

THESE présentée
pour l'obtention
du
DIPLOME de DOCTEUR de 3^e CYCLE
à
L'UNIVERSITE DE PARIS VI

spécialité : STATISTIQUE MATHEMATIQUE
mention : ANALYSE DES DONNEES



par M .Jean-Pierre ORFEUIL.....

Sujet de la thèse : Etude par Analyse des Données de la structure
de la haute atmosphère terrestre.

soutenue le : 12 Janvier 1976..... devant la Commission composée de :

M .DUGUE.....	Président
M .BENZECRI.....	Rapporteur
M .BARLIER.....	Examineur
M .MATHERON.....	Invité

REMERCIEMENTS

Il m'est agréable de remercier ceux qui m'ont aidé dans ce travail :

Monsieur J.P. BENZECRI, qui m'a enseigné plus qu'une méthode, une philosophie

Monsieur F. BARLIER, qui a su me faire partager sa passion pour l'étude de la haute atmosphère

Monsieur G. MATHERON, qui dirige mes travaux depuis cinq ans et qui m'a fait aimer l'étude rigoureuse des phénomènes naturels par des méthodes probabilistes

Monsieur D. DUGUE, qui a bien voulu présider mon jury

Mon camarade J.P. BORDET, qui m'a proposé ce travail et m'a mis en contact avec F. BARLIER

Mesdames BOSCH et PIPAULT qui ont réalisé la frappe de ce travail.

P L A N

I - OBJET DE L'ETUDE	page 2
II - LE FICHIER UTILISE	page 6
III - PLAN DE TRAVAIL	page 12
IV - ETUDE DESCRIPTIVE DES DEFORMATIONS	page 13
V - MODELISATION ET CARTOGRAPHIE	page 56
VI - ANALYSE GLOBALE DES EFFETS METHODE DU TRI CROISE	page 94
VII - ANALYSE MULTIDISJONCTIVE	page 119
VIII - BIBLIOGRAPHIE	page 137

I - OBJET DE L'ETUDE

Un certain nombre de géophysiciens ont pu étudier l'atmosphère terrestre d'une manière suffisamment précise pour pouvoir proposer des modèles permettant de calculer une densité atmosphérique théorique, dans des conditions données. Au premier rang de ceux-ci, l'américain JACCHIA a calculé un modèle de Haute Atmosphère qui sert de référence aux travaux de nombreux laboratoires, et en particulier à ceux du laboratoire de dynamique Spatiale. Il fait intervenir deux types de paramètre : Position et activité solaire.

POSITION : Latitude, heure locale, altitude, date.

ACTIVITE SOLAIRE : Flux moyen, variation de flux, activité géomagnétique.

L'activité solaire est soumise à deux types principaux de variations cyclique, appréhendés à partir de l'évolution du flux radio-électrique sur la longueur d'onde 10,7 cm : la première a un cycle d'environ 11 ans : on en rend compte par une grande moyenne mobile du flux journalier appelée flux moyen (noté FB). La seconde a un cycle d'environ 27 jours. On en rend compte par la variation de flux (notée DF), différence entre le flux journalier et le flux moyen.

En fin l'activité géomagnétique, liée à l'émission de particules par le soleil provoque un chauffage général de l'atmosphère et une perturbation du champ magnétique terrestre. On mesure cette activité par un paramètre "activité géomagnétique" qui rend compte de l'énergie reçue par l'atmosphère. Ces trois paramètres ne dépendent que de la date, et non de la position géographique du point.

Le modèle JACCHIA, comme la plupart des modèles de Haute Atmosphère, utilise les hypothèses simplificatrices suivantes :

- . Il existe une altitude de base à laquelle la densité est invariante.
On connaît, à cette altitude, les concentrations de différents constituants : Azote, oxygène moléculaire, oxygène atomique, hélium, hydrogène.
- . A partir de cette altitude, les différents constituants sont distribués selon la loi de l'équilibre de diffusion.
- . Tout se passe comme si chacun des constituants était seul présent : donc pas d'interactions.
- . Le profil des températures le long de la verticale est une exponentielle.
Le modèle procède en deux temps :
 - . Il calcule d'abord la température exosphérique, correspondant au point à l'infini de l'exponentielle. En fait cette température est largement atteinte entre 1500 et 2000 km, zone quasiment isotherme. Ce calcul fait intervenir l'ensemble des paramètres de position et d'activité solaire. Seule l'altitude, bien entendu, n'intervient pas.
 - . Moyennant la température exosphérique et les concentrations initiales, on peut calculer la densité atmosphérique en un point quelconque de la verticale.

On trouvera les équations détaillées des modèles JACCHIA 65 et JACCHIA 71 dans la thèse de J.L. FALIN ([1]). Nous ne les donnerons pas, ne souhaitant pas entrer dans la technique du modèle, mais seulement broser les grands traits d'un modèle très complexe.

La qualité d'un modèle doit bien entendu être confrontée à la réalité, et c'est là tout l'objet de l'étude qui suit. On dispose en effet de mesures très perfectionnées de la densité en Haute Atmosphère, bien que relativement indirectes. François BARLIER a eu l'idée de se servir du mouvement des satellites artificiels pour calculer les densités au périgée de leur trajectoire : En l'absence d'atmosphère, un satellite aurait une

trajectoire T. Evoluant dans une atmosphère raréfiée mais non négligeable, ce satellite subit diverses forces de frottement, qui le conduisent à décrire une trajectoire T_1 . On peut estimer la densité atmosphérique au périgée du satellite (point le plus bas de sa trajectoire, où les effets sont les plus sensibles puisque la densité atmosphérique chute considérablement avec l'altitude), à partir de la comparaison des orbites théoriques et observées; on recueille, pour chaque observation, une densité observée et l'ensemble des paramètres de position et d'activité solaire. On peut alors définir un coefficient correcteur f, quotient de la densité observée par la densité théorique.

Alors, plutôt que d'étudier directement les densités observées, on s'intéresse à la structure des déformations du modèle, c'est à dire à la structure du coefficient f. Pour donner une idée de la simplification qu'apporte un tel modèle, et de la qualité du modèle JACCHIA 71, indiquons des ordres de grandeur: Dans notre champ d'étude, les densités peuvent varier de 10^{-12} à 10^{-18} , soit un rapport de 1 à 1 million. Le coefficient f lui variera de 1/3 à 3, la grande majorité des points donnant un f compris entre 0,7 et 1,2. Il faut bien entendu préciser que le principal facteur de variabilité est l'altitude : Lorsqu'on fait varier, à l'altitude donnée, les différents paramètres, on peut avoir des variations de densité de 1 à 10. Par contre lorsqu'on passe de 300 à 900 km, tous les autres paramètres étant fixés, on peut obtenir des variations de 1 à 10 000, voire même plus grandes encore. C'est à dire que l'altitude joue un rôle spécifique, et on rencontrera cette particularité tout au long de l'étude.

Notre travail consistera donc essentiellement à étudier les déformations du modèle JACCHIA 71 : D'une part décrire les effets du modèle (sur ou sous estimation systématique dans certaines conditions), d'autre part proposer un modèle correctif, permettant de prévoir, dans des conditions

données, la valeur du coefficient f , et par suite la valeur de la densité.

Tout au long de cette étude, nous utiliserons beaucoup une méthode d'Analyse des Données, appelée l'Analyse des Correspondances. Cette méthode est due au professeur Jean-Paul BENZECRI. Laissant de côté la construction mathématique, nous essaierons dans le courant de l'étude de familiariser par la pratique le lecteur avec cette méthode.

II LE FICHIER UTILISE

Le fichier utilisé a été confié au laboratoire de Dynamique Spatiale par JACCHIA. Il comporte environ 50 000 observations portant sur 16 satellites. Nous en avons extrait les observations relatives à la période 1961 - 1970. Il comporte les données suivantes :

- . JOUR JULIEN de l'observation.
- . Ascension droite et déclinaison du satellite.
- . Ascension droite et déclinaison du soleil.
- . Densité effective et densité périgée. Ces densités correspondent, pour une même observation, à deux altitudes de référence. JACCHIA conseille d'utiliser les densités effectives, ce que nous ferons.
- . Altitude effective et altitude au périgée.

Le tableau suivant donne, pour chaque satellite, son nom, sa période d'activité, et le nombre d'observations. Nous n'avons retenu que les observations correspondant à la période 1961 - 1970, soit de 37 300 à 40 951 en jour JULIEN . En fait, cette période a été découpée en 4 tranches inégales :

- 1 : 1961 - 1962 : Forte activité. Orages magnétiques fréquents
- 2 : 1963 - 1966 : Faible activité
- 3 : 1967 : Activité intermédiaire
- 4 : 1968 - 1970 : Forte activité

NOM	DEBUT	FIN	NOMBRE D'OBSERVATIONS	OBSERVATIONS RETENUES
1958 ALPHA 1	35242	40666	5458	4534
1958 BETA 2	36340	37730	409	92
1959 ALPHA 1	36622	36647	2179	1520
1959 ETA	36834	37666	461	188
1960 XI 1	37246	41567	4144	3091
1061 DELTA 1	37354	38494	3167	3040
1962 BETA TAU 2	38013	39668	1975	1975
1963 IOTA 1	38128	38311	272	272
1963 53 A/002	38390	41553	4490	3578
1963 53 A/017	38386	41143	8949	8327
1964 4 A	38425	40377	3372	3372
1964 76 A/018	38726	40144	3299	3294
1964 76 A/019	38725	40077	4126	4119
1966 44 A	39279	41543	3760	2529
1966 70 A	40222	41594	1851	1033
1968 66 A	40084	41562	2981	2731

L'essentiel de notre travail porte sur les zones 2 et 4, périodes de minimum et de maximum du cycle solaire de 11 ans.

Les tableaux suivants sont des tris sur le fichier, donnant une idée de la répartition des points.

Année	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
Nb points	1816	1572	2538	3843	4717	5466	7089	6326	5092	4203

Mois	Janv.	Fev.	Mars	Avril	Mai	Juin	Juil.	Aout	Sept.	Oct.	Nov.	Dec.
Nb points	3633	3419	3635	3742	4153	3504	3694	3310	3165	3586	3424	3397

On verra dans la suite qu'on a été conduit à distinguer, pour des raisons physiques, 4 tranches d'altitude : De 250 à 320, de 320 à 530, de 530 à 750, de 750 à 1250. Les tableaux donnent, pour chacune des périodes étudiées et pour chaque altitude, la distribution des points en latitude en heure locale. Les heures vont de quatre en quatre. Les latitudes sont :

1ere tranche : de 45 à 90° .

2eme tranche : de 15 à 45°

3eme tranche : de -15 à +15°

4eme tranche : de -45 à -15°

5eme tranche : de -90 à -45°

1963 - 1966

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 --+90	110	99	105	95	108	45
+15 --+45	55	108	142	113	61	35
-15 --+15	29	70	76	119	60	69
-15 - -45	57	36	76	72	96	77
-45 --90	66	52	78	75	94	79

de 250 à 320 km.

1963 - 1966

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	24	8	16	31	18	8
+15 - +45	129	129	152	190	276	194
-15 - +15	117	135	135	143	256	235
-45 - -15	131	95	145	129	338	298
-90 - -45	7	27	15	15	9	27

de 320 à 530 km

1963 - 1966

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	205	242	190	163	319	125
+15 - +45	71	187	453	587	402	58
-15 - +15	54	142	194	1085	240	59
-45 - -15	73	118	584	384	377	78
-90 - -45	156	309	230	125	226	135

de 530 à 750 km.

1963 - 1966

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	138	217	120	34	240	169
+15 - +45	4	52	169	108	52	0
-15 - +15	0	64	90	111	34	0
-45 - -15	4	41	168	105	133	0
-90 - -45	186	269	154	22	280	262

de 750 à 1250 km.

1968 - 1970

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	77	85	26	36	33	52
+15 - +45	54	76	87	111	93	59
-15 - +15	19	43	27	22	51	31
-45 - -15	4	47	41	29	17	5
-90 - -45	62	50	58	71	75	54

de 250 à 320 km.

1968 - 1970

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	173	188	142	136	156	147
+15 - +45	147	163	222	314	168	136
-15 - +15	84	114	304	242	132	105
-45 - -15	73	132	276	231	138	160
-90 - -45	53	53	53	39	49	77

de 320 à 530 km.

1968 - 1970

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	85	94	74	52	63	26
+15 - +45	27	279	53	58	62	46
-15 - +15	63	146	132	110	46	48
-45 - -15	74	36	217	115	74	50
-90 - -45	34	68	39	34	64	56

de 530 à 750 km.

1968 - 1970

	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24
+45 - +90	258	335	426	348	290	335
+15 - +45	129	21	148	341	480	173
-15 - +15	72	94	148	278	541	214
-45 - -15	99	85	88	317	431	225
-90 - -45	239	232	224	330	193	300

de 750 à 1250 km.

La distribution des points est en gros satisfaisante. Il faut remarquer cependant le faible nombre de points aux pôles aux altitudes basses en faible activité, et d'une manière générale l'inégale répartition des points en altitude, comme le montre le tableau suivant, où sont confondues les 4 périodes d'activité.

ALTITUDE	de 250 à 320	de 320 à 530	de 530 à 750	de 750 à 1250
NB DE POINTS	4575	10035	13003	15027

Donnons maintenant quelques tableaux concernant la répartition de la variation du flux, du flux moyen et de l'activité géomagnétique pendant la période concernée.

VAR. DE FLUX : DF	< 30	-30,-20	-20,-10	-10,0	0,10	10,20	20,30	> 30
NB DE POINTS	1679	2150	5191	9949	13320	3733	2727	3513

FLUX MOYEN : FB	60,72	72,81	81,110	110,130	130,145	145,153	153
NB DE POINTS	1221	8558	7036	4168	3685	10489	4459

ACT. GEOMAGNET. : AP	0,7	7,15	15,30	30,45	45,60	60,90	> 90
NB DE POINTS	18989	12686	7559	1583	784	564	397

III PLAN DU TRAVAIL

Au départ, nous avons cherché à nous familiariser avec les problèmes posée par la Haute Atmosphère, et avec le fichier JACCHIA. Les quelques statistiques données précédemment relèvent de ce dernier souci. Nous avons ensuite cherché à tracer des "courbes moyennes" de f en fonction des différents paramètres descriptifs. Il s'est trouvé que certaines courbes (effet en latitude en particulier) ne ressemblaient pas du tout à celles qu'avaient obtenu précédemment de nombreux auteurs. Pour sortir de l'impasse, il a fallu serier les problèmes, se placer dans des conditions moins générales; c'est cette difficulté (sur laquelle on reviendra en détail au chapitre suivant) qui est à l'origine de la méthode que nous avons adoptée pour analyser les déformations du modèle JACCHIA 71 : L'Analyse Factorielle des Correspondances d'un ensemble de courbes. A la suite de cette étude, qui nous a permis d'acquérir un certain nombre d'idées - force concernant les déformations du modèle, nous avons cherché à modéliser les déformations, c'est à dire à proposer un algorithme qui associe à chaque "condition expérimentale" une valeur du coefficient f . Les difficultés rencontrées, la solution finalement adoptée, sont décrites dans la partie suivante, intitulée modélisation et cartographie. Enfin, nous désirons avoir un point de comparaison avec la méthode qu'avait adopté, pour un traitement de données semblables, Jean-Pierre BORDET (2). C'est pourquoi nous avons repris un traitement basé sur les mêmes principes, dont les résultats sont développés dans les parties VI et VII du présent rapport.

IV ETUDE DESCRIPTIVE DES DEFORMATIONS DU MODELE

JACCHIA 1971

I PRINCIPE DE LA METHODE

II APPLICATION : LES DEFORMATIONS ETUDIEES

III LA DEFORMATION EN LATITUDE

IV LA DEFORMATION EN HEURE LOCALE

V LA DEFORMATION SEMI-ANNUELLE

VI PRINCIPALES CONCLUSIONS

I PRINCIPE DE LA METHODE

Le lecteur rencontrera, dans les pages suivantes, un ensemble de courbes décrivant l'évolution du coefficient correcteur f en fonction d'un paramètre, les autres étant astreints à ne pas sortir d'un certain domaine de variations. Il convient d'abord de justifier le choix de cet objet d'étude, de préciser la richesse (mais aussi l'extrême pauvreté) de l'information ainsi recueillie et enfin de proposer une méthode automatique d'étude, nous permettant de préciser en quelques lignes la structure des déformations du modèle.

A/ POURQUOI ETUDIER CES COURBES ?

Le coefficient correcteur f dépend d'un certain nombre de variables, qui, bien que conceptuellement indépendantes, n'en ont pas moins des influences conjuguées sur l'erreur du modèle. Il n'est bien évidemment pas possible, dans le domaine de la Haute Atmosphère, de bâtir de savants plans d'expérience pour étudier l'influence simultanée de plusieurs facteurs, comme le ferait un statisticien chargé d'un contrôle de production : Toute nouvelle banque de données coûte très cher, dans la mesure où elle représente le lancement de plusieurs satellites. Il faut d'ailleurs remarquer que la plupart des données recueillies sont issues de lancements dont l'objectif premier n'est pas l'étude de la Haute Atmosphère. Enfin, il est clair que certains paramètres, comme les variables d'activité solaire, ne sont nullement maîtrisables par l'homme. Le statisticien doit exploiter au mieux un fichier que, dans une certaine mesure, il ne maîtrise pas. Il convient de préciser deux types de difficultés auxquelles on peut se heurter.

1) Le Problème des Couplages

Le problème a été largement soulevé par J.P. BORDET dans sa

thèse (2). Il a montré que certaines anomalies apparaissant dans son analyse étaient en fait dûes à une mauvaise repartition des points dans le fichier : il avait par exemple beaucoup trop de points en Septembre et Novembre appartenant à une tranche donnée de flux moyen. Il en avait par contre trop peu dans une autre tranche. Il est important de pouvoir repérer de tels couplages d'une part, d'éliminer leur influence d'autre part.

2) Le Problème des Moyennes

Apparenté au précédent, il se double d'un problème méthodologique. Supposons que le coefficient correctif f soit une fonction parfaitement déterminée des seuls paramètres présents dans le fichier (paramètres de position et d'activité solaire), ce qui est une hypothèse relativement peu réaliste, bien que quasiment obligatoire pour réaliser un traitement. Quel sens peut-on donner à l'expression "courbe de f en fonction de la latitude" ? D'une façon générale, à ce concept est attaché celui de moyenne, donc de mesure. Si f est par exemple fonction de deux paramètres x et y , et de la latitude, on définira :

$$f_{\mu}(l) = \frac{1}{\int \mu(dx, dy)} \int f(x, y, l) \mu(dx, dy)$$

Il est évident que la forme de la courbe dépend de la mesure utilisée. Ce qui peut n'apparaître que comme un scrupule de mathématicien est en fait justifié par l'expérience que nous avons faite au début de notre travail : pour avoir une idée de l'ampleur des déformations dûes à chacune des variables, nous avons tracé un certain nombre de courbes donnant l'évolution de f en fonction des principaux paramètres : pour reprendre l'exemple de la latitude, nous avons découpé l'intervalle de variation en 18 tranches de 10° , fait la moyenne des f dans chacune de ces tranches, et ainsi construit une courbe $f=f(\text{latitude})$. Alors que dans les travaux précédents, on constate toujours que cette courbe décroît de

l'équateur à chacun des pôles, nous constatons au contraire que la nôtre allait en croissant, de chacun des pôles à l'équateur. La raison en était en fait fort simple : Dans les fichiers traités précédemment, les satellites avaient une altitude qui dépassait rarement 500 km. Dans notre fichier, les altitudes varient de 250 à 1200 km. Or une étude plus précise montre que les courbes $f = f(\text{latitude})$ sont complètement différentes, selon que l'on se place au-dessus ou au-dessous de 500 km : Le comportement dynamique des molécules "lourdes" (azote, oxygène), qui composent l'atmosphère en dessous de 500 km est très différent de celui des atomes légers qui évoluent au dessus de 500 km (hélium, hydrogène). Comme nous avions plus d'observations au dessus de 500 km qu'en dessous, l'inversion était expliquée. Notons qu'avec une répartition à peu près égale des points, les deux effets auraient pu se compenser, donner une courbe moyenne plate, et nous aurions pu conclure à l'absence d'effets, ce qui aurait été évidemment erroné.

Cette expérience nous a incité à décrire l'évolution de f en fonction d'une variable, non pas à l'aide d'une seule courbe, mais à l'aide d'une famille de courbes correspondant à différentes valeurs des autres paramètres. C'est ainsi que nous avons par exemple tracé 32 courbes d'évolution de f en fonction de la latitude, correspondant à 32 situations différentes : L'activité solaire est faible ou forte, l'altitude est très basse, basse, haute ou très haute, la saison est l'hiver, le printemps, l'été ou l'automne. Cette façon de procéder a l'avantage d'être relativement précise, et d'être indifférente aux couplages des variables prises en compte. Elle nécessite par contre de très gros fichiers : Partant d'un fichier d'environ 40.000 observations, on dispose d'environ 1200 observations pour chacune des 32 situations décrites plus haut. Si on découpe les latitudes en 10 tranches, il ne reste qu'environ 120 observations par tranche en moyenne;

en fait on constate que certaines tranches sont beaucoup plus riches, mais que d'autres sont vides : C'est la raison de l'absence de certaines courbes sur les graphiques.

B/ RICHESSSE DE L'INFORMATION CONSERVEE

Il est clair que lorsqu'on substitue à un fichier de 40 000 observations portant sur 8 variables un ensemble d'une centaine de courbes à 5 ou 6 valeurs chacune, on perd une très grande quantité d'information. C'est, pensons nous, le prix à payer pour une meilleure compréhension des phénomènes : Tout modèle appauvrit considérablement la réalité dont il prétend rendre compte. Il importe cependant qu'il traduise les traits essentiels, et ce sont ces traits que nous chercherons à mettre en évidence par la comparaison simultanée des différentes courbes d'une famille. Notons que nous devons disposer d'une méthode efficace pour comparer entre elles ces diverses courbes : Malgré la perte d'information considérable dont nous avons parlé au départ, il faut remarquer que l'information est encore trop riche pour apparaître clairement à l'oeil : Le sentiment de monotonie qui se dégage de l'observation des 32 courbes citées plus haut pourrait nous faire penser à une absence de structure véritable : Une méthode automatique de comparaison, l'Analyse Factorielle des Correspondances ([3]), nous prouvera le contraire.

C/ METHODE DE COMPARAISON DES COURBES

Nous souhaitons disposer d'une méthode nous permettant de définir des relations de proximité entre courbes, et nous autorisant à les visualiser. Remarquons que d'un point de vue strictement mathématique,

on se donne le même objet lorsqu'on a 32 courbes à 6 valeurs ou 32 points dans un espace vectoriel de dimension 6. Mais 6 dimensions, c'est encore beaucoup pour l'oeil humain. On pourrait penser alors observer les plans obtenus en négligeant systématiquement 4 valeurs sur 6 : on y passerait beaucoup de temps, et les conclusions ne pourraient être que malaisées. En fait, s'il existe bien une structure, comme on le pense au départ, ces points ne sont pas répartis n'importe comment dans notre espace à 6 dimensions : Tel groupe de points se répartit très certainement selon telle ligne de force, qui l'oppose à tel autre groupe de points : D'où l'idée d'examiner notre ensemble de points sur des plans non pas choisis en fonction des axes initiaux, mais choisis à partir de la structure même de l'ensemble de points : On pense alors tout naturellement aux mécaniciens, qui cherchent toujours les directions principales d'allongement - baptisées axes principaux d'inertie - des solides dont ils étudient les mouvements. Pour transposer ce formalisme, il faut munir notre espace d'une métrique euclidienne, de manière que la notion de produit scalaire y ait un sens. Il convient que cette métrique soit adaptée à notre problème. Une courbe quelconque, une fois discrétisée, est le résultat d'un découpage en classes qui comporte nécessairement une part d'arbitraire : Il faut que la métrique choisie conduise à des résultats stables, c'est à dire aussi peu tributaires que possible du découpage. La métrique X^2 , utilisée par J.P. BENZECRI pour l'Analyse des Correspondances, possède cette intéressante propriété, qui s'énonce ainsi : Si l'on regroupe deux classes dont les profils sont identiques, alors on ne change rien à l'analyse.

Enfin, pour que l'analogie avec le problème mécanique soit totale, il convient d'attribuer un poids à chacune des courbes. Le formalisme de l'Analyse des Correspondances accorde à chaque ligne (ou à chaque colonne) d'un tableau un poids égal à la somme des éléments de la ligne (ou de la

côlonne) : Ainsi, en ce qui nous concerne, toutes les courbes auront sensiblement le même poids, ce qui paraît parfaitement raisonnable.

Les éléments du décor sont maintenant posés : Nous avons 32 points pesants répartis dans un espace euclidien de dimension 6 . Il nous est possible de rechercher les directions principales d'inertie de cet ensemble de points, et de visualiser les projections du nuage sur ses axes principaux : On visualise ainsi le nuage en le déformant le moins possible. Le lecteur soucieux d'approfondir la théorie pourra se reporter à 4 pour une présentation mathématique simple, ou à 3 pour une présentation complète illustrée de nombreux exemples d'application. Les autres, qui s'intéressent surtout à la pratique de l'Analyse, doivent essentiellement retenir les règles suivantes, pour l'étude d'un dessin quelconque :

- . Une courbe quelconque est représentée par un point.
- . Deux courbes dont les profils sont très ressemblants sont proches.
- . On peut représenter simultanément les courbes et les vecteurs de base; reprenons l'exemple des 32 courbes $f=f(\text{latitude})$. Elles sont définies en 5 points : Pôle Sud, Zone Tempérée Sud, Equateur, Zone Tempérée Nord, Pôle Nord. Ces 5 points constituent les 5 vecteurs de base, et ils sont visualisés sur les graphiques. Une courbe quelconque sera très proche d'un point si elle présente en ce point un maximum très accusé. Le point représentatif d'une courbe assez plate sera au contraire sensiblement au barycentre du système des points.

II LES DEFORMATIONS ETUDIEES

Les travaux récents du laboratoire de Dynamique Spatiale ont montré que le modèle JACCHIA 71, bon en moyenne, déformait la réalité non pas d'une façon purement aléatoire, mais d'une façon structurée : On a par exemple constaté que le modèle surestime au pôle, et sous estime à l'équateur, surestime en été, sous estime en hiver.

Nous nous sommes essentiellement intéressés à l'étude de trois types de déformation.

- . Evolution de f en fonction de la date.
- . Evolution de f en fonction de la latitude.
- . Evolution de f en fonction de l'heure locale.

Chacune de ces évolutions est décrite par un certain nombre de courbes, tracées pour différentes valeurs des autres paramètres.

PARAMETRE : ACTIVITE SOLAIRE

Nous avons distingué deux périodes d'activité solaire. La première, couvrant la période de 1963 à 1966, est une époque de faible activité : Le flux solaire moyen ne dépasse jamais 110 unités. La seconde va de 1968 à 1970 : Le flux solaire moyen ne tombe jamais en dessous de 130 unités. Nous avons volontairement laissé de côté les années 1961 et 1962, pendant lesquelles l'activité géomagnétique a parfois été très forte et l'année 1967, période de transition entre faible et forte activité, au cours de laquelle le flux solaire moyen prend des valeurs comprises entre 110 et 130 unités.

Nous distinguerons de façon systématique ces deux périodes d'activité, caractérisées par des flux moyens différents. Par contre l'activité géomagnétique prend des valeurs similaires dans les deux périodes. Elle y est en général faible (inférieur à 20 unités).

PARAMETRE : ALTITUDE

Nous nous sommes laissés guider par la composition chimique de la haute atmosphère : Plus on monte, plus les gaz lourds se raréfient. Nous avons défini quatre tranches, déterminées par leurs constituants majoritaires:

- 1 : de 220 à 320 km : Azote et Oxygène atomique
- 2 : de 320 à 530 km : Oxygène atomique
- 3 : de 530 à 750 km : Helium
- 4 : de 750 à 1250 km : Hydrogène atomique

Il faut dès maintenant souligner l'importance de ce paramètre : Le comportement dynamique des gaz est en relation plus ou moins directe avec la masse de ceux-ci : On peut presque dire que l'on change d'objet d'étude lorsqu'on passe 200 à 1200 km, puisqu'on ne s'intéresse plus à la dynamique du même gaz, ce qui n'est évidemment pas vrai quand on passe de minuit à midi, ou de janvier à juillet.

PARAMETRE : DATE

Nous distinguerons quatre périodes dans l'année, que nous appellerons conventionnellement saisons :

- 1 : Hiver : novembre, décembre, janvier, février.
- 2 : Printemps : mars, avril.
- 3 : Ete : mai, juin, juillet, août.
- 4 : Automne : septembre, octobre.

Le découpage est adapté à la variation de la densité : Celle ci passe par des extrémums plats aux solstices, et subit au contraire une évolution très rapide aux équinoxes.

PARAMETRE : LATITUDE

La latitude n'intervient comme paramètre que pour les courbes donnant l'évolution de f en fonction de la date. Nous avons distingué

cinq tranches de latitude, dont les bornes ont été calculées de manière à avoir suffisamment de points dans chaque tranche : Les observations sont très rares dans les zones polaires, ce qui explique que le "Pôle Nord" commence ..à Bordeaux.

- 5 : Pôle Nord : de +45 à +90 degrés
- 4 : Zone Tempérée Nord : de +15 à +45 degrés
- 3 : Equateur : de -15 à +15 degrés
- 2 : Zone Tempérée Sud : de -15 à -45 degrés
- 1 : Pôle Sud : de -45 à -90 degrés

On obtiendra sans doute par ce biais une sous estimation des effets polaires (quels qu'ils soient).

COURBES $F=F(\text{LATITUDE})$

Elles sont repérées sur les graphiques par un nombre de trois chiffres.

Le premier indique la tranche d'activité solaire : (1 : Faible activité; 2 : Forte activité); le second : La saison; le troisième : La tranche d'altitude.

Chaque courbe prend ses valeurs en cinq points :

- de -90 à -45 repéré par PSUD
- de -45 à -15 repéré par TESU
- de -15 à +15 repéré par EQUA
- de +15 à +45 repéré par TENO
- de +45 à +90 repéré par PNOR

COURBES $F=F(\text{HEURE LOCALE})$

Elles sont repérées sur les graphiques par un nombre de trois chiffres, le premier indique la tranche d'activité solaire, le second la saison, le troisième l'altitude.

Chaque courbe prend ses valeurs en 6 points :

de 0 à 4 Heures repéré par 0004

de 4 à 8 Heures repéré par 0408

de 8 à 12 Heures repéré par 0812

de 12 à 16 Heures repéré par 1216

de 16 à 20 Heures repéré par 1620

de 20 à 24 Heures repéré par 2024

COURBES $F=F(\text{MOIS})$

Elles sont repérées sur les graphiques par un nombre de trois chiffres : Le premier indique l'activité solaire, le second l'altitude et le troisième la latitude.

Chaque courbe prend ses valeurs en 6 points :

Janvier Février :repéré par JAFE

Mars Avril :repéré par MAAV

Mai Juin :repéré par MAJU

Juillet Août :repéré par JUAO

Septembre Octobre :repéré par SEOC

Novembre Décembre :repéré par NODE

III LA DEFORMATION EN LATITUDE

L'examen des courbes $f=f(\text{latitude})$ permet de mettre en évidence un certain nombre de traits de structure. On remarque par exemple qu'on obtient bien le comportement parabolique attendu aux altitudes basses et que l'effet est plus fort en période de forte activité. On remarque également que cette évolution s'inverse au delà de 500 km. D'autres effets existent, mais il est difficile de les percevoir. Pour avoir une vision synthétique de la structure, nous avons fait l'Analyse Factorielle des Correspondances de ces courbes, selon les principes exposés en I. Il n'était pas à priori évident que la structure soit la même en période de faible et de forte activité. C'est pourquoi nous avons fait une analyse globale, et une analyse par période d'activité. Notons qu'il nous manque quatre courbes en période de forte activité, car certains points de ces courbes n'ont pu être estimés, faute d'observations.

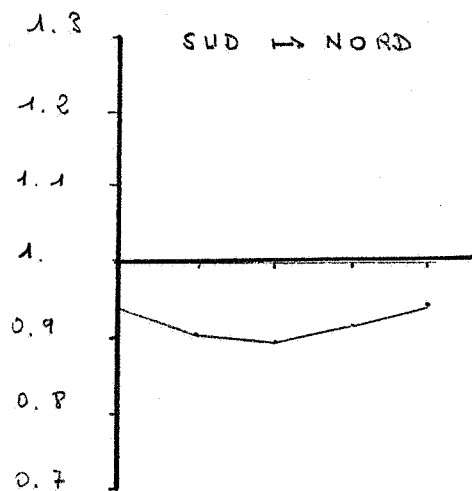
A/ RESULTATS GENERAUX

Notons d'abord les pourcentages d'inertie cumulés obtenus dans les différentes analyses :

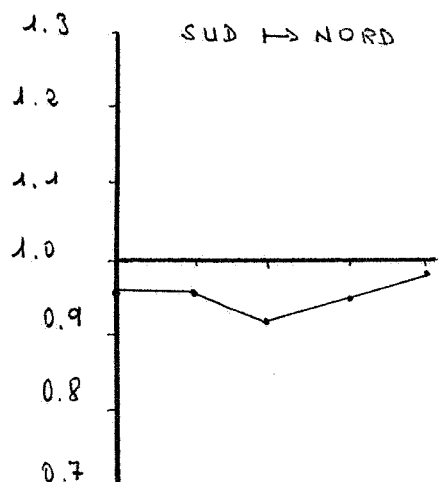
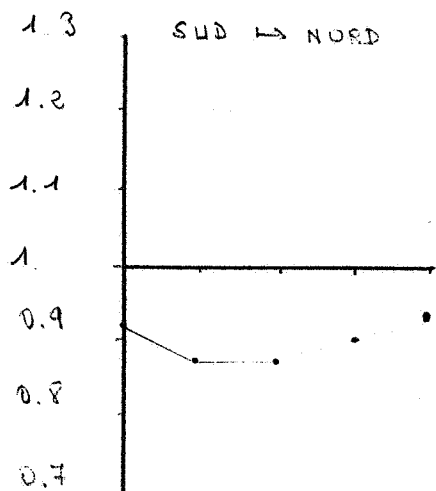
DIMENSIONS	1	2	3	4
analyse globale	65	89	97	100
faible activité (1)	59	86	96	100
forte activité (2)	73	93	99	100

Ces résultats nous incitent bien sûr à laisser de côté l'étude des axes 3 et 4, pour concentrer notre attention sur le plan des axes 1 et 2, qui rend compte à 90% environ de l'inertie des données.

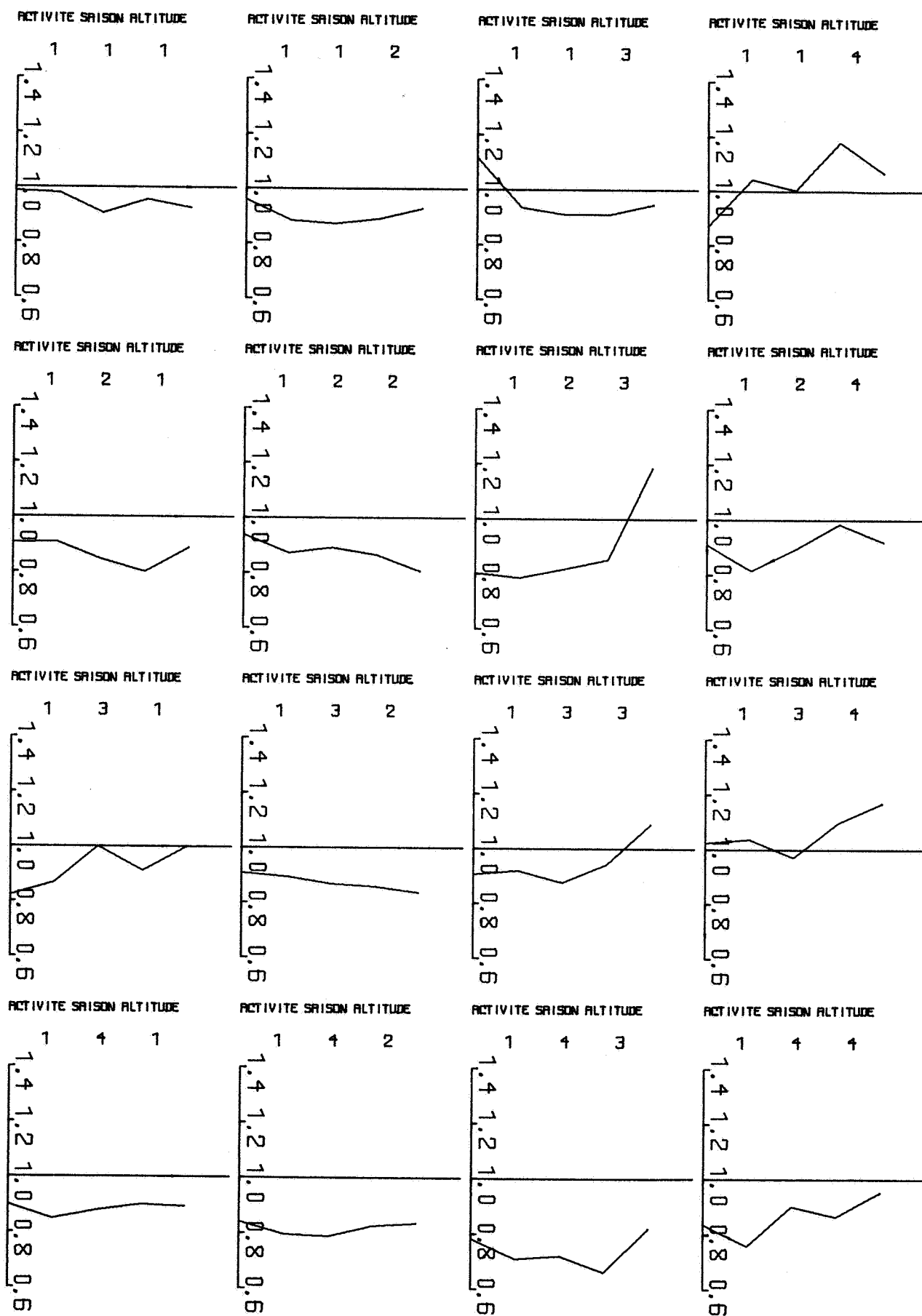
Dans ce plan la courbe joignant les caractères a une bonne



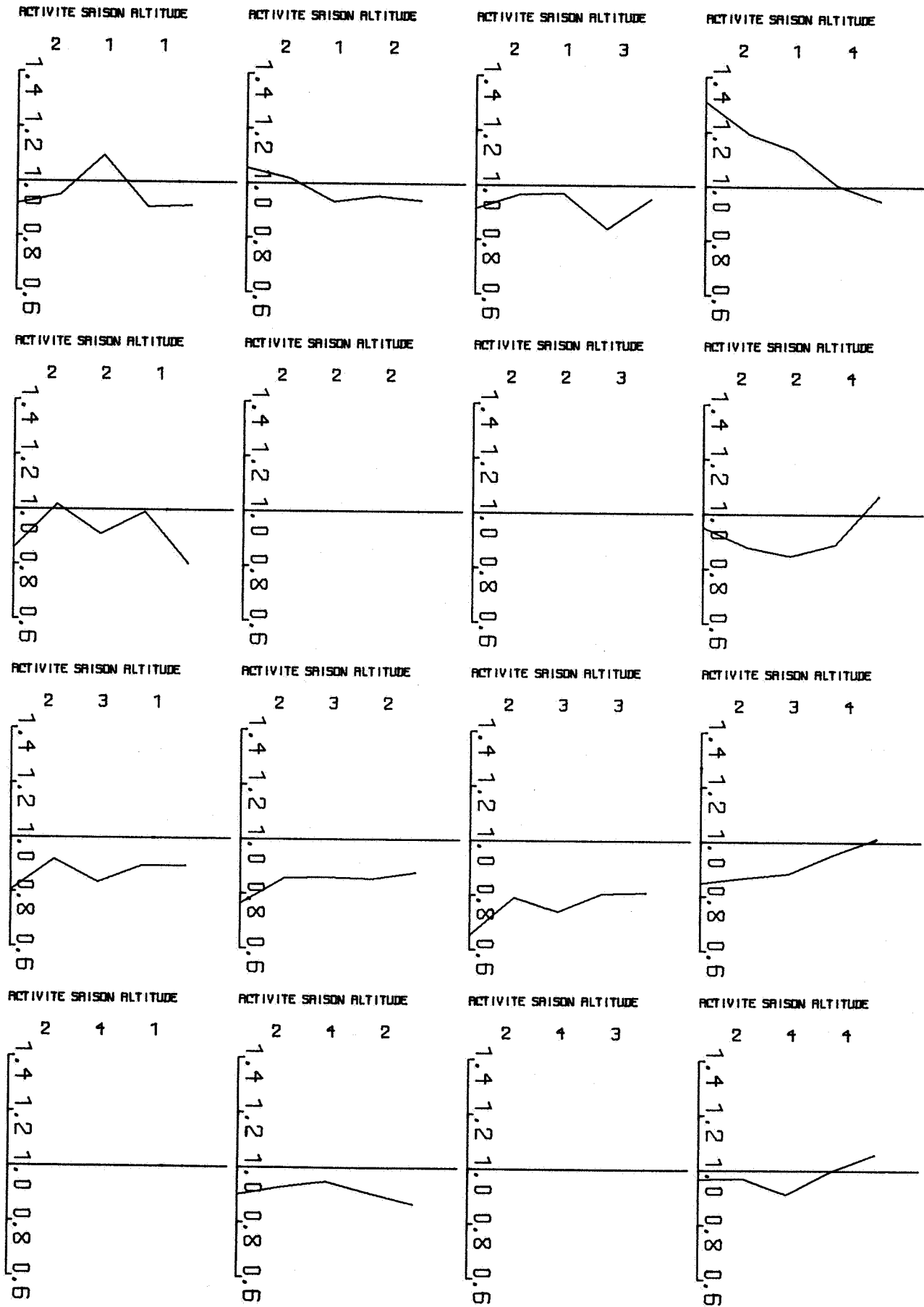
COURBES $F = f(\text{LATITUDE})$
 SUCCESSIVEMENT:
 COURBE MOYENNE
 COURBE MOYENNE FAIBLE ACTIVITE
 COURBE MOYENNE FORTE ACTIVITE



FAIBLE ACTIVITE



EVOLUTION DE F EN FONCTION DE LA LATITUDE : DU SUD AU NORD



EVOLUTION DE F EN FONCTION DE LA LATITUDE : DU SUD AU NORD

allure parabolique dans l'analyse globale et dans l'analyse de la période de forte activité. Elle présente par contre un pic assez curieux à l'équateur en faible activité. Il convient de noter que c'est la période de forte activité qui, bien que plus faiblement représentée (12 individus au lieu de 16) a imposé son type de courbe dans l'analyse globale :

On doit s'attendre à des coefficients correcteurs f beaucoup plus variables à mesure qu'on passe d'un minimum à un maximum solaire.

Précisons maintenant les faits, altitude par altitude.

B/ RESULTATS PAR TRANCHE D'ALTITUDE

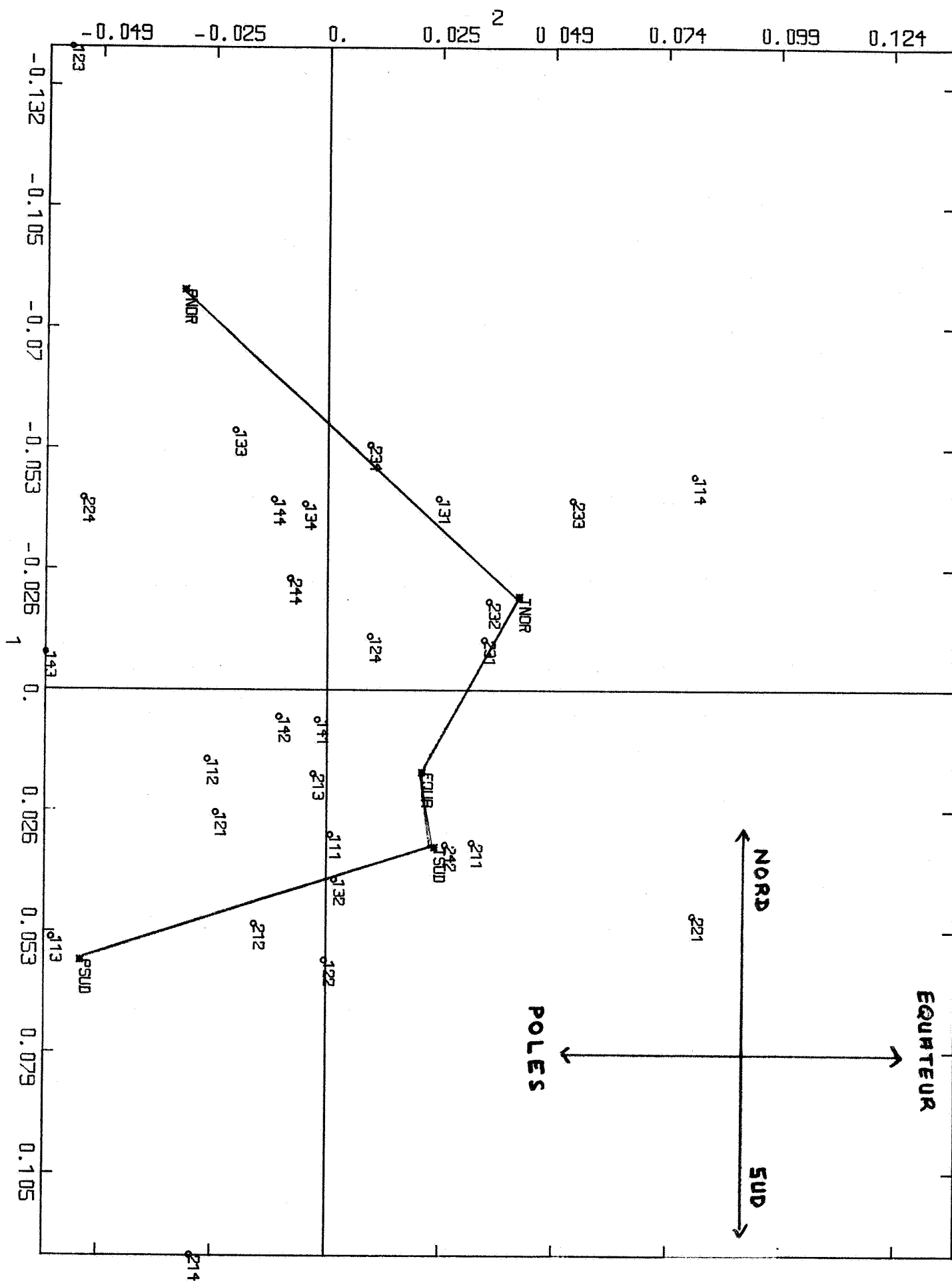
ALTITUDE 1 : DE 220 à 320 KM

Dans les 3 Analyses les points représentatifs de cette altitude se situent vers le sommet de la parabole, ce qui traduit le fait que les courbes $f=f(\text{latitude})$ croissent en général des pôles vers l'équateur. Les saisons modulent ce phénomène : Alors qu'il est décalé vers le sud en automne, hiver et surtout au printemps, il est nordique en été, pour la période de faible activité. En activité forte, le phénomène est sensiblement équivalent. Les maximums sont plus accusés, mais légèrement moins décalés vers le sud en hiver et au printemps; en été le maximum est du même ordre de grandeur, mais il est moins décalé vers le nord.

ALTITUDE 2 : DE 320 à 530 KM

Il convient, à cette altitude, de distinguer assez nettement les deux périodes d'activité solaire.

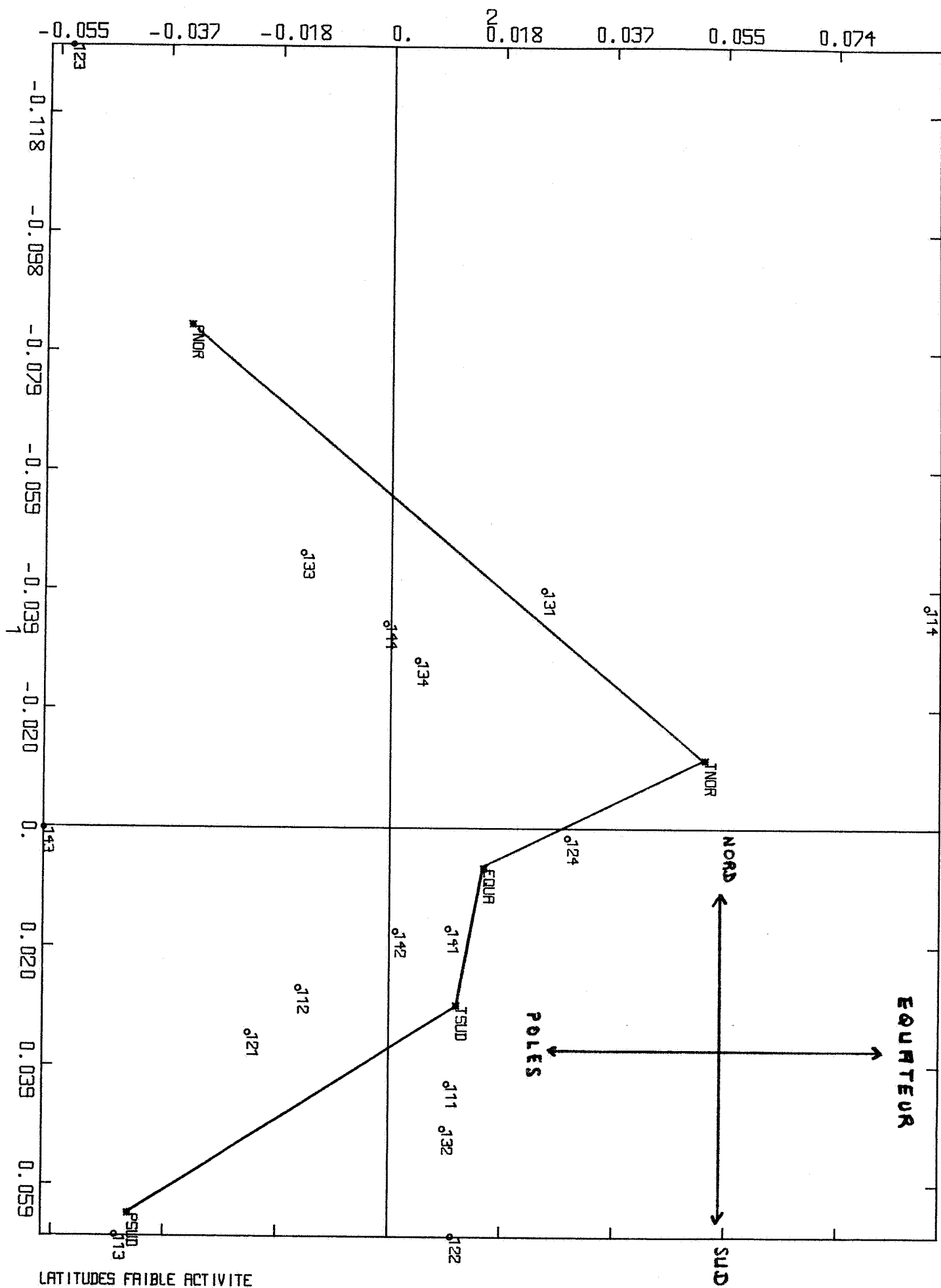
En faible activité, les courbes ont leur maximum au Sud : En septembre et en décembre la courbe ressemble déjà à celle de la tranche d'altitude supérieure (décroissante des pôles à l'équateur). Par contre au printemps et en été la courbe décroît assez régulièrement du pôle Sud au pôle Nord.



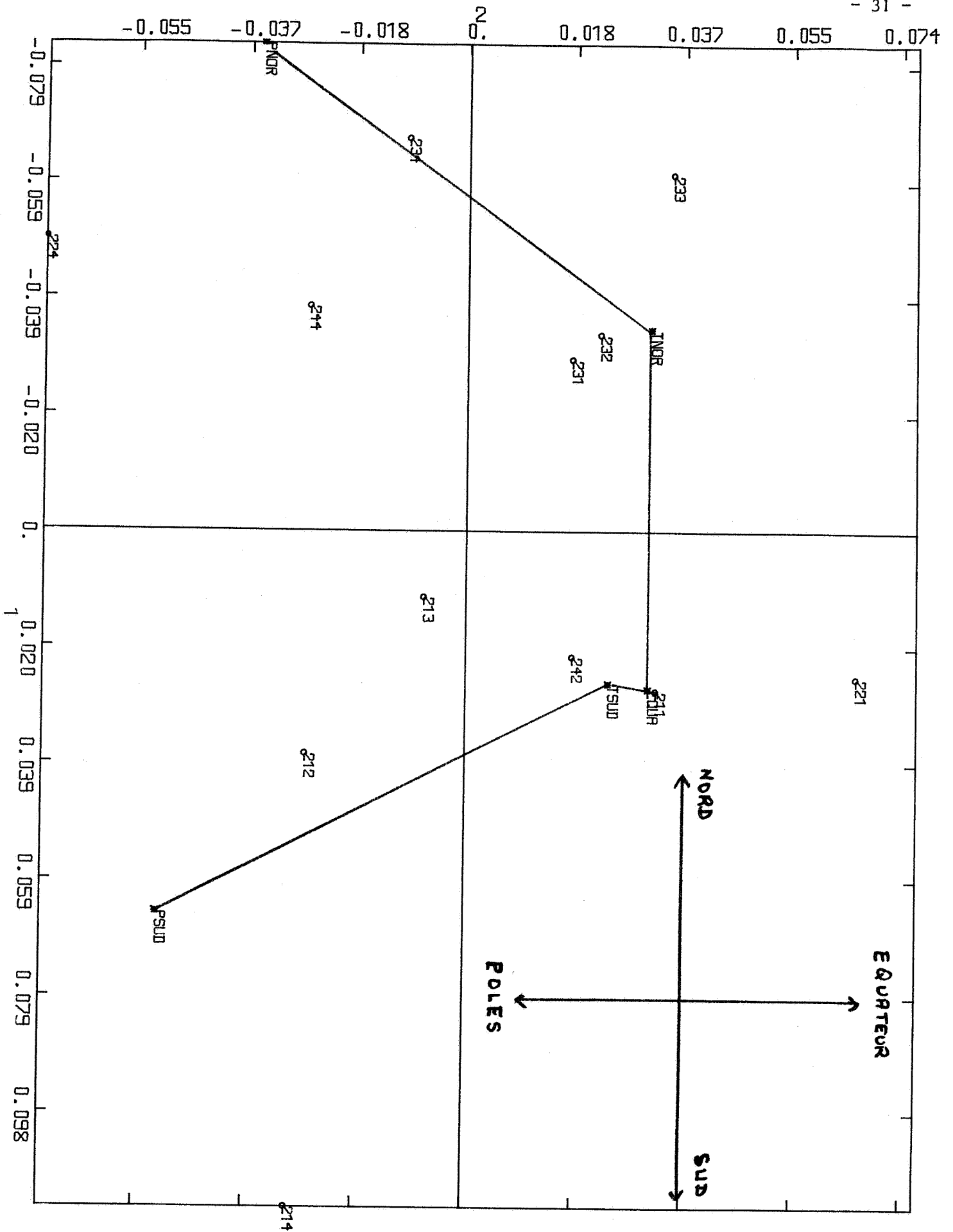
ANALYSE TOTALE.LATITUDES

AXE 1 : 53,7 %

AXE 2 : 26,5 %



LATITUDES FAIBLE ACTIVITE



LATITUDES FORTE ACTIVITE

En forte activité, on retrouve au contraire des traits de structure de la tranche d'altitude inférieure : Juin s'oppose à septembre et décembre : En juin, maximum prononcé en zone tempérée nord, en septembre maximum à l'équateur, mais légèrement décalé au sud, en décembre maximum très au sud. Cette tranche apparaît donc comme une charnière entre les zones où prédominent l'azote et l'oxygène atomique d'une part, et l'hélium d'autre part : L'atmosphère est susceptible de tassement ou d'expansion selon l'intensité de l'activité solaire : Sur une verticale, le comportement de l'atmosphère dépend essentiellement de la composition chimique, dont l'altitude ne rend compte qu'avec une certaine erreur : Selon que l'activité solaire est faible ou forte, la composition de l'atmosphère à une altitude donnée peut varier notablement.

L'ASYMETRIE NORD - SUD EN DEÇA DE 530 KM

On remarque, dans l'analyse générale, que tous les points relatifs aux altitudes 1 et 2 se placent en zone sud, sauf les points 131, 231, 232 représentatifs de l'été. Le comportement "sud" est général et prononcé à l'altitude 2 en faible activité (là, même en juin, le maximum est au sud), et à l'altitude 1, seul l'été fait exception. En forte activité, il faut être plus prudent car certains points manquent. On remarque néanmoins que seul l'été est décalé au nord, l'hiver, le printemps et l'automne, quand ils existent, se situent au sud.

ALTITUDE 3 : DE 530 à 750 KM

C'est la tranche qui a le comportement le plus nettement polaire, et dont la coefficient correcteur est susceptible des variations les plus fortes.

En faible activité, on a un maximum très accentué au pôle nord en mars et juin, au pôle sud en décembre, et un minimum équatorial en

septembre (l'atmosphère semble symétrique en septembre).

En forte activité, le maximum est légèrement au sud un hiver, plus nettement au nord en été, mais, curieusement, les effets en latitude semblent plus faibles qu'en période de faible activité.

ALTITUDE 4 : DE 750 à 1250 KM

Cette altitude est encore assez mal connue. La physionomie de l'atmosphère à cette altitude peut changer très vite. Dans le plan factoriel, les barycentres de deux périodes d'activité solaire sont presque confondus. L'évolution moyenne dans l'année semble identique, à l'exception toutefois de décembre : Maximum très accentué et très au sud en forte activité, très accentué et au nord en faible activité. En faible activité, le maximum se déplace lentement des zones tempérées nord vers le pôle nord de décembre à septembre.

C/ CONCLUSIONS SUR LA NATURE DE L'INFLUENCE DES DIFFERENTS PARAMETRES

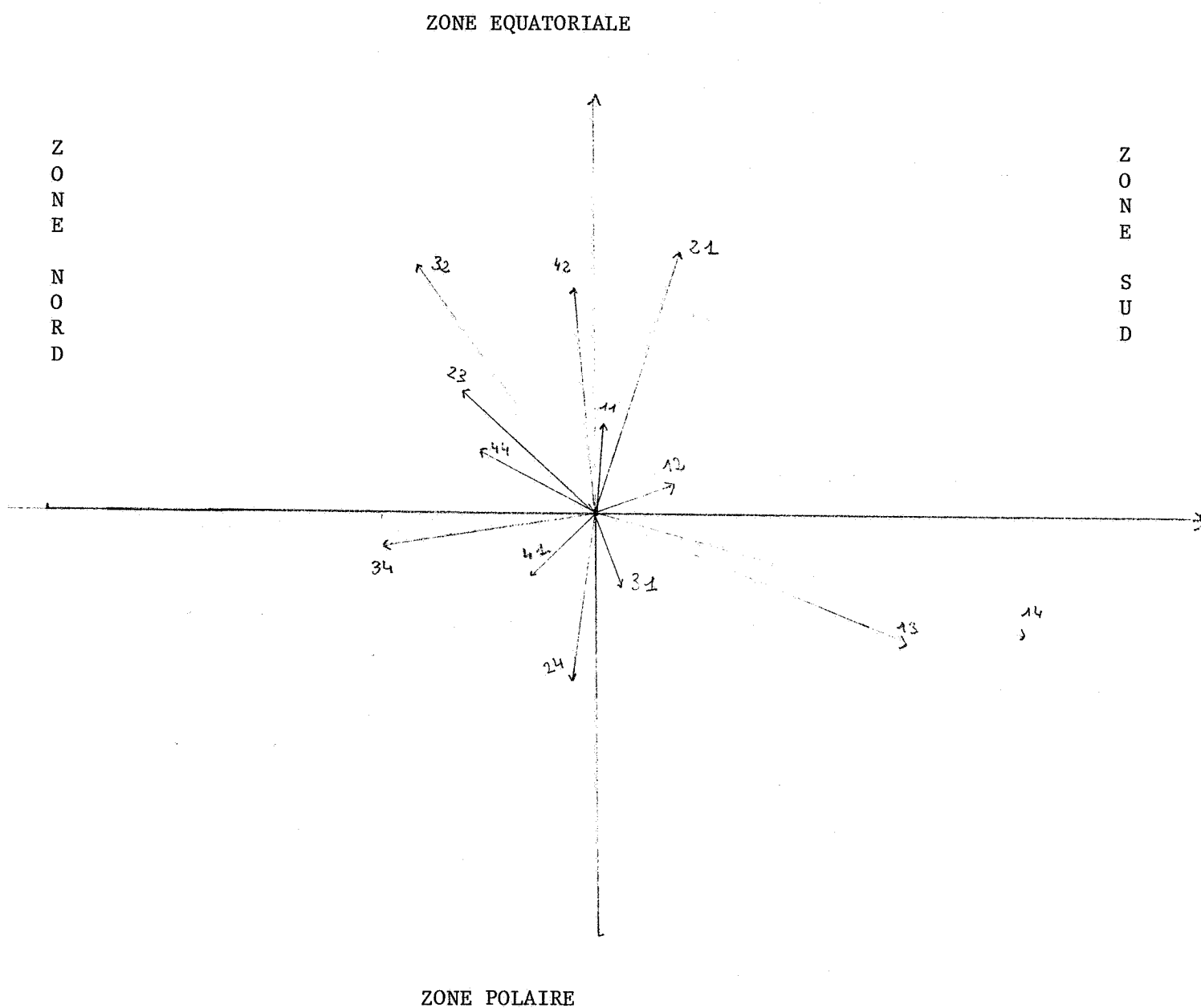
Il apparaît que l'altitude est le principal facteur de forme de l'effet en latitude. A cet égard, l'opposition entre la première et la troisième tranche, et le rôle de charnière de la deuxième tranche, sont les points importants mis en évidence. Il reste que l'altitude est un paramètre simple, facile à mesurer, mais qui ne synthétise pas de façon assez fine le paramètre le plus important : La composition chimique de l'atmosphère.

La période de l'année est un facteur de modulation de ces formes : Si l'on représente sur les graphiques une saison comme barycentre des points où cette saison intervient, les saisons viennent se confondre au centre de gravité du nuage des points, à l'exception cependant de l'hiver : Son point représentatif reste dans la zone sud. Le seul point "décembre" ne satisfaisant pas à la règle concerne les altitudes supérieures à 750 km en faible activité solaire. Ce qui laisse à penser

qu'en moyenne le modèle JACCHIA 71 - symétrique dans l'espace - temps, en ce qu'il postule que toutes choses égales par ailleurs l'atmosphère a la même structure en juin dans le nord et en décembre dans le sud est entaché d'un léger biais : Les données montrent qu'en fait le modèle sous estime plus en moyenne en décembre au sud qu'en juin au nord.

L'activité solaire elle ne modifie pas la structure des effets; elle a simplement tendance à les amplifier d'une façon générale, mais surtout dans les deux tranches d'altitude basse; au dessus de 580 km, certains effets sont amplifiés, d'autres sont au contraire réduits, et il n'est pas possible de dégager une ligne générale.

Sur le graphique de la page suivante nous avons représenté par des vecteurs issus de l'origine le déplacement des points saison - altitude quand on passe de faible à forte activité. Le déplacement des points est systématiquement vers l'équateur aux altitudes inférieures à 500 km (les courbes ont de plus en plus tendance à avoir un maximum accusé vers l'équateur), vers le sud en décembre, vers le nord en été, sauf pour les altitudes inférieures à 320 km.



DEPLACEMENT DES POINTS SAISON-ALTITUDE QUAND ON PASSE DE FAIBLE A FORTE
ACTIVITE SOLAIRE

IV LA DEFORMATION EN HEURE LOCALE

L'examen des courbes $f=f(\text{heure locale})$ montre une différence assez notable de comportement selon la période d'activité solaire. On a tendance à avoir un maximum de nuit en faible activité, et un maximum de jour en forte activité. Ces effets se compensent d'ailleurs assez bien, puisque la courbe moyenne des 32 courbes est très plate : f a un maximum de 0,92 à 10h et un minimum de 0,89 à 22h.

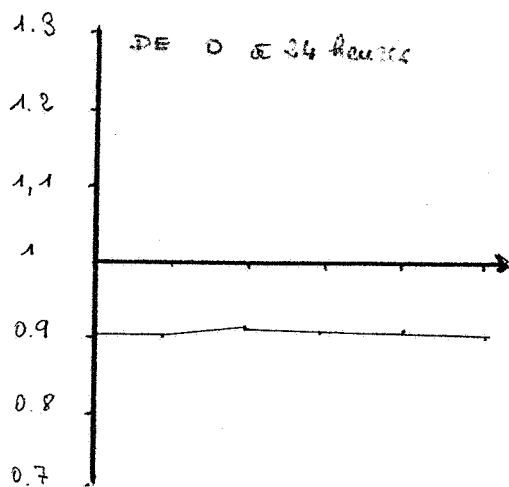
L'analyse des correspondances du jeu de courbes va nous faire découvrir des traits de structure plus fins.

A/ ANALYSE DES CORRESPONDANCES : RESULTATS GENERAUX

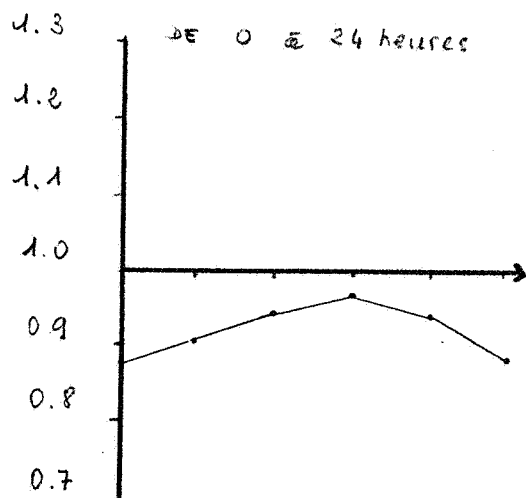
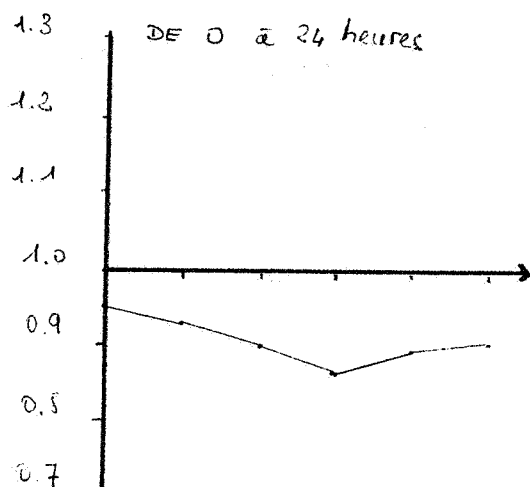
Nous avons réalisé une analyse du fichier complet, et une analyse par période d'activité solaire. Les pourcentages cumulés d'inertie sont les suivants.

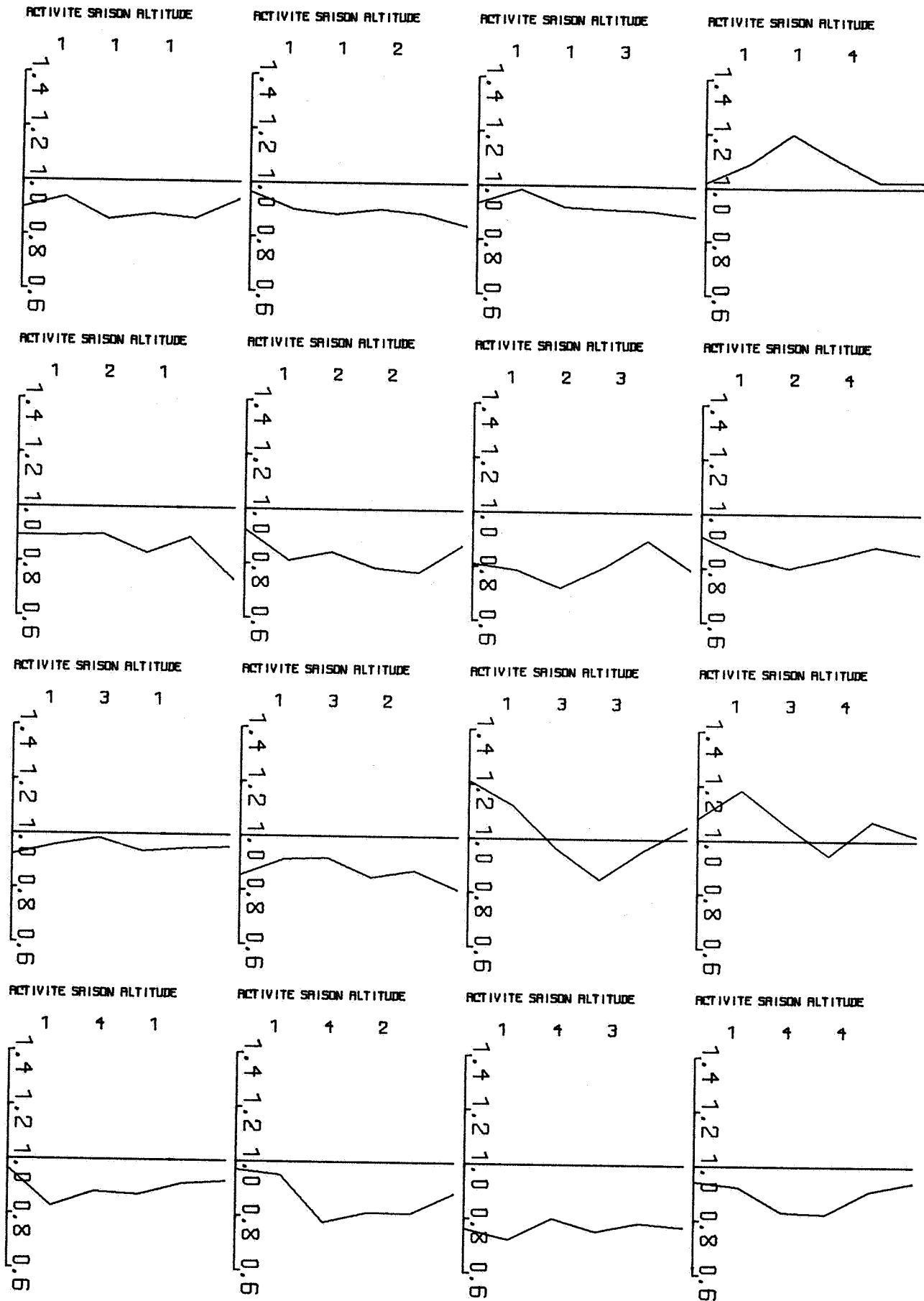
AXES	1	2	3	4	5
Fichier complet	46	71	85	95	100
Activité faible	48	71	86	95	100
Activité forte	45	69	88	97	100

Considérons d'abord l'analyse du fichier complet, on constate que le premier axe (qui prend en compte 46% de l'inertie) est parfaitement discriminant : Tous les points situés à gauche appartiennent à la période de faible activité (à l'exception de 241). Tous les points situés à droite (à l'exception de 131 et 143, légèrement à droite, et de 114 nettement à droite) appartiennent à la période de forte activité. Le trait majeur mis en évidence est donc celui que nous avons perçu directement à la lecture du jeu de courbe. Les courbes $f=f(\text{heure locale})$ présentent un maximum nocturne et un minimum diurne en faible activité, et

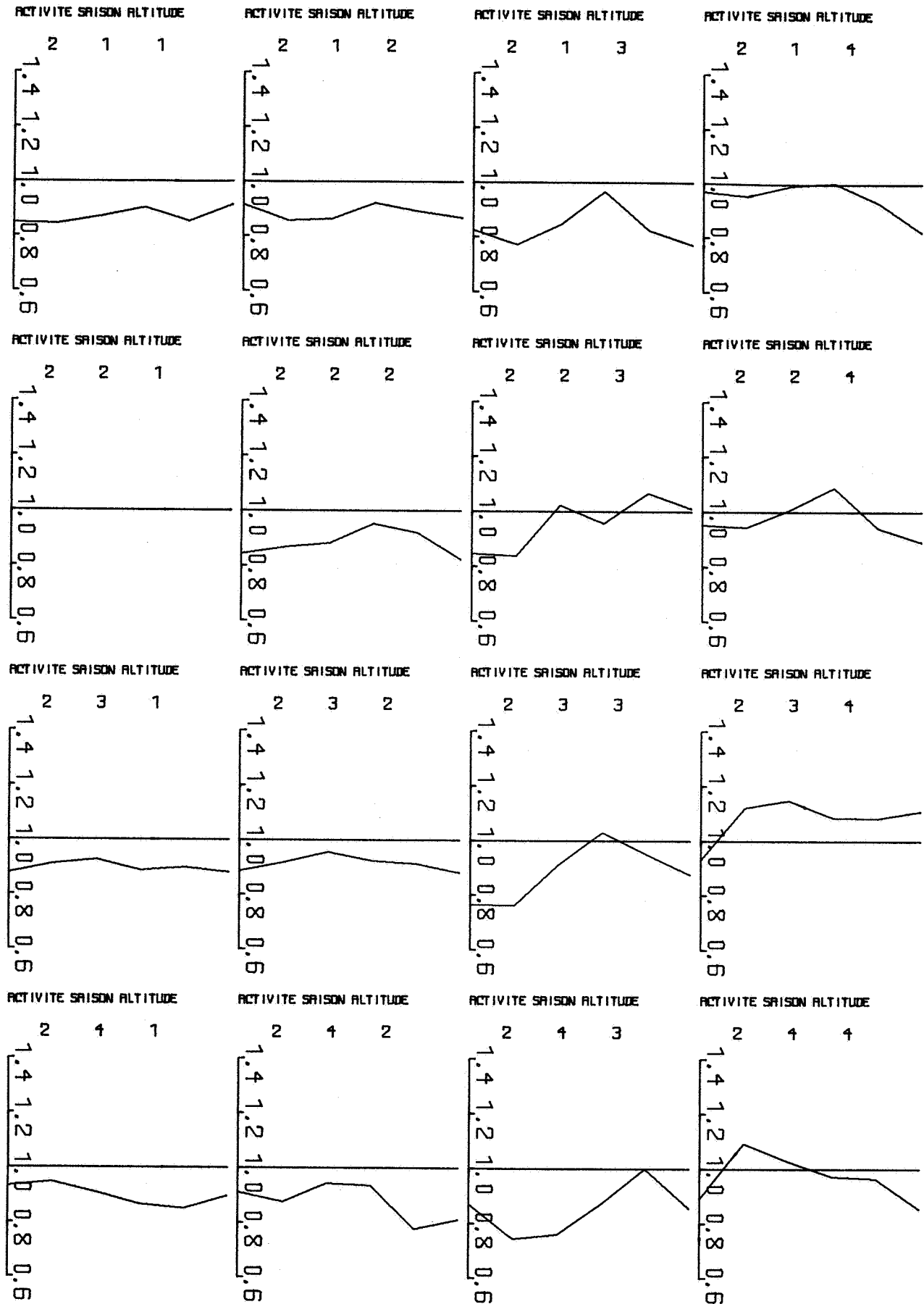


COURBES $F = F$ (HEURES LOCALES)
 SUCCESSIVEMENT
 COURBE MOYENNE
 COURBE MOYENNE FAIBLE ACTIVITE
 COURBE MOYENNE FORTE ACTIVITE





EVOLUTION DE F EN FONCTION DE L'HEURE LOCALE DE 0 A 24 HEURES



EVOLUTION DE F EN FONCTION DE L'HEURE LOCALE DE 0 A 24 HEURES

et présente au contraire un maximum diurne et un maximum nocturne en forte activité : JACCHIA 71 sous estime le jour en forte activité.

B/ L'EFFET DES AUTRES PARAMETRES: ANALYSE GLOBALE

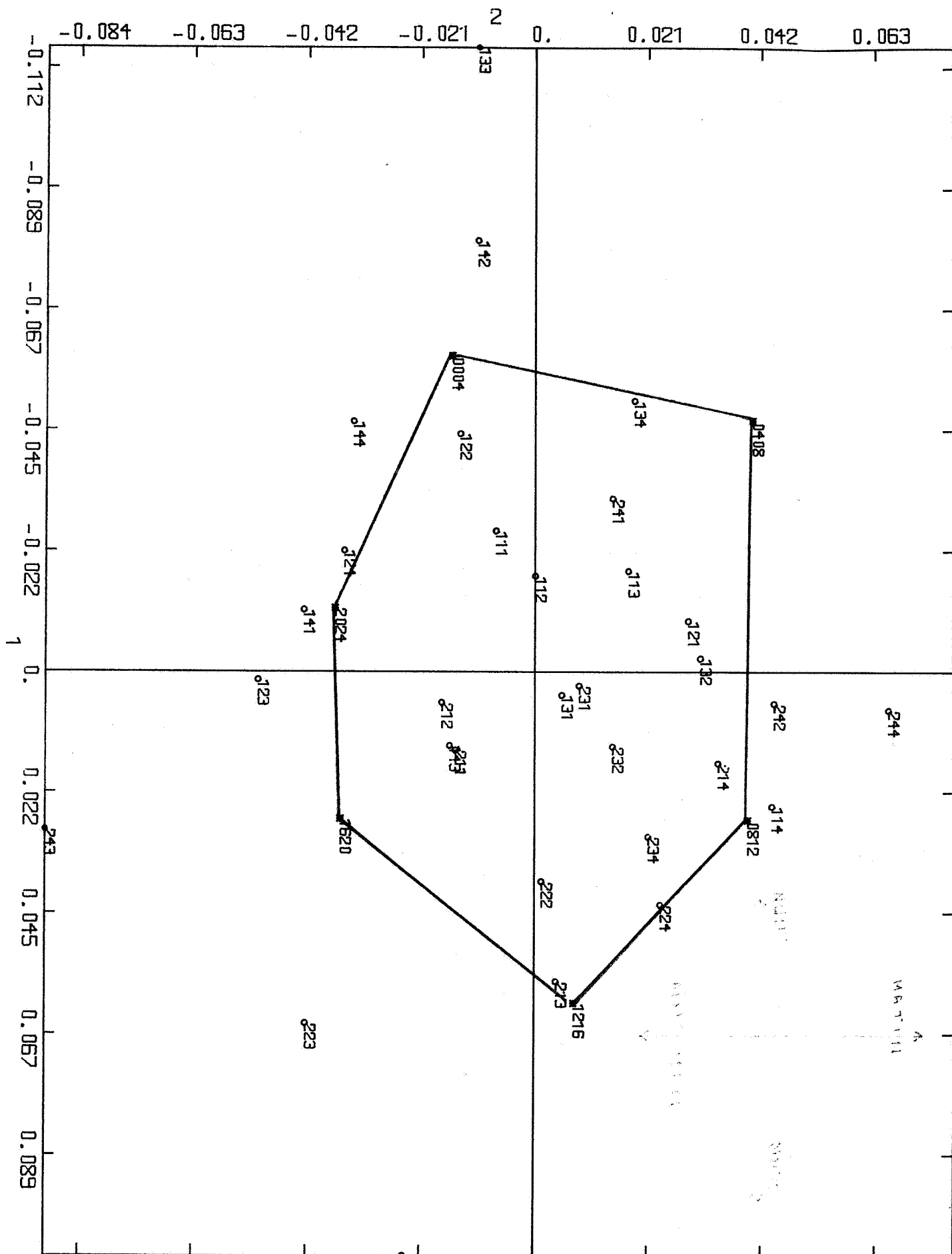
L'opposition entre les deux périodes d'activité solaire n'a pas du tout la même importance aux altitudes faibles et fortes.

Les points 1A1 et 2A1, représentant l'altitude 1 en faible et forte activité solaire, sont presque confondus : En dessous de 320 km, la différence est peu sensible. Les points 1A2 et 2A2 sont beaucoup plus nettement séparés, et les polygones de sommet 112, 122, 132, 142 d'une part, et 212, 222, 232, 242 d'autre part, ont une intersection vide. En faible activité, le maximum reste très nocturne (vers minuit), sauf en été local où il passe par un maximum vers 8h. En forte activité au contraire, le maximum se déplace continuellement de 20h en hiver local à 10h en automne local.

A l'altitude supérieure (de 530 à 750 km) l'effet est prédominant : Points 1A3 et 2A3 très éloignés, polygone 113, 123, 133, 143 et 213, 223, 233, 243, d'intersection vide. En faible activité, le maximum est vers 21h aux équinoxes, vers 3h aux solstices. En forte activité, les maximums restent situés vers 15h aux solstices, et 18h aux équinoxes.

Au dessus de 750 km, les points 214, 224, 234, 244 sont remarquablement groupés autour de leur barycentre situé vers 10h : En forte activité, le maximum reste vers 10h, indépendamment de la saison. En faible activité, les points restent groupés vers 2h, sauf en hiver, dont le point semble aberrant.

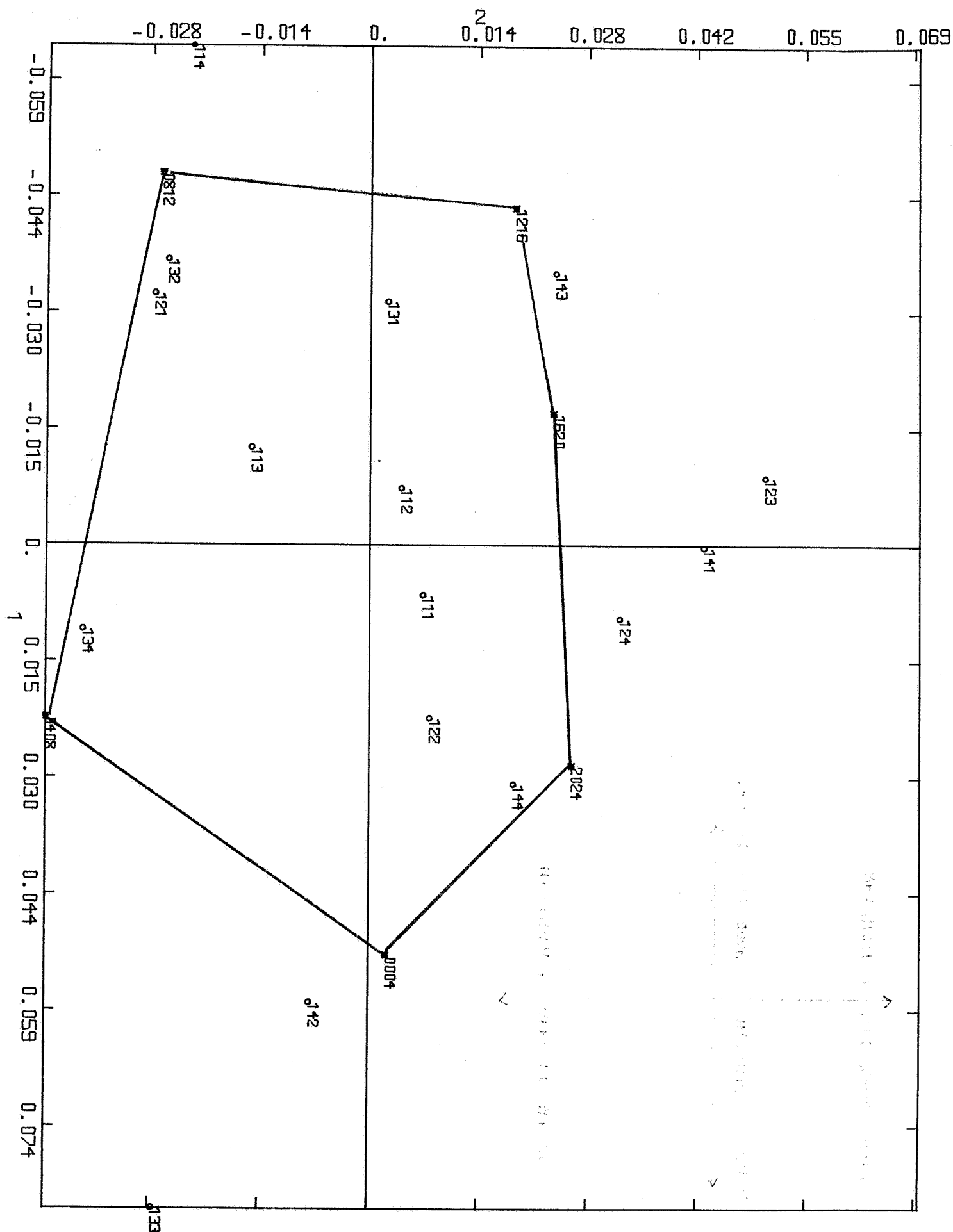
Les points représentatifs des saisons sont remarquablement groupés vers midi en forte activité; en faible activité par contre une légère différenciation existe entre les équinoxes (maximum vers minuit) et les solstices (maximum vers 4h).

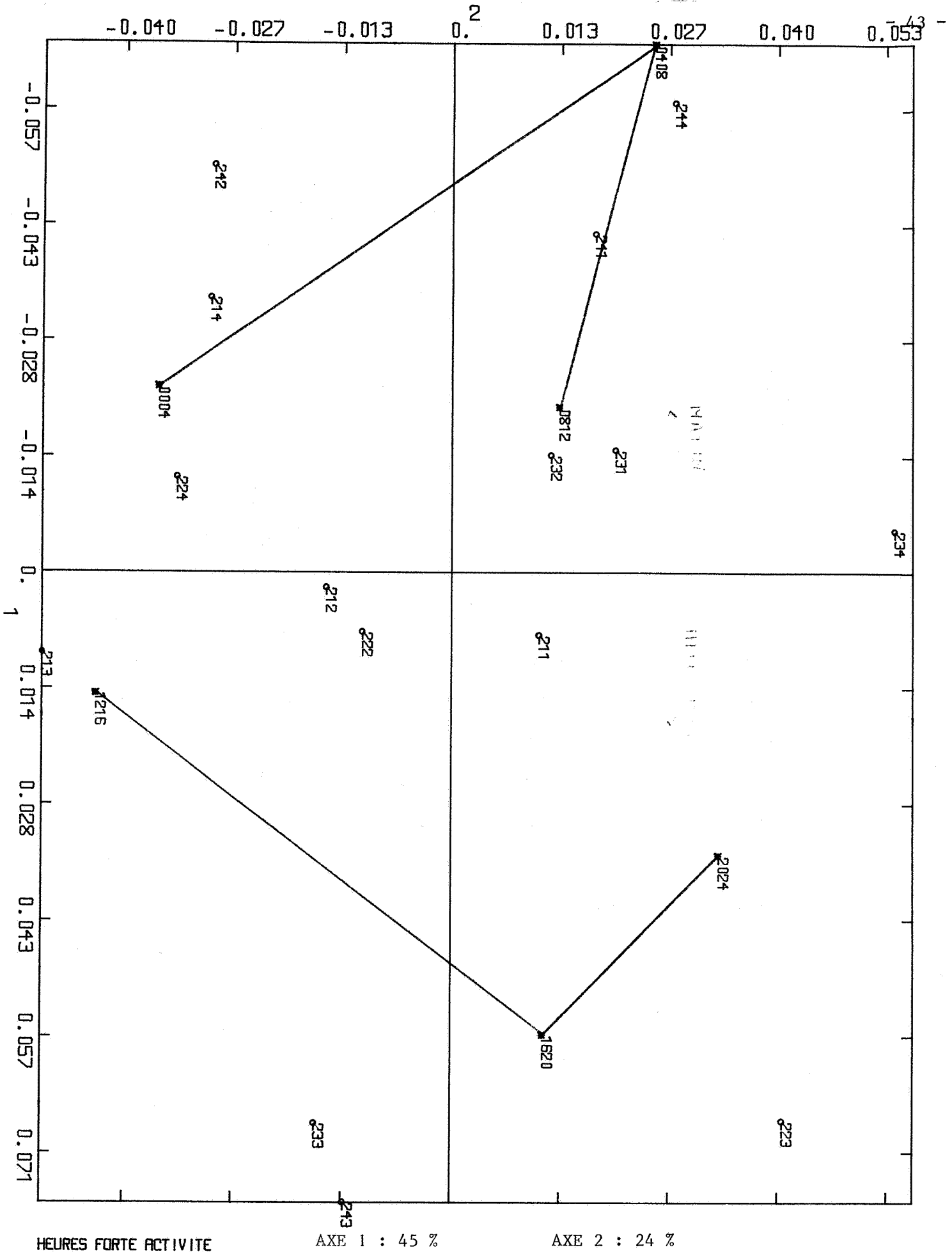


ANALYSE TOTALE. HEURES

AXE 1 : 45,9 %

AXE 2 : 25,3 %





C/ LES ANALYSES PAR PERIODE D'ACTIVITE

Dans les analyses, il faut prendre garde que la courbe de référence (courbe moyenne, qui, si elle était représentée, serait projetée au centre de gravité) n'est pas une courbe plate : Elle présente un maximum nocturne et un minimum diurne en faible activité, l'effet contraire en forte activité. Le "classement" proposé par les graphiques se situe par rapport à cette courbe.

En faible activité, il se produit un phénomène intéressant : La courbe joignant les différentes heures locales cesse de présenter l'aspect circulaire qu'elle avait dans les deux premières analyses pour se "casser" en deux, une courbe pour le matin, et une courbe pour l'après-midi. Rappelons que la courbe de référence (situé au centre de gravité) présente un maximum vers 14h. L'analyse de la forte activité oppose essentiellement les altitudes 3 et 4 : Entre 530 et 750 km, le maximum est déplacé dans l'après-midi vers 16h, alors que au dessus de 750 km il est déplacé dans la matinée vers 10h.

D/ CONCLUSION SUR LA NATURE DES DIFFERENTS PARAMETRES

L'activité solaire est le principal facteur de forme des effets. Le modèle JACCHIA 71, modèle moyen pour toute la période d'un cycle solaire de 11 ans, ne rend pas bien compte de l'influence de l'activité solaire sur la variation de la densité en heure locale. Il sous estime la nuit quand l'activité est faible, sous estime le jour quand l'activité est forte. L'effet est d'autant plus important qu'on s'élève en altitude : L'altitude est un facteur d'intensité, mais ne structure par elle même les effets. Les saisons introduisent une modulation relativement faible et erratique. On peut seulement dire qu'en faible activité; été et hiver se ressemblent beaucoup plus que printemps et automne.

V LA DEFORMATION SEMI-ANNUELLE

L'examen des courbes $f=f(\text{date})$ montre essentiellement une tendance à la sous estimation en décembre et janvier, à la surestimation en juin, juillet, août.

Nous reproduisons ci-après les moyennes de toutes ces courbes en faible et forte activité, et toutes périodes confondues.

On constate que cet effet est surtout sensible en forte activité. En faible activité par contre, on remarque que le modèle surestime surtout aux équinoxes. En moyenne sur les deux périodes, les effets s'ajoutent, et on observe une surestimation assez nette de mars à octobre.

Nous avons, comme pour les études précédentes, effectué trois Analyses de Correspondances sur les courbes : Une analyse en faible activité, une analyse en forte activité, et une analyse globale.

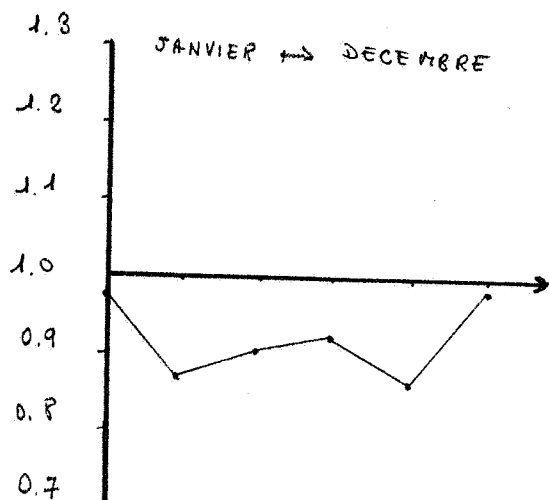
A/ RESULTATS GENERAUX

Nous reproduisons ci dessous les taux d'inertie cumulés relatifs aux axes factoriels.

DIMENSIONS	1	2	3	4	5
Faible activité	36	64	83	94	100
Forte activité	56	78	90	96	100
Analyse globale	47	69	82	93	100

Le plan des axes 1-2 restitue donc environ 70% de l'inertie. Nous ignorerons dans la suite les axes suivants.

L'analyse globale sépare d'une façon bien compréhensible les deux périodes d'activité. Le mécanisme en est le suivant : Les points relatifs à l'activité forte sont attirés par novembre, décembre, janvier, février ce qui est bien normal puisque la moyenne des courbes présente



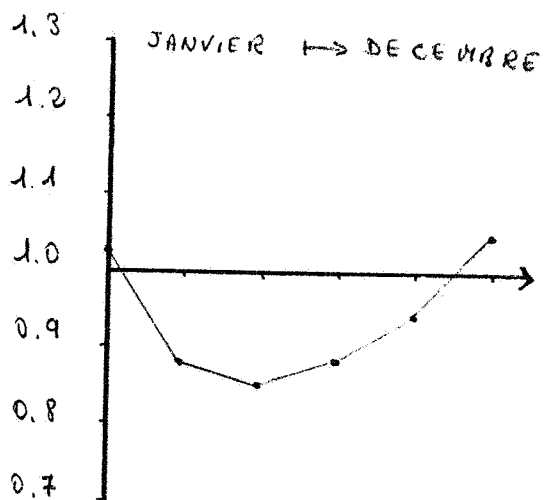
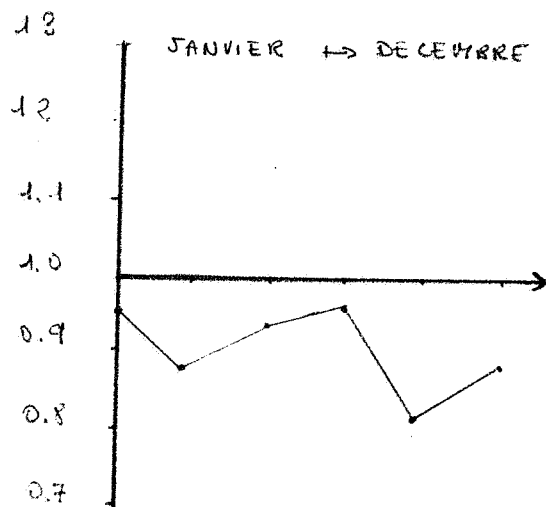
COURSES $F = F(\text{DATE})$

SUCCESSEMENT

COURSE MOYENNE

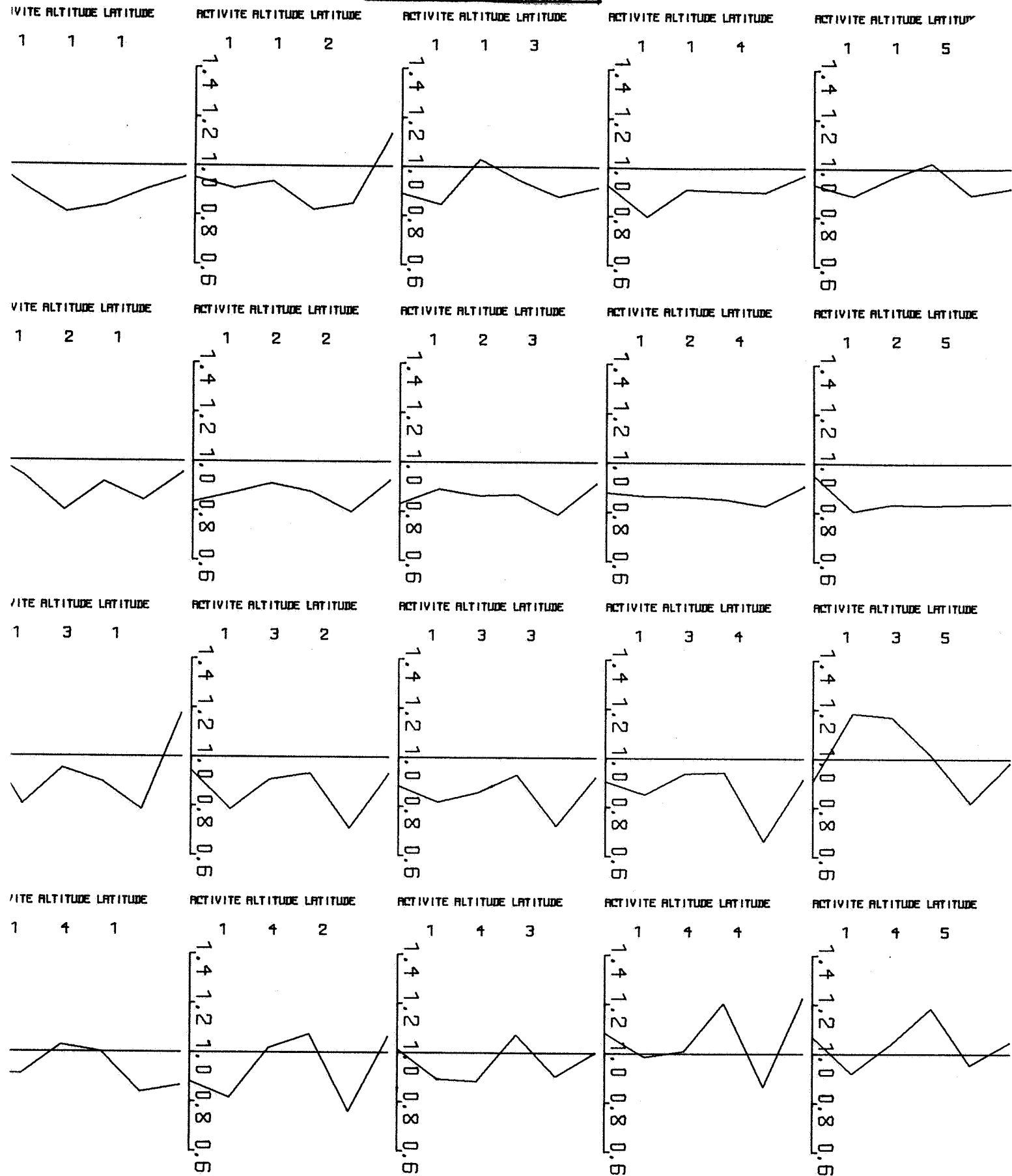
COURSE MOYENNE - FAIBLE ACTIVITE

COURSE MOYENNE - FORTE ACTIVITE



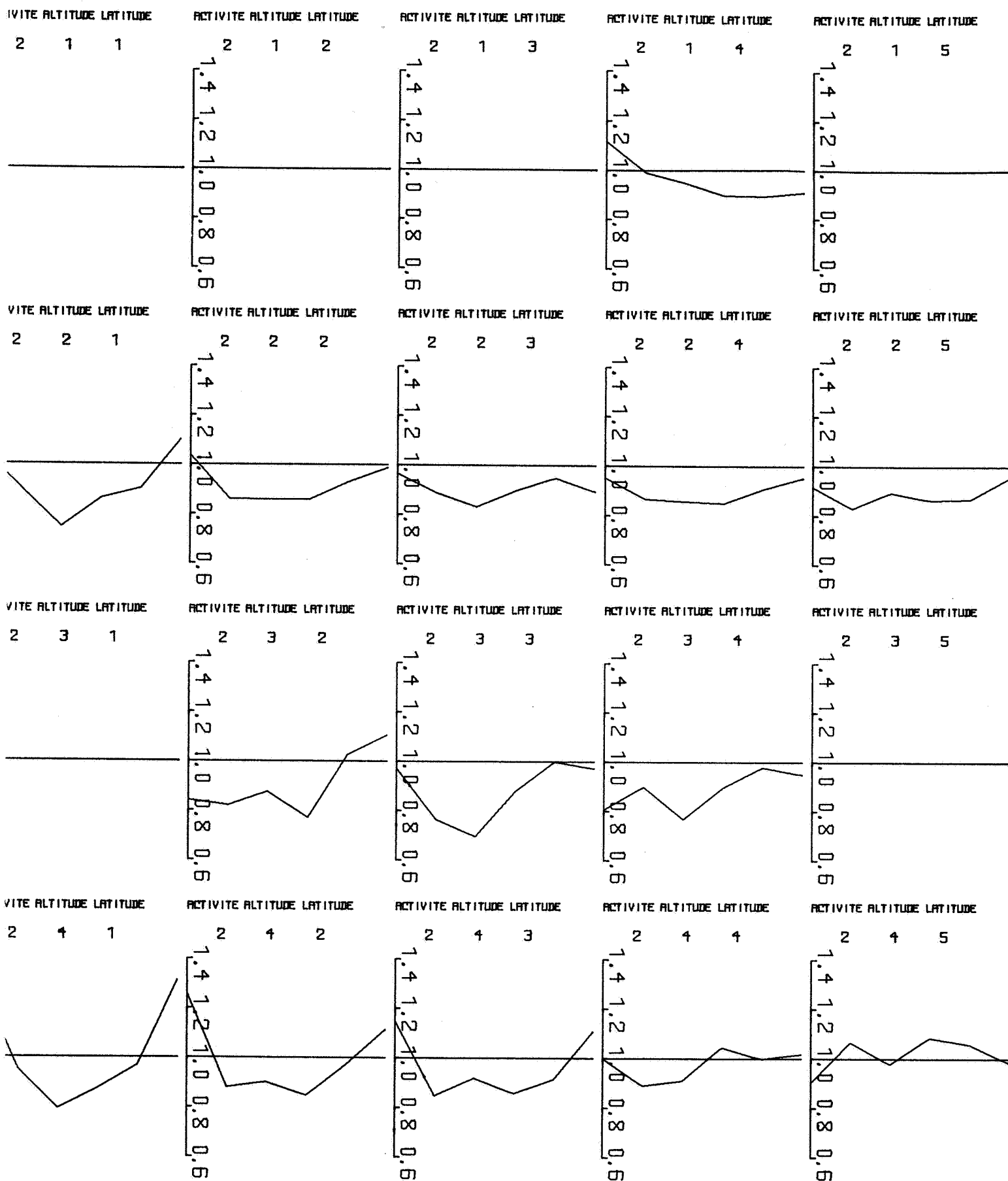
FAIBLE ACTIVITÉ

- 47 -



Evolution de F en fonction de la date de l'année à l'échelle
EVOLUTION DE F EN FONCTION DE LA DATE. DE JANVIER A DECEMBRE

EVOLUTION DE F EN FONCTION DE LA DATE. DE JANVIER A DECEMBRE



Evolution de F en fonction de la date. De janvier à décembre
 EVOLUTION DE F EN FONCTION DE LA DATE. DE JANVIER A DECEMBRE

EVOLUTION DE F EN FONCTION DE LA DATE. DE JANVIER A DECEMBRE.

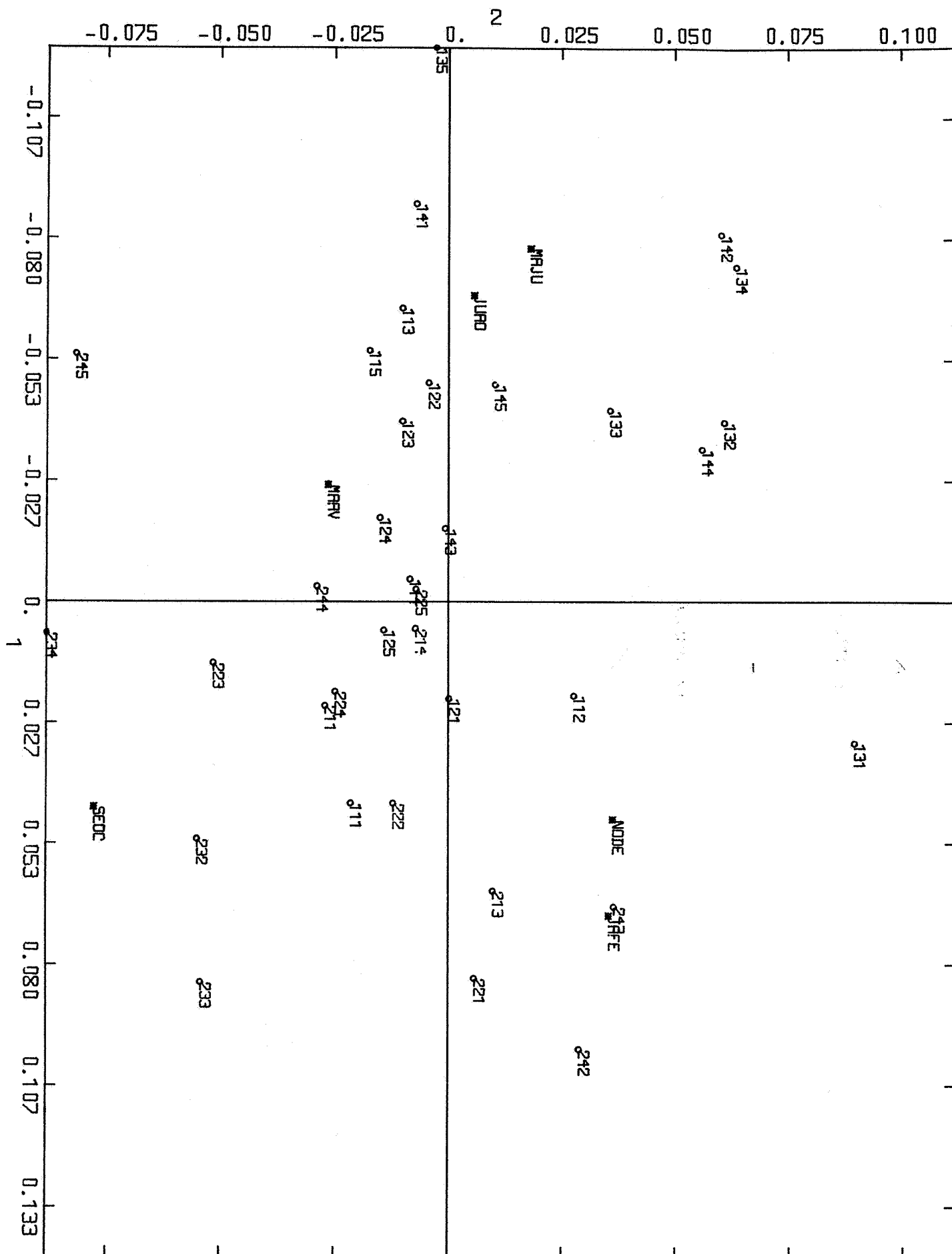
en ces mois un maximum assez prononcé. Ils sont aussi repoussés vers septembre, octobre par les points de la période 1 : Ceux-ci présentent en moyenne un minimum prononcé en septembre-octobre, alors qu'en forte activité la valeur de f en septembre - octobre, sans être maximale, est au-dessus de la moyenne générale.

Les points de la période 1 (activité faible) sont au contraire attirés par mai, juin, juillet, août, points où leur courbe moyenne présente un maximum.

On remarque enfin que si l'on veut définir 4 périodes dans l'année (comme nous l'avons déjà fait dans les analyses précédentes), il est bien judicieux de grouper novembre, décembre, janvier, février d'une part, et mai, juin, juillet, août d'autre part, les équinoxes jouant un rôle spécifique, mais nullement identique (mars - avril est assez loin de septembre - octobre).

B/ ETUDE DETAILLEE DE L'INFLUENCE DES DIFFERENTS PARAMETRES

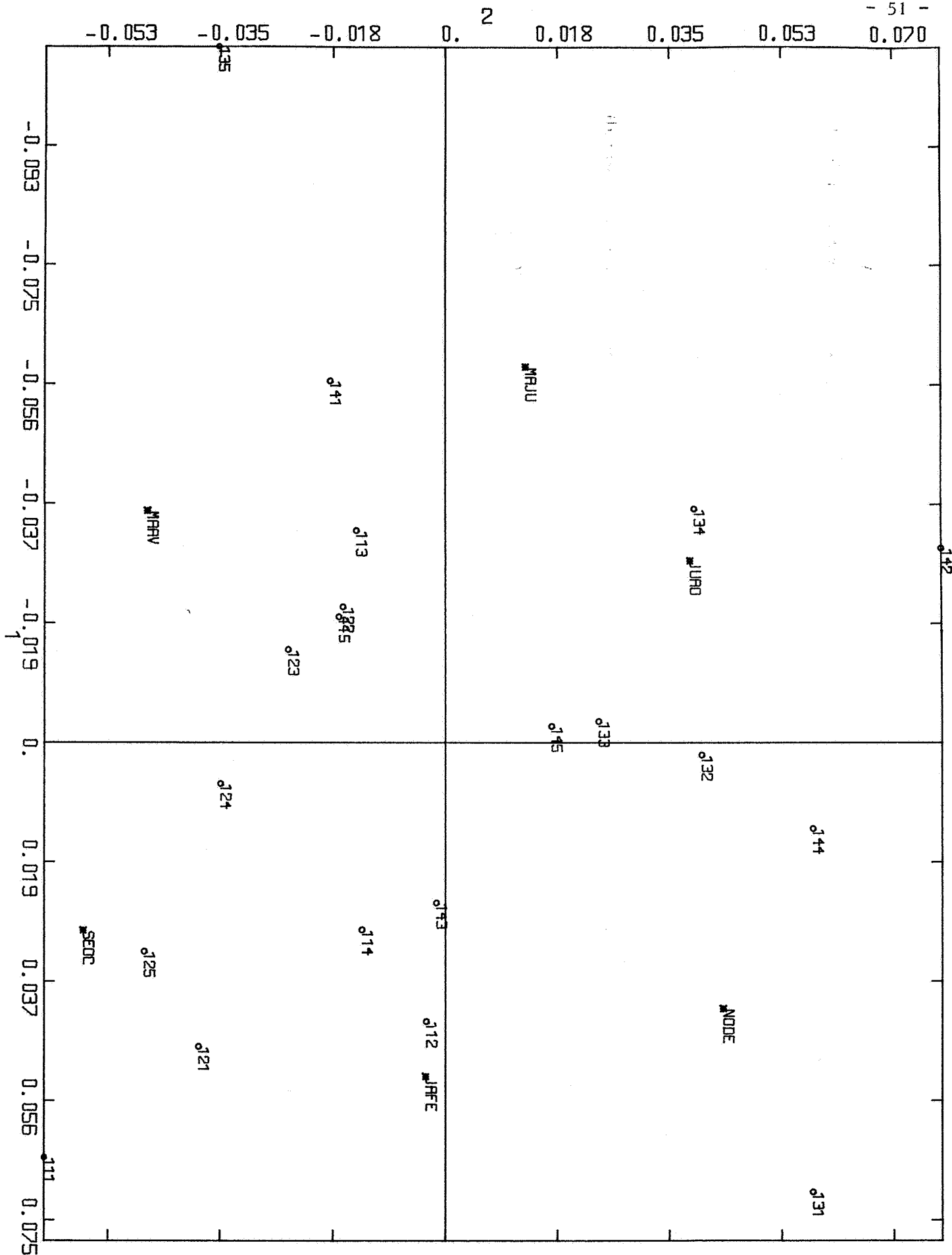
Il est intéressant de tracer, sur les plans factoriels, les courbes joignant les différentes latitudes, pour une altitude et une activité donnée on s'aperçoit alors du phénomène suivant : Aux altitudes 1 et 2, les points sont relativement bien groupés. A l'altitude 3, en faible activité la courbe décrit tout le plan factoriel, voisinant décembre au sud et juin au nord : On observe donc une sous estimation relative en été local. Ce phénomène se produit d'une façon strictement identique à l'altitude 4, et avec une ampleur comparable, en forte activité. Nous ne pouvons malheureusement rien dire quant à l'altitude 3 en forte activité, car les points concernant les pôles manquent. On voit également sur l'analyse globale qu'à l'altitude 4, en faible activité, on a, quelque soit la latitude, un coefficient f plus fort que la moyenne en mai, juin, juillet, août.



ANALYSE TOTALE.MOIS

AXE 1 : 47,4 %

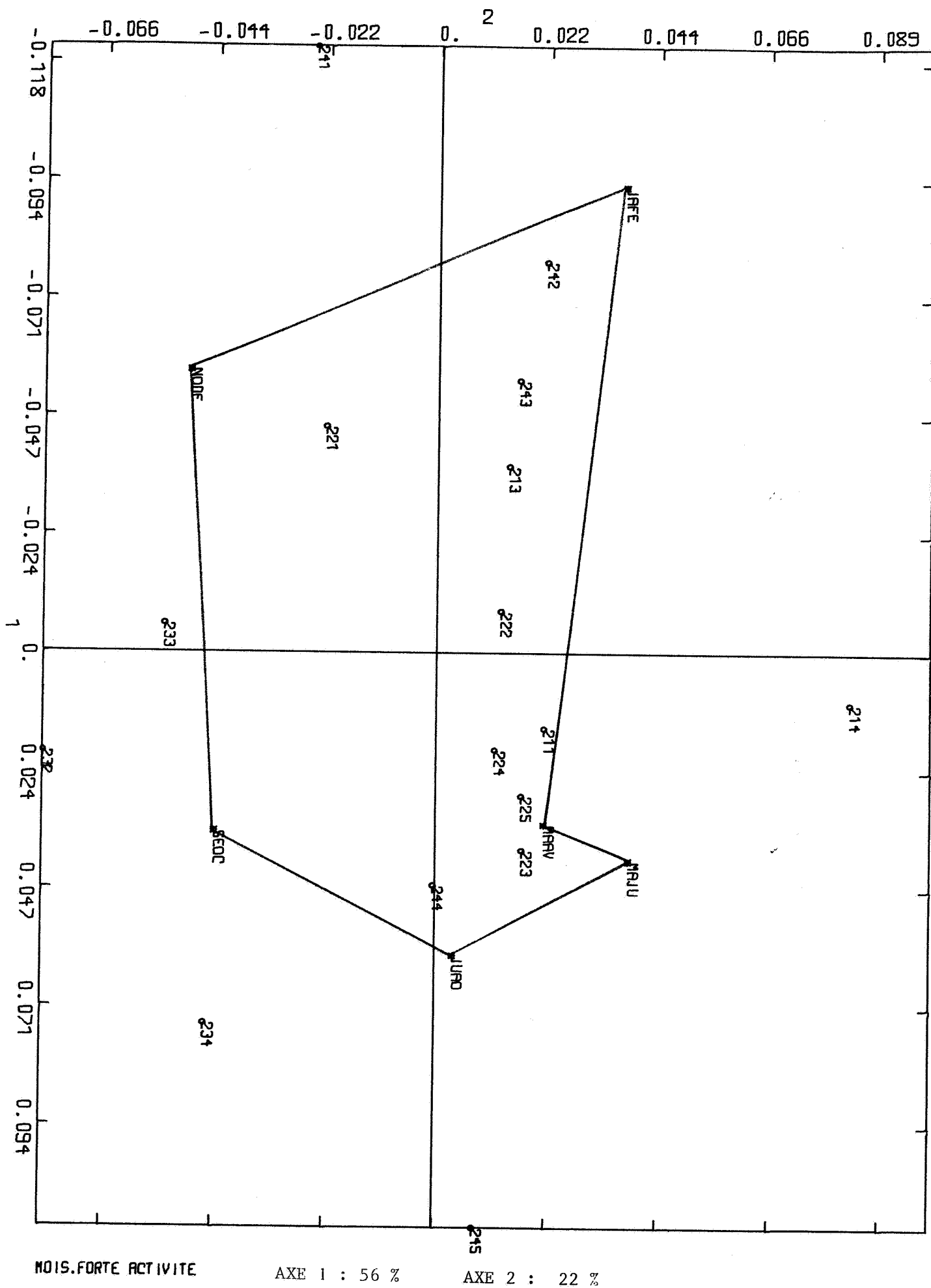
AXE 2 : 21,9 %



MOIS FAIBLE ACTIVITE

AXE 1 : 36 %

AXE 2 : 28 %



AXE 2 : 22 %

Revenons maintenant aux altitudes basses et examinons l'altitude 2 en forte activité : Le pôle sud est très proche de novembre, décembre, alors que toutes les autres latitudes sont groupées aux alentours de mars, avril, mai, juin. La signification en est la suivante (on peut d'ailleurs le vérifier en examinant les courbes correspondantes) f présente un maximum en décembre - janvier très accusé au pôle sud, alors qu'aux autres latitudes les courbes s'écartent peu de la courbe moyenne. Une carte de l'atmosphère en décembre fera donc apparaître un maximum très accusé au sud, alors qu'une carte de l'atmosphère en juin fera apparaître un maximum assez plat, bien reparté dans toute la zone nord. Il se produit le même phénomène à l'altitude 1 en faible activité.

C/ CONCLUSIONS

Il est difficile de tirer des conclusions simples des analyses de l'effet semi-annuel, pour deux raisons au moins : D'une part les courbes de référence (situées au centre de gravité) ne sont pas plates, dans aucune des trois analyses. D'autre part septembre - octobre joue manifestement un rôle très particulier dans l'analyse globale et l'analyse en faible activité : C'est le seul point qu'il faudrait déplacer sur les graphiques pour rendre au nuage des mois la symétrie circulaire qui lui est due; il est probable que la très forte surestimation de JACCHIA 71 en cette période en faible activité contribue à fixer ce point dans une position qui n'est pas logique.

Il semble que trois traits principaux soient à retenir :

- . surestimation aux équinoxes en faible activité, en juin en forte activité ; tempérant cet effet global, tendance à la sous estimation en été local, particulièrement forte à l'altitude 3
- . maximums plus accentués au sud qu'au nord aux altitudes 1 et 2.

VI PRINCIPALES CONCLUSIONS

On peut distinguer deux types d'enseignements de l'étude de ces courbes.

Les premiers sont d'ordre méthodologique : Certains paramètres, comme l'altitude, peuvent très bien donner l'impression de ne pas structurer l'évolution de la variable à expliquer : Une courbe $f=f(\text{altitude})$ est en général assez plate. Elle joue pourtant un rôle prédominant, dès qu'elle est associée à un autre paramètre : latitude ou date notamment. C'est pourquoi il nous semble très important de mener des études où plusieurs paramètres sont susceptibles de variation simultanée, les seules bornes que le statisticien doit se fixer lui étant imposées par la quantité de données dont il dispose.

Les seconds sont d'ordre géophysique : Nous avons systématiquement rencontré dans nos études des séparations très nettes en fonction de l'intensité de l'activité solaire et en fonction de l'altitude. Il convient donc de ne pas extrapoler des résultats d'une période sur l'autre, d'une altitude à l'autre. En particulier, il faut bien admettre que l'atmosphère à 600 km n'est pas "homothétique" de l'atmosphère à 300 km ; Le rôle des pôles en particulier n'est pas le même pour les différents constituants.

De même, selon que l'activité solaire est faible ou forte, il se passe des choses différentes le jour et la nuit, aux équinoxes, en juin, etc..

Les points qui nous semblent essentiels, et acquis avec certitude au cours de cette étude, sont les suivants :

. JACCHIA 71 sous estime à l'équateur et surestime aux pôles, aux alti-

tudes basses. Il surestime au contraire à l'équateur et sous estime aux pôles aux altitudes hautes. Cet effet est d'autant plus net que l'activité solaire est forte. Enfin, sud et nord ne semblent pas jouer des rôles absolument symétriques. Effets plus ou moins prononcés au sud en hiver qu'au nord en été.

. JACCHIA 71 sous estime la nuit en faible activité, le jour en forte activité, et cet effet est d'autant plus important que l'altitude correspondante est élevée.

. JACCHIA 71 surestime aux équinoxes en faible activité, en juin en forte activité.

V - MODELISATION ET CARTOGRAPHIE

- I. PRINCIPE DE LA METHODE
- II. MISE EN OEUVRE ET DIFFICULTES RENCONTREES
- III. CARTOGRAPHIE DE L'ATMOSPHERE

PRINCIPE DE LA METHODE

L'étude précédente a mis en évidence un certain nombre de phénomènes dont certains étaient bien connus (effet latitude aux basses altitudes par exemple, effet semi-annuel), et d'autres plus mal connues (inversion de l'effet jour-nuit avec l'activité solaire, effet latitude inversé au-dessus de 500 km). Il devient utile de pouvoir visualiser ces effets, non seulement au niveau d'une Analyse descriptive, mais aussi en cartographiant les densités "réelles" (c'est à dire les densités JACCHIA 71 corrigées par le coefficient f)

Il faut alors fixer un certain nombre de conditions : une carte est une représentation des densités en fonction de la latitude et de l'heure locale. Il convient donc de fixer les autres paramètres de position - altitude et date - et les paramètres d'activité solaire.

Mais on ne trouve évidemment pas dans le fichier un nombre de points suffisant répondant à des conditions précises concernant les 5 paramètres "exogènes" permettant de dresser telle ou telle carte... on est donc conduit à modéliser les déformations, c'est à dire à se définir une fonction f associant à tout 7-uplet de paramètres une valeur du coefficient f.

Un modèle, quel qu'il soit, repose sur certaines hypothèses. Nous sommes en effet dans l'ignorance totale de la forme de la fonction f : S'il est relativement facile d'estimer un développement approché d'une fonction f à un paramètre sur un ensemble de polynômes ou de fonctions trigonométriques, il n'en est évidemment pas de même quand on est en présence de 7 variables dont les interactions sont relativement mal connues.

Le modèle le plus simple - correspondant à l'hypothèse la plus forte - est issu de l'hypothèse suivante : les paramètres n'ont pas d'effet structuré sur le coefficient f. La meilleure estimation - indépendante de tous les paramètres - est la moyenne de f dans tout le fichier. Nous ne retiendrons évidemment pas cette hypothèse manifestement mise en défaut par les études qui ont précédé.

Une deuxième hypothèse pourrait être la suivante : les déformations constatées restent d'amplitude faible (au plus 15 à 20%). Aussi peut on assimiler la fonction f à son développement au premier ordre $f(x_1, x_2, \dots, x_7) \approx \bar{f} + \sum_i a_i (x_i - m_i)$. Cette hypothèse reste trop grossière pour un certain nombre de raisons : d'une part certaines variables (heure locale, date) sont périodiques. Il conviendrait donc d'utiliser, pour celles-ci, des fonctions trigonométriques, si on ne veut pas créer de discontinuité entre 23h59 et 0h1mm. Plus encore,

les effets mis en évidence n'ont pas une structure linéaire, et il est absurde de vouloir assimiler une parabole à une droite,...

Enfin, et nous pensons que c'est l'argument déterminant, un tel modèle ne rend pas compte des influences couplées des paramètres. Un changement d'altitude ne changera pas la forme de la carte de f , le changement étant réduit à une simple translation de valeurs : La structure des inversions (effet latitude, effet jour-nuit) ne pourra pas apparaître sur un tel modèle.

Nous avons adopté une hypothèse plus faible dès le départ, hypothèse qu'on peut assimiler à une hypothèse de continuité, et qui peut s'énoncer ainsi : l'évolution de f est bien représentée localement par un développement du type :

$$f(x_1, \dots, x_7) = \bar{f}_v + \sum a_i (x_i - \bar{x}_i)$$

$\bar{f}_v$ étant la moyenne de f dans le voisinage considéré, et $\bar{x}_i$ les moyennes de chacun des paramètres.

Encore faut-il pouvoir définir des voisinages opérationnels. L'étude descriptive nous sera alors d'un grand secours. Elle nous a appris essentiellement deux traits fondamentaux :

- * il faut distinguer nettement les périodes de minimum et de maximum solaire du cycle de 11 ans.
- * il faut distinguer nettement au minimum 4 zones d'altitude.
- * Des distinctions sont d'autre part imposées : il convient de générer des cartes aux équinoxes et aux solstices; enfin, il faut au minimum distinguer 3 tranches de latitude (2 pôles et une zone intermédiaire) et deux tranches d'heure locale.

On est donc conduit à partitionner le fichier en sous fichiers, chacun de ces fichiers rassemblent les observations d'un voisinage. En fait, nous n'avons pas partitionner physiquement le fichier, ce qui aurait manqué de souplesse, mais nous avons créé un fichier où chaque observation est précédée d'indicateurs permettant de l'affecter à tel ou tel sous fichier selon la partition souhaitée.

Le principe de la méthode est alors le suivant

A. CHOIX DE VOISINAGES FIXES

5 paramètres entrent en jeu :

X_1 : Activité solaire

X_2 : Altitude

X_3 : Date

X_4 : Latitude

X_5 : Heure locale

Un voisinage est défini par les conditions :

$$x_{i \min} \leq X_i < x_{i \max} \quad \forall i \in [1, 5]$$

Ayant choisi deux voisinages en activité solaire, quatre en altitude et quatre en date (solstices et équinoxes), nous avons déjà une partition en 32 ensembles. A cela s'ajoutent les partitions en latitude et heure locale : on voit qu'on arrive facilement à 400 sous fichiers, comportant donc en moyenne une centaine de points.

B. REGRESSION

Sur chacun de ces fichiers, on effectue une regression linéaire de f en fonction des paramètres descriptifs.

C. ESTIMATION

On peut alors se fixer à priori les sept paramètres, chercher à quel fichier appartient le 7-uplet, appliquer la formule de regression correspondant et obtenir immédiatement la valeur correspondante de f .

D. CARTOGRAPHIE

On peut alors générer, pour chacune des 32 situations de base, une grille régulière latitude heure locale et dresser la carte correspondante.

II. MISE EN OEUVRE ET DIFFICULTES RENCONTREES

A. CONCEPTION DU PROGRAMME UTILISE

Nous avons écrit un programme de regression linéaire adapté à nos objectifs. Il remplit les fonctions suivantes :

- a) Calcul, case par case, de statistiques globales et des coefficients de regression. Impression sur bande de ceux ci. Itération sur le nombre de partitions différentes souhaitées.
- b) Contrôle, pour chaque partition de la qualité du modèle : Nombre de points améliorés, nombre de points dégradés, histogramme des nouveaux coefficients correcteurs, etc..
- e) Génération des valeurs de f aux points d'une grille régulière (tous les 10 degrés et toutes les heures). Impression sur cartes perforées de ces valeurs, prêtes à entrer dans un programme de tracé automatique de courbes de niveau.

B. FONCTION DES DIFFERENTS SOUS-PROGRAMMES

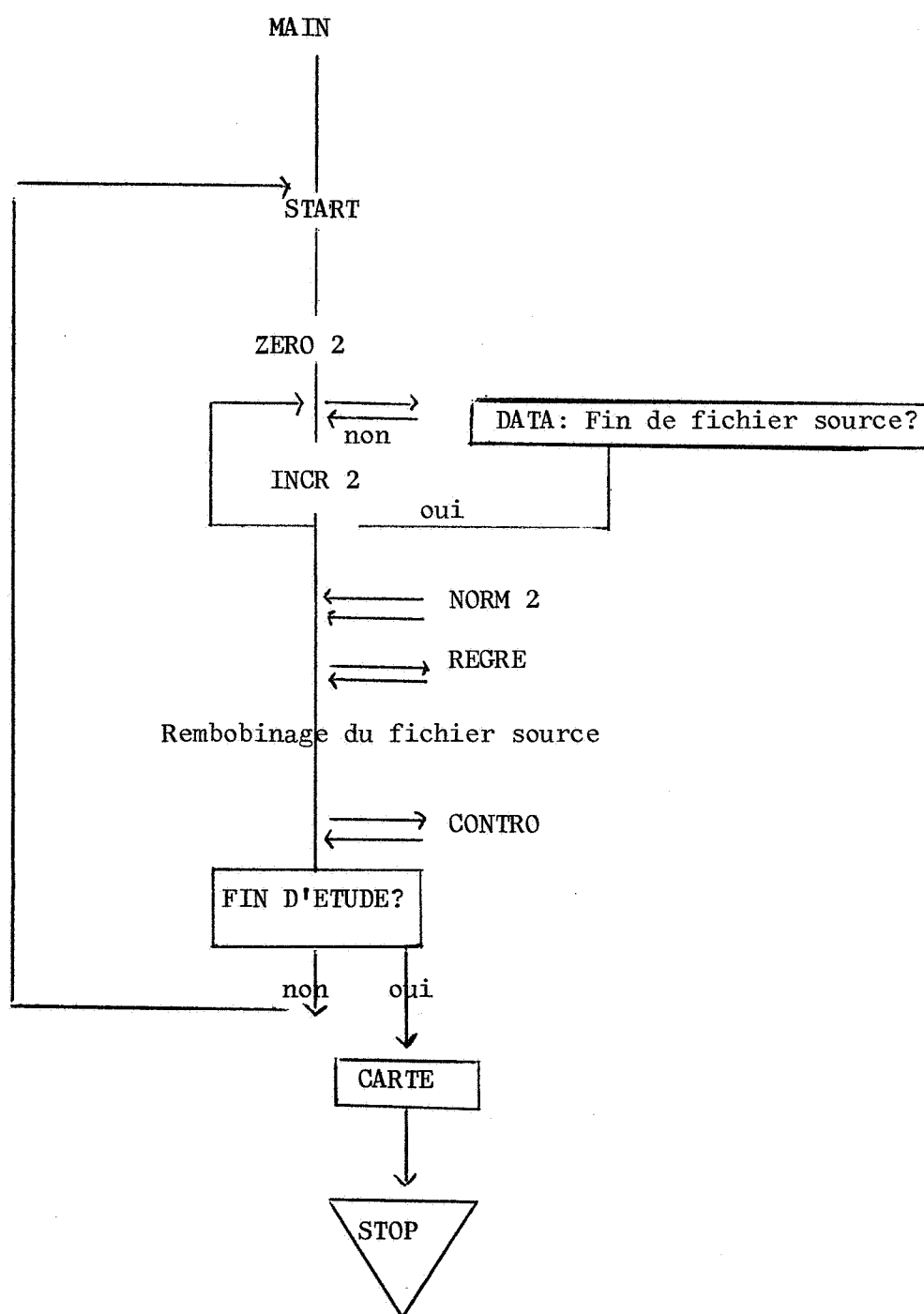
On trouvera, page suivante, un organigramme donnant les liens entre les différents sous programmes.

Précisons leurs fonctions :

MAIN	LECTURE des paramètres de l'étude appel à START
ZERO2	Initialisation à zéro des tableaux qui contiendront les moyennes, écarts types, corrélations, nombre de points des cases, à -10^{40} des tableaux contenant les maximums, à $+10^{40}$ des tableaux contenant les minimums des paramètres.
DATA	Lecture d'une observation. Affectation de celle ci à une case de l'étude.
INCR2	Incrémentation des tableaux définis dans ZERO2.
NORM2	Obtention des tableaux définitifs : moyennes à partir des sommes et des nombres, écarts types à partir des sommes de carrés et des nombres, corrélations à partir des covariances.
REGRE	Regression case par case. Impression des coefficients de regression sur bande.

CONTRO Test des regressions : amélioration comptable, histogrammes etc...

CARTE Obtention d'un fond de carte à maille régulière permettant le tracé automatique des courbes de niveau à l'aide du programme ISIS 73 de Michel ALBUISON (LABORATOIRE DE RESSOURCES TERRESTRES ECOLE DES MINES FONTAINEBLEAU)



C. LES DECOUPAGES UTILISES

Nous avons utilisé 3 types différents de découpage de l'espace, mais nous ne leur accorderons pas la même importance.

Découpage en 224 cases

C'est le découpage qui est le mieux adapté à la mise en évidence de l'effet en latitude.

L'activité solaire est découpée en 2, selon qu'on se trouve en forte ou faible activité : on a donc 112 cases par période d'activité.

Les latitudes sont découpées en 6 tranches :

$[-90, -58]$, $[-58, -28]$, $[-28, 0]$, $[0, 28]$, $[28, 58]$, $[58, 90]$
chaque tranche polaire est elle même découpée en 8 tranches, selon que l'altitude est inférieure ou supérieure à 530 km, et selon la saison. L'heure locale ne subit aucun découpage, mais elle n'est pas prise comme paramètre actif (au pôle, la notion d'heure locale perd son sens). Chaque tranche tempérée est découpée en 16 tranches : altitude inférieure ou supérieure à 530 km, saison, et heure locale inférieure ou supérieure à midi. L'heure locale est alors prise comme paramètre actif.

Enfin chaque tranche équatoriale est découpée en 32 parties : 4 saisons et 2 heures locales selon le principe des zones tempérées, mais 4 tranches d'altitude : de 250 à 320, de 320 à 530, de 530 à 750, de 750 à 1200 km.

Découpage en 192 cases

C'est celui qui met le mieux en évidence l'influence de l'Altitude; l'activité solaire est découpée en 2 tranches, les saisons en 4 tranches, l'altitude en 4 tranches, l'heure locale est découpée en 2 tranches (de 9 heures à 21 heures, et de 21 heures à 9 heures) et les latitudes en 3 tranches : de -900 à -450, de 450 à +450, de +450 à +900.

Découpage en 96 tranches

Les découpages précédents sont relativement bien adaptés au problème; ils laissent malheureusement un certain nombre de cases vides, et n'accordent pas une grande importance à l'heure locale. Pour rendre compte des variations de faible importance en heure locale, nous avons adopté le découpage suivant :

2 périodes d'activité, 2 altitudes (en dessous ou au dessus de 530), 3 latitudes ($[-90, -28]$, $[-28, +28]$, $[+28, +90]$) et 8 heures locales (de 0 à 24 heures par pas de trois). Bien entendu ce découpage étant surtout utilitaire, nous lui accorderons toujours une importance plus faible qu'aux deux premiers.

D. DIFFICULTES RENCONTREES

Nous laisserons de côté les difficultés d'ordre purement informatique pour nous consacrer essentiellement aux difficultés imputables à la méthode elle même. Rappelons brièvement le principe de l'estimation du coefficient f : une case étant définie, on sélectionne dans le fichier les points appartenant à cette case. Pour les deux premiers découpages, on dispose donc d'environ 150 points par case en moyenne. Il s'en faut de beaucoup cependant que le fichier soit équi-réparti, les cases vides ne sont pas rares, et il existe des cases contenant 1000 points et plus. Le nombre de case ou le nombre de points est inférieur à 50 est assez important : 20 à 30% selon les cas. Dans chaque cas, nous formons le meilleur estimateur linéaire du coefficient f en fonction des 7 paramètres déjà cités, et d'un huitième, la température exosphérique modèle JACCHIA 71, que nous avons adjoint aux sept premiers, car il nous avait semblé assez utile. Le meilleur estimateur est obtenu simplement en appliquant l'algorithme des moindres carrés. Pour contrôler la validité de ces regressions, nous avons essentiellement retenu deux critères :

1/ L'amélioration comptable

Soit f^* la valeur prédite, du coefficient, et f sa valeur réelle. Nous considérons qu'un point est amélioré si la valeur de la densité prédite par le modèle correctif est plus proche de la valeur réelle que ne l'est la densité prédite par JACCHIA 71, c'est à dire, en notant ρ_J la densité prédite par JACCHIA 71, et ρ_c la densité corrigée :

$$\left| \frac{\rho_{OBS}}{\rho_c} - 1 \right| < \left| \frac{\rho_{OBS}}{\rho_J} - 1 \right| , \text{ ou encore :}$$

$$\left| \frac{f}{f^*} - 1 \right| < |f - 1|$$

Un point est dégradé s'il n'est pas amélioré. Si N_A et N_D sont respectivement les nombres de points améliorés et dégradés, la quantité :

$$A C = 100 \frac{N_A - N_D}{N_A + N_D}$$

varie de -100 (si tous les points sont dégradés, modèle très mauvais) à +100 (si tous les points sont améliorés, modèle excellent) en passant par 0 (le modèle correctif n'est pas performant, mais il ne dégrade pas en moyenne la qualité du modèle primitif).

Les deux premiers découpages nous ont donné des améliorations comptables de l'ordre de 50 : en moyenne, on améliore trois fois plus de points qu'on en dégrade. Le troisième conduit à une amélioration comptable de 40, ce qui est encore fort bon.

2/ Histogramme et écart type

L'amélioration comptable ne prend pas en compte l'amplitude des améliorations, puisqu'on se contente de comptabiliser les améliorations positives et négatives, et d'établir un solde. Un utilisateur qui se servirait de notre modèle correctif aurait un coefficient correctif $f'' = \frac{f}{f^*}$. Nous avons donc comparé les histogrammes des f (erreur commise en utilisant JACCHIA 71) aux histogrammes des f'' . L'amélioration est là aussi considérable : d'une part l'histogramme des f'' est bien centré autour de 1, et il est d'autre part beaucoup plus "concentré" autour de sa valeur moyenne : environ 70% des points sont bons à 10%, l'écart type de la distribution est diminué de moitié.

En fait, ces résultats tout à fait remarquables dissimulent le fait que le modèle est tout à fait inadapté. Nous avons en effet utilisé ce modèle pour construire des cartes de f en fonction de l'heure locale et de la latitude. Nous étions donc conduits à fixer les valeurs des paramètres d'activité solaire, de date et d'altitude, et à étudier la variation de f en fonction de la latitude et de l'heure locale. Nous nous sommes alors aperçu que notre modèle correctif pouvait nous prédire une variation allant de 1 à 2 quand nous montions de 10 km en altitude, où de 10 degrés en latitude, et même nous prédire des coefficients f négatifs.... on ne pouvait pas en rester là, il fallait analyser les causes de ces absurdités, et autant que faire se peut y remédier.

E. LES CAUSES D'ERREUR

On peut distinguer au moins deux causes fondamentales, qui sont d'ailleurs assez souvent liées.

1/ Faiblesse du nombre d'observations par case

Nous avons déjà remarqué que certaines cases comportaient assez peu de points. Une regression portant sur 8 variables explicatives et 40 observations a bien peu de chances d'avoir un sens physique. La méthode des moindres carrés va accorder de l'importance à des écarts qui ne sont pas significatifs. Nous avons d'ailleurs remarqué que les résultats absurdes provenaient en général de cases dont le nombre de points était inférieur à 150.

2/ Mauvaise répartition des observations dans les cases

Nous allons prendre un exemple à un paramètre pour montrer qu'une mauvaise répartition des points peut être à l'origine de résultats absurdes. Supposons que l'on s'intéresse à l'évolution de f en fonction de la latitude à l'intérieur d'une case $[-40, +40]$. Supposons de plus que 90% des observations soient situées entre $[-5$ et $+5]$ degrés, et que dans ce petit domaine, la variation de f soit bien représentée par une droite croissant de 0,8 à 1,2. Les 10% de points restant contribueront relativement peu à la définition de la droite sur l'intervalle $[-40, +40]$, celle-ci s'obtenant grossièrement par prolongation de la première hors de son intervalle de variation. La droite a pour équation :

$$f - 1 = 0.04 \ell$$

A 40 degrés sud, le modèle prédira un f de -0,6... ce qui est absurde

En fait, cette mauvaise répartition d'un paramètre dans une case est peu fréquente. Elle est par contre systématique lorsqu'on s'intéresse à l'ensemble des paramètres. Supposons en effet, en restant très grossiers, qu'on découpe l'intervalle de variation de chaque paramètre en 3 parties, une première autour de la moyenne, les deux autres représentant les valeurs très différentes de la moyenne, par valeurs inférieures ou supérieures. Si l'on exige que des occurrences du type "tel paramètre est bas et tel autre moyen,..., et le dernier est haut", on doit disposer d'au moins 38 observations, soit 6561 observations par case.

Il existe donc par la force des choses des couplages dans chaque case : Telle case qui est censée représenter les hautes latitudes en forte activité, au printemps et à altitude basse, ne comporte peut-être que des points dont la variation est comprise entre 35 et 37. Il peut exister une corrélation nette

entre l'altitude et le flux moyen, si les données proviennent d'un satellite dont le périégée chute lentement : l'évolution de son altitude est pratiquement "reliée" à l'évolution du flux moyen, par le simple fait que les deux grandeurs varient lentement et simultanément dans le temps. Ces couplages possibles sont malheureusement dus à la méthode de collecte des données : lorsqu'on suit un satellite de jour en jour, l'altitude de son périégée, le flux moyen, la date, et bien souvent l'heure locale sont étroitement liés. Seule la multiplication du nombre de satellites permet d'éviter ce problème. Or le fichier JACCHIA ne comporte que 16 satellites, ce qui est relativement peu dans certaines zones : en effet chacun de ces satellites passe dans les zones équatoriales. Les zones polaires au contraire sont reconnues par un nombre beaucoup plus faible de satellite, et chaque satellite apporte beaucoup moins d'informations: il ne faut pas s'étonner alors que les phénomènes aux pôles soient beaucoup plus sujets à caution : la plus grande prudence doit être de mise.

F. CONSTRUCTION D'UN MODELE PLUS REALISTE

De l'Analyse qui précède, il ressort que le nombre élevé de variables prises en compte, et la faiblesse du nombre de données par cases et surtout leurs couplages mutuels sont les principaux responsables de l'inadéquation du modèle à la réalité dont il doit rendre compte. Il nous est malheureusement impossible d'augmenter le nombre total de données. Nous pourrions alors penser à réduire le nombre de cases : mais des effets importants pourraient alors être masqués. Il ne reste donc, dans l'optique de ce formalisme, qu'à réduire le nombre de variables explicatives . Nous sommes alors conduits à former le meilleur estimateur linéaire de f dans chaque case en fonction de la latitude et de l'heure locale. A première vue, on peut penser qu'il s'agit d'un appauvrissement considérable du modèle. En fait, nous ne le pensons pas pour une série de paramètres : la date, l'altitude, le flux moyen notamment. Il est en effet quelque peu illusoire de vouloir prédire une évolution d'un coefficient correcteur f (qui reste en général compris entre 0,7 et 1,3) sur une différence de quelques jours, ou de quelques kilomètres.

Il est raisonnable de penser au contraire que les zones choisies en nous laissant guider par l'Analyse des Données préalable présentent une certaine homogénéité. On a vu par ailleurs qu'il est assez illusoire de vouloir séparer l'influence de tel ou tel paramètre, le nombre des données ne le permettant pas,

et la structure des données nous ayant montré que certains paramètres n'agissent pas du tout de façon indépendante.

Nous regrettons par contre d'avoir éliminé l'Activité géomagnétique des facteurs explicatifs, car son rôle est assez important. Nous ne pouvions en fait guère faire autrement, car les effets sont surtout très sensibles en forte activité magnétique (A_p supérieure à 20). Or nous ne disposons que de peu de points dans ces zones, ce qui ne nous aurait pas permis des estimations fiables.

Les regressions obtenues en limitant les variables explicatives à la latitude et à l'heure locale ont dans leur très grande majorité un sens physique. Elles se présentent en effet sous la forme $f = f_0 + a_1 (\text{lati}) + a_2 (\text{heure})$ avec des coefficients a_1 et a_2 petits : nous avons donc atteint notre objectif initial, à savoir représenter les variations de f dans chaque case par de petites oscillations autour d'une valeur moyenne.

Il reste cependant environ 5% des cases où les coefficients a_1 et a_2 sont trop élevés (variation de f de plus de 15% pour 10 degrés de latitude, ou variation de f de plus de 10% pour deux heures. Ces cases sont celles dans lesquelles l'effectif des points est le plus faible. Nous avons alors substitué à la forme de regression obtenue l'assimilation pure et simple de f à sa valeur moyenne.

III. CARTOGRAPHIE DE L'ATMOSPHERE

A. OBTENTION DES CARTES

A partir des différents modèles de coefficient f que nous avons mis au point, nous avons généré un certain nombre de cartes de coefficient f , puis de densité. Nous génèrons systématiquement 32 cartes correspondant à :
 + 2 périodes du cycle solaire : faible et forte activité, c'est à dire flux moyen inférieur à 90 ou supérieur à 115. L'activité géomagnétique est inférieure à 20 dans tous les cas.

+ 2 équinoxes et 2 solstices

+ 4 catégories d'altitude

Pour chacune de ces conditions, nous generons une grille régulière de 17x25 points allant de -80 à +80 degrés de latitude par pas de 10 degrés, et de 0 à 24 heures par pas de 1 heure.

Les valeurs de f aux points de grille sont obtenues par l'algorithme suivant ; un point de grille étant défini, nous calculons la valeur de f prévu par chacun des trois modèles (96, 224 ou 192 cases). Nous faisons une moyenne pondérée de ces trois valeurs, en accordant un poids 1/6 au modèle à 96 cases, 1/3 au modèle à 224 cases et 1/2 au modèle à 192 cases, qui distingue le mieux les tranches d'altitude. Il arrive que le modèle crée des discontinuités à la jonction de deux cases, principalement à 9h et 21h, et aux alentours de -45 et +45 degrés, jonctions de zones du modèle à 192 cases. Nous régularisons alors le modèle par une moyenne mobile autour de ces valeurs.

Nous avons utilisé les valeurs de référence suivante :

	FAIBLE ACTIVITE	FORTE ACTIVITE
Année	1964	1969
Flux moyen	80	150
Variation de flux	0	0
Activité géomagnétique	3	7

Les altitudes de référence sont :

290, 430, 650, 900 km.

Le programme de cartographie automatique ISIS 73 est dû à Michel ALBUISSON.

Les cotes entières inscrites sur les cartes de coefficient f correspondent à 10000 x f.

Les cartes de densité sont obtenues de la manière suivante : pour chaque période d'activité, et pour chaque zone d'altitude, nous calculons la valeur minimale de la densité, soit ρ_{Min} . Nous cartographions alors $100 \cdot \rho / \rho_{\text{Min}}$, ce qui permet d'avoir des cartes beaucoup plus lisibles.

Les valeurs ρ_{Min} de référence sont les suivantes :

ALTITUDE ACTIVITE	290	430	650	900
FAIBLE	$0.571 \cdot 10^{-14}$	$0.1575 \cdot 10^{-15}$	$0.361 \cdot 10^{-17}$	$0.868 \cdot 10^{-18}$
FORTE	$0.174 \cdot 10^{-13}$	$0.100 \cdot 10^{-14}$	$0.215 \cdot 10^{-16}$	$0.217 \cdot 10^{-17}$

Il est à noter qu'on utilise la même valeur de référence pour toutes les saisons (à activité et à altitude donnée) ce qui permet d'avoir une idée de l'évolution annuelle de la densité.

B. LES CARTES OBTENUES

Nous dirons peu de choses des cartes de coefficient f. On y retrouve essentiellement les points mis en évidence dans l'étude descriptive.

- Variations plus fortes de f aux altitudes élevées qu'aux altitudes basses.
- Surestimation générale du modèle JACCHIA, sauf aux altitudes très hautes.
- Maximums équatoriaux aux altitudes basses, polaires aux altitudes hautes.
- Sous estimation à la fin de la nuit en faible activité, vers 16h en forte activité.
- Surestimation en été ou aux équinoxes selon qu'on est en activité forte ou faible.

Plus parlantes sans doute sont les cartes de ρ/ρ_{Min} . Elles représentent, dans des conditions données l'atmosphère à une homothétie près. Nous laissons bien entendu aux géophysiciens le travail d'interprétation détaillée. Néanmoins, conformément à notre démarche habituelle, nous nous permettrons de relever les points qui, à nos yeux, sont importants .

Sans habitudes ni préjugés, puisque ignorant la plupart de phénomènes qui régissent notre atmosphère, peut-être attirerons nous le regard du géophysicien vers des zones où il ne se serait pas posé.

C. REMARQUES SUR L'ASPECT DES CARTES

Une première remarque s'impose : les cartes représentent toute l'évolution d'un même paramètre: $100^* \rho/\rho_{\text{Min}}$.

Pour toutes les cartes, les niveaux de courbe vont de 10 et 10. D'une façon générale les courbes sont de plus en plus "noires" quand on passe de l'altitude 1 à l'altitude 3, puis à l'altitude 4, on observe une certaine décroissance : l'atmosphère varie beaucoup plus aux alentours de 600 km (on peut y voir des rapports de densité de 1 à 7 selon l'endroit où l'on se place) que vers 300 km (au plus un rapport de 1 à 2).

Le tableau suivant montre ce phénomène. On trouve, dans chaque case, la valeur minimale, la valeur maximale, et le quotient de ces deux valeurs.

		ALTITUDE 1	ALTITUDE 2	ALTITUDE 3	ALTITUDE 4
FAIBLE ACTIVITE	HIVER	1,40	1,30	1,50	1,40
		3,10	5,40	5,50	4,40
		2,21	4,15	3,66	3,14
	PRINTEMPS	1,30	1,60	1,60	1,30
		3,20	6,30	7,10	3,90
		2,46	3,93	4,43	3
	ETE	1,10	1,10	1,10	1,10
		2,60	4,30	6,10	4,10
		2,36	3,90	5,54	3,72
	AUTOMNE	1,40	1,30	1,20	1,50
		3,00	5,30	5,00	2,90
		2,14	4,07	4,16	1,93
FORTE ACTIVITE	HIVER	1,30	1,40	1,60	1,50
		2,30	4,10	10,50	6,80
		1,77	2,92	6,56	4,53
	PRINTEMPS	1,50	1,10	1,70	1,70
		2,70	3,10	11,90	7,90
		1,80	2,87	7,00	4,65
	ETE	1,10	1,10	1,10	1,10
		1,90	3,00	7,20	4,60
		1,72	2,72	6,54	4,18
	AUTOMNE	1,40	1,50	1,70	1,80
		2,20	4,10	12,0	7,00
		1,60	2,73	7,06	3,89

On remarque ensuite qu'à une altitude donnée, le minimum de référence est toujours obtenu en été.

Enfin, aux altitudes 1 et 2, l'amplitude de variation, pour une carte donnée, est plus forte en faible activité. Le contraire se passe aux altitudes 3 et 4 : Variations beaucoup plus importantes en forte activité.

D'une façon générale, les positions des maximums et des minimums sont plus accusés que ne le prévoit JACCHIA 71. Ce modèle suppose que le maximum de densité suit le soleil (-23° et $+23^\circ$ aux solstices, 0° aux équinoxes). Nous nous trouvons fréquemment avec un maximum aux alentours de $\pm 45^\circ$ selon la saison. Les minimums sont beaucoup plus plats en été qu'en hiver.

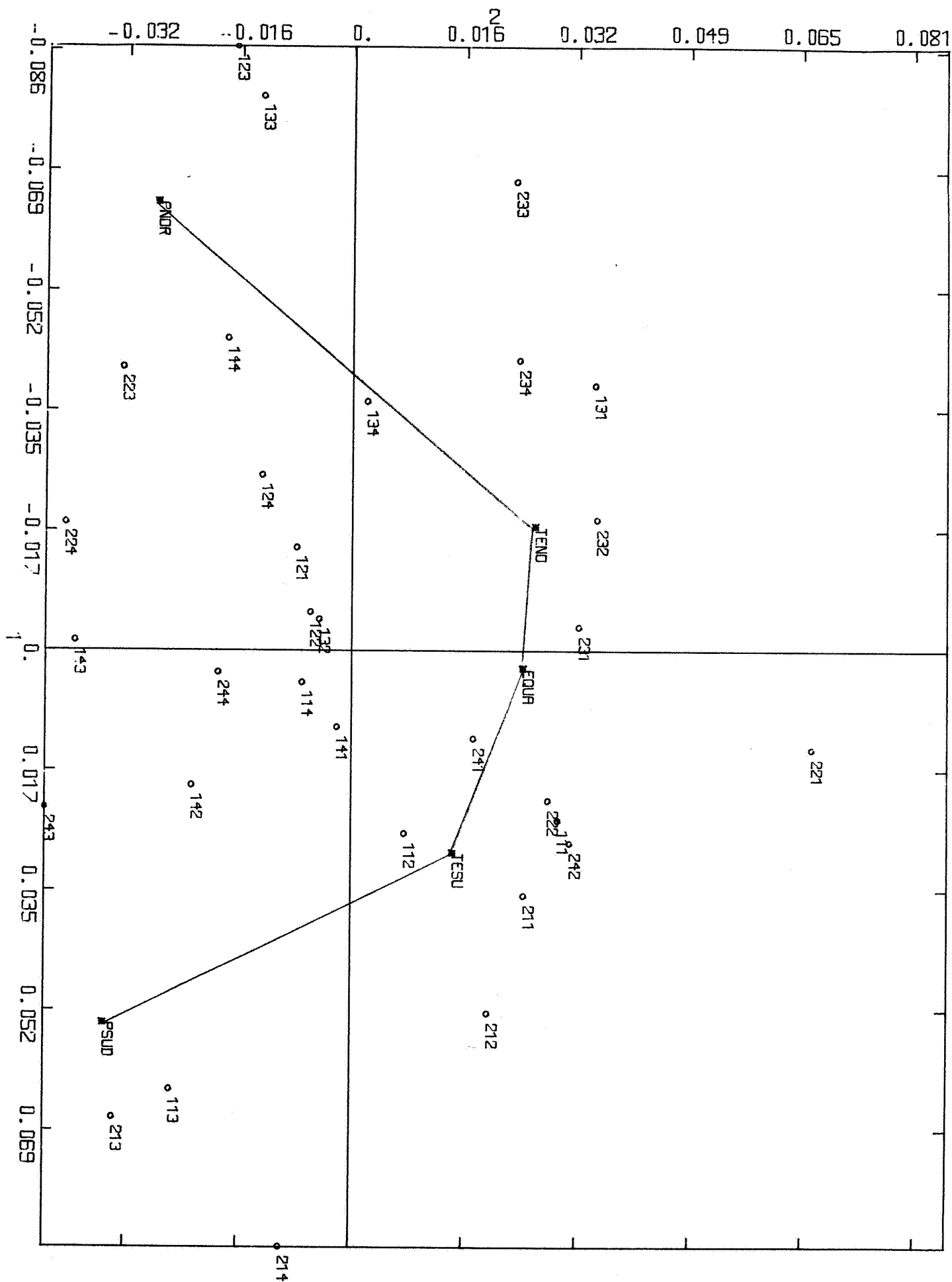
Pour essayer d'analyser ces cartes de la manière la plus objective possible, nous avons procédé à deux études complémentaires.

D'une part, pour chaque carte, nous avons cherché la position du centre des masses (en latitude). Les variations sont assez faibles : autour de -5 degrés en hiver, $+5$ en été, et nul aux équinoxes. Il semblerait donc que notre modélisation ne rende pas bien compte de l'asymétrie nord-sud que nous avons mis en évidence lors de l'étude des courbes $f = f$ (latitude)

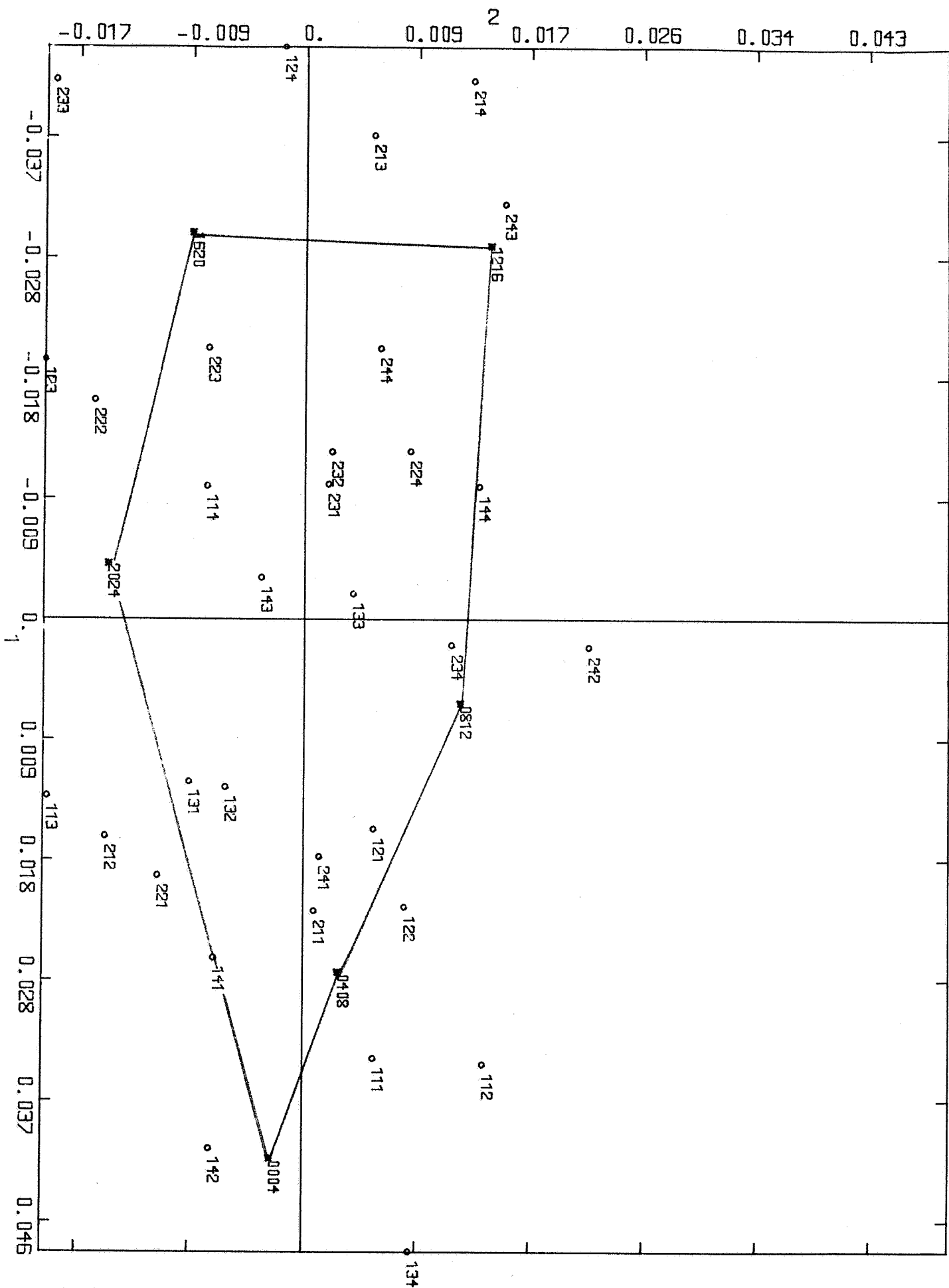
Nous avons ensuite associé à chaque carte une courbe moyenne $f = f$ (latitude) et $f = f$ (heure locale). Nous avons fait l'analyse des correspondances du jeu de courbes, selon des principes déjà exposés. On trouvera les résultats graphiques dans les deux pages suivantes.

Pour l'essentiel, la structure est conservée: mêmes pourcentages d'inertie, même "parabole" pour l'Analyse en latitude, même courbe fermée pour l'Analyse en heure. Apparaissent des changements de détail. L'anomalie à l'équateur (point Equateur situé sous les points tempéré Sud et tempéré Nord) disparaît ; dans l'Analyse des heures locales, les points 0004 et 004-08 s'opposent plus aux autres que dans l'Analyse précédente.

Les changements essentiels portent sur la position des points dans l'Analyse des latitudes : les points "HIVER" sont déplacés vers le Sud (quelles que soient les altitudes et l'activité), et les points "ETE" vers le Nord. L'automne est légèrement déplacée vers le Sud, et le printemps assez notablement vers le Nord.

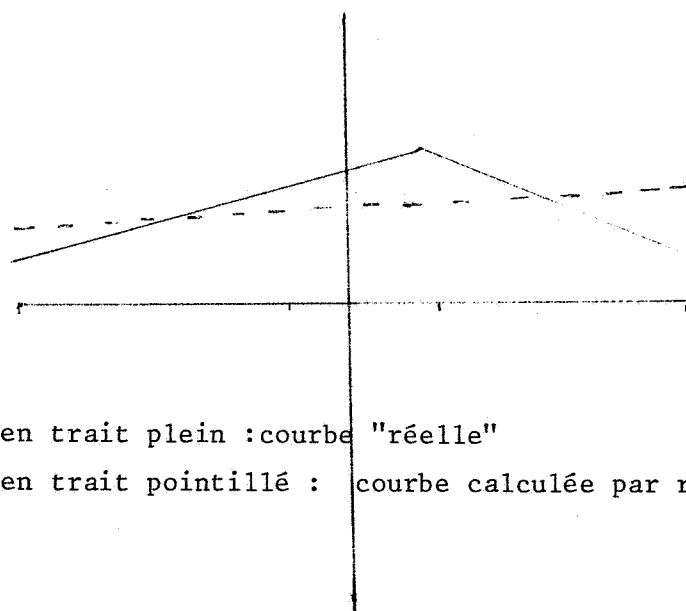


ANALYSE DES COURBES $F=f(\text{LATITUDE})$, MODELE



ANALYSE DES COURBES F-F (HEURE LOCALE) MODELE

Les déplacements aux équinoxes s'expliquent assez bien : En considérant pour Equinoxe les mois de mars et avril, on prend deux fois plus de points après l'équinoxe qu'avant. Le "PRINTEMPS" ressemble donc plus à l'été qu'à l'hiver. De même pour l'automne. Or les effets étant apparemment amplifiés aux solstices, les problèmes sont répercutés aux équinoxes. Par contre, la raison de l'amplification des effets aux solstices ne nous apparaît pas clairement. Il est possible que le découpage de l'espace qui intervient le plus (en 192 cases) en soit en partie responsable. Dans ce découpage en effet, on a une zone médiane allant de -45° à $+45^{\circ}$. Il se peut que des phénomènes avec maximum à -10° (en hiver par exemple) soient en fait assimilés, par le biais de la regression linéaire, à une évolution avec maximum à la borne de même signe (à savoir -45°), selon le schéma suivant :

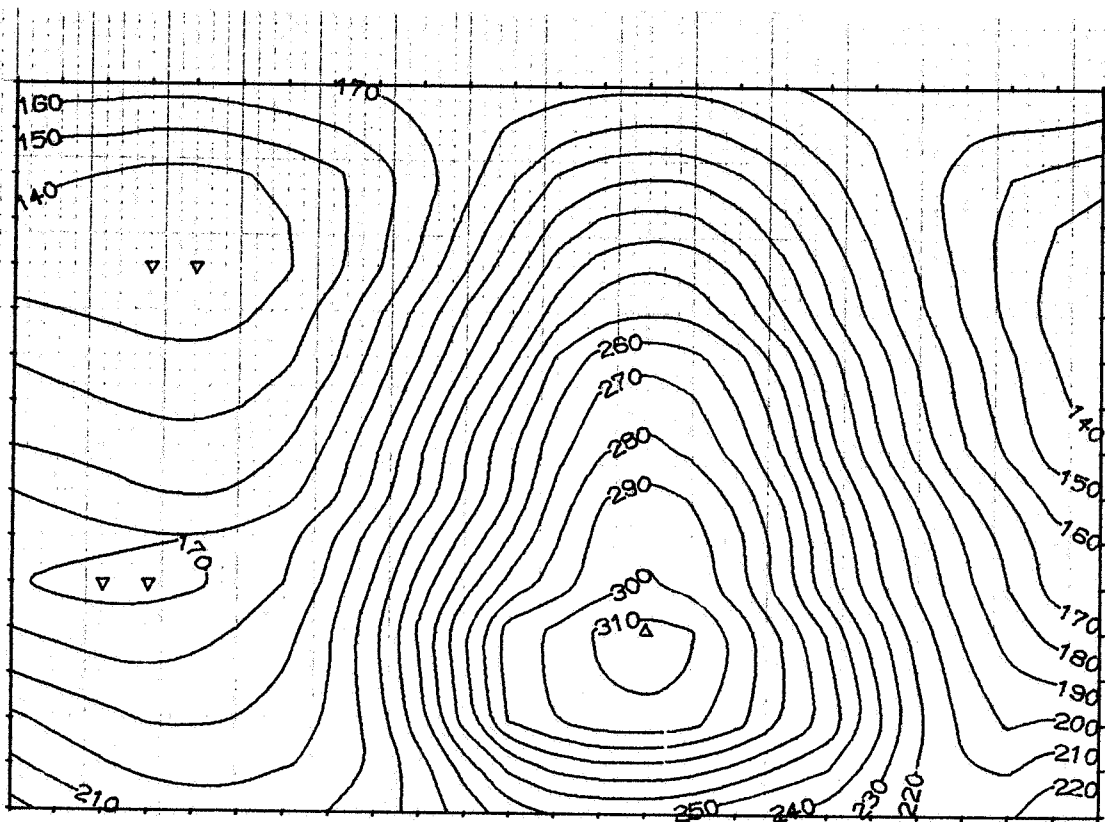


Notons enfin qu'une aberration au moins semble corrigée : Le point 132, très au sud au départ, trouve une position équatoriale après la regression.

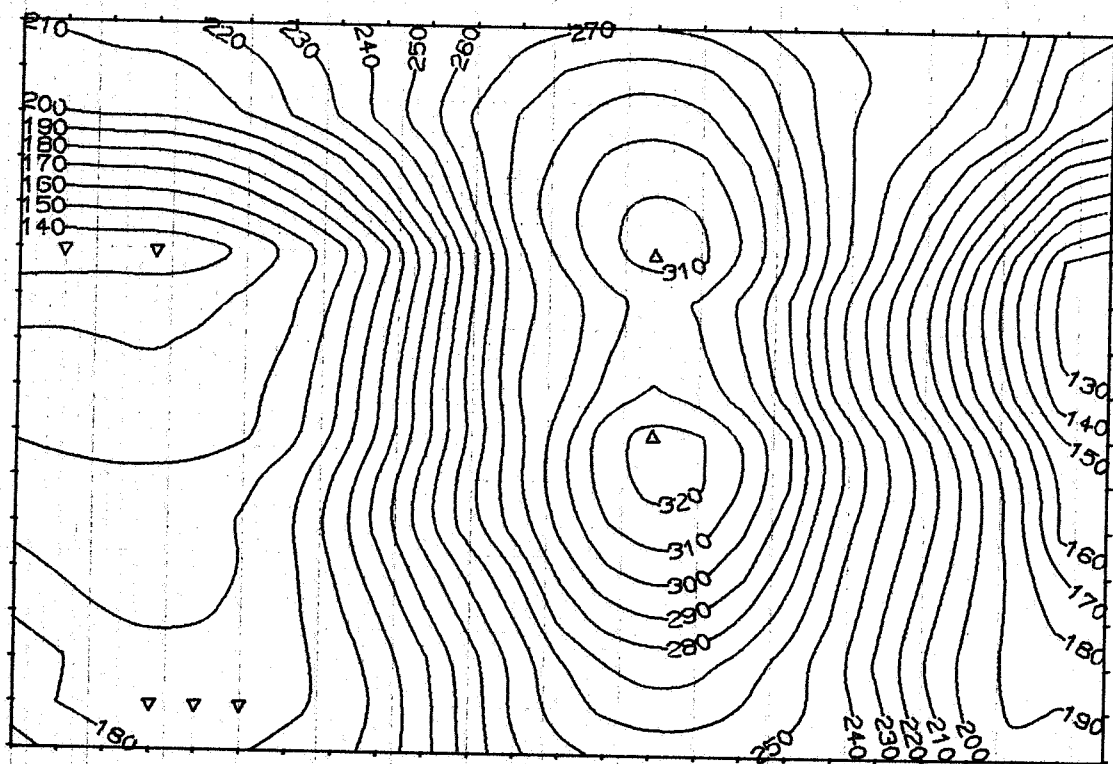
CARTOGRAPHIE DE RO/ROMIN

FAIBLE ACTIVITE (1963.1966)

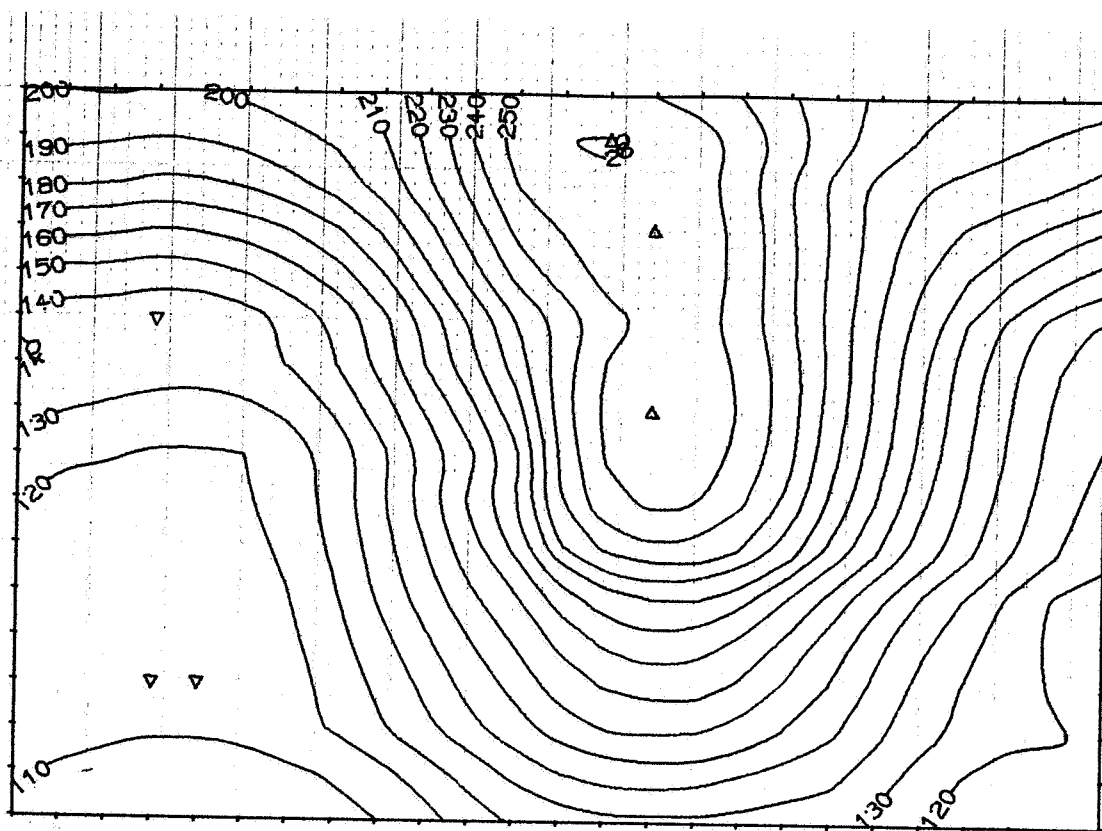
-:-:-



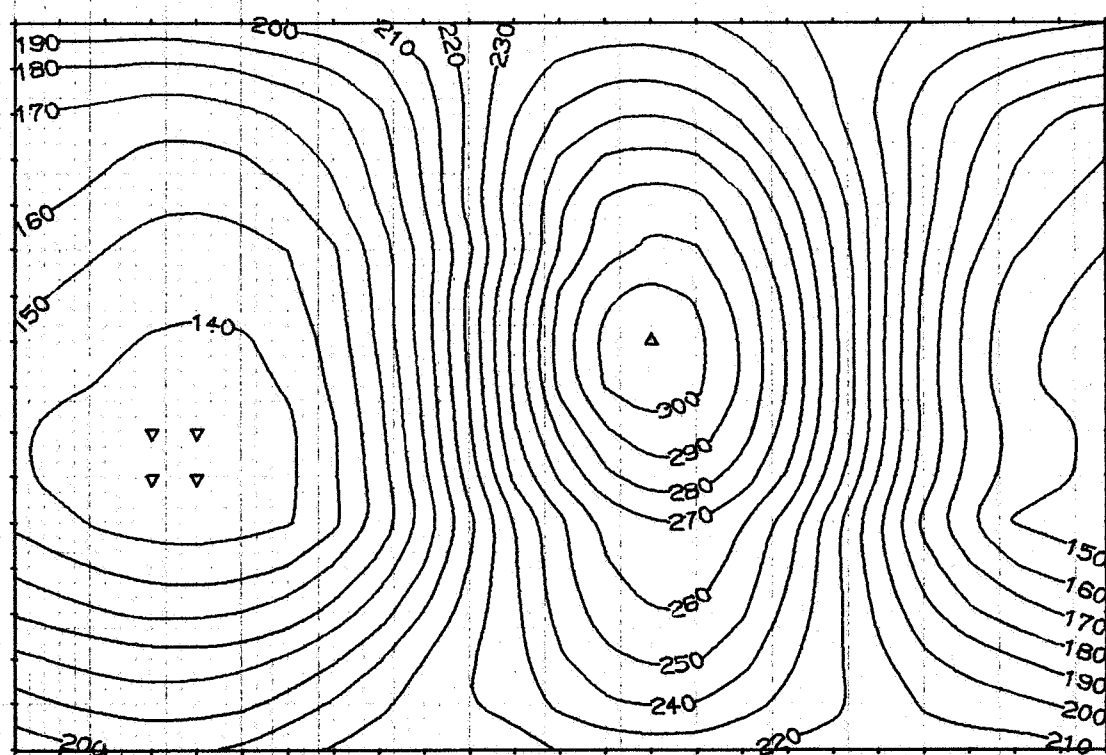
Hiver Altitude 1



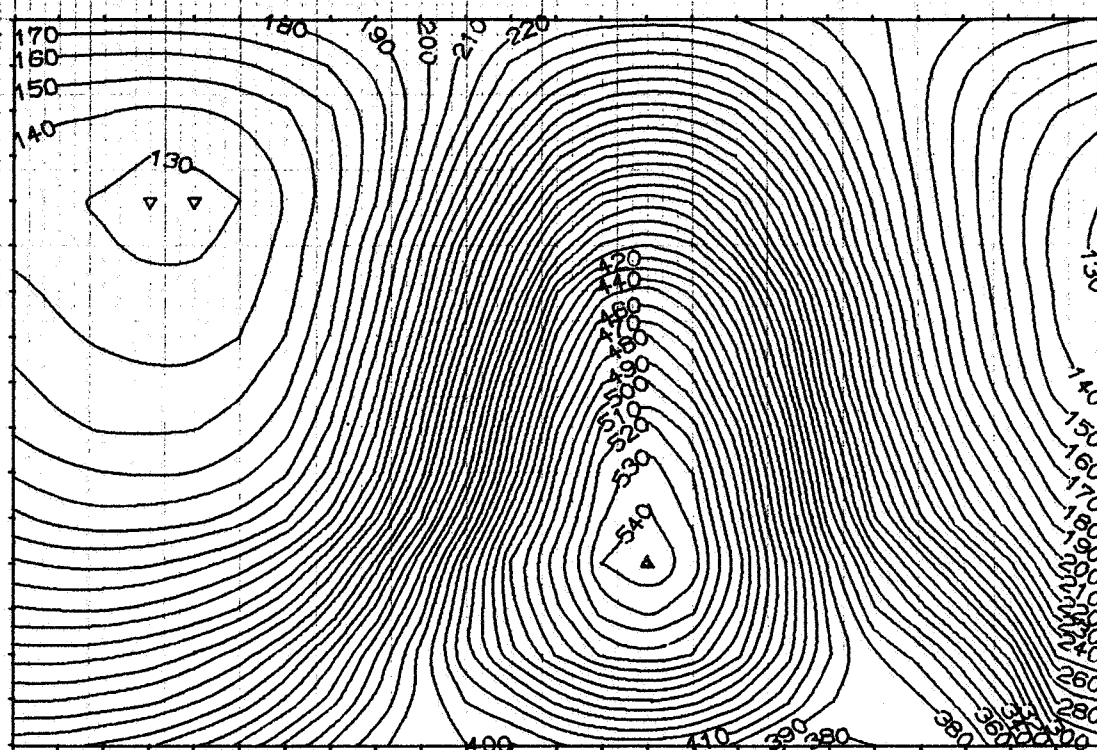
Printemps Altitude 1



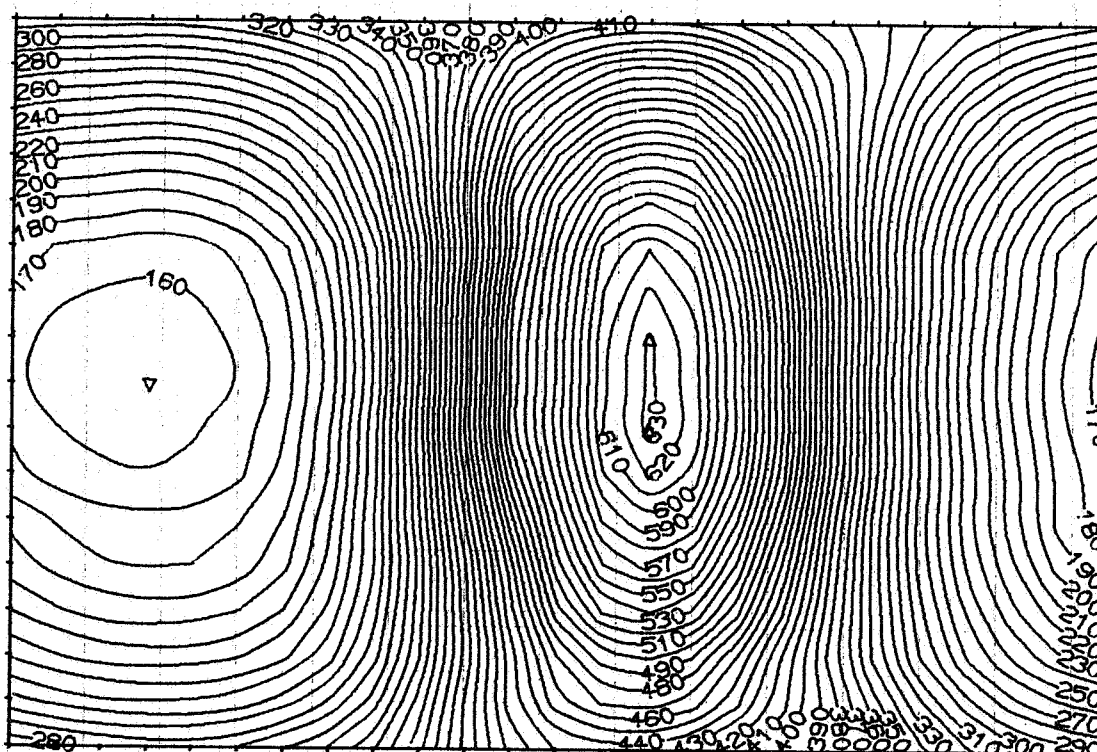
Eté Altitude 1



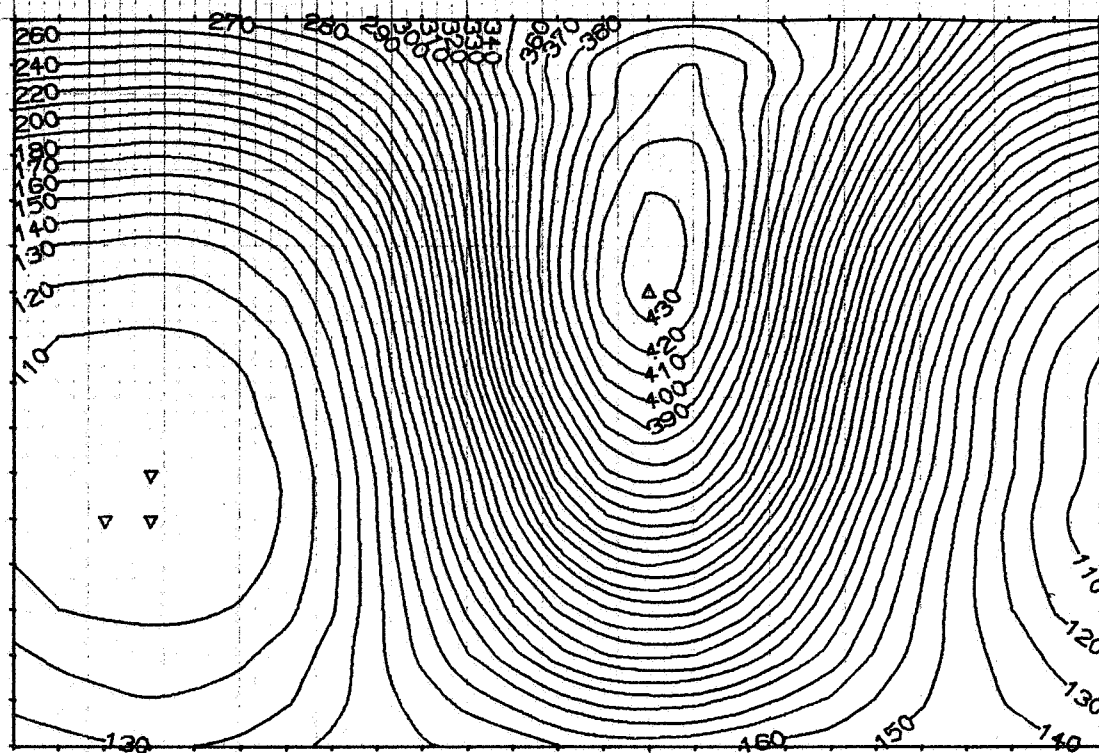
Automne Altitude 1



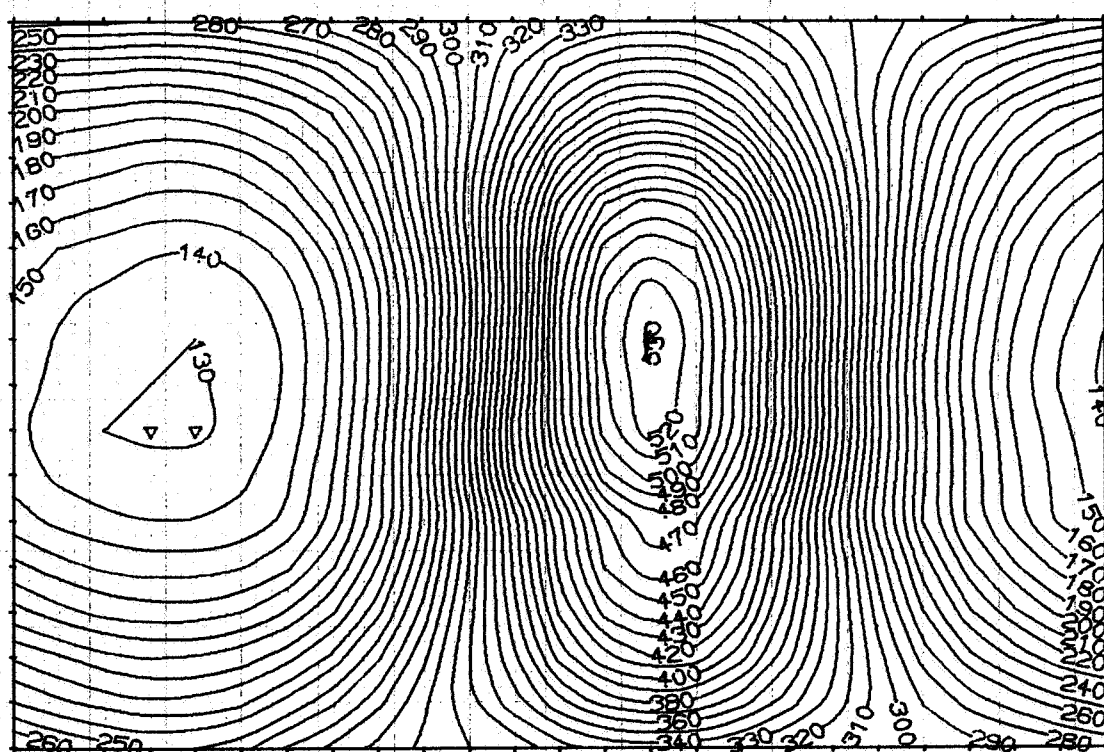
Hiver Altitude 2



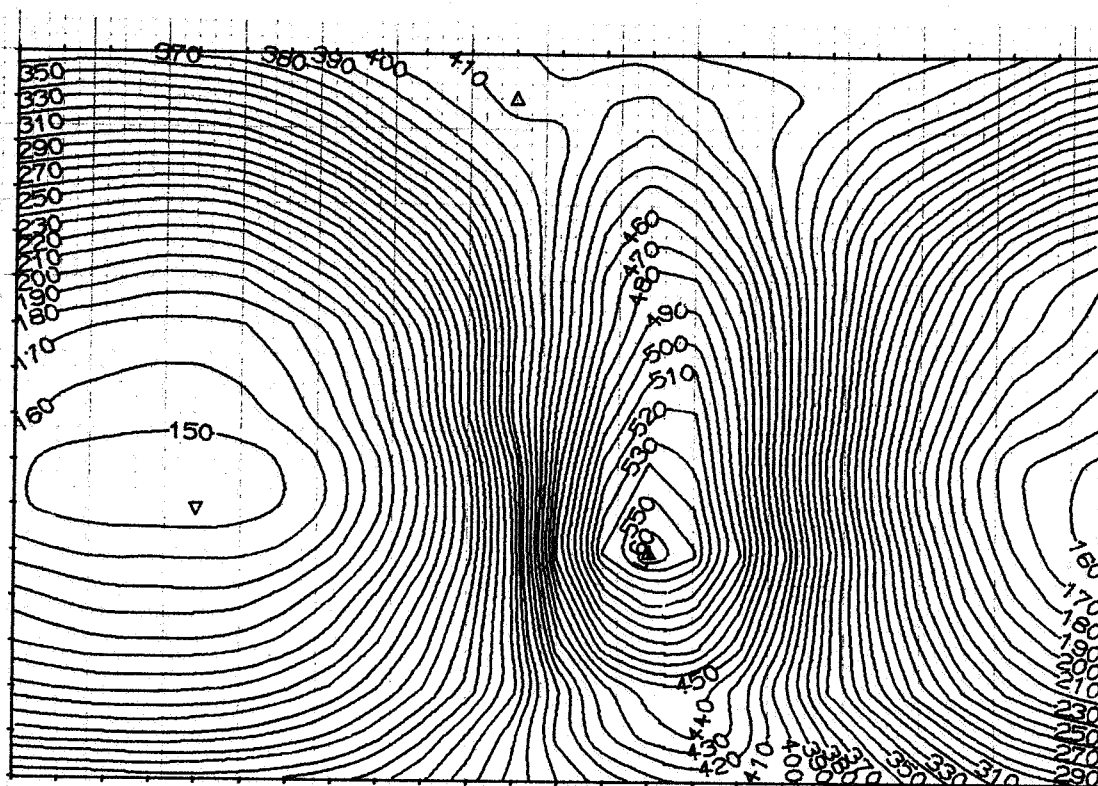
Printemps Altitude 2



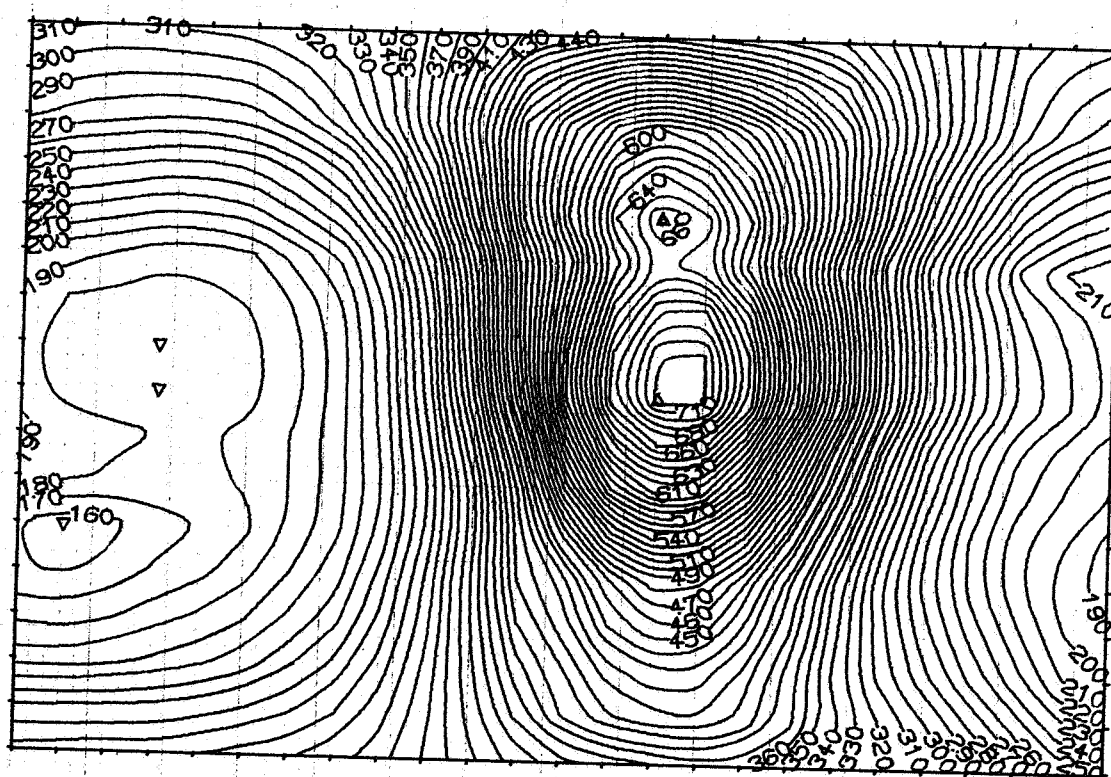
Eté Altitude 2



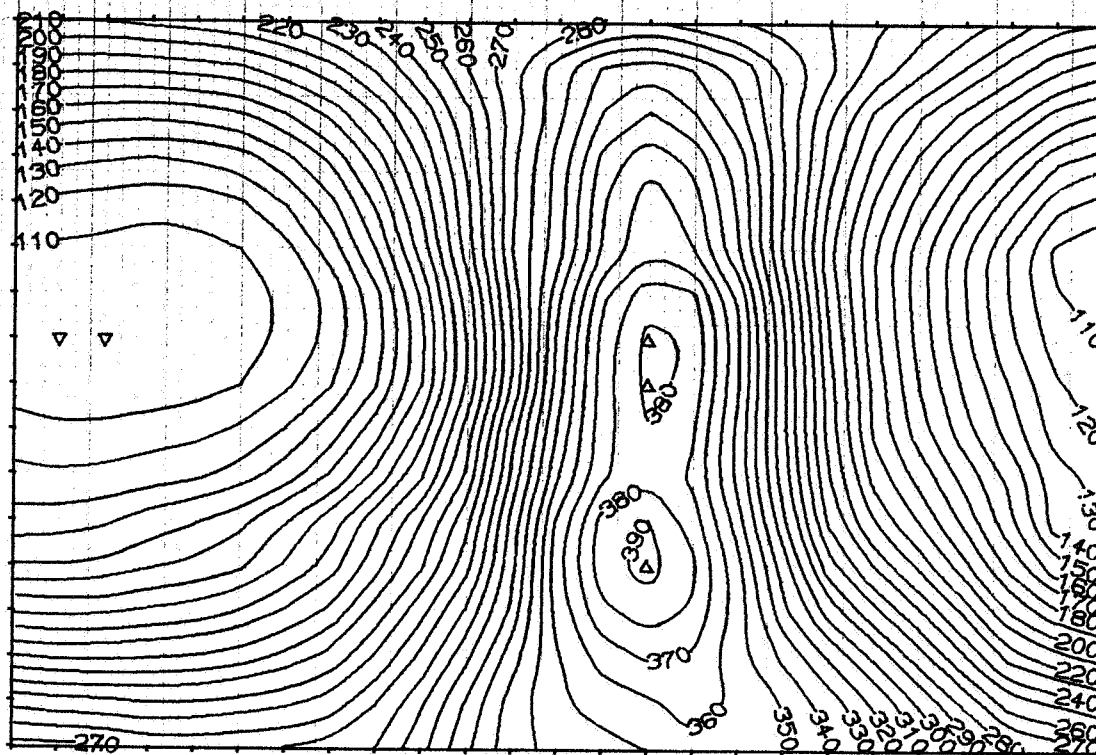
Automne Altitude 2



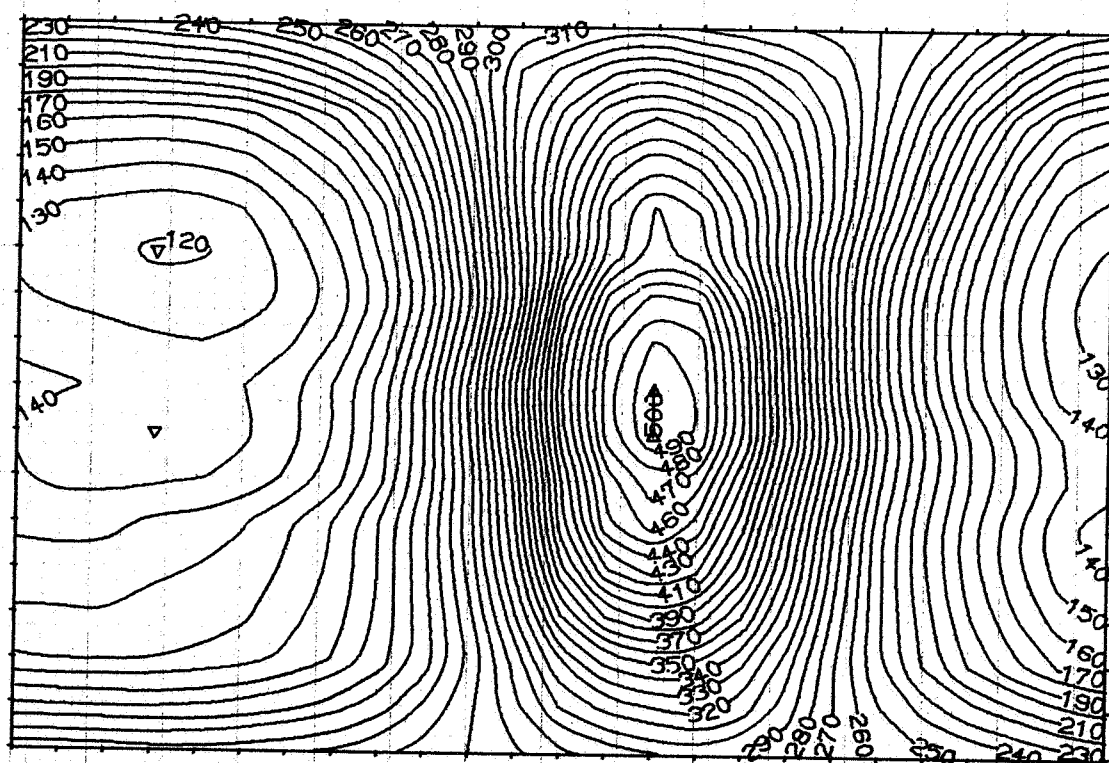
Hiver Altitude 3



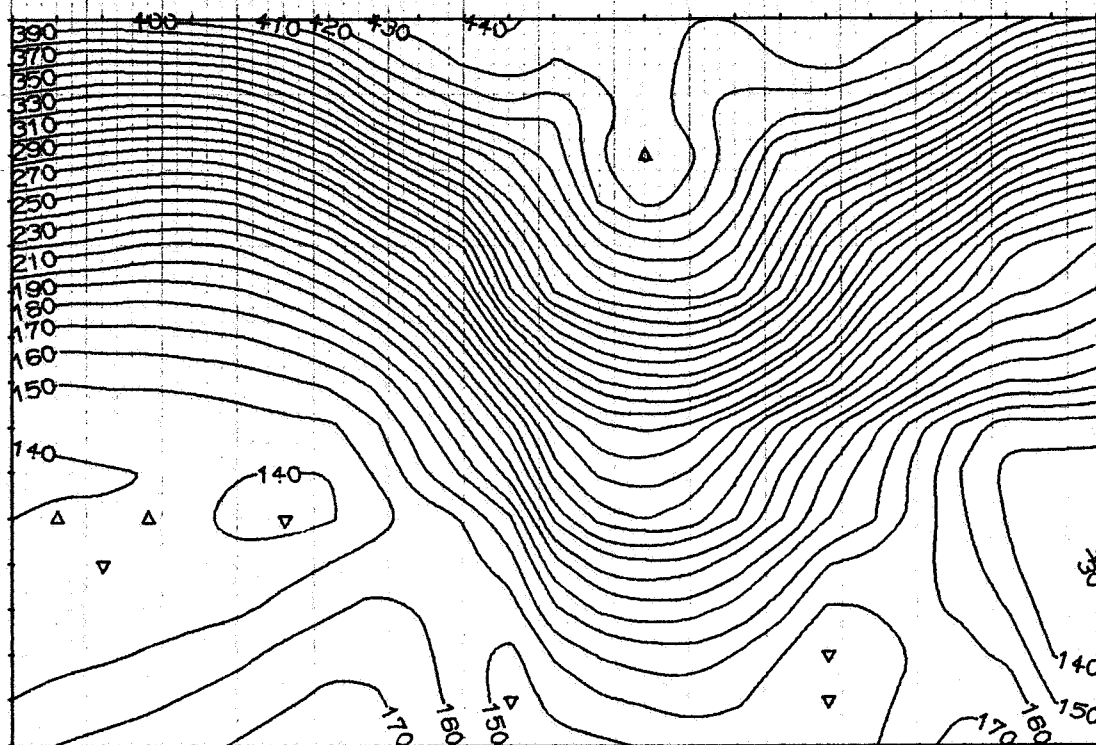
Printemps Altitude 3



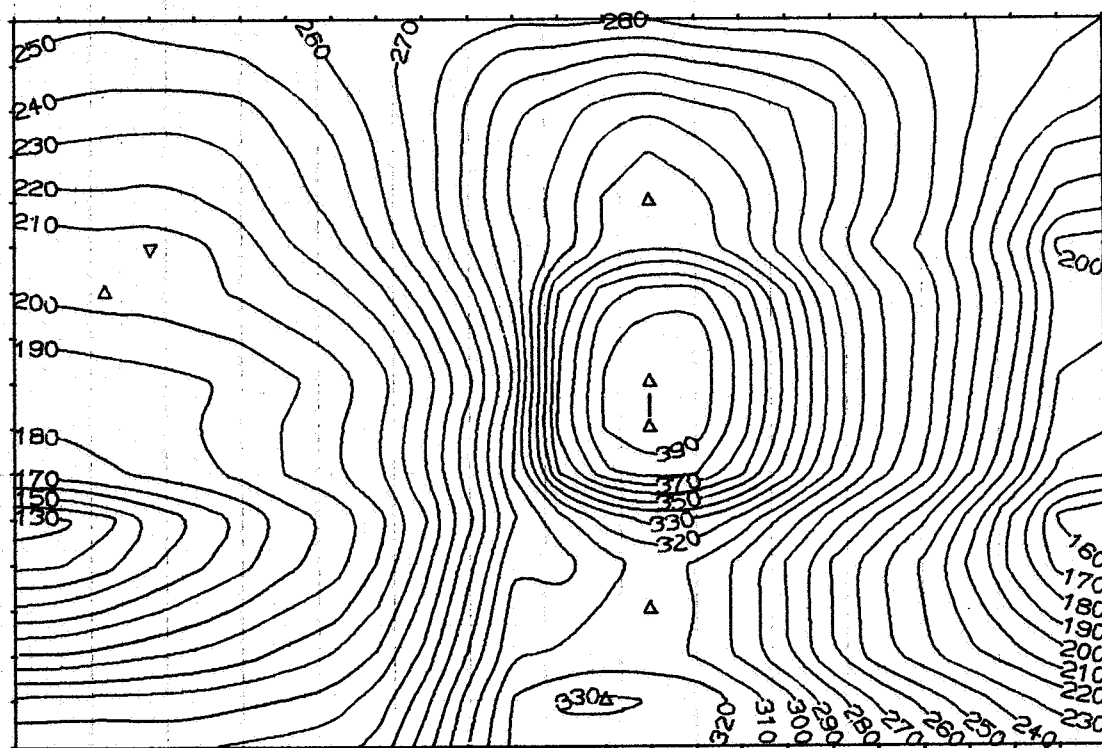
Eté Altitude 3



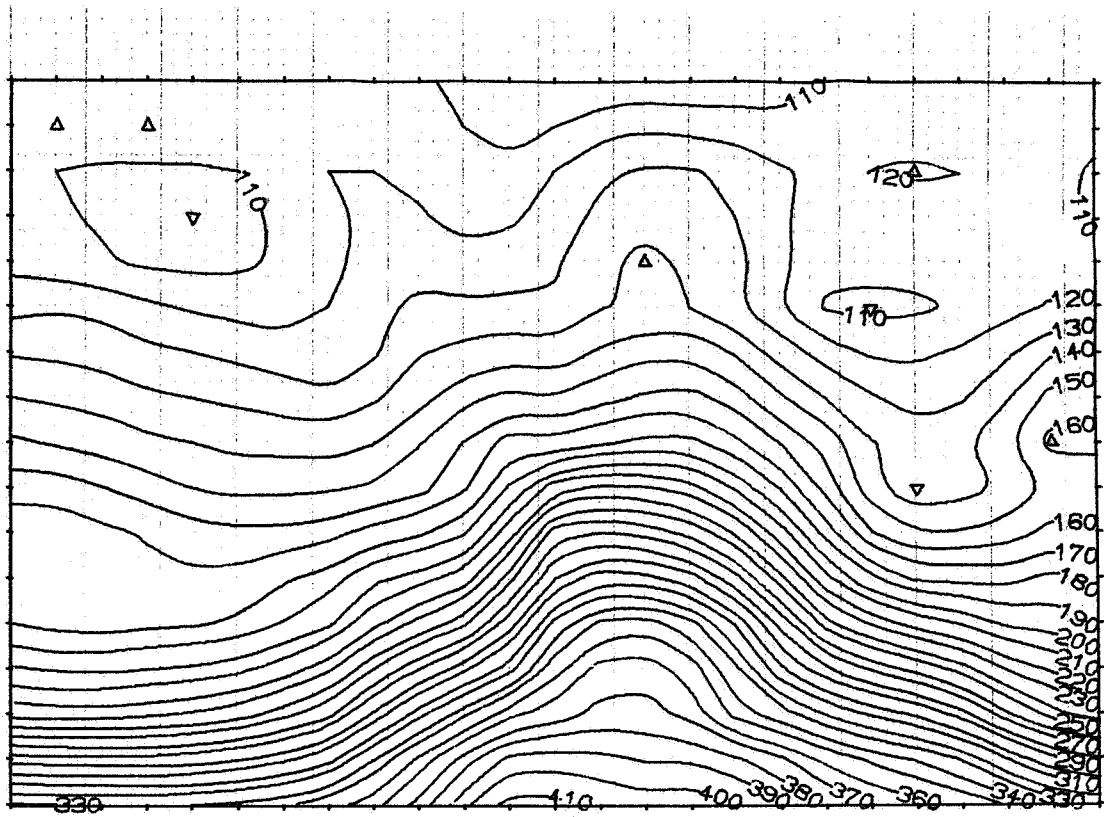
Automne Altitude 3



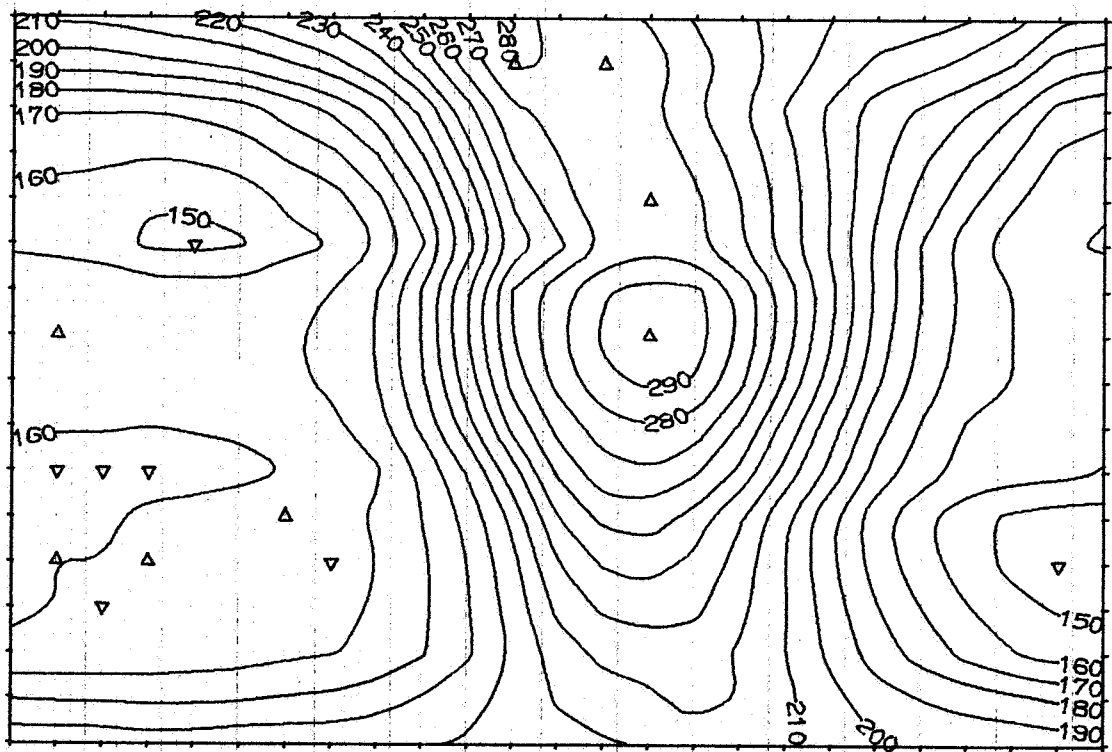
Hiver Altitude 4



Printemps Altitude 4



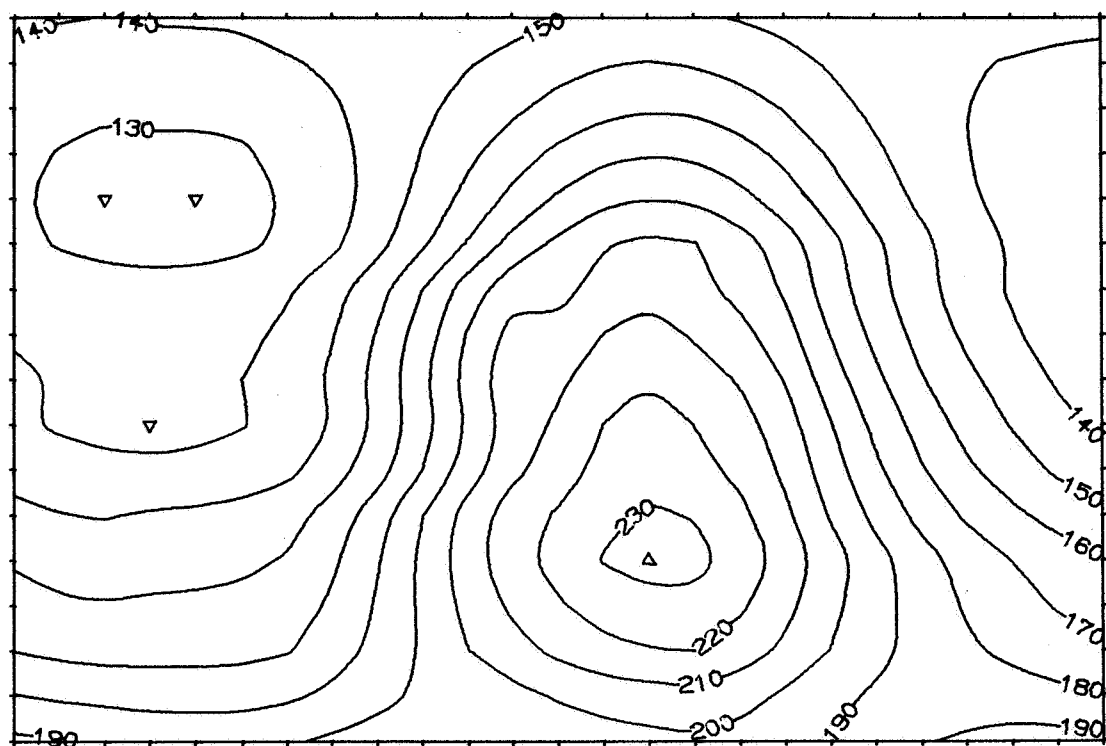
Eté Altitude 4



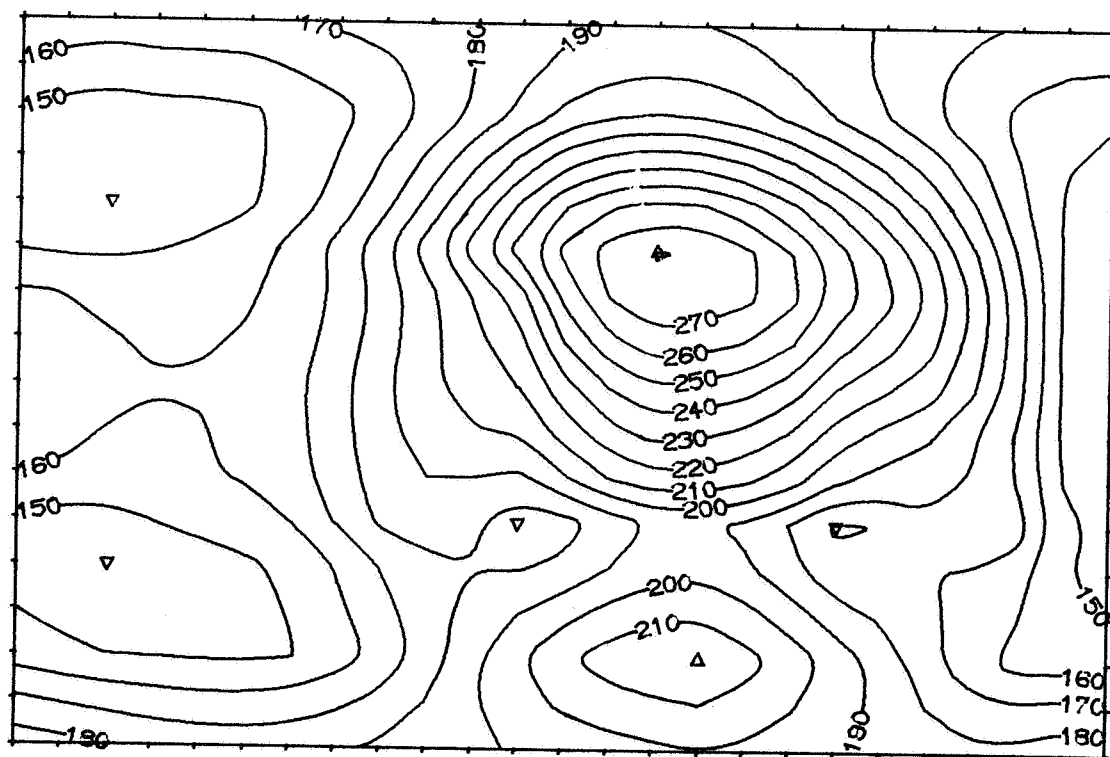
Automne Altitude 4

CARTOGRAPHIE DE RO/ROMIN
FORTE ACTIVITE (1968.1970)

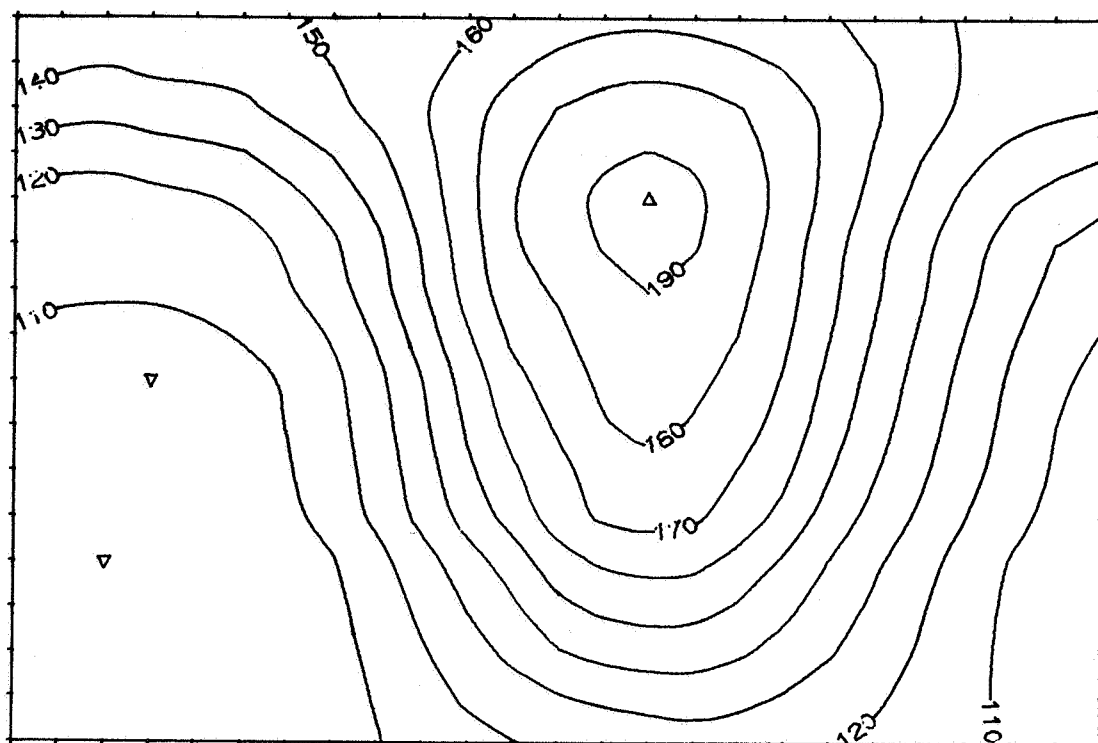
-:-:-



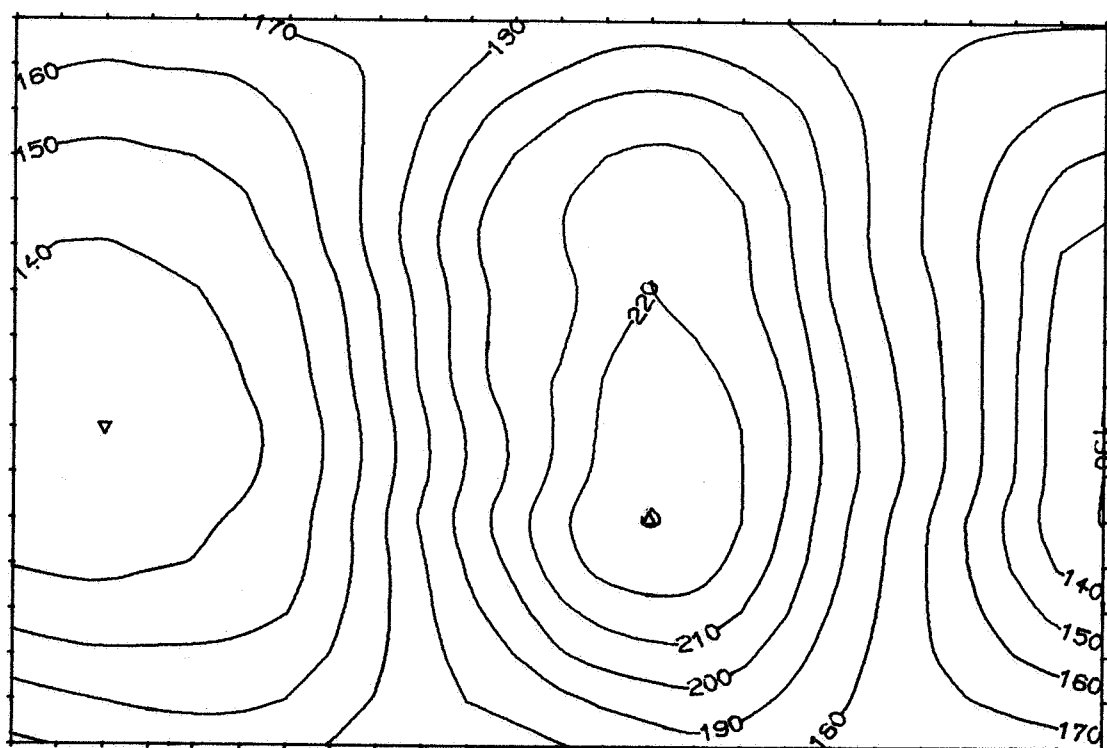
Hiver Altitude 1



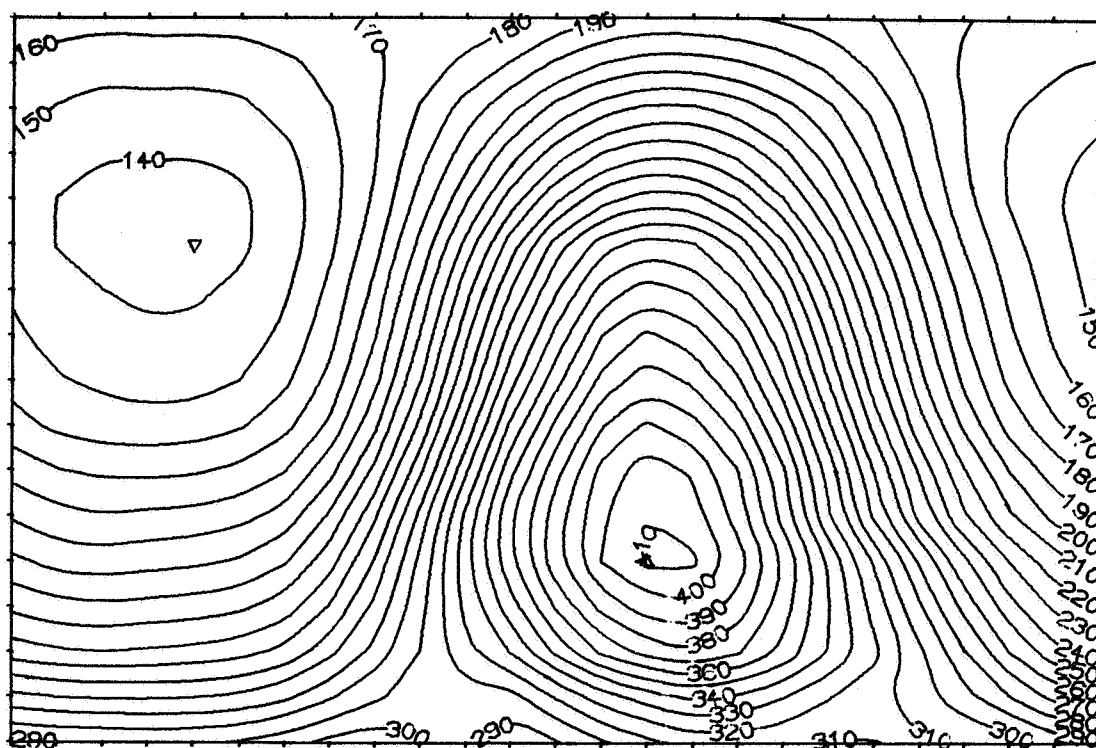
Printemps Altitude 1



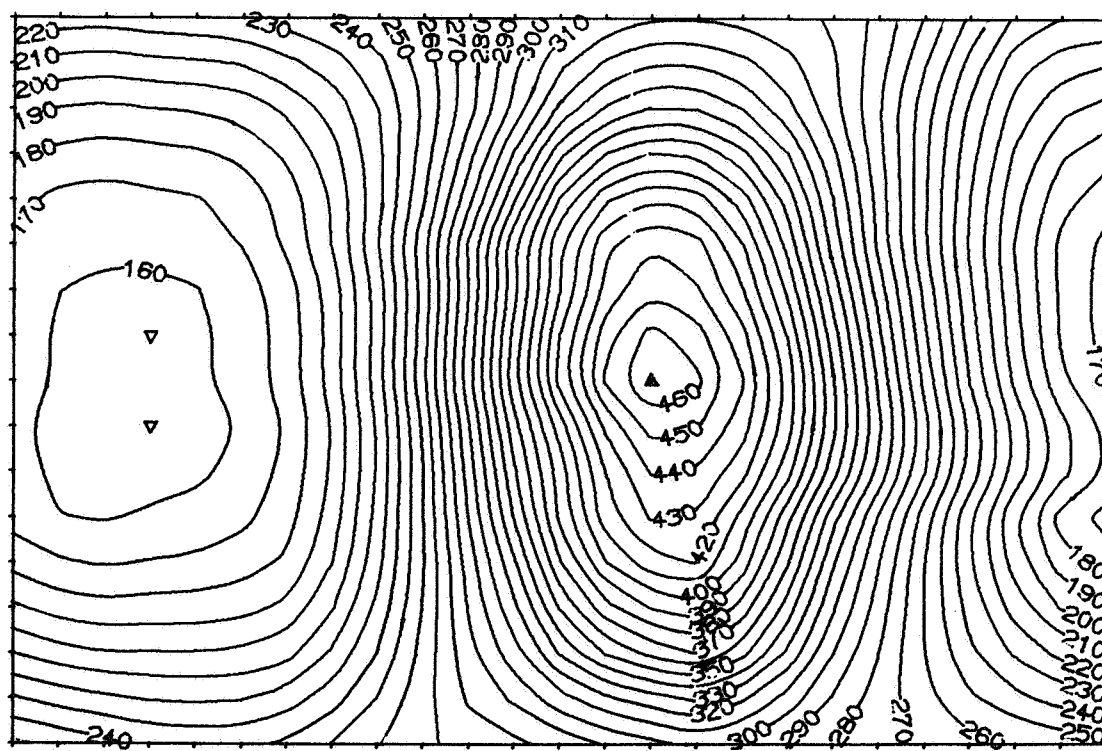
Eté Altitude 1



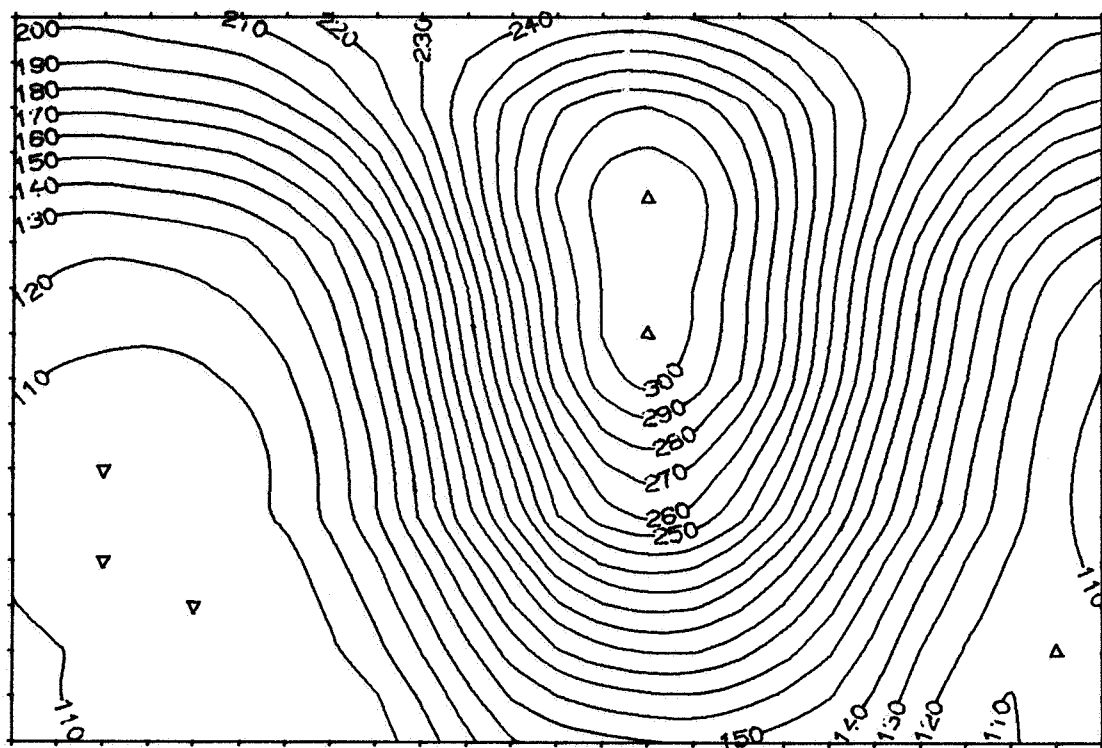
Automne Altitude 1



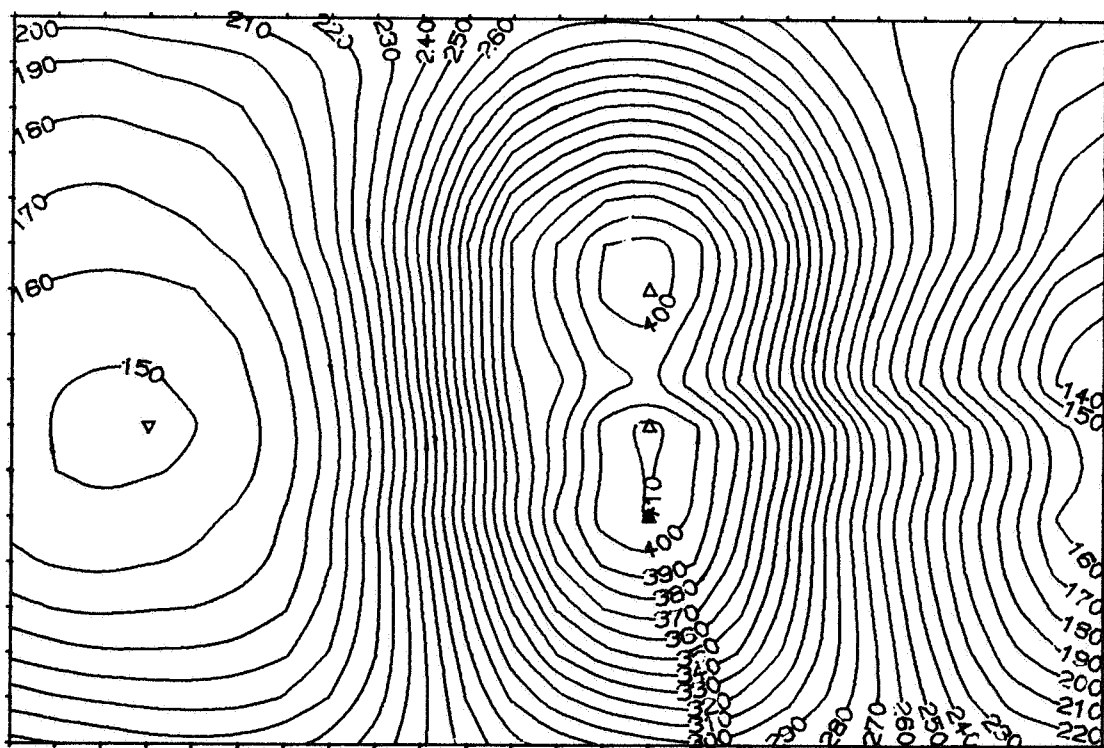
Hiver Altitude 2



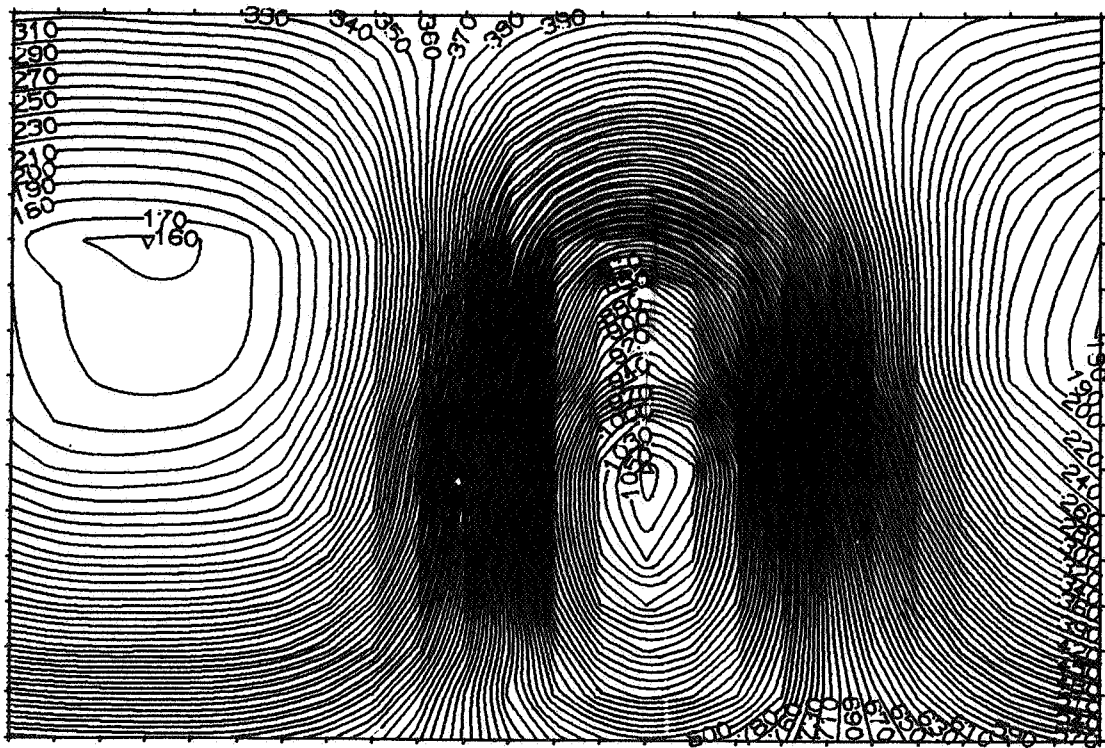
Printemps Altitude 2



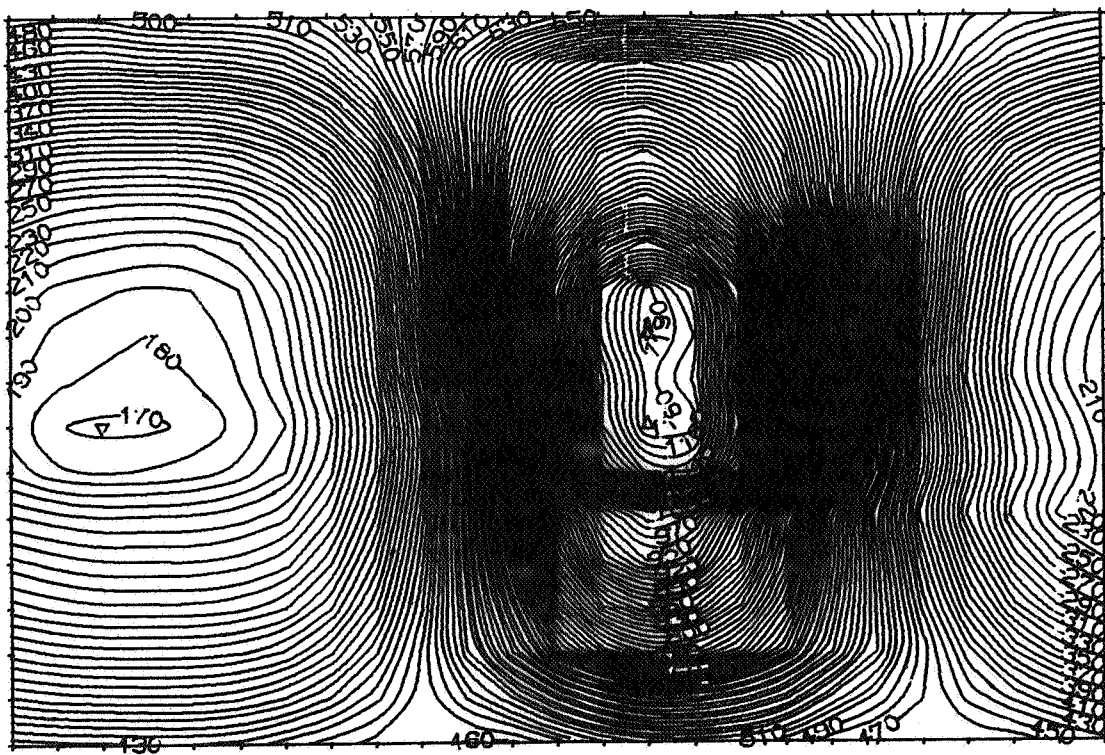
Eté Altitude 2



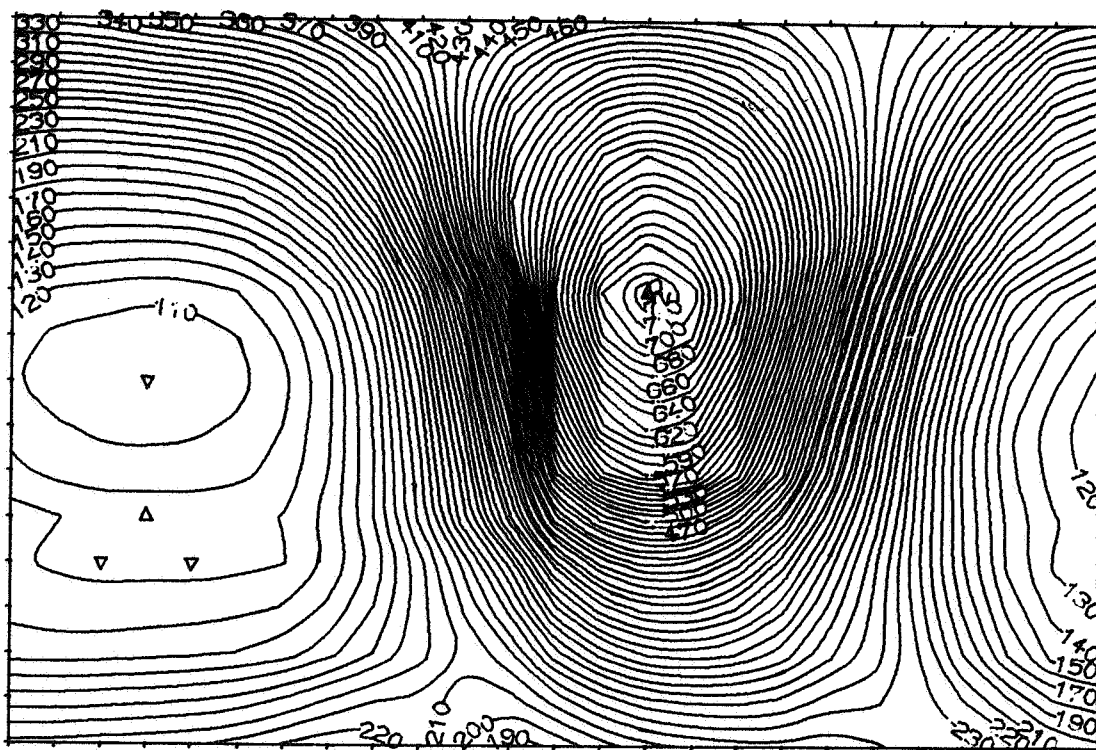
Automne Altitude 2



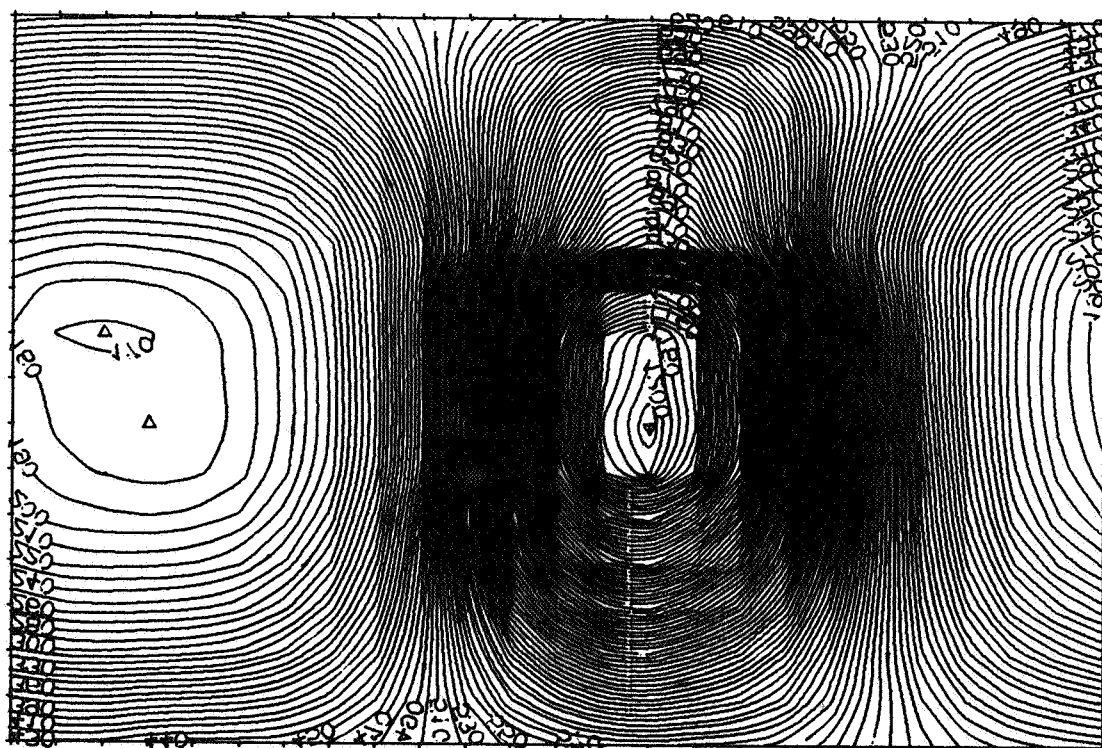
Hiver Altitude 3



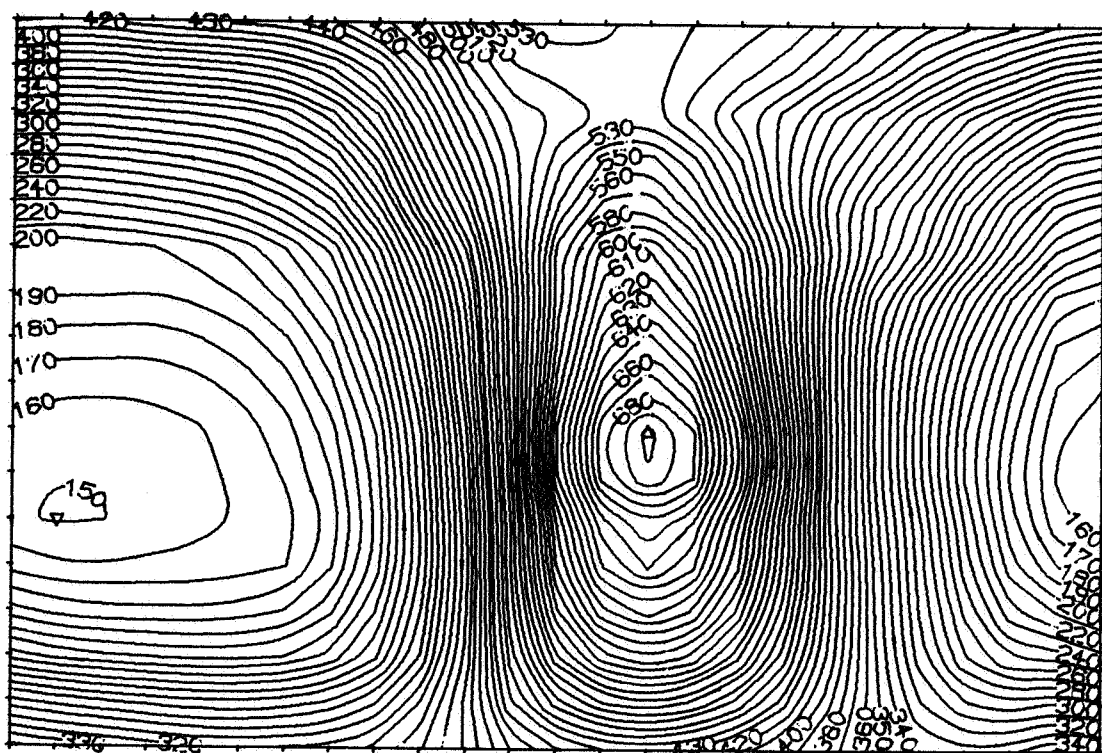
Printemps Altitude 3



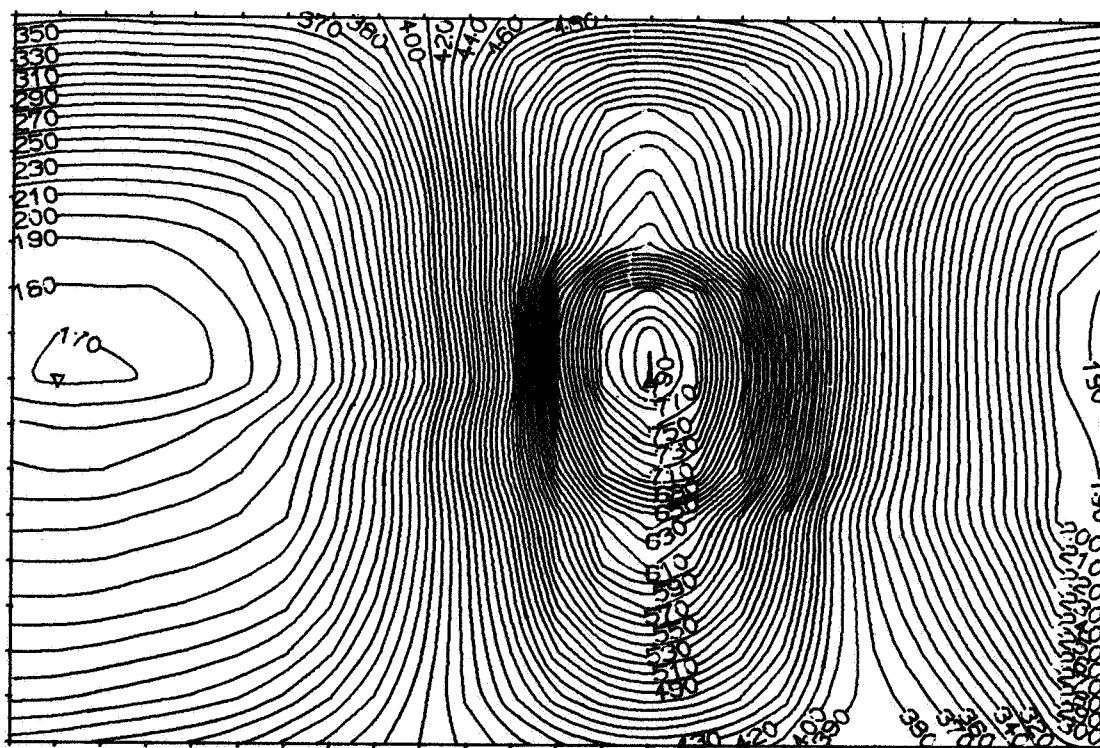
Eté Altitude 3



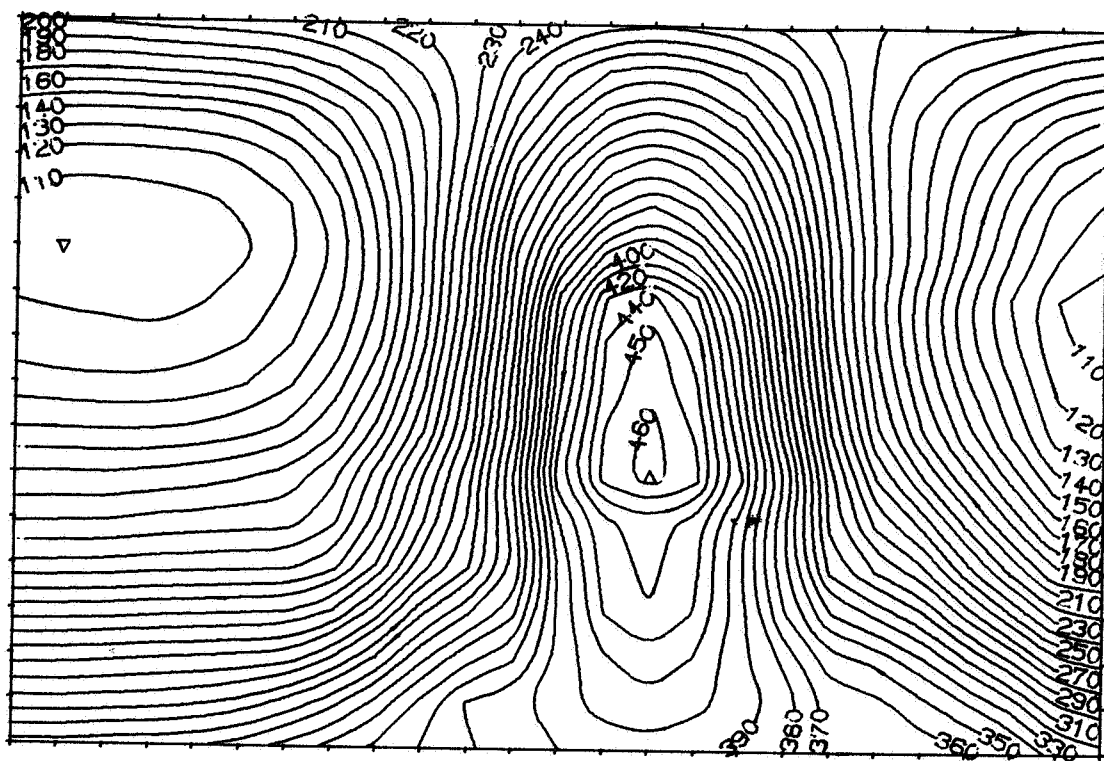
Automne Altitude 3



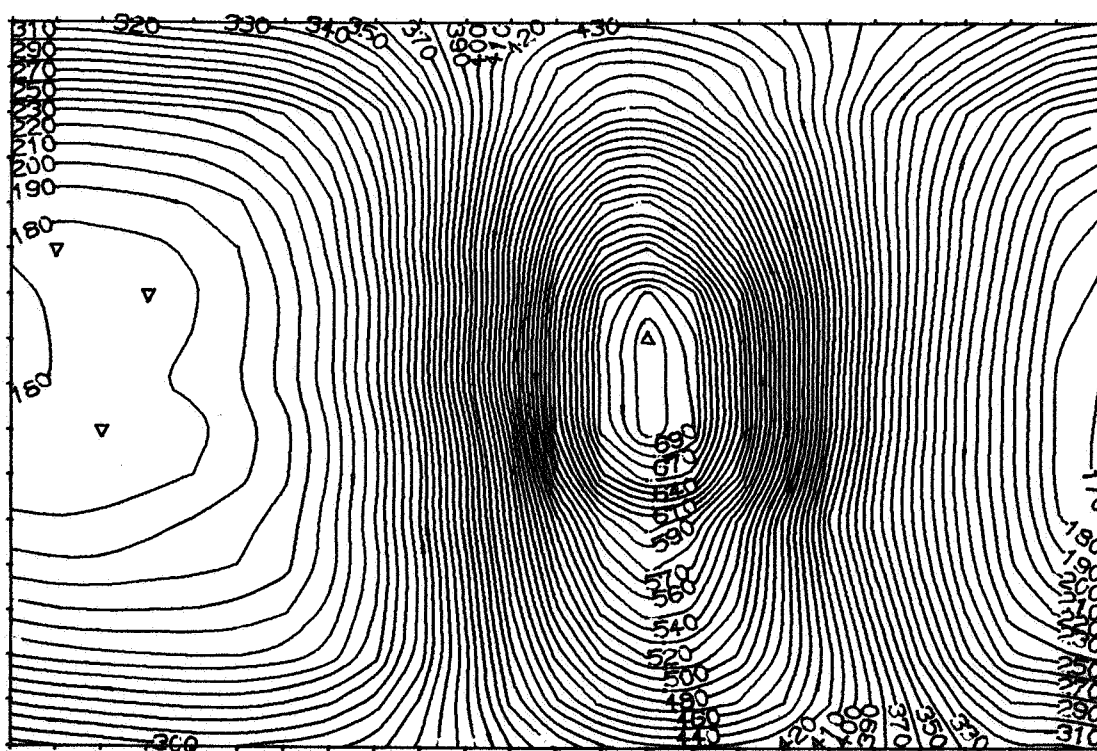
Hiver Altitude 4



Printemps Altitude 4



Eté Altitude 4



Automne Altitude 4

VI - ANALYSE GLOBALE DES EFFETS

METHODE DU TRI CROISE

- I PRINCIPE DE LA METHODE
- II MISE EN OEUVRE
- III RESULTATS GENERAUX
- IV RESULTATS PAR PARAMETRES
- V CONCLUSION

I PRINCIPE DE LA METHODE

La méthode du tri croisé est décrite par J.P. BORDET dans sa thèse, (cf [2]). Nous en rappellerons seulement le principe. Une variable, le coefficient correcteur f , dépend, par une relation à priori inconnue, d'un certain nombre d'autres variables. Pour essayer d'étudier la relation, on s'intéresse essentiellement aux fréquences d'événements du type " f est compris entre telle et telle valeur, et tel paramètre descriptif est compris entre telle et telle borne".

Il convient pour cela de découper en tranches aussi bien les paramètres descriptifs que la variable expliquée.

Convenons de noter $n_{ij}(k)$ le nombre d'événements où le coefficient f appartient à la tranche numéro i , et la variable numéro k à la tranche numéro j . Pour fixer les idées, nous allons construire un tel tableau. Supposons que l'intervalle de variation de f soit découpé en 3 tranches : De 0,5 à 0,8, de 0,8 à 1,2, de 1,2 à 1,5. Les variables explicatives sont la latitude (découpée en 3 tranches : De -90 à -30, de -30 à +30, et de +30 à +90) et l'heure locale (découpée en 2 tranches : De 6 à 18 heures, et de 18 à 6 heures.)

Supposons qu'on dispose de 10 observations

	f	Latitude	Heure locale
1	0,7	-80	12
2	0,82	+50	16
3	1,45	+10	5
4	1,03	+70	22
5	0,85	-60	19
6	0,60	+55	21

	f	Latitude	Heure locale
7	0,85	+12	22
8	0,58	-25	4
9	1,35	+17	12
10	1,12	+53	16

Le nombre d'observations où f appartient à la 1ere tranche et la latitude à la 1ere tranche est $n_{11}(1)=1$ (observation numero 1).

Le nombre d'observations où f appartient à la 2eme tranche et la latitude à la 3eme tranche est : $n_{13}(1)=3$ (observations 2, 5, et 10).

On obtient le tableau croisé suivant :

	LAT 1	LAT 2	LAT 3	HL 1	HL 2
F 1	1	1	1	1	2
F 2	1	1	3	2	3
F 3	0	2	0	1	1

Ce tableau croisé permet de mettre en évidence des relations simples entre les variables explicatives et la variable expliquée. Par exemple, conditionnellement au fait qu'on est dans la 3eme tranche de latitude, on a plus de chances (3/4) d'avoir un coefficient f compris entre 0,8 et 1,2 que d'avoir un coefficient f inferieur à 0,8 (1/4).

L'Analyse des Correspondances est la méthode adoptée à la visualisation de telles associations (cf [3]). Elle permet une description objective des écarts entre un ensemble de lois de probabilité et une loi de référence, qui est la loi de probabilité moyenne. Comme au chapitre précédent, nous renvoyons le lecteur à l'annexe pour avoir une description mathématique simplifiée de l'Analyse des Correspondances.

Pour analyser avec profit les diagrammes issus d'une Analyse des Correspondances, il faut retenir essentiellement les points suivants.

- . Deux tranches de paramètre sont proches si les histogrammes des f dans ces deux tranches sont approximativement identiques.
- . Une tranche de paramètre est d'autant plus proche d'une tranche de f que l'histogramme des f dans cette tranche est plus "concentré" autour de la tranche de f .

II MISE EN OEUVRE

Nous avons étudié la correspondance entre le logarithme du coefficient correcteur f et les sept paramètres déjà évoqués. Nous utilisons le logarithme de f car f est un facteur multiplicatif. Dans cette étude, f varie entre 0,33 et 3, et, dans ce champ de mesures, il n'est pas possible d'assimiler $\log(1+x)$ à x .

L'intervalle de variations de $\log f$ a été découpé en 20 tranches égales, de $\log(1/3)$ à $\log(3)$.

L'intervalle de variation de la date est découpé en 12 tranches de un mois.

L'intervalle de variation de l'heure locale est découpé en 12 tranches de 2 heures.

La latitude est découpée en 8 tranches : $(-90, -60)$; $(-60, -40)$; $(-40, -15)$; $(-15, 0)$; $(0, +15)$; $(+15, +40)$; $(+40, +60)$; $(+60, +90)$.

L'altitude est découpée en 4 tranches : De 250 à 320, de 320 à 530, de 530 à 750, et de 750 à 1250 km.

La variation de flux est découpée en 6 tranches : $(-50, -20)$; $(-20, -10)$; $(-10, 0)$; $(0, +10)$; $(+10, +30)$; $(+30, +80)$.

Le flux moyen FB est découpé en 9 tranches : $(40, 75)$; $(75, 80)$; $(80, 90)$; $(90, 110)$; $(110, 135)$; $(135, 145)$; $(145, 150)$; $(150, 156)$; $(156, 180)$.

L'activité géomagnétique AP est découpée en 7 tranches : $(0, 5)$; $(5, 7)$; $(7, 11)$; $(11, 17)$; $(17, 25)$; $(25, 40)$; $(40, 180)$.

Les tableaux suivants donnent les effectifs et la moyenne de $\log f$ dans chaque tranche.

Nom	Effectif	Moyenne de log F.	Moyenne de F.
Janvier	3375	0.005	1.01
Février	3643	0.004	1.01
Mars	3523	- 0.034	0.92
Avril	3855	- 0.056	0.88
Mai	4009	- 0.062	0.87
Juin	3651	- 0.060	0.87
Juillet	3700	- 0.039	0.91
Août	3311	- 0.027	0.94
Septembre	3161	- 0.070	0.85
Octobre	3584	- 0.048	0.89
Novembre	3425	- 0.021	0.95
Décembre	3403	0.004	1.01
0 - 2 heures	2187	- 0.041	0.91
2 - 4 heures	2197	- 0.047	0.90
4 - 6 heures	2629	- 0.044	0.90
6 - 8 heures	3346	- 0.030	0.93
8 -10 heures	4196	- 0.017	0.96

Nom	Effectif	Moyenne de log F.	Moyenne de F.
10-12 heures	4417	- 0.034	0.92
12-14 heures	4770	- 0.039	0.91
14-16 heures	4899	- 0.025	0.94
16-18 heures	4574	- 0.018	0.96
19-20 heures	4023	- 0.042	0.91
20-22 heures	3087	- 0.050	0.89
22-24 heures	2315	- 0.048	0.89
LAT 1: -90,-60	4083	- 0.028	0.94
LAT 2: -60-40	3860	-0.047	0.90
LAT 3: -40,-15	8222	- 0.041	0.91
LAT 4: -15,0	4421	- 0.043	0.90
LAT 5: 0,15	4660	- 0.040	0.91
LAT 6: 15,40	8466	- 0.036	0.92
LAT 7: 40,60	4415	- 0.022	0.95
LAT 8: 60,60	4513	- 0,010	0.98
ALT 1 250,320	4575	- 0.046	0.90
ALT 2 320,530	10035	- 0.052	0.89
ALT 3 530,750	13003	- 0.053	0.89

Nom	Effectif	Moyenne de log F.	Moyenne de F.
ALT 4 750,1200	15027	- 0.002	1.
DF 1 -50,-20	3829	- 0.062	0.87
DF 2 -20,-10	5191	- 0.029	0.94
DF 3 -10, 0	11458	- 0.042	0.91
DF 4 0,10	12380	- 0.033	0.93
DF 5 10,30	6228	- 0.007	0.98
DF 6 30,80	3554	- 0.036	0.92
FB 1 40,75	4109	- 0.078	0.84
FB 2 75,80	5671	- 0.059	0.87
FB 3 80,90	3550	- 0.049	0.89
FB 4 90,110	4464	- 0.001	0.99
FB 5 110,135	4106	0.019	1.04
FB 6 135,145	7969	- 0.020	0.95
FB 7 145,150	5278	- 0.020	0.95
FB 8 150,156	2280	- 0.033	0.93
FB 9 156,180	5213	- 0.069	0.85

Nom	Effectif	Moyenne de log F.	Moyenne de F.
AP 1 0, 5	9496	0.014	1.03
AP 2 5, 7	7019	- 0.020	0.95
AP 3 7, 11	9085	- 0.039	0.91
AP 4 11, 17	7906	- 0.054	0.88
AP 5 17, 25	4424	- 0.070	0.85
AP 6 25, 40	2623	- 0.073	0.84
AP 7 40, 180	2087	- 0.076	0.83

III RESULTATS GENERAUX

A Inertie portée par les axes

A priori, nos 58 points sont distribués dans un espace à 20 dimensions. L'analyse des correspondances extrait les axes principaux d'inertie du nuage : le premier facteur explique 61 % de l'inertie, le second 23,5 %, si bien que deux dimensions seulement expliquent 84,5 % de l'inertie.

Le phénomène se déploie donc en fait selon une structure bi-dimensionnelle. Donnons à titre indicatif les taux d'inertie expliqués par les 5 premiers axes.

AXE	1	2	3	4	5
%	61,2	23,4	4,5	3,5	1,9
% cumule	61,2	84,6	89,1	92,6	94,5

La chute très nette observée entre le deuxième et le troisième facteur prouve également la nature bidimensionnelle du phénomène.

B Explication de l'inertie d'un axe par les paramètres descriptifs

Si chacun des paramètres contribuait de la même façon à l'inertie d'un axe, chacun y contribuerait pour $1/7$, soit 14,3 %. On appellera contribution d'un paramètre à l'explication d'un axe la somme des contributions de chaque tranche de ce paramètre à l'inertie de l'axe, et contribution d'un paramètre à l'inertie du plan 1-2 la somme pondérée par les inerties et chacun des axes des contributions de ce paramètre à l'explication des axes. Le tableau suivant résume ces différentes contributions.

.../...

NOM	AXE 1	AXE 2	PLAN 1 - 2
DATE	18,6	6,3	15,3
Heure locale	3,7	1,8	3,1
Latitude	3,2	8,1	4,5
Altitude	16,2	44,3	24,3
Variation de flux	5,9	4,8	5,7
Flux moyen	25,4	13,6	21,5
Activité géomagnétique	27,0	21,1	25,6

Contribuent donc de façon significative à l'axe 1, et par ordre décroissant : l'AP, le flux moyen, la date et l'altitude.

Contribuent à l'axe 2 par ordre décroissant : l'altitude et l'AP.

Contribuent au plan : l'AP, l'altitude, le flux moyen et la date.

Il convient maintenant d'examiner le sens de chacun des axes.

C Le sens des axes

Le sens du premier axe est parfaitement clair: c'est simplement un facteur de niveau. Il oppose les tranches de f faible (situées à gauche) aux tranches de f fort (situées à droite) la progression est particulièrement nette et régulière de F 5 à F 15. La situation se brouillant un peu aux extrémités, car le nombre d'observations est relativement faible dans ces tranches.

La courbe des tranches de $\log f$ se déploie sur le plan factoriel selon un nuage à aspect parabolique, dont la signification est bien connue en analyse des données : plus on est à l'intérieur de la parabole, plus l'histogramme de f est plat. Plus on est à l'extérieur ou vers le sommet de la parabole, plus l'histogramme est "pointu" au voisinage de la tranche près de laquelle on se trouve. C'est la raison pour laquelle - entre autres - le point Alt 3 - représentant les altitudes comprises entre 530 et 750 km est très intérieur (on trouve fréquemment des valeurs de f à cette altitude), alors que Alt 1 (points d'altitude inférieure à 320 km) est très extérieur : A cette altitude, les effets sont considérablement plus faibles, l'histogramme des f est très concentré. On reviendra sur ces questions dans l'analyse paramètre par paramètre.

IV RESULTAT PAR PARAMETRE

A La date

Paramètre essentiellement significatif sur le premier axe, facteur de niveau. JACCHIA 71 sous estime assez notablement en décembre, janvier, février, surestime notablement en mai et juin et septembre. En fait, on avait vu lors de la première étude que JACCHIA 71 surestime en mai, juin, juillet, août en forte activité, et aux équinoxes en faible activité. Ici les deux effets mêlés montrent que les effets prédominants sont la surestimation en septembre en faible activité, et en mai-juin en forte activité.

Le paramètre est sans signification sur le deuxième axe, ce qui montre que la variabilité du coefficient f reste sensiblement la même au cours des saisons.

B L'heure locale

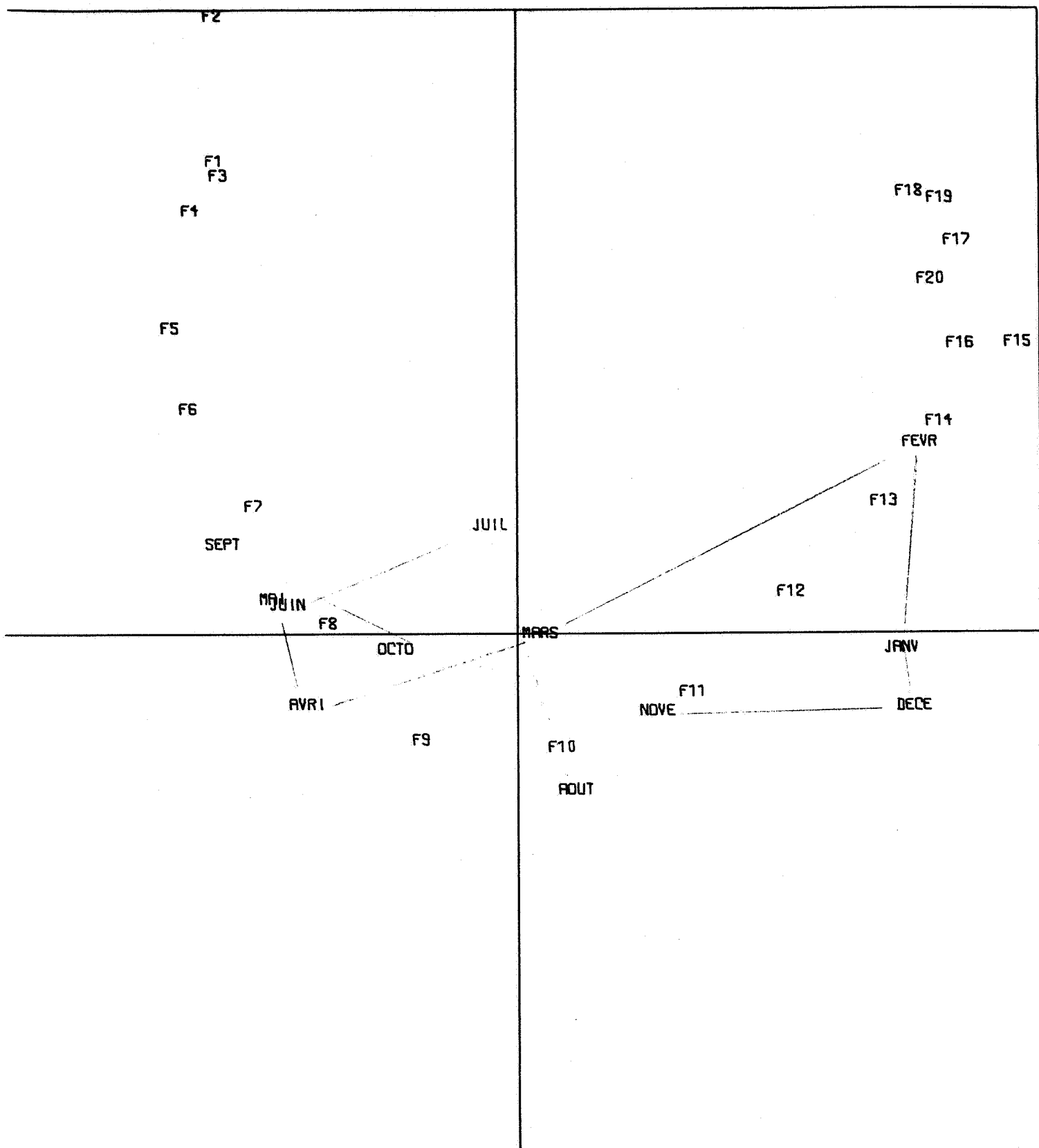
La paramètre ne semble pas significatif sur le plan factoriel. D'une façon générale la variation de f est faible avec l'heure locale, mais on a vu qu'elle avait tendance à s'inverser avec l'activité solaire. Ici prédomine un maximum de jour (sous estimation le jour) et un minimum de nuit, surtout accentué entre 20 et 24 heures.

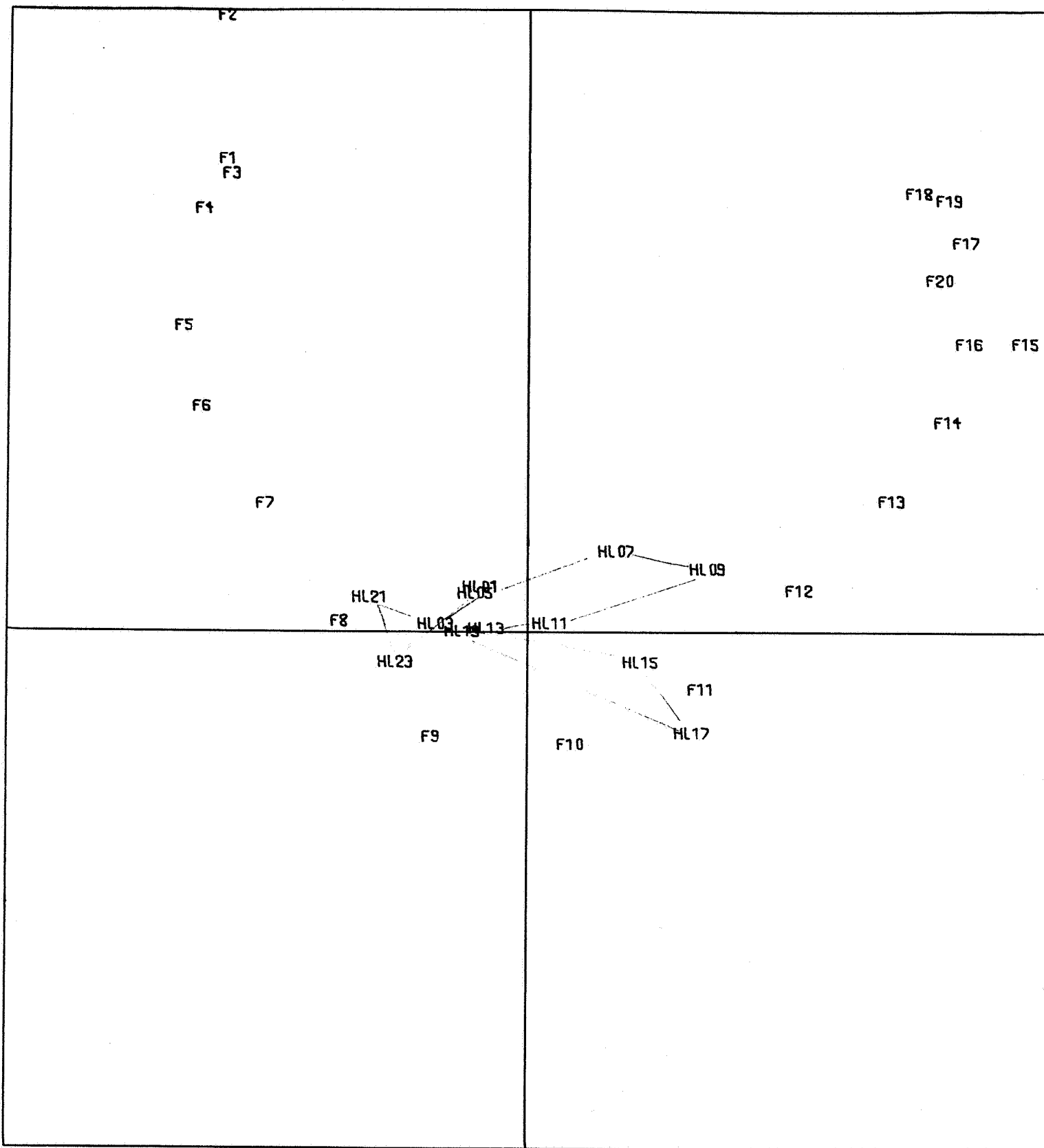
Dans l'ensemble, ce paramètre semble néanmoins d'une importance faible.

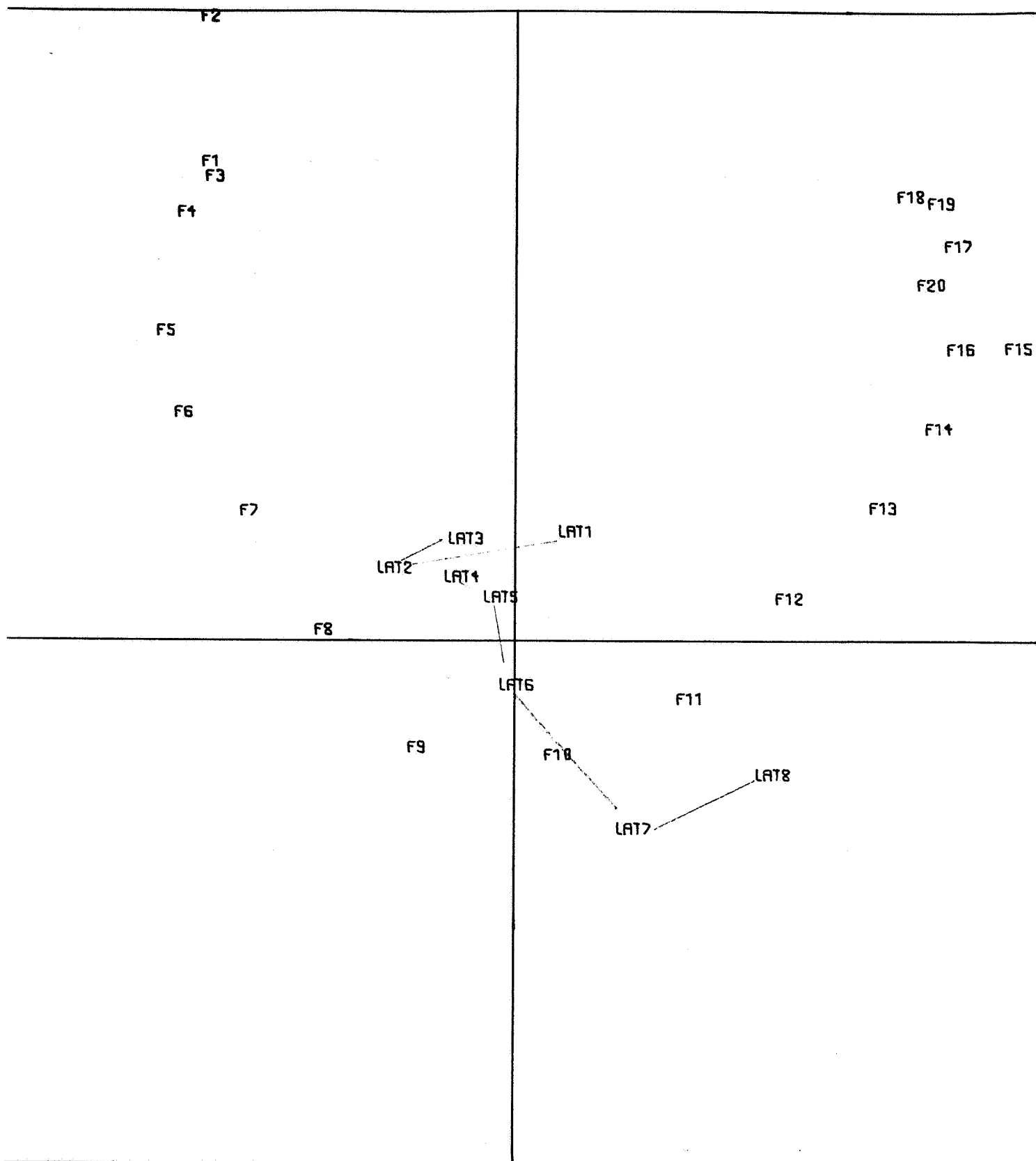
C La latitude

Bien qu'il contribue faiblement à la définition du plan factoriel, il est intéressant d'analyser ce paramètre, dont on a vu l'importance dans l'étude précédente. Les points de LAT 1 à LAT 5 sont bien groupés dans le cadran haut et gauche du plan factoriel, alors que les points LAT 6, LAT 7, LAT 8 se déploient avec une certaine régularité dans le cadran bas-droite du plan. Il n'y faut point voir une opposition de niveau - voisinage de F 10 et F 11 au nord, de F 9 et F 8 au sud -, mais une opposition de variabilité :

.../...







Il faut s'attendre à voir, de façon générale, une variabilité plus grande au sud qu'au nord, en fonction des autres paramètres. L'analyse précédente nous a montré que l'effet latitude dépendait essentiellement de l'altitude, cette analyse nous apprend que, d'une façon générale, les effets quels qu'ils soient (sur ou sous estimation) seront plus prononcés au sud. Il faut noter que la latitude était un paramètre très actif dans le travail qu'avait effectué J.P BORDET. Il cesse de l'être dans notre travail, car nous travaillons dans des gammes d'altitude beaucoup plus variées, les effets aux altitudes basses annulant les effets aux altitudes plus hautes. Le centre des masses de nos altitudes appartenant en fait aux altitudes hautes, nous obtenons un effet (non significatif) inverse de celui obtenu par J.P BORDET : à savoir sous estimation aux pôles et surestimation à l'équateur, ce qui correspond effectivement au comportement de l'hélium et de l'hydrogène atomique, présents respectivement à partir de 500 et 800 km.

D L'altitude

Présent au cours de toute cette étude, l'altitude, qui est manifestement le paramètre le plus important, joue un rôle actif mais non prédominant sur le premier axe, axe de niveau: celui-ci oppose les altitudes supérieures à 750 km, où JACCHIA 71 sous estime assez considérablement aux altitudes plus faibles, où JACCHIA 71 surestime de façon générale.

Mais son véritable rôle apparaît sur le deuxième axe : alors qu'en dessous de 530 km, les effets - quels qu'ils soient - sont d'amplitude assez faible, ils prennent une importance considérable entre 500 et 750 km. C'est à ces altitudes - et à elles seules pratiquement - que - que JACCHIA 71 peut commettre des erreurs du simple au double. Les effets perdent une partie de leur ampleur, qui reste forte, au dessus de 750 km. La zone de l'hélium semble donc la plus mal modélisée. Il convient de tempérer un peu ce jugement sévère en précisant bien le rôle de l'altitude: Facteur de niveau opposant la zone de l'hydrogène atomique aux autres, mais aussi et surtout facteur de forme, imposant aux effets en latitude en particulier une structure bien différente selon

F2

F1
F3

F4

F5

F6

F7

ALT3

F8

F9

ALT2

ALT1

F18 F19

F17

F20

F16 F15

F14

F13

F12 ALT4

F11

F10

que les constituants majoritaires sont l'oxygène et l'azote d'une part, ou l'hélium d'autre part.

E. LA VARIATION DE FLUX

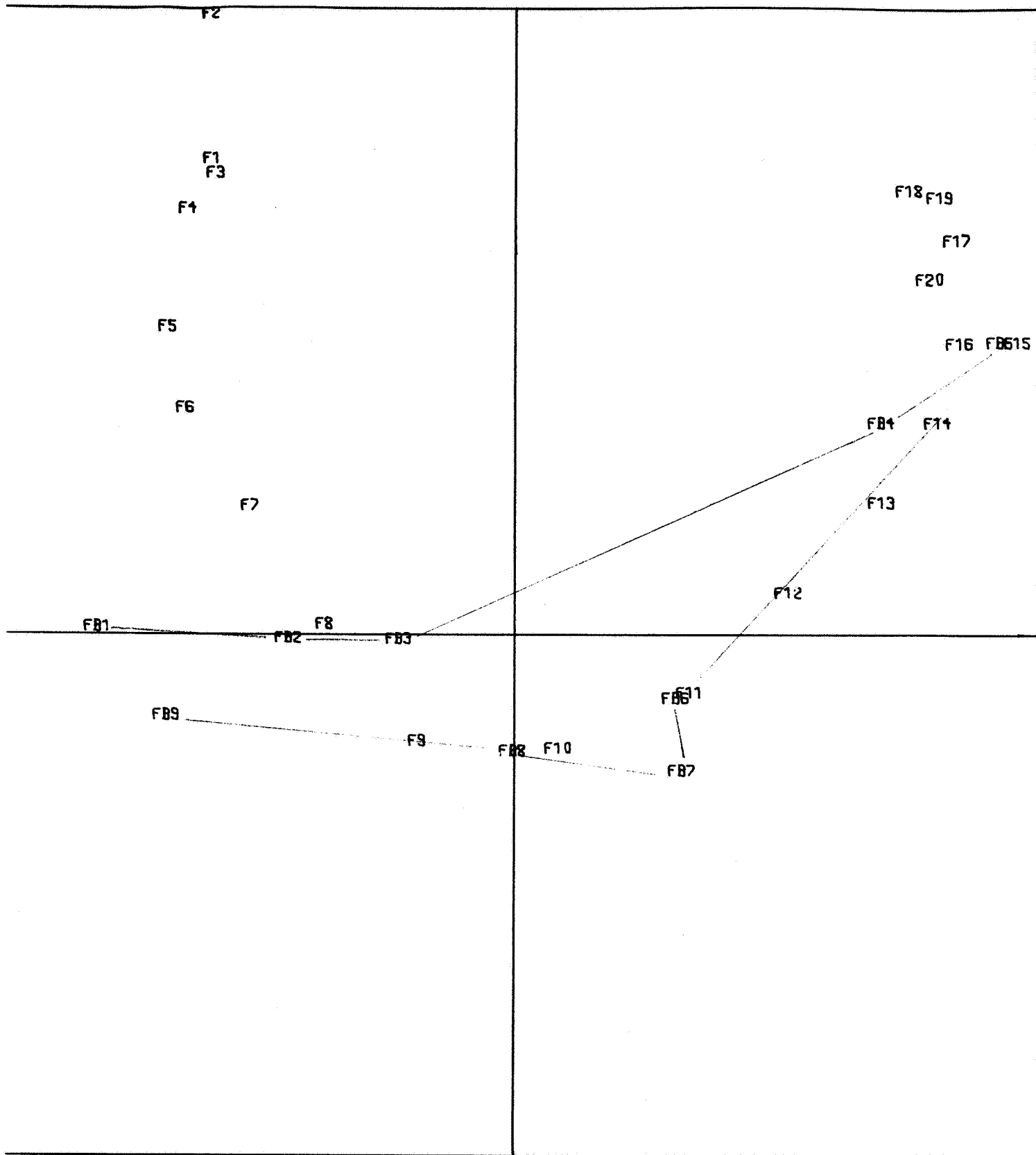
Ce paramètre ne contribue pas significativement à la définition du plan factoriel. Il faut néanmoins s'y arrêter un instant, Monsieur WACHTEL, chercheur au Laboratoire de Dynamique Spatiale, ayant mis en évidence une propriété assez curieuse confirmée par l'étude du plan factoriel : JACCHIA 71 a tendance à surestimer lorsque les variations de flux sont importantes en valeur absolue. La variation de flux contribue relativement peu au plan factoriel peut être plus parce que les points significatifs DF5, DF1 ont des effectifs assez faibles que par sa structure propre. Les variations de flux positives semblent être des causes d'incertitude plus grande que les variations négatives.

F. LE FLUX MOYEN

Ce paramètre très important contribue de façon très notable à l'axe 1. On observe une tendance forte à la surestimation en minimum solaire, à la sous estimation en période de transition (par exemple l'année 1967 dans notre fichier), et une tendance plus faible à la sous estimation en maximum solaire. La position du point FB9 paraît surprenante de prime abord, d'autant plus que ce point représente 5213 observations; il s'agit vraisemblablement d'un couplage avec les points d'AP très forte (AP5,6,7) surtout relatifs aux années 1961-1962, où de fréquents orages magnétiques perturbèrent la structure de la haute atmosphère. On retiendra surtout une surestimation importante pour des flux inférieurs à 20, et une surestimation plus faible au dessus de 100.

G. L'ACTIVITE GEOMAGNETIQUE

Paramètre très actif dans la définition du plan factoriel. La courbe des AP suit de façon surprenante la courbe des tranches de Log f, mais en sens inverse : plus l'AP est forte, et plus JACCHIA surestime : f décroît très notablement avec l'AP. C'est le seul paramètre qui semble expliquer correctement les très faibles valeurs de f, observées en période d'orage magnétique, de même que les flux moyens intermédiaires (de 70 à 135) semblent seuls responsables des très fortes valeurs de f. Il n'en reste pas moins qu'en période normale (AP inférieure à 20) il existe un effet en AP particulièrement net : de F13 pour AP1, ou passe au voisinage de F8 pour AP4. La régularité surprenante de l'effet



F2

F1
F3

F4

F5

F6

F7

F8

DF3

DF1

F9

DF6

DF4

F10

DF2

F11

F12

DF5

F13

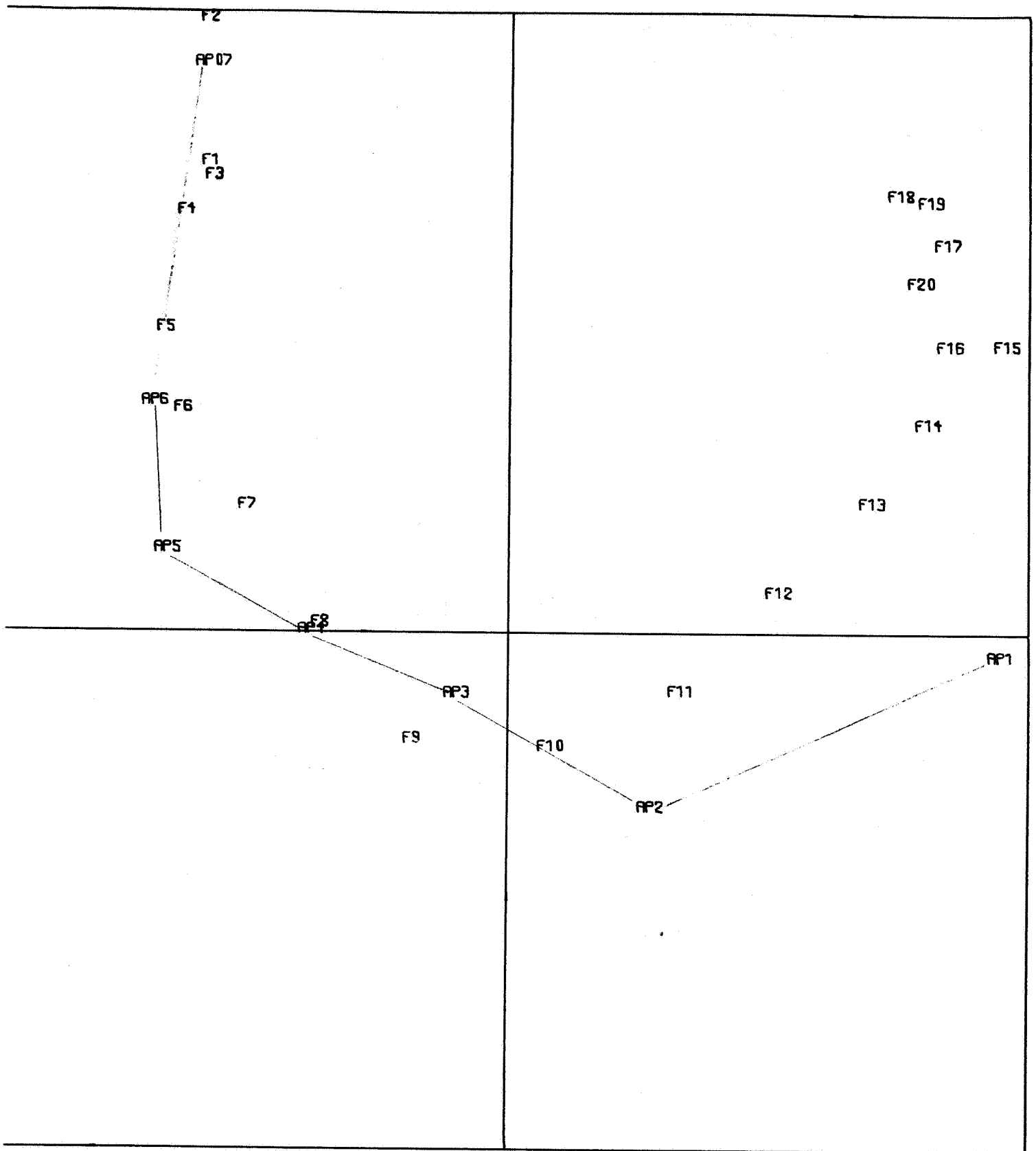
F18 F19

F17

F20

F16 F15

F14



permet de penser qu'il sera aisément corrigé. Nous ne l'avions pas pris en compte dans nos études précédentes, car il s'agit d'un paramètre instantané. Nous souhaitons au contraire nous placer dans des conditions d'une certaine régularité, et seul le flux moyen (par l'intermédiaire de la période du cycle solaire de 11 ans) a été pris en compte. L'absence de l'étude de l'AP est sans aucun doute une des lacunes les plus regrettables de la première étude.

V. CONCLUSION

A. REMARQUES METHODOLOGIQUES

Il convient, alors que deux traitements d'esprit et de nature différents ont été réalisés par Analyse des Correspondances, de comparer leurs performances respectives.

La méthode disjonctive complète qui permet de réaliser le tri croisé que nous venons d'analyser, est d'une grande généralité. Elle permet de prendre en compte simultanément un grand nombre de "variables explicatives". Nous n'en avons que sept, mais elle se serait volontiers accomodée d'une vingtaine ou d'une trentaine. C'est, à notre point de vue, son avantage primordial.

La méthode de l'Analyse des courbes est moins générale. Elle permet néanmoins une description plus riche des effets, dans la mesure où elle permet d'étudier l'influence conjuguée de différents paramètres. Elle permet par exemple de mettre en évidence l'inversion en altitude de l'effet latitude, qui n'apparaît pas dans la méthode disjonctive.

L'Analyse des courbes nécessite au préalable le concours de chercheurs de la discipline pour sélectionner, dans toute la gamme des effets couplés, ceux qui à priori leur semblent avoir un sens, et éliminer ceux dont, de toute façon, ils ne sauraient que faire. Enfin, un problème indiscutable de taille de fichier se pose, la méthode disjonctive peut s'accomoder de fichier d'importance modeste (quelques milliers de points) quel que soit le nombre des variables, dans la mesure où elle juxtapose un certain nombre d'histogrammes bidimensionnels. L'Analyse des Courbes exige des fichiers d'autant plus grands qu'on souhaite traiter d'autant plus de variables simultanément.

D'un point de vue mathématique, les choses se présentent de la façon suivante :

Dans la méthode disjonctive complète, on associe à chaque variable descriptive un espace de dimension au plus égale au nombre de tranches sur lesquelles cette variable est découpée. A chaque tranche d'un paramètre X on associe une variable aléatoire $1_{X_i}(\omega)$ qui vaut 1 si, pour l'observation ω , X appartient à la tranche i , et zéro ailleurs. On a les relations d'orthogonalité :

$$\langle 1_{X_i} , 1_{X_j} \rangle = 0. \quad \text{si } i \neq j$$

De plus, on considère les espaces engendrés par chaque paramètre comme orthogonaux, et donc :

$$\langle I_{Y_i}, I_{X_j} \rangle = 0 \quad \text{si } X \neq Y.$$

On travaille donc sur des σ - algèbres engendrées par un nombre de variables aléatoires égales à la somme du nombre de tranches en lesquelles chaque paramètre est découpée.

Dans la méthode de l'Analyse des courbes, on travaille sur des σ algèbres plus riches; on s'intéresse à des événements du type : X appartient à la tranche i, [et] Y appartient à la tranche j ; [et] Z appartient à la tranche k : on travaille sur des algèbres engendrées par un nombre de variables aléatoires égal au produit des nombres de tranche en lesquelles chaque paramètre est découpé : Si on a 10 paramètres découpés chacun en 4 tranches, on aura dans le premier cas $10 \times 4 = 40$ événements, dans le second $4^{10} = 1.048.576$ événements élémentaires....

On peut en fait vraisemblablement trouver un moyen terme entre les deux méthodes, à partir des études faites. L'Altitude, et à un moindre degré l'Activité Solaire, nous sont apparus comme les seuls paramètres importants structurant les effets à l'aide d'un autre paramètre. On peut alors imaginer des tris croisés recensant des événements du type :

L'observation est dans telle tranche d'altitude, telle tranche de flux, telle tranche d'Activité Géomagnétique, et dans telles tranches des autres paramètres. C'est ce que nous verrons au chapitre suivant.

B. CONCLUSIONS SUR LES EFFETS

Cette Analyse nous a permis de confirmer le rôle très important et très régulier de l'AP, déjà mis en évidence dans les mêmes conditions par J.P. BORDET. D'une manière générale, nous avons surtout pu préciser l'influence des paramètres d'activité solaire d'une façon beaucoup moins grossière que dans l'étude précédente. Les résultats concernant les paramètres de position sont très concordants en ce qui concerne l'altitude et la date; ceux concernant la latitude et l'heure locale sont plus faibles, mais ne contredisent en rien nos premières conclusions.

VII - ANALYSE MULTIDISJONCTIVE

=====

I - OBJECTIFS

II - CONSTITUTION DU TABLEAU

III - LES RESULTATS DE L'ANALYSE DU TABLEAU

IV - MODELISATION

I - OBJECTIFS

Les deux analyses précédemment exposées sont complémentaires. Nous avons essayé d'en bâtir une synthèse qui conserve les avantages de l'une (prise en compte d'interactions d'ordre élevé) et de l'autre (utilisation de tous les paramètres, possibilité de modélisation simple dans l'espace factoriel).

L'analyse du tri croisé nous a montré que les paramètres n'avaient pas des importances comparables : altitude, flux moyen et Activité Géomagnétique sont les plus déterminants. Nous sommes donc amenés à bâtir un tableau de correspondance qui privilégie ces trois paramètres, que nous appellerons désormais "paramètres primaires".

II - CONSTITUTION DU TABLEAU

II-1 - DEFINITION

Définissons une situation de base par un ensemble de modalités portant sur les trois paramètres primaires :

- + l'altitude est basse (< 530 km) ou haute.
- + le flux est faible (< 85 unités), moyen (compris entre 85 et 135 unités) ou fort.
- + l'Activité Géomagnétique est faible (inférieure à 3), moyenne (comprise entre 3 et 7) ou forte.

En tout donc 18 situations de base qui seront repérées sur les graphiques par un nombre de trois chiffres, le premier donnant le numéro de la classe d'altitude, le second le numéro de la classe de Flux, le troisième le numéro de la classe d'Activité Géomagnétique.

Pour chaque situation, on génère un tableau disjonctif complet croisant les tranches du coefficient f avec les paramètres restant, selon les principes suivants :

- + Au coefficient correcteur f, sont associées 12 tranches, les 10 tranches centrales étant égales aux tranches de l'Analyse précédente, et les deux tranches extrêmes regroupant respectivement les tranches 1 à 5 et 16 à 20 dont l'effectif était trop faible. Elles sont repérées sur les graphiques par un symbole Fi, i allant de 1 à 12.
- + L'heure locale est scindée en 4 tranches égales repérées sur les graphiques par leurs bornes : 00 06, 06 12, 12 18, 18 24.
- + La date est scindée en 4 tranches :
 - . HIVE de Novembre à Février
 - . PRIN de Mars à Avril
 - . ETE de Mai à Août
 - . AUTO de Septembre à Octobre
- + La latitude est scindée en 5 tranches :
 - . PSUD de - 90° à - 45°
 - . TSUD de - 45° à - 15°
 - . EQUA de - 15° à + 15°
 - . TNOR de + 15° à + 45°
 - . PNOR de + 45° à + 90°

+ La variation de flux est scindée en trois tranches :

- DF⁻ : variation inférieure à - 10 unités
- DF0 : variation comprise entre - 10 et + 10 unités
- DF⁺ : variation supérieure à 10 unités

En tout donc, 16 tranches pour les paramètres explicatifs secondaires. A chaque situation est donc associé un tableau disjonctif 12×16 , et la superposition de ces tableaux fournit un tableau 12×288 qu'on soumet à l'Analyse.

II-2 - REMARQUES SUR LA REPRESENTATIVITE DU TABLEAU

On travaille sur un fichier d'environ 42.000 observations, qu'on partitionne en 18 classes : chaque tableau disjonctif est constitué à partir de 2.330 observations en moyenne. En fait, la situation d'effectif le plus faible ne comporte que 857 observations. Le tableau croisé "le plus fin" (qui comporte le plus grand nombre de cases) est le tableau croisant F avec la latitude (60 cases) : en moyenne donc, 14 observations par case, dans le cas le plus défavorable. Ces chiffres sont raisonnables, mais montrent qu'il n'aurait pas fallu introduire un quatrième "paramètre primaire".

III - LES RESULTATS DE L'ANALYSE DU TABLEAU

III-1 - L'INERTIE EXPLIQUEE PAR LES AXES

Les premières valeurs propres et les pourcentages correspondants sont consignés dans le tableau suivant :

AXE	INERTIE	% EXPLIQUE	% CUMULE
1	0.1737	55,3	55,3
2	0.0823	26,2	81,5
3	0.0208	6,7	88,2
4	0.0145	4,6	92,8

Les deux premiers axes suffisent à eux seuls à expliquer plus de 80 % de l'inertie : nous concentrerons notre attention sur le premier plan factoriel.

Les valeurs propres sont fortes, indiquant une bonne description de l'évolution de F en fonction des paramètres retenus. A titre indicatif, la première valeur propre de l'analyse précédente ne valait que 0.04.

III-2 - LES TRANCHES DU COEFFICIENT F

Elles se répartissent dans le plan factoriel de façon régulière, comme dans l'Analyse précédente, sur une courbe à aspect parabolique. Le premier axe reste un facteur de niveau, et le second un facteur de variabilité. L'Analyse du nuage des variables explicatives nous permettra de préciser ces notions.

III-3 - LES VARIABLES EXPLICATIVES

Ils serait délicat d'interpréter les axes à l'aide des 288 modalités présentes. Or les 288 points du plan se répartissent en 18 groupes de 16 points, correspondant aux 18 situations de base. Considérons dans un premier temps les barycentres de chacun de ces 18 groupes.

III-3-1 - La position des barycentres est-elle significative ?

Considérons un point G, de coordonnées X et Y, barycentre de 16 points M_i , de coordonnées X_i et Y_i , tirés au hasard (sans remise) dans le lot des 288 points. Supposons, pour faciliter les écritures, que tous les M_i aient le même poids.

$$\text{Comme } X = \frac{1}{16} \sum_{i=1}^{i=16} X_i, \quad Y = \frac{1}{16} \sum_{i=1}^{i=16} Y_i,$$

il est clair que $E(X) = 0$, $E(Y) = 0$

$$E(X^2) = \frac{1}{256} \sum_{i=1}^{i=16} \sum_{j=1}^{j=16} E(X_i X_j)$$

$$E(X^2) = \frac{1}{256} \sum_{i=1}^{i=16} E(X_i^2) + \frac{1}{256} \sum_{i \neq j} E(X_i X_j)$$

$E(X_i^2)$ est égal, par définition, à la valeur propre λ_1 .

$E(X_i X_j)$ serait nul si les 16 points M_i étaient tirés avec remise. Ici, comme tous les M_i sont différents, on doit introduire une légère correction :

$$\text{Pour } i \neq j \quad E(X_i X_j) = E(X_i E(X_j | X_i))$$

$$\text{Or :} \quad E(X_j | X_i) = \frac{1}{287} \sum_{k \neq i} X_k = \frac{-X_i}{287}$$

$$\text{Donc pour } i \neq j : \quad E(X_i X_j) = \frac{-E(X_i^2)}{287}$$

$$\text{D'où finalement :} \quad E(X^2) = \frac{\lambda_1}{16} - \frac{240}{256} \frac{\lambda_1}{287}$$

$$\text{soit} \quad : \quad E(X^2) \sim 0.01$$

$$E(Y^2) \sim 0.005$$

Donc, lorsqu'on tire 16 points au hasard parmi les 288, l'écart-type de l'abscisse de leur centre de gravité est de 0.10, celui de l'ordonnée est 0.07. Sous une hypothèse gaussienne, 66 % des centres obtenus sont contenus dans l'ellipse dont le grand axe, porté par le premier facteur, a une longueur de 0.10 et dont le petit axe, porté par le second facteur, a une longueur de 0.07. De même 95 % des centres doivent tomber dans l'ellipse déduite de la précédente par homothétie de rapport 2.

Aucun des 18 centres de gravité ne tombe dans la première ellipse, et trois seulement appartiennent à la seconde : c'est dire que la position des centres est très significative.

III-3-2 - La cohésion des 18 groupes

Nous avons représenté sur la figure 1 les centres de gravité de chaque groupe entouré d'un cercle dont le rayon est égal à la distance quadratique moyenne des points du groupe à leur centre de gravité. On note, comme on pouvait s'y attendre, que la dispersion des points autour de leur centre de gravité n'est pas trop forte. Notons également que les groupes correspondant à l'altitude élevée sont les plus dispersés : comme toujours dans cette étude, les phénomènes sont de variance plus forte aux altitudes élevées. On notera également que les cercles les plus grands sont obtenus pour un flux moyen (tendance à la sous-estimation) et une activité géomagnétique forte (tendance à la surestimation) : dans ces deux modalités, ces deux paramètres ont une influence contraire. S'il semble qu'en altitude faible, l'activité géomagnétique l'emporte, il n'en est pas de même à l'altitude 2 : les autres paramètres seront nécessaires pour conclure.

III-3-3 - Une analyse de contrôle

Deux remarques nous suggèrent une construction nouvelle :

- + La position de tous les centres de gravité des 18 situations est significative.
- + Considérons une modalité d'un paramètre secondaire, par exemple AUTO. Elle apparaît 18 fois, (une fois par situation de base). Soit alors G le barycentre des 18 points correspondant. Aucun des barycentres correspondant aux 16 modalités des variables secondaires n'a une position significativement différente de l'origine.

Nous avons alors constitué un tableau 12×18 , croisant simplement les tranches de F avec chacune des situations de base. Le plan des axes 1-2 est pratiquement invariant : il faut bien admettre que les barycentres des 18 modalités suffisent à eux seuls à déterminer le plan.

III-4 - L'INTERPRETATION DU PLAN AVEC LES 18 BARYCENTRES

Pour lire les graphiques qui suivent, il faut se souvenir que le 1er chiffre indique la tranche d'altitude, le second la tranche de flux, le troisième la tranche d'Activité Géomagnétique. Chacun des graphiques représente l'évolution des centres de gravité quand l'un des paramètres primaires varie.

III-4-1 - L'altitude

Un point représentatif de l'altitude 1 a toujours un deuxième facteur plus faible que son homologue à l'altitude 2 : quelle que soit la situation en flux et en A.P., les variations de P sont toujours plus faibles en basse altitude. On note également que les vecteurs joignant les points homologues sont (à une exception près) orientés vers la droite : l'altitude est aussi un facteur de niveau, surtout lorsque le flux est moyen ou fort. L'Activité Géomagnétique tempère ce déplacement : son amplitude est d'autant plus forte que l'A.P. est faible.

III-4-2 - Le flux

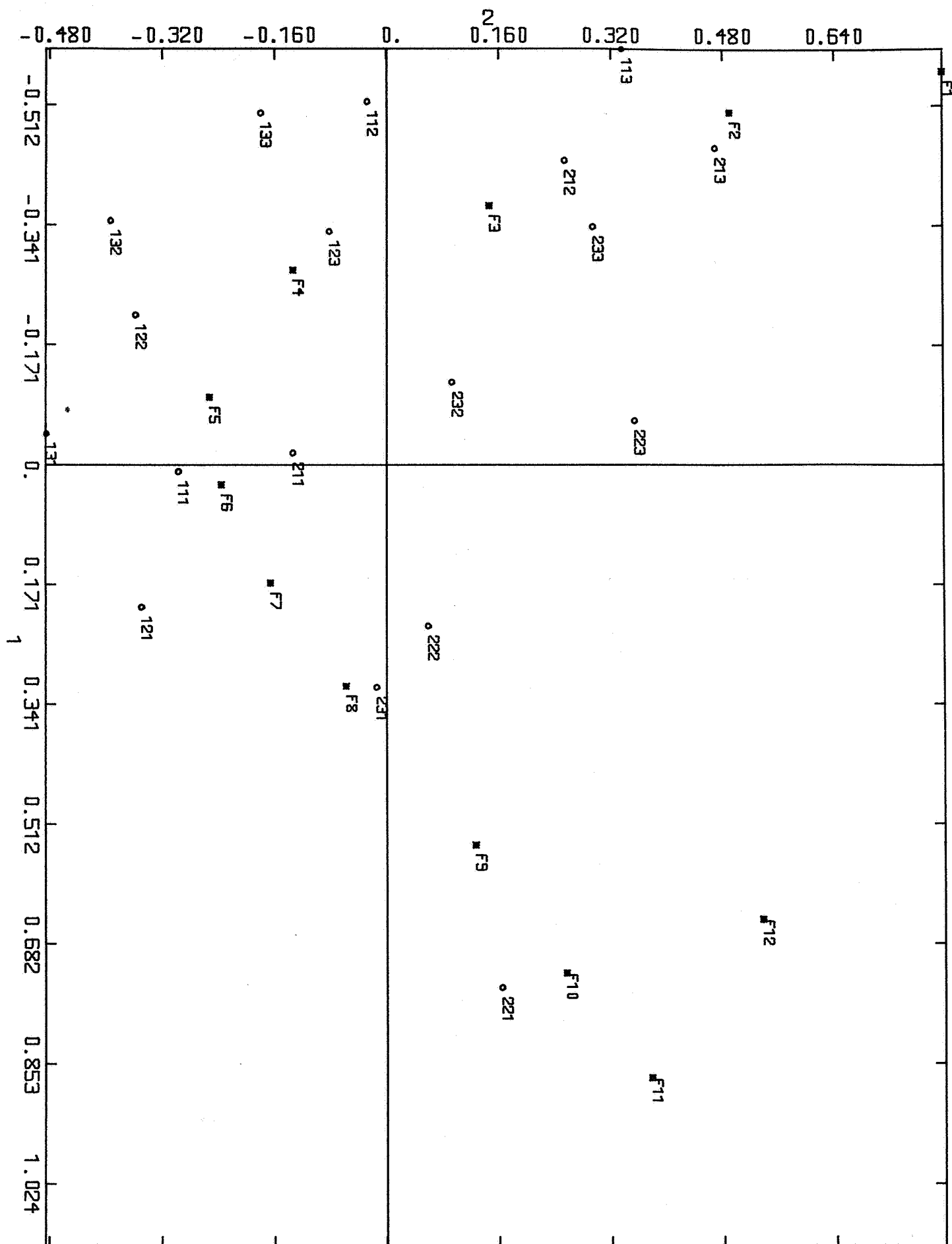
L'effet en flux est conforme à notre attente : f croît quand le flux croît jusqu'à sa valeur "moyenne", puis décroît. On remarque de plus qu'à mesure que le flux croît, le deuxième facteur décroît : la variabilité est d'autant plus faible que le flux est fort : comme l'altitude, le flux est à la fois un facteur de niveau et de variabilité.

III-4-3 - L'Activité géomagnétique

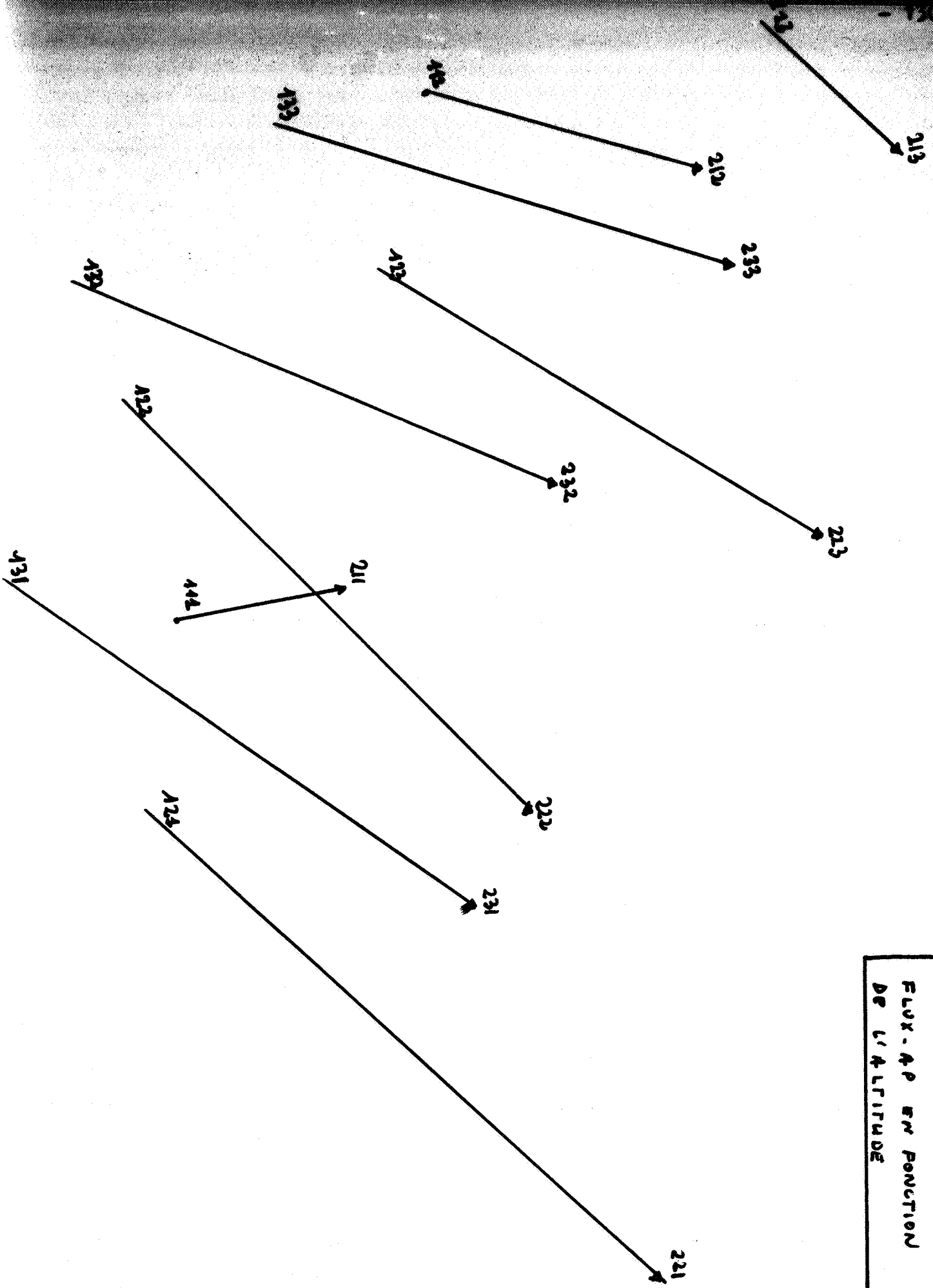
F décroît de façon systématique avec l'Activité Géomagnétique, pour toutes conditions de flux et d'altitude. On observe également que les points d'A.P. faible ont toujours un deuxième facteur plus faible que les points d'A.P. forte : bien que dans une moindre mesure, l'A.P. est aussi un facteur de variabilité.

III-4-4 - Conclusions sur les paramètres primaires

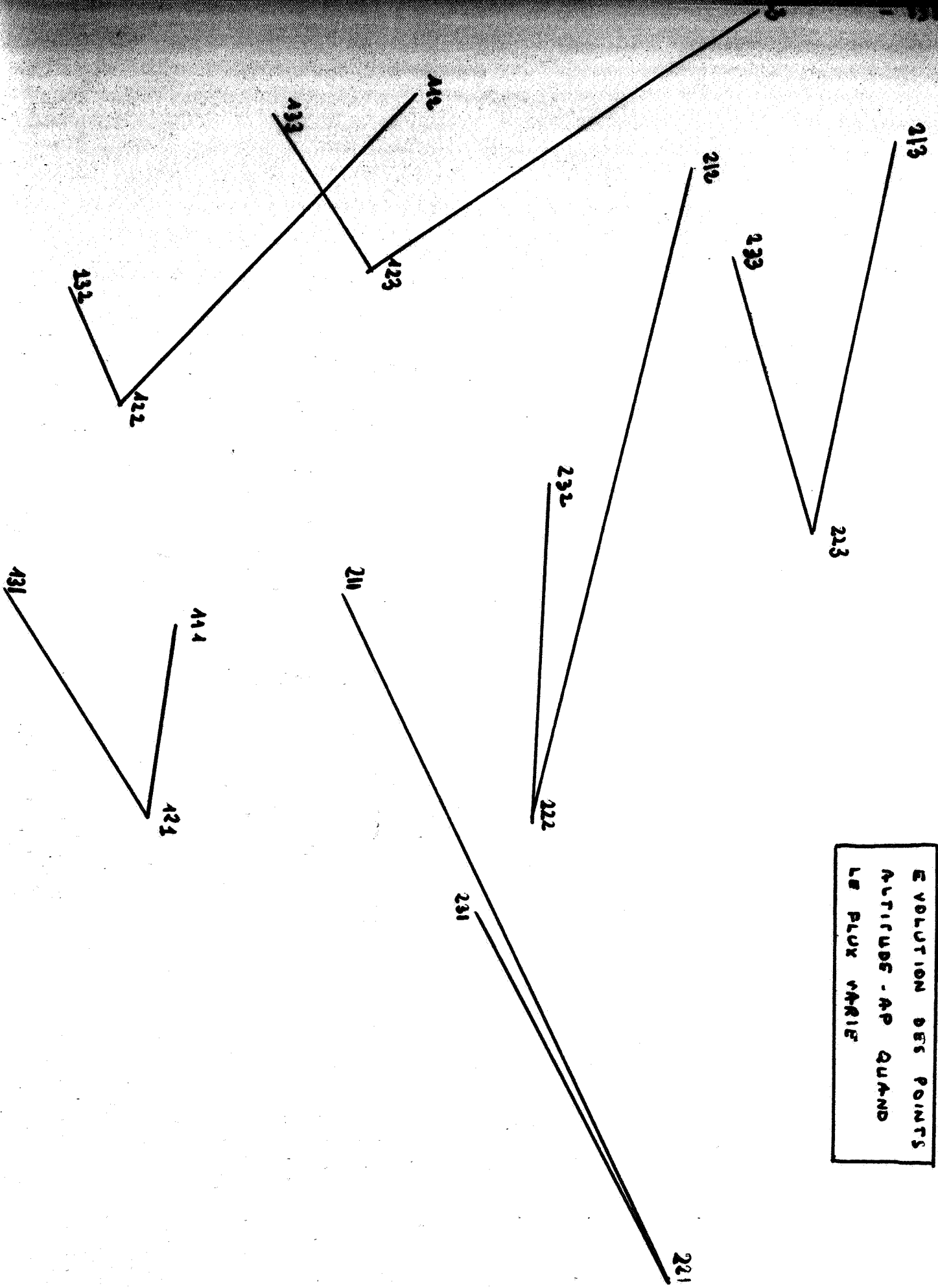
Tous les paramètres primaires sont à la fois des paramètres de niveau et de variabilité. Une analyse de ce type permet de préciser leurs interrelations, qui sont d'une nature assez complexe. Nous n'analyserons pas les 18 graphiques donnant la répartition des points autour de leur centre de gravité. Disons simplement qu'un certain nombre d'effets peuvent être retrouvés : la structure de l'effet latitude par exemple. Mais on n'apporte pas d'informations significatives véritablement nouvelles.



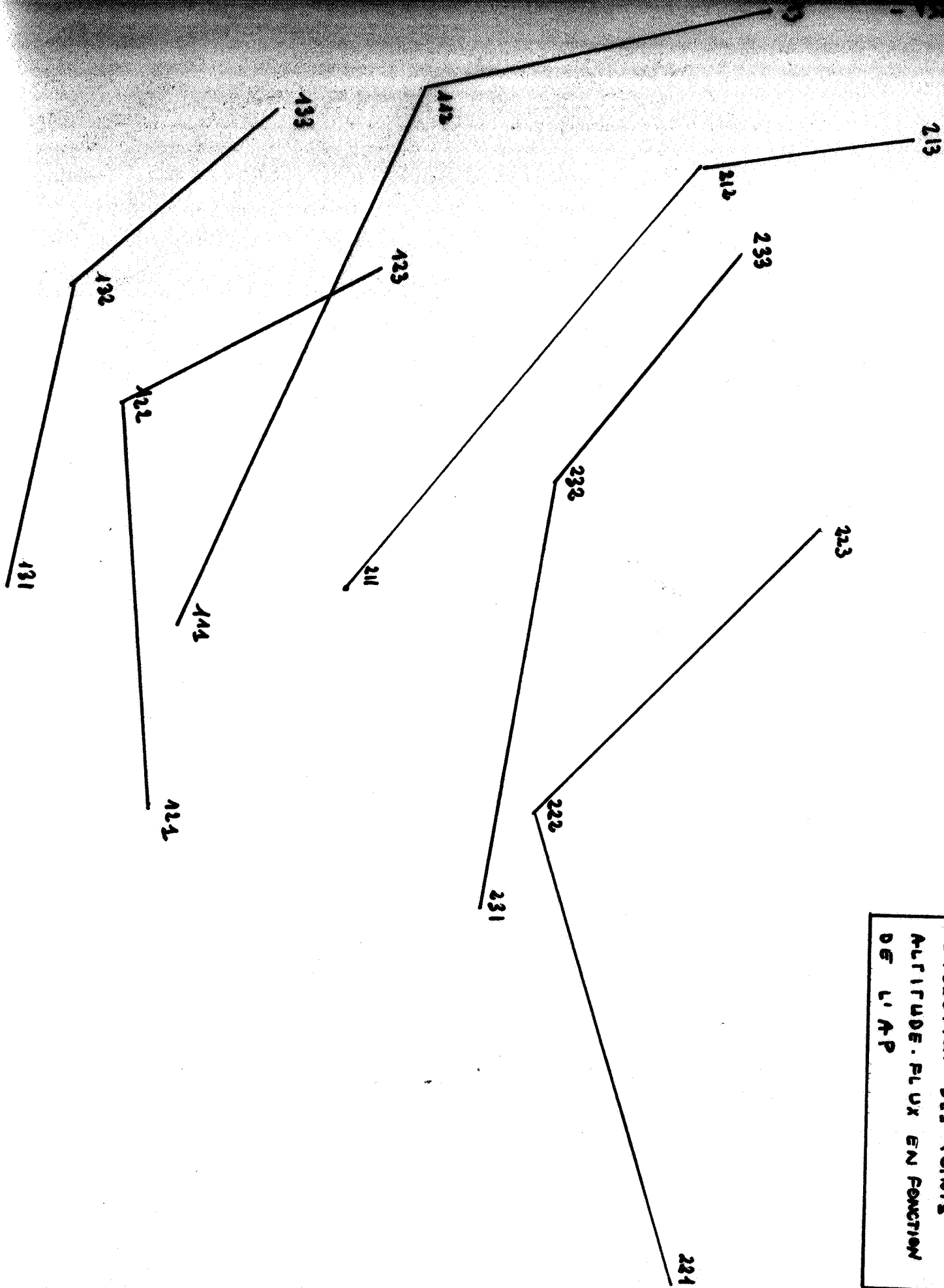
EVOLUTION OBS POINTS
FLUX-AP FW PRODUCTION
DE L'ALTITUDE



EVLUTION DES POINTS
ALTITUDE - AP QUAND
LE FLUX VARIE



EVOLUTION DES POINTS
ALTITUDE. FLUX EN FONCTION
DE L'AP



IV - MODELISATION

Cette analyse semble susceptible d'apporter une modélisation satisfaisante de l'évolution du coefficient correcteur f . Nous avons effectué d'une part une régression linéaire sur les deux premiers facteurs, et d'autre part une méthode de régression aléatoire simplifiée, inspirée directement de l'idée de Régression aléatoire de J.P. BORDET ([2]).

IV-1 - REGRESSION LINEAIRE

La régression donne le résultat suivant :

$$f = 0.950 + 0.113 \varphi_1 + 0.0000003 \varphi_2$$

φ_1 , φ_2 désignant les deux premiers facteurs. Seul le premier facteur apporte donc une information significative pour la régression linéaire. L'amélioration comptable, définie comme le quotient de la différence entre le nombre de points améliorés et le nombre de points dégradés au nombre total de points est de 13 %.

IV-2 - REGRESSION ALEATOIRE

Toute observation est affectée à une des 18 situations de base. Elle est ensuite codée en "polychotomique flou" selon les principes développés par J.P. BORDET. Elle est ensuite projetée en élément supplémentaire dans l'Analyse. Cette opération est assez rapide, car sur les 288 coordonnées affectées à tout point, seules huit sont non nulles. Nous quadrillons ensuite le plan et calculons les valeurs moyennes de f sur les carreaux. Dans une seconde étape, nous affectons à une observation tombée dans un carreau la valeur moyenne du carreau. Nous avons essayé un grand nombre de quadrillages possibles. En fait, ils ont tous donné des résultats équivalents. Avec 600 carreaux, on obtient une amélioration comptable d'environ 20 %.

IV-3 - COMPARAISON DES RESULTATS

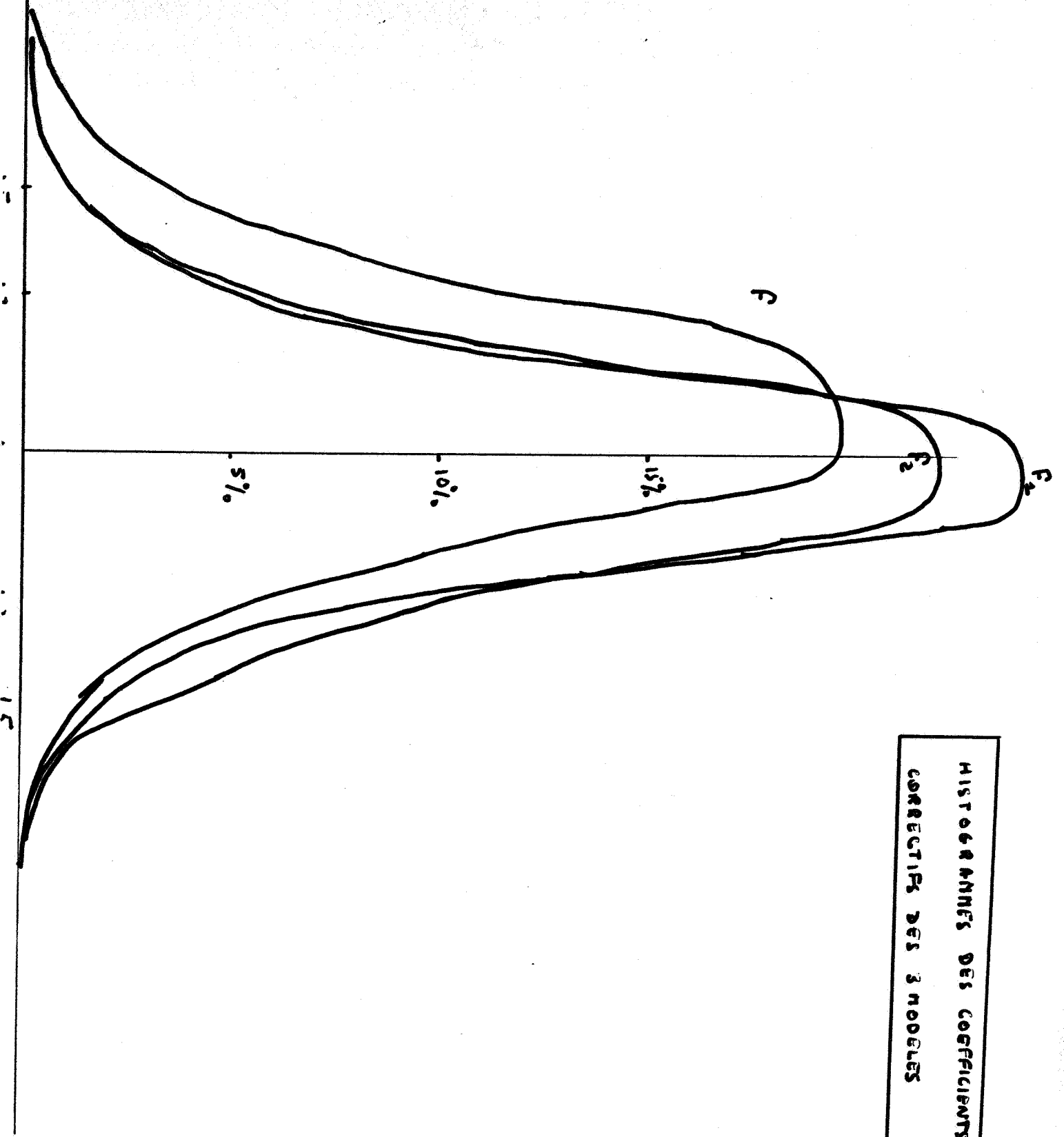
Nous avons tracé, à titre indicatif, les histogrammes des différents prédicteurs

+ histogramme de f , quotient de la densité observée par la densité JACCHIA

+ histogramme de f_1 , quotient de la densité observée par la densité JACCHIA corrigée par régression aléatoire.

+ histogramme de f_2 , quotient de la densité observée par la densité JACCHIA corrigée par régression linéaire.

HISTOGRAMMES DES COEFFICIENTS
CORRECTIFS DES 3 MODELES



CONCLUSION

Isolant les variables importantes des variables secondaires, révélant les influences conjuguées de celles-ci et de celles-là, l'Analyse des Correspondances nous guide à la découverte d'un ordre propre à la haute atmosphère. Grâce à elle, la structure latente devient manifeste. Là s'arrête notre travail : statisticien nous sommes, statisticien nous voulons demeurer. Aux physiciens, aux chimistes de confronter nos constructions à leurs hypothèses, à leurs modèles, à leurs intuitions. C'est d'eux et d'eux seuls, que pourra venir l'explication des phénomènes et des structures observées.

VIII - BIBLIOGRAPHIE

=====

- [1] J.L. FALIN Analyse des variations de la densité de la haute
 atmosphère neutre.

- [2] J.P. BORDET Etude de Données Géophysiques. Modélisations statis-
 tiques par régression factorielle. THESE PARIS 1973.

- [3] J.P. BENZECRI L'Analyse des Données. DUNOD 1973.