

Analyse de plans d'expériences en situation multivariée :

une contribution des méthodes multi-blocs.

G. Mazerolles

INRA, UMR « Sciences pour l'œnologie »,
F 34060 Montpellier, France.



Une situation fréquente : Plan d'expériences + méthode globale.

	Fac.A	Fac.B	← J variables →
exp1	niveau1	niveau1
exp2	niveau1	niveau1
exp3	niveau2	niveau1
exp4	niveau2	niveau1
exp5	niveau1	niveau2
exp6	niveau1	niveau2
exp7	niveau2	niveau2
exp8	niveau2	niveau2

Le tableau de données

P Exp. : A*B

IR, UV, RMN, LC/MS

➡ Analyse de variance multivariée

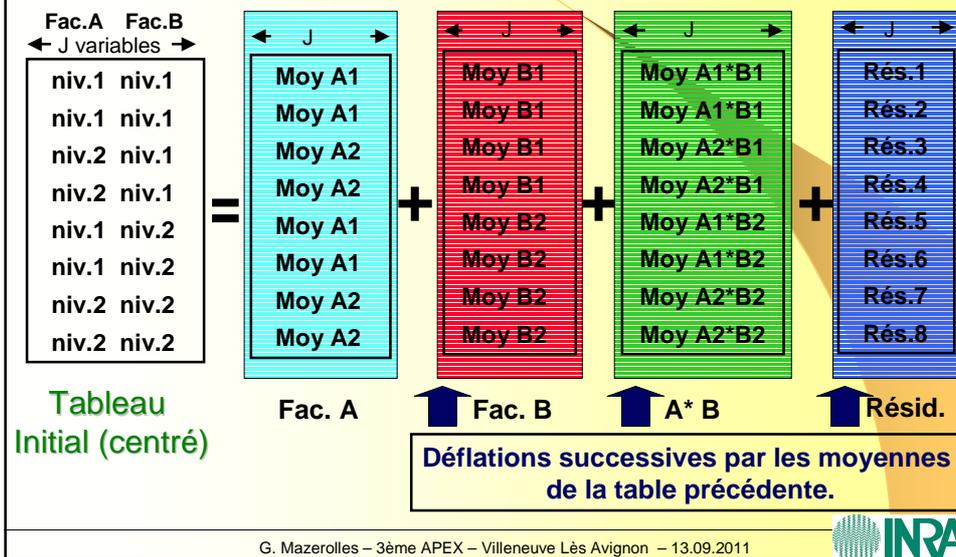
Solutions développées dans les « communautés » protéomique et métabolomique:

- Anova-Simultaneous Component Analysis (Smilde et al - 2005).
- Anova-PCA (Harrington et al - 2005).



**Anova-SCA, Anova-PCA et Analyse en Composante
Principales et Variables Instrumentales (ACP-VI):**

Répartir la variabilité initiale en contributions indépendantes:



Modèle d'analyse de la variance pour 2 facteurs croisés A et B (respectivement I et J modalités) n répétitions et 1 réponse y:

Décomposition de la somme des carrés totaux:

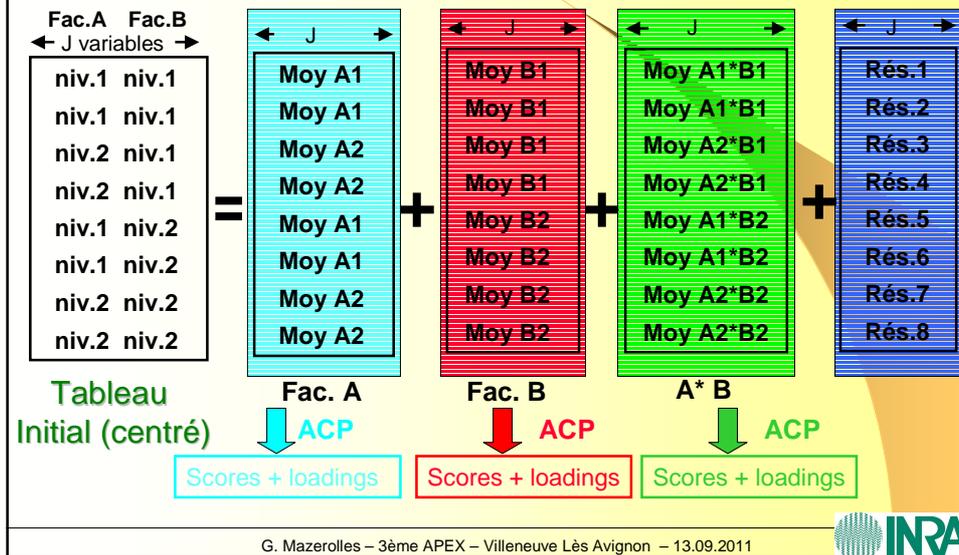
$$SS_T = SSA + SSB + SS(AB) + SSE$$

$$SSA = n * J \sum_{I=1}^I (\bar{y}_i - \bar{y})^2$$

$$SS(AB) = n \sum (y_{ij} - \bar{y}_i - \bar{y}_j - \bar{y})^2$$

$$SSE = \sum_{ijk} e_{ijk}^2$$

Anova-SCA, ACP-VI : Réaliser une Analyse en Composantes Principales (ACP) de chaque tableau: « synthèse des variations attribuables à chaque facteur du plan d'expériences »



ACP de chaque tableau associé à un des facteurs étudiés



Interprétation = Scores des modalités et loadings associés.

Validation :

- 1) Visualisation de sources de variabilité autre que le facteur étudié (résiduelle) : Dispersion autour des barycentres.
- 2) Bootstrap: estimation de la variabilité des moyennes.

Application :
Rôle de l'oxygène dans l'élevage de vins rosés : influence de la vinification et du type de bouchon.

➤ Plan d'expériences : 2 facteurs croisés:

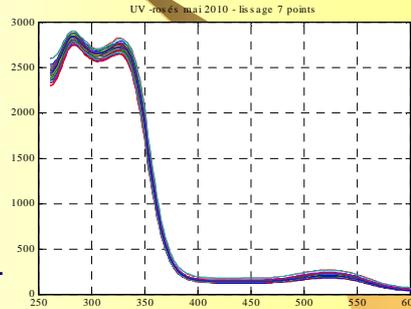
✓ Vinification : 4 niveaux:

	OXYDATION	REDUCTION
anti-ox. -	Rosé 1	Rosé 3
anti-ox. +	Rosé 2	Rosé 4

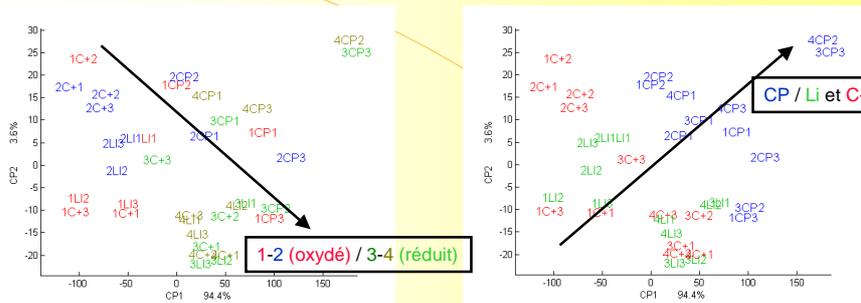
✓ type de bouchon : 3 niveaux de perméabilité à l'oxygène (CP, C+, Li).

✓ 3 bouteilles / vin * type de bouchon : 4 * 3 * 3 = 36 expériences

➤ Caractérisation par méthode rapide : spectrométrie UV



G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ACP sur les spectres des 36 vins: Composantes principales 1 et 2

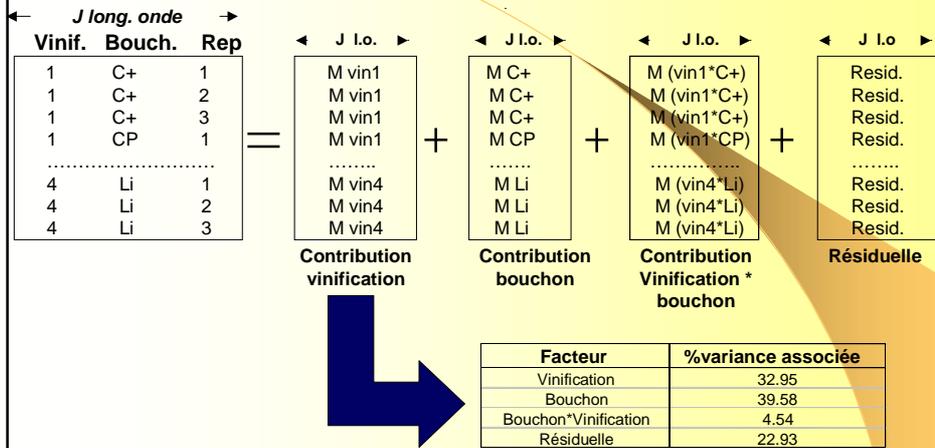
(codification individus : 1er caractère = type de vin; 2ème et 3ème caractères = type de bouchon; 4ème caractère = répétition).

Certaines modalités des facteurs du plan d'expériences mises en évidence.
 Interprétation? (pas de composante principale directement «liée» à un facteur).

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Le traitement Asca: Décomposition de la variabilité du tableau initial



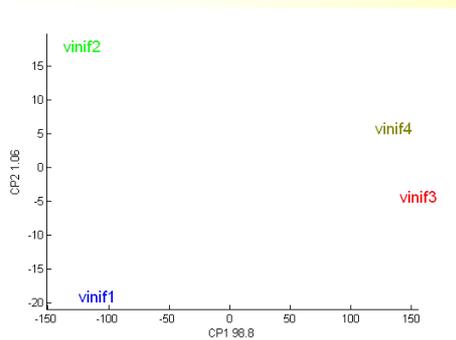
Gestion de la variabilité « individuelle » (intra-modalités):

- 1) ACP sur moyennes puis projections moyennes+résiduelles.
- 2) Bootstrap : comparaison de moyennes

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



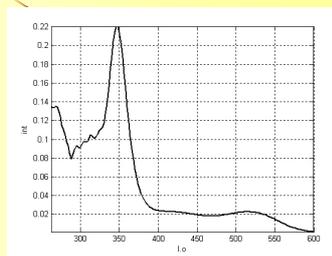
Asca : facteur vinification ACP des moyennes des 4 modalités.



Plan factoriel 1-2

OXYDATION	REDUCTION
anti-ox. - Rosé 1	Rosé 3
anti-ox. + Rosé 2	Rosé 4

Composante 1 : Conditions de vinification : opposition réductrices-oxydatives.
Composante 2 : rôle de l'antioxydant.

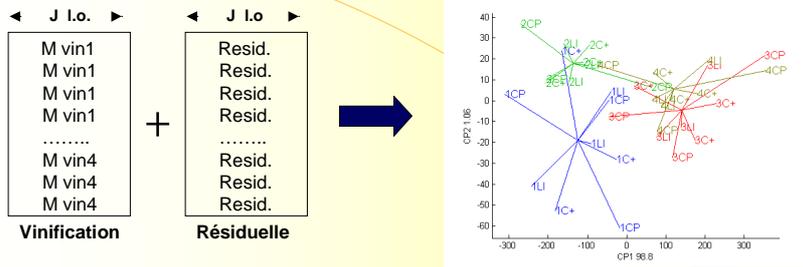


Vecteur propre associé à CP1

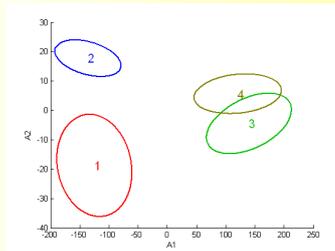
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Asca : test effet facteur vinification



Projections (moyennes + résidus) dans le référentiel moyennes vinif.

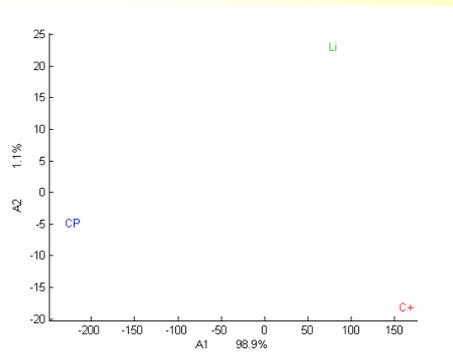


Bootstrap : Ellipsoïdes barycentres (5%)

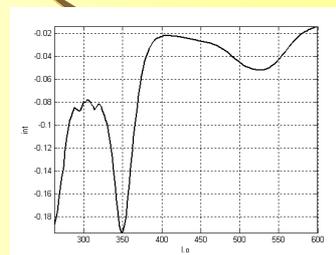
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Asca : facteur type de bouchon ACP des moyennes des 3 modalités.



Plan factoriel 1-2



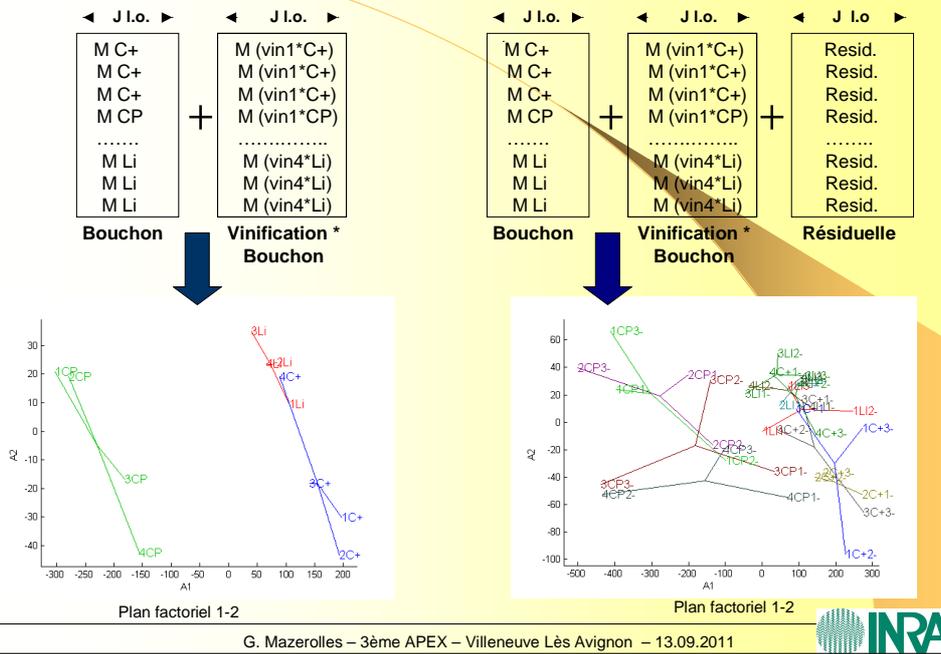
Vecteur propre associé à CP1

Composante 1 : Caractérisation d'un type de bouchon.
 CP : + de rouge (520 nm), + de jaune (350 nm) : vins + orangés
 Composante 2 : ?

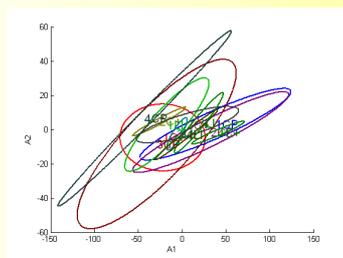
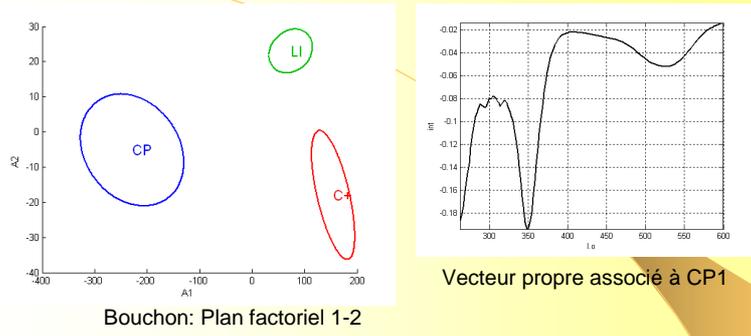
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Asca : type de bouchon : Interaction avec le facteur vinif.



Asca : test bouchon et interaction bouchon*vinif



Interaction Bouchon*vinif: Plan factoriel 1-2

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



En conclusion : ASCA = généralisation de l'analyse de variance.

Dans un cadre « opérationnel » : Élaboration des contributions en fonction des objectifs de l'étude.

Dans l'exemple précédent : test effet bouchon quelque soit vinification.

1) retrancher effet moyen vinification

2) Élaborer moyennes bouchons+interaction (12 moyennes) puis tester par rapport à la variabilité bouteille.

Bibliographie :

A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analysing designed metabolomics data, *Bioinformatics* 21 (2005) 3043.

J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA : analysis of multivariate data obtained from experimental design, *J. of Chemometrics* 19 (2005) 469.

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Application au contrôle anti-dopage : Dépistage de métabolites de la testostérone.

➤ Suivi réalisé sur 16 athlètes (9 reçoivent un placebo).

Semaine	1	2	3	4	5	6
Protocole Prise produit (7 athlètes)	 jours 01,03,05	 jours 08,10,12	 jours 15,17,19	 jours 22,24,26	✗	✗
Récolte Echantillons (ensemble des athlètes)	 tps0	✗	✗	 tps1 → tps4	✗	 tps5

➤ Dosage de 10 métabolites de la testostérone à chaque prélèvement (GC/MS) : TS, EpiS, DHEAS, AS, EtioS, TG, EpiG, DHEAG, AG, EioG.

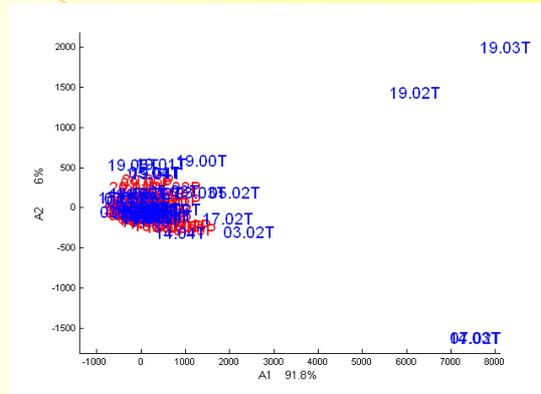
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Application au contrôle anti-dopage

← 10 variables →

A1	P	tps0
A2	P	tps0
A3	P	tps0
A4	P	tps0
.....		
.....		
.....		
A14	T	tps5
A15	T	tps5
A16	T	tps5



↓ Prélèvement
↓ Prise
↓ Numéro Athlète

Traitement par ACP

Le tableau de données

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Modèle d'analyse de la variance et tableaux correspondants:
2 facteurs croisés + 1 facteur hiérarchisé

Temps / **Traitement / Traitement*Temps** / Athlète / Athlète*Traitement* Temps

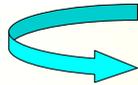
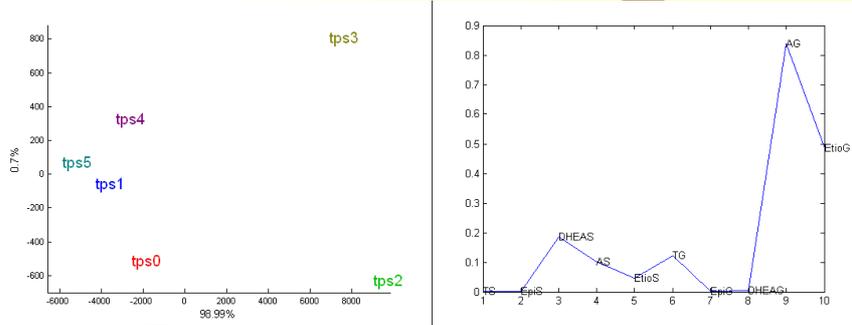
A1 P tps0	M tps0	M P*tps0	M A1	Res
A2 P tps0	M tps0	M P*tps0	M A2	Res
A3 P tps0	M tps0	M P*tps0	M A3	Res
A4 P tps0	M tps0	M P*tps0	M A4	Res
.....
.....
.....
A14 T tps5	M tps5	M T*tps5	M A14	Res
A15 T tps5	M tps5	M T*tps5	M A15	Res
A16 T tps5	M tps5	M T*tps5	M A16	Res

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ASCA - Facteur temps – Variance = 14.5 %

	Var explained (% / var tot)	2 PC demandés (% / var facteur)
Facteur temps	14.50	99
Facteur traitement + inter	22.40	99
Facteur athlète + inter	63.04	98

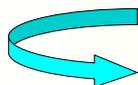
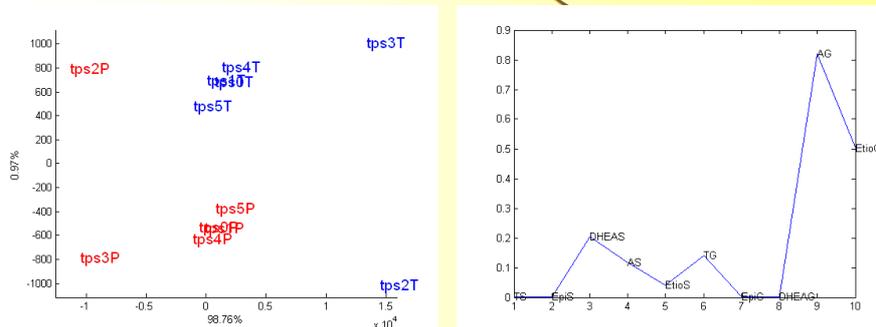


2 prélèvements (tps2 et tps 3) distingués sur dimension 1.

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ASCA - Facteur trait + trait*temps – Variance = 22.4 %

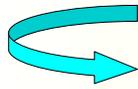
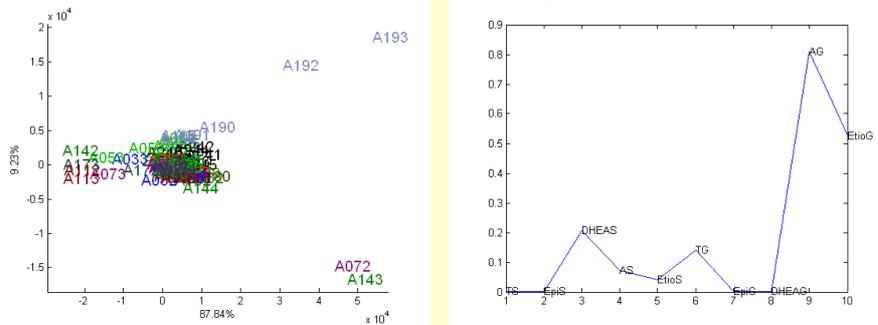


Traitement distingué sur dimension 1 pour tps2 et tps 3.

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ASCA - Facteur athlète + interaction – Variance = 63 %

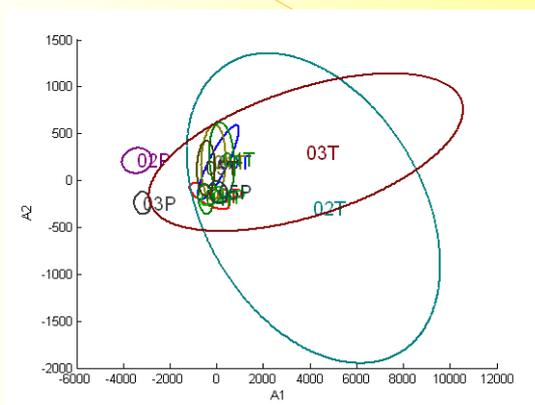


Athlètes distingués sur dimension 1 (interaction Athlète(traitement)*tps).

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ASCA - Test traitement+interaction*temps



(résiduelle athlète+ interaction temps*athlète(trait))

Les mêmes athlètes participent à l'ensemble des prélèvements à chaque réalisation bootstrap)

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ASCA: L'analyse de variance multivariée peut-elle être considérée seulement comme une extension de l'univariée?

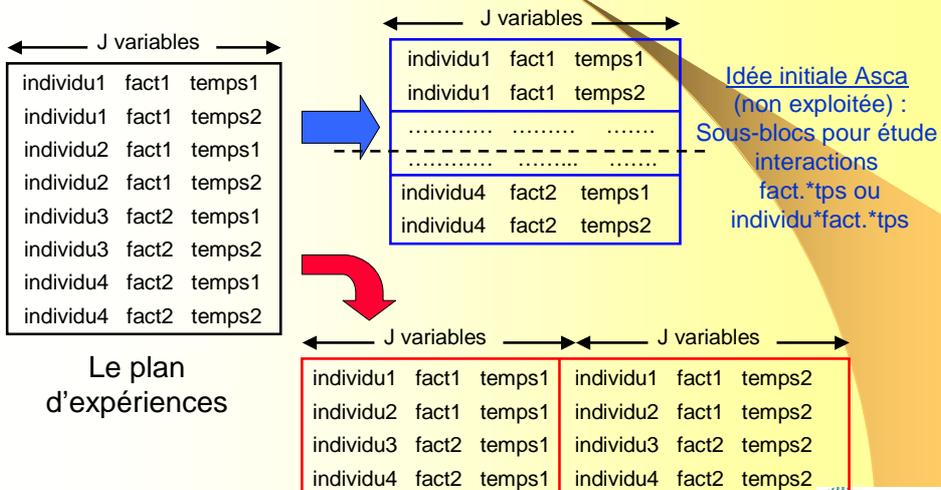
Les relations entre variables sont elles les mêmes pour chaque modalités de facteur ou combinaison de modalités?

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Plan d'expériences : Structuration en blocs suivant les modalités des facteurs du plan.

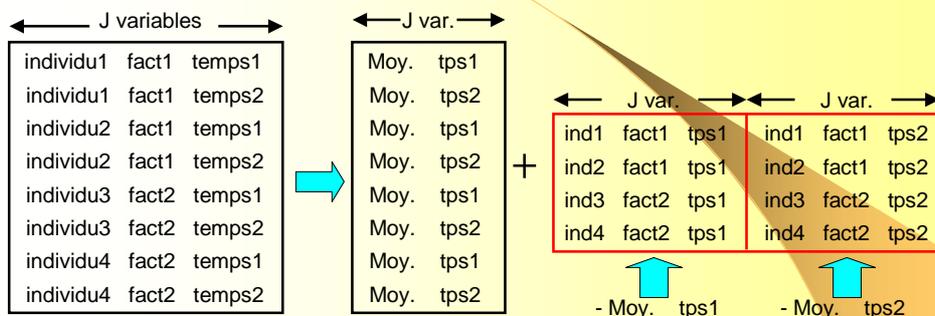
Un plus : prendre en compte la structure du jeu de données lors son analyse. **Méthodes multi-blocs**



G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



L'approche multi-bloc

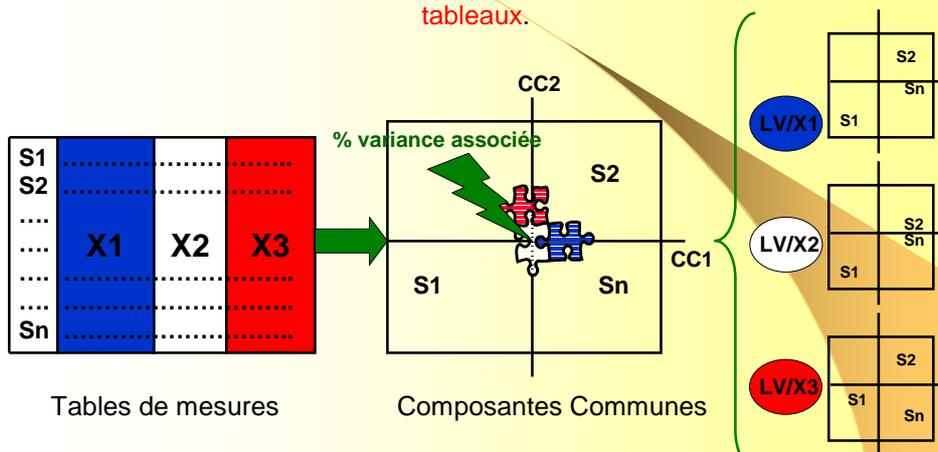


L'effet de fact. est il visible aux temps 1 et 2 et ce sur la base des mêmes variables?



Analyse en co inertie multiple (Acom) :

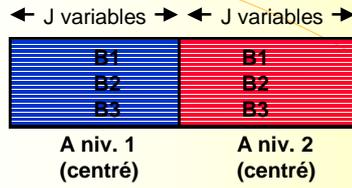
Définition d'un espace commun de représentation sur la base du lien entre tableaux.



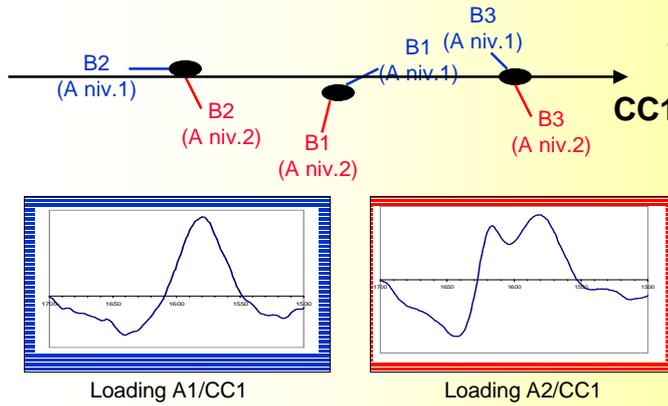
par table :
variables latentes +
loadings associés



Acom : les outils d'interprétation pour l'interaction A*B



NB: Le positionnement dans l'espace de représentation commune peut être vu comme le barycentre des représentations obtenues pour chaque table.



Suspicion d'interaction pour :
 - dispersion autour du barycentre
 - hétérogénéité des loadings

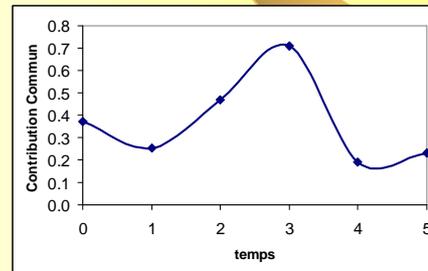
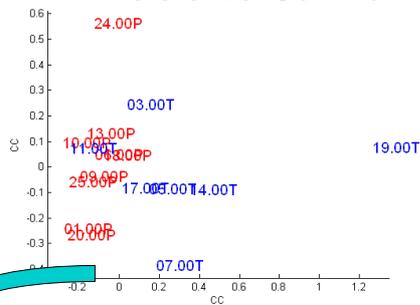
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Dopage : Analyse multi-blocs (6 blocs : 1 par temps)

← 10 var →							
A1 P tps0	A1 P tps1	A1 P tps1	A1 P tps5				
A2 P tps0	A2 P tps1	A2 P tps1	A2 P tps5				
.....
A15 T tps0	A15 T tps1	A15 T tps1	A15 T tps5				
A16 T tps0	A16 T tps1	A16 T tps1	A16 T tps5				

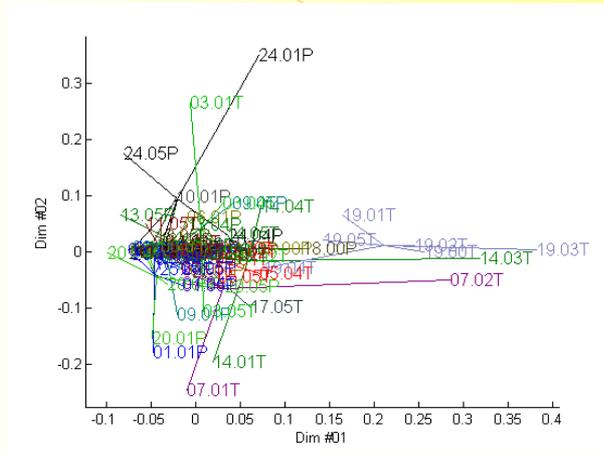
Référentiel Commun



Traitement distingué sur dimension 1 (temps 2 et 3).

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011

Dopage : Analyse multi-blocs

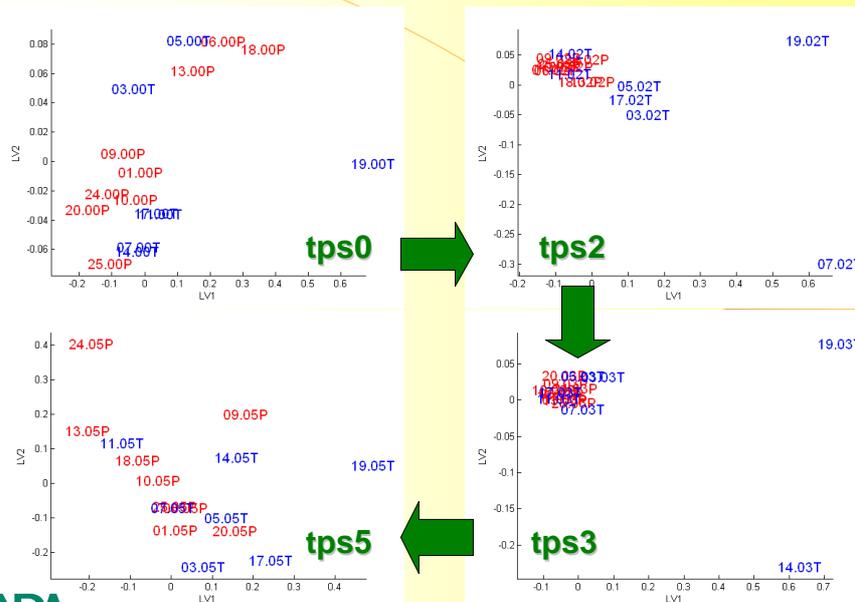


Traitement + interaction traitement*temps

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011

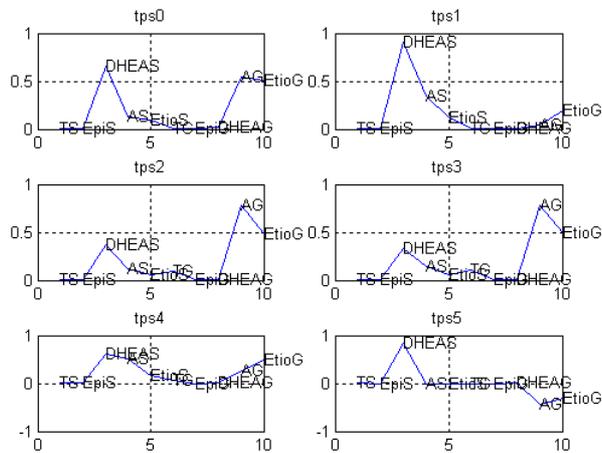


Analyse multi-blocs (6 blocs : 1 par temps) Contribution de chaque table au référentiel commun



G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011

Analyse multi-blocs (6 blocs : 1 par temps) Contribution des variables de chaque table



G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



**En conclusion : ACoM Complémentarité / Asca, ACP-VI:
possibilité de décrire en autorisant l'interaction à se
manifestar au niveau des variables.**

Bibliographie:

Mazerolles G., Boccard J., Hanafi M., Rudaz S., Analysis of experimental design with multivariate response: a contribution using multiblock techniques, *Chemometrics and Intelligent Laboratory System*, 106 (2011) 65-72

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Application :
Rôle de l'oxygène dans l'élevage de vins rosés : influence de la vinification et du type de bouchon.

➤ Plan d'expériences : 2 facteurs croisés:

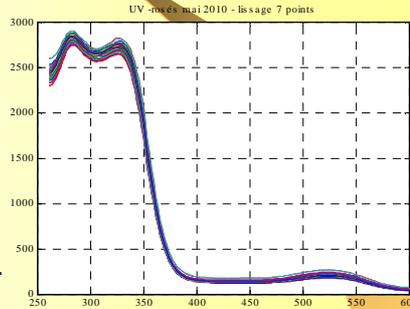
✓ Vinification : 4 niveaux:

	OXYDATION	REDUCTION
anti-ox. -	Rosé 1	Rosé 3
anti-ox. +	Rosé 2	Rosé 4

✓ type de bouchon : 3 niveaux de perméabilité à l'oxygène (CP, C+, Li).

✓ 3 bouteilles / vin*type de bouchon : $4 \times 3 \times 3 = 36$ expériences

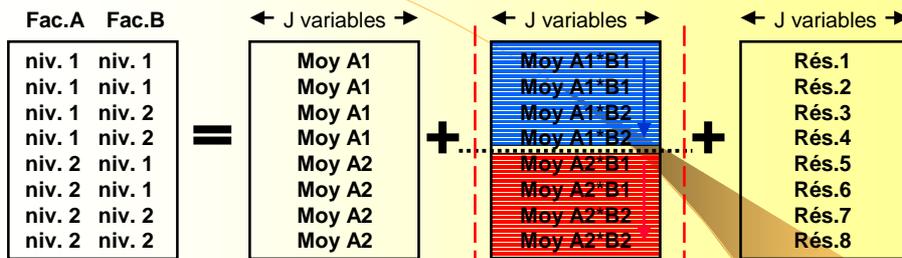
➤ Caractérisation par méthode rapide : spectrométrie UV



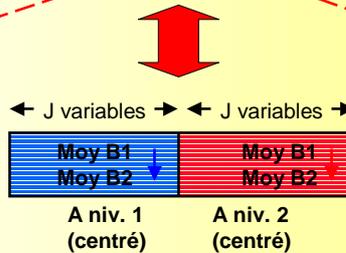
G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Idee initiale Asca (non exploitée) : Structuration des expériences en blocs suivant les modalités des facteurs du plan + Méthode multibloc (SCA).



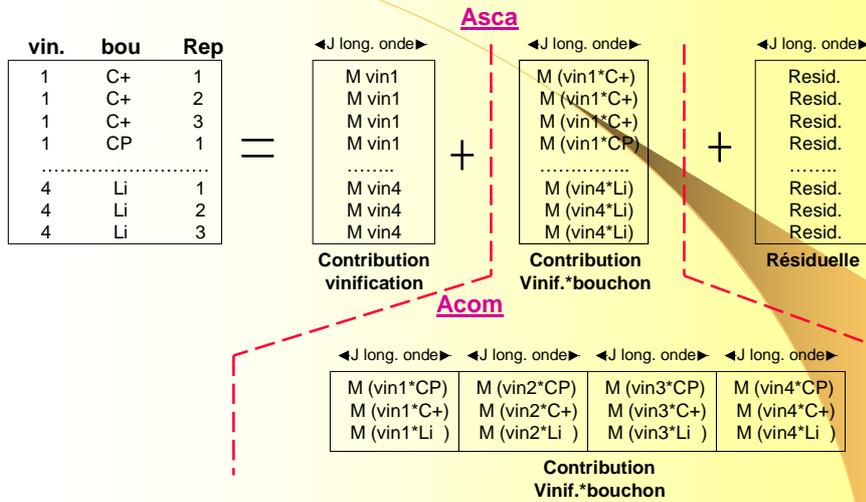
Alternative : Structuration et méthode multi bloc différentes (*Analyse en co inertie multiple*) pour appréhender une éventuelle modification du rôle des variables dans les différences observées entre modalités d'un facteur d'étude.



G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Les traitements réalisés par Asca et ACoM



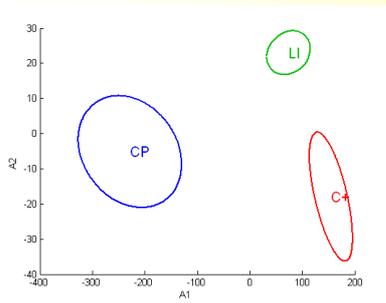
Gestion de la variabilité « individuelle » (intra-modalités):

Asca : ACP sur moyennes puis projections moyennes+résiduelles.

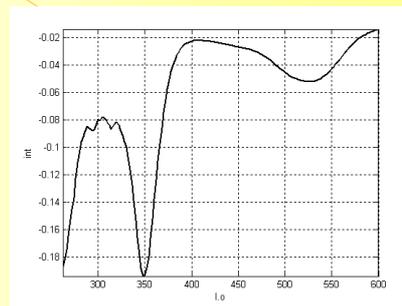
Acom : ACoM sur moyennes puis projections individus.



Asca : type de bouchon



Projection moyennes + résidus

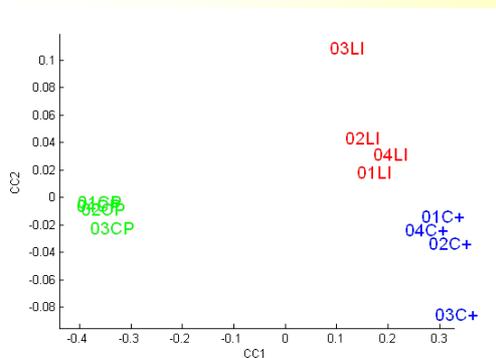


Loading associé à CP1

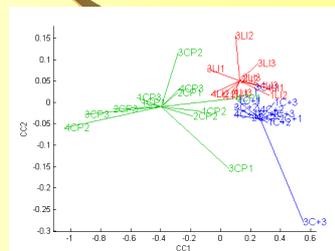
Interprétation : modalité CP : + de rouge (520 nm), + de jaune (350 nm) ➔ vins + orangés



ACoM : type de bouchon



Acom sur moyennes



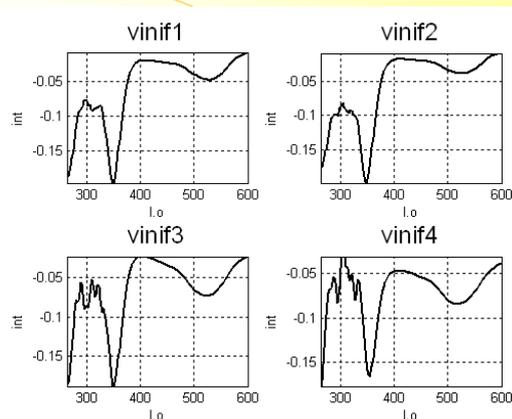
Projection individus

Composante 1 : Caractérisation type CP / Li et C+.
Composante 2 : ?

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



ACoM : type de bouchon



Loadings associés à la première composante pour les différentes tables

Interprétation : modalité CP : + de rouge (520 nm), + de jaune (350 nm)
vins + orangés (identique à Asca).



Mais différences induites par le type de bouchon dépendent du vin initial (CP pour les vins « conditions oxydatives » (1 et 2): rapport jaune/rouge + important).

G. Mazerolles – 3ème APEX – Villeneuve Lès Avignon – 13.09.2011



Conclusions et perspectives :

ASCA: approche pertinente, généralisation analyse de la variance.

➤ Intérêt / ACP : description plus aboutie des données (réponses à des questions).

Complémentarité ACoM/ Asca, Anova-Pca, ACP-VI: possibilité de décrire les interactions entre facteurs sans utiliser les mêmes variables.

Perspectives :

- plans déséquilibrés.
- Comparaison avec l'approche 3D (Parafac/Asca, Discrimination).

REMERCIEMENTS :

M. Hanafi, S.Ouertani (doctorante).

ONIRIS, UR «Sensométrie and Chimiométrie», F-44322 Nantes, France.

J. Boccard, S. Rudaz

Université de Genève, Université de Lausanne, Ecole des sciences pharmaceutiques, Suisse.

J. Wirth, A. Verbaere, V. Cheynier.

UMR SPO, INRA Montpellier, France.

M. Saugy, J.L. Veuthey, F. Badoud et E. Grata.

Université de Genève, Université de Lausanne, Ecole des sciences pharmaceutiques, Suisse
pour la fourniture des données analytiques relatives à la testostérone ayant permis d'illustrer ce travail.

La Société Nomacorc (Nomacorc SA, 4890 Thimister Clermont, Belgium) pour le financement de l'étude sur les bouchons.

D.J Vis, J.J. Jansen (Biosystems data Analysis group) et **D. Bertrand** (INRA, Nantes)
pour les codes Matlab nécessaires à la mise en œuvre de l'ASCA et de l'ACoM.