

Un outil graphique interactif d'aide à l'interprétation de résultats d'analyse de données : FGPlot

Auteur: Françoise GOUPIL-TESTU
Laboratoire LISE - CEREMADE (URA 748)
Université Paris IX-Dauphine
75775 PARIS CEDEX 16

Introduction

La pratique de l'analyse de données conduit souvent à une représentation des objets ou individus sur un plan : plan déterminé par 2 variables quelconques, plan factoriel, plan discriminant. La possibilité d'exploiter ces résultats en **mode graphique** est désormais considérée comme un complément précieux, voire indispensable. Parallèlement, la possibilité de modifier le choix des données et le mode de représentation à partir du graphique en cours, donc **interactivement** s'impose tout naturellement. Ceci doit en outre pouvoir s'effectuer autant que possible à partir des fichiers de résultats que produisent les logiciels d'analyse de données (LEAS [1], SICLA [2], SPAD [3] par exemple) donc sans requérir de traitement intermédiaire supplémentaire ni de manipulation importante. Enfin, la facilité d'utilisation et d'apprentissage, autrement dit la **convivialité**, est également un impératif majeur.

Ces principes ont guidé le développement de **FGPlot**, qui trouve son origine dans la volonté de faciliter le dépouillement répétitif de résultats dans le cadre d'un projet de recherche mené au Groupe de Recherche en Astronomie et Astrophysique du Languedoc [4], sous la direction de **M.O. Mennessier (URA 1368)**.

Environnement

FGPlot ne requiert qu'un environnement de type courant :

matériel

- micro-ordinateur compatible IBM PC
- écran graphique EGA ou VGA, couleur ou monochrome
- souris

logiciel système

- MS-DOS

fichiers de données

FGPlot est prévu pour traiter des fichiers ayant les caractéristiques suivantes :

- fichiers « texte » ASCII
- enregistrements de longueur fixe (le séparateur d'enregistrement peut indifféremment être CR/LF ou LF seul)

- jusqu'à 32000 octets par enregistrement
- jusqu'à 60 champs par enregistrement
- un ou plusieurs blancs comme séparateur de champs

FGPLOT peut traiter des fichiers contenant jusqu'à 32000 enregistrements

Tout fichier ASCII provenant d'une analyse de données, modifié éventuellement avec un éditeur de texte, peut être utilisé s'il suit les règles précédentes.

Impression des graphiques:

La touche **Impression Ecran** est utilisée pour l'impression directe ou pour la capture par un logiciel approprié (Pizzaz plus par exemple). On peut aussi sauvegarder le graphique dans un fichier au format PCX (voir Guide d'utilisation).

Fonctions statistiques

Représentation plane d'un nuage de points.

FGPLOT réalise la représentation d'un nuage de points dans un plan factoriel, dans un plan discriminant ou dans le plan de deux variables quantitatives quelconques (cf graphique 1).

Introduction d'une variable nominale.

L'attribution d'un symbole graphique à chaque modalité permet de visualiser la répartition des modalités de cette variable sur le plan sélectionné.

Cas particuliers :

- visualisation des classes associées aux modalités de la variable nominale sur le plan factoriel
- visualisation des classes associées aux modalités de la variable à expliquer dans le plan discriminant
- visualisation des classes obtenues par SICLA sur un plan factoriel, les composantes principales ayant pu être fournies par LEAS ou SPAD par exemple, cf graphique 2.

On peut également faire afficher les effectifs de chaque classe avec la légende des symboles

Introduction de deux variables nominales.

On ne peut affecter des symboles graphiques différents qu'à l'une des deux variables nominales sélectionnées qui sera dite variable qualitative principale.

Les différents symboles, combinés avec les possibilités d'affichage unitaire de chaque classe de modalité, permettent de visualiser les distributions conditionnelles successives d'une variable sur chacune des classes déterminées par l'autre, ce qui revient à **visualiser le tableau de contingence**. (cf graphique 3).

Cas particulier :

visualisation de la répartition d'une variable nominale dans les classes obtenues par SICLA (cf graphique 4).

Fonctions graphiques.

On fournit ci-dessous la liste des fonctionnalités du logiciel

- adaptation automatique à la structure du fichier de données
- libre choix des champs à utiliser :
 - ⇒ champ identificateur d'objet
 - ⇒ champ abscisse
 - ⇒ champ ordonné
 - ⇒ champ variable qualitative principale (optionnel)
 - ⇒ champ variable qualitative secondaire (optionnel)
- traitement automatique des axes et des échelles sur les axes
- interactivité : logiciel piloté par touches de fonction
- fonctions de base :
 - ⇒ différents symboles graphiques pour chaque valeur de la variable qualitative principale
 - ⇒ affichage d'une légende des symboles (mobile)
 - ⇒ affichage des limites d'abscisse et d'ordonnée (mobile)
 - ⇒ ajout éventuel d'un titre pour le graphique et pour chaque variable
- fonctions avancées :
 - ⇒ sélection de modalités spécifiques des variables nominales (principale, secondaire ou combinées)
 - ⇒ affichage d'identificateurs
 - affichage d'identificateur près d'un symbole (pointé, par la souris)
 - affichage d'identificateur près de chaque symbole
 - liste et total des identificateurs pour un groupe de symboles (cadre avec la souris)
 - ⇒ inversement, repérage d'un symbole à partir de l'identificateur
 - ⇒ zoom (avec recalcul automatique des axes et des échelles, et utilisable en cascade)
 - ⇒ réaffectation interactive des symboles (formes et couleurs)
 - ⇒ sauvegarde des graphiques en format **PCX**.

Guide d'utilisation.

Installation

FGPLOT est fourni sur une disquette 3 pouces ½ de 720 Koctets avec un fichier de données de test : **Testplot**. Il suffit de copier le contenu de la disquette dans le répertoire de son choix.

Principes

FGPLOT est piloté au moyen des touches de fonction (Fn). En raison du nombre limité des touches de fonction, elles sont combinées avec les touches Alt, Ctrl et -plus rarement- Maj.

Les touches valides (en fonction de l'environnement et des choix antérieurs) sont affichées sur une ligne d'aide en bas de l'écran. L'affichage en bas d'écran tient compte de ces dernières pour refléter le choix réel offert à un instant donné.

Certaines fonctions font ensuite appel à la souris ou aux touches de curseur. La procédure d'utilisation de la souris est affichée sur l'écran quand nécessaire.

Démarrage

Taper **FGPLOT** sur la ligne de commande DOS. Après le panneau d'accueil, dont on sort en appuyant sur Entrée, **FGPLOT** propose le choix entre travailler sur un nouvel environnement (option 1) ou un environnement sauvegardé (option 2). Cette seconde possibilité sera expliquée plus loin à propos de la sauvegarde d'environnement. **FGPLOT** demande ensuite le nom du fichier de données à exploiter. Il faut entrer le nom de fichier complet avec le chemin, si ce fichier n'est pas dans le répertoire en cours. **FGPLOT** analyse la structure du fichier, affiche une liste numérotée des champs composant les enregistrements et demande d'indiquer les numéros des champs suivants :

- champ identificateur d'objet - obligatoire
- champ de la variable abscisse - obligatoire
- champ de la variable ordonnée - obligatoire
- champ de la variable qualitative principale - optionnel
- champ de la variable qualitative secondaire - optionnel (n'est pas demandé si la variable qualitative principale n'a pas été sélectionnée).

Exemple de démarrage

Cet exemple concerne le démarrage du graphique 4

champs des enregistrements du fichier (exemple du 1^{er} enregistrement) :

1 : et001
2 : 10 7
3 : -63 9
4 : 96 4
5 : 76
6 : -1 082
7 : -1 036
8 : -2 420
9 : 1
10 : 1
11 : 4

numéro du champ pris comme identificateur ? 1

numéro du champ pris comme abscisse ? 6

numéro du champ pris comme ordonnée ? 7

numéro du champ pris comme variable qualitative principale (ou Entrée) ? 9

numéro du champ pris comme variable qualitative secondaire (ou Entrée) ? 11

Après le choix des paramètres, **FGPLOT** exécute une phase d'initialisation puis affiche un graphique initial

Exploitation du graphique

Les fonctions d'exploitation du graphique offrent de nombreuses possibilités pour adapter l'affichage aux souhaits de l'utilisateur

◆ F1 : Affichage des fonctions des touches (aide)

Affiche le résumé sous forme de tableau reproduit ci-dessous , des diverses actions déclenchées par les touches de fonction.

F1	base : affichage de l'aide (cet affichage) Alt : affichage de l'environnement en cours
F2	Variable qualitative principale : base : affichage par valeurs unitaires successives Alt : sélection directe de une ou plusieurs valeurs Ctrl : resélection de toutes les valeurs
F3	Variable qualitative secondaire : base : affichage par valeurs unitaires successives Alt : sélection directe de une ou plusieurs valeurs Ctrl : resélection de toutes les valeurs
F4	base : édition titre et/ou noms des coordonnées et var. qualitative Alt : réaffectation des symboles (si variable qualitative) Ctrl : resélection de toutes les valeurs
F5	base : déplacement légende des limites de X et Y Alt : déplacement légende des symboles (si var. qualitative) Ctrl : va et vient légende des limites de X et Y Maj : va et vient légende des symboles (si var. qualitative)
F6	Zoom : base : agrandissement (par étapes) Alt : réduction (retour par les mêmes étapes) Ctrl : annulation directe du zoom
F7	Identification : base : affichage de l'identificateur d'un symbole pointé Alt : affichage des ident. d'un groupe encadré (avec comptage) Ctrl : affichage des ident. de tous les symboles affichés
F8	base : repérage de la position d'un objet à partir de son nom
F9	base : rafraîchissement affichage Alt : va et vient affichage / effacement des libellés des clés Ctrl : échelles maximales sur chaque axe Maj : échelles identiques
F10	base : choix de nouveaux critères de sélection Alt : sauvegarde du graphique et de son environnement Ctrl : sortie du programme Maj : sauvegarde de l'image écran dans un fichier PCX

◆ **Alt-F1 : Affichage de l'environnement en cours**

Indique l'état de l'environnement : fichier utilisé, valeurs de paramètres sélectionnés, niveau de zoom.

◆ **F2 : Exploitation de la variable qualitative principale**

L'affichage est cyclique, modalité par modalité

Chaque action sur la touche sélectionne une nouvelle modalité en commençant par la plus faible. Le programme affiche alors la classe correspondante et ainsi de suite en rebouclant à la fin. Les valeurs n'existant pas dans le fichier sont sautées pour éviter les actions inutiles. Ce type d'affichage est pratique s'il n'y a pas trop de modalités.

◆ **Alt-F2 : affichage avec sélection directe des valeurs (une ou plusieurs)**

Limite l'affichage aux objets correspondant aux modalités de la variable qualitative indiquées par l'utilisateur. Ces valeurs sont demandées par le programme, lequel indique à l'utilisateur les diverses valeurs possibles (c'est à dire présentes dans le fichier). La légende des symboles indique les symboles associés aux valeurs retenues.

◆ **Ctrl-F2 : retour à l'affichage de toutes les valeurs**

Annule les sélections opérées par F2 ou Alt-F2 et provoque le retour immédiat à l'affichage de l'ensemble des valeurs.

◆ **F3 : Exploitation de la variable qualitative secondaire**

Mêmes fonctions que pour la variable qualitative principale, sauf pour la gestion des symboles, qui ne concerne pas la variable qualitative secondaire.

◆ **F4 : Edition**

Permet de donner un titre au graphique, ainsi que des noms aux coordonnées, et aux variables qualitatives si elles existent.

Par défaut, l'abscisse est désignée par X, l'ordonnée par Y, la variable qualitative principale est désignée par « var1 », la secondaire par « var2 ». Il n'y a pas de titre par défaut. Le titre et les noms de variables sont conservés jusqu'à la fin de la session, mais peuvent être modifiés à volonté durant cette session.

◆ **Alt-F4 : Réaffectation de symboles**

Permet de modifier les symboles affectés aux différentes valeurs de la variable qualitative principale, si le choix par défaut ne convient pas. La sélection s'effectue avec la souris et les instructions nécessaires sont affichées au fur et à mesure.

◆ **Ctrl-F4 : Affichage multi-symboles/mono-symbole**

En cas d'affichage multi-symboles (variable qualitative principale utilisée), ceci permet de réafficher avec un symbole unique, notamment pour mieux refléter la densité des objets dans les diverses zones quand ils sont nombreux. La même opération permet le retour au multi-symboles (fonctionne en va et vient).

Le symbole unique par défaut est le point mais il peut être modifié dynamiquement (voir Alt-F4).

◆ **F5 : Déplacement de la légende des limites X/Y**

Cette fonction permet de déplacer la légende comme suit :

- utiliser les touches de déplacement du curseur (la légende se déplace en conséquence sur l'écran)
- appuyer sur Entrée pour fixer la légende et réafficher le graphique.

La nouvelle position est conservée jusqu'à la fin de la session, sauf nouveau déplacement.

◆ **Alt-F5 : Déplacement de la légende des symboles**

Même méthode que pour la légende des limites X/Y

◆ **Ctrl-F5 : Va et vient de la légende des limites X/Y**

◆ **Maj-F5 : Va et vient de la légende des symboles**

◆ **F6 : Zoom**

Permet d'affecter la totalité de l'écran à une zone délimitée par l'utilisateur. Ceci peut servir à focaliser sur un groupe d'objets mais aussi à dégager des objets cachés.

La zone visée est délimitée au moyen de la souris. Les instructions nécessaires sont affichées par le programme au fur et à mesure.

Ce zoom peut être utilisé « en cascade » dans la limite de 10 zooms successifs. L'indication du « niveau de zoom » en cours figure en haut et à gauche de l'écran, sous la forme « Zn », (n=1, 10). Les limites de zones sont mémorisées à chaque étape, pour permettre le retour arrière par les mêmes étapes (par exemple pour revenir en arrière d'une étape en vue de rectifier le choix de zone).

Ce zoom n'est pas un simple agrandissement de zone : l'environnement est réadapté et recalculé, afin que les autres fonctions restent totalement utilisables et valides. Seule la mémorisation des identificateurs d'objets (voir ci-dessous) est perdue.

◆ **Alt-F6 : Zoom « arrière ».**

Ramène au niveau de zoom précédent.

◆ **Ctrl-F6 : annulation du zoom**

Ramène directement à l'échelle de départ (« niveau 0 »).

◆ **F7 : identification unitaire d'objets**

Par souci de clarté et pour cacher le moins d'objets possibles, les noms des objets ne sont pas affichés sur le graphique initial. Pour identifier un objet, on le pointe avec la souris. L'affichage est déclenché avec le bouton gauche de la souris. L'identificateur de l'objet est affiché à côté du symbole et la légende des limites X/Y est remplacée par les coordonnées de l'objet.

L'identificateur peut être déplacé avec la souris dont il suit le mouvement. Il peut être soit effacé (bouton droit), soit fixé et mémorisé (bouton gauche). On peut pointer et mémoriser autant d'objets que l'on veut. L'affichage et la mémorisation peuvent être annulés en pointant à nouveau l'objet.

◆ **Alt-F7 : affichage et comptage d'un groupe d'objets**

Permet d'afficher la liste des noms et le nombre total des objets figurant dans une zone délimitée au moyen de la souris (les instructions nécessaires sont affichées par le programme au fur et à mesure).

Si le nombre et la taille des identificateurs font que la taille de l'écran est dépassée, l'affichage des identificateurs s'arrête en fin d'écran, et se termine par « ... ». Le nombre affiché représente néanmoins la totalité des objets du groupe.

◆ **Ctrl-F7 : affichage de tous les identificateurs**

Permet d'afficher les identificateurs de tous les objets en une seule action.

◆ **F8 : repérage d'objet**

A l'inverse des fonctions précédentes, cette fonction permet de repérer la position d'un objet en donnant son identificateur. Le programme demande cet identificateur, puis pointe le symbole de l'objet correspondant avec la flèche de la souris (ou indique qu'il est introuvable).

◆ **F9 : Réaffichage**

Permet de rafraîchir l'écran pour annuler les affichages additionnels temporaires entraînés par certaines des fonctions.

◆ **Alt-F9 : Effacement des libellés des touches**

Annule l'affichage des libellés des touches figurant au bas de l'écran, notamment pour améliorer l'aspect du graphique en cas d'impression. La même opération permet de les réafficher (fonctionne en va et vient).

◆ **Ctrl-F9 : Echelles maximales sur chaque axe**

Impose la résolution maximale de l'écran suivant les deux axes pour utiliser au mieux la surface de l'écran

◆ **Maj-F9 : Echelles identiques sur chaque axe**

Impose la même résolution sur les deux axes (une même unité occupe une même longueur physique sur l'écran).

◆ **F10 : resélection des variables d'environnement**

Permet, sans avoir à sortir du programme et le relancer, de modifier la sélection des variables d'environnement : abscisse, ordonnée, variable qualitative principale, variable qualitative secondaire.

La séquence d'affichage de la structure du fichier et de choix des champs est identique à celle du début de programme (voir section démarrage de **FGPLOT**)

◆ **Alt-F10 : mémorisation de l'environnement**

Permet de mémoriser sur disque l'environnement pour pouvoir notamment :

1. interrompre le travail en cours et le reprendre ultérieurement au point exact où il en était
2. conserver cet environnement pour référence ultérieure.

Le programme demande un nom pour le fichier de sauvegarde (1 à 8 caractères). Il rajoute l'extension **FGS**.

Il faut bien noter que :

- il s'agit d'une sauvegarde d'environnement et non pas d'une simple sauvegarde d'image d'écran

- lors du rappel de cet environnement (voir section démarrage), on se retrouve exactement au même point du traitement, avec les mêmes possibilités d'interaction, comme si ce traitement n'avait pas été interrompu
- tout est sauvegardé : sélection de paramètres, zoom, titre, position des légendes, affichage des noms mémorisés, sélection des symboles, etc
- la place nécessaire sur disque est très faible (de l'ordre de 500 octets) alors que la sauvegarde d'une image d'écran VGA 640x480 16 couleurs -par exemple- nécessite environ 150 Koctets.

◆ **Ctrl-F10 : Sortie du programme**

Termine le programme et retourne au DOS. Une confirmation est demandée, pour éviter une sortie involontaire suite à une erreur de manipulation.

◆ **Maj-F10 : Sauvegarde de l'image écran dans un fichier PCX**

Transcrit l'image écran en fichier PCX, pour reprise ultérieure par les logiciels sachant exploiter ce type de fichiers (par exemple, pour inclusion du graphique dans un document sous Word).

Jeu d'essai

Description du fichier

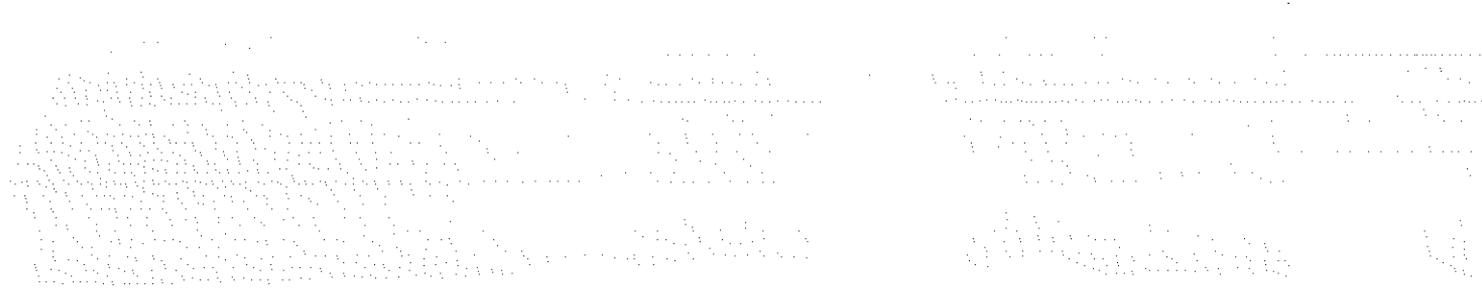
Le fichier TESTPLOT [4] contient 168 enregistrements relatifs à des étoiles

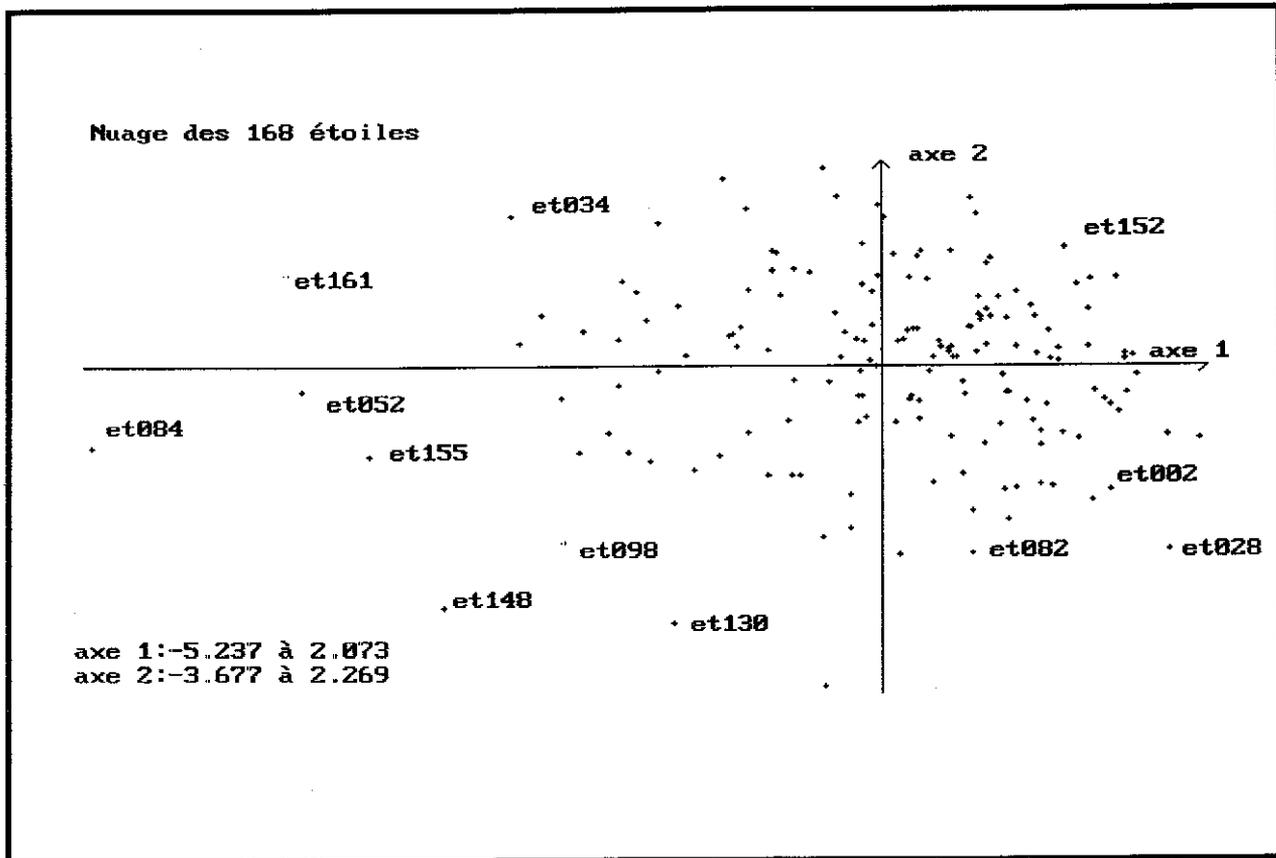
Chaque enregistrement est constitué de 11 champs

- le 1^{er} champ contient un identificateur de l'étoile. Les deux premiers caractères de cet identificateur sont « et ». Les suivants sont le numéro de l'étoile dans le fichier
- Les champs 2 à 5 sont les valeurs de variables quantitatives: deux vitesses dans le plan galactique, une vitesse perpendiculaire au plan galactique et la distance au plan galactique
- Les champs 6 à 8 sont les 3 premières composantes principales fournies par LEAS
- Les champs 9 à 11 contiennent des variables nominales :
 - * champ 9 : variable émission oh à 5 modalités ,notées oh1,oh2,oh3,oh4,oh5
 - * champ 10 : variable émission Sio à 2 modalités, notées sio0, sio1
 - * Champ 11 : variable nominale correspondant à une partition en 6 classes des 168 étoiles par SICLA.

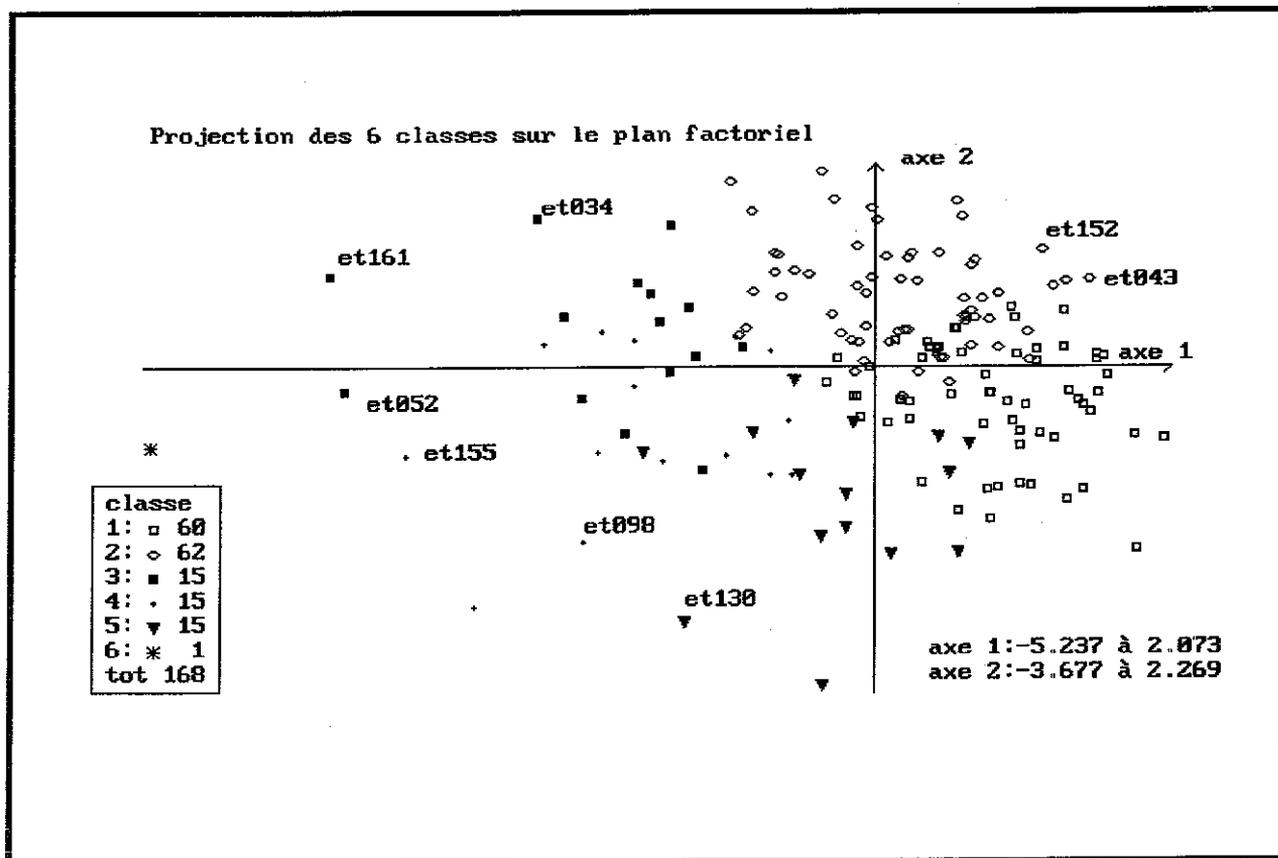
Graphique 1

Le graphique ci-dessous représente le nuage des 168 étoiles dans le plan d'une analyse factorielle avec identification de quelques étoiles.



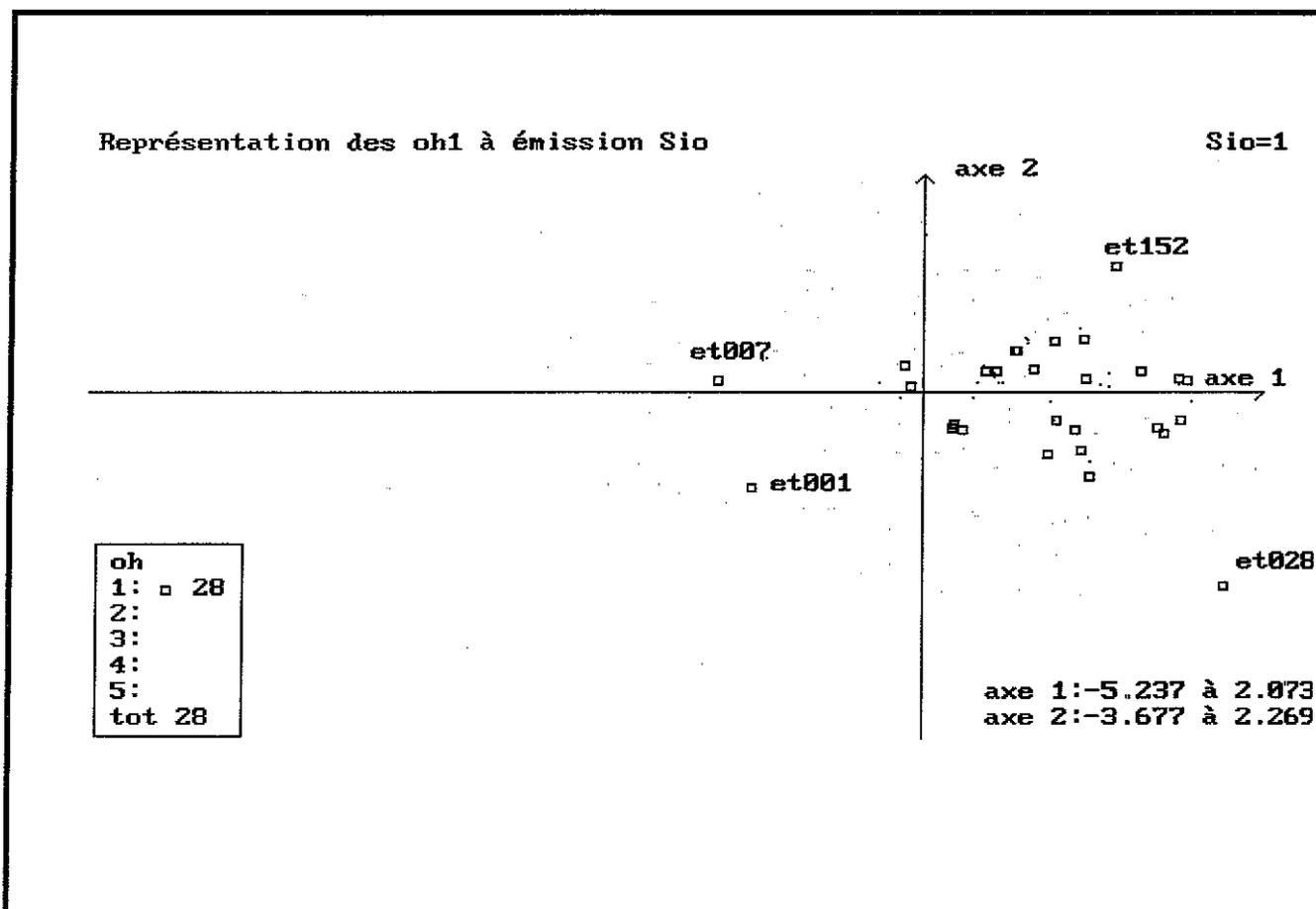


Graphique 2



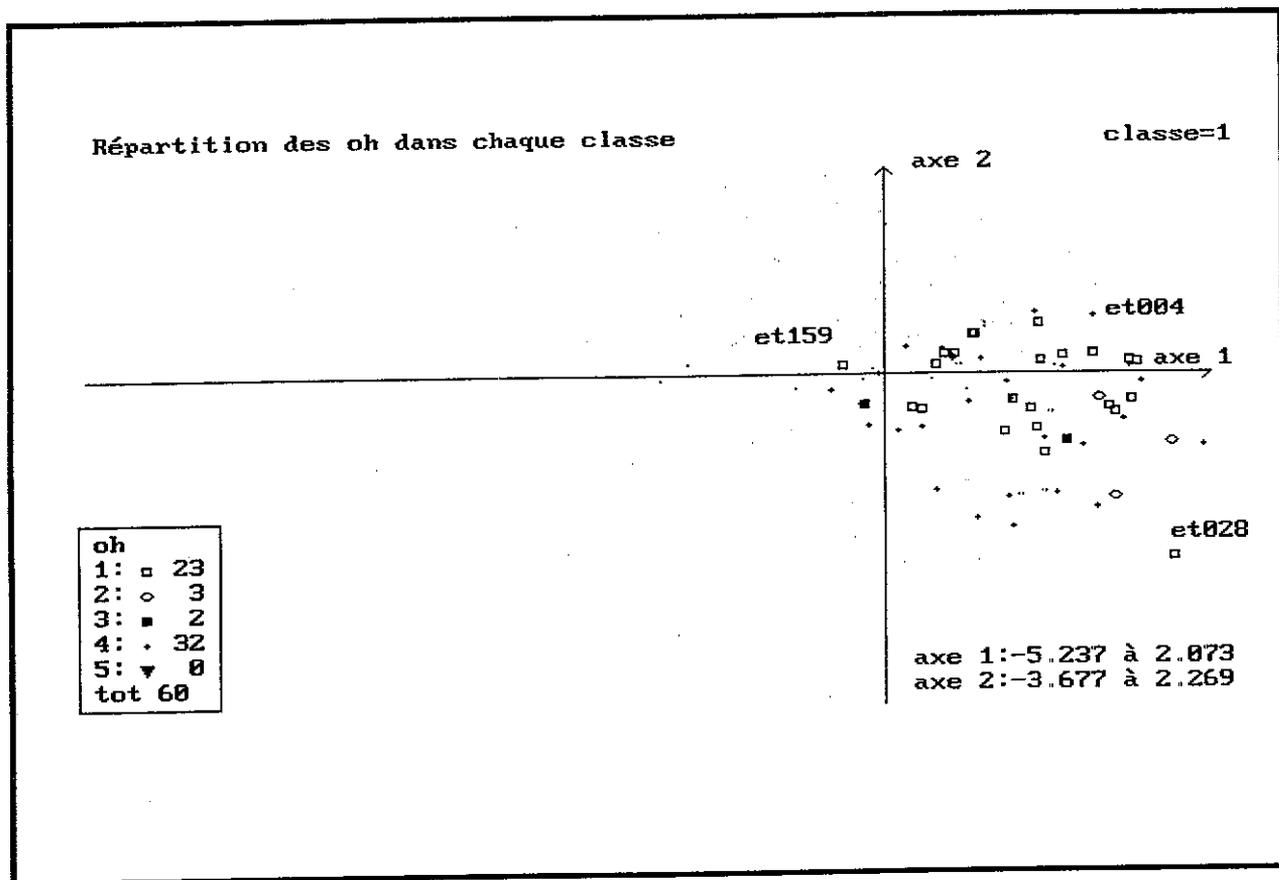
Graphique 3

Représentation sur le plan factoriel des étoiles possédant à la fois la modalité 1 de la variable oh (oh1) et la modalité 1 de la variable Sio (Sio1) dans le plan factoriel.



Graphique 4

Répartition de la variable oh dans la classe 1, visualisation sur le plan factoriel. Les étoiles représentées sont celles de la classe 1. Un symbole différent a été affecté à chacune selon sa valeur en oh. La légende des symboles fournit les effectifs de chaque modalité dans la classe 1.



BIBLIOGRAPHIE

- [1] LEAS : Logiciel d'enseignement et d'analyses statistique. Unité de Biométrie
9, place Pierre Viala - 34000 Montpellier.
- [2] SICLA: Système interactif de classification automatique. CISIA - 1, avenue Herbillon
94160 Saint-Mandé.
- [3] SPAD: Système portable d'analyse de données. CISIA - 1, avenue Herbillon
94160 Saint-Mandé.
- [4] F.GOUPIL-TESTU et M.O. MENNESSIER (1993)- Cinématique et émission OH,
rencontres de travail « Miras » - GRAAL: Groupe de Recherche en
Astronomie et Astrophysique. 34095 Montpellier.

COMMENT SE PROCURER CE PROGRAMME :

S'adresser à Yves Lechevallier à l'IN.R.I.A.