

Comparaison de deux tableaux homogènes de tests de toxicité

Résumé

La fiche porte sur la description des tableaux homogènes quantitatifs, en particulier de leur centrage. Le centrage multiplicatif est le plus adapté pour des tableaux homogènes de tests de toxicité. On illustre une analyse de co-inertie entre deux ACP avec un centrage multiplicatif commun aux deux tableaux. On met en évidence, au delà de la similitude apparente, une différence de fond entre résultats de tests *in vivo* et *in vitro*, différence systématique exprimée par le biais d'une analyse de co-structure.

Plan

| | |
|--|----|
| 1. Type de données..... | 3 |
| 2. Structure du tableau de tests <i>in vivo</i> | 3 |
| 2.1. Représentation des données..... | 3 |
| | 4 |
| 2.2. Modèles élémentaires | 5 |
| 2.3. Informations complémentaires..... | 10 |
| 3. Structure du tableau de tests <i>in vitro</i> | 12 |
| 4. Comparaison des tableaux..... | 16 |
| 4.1. Comparaison des niveaux de toxicité..... | 16 |
| 4.2. Comparaison des évolutions temporelles..... | 18 |
| Références | 24 |

D. Chessel & J. Devillers

1. Type de données

Les données traitées¹ sont présentes dans la fiche ToxiCornée de la pile ADE•Data :

| ADE-4•Data | | | |
|---|---------------------|----------------|--|
| Toxicologie sur la cornée (34-4)+(34-4) (F) | | | |
| toxicité de 34 p | | toxicité de 34 | |
| Liste des produit | | | |
| 1 0 0 0 | | 0 0 0 0 | |
| 0 0 0 0 | | 7 7 8 4 | |
| -1 -1 -1 | Sodium acetate, 10% | 5 3 | |
| 3 1 2 1 | Triacetin | 7 3 | |
| -1 0 0 2 | 1,2-Propanediol | 7 1 | |
| 3 4 4 4 | Tween 80 | 5 5 | |
| 15 5 4 3 | Dimethyl sulfoxide | 9 7 | |
| 9 6 1 1 | Glycerine | 7 37 51 | |
| 14 19 17 | Toluene | 1 46 53 | |
| 13 19 16 12 | | 13 15 19 24 | |
| 27 30 17 5 | | 21 21 28 35 | |
| 26 19 19 16 | | 21 24 29 32 | |
| 21 24 8 7 | | 16 16 17 12 | |

Deux tableaux ont 34 lignes et 4 colonnes. Ils sont appariés par les lignes mais le même nombre de colonnes dans les deux cas n'a pas de signification. Ils sont par contre tous les deux homogènes, c'est-à-dire que chaque cellule donne une mesure de la même grandeur.

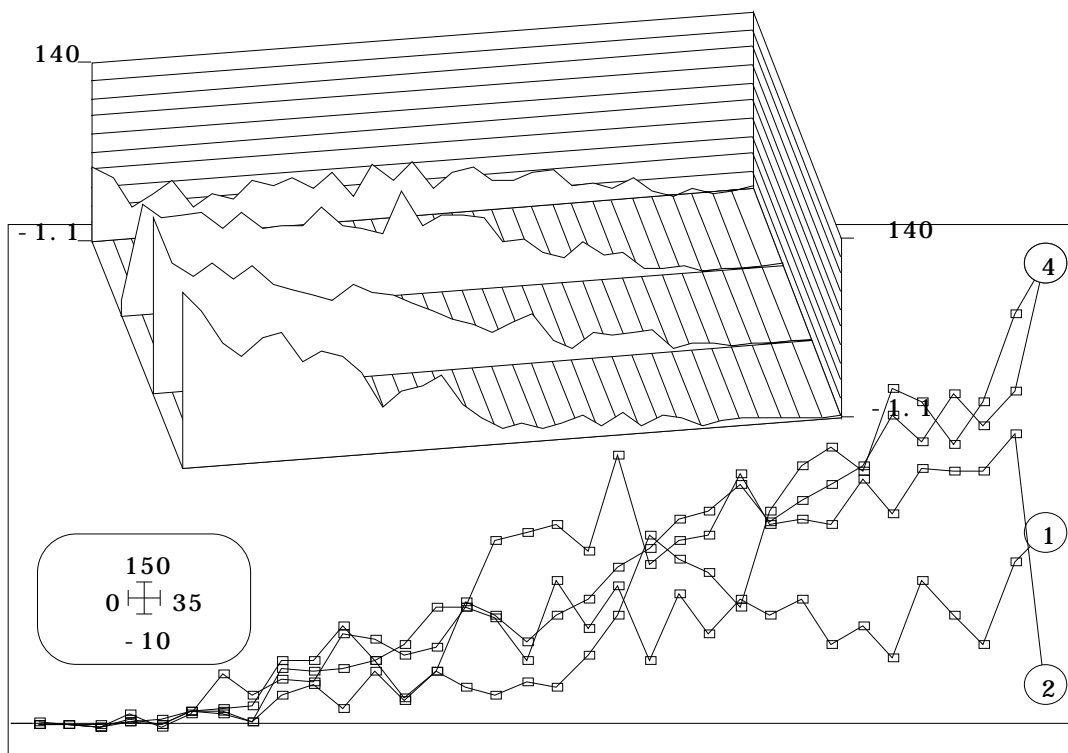


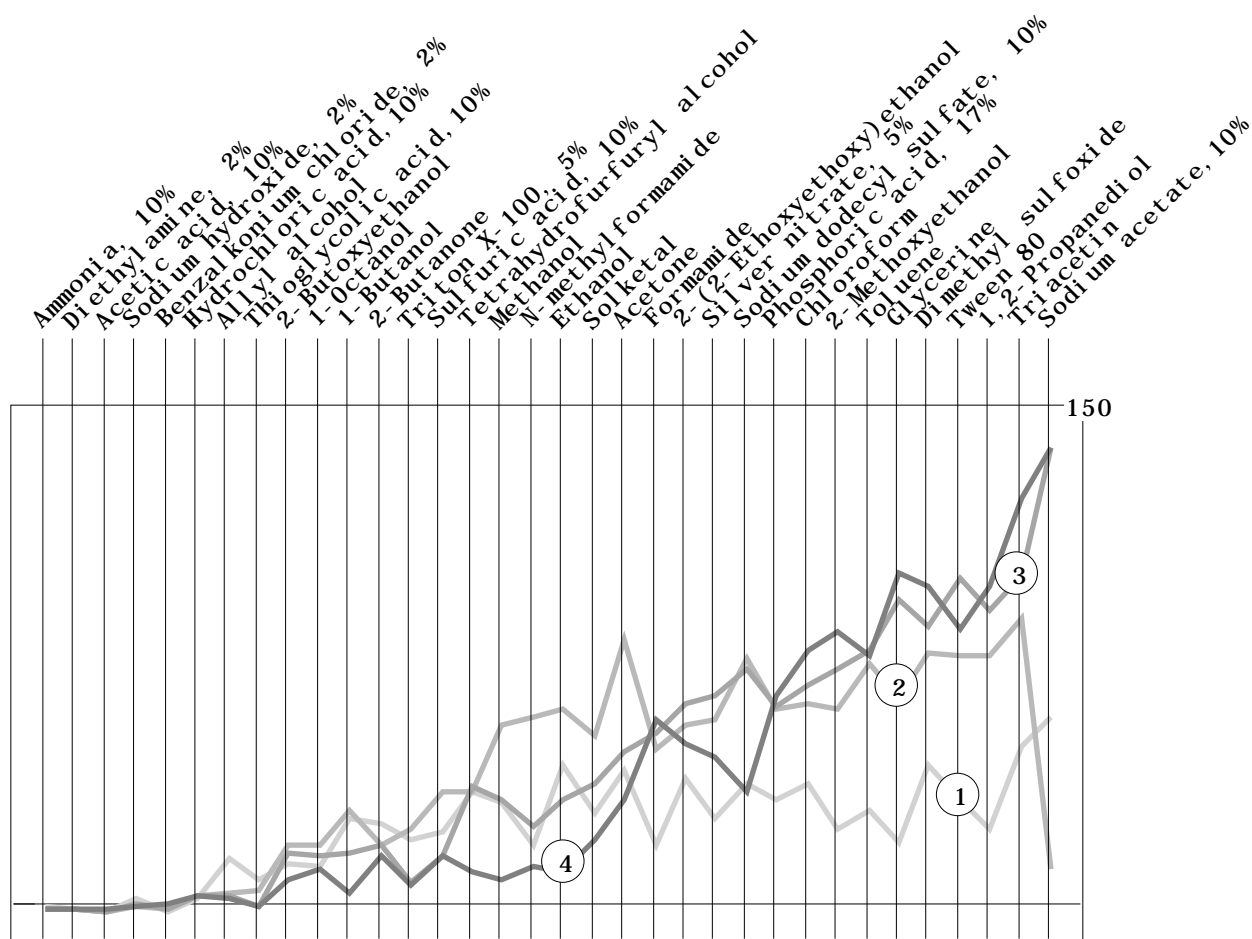
Figure 1 : Fichier Vivo vu par le module Table3D-Line (QuickBasic, version 3.7) et le module Curves : Lines (ADE-4). La toxicité dépend du produit (de 1 à 34 en abscisse) et de la durée (4 courbes).

Le champ de gauche donne un fichier texte Vivo.txt et un fichier binaire Vivo (34-4). On y trouve la toxicité (score calculé par rapport à un témoin) de 34 produits (code dans le champ du milieu : fichier Code_Prod) mesurée in vivo après 4h, 24 h, 48 h et 72 heures.

Le champ de droite donne un fichier texte Vitro.txt et un fichier binaire Vitro (34-4). On y trouve la toxicité (score calculé par rapport à un témoin) des mêmes 34 produits mesurée in vitro après 1/2 h, 1 h, 2 h et 4 heures. L'analyse statistique cherche à caractériser l'aptitude des tests in vitro à reproduire les toxicités mesurées in vivo (dans ce cas les coûts des expériences, liés en partie à leur durée, pourrait être sensiblement diminués). La résultats décrits dans cette fiche sont publiés dans ².

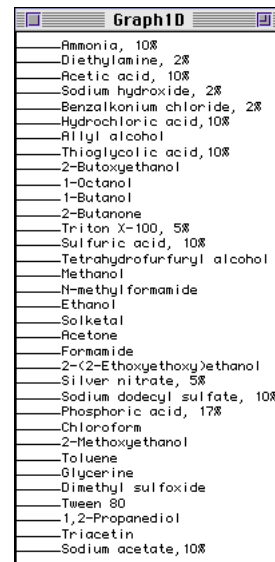
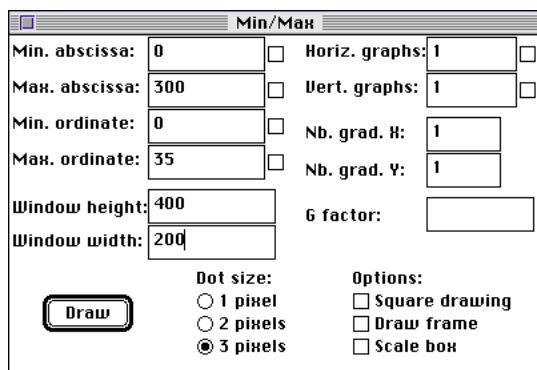
2. Structure du tableau de tests *in vivo*

2.1. Représentation des données



Pour préparer des légendes aux graphiques obtenus par Curves : Lines, utiliser GraphID : Labels :





Copier le résultat et coller directement dans une figure en cours. La complète compatibilité des modules graphiques entre eux permet toutes les illustrations avec un minimum de manipulation. Il suffira de conserver la dimension totale (ici 400 pixels verticalement pour la légende avec une rotation et 400 pixels horizontalement pour les courbes).

La structure du tableau est relativement simple. Les produits semblent classés par ordre de toxicité croissante (effet ligne) et les valeurs augmentent avec la durée d'exposition (effet colonne). On peut donc superposer les courbes lignes ou colonnes qui sont comparables. La meilleure représentation est obtenue avec Curves : Lines (Figure 1).

2.2. Modèles élémentaires

On cherchera d'abord un modèle élémentaire des données avec le module HTA. Conçu pour examiner le type de centrage à utiliser avant l'analyse d'un tableau homogène, le programme donne des résultats simples d'utilisation. Utiliser HTA : Edit inertia :



```
input file: vivo
Option 1 = No action (non centred PCA)          Inertia = 9551.68
Option 2 = Centred table (overall centred PCA)  Inertia = 4220.53
Option 3 = Centred (zero mean) columns         Inertia = 4014.18
Option 4 = Centred (zero mean) rows           Inertia = 1212.48
Option 5 = Additive model                      Inertia = 1006.13
Option 6 = Multiplicative model                Inertia = 687.313
```

L'objet de l'option est de donner une mesure de la pertinence de divers modèles simples, dits modèles évidents, du tableau. X est la matrice de départ, $x(i, j)$ la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes, p_i le poids associé à la ligne i , q_j le poids associé à la colonne j . Soit Z un

modèle du tableau, $m(i,j)$ la valeur du terme de la ligne i et de la colonne j , n le nombre de lignes et p le nombre de colonnes.

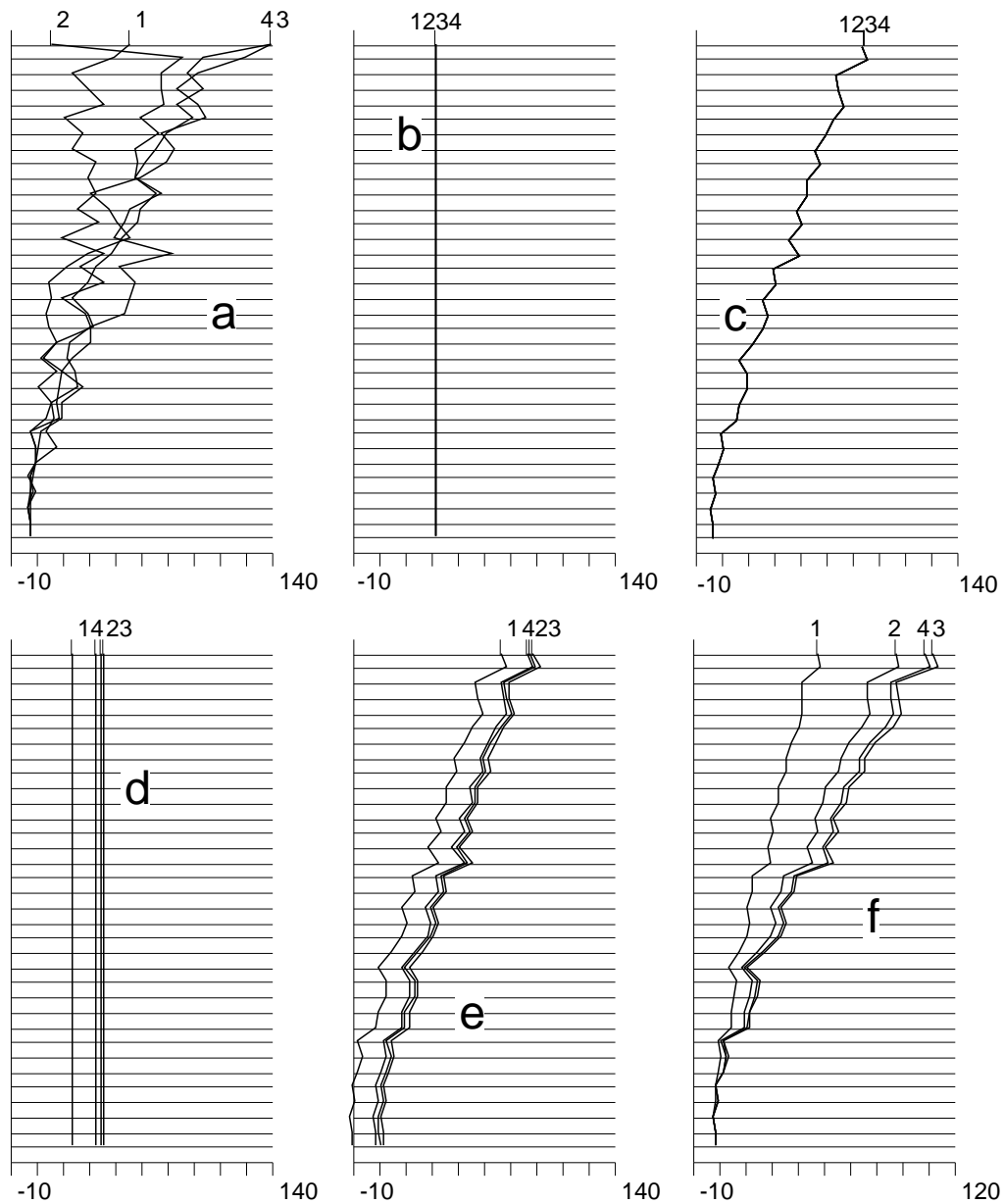


Figure 2 : Données et modèles élémentaires du tableau de tests in vivo. Lecture graphique du listing des options de HTA obtenue par **Table-Lines**. (L'équivalent avec Curves : Lines est dans la documentation du module HTA).

L'erreur globale commise avec ce modèle est l'inertie de l'ACP décentrée du tableau $\mathbf{X} - \mathbf{Z}$, soit :

$$E(\mathbf{Z}) = \sum_{ij} p_i q_j (x(i,j) - m(i,j))^2$$

Cette valeur est éditée pour les modèles suivants :

$$\begin{aligned}
m(i,j) &= 0 \quad (\text{modèle nul}) \\
m(i,j) &= m_{..} \quad (\text{modèle constant, estimation LS}) \\
m(i,j) &= m_{i.} \quad (\text{effet ligne simple, estimation LS}) \\
m(i,j) &= m_{.j} \quad (\text{effet colonne simple, estimation LS}) \\
m(i,j) &= m_{i.} + m_{.j} - m_{..} \quad (\text{effet additif, estimation LS}) \\
m(i,j) &= \alpha_i \beta_j \quad (\text{effet multiplicatif, estimation LS})
\end{aligned}$$

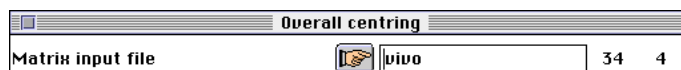
La notation LS est celle des moindres carrés (Least Squares). L'analyse du tableau $\mathbf{X} - \mathbf{Z}$, pour chacune des possibilités donne chacune des options du module. Le premier modèle élémentaire est celui d'une absence complète de structure dans le tableau. C'est le modèle nul par excellence qui s'écrit :

$$x(i,j) = a + e(i,j)$$

On notera $x(i,j)$ la valeur observée à la ligne i et à la colonne j , $m(i,j)$ la valeur prévue par un modèle et $e(i,j)$ l'erreur commise en utilisant ce modèle. p est le nombre de colonnes (ici 4) et n est le nombre de lignes (ici 34). L'erreur totale s'écrit :

$$E = \sum_{j=1}^p \sum_{i=1}^n e(i,j)^2 = \sum_{j=1}^p \sum_{i=1}^n (x(i,j) - m(i,j))^2$$

Pour ce premier modèle, la valeur de a qui minimise l'erreur est tout simplement la moyenne générale.



Pour chaque modèle, on conserve le fichier des valeurs prédites (mo) et celui des erreurs (ec). Le fichier --mo permet de représenter le modèle comme les données avec le même module.

```

-----
Homogeneous Table Analysis: option .cu
-----
Each row has 0.029412 weight (Sum = 1)
File vivo.cupl contains the row weights
It has 34 rows and 1 column

Each column has unit weight (Sum = 4)
File vivo.cupc contains the column weights
It has 34 rows and 1 column

Input file: vivo
Number of rows: 34, columns: 4
File vivo.cuta contains the centred (overall mean) table
It has 34 rows and 4 columns
File vivo.cumo contains the model
Model = overall mean
It has 34 rows and 4 columns
-----
Total inertia: 4220.53
-----

```

On ne s'intéresse pas à l'analyse du tableau centré mais simplement au modèle. On obtient la partie b de la figure 2. L'erreur totale est de 4220. L'ACP associée aux résidus de ce modèle est dite à centrage unique (cu).

Le second modèle élémentaire est celui d'un seul effet ligne (produit) qui s'écrit :

$$x(i, j) = a(i) + e(i, j)$$

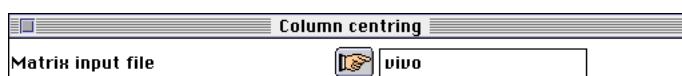


La meilleure estimation est obtenue avec la moyenne par ligne $a(i) = m_{i.}$. Le modèle est représenté par la figure 2-c et l'erreur est très sensiblement diminuée. L'ACP associée aux résidus de ce modèle est dite centrée par lignes (cl).

```
Input file: vivo
Number of rows: 34, columns: 4
File vivo.clta contains the (column) centred table
It has 34 rows and 4 columns
File vivo.clmo contains the model (Model = mean by row)
It has 34 rows and 4 columns
-----
Total inertia: 1212.48
-----
```

Le troisième modèle élémentaire est celui d'un seul effet colonne (durée) qui s'écrit :

$$x(i, j) = b(j) + e(i, j)$$

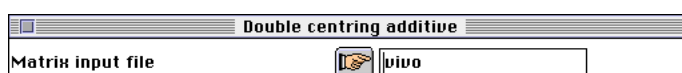


La meilleure estimation est obtenue avec la moyenne par colonne $b(j) = m_{.j}$. Le modèle est représenté par la figure 2-d et l'erreur est à nouveau très importante. L'ACP associée aux résidus de ce modèle est dite centrée par colonnes (cp, suffixe habituel de l'ACP centrée, implicitement par colonne-variable). C'est l'ACP traditionnelle dont l'usage ici serait une erreur (au demeurant habituelle dans la littérature).

```
File vivo.cpta contains the (column) centred table
It has 34 rows and 4 columns
File vivo.cpmo contains the model (Model = mean by column)
It has 34 rows and 4 columns
-----
Total inertia: 4014.18
-----
```

Le quatrième modèle élémentaire est celui d'un effet additif ligne+colonne (produit+durée) qui s'écrit :

$$x(i, j) = a(i) + b(j) + e(i, j)$$



La meilleure estimation est obtenue avec $m(i, j) = m_{i.} + m_{.j} - m_{..}$. Le modèle est représenté par la figure 2-e et l'erreur diminue à nouveau. L'ACP associée aux résidus de ce modèle est dite doublement centrée³ ou à centrage additif (cc).

```
Input file: vivo
```

```

Number of rows: 34, columns: 4
File vivo.ccta contains the doubly centred table
It has 34 rows and 4 columns
File vivo.ccmo contains the model
Model = mean by row + mean by column - overall mean
It has 34 rows and 4 columns
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: vivo.ccta
--- Number of rows: 34, columns: 4
-----
Total inertia: 1006.13
-----

```



Le cinquième modèle élémentaire est celui d'un effet multiplicatif ligne x colonne (produit x durée) qui s'écrit :

$$x(i, j) = a(i)b(j) + e(i, j)$$

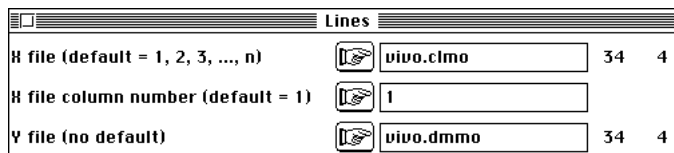
On utilise la meilleure estimation obtenue avec le premier facteur de l'ACP non centrée⁴ du tableau. Le modèle est représenté par la figure 2-f et l'erreur est minimale. L'ACP associée aux résidus de ce modèle est dite à centrage multiplicatif optimal (dm), qui donne les facteurs suivants de l'ACP non centrée(nc). On retrouve ici la dualité de l'analyse des données qui permet au choix de placer des éléments de décomposition dans la partie modèle ou dans la partie résidu⁵. Le listing indique où sont conservés les éléments de ce modèle multiplicatif optimal.

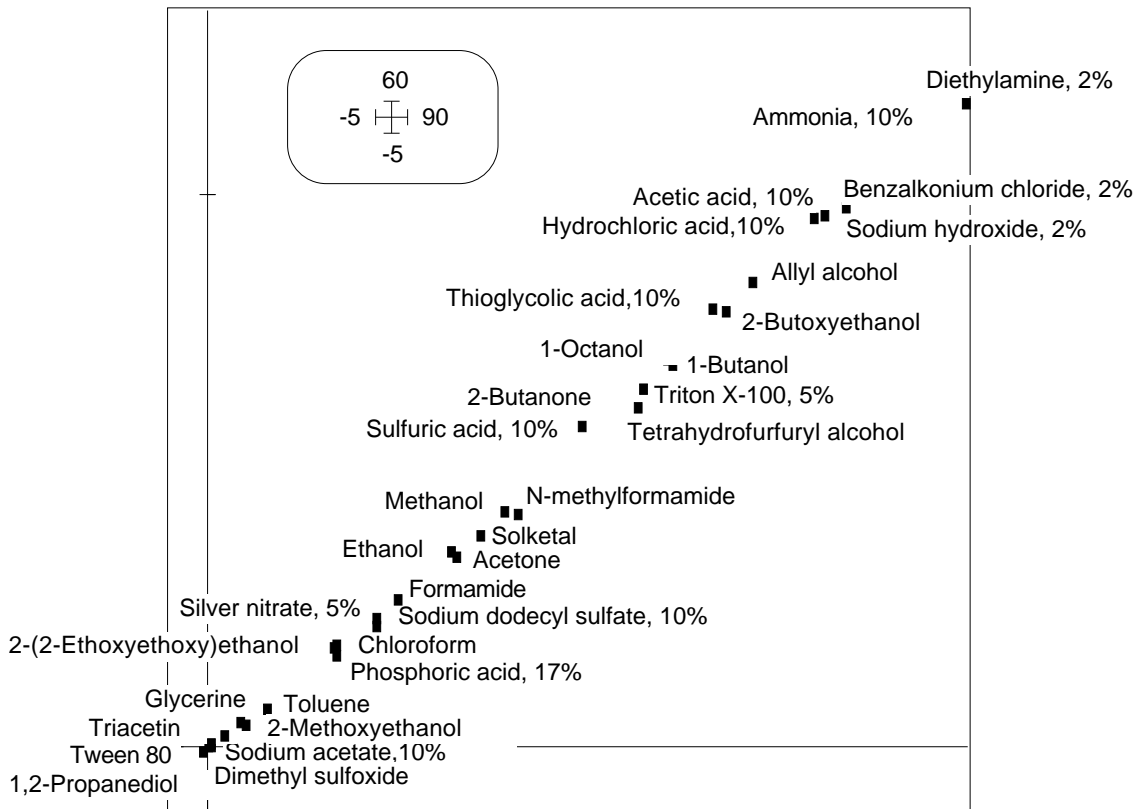
```

Input file: vivo
Number of rows: 34, columns: 4
File vivo.dmta contains the doubly centred table
It has 34 rows and 4 columns
File vivo.dmmo contains the model
Model = A(i)*B(j) optimal from the first axe of non centered PCA
It has 34 rows and 4 columns
-----
Total inertia: 687.313
-----

```

A une constante près multiplicative, les 4 colonnes de vivo.dmmo sont identiques et l'une quelconque d'entre elles (par exemple la première) donne l'estimation globale de la toxicité des produits vue par un modèle multiplicatif. Les 4 colonnes de vivo.clmo sont identiques et l'une quelconque d'entre elles (par exemple la première) donne l'estimation globale de la toxicité des mesures vue par un modèle additif. En fait les deux sont très voisines :

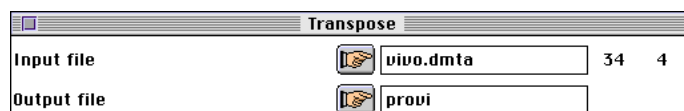




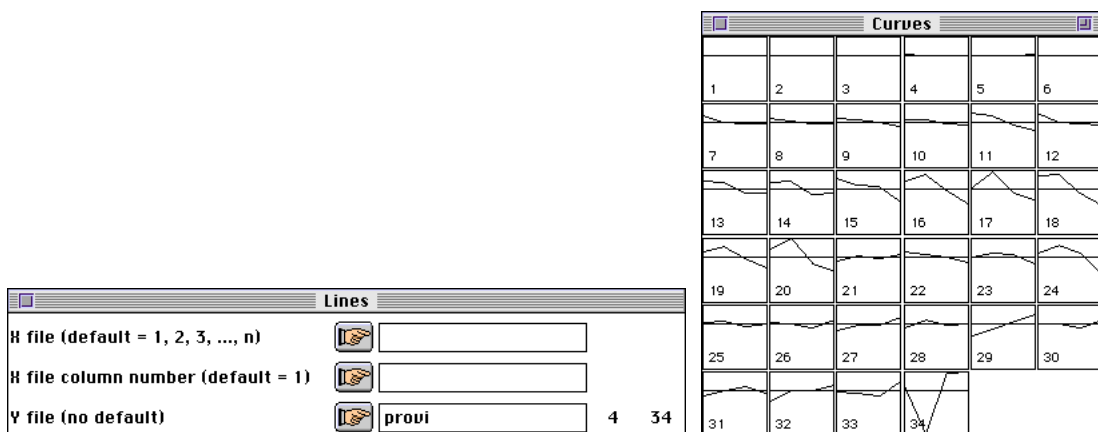
On conservera le schéma *Données = modèle évident + modèle caché + erreur* en cherchant dans les résidus une éventuelle typologie des produits.

2.3. Informations complémentaires.

Le fichier vivo.dmta donne les écarts du type *Données - modèle évident* dans lesquels on cherche le modèle caché. Représenter d'abord ces nouvelles mesures privées de l'effet principal, après transposition par FilesUtil : Transpose :



Préparer alors une collection de courbes avec Curves : Lines :



Tracer le plan factoriel de l'analyse à double centrage multiplicatif :



Illustrer pour obtenir la figure 3.

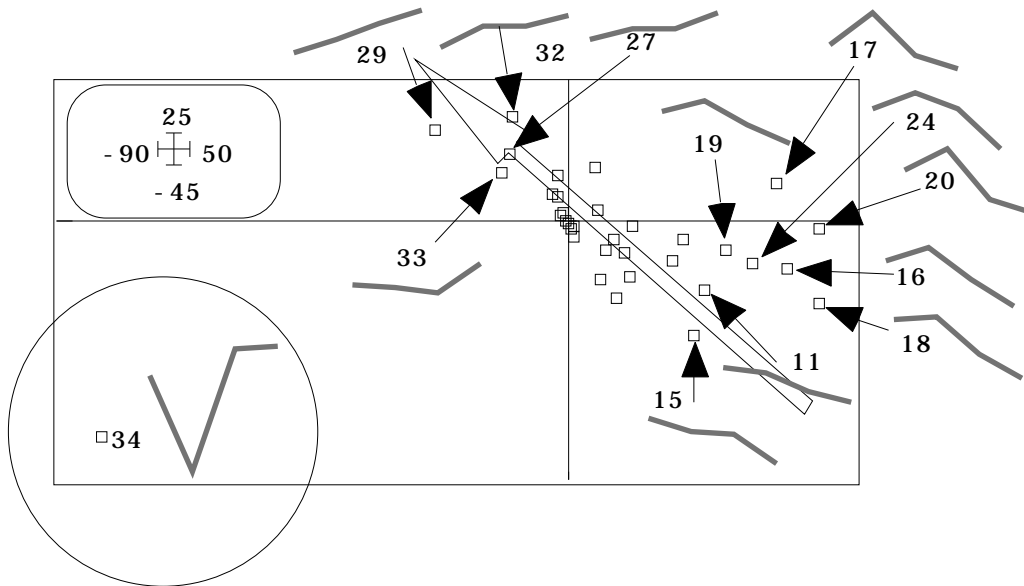


Figure 3 : dépouillement graphique d'un plan d'ACP à double centrage multiplicatif.

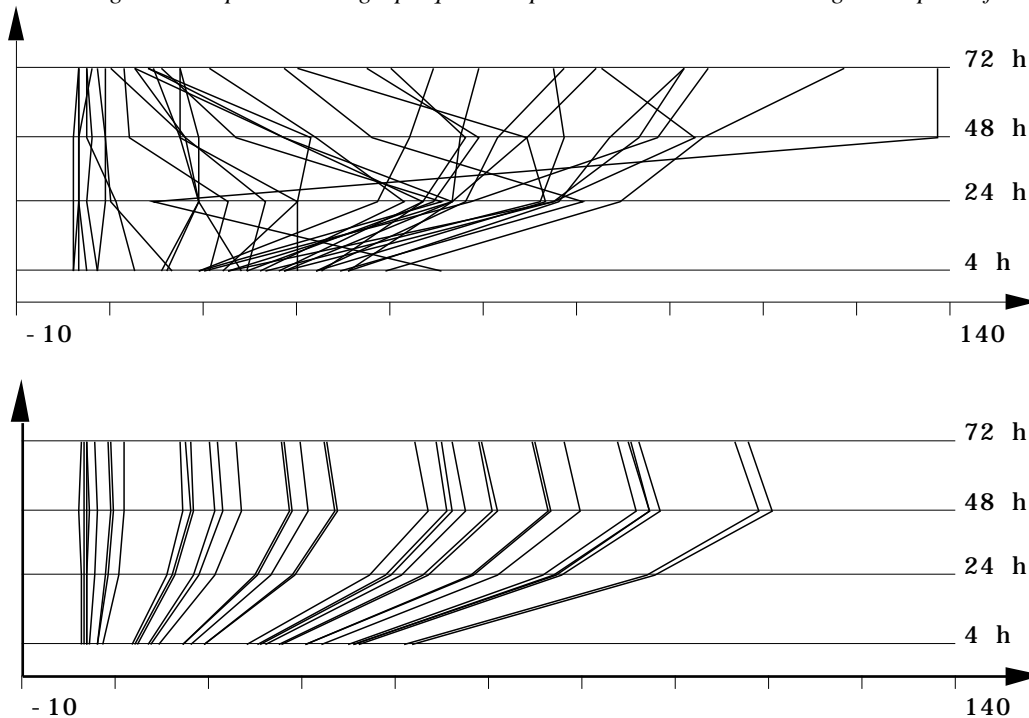
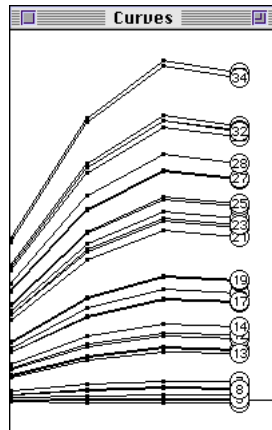


Figure 4 : En haut, données brutes (34 courbes de 4 points), en bas modèle multiplicatif optimal (observation = toxicité * durée). Exécution de **Table-Lines** sur Vivo et Vivo.dmmo.

Le plan factoriel s'interprètent sans difficulté. On obtient une typologie des formes des courbes de réponses indépendante du niveau de toxicité. On note le caractère aberrant de la seconde valeur pour Ammonia, 10% qui l'oppose aux courbes avec un maximum dans cette position. Sont isolés les acides qui présente un maximum d'irritation à 72 heures aux autres produits qui en général ont un niveau de réponse

optimum à 24 ou 48 h. La forme du plan est largement déterminée par la valeur aberrante soulignée par l'isolement du point sur la carte factorielle. La figure 3 exprime l'intégralité de l'écart qui existe entre les données brutes et le modèle multiplicatif, écart clairement exprimé dans la figure 4.

On peut obtenir le même graphique avec Curves : Lines en transposant le tableau (courbes en colonnes) et en utilisant la superposition des courbes.



3. Structure du tableau de tests *in vitro*

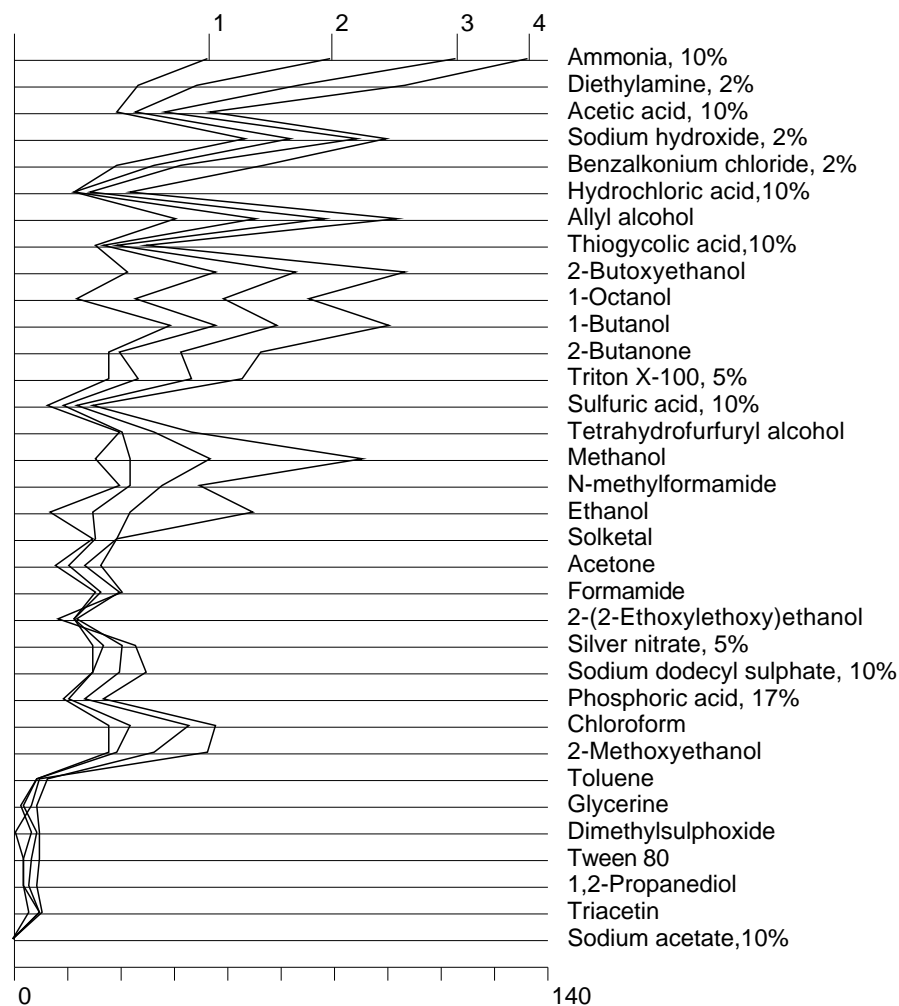


Figure 5 : Représentation des données brutes du tableau Vitro (comparer avec la figure 1).

La figure 5 reproduit pour le tableau Vitro la représentation obtenue dans la figure 1 pour le tableau Vivo. La comparaison des deux figures impose une idée de communauté de l'ordination des toxicités et une idée de différences de structures internes, le second étant plus régulier pour un produit donné et plus irréguliers entre produits. Du listing de l'exécution de HTA : Edit inertia on retiendra :

```
cu : Total inertia : 2999 (4221)
cl : Total inertia : 771 (1212)
cp : Total inertia : 2605 (4014)
cc : Total inertia : 377 (1066)
dm : Total inertia : 94 (687)
```

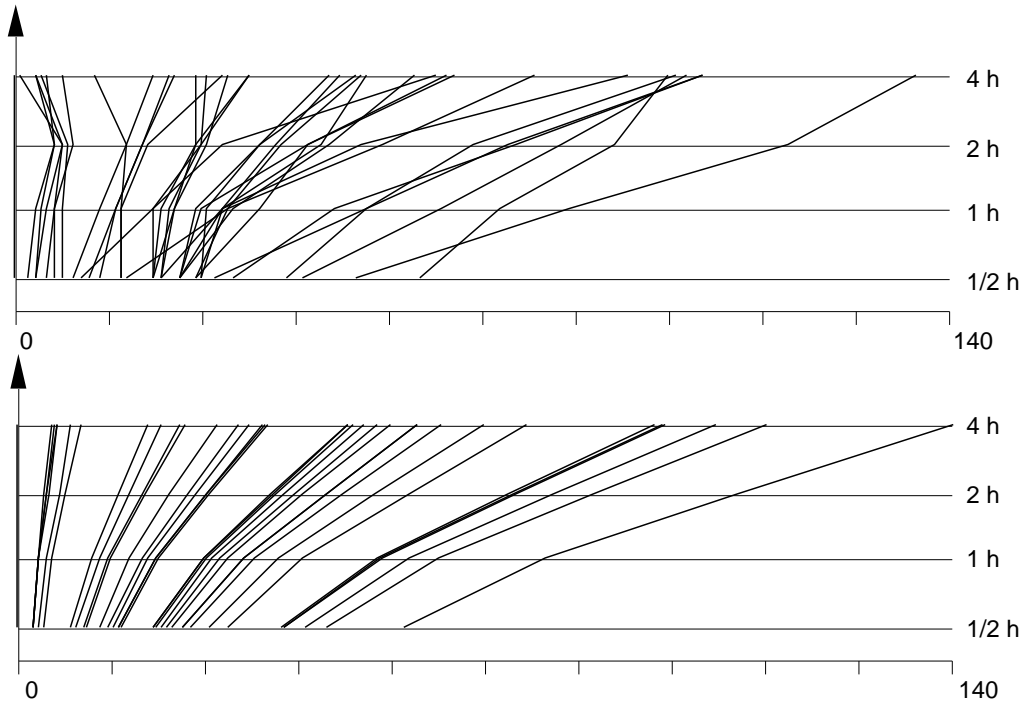


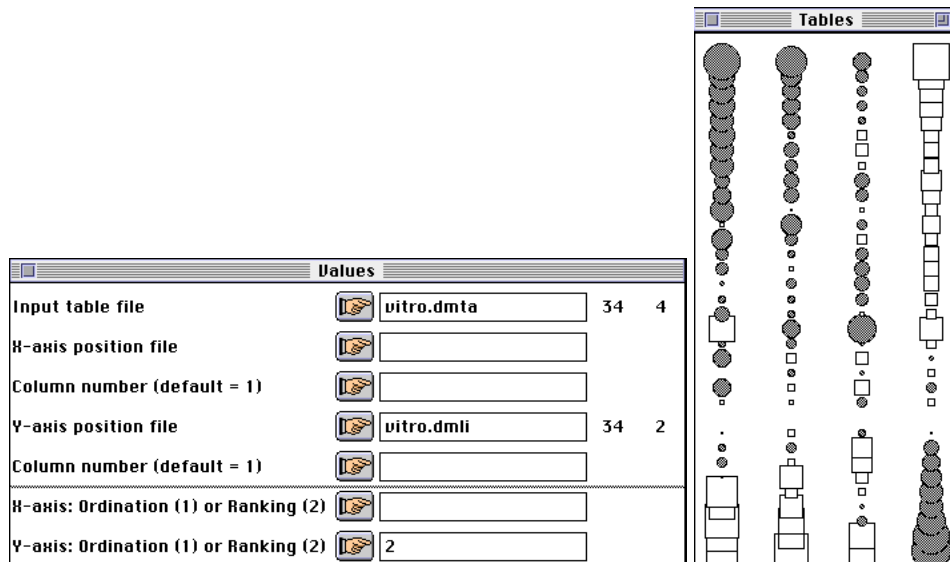
Figure 6 : En haut, données brutes (34 courbes de 4 points), en bas modèle multiplicatif optimal (observation = toxicité * durée). Exécution de **Table-Lines** sur Vitro et Vitro.dmmo. (comparer avec la figure 4).

On a reproduit entre parenthèses les valeurs obtenues pour le tableau Vivo. En partant d'une variabilité totale égale au trois quart de la première, on arrive à une erreur autour du modèle multiplicatif sept fois plus faible. Très sensiblement, les mesures *in vitro* sont beaucoup plus reproductibles que les mesures *in vivo*, ce qui semble assez normal. La figure 6 reproduit pour le tableau Vitro la représentation obtenue dans la figure 4 pour le tableau Vivo. L'optimum atteint pour 48 heures dans l'un n'a pas de correspondant dans l'autre. Au contraire c'est très généralement en 4 heures qu'on obtient le maximum et le modèle commun aux produits est une courbe monotone.

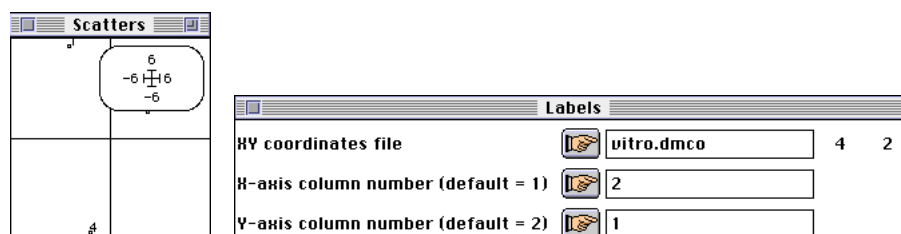
L'analyse à double centrage multiplicatif est exécutée :



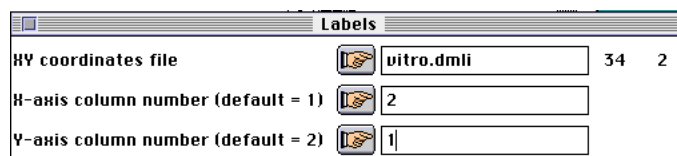
On utilise le module Tables : Values sur le tableau vivo.dmta (écarts aux prévisions du modèle multiplicatif) réorganisé par le facteur 1 de l'analyse (colonne 1 de vivo.dmti) pour lire très facilement la nature du facteur 1 (83% de l'inertie).



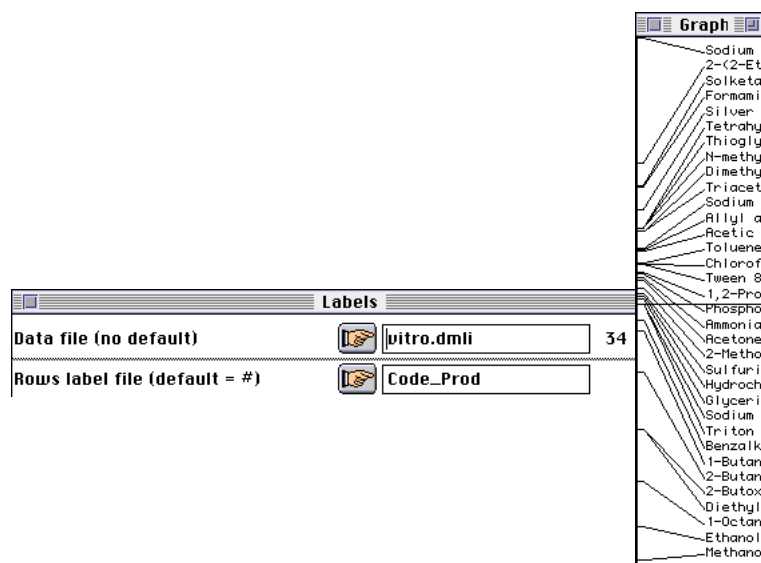
On utilise Scatters : Labels pour obtenir la carte des colonnes :



On utilise Scatters : Labels pour obtenir la carte des lignes en portant le facteur 1 verticalement pour faciliter la construction de la figure de synthèse :



On utilise Graph1D : Labels pour construire la légende du facteur 1 :



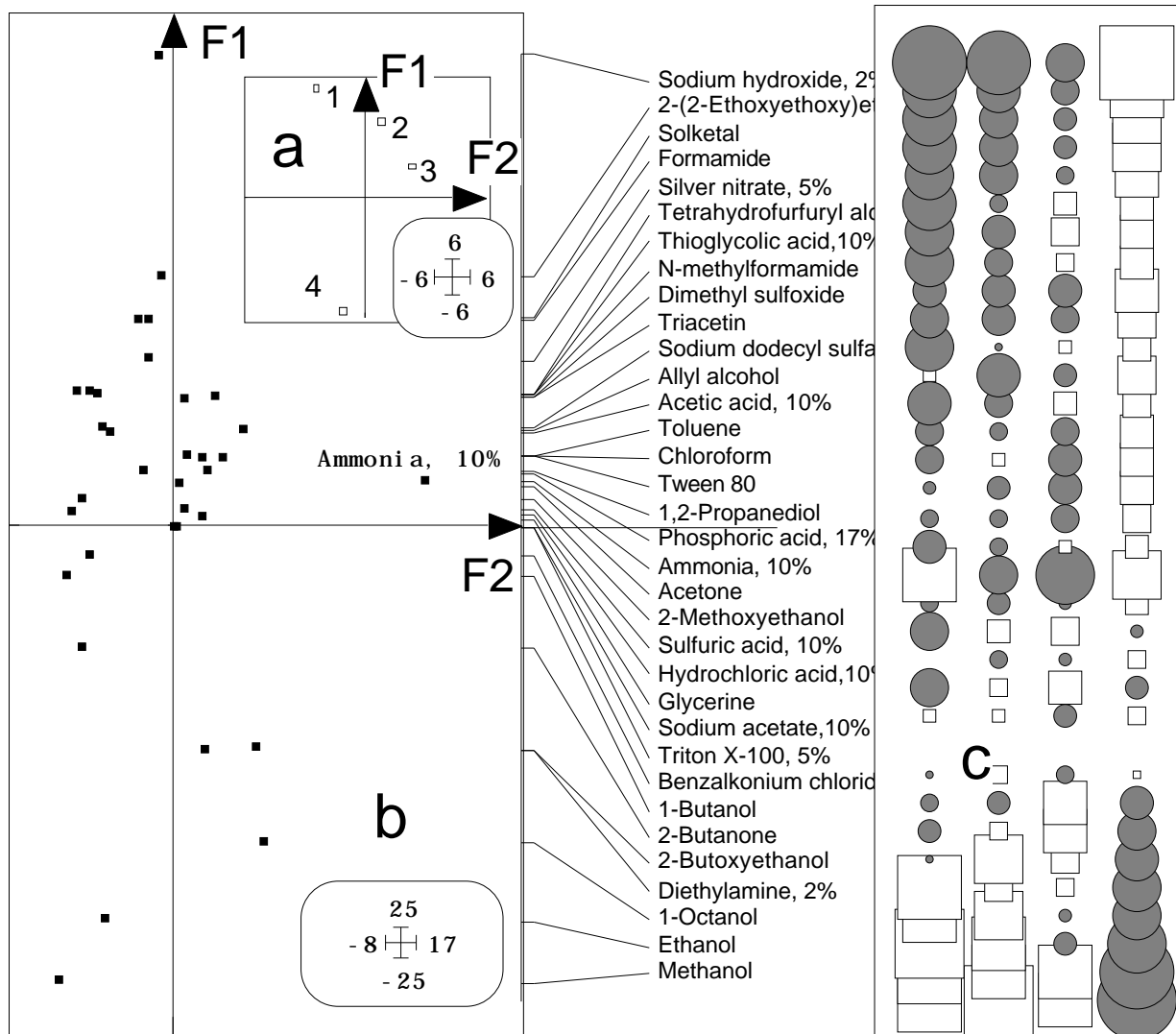


Figure 7: Typologie de produits par ACP doublement centrée. a - carte des colonnes (durée). b - carte des lignes (produits). c- Représentation des données traitées ordonnées par le score sur l'axe 1.

Le total est assemblé dans un logiciel de dessin. Noter l'efficacité des réécritures organisées des tableaux centrés⁶. On retrouve Ammonia, 10% comme point particulier et une ordination des produits depuis Sodium hydroxide, 2% provoquant une réaction rapide (maximum relatif en une 1/2 h) à Ethanol et Méthanol provoquant une réaction lente (maxima relatifs en 4 h). Niveau et vitesse de toxicité sont deux paramètres qui résument le second tableau. Le produit Ammonia, 10% est complètement original dans les deux séries et sera enlevé pour faciliter la discussion finale.

4. Comparaison des tableaux

4.1. Comparaison des niveaux de toxicité

Observons d'abord la relation entre niveaux de toxicités des produits dans chacun des types de mesures. Pris séparément, chaque tableau fournit un indice de toxicité des produits soit par la moyenne soit par le code numérique qui minimise l'erreur du

modèle $m(i,j) = a(i)b(j)$ (colonne 3 des mêmes fichiers). Précisons que ce type de modèle n'est pas défini de manière unique, puisque un modèle du type :

$$m(i,j) = ka(i) \frac{b(j)}{k}$$

a les mêmes performances que le précédent.

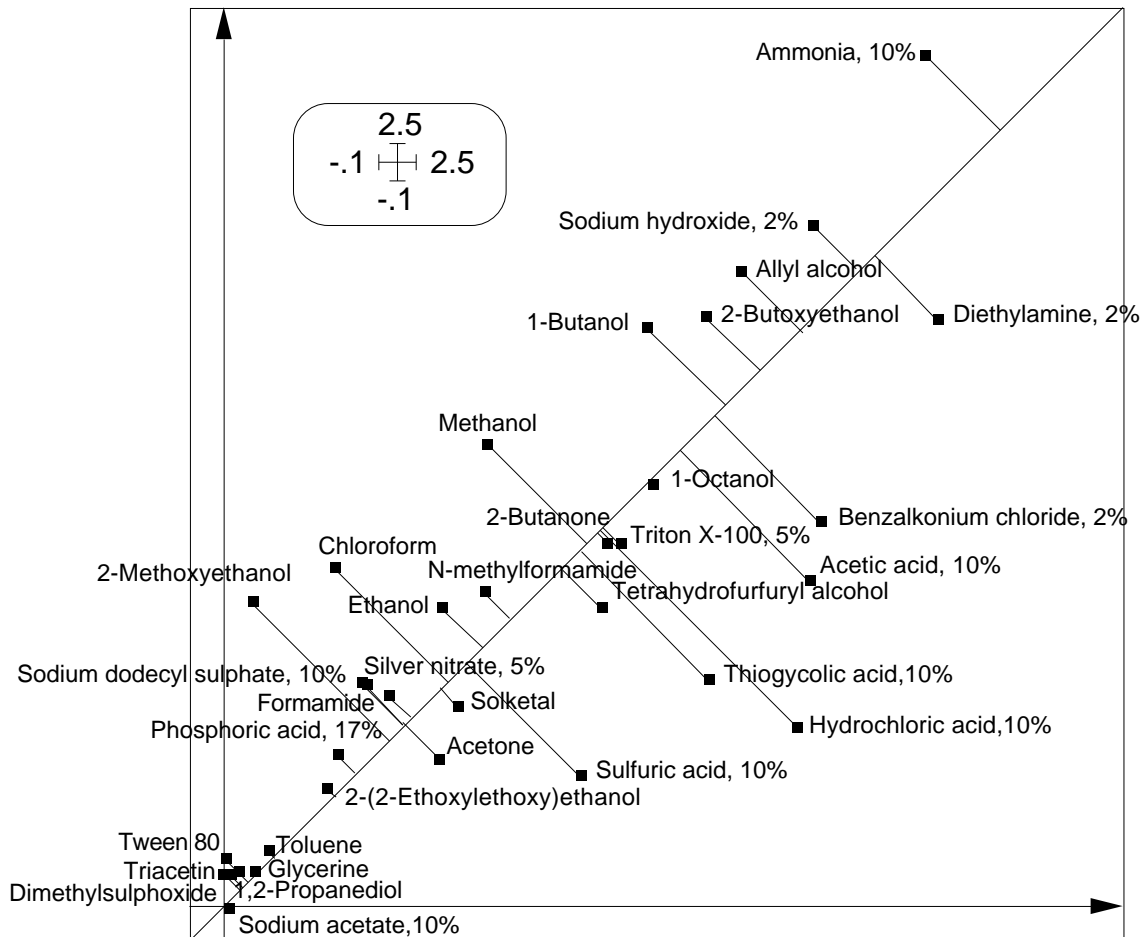


Figure 8 : En abscisse, niveau de toxicité dans le tableau Vivo. En ordonnée, niveau de toxicité dans le tableau Vitro. Utilisation des paramètres du modèle multiplicatif optimal. C'est parce que les codes sont standardisés que la droite $y=x$ minimise la somme des carrés des écarts aux projections.

Si on impose à $a(i)$ d'être de moyenne de carrés unité (norme 1 pour la pondération uniforme) et à $b(j)$ d'être de somme de carrés unité (norme 1 pour la pondération unitaire), on a alors le modèle unique déjà précité de la forme :

$$m(i,j) = \sqrt{\lambda} a(i)b(j) \quad \frac{1}{n} \sum_{i=1}^n a(i)^2 = 1 \quad \sum_{j=1}^p b(j)^2 = 1$$

Les codes $a(i)$ sont alors à la même échelle pour les deux tableaux. Pour récupérer ces vecteurs on peut faire les ACP non centrées des fichiers (HTA : No centring), ne garder qu'un facteur et normer la composante principale (DDUtil : Add normed scores) :

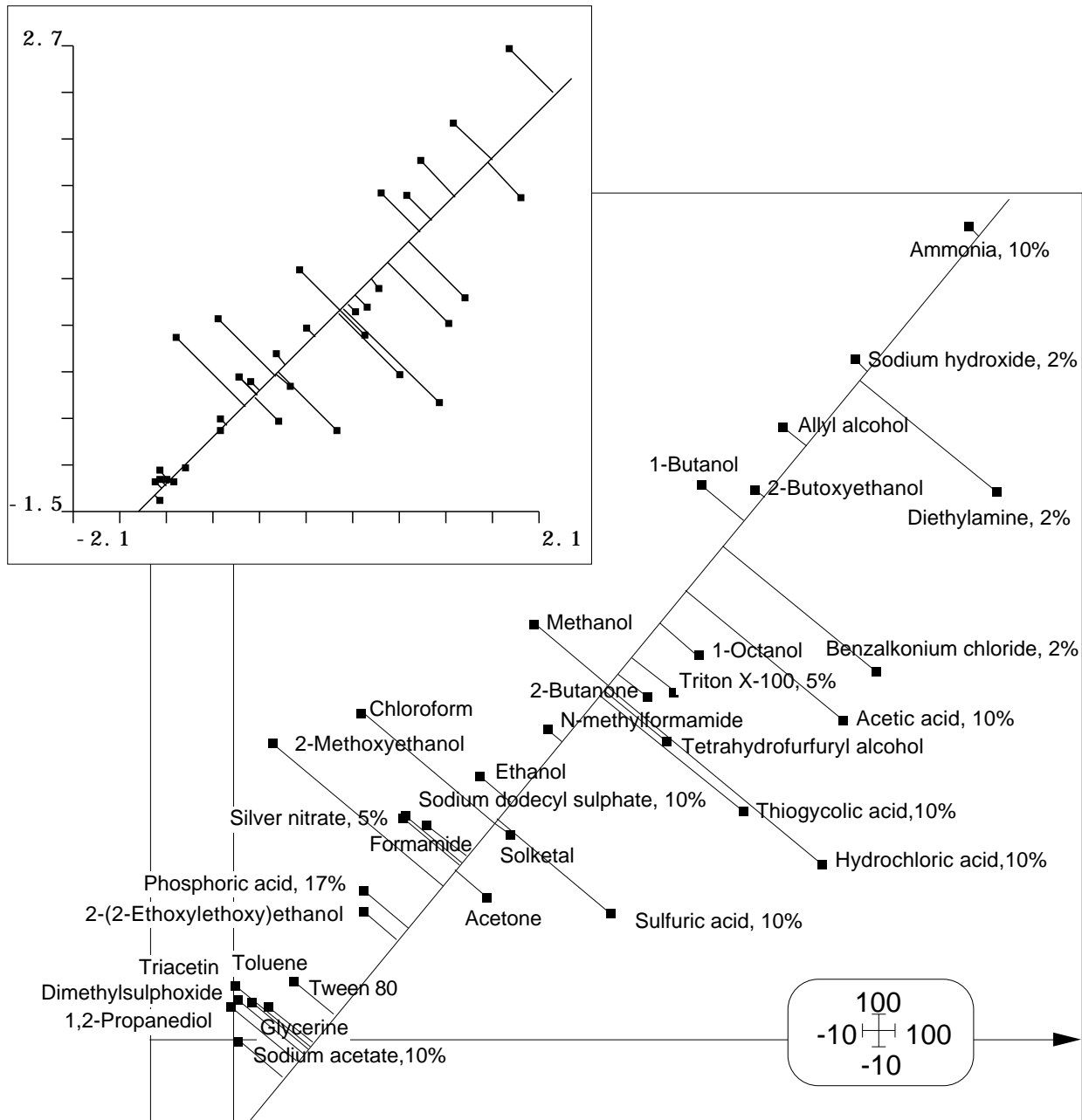


Figure 9 : En abscisse, niveau de toxicité dans le tableau Vivo. En ordonnée, niveau de toxicité dans le tableau Vitro. Utilisation des moyennes sur quatre tests (paramètre du modèle additif). En bas : nuage brut et axe principal du nuage centré. En haut : axe principal du nuage normalisé. On voit qu'un problème d'échelle se pose, en dépit des apparences. Les axes principaux expriment respectivement la covariance et la corrélation.

| | |
|-------------------|----------------|
| No centring | |
| Matrix input file | vivo 34 4 |
| No centring | |
| Matrix input file | vitro 34 4 |
| Add normed scores | |
| Input file | vivo.ncup 4 2 |
| Add normed scores | |
| Input file | vitro.ncup 4 2 |

| Lines | | | |
|------------------------------------|--|------------|------|
| X file (default = 1, 2, 3, ..., n) | | vivo.nc11 | 34 1 |
| X file column number (default = 1) | | | |
| Y file (no default) | | vitro.nc11 | 34 1 |

On obtient la figure 8.

Les codes $a(i)$ étant des vecteurs unitaires à 34 composantes, leur ressemblance se mesure par le cosinus des angles (0.92) et non leur corrélation (0.78), ce qui n'est pas dans les mœurs.

Par contre, les moyennes des quatre tests de toxicité par produits ne sont pas forcément à la même échelle (figure 9) et leur ressemblance se mesure par un coefficient de corrélation (0.79). Quelque soit le mode de comparaison, il est clair que les deux tableaux sont fortement lié par ce paramètre et l'article original discute complètement cet élément, en particulier le rôle des acides qui affaiblit la corrélation globale.

4.2. Comparaison des évolutions temporelles

On abordera enfin une question plus complexe. Une fois le niveau de toxicité éliminé, dans chaque tableau on peut faire une typologie d'évolution temporelle autour du modèle commun. Ces typologies sont elles cohérentes ? En d'autres termes, au delà de la corrélation en valeurs, y a t'il une corrélation en fonctionnement ? La question est posée dans la figure 10. Comme convenu on enlève des deux tableaux le produit Ammonia, 10% qui est isolé.

| Row-Col Selection | | | |
|--------------------------------------|--|------|--|
| Input file | | vivo | |
| Selection of rows (default = all) | | 1a33 | |
| Selection of columns (default = all) | | | |
| Output file | | VV | |

| Row-Col Selection | | | |
|--------------------------------------|--|-------|--|
| Input file | | vitro | |
| Selection of rows (default = all) | | 1a33 | |
| Selection of columns (default = all) | | | |
| Output file | | VT | |

On obtient les fichiers VV (*in vivo* 33-4) et VT (*in vitro* 33-4). On cherche un modèle commun de toxicité en accolant les deux fichiers (V 33-8) qui donne par HTA : Double centring multiplicative un fichier V.dmta qu'on sépare en deux parties (VV0 33-4 et VT0 33-4).

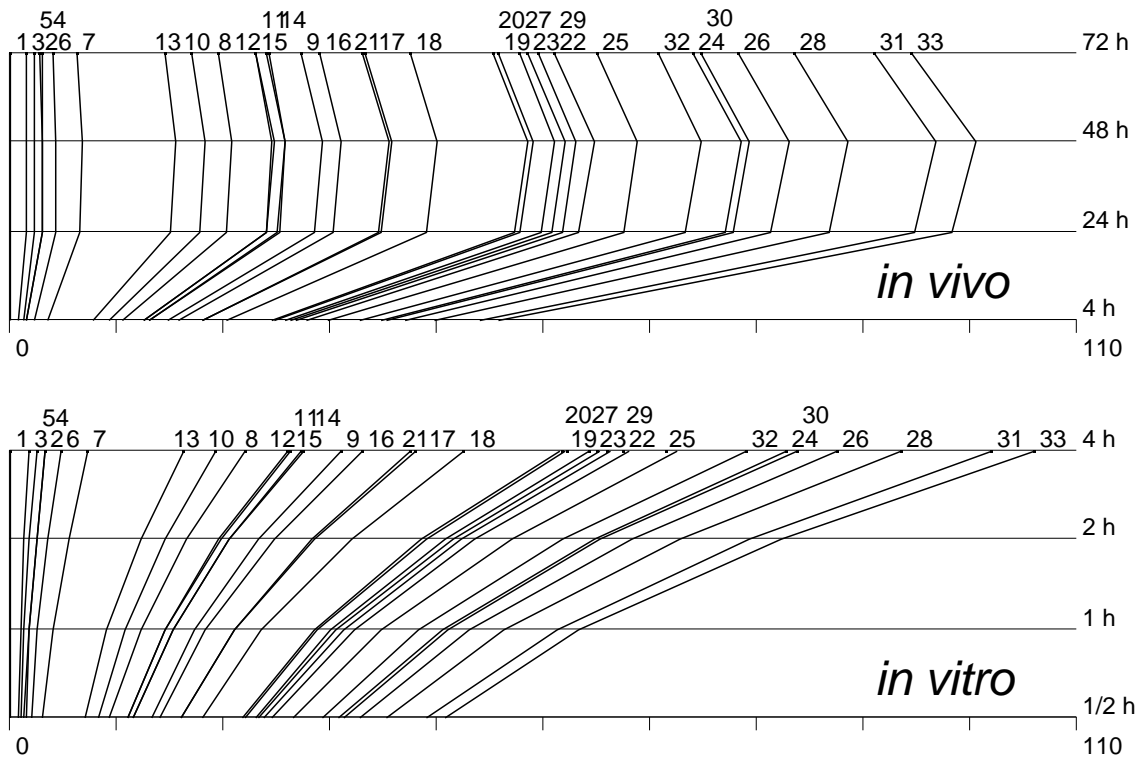
| PasteFiles-SameRow | | | |
|-----------------------------|--|----|------|
| Input BIN file 1 | | VV | 33 4 |
| Input BIN file 2 (optional) | | VT | 33 4 |

| Double centring multiplicative | | | |
|--------------------------------|--|---|------|
| Matrix input file | | V | 33 8 |

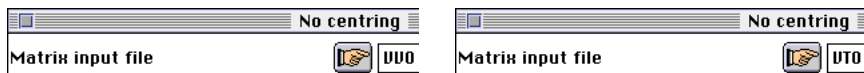
| Row-Col Selection | | | |
|--------------------------------------|--|--------|--|
| Input file | | V.dmta | |
| Selection of rows (default = all) | | | |
| Selection of columns (default = all) | | 1a4 | |
| Output file | | VV0 | |

| Row-Col Selection | | | |
|--------------------------------------|--|--------|--|
| Input file | | V.dmta | |
| Selection of rows (default = all) | | | |
| Selection of columns (default = all) | | 5a8 | |
| Output file | | VT0 | |

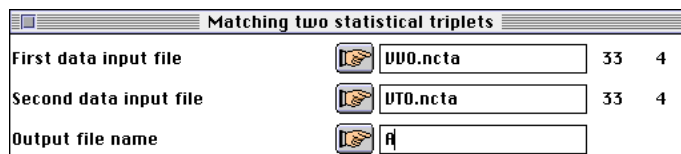
On vérifie que l'usage d'un même code de toxicité n'a pas modifié les modèles propres à chaque tableau :



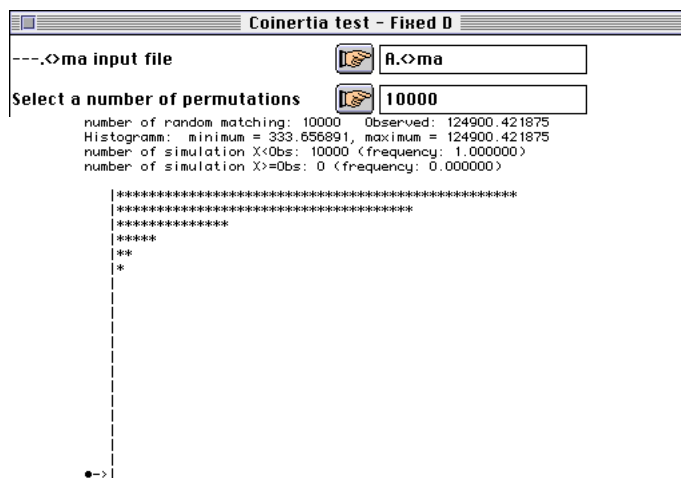
On effectue l'ACP non centrée de chacun des deux :



On associe les deux analyses :



On vérifie le niveau de signification de l'association :



On exécute l'analyse de co-inertie :



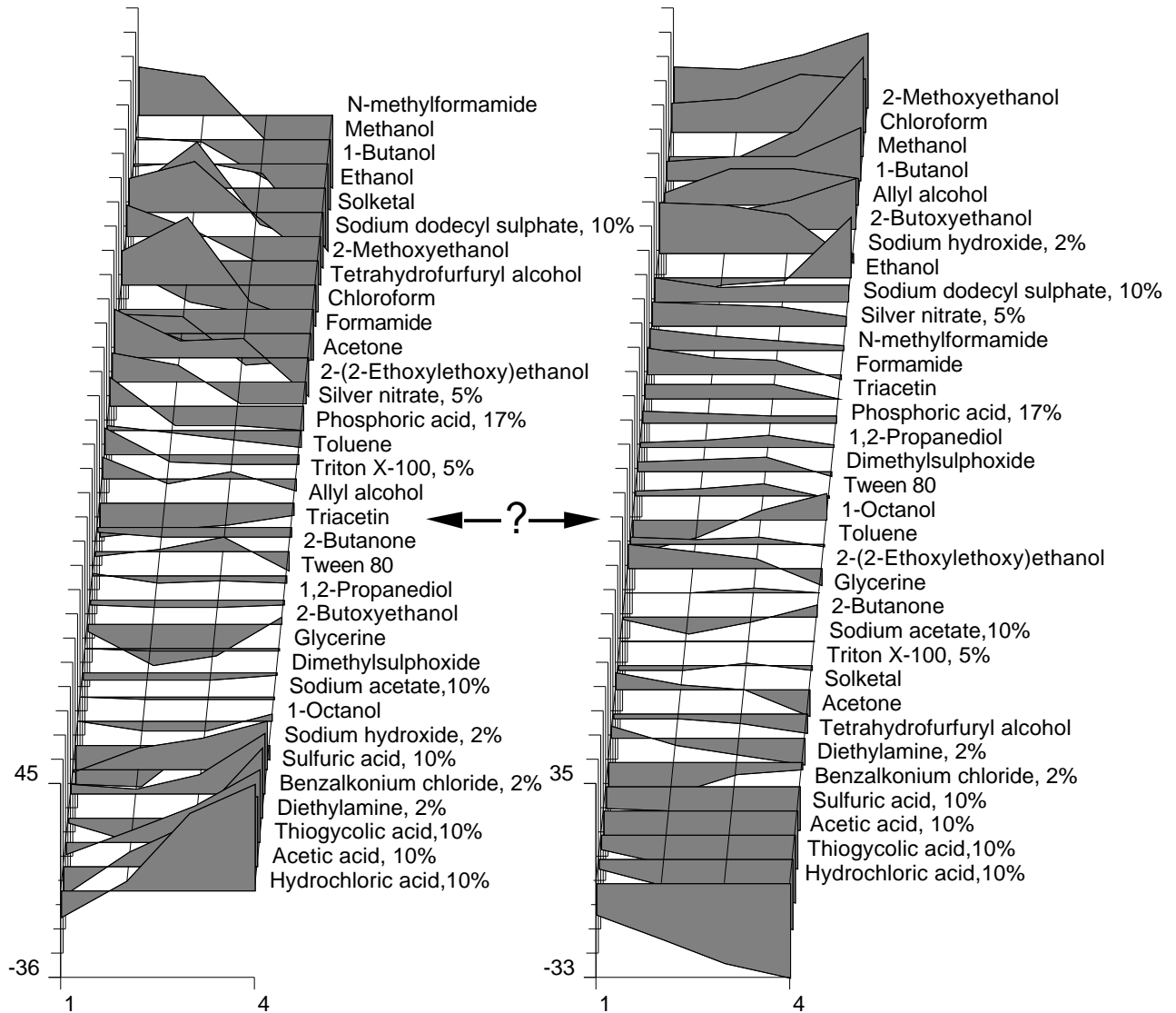


Figure 10 : Réécriture des tableaux de résidus du modèle multiplicatif commun ordonnés par leur facteur 1 respectif. Il est manifeste que la répartition des erreurs n'est pas du même type.

Le listing habituel demande des explications complémentaires car les deux tableaux ne sont pas centrés au sens traditionnel du terme.

DiagoRC: General program for two diagonal inner product analysis

Input file: A.<>ta

--- Number of rows: 4, columns: 4

Total inertia: 124900

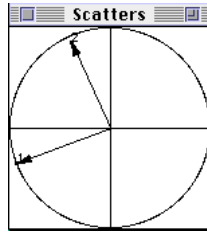
| Num. | Eigenval. | R.Iner. | R.Sum | Num. | Eigenval. | R.Iner. | R.Sum |
|------|-------------|---------|---------|------|-------------|---------|---------|
| 01 | +1.2427E+05 | +0.9949 | +0.9949 | 02 | +5.1621E+02 | +0.0041 | +0.9991 |
| 03 | +1.1154E+02 | +0.0009 | +1.0000 | 04 | +3.2958E+00 | +0.0000 | +1.0000 |

A.<>a1 is a binary file with 2 rows and 2 columns

It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 1)

File :A.<>a1

| Col. | Mini | Maxi |
|------|------------|------------|
| 1 | -9.229e-01 | -3.850e-01 |
| 2 | -3.552e-01 | 8.458e-01 |

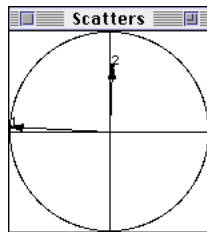


Le plan 1-2 de co-inertie est le plan 1-2 d'inertie après une légère rotation (*in vivo*).

A.<>a2 is a binary file with 2 rows and 2 columns
It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 2)

File :A.<>a2

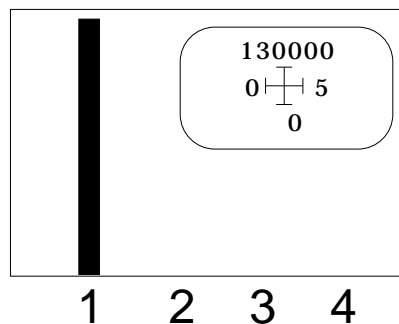
| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -9.990e-01 | 3.394e-02 |
| 2 | 4.550e-02 | 6.631e-01 |



L'axe 1 de co-inertie est l'axe 1 d'inertie (*in vitro*).

| Num | Covaria. | Varian1 | varian2 | Correla. | INER1 | INER2 |
|-----|----------|---------|---------|----------|-------|-------|
| 1 | 352.5 | 430.7 | 347 | 0.9119 | 480.3 | 347.6 |
| 2 | 22.72 | 169.5 | 40.08 | 0.2756 | 145.9 | 79.67 |

La notion de variance, covariance et corrélation n'est pas ici utilisable. Les axes 1 de co-inertie sont sensiblement de même valeur que les axes 1 d'inertie. Il n'y a qu'une seule valeur propre à prendre en compte.



Coinertia: 124900, RV coefficient: 0.694329

A.<>w1 is a binary file with 4 rows and 2 columns

It contains the canonical weights of the variables of table 1

File :A.<>w1

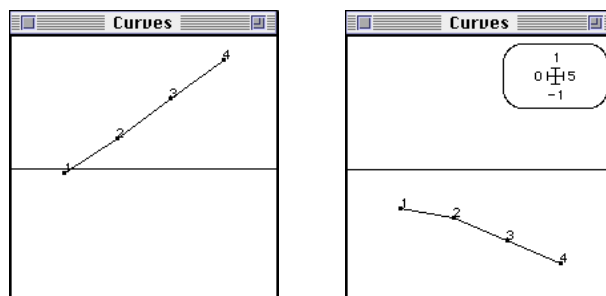
| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -3.771e-02 | 8.166e-01 |
| 2 | -5.757e-01 | 4.400e-01 |

A.<>w2 is a binary file with 4 rows and 2 columns

It contains the canonical weights of the variables of table 2

File :A.<w2

| Col. | Mini | Maxi |
|------|------------|------------|
| 1 | -7.052e-01 | -2.910e-01 |
| 2 | -9.334e-01 | 3.064e-01 |



Les poids canoniques sont de signes constants et évoluent en sens contraire et en intensité différente (l'échelle est commune). On retrouve ici les différences de structure interne des tableaux qui sont pourtant liées entre elles.

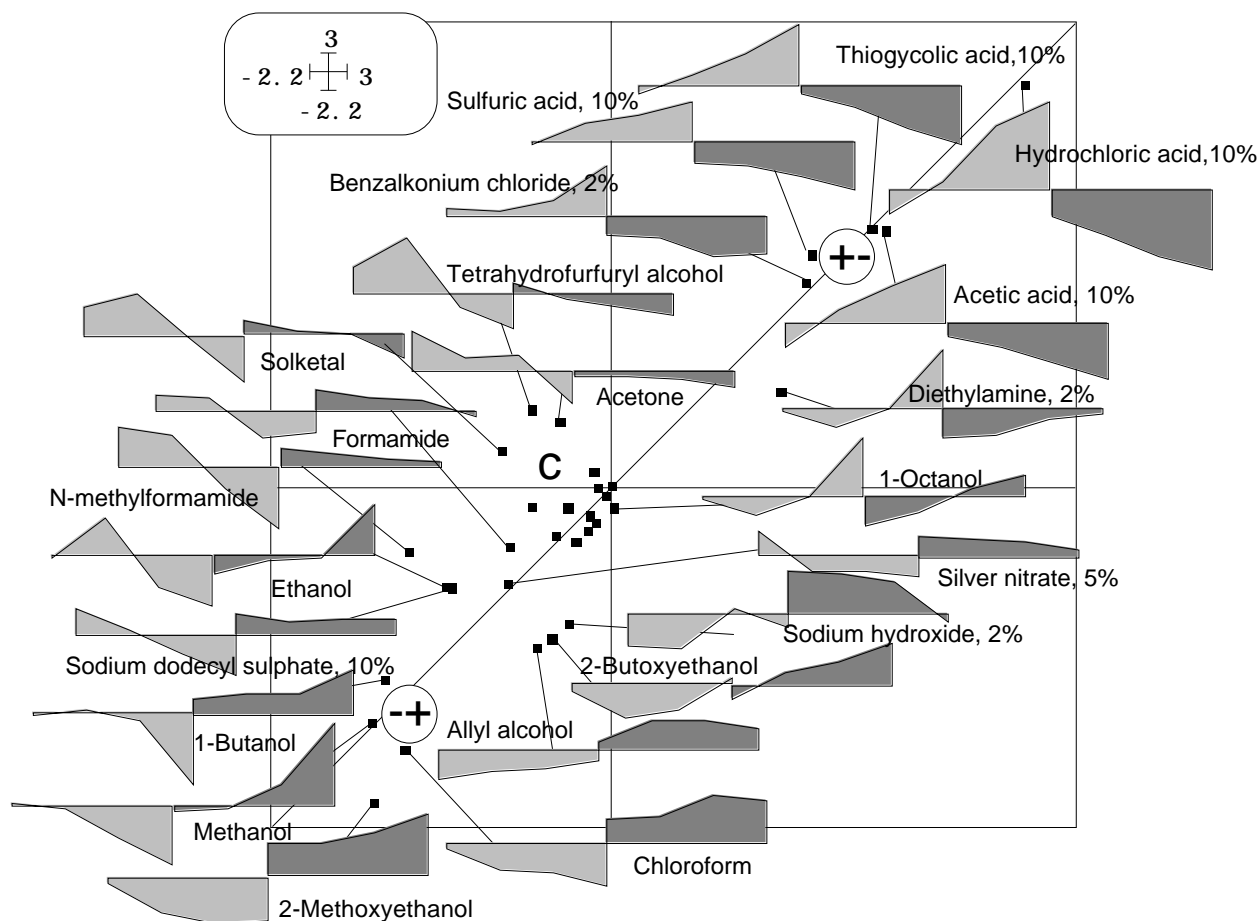


Figure 11 : Report des données traitées sur le graphe des scores de co-inertie. En abscisse score in vivo, en ordonnée score in vitro (scores de norme 1). Pour chaque point, résidus de prédiction par une valeur commune de toxicité (hachuré : valeurs in vivo, grisé : valeurs in vitro). On met en évidence les biais systématiques entre méthodes très sous-estimés par les mesures traditionnelles de corrélation.

A.<ml is a binary file with 33 rows and 2 columns

It contains the normalized coordinates of the rows (table 1)
In earlier version of ADE this file name uses suffix bl1

File :A.<>m1

| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -1.565e+00 | 2.666e+00 |
| 2 | -2.269e+00 | 1.499e+00 |

A.<>m2 is a binary file with 33 rows and 2 columns

It contains the normalized coordinates of the rows (table 2)

In earlier version of ADE this file uses suffix bl2

File :A.<>m2

| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -2.047e+00 | 2.599e+00 |
| 2 | -2.226e+00 | 2.257e+00 |

La figure 11 utilise ces codes pour ordonner les produits. En haut 4 des 5 acides expriment une structure du type - + avec une courbe *in vitro* croissante. En bas 4 produits caractéristiques et une tendance pour nombre d'autres expriment une structure du type - + avec un courbe *in vitro* décroissante. Il y a donc contradiction entre les deux tableaux autour de la notion de toxicité commune. La co-structure exprime que les produits lents dans les tests *in vivo* sont sous-estimés dans les tests *in vitro*, alors que les produits rapides dans les tests *in vivo* sont sur estimés dans les tests *in vitro*.

Références

¹ Jacobs, G.A. & Martens, M.A. (1990) Quantification of eye irritation based upon in vitro changes of corneal thickness. *ATLA* : 17, 255-262.

² Devillers, J. & Chessel, D. (1995a) Can the enucleated rabbit eye test be a suitable alternative for the in vivo eye test? A chemometrical response. *Toxicology Modelling* : 1, 21-34.

Devillers, J. & Chessel, D. (1995b) Comparaison of in vivo and in vitro toxicity tests from co-inertia analysis. In : *Computer-Aided Molecular Design. Applications in Agrochemicals, Materials and Pharmaceuticals*. Reynolds, C.H., Holloway, M.K. & Cox, H.K. (Eds.) ACS Symposium Series 589. American Chemical Society, Washington. 250-266.

³ Mandel, J. (1961) Non additivity in two-way analysis of variance. *Journal of the American Statistical Association* : 65, 878-888. MANDEL, J. (1971) A new analysis of variance model for non-additive data. *Technometrics* : 13, 1-18. OKAMOTO, M. (1972) Four techniques of principal component analysis. *Journal of the Japanese Statistical Society* : 2, 63-69.

⁴ Whittle, P. (1952) On principal components and least square methods of factor analysis. *Skandinavisk aktuarietidskrift* : 35, 223-239. GABRIEL, K.R. (1971) The biplot graphical display of matrices with application to principal component analysis. *Biometrika* : 58, 453-467.

⁵ Chessel, D. & Thioulouse, J. (1990) Auto-modélisation en analyse des données. In : *Modélisation confluent des sciences*. BRISSAUD, M., FORSÉ, M. & ZIGHED, A. (Eds.) Editions du CNRS. 71-86.

⁶ Devillers, J., Thioulouse, J. & Karcher, W. (1993) Chemometrical Evaluation of Multispecies-Multichemical Data by Means of Graphical Techniques Combined with Multivariate Analyses. *Ecotoxicology and Environmental Safety* : 26. 333-345.