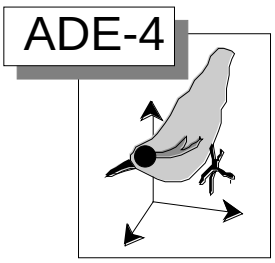


PLSgen2

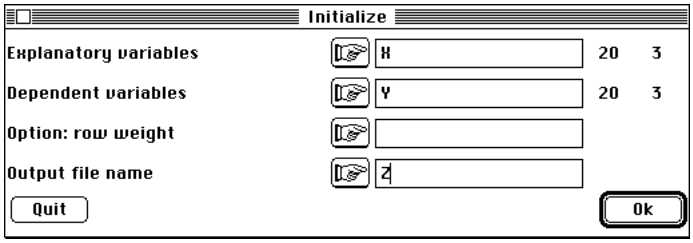
Luc MONIMEAU
Laboratoire de Phytopathologie, Centre ORSTOM
98848 NOUMEA Cedex, BP A5, Nouvelle-Calédonie.
Tel : (19) (687) 26 07 15
Fax : (19) (687) 26 43 26
e-Mail : monimeau@noumea.orstom.nc

PLSgen2 : Initialize

Utilitaire de contrôle de données : préparation des données pour effectuer une régression PLS de deuxième génération.

L'option s'emploie pour définir un jeu de paramètres qui sera utilisé par toutes les autres options de ce module. Ici, on s'intéresse aux modèles de prédiction élaborés à partir d'une régression PLS de 2ème génération. Nous serons donc en présence d'un groupe **de *m* variables explicatives quelconques** et de ***p*** **variables à prédire.**

L'option utilise une seule fenêtre de dialogue :



Nom du fichier binaire contenant les variables explicatives.

Nom du fichier binaire contenant les variables à prédire.

Fichier de pondération des lignes (par défaut, on utilise la pondération uniforme).

Nom générique du fichier de sortie.

Utiliser la carte Lineerud de la pile ADE-4•Data pour obtenir les fichiers X (20-3) et Y (20-3). Associer le fichier des variables prédictives (X) et des variables à prédire (Y) par la présente option.

```
-------------------------------------------------------
New TEXT file Z.reg contains the parameters:
----> Explanatory variables: X [20][3]
----> Dependent variable file: Y [20][3]
----> Row weight file: Uniform weight
-------------------------------------------------------
```

L'option enregistre ces paramètres dans un fichier texte. Le fichier ---.reg ainsi créé nous permettra d'utiliser les autres options du module.

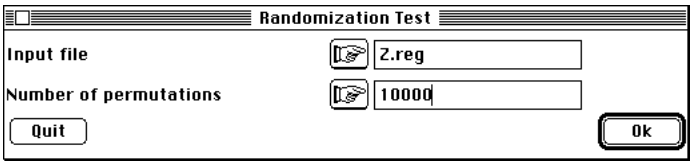
Le fichier ---.reg créé par cette option, pourra également être utilisé dans le module LinearReg pour les régressions multiples et PLS d'ordre 1.

PLSgen2 : RandomizationTest

Utilitaire pour le choix du nombre d'itérations et par conséquent de composantes dans une régression PLS de deuxième génération.

La régression PLS étant une méthode itérative, la question porte alors sur le nombre d'itérations à utiliser. Les tests de permutation[1] permettent de répondre à cette question, en prenant comme statistique à mesurer le paramètre qui caractérise la qualité de la régression[2]. Ainsi à chaque pas, on compte la fréquence des permutations qui donnerait un pourcentage d'explication aussi bon que celui observé. On ne retiendra pour PLSgen2 : Modelling que le nombre d'itérations franchement significatives.

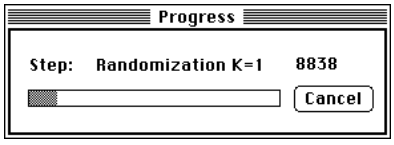
L'option utilise une seule fenêtre de dialogue :



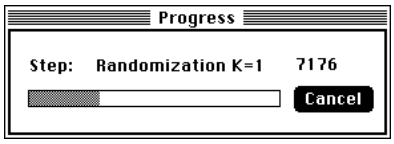
Fichier des paramètres créé par PLSgen2 : Initialize.

Nombre de permutations utilisées (n). A chaque itération, le nombre maximum étant fixé à 4, on fait n permutations.

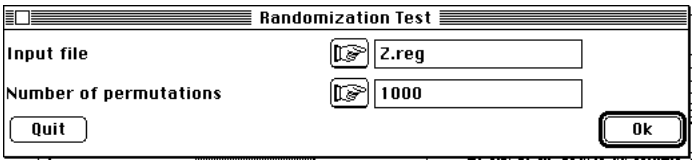
L'opération peut s'avérer assez longue en fonction de la taille des matrices qu'il faudra diagonaliser pour chaque permutation et ce à chaque itération. La fenêtre habituelle affiche la progression :



Cette opération va se répéter au plus quatre fois si les quatre premières composantes sont significatives. Si on pense avoir abuser, arrêter :



et relancer :



Utiliser l'exemple mis en place dans la fiche de PLSgen2 : Initialize :

```
-----------------------------------------
Explanatory variable file: X
It has 20 rows and 3 columns
-----------------------------------------
Dependent variable file: Y
It has 20 rows and 3 columns
--------------------------------------
-------- Dependent variable Y ---------
--------------------------------------
| Step | Nrepet | X>Xobs | Frequencies|
|     1|    1000|      57|   5.700e-02|
|     2|    1000|     605|   6.050e-01|
```

Le fait de ne trouver au premier tour, que 5.7% de permutations aléatoires dépassant l'observation, prouve que la prédiction a un sens. Au second tour on en trouve déjà 60%, ce qui implique que l'on peut tenir compte que d'une seule itération et par conséquent ne retenir qu'une seule composante pour la mise en place du modèle.

⚠ Le nombre maximum de pas d'une telle procédure est fixé à 4 s'il y a plus de 4 variables explicatives et au nombre d'explicatives dans le cas contraire. Par exemple, ici le nombre de pas est limité à 3. Toutefois pour éviter une perte de temps, l'algorithme s'arrête dès qu'une composante n'est pas significative.

[1] Good, P. (1994) Permutation tests. Springer-Verlag, New-York. 1-228.

[2] Tenenhaus, M. (1995) Nouvelles Méthodes de Régression PLS. CR 540/1995. Groupe HEC.

Tenenhaus, M., Gauchi, J.P. & Ménardo, C. (1995) Régression PLS et applications. *Revue de Statistique Appliquée* : 43, 7-63.

Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1994) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* : 20, 643-656.

Fraile, L., Escoufier, Y. & Raibaut, A. (1993) Analyse des correspondances de données planifiées : Etude de la chémotaxie de la larve infestante d'un parasite. *Biometrics* : 49, 1142-1153.

Wold, S. (1978) Cross-validation estimation of the number of components in factor and principal components models. *Technometrics* : 20, 397-405.

Cramer, R.D. III, Bunce, J.D., Patterson, D.E. & Frank, I.E. (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships* : 7, 18-25.

Gauchi, J.P. (1995) Utilisation de la régression PLS pour l'analyse des plans d'expériences en chimie de formulation. *Revue de Statistique Appliquée* : 43, 65-89.

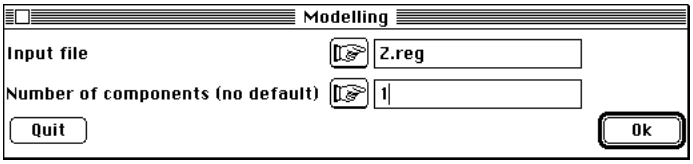
Ter Braak, C.J.T. & Juggins, S. (1993a) Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* : 269/270: 485-502.

PLSgen2 : Modelling

Régression PLS : méthode de régression linéaire permettant de relier un ensemble de variables à prédire à un ensemble de variables explicatives quand le nombre de variables à expliquer et/ou explicatives est grand par rapport au nombre d'individus. Cette méthode est notamment adaptée à la construction de modèles lorsque les variables explicatives sont fortement corrélées entre elles. Elle permet en effet d'obtenir une certaine cohérence entre les coefficients de corrélation et les coefficients de régression.

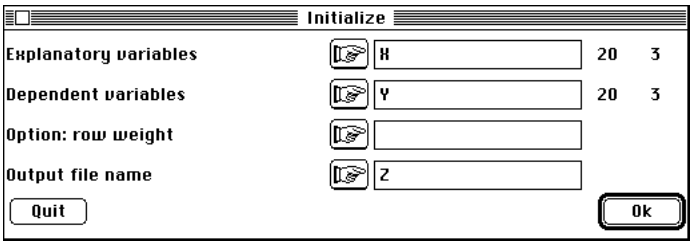
Le module exécute une régression PLS (Partial Least Squares ou partiellement aux moindres carrés) de deuxième génération. Il s'agit plus particulièrement d'une méthode dérivée de la PLS dite "classique" (synthèse complète dans Lindgren[1]), qui est adaptée à la mise en place de modèles. L'algorithme utilisé est décrit par Tenenhaus (1995)[2].

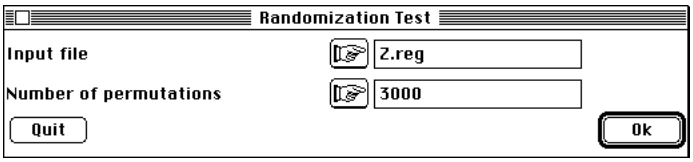
L'option utilise une seule fenêtre de dialogue :



- Fichier des paramètres crée par PLSgen2 : Initialize.
- Nombre de composantes utilisées (Voir PLSgen2 : Randomization test)

Utiliser l'exemple mis en place dans la fiche PLSgen2 : Initialize. Initialiser et tester le nombre de composantes:






```
-----------------------------------------
Explanatory variable file: X
It has 20 rows and 3 columns
-----------------------------------------
Dependent variable file: Y
It has 20 rows and 3 columns
--------------------------------------
-------- Dependent variable Y ---------
--------------------------------------
| Step | Nrepet | X>Xobs | Frequencies|
|     1|    3000|     172|   5.733e-02|
|     2|    3000|    1808|   6.027e-01|
```

Par conséquent nous allons utiliser qu'une seule composante pour exécuter la régression PLS.

Attention, ce programme présente une logique particulière associée aux choix de l'auteur. En prenant une composante explicative, on s'attend à ce que les fichiers de sortie contenant des aides à l'interprétation possède une colonne. Il n'en est rien.

Les composantes explicatives sont calculées et jusqu'à la quatrième (sauf s'il y a moins de quatre explicatives comme c'est la cas ici on en a trois). Le nombre de composantes explicatives calculées est donc de 4 systématiquement ou du nombre d'explicatives si celui-ci est inférieur à 4). Par contre le modèle (tableau modélisé et coefficients de régression) n'intègre que le nombre de composantes explicatives choisi par l'utilisateur.

Le listing contient les indications qui suivent. Se reporter aux travaux de M. Tenenhaus cités pour avoir la définition mathématique précise des quantités calculées.

Nom des fichiers utilisés :

```
----------------------------------------
Explanatory variable file: X
It has 20 rows and 3 columns
----------------------------------------
Dependent variable file: Y
It has 20 rows and 3 columns
----------------------------------------
```

Matrice des corrélations entre explicatives (1 à 3) entre expliquées (4 à 6) et entre les deux paquets de variables :

```
---------------------- Correlation matrix -------------------
[  1] 1000
[  2]  870 1000
[  3] -366 -353 1000
[  4] -390 -552  151 1000
[  5] -493 -646  225  696 1000
[  6] -226 -191   35  496  669 1000
------------------------------------------------------------
```

Pourcentage de variances expliquées pour chaque tableau. Il faut savoir que les composantes explicatives sont des combinaisons de variables explicatives qui servent à modéliser le tableau des variables dépendantes (Eo), ceci est attendu, et qui servent à modéliser le tableau des explicatives, ceci est moins ordinaire. En effet, la première composante explicative est retiré du tableau des variables explicatives par résidus de régression simple et la composante explicative suivante est une combinaison des variables ainsi modifiées (c'est pourquoi elle est non corrélée à la première).

```
|---------------|*** Eo ***|---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 6.948e-01| 6.948e-01| 6.948e-01|


|---------------|*** Fo ***|---------------------|
|Step| Variance | Explained|  Ratio   | Exp. Sum |
|   1| 1.000e+00| 2.094e-01| 2.094e-01| 2.094e-01|


|---------------| Fo Col:  1|---------------------|
|Step| Variance | Explained |  Ratio   | Exp. Sum |
|   1| 1.000e+00|  2.363e-01| 2.363e-01| 2.363e-01|


|---------------| Fo Col:  2|---------------------|
|Step| Variance | Explained |  Ratio   | Exp. Sum |
|   1| 1.000e+00|  3.506e-01| 3.506e-01| 3.506e-01|


|---------------| Fo Col:  3|---------------------|
|Step| Variance | Explained |  Ratio   | Exp. Sum |
|   1| 1.000e+00|  4.140e-02| 4.140e-02| 4.140e-02|
```

Pour ces 5 tableaux, la variance initiale de l'ensemble des variables à prédire, ou de la variable à prédire, vaut 1. La première composante explique 69.5% de la variance du tableau E_0, c'est à dire des variables prédictives et 20.9% de la variance des variables à prédire (Tableau F_0).

Plus particulièrement, la première composante explique 23.6% de la variance de la première variable à prédire, 35% de la seconde... La moyenne sur les trois composantes étant bien égale à 20.9%

Les coefficients du modèle global élaboré avec une seule composante sont conservés dans les deux fichiers suivants :

```
File Z_PLS2w1 contains the regression coefficients
It has 3 rows and 3 columns
File :Z_PLS2w1
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-2.635e-01| 8.161e-02|
|   2|-3.209e-01| 9.939e-02|
|   3|-1.103e-01| 3.415e-02|
|----|----------|----------|
File Z_PLS2w2 contains the regression coefficients on original data
It has 4 rows and 3 columns
File :Z_PLS2w2
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-4.350e-01| 2.920e+01|
|   2|-6.271e+00| 4.303e+02|
|   3|-1.766e+00| 1.505e+02|
|----|----------|----------|
```

Le premier de ces fichiers contient les coefficients du modèle de régression écrit sur les variables normalisées alors que le second fichier donne les coefficients pour les données d'origine. On obtient ainsi pour les données centrées réduites :

$$Chins^* = -0.20\ Weight^* - 0{,}26\ Waist^* + 0.08\ Pulse^*$$
$$Situps^* = -0.24\ Weight^* - 0.32\ Waist^* + 0.10\ Pulse^*$$
$$Jumps^* = -0.08\ Weight^* - 0.11\ Waist^* + 0.03\ Pulse^*$$

Le second fichier correspond aux coefficients calculés pour les données brutes

⚠ Attention, la première ligne de ce tableau contient les termes constants. On obtient ainsi pour les données d'origine :

Z_PLS2w2

	1	2	3
1	29.2002	430.2511	150.4807
2	-0.0431	-0.6220	-0.1752
3	-0.4350	-6.2712	-1.7662
4	0.0598	0.8625	0.2429

Chins = 29.20 - 0.04 Weight - 0,43 Waist + 0.05 Pulse
Situps = 430.25 - 0.62 Weight - 6.27 Waist + 0.86 Pulse
Jumps = 150.48 - 0.17 Weight - 1.76 Waist + 0.24 Pulse

Ces résultats sont ceux de [3] (p. 43) obtenu sur ces données avec le logiciel SIMCA[4]. En effet, la première composante t_1 de la PLS-4 est la même que celle de la PLS dite "classique". D'autre part, le modèle et le résidu, écrits pour les données d'origine, sont conservés :

```
-------------------------------------------
File Z_PLS2mod contains the component model
It has 20 rows and 3 columns
File :Z_PLS2mod
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 1.522e+00| 1.296e+01|
|   2| 3.127e+01| 1.961e+02|
|   3| 3.812e+01| 8.454e+01|
|----|----------|----------|
-------------------------------------------
File Z_PLS2res contains the residual matrix
It has 20 rows and 3 columns
```

```
File :Z_PLS2res
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.096e+01| 6.892e+00|
|   2|-8.612e+01| 8.395e+01|
|   3|-4.287e+01| 1.712e+02|
|----|----------|----------|
```

Le listing nous permet d'obtenir les différents vecteurs et différentes composantes entrant en compte dans l'algorithme de la PLS [2], soit :

— les vecteurs w_k :

```
File Z_PLS2.w contains the vectors w
It has 3 rows and 4 columns
File :Z_PLS2.w
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-7.713e-01| 2.389e-01|
|   2|-6.902e-01| 3.635e-01|
|   3|-7.467e-01| 6.408e-01|
|----|----------|----------|
```

— les composantes explicatives t_k(on peut vérifier que la première est toujours la première coordonnée sur le tableau des explicatives de l'analyse de co-inertie entre les ACP normées des explicatives et des dépendantes) :

```
File Z_PLS2.t contains the components t
It has 20 rows and 4 columns
File :Z_PLS2.t
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-4.504e+00| 1.993e+00|
|   2|-1.142e+00| 1.145e+00|
|   3|-1.277e+00| 4.648e-01|
|----|----------|----------|
```

— les vecteurs c_k* :

```
File Z_PLS2.c contains the vectors c
It has 3 rows and 4 columns
File :Z_PLS2.c
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 2.567e-01| 7.470e-01|
|   2| 1.446e-01| 7.485e-01|
|   3|-6.885e-01| 3.070e-01|
|----|----------|----------|
```

— les vecteurs u_h* :

```
File Z_PLS2.u contains the components u
It has 20 rows and 4 columns
File :Z_PLS2.u
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-2.280e+00| 3.113e+00|
|   2|-2.300e+00| 2.736e+00|
|   3|-1.581e+00| 2.034e+00|
|----|----------|----------|
```

— les vecteurs c_k :

```
File Z_PLS2.r contains the components r
It has 3 rows and 4 columns
File :Z_PLS2.r
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 1.430e-01| 4.161e-01|
|   2| 6.606e-02| 3.419e-01|
|   3|-5.967e-01| 2.660e-01|
```

```
|----|----------|----------|
```

Sont conservées ensuite les matrices normalisées

— d'une part des variables prédictives :

```
File Z_PLS2.Eo contains the normalized matrix X
It has 20 rows and 3 columns
```

— d'autre part des variables dépendantes :

```
File Z_PLS2.Fo contains the normalized matrix Y
It has 20 rows and 3 columns
```

Cette option nous permet également d'obtenir des fichiers destinés à la représentation graphique, avec la matrice des corrélations entre les explicatives et les composantes explicatives :

```
-------------------------------------------
File Z_PLS2.X+co contains the column scores of matrix X
It has 3 rows and 1 columns
File :Z_PLS2.X+co
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-9.620e-01| 5.108e-01|
|   2|-7.901e-01|-1.280e-02|
|   3|-3.389e-01| 1.391e-01|
|----|----------|----------|
```

la matrice des corrélations entre les variables dépendantes et les composantes explicatives :

```
-------------------------------------------
File Z_PLS2.Y+co contains the column scores of matrix Y
It has 3 rows and 1 columns
File :Z_PLS2.Y+co
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1| 2.035e-01| 5.921e-01|
|   2| 4.306e-02| 2.229e-01|
|   3|-2.314e-01| 1.032e-01|
|----|----------|----------|
```

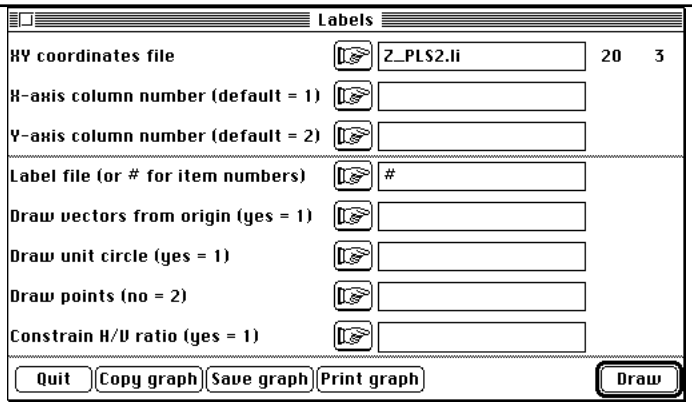
le regroupement des deux :

```
-------------------------------------------
File Z_PLS2.XY+co contains the column scores of matrix X and Y
It has 6 rows and 1 columns
File :Z_PLS2.XY+co
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-9.620e-01| 5.921e-01|
|   2|-7.901e-01| 2.229e-01|
|   3|-3.389e-01| 1.391e-01|
|----|----------|----------|
```

et le regroupement des composantes t et u, pour les composantes conservées :

```
-------------------------------------------
File Z_PLS2.tu contains components th and uh (for graphic (th,uh))
It has 20 rows and 2 columns
File :Z_PLS2.tu
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-4.504e+00| 1.993e+00|
|   2|-2.280e+00| 3.113e+00|
|----|----------|----------|
```

Le module Scatters va permettre d'obtenir les représentations graphiques classiquement utilisées dans l'interprétation :



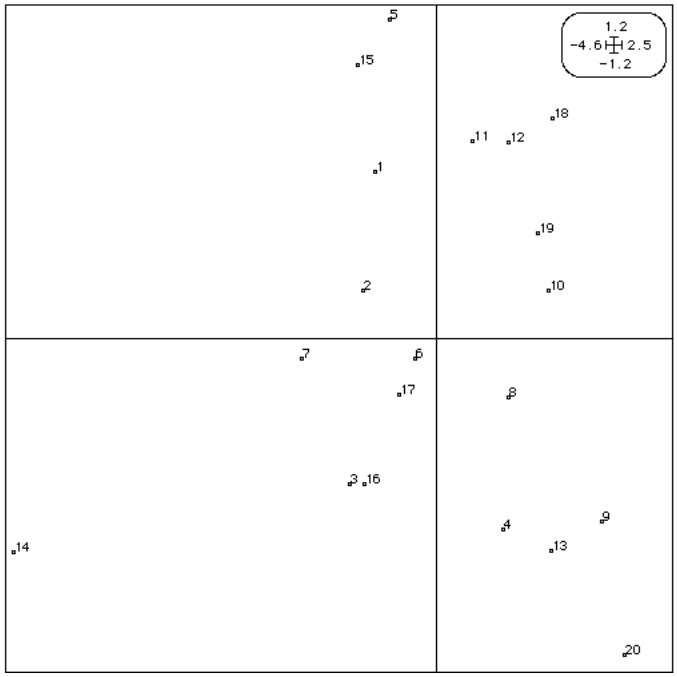
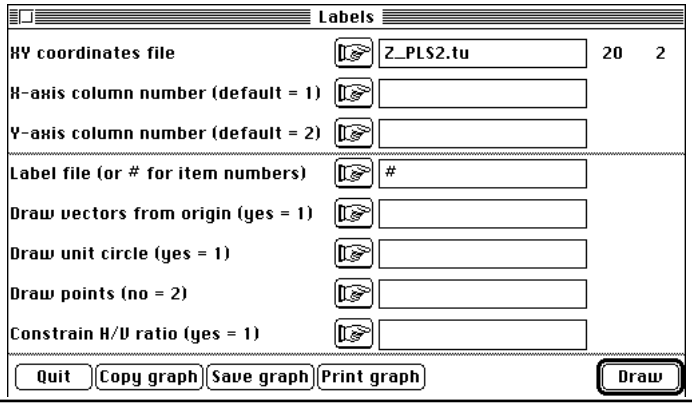


Figure 1 : Représentation des individus dans le plan (t_1 , t_2).

Le plan (t_1,t_2) est équivalent à un plan principal résumant au mieux les variables prédictives, tout en étant orienté vers l'explication des variables à prédire. Ainsi sur ce graphique, on voit clairement le rôle particulier joué par l'individu 14. En effet, celui-ci obtient les plus faibles valeurs pour les variables Chins (1) et Situps (50) et possède les plus fortes valeurs pour les variables explicatives Weight (247) et Waist (46).

Ce graphique met en évidence les trois individus 10, 14 et 20 qui vont jouer un rôle particulier dans la régression PLS. En effet, on s'aperçoit que les individus 10 et 20 obtiennent des résultats opposés aux différents exercices, bien qu'ayant des caractéristiques physiques semblables.



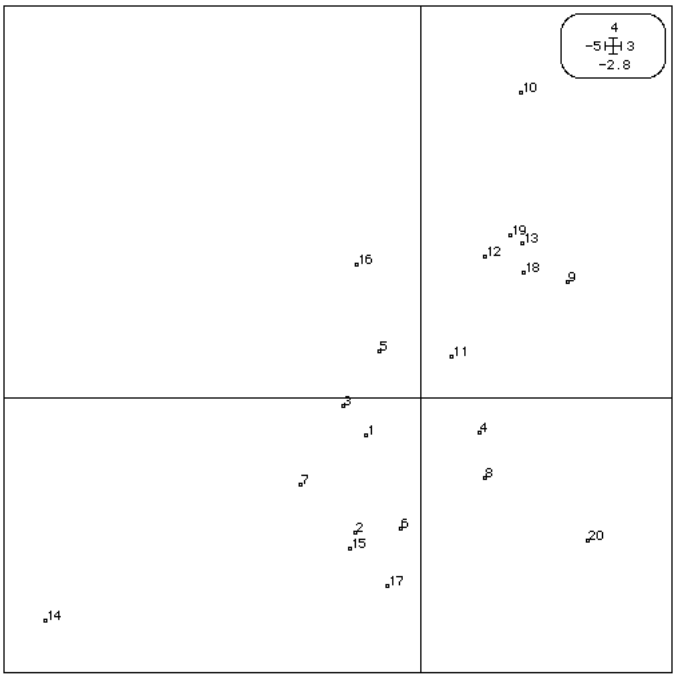


Figure 2 : Représentation des individus dans le plan (t_1 , u_1).

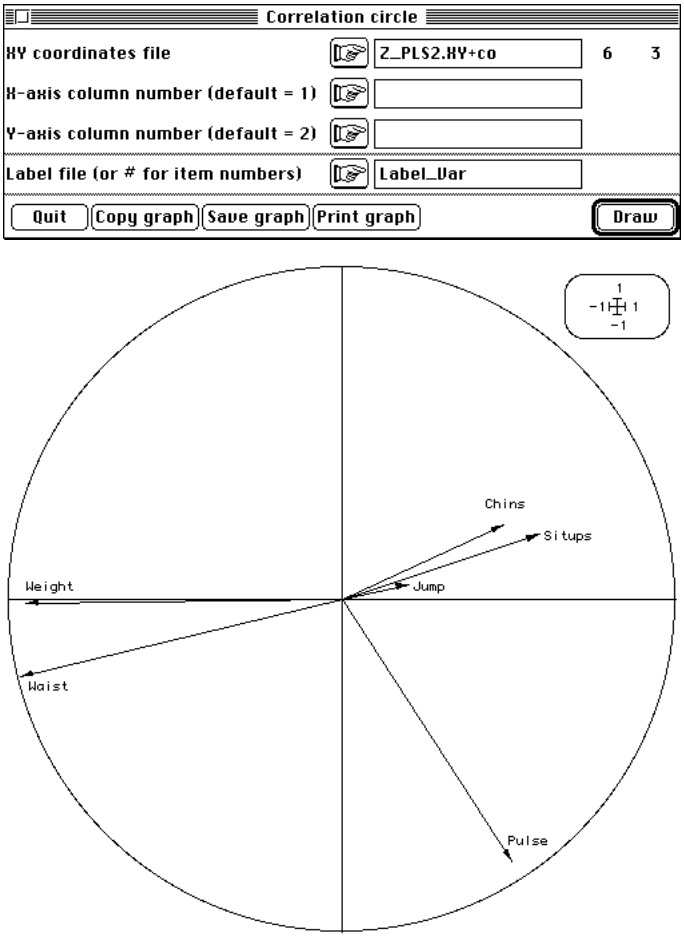
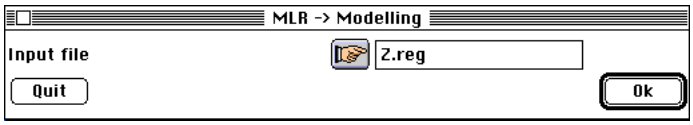


Figure 3 : Cercle des corrélations.

Le fichier Label-Var est un fichier texte contenant le nom des différentes variables (variables explicatives + variables à prédire).

Le cercle des corrélations traduit les corrélations au sein de chaque groupe de variables, ainsi que celles entre les groupes.

Il est intéressant de pouvoir comparer les résultats de la PLS avec ceux obtenus par une régression multiple (LinearReg : MLR -> Modelling). Pour cette dernière, on obtient :




```
R2 = Squared multiple correlation coefficient
Var.|   Mean     | Variance |     R2     |
   1| 9.450e+00| 2.655e+01| 3.396e-01|
   2| 1.456e+02| 3.719e+03| 4.365e-01|
   3| 7.030e+01| 2.498e+03| 5.390e-02|
----------------------------------------
File Z.MLRmod has 20 rows and 3 columns
It contains the linear models resulting
from separate multiple linear regression of each dependant variable
upon the set of explanatory variables
File :Z.MLRmod
|Col.|   Mini    |   Maxi    |
|----|----------|----------|
|   1|-4.734e-01| 1.531e+01|
|   2| 1.017e+01| 2.227e+02|
|   3| 3.836e+01| 8.695e+01|
|----|----------|----------|
----------------------------------------
File Z.MLRres has 20 rows and 3 columns
It contains (data - model) matrix
File :Z.MLRres
|Col.|   Mini    |   Maxi    |
|----|----------|----------|
|   1|-7.517e+00| 6.153e+00|
|   2|-7.419e+01| 8.164e+01|
|   3|-4.789e+01| 1.670e+02|
|----|----------|----------|
----------------------------------------
File Z.MLRw1 has 3 rows and 3 columns
It contains the regression coefficients
Rows : explanatory variables / Columns : dependant variables
Models for normalized (mean = 0 / variance =1) variables
File :Z.MLRw1
|Col.|   Mini    |   Maxi    |
|----|----------|----------|
|   1|-8.818e-01| 3.683e-01|
|   2|-8.898e-01| 2.872e-01|
|   3|-2.590e-01| 1.460e-02|
|----|----------|----------|
----------------------------------------
```

On peut éditer les coefficients des régressions multiples ordinaires, afin de pouvoir les comparer avec ceux obtenus par la PLS :

	Chins*	Situps*	Jumps*
Weight*	0.3683	0.2872	-0.2590
Waist*	-0.8818	-0.8898	0.0146
Pulse*	-0.0258	0.0161	-0.0546

Tableau 1 : coefficients des régressions multiples ordinaires.

Dans ce cas là, on remarque une certaine incohérence entre valeurs des coefficients de régression et valeurs des corrélations. En effet une corrélation négative peut s'accompagner d'un poids positif dans le modèle. La régression PLS va permettre de remédier à ce problème en rendant les équations de régressions cohérentes, au prix d'une légère diminution des pouvoirs explicatifs de chaque régression. Ainsi, dans un cas comme celui-ci, où les variables explicatives sont fortement corrélées entre elles, la PLS va s'imposer [5].

[1] Lindgren, F. (1994) *Third generation PLS. Some elements and applications*. Research Group for Chemometrics. Department of Organic Chemistry. Umeå University. S-901 87 Umeå, Sweden. ISBN 91-7174-911-X, 1-57 with five papers.

[2] Tenenhaus, M. (1995) Nouvelles Méthodes de Régression PLS. CR 540/1995. Groupe HEC.

[3] Tenenhaus, M., Gauchi, J.P. & Ménardo, C. (1995) Régression PLS et applications. *Revue de Statistique Appliquée* : 43, 7-63.

Tenenhaus, M. (1993) La Régression PLS. D 1614K93 Document Groupe HEC. Groupe HEC. 1-25.

Gauchi, J.P. (1995) Utilisation de la régression PLS pour l'analyse des plans d'expériences en chimie de formulation. *Revue de Statistique Appliquée* : 43, 65-89.

Geladi, P. (1988) Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics* : 2, 231-246.

Geladi, P. & Kowalski, B.R. (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta* : 1, 185, 19-32.

Höskuldsson, A. (1988) PLS regression methods. *Journal of Chemometrics* : 2, 211-228.

Kettaneh-Wold, N. (1992) Analysis of mixture data with partial least squares. *Chemometrics and Intelligent Laboratory Systems* : 14, 57-69.

Wold, S. (1995) PLS for multivariate linear modeling. In : *Chemometric Methods in Molecular Design.* van der Waterbeemd. (Ed.) VCH, Weinhein, Germany. 195-218.

Devillers, J. & Chessel, D. (1994) Graphical Analysis as an Aid in Medicinal Chemistry. In : *Chemometric Methods in Molecular Design*. van der Waterbeemd. (Ed.) VCH, Weinhein, Germany. 165-176.

Ter Braak, C.J.T. & Juggins, S. (1993a) Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* : 269/270: 485-502.

[4] SIMCA. (1991) Soft Independent Modeling of Class Analogy. Version 4.3R. Umetri AB Box 1456, S - 90124 Umea. -.

[5] Cramer, R.D. III, Bunce, J.D., Patterson, D.E. & Frank, I.E. (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships* : 7, 18-25.