

OrthoVar

OrthoVar : Initialize.....	2
OrthoVar : Modelling.....	6
OrthoVar : Partial test.....	7
OrthoVar : Subspace test.....	9
OrthoVar : Variable test.....	10

OrthoVar : Initialize



Utilitaire de contrôle de données : préparation de régressions multiples sur variables orthogonales.



L'option s'emploie pour définir un jeu de paramètres qui sera utilisé par toutes les autres options de ce module. On s'intéresse ici au cas d'un **groupe de variables explicatives orthogonales** (D-produits scalaires nuls, en particulier covariances nulles pour des variables centrées), de k variables à expliquer et de modèles de prédiction multivariés indépendants pour chacune des variables à prédire.



L'option utilise une seule fenêtre de dialogue :

X file: explanatory variables	<input type="text" value="DateGraph.gax"/>	19	4
Y file: dependent variables	<input type="text" value="Gre"/>	19	14
Y transformation (default = none)	<input type="text" value="1"/>		
Option: row weighting	<input type="text" value="DateGraph.gpl"/>	19	1
Output file name	<input type="text" value="R"/>		

Buttons: Quit, Ok

Nom du fichier binaire contenant les variables explicatives.

Nom du fichier binaire contenant les variables à prédire.

Type de transformation à faire subir aux variables explicatives :

Enter value:

Y transformation (default = none)
0 = no transformation
1 = D-centring
2 = D-standardization
3 = D-normalization

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

Nom générique des fichiers de sortie.



Utiliser la carte Grèbes de la pile ADE-4•Data pour obtenir les fichiers date (19-1) et Gre (19-14). Représenter (Curves : Lines) les variables de Gre en fonction de la variable unique de date (Cf. la fiche de UniVarReg : Initialize).

Calculer par Distances : Table To Distance Matrix la matrice des distances entre dates :

Input file	<input type="text" value="date"/>	19	1
Option: Output file	<input type="text"/>		
Option: default = between rows	<input type="text"/>		
Index use (no default)	<input type="text" value="2"/>		

Buttons: Quit, Ok

Calculer par Distances : Minimal Spanning Tree un graphe de voisinage entre dates :

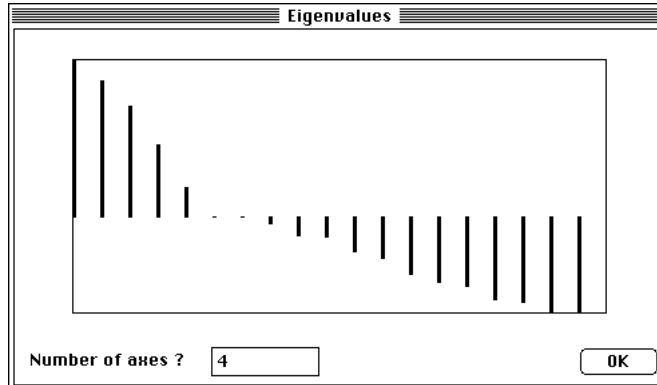
Distances input file	<input type="text" value="date_EU"/>	19	19
Component number (default=1)	<input type="text" value="2"/>		
Option: Output file	<input type="text" value="DateGraph"/>		

Buttons: Quit, Ok

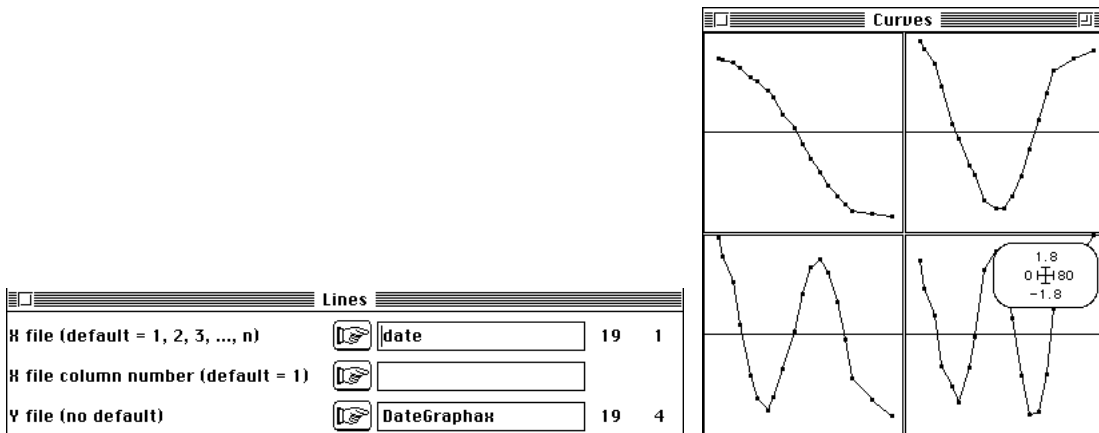
Calculer par NGStat : Moran EigenVectors les vecteurs propres associés à ce graphe de voisinage :



Conserver les quatre premiers :



Représenter ces quatre vecteurs en fonction du temps (Curves : Lines) :



Associer ces quatre variables explicatives orthonormales par la présente option (ci-dessus). On obtient :

```
-----
New TEXT file A.OVpa contains the parameters:
----> Explanatory variables: DateGraphax [19][4]
----> Dependent variable file: Gre [19][14]
----> Transformation used: 1
      0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
----> Row weighting file: DateGraphpl
-----
```

File A.OVcs contains the cosinus square between explanatory and dependent variables:

```
----> 4 rows (explanatory variables)
----> 14 columns (dependent variables)
```

```
*-----*
```

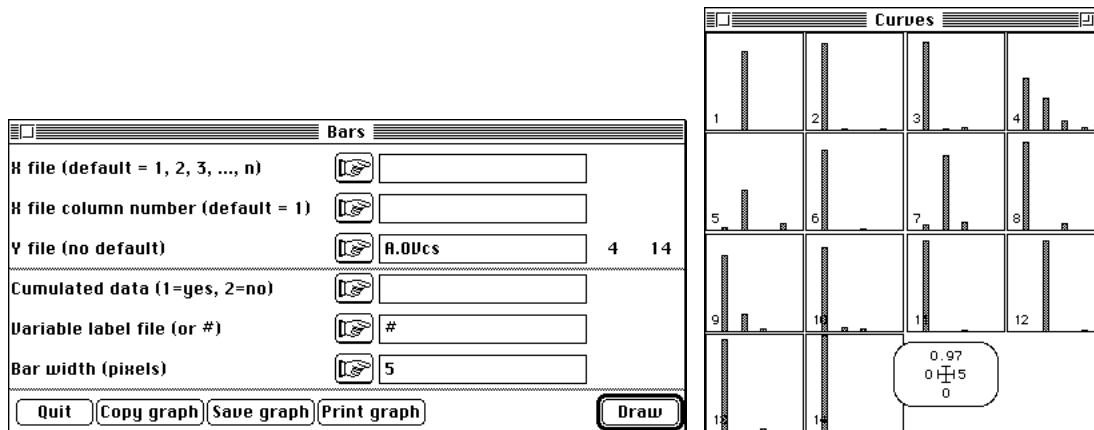
N°	Variance	Explained	Ratio
1	9.940e+02	7.884e+02	7.932e-01
2	2.401e+02	2.178e+02	9.071e-01
3	4.317e+02	4.073e+02	9.435e-01
4	2.095e+03	2.024e+03	9.664e-01
5	6.915e+02	3.519e+02	5.090e-01
6	1.016e+02	8.492e+01	8.362e-01
7	1.000e+02	9.036e+01	9.036e-01
8	1.309e+02	1.272e+02	9.712e-01
9	1.526e+02	1.484e+02	9.724e-01
10	3.054e+02	2.860e+02	9.366e-01
11	1.583e+02	1.488e+02	9.403e-01

12	3.737e+03	3.518e+03	9.414e-01
13	1.249e+02	1.209e+02	9.677e-01
14	8.986e+01	8.679e+01	9.659e-01

On entre dans les autres options du module par le fichier ---.OVpa ainsi créé.

Le listing donne le pourcentage de variance expliquée par la régression sur les quatre variables (minimum 51% pour la courbe 5, maximum 97% pour la 4). On peut s'interroger sur le rôle de chacune des explicatives dans la modélisation de chacune des chroniques. Les explicatives étant non corrélées, il n'y a aucune ambiguïté, car la variance expliquée s'additionne purement et simplement et on peut décortiquer la constitution du pourcentage de variance expliquée (Cf. OrthoVar : Variable test).

Le fichier ---.OVcs permet de représenter le rôle de chaque variable :



Les pourcentages de variance expliquée s'additionnent entre les explicatives, car elles sont orthogonales. La variable 1 sera prédite par le second vecteur propre, la variable 2 le sera par le premier, il en faudra plusieurs pour la variable 4 et la variable 5 ne sera pas bien prédite, ...



Le module permet de manipuler des explicatives orthogonales, c'est-à-dire, dans la majorité des cas des **variables explicatives non corrélées**. Ceci recouvre les vecteurs propres des graphes de voisinages¹ (qui généralisent les méthodes de Trend Surface Analysis²) et les coordonnées factorielles des analyses classiques. Le deuxième cas le plus connu est appelé régression sur composantes (PCR)³ étendue en écologie au cas de l'AFC dans ⁴. Ces régressions ont des propriétés de stabilité numérique qui les apparentent aux régressions PLS⁵ (Cf. LinearReg). Pour en souligner l'intérêt nous les avons mis en évidence dans un module séparé, bien qu'il s'agisse d'un cas particulier des régressions linéaires multiples (MLR, cf. LinearReg).



Les tests sur les variables explicatives dans ce module envoient les messages :

- Sorry: Null weight found for row number (poids nul) ;
- Sorry: null vector found in explanatory data file (variable nulle) ;
- Sorry: non orthogonal vectors found in explanatory data file
It may be useful to verify weighting file (variables non orthogonales) ;

Ne pas oublier que l'orthogonalité se définit par rapport à une pondération. En particulier, utiliser les pondération de voisinages pour les vecteurs propres des graphes de voisinages et la pondération de l'analyse pour la régression sur composantes après une AFC.

Si les explicatives sont centrées, ne pas oublier d'utiliser le centrage ou la normalisation sur les variables à expliquer ou centrer au préalable avec la même pondération (FilesUtil : Centring).



¹ Thioulouse, J., Chessel, D & Champely, S. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics* : 2, 1-14.

² Gittins, R. (1968) Trend-surface analysis of ecological data. *Journal of Ecology* : 56, 845-869.

Wartenberg, D.E. (1985a) Canonical trend surface analysis: a method for describing geographic pattern. *Systematic Zoology* : 34, 259-279.

Burrough, P. A. (1987) Spatial variation in vegetation, soil and precipitation across the Hawkesbury Sandstone Plateau in NewSouth Wales, Australia: an analysis using ordination and trend surfaces. In : *Data Analysis in Community and Landscape Ecology*. Jongman, R. H. G. , Ter Braak, C. J. F. & van Tongeren, O. F. R. (Eds.) Pudoc, Wageningen, The Netherlands. 252-257.

³ Næs, T. (1984) Leverage and influence measures for principal component regression. *Chemometrics and Intelligent Laboratory Systems* : 5, 155-168.

⁴ Roux, M. (1973) Estimation des paléoclimats d'après l'écologie des foraminifères [Paléoclimats]. *Les Cahiers de l'Analyse des Données* : 4, 61-79.

⁵ Ter Braak, C.J.F., Juggins, S., Birks, H.J.B. & Voet, H. Van der. (1993) Weighted averaging partial least squares regression (WA-PLS): definition and comparison with other methods for species-environment calibration. In : *Multivariate Environmental Statistics*. Patil, G.P. & Rao, C.R. (Eds.) Elsevier Science Publishers. 525-560.

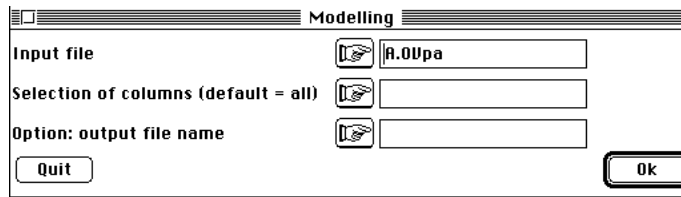
OrthoVar : Modelling



Régression sur variables orthogonales : calcul des prédictions.



L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par OrthoVar : Initialize.



Sélection de variables explicatives (format standard, par défaut toutes les variables sont sélectionnées).



Nom générique des fichiers de sortie (par défaut, il est extrait du nom du fichier d'entrée).



Utiliser l'exemple introduit dans la fiche de OrthoVar : Initialize :

```
-----> Explanatory variables: DateGraph.gax
-----> Dependent variable file: Gre
-----> Transformation used: 1
        0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
-----> Row weight file: DateGraph.gpl
-----> Selection of explanatory variables: 1a4
-----
```

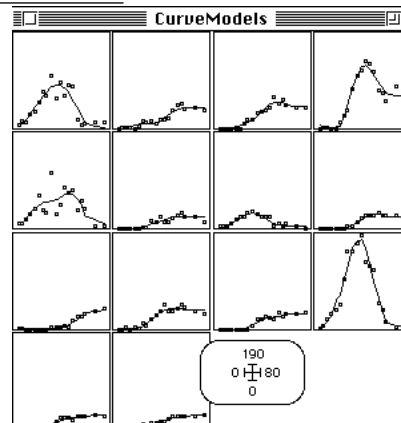
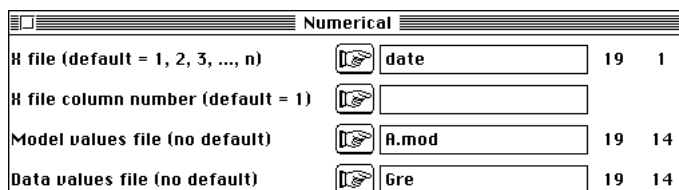
File A.mod has 19 rows and 14 columns
It contains the linear models variable resulting
from the separate multiple linear regression of each dependent variable
upon the set of explanatory variables

```
File :A.mod
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1| 1.430e-01| 8.518e+01|
...
| 14|-1.243e+00| 2.653e+01|
|----|-----|-----|
```

File A.res has 19 rows and 14 columns
It contains (data - model) matrix

```
File :A.res
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-2.686e+01| 2.827e+01|
...
| 14|-3.165e+00| 3.630e+00|
|----|-----|-----|
```

Représenter données et modèles par CurveModels : Numerical :



OrthoVar : Partial test



Régression sur explicatives orthogonales : test de signification partielle d'un sous-ensemble de variables explicatives sachant l'effet d'un autre sous-ensemble d'explicatives.

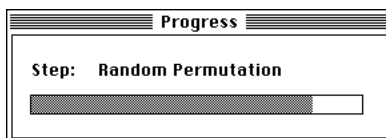


Test de randomisation : à chaque simulation (permutation aléatoire des lignes du tableau des explicatives), on calcule le pourcentage de variance expliquée par la régression multiple sur le sous-ensemble d'explicatives B du résidu de la prédiction par la régression multiple sur le sous-ensemble d'explicatives A (test de l'effet B/A) et on édite la fréquence des simulations donnant une meilleure prédiction que l'appariement réel.



L'option utilise une seule fenêtre de dialogue :

- Fichier des paramètres créé par OrthoVar : Initialize.
- Sélection de variables explicatives définissant le sous-espace A (format standard).
- Sélection de variables explicatives définissant le sous-espace B (format standard).
- Nombre de simulations utilisé.



Utiliser l'exemple introduit dans la fiche de OrthoVar : Initialize :

```
-----> Explanatory variables: DateGraphax
-----> Dependent variable file: Gre
-----> Transformation used: 1
          0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
-----> Row weight file: DateGraph.gpl
-----> Number of random permutations: 5000
-----> Effect A: selection of explanatory variables: 1;2
-----> Effect B: selection of explanatory variables: 3;4
-----
```

	VarY r2 observ.	mean sim.	normal I	X>Xobs	Frequency
1	7.137e-03	1.077e-01	-1.046e+00	4681	9.362e-01
2	1.848e-01	1.078e-01	7.782e-01	933	1.866e-01
3	3.796e-01	1.083e-01	2.873e+00	82	1.640e-02
4	7.884e-01	1.072e-01	6.819e+00	0	0.000e+00
5	1.400e-01	1.094e-01	3.144e-01	1490	2.980e-01
6	1.036e-01	1.084e-01	-5.124e-02	2067	4.134e-01
7	4.974e-01	1.073e-01	4.144e+00	18	3.600e-03
8	7.436e-01	1.083e-01	6.753e+00	0	0.000e+00
9	5.382e-01	1.081e-01	4.599e+00	10	2.000e-03
10	3.931e-01	1.082e-01	2.966e+00	79	1.580e-02
11	3.740e-01	1.098e-01	2.668e+00	113	2.260e-02
12	3.358e-01	1.080e-01	2.306e+00	180	3.600e-02
13	5.674e-01	1.087e-01	4.661e+00	5	1.000e-03
14	4.022e-02	1.082e-01	-7.247e-01	3643	7.286e-01

Outre la fréquence de dépassement de l'observation, le listing donne la moyenne observée sur l'ensemble des simulations et l'approximation normale (observation - moyenne / écart-type).

On permute ici les résidus de la prédiction par les variables 1 et 2 pour tester strictement l'effet des variables 3 et 4 sur le reste. L'effet n'est pas négligeable dans plusieurs des cas et on décide de conserver les quatre variables explicatives (OrthoVar : Modelling).



Voir un autre exemple dans la fiche 1 du fascicule 5 de la documentation thématique (p. 23).

OrthoVar : Subspace test



Régression sur explicatives orthogonales : test de signification d'un sous-ensemble de variables explicatives.



Test de randomisation : à chaque simulation (permutation aléatoire des lignes du tableau des explicatives), on calcule le pourcentage de variance expliquée par la régression multiple sur le sous-ensemble d'explicatives et on édite la fréquence des simulations donnant une meilleure prédiction que l'appariement réel.




L'option utilise une seule fenêtre de dialogue :

Subspace test

Input file

Selection of columns (default = all)

Number of permutations

 Fichier des paramètres créé par OrthoVar : Initialize.

 Sélection de variables explicatives testées (format standard).

 Nombre de simulations utilisé.

Progress

Step: Random Permutation



Utiliser l'exemple introduit dans la fiche de OrthoVar : Initialize :

```
-----> Explanatory variables: DateGraphax
-----> Dependent variable file: Gre
-----> Transformation used: 1
      0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
-----> Row weight file: DateGraph.gpl
-----> Number of random permutation: 5000
-----> Selection of explanatory variables: 1;2
-----
```

VarY r2 observ.	mean sim.	normal I	X>Xobs	Frequency
1 7.917e-01	1.054e-01	7.020e+00	0	0.000e+00
2 8.861e-01	1.056e-01	8.068e+00	0	0.000e+00
3 9.090e-01	1.052e-01	8.317e+00	0	0.000e+00
4 8.413e-01	1.068e-01	7.594e+00	1	2.000e-04
5 4.290e-01	1.062e-01	3.428e+00	38	7.600e-03
6 8.173e-01	1.060e-01	7.400e+00	0	0.000e+00
7 8.081e-01	1.061e-01	7.213e+00	0	0.000e+00
8 8.878e-01	1.067e-01	7.964e+00	0	0.000e+00
9 9.401e-01	1.064e-01	8.673e+00	0	0.000e+00
10 8.955e-01	1.064e-01	8.106e+00	0	0.000e+00
11 9.046e-01	1.069e-01	8.146e+00	0	0.000e+00
12 9.117e-01	1.059e-01	8.337e+00	0	0.000e+00
13 9.253e-01	1.069e-01	8.334e+00	0	0.000e+00
14 9.644e-01	1.055e-01	8.930e+00	0	0.000e+00

Outre la fréquence de dépassement de l'observation, le listing donne la moyenne observée sur l'ensemble des simulations et l'approximation normale (observation - moyenne / écart-type).

OrthoVar : Variable test



Régression sur explicatives orthogonales : test de signification des variables.



Test de randomisation : à chaque simulation (permutation aléatoire des lignes du tableau des explicatives), on calcule le pourcentage de variance expliquée par chacune des explicatives et on édite la fréquence des simulations donnant une meilleure prédiction que l'appariement réel.



L'option utilise une seule fenêtre de dialogue :

Variable test

Input file R.0Upa

Number of permutations 5000

Quit Ok

Fichier des paramètres créé par OrthoVar : Initialize.

Nombre de simulations utilisé.

Progress

Step: Random Permutation



Utiliser l'exemple introduit dans la fiche de OrthoVar : Initialize :

```
-----> Explanatory variables: DateGraphax
-----> Dependent variable file: Gre
-----> Transformation used: 1
      0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
-----> Row weight file: DateGraph.gpl
-----> Number of random permutations: 5000
-----
```

Dependent variable number: 1

VarX	r2	observ.	X>Xobs	Frequency
1	4.966e-03	3812	7.624e-01	
2	7.868e-01	0	0.000e+00	
3	1.448e-03	4368	8.736e-01	
4	3.807e-05	4911	9.822e-01	

Dependent variable number: 2

VarX	r2	observ.	X>Xobs	Frequency
1	8.709e-01	0	0.000e+00	
2	1.519e-02	2981	5.962e-01	
3	4.983e-03	3831	7.662e-01	
4	1.607e-02	2975	5.950e-01	

...

Dependent variable number: 4

VarX	r2	observ.	X>Xobs	Frequency
1	5.197e-01	6	1.200e-03	
2	3.215e-01	47	9.400e-03	
3	9.700e-02	953	1.906e-01	
4	2.814e-02	2401	4.802e-01	

...