

STATIS et ordination de plusieurs ensembles de relevés

Résumé

Sur proposition de leur auteur, les données ichtyologiques publiées dans la thèse de J. Verneaux (1973 - Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon. 257 p.) sont revisitées avec la méthode STATIS. La comparaison de trois groupes de variables (géomorphologiques, ichtyologiques et physico-chimiques) permet de souligner l'objectif particulier qu'on se donne en utilisant cette méthode multi-tableaux.

Plan

1 — Préparation des données.....	2
2 — Expression d'une structure stable.....	11
3 — STATIS comme analyse intra-classes achevée	14
4 — STATIS et analyse des correspondances.....	20
5 — Paramètres physico-chimiques.....	23
Annexe : Mettre en œuvre un STATIS version AFC	27
Références	34

D. Chessel, J. Thioulouse et M. Hanafi

1 — Préparation des données

Les travaux classiques de Huet ¹ montrent l'existence d'une zonation des peuplements ichthyologiques le long des cours d'eau connue sous le nom de règles des pentes. Ils distinguent les zones à Truite et à Ombre dans la partie amont (peuplement à Salmonidés) et les zones à Barbeau et à Brème dans la partie aval (peuplement à Cyprinidés). On désire analyser le niveau de reproductibilité de cette structure d'une rivière à l'autre, sa permanence ou sa variabilité et détecter des causes éventuelles de sa perturbation. On s'interroge en particulier sur la solidité de la liaison de cette structure faunistique avec l'évolution multivariée des paramètres morphologiques et physico-chimiques relevés sur des stations régulièrement disposés sur plusieurs rivières. La question est posée par J. Verneaux qui propose de revisiter les données acquises sur le réseau du Doubs. Il faut préciser qu'on assigne au traitement statistique des objectifs qu'il n'eut pas été raisonnable d'envisager au moment de l'acquisition.

On trouvera ici le point de vue du statisticien qui ne préjuge pas de l'interprétation des résultats qui sera faite. Du point de vue statistique, la présence de trois groupes de variables (respectivement faunistiques, géomorphologique et physico-chimiques) permet de tester le comportement de la méthode face à trois types de descripteurs qu'on sait avoir des structures très différentes. Les données traitées sont celles de la thèse citée mais ont subi certains traitements préliminaires.

1.1 — Choix des stations

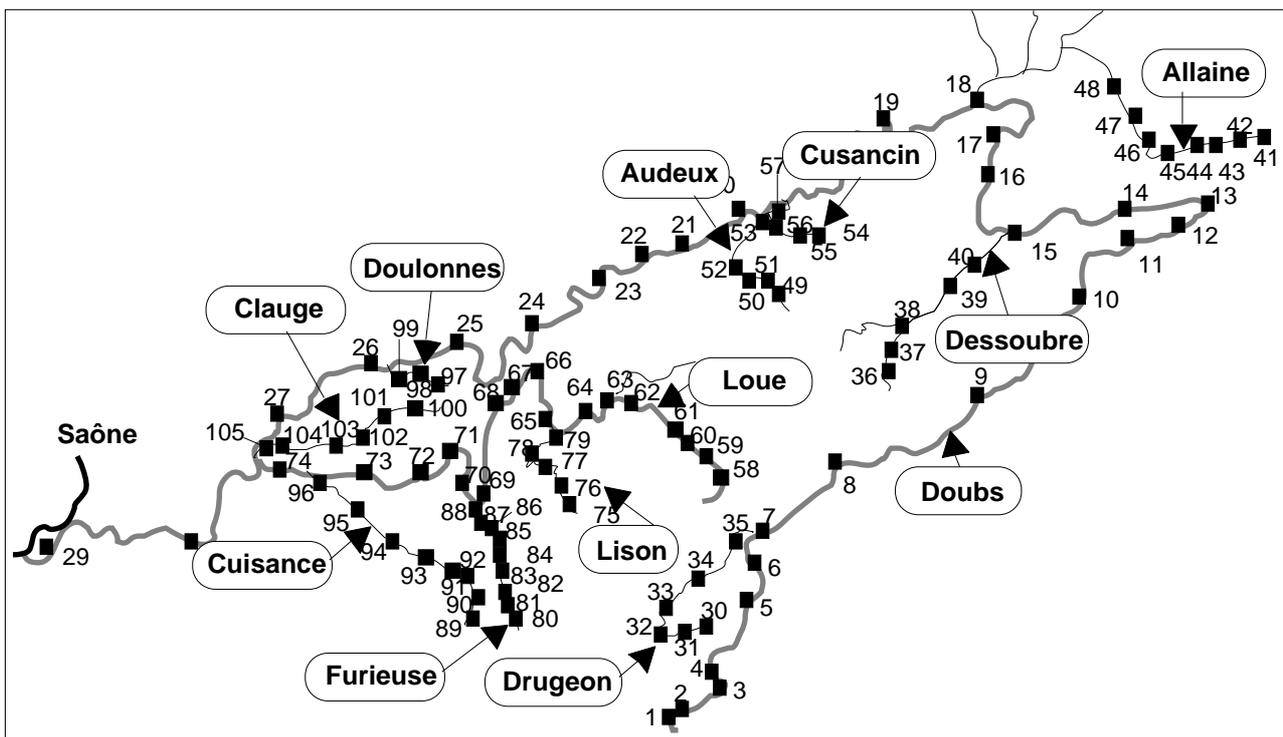


Figure 1 : code des rivières et des stations.

Il s'agit de comparer des structures et non des stations. Ce point est essentiel. La typologie des stations est une chose, celle de la structure des rivières en est une autre. Comme dans toute question de typologie, il convient de considérer un ensemble homogène. L'homogénéité recherchée est celle des niveaux de variabilité des peuplements et des paramètres à l'intérieur d'une rivière.

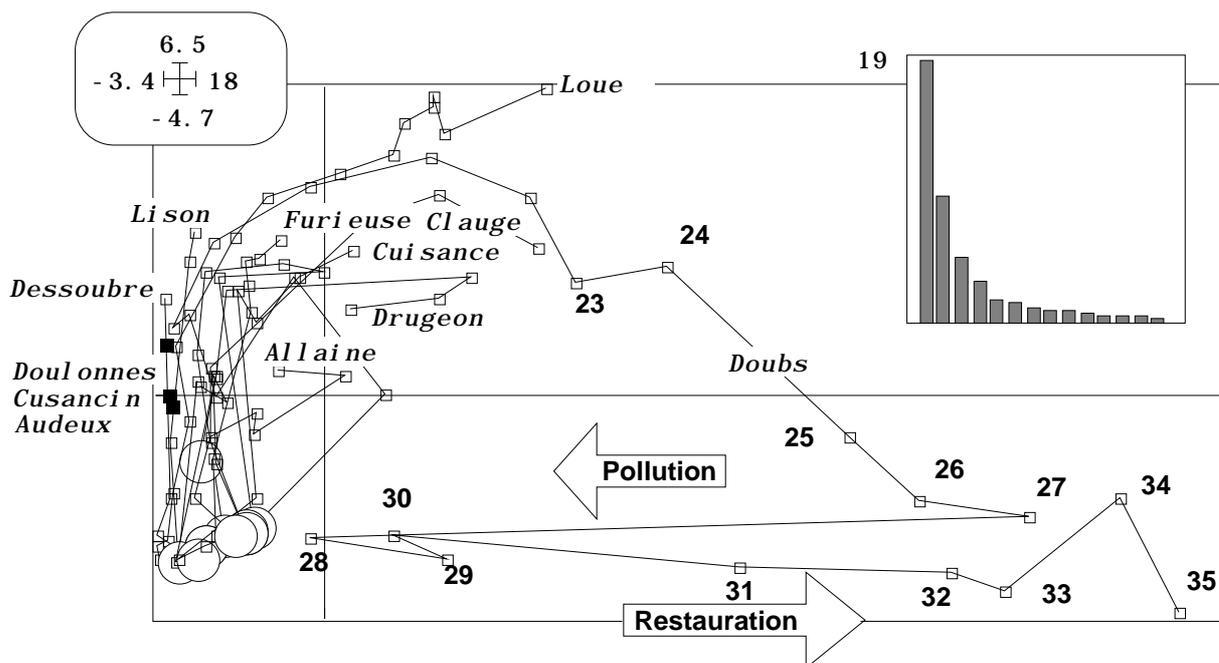


Figure 2 : Carte factorielle des relevés de l'analyse en composantes principales centrée du tableau 105 stations et 29 espèces. En haut valeurs propres de l'analyse.

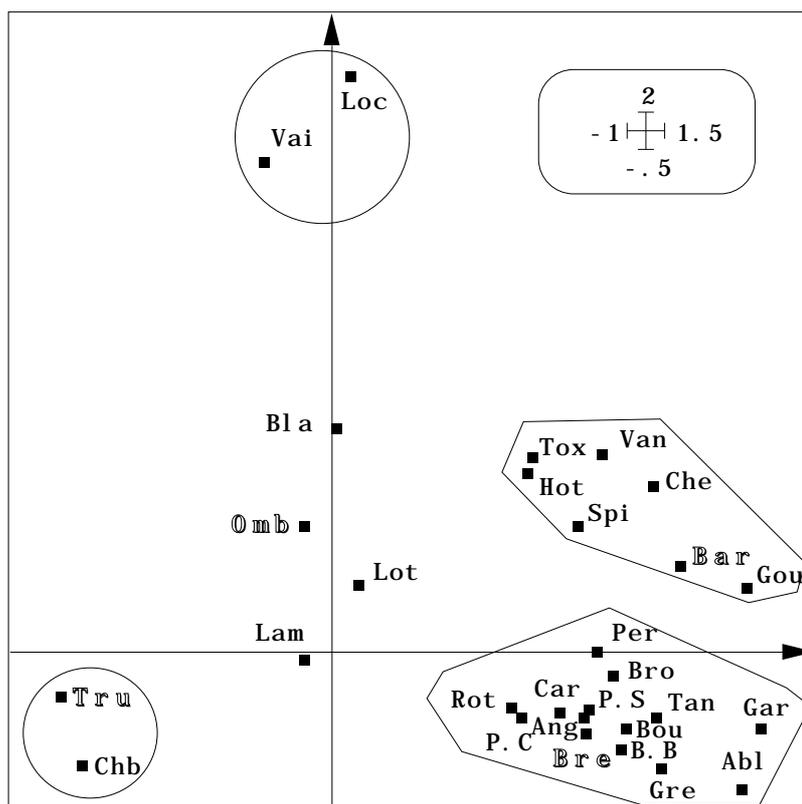


Figure 3 : Carte factorielle des colonnes de l'analyse en composantes principales centrée du tableau 105 stations et 29 espèces. En gras, les espèces étiquettes des zones de Huet.

Les 111 stations (figure 1) sont réparties le long de 12 rivières. Sont enlevées les 6 stations placées dans les retenues du haut Doubs et qui sont étrangères au problème posé. Les 105 stations comportent une note d'abondance pour 29 taxons. Une analyse en composantes principales (carte 1-2 des relevés : figure 2, carte 1-2 des espèces :

figure 3) sur ce premier tableau 105-29 soulignent deux faits d'importance. Le premier concerne le rôle du Doubs aval, seule partie du réseau contenant un peuplement typiquement à Cyprinidés. Le second concerne la prise en compte en ACP des questions d'abondance et la mise en évidence de la diminution sensible de l'abondance de toutes les espèces et de la restauration progressive de la rivière dans l'aval du Doubs.

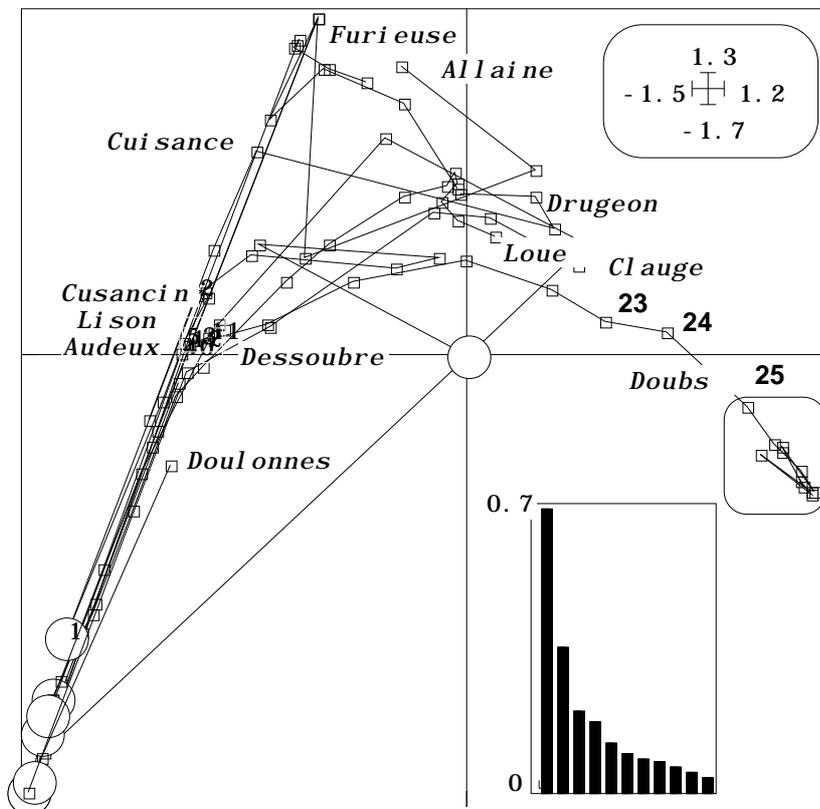


Figure 4 : Carte des relevés de l'analyse des correspondances du tableau 105 stations et 29 espèces.

En comparaison l'analyse des correspondances (figure 4), dont c'est la vocation par un double centrage implicite, identifie parfaitement que la liste faunistique est constante entre les stations 25 et 35. Dans les deux analyses, on sort d'un schéma global d'ordination commune à partir de la station 23 du Doubs. L'ACP dit que la densité y varie fortement sous l'effet de la pollution, tandis que l'AFC souligne qu'il n'y a pas modification de la liste d'espèces. Mais quel que soit le point de vue, les 11 stations aval sur le Doubs ne feraient que perturber une analyse précise de la structure longitudinale comparée des autres rivières, y compris le Doubs amont.

La carte des espèces de l'AFC (figure 5) est aussi différente que possible de son équivalente en ACP. On y voit généralement un effet Guttman. Ce n'est pas le lieu de discuter cet aspect ². Il intervient souvent quand il convient d'exprimer deux faits de structure articulés. Ici la partie potamique du Doubs et les Cyprinidés qui s'y tiennent en est un. L'évolution du peuplement le long des affluents en est un autre. L'articulation est évidente. Le rôle du Doubs aval, dont l'ACP décrit la variation de densité et l'AFC la permanence de la liste d'espèces, joue un rôle essentiel dans cette différence entre les deux analyses.

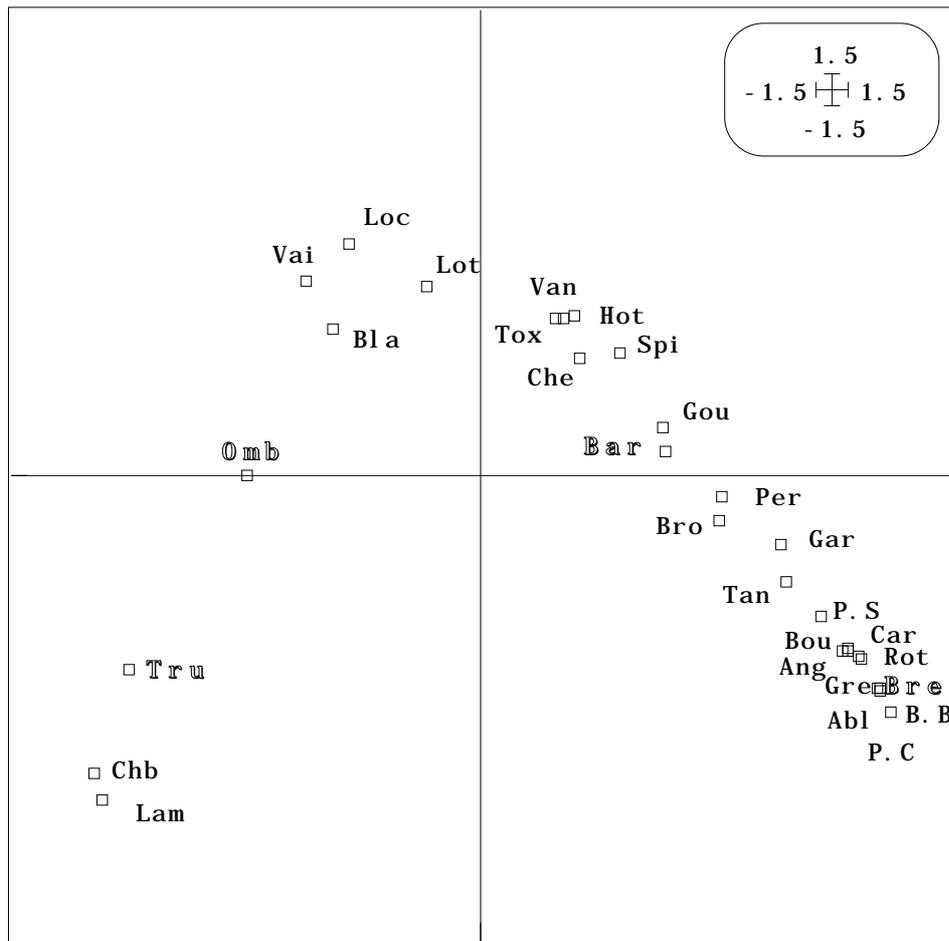


Figure 5: Carte des espèces de l'analyse des correspondances du tableau 105 stations et 29 espèces.

On s'en convainc facilement en éliminant du tableau les stations sur le Doubs en aval de la station 23. Il reste 92 stations. Les espèces présentant au plus 4 présences (seuil de 5%) sont enlevées. Il reste 19 taxons. L'ACP et l'AFC de ce nouveau tableau, qui restera celui des analyses qui suivent, donnent les cartes de la figure 6 (plan 1-2). Les différences sont très sensiblement moindres. On notera essentiellement, que d'un point de vue comme de l'autre, ces plans expriment un mélange intime de différences inter-rivières et de variabilité intra-rivière.

Les représentations des espèces restent, cependant différentes. On notera que la contrainte de centrage pour la carte espèce est propre à l'AFC, ce qui en présence d'un effet taille (augmentation simultanée de l'abondance de toutes les espèces dans certaines stations) sépare les deux analyses. Il reste un effet d'abondance considérable dans le tableau faunistique, vraisemblablement lié aux pollutions qui ne sont pas réservées à l'aval du Doubs. La présence de plus de deux facteurs (franchement 3 en ACP, peut-être 4 en AFC, figure 6) rend la comparaison difficile. En général, si l'abondance des individus capturés n'est pas maîtrisée dans un relevé l'AFC s'impose. Si, au contraire les variations d'abondance totale entre deux stations ont une signification expérimentale, l'ACP est préférable. Si les abondances totales des relevés sont sensiblement constantes (ce qui n'est pas le cas, cf. la première station vide sur l'Allaine), les résultats sont voisins. Dans la suite, les données seront recentrées par rivières et une bonne part des effets sur l'abondance seront éliminés. Par contre, nous n'éliminerons pas les effets intra-rivière de la pollution en choisissant de pratiquer des analyses du type ACP (sans centrage par relevé).

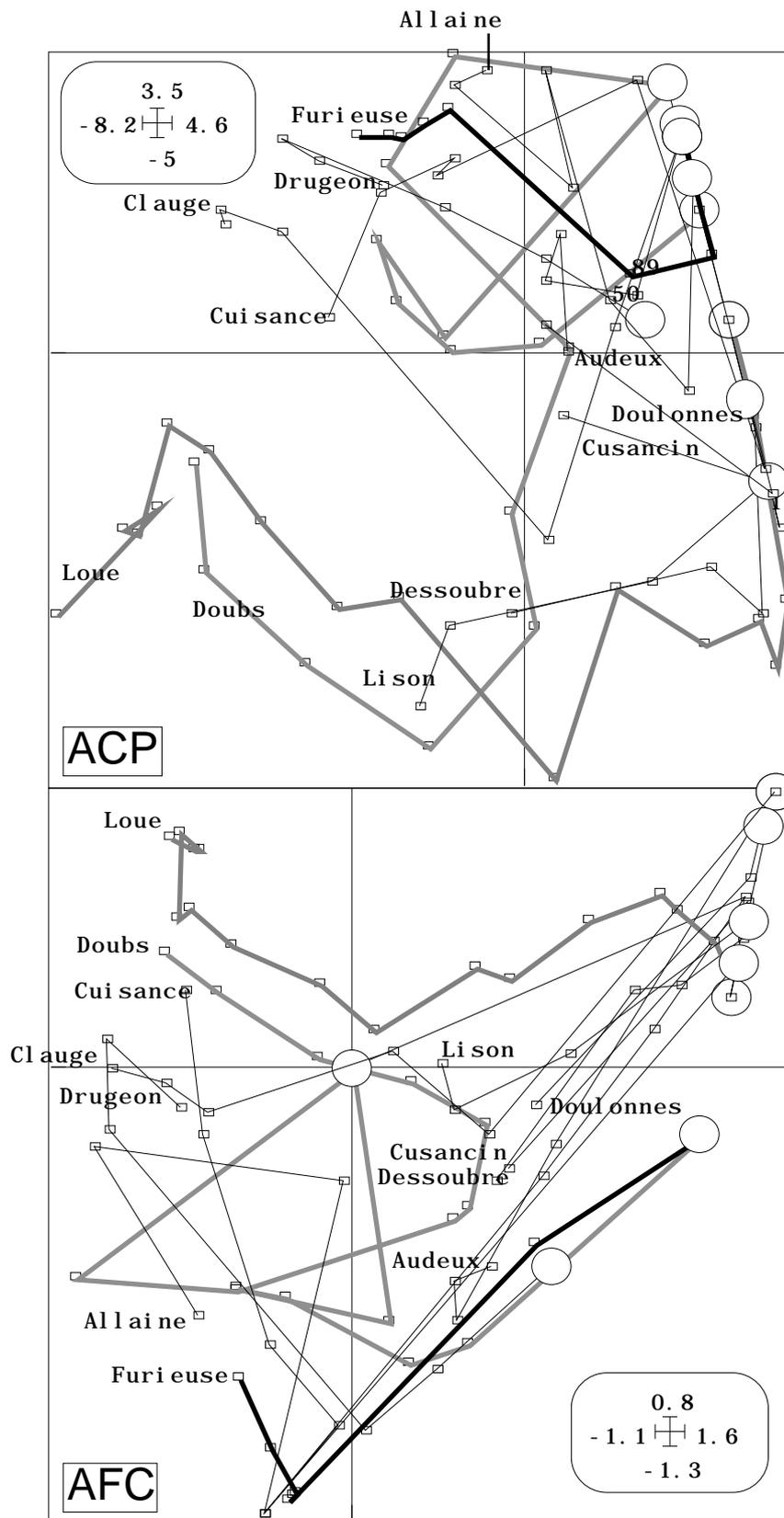


Figure 6 : Carte des relevés de l'analyse en composantes principales et de l'analyse des correspondances du tableau 92 stations et 19 espèces.

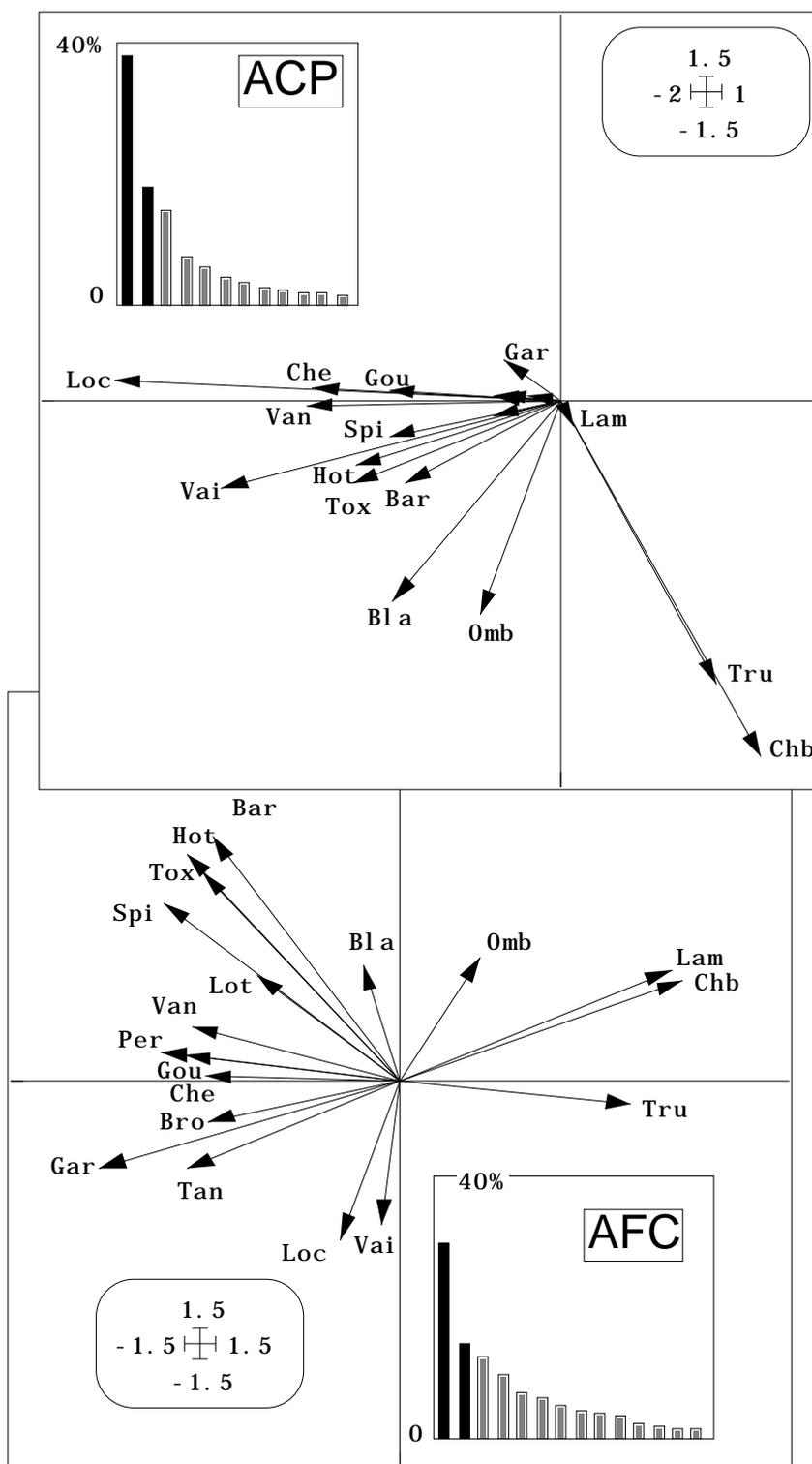


Figure 6 (suite) : Valeurs propres (pourcentages d'inertie) et carte des colonnes (plans 1-2) de l'analyse en composantes principales et de l'analyse des correspondances du tableau 92 stations et 19 espèces.

1.2 — Recodage des variables

Les variables retenues pour la description du milieu sont simplement recodées pour uniformiser les amplitudes de variation et rendre les distributions plus acceptables. Les variables peuvent prendre des valeurs comprises entre 0 et 9, valeurs associées aux numéros de classes.

Les données traitées sont reproduites dans le tableau 1. Les classes sont définies par des suites arithmétiques de raison e ou des suites géométriques de raison r suivant les cas (transformations linéaire ou logarithme). Les paramètres utilisées sont :

Repères géomorphologiques					
Nom	Code	Unités	Raison	Classe 0	Classe 9
Altitude	Alt	m	$e = 50$	< 150	> 1350
Distance à la source	Das	km	$r = 2.5$	< 0.4	> 810
Pente	Pen	$^{\circ}/^{\circ}$	$r = 2$	< 0.1	> 25.6
Section mouillée	Smm	m ²	$r = 3$	< 0.05	> 328
Débit moyen	Qmm	m ³ /s	$r = 2$	< 0.2	> 510.2
Paramètres physiques et chimiques					
Nom	Code	Unités	Raison	Classe 0	Classe 9
Vitesse (Qmm/Smm)	Vme	m/s	$e = 0.1$	< 0.1	> 0.9
Température	Tmm	$^{\circ}\text{C}$	$e = 2$	< 8	> 24
Conductivité	Con	$\mu\text{s}/\text{cm}$	$e = 75$	< 50	> 650
pH	pH		$e = 0.4$	< 6	> 9.2
Dureté	Dur	mg/l Ca	$e = 15$	< 10	> 130
Chlorures	Cl-	mg/l	$e = 3$	< 3	> 27
Sulfates	SO ₄ --	mg/l	$e = 10$	< 10	> 90
Phosphates	PO ₄ ---	mg/l	$r = 2$	< 0.005	> 1.280
Nitrates	NO ₃ -	mg/l	$r = 2$	< 0.05	> 12.8
Azote nitreux et ammoniacal	NO ₂ -/NH ₄ +	mg/l N	$r = 2$	< 0.004	> 1.024
Oxygène dissous	O ₂ %		$e = 5$	< 60	> 100
Oxydabilité	OXY		$e = 2$	< 1	> 17
DBO	DBO	mg/l	$e = 2$	< 1	> 17

1.3 — Cadre de représentation des données

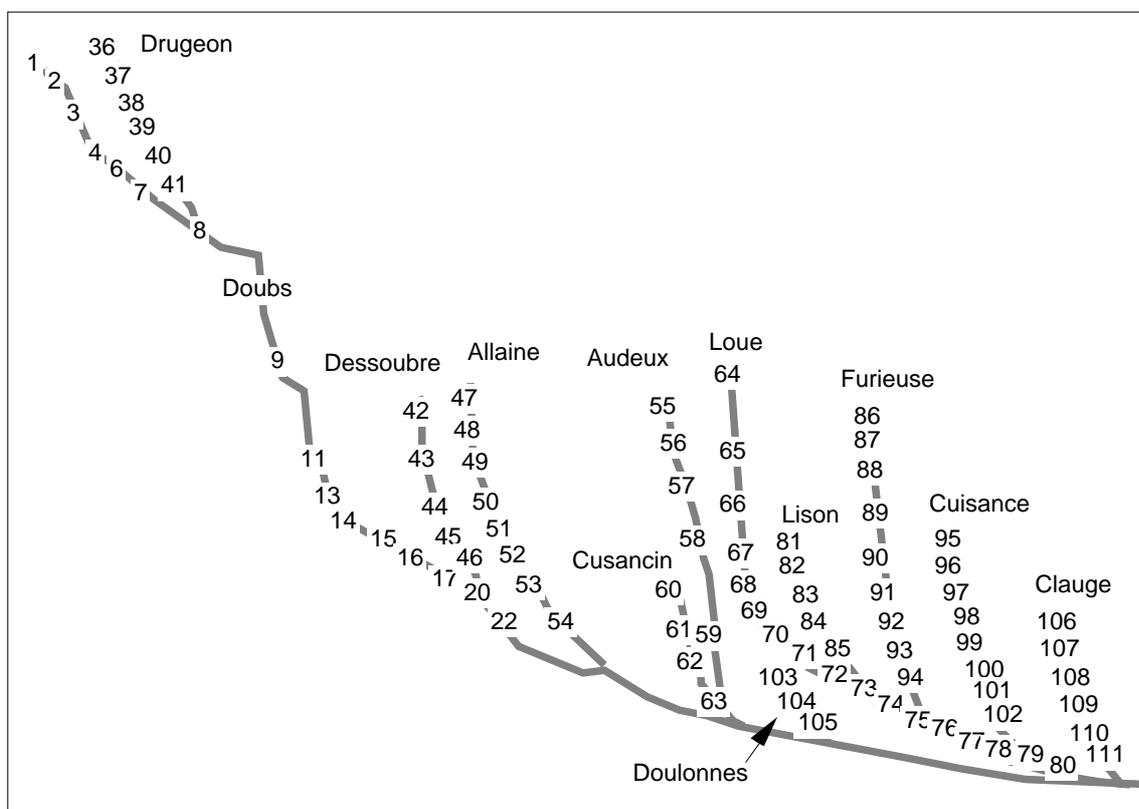


Figure 7 : Profil longitudinal simplifié pour l'expression des données.

L'expression cartographique des données et des résultats dans l'espace, sur le fond de carte de la figure 1, est peu satisfaisante. On lui préférera le cadre de la figure 7, expression du profil longitudinal du réseau hydrographique du bassin du Doubs (Verneaux, 1973, op. cit. p. 60). Les étiquettes des stations conserveront le code initial dans toutes les figures.

La position d'une station, en abscisse, est définie par la distance à la confluence avec le Doubs et le confluent est positionné par la distance à la source (intervalle de variation [0, 400 km]) et en ordonnée par l'altitude (intervalle de variation [175, 975 m]). Les valeurs sont légèrement modifiées pour faciliter la lecture des figures. La figure 8 contient la représentation des paramètres géomorphologiques centrés.

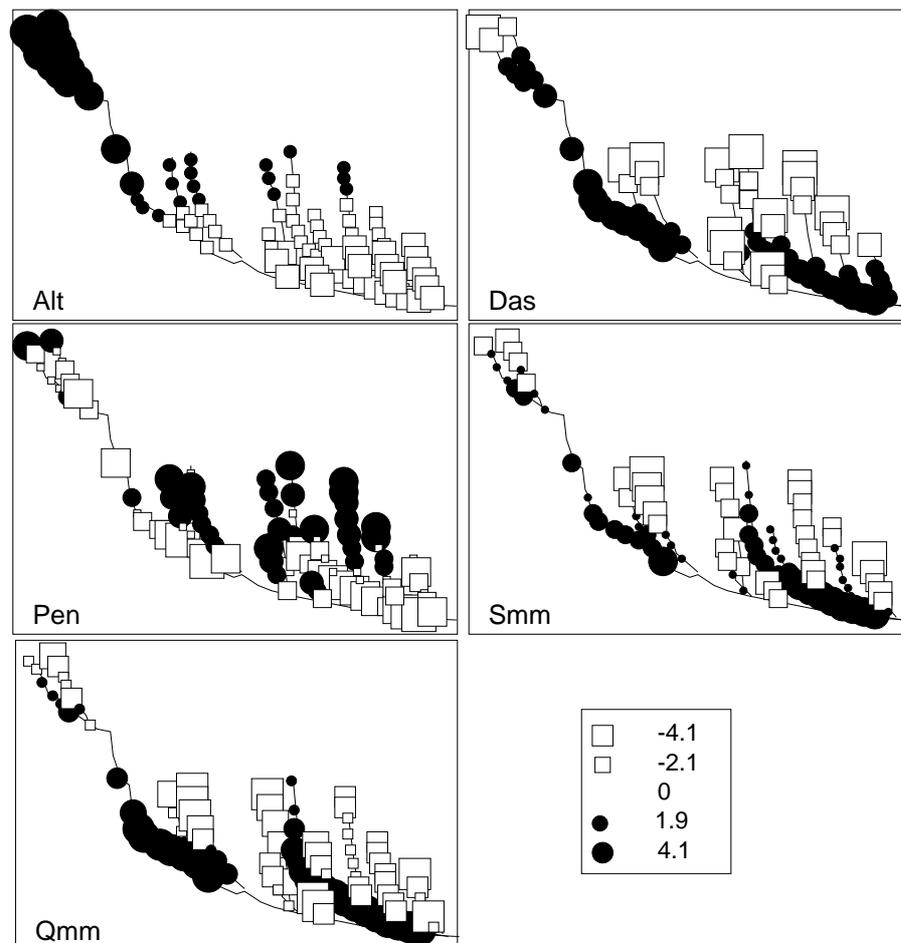


Figure 8 : représentation des paramètres géomorphologiques centrés.

2 — Expression d'une structure stable

L'intérêt des données est d'offrir trois types de descripteurs de nature et de fonction différentes. Le premier est celui de la géomorphologie des rivières. On utilise STATIS pour l'examiner et justifier le reste de l'expérience. En effet, il serait inutile d'observer la stabilité de la structure faunistique en dehors d'un espace écologique dont on peut dire qu'il offre des répliquats comparables entre eux. Est en cause le gradient amont-aval de plusieurs rivières. Nous avons éliminé la partie potamique du Doubs pour homogénéiser les données. Le reliquat est-il favorable à une comparaison de structures ?

Il ne s'agit pas là d'une observation pertinente sur les données (dans le gradient amont-aval l'altitude diminue, la distance à la source augmente, la pente diminue, la section et le débit augmentent) mais sur la méthode. Les relations entre variables sont stables d'une rivière à l'autre. Quand ces relations sont stables les faisceaux d'axes de même rang sont étroits, le compromis est fortement relié à chaque étude.

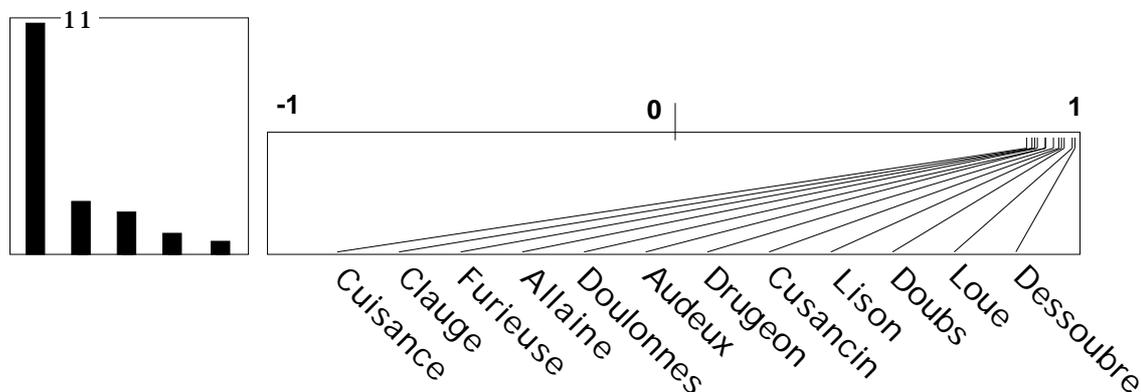


Figure 10 : A gauche, valeurs propres de l'analyse du compromis de Geo. A droite, projections des axes 1 (projection uni-dimensionnelle d'un ensemble de vecteurs unitaires sur un vecteur unitaire). Les axes 1 de chaque analyse forment un faisceau très étroit.

La méthode donne un axe commun (coordonnées des variables, soit à une constante près le poids de la variable dans la constitution des ordinations par tableau) et une typologie interne des relevés basée sur cet axe commun (figure 11). On peut y suivre l'évolution de chaque rivière. Ces observations montrent qu'on peut raisonnablement travailler une typologie interne des rivières d'un point de vue "toutes choses égales" pour ce qui est de la morphologie.

Cela ne prouve pas, au contraire, que la variabilité soit constante dans chaque rivière. Le nombre de stations, leur répartition, la longueur de la rivière, ... donne une hétérogénéité variant de 0.3 à 3.5 (carré des normes : Cf. tableau 3). La participation au compromis est par contre moins variable et la mesure de ressemblance au compromis proposé dans la fiche STATIS-1 prend des valeurs élevées pour pratiquement tous les tableaux.

Number	Rows	Weights	NS norm2	Cos2
1	16	4. 718e- 01	3. 484e+00	0. 846
2	6	1. 570e- 01	4. 311e- 01	0. 769
3	5	3. 897e- 01	2. 194e+00	0. 962
4	8	3. 922e- 01	3. 047e+00	0. 638
5	5	1. 302e- 01	3. 165e- 01	0. 678
6	4	2. 578e- 01	1. 098e+00	0. 821
7	17	3. 748e- 01	2. 051e+00	0. 945
8	5	3. 059e- 01	1. 512e+00	0. 818
9	9	2. 035e- 01	8. 383e- 01	0. 638
10	8	2. 001e- 01	8. 847e- 01	0. 574
11	3	1. 431e- 01	3. 911e- 01	0. 713
12	6	1. 663e- 01	6. 040e- 01	0. 567

Tableau 3 : paramètres de description des structures des tableaux : numéro d'ordre, nombre de stations, poids dans le compromis, carré de la norme de l'opérateur, cosinus carré de l'opérateur avec l'opérateur de rang 1 approximation du compromis.

Il convient de rechercher, par la même voie, une même homogénéité éventuelle de l'évolution du peuplement ichthyologique.

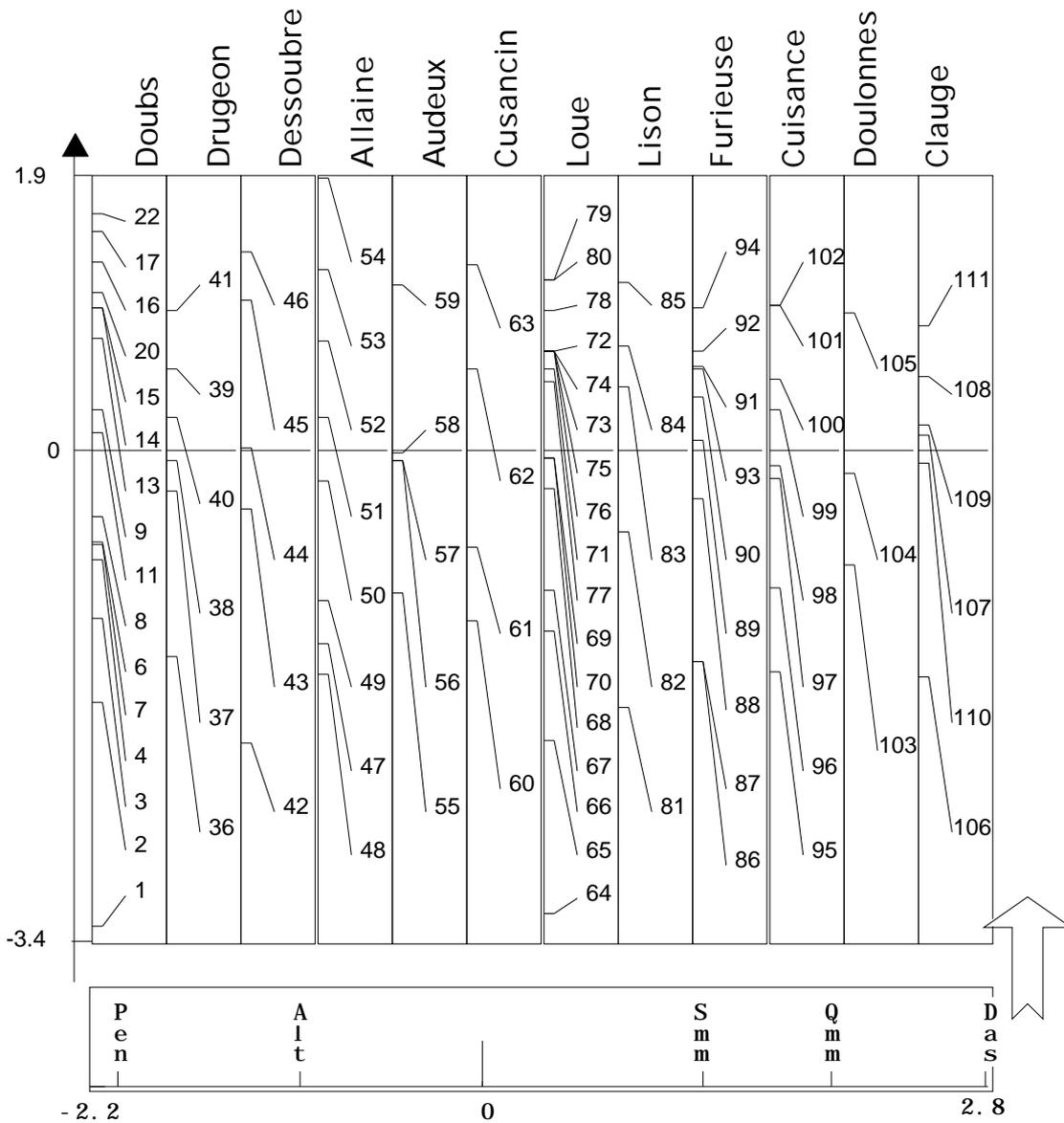


Figure 11 : En bas, coordonnées des variables de l'analyse du compromis de Geo (facteur 1). En haut, projections des lignes de chaque tableau.

3 — STATIS comme analyse intra-classes achevée

Le tableau Poi comporte 92 relevés et 19 taxons. Il est partitionné en 12 sous-tableaux centrés par espèces. Les coefficients R_v sont dans le tableau 4 et les paramètres de description de structure dans le tableau 5.

Une diminution des R_v est sensible. Les paramètres ont été largement modifiés. Les normes des opérateurs sont très variables (0.04 à 1.7). Les cosinus carrés mettent en évidence une rivière de référence, la Loue. C'est la structure interne de cette rivière qui s'approche au mieux de la structure compromis. L'Audeux (5) n'est pratiquement pas concernée.

La structure est essentiellement bidimensionnelle (figure 12). On s'aidera de la cartographie des espèces (figure 13) pour interpréter les résultats. Un fait marquant est la disparition de l'Ombre de la liste des espèces structurantes. Les différences entre la carte de STATIS et celle des analyses simples sont instructives.

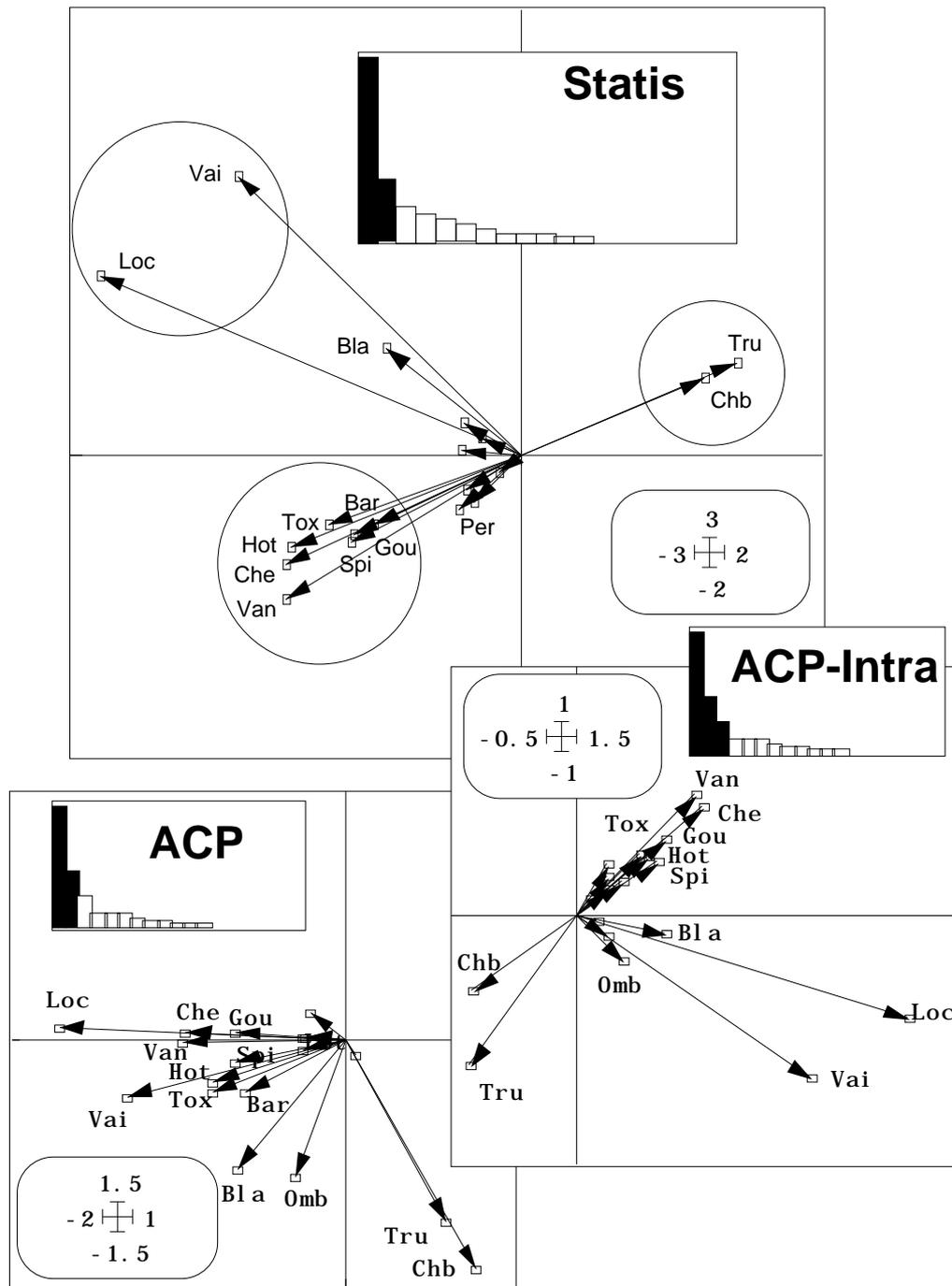


Figure 12 : valeurs propres et carte des variables (1-2) du compromis de Poi. Pour comparaison, cartes des variables de l'ACP (figure 6, en haut) et carte des variables de l'ACP intra-classes.

L'effet taille disparaît avec l'introduction de la notion de variabilité intra-rivière. La variation de densité de l'Ombre est d'abord une variation inter-rivières. Loue, Lison et Doubs (en bas) sont des rivières à Ombre.

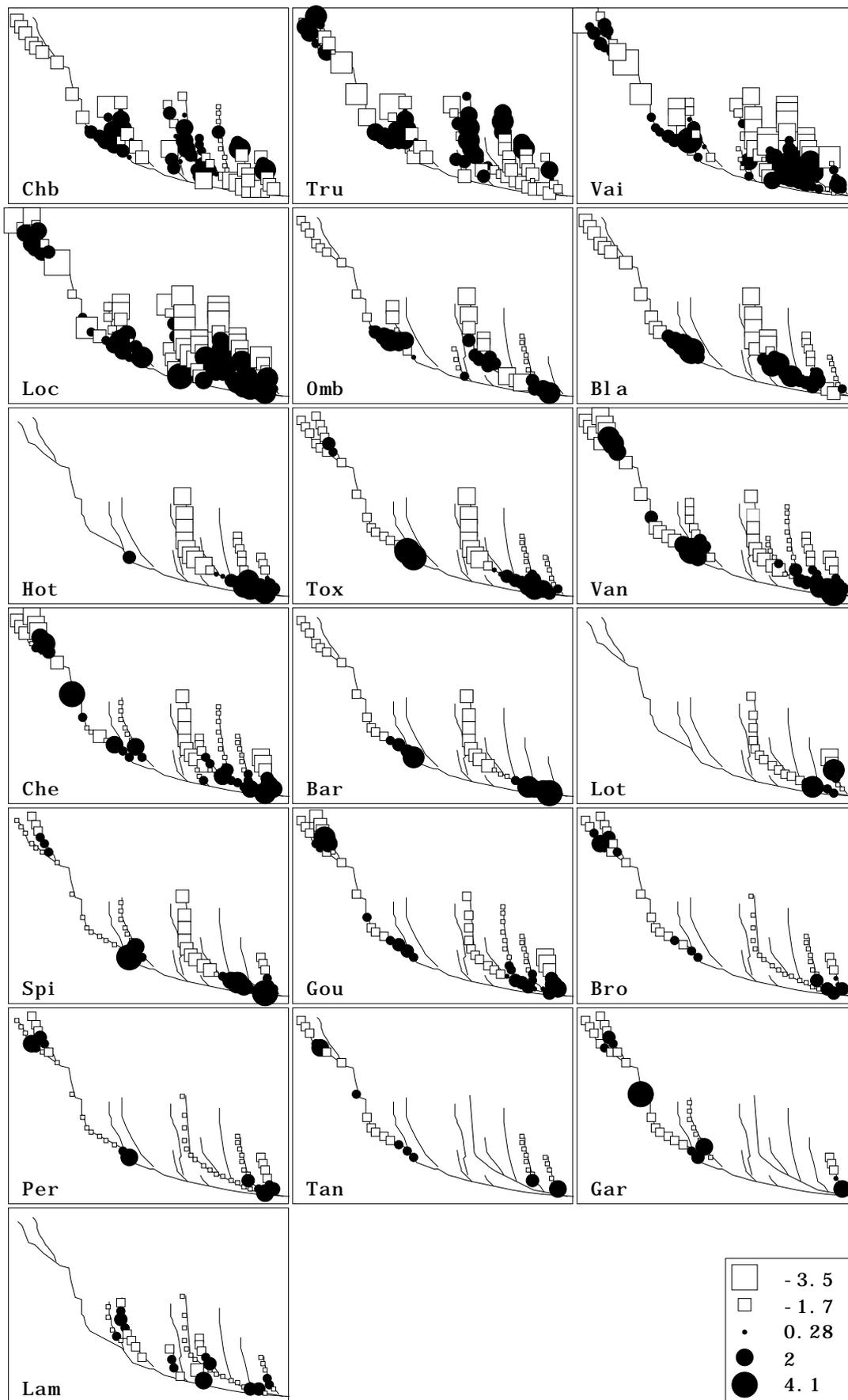


Figure 13 : représentation de l'abondances des espèces. Données centrées par rivières utilisées dans l'ACP intra-classes et dans STATIS.

```

-----
[ 1] 1000
[ 2] 421 1000
[ 3] 556 99 1000
[ 4] 416 330 426 1000
[ 5] 528 100 746 392 1000
[ 6] 265 144 410 675 447 1000
[ 7] 440 311 203 387 113 307 1000
[ 8] 418 74 481 231 397 398 360 1000
[ 9] 278 182 409 384 329 669 386 521 1000
[10] 420 221 460 488 305 596 649 631 777 1000
[11] 218 67 401 240 378 299 301 535 475 545 1000
[12] 460 515 291 414 257 235 519 290 326 493 331 1000
-----

```

Tableau 4 : Coefficients de corrélation vectorielle (RV) du tableau Poi.

Figurant encore un peu dans l'intra-classe, l'Ombre ne joue plus aucun rôle dans STATIS qui insiste sur la notion de covariances entre espèces. L'espèce varie *pour son compte*, c'est-à-dire ne rentre pas significativement dans une co-variation avec une ou plusieurs autres espèces. En ce sens, elle ne peut servir d'étiquettes de zonation. L'essentiel se passe entre le couple Truite-Chabot (auquel on peut associer le Blageon), le couple vairon-Loche et le reste (zone dite à Barbeau).

Number	Rows	Weights	NS norm2	Cos2
1	16	2.395e-01	3.913e-01	0.295
2	6	2.389e-01	6.379e-01	0.203
3	5	1.295e-01	2.473e-01	0.130
4	8	8.540e-02	7.001e-02	0.221
5	5	4.227e-02	4.474e-02	0.080
6	4	8.583e-02	9.186e-02	0.182
7	17	7.238e-01	1.707e+00	0.813
8	5	2.393e-01	4.952e-01	0.312
9	9	2.198e-01	3.806e-01	0.329
10	8	2.619e-01	2.865e-01	0.602
11	3	1.139e-01	1.637e-01	0.191
12	6	3.761e-01	7.168e-01	0.473

Tableau 5 : paramètres de description des structures des tableaux : analyse du tableau Poi.

STATIS peut ici être considérée comme une analyse intra-classe achevée, en ce sens qu'on aurait une bonne part des résultats avec une intra-classe, la qualité des aides à l'interprétation en termes de représentativité par blocs de la structure moyenne étant grandement enrichie. Car la structure triangulaire s'exprime plus ou moins dans chaque rivière.

La figure 14 donne l'information associée aux lignes des 12 tableaux (relevés). Elle contient la carte factorielle de chaque rivière sur laquelle on a représenté le facteur 1 de l'analyse STATIS sur le tableau Geo. Ceci permet d'obtenir l'orientation commune des cartes (valeurs négatives en tête de rivière, Cf. figure 11). Est mentionnée le cosinus carré de l'opérateur avec le compromis approximé. Est figurée encore la projection des trois premiers axes de l'ACP de chaque tableau. Il est impossible de faire mieux en termes d'expression simultanée des structures faunistiques par rivière.

Le Doubs apparaît comme profondément perturbé, l'Allaine comme beaucoup trop homogène, la Loue comme la structure de référence et il semble manquer des points

dans la Furieuse et la Cuisance. La zonation faunistique classique est un modèle simple qui se réalise avec beaucoup de variabilité.

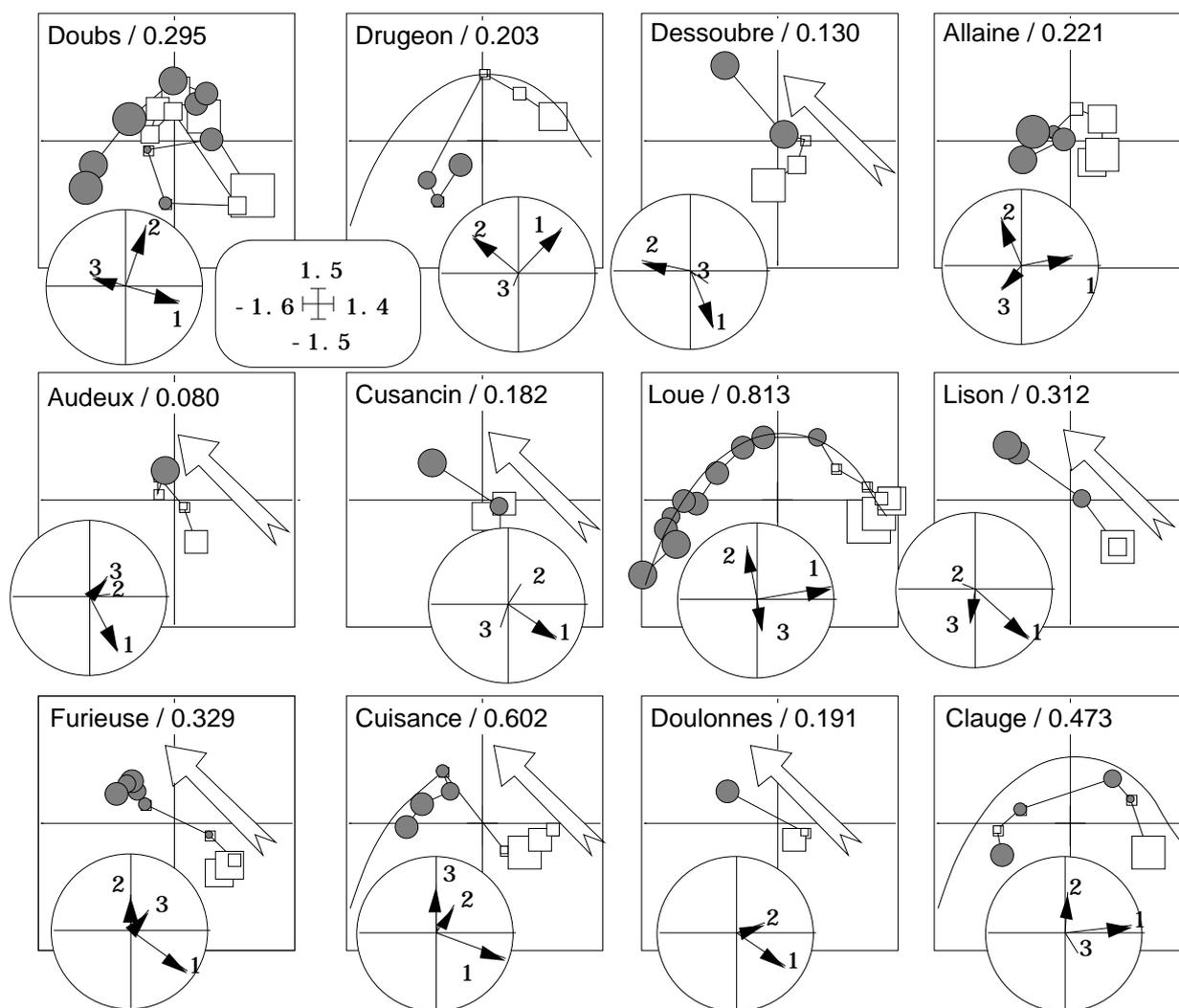


Figure 14 : STATIS du tableau Poi. Typologie simultanée de 12 tableaux faunistiques. Chaque station de chaque rivière est projetée sur le plan 1-2 du compromis. Le cartouche donne les bornes des coordonnées communes à tous les sous-tableaux. Les cercles gris et les carrés blancs donnent la première coordonnées de STATIS sur le tableau Geo, ce qui permet d'identifier les gradients amont-aval (Amont : valeurs négatives, carrés blancs ; aval : valeurs positives, cercles gris). Dans les cercles, projections des axes 1, 2 et 3 des analyse simples par tableau. La courbe parabolique indique la structure compromis globale réalisée exactement sur la Loue (rivière témoin). Les flèches indiquent la structure partielle réalisée dans une partie des rivières.

En couplant les figures 12, 13 et 14, on a toute l'information. On cherche alors une relecture des données en dehors du champ technique de l'analyse statistique.

La figure 12 invite à simplifier la faune en trois groupes A (truite + Chabot), B (Vairon + Loche) et C (le reste). La figure 14 invite à relire directement la variation du contenu faunistique d'une station à l'autre. On obtient une figure simple avec la représentation triangulaire (figure 15). On comprend alors pourquoi l'axe principal de chaque analyse est instable, et contrairement au cas des variables géomorphologiques, il n'y a pas de zonation stable d'une rivière à l'autre.

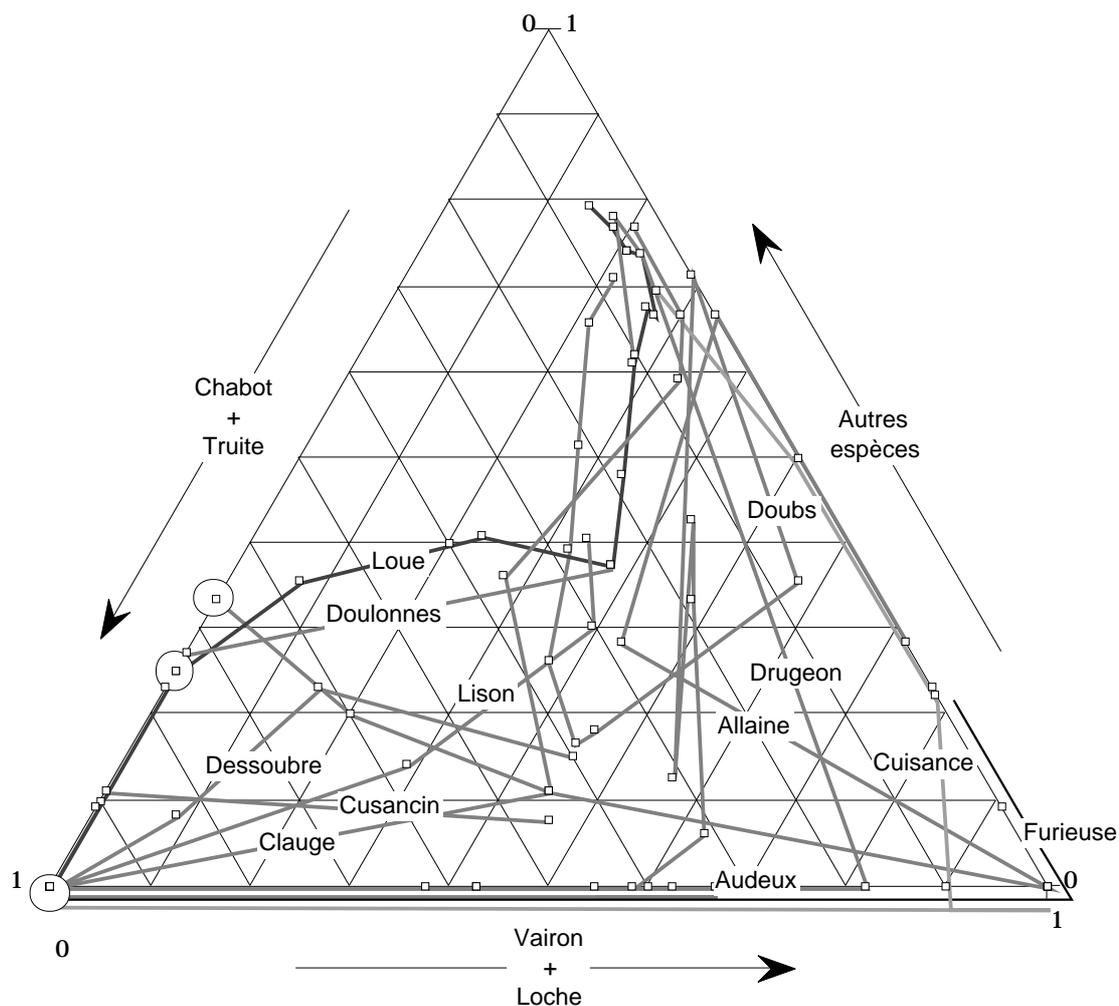


Figure 15 : représentation triangulaire des données faunistiques. En bas et à gauche une station ne contient que Truite et Chabot (point amont des rivières). En bas et à droite on n'y trouve que Vairon et Loche. En haut, il n'y aurait aucune des quatre espèces citées.

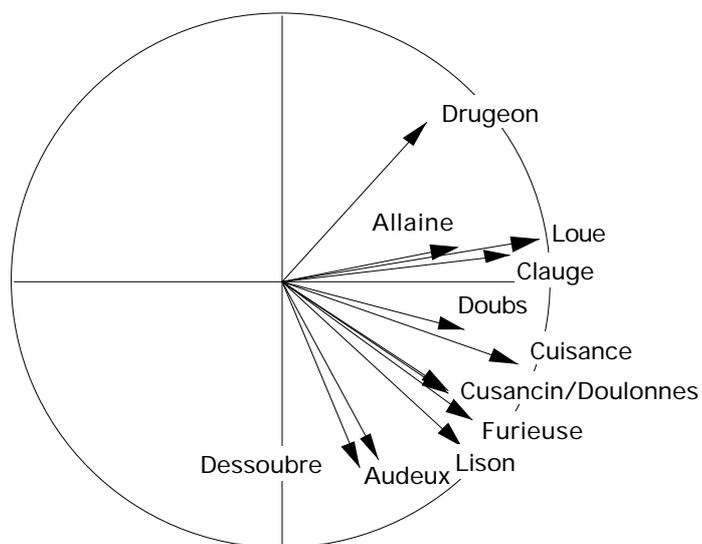


Figure 16 : projections des axes 1 des analyses séparées sur le plan 1-2 du compromis.

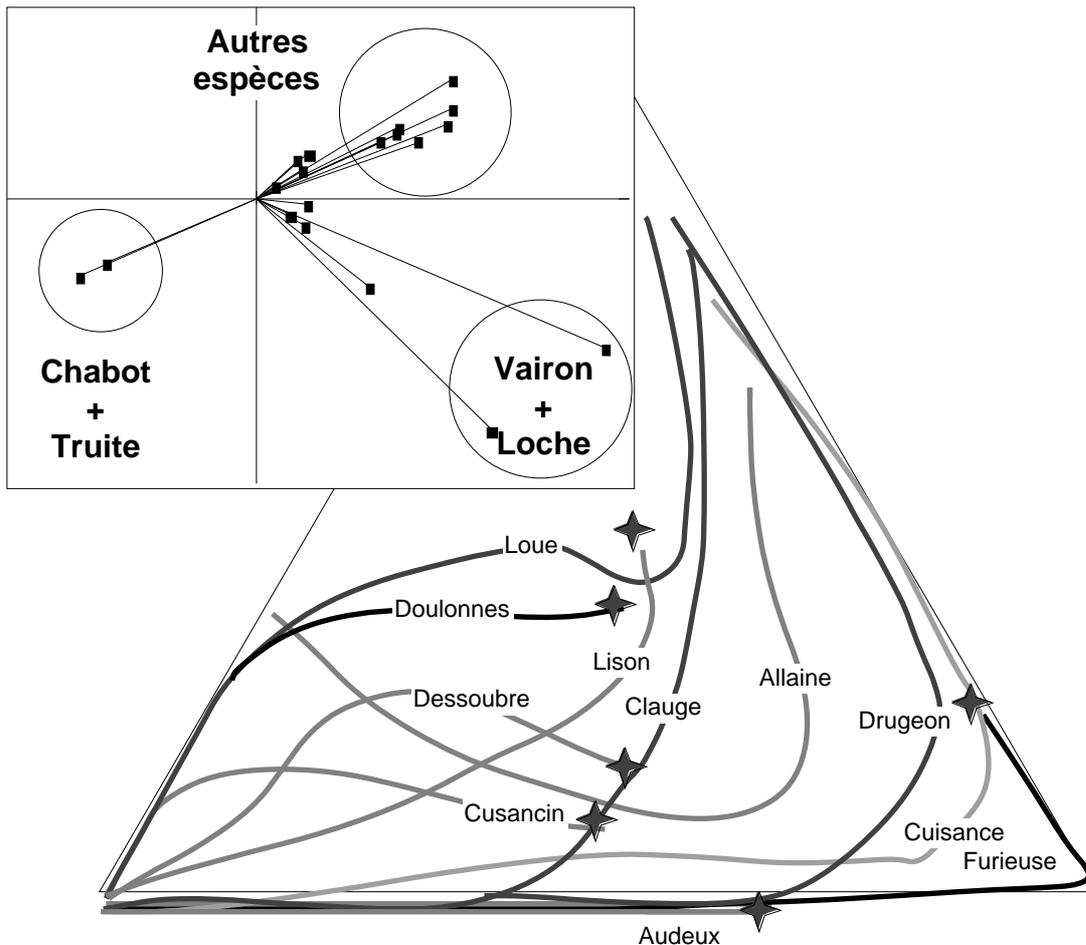


Figure 17 : résumé des variations faunistiques intra-rivières.

STATIS indique qu'il y a deux grands types de structure. Dans la Loue, la Clauge et le Drugeon, c'est la covariance négative entre Chabot-Truite et toutes les autres espèces qui fait l'essentiel de la covariation inter-spécifique. L'Allaine réalise partiellement cette évolution. Dans six rivières, Dessoubre, Audeux, Cusancin, Furieuse, Lison et Doulonnes, c'est l'opposition Chabot+Truite contre Vairon+Loche qui l'emporte. La Cuisance et le Doubs mélangent ces deux structures. Au total, on passe d'une typologie d'états au niveau de la station à une typologie de structures au niveau de la rivière.

La représentation triangulaire simplifiée (figure 17) correspond assez bien à la carte espèces de STATIS. Elle est cependant plus en accord avec un modèle d'averaging (relevés à la moyenne des espèces qu'on y trouve) et on peut se demander si une version "analyse des correspondances" de STATIS ne serait pas préférable.

4 — STATIS et analyse des correspondances

STATIS est d'abord un point de vue théorique radicalement original : on y compare des structures de tableaux via leurs opérateurs d'inertie. Si on désire comparer des structures, il faut employer STATIS, sinon il faut employer un autre outil. Comparer des structures est plus difficile que comparer des positions multivariées, ce qui est plus difficile que comparer des valeurs d'un paramètre.

```

-----
[ 1] 1000
[ 1] 1000
[ 2] 413 1000
[ 2] 421 1000
[ 3] 77 135 1000
[ 3] 556 99 1000
[ 4] 306 321 333 1000
[ 4] 416 330 426 1000
[ 5] 140 207 292 105 1000
[ 5] 528 100 746 392 1000
[ 6] 49 105 678 314 119 1000
[ 6] 265 144 410 675 447 1000
[ 7] 421 374 402 288 73 248 1000
[ 7] 440 311 203 387 113 307 1000
[ 8] 162 200 452 286 171 307 548 1000
[ 8] 418 74 481 231 397 398 360 1000
[ 9] 236 406 397 344 482 361 482 657 1000
[ 9] 278 182 409 384 329 669 386 521 1000
[10] 265 384 414 330 156 365 632 722 671 1000
[10] 420 221 460 488 305 596 649 631 777 1000
[11] 109 81 510 381 351 549 369 552 605 571 1000
[11] 218 67 401 240 378 299 301 535 475 545 1000
[12] 340 374 247 253 135 104 484 477 359 609 207 1000
[12] 460 515 291 414 257 235 519 290 326 493 331 1000
-----

```

Tableau 6: Coefficients de corrélation vectorielle (RV) : en caractères droits, option analyse des correspondances, en italique, pour mémoire, option analyse en composantes principales.

Ce point de vue signifie qu'il y a deux étapes. K tableaux définissent K structures. Chacune d'entre elle demande une analyse. Le besoin des les comparer demande qu'elles aient des points communs, en particulier les poids des colonnes (si les tableaux sont appariés par les colonnes, individus ou variables suivant les cas) et les types de centrage. STATIS définira indifféremment la moyenne de plusieurs analyses en composantes principales ou de plusieurs analyses des correspondances.

La première étape consiste donc à choisir pour chaque tableau un type d'approche, la seconde étape consiste à calculer une structure moyenne puis à comparer cette structure moyenne à chacune des structures de base.

Pour une version "analyse des correspondances", il y a une difficulté. L'analyse des correspondances d'un tableau définit les pondérations lignes et colonnes. L'analyse en composantes principales importe ces pondérations de l'extérieur. Il faut donc une version de l'analyse des correspondances qui accepte une marge imposée. L'analyse des correspondances floues, qui généralise l'analyse des correspondances multiples permet de faire cela⁴.

Le nouveau K-tableau est formé de sous-tableaux ayant en commun la pondération colonne qu'on peut choisir uniforme (chaque espèce a le poids 1/19). Les poids des lignes sont définis par tableau comme moyenne des profils-espèces. Les lignes vides ont un poids nul et une position à l'origine, comme si elles étaient absentes. Les analyses séparées travaillent sur les profils centrés du type :

$$\frac{P_{ij}}{P_{i.P.j}} - 1$$

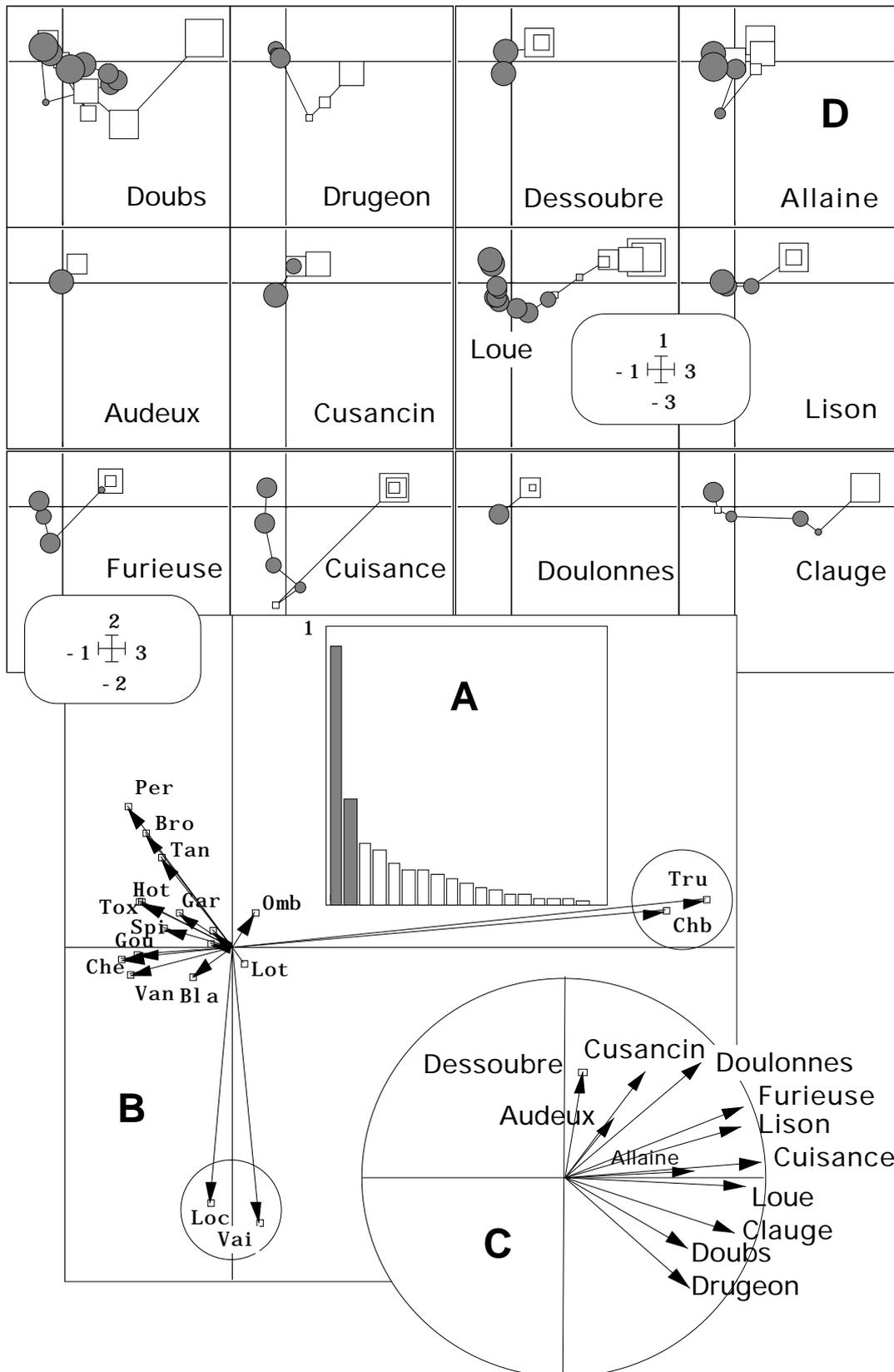


Figure 18 : Aides à l'interprétation graphique de l'analyse STATIS, version AFC, sur le tableau *Poi*.
 A - Valeurs propres : le compromis est bidimensionnel. B - Carte factorielle des espèces : le triangle à trois groupes est confirmé. C - Projections des axes 1 des analyses séparées : le faisceau est sensiblement plus resserré qu'en ACP (figure 16) et témoigne d'une meilleure appréciation de la notion de compromis. D - Cartes des relevés par variables : l'interprétation précédente n'est pas modifiée, mais la lecture est plus claire et très proche de celle de la représentation triangulaire.

On travaille de cette manière sur la distribution d'abondance de chaque taxon entre les relevés d'un même bloc. La cohérence est stricte avec la figure 17. Les résultats obtenus sont assez voisins des précédents en insistant sur la notion de profils d'espèces et en ignorant les questions d'abondance (un point vide ne contient plus d'information alors qu'il est actif en ACP).

k	n _k	Weights	NS norm2	Cos2	Weights	NS norm2	Cos2
1	16	2.395e-01	3.913e-01	0.295	3.164e-01	3.966e-01	0.130
2	6	2.389e-01	6.379e-01	0.203	1.713e-01	1.187e-01	0.240
3	5	1.295e-01	2.473e-01	0.130	7.489e-02	3.140e-02	0.146
4	8	8.540e-02	7.001e-02	0.221	1.805e-01	1.747e-01	0.152
5	5	4.227e-02	4.474e-02	0.080	7.635e-03	1.287e-03	0.015
6	4	8.583e-02	9.186e-02	0.182	8.271e-02	6.211e-02	0.087
7	17	7.238e-01	1.707e+00	0.813	4.331e-01	3.387e-01	0.535
8	5	2.393e-01	4.952e-01	0.312	1.394e-01	4.424e-02	0.512
9	9	2.198e-01	3.806e-01	0.329	3.009e-01	2.036e-01	0.430
10	8	2.619e-01	2.865e-01	0.602	6.178e-01	5.428e-01	0.851
11	3	1.139e-01	1.637e-01	0.191	8.602e-02	2.945e-02	0.305
12	6	3.761e-01	7.168e-01	0.473	3.727e-01	3.043e-01	0.461

Tableau 7: paramètres de description des structures des tableaux : en caractères droits, option analyse des correspondances, en italique, pour mémoire, option analyse en composantes principales.

La première différence concerne quelques coefficients Rv entre tableaux (tableau 6). Le tableau 7 permet de comparer les paramètres fondamentaux. Les changements sont globalement mineurs. La Loue n'est plus la station de référence mais le cosinus carré est maximum pour la Cuisance, ce qui est assez normal. La carte des espèces à trois pôles (Truite/Chabot, Vairon-Loche, le reste) est confirmée (figure 18). Les cartes des relevés sont parfaitement en accord avec la représentation triangulaire (modèle de reciprocal averaging), tout en étant centrées comme dans autant d'analyse des correspondances séparées. Les deux types de structures indiqués sur la figure 14 se retrouvent à l'identique dans le compromis AFC.

De quelque manière qu'on s'y prenne, on enregistre des structures faunistiques intra-rivières reproductibles, largement moins cependant que l'évolution longitudinale des paramètres géomorphologiques. C'est le couple Vairon-Loche qui introduit une variabilité, par son rôle plus ou moins grand. Il n'y a pas de raisons sérieuses d'évoquer dans ce système une zone à Ombre (il y a des rivières à Ombre) mais plutôt une zone à Vairon-Loche, parfois bien marquée (Furieuse, Cuisance), parfois imperceptible entre la zone à Truite-Chabot et la zone à Cyprinidés (Loue, Clauge).

5 — Paramètres physico-chimiques

Le compromis de structure est attendu et parfaitement exprimé pour les paramètres géomorphologiques. Il s'exprime, avec une certaine complexité pour la composition des peuplements. Qu'en est-il pour les paramètres environnementaux du dernier groupe ? Le tableau Mil comporte les mêmes 92 stations réparties entre 12 rivières et 13 paramètres.

Vme	Mean:	2.99	Variance:	5.88
Tmm	Mean:	4.01	Variance:	2.34
Con	Mean:	4.60	Variance:	2.57
pH	Mean:	5.48	Variance:	1.4
Dur	Mean:	6.15	Variance:	3.76
Cl -	Mean:	1.92	Variance:	4.55

S04-	Mean:	0.92	Variance:	4.14
P04-	Mean:	5.78	Variance:	2.97
N03-	Mean:	4.35	Variance:	3.68
N02/NH4	Mean:	5.14	Variance:	8.08
02%	Mean:	7.75	Variance:	4.25
OXY	Mean:	1.03	Variance:	1.31
DB0	Mean:	1.75	Variance:	1.77

Les variances sont d'ordre de grandeur comparable. On notera, par ailleurs, qu'il y a 54 % de variabilité inter-rivières et 46 % de variabilité intra. Les données sont centrées par variables et par rivières.

```

-----
[ 1] 1000
[ 2] 285 1000
[ 3] 242 255 1000
[ 4] 341 256 33 1000
[ 5] 379 425 172 216 1000
[ 6] 377 210 98 64 244 1000
[ 7] 569 280 143 201 328 560 1000
[ 8] 349 355 146 128 211 299 595 1000
[ 9] 242 192 95 319 104 66 252 166 1000
[10] 462 478 94 523 374 388 547 476 417 1000
[11] 344 457 263 121 118 63 110 102 45 93 1000
[12] 406 476 109 114 277 91 179 79 142 134 337 1000
-----

```

k	nk	Weights	NS norm2	Cos2	
1	16	1.599e-01	1.405e+00	0.119	
2	6	6.240e-02	3.997e-01	0.074	
3	5	9.258e-03	9.009e-02	0.003	
4	8	7.122e-01	7.901e+00	<u>0.676</u>	Allaine
5	5	2.369e-02	1.025e-01	0.032	
6	4	2.590e-02	4.435e-01	0.004	
7	17	8.890e-02	7.328e-01	0.062	
8	5	1.699e-02	6.330e-02	0.021	
9	9	6.639e-01	7.611e+00	<u>0.618</u>	Furieuse
10	8	7.710e-02	1.673e-01	<u>0.342</u>	Cuisance
11	3	2.803e-02	4.044e-01	0.011	
12	6	7.890e-02	1.426e+00	0.016	

Tableau 8 Coefficients de corrélation vectorielle (RV) et paramètres de l'analyse STATIS sur le tableau Mil

Les coefficients Rv sont souvent beaucoup plus faibles que précédemment. Les cosinus carrés sont négligeables pour la majorité des rivières. Il faut immédiatement abandonner l'idée de reproductibilité de structure d'évolution des paramètres environnementaux d'une rivière à l'autre.

La projection des axes 1 sur le plan du compromis (figure 19) confirme que le compromis ne concernent que trois sous-tableaux. Pourtant la carte des variables (figure 20) n'est pas étonnante. Les cartes factorielles des stations sont dans la figure 21 et les cartes des observations centrées par rivière sont dans la figure 22. L'interprétation souligne la faiblesse des co-structures intra-rivières sur l'ensemble des variables. La redondance ayant été limitée au maximum, les paramètres présentent des matrices de covariances peu stables. On notera la mise en évidence des points de pollution (stations 8 et 9 sur le Doubs, 51 et 54 sur l'Allaine), l'augmentation de la charge minérale sur la Furieuse, la réoxygénation sur la Clauge, une homogénéité certaine de plusieurs rivières.

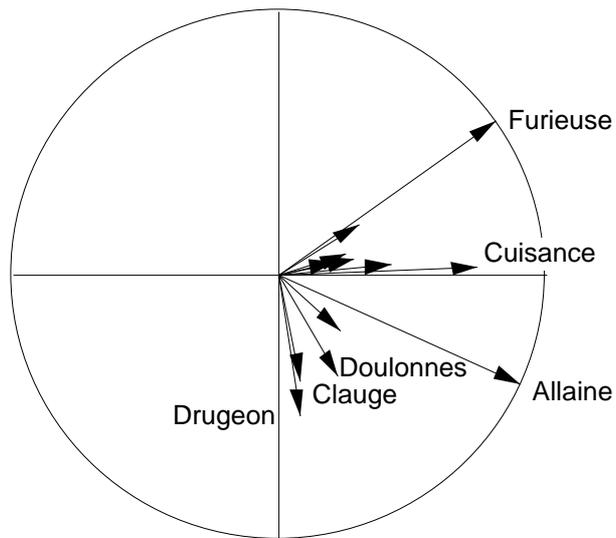


Figure 19 : Projection des axes 1 sur le plan du compromis du tableau Mil.

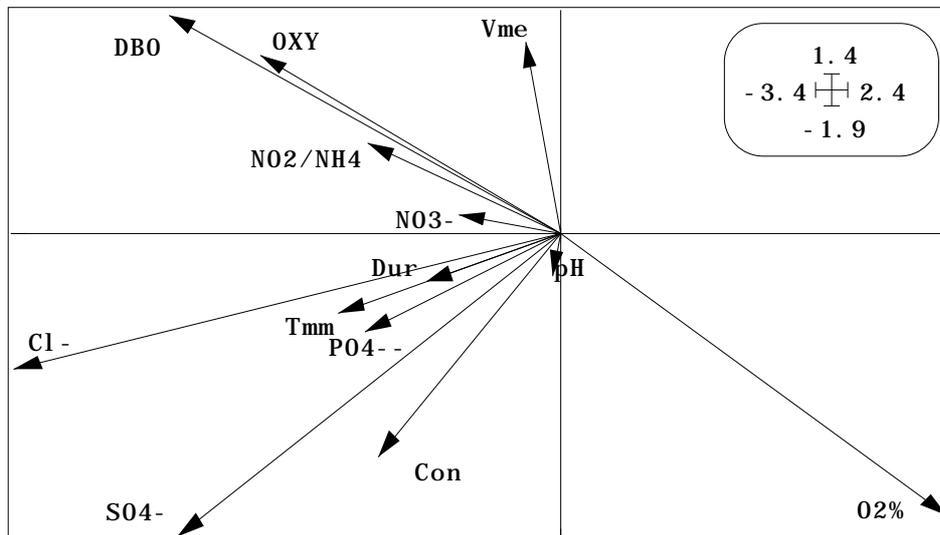


Figure 20 : Coordonnées des variables du tableau Mil. On reconnaît un gradient amont-aval (vers la gauche) de croissance de la charge minérale (en bas) et organique (en haut) de l'eau.

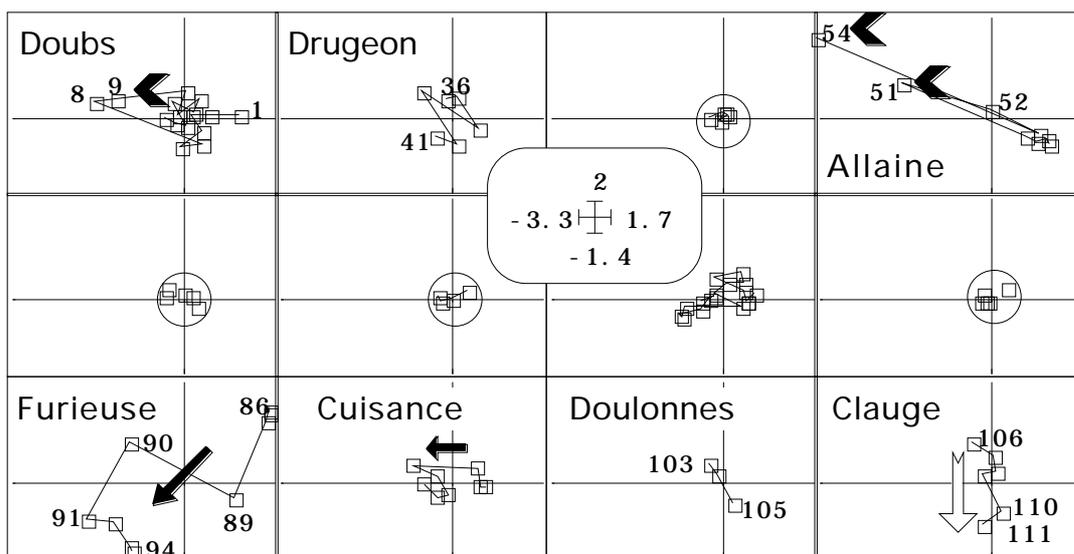


Figure 21: Projection des stations sur le plan du compromis du tableau Mil.

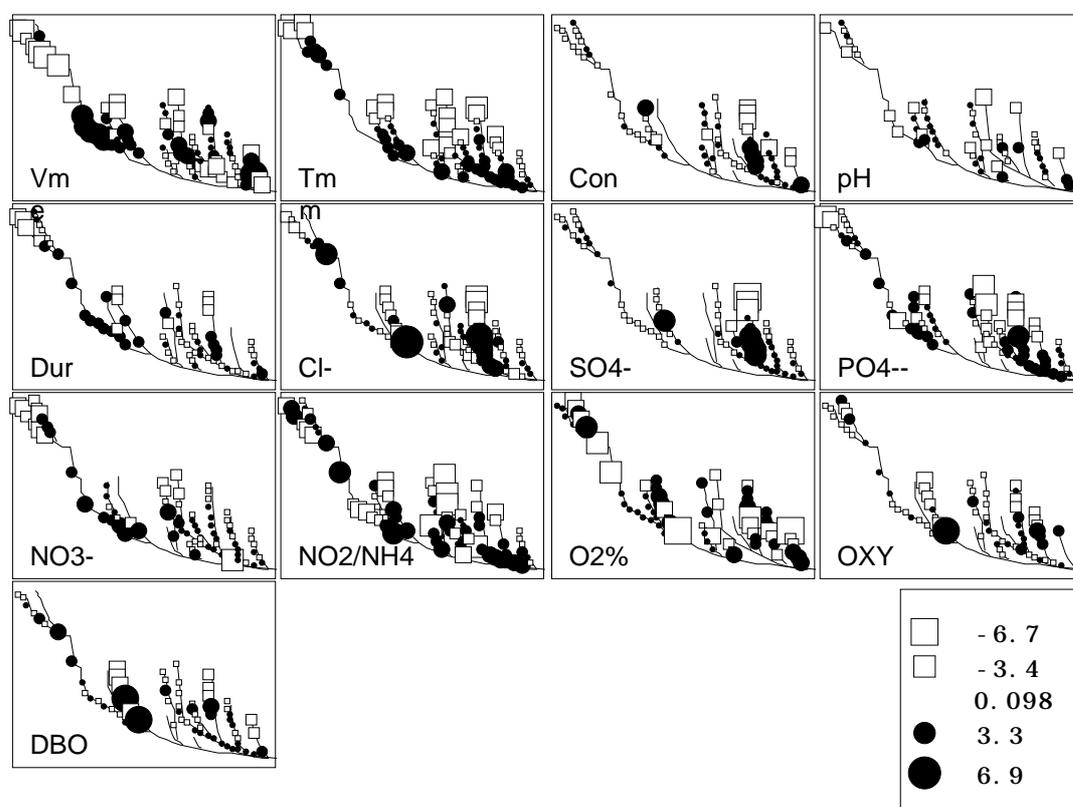


Figure 22 : Cartes des valeurs centrées par rivières des variables de Mil.

On peut s'étonner qu'une évolution aussi reproductible entre rivières que l'augmentation de la température dans le gradient amont-aval ne joue aucun rôle. C'est l'absence de covariances autour de cette variable qui l'explique : STATIS ne travaille que sur les opérateurs d'ACP par tableau. C'est pourquoi la covariation des chlorures et des sulfates le long de la Furieuse compte (et cette covariation est propre à cette rivière), alors que la variation interne de la température le long de chaque rivière n'est pas un fait marquant pour ce type d'analyse. STATIS ne s'occupe que de la reproductibilité des matrices de covariances. Les matrices de covariance entre abondance des taxons ont une bonne part de ressemblance, d'une rivière à l'autre. Les matrices de covariances entre

paramètres sont peu stables (et ce n'est pas une question de nombre de stations, sinon l'argument serait adapté aux deux ensembles de variables).

Si donc l'évolution des paramètres environnementaux joue un rôle dans l'établissement des structures des peuplements, ce n'est sans doute pas par le jeu d'une structure reproductible de l'évolution du milieu multivarié (structure qui n'existe pas) mais par le jeu de une ou plusieurs variables qui ont des rôles particuliers dans une ou plusieurs situations. La typologie des peuplements est reproductible, avec des variantes, d'une rivière à l'autre. Existe-t-il des variables explicatives de cette typologie, voire de ses variantes ? L'explication géomorphologique simple (la règle des pentes) ne peut suffire (un gradient simple ne peut expliquer la présence de ces variantes faunistiques). Les autres paramètres soit sont stables, ce qui renvoie à une explication purement biologique (River Continuum Concept), soit varient de manière peu reproductible. Il serait intéressant alors d'étendre STATIS de la notion de stabilité d'une structure à celle de stabilité (ou de variabilité) d'une co-structure.

Annexe : Mettre en œuvre un STATIS version AFC

Comme dans toutes ses options ADE-4 propose des modules paramétrables, ce qui signifie qu'une analyse d'un type défini peut s'exécuter sur tous les types de variables. Ceci est vrai pour le module STATIS. Dans les travaux cités, STATIS se veut une méthode de type ACP, qui manipule donc des variables quantitatives. Il n'est pas compliqué d'utiliser une version AFC, suggérée au paragraphe 4. Employer la carte Oiseaux de la pile ADE-4•Data⁵ :

The screenshot shows a window titled 'tableau faunistique multi-groupes (23-8+6+5)'. It contains a data matrix with 4 rows and 4 columns of numerical values. To the right of the matrix is a legend with four entries: '1-Fauvette orphée', '2-Fauvette des jardins', '3-Fauvette à tête noire', and '4-Fauvette babillarde'. Below the legend is a list of numbers: 0, 2, 2, 0, 2, 2, 0, 2, 2, 0, 3, 2, 0, 3, 2, 0, 2, 2, 0, 2, 3. On the far right, there are three small boxes containing the numbers 8, 6, and 5.

Récupérer les fichiers AviAtlas (23 lignes-districts, 19 colonnes-espèces), AviAtlasnum (3 lignes, nombre d'espèces par bloc) et Label_Esp. Préparer l'ACM floue de AviAtlas (FuzzyVar : Read Fuzzy File) :

The screenshot shows a dialog box titled 'Read Fuzzy File'. It has three input fields: 'Fuzzy variables: input file (---)' with the value 'AviAtlas', 'Category indication file' with the value 'AviAtlasnum', and 'Output file name (default = ---F)' which is empty. To the right of the first two fields are the numbers '23 19' and '3 1' respectively. There are 'Quit' and 'Ok' buttons at the bottom.

Input file: AviAtlas
 Row number: 23, column number: 19
 Number of modalities for each variable: file AviAtlasnum
 Row number: 3
 Missing data: 0

Variable number 1 has 8 categories

[1] Category: 1 Freq.: 0.0434

```
[ 2] Category: 2 Freq.: 0.305
...
```

Variable number 2 has 6 categories

```
[ 9] Category: 1 Freq.: 0.0755
...
```

Variable number 3 has 5 categories

```
[ 15] Category: 1 Freq.: 0.263
...
```

Exécuter l'analyse sans se préoccuper du résultat (MCA : Fuzzy Correspondence Analysis) :



Passer alors du niveau analyses simples au niveau K-tableaux par l'option KTabUtil : FuzzyToKTab :



Exécuter enfin STATIS sur le nouveau multi-tableau :



Option operator averaging
 Input file AviAtlasF
 Number of rows: 19, columns: 23

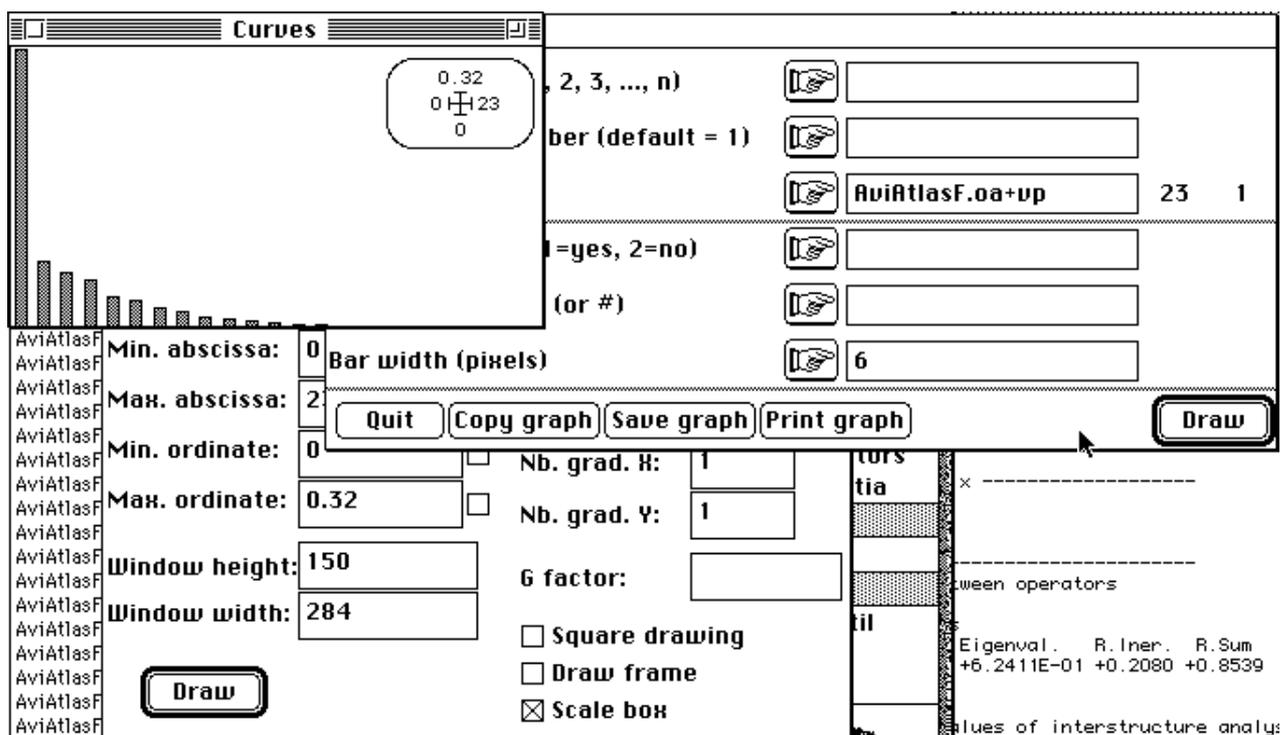
Les coefficients RV sont modérés et équilibrés :

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 553 1000
[ 3] 392 457 1000
-----
```

File AviAtlasF.oa+RV contains cosinus between operators
 It has 3 rows and 3 columns

Compromise with cosinus between operators

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1.9377E+00	+0.6459	+0.6459	02	+6.2411E-01	+0.2080	+0.8539
03	+4.3817E-01	+0.1461	+1.0000				



File AviAtlasF. oa+EV contains the eigenvalues of interstructure analysis

It has 3 rows and 1 columns

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+3.1912E-01	+0.4820	+0.4820	02	+7.6552E-02	+0.1156	+0.5977
03	+6.3645E-02	+0.0961	+0.6938	04	+5.5008E-02	+0.0831	+0.7769
...							
23	+0.0000E+00	+0.0000	+1.0000				

On peut hésiter pour ne garder qu'un seul facteur :

File AviAtlasF. oa+vp contains the eigenvalues of compromise analysis

It has 23 rows and 1 columns

Remarquer que les tableaux juxtaposés (lignes-relevés en commun) dans l'analyse floue sont superposés (colonnes-relevés en commun) dans le multi-tableau.

File AviAtlasF. oa+co contains column scores (analysis of the compromise)

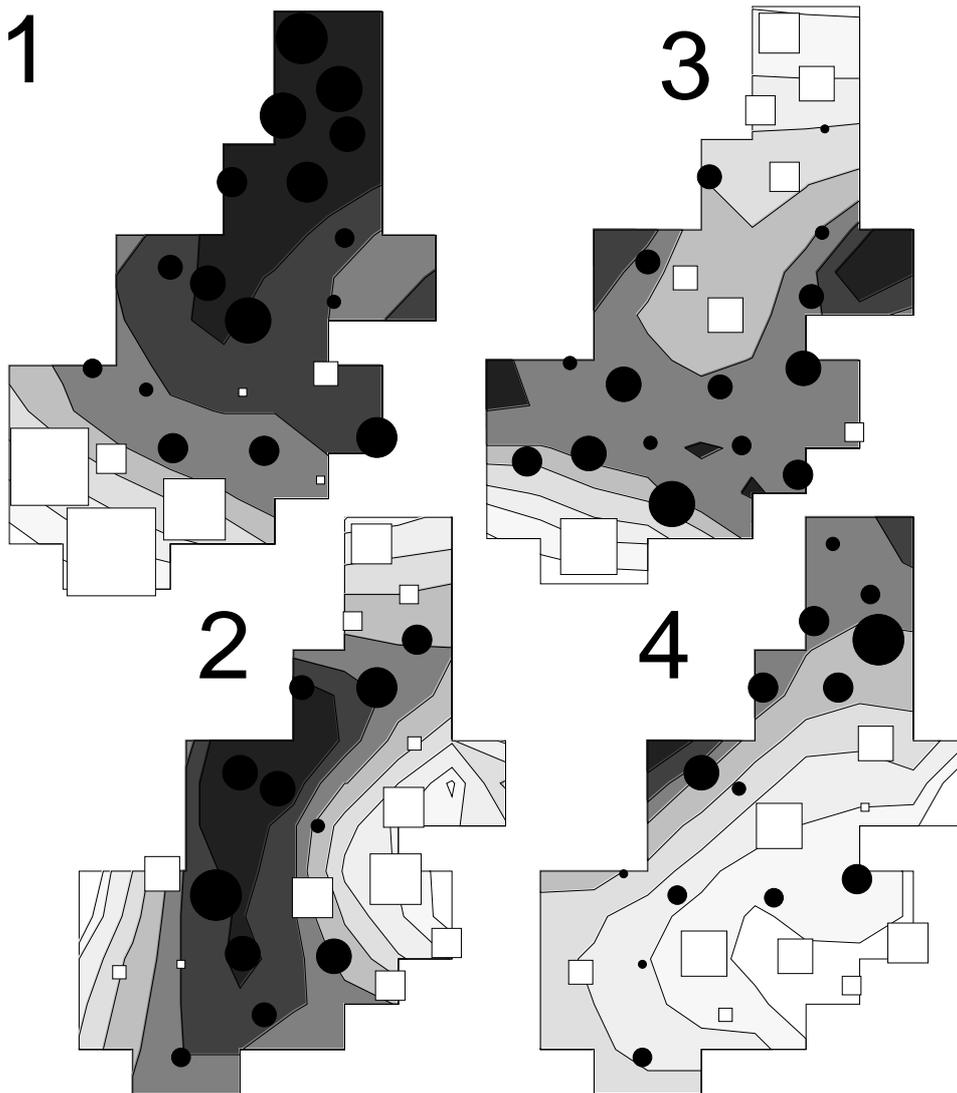
It has 23 rows and 4 columns

File : AviAtlasF. oa+co

----- Mini mum/Maxi mum:

Col. :	1	Mini = -1.8474	Maxi = 0.63044
Col. :	2	Mini = -0.5922	Maxi = 0.56951
Col. :	3	Mini = -0.72252	Maxi = 0.43554
Col. :	4	Mini = -0.50149	Maxi = 0.53995

Les coordonnées du compromis portent donc sur les districts. Récupérer les fichiers Atlas_Digi et Atlas_Carto par l'option Copy Files du bouton Data Folder. Utiliser Digit pour enregistrer les coordonnées des districts et préparer une grille pour courbe de niveaux. Tracer ces courbes de niveaux avec Levels et représenter les valeurs avec Maps, option values. Superposer pour valider. Les coordonnées présentent des structures spatiales fortes qui imposent de les prendre en considération :



File AviAtlasF. oaTl1 contains standard column scores with lambda norm

It has 19 rows and 4 columns

It is to be used with --Tl1.label and --Tl1.cat files

File : AviAtlasF. oaTl1

----- Minimum/Maximum:

Col.: 1 Mini = -3.2703 Maxi = 0.38017 ...

File AviAtlasF. oaT4a contains scores of axes of separate analysis

It has 12 rows and 4 columns

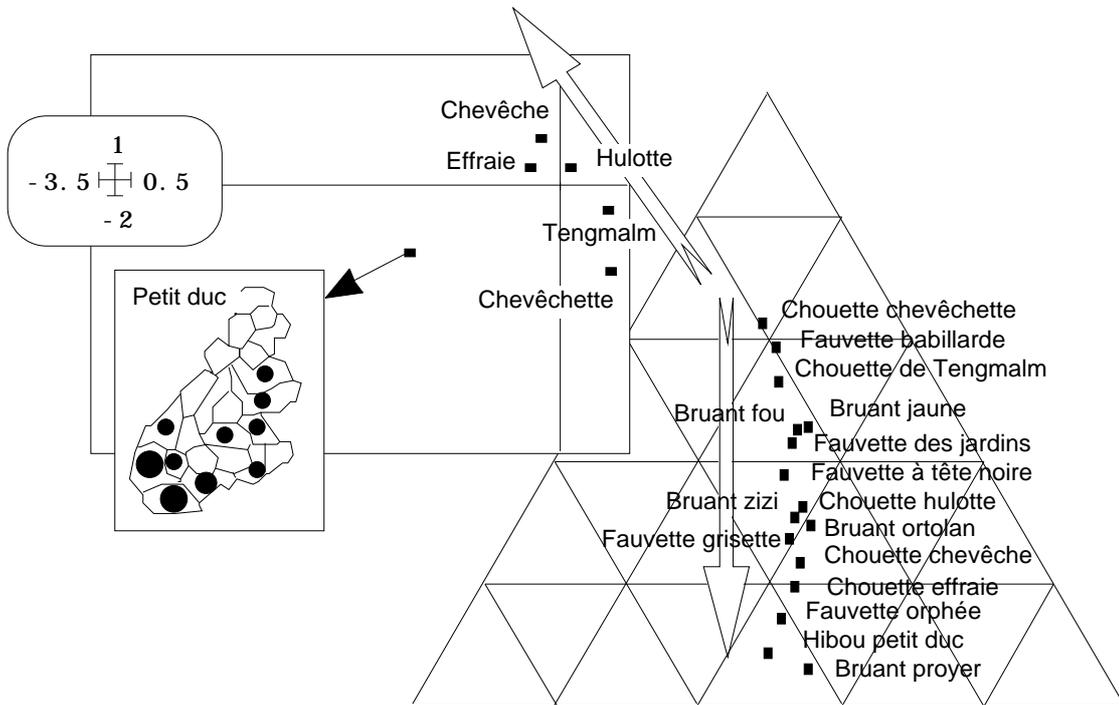
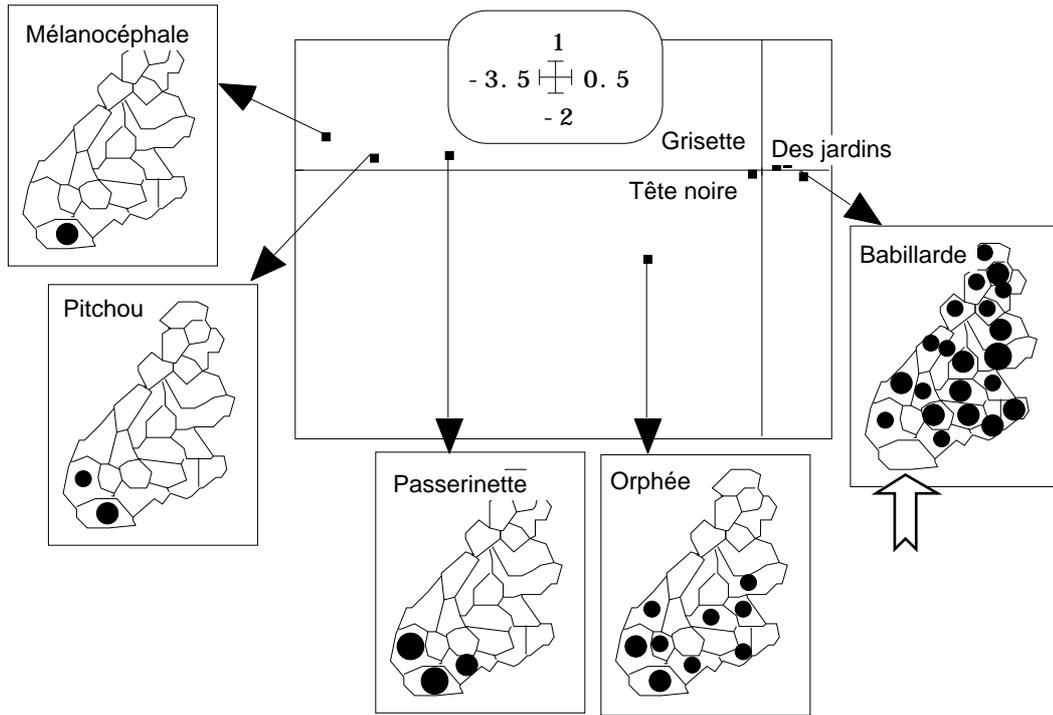
It is to be used with --T4a.label and --T4a.cat files

File : AviAtlasF. oaT4a

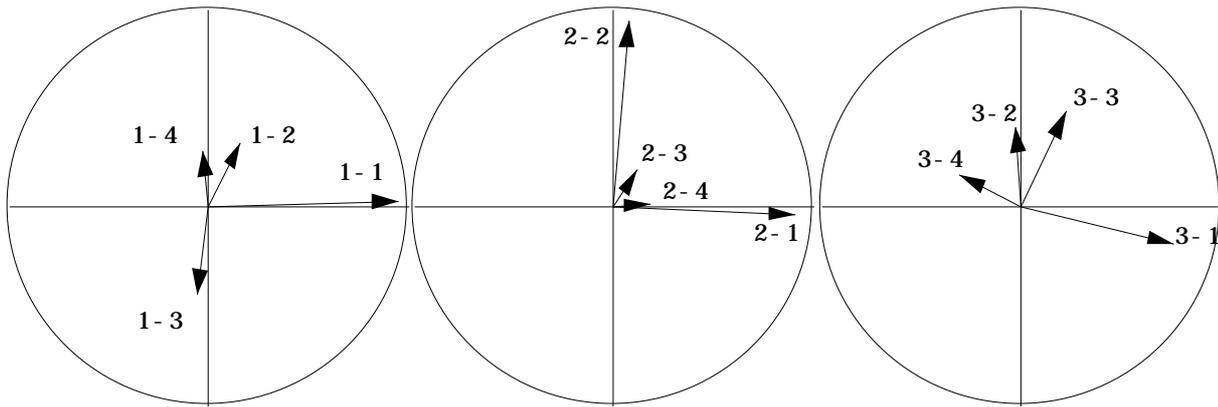
----- Minimum/Maximum:

Col.: 1 Mini = -0.308 Maxi = 0.95611 ...

Utiliser Scatters pour obtenir les cartes factorielles par tableau. On obtient typiquement des ordinations de type AFC (gradient, séparation de niche, privilège accordé aux espèces peu fréquentes). La carte des Fauvettes est très clairement cohérente avec la carte du premier facteur colonne :



Si l'axe 1 décrit l'influence méditerranéenne, l'axe 2 retrouve le gradient d'altitude sur lequel s'ordonne les rapaces nocturnes (ci-dessus, consulter la fiche sur la représentation triangulaire). Utiliser Scatters pour projeter les axes par tableau. Les axes 1 sont sensiblement communs ce qui justifie la taille de la première valeur propre. Pour les suivants, on n'observe rien de tel.



La cohérence de toutes les aides à l'interprétation souligne que STATIS additionne les structures en moyennant les opérateurs d'inertie. Les structures présentes dans plusieurs tableaux sont amplifiées. Sont ensuite décrites les typologies plus particulière et la méthode permet de noter la réaction ou l'absence de réaction de chaque groupe sur chaque gradient.

File AviAtlasF.oaTCrep contains column scores (analysis of the compromise)

Duplication of general scores

It has 69 rows and 4 columns

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

File : AviAtlasF.oaTCrep

----- Minimum/Maximum:

Col.: 1 Mini = -1.8474 Maxi = 0.63044

Col.: 2 Mini = -0.5922 Maxi = 0.56951

Col.: 3 Mini = -0.72252 Maxi = 0.43554

Col.: 4 Mini = -0.50149 Maxi = 0.53995

File AviAtlasF.oaTCest contains columns scores issued from each table by

It has 69 rows and 4 columns

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

File : AviAtlasF.oaTCest

----- Minimum/Maximum:

Col.: 1 Mini = -1.8474 Maxi = 1.0704

Col.: 2 Mini = -0.60065 Maxi = 0.58074

Col.: 3 Mini = -0.72252 Maxi = 0.43554

Col.: 4 Mini = -0.44431 Maxi = 0.48496

File AviAtlasF.oaTCtra contains columns scores issued from each table by

Modified Statist trajectories

It has 69 rows and 4 columns

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

File : AviAtlasF.oaTCtra

----- Minimum/Maximum:

Col.: 1 Mini = -2.1629 Maxi = 0.73034

Col.: 2 Mini = -0.70733 Maxi = 0.67073

Col.: 3 Mini = -1.8717 Maxi = 0.80764

Col.: 4 Mini = -0.74712 Maxi = 0.578

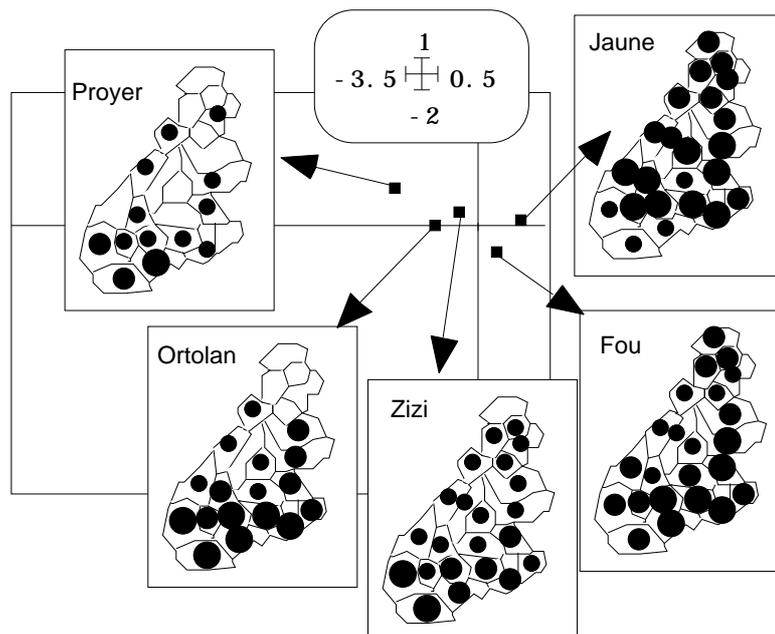
Les fichiers qui contiennent les coordonnées des relevés propres à chaque tableau sont de peu d'intérêt. Ce n'est pas l'objet de STATIS que de faire de la typologie de relevés avec les différents groupes faunistiques. Son objet est de faire de la typologie de taxons coordonnées par une même typologie de relevés. A ce sujet, on fera le contrôle suivant. Transposer le fichier AviAtlasF (tableau flou en pourcentage par bloc) en ATR

(19 espèces-lignes et 23 relevés colonnes). Calculer par le module Basic Row%Mean la position moyenne des espèces (en lignes dans ATR) sur les codes numériques en colonnes dans AviAtlasF.0a+co. Vérifier par Curves que ces moyennes sont proportionnelles aux coordonnées qui sont dans AviAtlasF.0aTLI. Les espèces de chaque groupe sont (à une constante près) à la moyenne des relevés. La dissymétrie introduite par la structure multi-tableau oblige à perdre le double averaging pour obtenir un averaging simultané optimal. STATIS version AFC est une “demi-AFC” (relevés -> espèces). On ne peut pas tout avoir, même en statistique.

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	8	5.833e-01	1.261e-01	0.797
2	6	6.071e-01	4.190e-02	0.638
3	5	5.397e-01	2.566e-02	0.400

Des trois tableaux, celui des Bruants est le moins structuré. On y retrouve cependant le gradient Nord-Sud commun au trois groupes.



STATIS a bien exécuté trois analyses des correspondances coordonnées.

Remerciements : Merci à E. Baran, M. Bournaud et S. Champely de leurs commentaires fructueux sur la première version de cette fiche.

Références

¹ Huet, M. (1949) Appréciation de la valeur piscicole des eaux douces. *Travaux de la station de Recherches de Groenendaal*, série D : 10, 1-52.

Huet, M. (1954) Biologie, profils en long et en travers des eaux courantes. *Bulletin français de pisciculture* : 175, 41-53.

Huet, M. (1959) Profiles and biology of western european streams as related to fish management. *Transactions of the American Fisheries Society* : 88, 3, 166-163.

² Wartenberg, D., Ferson, S. & Ohlf, F.J. (1987) Putting things in order : a critique of detrended correspondence analysis. *The American Naturalist* : 129, 3, 434-448.

Peet, R.K., Knox, R.G., Case, J.S. & Allen, R.B. (1988) Putting things in order : the advantages of detrended correspondence analysis. *The American Naturalist* : 131:924-934.

³ Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.

⁴ Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.

⁵ Lebreton, Ph. (1977) *Les oiseaux nicheurs rhônalpins. Atlas ornithologique Rhône-Alpes*. Centre Ornithologique Rhône-Alpes, Université Lyon 1, 69621 Villeurbanne. Direction de la Protection de la Nature, Ministère de la Qualité de la Vie. 1-354.