

# Une introduction à l'Analyse en Composantes Principales avec *SPSS pour Windows*

*Dominique DESBOIS*

*INRA-ESR Nancy et SCEES*

*4 avenue de Saint-Mandé, 75570 Paris Cedex 12.*

*Fax : +33 1 49 55 85 11      Mel : desbois@jouy.inra.fr*

## **RÉSUMÉ :**

Cette note initie l'utilisateur débutant à la mise en oeuvre de l'Analyse en Composantes Principales au moyen de la procédure d'Analyse Factorielle **FACTOR** du logiciel *SPSS pour Windows*. Cette mise en oeuvre concerne l'analyse multidimensionnelle des tableaux de données numériques quantitatives. Le listage des résultats obtenus est commenté par la présentation du formulaire de l'Analyse en Composantes Principales associé à chacun des résultats obtenus.

**MOTS-CLÉS :** Analyse en Composantes Principales, Analyse Factorielle, logiciel statistique, mise en oeuvre.

## **1. Introduction**

L'Analyse en Composantes Principales permet d'analyser des tableaux de données numériques quantitatives pour en réduire la dimensionnalité aux principaux facteurs d'interaction entre variables et en représenter graphiquement les interrelations. La mise en oeuvre d'une Analyse en Composantes principales (**ACP**) peut être effectuée au moyen de la procédure d'Analyse Factorielle de *SPSS* (**FACTOR**).

## **2. Un exemple simple de mise en oeuvre**

### **2.1. Les données**

Les données sont constituées par un tableau récapitulant la composition chimique d'un certain nombre d'eaux minérales classées par pays (cf. *Tomassone, Dervin & Masson 1993*). Le fichier *SPSS* **eauminer.sav** comporte les variables suivantes : la dénomination d'origine de la source (variable **origine**), un sigle d'identification de trois caractères (variable **sigle**), le pays d'origine (variable **p**), la composition en chlore (variable **cl**), en calcium (variable **ca**), en manganèse (variable **mg**), en nitrate (variable **na**), en soufre (variable **so4**) et en gaz carbonique (variable **hco3**).

The screenshot shows the SPSS 'Data View' window for a file named 'c:\dd\spss\acp\eauminer.sav'. The table contains the following data:

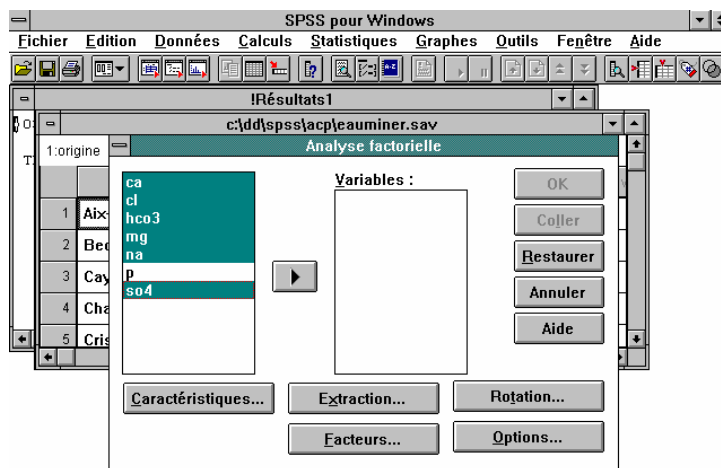
origine	sigle	hco3	so4	cl	ca	mg	na	p	f
1 Aix-les-Bains [F-73]	Aix	341	27	3	84	23	2	1	
2 Beckerish [F-68]	Bec	263	23	9	91	5	3	1	
3 Cayranne [F-84]	Cay	287	3	5	44	24	23	1	
4 Chambon [F-45]	Cha	298	9	23	96	6	11	1	
5 Cristal-Roc [F-72]	Cri	200	15	8	70	2	4	1	
6 Saint-Cyr [F-45]	Cyr	250	5	20	71	6	11	1	
7 Evian [F-74]	Evi	357	10	2	78	24	5	1	
8 Ferita [I]	Fer	311	14	18	73	18	13	2	
9 Saint-Hippolyte [F-3]	Hin	256	6	23	86	3	18	1	

## 2.2. La spécification des paramètres de l'analyse

### 2.2.1. Définition des variables actives de l'analyse

Afin d'afficher la boîte de dialogue principale de la procédure FACTOR, sélectionnez à partir du menu principal les options suivantes :

Statistiques  
Factorisation  
Analyse factorielle ...

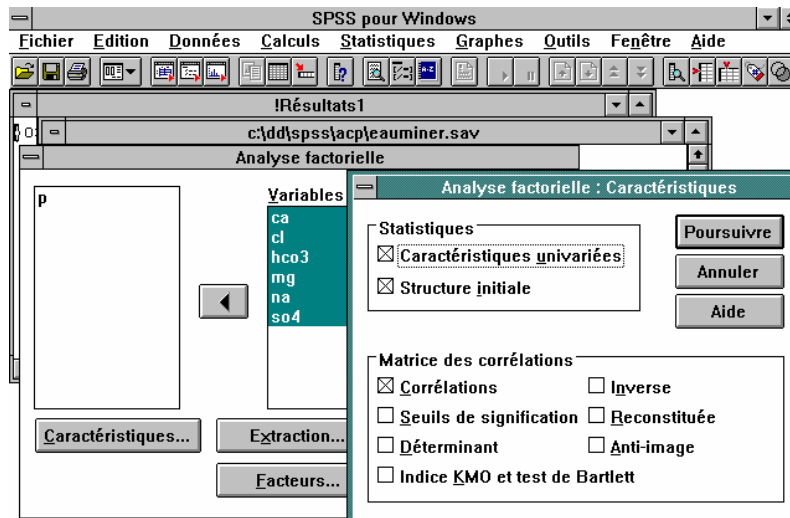


### 2.2.2. Sélection des variables

Sélectionner les variables numériques choisies pour l'ACP (minimum : 2 variables) parmi celles figurant dans la liste source en les transférant dans la liste des **Variables** à l'aide du bouton. Il suffit alors de cliquer sur le bouton **OK** pour effectuer une analyse factorielle avec les paramètres prévus par défaut. On obtient alors le listage de la structure initiale, la matrice des corrélations variables-facteurs et les statistiques concernant la structure finale.

### 2.2.3. Statistiques descriptives

Afin de choisir les statistiques optionnelles de la procédure FACTOR, cliquez sur le bouton **Caractéristiques ...** pour ouvrir la boîte de dialogue secondaire permettant d'effectuer ces choix.



### Statistiques

Vous pouvez choisir l'une ou plusieurs des statistiques suivantes :

- **Caractéristiques univariées.** Affichage du nombre d'observations valide, de la moyenne et de l'écart-type pour chaque variable.
- **Structure initiale.** Communautés de la solution initiale, valeurs propres et pourcentage d'inertie expliquée. Il s'agit de la statistique descriptive choisie par défaut.

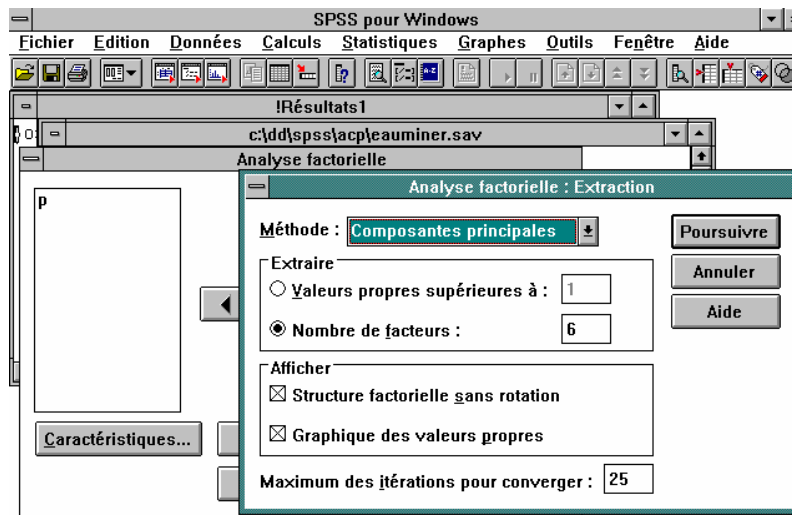
### Matrice des corrélations

Vous pouvez choisir l'un ou plusieurs des indicateurs statistiques suivants :

- **Corrélations.** Matrice des coefficients de corrélation pour les variables actives.
- **Seuils de signification.** Seuils unilatères de signification des coefficients de corrélations.
- **Déterminant.** Déterminant de la matrice des corrélations.
- **Indice KMO et test de Bartlett.** Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage et test de sphéricité de Bartlett.
- **Inverse.** Inverse de la matrice des corrélations.
- **Reconstituée.** Matrice des coefficients de corrélations reconstitués et leurs résidus. Les coefficients de corrélation sont affichés en dessous de la diagonale tandis que les résidus sont situés au-dessus.
- **Anti-image.** Anti-images des matrices de corrélation et de variance-covariance. La mesure de la qualité de l'échantillonnage pour chaque variable est affichée sur la diagonale de l'anti-image de la matrice des corrélations.

## 2.2.4. Extraction des facteurs

Afin de choisir une méthode d'extraction des facteurs, obtenir un histogramme des valeurs propres ou contrôler le nombre de facteurs à extraire, cliquez sur le bouton **Extraction ...** pour ouvrir la boîte de dialogue secondaire permettant d'effectuer ces choix.



### ↓ Méthode

Vous pouvez choisir une ou plusieurs des méthodes d'extraction suivantes :

**Composantes principales.** Analyse en composantes principales. Il s'agit de la méthode d'extraction par défaut.

**Moindres carrés non-pondérés.** Méthode des moindres carrés ordinaires (MCO).

**Moindres carrés généralisés.** Méthode des moindres carrés généralisés (MCG).

**Maximum de vraisemblance.** Méthode du maximum de vraisemblance (EMV).

**Factorisation en axes principaux.** Méthode de la factorisation en axes principaux.

**Alpha-maximisation.** Méthode d'alpha-maximisation.

**Factorisation en projections.** Méthode de la factorisation en projections.

### Extraire

Vous pouvez choisir l'un des critères d'extraction suivants :

- **Valeurs propres supérieures à.** Dans l'option par défaut, les facteurs correspondant aux valeurs propres supérieures à 1 sont extraits. Pour obtenir un nombre de facteurs différents, modifiez cette valeur par un nombre compris entre 0 et le nombre total de variables actives.
- **Nombre de facteurs.** Permet d'extraire un nombre de facteurs spécifié. Entrez un entier positif.

### Afficher

Vous pouvez choisir une ou plusieurs options d'affichage :

- **Structure factorielle sans rotation.** Coordonnées factorielles (matrice des corrélations variables-facteurs), communautés et valeurs propres pour la structure factorielle. Option par défaut.

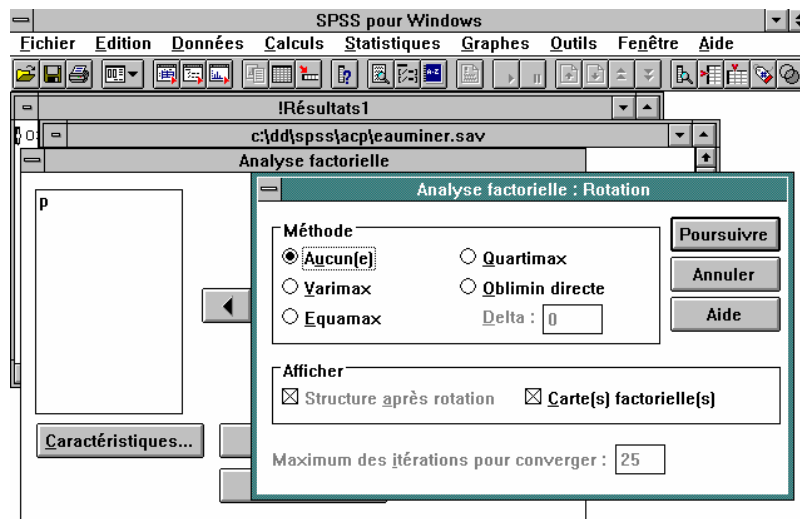
- **Graphique des valeurs propres.** Histogramme des valeurs propres triées par ordre décroissant. Le graphique affiche les facteurs après rotation si une rotation a été demandé (cf. section « Rotation des facteurs » ci-après).

L'option suivante est également offerte :

**Maximum des itérations pour converger.** Le nombre maximum d'itérations pour que la procédure d'extraction converge est fixé par défaut à **25**. Pour fixer une limite de convergence différente, entrez un entier positif. La valeur du critère de convergence pour l'extraction est égale à **0.001**.

### 2.2.5. Rotation des facteurs

Afin de sélectionner une procédure de rotation des facteurs, cliquez sur le bouton **Rotation ...** pour ouvrir la boîte de dialogue secondaire permettant d'effectuer ces choix.



**Méthode.** Aucune rotation n'est possible s'il n'y a qu'un seul facteur extrait. La standardisation de Kaiser est utilisée avec n'importe laquelle des méthodes de rotation. Vous pouvez choisir l'une des méthodes de rotation suivantes :

- **Aucune.** Aucune rotation n'est effectuée. C'est l'option par défaut.
- **Varimax.** Rotation orthogonale selon la méthode Varimax.
- **Equamax.** Rotation orthogonale selon la méthode Equamax.
- **Quartimax.** Rotation orthogonale selon la méthode Quartimax.
- **Oblimin directe.** Rotation oblique selon la méthode Oblimin.
- **Delta.** Pour modifier la valeur par défaut **0** du delta, entrez un nombre inférieur ou égal à 0,8 (**0.8**).

#### Afficher

Vous pouvez choisir une ou plusieurs options d'affichage :

- **Structure après rotation.** Affichage par défaut dès qu'une rotation est demandée. Pour les rotations orthogonales, les matrices de la configuration après rotation et de passage dans la nouvelle base sont affichées. Pour les rotations obliques, les matrices de configuration, de

structure et de corrélation des facteurs sont affichées. Pour supprimer l'affichage de la structure après rotation, dé-sélectionner cet item..

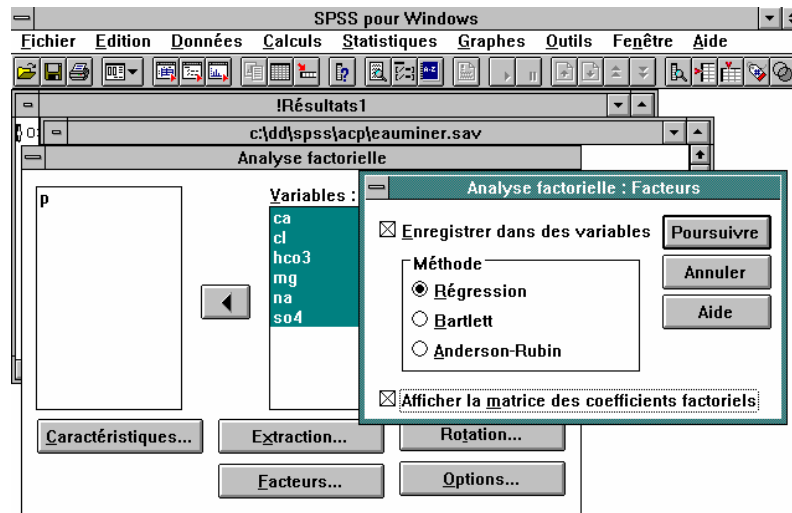
- **Carte(s) factorielle(s)**. Graphique à 3 dimensions pour les trois premiers facteurs. Pour une structure à deux facteurs, un graphique-plan est édité. Le graphique ne s'affiche pas s'il n'y a qu'un seul facteur extrait. Les graphiques affichent les structures après rotation, dès qu'une rotation a été demandée.

L'option suivante est disponible si vous avez demandé une rotation :

**Maximum des itérations pour converger**. Par défaut, un maximum de **25** itérations est prévu pour effectuer une rotation des facteurs. Pour spécifier un maximum différent, entrez un entier positif.

### 2.2.6. Coordonnées factorielles

Afin de sauvegarder les coordonnées factorielles pour les réutiliser dans d'autres analyses, cliquez sur le bouton **Facteurs ...** pour ouvrir la boîte de dialogue secondaire permettant d'effectuer ce choix.



Pour sauvegarder les coordonnées factorielles des individus, sélectionnez le choix **Enregistrer dans des variables**, puis choisissez une méthode de calcul de ces coordonnées factorielles :

- **Enregistrer dans des variables**. Sauvegarde les coordonnées factorielles comme variables du fichier *SPSS* courant. Le listage de l'exécution résume dans un tableau le nom de chacune de ces nouvelles variables ainsi créées et l'étiquette de variable indiquant la méthode de calcul utilisée.

#### Méthode

Ces options permettent de contrôler la méthode de calcul des coordonnées factorielles :

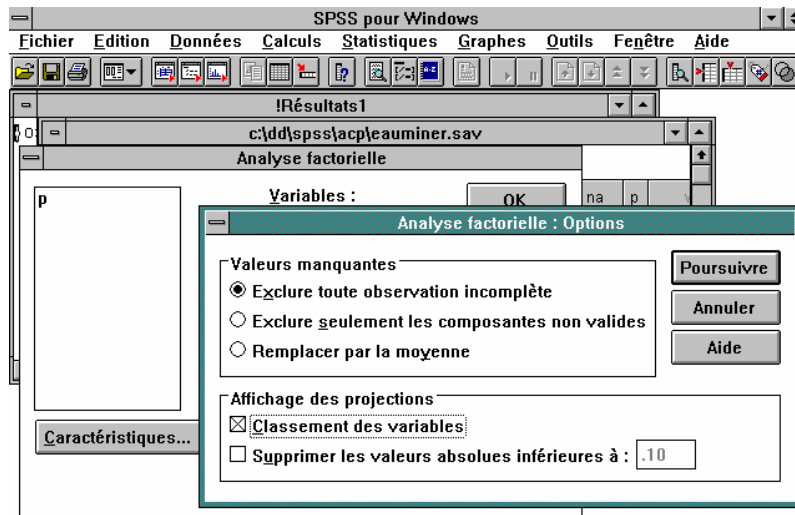
- **Régression**. Méthode de la régression. C'est l'option retenue par défaut.
- **Bartlett**. Méthode de Bartlett.
- **Anderson-Rubin**. Méthode d'Anderson-Rubin.

L'option suivante est également disponible :

- **Afficher la matrice des coefficients factoriels.** Cette option permet d'afficher la matrice des coefficients ainsi que la matrice de variance-covariance des coordonnées factorielles.

### 2.2.7. Options d'analyse

Pour changer le traitement des valeurs manquantes ou permuter l'affichage des matrices de facteurs selon leur ordre d'importance, cliquez sur le bouton **O**ptions ... pour ouvrir la boîte de dialogue secondaire permettant d'effectuer ces choix.



### Valeurs manquantes

Vous pouvez choisir l'une des options suivantes :

- **Exclure toute observation incomplète.** Seules les observations ayant des valeurs valides pour l'ensemble des variables actives sont retenues pour l'analyse.
- **Exclure seulement les composantes non valides.** Les observations sont exclues selon une approche bivariée : dans le calcul d'une corrélation, *SPSS* utilise toute observation ayant des valeurs valides pour les deux variables (même si certaines de ces observations possèdent des valeurs manquantes pour d'autres variables actives).
- **Remplacer par la moyenne.** Remplace les valeurs manquantes par la moyenne de la variable puis retient toutes les observations pour l'analyse factorielle.

### Affichage des projections

Vous pouvez choisir un ou plusieurs des formats d'affichage suivants pour les projections :

**Classement des variables.** Tri des tableaux de structure et de coordonnées factorielles pour que les variables soient triées selon la qualité de leurs projections sur chacun des facteurs.

**Supprimer les valeurs absolues inférieures à.** Supprime les coefficients dont la valeur absolue est inférieure à la valeur spécifiée. La valeur par défaut est **0.1**. Pour modifier cette valeur par défaut, entrez un nombre compris entre 0 et 1.

Vous pouvez combiner ces deux options en classant les variables et en omettant les coefficients à faible valeur absolue.

## 2.3 Spécifications optionnelles du langage de commande

Vous pouvez adapter la procédure FACTOR à votre usage personnel en collant les paramètres de votre sélection effectuée par l'intermédiaire des boîtes de dialogue dans la fenêtre de syntaxe (référez-vous au *Manuel d'utilisation, Système de base* pour plus d'information sur les fenêtres de syntaxe). Vous pourrez alors modifier le langage de commande résultant de ces choix pour :

- écrire une matrice de corrélation ou un tableau de coordonnées factorielles, voire lire une matrice pour l'utiliser à la place des données de base (avec la sous-commande MATRIX).
- spécifier le nombre de caractères des labels de valeur utilisés pour étiqueter les points dans les graphiques (avec la sous-commande PLOT).
- analyser différents ensembles de données au moyen d'une seule commande (avec la sous-commande ANALYSIS), voire procéder à plusieurs extractions et rotations de facteurs pour une même analyse (avec les sous-commandes EXTRACTION et ROTATION).
- sauvegarder un sous-ensemble des coordonnées factorielles ou fournir des noms hiérarchisés pour les variables contenant ces projections (avec la sous-commande SAVE).
- utiliser des critères d'extraction ou de rotation supplémentaires (avec la sous-commande CRITERIA).
- définir des valeurs particulières à l'utilisateur pour la diagonale de l'opérateur d'inertie dans la méthode de factorisation en axes principaux (avec la sous-commande DIAGONAL).

Pour une description complète de la commande FACTOR et des règles de syntaxe, consultez la section correspondante du manuel (*SPSS 6.1 Professional Statistics, Syntax Reference*).



### 3. Les résultats de l'Analyse en Composantes Principales

#### 3.1. Indicateurs statistiques univariés

Une première série de statistiques descriptives élémentaires sur les variables de l'échantillon des eaux minérales<sup>1</sup> peut être obtenue en choisissant l'option **Caractéristiques univariées**.

```

- - - - - F A C T O R   A N A L Y S I S - - - - -
Analysis number 1 Listwise deletion of cases with missing values

Label                Mean                Std-Dev
CA                   77.50000            48.08819
CL                   13.65000            10.58934
HCO3                 250.45000           102.61398
MG                   11.85000            11.09421
NA                   10.10000            7.31905
SO4                  42.40000            77.97125

Number of Cases = 20

```

Cette option nous donne :

- la **moyenne**  $\bar{x} = \sum_{i \in I} \mu_i \times x_i$  comme paramètre de position;
- l'**écart-type**  $\sigma_x = \sqrt{\sum_{i \in I} \mu_i \times (x_i - \bar{x})^2}$  comme paramètre de dispersion;

avec, dans le cas d'un échantillon pondéré par un système de poids  $\{p_i, i \in I\}$ ,  $P = \sum_{i \in I} p_i$ , le

**poids de l'échantillon** et  $\mu_i = \frac{p_i}{P}$  la masse au point définissant la **pondération de l'échantillon**  $\{\mu_i, i \in I\}$ .

#### 3.2. Matrice des corrélations

Le **coefficient de corrélation linéaire** est défini par:

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \times \sigma_y} = \frac{\sum_{i \in I} (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i \in I} (x_i - \bar{x})^2 \times \sum_{i \in I} (y_i - \bar{y})^2}}$$

<sup>1</sup> Par la suite, on notera  $I$ , l'ensemble des individus (e.g. les eaux minérales) et  $J$  l'ensemble des variables (e.g. les teneurs des eaux minérales en différents composés chimiques).

Les valeurs du coefficient de corrélation linéaire entre les variables du tableau sont données par la **matrice des corrélations** ("Correlation Matrix"):

Correlation Matrix:

	CA	CL	HCO3	MG	NA	SO4
CA	1.00000					
CL	.25204	1.00000				
HCO3	.85175	.12178	1.00000			
MG	.60568	-.12546	.73058	1.00000		
NA	-.19619	.66801	-.10883	-.09055	1.00000	
SO4	.73265	.04493	.47796	.67063	-.27878	1.00000

Dans l'ACP normée (i.e. sur données centrées réduites), la matrice des corrélations constitue l'opérateur d'inertie qui sera diagonalisé (i.e. extraction des valeurs propres).

### 3.3. Valeurs propres

Les principaux indicateurs statistiques concernant l'extraction des **facteurs principaux d'inertie**, appelés **composantes principales** en ACP, sont :

- la **valeur propre**  $\lambda_\alpha$  (colonne "Eigenvalue") représente l'inertie de l'axe principal de rang  $\alpha$ ;
- l'**inertie totale** est égale à la somme des valeurs propres  $I = \sum_{\alpha \in A} \lambda_\alpha$ ;
- la contribution de chaque axe principal à l'inertie totale est donnée par le **pourcentage d'inertie**  $\tau_\alpha = \lambda_\alpha / I$  (colonne "Pct of Var").

Extraction 1 for analysis 1, Principal Components Analysis (PC)

Initial Statistics:

Variable	Communality	Factor	Eigenvalue	Pct-of-Var	Cum-Pct
CA	1.00000	1	3.09409	51.6	51.6
CL	1.00000	2	1.68756	28.1	79.7
HCO3	1.00000	3	.59651	9.9	89.6
MG	1.00000	4	.50284	8.4	98.0
NA	1.00000	5	.09324	1.6	99.6
SO4	1.00000	6	.02576	.4	100.0

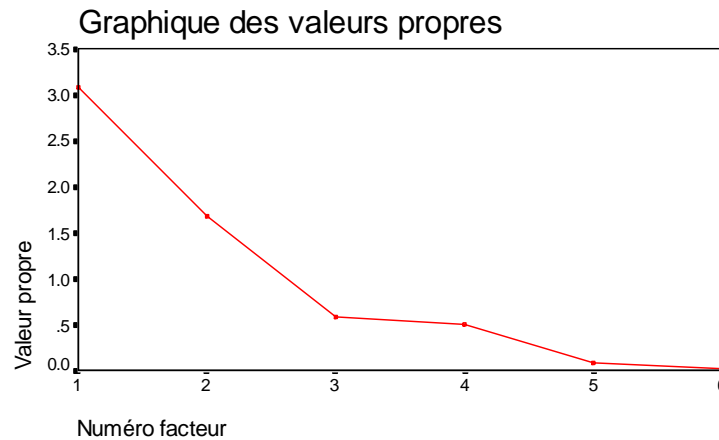
Le choix du nombre de composantes principales à retenir pour représenter les données dans la nouvelle base des vecteurs propres ou **axes principaux d'inertie** peut s'effectuer de différentes façons :

- en choisissant un niveau global de **pourcentage d'inertie cumulé** (colonne "Cum Pct"), permettant d'affirmer que l'on représente, par exemple, 80% de l'inertie totale;
- en sélectionnant les axes principaux dont l'inertie est supérieure à celle des variables originales. C'est à dire, en ACP normée, supérieure à 1 puisque chacune des variables originales a été réduite;
- en étudiant la distribution des valeurs propres (cf. ci-dessous "Graphique des valeurs propres") pour détecter une décroissance brutale de l'inertie permettant de considérer les axes résiduels comme négligeables.

L'existence du graphique des valeurs propres est signalée par le logiciel SPSS dans le listing des résultats par le message suivant :

**Hi-Res Chart # 1:Graphique des valeurs propres**

Il suffit alors d'ouvrir le *Carousel* pour consulter le graphique des valeurs propres :



L'examen de ce graphique montre que l'on peut se limiter à l'extraction des deux premières composantes principales qui permettent de prendre en compte environ 79,7% de l'inertie totale. Cependant, il est conseillé de vérifier si une variable possédant par exemple une valeur élevée pour l'une des composantes principales non extraites pourrait ne pas être bien représentée dans l'espace des  $q$  premiers facteurs.

De telles situations peuvent être détectées en examinant la matrice des corrélations variables/facteurs d'une part et les graphiques factoriels d'autre part (projections éloignées du cercle des corrélations).

### 3.4. Coordonnées factorielles

#### 3.4.1. Variables

Les coordonnées factorielles des variables du tableau sont données par la **matrice des corrélations variables/facteurs** ("Factor Matrix") dont le calcul s'effectue au moyen de la relation :

$$G_{\alpha}(j) = \rho(c_{\alpha}, x_j) = \sqrt{\lambda_{\alpha}} \times v_{j\alpha}$$

où  $v_{j\alpha}$  est la  $j^{\text{e}}$  composante du vecteur propre  $V_{\alpha}$  associé à la  $\alpha^{\text{e}}$  valeur propre  $\lambda_{\alpha}$ .

Factor Matrix:						
label	Factor-1	Factor-2	Factor-3	Factor-4	Factor-5	Factor-6
CA	.92163	.16628	-.24928	-.18634	.13031	-.09545
HCO3	.87597	.14270	.18788	-.40999	.00121	.09446
MG	.84667	-.04638	.46444	.21020	-.13364	-.05697
SO4	.82972	-.06053	-.26847	.47837	.05052	.06657
CL	.04448	.93620	-.30135	-.01518	-.17465	-.00190
NA	-.26475	.87021	.34710	.16365	.15919	.00638

En utilisant cette relation, on peut, à partir de la matrice des corrélations variables-facteurs, obtenir la matrice des vecteurs propres :

$$v_{j\alpha} = G_{\alpha}(j) / \sqrt{\lambda_{\alpha}}$$

par exemple,  $v_{Ca,1} = G_1(Ca) / \sqrt{\lambda_1} \approx 0,92163 / \sqrt{3,09409} \approx 0,524$

### Aide à l'interprétation

La qualité de la représentation d'un point-variable  $P^j$  par sa projection  $H_\alpha^j$  sur un axe  $\alpha$  est définie par le rapport de l'inertie de la projection  $H_\alpha^j$  du point  $P^j$  sur l'axe  $\alpha$ , à l'inertie totale du point  $P^j$  :

$$QLT_\alpha(j) = \frac{\|OH_\alpha^j\|^2}{\|OP^j\|^2} = \cos^2 \theta = \rho^2(c_\alpha, x_j)$$

Ce rapport s'exprime comme le cosinus carré de l'angle  $\theta$  entre les deux vecteurs  $OH_\alpha^j$  et  $OP^j$ . Il se calcule en prenant le carré du coefficient de corrélation entre la variable  $x_j$  et le facteur  $c_\alpha$ . Cette définition s'étend à la qualité de la représentation par la projection sur un sous-espace factoriel de dimension  $q$  en sommant ce rapport sur chacun des  $q$  facteurs.

Ainsi, en supposant que l'on ait choisi un modèle de représentation des données à trois facteurs ( $q = 3$ ), on peut vérifier la reconstitution du point-variable  $SO_4^-$  par la qualité de sa représentation dans ce sous-espace :

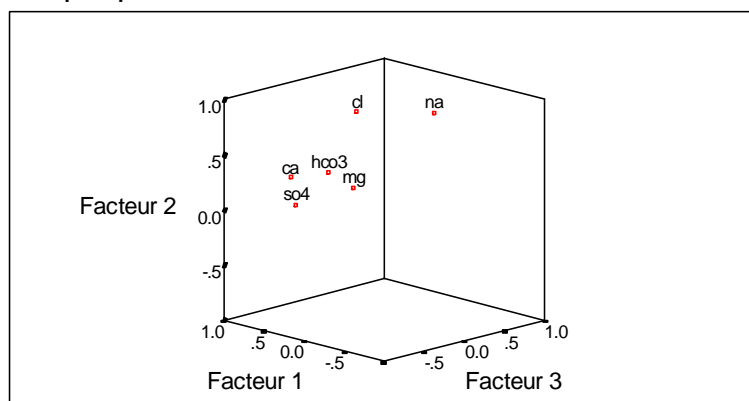
$$QLT_{F_1 \times F_2 \times F_3}(SO_4^-) = (0,82972)^2 + (-0,06053)^2 + (-0,26847)^2 \approx 0,764$$

qui est environ de 76,4%.

### Représentations graphiques

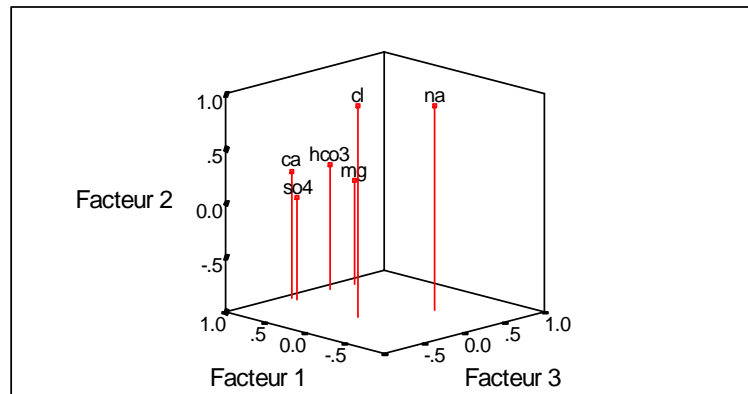
Les représentations graphiques fournies par la procédure FACTOR concernent uniquement l'espace des points-variables. Par défaut, si l'on extrait plus de deux composantes principales, le graphique factoriel représente les points-variables dans le repère orthonormé  $F_1 \times F_2 \times F_3$  des trois premiers axes principaux d'inertie.

Graphique factoriel



Une solution pour améliorer la lisibilité de ce graphique est de le modifier. Appuyez sur le bouton (Modifier) correspondant dans le carrousel de graphiques puis, dans la fenêtre graphique, utilisez le choix Options ... du menu Graphiques. Il faut ensuite choisir l'option Plancher dans la liste des options Projections de la boîte de dialogue Options diagramme de dispersion 3D. On obtient ainsi le graphique suivant :

## Graphique factoriel



Néanmoins, dès que le nombre de variables actives s'accroît cette solution devient vite impraticable. Une alternative consiste à utiliser les fonctions supplémentaires de la procédure FACTOR fournies par la syntaxe du langage de commandes SPSS : la sous-commande PLOT du bloc ANALYSIS permet de tracer des plans factoriels en spécifiant les numéros des axes principaux d'inertie en abscisse et en ordonnée.

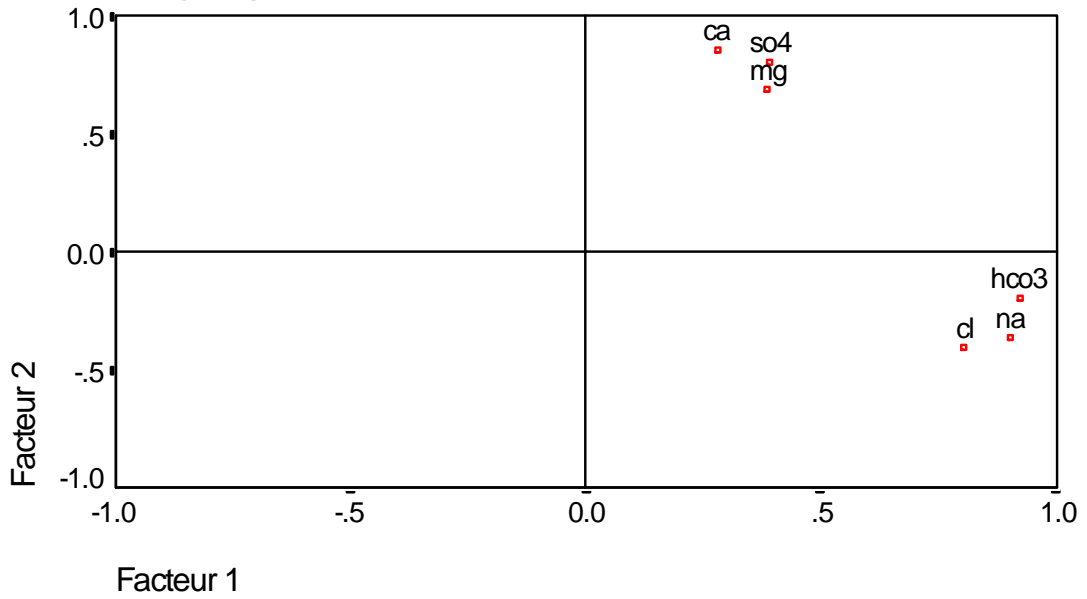
```

FACTOR
/VARIABLES ca cl hco3 mg na so4
/MISSING LISTWISE
/ANALYSIS ca cl hco3 mg
na so4
/PRINT UNIVARIATE CORRELATION
/PLOT EIGEN ROTATION (1,2) (3,4)
/CRITERIA FACTORS(4) ITERATE(25)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL) .

```

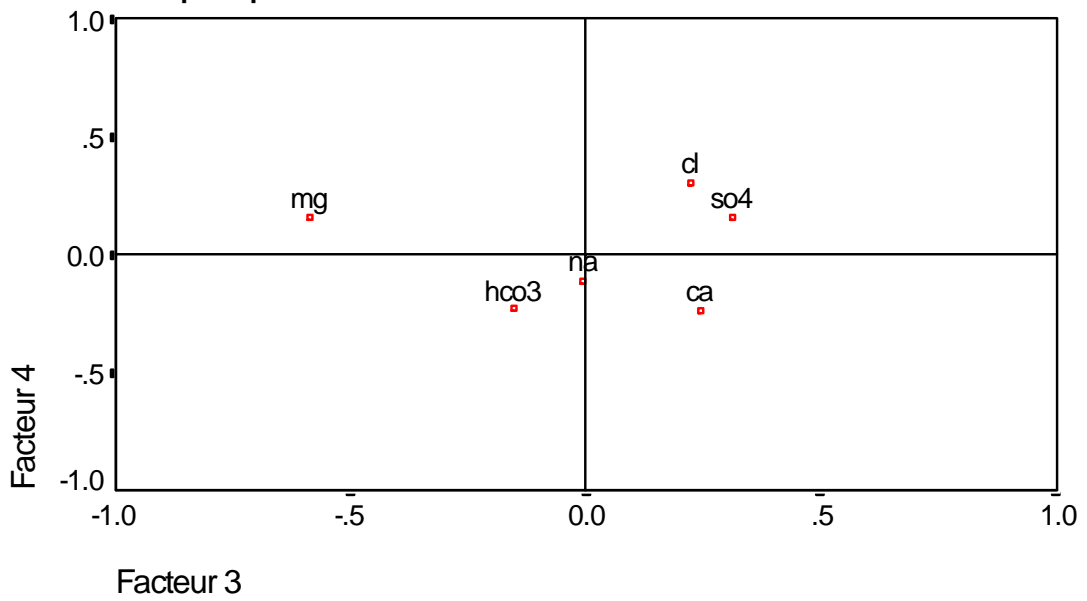
On obtient ainsi les projections des variables actives de l'ACP sur le graphique-plan  $F_1 \times F_2$  :

## Graphique factoriel



suivi d'une représentation graphique de ces mêmes variables dans le plan factoriel  $F_3 \times F_4$  :

## Graphique factoriel



### 3.4.2. Individus

On peut également obtenir la matrice des vecteurs propres (coordonnées des vecteurs propres dans la base des variables d'origine) à partir de la **matrice des coefficients factoriels** ("Factor Score Coefficient Matrix"). Il suffit de cliquer sur le bouton **Facteurs** dans la boîte de dialogue **Analyse factorielle** pour afficher la boîte de dialogue correspondant au calcul des facteurs et de cocher la sélection **Afficher la matrice des coefficients factoriels**.

**Factor Score Coefficient Matrix:**

Label	Factor-1	Factor-2	Factor-3	Factor-4	Factor-5	Factor-6
CA	.29787	.09853	-.41790	-.37057	1.39755	-3.70609
CL	.01438	.55477	-.50518	-.03019	-1.87317	-.07368
HCO3	.28311	.08456	.31496	-.81535	.01296	
MG	.27364	-.02749	.77859	.41802	-1.43334	
NA	-.08557	.51566	.58189	.32545	1.70735	
SO4	.26816	-.03587	-.45007	.95132	.54179	

Cette matrice de coefficients factoriels permet de calculer les coordonnées factorielles des individus à partir du tableau des données centrées réduites :

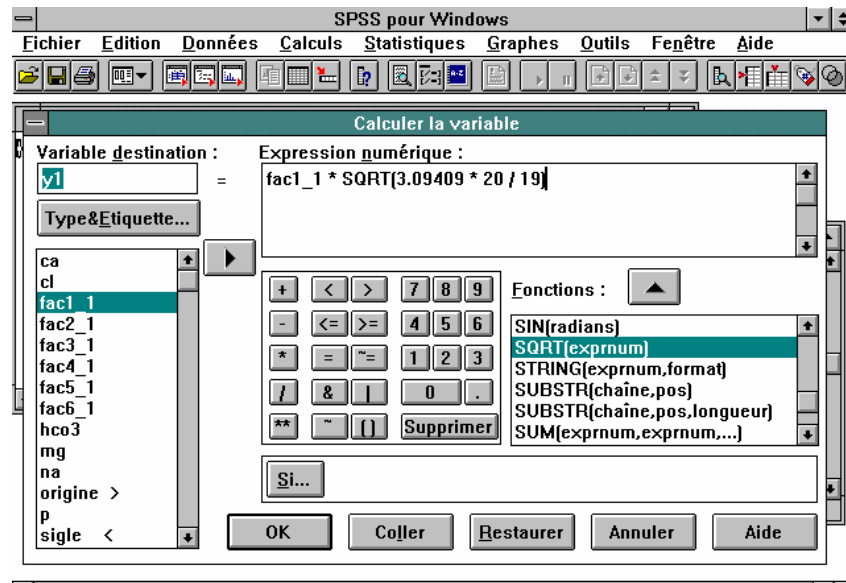
$$F_{\alpha}(i) = \sum_{j \in J} w_{\alpha j} z_{ji}$$

L'estimation de ces coefficients factoriels est effectuée par la procédure *SPSS FACTOR* selon trois méthodes (choix **R**égression, **B**artlett et **A**nderson-Rubin pour la sélection **M**éthode de la boîte de dialogue **Analyse factorielle : Facteurs**) qui sont équivalentes dans le cas particulier de l'Analyse en Composantes Principales.

Les projections des individus sur les graphiques-plans s'obtiennent à partir des coordonnées factorielles à un facteur d'échelle près :

$$G_{\alpha}(i) = F_{\alpha}(i) \sqrt{\lambda_{\alpha} n / (n-1)}$$

Ce calcul peut être effectué par l'option Calculer du menu Calculs à partir de l'interface des boîtes de dialogue :



Mais on peut procéder également à ce calcul en utilisant la syntaxe suivante du langage de commandes *SPSS* :

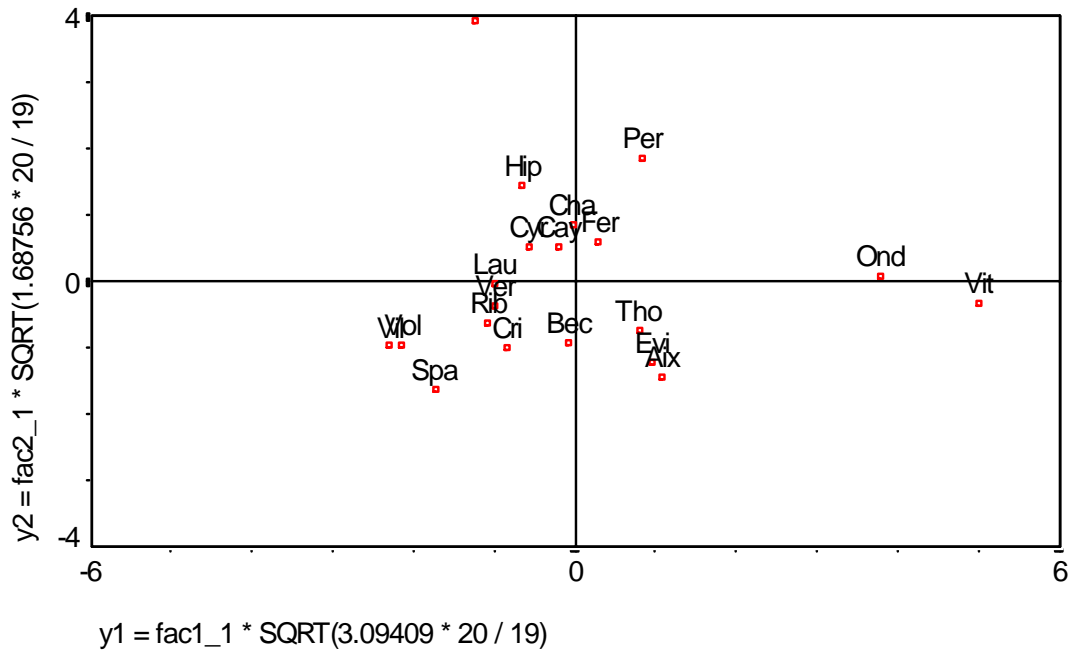
```
COMPUTE y1 = fac1_1 * SQRT(3.09409 * 20 / 19) .
COMPUTE y2 = fac2_1 * SQRT(1.68756 * 20 / 19) .
COMPUTE y3 = fac3_1 * SQRT(0.59651 * 20 / 19) .
COMPUTE y4 = fac4_1 * SQRT(0.50284 * 20 / 19) .
COMPUTE y5 = fac5_1 * SQRT(0.09324 * 20 / 19) .
COMPUTE y6 = fac6_1 * SQRT(0.02576 * 20 / 19) .
EXECUTE .
```

On peut alors utiliser la sous-commande *SCATTERPLOT* de la commande *GRAPH* pour afficher les projections des individus sur les premiers plans factoriels

```
GRAPH
/SCATTERPLOT(BIVAR)=y1 WITH y2 BY sigle (IDENTIFY)
/MISSING=LISTWISE .
```



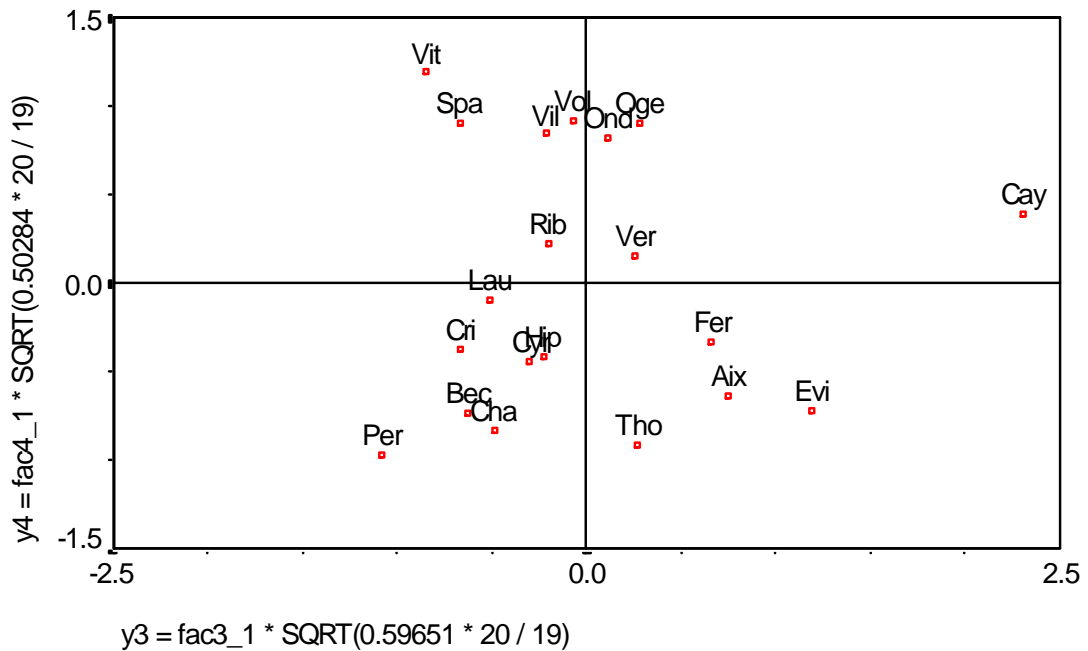
On obtient ainsi les projections des individus actifs de l'ACP sur le graphique-plan  $F_1 \times F_2$  :



En réitérant la commande GRAPH pour les axes factoriels suivants,

```
GRAPH
/SCATTERPLOT(BIVAR)=y3 WITH y4 BY sigle (IDENTIFY)
/MISSING=LISTWISE .
```

on obtient une représentation graphique des mêmes individus dans le plan factoriel  $F_3 \times F_4$  :



## 4. Références

- Tomassone R., Dervin C., Masson J.-P. 1993. *BIOMÉTRIE, Modélisation des phénomènes biologiques*, Masson, Paris, 553 p.
- SPSS Inc. 1994. *SPSS 6.1 Categories*, SPSS Inc., Chicago, 209 p.

## 5. Annexe

Tableau des données quantitatives : Composition chimique de 20 eaux minérales.

ORIGINE	SIGLE	HCO3	SO4	CL	CA	MG	NA
Aix-les-Bains (F-73)	Aix	341	27	3	84	23	2
Beckerish (F-68)	Bec	263	23	9	91	5	3
Cayranne (F-84)	Cay	287	3	5	44	24	23
Chambon (F-45)	Cha	298	9	23	96	6	11
Cristal-Roc (F-72)	Cri	200	15	8	70	2	4
Saint-Cyr (F-45)	Cyr	250	5	20	71	6	11
Evian (F-74)	Evi	357	10	2	78	24	5
Ferita (I)	Fer	311	14	18	73	18	13
Saint-Hyppolite (F-37)	Hip	256	6	23	86	3	18
Laurier (F-77)	Lau	186	10	16	64	4	9
Ogeu (F-64)	Oge	183	16	44	48	11	31
Ondine (F-95)	Ond	398	218	15	157	35	8
Perrier (F-30)	Per	348	51	31	140	4	14
Ribes (E)	Rib	168	24	8	55	5	9
Spa (B)	Spa	110	65	5	4	1	3
Thonon (F-74)	Tho	332	14	8	103	16	5
Veri (E)	Ver	196	18	6	58	6	13
Viladreau (E)	Vil	59	7	6	16	2	9
Vittel (F-88)	Vit	402	306	15	202	36	3

d'après [Tomassone, Dervin & Masson 1993], p.114.