0,366
1991	32,98	5,64	0,43	3,01	2,14	0,828	4,29	0,449
1992	34,76	5,14	0,38	0,35	2,21	0,955	4,99	0,501
1993	35,13	5,29	0,500	0,056	2,32	0,976	5,39	0,511

10

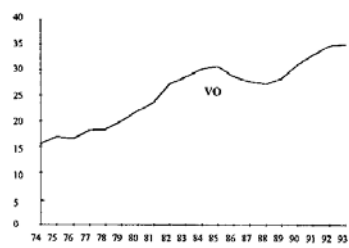


Fig. 5.8 Évolution de la variable VO (voles et recels) en France métropolitaine.

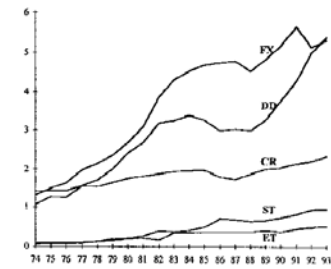


Fig. 5.9 Évolution des variables ST, ET, FX, CR et DD en France métropolitaine.

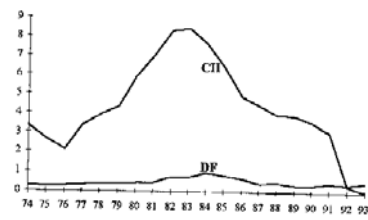


Fig. 5.7 Évolution des variables CH et DF en France métropolitaine (criminalité financière).

11

Tableau 5.2 Valeurs propres de l'interstructure déterminée par la DACP

Axes	Valeurs propres	% d'inertie	% cumulé
1	5,87	73,42	73,42
2	1,70	21,27	94,70
3	0,19	2,31	97,01
4	0,17	2,10	99,11

12

Tableau 5.3 Aides à l'interprétation de l'ACP des centres de gravité

Années	Axe 1				Axe 2			
	Coord	Ctr	Cos <sup>2</sup>	Qlt	Coord	Ctr	Cos <sup>2</sup>	Qlt
74	-3,85	12,6	96,3	96,3	-0,7	1,5	3,2	99,5
75	-3,58	10,9	92	92	-1,03	3,1	7,6	99,6
76	-3,58	10,9	90,2	90,2	-1,1	3,6	8,6	98,7
77	-2,97	7,5	95,7	95,7	-0,57	1	3,5	99,3
78	-2,75	6,4	98,7	98,7	-0,25	0,2	0,8	99,5
79	-2,16	4	98,6	98,6	-0,07	0	0,1	98,7
80	-1,47	1,8	78,6	78,6	0,54	0,9	10,7	89,3
81	-0,9	0,7	35,5	35,5	0,86	2,2	32,9	68,4
82	0,12	0	0,3	0,3	2,21	14,4	87,7	88
83	0,92	0,7	15,5	15,5	2,07	12,6	78,6	94,1
84	1,45	1,8	22,1	22,1	2,62	20,1	72,3	94,4
85	1,63	2,3	41,7	41,7	1,82	9,8	52	93,7
86	1,28	1,4	53,5	53,5	0,6	1,1	11,9	65,4
87	0,86	0,6	38,5	38,5	-0,24	0,2	2,9	41,5
88	1,01	0,9	73,8	73,8	-0,22	0,1	3,6	77,4
89	1,45	1,8	62,4	62,4	-0,79	1,8	18,3	80,7
90	1,9	3,1	71,8	71,8	-0,92	2,5	16,8	88,6
91	2,95	7,4	89,6	89,6	-0,92	2,5	8,8	98,4
92	3,56	10,8	73,4	73,4	-2,09	12,8	25,2	98,6
93	4,12	14,5	79,8	79,8	-1,83	9,8	15,7	95,4

Abréviations :

Coord : coordonnée du tableau sur l'axe;

Ctr : contribution relative du tableau à l'inertie expliquée par l'axe;

Cos<sup>2</sup> : qualité de représentation sur l'axe;

Qlt : qualité de représentation sur le sous-espace engendré par l'axe et les axes précédents

13

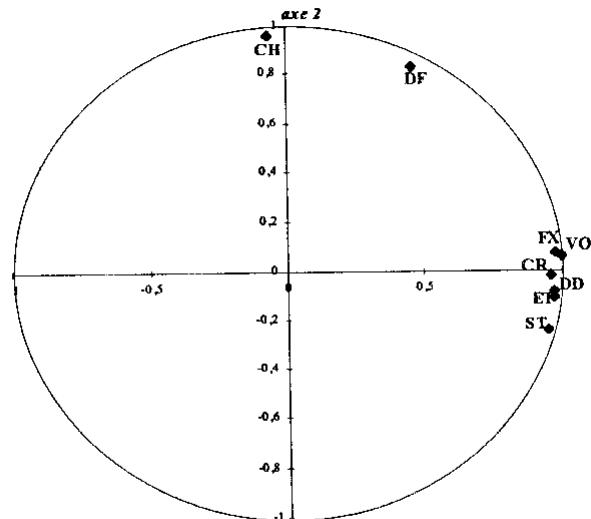
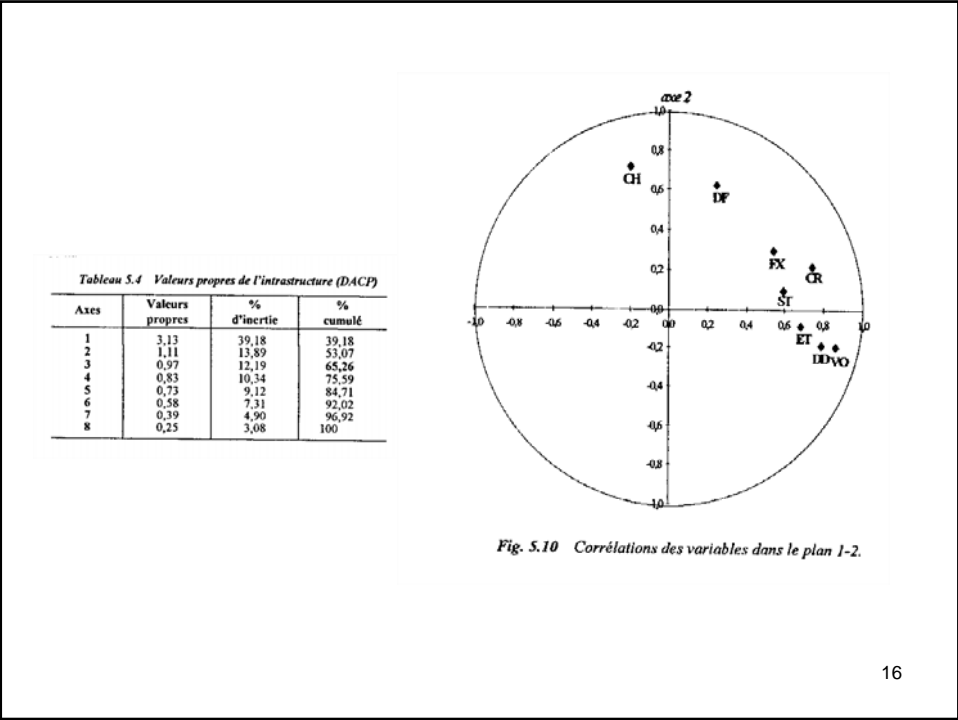
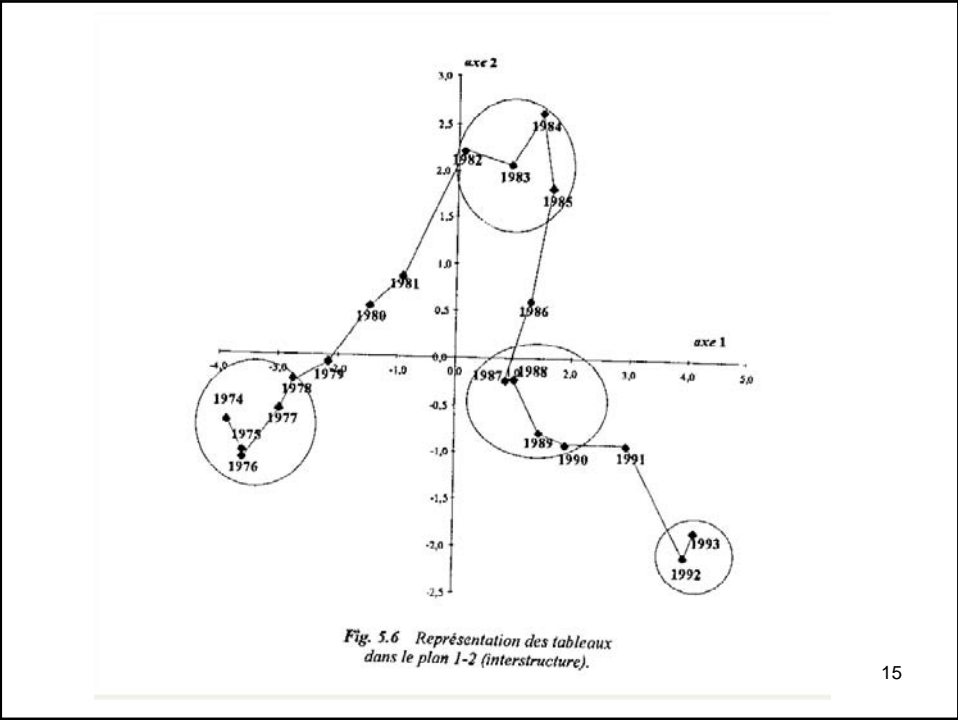
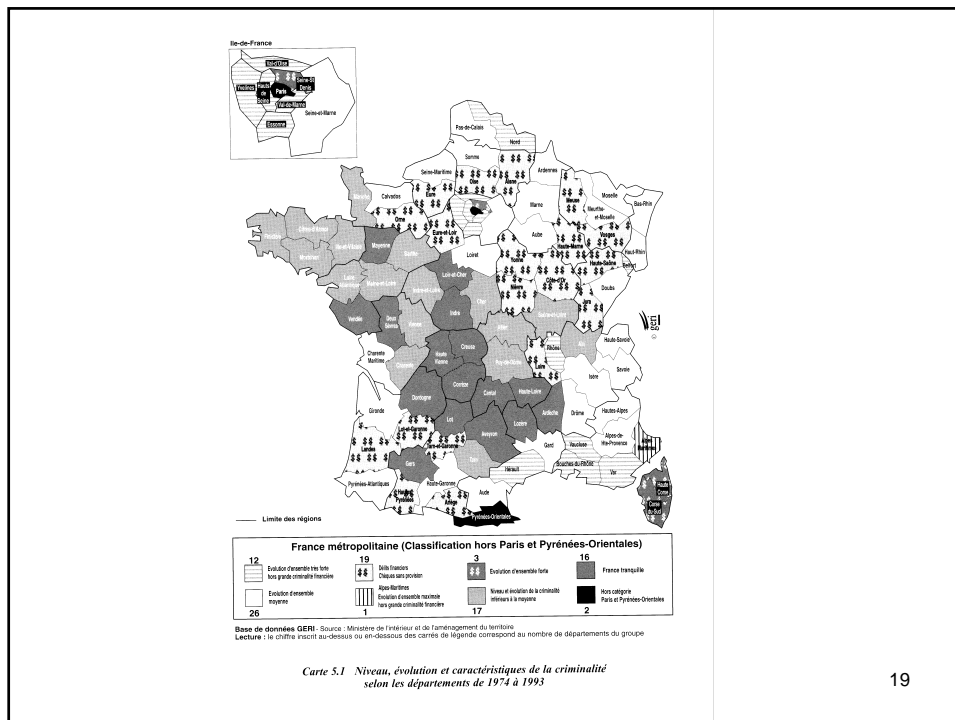


Fig. 5.5 Représentation des corrélations des variables dans le plan 1-2 (interstructure de la DACP).

14







19

### 3. La méthode STATIS (Escoufier et L'Hermier des Plantes (1976) + C.Lavit)

- Structuration de Tableaux A Trois Indices de la Statistique
- permet l'exploration simultanée de plusieurs tableaux de données quantitatives:
- n individus et p variables → STATIS  
– Privilégie la position des individus

20

# STATIS

- Idées essentielles 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 ou **intrastructure**

21

## Rappels d'ACP

- ACP classique : triplet ( $X$ ,  $M$ ,  $D$ )
    - facteurs principaux :  $MVu = \lambda u$
    - composantes principales :  $c = Xu$ 
      - $MX'DXu = \lambda u$
      - $XMx'DXu = \lambda Xu$
      - $WDc = \lambda c$
- $W$  =matrice des produits scalaires entre individus
- Composantes principales**  $WDc = \lambda c$

22

## 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

23

## Interstructure, objet représentatif: $W_t$

- $W_t = X_t M_t X_t'$  caractérise  $(X_t, M_t, D)$ 
  - $W_t$  contient les produits scalaires entre individus = tous les liens inter- individus
  - ses vecteurs propres sont les composantes principales
  - permet de reconstruire la géométrie du nuage
  - "Opérateur d'Escoufier"

24

- Métrique de l'ACP:

définit une distance entre objets

Produit scalaire de Hilbert Schmidt:

$$S_{t,t'} = \langle \mathbf{W}_t | \mathbf{W}_{t'} \rangle_{HS} = \text{Tr} (\mathbf{D}\mathbf{W}_t \mathbf{D}\mathbf{W}_{t'})$$

Si les objets  $W_t$  ont des normes très différentes, on utilise le cosinus associé:

$$RV_{t,t'} = \langle \mathbf{W}_t | \mathbf{W}_{t'} \rangle / \sqrt{\text{Tr}[(\mathbf{D}\mathbf{W}_t)^2] \text{Tr}[(\mathbf{D}\mathbf{W}_{t'})^2]}$$

25

- RV coefficient de corrélation vectoriel

- mesure la liaison entre deux tableaux de données portant sur les mêmes individus
- RV=1 si les nuages associés sont homothétiques
- RV positif ou nul

26

- STATIS: une ACP particulière

- Individu (ou "supervariable") =  $W_t$  = objet représentatif d'un tableau
- Matrice de corrélation =  $S$  = Matrice des coefficients RV (produits scalaires entre objets normés  $W_t$ )
- Composante principale = vecteur propre de  $S$
- Facteur taille car tous les RV sont positifs

27

## Interstructure image représentative des objets

- ACP de  $S$

Les vecteurs propres de  $S$  associés aux deux plus grandes valeurs propres permettent la représentation des objets  $W_1, \dots, W_T$  sur un plan factoriel :

- Les coordonnées des  $W_t$  sur l'axe  $i$  sont contenues dans  $c_{t,i}$

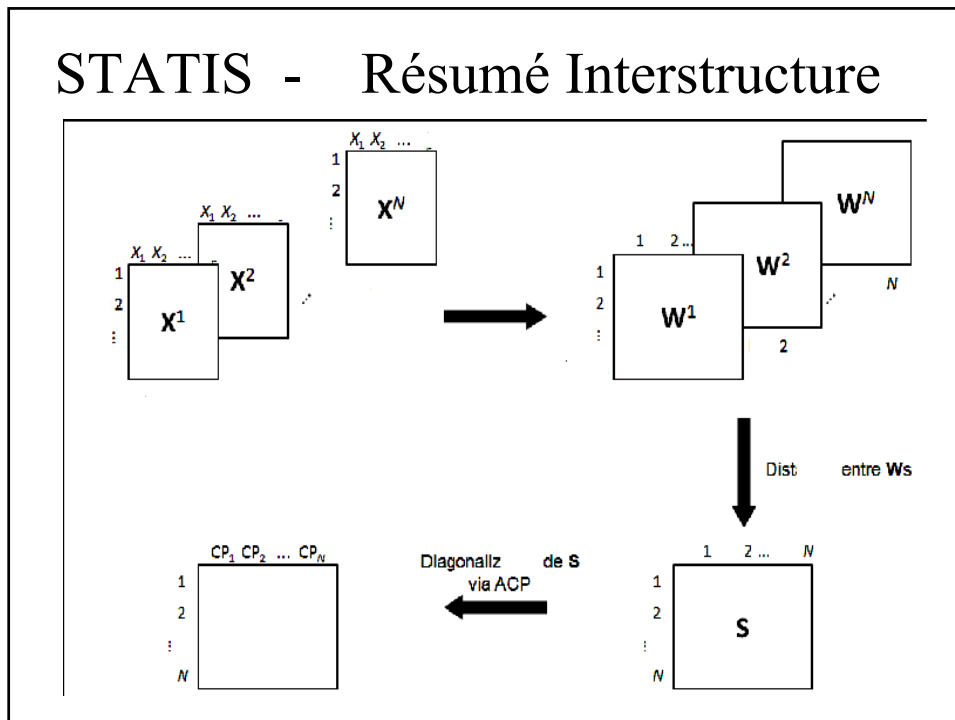
$$c_{t,i} = \sqrt{\lambda_i} u_i$$

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

$u_i$  :  $i^{\text{ème}}$  vecteur propre de  $S$

28

## STATIS - Résumé Interstructure



## 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  $W_t$

$$W_{co} = \sum_{t=1}^T \alpha_t W_t$$

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

31

En ACP usuelle, la première composante principale vérifie un principe d'association maximale

$$\max \sum_{j=1}^p r^2(c; x_j) \quad \text{avec } c = \sum_{j=1}^p \alpha_j x_j$$

Ici on prendra pour  $\alpha_j$  les composantes (toutes positives) du premier vecteur propre de S

32

## Recherche d'un compromis

Matrice compromis  $W_{CO}$

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

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

$W_t = X_t M X_t'$

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

1<sup>ère</sup> valeur propre de **S**

Matrice de données initiales

33

- $W_{CO}$  est encore un opérateur d'Escoufier car les  $\alpha_j$  sont non négatifs
- $W_{CO}$  est donc associé à un tableau de données "compromis"  $X$  tel que  $W_{CO} = X M X'$

- ACP de  $X = \left[ \sqrt{\alpha_1} X_1 \right] \left[ \sqrt{\alpha_2} X_2 \right] \dots \left[ \sqrt{\alpha_T} X_T \right]$

- Pondération des tableaux

34

## Recherche d'un compromis

4 cas:

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

35

## Statis: exemple

- Elections présidentielles de 1969 à 1988
  - résultats par département

36

En 1969 :

- GI1<sup>1</sup> : Vote Defferre
- DII : Vote Poher + Vote Pompidou
- GP1 : Vote Rocard + Vote Duclos + Vote Krivine
- AT1 : Vote Ducatel
- NP1 : Votes blancs et nuls + Abstention
- CN2<sup>2</sup> : Vote Poher
- D12 : Vote Pompidou
- NP2 : Votes blancs et nuls + Abstention

En 1974 :

- GI1 : Vote Mitterrand
- DII : Vote Giscard + Vote Muller + Vote Royer + Vote Chaban
- GP1 : Vote Laguiller + Vote Krivine
- DP1 : Vote Le Pen + Vote Renouvin *(droite protestataire)*
- AT1 : Vote Héraut + Vote Sebag + Vote Dumont
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- D12 : Vote Giscard d'Estaing
- NP2 : Votes blancs et nuls + Abstention

En 1981 :

- GI1 : Vote Mitterrand + Vote Crépeau
- DII : Vote Giscard + Vote Chirac + Vote Debré + Vote Garaud
- GP1 : Vote Laguiller + Vote Marchais + Vote Bouchardeau
- AT1 : Vote Lalonde
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- D12 : Vote Giscard d'Estaing
- NP2 : Votes blancs et nuls + Abstention

En 1988 :

- GI1 : Vote Mitterrand
- DII : Vote Barre + Vote Chirac
- GP1 : Vote Lajoinie + Vote Juquin + Vote Laguiller + Vote Bousset
- DP1 : Vote Le Pen
- AT1 : Vote Waechter
- NP1 : Votes blancs et nuls + Abstention
- GI2 : Vote Mitterrand
- D12 : Vote Chirac
- NP2 : Votes blancs et nuls + Abstention

37

Tableau 3.1 Statistiques élémentaires sur les tableaux de données

Scrutin 1969									
(en %)	GI1	DII	GP1	AT1	NP1	D12	CN2	NP2	
Moyenne	3,74	52,77	19,48	0,98	23,04	37,79	28,33	33,88	
Écart-type	1,07	5,83	5,18	0,23	1,68	5,35	4,37	5,36	

Scrutin 1974									
(en %)	GI1	DII	GP1	DP1	AT1	NP1	D12	GI2	NP2
Moyenne	36,33	43,41	2,46	0,78	0,98	15,81	44,19	42,86	12,95
Écart-type	5,55	5,58	0,49	0,18	0,36	1,68	5,83	6,13	1,42

Scrutin 1981									
(en %)	GI1	DII	GP1	AT1	NP1	D12	GI2	NP2	
Moyenne	23,12	39,61	14,84	1,21	19,36	40,35	44,10	15,54	
Écart-type	2,92	4,75	3,94	0,55	2,01	4,84	4,94	1,61	

Scrutin 1988									
(en %)	GI1	DII	GP1	DP1	AT1	NP1	D12	GI2	NP2
Moyenne	27,70	29,66	9,16	11,09	3,12	19,26	37,78	44,51	17,71
Écart-type	3,42	4,05	2,61	3,30	0,83	1,65	3,94	4,19	1,96

38

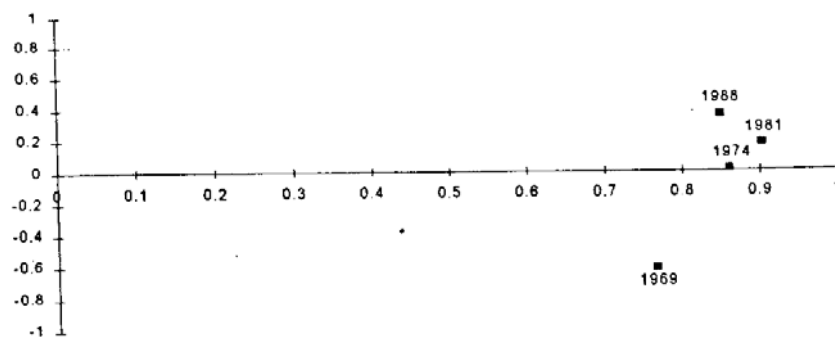
**Tableau 3.2** *Matrice des coefficients RV*

Années	1969	1974	1981	1988
1969	1,000	0,571	0,576	0,490
1974	0,571	1,000	0,710	0,622
1981	0,576	0,710	1,000	0,746
1988	0,490	0,622	0,746	1,000

**Tableau 3.3** *Valeurs propres de l'interstructure*

Axes	Valeurs propres	% d'inertie	% cumulé
1	2,87	71,66	71,66
2	0,54	13,53	85,19
3	0,36	9,04	94,24
4	0,23	5,76	100

39



**Fig. 3.10** *Représentation de l'interstructure.*

40

## Intrastructure image représentative des points compromis

- **ACP de  $W_{CO}$**
- Les vecteurs propres de  $W_{CO}$  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  $\mathbf{c}_{CO,i}$

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

$\swarrow$   $j^{\text{ème}}$  valeur propre de  $\mathbf{W}_{CO}$        $\searrow$   $j^{\text{ème}}$  vecteur propre de  $\mathbf{W}_{CO}$

41

Tableau 3.5 Valeurs propres de l'intrastructure

Axes	Valeurs propres	% d'inertie	% cumulé
1	0,753	31,68	31,68
2	0,480	20,19	51,88
3	0,257	10,81	62,68
4	0,208	8,74	71,42
5	0,166	6,97	78,39
6	0,106	4,44	82,83
7	0,097	4,07	86,90
8	0,074	3,12	90,02

42



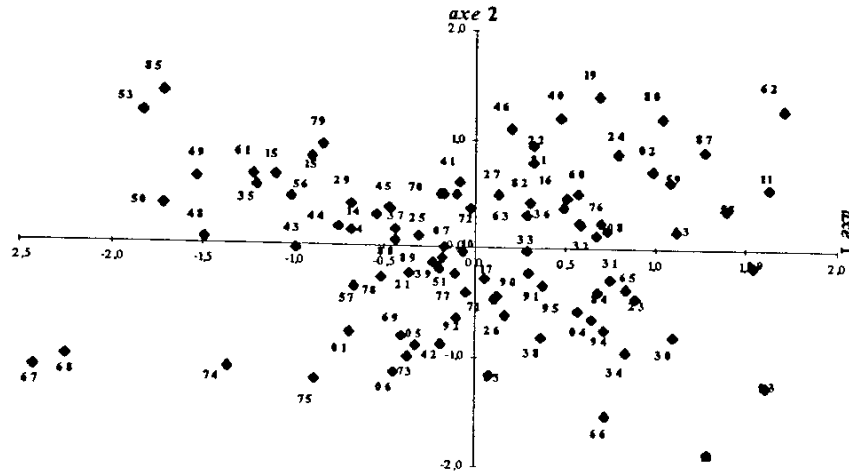
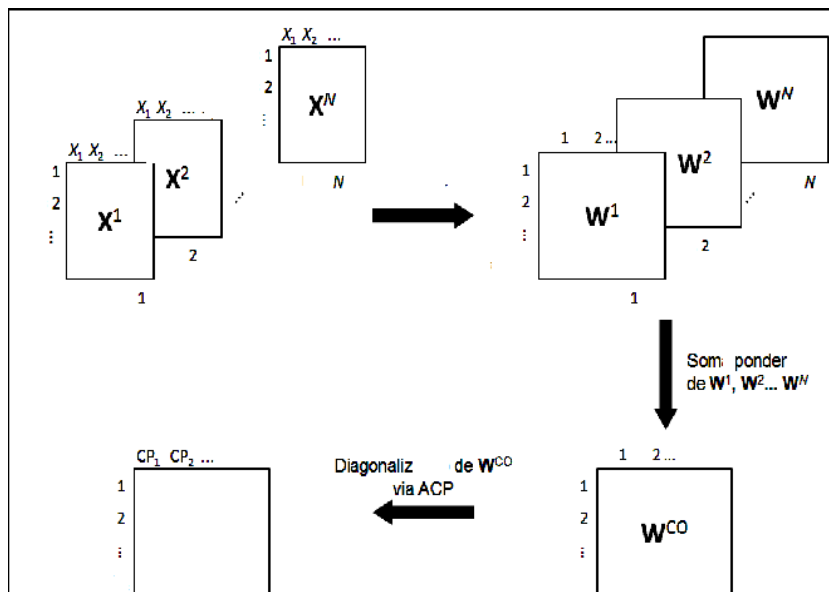


Fig. 3.13 Positions-compromis des individus dans le plan 1-2.

# STATIS - Infrastructure



## STATIS - Trajectoires des individus

- Méthode: technique des points supplémentaires:

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

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

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

47

## 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

48

## 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

Rque: 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

49

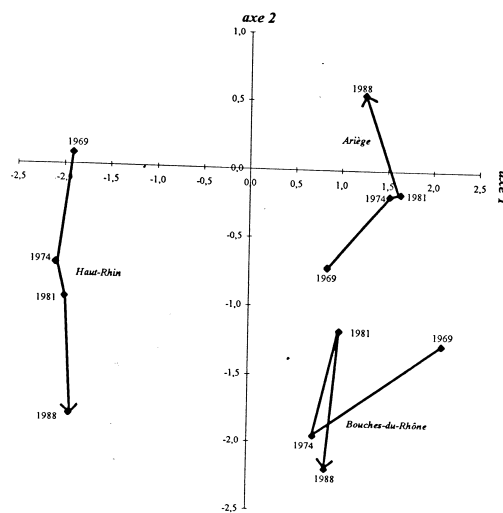
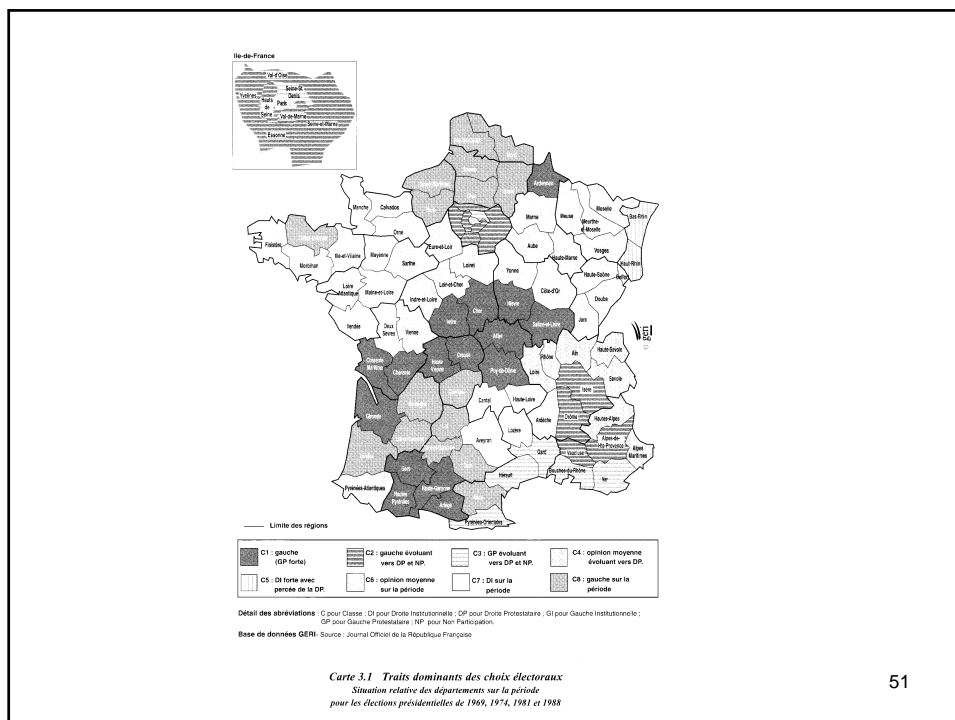


Fig. 3.15 Trajectoires de trois départements dans le plan 1-2.

50



51

## 4 . Méthode AFM

- **Analyse Factorielle Multiple** (B.Escofier et J.Pagès)
- permet l'étude d'un ensemble d'individus décrits par plusieurs groupes de variables **quantitatives ou qualitatives**:
- Groupes = mesures à différents instants  
sous tableaux d'un tableau  
(issus d'un regroupement selon certains critères)

52

## AFM : analyse factorielle multiple

- Comme STATIS:
  - Comparaison globale des groupes de variables: **interstructure**
  - Comparaison des nuages représentant le même ensemble d'individus: **intrastructure et image compromis**
  - Mais ici juxtaposition des tableaux avec :
- Idée de base = pondération des groupes de variables pour équilibrer leur influence dans l'analyse globale

53

## AFM : analyse factorielle multiple

- Dans la juxtaposition des N tableaux,
  - \* plus le nombre de variables est grand, plus le tableau aura une grande influence sur les résultats
  - \* plus les variables d'un tableau sont liées (forte structure) plus ce tableau influence le compromis
- Une autre manière de pondérer les tableaux: égaliser les premières valeurs propres de chaque ACP

54

# AFM

- Objet représentatif  $\frac{1}{\sqrt{\lambda_1^{(t)}}} \mathbf{W}_t$
- Compromis  $\sum_{t=1}^T \frac{1}{\sqrt{\lambda_1^{(t)}}} \mathbf{W}_t$
- Intrastructure ACP de

$$\mathbf{X} = \left[ \frac{1}{\sqrt{\lambda_1^{(1)}}} \mathbf{X}_1 \mid \frac{1}{\sqrt{\lambda_1^{(2)}}} \mathbf{X}_2 \mid \dots \mid \frac{1}{\sqrt{\lambda_1^{(T)}}} \mathbf{X}_T \right]$$

55

Tableau 4.1 Les quatre premières valeurs propres des objets  $W, D$

Tableau 1969

Axes	Inertie	Pourcentage	Cumul
1	2,25	45,03	45,03
2	1,32	26,37	71,41
3	0,77	15,36	86,77
4	0,66	13,23	99,99

Tableau 1974

Axes	Inertie	Pourcentage	Cumul
1	2,16	35,96	35,96
2	1,63	27,24	63,20
3	1,04	17,40	80,64
4	0,67	11,16	91,80

Tableau 1981

Axes	Inertie	Pourcentage	Cumul
1	2,07	41,44	41,44
2	1,31	26,15	67,59
3	0,98	19,52	87,11
4	0,64	12,89	99,99

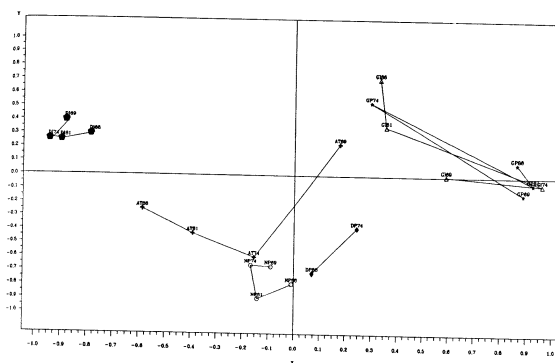
Tableau 1988

Axes	Inertie	Pourcentage	Cumul
1	2,16	35,96	35,96
2	1,79	19,79	65,76
3	0,95	15,85	81,61
4	0,71	11,82	93,42

56

**Tableau 4.2** Les 7 premières valeurs propres de l'infrastructure de l'AFM

Axes	Inertie	Pourcentage	Cumul
1	3,54	34,77	34,77
2	2,30	22,60	57,37
3	1,16	11,36	68,73
4	0,84	8,20	76,92
5	0,50	4,95	81,87
6	0,45	4,39	86,26
7	0,34	3,36	89,62



**Fig. 4.4** Corrélations des variables dans le plan 1-2.

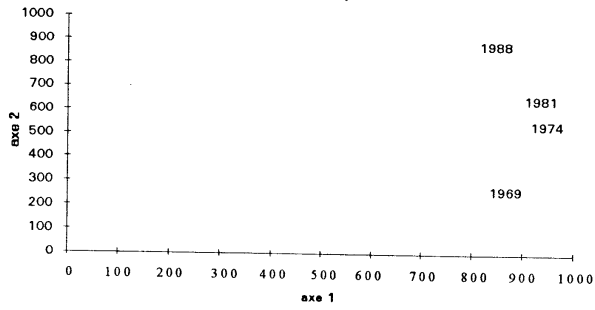


Fig. 4.6 Représentation de l'infrastructure de l'AFM dans le plan 1-2.

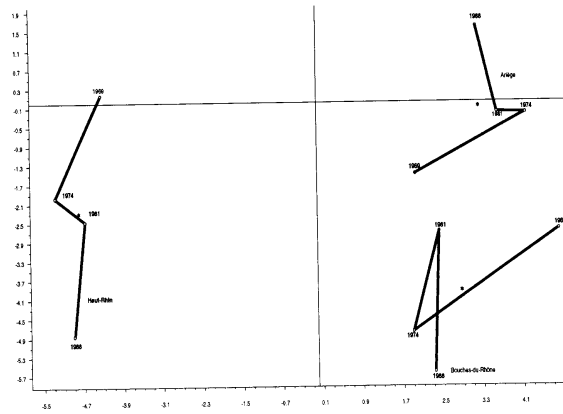


Fig. 4.9 Trajectoires issues de l'AFM.

## Remarques et extensions

- La double ACP : T tableaux où les individus ne sont pas forcément les mêmes, mais variables identiques
- STATIS et AFM T tableaux avec des variables éventuellement différentes mais les mêmes individus
- STATIS duale pour comparer des matrices de variance

61

## Références

- Bry, X. *Analyses factorielles multiples*, Economica, 1996
- Escofier, B., Pagès, J. *Analyses factorielles simples et multiples*, 4<sup>ème</sup> édition, Dunod , 2008
- Kroonenberg, P., *Applied Multiway Data Analysis*, Wiley, 2008
- Lavit, C., *Analyse conjointe de tableaux quantitatifs*, Masson, 1998

62