

STATICO : STATIS et Co-inertie

Résumé

La fiche décrit le mode d'emploi de l'analyse triadique partielle étendue aux opérateurs de co-inertie. La méthode, qu'on peut appeler STATICO, est une analyse triadique partielle sur les tableaux croisés dans les analyses de co-inertie. On combine alors la logique de STATIS (trouver ce qui dans plusieurs tableaux constitue le fond typologique commun, fond commun qui peut exister derrière de fortes particularités propres à chaque tableau) et la logique des couplages de co-inertie (trouver ce qui dans deux groupes de descripteurs engendre une typologie commune des objets décrits dans chacun des tableaux). Cette approche fait suite à une demande de l'équipe de D. Nandris (Phytopathologie, ORSTOM, Nouméa).

Plan

1 — Principe général.....	2
2 — Premier exemple de mise en œuvre.....	4
3 — Utilisation des résultats.....	8
3.1 — Paramètres globaux.....	8
3.2 — Compromis de co-structure.....	11
3.3 — Stabilité des réalisations du compromis.....	14
4 — Typologie stable et environnement fluctuant	18
5 — Moyennes et co-structures	23
5.1 — Variabilité environnementale.....	24
5.2 — Variabilité épidémiologique	27
5.3 — Co-Variabilité.....	31
5.4 — Co-structure des moyennes.....	36
Références	39

D. Chessel, M. Simier, M. Hanafi

1 — Principe général

On s'intéresse ici aux enquêtes écologiques à composantes temporelles. Le tableau faunistique X comprend les abondances de p taxons (en colonnes) relevées dans s stations (en lignes). Le tableau mésologique Y comprend les valeurs de q variables quantitatives (en colonnes) caractérisant chacune des s stations. Les mesures ayant été réalisées pour t dates différentes, les n ($n = st$) relevés sont regroupés par dates en t blocs, chacun d'entre eux comportant s stations. Plus généralement les blocs peuvent être incomplets.

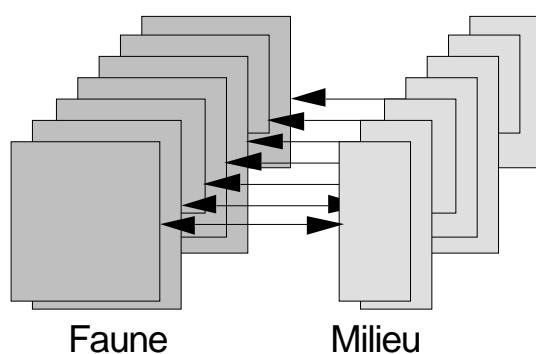
Caractériser les relations existant entre le tableau faunistique X et le tableau mésologique Y est un problème classique en écologie. Ce problème est au centre de toutes les études visant :

- de façon dissymétrique¹ à expliquer une typologie faunistique des stations à l'aide de plusieurs variables de milieu (notion d'écologie factorielle, particulièrement importante dans le suivi écologique des rivières²), ou inversement à prédire un ensemble de variables de milieu à l'aide d'une structure faunistique (notion de bio-indicateurs et reconstitution des paléo-environnements³).

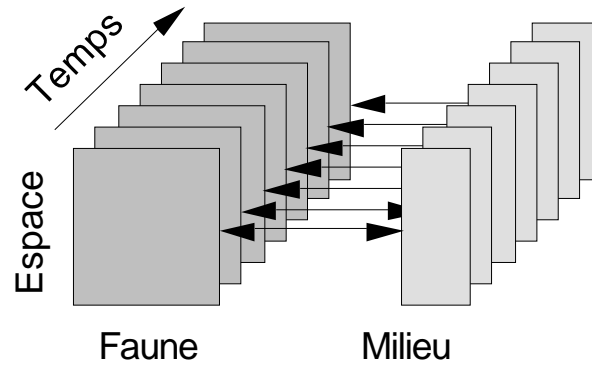
- de façon symétrique à établir la co-typologie des stations issue des données faunistiques et celle issue des données de milieu. Actuellement, les méthodes de couplage de deux tableaux (analyse de co-inertie) permettent la mise en évidence d'une co-structure entre relevés mésologiques et faunistiques pour tous les types de données⁴.

Caractériser la stabilité des relations existant entre le tableau faunistique et le tableau mésologique, quand l'observation a été renouvelée, est une question récente qui a été peu étudiée au plan méthodologique. On peut citer les analyses de co-inertie inter et intra-classes⁵. On aborde ici la question par le biais de la logique des méthodes STATIS⁶.

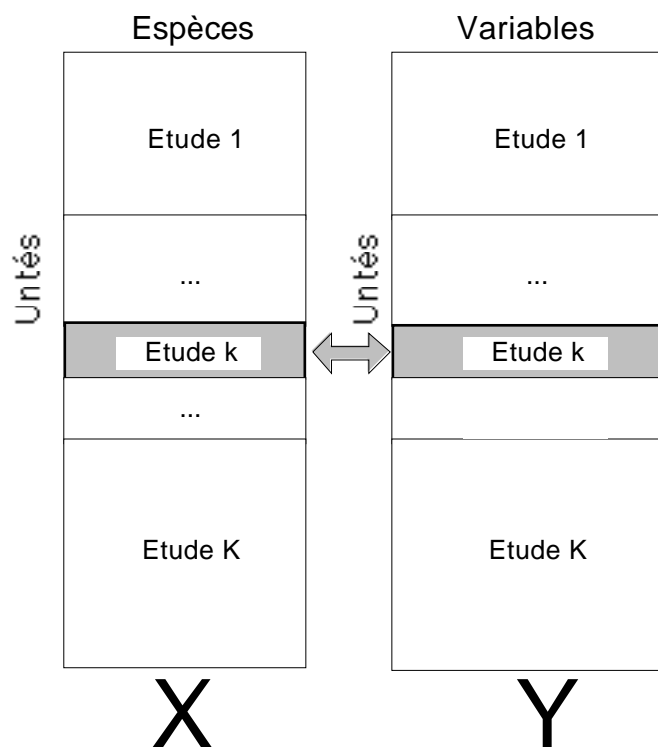
Les données traitées ont la forme suivante :



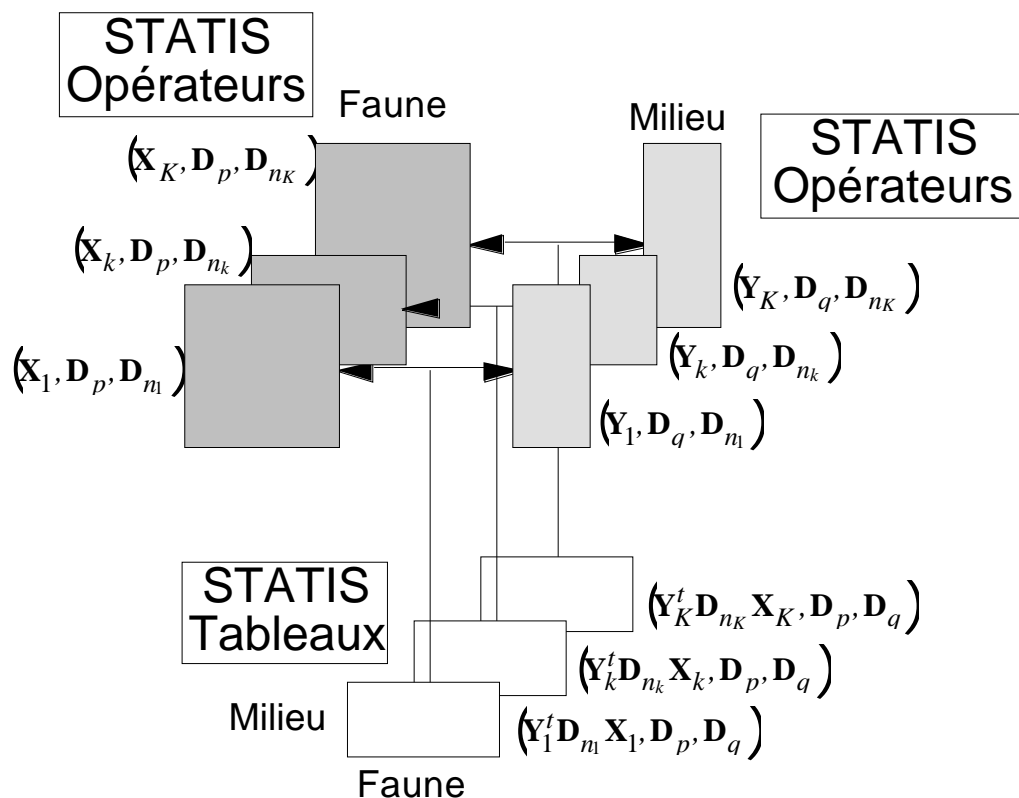
Dans le cas d'un échantillonnage complet de toutes les stations à toutes les dates de mesure, on obtient un cas particulier formé d'un couple de cubes de données :



On notera, dans tout ce qui suit 1, 2, ..., K les répétitions de l'observation, appelées encore études dans la terminologie de STATIS.



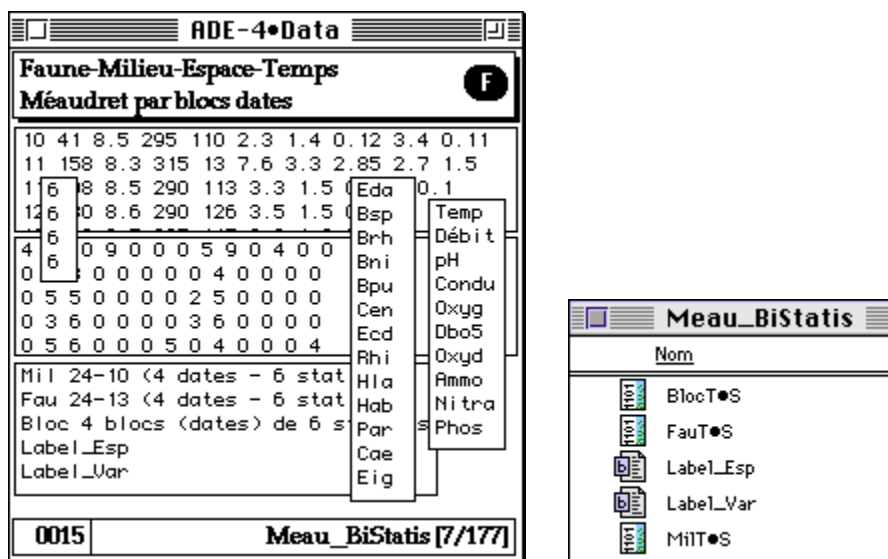
L'observation k est formée d'un tableau \mathbf{X}_k donnant l'abondance de p taxa (variables du groupe 1) dans n_k unités de mesure et d'un tableau \mathbf{Y}_k donnant la valeurs de q variables environnementales (variables du groupe 2). La liste des variables des deux groupes est la même pour toutes les répétitions. Les deux tableaux sont insérés dans un schéma de dualité⁷, respectivement $(\mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_{n_k})$ et $(\mathbf{Y}_k, \mathbf{D}_q, \mathbf{D}_{n_k})$. d'un type quelconque. Les métriques diagonales \mathbf{D}_p et \mathbf{D}_q sont fixées et indépendantes de la répétition. La métrique des poids des unités \mathbf{D}_{n_k} est commune au deux schémas, ce qui autorise l'analyse de co-inertie des deux tableaux. L'analyse de co-inertie associée à l'étude k est celle du schéma $(\mathbf{Y}_k^t \mathbf{D}_{n_k} \mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_q)$. Les k schémas de co-inertie partagent la même dimension des tableaux et les mêmes métriques.



L'extension de STATIS proposée ici est l'application de STATIS sur les tableaux (ou analyse triadique partielle⁸) aux tableaux croisés de chaque étude, qui sont alors des opérateurs de co-inertie. L'objectif est de mesurer la reproductibilité d'une co-structure faune-environnement vue comme compromis d'une série d'études de couple de tableaux.

2 — Premier exemple de mise en œuvre

Utiliser la carte Meau_BiStatist de la pile ADE-4•Data :



Le bouton F crée un dossier de travail contenant les fichiers binaires BlocT•S (4 fois 6 lignes), FauT•S (24 relevés de 10 variables) et MilT•S (24 relevés de macrofaune benthique de 13 espèces d'Ephéméroptères) et les fichiers texte Label_Var (étiquettes des variables) et Label_Esp (étiquettes des taxons)⁹.

Enregistrer le K-tableaux faunistique (KTabUtil : InitKTab) :

Centrer les données par bloc et par colonnes (espèces) par KTabUtil : CentringKtab :

On obtient le K-tableaux centré avec les fichiers X.ktpc (13-1), X.ktpl (24-1) et X.ktta (24-13). On pourrait exécuter ici un STATIS sur les tableaux (STATIS : Table averaging).

Ces données illustrent les AFC inter et intra dates dans ¹⁰

Enregistrer le K-tableaux mésologique (KTabUtil : InitKTab) :

Normaliser par KTabUtil : NormelizeKtab :

On a utilisé le centrage par bloc suivi de la normalisation globale qui ramène les variances globales à l'unité en conservant d'éventuelles modification dans le temps de la variance par variables et par blocs (Option de Bouroche¹¹). On obtient le K-tableaux normalisé avec les fichiers Y.ktpc (10-1), Y.ktpl (24-1) et Y.ktta (24-10). On pourrait

exécuter ici un STATIS sur les tableaux (STATIS : Table averaging). Ces données ont été décrites dans ¹².

Coupler les deux K -tableaux par KTabUtil : MatchTwoKTab :

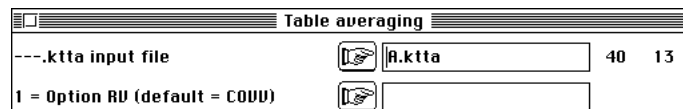


On obtient le descriptif de la structure d'une troisième K -tableaux comportant :

```
Output ASCII file: A!.ktp
-----
Qualitative variables file: A!TLl
Number of rows: 40, variables: 3, categories: 54
Auxiliary ASCII output file A!TLl.label: labelling file
Les 40 lignes de A!TLl correspondent aux 4 ( $K$ ) blocs de 10 ( $q$ ) variables.
-----
Qualitative variables file: A!TCc
Number of rows: 52, variables: 3, categories: 30
Auxiliary ASCII output file A!TCc.label: labelling file
Les 52 lignes de A!TCc correspondent aux 4 ( $K$ ) blocs de 13 ( $p$ ) espèces.
-----
Qualitative variables file: A!T4a
Number of rows: 16, variables: 2, categories: 8
Auxiliary ASCII output file A!T4a.label: labelling file
-----
File A.ktpl contains the row weights
It has 40 rows and 1 column
File A.ktpc contains the column weights
It has 13 rows and 1 column
File A.ktta contains the crossed K-table
It has 40 rows and 13 columns
```

Le tableau à 40 lignes et 13 colonnes est formé de la superposition des 4 (K) tableaux de 10 (q) lignes-variables et de 13 (p) colonnes-espèces. L'analyse de chacun de ces quatre tableaux est celle l'analyse de co-inertie couplant les tableaux à 6 (n_k) lignes-stations et 13 colonnes-espèces et à 6 (n_k) lignes-sations et 10 colonnes-variables.

Exécuter alors STATIS : Table averaging sur ce dernier K -tableaux :



```
Option : Table averaging
Input file A
Number of rows: 40, columns: 13
----- Correlation matrix -----
[ 1] 1000
[ 2] 450 1000
[ 3] 354 655 1000
[ 4] 187 354 643 1000
-----
File A.xa+RV contains cosinus between tables
It has 4 rows and 4 columns
File A.xa+CV contains inner products between tables
It has 4 rows and 4 columns
Compromise with COVV matrix
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.6901E+00  +0.8134  +0.8134  | 02  +2.2382E-01  +0.1077  +0.9211  |
03  +8.2595E-02  +0.0398  +0.9609  | 04  +8.1300E-02  +0.0391  +1.0000  |
```

File A.xa+EV contains the eigenvalues of interstructure analysis
It has 4 rows and 1 columns

File A.xa+IS contains the table scores of the interstructure analysis
It has 4 rows and 4 columns

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.5480E+00	+0.9159	+0.9159	02	+8.3825E-02	+0.0496	+0.9655
03	+5.0496E-02	+0.0299	+0.9954	04	+5.3826E-03	+0.0032	+0.9986
05	+1.2805E-03	+0.0008	+0.9993	06	+7.6555E-04	+0.0005	+0.9998
07	+2.2438E-04	+0.0001	+0.9999	08	+9.8368E-05	+0.0001	+1.0000
09	+2.4273E-05	+0.0000	+1.0000	10	+0.0000E+00	+0.0000	+1.0000

File A.xa+vp contains the eigenvalues of compromise analysis
It has 13 rows and 1 columns

File A.xa+co contains column scores (Norm = 1 for column weights)
in the analysis of the compromise
It has 13 rows and 2 columns

File :A.xa+co

Col.	Mini	Maxi
1	-1.775e+00	-1.015e-01
2	-2.293e+00	1.633e+00

File A.xa+li contains row scores (Norm = 1 for row weights)
in the analysis of the compromise
It has 10 rows and 2 columns

File :A.xa+li

Col.	Mini	Maxi
1	-9.733e-01	1.318e+00
2	-3.448e-01	2.553e+00

File A.xaTLl contains standard row scores with lambda norm
It has 40 rows and 2 columns
It is to be used with --TLl.label and --TLl.cat files

File :A.xaTLl

Col.	Mini	Maxi
1	-1.095e+00	1.546e+00
2	-1.765e-01	6.449e-01

File A.xaTCO contains columns scores issued from each table by
It has 52 rows and 2 columns
It is to be used with --TCc.label and --TCc.cat files

File :A.xaTCO

Col.	Mini	Maxi
1	-2.179e+00	1.595e-01
2	-5.052e-01	4.429e-01

Typological value indices
Rows = row number of each table
Weights = Weights of tables in the compromise
NS norm2 = Squared norm (Hilbert-Schmidt) of tables
Cos2 = Squared cosinus between table and approximed compromise

Number	Rows	Weights	NS norm2	Cos2
1	10	1.042e-01	1.088e-01	0.131
2	10	3.725e-01	4.125e-01	0.556
3	10	9.005e-01	1.402e+00	0.945
4	10	1.989e-01	1.545e-01	0.410

File A.xaTComp contains component scores issued from each table by separate analysis

It has 16 rows and 2 columns

It is to be used with --T4A.label and --T4A.cat files

File :A.xaTComp

Col.	Mini	Maxi
1	-7.012e-01	9.981e-01
2	-3.830e-01	9.071e-01

File A.xaTAxe contains axe scores issued from each table by separate analysis

It has 16 rows and 2 columns

It is to be used with --T4A.label and --T4A.cat files

File :A.xaTAxe

Col.	Mini	Maxi
1	-4.526e-01	9.927e-01
2	-4.570e-01	8.078e-01

File A.xaTXLl contains normed row scores (from table X.ktta)

It has 24 rows and 2 columns

It is to be used with FauT•S!TLl.label and FauT•S!TLl.cat files

File :A.xaTXLl

Col.	Mini	Maxi
1	-1.641e+00	2.595e+00
2	-2.635e+00	1.622e+00

File A.xaTYLl contains normed row scores (from table Y.ktta)

It has 24 rows and 2 columns

It is to be used with Milt•S!TLl.label and Milt•S!TLl.cat files

File :A.xaTYLl

Col.	Mini	Maxi
1	-1.230e+00	3.241e+00
2	-1.750e+00	1.625e+00

On reprend ci-dessous point par point l'interprétation des résultats.

3 — Utilisation des résultats

Comme dans toute stratégie du type STATIS, la méthode ici décrite, que nous pouvons appeler STATICO (pour Statis et Co-inertie), calcule une moyenne (compromis), analyse cette moyenne (analyse du compromis) et utilise les résultats pour l'examen des éléments constitutifs de cette moyenne (intra-structure).

3.1 — Paramètres globaux

Nous passons rapidement sur la première phase (inter-structure) en privilégiant la fonction de l'axe 1 de cette inter-structure, qui est toujours une fonction de définition d'un élément commun aux constituants du compromis, par rapport aux axes suivants qui ont une fonction d'expression d'éléments spécifiques et qui ne sont pas explicités.

A la date k , le tableau faunistique inséré dans le schéma $(\mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_{n_k})$ et le tableau environnemental inséré dans le schéma $(\mathbf{Y}_k, \mathbf{D}_q, \mathbf{D}_{n_k})$ donne le schéma de co-inertie

$(\mathbf{Y}_k^t \mathbf{D}_{n_k} \mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_q)$. Notons $\mathbf{Z}_k = \mathbf{Y}_k^t \mathbf{D}_{n_k} \mathbf{X}_k$. Trouver un compromis de co-inertie c'est trouver des nombres α_k tels que $\sum_{k=1}^K \alpha_k \mathbf{Z}_k, \mathbf{D}_p, \mathbf{D}_q$ présente une inertie maximum sous la contrainte $\sum_{k=1}^K \alpha_k^2 = 1$. Les nombres α_k sont les composantes du premier vecteur propre normé de la matrice des produits scalaires d'Hilbert-Schmidt des opérateurs de co-inertie définis par $(\mathbf{Z}_k | \mathbf{Z}_j)_{HS} = \text{Trace}(\mathbf{Z}_k^t \mathbf{D}_q \mathbf{Z}_j \mathbf{D}_p)$.

C'est pourquoi, dans STATIS sur les tableaux, il est logique de diagonaliser la matrice des COVV d'Escoufier¹³. Les tableaux de co-inertie portant sur les mêmes lignes et les mêmes colonnes, formés de covariances calculées sur un même nombre d'observations, peuvent avoir des inerties variables, mais cela est ici considérée comme une qualité propre des données. Ces covariances vont augmenter avec les variances mais nous avons conservé cette possibilité intentionnellement dans le tableau faunistique (qu'on ne normalise pratiquement jamais) et le tableau de milieu qui n'a pas été normalisé par dates à cet effet.

La matrice des RV est évidemment plus simple à lire, c'est pourquoi elle est directement éditée, mais on peut également lister la matrice des produits scalaires :

```
Option : Table averaging
Input file A
Number of rows: 40, columns: 13
----- Correlation matrix -----
[ 1] 1000
[ 2] 450 1000
[ 3] 354 655 1000
[ 4] 187 354 643 1000
-----
File A.xa+RV contains cosinus between tables
It has 4 rows and 4 columns
File A.xa+CV contains inner products between tables
It has 4 rows and 4 columns
```

Les tableaux numérotés de 1 à 4 se réfèrent aux saisons Printemps, Été, Automne et Hiver.

A.xa+CV-t			
0.1088	0.09534	0.13818	0.02419
0.09534	0.41247	0.4978	0.08925
0.13818	0.4978	1.4021	0.29944
0.02419	0.08925	0.29944	0.1545

On retrouve sur la diagonale les valeurs éditées dans le tableau :

Number	Rows	Weights	NS norm2	Cos2
1	10	1.042e-01	1.088e-01	0.131
2	10	3.725e-01	4.125e-01	0.556
3	10	9.005e-01	1.402e+00	0.945
4	10	1.989e-01	1.545e-01	0.410

La co-inertie est donc beaucoup plus forte en automne. Le module KTA : Separate analyses donnerait une vision complète de la variabilité à chaque date des deux tableaux :



On trouve les inerties 3.83, 4.05, 7.60 et 4.41.



On trouve les inerties 0.136, 0.314, 0.444 et 0.106. Simultanément la variabilité mésologique et la diversité faunistique sont faibles en hiver et au printemps, mais entre été et automne la différence de co-structure n'est peut-être pas seulement associée à des différences de structure. Des quatre saisons, l'automne est-elle le moment d'expression faunistique des contraintes environnementales ?

Pour mesurer la co-structure comme coefficient de corrélation dans la théorie des multi-tableaux, il faut se mettre dans un espace unique. Les RV édités ci-dessus sont des corrélations entre co-structure, basés sur :

$$\left(\mathbf{Z}_k | \mathbf{Z}_j \right)_{HS} = Trace \left(\mathbf{Z}_k^t \mathbf{D}_q \mathbf{Z}_j \mathbf{D}_p \right)$$

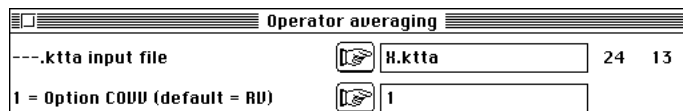
alors que maintenant nous voulons discuter de la co-structure comme d'une corrélation.

Rappelons que les tableaux \mathbf{X}_k sont dans $\mathbb{R}^{p^*} \times \mathbb{R}^{n_k}$, les tableaux \mathbf{Y}_k sont dans $\mathbb{R}^{q^*} \times \mathbb{R}^{n_k}$, les tableaux croisés sont dans $\mathbb{R}^{p^*} \times \mathbb{R}^q$. La norme HS de \mathbf{Z}_k est :

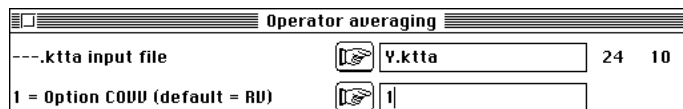
$$\|\mathbf{Z}_k\|_{HS}^2 = \left(\mathbf{Z}_k | \mathbf{Z}_k \right)_{HS} = Trace \left(\mathbf{Z}_k^t \mathbf{D}_q \mathbf{Z}_k \mathbf{D}_p \right) = Trace \left(\mathbf{X}_k^t \mathbf{D}_{n_k} \mathbf{Y}_k \mathbf{D}_q \mathbf{Y}_k^t \mathbf{D}_{n_k} \mathbf{X}_k \mathbf{D}_p \right)$$

$$\text{Donc } \|\mathbf{Z}_k\|_{HS}^2 = Trace \left(\mathbf{X}_k \mathbf{D}_p \mathbf{X}_k^t \mathbf{D}_{n_k} \mathbf{Y}_k \mathbf{D}_q \mathbf{Y}_k^t \mathbf{D}_{n_k} \right) = Trace \left(\mathbf{W}_X \mathbf{D}_{n_k} \mathbf{W}_Y \mathbf{D}_{n_k} \right)$$

Il faut donc comparer ce produit scalaire aux normes des opérateurs dans le STATIS sur les $\mathbf{W}\mathbf{D}$ équivalents (et non sur les \mathbf{X} et les \mathbf{Y} , assez curieusement). Il est donc logique d'utiliser STATIS : Operator averaging :



Number	Rows	Weights	NS norm2	Cos2
1	6	2.025e-01	6.093e+00	0.177
2	6	2.786e-01	5.810e+00	0.406
3	6	8.037e-01	2.170e+01	0.962
4	6	4.852e-01	9.325e+00	0.798



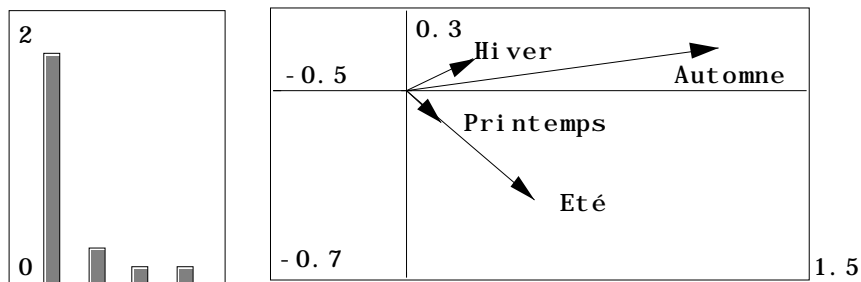
Number	Rows	Weights	NS norm2	Cos2
1	6	1.149e-01	8.317e-03	0.238
2	6	4.840e-01	4.695e-02	0.773
3	6	8.649e-01	1.210e-01	0.970
4	6	6.739e-02	4.921e-03	0.123

Ceci permet de calculer la corrélation vectorielle (RV) de co-inertie :

Date	COVV(X,Y)	VarV(X)	VarV(Y)	RV(X,Y)
printemps	0.1088	6.093	.00832	0.483
été	0.4125	5.810	.04695	0.790
automne	1.4021	21.70	.1210	0.865
hiver	0.1545	9.325	.00492	0.721

On voit donc, dans la covariance vectorielle, deux composantes, d'une part celle de l'intensité de la liaison espèces-environnement, relativement stable et d'autre part celle de l'intensité de l'expression de cette liaison qui évolue beaucoup. Printemps et hiver n'ont donc pas du tout le même statut. Variabilité faunistique et mésologique sont voisines (inertie de 3.83 et 4.41, 0.136 et 0.106). Le niveau de structures de cette variabilité sont déjà différentes (VV(X) 6.093 et 9.325, 0.0083 et 0.0049) mais la corrélation au printemps est plus faible qu'en hiver (0.48 contre 0.72) : nous verrons qu'elle est également nettement moins ajustée à la corrélation compromise. Avec les méthodes multivariées on passe de la mesure de la dispersion des valeurs à celle de la corrélation. Avec les méthodes multi-tableaux, on passe de la mesure de la variabilité par tableau à celle de l'intensité de la co-structure entre tableaux.

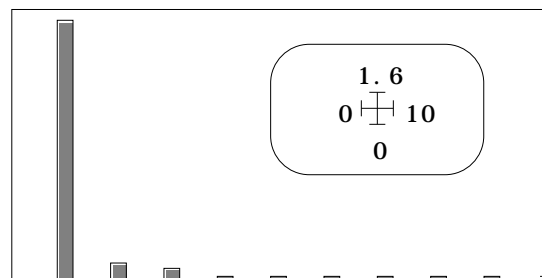
Les valeurs propres de inter-structure peuvent être récupérées directement dans A.xa+EV et l'image euclidienne associée dans A.xa+IS.



Nous n'accorderons pas trop d'importance à cette image euclidienne dont on connaît mal la valeur d'usage. Il doit y avoir des composantes particulières de la relation faune-milieu à chaque saison. On ne s'intéresse dans ce qui suit qu'à la composante stable exprimée par l'analyse du compromis.

3.2 — Compromis de co-structure

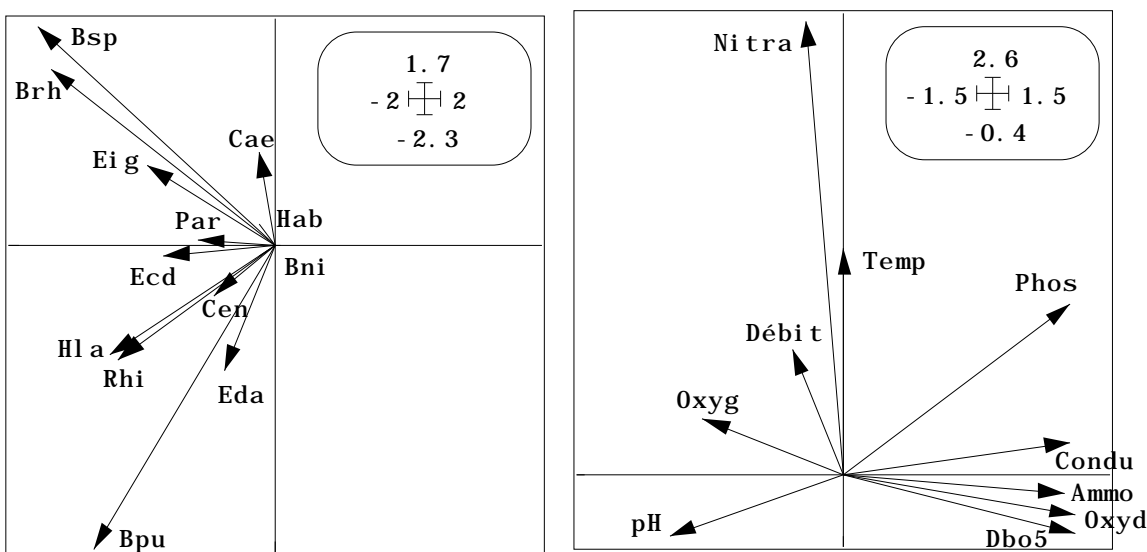
Le compromis est une analyse de co-inertie fictive dont le tableau croisé est une moyenne inter-dates de tableaux croisés. Ce compromis définit des valeurs propres dans A.xa+vp :



Une vision unidimensionnelle de la co-structure semble s'imposer (92%). Nous discuterons donc de la première composante. Mais l'utilisation de la seconde réserve bien des surprises et confirme, une fois de plus, combien la manipulation des statistiques d'inertie doit intervenir avec une bonne connaissance des supports. Le

compromis donne une ordination des taxons et une ordination des variables. Le fichier A.xa+co donne les scores des taxons (colonnes du tableau X donc colonnes du tableau Y^tDX) et le fichier A.xa+li donne les scores des variables (colonnes du tableau Y donc lignes du tableau Y^tDX). Ces scores sont les composantes des axes et composantes principales du tableau croisé compromis. On peut représenter graphiquement ces composantes (Scatters : Labels) pour comprendre comment seront constitués les scores des projections sur ces vecteurs) ou utiliser ces vecteurs pour voir comment s'organisent les nuages d'objets projetés. La cohérence mathématique de ces objets génère l'information.

On reconnaît ci-dessous sur les plans 1-2 un nuage de points-espèces caractéristique de l'effet taille et un nuage de points variables caractéristique des éléments pollution-restauration du cours d'eau. La pollution mesurée par des descripteurs fortement redondants (Ammoniaque, Oxygène, DBO, Oxydabilité) et son lien à la charge minérale et organique (Conductivité, Phosphore) génère de fortes inerties.



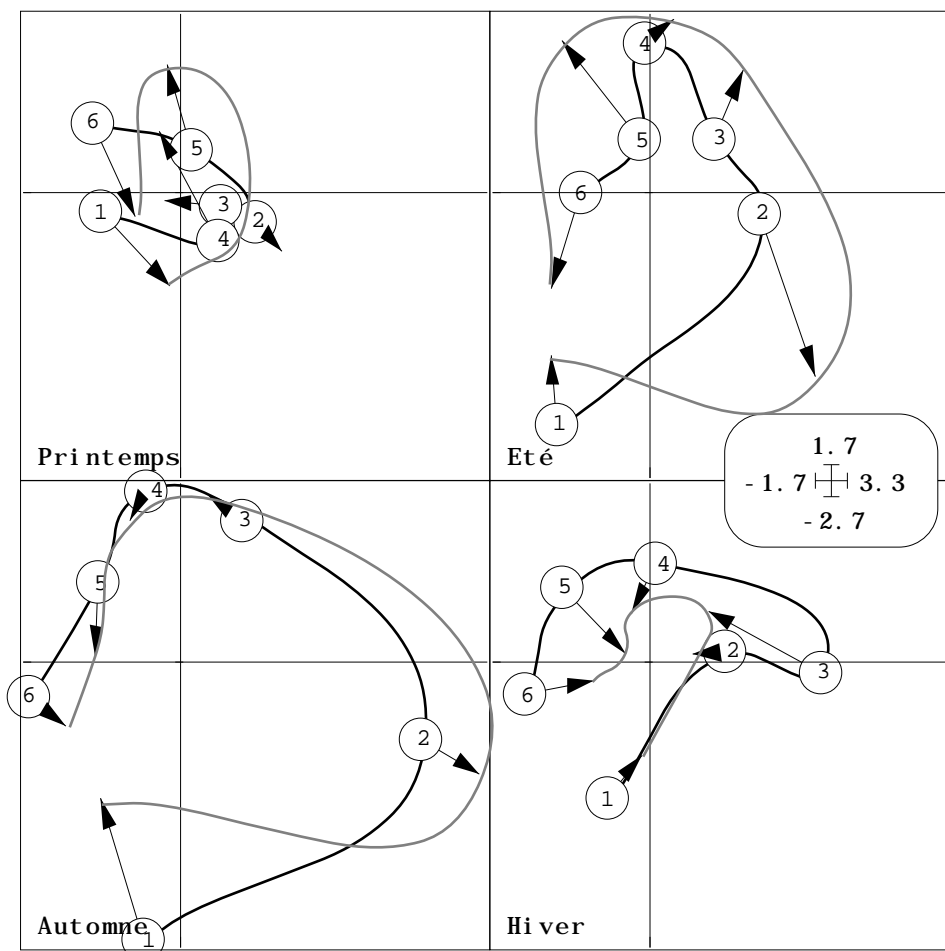
Match two scatters	
HY coordinates file	<input type="text" value="A.naTHLI"/> 24 2
X-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Second HY coordinates file	<input type="text" value="A.naTYLI"/> 24 2
Label file (or #) for rows	<input type="text" value="MIT•S!TLI.label"/>
Label file (or #) for groups	<input type="text" value="Label_Saison"/>

Row & col. selection	
Col. selection:	<input type="text"/>
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="MIT•S!TLI.ca"/>
Selection col. number:	<input type="text" value="1"/>

A la pollution est associée une diminution de l'abondance faunistique globale, qui entraîne une forte variabilité organisée des abondances faunistiques (elles diminuent toutes ensemble). Il s'en suit une accumulation massive des toutes les covariances espèces- variables associées à la diminution de la diversité sous l'effet de la pollution et un taux de co-inertie cumulée de 92%. A s'en tenir à ce point de vue, on ne verrait pas ce qu'exprime l'axe 2. Après la pollution, la rivière se restaure, mais son débit et sa température augmente et le contenu faunistique aval est distinct du contenu faunistique amont. D'où une structure suivant les axes pour les variables mésologiques et une structure suivant les bissectrices pour les variables faunistiques.

Ce qui permet d'affirmer cette interprétation, c'est la possibilité de projeter sur les axes faunistiques toutes les lignes du multi-tableau faunistique et sur les axes mésologiques toutes les lignes du tableau mésologique, de normer globalement les coordonnées pour retenir la partie corrélation dans la co-inertie et d'apparier les nuages

de lignes par dates. La figure est saisissante de signification écologique puisqu'on y voit, outre la justification du modèle décrit (résumé ci-dessous), pour la première fois, la dynamique temporelle d'une relation faune-milieu.

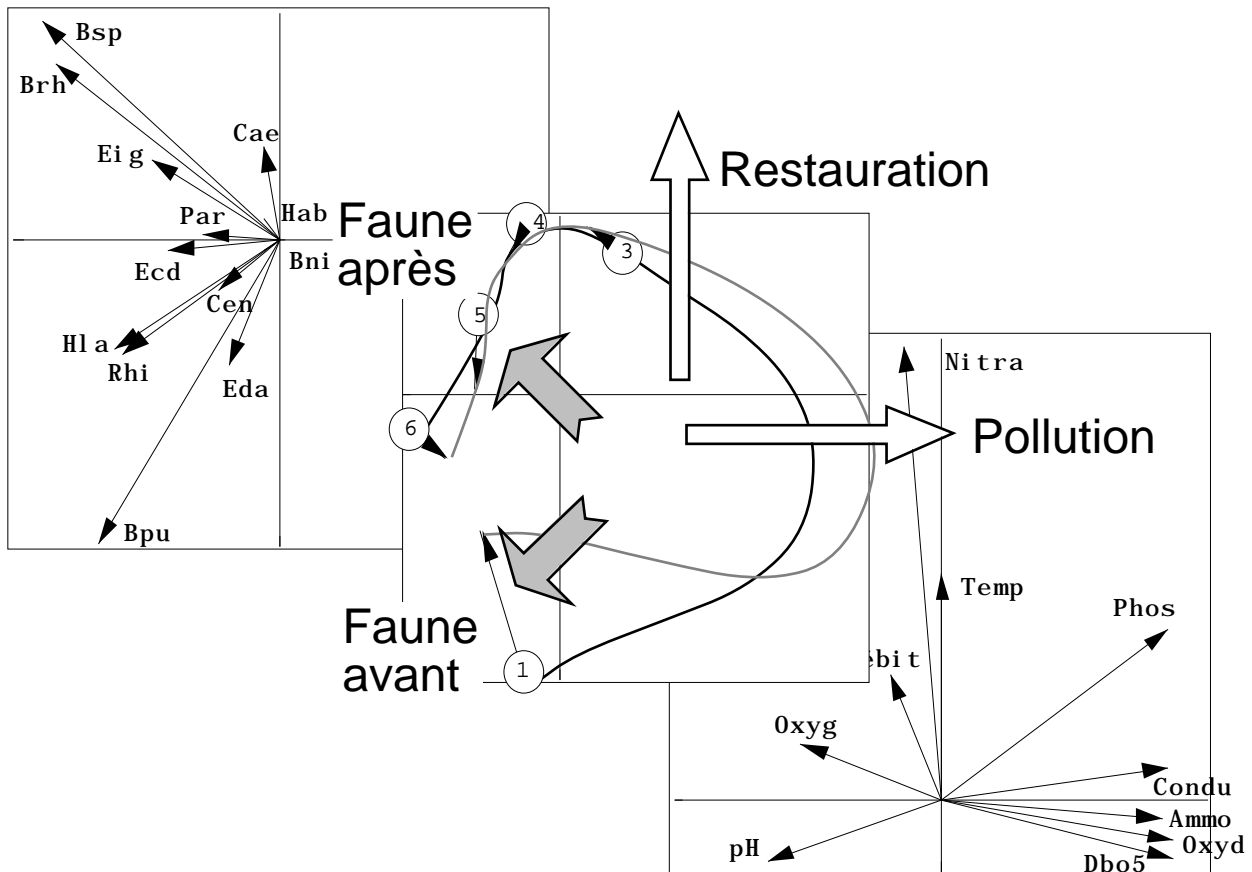


Les composantes de la figure de la page 13 est caractéristique de l'interaction pluridisciplinaire en jeu dans le logiciel ADE. La composante écologique est fortement présente dans les données qui donne un sens à la figure, la composante statistique apporte la possibilité numérique de tracer la figure en réunissant la logique STATIS et la logique co-inertie, la composante logicielle l'autorise par la structure même du fonctionnement du multi-fenêtrage (sélection par paquets de lignes) qui permet une même pratique de toutes les options d'un module. Noter l'usage des fichiers des deux niveaux de multi-tableaux, respectivement A pour les valeurs et X ou Y pour le multi-fenêtrage par couple de tableaux. Ceci est indiqué dans le listing par :

```
File A.xaTXLl contains normed row scores (from table X.ktta)
It has 24 rows and 2 columns
It is to be used with FauT•S!TLl.label and FauT•S!TLl.cat files
```

```
File A.xaTYLl contains normed row scores (from table Y.ktta)
It has 24 rows and 2 columns
It is to be used with MilT•S!TLl.label and MilT•S!TLl.cat files
```

Les contenus des deux fichiers FauT•S!TLl.label et MilT•S!TLl.label ou FauT•S!TLl.cat et MilT•S!TLl.cat sont évidemment identiques pour assurer l'analyse.



3.3 — Stabilité des réalisations du compromis

Nous l'avons déjà dit. STATIS, dans sa présentation classique, permet mal à l'utilisateur non professionnel d'identifier les fonctions de la méthode. L'inter-structure laisse à penser qu'on pourrait faire une typologie de structure. C'est vrai en partie. Ici la figure obtenue par A.xa+IS (page 12) oppose automne et été et indique qu'il y a des composantes de la relation faune-milieu originales à ces deux dates.

C'est peut-être le cas, mais aucun outil pour expliciter ces différences ne sont disponibles dans cette logique.

Par contre, ce qui est au cœur de la méthode c'est la possibilité de discuter de l'évolution de la réalisation d'une structure commune, donc ici d'une co-structure commune. Déjà observée sur les nuages de points-stations, cette possibilité est aussi en jeu pour les variables, les espèces et les axes des analyses de co-inertie des analyses séparées.

Pour les variables, on utilise :

```
File A.xaTl1 contains standard row scores with lambda norm
It has 40 rows and 2 columns
It is to be used with --Tl1.label and --Tl1.cat files
```

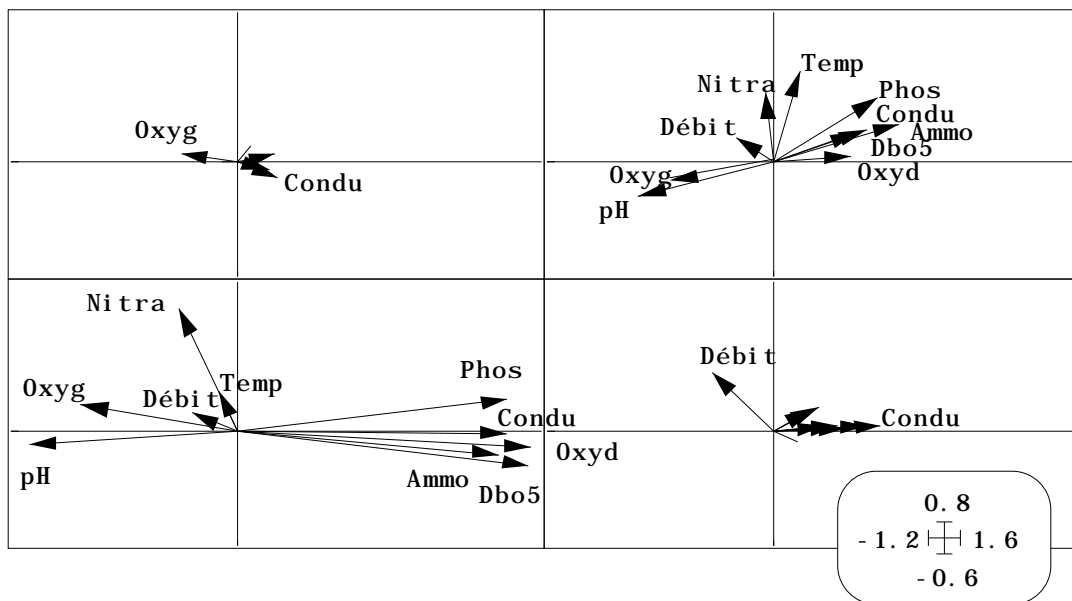
Chaque tableau croisé d'une analyse de co-inertie élémentaire a pour lignes les variables et chaque ligne a un profil de covariances avec toutes les espèces. C'est un point de R^p qui peut être projeté sur les axes principaux du tableau croisé compromis (individus supplémentaires). Dans la pratique des trajectoires, un point, ici une variable, est représenté pour chaque date et l'enchaînement de ces représentations définit une trajectoire¹⁴. On peut penser qu'il s'agit d'une erreur tactique. Le compromis est une moyenne de structure. L'intrastructure est analyse de la reproductibilité par étude de cette structure. Fondamentalement, il convient de représenter dans l'intrastructure des éléments d'expression du compromis, donc des nuages de points, des nuages de couples de points (ci-dessus, nous n'avons pas superposer des positions d'une station, mais

juxtaposer des couples de nuages de stations), des nuages de variables ou de taxons, etc. **Il faut systématiquement multifenêtrer par tableaux et non par objet pour respecter la logique conceptuelle de STATIS :**

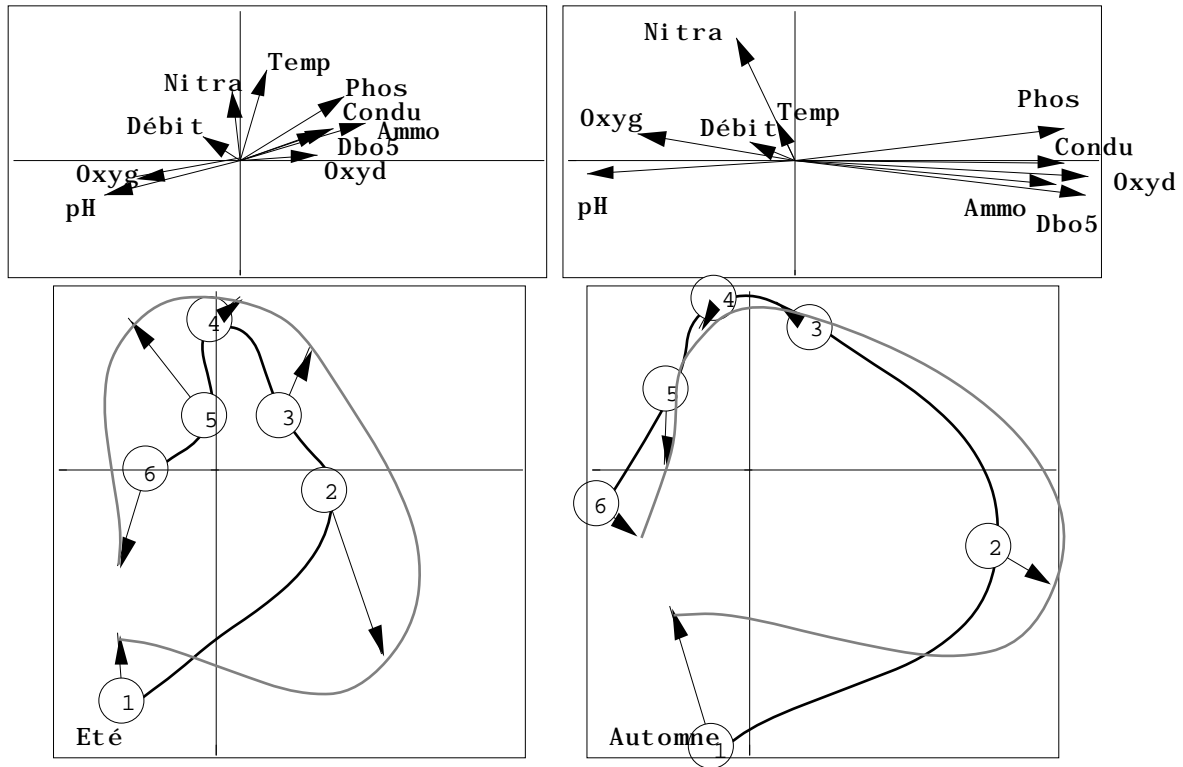
Labels		
X-Y coordinates file		R.naTLI 40 2
X-axis column number (default = 1)		
Y-axis column number (default = 2)		
Label file (or # for item numbers)		R!TLI.label_1
Draw vectors from origin (yes = 1)		1
Draw unit circle (yes = 1)		
Draw points (no = 2)		2

Row & col. selection	
Col. selection:	
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	R!TLI.cat
Selection col. number:	1

On obtient :



On retrouve les très forte différences inter-saisons de variabilités mésologiques mais on souligne ici les difficultés estivales de restauration la rivière :



Ce n'est pas la position d'une variable qui s'interprète dans STATIS mais la forme d'une structure, ici l'ensemble des covariances espèces-variables, exprimée par un nuage de variables dont la forme évolue.

Les éléments (stations, variables, espèces) sont des éléments constitutifs d'un ensemble relationnel. STATIS permet de discuter de la stabilité de cet ensemble autour de la moyenne (compromis) de référence. Ce qui bouge fortement entre été et automne c'est la position du couple Température-Nitrates.

En automne, l'importance des nitrates est plus grande et la station 5 pratiquement restaurée est proche de la station 6 (référence non polluée sur la Bourne) et de la station 1 (référence non polluée en amont du méandret).

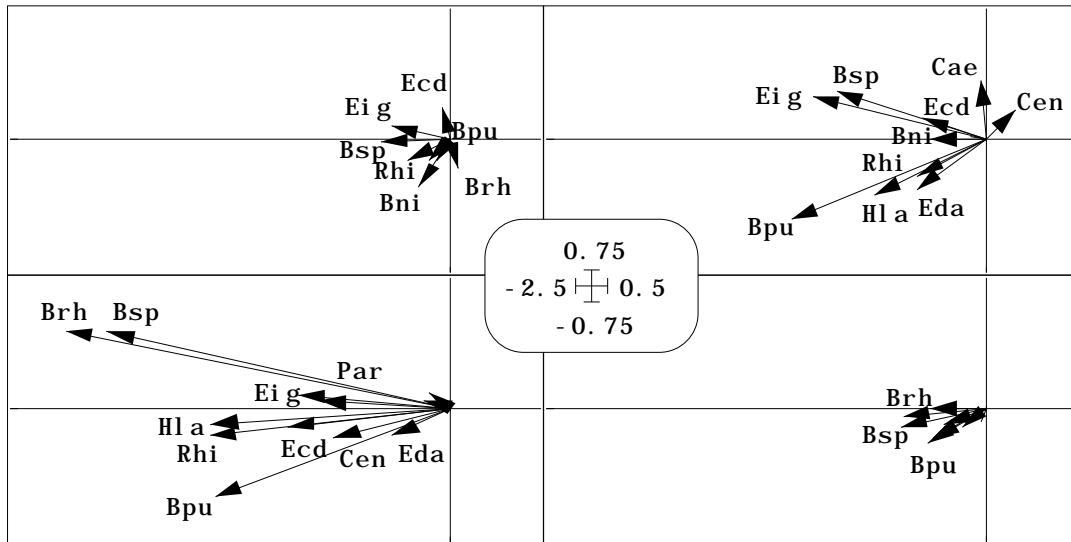
En été, la pollution en cours (tourisme) est plus forte et la restauration (en particulier dans la station 3) est faible mais la faune semble ne pas totalement l'enregistrer : les covariances sont plus faibles. En fait, on a peut-être là une illustration de la nature intégratrice des descripteurs faunistiques.

Labels			
XY coordinates file	<input type="button" value="..."/>	A.naTCO	52 2
X-axis column number (default = 1)	<input type="button" value="..."/>		
Y-axis column number (default = 2)	<input type="button" value="..."/>		
Label file (or # for item numbers)	<input type="button" value="..."/>	A!Tcc.label_1	

Row selection method: File
 Keyboard

Row selection file (.cat): A!Tcc.cat

Selection col. number:



Le dernier élément de comparaison, ici d'importance moindre à cause du nombre faible de points, est formé par la projection des vecteurs principaux des analyses séparées. A chaque date l'analyse de co-inertie définit les axes de co-inertie dans chaque espace, axes qui peuvent être projetés sur les éléments propres de l'opérateur compromis. Ce qui est ici étonnant, c'est qu'on peut projeter les axes faune d'une part et les axes milieu de l'autre sans pouvoir prévoir si la variabilité des représentations est la même dans les deux cas.

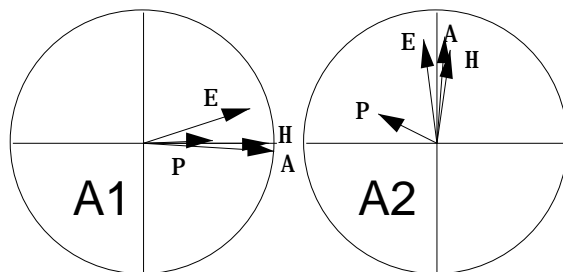
Labels	
XY coordinates file	<input type="text" value="А.хаТАне"/> 16 2
X-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="A!T4a.label"/>

Row selection method:	<input checked="" type="radio"/> File
	<input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="A!T4a.cat"/>
Selection col. number:	<input type="text" value="2"/>

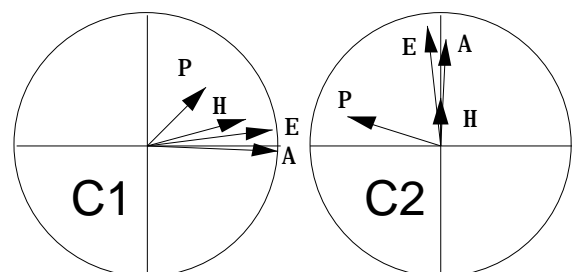
Labels	
XY coordinates file	<input type="text" value="А.хаТComp"/> 16 2
X-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="A!T4a.label"/>

Row selection method:	<input checked="" type="radio"/> File
	<input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="A!T4a.cat"/>
Selection col. number:	<input type="text" value="2"/>

Axes

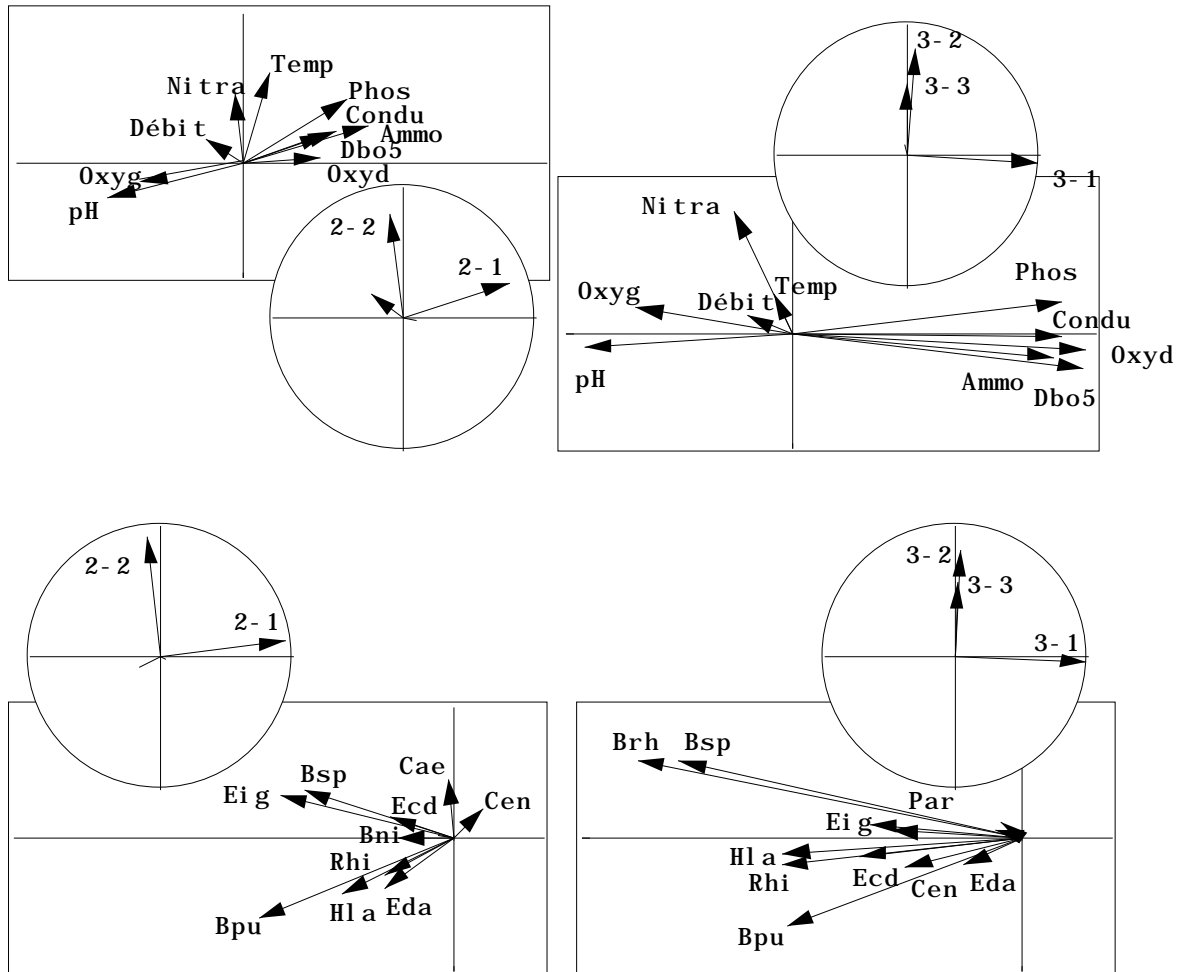


Composantes



Traditionnellement les axes d'une analyse sont dans le même espace que les lignes du tableau et les composantes dans le même espace que les colonnes. Dans les tableaux croisés de co-inertie, les lignes sont les colonnes du second tableau et sont dans l'espace des lignes du premier. Les axes sont donc les axes de co-inertie des tableaux faunistiques et les composantes sont les axes de co-inertie des tableaux mésologiques.

On peut les associer également par tableaux :



Comme on le sait déjà, l'analyse de l'automne a imposé son point de vue, mais curieusement la stabilité est plus grande pour les tableaux faunistiques, ce qui correspond aussi à la stabilité des descripteurs faunistiques intégrateur du temps qui sont de meilleurs descripteurs que les mesures instantanées du milieu aquatique. Remarquer que les axes et les composantes s'associent aux nuages de variables et non aux nuages de stations. Ce premier exemple montre que la notion de stabilité de la co-structure faune-milieu est accessible sans difficulté particulière.

4 — Typologie stable et environnement fluctuant

On utilise STATICO quand on veut examiner la reproductibilité d'une co-structure, comme on utilise STATIS quand on veut examiner la reproductibilité d'une structure. Ceci précise le rôle de ces méthodes. On en donne ici une autre illustration très simple.

Utiliser la carte Esolus de la pile ADE•Data. Pour toute information sur la signification de l'expérience, le matériel, la description des sites, on se reportera à la thèse de M.C. Forestier¹⁵.

0.7	0.3	0	0	0	0	0	10	0	0	2	0.7	0.3	0	0	0	0	0	3	0	0	1
0	0.4	0.2	0	0	0.4	0	28	14	1	2	0.1	0.4	0.1	0.1	0	0.4	0	20	2	2	2
0	0.5	0.4	0.1	0	0	0	30	22	0	2	0.1	0.5	0.3	0.2	0	0	0	14	11	1	1
0	0.5	0.4	0.2	0	0	0	30	22	0	2	0.1	0.5	0.3	0.2	0	0	0	14	11	1	1
0	0.2	0.3	0.4	0.2	0	0	20	27	0	1	0	0.4	0.2	0.3	0.2	0	0	13	13	1	1
0	0.2	0.2	0.2	0.2	0	0.3	25	47	1	1	0	0.1	0.2	0.2	0.3	0	0.3	15	30	2	0
0	0.1	0.1	0.1	0	0	0.8	20	50	2	0	0	0.1	0.1	0	0	0	0.9	11	82	1	0
0	0.7	0.4	0	0	0	0	32	14	0	2	0.4	0.3	0.3	0	0	0	0	4	0	0	1
0	0.3	0.2	0	0	0.6	0	55	31	0	2	0.2	0.3	0.1	0	0	0.4	0	16	0	1	1
0	0.4	0.4	0.2	0	0	0	50	60	0	2	0.1	0.6	0.2	0.1	0	0	0	14	3	1	2
0	0.4	0.4	0.2	0	0	0	50	60	0	2	0.1	0.6	0.2	0.2	0	0	0	8	8	1	2
0	0.2	0.2	0.6	0.2	0	0	45	74	0	2	0	0.4	0.2	0.3	0.2	0	0	10	5	1	1
0	0.1	0.1	0	0.3	0	0.6	55	111	0	1	0	0.2	0.2	0.2	0.2	0	0.3	8	38	2	1
0	0.1	0	0	0	0	0.9	45	110	0	1	0	0.1	0.1	0.1	0	0	0.8	5	54	1	0
0	0.4	0.5	0.1	0	0	0	19	14	0	0	0.1	0.6	0.4	0	0	0	0	8	0	0	2
0	0.3	0.3	0	0	0.4	0	39	21	0	1	0.2	0.3	0.1	0	0	0.4	0	10	0	1	2
0	0.4	0.4	0.3	0	0	0	39	36	0	1	0.3	0.5	0.2	0.1	0	0	0	10	0	0	2
0	0.3	0.4	0.4	0	0	0	42	30	0	1	0.2	0.5	0.2	0.1	0	0	0	8	0	0	2
0	0.2	0.2	0.6	0.2	0	0	38	46	0	1	0.1	0.4	0.2	0.3	0.2	0	0	4	3	1	1
0	0	0	0	0	0	0	1	50	57	0	0	0.2	0.2	0.2	0.2	0	0.3	5	19	2	1
0	0.1	0	0	0	0	0.9	38	70	0	0	0	0.1	0.1	0	0	0	0.8	5	14	1	1
0.1	0.6	0.4	0	0	0	0	20	10	0	1	0	0.7	0.4	0	0	0	0	10	0	0	2
0	0.4	0.2	0	0	0.4	0	40	21	0	1	0	0.3	0.2	0.1	0	0.5	0	22	9	0	1
0	0.4	0.4	0.2	0	0	0	42	22	0	1	0	0.5	0.3	0.2	0	0	0	29	27	0	1
0	0.4	0.4	0.2	0	0	0	47	22	0	1	0	0.4	0.4	0.3	0	0	0	34	25	0	1
0	0.2	0.3	0.4	0.2	0	0	35	29	0	1	0	0.2	0.3	0.4	0.2	0	0	25	20	0	1
0	0.1	0.1	0.1	0.3	0	0.5	51	56	0	0	0	0.1	0.1	0.2	0.3	0	0.3	25	43	1	1
0	0.1	0.1	0	0	0	0.8	34	62	0	0	0	0.1	0.1	0.1	0	0	0.8	22	49	1	1
0.7	0.3	0	0.1	0	0	0	5	0	0	1	0	0.3	0.7	0	0	0	0	40	17	0	0
0	0.4	0.2	0	0	0.4	0	25	19	1	1	0	0.3	0.2	0.1	0	0.5	0	55	23	0	2
0	0.4	0.4	0.2	0	0	0	15	22	0	1	0	0.3	0.4	0.4	0	0	0	70	50	0	1
0	0.4	0.4	0.2	0	0	0	15	22	0	1	0	0.3	0.4	0.4	0	0	0	60	52	0	1
0	0.3	0.3	0.3	0.2	0	0	10	27	1	1	0	0.2	0.2	0.5	0.2	0	0	50	56	0	1
0	0.1	0.2	0.2	0.3	0	0.3	10	67	2	0	0	0.1	0.1	0	0.3	0	0.5	58	69	0	1
0	0.1	0.1	0.1	0	0	0.8	15	82	1	0	0	0.1	0	0.1	0	0	0.9	50	94	0	0

Tableau 1 - 10 tableaux mésologiques (7 stations - 11 variables).

ADE-4•Data			
Esolus 70-7 70-11 et 10 dates / Marie-Catherine FORESTIER			
1	70-7	70-11	.7
2	DES (dates-es	DEV	.3
0	10 blocs cons	(dates-espace en	0
0		10 blocs cons	7
0	27/10/81	Limons	7
0	16/12/81	Sables	7
0	1/02/82	Graviers	7
13	22/03/82	Galets	7
15			10
27	10, 12, 613		7
34	8, 10, 492		7
33	8, 10, 2, 564		7
57	6, 2, 2, 734		7
10	14, 5, 12, 878		.4
18	24, 6, 30, 4, 948		.2
4	18, 19, 879		0
081		Esolus [101/177]	

Avec le champ de droite faire un fichier DEVar.import, l'importer par TextToBin : Import file qui trouve 770 nombres (10x7x11) dans Mil, le redimensionner par TextToBin : Changing row & col numbers :

Import file		
Text input file		DEVar.import
Binary output file		Mil

Changing row & col numbers		
Input file		Mil 770 1
New row number		70
New column number		11

Renvoyer le fichier Mil dans Excel par et l'éditer. On obtient le tableau 1. Les 7 premières variables de milieu (tableau 1) décrivent la granulométrie (1-limons, 2-sables, 3-graviers, 4-galets, 5-pierres, 6-blocs, 7-dalle, en pourcentage de recouvrement). La variable 8 est la hauteur d'eau (cm), la variable 9 est la vitesse du courant (cm/s), la variable 10 décrit le périphyton et la variable 11 concerne les débris organiques (0-absent, 1-présent, 2-abondant).

Créer le fichier Bloc (10 fois la valeur 7) et le fichier Label_Var. Les variables sont d'unités différentes et la normalisation s'impose, tout comme le centrage par tableau. On laisse faire l'hétérogénéité variable du milieu ("menu du jour" proposé à la population étudiée par ailleurs). Initier la structure des données :

InitKtab		
Matrix input file		Mil 70 11
Row indicator (Default = 1 class)		Bloc 10 1

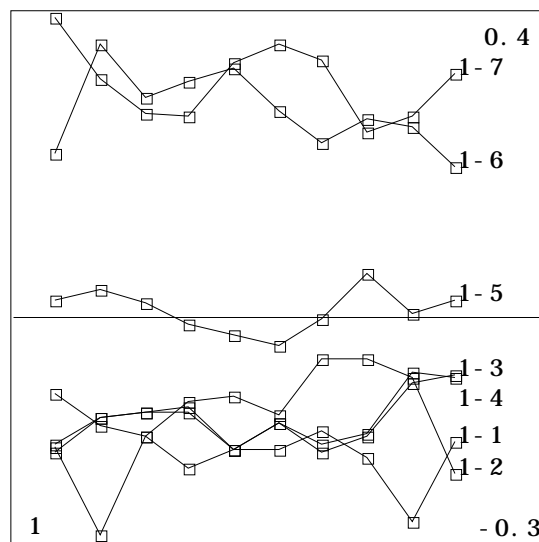
Normaliser :

NormalizeKtab		
---.ktp input file		Mil.ktp
Matrix input file		Mil 70 11
Row weighting (default 1/nk)		
Column weighting (default 1/pj)		
Row-Col option (no default)		3
Option: output file name		MBour

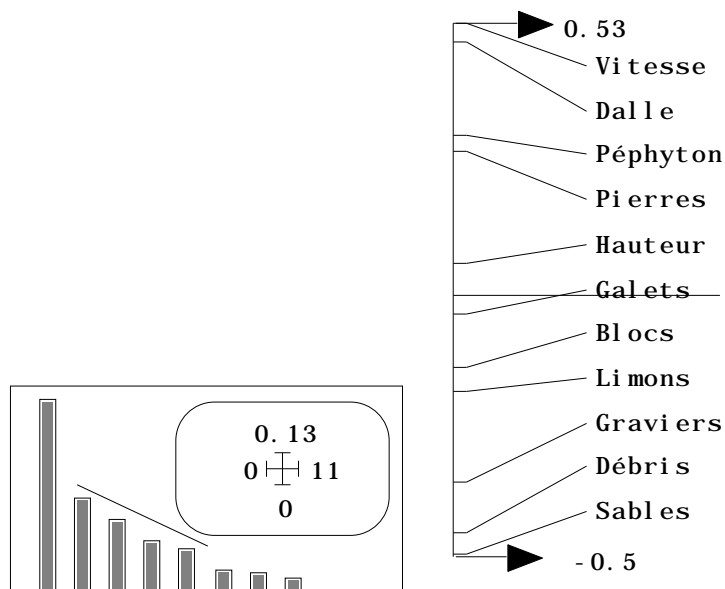
Exécuter son analyse (STATIS : Operator averaging) :

Operator averaging		
---.kta input file		MBour.kta 70 11
l = Option COUV (default = RV)		1

Lines	
X file (default = 1, 2, 3, ..., n)	Mil!TLI
X file column number (default = 1)	
Y file (no default)	MBour.oaTLI
Categories file (.cat)	Mil!TLI.cat
Categories file column number	3
Variable label file (or #)	#
Draw curves (1=yes, 2=no)	1
Draw points (1=yes, 2=no)	
Categories label file (or #)	Mil!TLI.label



Globalement à chaque date se reproduit la même typologie de stations à 3 groupes (1 à 4, 5 et 6-7). Cette typologie commune est essentiellement unidimensionnelle (fichier MBour.oa+vp, ci-dessous) :



Les variables qui jouent un rôle sont vues par le fichier MBour.oa+co (ci-dessus). On repère un gradient de granulométrie. On pourra comparer avec l'analyse intra-dates qui avait été faite dans les annexes de la thèse citée (fiches de la version 3.6) et reconnaître très clairement que STATIS est bien la forme achevée des analyses intra-classes. Ceci nous conduira à penser que STATICO est la forme achevée de l'analyse de co-inertie intra-classes.

Typological value indices
 Rows = row number of each table
 Weights = Weights of operators in the compromise
 NS norm2 = Squared norm (Hilbert-Schmidt) of operators
 Cos2 = Squared cosinus between operator and approximated compromise

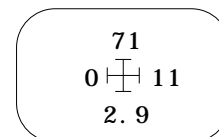
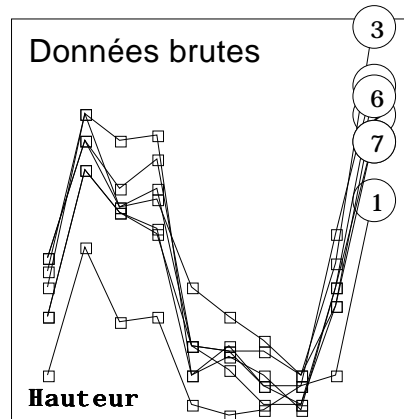
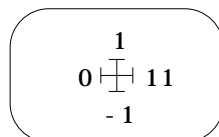
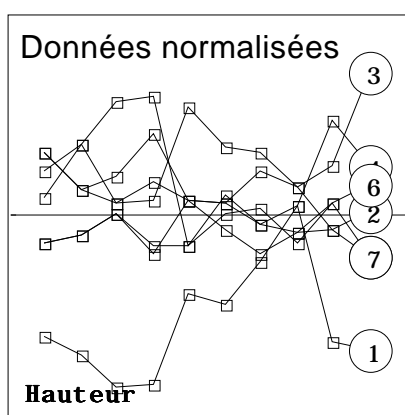
Number	Rows	Weights	NS norm2	Cos2
1	7	3.802e-01	4.862e-03	0.622
2	7	3.818e-01	4.662e-03	0.638
3	7	2.841e-01	3.491e-03	0.435
4	7	2.657e-01	2.317e-03	0.613

5	7	3.909e-01	4.507e-03	0.722
6	7	3.764e-01	4.226e-03	0.718
7	7	2.705e-01	2.226e-03	0.712
8	7	2.264e-01	2.164e-03	0.476
9	7	2.397e-01	2.025e-03	0.544
10	7	2.865e-01	3.254e-03	0.426

Les indicateurs numériques soulignent une bonne homogénéité de la structure mésologique. Cela ne signifie pas que le système est stable :

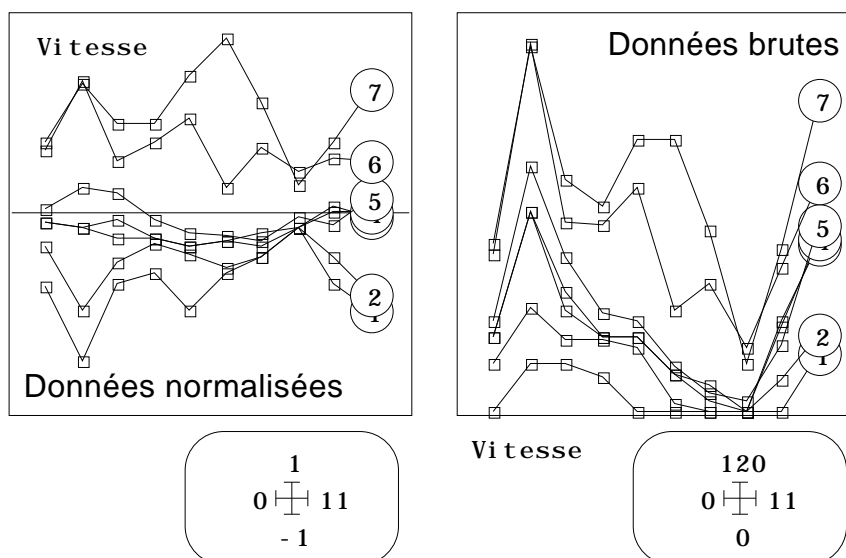
Lines	
# file (default = 1, 2, 3, ..., n)	Mil!TLI 70 3
# file column number (default = 1)	1
Y file (no default)	MBour.ktta 70 11
Cumulated data (1=yes, 2=no)	
Variable label file (or #)	Label_Var
Draw curves (1=yes, 2=no)	
Draw points (1=yes, 2=no)	
Row label file (or #)	
Number of curves by window	7

Row & col. selection	
Col. selection:	8
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	Mil!TLI.cat
Selection col. number:	3



Noter ainsi le rôle considérable du simple changement de variables en œuvre dans la normalisation de Bouroche. En milieu aquatique, il y a une forte corrélation entre l'état moyen du système (associé au débit de la rivière) et son hétérogénéité (variabilité entre les stations).

L'objectif expérimental lié à la pratique statistique de STATIS est très précis : il s'agit de rechercher la part stable de la typologie des stations. Cette part stable peut émerger d'une système fortement évolutif. Le changement de variable (positionner les points autour de la moyenne instantanée) et le principe de la méthode (caler plusieurs ACP) élimine autant que peut se faire les images de la variabilité de l'ensemble des stations (variabilité de valeur et variabilité de structure).



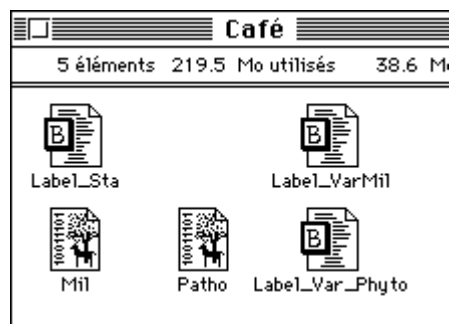
C'est évidemment dans sa fonction pédagogique que l'analyse de ce jeu de données par ces moyens se trouve justifiée.

5 — Moyennes et co-structures

Les méthodes STATIS, dont STATICO, permettent de discuter de moyennes de structures et de structure de moyennes. On utilise ici un extrait des données du laboratoire de phytopathologie de Nouméa (ORSTOM) représentatif de la problématique environnement-épidémiologie pour discuter de cet aspect de l'analyse des couples de cubes de données. Considérons n stations réparties dans un territoire donné. Plusieurs fois dans l'année, la visite de chaque station conduit à un enregistrement multivarié de l'état sanitaire de la culture d'une part, à un enregistrement multivarié de l'environnement de la culture d'autre part. Un élément particulier caractérise cette situation expérimentale : une partie des variables environnementales (météorologie, hydrologie) varie d'une date à l'autre, une partie de ces variables (qualités du sol, géomorphologie, altitude) est stable. Il est alors difficile, vue cette différence de nature de variables, de comparer les effets respectifs des paramètres environnementaux sur l'état phytosanitaire de la culture.

La tactique la plus accessible consiste à moyenner sur une année les valeurs des paramètres de milieu et à coupler avec les moyennes annuelles des variables épidémiologiques (ce qui élimine la logique propre de l'épidémie) pour faire un bilan en terme de co-structure entre moyenne. C'est la voie explorée dans ¹⁶ et commentée dans ¹⁷. On perd ainsi une part de l'information qu'on pense exister dans la dynamique temporelle des infections des cultures et des causes environnementales qu'on leur suppose. Il n'est pas question de résoudre ici cette question mais d'explorer ce qu'on obtient en pensant que le couplage entre les deux ensembles de variables se fait à chaque date et définit une co-structure moyenne plutôt qu'une co-structure entre moyennes. C'est ce qu'on commence à comprendre dans la comparaison entre STATIS et AFMULT récemment proposée dans ¹⁸. Un exemple est proposé dans la carte Café de la pile ADE•Data :

ADE-4•Data											
Milieu (120-8) Pathologie (120-6) F											
8 dates/15 stations											
76	23.6	44	1	1	5.9	ATEOU	1.03	1.07	1	8	0
299	20.2	63	3	1	7.5	BAN1	1	1.04	1.01	6	0
247	23.4	50	1	1	6.7	CAN1	1.07	1.02	1	8	0
375	19.9	25	1	1	6.7	EMA2	1.18	1	1	17	0
						ILP4	1	1.11	1.05	13	
						ILP6	1	1.01	1.05	0	
pl						KONE					
tmin						KOUAS					
fert											
struct											
cap											
pH											
8 dates						NSSROU					
15 stations						NSSANT					
Tableaux stations-vari						NSScerco					
Milieu 8 variables						%CFneoMAL					
Pathologie 6 variables						%CFTMal					
						%CFTSain					
152						Café [164/177]					



Cliquer sur le bouton F pour obtenir le dossier de travail correspondant. Ajouter un fichier Bloc contenant le nombre de stations visitées par dates (ici 8 fois la valeur 15). Tout ce qui suit supporte un nombre variable de stations visitées par dates, en particulier les données manquantes d'un plan d'observations prévu complet. Rien n'oblige à réduire les dates à un cycle annuel, mais ici on possède les "rondes" associées à un cycle de végétation.

Les variables de milieu sont :

- pl** pluviométrie cumulée (on a chaque jour la pluviométrie par 3h, on somme par jour puis par mois)
- tmin** température minimale (on fait la moyenne des T°C minimales récoltées entre deux rondes)
- fert** fertilité du sol (indice de synthèse des analyses de sol prenant les valeurs 1 - mauvais, 2 - passable, 3 - bon)
- struct** norme granulométrique (indice de synthèse prenant les valeurs 1 à 3)
- cap** capacité en eau : indice de rétention du sol donné par les analyses (1 15% ; 2 de 15 à 20 % ; 3 20%)
- pH** moyenne de nombreuses mesures
- alt** altitude du site en mètres
- ombr** degré d'ombrage (indice variant de 1 - couverture nulle ou faible à 4 - ombrage fort et homogène)

Les variables phytosanitaires sont :

- NSSROU** note sanitaire de la rouille, maladie la plus grave
- NSSANT** note sanitaire de l'antracnose
- NSScerco** note sanitaire de la cercosporiose
- %CFneoMAL** pourcentage cumulé de feuilles neo malades
- %CFTMal** pourcentage cumulé de feuilles disparues (indice de défoliation pathologique)
- %CFTSain** pourcentage cumulé de feuilles tombées saines (indice de défoliation physiologique par les éléments physiques)

5.1 — Variabilité environnementale

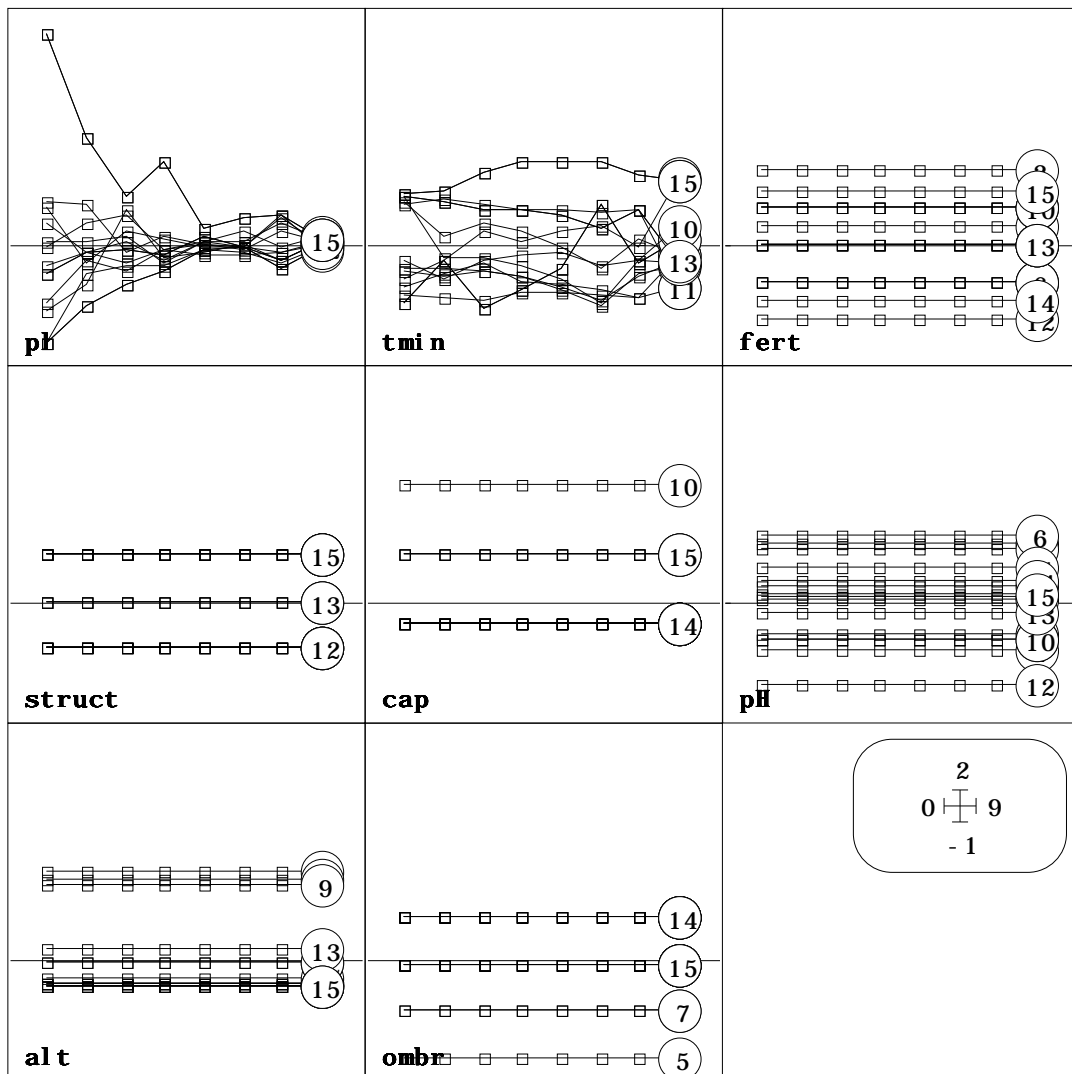
Implanter la structure du K-tableaux :

InitKTab			
Matrix input file	<input type="button" value="..."/>	Mil	120 8
Row indicator (Default = 1 class)	<input type="button" value="..."/>	Bloc	8 1
Col indicator (Default = 1 class)	<input type="button" value="..."/>		
Option: output file name	<input type="button" value="..."/>		

Normaliser les données après centrage par dates :

NormalizeKTab			
---.ktp input file	<input type="button" value="..."/>	Mil!.ktp	
Matrix input file	<input type="button" value="..."/>	Mil	120 8
Row weighting (default 1/nk)	<input type="button" value="..."/>		
Column weighting (default 1/pj)	<input type="button" value="..."/>		
Row-Col option (no default)	<input type="button" value="..."/>	3	
Option: output file name	<input type="button" value="..."/>		

Ceci permet de repérer l'évolution spatio-temporelle des variables :



Pour obtenir cette figure, dans Curves : Lines, utiliser le numéro de tableau comme abscisse, la variable normalisée comme ordonnée, la superposition des 15 courbes par station :

Lines	
X file (default = 1, 2, 3, ..., n)	Mil!TLI 120 3
X file column number (default = 1)	1
Y file (no default)	Mil.ktta 120 8
Cumulated data (1=yes, 2=no)	
Variable label file (or #)	Label_UarMil
Draw curves (1=yes, 2=no)	
Draw points (1=yes, 2=no)	
Row label file (or #)	
Number of curves by window	15

Utiliser le numéro de station par tableau pour multifenêtrer et caler les bornes :

Row & col. selection	
Col. selection:	
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	Mil!TLI.cat
Selection col. number:	3

Min/Max	
Min. abscissa:	0 <input checked="" type="checkbox"/> Horiz. graphs: 3 <input type="checkbox"/>
Max. abscissa:	9 <input checked="" type="checkbox"/> Vert. graphs: 3 <input type="checkbox"/>
Min. ordinate:	-1 <input type="checkbox"/> Nb. grad. H: 1
Max. ordinate:	2 <input type="checkbox"/> Nb. grad. V: 1
Window height:	400 G factor:
Window width:	400

On retrouve six des variables environnementales sans variation temporelle, ce qui suppose une excellente reproductibilité de la typologie inter-stationnelle, reproductibilité purement artefactuelle :

Operator averaging	
---.ktta input file	Mil.ktta 120 8
! = Option COUV (default = RV)	1

Le résultat ne manque pas de sel. Les corrélations entre tableaux sont excellentes, le contraire eut été fâcheux :

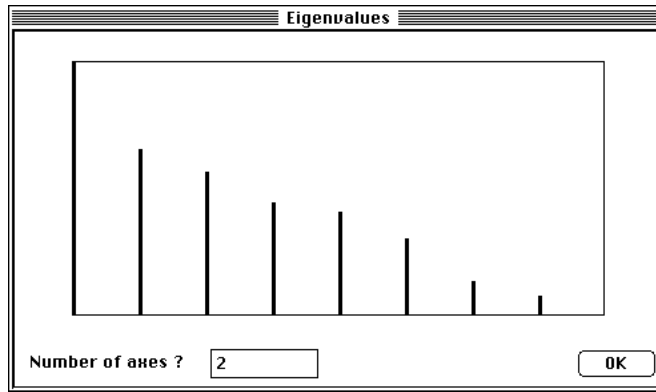
```
----- Correlation matrix -----
[ 1] 1000
[ 2] 819 1000
[ 3] 635 923 1000
[ 4] 668 945 975 1000
[ 5] 505 879 969 955 1000
[ 6] 482 860 921 922 969 1000
[ 7] 555 898 942 951 978 940 1000
[ 8] 493 857 946 931 977 923 971 1000
-----
```

La reproduction de la structure moyenne est cependant médiocre :

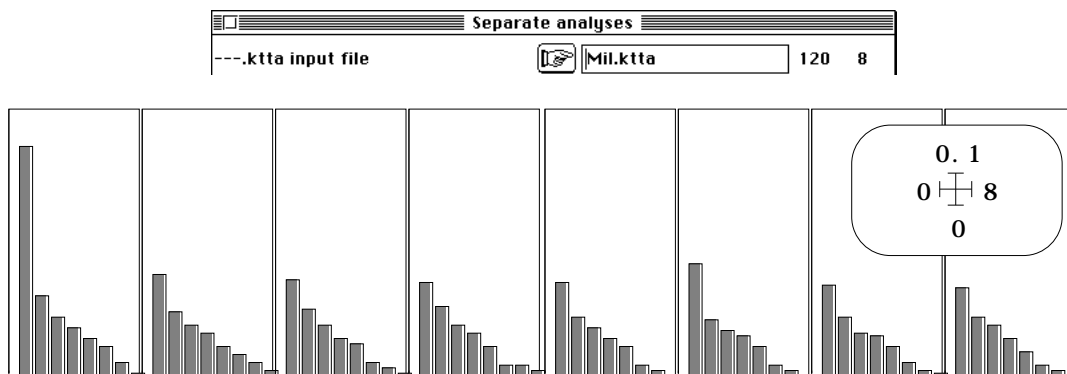
Typological value indices

Number	Rows	Weights	NS norm2	Cos2
1	15	5.291e-01	9.633e-03	0.397
2	15	3.598e-01	3.010e-03	0.666
3	15	3.311e-01	2.728e-03	0.593
4	15	3.396e-01	2.796e-03	0.659
5	15	3.046e-01	2.524e-03	0.530
6	15	3.199e-01	2.962e-03	0.510
7	15	3.004e-01	2.390e-03	0.540
8	15	2.834e-01	2.278e-03	0.492

Cela vient du fait que la structure moyenne, vue par les valeurs propres du compromis, brille par son absence :



Les tableaux à chaque ronde se ressemblent fortement mais la corrélation entre variables environnementales, dans le tableau compromis, est très faible. Cela prouve que les expérimentateurs ont débarrassé les mesures des facteurs environnementaux du maximum de redondance et que chaque variable apporte sa propre information, on peut dire sa propre typologie de station, sans restituer la même information que ses partenaires dans le tableau. Ceci est tout-à-fait remarquable et place le couple environnement-pathologie avec des variables explicatives sans redondance donc avec une capacité de modélisation optimale. On confirme par les analyses séparées (KTA : Separate analyses) :



Si co-structure il y a, la redondance des variables de milieu n'y jouera aucun rôle, puisqu'elle n'existe pratiquement pas. Soulignons à quel point une méthode d'analyse n'a pas de valeurs intrinsèques mais qu'une valeur relative pour chaque jeu de données.

5.2 — Variabilité épidémiologique

Introduire la structure du *K*-tableau :

Utiliser le même mode de normalisation :

NormalizeKtab

---.ktp input file

Matrix input file 120 6

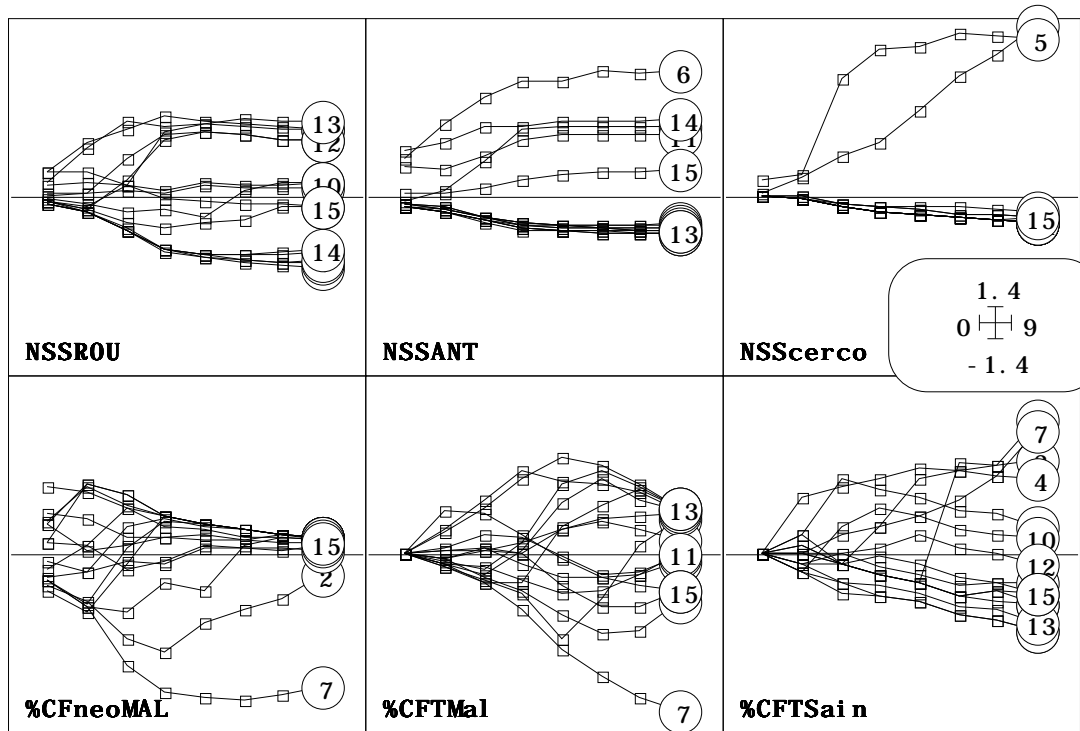
Row weighting (default 1/nk)

Column weighting (default 1/pj)

Row-Col option (no default)

Option: output file name

Représenter les données :



L'ambiance est totalement différente. Utiliser STATIS :

Operator averaging

---.kta input file 120 6

l = Option COUV (default = RV)

On obtient un message d'erreur d'impossibilité de diagonaliser une matrice. Cela vient de la première date où deux des variables sont uniformément nulles qui engendre une matrice 6-6 avec deux lignes et deux colonnes de zéros. Ces cas de mise en défaut de la procédure de diagonalisation¹⁹ sont rarissimes. On peut supprimer la date 1 ou plus simplement introduire sur la première ligne du fichier Patho.kta deux valeurs non nulles +0.000001 et -0.000001 à la colonne 5 et 6. Le module ADEBin est particulièrement pratique dans ce cas :

File: Patho.kta

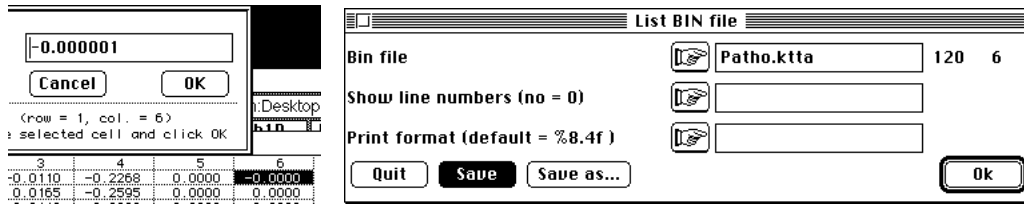
- Open [120-6]
- Print [120-6]
- List in windoid
- List in tabloid
- List with BBEdit

Enter value:

Cancel OK

Type the new value for cell (row = 1, col. = 5)
Enter the new value for the selected cell and click OK

1	2	3	4	5	6
-0.0544	-0.0638	-0.0110	-0.2268	0.0000	-
-0.0683	-0.0783	0.0165	-0.2595	0.0000	-
-0.0360	-0.0881	-0.0110	-0.2268	0.0000	-



Le résultat obtenu est totalement différent du précédent :

```

----- Correlation matrix -----
[ 1] 1000
[ 2] 928 1000
[ 3] 774 897 1000
[ 4] 610 782 953 1000
[ 5] 435 662 873 954 1000
[ 6] 346 573 812 914 981 1000
[ 7] 326 549 804 907 969 997 1000
[ 8] 268 477 761 851 900 952 970 1000
-----

```

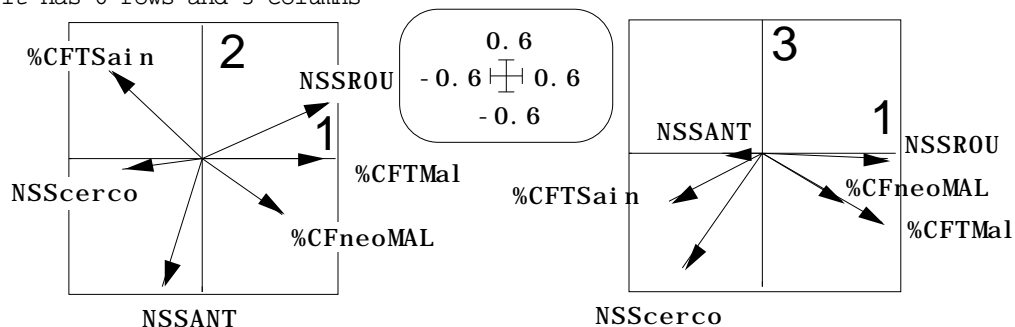
On obtient une véritable matrice d'autocorrélation temporelle entre tableaux avec un coefficient qui décroît avec l'intervalle de temps entre deux mesures.

Typological value indices

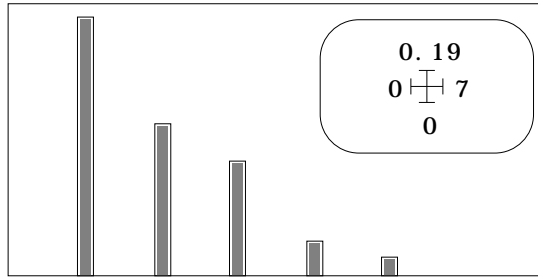
Number	Rows	Weights	NS norm2	Cos2
1	15	2.276e-02	1.672e-04	0.128
2	15	8.376e-02	9.440e-04	0.337
3	15	1.816e-01	2.286e-03	0.692
4	15	2.995e-01	5.141e-03	0.860
5	15	4.028e-01	8.541e-03	0.953
6	15	4.769e-01	1.165e-02	0.978
7	15	4.783e-01	1.167e-02	0.983
8	15	5.014e-01	1.361e-02	0.926

A l'inverse le poids du tableau et sa norme croissent tandis que son adéquation à la structure moyenne croit et atteint son optimum vers les dates 5 ou 6. Cela vient bien sur de la continuité des mesures de pathologie végétale et de l'autocorrélation des enregistrements. La typologie des variables est synthétisée par :

File Patho.oa+co contains column scores (analysis of the compromise)
It has 6 rows and 3 columns



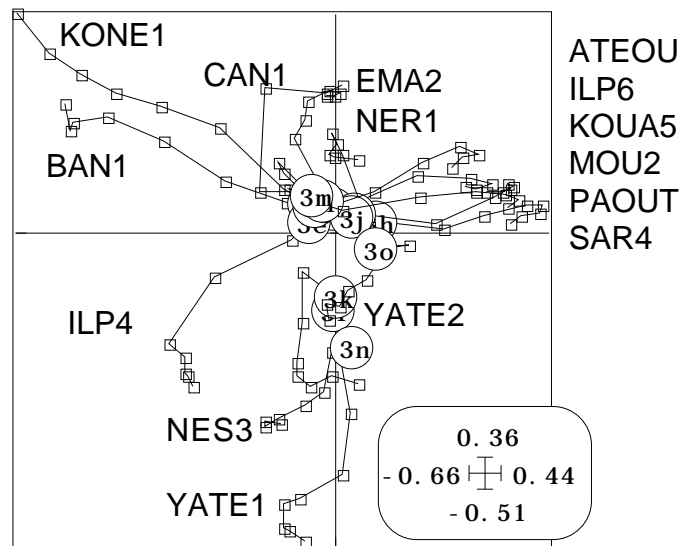
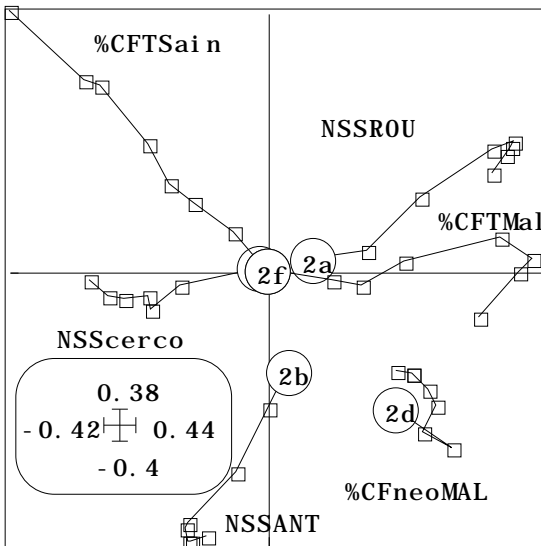
File Patho.oa+vp contains the eigenvalues of compromise analysis
It has 6 rows and 1 columns



Au total, on retrouve une indépendance assez grande entre les trois maladies avec une indication de liaison plus forte entre l'indice d'attaque de la rouille et les notes globales. La carte des variables est particulièrement expressive (à gauche) :

Trajectories	
HV coordinates file	<input type="text" value="Patho.aaTCtra"/> 48 3
H-axis column number (default = 1)	<input type="text"/>
V-axis column number (default = 2)	<input type="text"/>
Categories file (.cat)	<input type="text" value="Patho!Tcc.cat"/>

Row & col. selection	
Col. selection:	<input type="text" value="2"/>
Row selection method:	<input type="radio"/> File <input checked="" type="radio"/> Keyboard



Il en est de même de celle des stations (à droite) :

Trajectories	
HV coordinates file	<input type="text" value="Patho.aaTLI"/> 120 3
H-axis column number (default = 1)	<input type="text"/>
V-axis column number (default = 2)	<input type="text"/>
Categories file (.cat)	<input type="text" value="Patho!TLI.cat"/>

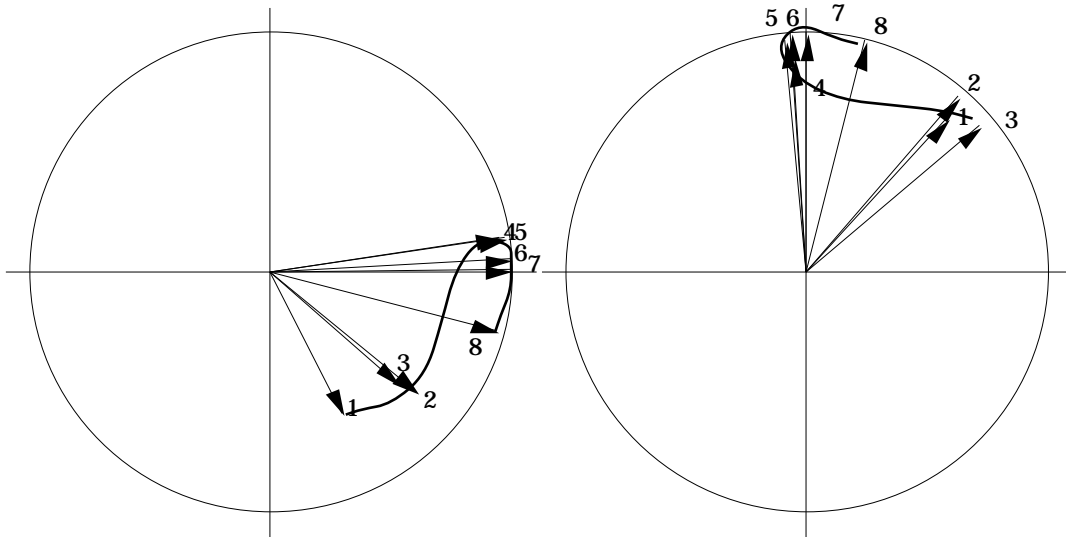
Row & col. selection	
Col. selection:	<input type="text" value="3"/>
Row selection method:	<input type="radio"/> File <input checked="" type="radio"/> Keyboard

La variable %CFneoMAL y apparaît comme une somme (rouille et antracnose), tandis que la variable %CFTMal est plus étroitement liée à la rouille. Dans les deux espaces les trajectoires divergentes indiquent l'augmentation de la variabilité inter-stationnelle avec le développement des maladies. La variance vectorielle augmente mais la corrélation inter-variables reste modeste.

La représentation des axes des analyses séparées confirme l'interprétation :

Labels		
HY coordinates file	<input type="text" value="Patho.aaT4a"/>	32 3
X-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="Patho!T4a.label"/>	
Draw vectors from origin (yes = 1)	<input type="text" value="1"/>	
Draw unit circle (yes = 1)	<input type="text" value="1"/>	

Row & col. selection	
Col. selection:	<input type="text"/>
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="Patho!T4a.ca"/>
Selection col. number:	<input type="text" value="2"/>



5.3 — Co-Variabilité

Associer les deux *K*-tableaux :

MatchTwoKtab		
First ---.ktt input file	<input type="text" value="Patho.ktta"/>	120 6
Second ---.ktt input file	<input type="text" value="Mil.ktta"/>	120 8
Output file name	<input type="text" value="R"/>	

Table averaging		
---.ktt input file	<input type="text" value="R.ktta"/>	64 6
I = Option RV (default = COUV)	<input type="text"/>	

La corrélation vectorielle du premier tableau se dégrade rapidement mais à partir de la date 3 l'autocorrélation vectorielle garde une valeur élevée.

```

----- Correlation matrix -----
[ 1] 1000
[ 2] 855 1000
[ 3] 376 630 1000
[ 4] 297 528 944 1000
[ 5] 210 423 815 934 1000
[ 6] 189 381 674 835 951 1000
[ 7] 220 390 752 883 960 940 1000
[ 8] 201 365 755 864 916 888 962 1000
-----

```

On a confirmation immédiate par les indices typologiques :

Typological value indices
Rows = row number of each table

Weights = Weights of tables in the compromise
 NS norm2 = Squared norm (Hilbert-Schmidt) of tables
 Cos2 = Squared cosinus between table and approximated compromise

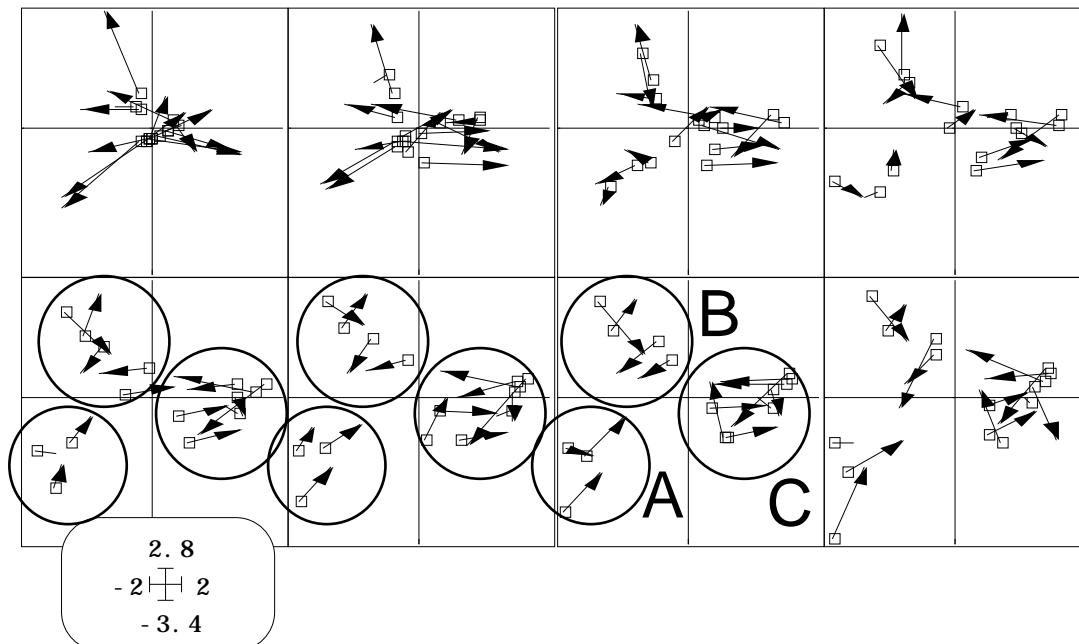
Number	Rows	Weights	NS norm2	Cos2
1	8	5.585e-02	5.670e-04	0.065
2	8	9.217e-02	4.914e-04	0.214
3	8	2.276e-01	9.416e-04	0.681
4	8	3.489e-01	1.716e-03	0.876
5	8	4.233e-01	2.288e-03	0.965
6	8	4.856e-01	3.198e-03	0.909
7	8	4.585e-01	2.703e-03	0.959
8	8	4.354e-01	2.562e-03	0.916

Les deux premiers tableaux ne participe pas à la définition du compromis, le troisième amorce la mise en place et de 5 à 8 la structure se reproduit à l'identique.

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+8.6828E-03	+0.7000	+0.7000	02	+2.6725E-03	+0.2155	+0.9155
03	+9.8458E-04	+0.0794	+0.9948	04	+4.8471E-05	+0.0039	+0.9988
05	+1.4106E-05	+0.0011	+0.9999	06	+1.3653E-06	+0.0001	+1.0000

On peut hésiter sur la signification de l'axe 3 de l'analyse du compromis. Nous l'ignorons.

File A.xaTXLl contains normed row scores (from table Patho.ktta)
 It has 120 rows and 3 columns
 It is to be used with Patho!TLl.label and Patho!TLl.cat files
 File A.xaTYLl contains normed row scores (from table Mil.ktta)
 It has 120 rows and 3 columns
 It is to be used with Mil!TLl.label and Mil!TLl.cat files



La double représentation des nuages de co-inertie donne une image particulièrement explicite de la mise en place de la co-structure. Aux dates 1 et 2, il n'y a aucune correspondance. La co-structure s'installe en 3 et 4. Elle est en place de 5 à 7 et se dégrade légèrement en 8.

Elle induit une partition à première vue en 3 groupes de stations. On examine en détail la disposition des stations :

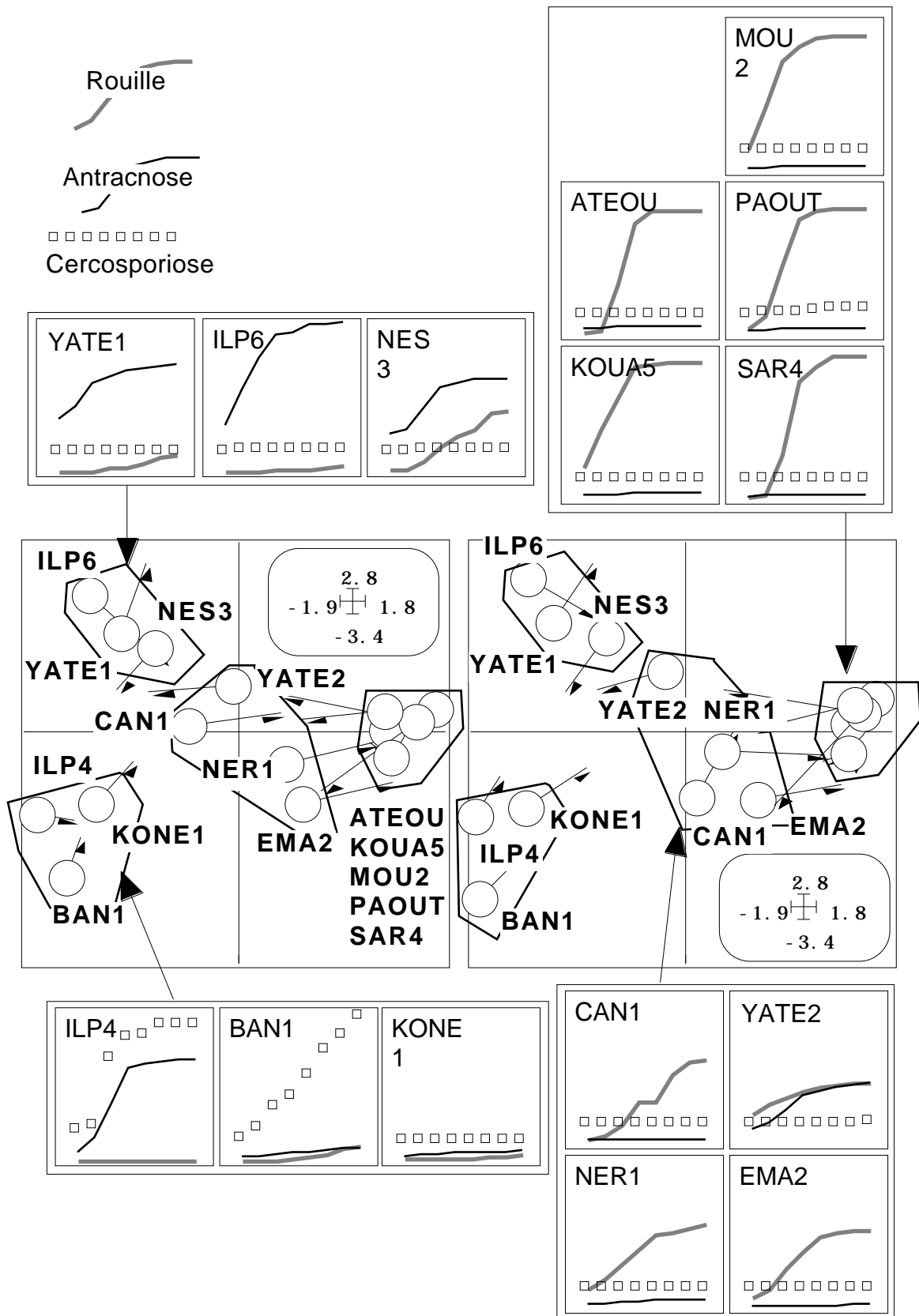


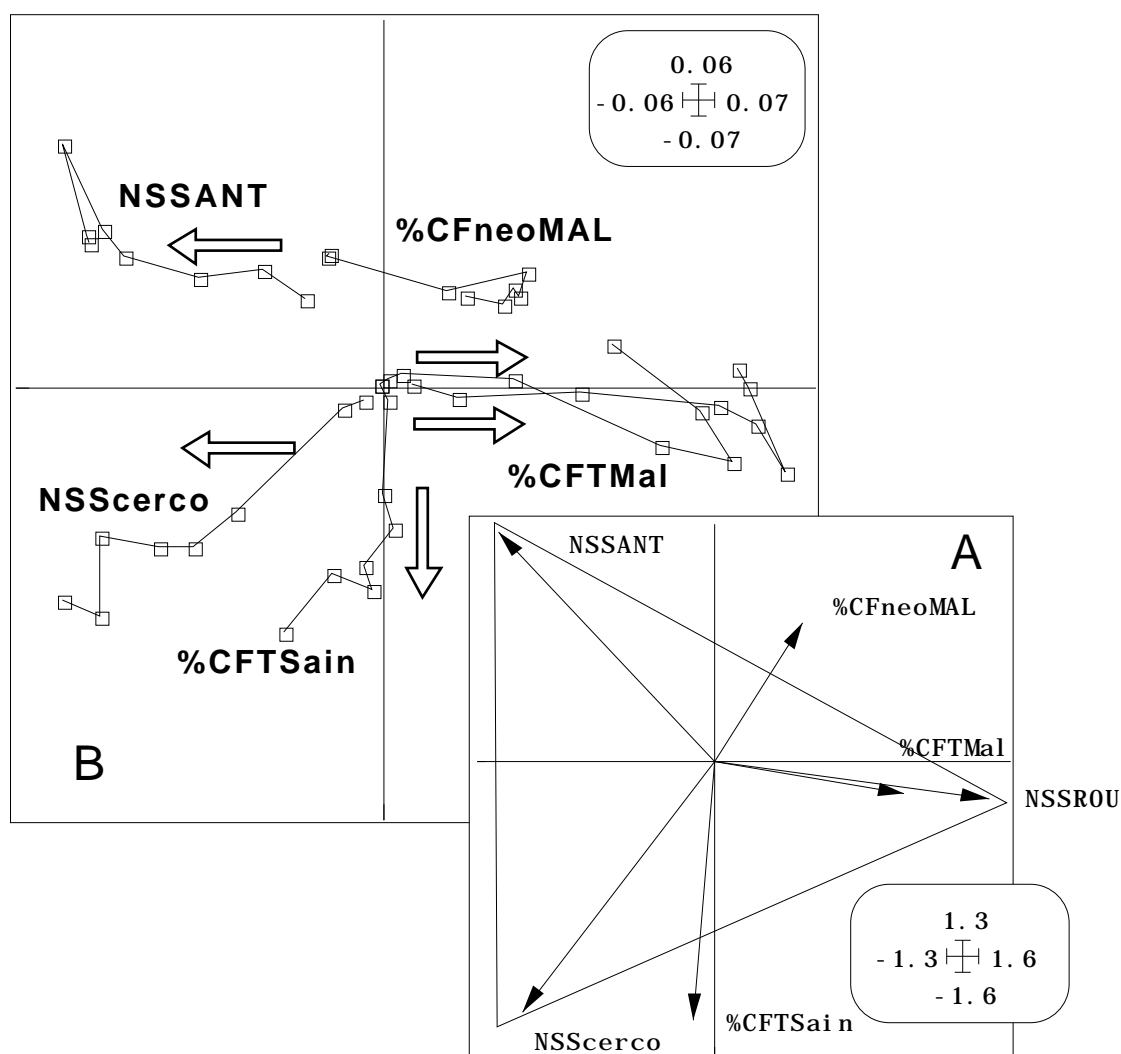
Figure 1 — Relecture de la partition des stations sur le graphe des données brutes.

Si on tient compte du fait que ces coordonnées sont normalisées (pour obtenir la superposition) mais que les inerties sur les deux axes sont très différentes (70% contre

22%), si on introduit la signification expérimentale des trois premières variables, il vaut mieux voir dans disposition des stations d'abord une ordination sur l'intensité de l'attaque de la rouille croissante avec la valeur du facteur 1. La séparation sur l'axe 2 ne portent que sur les 6 stations peu concernée par la rouille dans lesquelles on rencontre 4 situations (antracnose sans cercosporiose, cercosporiose sans antracnose, les deux à la fois et ni l'une ni l'autre). Il vaut mieux s'abstenir de chercher une explication à une partition de 6 objets en 4 classes avec plusieurs explicatives possibles. On ne tiendra compte que du facteur 1.

Les colonnes du *K*-tableau couplés sont les colonnes des **X**, soient les variables phytosanitaires. Les lignes sont les variables de milieu. Les scores de A.xa+co donne la partie A de la figure ci-dessous.

File A.xa+co contains column scores (Norm = 1 for column weights)
in the analysis of the compromise
It has 6 rows and 3 columns



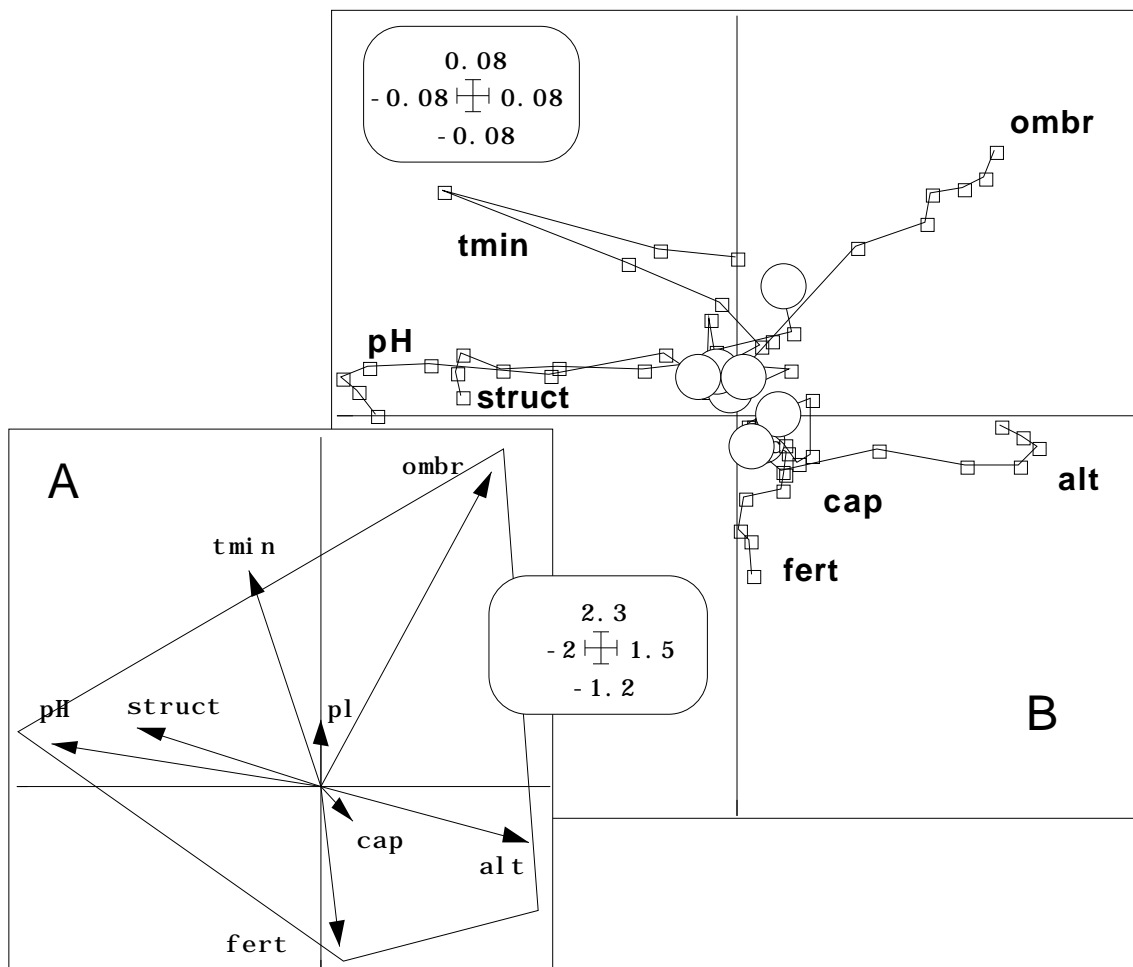
Les trois maladies ont des dynamiques largement indépendantes et les groupes correspondent globalement à des infections différentes, sans que la partition soit parfaite (figure 1 p. 34). En outre, il s'agit de poids (coordonnées normalisées).

File A.xaTCo contains columns scores issued from each table by
It has 48 rows and 3 columns
It is to be used with --Tc.label and --Tc.cat files

Les coordonnées standard (partie B de la figure ci-dessus) donne une image plus précise de la typologie des variables phytosanitaires. L'axe 1 est une ordination du niveau de rouille complétée par une intervention des deux autres maladies en l'absence de la première. La règle n'est pas absolue mais la liaison est franche. L'axe 2 est d'abord lié à la variable feuilles tombées saines, ce qui précise encore, qu'en terme de concordance pathologie-milieu, l'essentiel est sur le premier axe.

File A.xa+li contains row scores (Norm = 1 for row weights)
in the analysis of the compromise
It has 8 rows and 3 columns

File A.xaTl1 contains standard row scores with lambda norm
It has 64 rows and 3 columns
It is to be used with --Tl1.label and --Tl1.cat files



A.xa+li donne la partie A et A.xaTl1 donne la partie B dans une logique totalement équivalente à la figure précédente. Les covariances inversées entre les deux groupes est l'indicateur des facteurs limitant les infections. L'ombrage et l'altitude favorise la rouille, le pH et la structure du sol la limite. On pourrait penser, que si les variables de milieu sont stables, les coordonnées devraient l'être. C'est exact pour les coordonnées des points de mesure (station-date) mais les covariances entre variables de milieu et notes sanitaires augmentent avec le développement de la maladie et induise une nouvelle image de la croissance au cours du temps de la liaison entre les deux groupes de variables.

Utilisée sur plusieurs cycles successifs, la méthode donnera sans doute des renseignements pertinents sur le rythme d'expression de cette liaison.

5.4 — Co-structure des moyennes

Il est intéressant de comparer l'approche STATICO avec l'analyse de co-inertie inter-classes. Reprendre la procédure à partir des fichiers de données. Normaliser les deux tableaux (Bin->Bin : Centring) :

Centring	
Input file	Mil 120 8
Option: file for row weighting	
Option for H matrix (no default)	3
Output file	MN

Centring	
Input file	Patho 120 6
Option: file for row weighting	
Option for H matrix (no default)	3
Output file	PN

Créer un descriptif du plan d'observation sous forme de deux variables qualitatives (TextToBin : Create2Categ) :

Create2Categ	
Output file name	Plan
Variable 1	8
Variable 2	15

Moyenner par dates (FilesUtil : CateRowSum-Mean) :

CateRowSum-Mean	
Input file	MN 120 8
.cat file	Plan.cat
Column number for selection	2
Output file	MN_mean
Option: sum (1) or mean (2)	2

CateRowSum-Mean	
Input file	PN 120 6
.cat file	Plan.cat
Column number for selection	2
Output file	PN_mean
Option: sum (1) or mean (2)	2

Faire les deux ACP de tableaux de moyennes (PCA : Covariance matrix PCA). Coupler les deux analyses (CoInertia : Matching two statistical triplets) :

Matching two statistical triplets	
First data input file	PN_mean.cpta 15 6
Second data input file	MN_mean.cpta 15 8
Output file name	B

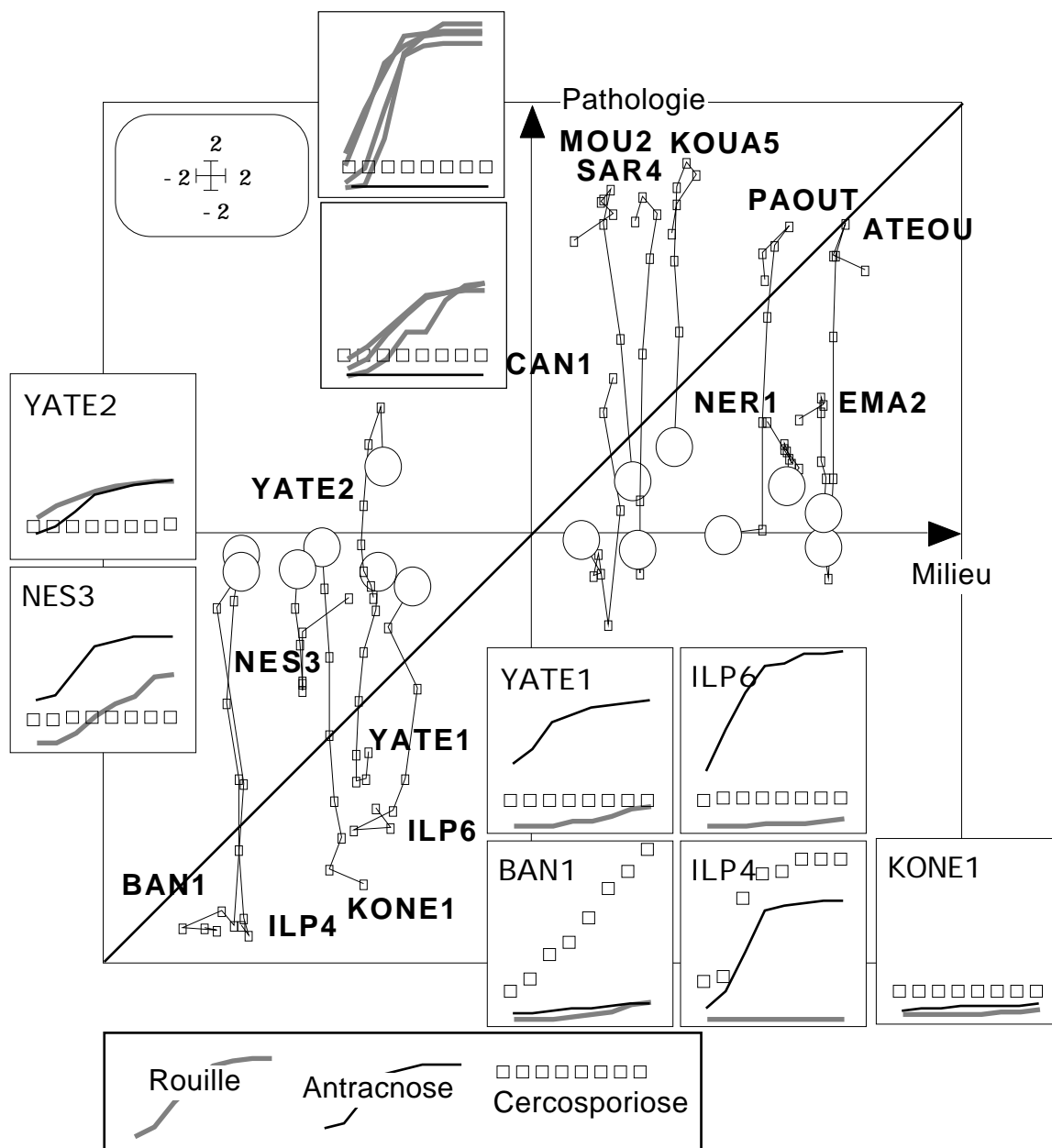
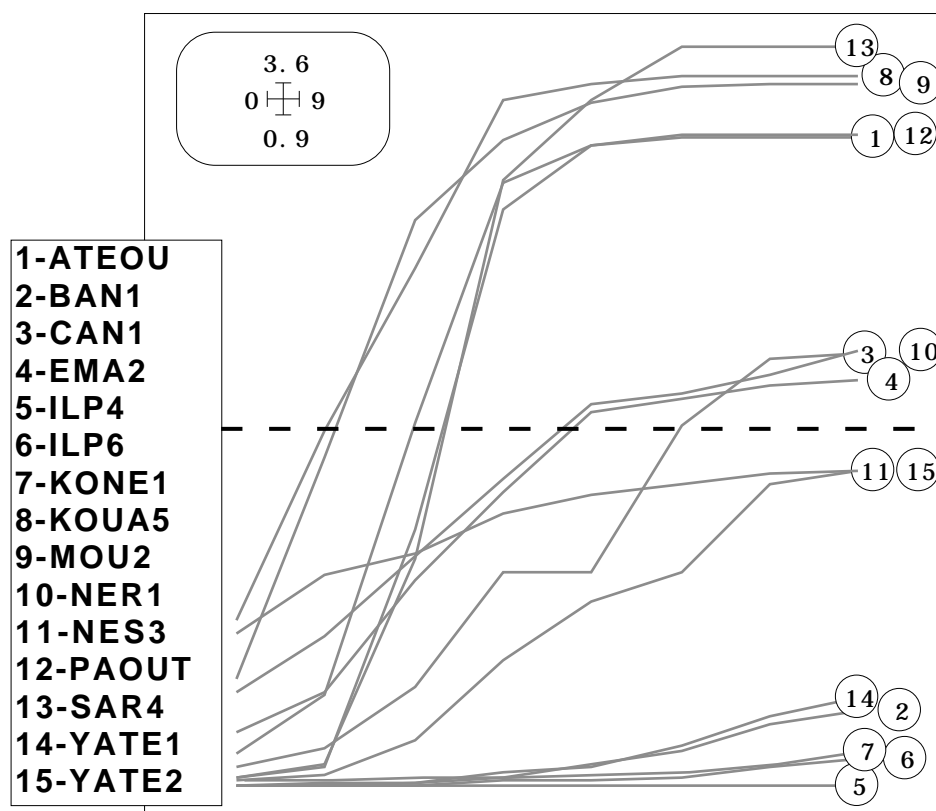


Figure 2 — Expression de la co-inertie vue par l'axe 1 de STATICO. Les 120 relevés (8 dates - 15 stations) sont positionnés en abscisse par leur projection sur l'axe 1 de co-inertie dans l'espace des variables de milieu (fichier A.xaTYLI) et en ordonnée par leur projection sur l'axe 1 de co-inertie dans l'espace des variables phyto-sanitaires (fichier A.xaTXLI). Les cercles sont les positions à la date 1, les trajectoires sont celles des stations. Il y a deux groupes de stations (milieu) qui renvoie à deux types d'évolution : rouille seule (moyenne ou forte) à droite, autres pathologies à gauche (rouille modérée et antracnose, antracnose forte, cercosporiose forte, antracnose et cercosporiose, aucune attaque).

Tester la co-inertie (CoInertia : Coinertia test - Fixed D) :



Le résultat est tout-à-fait satisfaisant :



Il y a trois classes de stations : on en fait une variable qualitative Cla (15-1 avec 3 modalités). On sélectionne les quatre variables efficaces dans un fichier MRed (15 stations et les variables Structure, pH, alt et ombr invariantes dans le temps). L'analyse de variance qui s'en suit (Discrimin : Anova1-FF) n'est significative que pour une seule des 4 variables et l'analyse discriminante associée (Discrimin : Initialize/LinkPrep et Discrimin : Discriminant analysis/Test) ne l'est même pas ! En ne gardant qu'une forme simple des éléments de l'analyse qui précède, on tombe sur un échec. La simplification est donc abusive, alors que le test p. 37 est convaincant. Qu'a t'on laissé de côté qui appauvrisse à ce point la signification des données ?

En fait, la co-structure propose du côté environnement une partition en deux classes (figure 2). La signification expérimentale de cette situation devra être discutée. On se contentera ici de souligner que les aides à l'interprétation dans la méthode STATIS étendue aux opérateurs de co-inertie devraient rendre des services par leur plasticité.

Références

¹ Ter Braak, C.J.F. (1987b) *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Box 100, NL-6700, AC Wageningen, The Netherlands. 1-152.

Lebreton, J.D., Sabatier, R., Banco, G. & Bacou, A.M. (1991) Principal component and correspondence analyses with respect to instrumental variables : an overview of their role in studies of structure-activity and species- environment relationships. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W. (Eds.) Kluwer Academic Publishers. 85-114.

Chessel, D. & Mercier, P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 15-44.

² Townsend, C. R., Hildrew, A. G. & Francis, J. (1983) Community structure in some southern english streams: the influence of physicochemical factors. *Freshwater Biology* : 13, 521-544.

Wright, J. F., Armitage, P. D., Furse, M. T. & Moss, D. (1989) Prediction of invertebrates communities using stream measurements. *Regulated rivers: research and management* : 4, 147-155.

³ Ter Braak, C.J.T., Juggins, S. (1993) Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, 269/270, 485-502.

⁴ Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.

⁵ Franquet, E. & Chessel, D. (1994) Approche statistique des composantes spatiales et temporelles de la relation faune-milieu. *C.R. Acad. Sci. Paris, Sciences de la vie* : 317, 202-206.

Franquet, E., Dolédec, S. & Chessel, D. (1995) Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia* : 300-301, 425-431.

⁶ Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.

Chessel, D. & Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : sous presse.

⁷ Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.

⁸ Thioulouse, J. & Chessel, D. (1987) Les analyses multi-tableaux en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Œcologica, Œcologia Generalis* : 8, 4, 463-480.

Kroonenberg, P.M. (1989) The analysis of multiple tables in factorial ecology. III Three-mode principal component analysis:"analyse triadique complète". *Acta Œcologica, Œcologia Generalis* : 10, 3, 245-256.

⁹ Pegaz-Maucet, D. (1980) Impact d'une perturbation d'origine organique sur la dérive des macro-invertébrés benthiques d'un cours d'eau. Comparaison avec le benthos. Thèse de 3^o cycle, Université Lyon 1. 130 p.

¹⁰ Dolédec, S. & Chessel, D. (1991) Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India* : 1, 133-155.

¹¹ Bouroche, J.M. (1975) *Analyse des données ternaires: la double analyse en composantes principales*. Thèse de 3^o cycle, Université de Paris VI. 1-57 + annexes.

¹² Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Œcologica, Œcologia Generalis* : 8, 3, 403-426.

¹³ Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.

¹⁴ Lavit, Ch. (1988) *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1-240.

¹⁵ Forestier, M.C. (1994) *Variabilité spatio-temporelle de distribution d'Esolus parallelepipedus (Müller, 1906) (Coleoptera, Elmidae) à différents échelles de l'hydrosystème fluvial*. Thèse de Doctorat, Université Lyon 1. 243 p. + annexes.

¹⁶ Lamouroux, N., Pellegrin, F., Nandris, D. & Kohler, F. (1995) The *Coffea arabica* fungal pathosystem in New Caledonia: interactions at two different spatial scales. *Journal of Phytopathology* : 143, 403-413.

¹⁷ Franquet, E. & Chessel, D. (1994) Approche statistique des composantes spatiales et temporelles de la relation faune-milieu. *C.R. Acad. Sci. Paris, Sciences de la vie* : 317, 202-206.

¹⁸ Pagès, J. (1995) Eléments de comparaison de l'Analyse Factorielle Multiple et de la méthode STATIS. In : *XXVIIe Journées de Statistique*, Jouy-en-Josas, 15-19 mai 1995. Groupe HEC, 1 rue de la libération, 78351 Jouy-en-Josas cedex, France. 492-496.

¹⁹ Lebart, L., Morineau, A. & Tabart, N. . (1977) Techniques de la description statistique, méthodes et logiciels pour la description des grands tableaux. Dunod, Paris. 1-351. VPROP et TRIDI p. 305-306.

