



**HAL**  
open science

# Progiciel INDA pour l'analyse des données : INtelligent Data Analysis

Djamel Belaïd

► **To cite this version:**

Djamel Belaïd. Progiciel INDA pour l'analyse des données : INtelligent Data Analysis. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1985. Français. NNT : . tel-00318680

**HAL Id: tel-00318680**

**<https://theses.hal.science/tel-00318680>**

Submitted on 4 Sep 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE

*présentée à*

**l'Université Scientifique et Médicale de Grenoble**

*et à*

**l'Institut National Polytechnique de Grenoble**

*pour obtenir le grade de*  
**DOCTEUR DE 3ème CYCLE**  
**«Mathématiques appliquées»**  
**(analyse numérique)**

*par*

**Djamel BELAID**



**PROGICIEL INDA POUR L'ANALYSE DES DONNEES:**

**INtelligent Data Analysis.**



**Thèse soutenue le 28 octobre 1985 devant la commission d'examen.**

<b>G. SAUCIER</b>	<b>Président</b>
<b>M. ADIBA</b>	
<b>J.R. BARRA</b>	
<b>M. BECKER</b>	<b>Examineurs</b>
<b>F. CHATELIN</b>	
<b>P. CHENIN</b>	



**UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE**

**Année universitaire 1982-1983**

**Président de l'Université : M. TANCHE**

**MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.**

**(RANG A)**

**SAUF ENSEIGNANTS EN MEDECINE ET PHARMACIE**

**PROFESSEURS DE 1<sup>ère</sup> CLASSE**

<b>ARNAUD Paul</b>	<b>Chimie organique</b>
<b>ARVIEU Robert</b>	<b>Physique nucléaire I.S.N.</b>
<b>AUBERT Guy</b>	<b>Physique C.N.R.S.</b>
<b>AYANT Yves</b>	<b>Physique approfondie</b>
<b>BARBIER Marie-Jeanne</b>	<b>Electrochimie</b>
<b>BARBIER Jean-Claude</b>	<b>Physique expérimentale C.N.R.S. (labo de magnétisme)</b>
<b>BARJON Robert</b>	<b>Physique nucléaire I.S.N.</b>
<b>BARNOUD Fernand</b>	<b>Biosynthèse de la cellulose-Biologie</b>
<b>BARRA Jean-René</b>	<b>Statistiques - Mathématiques appliquées</b>
<b>BELORISKY Elie</b>	<b>Physique</b>
<b>BENZAKEN Claude (M.)</b>	<b>Mathématiques pures</b>
<b>BERNARD Alain</b>	<b>Mathématiques pures</b>
<b>BERTRANDIAS Françoise</b>	<b>Mathématiques pures</b>
<b>BERTRANDIAS Jean-Paul</b>	<b>Mathématiques pures</b>
<b>BILLET Jean</b>	<b>Géographie</b>
<b>BONNIER Jean-Marie</b>	<b>Chimie générale</b>
<b>BOUCHEZ Robert</b>	<b>Physique nucléaire I.S.N.</b>
<b>BRAVARD Yves</b>	<b>Géographie</b>
<b>CARLIER Georges</b>	<b>Biologie végétale</b>
<b>CAUQUIS Georges</b>	<b>Chimie organique</b>
<b>CHIBON Pierre</b>	<b>Biologie animale</b>
<b>COLIN DE VERDIERE Yves</b>	<b>Mathématiques pures</b>
<b>CRABBE Pierre (détaché)</b>	<b>C.E.R.M.O.</b>
<b>CYROT Michel</b>	<b>Physique du solide</b>
<b>DAUMAS Max</b>	<b>Géographie</b>
<b>DEBELMAS Jacques</b>	<b>Géologie générale</b>
<b>DEGRANGE Charles</b>	<b>Zoologie</b>
<b>DELOBEL Claude (M.)</b>	<b>M.I.A.G. Mathématiques appliquées</b>
<b>DEPORTES Charles</b>	<b>Chimie minérale</b>
<b>DESRE Pierre</b>	<b>Electrochimie</b>
<b>DOLIQUE Jean-Michel</b>	<b>Physique des plasmas</b>
<b>DUCROS Pierre</b>	<b>Cristallographie</b>
<b>FONTAINE Jean-Marc</b>	<b>Mathématiques pures</b>
<b>GAGNAIRE Didier</b>	<b>Chimie physique</b>

.../...

GASTINEL Noël	Analyse numérique - Mathématiques appliquées
GERBER Robert	Mathématiques pures
GERMAIN Jean-Pierre	Mécanique
GIRAUD Pierre	Géologie
IDELMAN Simon	Physiologie animale
JANIN Bernard	Géographie
JOLY Jean-René	Mathématiques pures
JULLIEN Pierre	Mathématiques appliquées
KAHANE André (détaché DAFCO)	Physique
KAHANE Josette	Physique
KOSZUL Jean-Louis	Mathématiques pures
KRAKOWIAK Sacha	Mathématiques appliquées
KUPTA Yvon	Mathématiques pures
LACAZE Albert	Thermodynamique
LAJZEROWICZ Jeannine	Physique
LAJZEROWICZ Joseph	Physique
LAURENT Pierre	Mathématiques appliquées
DE LEIRIS Joël	Biologie
LLIBOUTRY Louis	Géophysique
LOISEAUX Jean-Marie	Sciences nucléaires I.S.N.
LOUP Jean	Géographie
MACHE Régis	Physiologie végétale
MAYNARD Roger	Physique du solide
MICHEL Robert	Minéralogie et pétrographie (géologie)
MOZIERES Philippe	Spectrométrie - Physique
OMONT Alain	Astrophysique
OZENDA Paul	Botanique (biologie végétale)
PAYAN Jean-Jacques (détaché)	Mathématiques pures
PEBAY PEYROULA Jean-Claude	Physique
PERRIAUX Jacques	Géologie
PERRIER Guy	Géophysique
PIERRARD Jean-Marie	Mécanique
RASSAT André	Chimie systématique
RENARD Michel	Thermodynamique
RICHARD Lucien	Biologie végétale
RINAUDO Marguerite	Chimie CERMAV
SENGEL Philippe	Biologie animale
SERGERAERT Francis	Mathématiques pures
SOUTIF Michel	Physique
VAILLANT François	Zoologie
VALENTIN Jacques	Physique nucléaire I.S.N.
VAN CUTSEN Bernard	Mathématiques appliquées
VAUQUOIS Bernard	Mathématiques appliquées
VIALON Pierre	Géologie

#### PROFESSEURS DE 2ème CLASSE

ADIBA Michel	Mathématiques pures
ARMAND Gilbert	Géographie

AURIAULT Jean-Louis	Mécanique
BEGUIN Claude (M.)	Chimie organique
BOEHLER Jean-Paul	Mécanique
BOITET Christian	Mathématiques appliquées
BORNAREL Jean	Physique
BRUN Gilbert	Biologie
CASTAING Bernard	Physique
CHARDON Michel	Géographie
COHENADLAD Jean-Pierre	Physique
DENEUVILLE Alain	Physique
DEPASSEL Roger	Mécanique des fluides
DOUCE Roland	Physiologie végétale
DUFRESNOY Alain	Mathématiques pures
GASPARD François	Physique
GAUTRON René	Chimie
GIDON Maurice	Géologie
GIGNOUX Claude (M.)	Sciences nucléaires I.S.N.
GUITTON Jacques	Chimie
HACQUES Gérard	Mathématiques appliquées
HERBIN Jacky	Géographie
HICTER Pierre	Chimie
JOSELEAU Jean-Paul	Biochimie
KERCKOVE Claude (M.)	Géologie
LE BRETON Alain	Mathématiques appliquées
LONGEQUEUE Nicole	Sciences nucléaires I.S.N.
LUCAS Robert	Physiques
LUNA Domingo	Mathématiques pures
MASCLE Georges	Géologie
NEMOZ Alain	Thermodynamique (CNRS - CRTBT)
OUDET Bruno	Mathématiques appliquées
PELMONT Jean	Biochimie
PERRIN Claude (M.)	Sciences nucléaires I.S.N.
PFISTER Jean-Claude (détaché)	Physique du solide
PIBOULE Michel	Géologie
PIERRE Jean-Louis	Chimie organique
RAYNAUD Hervé	Mathématiques appliquées
ROBERT Gilles	Mathématiques pures
ROBERT Jean-Bernard	Chimie physique
ROSSI André	Physiologie végétale
SAKAROVITCH Michel	Mathématiques appliquées
SARROT REYNAUD Jean	Géologie
SAXOD Raymond	Biologie animale
SOUTIF Jeanne	Physique
SCHOOL Pierre-Claude	Mathématiques appliquées
STUTZ Pierre	Mécanique
SUBRA Robert	Chimie
VIDAL Michel	Chimie organique
VIVIAN Robert	Géographie



**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**

**Année universitaire 1982-1983**

**Président de l'Université : D. BLOCH**

**Vice-Président : René CARRE  
Hervé CHERADAME  
Marcel IVANES**

**PROFESSEURS DES UNIVERSITES :**

<b>ANCEAU François</b>	<b>E.N.S.I.M.A.G.</b>
<b>BARRAUD Alain</b>	<b>E.N.S.I.E.G.</b>
<b>BAUDELET Bernard</b>	<b>E.N.S.I.E.G.</b>
<b>BESSON Jean</b>	<b>E.N.S.E.E.G.</b>
<b>BLIMAN Samuel</b>	<b>E.N.S.E.R.G.</b>
<b>BLOCH Daniel</b>	<b>E.N.S.I.E.G.</b>
<b>BOIS Philippe</b>	<b>E.N.S.H.G.</b>
<b>BONNETAIN Lucien</b>	<b>E.N.S.E.E.G.</b>
<b>BONNIER Etienne</b>	<b>E.N.S.E.E.G.</b>
<b>BOUVARD Maurice</b>	<b>E.N.S.H.G.</b>
<b>BRISSONNEAU Pierre</b>	<b>E.N.S.I.E.G.</b>
<b>BUYLE BODIN Maurice</b>	<b>E.N.S.E.R.G.</b>
<b>CAVAIGNAC Jean-François</b>	<b>E.N.S.I.E.G.</b>
<b>CHARTIER Germain</b>	<b>E.N.S.I.E.G.</b>
<b>CHENEVIER Pierre</b>	<b>E.N.S.E.R.G.</b>
<b>CHERADAME Hervé</b>	<b>U.E.R.M.C.P.P.</b>
<b>CHERUY Arlette</b>	<b>E.N.S.I.E.G.</b>
<b>CHIAVERINA Jean</b>	<b>U.E.R.M.C.P.P.</b>
<b>COHEN Joseph</b>	<b>E.N.S.E.R.G.</b>
<b>COUMES André</b>	<b>E.N.S.E.R.G.</b>
<b>DURAND Francis</b>	<b>E.N.S.E.E.G.</b>
<b>DURAND Jean-Louis</b>	<b>E.N.S.I.E.G.</b>
<b>FELICI Noël</b>	<b>E.N.S.I.E.G.</b>
<b>FOULARD Claude</b>	<b>E.N.S.I.E.G.</b>
<b>GENTIL Pierre</b>	<b>E.N.S.E.R.G.</b>
<b>GUERIN Bernard</b>	<b>E.N.S.E.R.G.</b>
<b>GUYOT Pierre</b>	<b>E.N.S.E.E.G.</b>
<b>IVANES Marcel</b>	<b>E.N.S.I.E.G.</b>
<b>JAUSSAUD Pierre</b>	<b>E.N.S.I.E.G.</b>
<b>JOUBERT Jean-Claude</b>	<b>E.N.S.I.E.G.</b>
<b>JOURDAIN Geneviève</b>	<b>E.N.S.I.E.G.</b>
<b>LACOUME Jean-Louis</b>	<b>E.N.S.I.E.G.</b>
<b>LATOMBE Jean-Claude</b>	<b>E.N.S.I.M.A.G.</b>

LESSIEUR Marcel	E.N.S.H.G.
LESPINARD Georges	E.N.S.H.G.
LONGEQUEUE Jean-Pierre	E.N.S.I.E.G.
MAZARE Guy	E.N.S.I.M.A.G.
MOREAU René	E.N.S.H.G.
MORET Roger	E.N.S.I.E.G.
MOSSIERE Jacques	E.N.S.I.M.A.G.
PARIAUD Jean-Charles	E.N.S.E.E.G.
PAUTHENET René	E.N.S.I.E.G.
PERRET René	E.N.S.I.E.G.
PERRET Robert	E.N.S.I.E.G.
PIAU Jean-Michel	E.N.S.H.G.
POLOUJADOFF Michel	E.N.S.I.E.G.
POUPOT Christian	E.N.S.E.R.G.
RAMEAU Jean-Jacques	E.N.S.E.E.G.
RENAUD Maurice	U.E.R.M.C.P.P.
ROBERT André	U.E.R.M.C.P.P.
ROBERT François	E.N.S.I.M.A.G.
SABONNADIÈRE Jean-Claude	E.N.S.I.E.G.
SAUCIER Gabrielle	E.N.S.I.M.A.G.
SCHLENKER Claire	E.N.S.I.E.G.
SCHLENKER Michel	E.N.S.I.E.G.
SERMET Pierre	E.N.S.E.R.G.
SILVY Jacques	U.E.R.M.C.P.P.
SOHM Jean-Claude	E.N.S.E.E.G.
SOUQUET Jean-Louis	E.N.S.E.E.G.
VEILLON Gérard	E.N.S.I.M.A.G.
ZADWORNY François	E.N.S.E.R.G.

#### PROFESSEURS ASSOCIES

BASTIN Georges	E.N.S.H.G.
BERRIL John	E.N.S.H.G.
CARREAU Pierre	E.N.S.H.G.
GANDINI Alessandro	U.E.R.M.C.P.P.
HAYASHI Hirashi	E.N.S.I.E.G.

#### PROFESSEURS UNIVERSITE DES SCIENCES SOCIALES (Grenoble II)

BOLLIET Louis  
Chatelin Françoise

#### PROFESSEURS E.N.S. Mines de Saint-Etienne

RIEU Jean  
SOUSTELLE Michel

#### CHERCHEURS DU C.N.R.S.

FRUCHART Robert  
VACHAUD Georges

Directeur de Recherche  
Directeur de Recherche

.../...

<b>ALLIBERT Michel</b>	<b>Maître de Recherche</b>
<b>ANSARA Ibrahim</b>	<b>Maître de Recherche</b>
<b>ARMAND Michel</b>	<b>Maître de Recherche</b>
<b>BINDER Gilbert</b>	
<b>CARRE René</b>	<b>Maître de Recherche</b>
<b>DAVID René</b>	<b>Maître de Recherche</b>
<b>DEPORTES Jacques</b>	
<b>DRIOLE Jean</b>	<b>Maître de Recherche</b>
<b>GIGNOUX Damien</b>	
<b>GIVORD Dominique</b>	
<b>GUELIN Pierre</b>	
<b>HOPFINGER Emil</b>	<b>Maître de Recherche</b>
<b>JOUD Jean-Charles</b>	<b>Maître de Recherche</b>
<b>KAMARINOS Georges</b>	<b>Maître de Recherche</b>
<b>KLEITZ Michel</b>	<b>Maître de Recherche</b>
<b>LANDAU Ioan-Dore</b>	<b>Maître de Recherche</b>
<b>LASJAUNIAS J.C.</b>	
<b>MERMET Jean</b>	<b>Maître de Recherche</b>
<b>MUNIER Jacques</b>	<b>Maître de Recherche</b>
<b>PIAU Monique</b>	
<b>PORTESEIL Jean-Louis</b>	
<b>THOLENCE Jean-Louis</b>	
<b>VERDILLON André</b>	

**CHERCHEURS du MINISTERE de la RECHERCHE et de la TECHNOLOGIE (Directeurs et Maîtres de Recherches, ENS Mines de St. Etienne)**

<b>LESBATS Pierre</b>	<b>Directeur de Recherche</b>
<b>BISCONDI Michel</b>	<b>Maître de Recherche</b>
<b>KOBYLANSKI André</b>	<b>Maître de Recherche</b>
<b>LE COZE Jean</b>	<b>Maître de Recherche</b>
<b>LALAUZE René</b>	<b>Maître de Recherche</b>
<b>LANCELOT Francis</b>	<b>Maître de Recherche</b>
<b>THEVENOT François</b>	<b>Maître de Recherche</b>
<b>TRAN MINH Canh</b>	<b>Maître de Recherche</b>

**PERSONNALITES HABILITEES à DIRIGER des TRAVAUX de RECHERCHE (Décision du Conseil Scientifique)**

<b>ALLIBERT Colette</b>	<b>E.N.S.E.E.G.</b>
<b>BERNARD Claude</b>	<b>E.N.S.E.E.G.</b>
<b>BONNET Rolland</b>	<b>E.N.S.E.E.G.</b>
<b>CAILLET Marcel</b>	<b>E.N.S.E.E.G.</b>
<b>CHATILLON Catherine</b>	<b>E.N.S.E.E.G.</b>
<b>CHATILLON Christian</b>	<b>E.N.S.E.E.G.</b>
<b>COULON Michel</b>	<b>E.N.S.E.E.G.</b>
<b>DIARD Jean-Paul</b>	<b>E.N.S.E.E.G.</b>
<b>EUSTAPOPOULOS Nicolas</b>	<b>E.N.S.E.E.G.</b>
<b>FOSTER Panayotis</b>	<b>E.N.S.E.E.G.</b>

.../...

GALERIE Alain	E.N.S.E.E.G.
HAMMOU Abdelkader	E.N.S.E.E.G.
MALMEJAC Yves	E.N.S.E.E.G. (CENG)
MARTIN GARIN Régina	E.N.S.E.E.G.
NGUYEN TRUONG Bernadette	E.N.S.E.E.G.
RAVAINE Denis	E.N.S.E.E.G.
SAINFORT	E.N.S.E.E.G. (CENG)
SARRAZIN Pierre	E.N.S.E.E.G.
SIMON Jean-Paul	E.N.S.E.E.G.
TOUZAIN Philippe	E.N.S.E.E.G.
URBAIN Georges	E.N.S.E.E.G. (Laboratoire des ultra-réfractaires ODEILLON)
GUILHOT Bernard	E.N.S. Mines Saint Etienne
THOMAS Gérard	E.N.S. Mines Saint Etienne
DRIVER Julien	E.N.S. Mines Saint Etienne
BARIBAUD Michel	E.N.S.E.R.G.
BOREL Joseph	E.N.S.E.R.G.
CHOVET Alain	E.N.S.E.R.G.
CHEHIKIAN Alain	E.N.S.E.R.G.
DOLMAZON Jean-Marc	E.N.S.E.R.G.
HERAULT Jeanny	E.N.S.E.R.G.
MONLLOR Christian	E.N.S.E.R.G.
BORNARD Guy	E.N.S.I.E.G.
DESCHIZEAU Pierre	E.N.S.I.E.G.
GLANGEAUD François	E.N.S.I.E.G.
KOFMAN Walter	E.N.S.I.E.G.
LEJEUNE Gérard	E.N.S.I.E.G.
MAZUER Jean	E.N.S.I.E.G.
PERARD Jacques	E.N.S.I.E.G.
REINISCH Raymond	E.N.S.I.E.G.
ALEMANY Antoine	E.N.S.H.G.
BOIS Daniel	E.N.S.H.G.
DARVE Félix	E.N.S.H.G.
MICHEL Jean-Marie	E.N.S.H.G.
OBLED Charles	E.N.S.H.G.
ROWE Alain	E.N.S.H.G.
VAUCLIN Michel	E.N.S.H.G.
WACK Bernard	E.N.S.H.G.
BERT Didier	E.N.S.I.M.A.G.
CALMET Jacques	E.N.S.I.M.A.G.
COURTIN Jacques	E.N.S.I.M.A.G.
COURTOIS Bernard	E.N.S.I.M.A.G.
DELLA DORA Jean	E.N.S.I.M.A.G.
FONLUPT Jean	E.N.S.I.M.A.G.
SIFAKIS Joseph	E.N.S.I.M.A.G.
CHARUEL Robert	U.E.R.M.C.P.P.
CADET Jean	C.E.N.G.
COEURE Philippe	C.E.N.G. (LETI)

.../...

<b>DELHAYE Jean-Marc</b>	<b>C.E.N.G. (STT)</b>
<b>DUPUY Michel</b>	<b>C.E.N.G. (LETI)</b>
<b>JOUVE Hubert</b>	<b>C.E.N.G. (LETI)</b>
<b>NICOLAU Yvan</b>	<b>C.E.N.G. (LETI)</b>
<b>NIFENECKER Hervé</b>	<b>C.E.N.G.</b>
<b>PERROUD Paul</b>	<b>C.E.N.G.</b>
<b>PEUZIN Jean-Claude</b>	<b>C.E.N.G. (LETI)</b>
<b>TAIEB Maurice</b>	<b>C.E.N.G.</b>
<b>VINCENDON Marc</b>	<b>C.E.N.G.</b>

**LABORATOIRES EXTERIEURS**

<b>DEMOULIN Eric</b>	<b>C.N.E.T.</b>
<b>DEVINE</b>	<b>C.N.E.T. (R.A.B.)</b>
<b>GERBER Roland</b>	<b>C.N.E.T.</b>
<b>MERCKEL Gérard</b>	<b>C.N.E.T.</b>
<b>PAULEAU Yves</b>	<b>C.N.E.T.</b>
<b>GAUBERT C.</b>	<b>I.N.S.A. Lyon</b>



*"Encore si, ce que l'on voit, on le voyait bien, ce serait toujours autant de connu mais on le voit tout autrement qu'il n'est. Ainsi les vrais philosophes passent leur vie à ne point croire ce qu'ils voient et à tâcher de deviner ce qu'ils ne voient point".*

*Fontenelle*

*"Entretiens sur la pluralité des Mondes".*



à ma famille,  
à Christine et mes amis.



Je tiens à exprimer ma profonde gratitude à Madame Françoise CHATELIN, pour m'avoir accueilli dans son équipe et dirigé ce travail. Je lui dois beaucoup.

Que Madame Gabrielle SAUCIER reçoive ma vive reconnaissance pour avoir accepté de présider ce jury.

Monsieur Jean René BARRA me fait un grand honneur de siéger à ce jury ; qu'il en soit remercié et qu'il reçoive ma gratitude pour les discussions fructueuses qu'il m'a accordées.

J'adresse à Madame Monique BECKER d'abord mes amitiés et puis ma reconnaissance pour faire partie du jury.

Je suis sensible à l'honneur que me fait Monsieur Michel ADIBA, Directeur de la MIAG de Grenoble, en acceptant de participer à ce jury.

J'adresse mes sincères remerciements à Monsieur Patrick CHENIN, pour sa présence parmi ce jury, ainsi qu'à son équipe qui m'a bien accueilli et m'a permis d'utiliser son matériel. Je tiens à citer deux noms dans cette équipe, Claire DI CRESCENZO et André EBERHARD pour qui j'exprime toute ma reconnaissance et mon profond respect.

Mes remerciements vont également à :

NAG (Numerical Algorithms Group Limited, Oxford, G.B.) pour leur support financier du projet.

Monsieur Gérard VEILLON, Directeur de l'ENSIMAG, pour m'avoir prêté le matériel de son Ecole.

Mes vifs remerciements vont à Madame Geneviève BICAIS pour sa grande compétence dans la dactylographie et à Messieurs Daniel IGLESIAS et Claude ANGUILLE, du service reprographie, pour l'excellente qualité de leur travail et pour la sympathie qu'ils m'ont témoignée.

Enfin je n'oublierai pas mes amis pour qui leurs places sont réservées. Leurs noms ne sont pas sur ce papier mais leurs images sont dans mon esprit.



## SOMMAIRE

INTRODUCTION -----	1
1 - PRESENTATION GENERALE DE INDA -----	5
2 - MODULE D'ACQUISITION DES DONNEES : INDA_AD ---	17
3 - MODULE D'EXECUTION DES METHODES D'ANALYSES : INDA_EM -----	29
4 - MODULE GRAPHIQUE : INDA_GR -----	55
CONCLUSION -----	67
TRAITEMENT D'UN EXEMPLE -----	71
BIBLIOGRAPHIE -----	81



## *INTRODUCTION*



Dans de très nombreux domaines scientifiques, tels que les sciences sociales (économie, psychologie...), les sciences de la nature (météorologie, géologie...), les sciences humaines et autres, les chercheurs sont amenés très souvent à utiliser des techniques statistiques et particulièrement l'analyse des données (ou méthodes statistiques descriptives multidimensionnelles).

Ces techniques d'analyse de données se regroupent principalement en deux familles, celle des méthodes de classification et celle des méthodes factorielles.

On présente dans ce travail un logiciel interactif sur micro-ordinateur, pour cette deuxième famille, destiné en particulier aux nombreux utilisateurs isolés qui ne peuvent bénéficier des services d'un centre de calcul, ainsi qu'à tous ceux qui sont confrontés au coût prohibitif et à l'absence d'interactivité des gros progiciels statistiques de traitement de données, du type SPAD, BMDP, SPSS.

C'est ainsi, dans le laboratoire LCS (Laboratoire Circuits et Systèmes, dirigé par le Professeur G. SAUCIER), des spécialistes de l'analyse des données, de l'analyse numérique et d'architectures informatiques ont décidé d'unir leurs efforts sur un projet de construction d'une machine spécialisée en analyse de données afin de contribuer à aider les utilisateurs isolés pour leur traitement de données.

La construction de cette machine, est un projet à long terme pour lequel dans un premier temps nous avons écrit un logiciel de techniques statistiques (méthodes factorielles) sur micro-ordinateur sous la direction de Madame F. CHATELIN pour l'aspect numérique. La deuxième phase consiste en l'écriture d'une interface d'aide au choix de la méthode à utiliser pour le traitement des données. Ce travail est fait par C. MAZEL, sous la direction de Madame M. BECKER et de Monsieur J.R. BARRA, sur financement partiel de NAG, Numerical Algorithms Group,

Oxford. Une autre phase est la réalisation d'un réseau systolique effectuant certains modules numériques de la bibliothèque INDAL du logiciel. Travail en cours de réalisation par T. DELESPIERRE (bourse DGRST) et G. CHEVALLIER (bourse C<sup>3</sup>), sous la direction de M. BECKER, F. CHATELIN, G. SAUCIER et M. TCHUENTE.

L'ensemble du logiciel porte le nom de INDA,  
Intelligent Data Analysis.

## CHAPITRE 1 :

### PRESENTATION GENERALE DE *INDA*

1.1 - Menu des méthodes -----	7
1.2 - Aide au choix -----	9
1.3 - Acquisition des données -----	10
1.4 - Exécution des méthodes -----	11
1.5 - Sortie graphique -----	12
1.6 - Fonctionnement de INDA -----	13
1.7 - Principe du calcul par blocs.-----	14



Nous donnons dans ce chapitre une présentation générale de INDA en décrivant globalement son fonctionnement et les modules qui le constituent.

### 1.1 - MENU DES METHODES :

Les méthodes factorielles contenues dans INDA sont les suivantes :

#### Groupe 1

ACP : Analyse en composante principale d'un nuage de points  
trois possibilités selon la norme choisie :

- ACP ..... CM=1
- ACP unitaire ..... CM=2
- ACP normée ..... CM=3

ACP : Analyse en composante principale d'un vecteur aléatoire  
relativement à une norme  
trois possibilités selon la norme choisie :

- ACP ..... CM=4
- ACP unitaire ..... CM=5
- ACP normée ..... CM=6

#### Groupe 2

AFC : Analyse factorielle des correspondances

- sur les lignes ..... CM=7
- sur les colonnes ..... CM=8

Corrélation canonique entre deux variables

aléatoires discrètes ..... CM=9

Groupe 3

AFD : Analyse factorielle discriminante ..... CM=10

Corrélation canonique entre un vecteur aléatoire  
et une variable aléatoire finie ..... CM=11

Groupe 4

Corrélation canonique entre deux vecteurs aléatoires CM=12

CM est un paramètre de type <integer> qui donne le choix de la méthode dans le MENU.

Les méthodes du MENU de INDA nécessitent des calculs algébriques communs [ 9 ].

- Factorisation de CHOLESKI
- Inversion de matrices triangulaires
- Calcul de quelques unes des valeurs propres dominantes et des vecteurs propres associés.

L'ensemble de ces procédures numériques constitue la bibliothèque INDAL du logiciel (INDA Library).

## 1.2 - AIDE DU CHOIX (INDA\_CM)

Le choix de la méthode, à utiliser pour analyser des données, est fixé par l'utilisateur, en donnant une valeur au paramètre CM, ou en appelant le module INDA\_CM d'assistance "intelligente". Ce module a été écrit par C. MAZEL [7], sous la direction de M. BECKER et de J.R. BARRA dans le cadre d'une collaboration avec NAG (Numerical Algorithms Group Limited, Oxford. Great Britain). Le module d'assistance "intelligente" permet de choisir entre les différentes techniques décrites au menu. Il est en quelque sorte une "interface intelligente" entre un utilisateur non expert en analyse de données et le module d'exécution des méthodes.

Son principe est de lire des fichiers questionnaires dont les noms comportent des indices, de noter les réponses de l'utilisateur, et de continuer le dialogue sur le fichier dont les indices correspondent aux réponses données.

Ainsi, il parcourt un graphe de questionnaires. Le fait de ranger les questionnaires dans des fichiers indépendants confère à ce module une grande généralité et une grande modularité. Le principe des systèmes experts, qui consiste à séparer la connaissance du raisonnement, est donc bien respecté.

### 1.3 - ACQUISITION DES DONNEES : (INDA AD)

Pour l'exécution d'une technique, un fichier standard contenant les données relatives à la méthode choisie est nécessaire.

Le rôle du module d'acquisition des données est de transformer un fichier écrit par l'utilisateur avec l'éditeur de texte sous la forme d'un fichier, standard linéaire de réels, dans lequel les données se présentent dans un ordre précis.

Au niveau de ce module, nous supposons au préalable que l'utilisateur dispose au moins d'un fichier initial de type <text> sur lequel il a rentré ses données.

Ce module demande à l'utilisateur s'il possède déjà un fichier standard, fichier crée par ce module lors d'une précédente acquisition, et directement utilisable par le module d'exécution. Si oui, il enregistre son nom. Sinon, il lui donne des informations quand aux données nécessaires à la méthode choisie, lui pose des questions relatives à la structure de son fichier <text> (i.e. positions des tableaux de données, tels que observations, norme, poids..., des uns par rapport aux autres), à la nature des tableaux de données et leurs dimensions. Enfin, quand la méthode choisie le permet, on termine en proposant à l'utilisateur de donner aux lignes et aux colonnes de la matrice des observations, des étiquettes de son choix ou des étiquettes standards (1,2,3,...).

Le résultat de l'exécution de ce module est un fichier linéaire de données, dit standard (voir la structure du fichier standard à la section 2.2), de type <file of real>, et éventuellement deux fichiers linéaires d'étiquettes, de type <file of string> , l'un pour les lignes, l'autre pour les colonnes.

Ces fichiers ne sont donc pas des fichiers de type <text> mais de type <file of ...> et ceci est dans le but d'avoir un stockage de données plus économique et de permettre une lecture beaucoup plus rapide, non forcément séquentielle, par le module d'exécution. Ces fichiers ne peuvent donc pas être lus directement par l'utilisateur.

1.4 - EXECUTION DES METHODES : (INDA\_EM)

Le module d'exécution des méthodes du menu comporte des calculs préliminaires qui dépendent de la technique, le noyau commun de calcul des valeurs et vecteurs propres, et des calculs de projections sur les axes factoriels.

Son traitement numérique est le suivant :

Soient  $X$  une matrice de taille  $n \times k$ ,  $n \geq k$ , représentant les points à étudier,  $A$  et  $B$  deux matrices, carrées symétriques et définies positives, d'ordre respectivement  $n$  et  $k$ , définissant les masses affectées aux points et la métrique de l'espace de ces derniers ; ces matrices sont définies par la technique ou données par l'utilisateur.

On décompose  $A$  (resp.  $B$ ) en  $A = A^{1/2} A^{1/2t}$  (resp.  $B = B^{1/2} B^{1/2t}$ ) où  $A^{1/2}$  (resp.  $B^{1/2}$ ) est triangulaire inférieure à diagonale positive.

Nous avons choisis la décomposition par la factorisation de CHOLESKI, les matrices  $A$  et  $B$  étant symétriques et définies positives.

On pose  $E = A^{1/2t} X^t B^{1/2}$ ,  $W = E^t E$ .

La matrice symétrique et définie positive  $W$  admet des valeurs propres positives  $\lambda_1 \geq \dots \geq \lambda_k \geq 0$  et des vecteurs propres associés  $w_i$  normalisés par  $w_i^t w_i = 1$ . On définit des formes linéaires, construites sur les vecteurs propres,  $f_i = \lambda_i^{1/2} B^{1/2} w_i$  et  $g_i = \lambda_i^{-1/2} A X^t f_i$   $i = 1 \dots k$ . Ces formes linéaires définissent comme on le verra les projections des colonnes (resp. des lignes) de  $X$  sur les vecteurs propres  $w_i$ , dits axes factoriels, relativement à la norme  $A$  (resp.  $B$ ).

Les techniques d'analyse de données nécessitent donc de calculer, pour différents choix des matrices  $A$  et  $B$ , la matrice  $W$ , quelques uns de ses éléments propres dominants  $(\lambda_i, w_i)$   $i = 1, \dots, r \leq k$  et les formes linéaires  $f_i$  et  $g_i$   $i = 1 \dots r$  associées aux éléments propres dominants.

### 1.5 - SORTIE GRAPHIQUE : (INDA\_GR)

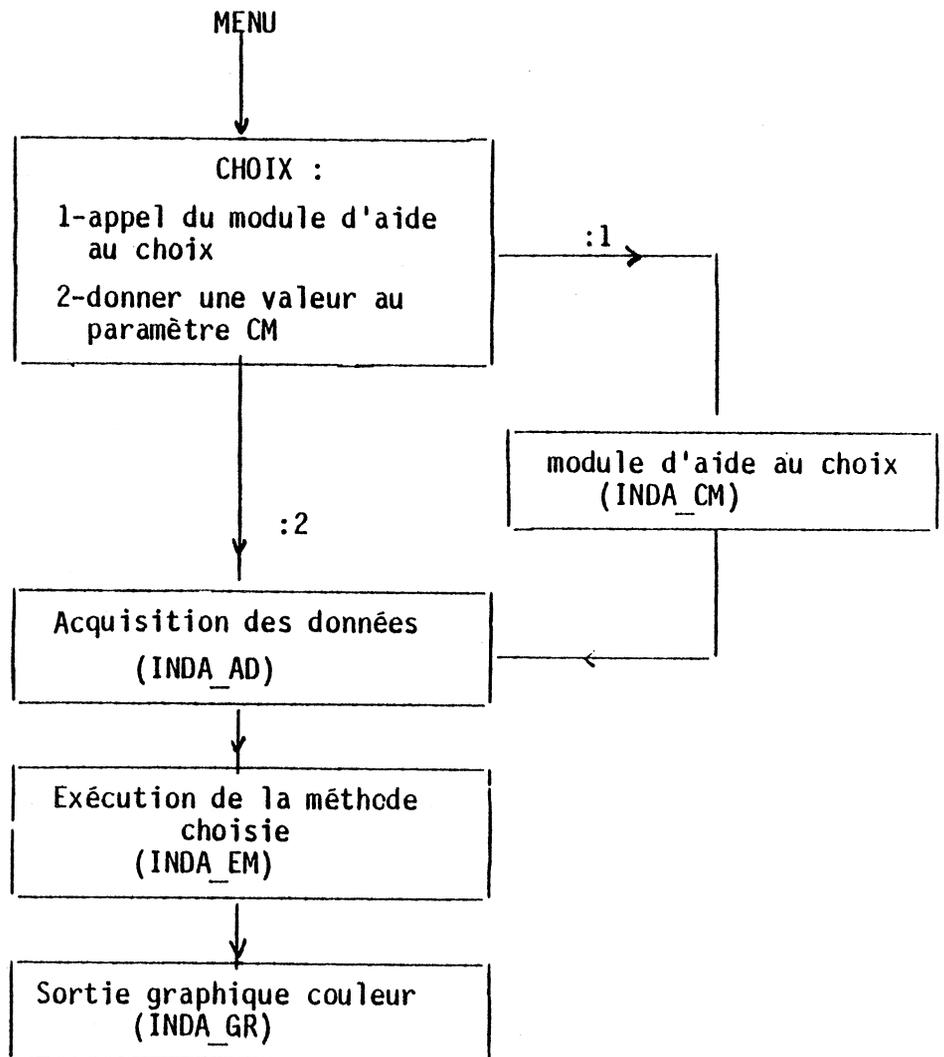
La fin du fonctionnement de INDA\_EM, produit un fichier réel NOMFICH.RES contenant les coordonnées des projections des données sur les axes factoriels. Le module INDA\_GR opère sur ce fichier, après le choix du plan factoriel de visualisation des projections par l'utilisateur. Il permet de tracer le nuage sur un écran graphique couleur ou noir et blanc.

Outre la visualisation du nuage, il nous offre la possibilité d'agrandir une zone de celui-ci (ZOOM graphique, voir section 4.2).

De même, il est possible par l'intermédiaire d'une fonction qu'on pourrait appeler ZOOM Calculatoire, de reconsidérer le nuage initial (données de départ) en redéfinissant les points principaux et les points supplémentaires par déplacement d'une fenêtre graphique et encadrement de ces premiers ou ces derniers par celle-ci. Sur cette nouvelle définition des données de départ, on réactive successivement INDA\_EM et INDA\_GR.

### 1.6 - FONCTIONNEMENT DE INDA

L'analyse d'un tableau de données s'effectue en activant INDA dont l'exécution est l'enchaînement des modules que nous avons décrit précédemment, d'une façon succincte, par le schéma suivant :



### 1.7 - PRINCIPE DU CALCUL PAR BLOCS

Le principe du calcul par blocs est d'astreindre certaines fonctions opérant sur des matrices de grandes tailles aux blocs de celle-ci.

Soit  $X$  une matrice de taille  $n \times k$ ,  $n \geq k$ .

On décompose  $X$  en  $NB$  blocs carrés d'ordre  $k$  et un dernier bloc de taille  $LDB \times k$  en définissant les paramètres  $NB$  et  $LDB$  par la procédure BLOC.

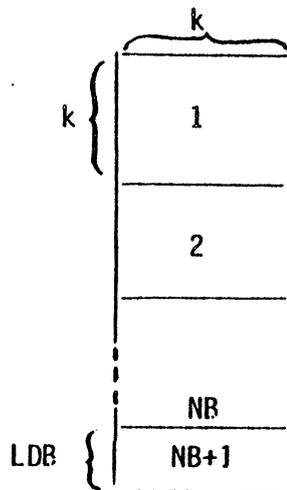
```
procedure BLOC (n:integer; var NB, LDB : integer) ;
```

```
begin
```

```
  NB:=n div k ;      (* nombre de blocs d'ordre k *)
```

```
  LDB:=n-NB*k      (* longueur, en lignes, du dernier bloc *)
```

```
end ;
```



Comme il est bien connu, les calculs matriciels peuvent s'effectuer par blocs, ce qui permet de diminuer la place nécessaire en mémoire centrale du micro ordinateur. Nous avons systématisé la décomposition et le calcul par bloc de la façon qui suit.

Le fonctionnement par blocs d'une action ACT sur une matrice  $M$  de taille  $n \times k$ ,  $n$  grand, stockée dans un fichier NFICH s'illustre, en utilisant une matrice carrée  $BM$  d'ordre  $k < n$ , par la procédure suivante :

```
procédure ACT ;  
var i : entier ;  
    procédure BACT (LB:entier) ; (* LB = longueur du bloc *)  
    début  
        lire (NFICH, BM, LB)      (* lecture dans le fichier NFICH de  
                                   la matrice BM de taille LB*k *)  
        ACT (LB,BM)  
    fin  
  
début  
    Bloc (n,NB,LDB) ;  
    ouvrir NFICH à la lecture ;  
    pour i=1 jusqu'à NB faire FACT (k) ;  
    BACT (LDB) ;  
    fermer NFICH  
fin
```

Ce principe s'applique aux calculs matriciels classiques, ainsi qu'aux autres calculs que nous aurons à effectuer, par exemple les sorties graphiques.

Donnons un exemple d'utilisation de ce principe dans INDA. Soit  $V$  un vecteur de taille  $n$  stocké dans un fichier NF. On désire calculer le minimum des composantes de ce vecteur dont la taille  $n$  peut être très grande par rapport à une constante  $k$ . Par application du principe de calcul par blocs, le minimum  $v_{\min}$  se calcule, très simplement, par la procédure MIN en utilisant seulement un vecteur  $BV$  de taille  $k$ .

```
procédure MIN (var vmin : réel) ;
var i : entier ;
    procédure BMIN (LB : entier) ;
    var j : entier ;
    début
        pour j:=1 jusqu'à LB faire
            si BV[j] < vmin alors vmin := BV[j]
        fin
    début
        Bloc (n,NB,LDB) ;
        ouvrir (NF) ;
        lire (vmin) ;
        fermer(NF) ;
        ouvrir (NF) ;
        pour i=1 jusqu'à NB faire
            début
                Lire (NF,BV,k) ;
                BMIN (k)
            fin
        lire (NF, BV, LDB) ;
        BMIN(LDB) ;
        fermer (NF)
    fin.
```

## CHAPITRE 2 :

### MODULE D' ACQUISITION DES DONNEES *INBA\_AD*

2.1 - Définition du fichier brut -----	21
2.2 - Structure du fichier standard -----	23
2.3 - Procédure TFBFS -----	24
2.4 - Acquisition des étiquettes -----	25
2.5 - Fonctionnement du module.-----	26



Après le choix d'une méthode du MENU, on suppose que l'utilisateur possède un fichier NFB contenant les données nécessaires à l'exécution de la méthode choisie. Ces données sont celles définies par les procédures de calculs préliminaires (voir section 3.3). Ce fichier qu'on appellera brut peut être de type <text> ou de type <file of real>.

Les calculs qu'effectue toute méthode du MENU, et en particulier les calculs préliminaires, utilisent le principe par blocs (voir section 3.3 pour le genre de calcul à effectuer). Afin de faciliter l'application de ce principe, nous devons donner une structure bien définie au fichier de données utilisé par INDA\_EM. Ce fichier structuré sera appelé fichier standard et identifié par NFB.DON.

Le rôle du module d'acquisition des données INDA\_AD est donc de transformer le fichier brut NFB en un fichier standard NFB.DON de type <file of real>.



## 2.1 - DEFINITION DU FICHER BRUT

Pour être accepté par INDA\_AD, le fichier brut, donné par l'utilisateur, doit contenir un minimum d'information, quant aux données nécessaires à l'exécution de la méthode choisie, afin qu'il puisse être transformé en un fichier standard. Des indications à ce propos lui sont données par la procédure INDIC que nous détaillerons par la suite.

Après lecture des indications données par la procédure INDIC, l'utilisateur aura à répondre à celle-ci, en appuyant pour n'importe quelle touche du clavier autre que 'Q', si son fichier brut satisfait les indications données ; INDA\_AD continue alors son travail. Sinon, le fichier brut ne répond pas aux indications données et l'utilisateur est invité à quitter le programme, en appuyant sur la touche 'Q', et modifier son fichier selon les indications reçus.

Les indications sur les données exigées dans le fichier brut diffèrent d'une méthode à l'autre; elles sont données par la même procédure INDIC qui dépend donc de CM, numéro de la méthode du MENU.

procédure INDIC(CM) ;

début

- le fichier brut ne doit contenir que :
    - \* le tableau des observations à étudier (noté O) ;
    - \* si CM = 1,2,3 :  
un vecteur des masses des individus (noté P) ;
    - \* si CM = 1,4 :  
un tableau carré représentant la norme de l'espace des individus (noté M) ;
    - \* si CM = 10, 11 :  
un vecteur représentant les modalités d'une variable discrète (noté J) ;
  - Ces tableaux peuvent être écrits dans n'importe quel ordre, les uns par rapport aux autres.
  - Les informations concernant le nombre d'individus, le nombre de variables, les étiquettes de ceux-ci ... ne doivent pas figurer dans le fichier brut.
- fin ; (INDIC)

Remarque :

- pour CM = 4,5,6 le vecteur P des masses n'est pas demandé ;  
INDA\_AD se charge de le rajouter,  $P[i] = 1 \quad \forall i$ .
- pour CM = 2,5 la norme est rajoutée par INDA\_AD et est égale à la matrice identité
- pour CM = 3,6 la norme est calculée par INDA\_AD et est égale à l'inverse de la variance des variables.

Si la réponse de l'utilisateur est telle que son fichier brut NFB vérifie les indications données, ce fichier est recopié intégralement dans un fichier, identifié par NFB.BRT, de type <file of real> sur lequel ont lieu les transformations pour aboutir au fichier standard. Le fichier NFB est ainsi sauvegardé.

Nous expliciterons par la suite la procédure TFBFS de transformation du fichier brut réel en un fichier standard réel ; mais définis-

## 2.2 - STRUCTURE DU FICHER STANDARD

Suivant les calculs à effectuer dans le progiciel, nous avons choisis une structure qui est la même, pour toutes les méthodes du MENU, modulo l'existence des tableaux masses et norme.

Explicitons cette structure pour chaque type de données.  
Notons par

- O : Tableau des Observations de taille  $N \times k$ ,  $N \geq k$  ;
- P : Vecteur des Poids, affectés aux individus du tableau, de taille N ;
- M : Norme de l'espace des N individus,  $\mathbb{R}^k$  ;
- J : Vecteur de partitionnement des N individus en n groupes,  $J[i] \in \{1, \dots, n\}$  pour  $i = 1, \dots, N$  ;

(\*) Donnée : Tableau O.

Structure: O[1] O[2]....  
..... O[N].

(\*\*) Données : Tableaux O, J.

Structure : O[1] J[1] O[2] J[2]...  
..... O[N] J[N].

(\*\*) Données : Tableaux O, P, M

Structure : O[1] P[1] O[2] P[2]...  
..... O[N] P[N] M[1] M[2]...  
..... M[k].

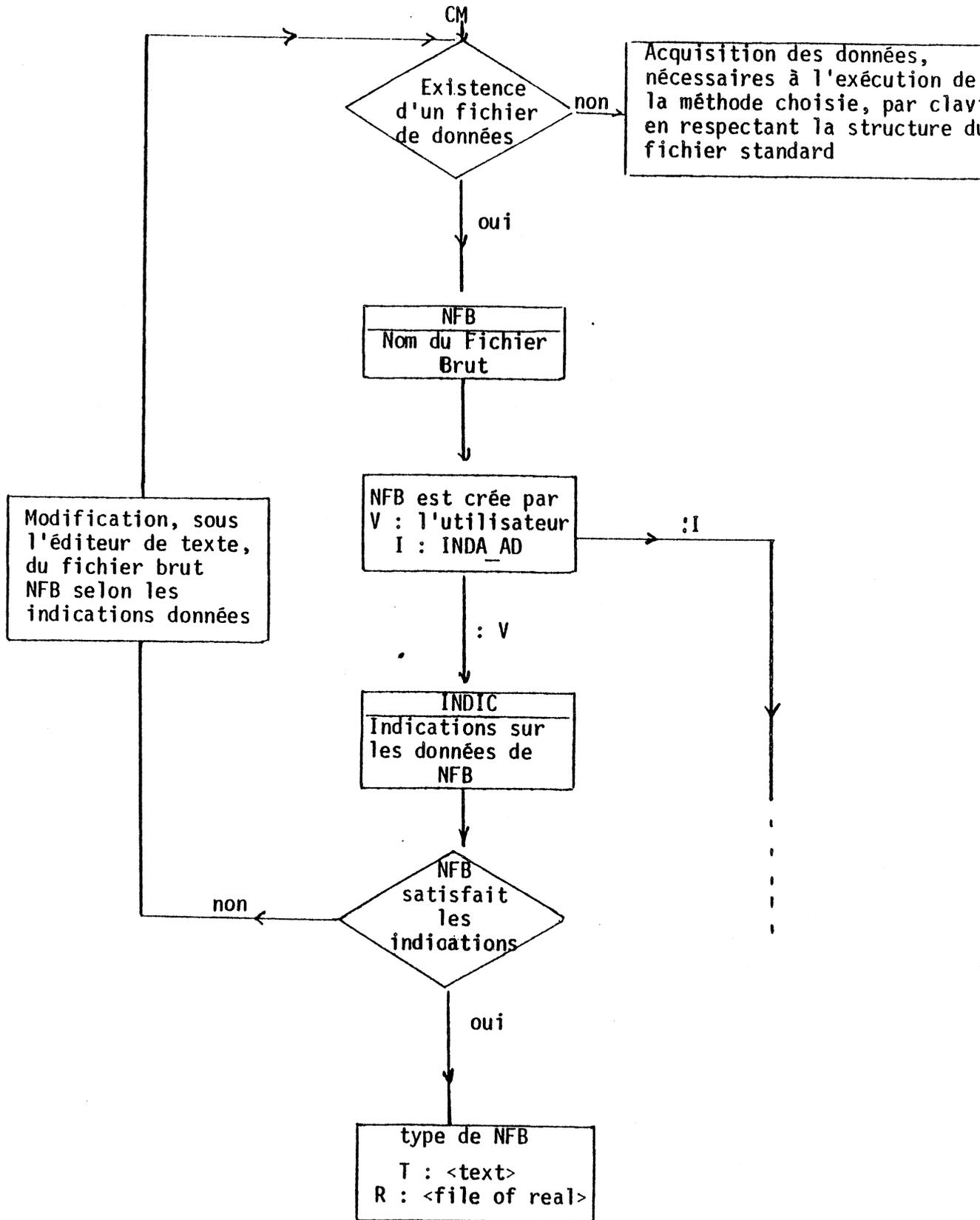


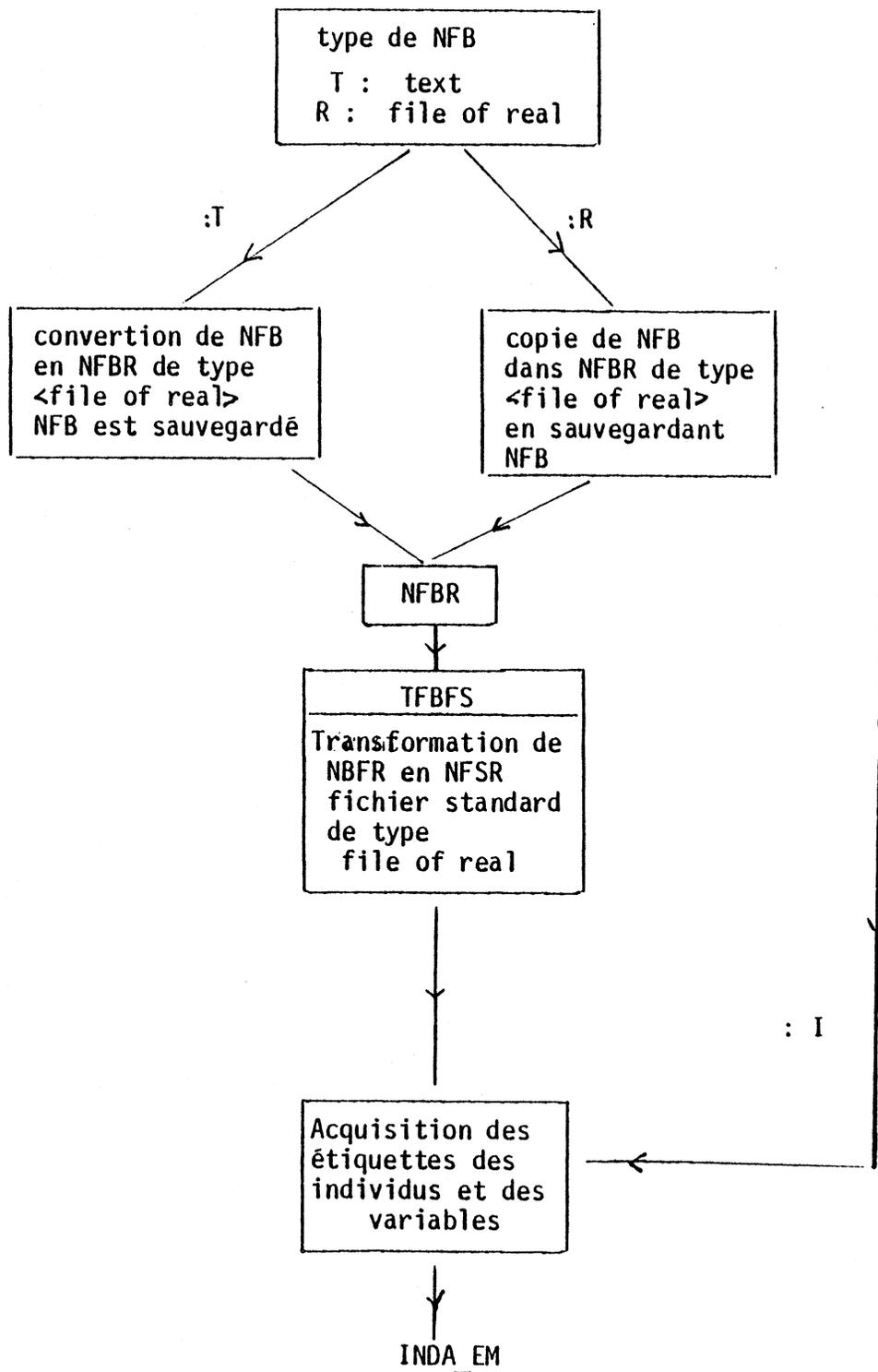
#### 2.4 - ACQUISITION DES ETIQUETTES

Dans tous les cas où l'échantillon n'est pas aléatoire, on peut vouloir donner des étiquettes aux individus. Dans le cas où l'utilisateur veut donner des étiquettes aux lignes ou/et aux colonnes de sa matrice des observations, une procédure permet de rentrer directement des étiquettes ayant au maximum quatre caractères.

L'utilisateur a aussi le choix de laisser le module INDA\_AD donner des étiquettes standards, c'est-à-dire les nombres 1,2,...,n. Le module range alors les étiquettes attribuées dans deux fichiers de type <file of string [4]>, l'un pour celles des lignes, l'autre pour celles des colonnes.

2.5 - FONCTIONNEMENT DE INDA\_AD







## CHAPITRE 3 :

### **MODULE D' EXECUTION DES METHODES D' ANALYSES *INDA\_EM***

3.1 - Bibliothèque INDAL -----	33
3.2 - Noyau commun aux techniques -----	36
3.3 - Calculs préliminaires -----	39
3.4 - Projections sur les plans factoriels	47
3.5 - Fonctionnement du module.-----	53



Pour chaque groupe de techniques, ce module opère sur un fichier standard de données, NOMFICH.DON, en effectuant certains calculs, préliminaires, propres à chaque groupe en utilisant la bibliothèque INDAL. Ces calculs produisent une matrice symétrique et définie positive pour laquelle on calcule les éléments propres. Les vecteurs propres de cette matrice correspondent aux axes factoriels sur lesquels on projette les points qu'on étudie.



### 3.1 - BIBLIOTHEQUE INDAL

La bibliothèque INDAL contient des procédures d'algorithmes d'algèbre linéaire et de sorties numériques sous forme de tableaux.

#### 3.1.1

Procédure FACTO (var A :RM ; N:integer ; var INDIC:integer) ;

Cette procédure permet la factorisation d'une matrice symétrique réelle et définie positive en un produit d'une matrice non singulière et triangulaire inférieure, notée  $A^{1/2}$ , par sa transposée.

Définition des paramètres :

entrée A : matrice réelle à factoriser  
N : ordre de A

sortie Indic : =  $\begin{cases} 1 & \text{si A est définie positive} \\ 0 & \text{sinon} \end{cases}$

A : sa partie inférieure contient les éléments de  $A^{1/2}$ ,  
sa partie supérieure stricte reste inchangée.

#### 3.1.2

Procédure INUSUP (var A:RM; N:integer ; Tol:real ; var Indic:integer) ;

Cette procédure permet d'inverser une matrice triangulaire supérieure qui est la partie supérieure d'une matrice carrée A, en stockant cette inverse dans la partie inférieure de A.

Définition des paramètres :

entrée : A : contenant dans sa partie supérieure la triangulaire  
à inverser.

N : ordre de A

Tol: constante réelle Tol =  $\inf\{\alpha : 1+\alpha \neq 1\}$ .

sortie : Indic : indicateur d'inversion

$$\text{Indic} = \begin{cases} 0 & \text{si } \exists i : A[i,i] < \text{Tol} \\ 1 & \text{sinon} \end{cases}$$

A : contient dans sa triangulaire inférieure la trans-  
posée de l'inverse.

La triangulaire supérieure stricte reste inchangée.

### 3.1.3

Procédure ELPRO (N,R:integer ; var A,VEP:RM ; var VAP:RV) ;

Cette procédure permet de calculer les valeurs et vecteurs propres dominants d'une matrice symétrique. La méthode utilisée est celle de Jacobi.

Définition des paramètres :

entrée : A : matrice carrée symétrique pour laquelle on cherche  
les éléments propres.

N : ordre de A

R : nombre d'éléments propres dominants recherchés.

sortie : VAP : VAP[i] pour  $i=1, \dots, R$  sont les R valeurs propres dominantes de A.

VEP : VEP[.,j] est le j<sup>ème</sup> vecteur propre associé à VAP[j],  $j=1, \dots, R$ .

### 3.1.4

Procédure TRTAB (var V:RV ; IN,SN,NCAR,NDEC:integer) ;

| Cette procédure donne une présentation lisible des résultats  
| numériques sous forme d'un tableau.

Définition des paramètres :

V : vecteur à lister

IN,SN : indices du premier et dernier élément de V à lister.

NCAR : nombre de caractères pour la représentation d'un nombre.

NDEC : nombre de décimales.

Procédure TRTAB (var V:RV ; IN,SN,NCAR,NDEC:integer) ;

var i,j : integer ;

begin

write(' \*") ;

for i:=IN to SN do

begin

for j:=1 to NCAR do write ('-' ) ;

write ('\*')

enf ;

writeln ;

write (' \*') ;

for i:=IN to SN do write (V[i]: NCAR : NDEC, '\*') ;

writeln

end ;

Remarque : Le paramètre (SN-IN) dépend évidemment de NCAR et du nombre de caractères d'une ligne de l'écran ou de l'imprimante.

### 3.2 - NOYAU COMMUN AUX TECHNIQUES

Soient la matrice  $X$  des observations, de taille  $k \times n$ , qu'on peut considérer comme  $k$  points de  $\mathbb{R}^n$  (resp.  $n$  points de  $\mathbb{R}^k$ ),  $A$  une matrice, symétrique définie positive, définissant un produit scalaire dans  $\mathbb{R}^n$  (voir section 3.4) (resp. les masses affectées aux points  $X[i,.]$  pour  $i = 1, \dots, k$ ), d'ordre  $n$ , et  $B$  une matrice d'ordre  $k$ , symétrique définie positive, représentant les masses affectées aux points  $X[.,j]$  pour  $j = 1, \dots, n$  (resp. un produit scalaire sur  $\mathbb{R}^k$ ).

$A$  et  $B$  étant symétriques définies strictement positives, on les factorise par la méthode de CHOLESKI en produits de matrices non singulières  $A = A^{1/2} A^{1/2t}$  et  $B = B^{1/2} B^{1/2t}$ , où  $A^{1/2}$  et  $B^{1/2}$  sont triangulaires inférieures.

Les méthodes d'analyse factorielle nécessitent toujours le calcul de valeurs et vecteurs propres des matrices, non nécessairement symétrique,  $U = XAX^t B$  d'ordre  $k$  et  $V = X^t BXA$  d'ordre  $n$ .

$U$  et  $V$  ont  $k$  valeurs propres en commun, qui sont celles de la matrice  $W = B^{1/2t} X A^{1/2} A^{1/2t} X^t B^{1/2}$  d'ordre  $k$  symétrique et définie positive [3]. Les valeurs propres sont donc réelles et positives.

On est donc amené à calculer les valeurs propres  $\lambda_i$  et vecteurs propres  $w_i$ , tels que  $w_i^t w_i = 1$ , de  $W = B^{1/2t} X A^{1/2} A^{1/2t} X^t B^{1/2}$ , et on déduit les vecteurs propres  $u_i$  de  $U$  et  $v_i$  de  $V$  par les formules :

$$u_i = B^{-1/2t} w_i \in \mathbb{R}^k \text{ avec } u_i^t B u_i = 1, u_i^t B u_j = 0 \quad i \neq j$$

dans le cas de  $n$  points de  $\mathbb{R}^k$ ,

et

$$v_i = \lambda_i^{-1/2} X^t B^{1/2} w_i \in \mathbb{R}^n \text{ avec } v_i^t A v_i = 1, v_i^t A v_j = 0 \quad i \neq j$$

dans le cas de  $k$  points de  $\mathbb{R}^n$ .

Remarque 1

Concrètement dans le logiciel, on ne calcule pas les  $u_i$  et  $v_i$  explicitement. Tout ce qui nous intéresse sur les  $u_i$  et  $v_i$  se déduit directement des  $w_i$  en utilisant les relations qui les lient et qu'on vient d'établir (voir section 4.3.1).

Les vecteurs propres  $u_i$  et  $v_i$  sont dits axes factoriels. Un plan factoriel est un plan engendré par deux axes factoriels.

Remarque 2 :

Calculer les valeurs et vecteurs propres de  $W$  au lieu de  $U$  et  $V$  est d'un intérêt numérique important. En effet, il suffit de calculer les valeurs et vecteurs propres d'une même matrice  $W$ , d'ordre  $k$ , symétrique, et définie positive, plutôt que d'exécuter deux fois l'algorithme de calcul des valeurs et vecteurs propres pour  $U$  d'ordre  $k$  et pour  $V$  d'ordre  $n$  non nécessairement symétriques et  $n$  pouvant être très grand par rapport à  $k$ .

Dans INDA, le calcul des éléments propres de  $W$  d'ordre  $k$ , utilise la procédure ELPRO de INDAL, et se fait par la procédure VAVEP. En outre, dans cette procédure, on calcule les pourcentages d'inertie de chaque valeur propre et les pourcentages cumulés.

Procédure VAVEP (k:integer; var W, VEP:RM; var VAP:RV) ;

début

ELPRO(k,k,W,VEP,VAP) ; (\* calcul des valeurs propres VAP et  
vecteurs propres VEP de W \*)

pour i = 1 à k faire

début

PI[i] := VAP[i] /  $\sum_{j=1}^k$  VAP[j]; (\* pourcentage d'inertie  
de VAP[i] \*)

PC[i] := PC[i-1] + PI[i]; (\* pourcentage cumulé de  
avec VAP[i] \*)

PC[1] := PI[1]

fin ;

fin de VAVEP ;

Les pourcentages d'inertie des valeurs propres traduisent la part explicative de chaque axe factoriel dans la dispersion du nuage de points. Cela permet de choisir le plan factoriel d'inertie maximale sur lequel on projette le nuage de points à analyser.

### 3.3 - CALCULS PRELIMINAIRES

Le modèle d'exécution des méthodes, INDA\_EM, utilise la bibliothèque INDAL contenant les procédures communes aux techniques du MENU et contient les procédures CPAFC, CPACP, CPAFD, CPCV, de calculs d'algèbre linéaire, propres à chaque groupe de techniques. Ces procédures consistent en un calcul préliminaire effectué sur la matrice des observations à analyser et éventuellement sur la matrice des masses et de la métrique pour aboutir à une matrice symétrique et définie positive  $W$  à partir de laquelle sont calculés les plans factoriels (voir section 3.2).

Dans ce qui suit on donne le détail des procédures de calculs préliminaires pour chaque groupe de techniques.

#### 3.3.1 - Groupe 1

Procédure CPACP ;

Données :

$S = [S_1, \dots, S_j, \dots, S_n]$ ,  $S_j \in \mathbb{R}^k$ ,  $n$  points dans  $\mathbb{R}^k$ ,

$A = \text{diag}(a_j)$ ,  $j = 1, \dots, n$ , matrice des masses,

$B$  une matrice d'ordre  $k$  symétrique et définie positive définissant une norme sur  $\mathbb{R}^k$ .

Calculs :

- Calcul du barycentre des  $n$  points  $S_j$  :

$$\bar{S} := \left( \sum_{j=1}^n a_j S_j \right) / p, \text{ avec } p = \sum_{j=1}^n a_j ;$$

- Centrage de  $S$  :

$$S_j := S_j - \bar{S} \quad j = 1, \dots, n ;$$

- Normalisation des masses :

$$a_j := a_j/P ;$$

- On pose  $X = S$ ,  $S$  centrée ;

- Calcul de la covariance empirique de  $S$  :

$$\Lambda_S = X A X^t ;$$

- Factorisation de CHOLESKI de  $B$  :

$B = B^{1/2} B^{1/2t}$ , où  $B^{1/2}$  triangulaire inférieure est stockée dans la partie inférieure de  $B$  ;

- Calcul de  $W$  :

$$W = B^{1/2t} \Lambda_S B^{1/2}.$$

Remarque :

Pour une ACP unitaire,  $B = I$ , matrice identité.

Pour une ACP normée,  $B = \text{diag}(b_i^{-1})$   $i = 1, \dots, k$ , avec

$$b_i = \left( \sum_{j=1}^n a_j (S_{ij})^2 \right) / P - (\bar{S}_i)^2.$$

3.3.2 - Groupe 2

Procédure CPAFC ;

Donnée : matrice des observations  $X$  de taille  $k \times n$ ,  $k \leq n$  $X_{ij}$  entiers positifs ou nuls pour  $i=1\dots k$ ,  $j=1\dots n$  $X$  représente  $n$  points de  $\mathbb{R}^k$  (resp.  $k$  points de  $\mathbb{R}^n$ ).

Calculs :

$$- S = \sum_{i,j} X_{ij} ;$$

$$- X_{ij} := X_{ij}/S ; \text{ d'où } \sum_{i,j} X_{ij} = 1 ;$$

- En sommant  $X$  par lignes, on obtient le vecteur

$$b^t = (b_1, \dots, b_i, \dots, b_k) \text{ avec } b_i = \sum_{j=1}^n X_{ij} > 0 ;$$

- En sommant  $X$  par colonnes, on obtient le vecteur

$$a^t = (a_1, \dots, a_j, \dots, a_n) \text{ avec } a_j = \sum_{i=1}^k X_{ij} > 0 ;$$

- Par commodité on pose  $X := X - b a^t$  (voir la remarque),

$$\text{c'est-à-dire } X_{ij} := X_{ij} - b_i a_j ;$$

- On pose :  $A^{-1} = \text{diag}(a_j)$  et  $B^{-1} = \text{diag}(b_i)$  ;

On obtient donc le triplet :

 $X$  : matrice des observations $A = \text{diag}(a_j^{-1})$  : masses (resp. norme de  $\mathbb{R}^n$ ) $B = \text{diag}(b_i^{-1})$  : norme de  $\mathbb{R}^k$  (resp. masses)- Calcul de  $W$  :

$$W = B^{1/2t} X A X^t B^{1/2} ; \text{ avec } B^{1/2t} = B^{1/2} = \text{diag}(b_i^{-1/2}).$$

Remarque

La déflation  $X := X - b a^t$  est faite dans le but d'éliminer la valeur propre triviale  $\lambda_1 = 1$  de  $W$ , de vecteur propre associé  $B^{1/2} b$ .

### 3.3.3 - Groupe 3

PROCEDURE CPAFD ;

DONNEES :

$$S = [S_1, \dots, S_\ell, \dots, S_N], \quad S_\ell \in \mathbb{R}^k, \quad \ell = 1 \dots N ;$$

$$J = [J_1, \dots, J_\ell, \dots, J_N], \quad J_\ell \in \{1, \dots, n\}, \quad \ell = 1 \dots N.$$

CALCULS :

- $F[j]$  = Fréquence de la valeur  $j$   
 = nombre de fois où  $J_\ell = j$   
 $j = 1, \dots, n$  et  $\ell = 1, \dots, N$

- Moyenne empirique de  $S$  :

$$\bar{S} = \left( \sum_{\ell=1}^N S_\ell \right) / N ;$$

- Centrage de  $S$  :

$$S_\ell := S_\ell - \bar{S}, \quad \ell = 1, \dots, N ;$$

- Matrice de variance empirique

$$\Lambda = \left( \sum_{\ell=1}^N S_\ell S_\ell^t \right) / N ;$$

- Calcul de la moyenne conditionnelle empirique centrée, multiplié par les probabilités marginales empiriques de  $J$  :

$$X_j = \left( \sum_{\ell: J_\ell=j} S_\ell \right) / (N \times F[j]), \quad j = 1 \dots n ;$$

- On pose :  $X = [X_1, \dots, X_j, \dots, X_n]$  matrice des centres de gravité de chaque groupe

$$A = \text{diag}(F[j]/N) \quad \text{matrice diagonale des masses affectées aux } X_j$$

$$B^{-1} = \Lambda ;$$

- On factorise  $B^{-1}$  par CHOLESKI :

$$B^{-1} = B^{-1/2} B^{-1/2t}$$

$B^{-1/2}$  est stockée dans la partie supérieure de  $B^{-1} = \Lambda$ , on inverse  $B^{-1/2}$  et on stocke  $B^{1/2t}$  dans la partie inférieure de  $B^{-1}$ .

On obtient  $B = B^{1/2t} B^{1/2}$

- Calcul de W :  $W = B^{1/2} X A X^t B^{1/2t}$ .

3.2.4 - Groupe 4

PROCEDURE CPCV ;

DONNEES :  $S = [S_1, \dots, S_\ell, \dots, S_N], S_\ell \in \mathbb{R}^k$  $T = [T_1, \dots, T_\ell, \dots, T_N], T_\ell \in \mathbb{R}^n ;$ 

CALCULS :

- Moyennes empiriques :

$$\bar{S} = \left( \sum_{\ell=1}^N S_\ell \right) / N, \quad \bar{T} = \left( \sum_{\ell=1}^N T_\ell \right) / N ;$$

- Centrage de S et T :

$$S_\ell := S_\ell - \bar{S}, \quad T_\ell := T_\ell - \bar{T}, \quad \ell = 1, \dots, N ;$$

- Matrices des variances empiriques :

$$\Lambda_S = \left( \sum_{\ell=1}^N S_\ell S_\ell^t \right) / N, \text{ d'ordre } k$$

$$\Lambda_T = \left( \sum_{\ell=1}^N T_\ell T_\ell^t \right) / N, \text{ d'ordre } n ;$$

- Calcul de X :

$$X = \left( \sum_{\ell=1}^N S_\ell T_\ell^t \right) / N, \text{ de taille } k \times n ;$$

- On pose :  $A^{-1} = \Lambda_T$  et  $B^{-1} = \Lambda_S$ - On factorise  $A^{-1}$  et  $B^{-1}$  par CHOLESKI :

$$A^{-1} = A^{-1/2} A^{-1/2t}, \quad A^{-1/2} \text{ est stockée dans la partie supérieure de } A^{-1}.$$

$$B^{-1} = B^{-1/2} B^{-1/2t}, \quad B^{-1/2} \text{ est stockée dans la partie supérieure de } B^{-1} ;$$

- On inverse  $A^{-1/2}$  et  $B^{-1/2}$  par la procédure INUSUP de INDAL  
et on obtient :

$A^{1/2t}$  stockée dans la partie inférieure de  $A^{-1}$   
 $B^{1/2t}$  stockée dans la partie inférieure de  $B^{-1}$

donc A et B sont factorisées par :

$$A = A^{1/2t} A^{1/2} \quad \text{et} \quad B = B^{1/2t} B^{1/2}$$

- Calcul de W :

$$W = B^{1/2} \times A^{1/2t} A^{1/2} \times B^{1/2t}$$

Remarque

Pour les quatre groupes, les calculs préliminaires sur les données reviennent à des multiplications matricielles, à des factorisations de CHOLESKI et des inversions de matrices triangulaires inférieures.

La répartition des techniques en groupes a été faite sur le critère de calculs communs, comme on vient de le voir, et sur le critère de complexité des calculs [9]. Cette complexité s'explique ainsi :

- Groupe 1 : Produits matriciels (P.M.)
- Groupe 2 : \* P.M.  
\* Choleski d'ordre k
- Groupe 3 : \* P.M.  
\* Choleski d'ordre k  
\* 1 inversion d'une triangulaire d'ordre k
- Groupe 4 : \* P.M.  
\* 1 Choleski d'ordre k  
1 Choleski d'ordre n (en général  $n \gg k$ )  
\* 1 inversion d'une triangulaire d'ordre k  
1 inversion d'une triangulaire d'ordre n.

### 3.4. - PROJECTIONS SUR LES PLANS FACTORIELS

On a vu que les quatre procédures de calcul préliminaires aboutissent toutes à une matrice  $W$  symétrique d'ordre  $k$ . A cette matrice, on applique la procédure VAVEP (voir section 3.2.), pour obtenir en sortie :

- les  $s \leq k$  plus grandes valeurs propres de  $W$ ,  $\lambda_1 \geq \dots \geq \lambda_s \geq 0$ .
- les parts explicatives des  $\lambda_i$ , soit  $\lambda_i / \sum_{j=1}^s \lambda_j$ ,  $i = 1 \dots s$ .
- les  $s$  vecteurs propres dominants  $w_i$ , associés aux  $s$  plus grandes valeurs propres qui correspondent aux axes factoriels qui engendrent les plans factoriels sur lesquels sera projeté le nuage de points. Ces vecteurs propres sont normés par  $w_i^t w_i = 1$ ,  $i = 1, \dots, s$ .

Dans la section qui suit, on donne les procédures du logiciel qui calculent les coordonnées des projections des points sur les axes factoriels, et on spécifie pour chaque technique la procédure utilisée.

On note  $\langle \dots \rangle_B$  le produit scalaire dans  $\mathbb{R}^k$  défini, par une matrice  $B$ , d'ordre  $k$  symétrique et définie positive, par  $\forall x, y \in \mathbb{R}^k$   
 $\langle x, y \rangle_B = x^t B y$ .

Ayant un point  $x$  de  $\mathbb{R}^k$ , sa projection sur un vecteur  $u$  de  $\mathbb{R}^k$  est donnée par  $\langle x, u \rangle_B u$ .

### 3.4.1 - Coordonnées des projections

Soit  $P$  une matrice de taille  $\ell \times c$ .

(\*)  $P$  peut représenter  $\ell$  points de  $\mathbb{R}^c$  muni d'un produit scalaire, défini par une matrice  $N$  d'ordre  $c$  symétrique et définie positive,  $\langle \cdot, \cdot \rangle_N$ . En affectant ces points de masses, représentées par une matrice  $M$  de taille  $\ell \times \ell$  symétrique et définie positive, on obtient  $P^t M$ .

On veut projeter ces  $\ell$  points de  $\mathbb{R}^c$  sur les  $s$  vecteurs propres dominants  $v_i$ , de la matrice  $V = P^t M P N$  de taille  $c \times c$ , qui sont déduits des vecteurs propres de  $W = M^{1/2 t} P N P^t M^{1/2}$  par la formule  $v_i = \lambda_i^{-1/2} P^t M^{1/2} w_i$  pour  $i = 1 \dots s$  (voir section 3.2) ; pour cela, on calcule le produit scalaire :

$$\begin{aligned} \langle P^t M, v_i \rangle_N &= (P^t M)^t N v_i = M P N v_i = \\ &= \lambda_i^{-1/2} M P N P^t M^{1/2} w_i = \lambda_i^{-1/2} M^{1/2} M^{1/2 t} P N P^t M^{1/2} w_i = \\ &= \lambda_i^{-1/2} M^{1/2} W w_i = \lambda_i^{-1/2} \lambda_i M^{1/2} w_i = \lambda_i^{1/2} M^{1/2} w_i = F_{li}, \end{aligned}$$

pour  $i = 1 \dots s$ .

Donc, pour  $j = 1 \dots \ell$ , les  $s$  coordonnées des projections de  $P[j, \cdot]$  sont  $F_{li}[j]$  pour  $i = 1 \dots s$ .

(\*\*)  $P$  peut représenter  $c$  points de  $\mathbb{R}^\ell$  muni d'un produit scalaire, défini par une matrice  $N$  d'ordre  $\ell$  symétrique et définie positive,  $\langle \cdot, \cdot \rangle_N$ . En affectant ces points de masses, représentées par une matrice  $M$  d'ordre  $c$  symétrique et définie positive, on obtient  $P M$ .

On veut projeter ces  $c$  points,  $P[\cdot, j]$   $j = 1, \dots, c$ , de  $\mathbb{R}^\ell$  sur les  $s$  vecteurs propres dominants  $u_i$  de la matrice  $U = P M P^t N$  d'ordre  $\ell$ , qui sont déduit des vecteurs propres de  $W = N^{1/2 t} P M P^t N^{1/2}$  par la formule  $u_i = N^{-1/2 t} w_i$  pour  $i = 1 \dots s$  (voir section 3.2) ;

pour cela, on calcule le produit scalaire :

$$\begin{aligned} \langle PM, u_i \rangle_N &= (PM)^t N u_i = MP^t N u_i = MP^t N^{1/2} N^{1/2} u_i \\ &= MP^t N^{1/2} N^{1/2} t_N^{-1/2} w_i = MP^t N^{1/2} w_i = \lambda_i^{-1/2} MP^t F1_i = F2_i \end{aligned}$$

pour  $i = 1 \dots s$ . Donc, pour  $j = 1, \dots, c$ , les  $s$  coordonnées des projections de  $P[.,j]$  sont  $F2_i[j]$  pour  $i = 1 \dots s$ .

### 3.4.2 - Procédures de projections

FOLIN1 ( $\lambda$ ,  $c$ ,  $P$ ,  $N$ ,  $M$ ) ;

```

début
  pour i = 1...s ,
    calcul de F1i =  $\lambda_i^{1/2} M^{1/2} w_i$  ;
fin

```

FOLIN2 ( $\lambda$ ,  $c$ ,  $P$ ,  $N$ ,  $M$ ) ;

```

début
  FOLIN1( $\lambda$ ,  $c$ ,  $P$ ,  $M$ ,  $N$ ) ;
  pour i = 1...s, calcul de F2i =  $\lambda_i^{-1/2} MP^t F1_i$  ;
fin

```

### 3.4.3 - Spécification des procédures de projections pour chaque technique du MENU.

A partir des procédures de calculs préliminaires, on va spécifier pour chaque technique du MENU, les procédures de projections utilisées et leurs paramètres.

#### (\*) Groupe 1

La procédure de calcul préliminaire CPACP nous donne une matrice  $X$  de taille  $k \times n$ , représentant  $n$  points  $X[.,j]$  de  $\mathbb{R}^k$  centrés et réduits, à partir de la donnée de  $n$  points  $S_j$  de  $\mathbb{R}^k$  affectés de masses représentées par une matrice  $A$  d'ordre  $n$ . Une norme sur  $\mathbb{R}^k$  est donnée par une matrice  $B$  d'ordre  $k$ .

- ACP d'un nuage de points : (CM = 1,2,3)

On projette les  $n$  points  $X[.,j]$ , sur les  $s$  vecteurs propres  $u_i$ ,  $i = 1, \dots, s$  affectés de masses unitaires représentées par la matrice identité en appliquant FOLIN2( $k, n, X, B, I$ ). On aura  $s$  vecteurs  $F2i = X^t B^{1/2} w_i$  pour  $i = 1 \dots s$  tels que le nuage des projections de coordonnées  $(F2i_1[j], F2i_2[j])$  pour  $j=1, \dots, n$  est tracé par le module graphique

- ACP d'un nuage de points relativement à une norme (CM = 4,5,6)

On projette les  $k$  points  $X[j,.]$  de  $\mathbb{R}^n$  sur les  $s$  vecteurs propres  $v_i$ ,  $i = 1, \dots, s$  en appliquant la procédure FOLIN1( $k, n, X, A, B$ ).  $A$  représente la norme de  $\mathbb{R}^n$  et  $B$  les poids affectés aux points. On obtient  $s$  vecteurs  $F1i = \lambda_i^{1/2} B^{1/2} w_i$  pour  $i = 1 \dots s$  tels que  $F1i[j]$  est la projection du point  $X[j,.]$  sur le  $i$ ème axe factoriel.

**(\*\*)** Groupe 2Paramètres :

- X , de taille  $k \times n$ , matrice des observations  
 A , d'ordre n, matrice symétrique et définie positive  
 B , d'ordre k, symétrique et définie positive.

- AFC sur les lignes : (CM = 7)

On projette les k points,  $X[j, \cdot]$   $j = 1 \dots k$ , de  $\mathbb{R}^n$  sur les s vecteurs propres  $v_i$ ,  $i = 1 \dots s$ , en appliquant la procédure FOLIN1(k,n,X,A,B), où A est le paramètre norme et B masses des points. On aura s vecteurs  $F1i = \lambda_i^{1/2} B^{1/2} w_i$  pour  $i = 1 \dots s$ , tels que le nuage des projections de coordonnées  $(F1i_1[j], F1i_2[j])$  pour  $j = 1, \dots, k$  est tracé par le module graphique.

- AFC sur les colonnes : (CM = 8)

On projette les n points,  $X[\cdot, j]$   $j = 1 \dots n$ , de  $\mathbb{R}^k$  sur les s vecteurs propres  $u_i$ ,  $i = 1 \dots s$ , en appliquant FOLIN2(k,n,X,B,A), où B est la norme et A les masses des points. Le tracé du nuage des projections de coordonnées  $(F2i_1[j], F2i_2[j])$   $j=1 \dots, n$  où  $F2i = AX^t B^{1/2} w_i$  pour  $i = 1 \dots s$ , est fait par le module graphique.

- Corrélation canonique entre deux variables aléatoires discrètes (CM = 9)

Pour cette technique, on représente sur le plan engendré par les vecteurs  $u_1$  et  $v_1$  le nuage de points de coordonnées  $(F11[i], F21[j])$  pour  $i = 1, \dots, k$  et  $j = 1, \dots, n$ .

F11 étant calculé par FOLIN1(k,n,X,A,B) et F21 étant calculé par FOLIN2(k,n,X,B,A). Le tracé du nuage de points est fait par le module graphique

**(\*\*\*)** Groupe 3

La procédure de calcul préliminaire CPAFD nous donne à partir d'une variable  $J \in \mathbb{R}^n$  où  $J[i] \in \{1, \dots, n\}$  et d'une matrice S de taille  $k \times N$ , d'observations, une matrice X de taille  $k \times n$ , une norme B d'ordre k et une matrice A d'ordre n, de masses affectées aux n points  $X[\cdot, j]$ . Les N points  $S[\cdot, j]$  étant affectés de masses unitaires.

AFD : (CM = 10)

On projette les  $n$  points  $X[.,j]$  sur les  $s$  vecteurs propres  $u_i$ ,  $i = 1 \dots s$  en appliquant  $FOLIN2(k,n,X,B,A)$ , ce qui nous ramène au même cas de la technique (2).

Les  $n$  points  $X[.,j]$  sont les centres de gravités des  $N$  points  $S[.,j]$  de  $\mathbb{R}^k$ .

On projette aussi les  $S[.,j]$ , muni des masses identités, sur les  $s$  vecteurs  $u_i$  en appliquant  $FOLIN2(k,N,S,B,I)$ .

Corrélation canonique entre un vecteur aléatoire et une variable aléatoire finie : (CM = 11)

Pour cette technique, on calcule à partir de  $S$  et  $X$ , et en appliquant  $FOLIN2(k,N,S,B,I)$  et  $FOLIN2(k,n,X,B,A)$  deux vecteurs  $F21 = S^t B^{1/2} w_1$   $F22 = AX^t B^{1/2} w_1$ . Avec ces deux vecteurs, on représente sur le plan engendré par  $u_1$  et  $v_1$  le nuage de points de coordonnées  $(F22[j], F21[i])$  pour  $j = 1 \dots n$  et  $i$  tel que  $J[i] = j$  en appliquant le module graphique.

(\*\*\*\*) Groupe\_4

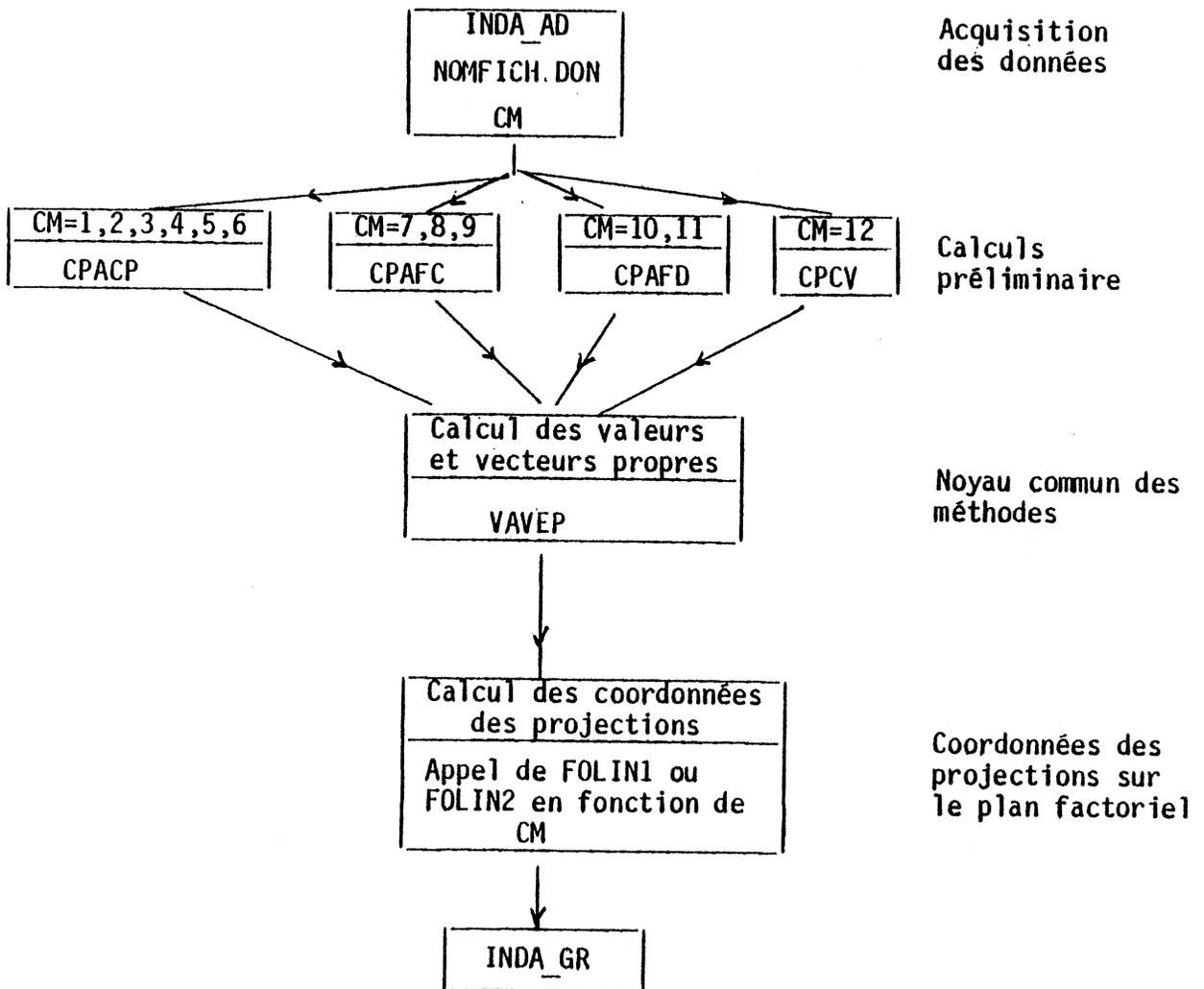
A partir de la donnée de  $N$  couples de points  $(S_j, T_j)$  pour  $j = 1, \dots, N$   $S_j \in \mathbb{R}^k$  et  $T_j \in \mathbb{R}^n$ , la procédure CPCV, de calcul préliminaire nous donne une matrice  $X$  de taille  $k \times n$ , une norme  $B$  sur  $\mathbb{R}^k$  et une matrice  $A$  de masses, d'ordre  $n$ .

Corrélation entre deux vecteurs : (CM = 12)

On cherche les coordonnées des projections, sur les  $s$  vecteurs propres  $u_i$ , des  $n$  points  $X_j$  de  $\mathbb{R}^k$  normé par  $B$ , munis des masses  $A$  en appliquant la procédure  $FOLIN2(k,n,X,B,A)$ . Le nuage des projections sera donné par les coordonnées  $(F2i[j])$  pour  $j = 1, \dots, n$  où  $F2i = AX^t B^{1/2} w_i$   $i = 1, \dots, s$ .

### 3.5 - FONCTIONNEMENT DE INDA\_EM

Le module INDA\_EM, d'exécution des méthodes fonctionne par l'enchaînement des procédures qu'on vient d'établir suivant le schéma suivant :



Remarque :

Pour le module INDA\_EM, on donne la possibilité à l'utilisateur de considérer certains de ses points de départ (données à analyser) comme supplémentaires et le reste comme points principaux.

Les procédures de calculs préliminaires et de calcul des éléments propres sont alors activés seulement sur le nuage de points principaux. Les points supplémentaires seront aussi projetés sur les axes factoriels sans être intervenus dans le calcul de ceux-ci.

## CHAPITRE 4 :

### MODULE GRAPHIQUE *INDA\_GR*

4.1 - Procédure TRN -----	58
4.2 - ZOOM graphique -----	62
4.3 - ZOOM calculatoire -----	63
4.4 - Fonctionnement du module.---	64



Après l'exécution du module INDA\_EM, sur le fichier standard NOMFICH·DON, on récupère un fichier NOMFICH·RES de type <file of real> contenant les coordonnées des projections des données sur les axes factoriels.

Procédure LPJ ;

Après le choix du plan de visualisation du nuage par l'utilisateur et ce en donnant les numéros des deux axes qui l'engendrent, la procédure LPJ lit les projections des données, sur les deux axes choisis, dans NOMFICH-RES et les stockes dans deux fichiers. La première projection est stockée dans le fichier réel NOMFICH-PJ1 et la seconde dans NOMFICH-PJ2.

Une procédure de tracé de nuage opère alors sur ces deux fichiers.

#### 4.1 - Procédure TRN (tracé de nuage de points).

Cette procédure visualise à l'écran le nuage de points projeté dont les abscisses sont dans NOMFICH.PJ1 et les ordonnées dans NOMFICH.PJ2. Elle fonctionne par l'enchaînement des sous-procédures qui suivent.

##### - Procédure EXT ;

Pour chacune des projections PJ1 et PJ2, on calcule leurs extrema, MINPJ1, MAXPJ1, MINPJ2, MAXPJ2, en utilisant la technique par blocs ; le nombre de points de nuage peuvent être grand, (voir section 1.7).

##### - Procédure CADREP ;

Calcule l'origine du repère et l'échelle graphiques à partir des coordonnées graphiques de la fenêtre dans laquelle sera visualisé le nuage. On désigne par (X1,Y1) les coordonnées du point bas gauche et par (X2,Y2) les coordonnées du point haut droit de la fenêtre de visualisation.

On a alors :

.Echelle de l'axe des abscisses :

$$ECHX = (X2 - X1) / (MAXPJ1 - MINPJ1) ;$$

.Echelle de l'axe des ordonnées :

$$ECHY = (Y2 - Y1) / (MAXPJ2 - MINPJ2) ;$$

Origine (X0,Y0) du repère graphique.

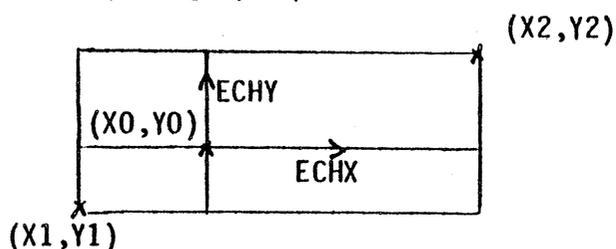
Si MINPJ1 > 0 alors X0=X1

sinon X0=X1+round(-MINPJ1\*ECHX) ;

Si MINPJ2 > 0 alors Y0=Y1

sinon Y0=Y1+round(-MINPJ2\*ECHY)

Le calcul de ces trois paramètres permet à ce niveau de tracer les axes du repère graphique



Un point de coordonnées réelles P1 et P2, a pour coordonnées graphiques GP1 et GP2 calculées par :

$$GP1 = X1 + \text{round}((P1 - \text{MINPJ1}) * \text{ECHX}) ;$$

$$GP2 = Y1 + \text{round}((P2 - \text{MINPJ2}) * \text{ECHY}).$$

Ce point graphique est visualisé à l'écran en couleur 'COL' par la procédure VISPT(GP1, GP2, COL).

Deux procédures essentielles de visualisation d'un nuage de points dans un plan, sont utilisées pour toutes les méthodes factorielles étudiées.

- Procédure TRN1 ;

On veut visualiser le nuage de point  $(u(i), v(i))$   $i = n_1, \dots, n_2$   $1 \leq n_1, n_2 \leq n$ , où  $u$  et  $v$  sont deux vecteurs de  $\mathbb{R}^n$ . Les vecteurs  $u$  et  $v$  sont stockés respectivement dans les fichiers NOMFICH.PJ1 et NOMFICH.PJ2. Le procédé de visualisation utilise le principe par blocs.

procédure TRN1(n1,n2,U,V,COL) ;

var i, GU, GV: integer ;

début

pour i=n1 à n2 faire

début

GV:=X1+round( (u(i)-UMIN)xECHX) ;

GV:=Y1+round( (v(i)-vmin)xECHY) ;

VISPT(GU,GV,COL)

fin

fin

## - Procédure TRN2

On veut visualiser le nuage de points  $(u(i), v(j))$  pour  $i = n_1 \dots n_2$   
 $j = k_1 \dots k_2$ , où  $u \in \mathbb{R}^n$ ,  $v \in \mathbb{R}^k$  et  $1 \leq n_1, n_2 \leq n$ ,  $1 \leq k_1, k_2 \leq k$ .  
 Les vecteurs  $u$  et  $v$  sont stockés respectivement dans les fichiers  
 NOMFICH.PJ1 et NOMFICH.PJ2.

Le principe bloc par bloc est utilisé pour la visualisation.

```

procédure TRN2 (n1, n2, k1, k2, u, v, col) ;
var i, j, Gu, Gv: integer
début
  pour i:=n1 à n2 faire
  début
    calcul de Gu ;
    pour j:=k1 à k2 faire
    début
      calcul de Gv ;
      VISPT(Gu, Gv, COL)
    fin de boucle j
  fin (boucle i)
fin (procédure)
  
```

Remarque

Des généralisations de TRN1 et de TRN2 à partir de celles-ci sont  
 utilisées.

$$\text{TRN1} : u = (u_{11}, \dots, u_{1k} ; u_{21}, \dots, u_{2n})$$

$$v = (v_{11}, \dots, v_{1k} ; v_{21}, \dots, v_{2n})$$

On visualise le nuage  $(u(i), v(i))$  en appliquant 2 fois TRN1 :

TRN1(1, k, u, v, COL1) ;

TRN1(k+1, n-k, u, v, COL2) ;

Cette généralisation est utilisée en AFD pour la visualisation des projections des individus et de leurs centres de gravités.

TRN2 :  $u = (u_1, \dots, u_k)$

$v = (v_{11}, \dots, v_{1s} ; v_{21}, \dots, v_{2t})$

Le nuage  $(u(i), v(j))$  est alors visualisé par :

TRN2(1, k, 1, S, u, v, COL1) ;

TRN2(1, k, S+1, t-S, u, v, COL2) ;

La corrélation canonique entre un vecteur aléatoire et une variable aléatoire finie utilise cette généralisation.

Après le tracé du nuage, l'utilisateur choisit une des recettes du MENU graphique qui offre :

- 1 : Copie graphique sur imprimante
- 2 : Projections sur d'autres axes
- 3 : ZOOM graphique
- 4 : ZOOM calculatoire.

Le premier choix permet d'avoir une copie verticale de l'écran graphique sur imprimante. Le second, permet de visualiser le nuage de points en choisissant deux autres axes factoriels.

Certains des points du nuage pouvant être superposés sur l'écran graphique sans avoir des coordonnées réelles égales, le graphe pouvant être localement dense, nous amène à considérer les fonctions ZOOM.

#### 4.2. - ZOOM GRAPHIQUE

Cette fonction sert à agrandir une partie du graphe. On définit une fenêtre graphique qui contient la partie du graphe à agrandir. Les coordonnées graphiques de cette fenêtre, c'est-à-dire le point bas gauche et le point haut droit, permettent de sélectionner les points du nuage dont les coordonnées graphiques sont comprises entre celles-ci.

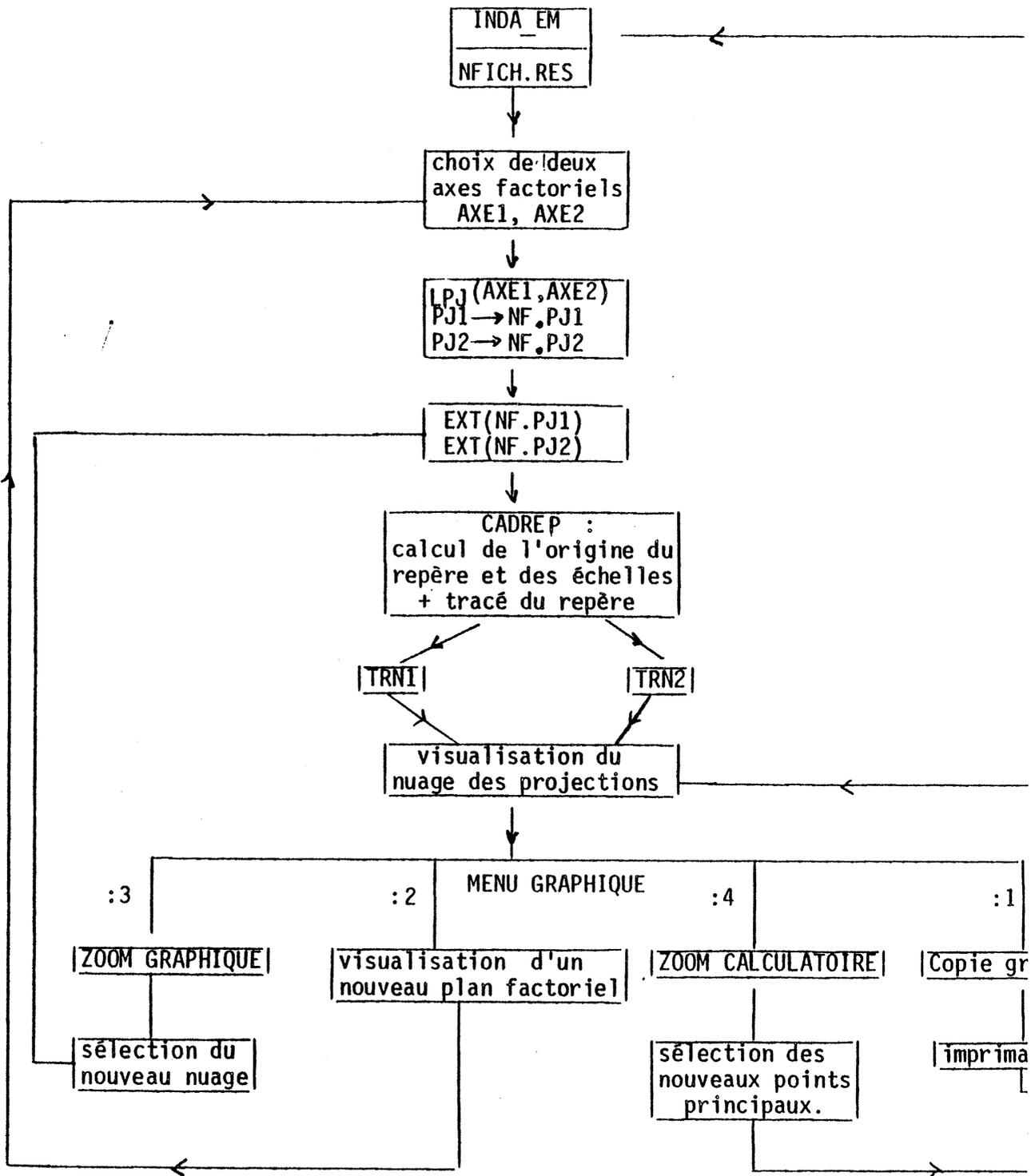
Le nouveau nuage étant défini, on lui applique la procédure TRN pour avoir un nouveau graphe constitué des seuls points sélectionnés par la fenêtre graphique.

#### 4.3 - ZOOM CALCULATOIRE

Après visualisation du nuage des projections des points de départ, l'utilisateur voudrait redéfinir les points principaux et les points supplémentaires de ses données initiales pour les retraiter sous cette nouvelle définition. Cette option est possible en utilisant la fonction ZOOM calculatoire.

On sélectionne une partie du nuage visualisé, de la même façon que dans la fonction ZOOM graphique, par une fenêtre. A partir du nuage, des projections, sélectionné, on remonte aux données de départ pour considérer les points initiaux correspondant comme points principaux et le reste comme points supplémentaires. Le module INDA\_EM est alors activé sur ces nouveaux points principaux afin de calculer les nouvelles composantes principales ; les points supplémentaires tout comme les points principaux sont projetés sur les nouveaux axes factoriels mais sans être intervenus au calcul de ceux-ci. La procédure TRN réappliquée sur ces nouvelles projections et visualise le nouveau nuage.

4.4 - FONCTIONNEMENT DU MODULE



Remarque

- Signalons que toutes les procédures graphiques, nécessitant l'appel de vecteurs de grandes tailles, fonctionnent suivant le principe par bloc (voir section 1.7). Cette politique est dans le but de pouvoir visualiser un grand nombre de points en gardant leurs coordonnées stockées dans un fichier.
- Remarquons que les fonctions graphiques de base sous DOS étant écrites en langage BASIC, on a utilisé une interface graphique écrite en COMPAS par A. EBERHARD.



**CONCLUSION**



1. Le progiciel INDA est actuellement opérationnel sur tout micro ordinateur compatible avec IBM-PC possédant un minimum de :
  - 256 k en mémoire centrale
  - deux unités périphériques (disquettes)
  - un écran monochrome
  - un coprocesseur 8086
  - système d'exploitation MS-DOS
  - langage PASCAL.
  
2. INDA est écrit sous COMPAS. Il a une capacité de traitement des données de taille  $n \times k$ , où  $k$  est le nombre maximum de variables et est égal à 100, et  $n$  est le nombre maximum d'individus et est aussi grand qu'on veut, limité par l'unité de stockage des données.
  
3. La plupart des progiciels sont constitués de plusieurs programmes que l'utilisateur doit appeler à chaque étape de traitement. INDA fait lui-même l'enchaînement de tous les modules nécessaires à l'analyse d'un tableau de données allant de l'aide au choix et de l'acquisition des données en passant par l'exécution de la méthode choisie et terminant par le graphique.
  
4. INDAL constitue le noyau numérique commun à toutes les techniques du MENU, cela a pour avantage de réduire considérablement la taille du module INDA\_EM d'exécution des méthodes. Un autre facteur agissant dans ce sens est la structure même de INDA\_EM qui en lui-même contient toutes les méthodes factorielles. La plupart des logiciels existant exécutent chaque technique du MENU par un programme différent ce qui les rend rigides d'utilisation

L'algorithme le plus cher pour toutes les méthodes factorielles est le calcul des éléments propres d'une matrice symétrique et définie positive. C'est ce fait qui a entraîné la non interactivité de la majorité des progiciels sur micro-ordinateur : le temps de calcul étant trop grand pour qu'il soit envisageable que l'utilisateur attende la réponse devant son écran. D'autre part, les méthodes numériques utilisées sont peu performantes.

Pour ce qui concerne INDA, la méthode de calcul des éléments propres utilisée ici est celle de Jacobi. Ce choix a été fait dans le but de pouvoir comparer les performances du logiciel avec celles de la maquette systolique. Notons cependant que ce choix n'est pas le meilleur ; en effet pour une matrice d'ordre  $k$  on calcule les  $k$  éléments propres alors qu'on n'a besoin en général que des dominants.

Il est prévu d'incorporer d'autres algorithmes de calcul des éléments propres, ce qui est très facile vu la modularité de INDA.

Le premier algorithme consiste en une généralisation de la méthode de la puissance, dite méthode des itérations de sous-espaces, accéléré par TCHEBYCHEV. Le second, est celui de LANCZOS qui est particulièrement intéressant pour des matrices de très grandes tailles (cf. travail de M. DIAMOUTANI).

5. Une comparaison des temps de calcul entre INDA et le progiciel CHADOC a été faite sur un échantillon d'un groupe de 37 étudiants de la MIAG prenant leurs notes d'examens dans neuf disciplines. On a donc testé une ACP sur cet exemple dont la matrice des observations est de taille  $37 \times 9$  et on a obtenu les résultats suivants : (la comparaison a été faite avec B. DURET, utilisateur de CHADOC).

temps total d'exécution	CHADOC	INDA
	8 mn	3 mn
calcul des éléments propres	2mn 30s	45 s

Signalons tout de même que CHADOC est écrit en BASIC.

*TRAIEMENT D' UN EXEMPLE*



## La structure fonctionnelle des dépenses de l'Etat (1872-1971) (en %)

	Pensions publiques PVP	Agriculture AGN	Commerce et industrie CMI	Transports TIA	Logement et aménagement du territoire LOG	Education et culture EDU	Action sociale ACS	Anciens combattants ACO	Défense DEF	Dette DEF	Bien-être DIV	Total
1872	18,0	0,5	0,1	6,7	0,5	2,1	2,0		26,4	41,5	2,1	100
1880	14,1	0,8	0,1	15,3	1,9	3,7	0,5		29,8	31,3	2,5	100
1890	13,6	0,7	0,7	6,8	0,6	7,1	0,7		33,9	34,4	1,7	100
1900	14,3	1,7	1,7	6,9	1,2	7,4	0,8		37,7	26,2	2,2	100
1903	10,3	1,5	0,4	9,3	0,6	8,5	0,9		38,4	27,2	3,0	100
1906	13,4	1,4	0,5	8,1	0,7	8,6	1,8		38,5	25,3	1,9	100
1909	13,5	1,1	0,5	9,0	0,6	9,0	3,4		36,8	23,5	2,6	100
1912	12,9	1,4	0,3	9,4	0,6	9,3	4,3		41,1	19,4	1,3	100
1920	12,3	0,3	0,1	11,9	2,4	3,7	1,7	1,9	42,4	23,1	0,2	100
1923	7,6	1,2	3,2	5,1	0,6	5,6	1,8	10,0	29,0	35,0	0,9	100
1926	10,5	0,3	0,4	4,5	1,8	6,6	2,1	10,1	19,9	41,6	2,3	100
1929	10,0	0,6	0,6	9,0	1,0	8,1	3,2	11,8	28,0	25,8	2,0	100
1932	10,6	0,8	0,3	8,9	3,0	10,0	6,4	13,4	27,4	19,2	0	100
1935	8,8	2,6	1,4	7,8	1,4	12,4	6,2	11,3	29,3	18,5	0,4	100
1938	10,1	1,1	1,2	5,9	1,4	9,5	6,0	5,9	40,7	18,2	0	100
1947	15,6	1,6	10,0	11,4	7,6	8,8	4,8	3,4	32,2	4,6	0	100
1950	11,2	1,3	10,5	12,4	15,8	8,1	4,9	3,4	20,7	4,2	1,5	100
1953	12,9	1,5	7,0	7,9	12,1	8,1	5,3	3,9	36,1	5,2	0	100
1956	10,9	5,3	9,7	7,6	9,6	9,4	8,5	4,6	28,2	6,2	0	100
1959	13,1	4,4	7,3	5,7	9,8	12,5	8,0	5,0	26,7	7,5	0	100
1962	12,8	4,7	7,5	6,6	6,8	15,7	9,7	5,3	24,5	6,4	0,1	100
1965	12,4	4,3	8,4	9,1	6,0	19,5	10,6	4,7	19,8	3,5	1,8	100
1968	11,4	6,0	9,5	5,9	5,0	21,1	10,7	4,2	20,0	4,4	1,9	100
1971	12,8	2,8	7,1	8,5	4,0	23,8	11,3	3,7	18,8	7,2	0	100

Source : C. ANDRÉ et R. DELORME, *L'évolution des dépenses publiques en France (1872-1971)*  
rapport CORDES, CEPREMAP, 1976.

remarque: Le fait d'avoir choisi les données en pourcentage, plutôt que les valeurs en francs, évite les variations de l'unité monétaire au fil des années .

Voici ce qui apparaît à l'écran lors de l'exécution de ce logiciel :

(réponses de l'utilisateur en majuscules)

Titre de votre exemple :>>DEPENSES DE L'ETAT

Menu :

- 1 ACP d'un nuage de points
- 2 ACP unitaire d'un nuage de points
- 3 ACP normée d'un nuage de points
- 4 ACP d'un vecteur aléatoire
- 5 ACP unitaire d'un vecteur aléatoire
- 6 ACP normée d'un vecteur aléatoire
- 7 AFC sur les lignes
- 8 AFC sur les colonnes
- 9 Corrélation canonique entre deux variables aléatoires discrètes
- 10 AFD
- 11 Corrélation canonique entre un vecteur aléatoire et une variable aléatoire discrète
- 12 Corrélation canonique entre deux vecteurs aléatoires

1:Choix d'une des méthodes du menu

2:Demande d'aide au choix

>>2

Les données se présentent elles sous forme d'une matrice  $X_0$  à termes entiers positifs ?

\*oui:0  
\*non:N

>> N

Se présentent elles alors sous l'une des formes suivantes :  
-- n vecteurs à k composantes réelles :  $S_1, \dots, S_n$  appelés "individus"  
-- ou un tableau  $k \times n$  à termes réels  $X_{ij}$  dont on peut appeler "individus" soit les lignes soit les colonnes ?

\*oui:0  
\*non:N

>> 0

Est ce qu'on peut définir concrètement k caractères qui soient les mêmes pour tous les individus, et tels que le terme  $X_{ij}$  de la matrice des données ( de la ième composante du jème vecteur ) soit la mesure quantitative du caractère i pour l'individu j ?

\*oui:0  
\*non:N

>> 0

Est ce que les individus forment un échantillon aléatoire sur  $IR^k$  de taille n ?

\*oui (c'est à dire qu'ils ont été tirés au hasard indépendamment dans une population théorique) :0

\*non (on peut alors attacher à chaque individu j une importance mesurée par une masse  $a_j > 0$ ) :N

\*je ne sais pas :X

>> N

Votre finalité est elle ?

\*inférentielle :I  
\*descriptive :D  
\*je ne sais pas :X

>> D

Comment peut on définir la distance entre deux individus ?

\*On dispose d'une norme sur IRk définie par une matrice B symétrique définie positive :A

\*En faisant la somme des carrés des différences de leurs k caractères ,c'est à dire en accordant la même importance à chaque caractère (norme unité) :B

\*En faisant la somme des carrés des différences de leurs k caractères ,pondérée par l'inverse des variance de ces caractères , c'est à dire en ramenant leurs mesures à la même échelle :C

>> C

Je vous propose d'estimer les axes principaux de dispersion de l'ensemble des n vecteurs munis de masses  $a_1, \dots, a_n$  dans IRk muni de la norme choisie .  
Cela vous convient il ?

\*oui:O

\*non:N

>> O

Méthode proposée : ACP normée d'un nuage de points CM = 3

Vos données :

-existent sous une forme standard (XXX.DON) : S

-existent sous une forme brute (NOMFICH) : B

>> B

Ce fichier brut ne doit contenir que :

-- Un tableau de données .

-- Un vecteur des masses des individus .

Ces données doivent être écrites à la suite les unes des autres , dans n'importe quel ordre .

Les autres informations (nombre de lignes,nombre de colonnes,étiquettes,...) seront demandées ultérieurement .

Si cela n'est pas le cas taper Q pour quitter le programme, et modifier votre fichier sinon taper n'importe quel autre caractère .

>>

Pour une ACP normée, il faut :

-- Un tableau de données (n individus \* k caractères) .

-- Un vecteur des n poids des individus .

Je vais lire votre fichier et le recopier sous une forme standard

Dans votre fichier brut les individus correspondent ils :

-- A des lignes : L

-- A des colonnes : C

>> L

Quel est le nombre N d'individus ?

>> 24

Quel est le nombre K de caractères ?

>> 11

Avez vous écrit :

-- Le tableau des données en premier : D

-- Le vecteur des poids en premier : P

>> D

Voulez vous vérifier la structure du fichier standard créé ?

>> N

Voulez vous donner des étiquettes aux individus ? (0/N)

>> 0

Que préférez vous ?

-- Donner vos propres étiquettes . :E

-- Donner des étiquettes standard (1,2,...,N) :S

>> E

Page 1

étiquette1: >>1872

étiquette2: >>1880

.

.

.

étiquette20: >>1959

Page correcte ? >> 0

Page 2

étiquette21: >>1962

.

.

étiquette24: >>1971

Page correcte ? >> 0

Voulez vous donner des étiquettes aux caractères ? (0/N)

.

.

.

oix de sortie des résultats :

- sur terminal : (Tapez 'T')
- sur imprimante : (Tapez 'I')

les sorties obtenues sur l'imprimante :

TRAITEMENT DE L'EXEMPLE : DEPENSES DE L'ETAT  
FICHER DE DONNEES : B:DEPETA.DON  
MATRICE DE CORRELATION :

```

*-----*
RVP* 1.00*
*-----*-----*
AGA*-0.08* 1.00*
*-----*-----*
CNI*-0.00* 0.60* 1.00*
*-----*-----*
TRA* 0.23*-0.28* 0.09* 1.00*
*-----*-----*
LOG* 0.04* 0.44* 0.89* 0.17* 1.00*
*-----*-----*
EDU*-0.15* 0.73* 0.47*-0.21* 0.23* 1.00*
*-----*-----*
ACS*-0.13* 0.81* 0.62*-0.20* 0.49* 0.87* 1.00*
*-----*-----*
ACO*-0.69* 0.04* 0.02*-0.31* 0.04* 0.16* 0.29* 1.00*
*-----*-----*
DEF* 0.10*-0.45*-0.54* 0.10*-0.38*-0.52*-0.57*-0.42* 1.00*
*-----*-----*
DET* 0.03*-0.70*-0.80*-0.15*-0.76*-0.67*-0.81*-0.05* 0.26* 1.00*
*-----*-----*
DIV* 0.15*-0.28*-0.35* 0.11*-0.44*-0.25*-0.53*-0.38* 0.02* 0.55* 1.00*
RVP  AGA  CNI  TRA  LOG  EDU  ACS  ACO  DEF  DET  DIV

```

VALEURS PROPRES :

```

*-----*-----*-----*-----*-----*
*          119.39411*          49.18483*          30.93500*          23.83410*          17.00295*

```

% D'INERTIE :

```

*-----*-----*-----*-----*-----*
*          0.45*          0.19*          0.12*          0.09*          0.06*

```

% CUMULE :

```

*-----*-----*-----*-----*-----*
*          0.45*          0.64*          0.76*          0.85*          0.91*

```

VECTEURS PROPRES :

```

*-----*-----*-----*-----*-----*
*          -0.07785*          0.51698*          -0.29997*          -0.10730*          -0.40487*
*-----*-----*-----*-----*-----*
*          0.36713*          0.00464*          -0.32248*          -0.15362*          0.03884*
*-----*-----*-----*-----*-----*
*          0.37370*          0.23774*          0.12557*          0.25826*          -0.17931*
*-----*-----*-----*-----*-----*
*          -0.06165*          0.44013*          0.33209*          0.28118*          0.64963*
*-----*-----*-----*-----*-----*
*          0.32378*          0.27761*          0.34073*          0.20754*          -0.31192*
*-----*-----*-----*-----*-----*
*          0.35290*          -0.09490*          -0.37420*          -0.11527*          0.37680*
*-----*-----*-----*-----*-----*
*          0.41846*          -0.06997*          -0.14626*          -0.15088*          0.12332*
*-----*-----*-----*-----*-----*
*          0.12951*          -0.56441*          0.32940*          0.20235*          -0.01384*
*-----*-----*-----*-----*-----*
*          -0.27463*          0.15052*          0.22744*          -0.63963*          0.18394*
*-----*-----*-----*-----*-----*
*          -0.39875*          -0.20986*          -0.13936*          0.17980*          -0.20970*
*-----*-----*-----*-----*-----*
*          -0.24552*          0.07909*          -0.47202*          0.50754*          0.21890*

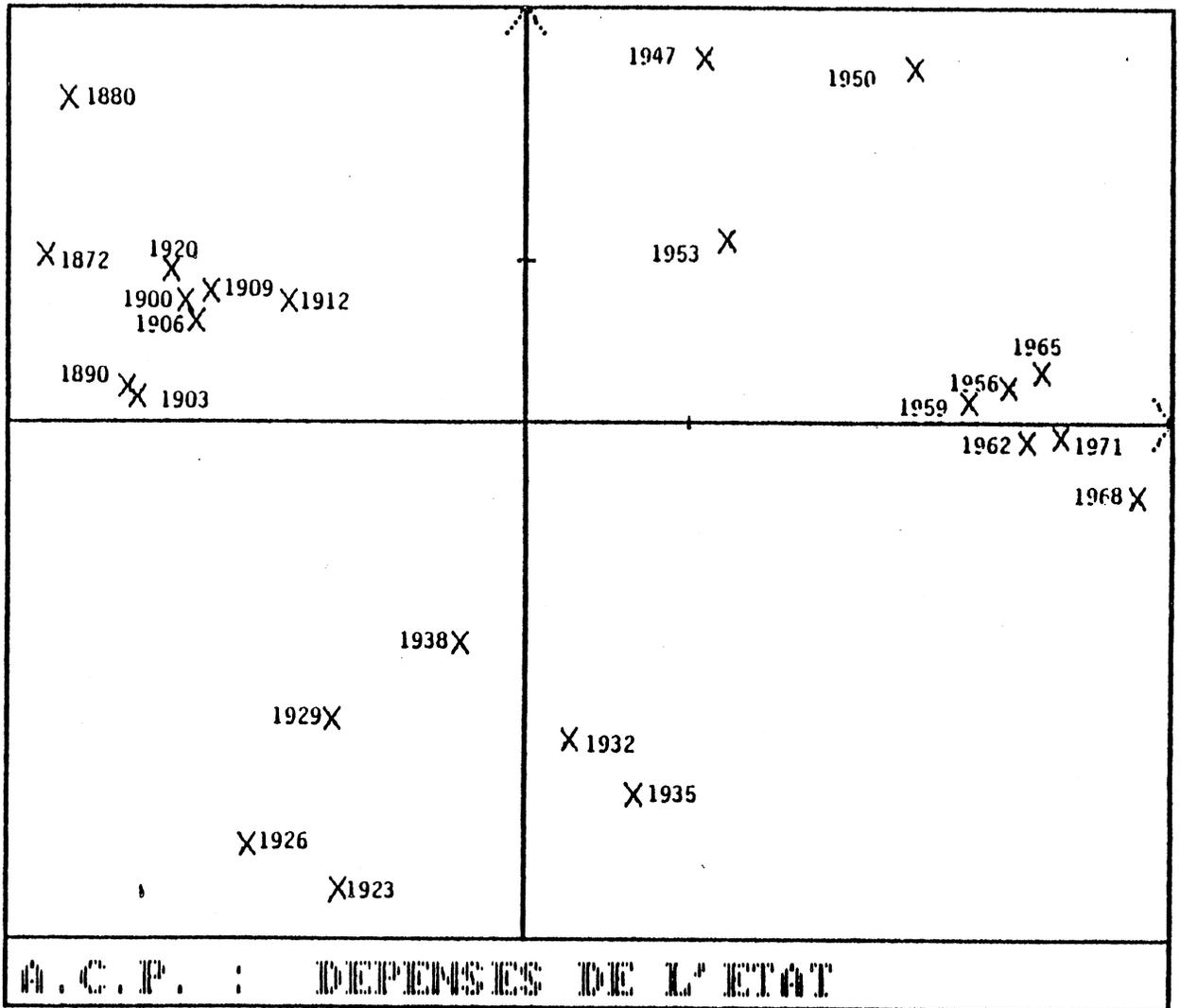
```

* -0.388219*	* 1.654704*	* -0.761445*	* -0.239072*	* -0.76191
* 2.437412*	* 0.019762*	* -1.089791*	* -0.455672*	* 0.09720
* 0.910769*	* 0.371896*	* 0.155784*	* 0.281227*	* -0.16491
* -0.272958*	* 1.250791*	* 0.748464*	* 0.556253*	* 1.08547
* 0.844944*	* 0.465555*	* 0.453172*	* 0.242291*	* -0.30756
* 0.738256*	* -0.127416*	* -0.398463*	* -0.107737*	* 0.29746
* 1.341363*	* -0.143947*	* -0.238641*	* -0.216083*	* 0.149174
* 0.340607*	* -0.952696*	* 0.440961*	* 0.237769*	* -0.013731
* -0.410540*	* 0.144418*	* 0.173060*	* -0.427211*	* 0.103763
* -0.356725*	* -0.120499*	* -0.063459*	* 0.071866*	* -0.070796
* -2.615306*	* 0.540741*	* -2.559352*	* 2.41556	* 0.879948

## COORDONNEES DES POINTS SUR LES AXES FACTORIELS

* -2.899857*	* 1.026953*	* -1.560052*	* 0.488915*	* -2.058912
* -2.767470*	* 2.012185*	* 0.174292*	* 1.482575*	* 1.235463
* -2.415937*	* 0.225037*	* -0.765564*	* -0.266150*	* -0.709893
* -2.056443*	* 0.755956*	* -1.007935*	* -0.520629*	* -0.503176
* -2.337478*	* 0.167773*	* -0.624317*	* 0.183305*	* 1.213872
* -1.985085*	* 0.626435*	* -0.694476*	* -0.708201*	* 0.151530
* -1.907241*	* 0.812844*	* -0.988194*	* -0.198542*	* 0.594596
* -1.431466*	* 0.176933*	* -0.197939*	* -1.289421*	* 0.781215
* -2.140116*	* 0.453170*	* 1.745004*	* -1.067078*	* 0.628052
* -1.142736*	* -2.884422*	* 0.864152*	* 0.435796*	* -0.540472
* -1.673280*	* -2.609311*	* -0.497169*	* 1.763380*	* -1.193910
* -1.173533*	* -1.831995*	* 0.608953*	* 1.155587*	* 0.695247
* 0.269699*	* -1.961556*	* 1.459098*	* 0.038248*	* 0.344518
* 0.658483*	* -2.297477*	* 0.658447*	* -0.308023*	* 0.823449
* -0.403117*	* -1.345034*	* 0.843511*	* -1.849742*	* 0.063884
* 1.079369*	* 2.248784*	* 1.278470*	* -0.234056*	* -0.373120
* 2.370770*	* 2.173546*	* 1.926733*	* 2.658372*	* -0.184836
* 1.201895*	* 1.131414*	* 1.662171*	* -0.755558*	* -0.959842
* 2.926493*	* 0.229701*	* 0.590818*	* -0.445222*	* -0.509861
* 2.684905*	* 0.139764*	* -0.071059*	* -0.690164*	* -1.212549
* 3.053571*	* -0.110680*	* -0.586814*	* -0.643739*	* -0.421138
* 3.142154*	* 0.312420*	* -1.411069*	* 0.766992*	* 0.928107
* 3.708216*	* -0.457460*	* -2.292605*	* 0.281674*	* 0.479704
* 3.238206*	* -0.085739*	* -1.114456*	* -0.278319*	* 0.728070

Voici les résultats obtenus sur l'écran graphique :





**BIBLIOGRAPHIE**



- [1] J.R. BARRA (1981 a)  
Mathematical basis of statistics  
Académic Press, New York.
  
- [2] J.R. BARRA (1981 b)  
Phase inductive et phase décisionnelle dans l'étude statistique  
d'un psychotrope.  
Grenoble 1981.
  
- [3] J.R. BARRA (1981 c)  
Approche inférentielle de l'Analyse des Données.  
Grenoble, Séminaire de statistique. Janvier 1981.
  
- [4] J.R. BARRA (1983)  
Statistical methods in psychiatry. Virtual Models.  
7ème congrès Mondial de Psychiatrie, Vienne 1983. Plenum Press.
  
- [5] J.R. BARRA, M. BECKER (1985)  
Conception d'un logiciel d'assistance intelligente en Analyse  
factorielle et canonique.  
Fourth International Symposium on Data Analysis and Informatics.  
Versailles Octobre 1985 (sera publié).
  
- [6] J.R. BARRA, M. BECKER, D. BELAID, F. CHATELIN (1984)  
Un logiciel d'Analyse Factorielle. Une maquette systolique  
programmable.  
R.R. 450, Juin 1984, IMAG Grenoble.
  
- [7] J.R. BARRA, M. BECKER, D. BELAID, F. CHATELIN, C. MAZEL (1985)  
Réalisation d'un logiciel d'Analyses Factorielles avec système  
d'assistance intelligente à l'utilisateur.  
R.R. 542, Juin 1985. IMAG Grenoble
  
- [8] J.R. BARRA, M. BECKER, F. CHATELIN (1984)  
A processor array for Data Analysis.  
Frontiers in computational Statics. A.S.A.-I.A.S.C.-SIAM.  
Conference, Boston, Octobre 1984.

- [9] J.R. BARRA, F. CHATELIN (1984)  
Propositions pour un logiciel d'Analyse de Données  
(Analyses Factorielles) sur micro-ordinateurs.  
IMAG, Grenoble, 1984
- [10] F. CHATELIN (1983)  
Spectral approximation of linear operators.  
Academic Press. New York.
- [11] F. CHATELIN (1986)  
Valeurs propres de matrices.  
Masson, Paris. A paraître.
- [12] J.P. CHEVALLIER (1977)  
Problèmes méthodologiques posés par l'utilisation de l'Analyse  
des Données.  
Psychologie médicale, 9-2.
- [13] Y. ESCOUFFIER(1982)  
L'analyse des tableaux de contingence simples et multiples.  
Metron XL, n° 1-2, p. 53-77.
- [14] R.A. FISHER (1956)  
Statistical methods for research workers.  
Ninth. ed Edinburg London.
- [15] T. FOUCART (1982)  
Analyse factorielle. Programmation sur micro-ordinateurs  
Masson. Paris.
- [16] L. LEBART (1983)  
EDA applied to large sets of data : a critical balance.  
Meeting 20. 44 ISI Madrid.
- [17] J. LEFEBVRE, M. BOITARD, J.F. REY (1981)  
Logiciels d'analyses statistiques multidimensionnelles sur  
micro-ordinateurs.  
Masson, Paris.

- [18] J. LEMAIRE (1984)  
CHADOC sous CP/M 86  
Brochure d'utilisation du progiciel CHADOC NICE - 1984
  
- [19] J.P. PAGES, F. CAILLIEZ, Y. ESCOUFFIER (1979)  
Analyse factorielle. Un peu d'histoire et de géométrie.  
Rev. Stat. Appl. 27, 6-28.
  
- [20] SCHNEIDER (1981)  
The role of hypotheses testing in clinical trials.  
Methods of information in medicine.  
Volume 20 n° 2.
  
- [21] J.W. TUKER (1977)  
Exploratory data analysis.  
Addison Wesley.
  
- [22] VICTOR (1982)  
Exploratory data analysis and clinical research.  
Methods of information in medicine. Vol. 21.



AUTORISATION DE SOUTENANCE

DOCTORAT 3<sup>ème</sup> CYCLE, DOCTORAT-INGENIEUR, DOCTORAT USMG

/u les dispositions de l'arrêté du 16 avril 1974,

/u les dispositions de l'arrêté du 5 juillet 1984,

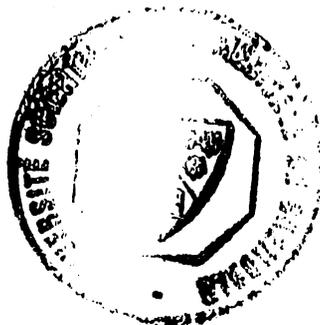
/u les rapports de M. *adame Chabelin* .....

M. ....

*Monsieur Djamel BELAÏD* ..... est autorisé  
à présenter une thèse en vue de l'obtention du *grade* .....  
*de Docteur de 3<sup>ème</sup> cycle* .....

Grenoble, le **21 OCT. 1985** .....

Le Président de l'Université Scientifique  
et Médicale



*M. Tanche*  
M. TANCHE



## RESUME

On présente dans ce travail un logiciel interactif sur micro-ordinateur, des méthodes factorielles d'Analyse de Données. Ce logiciel est composé des modules :

INDA\_CM : module d'aide au choix à l'utilisateur d'une des méthodes proposées.

INDA\_AD : module d'acquisition des données nécessaires à l'exécution de la méthode choisie.

INDA\_EM : module d'exécution des méthodes proposées.

INDA\_GR : module de visualisation à l'écran graphique d'un nuage de points.

INDA : programme de gestion des modules précités.

MOTS CLES : méthodes factorielles - CHOLESKI - JACOBI - inversion.