



**HAL**  
open science

# Contribution à la classification automatique de données de production en technologie de groupe

Abdellah Nadif

► **To cite this version:**

Abdellah Nadif. Contribution à la classification automatique de données de production en technologie de groupe. Sciences de l'ingénieur [physics]. Université Paul Verlaine - Metz, 1987. Français. NNT : 1987METZ016S . tel-01775729

**HAL Id: tel-01775729**

**<https://hal.univ-lorraine.fr/tel-01775729v1>**

Submitted on 24 Apr 2018

**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.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

THESE

présentée à

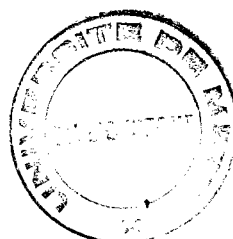
L'UNIVERSITE DE METZ

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITE DE METZ

par

Abdellah NADIF



-----  
CONTRIBUTION A LA CLASSIFICATION AUTOMATIQUE  
DE DONNEES DE PRODUCTION EN TECHNOLOGIE DE GROUPE  
-----

soutenue le 24 Septembre 1987 devant la commission d'examen:

MM. ARMBRUSTER  
GOVAERT  
MUTEL  
PROTH  
SAUVAIRE  
VERON

Rapporteur

Rapporteur

BIBLIOTHEQUE UNIVERSITAIRE - METZ	
N° inv.	1987 0325
Cote	S/M3 87/16
Loc	Magasin

## REMERCIEMENTS

Ce travail a été réalisé dans le cadre d'une étroite collaboration entre l'Agence Nationale pour le Développement de la Production Automatisée (A.D.E.P.A) et le Laboratoire d'Automatique et d'Electronique Industrielle (L.A.E.I.). Je tiens à remercier leurs directeurs, monsieur Mutel et monsieur Sauvaire, grâce à qui j'ai pu entreprendre ce travail et à toutes les personnes de ces organismes qui m'ont aidé à le mener à bien, et en particulier MM. Ballot, Chebi, Minot et Souvestre.

Je remercie vivement d'une part MM. Lechevallier et Ralambondrainy de l'INRIA (Recquencourt), et d'autre part MM. Dollé et Lambert de la DAST (R.N.U.R), pour l'intérêt qu'ils ont porté à mon travail et pour leur collaboration aimable et efficace.

Enfin, j'associe à ces remerciements:

MM. Armbruster	(Professeur, Directeur de l'ENSAIS)
Govaert	(INRIA, Professeur à l'Université de Metz)
Mutel	(Directeur du LAEI, Professeur à l'Université de Metz)
Proth	(Directeur de recherche à INRIA)
Sauvaire	(Directeur de l'ADEPA)
Véron	(Directeur du LACN, Professeur à l'Université de Nancy)

qui m'ont fait l'honneur de participer à ce jury.

## RESUME

### CONTRIBUTION A LA CLASSIFICATION AUTOMATIQUE DE DONNEES DE PRODUCTION EN TECHNOLOGIE DE GROUPE

La technologie de groupe se définit comme un concept d'analyse et un outil de rationalisation de la production.

Le système de production est considéré comme un processus complexe tenant compte de très nombreuses variables de natures diverses, prenant des valeurs très différentes et interagissant de manières multiples.

L'observation de ce système sur un horizon déterminé oblige à prendre en compte un ensemble important de données, dont la structuration dépasse les possibilités humaines.

La technologie de groupe consiste à regrouper les données techniques et économiques en familles homogènes (Classification) afin d'optimiser les flux de matières et faciliter la mise au point de solutions utilisant le savoir-faire de l'entreprise en tirant profit de l'analogie au niveau des formes, des dimensions, des devis,... et des processus de fabrication.

L'utilisation des systèmes de classification actuellement existants en technologie de groupe n'est pas satisfaisante. Elle nécessite beaucoup de temps et de clairvoyance de l'utilisateur, sans pour autant garantir un résultat objectif.

En dehors de ces systèmes, l'Analyse de Données propose un grand nombre d'algorithmes, parmi lesquels l'industriel n'a pas les moyens de choisir.

La présente étude tente de clarifier cette situation et propose une méthodologie, dans le but de permettre une utilisation rationnelle des méthodes d'Analyse des Données.

Cette méthodologie a donné lieu à l'élaboration du logiciel CL<sup>2</sup>AUDIA (Classification et Classement Automatique de Données Industrielles Analytiques).

Une application industrielle de ce logiciel à des données de production dans le domaine de l'Industrie Mécanique a été développée.

Des familles de pièces satisfaisantes ont ainsi pu être déterminées à l'aide du logiciel CL<sup>2</sup>AUDIA.

Les principales difficultés auxquelles se heurtaient les systèmes de classification traditionnels ont pu être aplanies.

## TABLE DES MATIERES

I/	GENERALITES .....	1
I-1/	INTRODUCTION .....	2
I-2/	TECHNOLOGIE DE GROUPE .....	2
	a) Définition .....	2
	b) Avantages .....	3
I-3/	IMPLANTATION DE LA TECHNOLOGIE DE GROUPE ET LIMITES DES SYSTEMES ACTUELS .....	6
I-4/	OBJECTIFS ET PRESENTATION DE CETTE ETUDE .....	12
II/	ANALYSE DES SYSTEMES ET DES TECHNIQUES DE TECHNOLOGIE DE GROUPE ....	15
II-1/	ANALYSE DES SYSTEMES DE CLASSIFICATION EN TECHNOLOGIE DE GROUPE ....	16
II-1-A/	Systèmes basés sur des tris .....	18
II-1-B/	Méthodes utilisant des mesures de ressemblance .....	19
II-2/	ETUDE DES PRINCIPAUX ALGORITHMES D'ANALYSE DE DONNEES POUR LA TECHNOLOGIE DE GROUPE .....	25
II-2-A/	Représentation graphique des données .....	30
II-2-B/	Obtention de familles (Classification) .....	33
II-2-C/	Regroupement de familles .....	35
II-3/	CONCLUSION .....	37
III/	METHODOLOGIE D'IMPLANTATION DE GROUPEMENTS TECHNOLOGIQUES .....	39
III-1/	COLLECTE DES DONNEES .....	40
III-1-A/	Choix des objets à analyser .....	42
III-1-B/	Choix des variables décrivant les objets .....	42
	a) Variables Quantitatives .....	43

	AND ET CAH .....	117
IV-3-A/	Classification automatique .....	117
IV-3-B/	Classement et Validation .....	126
V/	CONCLUSION .....	128
VI/	BIBLIOGRAPHIE .....	132
VII/	ANNEXES .....	143
VII-1/	EXEMPLES DE SYSTEMES DE CODIFICATION AUTRES QUE LE CODE MULTI-M ...	144
	Code OPITZ .....	145
	Code CETIM-PMG .....	148
	Code PEKLENIK .....	150
VII-2/	ALGORITHME DES K-MEANS DE MACQUEEN .....	152
VII-3/	EXEMPLES DE MESURES DE PROXIMITE CLASSIQUES .....	153
VII-4/	ALGORITHME DE FISCHER. EXEMPLE D'APPLICATION .....	158
VII-5/	TRANSCODAGE DU CODE MULTI-M .....	161
VII-6/	ALGORITHME DES NUEES DYNAMIQUES .....	166
VII-7/	RESULTATS DE LA CLASSIFICATION EFFECTUEE SUR DES DONNEES INDUSTRIELLES.....	170
VII-8/	RESULTAT DU REGROUPEMENT DES FORMES FORTES.....	175

	b) Variables Qualitatives Nominales .....	43
	c) Variables Qualitatives Ordinales .....	44
	d) Cas particulier des données de production .....	44
	d-1) Gammes de fabrication .....	45
	d-2) Autres données de production .....	45
III-2/	TRANSCODAGE .....	48
III-2-A/	Données ordinales incomplètes (cas des gammes de fabrication) .....	48
III-2-B/	Autres données .....	50
III-3/	CALCUL DE PROXIMITE .....	52
III-3-A/	Généralités .....	52
III-3-B/	Application aux données de production .....	58
	a) Choix d'un indice de proximité entre données ordinales incomplètes (application aux gammes de fabrication) .....	59
	a-1) Obtention de distances euclidiennes .....	60
	a-2) Etude sur un échantillon test .....	61
	a-3) Validation .....	69
	b) Choix d'un indice de proximité pour les autres données de production (application aux codes morphodimensionnels) .....	71
III-4/	REPRESENTATION GRAPHIQUE DES DONNEES .....	73
III-5/	CLASSIFICATION AUTOMATIQUE PAR LA METHODE DES NUEES DYNAMIQUES .....	75
III-6/	DESCRIPTION DES FAMILLES OBTENUES .....	81
III-7/	CLASSEMENT ET VALIDATION .....	84
III-8/	CONCLUSION .....	86
IV/	LOGICIEL CL <sup>2</sup> AUDIA ET APPLICATION INDUSTRIELLE .....	87
IV-1/	ANALYSE DE L'ECHANTILLON ET PREPARATION DES DONNEES .....	90
IV-1-A/	Définition du code MULTI-M .....	91
IV-1-B/	Procédure "PTG" .....	95
	a) MENU1.PTG .....	95
	b) MENU2.PTG .....	99
	b-1) Transcodage du code MULTI-M .....	101
	b-2) Matrice de distances entre codes .....	103
	b-3) Matrice de distances entre gammes .....	106
IV-2/	REPRESENTATION GRAPHIQUE DES REFERENCES ANALYSEES PAR ACPTD .....	107
IV-3/	CLASSIFICATION AUTOMATIQUE ET CLASSEMENT PAR LES PROCEDURES	



**CHAPITRE I**

**GENERALITES**

## I-1/ INTRODUCTION

Les besoins de productivité, de capacité d'adaptation aux variations de la demande, de respect des délais et de la qualité, la nécessité de diminuer les coûts de conception et de fabrication obligent les entreprises à intensifier les mesures de rationalisation à tous les échelons du processus de la production, afin d'accroître leur compétitivité.

Certains experts [1][2] estiment à 75% le taux de pièces fabriquées en mécanique en séries de moins de 50 exemplaires. Dans 95% des cas, au bureau d'étude pour la conception de pièces, il s'agit de travaux de recopie et d'exécution de variantes. L'usinage pour sa part, ne représente que 6% du temps total de la présence d'une pièce dans l'atelier.

Pour une entreprise travaillant en petites séries, chaque commande est particulière et seulement 10% des offres se transforment en commande ferme. Dans ce cas, on ne peut consacrer beaucoup de temps à effectuer les devis et établir les offres.

Une organisation basée sur les concepts de la Technologie de Groupe permet d'apporter une solution très intéressante à tous ces problèmes.

## I-2/ TECHNOLOGIE DE GROUPE [3 à 7]

### a) Définition

La technologie de groupe se définit comme un concept

d'analyse et un outil de rationalisation de la production. Elle consiste à regrouper les données techniques et économiques de l'entreprise en familles homogènes (Classification) afin d'optimiser leur flux. Elle facilite la mise au point de solutions utilisant le savoir faire de l'entreprise en tirant profit de l'analogie au niveau des formes, des dimensions, des tolérances, de la matière, du devis et du processus de fabrication.

#### b) Avantages

Une organisation basée sur la technologie de groupe permet d'obtenir des gains de temps et d'argent aussi bien au bureau d'études, au bureau des méthodes qu'aux ateliers (figure-I-1)[3].

-Au bureau d'études la technologie de groupe conduit à une standardisation des éléments de formes, des dimensions et des matières. A partir du modèle paramétré ou des pièces types caractéristiques d'une famille, le concepteur peut définir de nombreuses variantes en jouant sur les paramètres de définition de la pièce type ou du modèle (figure-I-2).

RD / Les tâches de conception deviennent moins répétitives, plus rapides, plus rationnelles et plus motivantes. L'analyse de la valeur peut alors être utilisée plus intensément pour la définition des produits nouveaux.

-Au bureau des méthodes, la technologie de groupe implique une démarche plus rigoureuse et plus précise dans l'élaboration des gammes d'usinage et de montage. Le principal objectif, ici,

Bureau d'études	Bureau des méthodes	Atelier	
Conception	Préparation du travail	Fabrication	Montage
<ul style="list-style-type: none"> <li>* Elaboration de fichiers de pièces répétitives et/ou similaires</li> <li>* Elaboration de dessins standards</li> <li>* Utilisation des principes de conception modulaire</li> <li>* Développement de CAO</li> <li>* Systématisation de l'élaboration des offres</li> </ul>	<ul style="list-style-type: none"> <li>* Elaboration de fichiers de dessins et gammes</li> <li>* Création de familles d'usinage et de montage indépendantes du produit</li> <li>* Développement des directives de durées, d'outillages et de montage standards</li> <li>* Elaboration de gammes d'usinage et de montage standards</li> <li>* Développement de méthodes pour l'élaboration automatique des gammes d'usinage et de montage</li> <li>* Systématisation du planning d'investissements</li> </ul>	<ul style="list-style-type: none"> <li>* Fabrication par familles</li> <li>* Ajustement des moyens de fabrication avec le spectre de production</li> <li>* Réalisation de formes d'organisation développées</li> <li>* Organisation en îlots de fabrication</li> </ul>	<ul style="list-style-type: none"> <li>* Montage de familles de produits</li> <li>* Ajustement des installations de montage avec le spectre produit</li> <li>* Réalisation d'une meilleure organisation</li> </ul>

Figure-1-1: Domaines d'application de la technologie de groupe [3]

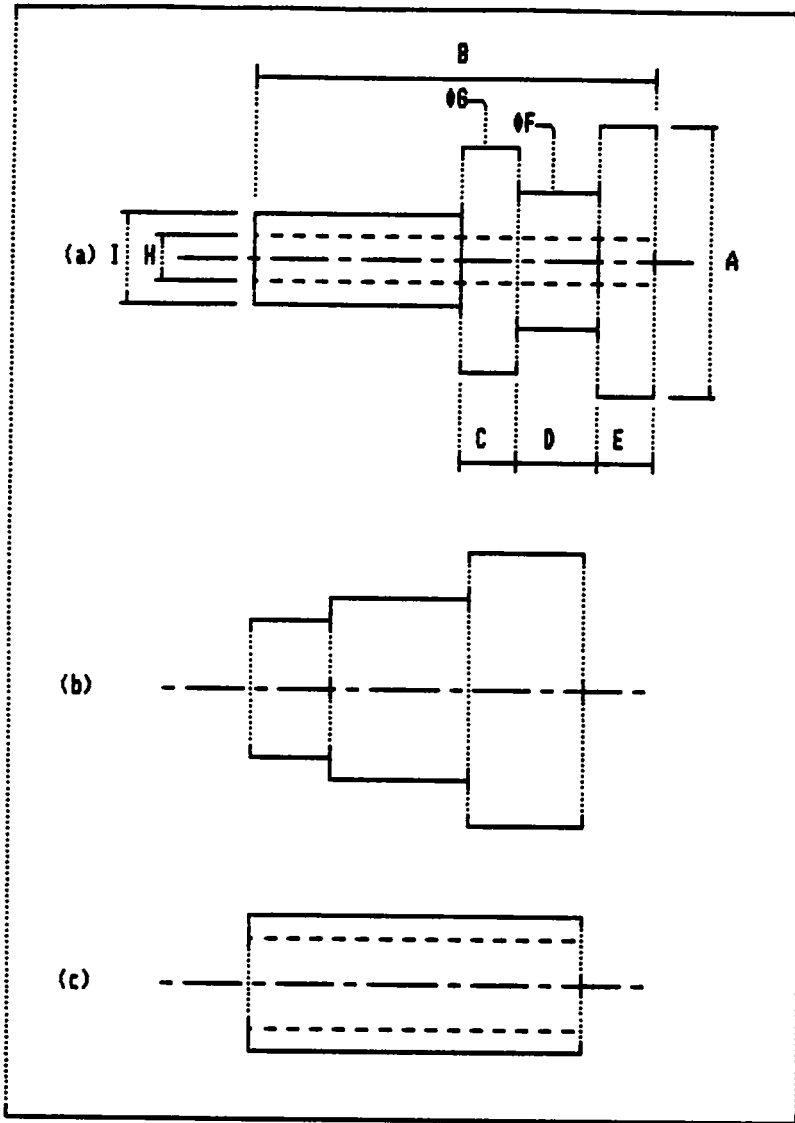


Figure-1-2: Modèle paramétré (a) et deux variantes (b et c).

est de créer des familles de pièces pour optimiser leur fabrication. La gamme standard de chaque famille sert de base à l'élaboration de la gamme de chaque élément nouveau de la famille (figure-1-3)[3].

Pour le préparateur, les choix sont moins arbitraires, le travail moins fastidieux, les solutions plus rationnelles et plus rapides. Il peut consacrer plus de temps à l'utilisation des techniques nouvelles pour améliorer encore plus la productivité.

-Aux ateliers, la technologie de groupe accroît les séries apparentes, ce qui diminue les frais d'outillage, les temps de réglage, les en-cours et les manutentions. Les machines sont groupées en îlots de fabrication au sein desquels il est possible d'usiner entièrement les pièces d'une même famille (figure-1-4) [2]. La gestion de production, qui s'intéresse avant tout à l'ordonnancement des tâches et aux flux de matière, se trouve ainsi décentralisée sur chaque îlot et donc plus souple et plus aisée à réaliser.

-Au service commercial, le calcul des temps et devis est réduit de façon substantielle par une exploitation systématique de tous les données déjà accumulées dans l'entreprise. Ainsi il est possible de rechercher des produits similaires déjà fabriqués. Ceci permettra une estimation rapide du coût d'un produit nouveau (figure-1-5)[3].

### I-3/ IMPLANTATION DE LA TECHNOLOGIE DE GROUPE ET LIMITES DES SYSTEMES ACTUELS

Le concept de la technologie de groupe a été primitivement

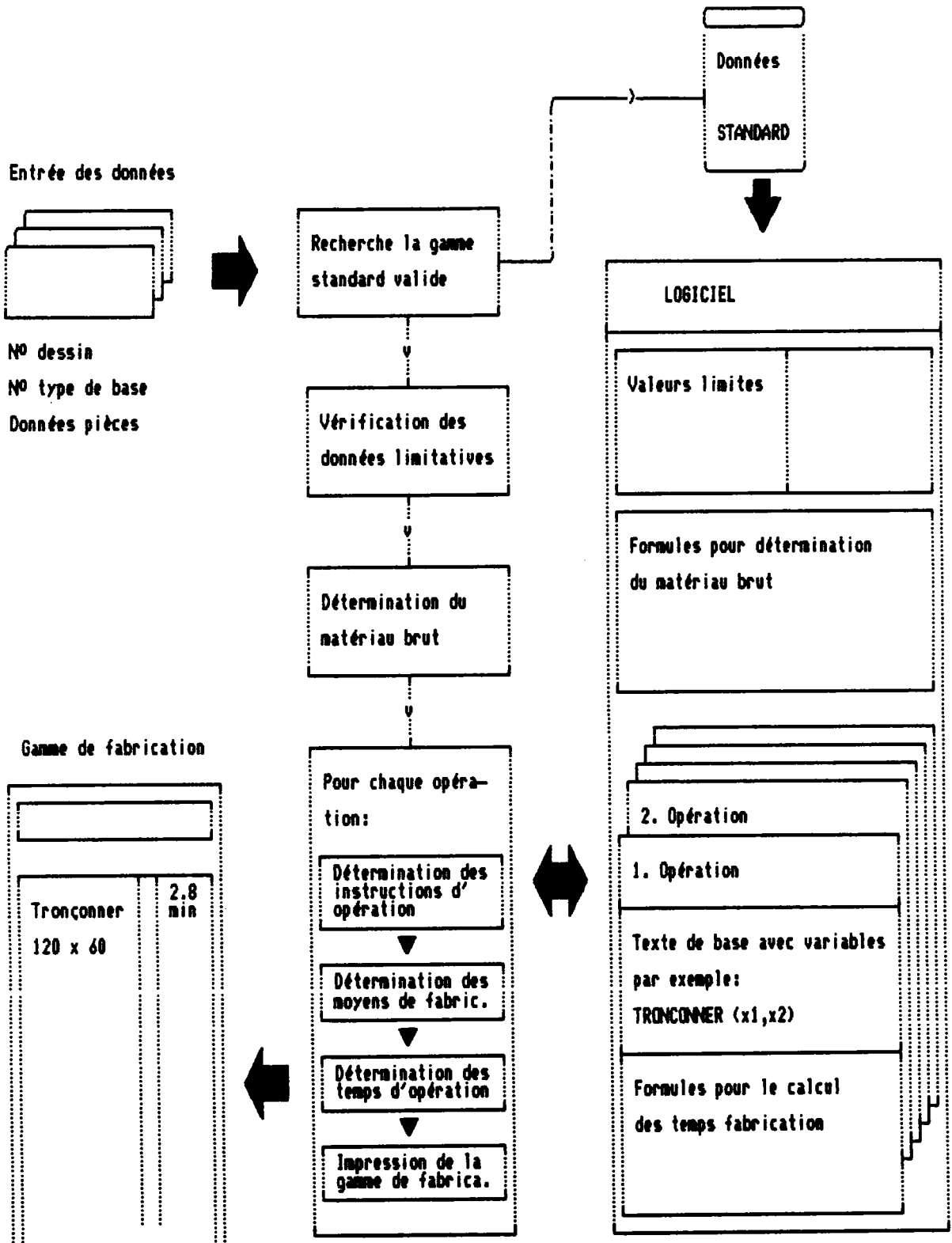


Figure-1-3: Exemple d'élaboration des gammes de fabrication [3].

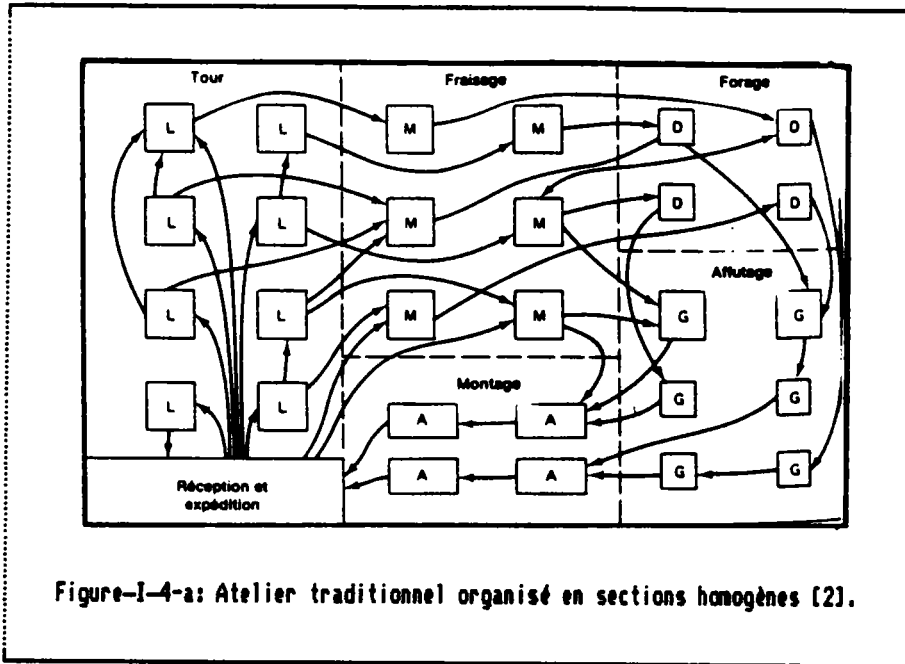


Figure-I-4-a: Atelier traditionnel organisé en sections homogènes [2].

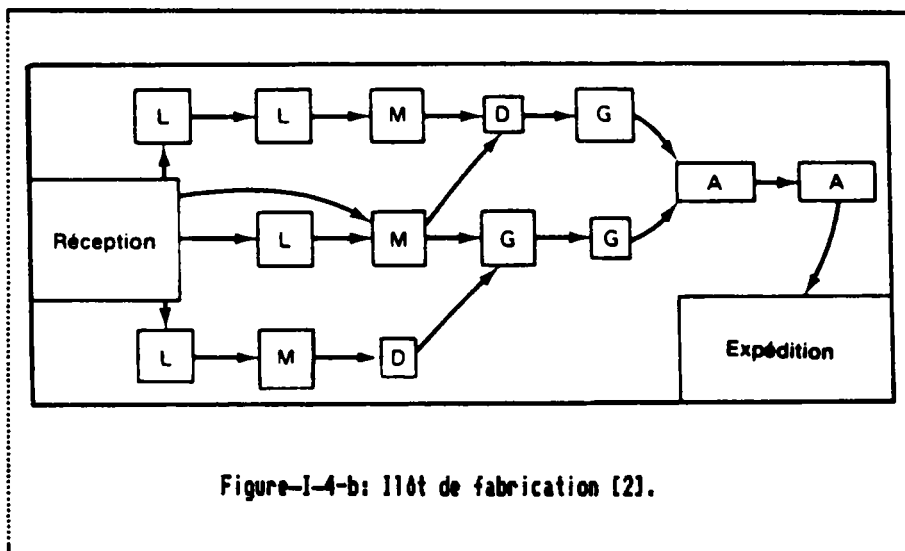


Figure-I-4-b: Ilôt de fabrication [2].



Figure-I-5-a:

Utilisation d'un système de classification des appareils et installations pour la détermination des caractéristiques d'un produit [3].

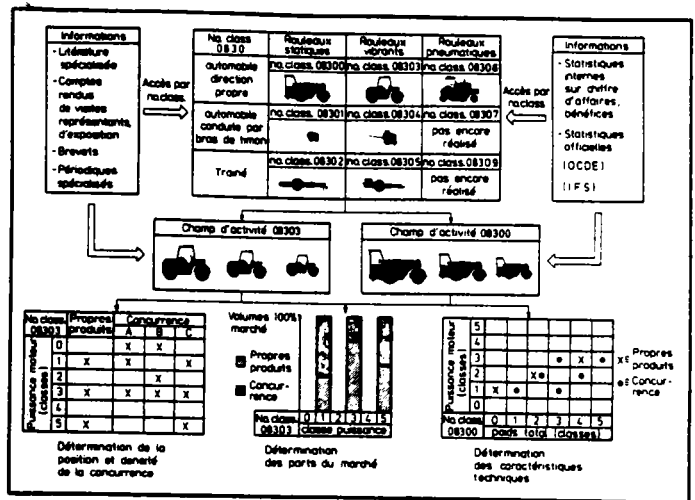
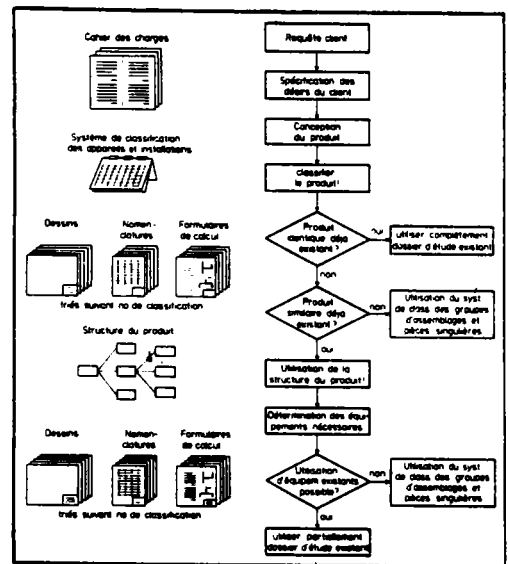


Figure-I-5-b:

Exemple de méthode pour l'élaboration systématique des offres [3].



développé par Mitrovanov [6]. A l'aide d'un système de classification, Mitrovanov regroupe les composants qui présentent des processus de fabrication similaires ou identiques. Pour chacun des groupes (des familles) il définit un processus standard de fabrication et des règles d'affectation des pièces à cette famille.

La description des pièces est faite sous la forme d'un code qui rassemble des informations morphodimensionnelles et technologiques. Pour une nouvelle pièce codée, le principe de la méthode est de rechercher la famille d'appartenance et d'éditer la gamme standard qui lui est associée. Il appartient à l'utilisateur (préparateur) de valider ou modifier éventuellement certains paramètres de cette gamme compte tenu des particularités de la nouvelle pièce.

Depuis le début des années 60, plusieurs systèmes de classification ont été développés pour définir des familles de produits.

Ces systèmes sont basés pour la plupart sur des algorithmes de tris à l'intérieur d'une banque de données [7]. Les résultats obtenus dépendent en grande partie de la clairvoyance de l'utilisateur, qui choisit les positions du code sur lesquelles s'effectuent les tris.

Aucun système de classification n'est universel. La classification utilisée au bureau d'études n'est, par ailleurs, pas toujours valable au bureau des méthodes (figure-I-6)[2]. En effet au bureau d'études, la forme ou la fonction de la pièce sont privilégiées par rapport au processus de fabrication, qui, lui intéresse pleinement le bureau des méthodes (figure-I-7)[2].

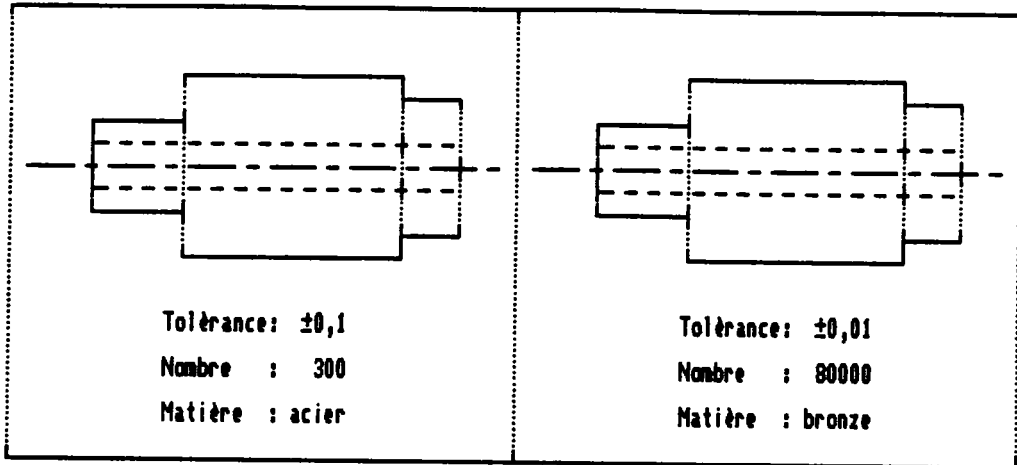


Figure-I-6: Pièces de formes et dimensions identiques mais de fabrication différente.

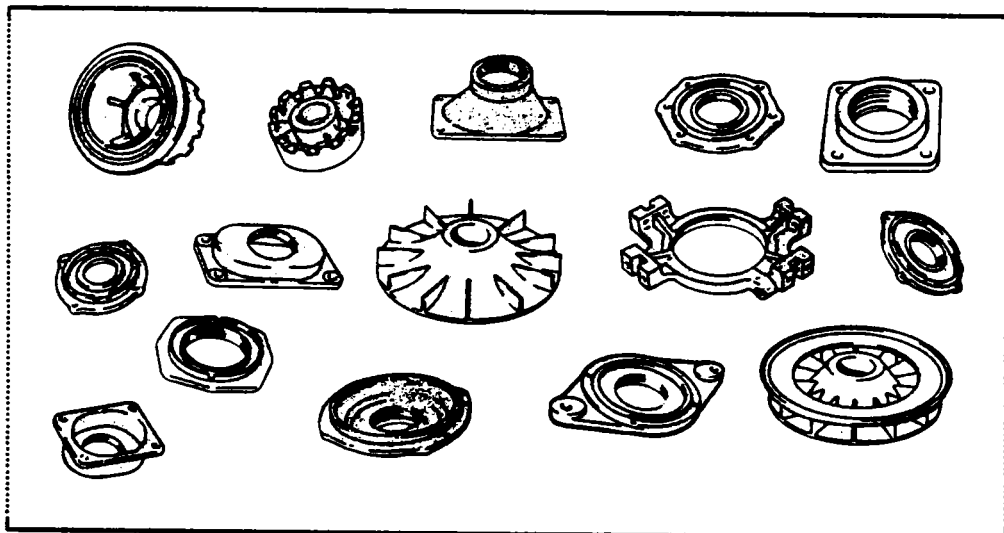


Figure-I-7: Pièces au processus de fabrication identique mais à conception différente [2].

Au bureau des méthodes, la classification porte sur les gammes de fabrication, celles-ci étant des suites ordonnées de postes de travail dont le nombre diffère d'une gamme à l'autre.

#### I-4/ OBJECTIFS ET PRESENTATION DE CETTE ETUDE

Le système de production est considéré comme un processus complexe tenant compte de très nombreuses variables de natures diverses, prenant des valeurs très différentes et interagissant de manières multiples.

L'observation de ce système sur un horizon déterminé oblige à prendre en compte un ensemble important de données, dont la structuration dépasse les possibilités humaines. Aussi le problème est-il de décomposer ce système complexe en sous-systèmes (familles) de composants identiques ou similaires.

L'utilisation actuelle des systèmes de classification en technologie de groupe n'est pas satisfaisante. Elle nécessite beaucoup de temps et de clairvoyance de l'utilisateur, sans pour autant garantir un résultat objectif.

En dehors de ces systèmes, l'Analyse de Données propose un grand nombre d'algorithmes, qui tels quels ne sont pas adaptés aux problèmes de production.

Cette étude présente une méthodologie générale d'implantation de la technologie de groupe basée sur la notion de ressemblance utilisée en Analyse de Données. Elle a conduit à la réalisation d'un système informatisé complet de détermination de groupement

technologique.

Le chapitre II présente une analyse des systèmes et techniques de Technologie de groupe actuellement disponibles en production. Après une revue des principales méthodes d'Analyse des Données destinée à mieux cerner leur champ d'application, le choix de celles qui seront utilisées dans la suite de l'étude est effectué.

Le Chapitre III propose une méthodologie, dans le but de permettre une utilisation rationnelle des méthodes d'Analyse des Données sélectionnées. Les différentes étapes de cette méthodologie sont les suivantes:

- collecte des données
- mise en forme des données
- choix de la mesure de proximité adaptée aux données
- représentation graphique
- classification automatique
- description des familles obtenues
- classement et validation.

Cette méthodologie a donné lieu à l'élaboration du logiciel CL2AUDIA (Classification et Classement Automatique de Données Industrielles Analytiques) qui est présenté au chapitre IV, où une application industrielle de ce logiciel à des données de production dans le domaine de l'Industrie Mécanique est

développée.

Cette étude s'insère dans un ensemble de travaux en technologie de groupe menés par le LAEI (Laboratoire d'Automatique et d'Electronique Industrielles) parmi lesquels ceux de J.Minot [7] qui ont montré l'intérêt de l'utilisation des méthodes d'Analyse de Données.

L'apport principal de ce travail est:

- l'analyse de données ordinales incomplètes, que l'on ne savait pas traiter jusqu'ici à notre connaissance. C'est le cas, par exemple, des gammes de fabrication,

- l'analyse simultanée sur plusieurs ensembles de données,

- l'analyse fiable de tous les types de codes même redondants, par l'introduction de la notion de transcodage,

- une classification fidèle, grâce à l'identification sûre du noyau des classes obtenues,

- une représentation plus efficace de chaque famille par un spectre binaire optimisé, facilitant le classement ultérieur des nouvelles pièces.

Ainsi, les principales difficultés auxquelles se heurtaient les systèmes de classification traditionnels ont pu être aplanies.

Enfin, le logiciel mis au point dans cette étude est d'une utilisation facile, bien adaptée aux conditions de la production.

## CHAPITRE II

### ANALYSE DES SYSTEMES ET DES TECHNIQUES DE TECHNOLOGIE DE GROUPE

II-1/ ANALYSE DES SYSTEMES DE CLASSIFICATION EN  
TECHNOLOGIE DE GROUPE

La technologie de groupe est basée sur la définition d'un système de classification (formation des familles) et de classement (affectation d'un nouvel objet à sa famille d'appartenance).

La classification repose sur un code [annexe 1] et sur un système de gestion de celui-ci basé sur un critère de ressemblance, indispensable à la constitution des familles.

Les principaux critères de codification envisagés sont définis dans la figure-II-1.

Nature	Définition
Critère fonctionnel	Description de la pièce par son nom et sa fonction (écrou, vis, matrice,...)
Critère morphodimensionnel	Définition - de la forme principale de la pièce (enveloppe) - des dimensions (longueur, largeur, diamètre) - des formes secondaires (chanfrein, rainures,...)
Critère de fabrication	Définition des contraintes nécessaires à la transformation de la pièce brute en produit fini (matière, nombre de pièces, tolérance, forme du brute, type de trous,...)

Figure-II-1: Principaux critères de codification



Il est à noter que les deux derniers critères ne sont pas entièrement indépendants. Par exemple l'existence d'un chanfrein, dans une pièce cylindrique qui relève de la morphologie, nécessitera l'utilisation d'un tour, ce qui fait intervenir le processus de fabrication.

La figure-II-2 présente, pour les principaux codes utilisés à l'heure actuelle, le poids de chaque critère [6], [8 à 14].

systèmes	codes	type de pièce	nombre de digits	critères en %			applications			référence
				morpho-dimensionnel	fonctionnel	de fabrication	possibles		connues	
							au B.E.	au B.M.		
NAKK	TEKLA	quelconque	12	33	0	67	X	X	X	[8]
OPITZ	OPITZ	"	9	33	0	67	X	X	X	[9]
SULTZER	SULTZER	"	25	12	0	88	X	X	X	[6]
VUSTE	VUSTE	"	4	75	0	25	X	X	X	[6]
ZAF0	ZAF0	"	26	100	0	0	X			[6]
PEKLENIK	PEKLENIK	"	10	100	0	0	X		X	[10]
TCLASS	CETIM-PM6	révolution	11	55	0	45	X	X	X	[11]
M.D.S.I.	CODE	quelconque	8	100	0	0	X		X	[12]
ACAPS	KK3	révolution	21	76	10	14				[12]
BRITISCH BURN	PNC	révolution	6	100	0	0	X		X	[13]
VETTER	VETTER	"	4	75	0	25	X	X		[6]
MIGROUPE	MICLASS*	quelconque	12	67	0	33	X	X	X	[13]
MULTIGROUPE	MULTI-M	"	19	57	10	33	X	X	X	[14]

\* Développé par TNO, transformé ultérieurement en Multi-M par OIR

Figure-II-2: Principaux codes utilisés.

D'après ce tableau, il apparaît que le critère fonctionnel ne

joue qu'un rôle tout à fait mineur pour la codification de pièces, bien qu'il puisse être utilisé pour certaines applications ponctuelles [15 et 16].

On distingue deux types de systèmes industriels existants. Les uns recherchent des familles simplement par tri manuel ou automatique sur les différents digits du code (information d'une idée a priori). Les autres basés sur des mesures de ressemblance, reconnaissent des familles par optimisation d'un critère mathématique.

### II-1-A/ Systèmes basés sur des tris

La figure-II-3 rassemble les principaux systèmes basés sur des tris et les codes correspondants [6],[8 à 14].

Systèmes	Codes	Références.
NAKK	TEKLA	[8]
OPITZ	OPITZ	[9]
SULTZER	SULTZER	[6]
VUSTE	VUSTE	[6]
ZAFO	ZAFO	[6]
TCLASS	CETIM-PM6	[11]
M.D.S.I.	CODE	[12]
ACAPS	KK3	[12]
BRITISCH BURN	PNC	[13]
VETTER	VETTER	[6]
MIGROUPE	MICLASS*	[13]
MULTIGROUPE	MULTI-M	[14]
COPIC-BRISCH	COPIC	[6]

Figure-II-3: Principaux systèmes de classification et de classement avec leur code respectif.

Les principaux inconvénients présentés par ces systèmes sont les suivants:

D'une part ce sont des systèmes heuristiques. Les clés des tris effectués sont choisies empiriquement par l'utilisateur. Les résultats dépendent donc essentiellement de son expérience.

D'autre part ces systèmes ne permettent pas le couplage des codes et des séquences de fabrication.

### II-1-B/ Méthodes utilisant des mesures de ressemblance

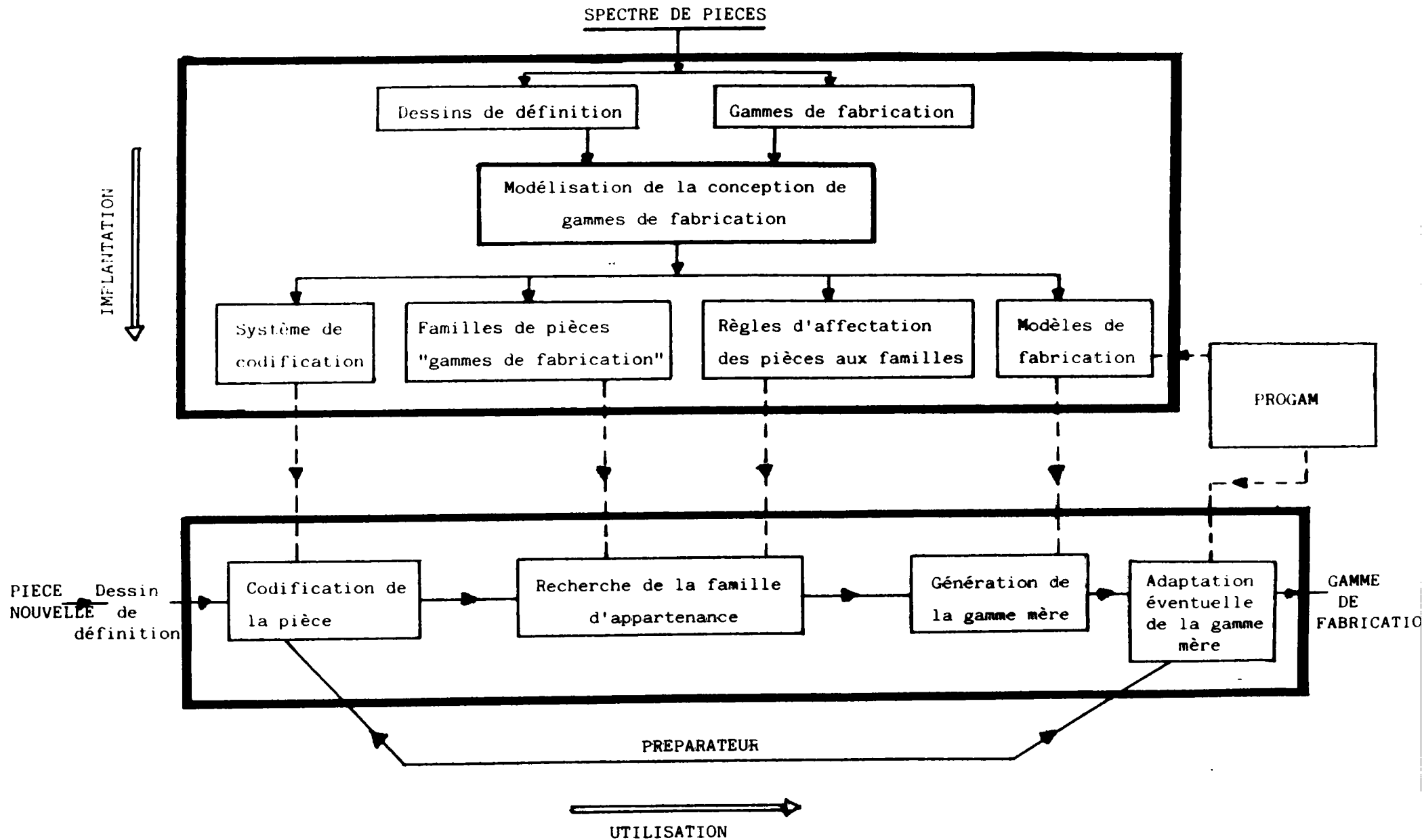
Les méthodes utilisant des critères mathématiques et la notion de ressemblance entre données codées ont été surtout développées à la fin des années 70.

Les principaux travaux sont:

1) l'étude menée par J. Minot [7] autour du logiciel PROGAM de l'ADEPA, qui fait appel à l'Analyse Factorielle des Correspondances pour la représentation graphique préliminaire des données à traiter, puis la méthode des Nuées Dynamiques est utilisée pour établir des familles.

Cette démarche a été appliquée industriellement indépendamment à des pièces codées selon le code Opitz et aux gammes de fabrication correspondantes en ne tenant pas compte de l'ordre des phases. La comparaison des résultats obtenus lors de ces deux traitements a conduit à la définition de fonctions de classement des nouvelles pièces.

L'architecture de ce système est schématisée à la figure ci-dessous.



Systeme développé par J. Minot [7]

2) la méthode de W.Eversheim et F.Lock [17 et 18], permettant une représentation automatique de données codées par réduction de l'espace multidimensionnel par une méthode multivariante statistique (figure-II-4). Cette méthode a été appliquée en technologie de groupe dans les domaines du dessin et de la planification.

Bien que ce travail repose sur des critères mathématiques (équivalence avec la méthode factorielle), la plus grande partie du traitement reste manuelle et l'interprétation des résultats devient difficile quand le nombre d'objets dépasse 3.

**Exemples**

a) Réduction multidimensionnelle de 3 objets:

	$\omega_1$	$\omega_2$	$\omega_3$
$\omega_1$	0	5	3
$\omega_2$	5	0	4
$\omega_3$	3	4	0

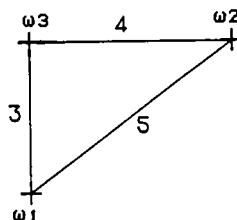


Tableau de ressemblances

Représentation graphique

b) Réduction multidimensionnelle de 4 objets:

	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$
$\omega_1$	0	6	7	5
$\omega_2$	6	0	3	6
$\omega_3$	7	3	0	3
$\omega_4$	5	6	3	0

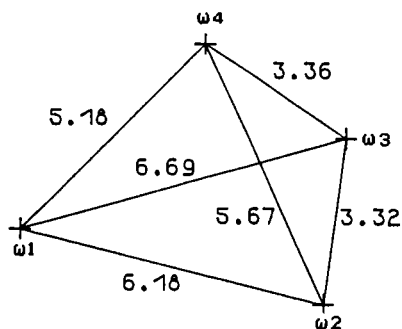
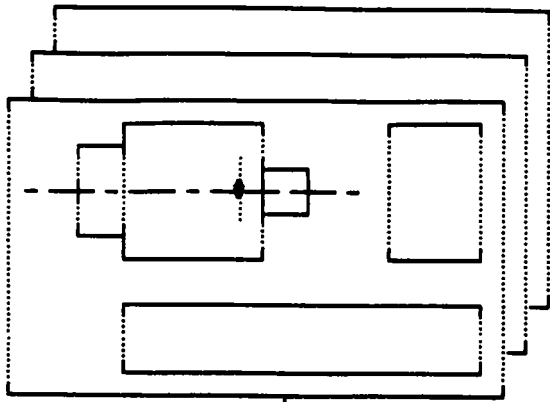


Tableau de ressemblances

Représentation graphique



définition des données  
technico-économiques de  
production.

$$\begin{array}{c}
 v \\
 \begin{array}{|c|} \hline \begin{array}{cccc} v_1 & \dots & v_j & \dots & v_p \\ \hline \omega_1 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ \omega_i & 1 & \dots & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots & & \vdots \\ \omega_n & 1 & \dots & 0 & \dots & 0 \end{array} \\ \hline \end{array}
 \end{array}$$

description mathématique  
des données techniques et  
économiques.

$$\begin{array}{c}
 v \\
 \begin{array}{|c|} \hline \begin{array}{cccc} \omega_1 & \dots & \omega_i & \dots & \omega_n \\ \hline \omega_1 & 0 & d_{1i} & \dots & d_{1n} \\ \vdots & \vdots & \vdots & & \vdots \\ \omega_i & d_{1i} & 0 & \dots & d_{in} \\ \vdots & \vdots & \vdots & & \vdots \\ \omega_n & d_{1n} & d_{in} & \dots & 0 \end{array} \\ \hline \end{array}
 \end{array}$$

définition et calcul de  
la ressemblance entre  
produits.

$$\begin{array}{c}
 v \\
 \begin{array}{|c|} \hline \begin{array}{ccccccc} \omega_i & & & & \omega_{15} & & \\ \omega_n & & \omega_2 & \omega_3 & \omega_9 & & \\ \omega_j & & \omega_1 & \omega_6 & \omega_8 & & \\ & & \omega_4 & \omega_{16} & & & \\ & & \omega_{55} & \omega_{22} & & & \\ & & \omega_5 & \omega_{17} & & & \\ & & \omega_{49} & & & & \end{array} \\ \hline \end{array}
 \end{array}$$

réduction de l'espace  
multi-dimensionnel (de  
dimension n) par une  
transformation des valeurs  
des mesures de ressemblance  
pour faire une représentation  
approchée et plane tout en  
minimisant la perte  
d'information.

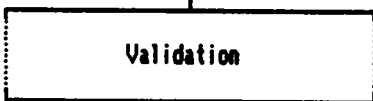


Figure-II-4: Système de W.Eversheim et F.Lock [17].

Ces deux exemples simples montrent que plus il y a d'objets, plus la représentation perd en exactitude.

3) les méthodes automatiques d'analyse par amas de Freist et Granow[19,20] et Kälberer [21,22] utilisant respectivement une classification hiérarchique et l'algorithme des K-means [annexe 2] appartenant aux méthodes des centroïdes [23].

Ces méthodes sont schématisées à la figure-II-5.

Ces méthodes ont été appliquées à l'utilisation de plans de travail standards dans le domaine de la fabrication.

Elles présentent un certain nombre d'inconvénients.

Tout d'abord, elles se limitent au traitement de données logiques exprimant la présence ou l'absence de phénomène.

Ensuite, la méthode de Kälberer fait appel à la méthode des K-means utilisant exclusivement la distance euclidienne, ce qui élimine d'entrée toute possibilité de choix d'un autre type de mesure de ressemblance.

Enfin, la méthode de Freist et Granow utilise une classification hiérarchique permettant de traiter uniquement des données peu nombreuses dans des temps raisonnables et avec des conditions d'interprétation satisfaisantes, celle-ci étant généralement effectuée à partir des sorties graphiques.

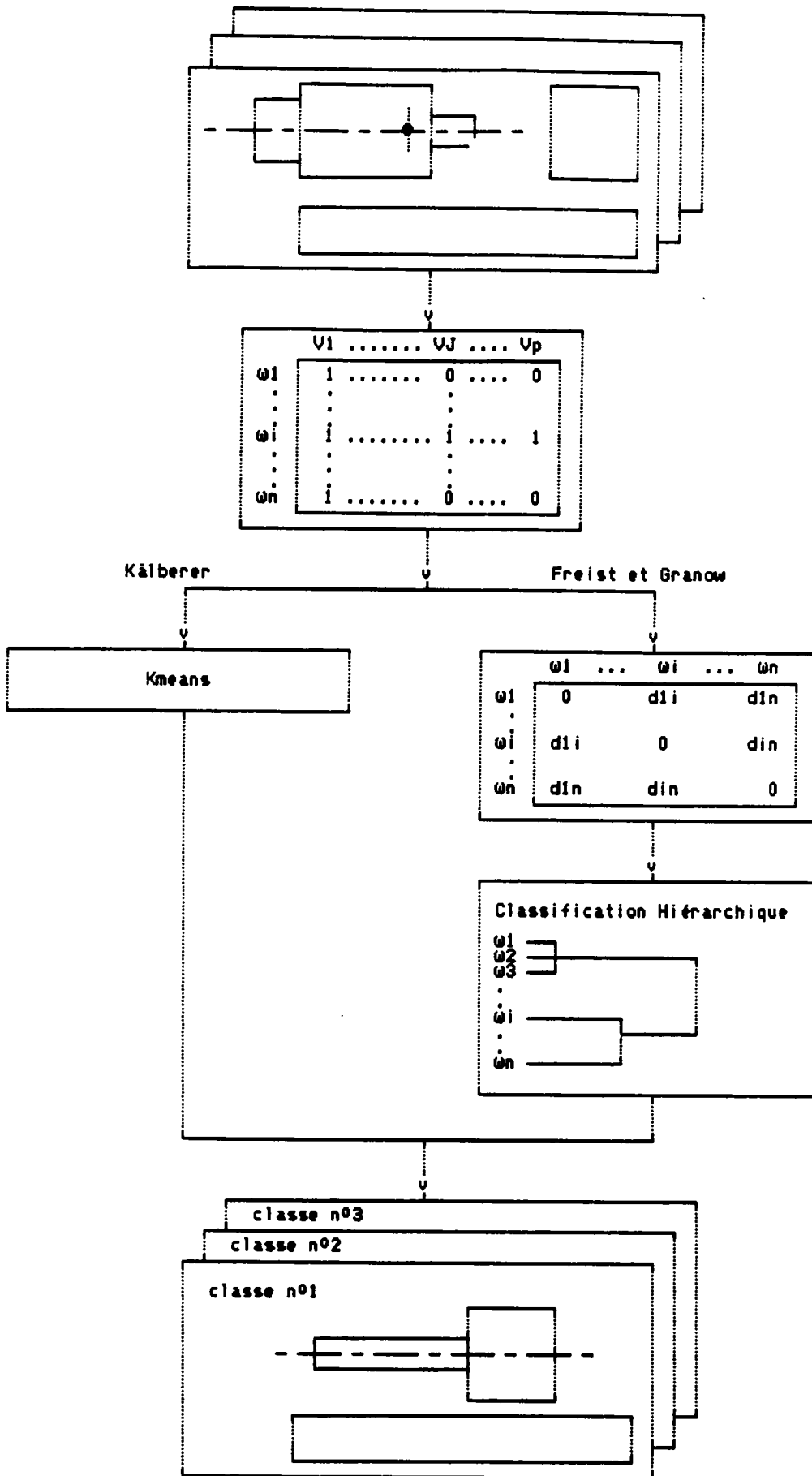


Figure-II-5: Classification par amas [19 à 22].



4) Le système semi-automatique d'analyse multi-objective par amas de C.Han et I.Ham [24] qui se ramène en fait à un pseudo-tri auquel s'ajoute la minimisation lexicographique d'une fonction vectorielle entière (figure-II-6).

$$U = \sum_{q=1}^m \sum_{p=1}^n ( Y_{pq} \cdot dpqk , k \in z )$$

avec

n = nombre de pièces

m = nombre de classes

dpqk = distance de la pièce p à la famille q pour les k digits

dpqk ≤ 0, pour tout p ≠ q, p=1,2,...,n

dpqk = 0, pour tout p = q, p=1,2,...,n et pour tout k ∈ z

Ypq = 1, si la pièce p appartient à la famille q

= 0, sinon

pour tout p variant de 1 à n  $\sum_{q=1}^m Y_{pq} = 1$

Z = code utilisé pour l'analyse

z = code rassemblant les digits qui vont servir au calcul (z c Z)

Figure-II-6: Analyse multi-objective par amas.

5) Le système de Peklenik [10,25,26], appliqué uniquement au bureau d'études. Cette méthode nécessite le choix a priori parmi les pièces à regrouper de pièces de référence auxquelles l'algorithme affecte les autres pièces. La comparaison entre pièces est basée sur la théorie des Fonctions Potentielles [59].

L'algorithme est présenté à la figure-II-7. Cet algorithme de classification n'est pas itératif, et les familles obtenues dépendent essentiellement du choix des pièces de référence, d'où une certaine rigidité.

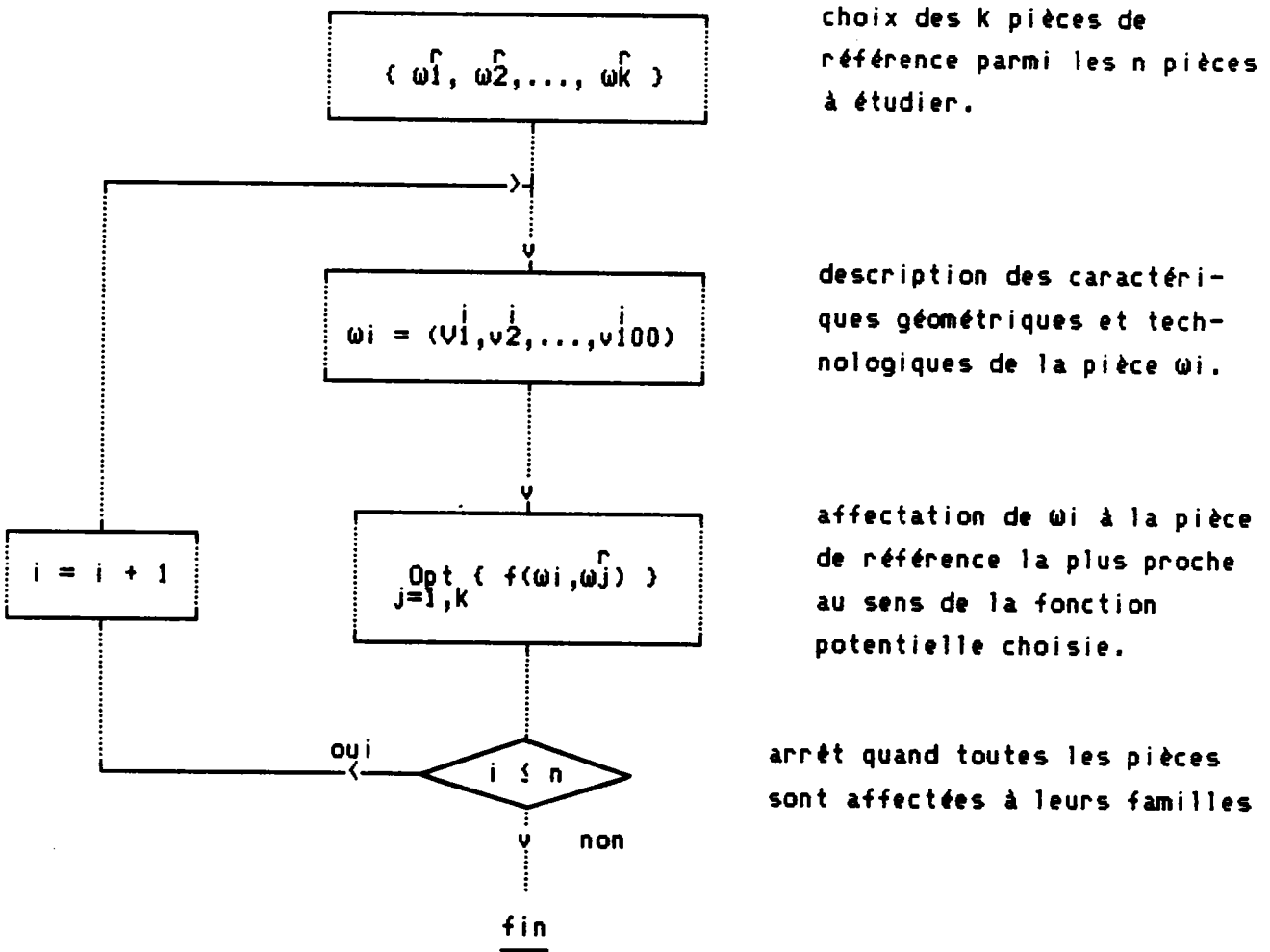


Figure-II-7: Algorithme de Peklenik [10,25,26].

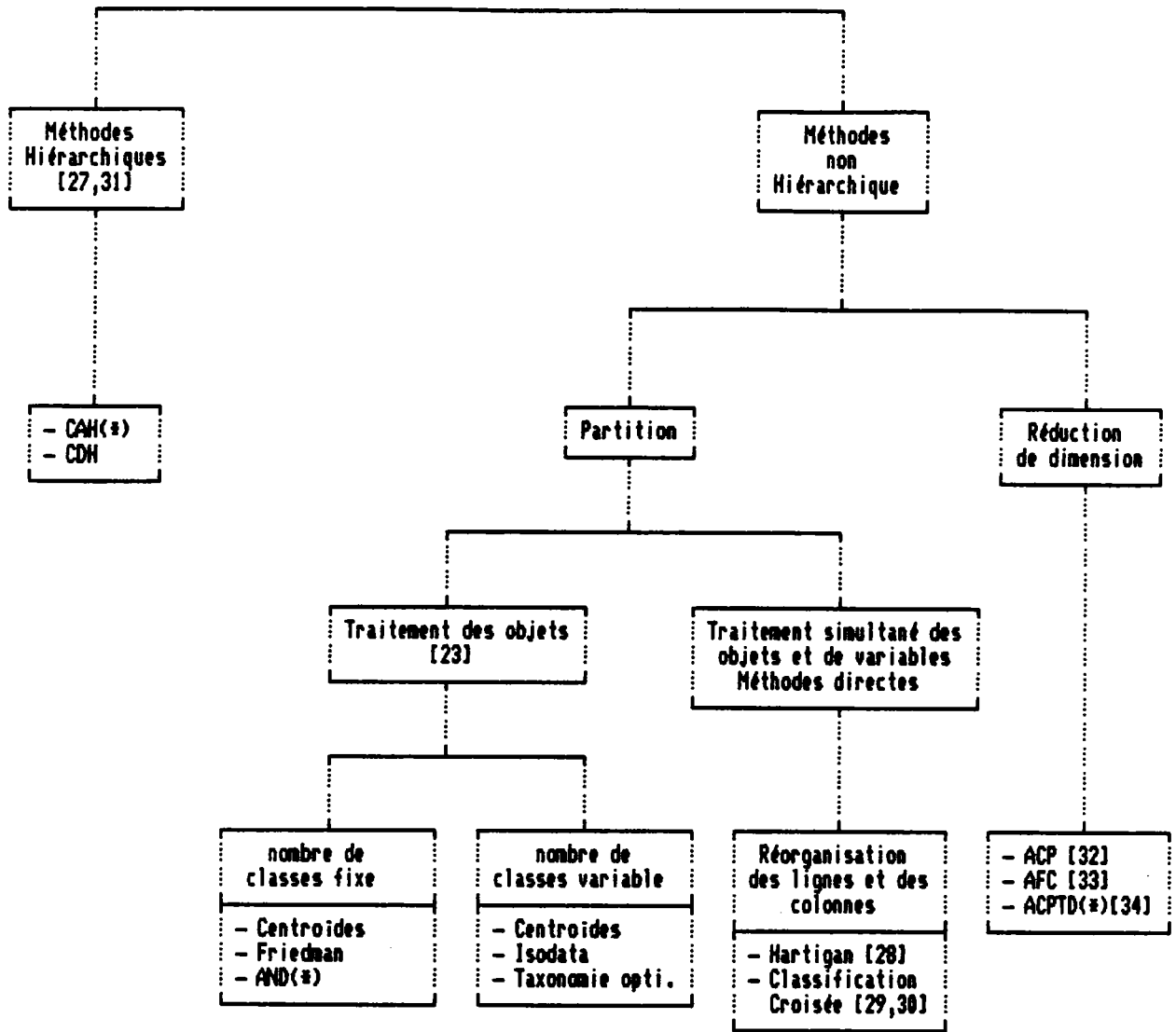
Les méthodes qui viennent d'être décrites travaillent sur des codes au sens large. Il peut s'agir de codes morphodimensionnels ou encore de séquences de fabrication. Dans ce dernier cas, c'est la présence ou absence de phase qui est codée, et il n'est pas possible de tenir compte de l'ordre des séquences.

D'autre part, elles ne permettent pas de conduire une analyse simultanée sur plusieurs ensembles de données (codes et gammes de fabrication). Or celle-ci est nécessaire pour pouvoir affecter par exemple une nouvelle pièce décrite par un code morphodimensionnel à sa famille d'appartenance obtenue à partir des séquences de fabrication. D'une façon plus générale, il est indispensable d'effectuer une telle analyse dès qu'il s'agit de réaliser une classification (formation de familles) sur un ensemble  $\Omega_2$  et d'opérer un classement (affectation) sur un ensemble  $\Omega_1$ .

## II-2/ ETUDE DES PRINCIPAUX ALGORITHMES D'ANALYSE DE DONNEES POUR LA TECHNOLOGIE DE GROUPE [23,27 à 35]

En Analyse des Données, de nombreux algorithmes de classification automatique ont été élaborés du fait du développement de l'informatique. Les principaux algorithmes sont rassemblés dans la figure-II-8. Les méthodes les plus intéressantes pour la technologie de groupe seront détaillées et sont marquées d'un astérisque dans la figure-II-8. La description de la plupart des autres méthodes a été effectuée dans une étude bibliographique préliminaire [35].

Principales Méthodes d'Analyse des Données



- CAH : Classification Ascendante Hiérarchique
- CDH : Classification Descendante Hiérarchique
- AND : Algorithme des Nuées Dynamiques
- ACP : Analyse en Composante Principale
- AFC : Analyse Factorielle des Correspondances
- ACPD: Analyse en Composante Principale sur un tableau de Distances

(\*) méthodes particulièrement adaptées en technologie de groupe

Figure-II-8: Principales méthodes d'Analyse des Données.

D'une façon générale, un système de technologie de groupe doit permettre (figure-II-9):

- de représenter les données pour en avoir une idée globale,
- d'obtenir une classification précise,
- éventuellement d'opérer des regroupements de familles.

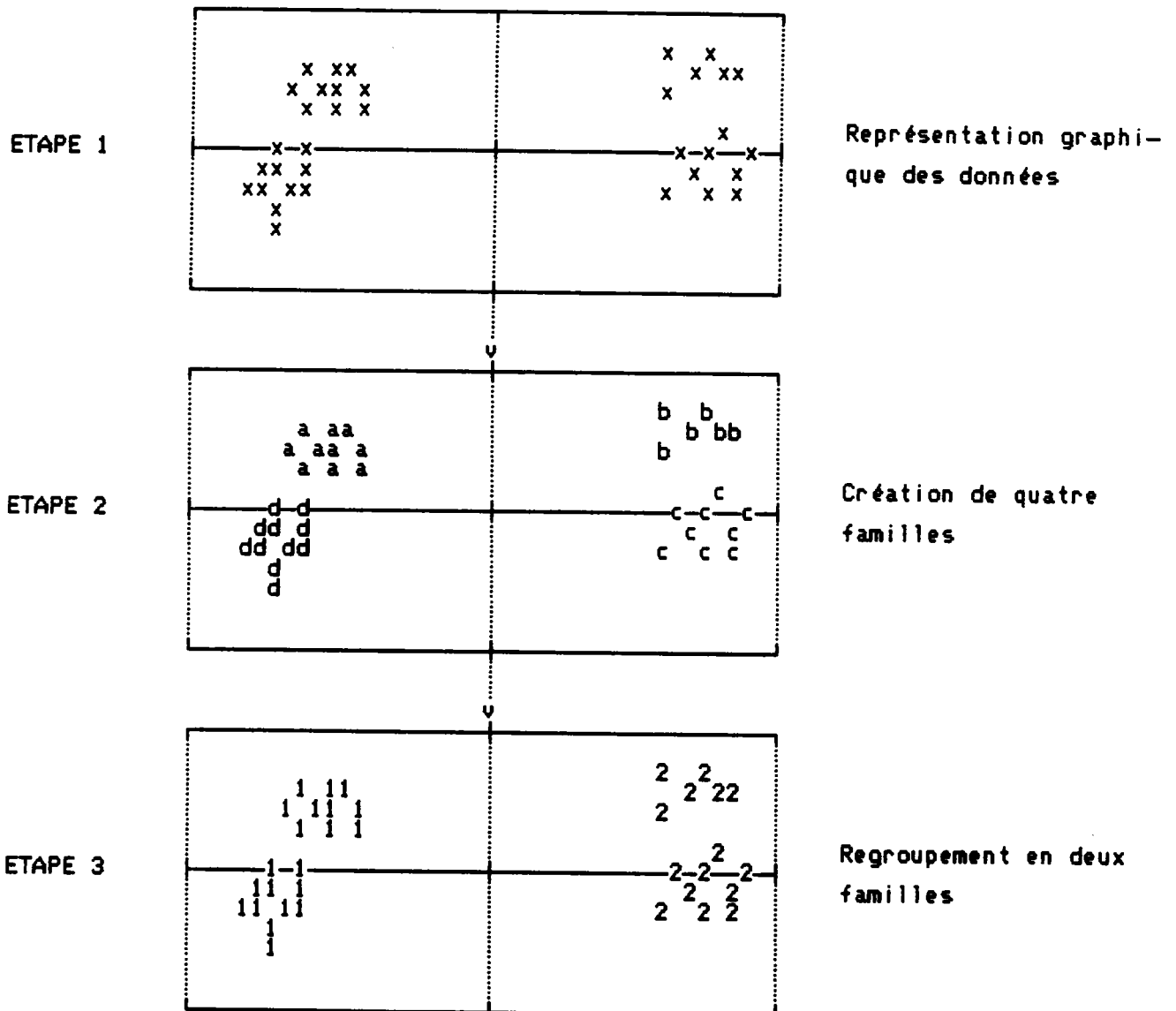


Figure-II-9: Principales étapes de la classification automatique en Technologie de Groupe.

Ensuite, il est nécessaire de vérifier la stabilité et la validité des familles obtenues.

II-2-A/ Représentation graphique des données[27,33,34,36]

La représentation graphique des données est obtenue par les méthodes d'ANALYSE FACTORIELLE qui permettent d'opérer une réduction de dimension. En effet, ces méthodes descriptives ont pour but de représenter un ensemble de points définis dans un espace à  $p$  dimensions dans un espace à  $q$  dimensions avec  $p > q$ , tout en conservant le plus possible d'information, en particulier les distances entre les points. La figure-II-10 montre un exemple de transformation d'un espace de trois dimensions en un espace de deux dimensions.

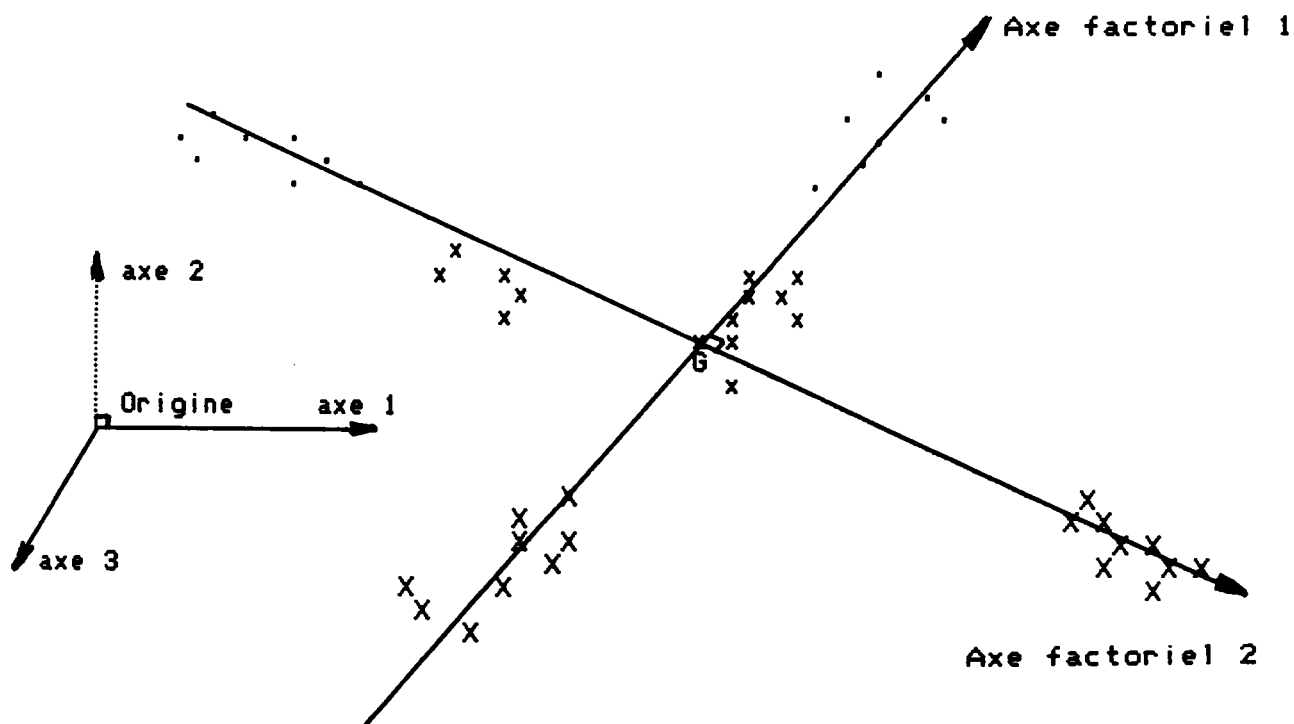


Figure-II-10: Représentation de données définies dans un espace tridimensionnel, dans un plan factoriel défini par deux axes factoriels

La qualité de l'analyse est évaluée à l'aide du taux d'inertie  $\tau$ , exprimé en pourcentage, relatif à chaque axe factoriel du nouvel espace. Ceux-ci sont classés par ordre d'inertie décroissant. Plus la somme des taux d'inertie des premiers axes factoriels est proche de 100%, meilleure est la représentation obtenue.

Les méthodes d'Analyse Factorielle existantes diffèrent essentiellement selon la nature des tableaux de données à expliquer et celle de la distance utilisée pour le traitement :

Méthode	Données	Distance
ACP	quantitatives	euclidienne
AFC	qualitatives	Khi-deux
ACPTD	quelconques	quelconque

Dans le cas de l'Analyse en Composante Principale ACP, le traitement est opéré sur un tableau de distances euclidiennes obtenu à partir de données quantitatives.

Au contraire, l'Analyse Factorielle des Correspondances AFC porte sur un tableau de distances du Khi-deux calculé à partir de données qualitatives.

Enfin, l'Analyse en Composante Principale sur un Tableau de Distances ACPTD permet de traiter un tableau obtenu à l'aide d'une distance quelconque.

Du fait de sa souplesse, c'est cette dernière méthode (ACPTD) qui a été choisie dans cette étude.

L'algorithme de l'ACPTD est présenté ci-dessous:

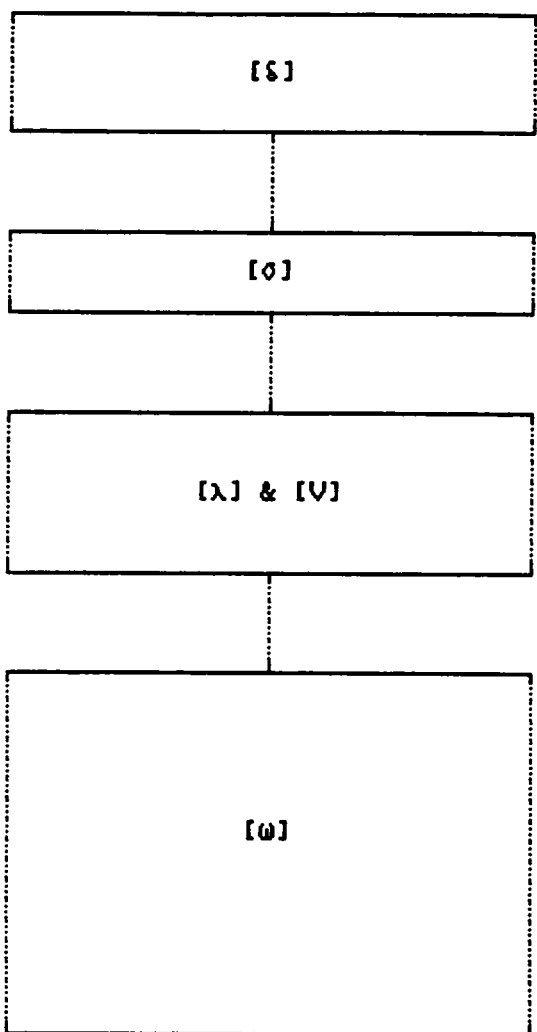


tableau d'une distance quelconque  $\delta$ :  $\delta_{ij}$  représente la distance entre l'objet  $\omega_i$  et l'objet  $\omega_j$

calcul du tableau  $[\sigma]$  de terme général  $\sigma_{ij}$

$$\sigma_{ij} = -\frac{1}{2}(\delta_{2ij} - \delta_{2i} - \delta_{2j} + \delta_{2..})$$

calcul des valeurs propres  $[\lambda]$  et des vecteurs propres  $[V]$ . Les vecteurs propres sont classés par ordre décroissant des valeurs propres

Les nouvelles coordonnées des de l'objet  $\omega_i$  s'expriment par la relation suivante:

$$\omega_i = \sqrt{\lambda_i} (v_{i1}, v_{i2}, \dots, v_{in})$$

Le taux d'explication  $\tau_i$  du  $i$ -ième axe factoriel est

$$\tau_i = \frac{\lambda^2_i}{\sum_{k=1}^p \lambda^2_k}$$



## II-2-B/ Obtention de familles (Classification)

D'une façon générale, les méthodes non hiérarchiques de recherche de partition ont pour but de regrouper les  $n$  objets d'un ensemble  $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_i, \dots, \omega_n\}$  en un nombre restreint de classes disjointes homogènes, de telle sorte que les objets d'une même classe soient fortement semblables.

Ces méthodes consistent à partir d'un certain nombre de représentants de classes, à leur affecter les autres objets et à recalculer à un stade plus ou moins avancé de l'algorithme les nouveaux représentants des classes.

Les méthodes avec nombre de classes variable (déterminé par l'algorithme) sont séduisantes. Cependant, il est très difficile d'estimer avec justesse certains paramètres qu'elles nécessitent: le seuil d'attraction au delà duquel deux familles seront réunies (Taxonomie Optimale [23]), le seuil de fusion et le seuil de division pour lesquels deux familles vont fusionner ou, au contraire, une famille va éclater (méthode Isodata [23]), ou encore les distances minimale et maximale entre classes (méthodes des Centroides [23]).

D'autre part, le choix de la distance ou de l'indice de similarité et du mode de représentation des classes est figé pour ces méthodes, ainsi que pour la méthode de Friedman [23] et la méthode des Centroides à nombre de classes fixe.

La méthode des Nuées Dynamiques [29,34,37], malgré l'inconvénient d'un nombre de classes fixe, déterminé a priori par l'utilisateur, présente un certain nombre d'avantages:

-elle peut traiter un tableau obtenu avec une distance quelconque (fonction d'affectation)

-une grande latitude existe dans le choix du mode de représentation de la classe (détermination du noyau par une fonction de représentation): les noyaux peuvent être des droites (figure-II-11-a), des groupes d'objets (figure-II-11-b) ou encore les barycentres des classes (figure-II-11-c), etc...

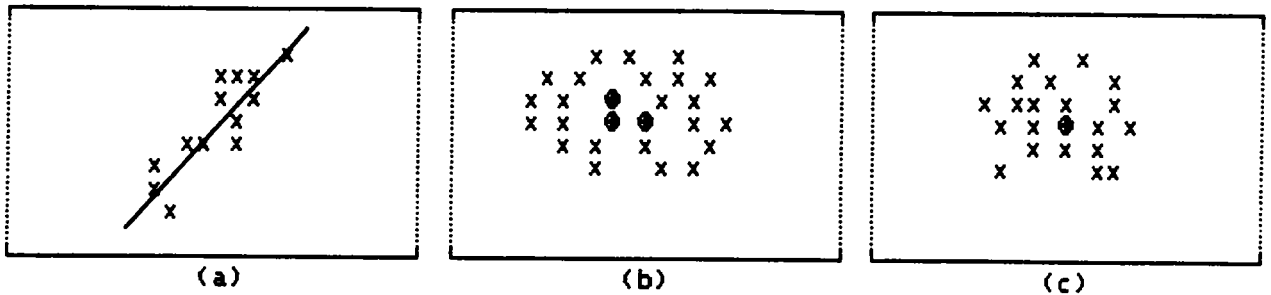


Figure-II-11: Exemples de choix du mode de représentation du noyau [34]:

- a- droite
- b- élément(s)
- c- barycentre

-le test d'arrêt, qui fait intervenir à la fois la fonction d'affectation et la fonction de représentation, est laissé au libre choix de l'utilisateur

-elle est séquentielle et ne nécessite pas la mémorisation des données.

Pour ces raisons, l'Algorithme des Nuées Dynamiques à été retenu pour effectuer la classification dans cette étude.

L'algorithme des Nuées Dynamiques est reporté à l'annexe 6.

II-2-C/ Regroupement de familles [23,27,31,38 à 41]

Si le nombre de classe trouvé par la méthode des Nuées Dynamiques est trop important pour être utilisable industriellement, il peut être utile de regrouper certaines classes entre elles.

Le regroupement peut être obtenu par les méthodes hiérarchiques. La classification peut être opérée soit d'une manière descendante soit d'une manière ascendante.

La Classification Descendante Hiérarchique CDH procède par dichotomies successives de l'ensemble de départ. L'emploi de cette méthode nécessite que l'utilisateur ait une idée juste et bien précise des critères selon lesquels ces dichotomies sont effectuées [27], ce qui est très difficile.

En revanche, il est plus sûr d'utiliser une Classification Ascendante Hiérarchique CAH, en constituant à partir des objets de l'ensemble de départ  $\Omega$  muni d'une distance des partitions de moins en moins fines par agrégation des plus proches voisins, chaque classe étant remplacée par un représentant.

Par exemple, la configuration représentée à la figure-II-12-a correspond à la hiérarchie tracée à la figure-II-12-b pour la distance euclidienne.

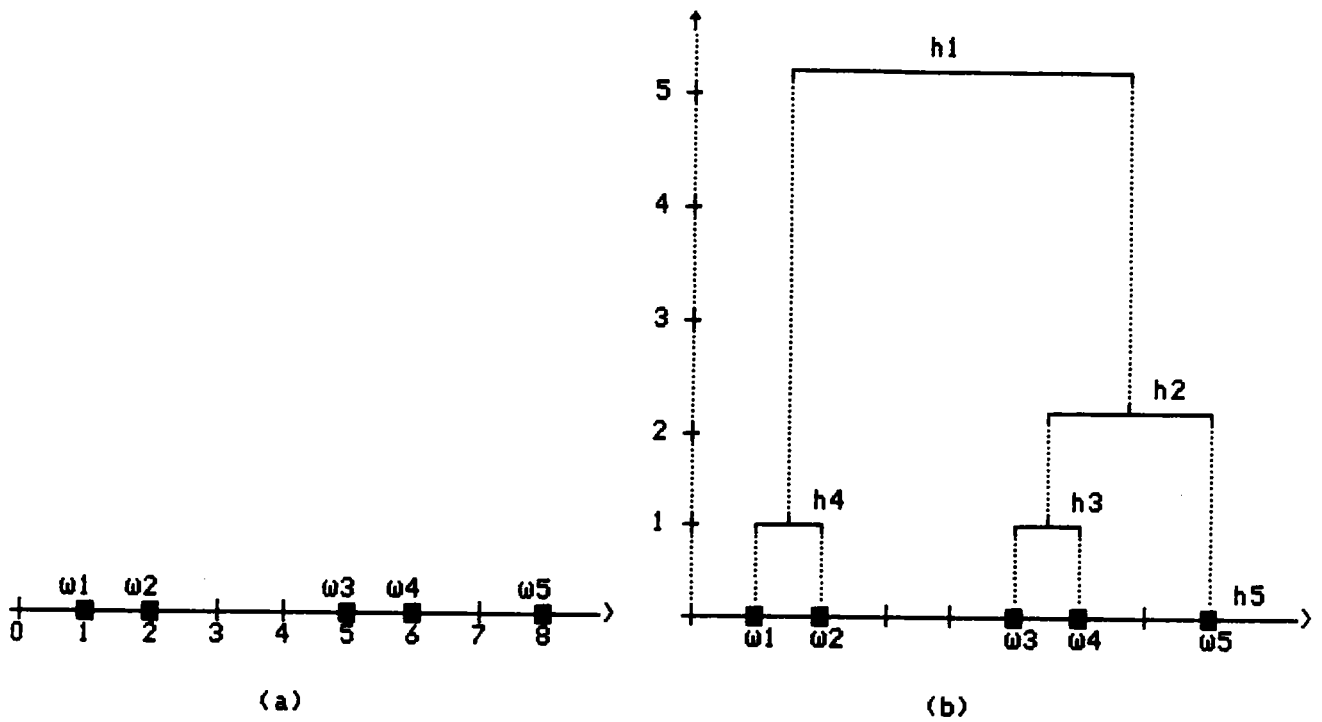
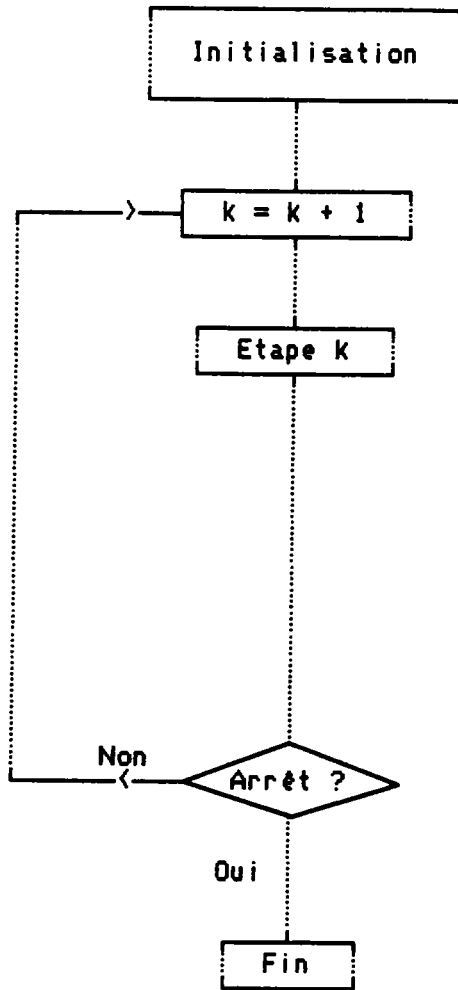


Figure-II-12: Exemple de CAH

La hiérarchie est composée de différents paliers, dont la hauteur exprime le degré d'agrégation du groupe. La partition formée par le palier  $h_1$ , le plus élevé, est la moins fine, puisqu'elle correspond à l'ensemble  $\Omega$  tout entier. La partition formée par le palier  $h_5$ , de hauteur nulle, est la plus fine: chaque élément  $\omega_i$  constitue une classe.

Cette méthode est très dépendante du critère utilisé dans le choix du représentant de classe (barycentre, l'un des éléments de la classe,...). D'autre part, la partition à l'étape précédente n'est jamais remise en cause. Enfin le temps de calcul est important.

Malgré ces inconvénients, la Classification Ascendante Hiérarchique reste une bonne méthode qui se situe à cheval entre les méthodes descriptives et les méthodes de classification automatique (figure-II-13) [38].



Partition la plus fine

Calcul des distances entre toutes les paires de classes de la partition obtenue à l'étape précédente. Recherche des deux classes  $C_i$  et  $C_j$  les plus proches.

Création de la partition de niveau  $k$  par fusion des classes  $C_i$  et  $C_j$ .

Lorsque la partition la moins fine est obtenue.

Figure-II-13: Algorithme de Classification Hiérarchique Ascendante [23].

C'est donc la CAH qui doit être retenue pour le regroupement des familles

### II-3/ CONCLUSION

Devant la diversité du nombre des méthodes d'Analyse de Données, il n'est pas toujours facile pour l'utilisateur industriel de savoir laquelle est adaptée au traitement de ses données, compte tenu des contraintes liées à la production.

Dans le cadre de cette étude, il s'agit de décomposer le système de production en familles moins complexes de composants identiques ou similaires.

Après analyse [35] des différentes méthodes d'Analyse de Données, les algorithmes suivants ont été retenus:

- Nuées Dynamiques AND.
- Analyse en Composantes Principales sur un Tableau de Distances ACPTD
- Classification Ascendante Hiérarchique CAH

D'une part, l'AND est un algorithme qui peut être utilisé avec toutes les mesures de ressemblance entre objets ayant au moins localement les propriétés d'une distance. Cette possibilité permet d'envisager de nombreuses variantes pour tenir compte des spécificités de chaque application.

D'autre part, c'est une méthode bien adaptée au traitement de grands tableaux de données.

En pratique, l'AND doit être couplé à des méthodes descriptives et interprétatives (ACPTD et CAH) afin de mieux appréhender les caractéristiques des données traitées et valider les résultats.

Une méthodologie de classification et de classement automatique a été mise au point dans cette étude et sera décrite dans le chapitre suivant.

**CHAPITRE III**

**METHODOLOGIE D'IMPLANTATION DE GROUPEMENTS TECHNOLOGIQUES**

Ce chapitre présente une méthodologie visant à permettre une utilisation rationnelle de l'Analyse de Données et qui s'articule en un certain nombre d'étapes essentielles pour mener à bien les opérations.

Très schématiquement, ces étapes sont les suivantes (figure-III-1):

- collecte et mise en forme (transcodage) des données
- choix des mesures de ressemblance et calcul des proximités
- représentation graphique des données
- classification automatique des données
- description des familles obtenues
- classement et validation

Les paragraphes suivants vont détailler les bases mathématiques fondamentales et l'analyse qui en est faite dans cette étude, sur lesquelles reposent ces différentes étapes.

### III-1/ COLLECTE DES DONNEES

Les données "objets" à analyser sont des plans, des gammes de fabrication, des outils de coupe, des devis ou des machines etc...

Un objet étant choisi, l'identification par ordinateur de celui-ci implique une codification à l'aide de "variables" (de caractéristiques) décrivant cet objet, c'est à dire une



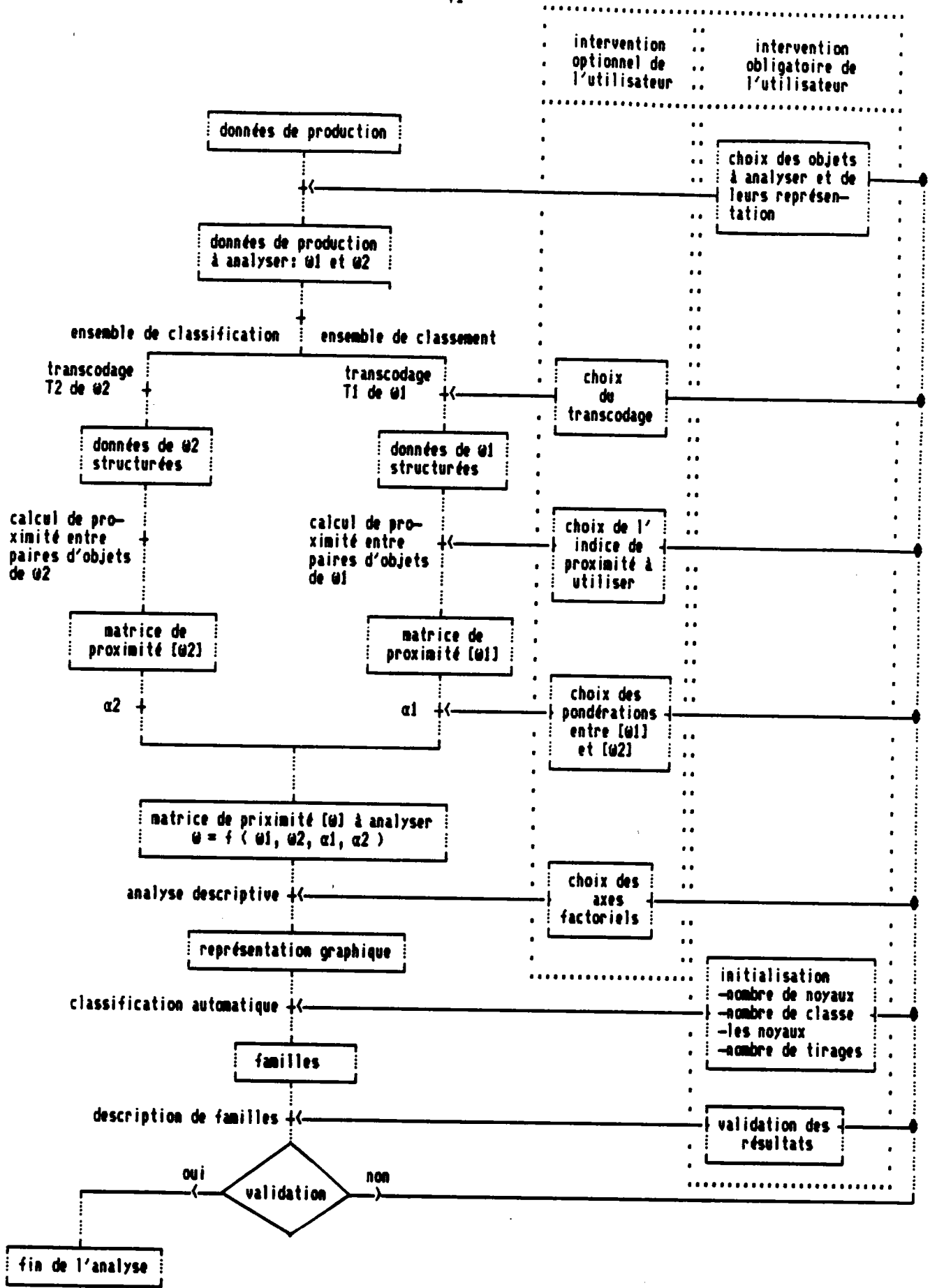


Figure-III-1: Schéma de la démarche de la méthodologie.

transformation des entités physiques ou économiques en grandeurs mathématiques.

### III-1-A/ Choix des objets à analyser

Si la population, c'est à dire l'ensemble d'objets, considérée est petite, l'utilisateur n'a pas à effectuer de sélection parmi les objets à analyser.

Cependant, dans la plupart des cas, les données technico-économiques de production sont trop nombreuses pour être simultanément analysées. Un échantillon représentatif de l'ensemble de la population doit être prélevé.

Un bon choix de l'échantillon d'étude est très important car il est à la base de l'analyse.

De plus, il faut choisir un ou plusieurs échantillons-tests qui permettront d'évaluer la qualité des résultats obtenus.

### III-1-B/ Choix des variables décrivant les objets [42]

Le choix des objets étant effectué, il faut définir leurs caractéristiques.

Ainsi, lors de l'analyse d'une population, on distingue deux ensembles  $\Omega$  et  $\emptyset$ .

$\Omega$  est l'ensemble des objets à analyser :

$\Omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_i, \dots, \omega_n)$  avec  $\text{card}(\Omega) = n$

$\omega_i$  ( $i=1, \dots, n$ ) représente par exemple le plan  $P_i$ , la gamme de fabrication  $G_i$ , l'outil de coupe  $O_i$  ou le devis  $D_i$ , etc...

$\Phi$  est l'ensemble des variables qui décrivent ces objets:

$\Phi = (\forall_1, \forall_2, \forall_3, \dots, \forall_j, \dots, \forall_p)$  avec  $\text{card}(\Phi) = p$

$\forall_j$  ( $j=1, \dots, p$ ) représente la variable  $j$  décrivant les  $n$  objets.  $\forall_j$  peut être par exemple la  $j$ -ième position du code d'un plan, le  $j$ -ième poste de travail d'un parc machine... On est alors amené à considérer plusieurs types de variables.

#### a) Variables Quantitatives

Une variable quantitative est une variable définie dans un espace métrique ( $\forall_j \in \mathbb{R}$ ), muni des opérations mathématiques (addition, soustraction, multiplication et division), et caractérisé par l'existence de l'élément zéro, qui indique l'absence de phénomène, par exemple pour un outil de coupe, la durée de vie nulle, pour une série de pièces la quantité nulle,...

- Exemple de variables quantitatives:

La longueur, la largeur, la valeur d'un diamètre, le nombre de trous dans une pièce, le prix unitaire d'un objet, la durée de vie d'un outil ou d'une machine, etc...

#### b) Variables Qualitatives Nominales

Une variable qualitative nominale est définie par des

symboles appelés modalités, qui ont pour rôle d'identifier les différents états des variables (la plupart du temps ces symboles appartiennent à  $\mathbb{N}$ ).

$$\forall j = (\mu_{j1}, \mu_{j2}, \mu_{j3}, \dots, \mu_{jf}, \dots, \mu_{jk}) \text{ avec } \text{card}(\forall j) = k$$

où  $\mu_{jf}$  représente la  $f$ -ième modalité de la  $j$ -ième variable.

Le nombre de modalités peut changer d'une variable à l'autre.

-Exemple de variables qualitatives nominales: Fontion, forme, matière, etc...

### c) Variables Qualitatives Ordinales

Une variable qualitative ordinale est une variable qualitative nominale dont l'ensemble des modalités est muni de la relation d'ordre total. Les valeurs numériques de  $\mu_{j1}, \mu_{j2}, \mu_{j3}, \dots, \mu_{jk}$  associées au  $k$  modalités de la variable  $\forall j$  doivent être telles que:

$$\mu_{j1} \leq \mu_{j2} \leq \mu_{j3} \leq \dots \leq \mu_{jf} \leq \dots \leq \mu_{jk}$$

-Exemple de variables qualitatives ordinales: variables dont les modalités sont obtenues après segmentation des valeurs quantitatives en un certain nombre d'intervalles.

### d) Cas particulier des données de production

Les données de production peuvent être des gammes de fabrication, des dessins de définition, des équipements, des

outillages, des postes de travail, etc...

#### d-1) Gammes de fabrication

La fabrication d'une pièce nécessite l'utilisation d'un certain nombre de machines ou postes de travail d'un parc machines.

Une gamme de fabrication est définie comme une suite ordonnée des postes de travail utilisés (phases). Cet ordre n'est pas complet: en effet, toutes les pièces ne passent pas par le même nombre de machines et une pièce peut passer deux fois par la même machine.

Les séquences de fabrication se définissent comme des données qualitatives ordinales incomplètes pour lesquelles la notion de modalité a disparu, au profit de l'ordre des phases.

A notre connaissance, les méthodes existantes d'analyse des données ne se sont jamais encore intéressées au traitement de ce type de données.

#### d-2) Autres données de production

On utilise dans l'industrie trois types de codes: les monocodes, les codes hiérarchiques et les codes mixtes ou polycodes.

-Un monocode est un code dans lequel la signification d'une position est définie d'une manière précise et indépendante des autres positions. C'est le cas du code Peklenik [26].

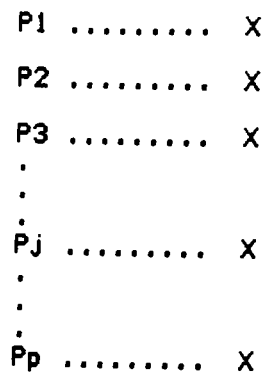


Schéma d'un monocode à p positions

-Un code hiérarchique est un code dans lequel la signification de la j-ième position est déterminée par la (j-1)-ième. C'est le cas par exemple du code Copic [6].

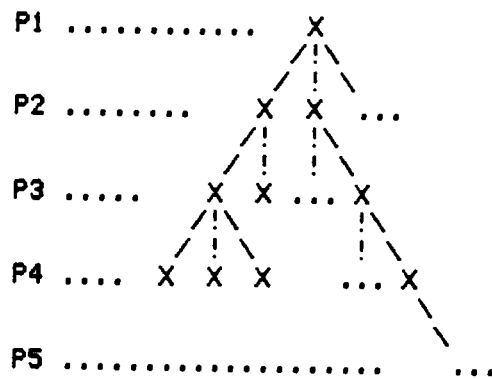


Schéma d'un code hiérarchique où les valeurs d'une position dépendent de la valeur de la position précédente.

-Un polycode est un ensemble de variables dont certaines sont autonomes (monocode) et d'autres sont dépendantes (code hiérarchique). C'est le cas par exemple du code Multi-M [43].

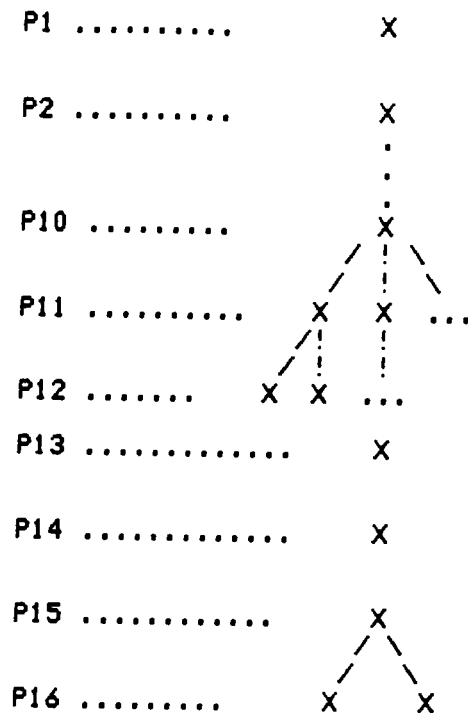


Schéma d'un polycode qui est la réunion de monocodes et de codes hiérarchiques.

D'un façon générale, il n'existe pas de code optimal. Chaque codage est orienté vers l'analyse de types de pièces ou de procédés de fabrication bien définis.

De plus, les codes utilisent en général un nombre surabondant de positions et de modalités pour éliminer les risques de lacunes d'information. Ceci explique le fait que l'on trouve des codes ayant jusqu'à une trentaine de positions. La redondance qui en découle biaise l'objectivité de l'analyse, qui en toute rigueur doit porter sur des variables (positions) totalement indépendantes, et donc sur un monocode.

D'autre part, compte tenu de l'hétérogénéité des variables les constituant, les codes industriels sont uniquement formés de variables qualitatives nominales. Ainsi les données quantitatives subissent une dégénérescence par découpage des échelles de mesure en plusieurs intervalles, ce qui entraîne une perte d'information.

En conclusion, il s'avère que, d'une part, certaines données industrielles de production sont des données ordinales incomplètes (par exemple les gammes de fabrication) que l'on ne savait pas traiter jusqu'ici, et d'autre part que l'analyse portant sur les autres types de données codées n'a de sens que si les variables intervenant sont indépendantes, alors que la plupart des codes ont une structure de polycode.

L'une des originalités de cette étude est de proposer:

-un transcodage des données codées pour rendre les variables aussi indépendantes que possible (transformation des polycodes en monocodes)

-dans le cas particulier des données ordinales incomplètes, un transcodage en données binaires.

### III-2/ TRANSCODAGE

Le transcodage a pour but de transformer l'ensemble  $\Phi$  des variables du code en un ensemble  $\mathcal{V}$  formé uniquement de variables indépendantes entre elles. Il est essentiel que les variables de l'ensemble  $\mathcal{V}$  aient une signification physique.

#### III-2-A/ Données ordinales incomplètes (cas des gammes de fabrication [44])

Le transcodage mis au point permet de transformer les données ordinales incomplètes en tableaux logiques composés uniquement de 0 (zéro) et de 1. Dans le cas de gammes de fabrication par exemple, soit  $M$  le parc machines:

$$\mathcal{M} = (M_1, M_2, M_3, \dots, M_j, \dots, M_q)$$



Une gamme  $G_i$  est caractérisée par une suite ordonnée de  $p_i$  postes de travail appelés phases [45]:

$$G_i = (\forall 1, \forall 2, \forall 3, \dots, \forall j, \dots, \forall p_i)$$

Le transcodage pour chaque gamme  $G_i$  comparée à une gamme  $G_i'$  consiste à construire un tableau  $T_i$ , qui est une matrice carrée de dimension égale au cardinal de la réunion des machines intervenant dans les deux gammes  $G_i$  et  $G_i'$ . Le terme général de ce tableau  $t_{jk}$  sera:

- $t_{jk}=0$  si la machine  $M_j$  n'apparaît pas dans la gamme  $G_i$ , ou si elle apparaît après la machine  $M_k$

- $t_{jk}=1$  si la machine  $M_j$  apparaît dans la gamme  $G_i$  avant la machine  $M_k$  ou en même temps qu'elle.

Le calcul de proximité entre gammes  $G_i$  et  $G_i'$  revient donc au calcul de proximité entre les tableaux  $T_i$  et  $T_i'$ .

L'exemple suivant montre le résultat du transcodage pour la gamme  $G_1=(A-B-C-D)$  et la gamme  $G_2=(D-C-E)$ .

Le fait que le nombre de postes de travail soit différent d'une gamme à l'autre, qui représentait a priori une difficulté pour la mesure de proximité entre séquences de fabrication, n'est plus contraignant après transcodage, puisque les tableaux à comparer deux à deux sont de même dimension, comme cela est montré dans l'exemple ci-dessous (figure-III-2).

	A	B	C	D	E
T1 =	1	1	1	1	0
	0	1	1	1	0
	0	0	1	1	0
	0	0	0	1	0
	0	0	0	0	0

	A	B	C	D	E
T2 =	0	0	0	0	0
	0	0	0	0	0
	0	0	1	0	1
	0	0	1	1	1
	0	0	0	0	1

FIGURE-III-2: Transcodage des gammes G1(A-B-C-D) et G2(D-C-E).

III-2-B/ Autres données [46, 47]

Le transcodage doit être adapté à chaque code utilisé; il n'y a donc pas d'algorithme général. Dans l'application industrielle du chapitre IV, un transcodage pour le code Multi-M est proposé.

D'autre part, l'ensemble  $\{V$  des variables qualitatives nominales transcodées étant obtenu, il y a lieu de les transformer en données binaires (Tableau Disjonctif Complet) afin de les traiter à l'aide des méthodes d'Analyse des Données. Le tableau disjonctif complet (figure-III-3) est un tableau dont le terme général  $\mu_{ij}$  prend la valeur 1 si l'objet  $\omega_i$  possède la modalité  $\mu_{jf}$  de la variable  $V_j$ , sinon  $\mu_{ij}$  s'annule.

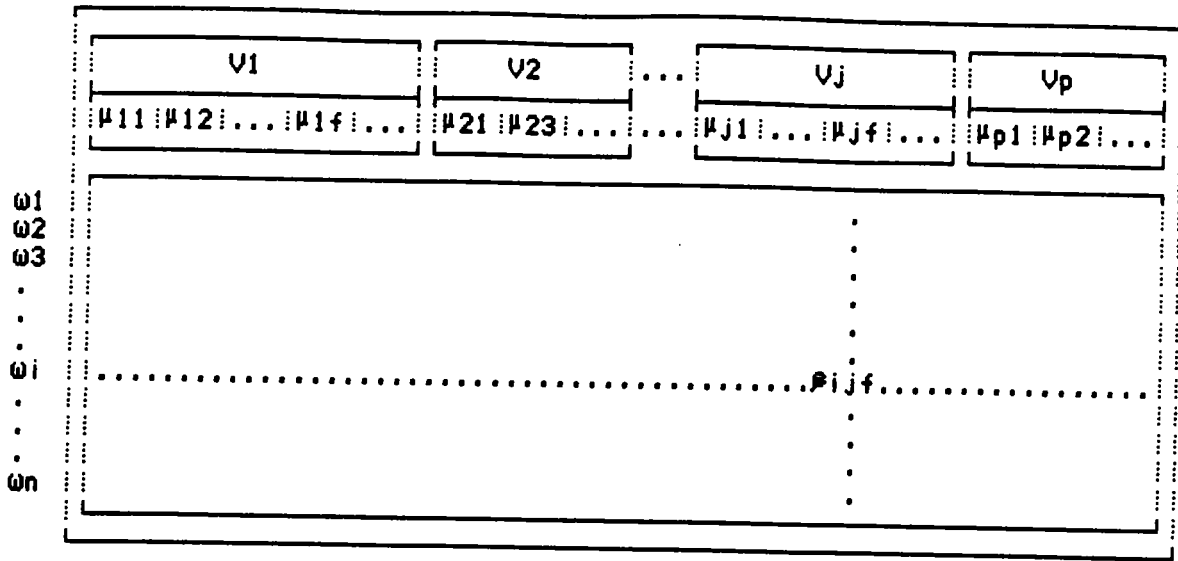


Figure-III-3: Tableau Disjonctif Complet [48].

Exemple

Soit  $\Omega$  l'ensemble de trois objets  $\Omega = (\omega_1, \omega_2, \omega_3)$  décrits par trois variables indépendantes  $V = (V_1, V_2, V_3)$ . Ces trois variables prennent un certain nombre de modalités:

$$V_1 = (\mu_{11}, \mu_{12}, \mu_{13})$$

$$V_2 = (\mu_{21}, \mu_{22})$$

$$V_3 = (\mu_{31}, \mu_{32})$$

Soient les codes associés aux trois objets de l'ensemble  $\Omega$

Objet	Code		
	V1	V2	V3
$\omega_1$	$\mu_{11}$	$\mu_{21}$	$\mu_{31}$
$\omega_2$	$\mu_{12}$	$\mu_{21}$	$\mu_{32}$
$\omega_3$	$\mu_{13}$	$\mu_{22}$	$\mu_{32}$

Le tableau disjonctif complet s'écrit:

	V1	V2	V3
	$\mu_{11} \quad \mu_{12} \quad \mu_{13}$	$\mu_{21} \quad \mu_{22}$	$\mu_{31} \quad \mu_{32}$
$\omega_1$	1 : 0 : 0	1 : 0	1 : 0
$\omega_2$	0 : 1 : 0	1 : 0	0 : 1
$\omega_3$	0 : 0 : 1	0 : 1	0 : 1

### III-3/ CALCUL DE PROXIMITE

#### III-3-A/ Généralités [34]

Savoir si un objet  $\omega_i$  est semblable à un autre  $\omega_j$  implique que l'on puisse mesurer le degré de similarité (de ressemblance) des objets entre eux. Ceci est effectué à l'aide d'une mesure de ressemblance encore appelée mesure de proximité  $r$  qui est définie comme une application de  $\Omega^2$  dans  $\mathbb{R}^+$ :

$$\forall (\omega_i, \omega_j) \in \Omega^2 : \quad \begin{array}{l} \Omega^2 \longrightarrow \mathbb{R}^+ \\ (\omega_i, \omega_j) \longrightarrow r(\omega_i, \omega_j) \end{array}$$

Les traitements par les méthodes d'analyse des données supposent que la mesure de proximité  $r$  réponde aux trois propriétés suivantes:

a) La positivité

Si  $\Omega$  est l'ensemble des objets, on vérifie :

$$\forall (\omega_i, \omega_j) \in \Omega^2, r(\omega_i, \omega_j) \in \mathbb{R}^+$$

b) La symétrie

$$\forall (\omega_i, \omega_j) \in \Omega^2, r(\omega_i, \omega_j) = r(\omega_j, \omega_i)$$

c) La triangularité au moins localement

$$\forall (\omega_i, \omega_j, \omega_k) \in \Omega^3, r(\omega_i, \omega_j) \leq r(\omega_i, \omega_k) + r(\omega_k, \omega_j)$$

On distingue trois types de mesure de proximité :

- indices de similarité  $\underline{s}$
- indices de dissimilarité  $\underline{\xi}$
- distances  $\underline{d}$

D'une façon générale, un indice de similarité vérifie les propriétés de positivité et de symétrie. De plus, il est normalisé :

$$\forall (\omega_i, \omega_j) \in \Omega^2 \text{ avec } \omega_i \neq \omega_j \\ s_{\max} = s(\omega_i, \omega_i) = s(\omega_j, \omega_j) > s(\omega_i, \omega_j)$$

En effet, un accroissement de l'indice de similarité  $s$  correspond à un accroissement de proximité entre objets.

Il existe dans la littérature un grand nombre d'indices de similarité [49, 50] sur variables binaires:

- |                               |                                  |
|-------------------------------|----------------------------------|
| 1) Jaccard (1908)             | $s_j = \frac{P}{P + N}$          |
| 2) Dice et Czekanowski (1945) | $s_d = \frac{P}{2P + N}$         |
| 3) Sokal et Sneath            | $s_s = \frac{P}{P + 2N}$         |
| 4) Kulczynski (1927)          | $s_k = \frac{P}{N}$              |
| 5) Sokal et Michener          | $s_m = \frac{P + A}{P + N + A}$  |
| 6) Russel et Rao              | $s_r = \frac{P}{P + N + A}$      |
| 7) Rogers et Tanimoto         | $s_t = \frac{P + A}{P + A + 2N}$ |

faisant intervenir les notions suivantes:

- Co-présence 'P': nombre de positions prenant simultanément la valeur 1 dans les deux codes comparés
- Co-absence 'A': nombre de positions prenant simultanément la valeur 0 dans les deux codes comparés
- Non-coïncidence 'N': autres positions

Supposons par exemple les deux codes binaires  $\omega_1$  et  $\omega_2$  suivants :

$$\begin{array}{l} \omega_1 = ( \boxed{0,0}, \boxed{1}, \boxed{1,1,1} ) \\ \omega_2 = ( \boxed{0,0}, \boxed{0}, \boxed{1,1,1} ) \end{array}$$

$\downarrow$        $\downarrow$        $\downarrow$   
 $A=2$     $N=1$     $P=3$

Leur co-présence vaut 3, leur co-absence vaut 2 et leur non-coïncidence vaut 1.

A ces indices de similarité correspondent des indices de dissimilarité  $S$  qui vérifient les propriétés de positivité et de symétrie décrites ci-dessus, mais pour lesquels la normalisation est différente :

$$\forall (\omega_i, \omega_j) \in \Omega^2 \text{ avec } \omega_i \neq \omega_j$$

$$S(\omega_i, \omega_i) = S(\omega_j, \omega_j) = 0 < S(\omega_i, \omega_j)$$

En effet, un accroissement de la valeur de l'indice de dissimilarité  $S$  correspond à une diminution de la proximité entre objets.

Une distance  $d$  est un indice de dissimilarité propre, c'est à dire vérifiant :

$$\forall (\omega_i, \omega_j) \in \Omega^2, S(\omega_i, \omega_j) = 0 \iff \omega_i = \omega_j$$

qui vérifie en plus de la relation de triangularité.

La distance euclidienne pour deux codes  $\omega_i$  et  $\omega_j$ , par exemple, peut être écrite en utilisant la notion de non-coïncidence  $N$  définie précédemment:

$$d^2(\omega_i, \omega_j) = \sum_{k=1}^p (\omega_i^k - \omega_j^k)^2 = N$$

D'un point de vue mathématique, il est plus intéressant d'utiliser une distance qu'un indice de similarité ou de dissimilarité, pour des questions de convergence.

Il est possible, d'une façon générale, et au prix d'une certaine distorsion, de transformer un indice de dissimilarité propre en distance [34].

Soit  $\delta$  un indice de dissimilarité propre. Si on pose:

$$d(\omega_i, \omega_j) = \delta(\omega_i, \omega_j) = 0 \quad \forall \omega_i \in \Omega$$

$$d(\omega_i, \omega_j) = \delta(\omega_i, \omega_j) + c, \quad \forall (\omega_i, \omega_j) \in \Omega^2 \text{ avec } \omega_i \neq \omega_j \\ \text{et } c \in \mathbb{R}^+$$

$c$ , constante réelle positive, doit être telle que l'inégalité triangulaire soit vérifiée:

$$\delta(\omega_i, \omega_j) + c \leq \delta(\omega_i, \omega_k) + c + \delta(\omega_k, \omega_j) + c$$



soit  $s(\omega_i, \omega_j) - s(\omega_i, \omega_k) - s(\omega_k, \omega_j) \leq c, \forall (\omega_i, \omega_j, \omega_k) \in \Omega^3$

Le fait d'ajouter une constante  $c$  à toutes les valeurs non nulles de l'indice de dissimilarité  $s$  introduit une discontinuité entre objets très semblables ( $d(\omega_i, \omega_j) \approx c$ ) et objets identiques ( $d(\omega_i, \omega_j) = 0$ ). La valeur prise par  $c$  doit être la plus faible possible.

Il apparaît donc que :

$$c = \text{Max}(s(\omega_i, \omega_j) - s(\omega_i, \omega_k) - s(\omega_k, \omega_j)), \forall (\omega_i, \omega_j, \omega_k) \in \Omega^3$$

est la valeur optimale qu'il convient d'ajouter à l'indice de dissimilarité  $s$ .

On vérifie facilement que  $\underline{s(\omega_i, \omega_j) + c}$  est une distance.

D'autres auteurs [51] montrent que pour un certain nombre d'indices de dissimilarité  $s, \sqrt{s}$  est une distance euclidienne (figure-III-4). La distorsion introduite par le passage de  $s$  à  $\sqrt{s}$  est alors continue.

Indice	définition de $S(x,y), x = y$	nature de S	nature de $\sqrt{S}$
Dice-Czekanowski	$\frac{N}{2P+N}$		euclidienne
Sokal-Sneath	$\frac{2N}{P+2N}$	distance	euclidienne
Jaccard	$\frac{N}{P+N}$	distance	euclidienne
Russel-Rao	$\frac{N+A}{P+N+A}$	distance	euclidienne
Roger-Tanimoto	$\frac{2N}{P+2N+A}$	distance	euclidienne

Figure-III-4: Nature de différents indices de dissimilarité et de leur racine carrée [51].

### III-3-B/ Application aux données de production

Afin d'établir une typologie pertinente des objets, pour chaque type de données de production il faut choisir une mesure de proximité qui reflète le mieux la réalité des ressemblances.

De ce point de vue il est nécessaire de faire la distinction entre le traitement des données ordinales incomplètes et celui des autres données de production.

a) Choix d'un indice de proximité entre données ordinales  
incomplètes (application aux gammes de fabrication)

Il est nécessaire de rappeler que dans le cas des gammes de fabrication le nombre de postes de travail est variable d'une gamme à l'autre. Le nombre de machines non utilisées qui correspond à la co-absence définie au paragraphe III-3-1 ne doit pas intervenir dans le calcul de la ressemblance. Seules les notions de co-présence et de non-coïncidence sont en effet significatives.

Les distances classiques [Annexe 3] telles que la distance du Khi-2 ( $\chi^2$ ) qui font intervenir la notion de co-absence ne sont donc pas adaptées.

D'autre part, l'indice de Kulczynski ( $s_k = P/N$ ) est d'un usage délicat puisqu'il tend vers l'infini pour deux objets identiques ( $N=0 \Rightarrow s_k \rightarrow \infty$ ). Les indices de similarité de Jaccard  $s_j$ , de Dice-Czekanowski  $s_d$  et de Sokal-Sneath  $s_s$  semblent plus adéquats.

La démarche utilisée dans cette étude pour trouver la meilleure mesure de ressemblance entre données ordinales incomplètes a consisté à :

-transformer les indices de similarité retenus en distances euclidiennes,

-conduire une analyse sur un échantillon test afin de pouvoir effectuer un choix comparatif entre les distances obtenues,

-définir un certain nombre de critères qualitatifs

industriels pour vérifier que la distance sélectionnée est bien adaptée aux données analysées.

a-1) Obtention des distances euclidiennes

Les indices de similarité  $s_j$ ,  $s_d$  et  $s_s$  ont été transformés en indices de dissimilarité  $\delta_j$ ,  $\delta_d$  et  $\delta_s$  puis en distances euclidiennes  $d_j$ ,  $d_d$  et  $d_s$  à l'aide de l'analyse de Fichet et le Calve (figure-III-5) [51].

	indice de similarité	indice de dissimilarité	distance
Dice-Czekanowski	$s_d = \frac{2P}{2P+N}$	$\delta_d = \frac{N}{2P+N}$	$d_d^2 = \frac{N}{2P+N}$
Sokal-Sneath	$s_s = \frac{P}{P+2N}$	$\delta_s = \frac{2N}{2N+P}$	$d_s^2 = \frac{2N}{2N+P}$
Jaccard	$s_j = \frac{P}{P+N}$	$\delta_j = \frac{N}{N+P}$	$d_j^2 = \frac{N}{N+P}$

Figure-III-5: Transformation des indices de similarité de Jaccard, Dice-Czekanowski et Sokal-Sneath en distances euclidiennes [51].

Ces distances ont leur valeur minimale égale à zéro pour une gamme comparée à elle-même ou à une autre identique ( $N=0$ ). Elles atteignent leur valeur maximale, égale à l'unité, pour deux gammes totalement différentes ( $P=0$ ).

a-2) Etude sur un échantillon test

Sur un échantillon de 13 gammes tests, correspondant à un parc machines comprenant 14 postes de travail, présentées dans la figure-III-6, les caractéristiques de ces distances ont été comparées.

Après transcodage et calcul des proximités entre gammes, une analyse en composante principale sur le tableau de distances a été effectuée (figure-III-7).

On s'aperçoit que le taux d'explication sur les premiers axes factoriels, défini au chapitre II, est plus élevé pour la distance de Dice-Czekanowski  $d_d$  que celles de Jaccard  $d_j$  et Sokal-Sneath  $d_s$  (figure-III-8).

Ceci signifie que la mesure  $d_d$  permet une analyse plus fine entre gammes que les deux autres mesures. Considérons par exemple les gammes de fabrication 5 à 8 de l'échantillon test, et comparons la gamme 5 à chacune des autres à l'aide des trois mesures  $d_d$ ,  $d_s$  et  $d_j$ :

Gammes	Comparées	$d_j$	$d_d$	$d_s$
5	U V W X	0.000	0.000	0.000
6	U V W	0.632	0.500	0.756
7	U V	0.837	0.734	0.907
8	U	0.949	0.905	0.973

ANALYSE DES CHARGES MACHINE

NUM. REF.	NOMBRE DE PIECES	TEMPS TOTAL (EN HEURES) DE			GAMMES DE FABRICATION							
		CHARGE	REGLAGE	TRAVAIL	POSTE-A-	POSTE-B-	POSTE-C-	POSTE-D-	POSTE-E-	POSTE-F-	POSTE-G-	POSTE-H-
1	1				POSTE-A-	POSTE-B-	POSTE-C-	POSTE-D-	POSTE-E-	POSTE-F-		
2	1				POSTE-A-	POSTE-B-	POSTE-C-	POSTE-D-	POSTE-F-	POSTE-E-		
3	1				POSTE-F-	POSTE-B-	POSTE-C-	POSTE-D-	POSTE-E-	POSTE-A-		
4	1				POSTE-B-	POSTE-C-	POSTE-E-	POSTE-D-	POSTE-F-			
5	1				POSTE-B-	POSTE-C-	POSTE-D-	POSTE-F-	POSTE-J-	POSTE-K-	POSTE-L-	
6	1				POSTE-U-	POSTE-U-	POSTE-W-	POSTE-X-				
7	1				POSTE-U-	POSTE-U-	POSTE-W-					
8	1				POSTE-U-	POSTE-U-						
9	1				POSTE-U-							
10	1				POSTE-I-	POSTE-J-	POSTE-K-	POSTE-L-				
11	1				POSTE-I-	POSTE-J-	POSTE-L-	POSTE-K-				
12	1				POSTE-L-	POSTE-J-	POSTE-K-	POSTE-I-				
13	1				POSTE-I-	POSTE-L-	POSTE-J-	POSTE-K-				
TOT.												

NUM. REF.	COD. POST. TRAVAIL	DESCRIPTION DES POSTES DE TRAVAIL	TOTAL		TEMPS TOTAL (EN HEURES) DE			
			OPERATION	PIECES	CHARGE	REGLAGE	TRAVAIL	
1	POSTE-A-	-		3	3			
2	POSTE-B-	-		5	5			
3	POSTE-C-	-		5	5			
4	POSTE-D-	-		5	5			
5	POSTE-E-	-		1	1			
6	POSTE-F-	-		8	8			
7	POSTE-J-	-		5	5			
8	POSTE-K-	-		5	5			
9	POSTE-L-	-		5	5			
10	POSTE-U-	-		4	4			
11	POSTE-U-	-		3	3			
12	POSTE-W-	-		1	1			
13	POSTE-X-	-		2	2			
14	POSTE-I-	-		4	4			
TOTAL				56	56			

Figure-III-6: Echantillon de gammes-tests.

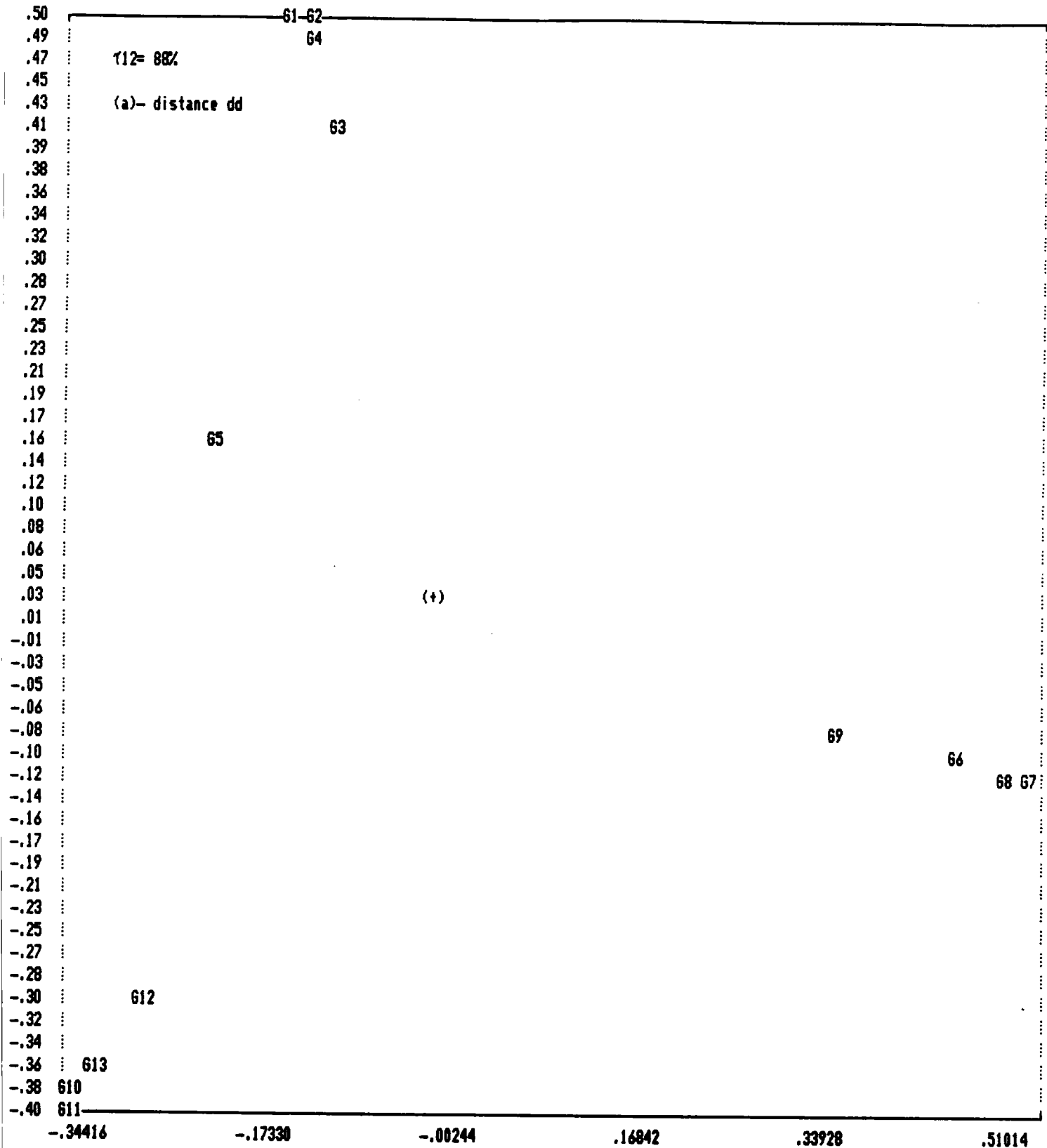
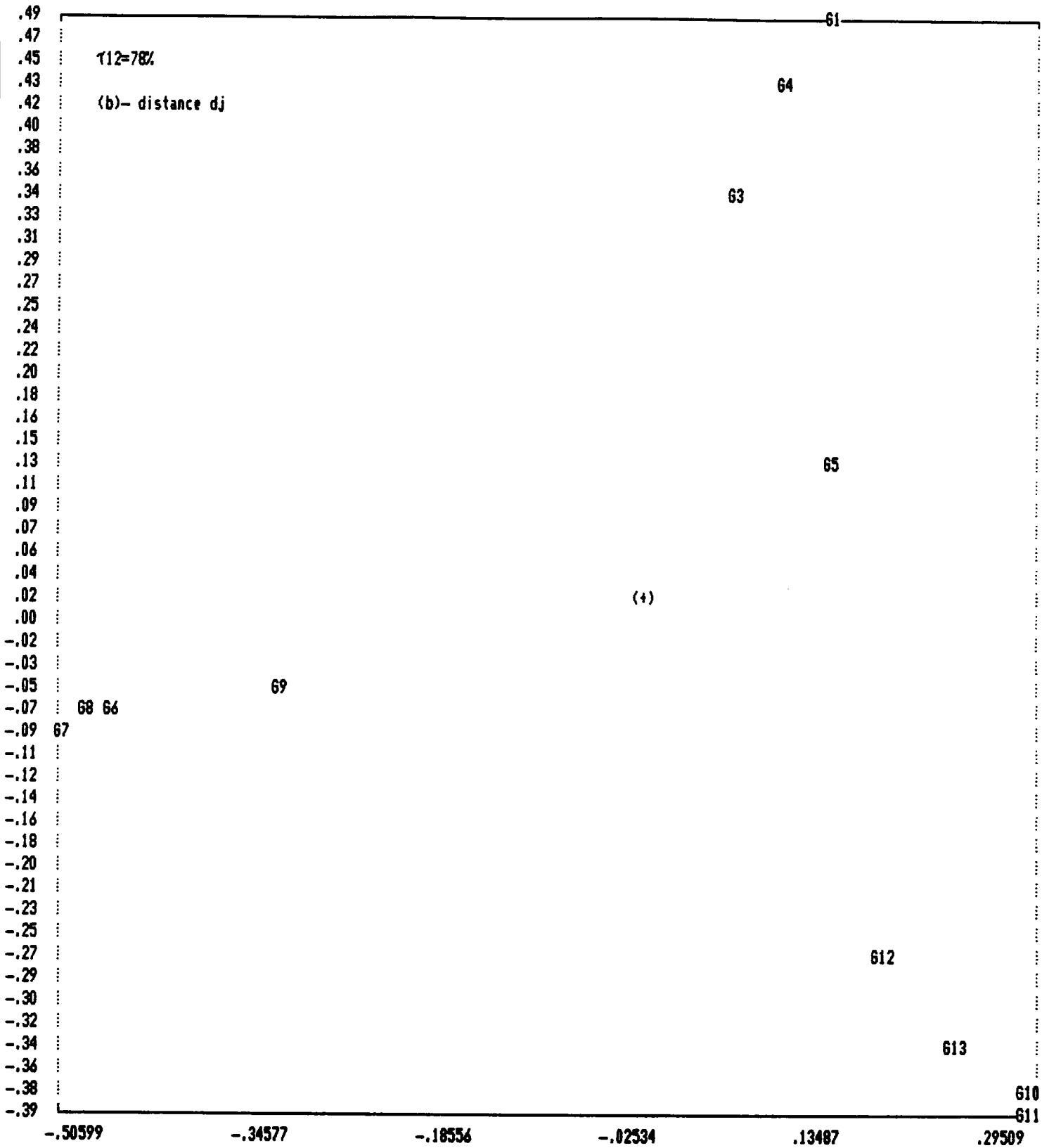


Figure-III-7: Représentation graphique des 13 gammes-tests sur les deux premiers axes factoriels.



POINTS MULTIPLES SONT:

1 : ( 61 , 62 )

Figure-III-7: Représentation graphique des 13 gammes-tests sur les deux premiers axes factoriels (suite).



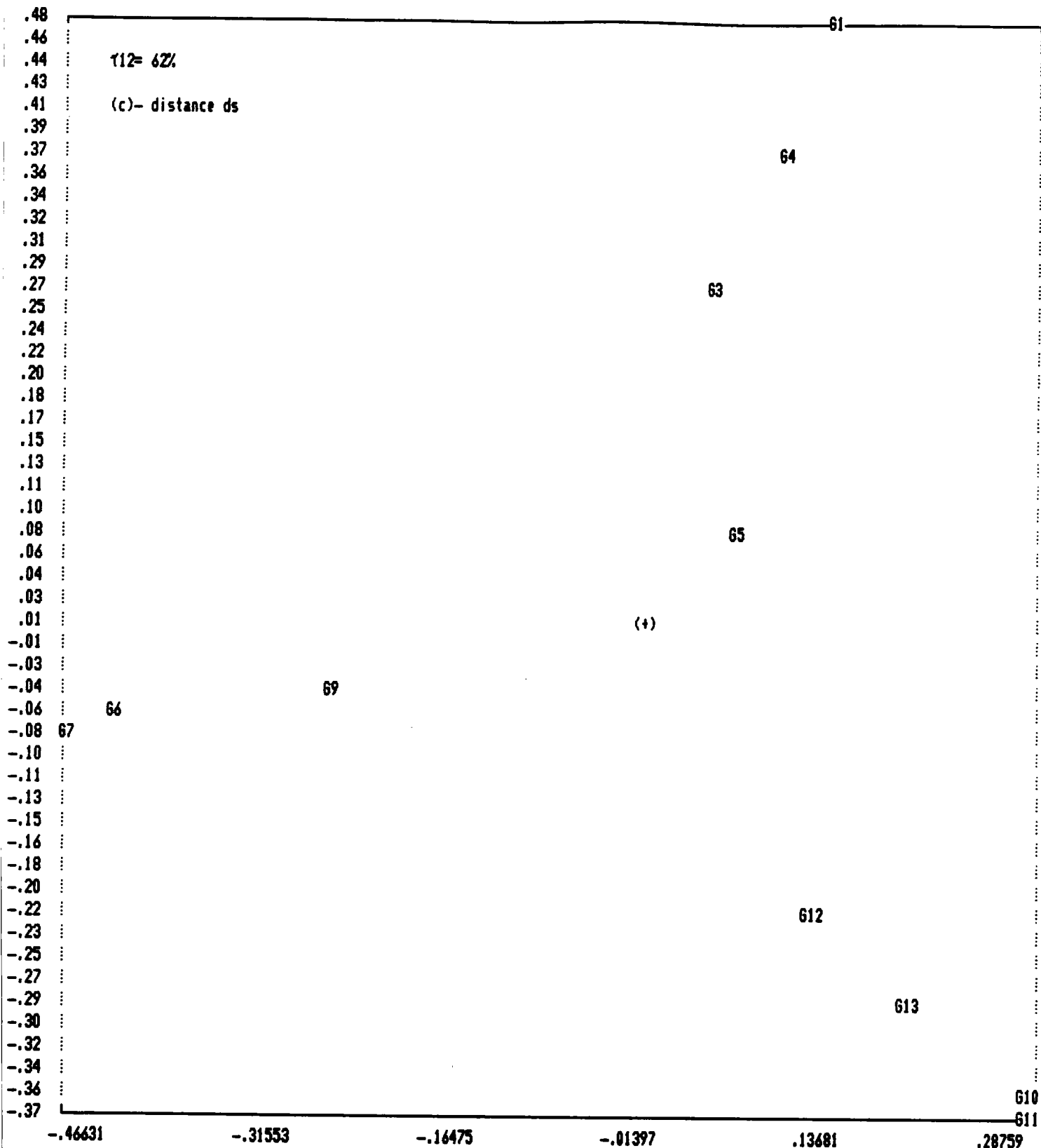


Figure-III-7: Représentation graphique des 13 games-tests sur les deux premiers axes factoriels (fin).

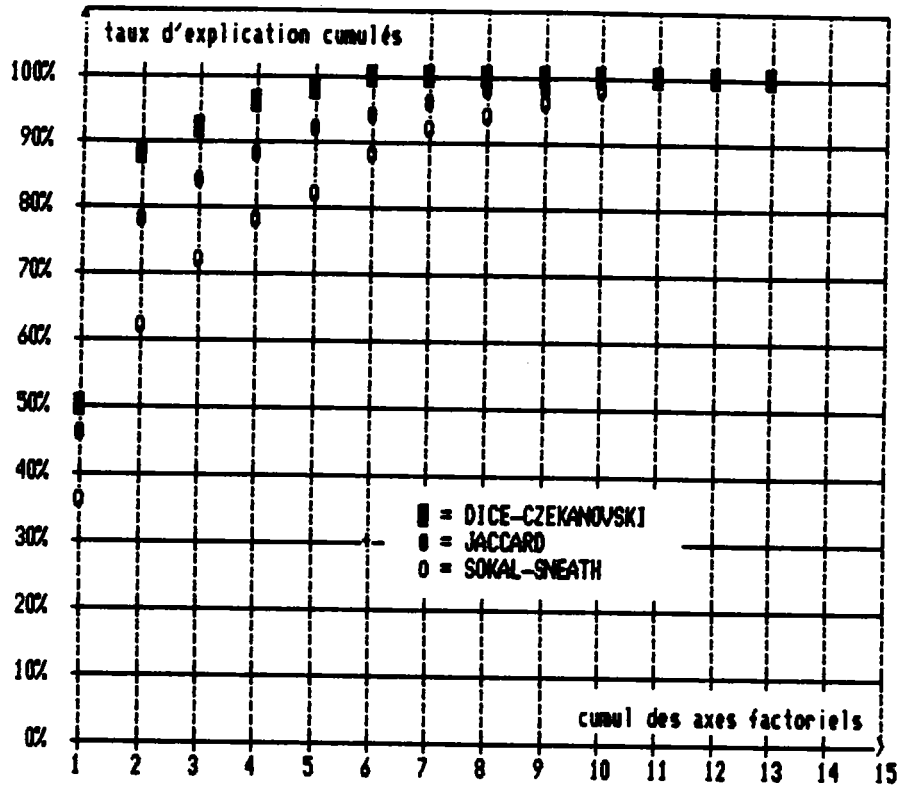


Figure-III-8: Représentation graphique des taux d'explication pour les distances  $d_d$ ,  $d_j$ , et  $d_s$ .

Deux remarques s'imposent: d'une part la distance  $dd$  reconnaît mieux que les distances  $ds$  et  $dj$  que les gammes U-V-W-X et U-V-W se ressemblent. D'autre part, l'utilisation des distances  $dj$  et surtout  $ds$  revient à considérer que vues de la gammes U-V-W-X, les gammes U-V et U sont presque aussi éloignées l'une que l'autre, alors que la distance  $dd$  permet un meilleur étalement et une meilleure discrimination, comme le montre la figure-III-9.

Mathématiquement ceci s'explique par le fait que ces trois distances peuvent se mettre sous la forme générale:

$$(1) \quad d = \sqrt{\frac{N}{\alpha P + N}}$$

avec respectivement  $\alpha=2$ ,  $\alpha=1$  et  $\alpha=\frac{1}{2}$  pour les mesures  $dd$ ,  $dj$  et  $ds$ . Plus  $\alpha$  est grand, plus la ressemblance entre gammes traduite par la co-présence  $P$  a un poids important. Notons enfin que, d'après Fichet et le Calve [51], l'expression (1) est une distance euclidienne pour  $\alpha \leq 2$ .

La valeur optimale  $\alpha=2$  correspond donc à la distance de Dice-Czekanowski. Pour toutes ces raisons, la mesure

$$dd = \sqrt{\frac{N}{2P + N}}$$

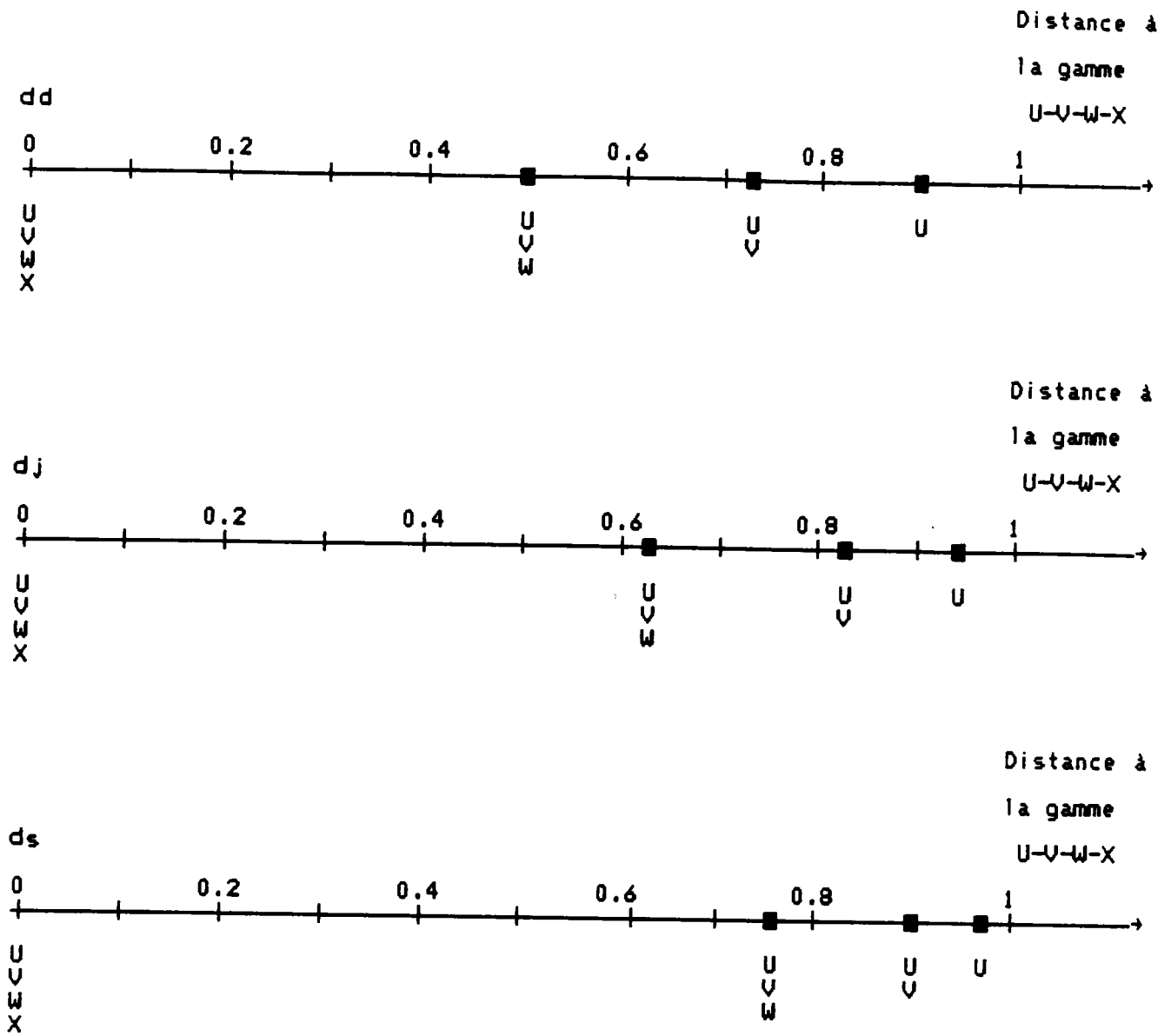


Figure-III-9: Représentation graphique des distances de Dice-Czekanowski, Jaccard et Sokal-Sneath entre la gamme U-V-W-X et les gammes U-V-W, U-V et U.

semble le mieux convenir.

### a-3) Validation

Il s'agit maintenant de vérifier que cette distance est bien adaptée à la mesure de proximité entre gammes de fabrication.

En effet, une bonne mesure doit prendre en compte l'ordre des phases dans la séquence, le nombre de phases relatif à chaque gamme, différencier une absence de phase d'une permutation et enfin ne pas trop pénaliser l'introduction d'une phase.

D'où l'idée de définir un certain nombre de critères:

C1: non-coïncidence

C2: ordre des phases

C3: différence de rang (permutation)

C4: nombre de phases par gamme

Le tableau suivant montre que la distance  $d_d$  vérifie bien ces critères, et qu'elle est donc tout à fait appropriée à la mesure de ressemblance entre gammes de fabrication:

Critère	Définition	Observation	Illustration		
			Gamme G1	Gamme G2	Distance
C1	non-coïncidence	plus la non-coïncidence croît, plus l'indice de proximité décroît	FBCDEA	BCEDF	0.667
			BCDFJKL	BCEDF	0.731
C2	ordre des phases	l'ordre des phases est fidèlement traduit	IJKL	IJLK	0.316
			IJKL	ILJK	0.447
C3	différence de rang (permutation)	une permutation de deux phases consécutives a moins d'incidence qu'une permutation entre deux phases éloignées	ABCDEF	ABCDFE	0.218
			ABCDEF	FBCDEA	0.655
C3		une permutation a plus d'incidence sur une petite gamme que sur une grande	ABCDEF	ABCDFE	0.218
			IJKL	IJLK	0.316
C4	nombre de phases par gamme	62 $\subset$ 61			
		-si (n1-n2) décroît, la distance entre G1 et G2 décroît	UWDX	UW	0.500
			UWDX	UW	0.734
		-si (n1-n2)=constante, plus le nombre de phases par par gamme est élevé, plus la distance décroît	UWDX	UW	0.500
			UW	UW	0.577
			UW	U	0.707

b) Choix d'un indice de proximité pour les autres données de production (application aux codes morphodimensionnels)

Contrairement au cas des données ordinales incomplètes, l'utilisation de n'importe quel indice de similarité peut convenir pour le traitement de codes constitués de variables indépendantes. D'une façon générale, il est préférable cependant d'utiliser une distance.

Les codes morphodimensionnels sont transformés dans une étape préalable en tableau disjonctif complet, comme cela est indiqué au paragraphe III-2, et ils se présentent donc sous la forme de données binaires qui possèdent un certain nombre de caractéristiques:

-le nombre total de modalités  $p$  prises par chaque objet est égal au nombre de variables décrivant les objets. Il est donc identique pour tous les objets:

$$(2) \quad \sum_{k=1}^p \sum_{f=1}^{n_k} \beta_{ikf} = \sum_{k=1}^p \sum_{f=1}^{n_k} \beta_{ikf} = p \quad \forall (\omega_i, \omega_j) \in \Omega^2$$

avec:

. $p$ : nombre de variables

. $n_k$ : nombre de modalités de la  $k$ -ième variable

. $\beta_{ikf}$ : terme général du tableau disjonctif complet:

$\beta_{ikf}=1$  si l'objet  $\omega_i$  possède la modalité  $k_f$  dans la variable  $V_k$

$\beta_{ikf}=0$  sinon

-le nombre total  $t$  de modalités prises par l'ensemble des objets est égal au nombre de colonnes du tableau disjonctif complet. Il peut se mettre sous la forme:

$$(3) \quad t = \sum_{k=1}^p n_k$$

Dans ces conditions la formulation de la non-coïncidence  $N$ , de la co-présence  $P$  et de la co-absence  $A$  entre deux objets  $\omega_i, \omega_j$  est relativement simple:

$$(4) \quad N = \sum_{k=1}^p \sum_{f=1}^{n_k} (\beta_{ikf} - \beta_{jkf})^2$$

$$(5) \quad P = \sum_{k=1}^p \sum_{f=1}^{n_k} (\beta_{ikf} \times \beta_{jkf})^2$$

$$(6) \quad A = \sum_{k=1}^p \sum_{f=1}^{n_k} (1 - \beta_{ikf}) \times (1 - \beta_{jkf})$$

Le calcul montre:

-que la distance euclidienne s'écrit:

$$\text{d'après (4)} \quad d^2(\omega_i, \omega_j) = N$$

-que les indices de dissimilarité de Dice-Czekanowski  $S_d$  et de Sokal-Michener  $S_{sm}$  appliqués à un tableau disjonctif complet sont proportionnels au carré de la distance euclidienne:



$$\delta_d = \frac{N}{2p} \quad (\text{d'après (2), (4) et (5)})$$

$$\delta_{sm} = \frac{N}{t} \quad (\text{d'après (2), (3), (4), (5) et (6)})$$

Lorsque l'analyse porte simultanément sur des données hétérogènes, il est indispensable d'utiliser la même distance pour l'ensemble du traitement.

Si, comme c'est le cas dans cette étude, elle est conduite à la fois sur des données ordinales incomplètes et sur des données codes, le choix de la distance est imposé par les données ordinales incomplètes qui ne peuvent être traitées que par une distance de la forme:

$$d = \sqrt{\frac{N}{\alpha P + N}}$$

la plus appropriée étant celle déduite de l'indice de Dice-Czekanowski ( $\alpha=2$ ).

C'est donc cette dernière distance qui a été retenue pour traiter l'ensemble des données. J

### III-4/ REPRESENTATION GRAPHIQUE DES DONNEES

Etant donné un ensemble d'objets  $\Omega$ , l'Analyse en Composantes

Principales permet à partir du tableau des distances entre les différents éléments de  $\Omega$  de visualiser les objets. En particulier elle permet d'obtenir leurs projections sur le plan déterminé par les deux premiers axes factoriels.

Lorsqu'une analyse simultanée de deux types de données  $\Omega_1$  et  $\Omega_2$  est conduite, la comparaison des deux images obtenues par ACPTD sur ces deux types de données montre qu'en règle générale elles ne sont pas identiques: il n'existe pas a priori de bijection entre les deux ensembles de données  $\Omega_1$  et  $\Omega_2$ .

La solution proposée dans cette étude pour traiter simultanément les données  $\Omega_1$  et  $\Omega_2$  est la constitution d'un tableau distances  $[\Omega]$  à partir des matrices de distances  $[\Omega_1]$  et  $[\Omega_2]$ :

$$[\Omega] = \frac{\alpha_1}{\alpha_1 + \alpha_2} [\Omega_1] + \frac{\alpha_2}{\alpha_1 + \alpha_2} [\Omega_2]$$

$\alpha_1$  et  $\alpha_2$  sont les poids correspondant respectivement aux données des ensembles  $\Omega_1$  et  $\Omega_2$ . Les valeurs de  $\alpha_1$  et de  $\alpha_2$  sont choisies par l'utilisateur. Pour  $\alpha_1=0$ , on se ramène à l'étude de l'ensemble  $\Omega_2$  seul et pour  $\alpha_2=0$ , à celle de l'ensemble  $\Omega_1$  seul.

Le traitement de ce tableau de distances  $[\Omega]$  par l'ACPTD permet d'obtenir la projection de l'ensemble des données sur le premier plan factoriel. Les axes factoriels étant classés par ordre d'inertie décroissant, le plan défini par les premiers axes factoriels est la plan d'inertie maximale. La qualité de la représentation est définie par le taux d'explication  $\tau$  qui est égal à la somme des inerties des deux premiers axes factoriels. Plus  $\tau$  est proche de 100%, meilleure est la représentation (cf chapitre II-2).

Bien qu'elle ne soit pas une image entièrement fidèle de la réalité, cette représentation graphique permet de mettre en évidence la répartition apparente des objets.

### III-5/ CLASSIFICATION AUTOMATIQUE PAR LA METHODE DES NUÉES DYNAMIQUES

Nous disposons d'un ensemble  $\Omega$  à  $n$  objets:

$$\Omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_j, \dots, \omega_n)$$

et nous cherchons une partition  $\{PK$  de  $\Omega$  en  $k$  classes disjointes:

$$\{PK = (P_1, P_2, P_3, \dots, P_j, \dots, P_k)$$

Chaque classe  $P_j$  sera représentée par son noyau  $\lambda_j$ .  $\Lambda_k$  est l'ensemble des  $k$  noyaux des  $k$  classes de la partition  $\{PK$ :

$$A_k = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_j, \dots, \lambda_k)$$

Dans le cas où  $A_k \subset \Omega$ , nous avons pour tout  $j$  variant de 1 à  $k$ ,  $\lambda_j \subset P_j$ .

L'utilisation de la méthode des Nuées Dynamiques suppose [52]:

a) L'initialisation du nombre de classes et des noyaux correspondants

Une initialisation rationnelle du nombre de classes et des noyaux associés peut être effectuée à partir de la représentation graphique des objets par l'ACPTD.

b) La définition d'une fonction d'affectation  
 $f: A_k \rightarrow \{P_k\}$

Elle permet d'allouer les objets de  $\Omega$  aux  $k$  noyaux. Un objet  $\omega_i$  sera affecté à la classe  $P_j$  si sa distance au noyau  $\lambda_j$  est inférieure à celle existant entre  $\omega_i$  et les autres noyaux:

$$d(\omega_i, \lambda_t) = \text{Min}_{r=1, k} (d(\omega_i, \lambda_r))$$

c) La définition d'une fonction de représentation  
 $g: \{P_k\} \rightarrow A_k$

Parmi les choix possibles pour le représentant de chaque classe, l'utilisation du barycentre comme noyau est l'une des possibilités les plus séduisantes.

Cependant, en règle générale, pour les données de production, la définition de la fonction de représentation  $g$  est délicate, car l'analyse par la méthode des Nuées Dynamiques est conduite non sur les coordonnées des objets mais sur les distances entre paires d'objets.

Il n'est pas possible dans ces conditions de calculer le barycentre de chaque classe.

La solution adoptée dans cette étude consiste à déterminer pour chaque classe d'objets l'élément le plus proche du barycentre en faisant appel au théorème de Huygens.

L'inertie d'un objet  $\omega_i$  muni d'une masse  $m_i$  par rapport à un objet  $\omega_x$  s'écrit :

$$I_{\omega_x}(\omega_i, m_i) = m_i \cdot d^2(\omega_i, \omega_x)$$

où  $d(\omega_i, \omega_x)$  désigne la distance entre les objets  $\omega_i$  et  $\omega_x$ .

De la même façon, l'inertie d'un ensemble  $\Omega$  de  $n$  objets  $((\omega_i, m_i)$  avec  $i$  variant de 1 à  $n$ ) par rapport à un objet  $\omega_x$  se définit comme la somme des inerties des objets de l'ensemble  $\Omega$  par rapport à l'objet  $\omega_x$  :

$$I_{\omega_x}(\Omega) = \sum_{i=1}^n m_i \cdot d^2(\omega_i, \omega_x)$$

D'après le théorème de Huygens,  $I_{\omega x}(\Omega)$  peut se mettre encore sous la forme:

$$I_{\omega x}(\Omega) = I_G(\Omega) + \left( \sum_{i=1}^n m_i \right) \cdot d^2(\omega x, G)$$

où  $I_G(\Omega)$  désigne l'inertie de l'ensemble  $\Omega$  par rapport à son barycentre  $G$ .

Il se déduit de ce théorème que l'inertie de l'ensemble  $\Omega$  prend sa valeur minimale au barycentre:

$$\forall \omega x \neq G, \quad I_{\omega x}(\Omega) > I_G(\Omega)$$

Par analogie et étant donné une classe  $P_j$  de  $n_j$  objets, l'objet  $\lambda_j$  le plus proche du barycentre  $G_j$  sera celui par rapport auquel l'inertie de la classe  $P_j$  sera la plus petite, d'où la fonction de représentation  $g$ :

$$g(P_j) = \text{Min}_{q=1}^{n_j} \left( \sum_{i=1}^{n_j} m_i \cdot d^2(\omega x, \omega q) \right)$$

d) Définition d'un critère  $W$  ( $W: A_K \times \{P_K \longrightarrow \mathbb{R}^+\}$ ) à optimiser, qui évalue l'homogénéité des classes.

Dans le cas où le barycentre de chaque classe est connu, l'homogénéité des classes est mesurée par l'inertie intra-classe  $W_1$ , qui se définit comme la somme de l'inertie de chaque classe

$P_j$  autour de son barycentre  $G_j$ . Plus l'inertie intra-classe est petite, plus les classes sont compactes:

$$W_1 = \sum_{j=1}^k \sum_{i=1}^{n_j} m_i \cdot d^2(\omega_i, G_j)$$

La minimisation de l'inertie intra-classe revient à maximiser l'inertie inter-classes  $W_2$ , égale à la somme des inerties des barycentres  $G_j$  des classes  $P_j$  par rapport au barycentre  $G$  de l'ensemble des objets  $\Omega$ . Plus l'inertie inter-classes est grande, plus les classes sont disjointes:

$$W_2 = \sum_{j=1}^k M_j \cdot d^2(G_j, G) \quad \text{avec} \quad M_j = \sum_{i=1}^{n_j} m_i$$

En effet, d'après la relation de Huygens, la somme de l'inertie intra-classe et de l'inertie inter-classes d'une partition d'un ensemble d'objets  $\Omega$  est égale à une constante, l'inertie totale de cet ensemble par rapport à son barycentre  $G$ :

$$\sum_{i=1}^n m_i \cdot d^2(\omega_i, G) = \sum_{j=1}^k \sum_{i=1}^{n_j} m_i \cdot d^2(\omega_i, G_j) + \sum_{j=1}^k M_j \cdot d^2(G_j, G)$$

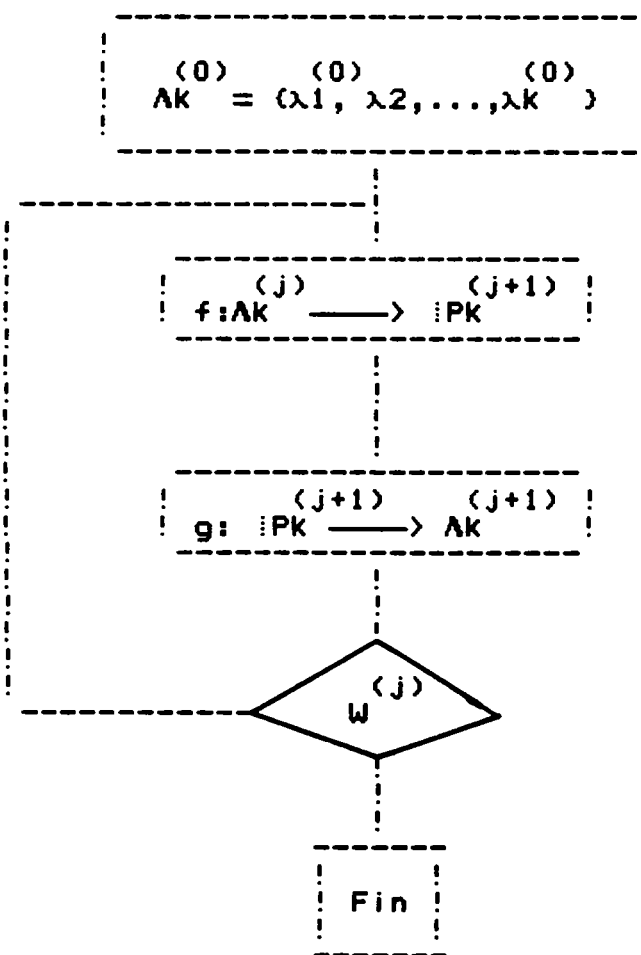
Le calcul de barycentre nécessite la connaissance des coordonnées des objets, dont on ne dispose pas dans cette étude où seuls les tableaux de distances entre objets sont disponibles. L'inertie intra-classe s'écrira alors en première approximation:

$$W_1 = \sum_{j=1}^k \sum_{i=1}^{n_j} m_i \cdot d^2(\omega_i, \lambda_j)$$

où  $\lambda_j$  désigne le noyau de la classe  $P_j$  et est l'élément le plus proche du barycentre de cette classe.

C'est ce critère qu'il faudra donc minimiser.

Une fois que ces quatre étapes préliminaires sont franchies, l'algorithme consiste à faire intervenir par itérations successives la fonction d'affectation  $f$  et la fonction de représentation  $g$  à partir d'un tirage initial  $\Lambda_K^{(0)}$  à  $K$  noyaux. Le critère  $W$  associé à  $f$  et  $g$  doit décroître pendant le déroulement de l'algorithme (figure-III-10).



Initialisation:

choisir les  $K$  noyaux de l'ensemble  $\Omega$ .

Affectation:

allouer les  $n$  objets aux  $K$  noyaux pour la formation de la partition  $P_K^{(j+1)}$ .

Représentation:

redéfinir les nouveaux noyaux  $\Lambda_K^{(j+1)}$ .

Le critère  $W$  évolue-t-il ?

Figure-III-10: Schéma de l'Algorithme des Nuées Dynamiques.



L'algorithme des Nuées Dynamiques est quasi-optimisant, car il converge vers un optimum local et non vers un optimum global. Cet optimum dépend de la partition initiale. Il est donc impossible de savoir si la partition obtenue est la meilleure.

Cependant, le résultat souhaité peut être approché en effectuant plusieurs tirages à  $k$  classes et en comparant les partitions obtenues: les objets qui ont toujours été classés ensemble définissent des Formes Fortes [53] qui sont les parties vraiment homogènes et stables de l'ensemble  $\Omega$ , puisqu'elles ont résisté aux aléas des tirages des noyaux initiaux. Le nombre de formes fortes est supérieur ou égal à  $k$ . Le cas échéant, il est possible d'utiliser une Classification Hiérarchique sur les formes fortes pour réduire le nombre de classes et ainsi aider à l'interprétation des résultats.

### III-6/ DESCRIPTION DES FAMILLES OBTENUES

#### (EXEMPLE DES DONNEES DE PRODUCTION)

La description des familles doit être adaptée à chaque cas spécifique traité. Elle sera détaillée ici exclusivement dans le cas des données de production: séquences de fabrication et codes morphodimensionnels.

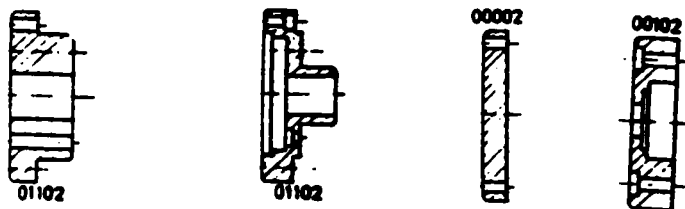
Les familles étant définies, chacune d'entre elles est représentée du point de vue des séquences de fabrication par sa gamme mère sous la forme d'un graphe. Ce graphe, qui rassemble tous les poste de travail intervenant dans la fabrication des différentes pièces de la famille, se déduit du tableau  $U$  dont le terme général  $u_{ij}$  est égal au nombre de fois où la machine  $i$  alimente la machine  $j$ . Le terme  $u_{ij}$  peut également représenter, au choix de l'utilisateur, le nombre de pièces ou de lots transmis

directement de la machine i à la machine j. Ce terme peut aussi être binaire, et prendre la valeur 1 dès que la machine i alimente la machine j pour l'une quelconque des pièces de la famille et la valeur 0 sinon.

Supposons qu'une famille soit composée de quatre pièces décrites avec le code Opitz [54] par exemple:

Références	Codes	Gammes
$\omega_1$	0 1 1 0 2 - x x x x	A-B-C-D
$\omega_2$	0 1 1 0 2 - x x x x	A-C-D-E
$\omega_3$	0 0 0 0 2 - x x x x	A-B-E
$\omega_4$	0 0 1 0 2 - x x x x	B-C-D-E

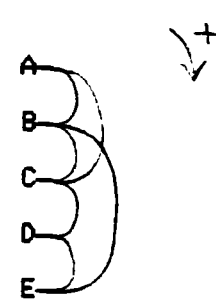
correspondant aux pièces suivantes:



Le tableau U correspondant à cette famille type, ainsi que la gamme mère qui s'en déduit sont représentés ci-dessous.

	A	B	C	D	E	
0	2	1	0	0		A
0	0	2	0	1		B
0	0	0	3	0		C
0	0	0	0	2		D
0	0	0	0	0		E

U =



Du point de vue du code, la famille est représentée par son spectre S, dont le terme général  $s_{ij}$  est égal au nombre de fois, exprimé en % ou en effectif, où la modalité  $j$  est prise pour la position  $i$ . Le terme  $s_{ij}$  peut également être binaire, et prendre la valeur 1 si la modalité  $j$  est prise dans la position  $i$  pour l'une quelconque des pièces de la famille, et la valeur 0 sinon.

Dans l'exemple type décrit ci-dessus, le spectre de la famille considérée est représenté en pourcentage sur la figure-III-11-a et en binaire sur la figure-III-11-b:

		MODALITES									
		0	1	2	3	4	5	6	7	8	9
POSITIONS	1	100									
	2	50	50								
	3	50	50								
	4	66	33								
	5			100							
	6	X	X		X	X					
	7	X				X		X			X
	8	X	X		X						
	9	X							X		

figure-III-11-a : exemple de spectre de famille

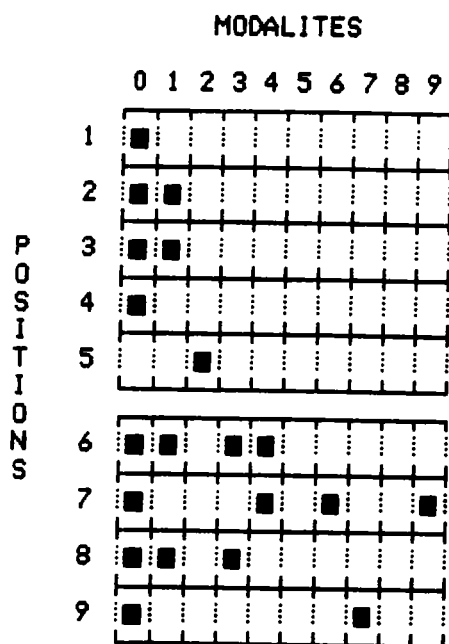


Figure-III-11-b: Exemple de spectre de famille binaire

### III-7/ CLASSEMENT ET VALIDATION

Lorsque les familles sont définies, ainsi que les spectres les décrivant, il s'agit maintenant de classer une pièce quelconque, de la classer dans l'une des familles.

Pour ce faire, il est nécessaire de représenter cette pièce à l'aide de son spectre. Celui-ci est obtenu à partir de son code de la même façon que cela est réalisé pour le spectre des familles:

$s_{ij} = 1$  si la modalité  $j$  est prise pour la position  $i$

$s_{ij} = 0$  sinon.

D'une manière classique, le spectre de la pièce, binaire par définition, est comparé aux spectres binaires des différentes familles issues de la classification  $(S^1, S^2, S^3, \dots, S^k, \dots, S^f)$ . Cependant, cette façon de procéder n'est pas nécessairement optimale, puisque dans le spectre binaire d'une famille  $k$ , il n'existe pas de différence entre une modalité prise fréquemment et une modalité prise rarement. Dans les deux cas, on a :

$$s_{ij}^k = 1$$

D'où l'idée de proposer une démarche plus élaborée, faisant intervenir un seuil  $\sigma$ , qui va être défini ci-dessous. Si on considère l'ensemble des spectres des familles résultant de la classification, exprimés en pourcentage, le seuil  $\sigma$  va servir à transformer ces tableaux en spectres binaires  $S^{\sigma k}$ , dont le terme général  $s_{ij}^{\sigma k}$  s'écrira :

$$\begin{aligned} & \text{pour } s_{ij}^k = 0 & s_{ij}^{\sigma k} &= 0 \\ & \text{pour } s_{ij}^k \neq 0 & s_{ij}^{\sigma k} &= 1 \text{ si } s_{ij}^k \geq \sigma \\ & & s_{ij}^{\sigma k} &= 0 \text{ sinon} \end{aligned}$$

Lorsque  $\sigma=0$ , les spectres  $S^{\sigma k}$  obtenus sont les spectres binaires définis au paragraphe précédent.

Ensuite, pour chaque valeur de  $\sigma$  comprise entre 0 et 100%, le spectre  $S_p$  de la pièce est comparé aux spectres binaires  $S^{\sigma k}$  des différentes familles issues de la classification et affecté à celle dont le spectre est le plus proche.

Il est alors possible, en s'intéressant aux pièces ayant servi à la classification, de tracer, en fonction de la valeur prise par le seuil  $\sigma$ , le pourcentage de ces pièces qui seront bien classées. La valeur de  $\sigma$  pour laquelle ce pourcentage sera maximal est ensuite adoptée pour effectuer le classement de pièces nouvelles dans les meilleures conditions.

### III-8/ CONCLUSION

La méthodologie qui vient d'être décrite permet le traitement automatique d'un type quelconque de données et l'analyse simultanée de plusieurs ensembles.

Grâce à l'utilisation rationnelle des algorithmes d'Analyse des Données, elle conduit à une classification et à un classement optimisés, bien adaptés aux problèmes industriels.

Une application de cette méthodologie au cas des données de fabrication (gamme d'usinage et codes morphodimensionnels Multi-M) va être détaillée au chapitre suivant.

## CHAPITRE IV

**LOGICIEL CL<sup>2</sup>AUDIA ET APPLICATION INDUSTRIELLE**

Le logiciel CL<sup>2</sup>AUDIA est un système de Classification et de Classement Automatique de Données Industrielles et Analytiques. Il peut être utilisé pour le traitement de tous les types de données de production sans restriction.

CL<sup>2</sup>AUDIA est écrit en FORTRAN77 sur un VAX sous VMS; la taille des données est illimitée étant donné que VMS travaille en mémoire virtuelle.

L'application du logiciel CL<sup>2</sup>AUDIA à des données industrielles, qui fait l'objet de ce chapitre, porte sur un échantillon de 70 références fourni par l'ADEPA. Une référence "R" est la réunion des données codes+gamme de fabrication.

Cet échantillon est relatif à la fabrication en petite série: ces 70 références correspondent à 203 pièces tridimensionnelles réparties en 100 lots.

L'objectif est la recherche de groupes technologiques au bureau des méthodes. Dans ces conditions  $\Omega_1$  est l'ensemble des codes "C" et  $\Omega_2$  l'ensemble des gammes de fabrications "G".

CL<sup>2</sup>AUDIA, dont l'architecture est représentée à la figure IV-1, regroupe quatre procédures dont les fonctions seront détaillées aux paragraphes suivants:

- "PTG": analyse de l'échantillon préparation des données,  
et choix des distances

- "ACP": représentation graphique des références analysées, à l'aide de l'Analyse en Composantes Principales sur un Tableau de Distances



- "AND" et "CAH": classification automatique par l'Algorithme des Nuées Dynamiques, éventuellement couplée à un regroupement de familles par Classification Ascendante Hiérarchique. Classement.

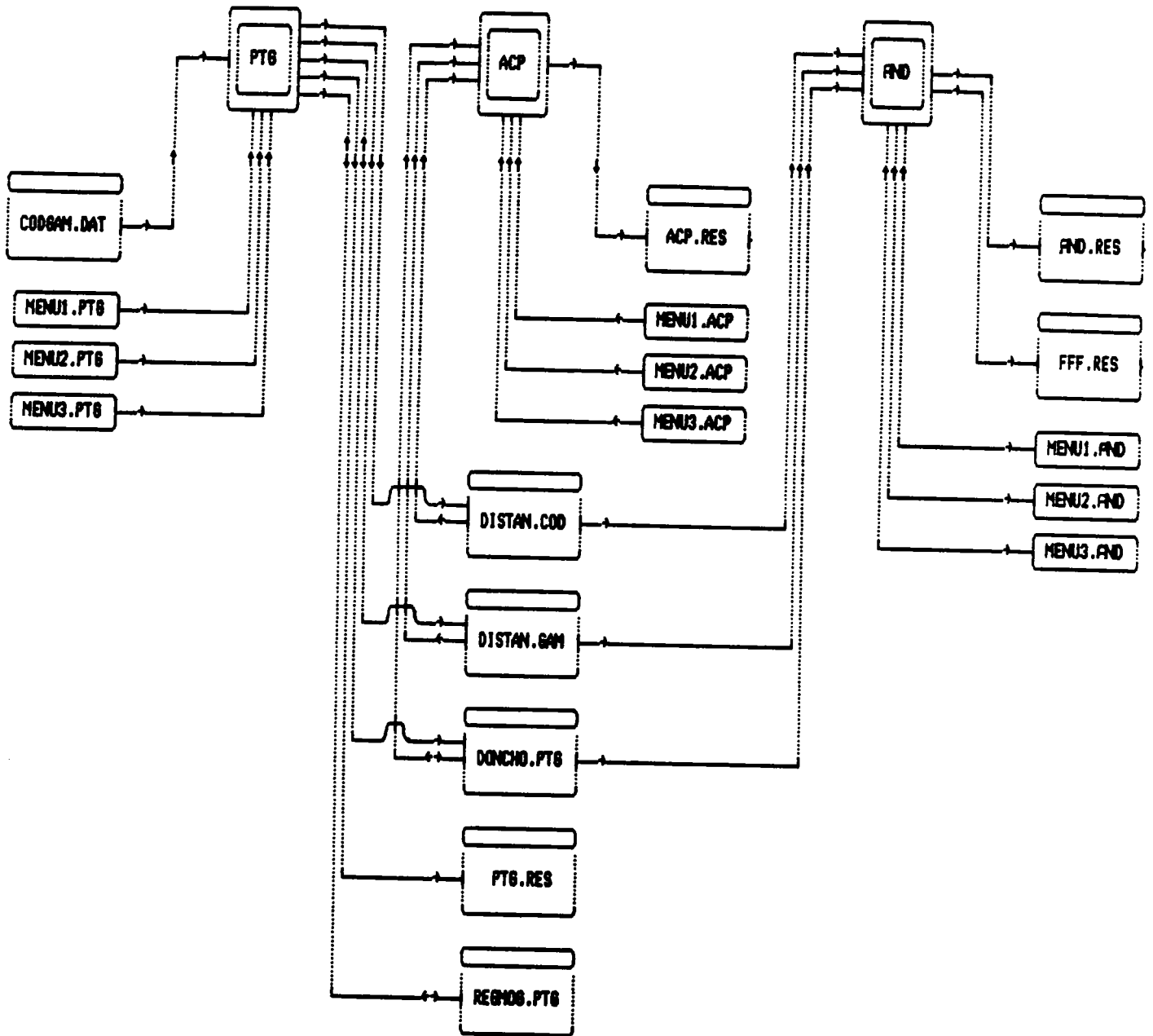


figure-IV-1: Organigramme général du logiciel CLAUDIA

IV-1/ ANALYSE DE L'ECHANTILLON ET PREPARATION DES DONNEES

Les données à traiter sont regroupées dans le fichier "CODGAM.DAT" décrit à la figure-IV-2, sous le format MTRAN [55] mis au point par l'ADEPA:

```
DEBUT
COMPTE      :
REF.PIECE   :512906/49
DESIGNATION1:ELEMENT MATRICE
CODE MULTI-M:1764341302316931409000000000000000
PREPARATEUR :
DATE CREAT. :02/01/80
N MODIF.    :1
MODIFICATION:CREATION
:
:
:
N LANCEMENT:935010
DELAI       :02/01/80
NB. LANC.   : 2
DATE LANC.  :07/01/80
QTE LANC.   : 1
ATIERE      :SANCY 2180C13
OBSERVATION1:1/65372
ATTENTE     :
QTE./AN     :      D
MILIEU
  05  FRAISAGE EBAUCHE COMPLETE AVEC RESERVE      1.80      1.80
  10  AJUSTAGE 2 M6                                0.30      0.50
  15  T.THERM. HRC 59/60
  20  RECT.PP. COMPLET SUIVANT PLAN                 1.50      1.90
FIN
```

figure-IV-2: Structure d'une référence du fichier "CODGAM.DAT"

Chaque référence regroupe le code Multi-M plus les données relatives à la fabrication de la pièce (gamme).

IV-1-A/ Définition du code Multi-M [43, 56]

Le code Multi-M est composé de 31 positions dont les 19 premières sont appelées code principal et les 12 restantes sont le code spécifique (Figure-IV-3)

Le code principal contient des informations communes à toutes les sociétés avec toutefois la possibilité d'adapter les positions relatives aux fonctions, aux tolérances et aux matières utilisées.

Le code spécifique contient des informations propres à chaque société.

Le code Multi-M fait partie d'un réseau de codification comprenant 9 codes (réseau Multi-Class). Pour cette raison, la première position du code principal est toujours égale à 1 et représente le code Multi-M.

La configuration globale des pièces est donnée au niveau des positions 2, 3 et 4.

Le code Multi-M répartit les pièces en 9 catégories qui sont représentées dans la position 2 (Figure-IV-4):

Les modalités 1 à 4 correspondent au groupe des pièces de révolution, 5 et 6 à celui des pièces de tôlerie. Enfin les modalités 7 à 9 concernent le groupe des pièces tridimensionnelles.

La position 2, position d'appartenance à l'une des 9 catégories principales, conditionne la signification des positions 3 et 4, dont le contenu varie de 0 à 9.

0 1 2 3 4 5 6 7 8 9

C O D E  
P R I N C I P A L

1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										

Code MULTI-M
Forme
Fonction
Dimensions
Tolérance
Matière
Forme du brut
Quantité produite et cadence
Complexité

C O D E  
S P E C I F I Q U E

20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										


Figure-IV-3: Code Multi-M

Contenu de la position 2	Signification	Groupe de pièces
1	pièces étagées d'un côté	rondes
2	étagées des deux côtés	
3	avec cône et/ou profil	
4	spéciales	
5	plates	tôlerie
6	déformées	tridimensionnelles
7	parallélépipédiques	
8	sculptées	
9	irrégulières	

Figure-IV-4: Types de pièces décrites par le code Multi-M

Les caractéristiques des diamètres extérieur et intérieur sont codifiées dans ces positions pour les quatre premières catégories de pièces rondes.

Pour les cinq catégories de pièces non rondes, les positions 3 et 4 contiennent les informations relatives au périmètre, à la symétrie des vues et aux éléments propres à caractériser les pièces.

Les informations concernant les usinages secondaires tels que percage, taraudage etc..., sont données à la position 5.

Les positions 6 et 7 représentent la fonction de la pièce.

Les positions 8 à 13 servent à codifier par trois doublets les informations relatives aux caractéristiques dimensionnelles

(dimension principale, dimension secondaire et dimension complémentaire).

Les tolérances d'usinage caractérisant la précision de la pièce sont codifiées à la position 14. Cette précision est considérée relativement aux :

- caractéristiques dimensionnelles,
- caractéristiques d'états (de surface...)
- caractéristiques d'assemblage.

On se sert des positions 15 et 16 pour codifier les matières utilisées pour la fabrication des pièces telles que : alliages légers, différentes nuances d'aciers etc...

La position 17 concerne la codification de l'origine de la pièce.

L'importance des lots de pièces (petits, moyens ou grands) lancés en fabrication est indiquée à la position 18.

La complexité traduite par le nombre des moyens de fabrication qu'une pièce nécessite est codifiée à la position 19.

Le code spécifique contient des données propres à chaque entreprise et portées aux positions 20 à 31. L'utilisateur peut y mettre des informations concernant les prix, le temps de fabrication etc...

J

IV-1-B/ Procédure "PTG"

a) MENU1.PTG

La procédure "PTG" lit les données contenues dans le fichier "CODGAM.DAT" et permet grâce au MENU1.PTG (figure-IV-5) de choisir le type de pièces auquel on va s'intéresser.

---

TYPE DE PIECES A ANALYSER

- (1)- RONDES ETAGEES D'UN COTE
  - (2)- RONDES ETAGEES DES DEUX COTES
  - (3)- RONDES AVEC CONE ET/OU PROFIL
  
  - (4)- RONDES SPECIALES ( RESSORT,...)
  
  - (5)- PLATES
  - (6)- DEFORMEES
  
  - (7)- PARALLELEPIPEDIQUES
  - (8)- SCULPTEES
  - (9)- IRREGULIERES
  
  - (0)- FIN DU TRAVAIL
  
  - VOUS POUVEZ COMBINER LES DIFFERENTS TYPES DE PIECES  
A L'INTERIEUR DE CHAQUE GROUPE
- QUEL EST VOTRE CHOIX ?
- 

Figure-IV-5: Menu du choix des pièces à analyser.

S'il s'agit d'une simple description de l'échantillon, l'utilisateur a la possibilité d'examiner un ou plusieurs types à la fois. Après sélection des données choisies, celles-ci sont stockées dans le fichier "DONCHO.PTG". Les résultats de cette

description sont les suivants:

1) analyse des charges machines (figure-IV-6 a et b)

1-a) pour chaque référence:

- le nombre de pièces
- les temps totaux de charge, de réglage et de travail
- la gamme.

ANALYSE DES CHARGES MACHINE

NUM. REF.	NOMBRE DE PIECES	TEMPS TOTAL (EN HEURES) DE			GAMMES DE FABRICATION								
		CHARGE	REGLAGE	TRAVAIL									
1	2				FRAISAGE	STABILI.	RECT.P.	FRAIS.C.	AJUSTAGE	T.THERM.	RECT.PP.		
2	2				RABOTAGE	RECT.P.	AJUSTAGE		T.THERM.	RECT.P.			
3	2				RABOTAGE	FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.				
4	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
5	2				FRAISAGE	STABILI.	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.		
6	2				FRAISAGE	STABILI.	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.		
7	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
8	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
9	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
10	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
11	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
12	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
13	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
14	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
15	2				FRAISAGE	STABILI.	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	ELEC.E.	
	2					AJUSTAGE	T.THERM.	RECT.PP.					
	2				RABOTAGE								
63	2				RABOTAGE	RECT.P.							
64	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
65	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
66	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
67	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
68	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
69	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
70	2				FRAISAGE	AJUSTAGE	T.THERM.	RECT.PP.					
TOT.	140												

Figure-IV-6-a: Analyse des charges machines par références.



1-b) pour chaque poste de travail:

- le code de ce poste
- sa description
- le nombre total de références où figure ce poste
- le nombre total de pièces passant par le poste
- les temps totaux de charge, de réglage et de travail.

NUM. REF.	COD. POST. TRAVAIL	DESCRIPTION DES POSTES DE TRAVAIL	TEMPS TOTAL (EN HEURES) DE		
			TOTAL OPERATION	TOTAL PIECES	CHARGE REGLAGE TRAVAIL
1	FRAISAGE		60	120	
2	STABILI.		6	12	
3	RECT.P.		27	54	
4	FRAIS.C.		2	4	
5	AJUSTAGE		66	132	
6	T.THERM.		70	140	
7	RECT.PP.		65	130	
8	RABOTAGE		21	42	
9			3	6	
10	FRAIS.P.		12	24	
11	RECT.IE.		3	6	
12	ELEC.E.		1	2	
13	DEBIT		2	4	
TOTAL			338	676	

Figure-IV-6-b: Analyse des charges machines par postes de travail.

L'analyse de ces tableaux montre que pour l'échantillon considéré, la réalisation des 70 références a nécessité l'utilisation de 13 postes de travail. Enfin, 18 gammes-types peuvent être dénombrées.

2) spectre des références étudiées:

Ces spectres sont donnés:

- 2-a) en effectifs
- 2-b) en pourcentages
- 2-c) en mode binaire.

La figure-IV-7 montre le spectre de l'échantillon de 70 références de pièces tridimensionnelles sur lequel a porté cette étude.

SPECTRE DES REFERENCES EN EFFECTIF

SPECTRE DES REFERENCES EN POURCENT

MAT.SIMIL.

	0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9
1:	0	70	0	0	0	0	0	0	0	0	1:	0	100	0	0	0	0	0	0	0	0	1:	.X.....									
2:	0	0	0	0	0	0	0	69	1	0	2:	0	0	0	0	0	0	0	98	1	0	2:	.....XX.									
3:	1	9	6	0	0	29	18	3	3	1	3:	1	12	8	0	0	41	25	4	4	1	3:	XXX..XXXX									
4:	0	43	3	1	18	0	3	0	0	2	4:	0	61	4	1	25	0	4	0	0	2	4:	.XXXX.X..X									
5:	0	1	5	64	0	0	0	0	0	0	5:	0	1	7	91	0	0	0	0	0	0	5:	.XXX.....									
6:	5	16	0	0	45	0	0	1	3	0	6:	7	22	0	0	64	0	0	1	4	0	6:	XX..X..XX.									
7:	11	41	2	4	2	5	1	1	3	0	7:	15	58	2	5	2	7	1	1	4	0	7:	XXXXXXXXXX.									
8:	0	6	26	27	9	2	0	0	0	0	8:	0	8	37	38	12	2	0	0	0	0	8:	.XXXXX....									
9:	13	7	11	5	1	8	1	8	10	6	9:	18	10	15	7	1	11	1	11	14	8	9:	XXXXXXXXXXXX									
10:	1	40	25	4	0	0	0	0	0	0	10:	1	57	35	5	0	0	0	0	0	0	10:	XXXX.....									
11:	4	9	7	5	13	10	4	6	4	8	11:	5	12	10	7	18	14	5	8	5	11	11:	XXXXXXXXXXXX									
12:	3	57	10	0	0	0	0	0	0	0	12:	4	81	14	0	0	0	0	0	0	0	12:	XXX.....									
13:	6	10	8	8	16	7	6	3	1	5	13:	8	14	11	11	22	10	8	4	1	7	13:	XXXXXXXXXXXX									
14:	0	0	0	0	0	0	0	0	0	70	14:	0	0	0	0	0	0	0	0	0	100	14:	.....X									
15:	0	0	0	68	0	0	2	0	0	0	15:	0	0	0	97	0	0	2	0	0	0	15:	...X..X...									
16:	22	31	9	7	1	0	0	0	0	0	16:	31	44	12	10	1	0	0	0	0	0	16:	XXXX.....									
17:	0	1	0	0	69	0	0	0	0	0	17:	0	1	0	0	98	0	0	0	0	0	17:	.X..X.....									
18:	52	15	2	1	0	0	0	0	0	0	18:	74	21	2	1	0	0	0	0	0	0	18:	XXXX.....									
19:	0	0	0	25	0	3	1	26	4	11	19:	0	0	0	35	0	4	1	37	5	15	19:	...X.XXXXX									

Figure-IV-7: Spectre des codes Multi-M à analyser

3) liste des codes correspondant aux références étudiées.

b) MENU2.PTG

Si l'utilisateur souhaite poursuivre l'analyse, celle-ci n'est possible que pour un type de pièces à la fois ou plusieurs types de pièces à l'intérieur d'un même groupe. Le choix s'effectue à l'aide du MENU1.PTG.

Après cette sélection, le MENU2.PTG apparaît (figure-IV-8).

---

CHOIX DE L'ANALYSE

- 1)- TRANSCODAGE DU CODE MULTI-M
- 2)- PREPARATION POUR LA CLASSIFICATION DES GAMMES
- 3)- PREPARATION POUR LA CLASSIFICATION DES CODES
- 4)- FIN DU TRAVAIL

QUEL EST VOTRE CHOIX ?

---

Figure-IV-8: MENU2.PTG

Ce menu permet en particulier d'effectuer une transformation des positions des codes des références choisies, en variables indépendantes, et de visualiser ces variables sous forme d'histogrammes. La figure-IV-9 montre l'histogramme relatif à la variable Dimension principale:

correspondant aux 19 premières positions, intervient en Analyse des Données.

L'examen des 19 positions de ce code (P1,P2,...,P19), montre que n'existent que 13 variables (V1,V2,...,V13) définies dans la figure-IV-10. En effet, la première position est toujours égale à 1, et les dimensions principale, secondaire et complémentaire, la fonction et la matière sont codées sur deux positions chacune.

Variable	Signification	Description
V1	P2	configuration globale
V2	P3	caractéristique morphologique 1
V3	P4	caractéristique morphologique 2
V4	P5	usinages secondaires
V5	P6 x10+P7	fonction
V6	P8 x10+P9	dimension caractéristique 1
V7	P10x10+P11	dimension caractéristique 2
V8	P12x10+P13	dimension caractéristique 3
V9	P14	tolérances
V10	P15x10+P16	matière
V11	P17	origine
V12	P18	importance des lots
V13	P19	complexité

Figure-IV-10: Variables correspondant aux 19 positions du code principal du code Multi-M.

Il est à noter que bien que certaines caractéristiques décrites soient de nature quantitative (par exemple dimensions ou quantités produites), les variables correspondantes sont

HISTOGRAMME DE VARIABLE INDEPENDANTE

LIBELLE DE CETTE VARIABLE INDEPENDANTE NON COMMUNIQUE

DESCRIPTION DES MODALITES	REF.	EFF.	POURCENTAGE DES PIECES				
			10%	30%	50%	70%	90%
34	13	1 *					
35	14	1 *					
36	17	4 ***					
37	20	1 *					
38	21	2 **					
39	22	6 *****					
40	25	2 **					
41	27	1 *					
42	28	8 *****					
43	29	6 *****					
44	30	11 *****					
45	31	1 *					
46	32	5 ****					
47	33	4 ***					
48	35	2 **					
49	37	2 **					
50	38	2 **					
51	40	1 *					
52	41	3 ***					
53	45	3 ***					
54	46	1 *					
55	47	1 *					
56	51	1 *					
57	55	1 *					

Figure-IV-9: Histogramme relatif à la variable indépendante Dimension principale.

b-1) Transcodage du code Multi-M [46, 47]

Le code Multi-M étant composé de deux parties, le code principal et le code spécifique, seul le code principal,



qualitatives nominales, les modalités étant obtenues après découpage de l'échelle de mesure en un certain nombre d'intervalles.

D'autre part, comme cela a été mentionné au paragraphe IV-1-1, plusieurs variables dépendent de la variable V1 (configuration globale), car suivant la modalité prise par cette variable, les 12 autres ont des significations totalement différentes. Par exemple, pour les pièces de révolution, V6 et V8 définissent la valeur des diamètres principal et intérieur, alors que pour les autres types de pièces, elles désignent la longueur principale et l'épaisseur.

Or pour utiliser les méthodes d'Analyse de Données, les variables décrivant les objets à traiter doivent être totalement indépendantes. Il a donc été nécessaire de mettre au point une mise en forme de ces variables que nous avons appelée transcodage (voir Chapitre III) et schématisée sur la figure-IV-11:

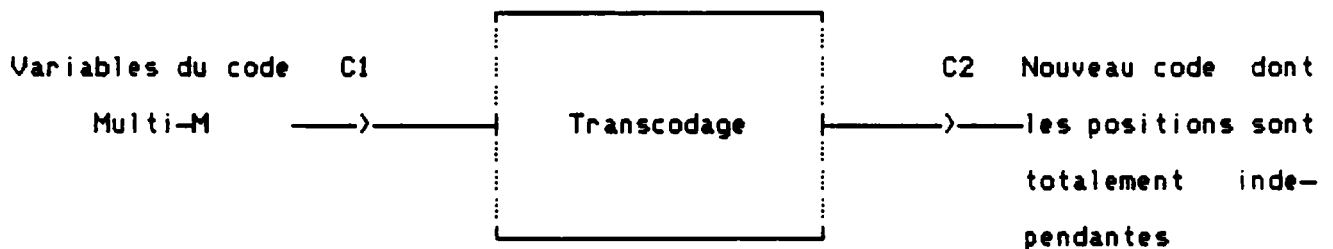


Figure-IV-11: Schéma de principe d'un transcodage.

Comme cela est exposé au paragraphe IV-1-1, le code Multi-M décrit trois types de pièces:

- les pièces de révolution, parmi lesquelles il faut distinguer les pièces rondes spéciales (ressorts...).

-les pièces de tôleries

-les pièces tridimensionnelles.

Un transcodage a été développé pour chacune de ces quatre familles (voir annexe 5).

b-2) Matrice de distances entre codes: établissement du  
Tableau Disjonctif Complet

Les différentes variables obtenues après transcodage sont des variables qualitatives nominales, et l'établissement du tableau disjonctif complet (nombre d'objets X nombre de modalités) est nécessaire à leur analyse (voir chapitre III).

Ce tableau peut être réalisé sur l'ensemble des variables comme il peut l'être sur un certain nombre de variables sélectionnées par l'utilisateur.

Cependant, ces variables peuvent correspondre à un nombre total  $t$  de modalités très important, supérieur à 600. La dimension de l'espace de représentation engendré est donc elle-même supérieure 600 ( $\mathbb{R}^1, t \geq 600$ ).

Il est donc souvent nécessaire de réduire le tableau disjonctif complet.

Le tracé d'histogramme pour chaque variable (figure-IV-12-a) donne une première idée de la dispersion des modalités décrivant pour cette variable l'ensemble des pièces à analyser. Il permet d'éliminer les modalités qui ne sont jamais prises, de déceler

celles qui véhiculent le moins d'information et de regrouper celles qui sont les plus proches.

Pour certaines variables, telles que les dimensions ou la matière, le regroupement peut se faire de façon purement logique. Pour d'autres variables, telles que la description du diamètre extérieur, il est nécessaire de faire appel à des critères technologiques de fabrication. Dans ce cas, la connaissance de la signification des modalités est indispensable.

Les figures IV-12-a à IV-12-c montrent un exemple-type des différentes étapes du traitement des histogrammes, ainsi que l'histogramme final, où les modalités 'deux diamètres usinés' et 'trois diamètres usinés' qui correspondent à un procédé de fabrication identique ont pu être regroupées en une seule 'plusieurs diamètres usinés'.

Ce logiciel donne à l'utilisateur la possibilité de procéder au regroupement de certaines modalités. Cette opération peut être réalisée par l'intermédiaire du fichier "FLIB2.PTG". Cependant, faute de connaître le libellé des modalités du code Multi-M relatives aux pièces traitées dans cette étude, aucun regroupement n'a été effectué. C'est pourquoi cette possibilité, bien qu'existante, n'apparaît pas en toutes lettres dans le MENU2.PTG.

La matrice de distances entre codes est obtenue à partir du tableau disjonctif complet. L'utilisateur a la possibilité de choisir une mesure de ressemblance à l'aide du MENU3.PTG

Le résultat de ce calcul est stocké dans le fichier 'DISTAN.COD'.



HISTOGRAMME DE VARIABLE INDEPENDANTE (a)

LIBELLE DE CETTE VARIABLE INDEPENDANTE NON COMMUNIQUE

DESCRIPTION DES MODALITES	REF. EFF.	POURCENTAGE DES PIECES				
		10%	30%	50%	70%	90%
0	0					
1	1					
2	2					
3	3 25	*****				
4	4					
5	5 3	***				
6	6 1	*				
7	7 26	*****				
8	8 4	***				
9	9 11	*****				

HISTOGRAMME DE VARIABLE INDEPENDANTE (b)

LIBELLE DE CETTE VARIABLE INDEPENDANTE NON COMMUNIQUE

DESCRIPTION DES MODALITES	REF. EFF.	POURCENTAGE DES PIECES				
		10%	30%	50%	70%	90%
0	3 25	*****				
1	5 3	***				
2	6 1	*				
3	7 26	*****				
4	8 4	***				
5	9 11	*****				

HISTOGRAMME DE VARIABLE INDEPENDANTE (c)

LIBELLE DE CETTE VARIABLE INDEPENDANTE NON COMMUNIQUE

DESCRIPTION DES MODALITES	REF. EFF.	POURCENTAGE DES PIECES				
		10%	30%	50%	70%	90%
0	3 25	*****				
1	5 4	***				
2	7 26	*****				
3	9 15	*****				

<-----5+6

Figure-IV-12: Exemple de regroupement de modalités.

---

CHOIX DES DISTANCES

- 1)- DISTANCE EUCLIDIENNE
- 2)- INDICE DE PROXIMITE DE JACCARD
- 3)- INDICE DE PROXIMITE DE DICE ET CZEKANOVSKI
- 4)- INDICE DE PROXIMITE DE SOKAL ET MICHENER
- 5)- INDICE DE PROXIMITE DE RUSSEL ET RAO
- 6)- INDICE DE PROXIMITE DE KULCZINSKI
- 7)- INDICE DE PROXIMITE DE SOKAL ET SNEATH
- 8)- INDICE DE PROXIMITE DE ROGERS ET TANIMOTO
- 9)- DISTANCE DU KHI-DEUX
- 0)- FIN DU TRAVAIL

QUEL EST VOTRE CHOIX ?

---

Figure-IV-13: Mesures de ressemblance.

b-3)Matrice de distances entre gammes

Le calcul de proximité entre gammes de fabrication est effectué après sélection par l'utilisateur de la mesure de ressemblance de son choix à l'aide du MENU3.PTG.

Le résultat du calcul est stocké dans le fichier "DISTAN.GAM".

Rappelons que, dans cette étude, la distance de Dice-Czekanowski a été choisie et pour les codes et pour les gammes.

L'ensemble des résultats relatifs à l'analyse de l'échantillon et à la préparation des données qui ont été décrits ci-dessus est regroupé dans le fichier "PTG.RES".

IV-2/ REPRESENTATION GRAPHIQUE DES REFERENCES ANALYSEES PAR  
ACPTD

Après constitution des tableaux de distances entre codes [C] et entre gammes [G] au moyen de la procédure "PTG", la représentation graphique des données est obtenue à l'aide de la procédure "ACP".

Le MENU1.ACP (figure-IV-14) permet de choisir la nature des données à analyser :

---

ANALYSE EN COMPOSANTES PRINCIPALES

- (1)- SUR LE TABLEAU DE DISTANCES ENTRE CODES SEULS
- (2)- SUR LE TABLEAU DE DISTANCES ENTRE GAMMES SEULS
- (3)- SUR LE TABLEAU DE DISTANCES ENTRE CODES + GAMMES

QUEL EST VOTRE CHOIX ?

---

Figure-IV-14: Choix des données à analyser

En effet, il est possible de faire porter l'analyse uniquement sur les codes ou sur les gammes, ou encore sur les références complètes (codes plus gammes).

Dans ce dernier cas, le MENU2.ACP (figure-IV-15) permet la constitution de la matrice de distances [R] qui est fonction des tableaux [C] et [G] selon la relation:

$$[R] = \frac{\alpha_C}{\alpha_C + \alpha_G} [C] + \frac{\alpha_G}{\alpha_C + \alpha_G} [G]$$

---

ANALYSE EN COMPOSANTES PRINCIPALES

POIDS DES DONNEES CODES ?

POIDS DES DONNEES GAMMES ?

QUEL EST VOTRE CHOIX ?

---

Figure-IV-15: Constitution de la matrice de distances

Ensuite, le choix du plan factoriel sur lequel vont être projetées les données, ainsi que des dimensions de la figure est effectué au moyen du MENU3.ACP (figure-IV-16):

---

ANALYSE EN COMPOSANTES PRINCIPALES

QUEL AXE FACTORIEL CHOISISSEZ VOUS COMME AXE DES X ?

QUEL AXE FACTORIEL CHOISISSEZ VOUS COMME AXE DES Y ?

DONNEZ LE NOMBRE DE PAGES GRAPHIQUES ( 1 OU 2 ) :

DONNEZ LE NOMBRE DE LIGNES ( ENTRE 10 ET 120 ) :

QUEL EST VOTRE CHOIX ?

---

Figure-IV-16: Choix du format de la représentation graphique

La figure-IV-17 montre la projection des codes sur le premier plan factoriel, avec un taux d'explication de près de 55%.

La figure-IV-18 présente celle des gammes sur le premier plan factoriel, avec un taux d'explication de près de 95% sur les deux premiers axes.

Cette différence entre les deux taux d'explication provient du fait que plusieurs références peuvent avoir la même gamme de fabrication (18 gammes-types pour 70 références), ce qui est conforme à la réalité industrielle: le nombre de références est toujours beaucoup plus important que le nombre de gammes.

ANALYSE EN COMPOSANTES PRINCIPALES SUR UN TABLEAU DES DISTANCES

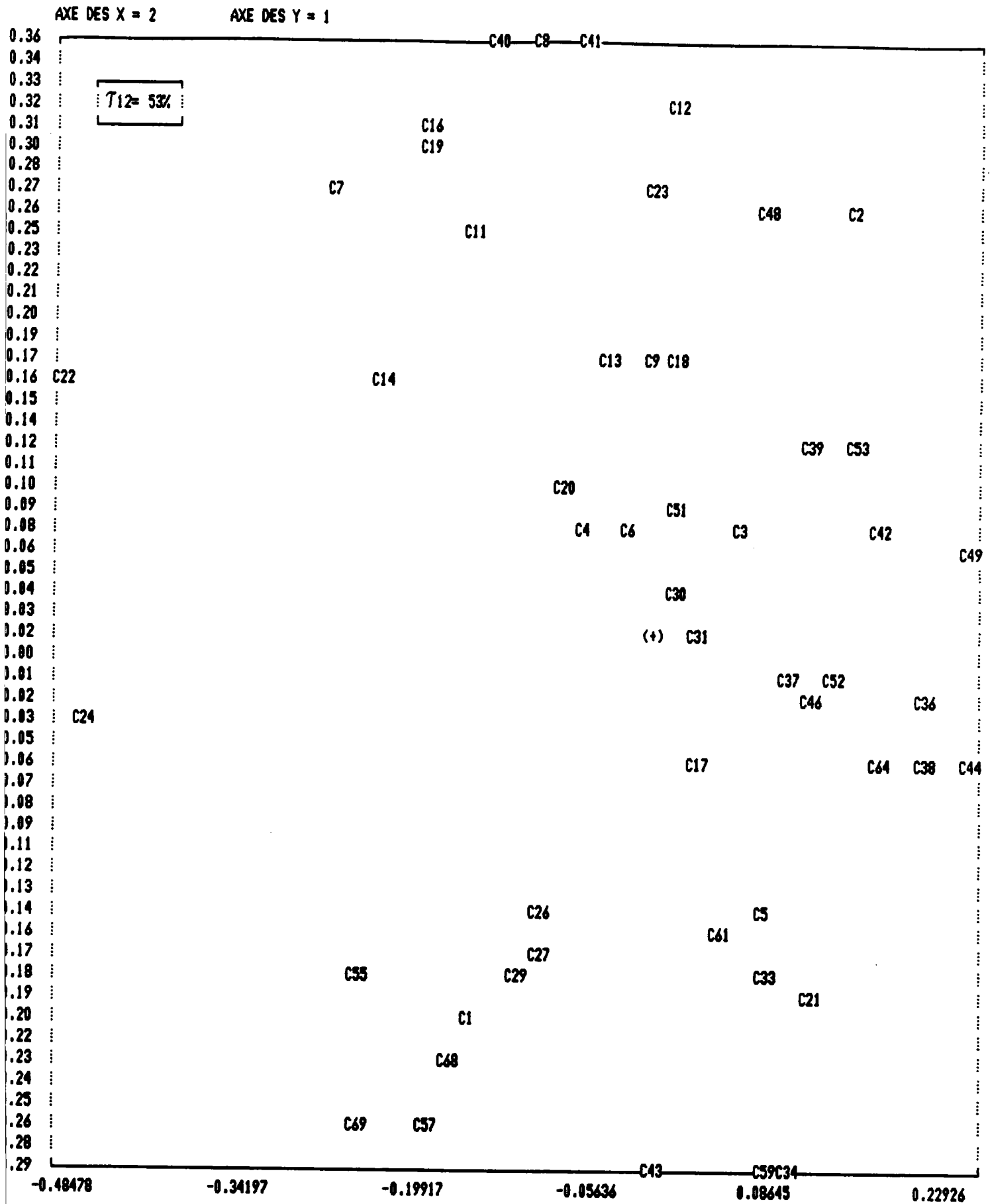


Figure-IV-17: Représentation graphique des codes par ACPTD

POINTS MULTIPLES :

- 1 : (C9 ,C10)
- 2 : (C13,C15)
- 3 : (C53,C54)
- 4 : (C42,C45)
- 5 : (C42,C47)
- 6 : (C42,C65)
- 7 : (C42,C66)
- 8 : (C49,C50)
- 9 : (C49,C67)
- 10 : (C31,C32)

- 11 : (C24,C25)
- 12 : (C44,C63)
- 13 : (C61,C62)
- 14 : (C27,C28)
- 15 : (C33,C35)
- 16 : (C55,C56)
- 17 : (C68,C70)
- 18 : (C57,C58)
- 19 : (C59,C60)

VALEURS EXTREMES	-0.48478E+00	0.24711E+00
VALEURS EXTREMES	-0.30123E+00	0.35609E+00

ANALYSE EN COMPOSANTES PRINCIPALES SUR UN TABLEAU DES DISTANCES

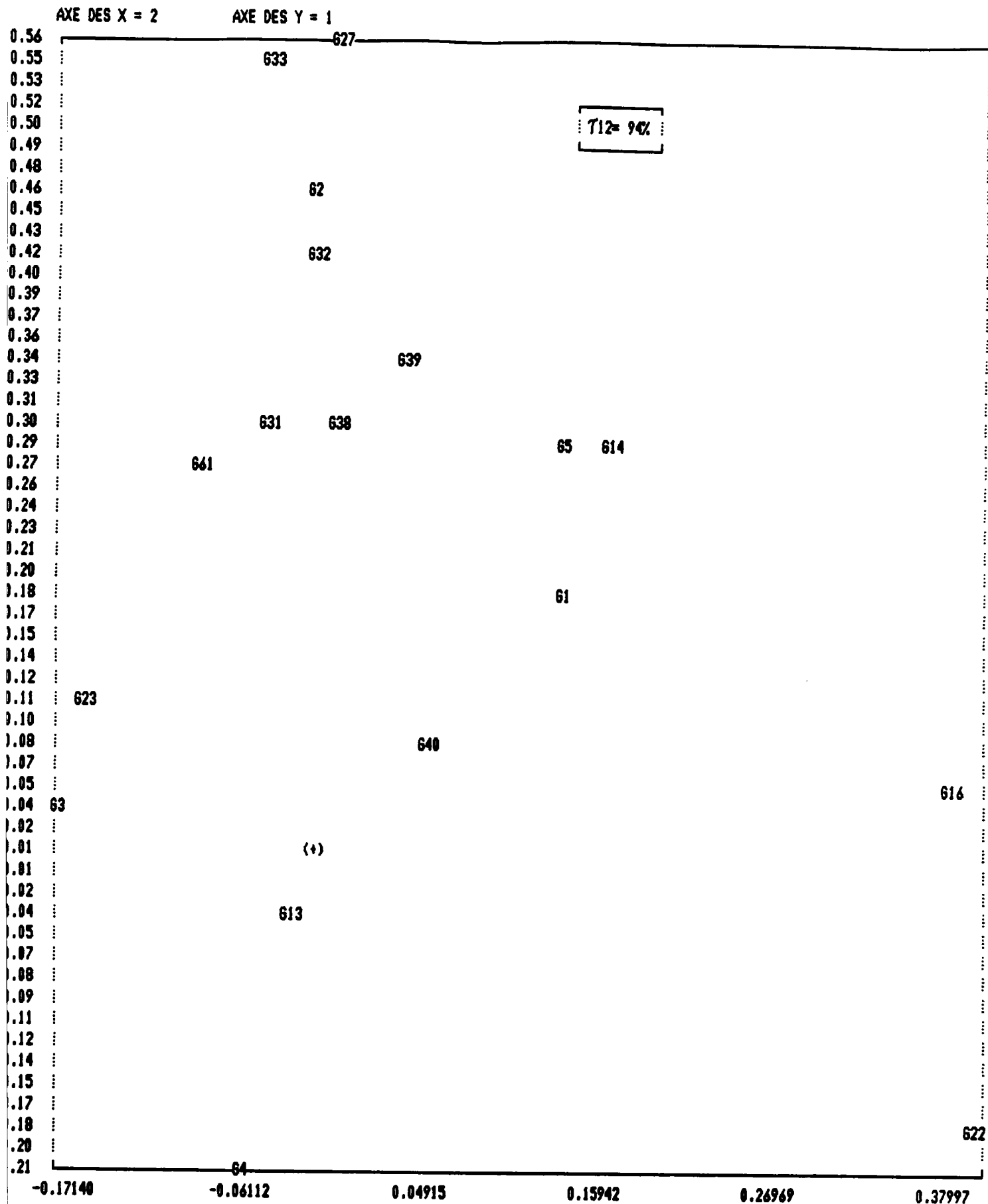


Figure-IV-18: Représentation graphique des gammes par ACPTD



- 1 : (627,628)  
 2 : (633,634)  
 3 : (633,635)  
 4 : (633,646)  
 5 : (633,655)  
 6 : (633,656)  
 7 : (633,657)  
 8 : (633,658)  
 9 : (65 ,66 )  
 10 : (661,662)
- 11 : (640,641)  
 12 : (63 ,617)  
 13 : (63 ,629)  
 14 : (63 ,630)  
 15 : (622,624)  
 16 : (622,625)  
 17 : (622,637)  
 18 : (622,647)  
 19 : (64 ,67 )  
 20 : (64 ,68 )
- 21 : (64 ,69 )  
 22 : (64 ,610)  
 23 : (64 ,611)  
 24 : (64 ,612)  
 25 : (64 ,615)  
 26 : (64 ,618)  
 27 : (64 ,619)  
 28 : (64 ,620)  
 29 : (64 ,621)  
 30 : (64 ,626)
- 31 : (64 ,636)  
 32 : (64 ,642)  
 33 : (64 ,643)  
 34 : (64 ,644)  
 35 : (64 ,645)  
 36 : (64 ,648)  
 37 : (64 ,649)  
 38 : (64 ,650)  
 39 : (64 ,651)  
 40 : (64 ,652)
- 41 : (64 ,653)  
 42 : (64 ,654)  
 43 : (64 ,659)  
 44 : (64 ,660)  
 45 : (64 ,663)  
 46 : (64 ,664)  
 47 : (64 ,665)  
 48 : (64 ,666)  
 49 : (64 ,667)  
 50 : (64 ,668)
- 51 : (64 ,669)  
 52 : (64 ,670)

VALEURS EXTREMES -0.17140E+00 0.39375E+00  
 VALEURS EXTREMES -0.22623E+00 0.56287E+00

La comparaison de ces deux figures peut permettre à l'utilisateur de mettre en évidence soit d'éventuelles erreurs de codification, soit des gammes de "remplacement", élaborées pour faire face à un surcroît de travail ou du fait de la panne d'une machine, mais qui n'ont plus de raison d'être. Dans cette étude, toutes les gammes ont été considérées comme justes.

La comparaison entre les figures IV-17 et IV-18 montre que, de façon générale, il n'a pas de bijection claire entre les données codes et les données gammes.

Or, le classement des pièces a lieu avec le code seul, et la nécessité de faire intervenir dans la classification à la fois les codes et les gammes apparaît sans équivoque. Ceci est réalisé dans cette étude en attribuant un poids égal aux codes et aux gammes ( $\alpha_c = \alpha_g = 1$ ).

La figure-IV-19 montre la projection des 70 références de l'échantillon étudié sur le plan défini sur les deux premiers axes factoriels. Le taux d'explication sur ces deux axes est de 77%, ce qui est satisfaisant.

Sur cette figure, il est possible de distinguer 4 groupes:

ANALYSE EN COMPOSANTES PRINCIPALES SUR UN TABLEAU DES DISTANCES

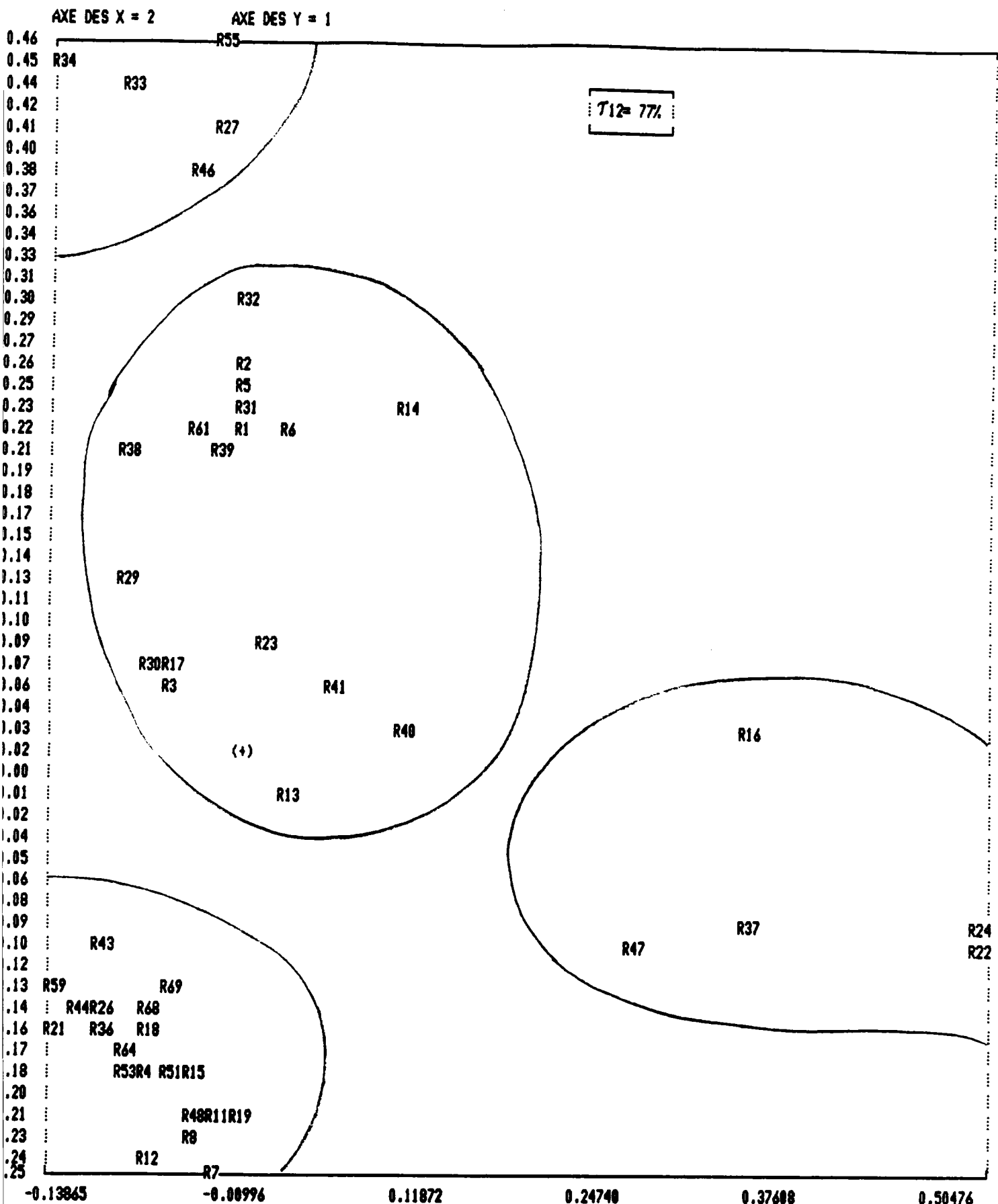


Figure-IV-19: Représentation graphique des références par ACPTD

POINTS MULTIPLES :

- 1 : (R55,R56)
- 2 : (R55,R57)
- 3 : (R55,R58)
- 4 : (R33,R35)
- 5 : (R27,R28)
- 6 : (R61,R62)
- 7 : (R24,R25)
- 8 : (R59,R60)
- 9 : (R44,R63)
- 10 : (R68,R70)

- 11 : (R36,R49)
- 12 : (R36,R50)
- 13 : (R36,R67)
- 14 : (R4 ,R9 )
- 15 : (R4 ,R10)
- 16 : (R15,R20)
- 17 : (R4 ,R42)
- 18 : (R4 ,R45)
- 19 : (R4 ,R52)
- 20 : (R53,R54)

- 21 : (R4 ,R65)
- 22 : (R4 ,R66)

VALEURS EXTREMES	-0.13845E+00	0.52085E+00
VALEURS EXTREMES	-0.26602E+00	0.46360E+00

C'est ce nombre de groupes qui a été choisi en pratique pour initialiser le nombre de classes dans la procédure AND, utilisée pour déterminer les familles de références.

Du fait de la fidélité de la description des références par l'ACPTD (taux d'explication supérieur à 70% sur le premier plan factoriel), les noyaux initiaux pour chacun des traitements AND sont sélectionnés à l'intérieur des 4 groupes schématisés sur la figure-IV-19.

L'ensemble des figures IV-17 à IV-19 est regroupé dans le fichier "ACP.RES".

### IV-3/ CLASSIFICATION AUTOMATIQUE ET CLASSEMENT PAR LES PROCEDURES AND ET CAH

#### IV-3-A/ Classification automatique

De la même façon que l'analyse par ACPTD, l'algorithme des Nuées Dynamiques (procédure "AND") peut s'appliquer (figure-IV-20) soit à l'étude des codes, soit à celle des gammes ou encore simultanément aux codes et aux gammes. Dans ce dernier cas, le poids respectif de ces deux types de données doit être choisi par l'utilisateur.

Celui-ci doit également préciser le nombre de tirages (<4) qu'il désire effectuer, le nombre de classes et les noyaux pour chaque tirage (figure-IV-21).

---

CLASSIFICATION AUTOMATIQUE PAR L'AND

- (1)- SUR LE TABLEAU DE DISTANCES ENTRE CODES SEULS
- (2)- SUR LE TABLEAU DE DISTANCES ENTRE GAMMES SEULS
- (3)- SUR LE TABLEAU DE DISTANCES ENTRE CODES + GAMMES

QUEL EST VOTRE CHOIX ?

---

Figure-IV-20: Choix des données à analyser par la procédure AND.

---

```
123456789 123456789 123456789 123456789 123456789 123456789 123456789 1234567890!  
3 <--- nombre de tirages  
4 <--- nombre de classes pour le premier tirage  
  <--  
  |  
  | les noyaux  
  <--  
4 <--- nombre de classes pour le deuxième tirage  
  <--  
  |  
  | les noyaux  
  <--  
4 <--- nombre de classes pour le troisième tirage  
  <--  
  |  
  | les noyaux  
  <--
```

---

Figure-IV-21: Initialisation de la procédure AND.

A l'issue de ces tirages, les formes fortes, c'est à dire les classes invariantes au cours des différents tirages sont calculées et les résultats stockés dans le fichier "FFFB.AND". L'analyse des charges machines, les éléments nécessaires à l'édition de la gamme mère et le spectre correspondant à chacune des familles obtenues sont stockés dans le fichier "AND.RES".

Le nombre de formes fortes est généralement beaucoup plus élevé que le nombre de familles souhaité par l'utilisateur. Si c'est le cas, il peut faire appel à la procédure "CAH" qui effectue une classification hiérarchique sur les formes fortes. Le menu "MENU.CAH" lui permet de préciser le nombre de classes désiré. Le fichier "CAH.RES" regroupe pour les familles obtenues les mêmes informations que le fichier "AND.RES" décrit ci-dessus.

En ce qui concerne l'échantillon des 70 références étudiées, 10 formes fortes ont été mises en évidence (figure-IV-22). Le détail des classes est donné à l'annexe 7.

Ces formes fortes sont très homogènes, comme le montre l'analyse des charges machines pour la forme forte n°1 (figure-IV-23) par exemple.

En comparant les figures IV-22 et IV-19, on s'aperçoit que trois des groupes apparents à la suite de la description des données par l'ACPTD sont restés pratiquement inchangés, alors que le groupe n°2 sur la figure-IV-19 s'avère non homogène.

Si un nombre de classes inférieur à celui des formes fortes est désiré, cela peut être réalisé en coupant l'arbre obtenu après classification hiérarchique sur les noyaux des dix formes

ANALYSE EN COMPOSANTES PRINCIPALES SUR UN TABLEAU DES DISTANCES

AXE DES X = 2

AXE DES Y = 1

T12 = 77%

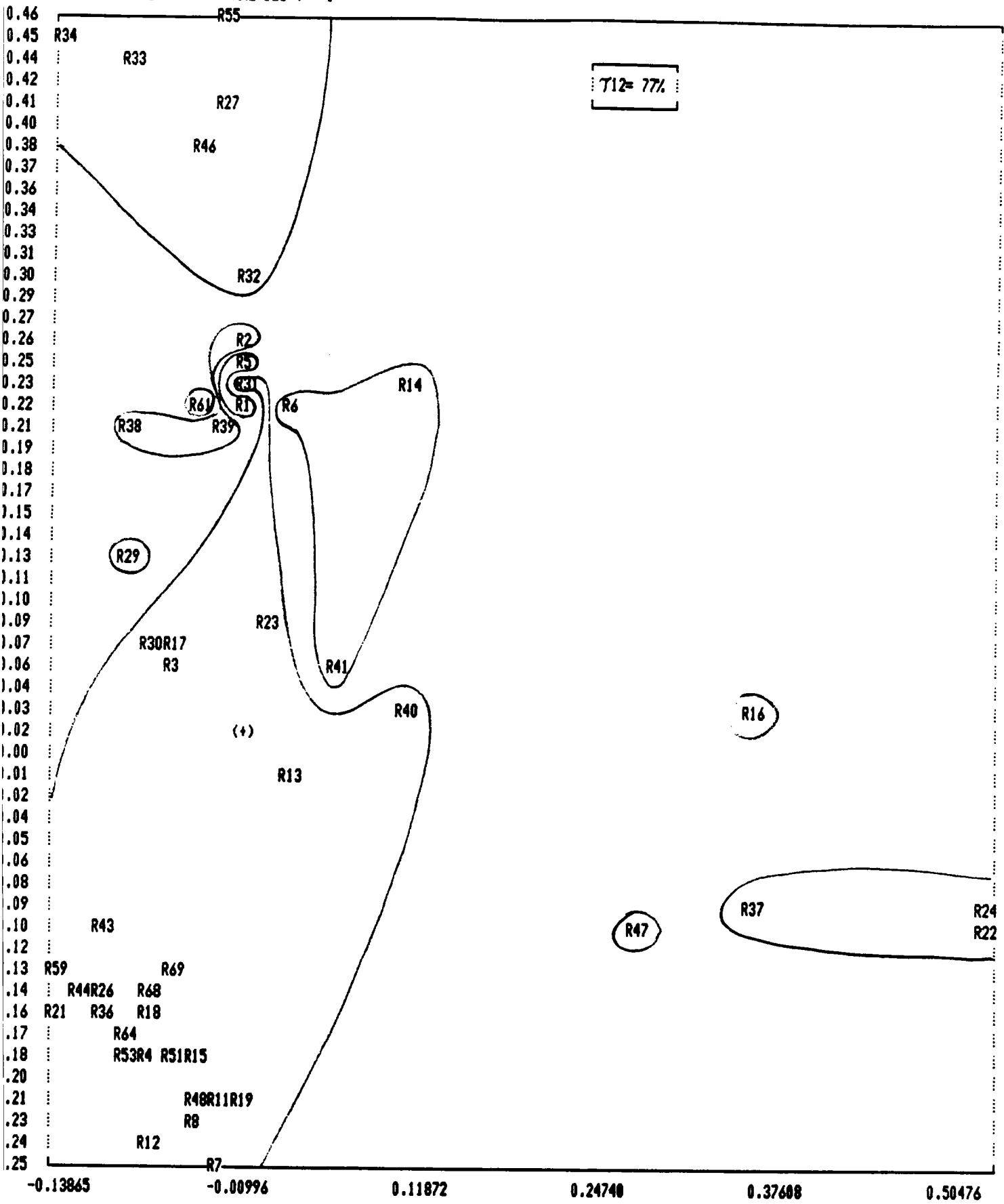


Figure-IV-22: Résultats de la classification par la procédure AND.



POINTS MULTIPLES SONT:

- 1 : (R55,R56)
- 2 : (R55,R57)
- 3 : (R55,R58)
- 4 : (R33,R35)
- 5 : (R27,R28)
- 6 : (R61,R62)
- 7 : (R24,R25)
- 8 : (R59,R60)
- 9 : (R44,R43)
- 10 : (R68,R70)

- 11 : (R36,R49)
- 12 : (R36,R50)
- 13 : (R36,R67)
- 14 : (R4 ,R9 )
- 15 : (R4 ,R10)
- 16 : (R15,R20)
- 17 : (R4 ,R42)
- 18 : (R4 ,R45)
- 19 : (R4 ,R52)
- 20 : (R53,R54)

- 21 : (R4 ,R65)
- 22 : (R4 ,R66)

VALEURS EXTREMES	-0.13865E+00	0.52085E+00
VALEURS EXTREMES	-0.26602E+00	0.46360E+00

ANALYSE DES CHARGES MACHINE

NUM. REF.	NOMBRE DE PIECES	TEMPS TOTAL (EN HEURES) DE			GAMES DE FABRICATION						
		CHARGE	REGLAGE	TRAVAIL	RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.P.	RECT.PP.
1	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.P.	
2	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.P.	
3	2				DEBIT	RABOTAGE	RECT.P.	FRAIS.C.	AJUSTAGE	T.THERM.	RECT.PP.
4	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
5	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
6	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
7	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
8	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
9	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
10	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
11	2				RABOTAGE	RECT.P.	FRAIS.P.	AJUSTAGE	T.THERM.	RECT.PP.	
TOT.	22										

NUM. REF.	COD. POST. TRAVAIL	DESCRIPTION DES POSTES DE TRAVAIL	TOTAL OPERATION	TOTAL PIECES	TEMPS TOTAL (EN HEURES) DE		
					CHARGE	REGLAGE	TRAVAIL
1	RABOTAGE		11	22			
2	RECT.P.		13	26			
3	FRAIS.P.		10	20			
4	AJUSTAGE		10	20			
5	T.THERM.		12	24			
6	DEBIT		1	2			
7	FRAIS.C.		1	2			
8	RECT.PP.		9	18			
TOTAL			67	134			

SPECTRE DES CODES MULTI-M A ANALYSER

SPECTRE DES REFERENCES EN EFFECTIF

SPECTRE DES REFERENCES EN POURCENT

MAT.SIMIL.

	0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9		0123456789
1:	0	11	0	0	0	0	0	0	0	0	1:	0	100	0	0	0	0	0	0	0	0	1:	.X.....
2:	0	0	0	0	0	0	0	11	0	0	2:	0	0	0	0	0	0	100	0	0	0	2:	.....X..
3:	0	4	1	0	0	2	3	0	1	0	3:	0	34	9	0	0	18	27	0	9	0	3:	.XX..XX.X.
4:	0	2	0	0	7	0	2	0	0	0	4:	0	18	0	0	43	0	18	0	0	0	4:	.X..X.X...
5:	0	0	0	11	0	0	0	0	0	0	5:	0	0	0	100	0	0	0	0	0	0	5:	...X.....
6:	0	4	0	0	7	0	0	0	0	0	6:	0	34	0	0	43	0	0	0	0	0	6:	.X..X.....
7:	2	6	0	2	0	1	0	0	0	0	7:	18	54	0	18	0	9	0	0	0	0	7:	XX.X.X.....
8:	0	0	1	9	0	1	0	0	0	0	8:	0	0	9	81	0	9	0	0	0	0	8:	...XX.X....
9:	2	1	5	1	0	1	0	1	0	0	9:	18	9	45	9	0	9	0	9	0	0	9:	XXXX.X.X...
10:	0	2	0	1	0	0	0	0	0	0	10:	0	18	72	9	0	0	0	0	0	0	10:	...XX.....
11:	0	2	0	1	1	0	2	0	2	3	11:	0	18	0	9	9	0	18	0	18	27	11:	.X..XX.X..XX
12:	0	5	6	0	0	0	0	0	0	0	12:	0	45	54	0	0	0	0	0	0	0	12:	.XX.....
13:	0	3	0	2	0	2	2	0	0	2	13:	0	27	0	18	0	18	18	0	0	18	13:	.X.X..XX..X
14:	0	0	0	0	0	0	0	0	0	11	14:	0	0	0	0	0	0	0	0	0	100	14:	.....X
15:	0	0	0	10	0	0	1	0	0	0	15:	0	0	0	0	90	0	9	0	0	0	15:	...X..X...
16:	2	4	3	2	0	0	0	0	0	0	16:	18	36	27	18	0	0	0	0	0	0	16:	XXXX.....
17:	0	0	0	0	11	0	0	0	0	0	17:	0	0	0	0	100	0	0	0	0	0	17:	...X.....
18:	11	0	0	0	0	0	0	0	0	0	18:	100	0	0	0	0	0	0	0	0	0	18:	X.....
19:	0	0	0	5	0	2	0	3	1	0	19:	0	0	0	45	0	18	0	27	9	0	19:	...X.X..XX.

Figure-IV-23: Analyse des charges machines pour la famille n°1.

fortes, représenté à la figure-IV-24:

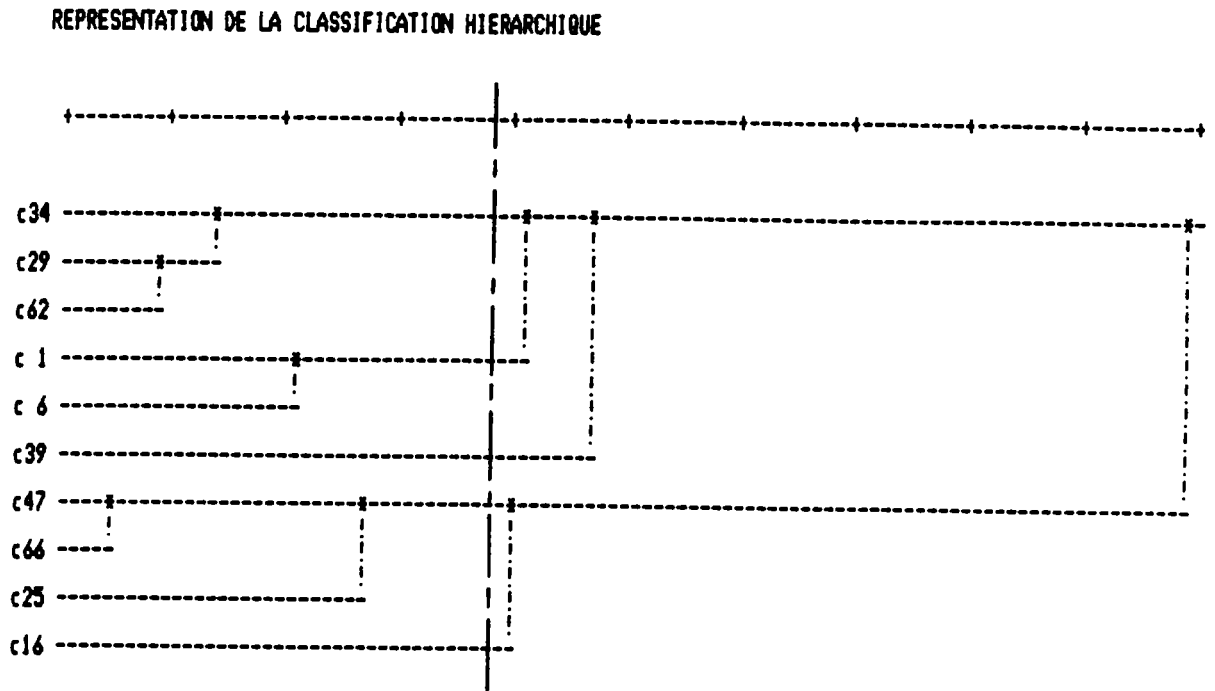


Figure-IV-24: Arbre obtenu par "CAH" sur les 10 formes fortes résultant de la classification automatique par la procédure "AND".

Pour la population étudiée, il est possible de couper l'arbre au niveau de 5 familles. L'arbre obtenu est représenté à la figure-IV-25. Le résultat est schématisé à la figure-IV-26. Le détail des familles est reporté à l'annexe 8.

REPRESENTATION DE LA CLASSIFICATION HIERARCHIQUE

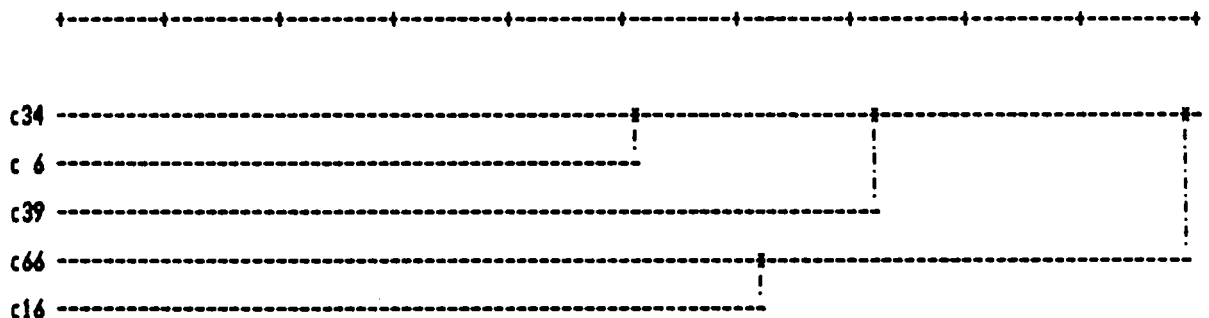


Figure-IV-25: Hiérarchie obtenue après coupure de l'arbre représenté à la figure-IV-24.

ANALYSE EN COMPOSANTES PRINCIPALES SUR UN TABLEAU DES DISTANCES

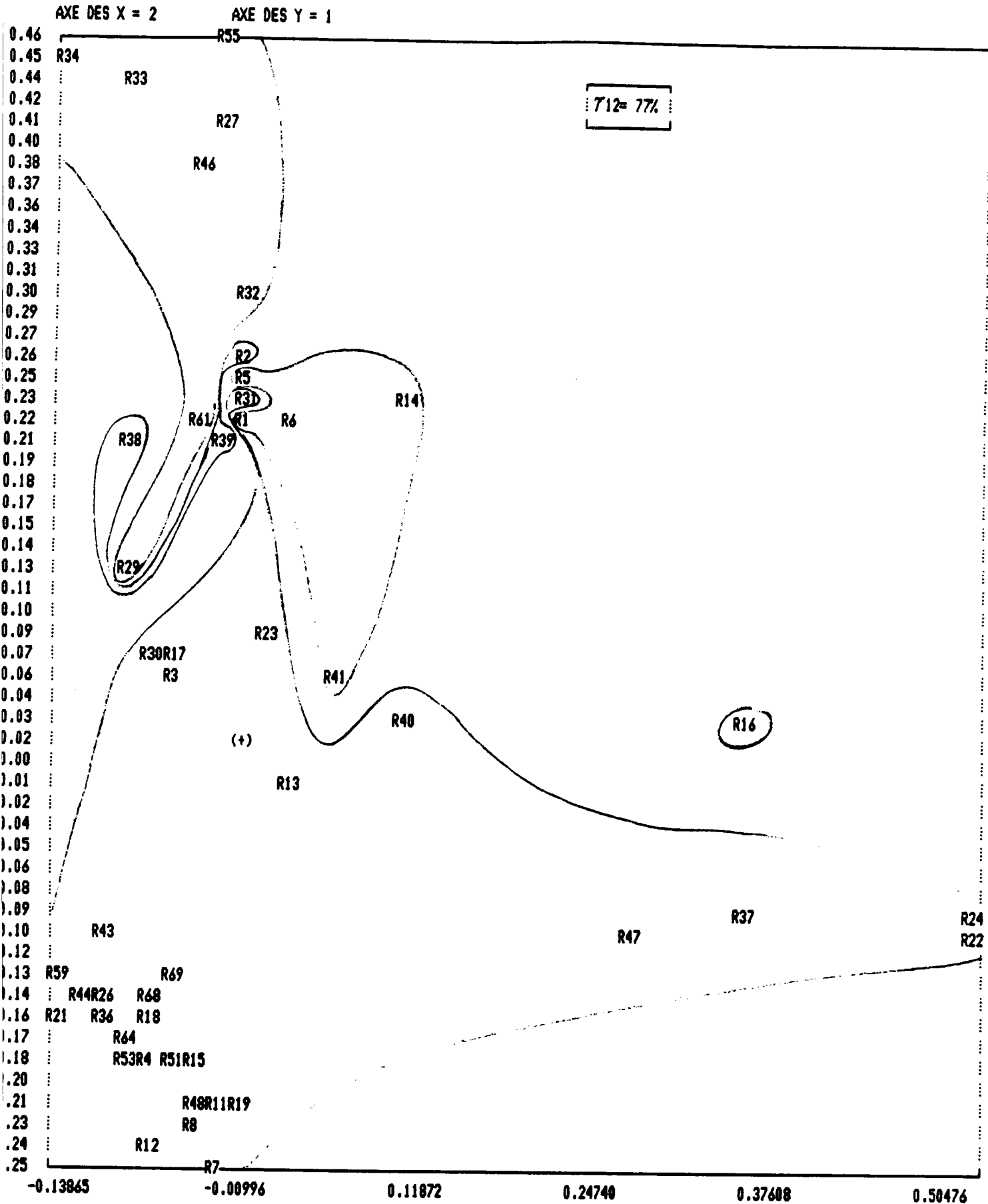


Figure-IV-26: Représentation graphique des 70 références en 5 familles.

POINTS MULTIPLES SONT :

- 1 : (R55,R56)
- 2 : (R55,R57)
- 3 : (R55,R58)
- 4 : (R33,R35)
- 5 : (R27,R28)
- 6 : (R61,R62)
- 7 : (R24,R25)
- 8 : (R59,R60)
- 9 : (R44,R63)
- 10 : (R68,R70)

- 11 : (R36,R49)
- 12 : (R36,R50)
- 13 : (R36,R67)
- 14 : (R4 ,R9 )
- 15 : (R4 ,R10)
- 16 : (R15,R20)
- 17 : (R4 ,R42)
- 18 : (R4 ,R45)
- 19 : (R4 ,R52)
- 20 : (R53,R54)

- 21 : (R4 ,R65)
- 22 : (R4 ,R66)

VALEURS EXTREMES	-0.13865E+00	0.52085E+00
VALEURS EXTREMES	-0.26602E+00	0.46360E+00

IV-3-B/ Classement et validation

Après classification, chaque famille obtenue est représentée par son spectre exprimé en pourcentages à partir des codes des éléments de la famille: le terme général de ce tableau  $s_{ij}$  est égal au pourcentage de fois où la modalité  $j$  est prise pour la position  $i$ .

A partir de ce spectre, il est possible, en faisant intervenir un seuil  $\sigma$ , de définir le spectre binaire de la famille, comme cela est montré au Chapitre III.

Le terme général de ce tableau,  $s_{ij}^{\sigma}$ , s'écrit:

-pour  $s_{ij} = 0$        $s_{ij}^{\sigma} = 0$

-pour  $s_{ij} \neq 0$        $s_{ij}^{\sigma} = 1$  si  $s_{ij} \geq \sigma$

$s_{ij}^{\sigma} = 0$  sinon

La détermination du seuil optimal est obtenue en réalisant le classement des pièces ayant servi à effectuer la classification par comparaison entre le spectre binaire de ces pièces et le spectre binaire des familles pour différentes valeurs de  $\sigma$ . Sur la figure-IV-20 est tracé le pourcentage de pièces bien classées en fonction de la valeur du seuil (classification en 10 formes fortes). D'après cette courbe, le seuil optimal est situé au alentours de 13%, valeur pour laquelle le pourcentage de pièces bien classées atteint 88%. La méthodologie mise au point pour effectuer la classification de pièces au Bureau des Méthodes est donc remarquablement performante, puisque les logiciels existants ne classent correctement que 40% des pièces ayant servi à la classification.

Si les pièces sont regroupées en 4 familles, obtenues par classification hiérarchique sur les noyaux des formes fortes, le pourcentage maximal de pièces bien classées atteint 80% pour un seuil de 10%.

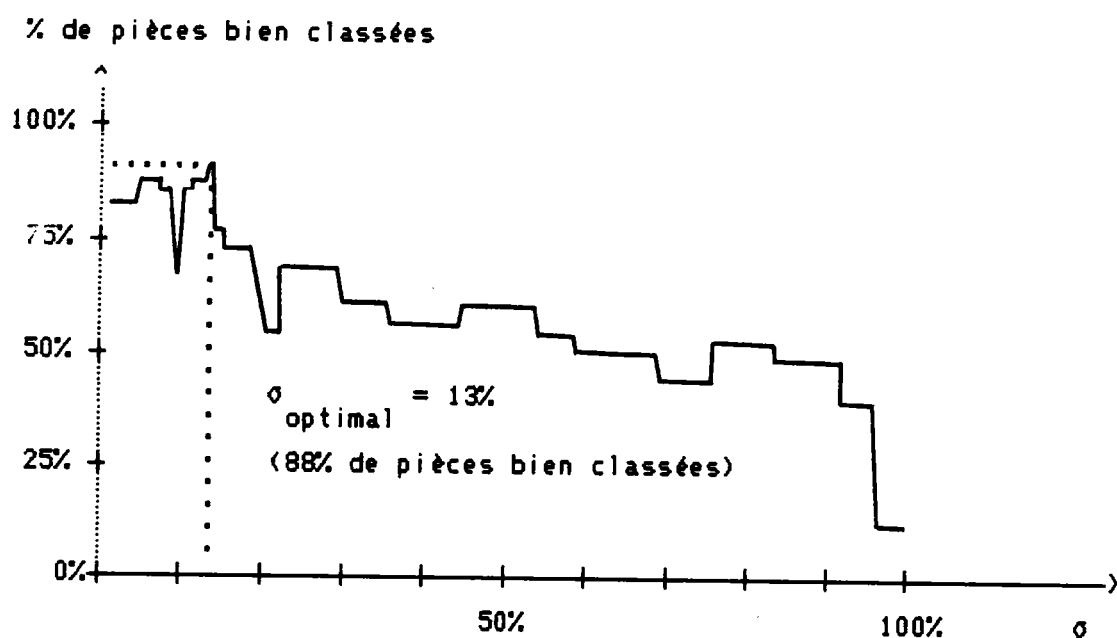


Figure-IV-20: Pourcentage de pièces bien classées en fonction du seuil  $\sigma$

**CHAPITRE V**

**CONCLUSION**



Cette étude a permis de définir une démarche permettant l'utilisation des méthodes d'Analyse des Données pour la classification et le classement automatiques de données.

La méthodologie mise au point a permis de résoudre un certain nombre de difficultés, à notre connaissance insurmontées jusqu'ici:

- analyse d'un type quelconque de données, y compris les données ordinales incomplètes, telles que les gammes de fabrication par exemple,
- classification et classement sur des ensembles distincts de données,
- analyse fiable sur tous les types de codes, par l'introduction de la notion de transcodage,
- classification fidèle grâce à la meilleure identification du noyau des classes formées,
- représentation optimisée des familles permettant un meilleur classement.

Cette méthodologie a donné naissance au logiciel CL<sup>2</sup>AUDIA, qui a été appliqué avec succès à des données industrielles de fabrication, conduisant à la classification et au classement automatiques de pièces codées au bureau des méthodes. Près de 90% des pièces ayant servi à la classification ont ainsi pu être bien classées, contre 40% dans le cas des logiciels existants de Technologie de Groupe, pour lesquels l'expérience de l'utilisateur joue un rôle considérable.

A la suite de cette étude, les principales difficultés auxquelles se heurtait l'implantation de la Technologie de Groupe en milieu industriel paraissent aplanies, d'autant qu'il serait possible d'installer sur micro-ordinateur le logiciel CL<sup>2</sup>AUDIA, allégeant ainsi la structure d'accueil indispensable.

Les sources d'amélioration semblent se situer plutôt en amont et en aval de la classification automatique.

En amont, tout d'abord, s'il ressort de cette étude que l'utilisation du code MULTI-M, moyennant un transcodage pour transformer ses positions en variables indépendantes, est satisfaisante pour la classification et le classement automatiques des pièces au bureau des méthodes, une amélioration sensible de la codification pourrait être obtenue en ce qui concerne les variables quantitatives, qui représentent à titre d'exemple 6 des 19 positions du code principal du code MULTI-M.

D'une façon générale, les valeurs prises par une variable quantitative (longueur, largeur, diamètre principal, etc...) sont découpées en un certain nombre d'intervalles. Dans de nombreux codes, y compris le code MULTI-M, les bornes relatives à ces intervalles sont figées et non pas adaptées à l'entreprise utilisatrice.

Il serait tout à fait possible d'adapter ce découpage à la population à traiter, soit en formant des intervalles à effectif identique en partant par exemple de tracés d'histogrammes, soit en utilisant un algorithme déterminant lui-même le nombre d'intervalles et leurs bornes, tel que celui de Fischer [28], présenté à l'Annexe 4.

L'utilisation de telles méthodes devrait conduire à une meilleure description de la réalité par les variables quantitatives, et donc à une efficacité accrue de la codification.

En aval, enfin, le recours à des méthodes d'Analyse Discriminante pour une optimisation du classement de nouveaux objets propose une solution intéressante.

D'autre part, dans le cas de données de production, l'obtention de la gamme-mère des classes issues de la classification, qui était jusqu'ici réalisée dans la plupart des cas sous la forme d'un graphe, comme cela est indiqué au paragraphe III-6, devrait être considérablement facilitée grâce à l'utilisation d'un système expert. L'amélioration de la procédure d'obtention des gammes standards apparaît comme un facteur fondamental de la progression de la Technologie de Groupe dans le monde des PME-PMI.

Ces deux derniers thèmes en particulier sont à l'étude au LAEI, où ils font l'objet de recherches actives.

**BIBLIOGRAPHIE**

- /1/ J.PETERS - W.DU MONG - M.DIEPERINCK  
Implantation d'un système de classification de pièces axé sur  
la fabrication. Technologie de Groupe  
CRIF, Section Construction Mécanique, MC 49, Jan 1974.
- /2/ M.P.GROOVER  
Fabrication assistée par ordinateur (FAO) volume 2  
HERMES PUBLISHING FRANCE, 1982, pp. 274-300.
- /3/ H.OPITZ  
Développements et domaines d'application de la Technologie de  
Groupe  
Mécanique, Matériaux, Electricité (GAMI), 282-283,  
juin-juillet 1973, pp. 3-10.
- /4/ A.NADIF - A.BRUYAND - B.MUTEL  
Technologie de Groupe: problèmes d'implantation et  
d'utilisation  
Mécanique, Matériaux, Electricité (GAMI), 410,  
janvier-février 1985, pp. 3-7.
- /5/ C.de l'ETANG - D.GOUTTE - Y.VENDEVILLE  
Organisation technique de la production à partir des  
groupements analogiques  
CETIM-Information, 37, oct 1974, pp. 15-28.
- /6/ C.de l'ETANG - Y.VENDEVILLE  
Systèmes de groupement analogiques  
Supplément au CETIM-Information, 34, févr 1974, pp.  
69-91.

- /7/ J.MINOT  
Contribution à l'étude de la reconnaissance de familles de pièces en fabrication mécanique. Application à la génération de gammes de fabrication.  
Thèse de Docteur Ingénieur, sept 1983.
- /8/ Application of computers to manufacturing  
Compte rendu du 13 Septembre 1965 Genève  
CIRP International Conference
- /9/ H.OPITZ  
Werkstücksystematik und Teilefamilienfertigung.  
Klassifizierungssystem  
Industrie-Anzeiger, vol 87, 103, déc 1965, pp. 457-468.
- /10/ J.PEKLENIK - J.GRUM  
Computer-aided design of the part spectrum data base and its application to design and production  
Annals of the CIRP, vol 31, 1, 1982, pp. 313-317.
- /11/ Y.VENDEVILLE  
Les groupements analogiques. Analyse générale des systèmes existants. Le système CETIM  
Document CETIM
- /12/ C.EMERSON - V.BOND - I.HAM  
Automated coding and process selection - ACAPS  
Manufacturing Engineering Transactions, 9th North American Manufacturing Research Conference, mai 1981, pp. 509-514.
- C.EMERSON - I.HAM  
An automated coding and process planning system using a DEC

PDP-10

Computer and Industrial Engineering, vol 6, 2, 1982, pp.  
159-168.

/13/ A.HOUTZEEL

Classification and coding

TNO, OIR (Organization for Industrial Research) Inc., USA

/14/ A.HOUTZEEL

MIPLAN - MULTIPLAN

OIR (Organization for Industrial Research) Inc., USA

/15/ D.COUE

WECO informatise l'élaboration de ses gammes de fabrication  
L'Usine Nouvelle, oct 1983, pp. 86-88.

/16/ M.BRUN - M.MORICEAU

SAVEGAM, logiciel d'élaboration et de valorisation des  
gammes de fabrication.

pp. 216-229.

/17/ W.EVERSHEIM

Zur Einführung eines Klassifizierungssystems

Industrie-Anzeiger, vol 89, 84, oct 1967, pp. 27-30.

/18/ W.EVERSHEIM - F.LOCK

Use of multivariate statistical methods for application of  
group technology in design and process planning departments  
Annals of the CIRP, vol 33, 1, 1984, pp.307-311.

/19/ C.FREIST - R.GRANOW

Aehnlichteilsuche mit Hilfe der Clusteranalyse

Teil 1: Grundlagen

VDI - Z, vol 124, 11, juin 1982, pp. 413-421.

/20/ C.FREIST - R.GRANOW

Aehnlichteilsuche mit Hilfe der Clusteranalyse

Teil 2: Das System Classic

VDI - Z, vol 124, 13, juil 1982, pp. 487-495.

/21/ G.KAELBERER

Untersuchung eines Produktspektrums mit Hilfe der Clusteranalyse

Z. für Operation - Research, Serie B, Praxis 21, 1977, pp.143-158.

/22/ G.KAELBERER

Automatische Klassifizierung von Arbeitsplänen mit Hilfe der Clusteranalyse

Z. für wirtsch.Fert., vol 75, 3, 1980, pp. 109-112.

/23/ J.L.CHANDON - S. PINSON

Analyse typologique. Théories et applications.

MASSON, 1981

/24/ C.HAN - I.HAM

Multiobjective cluster analysis for part family formations  
Journal of Manufacturing Systems, vol 5, 4, déc 1986, 8p.

/25/ J.PEKLENIK - J.GRUM

Investigation of the computer aided classification of parts

Annals of the CIRP, vol 29, 1, 1980, pp. 319-323.



- /26/ J.PEKLENIK - J.GRUM - B.LOGAR  
An integrated approach to CAD/CAPP/CAM and Group  
Technology by pattern recognition  
16th CIRP International Seminar on Manufacturing Systems,  
13-14 juil 1984, Tokyo, Japon.
- /27/ J.P.BENZECRI - F.BENZECRI  
Pratique de l'analyse des données  
DUNOD, 1984
- /28/ J.A.HARTIGAN  
Clustering algorithms  
WILEY, 1975.
- /29/ G.GOVAERT  
Classification croisée  
Thèse d'Etat, Université de Paris VI, Juin 1983.
- /30/ H.GARCIA - B.MUTEL - J.M.PROTH  
Classification automatique de données techniques de  
production.  
INRIA, RR n° 545, Juil 1986.
- /31/ J.P.BENZECRI et al  
L'analyse des données, tome 1: La taxinomie  
DUNOD, 1973.
- /32/ L.LEBART - J.P.FENELON  
Statistique et informatique appliquées  
BORDAS, 1975.
- /33/ J.P.BENZECRI et al

L'analyse des données, tome 2: L'Analyse des correspondances  
DUNOD, 1973.

/34/ E.DIDAY - J.LEMAIRE - J.POUGET - F.TESTU  
Eléments d'analyse des données  
BORDAS, 1982.

/35/ A.NADIF  
Méthodes typologiques d'analyse des données.  
Application industrielle  
Recherche bibliographique, ISMCM, juin 1983.

/36/ RONALD CEHESSAT  
Exercices commentés de statistique et informatique  
appliquées  
DUNOD, 1979.

/37/ E.DIDAY et al  
Optimisation en classification automatique. Tome 2.  
INRIA, 1980.

/38/ M.JAMBU  
Classification automatique pour l'analyse des données  
Tome 2: Logiciels  
DUNOD, 1978.

/39/ L.LEBART - A.MORINEAU - J.P.FENELON  
Traitement des données statistiques  
DUNOD, 1979.

/40/ I.C.LERMAN

Classification et analyse ordinale des données  
DUNOD, 1981.

/41/ H.RALAMBONDRAINNY - E.DIDAY

Optimisation d'une classification hiérarchique  
RAIRO (recherche opérationnelle), vol 18, 3, août 1984,  
pp. 279-295.

/42/ G.SAPOTRA

Traitement des variables qualitatives par codage  
COREF, note de travail n 1, Septembre 1976.

/43/ ADEPA-CEFIPA

MULTICLASS module de codification MULTI-M  
ADEPA-CEFIPA, mai 1983.

/44/ A.NADIF - M.COSTANTINI - B.MUTEL

Mesures de ressemblance de gammes de fabrication  
RAIRO, APII, vol 19, 5, 1985, pp. 455-470.

/45/ G.BRANGER

Guide du bureau des méthodes  
DESFORGES, Paris.

/46/ A.NADIF

Etude de la reconnaissance de familles de pièces décrites  
par le code MULTI-M  
Rapport ADEPA-LAEI, 1984.

/47/ A.NADIF

Application de l'analyse des données à la technologie de  
groupe

Note ADEPA-DAST (RNUR)-LAEI, sept 1984.

/48/ J.MINOT - B.MUTEL - Y.LEMOINE

Implantation assistée par ordinateur de la technologie de groupe

Congrès Automatique 1983, Productique et Robotique intelligente, Besançon, 15-17 nov 1983.

/49/ J.C.SIMON - E.DIDAY - J.P.MAROY - M.BERTHOD

Classification automatique et reconnaissance des formes

INRIA, RR n° 76, Juil 1974, pp. 13-18.

/50/ J.C.SIMON

La reconnaissance de formes par algorithmes  
MASSON, 1984, pp. 158-164.

/51/ B.FICHET - G.LE CALVE

Structure géométrique des principaux indices de dissimilarité sur signes de présence-absence  
Statistiques et analyse de données , vol 9, 3, 1984,  
pp. 11-44.

/52/ Y.LECHEVALLIER

Optimisation de quelques critères en classification automatique et application à l'étude des modifications des protéines sériques en pathologie clinique.

Thèse de Doctorat de 3ème cycle, Université de Paris VI, 17 juin 1974.

/53/ E.DIDAY

Optimisation en classification automatique et reconnaissance

des formes

RAIRO, vol 3, nov 1972, pp. 61-96.

/54/ H.OPITZ

Werkstücksystematik und Teilefamilienfertigung. Einführung  
in die Probleme der Werkstückklassifikation  
Industrie-Anzeiger, vol 89, 41, mai 1967, pp.19-22.

D.GRAESSLER

Aufbau und Systematik des VDW-Werkstückklassifizierungs-  
systems  
Industrie-Anzeiger, vol 89, 41, mai 1967, pp.22-28.

/55/ M.SOUVESTRE

MTRAN (utilitaire de formatage)  
ADEPA, jan 1985.

/56/ A.NADIF

Etude de la reconnaissance de familles de pièces au bureau  
des méthodes  
Mémoire de recherche, ISMCM, déc 1983.

/57/ Y.VENDEVILLE

Analyse des pièces élaborées dans les industries  
mécaniques  
CETIM-Informations, 49, févr 1977, pp. 25-34.

/58/ D.GOUTTE

Attribution rationnelle "pièces-moyens de production"  
CETIM-Informations, 46, juin 1976, pp. 21-30.

/59/ AIZERMAN, BRAVERMAN and ROZNER

The method of Potential Functions

Automation and Remote Control - 26, Nr11-1965.

**ANNEXES**

## ANNEXE 1

### Exemples de systèmes de codification autres que le code MULTI-M

- le code OPITZ (R.F.A), largement utilisé dans ce pays, aux Pays-Bas, au Japon, aux Etats-Unis et en Belgique [1, 3, 9, 54]

- le code CETIM-PMG (France), développé par le CETIM pour des entreprises françaises et sur lequel repose le système de classification T-CLASS [6, 57, 58].

- le code PEKLENIK (Yougoslavie), entièrement morphodimensionnel et présentant une grande originalité de codification [10, 25, 26].



CODE OPITZ (R.F.A.) [1, 3, 9, 54]

La description de la pièce est obtenue à l'aide d'un code (figure-A1-1) à neuf positions décimales, comprenant un code principal (les cinq premières positions) destiné à décrire la forme principale de la pièce finie ainsi que les usinages, et un code additionnel (les quatre dernières positions) contenant des informations complémentaires telles que les dimensions, les matières, les formes du brut et les tolérances.

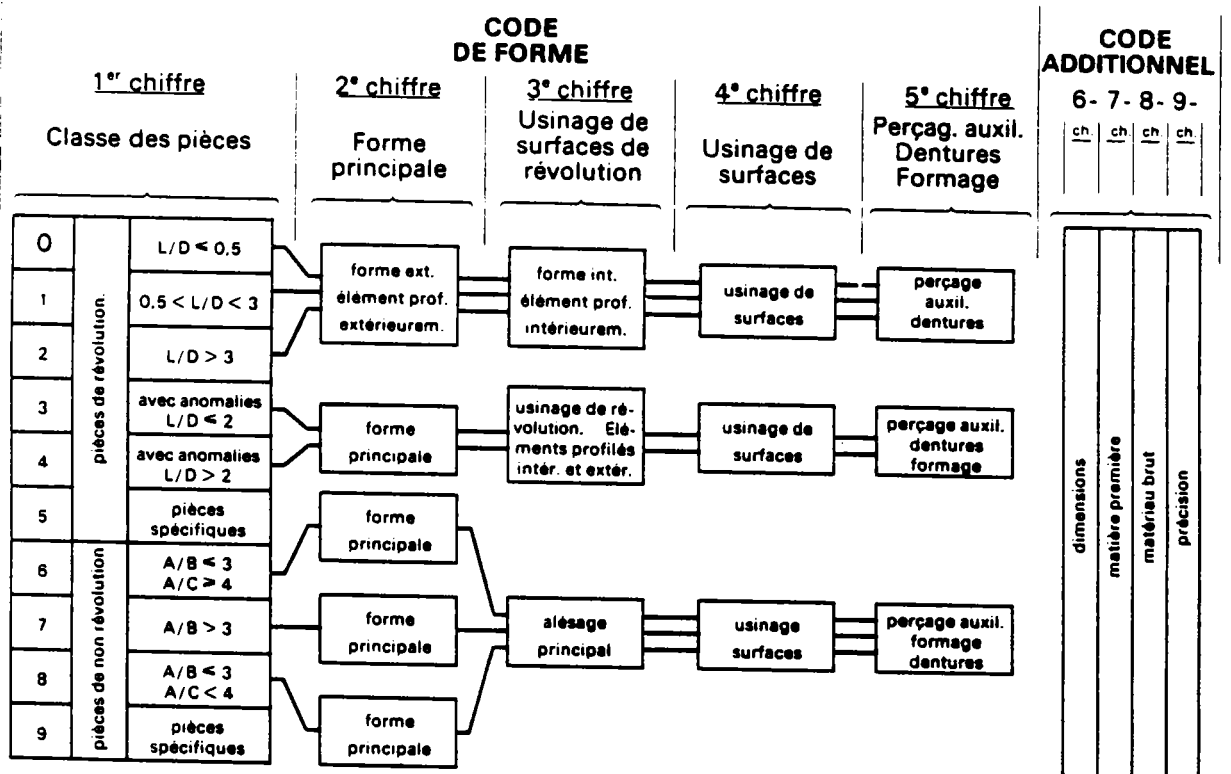
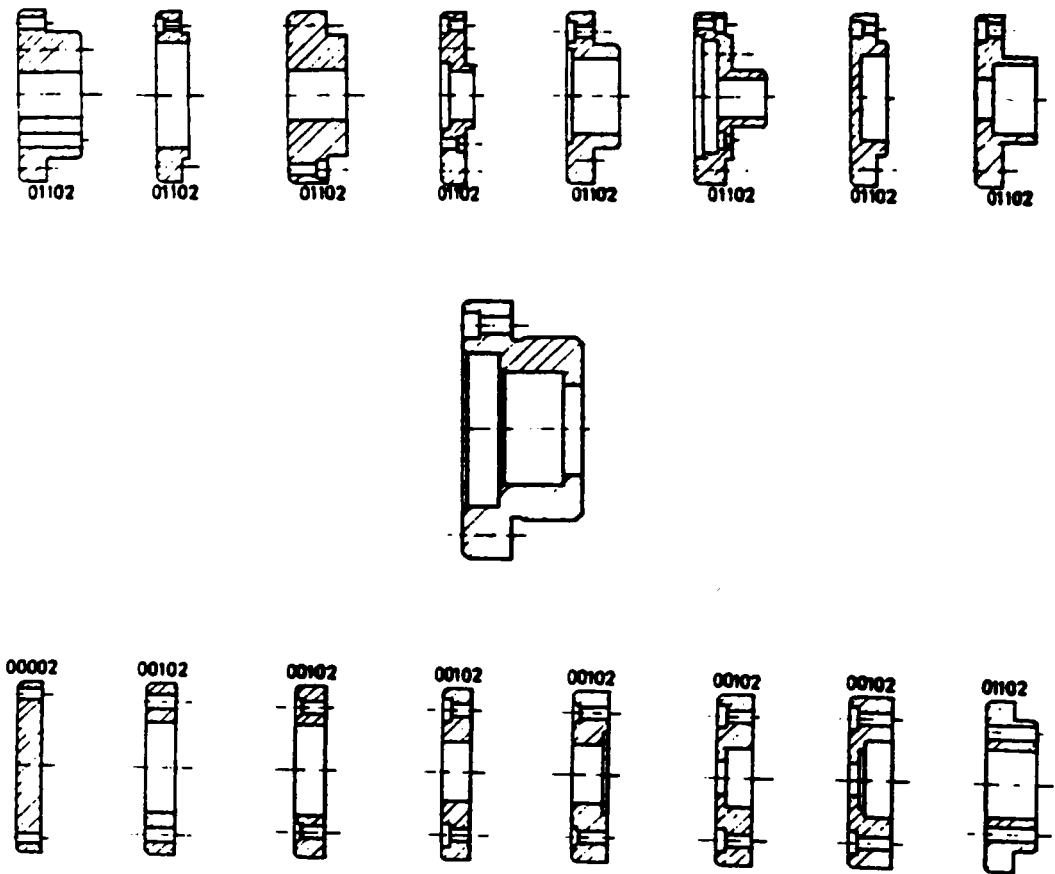


Figure-A1-1: Code Opitz [6].

L'efficacité du code OPITZ réside dans le fait que ses neuf positions sont entièrement indépendantes. Il est donc précis sans être redondant.

C'est un procédé très efficace pour la classification et le classement des pièces au Bureau d'Etudes.

Au Bureau des Méthodes, il présente cependant une certaine ambiguïté, liée à la classification primaire des pièces sur la première position (rapport  $L/D$  pour les pièces rondes,  $L/1$  pour les autres): une pièce ronde de longueur  $L$  et de diamètre  $D$  ne sera pas usinée de la même façon qu'une pièce ronde de longueur  $10L$  et de diamètre  $10D$ . Malgré cela, le code OPITZ constitue un bon moyen de classement au Bureau des Méthodes, en raison de son aptitude à décrire le processus de fabrication (poids du critère de fabrication: 67% (Chapitre II, figure-II-2)).



MODALITES

	0	1	2	3	4	5	6	7	8	9
1	■									
2	■		■							
3	■		■							
4	■									
5				■						
6	■	■	■	■	■					
7	■	■	■	■	■	■	■	■	■	■
8	■	■		■						
9	■	■	■		■	■		■		

Figure-A1-2: Exemple de spectre d'une famille de pièces codifiées avec le code OPITZ [17].

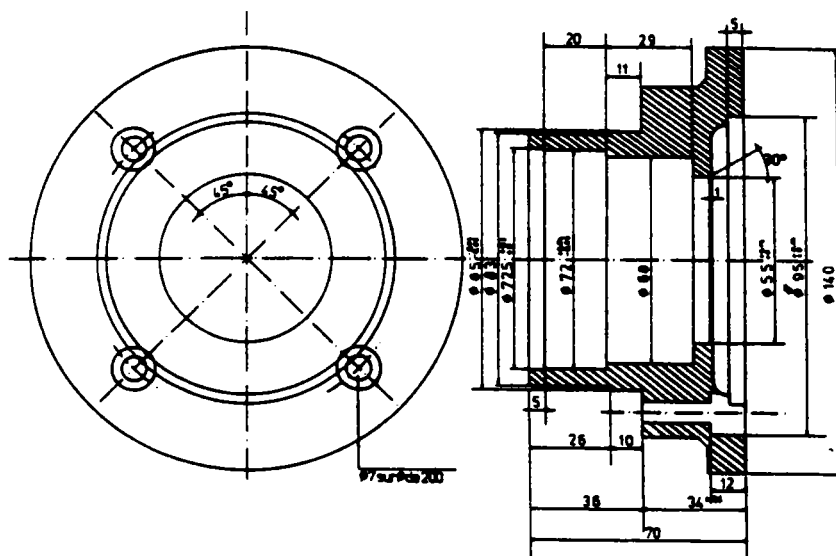
CODE CETIM-PMG [5, 6, 57, 58]

La description de la pièce est réalisée à l'aide d'un code à onze positions numériques décimales indépendantes entre elles, chacune d'entre elles offrant dix possibilités ou "modalités" (figure-A1-3):

		MODALITES										
		0	1	2	3	4	5	6	7	8	9	
0123456789	1											Morphologie de la pièce
	2											Dimension de la pièce
	3											Eléments de forme caractéristique
	4											Tolérance
	5											Matière
	6											Forme du brut
	7											Traitement thermique, Traitement de surface
	8											Quantité produite et Cadence
	9											
	10											
	11											

Figure-A1-3: Code CETIM-PMG [6, 57, 58]

La conception du code CETIM-PMG a été basée sur les résultats d'une enquête [57] menée par le CETIM auprès de vingt cinq entreprises françaises de mécanique générale, et qui a abouti à une codification des pièces relativement simple, ne nécessitant que des moyens informatiques de petite configuration.



MODALITES

	0	1	2	3	4	5	6	7	8	9
1				■						
2					■					
3						■				
4			■							
5	■									
6				■						
7						■				
8							■			
9								■		
10	■									
11		■								

Pièce de révolution étagée extérieurement d'un côté, étagée intérieurement des deux côtés sans chambrage	
$\phi$ compris entre 125 et 200 mm L compris entre 40 et 80 mm	$\frac{L}{D} > \frac{1}{3}$
Aucun élément de forme additionnel	
Trous hors de l'axe de révolution + encoches	
Pièce de qualité 4 à 5	
Pièce en fonte	
Pièce moulée en sable	
Sans traitement thermique	
Pièces lancées par lot de 6 à 25, < 3fois/an	

Figure-A1-4: Exemple de pièce codifiée à l'aide du code CETIM-PMG [58].

CODE PEKLENIK [10, 25, 26]

Ce code se décompose en trois codes morphodimensionnels distincts, réservés l'un aux pièces de révolution, l'autre aux pièces de tôlerie et le troisième aux pièces tridimensionnelles.

Chaque code comprend cent positions à une modalité, ce qui se ramène à un système binaire. Il se présente sous la forme d'un tableau carré de dimension 10x10.

La plus grande originalité réside dans la codification des pièces tridimensionnelles.

En effet, la pièce est décomposée par projection sur les six faces d'un parallélépipède rectangle (vue de face, de dessus, de côté,...) (figure-A1-5). Ensuite chaque face est codifiée séparément à l'aide d'un tableau 10x10 (figure-A1-6). Un tel procédé permet de coder des pièces complexes avec autant de facilité que des pièces simples.

Le code PEKLENIK est d'une grande précision, mais nécessite une mémoire de masse très importante (mémorisation de 600 chiffres pour une pièce tridimensionnelle, par exemple).

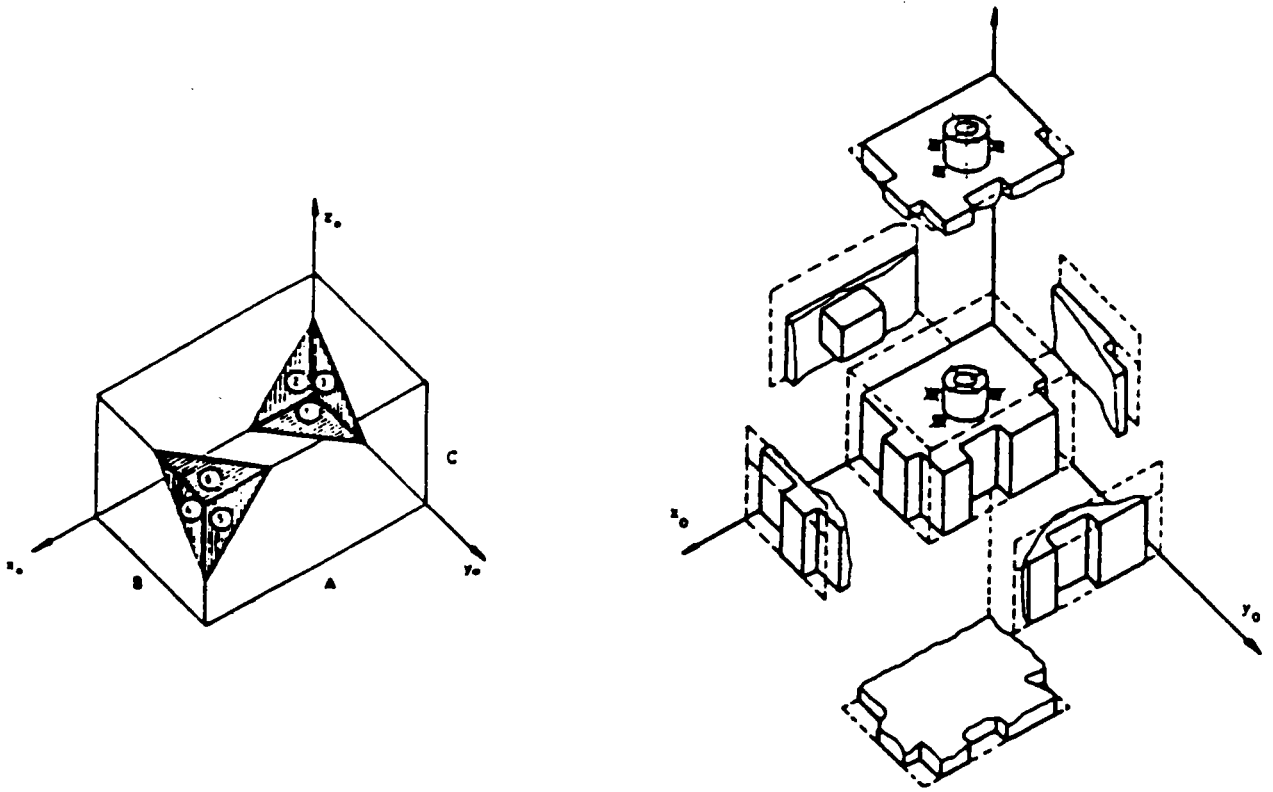


Figure-A1-5: Principe de codification des pièces tridimensionnelles par le code PEKLENIK: décomposition par projection sur les six faces d'un parallélépipède rectangle [26].

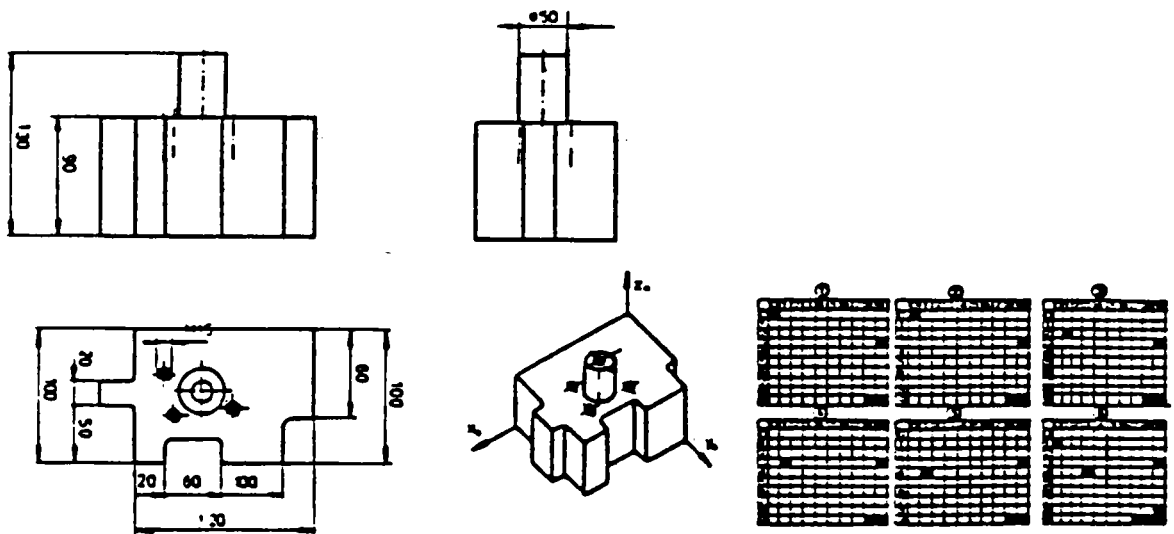
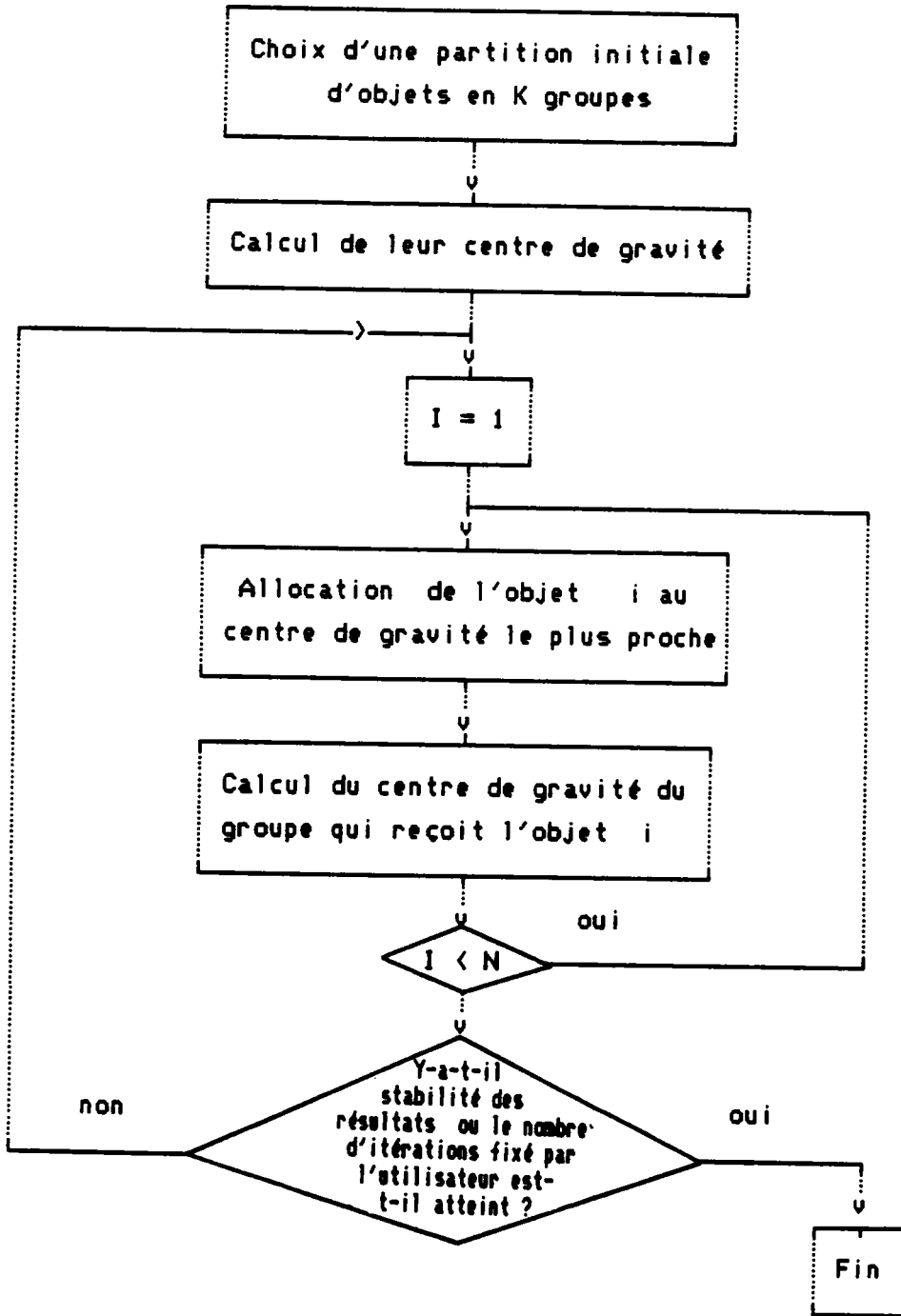


Figure-A1-6: Code PEKLENIK de la pièce représentée à la figure-A1-5 [26].

ANNEXE 2

ALGORITHME DES K-MEANS DE MACQUEEN (1967)





### ANNEXE 3

#### Exemples de mesures de proximité classiques [23]

- Distances euclidiennes
- Distances rectangulaires
- Distance de Chebychev
- Distance de Minkowski
- Distance adaptative de Rohlf
- Proximités utilisables sur des matrices de fréquence ou d'occurrence.

DISTANCES EUCLIDIENNES

$$d_{ik} = \left[ \sum_{j=1}^p w_j (x_{ij} - x_{kj})^2 \right]^{1/2}$$

D1 : Distance euclidienne non pondérée :  $w_j = 1$

$$d_{ik}^2 = \sum_{j=1}^p (x_{ij} - x_{kj})^2$$

D2 : Distance euclidienne pondérée par l'inverse de la variance :  $w_j = 1/\sigma_j^2$

$$d_{ik}^2 = \sum_{j=1}^p \frac{1}{\sigma_j^2} (x_{ij} - x_{kj})^2$$

D3 : Distance euclidienne pondérée par l'inverse de l'écart maximum :

$$w_j = 1/\max_{i,k} (x_{ij} - x_{kj})^2$$

$$d_{ik}^2 = \sum_{j=1}^p \frac{1}{\max_{i,k} (x_{ij} - x_{kj})^2} (x_{ij} - x_{kj})^2$$

D4 : Distance de CLARK :  $w_j = 1/(x_{ij} + x_{kj})^2$

$$d_{ik}^2 = \frac{1}{p} \sum_{j=1}^p \frac{1}{(x_{ij} + x_{kj})^2} (x_{ij} - x_{kj})^2$$

D5 : Distance euclidienne pondérée par la fiabilité des mesures :  $w_j = p_j$

$$d_{ik}^2 = \sum_{j=1}^p p_j (x_{ij} - x_{kj})^2$$

---

$X_{..}$  = Moyenne générale

$X_{.j}$  = Moyenne d'une colonne j

$X_{i.}$  = Moyenne d'une ligne i

D6 : Distance euclidienne pondérée moyenne :

$$d_{ik}^2 = \frac{1}{p} \sum_{j=1}^p w_j (x_{ij} - x_{kj})^2$$

D7 : Distance du Chi Deux

$$d_{ik}^2 = \sum_{j=1}^p \frac{x_{..}}{x_{.j}} \left( \frac{x_{ij}}{x_{i.}} - \frac{x_{kj}}{x_{k.}} \right)^2$$

D8 : Distance de MAHALANOBIS

$$d_{ik}^2 = \sum_{j=1}^p \sum_{l=1}^p w_{jl} (x_{ij} - x_{kj}) (x_{il} - x_{kl})$$

avec  $w_{jl}$  = élément de l'inverse de la matrice de co-variance

#### DISTANCES RECTANGULAIRES

$$d_{ik} = \sum_{j=1}^p w_j |x_{ij} - x_{kj}|$$

D9 : Distance rectangulaire non pondérée ou "city block" :  $w_j = 1$

$$d_{ik} = \sum_{j=1}^p |x_{ij} - x_{kj}|$$

D10 : Distance rectangulaire pondérée par le total colonne :  $w_j = 1/x_{.j}$

$$d_{ik} = \sum_{j=1}^p \frac{1}{x_{.j}} |x_{ij} - x_{kj}|$$

D11 : Distance de Canberra :  $w_j = 1/(x_{ij} + x_{kj})$

$$d_{ik} = \sum_{j=1}^p \frac{1}{(x_{ij} + x_{kj})} |x_{ij} - x_{kj}|$$

D12 : Distance de BRAY et CURTIS :  $w_j = 1 / \sum_{j=1}^p (x_{ij} + x_{kj})$

$$d_{ik} = \sum_{j=1}^p \frac{1}{\sum_{j=1}^p (x_{ij} + x_{kj})} |x_{ij} - x_{kj}|$$

D13 : Distance rectangulaire pondérée moyenne

$$d_{ik} = \frac{1}{p} \sum_{j=1}^p w_j |x_{ij} - x_{kj}|$$

#### DISTANCE DE CHEBYCHEV

D14 :  $d_{ik} = \max_j (x_{ij} - x_{kj})$

#### DISTANCE DE MINKOWSKI

D15 :  $d_{ik} = \left[ \sum_{j=1}^p w_j (x_{ij} - x_{kj})^r \right]^{1/r}$

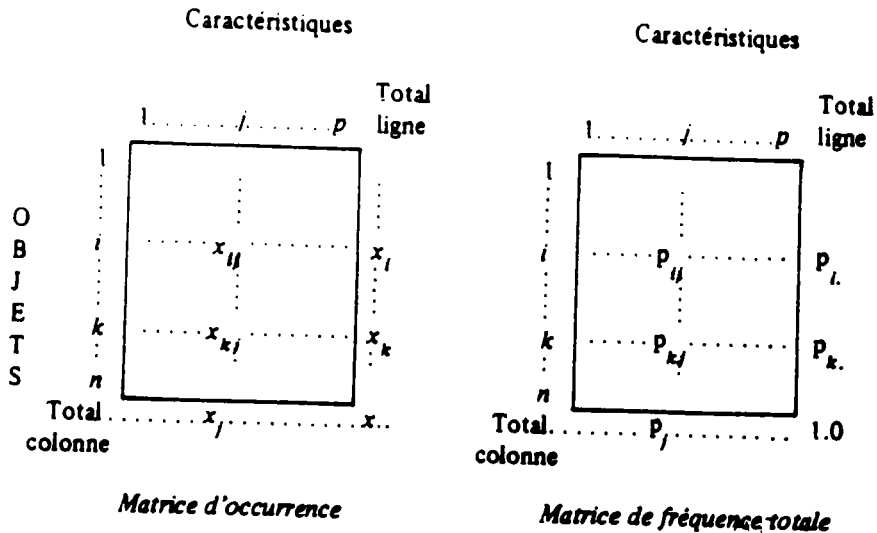
#### DISTANCE ADAPTATIVE DE ROHLF

D16 :  $d_{ik}^2 = (\bar{y}_k - \bar{y}_i)^t S_i^{-1} (\bar{y}_k - \bar{y}_i) |S_i|^{1/p}$

avec  $\bar{y}_i$  = vecteur des moyennes, interne à l'objet i ou au groupe d'objets i, des p variables.

$S_i^{-1}$  = inverse de la matrice de variance covariance des variables à l'intérieur du groupe d'objets i

$|S_i|$  = déterminant de  $S_i$



Matrices d'occurrence et de fréquence

Les proximités utilisables sur des matrices de fréquence ou d'occurrence.

Distance du Chi Deux

$$F1 : d_{ik}^2 = \sum_{j=1}^p \frac{1}{p \cdot j} \left( \frac{p_{ij}}{p_{i.}} - \frac{p_{kj}}{p_{k.}} \right)^2$$

Distance de Jambu

$$F2 : d_{ik}^2 = \sum_{j=1}^p \frac{1}{(p_{ij} + p_{kj})} \left( \frac{p_{ij}}{p_{i.}} - \frac{p_{kj}}{p_{k.}} \right)^2$$

Information mutuelle

$$F3 : S_{ik} = \sum_{j=1}^p \sum_{l=1}^n p_{jl} \log_2 \left[ \frac{p_{jl}}{p_{j.} \cdot p_{l.}} \right]$$

Distance de RAJSKI

$$F4 : d_{ik} = \frac{\sum_{j=1}^p \sum_{l=1}^n p_{jl} (\log_2 p_{j.} + \log_2 p_{l.} - 2 \log_2 p_{jl})}{-\sum_{j=1}^p \sum_{l=1}^n p_{jl} \log_2 p_{jl}}$$

Coefficient de cohérence de RAJSKI (similarité)

$$F5 : R_{ik} = \left[ 1 - d_{ik}^2 \right]^{1/2}$$

ANNEXE 4

Algorithme de FISCHER [28]. Exemple d'application.

Soit N objets partagés en K classes à l'aide d'une partition P(N,K).

L'hétérogénéité de la partition peut se mettre sous la forme :

$$W[P(N,K)] = \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

avec :

-  $n_i$  : nombre d'éléments de la classe i

$$\sum_{i=1}^K n_i = N$$

-  $\bar{x}_i$  : moyenne des valeurs prises par la variable x dans la classe i

-  $x_{ij}$  : valeur prise par la variable x pour l'élément j de la classe i.

La maximisation de l'homogénéité des classes qui composent la partition revient à la minimisation de l'hétérogénéité  $W$  de la partition dans son ensemble.

L'algorithme de FISCHER, qui est décrit à la figure-A4-1, a donc comme objectif de minimiser le critère  $W$ .

### ALGORITHME DE FISHER

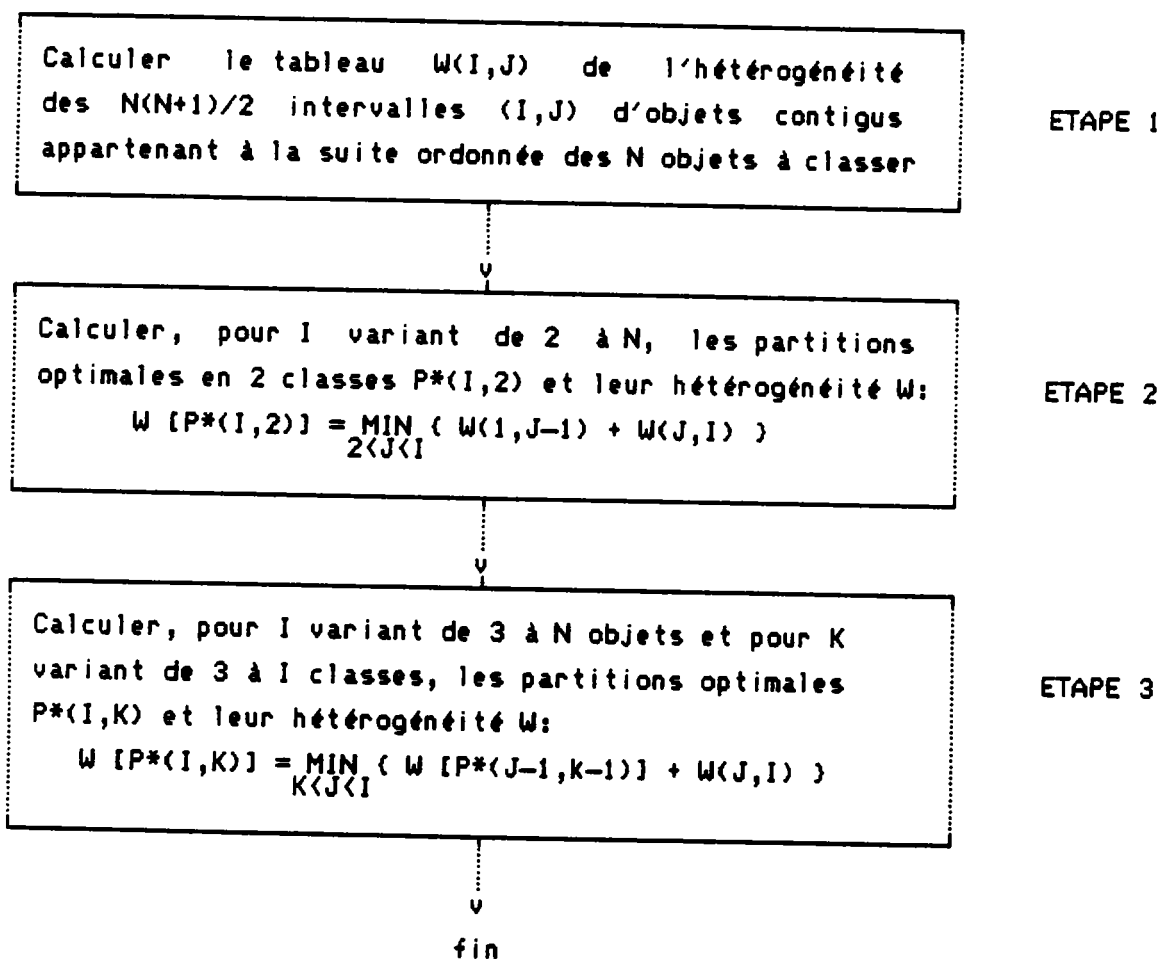


Figure-A4-1: Algorithme de FISCHER [28].

Exemple :

soit l'ensemble  $\Omega$  de six données (10.1,20,10,30.1,20.1,30), on veut trouver la partition optimale  $P(\Omega)$

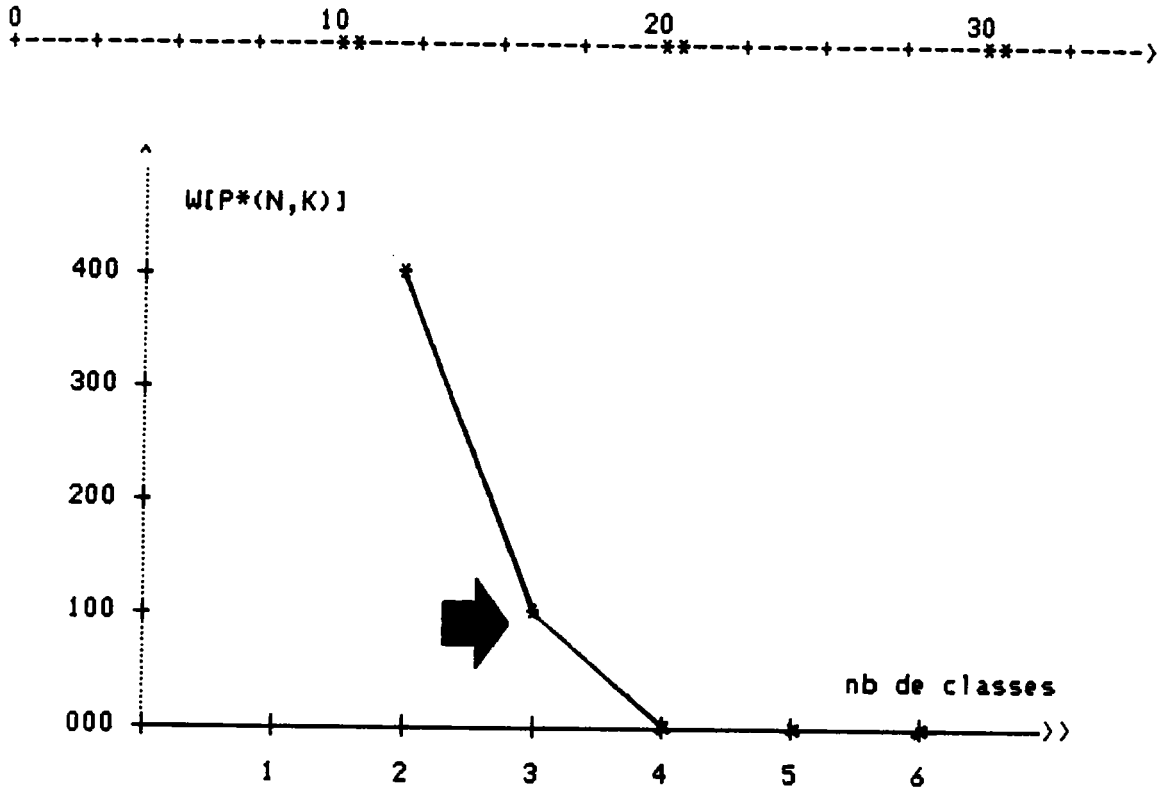


Figure-A4-2: Représentation du critère d'hétérogénéité  $W$  en fonction du nombre des classes.

La plus grande diminution de l'hétérogénéité intervient pour 3 classes

**PARTITION OPTIMALE :**

- 1 10
- 1 10.1
- 2 20
- 2 20.1
- 3 30
- 3 30.1



ANNEXE 5

Transcodage du code MULTI-M

a) Pièces de révolution

Selon le contenu de la variable  $V_1$ , on distingue trois types de pièces de révolution:

Contenu de la variable $V_1$	Signification	Groupe de pièces
1	pièces étagées d'un côté	rondes
2	étagées des deux côtés	
3	avec cône et/ou profil	

Les pièces rondes après transcodage sont décrites par 12 nouvelles variables indépendantes ( $NV_1, NV_2, \dots, NV_{12}$ ):

Nouvelles variables	Variables anciennes
$NV_1$	$V_1 \times 10 + V_2$
$NV_2$	$V_1 \times 10 + V_3$
$NV_3$	$V_4$
$NV_4$	$V_5$
$NV_5$	$V_6$
$NV_6$	$V_7$
$NV_7$	$V_8$
$NV_8$	$V_9$
$NV_9$	$V_{10}$
$NV_{10}$	$V_{11}$
$NV_{11}$	$V_{12}$
$NV_{12}$	$V_{13}$

b) Pièces rondes spéciales

Pour ce type de pièces ( $V1=4$ ), étant donné la complexité de codification des variables  $V1$ ,  $V2$  et  $V3$ , un transcodage particulier a été nécessaire.

Les pièces rondes spéciales sont décrites par 10 nouvelles variables indépendantes ( $NV1, NV2, \dots, NV10$ ):

Nouvelles variables	Variables anciennes
NV1	$V2 \times 100 + V3 \times 10 + V4$
NV2	V5
NV3	V6
NV4	V7
NV5	V8
NV6	V9
NV7	V10
NV8	V11
NV9	V12
NV10	V13

c) Pièces de tôlerie

Selon la valeur de la variable  $V_1$ , il y a deux sortes de pièces de tôlerie:

- $V_1=5$ : pièces plates

- $V_1=6$ : pièces déformées.

Les pièces de tôlerie après transcodage sont décrites par 12 nouvelles variables indépendantes ( $NV_1, NV_2, \dots, NV_{12}$ ):

Nouvelles variables	Variables anciennes
$NV_1$	$V_2$
$NV_2$	$V_3$
$NV_3$	$V_4$
$NV_4$	$V_5$
$NV_5$	$V_6$
$NV_6$	$V_7$
$NV_7$	$V_8$
$NV_8$	$V_9$
$NV_9$	$V_{10}$
$NV_{10}$	$V_{11}$
$NV_{11}$	$V_{12}$
$NV_{12}$	$V_1 \times 10 + V_{13}$

d) Pièces tridimensionnelles

Selon la valeur prise par la variable  $V_1$ , il y a 3 types de pièces tridimensionnelles:

Contenu de la variable $V_1$	Signification	Groupe de pièces
7	parallélépipédiques	tridimensionnelles
8	sculptées	
9	irrégulières	

Après transcodage, les pièces tridimensionnelles sont codées à l'aide de 12 nouvelles variables indépendantes ( $NV_1, NV_2, \dots, NV_{12}$ ):

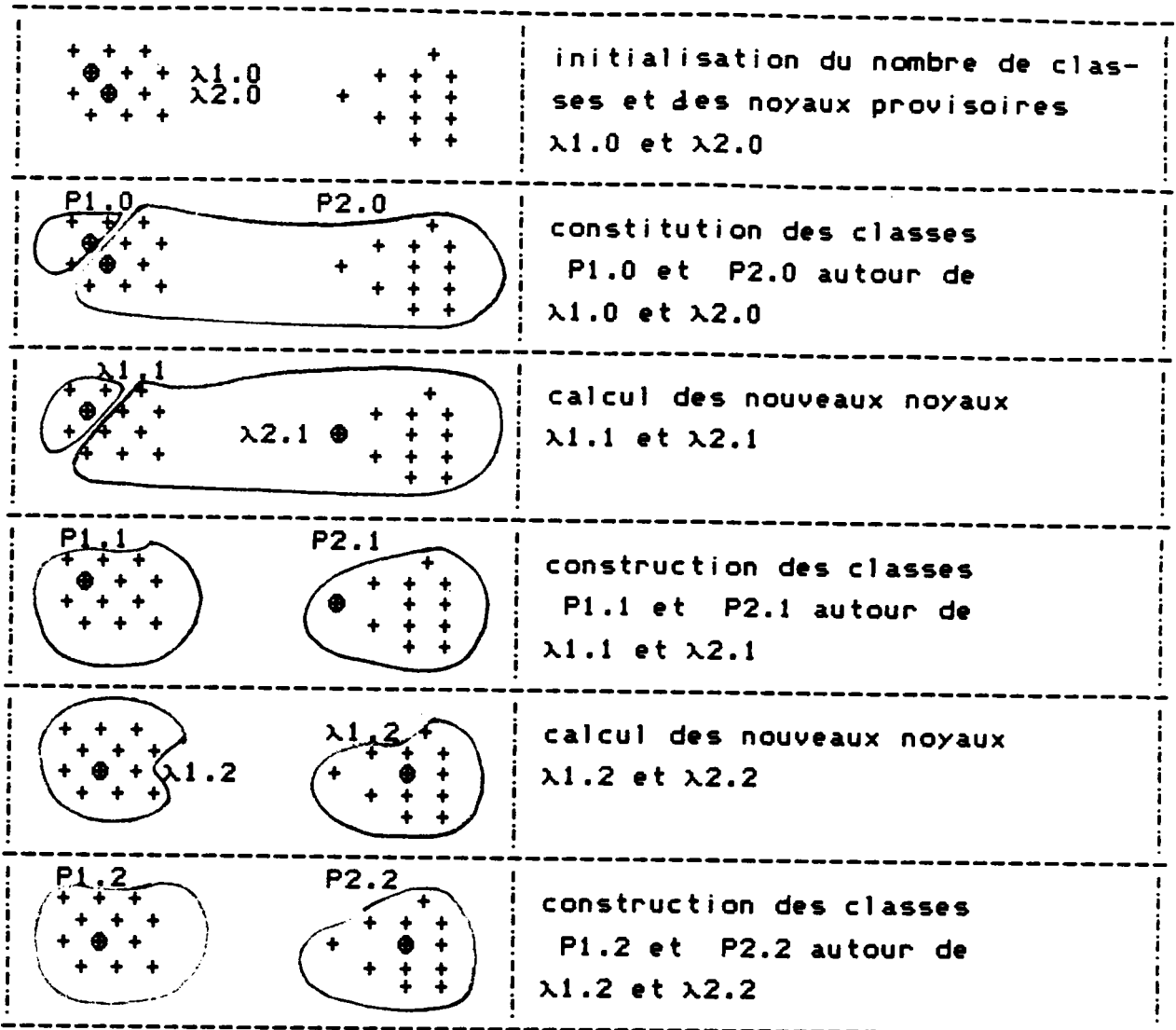
Nouvelles variables	Variables anciennes
$NV_1$	$V_1 \times 10 + V_2$
$NV_2$	$V_3$
$NV_3$	$V_4$
$NV_4$	$V_5$
$NV_5$	$V_6$
$NV_6$	$V_7$
$NV_7$	$V_8$
$NV_8$	$V_9$
$NV_9$	$V_{10}$
$NV_{10}$	$V_{11}$
$NV_{11}$	$V_{12}$
$NV_{12}$	$V_{13}$

## ANNEXE 6

### Algorithme des Nuées Dynamiques

Soit un ensemble  $\Omega$  à  $n$  objets ( $\Omega = (\omega_1, \omega_2, \dots, \omega_i, \dots, \omega_n)$ ) décrits par  $p$  variables. Nous cherchons une partition  $\{P_k$  de  $\Omega$  en  $k$  classes disjointes ( $\{P_k = (P_1, P_2, \dots, P_k)$ ). On suppose que l'espace  $\mathbb{R}^p$  est muni d'une distance euclidienne 'd'. Chaque classe  $P_j$  sera représentée par son noyau  $\lambda_j$ .  $A_k$  est l'ensemble des  $k$  noyaux des  $k$  classes de la partition  $\{P$  ( $A_k = (\lambda_1, \lambda_2, \dots, \lambda_k)$ ).

Les étapes de l'algorithme (voir la figure-III-10) sont illustrées ci-dessous:



ETAPE 0: On fixe les noyaux initiaux ou on les choisit au hasard. Ces noyaux  $(\lambda_{1.0}, \lambda_{2.0}, \dots, \lambda_{k.0})$  induisent une partition  $\mathcal{P}_{K.0}$  en  $k$  classes  $(P_{1.0}, P_{2.0}, \dots, P_{k.0})$ . Ainsi, l'objet  $\omega_i$  appartient à  $P_{k.0}$  s'il est plus proche de  $\lambda_{k.0}$  que tous les autres noyaux au sens d'une distance euclidienne.

ETAPE 1: On calcule les  $k$  nouveaux noyaux  $(\lambda_{k.1})$  à l'aide de la fonction de représentation qui peut être le centre de gravité de chaque classe par exemple.

ETAPE 2: On détermine la partition  $\mathcal{P}_{K.1}$  en affectant les  $i$  objets aux  $k$  noyaux les plus proches (fonction d'affectation).

L'algorithme s'arrête soit lorsque deux itérations successives conduisent à la même partition, soit lorsqu'un critère choisi convenablement ( par exemple la variance intra-classe ) cesse de décroître de façon sensible, soit encore parcequ'un nombre maximal d'itérations a été fixé a priori.

Montrons que la variance intra-classe ne peut que décroître ( ou rester stationnaire ) entre l'étape m et m+1. Des règles d'affectation ( qui sont des conventions de programmation propres à chaque spécification de l'algorithme ) permettent de faire en sorte que cette décroissance soit stricte, et donc de conclure à la convergence de l'algorithme puisque l'ensemble de départ  $\Omega$  est fini.

Supposons que les n objets soient munis de masses relatives  $p_i$  ( leur somme vaut 1 ), et soit  $d^2(\omega_i, \lambda_{k,m})$  le carré de la distance entre l'objet  $\omega_i$  et centre de la classe  $\lambda_k$  à l'étape m.

Nous nous intéresserons à la quantité critère:

$$v(m) = \sum_{k=1}^q \left( \sum_{\omega_i \in P_{k,m}} p_i d^2(\omega_i, \lambda_{k,m}) \right)$$

Rappelons qu'à l'étape m, la classe  $P_{k,m}$  est formée des objets plus proches de  $\lambda_{k,m}$  que de tous les autres centres ( ces centres de gravité des classes  $P_{k,m-1}$  de l'étape précédente ).

La variance intra-classe à l'étape m est donc la quantité:

$$V(m) = \sum_{k=1}^q \left( \sum_{\omega_i \in P_{k,m}} p_i d^2(\omega_i, \lambda_{k,m+1}) \right)$$



où  $\lambda_{k,m+1}$  est le centre de gravité de la classe  $PK.m$ . A l'étape  $m+1$ , la quantité critère s'écrit:

$$v(m+1) = \sum_{k=1}^g \left( \sum_{\omega_i \in PK.m+1} p_i d^2(\omega_i, \lambda_{k,m+1}) \right)$$

On va montrer que:

$$v(m) \geq V(m) \geq v(m+1)$$

ce qui établira la décroissance simultanée du critère et de la variance intra-classe.

Remarquons tout d'abord, en notant  $p(k)$  la somme des  $p_i$  pour  $\omega_i \in PK.m$ :

$$v(m) = V(m) + \sum_{k=1}^g p(k) d^2(\lambda_{k,m+1}, \lambda_{k,m})$$

d'après le théorème de Huygens, ce qui établit la première partie de l'inégalité.

La seconde partie découle du fait qu'entre les accolades qui apparaissent dans la définition de  $V(m)$  et  $v(m+1)$ , seules changent les affectations des objets aux centres. Puisque  $PK.m+1$  est l'ensemble des objets plus proches de  $\lambda_{k,m+1}$  que tous les autres centres, les distances n'ont pu que décroître ( ou rester inchangées ) au cours de cette réaffectation.

**ANNEXE 7**

**RESULTATS DE LA CLASSIFICATION EFFECTUEE SUR DES DONNEES INDUSTRIELLES**

classe numéro 1

ELEM	INERTIE
27	3.244
28	3.244
32	4.540
33	1.678
34	1.508
35	1.678
46	1.760
55	1.585
56	1.585
57	1.585
58	1.585

smin= 1.508 min=34  
smax= 4.540 max=32

classe numéro 2

ELEM	INERTIE
39	0.574
2	0.857
38	0.647

smin= 0.574 min=39  
smax= 0.857 max= 2

inertie du groupe par rapport à l'élément 39

l'élément 39 est le centre de classe

l'élément 2 est le plus éloigné de la classe

classe numéro 3

ELEM	INERTIE
29	0.000

smin= 0.000 min=29  
smax= 0.000 max=29

classe numéro 4

ELEM	INERTIE
5	0.364
1	0.364

smin= 0.364 min= 1  
smax= 0.364 max= 1

classe numéro 5

ELEM INERTIE  
-----

14	0.569
6	0.523
41	0.626

smin= 0.523 min= 6  
smax= 0.626 max=41

classe numéro 6

ELEM INERTIE  
-----

24	0.219
37	0.469
25	0.219
22	0.281

smin= 0.219 min=25  
smax= 0.469 max=37

classe numéro 7

ELEM INERTIE  
-----

16	0.000
----	-------

smin= 0.000 min=16  
smax= 0.000 max=16

classe numéro 8

ELEM INERTIE  
-----

61	0.000
62	0.000

smin= 0.000 min=62  
smax= 0.000 max=62

classe numéro 9

ELEM INERTIE  
-----

47	0.000
----	-------

smin= 0.000 min=47  
smax= 0.000 max=47

classe numéro10

ELEM INERTIE  
-----

12	7.201
30	10.618
31	18.260
13	14.291
3	10.604
15	4.366
4	5.455
36	5.118
17	11.866
18	4.509
19	5.898
40	12.897
20	5.131
42	4.352
43	6.145
44	4.810
45	4.352
21	6.233
26	6.131
48	5.115
49	4.430
50	4.430
51	4.756
52	5.043
53	5.586
54	5.586
7	8.145
23	15.123
8	5.532
9	4.765
59	6.045
60	6.045
10	4.765
11	5.603
63	4.810
64	4.743
65	4.352
66	4.352
67	4.430
68	6.449
69	6.480
70	6.449

smin= 4.352 min=66  
smax= 18.260 max=31

totsmin= 7.541 ← inertie intra-classe

totmin = 16.276 mintot=67 ← l'élément 67 est le noyau de  
totmax = 37.949 maxtot=28 ← l'ensemble des objets étudiés

l'élément 28 est le plus éloigné

b( 1)=34	effectif= 11
b( 2)=39	effectif= 3
b( 3)=29	effectif= 1
b( 4)= 1	effectif= 2
b( 5)= 6	effectif= 3
b( 6)=25	effectif= 4
b( 7)=16	effectif= 1
b( 8)=62	effectif= 2
b( 9)=47	effectif= 1
b(10)=66	effectif= 42

l'élément 34 est le noyau de la classe 1, qui a 11 éléments

**ANNEXE 8**

**RESULTAT DU REGROUPEMENT DES FORMES FORTES**

classe numéro 1

ELEM INERTIE  
-----

27	4.600
28	4.600
32	5.942
33	2.482
34	2.303
35	2.482
46	2.767
55	2.503
56	2.503
57	2.583
58	2.583
29	4.783
61	3.806
62	3.806

smin= 2.303 min=34  
smax= 5.942 max=32

classe numéro 2

ELEM INERTIE  
-----

14	1.268
6	0.958
41	1.506
5	1.238
1	1.504

smin= 0.958 min= 6  
smax= 1.506 max=41

classe numéro 3

ELEM INERTIE  
-----

38	0.647
39	0.574
2	0.857

smin= 0.574 min=39  
smax= 0.857 max= 2



classe numéro 4

ELEM	INERTIE
----	-----
25	19.991
22	18.987
24	19.991
37	15.912
47	11.912
12	9.478
30	13.099
31	21.426
13	16.990
3	12.961
15	6.109
4	7.424
36	7.018
17	14.353
18	6.418
19	7.765
40	15.325
20	6.942
42	5.973
43	8.013
44	6.710
45	5.973
21	8.306
26	8.156
48	6.859
49	6.261
50	6.261
51	6.552
52	6.784
53	7.554
54	7.554
7	10.325
23	18.000
8	7.555
9	6.728
59	8.118
60	8.118
10	6.728
11	7.471
63	6.710
64	6.483
65	5.973
66	5.973
67	6.261
68	8.372
69	8.299
70	8.372

smin= 5.973 min=66  
smax= 21.426 max=31

classe numéro 5

ELEM INERTIE  
-----

16 0.000

smin= 0.000 min=16  
smax= 0.000 max=16

totsmin= 9.808

totmin = 16.276 mintot=67  
totmax = 37.949 maxtot=28

b( 1)=34 effectif= 14  
b( 2)= 6 effectif= 5  
b( 3)=39 effectif= 3  
b( 4)=66 effectif= 47  
b( 5)=16 effectif= 1