

# Avant-après, amont-aval : les couples de tableaux totalement appariés

## Résumé

La fiche décrit les méthodes d'analyse des couples de tableaux complètement appariés (mêmes individus, mêmes variables). L'analyse de co-inertie prend dans cette situation des propriétés très particulière (Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* : 43, 109-121). Elle est complétée par une analyse des différences enrichie de nouvelles aides à l'interprétation. Des couples de tableaux de mesures acquises sur deux stations (Carrel, G. & Berget, J.C. , 1995. Suivi hydrobiologique du Rhône. Site de Cruas-Meysses (1982-1995). Cemagref Aix-en-Provence, Rapport à EDF, 1-44 + annexes.) ou à deux dates (Vespini, F., Légier, P. & Champeau, A. (1987) Ecologie d'une rivière non aménagée des Alpes du Sud : Le Buëch (France) I Evolution longitudinale des descripteurs physiques et chimiques. *Annales de Limnologie* : 23, 151-164) illustrent ces propositions.

## Plan

1 — La question .....	3
2 — Structure commune .....	4
2.1 — Co-structure d'ACP normées .....	4
2.2 — Co-structure totalement appariée .....	7
2.3 — Co-structure faunistique.....	11
3 — Recherche des différences.....	13
3.1 — Tableaux avant-après .....	13
3.2 — Normalisation des différences.....	16
3.3 — Structure spatiale des différences.....	19
Références .....	20

D. Chessel & G. Carrel

# 1 — La question

Elle est souvent posée dans les enquêtes sur l'environnement hydrobiologique où l'on dispose d'un tableau de mesures avant et après un aménagement. La même structure de données se présentent quand on examine en différents sites ou à différentes dates ce qui se passe dans une station en amont et en aval d'une centrale nucléaire, d'un barrage, d'un confluent, d'une station de pompage<sup>1</sup> ...

Dans tous les cas, on dispose d'un tableau X portant sur n individus-lignes et d'un tableau Y portant sur les mêmes individus (lignes) et les mêmes variables (colonnes). X et Y ont une structure commune au minimum associée à leur parenté expérimentale et présentent un ensemble de différences, ensemble dont l'expression peut être l'objectif principal de l'analyse.

Examinons le cas suivant résumé dans la figure 1 :

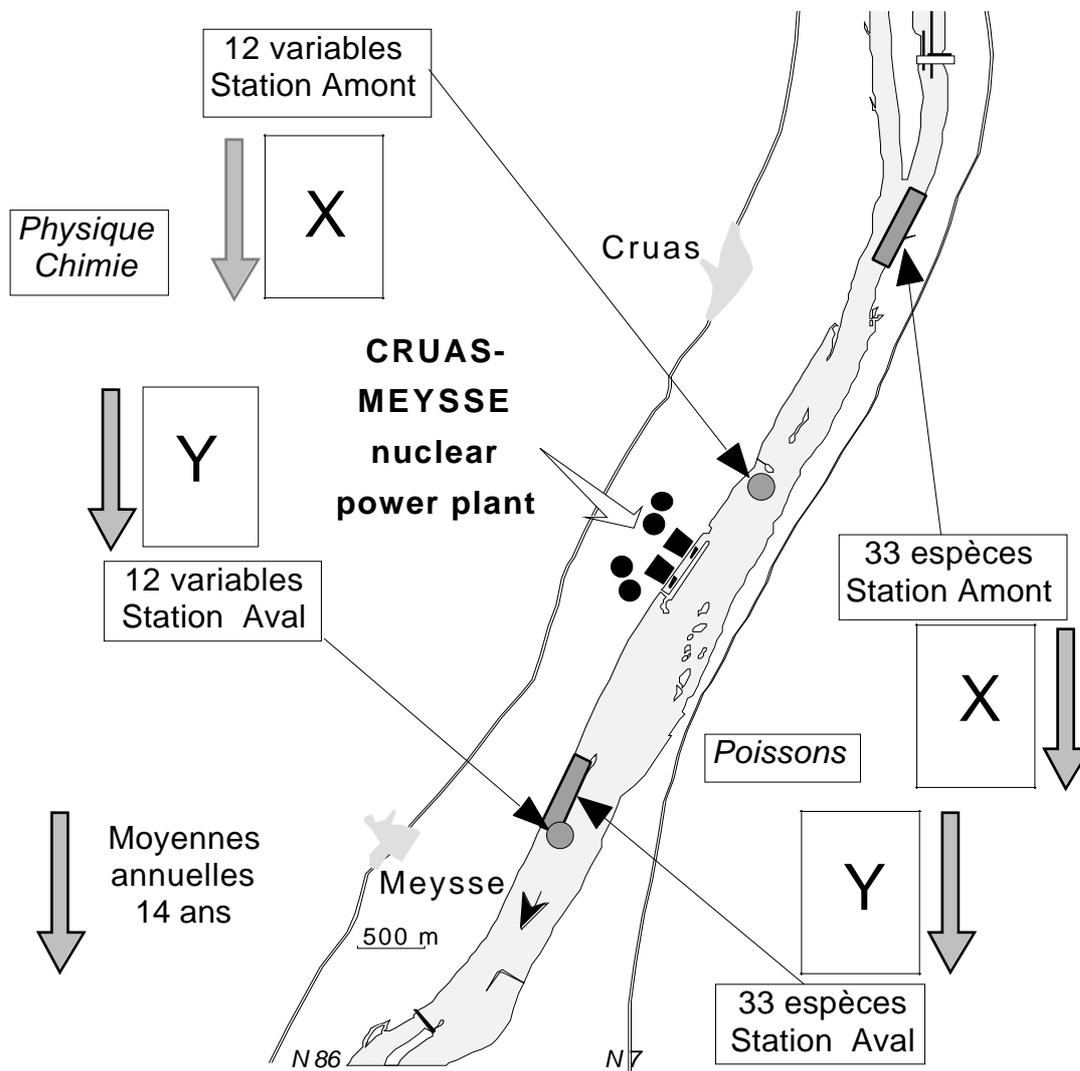


Figure 1 : Structure des données. Au centre, se trouve une centrale nucléaire. En amont, on mesure la physico-chimie de l'eau et l'abondance des Poissons. En aval, on fait de même. On recommence pendant 14 ans.<sup>2</sup>

A partir des indications de<sup>3</sup>, on examine ici les propositions programmées dans ADE-4 pour examiner ce type de données.

## 2 — Structure commune

### 2.1 — Co-structure d'ACP normées

Le tableau X s'appelle Mil1. Il a 14 lignes (années) et 12 colonnes (variables) :

7.9	347	67	0.2	4.3	0.0	0.4	158	68	17	2.6	2.4
7.9	333	70	0.1	4.6	0.0	0.5	147	58	17	1.7	2.4
8	371	12	0.3	4.7	0.0	0.7	153	68	22	2.5	2
8	393	23	0.2	5.3	0.1	0.6	160	66	27	2.9	2.5
8.1	363	18	0.2	5.8	0.1	0.5	148	66	23	3.1	2.4
8.1	361	83	0.3	4.2	0.1	0.3	164	55	18	3.2	2.9
8	358	34	0.1	4.8	0.0	0.3	152	61	18	2.8	2.3
8	404	16	0.2	5.2	0.1	0.4	159	76	25	2.3	3.6
8.1	357	21	0.1	4.8	0.1	0.3	152	67	20	2.3	3.3
8	360	44	0.2	6.1	0.1	0.3	153	60	19	3.1	3.3
8.1	346	48	0.1	5.2	0.1	0.3	149	56	17	3.1	2.8
8	382	17	0.2	5.6	0.1	0.2	161	59	20	2.7	2.4
8	356	45	0.1	5.9	0.1	0.1	156	56	15	3.2	2.5
8.1	362	32	0.1	4.9	0.1	0.1	158	50	18	3	3.5

Le tableau Y s'appelle Mil2. Il a 14 lignes (années) et 12 colonnes (variables) :

7.9	348	66	0.2	4.3	0.0	0.4	157	67	16	2.6	2.5
7.9	338	61	0.2	4.7	0.0	0.5	148	58	18	1.7	2.1
8	373	17	0.2	4.8	0.0	0.7	153	70	22	2.5	2.2
8	392	22	0.2	5.4	0.1	0.6	160	66	27	2.9	2.8
8.1	368	36	0.2	5.8	0.1	0.5	148	68	24	3.1	2.3
8.1	365	95	0.2	3.9	0.1	0.3	163	56	18	3.3	2.8
8.1	365	25	0.1	4.8	0.0	0.3	152	63	18	2.8	2.3
8.1	410	56	0.1	5.6	0.1	0.4	156	75	25	2.3	3.8
8.2	360	66	0.1	5	0.1	0.3	149	67	20	2.3	3.5
8.1	360	27	0.1	5.7	0.1	0.3	153	61	19	3	3.6
8.1	347	56	0.1	5.2	0.1	0.3	148	57	16	2.9	3
8.1	386	48	0.1	5.7	0.1	0.2	160	61	18	2.7	2.9
8	355	44	0.1	5.8	0.1	0.1	152	57	15	3.3	3.7
8.2	364	73	0.1	5.2	0.0	0.1	161	49	18	3.2	3.9

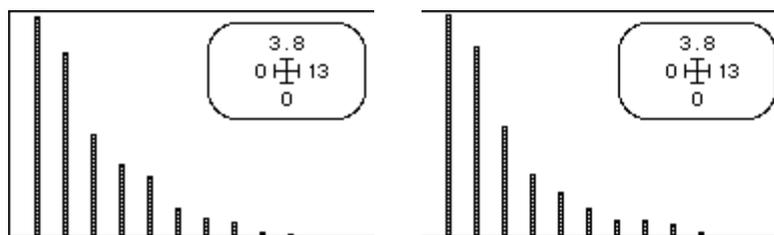
Les variables communes sont pH (pH), Conductivité (Cond), Matières en suspension (MES), Ammonium (NH<sub>4</sub>), Nitrates (NO<sub>3</sub>), Nitrites (NO<sub>2</sub>), Phosphates (PO<sub>4</sub>), Bicarbonates (HCO<sub>3</sub>), Sulfates (SO<sub>4</sub>), Chlorures (Cl), Silice (SiO<sub>2</sub>) et Demande Biologique en Oxygène (DBO<sub>5</sub>). La matrice de corrélation des variables du premier tableau est :

1.00	0.10	-0.24	-0.17	0.13	0.17	-0.29	0.02	-0.28	0.09	0.54	0.41
0.10	1.00	-0.65	0.33	0.27	0.63	0.22	0.57	0.51	0.82	0.08	0.23
-0.24	<b>-0.65</b>	1.00	0.11	-0.46	-0.42	-0.21	0.08	-0.46	-0.65	0.04	-0.04
-0.17	0.33	0.11	1.00	-0.35	-0.03	0.62	0.41	0.38	0.40	0.01	-0.34
0.13	0.27	-0.46	-0.35	1.00	0.57	-0.29	-0.17	-0.03	0.18	0.41	0.11
0.17	<b>0.63</b>	-0.42	-0.03	<b>0.57</b>	1.00	-0.31	0.39	0.15	0.40	0.10	0.48
-0.29	0.22	-0.21	<b>0.62</b>	-0.29	-0.31	1.00	-0.21	0.66	0.63	-0.43	-0.41
0.02	<b>0.57</b>	0.08	0.41	-0.17	0.39	-0.21	1.00	-0.02	0.18	0.28	0.22
-0.28	<b>0.51</b>	-0.46	0.38	-0.03	0.15	<b>0.66</b>	-0.02	1.00	0.67	-0.42	-0.03
0.09	<b>0.82</b>	<b>-0.65</b>	0.40	0.18	0.40	<b>0.63</b>	0.18	<b>0.67</b>	1.00	-0.11	0.07
<b>0.54</b>	0.08	0.04	0.01	0.41	0.10	-0.43	0.28	-0.42	-0.11	1.00	0.06
0.41	0.23	-0.04	-0.34	0.11	0.48	-0.41	0.22	-0.03	0.07	0.06	1.00

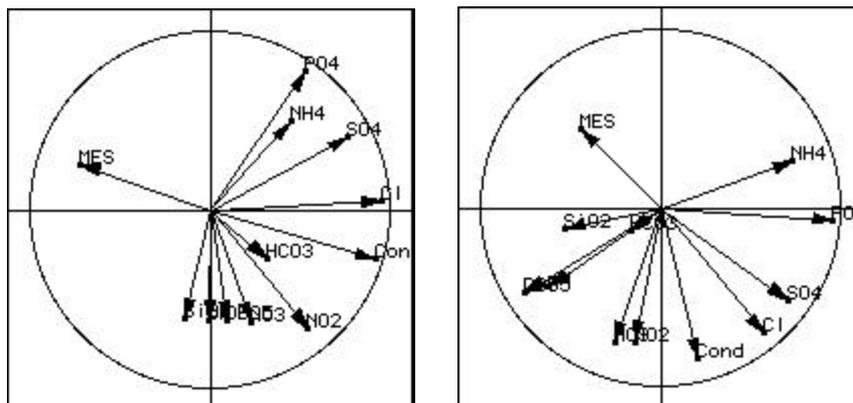
Elle présente un effectif non négligeable de valeurs significatives pour  $n = 14$  et l'ACP est légitime. Il en est de même pour le second :

1.00	0.28	0.18	-0.58	0.27	0.28	-0.47	0.10	-0.18	0.09	0.37	0.56
0.28	1.00	-0.26	0.11	0.38	0.53	0.21	0.47	0.53	0.74	0.04	0.25
0.18	-0.26	1.00	-0.04	-0.51	-0.15	-0.40	0.31	-0.42	-0.41	0.01	0.26
<b>-0.58</b>	0.11	-0.04	1.00	-0.54	-0.13	0.64	0.30	0.28	0.24	-0.07	-0.69
0.27	0.38	<b>-0.51</b>	<b>-0.54</b>	1.00	0.54	-0.24	-0.23	0.13	0.29	0.18	0.44
0.28	<b>0.53</b>	-0.15	-0.13	<b>0.54</b>	1.00	-0.15	0.17	0.31	0.29	0.09	0.44
-0.47	0.21	-0.40	<b>0.64</b>	-0.24	-0.15	1.00	-0.14	0.65	0.68	-0.47	-0.61
0.10	0.47	0.31	0.30	-0.23	0.17	-0.14	1.00	-0.20	0.09	0.38	0.24
-0.18	<b>0.53</b>	-0.42	0.28	0.13	0.31	<b>0.65</b>	-0.20	1.00	0.65	-0.42	-0.22
0.09	<b>0.74</b>	-0.41	0.24	0.29	0.29	<b>0.68</b>	0.09	<b>0.65</b>	1.00	-0.15	-0.07
0.37	0.04	0.01	-0.07	0.18	0.09	-0.47	0.38	-0.42	-0.15	1.00	0.31
<b>0.56</b>	0.25	0.26	<b>-0.69</b>	0.44	0.44	-0.61	0.24	-0.22	-0.07	0.31	1.00

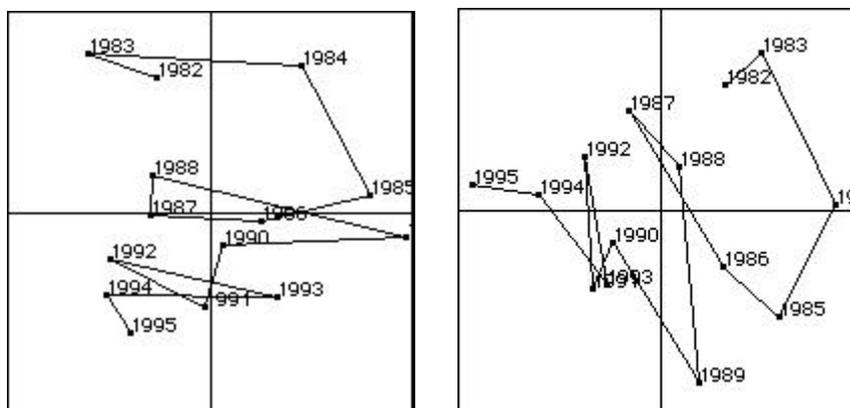
Après PCA : Correlation matrix PCA on garde deux facteurs (à gauche, Mil1, à droite Mil2), avec Curves : Eigenvalues :



Les coordonnées des variables sont (Scatters : Labels):



On reconnaît l'association des variables associés à la charge ionique (Chlorures, sulfates, conductivité) et une certaine identité de la structure de corrélation à une rotation près. Les coordonnées des lignes (Scatters : Trajectories) sont :

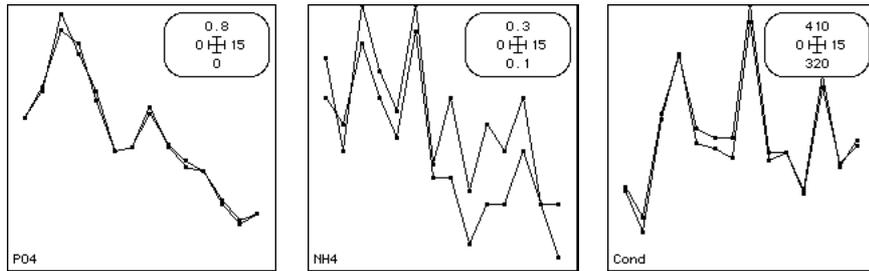


Les deux figures présentent, toujours à une rotation près une composante tendancielle sur la période et une composante alternée. La variabilité des données semble faible mais elle porte sur des moyennes et l'organisation des années ne peut être le résultat du hasard. On n'identifie pas immédiatement la partie commune de la relation lignes-colonnes et l'analyse simultanée des deux tableaux s'impose.

On peut utiliser l'analyse de co-inertie ordinaire<sup>4</sup> qui est alors une analyse inter-batterie<sup>5</sup>. C'est un cas typique d'appariement entre deux tableaux qui conduirait l'analyse canonique à une instabilité numérique chronique pour 14 lignes et 12 colonnes dans chaque tableau. Par CoInertia : Matching two statistical triplets et CoInertia : Coinertia test - Fixed D on obtient une signification certaine :



autres variables, la conductivité est indépendante de cette tendance et évolue de manière cyclique qui s'exprime simplement dans la représentation des données :



## 2.2 — Co-structure totalement appariée

L'examen de la partie commune des deux tableaux ne pose donc aucun problème particulier. On peut simplement utiliser cet exemple pour aller au bout de cette stratégie dans le cas de deux tableaux totalement appariés. En effet les deux nuages de 14 points sont dans le même espace. Quand on superpose les deux plans de co-inertie, on utilise abusivement une technique graphique. Les deux plans sont suffisamment proches dans l'espace de référence, ce qui autorise cet abus, mais ils ne sont pas confondus.

La question émerge alors de construire un axe, un plan, un sous-espace de co-inertie commun aux deux nuages, ce qui s'exprime dans la figure 2. Ceci est possible par la diagonalisation de la demi-somme des matrices de covariances inter-tableaux (op. cit. § 2). La méthode est programmée dans `CoInertia : Fully Matched Tables` :



```
First input statistical triplet: table Mil1.cnta
Number of rows: 14, columns: 12
Second input statistical triplet: table Mil2.cnta
Number of rows: 14, columns: 12
```

On calcule la matrice des covariances entre les variables du tableau 2 et celles du tableau 1. Comme les deux paquets sont identiques, c'est une matrice carrée :

```
File A_S.croi contains the crossed table YtDnX
It has 12 rows and 12 columns
```

<b>0.84</b>	0.26	-0.41	-0.35	0.21	0.44	-0.46	0.10	-0.19	0.11	0.35	0.64
0.11	<b>0.99</b>	-0.65	0.32	0.23	0.61	0.23	0.53	0.53	0.81	0.02	0.23
0.25	-0.28	<b>0.59</b>	-0.05	-0.53	-0.05	-0.41	0.33	-0.33	-0.41	-0.07	0.45
-0.30	0.10	0.28	<b>0.91</b>	-0.44	-0.30	0.65	0.28	0.24	0.24	-0.08	-0.59
0.13	0.41	-0.65	-0.43	<b>0.94</b>	0.60	-0.23	-0.17	0.10	0.32	0.22	0.18
0.10	0.56	-0.34	0.14	0.60	<b>0.93</b>	-0.16	0.32	0.29	0.39	0.17	0.34
-0.31	0.21	-0.17	0.60	-0.30	-0.29	<b>0.99</b>	-0.20	0.64	0.63	-0.43	-0.39
0.02	0.49	0.10	0.42	-0.23	0.29	-0.16	<b>0.94</b>	-0.16	0.19	0.28	0.19
-0.25	0.51	-0.50	0.41	0.02	0.15	0.67	-0.06	<b>0.99</b>	0.67	-0.37	-0.13
0.10	0.73	-0.60	0.35	0.16	0.29	0.68	0.07	0.65	<b>0.98</b>	-0.15	0.08
0.54	0.10	0.05	-0.01	0.34	0.07	-0.46	0.37	-0.46	-0.13	<b>0.98</b>	0.08
0.35	0.30	-0.17	-0.47	0.39	0.56	-0.62	0.34	-0.14	-0.04	0.27	<b>0.83</b>

Comme les variables sont normalisées, ces covariances sont des corrélations. Sur la diagonale on a la corrélation entre les deux mesures amont et aval de la même variable. On remplace ensuite la valeur  $Cov_{jk}$  par la demi-somme  $Cov_{jk} + Cov_{kj}$  et on obtient une matrice symétrique, donc diagonalisable.

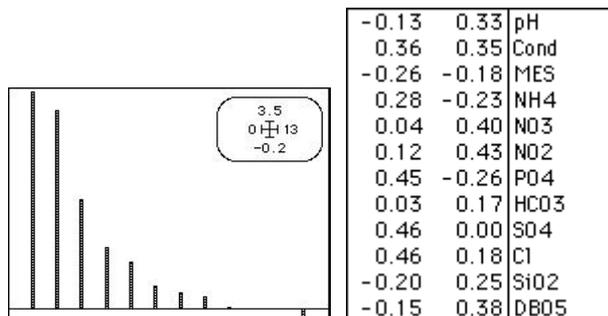
Symetric operator diagonalization

```

-----
Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |
001 3.475e+00 | 002 3.166e+00 | 003 1.761e+00 | 004 9.879e-01 |
005 7.517e-01 | 006 3.791e-01 | 007 2.723e-01 | 008 2.093e-01 |
009 3.398e-02 | 010 1.409e-02 | 011 -8.660e-03 | 012 -1.367e-01 |
File A_S.valp contains the eigenvalues
--- It has 12 rows and 1 column

```

Cette matrice n'est, par contre, pas positive et on obtient des valeurs propres négatives, ce qui ne pose pas de problèmes particuliers :



```

File A_S.vecp contains the eigenvectors (Dp-normed)
--- It has 12 rows and 2 columns

```

```

File :A_S.vecp
|Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -2.572e-01 | 4.632e-01 |
| 2 | -2.584e-01 | 4.295e-01 |
|-----|-----|-----|

```

Les composantes des vecteurs propres normés (ci-dessus, à droite) sont les coefficients des combinaisons linéaires des variables normalisées qui définissent les scores des relevés. Ces vecteurs sont des axes de co-inertie communs aux deux nuages de points, c'est à dire des axes qui servent à faire simultanément les deux ACP normée des deux tableaux. Ces deux axes définissent un plan sur lequel on projette les axes d'inertie des deux nuages :

```

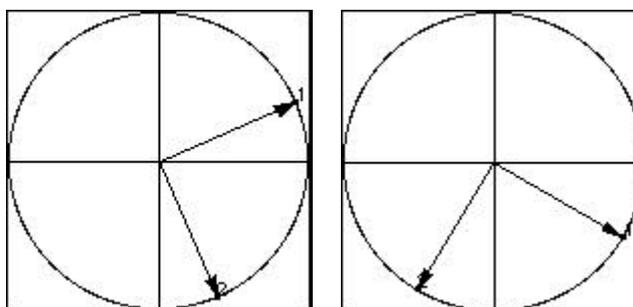
File A_S.cpl contains the scores of the projections of
inertia axes onto co-inertia axes (Tab 1)
--- It has 2 rows and 2 columns

```

```

File :A_S.cpl
|Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | 3.984e-01 | 9.090e-01 |
| 2 | -9.079e-01 | 3.900e-01 |
|-----|-----|-----|

```



```

File A_S.cp2 contains the scores of the projections of
inertia axes onto co-inertia axes (Tab 2)
--- It has 2 rows and 2 columns

```

```

File :A_S.cp2
|Col. | Mini | Maxi |
|-----|-----|-----|

```

1	-5.023e-01	8.546e-01
2	-8.484e-01	-4.984e-01

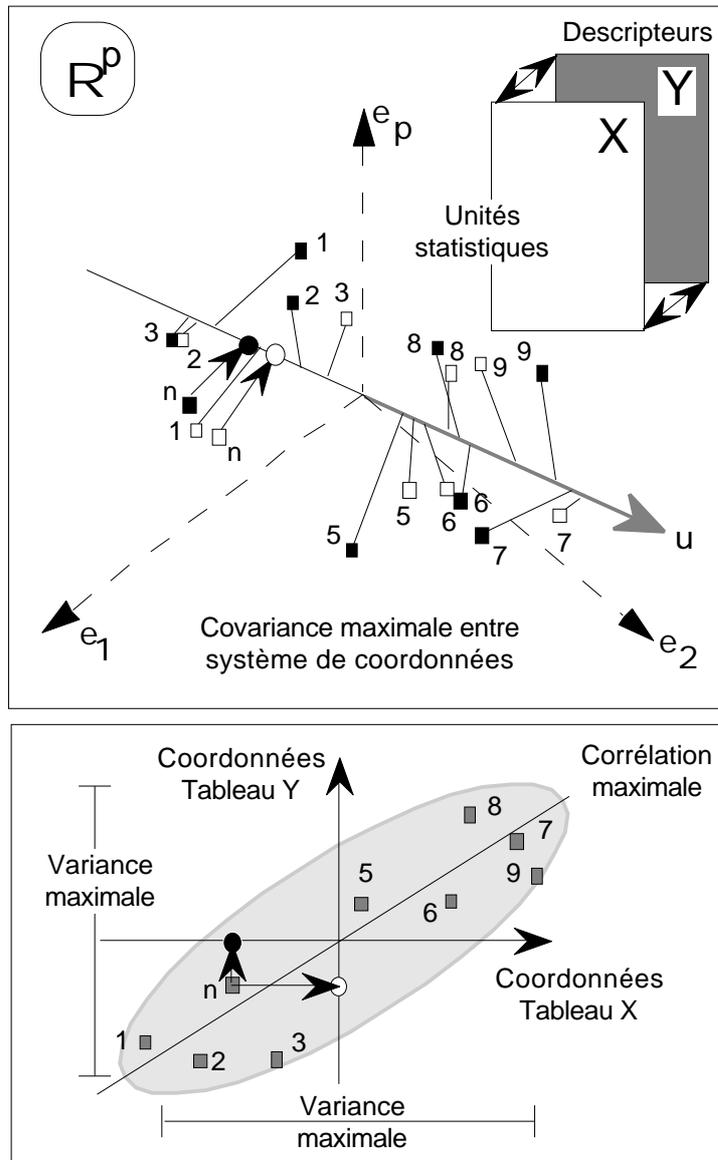


Figure 2 : Question posée dans l'analyse d'un couple de deux tableaux complètement appariés : existe-t-il un axe de co-inertie commun aux deux nuages ? (extrait de Torre & Chessel op. cit. p. 111)

On retrouve la rotation déjà mentionnée. Comme les deux premières valeurs propres des deux analyses séparées sont proches, il est clair que les axes d'inertie sont numériquement instables mais que le plan d'inertie lui-même est clairement défini. Il n'y a donc ici pas lieu d'interpréter les axes mais le plan qu'ils définissent.

Sur ce plan, on projette les deux nuages de points appariés :

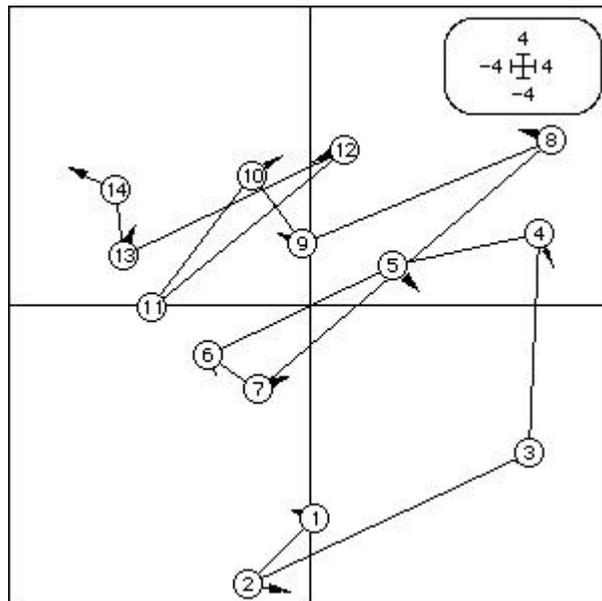
File A\_S.lil contains the coordinates of the rows of table 1  
 --- It has 14 rows and 2 columns  
 File :A\_S.lil

Col.	Mini	Maxi
1	-2.567e+00	3.267e+00
2	-3.685e+00	2.144e+00

File A\_S.li2 contains the coordinates of the rows of table 2  
 --- It has 14 rows and 2 columns  
 File :A\_S.li2

Col.	Mini	Maxi
1	-3.284e+00	3.219e+00
2	-3.890e+00	2.298e+00

On obtient une figure très voisine de celle de la co-inertie ordinaire. Les deux plans d'inertie, les deux plans de co-inertie et le plan de co-inertie unique sont donc tous très voisins :



Les coordonnées normalisées sont également conservées mais les deux tableaux sont suffisamment proches pour qu'ils soient ici inutiles.

File A\_S.mi1 contains the scores (unit norm) of the rows of table 1  
 --- It has 14 rows and 2 columns  
 File :A\_S.mi1

Col.	Mini	Maxi
1	-1.357e+00	1.727e+00
2	-2.082e+00	1.211e+00

File A\_S.mi2 contains the scores (unit norm) of the rows of table 2  
 --- It has 14 rows and 2 columns  
 File :A\_S.mi2

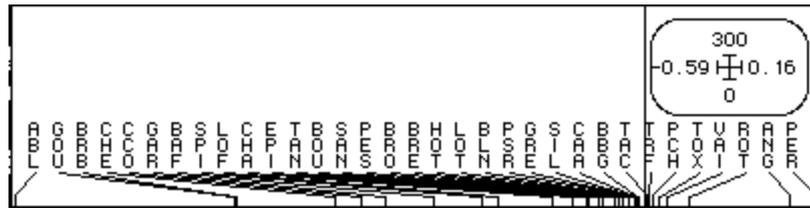
Col.	Mini	Maxi
1	-1.736e+00	1.702e+00
2	-2.148e+00	1.269e+00

Enfin les statistiques habituelles, calculées ici à partir d'un seul système d'axes, sont éditées et on note les très fortes corrélations obtenues :

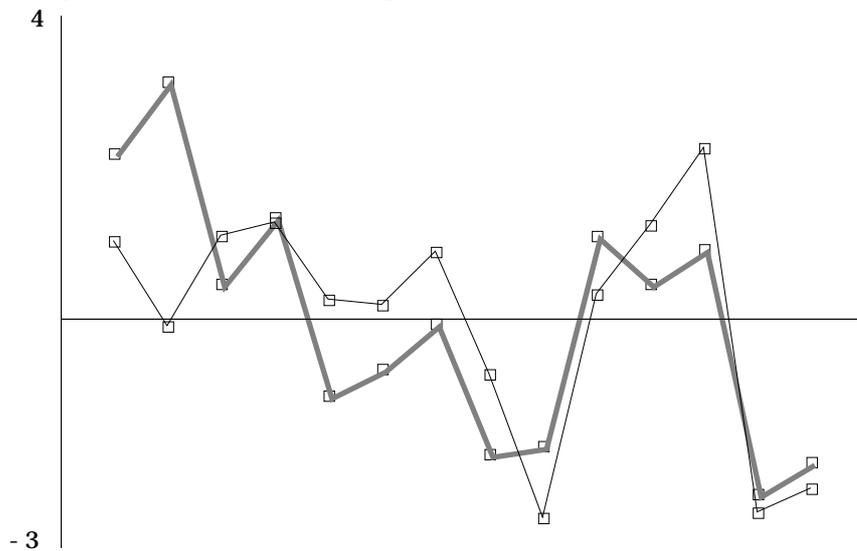
Inertia statistics

Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	3.475	3.58	3.578	0.9711	3.717	3.751
2	3.166	3.132	3.28	0.9876	3.101	3.218





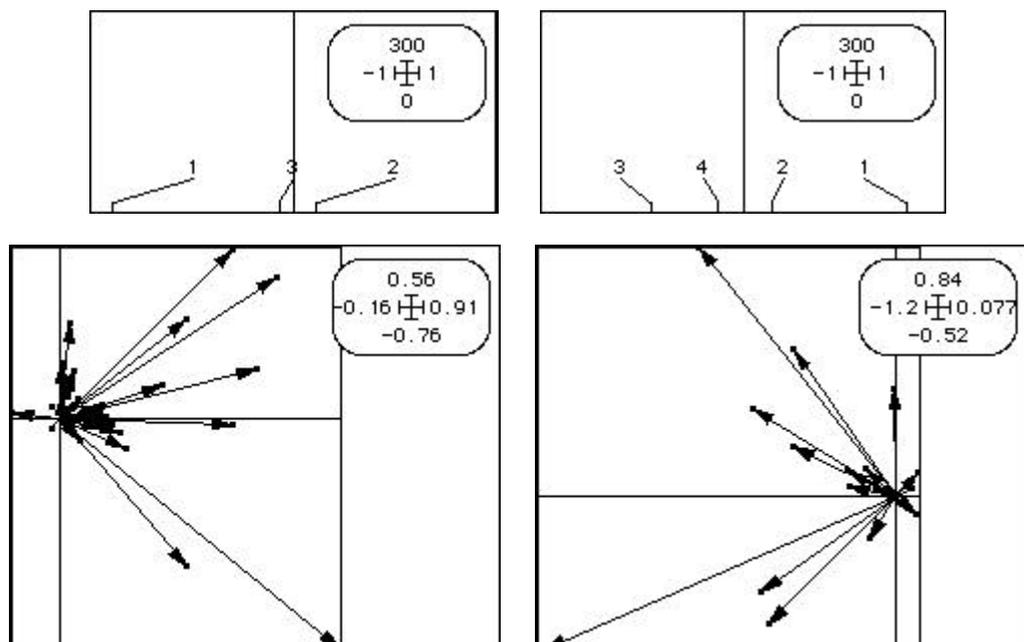
On a affaire à une variation simultanée de l'abondance des principales espèces. La cohérence de ces variations entre les deux stations est exprimée par la seule coordonnée conservée des lignes (coordonnées sur le premier axe de co-inertie) :



Cette cohérence est modérée ( $r = 0.69$ ) vu le nombre de variables disponibles :

Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	1.588	2.488	2.155	0.6857	3.043	2.75

Au total, la co-structure temporelle dans les deux stations est faible et le cortège faunistique reste en place pendant la période étudiée. On vérifie que l'axe de co-inertie unique est voisin des deux axes principaux des deux ACP séparés (on a gardé 3 axes dans la première et 4 dans la seconde) :

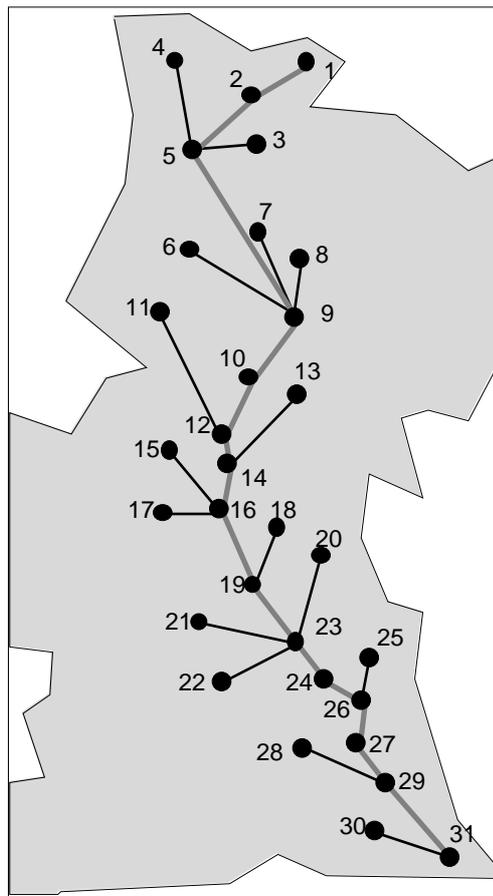


Les deux facteurs 1 sont des effets taille et l'analyse du couple a repéré une certaine similitude dans l'évolution de l'abondance totale. On ne peut en dire plus.

### 3 — Recherche des différences

#### 3.1 — Tableaux avant-après

On peut utiliser, pour compléter le discussion sur les couples de tableaux complètement apparié les données proposées par F. Vespini <sup>6 7</sup>. Elle sont disponibles dans la carte Buech de la pile ADE-4•Datb. Les lignes communes aux deux tableaux sont maintenant des stations :



Le premier tableau Mil1 contient 31 mesures sur 10 variables physiques et chimiques (1-PH, 2-Conductivité ( $\mu\text{S}/\text{cm}$ ), 3-Dureté Calcique ( $\text{mg}/\text{l CaCO}_3$ ), 4-Dureté Totale ( $\text{mg}/\text{l CaCO}_3$ ), 5-Alcalinité ( $\text{mg}/\text{l HCO}_3^-$ ), 6-Chlorures ( $\text{mg}/\text{l Cl}^-$ ), 7-MES ( $\text{mg}/\text{l}$ ), 8-Matières organiques ( $\text{mg}/\text{l}$ ), 9-Nitrates ( $\text{mg}/\text{l NO}_3^-$ ) et 10-Azote ammoniacal( $\text{mg}/\text{l NH}_4^-$ )). Il est acquis en juin. Le second tableau Mil2 donne les valeurs des mêmes variables sur les mêmes individus. Il est acquis en septembre (op. cit. p.162 et 163).

Les auteurs proposent deux ACP normée séparées que nous pouvons assurer simultanément (CoInertia : Fully Matched Tables) :

Fully Matched Tables			
First data input file	<input type="text" value="Mil1.cnta"/>	31	10
Second data input file	<input type="text" value="Mil2.cnta"/>	31	10
Output file name	<input type="text" value="M12"/>		



On ne garde qu'un axe principal commun. L'analyse permet de superposer les coordonnées des stations sur cet axe (fichiers M12\_S.li1 et M12\_S.li2) et de refaire d'un seul coup les graphes représentant "l'évolution de l'amont vers l'aval de la minéralisation des eaux du Buëch, codée par les coordonnées des stations sur le premier axe d'une ACP" (op. cit. figure 6 et 9). L'analyse montre la même structure pour les deux tableaux telle qu'elle avait été perçue séparément par les auteurs.

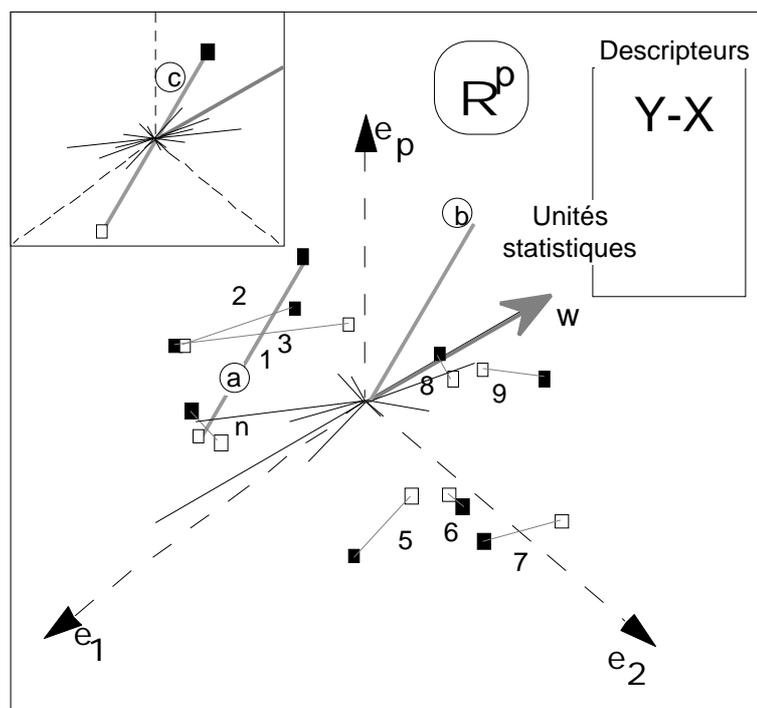
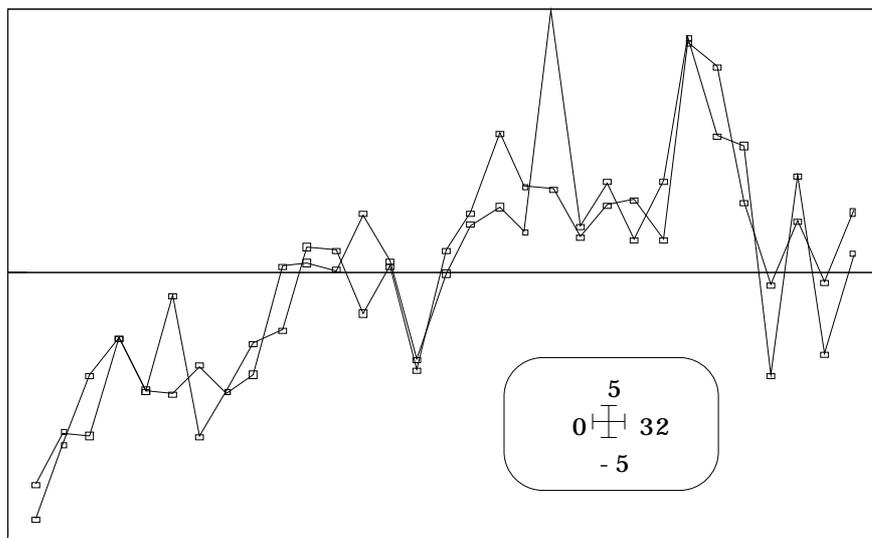


Figure 3 — Analyse des différences entre deux tableaux complètement appariés. L'axe principal du tableau  $Y-X$  est aussi celui de l'analyse intra-classe (en haut, à gauche). Extrait de Torre & Chessel op. cit. p. 115)

La question de la structure commune des deux tableaux renvoie à une question complètement disjointe : peut-on faire un bilan de leurs différences ?

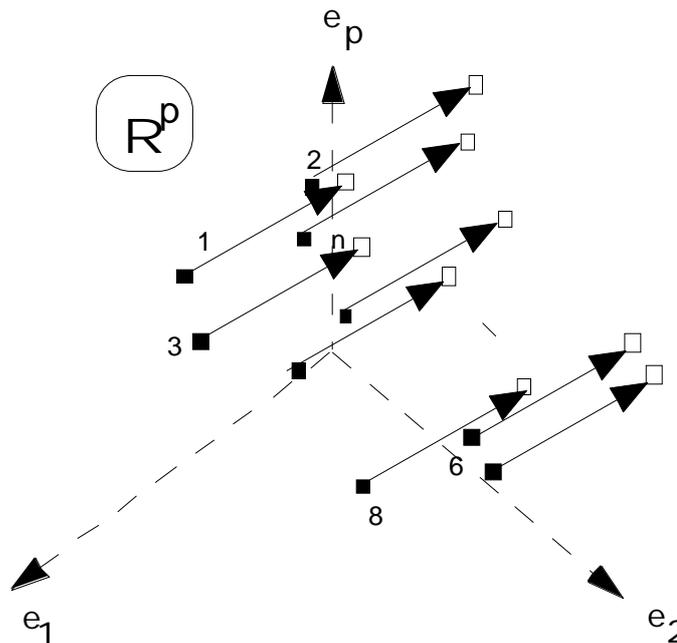


Figure 4 — Cas particulier d'une translation de nuage de points. Le recentrage des deux tableaux annule totalement l'expression de cette différence.

Nous avons abordé cette question par l'analyse des différences entre deux tableaux (figure 3). Il convient de préciser cette opération dans un cadre plus biologique. Qu'entend-on par différence de tableau ? Chaque station est un point dans l'espace  $\mathbb{R}^{10}$ . Avec deux tableaux on obtient deux représentations du point, donc un vecteur lié qui exprime la différence entre les deux positions.

Si on fait porter la différence sur les deux tableaux normalisés on élimine une éventuelle différence de moyennes entre les deux réalisations puisque le centre de gravité des deux nuages est ramené à l'origine (figure 4).

Mais si on utilise les données brutes pour calculer la différence des deux positions, un nouveau problème survient, qui est attaché aux différences d'unités entre variables. La normalisation dans l'ACP normée résout cette question, la variabilité de chaque variable étant ramenée à l'unité.

### 3.2 — Normalisation des différences

L'option PCA : Normed  $Y[i,j] - X[i,j]$  propose une méthode originale très simple pour contourner la difficulté dans le cas d'une différence de tableaux. Soient  $x_{ij}^1$  et  $x_{ij}^2$  les valeurs observées de la variable  $i$  sur l'individu  $j$  dans les tableaux 1 et 2. La quantité :

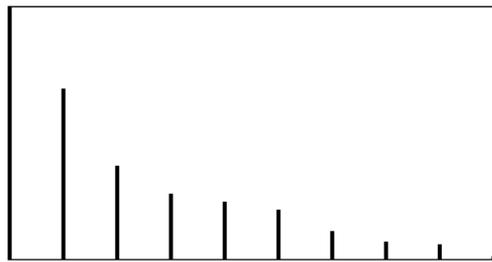
$$\frac{1}{n} \sum_{i=1,n} (x_{ij}^1 - x_{ij}^2)^2 = \vartheta(j)$$

exprime la variabilité inter-tableaux de la variable  $J$ . Transformons le tableau des différences par :

$$z_{ij} = \frac{x_{ij}^1 - x_{ij}^2}{\sqrt{d(j)}}$$

Chaque vecteur des différences est un vecteur normé de  $\mathbb{R}^n$  et justifie une ACP normée implicite mise en œuvre dans l'option :

Normed Y[i,j] - H[i,j]			
X matrix		Mil1	31 10
Y matrix		Mil2	31 10
Output file name		A	



Number of axes ?

On obtient un listing très classique. Les paramètres utilisés sont rappelés :

```
X Table: Mil1
It has 31 rows and 10 columns
Y Table: Mil2
It has 31 rows and 10 columns
Generic output file name: A
---- Row weight:
File A.cnpl contains the row weight
It has 31 rows and 1 column
Each row has 3.2258e-02 weight (Sum = 1)
---- Column weights:
File A.cnpc contains the column weights
It has 10 rows and 1 column
Each column has unit weight (Sum = 10)
---- Table:
File A.cnta contains the table Y-X
It has 31 rows and 10 columns
```

Le schéma est diagonalisés :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: A.cnta
--- Number of rows: 31, columns: 10
-----
Total inertia:      10
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +3.3400E+00 +0.3340 +0.3340 | 02  +2.2502E+00 +0.2250 +0.5590 |
03  +1.2381E+00 +0.1238 +0.6828 | 04  +8.7669E-01 +0.0877 +0.7705 |
05  +7.6389E-01 +0.0764 +0.8469 | 06  +6.6381E-01 +0.0664 +0.9133 |
07  +3.8168E-01 +0.0382 +0.9514 | 08  +2.3289E-01 +0.0233 +0.9747 |
09  +2.1035E-01 +0.0210 +0.9958 | 10  +4.2334E-02 +0.0042 +1.0000 |

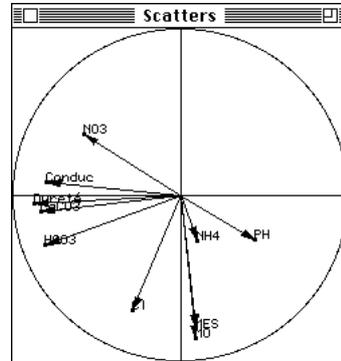
File A.cnvp contains the eigenvalues and relative inertia for each axis
--- It has 10 rows and 2 columns
```

Les coordonnées des colonnes donne un cercle de corrélation :

File A.cnco contains the column scores  
 --- It has 10 rows and 2 columns  
 File :A.cnco

Col.	Mini	Maxi
1	-8.707e-01	4.419e-01
2	-8.622e-01	3.606e-01

Labels	
HV coordinates file	<input type="text" value="A.cnco"/>
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="BuechCodeVar"/>
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>
Draw unit circle (yes = 1)	<input type="text" value="1"/>

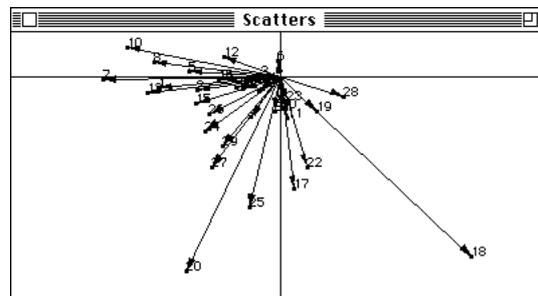


L'axe 1 correspond encore à la charge ionique et l'axe 2 à la charge particulaire. Les coordonnées des lignes donne une carte factorielle qui a un principe particulier :

File A.cnli contains the row scores  
 --- It has 31 rows and 2 columns  
 File :A.cnli

Col.	Mini	Maxi
1	-3.938e+00	4.284e+00
2	-4.395e+00	6.609e-01

Labels	
HV coordinates file	<input type="text" value="A.cnli"/>
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="#"/>
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>
Draw unit circle (yes = 1)	<input type="text"/>



L'origine est la représentation d'un point à la date 1 et l'extrémité est la représentation du même point à la date 2. Reste à l'origine un point ayant strictement la même position dans les deux tableaux. L'extrémité caractérise l'évolution d'une station entre les deux saisons. On peut identifier les cas typiques. La station 3 évolue peu, la station 20 évolue fortement sur les variables MES, MO et Cl, la station 7 évolue fortement sur les variables conductivité, dureté et HCO3. On vérifie ceci sur l'édition des données :

7.8	132	145	155	135	10	1	0.6	0.86	0.01	Sta 3/juin
7.9	213	145	160	140	7.5	3.8	0.5	0.88	0	Sta 3/Sept
8	381	180	230	190	<b>12.5</b>	<b>6.4</b>	<b>0.72</b>	0.54	0	Sta 20/juin
8	420	233	285	220	<b>25</b>	<b>52.2</b>	<b>2.4</b>	1.26	0.04	Sta 20/Sept
8	<b>183</b>	<b>40</b>	<b>65</b>	140	12.5	2.4	0.52	0.54	0.01	Sta 7/juin
7.8	<b>283</b>	<b>155</b>	<b>180</b>	165	10	4	0.45	0.9	0.02	Sta 7/Sept
PH	Conduc	CaCO3	Dureté	HCO3	Cl	MES	MO	NO3	NH4	

On conclut des cartes factorielles que la charge augmente généralement entre les deux saisons mais de deux façons différentes. La projection en individus supplémentaires des données transformées ayant généré par différence les axes factoriels, soit

$$\frac{x_{ij}^1}{\sqrt{\partial(j)}} \quad \text{et} \quad \frac{x_{ij}^2}{\sqrt{\partial(j)}}$$

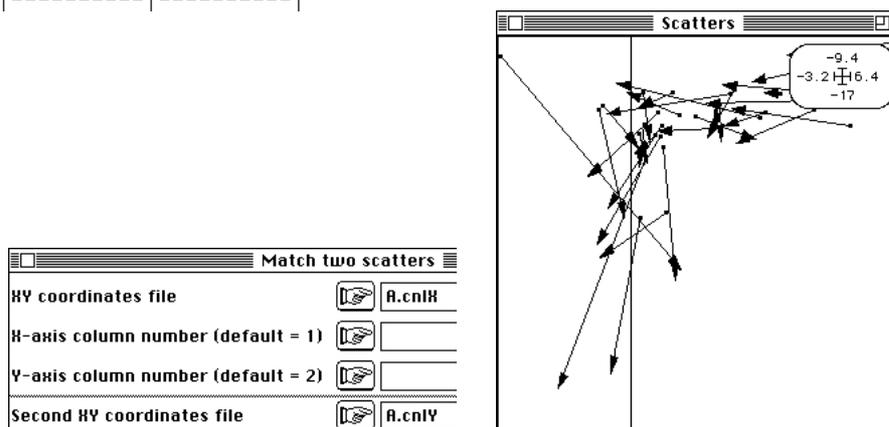
renforce cette appréciation :

File A.cn1X contains the row scores of the non centered table  
Number of rows: 31, columns: 2

```
File :A.cn1X
|Col.|      Mini      |      Maxi      |
|----|-----|-----|
|  1  | -3.114e+00 |  6.385e+00 |
|  2  | -1.287e+01 | -9.523e+00 |
|----|-----|-----|
```

File A.cn1Y contains the row scores of the non centered table  
Number of rows: 31, columns: 2

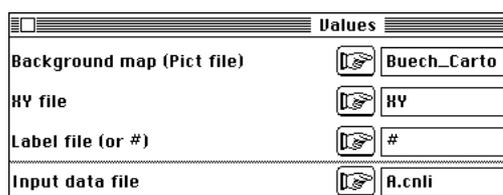
```
File :A.cn1Y
|Col.|      Mini      |      Maxi      |
|----|-----|-----|
|  1  | -1.741e+00 |  3.683e+00 |
|  2  | -1.611e+01 | -9.779e+00 |
|----|-----|-----|
```

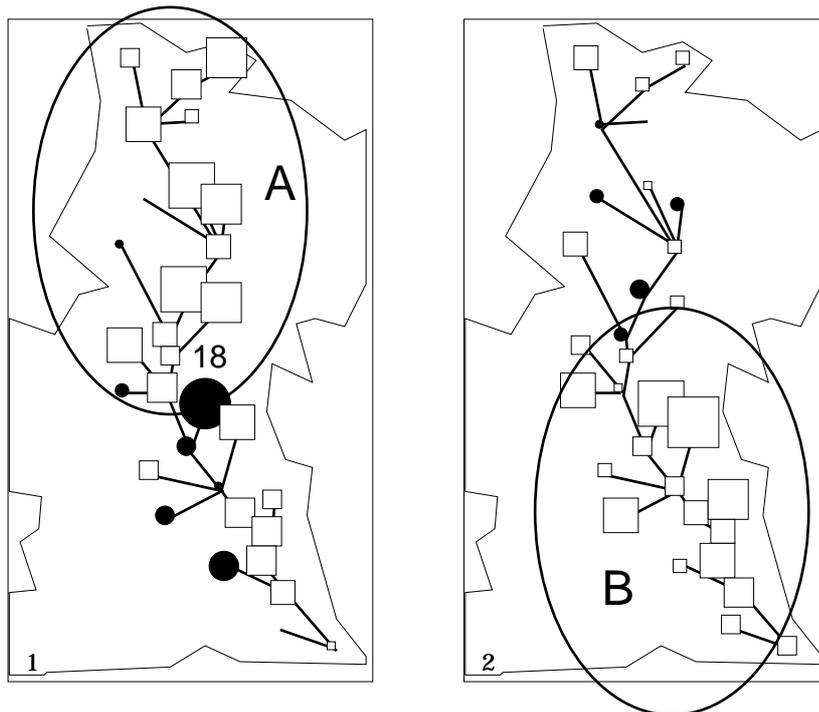
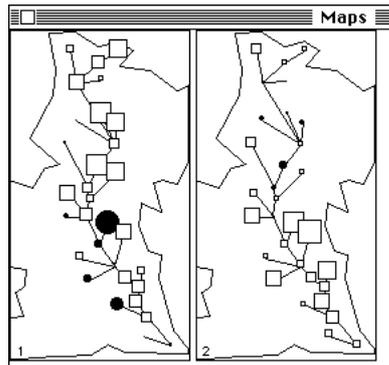


C'est l'opération qui consiste à replacer l'origine des vecteurs à l'origine des axes qui génère l'analyse d'inertie, comme dans une analyse des différences ordinaires (PCA : Decentring  $X[i,j] - Model[i,j]$ ) mais les unités hétérogènes entre variables impose ici l'usage de PCA : Normed  $Y[i,j] - X[i,j]$ .

### 3.3 — Structure spatiale des différences

On pense alors que les types d'évolution des variables eux-mêmes sont régionalisés et il est utile de cartographier les scores d'évolution :





On retrouve une augmentation générale de la charge en carbonates plus sensible dans la partie amont, vraisemblablement liée à une augmentation relative plus forte de la température, et une augmentation de la charge particulaire beaucoup plus forte en aval. Les valeurs observées à la station 18 sont sujettes à discussion.

Nous sommes passés d'une discussion sur la structure spatiale des variables (analyse appariée) à une mise en évidence de la structure spatiale de leur évolution.

Il se présente alors deux questions de généralisation des analyses sur tableaux totalement appariés. La première s'énonce ainsi : la généralisation de l'analyse de co-inertie de 2 à K tableaux <sup>8</sup> a-t-elle un équivalent qui permette de passer de la co-inertie de 2 tableaux complètement appariés à la co-inertie de K tableaux complètement appariés ? Dans l'affirmative, le résultat est-il équivalent à la méthode STATIS sur les tableaux (STATIS : Table averaging) ?

La seconde porte sur la structure de K tableaux totalement appariés définie en référence à un tableau de même type qui sert de structure modèle. On pourra alors parler de structures de données à K+1 tableaux. Quel bilan des ressemblances ou des différences au modèle est-elle possible ?

## Références

- <sup>1</sup> Castella, E., Bickerton, M., Armitage, P.D. & Petts, G.E. (1995) The effects of water abstractions on invertebrate communities in U.K. streams. *Hydrobiologia* : 308, 167-182.
- <sup>2</sup> Carrel, G. & J.C. Berget, J.C. , 1995. *Suivi hydrobiologique du Rhône. Site de Cruas-Meyssse (1982-1995)*. Cemagref Aix-en-Provence, Rapport à EDF, 1-44 + annexes.
- <sup>3</sup> Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* : 43, 109-121.
- <sup>4</sup> Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.
- <sup>5</sup> Tucker, L.R. . (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136.
- <sup>6</sup> Vespini, F. (1985) *Contribution à l'étude hydrobiologique du Buech, rivière non aménagée de Haute-Provence*. Thèse de troisième cycle, Université de Provence. 1-148 + bibliographie + annexes.
- <sup>7</sup> Vespini, F., Légier, P. & Champeau, A. (1987) Ecologie d'une rivière non aménagée des Alpes du Sud : Le Buëch (France) I Evolution longitudinale des descripteurs physiques et chimiques. *Annales de Limnologie* : 23, 151-164.
- <sup>8</sup> Chessel, D. & Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : 44, 35-60.

