

# Analyse en composantes principales pour des matrices à trois modes (3MPCA) : principes et applications

Jean-Luc Kop, Bruno Dauvier,  
Anne Congard, Vincent Berthet

*Université Nancy 2 et Université de Provence*

Troisième atelier Modevaia, Markstein, 28-30 juin 2011

## PLAN

1 mode, 2 modes, 3 modes....

Les modèles pour l'analyse de matrices à trois modes

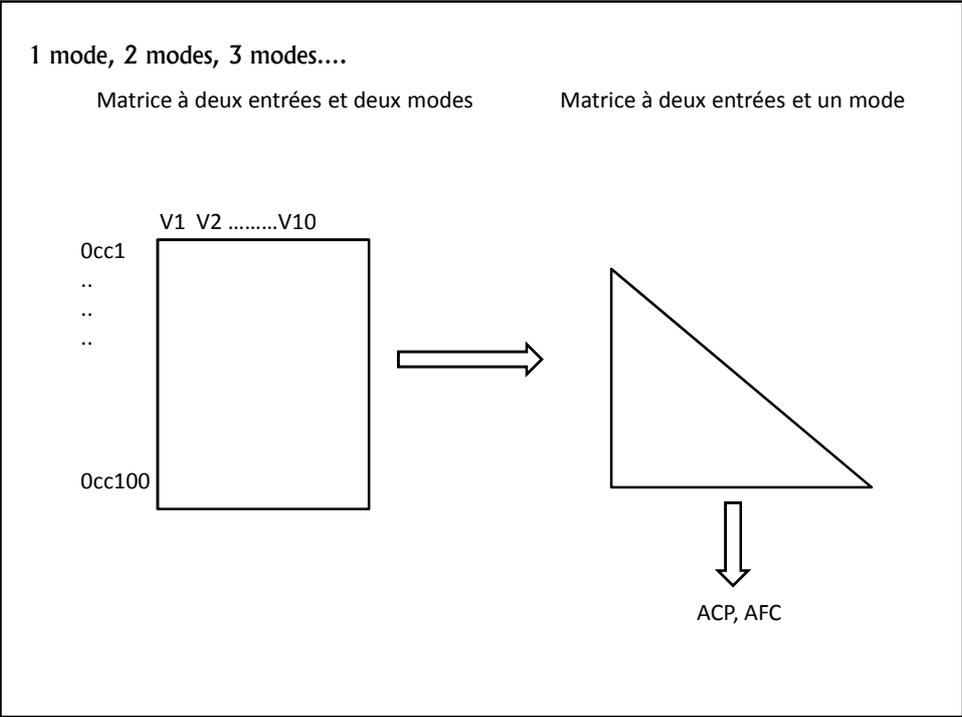
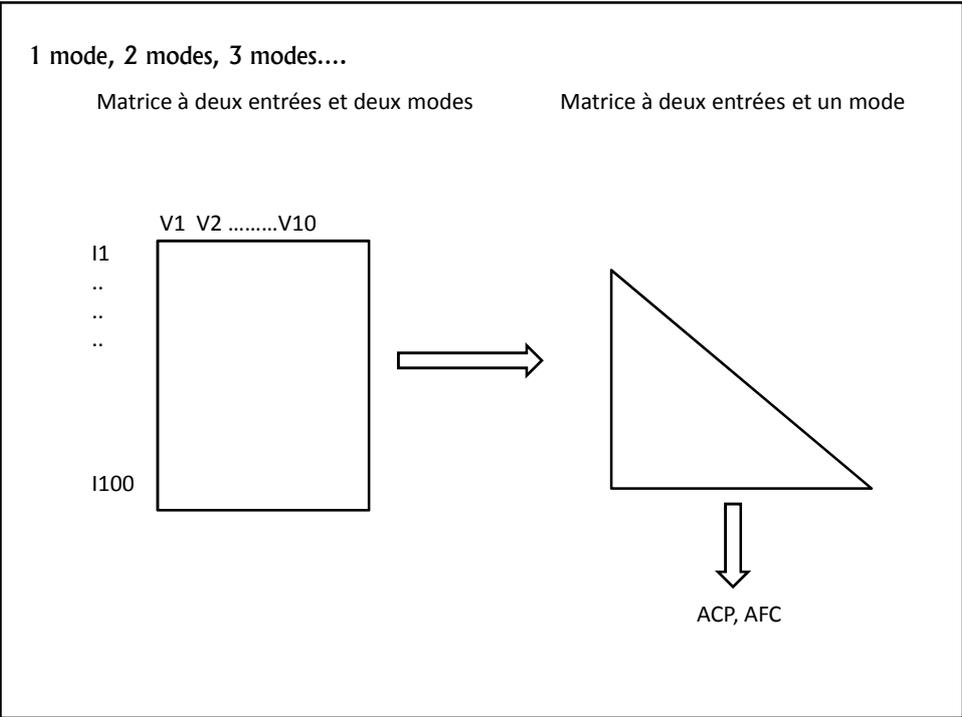
Le modèle en composantes principales

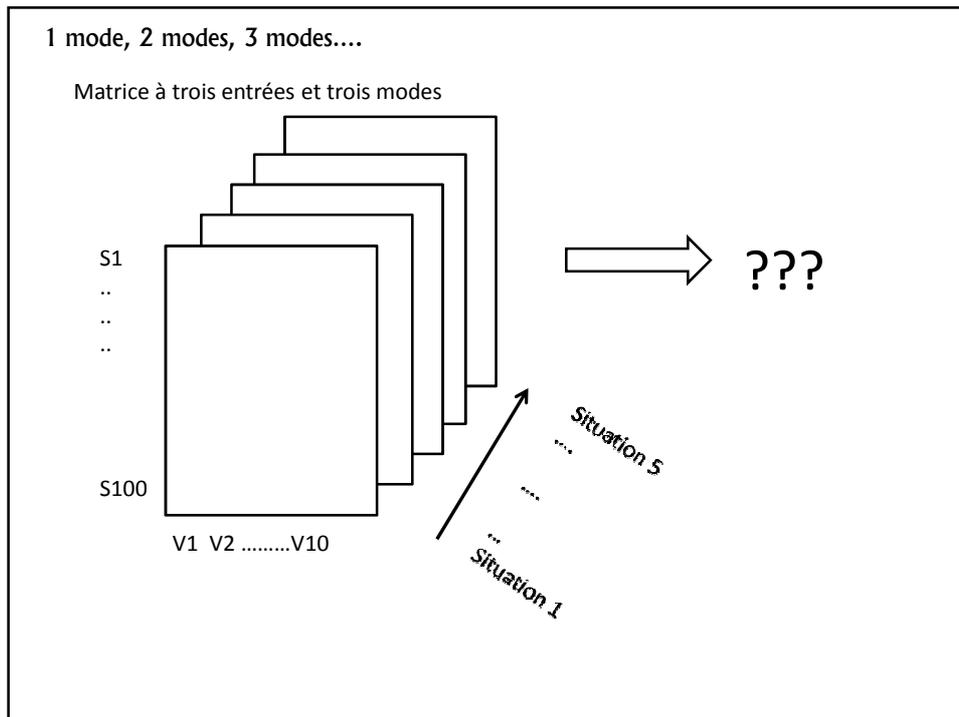
Exemple pédagogique

Les logiciels

Les « analyses trois modes » pour données expérimentales seulement ?

Un exemple pour données d'observation : affectivité et événements vécus





### Les modèles pour l'analyse de matrices à trois modes

Deux grandes familles pour l'analyse de matrices à trois modes (Gonzales, De Boeck, & Tuerlinckx, 2008) :

- celle issue de la logique de l'analyse en facteurs communs et des modèles structuraux (analogie avec matrices MTMM)
- celle issue des travaux de Tucker (1966) → généralisation de l'analyse en composantes principales (3MPCA : « three way principal components analysis »)

### Le modèle en composantes principales

En ACP traditionnelle (2 modes), on cherche :

- une matrice des composantes pour les variables
- une matrice de composantes pour les sujets

de manière à minimiser :

$$\sum_{i=1}^I \sum_{j=1}^J \left( x_{ij} - \sum_{q=1}^Q a_{iq} b_{jq} \right)^2$$

Avec :

$x_{ij}$  = score (standardisé) de l'individu  $i$  sur une variable  $j$

$q$  = nombre de composantes

$a_{iq}$  = score de l'individu  $i$  sur la composante  $q$

$b_{jq}$  = saturation de la variable  $j$  sur la composante  $q$

### Le modèle en composantes principales

#### 3MPCA

scores  $x_{ijk}$  = le score d'un individu  $i$ , à une variable  $j$ , dans une situation  $k$

On cherche trois matrices de composantes :

- La matrice A : matrice des relations entre les individus et les composantes
- La matrice B : matrice des relations entre les variables et les composantes
- La matrice C : matrice des relations entre les situations et les composantes

### Le modèle en composantes principales

#### Différence essentielle entre 2 MPCA et 3MPCA

Association univoque entre composantes des matrices A (sujets) et B (variable) dans 2 MPCA. Ce n'est plus le cas en 3 MPCA.

#### Conséquences (en 3MPCA) :

- le nombre de composantes peut être différent pour chaque mode
- une quatrième matrice est nécessaire pour représenter les relations entre les composantes de chacun des modes : la matrice noyau G (« *core matrix* »)

### Le modèle en composantes principales

$$x_{ijk} \cong \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr}$$

- $x_{ijk}$  = score de l'individu  $i$  sur une variable  $j$  dans une situation  $k$
- $a_{ip}$  = valeur de l'individu  $i$  sur l'une des  $p$  composantes résumant les individus
- $b_{jq}$  = valeur de la variable  $j$  sur l'une des  $q$  composantes résumant les variables
- $c_{kr}$  = valeur de l'occasion  $k$  sur l'une des  $r$  composantes résumant les occasions
- $g_{pqr}$  = éléments de la matrice noyau des relations entre les composantes des trois modes

## Le modèle en composantes principales

$$x_{ijk} \cong \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr}$$

**Cas particulier**

Si on impose que :

- le nombre de composantes est le même pour chaque mode
- que la composante de l'un des modes est lié à une seule composante d'un autre mode (i.e.  $g_{pqr} = 1$  lorsque  $p = q = r$  et  $g_{pqr} = 0$  dans les autres cas).

➔ Parallel Factor Analysis (PARAFAC )

- Avantage : le modèle a une solution unique (pas de rotation)
- Inconvénient : postulats peu plausibles (absence d'interactions entre les composantes de chacun des modes)

## Exemple pédagogique (Kiers &amp; van Mechelen, 2001)

- six sujets
- cinq variables : niveau de réponses dans cinq registres comportementaux (Emotionnel, Sensible, Bienveillant (*caring*), Conscientieux et Exactitude)
- quatre situations : passer un examen, faire un exposé, pique-nique familial, avoir un rendez-vous (galant)

Individu	Examen					Exposé					Pique nique					Rendez-vous				
	Em	S	B	C	Ex	Em	S	B	C	Ex	Em	S	B	C	Ex	Em	S	B	C	Ex
Anne	0,0	0,0	1,2	3,0	3,0	0,6	0,6	1,3	2,4	2,4	3,0	3,0	1,8	0,0	0,0	3,6	3,6	2,5	0,9	0,9
Benoit	0,0	0,0	0,8	2,0	2,0	0,2	0,2	0,8	1,8	1,8	1,0	1,0	1,0	1,0	1,0	1,2	1,2	1,4	1,8	1,8
Christophe	0,0	0,0	0,8	2,0	2,0	0,2	0,2	0,8	1,8	1,8	1,0	1,0	1,0	1,0	1,0	1,2	1,2	1,4	1,8	1,8
Dorothee	0,0	0,0	1,2	3,0	3,0	0,6	0,6	1,3	2,4	2,4	3,0	3,0	1,8	0,0	0,0	3,6	3,6	2,5	0,9	0,9
Emma	0,0	0,0	1,0	2,5	2,5	0,4	0,4	1,1	2,1	2,1	2,0	2,0	1,4	0,5	0,5	2,4	2,4	2,0	1,3	1,3
Françoise	0,0	0,0	1,2	3,0	3,0	0,6	0,6	1,3	2,4	2,4	3,0	3,0	1,8	0,0	0,0	3,6	3,6	2,5	0,9	0,9

## Exemple pédagogique (Kiers &amp; van Mechelen, 2001)

A : Individus	A1	A2
Anne	1,0	0,0
Benoît	0,0	1,0
Christophe	0,0	1,0
Dorothée	1,0	0,0
Emma	0,5	0,5
Françoise	1,0	0,0

A1 : féminité  
A2 : masculinité

B : Variables	B1	B2
Emotionnel	1,0	0,0
Sensible	1,0	0,0
Bienveillant	0,6	0,4
Conscientieux	0,0	1,0
Exactitude	0,0	1,0

B1 : émotionnalité  
B2 : contrôle de soi

C :	C1	C2
Occasion		
Passer un examen	1,0	0,0
Faire un exposé	0,8	0,2
Pique nique familial	0,0	1,0
Rendez-vous galant	0,3	1,2

C1 : situations de performance  
C2 : situations sociales

## Exemple pédagogique (Kiers &amp; van Mechelen, 2001)

Matrice noyau

	C1 : Situations de performance		C2 : Situations sociales	
	B1 : émotionnalité	B2 : contrôle	B1 : émotionnalité	B2 : contrôle
A1 : Féminité	0,0	3,0	3,0	0,0
A2 : Masculinité	0,0	2,0	1,0	1,0

- dans les situations de performance, les personnes féminines se comportent autant émotionnellement que les personnes masculines et les personnes féminines ont davantage de contrôle que les personnes masculines
- dans les situations sociales, les personnes féminines se montrent beaucoup plus émotives que les personnes masculines, alors que dans ces mêmes situations, les personnes féminines ont moins de contrôle que les personnes masculines

## Exemple pédagogique (Kiers &amp; van Mechelen, 2001)

## Réponses reproduites

	C1 : Situations de performance		C2 : Situations sociales	
	B1 : émotionnalité	B2 : contrôle	B1 : émotionnalité	B2 : contrôle
A1 : Féminité	0,0	3,0	3,0	0,0
A2 : Masculinité	0,0	2,0	1,0	1,0

A : Individus	A1	A2
Anne	1,0	0,0
Benoit	0,0	1,0
Christophe	0,0	1,0
Dorothée	1,0	0,0
Emma	0,5	0,5
Françoise	1,0	0,0

B : Variables	B1	B2
Emotionnel	1,0	0,0
Sensible	1,0	0,0
Bienveillant	0,6	0,4
Conscientieux	0,0	1,0
Exactitude	0,0	1,0

C :	C1	C2
Occasion		
Passer un examen	1,0	0,0
Faire un exposé	0,8	0,2
Pique nique familial	0,0	1,0
Rendez-vous galant	0,3	1,2

$$x_{ijk} \cong \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr}$$

→ Christophe (entièrement « masculin »), à la variable « sensible » (entièrement émotionnelle), à un pique nique familial (situation entièrement sociale) : réponse reproduite = 1

## Exemple pédagogique (Kiers &amp; van Mechelen, 2001)

## Réponses reproduites

	C1 : Situations de performance		C2 : Situations sociales	
	B1 : émotionnalité	B2 : contrôle	B1 : émotionnalité	B2 : contrôle
A1 : Féminité	0,0	3,0	3,0	0,0
A2 : Masculinité	0,0	2,0	1,0	1,0

A : Individus	A1	A2
Anne	1,0	0,0
Benoit	0,0	1,0
Christophe	0,0	1,0
Dorothée	1,0	0,0
Emma	0,5	0,5
Françoise	1,0	0,0

B : Variables	B1	B2
Emotionnel	1,0	0,0
Sensible	1,0	0,0
Bienveillant	0,6	0,4
Conscientieux	0,0	1,0
Exactitude	0,0	1,0

C :	C1	C2
Occasion		
Passer un examen	1,0	0,0
Faire un exposé	0,8	0,2
Pique nique familial	0,0	1,0
Rendez-vous galant	0,3	1,2

$$x_{ijk} \cong \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr}$$

→ Anne (entièrement féminine), à la variable « sensible » (entièrement émotionnelle), à un nouveau rendez-vous (mélange de performance et de social) : réponse reproduite =  $(0.3 \times 0) + (1.2 \times 3.0) = 3.6$

### Les logiciels

Package PTAK dans R (<http://cran.r-project.org/web/packages/PTAK/index.html>)

Logiciel 3WayPack (<http://three-mode.leidenuniv.nl/>)

Toolbox « tucker3 » de Kiers pour Matlab  
(<http://www.ppsw.rug.nl/~kiers/Threeway%20m-files.zip> )

Toolbox « N-way » de Anderson & Bro pour Matlab  
(<http://www.models.life.ku.dk/nwaytoolbox>)

Procédures MDS dans les logiciels ?? (INDSCAL et PARAFAC seraient équivalents ?)

Les « analyses trois modes » pour données expérimentales seulement ?

Les données typiques d'application 3MPCA (« expérimentales »)

	Situation 1				Situation 2				Situation 3			
	V1	V2	V3	V4	V1	V2	V3	V4	V1	V2	V3	V4
I1												
I2												
I3												
....												

Les « analyses trois modes » pour données expérimentales seulement ?

Des données typiques d'observation (« écologiques »)

- on enregistre les réponses des sujets à différentes occasions temporelles
- à chacune des occasions temporelles, on leur demande de :
  - répondre aux variables
  - caractériser les situations vécues

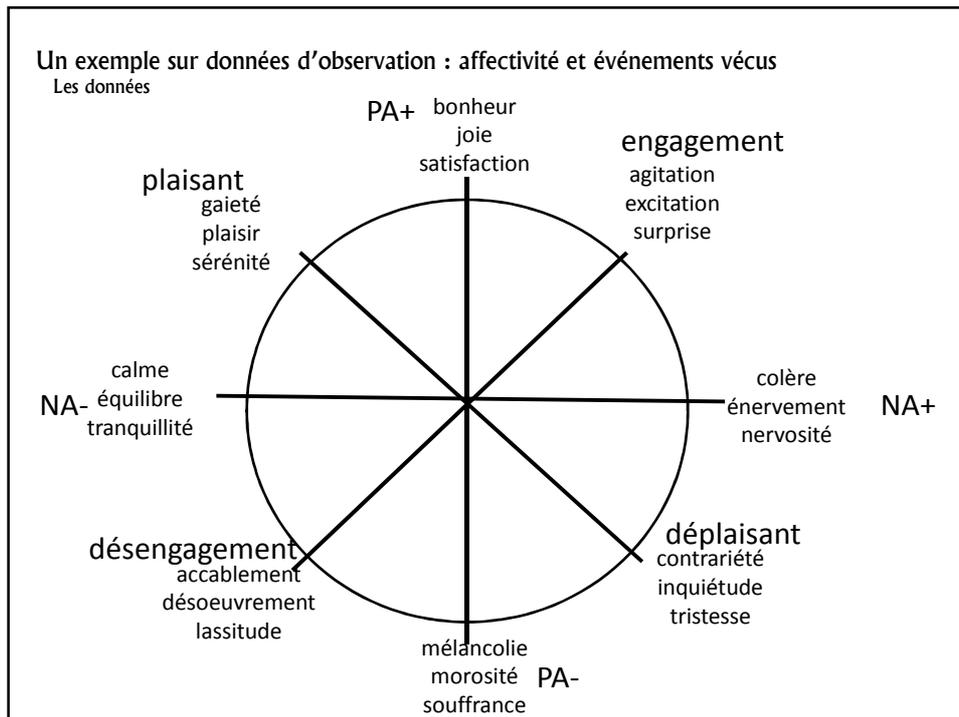
Problème : il n'y a pas trois modes, mais quatre (individus, variables, situations, moments)

→ La réduction à trois modes (par agrégation des moments relevant de la même situation) ne va pas de soit.... : p. ex. sur la durée de l'observation, une personne peut connaître très souvent une situation S1 et jamais une situation S2 et inversement pour une autre personne...

Un exemple sur données d'observation : affectivité et événements vécus

Les données

- 103 participants (17 à 61 ans)
- Interrogés 1 fois par jour pendant 30 jours (le soir, vers 19h)
- Questionnaire d'affectivité (24 items, échelle d'intensité)
- Questionnaire d'événements rencontrés dans la journée (domaine et impact)



Un exemple sur données d'observation : affectivité et événements vécus  
Les données

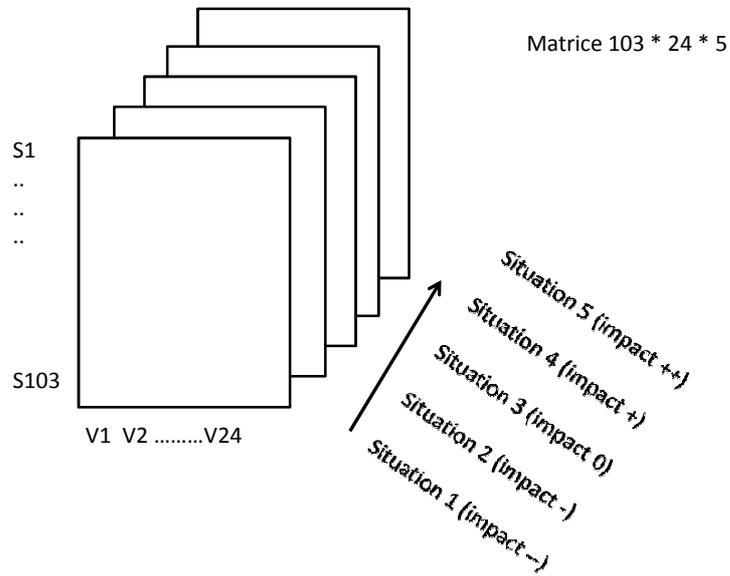
Agrégation des données en fonction de l'impact des événements :

- situation 1 : impact --
- situation 2 : impact -
- situation 3 : impact 0
- situation 4 : impact +
- situation 5 : impact ++

Chaque « modalité de situation » correspond pour un sujet donné à 6 observations (les 6 observations avec les impacts les plus négatifs ... les 6 observations avec les impacts les plus positifs)

### Un exemple sur données d'observation : affectivité et événements vécus

Les données



### Un exemple sur données d'observation : affectivité et événements vécus

Les résultats

#### 1. Analyse de variance

Somme des carrés des différents modes et de leurs interactions

SS_a	=	1850.025741	(	16.28%
SS_b	=	2464.286497	(	21.69%
SS_c	=	65.765221	(	0.58%
SS_ab	=	4149.500755	(	36.52%
SS_ac	=	248.848775	(	2.19%
SS_bc	=	1011.274944	(	8.90%
SS_abc	=	1572.701062	(	13.84%

## Un exemple sur données d'observation : affectivité et événements vécus

Les résultats

### 2. Pré-traitement des données

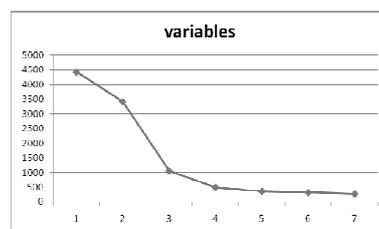
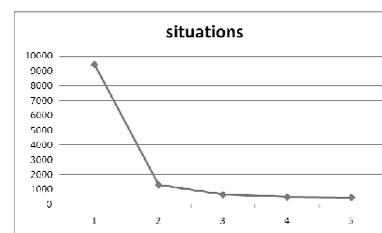
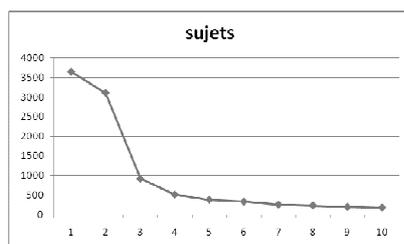
Les données sont :

- centrées sur les sujets
- normalisées sur les variables

## Un exemple sur données d'observation : affectivité et événements vécus

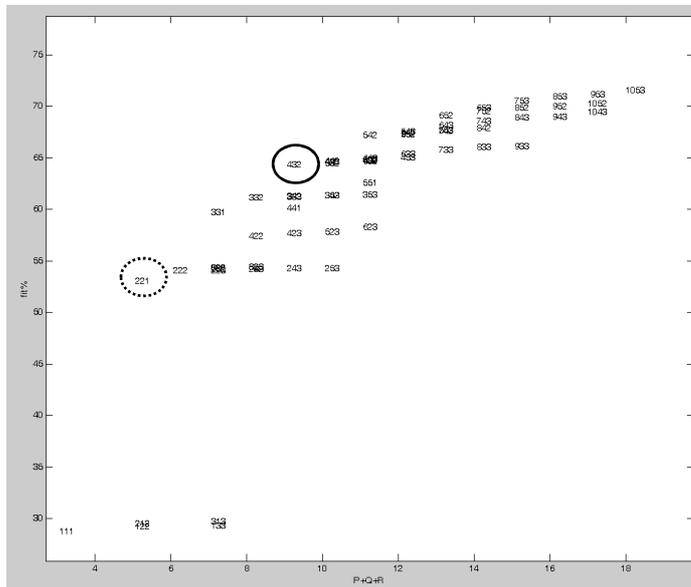
Les résultats

### 3. Choix de la dimensionnalité



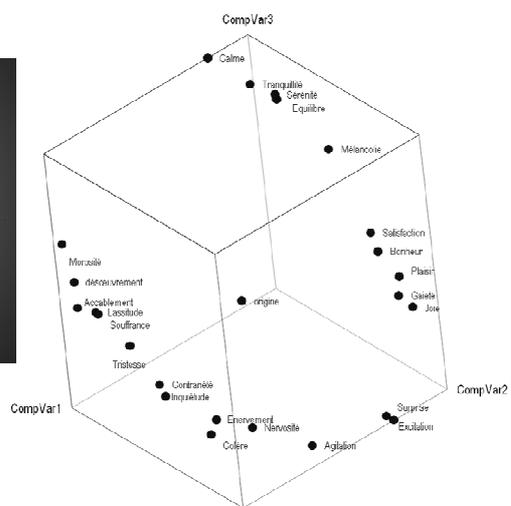
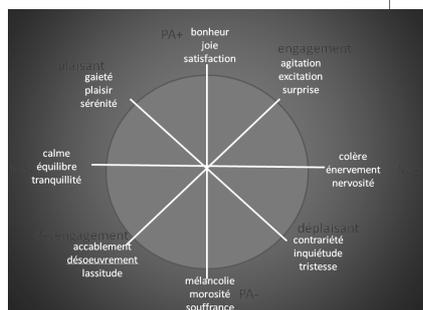
Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

3. Choix de la dimensionnalité



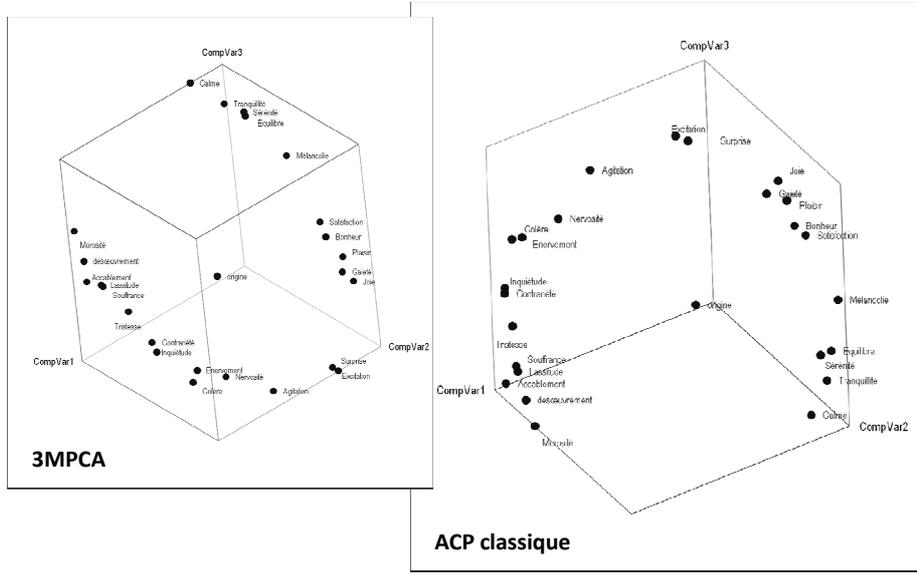
Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

4. Matrice des composantes pour les variables (solution 4,3,2)



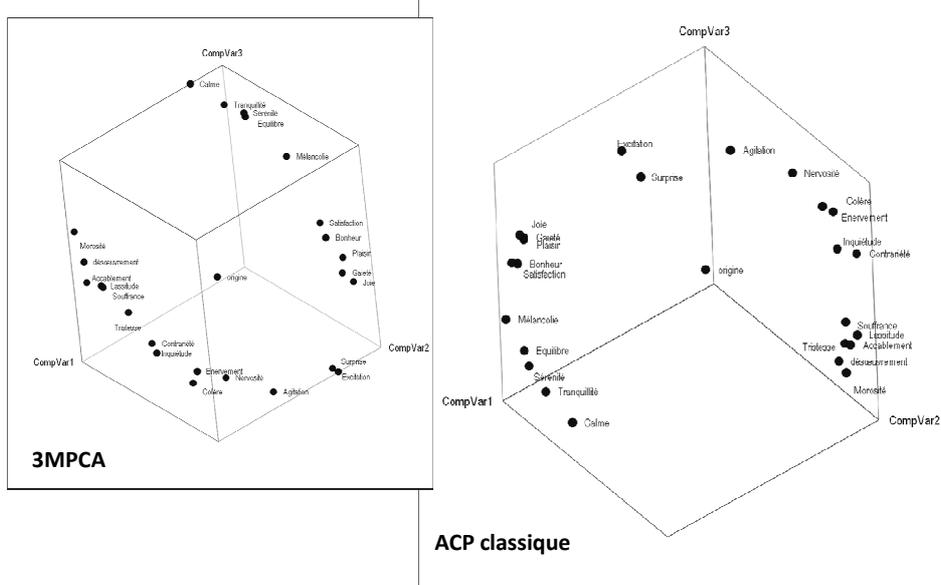
Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

5a. Comparaison avec des ACP classiques (données agrégées sur les situations)



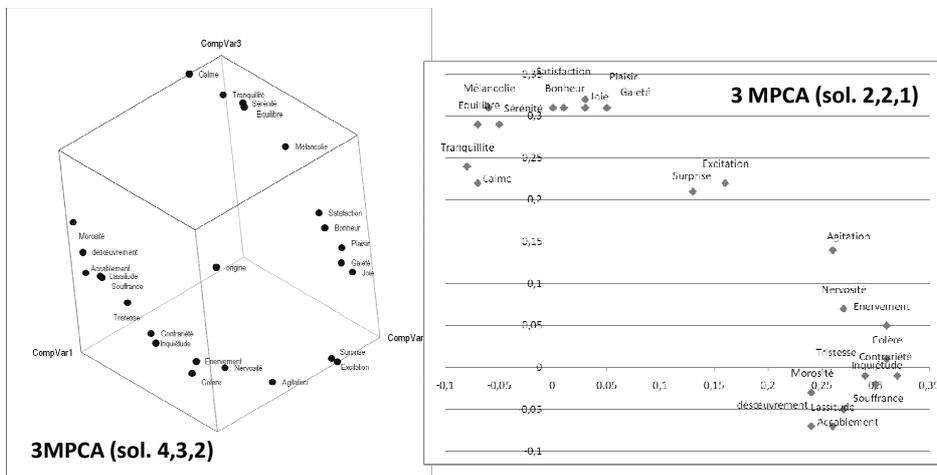
Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

5c. Comparaison avec des ACP classiques (sujets répétés)



Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

5d. Et avec seulement deux composantes 3MPCA ?



Un exemple sur données d'observation : affectivité et événements vécus  
Les résultats

6. Plus-value de 3MPCA : la matrice noyau

a) La matrice de composantes pour les variables

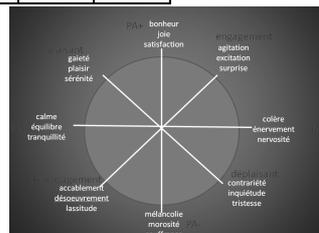
	B1	B2	B3
Surprise	-0.02	0.36	-0.09
Enervement	0.24	0.22	-0.02
Plaisir	-0.13	0.29	0.11
Tristesse	0.29	0.06	0.09
Mélancolie	-0.14	0.11	0.29
Equilibre	-0.10	0.03	0.38
Morosité	0.31	-0.08	0.25
Calme	-0.05	-0.09	0.44
Colère	0.24	0.20	-0.06
Agitation	0.13	0.34	-0.08
Bonheur	-0.12	0.25	0.15
Inquiétude	0.27	0.12	-0.00

	B1	B2	B3
Gaieté	-0.11	0.31	0.09
Lassitude	0.29	-0.03	0.12
Sérénité	-0.09	0.04	0.40
désœuvrement	0.30	-0.07	0.17
Excitation	-0.02	0.38	-0.09
Nervosité	0.18	0.24	-0.06
Satisfaction	-0.13	0.22	0.17
Joie	-0.13	0.32	0.06
Contrariété	0.28	0.12	0.03
Tranquillité	-0.09	-0.03	0.39
Souffrance	0.30	-0.01	0.13
Accablement	0.32	-0.04	0.14

B1 : de NA+ à désengagement (émotions négatives)

B2 : de NA+ à plaisant (émotions positives)

B3 : NA- (sérénité)



Un exemple sur données d'observation : affectivité et événements vécus

Les résultats

6. Plus-value de 3MPCA : la matrice noyau

b) la matrice de composantes pour les situations

Ev --	0.03	0.94
Ev -	0.41	0.20
Ev 0	0.46	0.12
Ev +	0.50	-0.01
Ev ++	0.61	-0.26

C1 : événements à impact « plutôt positif »

C2 : événements à impact fortement négatif

Un exemple sur données d'observation : affectivité et événements vécus

Les résultats

6. Plus-value de 3MPCA : la matrice noyau

c) la matrice noyau

	Événements à impact « plutôt positif »			Événements à impact fortement négatif		
	émotions négatives	émotions positives	sérénité	émotions négatives	émotions positives	sérénité
A1	1.31	-0.09	<b>-38.04</b>	-1.54	0.92	-16.89
A2	-0.94	-0.46	-2.18	<b>-36.08</b>	1.52	8.24
A3	0.76	<b>-46.01</b>	0.08	-0.43	-18.70	-0.85
A4	<b>-47.88</b>	1.12	2.51	-1.27	-4.19	-9.51

A4 : scores élevés → ressentent peu d'émotions négatives lors d'événements plutôt positifs

A3 : scores élevés → ressentent peu d'émotions positives lors d'événements plutôt positifs

A1 : scores élevés → ressentent peu de sérénité lors d'événements plutôt positifs

A2 : scores élevés → ressentent peu d'émotions négatives lors d'événements négatifs

Perspectives

Analyses 3MPCA de matrices individuelles (occasions, variables, situations)

Analyses 3MPCA de données expérimentales cognitives (ex : Stroop. A = individus ; B = blocs d'items ; C = type d'items [congruent, incongruent, neutre])

Analyses 3MPCA de la mesure des intérêts professionnels prenant en compte à la fois les activités et les contextes de travail

...