

STATISTIQUE ET ANALYSE DES DONNÉES

CLAUDE MANTE

**Analyse en composantes principales d'un processus
multiple non stationnaire : une application à
des données météorologiques**

Statistique et analyse des données, tome 14, n° 2 (1989), p. 25-53

http://www.numdam.org/item?id=SAD_1989__14_2_25_0

© Association pour la statistique et ses utilisations, 1989, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**ANALYSE EN COMPOSANTES PRINCIPALES D'UN PROCESSUS
MULTIPLE NON STATIONNAIRE : UNE APPLICATION A DES
DONNEES METEOROLOGIQUES**

Claude MANTE

Centre de Géomorphologie du C.N.R.S.
Rue des tilleuls - 14000 Caen

Résumé : On présente une méthode pour la recherche de séquences typiques dans une série chronologique multiple. Le problème étant d'abord posé dans le cadre de l'A.C.P. d'un processus, on proposera ensuite une technique plus appropriée : l'A.C.P. invariante sous l'action d'un groupe.

Abstract : A method to determine the evidence of typical sequences in a multiple time series is proposed. This question is first handled using the approach of P.C.A. for a stochastic process. A more suitable one is discussed afterwards : the shift-invariant P.C.A.

Mots clés : Processus multiple, variété d'orbites, pondération.

Indices de classification STMA : 06-070, 00-090, 00-060.

0 - INTRODUCTION

Les données servant de support à cette étude ont été recueillies à la crête de Vars (Alpes du sud), à 2500 m d'altitude. Les thermosondes de l'enregistreur sont placées en surface et à l'intérieur d'une paroi rocheuse, ainsi que dans l'air, à proximité de celle-ci. Seules trois sondes ont fonctionné tout au long de la période d'observation, du 21 août 1982 au 5 août 1985. Elles ont enregistré toutes les deux heures la température de l'air, et celle de la roche en sa surface et à 48 cm de profondeur.

Vue l'évidente non-stationnarité de cette série multiple, la F-ACP de Brillinger (1980) c.à.d., en bref, l'étude *fréquence par fréquence de la densité spectrale d'un processus stationnaire* n'est pas utilisable ici. De plus, notre objectif est la description "en

temps" du processus, et nous nous placerons dans l'espace des fréquences surtout pour des raisons pratiques (réduction du nombre de variables, "bon comportement" vis-à-vis des translations).

Notre approche initiale est plus apparentée à celle de Deville (1974), qui décompose un approximant de la fonction de covariance d'un processus (éventuellement vectoriel) de second ordre. La nature approximative de l'opérateur traité découle aussi bien de la nécessaire finitude du nombre de points de mesure (convergence en norme des opérateurs) que de celle de l'échantillon (convergence en moyenne quadratique des opérateurs). Deville soulève également le problème de l'interpolation du processus lorsqu'il n'est connu qu'en un nombre fini d'instant, et montre que les hypothèses faites sur la régularité des trajectoires conduisent à modifier de différentes manières la métrique de l'ACP. Besse (1979) poursuit dans cette direction, en prenant pour cadre privilégié de l'ACP les espaces de Sobolev (ACP filtrantes), ce qui est une manière de fixer la régularité des trajectoires. Les problèmes d'interpolation sont du même coup résolus, puisque ces espaces sont à noyau reproduisant ; le choix d'un espace se traduira encore par celui d'une métrique. Cette optique est particulièrement intéressante lorsque les données obéissent approximativement à une équation différentielle homogène de référence : $P(D)u = 0$. L'espace de Sobolev peut alors être scindé en deux sous-espaces : l'espace des "tendances" $\text{Ker}(P(D))$, et son orthogonal. Si l'on dispose de suffisamment de conditions aux limites ces trois espaces peuvent être munis de produits scalaires leur conférant la structure d'espace de Hilbert à noyau reproduisant, et l'ACP peut être entreprise dans l'un ou l'autre d'entre eux (Besse et Ramsay, 1986 ; Besse, 1988).

Dans le cas le plus simple la température $\theta(x,t)$ dans la paroi obéit à l'équation de la chaleur :

$$\partial\theta / \kappa\partial t - \partial^2\theta / \partial x^2 = 0 \quad \forall t > 0, \forall x > 0 .$$

Dans les faits, la diffusivité thermique κ n'a aucune raison d'être constante dans le temps et/ou l'espace, ce qui est une première difficulté. La deuxième réside dans le fait que nous n'avons ici qu'une contrainte (approximative) : il est en effet bien connu (cf. par exemple Carslaw and Jaeger, 1959) qu'à partir d'une profondeur p^0 d'une vingtaine de mètres la température est sensiblement constante (dans le sol), ce qui nous donne :

$$\partial\theta / \partial t(p^0,t) = 0 = \partial^2\theta / \partial x^2(p^0,t).$$

Etant donné que, de plus, nous ignorons la température à plus de 48 cm de profondeur, il semble difficile de traiter ces données dans le cadre de l'ACP filtrante

1 - CHOIX DE L'UNITE STATISTIQUE ET RECODAGE DES DONNEES (Manté, 1985)

La chaleur se propageant lentement dans la paroi, l'étude de profils instantanés ne présente guère d'intérêt. Un bloc de données quotidien serait a priori intéressant, mais pour la même raison les températures superficielles seront fort peu corrélées avec la température à 48 cm. En effet, la diffusivité thermique du calcaire étant de l'ordre de $0.007 \text{ cm}^2 \text{ s}^{-1}$ (Carslaw and Jaeger, 1959), une impulsion met environ 15h30 pour se propager au travers d'une couche de 48 cm de ce matériau. Il faut donc travailler sur des séquences plus longues, permettant de négliger ce déphasage. Nous avons choisi l'échelle décadaire qui présente deux avantages pratiques : d'une part il est commode de pouvoir mettre nos résultats en relation avec les statistiques décadaires publiées dans le Bulletin mensuel de la Météorologie Nationale, d'autre part un bloc de 11 jours correspond à 132 (environ 2⁷) mesures. On pourra donc recoder les décades en utilisant un programme de transformée de Fourier rapide (algorithme de Cooley-Tukey). Nous appellerons *décade-sonde* un ensemble de 128 mesures consécutives enregistrées par une thermosonde, et *décade* la "pile" de trois *décade-sondes* ainsi constituée.

Chaque *décade-sonde* est pré-traitée de la manière suivante :

- Transformation de Fourier discrète, après multiplication par une fenêtre de Hanning trapézoïdale.

- Filtrage préliminaire, par élimination d'un certain nombre (variable) d'harmoniques de faible énergie, assimilées au bruit de fond.

- Choix d'un nombre réduit d'harmoniques pour représenter les *décade-sondes*.

La sélection des harmoniques est le point le plus important. Elle est faite en conciliant deux points de vue complémentaires :

"Quantitatif" : on retient les harmoniques contribuant de manière importante au moment d'ordre deux (ACP non centrée) ou à la variance (ACP centrée) du spectre décadaire.

"Qualitatif" : le spectre des *décade-sondes* paraît être un mélange de plusieurs composantes à supports dans des bandes de fréquences distinctes. Il serait alors suffisant, pour avoir une description convenable de *l'évolution du processus*, de sélectionner quelques variables pertinentes dans chaque bande. L'analyse des correspondances met bien en évidence ce genre de structure (Manté et Ozouf, 1983). Nous avons analysé le tableau croisant les 213 spectres normés associés aux *décade-sondes* avec les 37 premières harmoniques, codées H0, H1, ..., H36. Le premier plan factoriel (63 % de l'inertie) montre l'existence de trois grandes composantes : la moyenne décadaire H0, un cycle décadaire à basse fréquence (H1,H2,H3) et un cycle quotidien associé aux harmoniques H9, H10 (période : 25.6h), H11 (23.3h), H12, H21 (12.2h), H22

(11.6h) et H32, de période 8h. Une classification des variables dans l'espace des 7 premiers facteurs (85 % de l'inertie) indique que la répartition suivante en cinq classes résume bien la structure générale des spectres (88 % de l'inertie résiduelle) : {H0}, {H1}, {H2}, {H10, H11, H21, H22, H32}, {autres harmoniques}.

Remarques :

Le choix d'un recodage par transformation de Fourier s'impose tout d'abord en raison de la périodicité marquée des données. Cette transformation a de plus un bon comportement vis-à-vis des translations (que nous exploiterons par la suite) et des dérivations (elle "diagonalise" les opérateurs différentiels à coefficients constants). On pourrait donc estimer (cf. Brillinger, 1980), à partir des transformées du signal enregistré à différentes profondeurs, les réponses spectrales et impulsionnelles de diverses "tranches" dans la paroi ; ces fonctions ne dépendent en principe que de κ , et en fournissent des estimateurs.

Pour d'autres données, on pourra bien entendu utiliser des polynômes orthogonaux différents, en fonction du support des courbes à traiter et des fonctions poids envisageables.

2 - L'ANALYSE EN COMPOSANTES PRINCIPALES DES DECADES :
(Manté et Coutard, 1988)

Notre fichier comporte 71 décades codées sous la forme ijCD où ij est le numéro de la décade et CD sont les deux premières lettres de son mois de début.

Notons $E(T)$ le complété de l'espace des fonctions continues sur $[0, T]$ et prenant la même valeur en 0 et T, pour le produit scalaire :

$$\langle f, g \rangle \equiv \frac{1}{T} \int_0^T f(t) g(t) dt .$$

$E(T)$ est un espace de Hilbert séparable admettant pour base $1_{[0, T]}$ et la famille de couples de fonctions :

$$\{ (\sqrt{2} \cos(2k\pi t / T), \sqrt{2} \sin(2k\pi t / T)) \mid k \geq 1 \} .$$

Le processus d'évolution des températures est considéré comme une variable aléatoire vectorielle associant à chaque décade un élément de l'espace $\mathbb{H} \equiv \bigoplus_{i=1}^3 E_1(T)$, avec

$E_i(T) = E(T)$, $\forall i = 1, 3$, qui est un espace de Hilbert pour la norme $\|f\|_H^2 = \sum_{i=1}^3 \|f_i\|_{E(T)}^2$. Pour faire l'ACP du nuage des décades dans H il est plus commode d'utiliser les coefficients de Fourier des decade-sondes : même lorsque le nombre des observations dans $[0, T]$ tend vers l'infini, nous pourrions nous contenter d'un nombre raisonnable d'harmoniques, et l'opérateur de covariance empirique approché ainsi construit est convergent (Manté, 1985).

A chaque axe factoriel de l'A.C.P. correspond, par transformée de Fourier inverse, une fonction propre (élément de H). Celles-ci sont des aides supplémentaires à l'interprétation des axes, et sont en fait plus "parlantes" que les habituelles listes de contributions. Les variables sont codées sous la forme sA_i ou sB_i , s étant le numéro de la sonde (1 = air, 2 = surface, 3 = 48 cm), i le numéro de l'harmonique, A désignant la partie paire et B la partie impaire.

Nous avons retenu les suivantes :

- pour la sonde 1 (air) : 1A0, 1A1, 1B1, 1A2, 1B2, 1A3, 1B3, 1A4, 1B4, 1A5, 1B5, 1A6, 1B6, 1A10, 1B10, 1A11, 1B11, 1A21, 1B21. Cela correspond à 96.1 % du moment d'ordre deux, et 96.7 % de la variance.

- Pour la sonde 2 (surface) : 2A0, 2A1, 2B1, 2A2, 2B2, 2A3, 2B3, 2A4, 2B4, 2A5, 2B5, 2A6, 2B6, 2A7, 2B7, 2A8, 2B8, 2A9, 2B9, 2A10, 2B10, 2A11, 2B11, 2A12, 2B12, 2A21, 2B21, 2A22, 2B22, 2A32, 2B32 ; soit 97.4 % du moment d'ordre deux et 96.2 % de la variance.

- Pour la sonde 3 (48 cm) : 3A0, 3A1, 3B1, 3A2, 3B2, 3A3, 3B3, 3A4, 3B4, ce qui donne 98.2 % du moment d'ordre deux et 98.3 % de la variance. Nous avons donc une bonne restitution avec 59 variables au lieu de 384. Remarquons que la température décadaire moyenne $A0$ contribue de manière très importante à l'inertie :

A0	AIR	SURFACE	48 cm
Moment d'ordre 2	69.8 %	66 %	91.3 %
Variance	75.7 %	67.5 %	89.4 %

Nous n'avons d'évidence pas affaire à un véritable échantillon ; par conséquent, nous avons d'abord soumis les données à une ACP non centrée. Ce traitement n'est pas équivalent à sa variante centrée (Dauxois et Pousse, 1976), sauf si le centre de gravité du nuage est orthogonal à toute réalisation centrée du processus, ce qui revient à dire qu'il est colinéaire à l'un des vecteurs propres de la matrice des moments. Le centre de gravité

n'est alors qu'un paramètre de translation, et l'on peut sans scrupule le prendre pour origine de l'espace (c'est ce qui se produit toujours en analyse des correspondances). Tel n'est pas le cas ici : le carré du cosinus entre la moyenne et l'axe 1 (resp. 2) est égal à 0.786 (resp. 0.211). La première fonction propre, engendrée essentiellement par les moyennes décadaires fait également intervenir le cycle quotidien. Les maxima étant généralement atteints vers 12 h, et les minima vers 2 h, la deuxième fonction propre (associée au cycle quotidien) semble avoir pour rôle essentiel de compenser les inversions de cycle liées aux changements de signe des températures moyennes, donc à l'opposition été-hiver (Manté, 1985). Ce phénomène est vraisemblablement en rapport avec la variation annuelle du flux lumineux solaire et de la durée du jour, qui modifient l'allure des cycles quotidiens (Kaiser B., 1987).

Critique de la méthode employée :

Tout d'abord, la variabilité des températures décadaires moyennes, qui est considérable, est susceptible de masquer des nuances plus subtiles. En particulier, on n'aboutit qu'à une classification peu satisfaisante des décades. Si l'on supprime les variables 1A0, 2A0 et 3A0 les résultats sont plus contrastés, mais ne semblent guère interprétables.

La méthode employée présente un deuxième défaut, qui tient à la manière naïve dont on a défini les décades. En effet, supposons que la morphologie du processus soit stable pendant plusieurs décades, ce qui revient à prolonger par périodicité la première d'entre elles. On aboutit alors à la situation suivante : certains blocs décadaires extraits de la longue période ainsi constituée peuvent être très éloignés de la décade initiale au sens de la métrique utilisée. Un phénomène analogue se produit également pour les fonctions propres (voir § 4). Cette incohérence résulte d'une conception trop restrictive de la morphologie. R. Thom (1977) définit la notion de forme de la manière suivante : "... si E est un espace topologique, G est un groupe (ou pseudo-groupe) opérant dans E , une G -forme est par définition une classe d'équivalence de fermés dans E modulo l'action de G ". L'adoption d'une notion de forme est donc équivalente au choix d'un groupe de transformations autorisées, une "forme" étant alors une orbite. De tels principes d'invariance ont déjà été associés aux techniques de l'analyse des données pour classer des réactions entre particules à haute énergie (Lütz et Maïti, 1978). On s'efforcera ici d'introduire des transformations qui, tout en respectant la structure temporelle des données, permettront de réduire la dispersion du nuage.

3 - PONDERATION VIA L'ANALYSE FACTORIELLE MULTIPLE CENTREE : (Escofier B. et Pagès J.P., 1984)

La température décadaire moyenne aux différentes profondeurs posant problème qu'elle soit ou non prise en compte il ne restait d'autre solution que d'avoir recours à un système de pondération convenable des deux groupes de variables en présence : celui associé à la "taille" (1A0, 2A0, 3A0), et celui associé à la "forme". L'A.F.M. permet précisément d'analyser un tel tableau, dans une optique descriptive (maximisation de la variance expliquée). Nous adopterons ensuite le système de pondération qui lui est attaché. Afin que la "forme" soit décrite au mieux, nous avons introduit quelques harmoniques complémentaires : H32 pour la sonde air, ainsi que H10, H11, H21 et H32 à 48 cm.

Nous avons en présence deux nuages \mathcal{N}_D^1 et \mathcal{N}_D^2 décrivant l'ensemble D des décades au moyen de chacun des deux groupes de variables. "L'AFM peut être vue comme une Analyse Multicanonique dans laquelle l'indice de liaison entre groupes de variables prend en compte la variance expliquée. Elle cherche des variables générales, (des directions communes aux nuages de variables), puis en déduit des variables canoniques qui traduisent cette direction pour chaque groupe et qui définissent des projections des \mathcal{N}_D^j susceptibles d'être superposées dans une représentation simultanée". Un groupe est ici d'autant plus important qu'il possède de nombreux degrés de liberté. Les moyennes décadaires étant très corrélées, le nuage des individus partiels associé au premier groupe est quasiment unidimensionnel : la première valeur propre ($\lambda_1^1 = 317.68$) correspond à 98.7 % de l'inertie. Le deuxième nuage est évidemment très différent, sa première valeur propre ($\lambda_1^2 = 22.3$) correspondant à 25.4 % de l'inertie, alors que les cinq premiers axes restituent 76 % de celle-ci.

Le tableau ci-après contient les coordonnées et les aides à l'interprétation des deux groupes.

GR	Q1	POID	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR	4=F	COR	CTR	5=F	COR	CTR	6=F	COR	CTR
GR 1	940	3	204	969	940	650	11	0	11	12	0	22	4	0	10	2	0	7	1	0	4
GR 2	938	66	795	521	133	349	982	473	988	544	145	977	458	103	989	358	63	992	204	20	995
				999	1490	1000	994	1000	556	1000	463	1000	463	1000	361	1000	205	999			

Il est clair que la première variable générale est la seule qui soit étroitement liée au premier groupe, sachant que la coordonnée du groupe j le long de la $s^{\text{ième}}$ variable générale peut s'interpréter de trois manières :

- c'est le poids (≤ 1) accordé à l'axe s par le groupe j
- c'est la contribution absolue du groupe à l'axe
- c'est une mesure de liaison entre le groupe et la variable générale.

Cela est confirmé par l'examen d'un deuxième tableau, dont les deux premières lignes recenseront les corrélations entre les variables canoniques attachées aux groupes et les variables générales, et dont la troisième ligne (Inertie inter / Inertie totale) permettra de déterminer dans quelle mesure les facteurs rendent compte des différences entre nuages partiels.

	F 1	F 2	F 3	F 4	F 5	F 6
GR 1	984	129	114	75	58	65
GR 2	949	995	980	990	992	997
Icg / Itot	874	503	484	492	491	499

La première variable générale apparaît comme un compromis entre "taille" et "forme" (fortes corrélations avec GR 1 et GR 2), tandis que les suivantes ne sont corrélées qu'avec les variables canoniques du 2^{ème} groupe. Il s'ensuit que la représentation des individus partiels relatifs au premier groupe n'est pertinente que le long de l'axe 1, ce qui se traduit par la baisse de l'indice Icg / Itot à partir de l'axe 2. Par contre, le long de l'axe 1, l'inertie du système des centres de gravité est égale à 87 % de Itot.

Première fonction propre (fig. 1-1) :

C'est donc un compromis entre les températures moyennes et la forme de décades. Engendrée par les variables 1A0, 2A0, 3A0 (dominantes), 1A1, 2A1, 3A1, 1A2, 2A2, 1A3, 2A3, 3A3, 1A4, 2A4, elle est approximativement paire. Cette variable est associée à l'opposition été-hiver, et tout particulièrement à des périodes de temps stable, présentant une légère oscillation décadaire.

Deuxième fonction propre (fig. 1-2) :

Elle est engendrée par 1B1, 2B1, 3B1 (dominantes), 1B2, 2B2, 3B2 et 3A1 ; elle est donc de type impair. Les décades liées à cet axe sont de périodes de changement : croissance des températures pour un point situé dans le demi-espace $F2 > 0$, décroissance sinon. On voit dans le premier plan (fig. 2) que les décades consécutives associées à cet axe sont fréquemment de signes opposés : cela correspond à des périodes d'une vingtaine de jours, où l'enveloppe des courbes de températures a une forme en cloche typique.

Figure 1-1

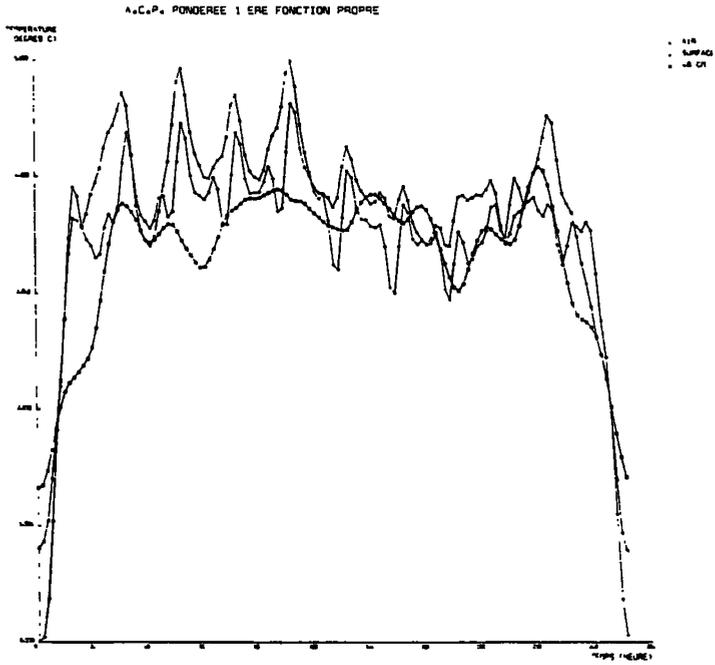


Figure 1-2

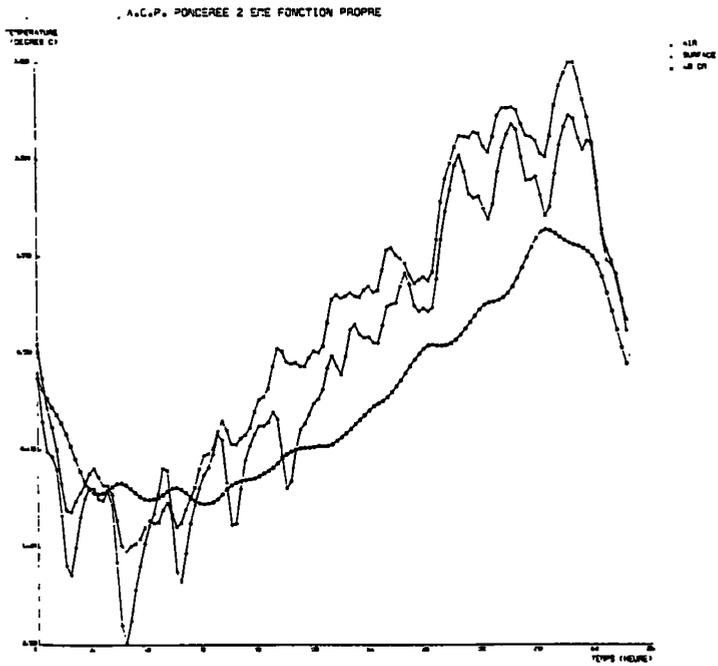


Figure 1-3

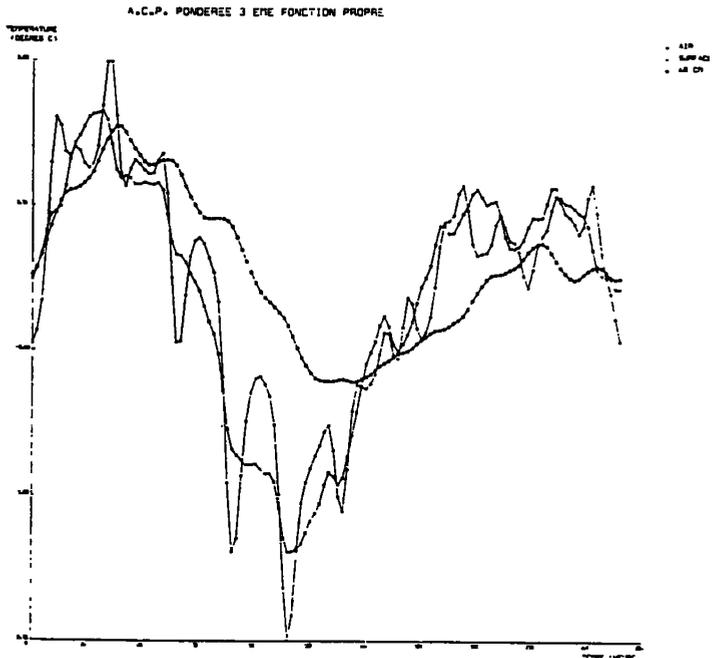


Figure 1-4

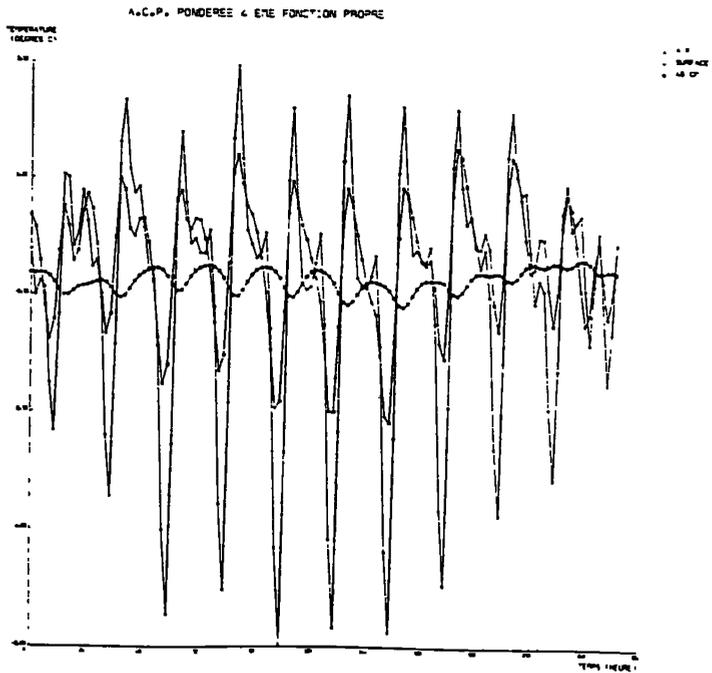


Figure 2

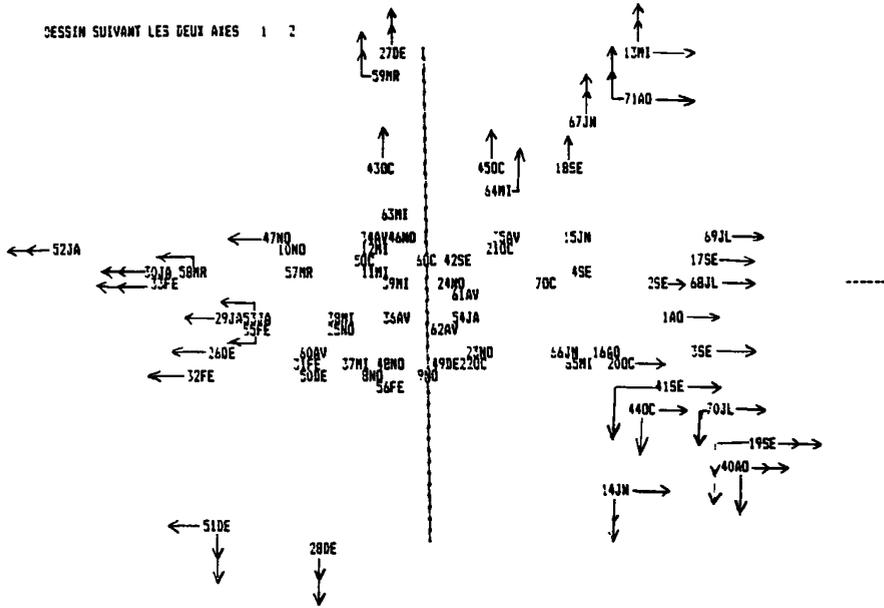
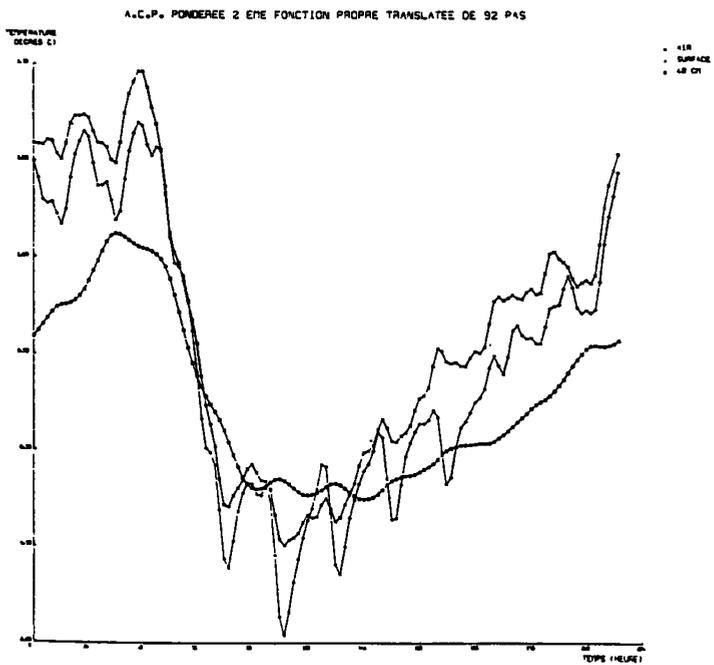


Figure 4



Troisième fonction propre (fig. 1-3) :

Elle est engendrée par : 1A1, 2A1, 3A1, 3B1, 1A2, 1B2, 2A2, 2B2, 3A2, 2A12.

Quatrième fonction propre (fig. 1-4) :

Elle est étroitement liée au cycle quotidien par les variables 1A10, 2A10, 1A11, 2A11, 1A21, 2A21, 1B21, 2B21, 2A22, 2B22 et 2B32. Outre la forme particulière de cette fonction en surface (faisant penser à un phénomène de modulation), on remarque l'écart considérable entre les minima en surface et dans l'air. Cet axe oppose deux groupes de décades : {20OC, 21OC, 22OC, 23NO, 24NO, 25NO, 26DE, 27DE} du côté $F_4 < 0$, et {40AO, 41SE, 42SE, 43OC, 44OC, 45OC} du côté $F_4 > 0$.

Le signe négatif du premier groupe implique une augmentation par rapport à la moyenne de l'écart entre maxima en surface et dans l'air. On trouvera dans le bulletin mensuel diffusé par la Météorologie Nationale les renseignements suivants, recueillis à la station météorologique d'Embrun :

octobre 1983 :

insolation : supérieure à la moyenne de 20 h

précipitations : déficit de 23 mm par rapport à la moyenne

température : supérieure de 0.8 °C à la moyenne

novembre 1983 :

insolation : excès de 43 h

précipitations : déficit de 10 mm

température : excès de 1.9 °C.

Nous avons affaire à une période très clémente pour la saison, mais entrecoupée assez régulièrement de courts épisodes de mauvais temps.

Quant au deuxième groupe nous avons :

septembre 1984 :

insolation : déficit de 11 h

pluviométrie : excès de 15 mm

température : déficit de 1.8 °C

octobre 1984 :

insolation : déficit de 3 h

pluviométrie : excès de 28 mm

température : 0.1 °C au-dessous de la moyenne.

Ce groupe se signale par un ensoleillement médiocre, associé à une pluviométrie supérieure à la moyenne. Cet axe semble donc associé à l'ensoleillement et à la pluviométrie.

En conclusion de ce paragraphe, on examinera (fig. 3) l'arbre hiérarchique obtenu en classant les décades dans l'espace des sept premiers facteurs (programme SIMCAH cf. Jambu et Lebeaux, 1978). Les simulations montrent une bonne stabilité de la partition en

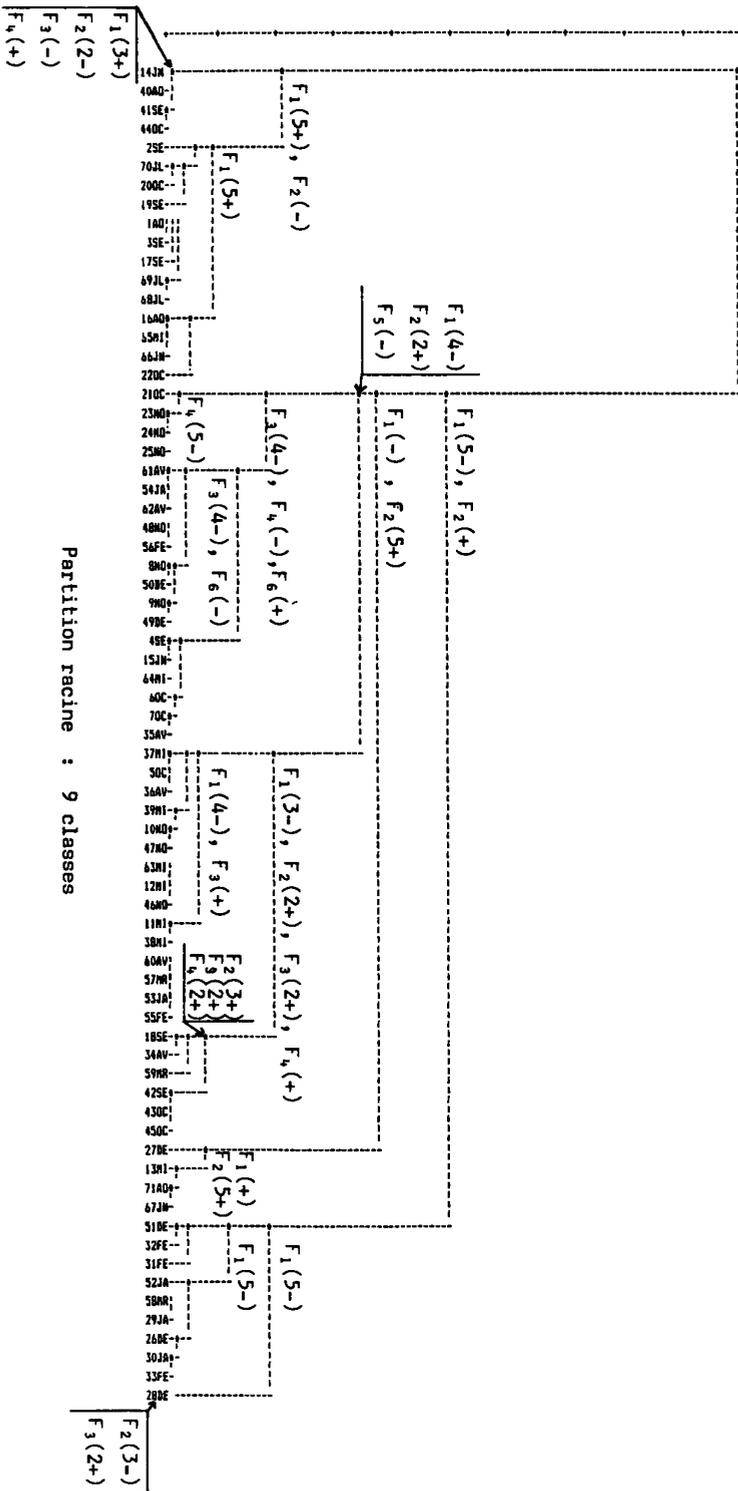
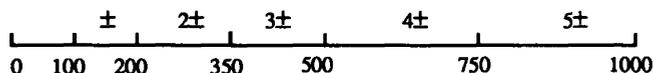


Figure 3

cinq classes ; une partition plus fine, en neuf classes semble également assez fiable. L'étiquetage de l'arbre repose sur l'échelle ci-dessous, proposée dans [Interpret. CAH] pour le critère COR :



4 - L'ANALYSE FACTORIELLE DANS LES ESPACES ORBITAUX

Le moment est venu d'appliquer les principes exposés à la fin du §2. Il est tout d'abord naturel d'exiger que les résultats de l'analyse factorielle soient invariants par translation. Or, les cosinus entre fonctions propres translatées (voir le tableau ci-dessous) montrent qu'il n'en est rien : en opérant un certain nombre de translations (de pas = 2h) sur la fonction propre en ligne, elle cesse d'être orthonormale à celle en colonne. Par exemple, la deuxième fonction propre décalée de 92 pas (fig. 4) est fort semblable à la troisième fonction propre (fig. 1-3).

		1	2	3	4	5	6	7
1	Cos	1.000	0.304	0.444	0.016	0.177	0.172	0.278
	Nb pas	0	83	71	53	57	118	121
2	Cos		1.000	0.840	0.179	0.558	0.430	0.210
	Nb pas		0	92	82	72	22	32
3	Cos			1.000	0.203	0.674	0.459	0.348
	Nb pas			0	119	114	66	57
4	Cos				1.000	0.336	0.505	0.480
	Nb pas				0	118	14	126
5	Cos					1.000	0.638	0.415
	Nb pas					0	80	77
6	Cos						1.000	0.483
	Nb pas						0	3

Nous avons donc là une importante source de redondance : un cercle dans le plan (2, 3) pourrait par exemple être associé à une unique "forme" (invariante par translation). Il est possible d'introduire des notions plus larges de "forme", par le biais d'autres groupes de symétrie. On peut par exemple souhaiter identifier entre elles des décades à tendance marquée, indépendamment du sens de cette tendance ; on fera alors intervenir le groupe multiplicatif $\{-1, +1\}$. Cette identification entre une décade D et son opposée -D n'est cependant pas réaliste. En effet, en raison de sa dissymétrie, la structure du cycle quotidien n'est pas conservée par cette transformation, même si l'on fait conjointement

agir le groupe des translations. De plus, cela conduirait à identifier entre elles décades froides et chaudes ! Ces obstacles peuvent être contournés dans l'espace des fréquences : il suffira de faire opérer la symétrie partielle (filtre) laissant invariantes, pour chaque décade-sonde, les harmoniques {H10, H11, H21, H22, H32} d'une part et H0 d'autre part (cf. §1).

Nous ferons donc agir deux groupes : un groupe de translations, noté G_1 , et un autre groupe, G_2 , qui sera le produit de G_1 et du groupe multiplicatif à deux éléments attaché à la symétrie partielle choisie. Comme dans le paragraphe précédent, les variables du groupe j sont pondérées par $(\lambda_1^j)^{-1/2}$.

4.1- LA PREMIERE DISTANCE ORBITALE

Le cycle quotidien étant une caractéristique importante, afin de limiter le volume des calculs, nous faisons agir le plus petit sous-groupe possédant la translation de 24h parmi ses éléments. G_1 est donc le sous-groupe (d'ordre $\mathfrak{O}(G_1) = 32$) engendré par la translation de 8h. Nous démontrons en annexe que, sous des hypothèses raisonnables, le nuage des points associés à la $s^{\text{ième}}$ sonde est situé dans un ouvert connexe par arcs de l'espace des fréquences, $D_1(K^s)$, qui ne dépend que de l'ensemble K^s constitué par les indices des harmoniques prises en compte pour la sonde s . Le $s^{\text{ième}}$ nuage de G_1 -décade-sondes est alors supporté par l'espace $D_1(K^s) / G_1$, qui peut être muni d'une structure de variété connexe. Les translations étant des isométries, on peut définir sur cet espace une distance orbitale entre décade-sondes (Lütz et Maïti, 1978) :

$$D_{or_1}(a, b) \equiv \text{Min} \{d(a, g(b)) \mid g \in G_1\}.$$

Nous montrons en annexe que c'est la distance géodésique entre l'orbite de a et celle de b -éléments de $D_1(K^s) / G_1$ - si cette variété est munie de la structure riemannienne attachée à la métrique identité. Chaque variété d'orbites a pour dimension le nombre de variables retenues, soit : 21 pour la sonde air, 31 pour la surface et 17 pour la dernière.

Il y a plusieurs manières de construire la variété des G_1 -décades. La plus immédiate serait de se placer sur la variété produit : $\frac{D_1(K^1)}{G_1} \times \frac{D_1(K^2)}{G_1} \times \frac{D_1(K^3)}{G_1}$.

Cette variété de dimension 69, à $\mathfrak{O}(G_1)^3$ feuilletés, est difféomorphe à la variété $(D_1(K^1) \times D_1(K^2) \times D_1(K^3)) / G_1^3$ (Dieudonné, 1970). La distance géodésique (pour la structure riemannienne produit) est alors définie par :

$$\delta_1(a, b) \equiv \sqrt{\text{Dor}_1^2(a_1, b_1) + \text{Dor}_1^2(a_2, b_2) + \text{Dor}_1^2(a_3, b_3)} .$$

Dans cet espace, la cohérence temporelle (plus généralement : relative à l'axe des abscisses) des mesures aux différents points n'est pas respectée. Afin de la préserver, nous nous placerons sur une variété plus "grande", définie ci-dessous.

Soit $\mathbb{G}_1 \equiv \{(g, g, g) \mid g \in G_1\}$ la diagonale de G_1^3 . C'est un sous-groupe de ce dernier, et il est évidemment sans point fixe dans $D_1(K^1) \times D_1(K^2) \times D_1(K^3)$, qui est un ouvert connexe par arcs de \mathbb{R}^{69} . On montrera donc sans difficulté que $(D_1(K^1) \times D_1(K^2) \times D_1(K^3)) / \mathbb{G}_1$ est une variété de dimension 69, à $\mathfrak{G}(G_1)$ feuilletés. La distance géodésique sur cette variété est :

$$\Delta_1(a, b) \equiv \sqrt{\text{Min} \left\{ \sum_{i=1}^3 d^2(a_i, g(b_i)) \mid g \in G_1 \right\}} .$$

Il n'y a en principe aucune raison pour que le nuage des G_1 -décades possède une image euclidienne, mais il est toujours possible de projeter la matrice des interdistances. Lorsque la plus petite des valeurs propres négatives est non négligeable, une double décomposition du tableau, sur la base du signe des valeurs propres, peut être intéressante (Lacourly et Lagos, 1981) ; on pourra aussi faire appel à d'autres techniques (cf. §4.3). Dans le cas contraire, il suffit d'une légère modification des interdistances pour rendre la matrice de Torgerson s.d.p. (Caillezeau et Pagès, 1976). Nous avons : $|\lambda_{71}| \approx \lambda_{28}$; sachant que la trace de la matrice est égale à la moitié de la moyenne des carrés des distances, $\overline{\Delta_1^2}$, le terme correctif $C^2 = 2 \cdot |\lambda_{71}|$ est de l'ordre de $\overline{\Delta_1^2} / 148$. La trace corrigée vaut : Trace + $70 \cdot |\lambda_{71}|$. Une preuve de l'efficacité de la méthode est la diminution de l'inertie totale lorsque l'on passe du nuage des décades brutes à celui des G_1 -décades : nous avons une chute de 35 % (sans correction) ou de 4.85 % (avec correction).

Nous n'avons plus de variables, donc plus de fonctions propres ; les seules aides à l'interprétation sont les décades elles-mêmes, au travers de leurs associations avec les axes. Les résultats antérieurs gardent donc tout leur intérêt, même si le cadre géométrique actuel est mieux adapté.

Nous ne nous étendrons pas sur les résultats de cette analyse, car ils sont très semblables à ceux que nous obtiendrons dans le paragraphe suivant.

4.2 - LA DEUXIEME DISTANCE ORBITALE

G_2 est le produit de G_1 par le groupe à deux éléments représentant la symétrie partielle. A la $s^{\text{ième}}$ décade-sonde est attaché l'ensemble $\text{Fix}(K^s)$ des indices des harmoniques de signe constant, et un domaine $D_2(K^s; \text{Fix}(K^s))$ ouvert et connexe par arcs, sans point fixe pour G_2 (voir annexe). Sous des hypothèses raisonnables, cet ensemble contient le $s^{\text{ième}}$ nuage de décade-sondes. Nous avons :

- $\text{Fix}(K^1) = \{0, 10, 11, 21, 32\}$
- $\text{Fix}(K^2) = \{0, 10, 11, 21, 22, 32\}$
- $\text{Fix}(K^3) = \{0, 10, 11, 21, 32\}$

Nous procéderons comme au §4.1, en plaçant les G_2 -décades sur la variété :

$$(D_2(K^s; \text{Fix}(K^s)) \times D_2(K^s; \text{Fix}(K^s)) \times D_2(K^s; \text{Fix}(K^s))) / \mathbb{G}_2$$

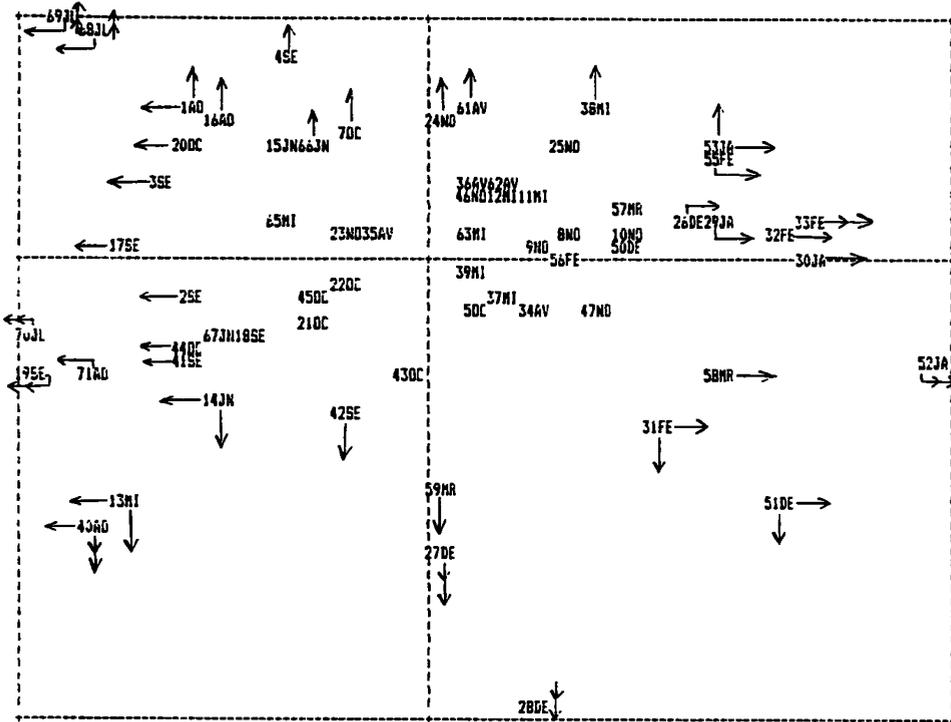
où \mathbb{G}_2 désigne la diagonale de G_2^3 . C'est une variété de dimension 69, à $2 \cdot \mathbb{G}(G_1)$ feuilletés. La distance géodésique est :

$$\Delta_2(a, b) \equiv \sqrt{\text{Min} \left\{ \sum_{i=1}^3 d^2(a_i, g(b_i)) \mid g \in G_2 \right\}}.$$

La décomposition du tableau d'interdistances montre qu'une plus grande compacité du nuage s'accompagne d'un meilleur ajustement par une image euclidienne. Les valeurs propres négatives sont très faibles en valeur absolue, et nous avons maintenant : $C^2 \simeq \overline{\Delta_1^2} / 221$; la baisse d'inertie relativement à l'ACP est égale à 38 % (sans correction), ou 19 % (avec correction).

Les premiers plans factoriels sont quasiment identiques pour les deux espaces d'orbites. Le premier axe oppose les périodes chaudes aux périodes froides, alors que le second oppose les décades à tendance marquée (du côté $F_2 < 0$) à celles qui sont plutôt "plates". On s'en convaincra en comparant le premier plan (fig. 5) à celui obtenu dans le cadre de l'ACP pondérée (fig. 2). Les trois éléments dont la tendance apparaissait comme particulièrement nette (27DE, 28DE, 59MR) se retrouvent voisins ; il en est de même pour le groupe {14JN, 40AO, 41SE, 44OC, 19SE, 70JL} (qui était de profil $F_1(3+)$ et $F_2(2-)$: chaleur et tendance décroissante) et le groupe {13MI, 67JN, 71AO} (de profil $F_1(+)$ et $F_2(5+)$: chaleur et tendance croissante), maintenant voisins dans le quadrant inférieur gauche.

Figure 5



Une fois les décades projetées dans l'espace des q premiers facteurs, on pourra sans difficulté les soumettre à tout programme opérant dans le cadre euclidien. On prendra garde que le résultat ne concerne que l'*image euclidienne* du nuage, car, comme le font remarquer Lütz et Maïti, on ne peut généralement pas définir la moyenne entre plus de deux orbites.

La hiérarchie obtenue en classant les orbites dans l'espace des 15 premiers facteurs de la même manière qu'au §3 (SIMCAH) montre une grande ressemblance entre l'arbre des G_1 -décades et celui des G_2 -décades, reproduit sur la figure 6. Dans les deux cas, les simulations montrent une bonne stabilité de la partition en 6 classes. Dans le tableau ci-dessous, nous comparons au moyen de la distance de Jaccard les partitions obtenues en classant les décades (A_1, \dots, A_6), les G_1 -décades (B_1, \dots, B_6), et les G_2 -décades (C_1, \dots, C_6).

	C1	C2	C3	C4	C5	C6
A1	0.53	1.0	0.91	1.0	1.0	1.0
A2	0.78	0.63	1.0	1.0	0.97	0.97
A3	1.0	1.0	0.80	1.0	1.0	0.75
A4	0.94	0.65	1.0	0.81	0.93	1.0
A5	1.0	1.0	1.0	0.88	0.50	0.86
A6	1.0	1.0	1.0	1.0	1.0	0.64
B1	0.39	1.0	1.0	1.0	1.0	1.0
B2	0.81	0.37	1.0	1.0	1.0	1.0
B3	1.0	1.0	0.0	1.0	1.0	1.0
B4	1.0	1.0	1.0	0.17	1.0	1.0
B5	1.0	0.94	1.0	0.95	0.27	0.96
B6	1.0	0.94	1.0	1.0	1.0	0.23

L'examen de la partie haute du tableau montre que la distance usuelle et la deuxième distance orbitale n'induisent pas la même structure hiérarchique sur l'ensemble des décades : la distance minimale entre parties est égale à 0.5, et il n'apparaît pas d'association biunivoque entre elles. Dans la partie basse du tableau, par contre, les classes peuvent être appariées, et la distance maximale sur la diagonale est inférieure à 0.4. Nous allons maintenant décrire les classes obtenues.

La classe C1 contient les décades les plus froides, et a pour profil : $F_1(5+)$; on y trouvera deux périodes stables :

29JA → 33FE (du 11 janvier au 6 mars 1984)

50DE → 53JA (du 17 décembre 1984 au 29 janvier 1985)

La classe C2, de profil $(F_1(2+), F_3(3+), F_7(2+))$ rassemble des G_2 -décades plutôt proches de l'origine de l'image euclidienne, et en particulier les suivantes :

50C → 70C (du 4 octobre au 5 novembre 1982)

34AV → 37MI (du 6 avril au 19 mai 1984)

450C → 49DE (du 23 octobre au 16 décembre 1984)

61AV → 64MI (du 17 avril au 30 mai 1985).

La classe C3, de profil $(F_2(4-), F_3(-))$ n'est composée que de 27DE et 28DE. Nous avons ici une situation exceptionnelle, avec des tendances très marquées.

La classe C4 qui s'agrègera plus tard à la précédente, se limite à la séquence 200C → 25NO, et a pour profil : $(F_3(3-), F_4(3-))$. Nous mettons ici en évidence un type de

REPRESENTATION DE LA CLASSIFICATION HIERARCHIQUE

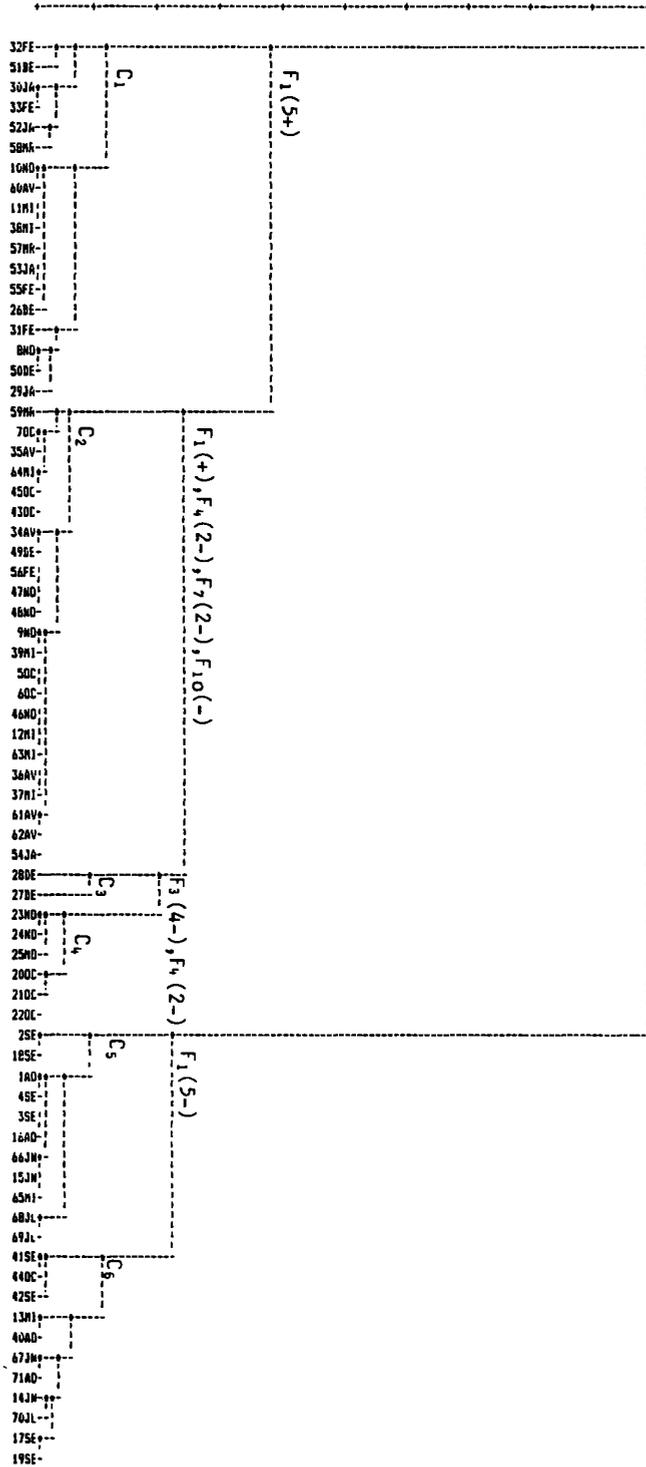


Figure 6

temps très persistant (apparemment associé au passage de trains de perturbations) tout au long de la période du 4 octobre au 8 décembre 1983.

La classe C5 a pour profil : $(F_1(4-), F_2(2+))$; elle contient donc des décades chaudes sans tendance marquée, soit :

1AO \rightarrow 4SE (du 21 août au 3 octobre 1982)

65MI \rightarrow 66JN (du 31 mai au 21 juin 1985)

68JL \rightarrow 69JL (du 3 au 24 juillet 1985).

La classe C6, enfin, a pour profil : $(F_1(5-), F_2(-))$. Ce sont des périodes très chaudes présentant une tendance relativement marquée. Cette classe contient les séquences ci-dessous :

13MI \rightarrow 14JN (du 31 mai au 21 juin 1983)

40AO \rightarrow 42SE (du 29 août au 30 septembre 1984)

70JL \rightarrow 71AO (du 25 juillet à la fin des mesures, le 16 août 1985).

4.3 - QUELQUES EXTENSIONS

La méthode présentée ci-dessus peut être généralisée, aussi bien dans ses aspects géométriques que statistiques.

Pour ce qui est du premier point, le support de chaque nuage approché peut être situé sur une sous-variété non triviale X de dimension p et de classe C^k de l'espace substrat \mathbb{R}^n , au lieu d'un simple ouvert. Par exemple, il est souvent utile de normer les objets étudiés lorsqu'on ne s'intéresse qu'à leur forme ; le nuage est alors porté par une sphère (cf. par exemple Le Nouvel, 1984 ; Manté et Coutard, 1988).

Le groupe d'isométries considéré est le reflet de l'ensemble des symétries admises. Si plusieurs symétries doivent être intégrées, le groupe résultant sera (comme G_2) le produit des groupes élémentaires, ou plus généralement une extension semi-directe de ceux-ci.

Soit \mathcal{G} ce groupe ; s'il est proprement discontinu sans point fixe sur un ouvert U de \mathbb{R}^n , U / \mathcal{G} est une variété séparée C^∞ de dimension n , telle que la surjection canonique Π est un revêtement à $\mathcal{G}(\mathcal{G})$ feuillet de U / \mathcal{G} (Berger et Gostiaux, 1972). En pratique U sera le complémentaire dans \mathbb{R}^n de l'ensemble des points fixes de \mathcal{G} , qui est fermé ($\mathcal{G} \subset \text{Diff}(\mathbb{R}^n)$). \mathcal{G} sera alors proprement discontinu sans point fixe sur $\tilde{X} \equiv X \cap U$, et $Y \equiv \tilde{X} / \mathcal{G}$ sera muni d'une structure de variété de classe C^k et de dimension p . Définissons sur cette variété une structure riemannienne en associant à tout point $y \in Y$ une forme quadratique $M_{(y)}$ définie sur l'espace tangent à Y en y . L'application tangente $\Pi'(\tilde{x})$ étant égale à l'identité $\forall \tilde{x} \in \tilde{X}$, on vérifie (voir en annexe) que tous les

relèvements sur \tilde{X} d'un chemin sur Y ont la même longueur. Si \tilde{X} est connexe, il en est de même pour Y , et il y a identité entre la distance orbitale et la distance riemannienne sur Y .

On se limitera au cas d'une métrique constante, le calcul des distances géodésiques devenant complexe (même avec M constante) dès que X n'est pas un ouvert de \mathbb{R}^n . Par exemple, si X est un ouvert de $S_{n-1}(0, r)$, et si $M \neq \text{Id}$, le calcul de la distance intrinsèque sur X ne sera pas trivial.

Remarquons enfin que le choix de la structure riemannienne produit, s'il est simple et consacré par l'usage en analyse des données, n'a rien d'intrinsèque!

Pour ce qui est des aspects statistiques, les tableaux d'interdistances orbitales sont également justiciables d'autres traitements que ceux qui ont été entrepris ci-dessus. La méthode de décomposition de la matrice d'interdistances ne se justifie pleinement que dans le cadre euclidien, mais a l'avantage d'être strictement compatible avec l'ACP, ce qui nous a permis de comparer les résultats des différentes analyses. D'autres méthodes factorielles, telles que MDSCAL seraient vraisemblablement mieux adaptées.

L'algorithme de classification utilisé présente les mêmes avantages et inconvénients. Vu la nature des distances représentées l'emploi du critère du plus proche voisin pour une éventuelle CAH serait tout à fait conséquent, ce qui nous amène tout naturellement aux représentations arborées (Barthélemy J.P. et Guénoche A., 1988), qui se sont déjà révélées très séduisantes en climatologie (Joly D. et Luong X., 1986).

5 - CONCLUSION

La méthode présentée convient surtout à l'étude d'un processus non stationnaire observé sur une longue période. Rien n'empêche cependant de traiter de la même manière un échantillon d'un processus, et les principes d'invariance introduits au §4 pourraient être très utiles pour la comparaison de réalisations asynchrones (pour une autre approche, voir Picard, 1987).

Si l'on a affaire à un processus stationnaire, le point de vue de la F-ACP est meilleur, puisque les "individus" forment un authentique échantillon, et que seules les caractéristiques des processus principaux sont significatives. Thiam (1984) a montré qu'alors, la matrice de covariance du processus possède théoriquement un spectre très particulier, puisque toutes ses valeurs propres (sauf éventuellement deux d'entre elles) ont une multiplicité paire, et sont identiques à celles de la F-ACP. Les vecteurs propres appariés (cas des valeurs propres doubles) sont liés à une même fréquence λ_k , les densités spectrales des composantes principales (stationnaires) associées sont identiques,

et de la forme $M_i(\lambda_k) \cdot \delta_{\lambda_k}(\lambda)$. Si le processus traité peut être modélisé comme somme d'une partie stationnaire et d'un trend, l'étude des résultats de l'ACP devrait permettre d'identifier et de séparer ces deux composantes, si aucune valeur propre n'est de multiplicité ≥ 4 .

BIBLIOGRAPHIE

- Barthélemy, J.P., Guénoche, A.**, Les arbres et les représentations des proximités, Masson, 1988.
- Berger, M., Gostiaux B.**, Géométrie différentielle, Armand Colin, 1972.
- Besse, P.**, Etude descriptive d'un processus, Thèse de 3^{ème} cycle, Toulouse, Université Paul Sabatier, 1979.
- Besse, P.**, Spline functions and optimal metric in linear principal component analysis, In : Component and correspondence analysis, Jan L.A. van Rijkevorsel and Jan de Leeuw Ed., John Wiley and Sons, p. 81-101, 1988.
- Besse, P., Ramsay, J.O.**, Principal component analysis of sampled functions, Psychometrika, 51, 2, p. 285-311, 1986.
- Brillinger, D.R.**, Time series : Data analysis and theory, Holden Day, 1980.
- Caillez, F., Pagès, J.P.**, Introduction à l'analyse des données, S.M.A.S.H., 1976.
- Carslaw, H.S., Jaeger, J.C.**, Conduction of heat in solids, Oxford University Press, 1959.
- Dauxois, J., Pousse, A.**, Les analyses factorielles en calcul des probabilités et en statistiques : essai d'étude synthétique, Thèse de Doctorat d'Etat, Toulouse, Université Paul Sabatier, 1976.
- Deville, J.C.**, Méthodes statistiques et numériques de l'analyse harmonique, Annales de l'INSEE, 15, p. 3-97, 1974.
- Dieudonné, J.A.**, Eléments d'analyse, tomes 3&4, Gauthier-Villars, 1970.
- Escofier, B., Pagès, J.P.**, L'analyse factorielle multiple, Cahier du BURO N° 42, Université Paris VI, 1984.
- [Interpret CAH], par **Benzécri, J.P., Benzécri, F., Cheung, Y.L., Maïza, S.**, Les Cahiers de l'analyse des données, 10, 3, p. 311-338, 1985.
- Jambu, M., Lebeaux, M.O.**, Classification automatique pour l'analyse des données, tome 2, 1978.
- Joly, D., Luong, X.**, Analyse des successions de types de temps fondée sur un nouvel algorithme de classification non hiérarchique, Cahiers de géographie de Besançon, Séminaires et notes de recherche N° 28, p. 317-322, 1986.

Kaiser, B., Les versants de la Vanoise. Enjeux traditionnels et fonctionnement morphoclimatique, Livre II : mesures climatiques et morphodynamiques, Thèse de Doctorat d'Etat, Université Paris VII, 1987.

Lacourly, N., Lagos, M., Représentation Lorentzienne d'une table de dissimilarités, Actes des journées de classification de Nancy, CNET, p. 213-224, 1981.

Le Nouvel, J., Analyse morphologique de courbes et application à un problème médical, Congrès européen de classification de Jouy-en-Josas, 1983.

Lütz, P., Maïti, D., Classification automatique d'après la distance entre orbites : application à la physique corpusculaire, Les Cahiers de l'analyse des données, 3, 4, p. 449-458, 1978.

Manté, C., Evolution du champ de température dans une paroi rocheuse naturelle : le cas de la Crête de Vars, Bulletin N° 30 du Centre de géomorphologie du CNRS, Caen, 1985.

Manté, C., Coutard, J.P., Recherches sur les situations thermiques types dans une paroi rocheuse. Détection de périodes propices à la cryoclastie, Bulletin N° 34 du Centre de géomorphologie du CNRS, Caen, 1988.

Manté, C., Ozouf, J.C., Approche statistique de la granulométrie, Actes des 12^{ème}, 13^{ème}, 14^{ème} colloques sur les méthodes mathématiques appliquées à la géographie, Cahiers de géographie de Besançon, Séminaires et notes de recherche N° 29, p. 139-160, 1988.

Picard, J., Classification des profils évolutifs incomplets et asynchrones, Revue de statistique appliquée, 35, 2, p. 27-37, 1987.

Thiam, S.S., ACP et F-ACP de séries chronologiques multidimensionnelles stationnaires, Thèse de 3^{ème} cycle, USM de Grenoble, 1984.

Thom, R., Stabilité structurelle et morphogénèse, Interéditions, 1977.

ANNEXE

A - INVARIANCE PAR TRANSLATION DES DECADE-SONDES

Nous avons au cours d'une décade $N = 2^p$ mesures. Soit $d \leq 1+N/2$ le nombre d'harmoniques retenues pour une sonde donnée, $K = \{k_1, \dots, k_d\}$ le d -uple des indices correspondants, avec $0 \leq k_j \leq N/2$. Nous définissons une opération du groupe additif $\mathbb{Z}_N \equiv \mathbb{Z}/N\mathbb{Z}$ sur $E(T)$ (cf. §2) de la façon suivante :

$$\begin{aligned} \mathbb{Z}_N \times E(T) &\rightarrow E(T) \\ (k, f) &\rightarrow S(k)[f] \end{aligned}$$

où k est l'unique représentant de la classe dans $[0, T]$ et $S(\dot{k})[f]$ désigne la fonction périodique sur $[0, T]$ ainsi définie :

$$\begin{aligned} \text{si } t \leq T(1-k/N) \quad S(\dot{k})[f](t) &= f(t+k.T/N) \\ \text{sinon : } S(\dot{k})[f](t) &= f(t-T+k.T/N) \end{aligned}$$

On vérifie facilement que :

$$\begin{aligned} S(\dot{0})[f] &= f \\ \forall \dot{k} \in \mathbb{Z}_N : S(\dot{k})[f](0) &= S(\dot{k})[f](T) \\ \forall \dot{k}_1, \dot{k}_2 \text{ dans } \mathbb{Z}_N : S(\dot{k}_1) \circ S(\dot{k}_2)[f] &= S(\dot{k}_1 + \dot{k}_2)[f]. \end{aligned}$$

\mathbb{Z}_N étant un groupe cyclique d'ordre 2^p , chacun de ses sous-groupes est isomorphe à un groupe $\mathbb{Z}_{k'}$, tel que $k' = 2^{p'}$ ($p' \leq p$), de générateur N°/k' .

Afin de limiter le volume des calculs, nous ferons opérer un sous-groupe strict de \mathbb{Z}_N noté G_1 , et nous travaillerons dans l'espace des décade-sondes approchées :

$E(K) \subseteq E(N) \subset E(T)$ de dimension $e \in [2(d-1), 2d]$; par commodité, nous nous placerons souvent dans \mathbb{R}^{2d} . $E(N)$ désigne le sous-espace pour lequel la résolution est maximale : $K = \{0, \dots, N/2\}$.

La transformation de Fourier se comporte très simplement vis-à-vis des translations :

$$\forall (n, k) \in \mathbb{Z}_N \times \mathbb{Z}, \mathcal{F}(S(n)[f])(k) = e^{-2inkn/N} \mathcal{F}(f)(k).$$

A l'opérateur $S(n)$ correspond donc dans l'espace des fréquences de $E(K)$ une matrice de type $e \times e$ diagonale par blocs (de taille 1 ou 2×2). Vue la nature cyclique de G_1 , l'orbite d'une décade-sonde approchée s'obtiendra itérativement, à partir de la matrice associée au générateur $g_0 (= 2^{p_0})$ de G_1 .

G_1 étant un groupe d'isométries de $E(T)$, il est possible de définir la distance entre deux orbites x et y de la manière suivante (Lütz et Maïti, 1978) :

$$\text{Dor}_1(x, y) \equiv \inf_{g \in G_1} d(x, g(y)).$$

La distance orbitale entre deux décade-sondes sera donc d'autant mieux estimée que $K \subseteq \{0, \dots, N/2\}$ et G_1 seront bien choisis, et que les mesures seront fréquentes.

Déterminons maintenant les points fixes de G_1 dans $E(K)$.

Soit $C(K)$ le d -uple de coefficients de Fourier complexes d'une décade-sonde. C'est un point fixe pour G_1 si et seulement si il existe $\lambda \in [1, 2^{p-p_0}]$ [tel que $g_0^\lambda(C(K)) = C(K)$], ce qui équivaut à :

$$\forall 1 \leq j \leq d, \exp(-2i\pi.k_j.\lambda.g_o / N).C_{k_j} = C_{k_j}.$$

Soit $\mathfrak{L}(k_j, g_o)$ le plus petit entier > 0 tel que $g_o.k_j.\mathfrak{L}(k_j, g_o) / N$ soit entier.

$\forall j, \exists r_j$ impair tel que $k_j = r_j.2^{q_j}$, avec $0 \leq q_j \leq p-1$, ce qui nous donne :

$$g_o.k_j.\mathfrak{L}(k_j, g_o) / N = \mathfrak{L}(k_j, g_o)r_j.2^{p_o+q_j.p}.$$

Si $p_o+q_j \geq p$, alors $\mathfrak{L}(k_j, g_o) = 1$; sinon $\mathfrak{L}(k_j, g_o) = 2^{p-p_o-q_j}$. Il faut donc en définitive que ℓ soit multiple de :

$$\mathfrak{L}(K, g_o) \equiv \text{PPCM} \{ \mathfrak{L}(k_j, g_o) ; 1 \leq j \leq d \}.$$

A moins que toutes les harmoniques d'indice impair n'aient été écartées, nous aurons, sauf si $C_{k_j} = 0 \forall k_j$ impair : $\mathfrak{L}(K, g_o) = \text{Max} \left(1, 2^{\left(p-p_o - \text{Min}_{1 \leq j \leq d} q_j \right)} \right) = \mathfrak{L}(G_1) = 2^{p-p_o}$. Les points fixes de G_1 doivent donc avoir toutes leurs composantes d'indice impair nulles, ce qui correspond à une structure dont la période fondamentale est de la forme $T/2^m$, pour $m > 0$. Ils sont donc inclus dans un sous-espace vectoriel strict de $E(K)$, que l'on supposera de probabilité nulle (il suffit que la loi de distribution des spectres soit absolument continue dans \mathbb{R}^e). Nous noterons $D_1(K)$ le supplémentaire de ce ss-ev dans $E(K)$.

1- $D_1(K)$ est un ouvert de \mathbb{R}^e :

Notons P_k la projection sur la $k^{\text{ième}}$ composante ; l'ensemble D_i des vecteurs de $E(K)$ dont l'une des composantes de Fourier (a_{k_i}, b_{k_i}) est non nulle est le complémentaire du fermé $P_{2i-1}^{-1}(0) \cap P_{2i}^{-1}(0)$. Par conséquent :

$$D_1(K) = \bigcup_{i=1}^d \{ D_i ; k_i \text{ impair} \} \text{ est ouvert dans } \mathbb{R}^e \blacksquare$$

2- $D_1(K)$ est connexe par arcs polygonaux :

Soient C et C' deux points de $D_1(K)$, et \mathcal{J} l'application ainsi définie :

$$D_1(K) \rightarrow \{0, \dots, d\}$$

$$C \rightarrow \mathcal{J}(C) \equiv \{ i \text{ tel que : } k_i \text{ impair et } C_{k_i} \neq 0 \}.$$

Si $\mathcal{J}(C) = \mathcal{J}(C')$, le segment $S(C, C') \equiv \{ \alpha.C + (1-\alpha).C' ; \alpha \in [0,1] \}$ peut éventuellement contenir un unique point hors de $D_1(K)$. Pour cela, il faut que :

$$\exists A > 0 \text{ tel que } \forall i \in \mathcal{J}(C), C_{k_i} = -A.C'_{k_i}.$$

Si $\alpha_o \in [0, 1]$ vérifie : $A = \alpha_o / (1 - \alpha_o)$, le point $\alpha_o.C + (1 - \alpha_o).C'$ n'est pas dans $D_1(K)$, qui n'est donc pas convexe.

Par contre, si $\mathcal{J}(C) \neq \mathcal{J}(C')$, $S(C, C') \subseteq D_1(K)$. Ce domaine est donc connexe par arcs polygonaux : lorsque $\mathcal{J}(C) = \mathcal{J}(C')$, en choisissant C'' tel que $\mathcal{J}(C) \neq \mathcal{J}(C'')$, on pourra joindre C et C' par le chemin $S(C, C'') \vee S(C'', C') \subset D_1(K)$ ■

Remarques :

- C'' peut être pris arbitrairement proche du segment $S(C, C')$, de sorte que la distance euclidienne entre C et C' est la distance géodésique dans $D_1(K)$ entre ces points.

- On peut construire de manière semblable l'ouvert connexe par arcs polygonaux $D'_1(K)$ des points de $E(K)$ possédant une harmonique d'indice *pair* non nulle.

3- $D_1(K) / G_1$ est une variété :

Il suffit de montrer que G_1 est proprement discontinu sans point fixe sur $D_1(K)$; alors, $D_1(K) / G_1$ pourra être muni d'une structure de variété séparée connexe ($D_1(K)$ l'est) de dimension e , telle que la surjection canonique soit un revêtement de $D_1(K) / G_1$ à $\mathfrak{d}(G_1)$ feuilletés (Berger et Gostiaux, 1972).

Il faut donc que les propriétés ci-dessous soient vérifiées :

(P1) : $\forall x \in D_1(K), \exists \mathcal{U}_{(x)}$ ouvert de $D_1(K)$ tel que :

$$\forall g \in G_1 - \{\text{Id}\}, g(\mathcal{U}_{(x)}) \cap \mathcal{U}_{(x)} = \emptyset.$$

(P2) : $\forall x$ et y dans $D_1(K)$ tels que $y \notin G_1(x)$, il existe des voisinages ouverts $\mathcal{U}_{(x)}$ et $\mathcal{V}_{(y)}$ tels que :

$$\forall g \in G_1, \mathcal{U}_{(x)} \cap g(\mathcal{V}_{(y)}) = \emptyset.$$

preuve :

(P1)-L'orbite de x compte exactement $\mathfrak{d}(G_1)$ points. Soit $0 < \delta \equiv \frac{1}{4} \cdot \text{Min} \{ \|x - g_0^n(x)\| ; 1 \leq n \leq \mathfrak{d}(G_1) \}$; G_1 étant un groupe d'isométries, nous avons : $g_0^n(B(x, \delta)) = B(g_0^n(x), \delta)$. Il suffit donc de prendre : $\mathcal{U}_{(x)} = B(x, \delta) \cap D_1(K)$ ■

(P2) - Soit $0 < \varepsilon \equiv \frac{1}{4} \cdot \text{Min} \{ \|x - g_0^n(y)\| ; 1 \leq n \leq \vartheta(G_1) \}$. Nous aurons alors : $B(x, \varepsilon) \cap g_0^n(B(y, \varepsilon)) = \emptyset$; il suffit donc de prendre $\mathcal{U}_{(x)} = B(x, \varepsilon) \cap D_1(K)$ et $\mathcal{V}_{(y)} = B(y, \varepsilon) \cap D_1(K)$ ■

4- Dor_1 est une distance géodésique sur $D_1(K) / G_1$:

Munissons cette variété d'une structure riemannienne de métrique $M_{(\gamma)}$. Soit un chemin de classe $C^1 : [0, 1] \xrightarrow{\gamma} D_1(K) / G_1$, tel que $\gamma(0) = a$ et $\gamma(1) = b$. Il existe $\vartheta(G_1)$ relèvements de γ sur $D_1(K)$, tous de longueur (pour la métrique transportée) égale à celle de γ (pour $M_{(\gamma)}$).

En effet, notons P la surjection canonique, P^*M la métrique transportée et Γ l'un des relèvements de γ ($\gamma = P \circ \Gamma$). Il vient :

$$L(\gamma) \equiv \int_0^1 \|\dot{\Gamma}(t)\|_{Q(t)} dt, \text{ avec } Q \equiv M_{\circ} \gamma$$

$$P^*(x) = \text{Id} \quad \forall x \in D_1(K) \Rightarrow P^*M_{(x)} = M_{P(x)}$$

$$d'ou : \int_0^1 P^*M_{\circ} \Gamma(t) (\Gamma'(t), \Gamma'(t)) dt = \int_0^1 \|\Gamma'(t)\|_{Q(t)} dt = L(\gamma).$$

D'après les remarques du §A.2, il est donc clair que $\text{Dor}_1(a, b) = \text{Inf}_{\gamma} L(\gamma)$ est la distance géodésique entre a et b .

B - INVARIANCE PAR TRANSLATION ET SYMETRIE PARTIELLE

Notons $\text{Fix}(K) \subseteq \{1, \dots, d\}$ l'ensemble des indices des harmoniques laissées fixes par la symétrie partielle choisie, et (U, \otimes) le groupe à deux éléments lié à celle-ci, \otimes désignant l'opération : multiplication composante par composante. Si G_2 désigne le groupe produit $U \times G_1$, montrons qu'il opère dans $E(K)$ de la manière suivante :

$$G_2 \times E(K) \rightarrow E(K)$$

$$((u, \dot{k}), f) \rightarrow (u, \dot{k}) f \equiv \mathcal{F}^{-1} (u \otimes \mathcal{F}(S(\dot{k}) [f])).$$

Remarquons que $\mathcal{F}(S(\dot{k}) [f])$ peut s'écrire sous la forme $\mathcal{F}(S(\dot{k})) \otimes \mathcal{F}(f)$, une translation revenant à une convolution par une masse de Dirac. On en déduit que :

$$((u_1, \dot{k}_1) (u_2, \dot{k}_2)) f \equiv \mathcal{F}^{-1} (u_1 \otimes u_2 \otimes \mathcal{F}(S(\dot{k}_1 + \dot{k}_2) [f]))$$

$$= \mathcal{F}^{-1}(u_1 \circ \mathcal{F}(S(k_1)[(u_2, k_2) f])) = (u_1, k_1) ((u_2, k_2) f).$$

Etant donné que de plus $(1, 0) f = f \quad \forall f \in E(K)$, il est donc établi que G_2 opère sur $E(K)$.

Les points fixes de G_1 sont évidemment fixes pour G_2 ; si u_0 est le générateur de U , les autres points fixes de G_2 doivent vérifier : $u_0 \circ g_0^n(C(K)) = C(K)$ avec $1 \leq n \leq \mathcal{Q}(G_1) - 1$. Nous nous restreindrons à l'ensemble presque certain $D_1(K)$. Soit \mathcal{P} l'application ainsi définie :

$$\begin{aligned} D_1(K) &\rightarrow \{0, \dots, d\} \\ C &\rightarrow \mathcal{P}(C) \equiv \{i \text{ tel que : } k_i \text{ pair et } C_{k_i} \neq 0\}. \end{aligned}$$

lemme : si $C \in D_1(K)$ est fixe pour G_2 , C vérifie :

$$\begin{aligned} \emptyset \neq \mathcal{J}(C) &\subseteq K\text{-Fix}(K) \\ \mathcal{P}(C) &\subseteq \text{Fix}(K). \end{aligned}$$

preuve : la première proposition découle directement de l'étude des points fixes de G_1 . Nous devons donc avoir : $\forall j \in \mathcal{J}(C), \exp(-2i\pi.k_j.g_0.n / N) = -1$, ce qui implique $n = 2^{p-p_0-1} = \mathcal{Q}(G_1) / 2$. On vérifie alors que : $\forall m \in \mathcal{P}(C), \exp(-2i\pi.k_m.g_0.n / N) = 1$ ■

Soit $0 < nf \equiv \text{Card}(\text{Fix}(K))$, et $nv \equiv d-nf$. Nous pouvons construire $D_1(\text{Fix}(K))$ (resp. $D_1(K - \text{Fix}(K))$), ouvert connexe par arcs polygonaux de \mathbb{R}^{2nf} (resp. \mathbb{R}^{2nv}). Sauf si la symétrie partielle a été bizarrement choisie de façon à laisser fixes les seules harmoniques d'indice pair, ces ensembles ne sont pas vides ; nous supposerons que tel est le cas.

Alors, $D_1(\text{Fix}(K)) \times \mathbb{R}^{2nv} \cup \mathbb{R}^{2nf} \times D_1(K - \text{Fix}(K))$ est un ouvert connexe de \mathbb{R}^{2d} ; nous baptiserons $D_2(K ; \text{Fix}(K))$ l'intersection de cet ensemble avec $D_1(K)$. En vertu de nos hypothèses, c'est un ouvert connexe non vide de \mathbb{R}^{2d} , et l'on montrera facilement que deux points quelconques C et C' de ce domaine peuvent être joints par un chemin du type :

$$S(C, C'') \vee S(C'', C) \subset D_2(K ; \text{Fix}(K)) \quad (\text{éventuellement } C'' = C').$$

Les propriétés de base étant établies, on montrera en suivant la même démarche qu'au §A que $D_2(K ; \text{Fix}(K)) / G_2$ peut être muni d'une structure de variété connexe et séparée de dimension e , telle que la surjection canonique soit un revêtement à $2.\mathcal{Q}(G_1)$ feuillets de $D_2(K ; \text{Fix}(K)) / G_2$. En munissant cette variété de la structure riemannienne correspondant à la métrique identité, on obtiendra une deuxième distance orbitale □.