Contribution à l'analyse des données en hydrométéorologie, la prévision des phénomènes accidentels et l'analyse des champs spatiaux : application à la prévision des avalanches à Davos et à l'analyse des épisodes pluvieux cévenols

Charles Obled

N° d'ordre

THÈSE

présentée à

L'UNIVERSITÉ SCIENTIFIQUE ET MÉDICALE

ET

L'INSTITUT NATIONAL POLYTECHNIQUE
DE GRENOBLE

pour obtenir le grade de

DOCTEUR D'ÉTAT ÈS-SCIENCES PHYSIQUES

par

Charles OBLED
Ingénieur ENSHG - IMAG

SUJET

Contribution à l'analyse des données en Hydrométéorologie

La prévision des phénomènes accidentels et l'analyse des champs spatiaux

Application à la prévision des avalanches à Davos et à l'analyse des épisodes pluvieux Cévenols

---

Soutenue le 14 Décembre 1979 devant la Commission d'Examen

M. G. LESPINARD	Président	
M. J. BERNIER	Examinateurs	
M. M. BOUVARD		
M. G. ROMIER		
M. W. GOOD	Invités	
M. D. DUBAND		

GRENOBLE

AVANT-PROPOS

Je remercie Monsieur le Professeur LESPINARD président du Jury, pour son soutien bienveillant et l'intérêt qu'il nous a manifesté tout au long de nos recherches.

Monsieur le Professeur BOUVARD, Directeur de l'Ecole d'Hydraulique, a constamment suivi et encouragé nos travaux. Je lui exprime toute ma reconnaissance.

Monsieur le Professeur SANTON, puis Monsieur VACHAUD, Maître de Recherche au C.N.R.S., nous ont accueilli au sein du Groupe d'Hydrologie et nous ont encouragés à y développer les approches statistiques. Je les en remercie.

Monsieur BERNIER, du Centre de Recherche E.D.F. à Chatou, a bien voulu accepter d'être mon rapporteur. Je le remercie très sincèrement pour ses stimulantes critiques, qui ont beaucoup contribué à améliorer l'exposé qui va suivre.

Nous avons envers Monsieur GUILLOT, chef du Service Ressources en Eau de la Division Technique Générale d'E.D.F., une dette particulière. C'est grâce à sa ténacité que notre équipe a pu naître, et il a su nous communiquer sa curiosité, son exigence scientifique et un certain sens critique.

Monsieur DUBAND, son collaborateur, a guidé nos premiers pas dans l'Analyse des Données. Confident de nos enthousiasmes et de nos déceptions, semant sans cesse le doute dans nos a priori les plus confortables, je ne sais comment le remercier de ce que je lui dois.

Monsieur le Professeur ROMMIER, directeur de l'Institut de Mathématiques en Sciences Sociales, ainsi que ses collaborateurs Messieurs CAILLOT et DROUET d'AUBIGNIES, nous ont, en tant que statisticiens, souvent aidé dans nos travaux. Je leur en suis très reconnaissant.

Monsieur le Professeur de QUERVAIN, Directeur de l'Institut Fédéral Suisse pour l'étude de la neige et des avalanches, nous a donné les moyens d'appliquer notre approche statistique à la prévision des avalanches et Monsieur Walter GOOD, son adjoint, en a assuré la mise en oeuvre et l'utilisation opérationnelle. Qu'ils en soient remerciés, ainsi que tout le personnel de leur Institut.

Ce mémoire est aussi le résultat d'un travail en équipe, et il doit beaucoup à la collaboration de Monsieur Philippe BOIS. Commencée dans les randonnées à ski, notre amitié s'est poursuivie dans l'hydrologie de montagne et l'hydrométéorologie ... Quant aux stagiaires de 3° cycle qui ont séjourné par nous, Messieurs FRITSCH, GOUSSEBAÏLE, LEBEL, et plus récemment, D. CREUTIN, ils ont contribué largement à nos travaux. Je remercie aussi nos collègues mathématiciens, en particulier Monsieur TEMPERVILLE, qui nous ont sorti de plus d'un mauvais pas.

De nombreuses personnalités extérieures nous ont, tout au long de ces recherches, manifesté leur intérêt et accordé leur soutien :

Monsieur de CRECY, Chef de la Division Nivologie du C.T.G.R.E.F. ne nous a pas ménagé son aide. Messieurs DELSOL puis MARBOUTY, directeurs du Centre d'Etudes de la Neige nous ont apporté la précieuse collaboration de la Météorologie Nationale. Monsieur le Professeur LLIBOUTRY, directeur du Laboratoire de Glaciologie, a toujours encouragé nos travaux et le développement des méthodes statistiques en Géophysique.

Messieurs MORLAT, Conseiller Scientifique à EdF, PAGES du CEA Saclay et CAZES de l'ISUP, nous ont encouragés à introduire l'Analyse des Données dans les sicences de la terre. J.P. DELHOMME du Centre d'Informatique Géologique de Fontainebleau, nous a initié avec beaucoup de patience aux variables régionalisées et l'équipe de M. DIDAY nous a grandement facilité l'accès à ses programmes.

Nous ne saurions oublier non plus la part ingrate qui revient aux responsables des Services de Calcul, Messieurs LAGARDE et RABATEL au Laboratoire, ou Messieurs GARCIA, LESPINASSE et bien d'autres au C.I.C.G. En butte à notre impatience et parfois à notre mauvaise humeur, réparant sans rancune les conséquences de nos erreurs, qu'ils acceptent ici ma très sincère reconnaissance.

Je remercie aussi Madame RICCIARDELLA, secrétaire de notre groupe pour sa compétence, son dévouement et sa gentillesse, Madame ELBERG a assuré avec beaucoup de soin la dactylographie du manuscrit, les dessinateurs du Laboratoire de Cartographie Climatique m'ont secondé pour la réalisation des figures et Messieurs BRAULT, PERROUX et DIOT en ont assuré le tirage.

"There are three kinds of lies :
lies, damned lies, and statistics ..."

DISRAELI

INTRODUCTION

Un titre, si long soit-il, ne résume qu'imparfaitement les pages qu'il recouvre, et les travaux en Analyse de Données se montrent toujours rebelles aux classifications habituelles ...

Nous ne sommes en effet ni des statisticiens, ni des informaticiens, et les phyciciens ne nous reconnaissent guère ... Les premiers voient en nous des néophytes peu soucieux de l'orthodoxie et ignorants des Tables de la loi, quelle qu'elle soit. Les seconds nous reprochent notre manque d'allégeance à la "Machine" et un irrespect fréquent pour le "Système". Quant aux derniers, ils restent partagés entre une certaine sympathie, pour ces amateurs fort curieux de comprendre les phénomènes naturels, et une réserve certaine, quant à la façon sommaire dont nos modèles traitent ces mêmes phénomènes.

Nous empruntons pourtant à ces trois disciplines mais en gardant à l'e prit que, pour résoudre un problème plus ou moins bien défini mais concret, nous disposons d'un volume d'informations donné et limité. Un des pionniers de l'Analyse des Données, John TUKEY, a d'ailleurs bien précisé notre but, dans un article prophétique intitulé "Future of Data Analysis". Considérant la capacité croissante des moyens de calcul, il souhaitait que des chercheurs, "plus intéressés par les sciences que les mathématiques, plus soucieux d'efficacité et d'utilité que de sécurité, plus intéressés par des techniques qui suggèrent que par celles qui concluent, et acceptant de se soumettre autant au jugement des faits qu'à la rigueur mathématique, concentrent leur attention sur les méthodes d'analyse des données et l'interprétation des résultats d'analyses statistiques ..." Ce besoin est particulièrement ressenti en Hydrométéorologie où l'effort des dernières decennies dans l'instrumentation et la collecte des informations conduit à des amoncellements de données parfois inquiétants. Un éminent météorologue anglais, J.M. CRADDOCK rappelait d'ailleurs, devant la "Royal Meteorological Society" que : "while we are adding to our background material as the years go

by, we also have the task to squeeze the last drop of information from the data we have ... " (1973 Annual Meeting).

C'est ce que nous tenterons de faire, dans les pages qui suivent, à propos de deux ensembles de données très représentatifs de ceux que l'on rencontre en Hydrométéorologie : Condenser l'information, en extraire le maximum, optimiser éventuellement le système de mesure, et mesurer son potentiel de prévision. Le premier exemple concerne la prévision des phénomènes "accidentels" connus le plus souvent de manière qualitative (occurence d'orages à grêle, nature des précipitations : pluie ou neige, risque de pollution, etc ...) parmi lesquels le problème de la prévision des avalanches n'avait pas encore fait l'objet d'un effort intensif. Un autre problème, très courant en Hydrologie, concerne les variables "régionalisées", c'est-à-dire mesurées sur un réseau où la proximité géographique entre stations a un sens et doit donc être prise en compte. L'exemple du réseau pluviométrique des Cévennes est aussi intéressant par sa densité que spectaculaire par l'importance des précipitations qu'il mesure.

Au passage, nous évoquerons, non pas l'ensemble des méthodes existantes en Analyse des Données, mais celles que nous avons personnellement utilisées, analysées, voire améliorées sur ces exemples. Grâce en particulier aux travaux de Pr. BENZECRI et de ses collaborateurs, l'Analyse des données n'a plus rien aujourd'hui d'un domaine grossièrement défriché, mais c'est encore un jardin à l'anglaise où de nombreuses zones d'ombre subsistent ... Nous avons cherché à en éclaircir quelques-unes.

Ces travaux constituent aussi une étape dans les recherches de notre équipe d'Hydrométéorologie : Ils font suite à la thèse de Ph. BOIS (1976) qui donnait un traitement très complet des méthodes de critiques des données et un premier ensemble de techniques de prévision,s'appuyant essentiellement sur la regression multiple. On retrouvera ici la même démarche en 3 temps :

- analyse théorique d'une technique et recherche des analogies pouvant enrichir son interprétation ;
- utilisation sur des données simulées et recherche ou vérification de propriétés que la spéculation théorique n'a pas forcément mises en évidence ;
- puis application à des exemples réels.

Sur un plan pratique, ce document est divisé en 5 parties pratiquement indépendantes (Un certain nombre de démonstrations accessoires ont été reportées dans une annexe qui sera publiée séparément). Exceptionnellement, dans l'interpolation optimale de variables régionalisées (Vème partie), nous ne présentons pas d'exemples réels traités : ceux ci ont été repris dans la thèse de

D. CREUTIN (1979) et développés dans une comparaison plus large des différentes méthodes d'interpolation utilisées pour les champs de précipitations.

Quant à la longueur de ce mémoire, elle ne s'explique pas seulement par la discussion, toujours longue, des résultats mais aussi par la difficulté de trouver un fil conducteur à des travaux entrepris indépendamment. Paraphrasant Pascal, reconnaissons aussi que le temps nous a manqué pour présenter un texte plus court ... mais un certain nombre de paragraphes (signalés par des astérisques) peuvent être omis en première lecture.

TABLE DES MATIERES

"Il y a une foule d'accidents qu'on ne peut prévoir : la grêle, la gelée quelquefois, la sécheresse, les pluies excessives ..."

Xénophon (V°siècle Av.J.C.)

L'Economique-Livre IV

PREMIERE PARTIE

LES PROBLEMES ET LEURS DONNÉES .

(PRÉTRAITEMENTS ET CODAGES)

On y décrit en détail les 2 problèmes qui ont servi de thème à nos travaux. Le premier chapitre précise les objectifs, les données disponibles et l'état antérieur de ces travaux d'après les références bibliographiques disponibles.

La nature extrèmement variée des variables dont nous disposons nous contraint dans le chapitre II à quelques réflexions théoriques sur ces variables, la mesure de leur liaison et l'effet de diverses transformations.

Le chapitre III en est l'application immédiate aux données que nous utiliserons par la suite .(données nivométéorologiques de Davos et précipitations sur les régions cévenoles).

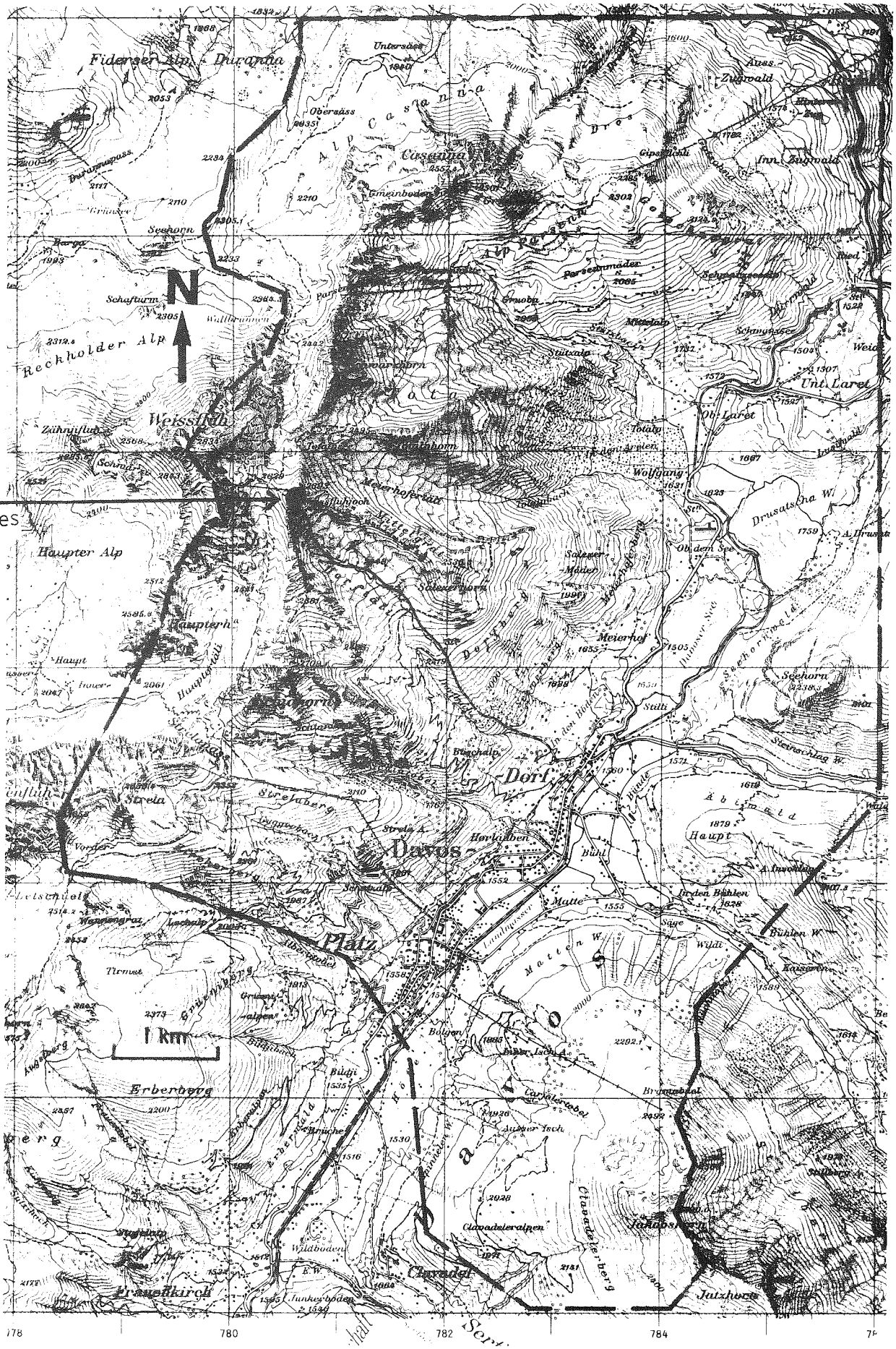


FIGURE I - 1 : Carte de la région de la Parsenn (Davos - SUISSE) .

CHAPITRE I

PROBLEMES ABORDES ET DONNEES DISPONIBLES

I.1 - La prévision des avalanches - Exemple de Davos

Ce problème de prévision d'un phénomène accidentel nous a été confié par l'Institut Fédéral Suisse pour l'Etude de la Neige et des Avalanches.

I.1.1. Les objectifs

Il s'agissait de mettre au point un schéma de prévision du danger d'avalanche fondé sur une approche numérique objective et sur les seules données observées. Il serait complémentaire de celui qui existe actuellement et repose uniquement sur l'expérience et l'intuition d'un prévisioniste très compétent, mais difficile à remplacer.

Précisons cependant un certain nombre de points :

a) Il ne s'agit pas en fait d'une prévision, mais, selon LACHAPELLE (1977) d'une évaluation du danger d'avalanche, compte tenu des conditions météorologiques du moment (supposées connues ou prévues de façon parfaite). Nous laissons donc de côté le problème de la prévision de ces variables nivométéorologiques.

b) Le problème peut ensuite se poser à différentes échelles d'espace et de temps.

Les méthodes diffèreront sensiblement selon que l'on veut évaluer le danger à l'échelle d'un pays entier (l'état du Colorado pour le Forest Service de Fort Collins, les Alpes et les Pyrénées françaises pour le Centre d'Etude de la Neige de St Martin d'Hères, etc...),d'un itinéraire particulier (la Transcanadian Highway dans les Rocheuses canadiennes, la Highway 550 dans les San Juan Mountains du Colorado, la route desservant la mine "ANDINA" près de Portillo au Chili, les voies ferrées menant aux mines d'uranium dans les montagnes d'Asie Centrale, diverses routes stratégiques du Tien Shan ou du Tibet, etc...), ou d'une petite région (vallée, parc naturel, etc...).

Dans chaque cas, le domaine géographique présente certaines particularités qui influeront sur la méthode de prévision. Dans notre exemple, nous nous limiterons à des zones relativement petites ($\approx$ 100 km^2) correspondant essentiellement au domaine skiable d'une station.

En ce qui concerne l'échelle de temps, certaines demandes de prévision, sont à l'échelle de l'année (pour les forestiers chargés du reboisement),mais plus souvent à l'échelle de quelques jours (le "week-end"),voire de quelques heures (quelle est l'heure la plus propice à un déclenchement artificiel ? etc...) Quant à nous, nous considérons le danger à l'échelle de la journée, ou plutôt des 12 heures diurnes du jour j, connaissant les conditions de ce jour j.

c) Un autre problème concerne la notion de danger et de son intensité. On peut en effet se demander si une avalanche de 200 m d'extension est moins dangereuse qu'une avalanche de 400 m et si une journée avec 10 avalanches observées est 2 fois plus dangereuse qu'une journée avec 5 avalanches. Enfin, si une seule avalanche se produit par beau temps dans un domaine ouvert au public, n'est-ce pas plus dangereux que 10 avalanches dans une station"fermée", et ne faut-il pas considérer avec circonspection toutes les zones d'exposition similaire, voire l'ensemble du domaine.

On pourra voir, dans la revue des travaux antérieurs, que certains auteurs ont cherché à prendre en compte cette notion d'intensité de l'activité avalancheuse.

d) Des problèmes se posent aussi au niveau des échantillons statistiques quant à des biais éventuels. En effet les facteurs physiographiques (pente, exposition, végétation) ne sont pas pris en compte. On suppose, par exemple, que toutes les orientations sont représentées en proportions sinon egales du moins suffisantes dans la region que nous considérons. (Si l'on ne dispose que de versants Est, il se peut qu'un jour donné, la situation soit considérée comme sûre, alors qu'elle serait dangereuse sur le premier versant Ouest venu). Enfin, on admet que toute avalanche descendue ne modifie en rien le potentiel avalancheux des jours suivants et qu'il se trouvera toujours une zone de caractéristiques voisines et non encore purgée. Cette considération est au contraire de première importance quand on prévoit au niveau d'un seul couloir. (Celui-ci étant d'autant plus sûr qu'il vient d'être purgé, même si les zones voisines sont encore très dangereuses).

Nous supposons donc la zone géographique choisie suffisamment vaste pour constituer une population où toutes les conditions physiographiques sont représentées. Par contre elle est supposée suffisamment petite et compacte pour que les variables nivométéorologiques mesurées ou calculées ponctuellement soient représentatives de l'ensemble de la zone.

Remarque . Ces hypothèses ne seraient pas vérifiées pour certaines zones apparemment réduites : les 2 côtés d'une route sur 100 km de long, ou pour certaines variables : les mesures de résistance du manteau ont une représentativité beaucoup plus locale (quelques dizaines de mètres).

e) Toujours au niveau des échantillons statistiques, des problèmes se posent quant aux avalanches qu'il nous faut considérer. En effet, dans les zones où la prévision présente un grand intérêt, on procède généralement à des déclenchements préventifs (par explosifs ou à ski). Pour notre part, nous avons considéré que le danger était matérialisé par des avalanches "naturelles". Or, il serait souhaitable, ne serait-ce que pour réduire l'incertitude d'observation de ces dernières, de ne travailler que sur les avalanches artificielles. Mais pour ce faire, il faudrait procéder de façon systématique, c'est-à-dire tirer des explosifs en de nombreux endroits et tous les jours sans distinction. On pourrait alors supposer que l'on a une bonne mesure de la stabilité du manteau, un peu pessimiste en ce qu'elle anticiperait légèrement les déclenchements naturels, et cette démarche serait idéale pour le traitement statistique.

Le problème est que l'on tire de manière trop épisodique : des avalanches naturelles apparaissent alors que l'on n'a pas jugé utile de tirer. Des journées classées sans avalanche auraient pu donner lieu à des déclenchements si l'on avait tiré, etc... Sans parler des avalanches qui se produisent "quelques temps" après que l'on ait tiré ... cf K. WILLIAMS (1978).

f) Enfin, notre but est de fournir une prévision continue, "de routine", chaque jour de la saison d'hiver, et non pas dans les seules situations "catastrophiques". Dans le cas de précipitations importantes par exemple, il est presque plus intéressant d'expliquer pourquoi il ne se produit pas, ou pas plus d'avalanches. Par contre, pour un type de situation considéré comme peu avalancheux, il est important de déceler si une avalanche isolée risque de se produire.

Ces raisons font que la prévision est présentée ici comme le pourcentage de probabilité, pour chaque jour donné, qu'une avalanche naturelle au moins apparaisse, dans les 18 H à 24 H considérées, sur la région étudiée. Celle-ci est la région de la Parsenn, qui comporte environ 100 km^2 autour du célèbre Institut du Weissfluhjoch (2670 m). La station de mesures nivométéorologiques, située à 2450 m est particulièrement représentative des zones de départ des avalanches (2000 - 3000 m) cf. figure I-1.

I.1.2. Travaux antérieurs (*)

a) Historique

Une synthèse partielle a déjà été présentée par E. LACHAPELLE (1976) mais elle pêchait un peu par ignorance en ce qui concerne les travaux menés en Union Soviétique. Une abondante bibliographie, malheureusement incomplète sur le sujet de la prévision a aussi été publiée par le "World Data Center A for Glaciology" en 1977.

Sans prétendre connaître l'ensemble des travaux effectués en ce domaine, nous nous proposons de les compléter à travers la chronologie suivante :

Ere Primaire : La prévision reposait entièrement sur l'expérience et l'intuition du montagnard, avec tous les biais que pouvaient introduire la transmission orale et les incertitudes concernant les zones peu fréquentées.

Bien que ces pratiques soient encore en vigueur dans certaines régions, l'observation systématique des avalanches a probablement commencé en France, dès le début du siècle, avec le réseau d'observations mis en place dans les Alpes par l'ingénieur P. MOUGIN, pionnier de l'hydrologie et de la restauration des terrains en montagne.

Ere Secondaire : Elle a consisté à mettre en regard, mais sans proposer de modèle, les avalanches observées et les phénomènes physiques qui pouvaient leur être associés. Même si divers observateurs ou géographes avaient bien suggéré ces rapprochements, mais sans les quantifier vraiment (cf. ALLIX A., 1925), les précurseurs semblent être essentiellement russes. On citera AKKURATOV V.N. qui, dès 1956, proposait d'utiliser la quantité de neige transportée par le vent et la température de contraction (?) de la neige comme index de danger. K.S. LOSEV en 1960, prenait les mêmes index

(accumulation de neige par précipitation et/ou transport). En 1963, ABDUSHELISHVILI K.L. tentait de déterminer pour le Caucase une quantité critique de neige fraiche X_{75} après que l'humidité relative de l'air ait dépassé 75%, choix qui fut très controversé ensuite ! Comme l'intérêt portait beaucoup sur l'heure de début du danger, une attention particulière fut accordée à l'intensité de la chute de neige, que nous retrouverons dans des modèles plus récents.

En 1965, le Symposium de Davos, organisé par l'A.I.H.S. permit de faire le point à un niveau international. C'est là quapparaissent les travaux de T.ZINGG, et de A. POGGI et P. PLAS, ces derniers essayant de quantifier la liaison entre le nombre d'avalanches observées dans une région des Alpes et les précipitations cumulées durant les 3 jours précédents.

Parallèlement à cette approche directement prévisionnelle, de nombreux travaux plus fondamentaux tentaient d'éclaircir la mécanique du manteau de neige. Il faut citer le travail de pionnier de G. SELIGMAN (1936), les nombreuses contributions de l'Institut Fédéral du Weissfluhjoch à Davos, et d'un certain nombre d'auteurs russes. On a bien sûr tenté d'appliquer ces résultats en prévision : il "suffisait" de prévoir quand les contraintes exercées sur le manteau excédaient sa résistance propre, ou son adhérence à la couche inférieure ... On trouve ainsi les travaux de SCHERBAKOV M.P. (1966), de KOSAREV M.V. (1969) et des préoccupations voisines chez ROCH A. (1965) ou quelques chercheurs américains (rassemblé dans MELLOR M., 1968), mais ces approches déterministes restèrent peu fructueuses jusqu'à ce jour.

Ere tertiaire : Celle-ci a vu apparaître les techniques statistiques comme outils pour quantifier la prévision. En 1969, KOSAREV M.V. définissait "La décision quant au moment où il y a danger d'avalanche (c'est-à-dire la probabilité d'avalanche pour le jour à venir sans préciser la localisation des coulées)..." . En Occident, R. PERLA (1970) commençait à exprimer le danger d'avalanche en terme de probabilité d'occurence et Ch. OBLED (1970) utilisant ce même index, ajustait pour la première fois un modèle statistique (Probits) entre la probabilité d'avalanche et les précipitations récentes. En 1972, P. BOIS et Ch. OBLED utilisèrent pour fournir la probabilité d'avalanche, une analyse discriminante avec plusieurs variables explicatives, utilisable en continu et non dans les seules périodes de précipitations. Leurs travaux ultérieurs (1973, 1974, 1975, 1976) sont décrits dans ce mémoire.

Leur approche fut reprise par d'autres chercheurs comme JUDSON A. et ERIKSON B.J. (1973) ou BOVIS M.J. (1977) avec des modifications de détail et quelques améliorations. Mais à notre avis, les connaissances actuelles sur les mécanismes de mise en condition du manteau et de déclenchement proprement dit sont encore trop mal connues et controversées pour que l'ère quaternaire commence à poindre.

Dans le futur, la compréhension de ces mécanismes étant quasi-complète, le choix entre des méthodes déterministes ou statistiques sera dû uniquement à des considérations d'emploi (modèles explicatifs ou opérationnels) et de disponibilité de données. Nous en sommes encore loin ...

b) Travaux récents

Parmi les tendances actuelles, on admet desormais que :

- Il est necessaire de distinguer differents types d'avalanches (en general avalanche de neige sèche et avalanche de neige humide)
- Il semble inconséquent de distinguer les journées à avalanche de neige sèche de celles à avalanches de neige humide pour considérer ensuite que les journées sans avalanches constituent une population homogène. On est donc conduit à les décomposer aussi et à définir des types de temps.
- Il semble nécessaire de prendre en compte le degré d'activité avalancheuse.

C'est ainsi que BOVIS M.J. (1977) utilise une stratification des journées avalancheuses en constituant son échantillon avalancheux de 4 façons différentes.

I - Toutes les journées à avalanches naturelles ou artificielles

II - Journées à avalanches naturelles (au moins 1) de magnitude 2 (cotation du U.S. Forest Service)

III - Journées à avalanches naturelles avec au moins 3 cas de magnitude 2

IV - Journées à avalanches naturelles de magnitude 3.

Bien que fondé sur des échantillons petits, il semble qu'il faille prendre un seuil d'activité minimum pour construire l'échantillon d'ajustement.

Si l'utilisation de l'analyse discriminante à 2 groupes (avalanche de neige sêche (resp. de neige humide) contre journée sans avalanche) semble devenue classique, elle est surtout utilisée pour un type bien précis de situation : les périodes avec fortes chutes de neige. C'est le cas de nombreux travaux russes (GRAKOVITCH V.F., ou KHOMENIOUK et al., in TOUCHINSKI, 1974, DROSDOVSKAYA, 1978, et divers articles dans les Compte-rendus de l'Académie des Sciences de l'U.R.S.S., Comité de Géophysique. N°31, Sept. 1977).

La prise en considération des autres types de temps est plus récente : BOIS et OBLED (1976) utilisent des méthodes objectives de classification en types de temps tandis que YEFIMOV M.K. et KOZIK Y.M. (1975) utilisent une classification empirique. Cette approche par type de temps est aussi proposée par GRIGORIAN S.S. (1974) dans une revue des Travaux de la Faculté de Géographie de Moscou. Au sein des types de temps, on peut alors procéder à une analyse discriminante avalanche/non avalanche (BOIS et OBLED) ou effectuer des regressions entre le nombre d'avalanches et des variables nivo-météorologiques (YEFIMOV et KOZIK). Enfin, on peut toujours au sein d'un type de temps, chercher à étudier les seules avalanches catastrophiques, atteignant le fond des vallées (P. FOHN, 1978).

Au lieu de chercher à prévoir la probabilité d'occurence, on peut aussi s'intéresser au nombre d'avalanches ou comme SALWAY A.A. (1976) à un index complexe combinant le nombre d'occurences, leurs extensions respectives, etc... Au lieu d'utiliser la régression habituelle, SALWAY A.A. traite ses observations comme la série chronologique d'un processus stochastique, mais sans admettre qu'il puisse y avoir plusieurs processus.

A l'heure actuelle, la plupart des équipes de recherches citées (Forest Service (Fort Collins), Université de Washington (Seattle) et de Colombie Britannique (Vancouver), Institut du Weissfluhjoch à Davos, Faculté de Géographie à Moscou avec ses 2 centres du Caucase (Elbrouz) et des Khibins (près de Mourmansk), Institut de recherches hydrométéorologiques d'Asie Centrale (Tachkent) poursuivent activement leurs travaux.

Note : On trouvera une bonne synthèse des diverses causes physiques des avalanches dans le "Avalanche Handbook" R.I. PERLA et M. MARTINELLI - USDA - Forest Service-Agriculture Handbook n°489 - 1973.

I.1.3. Données disponibles

a) Les données d'avalanches

Elles sont de deux types :

Avalanches déclenchées naturellement

On donne la date de l'observation (et non du déclenchement) car l'observateur parcourt certes tous les jours la région, mais certaines conditions (brouillard, blizzard) rendent l'observation impossible et des avalanches ne sont donc signalées que quelques jours après. Malheureusement, ces conditions défavorables sont assez fréquentes en période avalancheuse et il y a fréquemment des incertitudes de 1 à 2 jours sur les dates de déclenchement. Enfin l'observateur ne pouvant parcourir toute le région (topographie, remontées mécaniques disponibles ou non), certaines avalanches, très peu en fait, passent inaperçues. Par contre, on peut remarquer que la zone observée est énormément variable d'un jour à l'autre, ce qui est surtout gênant pour les journées faiblement avalancheuses qui peuvent être classées avalancheuses ou non selon la superficie observée.

Avalanches déclenchées artificiellement

C'est-à-dire celles déclenchées par grenades, lance-mines, roquettes, mais aussi par des skieurs ou des engins de damage. On connait généralement l'heure du déclenchement. Malheureusement, les essais de déclenchement ne sont pas systématiques à la fois dans le temps (on ne tire pas tous les jours et même si la situation semble avalancheuse) et dans l'espace (on ne tire que sur certains couloirs ou pentes particulières).

Enfin ces données ne sont archivées que depuis 1969-70 de façon exploitable.

Codage des avalanches

Chaque avalanche est affectée d'un code à 5 chiffres (cf. page suivante).

Seul les codes 4 et 5 seront utilisés dans notre étude. Les codes 3 n'apparaissent qu'en fin de saison, en dehors même des périodes que nous avons considérées (Décembre à Avril). Enfin les codes 0 et 1 se rencontrent souvent conjointement pour diverses avalanches de la même journée. Ils pourraient toutefois être assez utiles dans l'avenir.

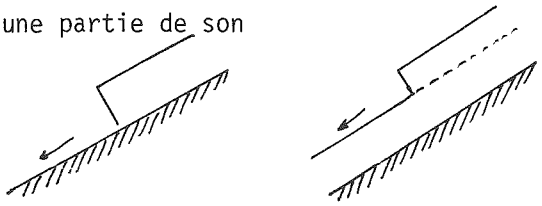
Description des codes :

0 : zone de déclenchement présentant une ligne de fracture

1 : départ ponctuel

2 : mise en mouvement du manteau sur une partie de son épaisseur

3 : glissement du manteau complet sur le sol



4 : la neige de l'avalanche dans la zone d'arrivée est sèche

5 : la neige de l'avalanche dans la zone d'arrivée est humide

6 : avalanche sur une pente

7 : avalanche de couloir

8 : écoulement avec aérosol (nuage de poudreuse)

9 : écoulement au sol

b) Données nivométéorologiques "brutes"

Il s'agit des paramètres nivométéorologiques effectivement mesurés sur le champ d'observation 2540 m ou à l'Institut (2670 m). (cf. Table 1)

Bien que dans des études antérieures nous ayons utilisé des données remontant à 1954, nous nous limiterons ici à la période postérieure à 1960 (hiver 1960-61 à 1973-74) pour des raisons d'homogénéité ou de disponibilité de certaines séries.

c) Variables explicatives "élaborées" (cf. TableII)

Il s'agit ici d'une étape qui soulève généralement des débats passionnés et dont le but est le suivant :

Construire à partir de variables brutes (qui ne sont pas toujours directement liées au phénomène avalanche), un ensemble d'index élaborés qui représenterait mieux les divers mécanismes physiques possibles.

Cette construction permet à notre avis d'introduire un bon nombre de considérations physiques, d'où apport d'information, et de construire des index mettant en jeu des seuils ou des fonctions non linéaires des variables brutes, ce qui est une façon détournée (avec les traitements par types de temps) de traiter les non-linéarités.

N° (article J.of G.)

1 Equivalent en eau de la couche de neige fraîche (exprimé en 1/10 de mm) mesuré à 08 H le jour j, donc tombé entre j-1 à 08 H et j à 08 H

2 Hauteur de neige fraîche (exprimée en mm) lue à 08 H le jour j à la planche à neige

3 Précipitation la plus importante de la journée (tempête ou "snowstorm"), ramenée à sa durée réelle intensité moyenne exprimée en mm/h. On affecte les chutes de neige, tombées entre j à 08 H et j+1 à 08 H, au jour j (Intensité associée à la période d'observation des avalanches du jour j).

4 Vent maximum du jour.
. de 1961 à 1969 : vent maximum instantané entre 0 et 24 H le jour j (sauf pannes 1962 et 1966)
. de 1970 à 1974 : maximum des 3 mesures (7 H 30, 13 H 30 et 21 H 30 ou 19 H)
On dispose maintenant des vents maximum instantanés pour la période 1970-74.

5 Rayonnement global mesuré en cal/cm^2 le jour j au pyranomètre.
(Certaines corrections ont été apportées pour corriger les détarages et des changements d'appareil).

6 Température à 7 H le jour j en 1/10 de °C.

7 Température à 13 H le jour j en 1/10 de °C.

8 Nombre d'heures d'insolation du jour j.

9 Nébulosité moyenne du jour j en dixièmes.

10 Hauteur de neige lue à la perche le jour j vers 09 H.

11 Pénétration de la sonde de battage sous son poids propre.

12 Température à -10 cm dans la neige (exprimé en valeur absolue et 1/10 de °C).

13 Equivalent en eau, lu à 08 H le jour j, de la précipitation recueillie dans le pluviomètre enregistreur (influencé par le vent et l'appareil).

14 Azimuth associé au vent maximum .
. de 1961 à 1971 : azimuth associé à la plus forte des valeurs de vent de 7 H, 13 H, 21 H.
. de 1972 à 1974 : azimuth associé au vent horaire maximum (≠ vent instantané)

15 Nombre d'observations de "chasse-neige)
Si pas de chasse-neige 0
Si chasse neige à 7 H et/ou 13 H et/ou 19 H 1, 2 ou 3
(Possibilité d'ajouter +1 si l'observateur a indiqué qu'il y en a eu beaucoup)

TABLE I - description des variables "brutes" .

N° Fichier 76	
1	Précipitation (08 H j-1 à 08 H j) + 1/2 (08 H j à 08 H j+1)
2	Hauteur de neige fraîche
3	Intensité
4	Densité de la couche de surface
5	Cumul des précipitations (sur 3 jours)
6	Vent du jour
7	Ecart en pourcentage entre valeur en eau au pluviomètre et à la planche à neige
8	Chasse neige de j et j-1
9	Vent * cos (-(AZ-45)) axe SW - NE
10	Vent * sin (-(AZ-45)) axe SE - NW
11	Accroissement du vent pendant ou après les précipitations
12	Température à 13 H
13	Variation des températures à 13 H entre j-1 et j
14	Température à 13 H + 3°C si < 0°
15	Température à 13 H + 3°C si > 0°
16	Température à 7 H
17	Variation des températures à 7 H entre j-1 et j
18	Nombre d'heures d'ensoleillement du jour j
19	Rayonnement solaire incident du jour j
20	Nébulosité moyenne du jour j
21	Pourcentage reçu du rayonnement potentiel du jour j
22	Nombre de séquences de précipitations
23	Quantité de précipitations depuis la fin de la dernière séquence
24	Nombre de jours sans précipitation depuis la fin de la dernière séquence
25	Nombre de jours sans précipitation avant la séquence précédente
26	Maximum enregistré depuis le début de l'hiver de la var.25
27	Nombre de cas de chasse-neige depuis le début de la séquence en cours
28	Nombre de jours où au moins 1 cas de chasse neige a été observé depuis le début de la séquence en cours
29	Précipitations pondérées cumulées (Storm Index de Judson)
30	Hauteur de neige fraîche "efficace"
31	Tassement depuis le dernier maximum d'épaisseur
32	Tassement relatif depuis le dernier maximum d'épaisseur
33	Tassement entre j-1 et j (si tassement)
34	Enfoncement de la sonde de battage
35	Durcissement entre j-1 et j
36	Cumul sur 5 jours des heures d'ensoleillement
37	Cumul sur 5 jours du rayonnement incident
38	Cumul sur 5 jours des degrés jour > 0°
39	Cumul proj. vent sur NE sur 3 jours
40	Cumul proj. vent sur SW sur 3 jours
41	Cumul proj. vent sur NW sur 3 jours
42	Cumul proj. vent sur SE sur 3 jours
43	Température 10 cm sous la surface de la neige
44	Variation de cette température entre j-2 et j
45	Variation de cette température entre j-1 et j
46	Variation relative de la température entre j-2 et j
47	Cumul sur 7 jours de cette température
48	Variations pondérées entre j-2 et j des températures à 7 H
49	Nombre de journées avalancheuses depuis le début de l'hiver
50	Nombre de journées avalancheuses depuis le début de l'hiver ramenées au nombre de séquences de précipitations.

-TABLE II - Description sommaire des variables "elaborées" .

Exemples

. La quantité de neige tombée en 24 H a moins d'importance que la quantité totale tombée depuis le début d'une séquence de précipitations.

. La hauteur totale de neige au sol a moins d'importance que ses variations d'un jour à l'autre.

. Le rayonnement solaire incident est un bon indice du type de temps mais pas du bilan thermique de la neige si l'on n'y ajoute pas l'albedo.

etc...

Nous ne donnons qu'une description sommaire de ces variables car les justifications seraient particulièrement volumineuses. Pour l'essentiel, nos choix recoupent ceux décrits par S.S. GRIGORIAN (1974). Par contre, nous avons cherché à être exhaustifs, en proposant le maximum de variables plausibles et en laissant à des algorithmes ultérieurs le soin de les sélectionner.

Toutes ces variables font l'objet d'un calcul automatique par programme ce qui exclut toute subjectivité.

Remarque : Certaines variables sont cumulées depuis le début de l'hiver, défini comme : "le début de la période sans précipitation suivant immédiatement la chute de neige ayant provoqué la formation d'un manteau saisonnier au champ de mesures (2450 m)"

On constate que ces variables peuvent se regrouper en 3 classes :

- Influences mécaniques : Précipitations, transport par le vent. Elles indiquent en partie aussi la situation atmosphérique (temps perturbé et venteux)

- Influences thermiques : Rayonnement, températures, qui influent sur le bilan thermique du manteau et indiquent aussi l'état thermique de l'atmosphère.

- Etat du manteau : surface (résistance, tassement, etc..) ou en profondeur (nombre de couches, état des intercouches, etc..) C'est assurément là qu'apparaissent les index les plus grossiers, mais peut-être les plus représentatifs spatialement.

Car si le reproche nous a souvent été fait de négliger les données bimensuelles de sondage du manteau, personne n'a pu nous indiquer clairement comment les traiter. (KOSAREV M.V. (1969), parle des sondages et de leur intérêt en ces termes :

"Cette situation incongrue ... s'explique par une obsession excessive quant aux propriétés physico-mécaniques de la neige ..." "Ces propriétés ... ne peuvent être utilisées pour la prévision sur des superficies assez grandes ..." (sous-entendu par rapport à la représentativité du sondage) et il conclut "qu'elles doivent certes être étudiées, mais pour d'autres raisons" ...)

Le débat reste donc ouvert entre les physiciens de la neige ..., et nous nous garderons d'y entrer.

En conclusion sur ces données, il est certain que ce sont les données d'avalanches qui sont les plus entachées d'incertitude et qu'en comparaison, les problèmes de précision ou de représentativité des variables explicatives semblent mineurs. Signalons pourtant que certaines séries (vent maximum, température de la neige, rayonnement) incomplètes ou partiellement erronées ont dû être reconstituées, avec plus ou moins de précision.

Note très importante sur les variables utilisées

Ces études s'étant étalées sur plusieurs années, un premier ensemble de 50 variables explicatives fut utilisé jusqu'en 1975. Leur description complète se trouve dans l'article de la Houille Blanche (BOIS Ph. et OBLED Ch., 1976). Certaines variables étant apparues inutiles ou redondantes, on les a remplacées par d'autres d'où l'ensemble final de 50 variables décrit plus haut.

Toutefois, nous donnons un certain nombre de résultats utilisant le premier ensemble (la plupart des analyses descriptives) et qui n'ont pas été repris avec l'ensemble final. Dans la mesure du possible, nous précisons bien sur quel ensemble a été faite l'analyse. La correspondance d'indices entre les deux ensembles est donnée dans le tableau

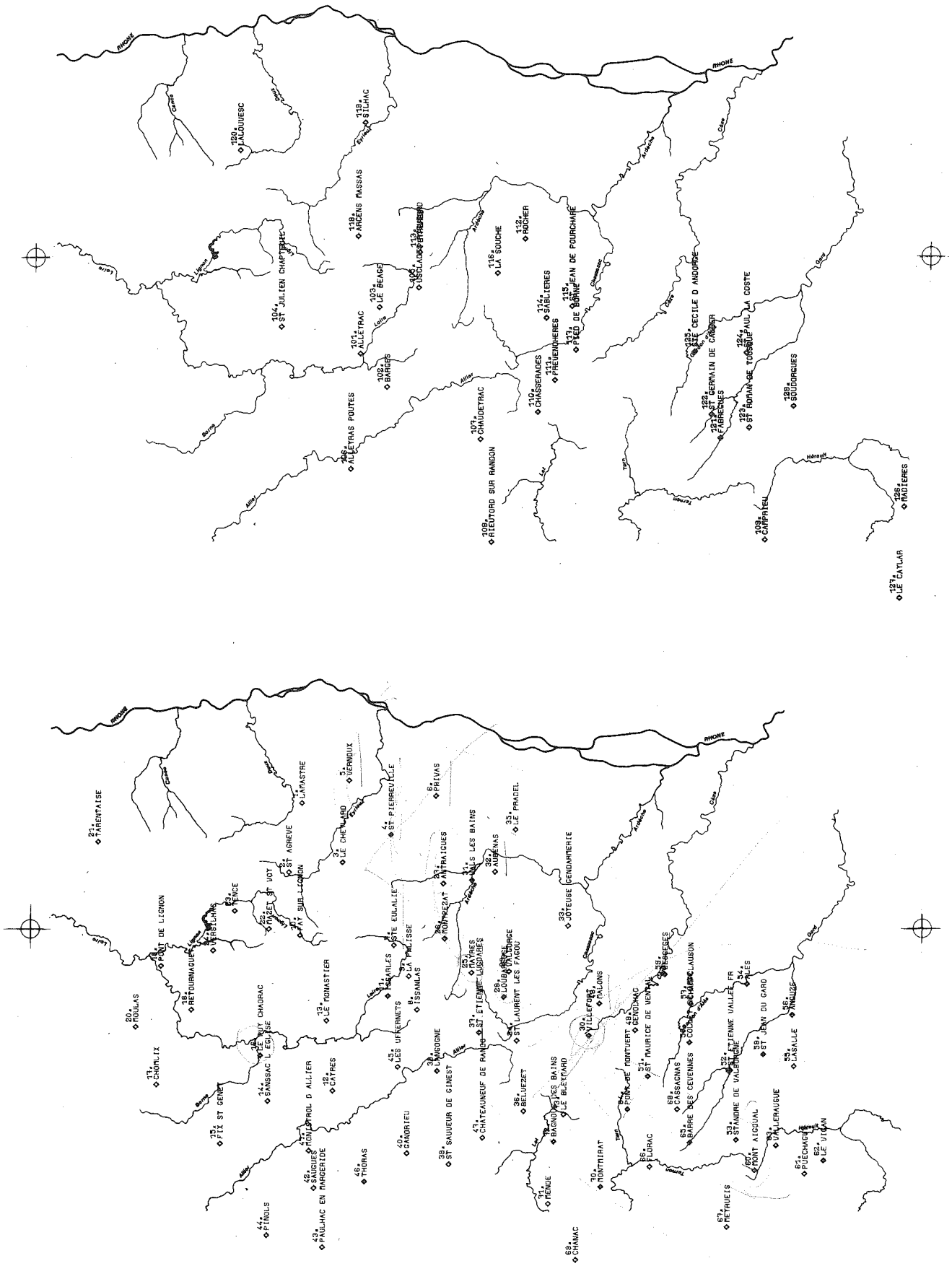


Figure I-2- localisation geographique des stations du reseau de base (73 stations) ,et du reseau test (28 stations) .

I.2 - Analyse des précipitations sur la bordure Sud-Est du Massif Central (Episodes Cévenols)

I.2.1. Problèmes posés

Notre groupe de recherche s'intéresse de longue date à l'hydrologie de ces régions. Après des études sur la propagation des crues en rivières, puis sur la formation des crues et l'étude des relations pluies-débits, il était normal que, remontant encore vers l'amont, on se pose le problème de la structure des pluies sur la région.

En effet, les modèles mathématiques pluie-débit utilisent, en entrée, la lame d'eau moyenne sur une parcelle ou un sous-bassin. Or c'est une donnée critique car, le modèle déterministe étant conservatif, toute la lame introduite doit transiter dans le modèle et une estimation erronée provoque des excédents ou des déficits systématiques, qu'il faut donc éviter. (Remarquons que ce n'est pas le cas dans les relations statistiques pluies-débits qui sont auto-adaptatives et n'utilisent la pluie que comme index).

Il se pose donc un problème d'interpolation optimale des valeurs ponctuelles de précipitations, ou de leur intégrale spatiale. En général, le réseau n'est pas assez dense, ou comme dans le bassin des Gardons, seule une partie est télémesurée et disponible immédiatement: Il faut donc reconstituer le champ à l'aide des seules stations disponibles. (Fig.I-2)

Comme dans le problème précédent, le déroulement de l'étude comportera d'abord l'analyse structurale du phénomène sur les données du passé, la mise au point de certains modèles puis leur application à des données test. Toutefois cette étude ayant donné lieu à des développements théoriques intéressants dont les applications pratiques sont encore en cours de mise au point (D. CREUTIN. Thèse d'Ingénieur-Docteur, à paraître en 1979) nous présenterons surtout les aspects méthodologiques suffisamment développés à l'heure actuelle.

I.2.2. Travaux antérieurs

Nous distinguerons deux types de travaux. Les premiers concernent les études climatologiques et météorologiques effectuées sur la région des Cévennes proprement dite. Les seconds rassemblent les essais d'analyse structurale des pluies, effectués en diverses régions du monde.

a) Aspects météorologiques des pluies sur la bordure Sud-Est du Massif Central

Cette région, connue de longue date pour les crues violentes qu'elle subit (cf. PARDE, 1961) a vu apparaître des études plus quantitatives, et un réseau de mesures plus conséquent, après les crues catastrophiques de Septembre-Octobre 1958. Mais ces phénomènes, sans être toujours catastrophiques sont relativement fréquents : J.M. CAROFF (1965) signale 171 cas en 10 ans où l'on a dépassé 40 mm en 24 H à l'un des postes de la région. De notre côté, sur un découpage un peu différent et sur 20 ans,

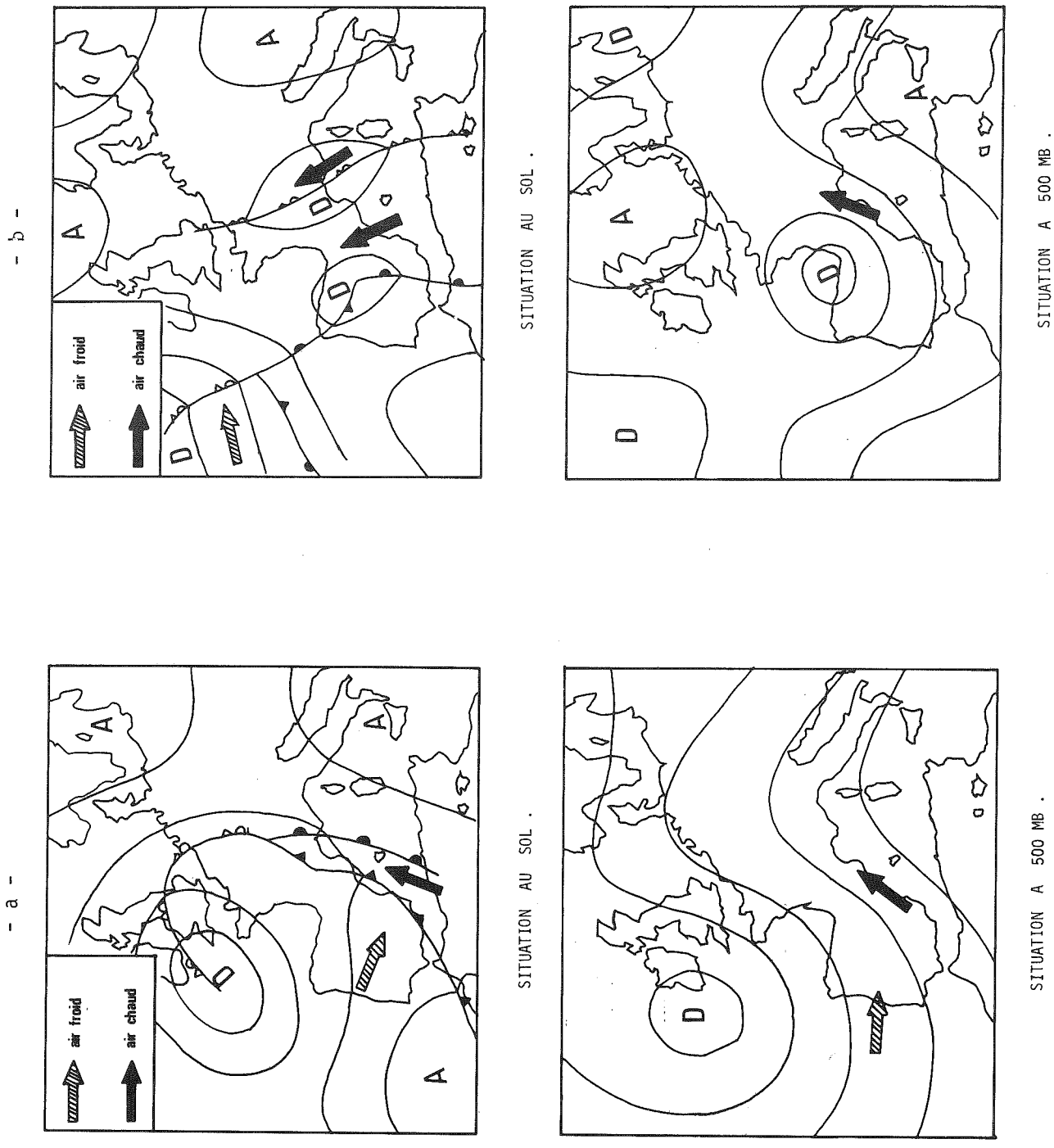


Figure I-3 - Situations metéorologiques types :

a) Perturbation Cevenole
b) Retour d'air mediterranéen

nous avons rassemblé une bonne centaine d'épisodes (pour la seule période Septembre-Octobre-Novembre) ayant donné plus de 50 mm en 2 ou 3 jours sur un de ces postes.

L'étude synthétique de J.M. CARROFF (1965) sur la météorologie des pluies de ces régions met en évidence 4 types de situations génératrices dont 2 sont largement dominantes, et également probables à l'échelle annuelle.

(I) - Les perturbations cévenoles proprement dites, caractérisées par :

- Une dépression barométrique au voisinage de l'Irlande à laquelle est associée une perturbation comportant 2 secteurs : le secteur chaud alimenté en air tropical, maritime ou méditerranéen instable envahit la vallée du Rhône et se trouve en général bloqué à l'Est par des hautes pressions allant de la Tunisie à la Scandinavie. Le secteur froid, d'origine atlantique et orienté à l'ouest, envahit progressivement le Sud-Ouest et le Massif Central et vient soulever brutalement la masse d'air chaud et humide "confinée" dans la vallée du Rhône (mais toujours alimentée) (Fig.I-3-a)

(II) - Retours d'air mediterranéens

La situation en altitude est assez similaire mais décalée sensiblement vers le Sud avec la dépression dans le Golfe de Gascogne et des hautes pressions de la Tunisie à la Mer du Nord. La dépression provoque un appel d'air chaud méditerranéen qui se trouve bloqué par la dorsale anticyclonique et soulevé assez rapidement, d'où une occlusion sur place des perturbations. (Fig.I-3-b)

Les 2 autres types signalés par CAROFF : régime de Nord-Ouest en altitude, et convergence horizontale de masses d'air instables, donnent des précipitations nettement moins abondantes.

Si on se limite à la période d'automne les fréquences respectives des deux types décrits passent à 2/3 1/3, avec des intensités et des volumes plus forts dans le type I que dans le type II. On remarquera que pour ce dernier, c'est le secteur chaud lui-même qui est générateur (même si ce n'est pas un front chaud au sens strict). Or les fronts chauds ont une pente assez faible ($\sim$ 1/300). Dans le premier type c'est un front froid (beaucoup plus raide: $\sim$1/60) qui soulève la masse d'air chaud et son passage donne lieu à une chute brutale des températures (cf tableau 1 dans l'article de J. JACQUET, 1959) et à de fortes intensités de pluies (atteignant parfois 100 mm en 1 H). Celles-ci s'expliquent non seulement par des facteurs météorologiques (forte instabilité de l'air chaud) mais aussi locaux.

La masse d'air froid, qui est montée progressivement sur le Massif Central, débouche sur une chute abrupte (1000 m en 15 à 20 km, soit $\sim$ 1/20). Il est possible que des effets de courants de densité accélèrent localement la vitesse de ce front, tandis que les effets de l'accélération de Coriolis (apparition de recirculations et d'une composante de Sud-Est) le devient latéralement.

Cela se traduit sur les pluviogrammes (cf P. GUILLOT, 1959) par une pluie relativement faible, d'origine orographique, tant que seule la masse d'air chaud escalade, plus ou moins franchement, le relief. Puis le passage du front froid donne, en quelques

heures (3 à 6) d'averse intense le maximum de l'épisode, suivi ensuite par des averses de traîne.

Les deux situations type décrites ci-dessus sont cependant un peu trop tranchées, et il arrive que l'on ait des situations intermédiaires, où un front froid peu actif et ondulant reste bloqué par le retour d'air méditerranéen pendant 1 ou 2 jours, donnant lieu à plusieurs averses intenses successives.

Dans notre cas, on constatera que pour un phénomène qui trouve son origine dans des mouvements à grande échelle, la localisation transversale est très précise et se produit toujours sur la rupture de pente des Cévennes, à quelques km près, car le phenomène de confinement de l'air chaud ne peut apparaitre que là.

Par contre, la localisation longitudinale est plus imprécise, les conditions topographiques se retrouvant sur plus de 100 km du Sud-Sud-Ouest au Nord-Nord-Est. Or, comme le dit J.JACQUET (1959) "un déplacement de 10 à 20 km du centre de gravité des averses exerce une influence déterminante sur la physionomie des crues et constitue un des obstacles majeurs à la prévision ou à l'avance des crues du Vidourle". C'est cette distribution spatiale d'épisodes qu'il nous faudra estimer à l'aide d'un réseau fixe, et localement clairsemé.

Dans notre cas, nous nous limiterons à l'analyse "statique" des épisodes, qui peut s'appliquer aussi à des précipitations mensuelles ou annuelles. Par contre, les courts pas de temps (3 H ou 1 H au moins) nécessitent la prise en compte des aspects "dynamiques" (déplacement des averses, etc...) qui font l'objet d'autres travaux en cours.

b) Traitement spatial des précipitations

On peut distinguer trois niveaux, qui nous amènent des champs spatiaux en général à celui des précipitations en particulier :

1 - Le traitement spatial s'est intéressé aux propriétés statistiques des champs spatiaux, considérés comme réalisations de fonctions aléatoires. On trouve une école anglaise, associée au nom de P. WHITTLE (1954) mais qui a connu un développement continu avec M.S. BARTLETT et aujourd'hui J. BESAG etc... Elle s'intéresse beaucoup aux aspects spatiaux en écologie, géographie humaine, agronomie, etc... Parallèlement, une école russe, conduite par YAGLOM, a développé les aspects théoriques sur les processus spatiaux et les applications dans le domaine de la turbulence. Enfin de façon complètement indépendante, G. MATTHERON (1965) a ouvert une voie nouvelle : la géostatique, avec des applications surtout dans le domaine minier.

2 - Le traitement spatial des champs météorologiques.

Les précurseurs sont indéniablement les chercheurs de Leningrad (DROSDOV, 1946, etc...) dont les travaux, synthétisés en 1963 par GANDIN, servent de référence pratique à l'O.M.M. Postérieurement des travaux de H.J. THIEBAUX (1973) au N.C.A.R. de Boulder (Colorado), ou à l'E.E.R.M. de la Météorologie Nationale ont repris et complété cette approche.

3 - Le traitement spatial dans le cas particulier des précipitations.

Nous passerons sous silence les nombreuses méthodes empiriques ou analytiques consistant à ajuster diverses surfaces plus ou moins sophistiquées à des données de pluies, mais sans hypothèses statistiques.

Une publication de P. HUTCHINSON (1972) fait entrer effectivement les méthodes d'analyse spatiale dans l'étude des champs de précipitations et J.P. DELHOMME (1973) applique la technique du krigeage à des champs de précipitations du Tchad. A côté de ces tentatives isolées, un groupe du M.I.T., dirigé par I. RODRIGUEZ-ITURBE (1974) a fait un important travail théorique et pratique dans l'analyse des pluies, mais en conservant malheureusement des hypothèses de stationnarité peu adaptées à nos région. Citons pour terminer une synthèse assez brève de R.T. CLARKE (1977).

On peut en conclure que la prise en compte des aspects spatiaux, après avoir atteint les grands champs météorologiques synoptiques, gagne peu à peu les champs plus "régionaux" des variables hydrométéorologiques.

I.2.3. Les épisodes cévenols. Données utilisées. Buts poursuivis

Pour des raisons matérielles, nous avons appelé "épisodes cévenols" les précipitations réparties sur 2 ou 3 jours et ayant produit, dans l'un au moins des bassins de la bordure Sud-Est du Massif Central (Gardon, Cèze, Ardèche, Doux ou Eyrieux) une lame d'eau supérieure ou égale à 50 mm.

On peut certes envisager d'autres choix, comme celui de prendre les pluies journalières maximales dans chaque épisode, mais le découpage de ces 24 H calendaires nous semble trop arbitraire et trop éloigné de la réalité. Il fallait donc prendre l'épisode globalement, ou descendre au niveau de phénomènes plus fins : les averses intenses de 2 à 6 H.

Ces épisodes sont caractérisés par la lame d'eau tombée aux 73 stations d'un réseau de base (cf. Fig.I-2) (ayant fonctionné sans arrêt sur la période 1956-1977). Un certain nombre d'autres stations, connues pour des périodes plus ou moins longues, serviront de réseau-test pour mesurer la qualité des reconstitutions.

On donne ci-joint un certain nombre de caractéristiques des données utilisées. (cf. Table III - date des episodes, durée, moyenne, valeur et localisation des mini & maxi)

Nos études auront 2 objectifs :

- Analyser le réseau considéré en vue de son optimisation. Nous verrons que le terme habituel d'optimisation des réseau de mesures peut recouvrir plusieurs notions distinctes.

- Analyser les épisodes disponibles et chercher à mettre en évidence leur structure spatiale. Cela nous amènera à rapprocher la méthode d'analyse en composantes principales et certains modèles de processus spatiaux. Cependant, l'intérêt pratique est d'estimer de façon optimale des valeurs manquantes ou non mesurées, ou des lames d'eau moyenne sur certains bassins.

N° épisode	Date	Durée en jours	Lame d'eau moyenne (en 1/10 mm)	Minimum mesuré (en 1/10mm)	N° et nom de la station minimale	Maximum mesuré (en 1/10mm)	N° et nom de la station maximale	Position dans le réseau
1	01/09/1956	3	740	89	40 GANDRIEU	2546	26 MONTPEZAT	N
2	25/09/1956	3	927	424	19 PONT DE LIGNON	2015	64 PONT DE MONTVERT	S
3	03/10/1956	2	192	0	53 STANDRE DE VALBORGNE	600	6 PRIVAS	N
4	21/09/1957	2	513	0	7 STE EULALIE	2630	52 ST ETIENNE VALLEE	S
5	05/11/1957	3	530	30	67 MEYRUEIS	1718	28 LOUBARESSE	N
6	09/11/1957	3	556	202	69 CHANAC	900	44 PINOLS	N
7	30/09/1958	2	1713	315	44 PINOLS	4290	52 ST ETIENNE VALLEE	S
8	04/10/1958	2	688	32	16 LE PUY CHADRAC	2367	48 MALONS	S
9	14/09/1959	5	447	27	17 CHOMLIX	1375	51 ST MAURICE DE VENTAL	S
10	17/10/1959	6	613	97	41 MONISTROL D'ALLIER	1829	60 MONT AIGOUAL	S
11	27/11/1959	6	1141	201	17 CHOMLIX	2806	25 MAYRES	N
12	17/11/1959	3	402	0	17 CHOMLIX	1291	60 MONT AIGOUAL	S
13	15/09/1960	5	1161	301	14 SANSSAC L'EGLISE	3176	6 PRIVAS	N
14	03/10/1960	4	1283	230	42 SAUGUES	3363	28 LOUBARESSE	N
15	20/10/1960	7	2179	369	44 PINOLS	5786	25 MAYRES	N
16	28/10/1960	3	534	151	21 TARENTAISE	1389	28 LOUBARESSE	N
17	05/11/1960	2	304	41	63 VALLERAUGUE	865	25 MAYRES	N
18	21/11/1960	3	732	170	45 LES UFFERNETS	1770	63 VALLERAUGUE	S
19	30/09/1961	4	593	44	6 PRIVAS	2389	60 MONT AIGOUAL	S
20	05/10/1961	3	997	346	42 SAUGUES	2454	48 MALONS	S
21	21/10/1961	8	840	82	70 MONTMIRAT	4231	28 LOUBARESSE	N
22	12/11/1961	2	726	163	69 CHANAC	1649	28 LOUBARESSE	N
23	26/11/1961	2	801	11	12 CAYRES	2392	60 MONT AIGOUAL	S
24	26/09/1962	4	486	208	17 CHOMLIX	1313	60 MONT AIGOUAL	S
25	12/10/1962	3	464	31	21 TARENTAISE	2314	60 MONT AIGOUAL	S
26	06/11/1962	6	1653	319	20 MOULAS	4090	60 MONT AIGOUAL	S
27	19/09/1963	3	596	98	1 LAMASTRE	1644	28 LOUBARESSE	N
28	31/10/1963	2	1525	259	24 VERSILHAC	6818	60 MONT AIGOUAL	S
29	03/11/1963	2	366	0	45 LES UFFERNETS	2271	27 ANTRAIGUES	N
30	05/11/1963	3	1128	324	44 PINOLS	2584	25 MAYRES	N
31	09/11/1963	5	314	25	16 LE PUY CHADRAC	1311	30 VILLEFORT	S
32	25/11/1963	3	271	15	12 CAYRES	981	26 MONTPEZAT	N
33	04/09/1964	3	539	237	15 FIX ST GENEY	1425	56 ANDUZE	S
34	01/10/1964	3	1005	318	44 PINOLS	2341	55 LASALLE	S
35	02/09/1965	5	654	132	62 LE VIGAN	1304	28 LOUBARESSE	N
36	09/09/1965	2	284	0	69 CHANAC	1051	57 CHAMPCLAUSON	S
37	24/09/1965	3	1579	193	17 CHOMLIX	5333	60 MONT AIGOUAL	S
38	30/09/1965	2	560	86	41 MONISTROL D'ALLIER	2347	25 MAYRES	N
39	16/10/1965	2	1041	147	5 VERNOUX	2948	60 MONT AIGOUAL	S
40	26/10/1965	4	593	17	13 LE MONASTIER	2685	61 PUECHAGUT	S
41	17/11/1965	2	403	75	16 LE PUY CHADRAC	1071	26 MONTPEZAT	N
42	29/09/1966	2	442	0	18 RETOURNAGUET	1647	60 MONT AIGOUAL	S
43	11/10/1966	3	233	16	43 PAULHAC EN MARGERIDE	703	2 ST AGREVE	N
44	14/10/1966	3	584	56	44 PINOLS	1582	30 VILLEFORT	S
45	19/10/1966	2	218	14	62 LE VIGAN	1300	7 STE EULALIE	N
46	24/10/1966	4	626	47	68 CASSAGNAS	1722	4 ST PIERREVILLE	N
47	02/11/1966	2	319	72	36 BELVEZET	720	32 AUBENAS	N
48	05/11/1966	2	321	0	15 FIX ST GENEY	1027	7 STE EULALIE	N
49	09/11/1966	2	704	234	42 SAUGUES	2110	60 MONT AIGOUAL	S
50	21/09/1967	2	262	58	54 ALES	799	27 ANTRAIGUES	N
51	05/11/1967	1	146	5	56 ANDUZE	580	29 VALGORGE	N
52	15/11/1967	3	1075	258	44 PINOLS	2943	27 ANTRAIGUES	N

N° épisode	Date	Durée en jours	Lame d'eau moyenne (en 1/10 mm)	Minimum mesuré (en 1/10mm)	N° et nom de la station minimale	Maximum mesuré (en 1/10mm)	N° et nom de la station maximale	Position dans le réseau
53	27/11/1967	2	346	115	71 MENDE	787	31 VALS LES BAINS	N
54	29/08/1968	3	658	157	69 CHANAC	2056	62 LE VIGAN	S
55	03/09/1968	3	489	115	44 PINOLS	1292	33 JOYEUSE GENDARMERIE	N
56	14/09/1968	3	286	47	44 PINOLS	1282	27 ANTRAIGUES	N
57	11/09/1968	1	202	15	67 MEYRUEIS	460	14 SANSSAC L'EGLISE	N
58	08/10/1968	2	316	32	10 FAY SUR LIGNON	2287	5 VERNOUX	N
59	25/10/1968	1	394	60	67 MEYRUEIS	1189	30 VILLEFORT	S
60	01/11/1968	4	1327	176	54 ALES	3877	9 LA PALISSE	N
61	14/09/1969	3	649	162	21 TARENTAISE	1432	60 MONT AIGOUAL	S
62	19/10/1969	4	825	4	19 PONT DE LIGNON	3777	60 MONT AIGOUAL	S
63	22/11/1969	3	921	0	54 ALES	2348	30 VILLEFORT	S
64	08/10/1970	4	2268	476	69 CHANAC	4954	26 MONTPEZAT	N
65	13/11/1970	2	560	0	63 VALLERAUGUE	1270	49 GENOLHAC	S
66	29/11/1970	2	245	0	53 STANDRE DE VALBORGNE	942	28 LOUBARESSE	N
67	20/09/1971	1	238	0	61 PUECHAGUT	682	4 ST PIERREVILLE	N
68	11/10/1972	4	1214	298	71 MENDE	2417	28 LOUBARESSE	N
69	26/10/1972	3	758	223	14 SANSSAC L'EGLISE	2463	30 VILLEFORT	S
70	21/09/1973	2	382	50	73 LE BLEYMARD	2720	3 LE CHEYLARD	N
71	02/10/1973	3	1152	390	21 TARENTAISE	1830	68 CASSAGNAS	S
72	03/11/1973	5	942	40	69 CHANAC	3520	30 VILLEFORT	S
73	17/09/1974	5	1055	230	20 MOULAS	3718	55 LASALLE	S
74	14/11/1974	6	902	200	42 SAUGUES	2920	30 VILLEFORT	S
75	10/09/1975	10	757	110	67 MEYRUEIS	2400	5 VERNOUX	N
76	01/10/1975	1	416	80	61 PUECHAGUT	1260	51 ST MAURICE DE VENTAL	S
77	29/08/1976	3	1480	181	69 CHANAC	3920	30 VILLEFORT	S
78	12/09/1976	2	1107	120	71 MENDE	3638	59 BESSEGES	S
79	25/09/1976	5	579	0	67 MEYRUEIS	1860	29 VALGORGE	N
80	01/10/1976	6	427	70	40 GANDRIEU	1581	32 AUBENAS	N
81	11/10/1976	5	786	183	44 PINOLS	1788	60 MONT AIGOUAL	S
82	24/10/1976	4	1698	160	21 TARENTAISE	3995	60 MONT AIGOUAL	S
83	28/10/1976	3	745	380	71 MENDE	1283	55 LASALLE	S
84	09/11/1976	5	1574	300	42 SAUGUES	4598	25 MAYRES	N

TABLE III - Liste des épisodes utilisés et principales caractéristiques .

Variable / Echelle	Continue	Discrète	Binaire
Absolue	Température en ° K Rayonnement en cal. Pression en mb Précipitation en mm	Nombre de couches dans un manteau de neige Nombre de jours dans une séquence de précipitation Remarque : pas de séquence de 0 jours	Coûts d'erreur dans une prévision d'occurence : $C_{A/B} = 1$ $C_{B/A} = 10$
Relative	Température en °C	Variable classée en classes d'égale amplitude	Nombre d'avalanches dans un couloir : 0 ou 1
Ordinale	Indice de stabilité d'un manteau de neige. Violence d'un orage à grêle	Rang d'une observation sur l'ensemble des obs. d'une variable. Variable classée en classe équiprobable Température <0, $=0$, >0	Valeurs fortes - faibles Température > 0 , $\leqslant 0$ Valeurs $>$ ou $\leqslant$ à la moyenne
Nominale	Direction du vent en ° d'angle	Département auquel appartient une station Type de temps	Occurence - non occurence Avec précipitation - sans précipitation

- TABLE IV - Classifications des variables .

CHAPITRE II

APERCU THEORIQUE SUR LES PROBLEMES DE CODAGE DES DONNEES. NOTIONS DE DISTANCES ENTRE INDIVIDUS

II.1 - Introduction

Les données, brutes ou élaborées, présentées dans le 1er chapitre, décrivent les phénomènes que l'on veut analyser et prévoir. Or les méthodes d'analyse multidimensionnelles que nous utiliserons supposent vérifiées certaines hypothèses. Par exemple, l'ensemble des variables est souvent présumé "homogène" et parfois multinormal. Les individus ou observations sont comparables et on peut mesurer les proximités entre individus ou entre variables.

Dans la réalité, les échantillons dont on dispose rassemblent :

- soit des "mélanges" de variables de natures diverses qu'il faut d'abord rendre comparables (cas de Davos)
- soit des variables homogènes et de même nature mais pas directement adaptées au problème. (Par exemple, ne permettant pas de comparer vraiment les individus). Dans ce cas des transformations sont nécessaires pour se rapprocher des propriétés souhaitées (cas des pluies cévenoles).

II.2 - Différents types de variables utilisées dans les analyses et leurs caractéristiques

II.2.1. Classification des types de variables

Nous empruntons à M.R. ANDERBERG (1973) les 2 classifications suivantes :

(a) La première s'appuie sur les propriétés analytiques de l'ensemble sur lequel les variables prennent leurs valeurs, d'où des variables :

- continues : sur R (sur $]-\infty, +\infty[$, $[0, +\infty[$, ou sur un intervalle borné $[a,b]$) avec éventuellement un ou plusieurs point d'accumulation. En général, pour un point quelconque de l'intervalle : $Pr\ (x < X \leqslant x + dx) \to 0$, et on appelle point d'accumulation a un point tel que : $Pr\ (X = a) = p_a$ valeur finie.

Exemples : . La température en °C est quasi-continue sur $]-\infty, +\infty[$ (en fait $]-273, +\infty[$, tandis que le rayonnement est continu sur $[0, Rmax]$

. La précipitation journalière a le point d'accumulation 0 sur $[0, +\infty[$

. et la nébulosité a les points d'accumulation 0 et 10 sur l'intervalle 0, 10 des dixièmes de couverture nuageuse.

- discrètes : sur un ensemble fini, ou infini mais dénombrable (pouvant être mis en correspondance avec l'ensemble des entiers naturels)
- dichotoniques ou binaires (cas particuliers du précédent).

(b) La seconde considère les propriétés algébriques de l'échelle de mesure.

- Echelle absolue : Si 2 éléments ont pour mesure x_a et x_b, on peut définir d'une part un ordre : $x_a > x_b$, d'autre part un écart (ou distance) $x_a - x_b$, et même un rapport x_a/x_b permettant de dire que A est x_a/x_b fois plus grand que B. Le 0 de l'échelle a donc un sens.

- Echelle relative ou mesure d'intervalle : Elle permet de définir un ordre $x_a > x_b$ et un écart $x_a - x_b$, éventuellement de comparer les écarts $x_a - x_b$ et $x_c - x_d$, mais le 0 de l'échelle est arbitrairement défini et n'a pas de sens absolu.

- Echelle ordinale: Définit une relation d'ordre $x_a = x_b$ ou $x_a > x_b$ ou $x_a < x_b$ mais l'écart $x_a - x_b$ n'a pas de sens.

Exemple : Journée à grêle avec

dégâts faibles	$x = 0$
dégâts moyens	$x = 1$
débâts forts	$x = 2$

- Echelle nominale : Entre 2 éléments A et B, il n'y a plus de relation d'ordre mais seulement : $x_a = x_b$ ou $x_a \neq x_b$

Exemple : Journée

sans précipitation	= 0
avec précipitation pluvieuse	= 1
avec précipitation neigeuse	= 2

Le tableau IV montre les relations entre ces 2 classifications et en donne d'autres exemples empruntés à nos données.

II.2.2. Changements d'échelle et codage des variables

Dans la mesure où les méthodes d'analyse supposent un type d'échelle homogène pour toutes les variables, nous sommes conduits à effectuer des transformations.

Une première technique d'homogénéisation, quand certaines variables du "paquet" sont ordinales, comportent des seuils ou des points d'accumulation, consiste à les ramener toutes en échelle ordinale par le biais d'un découpage en classes.

Or on peut considérer que les échelles absolue ⟶ relative ⟶ ordinale ⟶ nominale sont classées par information décroissante, et que tout passage de gauche à droite entraîne une perte, le passage inverse nécessitant un apport d'information extérieure.

Dans le cas fréquent qui nous intéresse : passer d'une variable continue, en

échelle absolue ou relative, à une variable discrète ordinale notre but sera de construire des classes en minimisant la perte d'information.

a - Classes d'égale amplitude

On divise l'écart entre la valeur maximum et la valeur minimum en un nombre fixé de segments. (Le problème est que la forme de l'histogramme de fréquences obtenu est très liée au nombre de classes choisies, surtout en cas de distribution plurimodale).

D'autre part, à nombre de classes donné, l'adjonction d'une valeur, si elle est extrémale, modifie complètement la forme de l'histogramme et peut conduire à des effectifs par classes très déséquilibrés.

b - Classes d'égal effectif

Elles évitent les inconvénients ci-dessus. De plus, elles peuvent être rendues encore plus explicites si on leur adjoint des hypothèses de distribution. En effet, on peut considérer que l'on cherche des classes à probabilité constante ou équiprobables. On peut alors prendre l'ensemble des points, lui ajuster un modèle probabiliste dont on calculera les paramètres, puis déduire de ce modèle les intervalles de classes équiprobables non plus au sens de la distribution empirique, mais au sens du modèle probabiliste choisi.

c - Classes d'égale dispersion

Il s'agit cette fois de classes telles que les distributions des valeurs au sein de chaque classe aient mêmes caractéristiques d'une classe à l'autre. Un premier problème est de définir un indice de dispersion qui sera en général :

soit $D_1^k = \left| x_{max\,/k} - x_{min\,/k} \right|$

soit $D_2^k = \frac{1}{n_k} \sum_{n_k} \left(x_{jk} - \bar{x}_k \right)^2$ $\longrightarrow$ variance de la classe k.

Le premier indice ou "distance" D_1 conduit à appliquer des méthodes de classification hiérarchique sur les données probablement classées:

Exemple d'algorithme (*)

- Classer les données par ordre croissant ou décroissant
- Examiner, tous les points 2 à 2 et agréger les 2 données qui définissent le couple le plus concentré.
- Aérer, en considérant que la distance de 2 groupes est celle des 2 individus (appartenant respectivement au 1er et au 2ème groupe) les plus proches.

Ceci revient à construire une ultramétrique qui se traduit par un arbre de hiérarchie. On choisit alors le nombre de classes puis on coupe l'arbre au niveau choisi. Evidemment, cela ne nous donne ni des classes d'égal effectif, ni d'égale amplitude. Par contre, les séparations entre classes correspondent aux "vides" de la distribution empirique.

La distance D_2 conduit à des algorithmes plus compliqués, analogues à ceux décrits dans la V^ème Partie Chap. II . On peut s'en faire une idée en considérant l'algorithme de Ward (in ANDERBERG ,1973) qui procède de manière partiellement hiérarchique .

Exemple d'algorithme : (*)

. Les N éléments constituent N groupes de variance nulle.

. On regroupe les 2 éléments donnant une variance minimale $\Longrightarrow$ i et j :

$$\sigma_{(i+j)} = \min_{l,k} \sigma_{(l+k)}$$

. On continue jusqu'à n'avoir qu'un seul groupe. Toutefois, le passage d'une étape à la suivante est très rigidement dépendant des étapes antérieures (effet de chaîne), au point que l'on peut parfois en modifiant astucieusement les groupes obtenus, améliorer sensiblement le résultat.

Une façon de l'améliorer, qui ne sera plus hiérarchique mais heuristique, consistera, en partant de cette partition en g classes, à considérer les g-1 frontières interclasses et à tenter, pour chacune d'elle :

. de déplacer à gauche la première observation de droite,

. effectuer la modification si celle-ci fait décroître la somme des variances intraclasses ou plutôt : $W = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k)^2$

. continuer les tentatives sur la frontière considérée tant que W décroît, sinon, passer à une autre frontière,

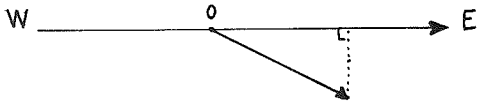
. s'arrêter quand W cesse de décroître.

Malheureusement, on aboutit aussi à un minimum local, faute d'effectuer un balayage complet de tous les groupements possibles, soit pour N observations à classer en g groupes : C_{N-1}^{g-1}

Les exemples précédents ne sont pas exhaustifs mais mettent en évidence la multiplicité des choix possibles quand il s'agit simplement de classer une variable continue.

Un cas semblable consistera à transformer des variables continues ou ordinales en variables nominales (où la notion d'ordre sur l'espace de départ ne se reflète plus de façon univoque sur l'espace d'arrivée).

Exemple : Vitesses du vent sur l'axe orienté Ouest-Est



Classes formées :

$[-3 , +3$ m/s$]$ vent faible

$[-15 , -3[$ et $]3,15]$ vent moyen < -15 et > 15 vent fort

On retrouvera ces problèmes dans la partie décisionnelle, quand on voudra, au vu d'une variable continue ou discrète en échelle absolue, effectuer des groupements :

Exemple :

Variable discrète	variable ordinale	Variable continue		variable discrète nominale	
pas d'avalanche observée	Classe 1	Précipitation	P 0	neige	C_2
1 ou 2	Classe 2		P = 0	sans	C_o
$\geqslant$ 3	Classe 3		P 0	pluie	C_1

II.2.3. Exemples simulés (*)

Pour illustrer ces diverses possibilités, nous donnons le détail des divers algorithmes évoqués sur un exemple simulé. Pour cela, on a généré 50 nombres aléatoires, pris dans une loi de Gauss $\mathcal{N}(150., 50.)$ que l'on se propose de regrouper en 10 classes.

(Dans les exemples concrets, où le nombre d'individus excède une centaine, il est malheureusement exclu de présenter les résultats intermédiaires).

① Classes d'égale amplitude

x_{min} = 56.17
x_{max} = 246.41
$\Rightarrow$ amplitude de classe = 19.02

Effectifs des groupes :

	1	2	3	4	5	6	7	8	9	10
n_k	2	2	8	4	4	13	7	3	4	3

② Classes d'égal effectif

Caracteristiques des groupes

Classe	1	2	3	4	5	6	7	8	9	10
Effectif	10	10	10	10	10	10	10	10	10	10
Amplitude	44.33	4.10	17.48	20.12	1.95	4.10	4.65	11.33	23.44	24.22

③ Classes d'égale hiérarchie ("single-linkage)

Effectifs des groupes

	1	2	3	4	5	6	7	8	9	10
n_k	1	1	2	8	4	17	10	4	2	1
Amplitude	0	0	7.42	11.72	7.81	19.92	28.71	16.99	.39	0
Ecart entre classes	9.179	12.89	14.84	8.39	13.476	8.98	10.54	11.33	7.62	

④ Classes d'égale hiérarchie (Méthode de Ward)

Effectifs des groupes

	1	2	3	4	5	6	7	8	9	10
n_k	2	2	8	4	4	13	7	3	4	3
Somme des carrés = variance du groupe x effectif	42.	27.6	120.1	36.8	28.2	129.6	139.	32.3	170.5	40.8

Total 7.67

Note : par un pur hasard, on retrouve au niveau 10 la même classification qu'en ① mais ce ne serait plus vrai aux niveaux 9 ou 11

⑤ Méthode d'amélioration itérative

A partir de la partition de Ward :

Effectifs des groupes

	1	2	3	4	5	6	7	8	9	10	
	2	2	8	4	4	13	6	4	4	3	Total
Somme des carrés	42.1	27.6	120.1	36.8	28.3	129.6	50.6	107.9	170.5	40.9	754.

A partir de classes d'égal effectif

Effectifs des groupes

	1	2	3	4	5	6	7	8	9	10	
	3	8	5	4	7	6	6	4	4	3	Total
Somme des carrés	245.8	303.1	149.5	28.3	10.4	12.7	50.6	107.9	170.6	40.9	1120

II.2.4. Caractéristiques statistiques des variables et transformations sur les variables

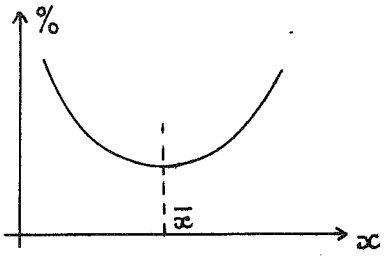
- Les variables sont caractérisées par leurs distributions statistiques, que l'on résume souvent à l'aide de quelques paramètres.

Moyenne : moment d'ordre 1 (caractéristique de position)

Variance: moment d'ordre 2 (caractéristique de dispersion)

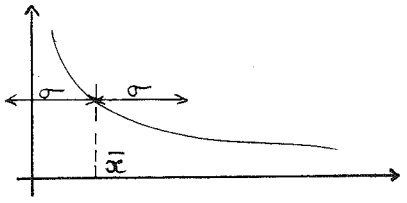
Toutefois, on peut se poser les problèmes suivants :

- la moyenne est liée à l'échelle choisie, de même que la variance, et elle est sensible à un changement d'origine.
- pour certaines distributions (distribution en U comme la nébulosité), la moyenne est la valeur la moins probable!



Dans ces cas particuliers , il vaut mieux considerer la distribution comme bimodale et donner les valeurs des 2 modes .

Dans le cas où les distributions sont dissymétriques, la moyenne n'est pas toujours un bon indice de position, surtout s'il y a des valeurs extrêmes possibles d'un seul côté de la distribution (Dans ce cas on peut lui préférer la médiane).



De même la variance n'est plus une bonne caractéristique de dispersion:

Pour ces distributions, on calcule : l'assymétrie, à l'aide du moment centrée d'ordre 3 μ_3 et on définit les coefficients :

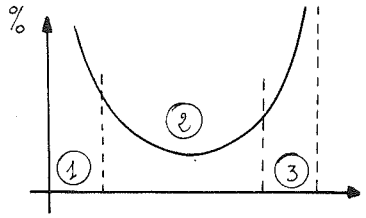
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \text{ou} \qquad \gamma = \sqrt{\beta_1} \qquad \text{avec} \quad \mu_2 = \sigma^2$$

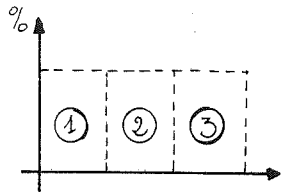
- Or, pour des raisons que l'on verra au paragraphe II-3, on cherche en général à utiliser des variables symétriques, donc à effectuer des transformations non linéaires pour modifier l'allule de la distribution.

On peut distinguer :

- les transformations fonctionnelles biunivoques (qui ne changent pas le type de la variable : une variable continue reste continue)
- les transformations non fonctionnelles, où l'on change d'échelle de mesure.

Exemple 1 : Une façon de recoder les variables consiste à les grouper en classes, puis à remplacer la valeur de la variable par le rang de sa classe d'appartenance. (Si les classes sont équiprobables, on obtient une distribution uniforme).





Le regroupement en classe et l'utilisation du rang de la classe plutôt que sa valeur moyenne entraîne cependant une perte d'information importante.

Une méthode plus fine consiste à remplacer la valeur de la variable par son rang dans l'échantillon. Cette "uniformisation" est très puissante pour des variables continues sans point d'accumulation, mais elle ne fonctionne pas si plusieurs valeurs sont identiques. De plus, l'adjonction d'une observation boulverse le classement précédent.

Exemple 2 : On dispose d'un échantillon d'une variable continue mais très dissymétrique (ex : les précipitations > 0).

On pense en général à préserver le caractère continu de la variable en effectuant une transformation fonctionnelle. On propose en général une transformation en racine carrée ou en logarithme. Pour la première, la valeur 0 ne pose pas de problème, par contre, pour la seconde, il faut envisager en plus un décalage d'origine: $\text{Log}(x - x_0)$

$\Longrightarrow$ Le cas particulier de la transformation en logarithme: (*)

Le problème est de choisir ce décalage x_0. On donne en annexe, une méthode simple mais assez grossière qui permet de redresser un échantillon donné. La méthode rigoureuse, mais assez lourde à mettre en oeuvre, est la suivante.

Si l'on suppose que la variable X suit une loi de Galton, c'est-à-dire :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}u^2} \, du \qquad \text{avec} \quad z = \alpha \, \text{Log}(x - x_0) + \beta$$

et α et β tels que z est $\mathcal{N}(0,1)$, il faut ajuster cette loi aux données observées, donc calculer les 3 paramètres α , β , et x_0.

On trouvera dans ROCHE M. (1963) l'ajustement par les moments et le maximum de vraisemblance. Dans notre cas, le paramètre clé est le décalage d'origine x_0 , car ensuite α et β interviennent de façon linéaire. La méthode des moment conduit à un calcul plus simple que le maximum de vraisemblance, car x_0 est solution de :

$$\frac{(\mu_x - x_0)^3}{\sigma_x^2 + 3(\mu_x - x_0)^2} = \frac{\sigma_x^4}{\mu_{3x}}$$

où μ_{3x} est le moment centré d'ordre 3 :

$$\mu_{3x} = E[x^3] - 3\mu_x \cdot \sigma_x^2 - \mu_x^3$$

Il suffit de résoudre l'équation du 3ème degré en $(\mu_x - x_0)$ soit par une méthode directe soit par la méthode de Bairstow.

On peut vérifier que $\mu_{3z} = 0$.

On se contente généralement de la transformation :

$$Y = \mathrm{Log}(X - x_0) \quad \text{ou} \quad \mathrm{Log}(X - \alpha)$$

où Y suit une distribution normale de moyenne μ_y et d'écart-type σ_y.

On démontre (cf annexe) que les relations sont :

$$\mu_x = a + e^{(\mu_y + \frac{1}{2}\sigma_y^2)}$$

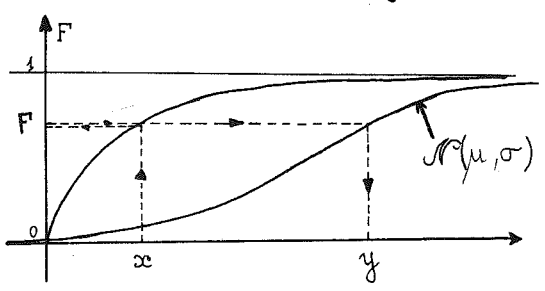
$$\sigma_x^2 = e^{2(\sigma_y^2 + \mu_y)} - e^{(\sigma_y^2 + 2\mu_y)}$$

On montre ainsi que le coefficient d'assymétrie :

$$\gamma_x = \frac{\mu_{3x}}{\sigma_x^3} = \frac{e^{3\sigma_y^2} - 3e^{\sigma_y^2} + 2}{(e^{\sigma_y^2} - 1)^{3/2}}$$

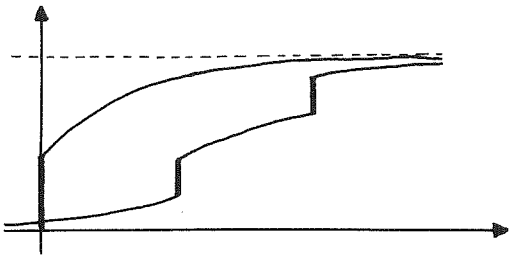
(AITCHINSON et BROWN , 1957, cité par MATALAS, 1967).

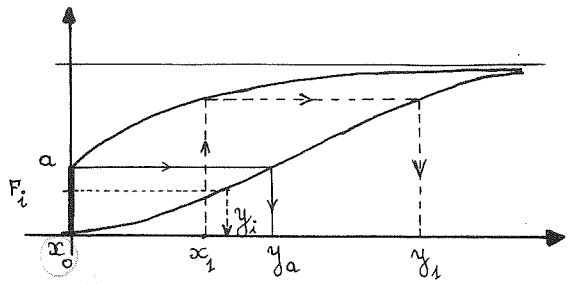
- Plutôt que de choisir un modèle pour la densité de probabilité, KUENY (1977) propose d'associer à la courbe de fréquences cumulées observée, une loi de Gauss (et de faire numériquement la relation entre la valeur x de probabilité F dans la loi considérée et la valeur y de même fréquence F dans la loi $\mathcal{N}(0,1)$).



Ceci évite le choix d'un modèle, son ajustement aux données x , et fournit directement un échantillon de y gaussien.

Ces transformations, de même que le remplacement de la variable par son rang, ne sont plus suffisantes quand la variable présente un ou des points d'accumulation.





Dans ce cas, GROSJEAN et KUENY* proposent d'associer à x_1 une valeur y_1 comme précédemment, mais à x_0, on associe un tirage aléatoire dans une loi uniforme sur $[0,a]$, soit F_i d'où y_i associée à x_0. Les valeurs $x = x_0$ génèrent donc un échantillon aléatoire continu, d'où une variable y gaussienne non tronquée, mais partiellement aléatoire. Il faut toutefois être très prudent si on veut utiliser cette variable y dans une corrélation !

II.3 - Distances entre 2 éléments (individus ou variables)

Nous nous limiterons aux plus usuelles que nous avons effectivement utilisées. Le but est de mesurer la proximité de 2 observations (individus) ou de 2 variables. Les 2 cas extrêmes que nous rencontrons sont :

- celui où les individus sont caractérisés par des variables continues ou discrètes associées à des échelles absolue, relative, voire ordinale,
- celui des variables binaires ou dichotoniques purement nominales.

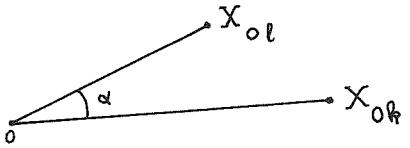
II.3.1. Distance associée à des variables d'intervalle et notion de corrélation

La distance "naturelle" entre 2 individus (observations) caractérisés par P variables est la distance euclidienne classique de $\mathbb{R}^P$:

$$d^2(i,j) = \sum_{l=1}^{P} (x_{il} - x_{jl})^2$$

Mais les intervalles de fluctuation respectifs des P variables peuvent être très différents et si les valeurs prises par une des variables peuvent être très grandes en valeur absolue, dans ce cas un terme dominant apparait dans la sommation : on mesure la distance surtout au sens de cette variable-là. D'où la nécessité de rendre les variables comparables en fluctuations pour éviter cet effet d'échelle.

De la même façon on peut considérer la variable X_l comme un point de l'espace $\mathbb{R}^N$ où N est le nombre d'observations. On peut alors parler de la distance entre 2 "points"-variables:



$$d^2(X_l, X_k) = \sum_{i=1}^{N} (x_{il} - x_{ik})^2$$

---

* Communication personnelle

Or, dans un espace euclidien, on sait que :

$$d^2(X_\ell, X_k) = \|X_\ell\| . \|X_k\| . \cos\alpha = X_{0\ell}^t . X_{0k}$$

la valeur de la distance peut être due soit à la norme des vecteurs (choix des échelles et de l'origine) soit à leur angle dans $\mathbb{R}^N$.

Une mesure intéressante de leur covariance sera la valeur de cet angle, qui nous libère des problèmes d'échelles.

Elle est invariante pour toute homothétie sur les variables. Par contre, elle est liée à l'origine choisie, et donc à une translation éventuelle.

On se libère de cette contrainte en utilisant des variables centrées :

$$X_k = \begin{bmatrix} x_{1k} - \bar{x}_k \\ x_{2k} - \bar{x}_k \\ \cdots\cdots \\ x_{Nk} - \bar{x}_k \end{bmatrix} \quad \text{auquel cas :} \quad \|X_k\| = \sum_i (x_{ik} - \bar{x}_k)^2 = N . \sigma_k^2$$

et $$X_\ell^t . X_k = \sum_{i=1}^N (x_{ik} - \bar{x}_k).(x_{i\ell} - \bar{x}_\ell) = N . cov(X_\ell, X_k)$$

d'où $$\cos\alpha = \frac{N . cov(X_\ell, X_k)}{N var(X_\ell) . N . var(X_k)} = \rho_{\ell k}$$

$\Longrightarrow$ Dans $\mathbb{R}^N$, on constate donc que si les variables sont centrées et réduites :

- les points variables sont sur une hypersphère de rayon 1
- la distance euclidienne entre 2 variables ℓ et k s'écrit :

$$d^2(X_\ell, X_k) = 2 - 2.\cos\alpha = 2.(1 - \rho_{\ell k})$$

où $\rho_{\ell k}$ est leur coefficient de corrélation habituel.

Cela est intéressant si l'origine des variables n'a qu'un sens relatif (auquel cas on prend la moyenne) et si on rend les échelles comparables en prenant comme unité, pour chaque variable, leurs variances respectives. On considère alors que le choix des échelles est arbitraire et qu'elles ont même variabilité intrinsèque.

Exemples : on pourra se poser les problèmes suivants :

- faut-il centrer les mesures de 2 stations de températures si la valeur 0°C joue un rôle dans le problème ?
- faut-il centrer les précipitations à 2 stations qui ont des moyennes très différentes, et quel phénomène physique élimine-t-on en faisant cela ?
- faut-il réduire (ou renormer à 10) une variable quasi-constante mais affectée d'un bruit de mesure ?
- 2 stations de précipitations dont les variances sont dans des rapports de 1 à 5 ont-elles même variabilité intrinsèque ?

Un autre problème est l'effet des transformations sur les variables.

Le coefficient de corrélation est insensible à toute transformation linéaire :

$$\rho(X_k, X_\ell) = \rho(a X_k + b, X_\ell)$$

Ceci n'est strictement vrai qu'au signe près car si a est négatif, le signe de ρ change. Nous verrons les problèmes que cela pose dans la IIème partie, § II-1.

Il n'en est pas de même si la transformation est non linéaire.

Exemple : transformation logarithmique. (*)

Si on considère les 2 variables X_1 et X_2 que l'on a normalisées par les transformations :

$$Y_1 = Log(X_1 - a_1) \qquad Y_2 = Log(X_2 - a_2)$$

Si on appelle μ_{Y_1}, μ_{Y_2}, σ_{Y_1}, σ_{Y_2} et ρ_Y les caractéristiques du couple Y_1 et Y_2 et de même pour le couple X_1, X_2, on montre que :

$$\rho_Y = \frac{1}{\sigma_{Y_1} \cdot \sigma_{Y_2}} \cdot Log\left[1 + \rho_X \sqrt{(e^{\sigma_{Y_1}^2} - 1)\cdot(e^{\sigma_{Y_2}^2} - 1)}\right]$$

Nous démontrons en annexe ce résultat, cité par MEIJA J.M. et al (1974).

On peut aussi n'effectuer la transformation que sur une seule des variables par exemple : $Y_1 = Log(X_1 - a_1)$ tandis que $Y_2 = X_2$

dans ce cas :

$$\rho_Y = cor.(Y_1, X_2) = \rho_X \cdot \frac{(e^{\sigma_{Y_1}^2} - 1)^{\frac{1}{2}}}{\sigma_{Y_1}}$$

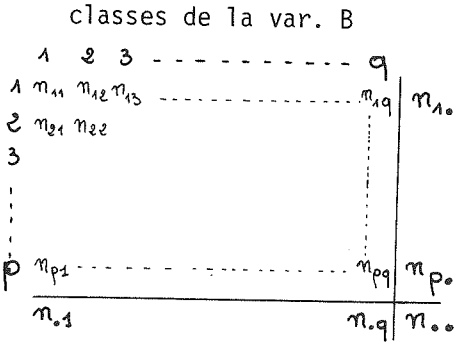
Remarque pratique

En général, on calcule d'abord μ_{x_j}, σ_{x_j} et $\rho(x_j, x_\ell)$ puis on décide éventuellement de faire une transformation $y_j = Log(x_j - a_j)$. Celle-ci suppose d'abord le calcul de a_j, qui peut se faire par différentes méthodes.

On pourrait en déduire μ_{y_j}, σ_{y_j}, puis $\rho(y_j, y_\ell)$ par les formules théoriques, mais celles-ci expriment des relations entre les paramètres vrais des populations, et non entre leurs estimations. Il y a alors un risque de trouver une matrice R qui ne soit plus semi-définie positive.

I.3.2. Distances associées à des variables nominales

ⓐ La mesure de distance entre des variables nominales a pour point de départ leur table de contingence :



où n_{ij} = nombre d'individus $\in$ [classe i de A $\cap$ classe j de B]

L'ordre des classes n'a théoriquement plus aucun rôle et on compare la fréquence observée $f_{ij} = \frac{n_{ij}}{N}$ à la fréquence que l'on pourrait attendre si les variables étaient indépendantes : $\frac{n_{i.} \times n_{.j}}{N}$ d'où :

$$\chi^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(n_{ij} - \frac{n_{i.} \times n_{.j}}{N}\right)^2}{\frac{n_{i.} \times n_{.j}}{N}} = N \left[\sum_i \sum_j \frac{n_{ij}^2}{n_{i.} \times n_{.j}} - 1 \right]$$

Comme cette mesure est proportionnelle à N on considère parfois la quantité : $\frac{\chi^2}{N}$.

ⓑ Un cas particulier intéressant est celui des variables binaires, que nous allons rencontrer par la suite. Si on appelle a, b, c et d les effectifs respectifs :

		var. B : 1	var. B : 0	
var. A	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	N

On trouve alors : $\chi^2 = N \frac{(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)}$

Or on peut aussi considérer la variable A comme un point de $\mathbb{R}^N$, qui aurait cette fois pour coordonnées (a+b):1 et (a+c):0. De même pour B.

La moyenne de A serait (a+b)/N

La moyenne de B serait (a+c)/N

et les produits croisés $\sum_i x_i y_i$ se résument à : a, et la corrélation classique s'exprime alors par :

$$\rho(A,B) = \frac{a - \frac{(a+b)(a+c)}{N}}{\sqrt{a+b-\left(\frac{a+b}{N}\right)^2} \cdot \sqrt{a+c-\left(\frac{a+c}{N}\right)^2}} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} = \frac{\chi_{AB}}{\sqrt{N}}$$

(compte tenu de $N = a + b + c + d$)

et les distances au sens du χ^2 ou au sens des corrélations (variables centrées réduites) se confondent.

Remarque : Les données n'ont même pas besoin d'être codées en 0 et 1 puisque le coefficient de corrélation est invariant dans toute transformation linéaire.

Ces notions seront reprises et développées dans le chapitre II de la 2ème Partie (Analyse des correspondances).

Pour être tout à fait complet, on peut citer encore comme mesures d'association convenant à des variables ordinales le coefficient de corrélation de rang de SPEARMAN et celui dû à KENDALL (τ , basé sur les différences de rangs entre les observations de 2 variables). Toutefois, nous n'aurons pas l'occasion de les utiliser par la suite.

CHAPITRE III

TRAITEMENT ET CODAGES APPLIQUES AUX PROBLEMES ETUDIES

Nous allons voir pourquoi les variables décrites au chapitre I nécessitent d'être recodées ou transformées à l'aide des techniques décrites dans le chapitre II.

III.1 - Premiers traitements sur les variables "explicatives" du phénomène avalanche

III.1.1. Constitution des échantillons

Compte tenu des importantes variations saisonnières, nous avions, dès le début de l'étude, décidé de désaisonnaliser en travaillant sur des périodes homogènes (cf thèse de Ph.BOIS, p.193). Malheureusement, si les périodes utilisées (mois par mois) réduisaient au maximum cette fluctuation, elles nous conduisaient à travailler sur des échantillons très faibles et posaient des problèmes de robustesse des modèles.

Nous sommes donc revenus à des périodes de 2 mois :

Novembre - Décembre
Janvier - Février
Mars - Avril
Mai - Juin

où les variables sont relativement stationnaires.

D'autre part, des problèmes subsistent quant à ce que l'on peut appeler le début et la fin de l'enneigement avec avalanches possibles, mais ces problèmes se posent dans le premier et le dernier bimestre, et nous ne considèrerons ici que les 2 bimestres Janvier-Février et Mars-Avril.

Nous disposons donc de :

14 bimestres Janvier-Février (59 ou 60 jours, sur les années 1961 à 1974)
soit 829 journées dont 154 avalancheuses.

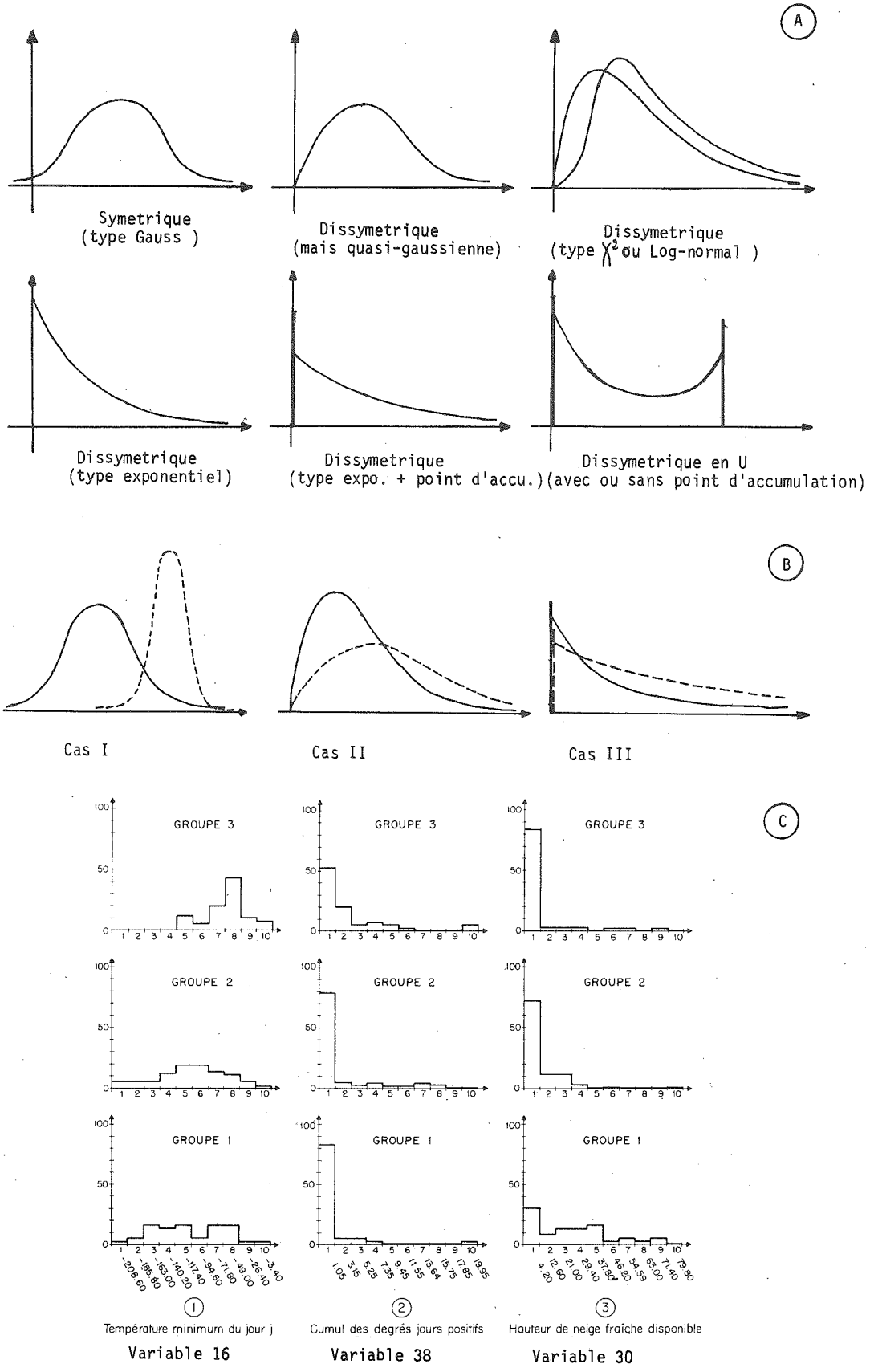
14 bimestres Mars-Avril (61 jours, sur les années 1961 à 1974)
soit 854 journées dont 153 avalancheuses.

De façon générale, les traitements descriptifs porteront sur l'ensemble des journées tandis que les modèles décisionnels seront systématiquement ajustés sur les années 1961-1972 et utilisés en test sur les périodes 1973 et 1974.

III.1.2. Traitement préliminaires

(a) Après désaisonnalisation, on peut s'inquiéter de la distribution statistique des variables. Celles-ci sont obtenues séparément pour les journées avalancheuses ou pour l'ensemble des journées du bimestre (avalanches incluses) mais leur forme varie peu selon l'échantillon choisi.

On trouve les formes suivantes (figure I-4a et tableau V):



Groupe : 1) avalanches de neige "sèche" 2) sans avalanche 3) avalanches de neige "humide"

Figure I-4 - Differentes courbes de densité de probabilité rencontrées :

a) formes possibles
b) differences possibles entre 2 groupes distincts
c) quelques exemples réèls

N°		
	1	dissymétrique type exponentiel
	2	dissymétrique type exponentiel
	3	dissymétrique type exponentiel
	4	dissymétrique type χ^2
	5	dissymétrique type exponentiel
	6	dissymétrique type χ^2 (mais quasi-gaussienne)
	7	dissymétrique mais avec des valeurs aberrantes
	8	dissymétrique type exponentiel
	9	symétrique (type gauss)
	10	symétrique
	11	dissymétrique (type χ^2 + point d'accumulation en 0)
	12	dissymétrique type χ^2 mais quasi-gaussienne
	13	symétrique (type gauss)
	14	dissymétrique type exponentiel
	15	dissymétrique type exponentiel
	16	dissymétrique type χ^2 mais quasi-gaussienne
	17	symétrique type gauss
	18	symétrique en $\cup$ plus point d'accumulation vers 0.
	19	symétrique type Gauss (ou χ^2)
	20	symétrique en $\cup$ plus point d'accumulation vers 0
	21	dissymétrique type χ^2 mais quasi-gaussienne
	22	symétrique type gauss
	23	dissymétrique type exponentiel
	24	dissymétrique type exponentiel
	25	dissymétrique type exponentiel
	26	dissymétrique type χ^2 (mais avec des valeurs erratiques)
	27	dissymétrique type exponentiel
	28	dissymétrique type exponentiel
	29	dissymétrique type exponentiel
	30	dissymétrique type exponentiel
	31	dissymétrique type exponentiel
	32	dissymétrique type exponentiel (plus 2 points d'accumulation 0 et 100)
	33	dissymétrique type exponentiel
	34	dissymétrique type exponentiel
	35	dissymétrique type exponentiel
	36	dissymétrique type χ^2 mais quasi-gaussienne
	37	dissymétrique type χ^2 mais quasi-gaussienne
	38	dissymétrique type exponentiel
	39	dissymétrique type exponentiel
	40	dissymétrique type exponentiel
	41	dissymétrique type exponentiel
	42	dissymétrique type exponentiel
	43	symétrique type gauss
	44	symétrique type gauss
	45	symétrique type gauss
	46	symétrique type gauss
	47	dissymétrique type χ^2 ou quasi-gaussienne
	48	symétrique type gauss
	49	symétrique type gauss
	50	symétrique type gauss

- TABLE V - Allure de la distribution des variables elaborées .

Parmi nos 50 variables, 21 sont donc assez symétriques et d'allure gaussienne, 27 sont d'allure exponentielle, dont une quinzaine présentent un point d'accumulation correspondant à un seuil, et les 2 restantes sont dissymétriques en U avec points d'accumulation aux extrémités .

(b) La séparation entre le groupe avalanche et l'ensemble des journées n'est pas toujours évidente, et elle est encore plus difficile quand les variables sont dissymétriques (Fig. I-4b).

D'autre part , la considération du seul groupe "avalanche" nous conduit à mélanger divers phénomènes, et un découpage en 2 sous-groupes (avalanches de neige sèche ou de neige humide) met déjà en évidence certaines réalités , malgré l'incertitude d'une telle classification (Fig.I-4c).

(c) Un problème peut aussi se poser dans les relations entre variables. Beaucoup des analyses ultérieures utiliseront la matrice des corrélations entre les variables. Si, pour les variables symétriques, ce coefficient a un sens, son interpretation se complique pour des variables dont l'une, ou les deux, sont dissymétriques.

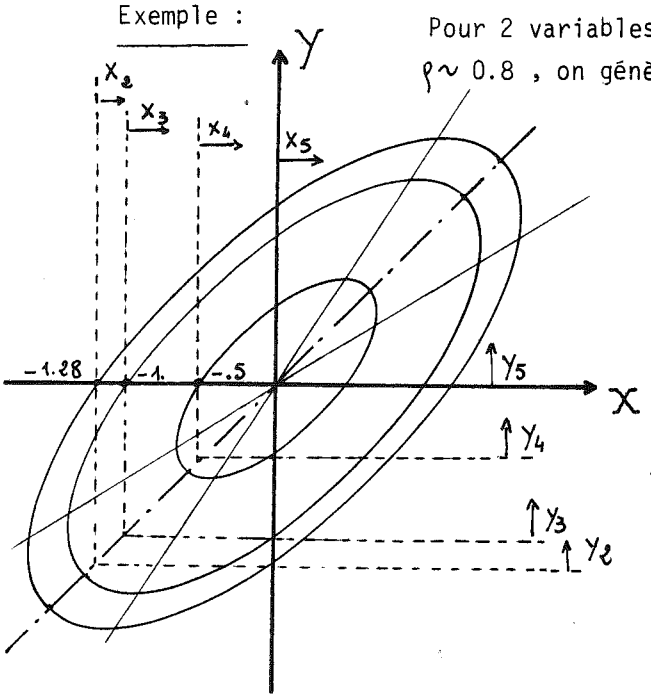
Exemple :

Precipitation du jour (dissymétrique) et Rayonnement solaire incident (symétrique): $-.250$

(d) Enfin,nombre de variables sont artificielles et comportent l'utilisation de seuils et de valeurs absolues qui modifient sensiblement la corrélation par rapport aux variables brutes de départ,et en général l'abaissent.

On peut en faire l'expérience sur un exemple artificiel:.

Exemple :



Pour 2 variables normales X_1, Y_1 initialement corrélées $\rho \sim 0.8$, on génère des variables X_k, Y_k en rendant égale au seuil S_k toute valeur $\leq S_k$:

$X_2, Y_2 \longrightarrow S_2 = -1.28$

$X_3, Y_3 \longrightarrow S_3 = -1.0$

$X_4, Y_4 \longrightarrow S_4 = -.5$

$X_5, Y_5 \longrightarrow S_5 = 0.$

D'où la matrice de correlation :

Y	X_1	X_2	X_3	X_4	X_5
1	0.786	0.777	0.763	0.725	0.655
2	0.782	0.782	0.774	0.746	0.681
3	0.770	0.775	0.770	0.751	0.691
4	0.727	0.741	0.744	0.743	0.702
5	0.653	0.674	0.684	0.709	0.683

Parmi nos variables "élaborées", c'est le cas des variables 12,14 et 15 (températures à 13h et selon qu'elles sont < 0 ou > 0). On trouve en Mars-Avril :

$$\rho_{12,14} = .353 \qquad \rho_{12,15} = .700 \qquad \rho_{14,15} = .450$$

En conclusion,il faudra regarder d'un oeil critique les analyses reposant sur les corrélations .

Dans certains cas de variables dissymétriques, on pourrait envisager soit une transformation du type Log ($x - a$) ou $\sqrt{x}$, soit une analyse de rang (la valeur de x_i est remplacée par son rang dans l'échantillon). Ce n'est plus possible lorsqu'en sus, les variables présentent un point d'accumulation : il vaut mieux alors mettre les variables en classes et utiliser les mesures vues en II.3.2.

III.2 - Premiers traitements et codages sur les données de pluies cévenoles

Les problèmes sont un peu différents car les variables sont évidemment beaucoup plus comparables entre elles. D'autre part, il ne s'agit pas non plus des séries pluviométriques journalières habituelles, d'où les considérations qui suivent.

III.2.1. Analyse de l'échantillon

Nous donnons d'abord quelques caractéristiques de l'échantillon d'épisodes utilisés, compte tenu de leur définition un peu particulière.

L'histogramme de leur durée (Fig.I-5a) montre que l'échantillon contient :

- 5 épisodes d'un jour, qui se retrouvent tous dans les classes correspondant à une faible lame d'eau et ont probablement un caractère orageux assez marqué;
- quelques épisodes particulièrement longs, que l'on pourrait redécouper en plusieurs phases intenses, l'épisode lui-même ayant été réellement continu ;
- mais les classes les plus nombreuses correspondent à la durée attendue de 2 à 3 jours, caractéristique des circulations cycloniques.

L'histogramme des moyennes spatiales des épisodes ($X_{i.} = \frac{1}{p} \sum_{j=1}^{p} X_{ij}$ sur les p valeurs de l'épisode i) a une allure assez lisse bien qu'assez dissymétrique (Fig.I-5b).

Par contre, les histogrammes du maximum ponctuel de l'épisode présentent 2 modes assez marqués (Fig.I-5d) même si le choix des classes a un effet assez sensible(Fig.I-5c). Ceci semble lié au nombre discret d'averses (1,2,3,parfois plus) ayant constitué l'épisode.

D'autre part, il semblerait que les épisodes ne présentent pas leur maximum ponctuel en n'importe quel point du réseau mais en 2 zones plus privilégiées :

Au Nord		Au Sud	
Loubaresse	9	Mt Aigoual	16
Villefort	8	Lasalle	3
Mayres	6	St Etienne V. Fse	2
Montpezat	4	St Maurice Ventalon	
Antrayges	4		
St Pierreville, Vernoux, Ste Eulalie, Aubenas, Privas, Malons	2		

Table VI : Nombre de cas où la station a reçu le maximum de l'épisode (parmi 84 épisodes)

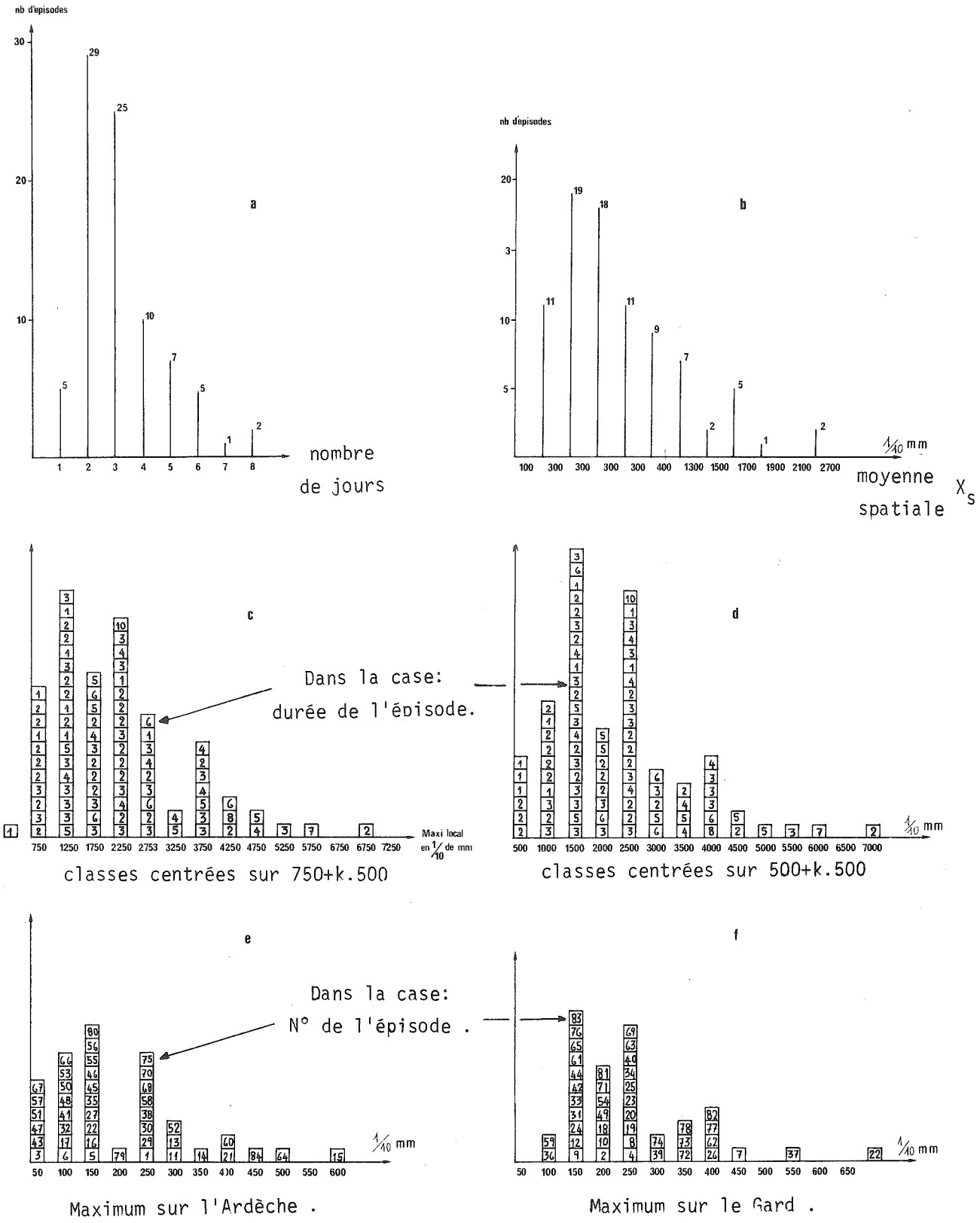


Figure I - 5 - Histogrammes des moyennes et des maxima des épisodes utilisés .

Si on fait l'histogramme des maximum en distinguant le Nord et le Sud du réseau, on retrouve les deux modes de I-5-d, et on constate un léger glissement des histogrammes : les épisodes sont en moyenne plus forts au Sud du réseau (fig.I-5-e et f).

III.2.2. Codage des variables

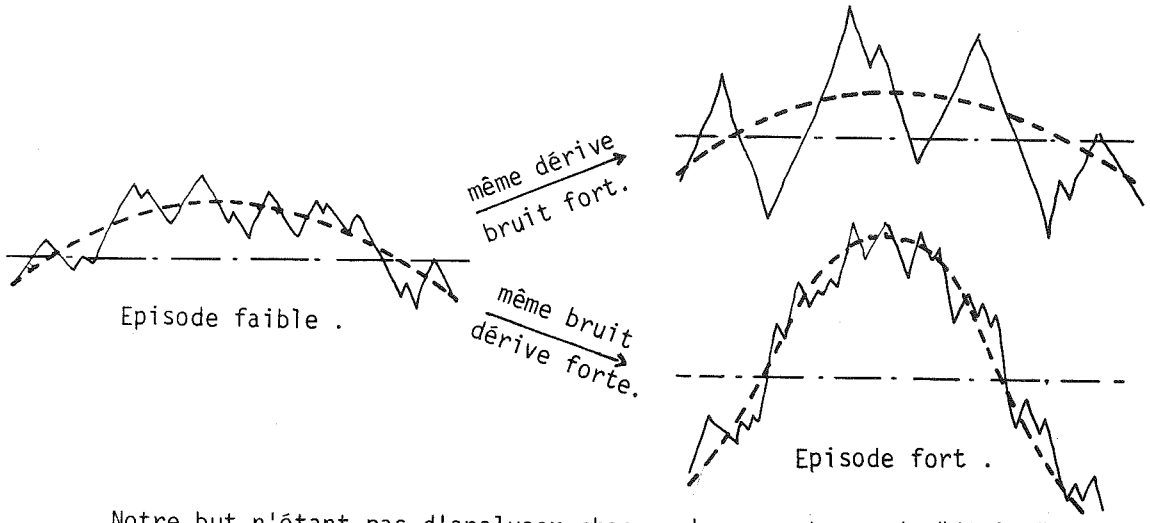
ⓐ Les premiers traitements furent effectués sur les valeurs brutes des épisodes. La nécessité d'utiliser des valeurs centrées et réduites est évidente, compte tenu de la forte liaison entre moyenne et écart-type climatologiques des stations (Fig. I-6a): Les zones de fortes valeurs moyennes ont aussi une plus forte variabilité.

ⓑ Cependant, notre but était d'analyser spatialement les données. Or là encore, il y a une liaison forte entre:

la moyenne spatiale d'un épisode : $\longrightarrow \quad X_{i\cdot} = \frac{1}{P} \sum_{j=1}^{P} X_{ij}$

et son écart-type spatial : $\longrightarrow \quad \sigma_s(i) = \sqrt{\frac{1}{P} \sum_{j} \left(X_{ij} - X_{i\cdot}\right)^2}$ (fig. I-6b)

$\Longrightarrow$ Un épisode est d'autant plus variable dans le champ qu'il est important. Remarquons cependant que cela peut avoir deux origines : soit un "bruit aléatoire" plus intense, soit tout simplement un facteur de "trend" ou de "dérive":



Notre but n'étant pas d'analyser chaque champ en terme de "dérive" et de "bruit" (cf CREUTIN , 1979), nous ne chercherons pas l'origine de la variance empirique du champ, mais par contre, si nous souhaitons comparer globalement les distributions spatiales de l'ensemble des champs, il nous faut éliminer la "taille" ou la "puissance" des épisodes. Cela revient à comparer les réseaux d'isohyètes , indépendamment des valeurs associées aux isohyètes, donc en valeurs adimensionnelles. Nous avons donc choisi de "normer" les valeurs de chaque champ par la moyenne arithmétique du champ (ou toute estimation de l'intégrale de la lame d'eau sur le champ). On obtient donc des données "profi-

On obtient donc des données "profilées" Y_{ij} qui expriment le poids en % de la station j dans l'épisode i.

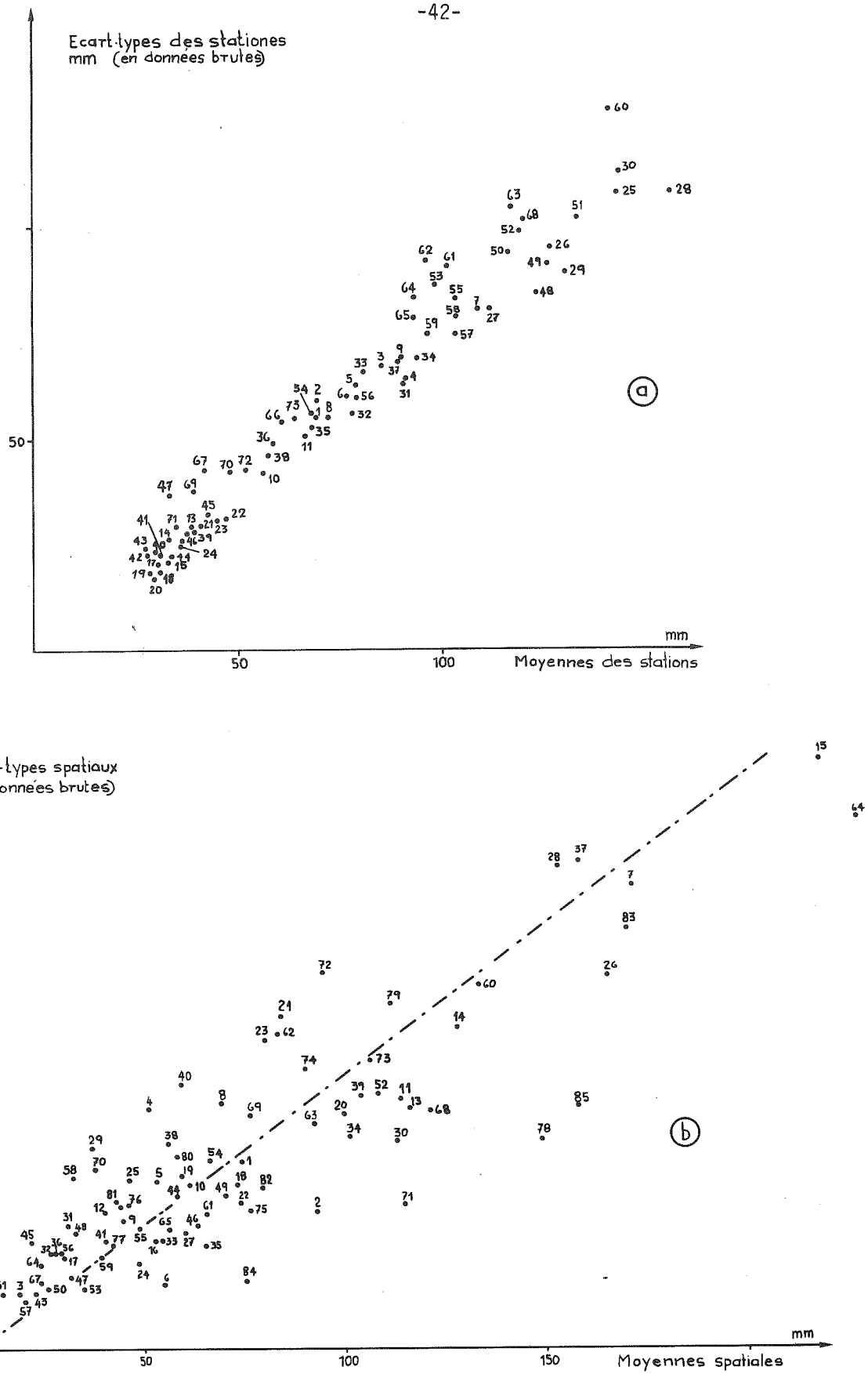
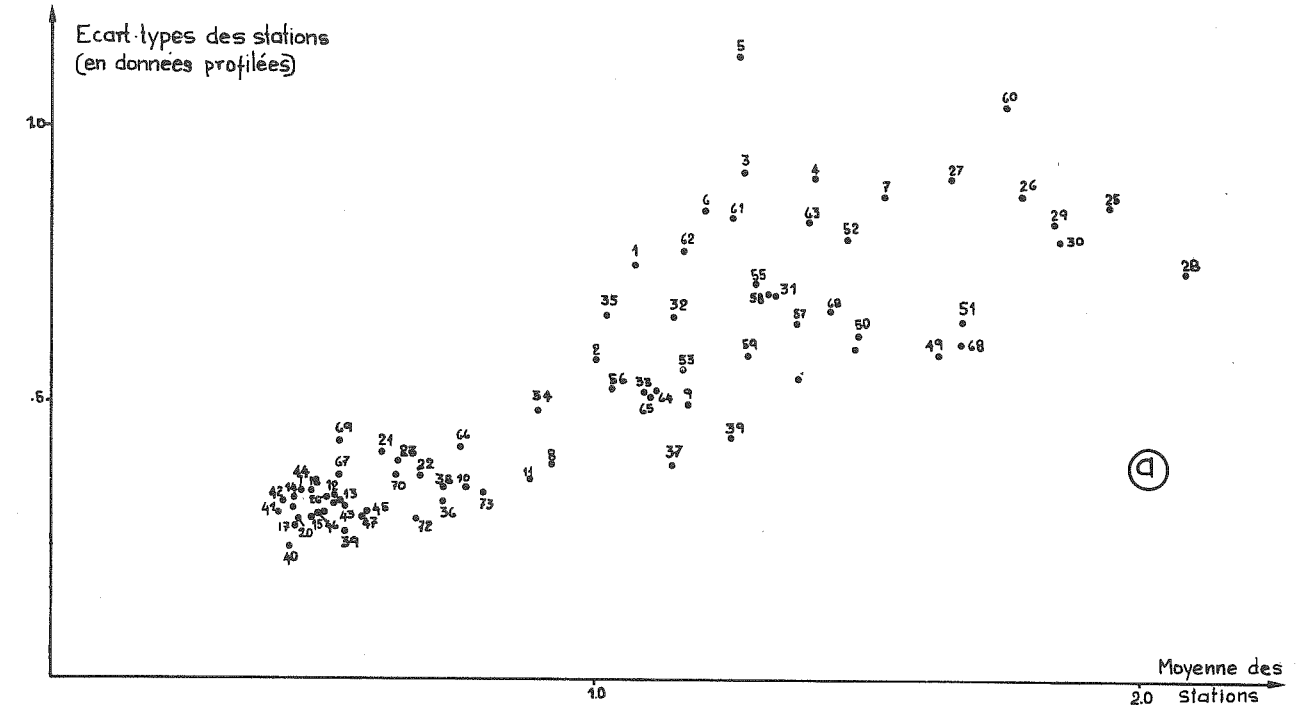


Figure I-6 -Corrélations entre moyennes et ecart-types des données brutes .

a) climatologiques (Chaque point est une station)
b) spatial (Chaque point est un épisode)



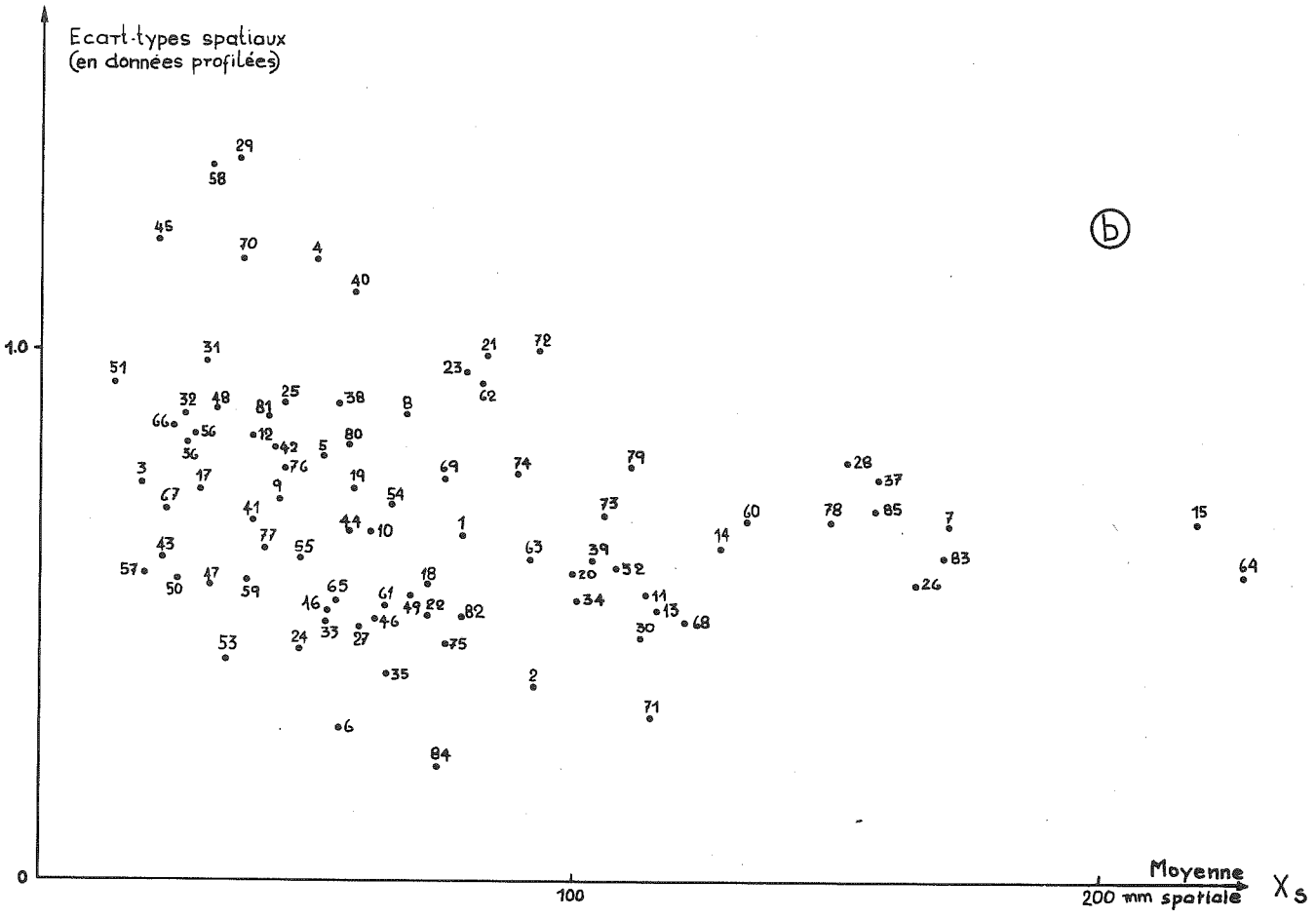


Figure I-7 -Corrélations entre moyennes et ecart-types des données profilées.

a) climatologiques (Chaque point est une station)
b) spatial (Chaque point est un épisode)

Compte tenu de la liaison entre moyenne spatiale et écart-type dans le champ, le fait de diviser par la moyenne donne pour les champs des valeurs profilées un écart-type dans le champ qui est quasi-constant (Fig.I-7a) $\sigma_s(i) \approx \sigma \; \forall$ l'épisode i .
Par contre si on regarde l'écart-type climatologique de chaque station en variable profilée on a encore une liaison, moins forte cependant entre σ_{Y_j} et $\overline{Y}_{.j}$ (Fig.I-7b).
Ceci met en évidence que les stations, en général plus fortes que les stations alentour, ont aussi une plus forte variabilité.

© Enfin le problème se pose, pour comparer les épisodes bruts, de savoir si un épisode de 50 et un épisode de 100 mm sont aussi distants qu'un épisode de 500 et un de 550 mm.

Cela se ramène à décider si une métrique euclidienne est valable en tous points de l'espace des points épisodes, ce qui est le cas pour un nuage multinormal.

Or chaque variable prise séparément est très dissymétrique (Fig.I-8-a) mais voisine d'une loi log-normale (Fig.I-8-b). On peut se poser le problème du décalage d'origine optimal à appliquer et utiliser les méthodes du chapitre II-2-4, mais les valeurs obtenues ne sont pas significativement différentes de zéro et n'améliorent pas sensiblement l'ajustement.

On vérifiera aussi que la simple transformation en $\sqrt{}$ améliore déjà sensiblement la dissymétrie de la distribution (Fig.I-8-c) mais que les variables profilées non transformées y parviennent tout autant sinon mieux (Fig.I-8-d).

En conclusion, la comparaison des épisodes se fera plutôt sur des données brutes transformées ou sur les données profilées.

On pourra au passage vérifier, sur les matrices de corrélations des variables brutes et transformées, les résultats théoriques du paragraphe II.3.1.

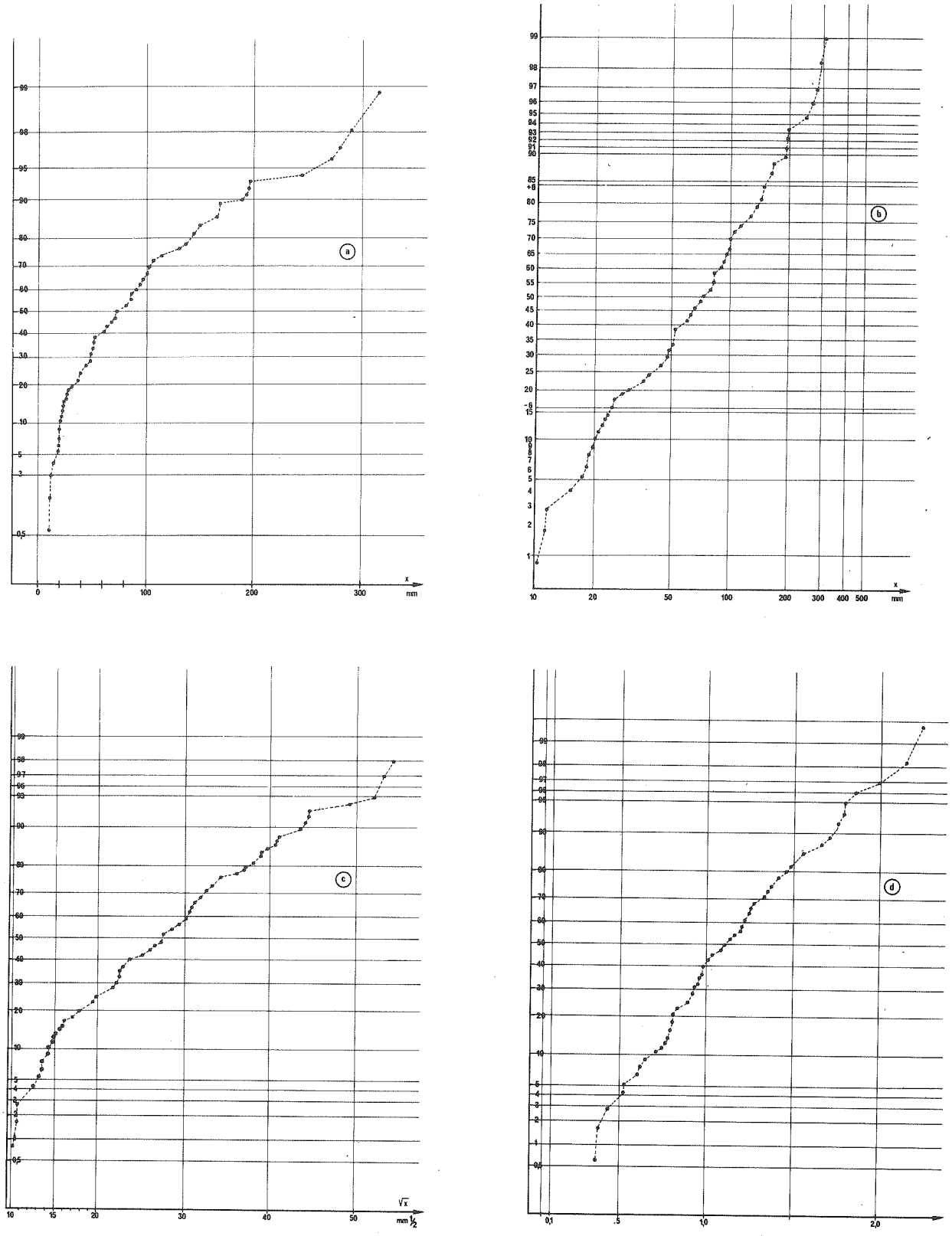


Figure I-8 -Distributions des valeurs brutes et transformées à S^{t}Etienne de Lugdarès.

a)Valeurs brutes sur papier normal .
b)valeurs brutes sur papier log-normal .
c)valeurs radicalisées sur papier normal .
d)valeurs profilées sur papier normal .

"Nul n'entre ici
s'il n'est géomètre"

Platon

DEUXIEME PARTIE

TECHNIQUES D'ANALYSES DES DONNÉES .

(INTERPRÉTATIONS-VISUALISATIONS-APPLICATIONS)

Les techniques que nous allons envisager ici sont essentiellement descriptives, et ont pour but de condenser l'information disponible dans notre paquet de données. Par descriptive, il faut entendre qu'elles réduisent au minimum les hypothèses ou les modèles extérieurs aux données initiales, mais précisons bien qu'il s'agit d'une description condensée de ces seules données initiales, et non du phénomène physique dont elles proviennent. Par condensation de l'information, on comprend qu'il y a optimisation de la description et l'optimisation s'applique à un critère, qui sera en général mais pas toujours, une notion d'inertie. Enfin, puisqu'"un dessin en dit plus qu'un long développement", elles fourniront généralement une représentation plane de l'ensemble multidimensionnel de départ.

Sur le plan théorique, elles ont fait l'objet de développements nombreux où les analogies géométriques se sont révélées très fructueuses, bien qu'un retour se dessine vers les interprétations probabilistes et fonctionnelles. Des efforts synthétiques visent à mettre en évidence les analogies de ces diverses techniques (cf J. DAUXOIS et A. POUSSE, 1976) en les considérant par exemple comme des cas particuliers d'analyse canonique.

Dans ce mémoire, notre but sera pratiquement inverse : loin de chercher par des généralisations parfois abstraites ce que ces méthodes ont en commun, nous analyserons plutôt ce qu'elles ont en particulier et ce qui permet à l'une ou à l'autre, de mieux prendre en compte les particularités de certaines données. Et, plutôt que de chercher, par des développements théoriques, à proposer de nouvelles méthodes, nous chercherons surtout à améliorer l'interprétation et l'utilisation des résultats des méthodes existantes. La présentation choisie est résolument classique, et nous n'utilisons pas ou peu les présentations condensées telles que les schémas de dualité (dont le potentiel synthétique indéniable nous a plutôt semblé utile pour résumer des résultats acquis que pour en entrevoir de nouveaux).

CHAPITRE I

ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

Cette méthode déjà ancienne (HOTTELING, 1933) est la plus connue en analyse des données car d'autres techniques, comme l'analyse des correspondances ou l'analyse discriminante, peuvent s'y ramener. De plus, elle est bien adaptée aux données que l'on rencontre en hydroclimatologie.

Sa formulation, désormais classique, ne sera décrite de façon détaillée que pour préciser des notations ; le lecteur déjà familiarisé peut ignorer les paragraphes I.1 et I.2. Dans le paragraphe I.3 nous nous attacherons à en interpréter les résultats tandis que le paragraphe I.4 présente de façon complète l'application qui en a été faite aux données de Davos. L'application aux pluies cévenoles est reportée dans la IVème partie.

Note : Dans la suite du texte, nous noterons A.C.P. l'analyse en composantes principales.

I.1 - Visualisation des individus

I.1.1. Effets d'échelle - Effets de taille

Un premier problème consiste, étant donné N individus caractérisés par p variables, à les comparer 2 à 2. Pour cela on peut calculer une distance entre ces individus, soit :

$$d^2(i,i') = d^2(X_{i\cdot}, X_{i'\cdot}) = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$$

Mais une première question concerne le choix des échelles au sein de la sommation.

Exemple : Soit 2 situations de précipitations journalières sur un réseau de pluviomètres. Si la station 1 transmet ses mesures en cm, tandis que les autres les transmettent en mm, cette station 1 recevra implicitement un poids de 10^{-2} dans la sommation et ne sera en fait pas utilisée dans la comparaison des 2 situations.

Ceci se produit en particulier quand les variables ne sont pas de même nature. Si les situations i et i' sont caractérisées par les précipitations x_1 et les températures x_2 relevées, il est difficile de mélanger dans $d^2(i,i')$ de tels paramètres sans les rendre comparables, c'est-à-dire adimensionnels.

La normalisation par l'écart-type n'est pas exempte d'arbitraire ni de modèles sous-jacents (essentiellement une loi normale) : on considèrera par exemple que la

distance entre 2 situations est la même si elles ont même précipitation et 1 écart-type de différence en température, ou même température et 1 écart type de différence en précipitation. Or on sait, vu la dissymétrie de la loi des précipitations journalières que + ou - 1 écart-type ne représente déjà pas la même chose.

Un autre problème, plusieurs fois évoqué dans ce mémoire, est celui des informations (ou variables) redondantes. Supposons les situations i et i' parfaitement bien caractérisées par 2 variables X_1 et X_2 statistiquement indépendantes, d'où :

$$d^2(i,i') = (x_{i_1} - x_{i'_1})^2 + (x_{i_2} - x_{i'_2})^2$$

Si, par souci de bien faire, on ajoute la variable X_3, en fait pratiquement identique à X_2, la distance devient :

$$d^2(i,i') = (x_{i_1} - x_{i'_1})^2 + [(x_{i_2} - x_{i'_2})^2 + (x_{i_3} - x_{i'_3})^2]$$

et là encore les distances représenteront surtout l'écart entre i et i' au sens de la variable 2 ou 3. Il s'agit cette fois d'un effet de taille des divers sous-ensembles présents dans l'ensemble des variables, qui modifieront les résultats des analyses utilisant de telles distances.

Enfin, quelle que soit la distance choisie, on se trouve devant une matrice de $P.(P-1)/2$ interdistances qu'il faut interpréter. Une méthode pour visualiser ces interdistances est l'analyse en composantes principales dans son aspect géométrique. Son aspect probabiliste, lié à une hypothèse de loi multinormale, est présenté par ANDERSON T.W. (1958) et utilisé par BOIS Ph. (1976) (p.52 et suivantes).

I.1.2. Présentation géométrique de l'analyse en C.P. (*)

Comme il est exclu de "voir" les N points et leurs interdistances, de manière parfaite, ailleurs que dans $\mathbb{R}^P$, on se contente de les voir imparfaitement, mais le mieux possible, dans un espace de dimension inférieure.

Un premier choix, arbitraire car on peut en imaginer d'autres (cf méthode de Sammon en II.3.1), consiste à utiliser pour passer d'un espace $\mathbb{R}^P$ à un espace $\mathbb{R}^{\ell < P}$, la projection orthogonale.

On cherche donc d'abord le sous-espace de dimension 1, c'est-à-dire la direction de vecteur directeur $\vec{u}$, telle que les distances entre les projections orthogonales des points sur cette direction soient les plus proches possible des distances dans $\mathbb{R}^P$.

Sur la direction u : $d^2_u(X_{1v}, X_{2v}) = [u^t.(X_{1v} - X_{2v})]^2$

et on choisira u tel que, en moyenne quadratique, (donc pour tous les points pris 2 à 2)

$$S^2 = \sum_{k,l=1}^{N} d^2_u(X_{kv}, X_{lv}) \qquad \text{maximum}$$

Ici deux remarques s'imposent :

- Tout d'abord, l'origine n'intervient pas directement dans le choix de u, qui dépend seulement des distances entre points.
- L'espace de départ peut être soit l'espace des variables brutes (ou seulement centrées), soit celui des données centrées réduites si on veut s'affranchir ainsi

des effets d'échelle. Cela revient à faire l'analyse avec une métrique, associée au produit scalaire et au calcul des distances, égale à I ou à :

$$D_\sigma = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_p^2} \end{bmatrix}$$

Mais le choix de vecteurs unitaires, sur les axes de départ associés aux variables, impose aussi, sur l'axe u , la contrainte :

$$\| u \| = u^t . u = 1$$

afin que les distances projetées sur les divers axes soient comparables.

D'autre part, le choix de maximiser les distances en moyenne quadratique n'est pas le seul possible (autre exemple $S = \sum_{k,l} d_u (X_{kv}, X_{lv})$: moyenne arithmétique).

Mais la dérivation par rapport à u est alors plus délicate et surtout on ne dispose plus de la règle de Pythagore valable pour le carré des distances.

$$d^2(X_{kv}, X_{lv}) = d^2_u (X_{kv}, X_{lv}) + d^2_{\perp u} (X_{kv}, X_{lv})$$

C'est l'une des propriétés essentielles du projecteur choisi (projection I orthogonale ou R orthogonale) linéarité et décomposition en somme directe (cf CAILLEZ, MAILLES et PAGES, 1973).

La solution du problème d'optimisation est classique et sa démonstration est évoquée en annexe I ou dans LEBART et FENELON (1973). On sait que cela revient à diagonaliser, selon l'espace de départ choisi, R ou T matrices de corrélation ou de variances covariances entre les données.

I.1.3. Aide à l'interprétation des axes, ou facteurs obtenus (*)

Le choix de l'origine étant sans importance (puisque seule compte la dispersion du nuage), si on la suppose au barycentre du nuage, alors toute variable X_j est centrée :

$$\sum_i x_{ij} = \overline{x_j} = 0$$

La coordonnée de l'observation i sur l'axe u_k s'écrit : $z_{ik} = X_{iv} . u_k$

et l'ensemble des observations de la nouvelle variable Z_k est le vecteur :

$$Z_{ok} = X_{ov} . u_k$$

On vérifie que Z_k est aussi centrée :

$$\overline{z_k} = \sum_i z_{ik} = \left| \sum_i X_{iv} \right| . u_k = \left| \sum_i x_{i1} \sum_i x_{i2} \cdots \sum_i x_{ip} \right| . u_k = \vec{0} . u_k = 0$$

et que sa variance s'écrit : $\frac{1}{N} Z_{ok}^t . Z_{ok} = \frac{1}{N} u_k^t X^t . X . u_k = u_k^t T u_k = \lambda_k$

La corrélation entre la variable X_{0j} de variance σ_j^2 et Z_{0k} est donc :

$$r\left(F_k, X_j\right) = r_{k,j} = \frac{1}{N} \cdot \frac{X_{0j}^t \cdot Z_{0k}}{\sigma_j \cdot \sqrt{\lambda_k}}$$

Et si on considère les corrélations avec l'ensemble des variables :

$$R_k = \begin{bmatrix} r_{k1} \\ r_{k2} \\ \vdots \\ r_{kp} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \frac{X_{01}^t}{\sigma_1} \\ \frac{X_{02}^t}{\sigma_2} \\ \vdots \\ \frac{X_{0p}^t}{\sigma_p} \end{bmatrix} \cdot \frac{Z_{0k}}{\sqrt{\lambda_k}} = \frac{X^t \cdot D_\sigma \cdot Z_{0k}}{N \sqrt{\lambda_k}} = D_\sigma T \cdot \frac{U_k}{\sqrt{\lambda_k}} = \sqrt{\lambda_k} \cdot D_\sigma U_k$$

Si les variables sont réduites :

$\sigma_j = 1$ alors $R_k = \sqrt{\lambda_k} \cdot U_k$ $\quad r_{kj} = \sqrt{\lambda_k}\, u_{kj}$

sinon : $R_k = \sqrt{\lambda_k} \cdot D_\sigma \cdot U_k$ $\quad r_{kj} = \frac{\sqrt{\lambda_k}\, u_{kj}}{\sigma_j}$

Le vecteur R_k est parfois appelé "structure du facteur F_k ".

I.1.4. Analyse en covariance ou en corrélation

On a vu en I.1.1 que selon les données, on pouvait être amené à analyser soit les données simplement centrées, et donc leur matrice de var-covariance T, soit les données centrées réduites, d'où leur matrice de corrélation R.

Avec $T = \frac{1}{N} X_{0v}^t \cdot X_{0v}$ $\qquad R = \frac{1}{N} D_\sigma \cdot X_{0v}^t \cdot X_{0v} \cdot D_\sigma$

Les 2 matrices n'ont évidemment pas les mêmes directions propres et les coordonnées des points observations sur les axes 1, 2 ... diffèrent selon que l'on choisit R ou T. Cependant, si l'on renorme les facteurs, de sorte que la variance des observations soit égale à 1 sur chaque facteur, alors on obtient un résultat intéressant, que nous utiliseront plus loin (analyse discriminante, Chap. IV.2) : la configuration des points observations transformés est la même, à une rotation près dans les axes de R ou T (cf Annexe II).

I.2 - Visualisation des variables

I.2.1. Présentation géométrique

On peut, de la même façon que pour les observations, considérer les p points variables caractérisés chacun par leur N coordonnées, déduites des valeurs initiales x_{ij} par :

r_{ij} = coordonnées de la variable j sur le $i^{ème}$ axe = $x_{ij} - \overline{x_j}$

avec $\overline{x}_j = \frac{1}{N} \Sigma\, x_{ij}$

ce qui équivaut à projeter les points variables sur un hyperplan perpendiculaire à la première bissectrice (dont le vecteur directeur est $\frac{1}{N}$) donc à un produit matriciel par $P = \{p_{ij}\} = \{\delta_{ij} - \frac{1}{N}\}$.

Toutefois, il s'agit d'un centrage dans l'espace des variables $\mathbb{R}^P$, où la somme des coordonnées d'une variable j est nulle : $\sum_i x_{ij} = 0$

mais il ne s'agit pas d'un centrage de nuage étudié dans $\mathbb{R}^N$ espace des observations car la somme des coordonnées des p points variables sur une direction, par exemple la $i^{ème}$, est :

$$\sum_j x_{ij} = \sum_j (x_{ij} - \bar{x}_j) \neq 0$$

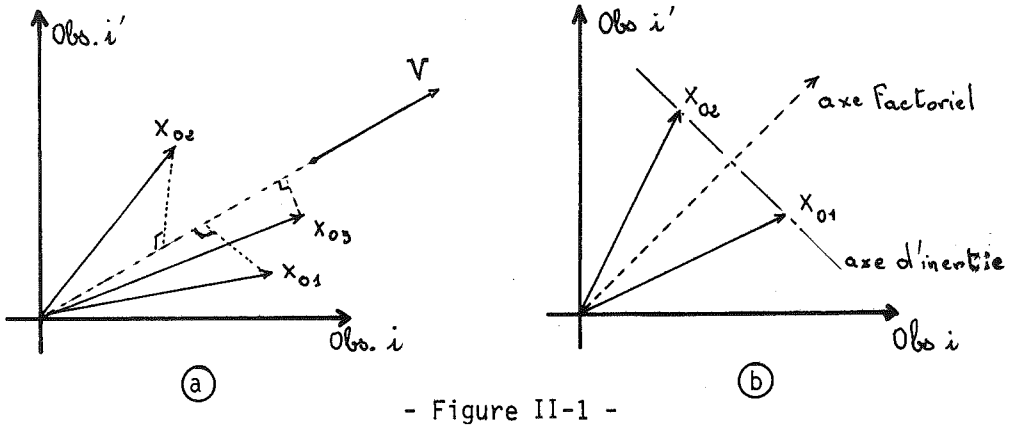
On peut alors chercher à effectuer l'analyse de ce nuage de la façon suivante :

- Chercher la droite, passant par l'origine, sur laquelle se projette au mieux le nuage: L'origine joue donc un rôle particulier, or celle-ci est arbitraire, au moins dans $\mathbb{R}^N$, puisque définie par une propriété qui n'est vraie que pour $\mathbb{R}^P$ (où l'origine est barycentre des points observations) (Fig. II-1a).

La droite de vecteur directeur $\vec{V}$, telle que la somme des distances des points à celle-ci soit minimum est aussi celle qui maximise la somme des carrés des projections des vecteurs $\overrightarrow{OX_{oj}}$:

$$\max \sum_{j=1}^{p} OX_{oj}^2 = \max \left(\sum_{j=1}^{p} \vec{V}.\overrightarrow{OX_{oj}} \right)^2 = \max V^t \left(\sum_{j,l=1}^{p} X_{oj}.X_{ol} \right) V = \max V^t X.X^t V$$

si l'on considère que X est la matrice des données déjà centrées (comme en I.1.1).



- Figure II-1 -

On remarquera que l'interprétation géométrique est différente entre le cas des variables (ci-dessus) et le cas des observations (en I.1.1). En effet, la droite obtenue n'est plus celle qui maximise la projection des distances interpoints, indépendamment de l'origine. L'axe V n'est plus axe d'inertie du nuage des points variables de $\mathbb{R}^N$: cela se vérifie aisément par le calcul et un exemple simple permet de s'en persuader (Fig.II-1b).

Enfin, on peut rappeler quelques propriétés importantes de la représentation dans $\mathbb{R}^N$:

- la distance de l'origine au point variable j,

$$OX_{oj} = \sum_{i=1}^{N} x_{ij}^2 = \sigma_j^2 = \text{variance de la variable}$$

Si les variables centrées sont de plus réduites, ou normées, tous les points variables sont sur une hypersphère de centre 0 et de rayon 1.

- La distance entre 2 points variables est égale à :

$$d^2(X_{o1}, X_{o2}) = \overrightarrow{OX}_{o1}^2 + \overrightarrow{OX}_{o2}^2 - 2\,\overrightarrow{OX}_{o1}.\overrightarrow{OX}_{o2} = \sigma_1^2 + \sigma_2^2 - 2\sigma_1.\sigma_2.r_{12}$$

où r_{12} est le coefficient de corrélation entre les variables 1 et 2, en même temps que le cosinus de leur angle dans $\mathbb{R}^N$. Si celles-ci sont normées on a : $d^2(X_{o1}, X_{o2}) = 2(1 - r_{12})$

L'analyse en composantes principales comme technique de concentration de l'information consiste, dans $\mathbb{R}^P$, à rechercher le nombre d'axes U_k (variables principales Z_k) réellement indépendants et significatifs ($\lambda \geqslant$ seuil à définir). Dans $\mathbb{R}^N$ cela revient à rechercher la dimension l du sous-espace de $\mathbb{R}^N$ engendré par les p points variables, plus l'origine. Or p+1 points engendrent au plus un espace de dimension p donc encore $l \leqslant p$. En fait, on verra au paragraphe suivant que l'analyse est la même dans $\mathbb{R}^P$ ou $\mathbb{R}^N$, et l'une ou l'autre de ces interprétations doit conduire au même résultat.

I.2.2. Résolution (*)

Le problème d'optimisation : $\max \; V^t.X.X^t.V$ avec $V^t.V = 1$

peut se résoudre comme en I-1 en cherchant les éléments propres de $X.X^t$, mais cela est inutile si on connaît déjà ceux de $X^t.X$, ou plutôt de $\frac{1}{N} X^t.X$, soit λ_k et U_k ($k = 1, \ldots p$). En prémultipliant par X :

$$\frac{1}{N} X^t.X\,U_1 = \lambda_1 U_1 \Longrightarrow \frac{1}{N}(X.X^t)X\,U_1 = \lambda_1 (X U_1)$$

ce qui traduit que $X.X^t$ a pour p premières valeurs propres celles de $X^t.X$ soit $N.\lambda_k$, les N-p restantes étant nulles, et pour vecteurs propres correspondant XU_k.

Si on veut qu'ils soient aussi normés, il suffit, compte tenu de :

$U_k^t.X^t.X.U_k = N.\lambda_k$, de prendre : $V_k = \frac{1}{\sqrt{N\lambda_k}} X.U_k$

La coordonnée d'une variable j sur l'axe 1 sera :

$$G_{1j} = X_{oj}^t.V_1 = \frac{1}{\sqrt{N\lambda_1}} X_{oj}.X.U_1 = \frac{1}{\sqrt{N\lambda_1}} \left(X_{oj}^t.X_{o1}\; X_{oj}^t X_{o2} \cdots X_{oj}^t.X_{op}\right).U_1$$

Or $X_{oj}^t.X_{ok} = N\sigma_j.\sigma_k.r_{jk}$ ou $N.r_{jk}$ si les variables sont normées, d'où :

$$G_{1j} = \sqrt{N}\,\sqrt{\lambda_1}\,\left(r_{1j}\; r_{2j} \cdots r_{pj}\right).U_1$$

C'est à un facteur près, le produit scalaire de la ligne j de R avec U_1.

Une autre expression permet d'obtenir les projections de toutes les variables sur l'axe 1 de $\mathbb{R}^N$:

$$\begin{bmatrix} G_{11} \\ G_{12} \\ \vdots \\ G_{1j} \\ \vdots \\ G_{1p} \end{bmatrix} = \frac{1}{\sqrt{N\lambda_1}} \cdot \begin{bmatrix} X_{01}^t \\ X_{02}^t \\ \vdots \\ X_{0j}^t \\ \vdots \\ X_{0p}^t \end{bmatrix} \cdot X \cdot U_1 = \frac{1}{\sqrt{N\lambda_1}} \cdot X^t \cdot X \cdot U_1 = \sqrt{N\lambda_1} \cdot U_1$$

Or on a vu en I-1 que la corrélation entre une variable X_{0j} et le facteur F_1 de $\mathbb{R}^p$ est $r_{F_1, X_j} = \sqrt{\lambda_1}\, u_{1j}$

Donc :

- les projections des points variables dans $\mathbb{R}^N$ sont, au facteur $\sqrt{N\lambda_1}$ près, les coordonnées du vecteur U_1 de $\mathbb{R}^p$, et au facteur $\sqrt{N}$ près, les coefficients de corrélations entre la variable principale Z_1 et les p variables X_j, $j = 1 \cdots p$ (Il faut y ajouter le facteur $\frac{1}{\sigma_j}$ si celles-ci ne sont pas normées).

La démarche des paragraphes I.1 et I.2 peut alors se resumer :

dans $\mathbb{R}^p$		dans $\mathbb{R}^N$	
axes initiaux	⟶ variables	axes initiaux	⟶ observations
points	⟶ observations (centrées) barycentre = origine	points	⟶ variables dans un hyperplan ⊥ à la 1ère bissectrice (nuage non centré) .
axes factoriels	⟶ maximisent la représentation du nuage des observations (de façon intrinsèque l'origine)	axes factoriels	⟶ ajustent le nuage au sens moindres carrés . (avec la contrainte de passer par l'origine)
vecteurs directeurs des axes factoriels U_k	vecteurs propres de la matrice de var-covariance $\frac{1}{N} X^t.X$	vecteurs directeurs	se déduisent simplement de ceux de $X^t.X$
nouvelles coordonnées des points observations	se calculent par projection sur les nouveaux axes	nouvelles coordonnées des points variables	se déduisent simplement des composantes des vecteurs propres U_k (ci-contre)
liaison entre axes factoriels et axes initiaux	Les coordonnées des points des observations sur l'axe initial j et sur l'axe factoriel k sont en corrélation : $\rho(Z_k, X_j) = \sqrt{\lambda_k}\, u_{kj}$	s'interprètent comme des corrélations entre variables et facteurs de $\mathbb{R}^p$, donc fournissent une représentation graphique de ces liaisons (ce qui est le but de l'opération, plus que de représenter le nuage de points variables.)	

Remarque . Dans $\mathbb{R}^N$ et sur 2 axes factoriels, on peut aussi interpréter la distance entre 2 points variables. On se rappellera toutefois que cette distance n'est pas représentée de façon optimale, et que la distance maximale est obtenue entre 2 variables parfaitement anticorrélées!

Extension. Si r_{jk} est le coefficient de corrélation entre la variable X_j et le facteur F_k, l'orthogonalité des facteurs permet d'écrire les corrélations multiples :

$$R^2_{X_j . F_k, F_\ell} = R^2_{j . k, \ell} = r_{jk}^2 + r_{j\ell}^2$$

et plus généralement :

$$R^2_{j . 1, \ldots q} = \sum_{k=1}^{q} r_{jk}^2$$

Si $q = p$, alors $R^2_{j . 1, \ldots . p} = 1$.

Cela se comprend aisément (cf BOIS Ph., 1976) car la corrélation multiple, considérée comme projection orthogonale dans $\mathbb{R}^N$ sur un sous-espace engendré par des variables elles-mêmes orthogonales, se ramène ici à une somme directe.

Cela fournit non plus la part de la variance d'une variable expliquée par 1 facteur mais par n'importe quel sous-ensemble de facteurs.

I.3 - Quelques problèmes associés à l'interprétation des facteurs

I.3.1. Interprétation des facteurs : Effets de taille, effet de forme

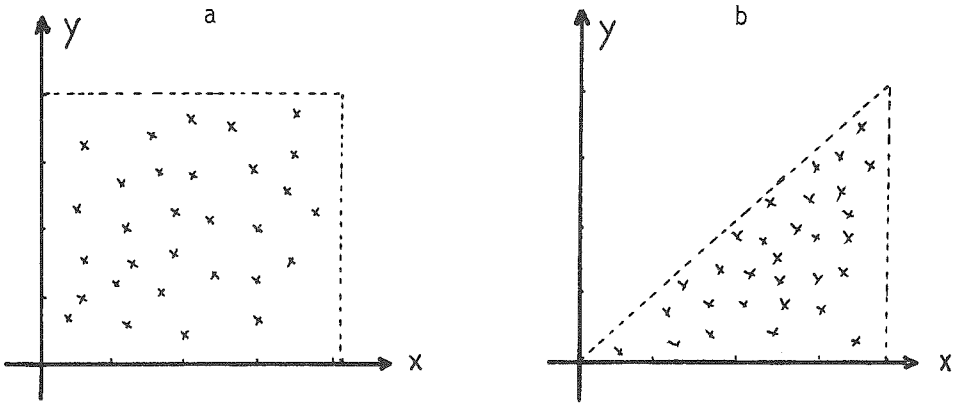
Il ne s'agit plus ici d'interpréter les facteurs au sens de leurs liaisons avec des variables initiales mais de chercher s'ils présentent un sens physique en eux-mêmes.

Inversement, on peut chercher à analyser une notion physique particulière, et une analyse en composantes principales convenablement menée peut condenser cette notion dans un seul facteur à condition d'avoir effectué au préalable un codage ad hoc. Nous en verrons un exemple concret (pluviométrie des Cévennes, IVème Partie) mais un exemple artificiel permet de s'en faire une idée.

Exemple simple : statistique des rectangles et des parallélépipèdes

On va regarder ce que donne l'A.C.P. de quelques formes géométriques simples, caractérisées par des variables correspondant à leurs dimensions habituelles (longueur et largeur) . On peut imaginer par exemple que l'administration des postes décident d'effectuer une analyse des lettres et des paquets déposés dans les boîtes aux lettres ! : Compte tenu des dimensions maximales et minimales autorisées et de la non-normalisation des emballages, on peut supposer que les dimensions sont tirées d'une loi uniforme ramenée à l'intervalle (0,1) et sont indépendantes.

a) Si on se place à 2 dimensions, et si on suppose que l'expérimentateur mesure sans distinction les deux dimensions, on a le nuage de la figure II-2 a), alors que s'il porte en X la plus grande des dimensions et en Y la plus petite, on a le nuage de la fig.II-2 b).



-Figure II-2

La corrélation linéaire est nulle dans le 1er cas, mais positive dans le second. Pour le démontrer, on peut vérifier que, si les densités de probabilité marginales sont constantes :

$$f(x)=1 \text{ si } 0<x<1 \text{ , } 0 \text{ sinon} \qquad g(y)=1 \text{ si } 0<y<1 \text{ , } 0 \text{ sinon}$$

et compte tenu de l'indépendance, la densité du couple (x,y) est :

$$h(x,y) = f(x).g(y) = 1$$

Or l'opération de sélection des dimensions a consisté à rabattre le triangle supérieur sur le triangle inférieur, chaque point situé au-dessus de la bissectrice étant remplacé par son symétrique par rapport à celle-ci. On en déduit que, sur le triangle inférieur :

$$h'(x,y) = 2.h(x,y) = 2 \quad \text{et} \quad h'(x,y) = 0 \quad \text{en dehors.}$$

Le calcul du coefficient de corrélation, présenté en annexe III, donne alors $r = 0.5$.

Donc le coefficient de corrélation entre 2 variables tirées indépendamment d'une loi uniforme, mais ensuite classées, devient positif.

b) Si on poursuit l'analyse , la matrice de corrélation:

$$\begin{bmatrix} 1. & 0.5 \\ 0.5 & 1. \end{bmatrix}$$ a les valeurs propres $\lambda_1 = \frac{3}{2}$ $\lambda_2 = \frac{1}{2}$

et les vecteurs propres: $u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ $u_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Remarque Dans cet exemple R et T ont mêmes vecteurs propres car $\sigma_x = \sigma_y$

Le premier axe que l'on obtient ($\frac{1}{\sqrt{2}}$ longueur + $\frac{1}{\sqrt{2}}$ largeur) est un effet de taille évident lié à la surface du rectangle (longueur et largeur fortes ou faibles en même temps).

Le second axe, par contre, sera un axe de "similitude" ou de forme. Au-dessus les rectangles trapus, en-dessous les rectangles élancés et on pourra, par exemple, sur cet axe, effectuer une typologie.

Toutefois, on peut se demander s'il est raisonnable d'effectuer cette analyse sur les valeurs brutes des longueurs et des largeurs :

. la "taille" du rectangle sera représentée par la quantité $\frac{X+Y}{\sqrt{2}}$,

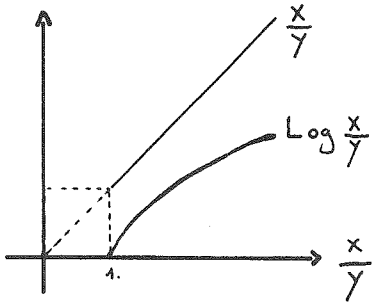
assez différente de la surface réelle X.Y.

. la forme, ou l'élancement du rectangle sera représentée par X-Y, assez différent de X/Y.

On voit immédiatement qu'il faudrait effectuer une transformation logarithmique sur les variables car :

$$\log X + \log Y \longrightarrow \log X.Y$$
$$\log X - \log Y \longrightarrow \log \frac{X}{Y}$$

Alors l'axe 1 rassemble complètement l'effet "taille", tandis que l'axe 2 explique en totalité la "forme". Remarquons toutefois qu'il faut ensuite appliquer la transformation inverse car sinon on atténue les valeurs extrêmes de l'élancement :



Une typologie des rectangles devrait par exemple se faire sur e^{Z_1} ou e^{Z_2}, Z_1 et Z_2 étant les C.P.

Par contre, dans l'analyse sur données brutes, on va trouver en même position sur l'axe 2 des rectangles ayant même différence X - Y mais des surfaces très différentes, ce qui correspond en fait à des élancements sensiblement différents. La valeur des résultats restera assez grossière à cause d'une linéarisation abusive du problème tandis que, après transformation des variables, le problème est parfaitement linéarisable.

On peut facilement étendre le problème à 3 dimensions (cf annexe III).

I.3.2. Un cas particulier intéressant : la matrice d'équicorrélation

Un cas extrême de l'effet "taille" peut se produire (ou au moins se simuler) si l'on suppose que les p variables sont en fait 1 seule et même information, Z, à laquelle on a ajouté respectivement un bruit ε_1 pour obtenir X_1, id. pour $X_2 \ldots X_p$, d'où :

$$X_1 = Z + \varepsilon_1 \qquad \text{avec} \quad \sigma_{X_1} = \sigma$$

et de même :

$$X_2 = Z + \varepsilon_2 \qquad \sigma_{X_2} = \sigma$$
$$\cdots\cdots\cdots \qquad \cdots\cdots$$
$$X_p = Z + \varepsilon_p \qquad \sigma_{X_p} = \sigma$$

Avec, dans ce cas particulier, $\sigma^2_{\varepsilon_i} = \sigma^2_{\varepsilon} = \sigma^2 - \sigma^2_Z \quad \forall i$

et les corrélations :

$$\rho(\varepsilon_i, Z) = 0 \ \forall i$$
$$\rho(\varepsilon_i, \varepsilon_j) = \delta_{ij}$$

ce qui fait que l'on a :

$$\rho(X_i, X_j) = \text{constante } \rho = 1 - \sigma^2_{\varepsilon} / \sigma^2$$

D'où une matrice de corrélation très particulière (équicorrélation) :

$$R = \begin{bmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \hline \rho & \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

On démontre alors (annexe IV) que l'on a une valeur propre dominante :

$$\lambda_1 = 1 + (p-1)\rho$$

et (p-1) valeurs propres égales:

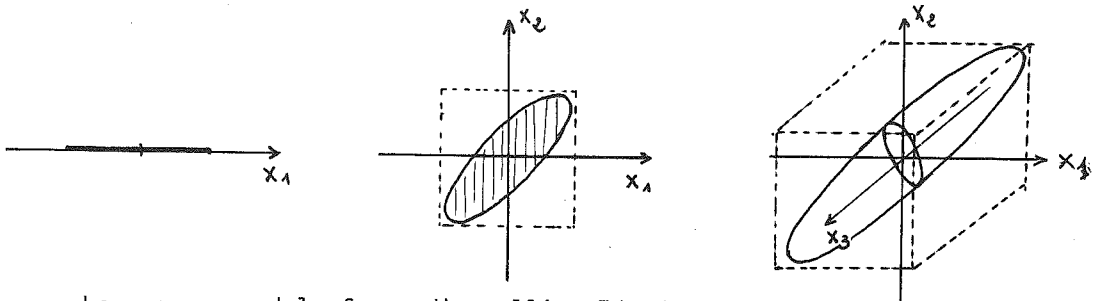
$$\lambda_2 = \cdots = \lambda_p = 1 - \rho$$

On vérifie aussi que le vecteur propre associé à λ_1 est $U_1 = \begin{bmatrix} 1/\sqrt{p} \\ 1/\sqrt{p} \\ \vdots \\ 1/\sqrt{p} \end{bmatrix}$

la première composante n'est autre que la moyenne des p variables, ce qui traduit bien un effet de taille.

Remarque. On peut s'étonner de n'obtenir que p composantes indépendantes alors que l'on a généré Z, puis p termes aléatoires ε_i mais Z n'apparaissant pas explicitement, c'est comme si on avait pris 1 variable particulière comme référence, puis généré p-1 termes aléatoires.

On peut se faire une idée intuitive et géométrique de l'équicorrélation en supposant que l'on génèrent X_1, variant entre $\pm$ 1, puis X_2 égale à X_1 mais à un terme aléatoire ε_2 près, etc...



Le nuage prend la forme d'un ellipsoïde de révolution autour de la diagonale de l'hypercube de R^p dont la longueur (si le côté vaut 1) est $\sqrt{p}$.

On voit que cette matrice d'équicorrélation présente des propriétés intéressantes. Il peut d'ailleurs être utile, quand N (nombre d'observations sur lesquelles on a calculé R) n'est pas trop faible, de tester l'hypothèse : R est une matrice d'équicorrélation.

On trouve dans D.F. MORRISON (1967, p.251) un test dû à LAWLEY qui teste l'hypothèse :

$$H_0 : \rho_{ij} = \rho \quad \forall i,j$$

ρ_{ij} étant estimé par r_{ij} sur un N-échantillon.

Un autre cas extrême est celui où toutes les variables sont indépendantes. Dans ce cas, tous les facteurs principaux ont une valeur propre associée égale à 1. Si on a estimé la matrice R sur un N-échantillon, on peut tester la circularité à l'aide du test :

$$\chi^2 = -\left(N - 1 - \frac{2p+5}{6}\right) \mathrm{Log}|R|$$

dû à BARTLETT, ou une approximation due à LAWLEY (cités par D.F. MORRISON, 1967)

$$\chi^2 = \left(N - 1 - \frac{2p+5}{6}\right) \sum_{i<j} \cdot \sum_{j} r_{ij}^2$$

distribuée selon un χ^2 à $p(p-1)/2$ degrés de liberté.

I.3.3. Effets de taille "emboités"

En fait, dans la pratique, on rencontrera plutôt des cas intermédiaires où l'on a des "paquets" de variables, très corrélées entre elles, et assez peu entre paquets. Par exemple, on aura une information repétée environ 15 fois avec des bruits divers, et une autre dix fois. Cela donnera 2 valeurs propres importantes $\lambda_1 \sim 12$ et $\lambda_2 \sim 8$, suivies de valeurs propres fortement décroissantes. Il s'agit alors de 2 effets de taille distincts, et on gardera par exemple 2 composantes principales, mais il ne faut plus associer à leurs valeurs propres, ni à l'ordre de celles-ci, une importance quelconque. Il vaut même mieux renormer à 1.0 les variables principales Z_k, plutôt que de leur donner, en les laissant normées par $\sqrt{\lambda_k}$, des poids complètement factices.

Nous verrons que ce problème se pose quand on traite des ensembles de variables de natures diverses, parfois très corrélées entre elles et généralement présumées explicatives d'un phénomène externe que l'on veut étudier.

Exemple. Supposons que l'on ait 2 paquets de variables de dimensions respectives p et q. Au sein des paquets on suppose les variables équicorrélées, tandis que les 2 paquets sont indépendants ($\forall\, l \in$ paquet 1 et $k \in$ paquet 2, $\rho_{lk} = 0$). La matrice de corrélation s'écrit donc :

$$R = \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix}$$

$$R_{11} = \begin{bmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_1 \\ & & \ddots & & \\ \rho_1 & \rho_1 & \rho_1 & \cdots & 1 \end{bmatrix} \quad \text{dim. } p$$

$$R_{22} = \begin{bmatrix} 1 & \rho_2 & \cdots & \rho_2 \\ \rho_2 & 1 & \cdots & \rho_2 \\ & & \ddots & \\ \rho_2 & \rho_2 & \cdots & 1 \end{bmatrix} \quad \text{dim. } q$$

et on suppose ρ_1 et $\rho_2 > 0$.

On peut vérifier que toute valeur propre λ_i de R_{11} est valeur propre de R associée au vecteur :

$$W = \begin{bmatrix} U_i \\ 0 \end{bmatrix} \begin{matrix} \} p \\ \} q \end{matrix}$$

En effet :

$$\begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix} \times \begin{bmatrix} U_i \\ 0 \end{bmatrix} = \lambda_i \cdot \begin{bmatrix} U_i \\ 0 \end{bmatrix}$$

De même toute valeur propre μ_j de R_{zz} associée au vecteur propre V_j est valeur propre de R associée au vecteur propre : $W = \begin{bmatrix} 0 \\ V_j \end{bmatrix}$

Comme la somme $\sum_{i=1}^{p} \lambda_i + \sum_{j=1}^{q} \mu_q = p+q$ on a trouvé toutes les valeurs propres.

Ceci s'étend au cas de 3, 4, ... n paquets. Dans ce cas, on peut vérifier que les 1ères valeurs propres de chaque paquet sortent en premier comme valeurs propres de R :

La 1ère valeur propre de n'importe quel paquet de dimension p et de corrélation ρ_i est supérieure à la valeur propre multiple d'un autre paquet de corrélation ρ_e car : $\lambda_{i1} = 1 + (p-1)\rho_i > 1 - \rho$

en effet $(p-1)\rho_i > -\rho \quad \forall \rho_i, \rho$ car ρ_i et $\rho > 0$

Par contre l'ordre d'apparition des 1ères valeurs propres ne dépend pas seulement de la dimension du paquet dont elles proviennent mais de sa corrélation propre.

On peut avoir : $1 + (p-1)\rho_i > 1 + (q-1)\rho_j$ même si $p < q$

On vérifie encore que l'ordre de sortie des 1ères valeurs propres et leurs valeurs ne sont pas significatifs. De plus, l'écart entre la dernière valeur propre significative $1 + (p-1)\rho_i$ et la première non significative $1 - \rho_j$ peut être faible (si on a un paquets avec ρ_i, ρ_j faible), et le passage se fait toujours au voisinage de 1, la valeur 1 étant toujours significative (paquet de 1 variable indépendante).

On voit aussi que si on a 2 variables faiblement corrélées :

$$\begin{bmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{bmatrix} \longrightarrow \lambda_1 = 1+\varepsilon \quad \lambda_2 = 1-\varepsilon$$

Il faut donc garder les 2 variables et donc les 2 valeurs propres (cf. ci-dessus) car ces 2 informations sont quasi-indépendantes.

En conclusion, on voit comment des redondances voulues ou non au sein de paquets de variables eux-mêmes indépendants créent des effets de taille en cascade. Dans la réalité on a rarement des paquets complètement disjoints et il faudra décider si la distribution décroissante des valeurs propres est associée à des effets de taille successifs ou non.

Nous arrêtons volontairement ici la présentation théorique de l'A.C.P., qui sera poursuivie dans l'analyse de la dimensionalité (2ème partie, chap.III) et dans le cas particulier de l'analyse de données homogènes correspondant à des champs spatiaux.

I.4 - Application aux données nivométéorologiques de Davos

L'application de l'A.C.P. au paquet de variables élaborées décrites dans la 1ère Partie (Chap. I.1.3) comportera 2 étapes : D'abord l'analyse de toutes les journées-observations, prises dans leur ensemble et l'interprétation des facteurs principaux. Puis l'utilisation de l'information complémentaire disponible : la journée est-elle avalancheuse ou non ?

Ces analyses ont été faites sur le premier ensemble de variables élaborées et ont en partie justifié sa modification.

I.4.1. Analyse de l'ensemble des journées - Interprétation des facteurs

(a) Il s'agit d'interpréter les résultats de l'analyse en C.P. de l'ensemble des individus des échantillons Mars-Avril et Janvier-Février. Nous sommes dans le cas d'un paquet de variables très mélangées et, sans résoudre complètement le problème de la dimensionalité (cf IIIème partie) il nous faut analyser et interpréter un certain nombre de facteurs.

Toutes les variables étant présumées explicatives, nous souhaiterions qu'elles soient toutes présentes, au moins en partie, dans les facteurs analysés ; les figures II-3 montrent le nombre n de variables qui nécessitent au moins k facteurs pour être représentées à α %.

Comme il est difficile d'analyser entre 15 et 20 facteurs, nous nous limiterons aux 5 premiers qui sont relativement interprétables.

On donne, dans les tableaux I et II , les coefficients de corrélations entre chaque variable et les plans factoriels F1/F2, F2/F3, F4/F5 et le sous-espace F1/2/3/4/5 et sur les figures II-4 les histogrammes des coefficients de corrélations multiples entre chaque variable et le sous-espace.

Enfin, on donne les corrélations de chaque variable avec les 16 premiers axes et, pour chaque axe, de 1 à 50, la corrélation maximum entre cet axe et une variable quelconque. (Tableau III)

Les conclusions sont les suivantes :

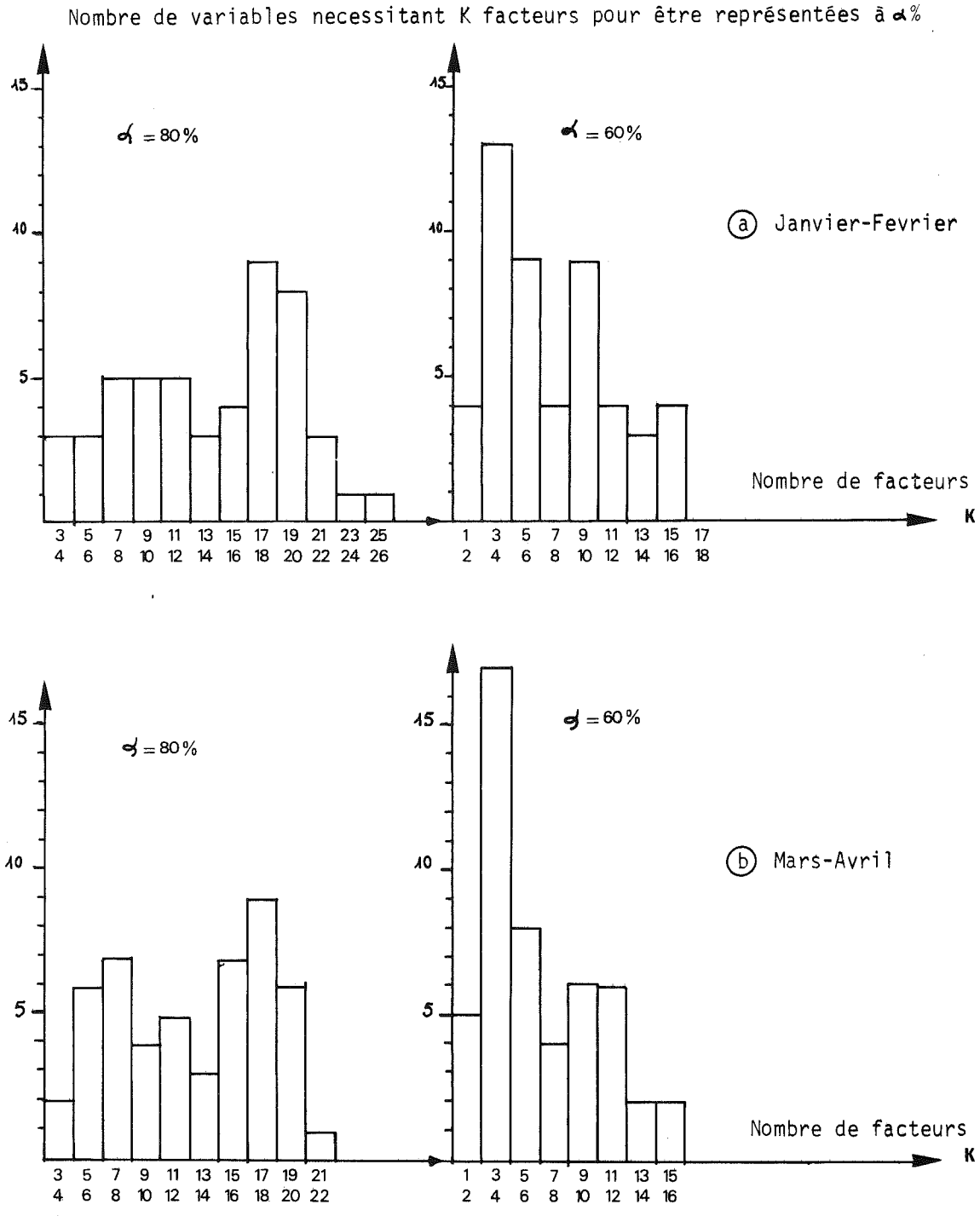
(1) Il faut environ une quinzaine de facteurs pour que chaque variable soit raisonnablement représentée dans l'analyse (cela sera repris ultérieurement)

(2) Il y a atomisation de certaines variables qui contribuent un peu à de nombreux facteurs.

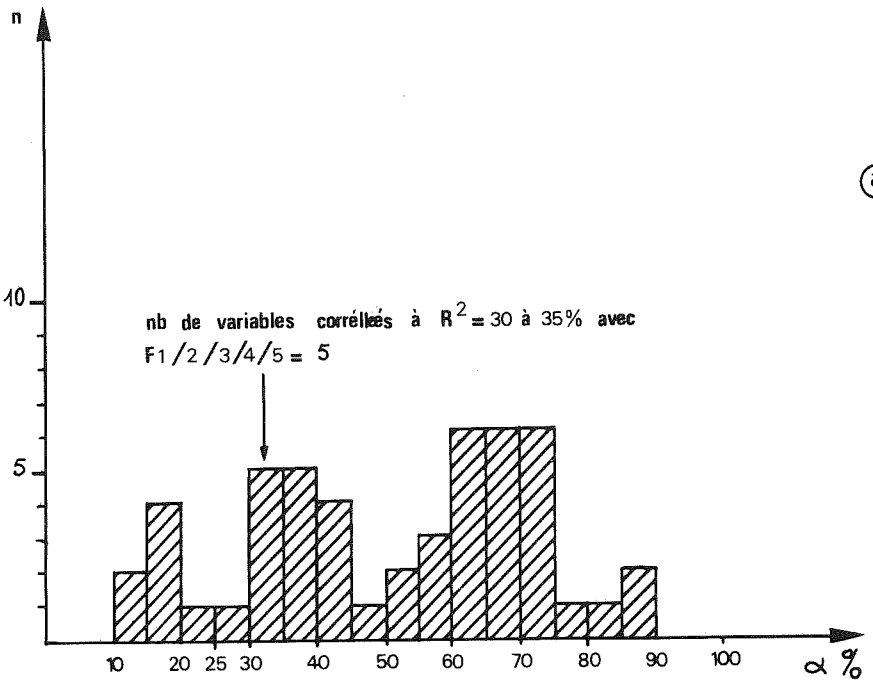
(3) Il y a des effets de taille qui sont parfois surprenants

Exemple . Sur le premier axe, on s'attendait à retrouver l'information précipitation, que l'on avait introduite sous de multiples formes. On la retrouve effectivement, mais les variables d'insolation, rayonnement, etc.. lui font un contrepoids tout aussi important.

Cet axe est donc seulement un axe beautemps sec ⟷ mauvais temps humide, mais il rassemble déjà 20% de la variance (tant en Janvier-Février qu'en Mars-Avril).

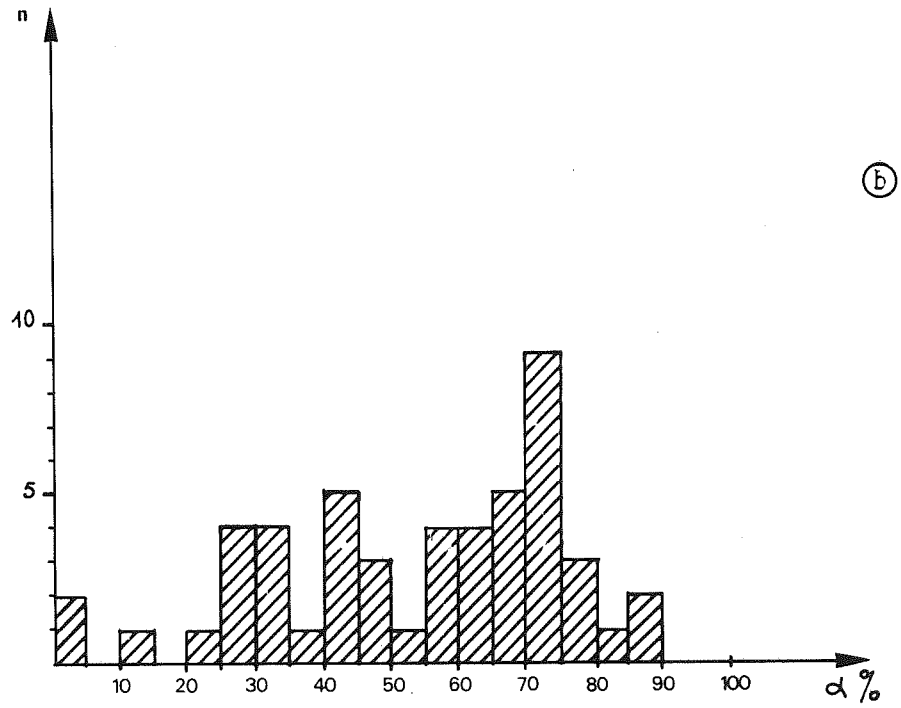


- Figure II-3 : Nombre de facteurs pour expliquer un pourcentage α donné de la variance des variables de départ (élaborées).



ⓐ Janvier - Fevrier

λ_1	9.53
λ_2	5.50
λ_3	4.78
λ_4	2.92
λ_5	2.72
Total	51 % de la variance



ⓑ Mars - Avril

λ_1	11.50
λ_2	4.79
λ_3	4.35
λ_4	3.34
λ_5	2.85
Total	54 % de la variance

- Figure II-4 - Représentation des variables dans les 5 premiers facteurs .

Numéro de la variable	F1/F2	F2/F3	F4/F5	F1-5
1	.576			.689
2	.609			.718
3				.386
4				.248
5	.625			.776
6	.388			.391
7				.135
8				.380
9	.471			.517
10				.324
11				.348
12	.424	.705		.889
13			.364	.479
14	.430	.676		.838
15				.367
16	.439	.778		.854
17			.390	.579
18	.627			.705
19	.549	.322		.678
20	.558			.667
21	.378	.276		.724
22	.622			.654
23		.281		.365
24	.505	.238		.638
25				.180
26				.104
27				.177
28	.406	.393	.167	.648
29	.368	.361	.162	.594
30	.362	.598	.001	.712
31	.442	.548	.017	.744
32		.239	.057	.431
33	.348		.082	.430
34			.157	.311
35				.329
36				.272
37				.186
38	.585			.615
39	.348		.205	.630
40			.171	.422
41			.211	.661
42	.442	.541		.702
43		.344	.303	.664
44		.256	.268	.533
45		.328	.291	.636
46			.231	.419
47		.315	.281	.603
48				.199
49		.268	.309	.597
50	.058		.260	.322
Seuil d'impression:	.350	.250	.150	

- Tableau I - Carré du coefficient de corrélation multiple entre la variable i et les plans factoriels F1/F2,F2/F3 , F4/F5 , puis le sous-espace F12345 .(Janvier - Fevrier).

Numéro de la variable	F1/F2	F2/F3	F4/F5	F 1,2,3,4,5
1	.550	.230		.725
2	.599			.753
3				.269
4				.279
5	.460			.766
6	.398			.487
7				.291
8				.305
9			.267	.603
10				.437
11	.317			.341
12	.721			.857
13			.273	.530
14	.564			.700
15	.529			.644
16	.695	.245		.855
17			.415	.598
18	.618		.191	.809
19	.387		.211	.736
20	.546	.269	.170	.721
21	.544			.716
22	.538	.330	.203	.742
23				.425
24		.414		.658
25				.036
26				.121
27				.208
28		.582		.701
29		.580		.635
30		.622		.740
31		.436		.699
32				.412
33	.381			.413
34				.301
35	.500			.606
36				.442
37			.186	.293
38	.411			.490
39	.391			.680
40	.581	.257		.702
41	.574	.163		.747
42	.611	.691		.782
43			.349	.456
44			.336	.398
45			.220	.312
46	.425	.548		.657
47			.450	.585
48				.049
49			.395	.578
50			.458	.578
Seuil d'impression:	.350	.250	.150	

- Tableau II -Carré du coefficient de corrélation multiple entre la variable i et les plans factoriels F1/F2,F2/F3 , F4/F5 , puis le sous-espace F12345 .(Mars - Avril) .

Numéro k ou j	Corrélation max de l'axe k	Corrélation de la var.j avec les 16 lers axes	k ou j	Corrélation max de l'axe k	Corrélation de la var. j avec les 16 lers axes
1	.780	.89	1	-.784	.87
2	.619	.92	2	-.727	.88
3	.647	.64	3	.610	.81
4	.516	.78	4	-.651	.76
5	.600	.86	5	-.655	.86
6	.617	.84	6	.518	.84
7	-.432	.76	7	.394	.75
8	-.637	.74	8	.607	.71
9	-.484	.78	9	.505	.81
10	.398	.69	10	-.401	.81
11	-.359	.61	11	.489	.69
12	-.502	.94	12	.584	.94
13	-.297	.73	13	.432	.80
14	.525	.93	14	.382	.88
15	-.362	.77	15	-.533	.78
16	-.358	.90	16	.467	.92
17	-.474	.85	17	.328	.85
18	.324	.91	18	-.336	.91
19	-.325	.91	19	-.259	.90
20	-.308	.85	20	-.334	.85
21	-.359	.80	21	.273	.79
22	.402	.88	22	-.315	.89
23	.381	.82	23	-.255	.69
24	-.282	.74	24	-.278	.74
25	.246	.79	25	.216	.86
26	.245	.74	26	.288	.82
27	-.250	.77	27	.197	.93
28	-.314	.92	28	.257	.91
29	.217	.90	29	-.201	.92
30	-.227	.88	30	.199	.89
31	.180	.87	31	-.290	.85
32	.199	.80	32	.171	.81
33	.170	.76	33	.185	.79
34	.210	.70	34	-.136	.76
35	-.182	.73	35	.132	.84
36	-.200	.79	36	.197	.75
37	-.131	.77	37	.100	.77
38	-.154	.84	38	-.145	.79
39	-.180	.89	39	.111	.87
40	.117	.79	40	.183	.86
41	-.155	.86	41	.123	.87
42	.147	.92	42	.206	.91
43	.137	.92	43	-.120	.85
44	.099	.67	44	.100	.66
45	.097	.90	45	.135	.67
46	-.086	.90	46	.114	.90
47	-.087	.70	47	.083	.70
48	.080	.67	48	-.070	.81
49	.038	.94	49	.064	.93
50	.000	.89	50	.000	.94

- Tableau III - Correlation maximale entre chaque axe et une variable quelconque, et correlation de chaque variable avec l'ensemble des 16 premiers axes .(exemple de Janvier-Fevrier)

Exemple. Les paramètres thermiques (air et neige réunis) sont pratiquement indépendants du premier axe (beau-temps / temps perturbé) en Janvier-Février.

On peut vérifier sur les corrélations :

		Janv.-Fév.	Mars-Avril
Température à 13H	nombre d'heures d'ensoleillement	.159	.345
	rayonnement incident	.080	.374

Par contre, en Mars-Avril, une dépendance apparaît : le beau temps est associé à de fortes valeurs des paramètres thermiques de l'air, mais pas de la neige ! Or il est peu probable que les nombreuses opinions émises sur l'origine des avalanches tiennent compte de ces intercorrélations, d'où leurs caractères parfois contradictoires.

4 Les facteurs ne se recouvrent pas exactement d'un bimestre à l'autre, ce qui traduit des effets saisonniers dans les intercorrélations.

5 Les facteurs d'ordre élevé présentent encore des corrélations significatives avec certaines variables (cf ρ_{jk}^{max} dans le Tableau III) même si globalement, celles-ci vont en décroissant.

Remarque. L'interprétation des facteurs sera résumée aussi dans les figures II-5 et II-6.

ⓑ En ce qui concerne les distributions des facteurs, on constate que les combinaisons linéaires de variables dissymétriques, même assez corrélées entre elles, sont sensiblement plus symétriques que les variables initiales.

Par exemple, l'effet de la valeur - seuil de 0.0 mm pour l'ensemble des variables "précipitations" est pondéré par la variabilité, sur ce même facteur, des variables "insolation" les jours où il n'y a pas précipitation, etc...

On dispose donc de variables sensiblement plus "gaussiennes", que l'on envisage d'utiliser dans les modèles décisionnels ultérieurs, compte tenu de leur orthogonalité.

Mais si l'on a annulé les intercorrélations il reste l'autocorrélation pouvant apparaître au sein de chaque composante. Car il ne s'agit pas de journées prises au hasard dans une population infinie, mais de journées successives, mises bout à bout par paquets de 2 mois.

On aurait pu effectuer l'analyse de l'autocorrélation sur les variables elles-mêmes, mais compte tenu de leurs distributions (certaines restent constantes pendant de longues périodes, etc...) le calcul sur les C.P. en donne une idée un peu "lissée".

Le tableau IV donne pour les 10 premières composantes de chaque mois, les coefficients d'autocorrélation calculés sur l'ensemble de l'échantillon (12 bimestres bout à bout, et les valeurs maximales et minimales du calcul bimestre par bimestre).

On constate qu'elles sont très variables d'une C.P. à l'autre et que, compte tenu des tailles d'échantillons, elles sont souvent très significatives. Toutefois, le calcul sur l'ensemble des échantillons est biaisé par les écarts de moyenne entre bimestres des années successives, ce qui augmente la corrélation (cf C.P. n°5 de Mars-Avril : r_1 moyen $> r_1$ max).

On vérifie dans le calcul par bimestre, que les corrélations au retard $k > 5$ sont généralement faibles, sauf cas pathologiques.

Mars-Avril

Composante principale n°	r_1	r_2	r_5	r_{1max}	r_{1min}	r_{5max}
1	.80	.57	.27	.86	.60	.60
2	.68	.43	.39	.73	.36	.59
3	.90	.78	.57	.97	.69	.88
4	.32	-.04	.12	.41	.17	.21
5	.70	.53	.52	.60	.21	.30
6	.73	.60	.49			
7	.44	.27	.17			
8	.72	.54	.34			
9	.59	.37	.19			
10	.54	.42	.27			

Janvier-Février

1	.82	.62	.27	.87	.53	.43
2	.70	.47	.33	.80	.44	.50
3	.76	.57	.44	.80	.33	.60
4	.68	.41	.24	.74	.42	.28
5	.49	.24	.29	.61	.12	.35
6	.74	.58	.40			
7	.53	.38	.30			
8	.79	.70	.50			
9	.63	.46	.21			
10	.51	.31	.22			

TABLEAU IV

Autocorrélation r_k avec un retard k des 10 premières variables orthogonales.

I.4.2. Introduction d'information exogène

On met en regard de nos 50 variables prises sous forme orthogonalisée, des informations exogènes qui seront, initialement, l'occurence ou la non-occurence d'avalanche le jour i considéré, et éventuellement d'autres informations complémentaires (cf Ière Partie, Chap.I-3).

Nous avons d'abord représenté systématiquement les observations dans les plans F1/F2, F2/F3 et F4/F5 en séparant journées avalancheuses et journées "normales". Nous les commentons succinctement :

ⓐ Janvier-Février (Fig.II-5)

- Dans F1/F2, nuage des points sans avalanche encore assez dissymétrique (à cause de la présence sur l'axe 1 de variables radiatives bornées qui interviennent négativement et s'opposent à des variables de précipitations qui n'ont pas de limite supérieure).
- Journées normales denses au-dessus de la 1ère bissectrice et assez rares dans le 2ème quadrant.
  Journées avalancheuses sous la 1ère bissectrice (les 2/3 environ) et en particulier dans le 2ème quadrant (environ 50 sur 122 journées avalancheuses) alors que les journées normales y sont rares (environ 70 sur 589).
- Ce 2ème quadrant correspond approximativement à des journées où il y a de la neige récente au sol, et un temps éventuellement beau mais froid et venteux. Pour beaucoup de ces journées avalancheuses, du chasse-neige a été observé le jour ou la veille.
- Par contre les avalanches de code 1, qui ne représentent que 7% du total, sont toutes au-dessus de l'axe F1. Elles correspondent à des déclenchements d'allure ponctuelle et associées à des valeurs élevées de F2, elles méritent le nom de "coulées". Les 2 seules avalanches cotées 5 (neige humide) se trouvent bien à la périphérie supérieure du nuage des points avalancheux.
- Dans F2/F3, on aboutit à des conclusions voisines sur le rôle du temps froid et venteux combiné aux fortes précipitations.

ⓑ Mars-Avril (Fig.II-6)

- Nuages assez elliptiques (tant pour les journées normales qu'avalancheuses) mais qui se superposent quasiment. Différences de densités difficiles à apprécier dans ce plan F1/F2.
- Par contre, le nuage des avalanches est bien séparé en avalanches humides (code 5) et sèches (code 4) de part et d'autre de l'axe 2. Cette partition recoupe presque le fait que le déclenchement a eu lieu en dehors d'une période de précipitation ou au contraire pendant (ou peu après).

- Les avalanches codées 1 (coulée à départ ponctuel) se rassemblent plutôt au-dessus de la 1ère bissectrice, et sont souvent codées humides (5). Elles représentent cette fois 25% des cas.
- Le plan F2/F3 non représenté ici, est plus intéressant qu'en Janvier-Février et une certaine discrimination apparait.
  Par contre, pour les 2 bimestres, le plan F4/F5 ne semble guère discriminant.
  nant .

Une autre méthode a consisté à visualiser les journées, normales et avalancheuses, ne presentant que certaines caractérisques, donc appartenant à un certain type de temps:

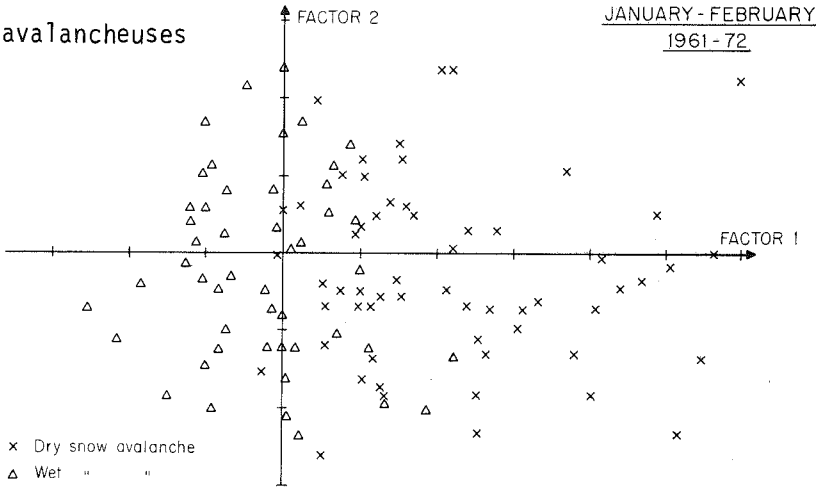
ⓒ Janvier-Février

- Journées avec chute de neige (> 5 cm) : Elles sont pratiquement toutes à droite de l'axe F2. Le 2ème quadrant rassemble alors 23 cas avalancheux (sur 38 au total) pour seulement 35 cas normaux. Cela se vérifie aussi dans F2/F3 où le 4ème quadrant rassemble 12 avalanches pour 19 journées normales. Cela confirme le rôle du vent et du temps froid.
- Journées sans précipitation, mais avec neige récente au sol. Ce cas ne diffère pas beaucoup du précédent et aboutit aux mêmes conclusions. Outre le rôle du vent sur la neige présente au sol, on perçoit aussi l'influence des variations de température et surtout le refroidissement de la neige.
- Journées sans précipitation récente : Seule 12 journées avalancheuses appartiennent à ce cas, mais elles se distinguent peu des journées normales sauf par des valeurs négatives de F2.

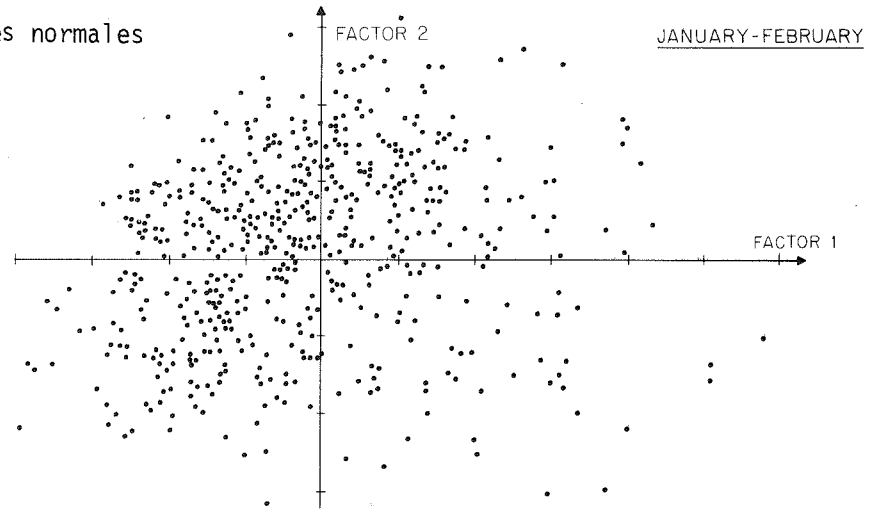
ⓓ Mars-Avril

- Journées avec précipitations > 3 cm :
  Le nuage des avalanches correspond à des valeurs positives de F1 mais se répartit de part et d'autre. C'est donc seulement la valeur absolue de la précipitation qui intervient. Dans F2/F3 on constate même que ces avalanches sont plus fréquentes par temps chaud avec beaucoup de précipitation.
- Journées consécutives à des précipitations:
  Elles sont relativement peu nombreuses dans l'ensemble (2 après chaque séquence, soit environ 125 journées sur 700). Mais elles donnent globalement beaucoup d'avalanches (40 environ sur les 122 du bimestre). Les avalanches se concentrent dans les 2ème et 3ème quadrants (beau temps froid et clair, refroidissement de la neige) alors que le beau temps chaud en occasionne peu. Le rôle du transport par le vent est encore très marqué.
- Journées sans précipitations récentes :
  Elles sont globalement assez rares (un peu plus d'une centaine) et donnent lieu à 27 avalanches. Mais uns distinction apparait, fonction des températures de l'air et de la neige, selon qu'elles correspondent à un beau temps froid ou au contraire plutôt chaud, avec fonte.

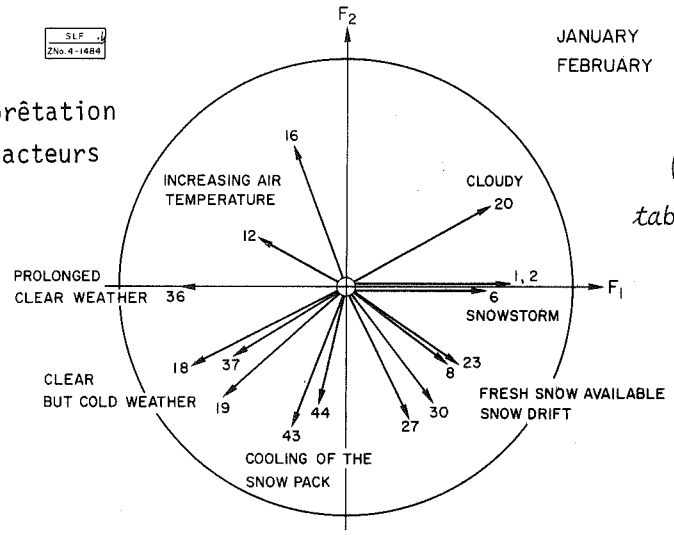
ⓐ Journées avalancheuses



ⓑ Journées normales

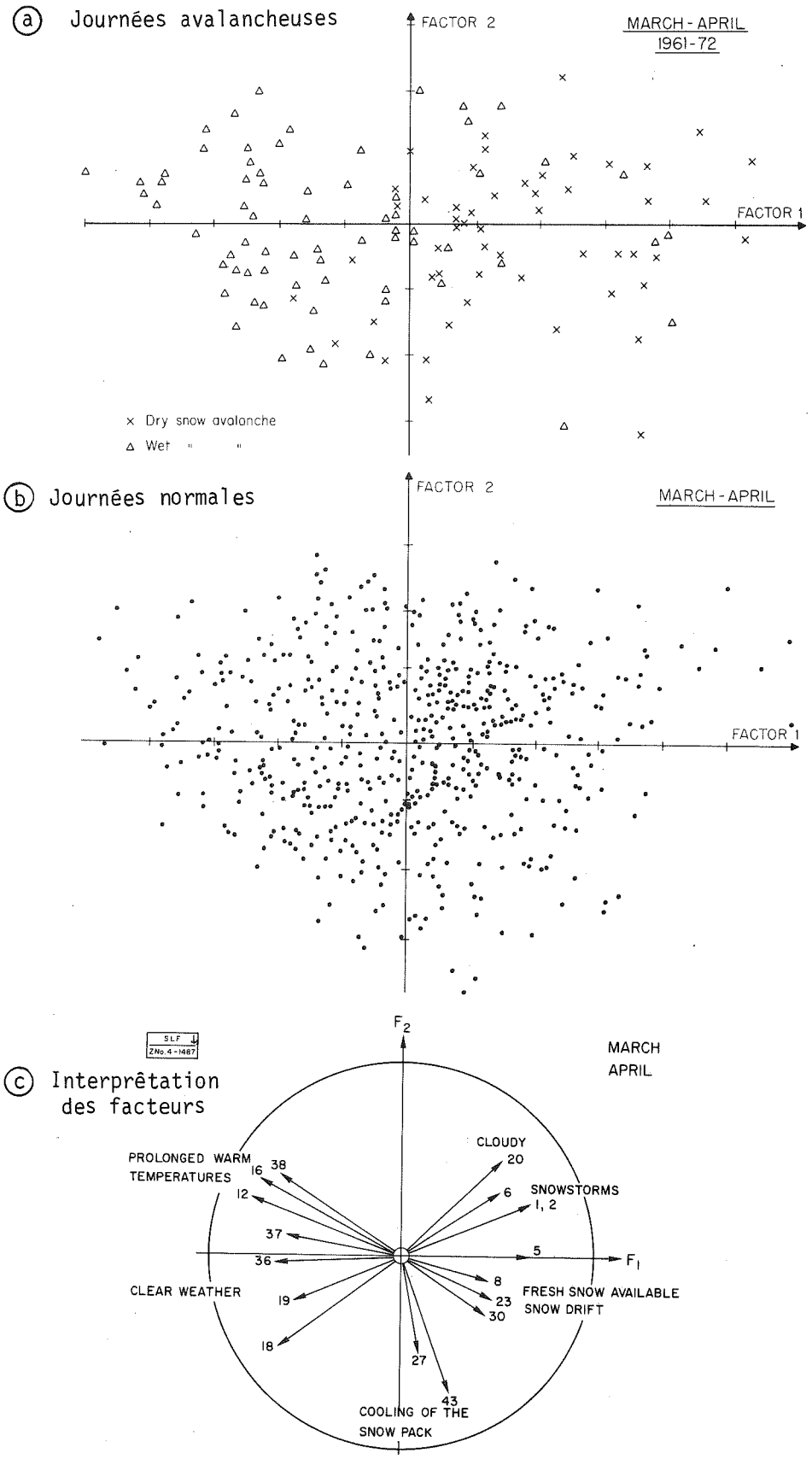


ⓒ Interprêtation des facteurs



(Les N° correspondent à la table II de la 1ère Partie- p.11)

- FIGURE II-5 -Projections des observations de Janvier - Fevrier .



- FIGURE II-6 -Projections des observations de Mars - Avril .

I.4.3. Analyse des trajectoires

Compte tenu de l'inertie de certains facteurs (autocorrélation) et aussi de l'incertitude qui pèse sur la date de l'avalanche, il était intéressant de considérer aussi les quelques journées précédant immédiatement celle où l'avalanche a été signalée.

On peut aussi admettre que l'avalanche n'est pas un processus instantané, lié au niveau absolu des variables le jour j , mais résulte aussi d'une évolution, où ce sont les variations de niveau qui interviennent.

En pratique, on a matérialisé les trajectoires en joignant le jour avalancheux aux 3 qui l'ont précédé. Les résultats sont assez marquants (cf. Figure II-8) :

(a) En Janvier-Février

On trouve dans le plan F1/F2 des trajectoires-types, très stables, parcourant dans le sens des aiguilles d'une montre les quadrants IV → I → II → III . Les premières avalanches correspondent à l'apparition de précipitations, souvent concomittantes à un réchauffement (de III ou IV → I). Ce temps est souvent suivi d'un refroidissement marqué et de vent fort, très propice aux avalanches (de I → II → III) précédant le retour d'un beau temps froid (III). La simple remontée des températures ne donne alors lieu qu'à de rares avalanches (III → IV).

Le plan F2/F3 complète ce résultat en insistant sur le rôle du chasse-neige. En dehors des périodes de précipitations, les cycles de températures : chaud → froid → chaud semblent critiques.

En conclusion, ces trajectoires coïncident assez bien avec les circulations météorologiques hivernales caractérisées par une succession de perturbations d'Ouest d'une durée de 3 à 4 jours, et associées à des masses d'air chaudes, en alternance avec des situations anticycloniques froides.

Les situations les plus critiques correspondent soit aux fortes chutes de neige (souvent associées à du vent) soit au cas de chutes de neige moyennes associées à des vents forts et à un refroidissement marqué (de la neige et pas forcément de l'air, du moins en ce qui concerne les températures maximales).

(b) En Mars-Avril

On retrouve dans ce plan des trajectoires analogues au cas de Janvier-Février (1), mais avec une tendance marquée en 2 grands types (2) et (3), qui recoupe en général les 2 types d'avalanches signalées (neige sèche ou humide).

Les avalanches survenant pendant des précipitations se caractérisent par une avancée sur l'axe F1, due aux précipitations mais avec des fluctuations sur F2 indiquant qu'elles peuvent en cette saison se traduire par un refroidissement ou un réchauffement. Par contre, les avalanches qui apparaissent après les précipitations ont toutes une trajectoire très stable parallèle à la 1ère bissectrice (I → III) qui correspond à un refroidissement par retour du beau temps clair, avec souvent du chasse-neige.

Enfin, les avalanches qui apparaissent en dehors des précipitations ont des trajectoires orientées négativement sur F1, ce qui correspond à des effets thermiques ou

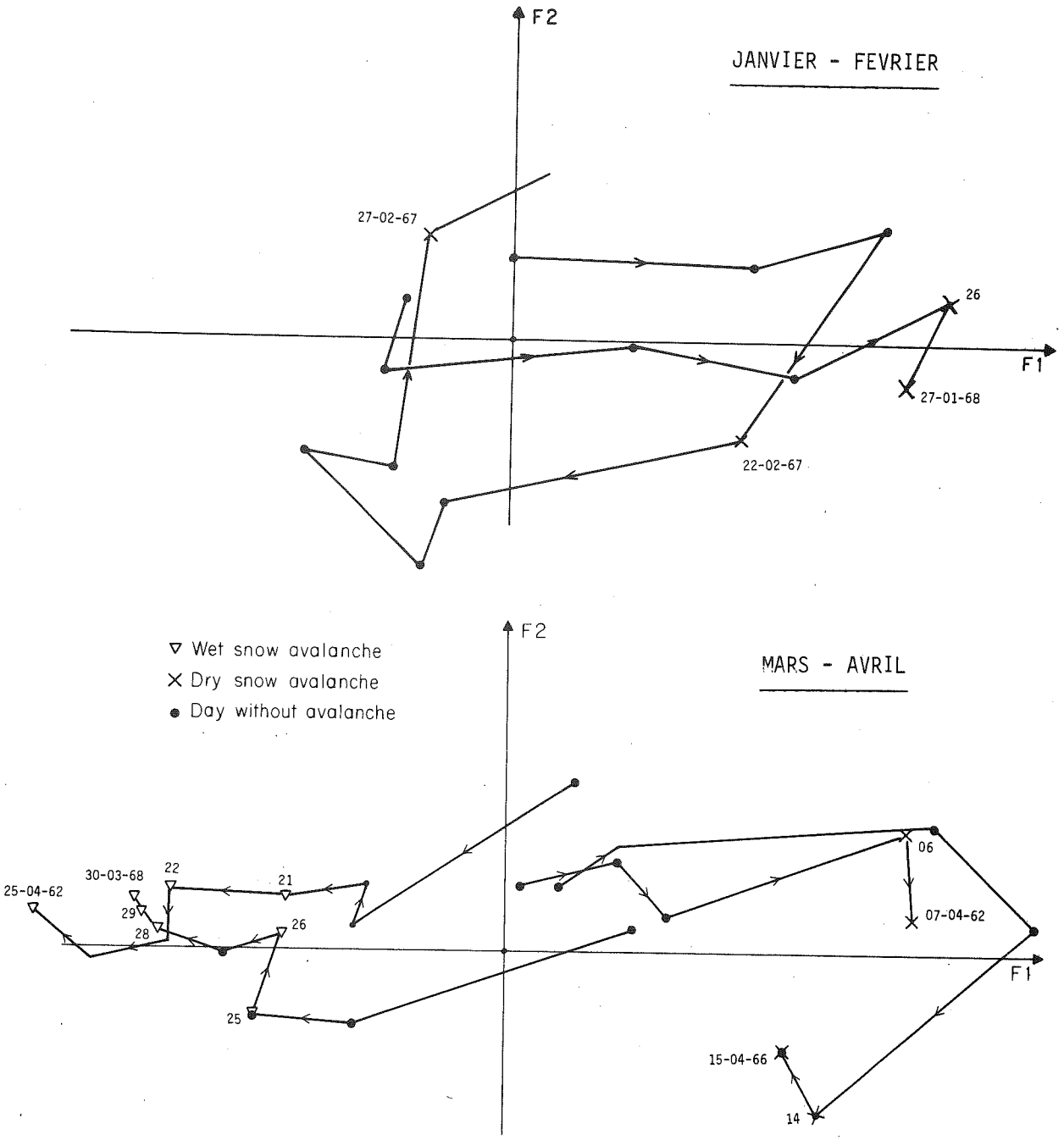


FIGURE II - 7 : Trajectoires des séquences ayant abouti à des journées avalancheuses .

mécaniques. Elles peuvent apparaître soit par températures négatives, soit plus souvent par températures positives, mais il est fréquent que la trajectoire ait peu varié au cours des jours précédant l'avalanche, ce qui traduit un effet d'accumulation plutôt que d'évolution.

On constate là aussi une plus grande diversité dans les situations propices aux déclenchements, cohérente avec la diversité des situations météorologiques qui peuvent aller du beau temps froid hivernal, semblable à celui du bimestre précédent, jusqu'à des pluies orageuses de caractère estival.

Pour tâcher de mettre ceci en évidence de façon plus quantifiée, nous avons utilisé le schéma suivant :

1) diviser le plan F1/F2 en 8 secteurs(et une zone centrale incertaine)
2) construire les matrices de passage d'un secteur à l'autre, d'une part pour les points non avalancheux, et d'autre part pour les points avalancheux, en ajoutant +1 dans la case (r,s) si le jour j était dans la secteur s sachant que j-2 était dans le secteur r.

La part du phénomène expliquée par la situation est appréciée par la valeur des termes diagonaux, tandis que la part d'évolution peut se mesurer à la valeur des termes non diagonaux.

On constate et on peut le vérifier à l'aide d'un test assez complexe (cf. GORDON HILTON , 1972) que ces matrices de passage diffèrent sensiblement pour les cas avalancheux et non avalancheux (

Si l'on se place, à titre d'exemple, dans les axes F1/F2 de Mars-Avril qui peuvent s'interpréter à l'aide du tableau V , on constate :

. au vu des fréquences marginales le jour j , que les journées avalancheuses se concentrent dans les secteurs 3, 4 et 7,8 alors que les journées normales sont mieux réparties entre les différents secteurs. En particulier le secteur 9 s'avère peu avalancheux.

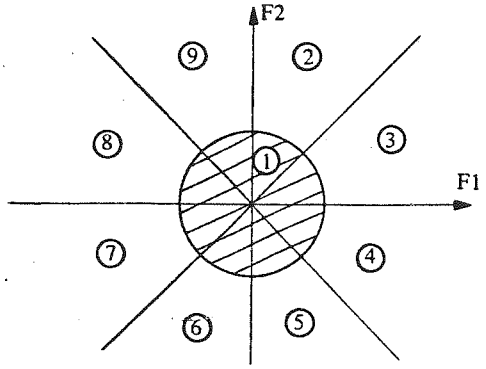
. que les journées avalancheuses du secteur :

3 proviennent de 2 et surtout 3
4 proviennent de 4 et surtout 3
5 proviennent de 4 et surtout 3.

Ce secteur 3 étant celui des précipitations récentes, cela veut dire qu'un certain nombre d'avalanches nécessite d'abord des précipitations, sur lesquelles un autre effet (refroidissement, vent, réchauffement) s'exercera.

. que les journées avalancheuses des secteurs 7 et 8 se caractérisent par un phénomène d'"accumulation" puisque la plupart étaient déjà 7 ou 8 l'avant-veille de l'avalanche.

. que, tant pour les journées avalancheuses que pour les autres, il y a plus de points au-dessus et à droite de la diagonale qu'en dessous. Ceci met en évidence une caractéristique climatologique intéressante. Les jours successifs ont tendance à suivre une trajectoire dominante dans le sens trigono-



Découpage du plan factoriel en
8 secteurs plus une zone incertaine.

Secteur auquel appartient la journée avalancheuse j

Avant-veille d'une journée avalancheuse	2	3	4	5	6	7	8	9	Total	%
2	1	2	0	0	0	1	1	0	5	5.5
3	1	8	9	4	2	1	0	0	25	27.5
4	0	1	3	2	3	1	0	0	10	11.0
5	0	3	2	0	0	1	0	0	6	6.6
6	1	3	0	0	2	4	1	0	11	12.1
7	2	0	0	0	0	10	6	1	19	20.9
8	0	0	0	0	0	1	12	0	13	14,3
9	0	0	0	0	0	0	2	0	2	2.2
Total	5	17	14	6	7	19	22	1	91	
%	5.5	18.7	15.4	6.6	7.7	20.9	24.2	1.1		100.0

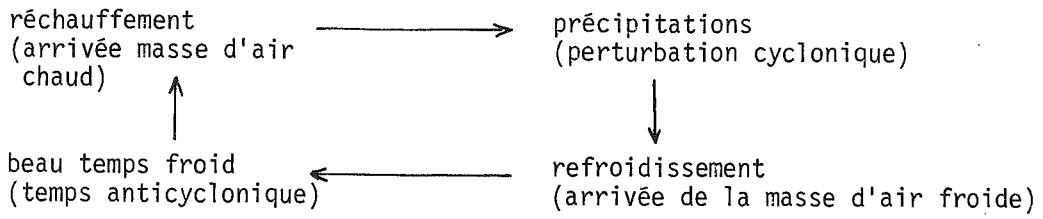
Secteur auquel appartient la journée quelconque j

Avant-veille d'une journée quelconque	2	3	4	5	6	7	8	9	Total	%
2	6	7	3	1	0	4	7	6	34	7.6
3	6	25	24	8	12	8	2	1	86	19.1
4	1	9	17	25	7	2	1	0	62	13.8
5	1	9	8	8	12	4	1	0	43	9.6
6	5	6	2	2	11	9	1	0	36	8.0
7	6	5	0	0	3	29	18	8	69	15.3
8	12	4	0	0	1	7	39	20	83	18.4
9	7	9	0	0	0	5	8	8	37	82
Total	44	74	54	44	46	68	77	43	450	
%	9.8	16.4	12.0	9.8	10.2	15.1	17.1	9.6		100.0

Note – On a éliminé les points du domaine (1) pour lesquels l'appartenance aux quadrants n'est pas significative.

- TABLEAU V -Matrices de passage pour les journées avalancheuses et pour l'ensemble.

- métrique qui peut se résumer en :



I.4.4. Conclusions préliminaires

Nous n'avons pas exploité plus avant cet aspect "dynamique" du phénomène avalancheux, considéré comme le résultat d'une évolution ou d'une séquence de situations assez particulière. Toutefois, le suivi, au jour le jour, de la position du point courant dans le plan F1-F2 s'est révélé très instructif.

Ces analyses nous ont de plus permis de constater que :

- Parmi les causes possibles de déclenchement, on peut distinguer celles qui correspondent à un effet d'accumulation (réchauffement et températures positives prolongées, précipitations ininterrompues)tandis qu'une autre famille rassemble les changements brusques (réchauffement brutal, arrêt des précipitation suivi d'un vent fort et d'une baisse des températures, etc ...).
- Malheureusement, il est difficile de trouver un plan principal où les situations avalancheuses se distinguent bien des situations normales,(ce n'est d'ailleurs pas le but de la méthode), et pour une situation avalancheuse donnée , on peut trouver des situations voisines non avalancheuses .
- Par contre, la densité de probabilité du phénomène avalancheux varie sensiblement selon les régions du plan F1-F2 par exemple, ce qui incite à penser qu'il y a plusieurs phénomènes distincts.

Ces conclusions nous conduiront d'une part à procéder à une classification des avalanches,d'autre part à développer des modèles par voisinage (cf. Vème Partie).

CHAPITRE II

ANALYSE FACTORIELLE DES CORRESPONDANCES (AFC)

Cette méthode, désormais classique, au moins en France, était initialement destinée au traitement des tableaux de contingence. Son intérêt en hydrométéorologie était alors limité et son utilisation hasardeuse. Ph.BOIS (1976) en donne quelques exemples d'application à des données brutes où déjà apparaît un besoin de recoder les données. Mais c'est avec le développement du codage disjonctif que cette méthode peut prendre, en climatologie, tout son intérêt.

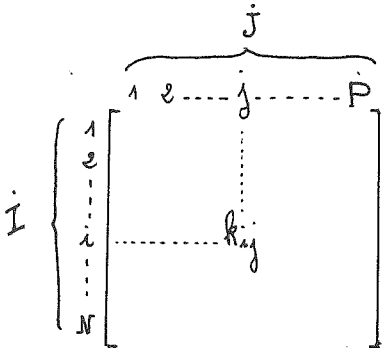
Là encore, nous donnons d'abord une présentation générale (II.1) qui peut être ignorée par le lecteur averti mais qui est nécessaire pour voir ce qu'apporte les paragraphes suivants.

C'est l'article de J.P. NAKACHE (1973) qui a éveillé notre intérêt pour cette méthode et nous a conduit à l'analyser plus finement dans les cas particuliers que nous rencontrons (FRITSCH et OBLED, 1974). Ils ont été repris depuis, en particulier par P. CAZES (1976) et sont désormais classiques.

II.1 - Présentation générale et notations

II.1.1. Notations

On considère des tableaux de fréquences croisant 2 ensembles de modalités I et J, où k_{ij} représente le nombre de cas où les modalités i et j étaient satisfaites conjointement.



On définit successivement :

$K = \sum_{i,j} k_{ij}$ nombre total d'évènements observés.

$P_{ij} = k_{ij}/K$ la fréquence du couple (i,j)

$P_{i.} = \sum_{j} P_{ij}$ la fréquence de la modalité i

$P_{.j} = \sum_{i} P_{ij}$ la fréquence de la modalité j

Si les 2 ensembles de modalités I et J étaient indépendants, on devrait avoir : $P_{ij} = P_{i.} \times P_{.j}$ et la quantité :

$$\chi^2 = \sum_{i,j} \frac{(P_{ij} - P_{i.} \times P_{.j})^2}{P_{i.} \times P_{.j}}$$

est classiquement utilisée pour mesurer la liaison entre les 2 ensembles de modalités I et J d'un tableau de contingence.

Notre but toutefois n'est pas de mesurer globalement la liaison entre (I) et (J) mais entre les éléments respectifs de I ou de J.

On peut donc encore considérer chaque point i comme un point de $\mathbb{R}^p$ caractérisé par p caractères $(P_{i1}, P_{i2}, \ldots, P_{ij}, \ldots P_{ip})$.

De même on peut chercher à représenter les caractères j dans $\mathbb{R}^N$ comme des points $(P_{1j}, P_{2j}, \ldots P_{ij} \ldots P_{Nj})$.

II.1.2. Représentations des points dans $\mathbb{R}^p$ et $\mathbb{R}^N$: (*)

On se place d'abord dans $\mathbb{R}^p$ où chaque "individu" i est représenté par p caractères. On pourrait considérer que i et i' sont identiques si : $P_{ij} = P_{i'j} / j = 1 \ldots p$ (j = 1...p). En fait si l'on remonte au tableau initial des k_{ij}, les 2 ensembles :

$$i \quad k_{i1}\ k_{i2} \ldots k_{ij} \ldots k_{ip}$$
$$i' \quad k_{i'1}\ k_{i'2} \ldots k_{i'j} \ldots k_{i'p}$$

correspondent aux histogrammes des caractères j, pour les 2 modalités ou individus i et i'. Or la comparaison des 2 histogrammes ne peut se faire qu'en se ramenant au même nombre d'individus. On divise donc :

$$k_{ij} \text{ par } \sum_j k_{ij}$$

ou, si l'on a déjà tout divisé par K, on divise

$$P_{ij} \text{ par } \sum_j P_{ij} = P_{i\cdot} \text{ et } P_{i'j} \text{ par } P_{i'\cdot}$$

ce qui consiste en quelque sorte à normer les individus.

On passe en fait de l'histogramme de i à la densité de probabilité empirique de i puisque :

$$\sum_j \frac{P_{ij}}{P_{i\cdot}} = 1 \quad \forall i$$

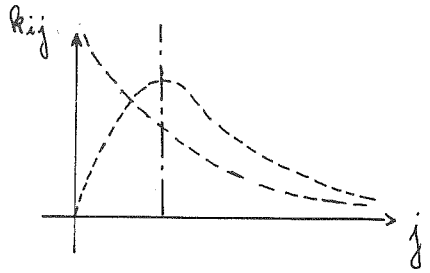
On considère donc maintenant $\mathbb{R}^p$ comme l'espace des $(i)\left\{\frac{P_{ij}}{P_{i\cdot}}, j = 1, \ldots p\right\}$

Donc 2 points i et i' confondus $\Longrightarrow$ ils ont des profils identiques (c'est-à-dire des histogrammes de départ semblables ou homothétiques). On pourrait alors mesurer leur distance par :

$$D^2(i,i') = \sum_j \left(\frac{P_{ij}}{P_{i\cdot}} - \frac{P_{i'j}}{P_{i'\cdot}}\right)^2 = \sum_j d_j^2(i,i')$$

Mais on rencontre alors le problème évoqué précédemment quand les termes de la sommation ne sont pas comparables. Dans notre cas par exemple, un écart d_j est d'autant plus significatif que le caractère considéré a globalement un poids faible sur l'ensemble des (i). Cela revient à donner d'autant plus de poids à un écart sur le caractère j que la "moyenne" (ou la fréquence , ou la variation ...) du caractère est faible .

Remarque : Si l'on se rappelle que k_{ij} est un effectif, donc $\geqslant 0$, de même que P_{ij}, la distribution de P_{ij} sur les individus i est dissymétrique et la distance de ...



0 à la moyenne peut être considérée comme un paramètre de variation.
Donc, on prend pour distance :

$$D^2(i,i') = \sum_j \frac{1}{P_{\cdot j}} \left(\frac{P_{ij}}{P_{i\cdot}} - \frac{P_{i'j}}{P_{i'\cdot}} \right)^2$$

ce qui revient implicitement à "normer" les caractères .

Enfin, si on construit $\mathbb{R}^P$ comme l'espace des points (i):

$$(i) \rightarrow \left\{ x_{ij} = \frac{P_{ij}}{P_{i\cdot}\sqrt{P_{\cdot j}}},\ j = 1 \dots P \right\}$$

alors la distance ci-dessus devient :

$$D^2(i,i') = \sum_j \left(x_{ij} - x_{i'j} \right)^2$$

c'est-à-dire la distance euclidienne usuelle.

Remarques: La distance

$$D^2(i,i') = \sum_j \frac{1}{P_{\cdot j}} \left(\frac{P_{ij}}{P_{i\cdot}} - \frac{P_{i'j}}{P_{i'\cdot}} \right)^2$$

est souvent appelée distance du χ^2, mais ne constitue pas à proprement parler une variable χ^2 classique, et diffère par exemple de celle que l'on calculerait sur le tableau de contingence croisant le couple (i,i') avec l'ensemble des modalités J.

A la différence de l'A.C.P., les points i et i' ont désormais des masses $P_{i\cdot}$ et $P_{i'\cdot}$, et l'analyse du nuage se fera ici au sens de l'inertie, c'est-à-dire en cherchant les axes du nuage des points pesants. Cette interprétation était déjà valable en A.C.P., mais se ramenait alors à la seule analyse des interdistances, ce qui n'est plus le cas ici.

On sait que cette analyse se ramène à la détermination des éléments propres d'une matrice $S = X^t . X$ avec $X = \{x_{ij}\}$ dim N x P
dont on trouvera le détail en annexe ou dans LEBART et FENELON. (1973).

L'analyse du nuage des variables j (ou modalités de J) s'en déduit ensuite très simplement et à la différence de l'A.C.P., l'origine ne joue plus de rôle particulier. Il y a identité complète entre les traitements dans $\mathbb{R}^P$ et $\mathbb{R}^N$, ce qui est normal, dans l'optique des tableaux de contingence, vu la symétrique des rôles des 2 ensembles I et J.

Ces brefs rappels exposent l'analyse des correspondances appliquée initialement aux seuls tableaux de contingence, peu courants en hydrologie.

Pourtant son application à des tableaux de nombres positifs pouvait donner des résultats assez intéressants, à condition de garder à l'esprit le rôle particulier joué par les marges du tableau, voire de le contrôler (cf Ph. BOIS, 1976 et FRITSCH et OBLED, 1974). Mais cette utilisation se limitait à des cas assez particuliers (réseaux de mesures).

Le traitement de tableaux de données plus générales nécessite un recodage plus énergique afin de se ramener à un tableau présentant pratiquement les propriétés d'un tableau de contingence.

II.2 - Codage disjonctif

Suggéré par J.P. NAKACHE (1973) nous allons le décrire rapidement, ainsi que quelques propriétés particulières que nous avons démontrées et qui sont utiles pour l'interprétation des résultats (FRITSCH et OBLED, 1974).

II.2.1. Buts et mise en oeuvre

Bien que les tableaux de mesures quelconques soient plutôt du ressort de l'A.C.P., celle-ci fait jouer un rôle particulier à la moyenne et à l'écart-type des variables, ce qui suppose implicitement des distributions plutôt symétriques. D'autre part, les variables peuvent être continues ou discrètes : la variable X_j peut valoir 1, 2, 3 , ou même 1 ou 0 si elle est dichotomique.

Dans ce cas, la "moyenne" et l'"écart-type", au sens de l'analyse en composantes principales, n'ont guère de sens, et la présence de 1 ou de 0, interprété en terme d'individu présent ou absent, nous amène plutôt à l'analyse en correspondance.

D'autre part, dans l'A.F.C. d'un tableau de contingence, la symétrie des rôles des ensembles I et J , qui n'a pas de raison d'être pour un tableau individus x variables, intervient surtout par l'utilisation que l'on fait des marges du tableau. Or si un codage adéquat permet d'éliminer l'effet des marges, par exemple en les rendant toutes identiques, alors l'A.F.C. peut s'appliquer sans effets parasites.

En pratique nous allons procéder en 2 étapes :

1) Classer les variables

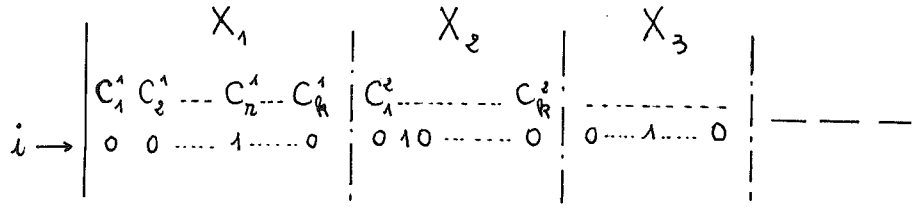
On passe d'une variable quantitative X à une partition $\mathcal{C}$ en k classes $\{\mathcal{C}_1, \mathcal{C}_2, \ldots \mathcal{C}_r, \ldots \mathcal{C}_k\}$ de fréquences respectives f_r $r = 1 \ldots k$

Utilisant le concept d'information de SHANNON, NAKACHE (1973) montre que la transformation d'un paramètre quantitatif en un caractère qualitatif provoque une perte d'information et que celle-ci est minimale quand les différentes modalités sont équiprobables, et quand le nombre de modalités est maximal.

Nous allons en général tâcher de respecter cette contrainte. Toutefois, sachant que l'individu i appartient à la classe $\mathcal{C}_r$, le problème est de coder son appartenance à $\mathcal{C}_r$.

2) Codage disjonctif

Plutôt que de lui affecter la valeur r , on va créer k variables $C_1, C_2, \ldots C_k$ associée à X et prenant la valeur 0 ou 1 selon que i appartient ou non à $\mathcal{C}_r$. Si maintenant on a p variables de départ, et qu'on les code toutes en k classes, on aura pxk variables et l'individu i aura pour coordonnées :



L'intérêt d'un tel codage est, outre d'être indépendant d'une échelle absolue ou d'intervalle, d'être indépendant aussi d'une échelle ordinale qui n'a peut être pas de sens.

On peut d'ailleurs prendre a et b au lieu de 0 et 1 , ou permuter l'ordre des classes, l'analyse n'est pas modifiée. Ce n'est pas la même chose si on conserve X_j codée 1, 2, ...r, ... k

Exemple : (cf R. CEHESSAT, 1976)

Si on extrait d'un tableau 4 individus caractérisés par 8 variables booléennes où le codage booléen est arbitraire, on constate que les distances du χ^2:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
i_1	1	1	1	1	1	1	0	1
i_2	1	1	1	1	1	1	1	0
i_3	0	0	0	0	0	0	0	1
i_4	0	0	0	0	0	0	1	0

$$D^2(i_1,i_2) = \sum_{j=1}^{8} \frac{1}{P_{.j}} \left[\frac{n_{1j}}{n_{1.}} - \frac{n_{2j}}{n_{2.}} \right]^2$$

$D^2(i_1,i_2)$ et $D^2(i_3,i_4)$ sont telles que $D^2(i_3,i_4) = 4\,9\, D^2(i_1,i_2)$

Or si l'on compare ces 2 couples, ils ont des comportements voisins (6 variables communes et 2 anomalies).

L'anomalie provient du rôle dissymétrique que jouent 0 et 1 dans les valeurs marginales $n_{1.}$ ou $n_{2.}$ par rapport à $n_{3.}$ et $n_{4.}$.

Dans certains cas, où la valeur 1 a un sens beaucoup plus fort que 0 (cas des études de biologie, écologie) où la présence d'une espèce a un rôle plus important que l'absence on peut et même on doit conserver ce codage. Dans le cas contraire, il introduit un biais que le codage disjonctif permet d'éviter. Ici, il suffit de dédoubler chaque variable pour que $D^2(i_3,i_4)$ devienne voisine de $D^2(i_1,i_2)$.

Dans le paragraphe suivant, nous analyserons l'effet du codage disjonctif sur l'A.F.C. d'un tableau d'observations de variables initialement quelconques (continues, discrètes, etc...). Cependant la portée de ce codage est plus générale, et on la retrouvera par exemple en analyse discriminante (IIe partie, chap.IV ou Ve partie, chap.I)

II.2.2. Propriétés du codage disjonctif

(a) Etant donné un tableau de N observations de p variables, celles-ci étant séparées en k classes équiprobables, on a donc kxp nouvelles variables.

Les classes étant incompatibles et équiprobables on aura pour une variable j donnée :

	X_j				
$i=1$	C_1	C_2	C_3	C_k
	0	0	1	0
i	0	1	0	0
N	0	0	0	0

. $\frac{N}{k}$ fois 1 dans 1 une colonne(en moyenne).

. 1 fois 1 et (k-1) fois 0 pour la partie de l'observation i correspondant à la variable j.

On aura au total $N \times P$ fois 1, et le tableau de fréquences aura pour particularités :

1) $P_{ij} = 0$ ou $P_{ij} = \frac{1}{N \times P}$

2) $P_{i\cdot} = \sum_{j=1}^{k.P} P_{ij} = P \times \frac{1}{N.P} = \frac{1}{N} \quad \forall i$

3) $P_{\cdot j} = \sum_{i=1}^{N} P_{ij} = \frac{N}{k} \times \frac{1}{N.P} = \frac{1}{k.P} \quad \forall j$

$\Longrightarrow$ Les fréquences marginales sont toutes identiques pour (I) et (J) respectivement.

On pourra remarquer que, dans le tableau codé, la moyenne d'une variable :

$$\overline{x_j} = \frac{1}{N} \sum_{i=1}^{N} x_{ij} = \frac{1}{N} \left(\frac{N}{k} \times 1 \right) = \frac{1}{k}$$

et une analyse en composantes principales non normée reviendrait à analyser le tableau :

$$x_{ij} - \overline{x_j} = x_{ij} - \frac{1}{k}$$

Or l'analyse des correspondances revient à analyser :

$$\frac{P_{ij} - P_{i\cdot} \times P_{\cdot j}}{\sqrt{P_{i\cdot} \times P_{\cdot j}}} = \frac{\frac{x_{ij}}{N \times P} - \frac{1}{k \times N \times P}}{\sqrt{\frac{1}{k \times N \times P}}} = \sqrt{\frac{k}{N.P}} \left(x_{ij} - \frac{1}{k} \right)$$

soit, à un facteur constant près, le même tableau .

$\Longrightarrow$ L'analyse des correspondances du tableau de codage disjonctif complet (avec même nombre de classes k quelque soit la variable et même effectif quelque soit la classe) se ramène à une simple A.C.P.

ⓑ Si l'on s'intéresse aux variables, il est intéressant de comparer les 2 nouvelles variables j et j' correspondant à 2 classes d'une même variable brute :

$$D^2(j,j') = \sum_{i=1}^{N} \frac{1}{P_{i\cdot}} \left(\frac{P_{ij}}{P_{\cdot j}} - \frac{P_{ij'}}{P_{\cdot j'}} \right)^2 = \frac{1}{P_{i\cdot} P_{\cdot j}^2} \sum_{i} \left(P_{ij} - P_{ij'} \right)^2$$

car $P_{i\cdot}$ et $P_{\cdot j}$ ne dépendent plus de i et j.

De plus, $P_{ij} = \frac{1}{N.P}$ pour $\frac{N}{k}$ observations i pour lesquelles $P_{ij'} = 0$ et réciproquement. On aura donc la sommation $2 \times \frac{N}{k}$ valeurs non nulles égales à $\left(\frac{1}{N \times P} \right)^2$

$$D^2(j,j') = N \left(k \times P \right)^2 \times 2 \frac{N}{k} \times \left(\frac{1}{N \times P} \right)^2 = 2k$$

$\Longrightarrow$ La distance entre 2 points correspondant à 2 classes d'une même variable brute est constante et égale à $\sqrt{2k}$.

ⓒ Si on calcule maintenant la distance du barycentre G au point j :

$$D^2(G,j) = \sum_{i=1}^{N} \frac{1}{P_{i\cdot}} \left(\frac{P_{ij}}{P_{\cdot j}} - P_{i\cdot} \right)^2 = \frac{1}{P_{i\cdot} \times P_{\cdot j}^2} \sum_{i=1}^{N} \left(P_{ij} - P_{i\cdot} P_{\cdot j} \right)^2$$

Or P_{ij} vaut $\frac{N}{k}$ fois 1 et $\frac{k-1}{k} \cdot N$ fois 0 d'où la sommation :

$$D^2(G,j) = N \times (k \cdot P)^2 \cdot \left[\frac{N}{k} \left(\frac{1}{N \times P} - \frac{1}{k \cdot N \cdot P} \right)^2 + \frac{k-1}{k} N \left(\frac{1}{k \cdot N \cdot P} \right)^2 \right]$$

$$D^2(G,j) = k \left(1 - \frac{1}{k} \right) + \frac{k-1}{k} = k - 1$$

$\Longrightarrow$ La distance entre le barycentre et une classe quelconque j de la variable brute considérée est constante et égale à $\sqrt{k-1}$.

Dans l'hypothèse d'équiprobabilité des classes, les points correspondant aux différentes classes d'une variable brute sont donc sur un polyèdre régulier (triangle équilatéral, tétraèdre, selon la valeur de) inscrit dans une hypersphère de centre G, commune à toutes les variables et de rayon .

Remarque. Un cas particulier intéressant est celui du codage en 3 classes (k=3). Les 3 points-classes d'une variable brute forment un triangle équilatéral, donc plan, et la variable est d'autant mieux représentée dans un plan factoriel (F_ℓ , F_m) qu'elle se projette selon un triangle équilatéral.

De plus, ce triangle a pour côté $\sqrt{2k} = \sqrt{6}$ et il est inscrit dans un cercle de rayon égal à $\sqrt{6} \times \frac{\sqrt{3}}{3} = \sqrt{2}$

Or le rayon de l'hypersphère , $\sqrt{k-1} = \sqrt{2}$ lui est égal, donc :

$\Longrightarrow$ le triangle, inscrit dans la sphère, se trouve de plus dans un plan diamétral.

On voit alors apparaître un problème de degré de liberté : Si les 2 classes j et j' , ℓ et ℓ' de 2 variables respectives sont confondues, alors les 3èmes classes de chaque variable, j" et ℓ'' , sont aussi confondues. On vérifiera que cela est vrai quel que soit k : Si k-1 classes de 2 variables sont confondues, les kèmes le sont aussi. Ceci provient de la contrainte du codage disjonctif :

$$\sum_i P_{ij} = P_{\cdot j} = P_{\cdot j'} = P_{\cdot j''} = \frac{1}{k \cdot P} \qquad \forall\, j, j', j''$$

ⓓ Si maintenant on considère la classe j d'une variable brute et la classe ℓ d'une autre variable brute, on peut calculer :

$$D^2(j,\ell) = \sum_{i=1}^{N} \frac{1}{P_{i\cdot}} \times \left(\frac{P_{ij}}{P_{\cdot j}} - \frac{P_{i\ell}}{P_{\cdot \ell}} \right)^2 = N \times (k \cdot P)^2 \cdot \sum_i \left(P_{ij} - P_{i\ell} \right)^2$$

Cette fois les valeurs 0 et 1 ne sont plus "disjointes" et on a la table de contingence avec les effectifs a, b, c, d, vérifiant :

Classe l de X_2 \ Classe j de X_2	1	0
1	a	b
0	c	d

$$a + b = a + c = \frac{N}{k}$$

$$b + d = c + d = \frac{k-1}{k} N$$

et surtout s'il y a b cas où $P_{ij} = 1/N.P$, alors que $P_{il} = 0$, alors il y a $c = b$ cas où $P_{il} = 1/N.P$ tandis que $P_{ij} = 0$ donc la table devient :

a	b	$\frac{N}{k}$
b	d	$\frac{k-1}{k} N$
$\frac{N}{k}$	$\frac{k-1}{k} N$	

et la distance s'écrit:

$$D^2(j,l) = N(k \times P)^2 \times 2b \times \left(\frac{1}{N.P}\right)^2 = \frac{2b.k^2}{N}$$

ou encore :

$$D^2(j,l) = 2\left(\frac{N}{k} - a\right).\frac{k^2}{N}$$

car on voit ainsi que si $a = \frac{N}{k}$, la distance devient nulle .

Si maintenant on considère la corrélation entre 2 variables binaires (cf Ière Partie) on trouve :

$$\rho_{j,l} = \frac{a.d - b.c}{\sqrt{(a+b)(a+c)(b+c)(c+d)}} = \frac{\left(\frac{N}{k} - b\right)\left(\frac{k-1}{k}N - b\right) - b^2}{\left(\frac{N}{k}\right).\left(\frac{k-1}{k}N\right)}$$

$$\rho_{j,l} = 1 - b\,\frac{k^2}{N(k-1)}$$

d'où $$\rho_{j,l} = 1 - \frac{D^2(j,l)}{2.(k-1)}$$

On remarquera que la distance entre les 2 points est différente de celle que l'on obtiendrait en composantes principales normées : $\rho_{jl} = 1 - \frac{D^2(j,l)}{2}$

Ici les variables ne sont pas normées mais ont même variance $\sqrt{(k-1)/k}$ et on peut donc encore interpréter leur distance en terme de corrélation.

Remarquons aussi que :

1. Si 2 classes j et j', l et l' de 2 variables initiales sont confondues, alors j'' et l'' sont confondues (cf ©)
2. Si 2 variables initiales sont coplanaires, les 3 interdistances sont égales.
3. Si 2 variables initiales sont quelconques, la connaissance de 2 distances $D(j,l)$ et $D(j',l')$ impose la 3ème $D(j'',l'')$.

Toutes ces propriétés sont liées aux k-1 degrés de liberté.

ⓔ Utilisation des résultats :

On s'intéresse à la similitude des variables, donc de leur représentation. Pour cela, il faut comparer 2 variables dans un plan factoriel où l'une au moins est bien représentée. Or la qualité de la représentation d'une variable est d'autant meilleure que celle-ci se projette selon un triangle équilatéral.(cf exemples dans le paragraphe II.4)

Signalons toutefois que le développement particulier que nous avons donné au cas n=3 se justifie par les considérations suivantes :

- savoir qu'une variable a une valeur forte moyenne ou faible est déjà très informatif,
- les constructions géométriques planes auxquelles nous aboutissons sont beaucoup plus faciles à interpréter que les projections planes de polyèdres à 3 dimensions ou plus,
- et surtout, nous rencontrons fréquemment des variables présentant un point d'accumulation qui regroupe environ 30% des observations.

Par exemple :

- la précipitation journalière (30 à 40% de valeurs nulles)
- mais aussi la nébulosité (1/3 des cas avec une nébulosité nulle, ou totale, ou intermédiaire, etc...)

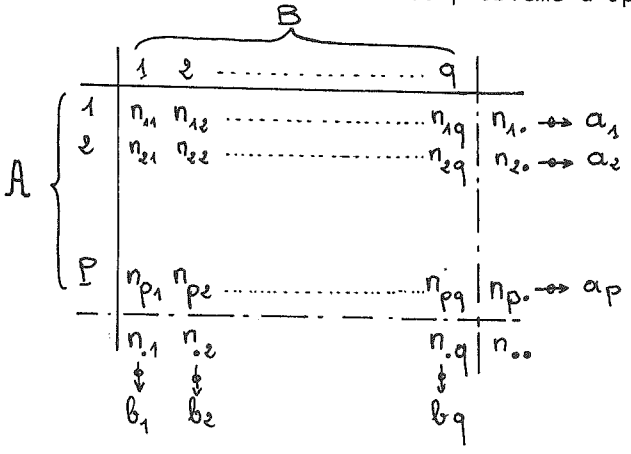
Cependant, pour tenter d'être complet sur une méthode relativement peu connue des utilisateurs (cf le titre de M.O. HILL, 1974 ...!) et avant de présenter quelques exemples de traitement, nous donnons quelques points de vue complémentaires pouvait aider à l'interprétation.

II.3 - Interprétations complémentaires de l'analyse en correspondances (*)

II.3.1. Comparaison de 2 variables classées (*)

ⓐ Etant données 2 variables A et B respectivement découpées en p et q classes (donc nominales), une méthode pour mesurer leur degré d'association, due à WILLIAMS (1952), présentée par ANDERBERG (1973), consiste à associer à chaque classe de A une valeur a_i, et à chaque classe de B une valeur b_j, et à calculer le coefficient de corrélation des a_i et des b_j. On choisit simplement les a_i et les b_j de façon à maximiser la corrélation entre A et B .

La solution de ce problème d'optimisation est assez laborieuse. Si l'on considère le tableau de contingence où les n_{ij}, $n_{i.}$, $n_{.j}$ sont des effectifs, on a :



$$r = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(a_i - \bar{a})(b_j - \bar{b})}{\sqrt{\sum_i n_{i.}(a_i - \bar{a})^2} \cdot \sqrt{\sum_j n_{.j}(b_j - \bar{b})^2}}$$

et l'invariance de r dans toute transformation linéaire nous oblige pour lever l'indétermination sur les a_i et b_i , à

imposer par exemple qu'ils soient centrés et réduits :

$$\bar{a} = \sum_i n_{i.} a_i / n_{..} = 0 \qquad \bar{b} = \sum_j n_{.j} b_j / n_{..} = 0$$

$$\sigma_a^2 = \sum_i n_{i.} a_i^2 / n_{..} = 1 \qquad \sigma_b^2 = \sum_j n_{.j} b_j^2 / n_{..} = 1$$

On va donc chercher à maximiser $|r|$, ou plutôt r^2 sous ces contraintes, d'où, avec les multiplicateurs de Lagrange λ_ℓ, la quantité à maximiser :

$$L = \left[\sum_{i=1}^{P} \sum_{j=1}^{q} n_{ij} a_i b_j / n_{..} \right]^2 - \lambda_1 \sum_{i=1}^{P} n_{i.} a_i / n_{..} - \lambda_2 \sum_{j=1}^{q} n_{.j} b_j / n_{..}$$
$$- \lambda_3 \left[\sum_{i=1}^{P} n_{i.} a_i^2 / n_{..} - 1 \right] - \lambda_4 \left[\sum_{j=1}^{q} n_{.j} b_j^2 / n_{..} - 1 \right]$$

La résolution (WILLIAMS (1952) et ANDERBERG (1973)) comporte les étapes suivantes :

- la Différentiation par rapport à chaque b_j
- la Sommation par rapport à j qui montre que $\lambda_2 = 0$
- le Produit de chaque équation par b_j et la sommation par rapport à j qui entraînent : $\lambda_4 = r^2$
- compte tenu des relations ci-dessus, on extrait :

$$\forall j \qquad b_j = \sum_{i=1}^{P} \frac{n_{ij} . a_i}{n_{.j} r}$$

et

$$r^2 = \sum_{j=1}^{q} \left(\sum_{i=1}^{P} n_{ij} . a_i \right) / n_{.j} n_{..}$$

Ceci résoud le problème si les a_i sont connus, par exemple si seule B est une variable nominale. Dans le cas contraire, on a trouvé pour chaque b_j et pour r une expression en fonction des a_i, que l'on reporte dans l'expression à optimiser, ainsi que $\lambda_2 = 0$ et $\lambda_4 = r^2$

- On dérive par rapport à chaque a_i d'où les équations :

$$2 \sum_{j=1}^{q} n_{ij} \sum_{k=1}^{P} n_{kj} a_k / n_{.j} n_{..} - \lambda_1 . n_{i.} / n_{..} - 2 \lambda_3 n_{i.} a_i / n_{..} = 0$$

- la sommation sur i montre que $\lambda_1 = 0$
- le produit par a_i et la sommation par rapport à i entraînent que $\lambda_3 = r^2$

Il reste à déterminer les a_i. On a les i équations (i=1 à p):

$$2 \sum_{j=1}^{q} n_{ij} \sum_{k=1}^{P} n_{kj} a_k / n_{.j} n_{..} - 2 r^2 n_{i.} a_i / n_{..} = 0$$

Si on multiplie par $\sqrt{\frac{n_{..}}{n_{i.}}}$ et $\sqrt{\frac{n_{k.}}{n_{k.}}}$, on obtient :

$$\sum_{j=1}^{q} \sum_{k=1}^{P} \left[\frac{n_{kj}}{\sqrt{n_{k.} \times n_{.j}}} \times \frac{n_{ij}}{\sqrt{n_{i.} \times n_{.j}}} \right] \sqrt{\frac{n_{k.}}{n_{..}}} \times a_k - r^2 \times \sqrt{\frac{n_{i.}}{n_{..}}} \times a_i = 0$$

En posant $m_i = \sqrt{\frac{n_{i.}}{n_{..}}}$, on cherche donc les a_i qui satisfont

$$\sum_{k=1}^{p}\left[\sum_{j=1}^{q}\frac{n_{kj}}{\sqrt{n_{k.}\times n_{.j}}}\times\frac{n_{ij}}{\sqrt{n_{i.}\times n_{.j}}}\right]m_k\cdot a_k - r^2 m_i\cdot a_i = 0$$

ou :

$$\sum_{k=1}^{p} t_{ik}\,(m_k\cdot a_k) - r^2(m_i\cdot a_i) = 0$$

Ce qui revient à chercher les valeurs propres et vecteurs propres de la matrice d'élément :

$$t_{ik} = \sum_{j=1}^{q}\frac{n_{kj}}{\sqrt{n_{k.}\,n_{.j}}}\,\frac{n_{ij}}{\sqrt{n_{i.}\,n_{.j}}} - \frac{\sqrt{n_{i.}\,n_{k.}}}{n_{..}}$$

Or, si on se rappelle que tous les n_{ij} sont des effectifs et que l'on divise numérateur et dénominateur par $n_{..} = N$, on trouve :

$$t_{ik} = v_{ik} = \sum_{j=1}^{q}\frac{p_{kj}}{\sqrt{p_{k.}\cdot p_{.j}}}\times\frac{p_{ij}}{\sqrt{p_{i.}\cdot p_{ij}}} - \sqrt{p_{i.}\,p_{k.}}$$

soit la matrice V trouvée en II.2.1. ⓒ lors de la résolution du problème d'analyse en correspondance dans $\mathbb{R}^p$. D'où une première propriété importante :

$\Longrightarrow$ L'affectation, à chacune des classes de 2 variables classées A et B, de valeurs a_i et b_i qui maximisent la corrélation entre A et B revient à faire une analyse des correspondances du tableau de contingence T de A et B.

Ou les "scores" , c'est-à-dire les nouvelles coordonnées affectées par l'analyse en correspondance du tableau de contingence de A et B aux modalités respectives de A et B sont tels que leur corrélation est maximale .

ⓑ Cela rejoint une autre présentation de l'analyse des correspondances (LEBART et FENELON, 1973) : la recherche de la meilleure représentation simultanée de 2 ensembles de modalités .

On cherche à représenter les diverses modalités i de A (i = 1 ... p) par des coordonnées Ψ_i et les modalités j de B (j = 1 ... q) par des coordonnées Φ_j. Chaque point i de A serait barycentre des points j affectés des poids $\frac{p_{ij}}{p_{i.}}$

$$\Psi_i = \sum_{j=1}^{q}\frac{p_{ij}}{p_{i.}}\,\Phi_j$$

et de même pour les points j barycentre des points i affectés des poids $\frac{p_{ij}}{p_{.j}}$

$$\Phi_j = \sum_{i=1}^{p}\frac{p_{ij}}{p_{.j}}\,\Psi_i$$

Comme il est impossible de satisfaire exactement ces 2 relations, on cherche à les satisfaire au mieux en introduisant un coefficient α aussi proche de 1 que possible et tel que :

$$\psi_i = \alpha \sum_{j=1}^{q} \frac{p_{ij}}{p_{i.}} \phi_j \qquad \alpha \leq 1$$

$$\phi_j = \alpha \sum_{i=1}^{p} \frac{p_{ij}}{p_{.j}} \psi_i$$

En remplaçant les ϕ_j dans ψ_i , on obtient :

$$\psi_i = \alpha^2 \sum_{j=1}^{q} \frac{p_{ij}}{p_{i.}} \left(\sum_{k=1}^{p} \frac{p_{kj}}{p_{.j}} \psi_k \right) \Longrightarrow \sum_{k=1}^{p} \sum_{j=1}^{q} \frac{p_{ij} \cdot p_{kj}}{p_{i.} \, p_{.j}} \psi_k = \frac{1}{\alpha^2} \psi_i$$

c'est-à-dire le problème de valeurs propres déjà rencontré en II.2, mais aussi celui obtenu au début de ce paragraphe.

Remarque 1 . Pour être cohérent avec II.2, il faut poser $\psi_k = \frac{u_k}{\sqrt{p_{k.}}}$ et chercher $\vec{u}$ tel que :

$$\sum_{k=1}^{p} \sum_{j=1}^{q} \frac{p_{kj} \cdot p_{ij}}{p_{i.} \cdot p_{.j}} \times \frac{u_k}{\sqrt{p_{k.}}} = \frac{1}{\alpha^2} \times \frac{u_i}{\sqrt{p_{i.}}}$$

qui se transforme en :

$$\sum_{k=1}^{p} \sum_{j=1}^{q} \frac{p_{kj}}{\sqrt{p_{k.} \cdot p_{.j}}} \times \frac{p_{ij}}{\sqrt{p_{i.} \, p_{.j}}} \; u_k = \frac{1}{\alpha^2} \times u_i$$

Remarque 2. La recherche de 2 ensembles de valeurs a_i et b_j (resp ψ_i et ϕ_j) caractérisant au mieux 2 ensembles de modalités A et B avait déjà été utilisée antérieurement de manière itérative. On choisissait arbitrairement les b_j (resp. ϕ_j), on calculait :

$$\psi_i = \sum_{j=1}^{q} \frac{n_{ij}}{n_{i.}} \phi_j$$

puis on utilisait les ψ_i pour calculer :

$$\phi_j = \sum_{i=1}^{p} \frac{n_{ij}}{n_{i.}} \psi_i$$

II.3.2. Matrice d'incidence (*)

On a vu toutefois que les problèmes qui nous intéressaient n'étaient pas constitués directement par des tableaux de contingence, mais croisaient plutôt les observations, sur un certain nombre d'individus, de différents caractères.

Mais on peut passer, dans un premier temps, d'un tableau de contingence à une matrice d'incidence.

Soit un tableau T reliant les variables classées A et B donc à 2 entrées. Si l'on considère chaque individu, on peut lui associer la modalité de A et celle de B auxquelles il appartient.

Exemple :

$$T = \begin{array}{c} \overbrace{}^{B} \\ 1\ \ 2\ \ 3 \end{array} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{matrix} 1 \\ 2 \end{matrix} \Big\} A \quad \Longrightarrow \quad I(T) = \begin{array}{cc|ccc} \overbrace{1\ \ 2}^{A} & & \multicolumn{3}{c}{\overbrace{1\ \ 2\ \ 3}^{B}} \\ \hline 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{array}$$

Et sur la matrice d'incidence, on voit que la liaison entre les 2 variables A et B se ramène à l'étude de la liaison entre 2 ensembles de variables $\{A_1, A_2, \ldots A_p\}$ et $\{B_1, B_2, \ldots B_q\}$ donc à une corrélation canonique.

$$I(T) = \begin{array}{c} \\ \text{Classes :} \\ \text{individu } k \\ \\ N \end{array} \begin{array}{c} \overbrace{1\ \ 2 \ldots i \ldots p}^{A} \quad \overbrace{1\ \ 2 \ldots j \ldots q}^{B} \\ \\ \begin{bmatrix} & & & & \\ & & & & \\ 0 \ldots 1 \ldots 0 & & 0 \ldots 1 \ldots 0 & & \\ & & & & \\ & & & & \end{bmatrix} \end{array}$$

Or celle-ci est la recherche de 2 vecteurs ou facteurs $\{a_1, a_2 \ldots a_p\}$ et $\{b_1, b_2 \ldots b_p\}$ tels que :

$$\sum_{i=1}^{p} x_{ik} \cdot a_i$$

et

$$\sum_{j=1}^{q} y_{jk} \cdot b_j$$

aient une corrélation maximale sur l'ensemble des individus .

Or ici : $x_{ik} = 0$ ou 1

$y_{jk} = 0$ ou 1

et l'individu k a un "score" sur le facteur $\{a_1 \ldots a_p\}$ associé à A :

$$\sum_{i=1}^{p} [I(T)]_{i,k}\ a_i$$

et

$$\sum_{j=1}^{q} [I(T)]_{j,k}\ b_j$$

sur celui associé à B.

On peut imposer, sans limiter la généralité, qu'ils soient centrés :

$$\sum_{k=1}^{N} \left(\sum_{i=1}^{p} x_{ik}\, a_i \right) = \sum_{i=1}^{p} a_i \sum_{k=1}^{N} x_{ik} = \sum_{i=1}^{p} n_{i.}\, a_i = 0$$

et de même : $\sum_{j=1}^{q} n_{.j}\, b_j = 0$ en ayant posé : $[I(T)]_{i,k} = x_{ik}$

$[I(T)]_{j,k} = y_{jk}$

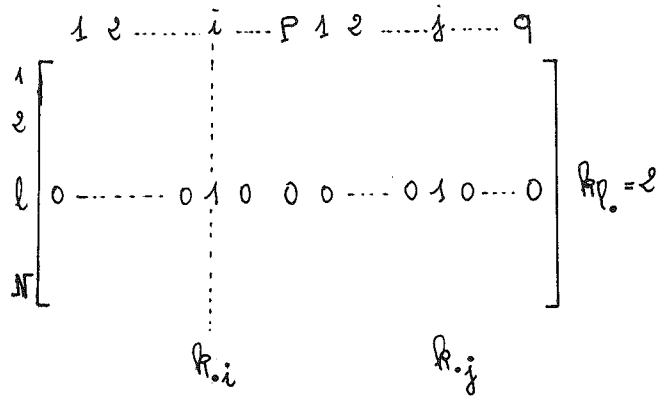
La corrélation entre les 2 facteurs s'écrit :

$$\rho = \frac{\sum_{k=1}^{N}\left(\sum_{i=1}^{P} x_{ik}\, a_i\right)\left(\sum_{j=1}^{q} y_{jk}\, b_j\right)}{\sqrt{\left[\sum_{k=1}^{N}\left(\sum_{i=1}^{P} x_{ik}\, a_i\right)^2\right]\times\left[\sum_{k=1}^{N}\left(\sum_{j=1}^{q} y_{jk}\, b_j\right)^2\right]}} = \frac{\sum_{i=1}^{P}\sum_{j=1}^{q} n_{ij}\, a_i\, b_j}{\sqrt{\left(\sum_{i=1}^{P} n_{i.}\, a_i^2\right)\left(\sum_{j=1}^{q} n_{.j}\, b_j^2\right)}}$$

et si on impose en plus aux scores la contrainte d'être réduits, on retrouve exactement le problème d'optimisation du paragraphe (a) .

$\Longrightarrow$ L'analyse canonique du tableau d'incidence I(T) est identique à l'analyse en corrélation de WILLIAMS, elle-même identique à l'analyse en correspondance du tableau T.

Considérons maintenant l'application de l'analyse des correspondances non pas à T , mais directement à I(T) et appelons $k_{\ell i}$ un de ses éléments.



On remarquera que

$k_{i\ell}$ = 1 si $\ell \in$ modalité i de A
0 sinon

$k_{j\ell}$ = 1 si $\ell \in$ modalité j de B
0 sinon

$k_{\ell.} = 2 \quad \forall \ell$

$k_{.i}$ = nombre de sujets présentant la modalité i de A.

On sait que l'analyse en correspondance va nous conduire à diagonaliser la matrice: $S = R^t . R$ (cf. LEBART et FENELON 1973)

avec $R = \{r_{\ell i}\}$ $\quad r_{\ell i} = \dfrac{k_{\ell i}}{\sqrt{k_{\ell.}\, k_{.i}}}$ $\qquad S = \{s_{ij}\}$ $\quad s_{ij} = \sum_{\ell=1}^{N} \dfrac{k_{\ell i} \times k_{\ell j}}{k_{\ell.}\sqrt{k_{.i} \times k_{.j}}}$

Or $\sum_{\ell=1}^{N} k_{\ell i} . k_{\ell j}$ = nombre d'individus présentant à la fois la modalité i de A et j de B = n_{ij} avec 2 possibilités supplémentaires:

$n_{ii'}$ où i et i' sont 2 modalités de A $= 0$
$n_{jj'}$ où j et j' sont 2 modalités de B

car les modalités s'excluent mutuellement.

n_{ii} = $k_{.i}$ ou $k_{.j}$: ensemble des individus présentant la modalité i ou j.
n_{jj}

Le tableau S associé à I(T) devient donc, compte tenu de $k_{\ell.} = 2$

$$S = \begin{bmatrix} \frac{1}{2}\frac{1}{2} & s_{ij} \\ s_{ji} & \frac{1}{2}\frac{1}{2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} I_A & C \\ C^t & I_B \end{bmatrix}$$

où C est le transformé pour l'analyse en correspondance $c_{ij} = n_{ij} / \sqrt{n_{i\cdot} n_{\cdot j}}$, du tableau de contingence $T = \{n_{ij}\}$.

Or l'analyse en correspondance de I(T) nous conduit à diagonaliser S, tandis que l'analyse du tableau de contingence nous conduit à diagonaliser $C^t . C$.

Si on appelle u un vecteur propre de S associé à la valeur propre λ, u de dimension p+q se décompose en $u_A + u_B$:

$$\frac{1}{2}\begin{bmatrix} I_A & C \\ C^t & I_B \end{bmatrix} \times \begin{bmatrix} u_A \\ u_B \end{bmatrix} = \lambda \begin{bmatrix} u_A \\ u_B \end{bmatrix} \Longrightarrow \begin{array}{l} \frac{1}{2}\left(I_A . u_A + C . u_B\right) = \lambda u_A \\ \frac{1}{2}\left(C^t . u_A + I_B . u_B\right) = \lambda u_B \end{array}$$

$$\Longrightarrow \begin{array}{l} C u_B = 2\left(\lambda - \frac{1}{2}\right) u_A \\ C^t u_A = 2\left(\lambda - \frac{1}{2}\right) u_B \end{array} \Longrightarrow \begin{array}{l} C^t C u_B = 4\left(\lambda - \frac{1}{2}\right)^2 u_B \\ C . C^t u_A = 4\left(\lambda - \frac{1}{2}\right)^2 u_A \end{array}$$

donc les 2 sous-vecteurs u_A et u_B sont vecteurs propres de $C.C^t$ et $C^t.C$ relativement à la valeur propre $4\left(\lambda - \frac{1}{2}\right)^2$, mais ils sont aussi vecteurs propres de S relativement à λ.

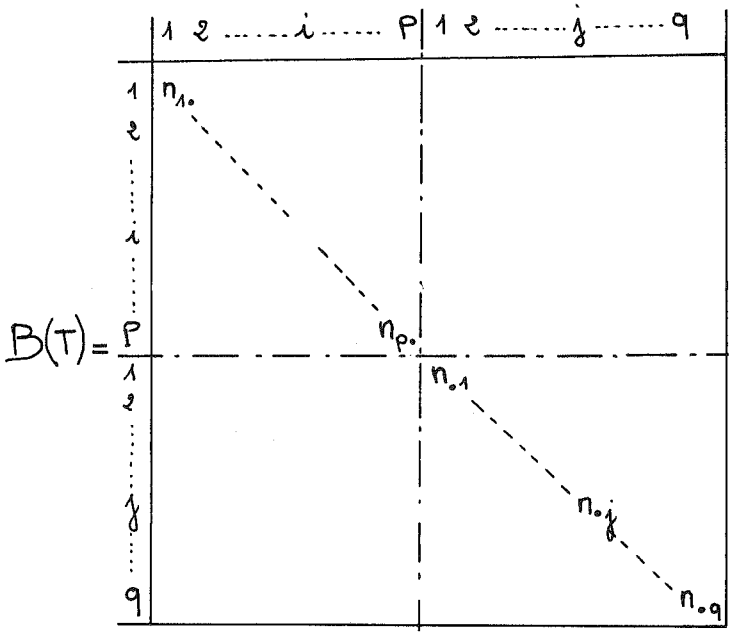
On trouvera par ailleurs (R. CEHESSAT, 1976) un certain nombre de propriétés. Pour nous, l'essentiel se résume à celles-ci :

$\Longrightarrow$ L'analyse en correspondance du tableau I(T) est équivalente à celle du tableau de contingence T qui croise les p modalités de A avec les q modalités de B. Les coordonnées des variables (modalités de A ou B) sont les mêmes (à des facteurs multiplicatifs près car si u est normé, u_A et u_B ne le sont pas). D'autre part, cette analyse bien que coûteuse (analyse d'un tableau de dimension p+q au lieu d'un tableau de dimension min(p,q) a l'avantage de fournir des représentations des variables i et j comparables entre elles et une représentation des individus.

On constate toutefois que tous les résultats obtenus en II.2.3.a) et b) ne s'appliquent qu'aux tableaux de contingence (ou à leurs matrices d'incidence associées) à 2 entrées, donc à 2 variables A et B.

II.3.3. Tableau de BURT (*)

Il existe une 3ème façon d'effectuer l'analyse en correspondance du tableau T croisant les variables A et B. Au lieu d'associer à T le tableau I(T) on lui associe le tableau de BURT B(T) :



Ce tableau est symétrique, et la somme des lignes ou des colonnes est égale à :
2 x ($n_{i.}$ ou $n_{.j}$)

Si on applique l'analyse des correspondances à ce tableau, on diagonalise une matrice :

$$P = Q.Q^t = Q^t.Q = Q^2$$ avec Q déduit de B(T) par :

$$q_{ij} = \frac{b_{ij}}{\sqrt{b_{i.}\, b_{.j}}} = \frac{n_{ij}}{2\sqrt{n_{i.}\, n_{.j}}}$$

avec : $n_{ii} = n_{i.}$ si i est une modalité de A et respectivement pour j.

$n_{ii'} = 0$ si i et i' sont 2 modalités différentes de A, et respectivement pour j

On constate donc que Q est strictement identique à S obtenu au paragraphe précédent et donc que : $P = S^2$

donc P et S ont mêmes vecteurs propres.

Les coordonnées des points j dans l'une ou l'autre analyse (I(T) ou B(T)) diffèrent seulement par le fait que la valeur propre associée au facteur k est λ_k dans la 1ère et λ_k^2 dans la seconde, d'où une distorsion éventuelle du nuage (cf R. CEHESSAT, 1976).

Cette présentation, ici encore, est limitée au cas de 2 variables à p et q modalités. Toutefois, nous allons voir que c'est elle qui permet l'extension à $n > 2$ variables.

Remarque. On a vu que, dans l'analyse en correspondance du tableau de contingence T, on avait, pour un facteur donné k les coordonnées :

$$\psi_{ik} = \varphi_k(i) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{q} \frac{u_{jk}}{\sqrt{P_{\cdot j}}} \times \frac{P_{ij}}{P_{i\cdot}} = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{q} \phi_{jk} \cdot \frac{P_{ij}}{P_{i\cdot}}$$

et

$$\phi_{jk} = g_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{p} \frac{v_{ik}}{\sqrt{P_{i\cdot}}} \times \frac{P_{ij}}{P_{\cdot j}} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{p} \psi_{ik} \cdot \frac{P_{ij}}{P_{\cdot j}}$$

Cette relation exprime que le point i de coordonnées ψ_{ik} est barycentre des points j affectés de la masse $P_{ij}/P_{i\cdot}$, et idem pour le point j barycentre des points i affectés de la masse $P_{ij}/P_{\cdot j}$.

Parallèlement, on avait trouvé que, si on affectait à chaque point i sa masse $P_{i\cdot}$:

$$\sum_{i=1}^{p} P_{i\cdot}\, \psi_{ik} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{p} P_{i\cdot} \sum_{j=1}^{q} \frac{u_{jk}}{\sqrt{P_{\cdot j}}} \times \frac{P_{ij}}{P_{i\cdot}} = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{q} u_{jk} \sqrt{P_{\cdot j}} = 0$$

Donc:

$\Longrightarrow$ Le barycentre des points ψ_i affectés de la masse $\frac{P_{ij}}{P_{\cdot j}}$ est ϕ_j, tandis que le barycentre des points ψ_i affectés du poids $P_{i\cdot}$ est 0.

Or, quand nous effectuons l'analyse de $\dot{I}(T)$ ou de B(T), on trouve un facteur associé au vecteur propre :

$$u_k^t = \left\{ u_{Ak}^t, u_{Bk}^t \right\} = \left\{ u_k^t, v_k^t \right\}$$

qui donne les coordonnées :

$$\left\{ \left\{ \psi_i \quad i = 1 \cdots p \right\}, \left\{ \phi_j \quad j = 1 \cdots q \right\} \right\}$$

qui vérifient à la fois, séparément sur $\{\psi_i\}$ et $\{\phi_j\}$ les propriétés de centrage et globalement sur l'ensemble $\{\psi_i\} \cup \{\phi_j\}$

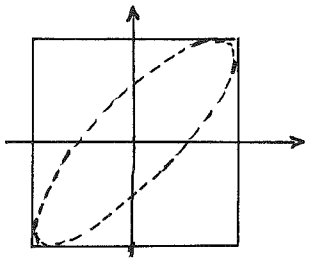
En effet, on a vu que u_A et u_B étaient les facteurs de T, qui vérifient la propriété, et u, facteur de I(T) ou B(T) par l'analyse en correspondance, la vérifie aussi.

On retrouve bien les propriétés vues en II.2.2.b) dans un cas particulier et montrant que l'origine est centre de gravité soit de chaque variable prise séparément (sur p modalités) soit de l'ensemble des variables.

II.3.4. Généralisation (*)

ⓐ Si l'on regarde de plus près la méthode décrite en II.3.1 a, on constate qu'elle a consisté à affecter à chaque classe i de la variable A et j de la variable B des valeurs a_i et b_j, telles que la corrélation, sur les n individus, soit maximale,

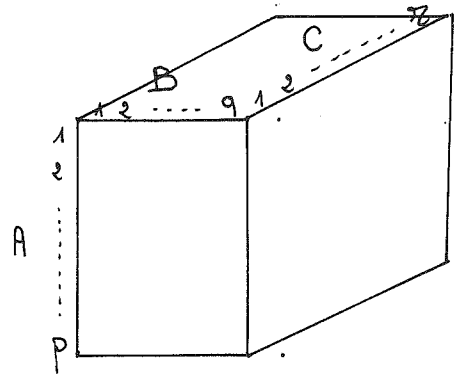
donc que les points dans le plan A, B (A et B étant centrées et réduites) soient le plus possible groupés sur la bissectrice. En fait, A et B jouent un rôle symétrique, et ce que l'on a en fait réalisé, c'est la maximisation de la première composante de A et B, grâce à un codage :



$$a_i = f(i) \qquad b_j = g(j)$$

tiré intrinsèquement des données disponibles. Les fonctions f et g sont non linéaires et définies empiriquement (points par points).

On peut étendre ces résultats à des tableaux de contingence à un nombre quelconque d'entrées. Soit le tableau à 3 entrées :



variable A à p modalités
variable B à q modalités
variable C à r modalités

On peut chercher à coder les modalités A, B et C en respectivement a_i, b_j, c_k telles que nuage des n points dans $\mathbb{R}^3$ soit le plus proche possible de la 1ère trissectrice. Cela revient encore à maximiser la 1ère composante principale nuage.

Ou encore, on cherche le codage :

$$a_i = f(i) \qquad b_j = g(j) \qquad c_k = h(k)$$

tel que la quantité :

$$\lambda = \frac{\text{variance}\,(u\,a_i + v\,b_j + w\,c_k)}{u^2\,\text{var}\,(a_i) + v^2\,\text{var}\,(b_j) + w^2\,\text{var}(c_k)}$$

soit maximum, pour toutes les valeurs possibles de u, v, w et a_i, b_j, c_k avec encore la contrainte de centrage sur a_i, b_j et c_k et on pressent que :

$\Longrightarrow$ Les analyses précédentes se ramènent en fait à une analyse en composantes principales de la matrice d'incidence et se généralisent à plus de 2 variables.

(b) Nous allons donner sans démonstration un aperçu plus rigoureux des résultats.

Si, sur un certain ensemble d'individus N on a mesuré r variables, chacune d'elles, par exemple la qème, ayant Jq modalités :

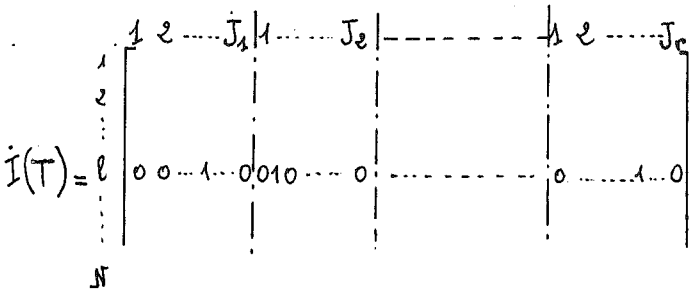
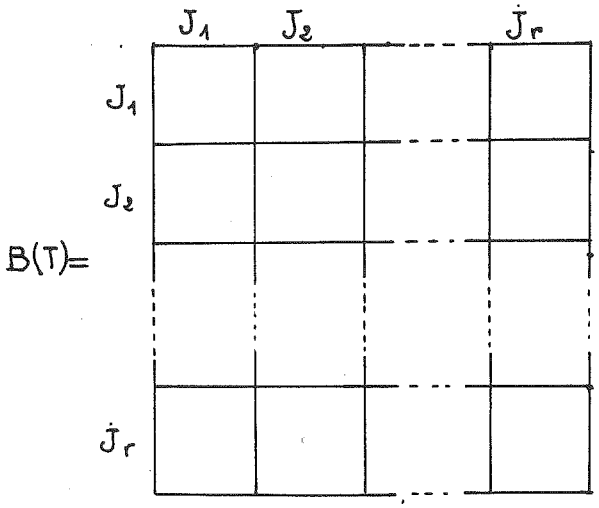


Tableau disjonctif complet ou matrice d'incidence I(T) associée à un tableau de contingence à r entrées .

On peut effectuer l'analyse en correspondance de ce tableau (à $J_1+J_2+\ldots J_r$ dimensions) qui, si les variables ont le même nombre de modalités et celles-ci le même effectif, se ramène à une analyse en composantes principales centrées.

Si maintenant on appelle tableau de BURT celui qui croise :



On montre (CAZES, 1976) que les facteurs de I(T) et B(T) sont les mêmes, à un facteur près, donc les coordonnées des variables (i.e. des modalités des variables) sont les mêmes dans les 2 analyses pour le même facteur considéré.

On vérifie aussi, comme dans le cas de 2 variables, que pour un facteur k donné, la totalité des variables est centrée :

$$\sum_{j=1}^{r} \left(\sum_{l=1}^{J_j} \psi_{kl}^{j} \times P_{\cdot l}^{j} \right) = 0$$

où pour le facteur k, ψ_{kl}^{j} est la coordonnées de la l ème modalité de la variable j et $P_{\cdot l}^{j}$ sa pondération (ou probabilité marginale), soit :

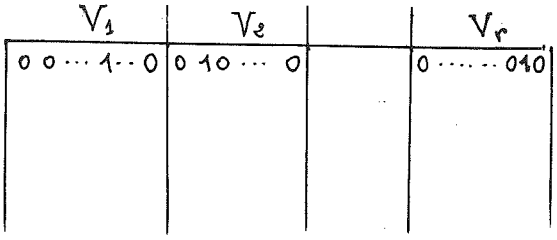
$$P_{\cdot l}^{j} = \frac{n_{\cdot l}^{j}}{n \cdot N}$$

De plus, cette propriété de centrage reste vraie au sein de chaque variable, c'est-à-dire que pour la variable j :

$$\sum_{l=1}^{J_j} \psi_{kl}^{j} \times P_{\cdot l}^{j} = 0$$

(c) Interprétation des valeurs propres associées aux facteurs.

Dans le cas où l'on effectue l'A.F.C. d'un tableau mis sous forme disjonctive complète, la matrice de données initiales devient :



$$R = \{r_{li}\} = \frac{k_{li}}{\sqrt{k_{l\cdot} \times k_{\cdot i}}}$$

Or si les classes sont équiprobables : $k_{\cdot i} = \frac{N}{q}$

q étant le nombre de modalités, et $k_{l\cdot} = r$ le nombre de variables.

Si, dans la matrice $S = R^t.R$ à diagonaliser, nous ne considérons que les termes diagonaux, on a :

$$s_{jj} = \sum_{l=1}^{N} \frac{k_{lj} \times k_{lj}}{k_{l\bullet} \times k_{\bullet j}} = \frac{\frac{N}{q}}{n \times \frac{N}{q}} = \frac{1}{n}$$

la trace est donc : $\sum_{j=1}^{n.q} \frac{1}{n} = q$

Si on lui enlève la valeur propre 1 non significative on a donc une "inertie totale" :

$$C_1 = q - 1$$

Cependant, A. LECLERC et P.AIACH (1978) mettent en garde contre l'utilisation de la quantité λ_j / C_1 pour mesurer l'importance du facteur j.

En effet les valeurs propres du tableau de BURT associé sont $\lambda_1^2 \ldots\ldots \lambda_j^2$ or le critère C_2 :

$$C_2 = \sum_j \lambda_j^2 - \frac{C_1}{n}$$

peut s'interpréter comme la somme des χ^2 associés aux tableaux non diagonaux du tableau de BURT.

La quantité $\frac{\lambda_j^2}{\sum_j \lambda_j^2 - \frac{q-1}{n}}$ est donc mieux appropriée.

II.4 - Exemples d'applications

On trouvera un certain nombre d'exemples dans FRITCH et OBLED (1974) et dans BOIS Ph. (1976). Nous nous limiterons donc aux applications faites sur les problèmes qui nous intéressent directement.

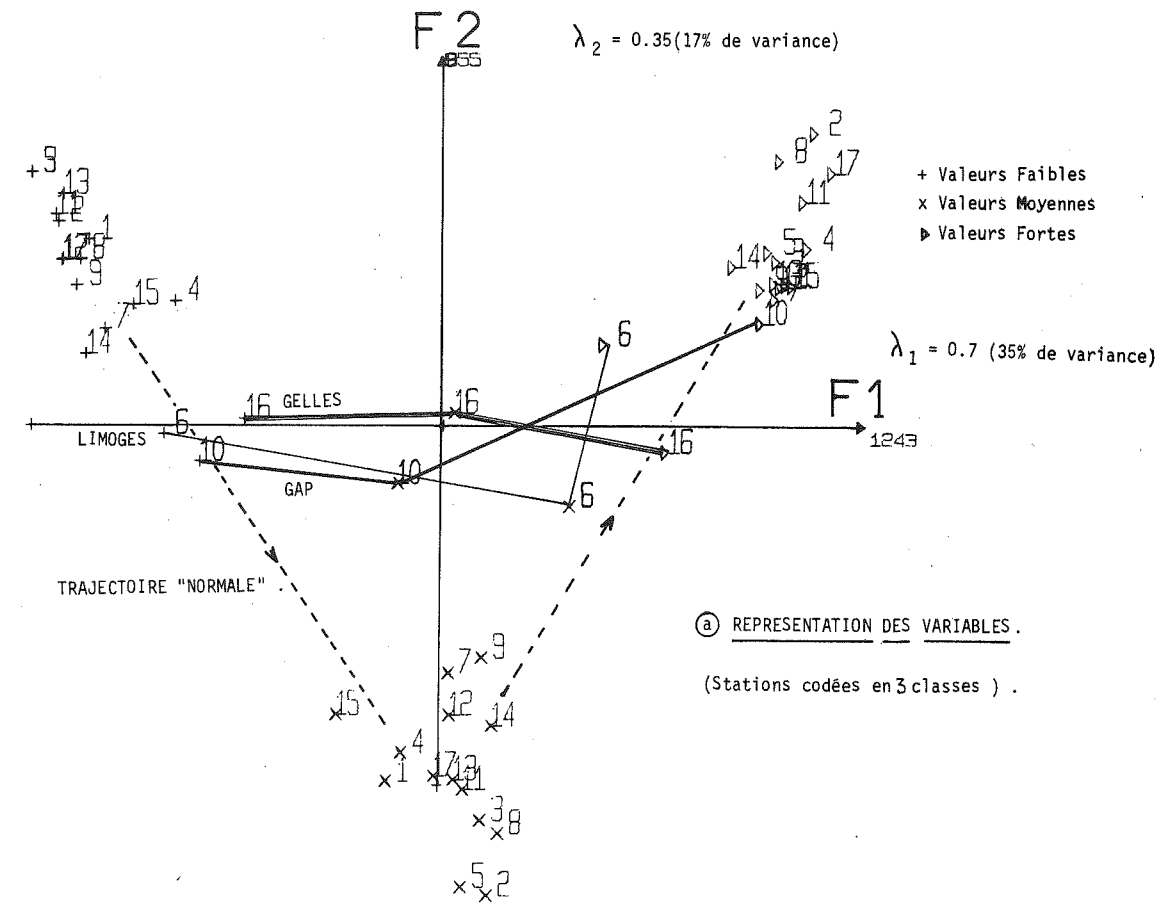
II.4.1. Analyse des liaisons entre variables

En général cette analyse s'effectue plutôt par la méthode des composantes principales, dans la mesure où le coefficient de corrélation linéaire a un sens. (cf d'ailleurs IIIème partie, Chap.II). Par contre quand les variables ne sont pas continues, ou sont très dissymétriques, on peut leur appliquer un codage disjonctif donc les éclater en plusieurs modalités et comparer les "trajectoires" (ensemble de modalités) associées à chaque variable.

Si on reprend l'exemple (cité dans BOIS Ph., 1976, p.73) des températures moyennes des 4 mois d'été (17 stations x 32 années, 1929-1960) on peut coder chaque température en "valeur faible, moyenne ou forte" donc en 3 modalités .

Les classes étant équiprobables et compte tenu d'un très fort effet de taille dû à l'homogénéité du champ en question, le plan F1-F2 est un plan de référence, qui rassemble 52% de la variance (Fig.II-8-a).

La plupart des variables sont représentées par des triangles équilatéraux, sauf les stations 6 10 et 16 . Comme leurs trajectoires sont aussi des triangles équilatéraux mais qu'ils se projettent mal dans F1-F2, cela signifie qu'elles ne cofluctuent pas



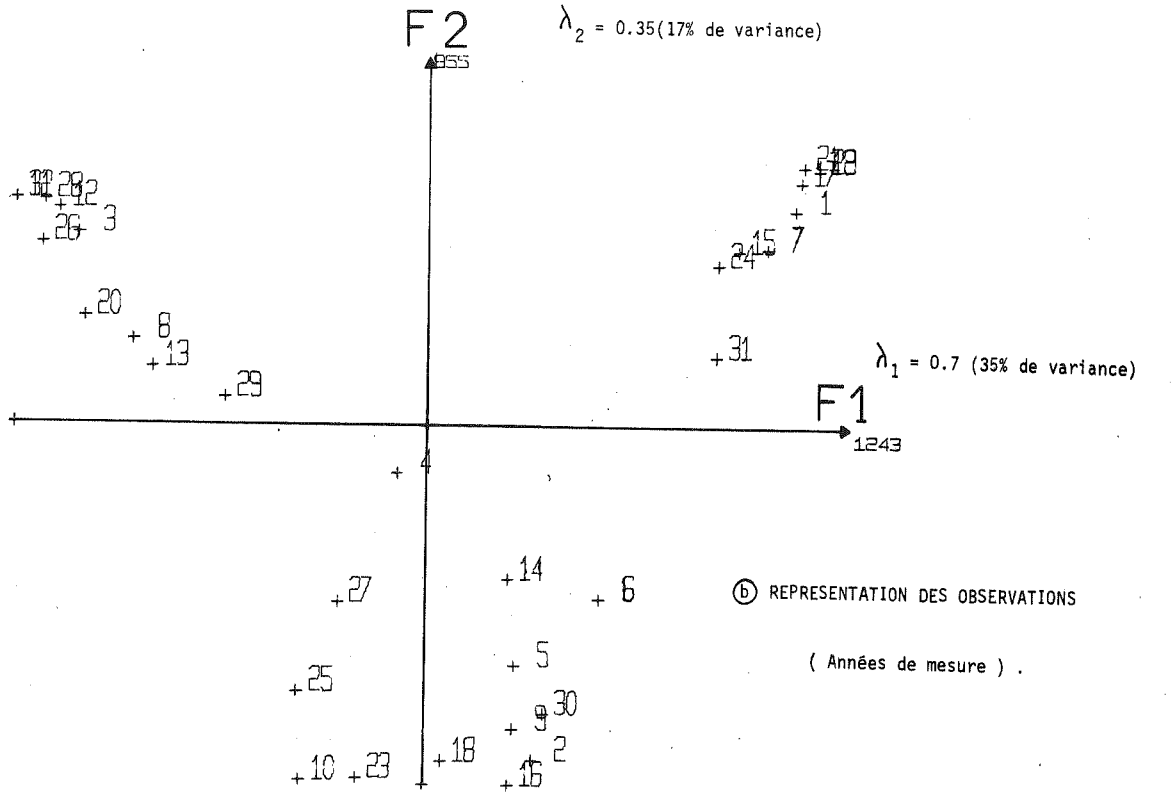


FIGURE II - 8 : Recherche des stations et des années anormales dans l'analyse des températures d'été.

avec le reste du réseau. Parallèlement on a pu classer les années en années froides, moyennes ou chaudes par rapport à l'ensemble des stations (Fig. II-8-b).

Cette technique présente donc un certain intérêt en critique des données, quand celles-ci sont par exemple catégorisées (beau temps, temps variable, mauvais temps) donc ne se prêtent pas à un traitement en corrélation, et que le champ présente une certaine homogénéité permettant de définir un plan de référence.

Par contre, dans des ensembles de variables moins structurés (Exemple : données de Davos), la comparaison des variables est plus délicate car les variables peuvent :

- se projeter selon des triangles équilatéraux voisins ou confondus $\Longleftrightarrow$ cofluctuation
- ne pas se projeter selon des triangles équilatéraux :
  - . parce qu'elles n'ont pu être codées en classes équiprobables
  - . parce qu'elles sont dans un autre plan.

Si le paquet à analyser comporte beaucoup de variables, le nombre de plans à considérer devient grand et vite impraticable. De plus, quand le nombre p de variables est grand, le passage en k modalités conduit à traiter kp variables d'où des matrices qui atteignent vite les limites de capacité de nos programmes.

A titre d'exemple, on a traité un sous-ensemble de 35 variables (extraites des 50 décrites en Ière Partie - Chap. II-1-3) sur le sous-échantillon des seules journées avalancheuses (de Janvier-Février puis de Mars-Avril). Le codage en 3 classes a conduit à un tableau 105x105 calculé sur 150 observations environ.

Comme les 35 variables proviennent des 50 variables élaborées après élimination des redondances essentielles (cf IIIème Partie) il faut s'attendre à des ressemblances assez faibles et à des grappes de variables assez réduites, d'où un nombre de facteurs significatifs assez important.

Si on se limite par exemple au plan F1-F2 on constate (Fig.II-9-a) b) c) d)) :

- En Janvier Février :

  Une bonne concordance des variables de précipitations (1 , 3 , 23). Par contre la densité superficielle 4 ne semble pas liée aux précipitations pour ses valeurs faibles ou moyennes. De même l'enfoncement de la sonde de battage 34 mesure en fait la couche de neige fraîche pour ses fortes valeurs, mais est indépendant des précipitations pour les valeurs moyennes ou faibles.

- En Mars-Avril :

  On constate que la densité superficielle 4 et enfoncement de la sonde 34 sont directement liées aux précipitations.

Nous ne nous étendrons pas plus longuement car le choix de l'échantillon (journées avalancheuses seulement) enlève à ces conclusions un caractère climatologique général, applicable à l'ensemble des journées de l'hiver.

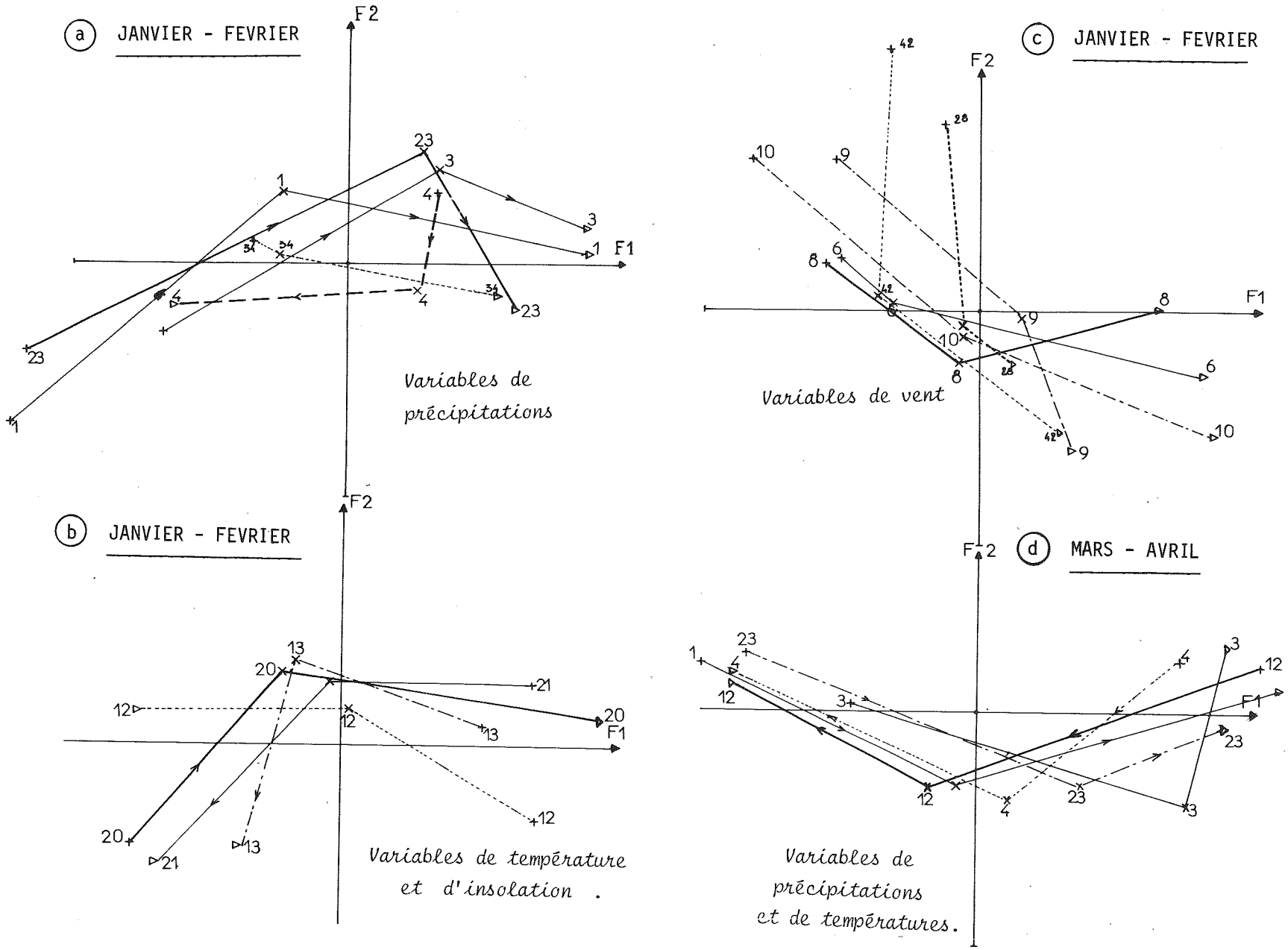
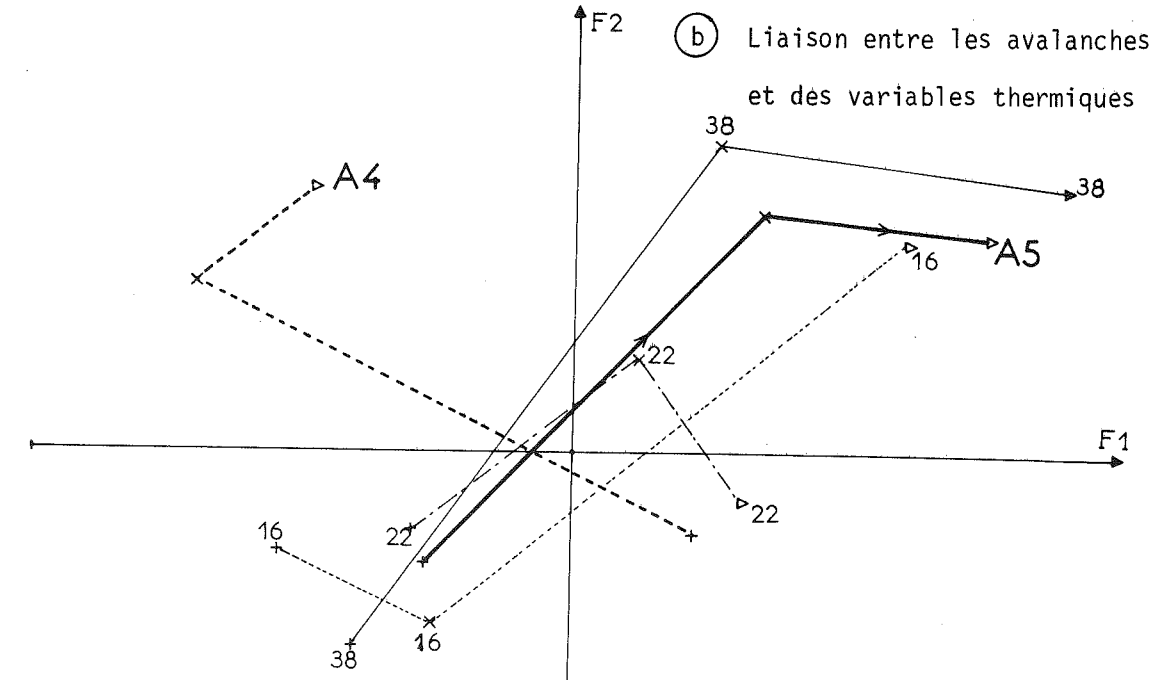


FIGURE II - 9 : Quelques de trajectoires de variables de même nature .

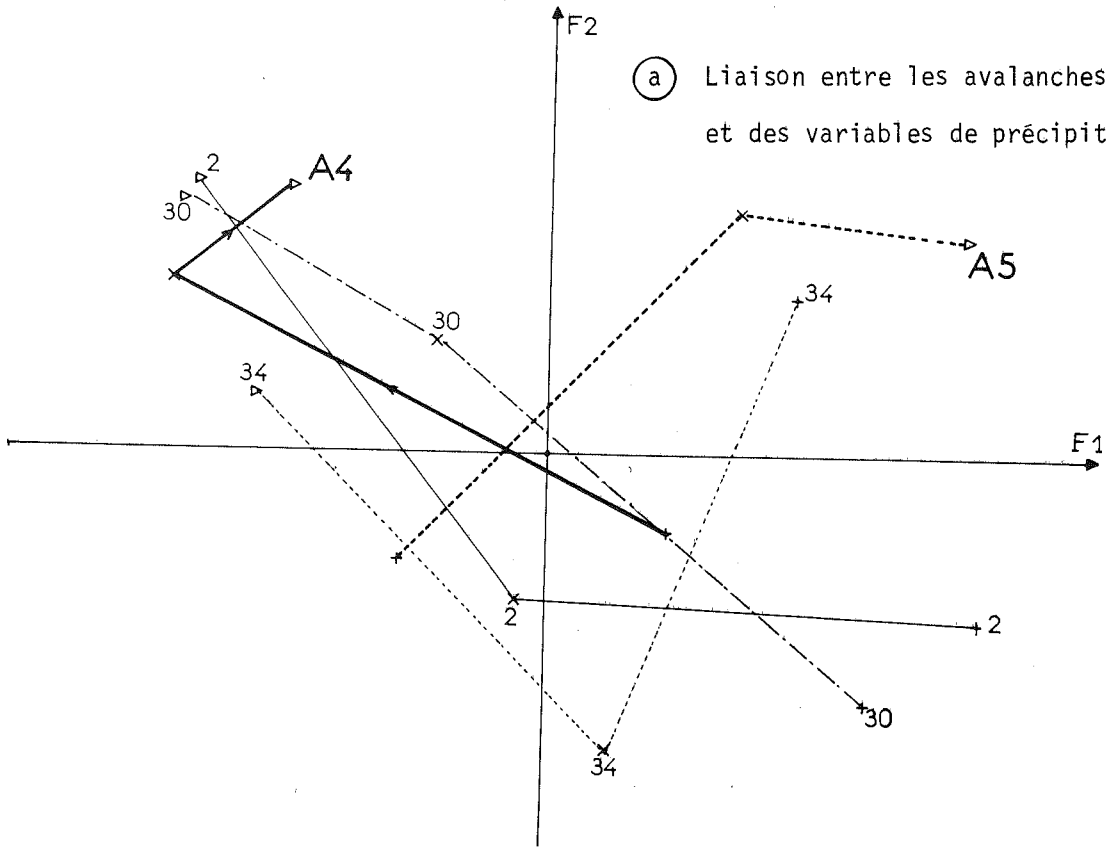
(Les N° correspondent à la table II de la 1ère Partie- p.11)



(b) Liaison entre les avalanches humides et des variables thermiques .

A 4 : Avalanches de neige sèche

A 5 : Avalanches de neige humide



(a) Liaison entre les avalanches sèches et des variables de précipitations .

Figure II-10 : Liaisons entre une variable éxogène et des variables par analyse en correspondance .

On retiendra quand même l'intérêt qu'il y a à visualiser non pas 2 variables prises globalement (cas de l'A.C.P.) mais les classes faible, moyenne ou forte de ces 2 variables de façon à voir si la liaison existe sur tout l'intervalle de variation de ces variables, ou seulement dans une partie de celui-ci.

Enfin, rappelons que cette méthode nous fournit pour les observations un ensemble de coordonnées ("Factor scores") qui présentent certaines propriétés optimales (cf § II-3) et que nous utiliserons ensuite (Vème Partie).

II.4.2. Introduction d'une variable exogène : la variable avalanche

On ajoute au tableau précédent une variable supplémentaire, codée de la même façon et représentant le phénomène à étudier. Dans ce cas l'étude des observations perd un peu de son intérêt (on introduit volontairement l'information que l'on voudrait voir apparaître). Par contre, l'analyse des similarités entre cette variable et les autres peut être instructive. C'est ce que P. CAZES (1976) appelle la régression par l'analyse en correspondance.

Nous avons réalisé de tels essais sur des échantillons un peu particuliers (essentiellement pour des raisons d'encombrement en mémoire) :

Mois de Janvier (période 1961-72)	Mois de Mars
- 57 journées avalancheuses	- 67
- 67 journées sans avalanche	- 48
33 variables "explicatives"	

Un problème apparaît dans le codage de la variable avalanche : si on veut la coder en classes équiprobables pour bénéficier des propriétés correspondantes , il faut se rappeler qu'il y a 1 journée sur 5 ou 7 qui est avalancheuse. Une solution consiste donc à tirer au hasard l'échantillon des journées "normales". On découpe ensuite les journées avalancheuses en 2 classes (1 ou 2 avalanches, $\geqslant$ 3 avalanches).

Les résultats les plus remarquables sont obtenus sur le mois de Mars où l'on a traité séparément les avalanches de neige sèche (A4) et humide (A5) (cf. Fig.II-10).

On visualise très nettement la liaison des premières avec les précipitations récentes, tant que les secondes sont liées à des températures élevées persistantes.

Malheureusement, l'impossibilité de travailler en classes équiprobables en particulier pour les variables avalanches, perturbe beaucoup les trajectoires. Dans le cas contraire, on pourrait analyser dans le plan où ces variables sont le mieux représentées, les autres variables explicatives.

CHAPITRE III

TECHNIQUES DIVERSES

III.1 - Méthode de SAMMON

Il s'agit d'une méthode non linéaire, proposée par J.W. SAMMON (1969) et citée entre autres par DER MEGREDITCHIAN G. (1973). Son but est de donner une représentation à 1, 2 ou 3 dimensions, (en général plane) d'une matrice de distances entre points. Elle s'applique par exemple aux matrices de corrélation conçues comme des matrices de distances dans mais elle peut s'appliquer indifféremment à des variables ou des individus, le point de départ étant la matrice d'interdistances.

III.1.1. Bref exposé

On dispose de P points $X_j \quad j = 1, \dots P$ dans $\mathbb{R}^N$:

$$X_j = X_{0j} = \left\{ x_{1j}, \dots x_{ij} \dots x_{Nj} \right\}$$

La distance de 2 points dans $\mathbb{R}^N$ étant :

$$d^{*2}_{ij} = \sum_{k=1}^{N} \left(x_{ki} - x_{kj} \right)^2$$

A chaque point :

$$X_j \in \mathbb{R}^N \quad \text{on associe} \quad Y_j = \left\{ \begin{matrix} y_{1j} \\ y_{2j} \end{matrix} \right\} \in \mathbb{R}^2$$

et la distance entre les points Y_i et Y_j dans $\mathbb{R}^2$ est notée :

$$d_{ij} = \sum_{k=1}^{2} \left(y_{ki} - y_{kj} \right)^2$$

J.W. SAMMON propose de chercher, dans $\mathbb{R}^2$, la configuration de points Y qui minimise le critère :

$$E = \frac{1}{c} \sum_{i<j} \frac{\left(d^*_{ij} - d_{ij} \right)^2}{d^*_{ij}} \qquad c = \sum_{i<j} d^*_{ij}$$

à l'aide d'un algorithme de descente.

Cet algorithme est le suivant:

On a P points Y donc $2.P$ inconnues $\left\{ y_{1j}, y_{2j} \right\}$. Comme l'origine n'a pas d'importance, le but étant de représenter les interdistances, on peut fixer les coordonnées du 1er point à $\left\{ 0, 0 \right\}$ d'où $2(P-1)$ inconnues. Si de plus on veut bloquer le nuage de points en rotation, on peut imposer une coordonnée (par exemple sur le point le plus éloigné du premier dans les distances initiales) égale à 0, d'où $2P-3$ inconnues. (Dans

la suite du développement, on laissera cependant P dans les sommations pour simplifier). A part cela, la configuration initiale peut être quelconque et, par exemple, générée aléatoirement.

Au bout de la m ème itération, on a :

$$E(m) = \frac{1}{C} \sum_{i<j}^{P} \frac{\left(d_{ij}^{*} - d_{ij}(m) \right)^{2}}{d_{ij}^{*}}$$

avec $$d_{ij}(m) = \left(\sum_{k=1}^{\ell} \left[y_{ki}(m) - y_{kj}(m) \right]^{2} \right)^{\frac{1}{2}}$$

Si on considère la qème coordonnée du point j, SAMMON propose :

$$y_{qj}(m+1) = y_{qj}(m) - MF. \Delta_{qj}(m) \qquad q = 1, \ell$$

avec : $$\Delta_{qj} = \frac{\partial E(m)}{\partial y_{qj}(m)} \Big/ \left| \frac{\partial^{2} E(m)}{\partial y_{qj}(m)^{2}} \right|$$

où MF est un "magic factor", en fait un coefficient empirique $\approx$ 0.3 ou 0.4.

On vérifie que :

$$\frac{\partial E}{\partial y_{qj}} = - \frac{2}{C} \sum_{\substack{i=1 \\ i \neq j}}^{P} \frac{d_{ij}^{*} - d_{ij}}{d_{ij} . d_{ij}^{*}} \left(y_{qj} - y_{qi} \right)$$

et que :

$$\frac{\partial^{2} E}{\partial y_{qj} \partial y_{si}} \begin{cases} \dfrac{2}{C} \left(y_{qj} - y_{qi} \right) \left(y_{sj} - y_{si} \right) \dfrac{1 - d_{ij}^{*}}{d_{ij}^{*} . d_{ij}} & s \neq q \\ \dfrac{2}{C} \left[\left(y_{qj} - y_{qi} \right)^{2} \dfrac{1 - d_{ij}^{*}}{d_{ij}^{*} . d_{ij}} - \dfrac{d_{ij}^{*} - d_{ij}}{d_{ij}^{*} . d_{ij}} \right] & s = q \end{cases}$$

mais SAMMON utilise seulement :

$$\frac{\partial^{2} E}{\partial y_{qj}^{2}} = + \frac{2}{C} \sum_{\substack{i=1 \\ i \neq j}}^{P} \left[\frac{\left(y_{qj} - y_{qi} \right)^{2}}{d_{ij}^{3}} - \frac{d_{ij}^{*} - d_{ij}}{d_{ij}^{*} . d_{ij}} \right]$$

III.1.2. Remarques

(a) Si on considère la méthode d'optimisation, on s'aperçoit qu'elle est un peu simpliste puisqu'elle se ramène à écrire :

$$F(y + \delta y) = F(y) + \delta y \frac{dF}{dy}$$

ou

$$\frac{\partial E(y + \delta y)}{\partial y} = \frac{\partial E(y)}{\partial y} + \delta y . \frac{\partial^{2} E}{\partial y^{2}} = 0$$

d'où δy.

Cela consiste donc à minimiser E , ou à annuler à dérivée, direction par direction, ce qui ne nous mènera que lentement vers le minimum de $E(y_{11}, y_{21}, \ldots, y_{2P})$

De plus, il est prudent, quand on a trouvé δy de ne pas faire le pas complet, d'où le rôle de ce "magic factor".

Pourtant, la méthode fonctionne relativement bien et on peut alors se demander pourquoi.

Si on considère une véritable méthode de gradient, comme la méthode de NEWTON-RAPHSON, on sait qu'elle conduit à résoudre un système linéaire :

$$\begin{bmatrix} \frac{\partial^2 E}{\partial y_{11}^2} & \frac{\partial^2 E}{\partial y_{11}\partial y_{21}} & \frac{\partial^2 E}{\partial y_{11}\partial y_{22}} & \cdots & \frac{\partial^2 E}{\partial y_{11}\partial y_{2P}} \\ \frac{\partial^2 E}{\partial y_{11}\partial y_{21}} & \frac{\partial^2 E}{\partial y_{21}^2} & \frac{\partial^2 E}{\partial y_{12}\partial y_{21}} & \cdots & \frac{\partial^2 E}{\partial y_{21}\partial y_{2P}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 E}{\partial y_{11}\partial y_{2P}} & \frac{\partial^2 E}{\partial y_{21}\partial y_{2P}} & \cdots & \cdots & \frac{\partial^2 E}{\partial y_{2P}^2} \end{bmatrix} \begin{bmatrix} \delta y_{11} \\ \delta y_{21} \\ \cdots \\ \delta y_{2P} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} \\ \frac{\partial E}{\partial y_{21}} \\ \cdots \\ \frac{\partial E}{\partial y_{2P}} \end{bmatrix}$$

Or, les formules donnant les dérivées seconde montrent que, en général, par la présence de la sommation dans les termes diagonaux :

$$\frac{\partial^2 E}{\partial y_{qj}^2} \gg \frac{\partial^2 E}{\partial y_{qj}\partial y_{si}}$$

Si la matrice est diagonalement dominante, la méthode de SAMMON revient à la supposer diagonale et à résoudre le système par la méthode de JACOBI avec une "sous relaxation".

En fait, cette méthode a le mérite de la simplicité mais on peut parfaitement en envisager d'autres et pour notre part, nous avons tenté, après convergence de l'algorithme de SAMMON, d'améliorer encore le critère par une autre méthode.

Des précautions sont à prendre car en dépit de son apparente simplicité, le critère E n'est pas une fonction convexe et la matrice hessienne n'est pas toujours définie positive.

(b) Par contre, on peut s'interroger sur la forme de ce critère. Il ressemble à la moyenne quadratique des erreurs relatives : $\frac{d_{ij}^* - d_{ij}}{d_{ij}^*}$

mais pondérées par le rapport de la distance d_{ij}^* à sa valeur moyenne $\sum_{i<j} d_{ij}^*$ d'où :

$$\sum_{i<j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^{*2}} \times \frac{d_{ij}^*}{\Sigma d_{ij}^*}$$

On aurait pu penser aussi au coefficient de corrélation entre d_{ij} et d^*_{ij} mais ces $\frac{P(P-1)}{2}$ valeurs ne sont pas indépendantes puisque la modification d'une coordonnée se répercute sur (P-1) distances. D'autre part, le sens statistique n'est pas évident.

Dans notre cas, où les matrices de distances proviennent de matrices de corrélation, il faudrait s'inspirer d'un test statistique. Par exemple, D.F. MORRISON (1967, p.152-153) propose un test pour l'égalité des matrices de covariance .

Si N_i est le nombre d'observations avec lequel on a calculé la matrice de covariance S_i (dans notre cas , on peut prendre :

$$N_1 = N_2 = N \qquad S_1 = R_1 = \{r^*_{ij}\}$$

$$S_2 = R_2 = \{r_{ij}\} = \left\{1 - \frac{d_{ij}^2}{2}\right\} \qquad \overline{R} = \frac{1}{2}(R_1 + R_2)$$

le test s'écrit :

$$F = 2.N \log|\overline{R}| - N \log|R_1| - N \log|R_2| = N . \log \frac{|R|^2}{|R_1|.|R_2|}$$

et suit, à un facteur près, une loi de FISHER quand $N \longrightarrow \infty$.

Le problème est que ce critère est difficile à optimiser car il contient des déterminants dans lesquels le rôle des variables y_{qj} n'est plus très simple.

D'autre part, la matrice R_2 n'est plus forcément définie, positive et peut aussi contenir des termes supérieurs à 1.0. On pourrait donc envisager cette méthode comme raffinement de la précédente, et avec des contraintes sur les d_{ij} (≤ 2).

(c) Un dernier problème est apparu lors des premiers essais d'application de cette méthode à des données réelles provenant d'un réseau de mesures hydrométriques : les projections obtenues ressemblaient fortement au plan F2/F3 de l'A.C.P. correspondante,et cela nous a conduit à en chercher la raison .

Or on a vu (Chpa.I-2 de cette partie) que l'A.C.P. du nuage des variables n'était pas tout à fait comparable à l'analyse des observations dans la mesure où l'origine jouait un rôle très particulier.

Dans le cas d'un réseau, où il y a un fort effet de taille, l'axe 1 de $\mathbb{R}^N$ ne cherche pas l'élongation maximale du nuage des P points variables mais plutôt ce qu'elles ont en commun.

Or dans ce cas, on a une valeur propre $\lambda_1 \# P$ et un vecteur propre $V_1 \approx \left\{\frac{1}{\sqrt{P}} \text{-------} \frac{1}{\sqrt{P}}\right\}$ car les coordonnées sont très voisines et il est unitaire. On on sait que l'on peut retrancher l'influence de la première composante de la matrice R en calculant, par déflation (cf aussi IVème Partie, Chap.II-1) :

$$\widetilde{R} = R - \lambda_1 V_1 . V_1^t$$

Mais dans notre cas : $\lambda_1 . V_1 . V_1^t \# \left\{ \frac{P}{\sqrt{P}\sqrt{P}} \right\} = \{1\}$

Or la matrice analysée par la méthode de SAMMON est justement :

$$D = \{ d_{ij}^* \} = \ell \left(\{ 1 \} - R \right) \approx - \ell \widetilde{R}$$

et on comprend pourquoi on retrouve l'allure des projections sur F2/F3 surtout si $\lambda_2 \# \lambda_3 \gg \lambda_4$

Précisons toutefois qu'il s'agit d'une approximation d'autant meilleure que l'effet de taille est fort. D'autre part cette interprétation disparait dans le cas de paquets de variables peu corrélées.

L'intérêt de la méthode est donc de nous présenter une vue globale des relations entre variables, un peu l'équivalent d'une vue en perspective par rapport aux 3 vues classiques. Par contre, la mesure de la qualité de la représentation est difficile à apprécier, à l'inverse des pourcentages d'inertie de l'A.C.P.

Par contre, dans le cas des réseaux, la différence avec la projection dans F2/F3 met en évidence la présence d'informations significatives sur les axes ultérieurs.

III.1.3. Exemples d'applications

On donne d'abord un exemple artificiel où les P points à représenter sont pris dans $\mathbb{R}^3$, le long d'une hélice circulaire d'équation :

$$Z(t) = \frac{\sqrt{2}}{2} (t-1) \qquad X(t) = \cos Z(t) \qquad Y(t) = \sin Z(t)$$

pour des valeurs entières de t = 1, 2, ... 30.

Les valeurs de Z varie de 0 à 20.5 tandis qu'en projection sur XOY les points se répartissent sur un cercle de rayon 1. On donne dans la figure II-11 quelques résultats.

On s'aperçoit que, pour des valeurs voisines du critère à la convergence on obtient des configurations assez différentes. Seule la dernière, qui correspond au meilleur critère obtenu, donne l'image assez régulière que l'on attendait. Il y a donc une sensibilité certaine au choix de la configuration initiale, qui nécessite des essais répétés pour tenter d'améliorer l'optimum, ce qui rend cette méthode très coûteuse.

Enfin, on donne un exemple simple de l'application que nous voulons en faire. Il concerne 13 stations de mesures de débits annuels dans les Pyrénées (il nous a été fourni par M. DUBAND , E.D.F.-D.T.G.). Les corrélations ont été calculées sur 24 années consécutives. On donne successivement (Fig.II-12) :

- les résultats de l'A.C.P. sur valeurs normées : tableau des corrélations des variables avec le 1er axe (qui met en évidence un fort effet de taille) et représentation dans F2/F3
- la meilleure représentation obtenue par la méthode de SAMMON, qui reproduit pratiquement la précédente.

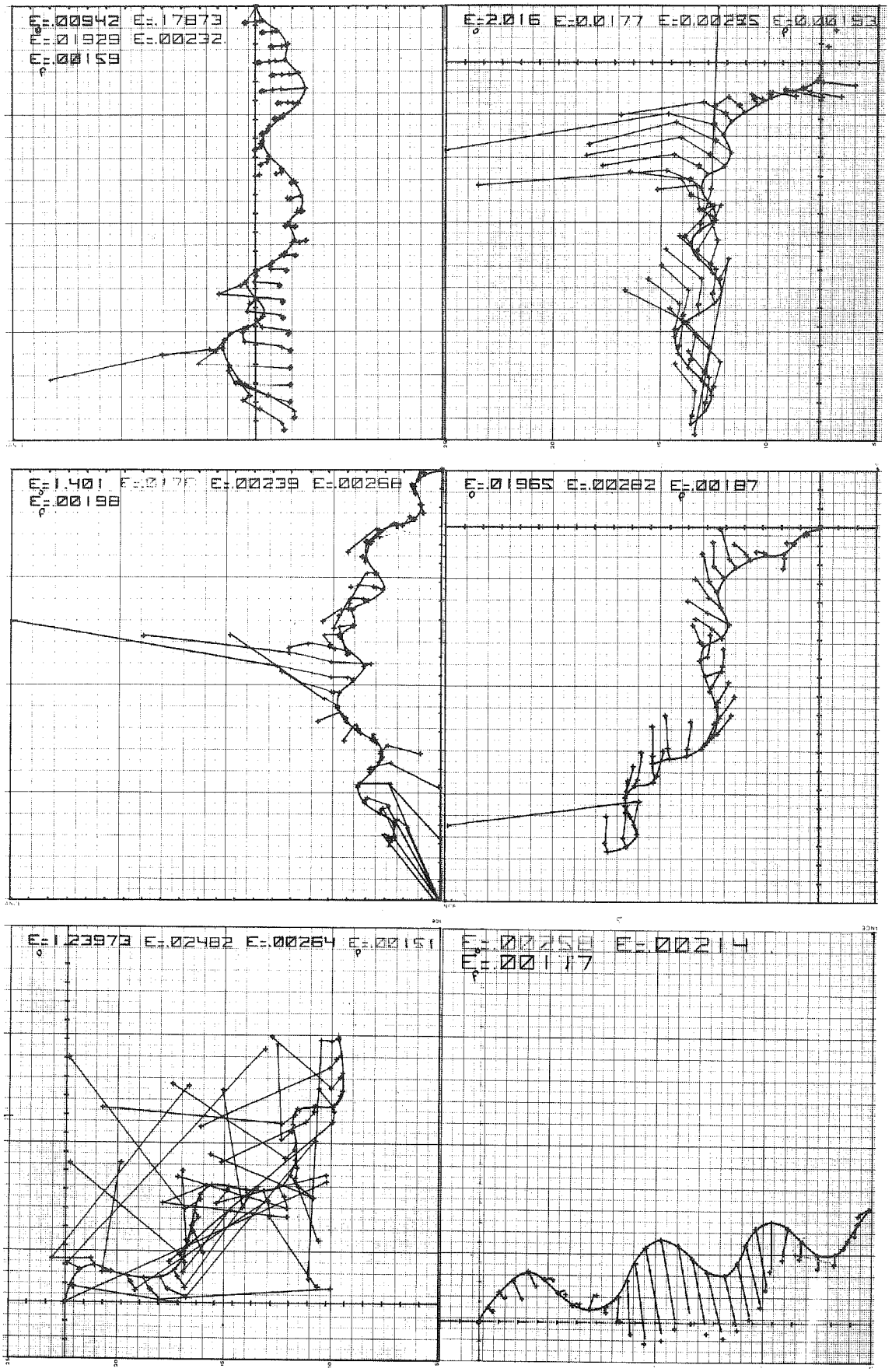


FIGURE III-11 : Méthode de SAMMON - Quelques exemples de courbes planes obtenues à partir d'une hélice dans R^3.

(E_0= critère associé au semis de points initial - E_f= critère final)

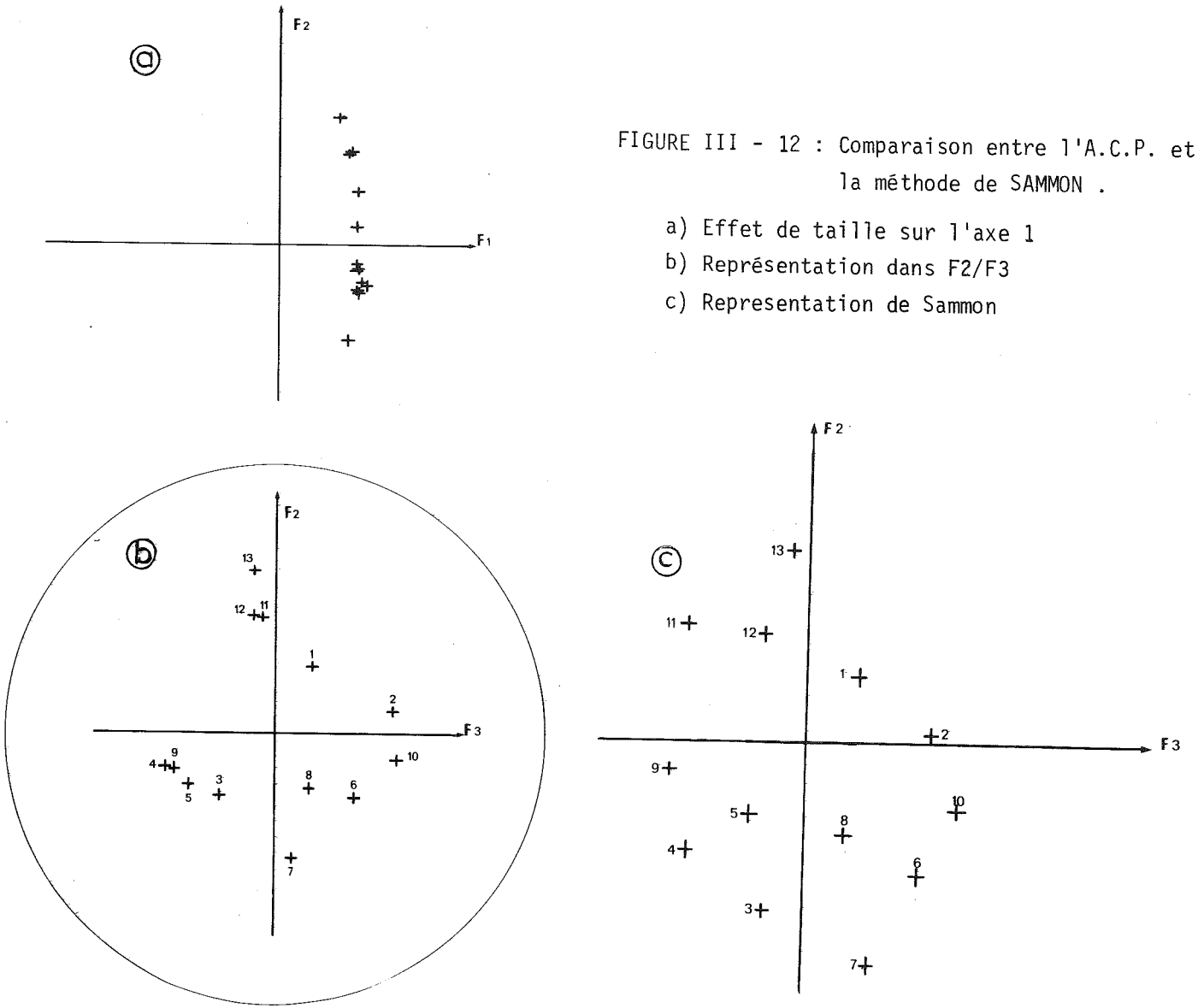


FIGURE III - 12 : Comparaison entre l'A.C.P. et la méthode de SAMMON .

a) Effet de taille sur l'axe 1
b) Représentation dans F2/F3
c) Representation de Sammon

III.2 - Méthode d'ANDREWS

Contrairement à la méthode de SAMMON, cette méthode n'a pas pour point de départ les représentations géométriques que l'on trouve dans les méthodes linéaires d'analyse des données.

Elle consiste à représenter un point d'un espace multidimensionnel (individu caractérisé par p variables,ou variable caractérisée par N observations) par la courbe du plan associée à la fonction d'une variable :

$$f(t) = x_1 \sin t + x_2 \cos t + x_3 \sin 2t + x_4 \cos 2t + \cdots$$

ou

$$\frac{1}{2} x_1 + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + \cdots$$

prise sur $-\pi \leq t \leq +\pi$

D.F. ANDREWS (1973) en donne quelques propriétés. Dans notre cas, nous envisagerons surtout l'application à des variables :

$$X_{0j} = \{x_{1j}, x_{2j} \cdots x_{Nj}\}$$

$$\longrightarrow f_j(t) = x_{1j} \sin t + x_{2j} \cos t + x_{3j} \sin 2t + \cdots + x_{Nj} \begin{cases} \sin \\ \text{ou} \\ \cos \end{cases}$$

et si on définit la distance entre 2 fonctions par :

$$\| f - g \| = \int_{-\pi}^{+\pi} [f(t) - g(t)]^2 \cdot dt$$

alors

$$\| f_j - f_k \| = \pi \sum_{i=1}^{N} (x_{ij} - x_{ik})^2$$

On remarquera que l'approche de ANDREWS est initialement analytique et se rattache à la transformation ou plus exactement à l'approximation de Fourier, dont les observations x_{ij} seraient les coefficients. Pourtant, l'interprétation des x_{ij} , $x_{i+1\,j}$ en terme d'énergie associée à une harmonique n'est pas très intéressante car, pour la variable aléatoire X_j, ce sont plutôt ses moments qui interviennent dans les fonctions caractéristiques ou spectrales.

Par contre, l'originalité consiste à ne plus projeter la variable dans un espace vectoriel "simple" $\mathbb{R}^N$, mais dans un espace de fonctions (dont $\mathbb{R}^N$ est un cas particulier si on prend comme fonction de base $e_i(t)$ la fonction impulsion égale à 1 pour l'observation i et à 0 ailleurs), et à représenter un élément (variable X_j) non plus par un point géométrique mais par une fonction définie sur $[-\pi, +\pi]$

Pourtant dans l'usage qu'il en fait D.F. ANDREWS revient à l'interprétation dans $\mathbb{R}^N$ puisqu'il considère, pour comparer plusieurs variables X_{01} , $X_{02} \cdots X_{0j} \cdots$ les valeurs de $f_1(t) \cdots f_j(t) \cdots$ à une abscisse donnée t_0.

Dans ce cas en effet, on peut interpréter $f_j(t_o)$ comme la projection, dans $\mathbb{R}^N$, de la variable j sur l'axe (sin t_o , cos t_o , sin $2t_o$, cos $2t_o$ ...) et plus le module de $f_j(t_o)$ est grand, plus X_{oj} est proche de cet axe.

On peut alors voir les variables qui sont fortes en même temps pour les mêmes valeurs de t et faire ainsi une typologie.

Malheureusement, cette interprétation présente 2 lacunes majeures :

- Dans $\mathbb{R}^N$ les variables centrées appartiennent à un hyperplan de vecteur directeur (1,1 ..... 1) qui n'est représenté par aucune valeur de t .
- Mais surtout le vecteur (sin t , cos t , sin $2t$ .....) ne balaie pas tout l'espace $\mathbb{R}^N$ mais seulement une courbe particulière de celui-ci. Cela se vérifie aisément à 3 dimensions.

Enfin et surtout, l'oeil interprête assez aisément la proximité de deux points et la structure d'un ensemble de P points. Il interprête déjà moins bien la proximité de 2 courbes et très difficilement un ensemble de P courbes.

C'est pourquoi nous verrons que l'on peut utiliser cette procédure à titre indicatif d'une éventuelle typologie mais sans qu'un résultat négatif prouve l'absence absolue de groupements.

Enfin, l'utilisation matérielle de cette méthode nécessite en pratique un programme interactif. On en donne une illustration dans la IIIème Partie (§ II.1.3).

III.3 - Méthode des faces de CHERNOFF

Nous citons pour mémoire cette méthode dûe à CHERNOFF et qui part d'un principe assez voisin. Etant donné un vecteur multidimensionnel (par exemple, une variable) on associe à chaque composante une fonction qui sera par exemple :

composante k : taille de la bouche dans un visage

composante k+1 : inclinaison des yeux

etc...

Et à chaque vecteur correspondra une face que l'on pourra comparer aux autres.

Il s'agit là encore d'une visualisation, très éloignée des techniques précédentes mais qui utilise la capacité du cerveau humain à mémoriser et à comparer 2 visages.

La programmation de cet algorithme étant assez lourde et nécessitant des écrans graphiques sophistiqués, nous en avons mis en oeuvre une version simplifiée (cf. un exemple de sorties en figure II-13).On trouvera un exmple d'application dans U. MAAG (1978).

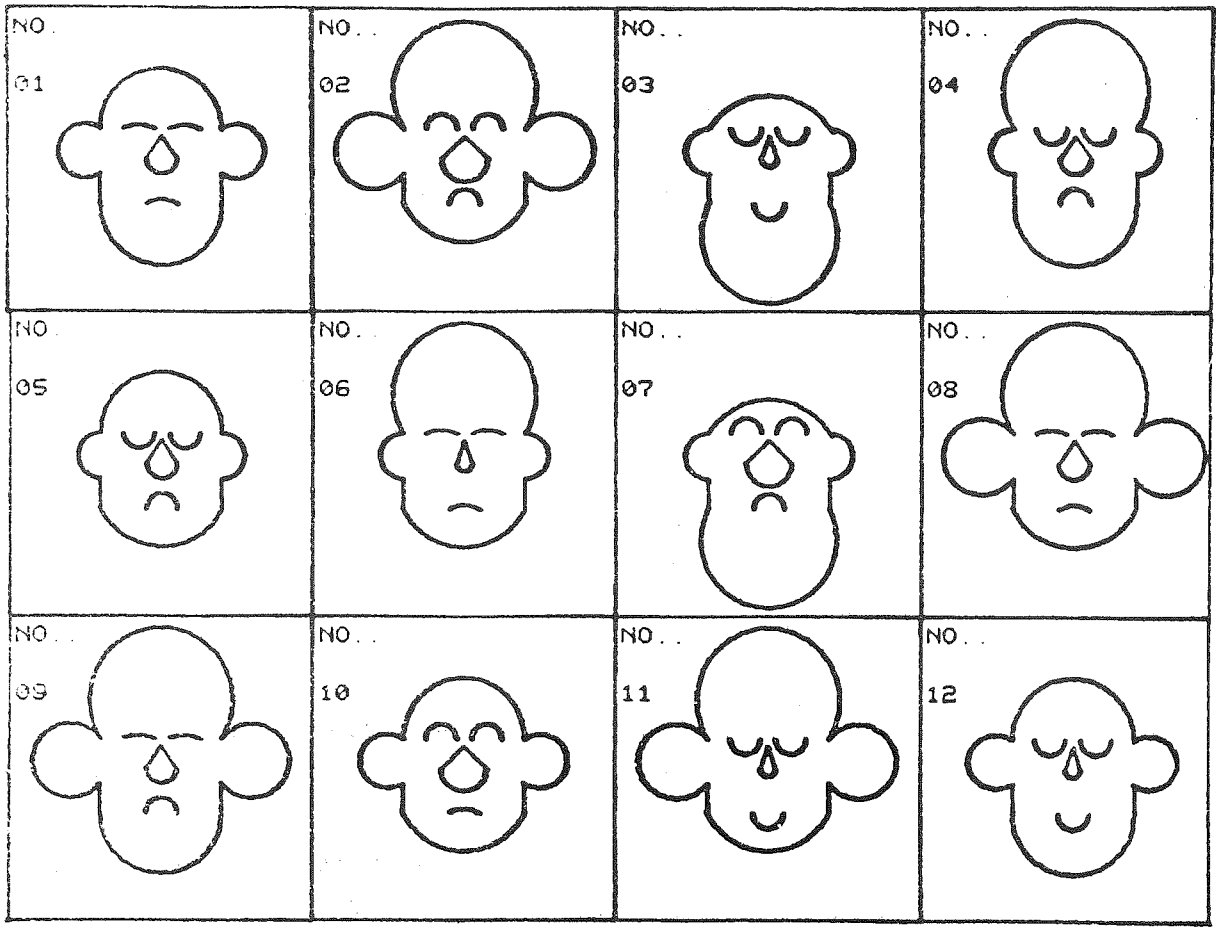


FIGURE II - 13 : Exemple de " faces de Chernoff".

CHAPITRE IV

ANALYSE FACTORIELLE DISCRIMINANTE

Un cas particulier de visualisation en présence d'information exogène

IV.1 - Position du problème et notations

On dispose, pour chacun des N individus caractérisés par P variables, d'une information supplémentaire : son appartenance à une classe (et une seule) parmi G.

Le problème géométrique consiste à trouver la meilleure représentation des G groupes dans un espace de faible dimension.

L'échantillon est traité ici comme une population finie, sans hypothèse statistique sur les distributions de points au sein des groupes et en utilisant l'ensemble des variables proposées. Les aspects décisionnels (affectation d'un nouvel individu, sélection de variables explicatives, etc .., seront abordés dans la Vème partie).

Nous présentons d'abord les notations, reprises de ROMEDER (1973) mais en les précisant.

On considère N individus caractérisés par P variables ($x_{i\ell}$ = la valeur de X_ℓ pour l'individu i) et répartis en G classes Y_k $(k = 1 \cdots G)$

A chaque individu est affecté un poids, ou une masse, égal à 1/N , ce qui nous conduit à considérer non plus des points moyens mais des barycentres.

Individu	Groupe
Masse : $x_i \in X \cdot m_{x_i} = \frac{1}{N} \quad m_X = 1$	Masse : classe Y_k d'effectif n_k $m_{Y_k} = \frac{n_k}{N} \qquad m_Y = \sum_k m_k = 1$
Centre de gravité : $\overline{x}_{\cdot j} = \frac{1}{m_X} \sum m_{x_i} \cdot x_{ij}$	Centre de gravité :
On vérifie que : $\overline{x}_{\cdot j} = \overline{y}_{\cdot j}$	- de la classe k : $\overline{y}_{kj} = \frac{1}{m_{Y_k}} \sum \left(m_{x_i} \cdot x_{ij} \mid x_i \in Y_k \right)$
	- de l'ensemble des groupes : $\overline{y}_{\cdot j} = \frac{1}{m_Y} \sum_k m_{Y_k} \cdot \overline{y}_{kj}$

Pour les variances, on regarde non pas la variance des variables (directions canoniques de $\mathbb{R}^P$) mais la variance d'une combinaison linéaire (ou d'une direction U de $\mathbb{R}^P$) $\longrightarrow$ $\vec{u}^t = \{u_1 \cdots u_P\}$

Variable U : $U_i = u^t \cdot X_{iV}$

de moyenne $\overline{U} = u^t \cdot \overline{X}_{\cdot V}$

Variance :

$$var_X(U) = \sum_{i=1}^{N} m_{x_i} \left\{ u^t (X_{i\cdot} - \overline{X}_{\cdot})' \right\}^2$$

= moment d'inertie du nuage complet sur la direction U.

Variance au sein du groupe Y_k :

$$var_{Y_k}(U) = \sum m_{Y_k} \left\{ u^t \cdot (X_{i\cdot} - \overline{X}_{\cdot}) \right\}^2 \qquad x_i \in Y_k$$

=moment d'inertie du groupe k .

Variance des groupes :

$$var_Y(U) = \sum_{k=1}^{G} m_{Y_k} \cdot \left\{ u^t \cdot (\overline{Y}_{kV} - \overline{Y}_{\cdot V}) \right\}^2$$

=moment d'inertie des groupes.

On démontre que :

$$var_X(U) = var_Y(U) + \sum_{k=1}^{G} var_{Y_k}(U)$$

qui n'est autre que le théorème de HUYGHENS.

Matrices de covariance :

totale T :

$$t_{lj} = \sum_{i=1}^{N} m_{x_i} (x_{il} - \overline{x_{\cdot l}})(x_{ij} - \overline{x_{\cdot j}})$$

intragroupe : W (within)

$$w_{lj} = \sum_{k=1}^{G} \sum_{i=n_k}^{n_k + n_k} m_{x_i} (x_{il} - y_{kl})(x_{ij} - y_{kj}) \qquad x_i \in Y_k$$

$$n_k = \sum_{\delta=1}^{k-1} n_\delta$$

interclasse : B (between)

$$b_{lj} = \sum_{k=1}^{G} m_{Y_k} (y_{kl} - \overline{y_{\cdot l}}) \cdot (y_{kj} - \overline{y_{\cdot j}})$$

Et le théorème de HUYGHENS s'écrit ici :

$$T = W + B$$

tandis que les variances selon la direction U vérifient :

$$var_X(u) = u^t . T . u$$

$$\sum_{k=1}^{G} var_{Y_k}(u) = u^t . W . u$$

$$var_Y(U) = u^t . B . u$$

On remarquera que, si $N > P$ on a en général $rang(T) = P$

Ceci est vrai aussi pour W , sauf si une ou plusieurs variables sont constantes au sein de chaque groupe. On notera que W n'est pas la matrice associée à un groupe mais une moyenne.

Par contre B ne peut être définie positive que si $G \geqslant P$. En effet les G points Y_k ne peuvent engendrer dans $\mathbb{R}^P$ qu'un sous-espace de dimension G - 1 (variété linéaire affine) et la forme quadratique $u^t . B . u$ sera nulle pour toute direction orthogonale à celui-ci.

Remarque sur les notations :

On a défini t_{lj} par : $t_{lj} = \frac{1}{N} \sum_{i=1}^{N} (x_{ij} - \bar{x}_{.j})(x_{il} - \bar{x}_{.l})$

mais en général, pour des considérations probabilistes, on prend $\frac{1}{N-1}$

De même pour W on prendra :

$$w_{lj} = \frac{1}{N-G} \sum_{k=1}^{G} \sum_{i=n_k+1}^{n_k+n_k} (x_{ij} - y_{kj})(x_{il} - y_{kl})$$

Dans ce cas les matrices T , W et B ne vérifient plus exactement la relation : $T = W + B$

Ce sont les matrices de "produits croisés" qui la satisfont. Mais quand $N \nearrow$ cela devient équivalent. Cela vient de ce que l'on mélange des préoccupations géométriques sur un nuage fin de points pesants avec des soucis d'inférence statistique et des modèles probabilistes considérant le nuage comme un échantillon tiré d'une population infinie.

IV.2 - Première formulation de l'analyse factorielle discriminante

IV.2.1. Recherche des axes discriminants

(a) Par analogie avec l'analyse en C.P. on va chercher ici l'axe qui maximise la dispersion entre groupes.

Une première façon consiste à déterminer l'axe factoriel qui maximise la variance interclasse par rapport à la variance intra-classe. En effet, la variance entre classes a d'autant plus de sens qu'elle est forte par rapport à la variance dans la classe et cela revient au fond à chercher la variable u qui a le meilleur t de STUDENT.

Cela s'écrit :

$$\frac{u^t . B . u}{u^t . W . u}$$

Compte tenu de : $T = W + B \implies u^t . B . u + u^t . W . u = u^t . T . u$

on voit aisément que cela revient à maximiser

$$\frac{u^t . B . u}{u^t . T . u}$$

On voit que si on remplace u par αu le résultat est inchangé, donc on peut poser la contrainte $u^t . T . u = 1$, d'où le problème :

maximiser $u^t . B . u$ sous la contrainte $u^t . T . u = 1$

Le multiplicateur de Lagrange λ nous conduit à chercher u tel que :

$$u^t . B . u - \lambda\, u^t . T . u \quad \max \implies B u = \lambda . T u$$

$$T^{-1} . B\, u = \lambda\, u$$

Le premier axe factoriel u_1 est le premier vecteur propre de $T^{-1} . B$ associé à la plus grande valeur propre λ_{T_1}.

On démontre d'ailleurs que ces valeurs propres sont ≥ 0 et < 1.

De même on cherchera le 2ème axe factoriel u_2 associé à λ_{T_2}, etc...

On peut vérifier que λ_{T_1} est la fraction de la variance totale le long de l'axe u_1, $u_1^t . T . u_1$, qui est due à la différence entre classes $u_1^t . B . u_1$. On l'appelle le pouvoir discriminant de l'axe 1 (et il varie entre 0 et 1).

(b) Dans la plupart des programmes existants et pour des raisons que l'on verra dans la 4ème partie, on utilise plutôt la matrice W que la matrice T. En effet, l'hypothèse sous-jacente, à savoir que les groupes ne diffèrent que par leur centre de gravité mais non par leur dispersion, fait que W matrice de dispersion des individus au sein d'un groupe quel qu'il soit, est plus caractéristique de la dispersion intrinsèque des individus que T.

On est conduit alors à chercher v_j tel que :

$$\frac{v_j^t . B . v_j}{v_j^t . T . v_j} \text{ max.} \Longrightarrow v_j$$

. vecteur propre de $W^{-1} . B$

. sous la contrainte $v_j^t . W . v_j = 1$

L'utilisation de $T = W + B$ et des contraintes respectives sur u_j et v_j permet de montrer que :

$$T^{-1} . B . v_j = \frac{\lambda_{W_j}}{1 + \lambda_{W_j}} v_j$$

$\Longrightarrow$ u_j et v_j sont tous les 2 vecteurs propres de $T^{-1}B$ ou $W^{-1}B$ mais avec des normes respectives différentes.

IV.2.2. Structure des facteurs et interprétation

On a obtenu dans $\mathbb{R}^p$ un nouveau système d'axes u_j (dont seuls les premiers ont des valeurs propres associées $\neq 0$) auxquels correspondent de nouvelles variables dites discriminantes FD_j, dont les observations sont :

$$fd_{ij} = X_{iV} . u_j \Longrightarrow FD_{0j} = X_{0V} . u_j$$

Et comme en A.C.P. il est intéressant de déterminer la structure des facteurs, c'est-à-dire les corrélations de cette variable FD_j avec les variables initiales, soit le vecteur :

$$S_j = \left\{ r\left(FD_j, X_k\right) \right\} = \frac{1}{N} X_{0V}^t FD_{0j} .$$

Or les variables X et FD étant centrées mais pas réduites, il importe donc de les normer.

Soit D_σ^{-1} la matrice diagonale de terme : $d_{ll} = \frac{1}{\sigma_{xl}}$

On sait aussi que la variance de FD_j est $\theta_j = \frac{1}{N} u_j^t . T . u_j$

d'où la matrice diagonale : $\Delta_\theta^{-\frac{1}{2}}$ de terme $\delta_{ll} = \frac{1}{\sqrt{\theta_l}}$

La structure S_j s'écrit alors :

$$S_j = \frac{1}{N} D_\sigma^{-1} . X_{0V}^t . X_{0V} . u_j . \Delta_\theta^{-\frac{1}{2}}$$

et on peut obtenir tous les facteurs discriminants :

$$S = \{ S_1, S_2, \ldots S_{G-1} \} = D_\sigma^{-1} . T . U . \Delta_\theta^{-1/2}$$

avec $U = \{ u_1, u_2, \ldots, u_j, \ldots u_{G-1} \}$

Dans le cas où l'on a choisi pour u_j les vecteurs de $T^{-1}B$, cela se simplifie car $u_j^t . T . u_j = \theta_j = 1 \Longrightarrow S = D_\sigma^{-1} . T . U$

Par contre, d'autres programmes (BMD 07 M en particulier) calculent les FD_j à l'aide du vecteur v_j, et il faut calculer : $\theta_j = v_j^t . T . v_j$ sachant que $v_j^t W . v_j = 1$

On montre encore, à l'aide des relations entre u_j, v_j, T et W, que $\theta_j = 1 + \lambda_{Wj}$

Remarques. Si on choisit la base $V = \{ v_1, v_2, \ldots v_{G-1} \}$ la variance des points, sur l'ensemble des individus est $\theta_j = 1 + \lambda_{Wj}$ donc > 1 sauf pour les axes non discriminants $\lambda_{W\ell} = 0$ où elle vaut 1.

Ceci provient du fait que dans cette projection, avec $v_j^t . W . v_j = 1$ les domaines d'équidensité de chaque groupe sont des cercles.

Cela est souvent faux dans la pratique où les groupes pris séparément ont souvent une allure elliptique mais il faut rappeler que W est une moyenne.

Enfin, l'interprétation des facteurs obtenus par BMD n'est pas immédiate car le programme fournit les v_{jk} mais pas directement la structure δ_{jk}.

IV.3 - Autres présentations de l'analyse factorielle

IV.3.1. L'A.F.D. comme cas particulier d'une A.C.P.

Nous donnons brièvement un résultat rarement présenté (cf ULMO 1973, CAILLEZ et al 1973) et qui peut aider à l'interprétation de l'analyse factorielle discriminante :

"L'A.F.D. se ramène à l'analyse en composantes principales du nuage des centres de classe Y_k affectés de leur masse m_{Y_k} ".

Pour cela, on peut procéder en 2 étapes :

- d'abord effectuer l'A.C.P. non normée de l'ensemble des individus, ce qui rend le nuage circulaire .
- calculer les barycentres de chaque groupe et analyser leur nuage en prenant en considération leurs masses respectives. Ceci se ramène à chercher les éléments propres de B dans un espace préalablement transformé par T^{-1}.

On en donne quelques détails en annexe.

IV.3.2. Analyse discriminante cas particulier d'une analyse canonique

On trouvera une présentation de l'analyse canonique dans LEBART et FENELON (1973) qui sera à nouveau évoquée dans la 3ème partie. Elle consiste en fait à chercher parmi 2 ensembles de variables Z_{0V} et X_{0V} sur N observations la combinaison linéaire des q variables de Z la mieux correllée avec la combinaison linéaire des P variables de X.

Dans ce cas particulier, on montre que si X est la matrice des données explicatives centrées, et si on construit Z comme une matrice d'indicatrices :

$$\underset{(N\times G)}{Z} = \{z_{ij}\} \qquad z_{ij} = \begin{cases} 1 & \text{si l'individu appartient au groupe} \\ 0 & \text{sinon} \end{cases}$$

donc non centrée.

L'analyse du tableau :

$$V = \frac{1}{N}\begin{vmatrix} X^t.X & X^t.Z \\ Z^t.X & Z^t.Z \end{vmatrix} = \begin{vmatrix} V_{xx} & V_{xz} \\ V_{zx} & V_{zz} \end{vmatrix}$$

au sens de l'analyse canonique conduit à chercher un vecteur u_1 de $\mathbb{R}^p$ tel que :

$$V_{xz}.V_{zz}^{-1}.V_{zx}.u_1 = \beta_1^2 V_{xx} u_1$$

On démontre simplement que : $V_{zz} = \frac{1}{N} Z^t.Z$

est diagonale et d'élément : $V_{zz}{}_{kk} = \frac{n_k}{N}$

que de même : $V_{xz}{}_{jk} = \frac{n_k}{N}\left(y_{kj} - x_{.j}\right)$

et que finalement :

$$V_{xz}.V_{zz}^{-1}.V_{zx} = B$$ matrice de var.-covar. intergroupe

Donc le vecteur u_1 cherché vérifie :

$$B u_1 = \beta_1^2 . T . u_1$$

ou encore :

$$T^{-1}.B.u_1 = \beta_1^2 . u_1$$

On voit immédiatement que c'est le problème résolu précédemment. Au niveau de l'interprétation, on en déduit que : $\lambda_{T_1} = \beta_1^2$

s'interprète comme le carré d'une corrélation canonique.

Dans $\mathbb{R}$ la variable canonique est la variable discriminante déjà obtenue :

$$F_{01} = X_{0V}.u_1$$

De même dans $\mathbb{R}^G$, espace des indicatrices, on trouve un vecteur s_1 lié à u_1 par

$$s_1 = \beta_1 . V_{xz}^{-1} . T . u_1$$

Ce vecteur définit la combinaison linéaire de variables non centrées Z_k la mieux corrélée aux variables X_j.

Enfin, si on se réfère à ce que l'on a vu en analyse des correspondances en II.2.3 on a vu que celle-ci se ramène à l'analyse canonique d'une matrice d'incidence sur 2 classes de modalités, tandis que l'analyse canonique classique s'applique sur 2 ensembles de variables. L'analyse discriminante est donc un cas intermédiaire où l'on a d'un côté une matrice d'incidence, de l'autre une matrice de données habituelle.

A ce sujet, l'article de J.M. BOUROCHE et G. SAPORTA (1978) donne une idée simple et claire des liens des diverses méthodes avec l'analyse canonique.

L'interprétation géométrique n'est pas très riche mais par contre, elle permet de ramener encore plus l'analyse discriminante aux autres méthodes d'analyse de données, en considérant les variables d'appartenance à des groupes comme de simples variables supplémentaires caractérisant l'individu, et en autorisant le traitement global de tous les individus indépendamment des groupes.

IV.4 - Applications (cf Vème Partie, Chap. III)

" Μηδὲν ἄγαν "

" Rien de trop..."

Les sept Sages
(Temple de Delphes.)

TROISIEME PARTIE

DIMENSIONALITE ET REDONDANCES .

(OPTIMISATION DES RÉSEAUX)

Dans un premier chapitre, nous tenterons de répondre à la question : "dans un ensemble de P variables, combien y-a-t-il de facteurs réellement indépendants à prendre en considération ?". Si l'on se limite aux analyses linéaires, cela revient à déterminer le nombre k de facteurs significatifs dans une A.C.P., problème déjà évoqué dans la IIème Partie. Nous verrons qu'en l'absence d'une réponse théorique satisfaisante, on peut quand même, en recoupant diverses méthodes heuristiques, proposer une valeur de k en général très inférieure à P .

On se pose alors la question de savoir s'il n'est pas possible d'abandonner certaines variables, et lesquelles, en réduisant au minimum la perte d'information. C'est l'objet des chapitres II et III.

Ce problème se pose parfois pour des raisons théoriques (cf les modèles à boules dans la Vème Partie) mais le plus souvent pour des raisons pratiques : impossibilité dans une étape de l'étude de considérer plus de $l < P$ variables. Un cas particulier où cette réduction est une fin en soi est celui de l'optimisation des réseaux de mesures : "quelles stations peut-on cesser d'exploiter tout en minimisant la perte d'information correspondante ?"

Sans lui donner une réponse définitive, nous essaierons d'en apporter quelques éléments.

Dans ces deux chapitres, nous définirons en général la ressemblance de 2 variables à l'aide de leur seul coefficient de corrélation linéaire, mais ce n'est ni la seule, ni toujours la meilleure mesure d'association possible.

Quant à la redondance entre 2 variables, utilisée au chapitre III, on peut la définir comme le pourcentage de variance de l'une qui est expliqué par l'autre. C'est donc le carré du coefficient de corrélation et nous généraliserons cette notion au chapitre III.

Le chapitre IV montre l'application de ces techniques aux variables explicatives pour la prévision d'avalanche, et surtout à un réseau de mesures de précipitations, extrait de celui décrit dans la Ière Partie.

CHAPITRE I

DIMENSIONNALITE D'UN ENSEMBLE DE VARIABLES

I.1 - Première approche de la dimensionnalité. Recherche du nombre de facteurs significatifs dans une A.C.P.

I.1.1. Règles empiriques et méthodes heuristiques

On utilise en général une A.C.P. pour détecter le nombre ℓ de "facteurs significatifs" contenus dans un ensemble de P variables ($P \gg \ell$). Le but est de n'utiliser ensuite que ces ℓ variables principales (ou toutes variables déduites de celles-là par des rotations orthogonales) et de négliger les autres, considérées comme des bruits. (Rappelons que ces ℓ composantes sont, à une rotation orthogonale près, les ℓ combinaisons linéaires des P variables de départ qui extraient, dans celles-ci, le maximum de variance). Le point-clé consiste donc à déterminer ℓ .

ⓐ Un certain nombre de "règles" ont été proposées, dont la plus connue, due à KAISER (1966, cité dans COOLEY et LOHMES, 1971, p.104), considère comme facteurs significatifs, ceux dont les valeurs propres associées sont $\geqslant 1.0$.

(Quand le nombre d'observations N est important, on relâche un peu cette contrainte en prenant ℓ tel que $\lambda_\ell \geq 0.8$ mais cette dernière pratique est en fait sujette à caution, cf paragraphe suivant).

Ce seuil de 1.0 correspond au souci de ne pas ignorer une variable X_k qui serait strictement indépendante des $P-1$ restantes, ce qui se traduirait dans la matrice de corrélation théorique, par une ligne et une colonne de 0 (à l'exception du terme diagonal) d'où une valeur propre théorique égale à 1.

Or on sait que, si ρ_{kj} , $j = 1, 2 \ldots k-1, k+1 \ldots P$ est nul dans la population, les estimations r_{kj} faites sur N observations, sont telles que $r_{kj} \longrightarrow 0$ quand $N \nearrow$ ce qui justifie cette règle.

Dans la réalité, le problème est différent. On ne dispose pas de la matrice de corrélation théorique mais d'une estimation. Et, on sait que si la valeur théorique d'une corrélation est $\rho = 0$, la valeur estimée sur un N-échantillon est en général $\neq 0$ (la variable $t = r\sqrt{\frac{N-2}{1-r^2}}$ suit une loi de Student à $N-2$ degrés de liberté, mais quand N est assez grand, la loi de r est voisine d'une loi normale dont l'écart-type dépend de N (cf VIALAR, 1956, Tome III, p.55 et suivantes)).

Et dans notre cas, une variable X_k théoriquement indépendante des autres variables, donnera, sur un N-échantillon, un ensemble de $P-1$ corrélations r_{kj} en général différentes de 0. Cela signifie qu'une part de la variance de X_k , certes

faible, est expliquée par les autres variables. En général, X_k restera associée à un seul facteur, mais de valeur propre $\lambda_k < 1.0$. On conçoit donc que si N augmente, on se rapporche de l'espérance mathématique $\lambda_k = 1$, ce qui devrait plutôt conduire à ramener le seuil vers 1.0 quand N augmente.

On verra ainsi en I.1.2 que l'écart-type d'échantillonnage de la valeur propre l_k autour de son espérance λ_k augmente en $\frac{1}{\sqrt{N}}$ et par conséquent, une estimation $l_k = 0.8$ peut aussi bien venir de $\lambda_k = 1$ que de $\lambda_k = 0.6$. Et si on veut à tout prix éviter d'introduire des variables de faible variance (bruits), mais au risque d'en négliger de significatives, il faut effectivement relever le seuil : c'est ainsi que l'on peut interpréter la règle de KAISER.

(b) Pour notre part, nous proposons une méthode qui permet d'adapter automatiquement la règle de KAISER à la taille N de notre échantillon. Pour cela, on génère une $(P+1)$ème variable, indépendante des P autres, et on regarde le rang de sortie du facteur principal qui lui est associé.

Pour ce faire, on aurait pu générer directement les $r_{j\,P+1}$ par tirages dans la loi de r (associée à $\rho = 0$ et à une taille N d'échantillons) et construire directement la ligne et la colonne supplémentaire de R_{P+1}.

Malheureusement, les coefficients de corrélation d'une $(P+1)$ème variable avec P variables fixées sont soumis à des contraintes supplémentaires. Si on se place dans $\mathbb{R}^N$ où l'on a déjà 2 variables X_{01} et X_{02} correlées à r_{12}, on peut générer r_{13} par exemple, mais alors r_{23} n'est plus quelconque et doit être compris entre :

$$\cos\left(\arccos r_{13} + \arccos r_{12}\right) > r_{23} > \cos\left(\arccos r_{13} - \arccos r_{12}\right)$$

Ceci représente la contrainte pour que R_{P+1} reste semi-définie positive (c'est-à-dire de la forme $\frac{1}{N} X^t . X$). Comme il n'est pas simple de générer les corrélations $r_{P+1,j}$ en respectant ces contraintes, on préfère travailler "en amont" du calcul de ces P corrélations, en adjoignant, au N échantillon de P variables, une $(P+1)$ème générée aléatoirement dans une loi normale $\mathcal{N}(0,1)$ et calculer les corrélations $r_{P+1,j}$ $\Longrightarrow$ Celle-ci sert de "marqueur" et on constate en effet que pour $N = 700$, certaines corrélations $r_{P+1,j}$ peuvent avoisiner 0.2 et qu'en général λ_k, associée à $X_{0\,P+1}$, est inférieure à 1 et voisine de 0,8.

Si on recommence avec un autre N échantillon de $X_{0\,P+1}$, la valeur λ_k varie légèrement mais son rang k reste très stable pour N assez grand. On pourrait tester cette stabilité en ajoutant non pas 1 mais 2, 3, ... variables X_{P+1}, X_{P+2}, etc ..... Malheureusement les intercorrélations qui vont apparaître entre elles accentuent artificiellement la variabilité de λ_k par rapport au cas où X_{P+1}, X_{P+2} .... sont ajoutées séparément aux P variables initiales (cf I.1.2).

Comme seuls les résultats de la diagonalisation de R_p nous intéressent, et que les diagonalisations de R_{p+1} peuvent être coûteuses , on ne fait en général qu' un ou deux essais.

Nous pensons donc que cette méthode permet de satisfaire la règle de KAISER (ne pas risquer de négliger une variable indépendante), tout en tenant compte de l'échantillonnage .

ⓒ Une méthode un peu différente consiste non pas à regarder où viendrait s'intercaler le facteur associé à une variable indépendante,dans les facteurs d'un ensemble de variables correlées, mais plutôt à regarder ce que deviendraient ces facteurs si les variables n'étaient pas correlées. Pour cela, on pourrait tirer au hasard dans la loi de chaque variable un N échantillon, indépendamment pour chaque variable, puis étudier la matrice de données obtenues. Une façon astucieuse de procéder (LEBART et FENELON , 1978, p.272) consiste à permuter aléatoirement les observations de X_{01} , puis de X_{02} , etc... et d'analyser les facteurs de ces nouvelles variables. En effet, les propriétés "marginales" des variables restent inchangées (même histogramme, etc...) mais leurs intercorrélations sont détruites. On peut même réaliser cette opération plusieurs fois d'où un intervalle de confiance sur le spectre des valeurs propres des variables décorrelées. Si ensuite on leur superpose le spectre des valeurs propres réellement observées, on peut en déduire si le spectre est globalement significatif, mais plus difficilement si chaque valeur propre est séparément significative.

On notera cependant que la méthode d'échantillonnage consistant à permuter aléatoirement le même ensemble de N valeurs est biaisée par rapport au tirage effectif de N autres valeurs dans la loi marginale.

D'autre part, les permutations aléatoires des observations pour chaque variable prise séparément deviennent assez coûteuses quand N est grand, de même que les diagonalisations des matrices $P \times P$. On en trouvera un exemple en III.5.1.

Enfin, nous reviendrons au paragraphe suivant sur la façon d'utiliser la méthode de simulation proposée par LEBART et FENELON.

I.1.2. Aperçus sur les distributions d'échantillonnage des valeurs propres

Celles-ci ont suscité un nombre réduit de publications dont les principales sont dues à D.N. LAWLEY (1956), T.W. ANDERSON (1963), G.A. ANDERSON (1965) et A.G. JAMES (1966). Les articles plus récents ont fait l'objet d'une synthèse par R.J. MUIRHEAD (1978) où il se confirme que, vue la complexité du problème, il y a peu de résultats applicables pratiquement par l'utilisateur. Nous allons en donner quelques exemples.

ⓐ Cas des matrices de covariances (*)

Si on calcule,sur un N échantillon d'une loi multinormale $\mathcal{N}(0,\Sigma)$, la matrice de covariance S , celle-ci suit une loi de WISHART (en fait c'est $(N-1).S$) et si on appelle :

Σ la matrice théorique, $\lambda_1 \ldots\ldots \lambda_p$ ses valeurs propres
$\alpha_1, \alpha_2, \ldots\ldots \alpha_p$ les valeurs propres de Σ^{-1}

S la matrice estimée, $l_1 \ldots l_p$ ses valeurs propres (calculées sur $N = n+1$ individus)

alors la fonction de densité des valeurs propres .... est : (R.G. MUIRHEAD, 1978)

$$f(l_1,\ldots l_p) = \frac{\left(\frac{1}{2}n\right)^{\frac{1}{2}pn} \times \pi^{\frac{1}{2}p^2}}{\Gamma_p\left(\frac{1}{2}n\right) \times \Gamma_p\left(\frac{1}{2}p\right)} \times \prod_{i=1}^{p} \alpha_i^{\frac{1}{2}n} \times \prod_{i=1}^{p} l_i^{\frac{1}{2}(n-p-1)} \times \prod_{\substack{i=1 \\ j<i}}^{p} (l_i - l_j) \times {}_0F_0\left(\frac{1}{2}nL, A\right)$$

où $L = \text{diag}\,(l_1, l_2, \ldots l_p)$

$A = \text{diag}\,(\alpha_1, \alpha_2, \ldots \alpha_p)$

et où ${}_0F_0$ est une fonction hypergéométrique assez difficile à évaluer. La plupart des travaux publiés concernent des approximations dans le cas où les α_i sont distinctes et bien espacées, ou dans le cas de racines multiples, etc...

Enfin, la plupart des résultats sont des résultats asymptotiques, valables pour $N \nearrow \infty$.

Dans un cas particulier qui nous intéressera ($\Sigma \equiv I$, les variables sont supposées gaussiennes et indépendantes), T.W. ANDERSON propose :

$$g(l_1,\ldots l_p) = \frac{\pi^{\frac{1}{2}p} \times \prod_{i=1}^{p} l_i^{\frac{1}{2}(n-p-1)} \times e^{\sum_{i=1}^{p} l_i} \times \prod_{i<j} (l_i - l_j)}{2^{\frac{1}{2}pn} \times \prod_{i=1}^{p} \left[\Gamma\left(\frac{1}{2}(n+1-i)\right) \cdot \Gamma\left(\frac{1}{2}(p+1-i)\right)\right]}$$

Il est toutefois assez difficile de tirer des conclusions pratiques de ces expressions de la fonction de densité et on regarde en général la fonction de vraisemblance, qui peut s'écrire :

$$\mathcal{L} = \prod_{i=1}^{p} \alpha_i^{\frac{1}{2}N} \times {}_0F_0\left(\frac{1}{2}N.L, A\right)$$

Dans des conditions asymptotiques, et pour des valeurs propres théoriques distinctes, G.A. ANDERSON (1965) propose l'approximation suivante :

$$\mathcal{L} = k \prod_{i=1}^{p} \alpha_i^{\frac{1}{2}N} \times e^{-\frac{1}{2}N\sum_{i=1}^{p} l_i\alpha_i} \times \prod_{i<j}^{p} C_{ij}^{-\frac{1}{2}} \times F$$

où k est une constante qui dépend seulement de p , et F un développement en série :

$$F = 1 + \frac{1}{2N}\sum_{i<j}^{p} C_{ij}^{-1} + \frac{9}{8N^2}\sum_{i<j}^{p} C_{ij}^{-2} + \cdots$$

qui tend vers 1 quand N est grand.

Dans cette fonction de vraisemblance, on voit que la partie associée à la valeur propre i est :

$$\alpha_i^{\frac{1}{2}N} \times e^{-\frac{1}{2}N l_i\alpha_i} \times \prod_{j} C_{ij}^{-\frac{1}{2}}$$

et se trouve d'autant plus influencée par les autres valeurs propres que $\frac{1}{\sqrt{(\alpha_j-\alpha_i)(l_j-l_i)}}$ est petit, donc que α_j et/ou l_j est proche de α_i et/ou l_i .

Cette relation sert d'ailleurs à déterminer α_i par la méthode du maximum de vraisemblance.

En effet, si on ne considère que la partie associée à α_i et qu'on en prend le logarithme :

$$\mathcal{L}og\,\mathcal{L} = \frac{1}{2}N\,\mathcal{L}og\,\alpha_i - \frac{1}{2}N l_i \alpha_i - \frac{1}{2}\sum_{j\neq i}\mathcal{L}og(\alpha_i-\alpha_j)$$

l'optimum correspond à :

$$\frac{\partial \mathcal{L}og\mathcal{L}}{\partial \alpha_i} = \frac{1}{2}N \times \frac{1}{\alpha_i} - \frac{1}{2}N l_i - \frac{1}{2}\sum_{j\neq i}\frac{1}{\alpha_i-\alpha_j} = 0$$

(avec $\alpha_i = \frac{1}{\lambda_i}$) et l'estimateur du maximum de vraisemblance est :

$$\lambda_i = l_i - \frac{1}{N}\sum_{j\neq i}\underbrace{\frac{1}{\frac{1}{\lambda_i}-\frac{1}{\lambda_j}}}_{\sim\ \frac{1}{l_i}-\frac{1}{l_j}}$$

d'où :

$$\lambda_i \sim l_i - \frac{l_i}{N}\sum_{j\neq i}\frac{l_j}{l_i-l_j}$$

On peut en tirer 2 conclusions pratiques :

1) Comme ce sont des termes en $\frac{1}{\alpha_i-\alpha_j}$ qui interviennent, ce seront les valeurs propres les plus proches qui influenceront le plus l'estimation de λ_i .

2) Bien que ces résultats supposent les λ_i distincts, on constate que plus α_i sera proche de α_j dans :

$$\lambda_i = \frac{1}{\alpha_i} = l_i - \frac{1}{N}\sum_{j\neq i}\frac{1}{\alpha_i-\alpha_j}$$

plus la valeur estimée l_i risque d'être dispersées car $\alpha_i - \alpha_j$ devient petit.

(b) Des approximations plus grossières, proposées par G.A. ANDERSON (1965) consistent à supposer que quand N est suffisamment grand, le terme F , mais aussi le produit $\prod_{i<j}(l_i-l_j)^{\frac{1}{2}}$, dans la fonction de densité, n'interviennent plus car ce produit devient très vite une bonne approximation de $\prod_{i<j}(\lambda_i-\lambda_j)^{\frac{1}{2}}$

Dans ce cas :

- les estimations l_i deviennent des variables aléatoires indépendantes, de même distribution,
- celle-ci est une loi de χ^2 qui, N étant grand, peut être approchée par une loi normale.

D'où :

- quand N est suffisamment grand, la distribution de l_i , estimation de λ_i , est une loi normale : $\mathcal{N}\left(\lambda_i, \lambda_i\sqrt{\frac{2}{N}}\right)$

Ce résultat suppose les racines distinctes, mais un résultat équivalent peut être obtenu pour une racine λ_i de multiplicité r . On en obtient en effet r réalisations l_{i+1} ..... l_{i+r} et si on estime λ par leur moyenne :

$$\bar{l}_i = \frac{1}{r} \sum_{q=1}^{r} l_{i+q}$$

on sait que l'écart-type de $\bar{l}_i$ est en $\frac{1}{\sqrt{r}}$ x celui de l_{i+q} $\forall q$, et on peut admettre que $\bar{l}_i$ est distribué en :

$$\mathcal{N}\left(\lambda_i , \lambda_i \sqrt{\frac{2}{N.r}}\right)$$

En général, on teste plutôt l'égalité des r dernières valeurs propres, et dans ce cas on garde la loi de χ^2 pour la variable :

$$\chi^2 = -N \sum_{i=p-r+1}^{p} \mathrm{Log}\, l_i - N.r. \mathrm{Log} \sum_i \frac{l_i}{r}$$

avec $\frac{1}{2} r(r+1) - 1$ degrés de liberté.

Dans le cas où $r = p$, cela rejoint le test de sphéricité de BARTLETT.

T.W. ANDERSON donne d'ailleurs dans ce cas la densité des valeurs propres estimées :

$$g(l_1, \ldots, l_p) = \frac{\pi^{\frac{1}{2}p}}{2^{\frac{p.N}{2}}} \times \prod_{i=1}^{p} \frac{l_i^{\frac{(N-p-1)}{2}}}{\Gamma\left[\frac{1}{2}(N-i-1)\right].\Gamma\left[\frac{1}{2}(p-i+1)\right]} \times e^{-\frac{1}{2}\sum_{i=1}^{p} l_i} \times \prod_{i<j} (l_i - l_j)$$

(c) Cas des matrices de corrélations

Il a été moins étudié, car on les rencontre moins fréquemment que les matrices de variance-covariance et les développements en sont encore plus compliqués. En effet, si dans le calcul de S , les moyennes seules sont des estimations (utilisées dans le calcul de s_{ij}) ici on estime d'abord les moyennes, puis les écart-types $\sigma_j = \sqrt{s_{jj}}$ pour estimer les coefficients de corrélation r_{ij} .

De plus, on introduit une contrainte sur les valeurs propres $\sum_{i=1}^{p} l_i = 1$, qui ne se relachera pas même pour N grand. Les valeurs ne seront donc pas asymptotiquement indépendantes.

Le seul cas où l'on dispose de résultats est celui où l'on a 2 valeurs propres distinctes seulement λ_1 et λ_2 de multiplicité q_1 et q_2 telles que $\lambda_1 . q_1 + \lambda_2 . q_2 = p$ (T.W. ANDERSON, 1963).

Dans ce cas, la valeur propre estimée :

$$\bar{l}_2 = \frac{1}{q_2} \sum_{i=q_1+1}^{p} l_i$$

est distribuée asymptotiquement selon une loi normale $\mathcal{N}\left(\lambda_2, \lambda_2 (p - \lambda_2 q_2) \sqrt{\frac{2}{N p q_1 q_2}}\right)$

Ces résultats tiennent compte de la contrainte en espérance mathématique, alors qu'elle est aussi vérifiée par les estimations.

Dans le cas où l'on a 1 seule valeur propre de multiplicité $r = p$ ce résultat est inapplicable directement puisque $\bar{l}$ n'a aucune variabilité et reste

égale à 1.

Note - La plupart de ces résultats sont résumés dans D.F. MORRISON (1967) p.247 et suivantes.

(d) Intérêt au niveau des applications

Remarquons d'abord qu'il s'agit de résultats asymptotiques pour N supposé grand, mais que nous n'avons trouvé nulle part d'ordre de grandeur, ou d'essais en simulation permettant de savoir à partir de quelles valeurs de N , et de P , on peut les appliquer.

De plus, il est rare que l'on ait une idée a priori des valeurs théoriques et, sauf cas particulier, on testera rarement si r racines $l_{q+1}, \ldots\ldots l_{q+r}$ ne correspondent pas par hasard à une seule valeur théorique λ_q.

Par contre il est plus courant de tester la sphéricité des $P-q$ dernières valeurs propres. Par exemple, si la matrice de covariance théorique est de la forme :

$$\Sigma = \Psi + \sigma^2 I_P$$

et Ψ de rang q , car dans ce cas : $\lambda_i(\Sigma) = \lambda_i(\Psi) + \sigma^2$

Exemple : Dans un réseau de stations, on peut imaginer qu'il y a une structure, limitée à quelques facteurs, plus des bruits de mesure indépendants mais de même variance σ^2 pour chaque station.

On peut alors tester si la structure se limite aux 2, 3, etc.. premiers facteurs.

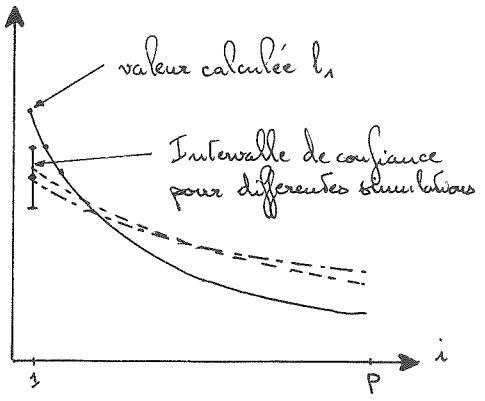
Dans le cas où l'on utilise la matrice de corrélation, le test est théoriquement inapplicable et T.W. ANDERSON (1973) propose simplement de vérifier que la variance résiduelle de toutes les variables sur les $P-q$ derniers facteurs est faible..!

Un autre résultat intéressant est l'explication du spectre relativement régulier que l'on obtient en général, même quand le spectre théorique présente une cassure nette (exemple : $\Psi + \sigma^2 I$). En effet les fluctuations d'échantillonnage tendent à lisser le spectre : A.T. JAMES (1966) a montré que ce résultat tient à la non-indépendance des valeurs propres et aux termes en $l_i - l_j$, ou $(l_i - l_j)^{\frac{1}{2}}$ selon que l'on considère les distributions exactes ou asymptotiques (On peut même le pressentir en les supposant indépendantes car alors une valeur anormale de l_i risque de la faire permuter avec l_{i-1} ou l_{i+1} dans le spectre des valeurs classées). Ceci explique pourquoi il est décevant de chercher une "cassure" dans le spectre, même si elle est théoriquement fondée.

I.1.3. Retour sur les simulations préconisées par Lebart et Fénelon :

La permutation aléatoire des observations fournit un ou plusieurs spectres ,qui est une réalisation du cas où Σ (respectivement R) ,matrice théorique ,est diagonale.

Différentes permutations permettent de définir un intervalle de confiance , et, si le spectre observé se trouve nettement au dessus de celui-ci,cela prouve bien que la



matrice n'est pas diagonale, et que la première valeur propre au moins est significative. Par contre, si on considère la q ème , il faut se rappeler que l'intervalle de confiance obtenu correspond à l'hypothèse :

$$\lambda_q = \lambda (= \lambda_1 = \cdots \lambda_P)$$

Or, si on considère désormais que la première, λ_1, est significativement différente de λ alors, il faut tester l'hypothèse $\lambda_2 = \lambda_3 = \cdots = \lambda_P = \lambda'$, où l'on doit avoir $\lambda' < \lambda$. De même, si on suppose λ_2, puis λ_3, $\cdots\cdots \lambda_{q-1}$ significatives..

Donc, hormis pour la première, l'intervalle de confiance obtenu est centré sur une valeur biaisée par excès et le test risque d'être systématiquement pessimiste sur le nombre de facteurs significatifs.

Ceci est d'autant plus vrai que l'on utilise une matrice de corrélation car on voit que la contrainte $\sum_{i=1} \lambda_i = P$ entraîne bien, si $\lambda_1 > 1$, $\lambda' = \frac{P-\lambda_1}{P-1} < 1$

Il existe pourtant une façon d'appliquer objectivement ce test :

Si l'on a accepté l_1 comme significative, cela revient à dire que la première composante Z_1 est significative, donc doit être préservée. Or chaque variable X_j peut s'écrire :

$$X_j = \rho_{1,j} \cdot \frac{Z_1}{\sqrt{\lambda_1}} + \varepsilon_{j\cdot 1} = V_{1j} \cdot Z_1 + \varepsilon_{j\cdot 1}$$

et on peut calculer, pour toutes les observations $i = 1, \ldots N$, les $\varepsilon_{ij\cdot 1}$ de toutes les variables: $j = 1, \ldots P \Rightarrow$ le tableau $\mathcal{E}_{OV.1} = \{\varepsilon_{ij\cdot 1}\}$

Si on analyse la matrice de variance-covariance associée, soit $\frac{1}{N} \mathcal{E}_{OV.1}^{t} \cdot \mathcal{E}_{OV.1}$ son spectre sera : $0, l_2, \cdots\cdots l_P$

On peut alors perturber aléatoirement les observations de chaque $\varepsilon_{j\cdot 1}$, d'où un nouveau spectre : $\{0, \mu_2, \mu_3, \cdots\cdots \mu_P\}$ et réaliser l'opération plusieurs fois. Si l_2 sort significativement de l'intervalle obtenu sur les 2, on accepte l_2 et on calcule tous les résidus aux composantes 1 et 2 : $\varepsilon_{ij\cdot 1,2}$ et on teste alors l_3, etc...

$$X_j = V_{1j} \cdot Z_1 + V_{2j} \cdot Z_2 + \varepsilon_{j\cdot 1,2}$$

Cela revient à faire des déflations successives sur la matrice initiale. On en verra l'utilisation en IV.2.

Cette façon de procéder teste effectivement chaque valeur propre séparément et successivement, en tenant compte de celles déjà retenues .On arrête au pas K quand les valeurs simulées ν_K ne diffèrent plus significativement de l_K.

Note : Cela suppose à chaque pas k perturbations (coûteuses si N est grand) de p variables et k diagonalisations (coûteuses si P est grand) de matrice de covariance. On donne en annexe un aperçu de l'organigramme et des méthodes employées.

I.1.4. La méthode LEV et quelques exemples de simulation

(a) Considérations théoriques sur la méthode LEV (Log-Eigen-Values)

- Il s'agit d'une méthode empirique proposée par des météorologistes anglais pour détecter le nombre de facteurs significatifs dans un ensemble de variables constituées par les mesures, en différentes stations, d'un même paramètre météorologique.

La règle proposée serait la suivante :

" Si on porte sur un graphique le logarithme des valeurs propres l_i en fonction de leur rang i , ce diagramme présente en général une partie médiane rectiligne. Seules les valeurs qui précèdent cette partie seront considérées comme significatives tandis que les autres seront rejetées (CRADDOCK J.M., 1973 ; FARMER S.A. 1971, et PROBERT JONES J.R., 1973)."

L'origine expérimentale de cette méthode est évidente :

- l'allure exponentielle du spectre suggère une transformation en logarithme (les l_i sont toujours > 0 et, s'il s'agit d'une matrice de corrélation, la valeur 1.0 joue un rôle particulier)
- on admet que, pour les données particulières auxquelles on l'a appliquée, (champs spatiaux) il y a une structure forte sur les quelques premiers facteurs, mais que chaque variable a ensuite une variabilité propre, du même ordre de grandeur pour chaque station. Ce sont ces variances individuelles (erreurs et bruits de mesures) qui apparaissent dans les facteurs suivants dont les valeurs propres associées sont théoriquement très voisines
- la recherche d'une "cassure" dans le spectre conduit à considérer non pas l'écart absolu entre les valeurs propres successives, mais leur décroissance relative (cf IIème partie, Chap.I-4). Tant que celle-ci reste constante :

$$\frac{l_{i-1}}{l_{i-2}} = \frac{l_i}{l_{i-1}} = \frac{l_{i+1}}{l_i} = \cdots\cdots$$

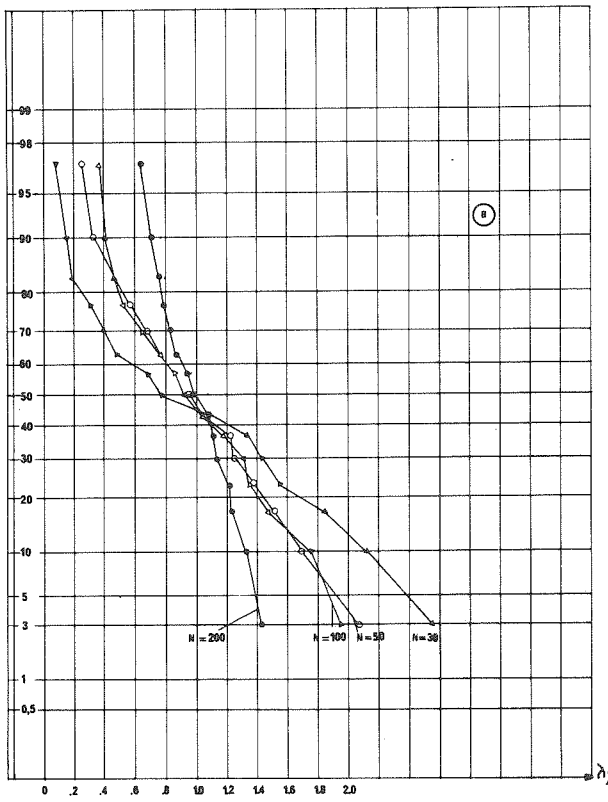
on ne peut considérer qu'il y a "cassure", et ces valeurs propres doivent être acceptées ou rejetées en bloc.

La question se pose de savoir si cette méthode empirique a des fondements théoriques et si elle s'applique non seulement aux données initialement considérées mais aussi à d'autres cas, comme les mélanges de variables hétérogènes.

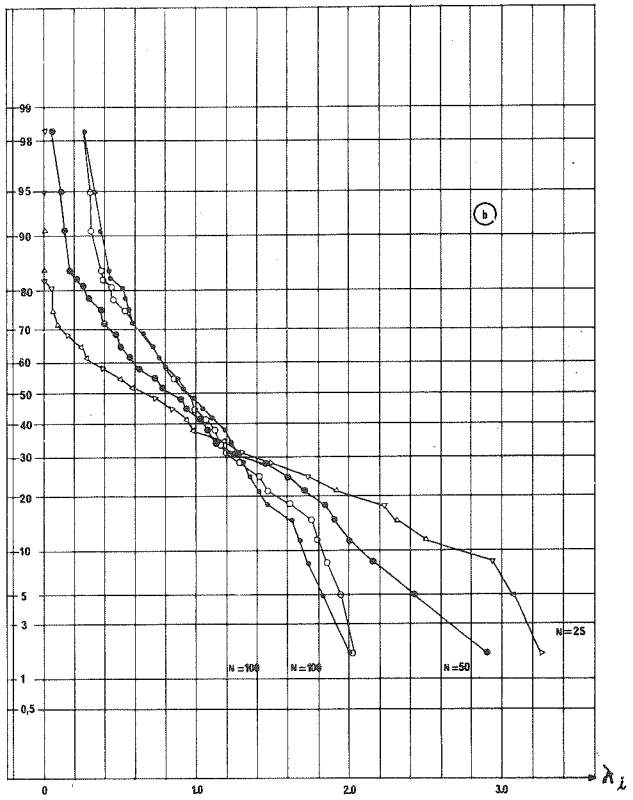
Si les considérations de FARMER et CRADDOCK sont essentiellement expérimentales, J.R. PROBERT JONES (1973) cherchait, de son côté, la distribution d'échantillonnage des valeurs propres d'une matrice de corrélation entre des variables aléatoires indépendantes. Il aboutit, par une démarche voisine de l'analyse dimensionnelle, à une expression théorique de la forme :

$$l_i = k(i-1).i \qquad \text{sauf pour } i = 1$$

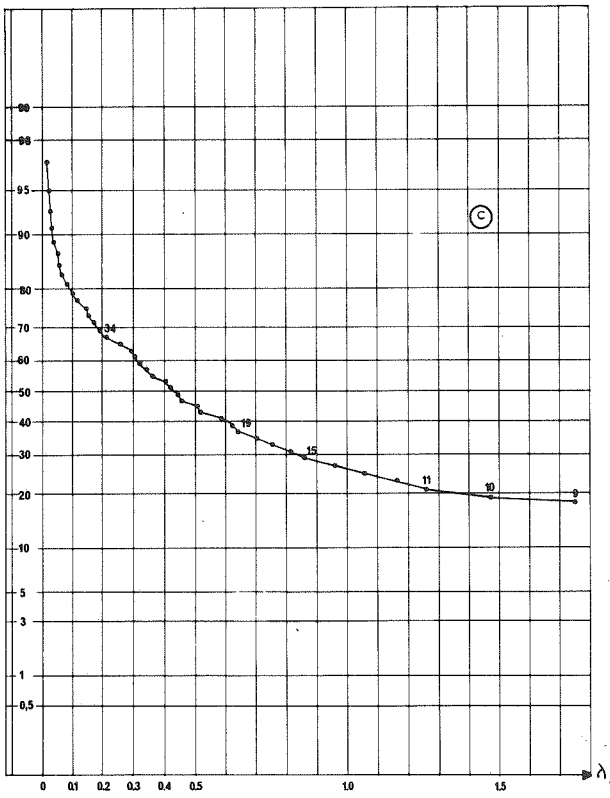
Malheureusement, les justifications sont obscures, voire erronées, et l'auteur semble ignorer les travaux de T.W.ANDERSON et autres cités en I.1.2. Nous proposons ci-dessous une autre justification, partiellement expérimentale elle aussi.



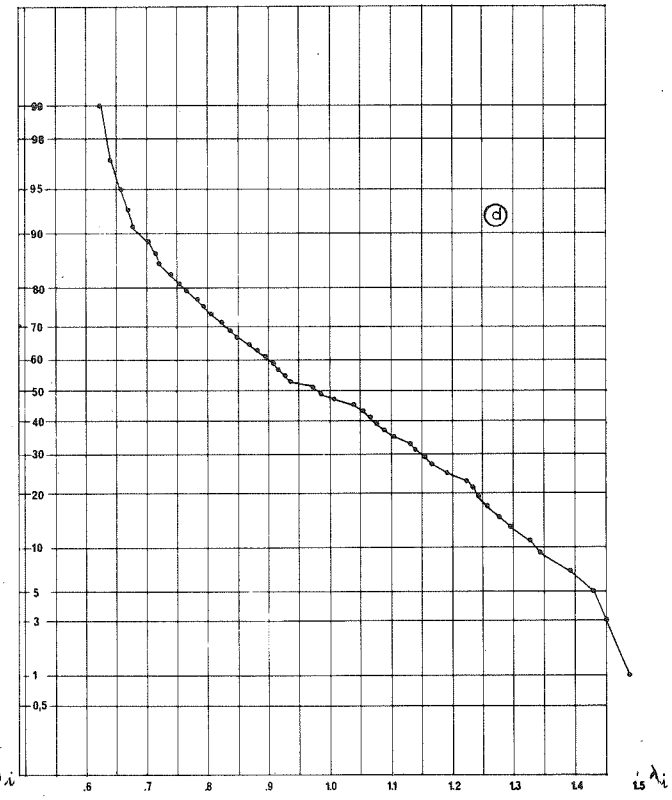
ⓐ p = 15 variables aléatoires indépendantes . Effet de l'échantillonage .



ⓑ p = 30 variables aléatoires indépendantes . Effet de l'échantillonage .



ⓒ p = 50 variables - Matrice originale de DAVOS , (bimestre Janvier-Fevrier)



ⓓ p = 50 variables - Mêmes données qu'en ⓒ mais perturbées aléatoirement .

FIGURE III - 1 : Distributions des valeurs propres de matrices de corrélation sur diagramme de GAUSS .

ⓑ Simulations de variables aléatoires indépendantes :

On a d'abord simulé des ensembles de P variables gaussiennes indépendantes, de moyenne nulle et d'écart type 1.0 (Sous-programme GAUSS de la Librairie Scientifique SSP d'IBM) sur des échantillons de tailles N . On en déduit ensuite des matrices de corrélation ($P \times P$) que l'on diagonalise.

On se trouve donc dans le cas d'une seule valeur propre théorique, égale à 1.0 et de multiplicité P . Les valeurs propres obtenues diffèrent sensiblement de 1.0 (bien que leur moyenne soit strictement égale à 1.0) et sont bornées par 0 et P . Or les résultats asymptotiques font pressentir que quand N est grand leur distribution devrait tendre à être normale autour de 1.0 .

C'est ce que l'on constate, pour $P = 15$, dès que $N \sim 50$ et de façon plus marquée, pour $N = 200$. Pour $P = 30$, on constate que l'alignement des points s'améliore entre $N = 25$ et $N = 100$ sans être encore parfait (Fig.III-1 a et b). Quant aux données de Davos décorrelées par permutation avec $N > 800$, on constate aussi une quasi-normalité de la courbe dans la partie médiane (bien que cette fois les variables ne soient pas elles-mêmes normales) (Fig.III-1 d) très différente du spectre initial (Fig.III-1 c).

Malheureusement les résultats asymptotiques du paragraphe précédent sont de peu d'intérêt ici puisqu'ils donnent la distribution d'échantillonnage de la valeur moyenne : $\overline{\ell}_q = \frac{1}{n} \sum_{i=1}^{n} \ell_{q+i}$ d'une valeur propre de multiplicité n , et non les ecarts-types respectifs des n réalisations obtenues. De plus ici, nous sommes dans le cas d'une matrice de corrélation, avec $n = P$ et ℓ n'a strictement aucune variabilité ($\overline{\ell} \equiv 1.0$).

Le tableau 1 donne quelques exemples des écarts-types $\sigma_\ell^2 = \frac{1}{P-1} \sum_i (\ell_i - 1)^2$ obtenus .

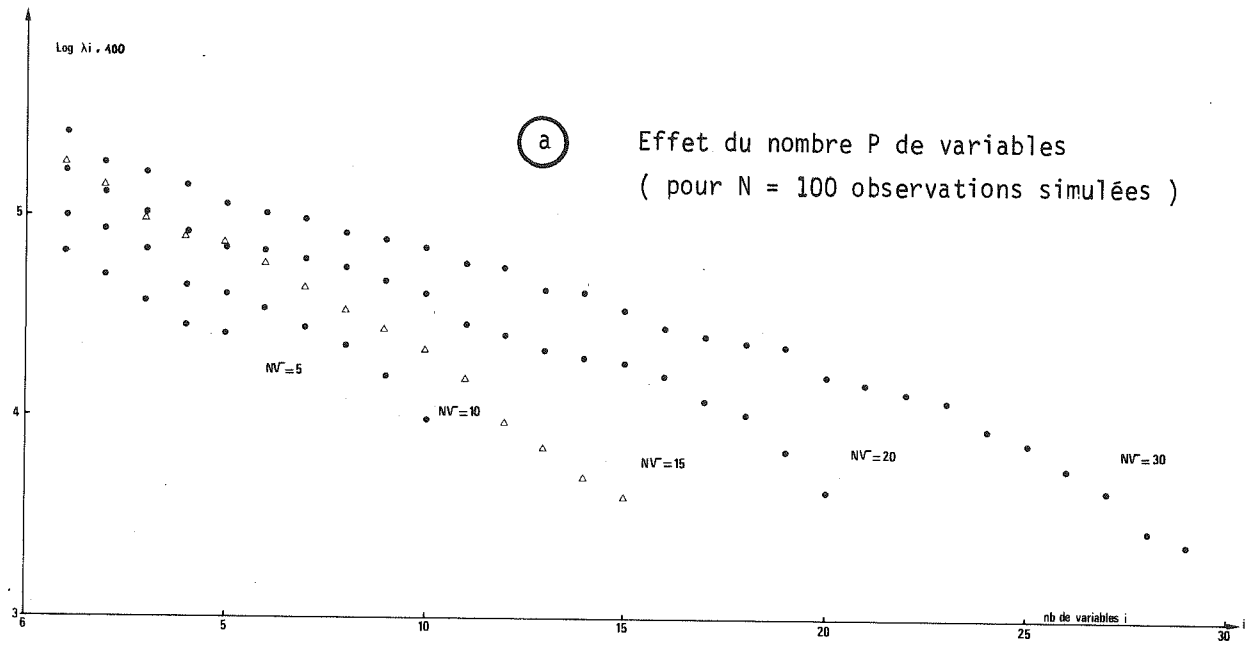
Tableau 1 - Ecart-type des valeurs propres en fonction du nombre d'individus N de l'échantillon et du nombre P de variables.

ⓐ Evolution avec N (P fixé)

$P = 15$	$N =$	25	50	100	200
	$\sigma_\ell =$	0.778	0.523	0.495	0.238
$P = 30$	$N =$	25	50	100	
	$\sigma_\ell =$	1.060	0.771	0.537	
$P = 50$	$N =$	829			
	$\sigma_\ell =$	0.236			

ⓑ Evolution avec P (N fixé): $N = 100$

$P =$	5	10	15	20	30
$\sigma_\ell =$	0.172	0.311	0.495	0.411	0.537



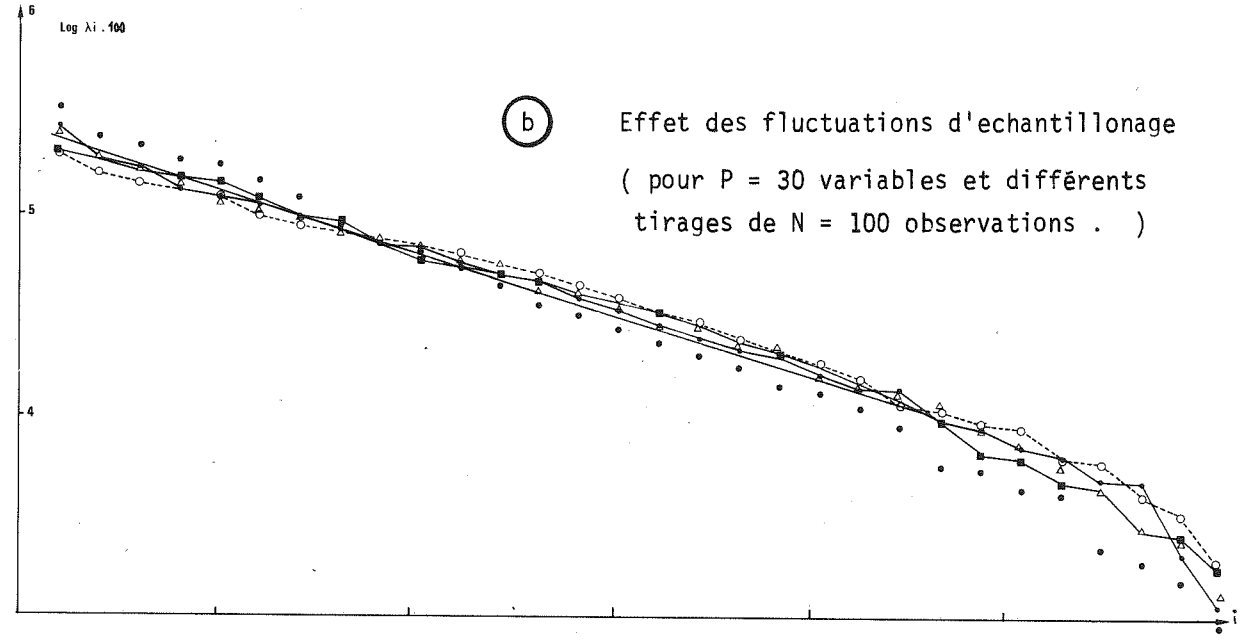
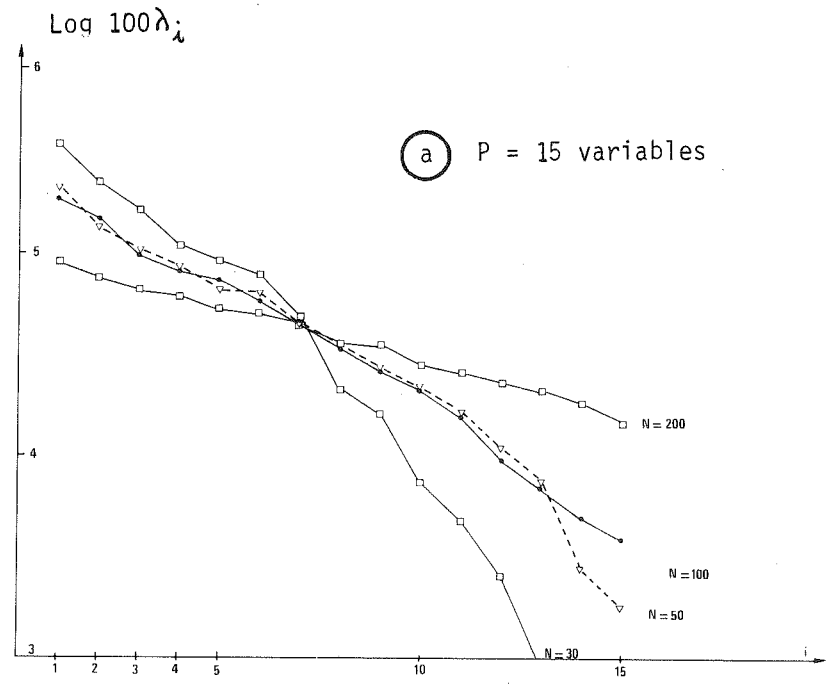


FIGURE III - 2 : Représentation en Logarithme (méthode L.E.V.)

du spectre de matrices de corrélation aléatoires.



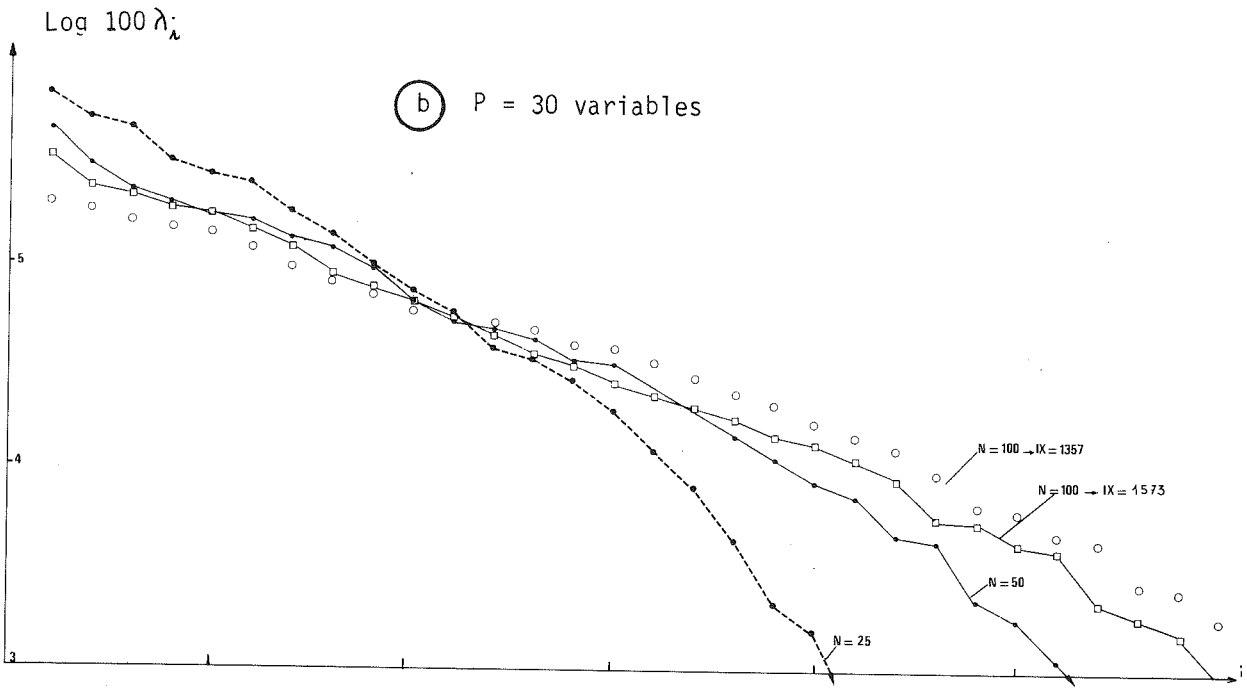


FIGURE III - 3 : Effet du nombre d'observations N sur la forme du spectre.

Dans le tableau 2, on a réalisé différentes simulations pour $N = 100$ et $P = 30$ d'où un aperçu des fluctuations d'échantillonnage d'une valeur propre de rang k fixé.

Tableau 2

$N = 100$ $P = 30$

	Simulation ↓	ℓ_1	ℓ_5	ℓ_{10}	ℓ_{15}	ℓ_{20}	ℓ_{25}	ℓ_{30}
σ_ℓ = .537	1	2.013	1.746	1.186	.975	.705	.443	.260
.493	2	2.019	1.626	1.264	.985	.711	.515	.266
.545	3	2.290	1.629	1.276	.927	.669	.464	.214
.539	4	2.256	1.582	1.267	.934	.666	.478	.228
	$\bar{\ell}_k$	2.145	1.646	1.248	.955	.688	.475	.242
	σ_{ℓ_k}	.149	.070	.042	.029	.024	.024	.025

On vérifie que pour les ℓ_k prises séparément leur écart-type d'échantillonnage σ_{ℓ_k} est faible par rapport à σ_ℓ. De plus, si on admet que la contrainte $\sum_1^P \ell_k = 10$ ne joue pas sur 1 valeur prise isolément, on peut rapprocher σ_{ℓ_i} du $\sigma_{\bar{\ell}}$ théorique pour une matrice de covariance soit $\sigma_{\bar{\ell}} = 1.0 \cdot \sqrt{\frac{2}{N.P}}$ soit ici 0.026.

Si maintenant on porte sur un diagramme les logarithmes des valeurs propres en fonction de leur rang, on constate un alignement assez surprenant de $\log \ell_i$ en fonction de i (Fig.III-2), plutôt meilleur en début de diagramme mais s'incurvant parfois vers la fin, pour les faibles valeurs propres .

Pour un nombre fixé de variables et pour un même nombre d'observations, il y a certes des fluctuations d'échantillonnage (Fig.III-2 b). Par contre pour un nombre de variables fixé, la taille de l'échantillon joue un rôle considérable (Fig.III-3 a et b).

(c) Essai de justification de la méthode

Cet alignement sur le diagramme en logarithme est assez surprenant, et il serait intéressant de lui trouver une justification théorique. Si on accepte l'hypothèse de normalité des valeurs propres, qui n'est vraie asymptotiquement que pour les matrices de covariance, le spectre correspond alors à valeurs extraites d'une loi normale et que l'on aurait classées. Or on connait la distribution de la kème valeur classée (analogue à celle obtenue en Annexe III au Chap.I de la Ière partie). On peut d'ailleurs calculer l'espérance de cette kème valeur $E[\ell_k]$ et celles-ci ont été tabulées par FISCHER et YATES (1953) pour $\ell \in \mathcal{N}(0,1)$.

Par exemple, si on tire P = 30 ou P = 50 valeurs on a :

Rang	k =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
p=30	$E(\ell_k)$=	2.04	1.62	1.36	1.18	1.03	.89	.78	.67	.57	.47	.38	.29	.21	.12	.04										
p=50	$E(\ell_k)$=	2.25	1.85	1.63	1.46	1.33	1.22	1.12	1.03	.95	.87	.80	.74	.67	.61	.55	.49	.44	.38	.33	.28	.23	.18	.13	.08	.03

(Note : la distribution étant symétrique , on n'en donne que la 1ère moitié)

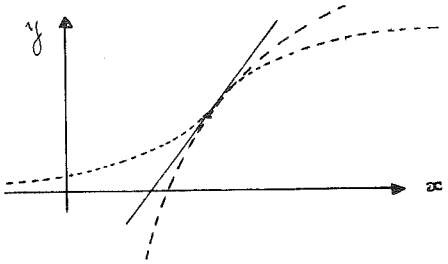
Pour se ramener au cas des valeurs propres qui nous intéresse, il faut remplacer la loi $\mathcal{N}(0,1)$ par une loi $\mathcal{N}(1.0,\sigma)$ avec σ tel que les valeurs $E[\ell_p]$ soient au minimum > 0. Dans le cas où $p = 30$ cela suppose au moins $\sigma < .5$ et pour $p = 50$ $\sigma < .4$.

Or on vérifie déjà bien, dans ces cas minimaux, que le graphique $\log(E[\ell_k])$ en fonction de k est déjà quasi rectiligne dans la partie médiane (Fig.III-4 a) et donc que les $E[\ell_k]$ sont en progression quasi-géométrique.

D'autre part, les valeurs ℓ_k, qui seraient elles normales, n'ont pas la même loi que $E[\ell_k]$ mais on démontre, comme pour la loi uniforme, que l'écart-type de la kème valeur classée, σ_{ℓ_k}, est très faible par rapport à l'écart type de ℓ, σ_ℓ (cf aussi tableau 2). Leur comportement, en particulier dans le diagramme en Log, doit donc être voisin.

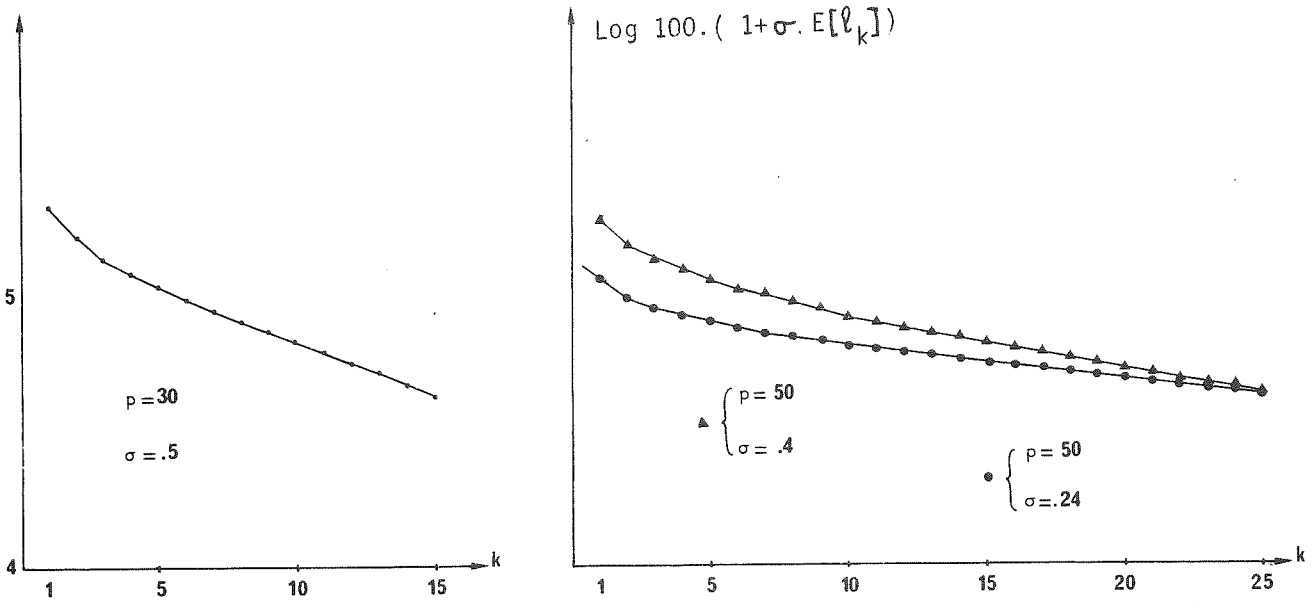
Si on revient maintenant aux valeurs propres observées, leur distribution n'est pas strictement normale. Toutefois elle s'en approche et ce d'autant plus que N est grand. (cf figures III-4-b et III-4-c). Pour les faibles valeurs de N, on constate seulement un biais en diagramme normal vers l'extrémité correspondant aux faibles valeurs propres, mais ce biais existe aussi dans le diagramme en log.

Enfin, quelque soit la distribution des variables il est évident que, dans la mesure où les écarts-types sont faibles, on peut approcher la fonction de répartition aussi bien par $y = Erf(x)$, $y \approx Log\, x$ que par $y \approx x$ au voisinage de 1.0.

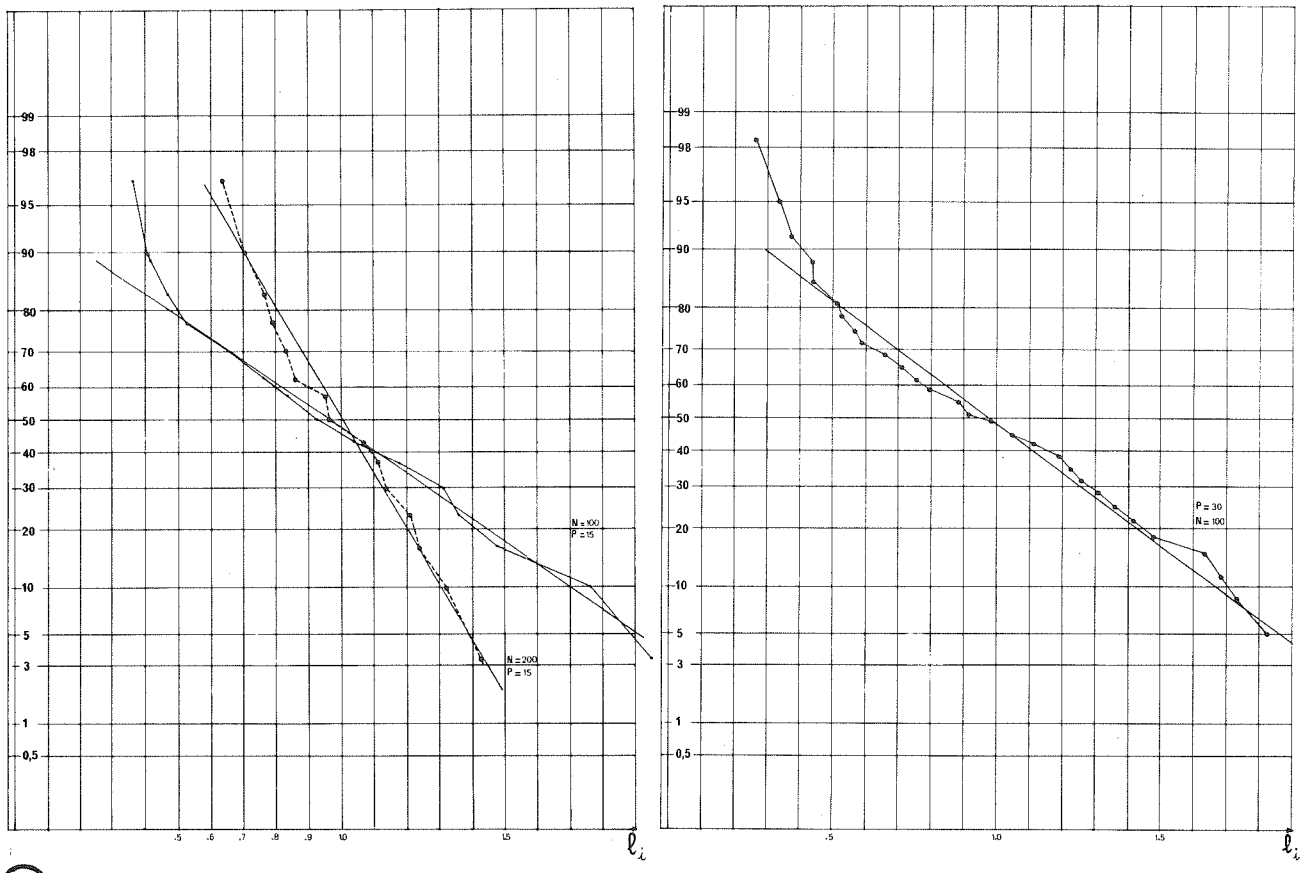


En conclusion, la transformation en logarithme n'a aucune vertu particulière sinon de ressembler à la transformation normale, qui est plus fondée théoriquement, au moins asymptotiquement. Quand N est trop faible relativement à P, l'effet de la contrainte de positivité se fait sentir pour les petites valeurs propres, mais autant en diagramme normal que logarithmique.

La méthode LEV semble donc reposer essentiellement sur la normalité approchée des valeurs propres.



(a) Logarithme des espérances de valeurs normales classées. ($\ell \in \mathcal{N}(0,1)$)



(b) Représentation en diagramme normal des valeurs propres de 15 variables indépendantes, (pour N=100 et N=200)

(c) Idem pour 30 variables (et N=100) .

FIGURE III - 4 : Liaison entre la transformation Log.(méthode LEV) et le Diagramme de Gauss .

(d) Simulations de variables aléatoires corréllées (par paquets)

On se donne d'abord le nombre $L < P$ de variables correlées que l'on souhaite et la matrice de corrélation théorique que l'on souhaiterait avoir entre ces L variables, soit R_L.

On diagonalise R_L d'où les valeurs propres λ_j^L et les vecteurs propres associés V_j^L.

On génère alors N observations de L variables indépendantes Z_j qui représenteront les composantes principales des L variables de R_L (z_{ij} = i ème observation de Z_j, doit être tirée dans une loi $\mathcal{N}(0, \sqrt{\lambda_j^L})$).

On recombine les variables $\{Z_j, j = 1, \ldots L\}$ par la matrice $Q = \{V_1^L, \ldots V_j^L, \ldots V_L^L\}$ d'où :

$$X_{OL} = Z_{OL} \cdot Q$$

et en espérance mathématique $\frac{1}{N} X_{OL}^t \cdot X_{OL} \longrightarrow R_L$. On ajoute ensuite, à X_{OL}, $P-L$ variables indépendantes d'où X_{OP} (dimension $N \times P$) qui fournit la matrice R_P que l'on étudie, sachant que certains sous-ensembles de variables sont correlés.

On constate (Fig.III-5) que le diagramme présente en général une partie médiane rectiligne précédée par un certain nombre de valeurs propres (voisin du nombre de paquets de variables correlées). A l'extrémité droite, on trouve au contraire des valeurs propres anormalement faibles mettant en évidence le nombre de variables déjà représentées par les 1er facteurs dominants.

Toutefois pour qu'un paquet de variables apparaisse bien parmi un ensemble de variables indépendantes, il faut qu'il soit suffisamment marqué, compte tenu de l'échantillonnage. Par exemple, le paquet :

$$\begin{matrix} 1.0 & .6 \\ .6 & 1.0 \end{matrix}$$

n'apparait pas beaucoup dans la Figure III-5 car il est noyé dans des variables indépendantes ayant entre elles des coefficients de corrélation atteignant .297 (pour $N = 100$). Donc, pour un paquet donné, la mise en évidence sera d'autant moins nette que l'échantillonnage est faible.

La règle qui considère la partie rectiligne comme aléatoire et la partie gauche de la courbe comme significative est donc un peu pessimiste (on décèle moins de paquets qu'il n'y en a en fait). Par contre, si on considère la courbe comme significative tant qu'elle reste rectiligne, on conserve bien 25 à 27 variables significatives.

En effet, le but n'est pas d'éliminer les influences aléatoires (dans ces cas simulés, il n'y a pratiquement que cela) mais de ne pas s'arrêter arbitrairement . Le test serait donc assez sensible, s'il n'y avait pas l'influence d'autres effets.

D'autre part, si on porte les valeurs propres sur un diagramme normal, on constate une certaine courbure (Fig.III-5 b) qui s'explique car certaines valeurs, les premières et les dernières, n'ont pas la même loi que les autres. Toutefois, la partie médiane du graphique est relativement rectiligne, ce que l'on peut vérifier en ne traçant que la distribution des ℓ_i pour i = 3 à 27, par exemple (Fig. III-5-c).

a) Représentation en diagramme Log (méthode LEV).

b) Représentation de l'ensemble en diagramme normal.

c) Représentation des 25 variables centrales (en éliminant les paquets).

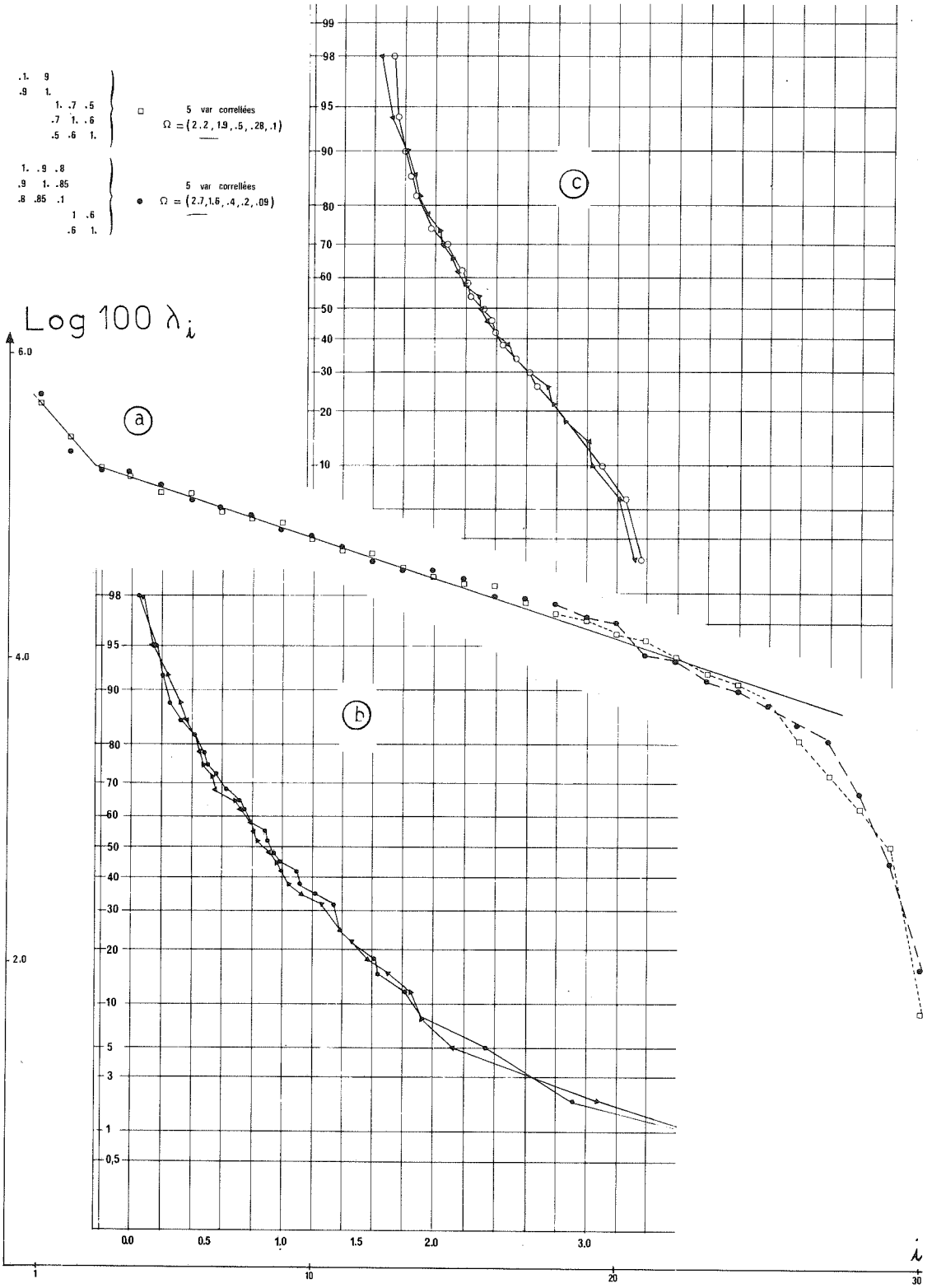


FIGURE III - 5 : Simulation de variables corrélées par paquets (30 variables dont 5 regroupées en paquets)

Donc expérimentalement, une valeur propre multiple, de multiplicité $n < p$ fournit encore n réalisations distribuées à peu près normalement.

On peut donc imaginer que des valeurs propres successives λ_i , λ_j , etc de multiplicités q , r , etc... donneront des paquets $l_{i+1} \ldots l_{i+q}$, $l_{j+1} \ldots l_{j+r}$ de moyennes $\overline{l}_i$, $\overline{l}_j$ ... et d'écarts-types $\sigma_{\overline{l}_i}$, $\sigma_{\overline{l}_j}$ ... différents, mais dont les distributions grossièrement normales, donneront dans le graphique en Log des tronçons de droites successifs. Pour le choix du nombre de facteurs, un tronçon devrait être inclus ou rejeté en bloc puisqu'il est associé à une même valeur propre multiple.

(e) Simulations de variables aléatoires indépendantes mais autocorrélées :

On simule encore p variables indépendantes entre elles (c'est-à-dire $\forall j \neq l \quad \rho_{jl} = E[x_j x_l] = 0$) mais au sein de la variable j , les observations sont liées, par exemple par un schéma d'autocorrélation simple:

$$x_{i,j} = \rho\, x_{i-1,j} + \varepsilon_i$$

avec $var[\varepsilon_i] = 1 - \rho^2$ et $x_0 = \varepsilon_0$

On génère donc seulement les ε_i et on construit ainsi les x_{ij}. Dans notre cas on a d'abord simulé la même autocorrélation au sein de chaque variable. Un exemple réel est celui des mesures d'une même variable sur un réseau de stations avec un échantillonnage dans le temps un peu trop fréquent par rapport à la fréquence du phénomène, d'où une certaine autocorrélation .

On constate que les Log. des valeurs propres restent alignées mais que la pente moyenne de la droite se redresse sensiblement quand ρ augmente (Fig.III-6). Par contre un effet parasite important est l'apparition d'une cassure dans le spectre, avec des premières valeurs propres anormalement élevées qui font croire à la présence de "grappes", complètement injustifiée dans notre cas.

On remarque aussi que cette cassure n'apparait pas dans le début du spectre tant que les autocorrélations n'atteignent pas .6 . Mais on a vu que dans le cas des données nivométéorologiques de Davos (cf IIème Partie, Chap.I.4.1) les autocorrélations atteignaient ces valeurs.

Un raison pouvant expliquer l'apparition de ces "groupes" est l'élévation, avec l'autocorrélation de la variance d'échantillonnage du coefficient de corrélation entre 2 variables théoriquement indépendantes. En effet, la présence d'une autocorrélation dans un N échantillon fait qu'il a en fait un nombre d'individus "équivalents"

$$N_{eq} < N$$

On trouvera cette notion développée dans MARCHENKO A.S. et MINAKOVA L.A. (1972) repris par BOIS Ph. (1976).

Par exemple, pour $N = 60$, le rapport $\lambda = \frac{N_{eq}}{N}$ en fonction de l'autocorrélation varie comme suit :

ρ =	.2	.4	.6	.8	.9
λ =	.92	.75	.55	.30	.15

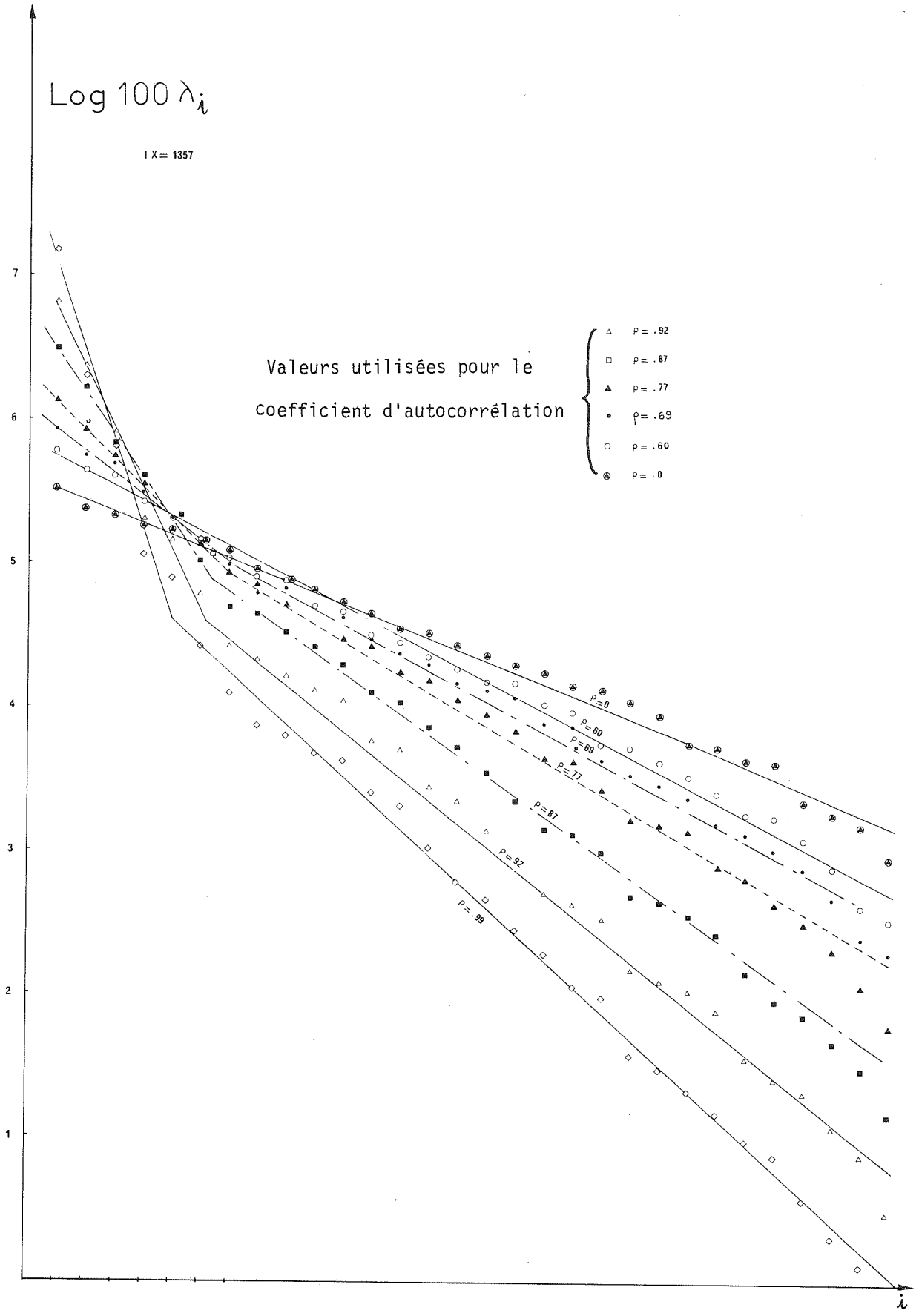


FIGURE III - 6 : Effet de l'autocorrélation sur le spectre de 30 variables indépendantes (pour 100 observations).

Dans notre cas cela se traduit pour N = 100 par des maxima de corrélation dans la matrice théoriquement aléatoire de :

ρ =	.0	.2	.4	.6	.7	.8	.9	.95
r max =	.477	.526	.574	.617	.642	.685	.778	.864

(Ceux-ci étant obtenus avec la même série de nombres aléatoires. On fait seulement croître ρ dans le schéma de MARKOV).

Cette diminution du N équivalent fait que la distribution ne peut plus être considérée comme normale (cf alinéa c précédent) d'où la courbe obtenue pour les Log l_i en fonction de i.

Enfin, on a comparé, pour les autocorrélations voisines de .9 , le cas où les 30 variables sont autocorrelées et celui où il y en a 15 seulement, parmi les 30. L'effet est très sensible, la figure se ramenant un peu à celle obtenue pour ρ = .6 sur 30 variables.

Cela montre la difficulté d'interprétation quand dans notre paquet, des variables sont autocorrelées mais avec des intensités très différentes.

I.1.4. Conclusions

Cette brève incursion dans le domaine des propriétés d'échantillonnage des valeurs propres d'une matrice de corrélation nous conduit aux conclusions suivantes :

- Les résultats théoriques sont rares souvent compliqués et inapplicables en pratique.
- Les méthodes de simulation permettent de se faire une idée heuristique de ces propriétés, mais elles sont lourdes à mettre en oeuvre.
- La règle empirique proposée par KAISER peut être améliorée par l' introduction d'une variable aleatoire indépendante servant de marqueur. Celle-ci peut même être générée au besoin avec une certaineautocorrélation. Les simulations préconisées par LEBART et FENELON deviennent beaucoup plus intéressantes si on les effectue conditionnellement aux valeurs propres déjà acceptées .
- La méthode LEV qui consiste à étudier le graphique du logarithme des valeurs propres en fonction de leur rang est empiriquement fondée car une valeur propre de multiplicité r donne r valeurs estimées distribuées quasi-normalement (pour N grand)et qui, après classement s'alignent à peu près en logarithme. Cette méthode permet de ne pas placer le seuil arbitrairement parmi des valeurs estimées qui risquent d'être associées à la même valeur théorique multiple, par exemple 1.0 si on a un paquet de variables indépendantes.
- La combinaison des effets d'échantillonnage et d'autocorrélation dans les variables rend difficile la détection objective des facteurs significatifs et la seule observation du spectre constitue un test très peu puissant.

I.2 - Retour sur la notion de dimensionalité (*)

I.2.1. Dimensionalité linéaire et dimensionalité intrinsèque (*)

Pour satisfaire le "principe de parcimonie" (minimiser le nombre de descripteurs d'un phénomène) ou plus prosaïquement pour construire un modèle avec un sous-ensemble de variables les plus indépendantes possible (amélioration de la robustesse des modèles) on est conduit à rechercher la dimension "vraie" de l'espace où l'on travaille.

Une approche familière consiste à regarder chaque variable comme un vecteur de $\mathbb{R}^N$ et à chercher la dimension ℓ de l'espace engendré par ces P variables (centrées et éventuellement normées). Il s'agit alors d'une dimensionalité linéaire :

$$\ell \leq \text{Min}\,(N-1, P)$$

qui exprime que l'on peut reconstituer linéairement toutes les variables avec seulement ℓ informations linéairement indépendantes.

On peut même assouplir cette règle et chercher à reconstituer "au mieux" les P variables, et donc, pour un seuil donné de reconstitution, la dimension ℓ' qui est nécessaire. On sait d'ailleurs que ce sont les ℓ' premières composantes principales qui optimisent cette reconstitution, et cela rejoint et justifie tous les essais du paragraphe précédent.

Une autre approche consiste à rechercher la dimensionalité "vraie" ou intrinsèque d'un ensemble de variables. A titre d'exemple, on pourra considérer chaque observation comme un signal, échantillonné dans le temps à des intervalles $t_1, t_2 \ldots t_P$, et l'observation i est :

$$X_{iV} = \left\{ f_i(t_1), f_i(t_2), \ldots, f_i(t_P) \right\}$$

L'ensemble des observations constitue donc un ensemble de signaux $F_i(t)$ et on peut imaginer que ces signaux soient engendrés par un même générateur ayant k paramètres indépendants $\varphi_1 \ldots \varphi_k$, qui pour le signal i , auraient été fixés à un ensemble de valeurs :

$$\Phi_i = \left\{ \varphi_1(i), \varphi_2(i), \ldots, \varphi_k(i) \right\}$$

Il est évident que la dimensionalité intrinsèque de cet ensemble de signaux est k et que, si le nombre de points d'échantillonnage P est suffisant : $P \geq \text{Max}(k,\ell)$ alors on a la relation entre les dimensionalités :

dimension vraie : $k \leq \ell$: dimension linéaire

Exemples

① Si on a pour générateur la fonction à 2 paramètres :

$$F(t,\varphi) = \varphi_1 . e^{-t} + \varphi_2 . e^{-2t}$$

échantillonné aux points $t_1 \ldots t_P$, on aura le tableau des données :

variables →	$X_1 \to t_1$	$X_2 \to t_2$	$X_P \to t_P$
observations 1	$a_1 e^{-t_1} + b_1 e^{-2t_1}$	$a_1 e^{-t_2} + b_1 e^{-2t_2}$	$a_1 e^{-t_P} + b_1 e^{-2t_P}$
	$a_2 e^{-t_1} + b_2 e^{-2t_1}$	$a_2 e^{-t_2} + b_2 e^{-2t_2}$	$a_2 e^{-t_P} + b_2 e^{-2t_P}$
	⋮	⋮	⋮
N	$a_N e^{-t_1} + b_N e^{-2t_1}$	$a_N e^{-t_2} + b_N e^{-2t_2}$	$a_N e^{-t_P} + b_N e^{-2t_P}$

mais qui peut encore s'écrire :

$$\begin{bmatrix} a_1, b_1 \\ a_2, b_2 \\ \text{----} \\ a_N, b_N \end{bmatrix} \times \begin{bmatrix} e^{-t_1} & e^{-t_2} & \text{----} & e^{-t_P} \\ e^{-2t_1} & e^{-2t_2} & \text{----} & e^{-2t_P} \end{bmatrix}$$

$$\dim(N \times 2) \times \dim(2 \times P) = N \times P$$

mais dont le rang sera $\leqslant Min(N, 2, P) = P$

et dans ce cas la dimensionalité linéaire sera égale à la dimensionalité intrinsèque parce que la fonction génératrice est linéaire sur les paramètres.

② Si maintenant on prend un générateur tout aussi simple :

$$F(t, \varphi) = \varphi_1 \, e^{-\varphi_2 \cdot t} \quad (1)$$

avec

$$F_i(t, \varphi) = \{ a_i.e^{-b_i t_1}, a_i.e^{-b_i t_2}, \text{-------}, a_i e^{-b_i t_P} \}$$

la dimensionalité linéaire sera beaucoup plus grande que 2, car on ne peut plus factoriser la matrice des données. De même si on prend en plus dans (1) la relation $\varphi_2 = g(\varphi_1)$ alors $k = 1$, mais l reste inchangé si g est non linéaire.

③ Un autre exemple couramment employé en simulation consiste à utiliser un générateur déterministe de nombres aléatoires (méthode de congruence) utilisant 1 seul paramètre $\varphi(i)$ pour passer d'un signal à un autre : le premier nombre impair fourni pour démarrer la série.

La matrice de données est donc constituée par N lignes de P nombres aléatoires indépendants :

	t_1	t_2	-------	t_P
$k_1 \longrightarrow$	$k_1(1)$	$k_1(2)$	- - - - -	$k_1(P)$
$k_2 \longrightarrow$	$k_2(1)$	$k_2(2)$	- - - - - - -	$k_2(P)$
$k_N \longrightarrow$	$k_N(1)$	$k_N(2)$	- - - - -	$k_N(P)$

et si $k_i \neq k_j \quad \forall i \neq j$ et k_i impair $\forall i$ alors la dimension linéaire tend vers P tandis que la dimension intrinsèque est égale à 1.

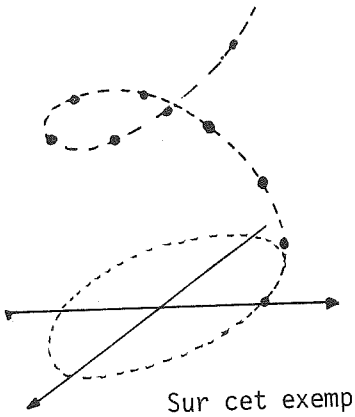
On se dirige donc vers une notion de dimensionalité qui cherche non pas le nombre l de paramètres linéairement indépendants dans l'ensemble des P variables, mais plutôt le nombre de paramètres qui interviennent, linéairement ou non, dans la fonction génératrice faisant passer d'un signal à l'autre pour engendrer les N signaux.

Dans $\mathbb{R}^N$, cela ne revient plus à chercher le sous-espace de dimension l tel que la projection orthogonale des variables soit satisfaisante, car cette transformation est linéaire.

Cela revient plutôt à chercher si le lieu des points est une courbe de $\mathbb{R}^N$ ($k = 1$), ou une surface ($k = 2$), ou un volume k-dimensionnel .

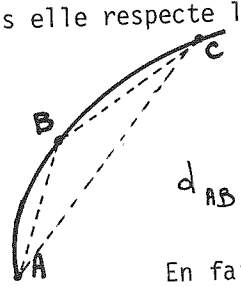
④ Si on génère les 3 variables X_1, X_2, X_3 par :

$$X_1 = \varphi \qquad X_2 = \sin \varphi \qquad X_3 = \cos \varphi$$



sur un échantillon $\varphi_1, \varphi_2, \ldots \varphi_N$, la dimension linéaire sera égale à 3 alors que les points sont répartis sur un arc d'hélice donc à 1 paramètre indépendant.

Sur cet exemple, l'idée semble naturelle d'étirer cet arc d'hélice de façon à le transformer en un segment de droite. Cette transformation n'est pas linéaire, mais elle respecte l'ordre des distances



$d_{AB} < d_{BC} < d_{AC}$

$d^*_{AB} < d^*_{BC} < d^*_{AC}$

En fait, c'est la seule contrainte, qui exprime que la fonction que l'on applique sur les distances doit être monotone, pour que $d_{ij} < d_{kl} \Rightarrow f(d_{ij}) < f(d_{kl})$

La recherche de la dimensionalité peut alors se ramener à la recherche de la fonction f monotone qui permet de minimiser la dimension de l'espace sans violer la relation d'ordre des interdistances.

Il n'est toutefois pas prouvé que, ce faisant, on trouve le nombre exact k de paramètres de la fonction génératrice, mais on trouve de toutes façons $k' < l$ (cf BENNETT, 1969).

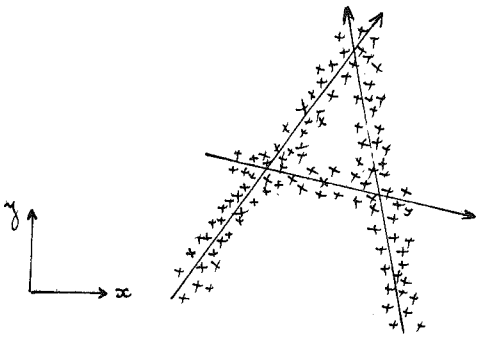
Par exemple, dans l'exemple 2 ou 4 on trouvera bien $k = 2$ ou $k = 1$. Par contre, pour l'exemple 3 il est peu probable que l'on trouve que les P points variables constituent une courbe dans $\mathbb{R}^N$.

Un exemple d'algorithme est celui proposé par SHEPARD (1962) et repris par d'autres auteurs (SHEPARD et CAROLL, 1966 ; BENNETT , 1969) . Il consiste à trouver empiriquement la fonction f en perturbant progressivement la matrice d'interdistance (ou de corrélation) de façon à abaisser son rang sans modifier trop sensiblement l'ordre initial des interdistances. Nous ne l'avons pas appliqué aux problèmes qui nous intéressent, mais il semble bien adapté au traitement des séries chronologiques. Il pourrait, par exemple, servir dans l'analyse des hydrogrammes de crues de fonte de neige (cf G. MORIN, 1974).

1.2.2. Dimensionalité locale - liaison avec l'analyse factorielle typologique (*)

Une dernière notion est celle de dimensionalité locale, développée par DIDAY E. (1973) dans le cas des ensembles de données hétérogènes. Cette méthode se rattache aux techniques d'aggrégation et d'analyse discriminante. Elle consiste à rechercher des noyaux de forte densité dans le nuage global et à rechercher la dimensionnalité linéaire de chacun de ces sous-nuages.

Exemple : Une application particulièrement adaptée est celle de l'analyse des caractères alphabétiques. Si un analyseur d'images fournit les coordonnées x et y des points sombres de la lettre A, il est clair que la dimensionnalité globale est 2, mais que l'on peut trouver 3 sous-ensembles de dimensionnalité 1.



On reverra cet algorithme dans la 5ème partie, mais il est clair que l'on a ici une notion de dimensionalité locale, ou conditionnelle (car liée à l'appartenance à un sous-ensemble).

CHAPITRE II

SELECTION ET ELIMINATION DE VARIABLES A PARTIR DE VISUALISATIONS OU DE NOTIONS DE DISTANCE

Comme nous l'annoncions en introduction, ce chapitre concerne l'aggrégation et la sélection de variables à partir de critères de similarité ou de distance entre variables, celle-ci étant souvent déduite de leur coefficient de corrélation.

Le premier paragraphe décrit un certain nombre de méthodes semi-empiriques ou manuelles tandis que le second concerne une approche purement algorithmique par classification hiérarchique.

Ces questions ont déjà été abordées dans le passé, par exemple par des géologues (in P. LAFITTE et al., 1972) qui, sous le nom d'"analyse des grappes", recherchaient les paquets existant dans un ensemble de variables. Ils se limitaient pourtant à l'utilisation des techniques de classification hiérarchique. Une étude plus complète a été menée par I.T. JOLLIFE (1972 et 1973), à partir d'un article de BEALE et al. (1967), et a servi de point de départ à la nôtre .

II.1 - Méthodes de visualisation - Méthodes utilisant les résultats d'une A.C.P.

Il s'agit ici donc de méthodes non décisionnelles, qui n'effectuent pas le choix des variables selon un algorithme précis, mais fournissent plutôt une certaine image des ressemblances entre variables, laissant à l'utilisateur le soin de les choisir.

II.1.1. Visualisation des ressemblances entre variables

a) Une première méthode consiste à permuter les lignes et colonnes de la matrice de corrélation de façon à regrouper les variables qui se ressemblent. Ceci est relativement simple quand on a peu de variables et des paquets relativement indépendants. Ce n'est plus très simple quand toutes les variables sont fortement correlées (cf IVème Partie - Cas des réseaux avec un fort effet de taille) ou quand leur nombre est trop élevé.

b) Toutes les méthodes d'analyse de données vues dans la 2ème partie et appliquées aux variables permettent de visualiser leurs proximités. L'indice de similarité n'est pas toujours le coefficient de corrélation, comme dans l'analyse des correspondance sur tableau disjonctif complet ou dans la méthode d'Andrews.

L'analyse en correspondance s'applique assez bien à des variables continues correspondant aux diverses stations d'un réseau. C'est le cas par exemple des températures des mois d'été, évoqué dans la IIème Partie (Chap.II-4 et aussi BOIS Ph., 1976). Elle

met en évidence la partie très homogène du réseau et les variables (stations) qui s'en écartent. Mais cela est plutôt utile en critique des données (recherche des anomalies) qu'en optimisation. En effet si l'ensemble de variables est constitué de plusieurs paquets homogènes, il faut considérer plusieurs systèmes d'axes (F1-F2, F1-F3, ... etc) et cela devient vite inextricable.

La méthode d'ANDREWS a l'avantage de représenter les variables dans un seul plan. Par contre, la représentation ne consiste plus en 1 point ou en une trajectoire simple comme en A.F.C. et il devient parfois difficile de les comparer quand elles n'ont pas de caractéristiques marquées. De plus, quand leur nombre est assez élevé, il faut un dispositif interactif pour constituer des groupes homogènes.

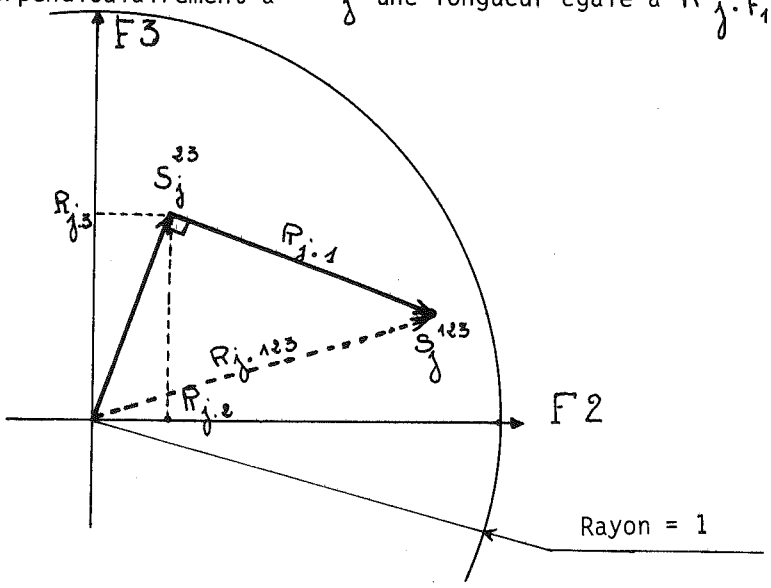
Une fois ces groupes constitués, on peut chercher à les représenter par un nombre réduit d'éléments bien représentatifs.

c) On a vu aussi que l'A.C.P. permettait de représenter les proximités entre variables. Malheureusement celle-ci se fait dans des plans factoriels successifs et même si leur nombre est réduit (5 ou 7, par exemple) il est difficile là aussi de conclure. Ceci se produit surtout pour les réseaux, quand les axes successifs ont des valeurs propres très voisines ce qui interdit de privilégier un plan factoriel plutôt qu'un autre.

La méthode de SAMMON, en dépit de certaines faiblesses sur le plan théorique, a l'avantage de prendre en compte les distances "vraies" de $\mathbb{R}^N$ et non des projections successives, et d'en fournir une représentation plane.

c) Les résultats de l'A.C.P. sont quand même utilisables si l'on arrive à concentrer dans une représentation plane des informations provenant de plusieurs plans factoriels. Une méthode, proposée par D. DUBAND ( ) et plus spécialement adaptée aux données de réseaux consiste en ceci :

- se mettre dans le plan F2/F3 car F1 est souvent associé à un effet de taille, et les valeurs propres λ_2 et λ_3 sont en général voisines. Soit S_j^{23} le point station.
- porter perpendiculairement à OS_j^{23} une longueur égale à $R_{j.F_1} = \sqrt{\lambda_1} . V_{1j}$



- On obtient un point S_j^{123}. Compte tenu de l'orthogonalité des facteurs 1, 2, 3, on a :

$$R^2_{j.123} = R^2_{j.1} + R^2_{j.2} + R^2_{j.3}$$

et par construction, OS_j^{123} représente donc le coefficient de corrélation multiple de X_j avec les 3 premiers axes. On peut poursuivre, par exemple en portant perpendiculairement à OS_j^{123} les coordonnées de X_j sur F4 puis sur F5.

L'intérêt de représenter F1 est de montrer la force de l'effet de taille sur la variable X_j. D'autre part, on voit que l'on se rapporche de plus en plus du cercle unité car $R^2_{j.123....k} \longrightarrow 1.$, ce qui permet de mesurer la qualité de la représentation.

Quand les effets de taille sont tous comparables sur F1 et qu'il reste peu de variance au-delà de F3, on peut aussi tracer des ellipses de densité dans F2/F3 et chercher le nombre et les groupements de variables les plus homogènes.

II.1.2. Méthodes "algorithmiques" utilisant les résultats d'une A.C.P.

I.T. JOLLIFE propose 4 méthodes (notées B1, B2, B3, B4), dont la première, itérative nécessite plusieurs analyses en C.P. Nous en reparlerons en II-3.

La méthode (B3) consistait à rejeter les variables dont la corrélation multiple avec les $P-k$ dernières composantes étaient les plus fortes, ou à garder celles dont les corrélations multiples avec les k premières composantes étaient les plus fortes. On voit bien que si un paquet est bien intercorrelé et a une taille importante, il définira les premiers facteurs. Toutes ses variables seront bien correlées avec ceux-ci et on gardera toutes les variables du paquet ...! ce qui est contraire au but recherché. I.T. JOLLIFE en donne une démonstration rigoureuse pour une matrice dont les blocs diagonaux seraient des matrices d'équi-corrélation, bordées de 0 (cf aussi IIème Partie, I.3.2).

Nous nous limiterons donc aux 2 méthodes restantes.

(B2) On associe à chacune des $P-k$ dernières composantes la variable qui lui est la mieux correlée, et on la supprime.

(B4) On associe à chacune des k premières composantes la variable la mieux correlée n'ayant pas déjà été retenue, et on la garde.

Après des essais sur des exemples simulés, JOLLIFE conclut à une légère superiorité de (B2). Sur des exemples concrets, (B2) et (B4) sont sensiblement équivalentes.

II.1.3. Application à un exemple simple : la piézométrie du bassin de l'Hallue (*)

Nous avons repris une étude déjà abordée par M. CANCEILL (1972) sur les données piézométriques du bassin de l'Hallue. Le nombre de données assez réduit (24 variables x 60 données mensuelles de 1966 à 1970 inclus) et les conclusions déjà

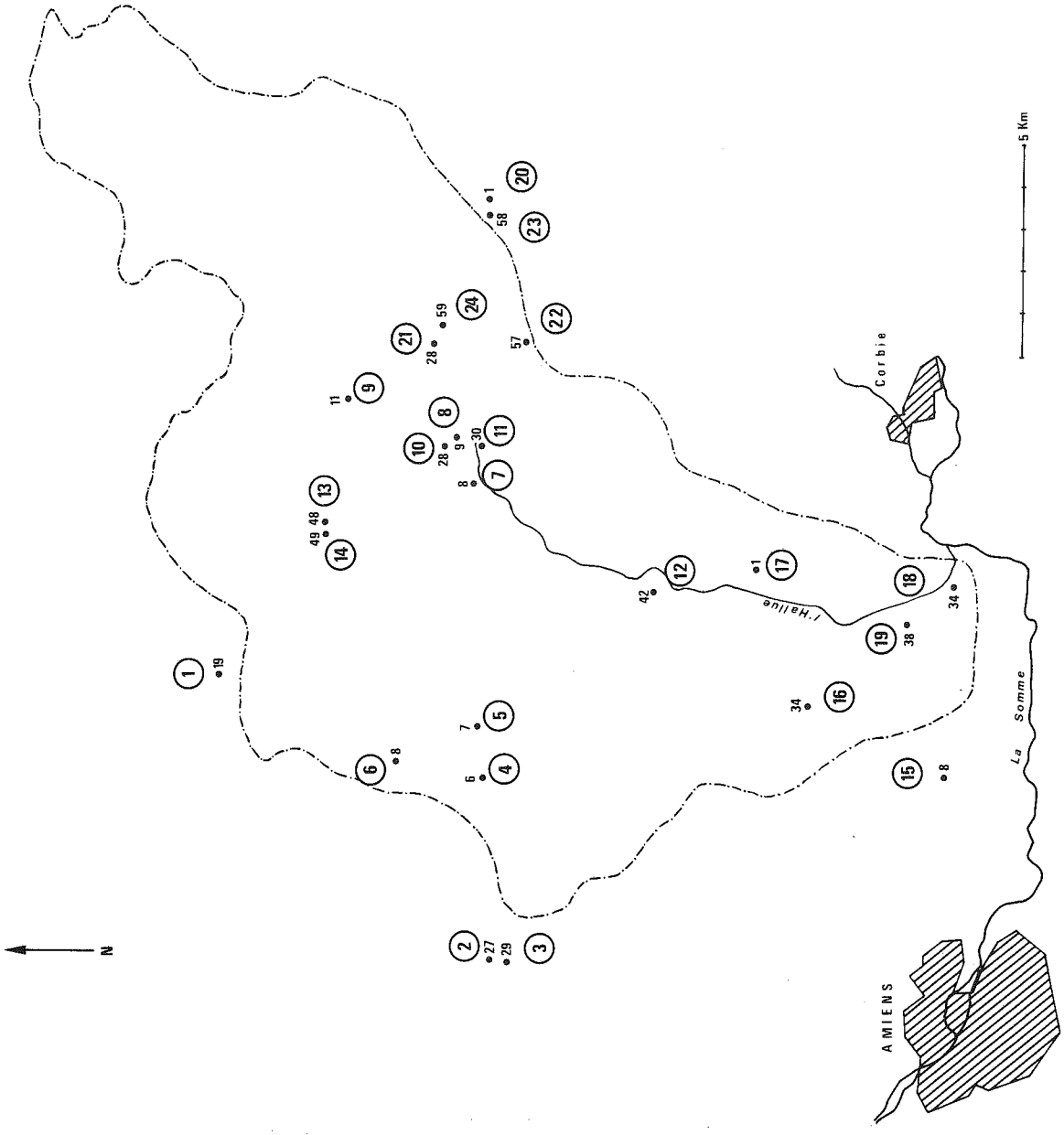


FIGURE III-7 : Carte du bassin de l'Hallue et positions des stations utilisées .

obtenues (cf rapport B.R.G.M.) nous ont permis de tester et de comparer les diverses méthodes décrites ci-dessus.

On trouvera sur la Fig.III-7 la carte du bassin et les localisations des piézomètres.

Les méthodes présentées ci-dessus conduisent aux conclusions suivantes.

(a) La simple observation de la matrice de corrélation (Tableau IV) met déjà en évidence un important paquet de variables intercorrelées: 1 4 5 6 7 8 9 10 11 12 13 14 16 19 20 21 23 24 , qui se subdiviserait éventuellement en 2.

Les variables: 2 , 3 , 17 , 18 et 22 ont des comportements particuliers.

Or, dans l'analyse en C.P. des données, le spectre décroît très rapidement et l'application des méthodes du Chap.I nous conduiraient à retenir 3 à 5 facteurs significatifs.

Et si on se limite à 1 variable associée à chaque facteur, donc à 3 variables, on ne prendra même pas en compte toutes les anomalies signalées.

(b) Si on regarde les projections classiques en A.C.P. (Fig.III-8 a), on observe un groupement important sur l'axe 1, à l'exception de 2 , 3 , 18 , 22 et 17 . De ce fait les axes F2 et F3 ne permettent pas de bien distinguer au sein du groupement dominant mais servent à représenter en fait ces autres influences.

Par contre, après élimination de ces 5 "anomalies" on arrive à redécouper le groupement, très homogène sur F1, en 4 groupes (Fig.III-8 b)

15 16 19	1 5 6 12
4 7 8 9 11 13 14	10 20 21 23 24

(c) L'analyse en correspondance sur les données codées en 3 classes met en évidence le groupement dominant, dont se distinguent, par ordre décroissant (Fig. III-9 a) :

3 18 17 2 22 15 19 16

La méthode de SAMMON, pour la meilleure convergence obtenue (Fig.III-9 b) isole :

2 3 15 17

un peu 18 22 et groupe 7 11 12 16 19 puis 8 10 23 24
et 1 4 5 6 9 13 14 20 21

La méthode d'ANDREWS se révèle déjà difficile à exploiter car on a du mal à comparer des courbes finalement assez voisines. On retrouve, plus qu'on ne trouve, les résultats déjà obtenus.

(d) Si on associe des variables aux premiers facteurs, on retiendra les variables (cf Tab. V)

(B4) 3 9 10 17 18 22

TABLEAU IV - Piezométrie de l'Hallue
Matrice de corrélation

1	1.0																						
2	.31	1.0																					
3	-.23	.59	1.0																				
4	.87	.60	-.9	1.0																			
5	.95	.47	-.17	.97	1.0																		
6	.96	.39	-.20	.94	.98	1.0																	
7	.85	.60	.10	.86	.90	.85	1.0																
8	.87	.38	-.20	.88	.92	.89	.88	1.0															
9	.93	.38	-.25	.92	.96	.94	.89	.96	1.0														
10	.84	.34	-.15	.84	.87	.84	.86	.91	.91	1.0													
11	.88	.49	-.16	.92	.92	.89	.85	.87	.94	.82	1.0												
12	.95	.37	-.27	.92	.97	.95	.86	.92	.94	.84	.90	1.0											
13	.81	.49	-.11	.94	.96	.93	.96	.94	.98	.92	.92	.93	1.0										
14	.93	.43	-.22	.95	.97	.95	.89	.94	.99	.91	.95	.94	.98	1.0									
15	.84	.32	-.06	.72	.81	.85	.78	.72	.77	.62	.73	.83	.77	.75	1.0								
16	.89	.46	.01	.83	.89	.90	.87	.79	.82	.73	.80	.91	.85	.82	.94	1.0							
17	.68	-.05	-.48	.56	.65	.67	.50	.60	.68	.53	.58	.68	.60	.63	.62	.57	1.0						
18	.25	-.03	.03	.05	.16	.22	.18	.13	.13	.07	.15	.24	.10	.10	.58	.41	.33	1.0					
19	.89	.17	-.23	.72	.84	.88	.73	.75	.80	.65	.73	.87	.76	.78	.94	.91	.73	.54	1.0				
20	.84	.41	-.27	.90	.90	.85	.86	.91	.96	.88	.92	.89	.94	.95	.67	.74	.68	.06	.69	1.0			
21	.88	.40	-.26	.90	.92	.88	.88	.94	.98	.89	.94	.91	.95	.97	.70	.77	.65	.09	.73	.97	1.0		
22	.59	.30	-.03	.50	.57	.55	.62	.69	.62	.48	.48	.60	.58	.56	.55	.58	.44	.29	.60	.56	.62	1.0	
23	.86	.38	-.29	.90	.91	.86	.87	.92	.97	.89	.93	.90	.95	.96	.68	.75	.67	.05	.71	.99	.99	.58	1.0
24	.83	.34	-.31	.86	.88	.84	.84	.91	.96	.85	.91	.87	.92	.93	.68	.71	.73	.11	.71	.97	.97	.56	.98	1.0

TABLEAU V - Variables associées (i.e.la plus corrélée)
aux facteurs principaux .

F 24	9	-.6345	-.0156
F 23	13	-.5823	-.0165
F 22	23	.6874	.0243
F 21	5	.6045	.0233
F 20	21	.4353	.0235
F 19	6	.5125	.0361
F 18	16	.4171	.0376
F 17	11	-.4663	-.0446
F 16	12	-.4577	-.0512
F 15	19	.5782	.0872
F 14	20	.5300	.0838
F 13	1	-.4718	-.0888
F 12	8	-.5469	-.1134
F 11	7	-.5177	-.1289
F 10	15	-.4275	-.1092
F 9	10	.2680	.0863
F 8	14	.3965	.1423
F 7	11	-.2660	.0850
F 6	10	-.4474	-.2518
F 5	17	-.3936	.2665
F 4	22	.8479	.7011
F 3	18	.6192	.8292
F 2	3	.5753	.8146
F 1	9	.2332	.9870
k	j	V_{kj}	ρ_{kj}

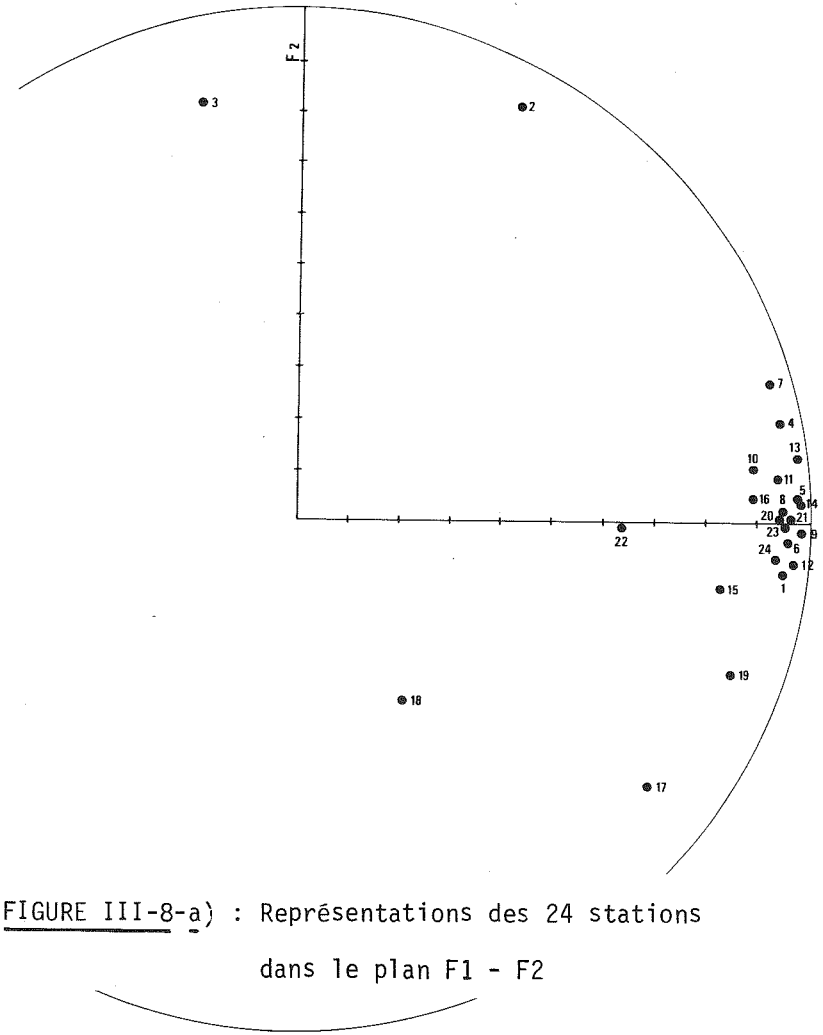
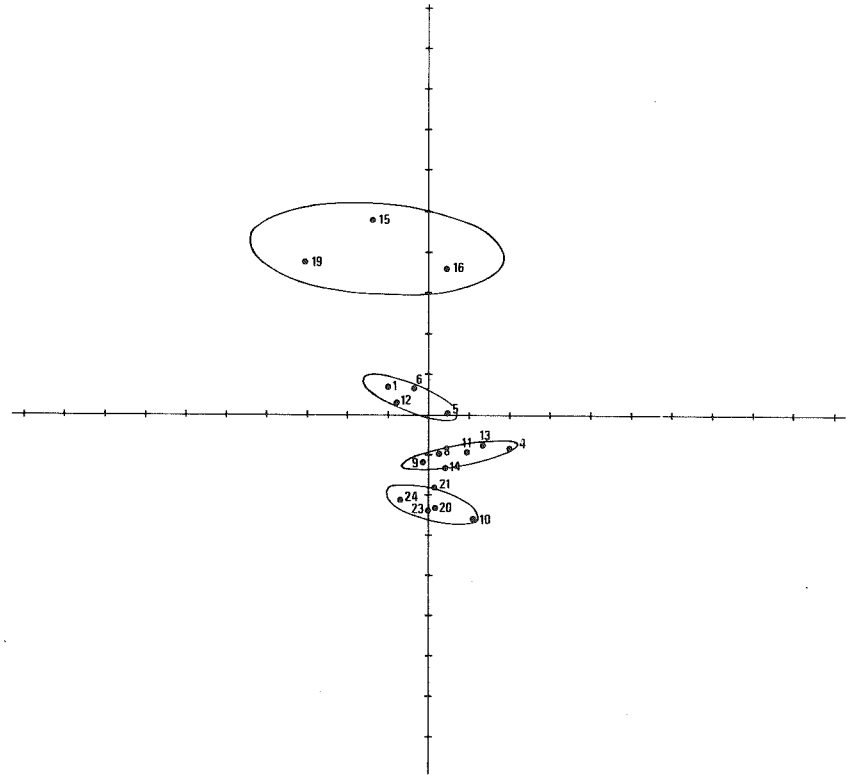


FIGURE III-8-a) : Représentations des 24 stations dans le plan F1 - F2



b) : Groupements dans F2 - F3 après élimination des stations 2 / 3 / 17 / 18 / 22 .

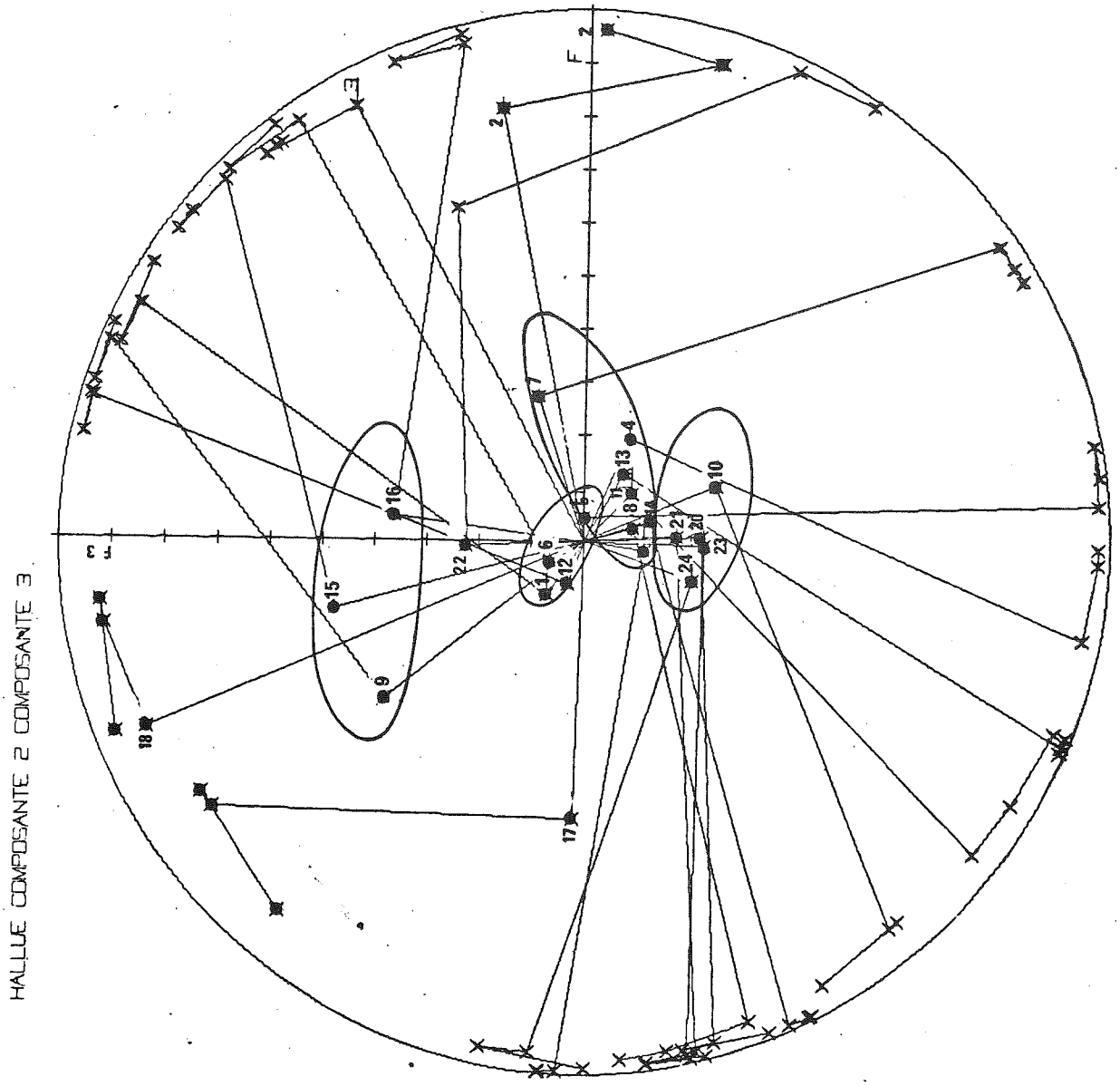
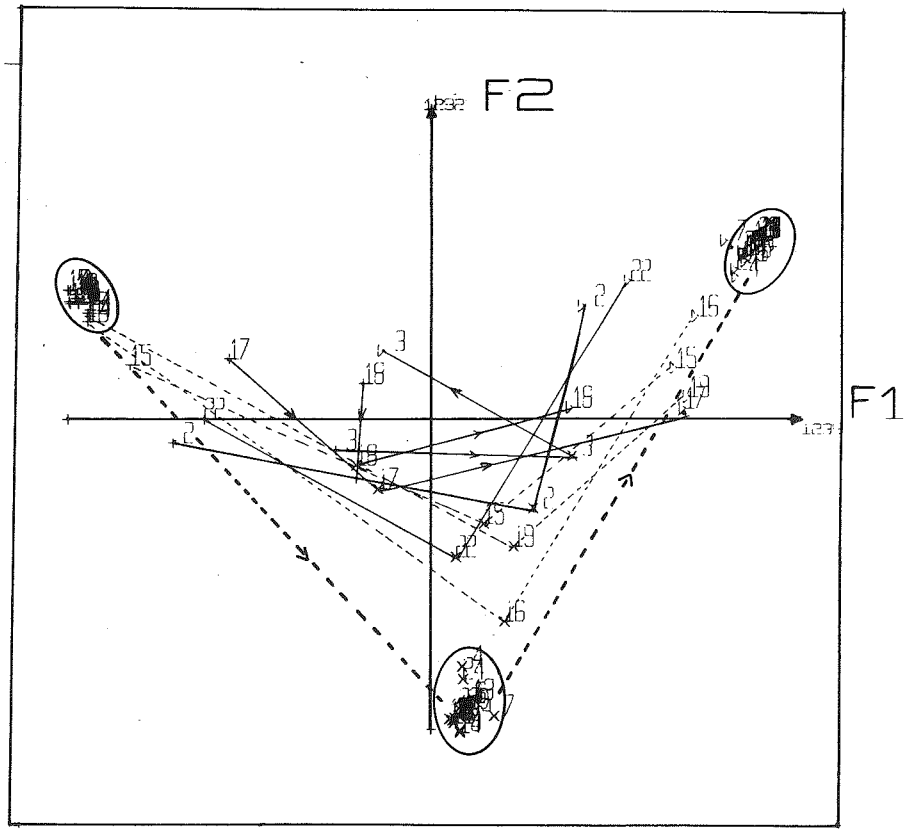
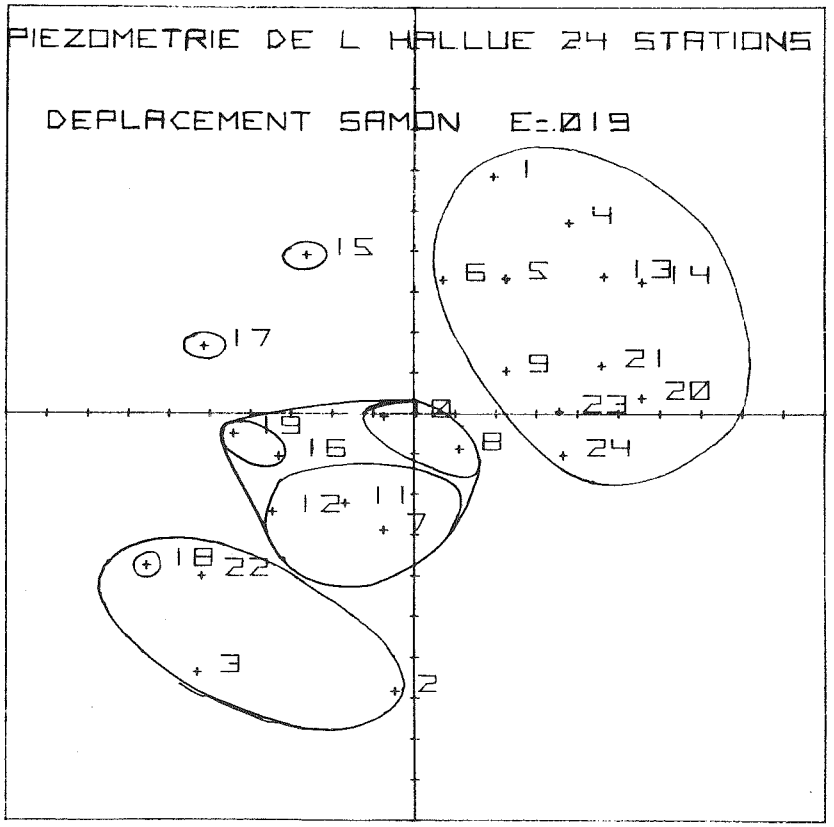


FIGURE III-8-c): Représentation des 24 stations dans F2-F3
avec en plus la corrélation avec F1,F4,etF5.





- →-- Trajectoire moyenne des variables "normales " .

---→-- Trajectoire "légèrement anormale"

——→ Trajectoire "anormale"

FIGURE III - 9 - a) : ANALYSE EN CORRESPONDANCE DES 24 STATIONS APRES CODAGE DISJONCTIF EN 3 CLASSES .

III - 9 - b) : REPRESENTATION NON-LINEAIRE (methode de SAMMON)

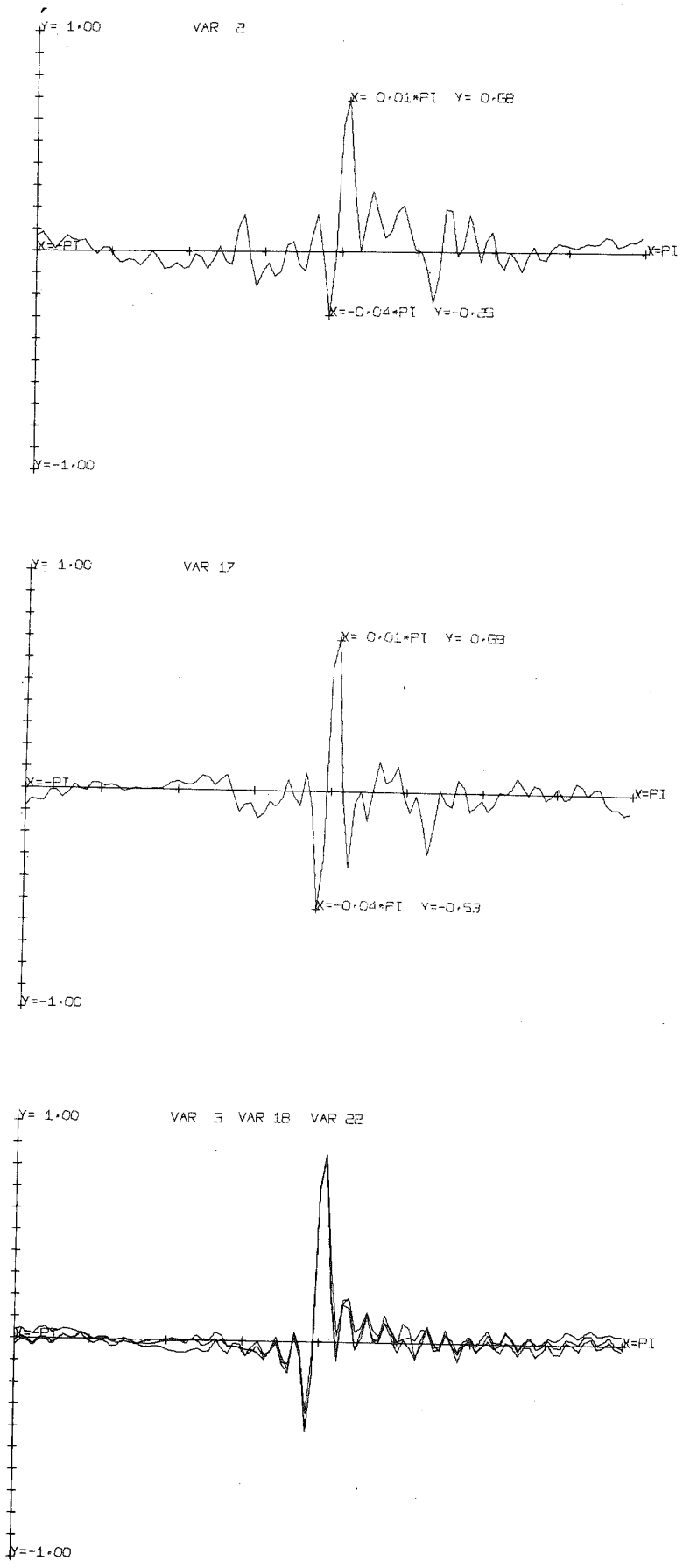
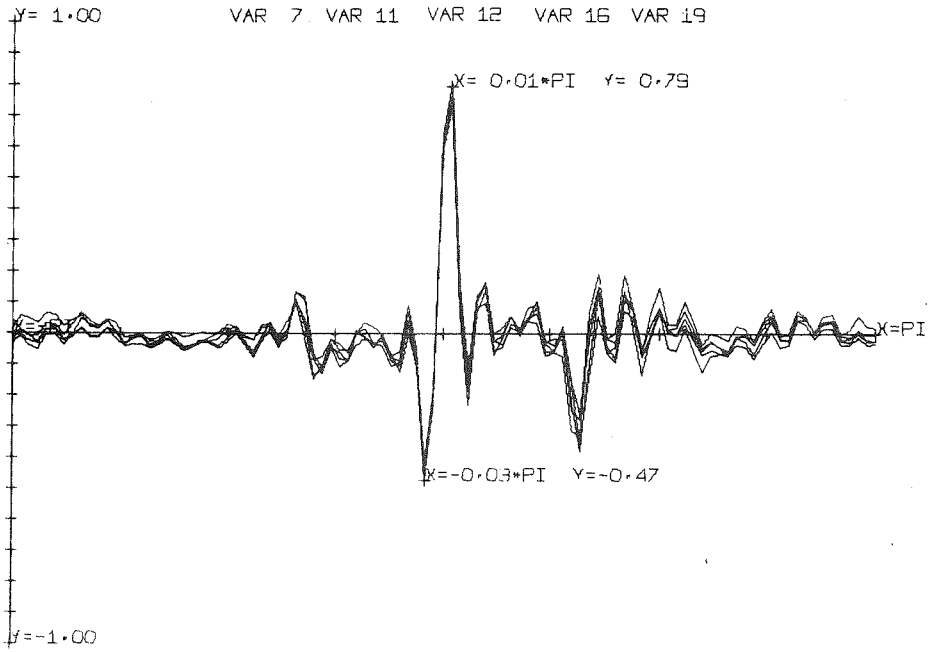


FIGURE III - 10 - a) : Méthode d'Andrew (Stations "anormales")

PIEZOMETRIE DE L HALLUE 24 STATIONS REPRESENTATION COSINUS



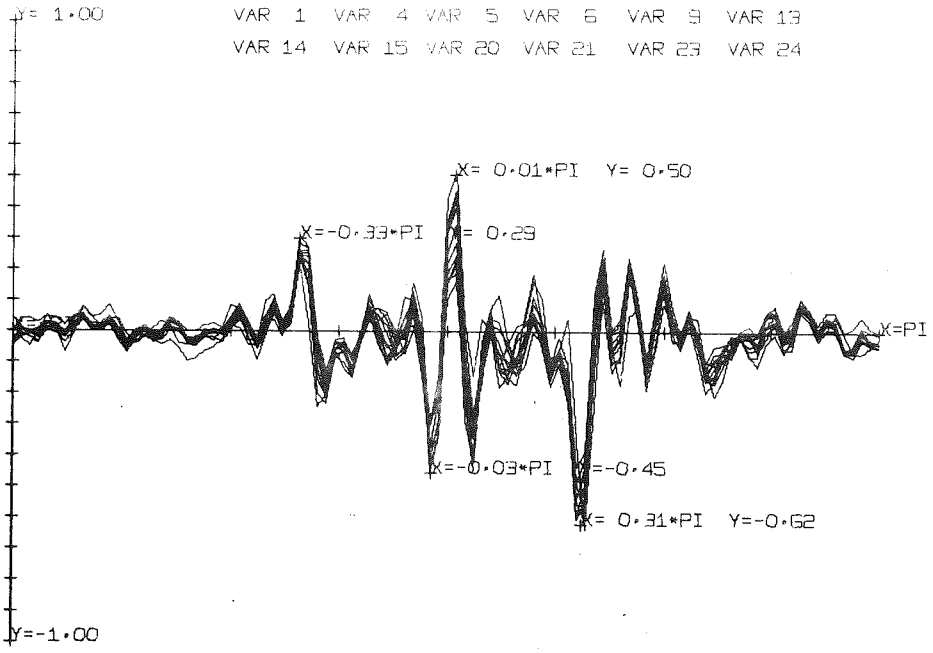


FIGURE III - 10 - b) : Méthode d'Andrew (groupements homogènes)

tandis que si on associe des variables aux facteurs de plus faible variance et qu'on les élimine, il reste :

(B2) 2 3 4 18 22 24

Un avantage reste à (B2), celui d'avoir sélectionné la 2 qui est très particulière.

On constate quand même que, pour 7 variables sélectionnées, (B2) donne les 5 variables particulières plus 1 dans chacun des groupements les plus importants.

Une remarque pourtant, met en évidence le caractère non absolu des critères choisis.

La première variable retenue par (B4) est la 9 , qui est aussi la première à être rejetée par (B2) ! D'autre part, certaines variables sont parfois quasiment ex aequo, par exemple sur le facteur 1. On pourrait l'éviter en appliquant un algorithme du type Varimax (rotation orthogonale-cf. COOLEY et LOHMES, 1971, p.144) pour maximiser les poids des variables les plus fortes sur chacun des facteurs.

(e) Conclusions sur les méthodes :

- Les méthodes graphiques déduites de l'A.C.P. détectent bien les variables particulières mais si un groupement dominant émerge, il vaut mieux en refaire une A.C.P. pour le subdiviser.

- Les méthodes qui visualisent les variables par des points (A.C.P. classiques ou SAMMON) sont préférables à celles qui les visualisent par des trajectoires. En effet, l'A.F.C. sur les données codées en 3 classes permet tout au plus de mettre en évidence les "anomalies". Quant à la méthode d'ANDREWS elle permet au mieux de vérifier les autres résultats, et ne présente que peu d'intérêt en elle-même.

- Les méthodes de sélection sur une analyse en C.P. donnent des résultats cohérents, avec une préférence pour la méthode (B2) . Par contre, il n'est absolument pas évident que le nombre de facteurs significatifs soit correctement choisi (3 dans notre cas) quand il y a un effet de taille important qui biaise les pourcentages d'inertie. Dans ce cas, il faudrait même envisager de l'enlever complètement (cf IVème Partie, Chap.II).

- Il est un peu choquant aussi que l'avantage aille à (B2) alors qu'il y a instabilité des derniers facteurs si l'on change d'échantillon. En fait, on verra sur un exemple que si les facteurs sont instables, l'ensemble des variables associées à l'ensemble des derniers facteurs est assez stable.

Par contre, la façon dont 1 variable se disperse sur différents facteurs est toujours assez délicate à interpréter.

Pour terminer, quelques remarques sur les résultats de cet exemple particulier :

- les variables présentant des comportements singuliers, en particulier 18 et 22 ont des variances très faibles par rapport aux autres variables. On peut

alors se demander s'il ne s'agit pas de variables quasi-constantes affectées d'un bruit de mesure.

- Les variables 2 et 3 sont un peu dans le même cas alors que les stations les plus proches, 4 5 et 6 ont des écart-types 4 à 5 fois plus grands.

- Cette dispersion des fluctuations conduit à rejetter une analyse sur matrice de corrélation telle que nous l'avons pratiquée à titre d'exemple, pour faire une analyse sur matrice de variance-covariance. Dans ce cas, la méthode (B2) garderait :

1 3 6 7 13 15 24

qui, par leur distribution, représentent assez correctement le réseau.

II.2 - Méthodes algorithmiques

Il s'agit cette fois de méthodes décisionnelles qui, à l'aide d'un algorithme et de quelques paramètres nécessaires à leur fonctionnement, fournissent une batterie de variables sélectionnées parmi l'ensemble de départ.

La plupart de ces techniques ressortent de la classification hiérarchique, qui peut se séparer en 2 grandes familles :

- les méthodes de segmentation où l'ensemble de départ forme initialement un tout que l'on subdivise progressivement, sans remettre en cause les subdivisions antérieures (méthode descendante)
- les méthodes d'aggrégation, où les individus de départ, considérés initialement comme indépendants, se regroupent progressivement en classes qui elles-mêmes peuvent fusionner jusqu'à ne plus en former qu'une seule (méthode ascendante).

Ces techniques ayant fait l'objet de nombreux développements, nous renvoyons aux synthèses existantes (par exemple : dans P. LAFITTE, 1972 ; ou ZIRPHILE, 1974).

II.2.1. Aperçu général

a) Rappel des méthodes

Les méthodes utilisées en analyse des grappes appartiennent au second groupe (aggrégation, ou classification hiérarchique ascendante) puisque l'on part d'éléments isolés que l'on veut regrouper. Celles-ci supposent définis :

(1) Un critère de distance, ou de similarité entre 2 éléments. Ici il s'agit de variables, dont on mesure la similarité par leur coefficient de corrélation r assimilable à une distance, en fait un carré de distance, dans $\mathbb{R}^N$:

$$d^2(X_1, X_2) = 2 \cdot (1 - r(X_1, X_2))$$

(2) Un critère de distance entre un élément et un groupe, ou entre deux groupes. C'est en partie sur ce critère que les méthodes se distingueront.

(3) Une procédure de fusion de 2 classes voisines. En général on fusionne les 2 plus proches au sens du critère défini en (2) . Dans d'autres méthodes on fusionne toutes celles qui sont plus proches qu'un seuil δ_o défini arbitrairement (Nous ne considèrerons que la première définition).

(4) Un critère d'arrêt de la classification pour éviter de fusionner 2 classes trop éloignées. Et un critère de sélection d'un ou plusieurs éléments représentatifs au sein de chaque groupe de la classification retenue. Nous envisagerons ces derniers problèmes ultérieurement.

En fait, la distinction se fera surtout sur la condition (3) , d'où les méthodes que nous appellerons :

(C1) : Méthode d'affinité singulière (Single linkage) où la distance de 2 classes est la distance séparant les 2 éléments les plus proches n'appartenant pas à la même classe :

Si on a 2 classes A et B , leur distance ou leur similarité sera :

$$R_{A,B} = \underset{i,j}{Max} \left\{ r_{ij} , \; i \in A \quad j \in B \right\}$$

où i et j sont les indices des variables dans A et B .
Et on regroupera les 2 classes D et F telle que :

$R_{D,F} = Max\{R_{AB}\}$ sur tous les couples (A,B) existant parmi ℓ classes.

C'est donc une méthode Max-Max.

(C2) : Méthode d'affinité moyenne (Average linkage). La distance de 2 classes A et B est la moyenne des distances entre tous les éléments de A et tous ceux de B d'où :

$$R_{AB} = \frac{\sum_{i \in A} \sum_{j \in B} r_{ij}}{n_1 \times n_2}$$

n_1 = cardinal (A)

n_2 = cardinal (B)

Et on regroupe les 2 classes qui sont les plus proches, d'où une méthode Max-Moy.

(C3) : Méthode d'affinité complète (Complete linkage). Cette fois la distance de 2 classes est la distance entre les 2 éléments les plus éloignés.

$$R_{AB} = \underset{i,j}{Min} \left\{ r_{ij} , \; i \in A \quad j \in B \right\}$$

La fusion des classes les plus proches en fait une méthode Max-Min.

Parmi leurs avantages et inconvénients, on peut retenir :

- Dans (C1) le fait de ne considérer que les meilleurs coefficients de corrélation peut mener à des groupes serpentiformes (effets de chaînes). Mais cette méthode permet de construire des groupes qui ne soient pas forcément plus ou moins ellipsoïdaux. Dans le cas d'un réseau de mesures, où les variables ont une évolution spatiale continue, on peut ainsi avoir, dans un même groupe, des variables très éloignées

les unes des autres bien que proches 2 à 2. Par contre, toute variable d'un groupe a au sein de celui-ci un voisin plus proche que tout autre élément n'appartenant pas au groupe.

- Dans (C3) , on distingue diverses variantes dans le calcul des similarités, après fusion de 2 classes.

Par exemple, on peut, après avoir décidé de regrouper A et B parce que le minimum, sur(A,B)de r_{ij} $i \in A$ $j \in B$ est maximum parmi toutes les classes prises 2 à 2, calculer la distance de AB à un autre élément (ou classe) C, comme :

$$R_{C,AB} = \text{Min} \left\{ R_{CA}, R_{CB} \right\}$$

ou au contraire :

$$R_{C,AB} = \text{Moyenne sur tous les éléments } r_{ij} \quad i \in C \quad j \in A \cup B$$

ce dernier choix, dû à SORENSEN (1948) (cité dans P. LAFITTE, 1972) évite une chute trop rapide des coefficients de similarité.

Dans le 1er cas on peut définir une enveloppe à la classe A,B, fusion des classes A et B, parce que:

$$\text{Min } r_{ij}(i \in A, j \in B) \leq \text{Min } (A, \{C, D \cdots\})$$

C'est l'intersection des 2 sphères de rayon R_{AB} centrées en A et B, indices associés au maximum.

L'avantage de cette méthode est de donner des classes assez compactes.

- Dans (C2) , où l'on effectue des moyennes de distances les interprétations en terme de borne inférieure ou supérieure ne tiennent plus. D'autre part, le critère retenu pour la similarité entre A et B :

$$R_{AB} = \frac{\sum_{i \in A} \sum_{j \in B} r_{ij}}{n_A \times n_B}$$

peut varier selon les auteurs.

ANDERBERG (1973) propose :

$$R_{AB} = \frac{\sum_{i,k \in A} r_{ik} + \sum_{j,l \in B} r_{jl} + \sum_{i \in A} \sum_{j \in B} r_{ij}}{(n_A + n_B) \times (n_A + n_B - 1)/2}$$

et SOKAL R.R. et SNEATH P.H. (1963) (cités dans P. LAFITTE, 1972) proposent, quand la similarité est exprimée par une corrélation :

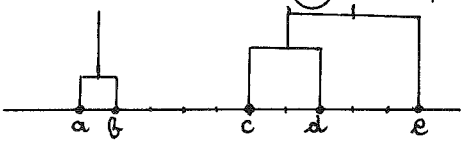
$$R_{AB} = \frac{\sum_{i \in A} \sum_{j \in B} r_{ij}}{\sqrt{n_A + 2 \sum_{i,k \in A} r_{ik}} \times \sqrt{n_B + 2 \sum_{j,l \in B} r_{jl}}}$$

qui prend en compte la "compacité" du groupe.

Expérimentalement, les diverses variantes de (C2) et (C3) donnent des résultats assez voisins.

b) Aspects particuliers aux cas des matrices de corrélation

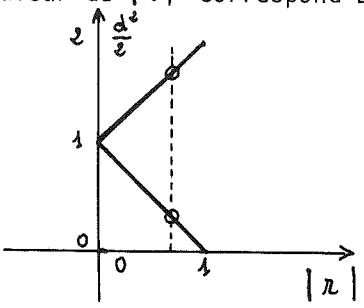
α - Pour appliquer ces méthodes, il suffit que l'on ait défini un coefficient de similarité, ou de dissimilarité, entre d'une part 2 éléments et d'autre part entre 2 classes. La seule contrainte est que le second critère devienne identique au premier quand 2 classes se réduisent à 2 éléments uniques. Toutefois cela ne va pas sans incohérence, comme le montre la méthode (C1) où on peut avoir :



c et e appartenant au même groupe alors que b et c, plus proches, appartiennent à des groupes distincts.

Si on veut de plus une interprétation statistique, ou de manière à peu près équivalente, une interprétation géométrique dans $\mathbb{R}^N$ puisqu'il s'agit de comparer des variables connues sur N observations, il faut que le critère de similarité puisse s'interpréter en terme de distance euclidienne.

Or ce n'est pas le cas ici puisque la similarité entre 2 éléments, ou l'intensité de leur liaison, se traduit par la valeur absolue de leur coefficient de corrélation et qu'à 1 valeur de $|r|$ correspond 2 distances possibles.



L'interprétation n'est donc possible que si tous les coefficients sont positifs et ce sont d'ailleurs les cas qu'ont traités, sans en donner la raison P. ISNARD et al. (in LAFITTE, 1972) et I.T. JOLLIFE (1972) qui ne génère que des variables correlées positivement. Ceci est en général vrai quand on analyse un réseau de mesure d'un paramètre hydrométéorologique assez homogène, mais ce n'est plus le cas dans les mélanges de variables utilisés dans des modèles explicatifs.

Dans ce cas, l'algorithme peut s'appliquer sur le critère :

$$C_{ij} = |r_{ij}|$$

Cependant, si l'on veut préserver l'interprétation géométrique, on peut essayer de faire disparaître, autant que faire se peut, les valeurs négatives de la matrice de corrélation.

Une technique consiste à :

1 - chercher le plus fort coefficient négatif

2 - inverser le signe de l'une des variables auxquelles il est associé: Par exemple celle qui a le plus de coefficients négatifs.

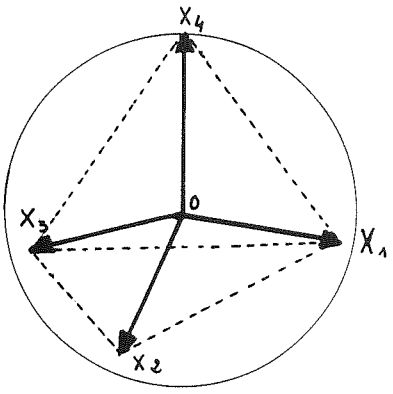
3 - itérer jusqu'à ce que l'on rencontre une impossibilité (le plus fort coefficient négatif qui apparaît est supérieur à celui que l'on fait disparaitre).

Or on peut vérifier, sur des contre-exemples, qu'il n'est pas possible, par de simples changements de signes des variables, de faire apparaître des signes tous positifs dans une matrice de corrélations.

Ou encore, étant donné un polyèdre quelconque inscrit dans une hypersphère de $\mathbb{R}^N$, il n'est pas suffisant de remplacer certains points par leur opposé sur la sphère pour que le polyèdre s'inscrive dans un cône d'angle au sommet $\frac{\pi}{2}$ issu de l'origine.

Exemples :

1 -



La transformation de X_j en $-X_j$ $\forall j$ ne rend pas tous les coefficients > 0.

2 - Ou cette matrice extraite d'un des exemples traités :

1.0				
+.953	1.0			
+.551	+.558	1.0		
-.238	-.283	-.205	1.0	
-.069	-.081	-.066	-.023	1.0

⇓

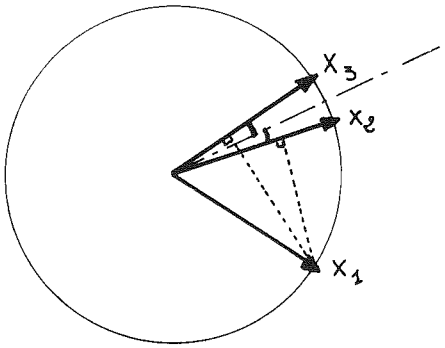
1.0				
.953	1.0			
.551	.558	1.0		
.238	.283	.205	1.0	
.069	.081	-.066	.023	1.0

3 - On a déjà vu l'inconvénient de (C1) pour les groupements qu'il fournit. La méthode (C3) est plus attrayante en ce sens qu'elle donne des paquets homogènes. Dans $\mathbb{R}^N$, le domaine dans lequel s'inscrit une classe est défini en plus par le fait que les variables sont sur la sphère de rayon 1.

Par contre, il n'est pas très satisfaisant de travailler sur les coefficients minimaux. En effet, on en arrive assez vite à envisager des valeurs qui, compte-tenu de l'échantillonnage, ne sont plus très significativement différentes de 0. De même dans les premiers regroupements de l'algorithme, on considère des coefficients en général voisins de 1.0 qui ne sont pas significativement différents les uns des autres. Ici encore, une bonne façon d'apprécier les fluctuations d'échantillonnage consiste à découper l'échantillon en 2 et à effectuer la classification sur les 2 matrices de corrélation ainsi estimées.

γ - Un autre problème provient, dans le cas de la méthode C2 par affinité moyenne, du calcul des similarités entre les 2 classes fusionnées et tout autre élément ou classe. La moyenne de coefficients de corrélation, ou de leur valeur absolue, n'a pas de sens statistique évident.

Dans $\mathbb{R}^N$, cela correspond à une moyenne de cosinus d'angles. Dans le cas le plus simple où l'on calcule la liaison d'une variable X_1 avec une classe de 2 variables (X_2, X_3), on voit que l'on n'obtient pas le cosinus de l'angle de X_1 avec la bissectrice de X_2, X_3, que l'on aurait pu souhaiter obtenir. Cette bissectrice est d'ailleurs la première C.P. de X_2, X_3. Cela peut nous conduire à choisir comme critère de liaison entre 2 classes, le cosinus de l'angle de leurs 1ères composantes respectives, c'est-à-dire un index qui ressemble au coefficient de corrélation canonique ... Nous retrouverons cette notion ultérieurement (Chap.III).

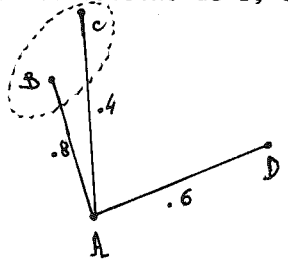


- Pour l'instant, le problème de la liaison d'une variable avec une autre a été plus ou moins bien résolu en prenant la valeur absolue de la corrélation, qui peut s'interpréter en terme de distance. Une façon d'éliminer les problèmes de signes peut consister à prendre les carrés de ces coefficients, qui s'interprêtent en pourcentages de variance commun aux 2 variables.

Malheureusement l'additivité des variances expliquées n'est pas valable si les variables sont intercorrelées. En général :

$$R^2_{1.2,3} \neq r^2_{12} + r^2_{13}$$

Il faut aussi noter que les résultats diffèreront selon que l'on prend le critère $|r|$ ou r^2. En effet l'utilisation des carrés favorise, dans les sommations, les termes voisins de 1, et tend alors à rapprocher la méthode du single-linkage :



Mais

$$d(A,D) = |.6| = .6$$
$$d(A,BC) = \frac{1}{2}(|.8| + |.41|) = .6$$

$$d(A,D) = (.6)^2 = .36$$
$$d(A,BC) = \frac{1}{2}(.64 + .16) = .40$$

A l'inverse les méthodes (C1) et (C3) sont insensibles à toute transformation qui ne modifie pas l'ordre des distances.

Un critère possible exprimant la liaison entre 1 variable X_1 et une classe (X_2, X_3) serait alors le coefficient de corrélation multiple. Le problème est

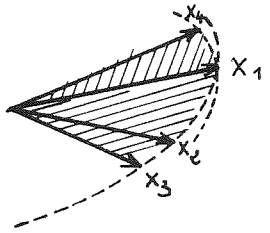
que ce coefficient vérifie:

$$R_{X_1 \cdot X_2, X_3} \geqslant Max \left(r_{X_1, X_2}, r_{X_1, X_3} \right)$$

au point que l'on peut avoir , dans l'exemple ci-dessous :

$$R_{X_1 \cdot X_2, X_3} > r_{X_1, X_4} \quad \text{(c'est le cas si } X_4 \in \text{plan } X_2 X_3)$$

alors qu'il semblerait naturel de regrouper d'abord, au niveau 2, X_1 et X_4 .



En effet, le critère sous-jacent n'est plus alors la distance moyenne aux éléments d'une classe mais la distance à l'hyperplan engendré par cette classe.

Les essais que nous avons réalisés montrent d'ailleurs bien une nette tendance à ne former qu'un seul groupe important, l'effet de chaîne étant accentué par ce critère.

$\mathcal{E}$ - Enfin, étant donné une classification ascendante, les derniers problèmes consistent à définir le niveau auquel on s'arrête, et les variables que l'on choisira pour caractériser les classes formées.

Le nombre de classes à envisager sera déterminé de façon assez arbitraire :

- il sera supérieur au nombre de facteurs considérés comme significatifs dans une A.C.P.
- on cessera de regrouper des classes quand la valeur du critère de similarité tombera en-dessous d'un certain seuil r_0 . Quand il s'agit de corrélation, cela peut correspondre à un seuil de signification compte-tenu de l'échantillonnage .

Sur des données simulées , I.T. JOLLIFE a déterminé empiriquement $r_0 \approx$ 0.5 à 0.6 pour N = 100 observations. Quant à la variable à retenir pour caractériser un paquet, il semble préférable de prendre l'une des premières apparues dans le paquet. Si celui-ci est assez gros, ou a fait l'objet d'une fusion récente, on lui adjoindra une des dernières variables entrées.

II.2.2. Une méthode de classification particulière : le procédé Iphigénie

Il s'agit d'une méthode qui ne ressort pas directement de la classification hiérarchique mais qui s'en rapproche et combine des aspects des méthodes de "single-linkage" et "complete-linkage".

On en trouvera une description détaillée dans Zirphile (1974) ou en annexe, car elle sera utilisée à nouveau dans les autres parties de ce mémoire. Sa particularité essentielle n'est pas l'algorithme d'aggrégation, analogue au "single-linkage" mais le critère d'arrêt : toute distance entre 2 éléments à l'intérieur d'une classe doit rester

inférieure à toute distance entre 2 éléments situés dans 2 classes différentes.

L'algorithme s'arrête à la première contradiction : quand il souhaite fusionner 2 classes proches parce que 2 de leurs éléments respectifs sont les plus voisins à ce stade de l'aggrégation, il vérifie aussi que les éléments les plus éloignés qui vont apparaître au sein de cette nouvelle classe ne le sont pas trop, c'est dire pas plus que tout autre couple d'éléments appartenant à des classes différentes.

Outre une programmation assez simple et l'absence de contradiction (possible avec la méthode (C1) ou (C2)), elle présente les avantages suivants :

1- Elle permet d'ignorer le signe du coefficient de corrélation, puisqu'elle travaille sur les distances seules et même seulement sur leur ordre.

2- Elle indique elle-même le nombre de classes à conserver, ce qui est très indicatif.

Par contre son aspect plus combinatoire (au niveau de la recherche des contradictions) la rend sensiblement plus coûteuse que (C1) ou (C2) en temps de calcul. Comme elle constitue une amélioration de (C1), nous l'appellerons quand même ainsi dans la suite.

CHAPITRE III

ELIMINATION ET SELECTION DE VARIABLES A PARTIR DE NOTIONS DE VARIANCES ET DE REDONDANCES

Les premières approches remontent encore à BEALE et al.(1967) et JOLLIFE (1972) mais nous avons été conduits à les améliorer et surtout à les compléter considérablement, en en précisant les critères et en les reliant à des développements récents de l'analyse des données

III.1 - Méthodes fondées sur les corrélations multiples

III.1.1. Méthodes utilisant les corrélations multiples

La première, que nous appellerons (A1) pour suivre les notations de JOLLIFE, consiste, si l'on veut garder K variables parmi P, à extraire le sous-ensemble de K variables tel que, pour les $P-K$ variables restantes, on ait :

$$\left[\underset{\substack{\ell \in P-K \\ \text{variables éliminées}}}{\text{Min}} \; R_{\ell.\{K\}} \right] \text{Maximum}$$

Donc parmi tous les sous-ensembles de $P-K$ variables, on choisit celui où la variable la moins bien reconstituée à partir de l'information retenue l'est pourtant le mieux. On se garantit donc sur la reconstitution minimum.

Ce critère Max-Min nous semble le plus réaliste pour l'utilisateur, par exemple dans le cas d'un réseau de mesures où l'on arrête certaines stations, sans savoir si on n'en aura pas à nouveau besoin quelques années après, ni laquelle il faudra éventuellement réutiliser. Ce critère Max-Min est donc le mieux adapté. Malheureusement, si on a beaucoup de stations, donc P élevé, et si on veut en retrancher beaucoup (10 à 20 %) le nombre d'essais est particulièrement élevé, puisqu'égal à C_P^{P-K}. S de plus on ne connaît pas K a priori, ou seulement approximativement, il faut faire plusieurs essais et les groupes de variables à $K-1$, K, $K+1$ ... éléments ne se recouvrent pas forcément. D'où un aspect combinatoire très coûteux en temps de calcul. (Dès que P dépasse 30 ou 50, le temps de calcul peut s'exprimer en heures, même sur un gros ordinateur).

Dans ce cas, la démarche classique est celle du pas à pas ("Stepwise") et une première méthode consiste ici (méthode (A2)) à éliminer la variable la plus corrélée aux autres, soit X_j, puis à chercher, parmi les $\{ P \text{ variables} - X_j \}$ la variable X_ℓ la mieux corrélée aux autres, soit $\{ P_{\text{variables}} - X_j - X_\ell \}$ et à l'éliminer.

Le critère d'arrêt peut être un seuil sur R_{Min} compte tenu de l'échantillonnage.

De façon plus précise, on procède comme suit :

1er pas : On élimine X_k tel que : $k \longrightarrow \underset{l \in E_0}{Max} \left\{ R_{l.\{E_0-l\}} = R_{l.1,2,\ldots l-1,l+1\ldots P} \right\}$

avec $E_0 = [1, 2, \ldots P] = \mathcal{C}$ $\longleftrightarrow$ ensemble total

2ème pas : On élimine X_l tel que :

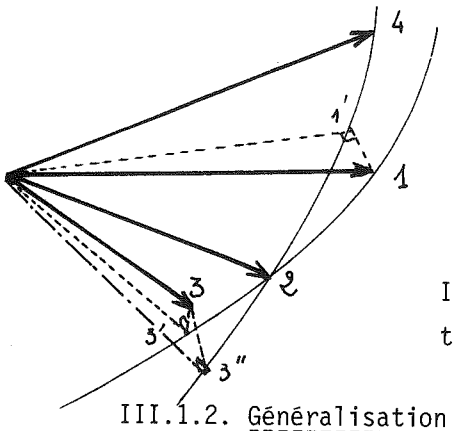
$l \longrightarrow \underset{i \in E_1}{Max} \left\{ R_{i.\{E_1-i\}} = R_{i.1,2\ldots i-1,i+1\ldots k-1,k+1\ldots P} \right\}$

avec $E_1 = [1,2\ldots k-1, k+1 \ldots P]$

Le principal avantage de la méthode est sa rapidité car on inverse une seule fois la matrice des corrélations et ensuite, à chaque pas, on enlève une ligne et une colonne et on modifie très facilement les autres.

D'autre part, si on arrête au pas K, on sait où l'on en est pour la variable $K+1$, reconstituée à $R^2_{K+1.1\ldots K}$, mais on ne peut pas affirmer qu' $\nexists$ $R_{l.1\ldots K}$, $X_l \in \mathcal{C} - E_k$ et éliminée précédemment, qui soit inférieur à $R_{K+1.1\ldots K}$.

Certes si on a 2 variables très correlées en corrélation totale, et indépendantes des autres, le fait d'en éliminer une fait que l'on conservera l'autre. Par contre si on met plusieurs variables en jeu ... (cf. fig.)



1er pas : 3 éliminée car bien expliquée par 1 et 2 (cf. 3').

2ème pas: 1 éliminée car bien expliquée par 2 et 4

Mais alors 3 est moins bien expliquée que 1 (cf. 3")

Il importe donc de contrôler simultanément la reconstitution de toutes les variables déjà éliminées .

III.1.2. Généralisation : Méthodes descendantes

On peut reprendre (A2) et prendre en compte ces variables déjà éliminées. Supposons qu'il reste k variables, soit $E_k = \{X_1 \ldots X_k\}$, et $\mathcal{C} - E_k$ l'ensemble éliminé.

On peut alors chercher un critère qui sélectionnera l tel que la variable la plus mal reconstituée, parmi l'ensemble $\mathcal{C} - E_k + X_l$ le soit encore convenablement.

On choisit alors X_l tel que :

$$\underset{X_j \in \mathcal{C}-E_k+X_l}{Min} \left[R^2_{j.\{E_k-X_l\}} \right] \quad \text{maximum}$$

Mais on pourrait aussi choisir un critère moyen qui, au lieu de se concentrer sur la plus mal reconstituée, chercherait :

$$\underset{X_j \in \mathcal{C} - E_k + X_\ell}{\text{Moy}} \left[R^2_{j.\{E_k - X_\ell\}} \right] \qquad \text{maximum}$$

On remarquera que si on raisonne non pas sur les $R^2_{j.}$ mais sur la variance résiduelle $1 - R^2_{j.}$ ces méthodes deviennent min-max ou min-moy. On a appelé (A5) et (A6) ces méthodes selon qu'elles utilisaient le critère du minimum ou de la moyenne.

Le premier pas est analogue à celui de (A2).

Sur le plan de la programmation, au lieu de considérer seulement l'ensemble $E_k = \{1 \ldots k\}$ dont on connait la matrice inverse $R^{-1}_{E_k}$ et de considérer, pour éliminer éventuellement X_ℓ, la seule corrélation multiple :

$$R^{-1}_{\{E_k - X_\ell\}} \cdot \beta_\ell = R_{X_\ell, \{E_k - X_\ell\}}$$

il faut considérer les $P - k + 1$ corrélations multiples :

$$R^{-1}_{\{E_k - X_\ell\}} \cdot \left(\beta_\ell \,\middle|\, \underbrace{\beta_P \,|\, \ldots\ldots \,|\, \beta_1}_{} \right) = \left(R_{X_\ell, \{E_k - X_\ell\}} \,\middle|\, R_{X_P, \{E_k - X_\ell\}} \,\middle|\, \ldots\ldots \,\middle|\, R_{X_{k+1}, \{E_k - X_\ell\}} \right)$$

β_ℓ : variable candidate à l'élimination dans E_k

$\beta_P \ldots \beta_1$: variables déja éliminées au pas k précédent .

et en prendre le minimum ou la moyenne, et scruter ainsi les k variables de E_k d'où $k \times (P - k + 1)$ corrélations multiples au pas k au lieu de k dans la méthode (A2).

Si on suppose $P = 2q$ et que l'on élimine la moitié des variables, cela équivaut pour :

(A2) à $q \cdot \dfrac{(3q+1)}{2}$

(A5) ou (A6) à $\dfrac{q.(q+1).(2q+1)}{6}$

} corrélations multiples

III.1.3. Généralisation : Méthodes ascendantes

On peut procéder de manière analogue en sélectionnant par exemple comme première variable celle qui reconstitue au mieux l'ensemble des autres, donc X_1 telle que :

$$\text{Min}_{i \in \mathcal{C} - X_1} r^2_{i,1} \longrightarrow \text{(A7)}$$

soit maximum

$$\text{Moy}_{i \in \mathcal{C} - X_1} r^2_{i,1} \longrightarrow \text{(A8)}$$

Cela s'écrit encore :

$$\left.\begin{array}{l} \text{Max}_i \left(1 - r^2_{i,1}\right) = \text{Max}_i \ \sigma^2_{i.1} \\ \text{ou} \quad \text{Moy}_i \ \sigma^2_{i.1} \end{array}\right\} \text{Minimum}$$

Au deuxième pas, on cherche X_2 telle que les variables $\{3, \ldots P\}$ soient bien expliquées donc que leur variance résiduelle à X_1 et X_2 soit minimale. Cela s'écrit donc :

$$\underset{\substack{\text{Moy} \\ i \neq 1,2}}{\text{Min}} \left[R^2_{i.1,2} \right] \quad \text{est Maximal}$$

ou $\underset{\substack{Moy \\ i \neq 1,2}}{Max} \left[1 - R^2_{i.1,2} \right] = \underset{Moy}{Max} \left[\sigma^2_{i.1,2} \right]$ est Minimal

Or si on ne remet pas 1 en cause, on sait que l'on peut écrire :

$$\sigma^2_{i.1,2} = \sigma^2_{i.1} \left(1 - r^2_{i,2.1} \right)$$

où $r_{i,2.1}$ est le coefficient de corrélation partielle de X_i avec X_2 compte tenu de 1.

L'algorithme est donc très simple car, pour choisir X_1 il a suffi de scruter la matrice $R_{\mathcal{C}}$ ligne par ligne, et prendre le maximum, ou la somme des carrés des éléments de chaque ligne, puis de les comparer.

Ensuite on calcule les variances résiduelles $\sigma^2_{i.1} \quad \forall i$ et la matrice des corrélations partielles compte tenu de X_1, de dimensions $(p-1)\times(p-1)$ soit : $R_{\mathcal{C}-1.1}$.

Au pas $k+1$, on a déjà selectionné : $E_k = \{1, 2, \ldots\ldots k\}$ et on connait

$$R_{\mathcal{C}-E_k.E_k} \quad \text{et} \quad \sigma^2_{j.E_k} \quad \forall j \notin E_k$$

On scrute alors chaque ligne de la matrice et on sélectionne X_{k+1} parce que :

$$\underset{\substack{Moy \\ j \notin E_k \text{ et } j \neq k+1}}{Max} \left[\sigma^2_{j.E_k} \left(1 - r^2_{j,k+1.E_k} \right) = \sigma^2_{j.E_k+X_{k+1}} \right] \text{ est Minimal}$$

Cet algorithme est très économique, puisqu'il ne nécessite pas d'inversion a priori, ni de calcul de corrélation multiple. D'autre part, le calcul de la matrice des corrélations partielles ne se fait qu'une fois à chaque pas, et ensuite on en balaye chaque ligne 1 fois et 1 seule.

III.1.4. Méthodes utilisant les corrélations partielles

a) Nous avons envisagé une autre méthode que nous appellerons (A3) et qui est en fait duale de (A2). Elle consiste à retenir les variables les moins expliquées par les variables restantes.

On procède comme suit :

1er pas : On retient X_k tel que :

$$R^2_{k.1,2,\ \ldots,k-1,k+1,\ \ldots,p} = R^2_{k.\mathcal{C}-k} = \underset{i \in E_o}{Min} \; R^2_{i.E_o-i}$$

avec $E_o = \mathcal{C} = \{1, 2, \ldots\ldots p\}$

2ème pas : Sachant que l'on conserve X_k, on ne considère plus que les informations indépendantes de X_k. Géométriquement, on se met dans un hyperplan de $\mathbb{R}^N$ orthogonal à X_k et on considère les résidus :

$$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1}, \varepsilon_{k+1}, \ldots \varepsilon_p$$

On devrait alors sélectionner X_ℓ tel que :

$$R_{\ell.1,2,\ldots k-1,k+1\ldots \ell-1,\ell+1\ldots p} = R^2_{\ell.E_1-\ell} = \min_{i\in E_1} R^2_{i.E_1-i}$$

$$E_1 = E_0 - k = \{1,2,\ldots k-1,k+1,\ldots p\}$$

puis calculer la matrice des corrélations partielles / X_k et X_ℓ, c'est-à-dire les intercorrélations des résidus de toutes les variables non encore retenues, X_k et X_ℓ étant supposées connues, etc...

Or, on s'aperçoit que dès le second pas :

$$R_{\ell.1,2,\ldots k-1,k+1\ldots \ell-1,\ell+1\ldots p/k}$$

n'est autre que :

$$R_{\ell.1,2,\ldots k-1,k,k+1,\ldots \ell-1,\ell+1\ldots p}$$

C'est-à-dire le coefficient de corrélation multiple de ℓ avec toutes les autres.

L'algorithme (A3) revient donc seulement à calculer la corrélation multiple de chaque variable sur les $P-1$ autres, à classer ces coefficients, et à sélectionner les K plus faibles puisque ce sont les variables qui, même connaissant toutes les autres, sont le plus mal reconstituées.

De façon analogue à l'algorithme (A2) celui-ci n'est pas optimal car il prend encore en compte toutes les variables, y compris celles que l'on va négliger. Mais il en est moins tributaire dans la mesure où l'on cherche justement X_i la plus indépendante possible des variables restantes (dont on va négliger certaines) mais compte tenu des variables déjà sélectionnées : $R^2_{i.1,2\ldots j\ldots p/k,\ell,\ldots m}$

Cependant, son utilisation en élimination de variables pose un problème que nous avons concrètement rencontré.

Supposons que nous ayons 2 variables X_k et X_ℓ quasi-indépendantes des autres mais très correlées entre-elles. Il semble logique d'en sélectionner rapidement une des deux et d'éliminer l'autre. En fait, à chaque pas, elles s'expliquent mutuellement presque complètement, donc aucune n'est sélectionnée sauf tout à la fin de l'analyse. Donc même si elle prend en compte 2 notions très attrayantes :

- l'orthogonalité maximum au paquet restant
- conditionnellement au paquet sélectionné,

le fait de considérer les variables individuellement conduit à un optimum qui n'est pas forcément celui recherché pour l'élimination de variables.

b) Une façon de s'affranchir de cela consiste à chercher non pas la plus indépendante du paquet restant, compte tenu de celles déjà retenues, mais à chercher directement la plus orthogonale à celles déjà retenues.

Nous appellerons cette méthode (A4), qui consistera donc, ayant sélectionné un ensemble E_m de variables $\{X_k, X_\ell, \ldots X_m\}$ à chercher dans l'ensemble restant $\mathcal{C} - E_m$ la variable X_i telle que :

$$R_{i.\,k,l,\ldots,m} = \underset{j \in \mathcal{C}-E_m}{Min} R_{j.\,k,l\ldots m}$$

ou encore celle dont la variance résiduelle, compte tenu des variables déjà sélectionnées est la plus forte.

On crée donc là encore un ordre et on construit, à partir des variables naturelles initiales la base la plus orthogonale possible. On notera au passage que l'utilisation du coefficient de corrélation multiple ou de la variance résiduelle est indifférente si les variables sont normées initialement à 1.0, mais si chacune est normée à $\sigma_k \neq \sigma_j \neq \ldots\ldots$ alors c'est sur les variances résiduelles qu'il faut sélectionner.

On remarquera aussi que le premier pas est analogue à celui de (A3) puisque l'on prend d'abord la moins correlée multiplement à l'ensemble et que l'algorithme ne démarre vraiment qu'au pas 2, en utilisant d'ailleurs à ce pas les corrélations totales, puis ensuite les corrélations partielles.

Enfin, cette méthode ne sélectionne pas les variables X_ℓ les plus indépendantes de l'ensemble $\mathcal{C}-X_\ell$, mais du sous-ensemble déjà retenu. Cela revient à construire pas à pas un sous-espace orthogonal et ce n'est probablement pas optimal au sens où, au pas k , ce n'est pas le sous-ensemble le plus orthogonal possible de k variables parmi P .

Elle serait aussi à comparer aux résultats d'une méthode Varimax qui effectue des rotations à partir des axes principaux d'une A.C.P. Cependant, même après rotation un axe peut rester associé à plusieurs variables (cas des effets de taille) entre lesquelles il faut choisir.

III.2 - Interprétations et analogie avec d'autres méthodes

III.2.1. Interprétation des méthodes présentées en III-1

(a) J.K. MILLER (1969) cité par COOLEY et LOHMES (1972) propose 2 notions distinctes de variance pour un ensemble de variables intercorrelées :

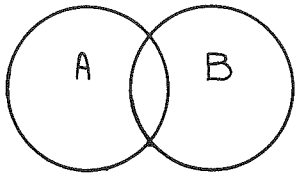
- la variance totale : telle qu'on la définit généralement en analyse factorielle. C'est la somme des variances individuelles, ou la trace de la matrice des covariances. (Si les P variables sont normées c'est simplement P),

- la variance complète : qui peut être définie comme la somme des composantes de variance indépendantes contenues dans l'ensemble considéré. Par exemple, dans le cas de 2 variables A et B on a 3 composantes indépendantes : la variance $S^2(A \cap B)$ commune à A et B, et les 2 résidus $S^2(A - A \cap B)$ et $S^2(B - A \cap B)$

Nous y ajouterons la notion de :

- variance spécifique : qui est, pour chaque variable, l'élément de variance indépendante qu'elle ne partage avec aucune autre. Son intérêt apparaît évidemment à partir de 3 variables et plus.

Cela s'explique assez bien à l'aide d'un diagramme ensembliste dit diagramme de Venn .

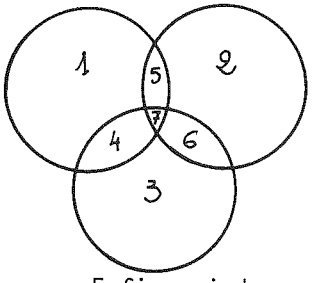


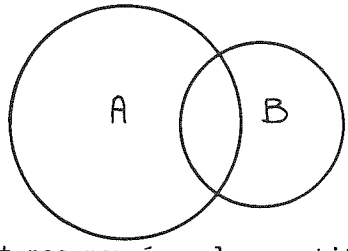
Si les variables sont normées, chacune correspond à un cercle de surface 1, et la variance totale vaut 2 tandis que la variance complète, surface de $A \cup B$, est comprise entre 1 et 2.

On remarquera que :

- il n'est pas possible sur un diagramme plan, de représenter les interrelations entre plus de 3 variables. Au-delà il faudrait considérer des sphères ou des hypersphères

- le nombre d'éléments de variance indépendants croît très vite avec le nombre de variables si elles sont intercorrelées.

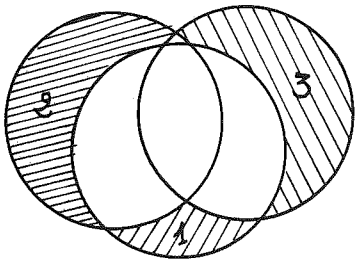




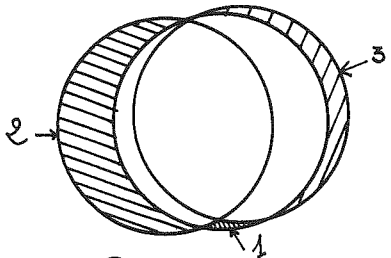
- Enfin, si deux variables ne sont pas normées, la quantité qu'elles ont en commun est le pourcentage de variance expliquée, mais la variance de A expliquée par B n'est pas la même que celle de B expliquée par A et dans ce cas on ne peut plus interpréter l'intersection $A \cap B$.

(b) Les différentes méthodes s'interprêtent alors aisément.

La méthode (A1) cherche à recouvrir le maximum de surface (resp. de volume ou d'hypervolume) avec un nombre fixé K de variables.

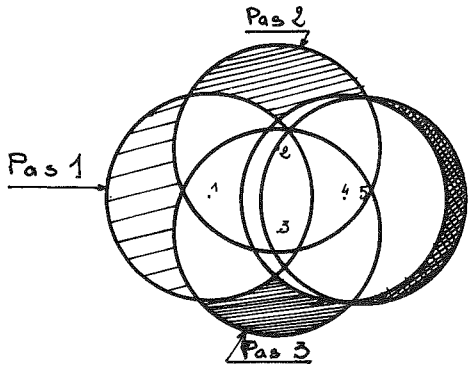


La méthode (A2) elle, regarde quel est le plus petit élément de variance spécifique dans l'ensemble $\mathcal{C}$, et on élimine l'ensemble de la variable correspondante. On fait de même à chaque pas, en réduisant donc le moins possible la variance complète du pas précédent.



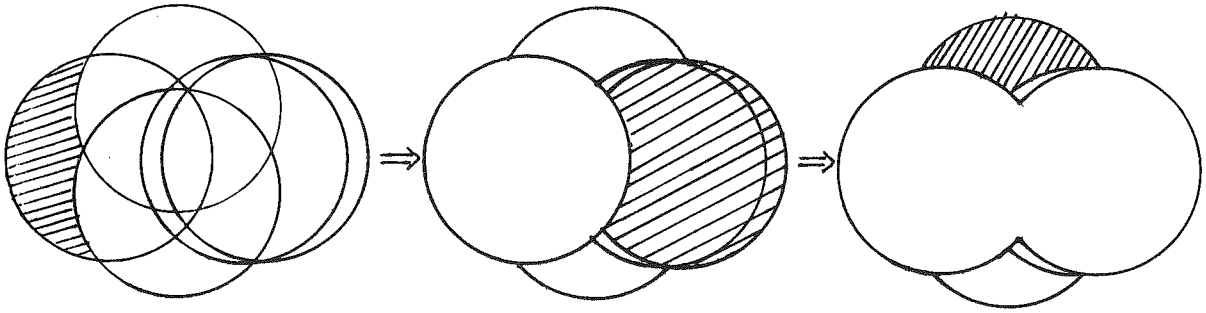
Mais on oublie la variance complète initiale et dans la configuration suivante, on voit qu'au pas 2 on peut éliminer soit la 2 soit la 3 mais si on choisit la 2 on ne reconstitue pas aussi bien la 1 que si on choisit la 3.

La méthode (A3) classe les variables par variances spécifiques décroissantes, donc ici 2, 3, 1. Mais ce n'est pas forcément en entrant ces variables dans cet ordre que l'on recouvre le maximum de variance complète à un pas donné.



On voit sur cette figure que si on sélectionne au pas 2 une variance spécifique associée à une variable assez liée à celle retenue au pas 1, on recouvre moins de variance complète que si on prenait 1 et 5.

Par contre la méthode (A4) satisfait ce critère :



recouvrir le maximum de variance complète pour un nombre fixé de variables. Néanmoins et surtout dans les premiers pas, on n'a aucune garantie quant à la partie déjà reconstituée des variables non encore sélectionnées.

C'est ce que vont prendre en compte les méthodes (A5) à (A8). Si l'on considère les méthodes Min-Max (A5) et (A7), on constate que la méthode descendante (A5) commence comme (A2). Par contre, dès le 2ème pas, elle peut en différer car le maximum de variance résiduelle aux variables restantes peut provenir d'une des variables déjà éliminées que (A2) ne considèrera pas. (Ceci est difficile à représenter sur le diagramme car il faut considérer 4 variables et les résidus de tous les couples, vis-à-vis du couple restant, ce qui est impossible à représenter dans le plan).

De son côté, il y a une analogie entre (A4) et (A7) dans leur souci d'intercepter le maximum de variance complète. Il y a cependant une différence dès le 1er pas car (A4) choisit la variable de plus grande variance spécifique, alors que (A7) cherche au contraire une variable plus "centrale" (qui reconstitue bien les autres). Ensuite, (A4) introduit la variable associée à la variance résiduelle maximale, alors que (A7) introduit celle qui minimisera la variance résiduelle maximale.

On notera aussi les différences entre :

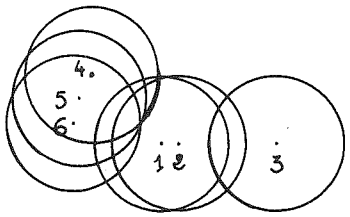
- Si on élimine 1 variable, c'est celle qui est la mieux reconstituée par l'ensemble des autres (A2 ou A5)

- Si on n'en conserve qu'une, c'est celle qui reconstitue le mieux chacune des autres (A7). Dans un cas, on élimine les plus expliquées, et dans l'autre on sélectionne les plus explicatives et non pas les moins expliquées (A4).

On a dans un cas un critère de variance expliquée, (qualité d'une projection orthogonale) et dans l'autre un critère de pouvoir explicatif (capacité de reproduire l'espace initial).

(C) Une dernière remarque concerne le choix de critères fondés sur la moyenne ou le minimum. Il est évident, par exemple dans les méthodes ascendantes que

le critère de la moyenne est sensible aux effets de taille : s'il existe un sous-ensemble assez fortement intercorrelé, c'est-à-dire une variable répétée plusieurs fois, elle prendra un poids plus grand dans ce critère moyen que pour le minimum.



Par exemple (A7) sélectionnera plutôt X_2 qui maximise l'explication de X_3 tandis que (A8) sera sensible au cumul des résidus de X_4, X_5 et X_6.

Le critère moyen travaille en fait sur la variance totale et dans (A8), on cherche effectivement la variable qui intercepte le plus de variance totale au premier pas. On a donc tendance à préserver les effets de taille. Par contre dans la méthode descendante, c'est l'inverse : un ensemble de variables très liées se verra rapidement décimé.

III.2.2. Liaison avec l'analyse canonique et la notion de "redondance"

(a) On supposera connus les principes de l'analyse canonique entre 2 ensembles de variables Z_1 et Z_2. On en trouvera un exposé détaillé dans COOLEY et LOHNES (1971) et dans LEBART et FENELON (1973), et un résumé succinct en annexe (dont on reprend les notations).

(b) Notion de redondance

Si l'on revient sur la notion de variance totale vue à la fin du paragraphe précédent (III.2.1) on constate que :

var. totale de Z_1 = P_1 = trace de R_{11}

Or le facteur X_k explique, ou extrait une proportion

$r^2_{k,z_{11}}$ de la variance de z_{11} $(=1)$

$r^2_{k,z_{1j}}$ de la variance de z_{1j} $(=1)$

etc...

Donc il extrait une variance :

$$r^2_{k,z_{11}} + r^2_{k,z_{12}} + \cdots r^2_{k,z_{1j}} + \cdots r^2_{k,z_{1p_1}} = s^2_{x_k}$$

dans l'ensemble Z_1, soit encore :

$$s^2_{x_k} = S^t_{k1} \times S_{k1}$$

où S_{k1} représente la "structure", ou le vecteur des corrélations du facteur X_k avec les variables de Z_1. Notons qu'il faut distinguer la variance du facteur (ici $\sigma^2_{x_k} = 1$, comme $\sigma^2_{y_k}$ d'ailleurs), de la variance extraite, ou expliquée par le facteur $s^2_{x_k}$.

De même, dans Z_2, le facteur correspondant Y_k extrait une variance :

Mais jusqu'à présent, on a défini d'une part la variance extraite par un facteur quelconque X_k de Z_1 et par le facteur Y_k associé dans Z_2, et d'autre part le kème coefficient de corrélation canonique $\Rightarrow R_{c_k}^2$.

On définit alors la redondance de l'ensemble 1 par rapport à l'ensemble 2 au niveau du facteur k :

Si $\delta^2_{X_k}$ = variance extraite par le facteur k dans Z_1

$\delta^2_{Y_k}$ = variance extraite par le facteur k dans Z_2

alors on peut appeler :

$$RD_{X_k} = \frac{\delta^2_{X_k}}{P_1} \times R_{c_k}^2$$

la redondance du facteur X_k par rapport à Y_k, c'est-à-dire la part de la variance de Z_1, extraite par X_k, qui est en plus expliquée par Y_k. Mais comme Y_k est un facteur orthogonal dans Z_2 et qu'aucun des autres facteurs orthogonaux à Y_k dans Z_2 n'est corrélé à X_k, on peut aussi bien dire que c'est, dans la part de variance que X_k extrait de Z_1, la proportion qui est expliquée non seulement par Y_k mais par l'ensemble Z_2 tout entier.

En effet :

Proportion de variance extraite par X_k
et expliquée par $Z_2 = \{z_{21}, z_{22} \dots z_{2P_2}\} =$

Proportion de variance extraite par X_k et expliquée par Y_1
+ Proportion de variance extraite par X_k et expliquée par Y_k
+ .....

Or $r(X_k, Y_l) = 0$ sauf $r(X_k, Y_k) = R_{c_k}$

Donc pour résumer :

- Variance extraite par X_k $\longrightarrow$ $\delta^2_{X_k} = S^t_{k1} \times S_{k1}$
- Proportion de variance de Z_1 extraite par X_k $\longrightarrow$ $\frac{\delta^2_{X_k}}{P_1}$
- Proportion de variance de Z_1 extraite par X_k et expliquée par Z_2 $\longrightarrow$ $\frac{\delta^2_{X_k}}{P_1} \times R_{c_k} = RD_{X_k}$

On remarquera que l'on peut appliquer la même démarche à Y_k et l'on trouvera:

$$RD_{Y_k} = \frac{\delta^2_{Y_k}}{P_2} \times R_{c_k} \neq RD_{X_k}$$

Donc ces 2 coefficients diffèrent, parce que les facteurs peuvent extraire des variances différentes, qui représentent elles-mêmes des proportions différentes de la variance totale de leur ensemble respectif Z_1 ou Z_2

Et on peut enfin définir la redondance totale de Z_1 par rapport à Z_2 (resp. de Z_2 par rapport à Z_1).

En effet, la base $\{X_1, X_2, \dots X_{P_1}\}$ (si $P_1 = Min(P_1, P_2)$)

étant orthogonale, il y a additivité des variances extraites par les facteurs respectifs :

$$var.(Z_1) = P_1 = var.(X_1) + var.(X_2) + \dots - var(X_{P_1}) = \sum_{k=1}^{P_1} \delta^2_{X_k}$$

et la proportion de variance de Z_1 expliquée par Z_2 est :

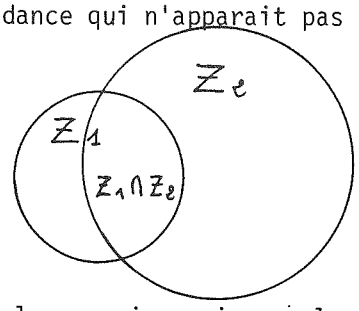
$$RD_{1/2} = \sum_{k=1}^{P_1} \frac{\delta^2_{X_k}}{P_1} \times R^2_{c_k} = \sum_{k=1}^{P_1} RD_{X_k}$$

De même :

$$RD_{2/1} = \sum_{k=1}^{P_1} RD_{Y_k}$$

(Naturellement, les sommations s'effectuent jusqu'à P_2 si $P_2 < P_1$).

Ces définitions, proposées pour la première fois par STEWART et LOVE (1968), et reprises par COOLEY et LOHNES (1971) ont l'avantage de mettre en évidence la dissymétrie de la redondance qui n'apparait pas dans les coefficients de corrélation canonique.



On peut, sur le croquis, voir que les 2 ensembles Z_1 et Z_2 ont une variance totale différente et que leur variance commune représente, par exemple, une part plus importante de Z_1 que de Z_2 .

Toutefois, le but des auteurs précédemment cités était surtout de définir un index, plus approprié que les coefficients R_{c_k}, de la liaison entre 2 ensembles de variables. A notre connaissance il n'a pas été utilisé directement en élimination de variables. On pourrait pourtant être tenté de chercher un sous-ensemble Z_1 ou E_1 de $\mathcal{C}$ tel que sa redondance avec le reste $\mathcal{C} - E_1$ soit maximale, et l'éliminer.

Nous allons voir que ce n'est autre alors que la méthode descendante (A6).

(C) En effet, si l'on interprète notre index de redondance en terme de corrélations multiples, on peut démontrer que la redondance définie pour l'ensemble Z_2 par exemple, soit $RD_{2/1}$ est la moyenne des carrés des coefficients de corrélations multiples obtenus en effectuant la régression de chaque variable de Z_2 sur <u>toutes</u> celles de Z_1.

En effet, si on calcule les corrélations de chaque variable de Z_2 avec le facteur X_k de Z_1 , on obtient :

$$\vec{r}_{Z_2, X_k} = \begin{bmatrix} r_{Z_{21}, X_k} \\ r_{Z_{22}, X_k} \\ \vdots \\ r_{Z_{2P_2}, X_k} \end{bmatrix} = \frac{1}{N} \, {}^{t}Z^{\circ}_{2v} \cdot \underbrace{\left(Z^{\circ}_{1v} \cdot U_k \right)}_{\text{ens. des obs. } X_k} = R_{21} \cdot U_k$$

Or la relation qui donne V_k en fonction de U_k est :

$$V_k = \frac{1}{Rc_k} \cdot R_{\ell\ell}^{-1} \cdot R_{\ell 1} \cdot U_k \quad \Longrightarrow \quad R_{\ell 1} \cdot U_k = Rc_k \times R_{\ell\ell} \cdot V_k$$

donc :

$$\overrightarrow{\pi}_{Z_\ell, X_k} = Rc_k \cdot \underbrace{R_{\ell\ell} \cdot V_k}_{\text{structure du facteur k dans } Z_\ell} = Rc_k \cdot S_{k\ell}$$

soit, $\forall\, z_{\ell j} \in Z_\ell \qquad r_{z_{\ell j}, x_k} = Rc_k \cdot r_{z_{\ell j}, y_k}$

Et la corrélation multiple de $z_{\ell j}$ avec toutes les variables, ou tous les facteurs orthogonaux, de Z_1 devient (puisque les X_k sont indépendants) :

$$R^2_{z_{\ell j} \cdot \{Z_1\}} = \sum_{k=1}^{P_1} r^2_{z_{\ell j}, x_k} = \sum_{k=1}^{P_1} Rc_k^2 \cdot r^2_{z_{\ell j}, y_k} = \sum_{k=1}^{P_\ell} Rc_k^2 \cdot r^2_{z_{\ell j}, y_k}$$

puisque $Rc_k \equiv 0 \qquad$ si $\quad k > Min\,(p_1, p_\ell)$

La moyenne de ces coefficients :

$$\frac{1}{P_\ell} \cdot \sum_{j=1}^{P_\ell} R^2_{z_{\ell j}, Z_1} = \frac{1}{P_\ell} \sum_{j=1}^{P_\ell} \sum_{k=1}^{P_\ell} Rc_k^2 \cdot r^2_{z_{\ell j}, y_k} = \frac{1}{P_\ell} \sum_{k=1}^{P_\ell} Rc_k \underbrace{\left(\sum_{j=1}^{P_\ell} r^2_{z_{\ell j}, y_k} \right)}_{= S^t_{\ell k} \cdot S_{\ell k} = \delta^2_{y_k}}$$

$$= \sum_{k=1}^{P_\ell} Rc_k \cdot \frac{\delta^2_{y_k}}{P_\ell} = \sum_{k=1}^{P_\ell} RD_{y_k} = RD_{\ell/1}$$

et par suite, la moyenne des carrés des corrélations multiples de chaque variable de Z_ℓ avec toutes celles de Z_1 n'est autre que la redondance définie en b) .

On a vu au passage que le fait d'expliquer une variable $z_{\ell j}$ par les variables originales de Z_1 ou par des facteurs canoniques revenait au même. Ceci reste vrai si on factorise Z_1 de n'importe quelle manière, par exemple par une A.C.P.

De même, on pourrait factoriser Z_ℓ, et chercher à expliquer ses facteurs, compte tenu alors des variances qu'ils extraient respectivement. On vérifierait alors que la redondance est la proportion de la somme des variances des facteurs de Z_ℓ expliquées par Z_1 (ou ses facteurs). On retrouvera cette notion en III.2.3.

d) Pour l'instant, on peut interpréter les méthodes (A6) et (A8) comme des méthodes consistant :

- pour (A6) : à extraire de $\mathcal{C}$ un ensemble E le plus redondant possible avec $\mathcal{C}-E$ et à l'éliminer (car en moyenne, l'explication de E par $\mathcal{C}-E$ est maximale)
- pour (A8) : à construire un ensemble S tel que la redondance de $\mathcal{C}-S$ avec S soit maximale.

Naturellement, il est possible d'utiliser des analyses canoniques pas à

pas pour bâtir ces ensembles (de même d'ailleurs que n'importe quelle factorisation) mais en pratique, il est inutile d'inverser à la fois R_{11} et R_{22}, et les algorithmes (A6) et (A8) sont les plus économiques.

III.2.3. Autres approches possibles et comparaison

(a) Les algorithmes présentés jusqu'ici répondent à la question : comment minimiser les redondances au sein d'un paquet de variables $\mathcal{C}$, ou comment extraire, avec un minimum de variables, le plus d'informations originales possible. Cela nous a conduit à sélectionner des sous-ensembles tels que toutes les variables qui n'y figurent plus peuvent être reconstruites au mieux, à l'aide de celui-ci.

L'intérêt de ces méthodes est de permettre d'éliminer des effets de taille indésirables dans des mélanges de variables hétérogènes, avant d'effectuer d'autres analyses. Dans le cas des réseaux de mesures, il peut permettre de réduire le nombre de stations en préservant à la fois les informations générales et les anomalies.

Mais certains posent le problème de manière très différente, par exemple pour les réseaux de mesure. Admettant que le réseau complet de P stations fournit des résultats très satisfaisants en A.C.P., on cherche un sous-réseau de q stations, dont l'A.C.P. fournirait des résultats les plus voisins possible de ceux obtenus sur le réseau complet.

(b) Le problème peut encore se poser de diverses manières. On appellera $\mathcal{Z}_1 = \mathcal{C}$ le paquet des P variables initiales, et $\mathcal{Z}_2 \subset \mathcal{Z}_1$ le sous-ensemble de q variables.

Une première solution consisterait à rechercher $\mathcal{Z}_2$ tel que l'A.C.P. de $\mathcal{Z}_2$ soit immédiatement superposable à celle de $\mathcal{Z}_1$, à une homothétie près. Cela supposerait que les facteurs se correspondent dans les 2 analyses : $F_k^{(1)}$ serait très corrélé à $F_k^{(2)}$ $\forall k$, et que les valeurs propres associées aux différents facteurs ou les poids de ces facteurs soient dans le même ordre : on préserve ainsi les effets de taille .

Cette approche pose quelques problèmes :

- d'abord on ne peut optimiser qu'une variable à la fois, et il faut donc définir un critère global de similarité. Il a fallu pour cela attendre l'apparition des opérateurs d'Escouffier (utilisés par SABATON et BENZECRI , 1977 , deuxième méthode)

- d'autre part, le souci de respecter les poids des facteurs conduit à reconstituer d'abord le 1er et,s'il est nettement dominant, à introduire plusieurs variables redondantes avant d'envisager de représenter le 2ème, ce qui semble souvent paradoxal à l'utilisateur.

- De plus, les résultats de l'A.C.P. de $\mathcal{Z}_1$ ne servent que de vérification et ne sont pas utilisés de façon optimale.

En effet, il arrive que ce ne soient pas les facteurs eux-mêmes que l'on souhaite retrouver, mais les configurations des points observations. On peut alors

admettre des rotations entre les $F_k^{(2)}$ et $F_k^{(1)}$, en plus des homothéties précédentes. Il faut donc être capable de reconstituer ces facteurs à une transformation linéaire près : ce n'est autre qu'une corrélation multiple.

Malheureusement celle-ci ne peut se faire directement car si on peut juger, dans l'A.C.P. de Z_1, de la reconstitution de X_j par k facteurs :

$$R^2_{X_j \cdot F_1^1, F_2^1, \ldots F_k^1} = \sum_{i=1}^{k} \lambda_i \cdot U_{ij}^2$$

il n'est pas possible de juger de la reconstitution d'un facteur par plusieurs variables intercorrelées :

Par exemple : $R^2_{F_j^1 \cdot X_1, X_2, \ldots X_k}$

$$R^2_{F_j^1 \cdot X_1, X_2} = \frac{r^2_{F_j^1, X_1} + r^2_{F_j^1, X_2} - 2\, r_{X_1, X_2} \cdot r_{F_j^1, X_1} \cdot r_{F_j^1, X_2}}{1 - r^2_{X_1, X_2}}$$

De plus, il faudrait pondérer la reconstitution du facteur par sa variance et donc chercher Z_2 tel que :

$$\left\{ \lambda_1 \times R^2_{F_1^1 \cdot \{Z_2\}} + \lambda_2 \times R^2_{F_2^1 \cdot \{Z_2\}} + \cdots\cdots + \lambda_p \times R^2_{F_p^1 \cdot \{Z_2\}} \right\} = \underset{Z_2 \subset Z_1}{\text{Max}}$$

© Compte tenu de la notion de redondance définie en III.2.2, on constate que la solution consiste à chercher Z_2 tel que la redondance :

$$RD_{Z1 \mid Z2} \quad \text{soit maximale}$$

avec la particularité qu'ici $Z_2 \subset Z_1$. Cette redondance est équivalente à :

$$\frac{1}{p} \sum_{k=1}^{p} \lambda_k \times R^2_{F_k^{(1)} \cdot \{Z_2\}} \quad \text{ou} \quad \frac{1}{p} \sum_{j=1}^{p} R^2_{Z_{1j} \cdot \{Z_2\}}$$

Mais en prenant cette dernière définition, on peut démontrer un résultat supplémentaire. En effet, la corrélation multiple d'une variable Z_j avec Z_2 a pour coefficients de régression le vecteur $b_j^{(1)}$: $\overrightarrow{b_j^{(1)}} = R_{22}^{-1} \cdot \overrightarrow{r_{j,Z_2}}$ $\qquad \overrightarrow{r_{j,Z_2}} = \{r_{Z_{1j}, Z_{2k}}\}$

et on obtient donc l'ensemble des coefficients de régression des variables de Z_1 sur celle de Z_2 par : $B^{(1)} = \left\{ b_1^{(1)} \cdots b_j^{(1)} \cdots b_p^{(1)} \right\} = R_{22}^{-1} \cdot R_{21}$

Or le coefficient de corrélation multiple d'une variable X_j de Z_1 avec toutes celles de Z_2 peut se déduire des coefficients de régression :

$$R^2_{Z_{1j} \cdot \{Z_2\}} = \sum_{k=1}^{q} b_{kj}^{(1)} \cdot r_{Z_{1j}, Z_{2k}}$$

On peut donc les obtenir globalement en

$$\begin{bmatrix} R^2_{Z_{11} \cdot \{Z_2\}} \\ \vdots \\ R^2_{Z_{1p} \cdot \{Z_2\}} \end{bmatrix} = \text{diagonale } B^{(1)t} \times R_{21} = \text{diagonale } R_{12} \times R_{22}^{-1} \times R_{21}$$

La redondance RD_{Z_1/Z_2} n'en est que la somme et devient ici :

$$\text{Trace}\left[R_{12} \times R_{22}^{-1} \times R_{21} \right]$$

et il faut trouver Z_2 qui maximise cette trace.

ⓓ Ceci se rattache à la technique de l'A.C.P. de Z_1 par rapport à Z_2 préconisée par RAO (1965) et reprise par ROBERT et ESCOUFIER (1976). Une présentation élémentaire consiste à dire que, en présence de 2 paquets X et Y, on n'analyse pas X mais la partie de X qui est reconstituée par Y soit $\hat{X}$. Si on appelle R_{11} la matrice de X, R_{22} celle de Y et R_{12} la matrice des corrélations entre X et Y on vérifie que :

$$\hat{X} = Y . R_{22}^{-1} . R_{21} \qquad Y = Y_{ov} = \{y_{ij}\}$$

et l'analyse de $\hat{X}^t . X = R_{12} . R_{22}^{-1} . Y^t . Y . R_{22}^{-1} . R_{21}$ se ramène à celle de $R_{12} . R_{22}^{-1} . R_{21}$ qui est une matrice de covariance.

Or on représente une part d'autant plus importante de la variance de X que : trace $\left[R_{12} . R_{22}^{-1} . R_{21} \right]$ est important, ou encore, on minimise d'autant plus la somme des résidus de X que l'on maximise la trace de $\hat{X}^t . \hat{X}$, et on retrouve bien notre critère quand X et Y correspondent à Z_1 et Z_2.

ROBERT et ESCOUFIER (1976) utilisent en fait une démarche plus générale qui définit la distance entre 2 configurations de points X et Y comme la norme de :

$$X . X^t - Y . Y^t \text{ (dimensions } N \times N) \Longrightarrow \text{ ici } X . X^t - \hat{X} . \hat{X}^t$$

Ils arrivent finalement à définir Z_2 comme le sous-ensemble qui maximise :

$$\left\{ \Sigma \lambda_i^2 \;/\; \text{trace}\left[R_{11}^2 \right] \right\}$$

or, R_{11}, matrice associée à Z_1 est une donnée, donc il reste à maximiser :

$$\Sigma \lambda_i^2$$

ou les λ_i sont les valeurs propres associées à : $R_{22}^{-1} . R_{21} . R_{12}$

Or on voit que maximiser $\Sigma \lambda_i^2$ est équivalent ici à maximiser $\Sigma \lambda_i$ donc $Tr\left[R_{22}^{-1} . R_{21} . R_{12} \right]$

Cette matrice diffère de celle rencontrée précédemment $R_{12} . R_{22}^{-1} . R_{21}$ mais comme $Tr[A.B] = Tr[B.A]$ on voit que la technique proposée par ESCOUFIER se ramène aussi à maximiser $Tr\left[R_{12} . R_{22}^{-1} . R_{21} \right]$: il y a identité entre les différentes méthodes.

Rappelons cependant que celle -ci maximise la redondance :

$$RD_{\mathcal{C} | Z_2} \quad \text{avec } Z_2 \subset \mathcal{C} = Z_1$$

alors que les méthodes utilisées en III-1 (A6 ou A8, par exemple) maximisent :

$$RD_{\mathcal{C}-Z_2 | Z_2}$$

CHAPITRE IV

APPLICATIONS

IV.1 - Applications à l'ensemble des variables explicatives des avalanches à Davos

Ces méthodes ont été utilisées par 2 fois pour améliorer les données présumées explicatives.

- Sur le fichier de 1974 (utilisé pour toutes les A.C.P. de la 2ème Partie), on s'est aperçu que certaines variables étaient très redondantes ou n'étaient jamais apparues dans les modèles décisionnels ultérieurs : on a donc voulu récupérer la place disponible pour introduire de nouvelles variables susceptibles d'apporter de nouvelles informations (d'où le fichier de 1976).

- Sur ce fichier de 50 variables, l'analyse en C.P. avait mis en évidence des effets de taille importants qui risquaient de biaiser les classifications ou plutôt les aggrégations, que nous voulions faire dans l'espace des variables principales. Les renormer à 1.0 était discutable dans la mesure où les derniers facteurs, difficilement interprétables et peut-être peu significatifs prenaient alors un poids énorme.

On a donc préféré utiliser des facteurs normés à leur valeur propre λ mais calculés sur un ensemble de variables plus réduit et dont les principales redondances auraient été enlevées.

IV.1.1. Choix du nombre de facteurs significatifs

Celui-ci a déjà été évoqué dans la 2ème partie, mais le choix final provient du recoupement entre les diverses méthodes suivantes :

a) nombre de facteurs nécessaires pour que toutes les variables soient correctement représentées : 10 à 20 facteurs selon le seuil choisi

b) position du facteur associé à une variable indépendante générée (dont la valeur propre associée est,pour l'échantillon qui nous intéresse, voisine de 0.8) : 15 à 18 (mais généralement 16). Notons cependant que la variable a été générée sans autocorrélation, or cette autocorrélation aurait eu pour effet d'élever son rang (autour de 18 ou 20)

c) perturbations des matrices de corrélations : les différentes simulations donnent des spectres très voisins entre eux (compte tenu de la taille élevée de l'échantillon et de l'absence d'autocorrélation) (Fig.III-1 c et III-1 d). Ceux-ci recoupent le spectre initial vers la valeur propre de rang 11 et la faible fluctuation d'échantillonnage des spectres dépourvus de corrélation montre que le spectre initial est très significatif au moins jusqu'à la valeur 10, c'est-à-dire que les valeurs propres sont significativement supérieures à 1.0. Par contre, cela ne nous indique rien quant aux éventuel-

les variables indépendantes, c'est-à-dire aux facteurs de valeurs propres voisines de 1.

d) La décroissance relative des valeurs propres reste à peu près constante (ou le diagramme Log λ_i en fonction de i reste linéaire) entre les facteurs 10 ou 11 et 18, et surtout entre 19 et 30 ou 35. Il est donc raisonnable de placer la coupure soit vers 18, soit vers 30 ou 35, mais alors dans ce dernier cas les valeurs propres associées tombent entre 0.7 et 0.3 et ne peuvent plus être considérées comme estimant 1.0.

En conclusion, on considèrera qu'il y a entre 10 et 18 facteurs significatifs, les 7 ou 8 derniers étant associés à des facteurs de valeurs propres voisines de 1. Ils ne correspondent cependant pas à des variables initiales bien individualisées et leurs pondérations se dispersent en général sur plusieurs d'entre elles, ce qui les rend difficiles à interpréter.

Il est cependant évident, même en admettant que chaque facteur retenu soit associé à 2 phénomènes physiques antagonistes qu'il y a beaucoup trop de variables et qu'il faut en garder entre 10 au minimum et 30 à 35 au maximum.

IV.1.2. Elimination des redondances : Méthodes fondées sur les A.C.P. ou les notions de distances

(a) Méthodes fondées sur les liaisons avec les facteurs principaux (B2) et (B4)

On donne dans le tableau VI les résultats pour l'analyse du bimestre Janvier-Février, d'une part sur l'échantillon complet, d'autre part sur le seul échantillon des journées avalancheuses. Pour chacun des facteurs (sauf le 50ème, numériquement indéterminé) on trouvera le n° et le coefficient de corrélation de la variable la mieux associée au facteur.

On peut constater :

1 - Une fluctuation importante entre les 2 échantillons, ce qui peut s'expliquer car l'échantillon avalanche n'est pas tiré au hasard.

2 - Une incertitude quant à la variable à associer au facteur: Il y en a parfois 2 ou 3 qui ont des coefficients de corrélation très proches.

3 - La faible valeur de ces coefficients dès que l'on s'éloigne des premiers facteurs.

4 - Des ambiguités entre la méthode (B2) et (B4) puisque la variable 2 , qui sort sur les facteurs 1 et 45, devrait donc être la première sélectionnée dans (B4) et l'une des premières éliminées dans (B2) !

De toutes façon la méthode (B2) repose sur des liaisons assez peu significatives entre variables et derniers facteurs, et si le rang de la matrice est inférieur à P , les derniers facteurs n'ont même aucune signification. Or si on veut éliminer des redondances, donc s'il y en a beaucoup, il est à craindre que cela se produise.

La méthode (B4) est plus satisfaisante quand on ne considère pas trop de

TABLEAU VI

Bimestre Janvier-Février

Facteur n°	Echantillon complet		Echantillon des journées avalancheuses	
	Variable j	$\rho(V_j, F_k)$	Variable j	$\rho(V_j, F_k)$
$k = 1$	2 (5)	.80	5	.75
2	43 (16)	.75	16	.82
3	31 (27)	.52	39	.60
4	49	.56	49	.78
5	47	.67	28	.46
6	37	.54	45	.55
7	50 (31)	.47	4	.44
8	38	.36	46	.39
9	24	.60	35	.42
10	26	.47	22	.51
11	4	.35	11 (15)	.43
12	35	.37	7	.44
13	7	.40	35	.41
14	35	.33	42	.43
15	7	.37	7	.59
16	7	.56	4	.32
17	11	.33	8	.48
18	33	.34	11	.43
19	3	.56	39	.33
20	26 (35)	.29	36	.36
21	48	.41	3	.35
22	11	.22	10	.23
23	31	.29	45	.24
24	4	.24	22	.22
25	6	.22	6	.28
26	45	.31	48	.19
27	34	.23	9	.21
28	25	.22	13	.17
29	9	.24	33	.16
30	41	.17	39	.15
31	3	.19	32	.17
32	38	.23	6	.17
33	6	.16	41	.18
34	17	.19	17	.14
35	42	.19	25	.16
36	10	.22	25	.12
37	48	.15	42	.16
38	5	.12	23	.12
39	29	.16	5	.13
40	5	.14	18	.15
41	28	.18	5	.08
42	30	.18	30	.08
43	46 (43)	.13	30	.08
44	27	.14	2	.09
45	2	.11	43	.06
46	44	.09	16	.07
47	16	.09	44	.07
48	19	.07	19	.08
49	19	.07	49	.06
50	--	--	--	--

facteurs, mais le problème des ambiguités entre variables se pose. Sur l'échantillon complet, les liaisons dominantes avec le facteur 1 sont :

N° variable :	1	2	5	6	23	36	41
Corrélation :	.776	.799	.795	.639	.666	-.691	.688

ce qui définit un axe de précipitation (1, 2, 5, 23) et de vent (6, 41) opposé au beau temps stable (36). Il peut être excessif de le caractériser par la seule variable 2 (neige fraîche du jour) alors qu'il y a manifestement une notion d'inertie (variables 5 23, 36 et 41). D'autre part ces variables ont entre elles des liaisons parfois faibles

$r(2,5) = .755 \qquad r(5,41) = .631$

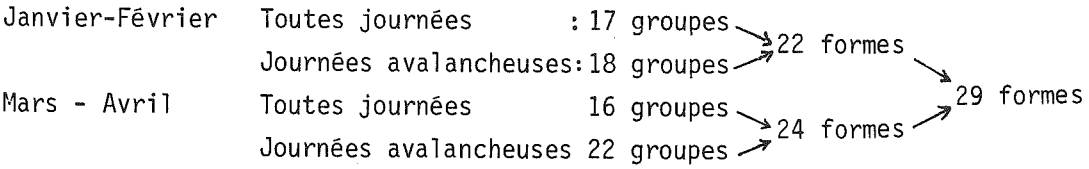
Et dans le cas d'axes représentant des influences antagonistes, il vaudrait mieux le caractériser par 2 variables (5,36), voire 3 (5, 36, 41). On pourrait même envisager des rotations orthogonales des facteurs qui maximiserait les liaisons fortes au dépens des plus faibles et fourniraient ainsi des groupements, associés aux facteurs 1, 2, ... Mais c'est en fait ce que font les méthodes suivantes.

(b) Méthodes fondées sur des notions de distances

La méthode (C1) utilisant le procédé Iphigénie, proche du "single-linkage clustering" présente le triple avantage de ne pas faire de moyenne sur les coefficients de corrélation, de n'utiliser que leur valeur absolue et de disposer d'un critère d'arrêt "objectif". Ses résultats sont donc à considérer avec attention :

- On constate une bonne stabilité entre l'échantillon complet et celui des seules journées avalancheuses.
- Les groupes sont assez faciles à interpréter, même s'ils mélangent parfois diverses influences.
- Il est très curieux de voir que le nombre de groupes obtenus au moment de l'arrêt est très proche du nombre de facteurs présumés explicatifs que nous avons laborieusement cherché à déterminer dans le paragraphe IV.1.1.

On a ensuite élaboré un système de formes fortes (au sens de DIDAY, cf Vème Partie) qui détermine les paquets de variables qui sont toujours apparues ensemble dans les analyses :



Comme certaines formes semblent devoir être caractérisées par plus d'une variable, on arrive à un total de 35 variables environ.

La méthode (C2) de "average-linkage" conduit à des résultats très voisins.

Le seuil de 0.4 comme critère d'arrêt conduit à des classifications en 16 à 20 groupes très comparables à ceux fournis par (C1). Mais l'intersection des résultats des 2 méthodes conduirait à 33 groupes.

Mars-Avril (C1) Iphigénie

Aval. seules

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
16	46	49	20	28	43	24	13	35	26	22	45	33	25	40	6	39	11	42	32	3	7
12	44	50	21	27	47		17									9			4		
14			18	29			48												31		
38			19																		
15			37																		
1			36																		
2																					
5																					
30																					
34																					
23																					
10																					
41																					
8																					

Toutes journées confondues

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
14	50	13	46	4	28	26	3	35	24	20	6	33	31	7	40
12	49	17	44		27				25	18	11		32		39
16		48	45		29					21					9
38					23					19					
15					30					37					
47					2					36					
43					1					22					
					5										
					34										
					10										
					41										
					42										
					8										

Janvier-Février (C1) Iphigénie

Aval. seules

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
14	50	48	35	2	40	38	36	25	32	26	31	15	22	7	9	6
12	49	17		1				24	4		33				39	11
16	21	13		5												
43	19			34												
47	20			28												
46	18			27												
44	37			30												
45				29												
				23												
				41												
				10												
				3												
				42												

Toutes journées confondues

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
14	46	34	35	32	22	50	13	39	15	26	11	31	33	25	24	7
12	44	2		4	20	49	17	9	38		6					
16	45	1			18		48	40								
43		5			21											
47		30			37											
		29			19											
		23			36											
		41														
		10														
		42														
		3														
		27														
		28														
		8														

Janvier - Février (C2)

Avalanches seules

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	4	6	7	8	9	10	12	15	18	22	23	24	26	31	35	43	44	49
2	32	11			39	41	14	38	20	37	29	25		33		47	46	50
5					40	42	16		19		30						45	
34							13		21		27							
3							17				28							
							48				36							

Toutes journées confondues

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	7	8	9	10	12	13	15	18	19	23	24	26	31	33	35	44	49
2	32	11			40	41	14	17	38	20	37	29	25					46	50
5					39	42	16	48		21	22	30						45	
34							43			36		27							
3							47					28							

Mars - Avril (C2)

Avalanches seules

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	3	4	7	9	10	12	13	18	22	24	26	27	33	44	49
2		32		39	41	16	17	21		25		28	35	46	50
5		31		40	42	14	48	20				29		45	
30						15		19							
34						38		37							
23						43		36							
8						47									
6															
11															

Toutes journées confondues

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	3	4	6	7	8	9	10	12	13	18	19	23	21	26	33	44	49
2		31	11			39	41	14	17	20	37	30	25		35	46	50
5		32				40	42	16	48	21	22	27				45	
								15		36		28					
								38				29					
								43									
								47									

IV.1.3. Méthodes fondées sur des notions de variances

A l'époque où l'étude a été faite, l'ensemble des méthodes décrites au chapitre III n'était pas disponible (parmi les méthodes A5 à A8 seule A6 comportait un programme, mais très peu performant)

ⓐ Méthodes fondées sur les corrélations multiples

La méthode (A2) propose d'éliminer un certain nombre de variables. Pour chaque échantillon, nous donnons les 20 premières variables à éliminer. On constate :

- une bonne stabilité selon que l'on prend l'échantillon avalancheux ou l'ensemble des journées
- si l'on compare aux groupements établis par les méthodes (C1) et (C2), que cela revient à éliminer des variables dans les groupes à forts effectifs
- que même après élimination de 20 variables les corrélations multiples restent encore élevées au sein du paquet restant et en particulier dans l'échantillon avalancheux (non aléatoire). Si l'on ajoute à cela que les variables à éliminer ont, dans les premiers pas, des corrélations fortes et très voisines, on voit que ces effets d'échantillonnages (échantillons non aléatoires et de taille réduite) modifient les choix effectués sur l'échantillon complet tant en Janvier-Février qu'en Mars-Avril.
- Par contre, si on compare,au pas 20,les variables à éliminer pour les échantillons complets de Janvier-Février et de Mars-Avril, ils ne diffèrent que par 3 variables, dont 2 sont en fait très voisines (36 et 38, 49 et 50).

Tableau VII

Pas n°		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Corrél. mult. max. au pas 21
Janvier Février	Avalanches seules	12	49	19	27	16	44	2	43	30	18	5	17	29	14	41	25	1	37	10	21	.737
	Toutes journées	14	12	19	49	2	44	43	27	30	18	5	36	29	16	17	10	20	1	41	37	.801
Mars Avril	Avalanches seules	16	14	50	19	43	27	18	5	38	2	12	37	41	30	29	17	10	20	32	36	.829
	Toutes journées	14	50	12	43	19	27	2	18	29	16	5	38	30	41	37	17	20	10	44	31	.788

Méthode (A2) - Minimisation des corrélations multiples maximales. Variables à éliminer

La méthode (A3) fondée sur la corrélation multiple minimale est sensiblement plus stable d'un échantillon à l'autre (toutes journées, ou journées avalancheuses). En effet l'ordre des coefficients de corrélations multiples les plus faibles est moins affecté par l'échantillonnage que celui des plus forts (cas de (A2)). On constate aussi qu'elle sélectionne à peu près les variables dans les groupes formés par les méthodes utilisant des distances (Tableau VIII).

Tableau VIII

	Janvier-Février (A3)		Janvier-Février (A4)		Mars-Avril (A4)	
	Avalanches seules	Toutes journées	Avalanches seules	Toutes journées	Avalanches seules	Toutes journées
1	7	7	7	7	3	26
2	35	35	15	22	50	46
3	4	24	41	9	46	17
4	32	25	50	35	48	25
5	38	4	45	34	24	22
6	11	32	33	46	28	40
7	6	11	22	38	47	8
8	26	6	36	13	26	3
9	33	3	35	28	39	4
10	39	33	47	26	7	35
11	9	26	13	24	35	29
12	40	8	26	11	4	7
13	3	39	4	4	11	42
14	10	9	11	47	42	45
15	42	40	9	33	45	49
16	41	42	34	21	13	11
17	8	10	24	48	22	15
18	24	41	3	50	21	21
19	25	34	8	42	33	33
20	31	38	20	3	40	34
21	34	5	48	25	8	39
22	5	23	29	8	25	47
23	23	31	38	23	34	32
24	30	30	31	45	15	36
25	29	29	46	31	9	9
26	45	45	42	15	6	24
27	48	48	39	39	23	48
28	20	17	32	32	1	6
29	18	13	6	40	31	1
30	36	20	19	6	44	44
31	1	18	40	37	36	31
32	2	21	1	10	32	10
33	21	36	10	1	20	23
34	17	2	28	20	10	20
35	13	1	25	17	17	13

C'est encore plus vrai de la méthode (A4) utilisant la variance partielle résiduelle. Si elle n'est que globalement stable quand on considère les 30 premières variables sélectionnées, on constate que chacun des groupes issus de C4 ou C2 contribue pour une variable et une seule jusqu'à un rang très avancé ($\sim$ 20). Cela conforte d'ailleurs l'intérêt de ces 2 méthodes.

Une dernière remarque qu'autorise (A4) provient des variances résiduelles calculées à chaque pas :

Pour que toutes les variables, même non sélectionnées, soient représentées à :

60% il faut sélectionner $\longrightarrow$ 30 variables en (A4)
et seulement 15 facteurs en A.C.P.

80% il faut sélectionner $\longrightarrow$ 40 variables en (A4)
et seulement 20 facteurs en A.C.P.

(b) Méthodes fondées sur les corrélations canoniques

On donne les résultats de la méthode (A6) dans le tableau suivant. Compte tenu du coût de la méthode à l'époque, on s'était contenté d'une quinzaine de pas.

Comme pour la méthode (A2) on constate que les redondances les plus fortes sont souvent assez voisines, ce qui induit des fluctuations selon les échantillons.

On notera que les premières variables éliminées sont analogues, pour Mars-Avril, à celle éliminées par (A2) . C'est un peu moins vrai pour Janvier-Février.

Note - La première variable éliminée devrait être la même par (A2) et (A6) puisqu'au pas 1, (A6) se réduit à une corrélation multiple. Mais l'inversion de la matrice dans les premiers pas peut être difficile et nous avons parfois dû éliminer manuellement une des variables pour pouvoir démarrer l'algorithme d'où ces anomalies. (Ces variables forcées sont soulignées dans les tableaux).

TABLEAU IX

Pas n°		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Janvier	Avalanches seules	12	49	19	44	16	27	2	30	43	18	5	17	29	25	41	(36)
Février	Toutes journées	12	16	19	49	2	44	43	27	30	18	5	36	10	17	29	(41)
Mars	Avalanches seules	<u>14</u>	50	19	16	43	27	5	18	38	2	41	30	29	36	12	(17)
Avril	Toutes journées	<u>14</u>	50	12	43	19	27	2	18	29	5	38	41	17	30	36	(16)

Méthode (A6) - Variables à éliminer

IV.1.4. Conclusion

En fait,les variables ont été choisies, au vu de ces diverses méthodes, de façon à ce que les plus indépendantes apparaissent, les paquets indépendants étant

a) et c) Distribution des valeurs propres en logarithme .

b) et d) Pourcentages de variance respectifs extraits par les 1ers facteurs.

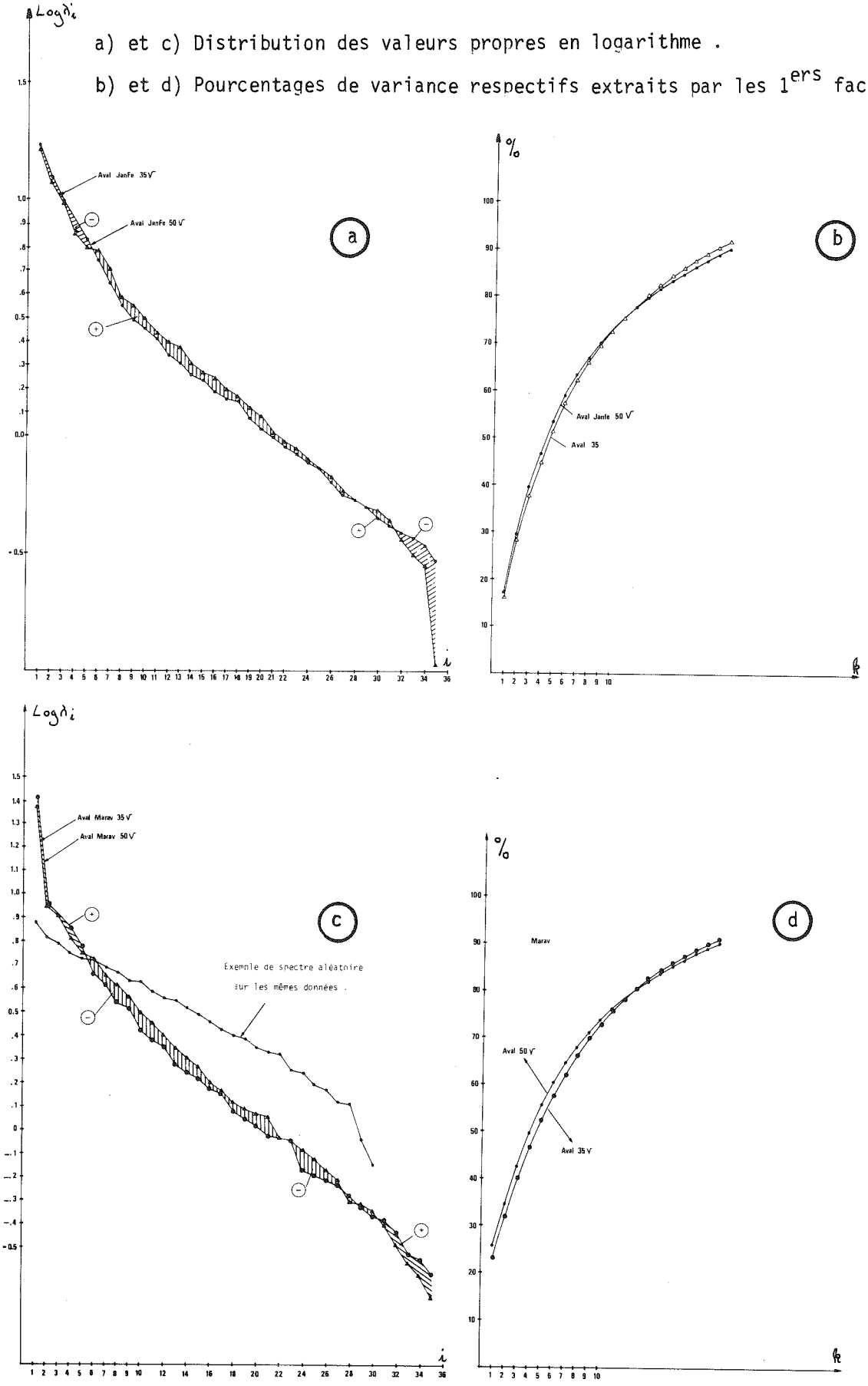


FIGURE III - 11 : Comparaison des A.C.P. des données de DAVOS avec l'ensemble complet de 50 variables et l'ensemble réduit de 35 variables:

représentés par 1 variable au moins et 2 s'ils contiennent des influences qui s'opposent. De plus certaines considérations de cohérence ou des connaissances physiques sur les phénomènes nous ont conduit à forcer certaines variables, d'où le sous-ensemble retenu:

1 3 4 6 8 9 10 11 12 13 20 21 22 23 24 25 26
28 31 32 33 34 35 37 38 39 40 41 42 44 45 46 47 48

+ 49 (Mars-Avril) ou 50 (Janvier-Février)

$\Longrightarrow$ soit 35 variables au total.

On donne, pour les échantillons des journées avalancheuses de Janvier-Février et Mars-Avril, les nouvelles distributions de valeurs propres obtenues par analyse de la matrice des 35 variables.

On constate (figure III-11 a et c) que les pourcentages d'inertie expliqués par les premiers axes sont plus faibles au début (abaissement des valeurs propres des premiers facteurs) et plus forts à la fin.

Note : On a comparé les pourcentages d'inertie sur les 35 variables avec ceux sur les 50 variables. En fait pour ces dernières, il aurait vallu porter le rapport des valeurs non pas à l'inertie totale (50) mais à l'inertie expliquée par les 35 premiers facteurs, soit :

Janvier-Février : 49.23 Mars-Avril : 49.29

Cela aurait encore légèrement écarté vers le haut les courbes associées à 50 variables dans les figures b et d).

On donne aussi les distributions sous la forme $\log_{10} 100 \frac{\lambda_i}{\sum \lambda_i}$, où $\sum \lambda_i$ vaut 35 ou 50 selon les cas (fig.11 b et d). On voit là encore que la pente (en logarithme) donc la décroissance du spectre est moindre avec 35 variables qu'avec 50.

Pourtant, on peut s'étonner de l'allure encore fortement décroissante de ce spectre, après élimination partielle des redondances. Il faut cependant se rappeler (cf IIIème Partie I.1.2) que la faible taille de l'échantillon conduit à des spectres de plus en plus exponentiels, même pour des variables théoriquement indépendantes. On a porté, à titre d'exemple, dans la figure III-11-c , la distribution obtenue avec 30 variables indépendantes et 100 observations. (Idéalement, il aurait fallu faire des simulations avec 35 variables et environ autant d'observations que de journées avalancheuses $\Longrightarrow \sim$ 150). On voit que l'on est loin de l'horizontale de $\log_{10} 100 \times \frac{1}{30} = .52$

On dispose donc désormais d'un ensemble de 35 variables, dont les redondances les plus marquantes ont été enlevées, et qui comporte environ 15 à 17 facteurs significatifs. Ceci sera utilisé dans la Vème Partie.

IV.2 - Applications à l'optimisation de réseaux pluviométriques

IV.2.1. Nombre de facteurs significatifs

Cette recherche s'est faite sur l'analyse du tableau des données brutes des épisodes. Le spectre est très caractéristique de ces données, avec un premier facteur

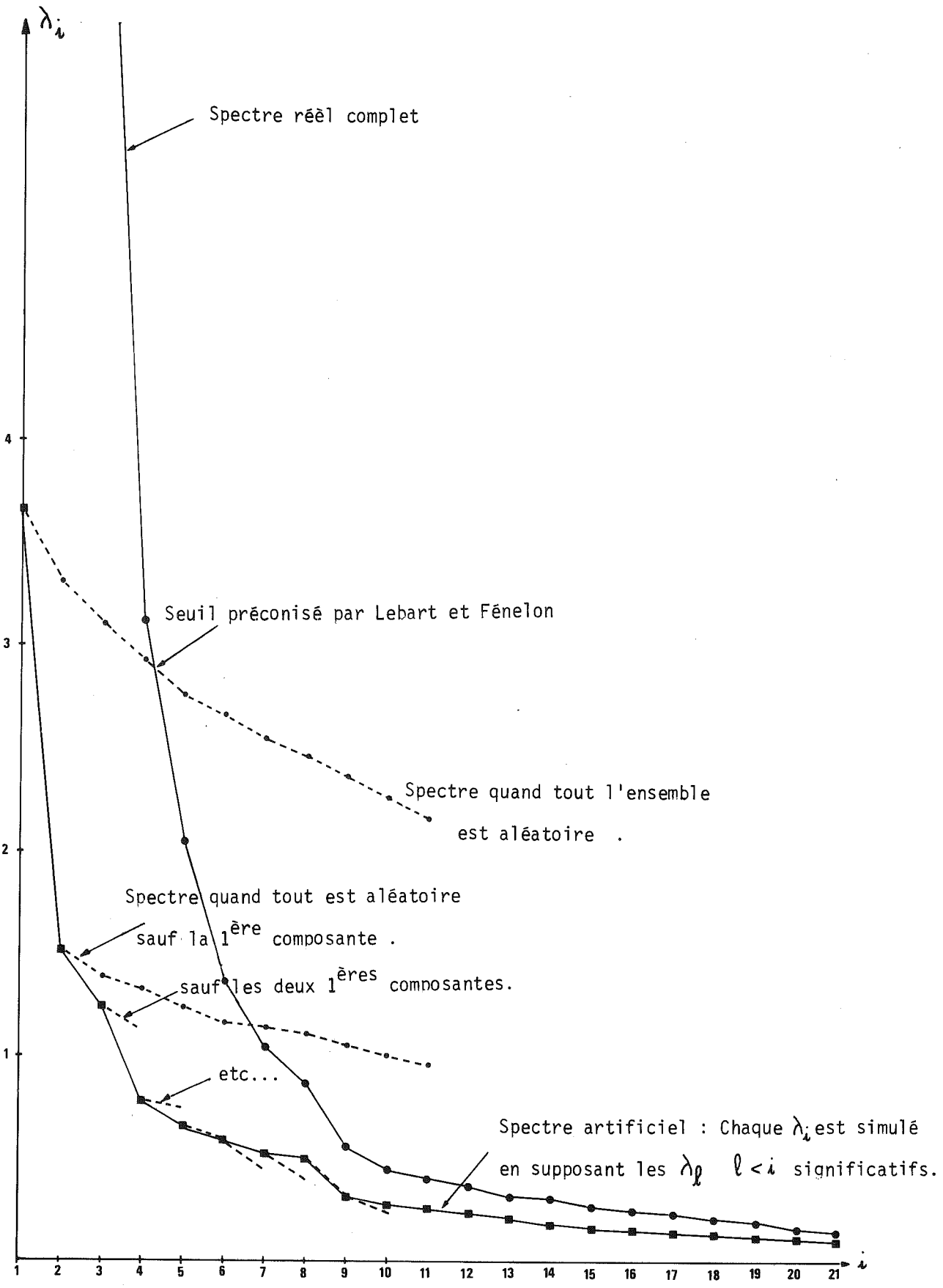


FIGURE III - 12 : Spectre de l'A.C.P. de 73 stations sur les données et recherche des facteurs significatifs.

qui représente 60% de la variance, essentiellement à cause de la variabilité interépisodes (cf IVème Partie).

L'application brutale de la règle de KAISER conduirait dans ce cas à 7 ou 8 facteurs. On constate aussi que le spectre $\lambda(i)$(Fig.III-12) se comporte comme la juxtaposition de 2 exponentielles (k = 1 à 8, puis k $\geqslant$ 9) ce qui dans le diagramme en Log, produit une cassure au niveau k = 9.

On donne aussi une courbe λ' qui n'est pas un spectre, obtenue par simulation en effectuant :

k = 1 Toutes les variables sont brassées aléatoirement $\Longrightarrow \lambda'_1$ 1ère valeur propre si la matrice de corrélation était aléatoire.

k = 2 Comme $\lambda_1 \gg \lambda'_1$, on accepte que le 1er facteur soit significatif, et on brasse les résidus au 1er facteur, puis on calcule la 2ème valeur propre λ'_2 (la 1ère étant λ_1) de cette matrice.

k = 3 Si $\lambda_2 \gg \lambda'_2$, on fait de même à partir du 3ème facteur...

On constate alors que :

- la règle de LEBART et FENELON est trop pessimiste car on ne garderait que 4 facteurs

- jusqu'au facteur k = 8, les valeurs λ'_{k+1} diffèrent selon que l'on considère que λ_k est significatif ou non. Par contre, les λ_k et λ'_k se confondent pratiquement à partir de k = 9, ce qui signifie que les λ évoluent de la même manière que les valeurs propres d'une matrice aléatoire (à l'exception de ses 8 premiers facteurs)

- si on porte de part et d'autre de la valeur λ_i son intervalle de confiance (dans l'hypothèse des matrices de variance-covariance et pour N grand, ce qui n'est pas le cas ici) soit $\ell_i = \lambda_i + \alpha_{80\%} \sqrt{\frac{2\lambda_i^2}{N}}$ on constate qu'ils se recouvrent partiellement à partir de k = 9 ou 10. Mais en fait, le λ_i obtenu est le maximum possible pour l'échantillon considéré et il faudrait prendre en compte leur dispersion plutôt en découpant l'échantillon.

- Néanmoins, la règle de KAISER donne une valeur très cohérente avec les techniques plus sophistiquées.

Par contre, il est dangereux de vouloir résumer un réseau comportant 8 à 10 facteurs significatifs par 10 stations seulement sur 73. L'analyse du graphique $\log 100 \frac{\lambda_i}{\Sigma\lambda_i}$ en fonction de k met en évidence outre la cassure au niveau 9, une autre vers le niveau 20 à 24 puis une perte de signification vers k $\approx$ 43.

IV.2.2. Optimisation du réseau par des méthodes fondées sur les distances

Remarque importante - Notre but était de tester ces méthodes entre elles et leur sensibilité selon les données utilisées, c'est-à-dire :

1- La matrice de corrélations des valeurs mensuelles d'Octobre et Novembre de 1961 à 1972 (fournie par M. D. DUBAND).

2 - La matrice de corrélation des valeurs radicalisées des 84 épisodes.

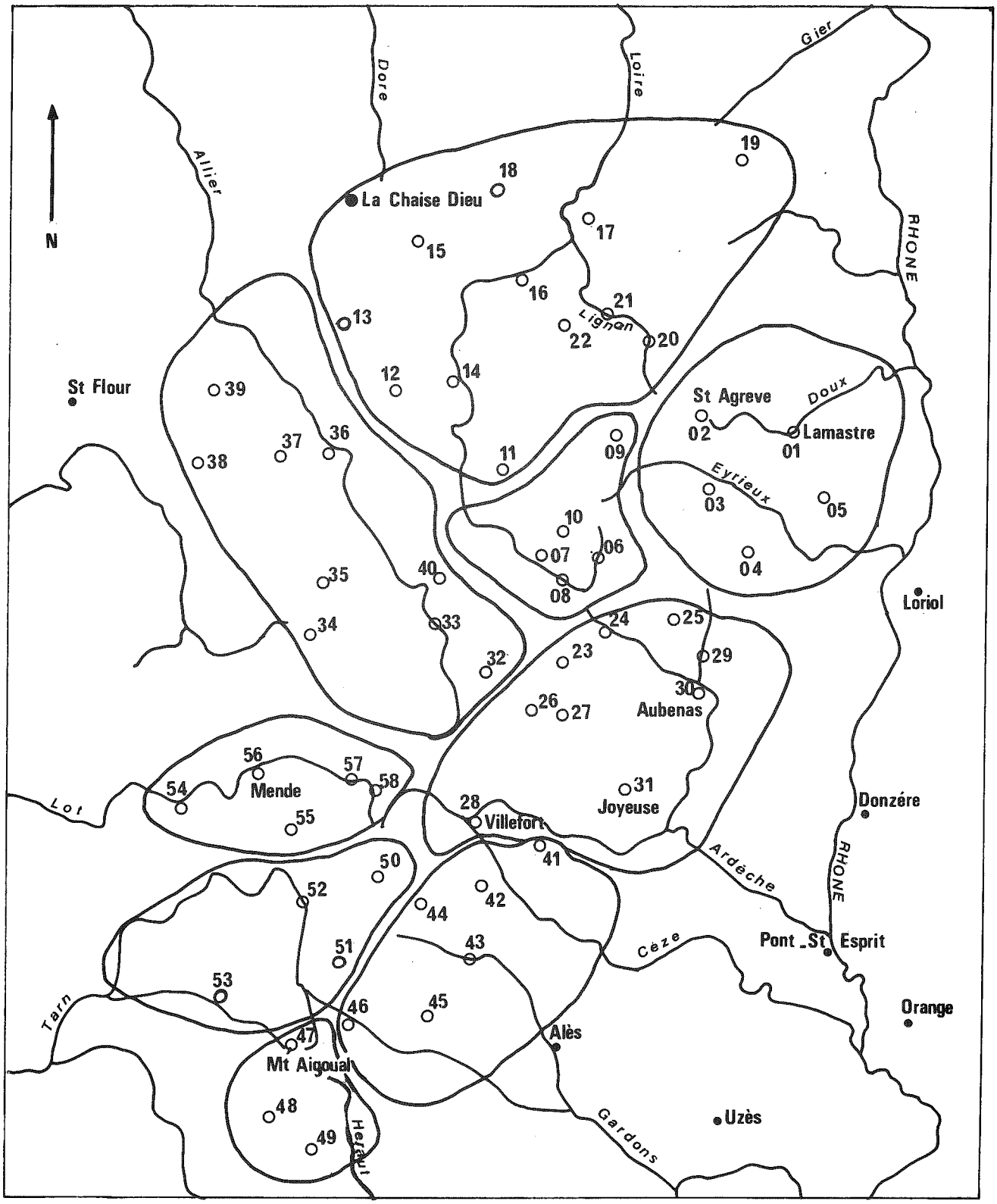


FIGURE III - 13 : Positions géographiques des 58 stations et regroupement en bassins hydrographiques.

L'intersection des 2 tableaux nous a donc limité à 58 stations seulement avec la numérotation suivante (Figure III-13).

- L'application de la méthode (C1) se heurte cette fois au problème de l'effet de chaîne qui nous conduit à un groupe dominant entouré de stations plus ou moins atomisées (Figure III-14-a).

Pour les épisodes par exemple, on trouve un groupe de 30 stations dont certaines, comme la 24 et la 56 ne sont correlées qu'à .27 ! Ceci est absolument inacceptable pour les réseaux.

- La méthode (C2) fournit des résultats acceptables en dépit d'un critère d'arrêt assez empirique. En utilisant pour distance la valeur absolue de la corrélation on constate que le critère reste assez longtemps élevé, puis se met à décroître rapidement. Cela se produit au niveau de 13 groupes pour les épisodes, et de 19 à 20 groupes pour les données mensuelles. De plus, on constate une certaine différence entre les 2 classifications, un peu plus erratique en valeurs mensuelles (cf. Figure III -15).

- La méthode (C3) présente un certain nombre d'avantages. Le problème de l'arrêt est encore résolu empiriquement en regardant évoluer le critère d'aggrégation (fig. III-14) mais celui-ci correspond effectivement au plus faible coefficient de corrélation existant au sein des groupes pour le niveau choisi. Cette évolution présente en général une ou plusieurs cassures nettes :

- au niveau 20 groupes pour les données mensuelles
- au niveau 33 ou 22 groupes pour les épisodes.

Si on accepte, par exemple, 22 groupes pour les épisodes, leurs compacités respectives font preuve d'une bonne homogénéité :

1	2	3	4	5	6	7	8	9	10	11
.851	.873	1.0	.848	.902	.843	.918	.887	.845	1.0	.859
12	13	14	15	16	17	18	19	20	21	22
.847	.873	.933	.929	.849	.879	.908	.908	1.0	1.0	.903

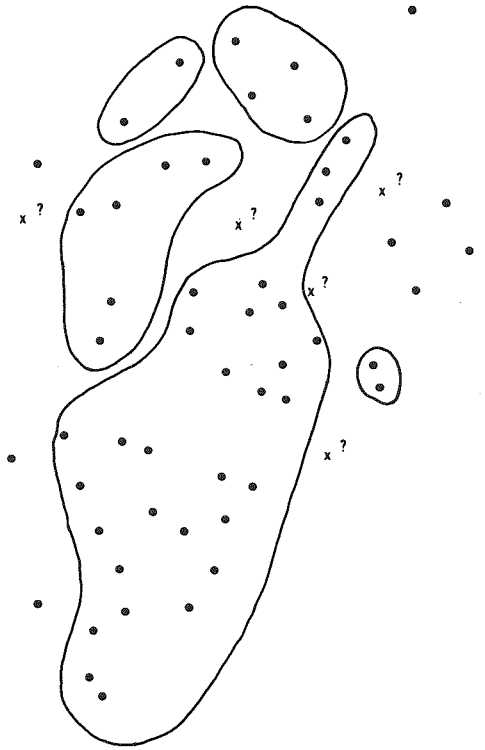
On notera aussi une chute en deçà de 13 groupes mais le critère est déjà un peu faible.

Ici aussi, les groupes diffèrent assez sensiblement selon que l'on prend les données par mois ou par épisode.

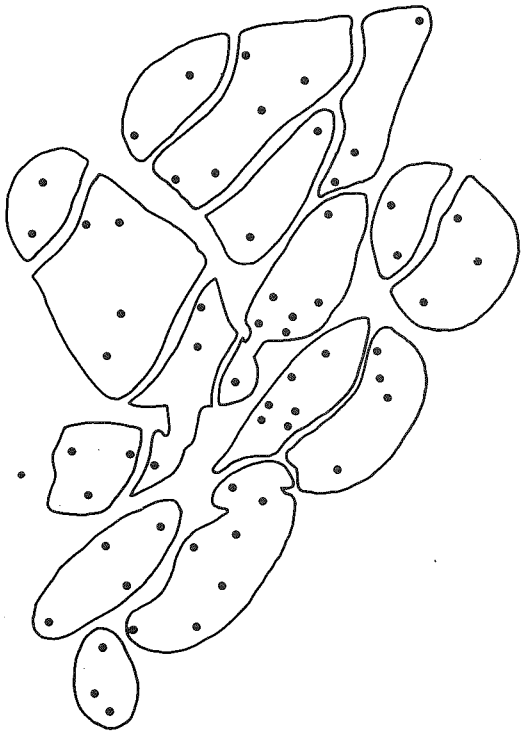
IV.2.3. Méthodes utilisant les variances

Au niveau des méthodes, on constate un certain nombre de propriétés un peu inattendues (cf figure III-16) :

(a) Si on regarde les méthodes qui prétendent minimiser le résidu maximal, on constate que la méthode descendante (A5) est sensiblement meilleure que la méthode ascendante (A7). En effet si on veut se limiter à 10% de variance résiduelle maximum

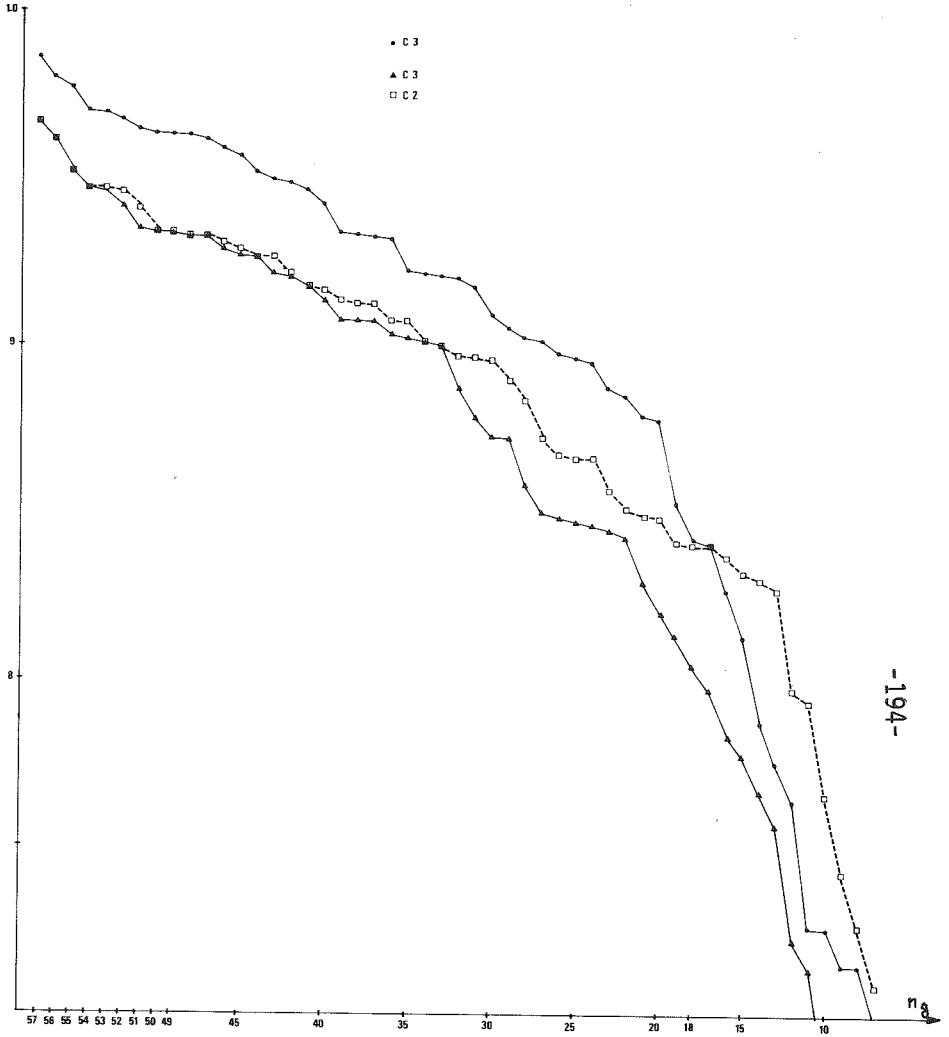


- groupements obtenus par C1



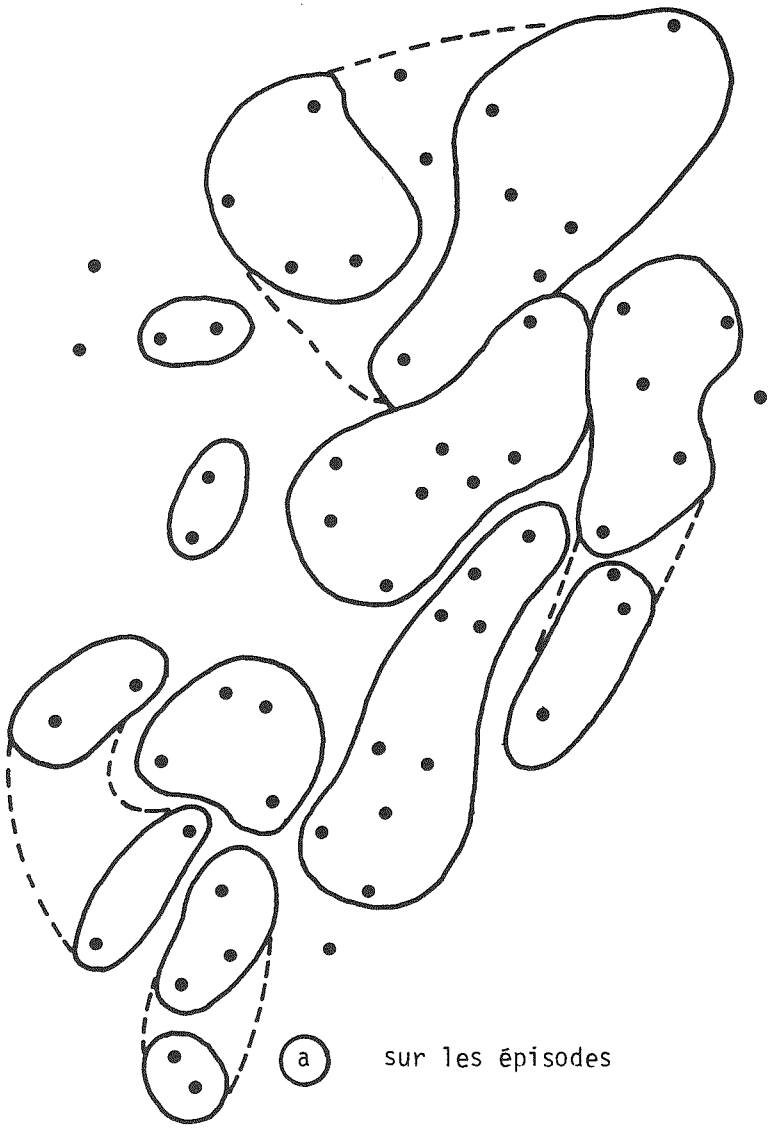
- groupements obtenus par C3

a) Regroupements obtenus par C_1 et C_3 sur les épisodes .

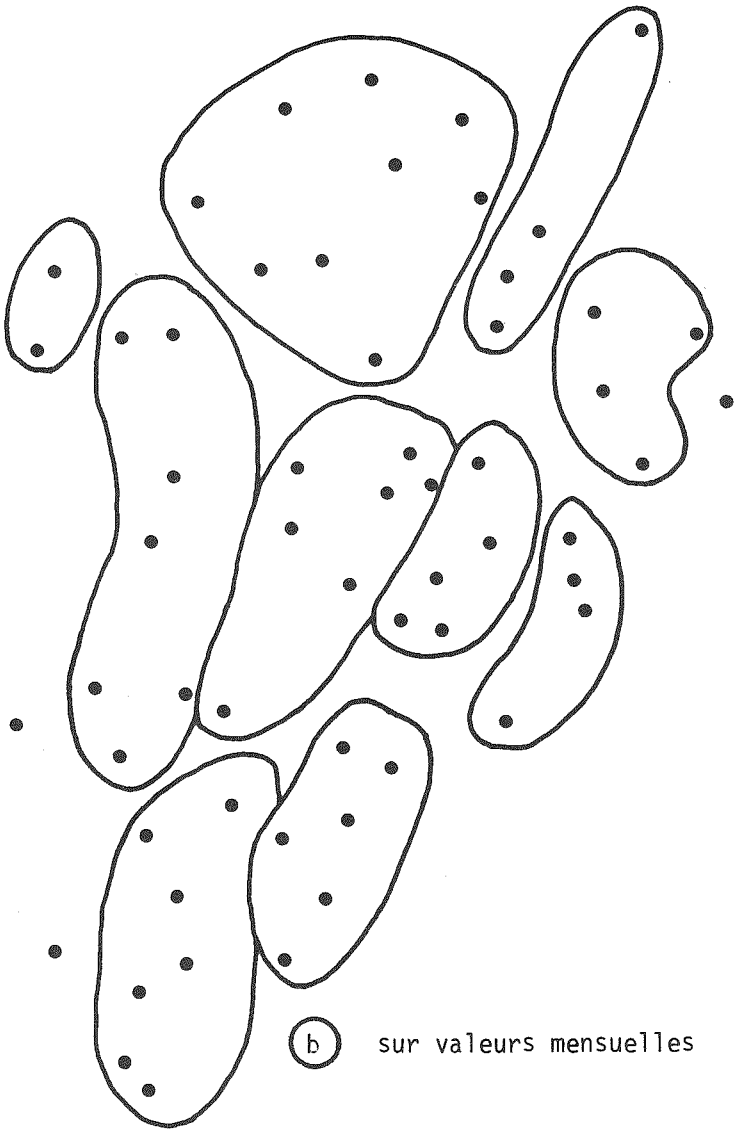


c) Evolution des critères respectifs des 3 méthodes en vue de detecter le nombre de groupes n_g.

FIGURE III - 14 : Groupements effectués par C1 ET C3



(a) sur les épisodes



(b) sur valeurs mensuelles

FIGURE III - 15 : Comparaison pour C2 des groupements obtenus sur les données mensuelles ou d'épisodes .

(non expliquée) il suffit de prendre 20 variables (partant de 58) dans (A5) contre 25 (partant de 0) dans (A7) pour les épisodes, et 15 dans (A5) contre 20 dans (A7) pour les données mensuelles. Cela semble privilégier les méthodes descendantes tant que l'on garde plus de 4 ou 5 variables.

Les temps de calcul sont, eux, comparables si on effectue le balayage complet de toutes les variables.

b) La méthode (A2), beaucoup plus rapide que (A5) ou (A6) dans la mesure où elle ne prend plus en compte les variables précédemment éliminées, fournit sensiblement les mêmes variables que la méthode minimax (A5).

(Sur les épisodes, il faut attendre le 22ème pas pour déceler une différence). Cela signifie que la reconstitution minimale est en général obtenue pour la dernière variable éliminée.

La similitude est moins bonne entre (A2) et (A6) car (A6) prend en compte une moyenne.

Néanmoins, la méthode (A2) appliquée à un réseau est assez satisfaisante et le risque de voir un résidu croître brusquement, quand on enlève une variable, est assez faible tant que le nombre de variables restantes est raisonnable.

c) La supériorité théorique des méthodes minimax, qui garantissent le résidu maximal, n'est pas évidente pratiquement. En effet, bien que les méthodes (A6) et (A8) minimisent le résidu moyen, on peut aussi considérer, dans cette moyenne (Fig. III-16 a et c) quel est le plus grand. On constate alors que le résidu maximal de (A8), sauf pour les premières variables entrées est parfois et même souvent inférieur au résidu maximal minimisé par (A7) pour le même pas ! (à partir de la 12ème variable).

La comparaison entre (A6) et (A5) est beaucoup moins nette, puisque le résidu maximal de (A6) n'est inférieur à celui de (A5) qu'à 2 ou 3 pas seulement.

Ceci est dû au caractère sous-optimal des méthodes pas à pas. Par exemple, au 1er pas, le résidu maximal de (A7) est certes inférieur à celui de (A8), mais ensuite ce n'est plus forcément vrai car les variables sélectionnées par les 2 méthodes diffèrent.

d) Toujours au niveau des méthodes, il faut noter les problèmes numériques qui peuvent apparaître. Les méthodes descendantes, qui nécessitent l'inversion préalable de la matrice, ont parfois des difficultés à démarrer. Mais surtout, compte tenu des fortes intercorrélations, les premières variables éliminées sont assez instables et peuvent être différentes pour de faibles perturbations numériques. Par exemple, si l'on applique (A5) et (A7) sur les mêmes matrices de corrélation tronquées à 3 décimales au lieu de 4, on constate des modifications.

Sur le plan de notre réseau proprement dit, on peut constater (Fig. III - 17 a,b,c,et d) que :

Pour les méthodes ascendantes (A7) et (A8), celle qui minimise le résidu maximal semble plus erratique dans le choix des stations. Par contre (A8) choisit d'abord une station centrale, la 32 puis balaie assez méthodiquement le pourtour du réseau (il

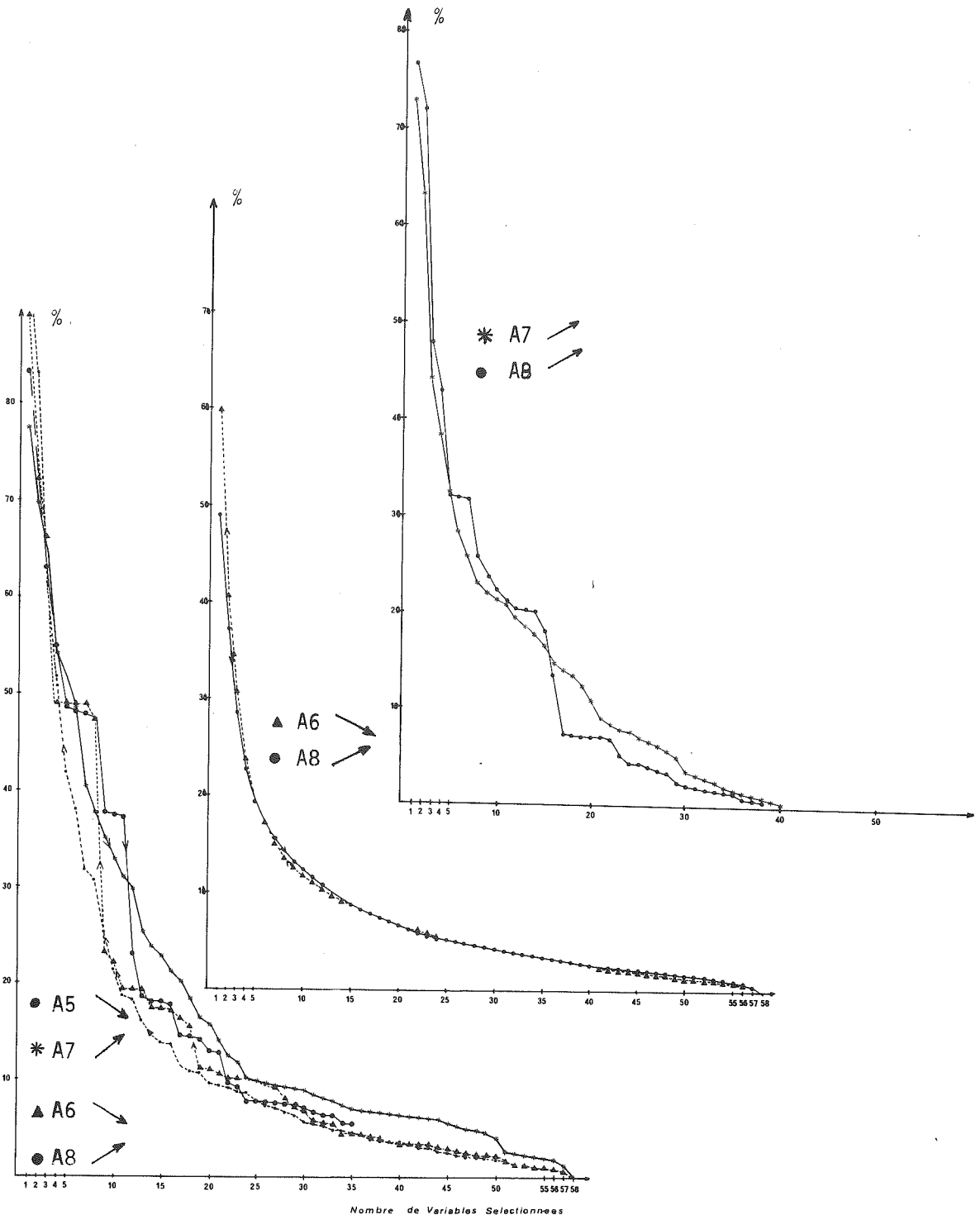


FIGURE III - 16 : Variance résiduelle maximale ou moyenne (en %)

a) valeurs maximales sur les épisodes .

b) valeurs moyennes sur les épisodes .

c) valeurs maxi en données mensuelles
(méthodes ascendantes seulement).

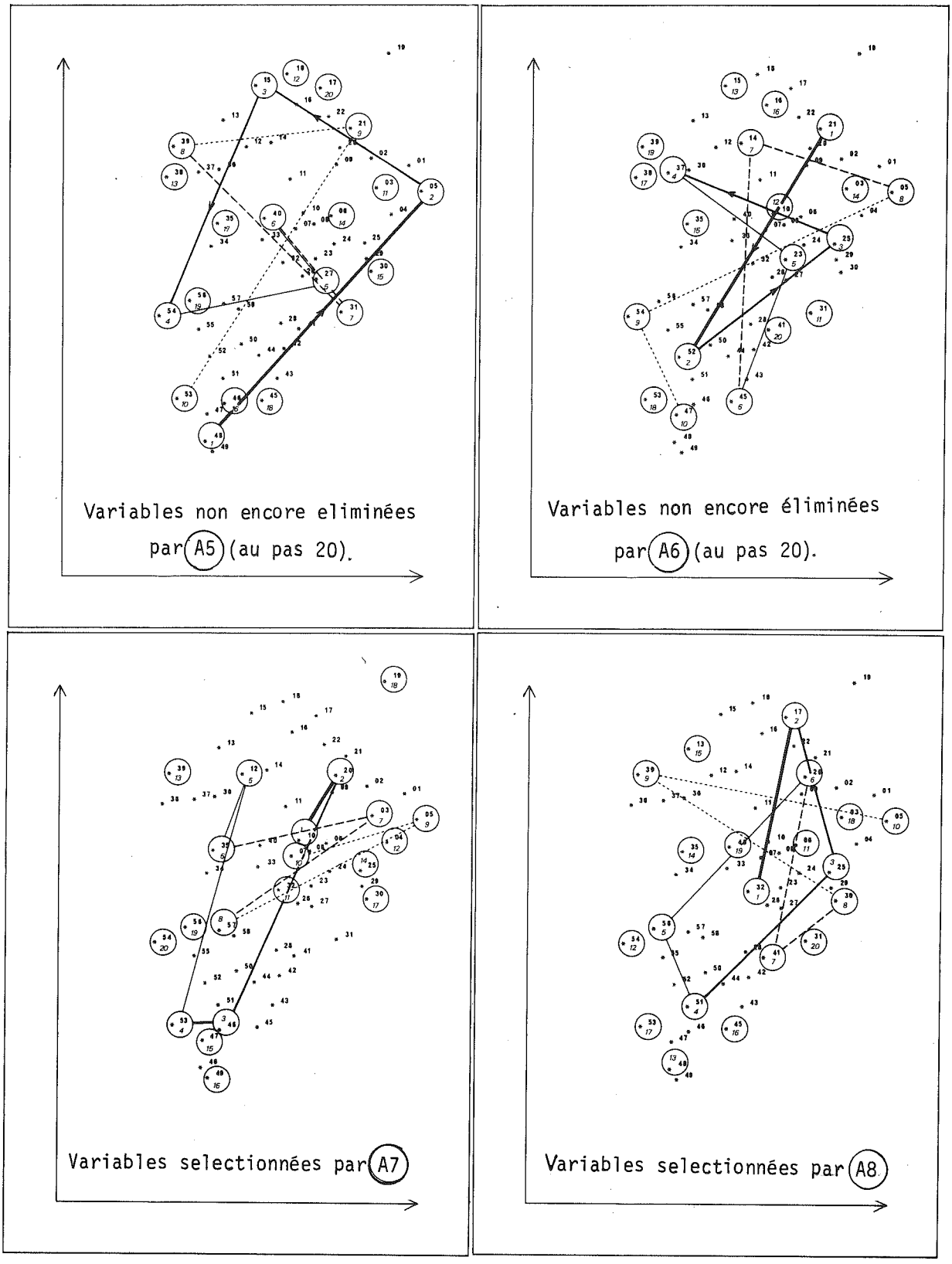


FIGURE III - 17 : Choix des 20 premières stations à retenir par A5 , A6 , A7 , et A8.
(on a joint les variables dans l'ordre de leur rang de sortie) .

faut attendre le 11ème pas pour revenir au centre). A l'inverse, (A7) donne un résultat final beaucoup moins bien réparti avec des groupes de stations un peu trop proches.

Les méthodes descendantes (A5) et (A6) présentent la même caractéristique : meilleure répartition par (A6). On constate aussi qu'elles éliminent en dernier les stations situées en bordure du réseau. Cela peut s'interpréter si on admet que l'on a un processus de nature Markovienne car alors, sur les bords, on ne dispose d'information que d'un seul côté. Leur reconstitution est donc moins bonne et il faut les garder, car elles expliquent bien les autres, mais sont mal expliquées par elles.

On constate que les répartitions obtenues par (A6) et (A8) sont très voisines avec même une meilleure distribution pour (A8) (supérieure en reconstitution à partir de 22 variables).

Le choix du nombre de variables peut se faire en se fixant un pourcentage de variance à reconstituer.

Si on choisit 90%, il faut 13 variables par (A6) ou (A8) (critère moyen) et 20 à 24 par (A5) ou (A7) (critère minimal). On constate que cela rejoint le nombre de groupes retenus précédemment. Si enfin on plaque les résultats de (C3) sur ceux de (A8), on constate que (A8) revient pratiquement à choisir une station pour chaque groupe proposé par (C3) d'où une très bonne cohérence.

Enfin tous les résultats ci-dessus correspondent aux données par épisodes. L'utilisation des données mensuelles modifie sensiblement les résultats. Les corrélations étant encore plus élevées, les méthodes descendantes ont toutes des problèmes de démarrage (il faut éliminer manuellement jusqu'à 15 variables). De plus, les choix se font sur des différences de variance expliquées très ténues et sont probablement moins stables. Par exemple, le réseau de 20 stations obtenu par (A8) en valeurs mensuelles est moins satisfaisant que celui obtenu en valeurs d'épisodes .

Il importe donc de bien choisir au départ la variable de travail (probablement le pas de temps le plus fin à utiliser ensuite).

D'autre part, il faut noter que l'on a considéré la variance expliquée de variables centrées réduites. Il resterait à voir s'il ne vaudrait pas mieux travailler en variance vraie.

IV.2.4. Conclusions

L'optimisation des réseaux se présente en pratique de manière beaucoup moins académique et doit prendre en compte beaucoup de contraintes matérielles qui ne sont pas envisagées ici. Néanmoins, l'utilisation complémentaire des diverses techniques présentées ici permet de cerner assez correctement :

- le nombre d'influences régionales indépendantes et significatives
- les groupements de stations interdépendantes
- et les stations les plus représentatives permettant de reconstituer les autres avec une bonne qualité.

Par contre, nous n'avons pas envisagé le problème de l'extension des réseaux, ni celui de la réduction des fréquences d'observations, etc...

"Parmi les trois étendues, il faut compter le temps, l'espace et le silence. L'espace est dans le temps, le silence est dans l'espace.."

Joseph Joubert (1754-1824)

QUATRIEME PARTIE

ANALYSE DE DONNEES REGIONALISEES ET ANALYSE HARMONIQUE DE PROCESSUS ALEATOIRES .

Elle concerne les aspects particuliers de l'analyse des données (surtout l'A.C.P. classique) quand celle-ci est appliquée à des données issues d'un processus continu. Le traitement des processus a été développé de façon considérable pour les processus temporels (séries chronologiques) et pour le traitement du signal. Sans en présenter la théorie complète, il sera nécessaire de rappeler certaines définitions ou notations. Mais, il nous arrive fréquemment de considerer des processus à 2 dimensions (champs spatiaux) dont le traitement, sensiblement plus complexe, n'a pas bénéficié d'un effort théorique équivalent à celui des processus temporels.

En dépit d'inévitables incursions dans ces aspects théoriques, le but de cette partie reste l'interprétation des analyses classiques appliquées à des processus, et la façon dont les propriétés du processus transparaissent dans ces analyses.

Le chapitre I met en évidence les relations entre les estimations climatologiques et les estimations spatiales des variances et corrélations .

Dans le chapitre II, on applique l'A.C.P. aux données des épisodes cévenols et on cherche à mettre en évidence une typologie. On fait aussi l'analogie entre les différents codages utilisés et des modèles simples d'analyse de la variance.

Le chapitre III considère quelques processus simples régis par des équations différentielles et les fonctions de corrélations classiques qui leur sont associées.

Le chapitre IV est sans doute le plus important car il traite l'A.C.P. des processus (encore appelée Analyse Harmonique dans ce cas).Les résultats théoriques, essentiels pour l'interprétation de ces analyses, serviront de base à une méthode d'interpolation optimale présentée en V^{ème} partie.

CHAPITRE I

RAPPELS ET DEFINITIONS

I.1 - Diverses notions de corrélation

I.1.1. Rappels

(a) On a considéré jusqu'à présent des ensembles de données regroupant, par exemple :

$\mathcal{N}$ $i=1...N$ { . valeurs (journalières ou mensuelles) ; . évènements ; . champs ; . réalisations } sur P $j = 1 \ldots P$ { . variables ; . stations ; . localisations géographiques }

Ces variables étaient considérées comme des variables aléatoires distinctes dont on calculait, sur le N-échantillon disponible :

m_j = moyenne de la variable j et r_{jl} = coefficient de corrélation

σ_j = écart-type " " entre les variables j et l .

Tous ces paramètres étaient donc calculés sur N tirages de la variable X_j, supposés indépendants sans que l'ordre des tirages influe sur le résultat.

Nous dirons que ces paramètres sont des estimateurs d'espérances climatologiques. En particulier, nous noterons $R_{jl} = E[r_{jl}]$ où E est l'espérance sur un grand nombre de tirages.

Mais il ne s'agit pas d'une espérance temporelle au sens strict, car même si ces tirages indépendants sont naturellement échelonnés dans le temps, celui-ci n'est pas supposé influer sur la variable aléatoire. (A la limite, on peut permuter aléatoirement les observations sans changer les résultats).

Dans d'autres cas, la variable aléatoire X_j devient vraiment fonction de t et on parlera d'une fonction, ou d'un processus aléatoire $X_j(t)$. On lui appliquera alors les notions classiques associées aux séries chronologiques comme par exemple la notion d'autocorrélation temporelle:

$$\mathcal{E}_t [X_j(t) \cdot X_j(t+k)] = \rho_t(k)$$

en lui réservant la notion d'espérance temporelle.

Dans le cas où t, ou x représente une variable autre que le temps, par

exemple l'espace $X(x,y)$ on parlera d'espérance spatiale.

On réservera la notion d'espérance spatio-temporelle aux cas où le processus est fonction à la fois de x et t (Par exemple la précipitation au sein d'un épisode, ou d'une averse : $X(x,t)$).

Remarque 1 : Dans la mesure où nous allons utiliser la terminologie des processus aléatoires, il importe d'en donner une définition.

La plus simple consiste à dire que si, pour tout ensemble quelconque de valeurs de la variable t soit $\{t_1, t_2, \ldots t_n\}$ on connait la fonction de répartition de $X(t_1)$, $X(t_2)$, ... $X(t_n)$, soit

$$Pr\left(X(t_1) < x_1 \cap X(t_2) < x_2 \cap \ldots\ldots X(t_n) < x_n \right)$$

alors cette famille de fonctions de répartition définit un processus aléatoire.

(b) Si on revient aux notions de corrélation, on définit, dans le cas des processus, la covariance ou fonction d'autocorrélation :

$$\Gamma(t_1, t_2) = E\left[X(t_1) . X(t_2) \right]$$

qui est en fait un produit croisé non centré ni réduit.

La fonction d'autocovariance est, elle, centrée :

$$C(t_1, t_2) \text{ ou } \mu(t_1, t_2) = E\left[\left(X(t_1) - m(t_1) \right) . \left(X(t_2) - m(t_2) \right) \right]$$

La fonction coefficient de corrélation, qui a réellement la dimension d'une corrélation, s'écrit :

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1) . C(t_2, t_2)}}$$

En fait, on verra que pour les phénomènes aléatoires supposés de plus stationnaires, toutes ces définitions peuvent être rendues équivalentes.

I.1.2. La corrélation spatiale

Nous allons maintenant appliquer ces définitions au cas des champs spatiaux.

(a) On considère par exemple des processus $X(t)$, ou $X(x)$ où la variable t ou x est une variable d'espace à 1, 2 ou 3 dimensions.

On dira alors qu'un champ observé (le long d'une ligne, d'une surface ou d'un volume) est une réalisation de la fonction aléatoire f (avec $f = f(x)$, $f(x,y)$ ou $f(x,y,z)$).

On trouvera d'ailleurs dans J. BASS (1962) une autre définition élémentaire d'une fonction aléatoire :

$$X(x) = f(x, A) = f(x, A_1, A_2, \ldots A_n \ldots)$$

où - x est une variable ordinaire
- $A_1, A_2, \ldots A_n \ldots$ un ensemble, fini ou non, de variables aléatoires qui, une fois fixées, définissent $X(x)$ comme une réalisation, ou une épreuve sur la fonction aléatoire $X(x, \omega)$.

Les moments de la fonction se définissent comme des moyennes d'ensemble :

$$\mu(x_0) = \langle X(x_0) \rangle \qquad \sigma^2(x_0) = \langle (X(x_0) - \mu(x_0))^2 \rangle$$

que l'on peut définir comme l'espérance mathématique, à x_0 fixé, sur un grand nombre de réalisations. De même, on peut appeler corrélation entre 2 points :

$$\rho(x_0, y_0) = \frac{\langle (X(x_0) - \langle X(x_0) \rangle)(X(y_0) - \langle X(y_0) \rangle) \rangle}{\sigma(x_0) \cdot \sigma(y_0)}$$

(b) On considèrera souvent que les phénomènes auxquels nous nous intéressons sont stationnaires. La définition stricte implique que tous les moments, au sens des moyennes d'ensemble, soient strictement indépendants de l'abscisse x.

$$\mu(x_0) = \mu \quad \forall x_0 \qquad \sigma(x_0) = \sigma \quad \forall x_0$$

et

$$\Gamma(x_0 + \Delta x, x_0) = \Gamma(\Delta x) \quad \forall x_0$$

On se contente souvent d'une stationnarité d'ordre 2, limitée aux 3 moments ci-dessus.

(c) L'application du principe ergodique nous conduit alors à approcher les moyennes d'ensemble par des moyennes au sein d'une seule réalisation (moyenne dans le champ, ou moyenne temporelle selon la nature de la variable x ou t). On trouvera la démonstration dans de nombreux manuels (GNEDENKO , 1976, et des considérations intuitives dans BASS, 1962).

Dans notre cas, nous noterons avec l'indice s les quantités ainsi définies spatialement. La stationnarité suppose :

$$\forall M = (x, y) \qquad m_s(M) = \mathcal{E}_s[X(M)] = m_s$$
$$\sigma_s(M) = \mathcal{E}_s[(X(M) - m_s)^2] = \sigma_s$$

et la covariance :

$$C_s(M_0, M) = C_s(\overrightarrow{M_0 M}) \quad \forall M_0$$

c'est-à-dire ne dépend que de la direction et de la norme du vecteur, mais pas de son point d'appui.

Dans le cas de 2 variables, on dit aussi que la fonction $X(x, y)$ est homogène. Si de plus, ses propriétés sont identiques dans toutes les directions de l'espace, c'est-à-dire :

$$C_s(\overrightarrow{M_0 M}) = C_s(\| M_0 M \|) = C_s(d)$$

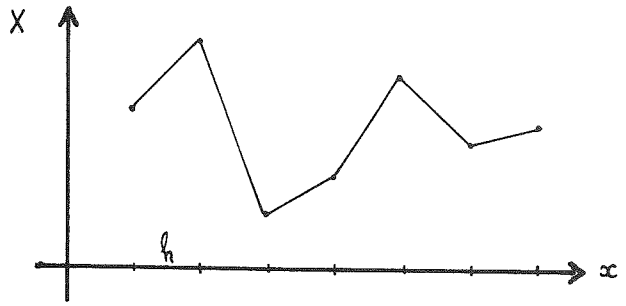
alors elle est isotrope.

Dans le chapitre suivant, nous verrons plus en détail comment on les estime.

En général, si on suppose le processus homogène et isotrope, la connaissance du corrélogramme (ou de la fonction d'autocorrélation) $\rho(d)$ est suffisante.

I.1.3. Le variogramme et la notion de dérive

(a) L'hypothèse la plus critique dans ce qui précède est celle de la stationnarité de la moyenne. Ce problème se rencontre assez souvent dans l'analyse des séries chronologiques affectées de tendance saisonnière ou à long terme, et la première façon d'y pallier est de travailler sur les écarts de la fonction entre les dates successives (cf JOHNSTON J., 1963 et BOIS Ph., 1976, ou différentes études E.D.F-D.T.G par D. DUBAND).



De même, une bonne façon de mesurer si une courbe est lisse, ou fortement autocorrelée positivement est de calculer la quantité :

$$\sum_{j=1}^{P-1} \left(X(x_j+h) - X(x_j) \right)^2$$

C'est à un terme constant $P.h^2$ près, la longueur de l'arc de courbe qui caractérise bien la variabilité. Naturellement, on la norme par le nombre de pas, d'où la moyenne quadratique de l'écart (différence première). Si le nombre P augmente, on tend vers :

$$\gamma(h) = \mathcal{E}\left[\frac{1}{2} \left(X(x+h) - X(x) \right)^2 \right]$$

C'est le demi-variogramme, ou encore la variance de la différence première.

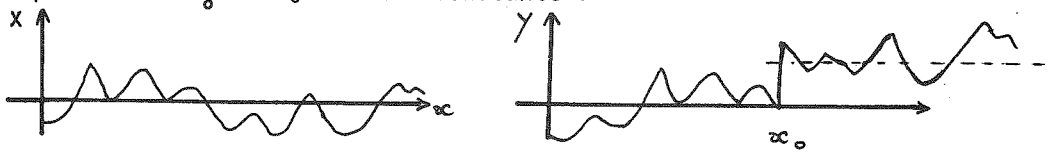
On peut naturellement le calculer pour h, $2h$, ... etc et tracer la courbe $\gamma(h)$ où h est la distance.

(b) L'intérêt du variogramme par rapport au corrélogramme provient de sa moins grande sensibilité aux non-stationnaı tés. Si le processus aléatoire X qui nous intéresse est stationnaire, et si le processus observé est Y, on cherche quels sont les effets de la présence de non stationnarité dans Y sur la connaissance de X.

Nous nous limiterons à supposer que Y est la superposition de X plus une tendance déterministe qui provoque la non-stationnarité. Naturellement, si cette tendance est connue ou identifiable, on peut facilement en déduire X. Mais le problème apparaît quand cette tendance n'est pas très évidente, auquel cas on traite Y comme un processus stationnaire, identique à X.

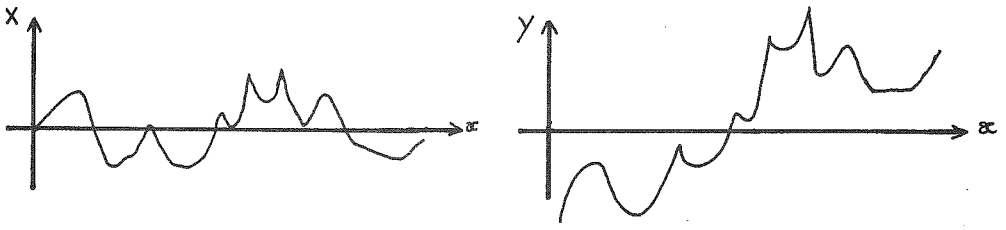
Nous allons anticiper un peu sur les techniques d'estimations, mais il est évident que si, connaissant une longueur donnée de la réalisation :

- A partir de x_0 on ajoute une constante :



L'estimation de m_Y sera décalée d'une constante de même que les moments d'ordre 2 (variance et corrélation) par rapport à ceux de X .
Par contre, au niveau des écarts, l'effet sera minime et le variogramme n'en sera pratiquement pas affecté.

- Si on ajoute maintenant une tendance linéaire, par exemple : $Y(x) = X(x) + \alpha x + \beta$



On constate que la moyenne de Y dépend du tronçon de réalisation utilisé, et que la variance calculée dans l'hypothèse stationnaire va dépendre de l'étendue du tronçon, ce qui affectera évidemment les corrélations.
Par contre le variogramme de $Y(x)$, ou variogramme <u>brut</u>, sera moins affecté puisque la moyenne ou l'espérance, de $[Y(x+h) - Y(x)]^2$ sera égale à celle de $[X(x+h) - X(x)]^2 + 3\alpha^2 h^2$ donc ne diffère du variogramme <u>vrai</u> que d'un terme constant. (Il y a seulement décalage d'origine.)

- La généralisation à une tendance de degré plus élevé est possible. Il suffit alors de considérer non plus la différence première $Y(x+h) - Y(x)$ mais l'accroissement du second ordre :

$$Y(x+h) - 2Y(x) + Y(x-h)$$

Sa variance, et celle du même accroissement associé à X , sont identiques pour des tendances linéaires ou par paliers, et ne diffèrent que d'une constante pour une tendance quadratique $\alpha x^2 + \beta x + \delta$.
C'est l'embryon de la notion de "fonction aléatoire intrinsèque d'ordre k": Etant donné un signal $Y(x)$ que l'on présume être la somme d'une dérive et d'un processus aléatoire stationnaire $X(x)$, alors on travaille sur les différences d'ordre croissant jusqu'à trouver l'ordre k pour lesquelles ces différences sont stationnaires (cf CHILLÈS J.P., 1977).

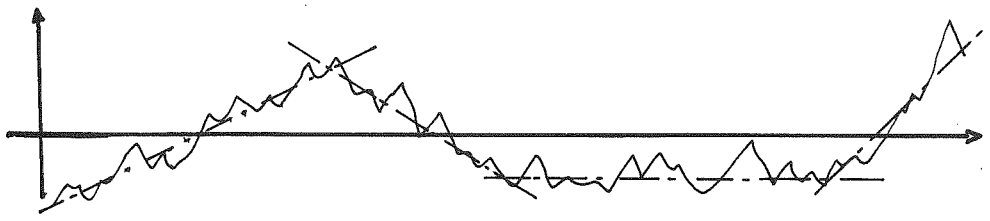
Ainsi, on peut considérer que sur le plan théorique, le corrélogramme filtre les polynômes de degré 0, le variogramme des polynômes de degré 1, etc...

Cette pratique est assez générale tant dans le domaine temporel que spatial. Elle consiste à analyser les phénomènes instationnaires en se ramenant à une somme de 2 processus, l'un déterministe et l'autre aléatoire, mais cette fois stationnaire. Par exemple, si $E[X(x)] = m(x)$ au sens de la moyenne d'ensemble, on va chercher à le modéliser par

$$X(x) = P(x) + Y(x)$$

avec $Y(x)$ stationnaire, et on appellera $P(x)$ la dérive.

Cette notion de dérive est assez délicate à justifier en pratique, quand on ne dispose que d'un tronçon fini d'une réalisation:



On considère qu'une partie du processus est déterministe si on peut l'expliquer par le passé lointain du phénomène, tandis que la partie aléatoire a un rayon de corrélation beaucoup plus petit.

Par exemple, si on considère le processus :

$$X(x) = a \cdot \cos(\omega_0 . x + V) + U(x)$$

où $U(x)$ est un bruit blanc.

Si on suppose que l'ensemble, c'est-à-dire X, est stationnaire, on calcule une autocorrélation spatiale

$$\rho_s(h) = \frac{a^2}{a^2 + 2\sigma_\mu^2} \cdot \cos h\omega_0$$

qui réaugmente quand la distance croît, ce qui est légitime mais choquant (l'aléa en x conditionne d'autant plus $X(x+h)$ que h ↗)

Et si on dispose d'une seule réalisation, on préfèrera considérer que l'on a une dérive "déterministe" valable en tout point de la réalisation, soit

$$P(x) = a \cos(\omega_0 . x + V)$$

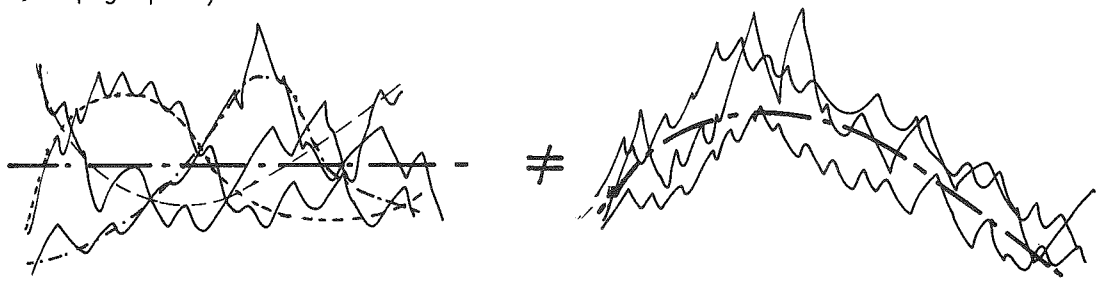
plus un bruit aléatoire (et ici non autocorrélé).

Par contre, si on a plusieurs réalisations, ou si on envisage l'ensemble des

réalisations, alors, on peut parler d'une dérive pour chaque réalisation, mais pas d'une composante déterministe car V peut être aléatoire et dans ce cas, les dérives ne se superposent pas d'une réalisation à l'autre : Il n'y a pas de dérive "climatologique" mais une composante aléatoire de grande portée (ou ici de grande longueur d'onde).

Exemple : Le champ d'un épisode pluvieux peut rarement être considéré comme une surface oscillant autour d'un plan horizontal. On considère alors qu'il y a une dérive spatiale, linéaire, ou quadratique, etc... et on entre dans le domaine d'application du krigeage universel ou des fonctions aléatoires intrinsèques.

Si par contre, pour tous les épisodes assimilés à des réalisations indépendantes, on obtient la même tendance, linéaire ou quadratique, dans le même repère fixe, alors il y a une dérive climatologique, qui est souvent d'origine déterministe (distance à la mer, topographie).



I.2 - Quelques problèmes d'estimation

I.2.1. Estimateurs divers

(a) Nous rappelons simplement les estimations climatologiques. Dans ce qui suit nous noterons E l'espérance climatologique ou la moyenne d'ensemble, d'où si on a N observations d'un champ de P stations, et si on suppose pour simplifier les données X_{ij} centrées réduites :

$$E[X_j] = E[X(x_j)] = \mu_j \qquad \text{estimé par} \quad X_{\cdot j} = \frac{1}{N}\sum_{i=1}^{N} X_{ij} \quad (=0)$$

$$E[(X_j - \mu_j)^2] = \sigma_j^2 \qquad " \qquad \sigma_j^2 = \frac{1}{N}\sum_i X_{ij}^2 \quad (=1)$$

$$E[(X_j - \mu_j)\cdot(X_k - \mu_k)] = R_{jk} \qquad " \qquad r_{jk} = \frac{1}{N}\sum_i X_{ij}\cdot X_{ik}$$

(b) Dans le domaine spatial, on définira de même, pour un épisode donné i :

$$\mathcal{E}_s[X(x)] = m_i \qquad \text{estimée par} \quad X_{i\cdot} = \frac{1}{P}\sum_{j=1}^{P} X_{ij}$$

$$C_s(0,i) = \mathcal{E}_s[(X(x) - m_i)^2] \qquad " \qquad \frac{1}{P}\sum_{j=1}^{P}(X_{ij} - X_{i\cdot})^2$$

$C_s(h,i)$ la covariance spatiale ou autocovariance, estimé ici par :

$$\frac{1}{\ell} \sum_{j,k} (X_{ij} - X_{i.})(X_{ik} - X_{i.})$$

sommé sur les ℓ couples (j,k) distants de h.
Et de même l'autocorrélation au sein de la réalisation i : $\rho_s(h,i) = \frac{C_s(h,i)}{C_s(0,i)}$

On calcule de même le demi-variogramme :

$$\gamma_i(h) = \frac{1}{2} E_s[X(x+h) - X(x)] \quad \text{par} \quad \frac{1}{\ell} \sum_{j,k} (X_{ij} - X_{ik})^2$$

sur les ℓ couples distants de h.

Et on peut montrer que, si le processus est stationnaire d'ordre 2, donc possède une fonction d'autocovariance, alors :

$$\gamma(h) = 2\, C_s(0).[1 - \rho_s(h)]$$

On fera l'analogie avec la relation entre la distance de 2 variables dans $\mathbb{R}^N$ et leur corrélation (cf Ière partie, Ch. II-3).

En pratique, il est rare d'avoir ℓ couples strictement distants de h, sauf pour les réseaux à mailles régulières, et on est conduit à faire des moyennes par classes de distances (cf D. CREUTIN, 1979).

I.2.2. Aperçus sur quelques relations entre les estimateurs climatologiques et spatiaux

On va supposer que les variables sont centrées et réduites (ou "standardisées") au sens habituel, c'est-à-dire :

$$X_{ij} \longrightarrow E[X_j] = 0 \qquad E[X_j^2] = 1 = \sigma_X^2$$

Dans ce cas, si on considère les estimateurs que l'on a décrit précédemment, on voit que :

(a) $$\frac{1}{N} \sum_{i=1}^{N} C_s(0,i) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{P} \sum_{j=1}^{P} (X_{ij} - X_{i.})^2 \right\} = \sigma_X^2 - \frac{1}{N} \sum_{i=1}^{N} X_{i.}^2$$

Soit :

La moyenne climatologique des variances spatiales $C_s(0,i)$ est égale à la variance climatologique de X moins la variance climatologique des moyennes spatiales $X_{i.}$.

(b) $$\frac{1}{N}\sum_{i=1}^{N} X_{i\cdot}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{P}\sum_{j=1}^{P} X_{ij}\right\}^{2} = \frac{1}{P}\,\sigma_x^{2} + \frac{2}{P^2}\sum_{j<k} r_{jk}$$

Or il y a $\frac{P.(P-1)}{2}$ couples (j,k) $j\neq k$ et le second terme est peu différent de la moyenne des r_{jk} sur le réseau.

La variance climatologique de la moyenne spatiale $X_{i\cdot}$ est donc égale à la moyenne spatiale des variances climatologiques plus la moyenne spatiale des coefficients de corrélation climatologiques.

(c) Plus généralement :

$$C_s(h,i) = \mathcal{E}_s\left[(X_{ij}-m_i).(X_{ik}-\underbrace{m_i}_{X_{i\cdot}})\right] = \mathcal{E}_s\left[X_{ij}.X_{ik}\right] - m_i^2$$

avec $h = d_{jk}$ distance entre les stations j et k.

Si on considère maintenant l'espérance climatologique :

$$E\left[C_s(h,i)\right] = E\left[\mathcal{E}_s\left[X_{ij}.X_{ik}\right]\right] - E\left[\left(\mathcal{E}\left[X_{ij}\right]\right)^2\right]$$

$$= \mathcal{E}_s\left[\underbrace{E\left[X_{ij}.X_{ik}\right]}_{r_{jk}}\right] - E\left[m_s^2\right]$$

soit :

$$E\left[C_s(h)\right] = \mathcal{E}_s\left[R(h)\right] - E\left[m_s^2\right]$$

On voit donc apparaître un effet de "taille" des évènements, lié au niveau moyen m_i du champ. Par contre, si $\mathcal{E}_s\left[X_{ij}\right] = m_i = 0 \quad \forall i$

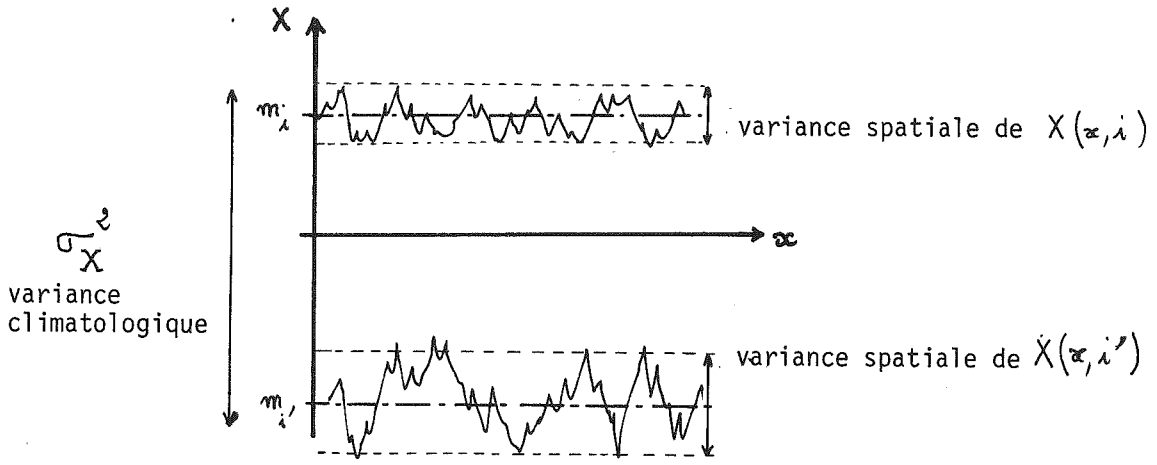
alors

$$E\left[C_s(h)\right] = \mathcal{E}_s\left[R(h)\right]$$

ce qui est une forme d'écriture du principe ergodique.

Celui-ci permet de calculer les moments du processus, (à condition qu'il soit stationnaire ou bien sur un ensemble de réalisations, ou bien sur une seule à condition qu'elle soit suffisamment étendue.

(d) Visualisation simplifiée : Si on se place à 1 dimension, on peut imaginer que l'on a N signaux ne différant que par leurs moyennes. Les aspects spatiaux s'observent en regardant parallèlement à l'axe des x , les aspects climatologiques parallèlement à l'axe des y.



Et si on considère chaque épisode i comme une classe où l'on a pris P échantillons, on retrouve la décomposition classique de la variance (cf IIème partie, Chap.IV).

$$T = W + B$$

En fait, cela suppose que, dans la classe i, le tirage des P échantillons soit aléatoire. Ce n'est pas vrai ici si on suppose qu'il y a une corrélation spatiale suffisante pour que les P stations soient liées entre elles.

I.2.3. Effet de la corrélation spatiale

Celle-ci influe sur l'estimation de la moyenne du champ m_i ou $\mathcal{E}_s[X]$ calculée par $\frac{1}{P}\sum_{j=1}^{P} X_{ij}$. On peut en effet calculer son écart-type d'estimation :

$$\sigma_m^2 = \sigma_{X_{i.}}^2 = \frac{\sigma_x^2}{P} \sum_{l=-P+1}^{P-1} \left(1 - \frac{|l|}{P}\right) \rho_l$$

(BAYLEY et HAMMERSLEY,1946 cités par G.H. JOWETT, 1955).

Ceci est vrai dans le cas d'une dimension et de stations régulièrement réparties. On voit que si $\rho_l = 0$ sauf $\rho_0 = 1$, on retrouve la variance habituelle de l'estimateur σ_x^2/P, mais que si $\rho_l \longrightarrow 1$ alors on tend vers σ_x^2 : le fait d'échantillonner P points très corrélés ou un seul revient au même. La corrélation dans le champ tend à accroître la variance de l'estimation de la moyenne du champ. Cela doit être pris en compte si on veut tester par exemple l'hypothèse : les champs ont des moyennes différentes.

JOWETT propose par exemple de calculer la moyenne des variances spatiales de chaque champ, soit :

$$v = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{P^2} \sum_{j,k=1}^{P} \frac{1}{2} (X_{ij} - X_{ik})^2 \right\}$$

En effet, la variance dans le champ i, $C_s(0,i)$ estimée par

$$v_i = \frac{1}{P} \sum_{j=1}^{P} (X_{ij} - X_{i.})^2$$

peut aussi s'écrire :

$$\nu_i = \frac{1}{2P^2} \sum_{j,k=1}^{P} \left(X_{ij} - X_{ik} \right)^2$$

Ensuite on cherche ν_{0i} , défini par $\frac{1}{l} \sum_{j,k} \frac{1}{2} \left(X_{ij} - X_{ik} \right)^2$ calculé sur l couples j, k tels que $d_{jk} \geqslant \delta_0$, et δ_0 étant tel que les points j , k ne soit plus corrélés ($\sim$ palier du variogramme ou extinction des corrélations). Alors ν_{0i} , ou $\frac{1}{N} \Sigma \nu_{0i} = \nu_0$ estime la variance d'ensemble et l'estimateur de σ_m^2 devient :

$$Est\ \sigma_m^2 = \nu_0 - \nu$$

traduit bien la relation classique :

$$\sigma_B^2 = \sigma_T^2 - \sigma_W^2$$

en faisant l'analogie $\sigma_B^2 \equiv \sigma_m^2$

L'extension à 2 dimensions est facile quand on suppose le champ homogène et isotrope. On trouve que :

$$\sigma_m^2 = \sigma_{x.}^2 = \frac{1}{P^2} \sum_{j,k} E\left[C_s(d_{jk}, i) \right]$$

et que :

$$\sigma_W^2 = \sigma_{x_i}^2 = E\left[C_s(0, i) \right] = \sigma_x^2 - \sigma_{x.}^2$$

que nous avons déjà rencontré en I.2.2 (b).

On trouvera, par exemple dans YEVJEVICH V. (1972 , page 50) la variance du coefficient d'autocorrélation pour des processus autoregressifs σ_{ρ_1} .

On remarquera aussi que ce problème est analogue à la recherche d'un nombre équivalent d'échantillons indépendants pour apprécier la variance d'échantillonnage d'un estimateur. Nous avons déjà rencontré ce problème par ailleurs.

CHAPITRE II

ANALYSE D'UN ENSEMBLE DE REALISATIONS AU SENS DE L'ANALYSE DES DONNEES (EXEMPLE DES EPISODES CEVENOLS)

II.1 - Premières analyses

II.1.1. Climatologiques

La plupart des modèles de processus, comme les modèles d'analyse de la variance, supposent que le processus est stationnaire ou qu'il est stationnaire au sein d'une réalisation, etc... Il est donc intéressant de regarder ce qu'il en est sur l'exemple des épisodes cévenols.

Si on considère la moyenne d'ensemble (climatologique) de nos épisodes, elle n'est manifestement pas constante dans le champ (fig.IV.1). Cela se voit aussi sur les transects NW-SE et SW-NE. On peut donc admettre l'existence d'une composante déterministe qui,dans le sens transversal, serait plutôt liée à la topographie (encore que la crête pluviométrique soit légèrement décalée vers l'Est de la crête géographique, sauf vers l'Aigoual) tandis qu'elle s'interprêterait, dans le sens longitudinal, comme une décroissance avec l'éloignement à la mer.

Le même phénomène apparaît pour les écarts-types climatologiques, qui sont très liés aux moyennes (cf 1ère Partie, Fig. I-6 a : R = .967).

Enfin, si on trace soit le champ des corrélations d'une station quelconque au reste du réseau, soit le corrélogramme empirique des r_{jk} on retrouve une anisotropie importante entre les directions transversale et longitudinale, mais surtout la décroissance de la corrélation avec la distance est extrêmement lente. (On atteint 0.5 au bout de 60 à 80 km, et l'absence de corrélation, difficile à apprécier ici requiert au moins 150 à 200 km). Ceci est manifestement dû à la variabilité interépisode, mais il est un peu choquant de parler d'une corrélation de .95 entre 2 stations éloignées de 10 km quand on sait que, au sein d'un épisode, 2 pluviomètres placés à 10 km l'un de l'autre divergent souvent de 100% et parfois plus.

II.1.2. Spatiales

Si on descend au niveau de l'épisode, on peut calculer de même la moyenne et l'écart-type spatial de l'épisode. On constate là encore une forte liaison (R = .897) qui tend à montrer que la variabilité dans l'épisode croît avec la moyenne de l'épisode. Il faut cependant prendre des précautions à cause des effets dus à la dérive climatologique, ou à celle de l'épisode proprement dit.

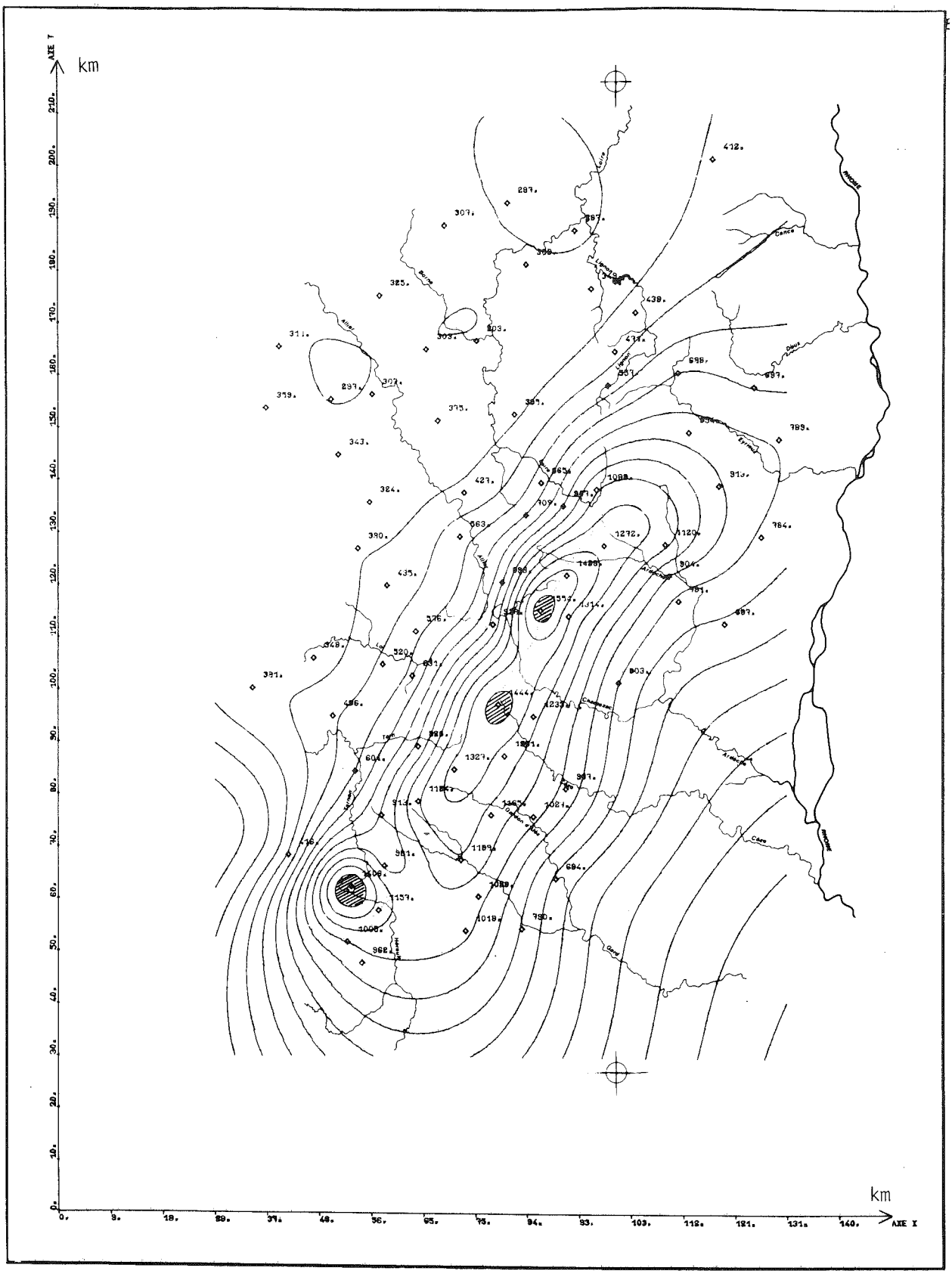


FIGURE IV - 1 : Carte de la moyenne des 84 épisodes cèvenols utilisés (interpolation spline - en 1/10 mm)

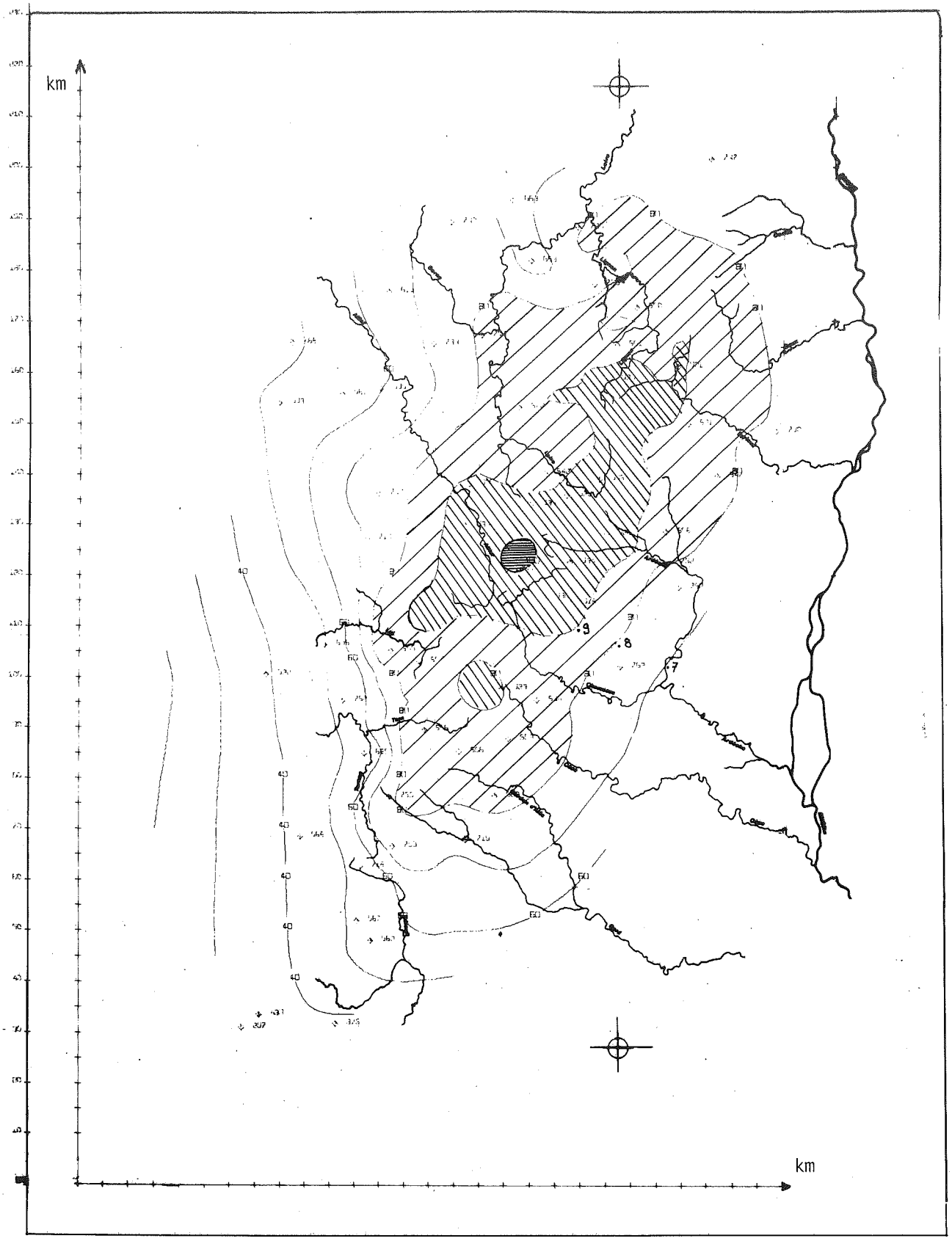
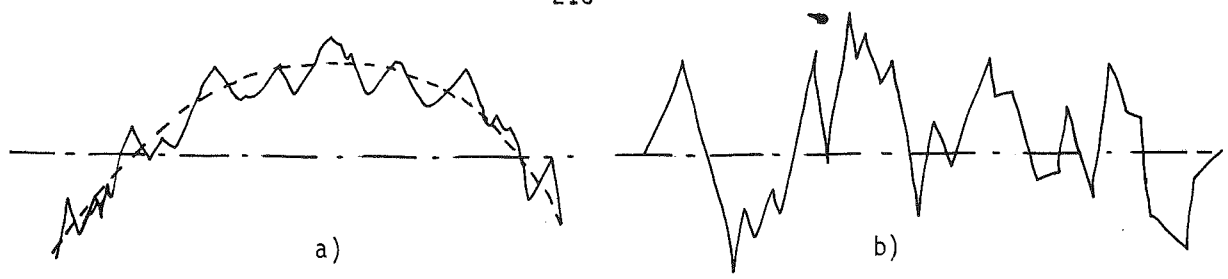


FIGURE IV - 2 : Carte des isocorrélations d'une station de référence (S^t Etienne de Lugdarès) avec l'ensemble des autres.



Bien qu'apparemment plus "lisse" que (b), le signal (a) aura le même écart-type spatial. Le même problème se pose, de façon encore plus aigüe, dans l'estimation du corrélogramme spatial. Enfin il faut garder à l'esprit la particularité de notre phénomène, qui n'est pas distribué symétriquement autour de la moyenne spatiale, surtout quand elle est faible, les valeurs étant tronquées à 0. Ceci tend à diminuer l'écart-type.et renforce donc la corrélation entre m_s et $C_s(0)$.

La présence de dérive rend donc difficile le tracé des corrélogrammes, mais aussi des variogrammes qui sont fortement biaisés. L'analyse de quelques épisodes présentant une faible dérive permet cependant d'estimer la portée du phénomène à une trentaine de km. Ceci est notoirement inférieur aux valeurs climatologiques et rejoint les valeurs obtenues par d'autres auteurs (15 km pour une région très montagneuse du bord de mer en Nouvelle Zélande (HUTCHINSON P., 1972) , contre 30 à 50 km dans la plaine normande (DELHOMME J.P., 1977)).

Malheureusement, en cas de dérive marquée,un choix se pose quant à ce que l'on considère comme la dérive. Ensuite, l'estimation du corrélogramme du spatial restant, ou du variogramme sous-jacent, requiert des techniques sophistiquées (cf J.P. DELHOMME, 1977). Cependant la plupart convergent vers une "portée" du phénomène (longueur de corrélation) qui est encore de l'ordre de 30 à 60 km (soit entre 1/5 et 1/3 de la valeur climatologique). Or cette portée est cohérente avec la physique du phénomène (cf AMOROCHO et WU, 1977) : existence de cellules de pluies de 4 à 10 km et d'une durée de vie de 15 à 20 minutes.

Il serait donc bon de ramener dans un cadre commun les corrélations calculées soit climatologiquement soit spatialement.

II.2 - A.C.P. et liaisons avec des modèles d'analyse de la variance

II.2.1. Sur les données brutes

On a vu précédemment que les moyennes climatologiques des stations variaient considérablement (de 29 à 157 mm) de même que les écart-types (de 19 à 128 mm) et même qu'ils cofluctuaient.

Il était donc exclu d'effectuer une A.C.P. non normée qui aurait donné des poids très différents aux différentes stations, encore que les résultats obtenus ne diffèrent pas très sensiblement du cas normé. Mais le souci de se rapprocher d'un

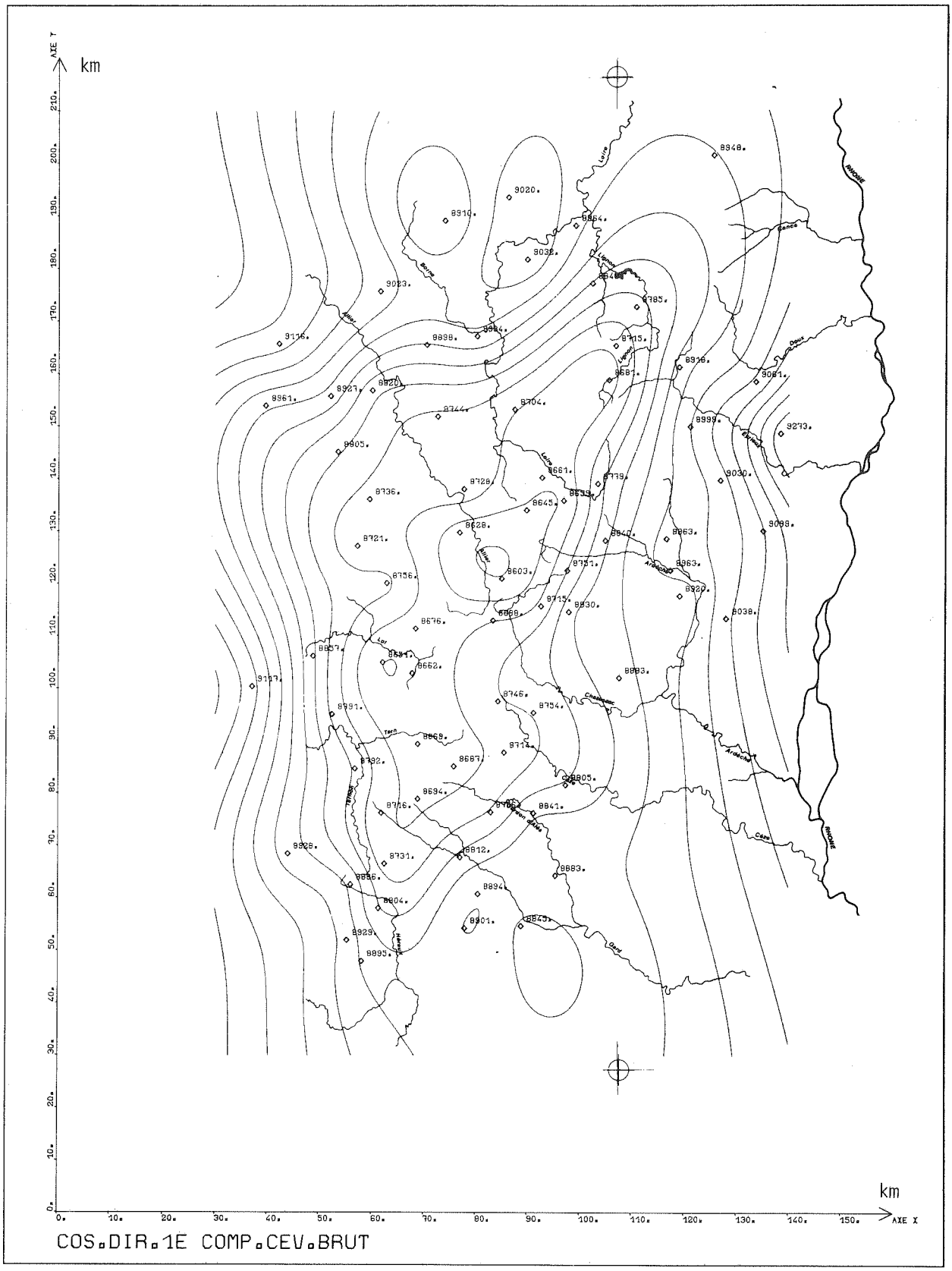


FIGURE IV - 3 : Carte des pondérations de la 1ère composante principale (ou de sa corrélation avec l'ensemble des stations).

processus stationnaire a conduit à utiliser les valeurs centrées et réduites.

(a) Les résultats de l'A.C.P., déjà partiellement décrits dans la IIIème Partie fournissent une première valeur propre très nettement dominante $\lambda_1 = 42.8$ (soit 59 % de la variance) et son vecteur propre associé présente des pondérations très voisines sur toutes les stations S_j (corrélations entre Z_1 et S_j toutes comprises entre 0.6 et 0.85).

On en donne la cartographie dans la figure IV-3. Ceci fait de la 1ère composante quelque chose de très voisin de la moyenne arithmétique des stations (en valeurs normées) et donc de la moyenne spatiale de l'épisode (sur les valeurs normées). On a d'ailleurs calculé les corrélations entre les observations à chaque station X_{ij} et les vraies moyennes arithmétiques de chaque épisode $m_i = \sum_j x_{ij}$, en valeurs brutes cette fois, mais on obtient un champ très voisin .

Le pourcentage de variance de cette composante, et donc de la moyenne spatiale, montre la prédominance de la variabilité interépisode sur la variabilité intraépisode.

Un autre phénomène curieux est la forte analogie qu'il y a entre la représentation des variables dans les axes 2 et 3 de $\mathbb{R}^N$ et leurs positions géographiques dans le plan x,y . Ceci s'expliquera dans le chapitre IV.

(b) Une façon d'interpréter ces analyses consiste à plaquer sur les données un modèle analogue à ceux utilisés en analyse de la variance. Par exemple, on peut imaginer que la valeur brute x_{ij} de la i ème observation s'écrit :

$$x_{ij} = \mu_j + \sigma_j . P_i + \sigma_j . \varepsilon_{ij}$$

avec μ_j de la forme $k . \sigma_j$.

Cela sous-entend que l'on a un processus sous-jacent $k + P_i + \varepsilon_{ij}$, qui est mesuré par la station S_j avec une amplification σ_j , variable d'une station à l'autre. Quand on revient en variables centrées réduites, on analyse donc :

$$X_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} = P_i + \varepsilon_{ij}$$

Si on estime P_i (ou si on pose) :

$$P_i = \frac{1}{P} \sum_{j=1}^{P} X_{ij} = \frac{1}{P} \sum_j \left(P_i + \varepsilon_{ij} \right)$$

cela impose : $\sum_{j=1}^{P} \varepsilon_{ij} = 0$

De plus : $\sum_{i=1}^{N} X_{ij} = 0 \Rightarrow \sum_i P_i + \sum_i \varepsilon_{ij} = \underbrace{\sum_i \left(\frac{1}{P} \sum_j X_{ij} \right)}_{=0} + \sum_i \varepsilon_{ij} = 0$

$$\Longrightarrow \sum_{i=1}^{N} \varepsilon_{ij} = 0 \quad \text{et} \quad \sum_{i=1}^{N} P_i = 0$$

Et quand on calcule la matrice de variances-covariances de X_{ij} (donc des corrélations de x_{ij}) on obtient :

$$r_{jl} = E\left[(P_i + \varepsilon_{ij})(P_i + \varepsilon_{il})\right] = \frac{1}{N} \sum_i (P_i + \varepsilon_{ij})(P_i + \varepsilon_{il})$$

Et le modèle suppose que P et ε sont indépendants, donc $E[P_i \cdot \varepsilon_{ij}] = 0$
Il reste alors :

$$r_{jl} = \frac{1}{N} \sum_i P_i^2 + \frac{1}{N} \sum_i \varepsilon_{ij} \cdot \varepsilon_{il}$$

Donc dans l'hypothèse d'un modèle où l'on considère le champ comme l'addition d'un terme moyen (spatialement) P_i , c'est-à-dire la puissance de l'épisode, plus un bruit aléatoire ε_{ij} de caractéristiques constantes, indépendant de P_i , alors la matrice de corrélation est la somme d'un terme représentant la variabilité des épisodes plus un terme spatial pur : $E[C_s(\varepsilon_j, \varepsilon_l)]$

$$r_{jl} = \sigma^2(P) + E[C_s(\varepsilon_j, \varepsilon_l)]$$

Si on s'intéresse seulement à cet aspect spatial, il suffirait d'analyser la matrice des écarts de la valeur centrée réduite à la moyenne spatiale $X_{i.}$ soit $X_{ij} - P_i$.

Un des intérêts serait de tracer l'espérance climatologique du corrélogramme spatial du processus $\varepsilon(x, y)$.

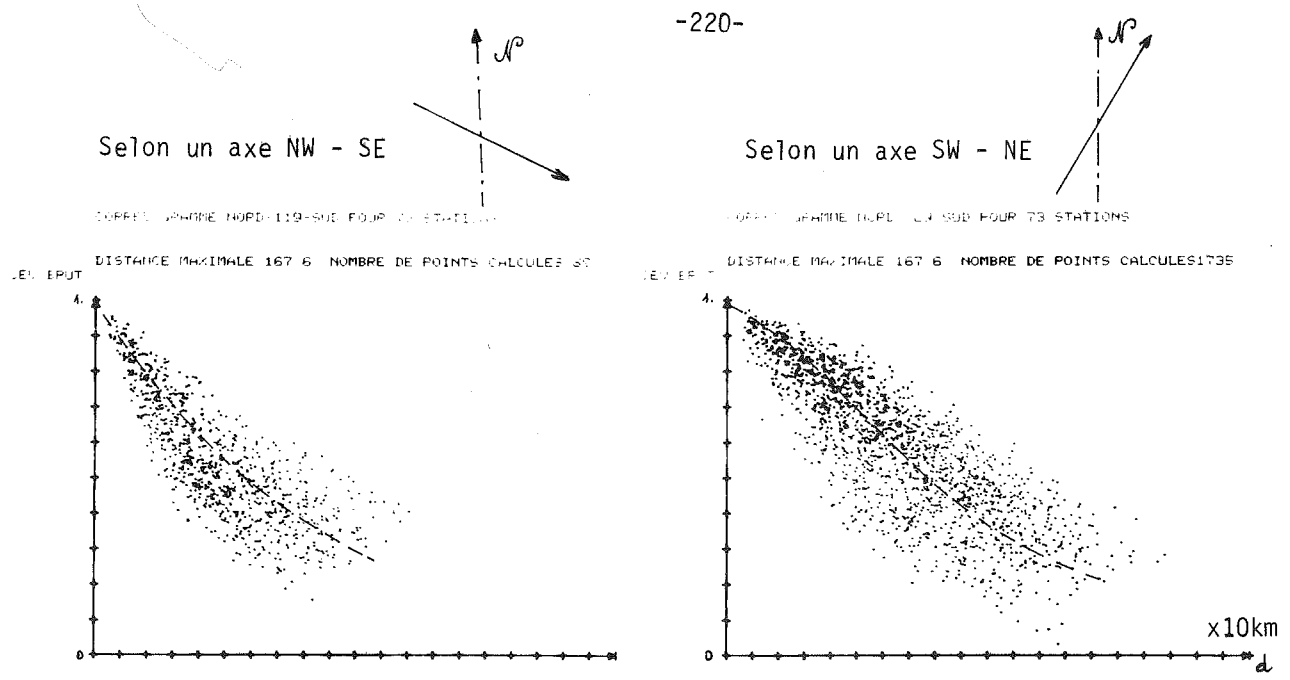
Une façon simple de l'obtenir consiste à remarquer que si la première composante Z_1, de vecteur propre V_1, de R est justement la moyenne spatiale $X_{i.}$, alors il suffit pour retrancher son influence de faire la déflation :

$$\left\{C_{jl}(\varepsilon)\right\} = \left\{R_{jl}(x)\right\} - \lambda_1 \cdot V_1 \cdot V_1^t$$

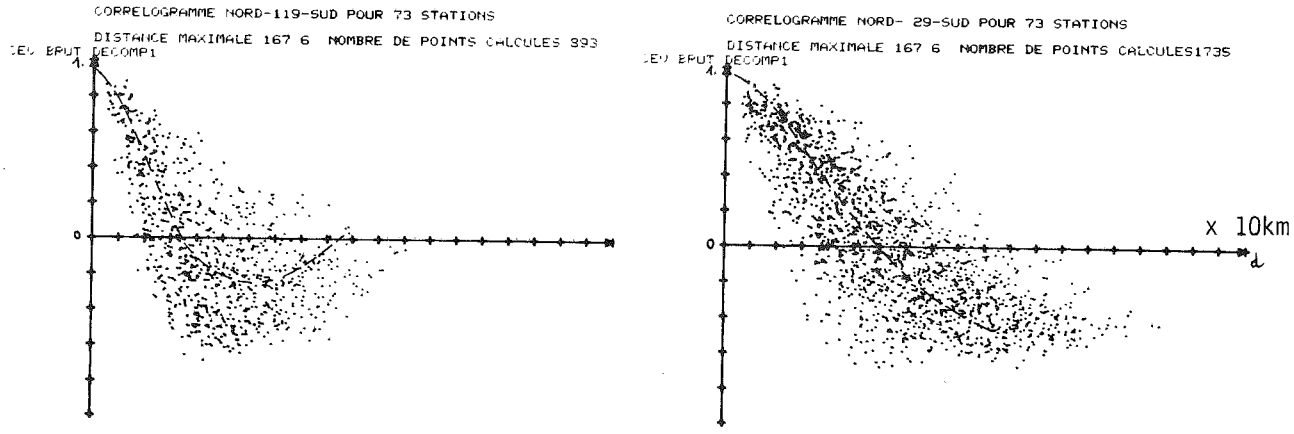
et de tracer $\rho_s(j,l) = \rho_s(d_{jl}) = C_{jl}$

(cf figure IV-4-b) et on voit bien que la corrélation spatiale s'éteint entre 30 et 60 km.

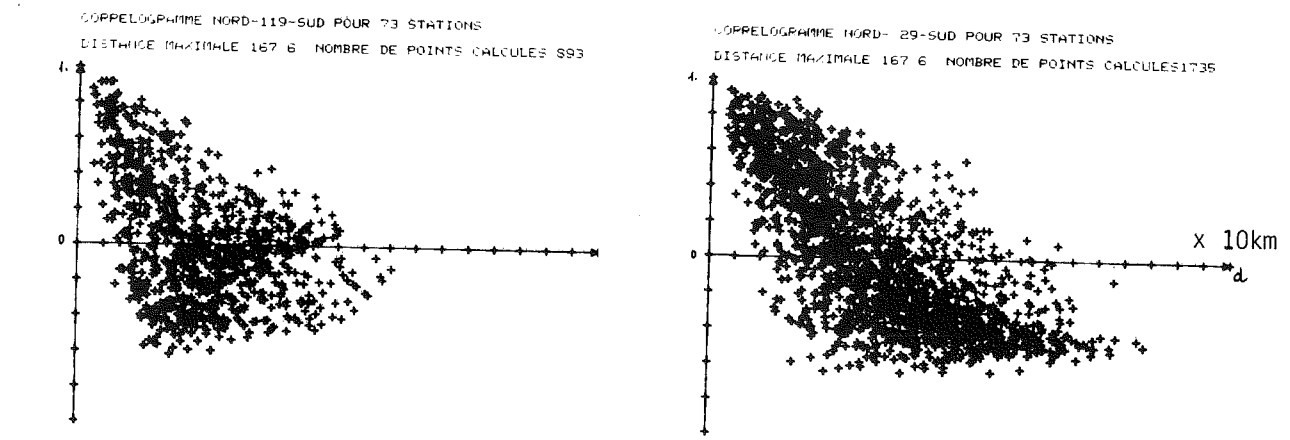
Remarquons pour terminer que l'analogie avec l'analyse de la variance se limite à un modèle très simple, et que les estimateurs employés sont très grossiers. Si par exemple on avait voulu tester la différence de moyenne entre tous les épisodes il aurait fallu prendre en compte la corrélation spatiale dans l'estimation de la moyenne, et affecter une surface d'influence à chaque station, etc... On en trouvera des exemples dans P. HUTCHINSON (1972) et CLARKE R.T. (1977). Le modèle utilisé est dit à effets fixés, en ce qu'il ne considère pas les $i = 1, \ldots N$ épisodes ni les $j = 1 \ldots P$ stations comme des échantillons aléatoires dans une population infinie.



a) Sur les données "brutes" .



b) Sur les données brutes mais après extraction de la 1ère composante.



c) Sur les données "profilées" .

FIGURE IV - 4 : Corrélogrammes "climatologiques " (calculés sur l'ensemble des épisodes) selon l'orientation et le codage des données.

Le développement dans le cas des données brutes s'applique aussi au cas où l'on effectue d'abord une transformation sur les variables brutes (en log. ou en racine carrée : données "radicalisées").

II.2.2. Sur les données profilées

Les hypothèses de base du modèle additif précédent ne sont toutefois pas complètement fondées, dans la mesure par exemple où l'écart-type spatial de l'épisode reste très lié à la moyenne de l'épisode (cf Ière Partie, fig. I-6-b) R = .897. La variabilité spatiale augmente donc avec la puissance de l'épisode. Ceci peut être dû en partie à la dissymétrie des valeurs et à la troncature à 0 pour les épisodes faibles, mais il est de toutes façons difficile de supposer le terme additif ε_{ij} de même variance $\forall P_i$.

On s'est donc tourné vers un modèle multiplicatif du type :

$$x_{ij} = P_i \times y_{ij}$$

et on a analysé les données "profilées". En effet y_{ij} traduit le pourcentage de l'épisode i recueilli à la station j. L'ensemble des valeurs $\{y_{ij}\}$ et $\{y_{i'j}\}$ sont directement comparables : 2 épisodes homothétiques mais de moyennes différentes auront même valeur de y_{ij} $\forall j$ donc même profil.

On vérifie que : $$\sum_{j=1}^{P} y_{ij} = 1 \quad \forall i$$

Par contre, les moyennes climatologiques $y_{\cdot j} = \eta_j$ et les écarts-types σ_j sont encore assez variables d'une station à l'autre et assez fortement corrélés. On peut donc appliquer aux y_{ij} un modèle de la forme :

$$y_{ij} = \eta_j + \sigma_j . \varepsilon_{ij}$$

avec $\eta_j = k . \sigma_j$ et l'analyse en corrélation donne des y_{ij} donne :

$$r_{jl} = E[\varepsilon_{ij} . \varepsilon_{il}]$$

car ici : $\begin{cases} \sum_j \varepsilon_{ij} = 0 \\ \sum_i \varepsilon_{ij} = 0 \end{cases}$

On a d'ailleurs vérifié que la variance spatiale des profils (de moyenne spatiale $\equiv 1$) n'est cette fois plus liée à la puissance de l'épisode P_i (cf Ière Partie fig.I-7-b).

Le tracé du corrélogramme donne encore une extinction aux environs de 30 à 50 km, et l'A.C.P. ne présente plus alors de valeur propre dominante : ($\lambda_1 \approx 21$, $\lambda_2 \approx 15$, --- etc). Les représentations dans F1/F2 sont proches des représentations dans F2/F3 de l'analyse sur données brutes.

En conclusion, nous n'avons pas réellement pratiqué une analyse de la variance au sens des tests d'hypothèse linéaire, mais nous avons montré comment des modèles linéaires simples permettaient de se ramener à ce que l'on croit être le phénomène "vrai"

(les ε_{ij}), cohérent avec la connaissance physique du phénomène pluie.

Certes on pourrait envisager des transformations encore plus élaborées pour rendre les épisodes comparables et proches d'un processus stationnaire, mais l'essentiel est d'éliminer d'abord l'effet de taille dû à la variabilité interépisode.

II.3 - Typologie des épisodes

II.3.1. Techniques de classification

a) On peut considérer que les analyses présentées dans la IIIème partie, chap. III-2) constituent une typologie des variables ou stations. De façon duale, on peut tenter d'effectuer une typologie des épisodes.

La raison concrète est la suivante : dans la partie prévisionnelle (Vème partie, Chap. IV) on verra que la caractéristique du champ que l'on utilise est le corrélogramme. Or il est à craindre que celui-ci varie selon certaines classes d'épisodes. Cela avait d'ailleurs été démontré par HENDRICK et COMER (1970) qui avaient comparé les corrélogrammes des pluies $\geqslant$ 2,5 mm, $\geqslant$ 12,5 mm et $\geqslant$ 25 mm (à une station au moins et en données journalières).

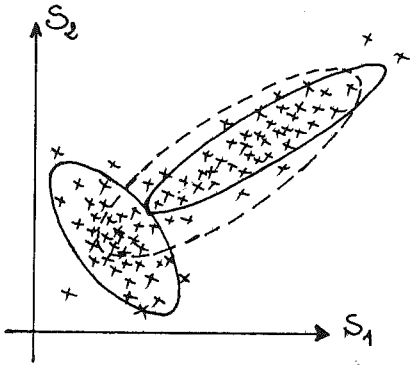
Nous verrons plus loin comment interpréter leurs résultats.

D'autre part, un problème se pose pour caractériser les épisodes. Au lieu de les représenter par les valeurs des 73 stations, on peut condenser l'information en ne prenant que leurs 5 ou 10 premières composantes principales. Mais alors on sait que pour nos données , la première composante n'est autre que la moyenne des 73 stations ou la puissance de l'épisode, or nous voulons plutôt comparer leur profil spatial.

On éliminera donc la 1ère C.P., ou on utilisera les données profilées.

b) Quant aux techniques d'aggrégation, on peut envisager soit des techniques de classifications hiérarchiques déjà utilisées sur les variables (IIIème Partie, Chap. II-2) soit des méthodes non hiérarchiques analogues à celles décrites en Vème Partie , Chap. II-1.

En fait, une variante de la méthode des nuées dynamiques, l'analyse factorielle typologique (DIDAY, 1973 ou DIDAY et SCHROEDER, 1974) semble particulièrement adaptée à notre cas. En effet, si l'on considère 2 stations seulement, on peut craindre que leur corrélation totale soit en fait une moyenne sur 2 groupes distincts. Et la connaissance éventuelle du groupe devrait être prise en compte dans l'interpolation d'un épisode. Généralisé à P stations, on chercherait donc si dans $\mathbb{R}^P$, il n'existe pas des sous-nuages denses, éventuellement de dimensions inférieures au nuage total (cf aussi IIIème partie, Chap. I-2-2).



La méthode en question cherche k variétés d'inertie minimum, ou k classes telles que leurs plans ou axes factoriels associés expliquent un pourcentage d'inertie plus important dans le sous-nuage que les axes de même rang dans l'analyse globale.

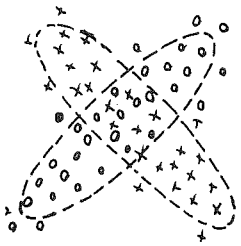
Un exemple déjà cité (description de la lettre A - cf DIDAY et Chap. I-2 de la IIIème partie) montre que le nuage total requiert 2 axes sur lesquels la variance se disperse (60%) , alors qu'avec 3 sous-nuages, leurs 1ers axes respectifs regroupent tous plus de 99% de la variance de leur classe.

Dans notre cas, le problème consistera à choisir le nombre d'axes et le nombre d'axes (ou la dimensionalité de ces groupes). On notera aussi que les méthodes de classification hiérarchique sont mieux adaptées à un ensemble fini (par exemple l'analyse du réseau de stations) qu'à un échantillon extrait d'une population infinie (notre ensemble d'épisodes par exemple).

III.3.2. Résultats

Comme toujours, ces résultats sont le fruit de recoupements entre méthodes et entre des analyses sur divers types de données. En résumé, on peut considérer que l'on a utilisé essentiellement les données profilées. La méthode hiérarchique Iphigénie a souvent été sujette à des effets de chaîne, mais l'analyse factorielle typologique ne peut se débarasser des effets de cylindre (si un tirage aléatoire introduit dans un noyau un point très éloigné, cela crée un axe d'allongement, ou d'inertie artificiel mais très fort et le groupe ne sera pratiquement plus remis en cause).

C'est ainsi qu'avec 2 groupes demandés, on obtenait pratiquement une configuration de type:

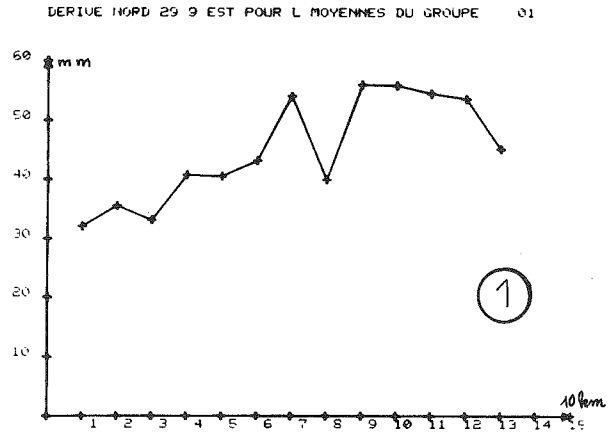


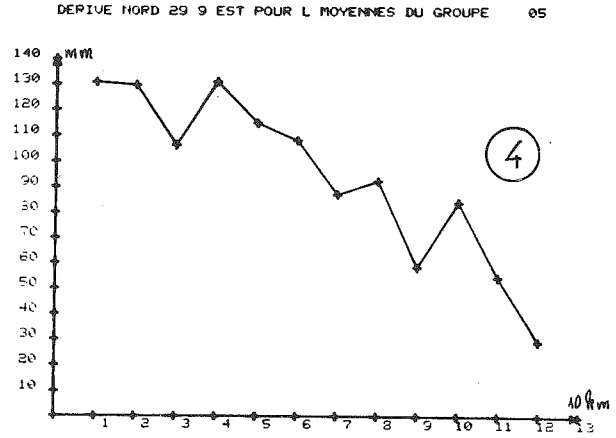
Finalement, on s'est arrêté sur 4 à 5 groupes - cf table.I

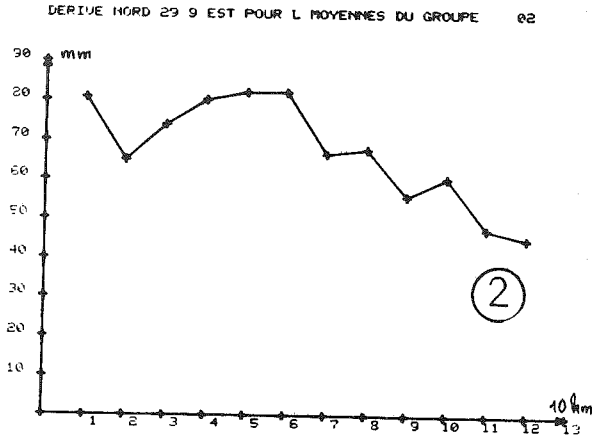
L'analyse de leur corrélogrammes respectifs met en évidence des différences (en particulier celui du groupe 1 reste très élevé) (fig.IV-5-a) et la distance d'annulation de la corrélation reste toujours élevée (100 à 200 km). On peut se demander alors quelle est la part de la fluctuation des moyennes. Si on retranche au sein de chaque groupe la 1ère C.P. du groupe (qui s'avère être encore approximativement la moyenne spatiale) on constate alors que les corrélogrammes se rapprochent sensiblement.

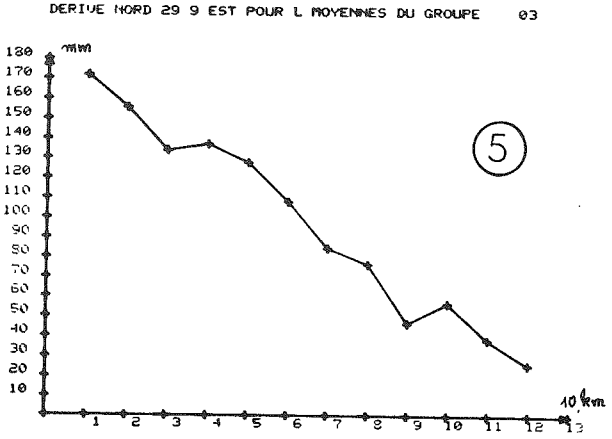
On constate que du groupe 1 au groupe 5 on a une évolution progressive depuis une forme exponentielle assez dispersée à des formes sinusoïdales amorties. Les distances d'annulation de la corrélation décroissent légèrement de 80 à 50 km.

Mais le plus frappant est le profil des isohyètes moyennes de chaque groupe. On constate que, du groupe 1 au groupe 5, on a des épisodes qui sont : (fig.IV-5-b)









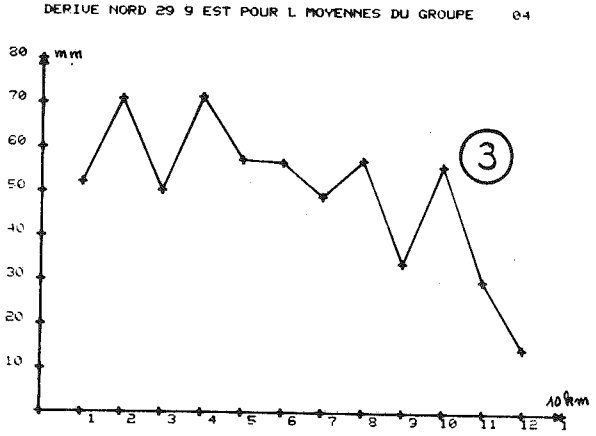
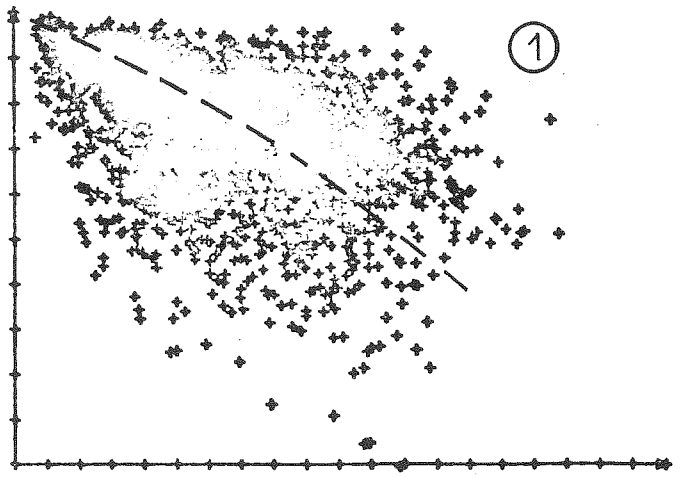


FIGURE IV - 5 : Typologie en 5 groupes .

a) Dérive moyenne selon un axe SW -NE

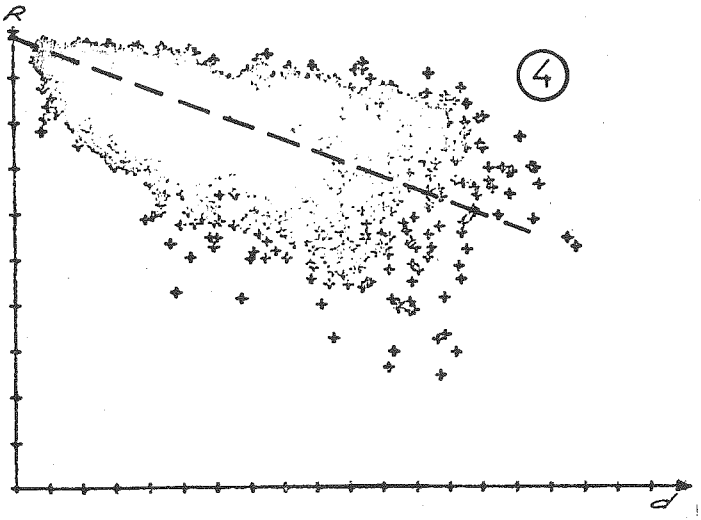
CORRELOGRAMME NORD- 29-SUD POUR 73 STATIONS

DISTANCE MAXIMALE 167.6 NOMBRE DE POINTS CALCULES1735



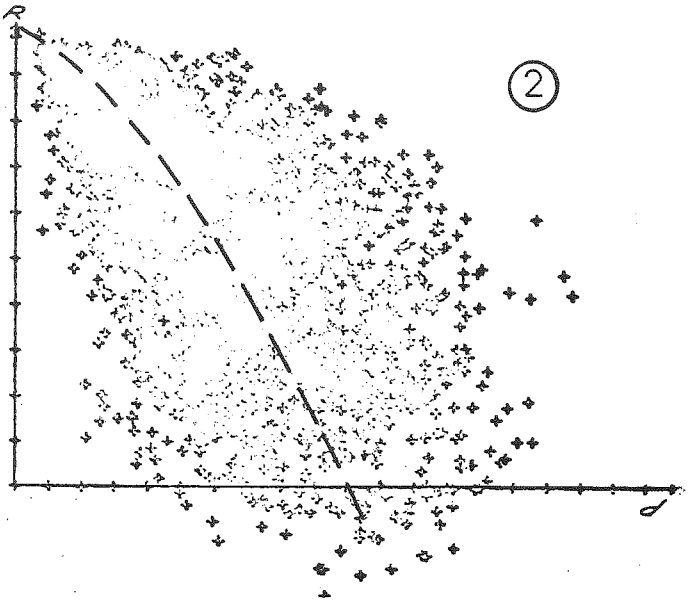
CORRELOGRAMME NORD- 29-SUD POUR 73 STATIONS

DISTANCE MAXIMALE 167.6 NOMBRE DE POINTS CALCULES1735



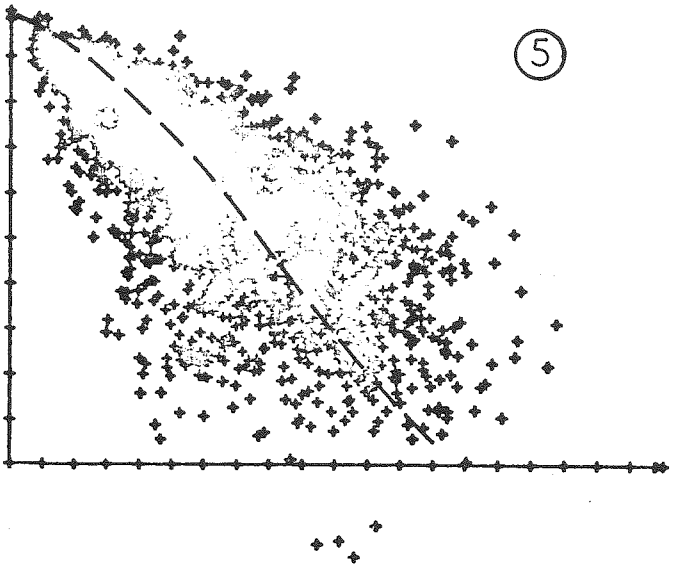
CORRELOGRAMME NORD- 29-SUD POUR 73 STATIONS

DISTANCE MAXIMALE 167.6 NOMBRE DE POINTS CALCULES1735



CORRELOGRAMME NORD- 29-SUD POUR 73 STATIONS

DISTANCE MAXIMALE 167 6 NOMBRE DE POINTS CALCULES1735



CORRELOGRAMME NORD- 29-SUD POUR 73 STATIONS

DISTANCE MAXIMALE 167 6 NOMBRE DE POINTS CALCULES1095
COUP A 65DEG

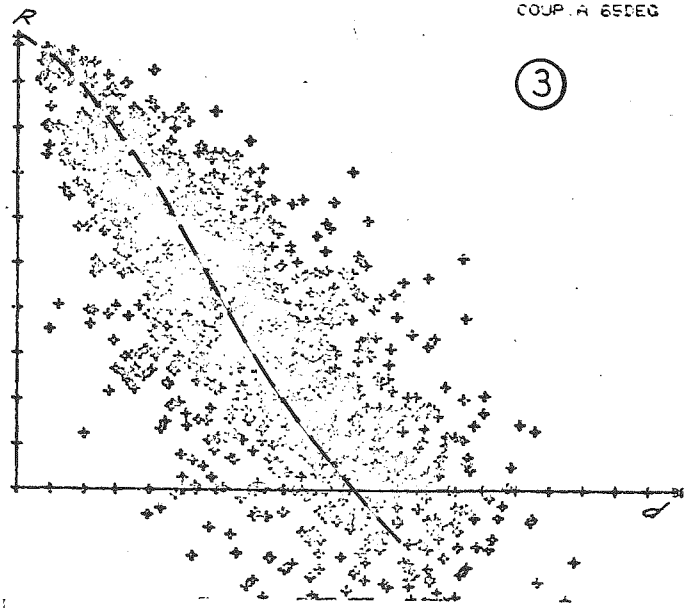


FIGURE IV - 5 : Typologie en 5 groupes .

b) Corrélogramme selon un axe SW - NE

Groupe ① : (ex.1) 12 épisodes

1/9/56	3/10/56	5/11/63	2/9/65	9/9/65
11/10/65	24/10/66	2/11/66	21/9/67	27/11/67
3/9/68	14/9/68			

Moyennes spatiales : 19 à 113 mm Moyenne : 46 mm

Groupe ② : (ex.2) 13 épisodes

25/9/56	9/11/57	30/9/61	26/9/62	19/9/63
9/11/66	11/9/68	25/10/68	1/11/68	2/10/73
10/9/75	1/10/75	28/10/76 <		

Moyennes spatiales 20 à 134 mm Moyenne : 68 mm

Groupe ③ : (ex.4) 17 épisodes

21/9/57	5/11/57	4/10/58	5/11/60	21/10/61
26/11/61	3/11/63	25/11/63	30/9/65	26/10/65
19/10/66	8/10/68	23/11/70	21/9/73	12/9/76
25/9/76	11/10/76			

Moyennes spatiales : 21 à 110 mm Moyenne : 51 mm

Groupe ④ : (ex.5) 22 épisodes

30/9/58	14/9/59	17/10/59	17/11/59	27/11/59
3/10/60	20/10/60	12/10/62	17/11/65	14/10/66
5/11/67	15/11/67	29/8/68	8/10/70	13/11/70
20/9/71	11/10/72	26/1P/72	3/11/73	14/11/74
23/8/76 <	9/11/76 <			

Moyennes spatiales : 24 à 226 mm Moyenne : 93 mm

Groupe ⑤ : (ex.3) 16 épisodes

28/10/60	5/10/61	12/11/61	6/11/62	31/10/63
4/9/64	1/10/64	24/9/65	16/10/65	29/9/66
5/11/66	14/9/69	19/10/69	22/11/69	17/9/74
24/10/76 <				

Moyennes spatiales : 33 à 170 mm Moyenne : 97 mm

TABLE I : Typologie des épisodes .

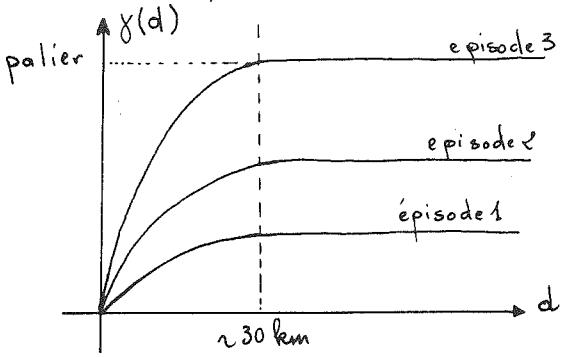
① centrés au Nord du réseau ② plutôt au centre et moyens ⑤ très au Sud et très forts
③ plutôt au Sud et moyens ④ franchement au Sud et forts

Il est probable que cette typologie est liée aux situations synoptiques correspondantes et à leurs interactions respectives avec le relief. Des recherches sont en cours sur ce sujet (cf aussi MARCH et al., 1979).

II.3.3. Aperçu sur d'autres possibilités de classifications

Les résultats ci-dessus montre que dans l'espace $\mathbb{R}^p$ les profils, c'est-à-dire l'allure générale des champs indépendamment de leurs niveaux, peuvent se regrouper selon leur pente moyenne S W - N E . On peut donc considérer que la classification s'est faite plutôt sur la dérive du champs que sur les corrélogrammes, même si ces derniers évoluent aussi progressivement d'un groupe à l'autre.

On pourrait par contre envisager une analyse purement spatiale du processus, en caractérisant chaque champ par son corrélogramme spatial, ou par son variogramme. Cela reviendrait à analyser le signal stationnaire que l'on trouve après extraction de la dérive et à chercher les épisodes lisses, ou bruités, à faible ou longue portée, etc... Malheureusement l'influence de la dérive sur l'estimation du corrélogramme nous conduit à employer des techniques sophistiquées qui sont développées dans J.P. DELHOMME (1977) et dont l'application à notre cas est détaillée dans D. CREUTIN (1979). Les résultats en sont d'ailleurs assez surprenants. Après extraction de la dérive, les variogrammes sont tous homothétiques et, si on les norme par la variance du signal stationnaire, ils deviennent quasiment identiques. (Cette norme revient en fait à "profiler" les données, mais après extraction de la dérive).



De plus, cette variance du signal stationnaire (ou "palier") bien que distincte de la variance spatiale du champ comme nous la calculions précédemment, est encore très liée au niveau moyen du champ.

$$\gamma(i,\infty) \# k \cdot P_i$$

Signalons aussi que, contrairement à ce que l'on attendait, il n'y a pas d'effets de pépite observable .

En conclusion, les propriétés statistiques du signal aléatoire stationnaire sont quasiment analogues pour tous les champs. Seules leurs dérives diffèrent et c'est bien sur celles-ci qu'il fallait faire une typologie.

Par contre, si on traite le champ comme stationnaire dans son ensemble (cas de la méthode de GANDIN, cf Vème Partie, Chap.IV-1), en mélangeant la dérive et le bruit, alors il n'y a plus de corrélogramme commun mais bien autant de corrélogrammes qu'il y a de types de dérive.

CHAPITRE III

MODELES DE PROCESSUS SPATIAUX (*)

III.1 - Compléments sur les processus stationnaires d'ordre 2 (*)

(a) On a déjà envisagé ce type de processus, au chapitre I, surtout du point de vue de la fonction d'autocorrélation. Rappelons que si on note $< , >$ la moyenne d'ensemble sur les réalisations ω, un processus stationnaire vérifie :

$$E[X(t)] = < X(t,\omega) > = m \quad \forall t$$

$$E\left[\left\{X(t,\omega) - E[X]\right\}^2\right] = \sigma_X^2 = \Gamma(0) \quad \forall t$$

Et surtout la fonction d'autocorrélation : $R(t,t') = < X(t,\omega) \cdot X(t',\omega) >$

se ramène, si $E[X] = 0$, à la fonction d'autocovariance : $\Gamma(t,t') = <(X(t) - E[X])(X(t') - E[X])>$

qui dans le cas stationnaire, vérifie :

$$\Gamma(t,t') = \Gamma(t-t') = \Gamma(\tau)$$

$$\Gamma(\tau) = \Gamma(-\tau) \qquad |\Gamma(\tau)| < \sigma_X^2$$

et $\forall \alpha_i$ et α_j, associés à t_i, t_j :

$$\sum_{i,j=1}^{N} \alpha_i \, \Gamma(t_i - t_j) \, \alpha_j = <\left(\sum_{i=1}^{N} \alpha_i \cdot X(t_i,\omega)\right)^2> \geqslant 0$$

donc la fonction est de type positif, ce qui limite déjà le choix des modèles pour $\Gamma(\tau)$. Si on veut de plus que la valeur soit un coefficient de corrélation au sens classique, il suffit de considérer

$$\rho(\tau) = \frac{\Gamma(\tau)}{\Gamma(0)}$$

(b) fonction de densité spectrale

C'est la transformée de Fourier de la fonction d'autocovariance :

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\lambda t} \cdot \Gamma(t) \, dt$$

Inversement :

$$\Gamma(t) = \int_{-\infty}^{+\infty} e^{+i\lambda t} \cdot f(\lambda) \, d\lambda$$

En fait,

$$\Gamma(\tau) = \int_{-\infty}^{+\infty} e^{i\lambda t} \cdot d\Psi(\lambda)$$

où $\psi(\lambda)$ est la fonction spectrale de $X(t)$, qui n'admet une densité $f(\lambda)$ que si ψ est continue.

On montre que :

- $f(\lambda)$ est positive (comme transformée de Fourier d'une fonction positive) et réelle (parce que $\Gamma(t)$ est paire)
- on montre aussi que $f(\lambda)$ est la densité d'énergie de $X(t)$ à la fréquence λ. Quand le processus a un spectre continu, cela suppose un passage à la limite: Si on filtre les fréquences extérieures à $\Delta = [\lambda, \lambda + d\lambda]$ alors le processus X_Δ filtré a une énergie : $E[X_\Delta(t)^2]$ égale à sa variance (on peut appliquer le théorème ergodique) qui s'écrit

$$\int_{\lambda}^{\lambda + d\lambda} f(\lambda)\, d\lambda$$

(cf ROZANOV, 1975, p.208).

(c) Un exemple qui nous sera utile ensuite est celui du processus :

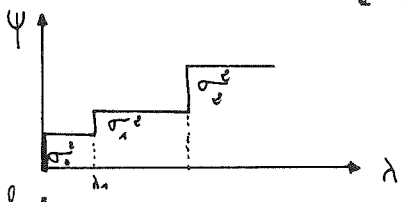
$$X(t) = \sum_{n=0}^{\infty} \left(A_n \cos \lambda_n t + B_n \sin \lambda_n t \right)$$

où les λ_i sont des nombres réels fixés et A_n et B_n des variables aléatoires indépendantes de même loi. Dans ce cas on montre que :

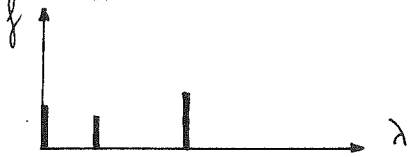
$$\Gamma(\tau) = \sum_{n=0}^{\infty} \sigma^2 \cos \lambda_n \tau$$

qui s'écrit aussi : $\Gamma(\tau) = \sum_{n=-\infty}^{+\infty} \frac{1}{2} \sigma_n^2 \, e^{i \lambda_n \cdot \tau}$ avec $\begin{cases} \sigma_{-n} = \sigma_n \\ \lambda_{-n} = \lambda_n \end{cases}$

Dans ce cas $\psi(\lambda)$ est de la forme :



et $f(\lambda)$:



On dit que l'on a un spectre de raies.

Naturellement beaucoup de processus continus ont des spectres continus, mais ces spectres de raies retrouvent de l'importance quand on échantillonne les processus.

(d) Dérivation et intégration de processus

Comme nous allons envisager des processus régis par des équations différentielles, il importe de définir d'abord la dérivée d'un processus et ses propriétés.

On définit d'abord la continuité du processus :

$$\lim_{t \to t_0} \left(X(t) - X(t_0) \right) = 0$$

mais cela dépend du mode de convergence choisi. On choisit en général la convergence au sens des moindres carrés (m.s.) qui s'écrit :

$$\lim_{t \to t_0} E\left[\left(X(t) - X(t_0) \right)^2 \right] = 0$$

et on montre qu'une condition nécessaire et suffisante est la continuité de $\Gamma(\tau)$ pour $\tau = 0$. (Par exemple un processus $X(t)$ où X vaut 0 ou 1 pendant les durées successives qui suivent une loi de Poisson est m.s. continue, cf BASS, 1962, p.127).

De même $X(t)$ sera m.s. différentiable si il existe $X'(t_0)$ tel que :

$$\lim_{t \to t_0} E\left[\left(\frac{X(t) - X(t_0)}{t - t_0} - X'(t_0) \right)^2 \right] = 0$$

Une condition nécessaire et suffisante pour qu'un processus du 2ème ordre soit m.s. différentiable est que $\Gamma(\tau)$ ait une dérivée première nulle et une dérivée seconde finie pour $\tau = 0$.

On démontre aussi (cf SOONG, p.98) que :

$$E\left[\frac{d^n X(t)}{dt^n} \right] = \frac{d^n E[X(t)]}{dt^n} \qquad E\left[X'(\delta) . X(t) \right] = \frac{\partial \Gamma(t,\delta)}{\partial t}$$

(e) Représentation du bruit blanc gaussien

On va être amené à considérer un signal gaussien continu, mais sans autocorrélation.

Comme cela est difficile à concevoir, on part du processus de mouvement brownien $B(t)$ de fonction d'autocorrélation $\Gamma_B(t,t') = 2D \min(t,t')$

et on calcule

$$\frac{\partial^2 \Gamma_B(t,t')}{\partial t . \partial t'} = E\left[B'(t) . B'(t') \right] = 2D . \frac{\min t,t'}{\partial t \, \partial t'}$$

$$= 2D \, \frac{\partial U(t-t')}{\partial t} = 2D . \delta(t-t')$$

où U est la fonction Heaviside et δ celle de Dirac.

On voit que l'on peut poser formellement :

$$\varepsilon(t) = \frac{d B(t)}{dt}$$

avec :

$$\Gamma_\varepsilon(t,t') = 2D . \delta(t-t') \quad \text{ou} \quad \rho_\varepsilon(t,t') = \frac{\delta(t-t')}{\delta(0)}$$

III.2 - Modèles de processus (*)

III.2.1. Modèles temporels ou unilatéraux (*)

(a) La théorie des processus s'est initialement développée à propos de séries chronologiques, dans lesquelles le temps joue un rôle particulier : le présent ne peut dépendre que du présent lui-même et du passé, mais pas de l'avenir.

Parmi ceux-ci, une classe particulièrement étudiée est celle des processus autoregressifs. Historiquement, on est parti de la relation discrète :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots \phi_m X_{t-m} + a_t$$

où a_t est une impulsion aléatoire de variance finie et surtout non autocorrelée

$$E[a_t \cdot a_{t+k}] = 0 \quad \forall k \neq 0$$

Cela exprimait que X_t dépendait des valeurs antérieures, plus d'un aléa instantané.

Analytiquement, on pouvait encore l'écrire sous forme d'une équation aux différences :

$$\Phi(B) X_t = a_t$$

où B est l'opérateur de retard : $B X_t = X_{t-1}$ et le passage à un processus continu, en supposant que les temps d'échantillonnage se rapprochent, ramène à une équation différentielle de la forme :

$$a_m \cdot \frac{d^m X}{dt^m} + a_{m-1} \frac{d^{m-1} X}{dt^{m-1}} + \cdots + a_1 \frac{dX}{dt} + a_0 X = \varepsilon(t)$$

Et le processus $X(t)$ s'interprète alors comme la réponse d'un système à un bruit blanc $\varepsilon(t)$ défini comme précédemment.

Ces modèles ont l'avantage de fournir assez simplement des expressions pour le processus lui-même, et ses fonctions d'autocorrélation et de densité spectrale : (cf par exemple : BOX et JENKINS, 1976 ou H.J. THIEBAUX, 1976).

Exemples

1er ordre :

$$\frac{1}{\nu} \cdot \frac{dX}{d\tau} + X(t) = \varepsilon(t) \longrightarrow X(t) = \int_0^{+\infty} \nu . e^{-\nu\tau} . \varepsilon(t-\tau) . d\tau$$

$$\rho(\tau) = e^{-\nu|\tau|} \qquad f(\lambda) = \frac{2\nu}{\nu^2 + \lambda^2}$$

2ème ordre :

$$a_1 \cdot \frac{d^2X}{dt^2} + a_2 \cdot \frac{dX}{d\tau} + X = \varepsilon(t) \longrightarrow X(t) = \int_0^{\infty} \frac{c^2 + a^2}{a} \cdot e^{-c\tau} \cdot \sin a\tau . \varepsilon(t-\tau) . d\tau$$

$$\rho(\tau) = e^{-c|\tau|}\left[\cos(a|\tau|) + \frac{c}{a}\sin(a|\tau|)\right] \qquad f(\lambda) = \frac{\cdots}{\left[c^2 + a^2 - \lambda\right]^2 + 4c^2\lambda^2}$$

(avec c et a tels que $a_1 = \frac{2c}{c^2+a^2}$ $a_2 = \frac{1}{c^2+a^2}$)

Remarquons que cela implique le calcul d'intégrale $\int f(t) . \varepsilon(t)\, dt$ où $\varepsilon(t)$ est un processus aléatoire. Or cette intégration est très délicate et ne peut se faire qu'au sens de ITO (cf SOONG, 1973 ou DELLEUR, 1977).

D'autre part, on a considéré l'équation différentielle comme limite d'une équation, ou d'un modèle, de différence stochastique. Or cette limite est différente (cf DELLEUR, 1977) selon que l'on prend la différence en avant ou en arrière. L'équation différentielle ne peut être considérée que comme la limite d'une équation aux différences en avant.

(b) La détermination de $\Gamma(\tau)$ peut se faire directement, en déterminant d'abord $X(t)$ puis en calculant $E[X(t).X(t')]$. Dans le cas d'un processus régi par une équation différentielle, on peut aussi déterminer $\Gamma(\tau)$ en remarquant qu'elle vérifie une équation différentielle un peu voisine.

Par exemple, si X satisfait :

(1) $$\frac{dX}{dt} + \alpha X(t) = \varepsilon(t)$$

On peut par exemple multiplier par $\varepsilon(t')$ et prendre l'espérance (cf PAPOULIS, 1965) d'où

$$< \frac{d\,(X(t).\varepsilon(t'))}{dt} > + \alpha < X(t).\varepsilon(t') > = < \varepsilon(t).\varepsilon(t') >$$

En permutant les 2 opérateurs différentiation et espérance :

(2) $$\frac{d\,R_{x\varepsilon}(t,t')}{dt} + \alpha\, R_{x\varepsilon}(t,t') = R_{\varepsilon\varepsilon}(t,t')$$

où le 2ème membre est connu, ce qui fournit $R_{x\varepsilon}(\tau)$.
Puis en reprenant l'équation mais en multipliant cette fois par $X(t')$, on obtient :

(3) $$\frac{d\,R_{xx}(t,t')}{dt} + \alpha\, R_{xx}(t,t') = R_{x\varepsilon}(t,t')$$

où le 2ème membre vient d'être calculé.

Dans le cas où $\varepsilon(t)$ est un bruit blanc :

$$R_{\varepsilon\varepsilon}(t,t') = \sigma_\varepsilon^2\, \delta(t-t')$$

$R_{x\varepsilon}$ est à un facteur près la fonction de GREEN de l'équation. Or on sait que la solution, pour un 2nd membre quelconque, s'obtient en convoluant la fonction de Green avec ce second membre. Et dans le cas de l'équation (3) qui nous intéresse, cela reviendra à la convoluer avec elle-même (On retrouve aussi cette approche dans SOONG, 1973, p.160 et suivantes).

III.2.2. Modèles spatiaux (*)

(a) Dans l'espace, une démarche similaire a fait passer des modèles discrets (équations aux différences ou équations de regression) aux expressions différentielles. Toutefois un problème souvent mal résolu est celui du rôle de la variable t , ou x , dans les modèles spatiaux. Contrairement au temps, la variable d'espace ne comporte plus de notion de passé et de futur.

Dans le cas d'une dimension par exemple, le véritable modèle spatial s'écrit :

(1) $$X_k = \beta \left(X_{k-1} + X_{k+1} \right) + \varepsilon_k$$

où ε_k est une variable indépendante. Dans ce cas, on peut modifier la relation en

$$X_{k+1} - 2 X_k + X_{k-1} - \frac{1 - 2\beta}{\beta} X_k = \varepsilon_k$$

et en faisant tendre le pas vers 0, on obtient :

(2) $$\frac{d^2 X}{dx^2} - \alpha^2 X(x) = \varepsilon(x)$$

donc la fonction spectrale est en $\frac{1}{\left(\lambda^2 + \alpha^2\right)^2}$ ce qui donne une fonction d'auto-corrélation assez compliquée .

M.S. BARTLETT (1975) l'appelle un modèle "simultané" ou bilatéral, en ce sens que X_k dépend simultanément de X_{k-1} et X_{k+1}.

Par contre, dans un modèle "conditionnel" la dépendance est unilatérale. Par exemple si on prend :

(3) $$X_k = \rho X_{k-1} + \varepsilon_k$$

on peut passer à l'expression :

(4) $$X_k = \frac{\rho}{1+\rho^2} \left(X_{k-1} + X_{k+1} \right) + \eta_k$$

mais alors les variables η_k ne sont plus indépendantes au sens des ε_k car :

$$\eta_k = \frac{\varepsilon_k - \rho \varepsilon_{k+1}}{1 + \rho^2}$$

et on ne peut pas passer à l'équation différentielle (2) mais seulement à celle déduite de (3) soit :

$$\frac{dX}{dx} - \alpha X(x) = \varepsilon(x)$$

de densité spectrale $\frac{1}{\lambda^2 + \alpha^2}$ et de fonction d'autocorrélation $e^{-\alpha |x|}$.

En fait l'expression (4) est conditionnelle au sens où :

$$E\left[X_k \mid X^*_{k-1}, X^*_{k+1} \right] = \frac{\rho}{1+\rho^2} \left(X^*_{k-1} + X^*_{k+1} \right)$$

où X_k est une variable aléatoire et X^*_{k-1}, X^*_{k+1} des valeurs observées.

C'est une régression qui peut servir à estimer le point central X_k si on ne connait que X^*_{k-1} et X^*_{k+1}. Mais le processus générateur est lui unilatéral car:

$$Pr[X_k \mid X_{k-1}, X_k] = Pr[X_k \mid X_{k-1}]$$

et si on veut générer le processus, on le fait effectivement de façon directionnelle et unilatérale.

BARTLETT propose l'analogie entre :

modèle → unilatéral → autoregressif → conditionnel
modèle → bilatéral → moyenne mobile → simultané

Il montre aussi que le processus simultané du 1er ordre (1) peut se ramener à un processus conditionnel du 2nd ordre :

$$X_k = 2\alpha X_{k-1} - \alpha X_{k-2} + \omega_k \qquad \text{avec } \beta = \frac{\alpha}{1+\alpha^2}$$

ce qui permet éventuellement d'en obtenir les propriétés.

(b) L'extension à 2 dimensions complique singulièrement les choses. Dans le cas des processus "simultanés" P. WHITTLE (1954 puis 1963) a proposé des solutions pour des équations aux dérivées partielles de la forme :

$$(\Delta - \alpha^2)^P X(x,y) = \varepsilon(x,y) \qquad \Delta = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$$

Dans le cas où $P = 1$, cela conduit à une fonction :

$$\rho(d) = \frac{d}{2\alpha} \cdot K_1(\alpha . d)$$

où K_1 est une fonction de Bessel modifiée de 2ème espèce. Elle diffère entre autres d'une exponentielle, qui serait solution du cas $P = 3/4$ (associé à un phénomène peu courant...).

On notera que ces résultats supposent évidemment l'homogénéité mais aussi l'isotropie du phénomène : les lignes d'isocorrélation sont des cercles dans (x,y). Un simple changement d'axe permet de s'y ramener quand elles ressemblent à des ellipses.

Enfin, V. HEINE (1955) a traité le cas général des processus simultanés régi par des équations aux dérivées partielles du 2nd ordre (qui ne sont pas forcément elliptiques).

Là où WITTLE utilisait la transformée de Fourier, HEINE utilise plutôt la transformation de Laplace. Sa démarche est analogue au cas de 1 dimension (III-2-1-b) et se résume ainsi :

(1) $$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) X(x,y) = \varepsilon(x,y)$$ (où L est un opérateur différentiel)

a une solution de la forme :

(2) $$X(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x-u, y-v) . \varepsilon(u,v)\, du . dv$$

où $G(x,y)$ est la fonction de Green du phénomène ou encore sa réponse impulsionnelle. Elle vérifie :

(3) $$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) G(x,y) = \delta(x) \cdot \delta(y)$$ et $G(x-u, y-v)$

est la réponse en x, y d'une impulsion unitaire en u, v.

Si $\varepsilon(x,y)$ est non correlé : $\Gamma_\varepsilon(x,y) = \sigma_\varepsilon^2 \cdot \delta(x) \cdot \delta(y)$

alors :

(4) $$\Gamma_x(x,y) = \sigma^2 \iint_{-\infty}^{+\infty} G(u,v) \cdot G(u-x, v-y) \cdot du \cdot dv$$

et $$\rho(x,y) = \rho(d) = \Gamma(x,y) / \Gamma(0,0)$$

Or si $g(p,q)$ est la transformée de Laplace de $G(x,y)$:

(3) $\Rightarrow$ $L(p,q)\, g(p,q) = p \cdot q$

et (4) $$\gamma(p,q) = \frac{g(p,q) \cdot g(-p,-q)}{p \cdot q}$$

d'où $\Gamma(x,y)$.

Dans le cas elliptique, HEINE retrouve le résultat de WHITTLE.

On peut imaginer des phénomènes se ramenant à ces modèles.

- Elliptique : déplacements verticaux d'une dalle supportant une charge $\varepsilon(x,y)$ aléatoire, ou répartition des températures stationnaires dans un milieu comportant des sources, stationnaires mais aléatoires $\varepsilon(x,y)$. On peut aussi imaginer la cote d'une nappe phréatique alimentée par des asperseurs aux débits constants mais aléatoires : $D \cdot \Delta H(x,y) = \frac{1}{S} \varepsilon(x,y)$
- Hyperbolique : Températures $T(x,y)$ dans un échangeur à flux croisés où le terme d'échange comporte un terme aléatoire (source ou puits) $\varepsilon(x,y)$.
- Parabolique : on peut imaginer d'introduire une dimension d'espace plus le temps, par exemple dans une nappe phréatique unidimensionnelle soumise à un terme d'évaporation ou de précipitation $\varepsilon(x,t)$:

$$\frac{\partial H(x,t)}{\partial t} - D \frac{\partial^2 H(x,t)}{\partial x^2} = \frac{1}{S} \varepsilon(x,t)$$

La corrélation $\rho(d,\tau)$ est alors assez compliquée (cf aussi B. SAGAR, 1978).

(c) Le cas des processus conditionnels est plus délicat et, bien que l'on s'intéresse à un phénomène spatial (ne comportant par exemple qu'une réalisation) il se traite en considérant la dimension temporelle.

P. WHITTLE (1962 et 1963) considère l'équation :

$$\frac{\partial X(s,t)}{\partial t} + \alpha\, X(s,t) - \frac{1}{c} \Delta X(s,t) = \varepsilon(s,t) \qquad s \to (x,y)$$

et regarde la solution stationnaire $X(s,\infty)$: en calculant la covariance spatiale $E[X(s+d,t).X(s,t)]$ il obtient une expression qui selon la dimension spatiale n fournit :

$$\Gamma(d) = \frac{e^{-d\sqrt{2\alpha}}}{e\sqrt{2\alpha}} \qquad \Gamma(d) = \frac{1}{2\pi} K_0(d\sqrt{2\alpha}) \qquad \Gamma(d) = \frac{e^{-d\sqrt{2\alpha}}}{4\pi d}$$

On remarquera que dans le cas d'une dimension spatiale cela ramène à $\rho(d) = e^{-d\sqrt{2\alpha}}$

J. BESAG (1972, cité par BARTLETT) interprète ceci comme la limite du modèle :

$$X_{k,l} = \beta(X_{k-1,l-1} + X_{k+1,l+1}) + \varepsilon_{k,l}$$

où $X_{k,l}$ dépend des valeurs de part et d'autre au temps $l-1$. On remarquera aussi que le résultat est cohérent avec l'approche directe du processus conditionnel associé (dans 1 dimension)

$$\frac{dX(x)}{dx} - \sqrt{2\alpha}\, X(x) = \varepsilon(x) \longrightarrow \rho(d) = e^{-d\sqrt{2\alpha}}$$

(d) En conclusion, notre but n'était pas de proposer de nouveaux modèles de processus, ni même d'appliquer ceux qui ont été étudiés. Mais les méthodes que nous utiliserons en Vème partie pour l'interpolation spatiale (ou dans CREUTIN, 1979) partent d'un modèle pour la fonction d'autocorrélation. On choisit en général une expression simple, qui, on l'a vu, provient en fait d'un modèle plus ou moins réaliste de processus. Par exemple l'exponentielle, dans le cas de 2 dimensions, ne peut pas se rattacher à des modèles différentiels "simples", or il vaut mieux savoir à quel modèle physique se rattache la loi théorique que l'on a ajustée.

CHAPITRE IV

ANALYSE HARMONIQUE DES PROCESSUS

Ce terme est relativement ambigu et sa définition varie selon les auteurs. La définition classique (cf par exemple, J. BASS, 1962) consiste à chercher le spectre d'une fonction "presque périodique" (somme de termes sinusoïdaux dont les périodes n'ont pas de sous-multiple commun). Dans ce cas les statisticiens parlent plutôt d'"analyse spectrale" avec tous les problèmes d'estimations que cela comporte. D'autres, par contre (DEVILLE, 1974) utilisent le terme d'analyse harmonique pour l'analyse en composantes principales d'échantillons prélevés dans un processus continu.

En fait, l'incertitude de la définition provient de ce que ces techniques sont toutes très voisines.

Dans ce chapitre, on cherchera surtout à voir comment peut s'interpréter l'A.C.P. classique dans un contexte de processus.

IV.1 - Processus continus

IV.1.1. Décomposition en série de Fourier

(a) Processus périodiques

D'après A. PAPOULIS (1965, p.367), un processus $X(t)$ est dit périodique au sens des moindres carrés si sa fonction d'autocorrélation (ou d'autocovariance) est périodique :

$$\exists T \qquad R(\tau+T) = R(\tau)$$

Dans ce cas, on peut la décomposer en série de Fourier :

$$R(\tau) = \sum_{n=-\infty}^{+\infty} \alpha_n e^{j n \omega_0 \tau} \quad \text{avec} \quad \omega_0 = \frac{2\pi}{T} \qquad \alpha_n = \frac{1}{T} \int_{-\frac{I}{2}}^{+\frac{I}{2}} R(\tau).e^{-jn\omega_0\tau} d\tau$$

Le spectre de $X(t)$ est alors une suite de raies :

$$S(\omega) = \sum_{n=-\infty}^{+\infty} \alpha_n . \delta[\omega - n.\omega_0]$$

On notera aussi que comme $R(\tau)$ est paire :

$$\alpha_n = \frac{1}{T} \int_{-\frac{I}{2}}^{+\frac{I}{2}} R(\tau).\cos n\omega_0\tau . d\tau + j \frac{1}{T} \underbrace{\int_{-\frac{I}{2}}^{+\frac{I}{2}} R(\tau) \sin n\omega_0\tau d\tau}_{\text{impaire}} \Longrightarrow \alpha_n \text{ réel}$$

D'où : $$R(\tau) = \sum_{n=0}^{+\infty} \varepsilon \, \alpha_n . \cos n\omega_0\tau$$

Dans le cas où le processus est périodique au sens des moindres carrés, $X(t)$ peut aussi être développé en série de Fourier :

$$X(t) = \sum_{n=-\infty}^{+\infty} \gamma_n \, e^{j n \omega_0 t} \qquad \text{sur } t \in]-\infty, +\infty[$$

et on a les relations suivantes:

$$E[\gamma_n \cdot \gamma_m^*] = 0 \qquad E[\gamma_n \cdot \gamma_n^*] = E[|\gamma_n|^2] = \alpha_n$$

et $$E[\gamma_0] = E[X] \qquad \gamma^* = \gamma \text{ conjugué}$$

Si le processus $X(t)$ est réel, alors la sommation ne comprend que des termes conjugués 2 à 2 :

$$\gamma_n \, e^{j n \omega_0 t} + \gamma^* e^{-j n \omega_0 t} \qquad \text{avec} \quad \gamma_n = c_n + j d_n$$

Dans ce cas le processus peut s'écrire :

$$X(t) = \sum_{n=0}^{+\infty} \left(a_n \cos n\omega_0 t + b_n \sin n\omega_0 t \right) \text{ avec } \begin{cases} a_n = 2 c_n \\ b_n = 2 d_n \end{cases}$$

et $$\gamma_n^2 = \frac{a_n^2 + b_n^2}{4}$$

$$\Longrightarrow E[\gamma_n^2] = \frac{E[a_n^2] + E[b_n^2]}{4} = \frac{2\sigma_n^2}{4} = \alpha_n$$

d'où le résultat remarquable :

$$R(\tau) = \sum_{n=0}^{+\infty} \sigma_n^2 \cdot \cos n\omega_0 \tau$$

(b) Processus non périodiques

Dans ce cas le développement du processus en :

$$X(t) = \sum_{n=-\infty}^{+\infty} \gamma_n \, e^{j n \omega_0 t}$$

est toujours possible, mais seulement dans un intervalle $\left[-\frac{T}{2}, +\frac{T}{2}\right]$ et on montre (PAPOULIS, 1965, p.455) que dans ce cas :

$$E[\gamma_n \cdot \gamma_m^*] \neq 0$$

c'est-à-dire qu'ils ne sont plus décorrelés au sens de la moyenne d'ensemble.

Cela provient de ce que l'on a remplacé la vraie fonction d'autocorrélation par son prolongement par périodicité en dehors de l'intervalle $\left[-\frac{T}{2}, +\frac{T}{2}\right]$.

(Toutefois, PAPOULIS montre que, quand T augmente, cette corrélation tend rapidement vers 0.)

IV.1.2. Développement de KARHUNEN-LOEVE

(a) On cherche alors un développement sur des fonctions autres que les fonctions trigonométriques ($\sin n\omega_0 t$ et $\cos n\omega_0 t$) tel que les coefficients de ce développement soient orthogonaux.

Il faut donc déterminer cet ensemble de fonctions $\varphi_1(t), \varphi_2(t) \cdots \varphi_n(t) \cdots$ qui conservent la propriété d'orthogonalité sur l'intervalle T

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} \varphi_n(t) \cdot \varphi_m^*(t)\, dt = \delta_{nm}$$

avec $$X(t) = \sum_{n=-\infty}^{+\infty} b_n \cdot \varphi_n(t) \quad \text{et} \quad b_n = \int_{-\frac{T}{2}}^{+\frac{T}{2}} X(t) \cdot \varphi_n^*(t) \cdot dt$$

mais vérifiant en plus : $E[b_n \cdot b_m^*] = \delta_{nm} \cdot \lambda_n$

On démontre simplement, en écrivant cette dernière condition que :

$$(1)\ \forall t_1,\ X(t_1) \times b_n^* = \sum_{m=-\infty}^{+\infty} b_m \cdot b_n^*\ \varphi_m(t_1)$$

$$\Longrightarrow E[X(t_1) \times b_n^*] = \sum_m E[\overbrace{b_m \cdot b_n^*}^{\delta_{mn}}] \varphi_m(t_1) = E[|b_n|^2]\, \varphi_n(t_1)$$

$$(2)\quad X(t_1) \times b_n^* = X(t_1) \times \int_{-\frac{T}{2}}^{+\frac{T}{2}} X(\ ) \cdot \varphi_n(\ ) \cdot d = \int_{-\frac{T}{2}}^{+\frac{T}{2}} X(\tau_1) \cdot X(\tau) \times \varphi_n(\tau) \cdot d\tau$$

et en prenant l'espérance :

$$\Longrightarrow E[X(t_1) \times b_n^*] = \int E[X(\tau_1) \cdot X(\tau)] \times \varphi_n(\tau)\, d\tau = \int R(t,\tau) \varphi_n(\tau) d\tau$$

et ce $\forall t_1$, donc φ_n est solution de l'équation intégrale :

$$(1)+(2) \Longrightarrow \int_{-T/2}^{+T/2} R(t,\tau) \cdot \varphi_n(\tau) \cdot d\tau = \lambda_n \cdot \varphi_n(t)$$

C'est une équation intégrale linéaire homogène de FREDHOLM de seconde espèce (KRASNOV et al, 1977).

λ_n est une valeur caractéristique ou valeur propre du noyau R et φ_n est la fonction propre associée.

KRASNOV et al (1977, p.73) montrent que si $R(t,t')$ est fonction de la seule différence des arguments $t - t' = \tau$ (processus stationnaire) et si $R(\tau)$ est prolongé par périodicité en dehors de l'intervalle $[-\frac{T}{2}, +\frac{T}{2}]$ ($\Longrightarrow$ processus rendu périodique) alors les fonctions propres sont :

$$\varphi_n^{(1)}(t) = \sin n\omega_0 t \qquad \varphi_n^{(2)}(t) = \cos n\omega_0 t$$

associées à :

$$\lambda_n = \frac{1}{2\pi} \int_{-\frac{T}{2}}^{+\frac{T}{2}} R(\tau) \cdot \cos n\omega_0 \tau \cdot d\tau \qquad \omega_0 = \frac{2\pi}{T}$$

mais cela suppose que $R(\tau)$ est prolongée par périodicité en dehors de l'intervalle $\left[-\frac{T}{2}, +\frac{T}{2}\right]$ et donc qu'elle est périodique au sens du § IV.1.1.

Un problème consiste à envisager ce que peuvent être les fonctions propres $\varphi(t)$ dans le cas d'un intervalle T infini et où $R(\tau)$ n'est pas périodique.

On remarquera dans ce cas que, contrairement aux développements de Fourier, les fonctions $\varphi(t)$ n'ont pas a priori d'expression analytique simple. Par contre, on peut les obtenir point par point en résolvant numériquement l'équation intégrale. C'est ce que nous ferons de façon plus détaillée dans la Vème Partie (Chap.IV).

Note : PAPOULIS (1965, p.461) montre aussi que pour un processus non périodique, on peut construire un développement

$$X(t) = \sum_{n=-\infty}^{+\infty} c_n \, e^{jn\omega_0 t}$$

qui est valable non plus entre $-T/2$ et $+T/2$ mais $\forall t$ à condition de définir c_n par:

$$c_n = \int_{-\infty}^{+\infty} X(t) \, \frac{\sin \frac{\omega_0 t}{2}}{\pi t} \, e^{-jn\omega_0 t} \, dt \qquad \left(\text{au lieu de } \int_{-\frac{T}{2}}^{+\frac{T}{2}} X(t) \, e^{-jn\omega_0 t} \, dt\right)$$

où cette fois ω_0 est arbitraire.

(b) On peut citer un certain nombre de résultats sur les fonctions propres.

. Le filtre passe bas :

$$S(\omega) = \begin{cases} S_0 & \text{si } \omega < \omega_c \\ 0 & \text{si } \omega > \omega_c \end{cases} \longrightarrow R(\tau) = \frac{\sin \omega_c \tau}{\pi \tau}$$

Pour un intervalle T quelconque, les solutions $\varphi_n(t, c)$ sont des fonctions sphériques dépendant de $c = \omega_c . T$ (cf PAPOULIS, 1965).

. Le processus de WIENER-LEVY (mouvement Brownien) de fonction d'autocorrélation (cf PAPOULIS, 1965 ou OBUKHOV, 1960) : $R(t,t') = \min(\alpha t, \alpha t')$

Bien que non stationnaire $\left(\sigma_W^2(t) = \alpha . t\right)$ on peut le développer selon KARHUNEN - LOEVE à l'aide des fonctions propres vérifiant :

$$\int_0^T R(t,t') . \varphi(t') dt' = \lambda \, \varphi(t)$$

$$\Longrightarrow \alpha \int_0^t t' . \varphi(t') dt' + \alpha t \int_t^T \varphi(t') dt' = \lambda \varphi(t)$$

En dérivant 2 fois par rapport à t on obtient :

$$\lambda \varphi''(t) + \alpha \varphi(t) = 0 \qquad \text{avec} \quad \varphi(0) = 0 \quad \varphi'(T) = 0$$

d'où

$$\varphi_n(t) = \sqrt{2} \sin \sqrt{\frac{\alpha}{\lambda_n}} \, t \qquad \lambda_n = \frac{\alpha T^2}{\pi^2 (n + \frac{1}{2})^2}$$

et on voit que les solutions, pour T fini sont quand même périodiques. OBUKHOV donne encore un autre cas particulier et on trouvera aussi des éléments dans B. LEVINE(1973).

On dispose cependant d'un résultat plus général pour toute une classe de processus.

IV.1.3. Recherche des fonctions propres dans le cas de spectres rationnels (Travaux de M.I. FORTUS)

M.I. FORTUS a cherché les solutions de l'équation

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} R(t-\tau).\varphi^T(\tau).d\tau = \nu^T.\varphi^T(t)$$

et les propriétés de ν^T et $\varphi^T(t)$ quand $T \nearrow \infty$.

(a) En remarquant d'abord que la densité spectrale du processus $X(t)$ s'écrit :

$$f(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i\lambda(\tau-t)}.R(t-\tau).dt$$

avec $R(t)$ paire d'où encore :

$$\int_{-\infty}^{+\infty} e^{-i\lambda t}.R(t-\tau).dt = 2\pi.f(\lambda).e^{-i\lambda\tau}$$

on vérifie à nouveau que les solutions pour $T\to\infty$ sont les fonctions trigonométriques $\cos\lambda t$ et $\sin\lambda t$ associées à la valeur propre $\nu^\infty(\lambda)$
La question étant de savoir ce qu'elles deviennent pour T fini.

Par contre la parité de $R(\tau)$ implique que les fonctions φ^T soient paires ou impaires (symétriques en valeur absolue). On montre aussi que toutes les valeurs propres sont doubles sauf la 1ère ($\sin 0.t$ n'est pas admissible comme vecteur propre) cf KRASNOV et al (1977, p.73).

(b) M. FORTUS montre d'abord que la 1ère valeur propre ν_1^T est toujours inférieure à $2\pi f(\lambda_0)$ (si $\lambda_0 \longrightarrow$ maximum de $f(\lambda)$) mais qu'elle tend vers cette valeur si $T \longrightarrow \infty$

(c) Ensuite il montre que si la densité spectrale $f(\lambda)$ est une fraction rationnelle :

$$f(\lambda) = \frac{Q_m(\lambda).Q_m^*(\lambda)}{P_n(\lambda).P_n^*(\lambda)}$$

P et Q étant des polynômes de degré n et m ($n>m$), le problème peut être résolu analytiquement.

Dans ce cas en effet, les fonctions $\varphi^T(t)$ sont de la forme :

$$\varphi^T(t) = \sum_{s=1}^{n} C_s\left(e^{k_s t} + e^{-k_s t}\right) \text{ ou } \sum C_s\left(e^{k_s t} - e^{-k_s t}\right)$$

Le calcul des ν^T, $k_s(\nu^T)$ et $C_s(\nu^T)$ est extrêmement laborieux et a été décrit par YAGLOM (1955). Il conduit à des équations transcendantes qu'il faut résoudre numériquement.

d) Toutefois, les calculs effectués pour les fonctions de corrélation les plus courantes font apparaître un certain nombre de points communs.

La première :

1) $R(\tau) = Ce^{-\alpha|\tau|} \longrightarrow f(\lambda) = \frac{C\alpha}{\pi(\alpha^2+\lambda^2)}$

a pour valeurs propres, sur l'intervalle $\left[-\frac{T}{2} + \frac{T}{2}\right]$

$\nu^T = \frac{2C\alpha}{\alpha^2+\lambda^2}$ et on montre qu'avec $k_n = \pm i\lambda_n$ solution de $\operatorname{tg}\lambda_n T = \frac{2\alpha\lambda_n}{\lambda_n^2-\alpha^2}$

Les fonctions propres deviennent :

$$\varphi_n^T(t) = \sqrt{\frac{2}{T+\nu_n^T}} \cdot \sin\left(\lambda_n t + n\frac{\pi}{2}\right)$$

c'est-à-dire une suite de <u>sinus</u> et de <u>cosinus</u> (la première, associée à λ_1 est un cosinus).

Pour ce cas particulier, on trouvera une autre approche par DAVENPORT et ROOT (1958 cité par PAPOULIS, p.472) qui aboutit au même résultat .

Pour les autres fonctions considérées par M. FORTUS :

2) $C.e^{-\alpha|\tau|}.(1+\alpha|\tau|)$ 3) $C.e^{-\alpha|\tau|}\cos\beta\tau$

4) $C.e^{-\alpha|\tau|}.\cos\alpha\tau$ 5) $C.e^{-\alpha|\tau|}\cos 2\alpha\tau$

les racines k_s ne sont plus imaginaires pures et les solutions sont des combinaisons linéaires de fonctions trigonométriques <u>amorties</u>.

Par contre, il montre comment la forme des fonctions propres, par exemple la 1ère $\varphi_1^T(t)$ est liée à T :

- si le spectre $f(\lambda)$ a son maximum en 0, alors elle est toujours paire, (et alternativement pour les autres) et la forme est stable $\forall T$ (cas des fonctions 1) et 2)). En particulier, la fonction $\varphi_1^T(t)$ a exactement j-1 zéros (recoupe j-1 fois l'axe des t) sur $\left[-\frac{T}{2}, +\frac{T}{2}\right] \forall T$

- par contre si $f(\lambda)$ a son maximum en $\lambda_0 \neq 0$ avec $\nu_1^\infty = 2\pi f(\lambda_0)$ on peut, selon la valeur de T, approcher ν_1^∞ par $\nu_1^T = 2\pi f(\lambda_1^T)$ avec $\lambda_1^T > \lambda_0$ ou $< \lambda_0$. Dans le second cas φ_1^T est impaire, mais cela suppose pour les fonctions d'autocorrélation considérées, $\alpha T > 1$.

Remarque : Pour cela, on montre que $\sum_{j=1}^{\infty} \nu_j^T = \int_{-\frac{T}{2}}^{+\frac{T}{2}} R(0).d\tau = T$

D'autre part, si $\lambda_0 \neq 0$ on a :

$\nu_1^\infty = 2\pi f(\lambda_0) > 2\pi f(0) = 1 \Longrightarrow$ le spectre est > 1 entre 0 et λ_0.

Or on approche ν_1^∞ par $\nu_1^T < \sum_j \nu_j^T = 1$ donc par la droite . Il faut donc au moins $\alpha T > 1$.

Cela explique la forte analogie des premières fonctions propres obtenues, quelle que soit l'allure de la fonction de corrélation dès que l'intervalle observé est inférieur à la longueur de corrélation (la première est toujours paire).

Cela n'est théoriquement prouvé que pour les fonctions d'autocorrélation à spectre rationnel, mais la plupart des fonctions observées peuvent se représenter ainsi.

L'extension à des champs bidimensionnels homogènes et isotropes a été proposée par M.I. FORTUS (1975) qui a mis en évidence le même type de résultats. Comme, de plus, la plupart des phénomènes météorologiques bruts ont un spectre qui tend à être maximal vers 0 (les basses fréquences ont plus d'énergie que les hautes) il est normal que l'on retrouve toujours le même type de forme pour les fonctions propres .

e) Enfin quand $T \longrightarrow \infty$, les fonctions propres convergent rapidement vers les fonctions trigonométriques quelque soit la fonction d'autocorrélation (ou la densité spectrale). En particulier si le maximum correspond à $\lambda_0 = 0$, la première fonction $e^{-i\lambda_0 t}$ est bien entendu une constante. (et se traduit par un effet de taille)

Remarque : On a parlé d'intervalle de corrélation, ou de longueur de corrélation que l'on peut définir par (MONIN et YAGLOM, 1975, p.35)

$$L = \frac{1}{R(0)} \int_0^\infty R(t)\, dt$$

qui donne un ordre de grandeur de la distance où la corrélation cesse d'être significative.

IV.2 - Cas des données discrètes - Processus échantillonné

IV.2.1. Analogie des formulations entre le cas discret et le cas continu

(a) Si on considère une fonction $f(t)$ et sa transformée de Fourier $F(\lambda)$ on a la relation

$$F(\lambda) = \int_{-\infty}^{+\infty} f(t).e^{-j\lambda t}.dt$$

On peut imaginer que l'on remplace l'intégrale par la somme :

$$\sum_{k=0}^{N-1} f(k.\Delta t)\, e^{-j\lambda k\Delta t}$$

ce qui conduit à la notion de Transformée de Fourier Discrète (cf RADIX J.C., 1970, p.141). A la suite $X_k = f(k.\Delta t)$ correspond la suite A_r :

$$X_k = f(k.\Delta t) \quad k = 0,\ldots N-1 \longrightarrow A_r = \sum_{k=0}^{N-1} X_k . e^{-\frac{2\pi r j k}{N}}$$

La formule inverse donne :

$$X_l = \frac{1}{N}\sum_{r=0}^{N-1} A_r\, e^{\frac{2\pi r j l}{N}}$$

La suite A_r constitue une analyse spectrale de la suite X_k . Elle ne représente que dans une certaine mesure celle de $X(t)$ car il y a eu 2 opérations où l'on a perdu de l'information :

- l'échantillonnage à la fréquence $\frac{1}{\Delta t}$
- l'utilisation d'un intervalle de temps fini $(0, N.\Delta t)$

Cela a pour effet d'éliminer les hautes fréquences (d'où repliement du spectre) et de prolonger implicitement le processus par périodicité (ce qui rend le spectre périodique).

(b) Pour rejoindre les notations de D.R. BRILLINGER, dont nous allons utiliser les résultats, nous écrirons plutôt, pour un processus X :

$$\left.\begin{array}{l} \frac{2\pi}{N} \longrightarrow \lambda \\ k.\Delta t \longrightarrow t \end{array}\right\} A_r \longrightarrow dF(\lambda) = \sum_{t=0}^{T-1} X(t)\, e^{-j.\lambda t}$$

et inversement :

$$X(t) = \frac{1}{T}\sum_{s=0}^{T-1} e^{j\frac{2\pi s t}{T}} . dF\left(\frac{2\pi s}{T}\right)$$

(On remplace N par T par analogie avec le cas continu mais T est ici le nombre de pas Δt).

En fait, pour se rapprocher encore du cas continu on peut supposer que le processus est connu entre $-T$ et $+T$, sur $2T+1$ points autour de 0. Dans ce cas :

$$dF(\lambda) = \sum_{t=-T}^{+T} X(t).e^{-j\lambda t}$$

La fonction d'autocorrélation du processus est, elle remplacée par une matrice définie posivite notée R ou Σ (selon que les valeurs sont normées ou non à 1) :

$$\begin{bmatrix} C(0) & C(1) & \cdots & C(2T) \\ C(-1) & C(0) & \cdots & C(2T-1) \\ \vdots & & & \vdots \\ C(-2T) & C(-2T+1) & \cdots & C(0) \end{bmatrix}$$

qui présente la particularité d'être de la forme TOEPLITZ (BRILLINGER, p.108).

IV.2.2. Recherche des éléments propres

(a) Dans le cas où T est grand, on donne des résultats théoriques approchés (qui sont détaillés dans BRILLINGER). La matrice Z :

$$Z = [z_{ij}] = [C(i-j) + C(i-j+2T)]$$

approche C quand $T \longrightarrow \infty$. Or Z est une matrice circulante et ses valeurs propres sont connues :

$$\mu_s = \sum_{t=-2T}^{+2T} C(t)\, e^{j\frac{2\pi s t}{2T+1}} \qquad s = -T \ldots 0 \ldots T$$

avec les vecteurs propres associés :

$$V_s = [v_{sl}] \qquad v_{sl} = \frac{1}{\sqrt{2T+1}}\, e^{-j\frac{2\pi s l}{2T+1}} \qquad l = -T \ldots 0 \ldots +T$$

Ces résultats sont donnés pour des matrices C hermitiennes mais ici, C est réelle définie positive, et ses valeurs propres doivent être réelles, de même que ses vecteurs propres.

Or si on regarde la sommation dans μ_s, on s'aperçoit que l'on a des termes :

$$C(-t)\, e^{-j\frac{2\pi s t}{2T+1}} + C(t)\, e^{j\frac{2\pi s t}{2T+1}} = 2.C(t).\cos\frac{2\pi s t}{2T+1}$$ C étant paire

donc μ_s est réelle.

$$\mu_s = C(0) + \sum_{t=1}^{2T} 2\, C(t).\cos\frac{2\pi s t}{2T+1}$$

De plus, on vérifie que $\mu_{-s} = \mu_s$ et il n'y a pas en fait $2T+1$ valeurs propres distinctes, mais 1 valeur simple μ_0 et T valeurs doubles.

Les 2 vecteurs propres complexes associés sont V_s et V_{-s}, mais toute combinaison linéaire de ces 2 vecteurs propres est vecteur propre. Si on prend les 2 combinaisons orthogonales :

$$V_s + V_{-s} \qquad \text{et} \qquad V_s - V_{-s}$$

les ℓ-èmes coordonnées deviennent, à un facteur près :

$$\cos \frac{2\pi s \ell}{2T+1} \qquad \sin \frac{2\pi s \ell}{2T+1}$$

$\Longrightarrow$ L'analogie avec le cas continu traité par FORTUS est parfaite dans le cas où $T \to \infty$: On trouve bien les fonctions trigonométriques comme vecteurs propres et la valeur propre associée est bien la densite spectrale du processus, ou la valeur de la transformée de Fourier de la fonction d'autocorrélation (dans le cas continu, $2\pi f(\lambda)$ comme dans le cas discret : $\sum_{t=-2T}^{+2T} C(t) e^{-j \frac{2\pi s t}{2T+1}} = 2\pi f\left(\frac{2\pi s}{2T+1}\right)$

(b) Dans le cas où T est petit, il existe très peu de résultats. La matrice Z diffère alors sensiblement de C ce qui exclut d'utiliser cet artifice. Les seuls résultats disponibles à notre connaissance sont dûs à KUENY J.L. (1977) dans le cas d'une fonction d'autocorrélation connue sur $[0, (p-1).\Delta t]$ et qui peut se mettre sous la forme :

$$\rho_\ell = \sum_{k=0}^{p-1} a_k \cos \frac{\pi k \ell}{p-1} = \sum_{k=-p+1}^{p-1} a_k . e^{j \frac{2\pi k \ell}{p-1}}$$

En cherchant des vecteurs propres V_s de la forme :

$$\sum_{\ell=0}^{p-1} b_\ell . \cos\left[\left(s - \frac{p+1}{2}\right).\frac{\pi \ell}{p-1}\right] \quad \text{ou} \quad \sum_{\ell=0}^{p-1} c_\ell . \sin\left[\left(s - \frac{p+1}{2}\right).\frac{\pi \ell}{p-1}\right]$$

et, en procédant par identification, KUENY a montré qu'il existait une solution exacte pour determiner les b_ℓ et c_ℓ.

Ceci est évidemment à rapprocher de la forme analytique :

$$\sum c_{\ell s}\left(e^{k_s t} + e^{-k_s t}\right) \quad \text{ou} \quad \sum c_{\ell s}\left(e^{k_s t} - e^{-k_s t}\right)$$

proposée par FORTUS pour φ_ℓ^T quand T est fini.

Les calculs d'identification des coefficients b_ℓ ou c_ℓ sont ici aussi très laborieux et pratiquement, il vaut mieux diagonaliser directement Σ. Mais cela explique encore le caractère périodique des fonctions propres que l'on observe sur des exemples.

(c) Comme dans le cas continu, la fonction d'autocorrélation $\rho(\tau) = e^{-\alpha|\tau|}$ ou la matrice R :

$$\begin{vmatrix} 1 & \rho & \rho^2 & \cdots & \rho^P \\ \rho & 1 & \rho & & \rho^{P-1} \\ \vdots & & & & \\ \rho^P & \rho^{P-1} & \rho^{P-2} & & 1 \end{vmatrix} \qquad \text{avec} \quad \rho = e^{-\alpha \cdot \Delta t}$$

permet d'obtenir des résultats théoriques simples (bien que légèrement approchés). Les démonstrations, reportées en Annexe fournissent comme valeurs propres :

$$\mu_s = \frac{1-\rho}{1 - 2\rho \cos \frac{\pi s}{P+1} + \rho^2}$$

$$\overrightarrow{V_s} = \{V_{sl}\} \qquad V_{sl} = \sin \frac{\pi s}{P+1} l$$

Nous en donnerons des exemples en IV-3.

On notera au passage que si, pour $T \longrightarrow \infty$, les valeurs propres sont doubles et associées respectivement à un sinus et un cosinus, et ce à la fois pour les cas continu et discrets, (quelque soit la fonction ou la matrice d'autocorrélation) cela cesse dès que T est fini. On peut le voir sur l'exemple $\rho(\tau) = e^{-\alpha|\tau|}$ où, même si les fonctions propres se ramènent à 1 sinus ou 1 cosinus, les valeurs propres ne sont plus doubles, et la première est associée à un cosinus.

(d) Enfin le passage à 2 dimensions est encore possible dans le cas $T^2 \longrightarrow \infty$ (où BRILLINGER montre que à partir de la matrice C, on peut construire Z qui sera cette fois circulante par <u>bloc</u>) et pour T^2 fini, KUENY a montré que les vecteurs propres étaient des combinaisons linéaires de sin . sin, sin . cos, ou cos . cos.

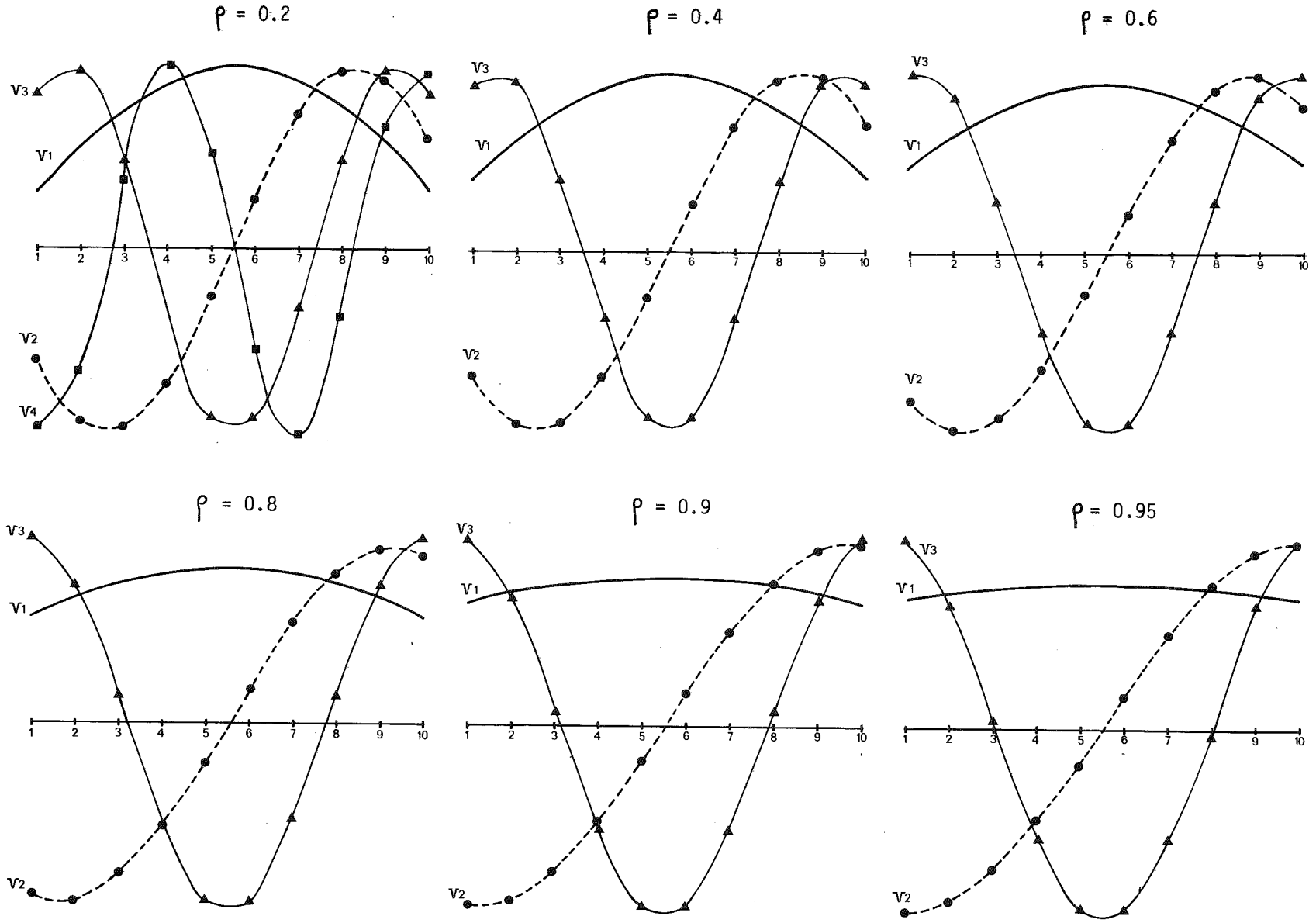


FIGURE IV - 6 : Premiers vecteurs propres d'un processus markovien sur un réseau unidimensionel régulier .

IV.3 - Exemples de simulation

IV.3.1. Modèles théoriques uni et bidimensionnels

On se propose d'étudier les éléments propres d'une matrice de corrélation quand ses éléments correspondent à des fonctions d'autocorrélation simples.

(a) On a d'abord considéré, sur un réseau hypothétique de P stations équidistantes le long d'une ligne, le modèle d'autocorrélation : $\rho(t) = e^{-\alpha|t|}$

D'où la matrice R de la page précédente, dont on calcule les éléments propres.
On vérifie que : (Fig. IV-6)

- V_1 a la forme d'un cosinus, plus ou moins aplati selon P et ρ :
  . quand P est fixé, la courbure augmente si ρ diminue. (En particulier on tend vers une constante si $\rho \longrightarrow 1$)
  . quand ρ est fixé, la courbure augmente si P diminue.
- De même V_2 ressemble à un sinus, etc... pour V_3 ---
- Le vecteur V_k recoupe k-1 fois l'axe. (Ils sont alternativement pair et impair).

Si on compare avec les résultats obtenus théoriques du IV.2.2, on constate, par exemple sur V_1, une légère divergence due aux effets de bord provenant de l'approximation utilisée sur R pour calculer les éléments propres.
De plus, le vecteur est ici renormé à 1 alors que : $\sum_{k=1}^{P} \sin^2 \frac{k\pi}{P+1} \neq 1$

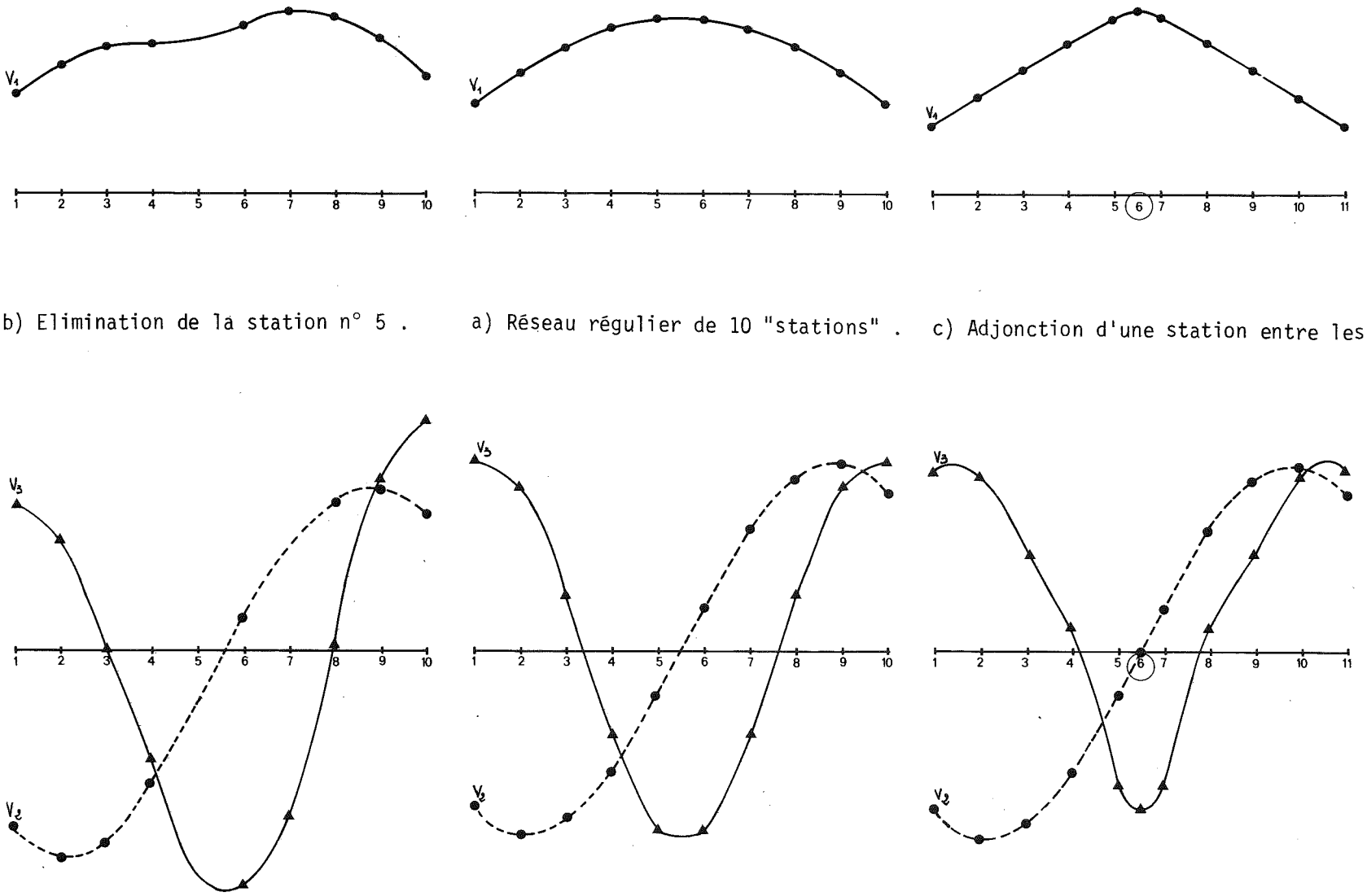
(b) Si l'échantillonnage sur les séries temporelles se fait souvent à pas constant, il est rare qu'un réseau spatial soit aussi régulier.

Nous donnons donc, pour $P = 10$, les 3 premiers vecteurs calculés en enlevant successivement la station 2, 3, ... 5 du réseau régulier (Fig.IV-7a). On constate que si ρ n'est pas très élevé ($< .8$) les déformations sont sensibles.

La même chose se produit si on ajoute une station par exemple entre 5 et 6. C'est le vecteur 1 qui est le plus affecté dans sa forme, mais les vecteurs 2 et 3 le sont aussi, et d'autant plus que ρ est faible (Fig. IV-7-c).

Enfin le cas le plus réaliste correspond à un réseau aléatoire. C'est ainsi que l'on a généré, pour le tronçon $10.\Delta x$,différents ensembles de 10 stations distribuées aléatoirement sur le tronçon (Fig.IV-8 a). On constate une forte sensibilité dès que ρ décroît, et ce même sur V_1. Bien que satisfaisant toujours le modèle théorique, on ne retrouve plus les tracés réguliers obtenus pour les stations équidistantes. Cela est vrai aussi à 2 dimensions, et explique par exemple pourquoi dans l'analyse des géopotentiels par exemple, on a des isolignes plus régulières en utilisant des champs interpolés sur des grilles régulières qu'en partant du réseau aléatoire des stations d'observations.

(c) On a d'ailleurs effectué des essais voisins en bidimensionnels. Le modèle théorique choisi pour la corrélation était : $\rho(d) = \alpha d \cdot K_1(\alpha d)$

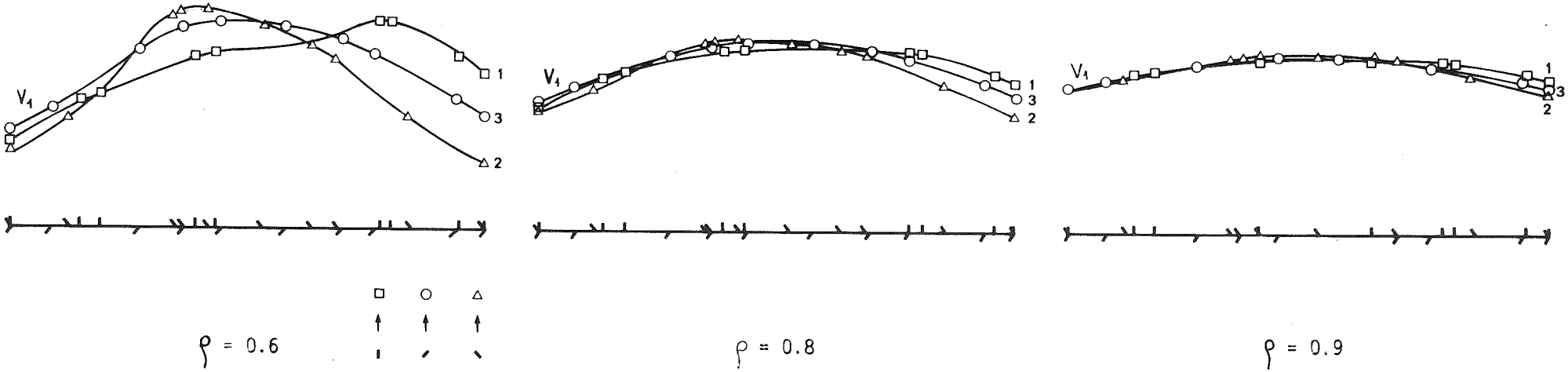


b) Elimination de la station n° 5 .

a) Réseau régulier de 10 "stations" .

c) Adjonction d'une station entre les n° 5 et 6 .

FIGURE IV - 7 : Perturbations dûes à des modifications du réseau (pour le cas $\rho = 0.6$)



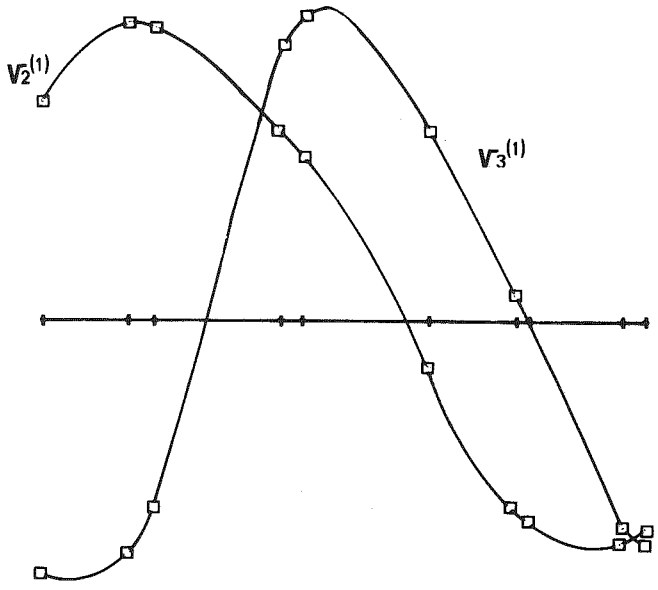
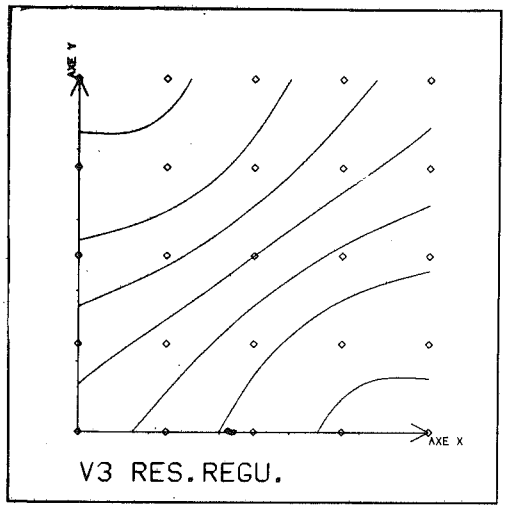
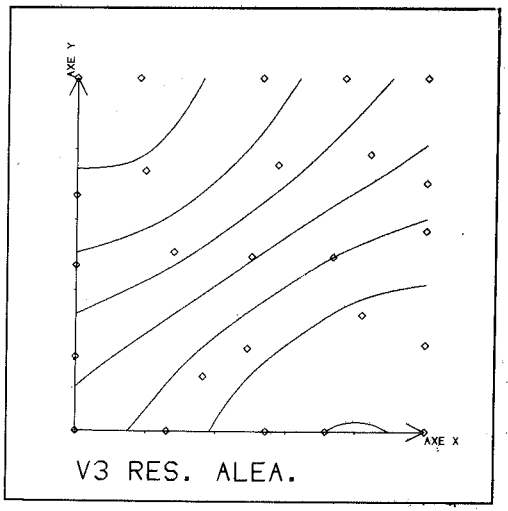


FIGURE IV - 8 : Premiers vecteurs propres d'un processus markovien sur un réseau unidimensionel aléatoire .

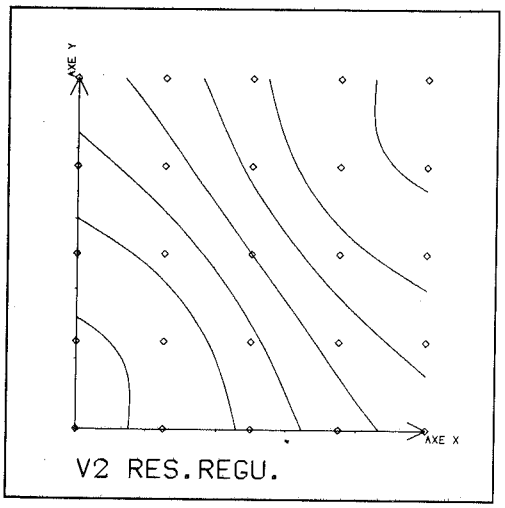
A l'exception des "stations" 1 et 10,la position des stations est tirée dans une loi normale : l'abcisse de la $j^{ème}$ est $x_j \in \mathcal{N}(j, 0.5)$
On donne 3 reseaux aléatoires différents ,et pour $\rho = .6$, .8 et .9



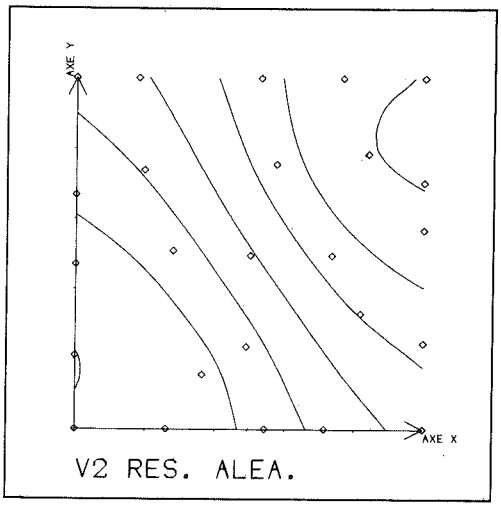
MCFISO3 LE 19/07/79 A 00:30:31



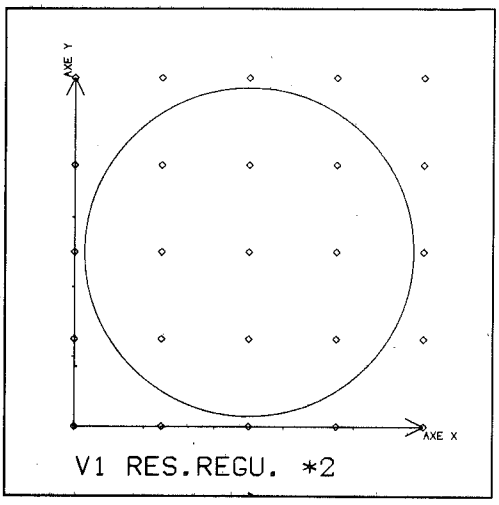
MCFISO3 LE 19/07/79 A 14:08:53



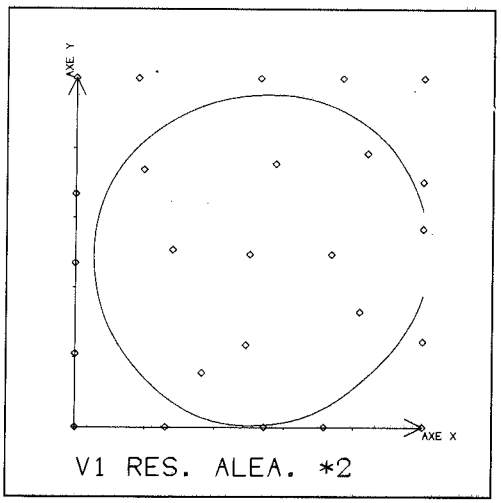
MCFISO3 LE 19/07/79 A 00:30:11



MCFISO3 LE 19/07/79 A 14:08:31



MCFISO3 LE 19/07/79 A 00:29:54



MCFISO3 LE 19/07/79 A 14:08:14

FIGURE IV - 9 : Premiers vecteurs propres d'un processus markovien bidimensionel,

a) sur une grille 5x5 régulière b) sur une grille aléatoire

associé à une équation aux dérivées partielles élliptique.

On n'a pas démontré mais seulement vérifié que la matrice de corrélation R^{-1} associée avait la même allure que celle associée à l'équation discrétisée, de ligne courante :

$$--- \quad \varepsilon \; 1 \; \varepsilon \; ------ \; 1 \; -4 \; 1 \; ------ \; \varepsilon \; 1 \; \varepsilon \; -----$$

Dans le cas théorique où $\varepsilon = 0$, on montre par analogie avec le cas unidimensionnel que les fonctions propres sont de la forme :

$$V_{kl}^{(pq)} = \sin k\left(\frac{p\pi}{m+1}\right) \cdot \sin l\left(\frac{q\pi}{m+1}\right)$$

associée à la valeur propre λ_{pq} ($p = 1,m$, $q = 1,n$).

Les valeurs numériques approchées en diffèrent sensiblement à cause des effets de bord, (surtout sur les petites grilles 5x5 utilisées) mais restent très régulières (Fig. IV-9 a).

On a aussi considéré l'effet d'un réseau légèrement perturbé (Fig. IV-9 b).

IV.3.2. Génération d'épisodes

Les essais précédents partaient non pas d'un phénomène, même idéalisé, mais d'un modèle théorique de corrélation. Pour nous approcher un peut de l'exemple des épisodes pluvieux, nous avons simulé, sur un réseau fixe de 10 stations, des observations (en général 100) selon le modèle suivant : (cf. Figure IV - 10)

- le centre de l'"épisode" suit une loi uniforme sur (0,110)
- la hauteur au centre est uniforme sur (0,400)
- la forme des épisodes est imposée et correspond à un cosinus .

Dans un premier temps , le rayon d'action de l'épisode était imposé et égal à:

$$R_x = 35., 50., 75. \text{ et } 100.$$

- Les hauteurs < 0 sont forcées à 0. On voit les résultats sur les figures IV-10. En particulier, quand R_x est de l'ordre de grandeur du réseau $R_x > 75$, on ne saisit plus la totalité de l'épisode, et la variabilité interépisode devient grande.

Les fonctions propres ressemblent alors beaucoup à celle du schéma markovien. Mais ce n'est plus vrai dès que $R_x < 75$. On voit donc le rôle de l'étendue du réseau par rapport à la période propre du phénomène. Certes il ne s'agit pas ici d'un signal périodique pur (les valeurs négatives sont écrêtées) mais de période fixée (R_x fixé).

Un cas encore plus réaliste a consisté à prendre un rayon d'action R_x variable et aléatoire $R_x \in \mathcal{N}(50, 75)$. On vérifie que l'on se rapproche alors beaucoup plus vite du schéma markovien.

On peut donc s'attendre, dans la réalité, à trouver des formes analogues à celles du schéma Markovien, même si le modèle théorique sous-jacent est différent. Nous n'avons pas fait d'essais à 2 dimensions mais il est probable que cela reste vrai, compte tenu des résultats expérimentaux sur des phénomènes variés (pluies cf Vème partie, Chap.IV, ou KUENY, 1977 ou DUBAND, 1973).

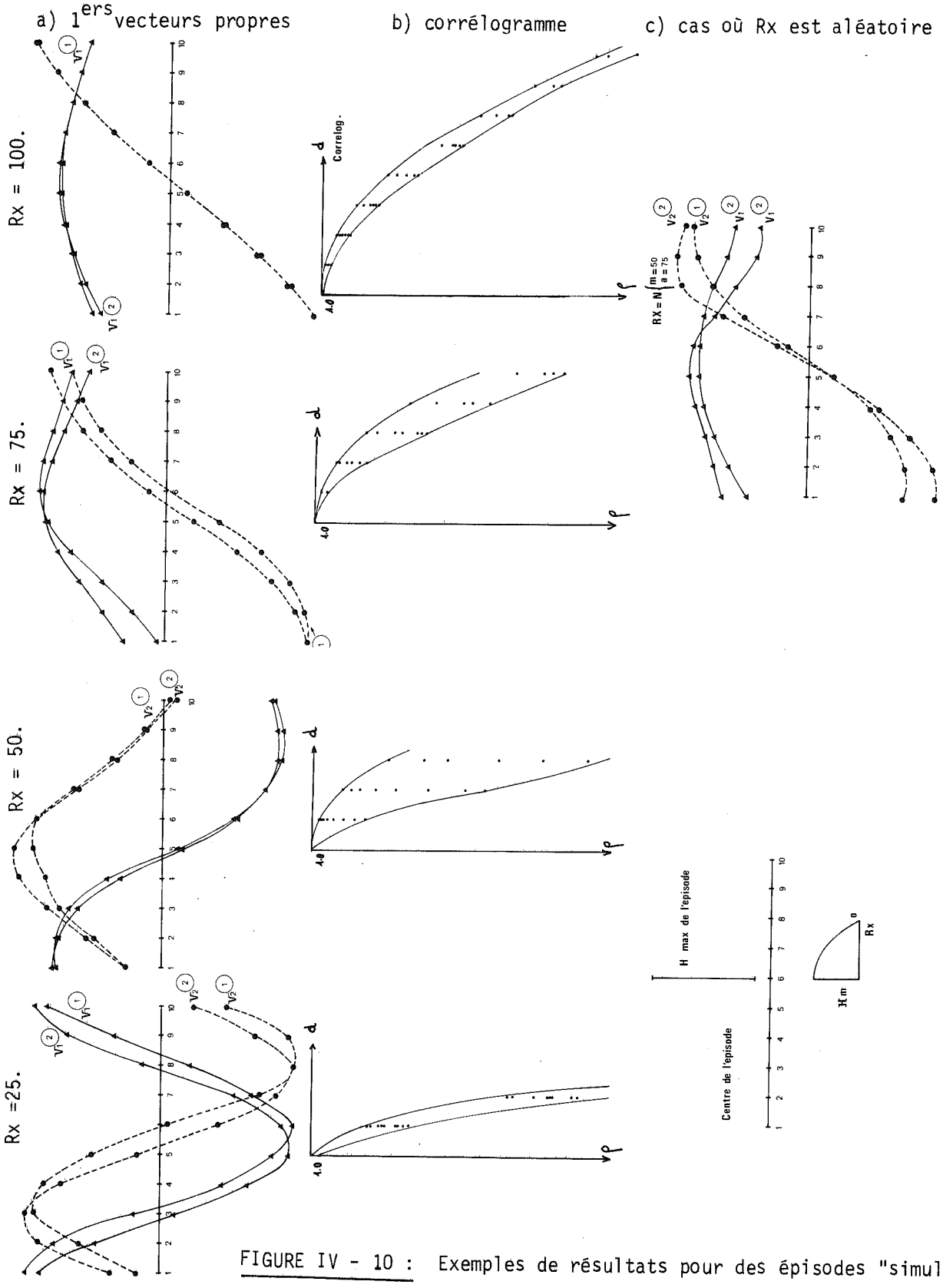


FIGURE IV - 10 : Exemples de résultats pour des épisodes "simulés" .

Le réseau fixe a une extension égale à 100 , et les épisodes ont un épicentre aléatoire sur le réseau , une "hauteur" au centre aléatoire , mais un "rayon d'action" fixe : Rx = 25 , 50 , 75 , ou 100

On donne aussi un exemple où Rx est lui aussi aléatoire : $Rx \in \mathcal{N}(50,75)$

IV.3.3. Processus simulé par des sommes de fonctions trigonométriques

Jusqu'à présent les modèles théoriques étaient solution d'équations différentielles ou aux dérivées partielles. Ici, nous allons considérer des processus théoriques qui sont essentiellement des sommes d'harmoniques indépendantes. Dans le cas d'un processus continu $X(t)$ sur $]-\infty, +\infty[$, on peut imaginer un spectre discontinu formé d'une infinité dénombrable de raies.

Nous en donnons d'abord la définition, puis des propriétés, à partir d'un exposé de MONIN et YAGLOM (1975) et enfin nous considérons le cas discret où le processus est échantillonné, puis ce que donne son analyse harmonique, c'est-à-dire l'A.C.P. classique expliquée à un tel processus.

(a) On a vu en IV.1.1 comment un processus périodique au sens des moindres carrés pouvait se mettre sous la forme :

$$X(t) = \sum_{n=-\infty}^{+\infty} \gamma_n e^{j n \omega_0 t} \qquad \omega_0 = \frac{2\pi}{T}$$

qui, s'il était réel, se réduisait à :

$$X(t) = \sum_{n=0}^{+\infty} \left(a_n . \cos n\omega_0 t + b_n . \sin n\omega_0 t \right)$$

Un autre cas, envisagé par MONIN et YAGLOM (1975) considère un processus :

$$X(t) = \sum_{k=1}^{n} \left(a_k . \cos \omega_k . t + b_k . \sin \omega_k . t \right)$$

où les valeurs données $\omega_1, \omega_2 \ldots \omega_n$ ne sont plus des multiples de ω_0.

Dans tous les cas, le processus étant réel, sa fonction de covariance (les espérances de chaque harmonique étant nulle $\forall t$ fixé) peut s'écrire :

$$C(t_1, t_2) = C(t_1 - t_2) = C(\tau) = \sum_{k=0}^{n} \sigma_k^2 \cos \omega_k . \tau$$

($\neq$ pour le cas périodique : $C(\tau) = \sum_{k=0}^{\infty} \sigma_k^2 \cos . k . \omega_0 . t$)

avec $\sigma_k^2 = E[a_k^2] = E[b_k^2]$ et en supposant $E[a_k . a_l] = E[b_k . b_l] = 0 \quad \forall k \neq l$

On peut même montrer que tout processus stationnaire peut se mettre sous cette forme à condition que $n \longrightarrow \infty$ et que les ω deviennent "denses" sur l'axe des ω.

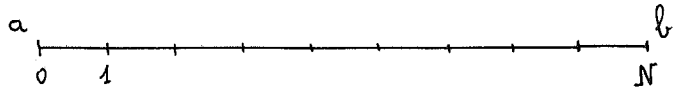
Et on arrive alors à la représentation de la fonction X en intégrale de Fourier.

Par contre, si le processus, quel qu'il soit, est échantillonné avec un pas Δt et n'est connu que sur un nombre fini de points, on peut le représenter exactement

par une somme finie d'harmoniques dont les fréquences sont multiples de la fréquence d'échantillonnage. Cela revient à rendre le processus périodique (prolongé par périodicité en dehors de l'intervalle où on l'a échantillonné) ainsi que sa fonction d'autocorrélation.

(b) Rappels sur l'interpolation trigonométrique.

Si l'on échantillonne une fonction $f(x)$ en N+1 points tels que N+1 soit pair (le cas impair est tout à fait analogue) $N+1 = 2L$



Et si on considère l'ensemble de fonctions :

$$1 \quad \cos x \quad \cos 2x \quad \cdots\cdots \quad \cos(L-1)x \quad \cos Lx \ (\equiv 1)$$
$$0 \quad \sin x \quad \sin 2x \quad \cdots\cdots \quad (L-1)x \quad \sin Lx \ (\equiv 0)$$

cet ensemble est orthogonal au sens discret sur l'ensemble de points suivants :

$$0 \quad \frac{\pi}{N} \quad \frac{2\pi}{N} \quad \cdots\cdots\cdots \quad \frac{2(N-1)\pi}{N}$$

Or si on considère la fonction échantillonnée sur l'intervalle $[a, b]$ et si l'on veut ramener cet intervalle à 2π, il suffit de faire le changement de variable :

$$x \longrightarrow x' = 2\pi . \frac{x-a}{b-a}$$

et si : $\left.\begin{aligned} x &= a + k.\Delta x \\ b-a &= 2L.\Delta x \end{aligned}\right\}$ alors : $x' = 2\pi \frac{k.\Delta x}{2L.\Delta x} = k\frac{\pi}{L}$

Cela suppose implicitement qu'on ne va pas au point d'abscisse 2L mais seulement de 0 à 2 L-1 car $f(2L) = f(0)$: la fonction est supposée périodique de période 2π (en fait $b-a$).

On vérifie alors que notre ensemble de fonctions de base :

$$1 \quad \cos\frac{\pi}{L}x \quad \cos\frac{2\pi}{L}x \quad \cdots\cdots \quad \cos\frac{L-1}{L}\pi x \quad \cos\frac{L}{L}\pi x$$
$$\sin\frac{\pi}{L}x \quad \sin\frac{2\pi}{L}x \quad \cdots\cdots \quad \sin\frac{L-1}{L}\pi x$$

est orthogonal sur les points $x = 0, 1 \ldots, 2L-1$.

$$\sum_{x=0}^{2L-1} \sin\frac{\pi}{L}kx \cdot \sin\frac{\pi}{L}mx = \begin{cases} 0 & k \neq m \\ L & k = m \neq 0 \end{cases}$$

$$\sum_{x=0}^{2L-1} \sin\frac{\pi}{L}kx \cdot \cos\frac{\pi}{L}mx = 0 \quad \forall k, m$$

$$\sum_{x=0}^{2L-1} \cos\frac{\pi}{L}kx \cdot \cos\frac{\pi}{L}mx = \begin{cases} 0 & k \neq m \\ L & k = m \neq 0 \\ 2L & k = m = 0 \end{cases}$$

On montre alors que la fonction $f(x)$ quelconque connue en $2L$ points peut s'écrire :

$$f(x) = \frac{1}{2} a_0 + \sum_{k=1}^{L-1} \left(a_k \cos k \frac{\pi}{L} x + b_k \sin k \frac{\pi}{L} x \right) + \frac{1}{2} a_L \cos \pi x$$

En fait, si on se ramène à la théorie de l'interpolation, on a $2L$ points et $2L$ fonctions de base, indépendantes au sens des vecteurs et, dans notre cas, orthogonales, le système est donc bien déterminé.

De plus, l'orthogonalité de la base permet d'introduire les notions de projection orthogonale et de produit scalaire.

On trouve donc les valeurs de a_k et b_k par :

$$a_k = \frac{1}{L} \sum_{x=0}^{2L-1} f(x) \cdot \cos \frac{\pi}{L} . k x \qquad k = 0, \ldots L$$

$$b_k = \frac{1}{L} \sum_{x=0}^{2L-1} f(x) \cdot \sin \frac{\pi}{L} . k x \qquad k = 1, \ldots L-1$$

Enfin, on montre que, en plus de l'interpolation exacte, la suppression de certaines harmoniques ou la prise en compte des M premières seulement donne une erreur quadratique moyenne sur les $2L$ points facile à évaluer par :

$$\sum_{x=0}^{2L-1} \left(f - f_M \right)^2 = \sum_x f(x)^2 - L \left[\frac{a_0^2}{2} + \sum_{k=1}^{M} \left(a_k^2 + b_k^2 \right) \right]$$

Tous ces résultats se trouvent démontrés dans SCHEID F. (1968) ou HAMMING R.W. (1962). On trouvera aussi l'analogie pour le cas où $N+1 = 2L+1$ (nombre impair de points).

Dans notre cas, chaque réalisation (ou observation) i, du processus (en fait un tronçon de réalisation) pourra s'interpoler exactement à l'aide des fonctions de base ci-dessus, d'où :

$$f^{(i)}(x) = \frac{1}{2} a_0^{(i)} + \sum_{k=1}^{2L-1} \left[a_k^{(i)} \cos \frac{\pi}{L} kx + b_k^{(i)} \sin \frac{\pi}{L} kx \right] + \frac{1}{2} a_L^{(i)} \cos \frac{\pi}{L} x$$

Naturellement, cette représentation $f^*(x)$ diffère du vrai processus $f(x)$ et en particulier les propriétés spectrales du processus interpolé diffèrent de celles du processus original (DEVILLE J.C. 1974)

(c) Aspect statistique (N observations) - Corrélation.

On considère maintenant que l'on a N réalisations connues sur $2L$ points :

$f^{(1)}(0)$	$f^{(1)}(1)$	----------	$f^{(1)}(2L-1)$
$f^{(2)}(0)$	$f^{(2)}(1)$	----------	$f^{(2)}(2L-1)$
$f^{(i)}(0)$	$f^{(i)}(1)$	----------	$f^{(i)}(2L-1)$
$f^{(N)}(0)$	$f^{(N)}(1)$	----------	$f^{(N)}(2L-1)$

ces valeurs étant centrées réduites au sens habituel.

Comme on a : $a_{k_1}^{(i)} = \frac{1}{L} \sum_{x=0}^{2L-1} f^{(i)}(x) \cdot \cos \frac{\pi}{L} k x$

$$E\left[a_k^{(i)}\right] = \frac{1}{N} \sum_{i=1}^{N} a_k^{i} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{L} \sum f^{(i)}(x) . \cos \frac{\pi}{L} . k . x$$

$$= \frac{1}{L} \sum_{x=0}^{2L-1} \cos \frac{\pi}{L} . k x \times \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} f^{(i)}(x) \right)}_{E\left[f^{(i)}(x)\right] = 0 \ \forall x}$$

$$= 0$$

De même pour $E\left[b_k^{(i)} \right] = 0$

Et le fait d'avoir forcé, sur N échantillons finis la moyenne arithmétique à 0 pour $f(x)$ donne bien la nullité de $E[a_k]$ et $E[b_k]$.

Le problème est un peu plus compliqué pour $E\left[a_k^{(i)} . b_k^{(i)} \right]$

En effet :

$$a_k^{(i)} . b_k^{(i)} = \frac{1}{L} \left[1 . f^{(i)}(0) + f^{(i)}(1) \cos \frac{k \pi}{L} + f^{(i)}(2) \cos \frac{2 k \pi}{L} + \cdots f^{(i)}(2L-1) \cos \frac{2L-1}{L} k \pi \right]$$

$$\times \left[0 . f^{(i)}(0) + f^{i}(1) \sin \frac{k \pi}{L} + f^{(i)}(2) \sin \frac{2 k \pi}{L} + f^{(i)}(2L-1) \sin \frac{2L-1}{L} k \pi \right]$$

et quand on fait la sommation, on obtient des termes de la forme :

$$f^{(i)}(l)^2 \cos \frac{k \pi}{L} l . \sin \frac{k \pi}{L} l$$

$$f^{(i)}(l) \times f^{(i)}(m) \cos \frac{k \pi}{L} l . \sin \frac{k \pi}{L} m$$

$$f^{(i)}(l) \times f^{(i)}(m) \sin \frac{k \pi}{L} l . \cos \frac{k \pi}{L} m$$

Les premiers termes sommés sur i donnent :

$$\sum_{j=0}^{2L-1} \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} f^{(i)}(k)^2 \right)}_{\sigma_f^2 = 1} \cos \frac{k \pi}{L} . j \times \sin \frac{k \pi}{L} . j = \sum_{j=0}^{2L-1} \cos \frac{k \pi}{L} j . \sin \frac{k \pi}{L} j = 0 \quad \forall k, j$$

(d'après les relations d'orthogonalité)

C'est-à-dire que la réduction des variables : $\frac{1}{N} \sum_{i} f^{(i)}(x)^2 = 1 \ \forall x$, qui recouvre condition en espérance : $E[f(x)^2] = 1 \ \forall x$ est exacte sur notre N échantillons et entraîne la nullité des termes carrés.

Par contre, les termes croisés ne s'annullent pas par construction sur un N échantillon, et les termes estimés a_k et b_k , ou a_k et b_l ne seront pas indépendants sauf s'il le sont dans le processus lui-même, par exemple si celui-ci est généré comme somme de L harmoniques de périodes multiples de $\frac{\pi}{L}$.

Dans le cas contraire, cela sera d'autant plus vrai que L sera grand (tronçon suffisamment long de réalisation, échantillonné à pas petit, cf aussi IV.1.1(b)) et que N sera grand.

(d) Analyse en composantes principales

Si on considère l'observation interpolée sur les 2L points variables :

$$y^{(i)}(x) = \frac{a_0^{(i)}}{2} + \sum_{k=1}^{L-1} \left(a_k^{(i)} \cos \frac{\pi}{L} kx + b_k^{(i)} \sin \frac{\pi}{L} kx \right) + \frac{a_L}{2} \cos \pi x$$

et si on regarde comment s'écrit le vecteur observation i :

$$Y_{iV} = \left[y_{i0}\ y_{i1} \text{ ---- } y_{ij} \text{ ------ } y_{i\,2L-1} \right]$$

on peut vérifier que cela peut se mettre sous la forme :

$$Y_{iV} = \left[\frac{a_0}{2}\ a_1 \text{ --- } a_k \text{ --- } a_{L-1}\ a_L\ b_1\ b_2 \text{ ---- } b_k \text{ --- } b_L \right] \times \begin{bmatrix} 1 & 1 & 1 & \text{----} & 1 & \text{------} & 1 \\ 1 & \cos\frac{\pi}{L} & \cos 2\frac{\pi}{L} & \text{----} & \cos j\frac{\pi}{L} & \text{------} & \cos(2L-1)\frac{\pi}{L} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 1 & \cos k\frac{\pi}{L} & \cos 2k\frac{\pi}{L} & \text{----} & \cos jk\frac{\pi}{L} & \text{------} & \cos(2L-1)k\frac{\pi}{L} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 1 & -1 & 1 & \text{----} & (-1)^j & \text{------} & (-1)^{2L-1} \\ 0 & \sin\frac{\pi}{L} & \sin 2\frac{\pi}{L} & \text{----} & \sin j\frac{\pi}{L} & \text{------} & \sin(2L-1)\frac{\pi}{L} \\ 0 & \sin\frac{2\pi}{L} & \sin 2\frac{2\pi}{L} & \text{----} & \sin j\frac{2\pi}{L} & \text{------} & \sin(2L-1)\frac{2\pi}{L} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & \sin k\frac{\pi}{L} & \sin 2k\frac{\pi}{L} & \text{----} & \sin jk\frac{\pi}{L} & \text{------} & \sin(2L-1)k\frac{\pi}{L} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & \sin\frac{L-1}{L}\pi & \sin 2\frac{L-1}{L}\pi & \text{---} & \sin j\frac{L-1}{L}\pi & \text{------} & \sin(2L-1)\frac{L-1}{L}\pi \end{bmatrix} \begin{matrix} 0 \\ 1 \\ \vdots \\ L \\ 1 \\ 2 \\ \vdots \\ L-1 \end{matrix}$$

Que l'on peut écrire matriciellement : $Y_{iV} = \vec{A}_i \cdot M$

Et l'ensemble des observations s'écrit : ---------- $Y_{0V} = \begin{bmatrix} \vec{A}_1 \\ \vdots \\ \vec{A}_i \\ \vdots \\ \vec{A}_N \end{bmatrix} \cdot M = A \cdot M$

et, si on suppose les variables déjà centrées, on a la matrice de covariance :

$$\frac{1}{N} Y_{0V} \cdot Y_{0V} = \frac{1}{N} \cdot M^t \cdot A^t \cdot A \cdot M$$

avec :

$$A^t \cdot A \begin{bmatrix} \sum a_0^{(i)2} & \sum a_0^{(i)} a_1^{(i)} & \text{----} \sum a_0^{(i)} a_L^{(i)} & \sum a_0^{(i)} b_1^{(i)} & \text{----} & \sum a_0^{(i)} b_{L-1}^{(i)} \\ & \sum a_1^{(i)2} & \text{----} \sum a_1^{(i)} a_L^{(i)} & \sum a_1^{(i)} b_1^{(i)} & \text{----} & \sum a_1^{(i)} b_{L-1}^{(i)} \\ & & \text{---} & \text{---} & \text{---} & \text{---} \\ & & & \sum b_1^{(i)2} & \text{----} & \sum b_1^{(i)} b_{L-1}^{(i)} \\ & & & & \text{---} & \text{---} \\ & & & & & \sum b_{L-1}^{(i)2} \end{bmatrix}$$

Or, si les hypothèses d'indépendance sont vérifiées :

$$E[a_k \cdot b_k] = E[a_k \cdot b_l] = E[a_k \cdot a_l] = E[b_k \cdot b_l] = 0$$

alors, $A^t \cdot A$ est diagonale. La matrice des vecteurs propres est M^t, et on vérifie que $\frac{1}{N} M^t \cdot M$ est la matrice unité I (car on retrouve toutes les relations d'orthogonalité.).

Dans l'écriture ci-dessus, les termes diagonaux ne sont pas classés par valeurs décroissantes, mais il suffit d'effectuer des permutations sur A et M pour les amener en position adéquate (Sur M^t cela revient à permuter vecteur par vecteur).

On a donc trouvé à la fois la matrice diagonale et la matrice de passage associées à la matrice de variance-covariance.

Remarque numérique :

On remarquera aussi que les vecteurs de M^t ne sont pas normés à 1. Pour les renormer, il faut :

- Pour le 1er, tenir compte de : $\sum_{k=0}^{2L-1} 1^2 = 2L \Longrightarrow$ diviser par $\frac{1}{2L}$
- Idem pour : $0 \quad \cos\pi \quad \cos 2\pi \quad \cos 3\pi \quad \ldots\ldots \cos(2L-1).\pi$

$$\sum_{k=0}^{2L-1} [-1^k]^2 = 2L$$

- Par contre, pour les autres :

$$\sum \sin k\left(\frac{m\pi}{L}\right).\sin k\left(\frac{n\pi}{L}\right) = \delta_{mn} L \,, \quad \sum \cos k \frac{m\pi}{L}.\cos k \frac{n\pi}{L} = \delta_{mn}.L$$

il faut seulement diviser par L

Une façon de faire est de diviser par $\sqrt{L}$ la matrice M^t, qui devient alors orthonormée, et de remultiplier $A^t.A$ par L.

D'où :

$$A^t A = \mathrm{Diag}\left[\frac{1}{N}.L\sum_{i=1}^{N} a_k^{(i)2} \text{ ou } b_k^{(i)2}\right]$$

$$M = \begin{bmatrix} \frac{1}{\sqrt{L}\sqrt{2}} & \frac{1}{\sqrt{L}\sqrt{2}} & \cdots\cdots\cdots & \frac{1}{\sqrt{L}\sqrt{2}} \\ \frac{1}{\sqrt{L}} & \frac{1}{\sqrt{L}}\cos\frac{\pi}{L} & \cdots\cdots & \frac{1}{\sqrt{L}}\cos(2L-1)\frac{\pi}{L} \\ \frac{1}{\sqrt{L}} & \frac{1}{\sqrt{L}}\cos\frac{L-1}{L}\pi & \cdots\cdots & \frac{1}{\sqrt{L}}\cos(2L-1).\frac{L-1}{L}\pi \\ \frac{1}{\sqrt{L}\sqrt{2}} & \frac{-1}{\sqrt{L}\sqrt{2}} & \cdots\cdots & \frac{-1}{\sqrt{L}\sqrt{2}} \\ 0 & \frac{1}{\sqrt{L}}\sin\frac{\pi}{L} & \cdots\cdots & \frac{1}{\sqrt{L}}\sin(2L-1)\frac{\pi}{L} \\ 0 & \frac{1}{\sqrt{L}}\sin\frac{L-1}{L}\pi & \cdots\cdots & \frac{1}{\sqrt{L}}\sin(2L-1)\frac{L-1}{L}\pi \end{bmatrix}$$

Pour les termes a_0, a_L, b_0 il y a un facteur $\sqrt{2}$ supplémentaire.

(e) Simulations

Sur un réseau fictif de P = 12 noeuds, on a généré un certain nombre de réalisations (100 ou plus) de processus par :

i-ème réalisation
j-ème noeud

$$X(i,j) = \frac{1}{2}a_0(i) + \sum_{k=1}^{NH} [a_k(i).\cos\theta_{kj} + b_k(i)\sin\theta_{kj}]$$

avec $\theta_{kj} = 2\frac{k\pi}{P}(j-1)$

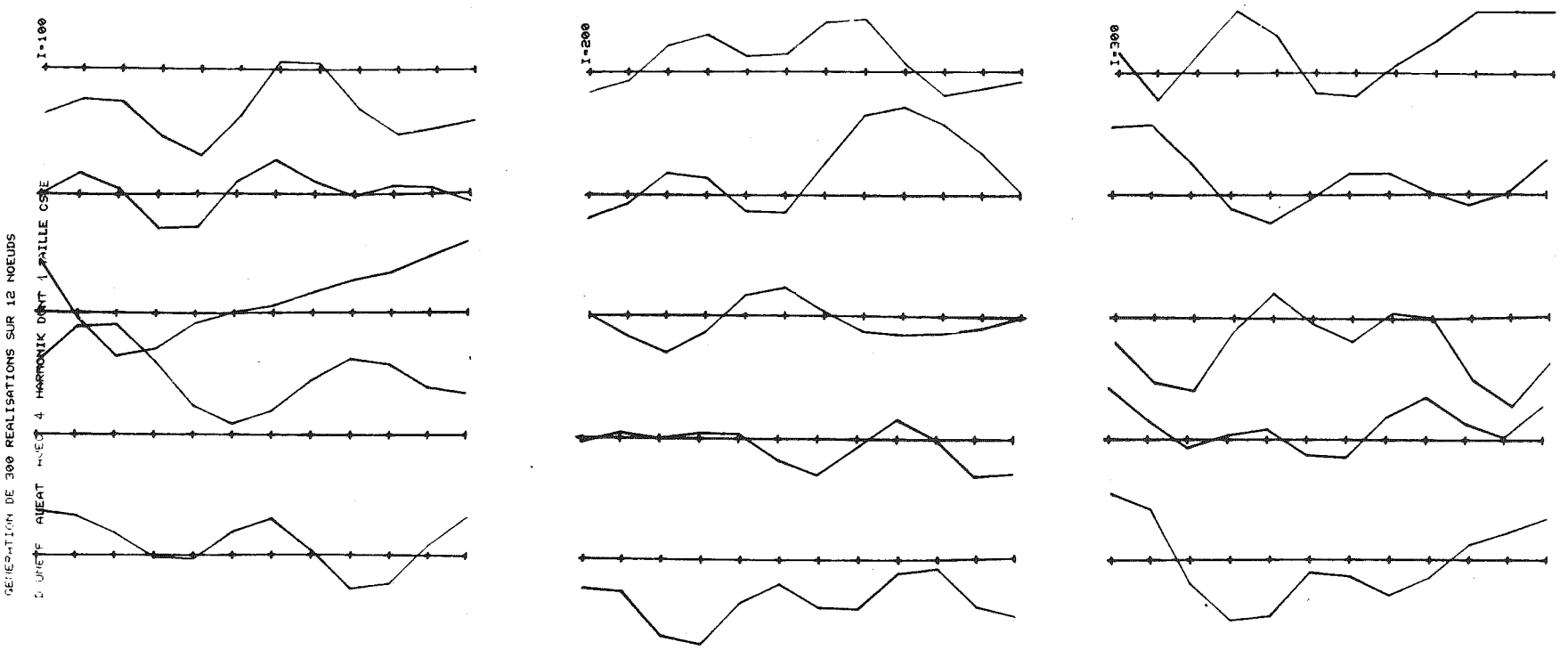


FIGURE IV - 11 : Exemples de réalisations de processus trigonométriques .

(On donne ici une simulation toutes les 20 , parmi les 300 effectuées.
Bien que n'utilisant que 4 harmoniques , on notera l'analogie de forme avec la dérive moyenne selon l'axe SW-NE des épisodes cèvenols) .

NH étant le nombre d'harmoniques (≤ 6).

Les valeurs a_{kk} et b_{kk} étant extraites d'une loi $\mathcal{N}(0, \sigma_k)$

On vérifie alors que :

- Il y a 2 x NH valeurs propres non nulles (+1 si on a mis un effet de taille a_o). Naturellement pour vérifier cela on a généré moins d'harmoniques que le maximum autorisé.
- Ces valeurs propres sont très voisines des variances demandées pour a_{kk} et b_{kk} soit σ_k^2. En fait, compte tenu de la façon dont on a renormé, on doit trouver $L.\sigma_k^2$
- Les valeurs propres sortent par "paire", et par ordre de variances décroissantes σ_k^2. En particulier, l'harmonique a_o sort bien en premier quand on a mis un effet de taille dominant.
- Les vecteurs propres sont voisins de $\left\{\frac{1}{\sqrt{L}}.\cos\frac{k\pi}{L}(j-1)\right\}$ $\left\{\frac{1}{\sqrt{L}}.\sin\frac{k\pi}{L}(j-1)\right\}$

On donne dans le tableau I les résultats d'un exemple:

GENERATION DE 300 REALISATIONS SUR 12 NOEUDS N = 300 2L = 12

CARACT. DES HARMO GENEREES

	Ecart-types demandés	$\overline{a_{kk}}$ ou $\overline{b_{kk}}$ observées	$\overline{a_{kk}^2}$ ou $\overline{b_{kk}^2}$ observées	$r(a_k, b_k)$
HARMO N = 0	σ_0 THEO = 200.0	MREEL = -12.27	VREEL = 42336.28	
HARMO N = 1	σ_1 THEO = 70.7	MAREEL = -3.95 MBREEL = 6.43	VAREEL = 4625.23 VBREEL = 5315.36	RAB = 0.090
HARMO N = 2	σ_2 THEO = 50.0	MAREEL = 2.51 MBREEL = -1.14	VAREEL = 2402.10 VBREEL = 2552.18	RAB = 0.044
HARMO N = 3	σ_3 THEO = 35.3	MAREEL = 0.46 MBREEL = -0.58	VAREEL = 1363.35 VBREEL = 1197.38	RAB = -0.047

MOYENNES DES STATIONS

-7.11 -6.66 -5.25 -1.63 1.60 2.16 -0.13 -5.07 -11.50 -15.66
-14.42 -9.95

STD DES STATIONS

145.30 146.63 143.77 139.80 134.12 127.52 128.30 137.90 145.31 141.44
139.30 139.38

MATRICE DE CORRELATION

1.0000									
0.8646	1.0000								
0.6915	0.8489	1.0000							
0.4457	0.5711	0.8404	1.0000						
0.4391	0.4409	0.5595	0.8290	1.0000					
0.4051	0.4472	0.4344	0.5268	0.8081	1.0000				
0.3636	0.4161	0.4279	0.3694	0.4641	0.7984	1.0000			
0.4010	0.3656	0.3750	0.3738	0.3449	0.4710	0.8076	1.0000		
0.4322	0.3804	0.3286	0.3614	0.3868	0.3793	0.5388	0.8558	1.0000	
0.4328	0.4123	0.3716	0.3426	0.3883	0.4366	0.4550	0.5882	0.8406	1.0000
0.8485	0.4274	0.4179	0.3775	0.3534	0.4148	0.4486	0.4277	0.5432	0.8396
1.0000									
0.8425	0.5837	0.4486	0.4297	0.4008	0.3702	0.3981	0.4232	0.4335	0.5693
0.8954	1.0000								

VALEURS PROPRES

128122.8	33327.6	26360.2	15606.6	13922.7	8229.0	7013.7	0.0	0.0	0.0
$2L.E\left[\frac{a_o^2}{4}\right] = \frac{L}{2}E\left[a_o^2\right]$	$L.E\left[a_1^2 \text{ ou } b_1^2\right]$		$L.E\left[a_2^2 \text{ ou } b_2^2\right]$		$L.E\left[a_3^2 \text{ ou } b_3^2\right]$				
$= 3\times(200)^2 = 120.000.$	$= 6\times(70.7)^2 = 30.000$		$= 6\times(50)^2 = 15.000$		$= 6\times(35.3)^2 = 7500$				

VECTEURS PROPRES

V(1)	0.316	0.319	0.303	0.282	0.262	0.248	0.261	0.274	0.297	0.300	0.294	0.301
V(2)	-0.169	-0.328	-0.391	-0.357	-0.236	-0.054	0.156	0.344	0.430	0.361	0.237	0.227
V(3)	-0.394	-0.228	0.021	0.228	0.358	0.380	0.337	0.245	0.096	-0.102	-0.302	-0.421

On constate qu'en pratique :

- Les "variables" générées X_j correspondant à la station j n'ont pas même variance (il ne faut donc pas utiliser la matrice de corrélation)
- De plus, elles n'ont pas une moyenne nulle (bien que l'espérance en soit nulle). Le centrage classique introduit donc un biais et il vaut mieux travailler sur la matrice des produits croisés.
- Bien entendu, il y a une corrélation entre a_k et b_l même si $E[a_k . b_l] = 0$, ce qui rend $A^t . A$ non diagonale, mais à diagonale nettement dominante.
- Les vecteurs propres correspondent bien à des sinus ou des cosinus mais sont légèrement déphasés par rapport aux valeurs théoriques. Toutefois l'effet de taille engendre bien un vecteur propre quasi-constant.

On donne aussi quelques exemples de réalisations du processus (Fig. IV-11) que l'on pourra comparer aux profils longitudinaux de nos épisodes cévenols.

IV.4 - Conclusions

Ce chapitre nous a permis de faire le point sur les analogies entre l'A.C.P. classique d'un tableau de données (provenant éventuellement de l'échantillonnage d'un processus) et l'analyse spectrale des processus.

En allant du théorique au pratique :

(a) On constate que l'analogie n'est totale que :
- si l'on traite le processus sous forme complexe
- sur une étendue T de $-\infty$ à $+\infty$.

Dans ce cas, que l'on traite le processus en continu, ou sous forme discrète, on trouve bien pour valeur double λ_ω l'énergie associée à l'harmonique de pulsation ω et pour fonctions propres les fonctions trigonométriques correspondantes ($\sin \omega t$ et $\cos \omega t$, en fait $e^{\pm j\omega t}$)

(b) Si le processus est réel et s'il est harmonique, les valeurs propres se dédoublent encore en λ_{ω_n} associée à $\sin \omega_n t$ et $\cos \omega_n t$ et sont égales à la 1/2 énergie de l'harmonique correspondante .

(c) Si le processus est échantillonné, et si l'étendue T considérée est finie, alors intervient la fréquence f_e d'échantillonnage. En effet si les fréquences "vraies" contenues dans le processus ne sont pas multiples de f_e, il y a diffusion des fréquences (analogue aux biais dans l'estimation des spectres).

(d) Si le processus n'est pas harmonique, alors l'A.C.P. n'est plus une analyse spectrale au sens strict. Les valeurs propres ne sont plus doubles (en espérance) et les fonctions propres ne sont , en général, plus des fonctions trigonométriques simples, sauf si l'étendue $T \longrightarrow \infty$.

(e) En pratique, les fonctions (ou vecteurs) propres sont soumises à de nombreux biais ou fluctuations d'échantillonnage (nombre limité de réalisations, nombre limité et position aléatoire des points de mesure) etc... On peut quand même considérer

grossièrement la fonction propre comme une espèce d'"harmonique empirique" (en fait de 1/2 harmonique) et la valeur propre comme une "énergie" associée.

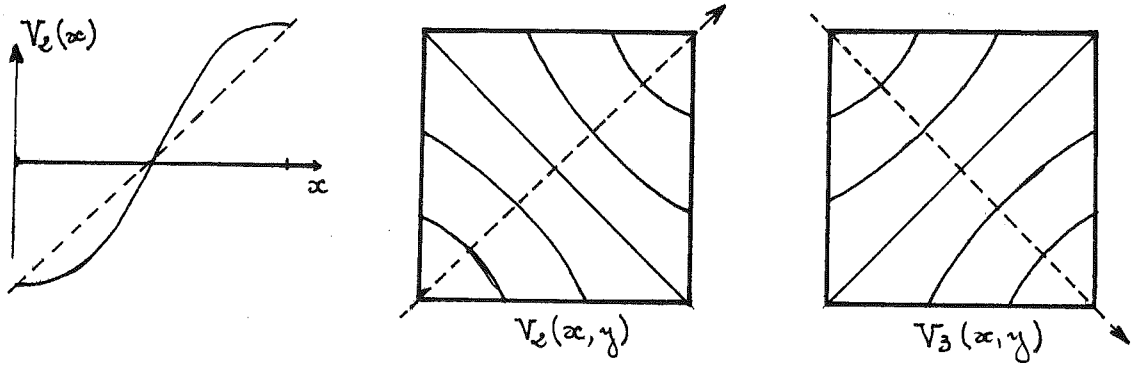
(f) La plupart de ces résultats proviennent de considérations à 1 dimension, mais s'étendent sans trop de problèmes à 2 dimensions.

Ils supposent toutefois que le processus sous-jacent est homogène. L'A.C.P. d'un réseau où un groupe de stations a un comportement tout à fait particulier, associé à un autre phénomène, n'est plus une A.C.P. de processus et ne permet plus d'interpréter les fonctions propres en terme d'"harmonique".

(g) Enfin, on a vu que les vecteurs propres d'une matrice de covariance empirique ne sont qu'une estimation des fonctions propres associées à la fonction de corrélation qui caractérise le processus. On y reviendra dans la Vème partie.

Remarque - Dans la plupart des "processus" que nous rencontrons en pratique, le spectre est monotone décroissant en fonction de la fréquence, ce qui signifie que ce sont les "harmoniques" de plus basse fréquence qui transportent le plus d'énergie.

Cela explique pourquoi le 1er vecteur propre ressemble souvent à une moyenne arithmétique, et le 2ème à un sinus. Toutefois sur 1/2 période $[-\frac{\pi}{2}, +\frac{\pi}{2}]$ il y a peu de différence entre $\sin x$ et x, c'est pourquoi on interprète souvent ce vecteur comme un "gradient" entre les extrémités du champ.



Dans le cas de 2 dimensions d'espace, ce sont les vecteurs V_2 et V_3 qui correspondent à un gradient selon les 2 directions (convenablement choisies). Si on se rappelle que $\sqrt{\lambda_2}\, V_2$ et $\sqrt{\lambda_3}\, V_3$ sont les coefficients de corrélation de la station j avec les axes F2 et F3 et si les fonctions $V_2(j)$ et $V_3(j)$ sont quasi-équivalentes à x_j et y_j, il ne faut pas s'étonner de retrouver, quand on représente les variables X_j dans les axes 2 et 3 de $\mathbb{R}^N$, la position géographique des stations S_j dans (x,y).

Ce fait est surtout vrai pour les processus quasi markoviens (car même sur un intervalle fini, et en processus discret, les fonctions propres restent des sinus ou cosinus purs). Il disparait quand on transforme les données et donc le spectre (renormalisation, ecarts à une "normale" etc...)

"Etudie le passé si tu veux prévoir le futur"
Confucius
"On veut savoir plus qu'on ne voit, c'est là la difficulté"
Fontenelle

CINQUIEME PARTIE

PREVISION ET ESTIMATION : PREVISION DES PHENOMENES ACCIDENTELS PAR ANALYSE DISCRIMINANTE ET INTERPOLATION OPTIMALE DE PROCESSUS

Cette partie capitalise les résultats des précédentes et montre que si des analyses suffisamment fines ont permis de bien poser un problème, il existe en général des techniques rapides et efficaces permettant de le résoudre.

Le cas des phénomènes accidentels est particulièrement exemplaire. Si on a pu montrer que, à l'aide de variables convenablement choisies, les situations normales et accidentelles constituent des populations distinctes, alors les techniques d'analyse discriminante, dans leurs développements les plus récents, proposent des algorithmes de résolution très performants.

Quand, par contre, le problème est moins bien posé, et que la physique ne peut fournir toutes les données souhaitables ni les schémas explicatifs possibles, l'existence de classes n'est plus apparente a priori et peut être recherchée par des algorithmes, plus ou moins robustes hélàs.

Les problèmes d'estimation ne se limitent pas à la prévision du futur possible, car bien souvent il faut compléter la connaissance partielle d'un état présent insuffisamment décrit par un nombre limité de points de mesure: C'est le cas des champs de variables météorologiques, mesurés par un réseau ponctuel plus ou moins lâche, ce qui nous conduit à estimer soit des points manquants , soit des intégrales spatiales .

CHAPITRE I

LES TECHNIQUES D'ANALYSE DISCRIMINANTES

I.1 - Les phénomènes "accidentels"

(a) Ils s'opposent essentiellement à ceux que l'on peut qualifier de continus (débits, durée d'insolation, température), pouvant prendre n'importe quelle valeur réelle sur un intervalle, borné ou non.

Ils se caractériseront, dans le cas le plus simple, par l'occurence ou la non-occurence et dans un cas plus complexe, par l'appartenance à une classe C_k d'un ensemble fini de classes mutuellement exclusives.

Exemples

- de phénomènes accidentels :
  orages à grêle, journée avec verglas, journée avec avalanches
- ou pouvant s'y ramener :
  pluie - non pluie, temps clair/ variable/ couvert.

Certes le phénomène n'est pas nécessairement purement booléen, et il arrive que l'on puisse coter son intensité de manière discrète ou continue au sein de 1 ou de chacune des classes.

Exemples

- intensité discrète : journée avec 1, 2, ... p avalanches (le problème est de savoir s'il y a continuité avec les journées à 0 avalanches)
- intensité continue dans une classe : journée de pluie avec x_{mm} observés $0.1 < x < \infty$
- intensité continue d'une classe à l'autre :
  temps clair, variable ou couvert avec h heures d'insolation.

On pourra se ramener à un traitement en phénomène accidentel soit quand la "variable" qui le caractérise est dichotomique, discrète ou ordinale, soit, si elle est continue, dès qu'elle a une distribution discontinue, avec point d'accumulation, ou simplement plurimodale.

(b) Le but final étant de prévoir ce phénomène, à l'aide d'autres variables plus aisément accessibles ou prévisibles, une technique courante en hydrométéorologie est la régression multiple. Malheureusement celle-ci fait l'hypothèse sous-jacente d'une relation continue, voire linéaire, entre la variable à expliquer et les variables explicatives. Cela suppose que ce sont toujours les mêmes mécanismes physiques qui interviennent.

Exemples

Plus les températures de l'air sont élevées, plus il y a de fonte dans un manteau de neige.

Or cette relation n'est plus vraie quand on passe en-dessous de 0°C, où d'autres mécanismes interviennent.

De même, on peut chercher à prévoir des températures de l'air à partir de champs météorologiques de pression, mais en présence de fronts, le fait de se trouver dans la masse d'air chaud ou froid est dominant.

La régression s'accommode éventuellement de certaines non-linéarités (régression polynomiale, transformation de variables) et accepte même les variables discrètes (mais ordinales).

Ce n'est souvent plus le cas quand le phénomène met en jeu des mécanismes complètement différents selon la valeur des variables explicatives. Même un codage discret devient arbitraire s'il n'y a plus de relation d'ordre.

(Exemples : temps clair/nuageux/pluvieux/brouillard).

Il est alors préférable d'utiliser les techniques d'analyse discriminante. De même dans le cas de fortes non-linéarités, le codage en modalités disjointes et leur traitement en analyse discriminante peut être préférable.

(c) On trouvera dans la thèse de Ph. BOIS (1976) le traitement dans le cas de 2 groupes, en général occurence ou non occurence que nous ne ferons que rappeler.

Nous développerons plutôt le cas de k groupes ($k > 2$) problème qui peut apparaître pour diverses raisons, même quand seule l'occurence ou la non-occurence d'un phénomène nous intéresse. Le cas le plus fréquent est celui où le phénomène est un effet, pouvant avoir plusieurs causes radicalement différentes.

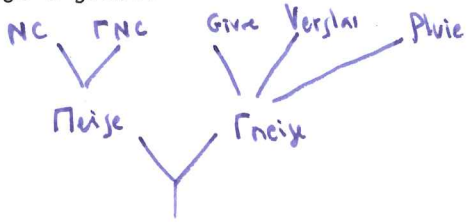
Exemple : Occurence Pluie/Non-Pluie.

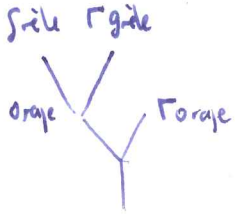
Le beau temps est en général lié à une situation anticyclonique, tandis que le déclenchement des précipitations peut être dû soit à une perturbation cyclonique, soit à une situation instable donnant des pluies convectives. Il peut donc être intéressant de traiter les 3 phénomènes de façon distincte, quitte à en regrouper certains par la suite.

(d) Un autre problème est le cas des phénomènes mettant en jeu divers mécanismes successifs, et présentant une structure arborescente, d'où une notion de niveaux successifs qui devraient apparaître dans le traitement.

Exemple :

Une situation météorologique est propice ou non à la formation d'orages, et un orage peut engendrer ou non des chutes de grêle ; mais il faut qu'il y ait orage pour que l'on ait orage à grêle.





Mais il n'est pas toujours évident que le phénomène soit à plusieurs niveaux (ex : l'occurence de pluie), et même s'il l'est, qu'il puisse être "perçu" comme tel par l'ensemble des variables dont nous disposons.

Nous vérifierons seulement, dans les exemples du chapitre III si l'introduction de plusieurs niveaux améliore ou non les performances.

Nous supposerons dans ce chapitre que la population est répartie en différents groupes connus a priori, et nous envisagerons au chapitre II le cas où ils ne le sont pas.

I.2. Analyse discriminante linéaire dans le cas de 2 groupes.

T matrice de var.covariance
Données centrées

On la trouvera exposée de façon détaillée dans BOIS Ph. (1976) ou ROMEDER (1973). Elle comporte 2 approches, l'une géométrique, l'autre probabiliste, qui en fait se rejoignent.

Dans la première, on définit une distance de Mahalanobis entre les groupes : $D^2_T = (Y_2 - Y_1)^t . T^{-1} . (Y_2 - Y_1)$ ou $D^2_W = (Y_2 - Y_1)^t . W^{-1} . (Y_2 - Y_1)$

(selon que l'on choisit les métriques T ou W, reliées par T = W+B) et on calcule la distance d'un individu à un groupe par : $d^2_T(X, Y_1) = (X - Y_1)^t T^{-1} (X - Y_1)$

dans laquelle le terme $X^t . T^{-1} . X$ ne dépend pas du groupe. On vérifie alors que l'affectation au groupe 1 plutôt qu'au groupe 2 :

$$d^2_T(X, Y_1) < d^2_T(X, Y_2) \iff (Y_1 - Y_2)^t . T^{-1} . X > \frac{1}{2} (Y_1 - Y_2)^t T^{-1} (Y_1 + Y_2)$$

Le terme de gauche est appelé fonction discriminante : $F_T(X) = (Y_1 - Y_2)^t . T^{-1} . X$

On vérifie que la forme linéaire $T^{-1} . (Y_1 - Y_2)$ est vecteur propre de $T^{-1}B$ associée à λ_1, pouvoir discriminant. On vérifie aussi que $\lambda_1 = \frac{n_1 . n_2}{N} . D^2_T$ ou compte tenu des relations entre T et W, que la valeur propre de $W^{-1} . B$, $\mu_1 = \frac{\lambda_1}{1 - \lambda_1} = \frac{n_1 . n_2}{N^2} . D^2_W$

Le vecteur propre $T^{-1}B$ ou $W^{-1}B$ définit l'axe discriminant, et on utilise aussi la cloison séparatrice, lieu des points équidistants au sens de D^2_W, des 2 centres de groupes (c'est la direction conjuguée de la droite des centres).

Enfin, l'approche probabiliste consiste à associer aux 2 populations un modèle normal qui permet de raisonner non en terme de distance mais de probabilité d'appartenance. On le reverra dans le paragraphe suivant.

Rappelons aussi l'analogie, dans le cas de 2 groupes entre la discrimination et la regression sur une variable qualitative 0-1, qui permet entre autres d'interpréter les coefficients de la fonction discriminante en terme de corrélations partielles.

1.3 - Analyse discriminante linéaire multigroupe

Cette analyse se rattache, pour ses aspects décisionnels, au cas de 2 groupes, mais évidemment aussi à l'analyse factorielle discriminante sur plusieurs groupes.

L'hypothèse la plus critique pour cette approche est celle de l'égalité des matrices de variance covariance pour tous les groupes considérés.

I.3.1. Fonctions discriminantes et cloisons séparatrices

(a) L'approche géométrique est immédiate. La distance à un groupe quelconque k s'écrit :

$$d^2(X, Y_k) = (X - Y_k)^t W^{-1} (X - Y_k)$$

et l'affectation au groupe k plutôt qu'au groupe j conduit à comparer :

$$Y_k^t W^{-1} X + \frac{1}{2} Y_k^t W^{-1} Y_k \leqslant Y_j^t W^{-1} X + \frac{1}{2} Y_j^t W^{-1} Y_j$$

Par analogie avec le cas de 2 groupes, où l'on appelait fonction discriminante la quantité :

$$(Y_k - Y_j)^t . W^{-1} . X$$

que l'on comparait à un terme constant, on est conduit ici :

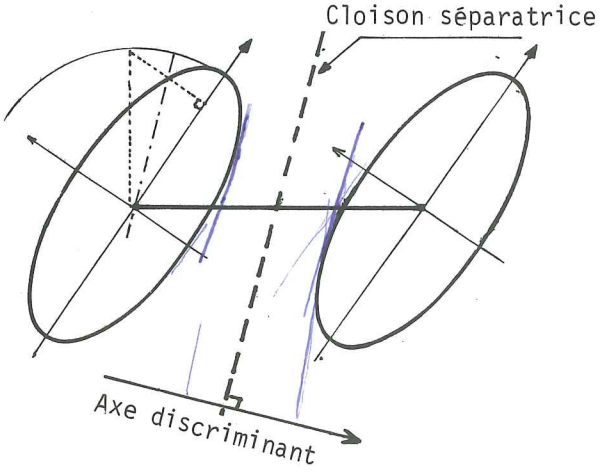
- à définir une fonction discriminante pour chaque groupe dans laquelle est inclus le terme constant :

$$Y_k^t . W^{-1} X + \frac{1}{2} Y_k^t . W^{-1} . Y_k$$

- et à effectuer les comparaisons des groupes 2 à 2.

Il est aisé de définir les cloisons séparatrices entre 2 groupes. C'est toujours l'hyperplan conjugué (au sens de W^{-1}) de la droite des centres. Cela donne le schéma suivant :

a) Rappel, dans le cas de 2 groupes, la construction de la séparatrice .



b) Cas de G (> 2) groupes .

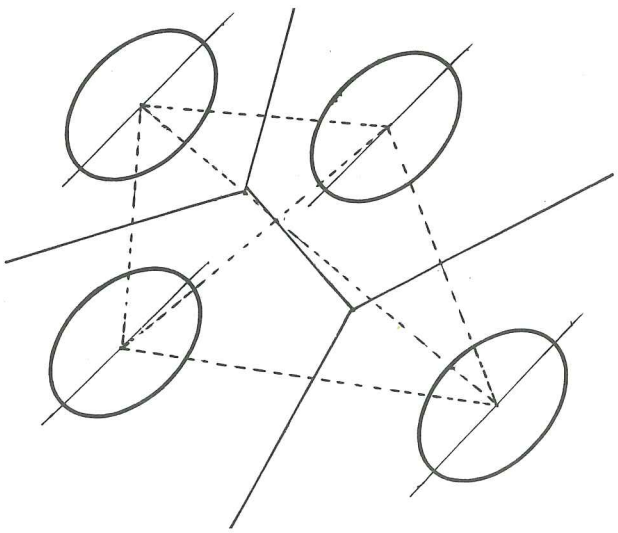


FIGURE V - 1 : Cloisons séparatrices .

Les combinaisons 2 à 2 donnent les régions respectives d'affectation à chaque groupe. Au delà de 3 groupes ceux-ci ne sont plus nécessairement dans un plan et on ne peut plus représenter ces cloisons en projection plane.

Note. Il y a un petit point litigieux dans ROEMEDER qui dit que les séparatrices sont les médiatrices, donc orthogonales à la droite des centres. C'est faux en général sauf si on se place dans un espace W^{-1} transformé. De même l'axe discriminant n'est pas la droite des centres en général, mais la perpendiculaire à la séparatrice.

(b) La présentation probabiliste complète l'approche géométrique et permet d'introduire les notions de probabilités a priori et de coût d'erreur utilisées par les règles de décision bayesiennes.

On comparera en général, pour affecter au groupe k, l'espérance du coût si on met X dans C_k à l'espérance du coût si on le met dans tout autre groupe C_j :

$$(1) \quad \sum_{\substack{l=1 \\ l \neq k}}^{G} \pi_l \cdot P_l(X) \cdot C(k|l) \overset{?}{\leqslant} \sum_{\substack{l=1 \\ l \neq j}}^{G} \pi_l \cdot P_l(X) \cdot C(j|l)$$

π_l, $P_l(X)$??

Dans le cas particulier où les coûts d'erreurs sont égaux, la condition $\sum_{l=1}^{G} \pi_l \cdot P_l(X) = 1$ montre qu'il est équivalent d'effectuer l'affectation en comparant :

$$\pi_k \cdot P_k(X) > \pi_j \cdot P_j(X)$$

L'hypothèse multinormale permet, en prenant les logarithmes et en éliminant les termes identiques, de comparer :

$$F_k(X) = Y_k^t W^{-1} X - \frac{1}{2} Y_k^t W^{-1} Y_k + \text{Log } \pi_k$$

avec $F_j(X)$... etc.

Ce qui est encore une fonction linéaire. C'est en général l'affectation proposée par les programmes courants.

Par contre si les coûts d'erreurs sont différents on ne peut plus prendre les logarithmes de la somme dans l'inégalité (1), sauf si l'on se réduit à 2 groupes. Mais il existe encore des simplifications même dans ce cas (cf ULMO et BERNIER, 1973, p.54).

1.3.2. Distance de MAHALANOBIS généralisée - Sélection de variables

Il ne s'agit pas en fait de la généralisation de la notion de distance mais de celle d'inertie. En effet, on a vu dans l'analyse en C.P. (IIème Partie, Chap.I.1.1) qu'une des interprétations de cette analyse consistait à définir l'inertie du nuage des individus et à chercher les axes qui permettaient d'extraire successivement le maximum d'inertie.

On sait aussi que si on a p axes orthogonaux $Z_1, Z_2 \dots Z_p$, l'inertie par rapport au barycentre G est proportionnelle à la somme des inerties des axes successifs :

$$(P-1)\, I_{/G} = I_{/GZ_1} + I_{/GZ_2} + \dots + I_{/GZ_P}$$

Dans notre cas, où l'on assimile les groupes à des points pesants Y_k de masse $\frac{n_k}{N}$, on peut définir les distances entre les groupes pris 2 à 2

$$d^2(Y_k, Y_j) = (Y_k - Y_j)^t T^{-1} (Y_k - Y_j)$$

mais où la masse du groupe n'intervient pas, et la somme de ces interdistances n'a pas de sens (contrairement à l'ACP) dans la mesure où les groupes ont des masses différentes.

Par contre, on peut définir l'inertie du système de points pesants par :

$$I = \sum_{k=1}^{G} \frac{n_k}{N} (Y_k - Y_.)^t T^{-1} (Y_k - Y_.) \sim D_G^2$$

C'est ce que l'on appelle encore la distance de MAHALANOBIS généralisée (à un facteur près éventuellement).

D'ailleurs l'analyse factorielle discriminante n'est autre qu'une ACP des centres de groupes Y_k affectés de leurs masses $\frac{n_k}{N}$ et consiste à chercher des axes présentant successivement le plus possible d'inertie. Comme ils sont orthogonaux, la somme de leurs inerties est proportionnelle à I.

Or on sait calculer leurs inerties, puisque cette ACP conduit à calculer les valeurs propres λ_j de $T^{-1}.B$, dont la somme n'est autre que la trace.

On a donc la proportionnalité :

$$D_G^2 \sim \text{Trace}\,(T^{-1}.B)$$

Ceci est utilisé en sélection de variables pour choisir le sous-ensemble de P variables qui maximise l'inertie du nuage des Y, donc la discrimination.

On sait aussi que maximiser Trace $(T^{-1}.B)$ revient aussi à maximiser Trace $(W^{-1}.B)$ et l'on définit généralement :

$$D_G^2 = (N - 1 - G.P).\text{Trace}\,(W^{-1}.B)$$

qu'il faut maximiser. L'accroissement $D_G^2(P+1) - D_G^2(P)$ suit une loi de χ^2 qui permet de tester sa signification.

Un des intérêts de cette relation entre D_G^2 et la trace est de montrer le rôle des effectifs n_k dans la sélection des variables.

Dans l'approche probabiliste bayesienne, on suppose connus les centres de classes, les dispersions des classes et leur probabilité a priori. L'échantillon disponible sert en général à estimer les 2 premiers mais pas nécessairement les probabilités a priori. Et les effectifs des classes ne jouent normalement que sur la

qualité des estimations des paramètres de ces classes.

Par contre, on constate ici que quand on choisit le sous-espace des variables les "plus" explicatives à l'aide du critère D^2_G, les effectifs interviennent et on a tendance à privilégier les variables qui discriminent bien les groupes présentant un fort effectif dans l'échantillon (ce, quelque soit leur probabilité a priori réelle). Il faut donc prendre garde à ce biais. (cf MILLER R.G., 1962 et D.G. DE COURSEY, 1970).

C'est la raison pour laquelle d'autres critères ont été utilisés, comme celui de BMD07M décrit en détail par ROMEDER (p.82), qui teste l'égalité des moyennes conditionnelles, ou ce qui revient au même, l'égalité des matrices T et W

I.3.3. Problèmes en analyse discriminante linéaire (*)

Compte tenu du nombre d'hypothèses sous-jacentes assez fortes, les problèmes qui se posent en analyse discriminante linéaire sont nombreux. Nous les évoquons sommairement, pour ne développer que ceux qui sont particuliers à nos applications.

On peut donc considérer :

(a) Le rôle de l'échantillonage dans la détermination des moyennes des différentes populations et dans celle de la matrice de dispersion W. En ce qui concerne les moyennes, on sait que l'écart-type sur une composante est en $1/\sqrt{n_k}$ (pour une population normale) mais le nombre d'individus par groupe peut être très différent d'un groupe à l'autre. En général, il est préférable d'avoir des effectifs comparables (cf T. CACOULOS, in CACOULOS, 1972).

On trouvera dans T.W. ANDERSON (1958) et dans l'ouvrage dirigé par CACOULOS (1972) (en particulier dans la synthèse bibliographique de S. DAS GUPTA) des considérations sur l'échantillonage des fonctions discriminantes et sur l'estimation des probabilités de mauvaise classification pour diverses situations d'échantillonage.

(b) La non-satisfaction d'hypothèses comme celle de l'égalité des matrices de variances-covariances. Celle-ci peut être validée par divers tests (BARTLETT ou BOX). Dans le cas où elle est trop fortement rejetée, on peut utiliser des techniques quadratiques (cf § I.4).

Par contre les hypothèses de normalité ne sont pas trop critiques dans la mesure où le test sur la distance de MAHALANOBIS qui se ramène à un T^2 de HOTTELING est relativement robuste vis-à-vis de ces hypothèses (cf K.V. MARDIA, 1975). Il est clair cependant que l'assymétrie est plus gênante que l'aplatissement.

Dans le cas où les distributions ne peuvent pas être considérées comme normales même très grossièrement, on peut chercher à développer des techniques particulières (cf I.4).

(c) Un problème souvent rencontré est celui de la sélection des variables. Outre les problèmes classiques des stratégies de sélection (ascendantes, descendantes,

avec ou sans remise en cause des variables introduites, et par des méthodes non paramétriques comme celle de l'échantillon-test ROMEDER (1973)) il faut bien voir le rôle particulier de l'échantillon d'ajustement utilisé.

La position respective des groupes dans l'espace $\mathbb{R}^p$ (colinéarité ou au contraire, aux sommets d'un polyëdre quasi-régulier) peut influencer sensiblement le nombre de variables à utiliser. C'est aussi le cas où l'on a des phénomènes quasi "orthogonaux" 2 à 2 : La description de chaque relation nécessite certaines variables inutiles vis-à-vis des autres groupes, mais qui une fois réunies, constituent un ensemble de variables trop nombreux pour être ajusté de façon robuste.

Comme en régression, on considère qu'il faut au maximum introduire r variables avec $2r < N$ nombre total d'individus. Il s'agit là d'une règle empirique, valable en discrimination linéaire.

ROMEDER (1973) propose une théorie,et a effectué des simulations,dans le cas de 2 classes équiprobables qui conduisent à des valeurs de N généralement supérieures à $2r$, pour introduire r variables et garantir (à un niveau donne de probabilité) que la discrimination obtenue n'est pas le fruit du hasard.

Seuil 5% $\quad N \sim 2{,}25\ r + 9$

1% $\quad N \sim 2{,}35\ r + 12$ pour $\begin{cases} N < 1000 \\ r < 50 \end{cases}$

Par contre, le cas où $N = N1 + N2$ avec N1 et N2 très différents, de même que le cas de plusieurs groupes n'a pas été tellement abordé dans la littérature.

On peut cependant imaginer que prenant les groupes 2 à 2, on veut éviter que la frontière entre 2 groupes, parmi k, soit purement aléatoire. Dans ce cas il faudra au moins que l'effectif n_k supposé le même pour tous les groupes soit

$$n > r \qquad (\sim 1{,}15\ r + 5 \text{ ou } 6)$$

Par contre, si on admet que certains groupes ne soient pas séparés mais qu'il y ait au moins 2 grands paquets séparés de façon significative, on pourra aller jusqu'à r tel que :

$$\frac{N}{2} > r \qquad N = \text{nombre total}$$

En fait, on sera parfois conduit à introduire pour G groupes un nombre r de variables dont on sait qu'il est excessif pour discriminer entre 2 groupes donnés pris parmi les G .

Il faut d'ailleurs insister aussi sur la sensibilité de certaines méthodes de sélection aux effectifs respectifs des différents groupes. La sélection utilisant la distance généralisée de MAHALANOBIS y est en particulier très sensible. On a vu que cette distance est en fait l'inertie d'un nuage de points ayant pour masses les effectifs des échantillons d'ajustement.

Les effectifs importants y sont donc une contribution plus importante, et chercher les variables qui maximisent cette inertie revient à avantager, ou à discriminer préférentiellement, entre les groupes de fort effectif .

Or les effectifs des différents groupes peuvent n'avoir aucune signification statistique et ne provenir que de contraintes matérielles (disponibilité, facilité

de mesure, etc...). Même si ces effectifs sont représentatifs des probabilités a priori des diverses populations, on peut vérifier que ce biais n'a pas lieu d'être, et que ce n'est pas l'endroit où il faut les prendre en compte. En effet, au niveau de l'estimation des densités $f_j(X)$ cela conduit à introduire des variables assez peu discriminantes pour les groupes à faibles effectifs et donc des densités assez proches de celles des groupes forts.

Ensuite, l'allocation à un groupe par :

$$\pi_k . f_k(X) \geq \pi_j f_j(X) \quad \forall j \neq k$$

favorisera encore les groupes à forts effectifs.

Or il est courant que ce soit les phénomènes relativement rares qui soient les plus intéressants à prévoir, d'où la nécessité de rendre les effectifs comparables au niveau de l'échantillon d'ajustement.

La méthode non paramétrique de l'échantillon-test souffre du même inconvénient si l'échantillon-test n'est pas correctement équilibré (même s'il n'est plus alors représentatifs de la population globale en terme de probabilités a priori).

Par contre, dans la méthode utilisée par BMD07M (test de la dispersion des moyennes conditionnelles, les effectifs des échantillons d'ajustement n'interviennent pas.

Remarque : On trouvera dans NAKACHE et DUSSERRE (1975) une étude comparative, dans le cas de 2 groupes, qui conclut à un léger avantage de la méthode utilisée par le programme BMD07M.

(d) Comme pour les analyses en corrélation entre 2 variables continues, on peut se demander ce que valent les estimateurs utilisés quand les échantillons utilisés ne sont pas constitués d'individus indépendants mais d'observations effectuées en séquence dans le temps et plus ou moins autocorrelées (en fonction de la fréquence d'observation par rapport à l'inertie du phénomène).

Les travaux sur ce sujet, bien qu'assez dispersés, reviennent tous à définir un nombre équivalent n_e d'observations indépendantes. Malheureusement, celui-ci est souvent associé à 1 paramètre particulier, par exemple le coefficient de corrélation simple entre 2 séries, ou les coefficients de régression partielle: on considère alors la formule donnant la variance théorique de cet estimateur en fonction de n observations indépendantes. Comme la variance vraie est en général plus grande quand les données sont autocorrelées, on cherche le nombre n_{eq} d'observations indépendantes qui auraient donné cette variance vraie en utilisant la même formule de variance théorique, soit n_e.

On trouvera des détails dans EZECHIEL et FOX (1970) , et une généralisation, pour la corrélation multiple, dans MARCHENKO et MINAKOVA (1972 cités par BOIS 1976).

Ces schémas peuvent à la rigueur s'appliquer à l'analyse discriminante à 2 groupes dans la mesure où elle se ramène à une régression sur une variable Y valant

0 ou 1. Malheureusement, on ne s'intéresse en général pas à la variance résiduelle de Y mais à d'autres paramètres comme la probabilité de mauvais classement ou la variance de la fonction discriminante (estimées sur des échantillons autocorrelés).

De plus, il est rare que les échantillons utilisés en discrimination soient des séries chronologiques séquentielles, contrairement à la régression.

Exemple 1 : On a des observations de p variables et on sait qu'il y a eu un changement brutal dans le temps donc 2 populations. On prend donc pour l'ajustement 1 année complète avant la discontinuité, et 1 année après et on cherche a réaffecter les observations, pour l'année de la discontinuité, à l'une ou l'autre des populations.

Dans un cas comme celui-ci, et avec des hypothèses très fortes, BASU et ODELL (1974) montrent que la variance d'échantillonage de la fonction discriminante :

$$F(X) = (Y_1 - Y_2)^t W^{-1} \left(X - \frac{1}{2}(Y_1 + Y_2)\right)$$

est multipliée par $\frac{1}{(1+\rho)^2}$ dans le cas où :

- les variables ont toutes même autocorrélation $\rho = \rho_{x_1} = \rho_{x_2} = \quad \forall x_j$
- selon un schéma équicorrelé, c'est-à-dire quelque soit l'ordre :

$$\rho_1 = \rho_2 = \cdots \rho_k = \rho \quad \forall k$$

Ceci est peu réaliste mais s'explique par la complexité du cas markovien, plus proche de nos préoccupations.

Ces hypothèses ne sont cependant pas représentatives des problèmes habituels d'analyse discriminante en météorologie, où l'on a plutôt des séquences, plus ou moins longues mais indépendantes, de journées appartenant à l'une ou l'autre des populations.

Exemple 2 : Les journées de verglas viennent par séquence de 2 ou 3 jours, où les conditions sont très voisines. De même pour les journées avec précipitation, ou brouillard (séquence de 2, 3 voire 5 ou 6 jours). A la limite chaque séquence ne représenterait qu'un seul individu et on a donc une borne inférieure pour le nombre équivalent de journées en occurence. Comme ces phénomènes sont relativement rares, on dispose par contre souvent d'un fichier pléthorique pour les cas de non-occurence. On peut alors éliminer l'autocorrélation par échantillonage au hasard dans les cas disponibles.

Enfin, bien que très partiels, les résultats publiés ne concernent que l'analyse discriminante à 2 groupes et non pas le cas de plusieurs groupes avec des probabilités a priori et des autocorrélations internes différentes.

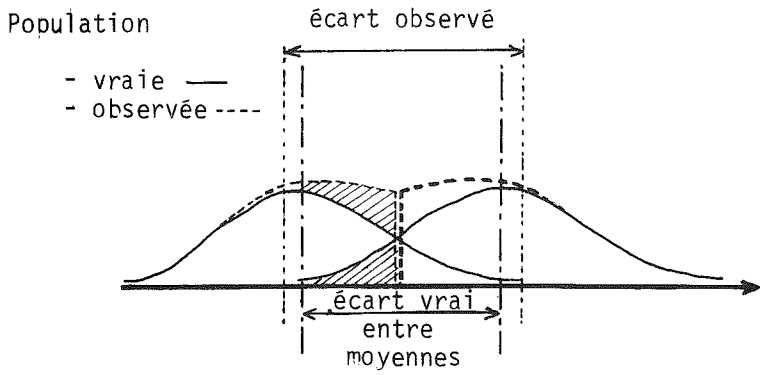
(e) Un dernier problème, fréquent en hydrométéorologie, est celui des erreurs dans les fichiers d'observations et de leur influence sur les paramètres du modèle ajusté et ses performances en prévision.

Si l'on se ramène au cas simple de la régression entre Y et X et si l'on suppose que les y_i sont affectés d'erreurs purement aléatoires, on peut vérifier que les coefficients de régression ne changent pas même si le coefficient de corrélation diminue (J. JOHNSTON, 1963). La même chose se produit en analyse discriminante à 2 groupes si on se trompe, de la même façon aléatoire, sur la variable d'appartenance à un groupe.

En effet, on réduit les distances entre les centres de groupe, on accroît les variances intra-groupe, mais les cloisons séparatrices restent inchangées. On montre de plus que les probabilités de mauvais classements restent inchangées, c'est-à-dire n'augmentent pas (P.A. LACHENBRUCH, 1974).

On constate que cela n'est plus vrai dès que le nombre de groupes atteint 3 sauf dans des conditions très particulières, car la configuration des centres de groupes intervient.

De même quand les erreurs ne sont pas aléatoires (par exemple quand l'observateur qui décide de l'affectation à C_1 ou C_2 utilise déjà implicitement les variables X_j qui seront utilisées dans le modèle : il y a alors tendance à négliger les extrémités des distributions du côté de la cloison séparatrice). Ceci conduit par exemple à écarter les moyennes, réduire les variances et, si la cloison séparatrice est inchangée ici (dans le cas de 2 groupes), les probabilités de mauvais classement estimées sur cet échantillon sont irréalistes car trop optimistes.



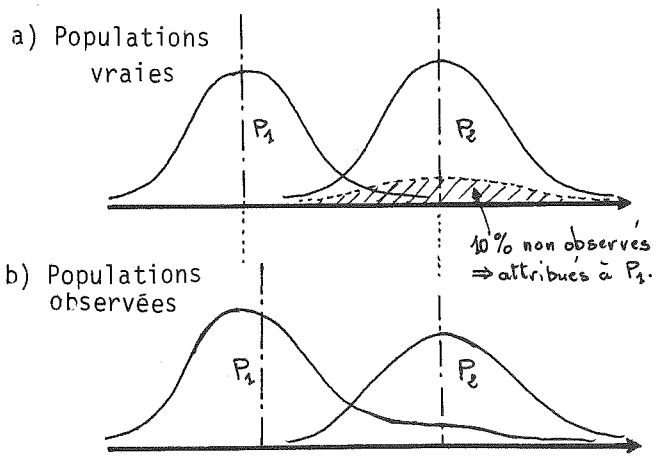
P.A. LACHENBRUCH donne des exemples de simulation dans le cas de 2 groupes, avec les types d'erreur décrits ci-dessus. Le cas de $G > 2$ groupes ne semble pas avoir été abordé.

D'autre part, le cas que l'on peut rencontrer en hydrométéorologie est celui où le biais (erreur d'observation) est toujours dans la même population.

Exemple : On peut facilement sélectionner des journées sans avalanche, par contre il peut arriver qu'une avalanche ne soit observée que le lendemain (et jamais la veille !) de son déclenchement, or cette journée, classée avalancheuse est en fait une journée normale. L'inverse est beaucoup moins probable (journée avec avalanche mais non observée, et retenue dans l'échantillon des journées normales).

Par contre, dans une application sur les orages à grêle, où l'on cherche à comparer des orages et des orages à grêle, c'est l'échantillon des orages sans grêle qui peut contenir des cas où la grêle a eu lieu mais n'a pas été signalée.

Ce type d'erreur va biaiser une seule population et si on suppose par exemple 2 populations $\mathcal{C}_1$ et $\mathcal{C}_2$ de mêmes densités de probabilité vraies, mais où l'échantillon de C_2 contient 10% de cas attribués par hasard à C_1 ; on voit que cela va biaiser la position de la cloison séparatrice et modifier le calcul de W. L'utilisation en prévision va donc augmenter les erreurs de classification des individus de C_2 et si les probabilités a priori sont déduites de l'échantillon, π_2 est sous estimée et π_1 surestimée ce qui défavorise encore plus C_2.



On pourrait envisager une procédure d'estimation qui prenne en compte pour chaque individu $X_i^{(k)}$ une certaine probabilité qu'il appartienne en fait à une autre classe que C_k.

		C_1	C_2	---	C_j	---	C_k	---	C_G
Pour une donnée sûre, on aurait :	$P_{jk}(i)$	0	0		0		1		0
Pour une journée douteuse :	$P_{jk}(i)$	.1	.0		.2		.6		.1

avec $\sum_j P_{jk}(i) = 1$

On trouvera dans S.J. PRESS (1968) une méthode qui pourrait s'étendre au cas envisagé ci-dessus. Cela revient par exemple à faire figurer les individus dans chaque population, mais pondérés proportionnellement à leur probabilité, et le problème est d'estimer ces P_{jk}.

I.4 - Analyse discriminante non linéaire, ou dans des hypothèses non normales

I.4.1. Cas où les lois ne sont pas multinormales

Il est fréquent que l'on ne puisse admettre que 2 populations suivant des lois multi-normales, même si elles ne diffèrent que par leurs moyennes. Bien que la méthode linéaire soit robuste vis-à-vis des hypothèses de normalités (K.V. MARDIA, 1975) on peut tenter d'adapter la méthode à des lois différentes, si elles sont connues.

C'est ce qu'on fait CHLIKARA et ODELL (1973) pour des lois de type exponentiel $f^{(r)}(X) = K_r e^{-c\|X\|_r}$ où la norme $\|X\|_r = \sum_{j=1}^{p} |X_j|^r$

Une autre méthode, plus générale encore est l'analyse logistique développée par COX D.R. (1968) et J.A. ANDERSON (1972) . L'hypothèse de base est que seul le rapport des densités des diverses populations interviennent, et qu'il peut se mettre sous la forme exponentielle :

$$\frac{\pi_i f_i(X)}{\pi_k f_k(X)} = e^{-\beta_{ik}^t \cdot X}$$

On trouvera d'autres aperçus dans l'analyse bibliographique de DAS GUPTA (1973 in T. CACOULOS) mais nous n'avons pas utilisé ces techniques.

I.4.2. Analyse discriminante quadratique

Même si les populations considérées ne sont pas trop éloignées de l'hypothèse normale, une autre hypothèse critique dans les méthodes linéaires classiques est l'égalité des matrices de covariance W_k (k = 1 ... G) (à P variables).

On peut tester cette égalité à l'aide d'un test dû à BOX (cité dans D.F. MORRISON , 1967 p.153). Si on dispose de G échantillons de tailles respectives N_k (k = 1, G), on calcule :

$$W = \sum_{k=1}^{G} \frac{n_k . W_k}{\sum n_k}$$

W_k = matrice de covariance estimée sur l'échantillon k

$n_k = N_k - 1$

BOX a montré que la quantité $M . C^{-1}$

$$M = \left(\sum_{k=1}^{G} n_k\right) . \text{Log}|W| - \sum_{k=1}^{G} \left(n_k . \text{Log}|W_k|\right)$$

$$C^{-1} = 1 - \frac{2p^2+3p-1}{6(p+1)(G-1)} \left(\sum_{k=1}^{G} \frac{1}{n_k} - \frac{1}{\sum n_k} \right)$$

suivant une loi de χ^2 ou F de FISHER (selon les valeurs de p, G, n_k)....

En fait, on applique rarement ce test, et on évite les méthodes quadratiques, qui nécessitent le calcul de nombreux paramètres, sauf si la non-égalité des matrices est flagrante.

Cela se voit en général sur les nuages de points projetés dans les premiers axes factoriels discriminants.

On peut alors utiliser 2 approches :

(a) Dans le cas où l'on garde l'hypothèse multinormale on peut encore tracer des cloisons séparatrices mais qui seront des quadratiques. La règle d'affectation entre 2 populations 1 et 2 s'écrira :

$$\pi_1 . f_1(X) > \pi_2 . f_2(X)$$

$$\frac{\pi_1}{(\sqrt{2})^p |W_1|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(X-\gamma_1)^t W_1^{-1}(X-\gamma_1)\right] \gtrless \frac{\pi_2}{(\sqrt{2})^p |W_2|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(X-\gamma_2)^t W_2^{-1}(X-\gamma_2)\right]$$

d'où, en passant aux logarithmes :

$$(X-\gamma_1)^t W_1^{-1}(X-\gamma_1) - \text{Log}|W_1| - 2\,\text{Log}\,\pi_1 \gtrless (X-\gamma_2)^t W_2^{-1}(X-\gamma_2) - \text{Log}|W_2| - 2\,\text{Log}\,\pi_2$$

Dans le cas où les probabilités a priori sont égales, on constate que cela revient à comparer des distances de la forme $(X-\gamma_k)^t W_k^{-1}(X-\gamma_k)$ mais où apparait en plus la constante $\text{Log}|W_k|$ qui est directement liée au choix d'une loi normale.

On voit immédiatement les problèmes d'échantillonage que pose une telle approche, puisqu'il faut estimer non plus 1 mais G matrices de covariance. D'autre part, dans le cas où 1 variable a une variance nulle dans un groupe, ce qui la rend en général très discriminante, on ne peut plus inverser W dans ce groupe, et il faut l'éliminer, ce qui est un peu choquant, ou la traiter au préalable.

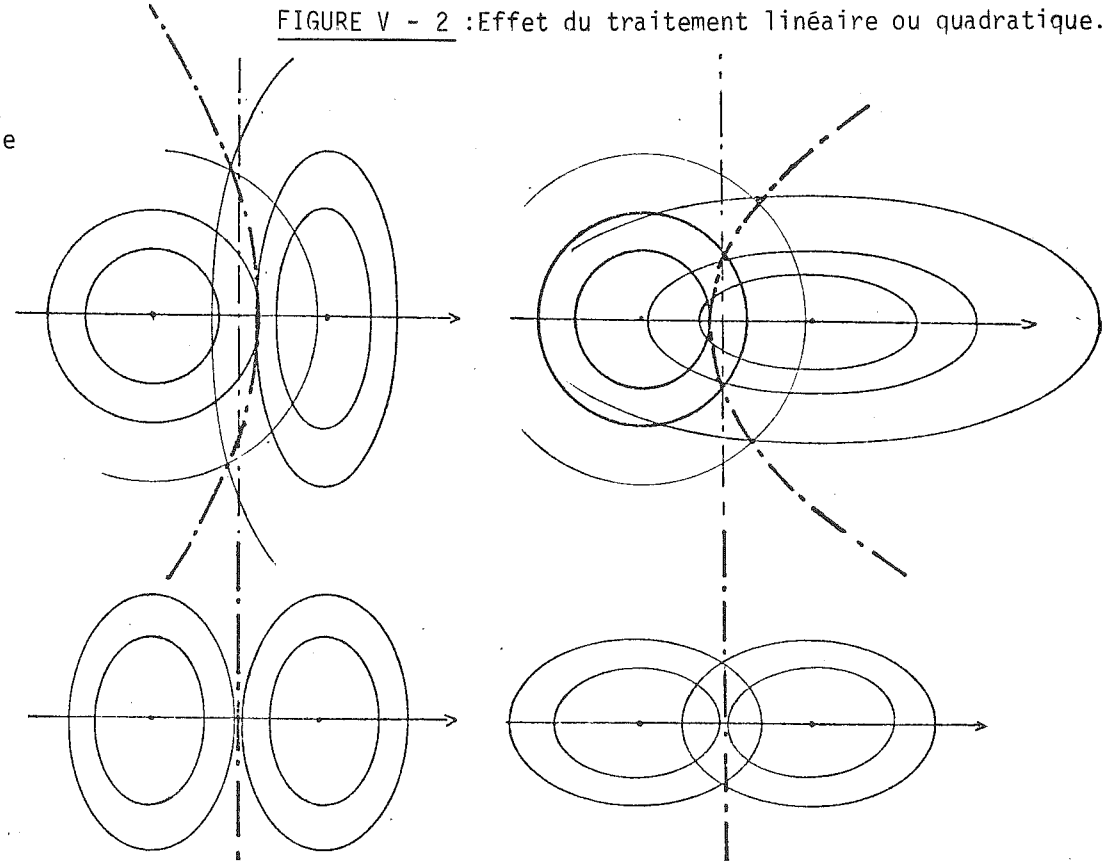
Enfin, il faut bien voir que même si l'hypothèse de non-égalité des matrices de covariance est vérifiée, le traitement quadratique n'améliore les performances par rapport au schéma linéarisé que plus ou moins selon la disposition des groupes (cf.Fig.V-2)

On voit qu'il y a un effet très différent selon les positions relatives des axes principaux locaux des populations par rapport aux axes discriminants globaux de l'ensemble des populations.

FIGURE V - 2 :Effet du traitement linéaire ou quadratique.

a) Traitement quadratique

b) Traitement linéaire



Cas où les performances diffèrent assez peu

Cas où les performances diffèrent sensiblement

(b) Une autre approche due à SEBESTYEN G.S. (1962) et reprise dans ROMEDER J.M. (1973) se veut non-paramétrique et procède comme suit :

1 - Définitions:

Distance ou similitude d'un individu à un groupe C_k , dans une métrique $Q_k = T_k^t . T_k$ donnée :

$$S(x, C_k) = \frac{1}{n_k} \sum_{i=1}^{n_k} d^2(x, X_i^k (\in k))$$

similitude d'un individu X_ℓ avec son groupe :

$$S(X_\ell^k, C_k) = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} d^2(X_\ell^k, X_i^k) \qquad i \neq \ell$$

aggrégation du groupe C_k :

$$D_k^2 = \frac{1}{n_k} \sum_{\ell=1}^{n_k} S(X_\ell^k, C_k) = \frac{1}{n_k (n_k - 1)} \sum_{\ell} \sum_{i \neq \ell} d^2(X_\ell^k, X_i^k)$$

2 - Optimisation :

Recherche de la matrique Q_k qui maximise l'aggrégation du groupe C_k.

Comme il faut bien une contrainte sur Q (sinon on prend par exemple une matrice diagonale de terme σ_{ii} très petit, tendant vers 0 !) SEBESTYEN impose :

$$\det(Q_k) = |Q_k| = 1$$

et démontre que la matrique optimale est :

$$Q_k = |W_k|^{\frac{1}{p}} . W^{-1}$$

Remarque : On peut s'étonner du choix de la contrainte dans SEBESTYEN. En effet, la contrainte essentielle est de prendre $\det Q$ = cste. Si on choisit $\det Q = \det W^{-1}$ on retombe alors sur la m trique $Q = W^{-1}$ de MAHALANOBIS sans coefficient multiplicatif parasite. Par contre, avec $\det Q = 1$ on trouve $Q = |W|^{\frac{1}{p}} . W^{-1}$

Dans le cas d'une variable cela remplace la distance au centre du groupe : $D^2(x, y_1) = \frac{(x - y_1)^2}{\sigma^2}$

par la similitude : $S(x, y_1) = (x - y_1)^2 + \sigma_1^2$

dont l'interprétation est sensiblement différente.

L'intérêt de la transformée à volume constant n'est pas évident dès que l'on change les axes (on fait des rotations) et surtout on fait des homothéties sur les axes.

La similitude d'un nouvel individu x à un groupe s'écrit ensuite :

$$S(x, P_k) = \frac{1}{n_k} \cdot |W_k|^{\frac{1}{P}} \cdot \sum_{i=1}^{n_k} (x - X_i^k)^t W_k^{-1} (- X_i^k)$$

dont on montre qu'elle se ramène à :

$$S(x, P_k) = |W_k|^{\frac{1}{P}} \cdot [P + (X - Y_k)^t W_k^{-1} (X - Y_k)]$$

L'affectation se fera alors selon le critère :

$$X \in P_k \quad \text{si} \quad S(X, P_k) = \min_j S(X, P_j) \quad j \neq k$$

On constate que ce critère ne recoupe pas exactement le précédent, mais que l'un comme l'autre se ramène à la discrimination linéaire classique quand $W_k = W \ \forall k$

Le rôle des constantes, surtout de $|W_k|^{\frac{1}{P}}$ apparaît surtout dans l'aggrégation (qui augmente si $|W|$ diminue) mais on voit que, toutes choses égales par ailleurs, la similitude de x avec P_k diminue si P augmente. Malheureusement, cet effet de la dimensionnalité est très liée aux choix qu'a faits SEBESTYEN .

Si on compare l'affection pour 2 groupes, on voit que :

sous hypothèse normale :

$$(X - Y_1)^t \cdot W_1^{-1} \cdot (X - Y_1) > (X - Y_2)^t W_2^{-1} (X - Y_2) + Log \frac{|W_2|}{|W_1|}$$

sous hypothèse de SEBESTYEN :

$$(X - Y_1)^t \cdot W_1^{-1} \cdot (X - Y_1) > (X - Y_2)^t \cdot W_2^{-1} \cdot (X - Y_2) \times \sqrt[P]{\frac{|W_2|}{|W_1|}}$$

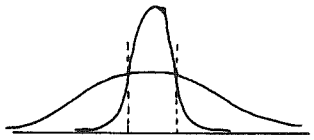
et dans le cas où $|W_1| \# |W_2|$ (égalité des volumes, mais pas des axes principaux) les 2 formules donnent des résultats voisins.

En fait des essais sur des exemples réels (ceux de ROMEDER) ne mettent pas en évidence des différences notables entre les 3 schémas suivants :

- multinormal
- SEBESTYEN
- MAHALANOBIS simple $\left(d^2(X, P_k) = (X - Y_k)^t W_k^{-1} (X - Y_k) \right.$

De même la suppression de la constante P dans SEBESTYEN ne modifie pas sensiblement la classification. En effet, les valeurs des constantes parasites P, $Log\ |W_k| / |W_j|$ restent en général faibles devant les distances de MAHALANOBIS et $(|W_k| / |W_j|)^{\frac{1}{P}}$ varie de 1 à 2, parfois 3, sauf cas spécial (une variable tend vers une variance nulle).

Remarque : Ces méthodes pourraient convenir même dans le cas de moyennes identiques



I.5 - Analyse discriminante sur variables qualitatives

La plupart des techniques précédentes supposent les variables continues et ayant une densité de probabilité $P(x)$ continue, même si elles sont relativement indépendantes de la loi $P(x)$.

Pourtant un cas fréquent est celui où les variables sont qualitatives : la variable prend un nombre fini de valeurs, par exemple 1, 2 ... k avec des probabilités $p_1, p_2 \ldots p_k$.

I.5.1. Aperçu sur le traitement complet

Un cas particulier est celui des variables binaires ou dichotoniques $X_j = 0$ ou 1 avec $Pr(X_j = 1) = p_j$. Dans ce cas, l'individu X_{iv}, caractérisé par q variables de Bernouilli $X_{iv} = \{x_{i1} \ldots x_{iq}\}$ suit une loi de probabilité multinomiale (cf LEBART et FENELON, 1973).

Celle-ci peut être caractérisée par :

$$p_j = E[X_j] \quad j = 1 \ldots q \qquad \text{d'où} \quad z_j = \frac{x_j - p_j}{\sqrt{p_j(1-p_j)}}$$

et les moments :

$$r_{jl} = E[X_j . X_l]$$

$$r_{jlm} = E[X_j . X_l . X_m] \quad \ldots\ldots \quad r_{1 \ldots j \ldots l \ldots m \ldots q}$$

Un certain nombre de méthodes ont été proposées, qui supposent les variables indépendantes, ou prennent en compte les corrélations, voire la totalité des moments.

Nous n'avons pas nous-mêmes appliqué ces méthodes car plusieurs auteurs ont comparé leurs performances à l'application directe, sur les données binaires, de l'analyse classique, linéaire ou quadratique. Parmi ceux-ci, GILBERT E.S. (1968) et MOORE D.H. (1973) ont utilisé des méthodes de Monte Carlo, allant jusqu'à 6 variables, pour 2 populations, et avec diverses intercorrélations. Ils concluent à la robustesse des méthodes classiques, avec un certain avantage pour la méthode linéaire.

Il est donc possible de l'appliquer telle quelle, sur des données binaires ou même sur un mélange comportant des données continues, discrètes ou binaires.

I.5.2. Cas du codage disjonctif

Une extension intéressante consiste à l'appliquer à des données continues préalablement transformées en données binaires par codage disjonctif complet.

Dans ce cas, chaque modalité d'une variable initiale X_j, soit X_j^k la kème, est une variable binaire.

On peut appliquer à ces variables une analyse discriminante linéaire avec sélection de variables et s'apercevoir que pour la variable X_j, seule la modalité ℓ est retenue : le classement dans P_1 ou P_2 se fera seulement sur une modalité de la variable X_j (P_2 si $X_j^\ell = 1$, P_1 sinon).

Exemple : On discrétise la variable température en 10 classes, mais seule la modalité $X^\ell = 1$ température dans la classe $[-2, +2°C]$

$= 0$ température en dehors de la classe

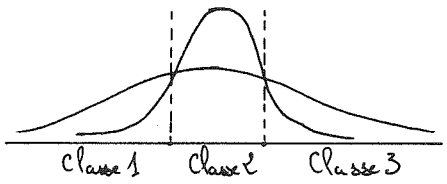
intervient pour prévoir les occurences de verglas.

On voit que cela peut amener, au niveau de la collecte de l'information significative, une économie considérable.

Par contre, il n'existe pas d'algorithme permettant réciproquement d'optimiser le codage (choix des limites de classes) pour maximiser la discrimination.

La question se pose sur le plan décisionnel de savoir s'il faut introduire toutes les modalités d'une variable, ou seulement certaines. Dans la mesure où le découpage en modalité n'est pas optimal pour la discrimination, et où 1 seule modalité ($X_j^k = 0$ ou 1) contient de l'information sur toute la variable (X_j appartient ou non à la classe k) il n'est pas utile de les introduire toutes.

Exemple :



Dans un but décisionnel, on voit que la seule modalité X_j^2 est suffisante.

De plus, il y a toujours intérêt dans un tel modèle à réduire au maximum le nombre de paramètres à estimer.

I.5.3

Par contre, sur le plan descriptif on peut vouloir garder, pour les variables les plus discriminantes, l'ensemble de leurs modalités. Cela oblige à sélectionner non plus des variables (fussent-elles binaires) mais des ensembles de variables (modalités) ce qui peut se faire (dans le cas de modalités exclusives ici) à

l'aide du coefficientde TSCHUPROW (cf G. SAPORTA, 1977).

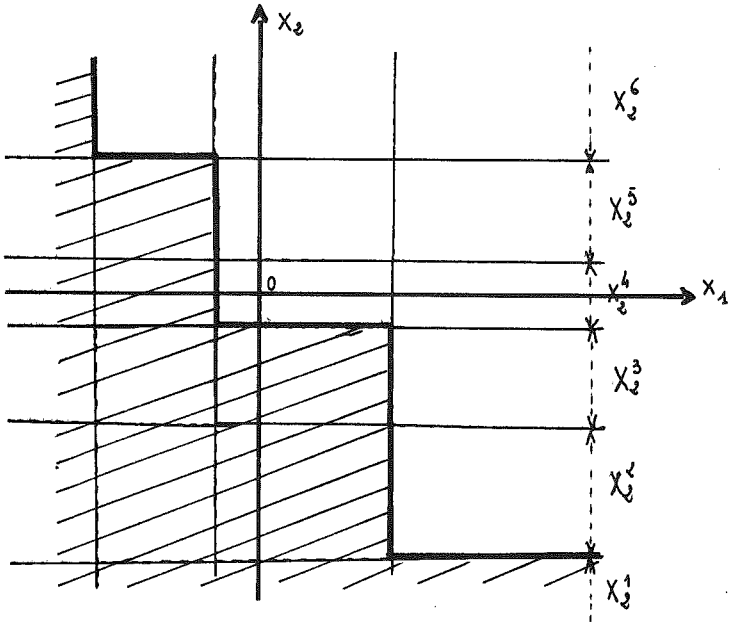
Si on considère le cas de 2 groupes et de 2 variables à K modalités, la discrimination linéaire fournit une fonction linéaire discriminante :

$$F(X) = (y_1 - y_2)^t W^{-1} X - \frac{1}{2}(y_1 - y_2)^t W^{-1}(y_1 + y_2)$$

avec $$X^t = \{ X_1^1, \ldots X_1^K, X_2^1, \ldots X_2^K \} \qquad X_j^k = 0 \text{ ou } 1$$

L'allocation se faisant à P_1 ou P_2 selon que $F(X)$ est > 0 ou < 0.

Or si on regarde dans $\mathbb{R}^2$, on constate que le codage revient à découper le plan en petits rectangles où 2 modalités seulement valent 1, et toutes les autres 0.



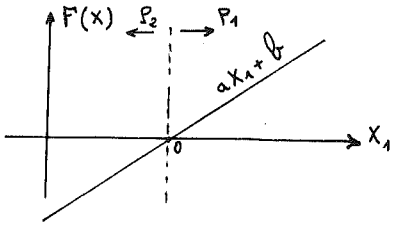
On peut alors calculer dans chaque rectangle la valeur (constante dans le rectangle) de $F(X)$ et en déduire les zones où l'on alloue à P_1 ou à P_2.

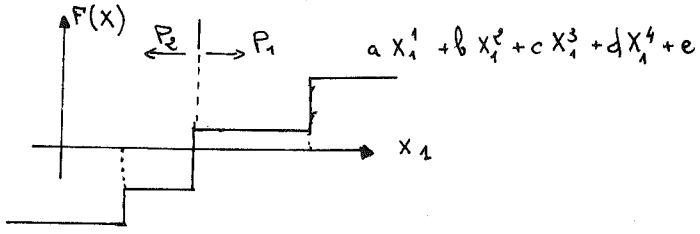
La, ou les séparatrices correspondent évidemment aux limites des rectangles (où la fonction est discontinue) et peuvent être quelconques

Il s'agit donc là d'une manière très économique de faire de la discrimination non linéaire (et sans se limiter au cas particulier quadratique).

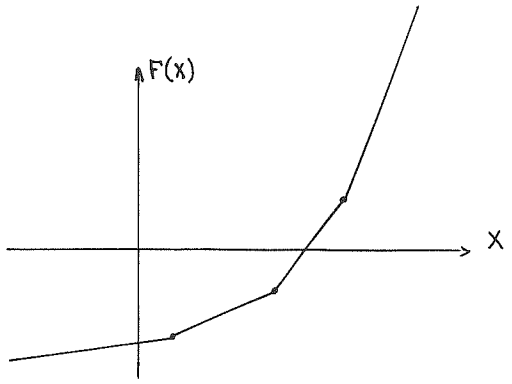
L'extension à plus de 2 variables est immédiate. Dans $\mathbb{R}^p$ on définit des petits domaines rectangles où $F(X)$ est constante et les surfaces séparatrices sont des surfaces en "escalier" (succession de facettes).

Diverses extensions de cette approche sont actuellement en cours. Si, en analyse discriminante linéaire, la fonction discriminante $F(X)$ est une droite, elle





devient, dans le cas discret une fonction en escalier, (polynômes de deg 0 par morceaux)



mais on peut chercher des polynômes de degrés supérieurs (linéaire par morceaux, etc.., cf LAFAYE DE MICHEAUX, 1978).

I.6 - Analyse discriminante non paramétrique (locale)

Une dernière famille de méthode rassemble les techniques d'analyse discriminante locale et non paramétrique.

I.6.1. Aperçu sur les différentes méthodes

On peut les classer approximativement en :

(a) Estimation locale de la densité de chaque population prise séparément. On applique alors une technique d'estimation (ex: PARZEN) de $f_1(x)$ et $f_2(x)$ aux 2 populations, puis on applique la règle de BAYES :

$$X \in P_1 \quad \text{si} \quad \pi_1 . f_1(x) > \pi_2 . f_2(x)$$

Cela permet d'estimer les 2 densités séparément, sur des échantillons partiels et de tailles différentes pour les 2 populations.

PATRICK et FISHER (1970) proposent une méthode intéressante de sélection des variables quand X est initialement à P dimensions. Ils proposent de rechercher un espace à ℓ dimensions dans lequel les distributions marginales des 2 populations, estimées par la méthode de PARZEN, sont le mieux séparées au sens d'une certaine distance. Cela répond au souci de réduire l'espace de travail aux dimensions les plus discriminantes.

Bien que non paramétriques, ces méthodes reviennent quand même à considérer l'ensemble des populations puisqu'elles approchaient $f(X)$. Elles ne sont donc pas à proprement parler locales.

(b) Une autre famille de méthodes consiste à découper l'espace $\mathbb{R}^P$ en régions, et selon celles où apparaît l'observation X à classer, on l'affecte à P_1 , à P_2, ou on réserve la décision.

Une méthode proposée par GESSAMAN et GESSAMAN (1972) consiste à découper l'intervalle de variation de la première variable X_1 de $\vec{X}$ en K classes (sur l'échantillon d'ajustement x_{11} à x_{1N}). Ensuite dans chacune des classes définie sur X_1 par exemple X_1^i , on découpe X_2 en K classes , etc... d'où une partition de en blocs statistiquement équivalents (car possédant le même nombre d'observations).

On décide ensuite pour chaque bloc de la population dominante.

On peut aussi faire le découpage en bloc sur le seul échantillon de la population P_1, puis porter les individus de P_2 et décider de l'affectation des régions. (On prendra par exemple pour P_1 l'échantillon le plus fourni). On pressent naturellement que plus le nombre de dimensions P est élevé, moins on peut faire de classes K ou plus le nombre d'individus par bloc est petit.

Et comme on l'a vu dans la Ière Partie (Chap.II-2), le codage en classes équiprobables risque de masquer les modes de la densité $f_1(x)$ si le nombre de classes est mal choisi. D'autres auteurs proposent donc d'effectuer une classification automatique sur chaque population, afin de détecter les noyaux de densité et de superposer les 2 classifications pour faire les blocs.

Ces méthodes sont réellement non paramétriques et locales. Par contre, elles sont statiques : une fois définis, les blocs ne varient plus, et elles ne permettent pas de sélectionner les variables.

(c) Les méthodes du "plus proche voisin" (Nearest Neighbor).

Elles consistent, dans leur forme la plus simple, à classer X dans la même population que son voisin le plus proche dans l'échantillon d'ajustement. Il peut sembler paradoxal de ne considérer qu'un seul voisin, mais COVER et HART (1967) ont montré que pour de grands échantillons, cette règle était optimale. En espérance mathématiques on montre que le risque d'erreur R (probabilité compte tenu éventuellement du coût) est compris entre :

$$R^* \leqslant R \leqslant 2.R^*(1-R^*)$$

où R^* est l'optimum bayesien (en supposant connues les distributions). Dans le cas où l'on a G populations ($G > 2$) alors :

$$R^* \leqslant R \leqslant R^*\left(2 - \frac{G}{G-1} R^*\right)$$

Il s'agit là du risque global sur l'ensemble du domaine, le risque local étant maximum là où les populations se recouvrent.

I.6.2. La méthode des K voisins les plus proches

(Méthode de FIX et HODGES ou méthode de la boule)

(a) Cette méthode tente d'estimer l'espérance conditionnelle locale $P_r(X \in P_1 \mid X)$ à l'aide des K plus proches voisins de l'observation à classer.

La démarche diffère légèrement de la précédente car elle suppose que pour la position X on a des densités $f_1(X)$ et $f_2(X)$ d'appartenir à P_1 ou P_2. Si l'échantillon était très grand, on aurait, pour la même abscisse (multidimensionnelle) X, K points dont $K_1 \in P_1$ et $K - K_1 \in P_2$

On estimerait alors :

$$q_x = Pr(X \in P_1 | X) \approx \frac{K_1}{K}$$

et on sait que la variance de cette estimation serait :

$$\text{var } q_x = \frac{q(1-q)}{K}$$

Et si on veut une précision de 0.1 (en écart-type) sur les valeurs de q_x, on voit qu'il faut prendre $K \approx 25$.

Mais cette estimation n'est vraie que si les K points ont même position X. Or si le fichier est de taille relativement faible (1000 ou 2000), il faut alors prendre un voisinage $v(X)$ où se trouveront K voisins. Mais plus ce voisinage sera grand, plus l'estimation, à K fixé perdra de la précision. Il y a donc un optimum à trouver dans les fichiers réels.

Celui-ci dépend de K, qui sera évidemment lié à N, taille de l'échantillon d'ajustement car si N est grand, on pourra augmenter K sans s'éloigner trop de X. Mais cet optimum dépend aussi de la dimension ℓ de l'espace de travail car, si K est fixé, la dispersion de ces K points tend à augmenter si ℓ augmente.

DER MEGREDITCHIAN (1969) cite une formule approchée de MECHALKINE :

$$K \approx \left(\frac{N}{2}\right)^{\frac{4}{\ell+4}}$$

qui donne un ordre de grandeur, et surtout met en évidence le rôle de ℓ.

Des exemples concrets ont montré, par tatonnement, que la valeur optimale de K est plutôt supérieure. Si on veut $K \approx 25$ et si on a $N = 800$, il ne faudrait pas dépasser 3 ou dimensions ! d'où la nécessité d'éliminer les redondances, et de sélectionner, parmi des variables indépendantes, les plus discriminantes.

Remarque : L'espérance conditionnelle tient compte des probabilités a priori des populations

$$Pr(X \in P_1 | X) = \pi_1 . f_1(X)$$

Il faut donc utiliser comme fichier d'ajustement 2 échantillons de P_1 et P_2 qui respectent le rapport des probabilités a priori. En général, en météorologie, on prend le fichier complet des observations d'une période donnée.

(b) Cette méthode se rattache aux estimations non paramétrique de densité puisque, si X est le point à classer, J.P. NAKACHE (1978) donne le théorème selon lequel :

$$f_N(X) = \frac{K-1}{N} \times \frac{1}{v(K,N,X)}$$

est bien un estimateur de $f(x)$.

Si N est le nombre total d'individus, K le nombre de voisins utilisés, et ϑ le volume de la sphère dans laquelle ils sont contenus.

Dans le cas des avalanches, on a K_a et K_s le nombre de voisins avec et sans avalanche (avec $K_a + K_s = K$ fixé). Le volume est le même pour les 2 populations et N_a et N_s sont les nombres totaux de journées avec ou sans avalanche dans l'échantillon (avec $N_a + N_s = N$).

On en tire :

$$Pr\left(X \mid Avalanche\right) = \frac{K_a - 1}{N_a} \times \frac{1}{\vartheta} \quad \Longrightarrow Pr(A|X) = Pr(A) \times Pr(X|A)$$
$$= \frac{N_a}{N} \times \frac{K_a - 1}{N_a} \times \frac{1}{\vartheta}$$

de même pour $Pr\left(X \mid S_{ans}\right)$ et la règle bayesienne conduit à déclarer avalanche si

$$\frac{K_a - 1}{K_s - 1} > \frac{N_a}{N_s}$$

(c) Enfin le problème de la distribution d'échantillonage des estimateurs utilisés devient particulièrement complexe. On trouvera quelques éléments dans G. COLLOMB (1978). Sur un plan pratique, on retrouve des expressions, pour les variances, où la dimension ℓ figure dans des exposants en $\frac{\ell+4}{4}$ (analogue à ceux cités par DER MEGREDITCHIAN). Cela souligne l'intérêt de réduire au maximum le nombre de dimensions.

CHAPITRE II

METHODES D'AGREGATION

II.1 - Introduction

(a) A plusieurs reprises dans ce mémoire, nous nous sommes posés la question, devant un ensemble d'individus a priori comparables, de savoir s'il n'existait pas en fait plusieurs groupements homogènes au sein de cet ensemble. Cette question peut avoir plusieurs origines :

- ayant sélectionné un échantillon présumé homogène, on se met à douter, par exemple après une visualisation, de cette homogénéité. Ce fut le cas pour les épisodes Cévenols (ou sélectionnés comme tels)

- au contraire, disposant a priori d'une classification (fondée sur des aspects physiques) en de multiples groupes, on se demande si celle-ci peut être perçue à l'aide des seules variables dont on dispose, ou si elles ne contiennent qu'une classification moins fine dans laquelle certains groupes initiaux ne peuvent être distingués. C'est le cas des types de temps météorologiques si on ne dispose pour les classer que des seuls champs de pression. C'est aussi le cas pour les divers types d'avalanches si l'on ne dispose que de variables météorologiques mais d'aucune caractéristique interne du manteau.

(b) Comme nous l'avons déjà évoqué dans la IIIème Partie, il existe 2 grandes familles de méthodes de classification:

- les méthodes de classification hiérarchique, qui s'intéressent moins au choix d'une partition finale en K groupes à définir qu'à une certaine hiérarchie d'association de ces individus. Ces méthodes proposent d'ailleurs une suite de partitions emboîtées, et considèrent les associations entre tous les individus pris 2 à 2 (d'où certaines limitations quant à la taille des problèmes à traiter). De plus, elles tendent à considérer l'échantillon comme une population complète à partitionner en classes disjointes, et nombre de méthodes s'intéressent plus aux propriétés "frontières" des classes, aux limites entre classes (cas des méthodes de "single-linkage" ou assimilées) qu'à leurs caractéristiques moyennes. De plus elles ne considèrent que rarement les problèmes d'échantillonnage et se prêtent peu à l' inférence statistique

- les méthodes non hiérarchiques, encore appelées méthodes d'agrégation. Dans ce cas, on ne considère plus un ensemble de partitions situées à différents niveaux dans une hiérarchie mais une seule, souvent caractérisée par un nombre donné K de groupes. Contrairement au cas précédent, ce ne sont pas des groupes entiers qui fusionnent, mais

au contraire des groupes initiaux (noyaux, étalons) caractéristiques des centres de classes, qui attirent à eux les individus pris isolément. La plupart des méthodes aboutissent à des classes disjointes, mais on peut en général plaquer sur les résultats des hypothèses probabilistes et admettre que certains éléments appartenant à 1 groupe ont une probabilité non négligeable d'appartenir à un autre. On peut facilement aussi considérer que l'ensemble d'individus dont on dispose n'est qu'un échantillon d'une population infinie, où il peut y avoir recouvrement partiel des différentes sous-populations. (Mais en admettant que l'échantillon disponible ne contient pratiquement pas de recouvrement).

Certes il existe un certain nombre de méthodes hybrides entre ces 2 familles (la méthode Iphigénie en est une) mais les ambiguités se limitent à celles qui travaillent sur des distances individuelles. Nous les ignorerons dans ce qui suit.

Pour replacer les méthodes que nous allons décrire dans un panorama plus complet, nous renvoyons soit à l'ouvrage de M. ANDERBERG (1973) soit à la synthèse bibliographique de J. ZIRPHILE (1974).

II.2 - Quelques méthodes d'agrégation

II.2.1. Principes généraux et méthodes classiques

— Les premières méthodes recherchent, en partant d'une partition aléatoire, à améliorer un critère. Pour cela, on balaye chaque élément de l'ensemble et on propose de le déplacer. Si ce déplacement améliore le critère, on continue. Ce critère peut être :

$$\text{Trace}(W) \qquad \text{Trace}(W^{-1}.B) \quad \text{ou} \quad \frac{\det(T)}{\det(W)}$$

Il n'existe malheureusement pas d'algorithme, autre que l'énumération, qui permette de calculer un optimum global : lorsque le balayage ne conduit à aucune réaffectation, il n'y a donc pas certitude que cet optimum soit global. (On notera de fortes analogies avec la méthode de WARD évoquée dans la Ière Partie, Chap.II-2). Toutefois ces méthodes ne font pas jouer de rôle particulier à certains points de l'ensemble, mais elles sont en général assez coûteuses en calcul car très combinatoire.

— Un deuxième ensemble de méthodes consiste à caractériser chaque groupe par un étalon. Cela peut se faire :

- par tirage au hasard ou choix arbitraire
- en utilisant une première partition et en utilisant les centres de classes de celle-ci (obtenue par ailleurs : classification hiérarchique, etc...)

Disposant de ces k points étalons (qui sont ou non des points figurant réellement dans l'ensemble à classer), on leur fait correspondre une partition en k groupes. Dans chacun de ceux-ci on reprend 1 étalon et on itère le procédé jusqu'à la convergence.

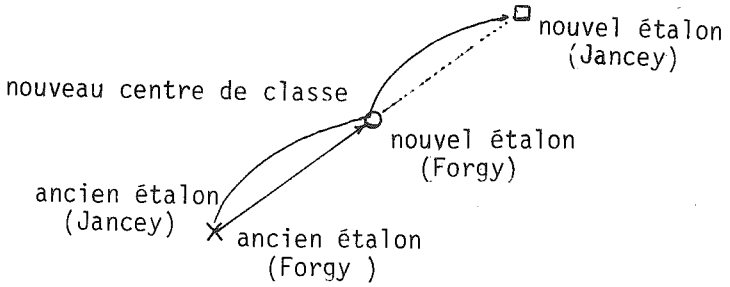
Parmi ces méthodes nous en avons utilisées 3 qui procèdent ainsi :

(a) Méthode de FORGY (1965) variante de JANCEY (1966) (citées par ANDERBERG, 1973)

Selon que l'on se donne une première partition ou k étalons, on part du pas 2 ou 1.

1. On agrège les N points aux K étalons
2. On prend comme nouveaux étalons les centres des classes obtenues
3. On teste si les classes obtenues sont stables depuis la dernière itération, sinon on retourne au pas 1.

La variante de JANCEY (utilisant le principe de la surrelaxation, ou de certaines méthodes de gradient) consiste, une fois calculé le nouveau centre de classe, à prendre comme étalon non pas celui-ci mais le symétrique de l'ancien étalon par rapport au nouveau centre de classe.



Ceci a pour but de compenser le biais dû à ce que, au cours du pas n°1, et bien que l'on puisse déjà voir que la classe se déplace à droite, on continue à agréger sur l'ancien étalon. JANCEY (1966, cité dans ANDERBERG, 1973) pense accélérer ainsi la convergence et éviter certains extréma locaux, mais des cas de divergences peuvent apparaître.

On peut montrer que ces méthodes reviennent à minimiser la somme des carrés des distances intra-groupes ou encore trace (W).

Les surfaces séparatrices des groupes sont des hyperplans orthogonaux aux segments de droite reliant les groupes. On remarquera que ces méthodes ne mettent en jeu que des distances entre points et étalons. Il n'y a donc pas de métrique naturelle autre que la métrique euclidienne I. Mais si on applique ensuite une analyse discriminante linéaire, voire quadratique à la partition obtenue, certains points apparaissent mal classés.

Par contre, le fait que l'étalon ne change pas pendant l'agrégation rend ces méthodes insensibles à l'ordre de balayage.

(b) Méthode de MACQUEEN

Elle remet en cause ceci et effectue 2 balayages successifs. Partant de k étalons, elle consiste :

1. à agréger chaque élément avec l'étalon E_j de la classe la plus proche par exemple C_j. Aussitôt on recalcule un nouvel étalon E'_j de cette classe compte tenu du nouvel élément.

2. Arrivé à une partition complète mais avec des étalons qui ont évolué, on refait une agrégation sur ces nouveaux étalons, maintenus fixes pendant ce pas.
3. On compare avec la partition du pas 2 de l'itération antérieure et éventuellement on itère. Cette méthode a connu diverses variantes. Toutes ont l'avantage d'être très économiques. Par contre, l'ordre de balayage des individus au pas 1 peut influer sur le résultat.

II.2.2. La méthode des nuées dynamiques

La plupart de ces méthodes ont été proposées dans les années 1964 à 1969 et le seul essai de formalisation fut celui de LANCE et WILLIAMS (1968).

E. DIDAY, à partir de 1971, a tenté de proposer une méthode, celle des "nuées dynamiques" qui tentait de remédier à la plupart des inconvénients rencontrés et de s'appuyer sur une formalisation poussée afin de pouvoir inscrire un certain nombre de variantes dans un seul cadre très général.

Nous en décrivons la variante la plus simple (cf DIDAY, 1974 et LECHEVALLIER Y., 1974).

(a) Idées de départ et définitions

On cherche toujours à caractériser les classes par un sous-ensemble d'éléments, mais qui ne se réduit plus ici à 1 seul étalon.

Soit: E ou X l'ensemble des éléments à classer x_i

K le nombre de classes demandées

C_k une classe contenant n_k élément x_{ik} ($i = 1 \ldots n_k$)

$\mathcal{P}_k$ l'ensemble des partitions de X en K classes disjointes.

$$\forall P \in \mathcal{P}_k \qquad P = \{C_1, C_2, \ldots C_K\}$$

avec : $E = \cup C_i$ et $C_i \cap C_j = \phi \quad \forall i \neq j$

$\mathcal{L}_k$ ensemble des noyaux ou étalons :

$$L \in \mathcal{L}_k \qquad L = \{L_1, L_2, \ldots L_K\}$$

avec cette fois : $\text{card}(L_j) = h_j \geq 1$

et $L_i \cap L_j = \phi \quad \forall i \neq j$

(On cherche donc à caractériser la classe P_k par un ensemble d'étalons L_k ou noyau, qui ne se réduit pas forcément à 1 seul élément).

On se définit les applications suivantes :

. d distances entre individus:

cette distance sera par exemple quadratique :

$$d^2(x, y) = (x-y)^t Q (x-y)$$

. D distance d'un individu à un groupe:(ce groupe étant soit une classe C_j de P, P étant une partition de $\mathcal{P}_K$, soit un noyau L_j de $\mathcal{L}_K$). La

distance la plus courante est

$$D^2(x, Y) = \frac{1}{card(Y)} \sum_{x_j \in Y} d^2(x, x_j)$$

on en trouvera d'autres, ainsi que leurs effets comparés sur les résultats, dans Y. LECHEVALLIER (1974).

. f "fonction" qui fait passer de $\mathcal{L}_K$ dans $\mathcal{P}_K$ c'est-à-dire qui, à partir d'un ensemble de noyaux $L = \{L_1, L_2, \ldots L_K\}$ fournit une partition P de $\mathcal{P}_K$. Pour cela, on agrège sur les noyaux les éléments n'appartenant pas déjà à ces noyaux par :

$$C_j \in P \quad C_j = \{ x \in E \text{ tels que } D(x, L_j) < D(x, L_k) \ \forall k \neq j, k = 1 \ldots K \}$$

donc C_j est l'ensemble des éléments de E les plus proches, au sens de D, du noyau L_j.

. g "fonction" qui fait passer de $\mathcal{P}_K$ dans $\mathcal{L}_K$ c'est-à-dire qui extrait, d'une partition P de l'ensemble E, $P = \{P_1, \ldots P_K\}$ un ensemble de noyaux : $L = \{L_1, L_2 \ldots L_K\}$

On extrait ainsi de C_j un noyau L_j par :

$$L_j = \left\{ x_i \in C_j \text{ tels que } \underset{i=1,\ldots h_j}{D(x_i, P_j)} = \min_{x \in C_j} D(x, C_j) \right\}$$

Remarques - Parmi les nombreuses variantes possibles celle que nous étudions est dite à masse constante (tous les points ont même pondération).

La fonction f est bien celle définie ci-dessus, mais la fonction g se réduit à :

$$\left\{ x_i \in C_j \text{ tel que } d(x, G(j)) = \min_{x \in P_j} \right\}$$

où $G(j)$ est le centre de gravité de C_j défini par :

$$G(j) = \frac{1}{card(C_j)} \sum_{x_i \in C_j} x_i$$

et $G(j)$ n'est en général pas un élément de E.

Mais on prend comme noyau de la classe C_j les h_j éléments les plus proches de son centre de gravité.

Si de plus on suppose la distance d quadratique, alors on peut vérifier que :

$$D(x, L_j) = \frac{1}{card(L_j)} \sum_{x_i \in L_j} (x - x_i)^t M (x - x_i)$$
$$= var(L_j) + d^2(x, G(L_j))$$

avec :

$$var(L_j) = \frac{1}{card(L_j)} \sum_{x \in L_i} d^2(x, G(L_j))$$

(b) Fonctionnement de l'algorithme

Etant donné une partition P et un ensemble de noyaux L (qui peuvent avoir été choisis indépendamment) de la partition, on se définit une fonction d'agrégation-écartement R qui mesure le degré d'association d'un individu avec un groupe de la partition

$$R(x, j, P) = D^2(x, C_j) \times \frac{card(C_j)}{card(L_j)}$$

et un critère de qualité des noyaux par rapport à la partition :

$$W = \sum_{j=1}^{K} \left(\sum_{x_i \in L_j} R(x_i, j, P) \right) = W(L, P) = W(v)$$

en posant : $v = \{L, P\}$

Ceci devient dans notre cas :

$$W = \sum_{j=1}^{K} D^2(L_j, C_j) = \sum_{j=1}^{K} card(C_j) \cdot \left[var(L_j) + var(C_j) + d^2(G(L_j), G(C_j)) \right]$$

On voit que dans ce cas, le critère est faible si les classes sont bien agrégées et si la distance de la classe au noyau associé est faible, donc si le noyau représente bien la classe.

Remarque 1 - Le critère W ainsi choisi ne considère que l'agrégation des classes. D'autres critères, utilisant d'autres fonctions R ont été proposés, par exemple :

$$R(x, j, L) = \frac{D^2(x, L_j) \cdot D^2(x, C_j)}{\sum_{i=1}^{K} D^2(x, L_i)}$$

Dans ce cas le minimum du critère suppose en plus que les noyaux soient bien séparés. E.DIDAY a démontré les propriétés que devaient posséder une fonction R pour être acceptable.

L'algorithme va donc procéder comme suit.

(1) On part d'une partition quelconque $P^{(0)}$ et d'un ensemble de noyaux $L^{(0)}$.
En fait $P^{(0)}$ n'a pas besoin d'être précisée.
$L^{(0)}$ peut être soit tiré au hasard soit fourni par l'utilisateur.
On a alors le doublet $v^{(0)} = \{ L^{(0)}, P^{(0)} \}$ et en général $W(v^{(0)})$ est mauvais.
On calcule donc immédiatement :

$$P^{(1)} = f(L^{(0)})$$

partition associée aux étalons de départ et

$$L = g(P^{(1)})$$ les étalons extraits de $P^{(1)}$

② On itère le processus

$v^{(\ell)} = \{P^{(\ell)}, L^{(\ell)}\}$ avec $P^{(\ell)} = f(L^{(1)})$, $L^{(\ell)} = g(P^{(\ell)})$

et on compare $W(v^{(\ell)})$ à $W(v^{(1)})$

③ On s'arrête quand

$W(v^{(n+1)})$ n'est plus sensiblement inférieur à $W(v^{(n)})$

et on garde $v^{(n+1)} = \{L^{(n+1)}, P^{(n+1)}\}$

E. DIDAY a montré que la suite $W(v^{(k)})$ décroît et converge vers un optimum local.

Remarque 2 - La forme du critère W (donc de la fonction R) n'intervient pas sur le processus d'agrégation qui ne dépend que de D et d, mais sur l'arrêt de l'algorithme. Selon la forme de W, à un pas donné des intérations supplémentaires peuvent encore l'améliorer ou non, c'est pourquoi la partition obtenue variera selon W.

Remarque 3 - En dépit de l'appareil mathématique dont elle est entourée, il est difficile à l'utilisateur de caractériser la qualité de la partition obtenue.

En effet, le critère W ne dépend pas toujours de l'écartement des classes, or cela serait souhaitable. Et s'il dépend de leur agrégation, donc s'il détecte les zones denses du nuage, les termes de distances entre centres de gravité des noyaux et centre de gravité des classes semblent un peu parasites car souvent l'utilisateur n'a que faire des noyaux pour la suite de son étude.

Cette méthode ne se compare donc pas directement avec celles cherchant à maximiser des critères comme trace($W^{-1}.B$), etc... Par contre, elle se compare avec les méthodes de FORGY ou MAC QUEEN, où là encore on ne considère que l'agrégation des groupes.

(c) Paramètres de l'algorithme

Ce sont :

NI le nombre d'initialisations effectuées, donc le nombre d'optima obtenus

K le nombre de classes demandées

n_e le nombre d'étalons par noyau (en général 10 à 20% des éléments de la classe).

Un autre paramètre est le choix de la distance. Dans notre cas (données de départ continues) nous l'avons prise euclidienne. Dans le cas où les variables ont une structure de corrélation stable d'un groupe à l'autre, on peut prendre, mais seulement à partir du second pas, une métrique de MAHALANOBIS W^{-1}. En fait dans notre cas, nous avons toujours trouvé préférable, vu le nombre de distances à évaluer, de travailler dans un espace $\mathbb{R}^{\ell}$ où ℓ est le plus faible possible, par exemple dans l'espace des premiers facteurs principaux.

On verra dans III-2 une autre approche.

(d) Les formes fortes

En plus de la valeur du critère et de la partition obtenue, qui peut varier d'un tirage à l'autre, on peut chercher ce qu'un certain nombre de tirages ont en commun.

On appellera "forme forte" un ensemble d'éléments qui, dans tous les tirages, sont toujours apparus groupés dans la même classe.

Evidemment, si on effectue un seul tirage, les formes fortes coïncident avec les classes de la partition obtenue. Au delà, pour $NI > 1$, elles ne peuvent être que des sous-ensembles, de plus en plus nombreux et de cardinaux de plus en plus faibles, des classes obtenues.

Pour NI tirages, une forme forte sera la partie commune (intersection) des classes qui se ressemblent le plus à raison d'une par tirage successif.

Pour les détecter, on peut construire le tableau qui à l'élément x associe le n° de classe dans laquelle il se trouvait au ième tirage.

$$x \longrightarrow c(x) = \left\{ \begin{array}{cccccccc} 1 & 2 & \cdots & i & \cdots & NI & \leftarrow \text{N° du tirage} \\ 2 & 5 & \cdots & k & \cdots & l & \leftarrow \text{N° de la classe} \end{array} \right\}$$

On peut définir un indice entre 2 éléments x et y :

$$\Delta(x,y) = NI - \{ \text{nombre de fois où } c(x) = c(y) \}$$

et $\Delta(x,y) = 0 \Longrightarrow x$ et $y \in$ même forme forte

et les formes fortes sont les classes d'équivalence de $E/\mathcal{R}$ où $\mathcal{R}$ est la relation $x \mathcal{R} y \Longleftrightarrow c(x) = c(y)$. En fait on peut, en utilisant Δ comme une distance et donc le tableau des $N(N-1)/2$ interdistances entre les éléments de E construire un arbre (dit des connexités descendantes) dont en général seul les derniers niveaux nous intéressent.

LECHEVALLIER propose d'étudier la stabilité de ces formes en comparant leur vraisemblance à ce que l'on obtiendrait en supposant les partitions complètement aléatoires. Malheureusement, il faudrait faire un nombre de tirages important pour que la différence soit très significative.

Il n'en reste pas moins que cette notion de formes fortes est l'élément le plus intéressant de la méthode, dans la mesure où elle introduit une notion de fluctuation d'échantillonage des partitions obtenues et élimine partiellement le rôle de l'initialisation.

II.3 - Essais d'évaluation de la méthode des nuées dynamiques

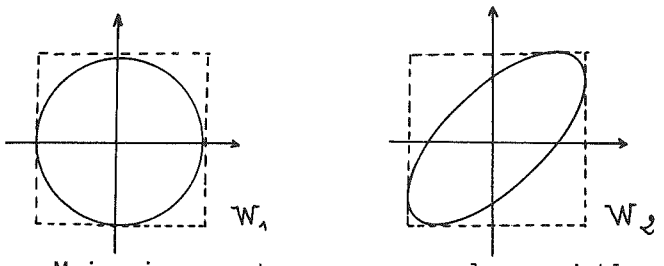
II.2.1. Détection du nombre réel de groupes : Méthodes possibles

Le problème le plus délicat en classification est de déterminer le nombre K de groupes à construire. En effet, si les données sont réellement multidimensionnelles, et en l'absence d'information extérieure, il est très difficile de pressentir ce nombre.

(a) La seule méthode proposée dans la littérature est due à VOGEL et WONG (1975)

Des tentatives antérieures cherchaient à utiliser det (W) comme mesure de l'agrégation dans une partition donnée .(En fait c'est $\frac{|W|}{|T|}$, c'est-à-dire le critère de WILKS Λ . Mais comme det(T) est inchangé on peut ne considérer que det(W). Malheureusement, quand on augmente K , det (W) ne peut que décroître.

Un critère de moins bonne qualité que det(W) et trace(W) qui ne prend pas en compte les intercorrélations entre variables.



$$\det(W_1) > \det(W_2)$$

$$\text{trace}(W_1) = \text{trace}(W_2)$$

Mais si on peut supposer que les variables sont raisonablement indépendantes alors on peut utiliser trace(W) comme critère, ce qui facilite grandement les calculs.

D'autre part, il faut un critère qui soit comparable d'un nombre de groupes à un autre, par analogie avec l'analyse de la variance à 1 dimension entre K classes, où l'on considère le rapport :

$$\frac{SSB}{SSW} \times \frac{N-K}{K-1}$$

Celui-ci suit une loi de FISHER avec ($K-1$, $N-K$) degrés de liberté (cf LEBART et FENELON, 1973). Plus F sera grand, plus la dispersion entre groupes sera importante et significative.

L'extension utilisée par VOGEL et WONG consiste donc à prendre comme variable la distance euclidienne (quelle que soit son orientation dans $\mathbb{R}^P$) d'où les sommes de carrés :

$$SST_e = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(\sum_{j=1}^{P} (X_{ij} - \bar{X}_j)^2 \right) = \sum_{i=1}^{N} d^2(x_i, G)$$

$$SSW_e = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(\sum_{j=1}^{P} (X_{ij} - \bar{X}_j^k)^2 \right) = \sum_{k=1}^{K} \sum_{i \in C_k} d^2(x_i, G(C_k))$$

$$SSB_e = \sum_{k=1}^{K} n_k \sum_{j=1}^{P} \left(\bar{X}_j^k - \bar{X}_j \right)^2 = \sum_{k=1}^{K} n_k \, d^2(G(C_k), G)$$

où G est le centre de gravité global (souvent pris pour origine).

On a alors un pseudo F ou $PFS = \frac{SSB_e}{SSW_e} \times \frac{N-K}{K-1}$

avec $SSW_e + SSB_e = SST_e$

On notera que ces quantités représentent, à des facteurs près, les traces des matrices de variance-covariance correspondantes.

Mais surtout, ce critère est débiaisé vis-à-vis du nombre de groupes et permet de comparer 2 partitions à K et $K+1$ groupes.

On fera donc varier le nombre de groupes K_d démandés, la valeur vraie K_v devant donner un maximum pour PFS.

Remarque - On vérifiera que la somme SSW_e n'est autre que :

$$\sum_{k=1}^{K} n_k \cdot var(C_k)$$

où $var(C_k)$ est la variance de la classe C_k définie en II.1.2(a).

Des essais ont montré que le maximum de F en fonction de K_d est cependant assez plat.

(b) De façon plus heuristique, on peut penser détecter le nombre de groupes vrai K_v en demandant 1, 2, ... K_d groupes.

On sait que si K_d est trop grand, certaines classes seront vides et le nombre de groupes obtenus tend à plafonner.

(c) Un meilleur critère est plutôt le nombre de formes fortes obtenu pour un nombre fixé NI d'initialisations (car ce nombre augmente avec NI).

On a vu que si l'on a K_v groupes bien séparés et que l'on demande une partition en K_v classes les formes fortes seront très stables et peu nombreuses. Par contre si on demande $K_d > K_v$ groupes, le découpage à travers les groupes existants se fera aléatoirement à chaque partition et le nombre de formes fortes devrait croître de façon très sensible.

II.3.2. Quelques résultats en simulation

Nous donnons ici quelques résultats de simulation, dans le cas de 2 et 3 groupes vrais, avec la méthode des nuées dynamiques.

On a simulé successivement des groupes de données gaussiennes dont la variance est $\sigma_x = \sigma_y = \sigma = 50$ dans $\mathbb{R}^2$, et dont la distance entre centres $\Delta = 2,3$, ou 4σ.

On s'est limité à 8 initialisations aléatoires par configuration, et 10 itérations pour améliorer le critère.

(a) Résultats dans le cas de 2 groupes

Dans ce cas, seule la variable x sépare les groupes.

On constate que pour $K_d = 2$, le nombre de formes fortes est petit et qu'il croît brutalement pour $K_d = 3$.

Par contre les classes vides apparaissent surtout pour $K_d = 4$ et relativement peu pour $K_d = 3$ même si la distance entre centres $\Delta = 4\sigma$.

Enfin, le PFS détecte toujours 2 groupes, mais de façon beaucoup moins nette quand $\Delta = 2\sigma$ (et dans ce cas on a même PFS(4) > PFS(3) mais il faut noter que le hasard de l'initialisation peut conduire à une très bonne configuration sur $K_d = 4$ et à une moins bonne pour $K_d = 3$.).

On a voulu tester l'influence d'une corrélation au sein des groupes en mettant, dans le cas $\Delta = 3\sigma$, $\rho_1 = \rho_2 = .6$ et $\rho_1 = -\rho_2 = .6$

On peut s'attendre dans le 1er cas à une meilleure agrégation des groupes, qui apparait dans le PFS mais pas au niveau des formes fortes ni du nombre de classes vides.

Par contre, dans le cas de 2 groupes anticorrelés, où l'on aurait pu craindre un certain recouvrement, la détection n'est pas très différente, sauf pour les PFS moins franchement décroissants.

Notons cependant que ceci a été obtenu en utilisant une métrique euclidienne I, alors que pour le 1er essai, une métrique de MAHALANOBIS serait optimale, et que le 2nd nécessiterait des métriques locales.

Exemple de simulation : $K_V = 2$ groupes $n_1 = n_2 = 60$ individus $\Delta = 4\sigma$

Nb de groupes demandés	Nb de classes vides pour NI = 10	Nb de formes fortes obtenues	Formes fortes obtenues (par nombre d'individus décroissants)
2	273.4	0	2	62 58
3	193.2	1	11	37 26 22 14 6 4 3 3 2 2 1
4	174.7	6	21	22 19 16 15 7 7 6 6 4 3 3
P F S ↑	(associé au w_{min})			2 2 1 1 1 1 1 1 1 1

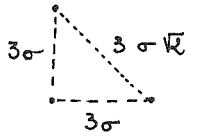
(b) Résultats dans le cas de 3 groupes

Ici le nombre de configurations devient un peu plus combinatoire, depuis 3 groupes alignés jusqu'au cas où ils sont aux sommets d'un triangle équilatéral, et ce pour diverses interdistances.

Nous constatons que cette configuration est essentielle et que, si les groupes ne sont pas équidistants, la méthode tend à agréger ceux qui sont relativement proches d'où une évolution du PFS plus plate au voisinage du maximum et une structure des formes fortes parfois plus stable pour $K_d < K_V$, mais dont le nombre augmente toujours assez brutalement pour $K_d > K_V$. De plus dans le cas où les groupes sont relativement équidistants les classes vides n'apparaissent surtout que pour $K_d = 5$....!

Enfin, les 2 variables x et y sont dans ce cas discriminantes, aussi avons nous testé l'influence d'une variable z non discriminante mais utilisée dans le calcul des distances. On pressent que la dispersion des points selon une 3ème dimension va augmenter l'incertitude sur les groupes, donc le nombre de formes fortes.

Dans ce cas on ne constate plus une augmentation brutale de leur nombre quand $K_d > K_V$ mais plutôt une évolution progressive et tous les critères pratiquement deviennent flou ou sont pris en défaut.



Exemple de simulation : Kv =3 $n_1 = n_2 = n_3 = 40$

Nb de groupes demandés	Nb de classes vides pour NI = 8	Nb de formes fortes	P F S		
			2 var.	3 var. ($\sigma_z = 2\sigma$)	3 var. ($\sigma_z = 3\sigma$)
2	0	4	105.2	32.8	23.2
3	1	8	150.2	44.4	30.8
4	1	25	128.0	45.1	35.0
5	5	23	114.1	-	-

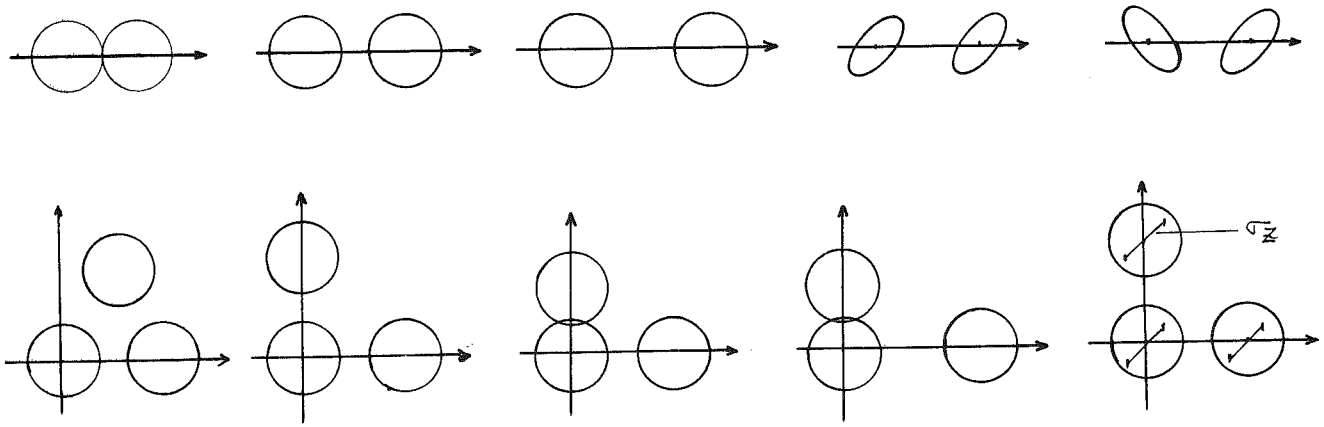


FIGURE V - 3 : Types de configurations simulées . ($\sigma = \vdash\!\dashv$)

c) Conclusions

Les aspects "statistiques" de la méthode des nuées dynamiques :

- initialisations successives
- notions de formes fortes

fournissent effectivement des informations précieuses.

Parmi les critères utilisés :

1- le nombre de classes vides:est peu "pointu" car elles n'apparaissent que pour $K_d \geq K_v + 1$, et à la condition que les groupes soient bien séparés.

2- le nombre de formes fortes : augmente assez brutalement quand $K_d > K_v$, et ce d'autant plus que les groupes sont moins séparés (car il y a plus de possibilités d'intersection). Il tend à être constant tant que $K_d < K_v$.

3- le critère PFS est le plus précis des 3 pour détecter le nombre de classes.

Ces remarques sont vraies tant que les variables utilisées sont discriminantes .Quand on introduit des variables supplémentaires non discriminantes, tous les critères deviennent flous, surtout si ces nouvelles variables ont une forte variabilité. Dans ce cas, le nombre de formes fortes peu être grand même si $K_d = K_v$ et on inclinera à choisir K trop faible, alors que le critère PFS indiquera plutôt une valeur de $K > K_v$.

Ceci met en évidence le rôle critique des variables inutiles, malheureusement inconnues a priori dans les cas réels.

La configuration des groupes, si elle intervient déjà dans le nombre de dimensions significatives (axes discriminants), influe aussi sur la sensibilité de la méthode pour détecter le nombre de groupes. Il semble préférable quand on pressent l'existence de "paquets" de demander peu de classes K_d, qui correspondront à ces paquets, quitte à les redécouper ensuite séparément.

Un dernier point concerne l'influence des corrélations intra-groupes, qui n'est pas très clair. Autant que possible il vaudra mieux utiliser la distance euclidienne $\dot{I}$, plutôt qu'une métrique de MAHALANOBIS moyenne qui peut mélanger des cas extrêmes. Notons cependant que l'utilisation de variables globalement indépendantes sur l'ensemble (composantes principales, par exemple) ne présume en rien des corrélations intra-groupes que l'on pourra obtenir.

II.3.3. Quelques variantes de la méthode

Des essais préliminaires ayant montré que les groupes obtenus pouvaient avoir des matrices de covariance très différentes, il nous a semblé intéressant d'en tenir compte et d'utiliser les distances propres à chaque groupe. G. GOVAERT (1975) a proposé l'algorithme suivant :

- soit $v^{(n)}$ le triplet $v^{(n)} = \{P^{(n)}, L^{(n)}, \Delta^{(n)}\}$ où $P^{(n)}$ est une partition de E en K classes, $L^{(n)}$ un ensemble de K noyaux et $\Delta^{(n)}$ un ensemble de K distances, ce à l'itération n, et partant d'une configuration $v_0 = \{P^{(0)}, L^{(0)}, \Delta^{(0)}\}$ (où la distance peut être euclidienne et la partition et les noyaux aléatoires)
- on passe à $v^{(n+1)} = \{P^{(n+1)}, L^{(n+1)}, \Delta^{(n+1)}\}$ par $P^{(n+1)} = f(L^{(n)}, \Delta^{(n)})$ (avec les distances Δ_n)

  $L^{(n+1)} = g(P^{(n+1)}, \Delta_n)$ noyaux caractéristiques de $P^{(n+1)}$ au sens de $\Delta^{(n)}$.

  $\Delta^{(n+1)} = e(P^{(n+1)}, L^{(n+1)})$ ensemble de distances d_k qui maximise l'agrégation du groupe C_k sur le noyau L_k.

Cette méthode, très attrayante car dans les exemples réels, les distances ne sont pas les mêmes d'un groupe à l'autre, présente 2 inconvénients :

- en cas de matrice de covariance non inversible : dans un groupe C_k, la variable X_j est telle que $\sigma_j(C_k) = 0$
- au niveau de la convergence : l'algorithme peut être conduit à calculer des métriques très "aplaties" en cas de points "anormaux" d'où des difficultés en cas d'initialisations aléatoires (effets de "cylindre"). Cette technique ne peut donc être utilisée qu'en phase finale, pour raffiner une partition obtenue par la méthode normale.

Nous ne l'utiliserons d'ailleurs pas et compte tenu des incertitudes dues à la métrique retenue, nous préférerons procéder comme suit :

- déterminer le nombre de groupes probable dans E, soit K.
- effectuer un certain nombre d'initialisations avec la métrique $\dot{I}$, d'où un ensemble de formes fortes. On garde les K formes les plus caractéristiques des K groupes.
- on réaffecte les points restants par analyse discriminante linéaire ou quadratique.

CHAPITRE III

APPLICATION A LA PREVISION DES AVALANCHES A DAVOS

III.1 - Analyse à 2 groupes (Modèles de type I)

Nous abordons le problème de manière élémentaire en considérant qu'il y a occurence ou non occurence d'avalanche. Ce faisant nous ne considérons que les effets observés, sans imposer de modèle quant aux causes possibles. Des tentatives antérieures (Ph. BOIS, 1976) avaient considéré 2 types d'avalanche pour Mars et Avril (neige sèche ou neige humide) mais déjà là se posent des problèmes quant à la qualité des observations, et au choix de l'échantillon sans avalanche. On se limite donc au cas le plus simple.

III.1.1. Utilisation de variables continues

On donne les résultats obtenus pour Janvier-Février et Mars-Avril. Dans les 2 cas on a considéré que les coûts étaient tels que : $\pi_1 . C_1 = \pi_2 . C_2$ ce qui revenait à mettre des coûts égaux et mêmes probabilités a priori.

Dans le cas de 2 groupes, l'analogie entre analyse discriminante et régression multiple permet une certaine interprétation des variables. En effet, l'analogie est immédiate si on utilise la matrice T qui est la matrice de corrélation totale (cf LEBART et FENELON, 1973, p.287). Dans le cas où l'on utilise W , le résultat est identique mais sa démonstration est un peu plus délicate (cf GROSJEAN, cité par KUENY J.L., 1976). Les coefficients de régression sont donc proportionnels aux composantes du vecteur discriminant : $\left(\bar{y}_2 - \bar{y}_1 \right)^t . W^{-1}$

Remarque - Si on code $y(X) = 0$ ou 1 selon que $X \in P_1$ ou P_2 alors il faut multiplier par $\pi_1 . \pi_2 = n_1 . n_2 / N^2$, tandis que si on code $y(X) = \sqrt{n_2 / n_1}$ et $- \sqrt{n_1 / n_2}$, les coefficients de régression sont strictement égaux à $\left(\bar{y}_2^2 - \bar{y}_1 \right) . T^{-1}$

Or on sait que si ces coefficients de régression ne sont pas reliés simplement aux coefficients de corrélation partielle, ils sont de même signe et donc, le signe du coefficient c_j de la variable X_j dans la fonction discriminante indique si y augmente ou non avec la variable X_j, les autres variables étant supposées constantes.

Les signes indiqués dans les tableaux ci-contre indiquent la liaison avec le groupe "avalanche". On rencontre cependant les mêmes problèmes qu'en corrélation multiple : instabilité ou compensation entre variables correlées 5, 30, 33 pour Janvier-Février 36, 37, 38 pour Mars-Avril.

De plus on a mélangé ici toutes les avalanches d'où par exemple le rôle de la variable 46 (qui accroît le danger d'avalanche s'il y a réchauffement en Janvier-Février). Or nous verrons que ce rôle est très différent selon les types de temps avalancheux (cf III-2).

TRAITEMENT SUR VARIABLES CONTINUES

	Janvier-Février		Mars-Avril	
Nombre de variables utilisées :	10		12	
Pouvoir discriminant :	37%		39 %	
Variables utilisées :	5 (+)	31	30 (+)	39
	36 (-)	27	46 (-)	34
	39 (+)	27	50 (+)	31
	30 (+)	25	3 (-)	29
	25 (+)	25	37 (-)	27
	35 (+)	23	15 (+)	23
	46 (-)	23	25 (-)	21
	21 (+)	23	16 (+)	22
	33 (+)	22	36 (+)	19
	16 (-)	21	5 (+)	21
			38 (+)	20
			26 (+)	17
Nombre de journées testées :	240			
Nombre d'alertes déclarées :	88			
dont journées avalancheuses:	41			
Journées avalancheuses non détectées :	7			

TRAITEMENT SUR VARIABLES DISCRETES

	Janvier-Février		Mars-Avril	
Modalité:	3 de la variable 30 (+)		1 de 22 (-)	
	2 42 (-)		3 46 (-)	
	3 22 (-)		3 30 (+)	
	3 4 (-)		2 13 (+)	
	1 36 (+)		2 37 (-)	
	2 30 (+)		2 32 (-)	
	1 46 (+)		3 33 (+)	
	1 44 (-)		3 28 (+)	
	3 35 (+)		1 8 (-)	
	1 12 (-)		2 10 (-)	
			3 9 (+)	
			3 20 (-)	
Erreur de classification sur l'échantillon d'ajustement	15 %		20 %	

TABLE I : RESULTATS POUR L'ANALYSE A 2 GROUPES

Les résultats de ces modèles sont pourtant assez satisfaisants (cf figures V-7-a) et V-8-a), pages 318 et suivantes) et sont sommairement résumés en table I .

III.1.2. Utilisation de variables discrètes

Les résultats ne sont pas tout à fait comparables au cas précédent car le programme BMD07M n'acceptant que 80 variables au maximum, nous nous sommes donc limités à 26 variables élaborées, codées en 78 modalités.

Les résultats apparaissent dans le tableau I et sur les figures V-7-a et V-8-a. On peut constater que l'on sélectionne assez rarement plusieurs modalités de la même variable. Un cas particulier intéressant est celui de la variable 30 (neige fraîche disponible) qui, en Janvier-Février, accroît la fonction discriminante de :

1.35 si $X_{30} \in$ modalité 2 de moyenne $\sim$ 5 cm

3.03 " " " " 3 " " $\sim$ 40 cm

ce qui met bien en évidence une certaine non linéarité.

Pour les autres variables, c'est en général une des classes extrêmes (1 ou 3) qui apparaît.

Les performances de la méthode sont très satisfaisantes en Janvier-Février, où le nombre de fausses alertes est même plutôt réduit. Par contre, en Mars-Avril, les performances sont nettement moins bonnes que dans le cas continu.

La réponse est en général plus brutale que dans le cas continu, car le changement de classe active brutalement un nouveau coefficient. Mais le choix de classes, effectué arbitrairement (ici en 3 classes équiprobables), n'est probablement pas optimal. Par exemple en Mars-Avril, les classes de températures de l'air à 13 H (X_{12}) sont à cheval sur le seuil de 0°C (qui appartient à la classe 2.)

III.2 - Analyse multigroupe

Dans cette approche, on considère que les mécanismes qui rendent le manteau instable et propice aux avalanches peuvent différer selon les conditions, et qu'il existe, de fait, plusieurs "types de temps avalancheux". Une telle classification (dite "génétique") a d'ailleurs été proposée par la Commission Internationale Neige et Glace de l'A.I.H.S. (DE QUERVAIN et al, 1973), mais elle n'est pas utilisable tel que car :

- elle cherche à être exhaustive et considère des types qui ne se produisent pas à Davos
- elle s'appuie sur des informations qui ne sont pas forcément présentes dans l'ensemble des variables dont nous disposons.

III.2.1. Typologie des avalanches

Nous allons donc chercher si nos données font apparaître une classification intrinsèque (au vu des seules variables disponibles, et sur l'échantillon existant) et si on peut l'obtenir de manière objective et reproductible, par des algorithmes numériques.

La méthode utilisée est essentiellement celle des nuées dynamiques.

(a) Conditions d'applications

Les données utilisées ont été :

- soit les 15 premières composantes principales des 35 variables obtenues après élimination des principales redondances, (cf. III^{ème} partie - Chap IV-1)
- soit les 15 premiers facteurs de l'AFC de ces 35 variables après codage disjonctif en 3 classes.

Les facteurs restaient systématiquement normés par leur valeur propre.

Les échantillons utilisés étaient les seules journées avalancheuses de Janvier Février ou Mars-Avril, car le but était de voir si une classification émerge "naturellement" au sein des avalanches observées.

On a systématiquement effectué :

- NI = 10 initialisations au hasard (sauf si au delà de 7 ou 8 le nombre de formes fortes devenait excessif)
- NE = 10 étalons/groupes

et en demandant :

- NG = 2, 3, 4 ou 5 groupes.

(b) Résultats pour Janvier-Février

Les critères détectent 2 ou 3 groupes (le nombre de formes fortes croît brutalement pour 4 groupes). On constate cependant que les journées sans précipitation restent toujours bien agrégées dans une forme forte (de 35 individus environ) très stable. On peut donc craindre que ce groupe, bien séparé du reste, ne masque des séparations plus fines et on les retranche pour analyser les journées avec précipitations.

Celles-ci se décomposent assez bien en 2 groupes sur les facteurs d'A.C.P.

mais le recoupement avec les résultats en A.F.C. ne sont pas parfaits. On voit là les effets des différents codages.

ⓒ Résultats pour Mars-Avril

Ici encore la classification doit se faire en 2 temps. On met d'abord en évidence la séparation entre journées avec et sans précipitation, et au sein de ces 2 classes on redivise encore en 2,d'où 4 groupes. Les critères (PFS, nombre de formes fortes) concordent généralement et on voit même apparaître des classes vides quand on demande trop de groupes.

ⓓ Utilisation d'autres méthodes

Quelques essais ont été réalisés avec les méthodes de FORGY et de JANCEY. Les résultats obtenus sont très comparables à ceux de la méthode des nuées dynamiques. Toutefois, les coûts étant comparables et les programmes disponibles étant peu fournis en traitement des résultats (formes fortes, etc...) nous n'avons pas mené de comparaison systématique.

ⓔ Conclusions sur la typologie obtenue

Les groupes obtenus sont en général mieux structurés sur les facteurs de l'AFC que sur ceux de l'ACP. Ceci montre l'intérêt du codage disjonctif, qui filtre la trop grande variabilité des précipitations par exemple, encore présente dans les résultats d'A.C.P.

Les facteurs utilisés étaient normés par leur valeur propre d'où des poids décroissants, qui ont influencé la classification. Mais des essais, effectués après les avoir renormés à 1, furent peu concluants, car sans doute donnait-on trop de poids relatif à des bruits ou à des facteurs non discriminants. On a aussi réduit le nombre de facteurs, mais les premiers sont les plus discriminants et la classification reste stable.

Si la détection du nombre de groupes semble satisfaisante, mais moins nette que dans les exemples simulés, le problème des variables classifiantes reste mal résolu.

Pour conclure, nous nous limitons à :

3 groupes pour Janvier-Février

4 groupes pour Mars-Avril

De plus, nous n'allons pas considérer les groupes formés par la méthode mais plutôt prendre comme noyaux les 3 (resp.4) formes fortes les plus caractéristiques et agréger les autres points par analyse discriminante.

III.2.2. Modèles discriminants à 2 étages (Modèles de Type II)

ⓐ Construction des types de temps avalancheux

Partant de ces formes fortes, on a agrégé les journées avalancheuses restantes et construit des groupes, utilisant cette fois une métrique de MAHALANOBIS avec sélection de variables. On a itéré plusieurs fois jusqu'à stabilisation des groupes.

Résultats pour Janvier-Février

- 3 itérations du processus ont suffi pour minimiser le nombre de mal classés
- 11 variables semblent suffisantes
- les groupes se stabilisent à AG 1 = 34
  AG 2 = 75
  AG 3 = 45
- l'analyse factorielle discriminante fournit une interprétation des groupes assez délicate à exploiter car limitée aux seules variables sélectionnées.

Nous proposons :

Ensemble des journées avalancheuses

Groupe 1	Groupe 2	Groupe 3
. Fortes précipitations journalières	. Précipitations journalières et neige récente	. Pas de précipitations et neige déjà tassée
. Faible densité superficielle	. Faible densité superficielle	. Températures relativement élevées (mais $\lesssim$ 0) et se réchauffant
. Vent fort et chasse neige (Régime NNW, NNE)	. Vent moyen ou faible (régime NW et surtout SW à SE)	. Régimes plutôt NW ou NE
. Températures de l'air basses et se refroidissant	. Températures de l'air assez basses et se refroidissant	. Neige froide et se refroidissant
. Température de la neige basse et se refroidissant	. Léger réchauffement de la neige.	. Nombreuses avalanches antérieures.
. Nombreuses avalanches antérieures		

(On pourra noter que, dans la classification proposée par BOIS et OBLED, 1976, les groupes 2 et 4 s'agrègent ici en 2).

D'ailleurs seul le groupe 3 est réellement stable dans cette classification. Les journées avec précipitations sont celles qui contiennent le plus d'incertitudes d'observation, et le codage choisi pour les précipitations est aussi très sensible.

La discrimination entre ces groupes d'avalanches ,présentée en Figure V-4-a) aboutit au modèle suivant:

N° des variables	32	39	20	8	31	10	45	23	30	6	21
% d'erreur	54%	27%	18%		8%						3%

Pouvoir discriminant axe 1 : 77 % axe 2 : 64 %

On utilise ensuite ce modèle discriminant pour associer toutes les journées avalancheuses,et surtout non avalancheuses, aux 3 types obtenus.

La répartition se fait comme suit :

	Type 1	Type 2	Type 3
Total	82	374	373
dont avalancheuses-----	34	75	45
	(23)	(42)	(34)

Dont on tire les probabilités a priori :

	Type 1	Type 2	Type 3
Probabilité du type	.099	.451	.450
A.G.	.041	.090	.054
N.G.	.058	.362	.396
A.G.|Type	.415	.201	.121

Dans les journées sans avalanches du type 1, on extrait un échantillon aléatoire, de même dans les types 2 et 3 d'où les groupes sans avalanche :

NG 1 = 21 NG 2 = 72 NG 3 = 71

On donne en Figure V-4-b) laposition des groupes avalancheux et non avalancheux.

Remarque - Les journées non avalancheuses sont relativement indépendantes entre elles (sauf dans NG 1) mais ce n'est pas le cas pour les journées avalancheuses. On a indiqué entre parenthèses le nombre de séquences avalancheuses indépendantes.

Résultats pour Mars-Avril

- 2 itérations suffisent pour stabiliser les groupes à partir des formes fortes, d'où AG 1 = 27 AG 3 = 42
  AG 2 = 43 AG 4 = 41

- 10 variables suffisent dont l'interprétation est délicate:

Ensemble des journées avalancheuses

Avec précipitations asses récentes (le jour ou la veille)

Sans précipitation en cours

Groupe 1	Groupe 2	Groupe 3	Groupe 4
. Fortes précipitations journalières ou récentes.	. Précipitations moyennes	. Pas de précipitation mais ciel couvert ou changeant	. Beau temps forte utilisation
. Faible densité superficielle	. Densité légèrement plus forte	. Densité élevée	. Densité élevée
. Vent fort et chasse neige important. (Régimes NNW, NNE)	. Vent moyen. Peu de transport. Régimes NNW mais aussi SE, SW	. Vent moyen Régimes de SSE à SSW	. Vent faible
. Températures de l'air basses et se refroidissant	. Températures plus élevées et se refroidissant légèrement	. Températures assez élevées et en hausse	. Températures positives et en hausse
. Neige froide et se refroidissant	. Neige froide mais se réchauffant	. Réchauffement de la neige . Nombreuses avalanches antérieures	. Effet thermiques cumulés

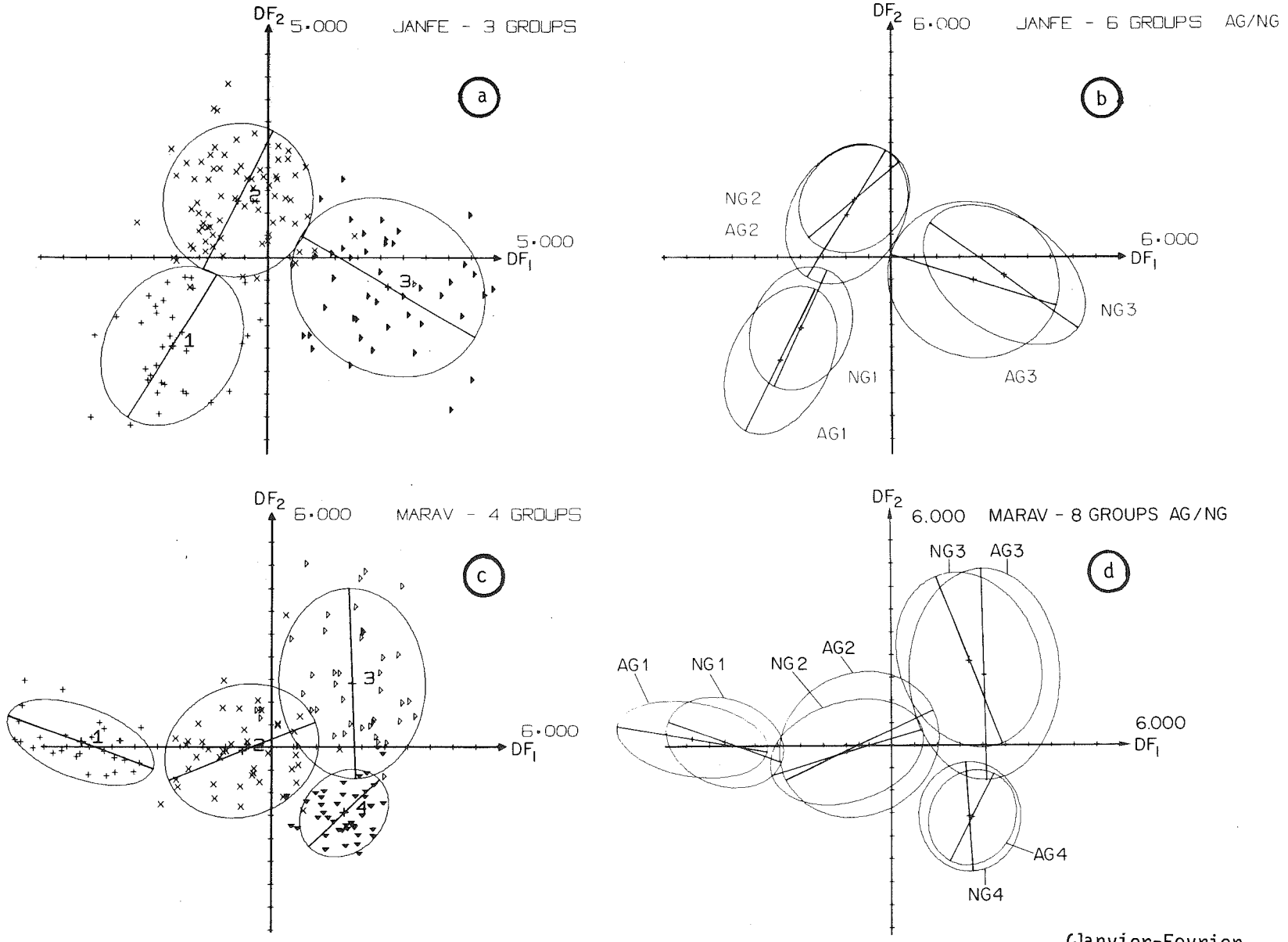


FIGURE V - 4 : a) et c) Décomposition en 3 ou 4 "Types de temps" des journées avalancheuses { Janvier-Fevrier / et Mars-Avril

b) et d) Comparaison entre journées avalancheuses et des échantillons non - avalancheux appartenant à ces mêmes "Types de temps" .

Le modèle discriminant entre types avalancheux(cf. Fig.V-4-c) utilise:

N° des variables	5	10	18	14	42	37	32	8	23	41
% d'erreur	50	24	22	16	12			6		4%

Pouvoir discriminant axe 1 : 87 % axe 2 : 62 % axe 3 : 40%

Quand on applique ce modèle à l'ensemble des journées, on obtient :

	Type 1	Type 2	Type 3	Type 4
Total	92	318	186	258
dont avalancheuses	27	43	42	41

dont on tire les probabilités a priori :

		Type 1	Type 2	Type 3	Type 4
	du type	.108	.372	.218	.302
Probabilité	AG	.032	.050	.049	.048
	NG	.076	.322	.169	.254
	AG/type	.293	.116	.226	.159

Là encore, on extrait des journées sans avalanche les échantillons respectifs :

NG 1 = 31 NG 2 = 56 NG 3 = 43 NG 4 = 50

Les positions respectives des groupes avec et sans avalanche apparait sur la figure V-4-c) .

(b) Modèles avalanches - non avalanche, par types de temps:

On peut résumer dans un tableau les modèles avalanches - non avalanche au sein des types de temps. On propose les interprétations suivantes .

En Janvier-Février:

. Type 1 : Le danger d'avalanche croît avec les chutes de neige et le transport par le vent, qui définissent d'ailleurs le type. Mais il est aussi accru par la formation de plaque (variable 35) par un refroidissement plus marqué de la neige (var. 45, 46) et par la fragilité intrinsèque du manteau (var. 49,50).

. Type 2 : On constate que la discrimination n'est pas très bonne, et que les avalanches correspondent à des précipitations cumulées plus fortes, à un peu de plus de vent et de chasse-neige, à des températures plus basses, une neige plus froide et éventuellement des plaques.

. Type 3 : Ici la discrimination est un peu meilleure. Les avalanches correspondent à un temps plus froid avec une neige plus légère, un vent de NW plus soutenu et un manteau un peu plus fragile.

On notera que c'est au sein du type I, concomittant à de fortes précipitations, que la discrimination est la plus facile, d'où le succès des auteurs qui se limitent à ce cas-là.

En Mars-Avril:

. Type I : Ici aussi, c'est plutôt l'accroissement des variables définissant le type (précipitations, vent et transport) qui accroissent le danger.

. Type II : La quantité de neige récente disponible et le réchauffement plus marqué après la séquence de précipitation augmentent le danger ainsi que la fragilité du manteau.

. Type III : Ici aussi le réchauffement de la neige intervient mais dans le sens de la stabilité. Par contre la variable importante est la fragilité antérieure du manteau (var.50) et la densité superficielle (4) qui indique peut-être des plaques.

. Type IV : La fragilité antérieure du manteau est encore le facteur le plus aisément interprétable en plus des variables définissant le type (températures élevées et effets thermiques cumulés).

On peut conclure que, si l'on discrimine bien les types de temps, on discrimine mal entre avalanche et non avalanche dès que les causes ne sont pas d'origine météorologique, mais plutôt structurale.

(c) Mise en oeuvre de ces modèles (cf. figures V-7-b et V-8-b), pages 318 et suivantes)

Elle comporte 2 étapes :

- la détermination du type de temps le plus probable
- puis, au sein du type de temps choisi, le calcul de la probabilité d'avalanche.

La règle de décision est un peu plus compliquée que pour un modèle à 1 seul étage mais peut se construire aisément. On admet, ce qui se vérifie dans la réalité que la journée X, à classer, a une probabilité proche de 1 d'appartenir au type de temps TT_k et une probabilité nulle d'appartenir à TT_ℓ, $\ell \neq k$. Dans ce cas, il n'y a pas d'erreur possible sur le type de temps.

La règle bayesienne revient à déclarer qu'il a risque si

$$C_{A_v} \times \pi\left(A_v \mid TT_k\right) \times Pr_k\left(A_v \mid X\right) > C_{\overline{A_v}} \times \pi\left(\overline{A_v} \mid TT_k\right) \times Pr_k\left(\overline{A_v} \mid X\right)$$

avec C_{A_v} coût de non alerte , $C_{\overline{A_v}}$ coût de fausse alerte

$Pr_k(A_v|X) = 1 - Pr_k(\overline{A_v}|X)$ probabilité conditionnée par X, et par le modèle avalanche/non avalanche du type de temps TT_k.

$\pi(A_v|TT_k)$ = probabilité a priori d'avalanche sachant que l'on est dans le type TT_k.

L'alerte se fera donc, dans le type TT_k, dès que le modèle fournira :

$$Pr_k(A_v|X) > \frac{1}{1 + \frac{C_{A_v}}{C_{\overline{A_v}}} \times \frac{\pi(A_v|TT_k)}{\pi(\overline{A_v}|TT_k)}}$$

Or $\pi(A_v) = \pi(A_v|TT_{Rk}) \cdot \pi(TT_{Rk}) + \pi(A_v|\overline{TT_{Rk}}) \times \pi(\overline{TT}_{Rk})$

de même pour $\pi(\overline{A_v})$, d'où finalement

$$Pr_{Rk}(A_v|X) > \frac{}{1 + \frac{C_{Av}}{C_{\overline{Av}}} \times \frac{\pi(A_v) - \pi(A_v|TT_{Rk}) \cdot \pi(TT_{Rk})}{\pi(\overline{A_v}) - \pi(\overline{A_v}|\overline{TT_{Rk}}) \cdot \pi(\overline{TT}_{Rk})}}$$

Exemple : Mars-Avril - Type 4

$\pi(A_v) = .179 \Rightarrow \pi(\overline{A_v}) = .821$

$\pi(TT_4) = .302 \Rightarrow \pi(\overline{TT_4}) = .698$

$\pi(A_v|TT_4) = .159$ $\quad \pi(A_v|TT_4) = .812$

$C_{Av} = 10$ $\quad C_{\overline{Av}} = 1$

Le seuil d'alerte est donc de : .162 . Il passe à .28 si le rapport des coûts est de 5, et .66 s'ils sont égaux.

III.2.3. Comparaison entre modèles à 1 ou 2 étages, et entre variables continues ou discrètes

(a) Modèle à 1 étage

La modélisation à 2 niveaux suppose implicitement qu'il y a des types de temps distincts et que dans chaque type de temps, il y a un processus de déclenchement différent.

Sur le plan pratique, cela permet d'utiliser un nombre élevé de variables sans compromettre trop la qualité des estimateurs : les variables du modèle AG1/NG1 diffèrent de celles de AG2/NG2 mais le nombre de variables du modèle AG1/NG1 est petit par rapport aux nombres de degrés de liberté.

Dans le modèle à 1 étage (où l'on traite en même temps tous les groupes avalanche et non avalanche) on suppose plus fortement l'hypothèse d'un continuum, les mêmes variables étant toujours utilisées conjointement. Mais la nécessité d'utiliser beaucoup de variables dans le même modèle conduit à se rapprocher dangereusement des effectifs de certains groupes.

Enfin, l'interprétation physique devient quasi impossible. Sur les figures V-5 on a reporté, pour Janvier et Février les probabilités d'avalanches obtenues par l'un et l'autre modèle. On constate que le modèle à 2 étages présente un léger avantage. En fait, il serait certainement plus marqué si l'on disposait des variables nécessaires pour les modèles avalanche - non avalanche.

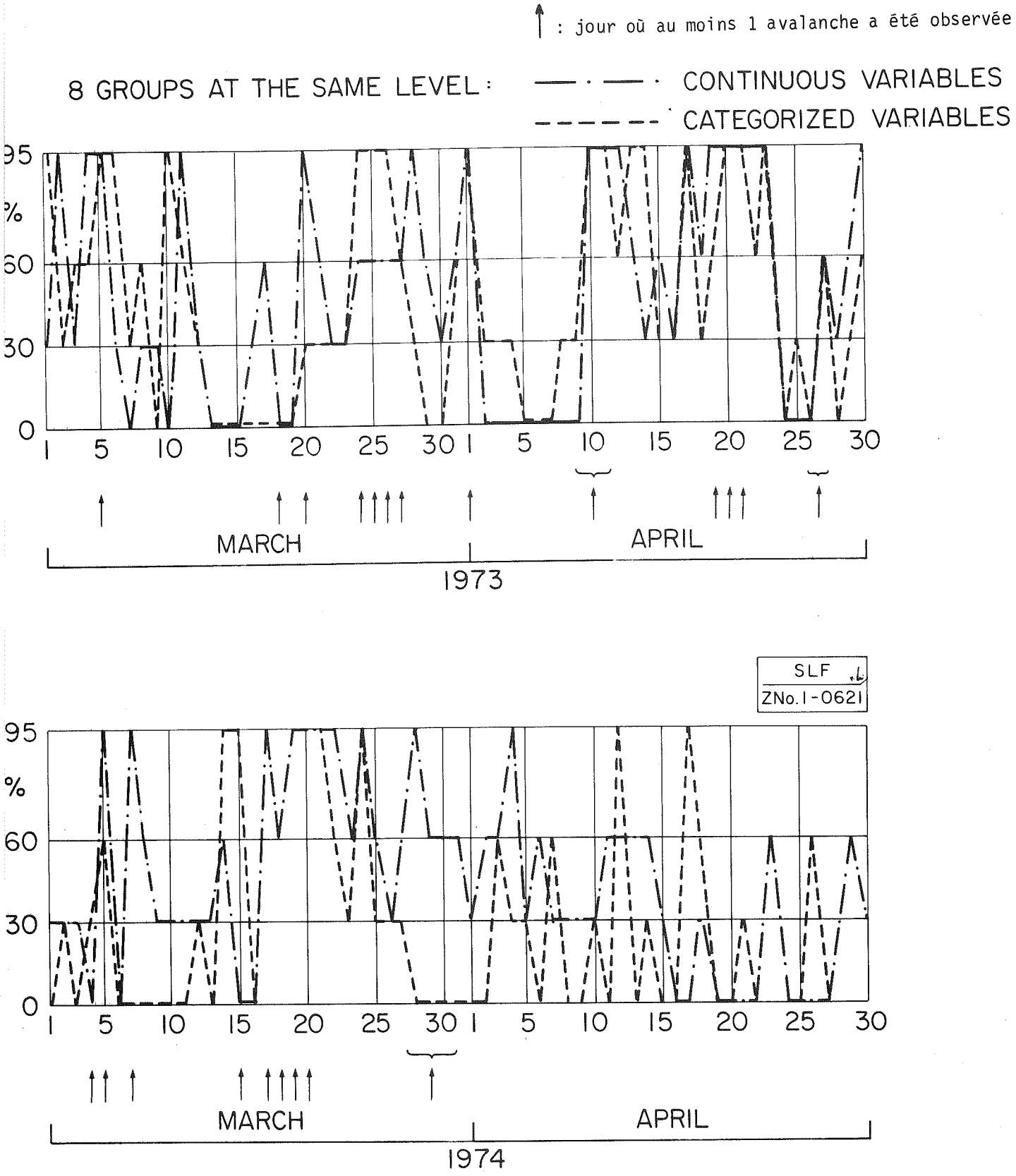


FIGURE V - 5 : Comparaison, sur une période "test" , entre l'approche par variables continues et celle par variables discrètes .

Modèle à 1 seul niveau sur 8 groupes : 4 types de temps comportant chacun { 1 groupe avalanche / 1 groupe sans }

Exemple : Sur le type de temps 1 de Janvier-Février, il faut, avec le modèle à 1 étage, plus de 10 variables pour avoir 69% de bien classés entre AG 1 et NG 1 alors que le modèle sur AG 1 et NG 1 pris isolément donne 93% de bien classés avec 5 variables seulement.

Le modèle à 2 étages explore donc mieux la structure du nuage.

	AG1	NG1	AG2	NG2	AG3	NG3	
AG 1	30	4					5 var.
NG 1	1	20					
AG2			55	20			6 var.
NG 2			12	60			
AG 3					34	11	6 var.
NG 3					8	63	

AG1	NG1	AG2	NG2	AG3	NG3	
22	11	1				. Janvier Février 15 variables
4	15		2			
	3	23	46	1	2	
		1	71			
			3	19	23	
			1	3	67	

2 étages

1 étage

	AG1	NG1	AG2	NG2	AG2	NG3	AG4	NG4	
AG 1	24	3							6 var.
NG 1	4	27							
AG 2			30	13					4 var.
NG 2			12	44					
AG 3					35	7			6
NG 3					7	36			
AG 4							34	7	3
NG 4							18	32	

AG1	NG1	AG2	NG2	AG3	NG3	AG4	NG4	
23	4							Mars-Avril 15 variables
6	25							
	2	33	7	0	1			
	1	13	42					
		1	1	29	8	1	2	
				9	34			
						31	10	
					1	16	33	

(b) Utilisation de variables discrètes

Comme nous l'avons signalé, les variables proposées ne sont pas les mêmes que dans le cas continu. Avec 21 variables binaires, issues de 16 variables continues, on arrive à des performances légèrement inférieures au cas continu (cf figure V-5).

III.3 - Analyses non-paramétriques

Les résultats sont ici un peu plus succincts car compte-tenu de la faiblesse relative du support théorique, on donne les résultats obtenus expérimentalement sur un fichier assez réduit.

III.3.1. Méthode de FIX et HODGES (Modèles de Type III)

Nous l'avons appliquée sur 15 facteurs principaux normés par $\sqrt{\lambda_k}$ ou par 1.0, avec un nombre variable de voisins. (cf. Figures V-7-c) et V-8-c), pages 318 et suivantes).

La formule approchée du paragraphe I.5 nous suggère $\sim$ 4 voisins, et nous avons fait des essais avec 5, 10, 15, 20 et 40 voisins.

Résultats pour Janvier-Février

Les meilleurs résultats sont obtenus avec 5 ou 10 voisins (on a choisi 10 car les fluctuations sont un peu plus significatives) et avec les données normées par $\sqrt{\lambda_k}$. Ceci avantage légèrement les journées avec précipitations, mais dans ce bimestre 2/3 des journées avalancheuses sont concomittantes ou consécutives à des précipitations. Le codage utilisé pour rendre les résultats comparables aux autres analyses a été :

Nombre de journées avalancheuses sur les 10 voisins	0 ou 1	Probabilité 0%
	2 ou 3	30%
	4 à 6	60%
	$>$ 6	100%

En fait, la règle bayesienne stricte conduit à dire avalanche dès qu'il y a 2 voisins avalancheux pour K = 10 (resp.5 pour K = 40)

On remarquera que le K optimum approché est proche (quoique légèrement inférieur) de l'optimum expérimental K = 5 à 10, et que la méthode du plus proche voisin (1 seul voisin est considéré) n'est pas incohérente non plus.

Probabilités pour Mars-Avril

Les résultats sont plus décevants, spécialement pour l'année 1974. On a vu que les journées avalancheuses se répartissent ici en 2 populations à peu près comparables selon qu'elles sont ou non associées à des précipitations. Cela explique sans doute qu'il n'y ait pas de métrique supérieure à une autre, les données normées à 1.0 n'étant que très légèrement préférables.

Il a de plus fallu utiliser 20 ou 40 voisins pour avoir des résultats un peu cohérents.

Par contre, on a pu constater que pour beaucoup de séquences avalancheuses par exemple du 16 au 23 avril 1973, ou du 16 au 17 Mars 1974 les distances nécessaires pour rassembler 40 voisins étaient anormalement élevées.

Si on donne l'histogramme des distances du plus proche voisin pour les journées de 1973 et 1974 (dans des coordonnées normées à 1) on trouve :

d	.05	.05-.1	.10-.15	.15-.20	.20-.25	.25-.30	.3
n	0	27	43	25	15	5	7
%	0	22	35	20	12	4	6

Or du 16 au 23 Avril 1973, on a la séquence:

.237 , .234 , .231, .328, .248, .285, .232, .280

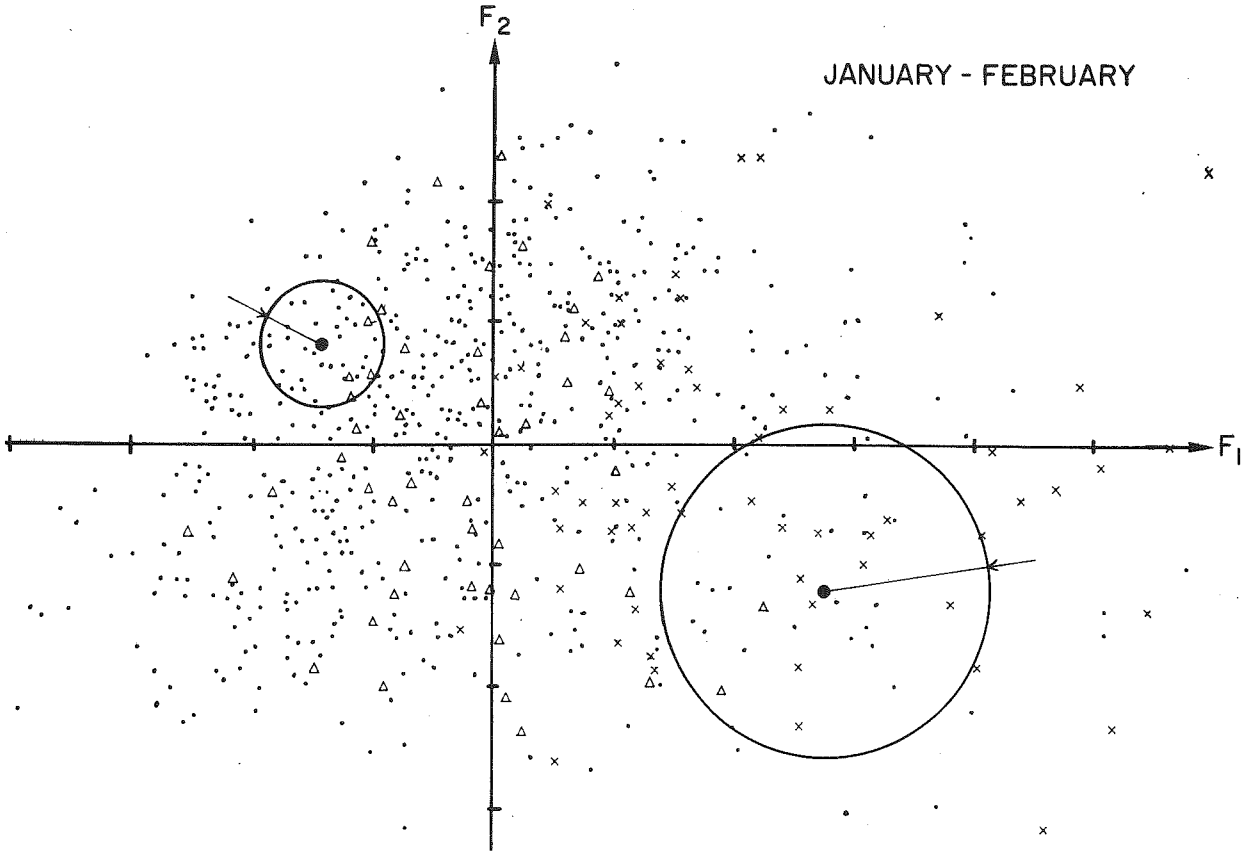
De même en Mars 1974, du 16 au 19 on relève : .347, .611, .252, .127

Cela signifie que l'on est dans une zone très mal représentée dans l'échantillon d'ajustement et peu dense. Il y a aussi de fortes chances pour qu'en plus, les points voisins soient mal distribués dans la sphère, ce qui réduit encore la validité de l'estimation locale.

III.3.2. Ajustement d'un modèle local

Pour chacun des points tests comportant au moins 8 journées avalancheuses parmi 40 voisins, on a effectué une analyse discriminante linéaire, avalanche/non avalanche, en utilisant 5 des 50 variables élaborées.

Même pour le bimestre Janvier-Février, où les voisins semblent représentatifs les résultats ne sont guère concluants. Il faut reconnaître que le problème est assez mal posé car on n'introduit pas d'information nouvelle au niveau de la discrimination, les 50 variables élaborées étant contenues dans les 15 facteurs qui ont servi à définir le voisinage.



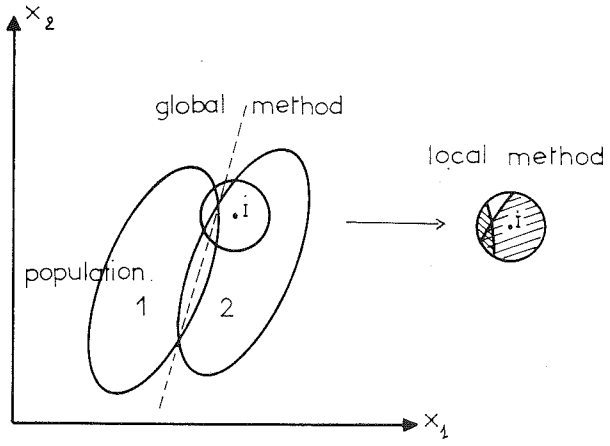


FIGURE V - 6 : Exemple de sélection par la méthode de la "boule" des 40 voisins les plus proches du jour j, et ajustement éventuel d'un modèle local dans la boule associée au jour j .

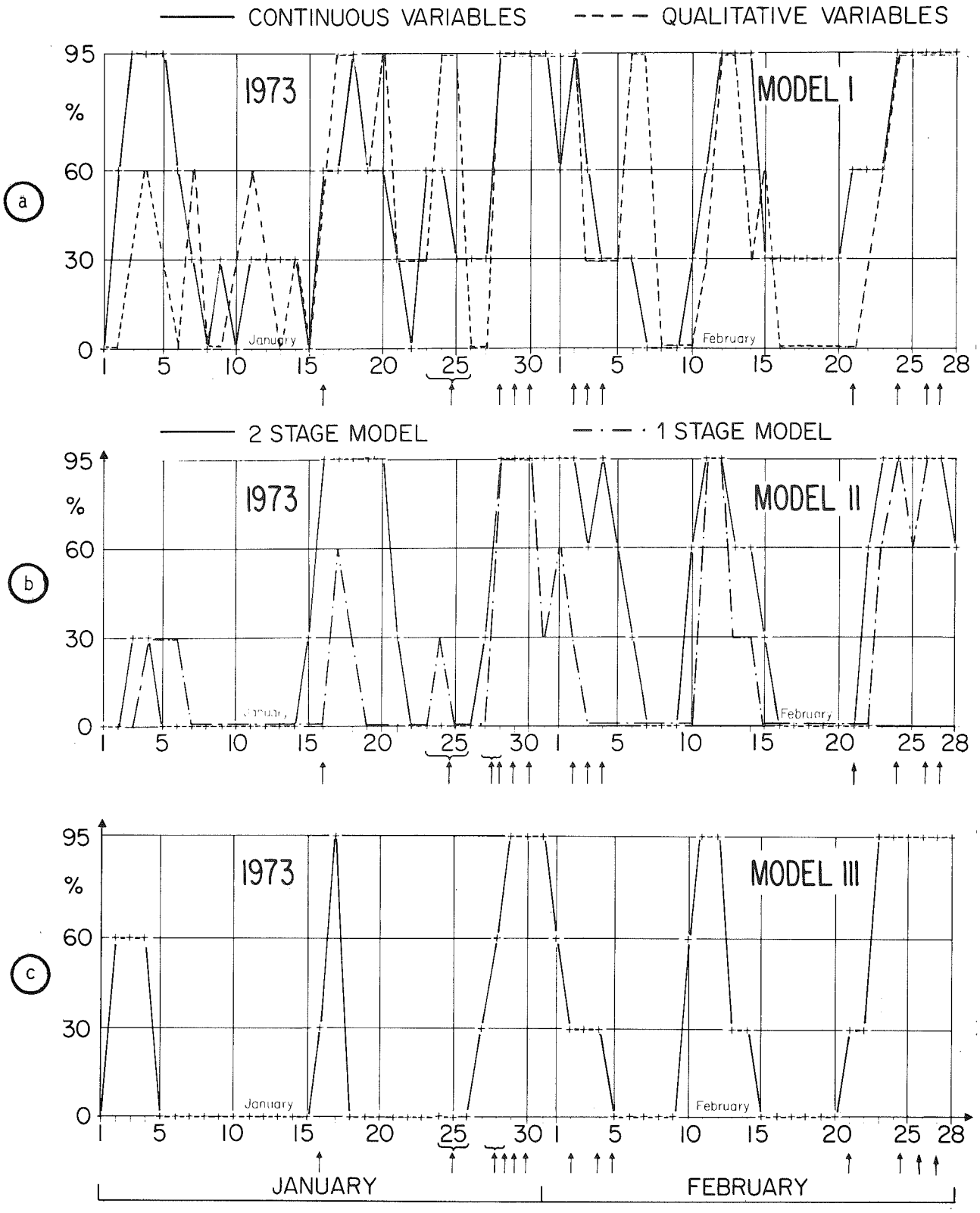
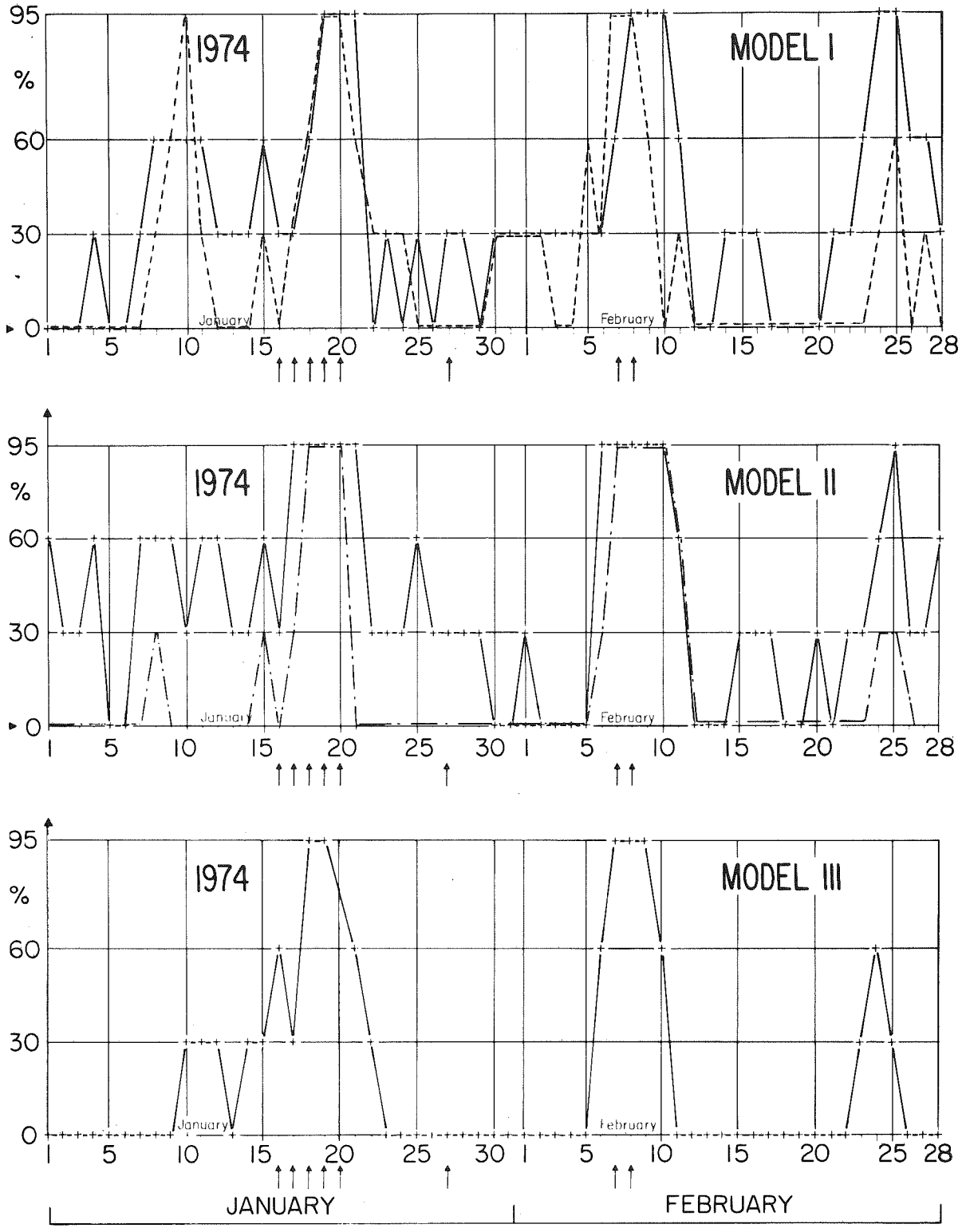


FIGURE V - 7 : Résultats des différents modèles pour les 2 années "test" 1973 et 1974 .

(Mois de Janvier et Février)

↑ : journée avec au moins 1 avalanche observée .



a) - Comparaison entre l'approche continue et discrète pour les modèles de type I .

b) - Comparaison entre le modèle à 2 étages (Sélection d'un type de temps puis modèle avalanche/non avalanche au sein du même type), et le modèle à 1 seul étage.

c) - Modèle de type III .

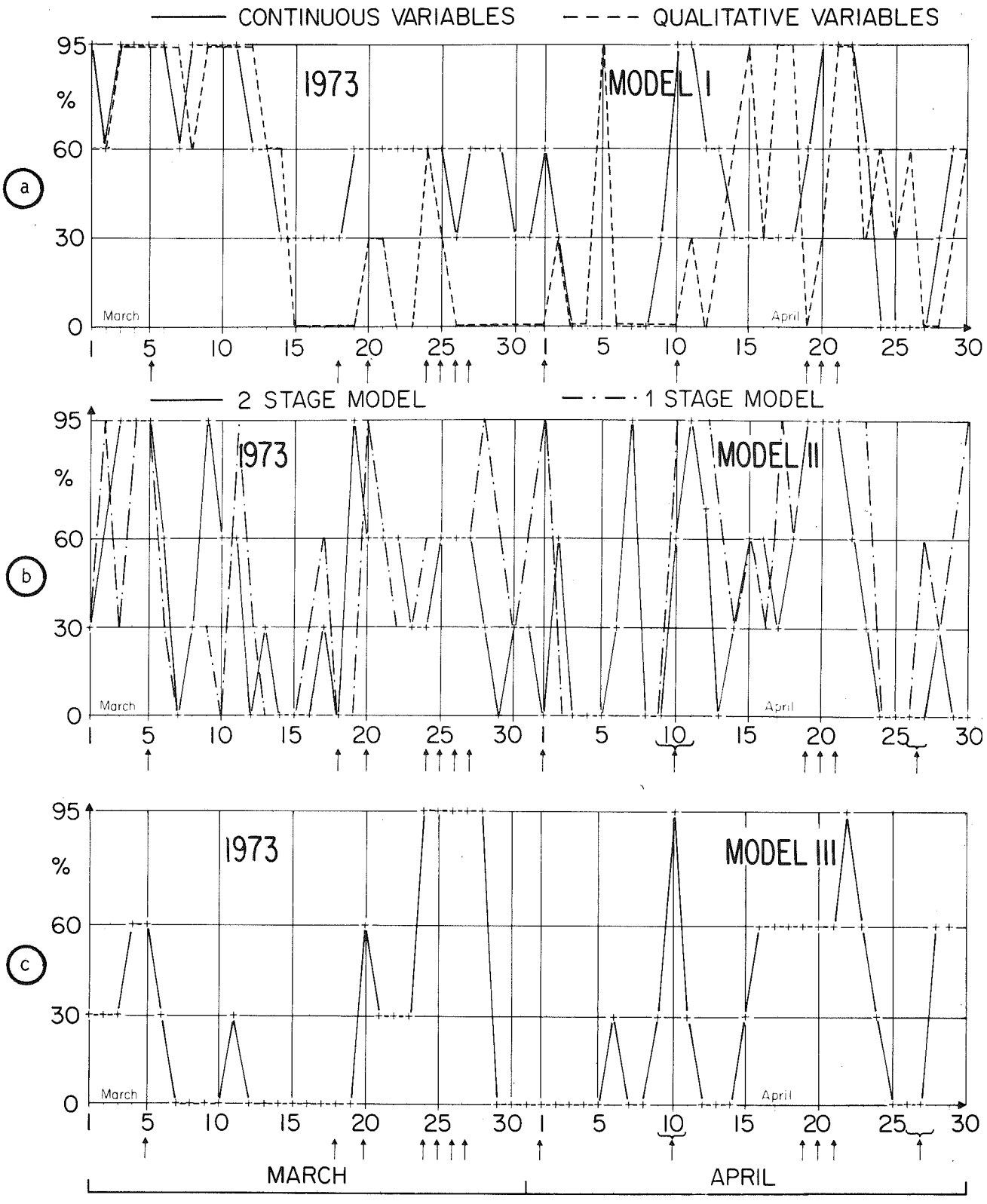
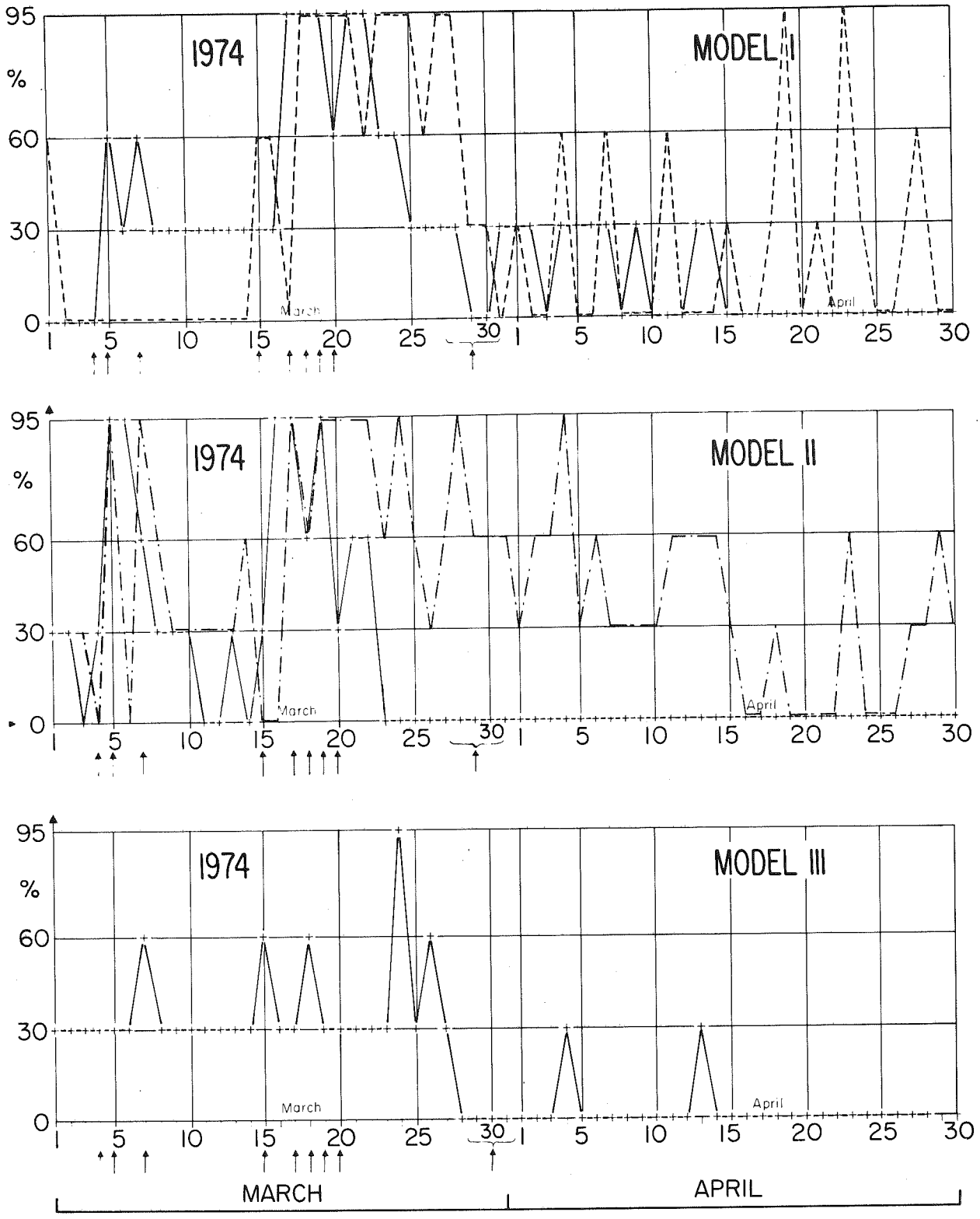


FIGURE V - 8 : Mêmes résultats qu'en figure V-7 pour les mois de Mars - Avril .


95
%
60
30
0
1974
MODEL I
March
April
1 5 10 15 20 25 30 1 5 10 15 20 25 30
95
%
60
30
0
1974
MODEL II
March
April
1 5 10 15 20 25 30 1 5 10 15 20 25 30
95
%
60
30
0
1974
MODEL III
March
April
1 5 10 15 20 25 30 1 5 10 15 20 25 30
MARCH
APRIL

III.4 - Conclusions sur la prévision des avalanches

L'analyse discriminante à deux groupes (avalanche - non avalanche) fournit des résultats satisfaisants surtout en Janvier-Février, quand il y a un type d'avalanche dominant. Par contre, en Mars-Avril où l'on a au moins deux types distincts, qui sont arbitrairement confondus dans le modèle, la discrimination tend à être plus floue. D'autre part, la réponse de ces modèles tend à être assez molle, avec beaucoup de valeurs intermédiaires qui risquent d'être gênante dans l'utilisation opérationnelle.

Quand le problème est relativement simple, comme en Janvier-Février, l'utilisation de variables codées discrètement ne réduit pas sensiblement voire même améliore la décision. Ceci est très intéressant pratiquement et montre que l'information significative est déjà contenue dans des mesures assez grossières et facilement accessibles : Le vent en Beaufort, l'état du ciel en 3 ou 4 classes, etc... sont presqu'aussi utiles qu'une vitesse en mètres-seconde ou un rayonnement global en 1/10 de Joule. Ceci permet d'envisager l'extension de ces méthodes à de nombreuses régions, sous réserve de bonnes observations d'avalanche.

Sur le plan théorique des recherches sont cependant nécessaires pour déterminer le codage optimum c'est-à-dire la réalisation de classes, en nombre limité, maximisant la discrimination.

L'analyse discriminante multigroupe est à 2 niveaux (par type de temps), présente de nombreux avantages :

- elle se rapproche plus de la réalité physique du phénomène et s'interprète d'ailleurs beaucoup plus finement, ce qui est appréciable pour un utilisateur
- le découpage en types de temps et leur modélisation séparée, outre sa signification physique, revient aussi à effectuer une désaisonnalisation que le découpage en bimestres n'avait qu'imparfaitement effectué en Mars-Avril. On remarquera que cette desaisonnalisation n'a plus lieu quand on travaille à 1 seul niveau, d'où les réponses anormalement élevées du modèle à 1 niveau en Avril 1974.

Parmi ses inconvénients, on peut citer :

- Le fait que les types de temps soient un peu trop influencés par les variables disponibles et la métrique choisie
- La nécessité d'ajuster, sur le même nombre de données, un nombre beaucoup plus élevé de paramètres. Ceci est encore plus net si on emploie des méthodes quadratiques. Or les performances en test ne semblent pas s'en ressentir, ce qui pourrait s'expliquer par la meilleure représentation de la réalité physique, qui viendrait compenser l'échantillonage plus réduit.

Des améliorations sont à apporter au niveau de la typologie, en éliminant peut être les variables d'état du manteau qui ne seraient prises en compte que dans les modèles avalanche-non avalanche. En effet, si les "situations" au sens nivométéorologique ont une échelle de temps de quelques jours, les variables d'état elles ont une échelle beaucoup plus longue. Il est certain même que la saison d'enneigement dans son ensemble peut avoir une originalité propre, même si le manteau voit défiler des situations déjà rencontrées dans d'autres hivers. Cette non stationnarité d'un hiver à l'autre peut être grossièrement prise en compte en déplaçant le seuil d'alerte. Mais si on admet que chaque hiver est particulier pour l'ensemble des variables il devient difficile d'y appliquer un modèle statistique. En fait, seules certaines variables présentent cette particularité et doivent être prises en compte séparément.

L'analyse discriminante non paramétrique est plutôt moins concluante. Elle est certes robuste quand on passe à l'échantillon test, et facile à mettre en oeuvre. En contre-partie, elle est relativement peu étayée sur le plan théorique:

- les incertitudes d'échantillonnages sont difficiles à apprécier empiriquement (car le rayon de la boule varie) et inextricables mathématiquement
- elle ne prend pas en compte le rayon de la boule nécessaire pour rassembler les voisins, ni leurs distances au point à classer. Or ses performances diminuent dans les zones peu denses et aux frontières du nuage d'ajustement,
- matériellement, elle suppose un fichier d'ajustement particulièrement fourni.

De plus, pour être non paramétrique, elle n'en est pas moins liée à des hypothèses. Le choix de la distance et de l'espace de travail est absolument critique et il n'y a pas à proprement parler d'ajustement d'un modèle puisque cette distance ou les variables utilisées, ne sont pas sélectionnées au préalable pour maximiser la discrimination.

Notre opinion est qu'il faut ici encore travailler à deux niveaux. La sélection des voisins les plus proches se ferait à l'aide de quelques variables caractérisant la situation météorologique récente, tandis que la discrimination avalanche/ non avalanche se ferait par un modèle linéaire adapté aux journées présentant le même type de situations. Ce serait en quelque sorte une version plus souple des modèles de type II qui remplacerait la définition rigide des types de temps et des modèles correspondants mais fixes, avalanche/ non avalanche.

Mais il ne nous parait plus souhaitable de mélanger dans une distance arbitraire des variables disparates, caractérisant la situation météorologique ou l'état du manteau, et considérées arbitrairement au même niveau.

En conclusion, nous pensons avoir démontré le potentiel des méthodes (statistiques) discriminantes dans la prévision des avalanches.

Ce sont d'ailleurs les seules possibles, en l'absence d'un modèle déterministe global adapté à toutes les situations, et leur usage tend à se répandre.

Ces méthodes sont apparues au moins aussi bonnes qu'un présisioniste expérimenté, si tant est qu'un contrôle objectif soit possible. Elles ne prétendent d'ailleurs qu'à quantifier sa démarche et à se rapprocher le plus possible, comme lui, d'une réalité physique fort diverses. Elles ont l'avantage de pouvoir se transposer en différents points sous réserve de disposer des données.

Parmi celles-ci les observations d'avalanche sont absolument critiques, tout biais systématique (retard d'observation, défaut d'observation, etc...) pouvant conduire à des conclusions fantaisistes. Il est donc nécessaire de mettre sur pied des protocoles d'observation rigoureux, incluant éventuellement des tentatives systématiques de déclenchements artificiels.

Néanmoins, l'utilisation de fichiers ne rassemblant pas toutes ces caractéristiques est déjà encourageante, comme les systèmes utilisés à Davos ou Fort Collins en font foi .

CHAPITRE IV

INTERPOLATION OPTIMALE DES CHAMPS
(au sens climatologique)

Une première application de l'analyse structurale des données régionalisées peut consister en une typologie des différentes réalisations observées comme nous l'avons vu au chap.II de la IVème partie. Néanmoins les buts essentiels sont soit l'estimation de valeurs manquantes, soit la cartographie des champs, soit encore l'évaluation d'intégrales sur une surface donnée (exemple : pluie moyenne sur un bassin). Et une manière de résoudre les 2 derniers consiste à évaluer la variable sur une grille très fine, ce qui nous ramène au premier cas.

Enfin, seule l'estimation de valeurs manquantes peut faire l'objet d'une vérification objective, par exemple sur un réseau test, (alors qu'il n'existe pas de mesure réelle de l'intégrale sur un bassin).

De plus, nous nous limitons aux seules méthodes que nous appelons "climatologiques" au sens où elles utilisent les propriétés statistiques déduites de l'observation d'un grand nombre de réalisations. (Par opposition aux méthodes spatiales qui n'en envisagent qu'une).

IV.1 - Interpolation locale (Méthode de GANDIN)

Nous appellerons ainsi une méthode qui ne se sert que des points mesurés les plus proches du point à interpoler pour effectuer l'interpolation. Elle utilise donc seulement un "voisinage", même si ce voisinage peut correspondre à l'ensemble des points mesurés. Elle peut prendre diverses formes:

IV.1.1. Ecriture du système d'interpolation (GANDIN 1965 et 1970)

Si on considère que l'on a un processus $X(\omega,t)$ où ω indique la réalisation, et t les coordonnées géographiques ($t \rightarrow x$ ou (x,y) ou (x,y,z)), on se propose, à l'aide des observations en P points $\{t_1, t_2 \dots t_j \dots t_P\}$ de déterminer la valeur au point t_0.

On suppose connue en tout point la moyenne du processus $\langle X(\omega,\tau)\rangle$, en particulier au point t_0, et on peut supposer que les X sont centrés et représentent en fait des écarts à cette moyenne.

On va donc chercher une estimation linéaire de $X(\omega,t_0) = X_0$ par :

$$\hat{X}_0 = \sum_{j=1}^{P} \lambda_j X_j$$

en imposant à cette reconstitution (donc aux λ_j) d'être "optimale" au sens où, en

espérance mathématique, l'erreur quadratique moyenne est minimale :

$$\delta_o = \langle (X_o - \hat{X}_o)^2 \rangle = E[(X_o - \hat{X}_o)^2] \quad \text{minimale}$$

Pour cela on développe l'expression :

$$\delta_o^2 = E[X_o^2] - 2\sum_{j=1}^{P} \lambda_j E[X_o . X_j] + \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i . \lambda_j . E[X_i . X_j]$$

et on dérive par rapport aux λ_j d'où le système :

$$j = 1 \ldots P \left\{ \frac{\partial}{\partial \lambda_j} (\delta_o^2) = -2 E[X_o . X_j] - 2\sum_{i=1}^{P} E[X_i . X_j] \right.$$

Les valeurs X étant centrées, $E[X_i . X_j]$ n'est autre que la covariance entre X_i et X_j soit σ_{ij}, d'où le système :

$$\begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1j} & \cdots & \sigma_{1P} \\ \vdots & & \vdots & & \vdots \\ \sigma_{j1} & \cdots & \sigma_{jj} & \cdots & \sigma_{jP} \\ \vdots & & \vdots & & \vdots \\ \sigma_{P1} & \cdots & \sigma_{Pj} & \cdots & \sigma_{PP} \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_j \\ \vdots \\ \lambda_P \end{bmatrix} = \begin{bmatrix} \sigma_{10} \\ \vdots \\ \sigma_{j0} \\ \vdots \\ \sigma_{P0} \end{bmatrix} \quad \text{ou} \quad \Sigma . \vec{\Lambda} = \vec{\Sigma}_o$$

Cette approche peut être :

(a) soit climatologique : dans ce cas, on se place en 1 point fixé du champ, et on suppose connu un grand nombre de réalisations.

Le système obtenu n'est rien d'autre que le système classique de régression multiple, mais écrit en espérance (climatologique) et non à l'aide des estimateurs classiques de la covariance (sur N observations, par exemple).

En effet, le problème est, au second membre d'estimer la covariance entre t_j et t_o en l'absence de mesures en t_o. On compense alors cette absence d'information par l'introduction d'un modèle, en général :

$$\sigma_{ij} = f(d_{ij}) \quad d_{ij} \text{ distance entre } t_i \text{ et } t_j$$

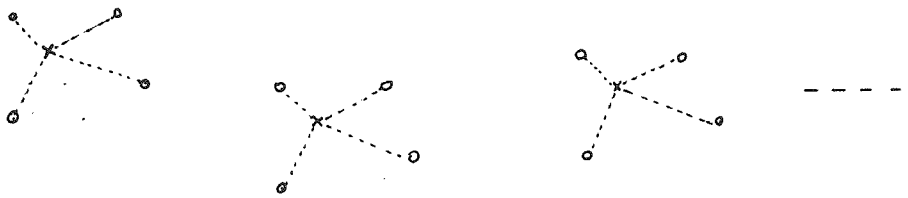
C'est le corrélogramme (ou la fonction de covariance climatologique), obtenu en ajustant un modèle théorique (cf IVème partie - chapitre III) aux nuages des σ_{ij} estimés sur les N observations, pour tous les couples de stations disponibles (cf D. CREUTIN, 1979).

En fait, on se limite à quelques modèles simples de fonctions définies positives, et on choisit sur la forme générale, l'allure à l'origine, etc...

Et c'est à l'aide de ce modèle que l'on remplit l'ensemble du système. (Mais on pourrait très bien remplir la matrice Σ avec les valeurs de σ_{ij} estimées directement).

(b) soit spatiale : on peut imaginer alors que l'on va se placer en divers points t_o, avec pour chacun un voisinage ayant toujours la même structure

géographique, et on veut qu'en espérance dans le champ, l'erreur quadratique soit minimale



Dans ce cas, le système est écrit en espérance spatiale : $\sigma_{ij} \rightarrow C_s(d_{ij})$
Là encore, pour pouvoir remplir le système il faut introduire un modèle d'autocovariance, et pour estimer ce dernier, on a besoin d'hypothèses très fortes d'isotropie et d'homogénéité.

C'est pour s'en libérer partiellement que l'on a été conduit à formuler le krigeage simple (cf DELHOMME J.P., 1976) qui suppose seulement que les variations du processus $X(t) - X(t+\Delta t)$ sont stationnaires (Hypothèses intrinsèque).

Dans la pratique, la méthode a surtout été appliquée dans son sens climatologique. La stationarité du processus brut n'est pas absolument requise, dans la mesure où on travaille sur les valeurs centrées (en supposant connue la moyenne, ou "normale climatologique", qui n'est pas forcément constante). De même si les variances ne sont pas constantes, on peut travailler sur les valeurs centrées réduites.

Et on effectue ainsi toutes les transformations nécessaires pour trouver un processus homogène et isotrope auquel ajuster un modèle. Le problème est éventuellement de trouver $m_o = E[X_o]$ ou $\sigma_{oo} = E[X_o^2]$.

Remarquons aussi que :

- La méthode est une méthode d'interpolation : si on remplace t_o par un point t_j quelconque on voit, sur $\vec{\Lambda} = \vec{\Sigma}_o \cdot \Sigma^{-1}$ que dans ce cas on a $\lambda_k = 0$ $\forall k \neq j$ et $\lambda_j = 1$ donc $\hat{X}_j = X_j$.
- Dans la formule de l'erreur quadratique moyenne :

$$\delta_o^2 = \sigma_{oo}^2 - 2 \sum_{j=1}^{P} \lambda_j . \sigma_{jo} + \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_i \lambda_j \sigma_{ij}$$

le dernier terme vérifie : $\sum_{i=1}^{P} \lambda_i \sigma_{ij} = \sigma_{jo}$

d'où l'expression : $$\delta_o^2 = E[(X_o - \hat{X}_o)^2] = \sigma_{oo}^2 - \sum_{j=1}^{P} \lambda_j \sigma_{jo}$$

IV.1.2. Diverses variantes (*)

a) Une première variante apparaît quand la moyenne est inconnue, ou difficile à interpoler mais quand le voisinage utilisé est suffisamment petit pour

qu'on admette que la moyenne m inconnue est la même partout dans le voisinage. On travaille alors en valeurs brutes X_j, et on impose que les poids λ_j dans :

$$\hat{X}_o = \sum_j \lambda_j X_j$$

satisfassent la relation en espérance :

$$m = E[X_o] = \sum \lambda_j E[X_j] = \sum \lambda_j m$$

d'où la contrainte : $\sum_{j=1}^{P} \lambda_j = 1$

Parallèlement, on peut écrire :

$$X_j = m + x_j \qquad \text{d'où} \qquad \hat{X}_o = \sum_{j=1}^{P} \lambda_j x_j + m$$

et l'erreur se ramène à :

$$\delta_o^2 = E\left[(X_o - \hat{X}_o)^2\right] = E\left[(m + x_o - m\sum_j \lambda_j - \sum_j \lambda_j x_j)^2\right]$$

d'où la même forme que précédemment, les x_j étant les valeurs centrées.

L'optimisation, compte tenu de la contrainte $\sum \lambda_j = 1$ porte sur le critère :

$$\delta_o + \mu\left(\sum \lambda_j - 1\right)$$

où μ est un multiplicateur de Lagrange d'où le système :

$$\sum_{i=1}^{P} \lambda_i \sigma_{ij} + \mu = \sigma_{jo}$$
$$\sum_{i=1}^{P} \lambda_i = 1$$

$$\begin{bmatrix} \sum & \vdots & 1 \\ \cdots & \cdots & \cdots \\ 1 \cdots 1 & \vdots & 0 \end{bmatrix} \begin{bmatrix} \vec{\Lambda} \\ \cdots \\ \mu \end{bmatrix} = \begin{bmatrix} \vec{\Sigma}_o \\ \cdots \\ 1 \end{bmatrix}$$

La matrice $\sum$ reste la matrice de covariance et l'erreur devient :

$$\delta_o^2 = \sigma_{oo}^2 - \sum_{j=1}^{P} \lambda_j \sigma_{jo} + \mu$$

On notera que c'est cette démarche, appliquée à une moyenne qui n'est plus forcément constante mais polynômiale, qui constitue la base du krigeage universel (cf DELHOMME, 1976). De même que dans ce cas, le problème est d'identifier le variogramme sous-jacent, le problème, ici est d'identifier $\sum$ si la moyenne est inconnue.

b) Une autre variante consiste à admettre que chaque station a une variabilité propre, ou une part de sa variance ε_i^2, qui n'est pas liée à la structure générale du champ mais à des effets microclimatiques ou à des erreurs de mesure.

Dans ce cas, l'espérance de l'erreur δ_o^2 s'écrit :

$$\delta_o^2 = \sigma_{oo}^2 - 2\sum_{j=1}^{P} \lambda_j \sigma_{jo} + \sum_i \sum_j \lambda_i \lambda_j \sigma_{ij} + \sum \lambda_j^2 \varepsilon_i^2$$

et le système devient :

$$\sum_{j=1}^{P} \lambda_i \sigma_{ij} + \lambda_i \varepsilon_i^2 = \sigma_{io}$$

$$\left(\Sigma + \begin{vmatrix} \varepsilon_1^2 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_p^2 \end{vmatrix} \right) . \vec{\Lambda} = \vec{\Sigma}_o$$

Cela revient en fait à perturber la diagonale de la matrice de covariance par des termes de "variance spécifique".

Les ε_i sont déterminés soit a priori (en considérant qu'une mesure est exacte à $\pm$ 5% par exemple) soit à partir du corrélogramme $\sigma(d_{ij})$ s'il ne tend pas vers 1.0 quand $d \longrightarrow 0$. Dans ce cas on ne peut déterminer qu'une valeur commune $\varepsilon \ \forall i$ et la matrice du système devient $\Sigma + \varepsilon^2 I$.

On vérifiera que dans ce cas, on n'interpole plus exactement mais on lisse.

Cette notion est tout à fait analogue à l'effet de pépite dans certains variogrammes, qui vient modifier le système de krigeage.

Enfin, dans la mesure où l'on peut faire varier ε , cela n'est pas sans rappeler les techniques de "ridge regression"
qui vise à améliorer la robustesse des λ_j estimés.

Nous n'envisagerons pas certaines autres variantes, utilisées surtout en exploitation opérationnelle(cf T. SCHLATTER et al, 1976 ; THIEBAUX J., 1975 ; etc...)

IV.1.3. Conclusions

La méthode de GANDIN n'est autre qu'une corrélation multiple classique notée habituellement : $R_{xx} . \Lambda = R_{xy}$ où l'on ne connaitrait pas R_{xy} . On le remplace donc par un modèle.

Outre le fait que la méthode est locale, et ne fournit pas par exemple une équation générale de la surface interpolée (quand on veut estimer le champ en tout point) le point délicat est le rôle de ce modèle.

En effet, le tableau R_{xx} , ou Σ fournit souvent un nuage σ_{ij} en fonction de d_{ij} assez grossier (cf figure IV-4 dans la IVème partie) dans lequel plusieurs modèles peuvent être ajustés. Le modèle retenu sert aussi à "lisser" le membre de gauche du système, sans que l'on ait aucune caractérisation de l'information "perdue" par ce lissage du corrélogramme.

Certes il ne s'agit pas forcément d'une perte d'information, mais d'un lissage des fluctuations d'échantillonnage et d'un filtrage des effets de microéchelle, qui n'est toutefois pas quantifié.

IV.2 - Approximation globale (l'analyse harmonique et les interpolations ou lissage par les fonctions orthogonales empiriques

Contrairement aux techniques précédentes, elle ne s'attache pas à estimer un point particulier à l'aide de son voisinage propre, mais elle fournit une approximation valable pour toute l'étendue du champ.

Comme dans le paragraphe précédent, on supposera que les champs des moyennes et écarts-types sont connus et donc que l'on travaille sur un processus stationnaire (ou en variables centrées réduites).

IV.1.4. Approche simple dans le cas d'une dimension

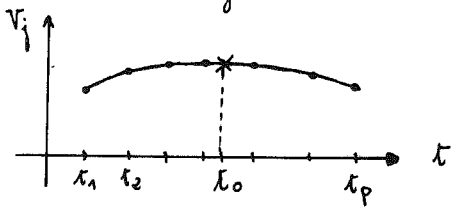
(a) Historiquement, l'idée de la méthode provient de la parfaite symétrie entre les rôles des variables X et des composantes Z dans l'ACP.

Une observation $X_i = X_{iV} = \{x_{i1}, x_{i2}, \ldots x_{ip}\}$ peut être parfaitement reconstituée si on connait les composantes principales

$$Z_{iV} = \{z_{i1} \ldots z_{ik} \ldots z_{ip}\} \Longrightarrow X_{iV} = Z_{iV} \cdot \begin{vmatrix} \vec{V}_1^t \ldots \\ \ldots \\ V_k^t \ldots \\ \ldots \\ V_p^t \ldots \end{vmatrix}$$

Si on se limite aux K premières composantes, on ne reconstitue qu'imparfaitement X_{iV} mais on sait qu'elles fournissent le maximum que l'on peut obtenir avec des combinaisons linéaires des variables initiales.

Dans le cas où X est un processus $X(t)$ échantillonné aux temps $t_1, t_2, \ldots t_p$ on a vu que les vecteurs propres, par exemple $\vec{V}_j$, présentaient une allure très régulière (cf. Chap IV, IV^{ème} partie); Et si on voulait insérer un nouveau point t_o, il est probable que la forme de V_j varierait peu et que $V_j(t_o)$ peut se déduire des $\{V_j(t_1) \ldots V_j(t_p)\}$ déjà connus.



En faisant de même pour $V_1(t_o), V_2(t_o) \ldots V_k(t_o)$ et en l'insérant dans le système

$$\left[X_i(t_1)\ X_i(t_2) \ldots\ X_i(t_o) \ldots X_i(t_p)\right] = \left[Z_{i1} \ldots Z_{ik}\right] \cdot \begin{bmatrix} V_1(t_1) \ldots V_1(t_o) \ldots V_1(t_p) \\ \ldots \\ V_k(t_1) \ldots V_k(t_o) \ldots V_k(t_p) \end{bmatrix}$$

on pourrait ainsi reconstruire une valeur approchée $X_i(t_o)$. Malheureusement un certain biais apparaît car après adjonction de $V_j(t_o)$ le vecteur V_j n'est plus normé et les vecteurs ne sont plus tout à fait orthogonaux 2 à 2. Néanmoins, si P est assez grand (> 10) les résultats sont déjà satisfaisants.

(b) Une autre approche, beaucoup plus rigoureuse, découle de la définition des fonctions propres d'un processus associées à sa fonction d'autocorrélation $R(t,t')$. Si on travaille sur un intervalle fini $[0,T]$, elles vérifient :

$$\int_0^T R(t,t').\varphi_k(t').dt' = \lambda_k.\varphi_k(t)$$

La résolution numérique de l'équation intégrale peut se faire en discrétisant la somme du premier membre en :

$$\sum_{j=1}^{P} R(t,t_j).\varphi_k(t_j).\Delta t_j = \lambda_k.\varphi_k(t)$$

et ce en autant de points t que l'on veut, par exemple pour $t = t_1, t_2, \ldots t_P$, d'où un système aux valeurs propres de P équations à P inconnues $\varphi_k(t_i)$:

$$i = 1 \ldots P \left\{ \sum_{j=1}^{P} R(t_i,t_j).\varphi_k(t_j).\Delta t_j = \lambda_k.\varphi_k(t_i) \right.$$

Malheureusement, la matrice cesse d'être symétrique, et on préfère multiplier les 2 membres par $\sqrt{\Delta t_i}$, d'où :

$$\sum_{j=1}^{P} \sqrt{\Delta t_i}.R(t_i,t_j).\sqrt{\Delta t_j}.\varphi_k(t_j).\sqrt{\Delta t_j} = \lambda_k.\varphi_k(t_i).\sqrt{\Delta t_i}$$

Si on pose : $\Gamma = \left\{ \Gamma_{ij} = R_{ij}.\sqrt{\Delta t_i.\Delta t_j} \right\}$ et $\overrightarrow{V_k} = \left\{ V_{ki} = \varphi_k(t_i).\sqrt{\Delta t_i} \right\}$

on voit qu'il suffit de diagonaliser la matrice symétrique $\Gamma \Longrightarrow \Gamma.\overrightarrow{V_k} = \lambda_k.\overrightarrow{V_k}$

Comme les fonctions φ_k doivent être orthonormées :

$$\int_0^T \varphi_k(t).\varphi_l(t).dt = \delta_{kl}$$

ce qui se discrétise en : $\sum_{i=1}^{P} \varphi_k(t_i).\varphi_l(t_i).\Delta t_i = \sum_{i=1}^{P} \varphi_k(t_i).\sqrt{\Delta t_i}.\varphi_l(t_i).\sqrt{\Delta t_i}$

$$= \sum_{i=1}^{P} V_{ki}.V_{li} = \delta_{kl}$$

$\Longrightarrow$ et on voit qu'il suffit de prendre les vecteurs $\overrightarrow{V_k}$ normés à 1 pour que cela soit vérifié.

Le calcul de la k ème composante principale pour la i ème réalisation du processus s'écrit alors :

$$Z_{ik} = \int_0^T X_i(t).\varphi_k(t).dt$$

soit en différences finies :

$$Z_{ik} = \sum_{j=1}^{P} X_{ij}.\varphi_{kj}.\Delta t_j = \sum_{j=1}^{P} X_{ij}.V_{kj}.\sqrt{\Delta t_j}$$

Inversement : $X_i(t) = \sum_{k=1}^{P} Z_{ik}.\varphi_k(t)$

et on peut montrer que le développement, limité à K termes $\hat{X}_K(t)$ minimise :

$$\int_0^T E\left[\left(X(t) - \hat{X}_K(t)\right)^2\right] . dt$$

parmi l'ensemble des systèmes de fonctions orthogonales .

C'est cette approche que l'on trouve chez COHEN et JONES (1969), HOLMSTROM I. (1963) et BUELL C. (1971).

IV.2.2. Généralisation

Elle a été proposée par J.C. DEVILLE (1973 et 1974) et part de la théorie classique de l'interpolation deterministe (cf par exemple J. LEGRAS , 1965).

Soit une fonction $X(t)$ connue sur les P couples $\{X(t_i), t_i\}$ et soit $\mathcal{C}$ une classe de fonctions ayant une structure d'espace vectoriel, on va interpoler $X(t)$ en la remplaçant par une fonction $\phi_x(t) \in \mathcal{C}$ et définie par :

$$\phi_x(t) = \sum_{j=1}^{P} y_j . \varphi_j(t) \qquad \text{et} \quad \phi_x(t_i) = X_i \quad \forall i = 1, \ldots P$$

où les $\varphi_j(t)$ constituent une base de $\mathcal{C}$. Un cas particulier important est celui où l'on prend pour fonctions de base des fonctions simples, ou élémentaires $e_i(t)$ telles que :

$$\{e_i(t_1), \ldots, e_i(t_i), \ldots, e_i(t_P)\} = \{0, \ldots, 1, \ldots, 0\} \Longrightarrow e_i(t_j) = \delta_{ij}$$

Dans ce cas :

$$\phi_x(t) = \sum_{j=1}^{P} X(t_j) . e_j(t)$$

Si on considère maintenant $X(t) = X(\omega, t)$ comme un processus, $\phi_x(t)$ est un autre processus, qui vérifie pourtant :

$$E[\phi_x(t)] = \sum_{j=1}^{P} E[X(t_j)] . e_j(t)$$

$$R_\phi(t, t') = E[\phi_x(t) . \phi_x(t')] = \sum_{i,j} R_{ij} . e_i(t) . e_j(t')$$

avec $R_{ij} = E[X(t_i) . X(t_j)]$ = covariance de X

La covariance R_ϕ du processus interpolé apparaît donc comme l'interpolation de la covariance du processus, et converge d'ailleurs vers celle-ci si $X(\omega, t)$ est continu en moyenne quadratique.

Le problème aux valeurs propres devient alors la recherche des fonctions $\varphi_k(t)$ associées non pas au noyau $R_x(t, t')$ (connu seulement par points) mais au noyau interpolé R_ϕ, soit :

$$\int_0^T R_\phi(t, t') . \varphi_k(t') . dt' = \lambda_k . \varphi_k(t)$$

On remplace R_ϕ par son expression et on va évidemment chercher $\varphi_k(t)$ dans la classe $\mathcal{C}$, sous la forme :

$$\varphi_k(t) = \sum_{i=1}^{P} \xi_{ik} \cdot e_i(t)$$

d'où :

$$\int_0^T \left[\sum_{ij} R_{ij} \cdot e_i(t) \cdot e_j(t') \right] \cdot \left[\sum_{l=1}^{P} \xi_{lk} \cdot e_l(t') \right] \cdot dt' = \lambda_k \cdot \sum_i \xi_{ik} e_i(t)$$

soit encore :

$$\sum_{i=1}^{P} e_i(t) \cdot \sum_{j,l} R_{ij} \cdot \xi_{lk} \int_0^T e_j(t') \cdot e_l(t') dt' = \lambda_k \cdot \sum_{i=1}^{P} e_i(t) \cdot \xi_{ik}$$

et en identifiant les termes en $e_i(t)$, pour $i = 1$ à P on a le système :

$$\begin{bmatrix} R_{i1} \cdots R_{ij} \cdots R_{iP} \end{bmatrix} \times \begin{bmatrix} E_{j1} \cdots E_{jl} \cdots E_{jP} \end{bmatrix} \times \begin{bmatrix} \xi_{1k} \\ \vdots \\ \xi_{lk} \\ \vdots \\ \xi_{Pk} \end{bmatrix} = \lambda_k \begin{bmatrix} \xi_{1k} \\ \vdots \\ \xi_{lk} \\ \vdots \\ \xi_{Pk} \end{bmatrix}$$

soit encore : $R \times E \times \xi = \lambda \times \xi$

avec $E = \{ E_{jl} \}$ $\quad E_{jl} = \int_0^T e_j(t) \cdot e_l(t) \cdot dt$

La matrice E est une matrice de produits scalaires donc définie positive et elle admet une racine carrée positive unique $E^{\frac{1}{2}}$ (que l'on peut calculer en diagonalisant E . $= E^{\frac{1}{2}} \cdot E^{\frac{1}{2}t}$

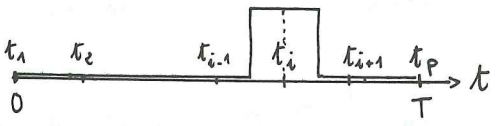
On préfèrera donc résoudre le problème symétrique

$$(E^{\frac{1}{2}t} R E^{\frac{1}{2}}) \cdot (E^{\frac{1}{2}t} \cdot \xi) = \lambda (E^{\frac{1}{2}t} \cdot \xi)$$

ou avec des notations évidentes : $\Gamma \cdot V = \lambda V$

La résolution numérique dépend évidemment des fonctions élémentaires $e_i(t)$ choisies :

(a) Constantes par morceau :



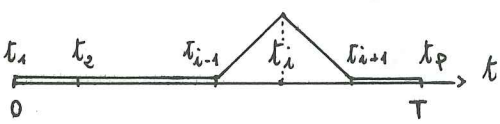
$$e_i(t) = \begin{cases} 1 \text{ entre } t_i - \frac{\Delta t_i}{2} , t_i + \frac{\Delta t_i}{2} \\ 0 \text{ ailleurs} \end{cases}$$

Dans ce cas la matrice E est diagonale,

car : $E_{ij} = \int_0^T e_i(\tau) \cdot e_j(t) \, dt = \delta_{ij} \cdot \Delta t_i$

On retrouve comme cas particulier l'approche du IV.2.1.

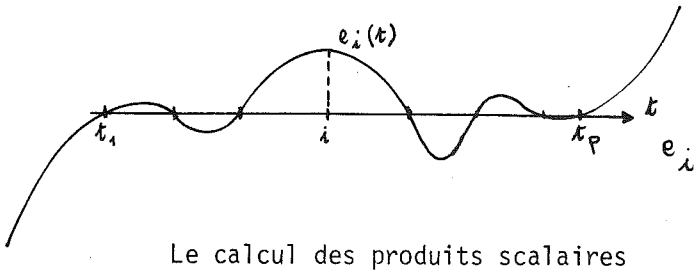
(b) Linéaires par morceaux :



$$e_i(t) \begin{cases} \frac{t - t_{i-1}}{t_i - t_{i-1}} & \text{sur } [t_{i-1}, t_i] \\ \frac{t_{i+1} - t}{t_{i+1} - t_i} & \text{sur } [t_i, t_{i+1}] \\ 0 & \text{ailleurs} \end{cases}$$

Le traitement dans ce cas peut d'ailleurs être simplifié en introduisant $\mathcal{P}-1$ fonctions, c'est-à-dire en découpant $e_i(t)$ en $e_i^-(t)$ et $e_i^+(t)$ à droite et à gauche (DEVILLE, 1973), mais ceci n'a d'intérêt qu'à une dimension.

c) Polynômiales



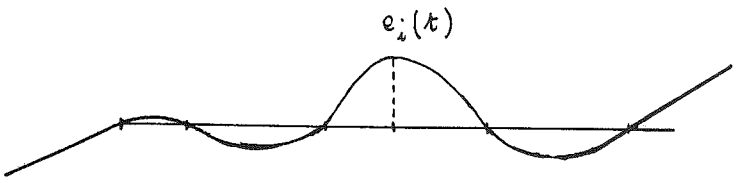
Les fonctions élémentaires sont alors les classiques polynômes de Lagrange:

$$e_i(t) = \mathcal{L}_i(t) = \prod_{j \neq i} \frac{t - t_j}{t_i - t_j}$$

Le calcul des produits scalaires $\int_0^T \mathcal{L}_i(t).\mathcal{L}_j(t).dt$ est possible mais laborieux aussi préfère-t-on faire un changement de base pour repasser dans la base canonique et calculer des intégrales (cf J. LEGRAS, 19): $\int_0^T t^r.t^s.dt$

d) Fonctions Splines

Les fonctions élémentaires sont alors des splines de base: $\sigma_i(t)$



Ce sont par exemple des morceaux de polynômes du 3ème degré entre 0 et T (des droites en dehors). Là aussi le calcul des produits scalaires peut être envisagé analytiquement.

IV.2.3. Cas de 2 dimensions

Le passage à 2 dimensions est relativement direct. On a un processus que l'on peut noter $X(P) = X(x, y)$ et soit $dP = dx.dy$

On introduit là aussi des fonctions de base :

$$e_i(P) \quad \text{telles que} \quad e_i(P_j) = \delta_{ij}$$

et on travaille sur un domaine Ω dans le plan (x, y).

Le traitement numérique est alors absolument analogue au cas de 1 dimension. L'équation intégrale:

$$\iint_\Omega R(x,y,x',y').\varphi(x',y')\,dx'dy' = \lambda.\varphi(x,y) \quad \text{ou} \quad \int_\Omega R(P,Q).\varphi(Q).dQ = \lambda.\varphi(P)$$

devient pour le processus interpolé :

$$\int_\Omega \left[\sum_{i,j} R_{ij}\, e_i(P).e_j(Q)\right]\left[\sum_\ell \varphi_{\ell k}\, e_\ell(Q)\right].dQ = \lambda_k \cdot \sum_i \varphi_{ik}\, e_i(P)$$

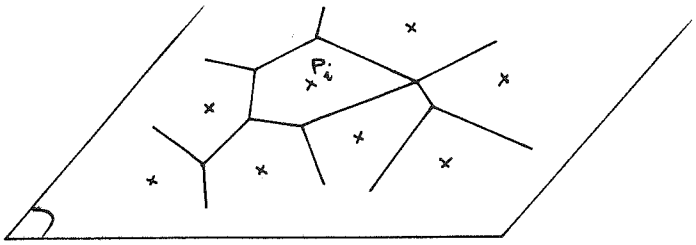
qui fournit le même système aux valeurs propres : $R.E.\varphi = \lambda.\varphi$

avec $E = \{E_{ij}\}$ où $E_{ij} = \int_\Omega e_i(P).e_j(Q).dP$

et que l'on résoud de la même manière que précédemment en passant à une matrice symétrique.

Le traitement numérique est légèrement plus compliqué qu'à 1 dimension et selon la base de fonctions choisies :

(a) Constantes par éléments :



On associe à chaque point P_i un élément de surface ΔP_i et :

$$e_i(P) = \begin{cases} 1 \text{ sur } \Delta P_i \\ 0 \text{ ailleurs} \end{cases}$$

La méthode classique utilise pour ΔP_i le polygone de THIESSEN associé au point P_i. Des problèmes peuvent se poser aux frontières du domaine (cf applications).

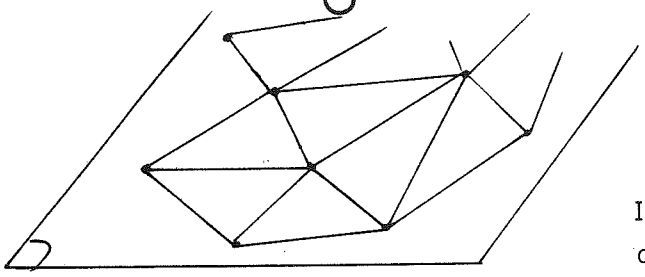
Les produits scalaires se ramènent trivialement aux éléments de surface eux-mêmes :

$$\int_\Omega e_i(P).e_j(P).dP = \delta_{ij}.\Delta P_i$$

La matrice à diagonaliser est donc seulement celle des $R_{ij}.\sqrt{\Delta P_i.\Delta P_j}$ et on voit que l'on compense bien, par cette pondération, les anomalies de densité d'un réseau de stations aléatoirement distribuées.

Notons aussi que la trace est égale à la surface du domaine Ω.

(b) Linéaires par facettes



On effectue alors une triangulation du réseau ; c'est un maillage classique, du type éléments finis, qui peut être fait automatiquement.

Il n'y a plus de problème aux frontières car celle-ci est le périmètre joignant les stations les plus extérieures du réseau.

La fonction $e_i(P)$ vaut 1 au point i et décroît linéairement jusqu'à 0 sur les différentes facettes issues de .

Le calcul des produits scalaires est assez laborieux mais peut encore être effectué analytiquement dans la mesure où 2 noeuds adjacents partagent au plus 2 faces.

On notera aussi que, dans la mesure où le système à résoudre est issu d'une équation intégrale, elle-même associée à un problème d'optimisation, il y a de fortes analogies avec la technique des éléments finis.

(c) Splines type plaque mince

Il s'agit cette fois de fonctions continues sur le domaine et présentant certaines propriétés optimales.

Par contre, à 2 dimensions, leur expression analytique n'est plus aussi simple ce qui rend impossible le calcul analytique des produits scalaires. Dans la mesure où l'évaluation de la fonction elle-même est simple, on fera le calcul numériquement pour :

$$E_{ij} = \int_\Omega \sigma_i(P).\sigma_j(P).dP$$

IV.2.3. Applications

(a) Comme on l'a vu déjà en IV.2.1, les composantes principales se calculent en projetant le processus (ou sa i-ème réalisation) sur la fonction propre correspondante, d'où :

$$Z_{ik} = \int_\Omega X_i(P) \cdot \varphi_k(P) \cdot dP$$

soit encore :

$$Z_{ik} = \int_\Omega \sum_j X_i(P_j) e_j(P) \cdot \sum_\ell \varphi_{\ell k} \cdot e_\ell(P) \cdot dP$$

ce qui refait apparaître les produits scalaires $E_{j\ell}$.

Une fois calculées les Z_{ik} , $k = 1, \ldots, P$, on peut évaluer $X(P)$ en un point quelconque P_o par :

$$X_i(P_o) = \sum_{j=1}^{P} Z_{ik} \cdot \varphi_k(P_o) \qquad (1)$$

où $\varphi_k(P_o)$ est estimée par :

$$\varphi_k(P_o) = \sum_\ell \varphi_{\ell k} \cdot e_\ell(P_o)$$

L'intérêt , par rapport à une interpolation classique :

$$X_i(P_o) = \sum_{j=1}^{P} X_i(P_j) \cdot e_j(P_o) \qquad (2)$$

provient de ce que, dans (1), on peut limiter la sommation aux $K < P$ premiers termes. Ce n'est pas tant l'économie de calcul que le filtrage qui est recherché. En effet, on a vu que les composantes de rang élevé ont une énergie faible, et sont en général associées à des fréquences élevées et considérées comme des bruits indésirables.

Il faut toutefois être très vigilant sur le choix de K et ne pas conserver une fréquence tout en éliminant arbitrairement une fréquence voisine. Cela nous ramène au problème du choix du nombre de facteurs évoqué dans la IIIème Partie (Chap.I).

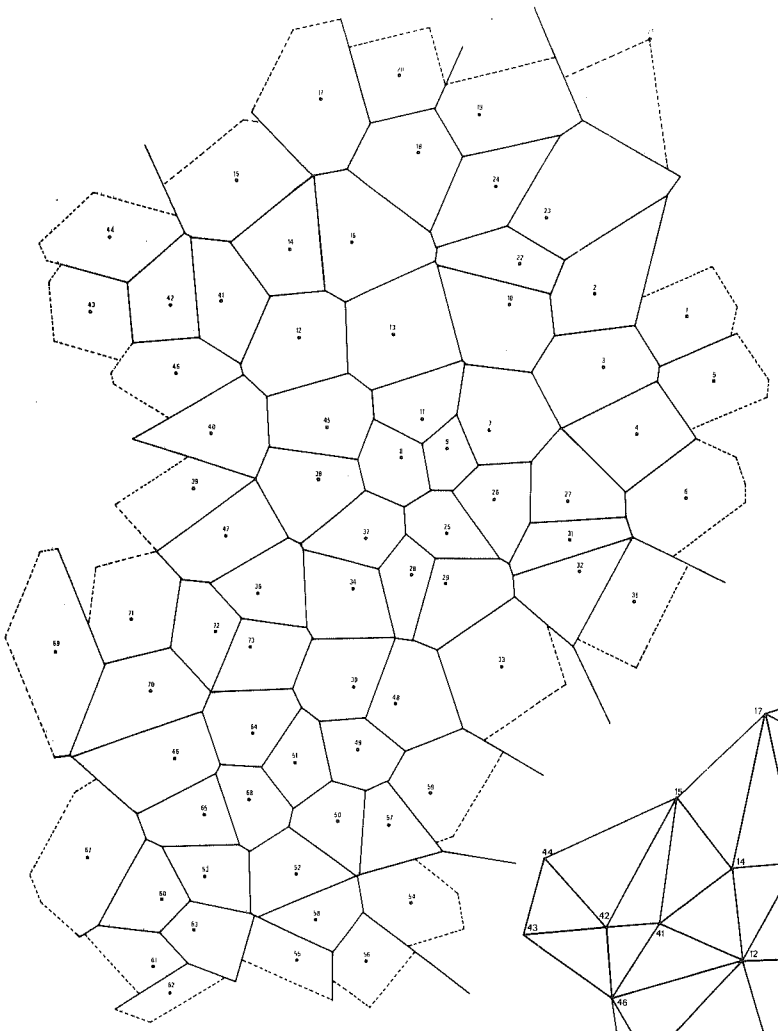
On constatera aussi qu'il doit y avoir cohérence entre la base choisie pour l'interpolation des φ_k et les produits scalaires utilisés dans l'évaluation des valeurs de $\varphi_k(P_j)$ aux stations (système aux valeurs propres). Il est discutable, par exemple, de diagonaliser la matrice de corrélation brute, puis de lisser les vecteurs propres obtenus par des techniques sophistiquées.

(b) Dans le cas où l'on choisit des fonctions $e_j(P)$ constantes par intervalle, on utilise un découpage en éléments du type THIESSEN(cf. Fig.V-9-a) .

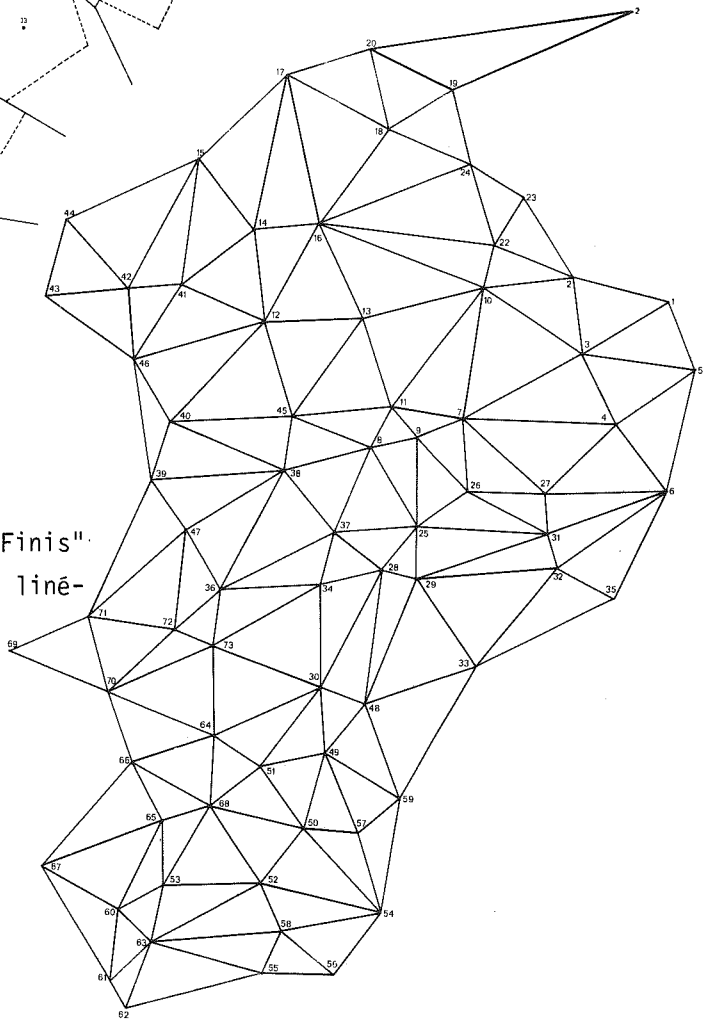
Les fonctions étant supposées constantes par morceaux, il est évident que si l'on estime $X(P_o)$ à l'aide de <u>toutes</u> les fonctions propres, on reconstitue $X(P_j)$ si $P_o \in$ élément ΔP_j et la méthode rejoint paradoxalement la technique du plus proche voisin !

Ce qui est intéressant dans ce cas est donc seulement l'effet du filtrage en n'utilisant qu'un nombre réduit de fonctions K .

(c) Le cas de l'interpolation linéaire est plus satisfaisant. Il utilise des éléments triangulaires (cf. Fig.V-9-b) et l'interpolation $X(P_o)$ dépend alors à la fois du nombre de fonctions choisies et de l'interpolation linéaire entre les 3 points de l'élément.



a) - Maillage du type "Thiessen" pour l'approximation par des fonctions constantes par intervalle .



b) - Maillage du type "Eléments Finis" pour approximation par fonctions linéaires par facettes .

FIGURE V - 9 : Découpages utilisés pour l'interpolation par "Analyse Harmonique".

IV.3 - Comparaison et évaluation sur les épisodes cévenols

On les trouvera dans D. CREUTIN (1979) ainsi que des méthodes purement spatiales comme le krigeage universel, ou plus triviale comme la moyenne arithmétique.

Il faut cependant signaler pour conclure que la méthode des fonctions orthogonales empiriques, contrairement à l'interpolation de GAUDIN ou au krigeage, n'a pas encore atteint un développement théorique pleinement satisfaisant et nos recherches continuent dans ce domaine.

"If a man will begin with certainties, he shall-end in doubts ; but if he will be content to begin with doubts, he shallend in certainties"

F. BACON

CONCLUSIONS GENERALES

Un certain nombre de conclusions ont déjà été présentées, au fil des pages, pour chaque méthode ou problème particulier. De manière plus générale, nous espérons avoir montré l'intérêt de l'approche statistique, tant dans les problèmes de prévision que dans les problèmes d'exploitation des réseaux de mesures.

Pour les premiers, on peut considérer cette approche comme un moyen de "courtcircuiter" un modèle physique trop compliqué, ou de pallier au contraire un manque de connaissance des mécanismes, en mettant directement en regard l'effet que l'on veut prévoir, et ses causes présumées. Dans le second type de problèmes, elle permet de décrire et de prendre en compte, par exemple dans un processus d'interpolation, les propriétés statistiques du phénomène mesuré par un réseau. Dans les deux cas, elle fournit, à la différence des modèles déterministes, des résultats gradués en probabilité ou affectés d'une variance d'estimation, ce qui est très utile pour les prises de décision, ou les analyses économiques ultérieures.

Sur un plan pratique, ces expériences nous ont suggéré un schéma méthodologique pour aborder les problèmes d'analyse de données: D'abord il est rare qu'un problème soit suffisamment bien posé pour qu'une méthode précise s'impose a priori. D'où l'intérêt des techniques qui visualisent les données avec le minimum d'hypothèses sous jacentes. Ensuite, quand une méthode a été retenue, il est souvent nécessaire de la modifier, ou de l'adapter, mais surtout, il faut, avant de l'utiliser sur un problème délicat, en explorer les possibilités soit sur des exemples simples, soit sur des exemples simulés. Et il n'est pas rare qu'une méthode mathématiquement attrayante se révèle d'un médiocre intérêt pratique.

Enfin, il est piquant de constater que, parti des notions ponctuelles et géométriques associées aux nuages de points dans R^p ou R^N, ce périple à travers l'Analyse des Données nous ait ramené aux aspects continus et analytiques des

fonctions aléatoires. Cette convergence avec les techniques de traitement du signal encore incomplètement explorée, ne sera sans doute jamais totale étant donné les particularités des échantillons dont on dispose en hydrométéorologie.

Mais cela illustre combien il peut être enrichissant, tant au niveau théorique que dans l'interprétation des résultats, de développer les analogies entre les différentes techniques dont on dispose actuellement.

BIBLIOGRAPHIE DE LA PREMIERE PARTIE

CHAPITRE I

I.1 - Prévision d'avalanche

LACHAPELLE E. (1977)
"Snow Avalanches : Current research and application"
Symposium of applied Glaciology. Cambridge 1976
J. of Glaciology - vol.19 n°81, 1977.

WILLIAMS K. (1978)
"Post control avalanche releases"
2ème Rencontre Internat. sur la Neige et les Avalanches - ANENA-
12-14 Avril 1978 - Grenoble.

WORLD DATA CENTER A FOR GLACIOLOGY
"Avalanches" - Report GD 1 1977 - INSTAAR - Boulder Colorado 80309 - USA

ALLIX A. (1925)
"Les Avalanches" - Revue de Géographie Alpine, vol.13 N°1 - 1925.

AKKURATOV V.N. (1963)
"Forecasting the onset of avalanche danger from the amount of wind transport and the temperature contraction of the snow"
In : Questions on snow utilisation and the prevention of snowdrifts and avalanches. Publ. Acad. Nauka. Moscou , 1966.

ABDUSHELISHVILI K.L. et TSOMAIA V.S. (1963)
"On an experiment of forecasting avalanches of freshly fallen snow in the Caucasus"
First Conference on Studying the formation and release of avalanche.
Tachkent , 1963.

ZINGG T. (1965)
"Relation between weather situation, snow metamorphism and avalanche activity"
Symposium AIHS - Davos 1965 - Publication n°69.

POGGI A. et PLAS J. (1965)
"Conditions météorologiques critiques pour le déclenchement des avalanches"
Symposium AIHS - Davos 1965 - Publication n°69.

SELIGMAN G. (1936)
"Snow structures and Ski fields"
London, Mac Millan Ed., 1936.

SCHERBAKOV M.P. (1966)
"Method of predicting avalanche danger from snowfall intensity"
Izvestia Serie Geog. N°3, Mai-Juin 1966 (Traduction de R. PERLA - USDA Alta Avalanche Study Center).

KOSAREV M.V. (1969)
"Main results of a study of avalanche formation conditions on the southern slopes of the western Tien Shan"
Soviet Hydrology : Selected papers. Vol.3 - 1969.

ROCH A. (1963)
"Les variations de la résistance de la neige"
Symposium AIHS - Davos, 1965 (op.cité)

MELLOR M. (1968)
"Avalanches"
CRREL report. US Army Hanover - New Hampshire.

PERLA R. (1970)
"On contributory factors in avalanche hazard evaluation"
Revue Canadienne de Géotechnique, vol.7 n°4, 1970.

OBLED Ch. (1970)
"Vers une prévision numérique des risques d'avalanches"
Rapport DGRST, Grenoble 1970. Présenté à la Commission de Glaciologie de la SHF. Mars 1971.

JUDSON A. et ERIKSON B.J. (1973)
"Predicting avalanche intensity from weather data : a statistical analysis"
U.S.D.A. Forest Service - Research Paper RM - 112.

BOVIS M.J. (1977)
"Statistical forecasting of snow avalanches. San Juan Mountains-Southern Colorado. USA"
J. of Glaciology, Vol.18, n°78, 1977.

TOUCHINSKY G.K. et TROCHKINA (1974)
"Avalanches de neige. Prévision et Protection"
1 vol. Faculté de Géographie. Edition de l'Université de Moscou. 1974, comprenant :

- GRAKOVITCH V.F. : "Utilisation de la méthode d'identification des types dans l'évaluation d'une situation avalancheuse à partir d'information météorologiques"
- KHOMENIOUK Y.V. et al. : "Prévision du danger d'avalanche dans le temps et dans l'espace".

DROSDOVSKAYA N.F. (1977)
"Using linear descriminant analysis to classify snowfall situations into avalanching and non-avalanching ones"
J. of Glaciology, vol.19, n°81, p.679.

YEFIMOV M.K. et KIOZIK Y.M. (1975)
"Relation of some meteorological elements to avalanching in the Dukant River Basin (Western Tien Shan)"
Soviet Hydrology. Selected papers. n°4 - 1975.

GRIGORIAN S.S. (1974)
"Mechanics of snow avalanches"
Symposium on Snow Mechanics. Grindelwald 1974. IAHS publ. n°114 (1975)

BOIS Ph. et OBLED Ch. (1973)
"Vers un système opérationnel de prévision des avalanches par des méthodes statistiques"
Bulletin AIHS, Vol.18, n°24, pp 419-429.

BOIS Ph., OBLED Ch. et GOOD W. (1974)
"Discriminant analysis as a tool for day by day avalanche forecast"
Symposium on Snow Mechanics. Grindelwald 1974, IAHS Publication n°114.

BOIS Ph., OBLED Ch. (1976)
"Prévision des avalanches par des méthodes statistiques. Aspects méthodologiques et opérationnels"
La Houille Blanche n°6/7, 1976, pp 509-531.

FOEHN P., GOOD W., BOIS Ph. et OBLED Ch. (1976)
"Evaluation and comparison of conventional and statistical methods to forecast avalanche hazard"
Symposium of Applied Glaciology. Cambridge 1976. Publié dans J. of Glaciology, vol.19, n°81, 1977.

OBLED Ch. et GOOD W. (1979)
"Recent developpements of avalanche forecasting by discriminant techniques : a methodological review and some applications of the Parsenn area (Davos - Switzerland)"
J. of Glaciology (sous presse)

FOHN P., HAECHLER P. (1978)
"Prévision des grosses avalanches au moyen d'un modèle déterministe statistique"
2nde rencontre internationale sur la neige et les avalanches - ANENA - Grenoble, 1978.

SALWAY A.A. (1976)
"Statistical estimation and prediction of avalanche activity from meteorological data"
Thèse de doctorat en philosophie. Université de Colombie Britannique. 1976.

I.2. - Analyse spatiale des pluies

PARDE M. (1961)
"Sur la puissance des crues en diverses parties du monde"
Geografica Ano VIII - Enero - Diciembre 1961 Saragosse - Faculté des lettres.

CAROFF J.M. (1965)
"Situations météorologiques provoquant de fortes précipitations sur le cours supérieur de la Loire"
Rapport HYD 65 / n°17 CREC - Chatou.

JACQUET J. (1959)
"Les Crues d'Automne 1958 sur le Vidourle"
La Houille Blanche, N° Spécial A , 1959.

GUILLOT P. (1959)
"Aspect hydrométéorologique des crues cévenoles des 30 Septembre et 4 Octobre 1958".
La Houille Blanche, Numéro I, 1959.

WHITTLE P. (1954)
"On Stationary processes in the plane"
Biometrika, Vol.41, pp 434-449, 1954.

AMOROCHO J. et WU B. (1977)
"Mathematical Models for the simulation of cyclonic storm sequences and precipitation fields"
J. of Hydrology, vol.32, pp 329-345.

GANDIN L.S. (1968)
"Objective analysis of meteorological fields"
Leningrad. Traduit en 1965 Israel Program for Scientific Translation

MATTHERON G. (1965)
"Théorie des variables généralisées et leur estimation"
Masson Ed.

HUTCHINSON P. (1972)
"The use of a modified time serie analysis technique for the determination of areal precipitations accuracies"
WMO/OMM Symposium de Geilo "Distribution of precipitation in mountainous areas"
Vol.II, pp 565-587.

RODRIGEZ-ITURBE I. et MEIJA J.M. (1974)
"The design of rainfall network in time and space"
Water Res. Res., Vol.10, N°4, 1974.

CLARKE R.T. (1976)
"Statistical Methods for the study of spatial variation in Hydrology variables"
Chap. 11, pp 299 à 314 in "Facets in Hydrology" publié par J.C. Rodda - John Wiley & Sons, 1976.

DELHOMME J.P. et DELFINER (1973)
"Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride" Colloque UNESCO-OMM-AIHS sur l'élaboration des projets d' utilisation des ressources en eau sans données suffisantes.Madrid 1973.
Actes du colloque -TomeI pp191-210.

CHAPITRE II

ANDERBERG M.R. (1973)
"Cluster analysis for applications"
Academic Press, 1973.

ROCHE M. (1963)
"Hydrologie de Surface"
Gauthiers-Villars Ed., Paris 1963.

MATALAS N.C. (1967)
"Mathematical assessment of synthetic hydrology"
Water Res. Res. vol.3, n°4, 1967.

KUENY J.L. (1977)
"Contribution au traitement statistique de données météorologiques"
Thèse D.I. Automatique, Grenoble 1977.

MEIJA J.M., RODRIGUEZ-ITURBE I et CORDOVA J. (1974)
"Multivariate generation of mixtures of Normal and Log-normal variables"
Water Res. Res. vol.10, n°4, 1974.

BIBLIOGRAPHIE DE LA DEUXIEME PARTIE

DAUXOIS J. et POUSSE A. (1976)
"Les analyses factorielles en analyse des probabilités et en statistiques : essai d'étude synthétique"
Thèse de doctorat d'état ès-sciences mathématiques. Université Paul Sabatier. Toulouse.

CHAPITRE I

BOIS PH. (1976)
Thèse de doctorat ès sciences. op. cité.

ANDERSON T.W. (1958)
"An introduction to multivariate statistical methods"
John Wiely & Sons Ed.

CAILLEZ F., MAILLES J.P., NAKACHE J.P. et PAGES J.P. (1973)
"Analyse de données multidimensionnelles"
publié par E.E.E., 3 Tomes.

DHUIME G. (1974)
"L'analyse en composantes principales"
in Cahiers IRIA 1974 n°5 "Contribution à l'enseignement assisté par ordinateur".

DENIAU C., OPPENHEIM G. et LEROUX B. (1972)
"Deux méthodes d'analyse factorielle" in "Analyse des données en architecture et en urbanisme".
Colloque de l'Institut de l'Environnement. Ministère des Affaires Culturelles, Avril 1972, pp 123-168.

LEBART L. et FENELON J.P. (1973)
"Statistique et information appliquées"
Dunod 2è. ed.

MORRISON D.F. (1967)
"Multivariate statistical methods"
Mac Graw Hill Ed.

HILTON G. (1972)
"An algorithm for detecting differences between transition probability matrices"
Applied Statistics, p.81-86.

HOTTELING H. (1933)
"Analysis of a complex of statistical variables into principal components"
J. of Educational Psychology, n°24.

CEHESSAT R. (1976)
"Exercices commentés de statistique et informatique appliquées". Dunod Ed.

CHAPITRE II

HILL M.O. (1974)
"Correspondence analysis : a neglected multivariate method"
J. of Appl. Statist. , Vol.23, n°3.

CAZES P., BRENOT J. et LACOURBY N. (1976)
"Régression par boule et par l'analyse en correspondances"
Rev. de Statist. Appl. , vol.XXIV, n°4

LECLERC A. (1975)
"L'analyse des correspondances sur juxtaposition de tableaux de contingence"
Rev. de Statist. Appl. 1975, vol. XXIII, n°3.

NAKACHE J.P. (1973)
"Influence du codage des données en analyse factorielle des correspondances"
Rev. de Statist. Appl. 1973., vol.XXI, n°2.

Ph. BOIS (1976)
Thèse de Doctorat (op. cité).

CEHESSAT R. (1976)
"Exercices commentés et statistique et informatique appliquées"
1ère ed. Dunod.

FRITSCH A. et OBLED Ch. (1974)
"Analyse des Correspondances : Etude théorique et possibilités d'applications en hydrométéorologie"
Rapport interne de D.E.A.

LECLERC A. et AIACH P. (1978)
"Mesure de l'importance des valeurs propres en analyse des données. Application à l'A.C.P. de variables catégorisées"
Revue de Stat. Appli., vol.XXVI, n°1.

CHAPITRE III

SAMMON J.W. (1969)
"A non linear mapping for data structure analysis"
I.E.E.E. Transactions on computers, Vol.C 18, N°5, May 1969.

DER MEGREDITCHIAN G. (1973)
"Eléments de statistique multidimensionnelle appliquée à la météorologie"
Cours polycopié. Météo. Nat. Paris-Gran.

MORRISON D.F. (1967)
"Multivariate statistical Methods"
Mac Graws Hill Ed.

ANDREW D.F. (1973)
"Graphical techniques for high dimensional data" in "Discriminant Analysis and Applications"
Ed. by T. CACOULOS Academic Press.

MAAG U. (1978)
"Analyses statistiques de données de pollution atmosphérique de la région de Montréal"
Communication aux journées de statistiques. Nice : Mai 1978.

CHAPITRE IV

ROMEDER M. (1973)
"Méthodes et programmes d'analyse discriminante" Dunod Ed.

ULMO J. (1973)
"Différents aspects de l'analyse discriminante"
Revue de Statistique Appliquée, vol.XXI, n°2, pp 17-55.

BIOMEDICAL COMPUTER PROGRAMS BMD (1970)
Manuel d'utilisation. Edité par W.J. Dixon. University of California Press.

BOUROCHE J.M. et SAPPORTA G. (1978)
"L'analyse des données"
Pour La Science - N°5

CAILLEZ F., MAILLES J.P., NAKACHE J.P., PAGES J.P. (1971)
"Analyse des données multidimensionnelles"
Publié par Centre d'Etudes Economiques d'Entreprises, 116 Bd Pereire, Paris 17°.

BIBLIOGRAPHIE DE LA TROISIEME PARTIE

CHAPITRE I

COOLEY W. et LOHNES P.R. (1971)
"Multivariate data analysis"
John Wiley Ed. New York

VIALAR J. (1956)
"Calcul des probabilités et statistiques"
Cours de la Météorologie Nationale - Octobre 1956 - 4 tomes.

LEBART L. et FENELON J.P. (1973)
"Statistiques et informatique appliquées"
2ème ed. Dunod.

LAWLEY D.N. (1956)
"Test of significance for the latent roots of covariance and correlation matrices"
Biometrika, vol.43, p.128-136.

ANDERSON T.W. (1963)
"Asymptotic theory for principal component analysis"
Ann. Math. Stat. vol.34, pp 122-148.

ANDERSON G.A. (1965)
"An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix"
Ann. Math. Stat. vol.36, pp 1153-1173.

JAMES A.T. (1966)
"Inference on latent roots by calculation of hypergeometric functions of matrix arguments"
dans Multivariate Analysis (P.R. Krishnaiah Ed.) Academic Press - New York , pp 209-235.

MUIRHEAD R.J. (1978)
"Latent roots and matrix variates : a review of some asymptotic results"
Annals of Stat. vol.6 , n°1 (p.5-33).

MORRISON D.F. (1967)
"Multivariate statistical methods"
Mac Graw Hill Ed.

CRADDOCK J.M. (1973)
"Problems and prospects for eigen vector analysis in Meteorology"
The Statistician, Vol.22, N°2, pp 133-145.

FARMER S.A. (1971)
"An investigation into the results of principal components analysis of data derived from random numbers"
The statistician. Vol.20, n°4, p.63-72.

PROBERT-JONES J.R. (1973)
"Orthogonal pattern eigenvectors analysis of random and partly random fields"
3rd Conf. on probability and statistics in atmospheric science.
American Meteorological society. Boulder Co. June 1973.

FISHER et YATES (1953)
"Statistical tables for a biological, agricultural and medical research"
Oliver and Boyd Ed.

MARCHENKO A.S. et MINAKOVA L.A. (1972)
"Effect of sample connectedness on the accuracy of linear statistical forecasting and optimum predictor dimension"
Izvestia Atmospheric and Ocanic Physics, Vol.8, n°11, pp 1143-1153.

SHEPARD R.J. (1962)
"The analysis of proximities : Multidimensional scaling with an unknown distance function"
Psychometrika. Vol.27, N°2 et 3, pp 125-140 et 219-246)

SHEPARD R.N. et CAROLL J.D. (1966)
"Parametric representation of non linear data structure"
in Multivariate Analysis, P.R. Krisnaiah. Ed. Academic Press, New York, pp 561-591.

BENNET R.S. (1969)
"The intrinsic dimensionality of signals collections"
IEEE Trans. on Information Theory. Vol.IT 15, N°5, p.517-525.

MORIN G. (1974)
"Génération de chroniques de débit. Conception nouvelle utilisant les fonctions orthogonales naturelles"
Thèse de D.I., Grenoble, Juin 1974.

DIDAY E. (1973)
"Introduction à l'analyse factorielle typologique"
Rapport de recherche IRIA, n°27, Août 1973.

CHAPITRE II

LAFITTE P. (1972)
"Traité d'informatique géologique"
Ouvrage collectif. Masson ed.
Chap. 9 : "Corrélations géologiques et méthodes statistiques de traitement des données" par P. ISNARD, J.L. MALLET, P. CAZES et V. SATTRAN.

JOLLIFE I.T. (1972 et 1973)
"Discording variables in a principal component analysis. I - Artificial Data. II - Real Data."
Applied Statistics , Vol.21, n°2, pp 160-173 et vol22 n°1, pp 21-31.

BEALE E.M., KENDAL M.G., & MANN D.W. (1967)
"The discording of variables in multivariate analysis"
Biometrika, Vol.54, pp 357-366.

DUBAND D. (1974)
"Analyse en composantes principales des séries pluviométriques des Alpes , de l'Est du Massif Central et des Cévennes."
Rapport interne E.D.F.-D.T.G. Grenoble.

Rapport B.R.G.M. (1972)
"Bassin expérimental de l'Hallue. Résultats du traitement des données acquises de 1965 à 1970"
Avril 1972. Rapport 72 SGN 252 AME.

ZIRPHILE J. (1974)
"La classification automatique: étude bibliographique"
Publication interne. Société Alsthom DRE. Grenoble.

BESSON M. (1973)
"IPHIGENIE ou un nouveau procédé de typologie ." Rapport interne .
Ecocentre Mac Grégor COMARAIN - Ville d'Avray - Janvier 1973.

CHAPITRE III

COOLEY W.W. et LOHNES P.R. (1971)
"Multivariate Data Analysis"
John Wiley and sons Ed.

MILLER J.K. (1969)
"The development and application of bimultivariate correlation : a measure of statistical association between multivariate measurement sets"
Ph. D. State Univ. of New York. Buffalo.

ROMANOV L.N. et VINOGRADOVA G.M. (1975)
"Orthogonal Expansions of synoptic situations"
Izvestia Atmos. and Oceanic Physics. Vol.11, n°2, pp 118-124.

LEBART L. et FENELON J.P.(1973)
"Statistique et informatique appliquées"
2ème édition. Dunod Ed.

PECK E. (1972)
"Relation of orographic winter precipitation patterns to meteorological parameters"
Symposium on the distribution of precipitation in mountainous areas.
Geilo-Norway. Technical Papers WMO/OMM n°326 - Genève Suisse.

STEWART D. et LOVE W. (1968)
"A general canonical index"
Psychological Bulletin. Vol.70, n°3, p.160-163.

ROBERT P. et ESCOUFIER Y. (1976)
"A unifying tool for linear multivariate statistical methods : The RV - coefficient"
Applied statistics. Vol. 25, N°3, pp 257-265.

BIBLIOGRAPHIE DE LA QUATRIEME PARTIE

CHAPITRE I

BASS J. (1962)
"Eléments de calcul des probabilités" Ed. Masson et Cie

GNEDENKO B. (1976)
"The theory of probability" Ed. Mir.Moscow . Edition anglaise.

JOHNSTON J. (1963)
"Econometric Methods". International Student Edition. Mac Graw Hill Ed.

CHILES J.P. (1977)
"Géostatique des phénomènes non-stationnaires"
Thèse de Docteur Ingénieur. Université de Nancy, 1977.

CREUTIN D. (1979)
"Méthodes d'interpolation optimale de champs hydrométéorologiques. Comparaison et applications à une série d'épisodes pluvieux cévenols"
Thèse de Docteur Ingénieur - U.S.M.G., Déc. 1979.

CHAPITRE II

JOWETT G.H. (1955)
"The comparison of means of sets of observations from sections of independent stochastic series".
J. of Royal Statistical Society. Serie B - Vol.17 n°2 pp 208-227.

HUTCHINSON P. (1972)
"The use of a modified time series analysis technique for the determination of areal precipitation accuracies"
Symposium de Geilo - WMO/OMM . Distribution des précipitations dans les régions montagneuses. Pub. N°326 - Genève, 1972.

YEVJEVICH V. (1972)
"Stochastic Processes in Hydrology"
Water Ressources Publications. Fort Collins Colorado. 276 p.

DELHOMME J.P.(1977)
"Etude des relations pluies débits sur 3 sous-bassins de l'Orne"
Rapport interne. Centre d'informatique géologique. Fontainebleau.

CLARK R.T. (1977)
"Statistical methods for the study of spatial variation in hydrological variables" pp 299-314
in:New Facets in Hydrology. J.C. RODDA editor.

AMOROCHO J. et WU B. (1977)
"Mathematical models for the simulation of cyclonic storm sequences and precipitations fields"
J. of Hydrology, vol.32, pp 329-345.

HENDRICK R.L. et COMER G.H. (1970)
"Space variations of precipitation and implications for raingauge network design"
J. of Hydrology, vol.10 pp 151-163.

DIDAY E. (1973)
"Introduction à l'analyse factorielle typologique"
Rapport de recherche IRIA - Laboria n°27.

DIDAY E. et SCHROEDER A. (1974)
"The dynamic clusters methods in pattern recognition"
in Information Processing 74 - North Holland Publishing Company, pp 691-697.

MARCH W.J., WALLACE J.R. and SWIFT L.W. (1979)
"An investigation of storm type on precipitation in a small mountain watershed"
Water Res. Research, Vol.15 n°2, pp 298-304.

CHAPITRE III

BOX G.E.P. et JENKINS (1976)
"Time series analysis forecasting and control" Holden Day editor.

THIEBAUX H.J. (1976)
"Anisotropic correlation functions for objective analysis"
Monthly weather review, vol.104, pp 994-1002.

SOONG T.T. (1973)
"Random differential equations in science and engineering"
Academic Press. 327 p.

DELLEUR J.W. (1977)
"Equations différentielles stochastiques et leur application à la diffusion dans un champ aléatoire"
Rapport Interne E 45/77-15. Centre de Recherche E.D.F. Chatou.

WHITTLE P. (1954)
"On stationnary processes in the plane" Biometrika, vol.41, pp 434-449.

WHITTLE P. (1963)
"Stochastic processes in several dimensions"
Bulletin de l'Institut International de Statistique, vol.40, pp 974-994.

HEINE V. (1955)
"Models for 2-dimensional stationary stochastic processes"
Biometrika, vol.42, pp 170-178.

SAGAR B. (1978)
"Analysis of dynamic aquifers with stochastic forcing function"
Wat. Res. Research, vol.14 n°2, pp 207-216.

CHAPITRE IV

PAPOULIS A. (1965)
"Probability, random variables and stochastic processes "
International Student Edition, Mac Graw Hill ed.

KRASNOV M., KISSELEV A., MAKARENKO G. (1977)
"Equations intégrales" Editions Mir 1976. Moscou. Traduction française.

LEVINE B. (1973)
"Fondements théoriques de la radiotechnique statistique"
Editions Mir, Moscou 1973 (2 tomes)

FORTUS M.I. (1973)
"Statistically orthogonal functions for finite segments of a random process"
Izvestia Atmospheric and Oceanic Physics, vol.9 N°1, pp 34-46.

YAGLOM A.M. (1955)
"Extrapolation, interpolation et filtrage de processus aléatoires stationnaires à densité spectrale rationelle"
Trudy Mosk. Matem. Obshch. n°4, 1955, Traduit

OBUKHOV A.M. (1960)
"The statistically orthogonal expansion of empirical functions"
Izvestia Geophysical Serie - English Translation AGU, Novembre 1960 pp 288-291.

FORTUS M.I. (1975)
"Statistically orthogonal functions for a random field specified in a finite region of a plane"
Izvestia, Atmospheric and Oceanic physics, vol.11, N°11, pp 1107-1112.

MONIN A.S. et YAGLOM A.M. (1965)
"Statistical Fluid Mechanics : Mechanics of turbulence"
1965 en Russe. Edité par J.L. Lumley, MIT Press - 1975 - Vol.2.

BRILLINGER D.R. (1975)
"Time Series - Data Analysis and theory"
Internat. Series in Decision Processes. Holt, Reinhart et Winston Ed.

RESCH F.J. et ABEL R. (1975)
"Spectral analysis using Fourier transform techniques"
Internat. Journal for numerical methods in Engineering. Vol.9, pp 869-902.

RADIX J.C. (1970)
"Introduction au filtrage numérique". Ed. Eyrolles, 240 p.

HAMMING R.W. (1962)
"Numerical methods for scientists and engineers"
International Student Edition. Mac Graw Hill.

KUENY J.L.(1977)
"Contribution au traitement statistique de données météorologiques"
Thèse de Docteur-Ingénieur - Automatique - Grenoble.

SCHEID F. (1968)
"Numerical analysis". Schaum's outline series. Mac Graw Hill.

DEVILLE J.C. (1974)
"Méthodes statistiques et numériques de l'analyse harmonique"
Annales de INSEE. N°15. Jan.-Avril 1974, pp 1-101.

BASS J. (1968)
"Cours de Mathématiques". Masson et Cie Ed. 2 tomes.

BIBLIOGRAPHIE DE LA CINQUIEME PARTIE

CHAPITRE I

BOIS Ph. (1976)
Thèse de Doctorat op.cit.

ROMEDER J.M. (1973)
"Méthodes et programmes d'analyse discriminante" Dunod Ed. (274 p.)

KUENY J.L. (1977)
"Contribution au traitement statistique de données météorologiques"
Thèse de Docteur Ingénieur. INPG - Grenoble 1977.

MILLER R.G. (1962)
"Statistical Prediction by Discriminant Analysis"
Meteorological Monographs. Vol.4 N°25. Octobre 1962.

DE COURSEY D.G. (1970)
"Use of multiple discriminant analysis to evaluate the effects of land use change on the simulated yield of a watershed"
Ph. D. Thesis - Georgia Institute of Technology - Atlanta 1970.

MARDIA K.V. (1975)
"Assessment of Multinormality and the Robustness of Hotelling T^2 test"
Applied Statistics Vol.24 n°2 (1975).

CACOULOS T. (Editeur) (1973)
"Discriminant analysis and applications"
Academic Press 1973 (Proceedings of a Nado Advance Study Institute - Juin 1972 - Athènes)

ULMO J. (1973)
"Différents aspects de l'analyse discriminante"
Revue de Statistique Appliquée. Vol.XXI n°2, p.17-55.

BERNIER J. et ULMO J. (1973)
"Eléments de décision statistique" PUF Editeur.

NAKACHE J.P. et DUSSERRE L. (1975)
"Etude de problèmes posés par l'analyse discriminante linéaire en pas à pas"
Revue de Statistique Appliquée. Vol.XXIII N°3.

EZEKIEL M. et FOX K.A. (1970)
"Methods of correlation and regression analysis"
John Wiley and sons Ed. New York, 10° édition (546 pages)

MARCHENKO A.S. et MINAKOVA L.A. (1972)
"Effect of sample connectedness on the accuracy of linear statistical forecasting and optimum predictor dimension"
Izvestia, Atmospheric and Oceanic physics. Vol.8 n°11, p 1143-1153.

BASU J.P. et ODELL P.L. (1974)
"Effect of intraclass correlation amont training samples on the misclassification probabilities of Bayes' procedure"
Pattern Recognition. Pergamon Press. Vol.6 , p.13-16.

LACHENBRUCH P.A. (1974)
"Discriminant analysis when the initial samples are misclassified"
Technometrics. Vol.15 n°3, p.419-424.

CHHIKARA R.S. et ODELL P.L. (1973)
"Discriminant analysis using certain normed exponential densities with emphasis on remonte sensing application"
Pattern recognition. Pergamon Press. Vol.5, p.259-272.

PRESS S.J. (1968)
"Estimation from misclassified data" JASA Vol.63 N°Mars 1968 - p.123-133

ANDERSON J.A. (1973)
"Logistic discrimination" in CACOULOS op. cité - 1972.

ANDERSON J.A. (1974)
"Diagnosis by logistic discriminant functions : further practical problems and results".
Applied Statistics. Vol.23, N°3, 1974.

SEBESTYEN G.S. (1962)
"Decision making processes in pattern recognition"
Mac Millan Company Ed. (1962).

SAPPORTA G. (1977)
"Une méthode et un programme d'analyse discriminante pas à pas sur variables qualitatives"
Compte-rendus - 1ères journées internationales sur Analyse des Données et Informatique. Paris IRIA , pp 201-210, Septembre 1977.

LAFAYE de MICHEAUX D. (1978)
"Approximation d'analyses canoniques non linéaires de variables aléatoires et analyses factorielles privilégiantes"
Thèse de docteur ingénieur. Nice . IMAN - 1978.

PATRICK E.A. et FISHER F.P. (1969)
"Non parametric feature selection"
IEEE Trans. on Information Theory. Vol. IT 15 N°5. Sept. 1969.

GESSAMAN M.P. et GESSAMAN P.H. (1972)
"A comparison of some multivariate discrimination procedures"
J. of the American Statistical Association. Vol.67 N°338 p.468-472.

COVER T.M. et HART P.E. (1967)
"Nearest Neighbor classification
IEEE Trans. on Information Theory. Vol.13 n°1, 1967.

DER MEGREDITCHIAN (1972)
"Application de l'analyse discriminante à la prévision du verglas"
Note EERM - Météorologie Nationale - 1972.

NAKACHE J.P. (1978)
"Méthodes multidimensionnelles de classement"
Rapport Interne. Centre de Statistique Appliquée. Université de Paris VI. Mars 1978.

COLLOMB G. (1978)
"Estimation non paramétrique de la régression : regressogramme et méthode du noyau"
Publ. du Labo. de Stat., Univ. Paul Sabatier, n°07-78 (59 p.)

CHAPITRE II

ANDERBERG M.R. (1973)
"Clustering analysis for applications"
Academic Press. 1973 (359 p.)

ZIRPHILE J. (1974)
"Classification automatique : étude bibliographique"
Rapport Interne - Alsthom DRE - Grenoble 1974.

LANCE G.N. et WILLIAMS W.T. (1967)
"A general theory of class-sorting strategies : II Clustering systems"
Computer Journal Vol.10 n°3, 1967.

DIDAY E., SCHROEDER A. et OK Y. (1974)
"Dynamic clusters methods in pattern recognition"
in Information Processing North Holland Publishing Company - 1974 - p.691-697.

LECHEVALLIER Y. (1974)
"Optimisation de quelques critères en classification automatique"
Thèse de 3ème cycle. Université de Paris VI. Juin 1974.

VOGEL M.A. et WONG A.K.C. (1975)
"PFS optimal clustering method"
Technical Report. Biotechnology program. Carnegie Mellon University.
Pittsburg (Pa). 1975.

DIDAY E. (1976)
"Sélection typologique de paramètres"
Rapport de recherche IRIA, n°188, Août 1976.

GOVAERT G. (1975)
"Classification automatique et distances adaptatives"
Thèse de 3ème Cycle. Université de Paris VI. Mai 1975.

CHAPITRE III

Pour mémoire (pas de références)

CHAPITRE IV

GANDIN L.S. (1965)
"Objective analysis of meteorological fields"
Leningrad 1963. Israel Program for Scientific Translation. Jerusalem 1965

GANDIN L.S. (1970)
"The planning of meteorological stations networks"
Technical note n°111. OMM-WMO N°265 - TP 145.

CREUTIN D. (1979)
"Méthodes d'interpolation optimale de champs météorologiques. Comparaison et Application à une série d'épisodes pluvieux cévenols"
Thèse de Doct.-Ingénieur - Décembre 1979.

DELHOMME J.P. (1976)
"Applications de la théorie des variables régionalisées dans les sciences de l'eau".
Thèse de Docteur Ingénieur. Université P. et M. Curie. Paris VI.

SCHLATTER T.W. , BRANSTATOR G.W. et THIEL L.G. (1976)
"Testing a global multivariate statistical objective analysis scheme with observed data"
Monthly Weather Review. Vol.104 N°6 p.765-783.

THIEBAUX H.J. (1975)
"Experiments with correlation representations for objective analysis"
Monthly weather Review. Vol.103 N° p.617-626.

COHEN A. et JONES R.H. (1969)
"Regression on a random field"
American Statistical Association Journal. Dec. 1969. p.1172-1182.

HOLMSTROM I. (1963)
"On a method for parametric representations of the state of the atmosphere"
Tellus, Vol.15, p.127-149.

BUELL C. (1971)
"Integrale equation representation for factor analysis"
J. of Atmospheric Sciences. Vol.28 p.1502-1505 .

DEVILLE J.C. (1973)
"L'analyse harmonique dans le cas de données discrètes"
Note interne. INSEE - 25p.

DEVILLE J.C. (1974)
"Méthodes statistiques et numériques de l'analyse harmonique"
Annales de l'INSEE n°15. Janvier-Avril 1974. p.3-101

LEGRAS J. (1971)
"Méthodes et techniques de l'analyse numérique"
Dunod ed. (323 p.)

dernière page de la thèse

AUTORISATION DE SOUTENANCE

VU les dispositions de l'article 5 de l'arrêté du 16 Avril 1974,

VU les rapports de Messieurs :

- J. BERNIER, Ingénieur, Chef de Service à E.D.F. - CHATOU -
- M. BOUVARD, Professeur à l'Institut National Polytechnique de GRENOBLE
- G. ROMIER, Professeur à l'Université des Sciences Sociales de GRENOBLE

Monsieur Charles O B L E D

est autorisé à présenter une thèse en soutenance pour l'obtention du grade de DOCTEUR d'ETAT, discipline SCIENCES.

Grenoble, le 29 Novembre 1979

Le Président de l'U.S.M.G.

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Le Président de l'I.N.P.G.

Ph. TRAYNARD
Président
de l'Institut National Polytechnique