

UNIWIN VERSION 9.7.0

ANALYSE EN COMPOSANTES INDEPENDANTES

Révision : 02/09/2023

Définition.....	1
Entrée des données	1
Données manquantes	3
Exemple 1 : Fichier FastICA.....	3
L'option Rapports	6
L'option Graphiques	8
Exemple 2 : Fichier Cocktail	11
Exemple 3 : Fichier ECG	19
Les variables internes créées par la procédure	22
Références	23

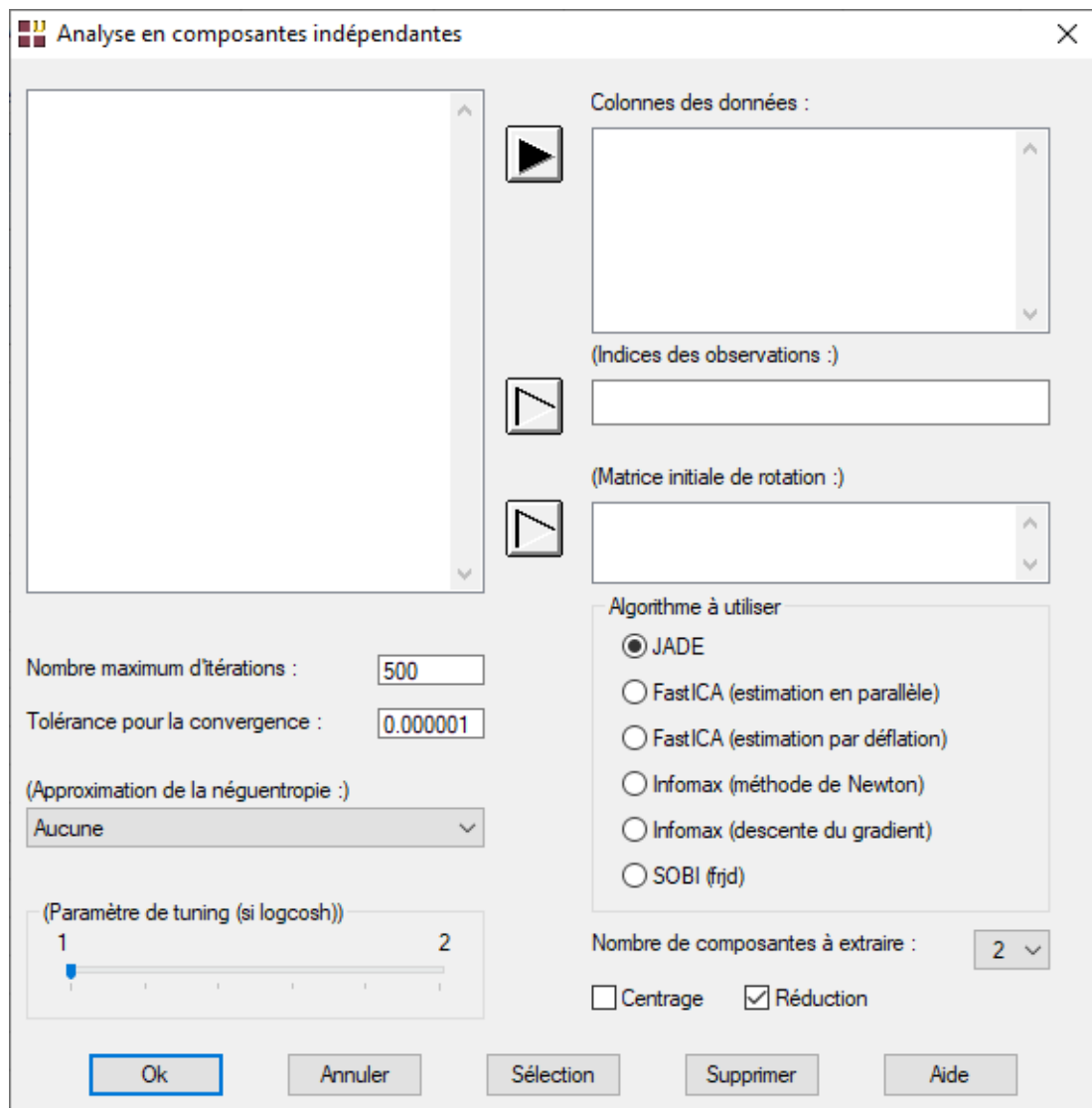
Définition

L'Analyse en composantes indépendantes (Independent Component Analysis - ICA) est une méthode d'analyse des données qui relève des statistiques, des réseaux de neurones et du traitement du signal. Elle est notoirement et historiquement connue en tant que méthode de séparation aveugle de signaux mélangés dans des proportions inconnues et est aujourd'hui appliquée à de nombreux problèmes, notamment en chimométrie pour l'analyse de spectres (source Wikipédia).

Cette procédure du logiciel utilise les packages R 'ica' et 'JADE'.

Entrée des données

Cliquons sur l'icône ICA dans le ruban Décrire. La boîte de dialogue montrée ci-dessous s'affiche :



Cette boîte de dialogue permet de préciser les colonnes des données à analyser, les noms des variables optionnelles contenant les indices des observations et la matrice initiale de rotation.

Elle permet également de préciser :

- l'algorithme à utiliser : JADE (Joint Approximate Diagonalization of Eigenmatrices), Fastlca (estimation en parallèle ou par déflation), Infomax (Information Maximization, méthode de Newton ou descente du gradient) ou SOBI (Second Order Blind Identification)
- le nombre de composantes indépendantes à extraire
- le centrage et/ou la réduction des données ou non
- le nombre maximum d'itérations de l'algorithme
- la tolérance pour la convergence de l'algorithme
- la technique d'approximation de la néguentropie pour les algorithmes Fastlca et Infomax
- le paramètre de tuning si l'option Fastlca - logcosh est utilisé pour approximer la néguentropie

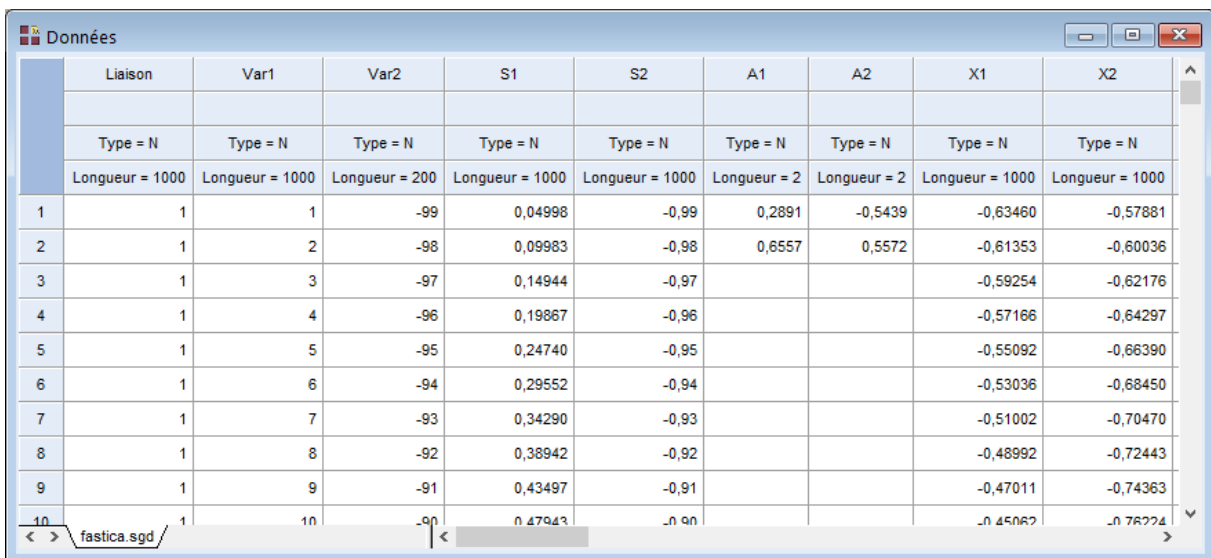
Données manquantes

Les données manquantes ne sont pas autorisées dans les colonnes des données, dans les indices des observations et dans la matrice initiale de rotation.

Exemple 1 : Fichier FastICA

Illustrons cette technique par un premier exemple dans lequel les colonnes S1 et S2 du tableur ci-dessous contiennent les signaux sources inconnus, les colonnes A1 et A2 la matrice de mélange de ces signaux inconnus et les colonnes X1 et X2 les signaux mélangés observés (source : <http://rpubs.com/skydome20/93614>).

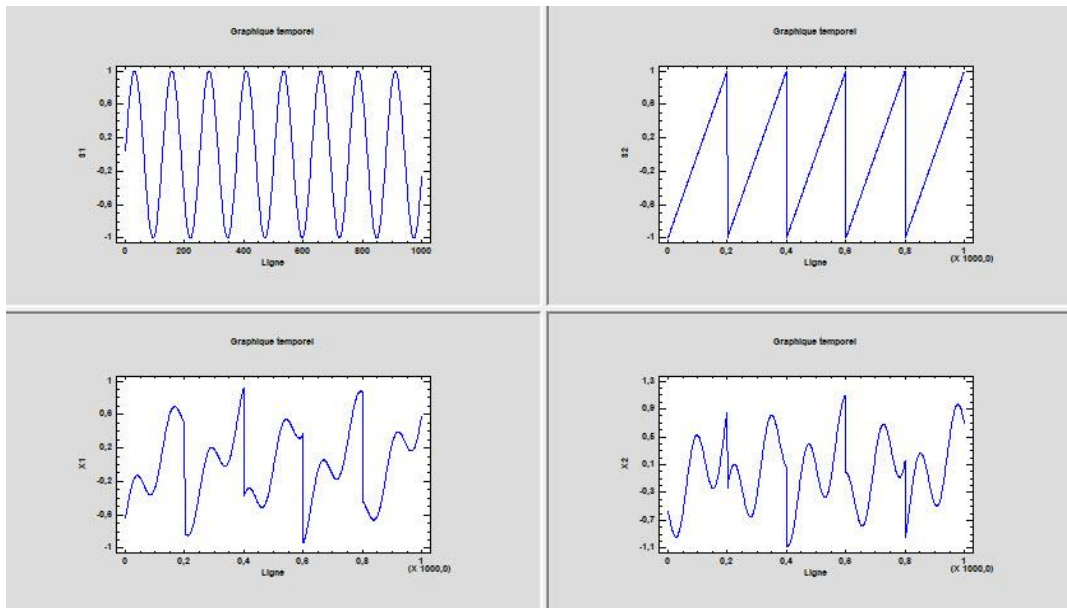
Les données du tableur sont contenues dans le fichier FastICA.sgd.



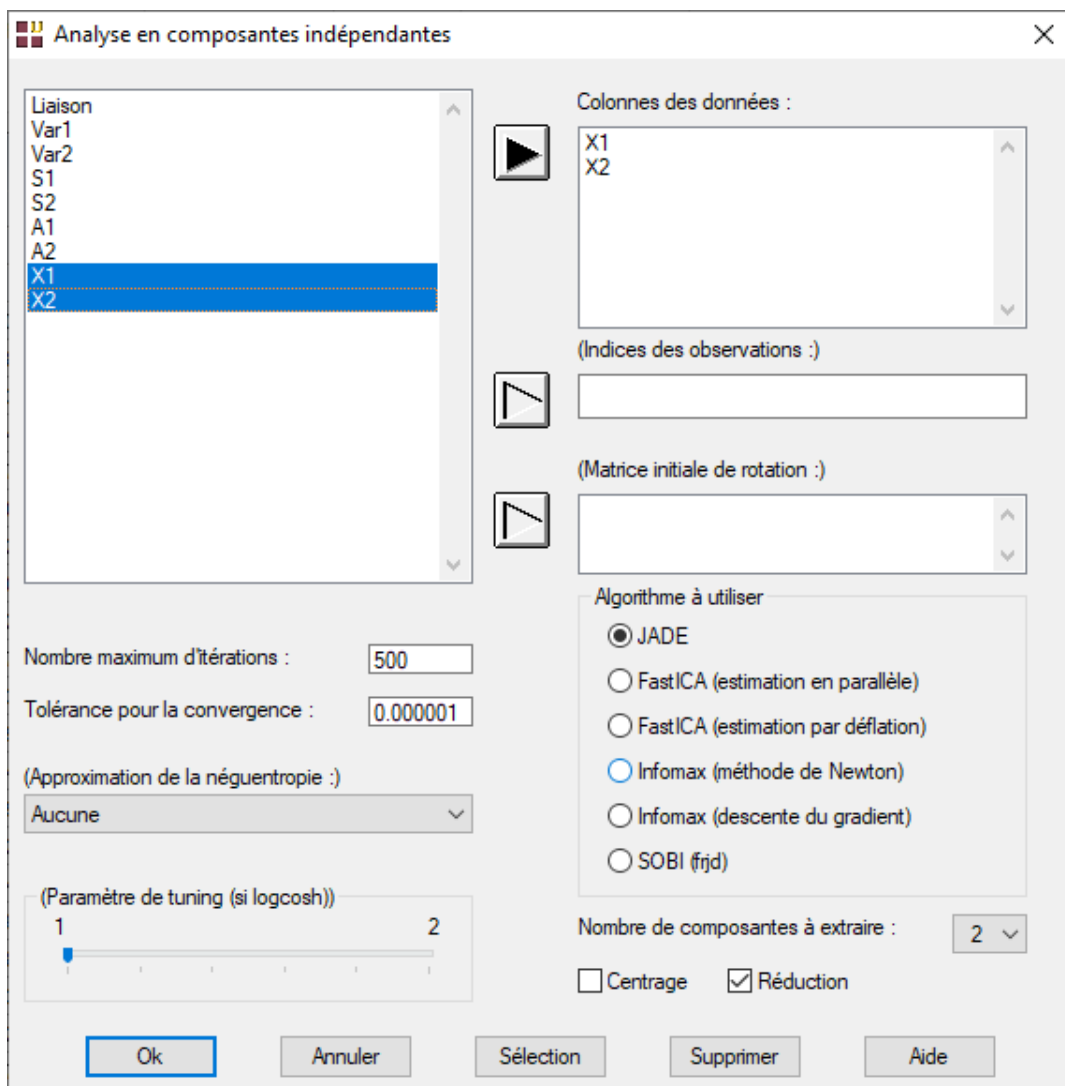
	Liaison	Var1	Var2	S1	S2	A1	A2	X1	X2
	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N	Type = N
	Longueur = 1000	Longueur = 1000	Longueur = 200	Longueur = 1000	Longueur = 1000	Longueur = 2	Longueur = 2	Longueur = 1000	Longueur = 1000
1	1	1	-99	0,04998	-0,99	0,2891	-0,5439	-0,63460	-0,57881
2	1	2	-98	0,09983	-0,98	0,6557	0,5572	-0,61353	-0,60036
3	1	3	-97	0,14944	-0,97			-0,59254	-0,62176
4	1	4	-96	0,19867	-0,96			-0,57166	-0,64297
5	1	5	-95	0,24740	-0,95			-0,55092	-0,66390
6	1	6	-94	0,29552	-0,94			-0,53036	-0,68450
7	1	7	-93	0,34290	-0,93			-0,51002	-0,70470
8	1	8	-92	0,38942	-0,92			-0,48992	-0,72443
9	1	9	-91	0,43497	-0,91			-0,47011	-0,74363
10	1	10	-90	0,47943	-0,90			-0,45062	-0,76224

L'objectif est à partir de ces signaux X1 et X2 observés de retrouver au mieux les signaux inconnus d'origine S1 et S2.

Visualisons les signaux inconnus (première ligne) et les signaux observés (deuxième ligne) :

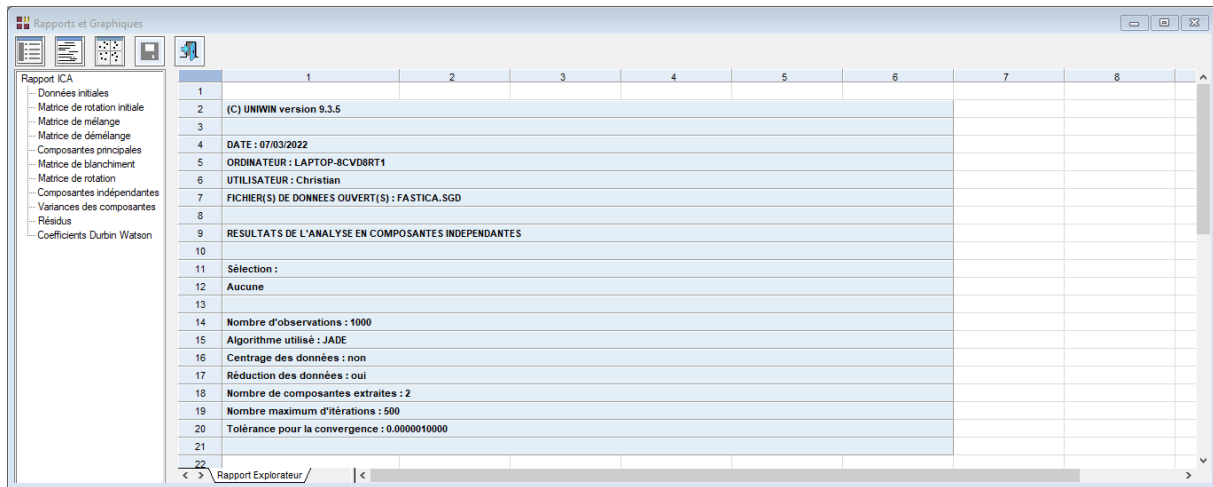



Cliquons sur l'icône 'ICA' dans le ruban 'Décrire' pour renseigner la boîte de dialogue d'entrée des données comme montré ci-après.




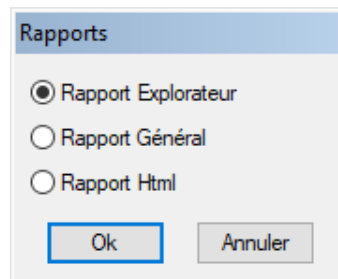
Choisissons l'algorithme JADE, les autres options aux valeurs par défaut et cliquons sur Ok.


Après quelques instants, la fenêtre Rapports et Graphiques s'affiche :

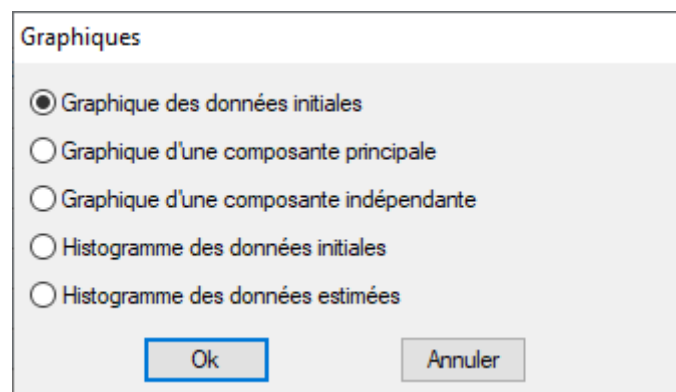



La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

L'icône 'Rapports'  affiche la boîte de dialogue des options pour les rapports :



et l'icône 'Graphiques'  affiche la boîte de dialogue des options pour les graphiques :



L'icône 'Enregistrer'  permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

Enregistrement des résultats (1/2)

Enregistrer

Composantes indépendantes

Matrice de démixage estimée

Liaisons entre les observations

Indices des observations

Composantes principales

Matrice de mélange estimée

Matrice de blanchiment

Matrice de rotation

Matrice des résidus

Coefficients de Durbin Watson

Noms attribués aux variables cibles

CI_1

W_1

liaisons

indices

CP_1

M_1

Q_1

R_1

RESID_1

DW

Ok Plus Annuler

Note : le bouton 'Plus' permet d'afficher la suite de la liste des variables.

L'icône 'Quitter'  permet de quitter l'analyse.

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

Voici trois exemples du rapport pour notre ICA : Explorateur, Général, HTML.

Rapports et Graphiques

Rapport ICA

- Données initiales
- Matrice de rotation initiale
- Matrice de mélange
- Matrice de démixage
- Composantes principales
- Matrice de blanchiment
- Matrice de rotation
- Composantes indépendantes**
- Variances des composantes
- Résidus
- Coefficients Durbin Watson

	1	2	3	4	5	6	7	8
1								
2	MATRICE DES COMPOSANTES INDEPENDANTES (DONNEES SOURCES)							
3								
4								
5		CL1	CL2					
6	1	1,71380	0,08111					
7	2	1,69560	0,15132					
8	3	1,67740	0,22118					
9	4	1,65921	0,29052					
10	5	1,64103	0,35915					
11	6	1,62285	0,42891					
12	7	1,60469	0,49363					
13	8	1,58655	0,55914					
14	9	1,56842	0,62328					
15	10	1,55032	0,68589					
16	11	1,53223	0,74680					
17	12	1,51417	0,80587					
18	13	1,49613	0,86295					
19	14	1,47812	0,91790					
20	15	1,46013	0,97058					
21	16	1,44218	1,02085					
22	17	1,42426	1,06859					

Rapport Explorateur

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	(C) UNIWIN version 9.3.3												
3													
4	DATE : 16/11/2021												
5	ORDINATEUR : LAPTOP-8CVD8RT1												
6	UTILISATEUR : Christian												
7	FICHIER(S) DE DONNEES OUVERT(S) : FASTICA.SGD												
8													
9	RESULTATS DE L'ANALYSE EN COMPOSANTES INDEPENDANTES												
10													
11	Sélection :												
12	Aucune												
13													
14	Nombre d'observations : 1000												
15	Algorithme utilisé : JADE												
16	Centrage des données : non												
17	Réduction des données : oui												
18	Nombre de composantes extraites : 2												
19	Nombre maximum d'itérations : 500												
20	Tolérance pour la convergence : 0.0000010000												
21													
22	Rapport Général /												

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	(C) UNIWIN version 9.3.5												
3													
4	DATE : 07/03/2022												
5	ORDINATEUR : LAPTOP-8CVD8RT1												
6	UTILISATEUR : Christian												
7	FICHIER(S) DE DONNEES OUVERT(S) : FASTICA.SGD												
8													
9	RESULTATS DE L'ANALYSE EN COMPOSANTES INDEPENDANTES												
10													
11	Sélection :												
12	Aucune												
13													
14	Nombre d'observations : 1000												
15	Algorithme utilisé : JADE												
16	Centrage des données : non												
17	Réduction des données : oui												
18	Nombre de composantes extraites : 2												
19	Nombre maximum d'itérations : 500												
20	Tolérance pour la convergence : 0.0000010000												
21													
22	MATRICE DES DONNEES INITIALES												
		X1	X2										
1		-0,63460	-0,57881										
2		-0,61353	-0,60036										

Ce rapport affiche :

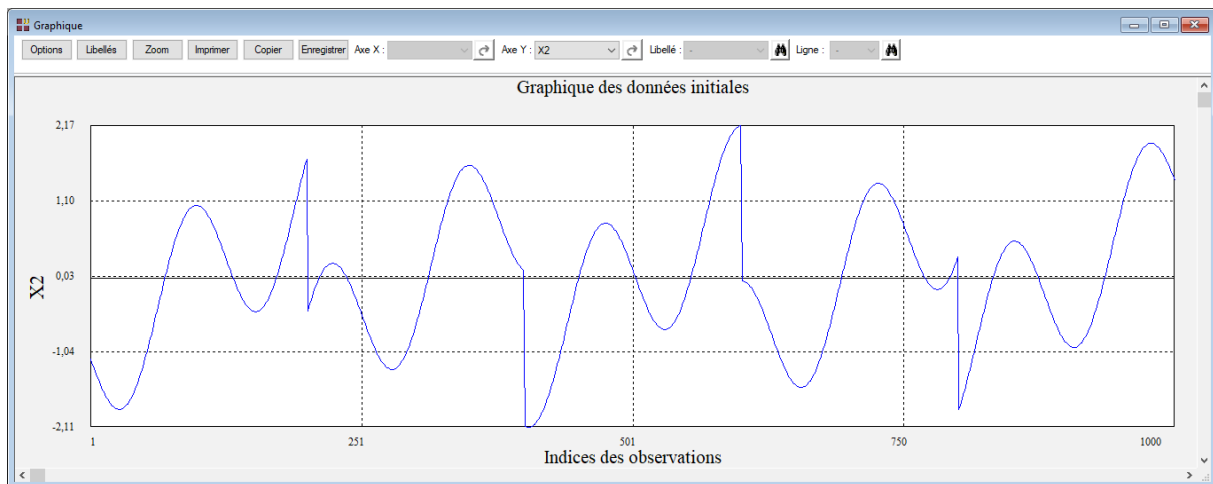
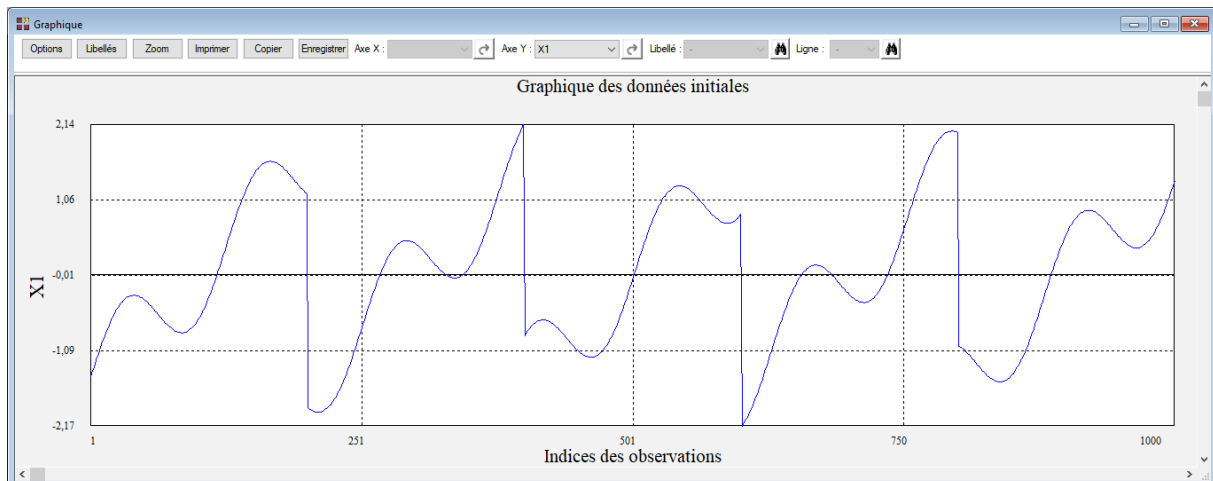
- les données initiales
- les données initiales centrées (si cette option a été choisie)
- la matrice de rotation initiale
- la matrice de mélange estimée
- la matrice de démixage estimée
- les composantes principales (données blanchies)
- la matrice de blanchiment estimée
- la matrice de rotation estimée
- les composantes indépendantes (données sources)
- les variances des composantes indépendantes
- les résidus
- les coefficients de Durbin Watson

L'option Graphiques

Cette option permet d'obtenir divers graphiques pour l'analyse ICA. Ils sont présentés ci-après.

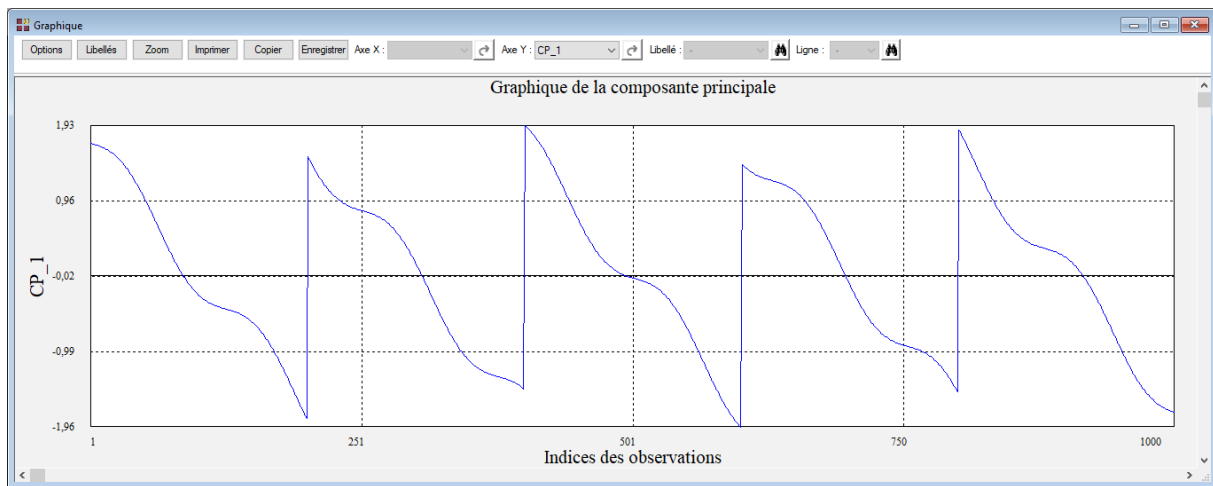
- Graphique des données initiales

Ce graphique permet de visualiser les données initiales à analyser à savoir dans cet exemple les variables X1 et X2.



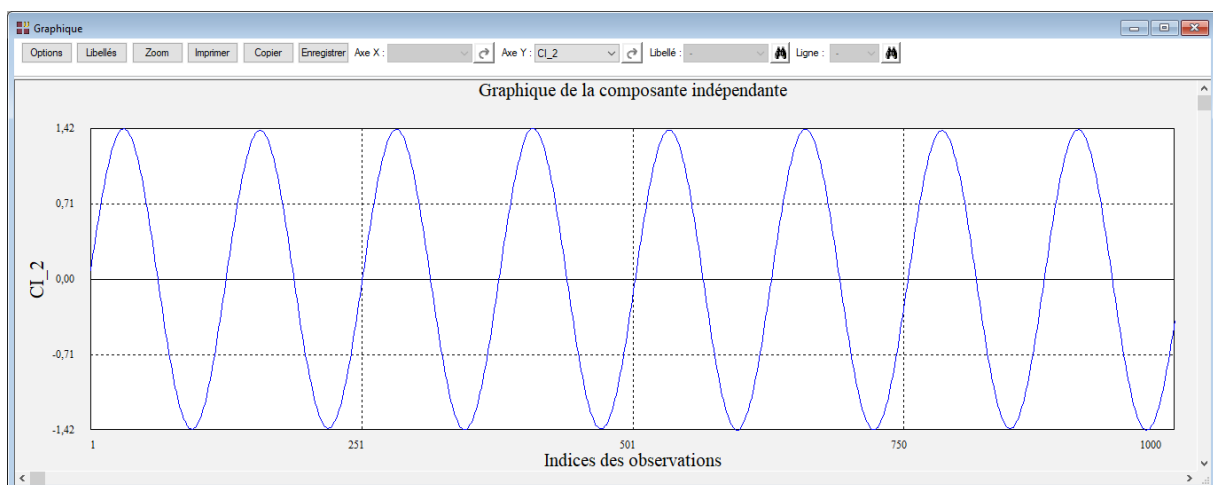
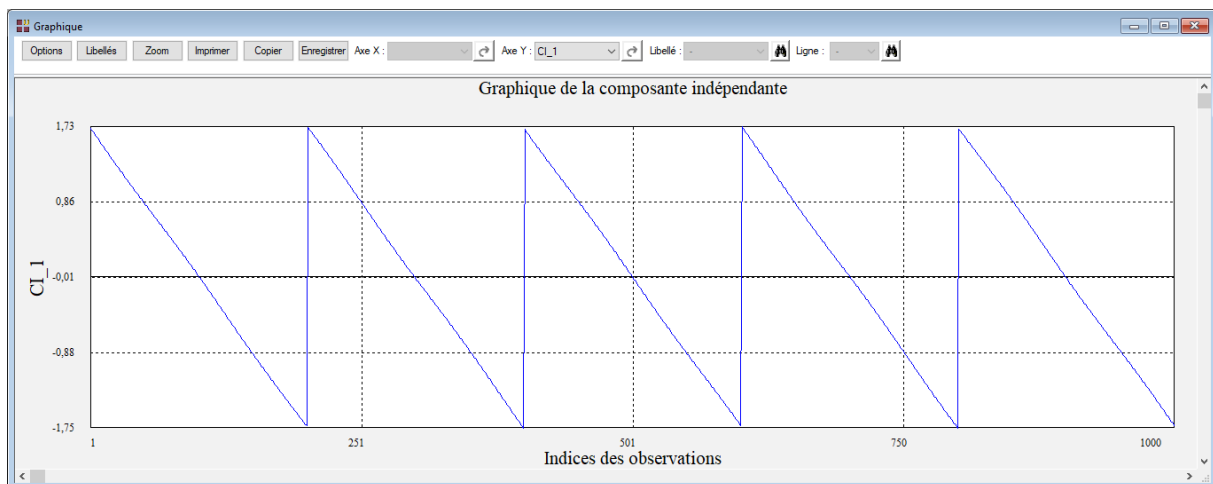
- Graphique d'une composante principale

Ce graphique affiche chacune des composantes principales issues de l'analyse.



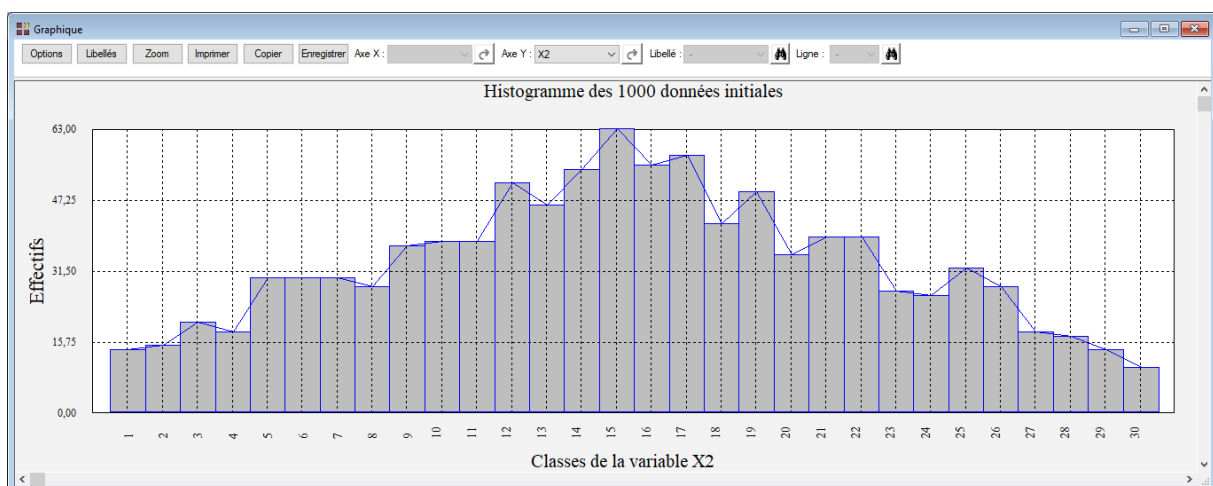
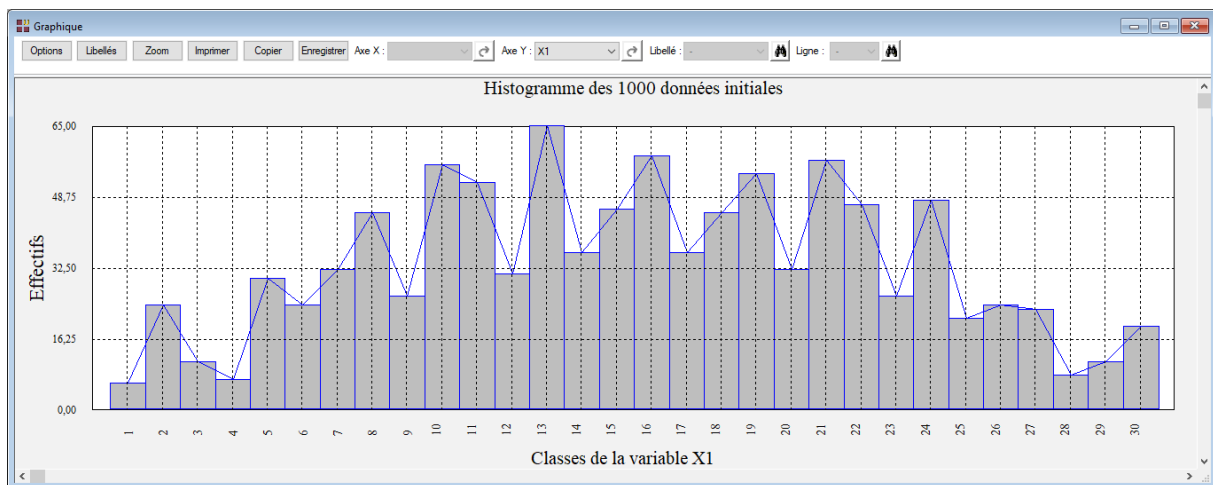
- Graphique d'une composante indépendante

Ce graphique affiche chacune des composantes indépendantes issues de l'analyse.



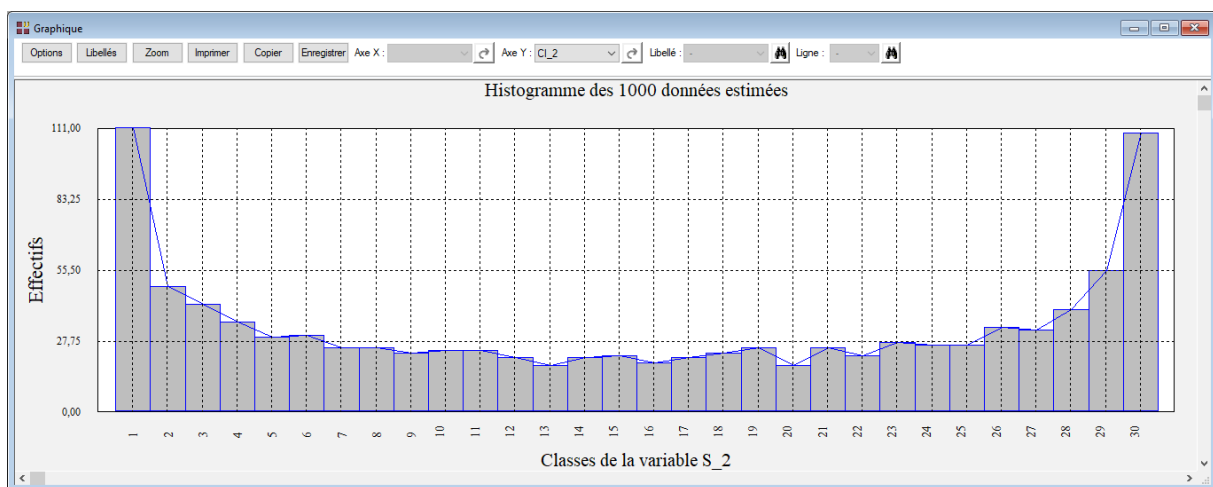
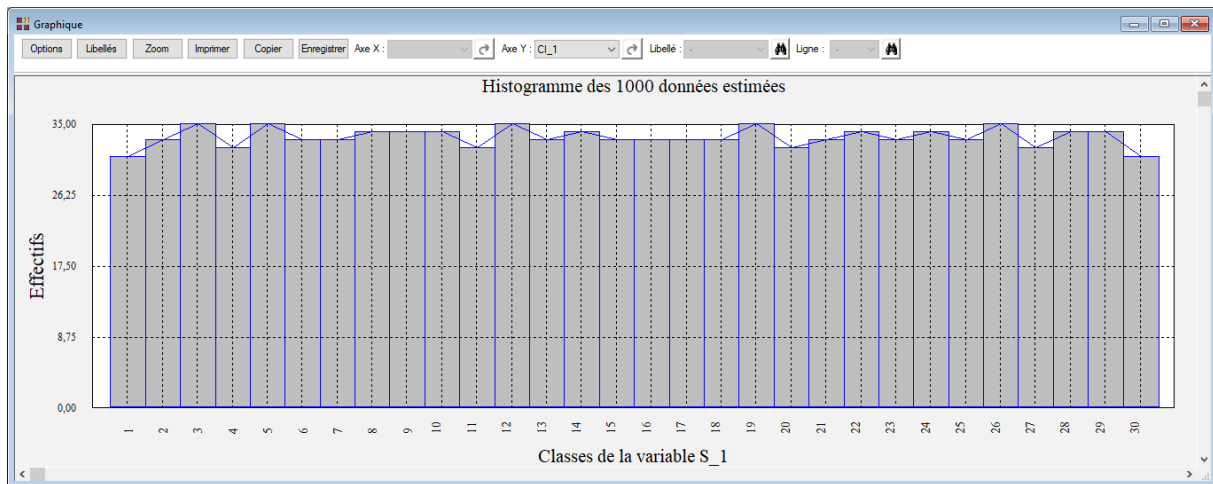
- Histogramme des données initiales

Ce graphique affiche un histogramme pour chacune des variables contenant les données initiales.



- Histogramme des données estimées


Ce graphique affiche un histogramme pour chacune des variables contenant les données estimées.



Exemple 2 : Fichier Cocktail

Un exemple classique d'application de la séparation aveugle de sources est le problème de la soirée cocktail. Pour séparer les voix de p personnes, il faut p microphones placés en différents endroits de la pièce. Les microphones enregistrent les mélanges des voix des p personnes et l'objectif est alors de retrouver les discours individuels à partir de ces mélanges.

Dans cet exemple, nous mélangeons les trois fichiers sons d'origine source5.wav, source7.wav et source9.wav, disponibles avec le package R 'JADE' et placés dans le répertoire des exemples d'UNIWIN, y ajoutons une composante de bruit et essayons de retrouver ces fichiers d'origine.

Pour écouter ces fichiers sons, vous pouvez cliquer sur l'icône  dans le ruban Accueil.

Pour utiliser ces fichiers dans UNIWIN, il faut préalablement les transformer en données numériques en utilisant la fonction readWave du package R 'tuneR'.

Pour cela, il faut exécuter les commandes suivantes dans lesquelles 'chemin' indique le chemin vers le répertoire contenant les fichiers :

```
if(!require('tuneR')) {install.packages('tuneR')}  
library("tuneR")  
S1 <- readWave("chemin/source5.wav")  
S2 <- readWave("chemin/source7.wav")  
S3 <- readWave("chemin/source9.wav")
```

Un composant de bruit est ajouté aux données :

```
set.seed(321)  
NOISE <- noise("white", duration = 50000)
```

Les données sont transformées pour avoir des variances unitaires :

```
Source <- cbind(S1@left, S2@left, S3@left, NOISE@left)  
Source <- scale(Source, center = FALSE, scale = apply(Source, 2, sd))  
Sourcets <- ts(Source, start = 0, frequency = 8000)
```

Les sources sont mélangées en utilisant une matrice de rotation dont les éléments sont générés à partir d'une loi uniforme :

```
p <- 4  
Rotation <- matrix(runif(p^2, 0, 1), p, p)  
Melange <- tcrossprod(Sourcets, Rotation)  
Melangets <- as.ts(Melange)
```

Les graphiques montrés ci-après affichent les données des sons d'origine et des sons après mélanges. Ils sont obtenus par les commandes R suivantes :

```
plot(Sourcets, main = "Données sources")  
plot(Melangets, main = "Données mélangées")
```

Pour cet exemple, supposons que nous avons observé les mélanges de signaux affichés dans la figure 2.

L'objectif est alors d'estimer les signaux d'origine (Figure 1) en se basant sur ces données observées (Figure 2).

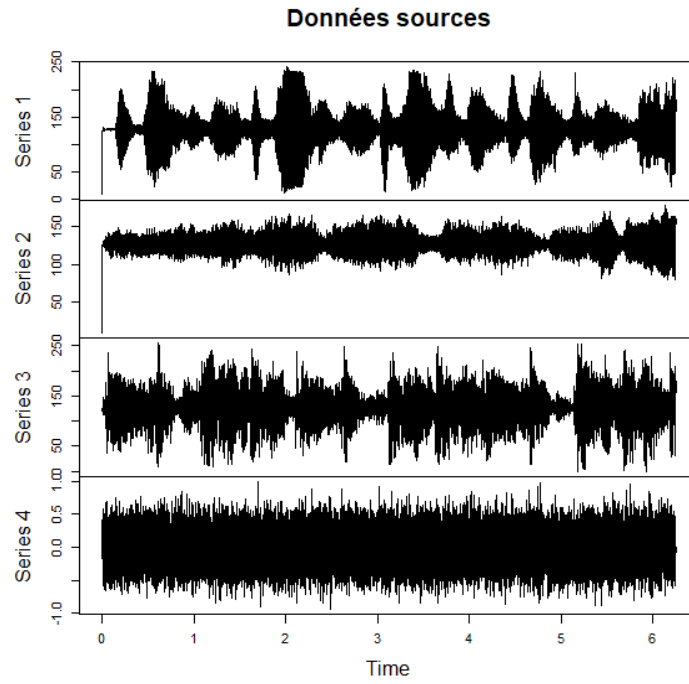


Figure 1 : Donnees sources + Composante de bruit

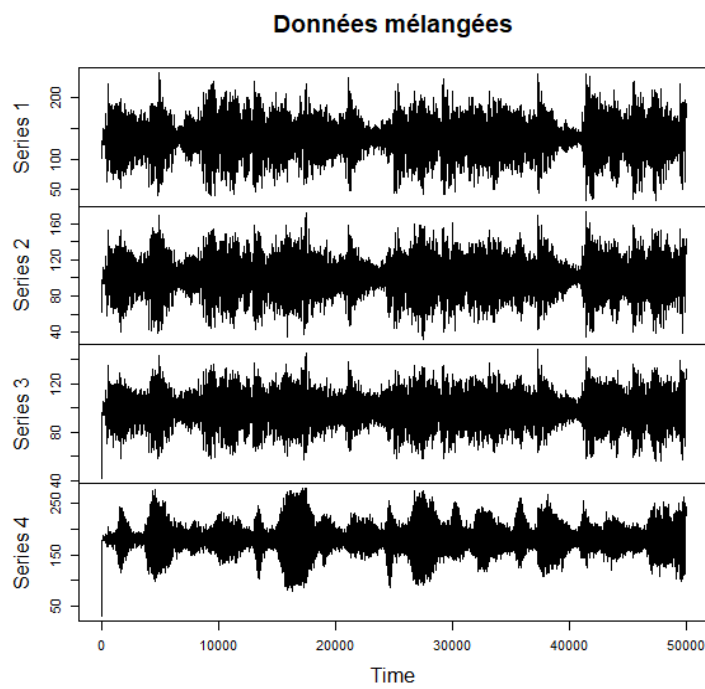


Figure 2 : Donnees melanges

Pour écouter les sons mélangés, utiliser les commandes R suivantes :


```
x1 <- normalize(Wave(left = Melange[, 1], samp.rate = 8000, bit = 8), unit = "8")
x2 <- normalize(Wave(left = Melange[, 2], samp.rate = 8000, bit = 8), unit = "8")
x3 <- normalize(Wave(left = Melange[, 3], samp.rate = 8000, bit = 8), unit = "8")
x4 <- normalize(Wave(left = Melange[, 4], samp.rate = 8000, bit = 8), unit = "8")
```

```
play(x1)
play(x2)
play(x3)
play(x4)
```

Les sons mélangés peuvent être enregistrés dans des fichiers par les commandes R suivantes utilisant le package 'tuneR' :

```
writeWave(x1, "chemin/x1.wav")
writeWave(x2, "chemin/x2.wav")
writeWave(x3, "chemin/x3.wav")
writeWave(x4, "chemin/x4.wav")
```

Ces fichiers sont placés dans le répertoire des exemples d'UNIWIN.

Pour écouter ces fichiers, vous pouvez cliquer sur l'icône  dans le ruban Accueil.

Cliquons sur l'icône 'ICA' dans le ruban 'Décrite' pour renseigner la boîte de dialogue d'entrée des données comme montré ci-après.

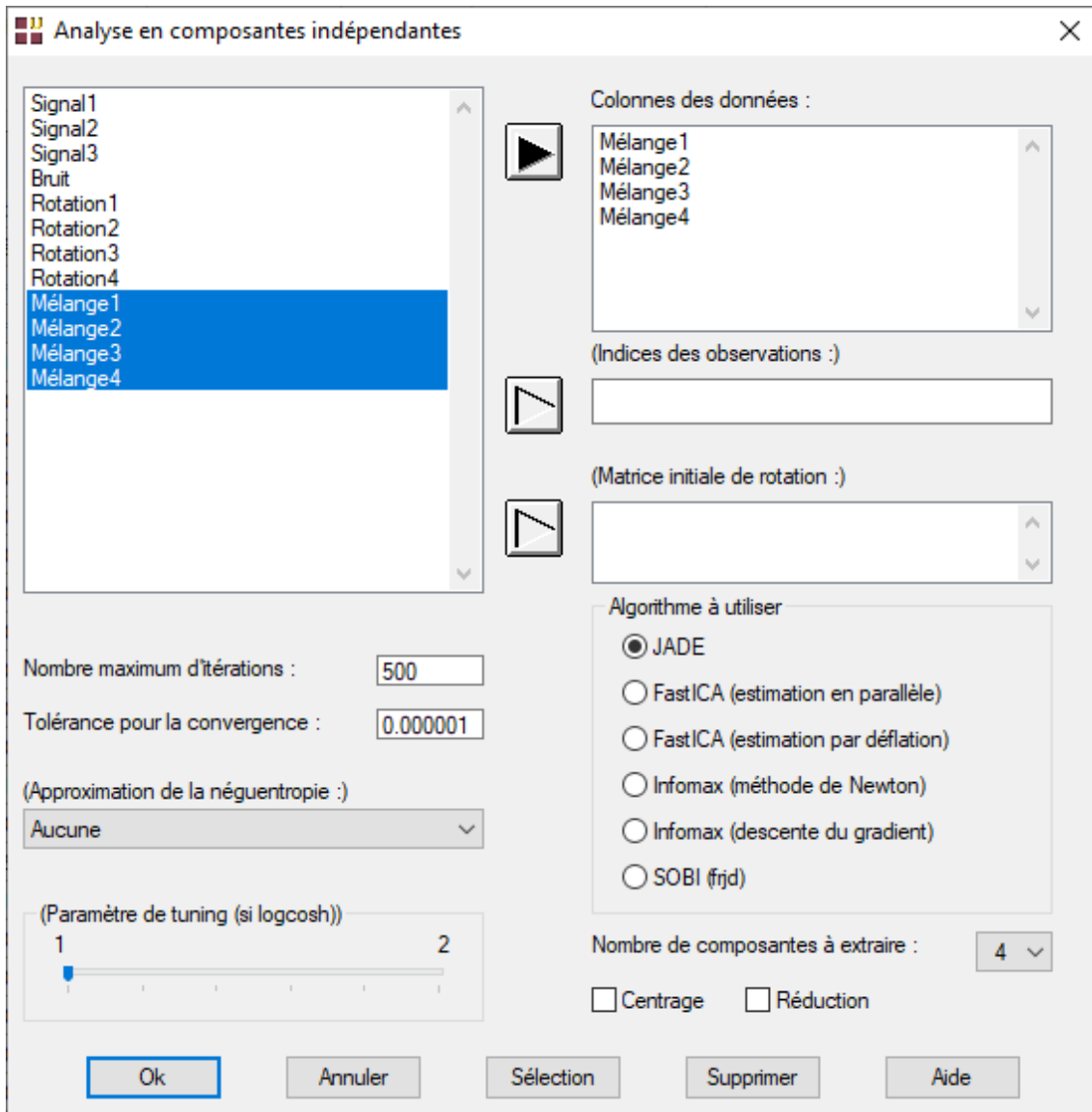
Les colonnes des données (signaux mélangés) sont les colonnes 'Mélange1', 'Mélange2', 'Mélange3' et 'Mélange4', chacune composée de 50000 observations.

La matrice initiale de rotation n'étant pas précisée, une matrice unitaire diagonale est utilisée.

Utilisation de l'algorithme JADE

Quatre composantes sont extraites et les données ne sont ni centrées ni réduites car les données initiales sont déjà réduites.

Après quelques instants pour effectuer les calculs, le rapport et les graphiques sont disponibles.



Rapports et Graphiques

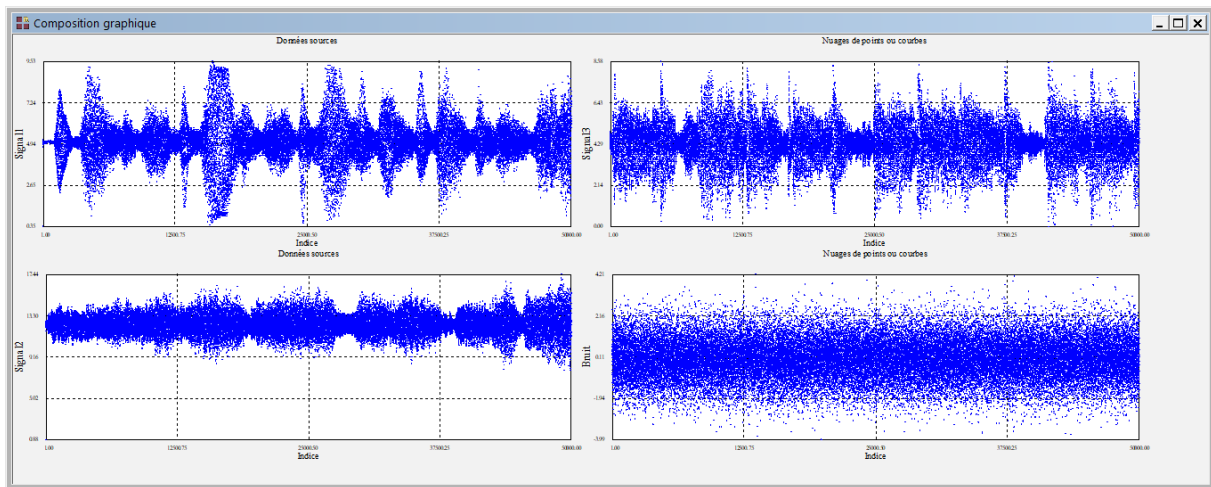
Rapport ICA

- Données initiales
- Matrice de rotation initiale
- Matrice de mélange
- Matrice de démélange
- Composantes principales
- Matrice de blanchiment
- Matrice de rotation
- Composantes indépendantes**
- Variances des composantes
- Résidus
- Coefficients Dubin Watson

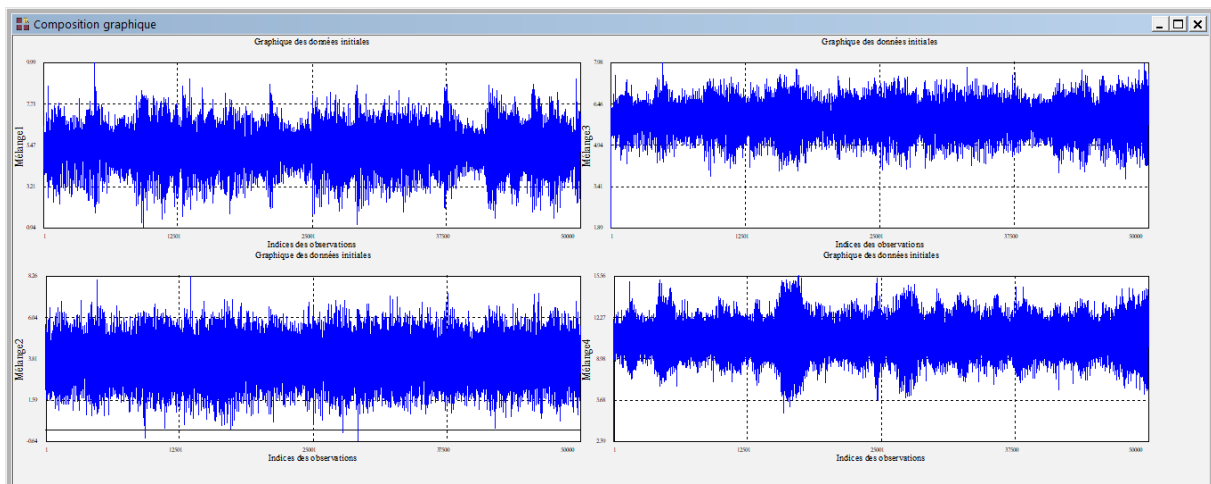
	1	2	3	4	5	6	7	8
1								
2	MATRICE DES COMPOSANTES INDEPENDANTES (DONNEES SOURCES)							
3								
4								
5		CL1	CL2	CL3	CL4			
6	1	-0,24329	1,84701	0,35996	3,58288			
7	2	-0,94291	-0,71351	-0,04182	-0,05473			
8	3	-1,01308	-0,28607	-0,05701	-0,23537			
9	4	-0,97624	-0,11263	-0,05429	0,14011			
10	5	-1,00795	-0,12941	-0,08646	-0,17922			
11	6	-0,97439	0,27360	-0,05495	0,09082			
12	7	-1,02238	0,73310	-0,05059	0,10088			
13	8	-0,99588	0,24982	-0,02790	0,37875			
14	9	-1,03962	0,35918	0,01742	0,46931			
15	10	-1,00004	-0,53853	-0,00542	0,31916			
16	11	-1,02466	0,35499	0,02367	0,15132			
17	12	-1,01237	1,50081	0,05597	0,35542			
18	13	-1,01177	0,19882	-0,04424	0,22100			
19	14	-1,00656	2,45541	-0,02082	0,21520			
20	15	-1,01467	-1,14074	-0,01273	0,31402			
21	16	-1,01262	-0,78730	0,00193	0,42588			
22	17	-1,00546	0,46728	-0,00248	0,24175			

Rapport Explorateur /

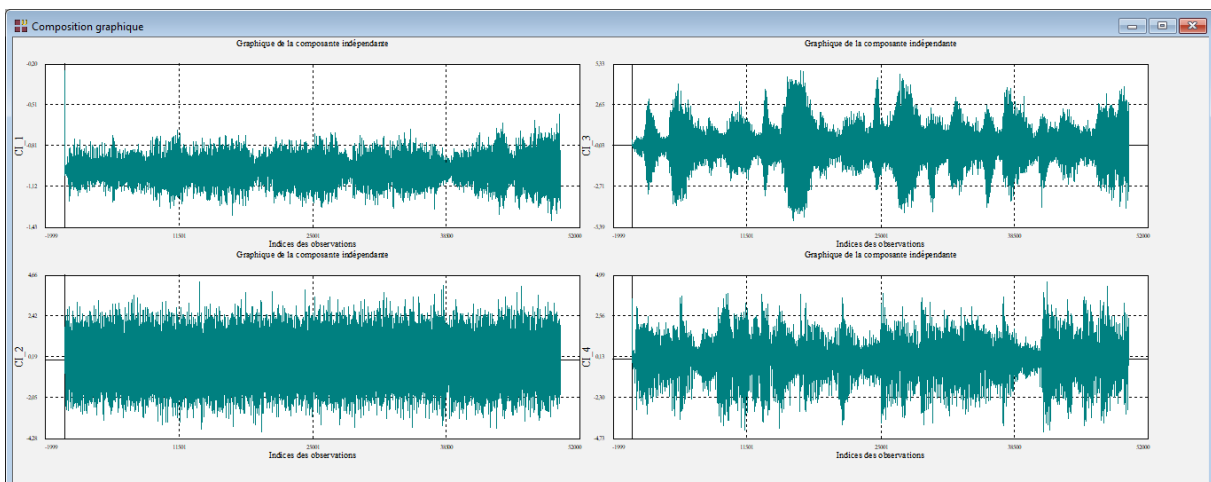
Voici les quatre graphiques des signaux sources affichés en utilisant la procédure 'Composition graphique' :



Voici les quatre graphiques des signaux mélangés :




Voici les quatre graphiques des signaux estimés :



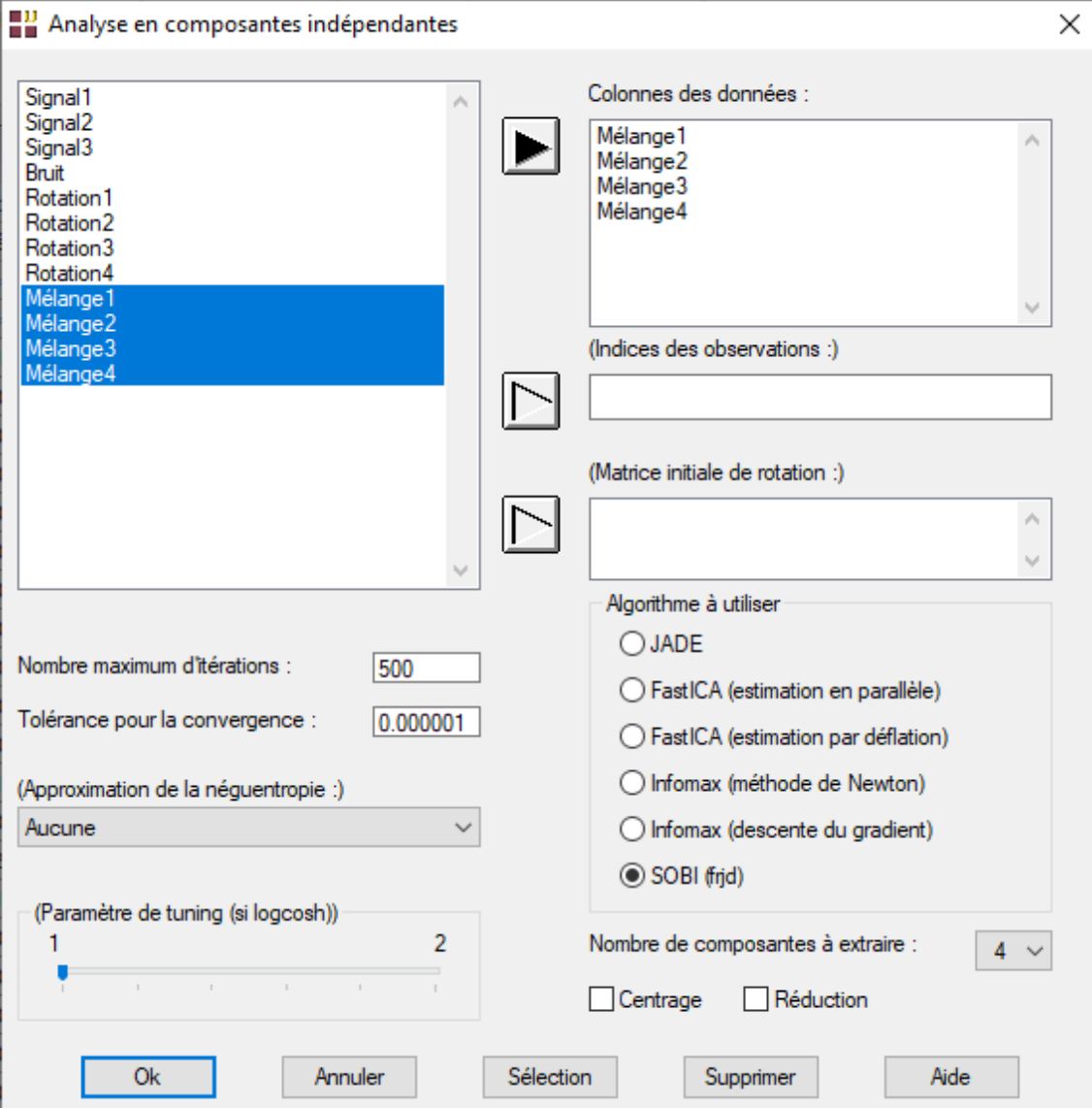
Ces quatre fichiers estimés sont placés dans le répertoire des exemples d'UNIWIN :

y1.wav
y2.wav
y3.wav
y4.wav

Pour écouter ces fichiers, vous pouvez cliquer sur l'icône  dans le ruban Accueil.

Utilisation de l'algorithme SOBI

Traitons ce même exemple en utilisant l'algorithme SOBI.



Analyse en composantes indépendantes

Signal1
Signal2
Signal3
Bruit
Rotation1
Rotation2
Rotation3
Rotation4
Mélange1
Mélange2
Mélange3
Mélange4

Colonnes des données :
Mélange1
Mélange2
Mélange3
Mélange4

(Indices des observations :)

(Matrice initiale de rotation :)

Algorithme à utiliser

JADE
 FastICA (estimation en parallèle)
 FastICA (estimation par déflation)
 Infomax (méthode de Newton)
 Infomax (descente du gradient)
 SOBI (frjd)

Nombre de composantes à extraire : 4

Centrage Réduction

Ok Annuler Sélection Supprimer Aide

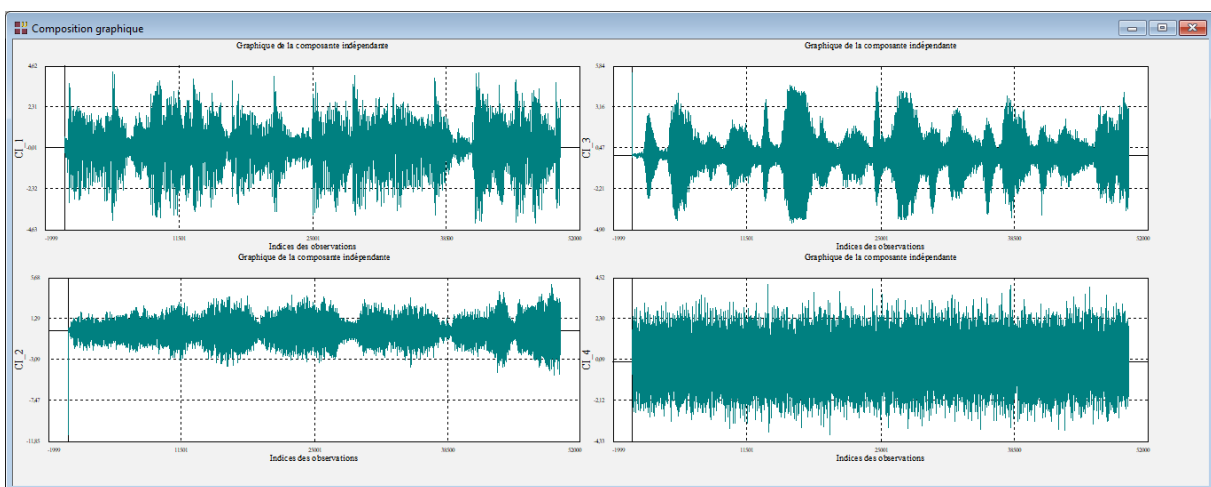
Voici ci-dessous les résultats obtenus.

Rapports et Graphiques

Rapport ICA

- Données initiales
- Matrice de démélange
- Composantes indépendantes

	1	2	3	4	5	6	7	8
1								
2	MATRICE DES COMPOSANTES INDEPENDANTES (DONNEES SOURCES)							
3								
4								
5								
		CI_1	CI_2	CI_3	CI_4			
6	1	-0,09262	-11,20194	-5,44438	1,69498			
7	2	-0,29120	-0,61703	-0,31105	-0,71023			
8	3	-0,12412	0,32621	0,15407	-0,27642			
9	4	0,04497	-0,33988	-0,12300	-0,11998			
10	5	-0,09004	0,22866	0,14675	-0,12272			
11	6	0,01234	-0,33934	-0,12300	0,26814			
12	7	0,24759	0,22514	0,16037	0,72567			
13	8	0,34827	-0,24872	-0,06763	0,23070			
14	9	0,61739	0,22344	0,13063	0,33517			
15	10	0,27973	-0,15105	-0,06481	-0,55384			
16	11	0,28213	0,23003	0,08184	0,34623			
17	12	0,45383	-0,05494	-0,04839	1,48170			
18	13	0,28026	0,03553	0,06814	0,18667			
19	14	0,35354	-0,05883	0,02966	2,44121			
20	15	0,31137	0,03732	0,02652	-1,15548			
21	16	0,41335	-0,05820	-0,01569	-0,80752			
22	17	0,28211	-0,05614	-0,01641	0,45433			



Dans cet exemple, nous connaissons la matrice de mélange initiale. Nous pouvons donc calculer le produit matriciel de la matrice de mélange initiale (Rotation) par la matrice de démélange estimée (W) de façon à voir si la matrice obtenue n'a à peu près qu'un seul élément non nul en ligne et en colonne :

```

R Console
> Rotation
      [,1]      [,2]      [,3]      [,4]
[1,] 0.1989445 0.066042430 0.7960446 0.4074442
[2,] 0.3164456 0.007431646 0.4714148 0.7279637
[3,] 0.1746490 0.294247244 0.3067949 0.1701618
[4,] 0.7910636 0.476462182 0.1508696 0.6219070
> Res$W
      [,1]      [,2]      [,3]      [,4]
[1,] 1.930935 -0.9493257 -0.2541069 -0.0801665
[2,] -2.716585 1.1377401 5.8263278 -1.1454935
[3,] -3.092835 2.9243862 4.7697215 -2.7058202
[4,] -2.708573 3.3364635 2.4661190 -1.1977056
> round(Res$W%*%Rotation, 4)
      [,1]      [,2]      [,3]      [,4]
[1,] -0.0241 0.0075 0.9995 0.0026
[2,] -0.0690 0.9976 -0.0115 0.0004
[3,] -0.9973 -0.0683 -0.0283 -0.0025
[4,] 0.0002 0.0009 -0.0074 1.0000
> |

```

Exemple 3 : Fichier ECG

Dans ce troisième exemple, nous utilisons un jeu de données contenant des enregistrements d'électrocardiographie d'une femme enceinte (ECG).

Ces données mesurent le potentiel électrique généré par le muscle cardiaque à partir de la surface du corps.

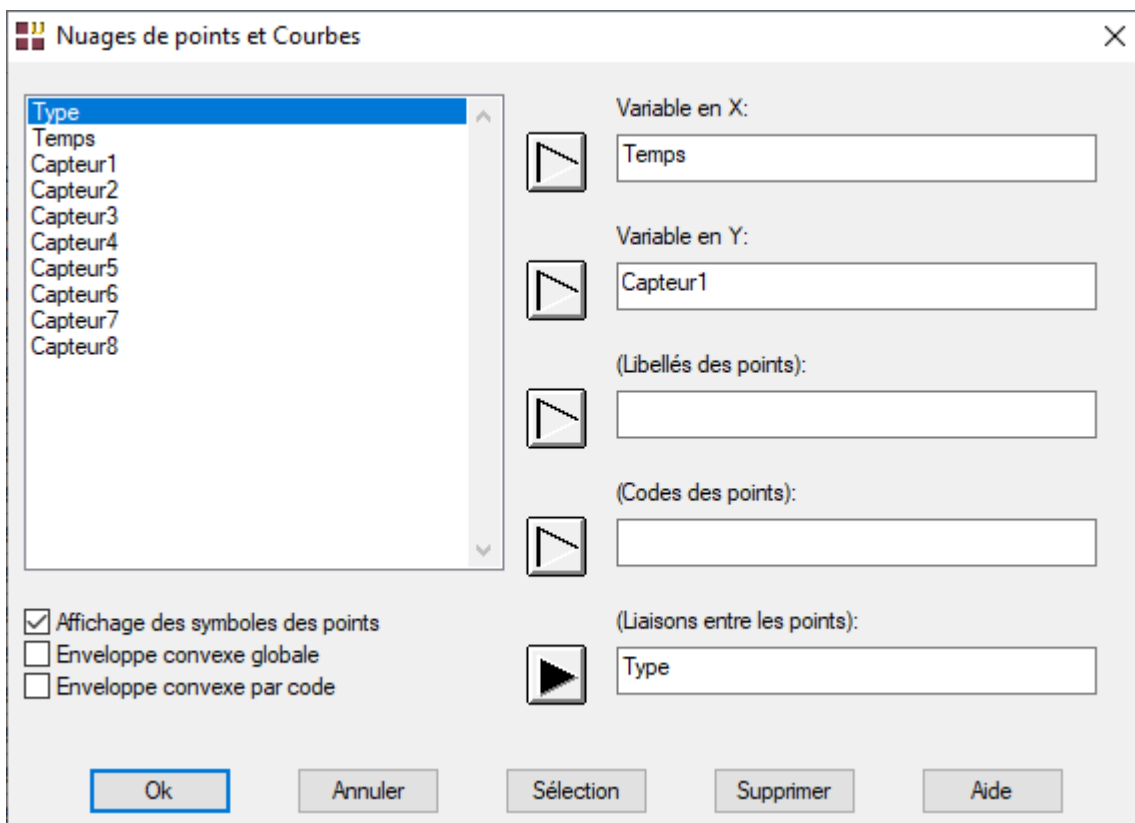
L'activité électrique produite par les battements cardiaques d'un fœtus peut alors être détectée en mesurant le potentiel sur la peau de la mère.

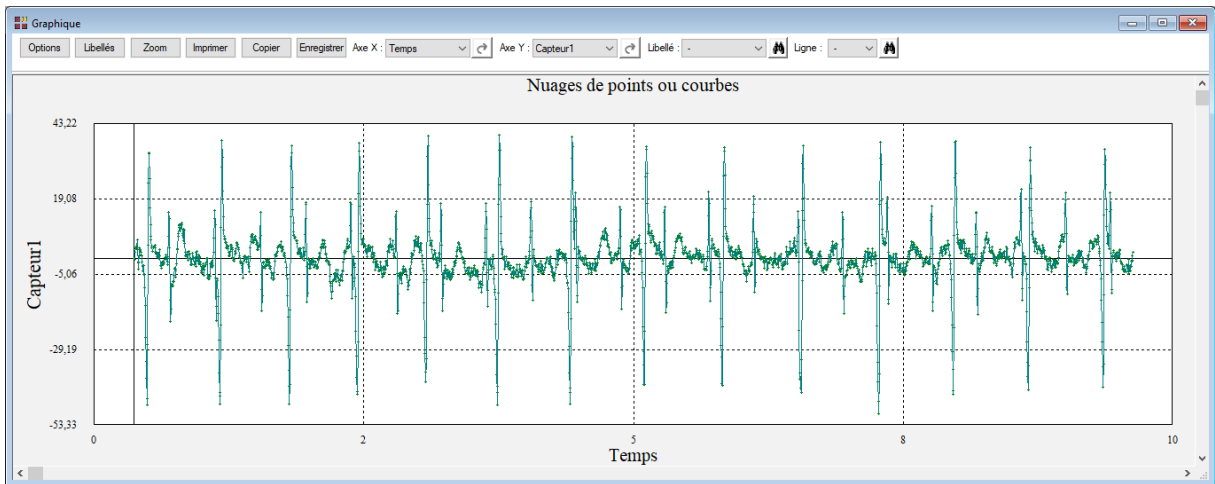
Comme les signaux mesurés sont des mélanges des battements cardiaques du fœtus et de la mère, le but est d'utiliser la méthode JADE pour séparer ces deux battements cardiaques ainsi que certains artefacts les uns des autres.

Dans ce contexte, il est utile de savoir que le cœur d'un fœtus est censé battre plus rapidement que celui de la mère.

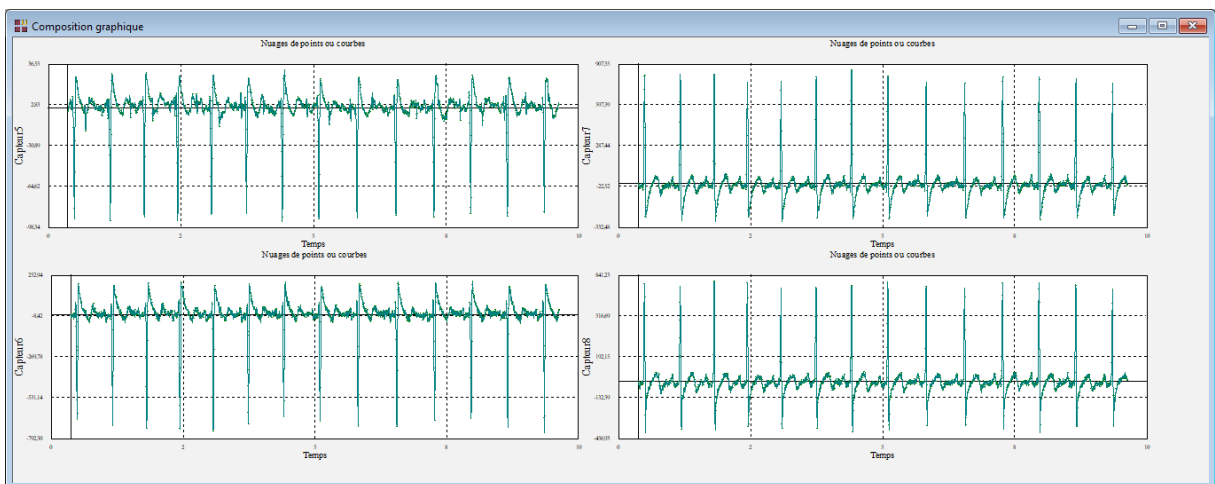
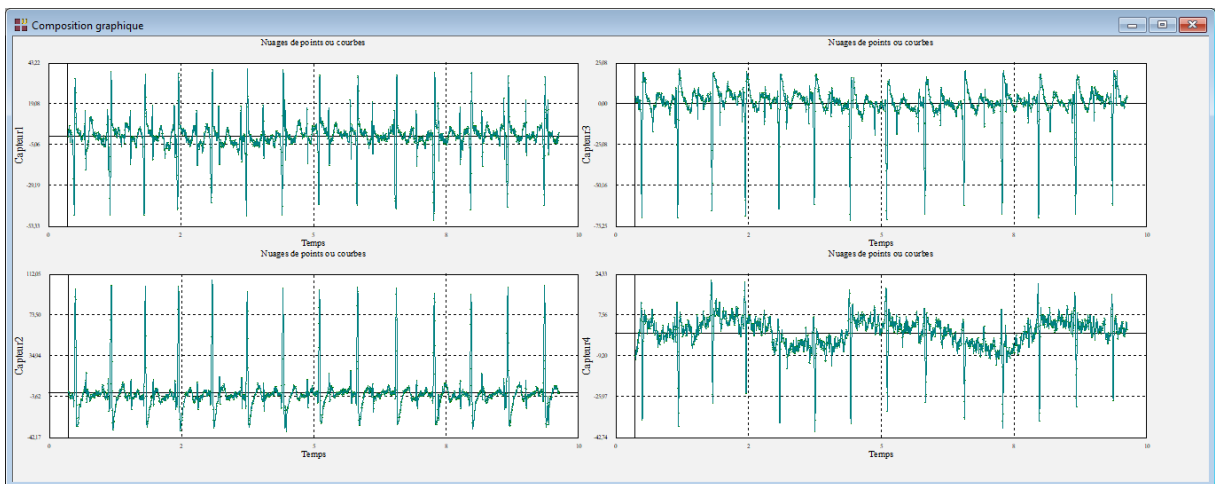
Dans cet enregistrement ECG, huit capteurs ont été placés sur la peau de la mère, les cinq premiers dans la région abdominale et les trois derniers dans la région thoracique.

Visualisons les données dans UNIWIN en utilisant la procédure 'Nuages de points et Courbes' du ruban 'Graphiques' :

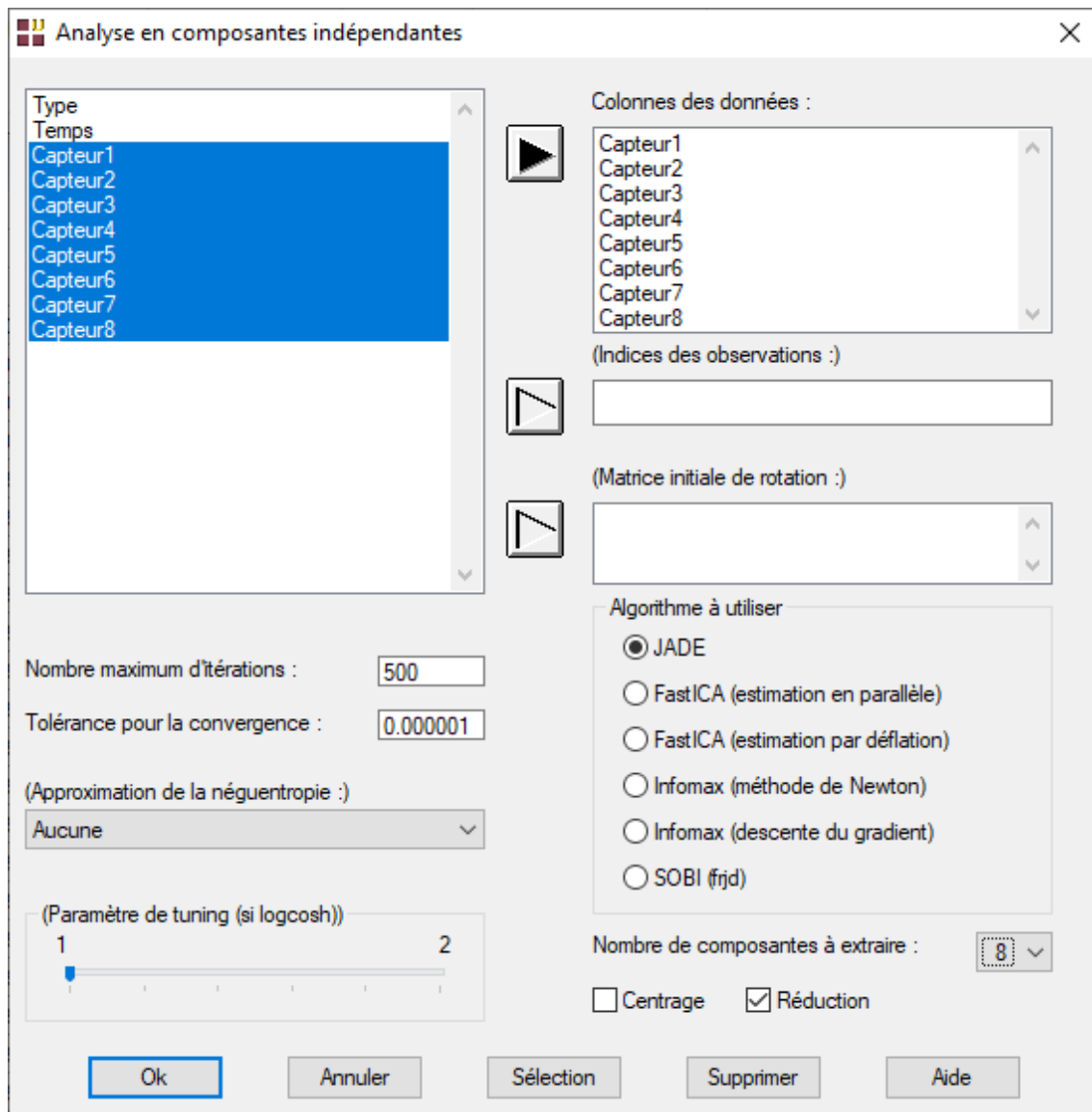




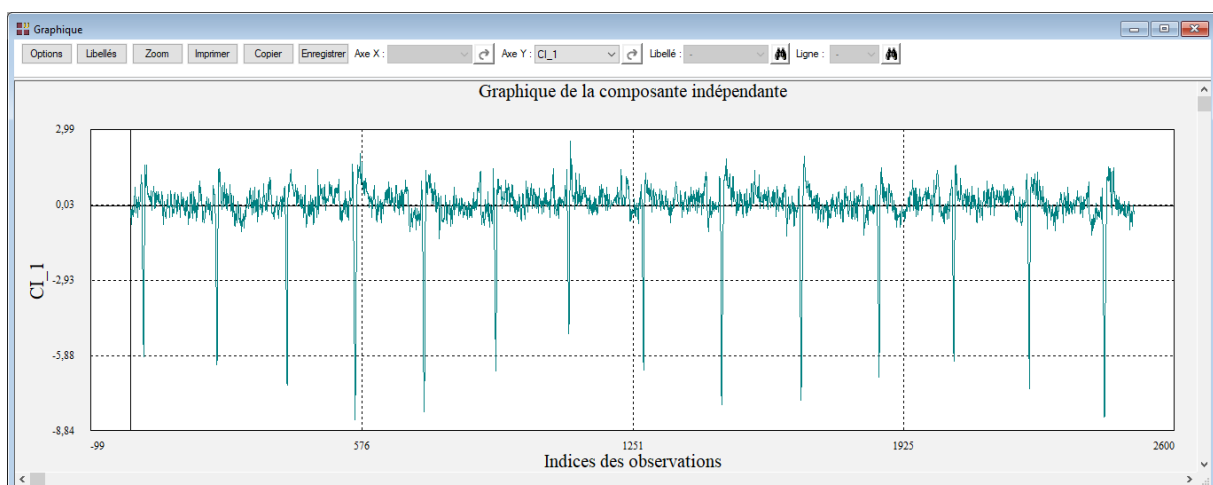
En utilisant l'option 'Composition' du ruban 'Graphique', visualisons les enregistrements des huit capteurs :



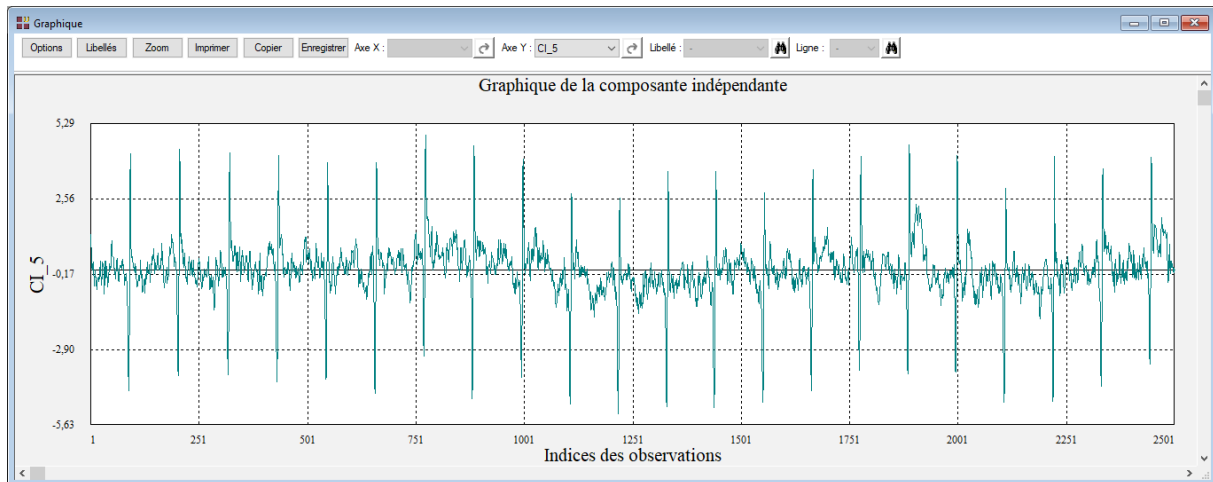
Exécutons la procédure JADE sur les données non centrées :



Parmi d'autres, la première composante indépendante possède 14 pics et représente les battements cardiaques de la mère :



La cinquième composante indépendante est celle d'intérêt dans cet exemple. Elle comporte 22 pics et représente les battements cardiaques plus rapides du fœtus :



Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure.

A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

<i>Variable</i>	<i>Contenu</i>
CI	Composantes indépendantes
W	Matrice de démélange estimée
liaisons	Liaisons entre les observations
indices	Indices des observations
CP	Composantes principales
M	Matrice de mélange estimée
Q	Matrice de blanchiment
R	Matrice de rotation
RESID	Matrice des résidus
DW	Coefficients de Durbin-Watson
vafs	Variances des composantes indépendantes
iter	Itérations de l'algorithme

Références

Documentation du package R 'ica' (2018)

<https://cran.r-project.org/web/packages/ica/ica.pdf>

Documentation du package R 'JADE' (2020)

<https://cran.r-project.org/web/packages/JADE/JADE.pdf>

Documentation du package R 'tuneR' (2018)

<https://cran.r-project.org/web/packages/tuneR/tuneR.pdf>

Exemple 1

<http://rpubs.com/skydome20/93614>

Exemples 2 et 3

Blind Source Separation Based on Joint Diagonalization in R: The Packages JADE and BSSasyp

Jari Miettinen, Klaus Nordhausen, Sara Taskinen - August 19, 2019

Signal Separation Methods based on Singular Value Decomposition and their Application to the Real-Time Extraction of the Fetal Electrocardiogram from Cutaneous Recordings

Dirk Callaerts, Ph.D. Thesis, K.U.Leuven - E.E. Dept., Dec. 1989.

Fetal Electrocardiogram Extraction by Blind Source Subspace Separation

IEEE Trans. Biomedical Engineering, Vol. 47, No. 5, May 2000, Special Topic Section on Advances in Statistical Signal Processing for Biomedicine, pp. 567-572.

L. De Lathauwer, B. De Moor, J. Vandewalle.