



HAL
open science

Application de l'analyse des données à l'exploration pétrolière : caractérisation des facies sismiques par analyses statistiques multivariées

Frédérique Fournier

► **To cite this version:**

Frédérique Fournier. Application de l'analyse des données à l'exploration pétrolière : caractérisation des facies sismiques par analyses statistiques multivariées. Sciences de la Terre. Institut National Polytechnique de Lorraine, 1990. Français. NNT : 1990INPL053N . tel-01776164

HAL Id: tel-01776164

<https://hal.univ-lorraine.fr/tel-01776164v1>

Submitted on 24 Apr 2018

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

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



AVERTISSEMENT

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

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

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

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

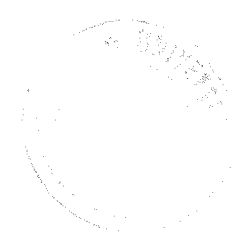
http://www.cfcopies.com/V2/leg/leg_droi.php

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

THÈSE

PRESENTEE
A L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE
POUR OBTENIR
LE GRADE DE DOCTEUR DE L'I.N.P.L.
PAR

Frédérique FOURNIER



Sujet de la thèse :

**APPLICATION DE L'ANALYSE DES DONNEES
A L'EXPLORATION PETROLIERE
CARACTERISATION DES FACIES SISMIQUES
PAR ANALYSES STATISTIQUES MULTIVARIABLES**

Service Commun de la Documentation
INPL
Nancy-Brabois

Soutenue publiquement le 6 juillet 1990 devant la Commission d'Examen :

MM. A. JOURNEL

*Président
et rapporteur*

J.J. ROYER

Directeur

C. CHAMBON

Rapporteur

J. LACAZE

M. WAESELYNCK

} *Examineurs*

No, I don't like work. I had rather laze about and think of all the fine things that can be done. I don't like work, -no man does- but I like what is in the work, -the chance to find yourself. Your own reality -for yourself, not for others- what no other man can ever know. They can only see the mere show, and never can tell what it really means.

Joseph CONRAD, Heart of Darkness.

AVANT-PROPOS

Le travail publié dans ce mémoire reprend en partie les recherches que j'ai menées pour le compte de la Société Nationale Elf Aquitaine Production (S.N.E.A.(P)) durant la période 1984-1988. Je tiens donc avant tout à exprimer ma gratitude à la direction de la S.N.E.A.(P) pour m'avoir permis d'utiliser ces résultats à des fins universitaires.

Je voudrais également témoigner de mon amitié pour Jean Dumay, ingénieur géologue à la S.N.E.A.(P), sans qui le projet "Alias" n'aurait jamais vu le jour et qui a su, tout en m'encourageant, me laisser l'autonomie et la liberté nécessaires à la conduite de tout travail de recherche.

Je remercie vivement Jean-Jacques Royer, pour avoir accepté de diriger ce travail. Ses conseils scientifiques m'ont été très profitables et j'ai grandement apprécié sa patience à l'égard du "vieil" étudiant que je suis, "déformé", de surcroît, par cinq ans d'activités industrielles.

Je souhaite exprimer ma profonde gratitude à André Journel, professeur à l'Université de Stanford, qui, malgré ses nombreuses occupations et l'éloignement géographique, a bien voulu examiner ce travail et m'honorer de sa présence dans ce jury. Je pense qu'André représente particulièrement bien le cadre interdisciplinaire géologie/géophysique de réservoir et géomathématiques, ainsi que le double aspect recherche/applications industrielles dans lesquels j'aimerais tant inscrire ces travaux ainsi que mes recherches futures.

Que Monsieur Chambon, professeur à l'Institut National Polytechnique de Lorraine (I.N.P.L.), soit également remercié pour avoir accepté d'être rapporteur de ce mémoire.

Jean Lacaze, chef de la Division des Recherches Géologiques et Géophysiques de la S.N.E.A.(P), m'a fait le plaisir de juger ce manuscrit du point de vue de l'expert interpréteur sismicien qu'il est. Qu'il trouve l'expression de mon amicale reconnaissance.

Michel Waeselynck, expert géophysicien auprès de la Division des Recherches Géologiques et Géophysiques de la S.N.E.A.(P), retiré depuis peu, a bien voulu apporter son savoir-faire en traitement du signal à l'examen de ce travail. Je l'en remercie ainsi que de toutes les discussions techniques que nous avons eues ensemble qui m'ont permis d'apprécier sa compétence scientifique, sa disponibilité et son extrême gentillesse.

J'adresse mes remerciements à l'Institut Français du Pétrole (I.F.P.) qui a pris en charge la frappe de ce mémoire, sans quoi sa réalisation pratique aurait été très compromise.

Ce travail est dédié à mes amis et à ma famille, plus particulièrement à ma mère Camille-Marie dont la tendresse vigilante m'a toujours accompagnée ainsi qu'à Yvan, qui a sacrifié, pour m'aider, tant de passionnants week-ends de chasse à l'araignée.

SOMMAIRE

INTRODUCTION	1
PREMIERE PARTIE : LES DONNEES SISMIQUES : ACQUISITION, TRAITEMENT ET INTERPRETATION	15
1 - INTRODUCTION	16
2 - LES PRINCIPES DE LA SISMIQUE REFLEXION	17
3 - MISE EN OEUVRE DE LA SISMIQUE REFLEXION	25
3.1 - Principe de mise en oeuvre	25
3.2 - Principales sources utilisées	28
3.3 - Développements récents de la mise en oeuvre de la sismique réflexion	30
4 - TRAITEMENT DES DONNEES DE SISMIQUE REFLEXION	32
4.1 - Introduction	32
4.2 - Traitement des amplitudes	33
4.2.1 - Récupération du gain	33
4.2.2 - Traitement en amplitudes égalisées	33
4.2.3 - Traitement en amplitudes préservées	34
4.3 - Mise en point milieu commun	35
4.4 - Corrections statiques	35
4.4.1 - Amélioration des statiques	36
4.4.2 - Corrections statiques résiduelles	37
4.5 - La déconvolution	38
4.5.1 - Déconvolution non prédictive et prédictive	39
4.5.2 - Evaluation du signal	39
4.5.3 - Problème de la phase	40
4.5.4 - Les paramètres de la déconvolution	40

4.6 - Les analyses de vitesse	41
4.6.1 - Principe de la détermination des corrections dynamiques	41
4.6.2 - Principe de l'analyse des vitesses	46
4.7 - La sommation	46
4.8 - La migration	48
4.9 - Conclusion	50
5 - INTERPRETATION STRUCTURALE DES DONNEES SISMIQUES	51
6 - INTERPRETATION STRATIGRAPHIQUE DES DONNEES SISMIQUES	52
6.1 - Introduction	52
6.2 - Processus d'interprétation	52
6.2.1 - Reconnaissance des différentes séquences sismiques	52
6.2.2 - Détermination de l'âge des séquences	55
6.2.3 - Construction d'une section chronostratigraphique	55
6.2.4 - Analyse des faciès sismiques	57
6.3 - Séquences sismiques et séquences stratigraphiques	57
6.4 - Analyse des faciès sismiques	63
6.4.1 - Configuration des réflexions	63
6.4.2 - Forme externe	63
6.4.3 - Paramètres internes	65
6.4.4 - Conclusions	69
7 - INTERPRETATION STRATIGRAPHIQUE OU LITHOLOGIQUE ET ANALYSE AUTOMATIQUE	73

**DEUXIEME PARTIE : ANALYSE DU FACIES SISMIQUE SUR
UNE FENETRE TEMPORELLE
D'ENVIRON 100 MS** 76

1 - POSITION DU PROBLEME 77

1.1 - Introduction 77

1.2 - Etat de l'art 77

1.2.1 - Analyse d'attributs sismiques classiques 77

1.2.2 - Utilisation plus quantitative d'attributs sismiques classiques : vitesses
d'intervalle, amplitudes, fréquences instantanées 78

1.2.3 - Etude quantitative d'attributs mesurés sur l'autocorrélation et son
spectre 79

1.3 - Critique des variables utilisées 83

**2 - CARACTERISATION D'UN ENVIRONNEMENT GEOLOGIQUE PAR LA
FONCTION D'AUTOCORRELATION ET SON SPECTRE** 86

2.1 - Introduction 86

2.2 - Présentation du modèle markovien 86

2.3 - Autocorrélation d'une séquence sédimentaire markovienne de premier ordre 89

2.3.1 - Modèle utilisé 89

2.3.2 - Evaluation du temps de transit dans une couche de lithologie i 89

2.3.3 - Evaluation de la probabilité d'observer au temps t une transition $k \rightarrow l$ 90

2.3.4 - Evaluation de la fonction d'autocorrélation des coefficients de réflexion 92

2.3.5 - Cas de deux lithologies 93

2.3.5.1 - Calcul de l'autocorrélation et de son spectre 93

- Calcul de la matrice $\left\{ \tilde{\pi}(s)^{-1} \cdot T \right\}^{-1}$
- Calcul de $R(t)$
- Calcul de la valeur de l'autocorrélation au temps $t = 0$
- Calcul du spectre de puissance

2.3.5.2 - Etude des courbes $R(t)$ et $P(f)$ 96

2.3.5.3 - Influence des divers paramètres 98

- Influence de $\left(\frac{\sigma_{12}}{\tau_{12}} \right)$
- Influence du paramètre γ

2.3.5.4 - Réalisation d'abaques	99
2.3.6 - Cas de plus de deux lithologies	104
2.3.6.1 - Evaluation de l'autocorrélation et de son spectre d'amplitude	104
2.3.6.2 - Calcul de la valeur $R(0)$	106
2.3.7 - Conclusion - Discussion	106
2.4 - Les environnements géologiques du delta du Niger comme cadre d'application de certaines méthodes	107
2.4.1 - Ensembles de lithofaciès	107
2.4.2 - Principales séquences lithologiques	109
2.4.2.1 - La séquence négative littorale	109
2.4.2.2 - La séquence positive de chenal	112
2.4.2.3 - La séquence positive transgressive	113
2.4.2.4 - Association de séquences	113
2.4.3 - Application à l'étude des séries rencontrées dans les forages du delta du Niger	113
2.5 - Application des résultats théoriques sur l'autocorrélation au delta du Niger	116
2.5.1 - Cas de deux lithologies	120
2.5.2 - Cas de quatre lithologies	122
2.6 - Conclusions	130
3 - ANALYSE DE LA REPOSE SISMIQUE DES ENVIRONNEMENTS DU DELTA DU NIGER	136
3.1 - Introduction	136
3.2 - Aspect sismique des différents groupes de dépôt	138
3.2.1 - Groupe plate-forme	138
3.2.2 - Groupe front de delta	141
3.2.3 - Groupe margino-littoral : chenaux et plaine d'inondation	141
3.2.4 - Bilan sismique des groupes	146
3.3 - Calcul de paramètres sismiques pour la caractérisation des traces à partir de leurs représentations temporelles et fréquentielles	146
3.3.1 - Variables calculées sur l'autocorrélation	151
3.3.2 - Variables sur le spectre d'amplitude	152

3.3.3 - Variables sur la trace réelle	153
3.3.4 - Variables sur le signal analytique	158
3.3.5 - Récapitulation	160
3.4 - Pouvoir séparateur des variables	161
3.4.1 - Analyses discriminantes sur 43 traces les plus typiques des divers groupes	161
3.4.2 - Recherche des variables les plus discriminantes (test de Student)	166
3.4.3 - Analyse factorielle discriminante sur les facteurs d'une analyse en composantes principales	169
3.4.4 - Analyse factorielle discriminante pas à pas	171
3.5 - Conclusion à l'analyse des traces synthétiques des puits du delta du Niger	180
3.6 - Caractérisation de traces sismiques réelles	181
3.7 - Conclusion	194
4- MISE AU POINT D'UNE METHODOLOGIE MULTIVARIABLE POUR LA CARACTERISATION DES FACIES SISMIQUES	196
4.1 - Introduction	196
4.2 - Rappels sur les méthodes utilisées	198
4.2.1 - Analyse en composantes principales	201
4.2.1.1 - Nature du problème	201
4.2.1.2 - Solution au problème	203
4.2.1.3 - Définition de l'analyse en composantes principales	204
4.2.1.4 - Représentations graphiques et interprétation	205
4.2.1.4.1 - Construction d'une image du nuage de points initial	205
a - Image fidèle	
b - Image normalisée	
4.2.1.4.2 - Représentations graphiques	206
a - Représentation des échantillons	
b - Représentation des variables	
4.2.1.4.3 - Remarques sur l'interprétation des résultats de l'A.C.P.	207
a - Choix des données et de leur poids dans l'analyse	

	<ul style="list-style-type: none"> • Choix des variables initiales • Point de vue des échantillons 	
	b - Analyse des tableaux et des représentations graphiques	
	<ul style="list-style-type: none"> • Pourcentage de variance exprimée par chaque facteur et pourcentage de variance cumulée • Vecteurs propres • Facteur "forme" • Etude des corrélations simples et multiples • Analyse graphique 	
4.2.2	- Analyse factorielle discriminante	210
	4.2.2.1 - Position du problème	210
	<ul style="list-style-type: none"> a - Discrimination à but descriptif b - Discrimination à but décisionnel ou identification 	
	4.2.2.2 - Discrimination linéaire	212
	<ul style="list-style-type: none"> a - Notations et définition b - Recherche des axes factoriels discriminants c - Retour sur la notion de discrimination linéaire 	
	4.2.2.3 - Méthode de classement	216
	4.2.2.4 - Procédures de pas à pas	217
	<ul style="list-style-type: none"> a - Définition et intérêt des procédures de pas à pas b - Pourcentage de bien classés c - Trace ($T^{-1} B$) d - Autres critères pour le pas à pas 	
4.2.3	- Classifications automatiques	221
	4.2.3.1 - Introduction	221
	4.2.3.2 - Classifications hiérarchiques	223
	<ul style="list-style-type: none"> a - Généralités b - Classification hiérarchique basée sur le critère de l'inertie expliquée (Ward) c - Limitations des méthodes de classification hiérarchique 	
	4.2.3.3 - Méthodes non hiérarchiques ou de partitionnement	225
	<ul style="list-style-type: none"> a - Généralités 	

b - Méthode des "k-means" ou algorithme de "réallocation-recentrage" autour des centres mobiles de gravité	
c - Méthode des "nuées dynamiques"	
- Choix d'un mode de représentation	
- Déroulement de l'algorithme	
d - Recherche des classes typiques par les "formes fortes"	
4.2.4 - Conclusions	230
4.3 - Méthodologie statistique développée	230
4.3.1 - Contraintes à prendre en compte	230
4.3.2 - Cas où il existe des puits	231
4.3.2.1 - Définition de traces sismiques d'apprentissage	233
4.3.2.2 - Validation du choix des paramètres discriminants	235
4.3.2.3 - Etape prédictive	237
4.3.2.4 - Validation des faciès sismiques par classifications automatiques	239
4.3.2.5 - Synthèse de la méthodologie (information géologique disponible)	239
4.3.3 - Cas où il n'y a pas d'information géologique disponible	241
4.4 - Discussion sur les limitations de la méthodologie proposée. Conclusions	241
5 - ETUDES DE CAS	255
5.1 - Introduction	255
5.2 - Etude des faciès sismiques du Dogger carbonaté	255
5.2.1 - Cadre géologique	255
5.2.2 - Phase d'apprentissage	255
5.2.2.1 - Sélection des traces de référence	255
5.2.2.2 - Sélection des paramètres discriminants	258
5.2.2.3 - Effet de la longueur de la fenêtre sur le choix des variables discriminantes	259
5.2.2.4 - Effet du choix des traces de référence sur la sélection des variables discriminantes	265
5.2.2.5 - Conclusion	266
5.2.3 - Phase de prédiction	267

5.2.4 - Phase confirmatoire à l'aide des techniques de classification	267
5.2.5 - Comparaison des résultats avec les interprétations antérieures	269
5.3 - Etude des faciès sismiques en environnement turbiditique argilo-gréseux	273
5.3.1 - Cadre géologique	273
5.3.2 - Données sismiques et calages aux puits	273
5.3.2.1 - Données sismiques étudiées	273
5.3.2.2 - Analyse de traces synthétiques	275
5.3.3 - Phase d'apprentissage	280
5.3.3.1 - Choix des traces de référence	280
5.3.3.2 - Etude des traces de référence	285
a - Calcul de l'autocorrélation et du spectre d'amplitude	
b - Calcul des paramètres sismiques	
c - Analyses factorielles	
5.3.4 - Etape de modélisation pour la validation géologique des paramètres discriminants	293
5.3.4.1 - Premier modèle	293
5.3.4.2 - Influence respective des ensembles inter R1/R2, R2 et R3 sur la réponse sismique du modèle	297
a - Modèle 2	
b - Modèle 3	
5.3.4.3 - Conclusions de la modélisation	310
5.3.5 - Etape prédictive	310
5.3.5.1 - Tests sur la ligne 152	311
5.3.5.2 - Cartes d'extension des faciès étudiés	311
5.3.5.3 - Contrôle avec les puits	312
5.3.6 - Conclusion	315
5.4 - Conclusion aux deux études de cas présentés	315

**TROISIEME PARTIE : ADAPTATION DE LA METHODOLOGIE
DE RECONNAISSANCE DES FACIES
SISMIQUES POUR UNE COURTE
FENETRE TEMPORELLE** 316

1 - INTRODUCTION 317

**2 - UTILISATION DE LA REPRESENTATION TEMPS-FREQUENCE POUR
LA CARACTERISATION DES STRUCTURES FINES SISMIQUES** 318

2.1 - Généralité sur la représentation conjointe temps-fréquence (R.C.T.F.) 318

2.1.1 - Définition 318

2.1.2 - Propriétés de la R.W.V. 320

2.1.3 - Etude numérique de la R.W.V. 321

2.1.4 - Discrétisation de la représentation de Wigner-Ville 322

2.2 - R.W.V. d'un signal modulé linéairement en fréquence : influence de la fenêtre de pondération 323

2.3 - Sensibilité et caractère local de la R.W.V. 328

2.3.1 - Sensibilité 328

2.3.1.1 - Sensibilité à une variation de vitesse 328

2.3.1.2 - Sensibilité à une variation temporelle 333

2.3.2 - Caractère local 333

2.4 - Etude des caractéristiques sismiques du Callovien Inférieur avec la R.W.V. 333

2.4.1 - Position du problème 333

2.4.2 - Description 339

2.4.3 - Calcul des R.W.V. et de leurs différences 341

2.4.4 - Fluctuation autour des modèles 344

2.4.4.1 - Transition oolites sommitales-argiles bioclastiques 344

2.4.4.2 - Base des calcaires bioclastiques 344

2.4.5 - Composition des modèles de base 344

2.4.6 - Comparaison entre les modèles et les traces synthétiques 344

2.4.7 - Conclusions à l'étude des modèles 348

2.4.8 - Etude des R.C.T.F. des traces synthétiques 350

2.4.8.1 - Analyse en composantes principales des paramètres précédents	350
2.4.8.2 - Analyse factorielle discriminante	350
- Argiles bioclastiques	
- Calcaires bioclastiques	
2.4.8.3 - Interprétation des résultats	352
2.5 - Conclusion sur l'application de la R.W.V. pour la caractérisation fine de la trace	352
3 - ANALYSE SPECTRALE PAR MAXIMISATION DE L'ENTROPIE	353
3.1 - Position du problème	353
3.2 - Méthode de l'entropie maximale	355
3.2.1 - Méthode de Yule-Walker	355
3.2.2 - Principe de l'entropie maximale	357
3.2.3 - Détermination de l'ordre du modèle autorégressif	360
3.3 - Algorithmes d'estimation des paramètres autorégressifs	362
3.3.1 - Algorithme de Yule-Walker	362
3.3.2 - Algorithme de Burg	363
3.3.3 - Algorithme de Barrodale-Erickson	364
3.3.4 - Algorithme de Marple	364
3.3.5 - Récapitulatif sur les algorithmes	364
3.4 - Essais sur des traces synthétiques	365
3.5 - Comparaison des résultats obtenus avec ceux des études antérieures	377
3.5.1 - Résultats antérieurs	377
3.5.2 - Comparaison des résultats antérieurs avec les nôtres	381
4 - CARACTERISATION DE LA TRACE PAR LA SUITE DES AMPLITUDES	384
4.1 - Principe	384
4.2 - Tests de faisabilité de la méthode	386
4.2.1 - Données analysées	386
4.2.2 - Premier essai de l'analyse en composantes principales	388
4.2.3 - Choix des paramètres	393

4.2.3.1 - Influence du pas d'échantillonnage	393
4.2.3.2 - Nombre de facteurs à retenir	393
4.2.3.3 - Diminution de la taille de la fenêtre	396
4.2.3.4 - Effet d'une erreur de pointé	399
4.3 - Conclusions	402
5 - COMPARAISON DE L'UTILISATION DE LA METHODE SPECTRALE PAR MAXIMISATION DE L'ENTROPIE ET DE L'A.C.P. POUR LA CARACTERISATION DES TRACES SISMIQUES	404
5.1 - Introduction	404
5.2 - Choix de l'ordre du modèle	404
5.3 - Classifications	404
5.4 - Conclusion	407
6 - METHODOLOGIE STATISTIQUE SUR "FENETRE COURTE"	409
6.1 - Description de la méthodologie statistique	409
6.1.1 - Introduction	409
6.1.2 - Cas où une information géologique est disponible	410
6.1.3 - Cas où aucune information géologique n'est disponible	412
6.2 - Discussion des limites de la méthode	412
6.3 - Une application de la méthodologie à un problème de réservoir	413
6.3.1 - Cadre géologique de l'étude	413
6.3.2 - Données disponibles	417
6.3.2.1 - Données géologiques	417
6.3.2.2 - Données sismiques	420
6.3.3 - Etude des traces sismiques synthétiques	422
6.3.3.1 - Analyse des traces sismiques synthétiques des puits	426
a - Etude préliminaire	
b - Analyse des traces synthétiques sur une fenêtre de 28 ms	
• Analyse en composantes principales	
• Essai de classification	

c - Analyse des traces synthétiques sur une fenêtre de 35 ms	
. Analyse en composantes principales	
. Classification à 5 groupes	
d - Comparaison des classifications	
6.3.3.2 - Analyse des traces sismiques modèles	438
6.3.3.3 - Conclusion à l'analyse des traces synthétiques	438
6.3.4 - Etude des traces réelles	440
6.3.4.1 - Représentation des traces dans l'espace sismique	440
6.3.4.2 - Classifications automatiques	440
6.3.4.3 - Analyse factorielle discriminante	440
6.3.4.4 - Affectation des traces synthétiques aux groupes de traces réelles d'apprentissage	442
- Homogénéisation des données	
- Affectation des traces synthétiques	
6.3.5 - Synthèse de l'analyse du domaine sismique	446
7 - CONCLUSION	447
<i>CONCLUSION GENERALE</i>	450
ANNEXES	457
. Multivariate statistical analyses applied to seismic facies recognition, J. Dumay, F. Fournier, 1988.	
. Extraction of quantitative geologic information from seismic data with multidimensional statistical analyses.	
Part I : methodology	
Part II : a case history	
F. Fournier, 1989	

LISTE DES FIGURES

- FIGURE 1 :** Géométrie des objets géologiques.
a - Coupe transversale du Gisement de Kangan, Iran (d'après Reyre, 1975).
b - Carte et coupe du champ de Groningue , Pays-Bas (d'après Stauble et Milos, 1970).
- FIGURE 2 :** Interprétation stratigraphique et lithologique de la sismique en complexe deltaïque.
- FIGURE 3 :** Coupe verticale schématique montrant les différents types de pièges à hydrocarbures.
- FIGURE 4 :** Exemple d'interprétation structurale de la sismique.
- FIGURE 5 :** Exemple d'interprétation stratigraphique de la sismique.
- FIGURE 6 :** Exemple d'interprétation lithologique de la sismique.
- FIGURE 7 :** Interprétation stratigraphique en environnement fluvio-deltaïque.
- FIGURE 8 :** Principe des méthodes sismiques.
- FIGURE 9 :** Principe du film synthétique.
- FIGURE 10 :** Mise en oeuvre d'une couverture simple et d'une couverture multiple.
- FIGURE 11 :** Filtrage des bruits organisés.
- FIGURE 12 :** Atténuation des réflexions multiples par la couverture multiple.
- FIGURE 13 (a, b, c) :**
Temps d'arrivée des événements sismiques.
- FIGURE 14 :** Principe des analyses de vitesse.
- FIGURE 15 :** Nécessité de la migration.
- FIGURE 16 :** Ordonnancement d'une étude de sismique stratigraphique.

FIGURE 17 : Nomenclature des terminaisons de réflexion.

FIGURE 18 : Notion de séquence de dépôt.

FIGURE 19 : Relation entre les strates aux frontières des séquences.

FIGURE 20 : Exemple de répartition spatiale des divers types de terminaisons latérales.

FIGURE 21 (a, b, c) :

Différentes évolutions sédimentaires en fonction des variations relatives du niveau de la mer.

FIGURE 22 : Configuration des réflexions.

FIGURE 23 : Evolution de l'amplitude en fonction du pourcentage d'argiles de la série modélisée.

FIGURE 24 :

et

Effet des bancs minces.

FIGURE 25 :

FIGURE 26 : Liaison entre les épaisseurs et les fréquences apparentes.

FIGURE 27 : Enchaînement séquentiel vertical des unités de faciès sismique.

FIGURES 28 A 31 :

Courbes du spectre de puissance paramétrées par γ

$$\text{Figure 28 : } \frac{\sigma_{12}}{r_{12}} = 0.1$$

$$\text{Figure 29 : } \frac{\sigma_{12}}{r_{12}} = 0.2$$

$$\text{Figure 30 : } \frac{\sigma_{12}}{r_{12}} = 0.3$$

$$\text{Figure 31 : } \frac{\sigma_{12}}{r_{12}} = 0.6$$

FIGURE 32 : Séquence paysage du delta du Niger.

- FIGURE 33 :** Séquences élémentaires observées dans le delta du Niger.
- FIGURE 34 :** Association de séquences.
- FIGURE 35 :** Expression sismique de la zone de plate-forme.
- FIGURE 36 :** Expression sismique de la zone de front de delta.
- FIGURE 37 :** Expression sismique de la zone margino-littorale.
- FIGURE 38 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
2 lithologies, hypothèse moyenne sur les épaisseurs.
- FIGURE 39 :** Spectre de puissance du modèle markovien appliqué aux série du delta du Niger.
2 lithologies, hypothèse maximale sur les épaisseurs .
- FIGURE 40 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, épaisseurs constantes égales à 10 m.
- FIGURE 41 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, épaisseurs constantes égales à 20 m.
- FIGURE 42 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, épaisseurs constantes égales à 30 m.
- FIGURE 43 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, épaisseurs constantes égales à 40 m.
- FIGURE 44 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, épaisseurs constantes égales à 50 m.
- FIGURE 45 :** Spectre d'amplitude normé du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, hypothèse moyenne sur les épaisseurs.

- FIGURE 46 :** Spectre d'amplitude normé du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, hypothèse maximale sur les épaisseurs.
- FIGURE 47 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, hypothèse moyenne sur les épaisseurs.
- FIGURE 48 :** Spectre de puissance du modèle markovien appliqué aux séries du delta du Niger.
4 lithologies, hypothèse maximale sur les épaisseurs.
- FIGURE 49 :** Signal utilisé dans le calcul des traces synthétiques.
- FIGURE 50 :** Echantillon du groupe plate-forme : log de vitesse et traces synthétiques impulsionnelle et filtrée.
- FIGURE 51 :** Autocorrélation de la trace synthétique filtrée de l'échantillon du groupe plate-forme.
- FIGURE 52 :** Echantillon du groupe front de delta : log de vitesse et traces synthétiques impulsionnelle et filtrée.
- FIGURE 53 :** Echantillon du groupe front de delta : log de vitesse et traces synthétiques impulsionnelle et filtrée.
- FIGURE 54 :** Echantillon du groupe margino-littoral : log de vitesse et traces synthétiques impulsionnelle et filtrée.
- FIGURE 55 :** Echantillon du groupe margino-littoral : log de vitesse et traces synthétiques impulsionnelle et filtrée.
- FIGURE 56 :** Traces sismiques synthétiques des 3 groupes.
- FIGURE 57 (a, b) :**
Autocorrélation des deux échantillons du groupe margino-littoral présentés.

FIGURE 58 (a, b) :

Autocorrélation des deux échantillons du groupe front de delta présentés.

FIGURE 59 (a, b) :

Spectre d'amplitude et sa courbe cumulée pour un échantillon du groupe plate-forme.

FIGURE 60 : Courbe cumulée des amplitudes du spectre pondérées par les fréquences pour un échantillon du groupe plate-forme.

FIGURE 61 : Variables calculées sur la trace sismique.

FIGURE 62 : Variables calculées sur le module du signal analytique.

FIGURE 63 : Analyse factorielle discriminante - 43 traces synthétiques - variables sur le spectre d'amplitude.

FIGURE 64 : Analyse factorielle discriminante - 43 traces synthétiques - variables sur l'autocorrélation.

FIGURE 65 : Analyse factorielle discriminante - 43 traces synthétiques - variables sur le module du signal analytique.

FIGURE 66 : Analyse factorielle discriminante - 43 traces synthétiques - variables des trois familles.

FIGURE 67 : Analyse factorielle discriminante - 88 traces synthétiques - 10 variables sélectionnées par le test de Student.

FIGURE 68 : Analyse factorielle discriminante sur les composantes principales d'une A.C.P. - 43 traces synthétiques.

FIGURE 69 : Analyse factorielle discriminante sur les composantes de 3 A.C.P. (famille par famille de variables) - 43 traces synthétiques.

FIGURE 70 : Analyse factorielle discriminante pas à pas - 88 traces synthétiques.

FIGURE 71 : Analyse factorielle discriminante pas à pas - 43 traces synthétiques.

FIGURE 72 : Section sismique 12 recoupant les faciès du delta du Niger.

- FIGURE 73 :** Analyse factorielle discriminante pas à pas - traces sismiques réelles du profil 12.
- FIGURE 74 :** Section sismique 19 recoupant les faciès du delta du Niger.
- FIGURE 75 :** Analyse factorielle discriminante pas à pas - traces sismiques réelles du profil 19.
- FIGURE 76 :** Variations géologiques latérales d'un réservoir carbonaté.
- FIGURE 77 :** Principales méthodes statistiques multivariées utilisées.
- FIGURE 78 :** Analyse en composantes principales.
- FIGURE 79 :** Analyse factorielle discriminante.
- FIGURE 80 :** Classifications automatiques.
- FIGURE 81 :** Méthodologie statistique Alias A.
- FIGURE 82 :** Analyse factorielle discriminante pas à pas sur les traces d'apprentissage.
- FIGURE 83 :** Analyse en composantes principales sur les traces d'apprentissage.
- FIGURE 84 :** Carte de faciès sismique obtenue par analyse factorielle discriminante - 5 groupes d'apprentissage.
- FIGURE 85 :** Carte de faciès sismique obtenue par classification automatique - 5 groupes.
- FIGURE 86 :** Affectation des traces par analyse discriminante sur la ligne 23 - faciès bon réservoir et réservoir moyen à mauvais.
- FIGURE 87 :** Carte du faciès géologique déduite de l'analyse sismique - faciès bon réservoir opposé au faciès réservoir moyen à mauvais.
- FIGURE 88 :** Données géologiques correspondant à la première étude de cas - Dogger carbonaté du Bassin de Paris.
- FIGURE 89 :** Traces d'apprentissage des pôles "dune oolithique" et "platier".

- FIGURE 90 :** Enveloppes et cumuls des enveloppes pour les traces d'apprentissage des pôles dune et platier.
- FIGURE 91 :** Traces moyennées latéralement joignant les deux pôles d'apprentissage.
- FIGURE 92 :** Evolution de la variable E2 sur la section 12 joignant les deux pôles - fenêtre de 120 ms.
- FIGURE 93 :** Traces d'apprentissage pour une fenêtre d'étude de 60 ms.
- FIGURE 94 :** Evolution de la variable E2 sur la section 12 joignant les deux pôles - fenêtre de 60 ms.
- FIGURE 95 :** Affectation des traces de la ligne 12 - différents essais en analyse discriminante et en classification.
- FIGURE 96 :** Carte des faciès sismiques sur la zone étudiée.
- FIGURE 97 :** Réponse en impédance acoustique des deux faciès géologiques étudiés.
- FIGURE 98 :** Faciès sismiques déterminés par la méthodologie statistique.
- FIGURE 99 :** Comparaison des unités réservoirs aux puits T1 et M1.
- FIGURE 100 :** Calage des données sismiques avec les données de puits.
- FIGURE 101 :** Portions de traces étudiées autour des puits T1 et M1.
- FIGURE 102 :** Traces synthétiques du puits T1.
- FIGURE 103 :** Traces synthétiques du puits M1.
- FIGURE 104 :** Traces et moyennes des traces d'apprentissage pour le faciès T.
- FIGURE 105 :** Etude de la moyenne des traces d'apprentissage du faciès T.
- FIGURE 106 :** Traces et moyennes des traces d'apprentissage pour le faciès M.
- FIGURE 107 :** Etude de la moyenne des traces d'apprentissage du faciès M.

- FIGURE 108 :** Faciès T : autocorrélation et spectre de la trace synthétique du puits T1 ainsi que de la moyenne des traces d'apprentissage.
- FIGURE 109 :** Faciès M : autocorrélation et spectre de la trace synthétique du puits M1 ainsi que de la moyenne des traces d'apprentissage.
- FIGURE 110 :** Evolution de la variable ETNT entre les deux pôles d'apprentissage - ligne 152.
- FIGURE 111:** Evolution de la variable AM1 entre les deux pôles d'apprentissage - ligne 152.
- FIGURE 112 :** Evolution de la variable Q4 entre les deux pôles d'apprentissage - ligne 152.
- FIGURE 113 :** Evolution de la variable Q8 entre les deux pôles d'apprentissage - ligne 152.
- FIGURE 114 :** Modèle 1 d'évolution latérale entre les puits M1 et T1.
- FIGURE 115 :** Comparaison trace modèle 1 et trace synthétique au puits T1.
- FIGURE 116 :** Comparaison trace modèle 1 et trace synthétique au puits M1.
- FIGURE 117 :** Autocorrélation des traces modèles pour les faciès T et M.
- FIGURE 118 :** Evolution de la variable ETNT sur le modèle 1.
- FIGURE 119 :** Evolution de la variable AM1 sur le modèle 1.
- FIGURE 120 :** Evolution de la variable Q4 sur le modèle 1.
- FIGURE 121 :** Evolution de la variable Q8 sur le modèle 1.
- FIGURE 122 :** Modèle 2 d'évolution latérale entre les puits M1 et T1.
- FIGURE 123 :** Modèle 3 d'évolution latérale entre les puits M1 et T1.
- FIGURE 124 :** Caractéristiques des traces des puits T1 et M1 pour le modèle 3.
- FIGURE 125 :** Evolution de la variable ETNT sur le modèle 3.

FIGURE 126 : Evolution de la variable AM1 sur le modèle 3.

FIGURE 127 : Evolution de la variable Q4 sur le modèle 3.

FIGURE 128 : Evolution de la variable Q8 sur le modèle 3.

FIGURE 129 : Carte des faciès T et M avec l'analyse discriminante : variables AM1, Q4.

FIGURE 130 : Comparaison des deux types de faciès différenciés par l'étude.

FIGURE 131 : Représentation temps - fréquence de Wigner-Ville de deux signaux.
a - Signal modulé linéairement en fréquence.
b - Trace sismique réelle.

FIGURE 132 : R.W.V. d'un signal modulé linéairement en fréquence 10-70 Hz.
a - Fenêtre de pondération de 256.
b - Fenêtre de pondération de 64.

FIGURE 133 : R.W.V. d'un signal modulé linéairement en fréquence 10-30 Hz.
a - Fenêtre de pondération de 256.
b - Fenêtre de pondération de 64.

FIGURE 134 : R.W.V. d'une portion de trace sismique.
a - Fenêtre de pondération de 256.
b - Fenêtre de pondération de 64.

FIGURE 135 : Sensibilité de la R.W.V. à une variation de vitesse.

FIGURE 136 : Sensibilité de la R.W.V. à une variation de vitesse - différences normée et non normée des R.W.V. de la figure 135.

FIGURES 137 A 140 :

R.W.V. d'une portion de trace sismique : étude de l'influence des échantillons extérieurs à la fenêtre d'analyse.

FIGURE 141 : Principaux ensembles du Bathonien et Callovien Inférieur du Bassin de Paris.

FIGURE 142 : Modèles de l'ensemble bioclastique étudiés avec la R.W.V.

- FIGURE 143 :** R.W.V. des 4 modèles étudiés.
- FIGURE 144 :** Différences normées des R.W.V. des modèles étudiés.
- FIGURE 145 :** R.W.V. de modèles variant au niveau de la transition oolithes sommitales/argiles bioclastiques.
- FIGURE 146 :** R.W.V. de modèles variant au niveau de la transition calcaires bioclastiques/Bathonien.
- FIGURE 147 :** R.W.V. de la composition des modèles de base.
- FIGURE 148 :** Différences des R.W.V. des traces sismiques synthétiques aux puits étudiés et du modèle de base.
- FIGURE 149 :** Paramètres mesurés sur les R.C.T.F. des traces synthétiques étudiées.
- FIGURE 150 :** Analyse en composantes principales des paramètres mesurés sur les R.C.T.F.
- FIGURE 151 :** Spectre de puissance : algorithme de Yule-Walker (N=50).
- FIGURE 152 :** Spectre de puissance : algorithme de Burg (N=50).
- FIGURE 153 :** Spectre de puissance : algorithme de Marple (N=50).
- FIGURE 154 :** Spectre de puissance : algorithme de Burg (N varie, M = 7)
- FIGURE 155 :** Spectre de puissance : algorithme de Marple (N varie, M = 7).
- FIGURE 156 :** Spectre de puissance : FFT ($\Delta t = 2$ ms, N varie).
- FIGURE 157 :** Spectre de puissance : algorithme de Marple.
- FIGURE 158 :** Spectre de puissance : FFT ($\Delta t = 2$ ms, N varie).
- FIGURE 159 :** Comparaison algorithme de Burg/FFT (d'après Ulrych, 1972).
- FIGURE 160 :** Essais de la méthode d'analyse spectrale par l'entropie (d'après Kane et Trivedi, 1982).

- FIGURE 161 :** Utilisation des coefficients autorégressifs dans un algorithme de reconnaissance des formes (d'après Bois, 1980, 1981, 1982).
- FIGURE 162 :** Principe de la caractérisation des traces à partir de leur représentation temporelle.
- FIGURE 163 :** Comparaison des Sonics des puits XG1, XG2, XP1, XP2 - Bassin de Paris.
- FIGURE 164 :** Impédances acoustiques du modèle d'évolution latérale entre les puits XG2, XG1, XP1, XP2.
- FIGURE 165 :** Section sismique synthétique du modèle d'évolution latérale entre les puits XG2, XG1, XP1, XP2.
- FIGURE 166 :** A.C.P. sur la fenêtre 892-932 ms - distribution de l'inertie sur les facteurs.
- FIGURE 167 :** Classification des traces du modèle avec les facteurs de l'A.C.P.
- FIGURE 168 :** A.C.P. sur la fenêtre 892-932 ms - influence du pas d'échantillonnage.
- FIGURE 169 :** Classification des traces du modèle avec les facteurs de l'A.C.P. - influence du pas d'échantillonnage.
- FIGURE 170 :** A.C.P. étude de l'influence de la taille de la fenêtre.
- FIGURE 171 :** Classification des traces du modèle avec les facteurs de l'A.C.P. - influence de la taille de la fenêtre.
- FIGURE 172 :** A.C.P. étude de l'influence d'une erreur de pointé.
- FIGURE 173 :** Classification des traces du modèle avec les facteurs de l'A.C.P. - influence d'une erreur de pointé.
- FIGURE 174 :** Spectre de puissance des traces synthétiques 7 et 57 - algorithme de Marple.
- FIGURE 175 :** Classification des traces du modèle avec les 4 coefficients autorégressifs.

FIGURE 176 : Réservoirs rencontrés aux puits YP6 et Y142 de l'étude gisement.

FIGURE 177 : A.C.P. sur les données de puits.

FIGURE 178 : Classification en 4 groupes des puits.

FIGURE 179 (a, b, c) :

Comparaison des films synthétiques des puits et des traces modèles pour les 3 principales zones de faciès géologique.

FIGURE 180 : Principaux faciès géologiques et leur réponse sismique.

FIGURE 181 : Distribution spatiale de la durée temps double correspondant à l'intervalle A.

FIGURE 182 : Classification à 5 groupes des traces synthétiques des puits - fenêtre de 28 ms.

FIGURE 183 (a, b) :

Groupes de traces synthétiques déterminés par classification automatique.

LISTE DES TABLEAUX

- TABLEAU 1 :** Différents types de terminaison des réflexions.
- TABLEAU 2 :** Delta du Niger - séquence négative littorale.
- TABLEAU 3 :** Delta du Niger - séquence positive de chenal.
- TABLEAU 4 :** Delta du Niger - séquence de chenal en amont du delta.
- TABLEAU 5 :** Delta du Niger - épaisseurs et paramètres γ de la zone margino-littorale.
- TABLEAU 6 :** Delta du Niger - épaisseurs et paramètres γ de la zone du front de delta.
- TABLEAU 7 :** Delta du Niger - épaisseurs et paramètres γ de la zone de plate-forme interne.
- TABLEAU 8 :** Variables calculées sur les traces sismiques.
- TABLEAU 9 :** Analyse factorielle discriminante pas à pas sur 88 traces sismiques synthétiques.
- TABLEAU 10 :** Analyse factorielle discriminante pas à pas sur 43 traces sismiques synthétiques.
- TABLEAU 11 :** Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 12 - pas n° 3.
- TABLEAU 12 :** Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 12 - pas n° 5.
- TABLEAU 13 :** Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 12 - pas n° 3 - affectation d'individus anonymes.
- TABLEAU 14 :** Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 19 - pas n° 5.

- TABLEAU 15 :** Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 19 - pas n° 5 - affectation d'individus anonymes.
- TABLEAU 16 :** Interprétation du cercle des corrélations en analyse en composantes principales.
- TABLEAU 17 :** Validation pratique des résultats obtenus en reconnaissance automatique des faciès sismiques avec la connaissance géologique a priori donnée par les puits.
- TABLEAU 18 :** Validation pratique des résultats obtenus en reconnaissance automatique des faciès sismiques avec la connaissance géologique a priori donnée par les puits - opposition "bon" réservoir à réservoir "moyen à mauvais".
- TABLEAU 19 :** Analyse factorielle discriminante pas à pas sur les traces d'apprentissage des pôles "dune oolithique" et "platier oncolithique" - Dogger carbonaté du Bassin de Paris.
- TABLEAU 20 :** Variables discriminantes sélectionnées pour les différents pôles "dune oolithique" et "platier oncolithique" - Dogger carbonaté du Bassin de Paris.
- TABLEAU 21 :** Analyse factorielle discriminante pas à pas sur les traces d'apprentissage des pôles M et T - réservoirs turbiditiques argilo-gréseux du Gabon.
- TABLEAU 22 (a, b) :**
Récapitulation des tests sur l'estimation spectrale par la M.E.M. sur des traces synthétiques.
- TABLEAU 23 :** Corrélations significatives entre facteurs principaux et variables initiales - fenêtre 892-932 ms - $\Delta t = 2$ ms ou $\Delta t = 4$ ms - influence du pas d'échantillonnage.
- TABLEAU 24 :** Corrélations significatives entre facteurs principaux et variables initiales - $\Delta t = 4$ ms - influence de la taille de la fenêtre.
- TABLEAU 25 :** Corrélations significatives entre facteurs principaux et variables initiales - $\Delta t = 4$ ms - pointé dégradé.

- TABLEAU 26 :** Variables géologiques analysées dans l'étude gisement présentée.
- TABLEAU 27 :** Analyse en composantes principales sur les traces synthétiques calculées avec les données de puits - fenêtre de 28 ms.
- TABLEAU 28 :** Analyse en composantes principales sur les traces synthétiques calculées avec les données de puits - fenêtre de 35 ms.
- TABLEAU 29:** Analyse en composantes principales sur les traces synthétiques calculées avec les données de puits ou modélisées à partir des principaux faciès géologiques.
- TABLEAU 30 :** Analyse en composantes principales sur les traces sismiques réelles.
- TABLEAU 31 :** Analyse factorielle discriminante sur les traces sismiques réelles - 5 groupes d'apprentissage définis après une classification automatique.
- TABLEAU 32 :** Comparaison de l'amplitude au toit du Cénomaniens entre les données sismiques synthétiques et les données sismiques réelles.
- TABLEAU 33 :** Affectation des traces sismiques synthétiques dans les groupes de traces réelles d'apprentissage - discrimination à 5 groupes.

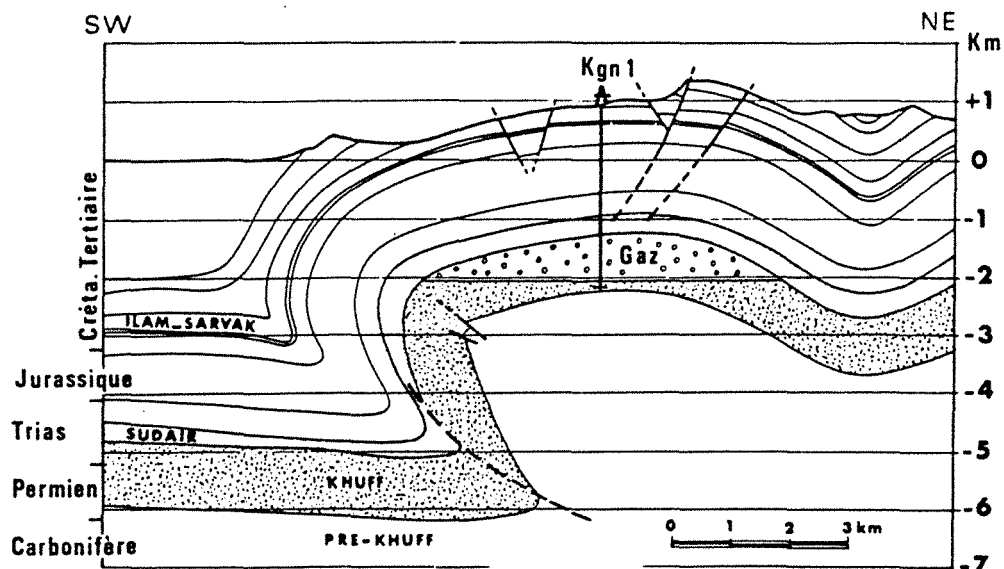
INTRODUCTION

Les données sismiques constituent une des aides fondamentales pour l'exploration pétrolière, aussi bien au stade de la découverte du gisement qu'aux stades de son appréciation (définition de son extension) ou de son développement (mise en production).

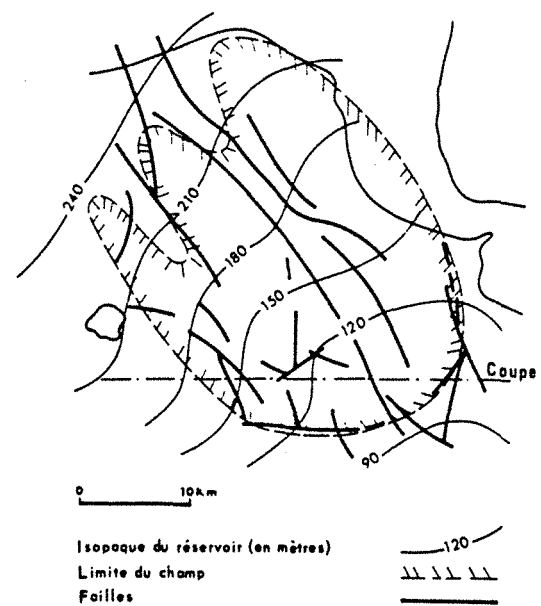
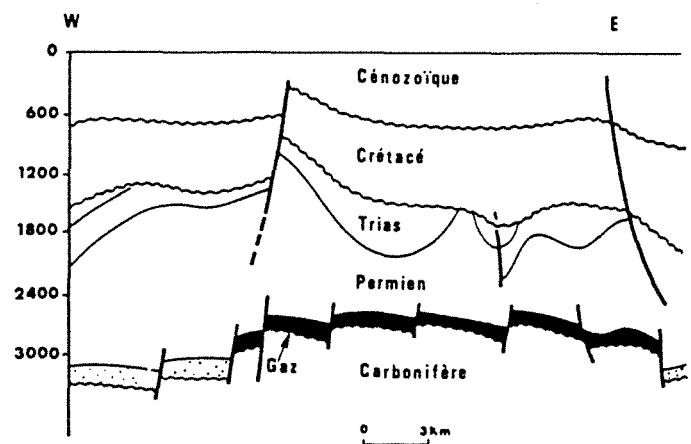
Les données sismiques fournissent en effet une image du sous-sol, certes imprécise si on compare cette information aux informations données par les puits, mais à forte densité spatiale et couvrant une large surface. Traditionnellement, l'image sismique du sous-sol est utilisée afin de définir la **géométrie** des objets géologiques, par exemple, la forme d'un anticlinal qui constituerait un excellent piège pétrolier (Figure 1a), ou encore la forme et la pente d'un horizon (ou couche géologique) proche du toit d'un réservoir potentiel. **L'interprétation structurale** de la sismique permet aussi de définir les failles majeures dont le comportement peut influencer l'existence ou les caractéristiques du réservoir (Figure 1b). Cette approche exclusivement géométrique des données sismiques a prévalu jusque vers les années 70. Le coût des puits étant extrêmement élevé par rapport aux coûts d'acquisition et de traitement d'une campagne sismique, l'interprétation structurale des données sismiques permet déjà une optimisation des nouvelles implantations de forage et offre une aide précieuse dans le calcul du volume de la structure d'où découle la décision de développer ou de ne pas développer le champ.

Dans la décennie 1970, une nouvelle façon de considérer et d'interpréter les données sismiques est apparue (Figure 2), et ceci sous la poussée simultanée de trois événements majeurs que nous allons évoquer successivement.

Tout d'abord, les données sismiques contiennent beaucoup plus que la seule information structurale. Si le temps de trajet de l'onde acoustique entre la surface du sol et les principaux "marqueurs géologiques" permet de positionner en temps, puis en profondeur, les objets géologiques, l'étude de la **forme** du signal sismique reçu, après transmission et réflexion, donne des renseignements très précieux sur la nature du sous-sol et des interfaces rencontrées par l'onde sismique. Plus précisément, la morphologie des traces sismiques est à relier aux propriétés pétrophysiques des couches (lithologies, porosités, perméabilités, contenu en fluides, etc.). Cette approche **lithologique** des données sismiques a été rendue plus facile suite à des progrès techniques dans le domaine de l'acquisition et du traitement des signaux



a — Coupe transversale du gisement de Kangan, Iran (d'après REYRE, 1975).



b — Carte et coupe du champ de Groningue, Pays-Bas (d'après STAUBLE & MILIOS, 1970).

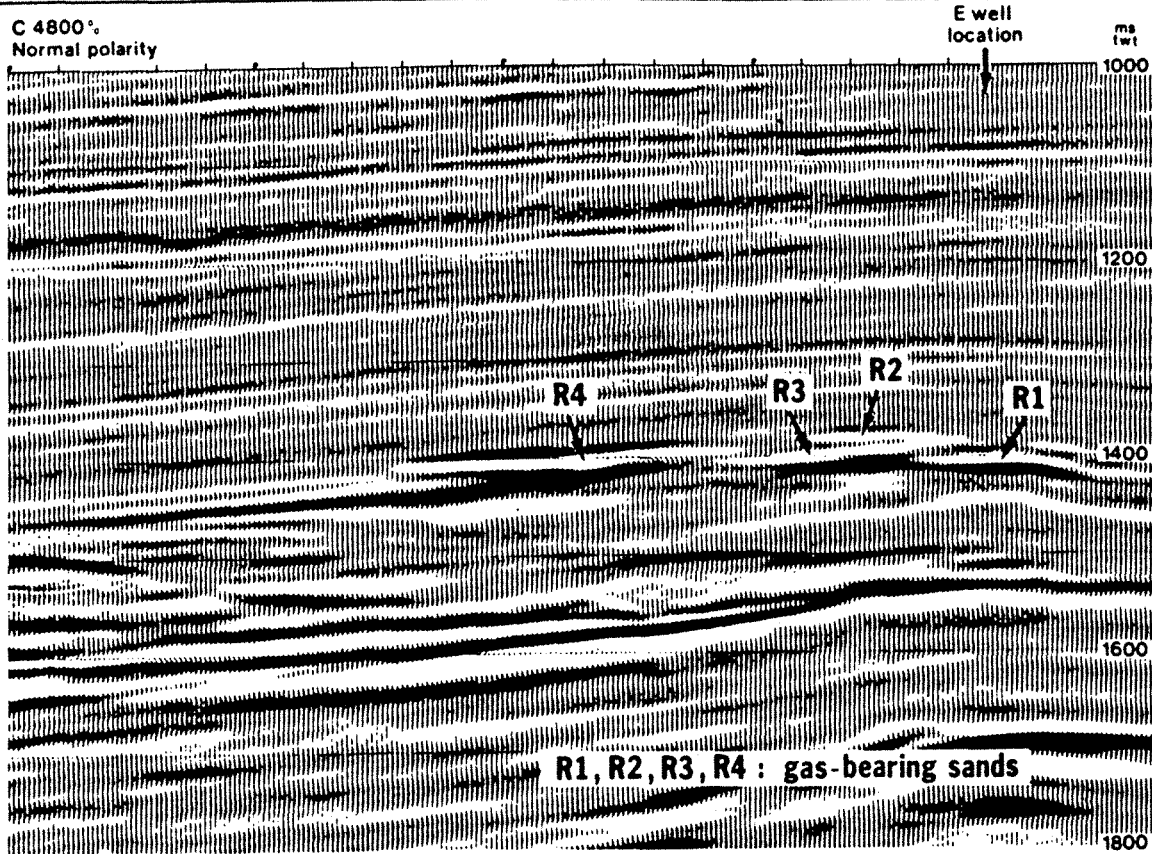
sismiques : sources plus pénétrantes, acquisition puis traitement avec une bande spectrale élargie, corrections statiques améliorées.

Par ailleurs, au tournant des années 70, la plupart des champs géants et en situation géologique simple (vastes anticlinaux en domaine de plate-forme, par exemple) avaient été découverts. Une nouvelle catégorie d'objets géologiques est apparue comme pouvant constituer des pièges intéressants : les pièges **stratigraphiques**, par opposition aux pièges **structuraux** précédents (Figure 3). Les pièges stratigraphiques sont souvent de plus petite taille et liés à des événements plus lithologiques que tectoniques (Figure 3) :

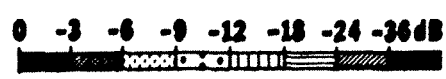
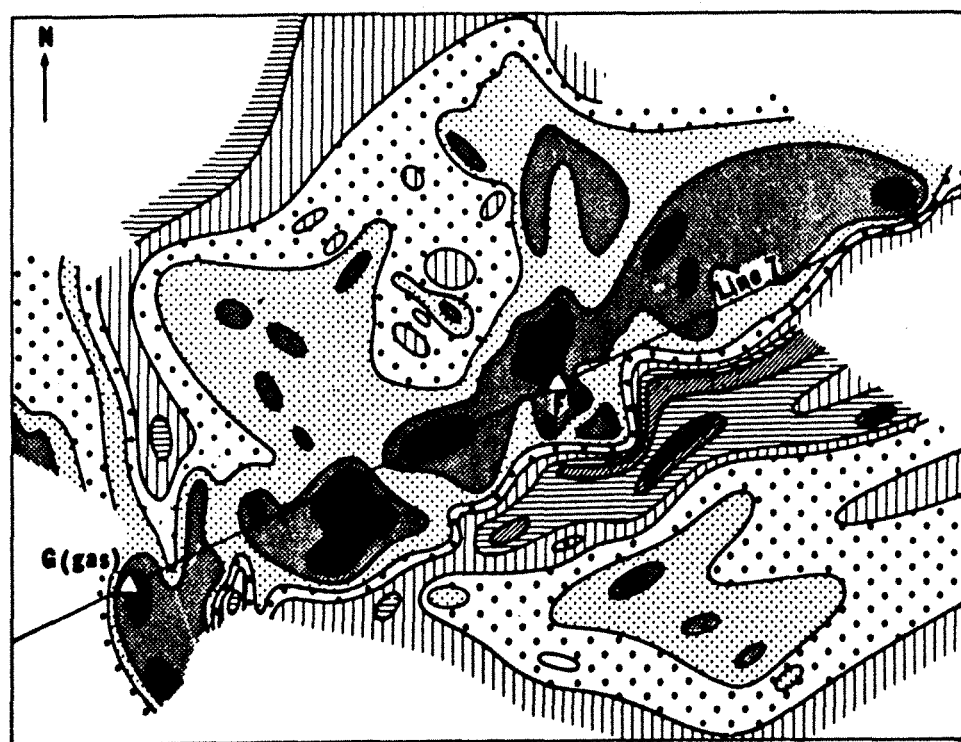
- . zone poreuse, liée à une construction récifale, drapée par un encaissant imperméable,
- . discordance étanche recouvrant des niveaux poreux,
- . variations lithologiques latérales (biseaux, chenaux gréseux dans l'encaissant argileux d'une plaine fluviale, exemple de la figure 2).

La mise en évidence des pièges stratigraphiques est beaucoup plus délicate (d'où le nom qui leur est donné parfois de "pièges subtils") et n'a pu se faire que grâce à la nette amélioration du traitement des données sismiques ainsi qu'au développement de nouveaux modèles géologiques globaux proposés par des géologues et plus en accord avec l'échelle des données sismiques (Vail, Sangree, Mitchum, 1977).

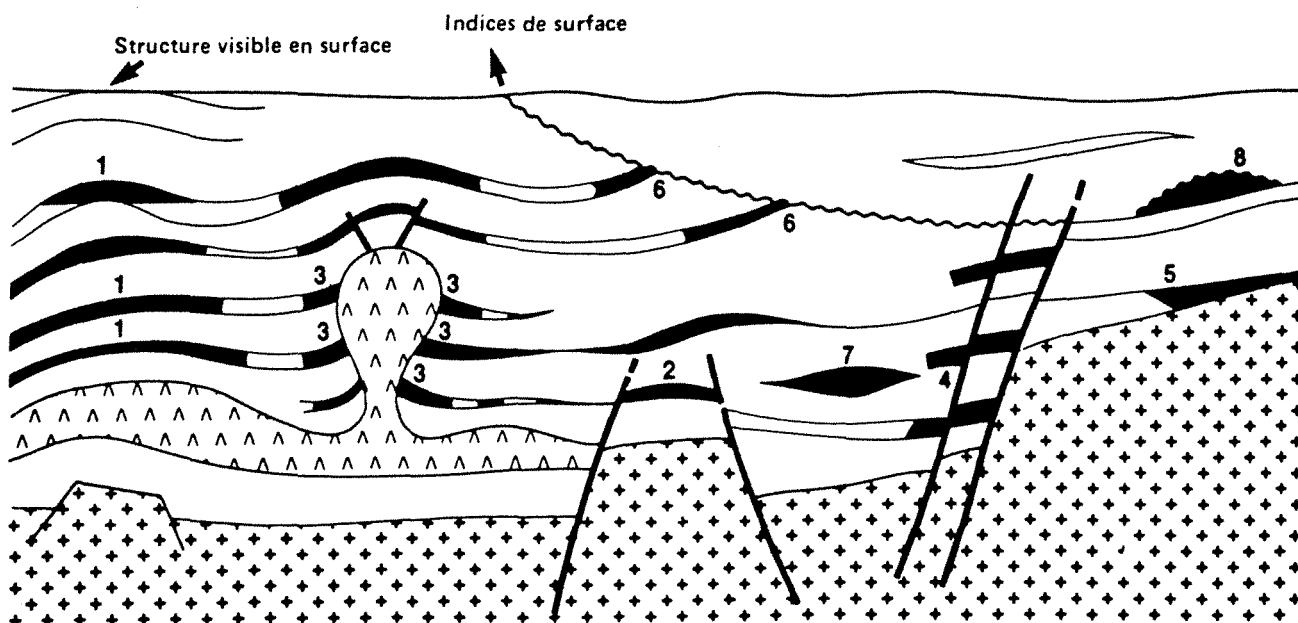
D'autre part, l'augmentation des coûts d'exploration (zones nouvelles difficiles d'accès, forages profonds, etc.) ainsi que des prix très fluctuants du pétrole expliquent un intérêt accru pour une meilleure exploitation des réserves **déjà trouvées**. Très souvent, moins d'un tiers de l'huile en place est récupéré et ceci suite à une mauvaise connaissance du gisement : variations de porosités mal connues, barrières de perméabilité non détectées... L'utilisation des données sismiques peut apporter des éléments fondamentaux à la connaissance du gisement, et donc à son exploitation. Le coût de l'acquisition et du traitement des données sismiques est très faible, le gain lié à une meilleure récupération est important. Ces deux éléments sont à comparer aux coûts énormes des forages d'exploration dans des zones nouvelles pour des résultats très incertains. Une meilleure exploitation des gisements reste la préoccupation majeure des années 90 et l'interprétation sismique semble s'orienter résolument vers des fins "réservoirs". Au dernier congrès mondial de la Society of Exploration Geophysicists (Dallas, novembre 1989), beaucoup d'exposés étaient consacrés à la géophysique de gisement et de plus en plus de logiciels sismiques sont destinés à l'interprétation **structurale et lithologique des réservoirs**.



Section PAM migrée profil 4.



Carte d'amplitude de la réflexion sur un banc de sable.



PIÈGES STRUCTURAUX

- 1 _____ Anticlinal
- 2 _____ Horst
- 3 _ Flanc de dôme de sel ou d'argile
- 4 _____ Piège contre faille

PIÈGES STRATIGRAPHIQUES

- 5 _____ Biseau sédimentaire
- 6 _____ Biseau d'érosion
- 7 _____ Lentille (de sable ou de grès)
- 8 _____ Récif



Socle cristallin ou substratum de la série sédimentaire

Coupe verticale schématique montrant les différents types de pièges à hydrocarbures

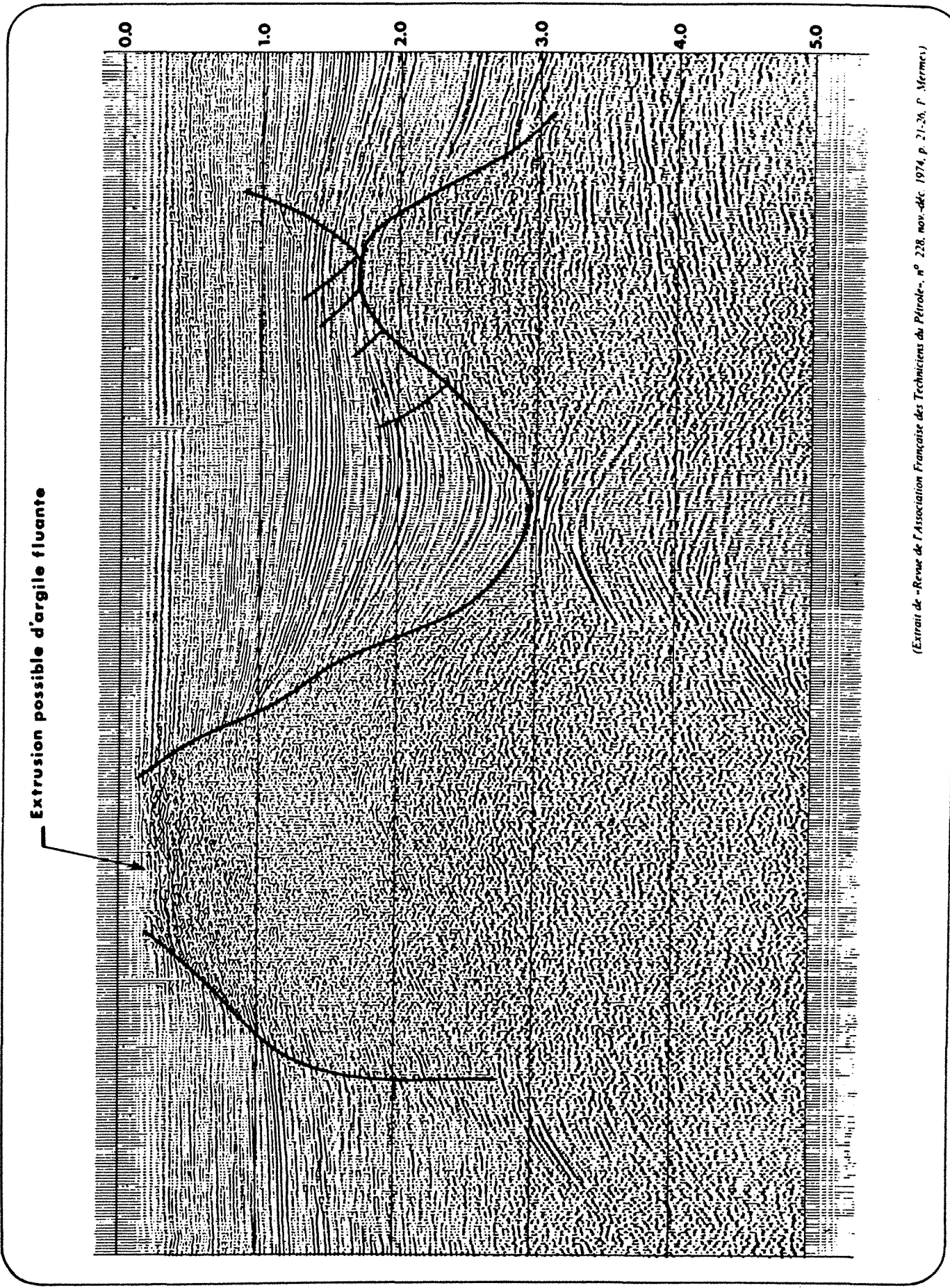
L'interprétation structurale des données sismiques consiste à relier les principaux marqueurs observés sur la section sismique (Figure 4), le plus souvent en position temps, aux couches géologiques repérées en profondeur au niveau des puits. Ce lien peut s'établir, par exemple, en utilisant le log de vitesse des ondes P, dans les différents terrains, enregistré aux puits. Les événements géologiques importants (discordance, horizon réservoir, etc.) étant repérés sur la section sismique, ils sont ensuite pointés ainsi que les failles qui peuvent affecter leur position. Ce travail permet d'aboutir à un carté -en temps ou en profondeur- de ces marqueurs géologiques (cf. carte structurale de la figure 1b, où sont reportées sur les courbes isopaches du réservoir les failles majeures).

L'interprétation stratigraphique des données sismiques étudie les variations de **forme du signal sismique** qui "habille" les différents horizons car elles peuvent traduire des évolutions des propriétés pétrophysiques des milieux de part et d'autre des interfaces. Dans une zone définie entre deux marqueurs sismiques, on étudie également les caractéristiques morphologiques du signal, ceci afin d'essayer de remonter aux propriétés lithologiques du milieu. L'interprétation stratigraphique permet de découper l'image sismique en grands ensembles géologiques et de proposer une interprétation géologique des différentes zones au sein de chaque ensemble (Figure 5).

L'interprétation lithologique (Figure 6) des données sismiques essaie d'aller encore plus loin en détectant la lithologie, le contenu en fluides, les propriétés pétrophysiques à travers l'analyse d'une fine portion de trace sismique appartenant à l'une des zones détectées au cours de l'interprétation stratigraphique. La figure 6 présente une modélisation sismique à partir des données de puits pour comprendre l'effet du gaz sur la réponse sismique du réservoir et prédire, par la suite, la présence ou non du gaz dans les sables.

Ces trois types d'interprétation se complètent et souvent s'enchaînent chronologiquement avec l'intensification de l'exploration dans une zone. Aux stades de la découverte ou de l'appréciation, ce sont les analyses structurale et stratigraphique qui vont prévaloir alors qu'au stade de développement d'un gisement, une expertise lithologique est souvent recherchée.

Ce mémoire présente un ensemble de méthodes de traitement appliquées à la **caractérisation automatique** et à l'**analyse lithologique des faciès sismiques**. Ces méthodes ont été développées dans le cadre d'un projet de recherche à la Société Nationale Elf-Aquitaine, initié



(Extrait de -Revue de l'Association Française des Techniciens du Pétrole-, n° 228, nov.-déc. 1974, p. 21-26, F. Mermec.)

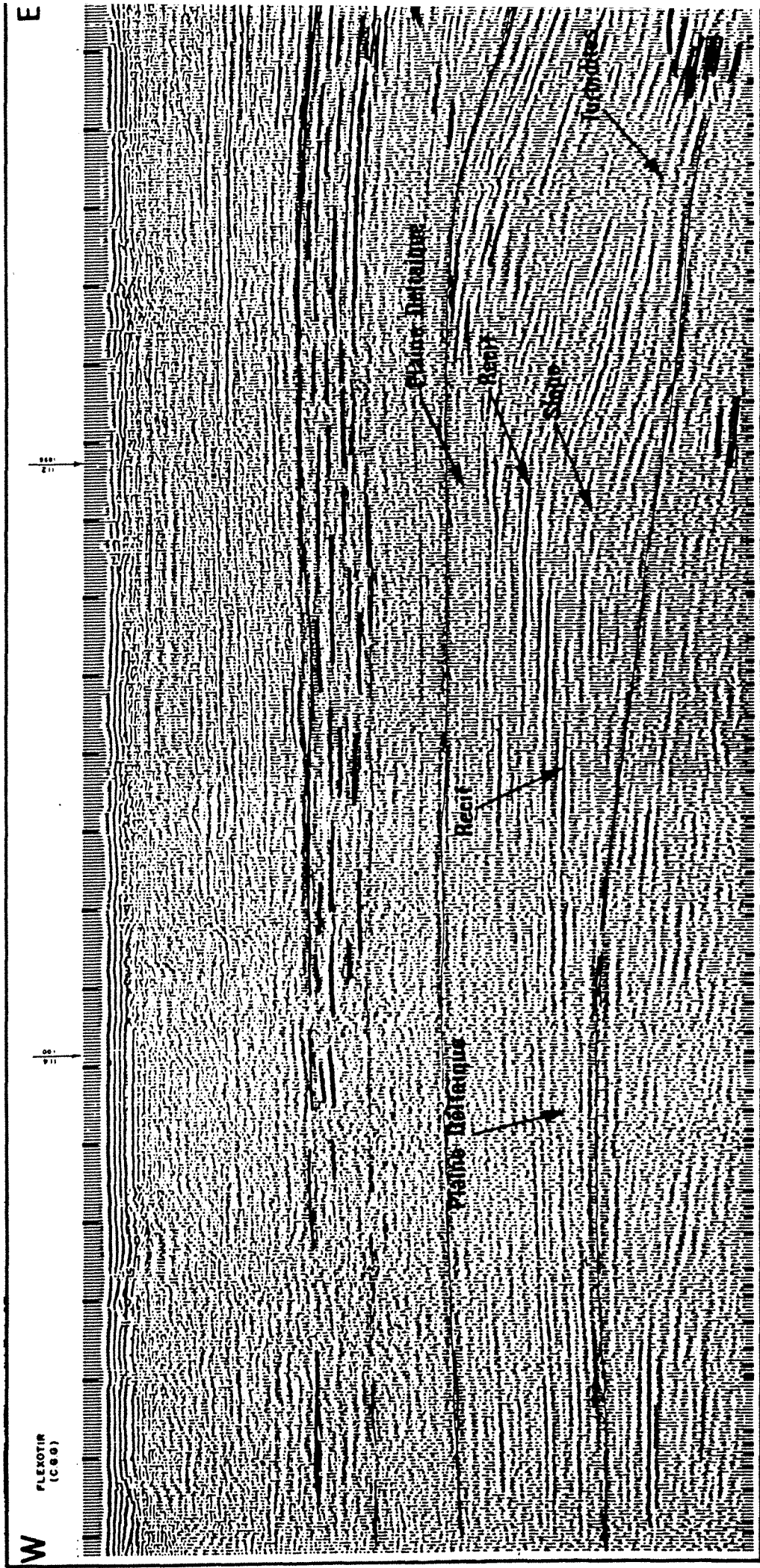
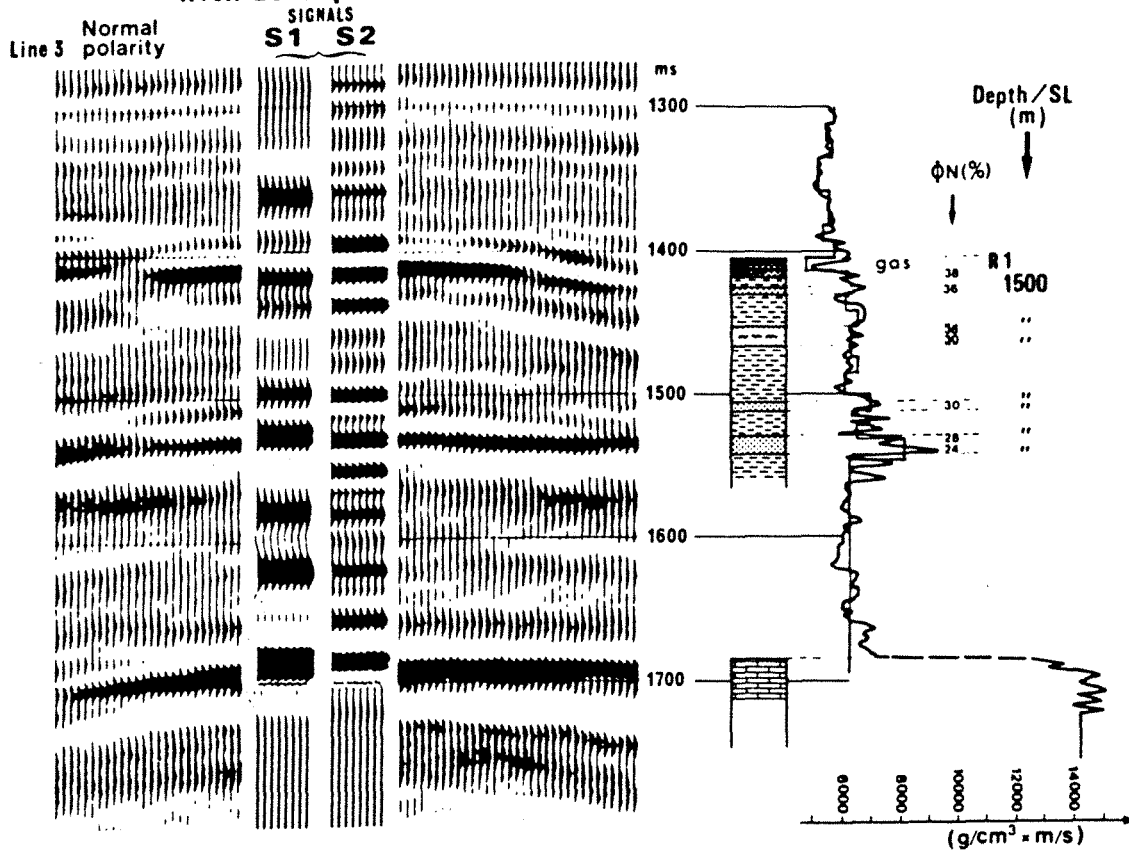


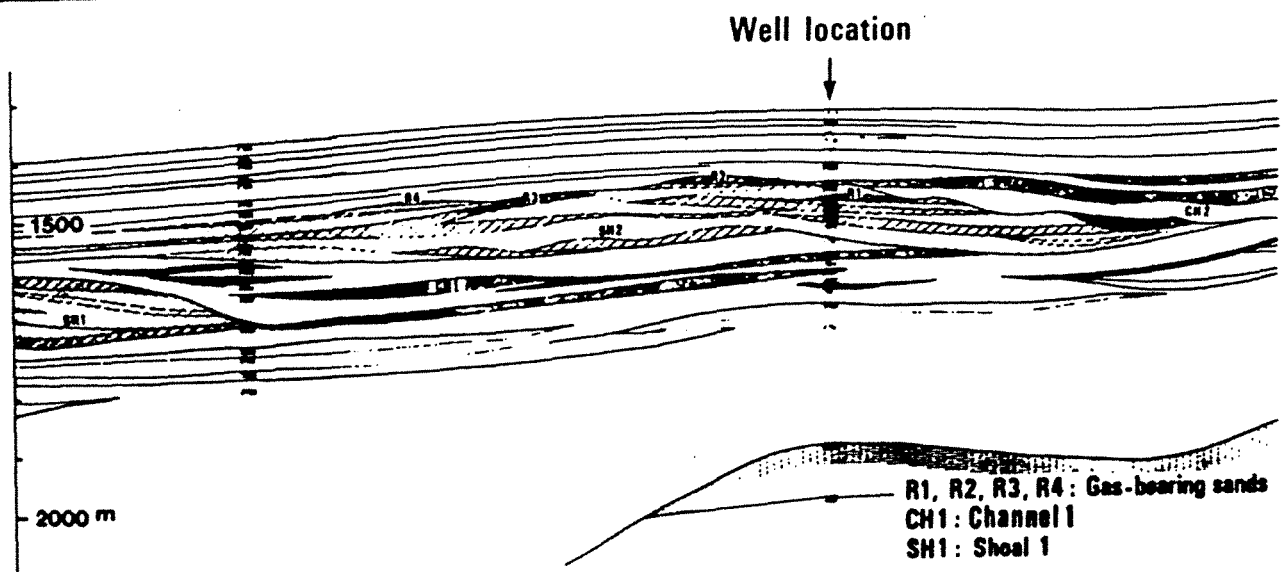
Fig. 5

Exemple d'une ligne sismique de la Mahakam pointée en séquence.

**Synthetic Seismograms
with zero phase**



*Simplification du log d'impédance du forage
et comparaison traces synthétiques-section réelle.*



Profil 3. Modèle profondeur.

en 1984 par Jean Dumay, géologue expérimenté en interprétation stratigraphique des données sismiques.

J. Dumay, s'appuyant sur sa propre expérience, a fait les remarques suivantes :

- . l'analyse des faciès sismiques est basée sur la caractérisation morphologique automatique faite par l'oeil de l'interpréteur; plusieurs paramètres sont ainsi automatiquement intégrés et reconnus systématiquement sur toute la zone d'étude pour déterminer les faciès sismiques ;
- . cette approche reste très qualitative et subjective ; d'un interpréteur à l'autre, des variations de diagnostic importantes existent ;
- . les caractéristiques des traces voisines de puits où l'information géologique est connue sont intégrées par l'interpréteur afin d'essayer de prédire en toute trace les propriétés géologiques correspondantes.

L'analyse des faciès sismiques et leur liaison avec les propriétés géologiques sous-jacentes apparaissent donc comme une **démarche essentiellement multivariable**.

La figure 7 présente une série argilo-gréseuse dans un environnement fluvio-deltaïque. Trois zones principales de dépôt ont été distinguées :

- une zone de plate-forme interne ;
- une zone de barrières littorales ;
- une zone plus amont de plaine deltaïque.

L'expression sismique de ces trois zones est donnée en regard, sur la figure 7. Si l'on utilise la continuité comme paramètre caractérisant le faciès sismique, il est possible de distinguer le faciès plaine deltaïque (zone sourde) des deux autres faciès (nombreux réflecteurs continus). Ce paramètre ne permet pas toutefois d'individualiser ces deux derniers faciès. Si l'on introduit le paramètre supplémentaire qu'est la fréquence apparente des diverses réflexions, alors la zone de barrières littorales se caractérise par rapport à la zone de plate-forme : le complexe de barrières littorales génère des réflexions plutôt "haute fréquence" alors que les réflexions de la zone de plate-forme sont plutôt "basse fréquence" (Figure 7).

SEISMIC STRATIGRAPHIC INTERPRETATION REGRESSIVE SEQUENCE IN DELTAIC ENVIRONMENT

PARAMETERS TO

CONSIDER :

- ENERGY
- FREQUENCIES

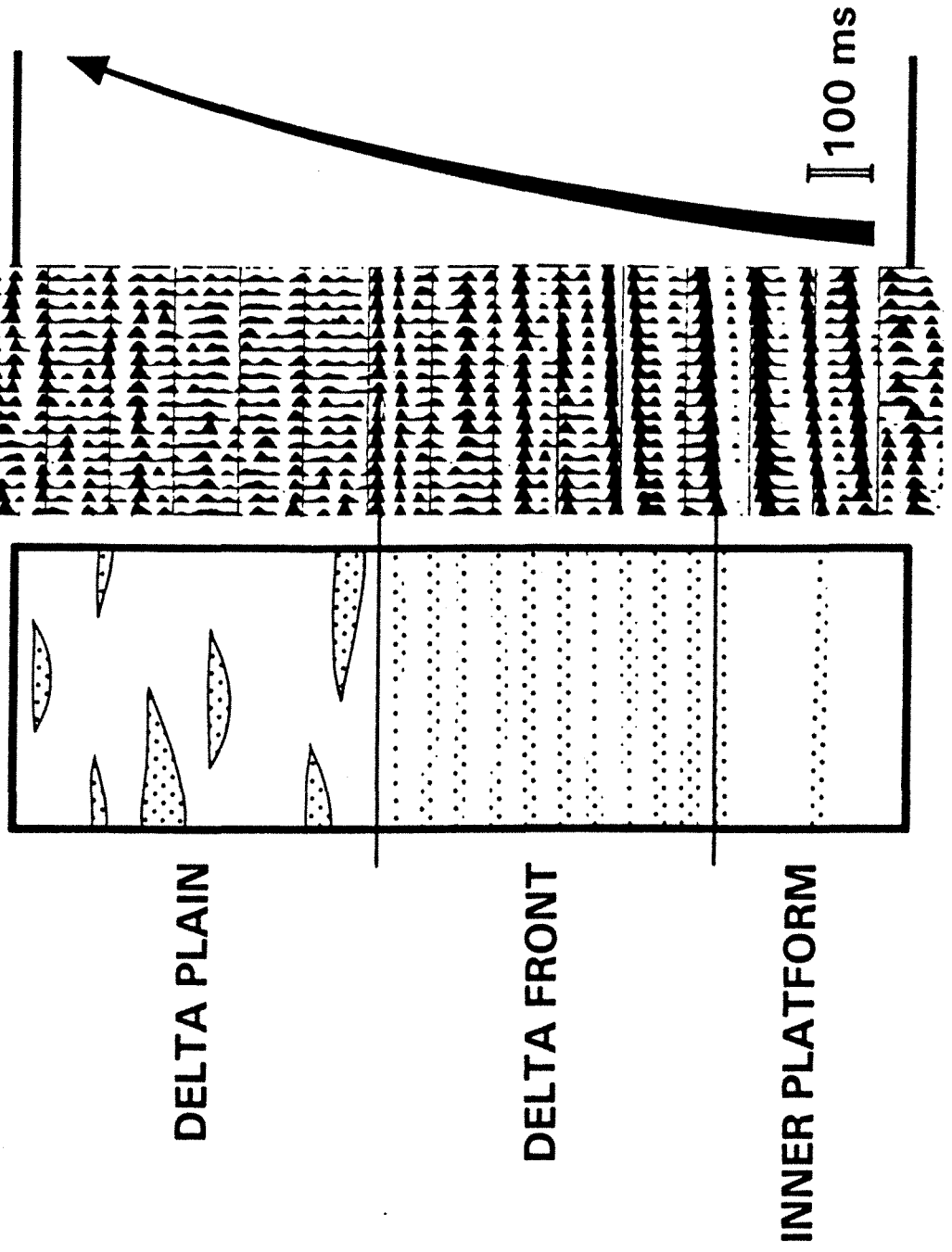


Fig. 7

Deux paramètres permettent donc de caractériser ces trois faciès sismiques (Figure 7).

Les objectifs que nous nous sommes fixés, dans le cadre de ce projet de recherche, sont les suivants:

- automatiser une démarche manuelle lourde et très subjective jusqu'à présent ;
- améliorer cette démarche en essayant de quantifier l'information sismique utile à l'analyse des faciès et de choisir les paramètres les plus discriminants par rapport aux problèmes posés ;
- aller au-delà de la simple analyse qualitative des faciès en en proposant une interprétation géologique (lithologique) en intégrant, de façon un peu plus quantitative, l'information géologique aux puits aux données sismiques.

Deux questions sont à résoudre au préalable :

- comment caractériser les traces sismiques étudiées ?
- quels sont les outils permettant de prendre en compte l'information sismique et d'automatiser son analyse ?

En 1984, très peu de travaux existaient dans le domaine de la reconnaissance automatique des faciès sismiques. Les paramètres proposés par les auteurs étaient soit les attributs classiques calculés sur les traces sismiques (vitesses d'intervalle, fréquences instantanées, Conticini, 1984) soit des attributs calculés à partir de la fonction d'autocorrélation des traces sismiques (Khattri et al., 1976, 1979, 1983, 1984). Mais le pouvoir discriminant de ces derniers paramètres n'était démontré que sur des données sismiques synthétiques (Velzeboer, 1981, Khattri et al., 1976, 1979, 1983, 1984).

Pour ce qui est de la prise en compte de l'information sismique, c'étaient des statistiques classiques qui étaient utilisées la plupart du temps (histogrammes, "cross-plots", diagrammes en étoile), quelques références existaient quant à l'utilisation de l'analyse factorielle discriminante. Or, il est clair que les techniques d'analyses statistiques multivariées sont des outils bien adaptés pour traiter la masse énorme d'information sous-jacente aux traces sismiques et sélectionner les éléments pertinents pour l'interprétation des faciès.

Nous avons donc développé le calcul de divers attributs à partir des traces sismiques et testé leur signification à la fois sur des **données synthétiques** et sur des **données sismiques réelles**. Parallèlement, nous avons exploré l'utilisation de diverses méthodes d'analyses multivariées. Ces travaux ont abouti en 1986 à une première méthodologie pour la caractérisation des faciès sismiques. Il est amusant de constater, qu'à la même époque, une équipe de chercheurs de Control Data Corporation aboutissait à une méthode très voisine et qu'en 1987, au congrès annuel de la Society of Exploration Geophysicists, un "workshop" entier était consacré à la reconnaissance des formes, basée sur des méthodes statistiques, en sismique pétrolière.

La première méthodologie proposée en 1986 a été enrichie ultérieurement pour des applications diverses (présence ou pas de puits pour une calibration géologique, fenêtre temporelle d'analyse des traces très courte) et testée dans un cadre industriel.

L'intérêt industriel de cette méthode ayant été démontré, elle est actuellement en cours d'installation sur des stations d'interprétation géophysique. Nous travaillons aussi à son habillage par des techniques d'intelligence artificielle.

De nouveaux travaux sont en cours, mais dans un cadre plus orienté pour le gisement (problèmes très fins de réservoir). Nous étudions des procédés d'"inversion statistique" des traces sismiques permettant d'extraire des données sismiques, avant ou après sommation, des informations géologiques **quantitatives** sur les propriétés moyennes du réservoir correspondant à la très fine portion de trace sismique analysée.

Dans ce mémoire, nous ne reprenons que les résultats obtenus de 1984 à 1987 sur l'**analyse automatique des faciès sismiques**.

Nous présenterons dans la première partie les données sismiques étudiées (acquisition et traitement standard) ainsi que l'interprétation qui peut en être faite dans le cadre de la sismique stratigraphique et lithologique.

Puis nous développerons (deuxième partie) la caractérisation des traces sismiques au moyen d'attributs ainsi que le traitement statistique des traces analysées. Nous discuterons les limites de la méthode puis proposerons deux études de cas.

La troisième partie est consacrée aux développements faits pour appliquer la méthodologie statistique à des portions de traces correspondant à des fenêtres temporelles beaucoup plus petites. Nous exposerons la nécessité de ces développements, puis proposerons une comparaison entre diverses méthodes d'analyse spectrale. Nous étudierons la possibilité de caractériser les traces sismiques sur une courte fenêtre au moyen de la représentation temps-fréquence de Wigner-Ville. Puis, nous proposerons une caractérisation des traces sismiques basée sur la succession des amplitudes de la trace et montrerons, dans ce cadre, une application à des données réelles.

En conclusion, nous synthétiserons les points clés de la méthode statistique de reconnaissance automatique des faciès sismiques. Puis nous évoquerons les travaux de recherche en cours (inversion statistique) pour une interprétation directe des traces sismiques en terme de valeurs moyennes d'attributs géologiques.

PREMIERE PARTIE

**LES DONNEES SISMIQUES : ACQUISITION, TRAITEMENT ET
INTERPRETATION**

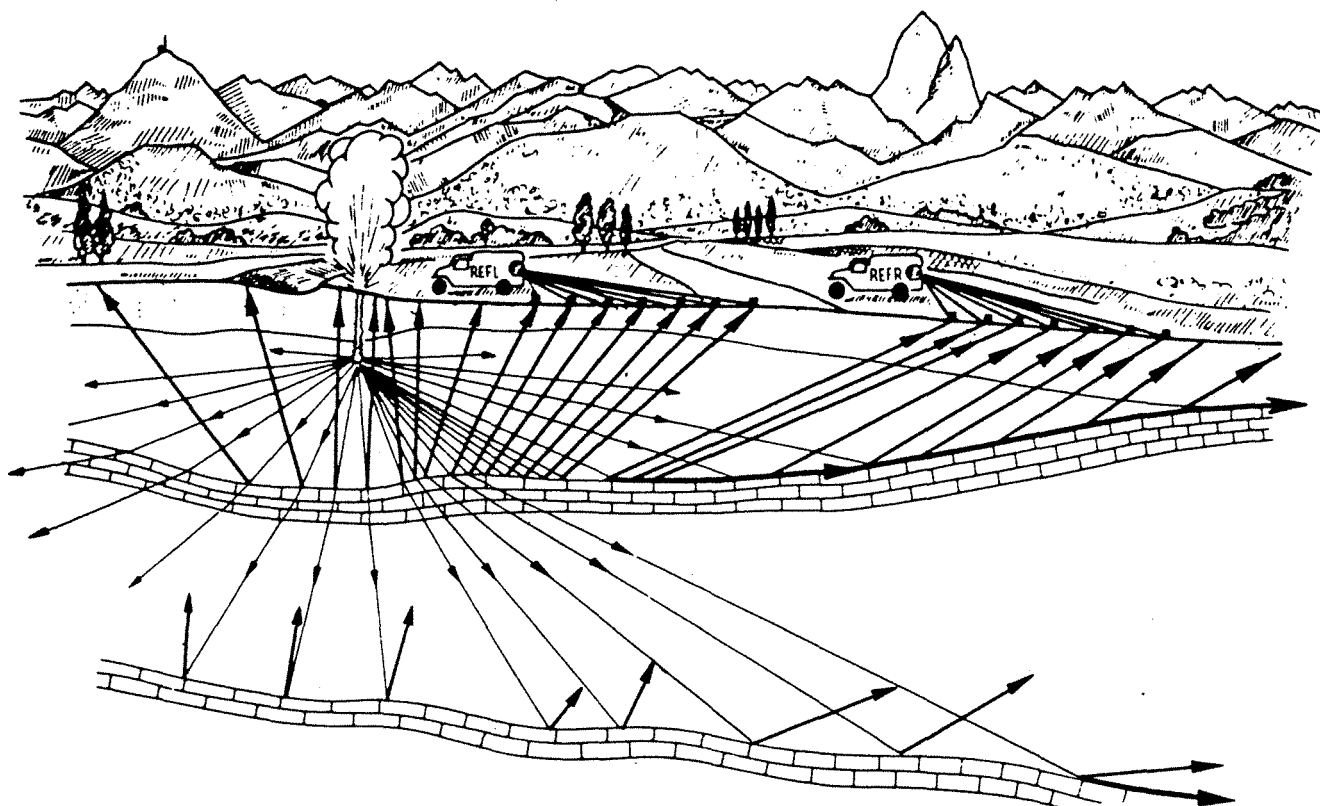
1. INTRODUCTION

La quasi totalité de la géophysique pétrolière est constituée par la méthode de **sismique réflexion** (plus de 90 % des dépenses de la géophysique d'exploration en 1986), bien que la prospection sismique ait été initialisée par des **méthodes de réfraction** (premiers brevets, en 1919, de l'allemand Mintrop). Ces méthodes de sismique réfraction avaient permis, dans les années 1920, de découvrir plusieurs dômes de sel aux Etats-Unis, dont certains liés à des gisements d'hydrocarbures. Jusqu'au début des années 30, la sismique réfraction, ainsi que les méthodes gravimétriques, magnétiques et électriques, sont à l'origine de nombreux succès (1923, dôme d'Aricești en Roumanie découvert avec des méthodes électriques, dômes de Nash et de Orchard au Texas par sismique réfraction).

La sismique réflexion fait une entrée en force à partir des années 1930 et supprime très rapidement les autres méthodes géophysiques. En France, son premier succès a été la découverte du champ de gaz géant de Lacq en 1950, suivie par celle du champ de Parentis en 1954.

Toutefois, la gravimétrie, le magnétisme, les méthodes électriques diverses, la magnéto-tellurique continuent à être utilisés actuellement comme méthodes d'appoint. En raison de leur coût beaucoup moins élevé, elles interviennent dans le processus de l'exploration, soit comme précurseurs pour préparer une judicieuse et économique utilisation de la sismique réflexion, soit à titre d'accompagnement de cette dernière pour en permettre une meilleure interprétation.

PRINCIPE DES MÉTHODES SISMIQUES



En un point d'émission, on ébranle le sol, au moyen d'explosifs par exemple : l'onde se propage alors dans toute les directions *se réfléchit et/ou se réfracte* sur les différentes couches et revient en surface au bout d'un temps plus ou moins long, fonction de la **VITESSE DE PROPAGATION DES ONDES** dans les terrains et de la longueur du trajet suivi.

Posés en surface où ils constituent le **DISPOSITIF DE RÉCEPTION**, des **GÉOPHONES** reçoivent les mouvements du sol et les transmettent aux enregistreurs (**RÉFLEXION** ou **RÉFRACTION**).

2. LES PRINCIPES DE LA SISMIQUE REFLEXION

La méthode consiste à provoquer un ébranlement du sol ou une onde de pression dans l'eau qui se transmet au sous-sol. La formation d'un ébranlement à proximité d'un tir de dynamite, que l'on peut considérer comme ponctuel, est souvent mal connue car les matériaux y sont soumis à des déformations non linéaires très complexes. Cependant, on peut admettre, qu'à une distance de l'ordre de quelques mètres du tir, les ondes émises ("signature" de la source) sont des **ondes élastiques**. Les principales discontinuités géologiques ou "marqueurs" réfléchissent ces ondes. Les signaux **réfléchis** sont captés à la surface du sol, traités puis analysés.

Les capteurs utilisés sont soit des vélocimètres, géophones posés à terre : ils produisent une tension électrique de sortie proportionnelle à la vitesse du mouvement du sol produit par un ébranlement, soit des hydrophones immergés à quelques mètres de profondeur en mer et qui sont sensibles aux variations de pression engendrées par un ébranlement grâce à leurs propriétés piezo-électriques.

La figure 8 illustre le principe des méthodes sismiques et donne une idée des domaines respectifs de la sismique réflexion et réfraction.

Les ondes émises par la source sont, d'une part, des ondes de **compression** (ondes P) ou ondes longitudinales pour lesquelles la vibration engendrée s'exerce parallèlement à la direction de propagation, d'autre part, des ondes de **cisaillement** (ondes S) ou transversales pour lesquelles la vibration s'exerce perpendiculairement à la direction de propagation. La vitesse des ondes de cisaillement est inférieure, dans un rapport voisin de 2, à celle des ondes de compression. Elles ne se propagent qu'en milieu anisotrope : en particulier, l'eau ne permet pas leur transmission. Chacun de ces types d'ondes peut subir des conversions dans l'autre type à la suite d'une réflexion sur un dioptré. Nous nous focaliserons sur le traitement et l'interprétation des données sismiques correspondant aux ondes de compression, qui jusqu'à présent constituent plus de 98 % des données sismiques enregistrées.

Cependant, les données sismiques des ondes S permettent d'accéder à des paramètres supplémentaires du sous-sol, non fournis par les ondes P. On peut penser que les méthodes d'acquisition et de traitement des ondes S vont prendre de l'importance industrielle dans un proche avenir.

Les ondes P et S ne sont pas les seules ondes émises à l'acquisition. D'autres ondes, considérées comme du bruit, et que l'on cherchera donc à éliminer, viennent perturber les enregistrements. Les ondes de Rayleigh ou de Love sont très superficielles ("ground roll") ;

elles se propagent horizontalement près de la surface et correspondent en quelque sorte à des "vagues" produites dans la partie souple des terrains. Des ondes dérivées des précédentes, résultant de diffractions sur des irrégularités des milieux superficiels, viennent s'y ajouter.

Les activités industrielles peuvent également générer du bruit au moment de l'acquisition. Signalons aussi que les réflexions multiples, c'est-à-dire des échos seconds, triples, ou davantage, des réflexions entre la surface du sol, ou d'autres dioptries en profondeur, sont considérées comme faisant partie du bruit.

La méthode sismique -émission d'ondes élastiques puis réception, traitement et analyse des ondes transmises et réfléchies dans le sous-sol- peut se définir comme l'enregistrement du filtrage de la signature par le sous-sol et l'interprétation de ce filtrage.

Le filtrage sismique du sous-sol est observé sur l'enregistrement de la signature et sur celui des ondes, en surface, qui se sont ensuite formées par réflexions, réfractions et diffractions dans le sous-sol. Selon une terminologie usuelle, le sous-sol est le **filtre**, la signature que l'on émet est l'**entrée** ou le signal de ce filtre et la surface sismique échantillonnée sous forme de traces en est la **sortie** ou la **réponse**. Cette réponse (filtrage du sous-sol) est la solution du **problème sismique direct**.

Ce problème peut être approché en se plaçant dans le cas d'ondes élastiques quelconques : le sous-sol est alors considéré comme un ensemble de milieux linéaires.

La simulation du problème direct, dans ces conditions, consiste d'abord à calculer la réponse impulsionnelle du sous-sol $w(X, Y, Z, x_0, y_0, z_0, t)$ en tout point $M(X, Y, Z)$ à l'impulsion de Dirac $\delta(x_0, y_0, z_0, t)$ émise au point de tir $S(x_0, y_0, z_0)$, connaissant la distribution $\Phi(X, Y, Z, \lambda, \mu, \rho, \alpha)$ où λ et μ sont les coefficients de Lamé, ρ la masse spécifique et α le coefficient d'absorption au point M .

En milieu homogène, isotrope et élastique ($\alpha = 0$), la distribution Φ et la réponse impulsionnelle sont liées par l'équation d'onde :

$$(\lambda + \mu) \overrightarrow{\text{grad}}(\text{div. } \vec{u}) + \mu \Delta \vec{u} = \rho \frac{\partial^2 \vec{u}}{\partial t^2}$$

où \vec{u} est le déplacement du point M au passage de l'onde.

En milieu inélastique, sous certaines conditions, le terme d'absorption α peut être introduit dans la propagation, par l'utilisation d'une vitesse de propagation complexe.

La simulation se termine par le calcul de la réponse $v(X, Y, Z, x_0, y_0, z_0, t)$ à la signature $s(x_0, y_0, z_0, t)$ émise en S, c'est-à-dire par la convolution de s par w .

$$v(X, Y, Z, x_0, y_0, z_0, t) = s(x_0, y_0, z_0, t) * w(X, Y, Z, x_0, y_0, z_0, t)$$

(où * représente l'opération de convolution sur la variable temps).

Ainsi, les traces sismiques sont un échantillonnage de la réponse ou "surface sismique" $v(x, y, z, x_0, y_0, z_0, t)$, x, y, z représentant les valeurs de X, Y, Z en surface.

L'interprétation de la surface sismique $v(x, y, z, x_0, y_0, z_0, t)$ ou solution du **problème inverse** donnerait, dans une première phase, par déconvolution des traces par la signature, la réponse impulsionnelle $w(x, y, z, x_0, y_0, z_0, t)$ puis, dans une seconde phase, remonterait à la représentation spatiale des couches avec les caractéristiques physiques, soit à la distribution Φ .

Le problème direct est résolu mais non le problème inverse. Les solutions données au problème direct sont appelées les "modèles". Le modèle le plus simple et le plus utilisé industriellement (en particulier pour le calage des données sismiques avec les événements géologiques, calage préalable à l'interprétation) est la solution de Wuenschel (1960) ou **film synthétique** (Figure 9). Le film (ou trace) synthétique est la solution au problème direct, à une dimension spatiale, dans l'hypothèse de couches élastiques horizontales et d'ondes planes à l'incidence normale.

Les discontinuités élastiques, autrement dit les dioptrés, qui provoquent les réflexions, sont dues à des différences de vitesse de propagation du son, et de densité, liées aux caractéristiques pétrographiques des terrains. En fait, les "marqueurs" géologiques traduisent un **contraste d'impédance acoustique** entre deux couches successives, l'impédance acoustique Z_i de la couche i étant définie par le produit de la vitesse du son par la densité de la couche i :

$$Z_i = \rho_i v_i$$

v_i : vitesse du milieu i ,

ρ_i : densité du milieu i .

Le coefficient de réflexion R_i , sur l'intervalle de temps t et $t+\Delta t$ pour le dioptre séparant un milieu i d'un milieu $(i + 1)$, peut être défini, pour une onde plane à incidence normale, par la relation de Gutenberg :

$$R_i \Delta t = \frac{\rho_{i+1} v_{i+1} - \rho_i v_i}{\rho_{i+1} v_{i+1} + \rho_i v_i} = \frac{Z_{i+1} - Z_i}{Z_{i+1} + Z_i}$$

En transformant cette relation et en négligeant les termes du second ordre en $\Delta v_i \Delta \rho_i$ (avec $\Delta v_i = v_{i+1} - v_i$ et $\Delta \rho_i = \rho_{i+1} - \rho_i$), il vient :

$$R_i \Delta t = \frac{\Delta(\rho_i v_i)}{2 \rho_i v_i + \Delta(\rho_i v_i)} \approx \frac{\Delta(\rho_i v_i)}{2 \rho_i v_i} \approx \frac{\Delta Z_i}{2 Z_i} = \frac{1}{2} \Delta \text{Log } Z_i$$

$$R_i = \frac{1}{2} \frac{d}{dt} \text{Log } (\rho_i v_i) = \frac{1}{2} \frac{d}{dt} \text{Log } Z_i$$

Le coefficient de réflexion R_i apparaît comme la dérivée par rapport au temps du logarithme de l'impédance acoustique.

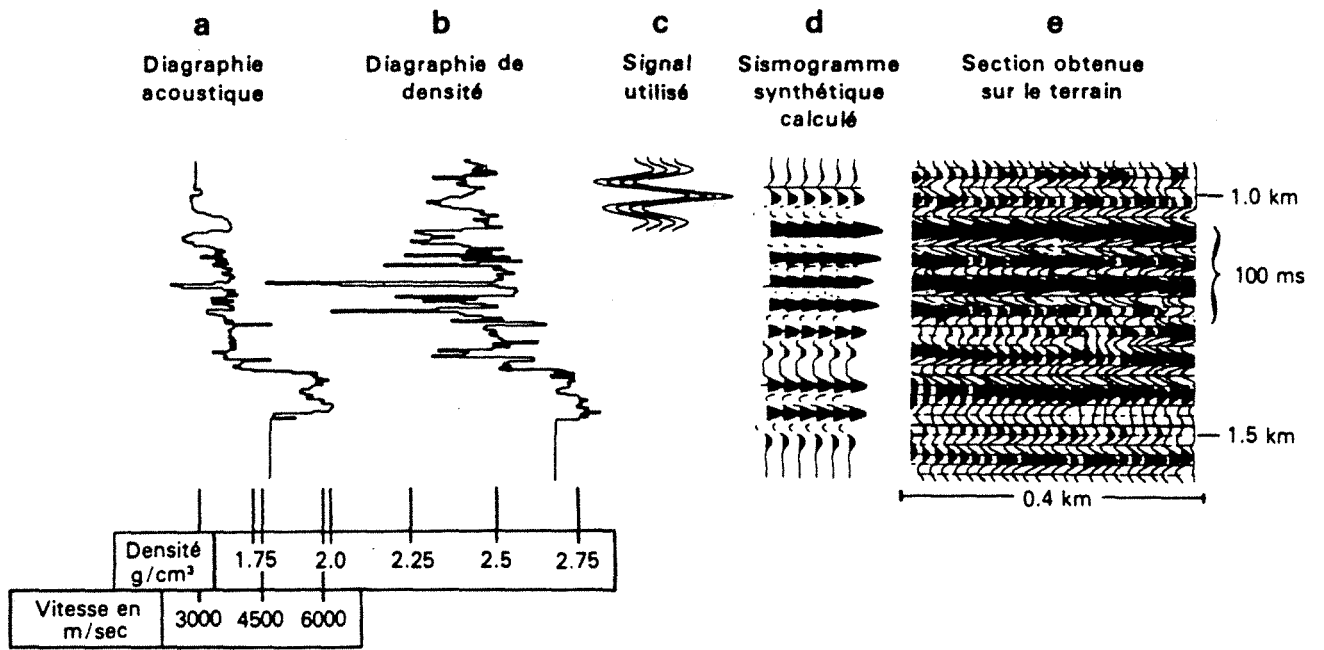
Pour connaître la suite des coefficients de réflexion, nous avons besoin des vitesses et densités des différents milieux.

Un modèle simple de la réponse sismique du sous-sol peut être calculé en utilisant les données de vitesses et densités fournies par les enregistrements diagraphiques aux puits (Figure 9).

Le log Sonic mesure en continu le temps de transit d'une onde acoustique dans la formation adjacente entre un émetteur et un récepteur à distance fixe. La diagraphie Sonic fournit ainsi un échantillonnage de la vitesse des ondes P au voisinage de la paroi du trou. La densité est par ailleurs également mesurée en continu à l'aide d'émetteurs et récepteurs de rayons gamma (diagraphie de densité).

La trace sismique synthétique (Figure 9) est calculée à partir des données de puits de la façon suivante.

La vitesse Sonic disponible est une vitesse instantanée dz/dt ; la variable "vitesse" au puits peut se mettre sous la forme d'une matrice à 3 colonnes :



- Film synthétique

- en colonne 1 : temps t,
- en colonne 2 : profondeur z,
- en colonne 3 : vitesse d'intervalle dz/dt.

A ce stade, l'échantillonnage n'est pas réalisé à pas constant.

On réalise, en premier lieu, un échantillonnage de la vitesse. En fait, on échantillonne la fonction temps-profondeur (soit les deux premières colonnes de la vitesse) à pas de temps constant. Les profondeurs associées aux temps T des échantillons constituent la variable PR.

On effectue ensuite un échantillonnage de la pression géostatique ; la pression géostatique est calculée à partir de la densité des divers milieux traversés.

Soit ρ la densité (moyenne par intervalle).

La pression géostatique à la profondeur z vaut : $p = \int_0^z \rho \, dz$

La variable Di correspond à l'information densité à 3 colonnes :

- en colonne 1 : profondeur z,
- en colonne 2 : pression géostatique,
- en colonne 3 : densité.

La fonction profondeur-pression géostatique est alors échantillonnée aux points d'abscisses PR précédemment calculées.

Les pressions géostatiques obtenues à la suite de cet échantillonnage constituent la variable P.

La pression géostatique est utilisée pour calculer l'impédance acoustique (le calcul direct est impossible car vitesse et densité ne sont pas échantillonnées aux mêmes points).

$$Z = \rho v = \frac{\rho dz}{dt} = \frac{dp}{dt}$$

On va donc former les accroissements de la variable P :

$$Z[N] = \frac{P[N] - P[N-1]}{\Delta T}$$

Le coefficient de réflexion $R[N]$, relatif à l'échantillon N , a pour expression :

$$R[N] = \frac{Z[N+1] - Z[N]}{Z[N+1] + Z[N]}$$

$$R[N] = \frac{P[N+1] - 2P[N] + P[N-1]}{P[N+1] - P[N-1]}$$

Si l'on ne dispose pas de densité, on la suppose constante, ce qui revient à remplacer, dans la variable R , la variable P par PR directement.

Le **film synthétique impulsionnel** est donné directement par la suite des coefficients de réflexion.

Il faut ensuite faire une hypothèse sur la forme et le contenu fréquentiel du signal à l'émission.

La **trace synthétique filtrée** est alors obtenue en convoluant les coefficients de réflexion R par ce signal.

Ce modèle simple de réponse sismique synthétique -modèle convolutif 1D- permet, à partir d'un forage, de caler et d'interpréter géologiquement les marqueurs observés sur une section sismique (Figure 9). Il nous permettra aussi d'étudier les caractéristiques de la réponse sismique de zones géologiques différentes grâce à des données de puits les recoupant.

3. MISE EN OEUVRE DE LA SISMIQUE REFLEXION

3.1. Principe de mise en oeuvre

Nous commencerons par décrire la mise en oeuvre la plus simple, dans une forme initiale qui n'est actuellement plus pratiquée.

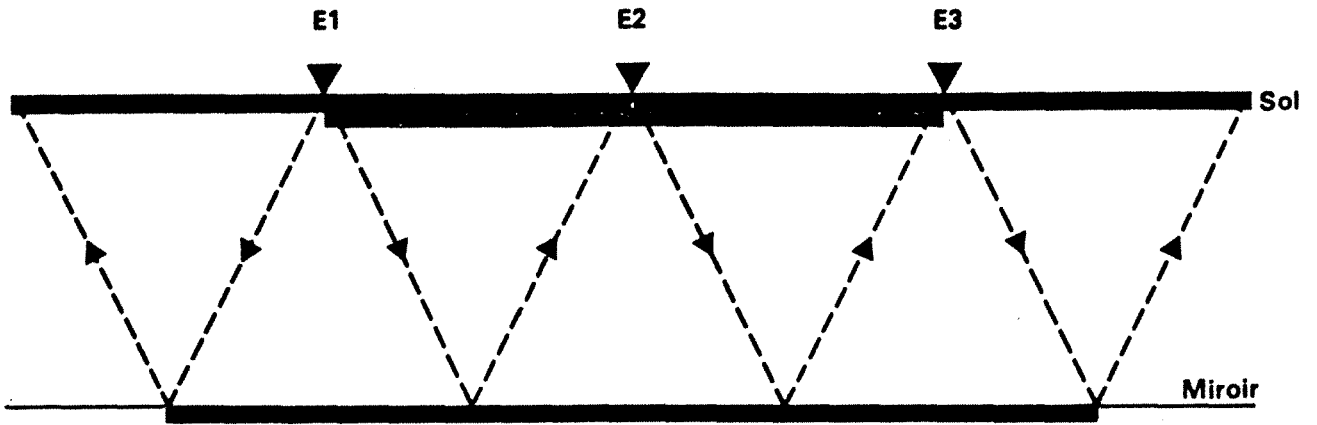
Des géophones sont disposés à intervalle régulier de part et d'autre de la source sismique (Figure 10a) pour couvrir une longueur de l'ordre du kilomètre. Après chaque ébranlement, on fait progresser ce dispositif de la moitié de sa longueur (Figure 10a) en même temps que l'on fait progresser d'une longueur égale la source. On réalise ainsi une "couverture" continue des réflexions qui se manifestent (Figure 10 a, couverture simple).

On obtient une section, ou coupe temps, juxtaposition des enregistrements obtenus à la suite des tirs. La durée d'écoute est en général de 5 à 6 secondes, ce qui permet d'obtenir en principe des réflexions provenant de couches situées parfois jusqu'à près de 10.000 m de profondeur. Cette coupe temps peut être convertie en profondeur en utilisant des informations sur les vitesses de propagation dans les terrains traversés. Toutefois, en première approximation, l'image temps est assez proche d'une coupe géologique verticale passant par le profil.

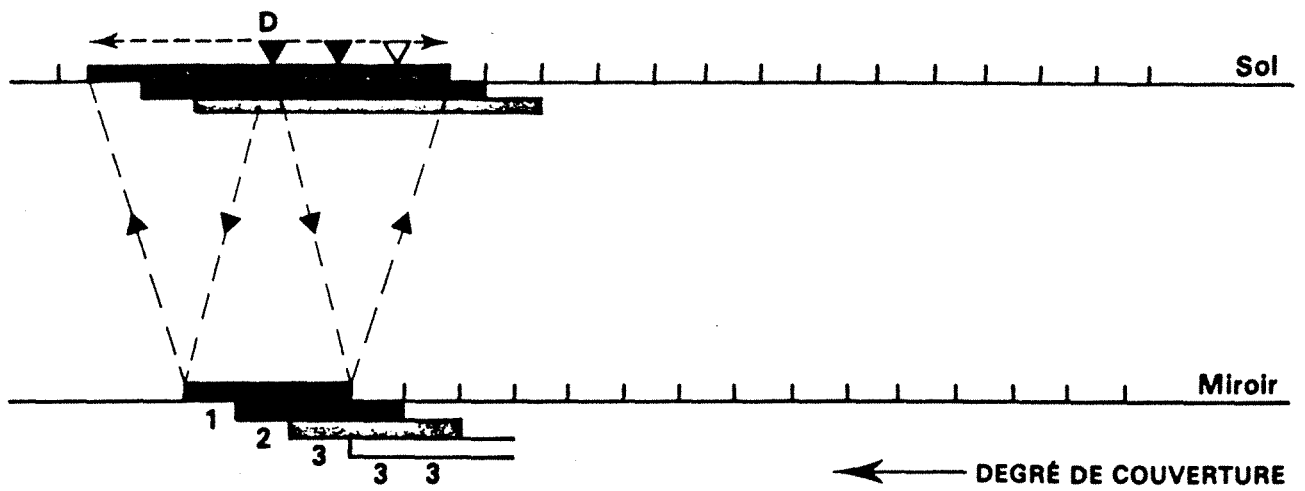
Un point clé dans la méthode sismique est l'optimisation du rapport signal/bruit, d'abord à l'acquisition, puis au traitement, afin de faciliter l'interprétation des événements sismiques les plus liés aux propriétés du sous-sol. Le premier grand pas de l'évolution de la sismique réflexion est d'ailleurs l'amélioration du rapport signal/bruit par l'invention du procédé de la **couverture multiple** (Waters, 1978).

Jusqu'en 1953, le filtrage du bruit ne s'opérait que par l'intermédiaire de filtres électriques incorporés dans les enregistreurs, utilisant le fait que le spectre de fréquences de certains bruits se situe en dehors du spectre utile du signal, soit 15 à 70-80 Hz.

Pour filtrer les bruits à l'acquisition, le géophone unique a été remplacé par plusieurs géophones, progressivement jusqu'à 48 et davantage, tous connectés au même canal et répartis sur une distance de 50 à 100 m, ou une surface de 1.000 à 3.000 m², de manière à atténuer les bruits qui se propagent horizontalement et à renforcer, en même temps, les ondes utiles réfléchies qui, elles, se propagent plus ou moins verticalement (Figure 11, Waters, 1978). En

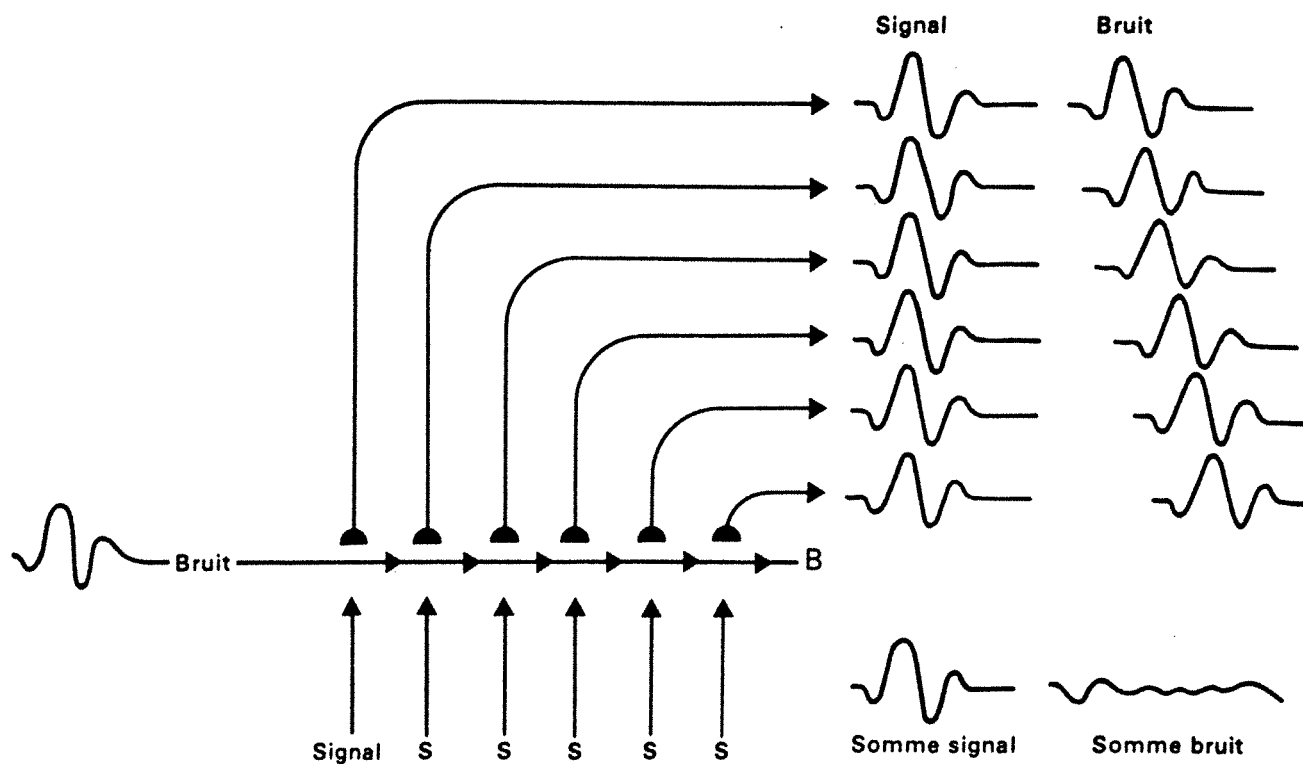


a. Principe de mise en œuvre en couverture simple



b. Mise en œuvre de la couverture multiple. Exemple de couverture 3

Fig. 11



Dans la figure théorique ci-dessus, le SIGNAL, S (RÉFLEXION) arrive aux CAPTEURS verticalement ou presque alors que le BRUIT ORGANISÉ, B y arrive horizontalement, de sorte qu'une TRACE constituée de 6 capteurs par exemple atténue le bruit organisé relativement au signal.

- Filtrage des bruits organisés

outre, par leur nombre, ils atténuent le bruit aléatoire, résultant de diffractions superficielles diverses.

De même, la source unique a été remplacée par plusieurs points répartis sur une distance ou sur une surface du même ordre que pour les géophones (Waters, 1978). Une amélioration notable du rapport signal/bruit des enregistrements a été ainsi obtenue.

Mais le problème de l'élimination des réflexions multiples restait entier.

Ce problème a été en grande partie résolu par le procédé dit de la **couverture multiple** (Waters, 1978). La couverture multiple consiste à couvrir les réflecteurs (Figure 10b) non seulement par la juxtaposition de traces uniques, mais par superposition d'un nombre plus ou moins important de traces, ce nombre définissant le degré de couverture (6, 12, 24, 48 et même 96). Ainsi, un point miroir (Figure 12) sera perçu à partir de plusieurs trajets différents "point de tir -groupe de récepteurs", la différence d'obliquité des rayons permettant d'atténuer les réflexions multiples à l'avantage des réflexions simples, à condition toutefois que la vitesse des terrains traversés augmente approximativement avec la profondeur (Figure 12). Les réflexions simples bénéficient, en outre, d'une amélioration du rapport signal/bruit aléatoire dans un rapport voisin de \sqrt{N} , où N est le degré de couverture.

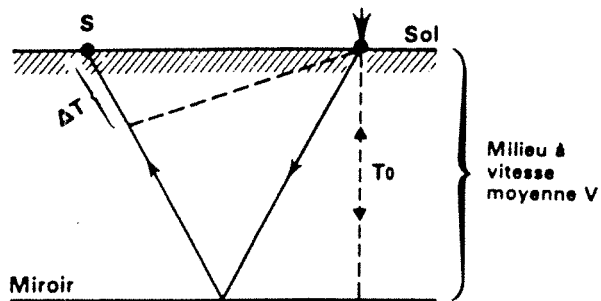
La couverture multiple a permis également une meilleure connaissance des vitesses ainsi qu'une conversion plus précise des coupes temps en coupes profondeur.

3.2. Principales sources utilisées

Les sources terrestres les plus utilisées actuellement sont l'explosif et le vibreur. Le signal émis par l'explosif est proche d'un signal à phase nulle, d'énergie importante. Si les charges sont enterrées sous la zone d'altération superficielle, qui s'étend en général jusqu'à la nappe phréatique, il y a très peu d'ondes de surface générées, donc de bruit.

Le vibreur, invention de la Continental Oil sous le nom de Vibroseis dans la décennie 1960, est basé sur une répartition de l'énergie de l'ébranlement sur une durée de plusieurs secondes (6 à 8) en faisant progressivement croître ou décroître la fréquence de la vibration jusqu'à couvrir toute la bande généralement utile en sismique, par exemple entre 20 et 80 Hz (c'est ce que l'on appelle le balayage ou le "sweep"). Cet effet est obtenu par un générateur transmettant

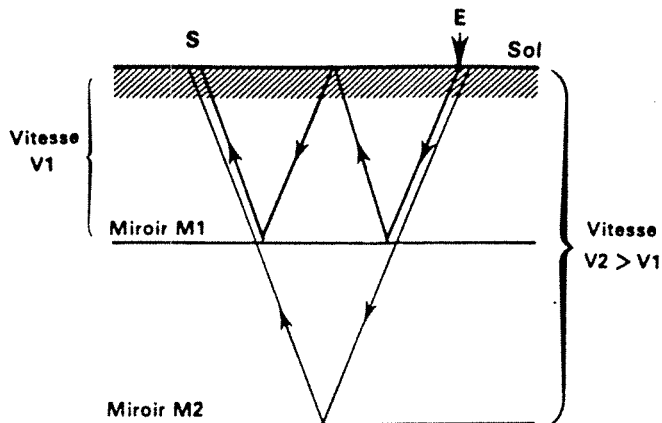
Fig. 12



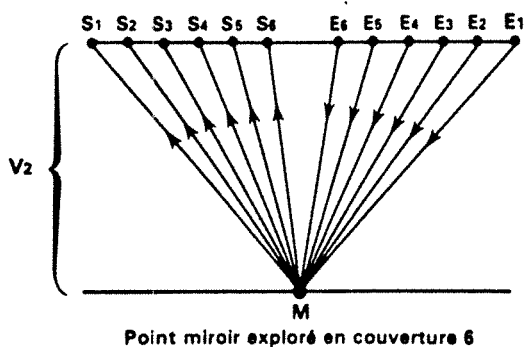
La correction d'obliquité (dynamique) ΔT à un temps T_0 à appliquer à chaque trace est fonction de la vitesse moyenne des terrains traversés et de la distance de la trace au point d'émission suivant la formule

$$\Delta T \sim \frac{x^2}{2 V^2 T_0}$$

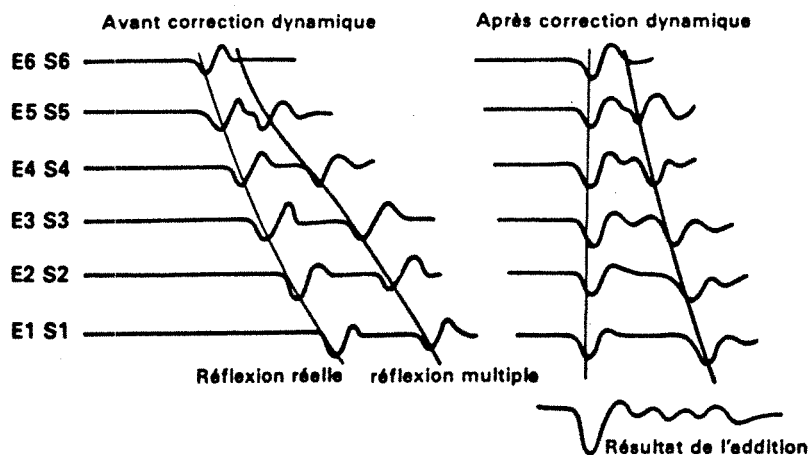
(en négligeant un terme du second ordre).



La réception de la réplique (multiple du 1^{er} ordre sur la surface du sol) sera plus tardive à un géophone éloigné de la source, que celle de la réflexion simple du fait de la différence de vitesse moyenne (voir formule ci-dessus).



En couverture multiple, un point miroir est exploré par des trajets d'obliquité différente. L'addition de ces enregistrements après correction dynamique calculée avec la vitesse V_2 sera favorable à la réflexion simple mais défavorable à la réflexion multiple, comme le montrent les sismogrammes théoriques ci-dessous.



- Atténuation des réflexions multiples par la couverture multiple

hydrauliquement ses vibrations au sol par l'intermédiaire d'une plaque que le poids du véhicule porteur (≈ 10 tonnes) maintient appliquée contre le sol.

Par la suite, on utilise l'intercorrélation entre la réponse reçue par les géophones et le signal émis -enregistré par ailleurs- pour obtenir un enregistrement comparable à ce qu'il aurait été sous l'effet d'une impulsion unique, telle qu'une charge d'explosif. Pour obtenir un rapport signal/bruit suffisant, il faut multiplier et ensuite sommer un nombre de vibrations suffisant, réparties sur une certaine surface considérée comme un point de tir. Le vibreur permet une bonne maîtrise du spectre de fréquences d'émission mais les corrections statiques (cf. § 4.4) sont un peu plus délicates à effectuer qu'avec l'explosif.

Les sources marines sont plus diversifiées que les sources terrestres. L'explosif a été très vite abandonné pour des raisons écologiques mais aussi techniques. Tout d'abord, des sources unitaires moins puissantes ont été utilisées : elles émettent simultanément, mais de façon répartie et pondérée, à l'aide de l'air comprimé ou de l'explosion de gaz en chambre (canons à air).

D'autres sources font appel, non plus à une explosion, mais à une implosion (exemple du Vaporchoc de la Compagnie Générale de Géophysique). Une bulle de vapeur d'eau surchauffée créée dans l'eau se contracte brutalement provoquant l'ébranlement souhaité. L'effet "bulle" (impulsions secondaires produites par l'oscillation de la bulle d'air dans l'eau) est totalement supprimé ; il ne subsiste qu'un faible signal précurseur dû au déclenchement de l'émission de vapeur.

3.3. Développements récents de la mise en oeuvre de la sismique réflexion

Signalons des développements pour augmenter le pouvoir de pénétration des signaux sismiques permettant ainsi l'étude des structures profondes (programme ECORS, C. Bois, 1989) ou des structures cachées par des zones superficielles intensément tectonisées. Un effort pour améliorer le pouvoir de résolution des données est aussi réalisé, mais dans un autre contexte : amélioration de l'interprétation lithologique de structures superficielles en tectonique relativement calme.

Des développements ont aussi trait à la logistique. L'exploration s'est étendue à des espaces de plus en plus difficilement accessibles : forêt dense équatoriale, zones marécageuses, mer profonde, et ceci avec des exigences de rendement et de qualité croissantes.

La technique de sismique réflexion à 3 dimensions (3D) permet depuis quelques années d'améliorer la densité et la répartition de l'information recueillie ainsi que sa position dans l'espace (Nelson, 1983). Cette technique consiste à acquérir les données, non plus suivant une ligne mais en étalant, par exemple, des dispositifs en croix, avec des lignes de sources perpendiculaires aux lignes de détecteurs.

La sismique 3D a d'abord été appliquée à l'estimation des gisements off-shore. Le coût des forages en mer est très élevé ; il faut donc apprécier les gisements découverts par un premier forage de façon très soigneuse car leur mise en production implique des investissements considérables. L'acquisition de campagnes 3D tend à se généraliser pour résoudre des problèmes de production rencontrés au cours de la vie du gisement.

La sismique réflexion classique, 2D, procède par l'intermédiaire de profils, le plus possible rectilignes, qui fournissent une information de caractère linéaire sous forme de coupes pseudo-verticales. Pour ce qui est de la détermination des pendages temps, il y a une méconnaissance de la composante orthogonale au profil. Cette composante qui permet de définir le pendage dans l'espace provient d'informations issues de profils différents et très proches. Pour la définir, la densité d'information à recueillir est beaucoup plus importante nécessitant une mise en oeuvre spéciale mais aussi des moyens de traitement plus importants et mieux adaptés. La précision de la position de chaque trace est indispensable et elle est maintenant possible grâce aux progrès réalisés dans les systèmes de radiolocalisation et aussi dans des systèmes de repérage de la position des traces le long du câble tracté par le bateau.

Dans la mise en oeuvre 3D, on obtient un découpage du volume à explorer en parallélépipèdes élémentaires d'information, qui, après des traitements classiques en 2 dimensions, seront mis en place dans l'espace (migration) pour tenir compte du pendage réel observé à partir de ses **deux composantes**. On peut alors réaliser des coupes dans tous les sens possibles, y compris le sens horizontal.

4 . TRAITEMENT DES DONNEES DE SISMIQUE REFLEXION

4.1. Introduction

Une présentation à la fois très complète et pragmatique des différentes étapes du traitement sismique, avec beaucoup d'exemples réels, est donnée par Yilmaz (1987) dans son ouvrage *Seismic Data Processing*, édité par la *Society of Exploration Geophysicists*.

Les éléments que nous indiquons ne constituent qu'un survol très bref de ce domaine.

Les objectifs principaux du traitement des données sismiques sont :

- . l'amélioration du rapport signal/bruit (élimination de tous les éléments sismiques qui pourraient perturber l'interprétation géologique),
- . l'amélioration de la résolution au niveau des objectifs,
- . la recherche d'une image correcte des structures géométriques du sous-sol (forme, mise en place) puis leur position sur la section **temps**,
- . la restitution des amplitudes les plus "proches" des propriétés géologiques du milieu traversé par les ondes acoustiques.

La séquence de traitement la plus usuelle comprend la récupération des amplitudes, les corrections statiques (terrestre), la mise en point miroir commun, la déconvolution, les analyses de vitesse, les corrections dynamiques, des corrections statiques résiduelles, la sommation, une déconvolution après sommation et éventuellement une migration.

Avant de débiter la chaîne de traitement, une mise en forme des données est effectuée avec les opérations de démultiplexage et de corrélation des signaux longs (si la source utilisée est un vibreur).

Le **démultiplexage** consiste à réarranger les échantillons enregistrés afin de rassembler séquentiellement la suite des échantillons correspondant à un même récepteur. Le pas d'échantillonnage est souvent de 4 millisecondes, en sismique classique, de 1 ou 2 millisecondes en sismique à haute résolution.

La **corrélation des signaux longs** permet, en **vibrosismique**, de comprimer la signature de la source.

Nous insisterons plus particulièrement sur les étapes du traitement qui s'avèrent capitales pour l'interprétation stratigraphique et lithologique de la sismique.

4.2. Traitement des amplitudes

Les manipulations d'amplitude sont déterminantes dans l'aspect des réflexions et conditionnent ainsi l'interprétation structurale (corrélations de part et d'autre des discontinuités, en particulier) mais surtout les conclusions concernant les faciès, la lithologie, les fluides.

Or souvent les paramètres du traitement des amplitudes sont laissés pour compte par rapport, par exemple, aux paramètres de la déconvolution.

Il faut garder présent à l'esprit que les **amplitudes constituent l'essentiel de l'analyse des traces qui est faite dans ce mémoire.**

4.2.1. Récupération du gain

Un échantillon sismique est représenté sur la bande sous forme de 2 nombres : l'amplitude brute, ou mantisse, et le gain.

La récupération du gain consiste à diviser l'amplitude brute par la valeur du gain pour obtenir l'amplitude vraie de l'échantillon.

A partir de là, deux traitements sont possibles qui ont pour objectif commun de compenser la perte d'énergie sismique avec le temps, liée à la divergence sphérique, l'absorption et les pertes de transmission.

4.2.2. Traitement en amplitudes égalisées

Une égalisation dynamique est appliquée à chaque trace sismique. La courbe de gain est calculée à partir de l'amplitude sismique moyenne dans une fenêtre glissante de longueur fixée. Les valeurs moyennes sont les coefficients diviseurs à appliquer aux amplitudes vraies.

L'égalisation est d'autant plus efficace que la fenêtre glissante est courte. Son intérêt réside dans l'amélioration des amplitudes des traces peu énergiques et l'atténuation des bruits très

forts. Les amplitudes de toutes les traces d'un même point de tir deviennent comparables aux mêmes temps, de sorte qu'on peut espérer bénéficier de toute l'amélioration du rapport signal/bruit aléatoire permise par l'addition en couverture multiple.

L'effet de l'égalisation dynamique n'est pas toujours bénéfique : elle produit, en particulier, un blanchiment des traces (zones "sourdes" artificielles) au-dessus et en-dessous des marqueurs très énergiques.

Les inconvénients majeurs de cette égalisation sont les suivants :

- les rapports d'énergie entre événements survenant à des temps différents ne sont pas conservés,
- la courbe de gain est différente pour chaque trace, de sorte que les variations d'énergie le long d'un horizon sont également perdues.

Ce traitement est donc à **proscrire pour toute interprétation stratigraphique ou lithologique.**

4.2.3. Traitement en amplitudes préservées

Cette appellation est un peu trompeuse car il n'y a pas d'amplitudes "vraies" en sismique, elle est à comprendre par opposition à l'égalisation dynamique.

Ce traitement ne fait appel qu'à des manipulations de gain **connues**, simples et **identiques** pour toutes les traces ou pour toutes les traces situées à une distance donnée de la source.

Cette courbe de gain peut être linéaire ou exponentielle, appliquée après avoir nettoyé la trace des très fortes anomalies d'amplitude (souvent fréquentes en acquisition terrestre). Le réglage de ce gain peut être délicat à faire ; d'autre part, une égalisation dynamique peut être nécessaire pour la sommation, en acquisition terrestre, que l'on réapplique ensuite pour revenir aux amplitudes préservées.

Malgré le coût supplémentaire du traitement "amplitudes préservées", il tend à se généraliser de plus en plus.

Toutes les données sismiques sur lesquelles s'appliquent les méthodes qui font l'objet de ce travail ont été étudiées en **amplitudes préservées.**

4.3. Mise en point milieu commun

Les traces sismiques, au moment de l'acquisition, sont enregistrées par regroupement pour un même point tir. Il faut donc les remettre en position "point milieu commun", c'est-à-dire les réarranger de telle sorte que tous les enregistrements qui concernent un même point du sous-sol soient réunis. La couverture multiple consiste, en effet, à sommer toutes les traces sismiques qui ont "éclairé" la même zone. Il faut remarquer que lorsque les réflecteurs sont pentés, il n'existe pas de point miroir commun, car chaque couple émetteur-récepteur donne naissance à un point miroir différent. C'est pourquoi, il est préférable de parler de point milieu commun.

4.4. Corrections statiques

Les corrections statiques interviennent principalement en **sismique terrestre**. Elles consistent à corriger les anomalies de temps de parcours introduites par les variations d'altitude des géophones et les variations de vitesse dans les terrains superficiels.

Les corrections statiques **terrain**, faites au moment de l'acquisition, consistent à ramener les temps de parcours à ce qu'ils seraient si sources et géophones étaient situés sur un plan de référence arbitraire ("datum plane"), un peu en-dessous de la zone altérée.

Les corrections statiques, **au centre de traitement**, interviennent dans le traitement sismique à trois reprises :

- les corrections calculées sur le terrain par la mission sismique sont améliorées (méthodes basées sur la réflexion ou sur la réfraction) afin de les amener au-dessus du seuil d'efficacité des statiques résiduelles ;
- des corrections statiques "résiduelles" sont ensuite calculées afin d'améliorer la sommation : elles sont dites "haute fréquence" car elles corrigent des anomalies superficielles dont la longueur est inférieure à la longueur du dispositif d'enregistrement ;
- des corrections "basse fréquence" sont parfois appliquées pour appréhender les anomalies de longueur supérieure à celle du dispositif.

Aucune de ces méthodes ne donne de renseignement sur les **temps absolus**, mais seulement des différences de trace à trace. Aucune ne distingue entre pendage des miroirs et variations d'épaisseur et/ou de vitesse dans les couches lentes superficielles.

Les programmes de corrections statiques utilisent tous la notion de **compatibilité des corrections en surface** ("surface consistent statics", Hileman et al., 1968, Taner et al., 1974). Une correction statique **unique** est liée à une position en surface, émission ou réception, donnée, quelle que soit l'inclinaison du rayon correspondant par rapport à la verticale. Cette condition est approchée de façon satisfaisante dans la mesure où la couche superficielle a une vitesse faible par rapport aux couches sous-jacentes.

On fait aussi l'hypothèse que les statiques sont des décalages de temps identiques pour toutes les fréquences.

4.4.1. Amélioration des statiques

Deux méthodes de terrains sont particulièrement utilisées, basées toutes deux sur l'étude des **arrivées réfractées**.

Le "stack dromo" (procédé de la Compagnie Générale de Géophysique) améliore la qualité des arrivées réfractées par une sommation (comme pour la réflexion) en point surface commun. L'addition est précédée d'une correction en x/v (équivalente des corrections dynamiques en sismique réflexion). La vitesse v est déterminée par un "balayage" du champ des vitesses potentielles. La vitesse v retenue est celle qui donne les arrivées réfractées les plus lisibles sur la section.

Les courbes de délai observées pour le marqueur étudié donnent les corrections statiques : différence, lue en position surface, entre marqueur et courbes directes et inverses.

Cette méthode a été remise en cause récemment : les arrivées réfractées sont additionnées **avant** application de statiques résiduelles (ce sont ces corrections qu'on se propose de calculer, par ailleurs). L'addition peut donner de mauvais résultats, et il faut réitérer la méthode. Une amélioration a été proposée pour remédier à cet inconvénient : la méthode des corrélations. On calcule les intercorrélations entre arrivées réfractées corrigées de x/v des traces successives des enregistrements individuels **sans addition**. L'addition est faite sur les corrélations, en position surface commune, des portions de dromochroniques provenant du même marqueur.

Le cumul des écarts ainsi déterminés donne les courbes de délai (en direct et en inverse) dont on prend la moyenne. La détermination précise de la vitesse du marqueur n'est plus indispensable. On lit les différences entre cette moyenne et une courbe simple qu'on se donne pour représenter le toit du marqueur suivi.

4.4.2. Corrections statiques résiduelles

La plupart des programmes de calcul des corrections statiques résiduelles utilisent les traces regroupées **en point milieu commun** car elles ont l'avantage capital de ne pas faire entrer le pendage en jeu. Ils mettent tous en jeu deux phases distinctes (Hileman et al., 1968, Taner et al., 1974).

La première phase consiste à déterminer par intercorrélacion les décalages entre traces individuelles et trace modèle. Les corrections qui en découlent sont donc liées à une trace donnée, influences du point d'émission et du point de réception confondues.

La trace modèle de la phase 1 est souvent la trace somme précédente : on intercorrèle les traces individuelles du point n avec la trace somme du point miroir n-1 (ou une addition pondérée des traces sommes n-2, n-1, n, n+1, n+2).

On obtient par application des corrections ainsi déterminées et addition en couverture multiple une coupe modèle préliminaire. Une réitération est alors faite : on intercorrèle les traces individuelles non corrigées avec la trace somme n corrigée correspondante.

Les corrections obtenues dans la phase 1 sont éclatées, au cours d'une phase 2, afin de donner une valeur de statique **unique** pour une position émission ou une position réception donnée.

Ce problème se réduit à un système d'équations du type :

$$T_{ij} = S_i + R_j + G_k + N_k X^2_{ij}$$

où T_{ij} est le temps de parcours entre source en position i et récepteur en position j,

S_i est le temps de parcours entre source en position i et plan de référence,

R_j est le temps de parcours entre le plan de référence et le récepteur en position j,

G_k est le temps aller-retour du plan de référence au point k du miroir,

$N_k X^2_{ij}$ est l'indicatrice résiduelle, liée au carré de la distance i, j.

Pour chaque couple source-récepteur, on dispose d'une équation de ce type, dont tous les paramètres sont à déterminer. Ce système linéaire est résolu par moindres carrés, des contraintes supplémentaires sont introduites afin d'obtenir une solution réaliste. L'algorithme de Gauss-Seidell (Wiggins et al., 1976) est le plus utilisé pour le résoudre. La solution converge en une seule itération pour les longueurs d'onde spatiales inférieures à la longueur du dispositif (les corrections à grande longueur d'onde ne sont donc pas traitées).

Le calcul donne un résidu non éclatable, composé de la dynamique résiduelle et d'une part, dite structurale, qui reste liée à la trace, mais des réitérations sont possibles, par retour à la phase 1 et reprise des analyses de vitesse (cf. § 4.6).

Le résidu doit diminuer au profit de la part explicable des statiques. Les corrections qui sont alors appliquées aux traces sont les valeurs éclatées en surface. On exclut d'appliquer les résidus car ceci nous ramènerait aux décalages de la phase 1.

De "bonnes" corrections statiques résiduelles ne sont possibles que si les corrections statiques terrain et améliorées sont déjà de bonne qualité. Leur rôle est déterminant dans l'amélioration de la sommation (meilleur pointé des vitesses, sommation plus efficace). Plus le problème étudié est fin -c'est le cas de notre étude du faciès d'un point de vue lithologique- plus les statiques résiduelles demandent attention et plus elles influent sur l'interprétation des traces sismiques qui sera faite. Très souvent, dans ce cas là, le pointé des analyses de vitesse (§4.6) est réeffectué après application des statistiques résiduelles. Les nouvelles vitesses sont utilisées pour recalculer les statiques résiduelles.

4.5. La déconvolution

Le choix de l'opérateur de la déconvolution appliquée aux traces sismiques aussi bien **avant qu'après sommation** est déterminant pour le pointé des événements et l'interprétation des faciès.

L'objectif de la déconvolution est, en premier lieu, d'estimer puis de comprimer le signal afin d'atténuer son influence sur la réponse impulsionnelle du sous-sol qui nous intéresse.

D'autres éléments perturbent la réponse impulsionnelle du sous-sol ; on peut aussi chercher à les atténuer à l'aide de la déconvolution, c'est le cas des multiples -fonction de démultiplication- ou des bruits organisés en bande étroite (ondes de surface, parasites industriels, etc.) -fonction de filtrage de la déconvolution-.

La déconvolution la plus utilisée en sismique est la déconvolution de Wiener-Levinson (1949). Les coefficients du filtre de Wiener (filtre de longueur déterminée) correspondent à la solution optimale au sens des moindres carrés. Le filtre de Wiener est causal ; il n'introduit donc pas de précurseur. L'estimation de l'**autocorrélation** du signal de départ et de son **intercorrélation** avec le signal désiré est suffisante pour calculer les coefficients du filtre (équations normales).

4.5.1. Déconvolution non prédictive et prédictive

On simplifie encore le problème posé dans les équations normales en limitant les types de signaux de sortie désirés. En déconvolution **non prédictive**, le signal de sortie souhaité est un Dirac ; elle tend donc à remplacer un signal dont le spectre contient des creux et des bosses par un signal dont le spectre est plat ("whitening déconvolution").

En déconvolution **prédictive**, on cherche à obtenir un signal de sortie identique au signal d'entrée pendant une certaine durée, appelée distance de prédiction, puis identiquement nul au delà. Le filtre n'est donc actif qu'après la distance de prédiction. On peut considérer que le filtre de Wiener prédit les valeurs futures de la trace à partir des valeurs passées. Ce futur correspond à la distance de prédiction et l'erreur de prédiction est la trace déconvoluée.

4.5.2. Evaluation du signal

Deux moyens sont utilisés en pratique pour obtenir l'autocorrélation du signal qui intervient dans les équations normales.

Le premier moyen est déterministe et utilise la signature de la source enregistrée, c'est un cas fréquent en acquisition marine. La déconvolution par la signature n'est efficace que sur le signal (et sa réflexion sur la surface de la mer ou "fantôme" de la source).

Le second procédé est le plus utilisé : on assimile la fonction d'autocorrélation du signal à celle de la trace sismique. L'hypothèse faite est que la série des coefficients de réflexion est totalement aléatoire, l'autocorrélation de la trace ne reflète donc que le signal. On peut espérer, en plus de la contraction du signal, une certaine atténuation des multiples mais le calcul de l'opérateur doit se faire sur une portion de trace suffisamment longue pour diminuer la probabilité d'une périodicité des coefficients de réflexion.

La fonction de filtrage est aussi l'apanage de la déconvolution statistique : le spectre d'amplitude de la trace déconvoluée est plus plat que celui de la trace brute. Si la trace brute contient un bruit en bande étroite très puissant dont la fréquence domine le spectre, la déconvolution, en remontant les autres fréquences au même niveau que le bruit, l'atténue en fait.

4.5.3. Problème de la phase

L'algorithme de Wiener-Levinson ne fait intervenir que l'autocorrélation du signal : toute information sur sa phase est perdue. Or la phase contrôle la forme de la trace sismique ainsi que l'interprétation géologique des marqueurs sismiques. En particulier, le pointé fin d'un horizon (pour un objectif stratigraphique ou lithologique) exige d'étudier une section sismique à **phase nulle**.

D'autre part, il est admis que la déconvolution ne fonctionne efficacement que si le signal d'entrée est à phase minimum (énergie concentrée au début).

En pratique, dans beaucoup de cas, on présente à l'entrée un signal dont la phase est mal connue et on obtient, après déconvolution, un autre signal dont la phase est différente, mais tout aussi mal définie. Ceci explique les difficultés de calage, très souvent rencontrées, des données sismiques au voisinage des puits, avec la trace synthétique calculée à partir des diagraphies Sonic et Densité.

4.5.4. Les paramètres de la déconvolution

La déconvolution remonte à un même niveau d'amplitude toutes les fréquences du spectre acquis, y compris celles qui correspondent à du bruit. La déconvolution entraîne toujours une élévation du niveau du bruit, d'autant plus importante que le rapport signal/bruit est plus faible au départ. Le choix des paramètres apparaît comme un compromis délicat entre le niveau de bruit et les effets attendus dans le domaine de la résolution et de la démultiplication.

Ces paramètres sont le type de déconvolution, le stade d'application, la longueur de l'opérateur, la distance de prédiction, la longueur de la fenêtre de calcul, sa position, le nombre d'opérateurs successifs (déconvolution variable), la position de la ou des fenêtres d'application (pas

nécessairement identiques à celles de calcul de l'opérateur), le pourcentage de bruit blanc, les traitements de la trace avant déconvolution : transformation de la phase, ou filtrage passe bande.

La **déconvolution stratigraphique** est un type de déconvolution très utilisé au préalable à l'analyse stratigraphique : l'opérateur calculé par moindres carrés assure la meilleure ressemblance possible entre le log d'impédance d'un puits et la trace sismique enregistrée. Toutefois, si plusieurs puits sont utilisés, il est souvent très difficile de trouver un opérateur unique à appliquer sur la zone, et impossible d'assurer une déconvolution cohérente avec un opérateur variant spatialement. Outre le spectre d'amplitude, la phase du signal peut être également déterminée à partir de la corrélation avec le log d'impédance. Pratiquement, cette recherche de la phase est très délicate, l'interprétation des données est très influencée par cette estimation, d'autant plus que l'objectif étudié est fin (c'est le cas des problèmes d'interprétation sismique pour le gisement).

Quant au stade d'application, une déconvolution au moins est toujours faite **avant** corrections dynamiques qui peuvent introduire des distorsions génératrices de bruit.

Théoriquement, la déconvolution après addition est plus efficace car elle porte sur des données dont le rapport signal/bruit a été amélioré par l'addition en couverture multiple. Pratiquement, il y a souvent production d'un niveau de bruit élevé et il y a un risque de faire disparaître l'information profonde, dont le niveau est faible, dans le bruit de la déconvolution.

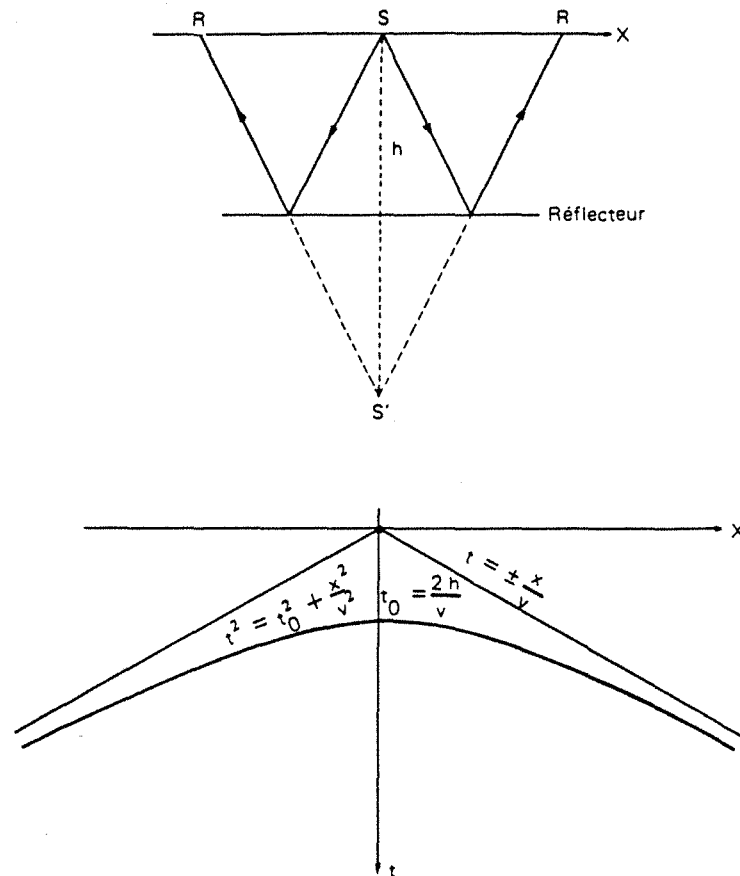
Une procédure souvent employée est de **compresser le signal avant addition** (déconvolution non prédictive, par exemple) puis de faire des essais de **démultiplication après addition**, avec un opérateur plus long et une distance de prédiction tenant compte de la longueur de l'opérateur employé avant addition.

La déconvolution peut apparaître comme une opération mathématique relativement simple, mais les résultats obtenus sont souvent surprenants. Le réglage des paramètres de déconvolution nécessite un nombre important d'essais, à discuter en fonction du problème géologique étudié.

4.6. Les analyses de vitesses

4.6.1. Principe de la détermination des corrections dynamiques

Dans le cas d'un réflecteur horizontal dans un milieu homogène (Figure 13a), le temps de trajet aller-retour de l'onde réfléchie s'écrit :



a - Temps d'arrivée des événements sismiques dans l'hypothèse d'un réflecteur horizontal dans un milieu homogène.

$$v^2 t^2 = x^2 + 4 h^2$$

où v est la vitesse de propagation,
 h la profondeur du miroir,
 x la distance source-détecteur.

La courbe $t(x)$ est une hyperbole d'équation :

$$t^2 = \frac{x^2}{v^2} + \frac{4h^2}{v^2}$$

admettant pour asymptotes les deux droites $t = \pm \frac{x}{v}$, qui, en fait, représentent les ondes directes se propageant à la vitesse v de la source aux géophones, le long de la surface du sol.

Lorsque le réflecteur est incliné avec un pendage α (Figure 13b), la vitesse de propagation v étant toujours supposée constante, le temps de trajet aller-retour de l'onde réfléchie est donné par la formule :

$$v^2 t^2 = x^2 + 4 h^2 + 4 h x \sin \alpha$$

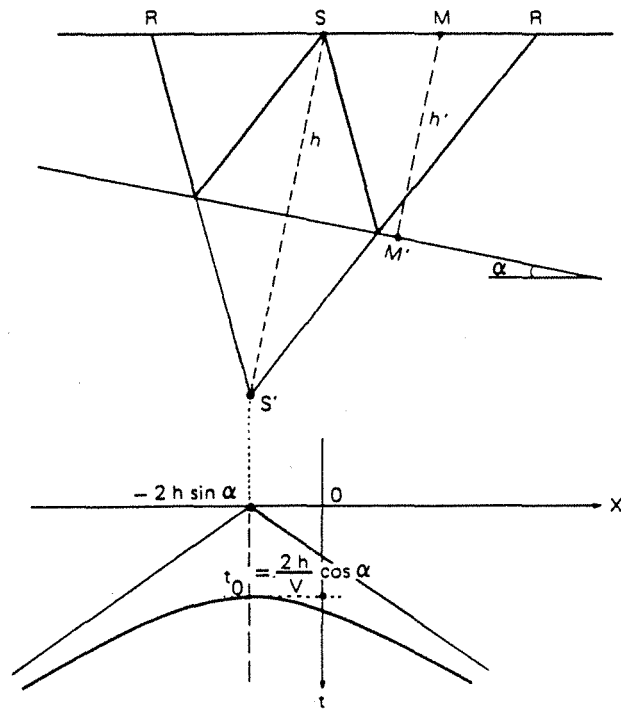
où h désigne la profondeur du réflecteur comptée perpendiculairement à partir de S (Figure 13b).

En remplaçant la variable h par $h' = h + \frac{x}{2} \sin \alpha$, le temps de trajet aller-retour est l'hyperbole d'équation :

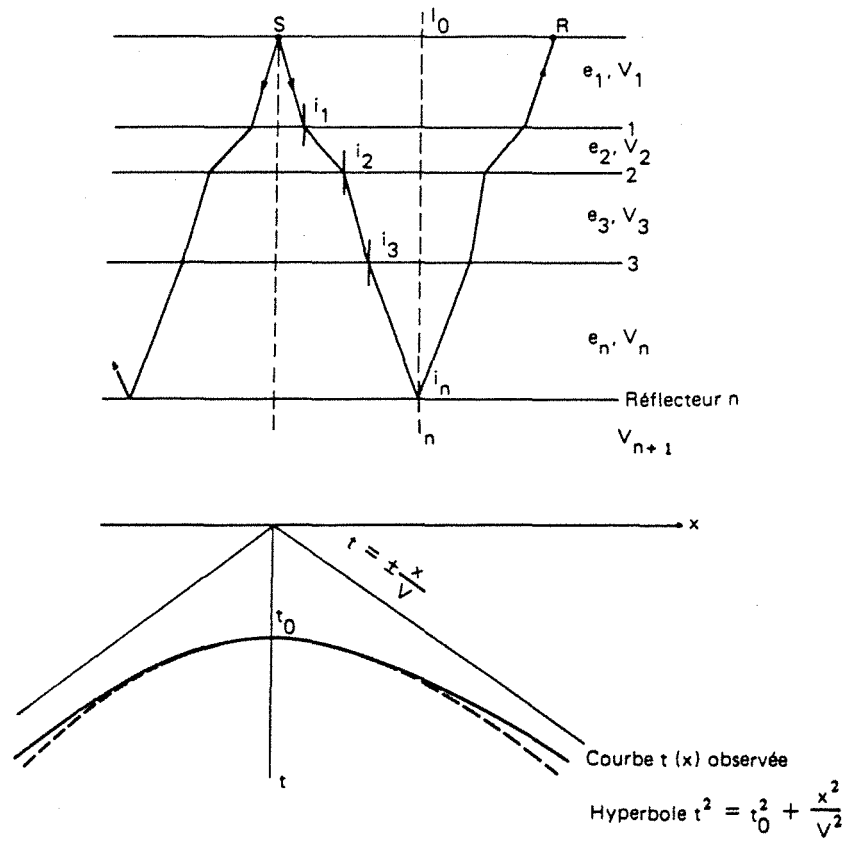
$$t^2 = \frac{x^2}{v_a^2} + \frac{4 h'^2}{v^2}$$

avec $v_a = \frac{v}{\cos \alpha}$.

Dans la pratique, l'hypothèse de vitesse constante n'est pas valable. Les vitesses des terrains varient verticalement et horizontalement ; le plus souvent, les variations verticales sont plus importantes que les variations horizontales. Sur toute la longueur du dispositif, on peut, par exemple, considérer en première approximation que les variations horizontales de vitesse sont négligeables et faire l'hypothèse d'un sous-sol à stratification tabulaire horizontale.



b - Temps d'arrivée des événements sismiques dans l'hypothèse d'un réflecteur incliné dans un milieu homogène.



c - Temps d'arrivée des événements sismiques dans l'hypothèse d'un empilement de couches horizontales.

Si le sous-sol est constitué d'un empilement de couches horizontales d'épaisseurs l_i et de vitesse v_i (Figure 13c), la courbe $t(x)$ du temps de trajet aller-retour entre S et R n'est plus rigoureusement une hyperbole, comme dans le cas du sous-sol homogène, mais une courbe à allure hyperbolique appelée l'indicatrice (Figure 13c).

Au voisinage de $x = 0$, l'indicatrice $t(x)$ est assez voisine d'une hyperbole d'équation :

$$t^2 = t_0^2 + \frac{x^2}{v^{*2}}$$

où t_0 est le temps de propagation verticale aller et retour de la source au miroir n et v^* une vitesse de propagation moyenne.

On peut montrer (Cordier, 1983) que :

$v^* = \bar{v}(n)$, vitesse quadratique moyenne pour le miroir n :

$$\bar{v}(n) = \sqrt{\frac{\sum_{i=1}^n v_i^2 \tau_i}{\sum_{i=1}^n \tau_i}}$$

où $\tau_i = \frac{2l_i}{v_i}$ est le temps de propagation verticale aller-retour dans chaque couche l_i de vitesse v_i .

Remarques : il ne faut pas confondre la vitesse quadratique moyenne $\bar{v}(n)$ avec la vitesse moyenne $V_M(n)$:

$$V_M(n) = \frac{\sum_{i=1}^n v_i \tau_i}{\sum_{i=1}^n \tau_i}$$

Dans les milieux stratifiés habituellement rencontrés dans les bassins sédimentaires, l'écart entre v_M et \bar{v} est de l'ordre de quelques pourcents (\bar{v} est toujours un peu plus élevée que v_M , Cordier, 1983).

Les vitesses moyennes quadratiques permettent d'obtenir les vitesses "d'intervalle" de chacune des couches de l'empilement, par la formule de Dix (1955) :

$$v_n^2 = (\bar{v}^2(n) t_n - \bar{v}^2(n-1) t_{n-1}) / (t_n - t_{n-1}).$$

4.6.2. Principe de l'analyse des vitesses (Figure 14)

On sélectionne sur la section sismique un certain nombre de points milieux communs, régulièrement espacés ou non, sur lesquels la loi des vitesses moyennes quadratiques en fonction du temps sera terminée.

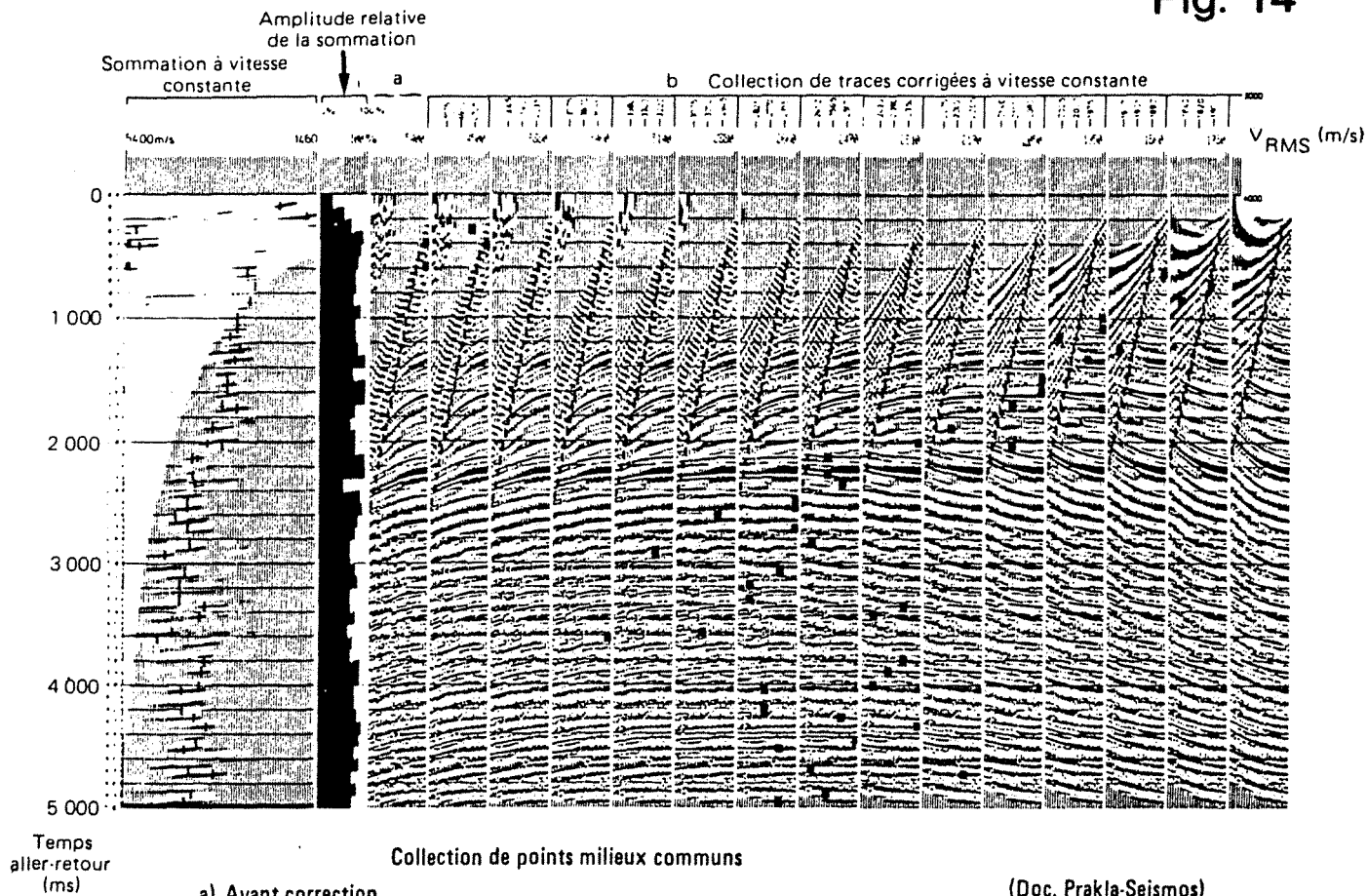
Le principe de cette détermination est le suivant : les traces constituant le point milieu commun étudié sont corrigées de la courbure de l'indicatrice pour toute une succession de vitesses moyennes quadratiques, depuis la plus lente jusqu'à la plus rapide qu'on puisse rencontrer dans la section étudiée (Figure 14). Si la vitesse de correction est la vitesse vraie, le réflecteur apparaît horizontal après correction de l'indicatrice ; sinon, il apparaît incliné vers le bas (vitesse trop forte) ou vers le haut (vitesse trop faible). Il est ainsi possible d'évaluer à l'oeil la vitesse moyenne quadratique de chaque réflecteur.

On peut également utiliser une méthode automatique d'évaluation de la vitesse moyenne quadratique qui consiste à faire la sommation ou à mesurer la cohérence des traces corrigées. Il y a un maximum d'amplitude ou de cohérence pour les vitesses correctes. Le calcul de l'énergie de la sommation ou de la cohérence est effectué en général dans des fenêtres d'une centaine de millisecondes.

Les analyses de vitesse permettent de déterminer la loi $\bar{v}(t)$ pour les points milieux communs étudiés, mais sont aussi très efficaces pour déceler les réflexions multiples.

4.7. La sommation

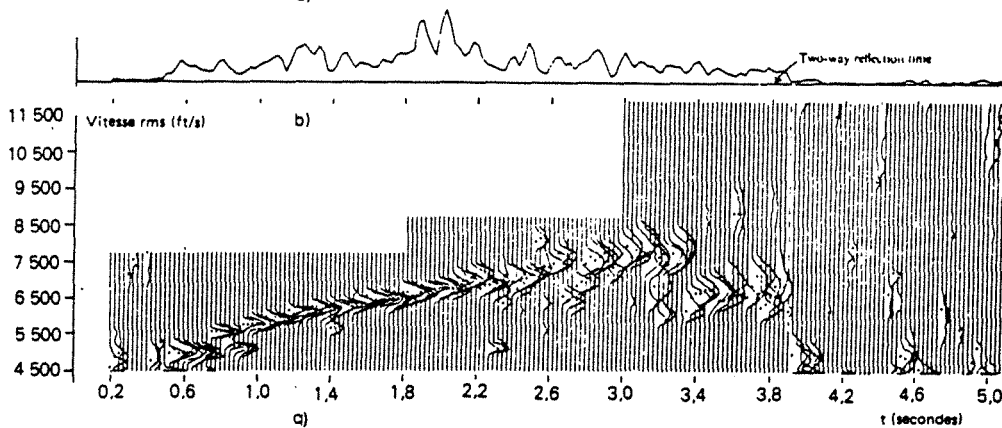
A partir des lois de vitesse moyenne quadratique $\bar{v}(t)$ déterminées sur un échantillonnage des points milieux de la section, la loi $\bar{v}(x,t)$ est calculée par interpolation (linéaire, le plus souvent)



a) Avant correction

b) Après correction de courbure des réflexions avec différentes vitesses.

(Doc. Prakla-Seismos)



Analyse de vitesse par mise en collection de points milieux communs, correction de courbure et sommation. a) Collection de points milieux communs.

b) Amplitude maximale de la sommation (c).

c) Amplitude des traces sommées en fonction de la vitesse dans des fenêtres de 100 millisecondes.

(Tiré de Telford et al., 1976, Applied Geophysics, Cambridge University Press, Cambridge, UK).

entre ces points. On peut alors appliquer à chaque collection de traces les corrections d'indicatrice -ou corrections dynamiques- et effectuer la sommation des traces d'un même point milieu commun (cf principe de la couverture multiple).

Nous rappelons que cette sommation améliore le rapport signal/bruit incohérent dans le rapport \sqrt{n} , n étant le degré de couverture, et permet aussi d'atténuer certaines réflexions multiples.

La section somme (ou section "stack") est l'équivalent de la section fictive qu'on obtiendrait en plaçant sources et détecteurs au même point, les rayons sismiques revenant sur eux-mêmes après réflexion en incidence normale sur les miroirs.

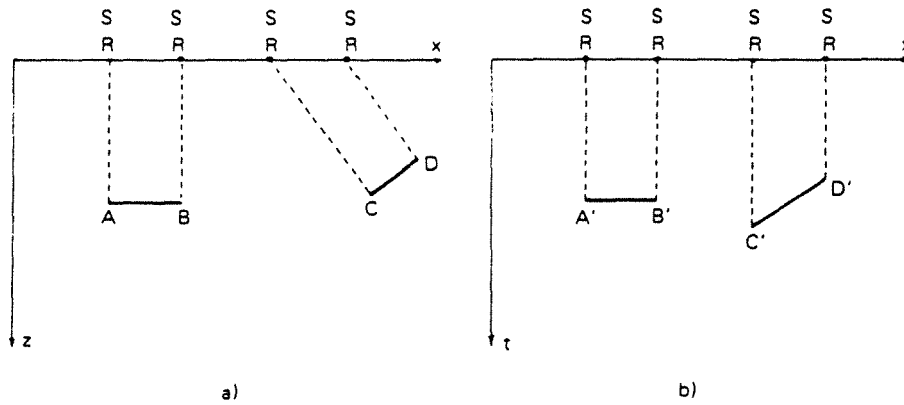
Avant la sommation, certaines portions de traces sont éliminées -"mute" ou remise à zéro- où le signal risque de subir des déformations trop grandes par la correction d'indicatrice. Les corrections dynamiques importantes concernent les horizons peu profonds et les traces éloignées du point de tir. Si la correction dynamique est très forte, la trace est étirée temporellement et déformée jusqu'à plusieurs millisecondes. On décide alors de mettre à zéro cette portion de la trace.

Un autre intérêt du mute est la suppression des arrivées réfractées parasites, mais aussi de toute partie de la trace perturbée par du bruit.

La définition du mute est particulièrement importante pour la sismique terrestre où les traces présentent souvent des anomalies d'amplitude très importantes. Il est indispensable d'éliminer ces portions d'enregistrement avant la sommation. L'ordre de couverture étant généralement assez élevé (48 à 96, par exemple), il reste toujours un nombre suffisant de traces pour donner l'information réfléchi. Il faut toutefois, par une pondération appropriée, tenir compte des portions de traces éliminées pour conserver les amplitudes vraies des réflecteurs après sommation.

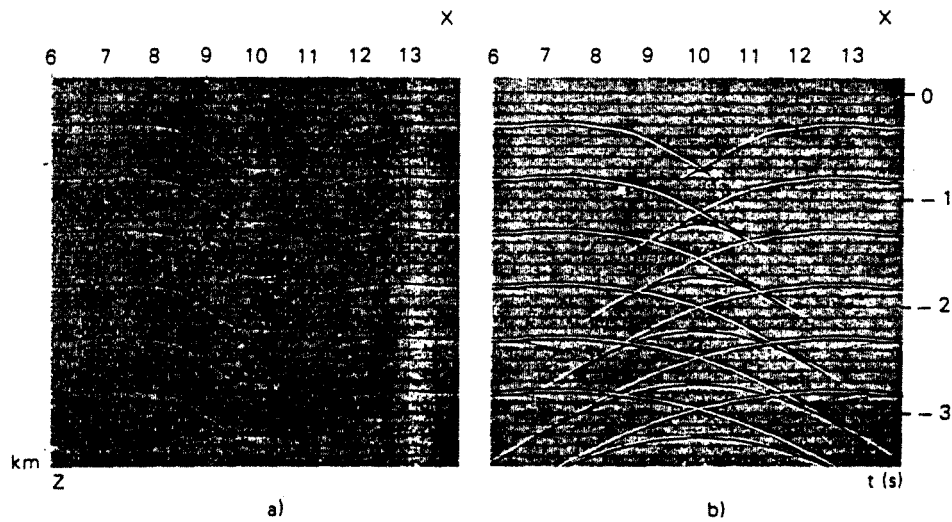
4.8. La migration

Sur la section somme, les réflecteurs inclinés ne sont pas représentés en position correcte (Figure 15). Les anticlinaux et les synclinaux donnent également des figures inexacts dans la représentation (x,t) (Figure 15). A la limite, les synclinaux peuvent prendre l'apparence d'anticlinaux, lorsque leur rayon de courbure est inférieur à celui du front d'onde incident ; on observe également des points de rebroussement dus à la présence de plusieurs trajets sismiques possibles.



Représentation d'un miroir horizontal AB et d'un miroir incliné CD en section profondeur xz et en section temps xt

- a) Section profondeur
- b) Section temps.



Représentation de synclinaux en section profondeur xz et en section temps xt.

- a) Section profondeur
- b) Section temps.

(Doc. Prakla-Seismos)

Le but de la migration est essentiellement de replacer les réflexions inclinées en bonne position par rapport au point milieu commun, sur la section (x,t) . Le passage à la section profondeur, à condition de connaître le champ de vitesses, n'est alors qu'une dilatation des échelles verticales.

Plusieurs techniques de migration existent, comme la méthode de sommation le long des hyperboles de diffraction et la méthode de l'équation d'ondes. La migration est une opération coûteuse mais pratiquement obligatoire dès que les pendages dépassent quelques degrés, en particulier pour les horizons profonds.

Les problèmes stratigraphiques et lithologiques étudiés dans ce mémoire s'inscrivent dans un cadre tectonique calme et superficiel, la plupart des données étudiées n'ont pas été migrées. C'est pourquoi, nous ne développons pas plus le sujet de la migration.

4.9. Conclusion

La plupart des étapes de la chaîne de traitement des données sismiques ont une influence déterminante sur l'interprétation des données qui sera faite. A chaque niveau, beaucoup de paramètres sont à choisir ; ce choix est **déjà une interprétation** et il est intéressant que le futur interpréteur des données y participe pour spécifier ses objectifs (structuraux, stratigraphiques, etc.) et le contexte géologique de son étude.

Après tous ces traitements où souvent une approche heuristique dans la détermination des paramètres prévaut, il est presque étonnant de pouvoir utiliser les données sismiques pour une interprétation aussi fine qu'est la sismique stratigraphique ou lithologique. Il faut donc garder présent à l'esprit le poids de toutes ces étapes dans le traitement ultérieur des données qui sera proposé dans ce mémoire, avec une attention spéciale aux problèmes de la récupération des amplitudes, de la déconvolution, des statiques et des analyses de vitesse, particulièrement en sismique terrestre.

5 . INTERPRETATION STRUCTURALE DES DONNEES SISMIQUES

C'est l'interprétation la plus traditionnelle qui est faite des données sismiques. Elle reste une étape indispensable à tous les stades d'étude d'un champ (Figure 4).

Il s'agit de définir la géométrie des objets géologiques comme dômes, anticlinaux (Figure 1a), failles (Figure 1b), toit du réservoir, couvertures, discordance, etc. Cette interprétation peut elle-même exiger un traitement très sophistiqué dans le cas des zones à tectonique complexe où la structuration superficielle intense masque ou fausse l'interprétation des structures profondes.

6. INTERPRETATION STRATIGRAPHIQUE DES DONNEES SISMIQUES

6.1. Introduction

L'interprétation géologique des sections sismiques a été développée dans la décennie 70 par une équipe de sédimentologues d'EXXON : J.B. Sangree, P.R. Vail, R.M. Mitchum, D.E. Frazier, R.G. Tood qui en ont posé les bases dans un mémoire désormais célèbre pour la profession : le mémoire AAPG (American Association of Petroleum Geologists) n° 26, Stratigraphic Interpretation of Seismic Data (1977).

Ce type d'interprétation exige souvent un retraitement spécifique des données sismiques, "orienté stratigraphique". En particulier, les amplitudes doivent être traitées avec grand soin : en effet, on ne se contente pas seulement d'analyser la forme des réflexions, mais on espère tirer partie des amplitudes des traces sismiques elles-mêmes pour retrouver les propriétés géologiques du sous-sol. Il est aussi souvent souhaitable de faire une déconvolution stratigraphique des données (utilisation du log d'impédance acoustique d'un puits) afin d'améliorer la résolution et le rapport signal/bruit au niveau de l'objectif d'étude, ainsi que de ramener le signal à phase nulle.

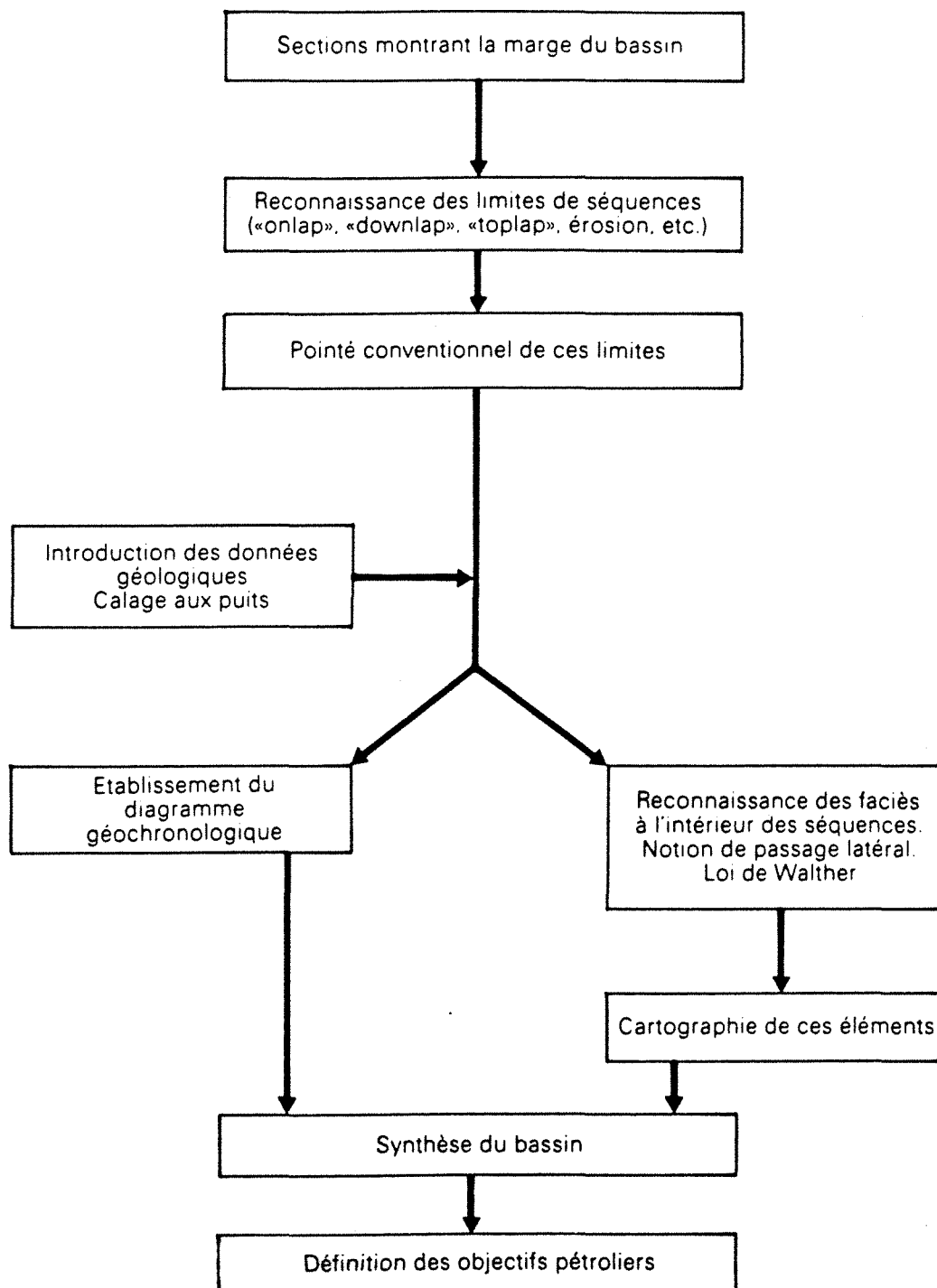
Les éléments qui suivent sont repris du mémoire AAPG n° 26 en insistant plus particulièrement sur l'analyse des faciès sismiques qui est l'objectif de la méthodologie statistique que nous avons développée.

6.2. Processus d'interprétation

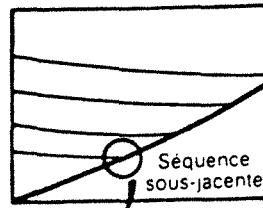
L'interprétation (Figure 16) consiste à subdiviser la section sismique en séquences (correspondant à des séquences de dépôt) puis à les dater à l'aide d'information de puits ou directement à partir des données sismiques. On peut distinguer quatre stades à cette interprétation.

6.2.1. Reconnaissance des différentes séquences sismiques

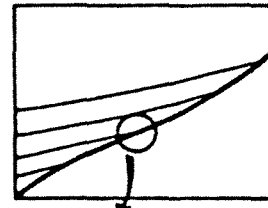
Une séquence sismique est un ensemble de réflexions concordantes limitées à la base et au sommet par des discontinuités reconnues à la terminaison latérale des réflexions. La figure 17 et le tableau 1 font la synthèse des différents types de terminaison ainsi que de leur signification géologique.



Ordonnement d'une étude de sismo-faciès.

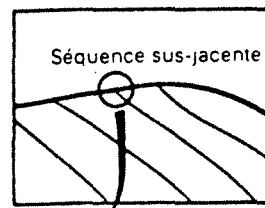


«Onlap»
a) «Onlap» horizontal

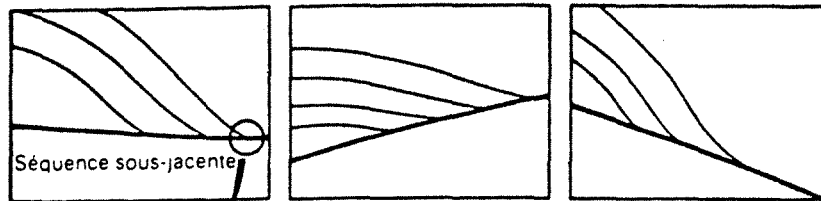


«Onlap»
b) «Onlap» d'amont-pendage

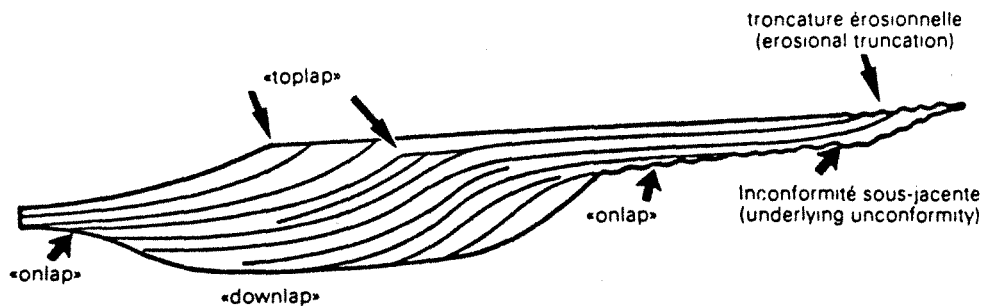
1 «ONLAP»



«toplap»
2 «TOPLAP»



«downlap»
3 «DOWNLAP»



(D'après Mitchum et al.)

Nomenclature des terminaisons de réflexion.

ABREVIATION	TERMINAISON	SIGNIFICATION
On	Onlap	Aggradation
Top	Toplap	Hiatus sédimentaire
Dow	Downlap	Progradation
Te	Troncature d'érosion	Erosion

Tableau 1 - Différents types de terminaison des réflexions

Une séquence de dépôts sismique correspond souvent à un cycle eustatique : élévation puis diminution du niveau marin.

Le tracé des discontinuités est étendu à l'ensemble des sections afin d'avoir une représentation spatiale de chaque séquence.

6.2.2. Détermination de l'âge des séquences

Cette détermination se fait souvent à partir du calage stratigraphique fourni par les puits (Figure 16).

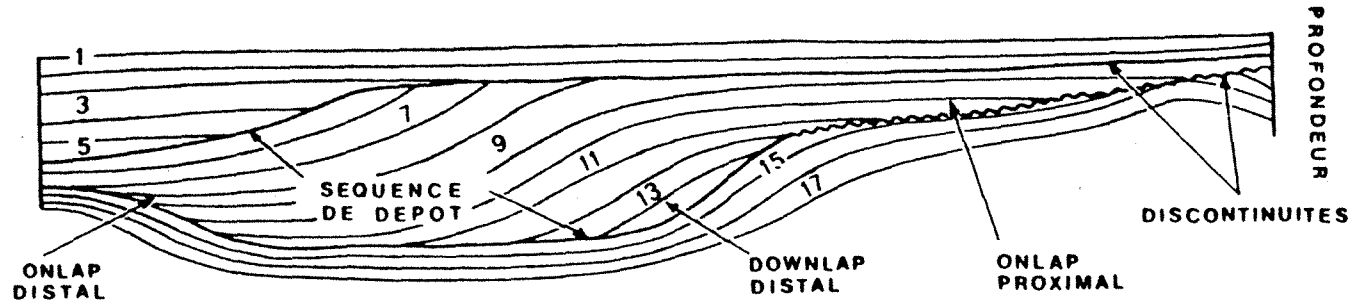
6.2.3. Construction d'une section chronostratigraphique (Figure 18)

Il s'agit d'une représentation où l'échelle verticale est graduée en périodes géologiques suivant le principe que les réflexions sont assimilables à des lignes temps géologiques. La section chronostratigraphique (Figure 18) résume :

- la nature des discontinuités,
- l'extension et la nature de la terminaison latérale des réflexions,
- les corrélations de séquences avec les périodes géologiques,
- l'importance des hiatus le long des discontinuités,
- la distribution des faciès et les milieux de dépôts.

SEQUENCE DE DEPOT

SECTION STRUCTURALE ET GEOLOGIQUE



SECTION CHRONO STRATIGRAPHIQUE

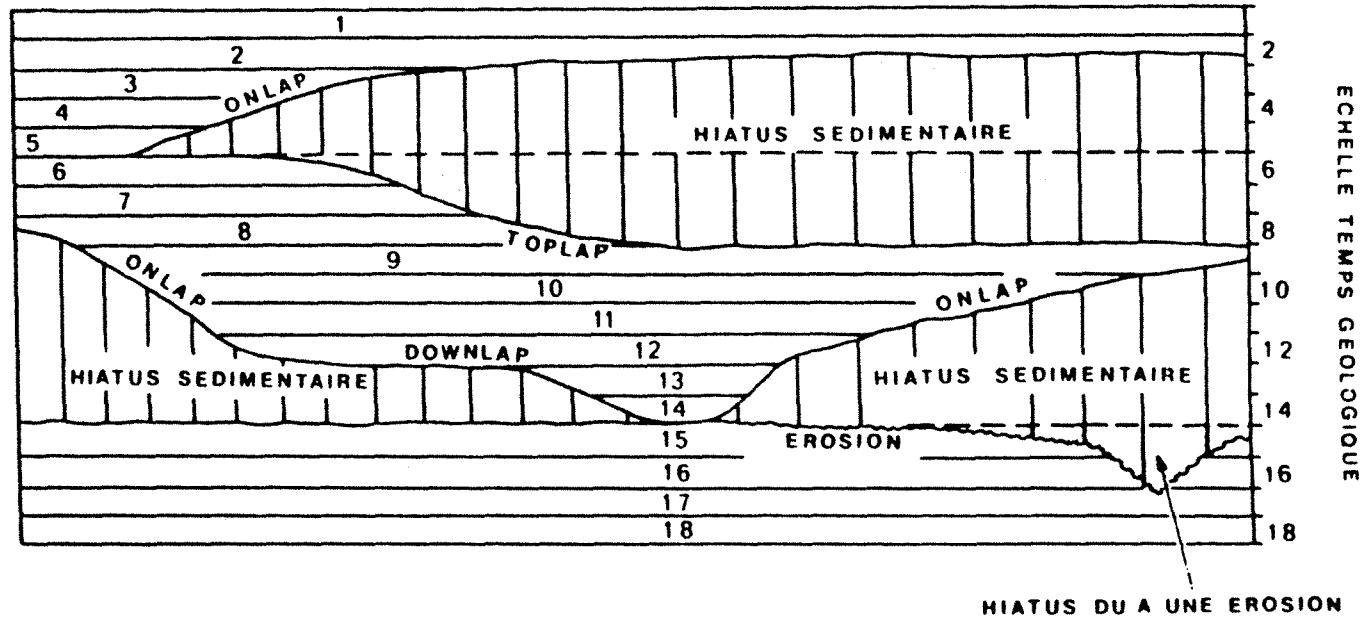


Fig. 18

6.2.4. Analyse des faciès sismiques

Il s'agit de reconnaître, cartographier et interpréter les unités de faciès sismiques. C'est pour cet objectif que nous avons développé une méthodologie statistique.

Une unité de faciès sismique est composée d'un groupe de réflexions dont les caractéristiques sont proches et différentes des unités adjacentes. L'interpréteur base son analyse sur les paramètres suivants : continuité, configuration des réflexions, amplitude, fréquence apparente et vitesse d'intervalle. Ces paramètres sont reliés à la lithologie, à la stratification et au mode de dépôt des sédiments.

La forme externe des unités de faciès sismique peut être aussi cartographiée.

Même si les limites de ces unités peuvent recouper les limites de séquence, ces unités doivent être interprétées dans le cadre de l'analyse séquentielle.

6.3. Séquences sismiques et séquences stratigraphiques

Les séquences sismiques sont déterminées entre les discontinuités qui les encadrent (Figure 18). Ces discontinuités se traduisent par la terminaison latérale des réflexions (Tableau 1 et Figure 17).

Différents types de terminaisons sont reconnues au sommet ou à la base des séquences (Figure 19) dont la signification géologique est bien établie (Tableau 1).

La cartographie de l'extension des différentes catégories de terminaisons latérales peut déjà être révélatrice de l'architecture du bassin, en désignant les zones de plate-forme, de pente et de bassin (Figure 20).

D'autre part, le type de terminaison est souvent relié aux variations relatives du niveau de la mer. Des "onlaps" côtiers vont traduire une élévation **relative** du niveau de la mer (Figure 21a) (transgression de la ligne de rivage avec approfondissement si les faciès littoraux migrent vers le continent, régression et remontée du fond de la mer dans le cas contraire).

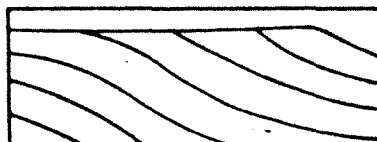
Une période de stabilité du niveau de la mer se traduit par l'absence d'"onlap" côtier mais la présence de "toplaps" côtiers (Figure 21b). Une chute brusque du niveau de la mer entraîne souvent une érosion des "onlaps" côtiers (Figure 21c).

RELATION ENTRE LES STRATES AUX FRONTIERES DES SEQUENCES

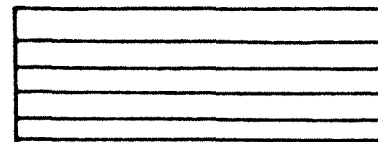
LIMITE SUPERIEURE



DISCORDANCE PAR EROSION

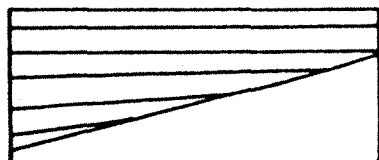


"TOPLAP"

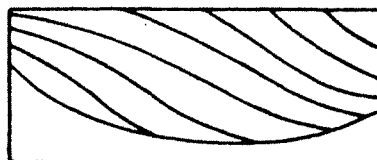


CONCORDANCE

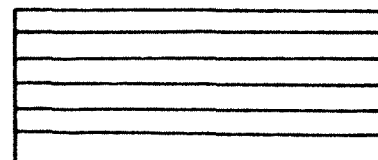
LIMITE INFERIEURE



"ONLAP"



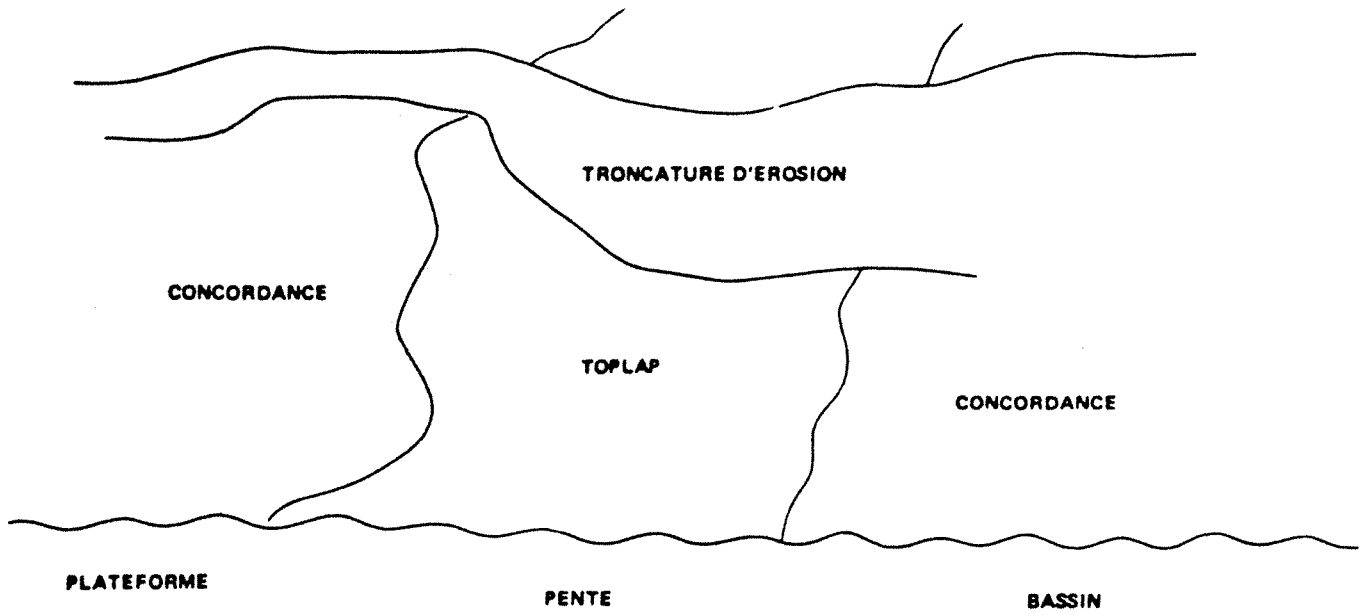
"DOWNLAP"



CONCORDANCE

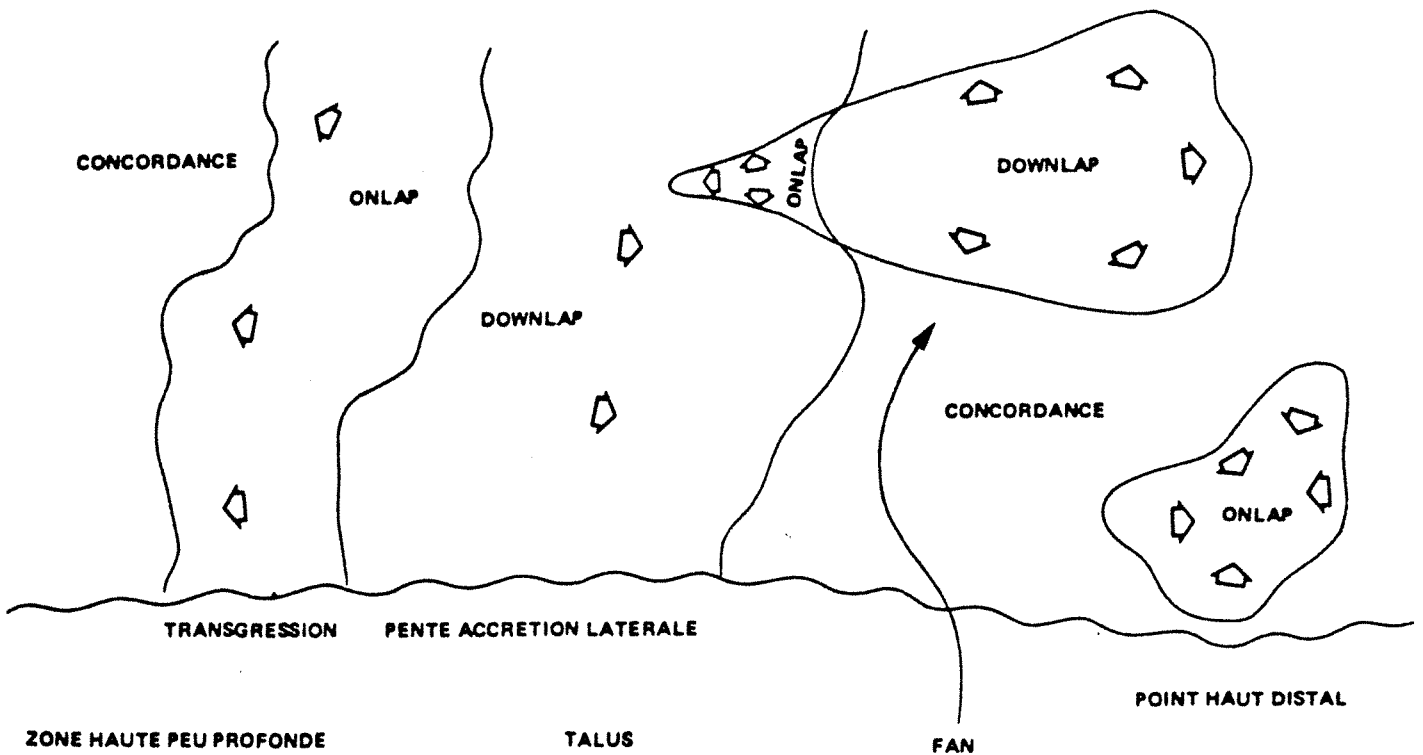
EXTENSION DES TYPES DE TERMINAISONS

LATERALES DES REFLEXIONS AU SOMMET DES SEQUENCES



EXTENSION DES TYPES DE TERMINAISONS

LATERALES DES REFLEXIONS A LA BASE



EXEMPLE DE REPARTITION SPATIALE DES DIVERS
TYPES DE TERMINAISONS LATERALES

BASSIN

EVOLUTION SEDIMENTAIRE EN PERIODE DE MONTEE RELATIVE DU NIVEAU DE LA MER

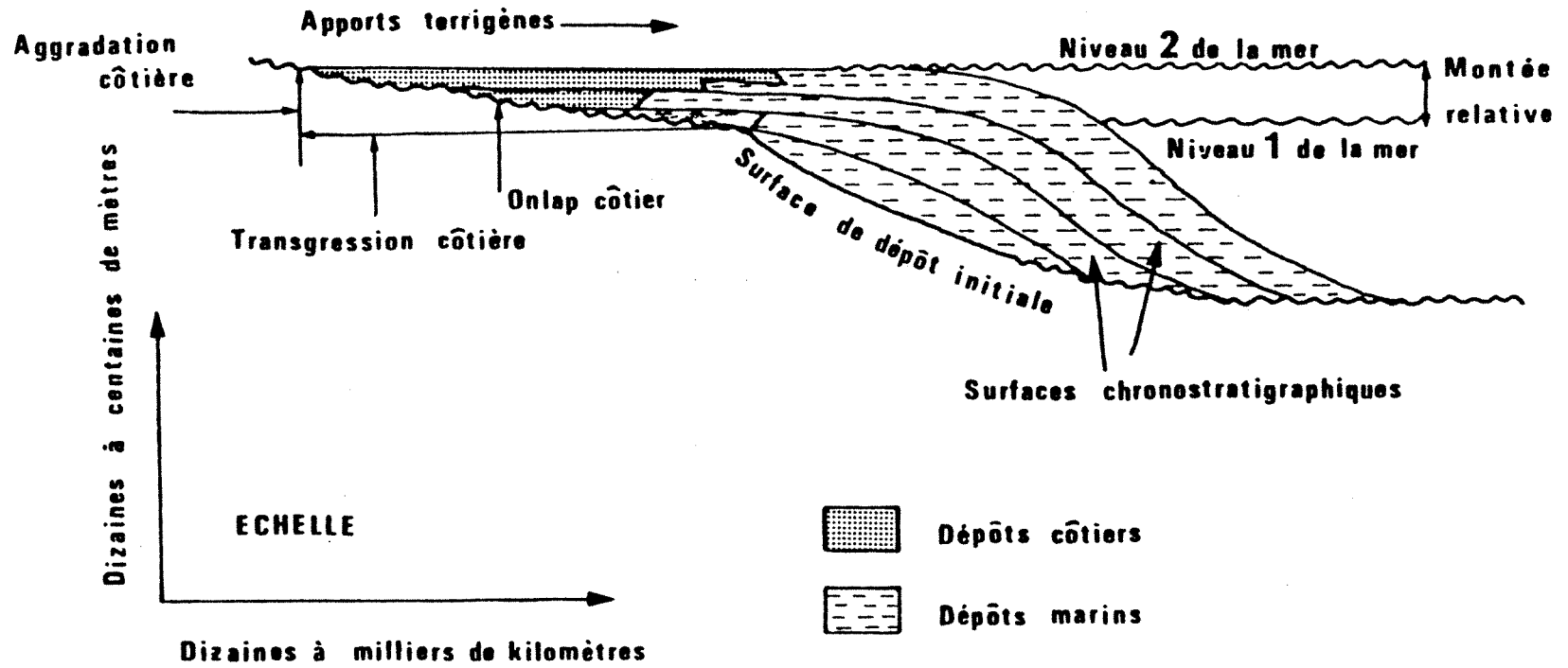


Fig. 21a

EVOLUTION SEDIMENTAIRE EN PERIODE DE STABILITE RELATIVE DU NIVEAU DE LA MER

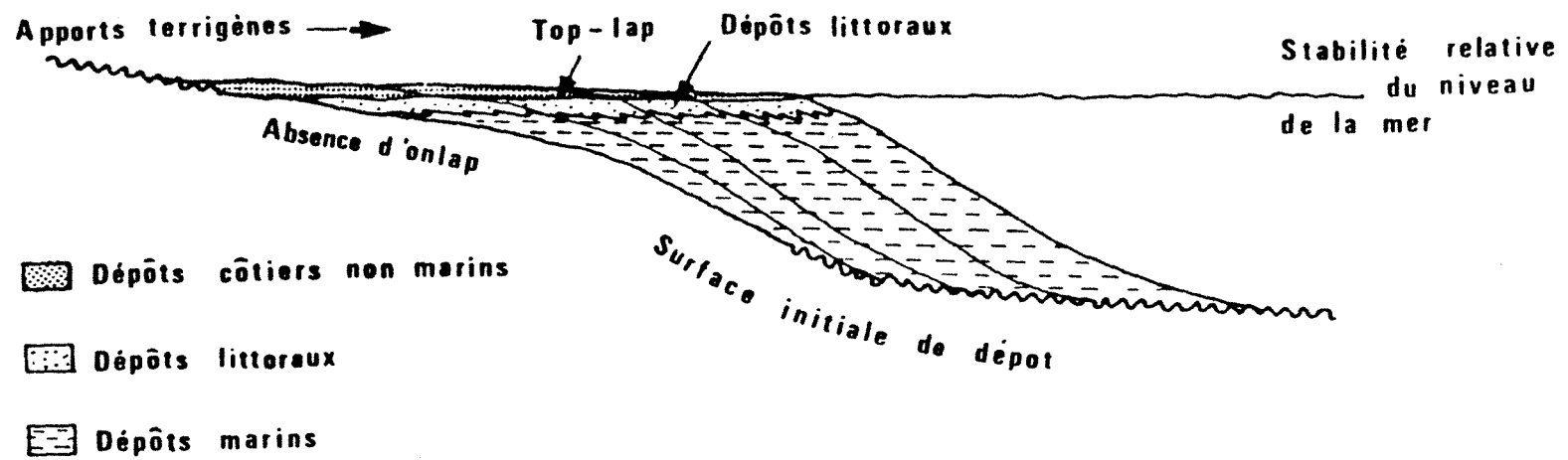
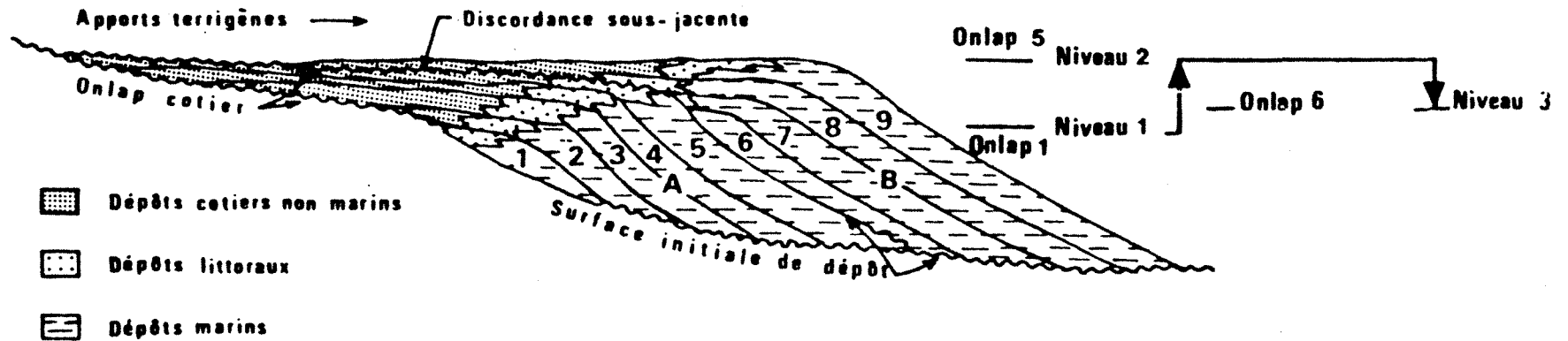


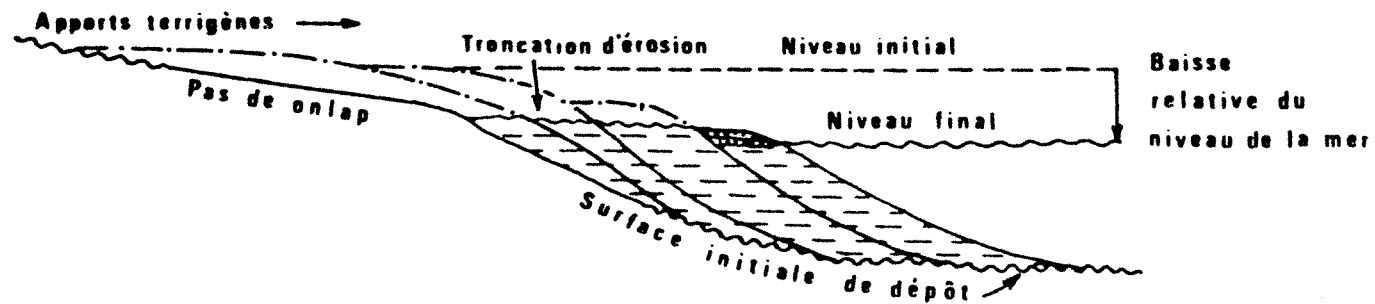
Fig. 21 b

MIGRATION DES ONLAPS COTIERS

a) UN DEPLACEMENT VERS LE BAS DES ONLAPS COTIERS
INDIQUE UNE BAISSSE RAPIDE



b) UN DEPLACEMENT VERS LE BAS DANS LE CLINIFORME
INDIQUE UNE BAISSSE GRADUELLE



L'analyse des séquences sismiques fournit donc des indications très importantes quant à la position relative du niveau de la mer par rapport à la plate-forme. Ces indications sont précieuses pour déterminer les zones où sont susceptibles d'apparaître des réservoirs, des couvertures, enfin des pièges intéressants. L'environnement géologique ainsi reconstruit permet aussi de déterminer le **type** de piège que l'on est en droit d'attendre. Par exemple, les périodes où la plate-forme est immergée correspondent, en général, au développement de dépôts de plate-forme du type deltas ou "off-shore bars". Tous les terrigènes se déposent sur la plate-forme ; si l'afflux de ces éléments est très important, les appareils deltaïques peuvent arriver en bordure du talus et être relayés par des éventails détritiques profonds. Au contraire, quand la plate-forme est émergée, elle est soumise à l'érosion ; les sédiments ne font que transiter par l'intermédiaire de canyons pour se déverser dans le bassin sous forme d'éventails profonds.

6.4. Analyse des faciès sismiques

L'analyse des faciès sismiques (Figure 16) est faite en étudiant la configuration des réflexions, leur amplitude, leur continuité, leur fréquence et leur vitesse d'intervalle.

6.4.1. Configuration des réflexions (Figure 22)

Ce paramètre est à relier au mode de dépôt. Les configurations suivantes sont possibles :

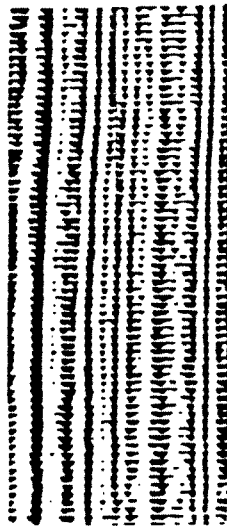
- réflexions parallèles,
- réflexions divergentes,
- réflexions progradantes, obliques, planes ou sigmoïdes,
- configuration en bosse et en creux,
- configuration chaotique,
- zone sourde.

Chaque type de configuration se rencontre dans des milieux bien particuliers.

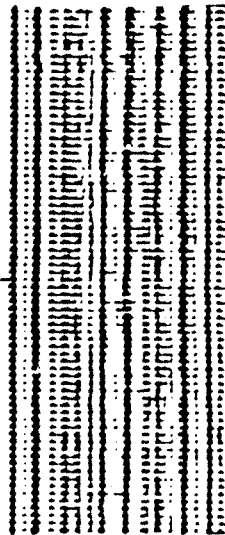
6.4.2. Forme externe

Cette forme externe donne également des indications sur le milieu et le type de dépôt.

ZONE STRATIFIEE



Simple



Parallèle

Complexe

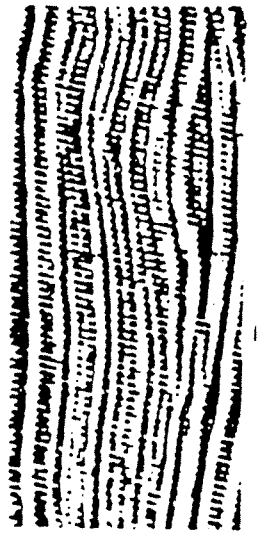


Oblique

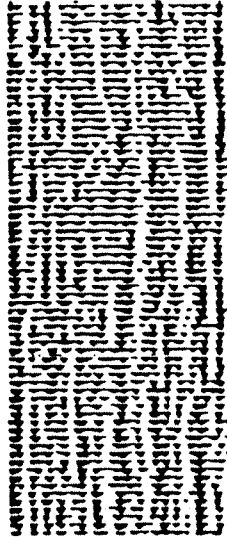
Divergent



Sigmoïde



ZONE SOURCE



CONFIGURATION DES REFLEXIONS

Fig. 22

Les formes en feuillets parallèles ou divergents sont les plus communes sur la plate-forme. Les formes en feuillets drapants indiquent une faible énergie de dépôt en eau profonde. Les lentilles se rencontrent dans les dépôts de pente. Les monticules sont souvent liés à un niveau d'énergie élevé : constructions organiques, turbidites, lobes deltaïques des "off-shore bars", etc. Les creux correspondent à des remplissages de dépression (canyons, chenaux, bassins, bas de pente, etc.).

6.4.3. Paramètres internes

Il existe une relation entre l'énergie du transport du dépôt et les réflexions issues des sédiments. Un haut niveau d'énergie de dépôt tend à produire des réflexions complexes, souvent discontinues, variables en amplitude, fréquence et phase. Inversement, un faible niveau d'énergie de dépôt se traduira par des réflexions très continues en amplitude, fréquence et phase sur de grandes surfaces. La **continuité** est directement liée à la continuité de la strate : elle augmente quand le niveau d'énergie du milieu diminue.

L'**amplitude** est liée au contraste d'impédance acoustique de part et d'autre de l'interface mais aussi à la composition interférentielle de tous les coefficients de réflexion convolués par le signal sismique. Ainsi, l'amplitude peut traduire le pourcentage et la distribution verticale des lithologies, au sein d'un intervalle donné.

Par exemple, le banc de 5 m de la figure 23, encadré d'argiles A, s'enrichit progressivement avec des intercalations d'argiles A. La longueur d'onde du signal utilisé est de 25 m. L'amplitude de l'ondelette varie avec le pourcentage entre les lithologies A et B mais est indépendante de la distribution des lithologies, **dans ce cas particulier**.

De fortes amplitudes peuvent aussi traduire un effet de résonance entre la série des coefficients de réflexion du sous-sol et le signal d'acquisition sismique. L'effet des bancs minces explique certaines anomalies d'amplitude ("tuning effect"). Lorsque l'épaisseur d'un banc est inférieure à la longueur d'onde du signal (Figure 24), il y a d'abord un effet de composition positive entre la réflexion au toit du banc et celle au mur ; cet effet atteint un maximum quand l'épaisseur du banc est égale à la demi-longueur d'onde du signal (Figures 24 et 25). L'amplitude de l'ondelette résultante peut être supérieure de 50 à 90 % à l'amplitude de l'ondelette initiale au toit du banc. A partir de ce maximum, l'amplitude décroît proportionnellement à la diminution de l'épaisseur. La fréquence apparente du signal est plus élevée (dérivée du signal entraînant un

Fig. 23

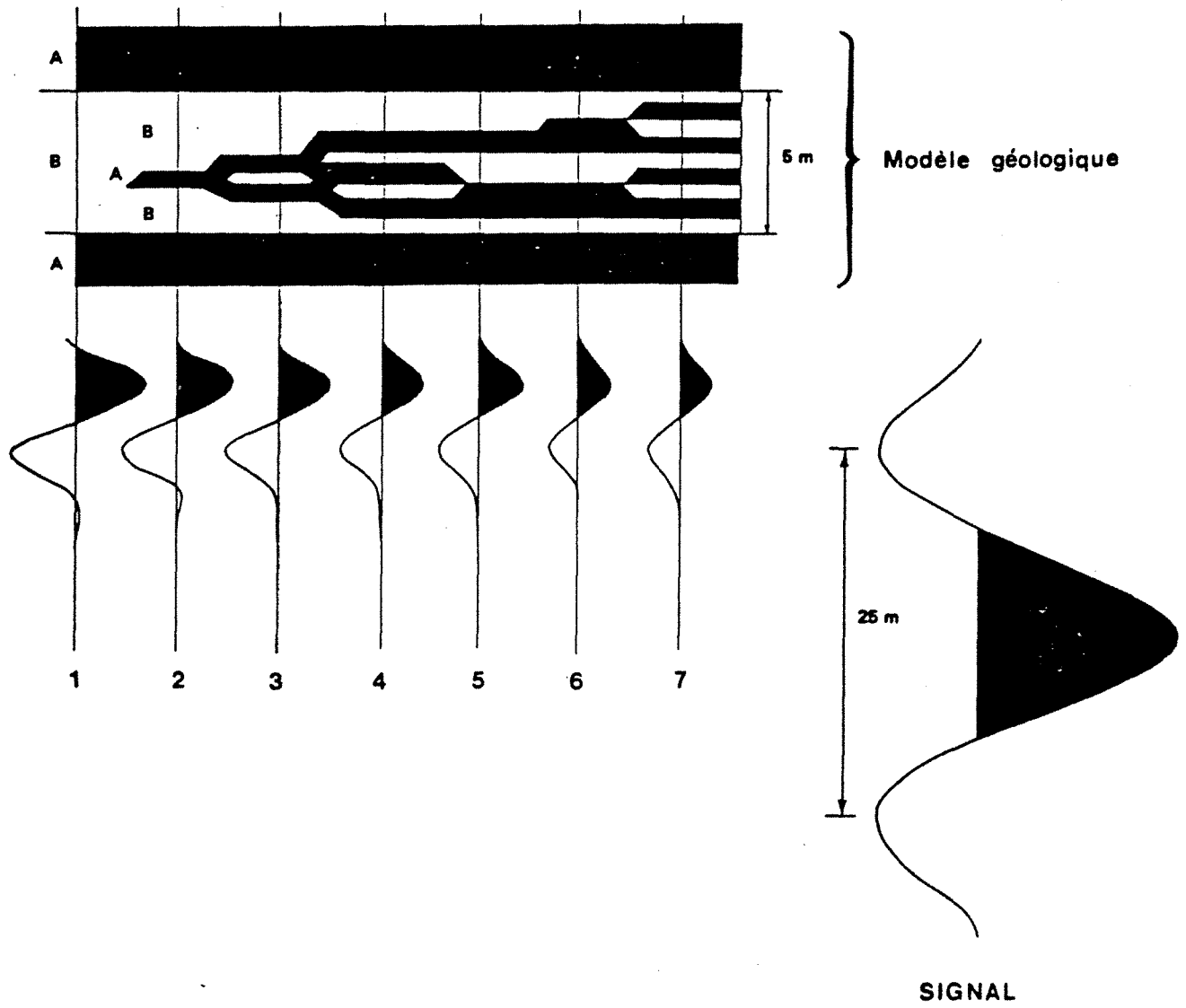
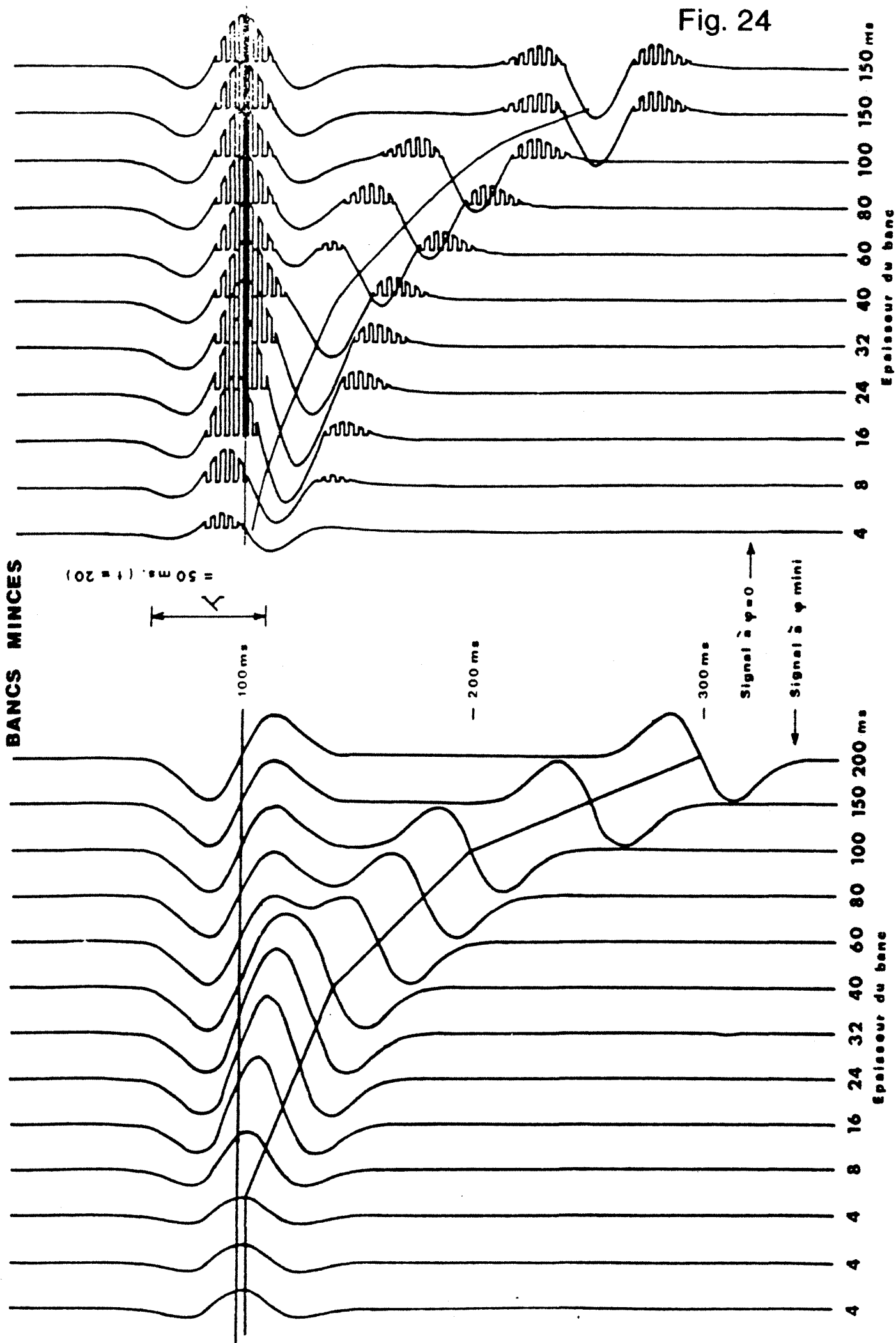
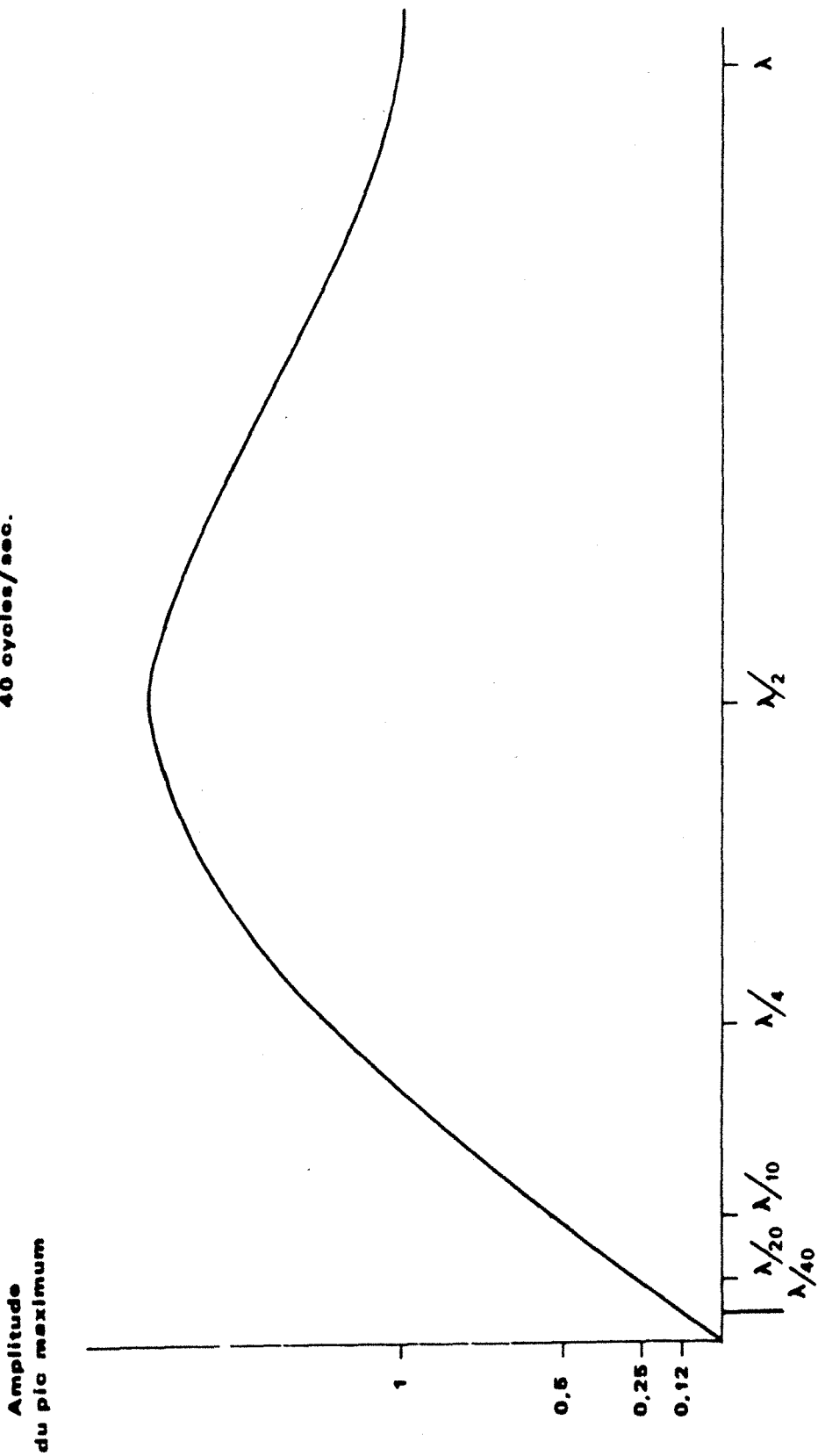


Fig. 24



BANCS MINCES

Un banc de dolomie de 1 m. d'épaisseur (0,5 ms TD) donnera la même amplitude qu'un contraste argille sur sable (2000 m/s : 2,45 / 2500 m/s : 2,10) en série deltaïque avec un signal de fréquence apparente de 40 cycles/sec.



Epaisseur du banc par rapport à la longueur d'onde apparente du signal

déplacement du spectre vers les hautes fréquences). La forme est aussi modifiée par rapport à celle de l'ondelette résultante.

La **phase** est un paramètre intéressant pour caractériser le faciès sismique. Une réflexion est en fait toujours la somme de plusieurs réflexions issues de surfaces de strates très proches les unes des autres. La composition de ces réflexions influe sur la forme de l'ondelette résultante.

La **fréquence** caractérise aussi le faciès sismique. Très globalement, les bancs minces se traduisent par des hautes fréquences apparentes, (dérivation du signal, Figure 26) et les zones de transition (effet de rampe) par des basses fréquences apparentes (intégration du signal, Figure 26). La distribution verticale des coefficients de réflexion des strates se traduit sur le spectre d'amplitudes. Toutefois, il ne faut pas oublier que l'aspect fréquentiel est **relatif** car il dépend du signal utilisé à l'acquisition. D'autre part, si l'on compare **latéralement** l'évolution du contenu fréquentiel, il faut faire l'hypothèse que les séries sus-jacentes n'ont pas introduit d'absorption différentielle des fréquences d'une zone à l'autre, afin de pouvoir relier les caractéristiques observées directement aux propriétés géologiques de la zone concernée.

Un ensemble de basses fréquences apparentes, sans variation latérale, correspondra souvent à des dépôts de faible énergie (argiles avec intercalations de silts dans un domaine marin, pro-delta ou turbidites distales). Un ensemble de basses fréquences apparentes intercalées avec des fréquences apparentes plus élevées, sans variation latérale, traduira, par exemple, des alternances de dépôts marins argileux et sables littoraux. Un ensemble à intercalations de zones sourdes et de marqueurs à fréquences normales ou élevées, d'une grande continuité latérale, traduira des dépôts de faible énergie sur une plate-forme stable, régulièrement subsidente (argiles avec intercalations de bancs minces de carbonates ou silts).

Les **vitesse d'intervalle** peuvent varier latéralement traduisant des variations lithologiques, de porosité, de fracturation ou de fluides.

6.4.4. Conclusion

L'analyse des faciès sismiques comprend cinq stades :

- étude des paramètres internes,
- délimitation des zones présentant les mêmes paramètres (ce sont les unités de faciès sismiques),
- cartographie de l'enveloppe de chaque unité de faciès sismique,

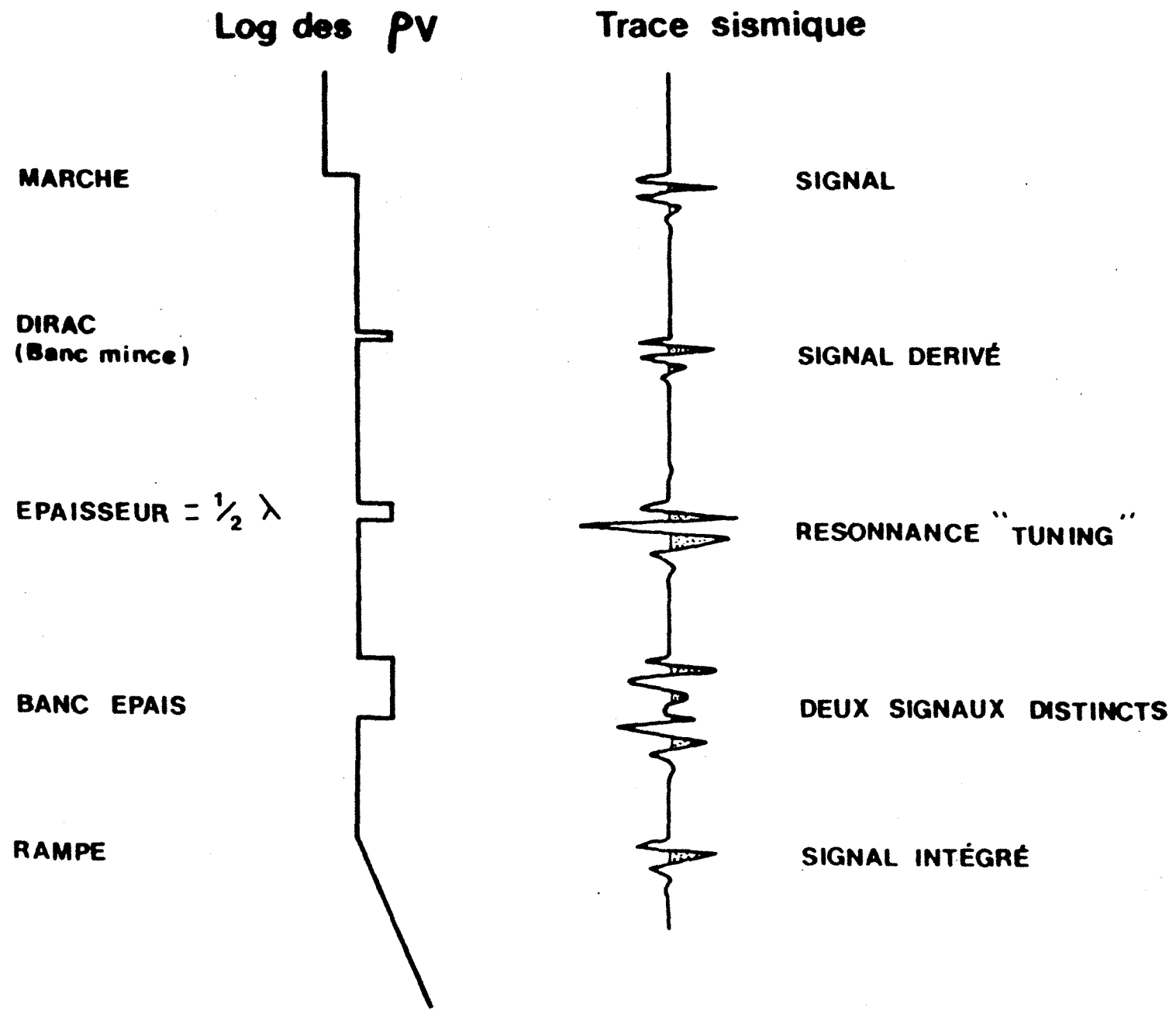
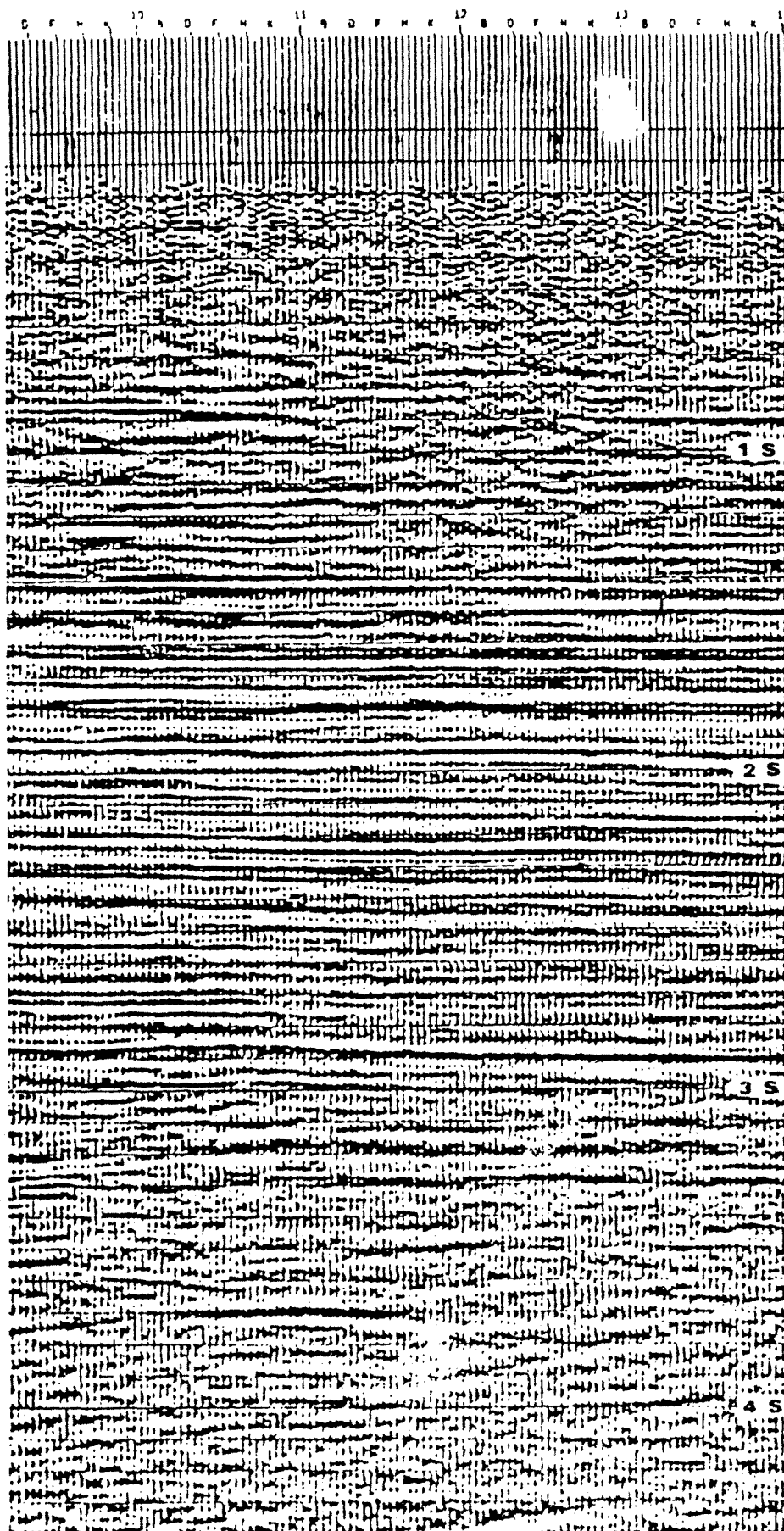


Fig. 26

- étude des associations latérales entre les différentes unités,
- étude des associations verticales des différentes unités entre elles (Figure 27) avec interprétation en liaison avec les séquences sismiques.

ENCHAINEMENT SEQUENTIEL VERTICAL
DES UNITES DE FACIES SISMIQUE

Fig. 27



Réflexions discontinues
variables en fréquence,
phase et amplitude

FORTE ENERGIE

Facies fluviatile
méandrique ou facies
margino - littoral

Réflexions de bonne continuité
variant peu latéralement.
Alternances de HF apparentes et
BF apparentes plus fréquentes
vers la base

ENERGIE DE DEPOT VARIABLE
Facies littoraux, sables et argiles

"Bright spot" lithologique
argile marine / barrier foot + (BF
transition)
Zone de transition aux argiles
sous compactées (BF apparent)

Zone sourde
argile sous compactée
+ bright spot ; poche de gaz
surcomprimé

FAIBLE ENERGIE

facies marins
+ sables de pieds de talus

EXEMPLE D'UNE SEQUENCE REGRESSIVE

7. INTERPRETATION STRATIGRAPHIQUE OU LITHOLOGIQUE ET ANALYSE AUTOMATIQUE

L'interprétation stratigraphique des données sismiques (Figures 5 et 27) s'applique à des échelles très variées, fait appel à des concepts différents et met en oeuvre des techniques très diverses.

La **détermination des séquences sismiques** est fondée sur le **pointé** des réflexions et l'analyse de leurs terminaisons latérales. Ce pointé est une action géométrique dont l'automatisation correspond à la technique de "pointé automatique" déjà disponible sur la plupart des stations d'interprétation en exploration pétrolière.

L'**interprétation des séquences** par rapport aux variations relatives du niveau de la mer, et la déduction qu'il en est faite des paléo-environnements de dépôts (Figure 2) relèvent essentiellement d'une expertise géologique. On ne peut pas vraiment parler d'automatisme pour cette tâche. Par contre, on peut imaginer rentrer les règles d'expertise dans un système expert aidant au diagnostic.

L'**analyse séquentielle sismique** correspond à un découpage à grande échelle de l'image sismique (Figure 27).

Dans l'**analyse des faciès sismiques**, l'échelle de travail est plus réduite. Ce sont surtout les variations latérales de l'aspect des traces sismiques, dans une moindre mesure les variations verticales, qui sont étudiées **au sein d'une unité séquentielle** (Figure 27). La forme externe des unités de faciès et leur intégration dans le cadre géologique sont encore des problèmes géométriques puis de modèle conceptuel géologique. Par contre, l'analyse des caractéristiques internes des faciès (fréquence, phase, amplitude), en vue d'**une interprétation lithologique** du milieu rencontré par l'onde acoustique, bénéficierait grandement d'une automatisation pour renforcer l'objectivité de l'analyse, quantifier les paramètres, intégrer plus systématiquement -comme données de calibrage- l'information géologique disponible aux puits. Cette étude lithologique des traces sismiques a lieu très souvent à une échelle beaucoup plus petite que l'analyse séquentielle (≈ 100 ms, temps double). D'autre part, c'est bien cette étape de l'interprétation qui est typiquement multivariable : une unité de faciès sismique est définie par un **ensemble** de paramètres internes différents de l'unité adjacente.

Le travail présenté dans ce mémoire porte sur la **caractérisation interne des faciès sismiques** en vue d'une liaison avec les propriétés géologiques de la série (lithologie,

pétrophysique, fluides). Il ne concerne donc qu'une petite partie de l'interprétation stratigraphique des sections sismiques ; mais cette partie est déjà tournée vers les problèmes très fins d'**analyse lithologique** rencontrés en **géophysique de gisement**, au stade du développement d'un champ. Ceci explique pourquoi de ces travaux ont dérivé des recherches actuelles appliquées à des problèmes de gisement sur "l'inversion statistique" des traces sismiques.

DEUXIEME PARTIE

**ANALYSE DU FACIES SISMIQUE SUR UNE FENETRE
TEMPORELLE D'ENVIRON 100 MS**

1. POSITION DU PROBLEME

1.1. Introduction

En analyse interne des faciès sismiques, les paramètres utilisés sont l'amplitude, la phase, la fréquence, la vitesse d'intervalle et la continuité des réflexions. Certains de ces paramètres sont redondants et tous, dans l'analyse stratigraphique classique, sont utilisés de façon empirique.

Afin d'automatiser la reconnaissance du faciès sismique et d'aller vers une approche plus quantitative de ses caractéristiques en liaison avec les propriétés géologiques, il faut au préalable explorer le plus exhaustivement possible les paramètres sismiques permettant de caractériser la **morphologie** de la trace (c'est-à-dire son faciès). Dans ce cadre de **l'analyse lithologique**, nous supposons le contexte séquentiel et le pointé de la zone des réservoirs (ou de l'objectif) connus. Les traces sont donc analysées, au niveau de l'objectif, sur une portion temporelle déterminée par ce pointé.

Dans ce chapitre, nous faisons le point sur les travaux **antérieurs** ou **contemporains** au début de notre étude, concernant le problème de l'extraction de paramètres sismiques traduisant les propriétés géologiques. La critique de ces travaux permet d'expliquer l'approche qui a été suivie.

Des travaux plus récents seront analysés pour comparaison avec la solution méthodologique que nous proposons, dans le paragraphe 4.4 de cette partie : Discussion sur les limitations de la méthodologie proposée - Conclusions.

1.2. Etat de l'art

1.2.1. Analyse d'attributs sismiques classiques

En analyse classique des faciès sismiques, les paramètres utilisés par l'interpréteur pour la classification des faciès sont des paramètres de continuité, de fréquences apparentes. Une étude (Conticini, 1984) reprend ces paramètres, dans le cadre de complexes fluviatiles (systèmes méandriformes), pour les soumettre à des analyses statistiques classiques (diagrammes de corrélation, diagrammes en étoile) afin d'essayer de différencier les types de séquences.

Ces analyses quantitatives sur l'amplitude, la vitesse d'intervalle, la continuité (évaluée à partir de la corrélation de la phase instantanée de deux traces contigues), la fréquence instantanée ainsi que l'étude qualitative de sections mettant en valeur un attribut sismique particulier (ex :

amplitude de l'enveloppe, phase et fréquence instantanées) permettent d'individualiser les faciès. En particulier, Conticini montre qu'il est possible de faire la distinction entre la séquence simple de "point-bar" et levée naturelle, la séquence plus complexe correspondant à un empilement de "point-bars", les dépôts liés aux ruptures des berges des chenaux et les argiles de plaine fluviale.

1.2.2. Utilisation plus quantitative d'attributs sismiques classiques : vitesses d'intervalle, amplitudes, fréquences instantanées

Mathieu et Rice (1969) s'intéressent aux vitesses d'intervalle des formations, évaluées sur des intervalles de 10 pieds. Leur étude porte sur une formation argilo-sableuse qui passe latéralement d'une dominante sableuse à une dominante argileuse. Dans un premier temps, ils travaillent sur des traces sismiques synthétiques calculées aux puits disponibles. L'étude de ces traces **non filtrées** par analyse factorielle met en évidence la séparation qui existe entre le groupe des traces représentant la zone sableuse et le groupe des traces correspondant à la zone argileuse.

Les traces synthétiques impulsionnelles (non filtrées) sont ensuite filtrées dans une bande passante de 15-50 Hz. Les variables choisies pour caractériser les traces ne sont plus les vitesses d'intervalle mais les amplitudes des 10 premiers échantillons (pas d'échantillonnage de 2 ms) à partir d'un marqueur de référence. Une analyse discriminante réalisée entre les deux groupes (traces "sableuses" - traces "argileuses") montre que leur séparation est significative (probabilité de mal classer un échantillon inférieure à 1 %).

L'étape suivante a consisté à classer 77 traces sismiques **réelles**, provenant d'un champ d'huile produisant à partir de cette unité sableuse, dans l'un des deux groupes définis à partir des traces synthétiques. On a pu séparer les traces réelles correspondant à la zone productrice de celles de la zone argileuse, ce qui confirme l'intérêt d'une analyse multivariable sur les amplitudes pour la discrimination sables/argiles.

Hagen (1982) a utilisé les sections "fréquence instantanée" pour mettre en évidence une forte variation de porosité dans une zone réservoir d'épaisseur voisine de 150 ms. Des puits de contrôle donnant une idée de l'extension de la zone poreuse ainsi que l'étude comparative des sections sommes classiques, des sections enveloppe et fréquence instantanée ont permis de retenir ce dernier paramètre (fréquence instantanée) pour caractériser les traces. C'est, en effet, sur les sections fréquence instantanée que le contraste entre la zone poreuse et la zone non poreuse apparaît le plus clairement.

Une analyse en composantes principales est alors réalisée puis une classification automatique des traces sismiques en utilisant comme variables les 3 premières composantes principales. Cette classification automatique en deux groupes est initialisée à partir de traces appartenant aux deux zones : poreuse et non poreuse. La probabilité de bonne classification a été également calculée.

La classification a mis très nettement en évidence l'anomalie de porosité ; d'autre part, la majorité des puits a été correctement affectée.

1.2.3. Etude quantitative d'attributs mesurés sur l'autocorrélation et son spectre

Khattri et al. (1976, 1979, 1983, 1984) ont réalisé de nombreux travaux sur l'utilisation de paramètres mesurés sur la fonction d'autocorrélation et son spectre pour caractériser les propriétés lithologiques des séries traversées. Ces travaux sont tous basés sur l'analyse de la réponse sismique synthétique de modèles sédimentaires.

Khattri et Gir (1976) ont étudié la signature sismique de modèles sédimentaires en prenant en compte les multiples. Les traces synthétiques, obtenues avec un signal Ricker dont la période moyenne est de 20 ms (soit une fréquence moyenne de 50 Hz), ont été étudiées aussi bien dans l'espace temps que dans l'espace des fréquences. L'objectif est de quantifier l'importance de différents paramètres (pourcentage d'argiles, épaisseurs) sur la réponse sismique, notamment sur les amplitudes et le spectre de la trace. Pour cela, 4 modèles ont été étudiés par les auteurs.

Le premier modèle correspond à des alternances argilo-sableuses. La longueur totale de la trace est constante (216 ms), mais plusieurs répartitions des lithologies constitutives ont été envisagées en jouant sur l'épaisseur des unités argileuses. L'analyse des divers cas met en évidence l'effet de résonance provoqué par la disposition répétitive des strates, à moins que la longueur de l'impulsion soit supérieure à 20 % du temps double du trajet dans les strates, ou que ce temps de trajet soit supérieur à la longueur du signal. Une étude des fréquences correspondant au pic principal d'amplitude suggère une corrélation entre cette fréquence et l'épaisseur maximale des couches (plus cette épaisseur est grande, plus la fréquence est basse).

Le second modèle ("argiles/charbon") représente une séquence constituée de fins lits d'argiles, de carbonates et de lignite. Les argiles peuvent passer progressivement à des silts et à des sables, mais les contrastes acoustiques étant faibles, les auteurs ont considéré une seule unité appelée "argiles". L'épaisseur totale restant constante (336 ms), l'épaisseur des niveaux ligniteux varie ainsi que le rapport charbon/argiles. Une caractéristique commune aux divers cas

étudiés est la forte atténuation de l'amplitude après les quatre premières couches de charbon. Ceci s'explique par le très fort coefficient de réflexion argiles-charbon qui provoque une réflexion d'une fraction importante de l'énergie incidente. Sur le spectre de la trace impulsionnelle, on observe une augmentation de l'énergie avec la fréquence alors que sur la trace filtrée, l'énergie relative aux hautes fréquences est atténuée conformément au spectre de fréquence du signal. On peut encore corrélérer la plus basse fréquence correspondant à un pic d'amplitude avec l'épaisseur maximale des couches du modèle.

Le troisième modèle, de transgression et de régression, est une modélisation des faciès littoraux. Le modèle régressif correspond à une séquence de progradation du littoral (argiles marines, sables littoraux, argiles de lagon, argiles et sables de la plaine fluviatile), tandis que le modèle transgressif correspond à la séquence de polarité opposée. Dans les deux cas, l'épaisseur des strates constitutives est la même. Les sismogrammes obtenus sont symétriques l'un de l'autre, la séquence transgressive présente un sommet bien défini par deux fortes impulsions sur la trace filtrée, alors que la séquence régressive montre, au contraire, une base nette.

La quatrième modèle correspond au cas du "sabot de transgression", épandage de sables remaniés lors d'une transgression et recouverts d'argiles marines. Ce modèle est constitué d'un substratum granitique, d'une couche de sables puis d'une couche d'argiles. Deux valeurs de la vitesse pour les sables et pour les argiles sont proposées. Les réponses impulsionnelles des deux cas sont dominées par la réflexion au toit du granite. Un des spectres présente un caractère oscillatoire très marqué à allure quasi-sinusoïdale tandis que l'autre ne montre que des oscillations de très faible amplitude.

En conclusion, les auteurs soulignent l'importance de la stratification et de la distribution des épaisseurs sur la réponse sismique du sous-sol, toutes choses étant égales par ailleurs, aussi bien dans le domaine temps que dans le domaine fréquence. Les études visant à déterminer le rapport sables/argiles, à partir des caractères des réflexions, doivent donc s'appuyer sur des analyses statistiques nombreuses.

Malgré tout, il est possible de faire la distinction entre des séries argilo-sableuses et des séries argilo-ligniteuses, ce qui est particulièrement intéressant pour l'exploration pétrolière. De même, il est possible de caractériser les sables de transgression qui peuvent constituer un piège stratigraphique intéressant.

Le spectre d'amplitude est prometteur quant à la nature du litage, puisque les plus basses fréquences correspondant à des pics sont liées à l'épaisseur maximale des couches dans la

section, et l'énergie de ce pic au nombre de couches possédant cette épaisseur. Les pics aux plus hautes fréquences pourraient indiquer la présence des couches d'épaisseur plus faible.

Nous avons là une première étape importante dans la détermination des caractéristiques lithologiques des séries à partir des caractères sismiques.

Dans un autre travail, Khattri, Sinvhal et Awasthi (1979) étudient les paramètres sismiques pouvant discriminer des modèles sédimentaires générés par la méthode de simulation de Monte-Carlo.

Les auteurs ont généré des séquences à dominante sablo-argileuse, ou argilo-ligniteuse dans lesquelles l'épaisseur de chaque strate ainsi que sa lithologie sont déterminées aléatoirement (méthode de Monte-Carlo). Les vitesses et les densités des sables, argiles, charbons, constantes pour une lithologie, ont été choisies à partir de données de champs. Les séries sédimentaires obtenues sont encadrées par deux couches homogènes d'épaisseur 200 m afin de les isoler.

110 modèles ont ainsi été fabriqués pour la séquence sables/argiles ainsi que pour la séquence argiles/charbons ; ils ont été classés en quatre groupes :

- groupe A : modèles avec du sable allant de 20 à 50 % de l'épaisseur totale,
- groupe B : modèles avec du sable allant de 50 à 80 % de l'épaisseur totale,
- groupe C : modèles où l'épaisseur de lignite varie de 20 à 50 % de l'épaisseur totale,
- groupe D : modèles où l'épaisseur de lignite varie de 50 à 80 %.

Les auteurs ont calculé les réponses impulsionnelles et leur fonction d'autocorrélation en tenant compte des multiples internes. Les paramètres sismiques étudiés sont les suivants :

- pour le contenu fréquentiel du film synthétique, on prend les rapports A_1/A_0 , A_2/A_1 , A_2/A_0 , rapports entre le pic d'amplitude de l'autocorrélation A_0 et les amplitudes des deux rebonds suivants A_1 , A_2 ;
- on prend la fréquence f_e qui sépare la zone de haute énergie de celle de basse énergie sur le spectre de l'autocorrélation ;
- sur le spectre de l'autocorrélation, on choisit la fréquence f_m correspondant au pic maximum d'énergie, et la fréquence f_p correspondant au premier pic significatif d'énergie.

Dans presque tous les cas, on observe que les réponses impulsionnelles présentent les plus fortes amplitudes à leur début et à la fin. La première impulsion représente la réflexion au sommet de la formation et la forte amplitude à la fin de l'enregistrement correspond à une combinaison de multiples internes. Le groupe A présente rarement une diminution de l'énergie avec le temps car le coefficient de réflexion sables/argiles est faible, à la différence du groupe C.

Les six variables précédemment définies sont soumises à des analyses statistiques classiques. Les histogrammes présentent des différences visuelles pour les paramètres f_e , f_p , f_m pour les quatre classes de modèles. En appliquant le test de Kolmogorov-Smirnov dans le but de savoir si les différences sont significatives avec un niveau de confiance à 95 %, on constate que :

- le paramètre f_e sépare les modèles A et B des modèles C et D, les modèles A des modèles B, mais ne permet pas la distinction entre les modèles C et D ;
- les paramètres f_m et f_p séparent seulement les groupes A et B des groupes C et D.

On applique le test sur les paramètres de l'autocorrélation A_2/A_0 , A_2/A_1 , A_1/A_0 : les résultats sont les mêmes pour les trois paramètres. On distingue les modèles sablo-argileux des modèles argilo-ligniteux mais, à l'intérieur de cette classification, on ne peut pas pousser la subdivision.

La fonction d'autocorrélation et son spectre d'amplitude fournissent donc des paramètres discriminants en terme de contrastes de lithologie, puisque, en moyenne, le schéma de stratification est le même pour les deux familles de modèles. Il devient plus difficile, à l'intérieur d'une famille, de séparer les types en fonction du pourcentage d'une lithologie.

L'étude a été conduite sur des traces synthétiques impulsionnelles. Dans le cas réel, suite à la réduction de la bande passante, la fonction d'autocorrélation semble être moins prometteuse : à moins de soustraire l'effet de convolution du signal, les paramètres précédents seront peu discriminants.

Khatti et Sinval (1983) ont également étudié la réponse sismique de modèles sédimentaires générés par des chaînes markoviennes de premier ordre.

L'étude réalisée porte sur une formation argilo-gréseuse, de 200 à 300 m d'épaisseur, comprise entre deux épaisses formations argileuses, et appartenant à un bassin sédimentaire tertiaire et quaternaire de l'Inde. Les auteurs ont distingué une zone X à dominante sableuse (> 56 % de sables, en épaisseur) et une zone Y à dominante argileuse (> 60 % d'argiles). La zone X

présente de grandes variations d'épaisseur et correspond à des faciès de plate-forme interne, littoraux et de lagon ; la zone Y correspond à des faciès de milieu marin ouvert.

Les matrices de transition caractérisant les zones X et Y ont été déterminées et on a généré des séquences sédimentaires synthétiques, lesquelles ont été incluses entre deux couches homogènes épaisses.

Avec des vitesses et des densités déduites de données de puits, les auteurs ont calculé 508 films synthétiques avec un signal de longueur 40 ms, de période moyenne 60 Hz (on aura donc une sismique d'assez haute résolution), en ajoutant 10 % de bruit aléatoire pour se rapprocher des conditions réelles. Les fonctions d'autocorrélation ont été calculées ainsi que leur spectre d'amplitude. Une analyse factorielle discriminante a été effectuée sur les paramètres suivants :

- A_1/A_0 , A_2/A_0 , A_3/A_0 ;
- les trois premiers temps de passage à zéro de l'autocorrélation, T_1 , T_2 , T_3 ;
- le temps correspondant au premier minimum de l'autocorrélation.

Sur le spectre d'amplitude de l'autocorrélation, on a choisi :

- la fréquence correspondant au pic d'énergie, f_M ;
- les quartiles de la distribution des amplitudes ;
- les quartiles de la distribution des amplitudes pondérées par les fréquences ;
- la fréquence f_m la plus basse à laquelle le logarithme du spectre d'amplitude de l'autocorrélation décroît à la moitié de sa valeur maximale.

Les résultats de l'analyse factorielle discriminante montrent que 70 % des sismogrammes sont correctement classés. Les variables les plus discriminantes sont A_1/A_0 (38,4 %), f_M (30,5 %), T_2 (24,5 %) et f_m (18,7 %). Un essai de classification de sismogrammes inconnus a prouvé que, dans 75 % des cas, ils étaient rattachés au bon modèle.

L'étape suivante a consisté à analyser des traces sismiques réelles dans les zones X et Y. L'analyse factorielle met en évidence une contribution très forte (95,3 %) de la variable f_m ; à titre de vérification, 90 % des sismogrammes réels ont été correctement classés.

1.3. Critique des variables utilisées

L'analyse faite par Conticini (1984) s'inscrit plutôt dans le cadre "analyse séquentielle". Les paramètres utilisés sont des attributs classiques : fréquence instantanée, continuité... qui ne

permettent pas réellement de caractériser **systématiquement** et **globalement** une portion de trace sismique. Le pouvoir discriminant de ces attributs n'est montré que de façon qualitative et ceci sur un cas très particulier.

Mathieu et Rice (1969) utilisent directement les amplitudes des 10 premiers échantillons de la trace sismique dans une analyse discriminante. Ces paramètres rendent bien compte de la morphologie de la trace mais non globalement et sont très sujets au bruit (qui peut entacher précisément la valeur d'un échantillon) ainsi qu'au problème de la taille de la fenêtre d'analyse (nous retrouverons ce problème ultérieurement).

Le paramètre retenu par Hagen (1982) est la fréquence instantanée, mais ceci nous paraît assez pernicieux. En effet, il dispose déjà d'une interprétation très détaillée de la zone en terme de porosité, ce qui lui a permis de constater que c'est la fréquence instantanée qui souligne le mieux les variations de ce paramètre. Il nous semble plus raisonnable de caractériser les traces par un **grand nombre** de variables et de remettre le choix **du ou des** paramètres pertinents à l'étape "traitement statistique".

Les travaux menés par l'équipe de Khatri (1976, 1979, 1983, 1984) soulignent dans divers environnements mais toujours sur des **traces synthétiques** l'intérêt de la fonction d'autocorrélation et du spectre de puissance (carré du spectre d'amplitude) pour la caractérisation lithologique des faciès sismiques.

Une synthèse de ces études met en évidence trois familles possibles de paramètres :

- des paramètres sur les traces elle-mêmes (**amplitudes** ou **temps correspondants**) qui sont liés à l'énergie et à sa répartition dans l'intervalle considéré ;
- des paramètres sur l' **autocorrélation** de la trace laquelle traduit dans une certaine mesure la cyclicité de la sédimentation ;
- des paramètres sur le **spectre de fréquence** des traces exprimant les épaisseurs moyennes des strates constitutives de l'intervalle, et surtout leur distribution.

On peut remarquer que la plupart des auteurs n'ont pris en compte qu'une seule famille de paramètres, et espérer, au vu des résultats déjà intéressants obtenus, améliorer l'efficacité de la caractérisation des faciès en considérant des paramètres appartenant aux trois groupes précédents.

Il faut également remarquer que ces paramètres sont redondants. De toute façon, la morphologie de la trace sismique, analysée dans le domaine après sommation, revient à l'analyse de la

représentation temporelle des données sismiques. L'analyse de la trace dans le domaine fréquentiel est possible aussi mais nous savons que, sous certaines hypothèses, la représentation fréquentielle est strictement équivalente à la représentation temporelle.

Nous allons maintenant étudier, dans un cadre synthétique théorique, le pouvoir discriminant, en terme d'environnement géologique, de la fonction d'autocorrélation et de son spectre.

2. CARACTERISATION D'UN ENVIRONNEMENT GEOLOGIQUE PAR LA FONCTION D'AUTOCORRELATION ET SON SPECTRE

2.1. Introduction

L'autocorrélation de la suite des coefficients de réflexion et son spectre sont les deux fonctions à partir desquelles Khattri et al. (1976, 1979, 1983, 1984) ont défini des paramètres sismiques discriminant les faciès géologiques.

Velzeboer (1981) propose des éléments de calcul de la fonction d'autocorrélation théorique d'une séquence sédimentaire binaire (deux lithologies constitutives) générée par un processus markovien de premier ordre. Nous avons repris et développé ces éléments dans le cadre de plus de deux lithologies. Nous présenterons une application aux faciès du delta du Niger.

L'objectif de cette étude est de corroborer la signification géologique et le pouvoir discriminant de la fonction d'autocorrélation en terme de propriétés géologiques.

2.2. Présentation du modèle markovien

Le modèle markovien a été beaucoup utilisé pour l'étude quantitative des séries sédimentaires car il permet de générer des séquences lithologiques synthétiques susceptibles de reproduire les caractéristiques des séquences réelles du point de vue **enchaînement** des faciès.

Une chaîne de Markov est une succession d'états s'excluant mutuellement, pour laquelle l'état en un point dépend partiellement, dans un sens probabiliste, des états précédents. Le modèle markovien rend compte de la probabilité des transitions d'un état vers un autre mais perd la notion de position des observations dans la succession.

A partir d'un nombre n fini de lithologies observées dans une série sédimentaire, un pas d'échantillonnage en profondeur ayant été sélectionné, il est possible de construire une matrice carrée $n \times n$ représentant les fréquences expérimentales des transitions d'une lithologie vers une autre.

Soit $T = (\alpha_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ cette matrice,

α_{ij} est le nombre de fois où la transition de la lithologie i vers la lithologie j a été observée, divisé par le nombre total de transitions de la lithologie i vers n'importe quelle autre lithologie.

La fréquence expérimentale α_{ij} est en fait une estimation de la probabilité conditionnelle $P(j/i)$ ou $P(i \rightarrow j)$.

On peut remarquer que : $\sum_{j=1}^n \alpha_{ij} = 1$ et que la matrice T n'est en général pas symétrique.

La proportion relative des n lithologies dans la séquence s'obtient en divisant le nombre de transitions de la lithologie i vers toute autre lithologie par le nombre total de transitions observées. Cette proportion est une estimation de la probabilité de la lithologie i :

$$P(i) = \frac{\sum_{j=1}^n N_{ij}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}}$$

Si les états sont indépendants, toutes les lignes de la matrice de transition sont identiques au vecteur $(P(1), P(2), \dots, P(n))$ des probabilités des différentes lithologies.

$$\forall_i \quad 1 \leq i \leq n$$

$$\forall_j \quad 1 \leq j \leq n \quad \alpha_{ij} = P(j).$$

En effet :

$$\begin{aligned} \alpha_{ij} &= P(j/i) = \frac{P(i \text{ et } j)}{P(i)} \\ &= \frac{P(i) P(j)}{P(i)} = P(j) . \end{aligned}$$

On peut appliquer un test du χ^2 pour valider l'hypothèse selon laquelle l'apparition des lithologies est indépendante des états précédents.

Dans le cas contraire, le "degré de dépendance" entre les lithologies successives définit l'ordre de la chaîne de Markov. Ainsi, une chaîne de Markov de **premier ordre** rend compte d'une dépendance statistique entre les points et leurs prédécesseurs immédiats (entre l'état n et l'état $(n - 1)$). Pour une chaîne de Markov de **second ordre**, il y a encore une dépendance significative entre des états séparés de deux intervalles.

L'ordre du modèle markovien à utiliser peut se déterminer expérimentalement en élevant à des puissances de plus en plus élevées la matrice de transition T . T^k donne les probabilités de transition entre deux états distants de k intervalles. A partir d'une certaine valeur k_1 de k , la matrice T^k se stabilise autour des valeurs $P(1), P(2) \dots P(n)$ correspondant aux probabilités d'apparition des différentes lithologies sous l'hypothèse de leur indépendance.

La façon dont ont été déterminés les termes (α_{ij}) de la matrice T présuppose d'avoir défini un pas d'échantillonnage de la série sédimentaire étudiée. Ce pas d'échantillonnage influence beaucoup les valeurs α_{ij} . En particulier, la probabilité α_{ij} de transition d'une lithologie i vers elle-même est très liée aux épaisseurs des couches de lithologie i . Avec un pas d'échantillonnage trop petit, la matrice de transition ne reflétera que les épaisseurs des unités les plus massives. Si le pas d'échantillonnage est trop grand, les fines unités ne seront pas reconnues.

C'est pourquoi, il existe une seconde façon de définir les termes α_{ij} en ne considérant que les transitions entre deux états **différents** (chaîne de Markov à temps continu). La définition d'un pas d'échantillonnage n'est plus nécessaire puisque on ne compte que les transitions d'une lithologie vers une autre. Dans ce cas là, les termes diagonaux de la matrice, α_{ii} , sont nuls :

$$\forall i, \quad 1 \leq i \leq n, \quad \alpha_{ii} = 0.$$

Les autres termes de la matrice se calculent de façon identique et la propriété :

$$\forall i, \quad 1 \leq i \leq n \quad \sum_{j=1}^n \alpha_{ij} = 1 \text{ est encore vérifiée. Par contre, on ne dispose pas d'estimation}$$

directe des probabilités $P(i)$ de chacune des lithologies.

C'est cette façon de définir la matrice de transition que nous avons utilisée pour décrire la séquence sédimentaire pour laquelle nous calculons, par la suite, l'autocorrélation des coefficients de réflexion.

2.3. Autocorrélation d'une séquence sédimentaire markovienne de premier ordre

2.3.1. Modèle utilisé

La **séquence sédimentaire** est décrite comme une chaîne de Markov de premier ordre, de matrice stochastique T.

Nous avons choisi une loi de Poisson pour la **distribution des épaisseurs**. Ce choix suppose qu'à partir du moment où une couche de lithologie donnée commence à se déposer, son dépôt peut s'arrêter à tout moment avec une probabilité constante (par unité de temps) et indépendante du temps. Le processus de dépôt des couches est alors de type poissonien.

Nous supposons que les **vitesse**s et les **densités** des différentes lithologies suivent des lois normales. Leur distribution est donc entièrement déterminée par la connaissance des moyennes et écarts-types de ces paramètres pour chaque lithologie.

Dans ces conditions, on peut montrer, au second ordre près, que les coefficients de réflexion suivent également une loi normale (Velzeboer, 1981).

2.3.2. Evaluation du temps de transit dans une couche de lithologie *i*

Dans le cadre des hypothèses précédemment formulées, la loi de probabilité du temps de transit (temps de séjour ou encore de transfert) de l'onde sismique dans une lithologie donnée *i* est

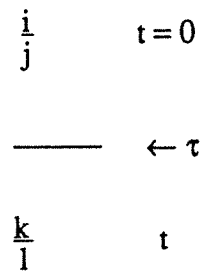
une loi exponentielle de paramètre $\frac{v_i}{D_i}$ dont la densité s'écrit :

$$\pi_i(t) = \frac{v_i}{D_i} e^{-t \frac{v_i}{D_i}}$$

où v_i est la vitesse moyenne de la lithologie *i* et D_i l'épaisseur moyenne.

La forme exponentielle de cette loi provient de l'utilisation d'une loi de Poisson pour rendre compte de la distribution des épaisseurs et du fait que l'on a pris une loi de Gauss pour la distribution des vitesses à l'intérieur d'une lithologie.

2.3.3. Evaluation de la probabilité d'observer au temps t une transition $k \rightarrow l$.



Notons $\pi_{k,l}^{i,j}(t)$ la probabilité conditionnelle d'observer entre t et $t + dt$ une transition k,l sachant qu'une transition i,j a été observée au temps $t = 0$.

Notons $\pi_{j,k}(t)$ la probabilité conditionnelle de sortie de la séquence k au temps t ($\in [t ; t + dt]$) sachant que l'on rentre dans j au temps $t = 0$.

Avec ces notations, on a évidemment :

$$\pi_{k,l}^{i,j}(t) = \pi_{jk}(t) a_{kl}$$

avec $l \neq j \neq k$.

De plus, cette probabilité de transition vaut :

$$\pi_{jk}(t) = \sum_m \int_0^t \pi_{jm}(\tau) \alpha_{mk} \pi_k(t - \tau) \delta\tau + \delta_{jk} \pi_j(t)$$

$$\begin{aligned} \text{avec } \delta_{jk} &= 0 && \text{si } k \neq j \\ &= 1 && \text{sinon.} \end{aligned}$$

Introduisons les matrices de transition suivantes :

$$\begin{aligned} \Pi(t) &= [\pi_{jk}(t)] \\ \pi(t) &= \begin{bmatrix} \pi_1(t) & & 0 \\ & \pi_i(t) & \\ 0 & & \pi_n(t) \end{bmatrix} \end{aligned}$$

Le produit de convolution ci-dessus s'écrit sous forme matricielle :

$$\Pi(t) = (\Pi T) * \pi + \pi$$

où l'inconnue est la matrice Π , T étant la matrice stochastique décrivant la séquence sédimentaire.

Ce produit de convolution s'exprime simplement à condition de considérer sa transformée de Laplace.

On calcule tout d'abord la transformée de Laplace de la fonction de transit $\pi_i(t)$ qui représente la probabilité que le temps de séjour dans la couche i soit de $[t ; t + dt[$:

$$\tilde{\pi}_i(s) = \int_0^{\infty} \pi_i(t) e^{-st} dt$$

où s est un nombre complexe.

$$\tilde{\pi}_i(s) = \int_0^{\infty} \pi_i(t) \left[1 - st + \frac{s^2 t^2}{2!} + \dots (-1)^n \frac{s^n t^n}{n!} + \dots \right] dt .$$

Or $\int_0^{\infty} \pi_i(t) t^j dt$ est le moment d'ordre j , m_{ji} de la densité de probabilité $\pi_i(t)$,

donc
$$\tilde{\pi}_i(s) = \sum_{j=0}^{\infty} \left(\frac{(-s)^j}{j!} \right) m_{ji} .$$

Soit $\tilde{\pi}(s)$ la matrice complexe n,n suivante :

$$\tilde{\pi}(s) = \begin{pmatrix} \tilde{\pi}_1(s) & & 0 \\ & \tilde{\pi}_i(s) & \\ 0 & & \tilde{\pi}_n(s) \end{pmatrix} .$$

La transformée de Laplace de $\Pi(t)$ s'exprime alors par :

$$\tilde{\Pi}(s) = \tilde{\Pi}(s) T \tilde{\pi}(s) + \tilde{\pi}(s)$$

dont la solution est :

$$\tilde{\Pi}(s) = (\tilde{\pi}(s)^{-1} - T)^{-1}, \text{ si } \tilde{\pi}(s) \text{ est inversible.}$$

Il s'ensuit la relation cherchée pour $\pi_{k,l}^{i,j}(t)$.

En effet, la probabilité d'observer un interface $k \rightarrow l$ au temps t , sachant qu'au temps $t = 0$, on a observé un interface $i \rightarrow j$, devient, dans le domaine complexe :

$$\tilde{\pi}_{k,l}^{i,j}(s) = \alpha_{k,l} \left[\tilde{\pi}(s)^{-1} - T \right]_{j,k}^{-1}$$

où les indices j,k désignent l'élément de la j ème ligne et k ème colonne de la matrice complexe $(\tilde{\pi}(s)^{-1} - T)^{-1}$.

2.3.4. Evaluation de la fonction d'autocorrélation des coefficients de réflexion

On suppose que la série étudiée est infinie. D'autre part, les coefficients de réflexion sont des fonctions du temps de transit.

L'expression de l'autocorrélation $R(\tau)$ s'obtient en calculant l'espérance mathématique du produit $r(t) \cdot r(t + \tau)$, où r est le coefficient de réflexion, à condition de supposer que $r(t)$ est stationnaire :

$$R(\tau) = E \{ r(t) \cdot r(t + \tau) \}.$$

Les variables aléatoires $r(t)$ et $r(t + \tau)$ sont indépendantes, on peut donc écrire :

$$R(\tau) = E \{ r(t) \} \cdot E \{ r(t + \tau) \}.$$

Là encore, on considère la transformée de Laplace de cette expression.

L'expression de l'autocorrélation dans le domaine complexe est la suivante (Velzeboer, 1981).

$$\tilde{R}(s) = \sum_{ijkl} \left[(T^* + U)^{-1} \cdot E \right]_i \cdot \alpha_{ij} \alpha_{kl} \left\{ \tilde{\pi}(s)^{-1} - T \right\}_{jk}^{-1} r_{ij} r_{kl}$$

où T^* est la matrice transposée de T ,

$$U \text{ est la matrice } \begin{pmatrix} 0 & 1 & \dots\dots 1 \\ 1 & 0 & \dots\dots 1 \\ 1 & 1 & \dots\dots 0 \end{pmatrix},$$

$$E = \begin{pmatrix} 1 & \dots\dots 1 \\ 1 & \dots\dots 1 \end{pmatrix},$$

r_{ij} , r_{kl} sont les moyennes des coefficients de réflexion des interfaces $i \rightarrow j$ et $k \rightarrow l$.

L'expression $\tilde{R}(s)$ est stable sauf pour $s = 0$.

Dans le domaine temporel, l'autocorrélation présente une discontinuité en $t = 0$.

A partir de cette expression, il est possible de calculer l'expression réelle de l'autocorrélation et son spectre d'amplitude dans le cas de **deux lithologies** (Velzeboer, 1981). Nous avons appliqué cette étude à des séquences binaires (argiles-sables) modélisées avec les données du delta du Niger.

En reprenant l'expression générale complexe de l'autocorrélation, que nous avons inversée **numériquement** dans le cas de **plus de deux lithologies**, nous avons calculé le spectre de puissance théorique correspondant en considérant **quatre lithologies**, ce qui permet de mieux rendre compte de la réalité géologique.

2.3.5. Cas de deux lithologies

2.3.5.1. Calcul de l'autocorrélation et de son spectre

Dans le cas de deux lithologies, la matrice de transition se simplifie :

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = T^*$$

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = T$$

$$E = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$(T^* + U)^{-1} \cdot E = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = V$$

$$\tilde{R}(s) = \sum_{ijkl} V(i) \alpha_{ij} \alpha_{kl} [\tilde{\pi}(s)^{-1} - T]_{jk}^{-1} r_{ij} r_{kl}$$

pour $\begin{cases} j = i & \alpha_{ij} = \alpha_{kl} = 0 \\ l = k \end{cases}$; donc les seuls termes qui

interviennent dans $\tilde{R}(s)$ sont ceux pour lesquels $i \neq j$ et $k \neq l$.

Tous calculs faits :

$$\tilde{R}(s) = \frac{1}{2} r_{12}^2 \left\{ \begin{array}{l} (\tilde{\pi}(s)^{-1} - T)_{21}^{-1} + (\tilde{\pi}(s)^{-1} - T)_{12}^{-1} \\ - (\tilde{\pi}(s)^{-1} - T)_{22}^{-1} - (\tilde{\pi}(s)^{-1} - T)_{11}^{-1} \end{array} \right\}$$

Calcul de la matrice $\{\tilde{\pi}(s)^{-1} - T\}^{-1}$

$$\tilde{\pi}(s) = \begin{pmatrix} \frac{1}{1 + s \frac{D1}{v1}} & 0 \\ 0 & \frac{1}{1 + s \frac{D2}{v2}} \end{pmatrix}$$

$$\tilde{\pi}(s)^{-1} = \begin{pmatrix} 1 + s \frac{D1}{v1} & 0 \\ 0 & 1 + s \frac{D2}{v2} \end{pmatrix}$$

Soit $T_i = \frac{D_i}{v_i}$, T_i est un temps de transit moyen pour la lithologie i .

$$\tilde{\pi}(s)^{-1} = \begin{pmatrix} 1 + s T_1 & 0 \\ 0 & 1 + s T_2 \end{pmatrix}$$

$$\tilde{\pi}(s)^{-1} - T = \begin{pmatrix} 1 + s T_1 & -1 \\ -1 & 1 + s T_2 \end{pmatrix}$$

Soit Det le déterminant de la matrice $\{\tilde{\pi}(s)^{-1} - T\}$:

$$\text{Det} = sT_1 + sT_2 + s^2T_1 T_2 .$$

A priori, la matrice $\{\tilde{\pi}(s)^{-1} - T\}$ est toujours inversible sauf pour $s = 0$. Il y a une instabilité de $\tilde{R}(s)$ pour $s = 0$; l'inversion de $\tilde{R}(s)$ donnera l'expression $R(t)$ sauf au temps $t = 0$.

Calcul de $R(t)$

$$\tilde{R}(s) = \frac{1}{2} r_{12}^2 \left(\left(\frac{2}{\text{Det}} \right) - \left(\frac{1 + sT_2}{\text{Det}} \right) - \left(\frac{1 + sT_1}{\text{Det}} \right) \right)$$

$$\tilde{R}(s) = -\frac{1}{2} r_{12}^2 \left(\frac{T_1 + T_2}{T_1 + T_2 + sT_1 T_2} \right) .$$

Soit :

$$\gamma = \frac{1}{T_1} + \frac{1}{T_2} .$$

En introduisant T^* , la moyenne harmonique des temps de transit T_1 et T_2

($\frac{1}{T^*} = \frac{1}{T_1} + \frac{1}{T_2}$), il apparaît que γ est l'inverse de la moyenne harmonique des temps de transit moyens des deux lithologies : $\gamma = \frac{1}{T^*}$.

$\tilde{R}(s)$ s'exprime simplement à l'aide du paramètre γ .

$$\tilde{R}(s) = -\frac{1}{2} r_{12}^2 \left(\frac{\gamma}{\gamma + s} \right) .$$

Donc, pour $t \neq 0$, $R(t) = -\frac{1}{2} r_{12}^2 \gamma e^{-\gamma t}$.

Calcul de la valeur de l'autocorrélation au temps $t = 0$

$R(0) = r_{12}^2 + \sigma_{12}^2$ où σ_{12} est l'écart-type du coefficient de réflexion r_{12} ,

d'où :

$$R(t) = (\sigma_{12}^2 + r_{12}^2) \delta(t) - \frac{1}{2} r_{12}^2 \gamma e^{-\gamma |t|}.$$

Calcul du spectre de puissance

La transformée de Fourier de l'expression ci-dessus est une fonction réelle, donc c'est directement le spectre d'amplitude de l'autocorrélation.

$$P(f) = \sigma_{12}^2 + r_{12}^2 - \frac{r_{12}^2 \gamma^2}{\gamma^2 + 4\pi^2 f^2}$$

$$P(f) = r_{12}^2 \left(\frac{\sigma_{12}^2}{r_{12}^2} + \frac{1}{\left(\frac{\gamma}{2\pi f}\right)^2 + 1} \right).$$

2.3.5.2. Etude des courbes $R(t)$ et $P(f)$

- $R(t)$ est une fonction paire et discontinue en $t = 0$.

$$R(0) = (\sigma_{12}^2 + r_{12}^2).$$

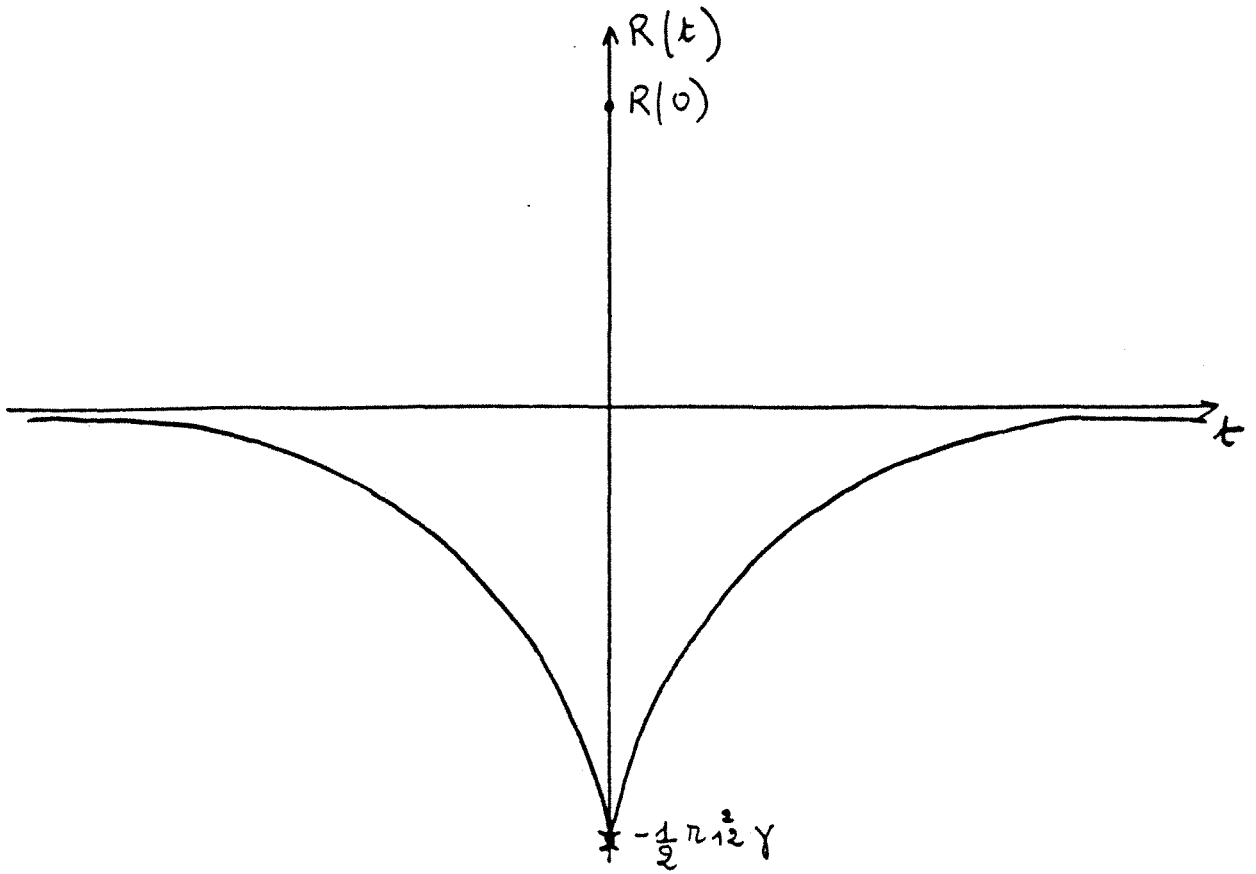
$$\lim_{t \rightarrow 0} R(t) = -\frac{1}{2} r_{12}^2 \gamma.$$

$$t \rightarrow 0$$

Pour tout $t \neq 0$, $R(t)$ est toujours négative et strictement croissante.

$$\lim_{t \rightarrow \pm \infty} R(t) = 0^-$$

$$t \rightarrow \pm \infty$$

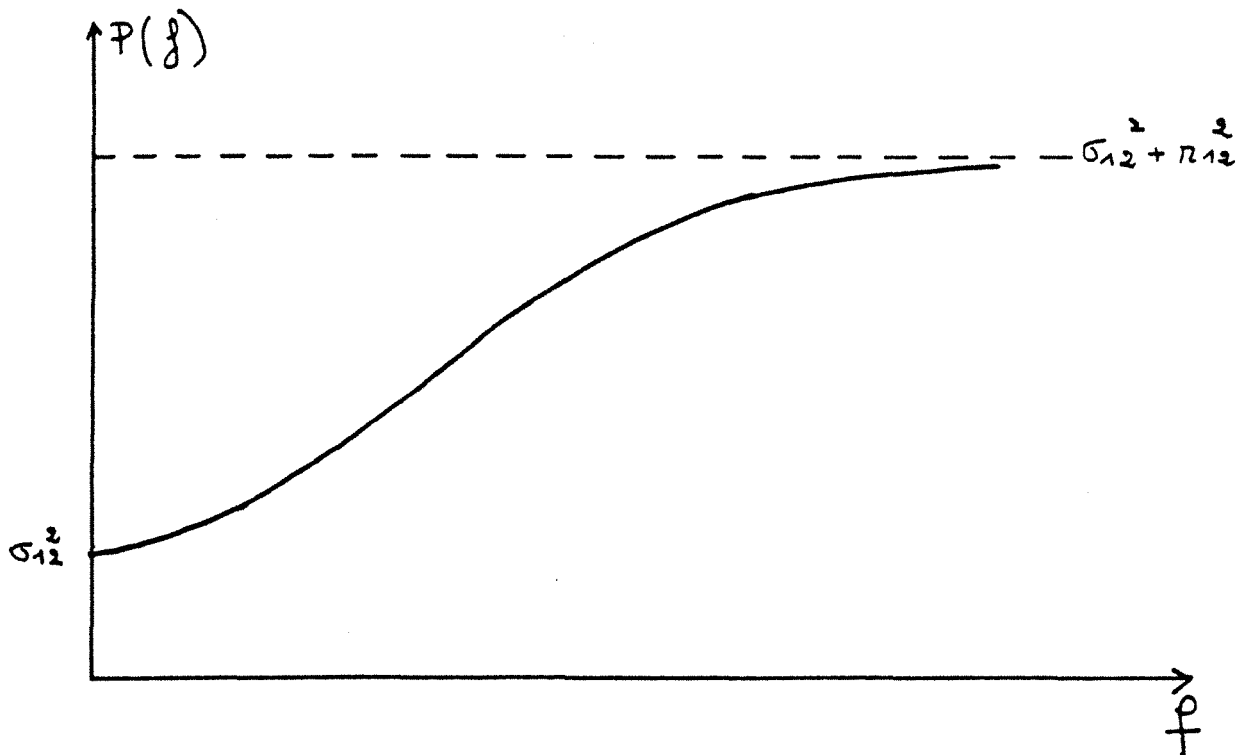


• $P(f)$ est une fonction croissante et bornée ; elle est continue en tout point.

$$\lim_{f \rightarrow 0} P(f) = \sigma_{12}^2 = P(0) ; \lim_{f \rightarrow +\infty} P(f) = \sigma_{12}^2 + \tau_{12}^2 .$$

$f \rightarrow 0$

$f \rightarrow +\infty$



2.3.5.3. Influence des divers paramètres

- Influence de $\left(\frac{\sigma_{12}}{r_{12}}\right)$

Lorsque la dispersion du coefficient de réflexion augmente, la courbe $P(f)$ s'aplatit, c'est-à-dire que le spectre devient plus blanc.

A la limite, si $\sigma_{12} \gg \gg r_{12}$, $P(f) = \text{ctante} = \sigma_{12}^2$.

Le spectre est constant, ce qui est prévisible puisque l'autocorrélation tend vers un Dirac d'amplitude σ_{12}^2 .

- Influence du paramètre γ

Ce paramètre est déterminant puisqu'il gouverne la relaxation de l'exponentielle qui intervient dans l'expression de l'autocorrélation.

Lorsque $\gamma \rightarrow +\infty$ $\gamma e^{-\gamma t} \rightarrow 0$

$R(t) \rightarrow (\sigma_{12}^2 + r_{12}^2) \delta(0)$,

d'où $P(f) \rightarrow \text{ctante}$.

Lorsque le paramètre γ croît, le spectre se rapproche d'un spectre constant, l'autocorrélation devient diracquienne. Ce résultat est aussi prévisible puisque γ est inversement proportionnel au temps de transfert ; à des couches minces, correspond un paramètre γ important. Comme γ résulte d'une moyenne harmonique ($\gamma = \frac{1}{T_1} + \frac{1}{T_2}$), il suffit même qu'une des lithologies soit, en moyenne, distribuée

suivant des épaisseurs faibles pour que γ soit très grand devant $2\pi f$ et que le spectre devienne constant.

En supposant que la gamme 0 – 100 Hz soit accessible à l'observateur, quel est le domaine de variation de γ dans lequel le spectre $P(f)$ n'est pas constant ?

$$\gamma > 1000, f = 10, \frac{\gamma}{2\pi f} > \frac{1000}{2\pi \times 10} \approx 16$$

$$\left(\frac{\gamma}{2\pi f}\right)^2 \approx 250$$

$$P(f) \approx r_{12}^2 \left(\left(\frac{\sigma_{12}}{r_{12}}\right)^2 + \frac{1}{1 + \left(\frac{\gamma}{2\pi f}\right)^2} \right)$$

$$P(f) \approx r_{12}^2 \left(\left(\frac{\sigma_{12}}{r_{12}}\right)^2 + \left(\frac{2\pi f}{\gamma}\right)^2 \right)$$

$$P(f) \approx r_{12}^2 \left(\left(\frac{\sigma_{12}}{r_{12}}\right)^2 + 4.10^{-3} \right) \approx (10^{-2} + 4.10^{-3}) .$$

$P(f)$ est encore influencé par le terme en $\frac{1}{\left(\frac{\gamma}{2\pi f}\right)^2}$ mais c'est la limite.

On peut raisonnablement étudier la courbe $P(f)$ paramétrée par γ à condition de rester dans la plage 0 – 1000 Hz pour les variations de γ ; au-delà, on peut affirmer que, quelle que soit la valeur de γ , le spectre de puissance théorique est celui d'un Dirac.

2.3.5.4. Réalisation d'abaques

Nous avons tracé les courbes $20 \log \frac{P(f)}{P(100)}$ pour γ variant de 10 à 300, et ceci dans les

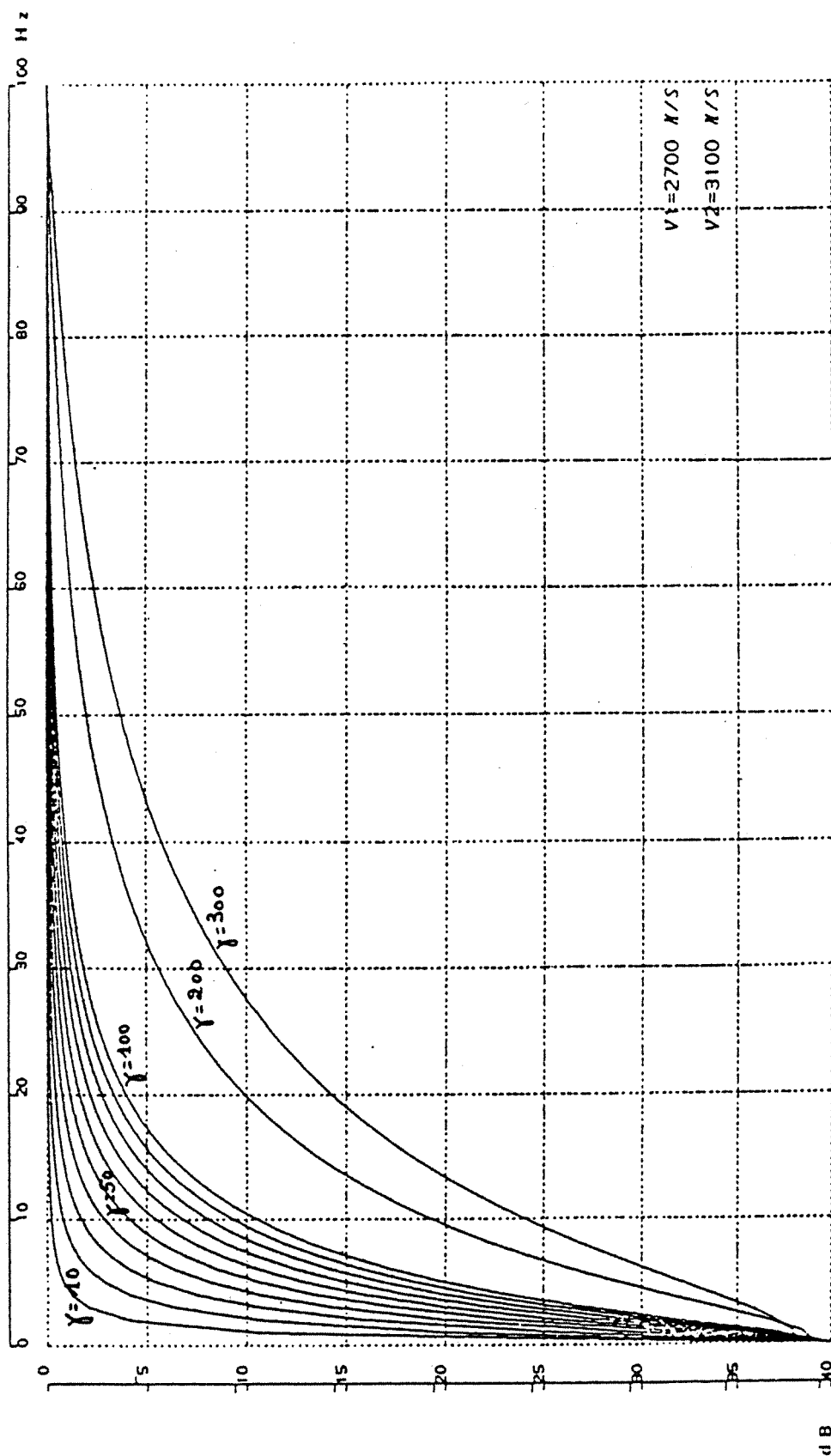
hypothèses $\left(\frac{\sigma_{12}}{r_{12}}\right) = 0,1$; $\left(\frac{\sigma_{12}}{r_{12}}\right) = 0,2$; $\left(\frac{\sigma_{12}}{r_{12}}\right) = 0,3$; $\left(\frac{\sigma_{12}}{r_{12}}\right) = 0,6$ (cf. Figures 28 à 31).

On retrouve bien le fait que lorsque γ augmente, le spectre est de moins en moins contrasté.

L'augmentation de la dispersion du coefficient de réflexion n'affecte l'autocorrélation qu'au point de discontinuité $t = 0$, mais on constate que le spectre de fréquence est alors beaucoup plus "blanc" même pour des valeurs petites de γ .

CAS DE DEUX LITHOLOGIES
 SPECTRE DE PUISSANCE NORME A 100 HZ. EN DB

SIGMA/REFLEX=0.1
 COURBES PARAMETREES PAR GAMMA



CAS DE DEUX LITHOLOGIES
SPECTRE DE PUISSANCE MORNE A 100 HZ. EN DB

SIGMA/REFLEX=0.2
COURBES PARAMETREES PAR GAMMA

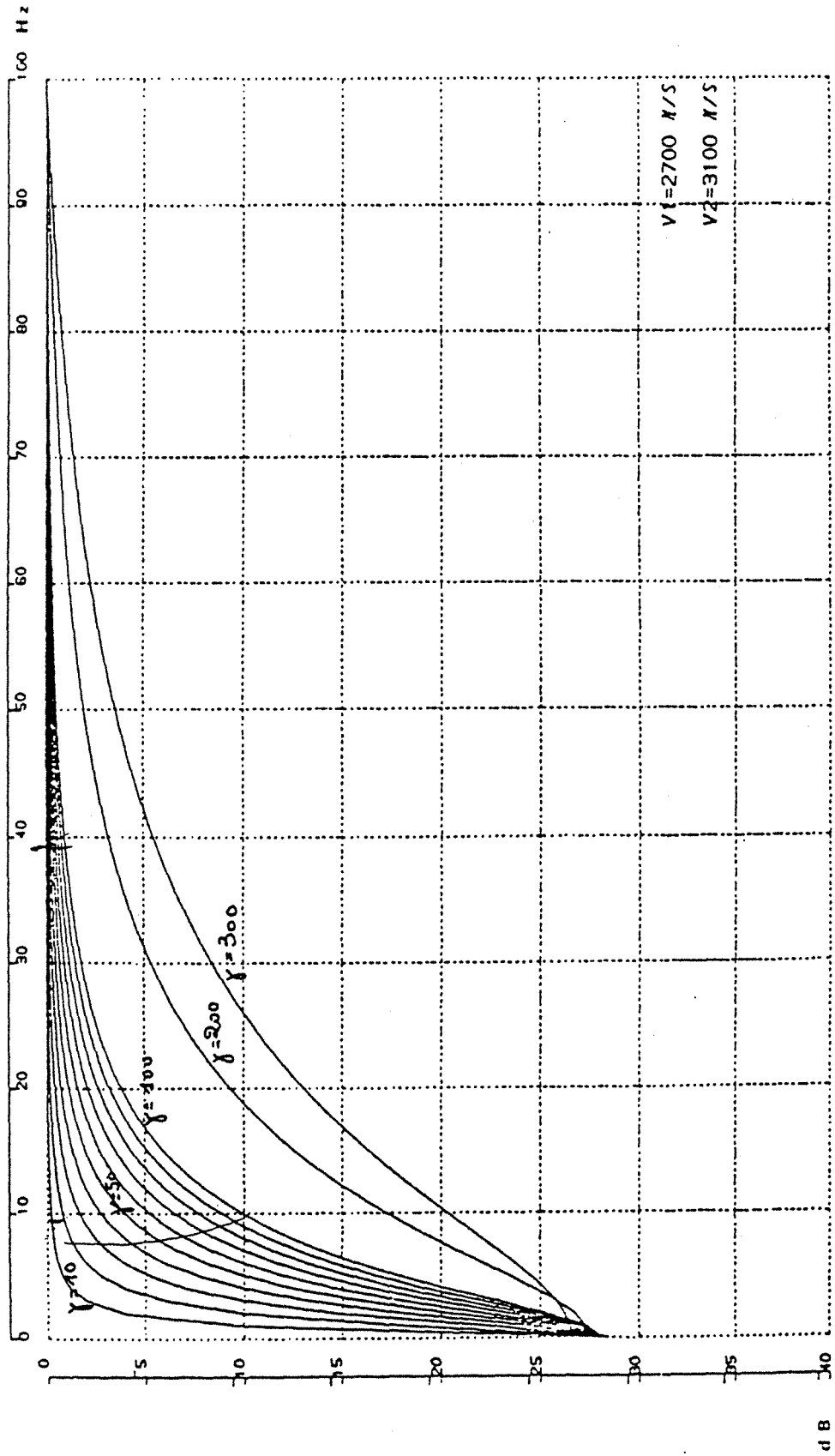
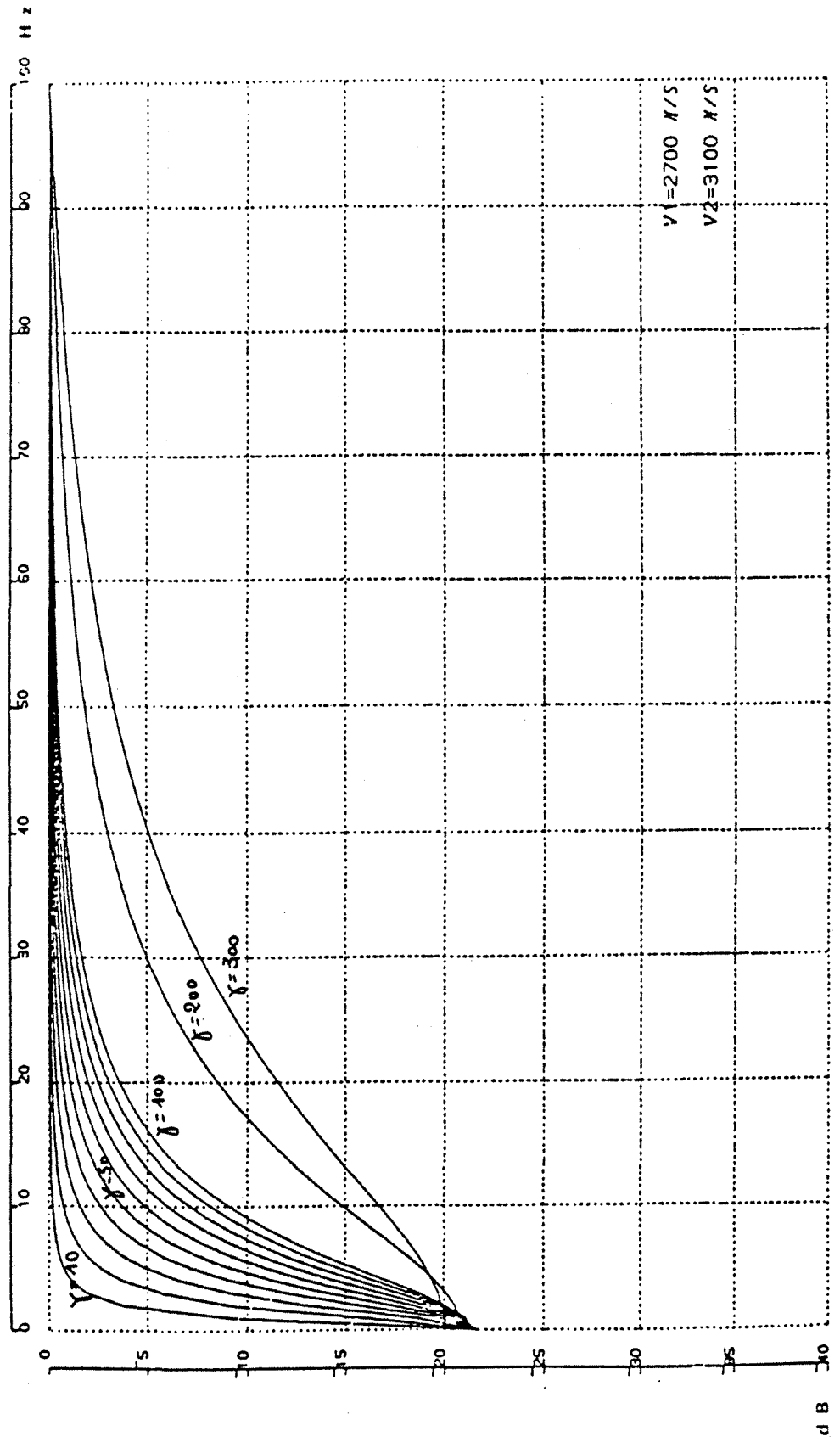


Fig. 30

CAS DE DEUX LITHOLOGIES
SPECTRE DE PUISSANCE MORME A 100 HZ. EN DB

SIGMA/REFLEX=0.3
COURBES PARAMETREES PAR GAMMA

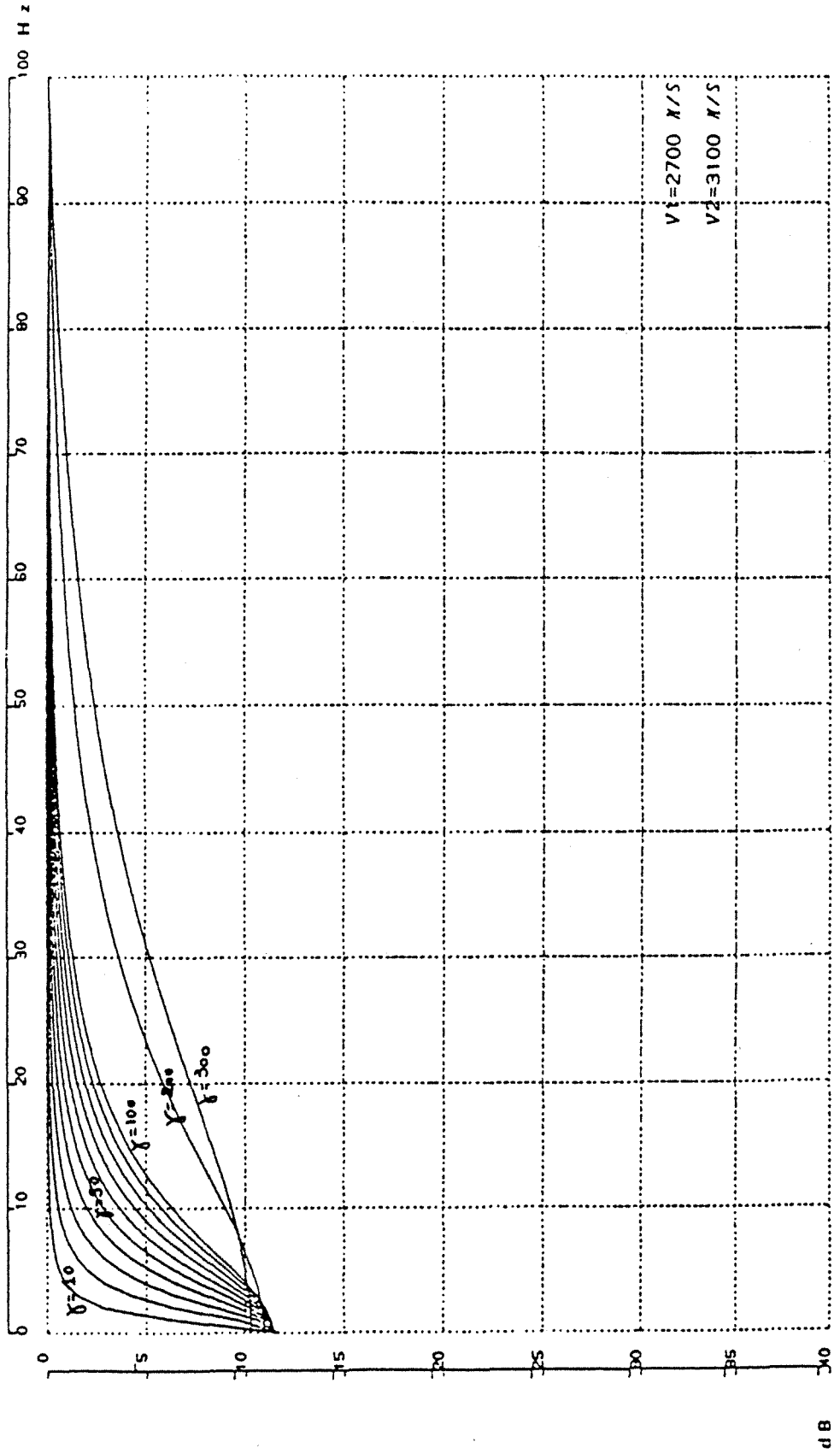


dB

Fig. 31

CAS DE DEUX LITHOLOGIES
SPECTRE DE PUISSANCE MORNE A 100 HZ. EN DB

SIGMA/REFLEX=0.6
COURBES PARAMETREES PAR GAMMA



2.3.6. Cas de plus de deux lithologies

2.3.6.1. Evaluation de l'autocorrélation et de son spectre d'amplitude

Dans le cas de plus de deux lithologies, il n'est pas possible de calculer directement la transformée de Laplace inverse de l'autocorrélation $\tilde{R}(s)$.

Nous avons donc développé son expression complexe avant de calculer par un algorithme numérique une approximation de sa transformée inverse.

Connaissant alors l'expression temporelle de l'autocorrélation, nous avons calculé son spectre d'amplitude par l'algorithme de Cooley-Tukey.

Rappelons l'expression générale de la matrice $\tilde{\pi}(s)$:

$$\tilde{\pi}(s) = \begin{pmatrix} \tilde{\pi}_1(s) & & 0 \\ 0 & \tilde{\pi}_i(s) & \\ 0 & 0 & \tilde{\pi}_n(s) \end{pmatrix}$$

avec $\tilde{\pi}_i(s)$ transformée de Laplace de la distribution $\pi_i(t)$.

Puisque nous avons fait l'hypothèse que la loi de l'épaisseur de chaque strate est une loi

exponentielle, $\pi_i(t) = \frac{v_i}{D_i} e^{-\frac{tv_i}{D_i}}$, il s'ensuit que sa transformée de Laplace vaut :

$$\tilde{\pi}_i(s) = \frac{1}{1 + \frac{sD_i}{v_i}} .$$

$$\tilde{\pi}(s) = \begin{pmatrix} \frac{1}{1 + \frac{sD_1}{v_1}} & 0 \\ 0 & \frac{1}{1 + \frac{sD_n}{v_n}} \end{pmatrix}$$

$$\tilde{\pi}(s)^{-1} = \begin{pmatrix} 1 + \frac{sD_1}{v_1} & 0 \\ 0 & 1 + \frac{sD_n}{v_n} \end{pmatrix} = \begin{pmatrix} 1 + sT_1 & 0 \\ 0 & 1 + sT_n \end{pmatrix}$$

où T_i est le temps de transit moyen dans la lithologie i .

Pour obtenir une expression explicite de $\tilde{R}(s)$, il suffit d'inverser la matrice $\{\tilde{\pi}(s)^{-1} - T\}$:

$$\begin{pmatrix} 1 + \frac{sD_1}{v_1} - \alpha_{12} & -\alpha_{1n} \\ -\alpha_{n1} & 1 + \frac{sD_n}{v_n} \end{pmatrix}$$

et de calculer $\tilde{R}(s)$ pour diverses valeurs complexes de s avec $s \neq 0$.

Les valeurs de s sont échantillonnées de telle sorte que $R(t)$ soit évaluée au pas de 2 ms entre 2 et 512 ms, grâce à un algorithme qui fournit une approximation numérique de la transformée de Laplace inverse.

L'autocorrélation présente une discontinuité en $t = 0$.

2.3.6.2. Calcul de la valeur $R(0)$

Soit $R = [r_{ij}]$ la matrice des moyennes des coefficients de réflexion et $S = [\sigma_{ij}]$ la matrice des écarts-types des coefficients de réflexion.

$R(i,j)$ est une matrice antisymétrique car $r_{ij} = -r_{ji}$ tandis que S est une matrice symétrique.

$$R(0) = \frac{\sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (r_{ij}^2 + \sigma_{ij}^2)}{\sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (r_{ij}^2 + \sigma_{ij}^2).$$

2.3.7. Conclusion-discussion

Le modèle markovien présenté par Velzeboer (1981) cherche à rendre compte des caractéristiques spectrales de la série des coefficients de réflexion, en particulier pour les temps proches de zéro. En effet, en traitement du signal classique, pour les problèmes d'estimation du signal et de déconvolution, la séquence primaire des coefficients de réflexion $\{r(t)\}$ est assimilée à un bruit blanc (variables aléatoires indépendantes, de distribution identique). La fonction d'autocorrélation est donc identiquement nulle sauf au temps $t = 0$, le spectre de puissance est plat jusqu'à la fréquence de Nyquist.

En fait, il est reconnu, depuis déjà un certain temps, que la fonction d'autocorrélation des coefficients de réflexion calculés à partir de logs enregistrés sur des puits est loin de présenter un spectre plat. En particulier, la F.A.C. (fonction d'autocorrélation) a souvent des valeurs très **négatives** pour les premiers intervalles temps où elle est estimée. O'Doherty et Anstey (1971) donnent comme explication à ce comportement des interférences constructives entre un grand nombre de multiples internes courts ("peg-legs") qui corréleront positivement avec les coefficients de réflexion primaires. Cette corrélation **positive** entre les multiples "peg-legs" et les coefficients primaires, qui restaure en partie l'énergie des primaires qui seraient autrement très rapidement atténuées par les pertes en transmission, entraîne que l'autocorrélation est le plus souvent **négative** pour les temps faibles (proches de l'origine).

Velzeboer (1981) retrouve ce comportement de la fonction d'autocorrélation au voisinage de l'origine. Par contre, Godfrey, Muir et Rocca (1980), qui utilisent un modèle markovien pour le logarithme de l'impédance mais une autocorrélation purement exponentielle, ne le mettent pas

en évidence. Cependant leur modèle correspond mal au comportement réel de la F.A.C. observé sur des puits, pour les temps courts. L'accord entre le modèle et les données réelles n'est satisfaisant que loin de l'origine.

Plus récemment, Walden et Hosken (1985) utilisent un modèle ARMA pour les coefficients de réflexion primaires. Ce modèle rend très bien compte des spectres réels observés sur un grand nombre de puits et permet une classification des séquences en deux grandes classes : séries très cycliques (sédimentation répétitive) opposées aux séries beaucoup plus aléatoires. Les paramètres ϕ_1 et θ_1 du modèle ARMA (1,1) déterminent le comportement de la F.A.C., qui décroît exponentiellement à partir de $r(t=1)$; le signe de $r(t=1)$ est donné par le signe de $(\phi_1 - \theta_1)$, indiquant à partir de quel côté, par rapport à zéro, la décroissance se produit.

Nous avons privilégié dans notre analyse le modèle de Velzeboer (1981) car il semblait particulièrement adapté aux séries argilo-sableuses deltaïques. Comme le modèle de Walden et Hosken (1985), il rend bien compte du comportement de l'autocorrélation au voisinage de l'origine.

Le modèle markovien présenté a donc été utilisé pour étudier la signature sismique théorique de divers environnements géologiques du delta du Niger. Avant de discuter les résultats de cette application, nous présentons de façon assez détaillée ce cadre géologique particulier, car des tests sur d'autres paramètres sismiques, mais dans ce **même environnement**, seront présentés ultérieurement.

2.4. Les environnements géologiques du delta du Niger comme cadre d'application de certaines méthodes

La mégaséquence régressive observée dans le delta du Niger est décrite sur la figure 32.

2.4.1. Ensembles de lithofaciès

On peut distinguer cinq ensembles de lithofaciès.

- Le premier ensemble comporte des sables hétérogènes grossiers propres en niveaux massifs (épaisseur souvent supérieure à 100 m) et des argiles "carton" hétérogènes très pures. Les contacts inférieurs et supérieurs des niveaux sableux sont francs.
- Le second ensemble présente des alternances de sables hétérogènes grossiers, de sables fins et d'argiles hétérogènes. Les sables fins forment rarement des niveaux massifs ; le plus

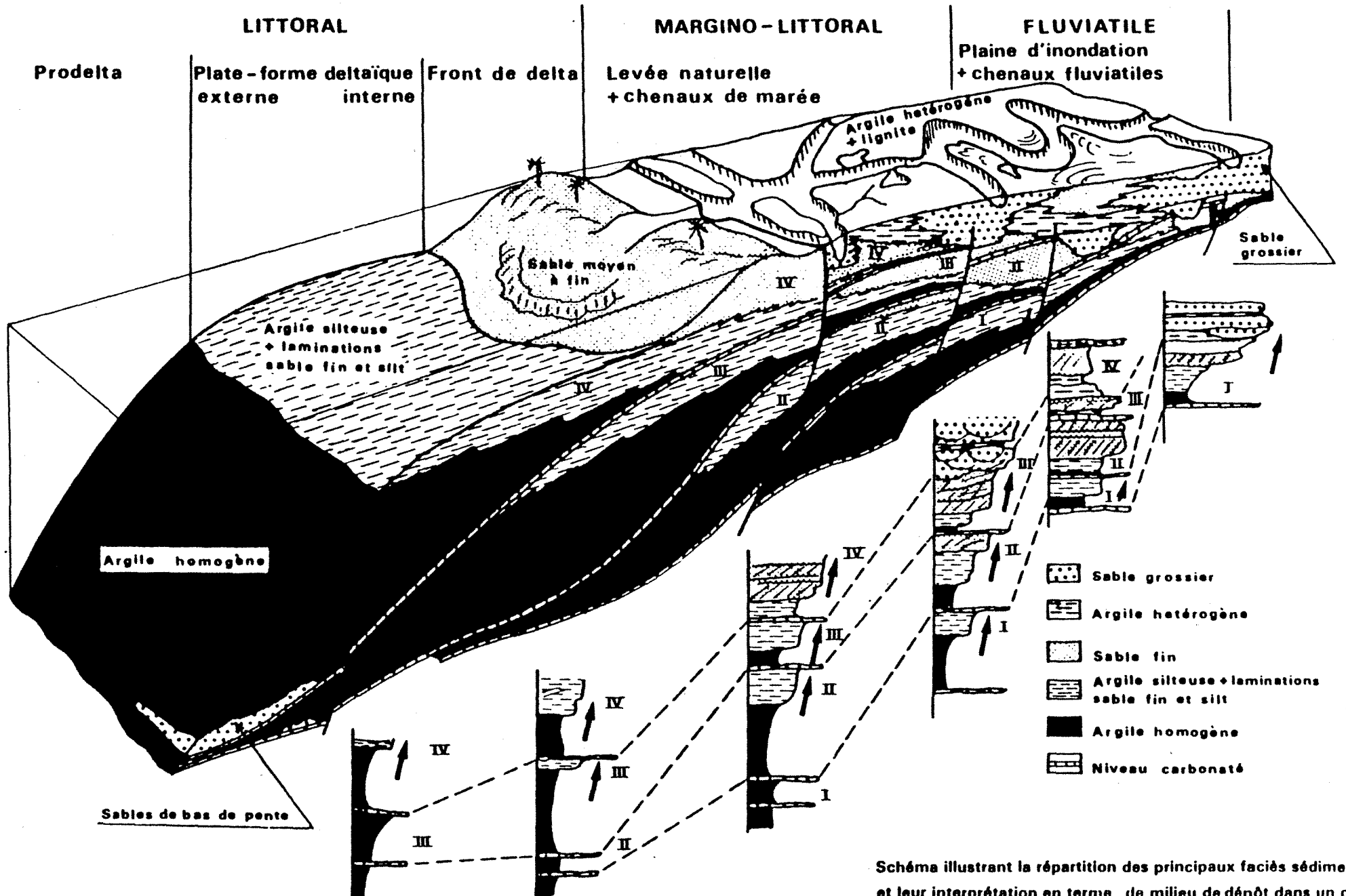


Fig. 32

Schéma illustrant la répartition des principaux faciès sédimentaires et leur interprétation en terme de milieu de dépôt dans un delta.

souvent, on observe des alternances avec des couches argileuses riches en débris végétaux. Cette transition est brutale.

- Le troisième ensemble comprend des sables fins propres ou argileux ; les bioturbations sont nombreuses ainsi que les niveaux de calcite ou sidérite. Le passage des sables argileux aux sables propres se fait de manière continue. Cet ensemble évolue souvent de façon séquentielle négative.
- Le quatrième ensemble correspond à des alternances d'argiles sableuses ou silteuses et de niveaux de sables fins ou de silts argileux. Les caractéristiques principales sont la bioturbation importante –surtout dans les niveaux de silts et sables fin argileux–, la faune très abondante et les nombreux niveaux carbonatés.
- Le cinquième ensemble englobe des alternances d'argiles homogènes et de niveaux d'argiles silteuses.

Ces divers ensembles ont des réponses diagraphiques particulières ainsi que les séquences formées par leur association.

2.4.2. Principales séquences lithologiques

On distingue les séquences positives ("fining-upwards") des séquences négatives ("coarsening-upwards").

2.4.2.1. La séquence négative littorale

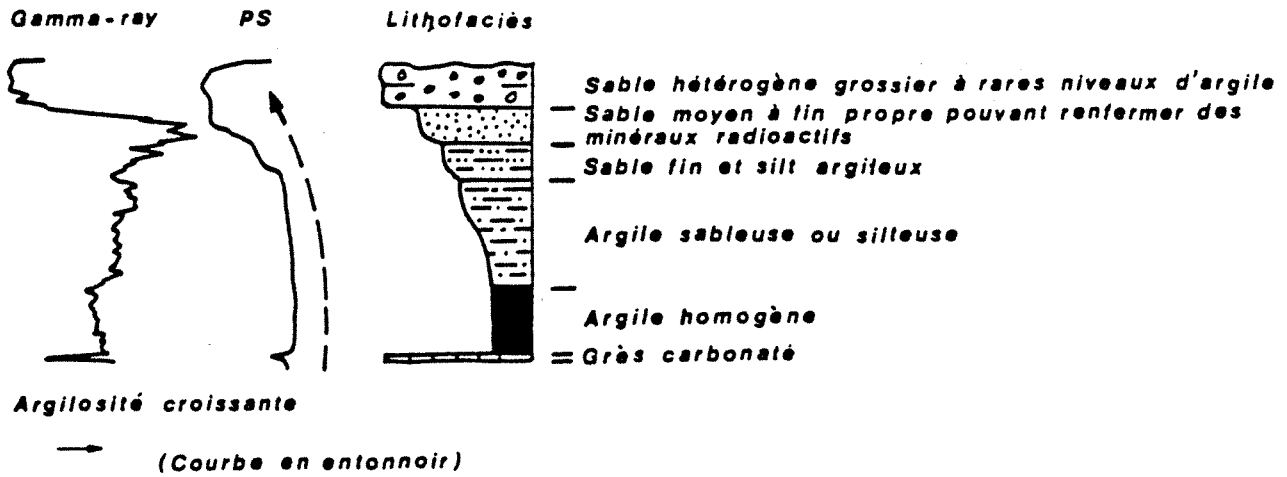
On se rapportera à la figure 33a pour la schématisation de cette séquence.

LITHOFACIES	MILIEUX DE DEPOT
<ul style="list-style-type: none"><li data-bbox="384 432 826 465">– sables fins, propres ou argileux <li data-bbox="384 719 820 842">– argiles sableuses alternant avec des niveaux de sables fins silts argileux <li data-bbox="384 909 890 987">– argiles homogènes alternant avec des niveaux d'argiles silteuses <li data-bbox="384 1055 667 1088">– niveaux carbonatés	<p data-bbox="1007 432 1453 651">dépôt de front de delta regroupant les dépôts de plage, d'avant-plage, parfois de barrière et pied de barrière, de chenier, de "split" et de barre d'embouchure</p> <p data-bbox="1007 719 1390 752">dépôt de plate-forme deltaïque</p> <p data-bbox="1007 909 1342 943">dépôt de pente de prodelta</p> <p data-bbox="1007 1055 1326 1088">niveau marin transgressif</p>

**Tableau 2 : séquence négative littorale
d'après Louis, Lacaze, Sulpice, Un delta subsident à travers
quelques techniques d'exploration (1978)**

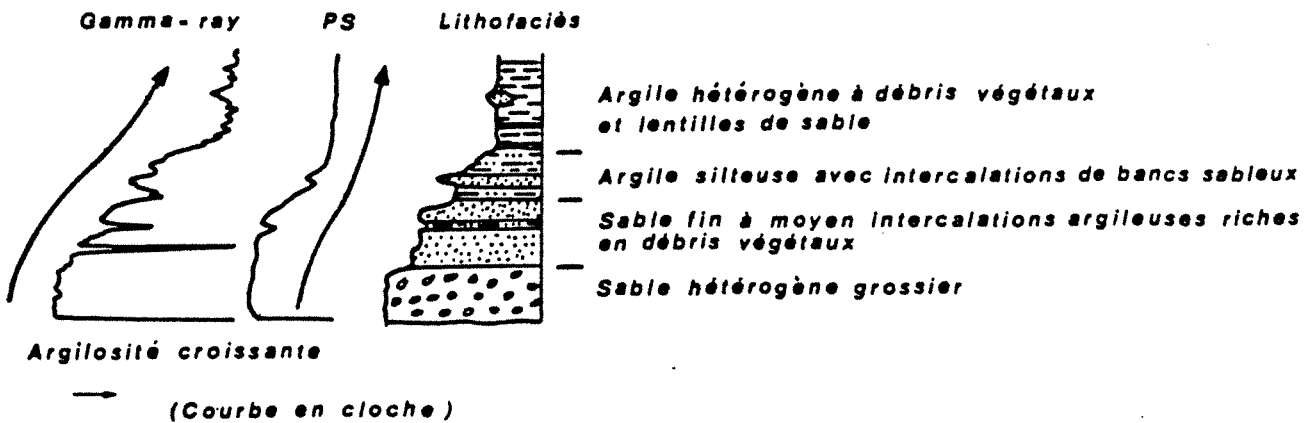
Cette séquence se caractérise par un accroissement de la granulométrie de bas en haut. C'est la séquence régressive deltaïque typique où l'on passe, depuis la base jusqu'au sommet, des faciès marins d'argiles homogènes, à des alternances sables-argiles, puis à des sables fins correspondant à des dépôts de front de delta.

Q. SEQUENCE ELEMENTAIRE NEGATIVE
(coarsening-upwards séquence)

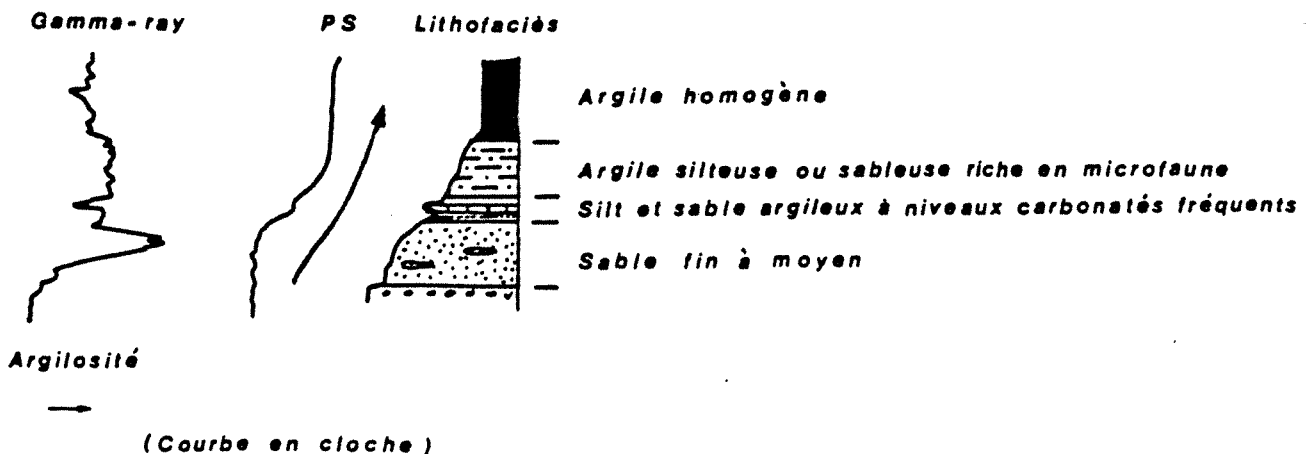


b+c. SEQUENCE ELEMENTAIRE POSITIVE
(fining-upwards séquence)

b. SEQUENCE POSITIVE A BASE COMPOSEE DE SABLE GROSSIER



C. SEQUENCE POSITIVE A BASE COMPOSEE DE SABLE FIN



d'après LOUIS, LACAZE, SULPICE.

Un delta subsident à travers quelques techniques d'exploration

2.4.2.2. La séquence positive de chenal

Cette séquence est schématisée sur la figure 33b.

LITHOFACIES	MILIEUX DE DEPOT
<ul style="list-style-type: none">- argiles hétérogènes riches en débris de plantes, traces de racines- sables fins alternant avec des passages argileux- sables grossiers hétérogènes	<p>plaine d'inondation</p> <p>levée naturelle</p> <p>chenal</p>

**Tableau 3 : séquence positive de chenal
d'après Louis, Lacaze, Sulpice, Un delta subsident à travers quelques
techniques d'exploration (1978)**

La granulométrie décroît depuis la base jusqu'au sommet. La séquence complète présente à sa base des sables grossiers hétérogènes, puis des alternances de sables fins et d'argiles riches en débris végétaux. Au sommet, on observe des argiles hétérogènes à débris végétaux.

Cette séquence est typique des chenaux méandriformes de marée.

On observe une séquence très voisine, mais tronquée vers le haut, dans les parties plus amont du delta, dans la zone où se développe un réseau fluvial en tresse.

L'évolution séquentielle est peu marquée et ceci s'explique par un empilement de chenaux qui s'érodent successivement.

LITHOFACIES	MILIEUX DE DEPOT
- argiles hétérogènes	Zone interchenaux (?) Niveau de débordement des levées
- sables grossiers	Lit du chenal

**Tableau 4 : séquence de chenal en amont du delta
d'après Louis, Lacaze, Sulpice, Un delta subsident à travers quelques
techniques d'exploration (1978)**

2.4.2.3. La séquence positive transgressive

Elle est schématisée sur la figure 33c.

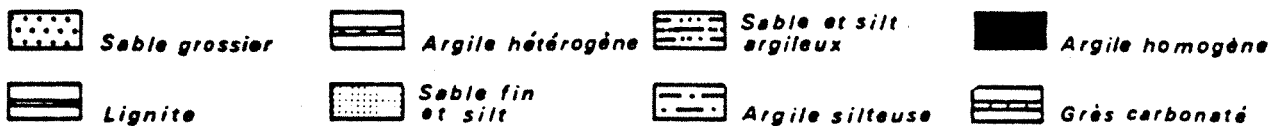
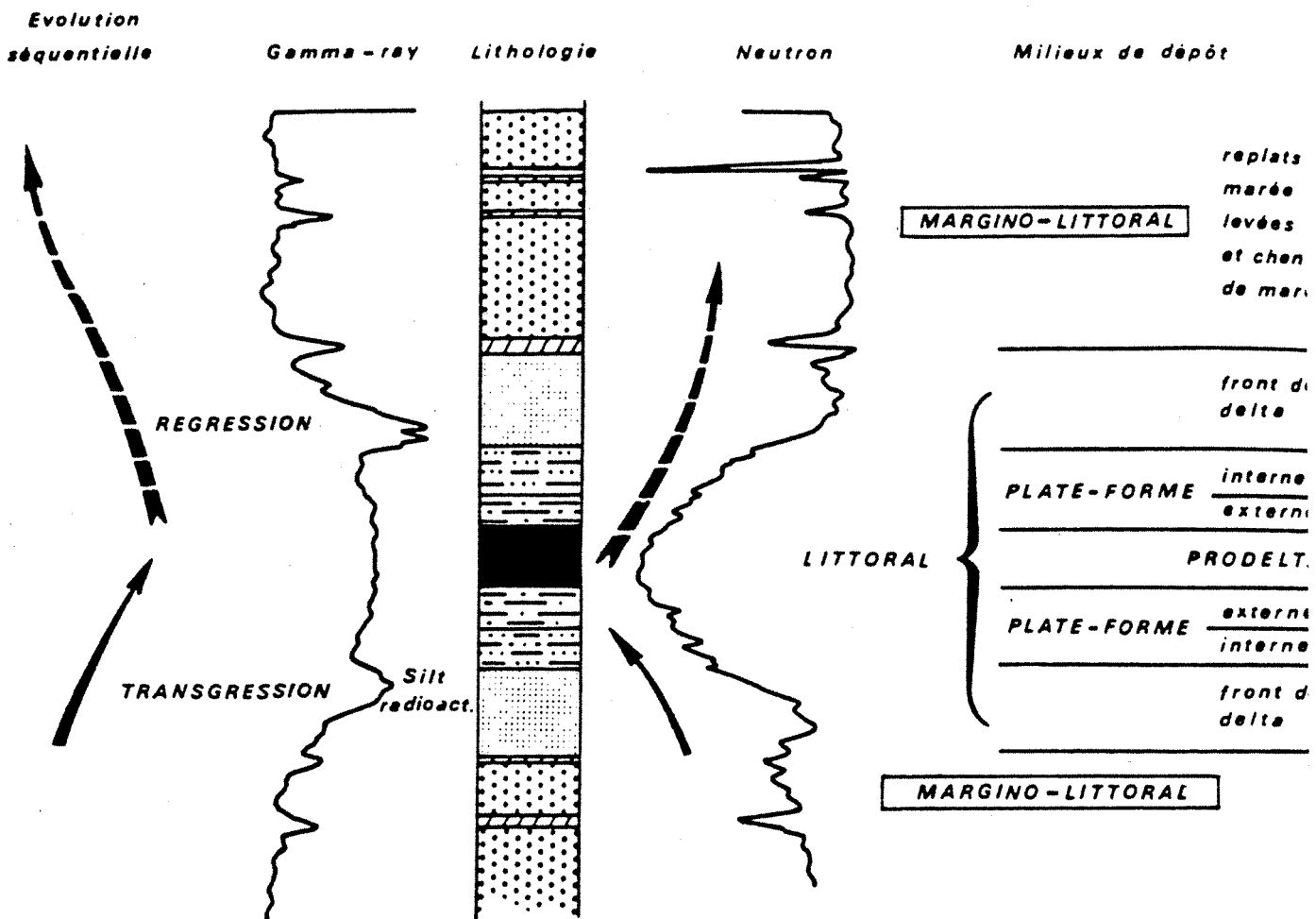
C'est une séquence positive dont la base est constituée de sables fins et qui évolue vers des argiles homogènes. Cette séquence correspond à une transgression ; elle s'accompagne d'un remaniement intense des séries antérieures avec formation de barres "off-shore" progressivement ennoyées.

2.4.2.4. Association de séquences

La figure 34 illustre les possibilités d'association des séquences étudiées précédemment pour la constitution de cycles, à signification transgressive-régressive, par exemple (Figure 34).

2.4.3. Application à l'étude des séries rencontrées dans les forages du delta du Niger

L'interprétation des séquences en terme de paléo-environnement est basée sur la "séquence paysage" (Figure 32). Au cours d'un forage, les caractéristiques diagaphiques de chaque



Ce cycle comprend deux séquences élémentaires :

- séquence positive à la base correspondant à un épisode transgressif
- séquence négative sommitale, équivalent de la séquence régressive deltaïque classique

Schéma du cycle Transgression-Régression

Extrait de LOUIS, LACAZE, SULPICE

un delta subsident à travers quelques techniques d'exploration

séquence sont analysées ainsi que les différents lithofaciès qui les constituent. Chaque séquence est donc replacée dans son paléo-environnement.

Sur plusieurs permis du delta du Niger, nous avons utilisé ces informations pour découper la colonne lithologique forée en intervalles profondeur correspondant à un même paléo-environnement. C'est la réponse sismique synthétique de ces intervalles qui a ensuite été analysée. En fait, nous avons considéré trois principaux environnements géologiques (correspondant à trois subdivisions majeures de la séquence paysage) :

- tous les intervalles prélevés depuis le prodelta jusqu'à la plate-forme interne ont été regroupés sous le terme "plate-forme",
- les intervalles découpés dans les milieux de plate-forme interne à front de delta ont été regroupés dans un ensemble appelé "front de delta",
- du point de vue lithofaciès et aspect séquentiel, nous n'avons vu que peu de différences entre la zone margino-littorale et la plaine fluviatile. Tous les échantillons prélevés dans ces deux environnements ont été affectés à un même groupe appelé "complexe de chenaux et plaine d'inondation".

Le premier groupe, "plate-forme", comprend des intervalles à dominante argileuse (plus de 70 %) avec des passées d'argiles silteuses. Les séquences sont assez discrètes, de taille moyenne (80 m).

Le second groupe comprend tous les faciès de barrière (à dominante sableuse) et les faciès de pied de barrière (dominante silteuse). L'aspect séquentiel est très caractéristique : on observe une succession de courtes séquences régressives (environ 30 m d'épaisseur) allant des argiles silteuses aux sables fins propres, les séquences se terminant souvent par un banc carbonaté.

Le troisième groupe comprend des séries riches en bancs de sables grossiers, d'épaisseur variable mais souvent forte (> 80 m), avec quelques niveaux d'argiles hétérogènes. On ne voit presque aucune évolution séquentielle.

Ces regroupements sont non seulement justifiés du point de vue sédimentologique mais correspondent aussi à des faciès sismiques typiques.

Le groupe **plate-forme deltaïque** s'exprime sismiquement par des réflexions parallèles continues, d'amplitude forte à moyenne et variant peu latéralement (Figure 35). Les fréquences apparentes sont plutôt basses (Figure 35).

Les alternances de sables fins et d'argiles silteuses déposées au niveau du **front de delta** génèrent des réflexions parallèles, de continuité bonne à moyenne avec de fortes amplitudes alternant avec des niveaux de faibles amplitudes (Figure 36). Le niveau d'amplitude varie peu, les phases et les fréquences sont relativement stables latéralement. Verticalement, on observe des alternances de zones à moyenne ou haute fréquence apparente et de zones à basse fréquence apparente (Figure 36).

Les alternances de sables grossiers et argiles hétérogènes déposées dans le **complexe margino-littoral** (chenaux méandriques de marée et plaine deltaïque) sont à l'origine de réflexions parallèles et discontinues, d'amplitudes fortes à faibles avec des variations rapides (Figure 37). On observe également des modifications rapides de la phase et de la fréquence (Figure 37).

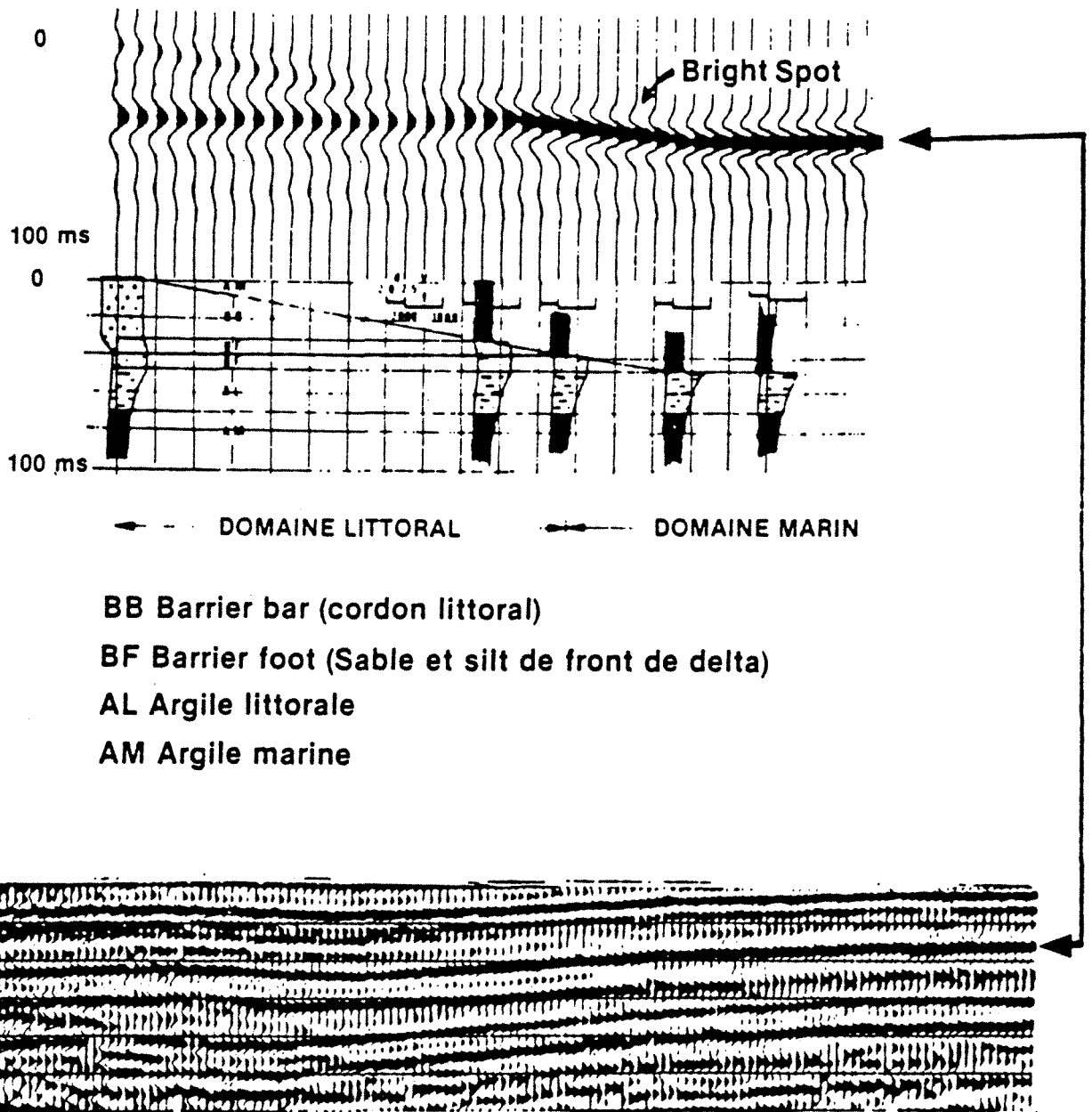
Ces divers faciès sismiques expliquent le regroupement que nous avons fait : le groupe "front de delta" s'oppose par sa forte cyclicité et courte période au groupe "plate-forme" où les séquences sont plus longues et moins marquées, à dominante plutôt argileuse. Leurs réponses sismiques sont bien différentes.

2.5. Application des résultats théoriques sur l'autocorrélation au delta du Niger

Le modèle markovien présenté dans le paragraphe 2.3 a été appliqué aux données de puits du delta du Niger pour les trois principaux environnements :

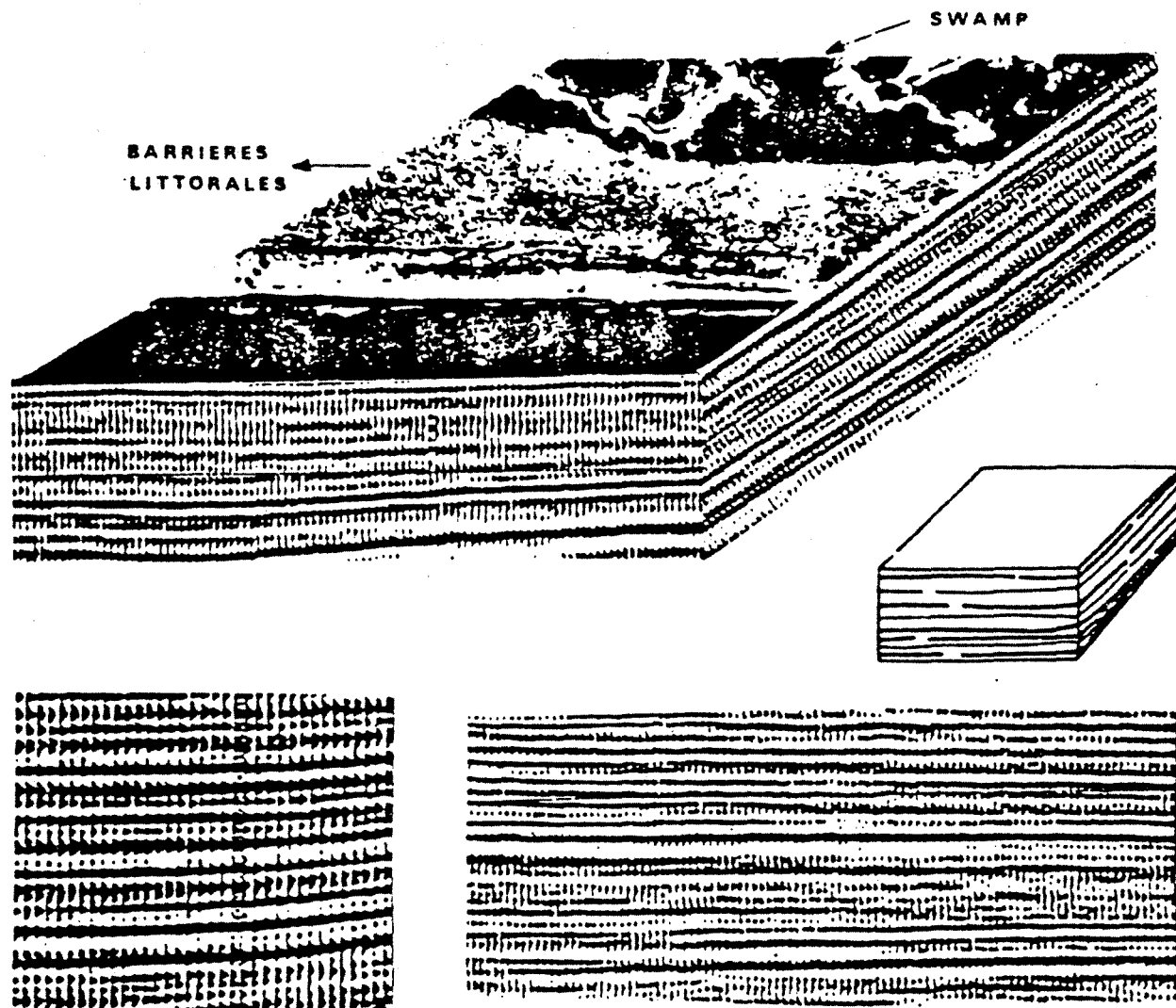
- margino-littoral,
- front de delta,
- plate-forme.

Il s'agissait de voir si le spectre de puissance théorique offrait une caractérisation des trois zones.

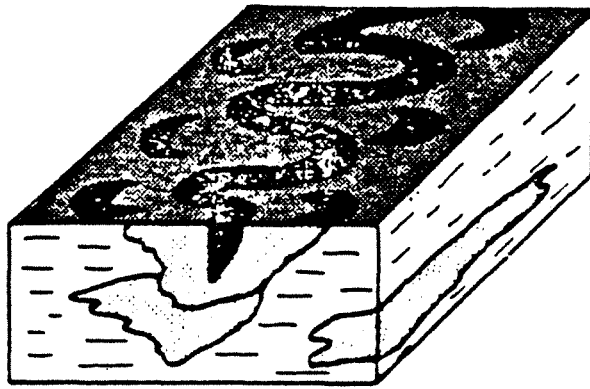


- BB Barrier bar (cordon littoral)
- BF Barrier foot (Sable et silt de front de delta)
- AL Argile littorale
- AM Argile marine

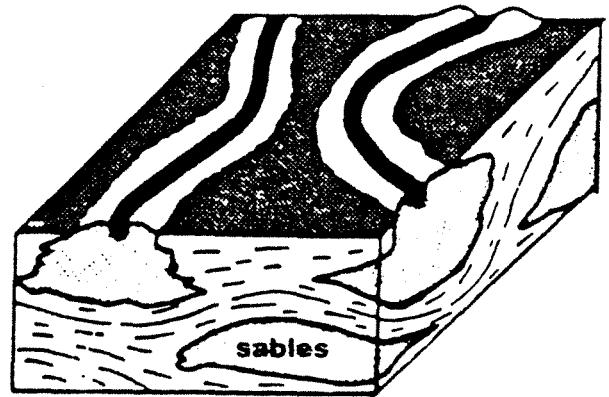
D'après LOUIS, LACAZE, SULPICE : Un delta subsident à travers quelques techniques d'exploration



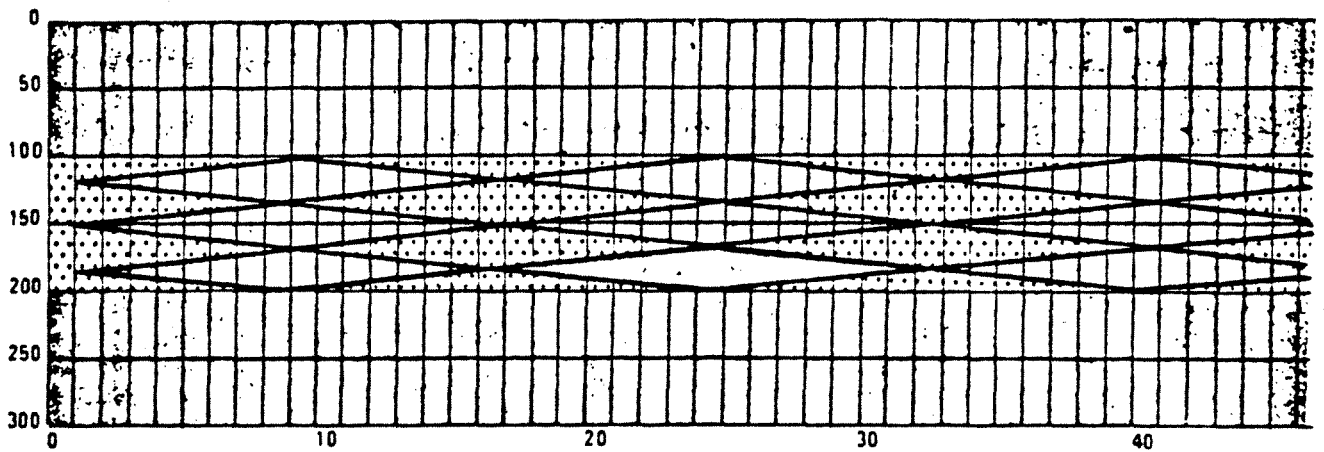
D'après LOUIS, LACAZE, SULPICE : Un delta subsident à travers quelques techniques d'exploitation



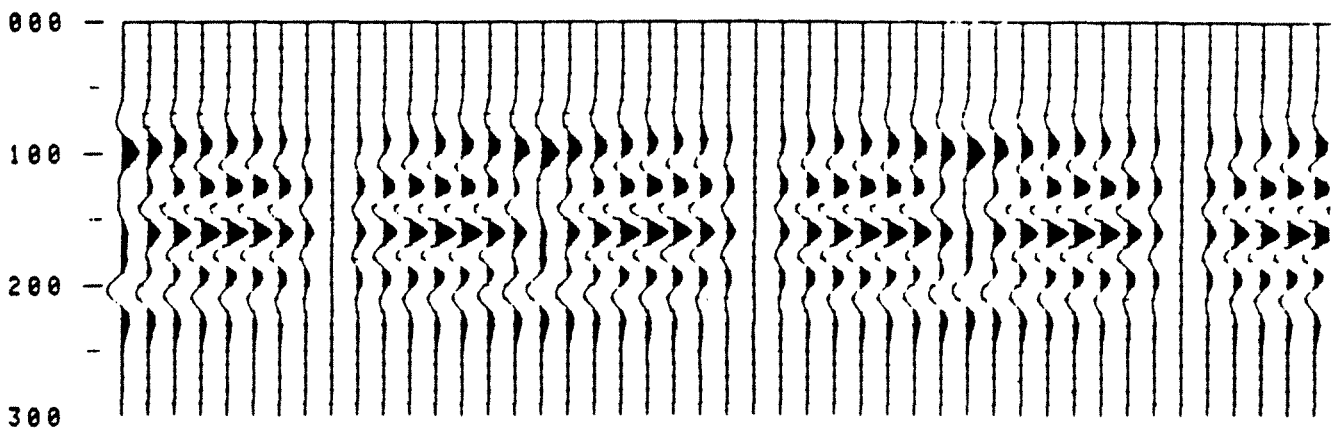
**SYSTEME FLUVIATILE
MEANDRIFORME**



PLAINE DELTAIQUE



Argiles Sables



Réflexions discontinues avec phase, fréquence et amplitude variables

D'après LOUIS, LACAZE, SULPICE : Un delta subsident à travers quelques techniques d'exploration

2.5.1. Cas de deux lithologies

Les deux lithologies considérées sont sables et argiles. Les vitesses sont évaluées à partir d'enregistrements diagraphiques aux puits. Nous n'avons pas tenu compte des densités pour calculer les coefficients de réflexion.

$$v_{\text{sables}} = 3100 \text{ m/s}$$

$$v_{\text{argiles}} = 2700 \text{ m/s}$$

$$\text{soit } r_{12} = 0,069.$$

$$\text{Nous avons choisi : } \frac{\sigma_{12}}{r_{12}} = 0,10.$$

L'épaisseur des diverses lithologies a été étudiée statistiquement pour les divers environnements à partir de l'étude sédimentologique des puits (Tableaux 5 à 7).

	Epaisseur moyenne	Epaisseur maximale
Sables } épaisseurs Argiles } (m)	19,156 3,577	81,62 8,128
γ (Hz)	916,7	368,9
$1/\gamma$ (ms)	2,71	1,09

Tableau 5 : zone margino-littorale

	Epaisseur moyenne	Epaisseur maximale
Sables } épaisseurs Argiles } (m)	13, 878 9,072	102,61 40,26
γ (Hz)	521,0	97,3
$1/\gamma$ (ms)	10,28	1,92

Tableau 6 : zone du front de delta

	Epaisseur moyenne	Epaisseur maximale
Sables } épaisseurs Argiles } (m)	5,08 9,008	14,22 42,67
γ (Hz)	910,0	281,3
$1/\gamma$ (ms)	3,55	1,10

Tableau 7 : zone de la plate-forme interne

La matrice de transition est invariante (cas de deux lithologies), le seul paramètre qui va influencer sur l'autocorrélation et son spectre est donc γ qui dépend directement des épaisseurs moyennes rencontrées dans les divers milieux.

Le paramètre γ est assez voisin pour les milieux margino-littoral et de plate-forme. Les réponses sismiques seront donc voisines.

C'est bien ce que l'on observe sur les spectres d'amplitude normés à 100 Hz des modèles : seule la courbe correspondant au front de delta s'individualise nettement (Figures 38 et 39).

2.5.2. Cas de quatre lithologies

Pour les trois principaux paléo-environnements étudiés à partir de l'étude complète de trois puits, nous avons déterminé les matrices de transition dans le cadre de quatre lithologies dont les vitesses sont les suivantes :

argiles	$v = 2700$ m/s
argiles silteuses	$v = 2900$ m/s
sables	$v = 3100$ m/s
sables grossiers	$v = 3500$ m/s.

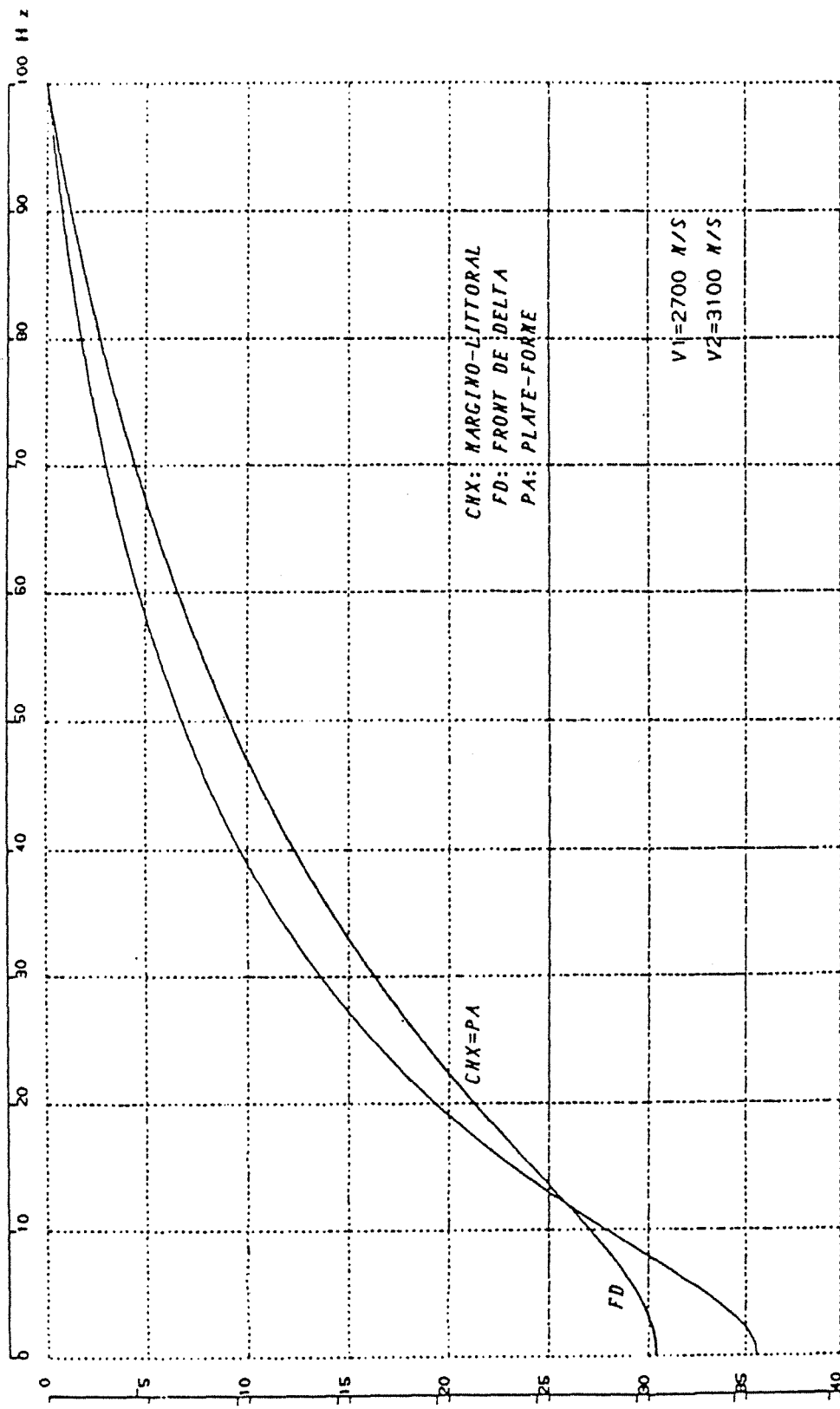
Pour les trois puits, les matrices de transition d'un même environnement étaient relativement comparables, ce qui nous a permis de les moyenner. Pour chaque paléo-environnement, nous disposons d'une matrice de transition, ainsi que de la statistique des épaisseurs des diverses lithologies.

Dans un premier temps, les épaisseurs sont **constantes, quel que soit le paléo-environnement.**

L'écart-type vaut approximativement 10 % du coefficient de réflexion. Le seul paramètre qui diffère d'un environnement à l'autre est la **matrice de transition.**

Pour une épaisseur donnée (Figure 40), on s'aperçoit que l'influence de la matrice de transition est relativement faible. Cependant, le faciès margino-littoral présente un spectre moins contrasté que front de delta et plate-forme interne.

CAS DE DEUX LITHOLOGIES
APPLICATION AUX SERIES DU DELTA DU NIGER
SPECTRES DE PUISSANCE NORMES A 100 HZ. EN DB
HYPOTHESE MOYENNE SUR LES EPAISSEURS



CAS DE DEUX LITHOLOGIES
 APPLICATION AUX SERIES DU DELTA DU NIGER
 SPECTRES DE PUISSANCE NORMES A 100 HZ. EN DB
 HYPOTHESE MAXIMALE SUR LES EPAISSEURS

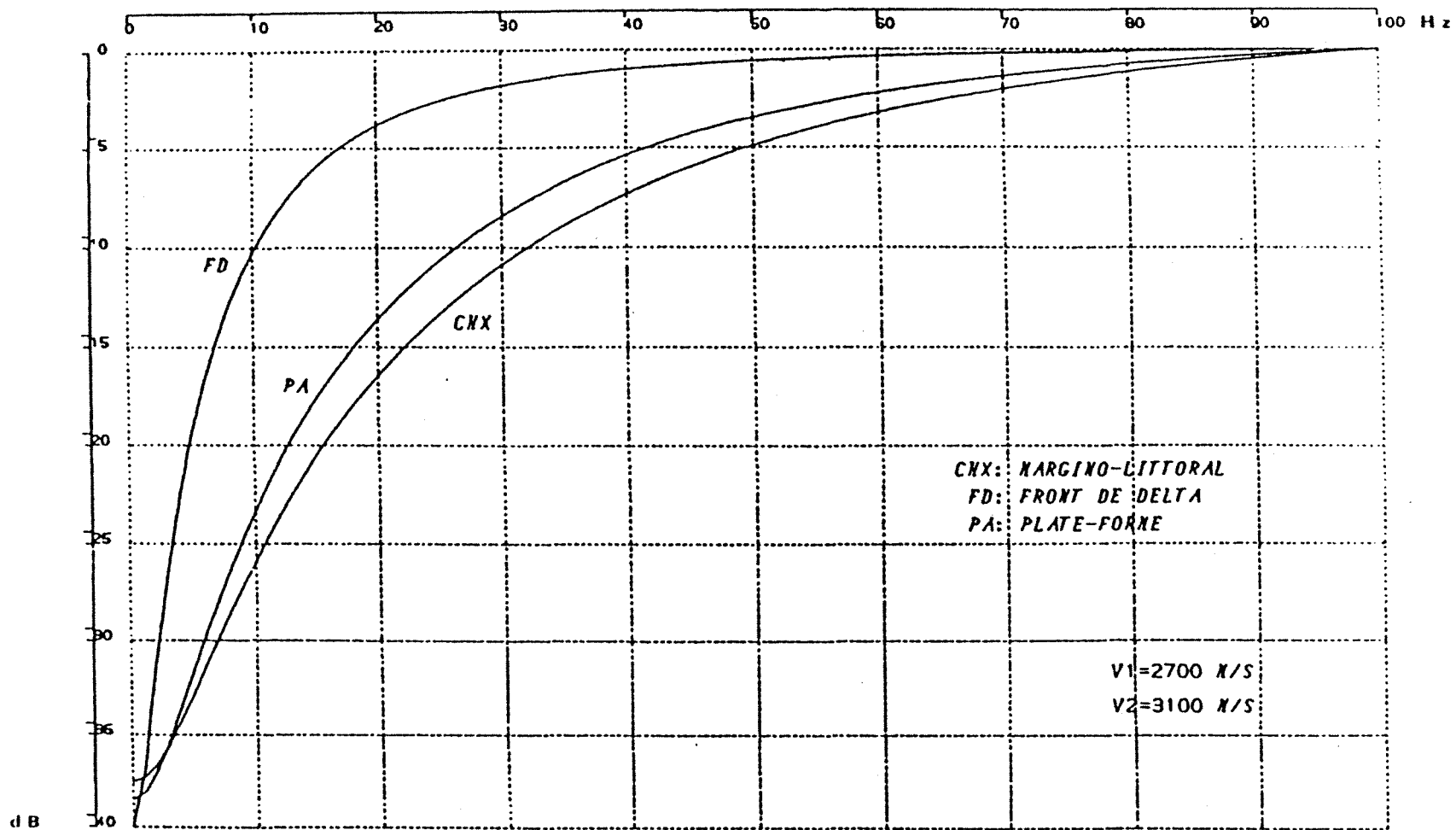


Fig. 39

Fig. 40

CAS DE QUATRE LITHOLOGIES
 INFLUENCE DU PALEO-ENVIRONNEMENT
 EPAISSEURS DES DIVERSES LITHOLOGIES
 TOUTES EGALES A: 10 M

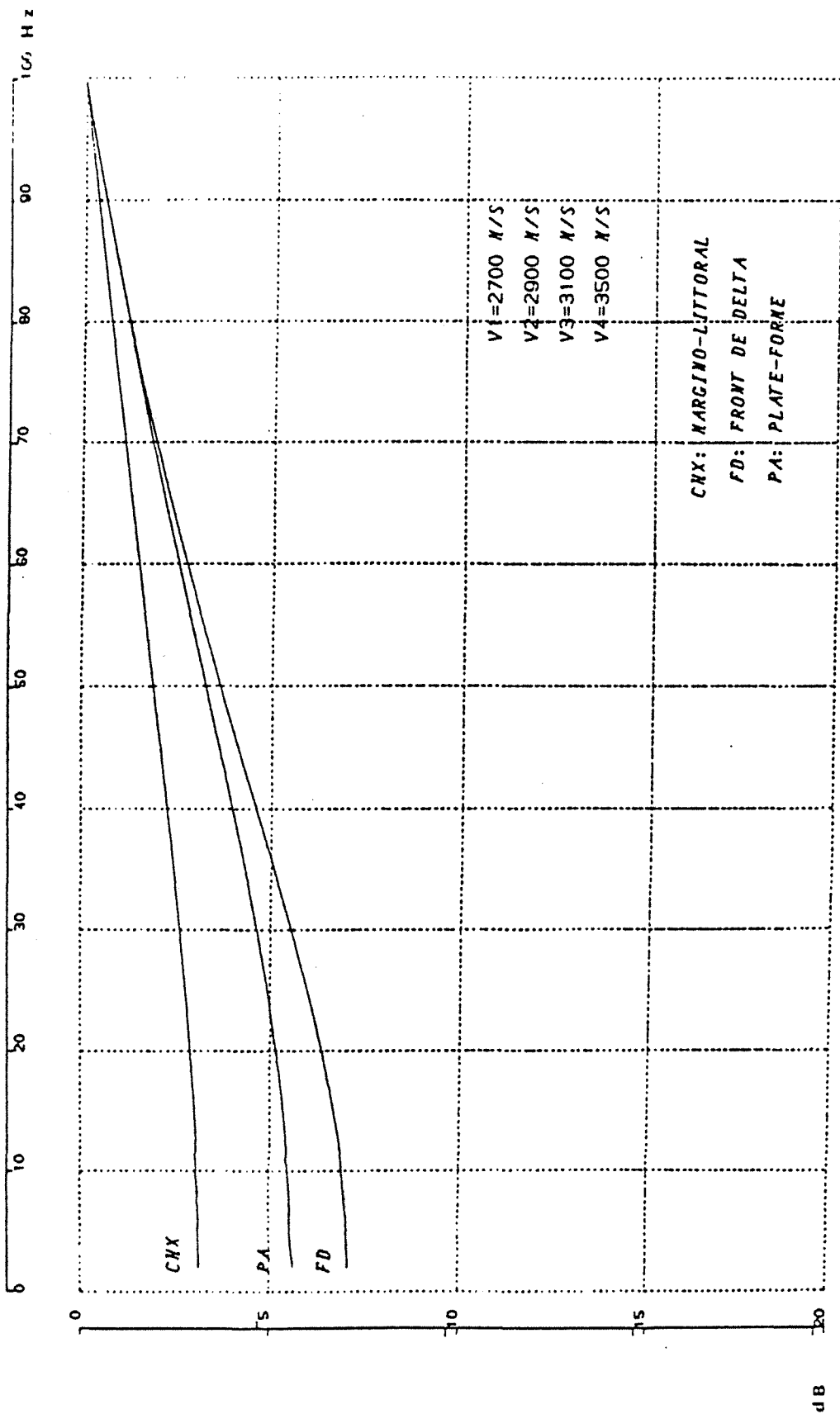


Fig. 41

CAS DE QUATRE LITHOLOGIES
 INFLUENCE DU PALEO-ENVIRONNEMENT
 EPAISSEURS DES DIVERSES LITHOLOGIES
 TOUTES EGALES A: 20 M

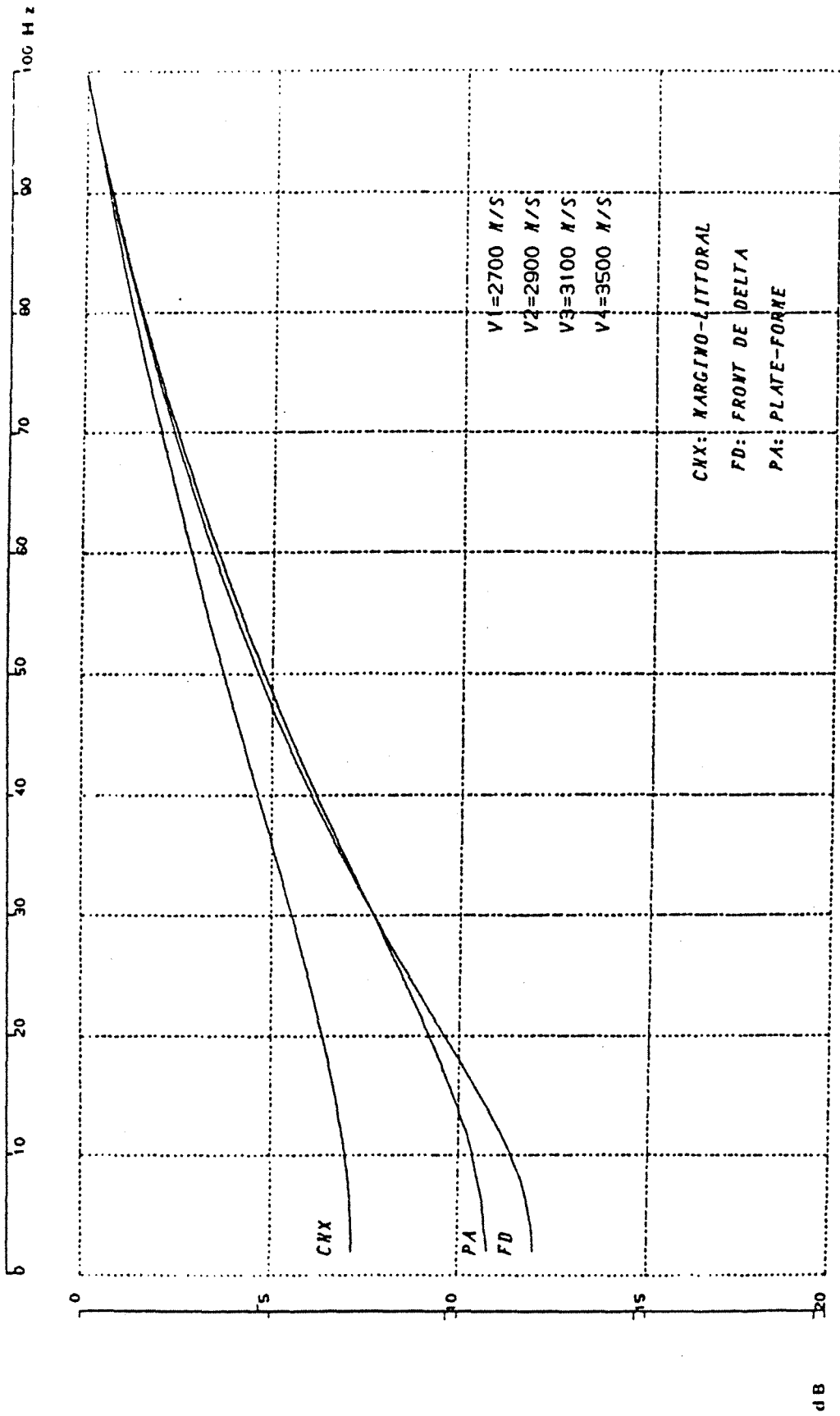
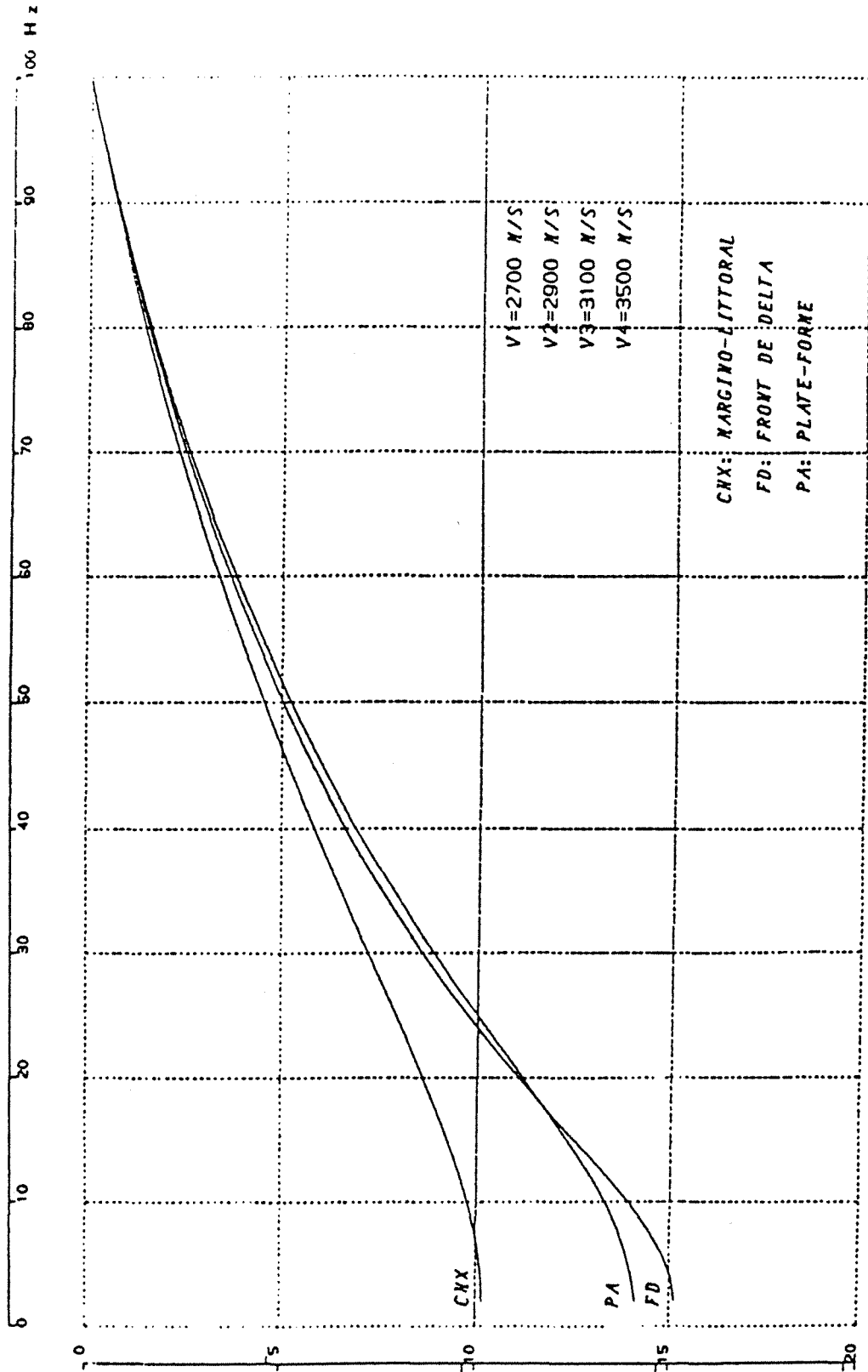


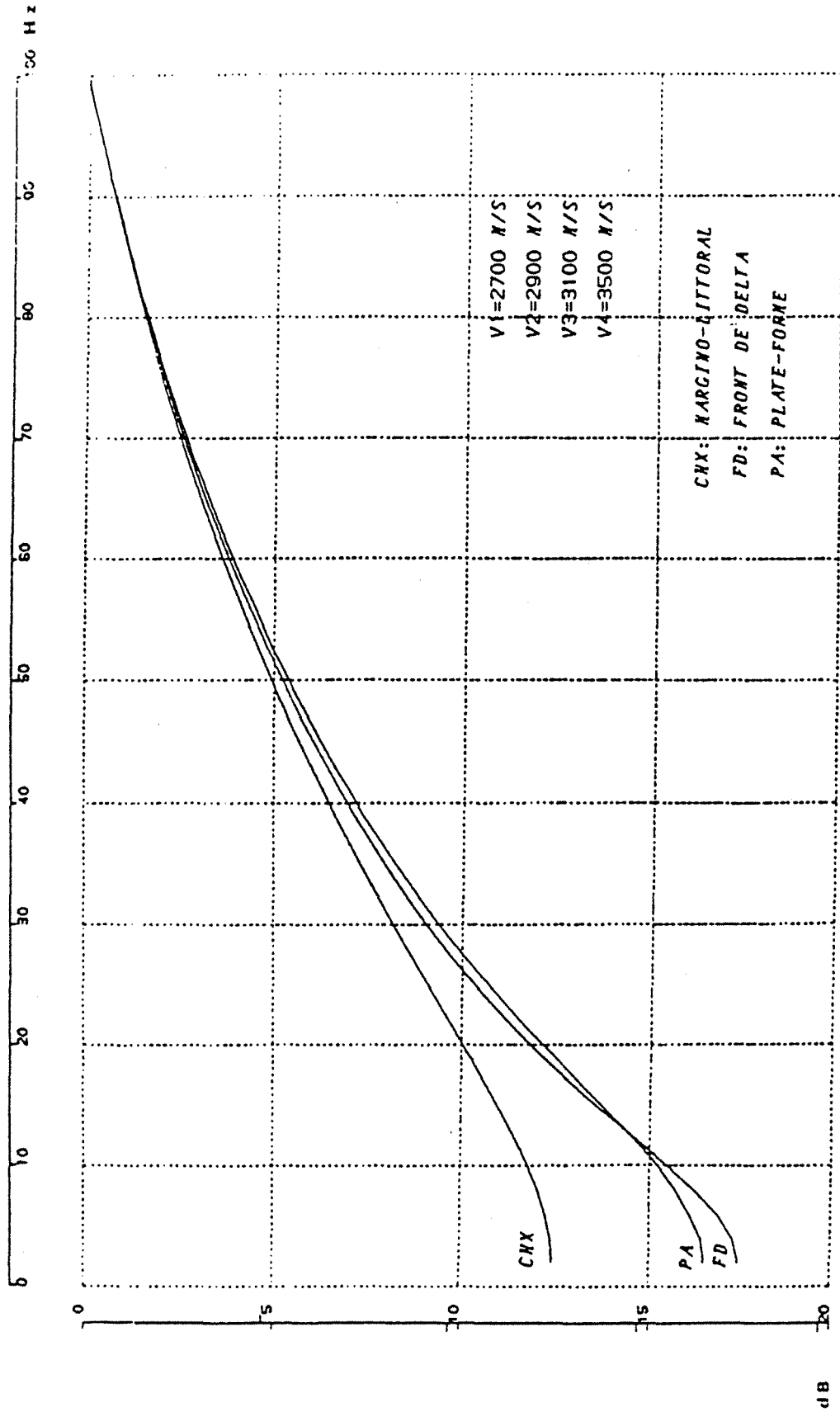
Fig. 42

CAS DE QUATRE LITHOLOGIES
 INFLUENCE DU PALEO-ENVIRONNEMENT
 EPAISSEURS DES DIVERSES LITHOLOGIES
 TOUTES EGALES A: 30 M

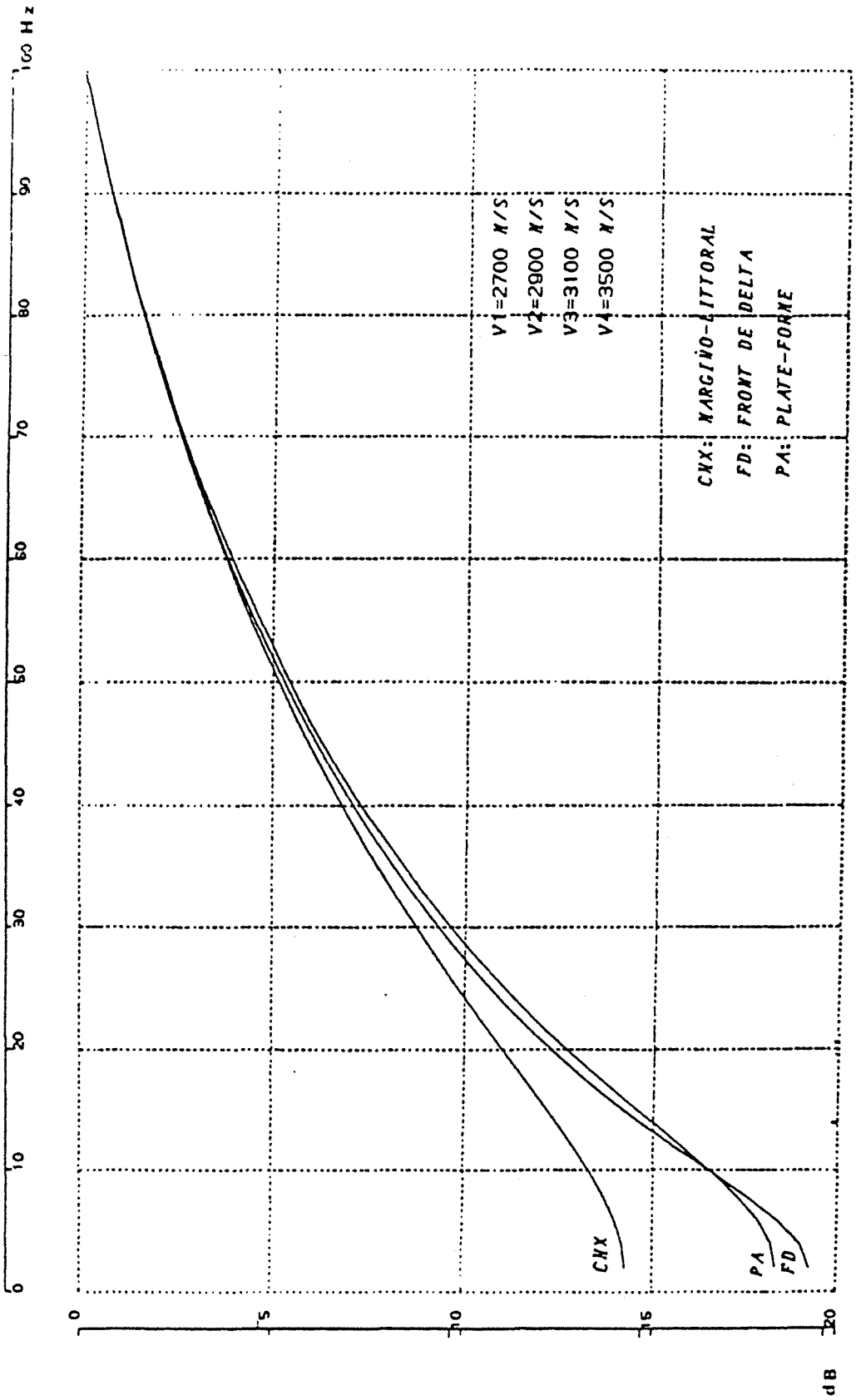


dB

CAS DE QUATRE LITHOLOGIES
 INFLUENCE DU PALEO-ENVIRONNEMENT
 EPAISSEURS DES DIVERSES LITHOLOGIES
 TOUTES EGALES A: 40 M



CAS DE QUATRE LITHOLOGIES
 INFLUENCE DU PALEO-ENVIRONNEMENT
 EPAISSEURS DES DIVERSES LITHOLOGIES
 TOUTES EGALES A: 50 M



Au fur et à mesure que l'épaisseur augmente (Figures 40 à 44), le spectre d'amplitude est de plus en plus contrasté (spectre enrichi en basses fréquences) mais l'écart entre les trois groupes diminue. Dans le domaine accessible à l'observateur, en particulier, les trois courbes sont presque confondues.

Ainsi, on constate que la matrice de transition a relativement peu d'influence sur les courbes théoriques de l'autocorrélation et du spectre de puissance.

On a repris l'étude des trois principaux environnements avec leur matrice de transition propre mais en prenant comme épaisseur des lithologies constitutives, les valeurs moyennes des épaisseurs observées aux puits dans chaque environnement (Figures 45 à 48).

Les spectres d'amplitude théoriques, normés à 100 Hz et exprimés en dB, des zones margino-littorales et de la plate-forme interne sont très semblables, presque blancs (Figures 47 et 48), seul s'individualise le front de delta. Cette différence est d'autant plus sensible que l'on prend pour épaisseur moyenne des diverses lithologies une valeur forte (comparer les Figures 47 et 48).

2.6. Conclusions

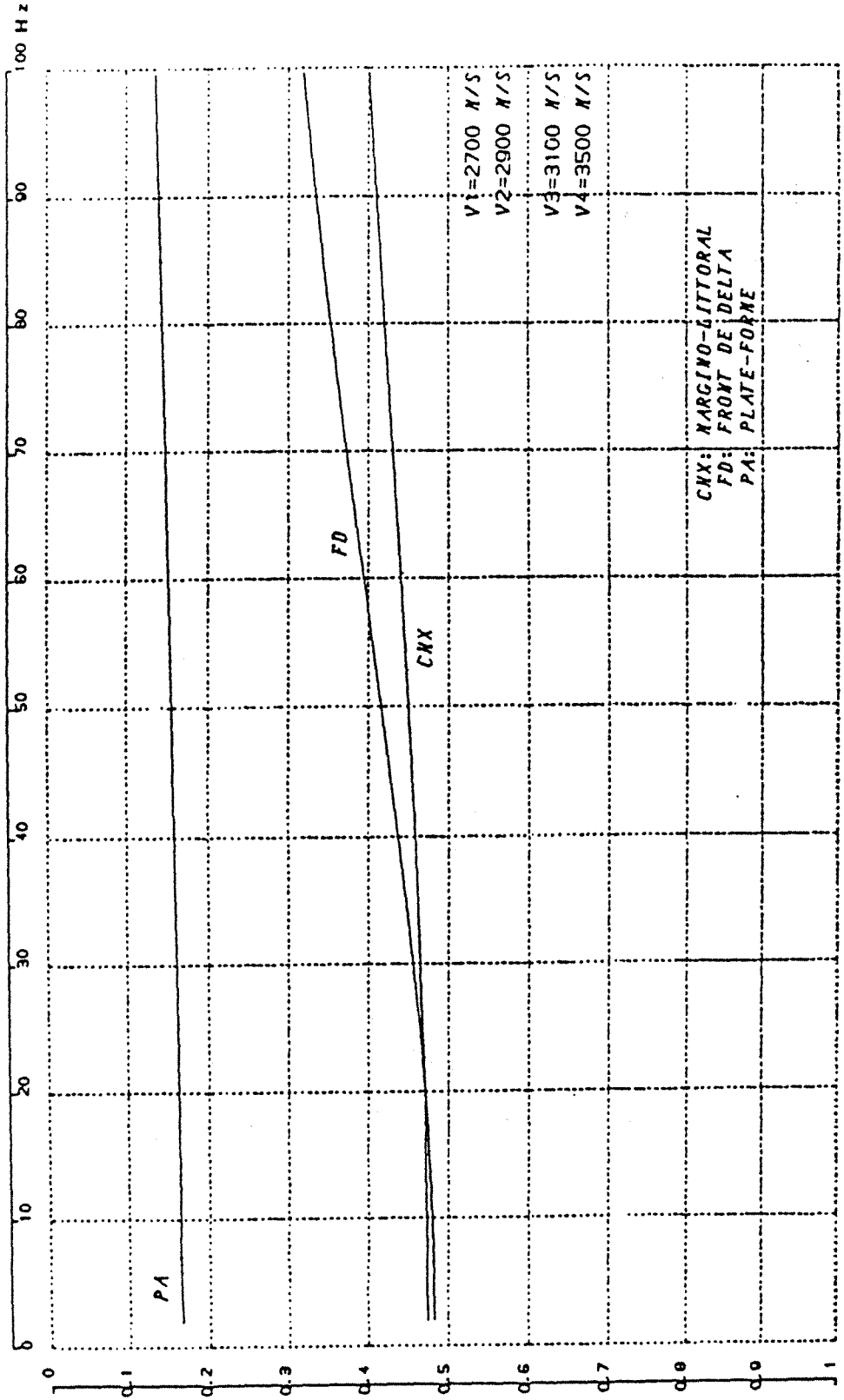
Dans le cas de deux lithologies ou de quatre lithologies, les études paramétriques des courbes du spectre de puissance théorique montrent l'influence prédominante des épaisseurs sur la réponse sismique théorique de la séquence. En particulier, pour deux lithologies, on a bien mis en évidence le rôle déterminant du paramètre γ (γ s'exprime en Hz, c'est l'inverse d'un temps de transfert moyen, $\gamma = \frac{v_1}{D_1} + \frac{v_2}{D_2}$). La réponse sismique, à travers le spectre de puissance, sera caractéristique de l'environnement pour des valeurs de γ inférieures à 500 Hz. Au-delà, la partie exponentielle de l'autocorrélation est négligeable et on obtient un Dirac au temps $t = 0$ dont le spectre d'amplitude est constant.

Dans le cadre simplificateur des hypothèses adoptées, les résultats obtenus montrent que le spectre de puissance peut aider à la caractérisation de l'environnement géologique, ceci dans une certaine gamme d'épaisseurs des lithologies constitutives du milieu.

Cependant, ces hypothèses sont assez restrictives par rapport à la réalité géologique complexe comme le montre, par exemple, la comparaison de la fonction d'autocorrélation d'une trace synthétique (calculée à partir des données Sonic et Densité d'un puits) et de l'autocorrélation théorique dans un **même environnement géologique**.

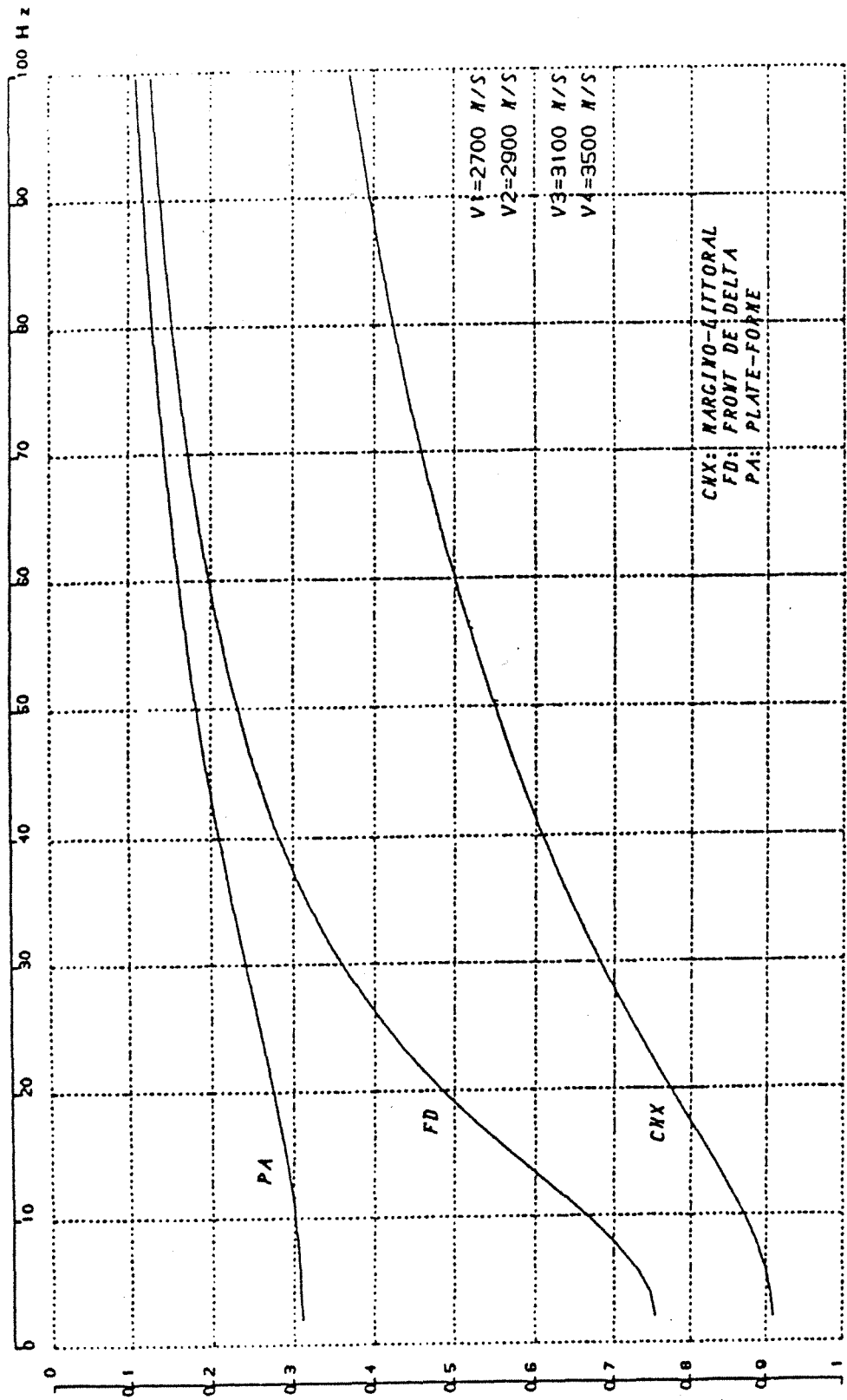
CAS DE QUATRE LITHOLOGIES
APPLICATION AUX SERIES DU DELTA DU NIGER

SPECTRES D AMPLITUDE
HYPOTHESE MOYENNE SUR LES EPAISSEURS



CAS DE QUATRE LITHOLOGIES
APPLICATION AUX SERIES DU DELTA DU NIGER

SPECTRES D'AMPLITUDE
HYPOTHESE MAXIMALE SUR LES EPAISSEURS



CAS DE QUATRE LITHOLOGIES
APPLICATION AUX SERIES DU DELTA DU NIGER

HYPOTHESE MOYENNE SUR LES EPAISSEURS

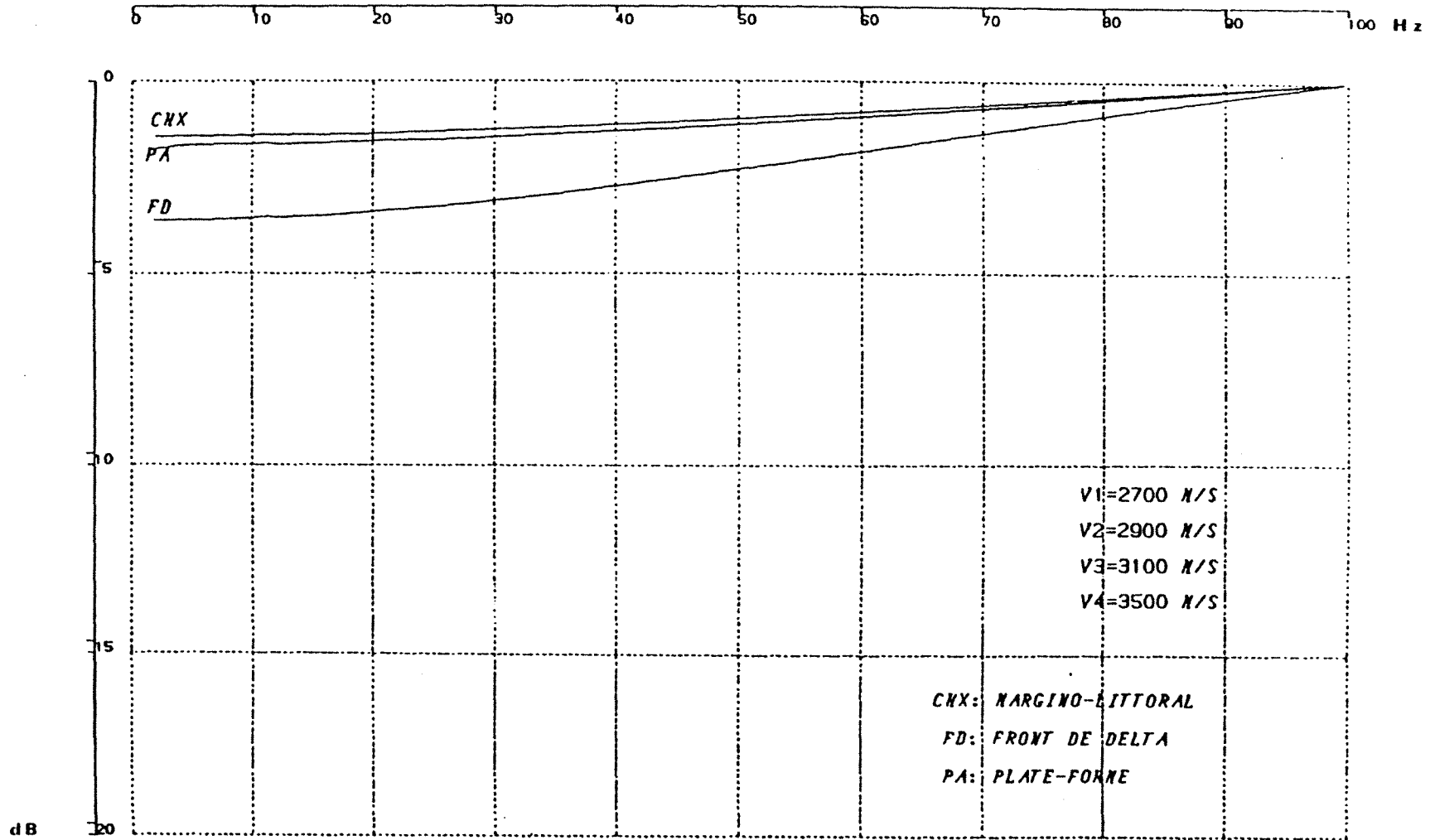


Fig. 47

CAS DE QUATRE LITHOLOGIES
 APPLICATION AUX SERIES DU DELTA DU NIGER

HYPOTHESE MAXIMALE SUR LES EPAISSEURS

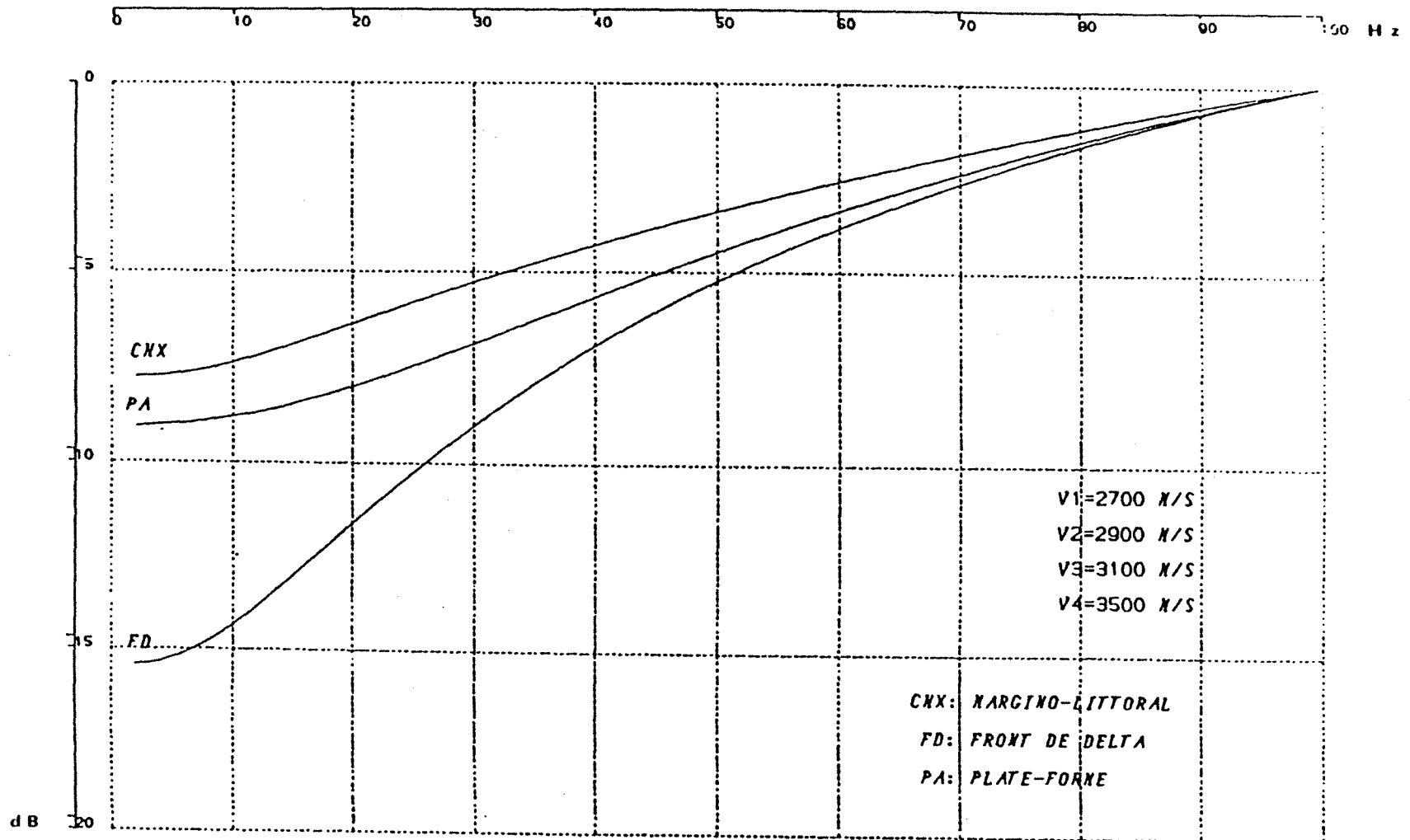


Fig. 48

Pour vérifier le réel pouvoir de caractérisation de la fonction d'autocorrélation, il semble plus intéressant de calculer un grand nombre de paramètres sur cette fonction et son spectre et d'appliquer des méthodes multivariées sur une population de traces pour lesquelles on dispose d'une information a priori quant à leur environnement géologique. Cette approche, déjà suivie par Khattri et al. (1976, 1979, 1983, 1984) permettra de quantifier le pouvoir discriminant des fonctions considérées (autocorrélation, spectre de puissance, représentation temporelle de la trace...).

C'est ce type d'analyse que nous allons privilégier dans les paragraphes suivants pour nous aider à déterminer les paramètres sismiques pertinents pour l'analyse du faciès.

L'approche théorique du pouvoir discriminant de l'autocorrélation, utilisant le modèle proposé par Velzeboer (1981), est trop liée au modèle utilisé pour décrire la séquence sédimentaire. Les hypothèses de ce modèle sont par ailleurs tellement nombreuses qu'on peut s'interroger sur son adéquation avec la réalité géologique.

3. ANALYSE DE LA REPONSE SISMIQUE DES ENVIRONNEMENTS DU DELTA DU NIGER

3.1. Introduction

Le cadre géologique du delta du Niger a déjà été présenté au cours de l'application du modèle markovien à ces séries.

A partir de l'analyse séquentielle d'un certain nombre de puits du delta du Niger, nous avons découpé des zones verticales ou intervalles profondeur, en fonction du paléo-environnement qu'elles représentent.

Les intervalles que nous allons analyser appartiennent à l'un des trois paléo-environnements suivants :

- plate-forme interne,
- front de delta,
- margino-littoral.

Nous avons calculé les traces sismiques synthétiques impulsionnelles de ces différents intervalles (le principe de ce calcul a été rappelé dans la première partie), uniquement à partir de la vitesse Sonic car les enregistrements Densité n'avaient pas été réalisés pour tous les puits étudiés.

Le signal choisi pour calculer les traces synthétiques **filtrées** est un signal standard type "Ricker". Le Ricker d'ordre 0 est la courbe de Gauss, le Ricker d'ordre n est sa dérivée nième. Plus on dérive cette courbe, plus le nombre d'arches secondaires augmente, la phase étant modifiée de $\frac{\pi}{2}$ à chaque dérivation.

Le signal que nous avons sélectionné est un Ricker d'ordre 2 de pseudo-période 32 ms, de longueur totale 80 ms. La figure 49 représente le signal dans le domaine temps et dans le domaine transformé avec spectre d'amplitude et spectre de phase. L'origine du signal a été choisie à -40 ms de telle sorte que les traces synthétiques soient toutes à phase nulle.

Nous utiliserons notre connaissance de l'environnement géologique des différentes traces synthétiques pour étudier le pouvoir de caractérisation de la fonction d'autocorrélation et de son spectre en fonction de cette information a priori.

CARACTÉRISTIQUES DU SIGNAL UTILISÉ

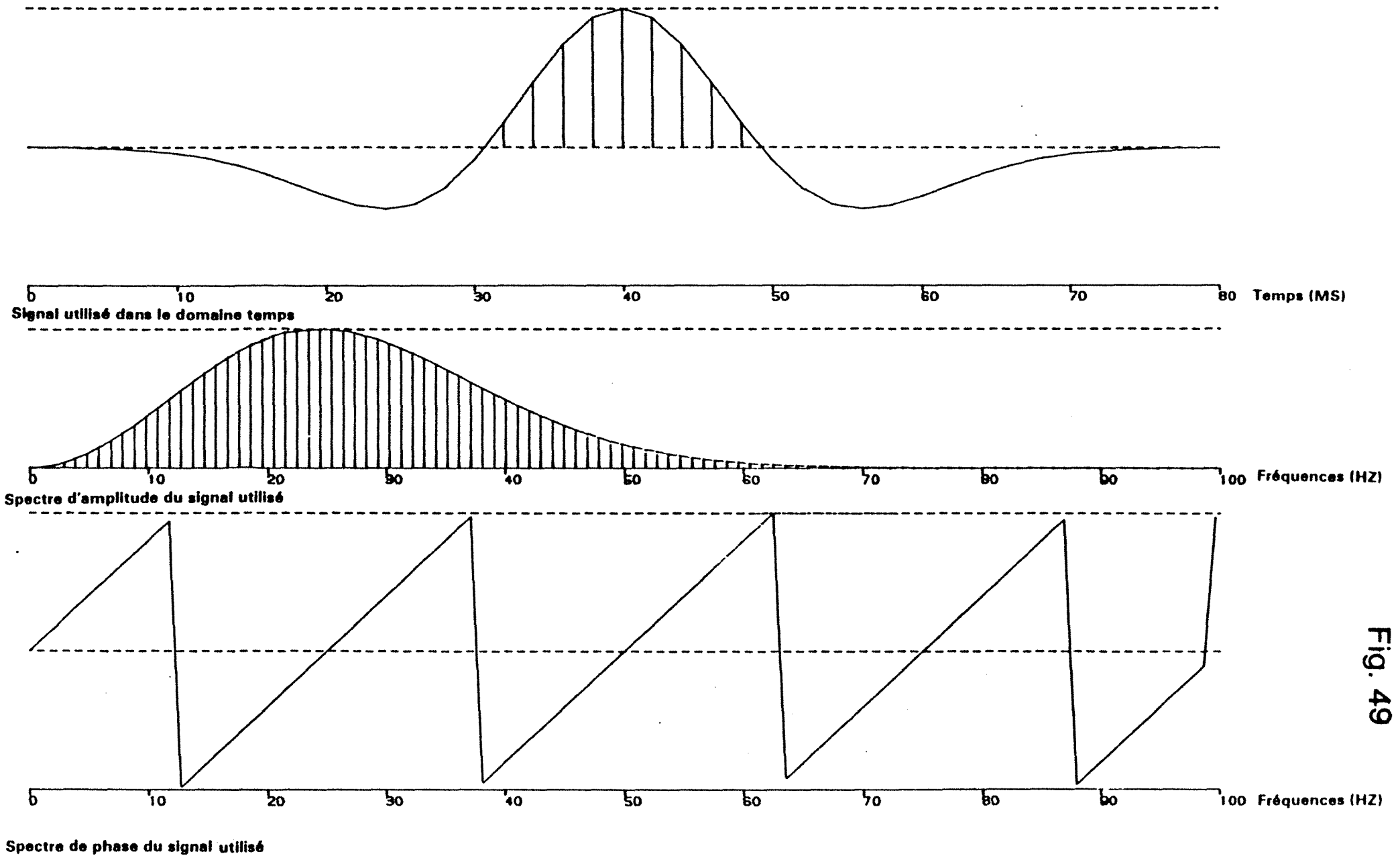


Fig. 49

Puis, en reprenant les principales familles de variables qui apparaissent au travers de la bibliographie (cf. paragraphe 1), nous caractériserons les traces synthétiques par un certain nombre d'attributs sismiques dont nous étudierons le pouvoir discriminant par rapport à l'information a priori.

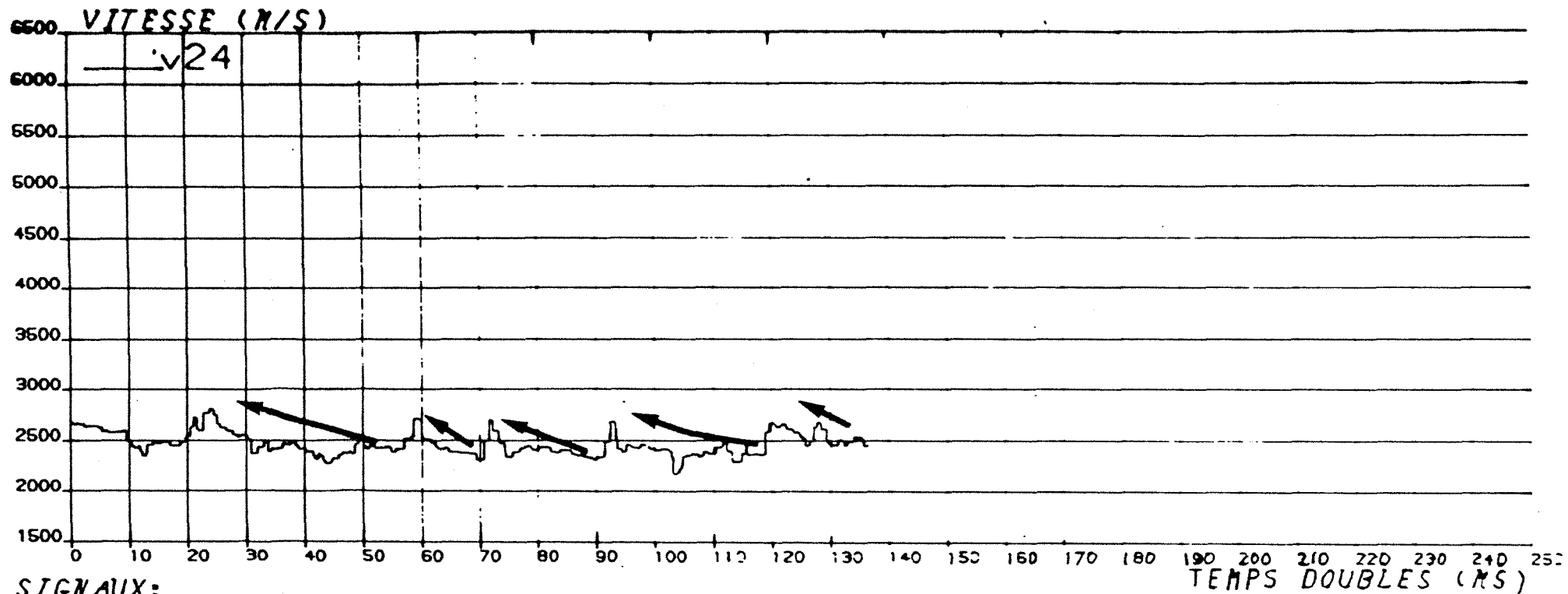
Nous vérifierons ces résultats sur des traces sismiques **réelles** prélevées dans les mêmes environnements du delta.

3.2. Aspect sismique des différents groupes de dépôt

Nous allons présenter, pour les trois paléo-environnements retenus, les traces synthétiques impulsionnelles et filtrées obtenues sur des "échantillons" très caractéristiques du groupe. Par le terme "échantillon", nous entendons la réponse sismique synthétique d'un intervalle profondeur, découpé à partir de l'étude sédimentologique des faciès rencontrés en sondage et qui est homogène du point de vue paléo-environnement.

3.2.1. Groupe *plate-forme*

L'échantillon présenté, Figure 50, est situé, du point de vue paléo-environnement, entre la plate-forme externe et la plate-forme interne. Du point de vue séquentiel, on observe sur le log lithologique des séquences de l'ordre de 50 m allant des argiles marines homogènes aux argiles silteuses. Cette expression séquentielle se retrouve sur le log de vitesse, où l'on observe des séquences de 20 ms temps double, en moyenne, à vitesse croissante vers le sommet. La trace sismique correspondante est assez contrastée, sa fonction d'autocorrélation est représentée sur la figure 51.



SIGNAUX:



TRACES:

1. F240000 2. F240001



3.

4.

5.

INTERVALLE 2180-2350 M PLATEFORME INTERNE A EXTERNE

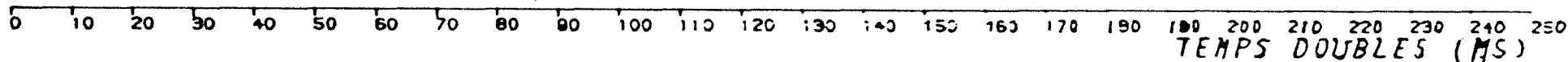
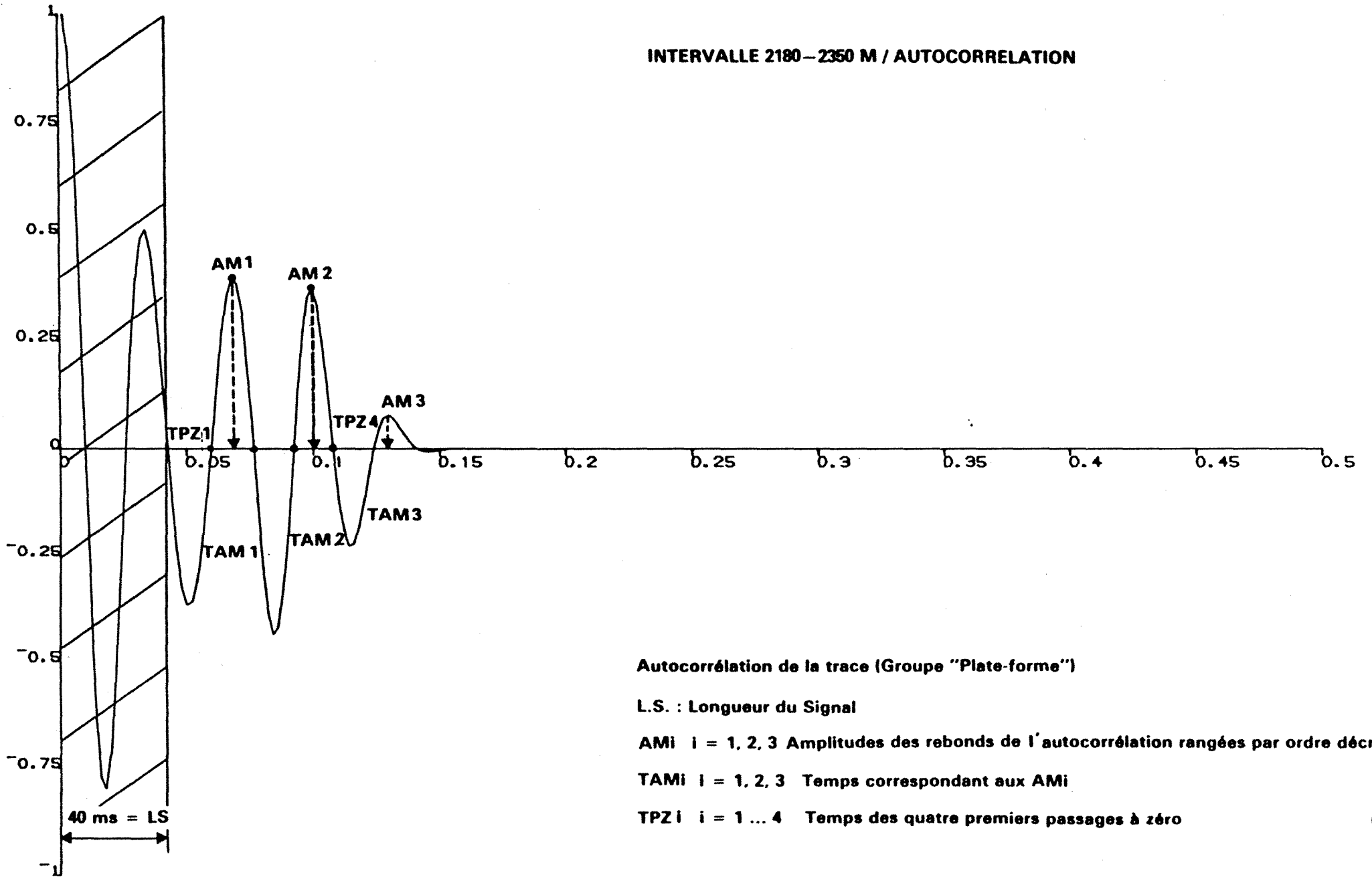


Fig. 50

INTERVALLE 2180-2350 M / AUTOCORRELATION



Autocorrélation de la trace (Groupe "Plate-forme")

L.S. : Longueur du Signal

AM_i $i = 1, 2, 3$ Amplitudes des rebonds de l'autocorrélation rangées par ordre décroissant

TAM_i $i = 1, 2, 3$ Temps correspondant aux AM_i

TPZ_i $i = 1 \dots 4$ Temps des quatre premiers passages à zéro

3.2.2. Groupe front de delta

Nous présentons deux échantillons assez différents, pour illustrer les hétérogénéités qui peuvent exister même au sein d'un groupe.

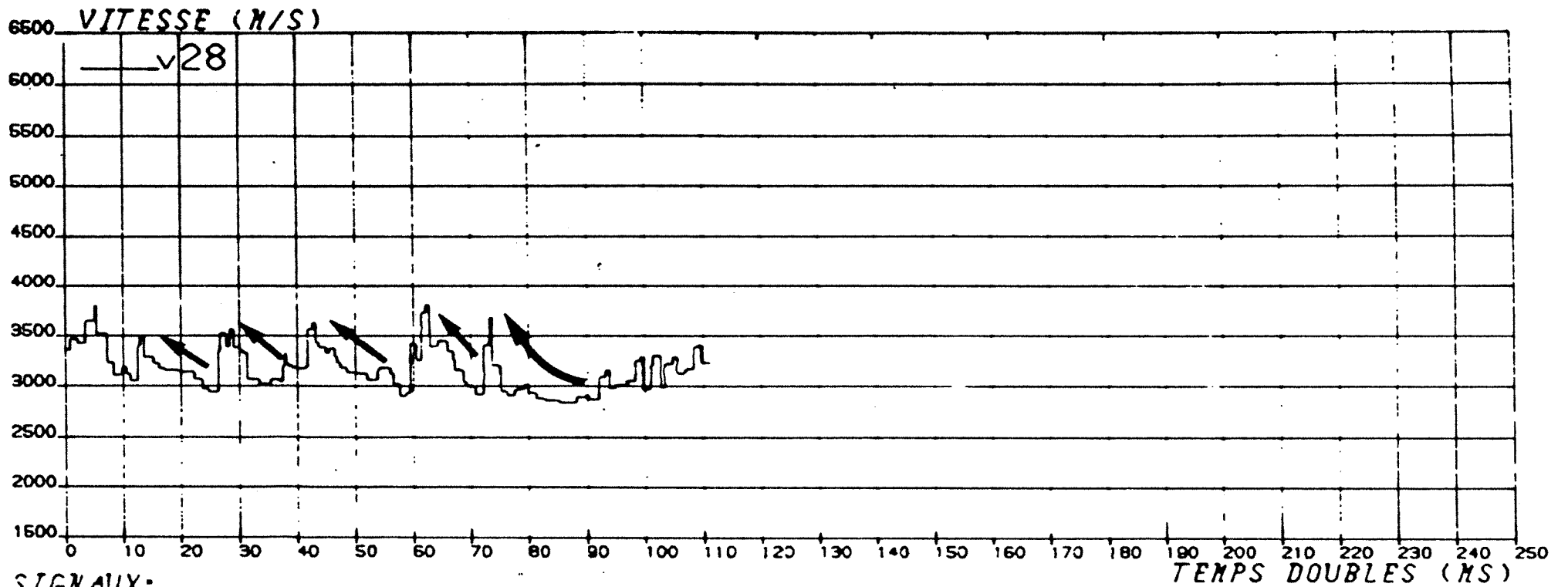
Le premier échantillon (Figure 52) est particulièrement caractéristique de la plupart des individus de "front de delta". Il est constitué de courtes séquences régressives allant des argiles silteuses aux sables fins. Ces séquences se retrouvent sur le log de vitesse (Figure 52) et, par rapport aux séquences observables sur l'échantillon présenté pour la plate-forme interne (Figure 50), elles sont plus courtes et surtout beaucoup plus contrastées, d'autant plus que les séquences peuvent se terminer sur un banc carbonaté qui entraîne l'apparition d'un pic de vitesse.

Le deuxième échantillon (Figure 53) n'est pas du tout caractéristique, du point de vue séquentiel, du front de delta. Il correspond à un important paquet de sables fins argileux à propres, constitué d'une répétition du faciès "barrière". Il s'agit, en fait, d'un empilement de barrières sur elles-mêmes. Il était difficile de négliger ce type d'échantillons, malgré son aspect séquentiel peu caractéristique, car du point de vue pétrolier, il constitue des réservoirs très intéressants. Le log de vitesse présente des vitesses assez élevées et constantes surtout au niveau du paquet de sables principal. On note la présence de quelques bancs carbonatés.

Malgré les différences morphologiques importantes de ces traces, nous pensons qu'il existe des paramètres communs à ces deux intervalles découpés dans l'environnement de front de delta, au niveau de la réponse sismique. Ce n'est qu'en quantifiant celle-ci et en analysant statistiquement les attributs sismiques qu'il sera possible de déterminer ces ressemblances, qui, à première vue, ne sont pas évidentes.

3.2.3. Groupe margino-littoral : chenaux et plaine d'inondation

Les deux échantillons présentés (Figures 54 et 55) correspondent tous deux à des bancs épais (≈ 80 m) de sables grossiers hétérogènes avec quelques niveaux d'argiles hétérogènes. Les vitesses sont relativement faibles et homogènes sur tout l'intervalle, avec quelques creux correspondant aux bancs d'argiles.



SIGNAUX:



TRACES:

1. F280000

2. F280001



3.

4.

5.

INTERVALLE 2840-3020 M/FONT DELTA

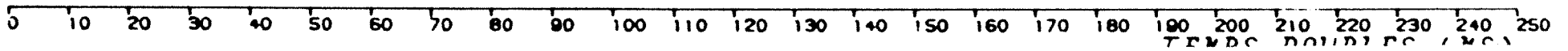
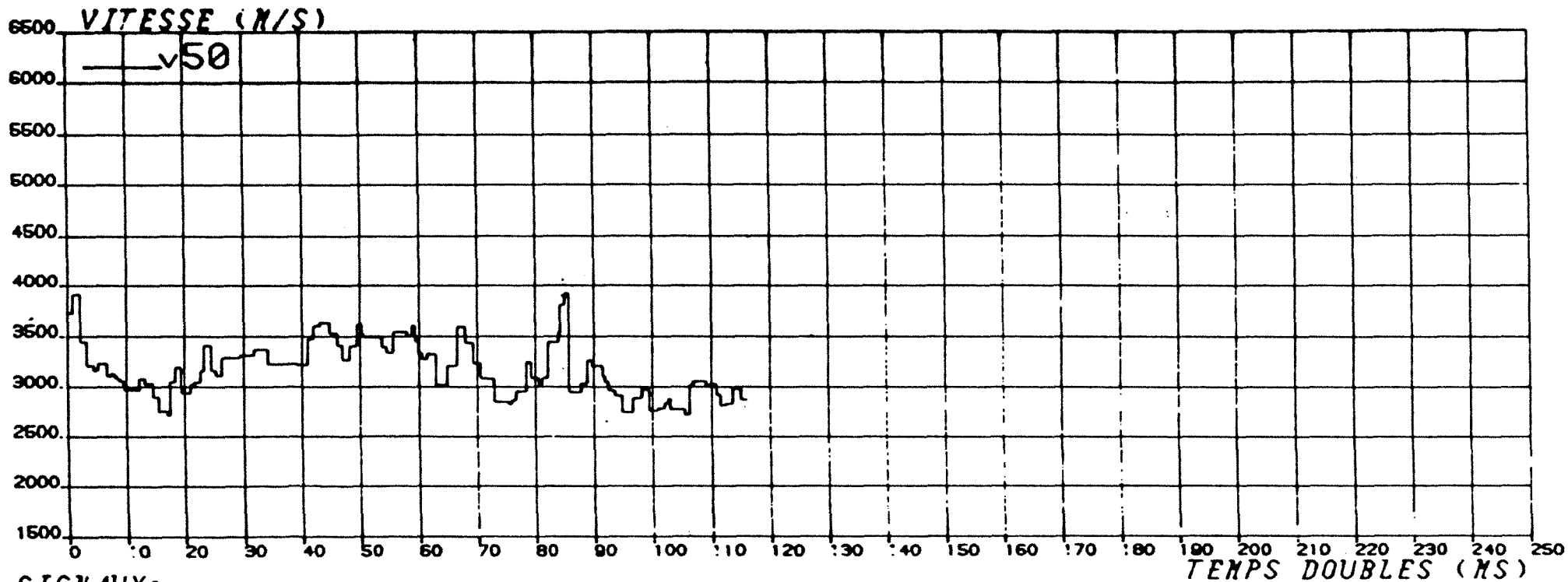


Fig. 52



SIGNAUX:



TRACES:

1. F500000 2. F500001



3.

4.

5.

INTERVALLE 2283-2472 M/FRONT DELTA

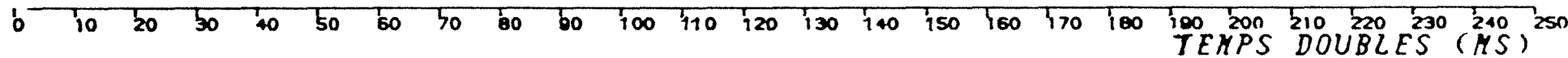
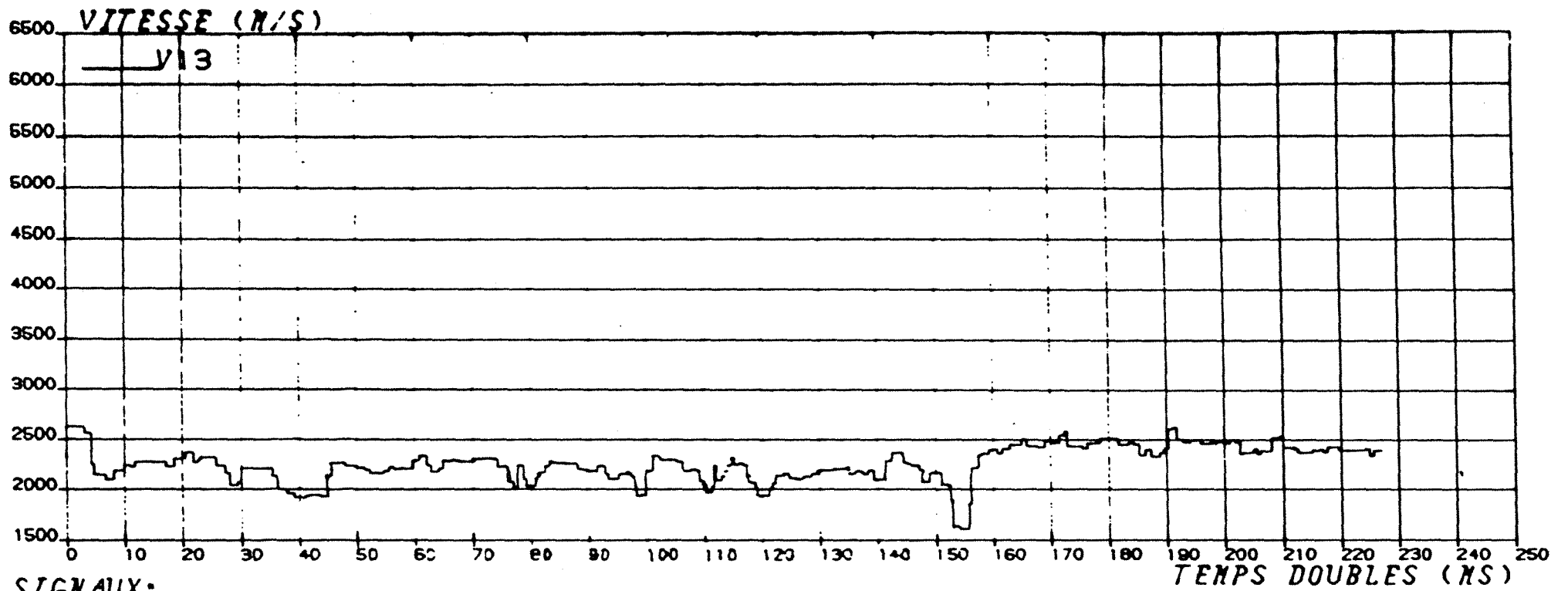


Fig. 53



SIGNAUX:



TRACES:

1. F130000 2. F130001

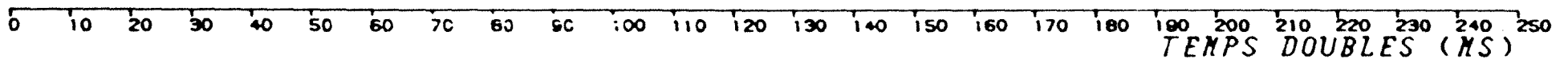


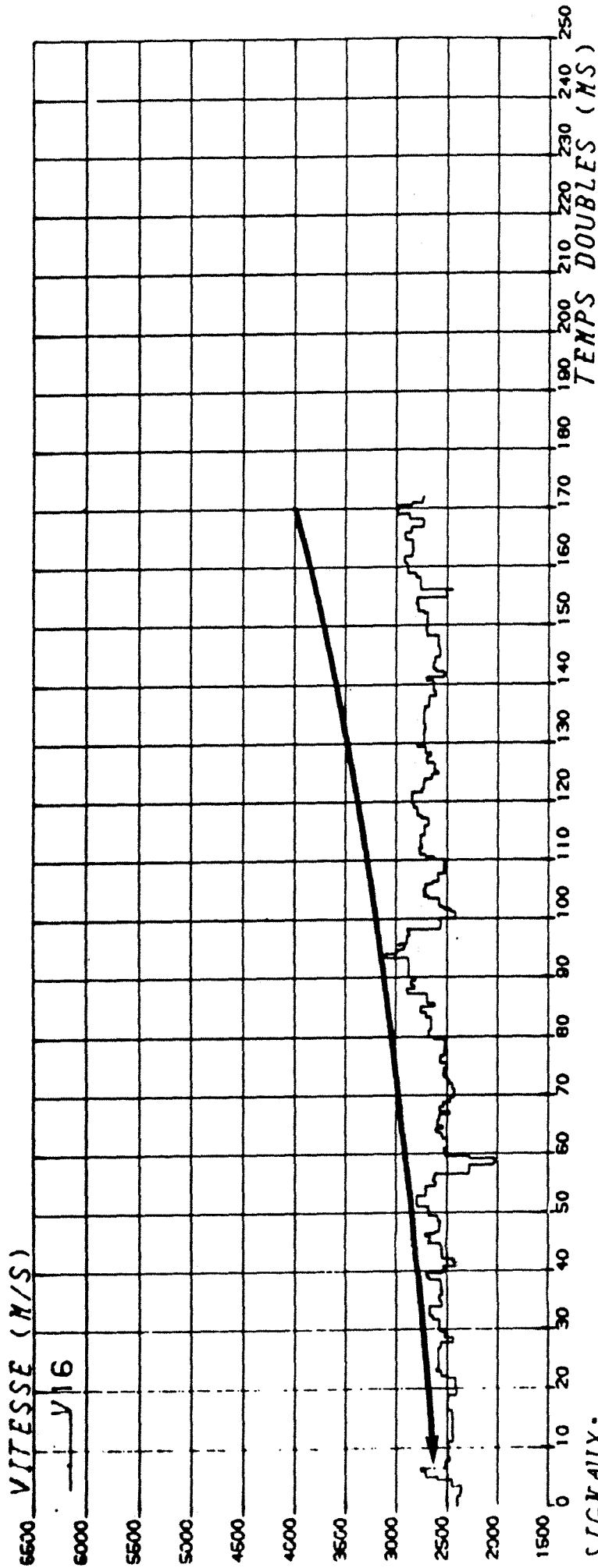
3.

4.

5.

INTERVALLE 920-1185 M/CHENAUX FLUVIALES ET DE MAREE (LEVEES)





SIGNAUX:



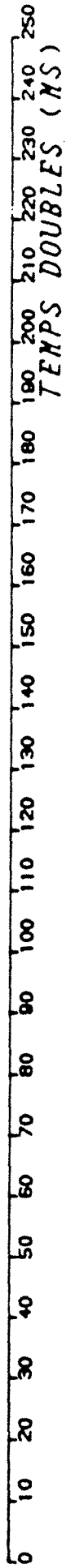
TRACES: 1. F160000 2. F160001



3.

4.

5.



INTERVALLE 1600-1829 M/CHENAUX FLUVIALES (LEVEES)

L'échantillon de la figure 55 présente une certaine évolution de la vitesse, de la base vers le sommet, que l'on peut interpréter comme une augmentation de la consolidation avec la profondeur car ces terrains sont assez superficiels.

3.2.4. Bilan sismique des groupes

Les traces sismiques synthétiques des logs géologiques des trois groupes sont présentées groupe par groupe (Figure 56). Toutes ces traces ont été normées avec l'amplitude maximale enregistrée sur leur ensemble (soit $A_{max} = 2068$), ce qui permet donc de les comparer directement. Les autocorrélations et les spectres de puissance de ces traces ont été calculés. A partir de la simple observation des groupes sismiques, il est difficile de savoir si les groupes sont suffisamment homogènes (i.e. traces semblables) et correctement individualisés par leur spectre de puissance, ou leur autocorrélation (Figures 51, 57 et 58). D'un groupe à l'autre, il existe des différences importantes mais des variations non négligeables apparaissent au sein d'un même groupe.

Nous allons donc chercher à quantifier la morphologie des traces afin de mesurer leur ressemblance (ou dissemblance). Pour cela, nous allons calculer des attributs sismiques à partir des représentations temporelles et fréquentielles des traces.

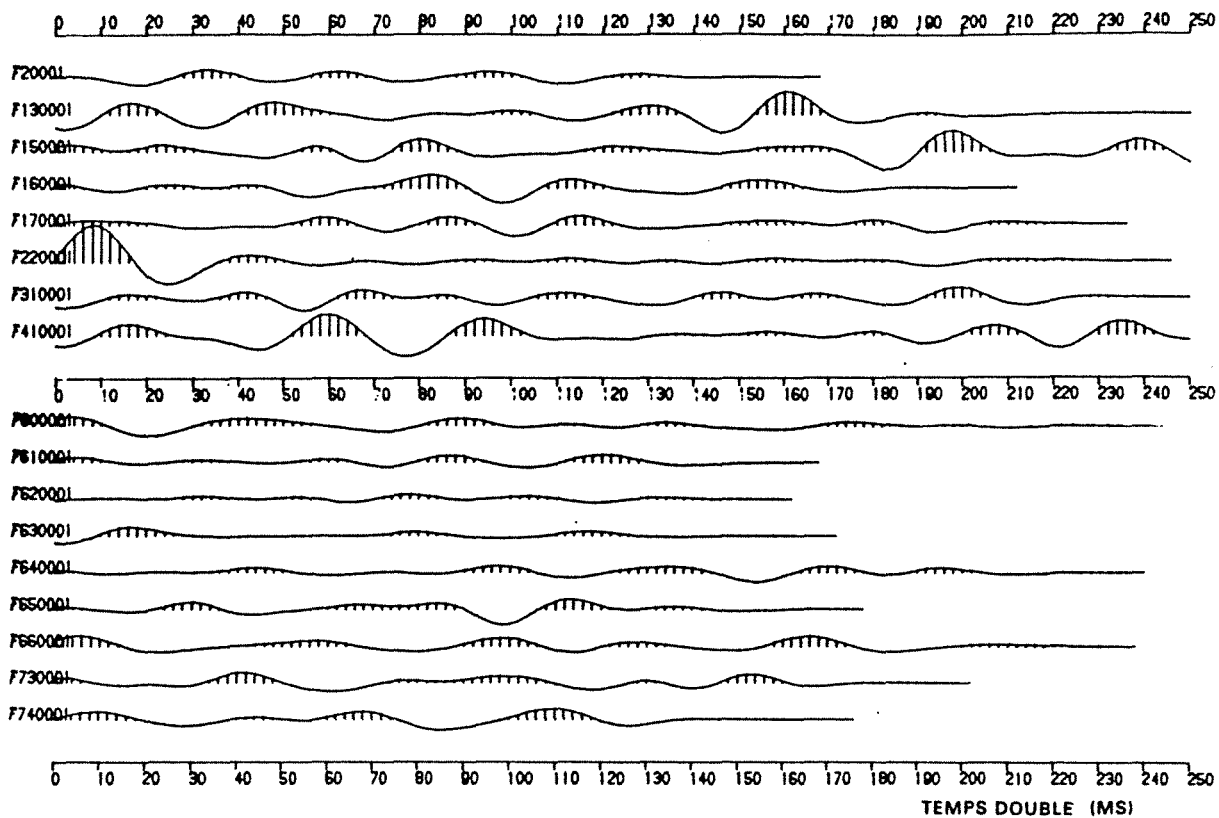
3.3. Calcul de paramètres sismiques pour la caractérisation des traces à partir de leurs représentations temporelles et fréquentielles

Toute trace sismique peut être considérée soit dans le domaine temporel, soit dans le domaine fréquentiel. Les deux domaines sont strictement équivalents mais certaines variables sont plus accessibles dans un domaine que dans l'autre.

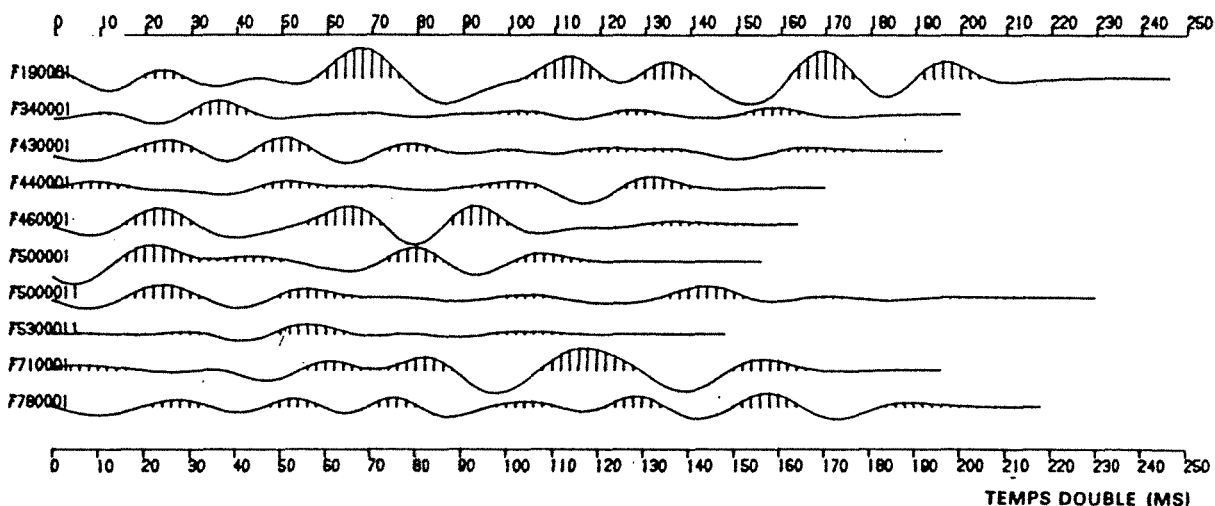
Dans le domaine temporel, on peut considérer la trace $y(t)$ ainsi que son autocorrélation $E[y(t) \cdot y(t + \tau)]$.

Dans le domaine fréquentiel, on dispose des spectres de phase et d'amplitude, calculés sur la transformée de Fourier de la trace ($y(t) = A(t) \sin(\omega t + \varphi)$), ainsi que du signal analytique ($s(t) = y(t) + iz(t)$ où $z(t)$ est la transformée de Hilbert de la trace).

Nous avons vu que toutes les variables possibles peuvent se regrouper en trois familles (paragraphe 1).

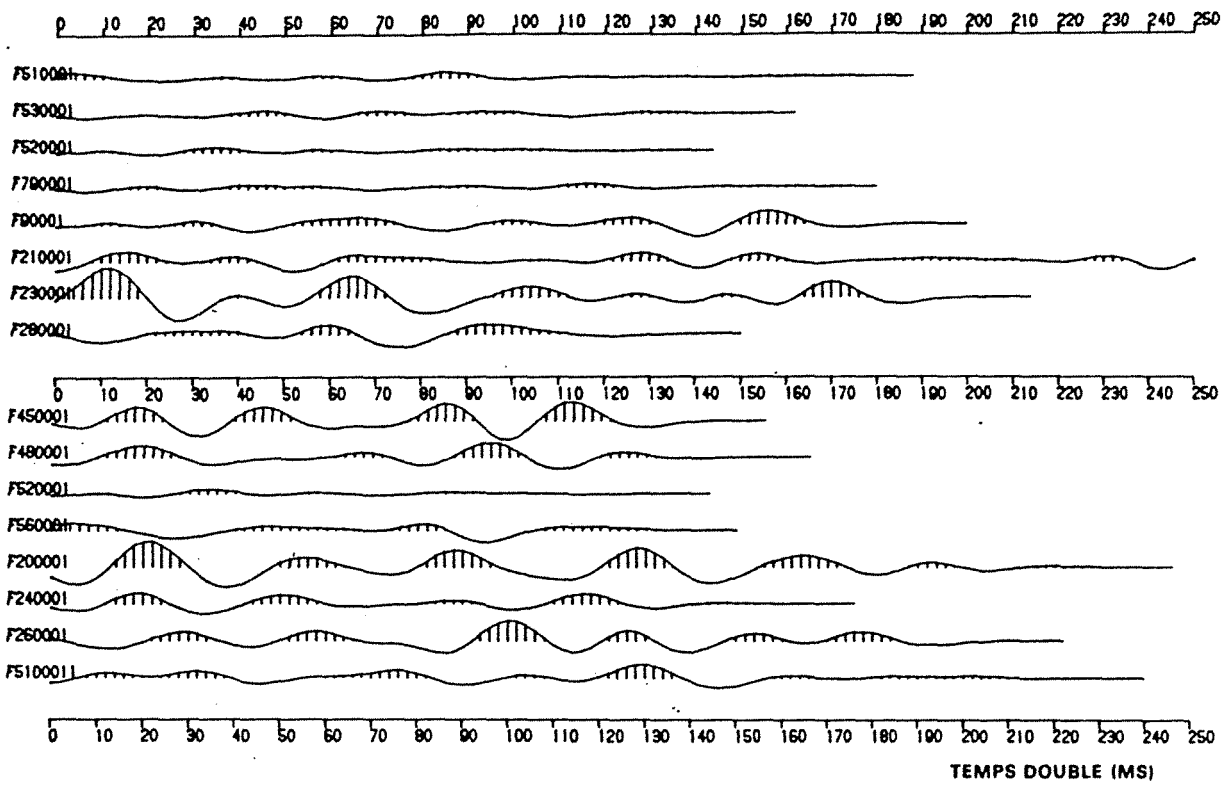


A - 17 traces sismiques calculées sur des intervalles des milieux fluviatiles et margino-littoraux

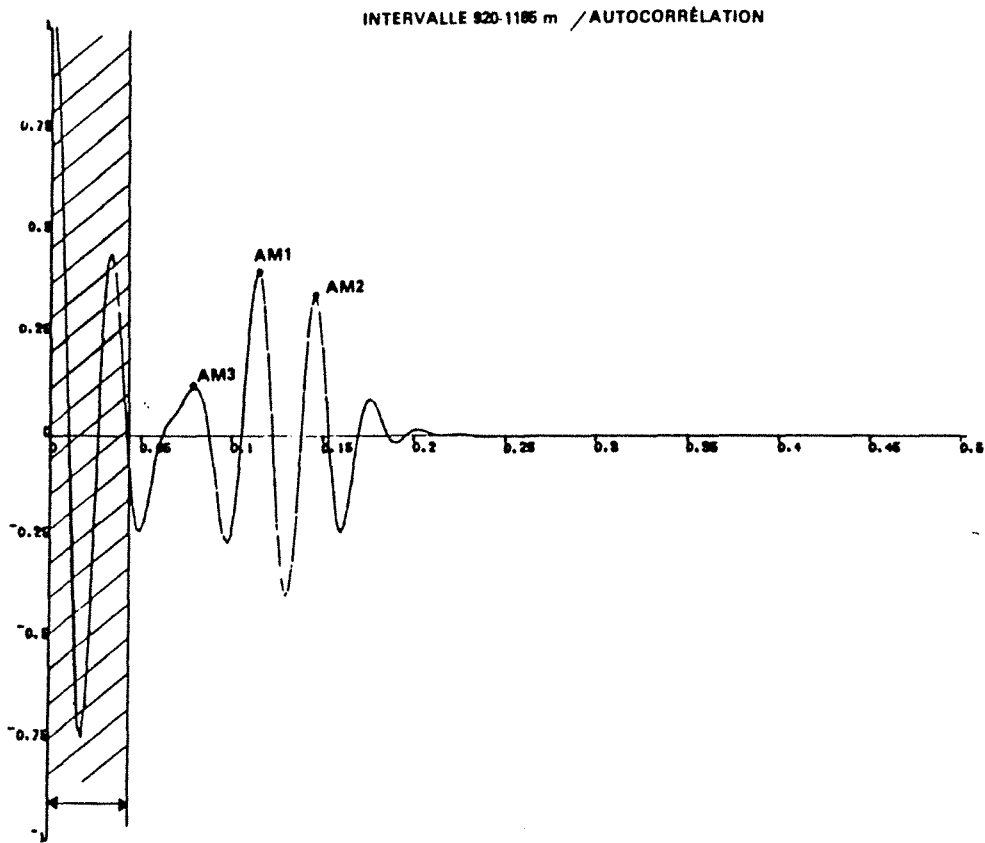


B - 10 traces synthétiques filtrées par S1 calculées pour des échantillons du groupe "Front de Delta"

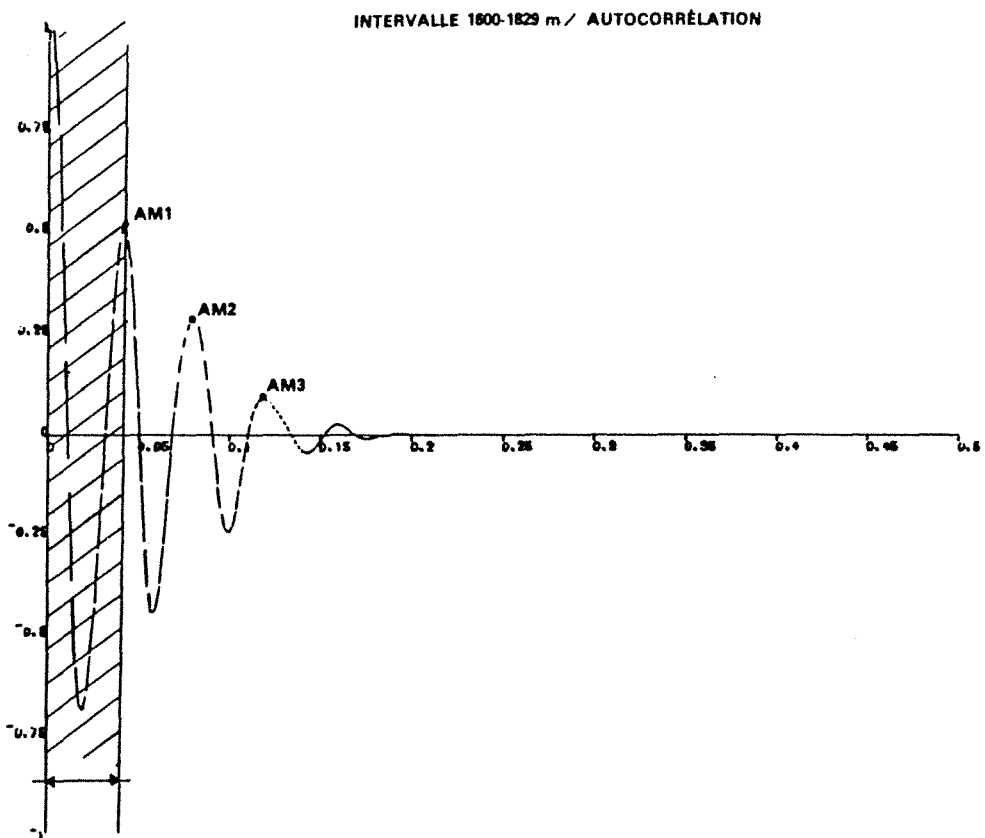
Fig. 56 c



C. traces calculées sur des échantillons du groupe plate-forme. Les 4 traces du haut correspondent aux échantillons de plate-forme externe

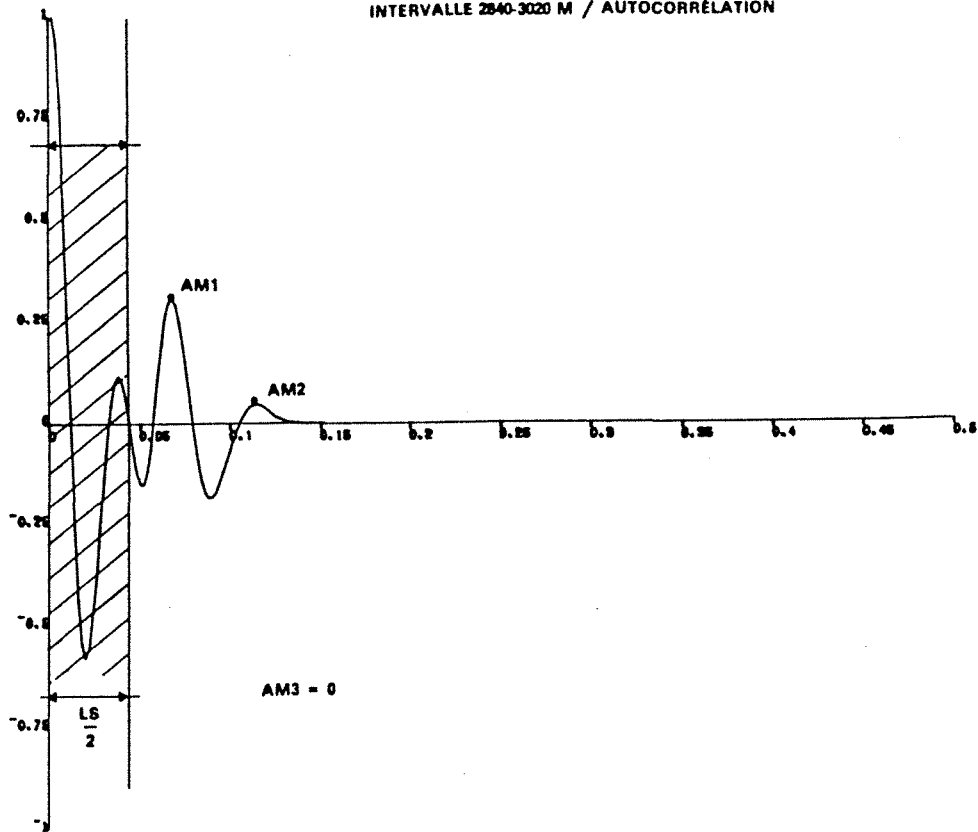


Q. Autocorrélation de la trace filtrée d'un échantillon du groupe « Chenaux et plaine d'inondation »



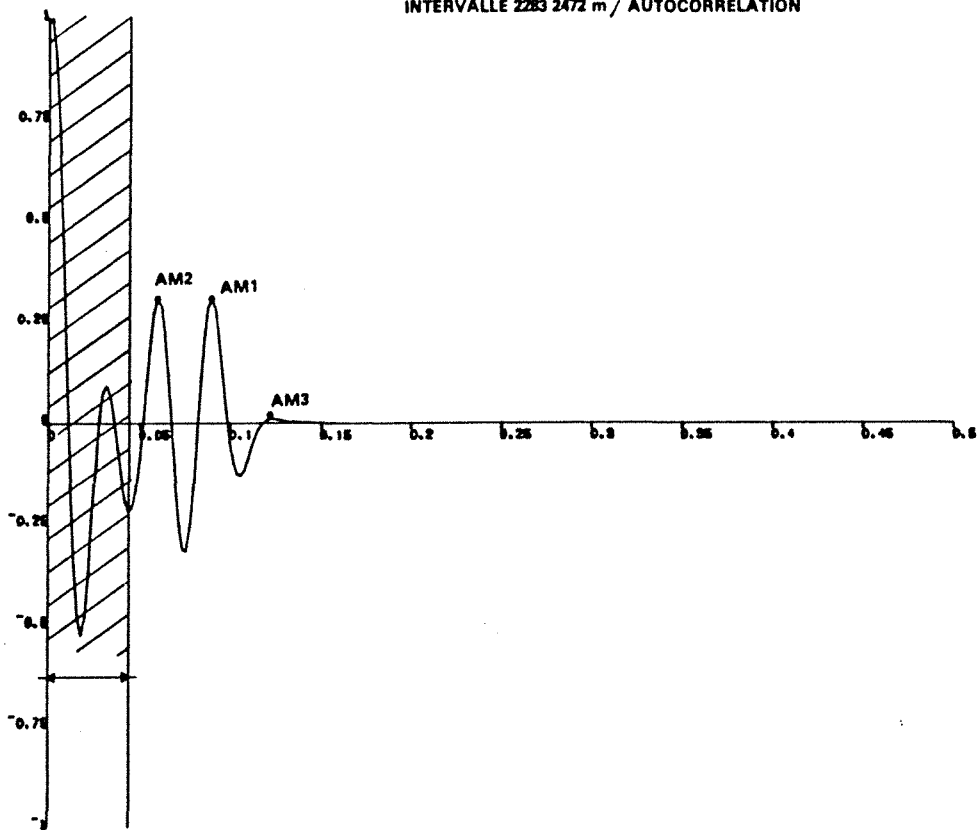
b. Autocorrélation d'une trace appartenant au groupe « Chenaux et plaine d'inondation »

INTERVALLE 2840-3020 M / AUTOCORRÉLATION



a. Autocorrélation d'une trace filtrée appartenant au groupe « Front de Delta »

INTERVALLE 2283 2472 m / AUTOCORRÉLATION



b. Autocorrélation d'une trace filtrée appartenant au groupe « Front de Delta »

Les travaux de Khattri et al. (1976, 1979; 1983, 1984) rattachent l'autocorrélation de la trace à la cyclicité de la sédimentation, et le spectre d'amplitude à l'épaisseur moyenne des bancs, leur fréquence et leur répartition dans l'intervalle étudié.

La troisième famille de variables est liée à la distribution de l'énergie de la trace, que l'on peut quantifier soit directement sur la trace, soit en utilisant le module du signal analytique.

Par conséquent, nous avons calculé des variables sur :

- l'autocorrélation,
- le spectre d'amplitude,
- la trace réelle,
- le module du signal analytique.

3.3.1. Variables calculées sur l'autocorrélation

Classiquement, l'autocorrélation de la trace permet au géophysicien d'approcher le signal et de connaître les multiples éventuels (cf. première partie, paragraphe 4.5). La répartition des coefficients de réflexion, directement liée à la sédimentation, est alors considérée comme aléatoire, à condition de la calculer sur une fenêtre de grande durée.

Or, du point de vue géologique, une mégaséquence est toujours constituée d'un enchaînement de séquences d'ordre inférieur qui sont elles-mêmes décomposables. Ainsi, sur une fenêtre relativement courte, la sédimentation est cyclique et on ne peut plus considérer comme aléatoire la répartition des coefficients de réflexion (cf. remarques, paragraphe 2.3.7.). Les rebonds de l'autocorrélation sont alors liés à la répétitivité des séquences. Tous nos films synthétiques ayant été calculés sans multiple, les rebonds éventuels reflèteront donc une information purement géologique.

Sur la figure 51 est représentée l'autocorrélation de l'échantillon dont on a reproduit la trace synthétique, Figure 50. On a reporté sur la figure 51 les points particuliers qui ont été pris en compte sur l'autocorrélation.

Ce sont les deux ou trois premiers rebonds, retenus en fonction de leur amplitude (pour la trace filtrée, on a supprimé la partie de l'autocorrélation qui correspond à l'autocorrélation du signal) et leurs temps correspondants ainsi que les premiers passages à zéro.

A priori, ce sont les paramètres mesurés sur le rebond le plus important qui vont refléter une cyclicité réelle. Notamment, l'amplitude du premier rebond devrait être particulièrement significative pour opposer une sédimentation cyclique à une sédimentation acyclique ; dans ce dernier cas, l'amplitude du rebond le plus important devrait être très faible.

A titre de comparaison, les figures 57 et 58 montrent l'autocorrélation obtenue sur les traces filtrées des échantillons de front de delta, chenaux et plaine d'inondation, présentés dans les paragraphes précédents. Dans tous les cas, l'origine du temps a été fixée au niveau du maximum d'amplitude de l'autocorrélation. Les temps sont exprimés en millisecondes (ms), temps double.

Les variables calculées sur l'autocorrélation sont au maximum au nombre de dix.

3.3.2. Variables sur le spectre d'amplitude

Nous n'avons pas cherché à extraire une information géologique du spectre de phase car la rotation très rapide de la phase semble délicate à interpréter.

Il est évident que dans les hypothèses classiques de la géophysique où l'on suppose une répartition aléatoire des coefficients de réflexion sur une longue fenêtre, toutes les distances entre réflecteurs étant possibles et équiprobables, toutes les fréquences devront être semblablement représentées dans le spectre d'amplitude, calculé sur un film synthétique impulsionnel.

Par contre, sur des courtes fenêtres, la répartition des épaisseurs des bancs va marquer sur le spectre d'amplitude :

- les bancs épais correspondent aux basses fréquences,
- les bancs minces correspondent aux hautes fréquences.

Les fréquences pour lesquelles on observe de fortes amplitudes correspondent à l'épaisseur des bancs statistiquement dominants.

Dans le cas d'une trace synthétique (film synthétique impulsionnel filtré), les bancs dont les épaisseurs correspondent à des fréquences hors de la bande passante sismique n'apparaîtront

pas. Ce seront les bancs très épais et les bancs très minces ; en plus, les amplitudes des fréquences situées sur les pentes du filtre seront artificiellement diminuées.

Plutôt que d'utiliser le spectre d'amplitude tel quel (Figure 59a), nous avons préféré utiliser la courbe des amplitudes cumulées en fonction des fréquences. Cette courbe est présentée, Figure 59b, pour un des échantillons du groupe plate-forme. A partir de cette courbe, nous avons évalué les fréquences correspondant aux déciles de la distribution des amplitudes (Figure 59b).

On a recalculé une courbe analogue en pondérant les amplitudes par les fréquences et déterminé, à partir de cette courbe, de nouvelles variables : les fréquences correspondant aux déciles de la distribution des amplitudes pondérées (Figure 60).

Le fait de se servir de ces courbes cumulatives plutôt que de calculer directement les variables sur le spectre d'amplitude revient à effectuer une intégration de l'information contenue dans le spectre d'amplitude. On peut ainsi éviter de prendre un pic qui ne correspondrait qu'à du bruit, puisqu'on lisse le spectre.

On a tout de même retenu, comme variable directement calculée sur le spectre d'amplitude, la fréquence f_M correspondant à l'amplitude maximale (Figure 59a).

3.3.3. Variables sur la trace réelle

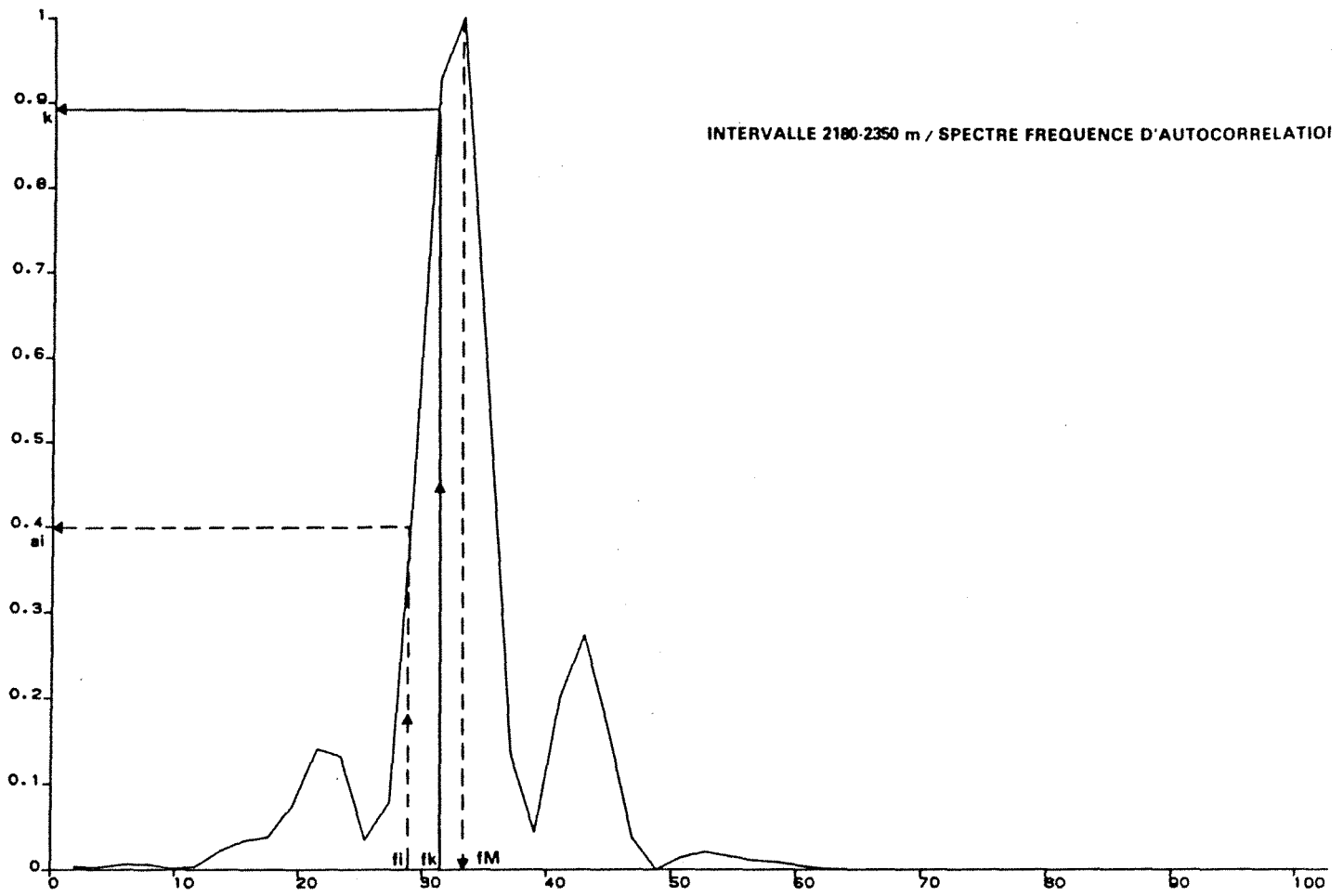
Nous n'avons pas encore abordé la quantification de la réflectivité de la série. En effet, une série homogène, sans contraste (empilement de chenaux, argiles de bassin) aura beaucoup moins de relief que des alternances argiles-sables-carbonates de barrière.

D'autre part, il est aussi intéressant de connaître la répartition de l'énergie sur un intervalle, l'énergie distribuée en tête de séquence, plutôt qu'en queue, donnera peut-être une signification géologique différente à cette séquence.

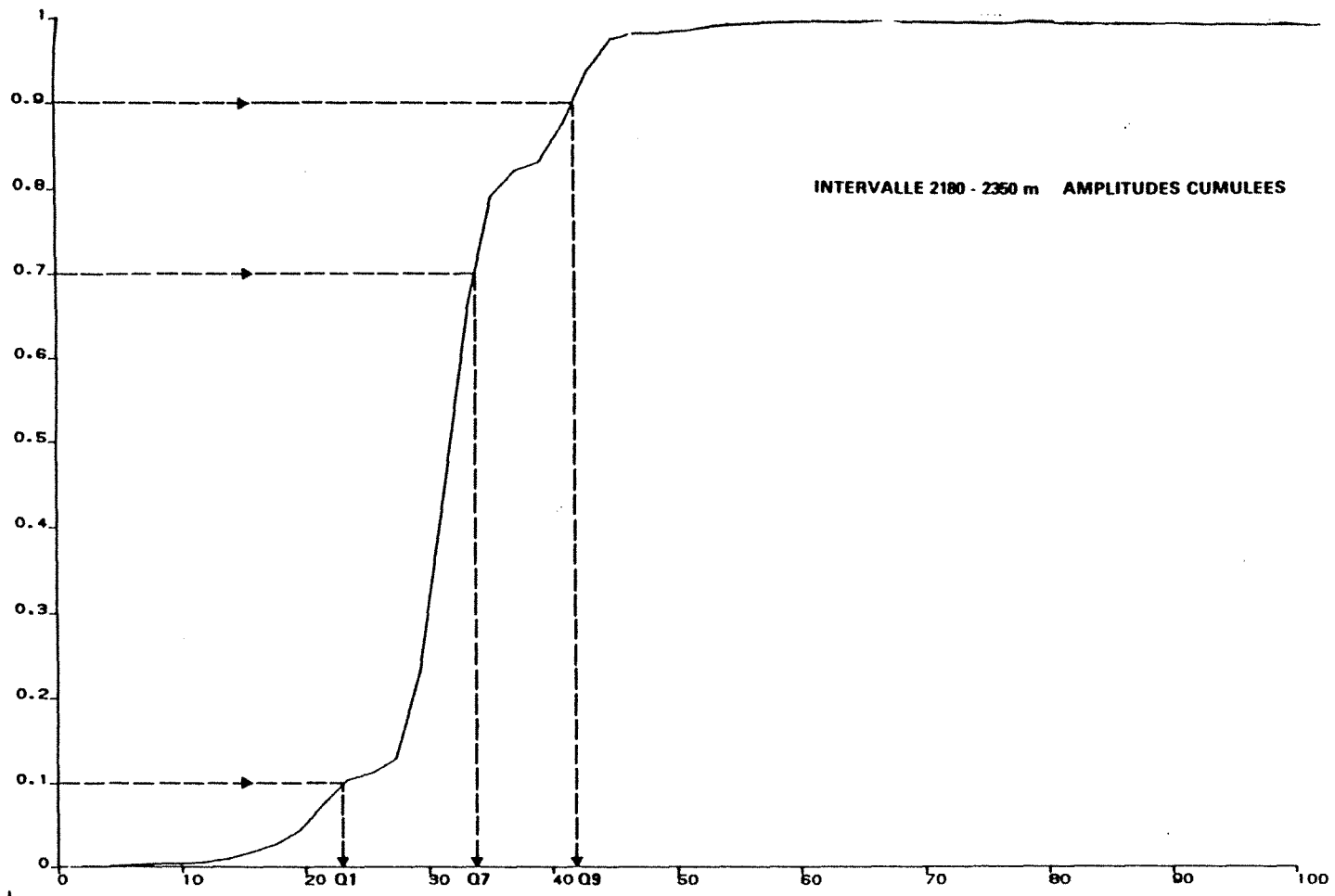
Pour pouvoir comparer la répartition de l'énergie sur des intervalles de longueur, en temps double, différente, on a réalisé une homothétie de l'échelle temps pour se ramener à un intervalle de longueur 100 ms, quel que soit l'échantillon (Figure 61).

On a effectué un redressement de la trace et cumulé les valeurs absolues des amplitudes ; puis on a estimé les temps correspondant aux déciles de cette distribution (figure 61).

Fig. 59a



A - Spectre d'amplitude de l'autocorrélation d'un échantillon du groupe "Plate-forme"



b - Courbe représentant les amplitudes cumulées en fonction des fréquences $Q_i : i = 1, \dots, 9$ Fréquences correspondant aux valeurs d'amplitudes cumulées de 10 %, 20 %, ... 90 %

Soit (f_i, a_i) les coordonnées d'un point sur le spectre d'amplitude

Le k^{ème} point de la courbe

aura pour coordonnées : $X : X_k = f_k - Y : Y_k = \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^n a_i}$

Fig. 59 b

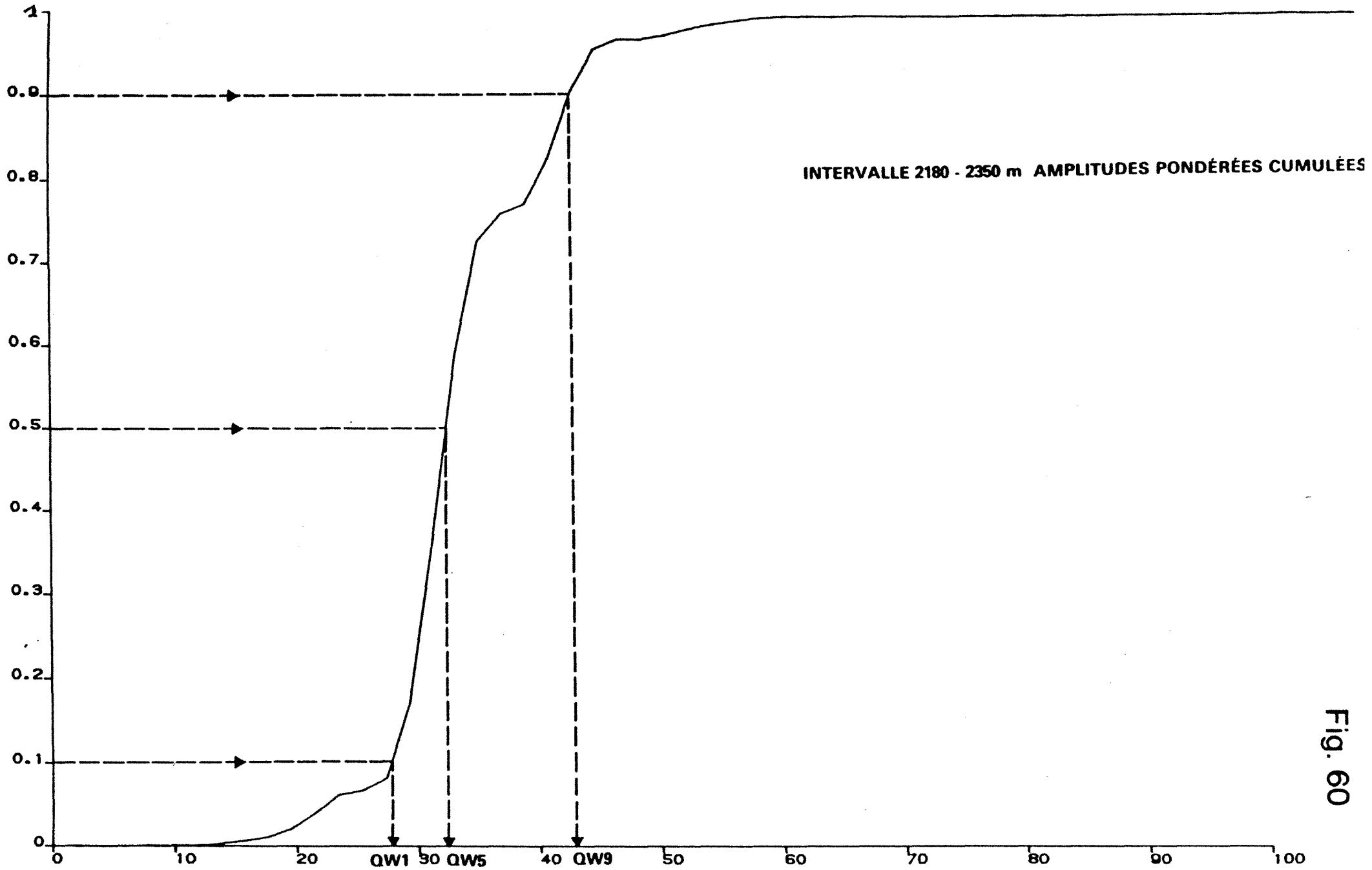


Fig. 60

Courbe représentant les amplitudes pondérées par les fréquences et cumulées en fonction des fréquences $QW_i \quad i = 1, \dots, 9$ Fréquences correspondant aux valeurs d'amplitudes pondérées cumulées de 10 %, 20 %, ... 90 %

Soit (f_i, a_i) les coordonnées d'un point sur le spectre d'amplitude (Fig 59)

$\sum_{i=1}^k a_i f_i$

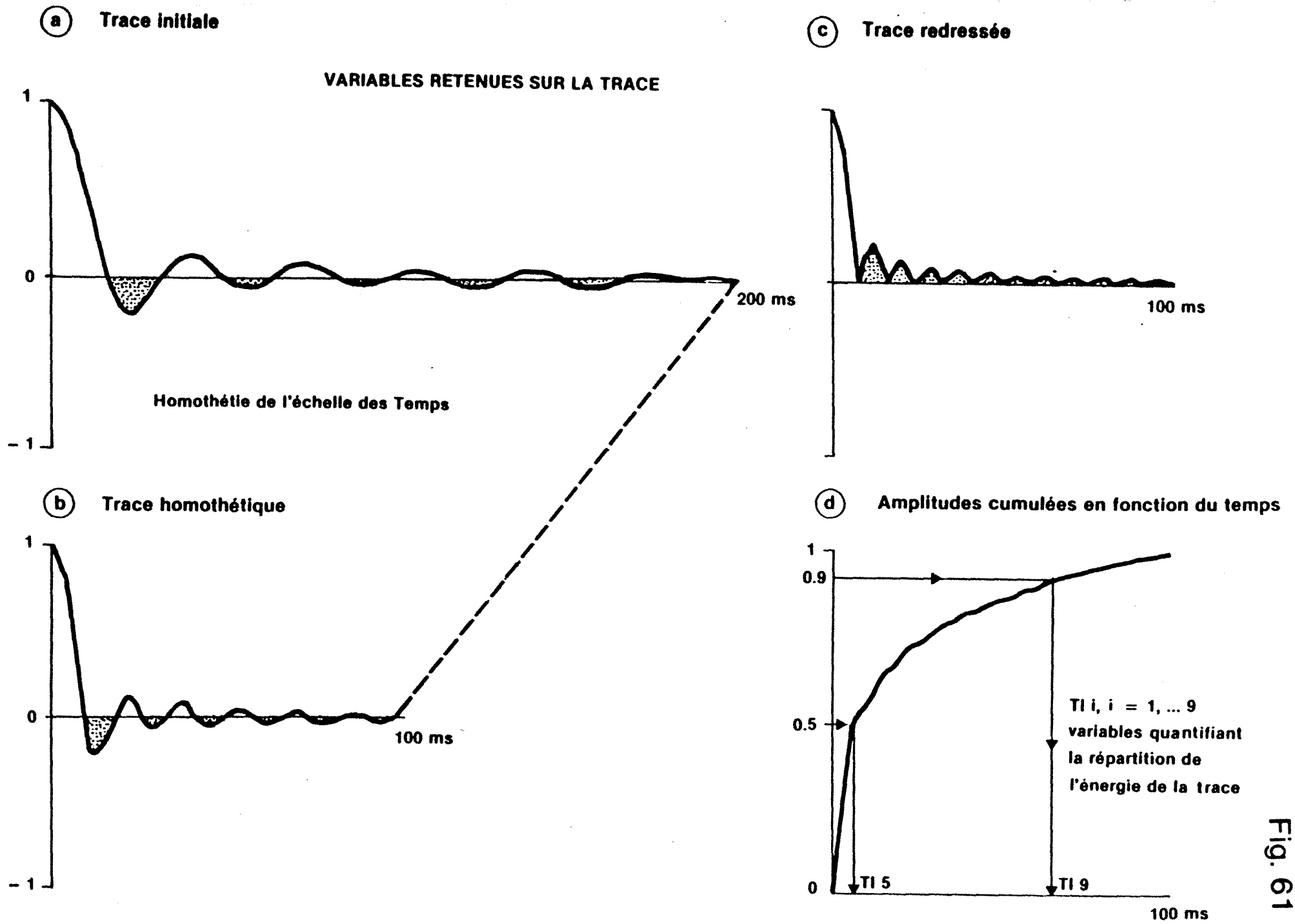


Fig. 61

On peut ainsi évaluer la répartition de l'énergie sur l'intervalle temps.

3.3.4. Variables sur le signal analytique

On a calculé la transformée de Hilbert de la trace sismique, puis la trace complexe ; nous avons utilisé son module.

Soit $s(t)$ la trace complexe :

$$s(t) = y(t) + iz(t)$$

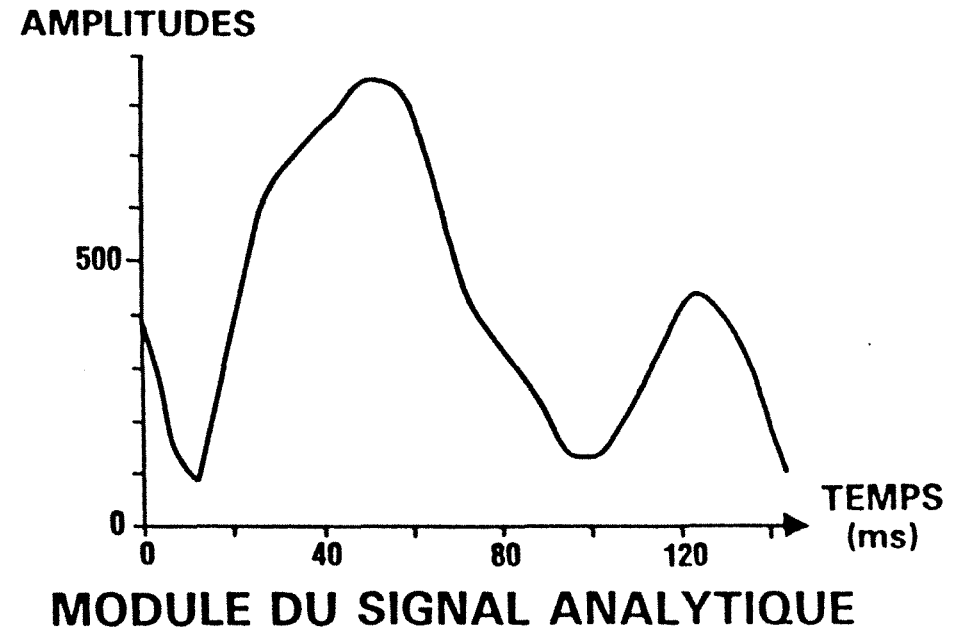
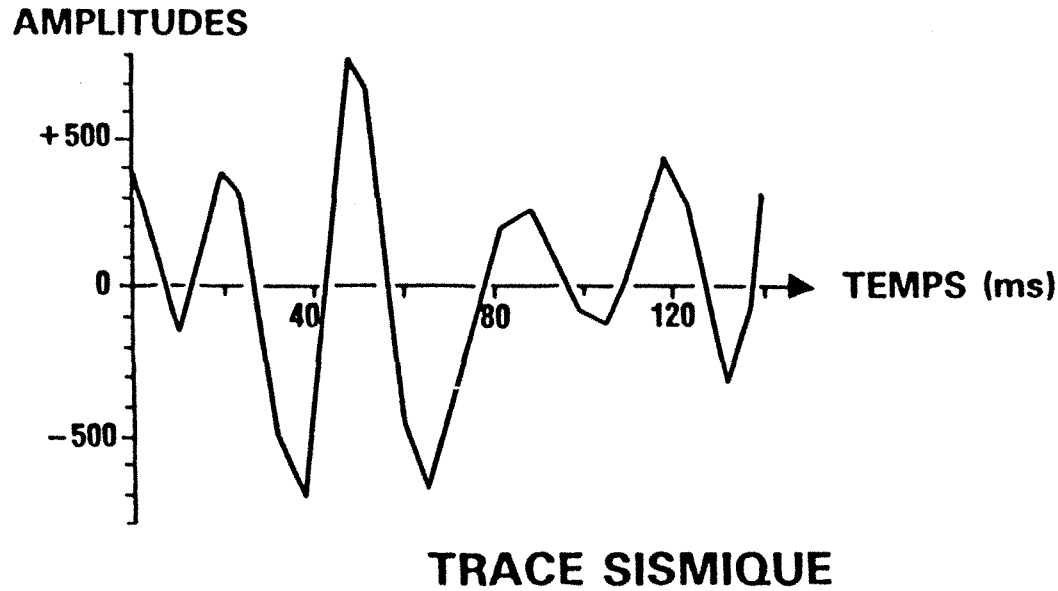
où y est la trace
 z sa transformée de Hilbert.

Le module complexe est $E(t) = \sqrt{y^2 + z^2}$.

Le module complexe d'un échantillon du groupe plate-forme est représenté sur la figure 62. Nous avons calculé l'intégrale de cette courbe qui permet de connaître la racine carrée de l'énergie totale de l'intervalle (variable ETNT).

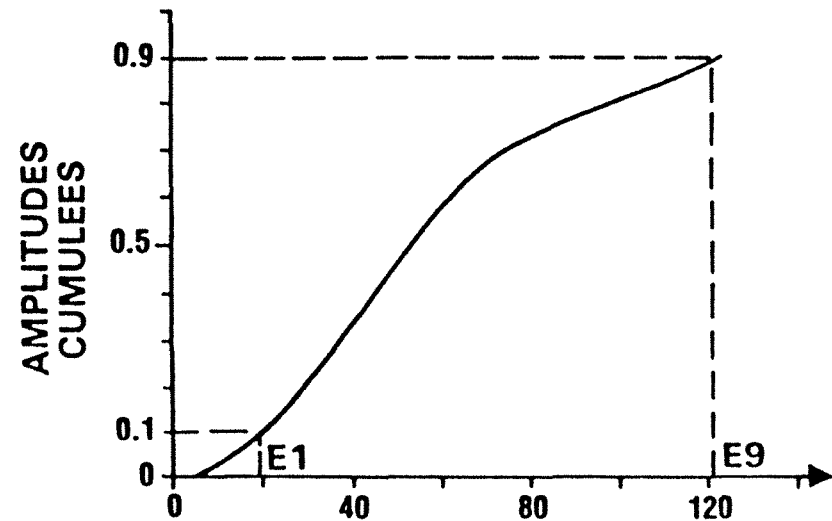
Nous avons ensuite, comme pour la trace réelle, cumulé les amplitudes du module et construit la courbe des amplitudes cumulées en fonction du temps. Les variables retenues sont les temps correspondant aux déciles de cette distribution (Figure 62). Comme dans le cas de la trace, ces variables nous permettent d'approcher la répartition temporelle de l'énergie.

PARAMETRES CALCULES SUR LE MODULE DU SIGNAL ANALYTIQUE



**E1 : TEMPS CORRESPONDANT
A 10 % DE L'ENERGIE TOTALE**

**E9 : TEMPS CORRESPONDANT
A 90 % DE L'ENERGIE TOTALE**



3.3.5. Récapitulation

Le tableau 8 récapitule l'ensemble des variables calculées qui caractérisent la morphologie des traces et sont à relier à la cyclicité de la sédimentation, la distribution de l'épaisseur des bancs et les contrastes acoustiques de la série.

FAMILLE	DESCRIPTION	SYMBOLE
AUTOCORRELATION	Rebonds principaux de l'autocorrélation rangés par amplitude décroissante	AM_i $i = 1, \dots 3$
	Temps correspondant aux rebonds principaux	TAM_i $i = 1, \dots 3$
	Premiers temps des passages à zéro	TPZ_i $i = 1, \dots 4$
SPECTRE	Fréquence correspondant au pic d'amplitude du spectre d'amplitude	FM
	Fréquences correspondant aux déciles de la distribution des amplitudes	$Q_i, i = 1, \dots 9$
	Fréquences correspondant aux déciles de la distribution des amplitudes pondérées par les fréquences	$QW_i, i = 1, \dots 9$
MODULE DU SIGNAL ANALYTIQUE	Aire comprise sous la courbe du module	ETNT
OU	Temps correspondant aux déciles de la distribution des amplitudes du module normé	$E_i, i = 1 \dots 9$
TRACE	Temps correspondant aux déciles de la distribution des valeurs absolues des amplitudes de la trace	$T_i, i = 1, \dots 9$

Tableau 8 : variables calculées sur les traces sismiques

Nous allons maintenant étudier si ces variables permettent d'individualiser les groupes de traces correspondant aux différents paléo-environnements.

3.4. Pouvoir séparateur des variables

Le pouvoir séparateur des variables a été testé essentiellement en réalisant des **analyses factorielles discriminantes** entre les groupes de traces des divers environnements.

Il est apparu très rapidement que les variables calculées sur la trace "redressée" (paramètres T_i) quantifiaient la même information que les variables du module du signal analytique (paramètres E_i). Prendre les valeurs absolues de la trace est aussi une des façons d'estimer une "enveloppe" de la trace. Mais l'enveloppe donnée par le signal analytique est moins sensible au bruit. C'est pourquoi nous avons conservé les paramètres E_i pour quantifier l'énergie de la trace et sa distribution, au détriment des variables T_i .

3.4.1. Analyses discriminantes sur 43 traces les plus typiques des divers groupes

Nous avons d'abord testé le pouvoir discriminant des variables en ne faisant appel qu'à une seule famille à la fois.

Les 19 variables calculées sur le **spectre d'amplitude** de l'autocorrélation permettent un certain regroupement des individus, en fonction de leur faciès, mais les barycentres des classes sont très proches, le nuage projeté très compact (Figure 63).

Les 10 variables calculées sur l'**autocorrélation** de la trace ne permettent qu'une opposition, grâce à l'axe 1 dont le pouvoir discriminant est très fort, entre les individus du faciès plateforme et les autres. Les groupes barrière et chenaux sont extrêmement mélangés, l'axe 2 permettant seulement une très faible discrimination (Figure 64).

Les 10 variables quantifiant l'énergie, énergie totale du module et les déciles de la distribution des amplitudes du module, n'ont permis pratiquement aucune séparation des groupes de faciès (Figure 65).

Une analyse factorielle discriminante classique avec les variables des **trois familles** permet une excellente séparation des trois groupes de faciès (Figure 66). Cependant, la fiabilité de ces résultats est douteuse car le nombre de variables utilisées, par rapport au nombre d'individus, est trop élevé.

ANALYSE FACTORIELLE DISCRIMINANTE

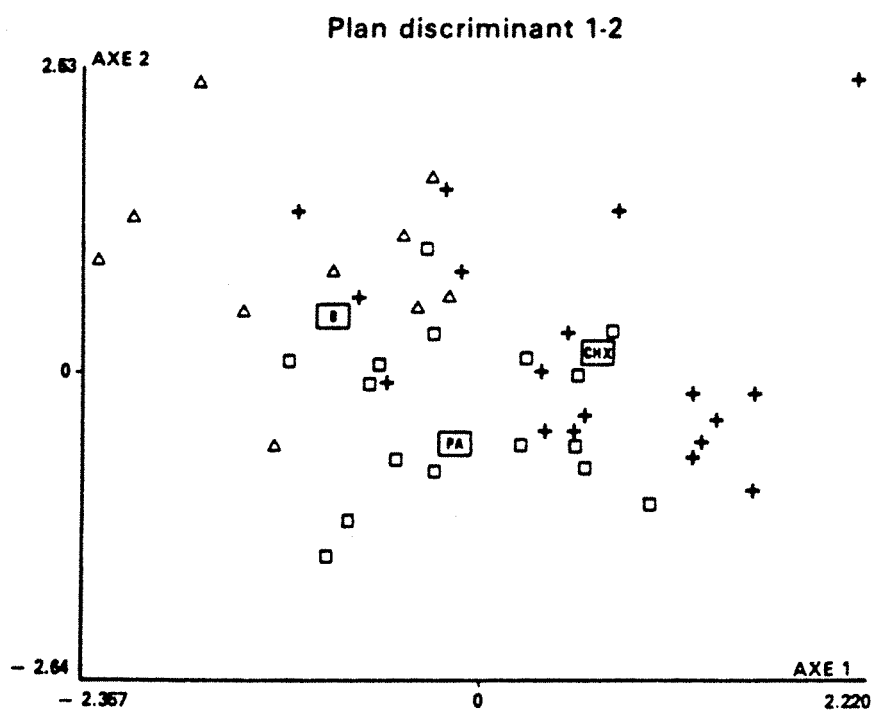
Traces Synthétiques

43 Échantillons

3 Groupes

19 Variables FM, Qi,i = 1,...9, QWi,i = 1,...9

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.598	0.598
2	0.402	1.000



ANALYSE FACTORIELLE DISCRIMINANTE

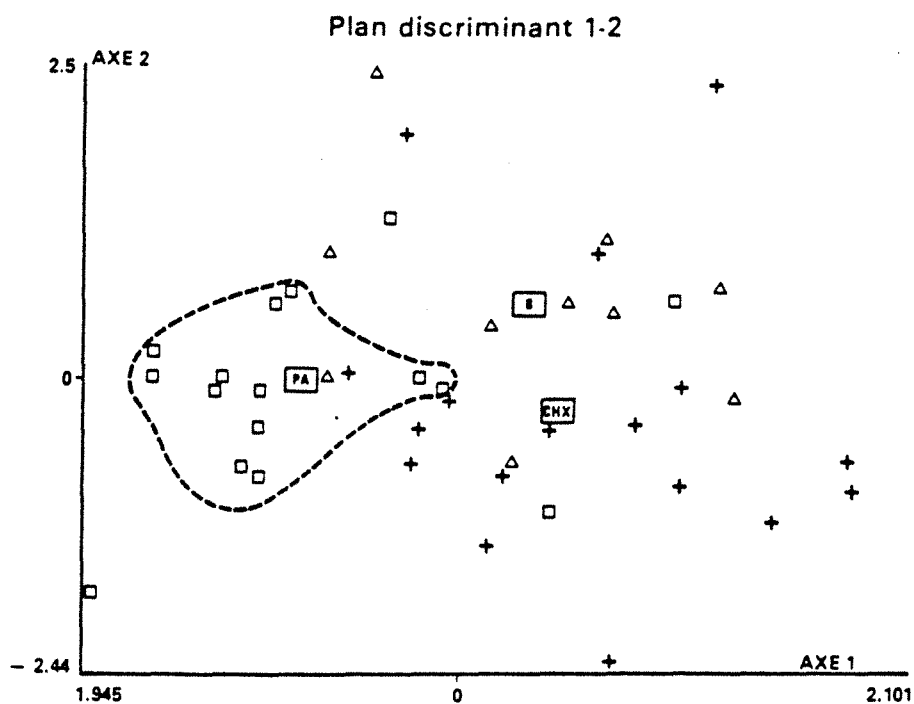
Traces Synthétiques

43 Échantillons

3 Groupes

10 Variables $AM_{i,j} = 1,2,3$, $TAM_{i,j} = 1,2,3$, $TPZ_{i,j} = 1,\dots,4$

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.761	0.761
2	0.239	1.000



ANALYSE FACTORIELLE DISCRIMINANTE

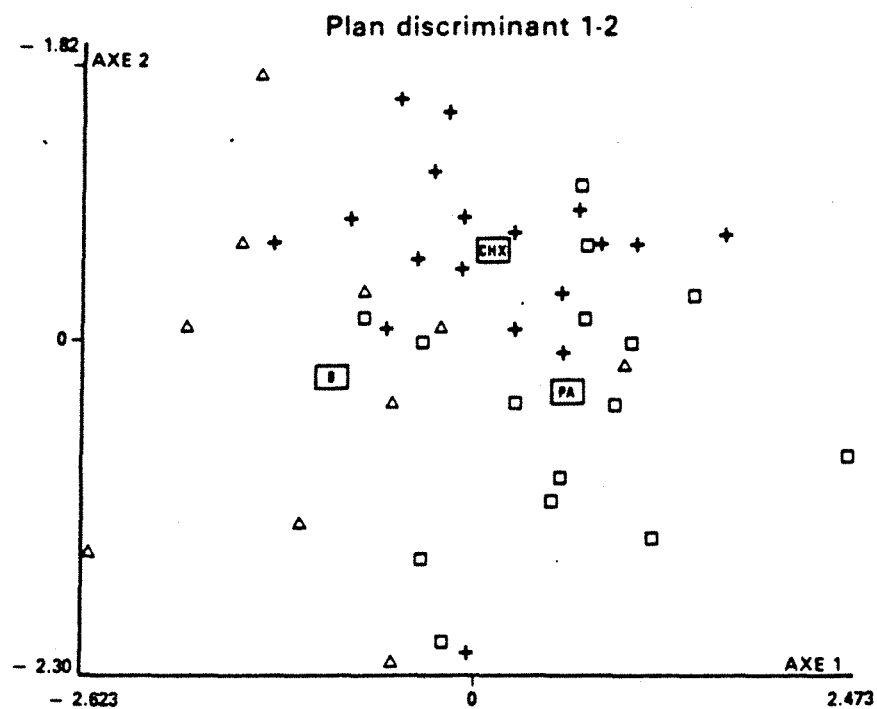
Traces Synthétiques

43 Échantillons

3 Groupes

10 Variables ETNT, E_i , i = 1,...,9

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.663	0.663
2	0.337	1.000



ANALYSE FACTORIELLE DISCRIMINANTE

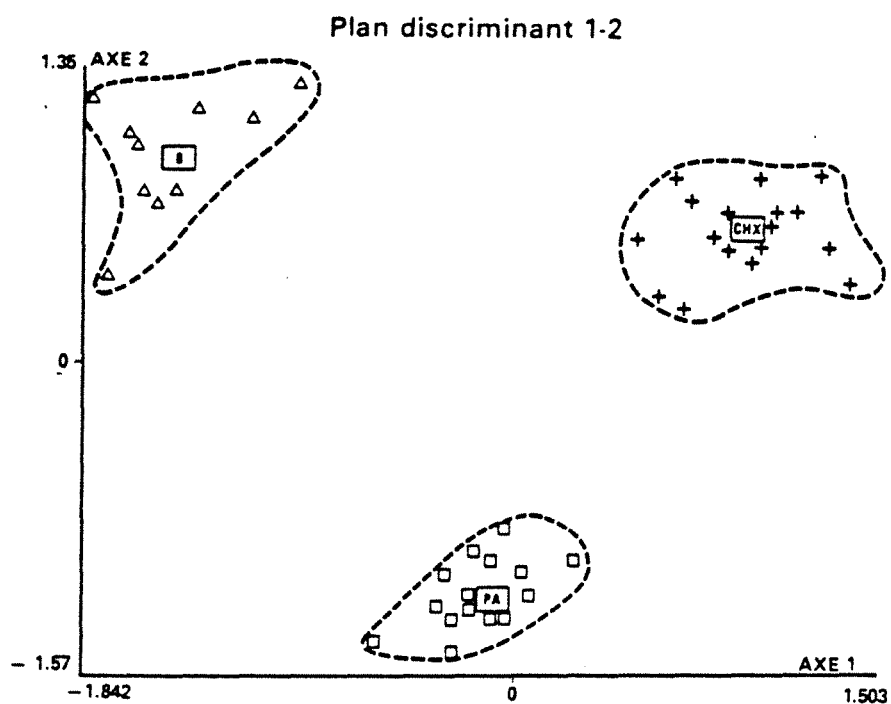
Traces Synthétiques

43 Échantillons

3 Groupes

39 Variables FM, Qi, i = 1, ... 9, QWi, i = 1, ... 9, TPZ i, i = 1, ... 4, AMi, i = 1, 2, 3, TAMi, i = 1, 2, 3, Ei, i = 1, ... 9, ETNT

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.497	0.497
2	0.503	1,000



3.4.2. Recherche des variables les plus discriminantes (test de Student)

La fiabilité de la discrimination sera meilleure si on considère un petit nombre de variables par rapport au nombre d'individus. Dans le cas de deux groupes d'individus, par exemple, on trouvera plus facilement un hyperplan séparateur dans un espace de dimension élevée que dans un espace de faible dimension.

Cependant, il existe très peu d'études permettant de définir, pour un nombre donné de variables et un nombre donné de groupes, à partir de combien d'individus la discrimination sera significative. Il est évident que, pour un nombre donné de variables, plus il y aura de groupes, plus il faudra d'individus pour obtenir une discrimination qui ait un sens statistique. Romeder (1973), dans le cadre d'hypothèses sur la probabilité d'existence d'une dichotomie aléatoire, a proposé une relation entre le nombre de variables et le nombre d'individus, dans le cadre de **deux groupes** et pour un risque de 1 ou de 5 %. Ainsi, avec 43 individus, pour le risque de 1 % et deux groupes, il faut utiliser au plus 14 variables. Avec trois groupes, il faudra en utiliser encore moins. Avec 88 individus et deux groupes, au risque de 1 %, on pourrait utiliser jusqu'à 32 variables.

La relation proposée par Romeder suppose, d'une part, qu'il y a équiprobabilité a priori des deux classes et que, d'autre part, on a pu aboutir à une séparation totale des deux groupes.

Dans la réalité, on en sépare seulement un fort pourcentage, de l'ordre de 90 %. La généralisation à plus de deux groupes est tout de suite beaucoup plus complexe : il faudrait considérer le nombre de configurations séparables par trois hyperplans concourants.

Cette relation présente toutefois l'avantage de préciser les ordres de grandeur du nombre de variables et du nombre d'échantillons afin que la discrimination soit significative.

Le test de Student est un test paramétrique de signification, c'est-à-dire qu'il cherche à vérifier si un écart entre un même paramètre mesuré sur deux échantillonnages de population différents est significatif, à un seuil de signification donné.

Soient n_1 , n_2 , les tailles des populations étudiées (ici, ce sont les deux groupes constitués, par exemple, par les individus de faciès plate-forme et de faciès front de delta).

Soient m_1 , m_2 , les moyennes des variables mesurées sur chaque groupe, et s_1 , s_2 les écarts-types correspondants.

$$d = m1 - m2$$

$$s^2 = \frac{n1 s1^2 + n2 s2^2}{n1 + n2 - 2}$$

$$sd = s \sqrt{\frac{n1 + n2}{n1 n2}}$$

$$t = \frac{d}{sd} .$$

Les effectifs étudiés sont de petite taille (c'est notre cas) ; alors, à condition de considérer les populations parentes normales et de même variance, on démontre que t suit une loi de Student avec $\nu = n1 + n2 - 2$ degrés de liberté. On détermine alors t et l'intervalle d'acceptation $[-t_\alpha ; t_\alpha]$, c'est-à-dire l'intervalle dans lequel doit être situé le paramètre t calculé dans le test, pour que l'on accepte l'hypothèse nulle, à savoir que les écarts entre les deux populations testées sont significatifs.

Les intervalles d'acceptation ont été calculés pour les variables mesurées sur les groupes pris deux par deux et pour un risque de première espèce de 5 %.

Les variables les plus significatives sont les suivantes : Q1, Q2, Q6, Q7, Q9, QW4, QW5, QW6, AM2, ETNT, soit 10 variables au total.

Des analyses factorielles discriminantes portant sur ces variables ont été refaites, soit sur la population des 43 traces synthétiques déjà étudiées, soit sur celle de toutes les traces synthétiques disponibles (88).

La régionalisation du plan factoriel en trois domaines qui est obtenue (Figure 67) est assez satisfaisante malgré un certain recouvrement des groupes.

Ces diverses analyses, effectuées après des tests de Student sur les variables, montrent qu'il est encore possible de séparer les trois groupes principaux de faciès (plate-forme, front de delta, margino-littoral) avec un nombre de variables **très réduit** (Figure 67), mais représentant **toutes** les familles de paramètres.

Cependant, la ségrégation obtenue est bien moins bonne (Figure 67) ; en particulier, elle n'est jamais totale, il existe toujours sur le plan factoriel une zone où les individus sont mélangés. Mais nous pensons qu'elle est beaucoup plus fiable et correspond bien plus à une réalité.

ANALYSE FACTORIELLE DISCRIMINANTE

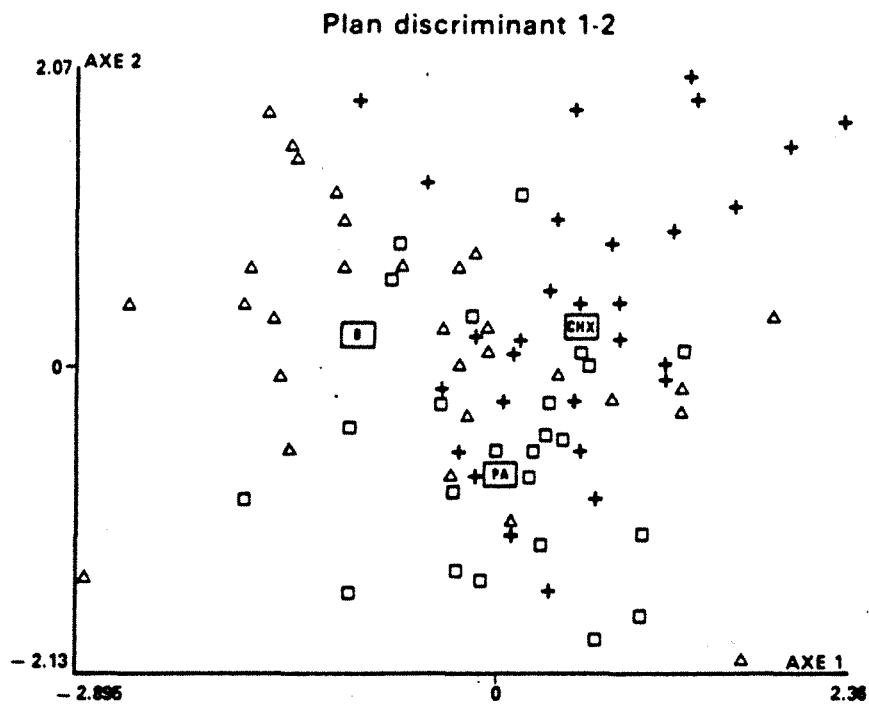
Traces Synthétiques

88 Échantillons

3 Groupes

10 Variables Q1, Q2, Q6, Q7, Q9, QW4, QW5, QW6, AM2, ET NT

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.676	0.676
2	0.324	1.000



Nous avons souhaité conforter ces résultats en discrimination par deux autres méthodes car le test utilisé pour sélectionner les variables est paramétrique. Les hypothèses sont probablement mal vérifiées : beaucoup de ces variables peuvent ne pas être normales et leurs variances sont certainement très différentes d'un groupe à l'autre.

3.4.3. Analyse factorielle discriminante sur les facteurs d'une analyse en composantes principales

Nous n'avons étudié que la population des 43 traces synthétiques, filtrées par le signal choisi, les plus typiques du point de vue faciès sédimentologique.

On a repris les 39 variables appartenant aux diverses familles : ce sont 19 variables sur le spectre de fréquence, FM et les divers déciles de la distribution des amplitudes du spectre d'amplitude ; il y a 10 variables calculées sur l'autocorrélation (rebonds principaux et temps correspondants, passages à zéro) et 10 variables sur le signal analytique. On a fait une analyse en composantes principales Varimax et conservé les 9 composantes principales qui expliquent environ 80 % de l'inertie totale du nuage (les résultats de cette analyse ne sont pas représentés).

Une analyse factorielle discriminante, effectuée sans réduire les nouvelles variables, avec ces 9 composantes principales donne une répartition assez satisfaisante des individus sur le plan factoriel. L'axe 1 a un pouvoir discriminant très supérieur à l'axe 2 (0,782) et permet une opposition nette entre faciès margino-littoraux et faciès de plate-forme. Par contre, les individus du groupe barrière sont assez dispersés. L'image obtenue, malgré tout, reste convenable du point de vue statistique (Figure 68).

Une analyse en composantes principales classique avec les 39 variables a permis l'extraction de 10 composantes principales expliquant 91 % de l'inertie totale du nuage initial. On a fait ensuite une analyse factorielle discriminante sur les 10 composantes, la population étant encore répartie en trois groupes. Les résultats obtenus sont peu satisfaisants car seul l'axe 1 a un pouvoir discriminant non négligeable et permet la séparation entre le groupe des individus de plate-forme et tous les autres. Les individus des groupes barrière et front de delta sont extrêmement mélangés.

On a continué ces études sur la population des 43 traces en effectuant des analyses en composantes principales sur les variables prises famille par famille. Ainsi, on a retenu 2 composantes principales pour résumer l'ensemble des variables calculées sur le spectre

ANALYSE FACTORIELLE DISCRIMINANTE

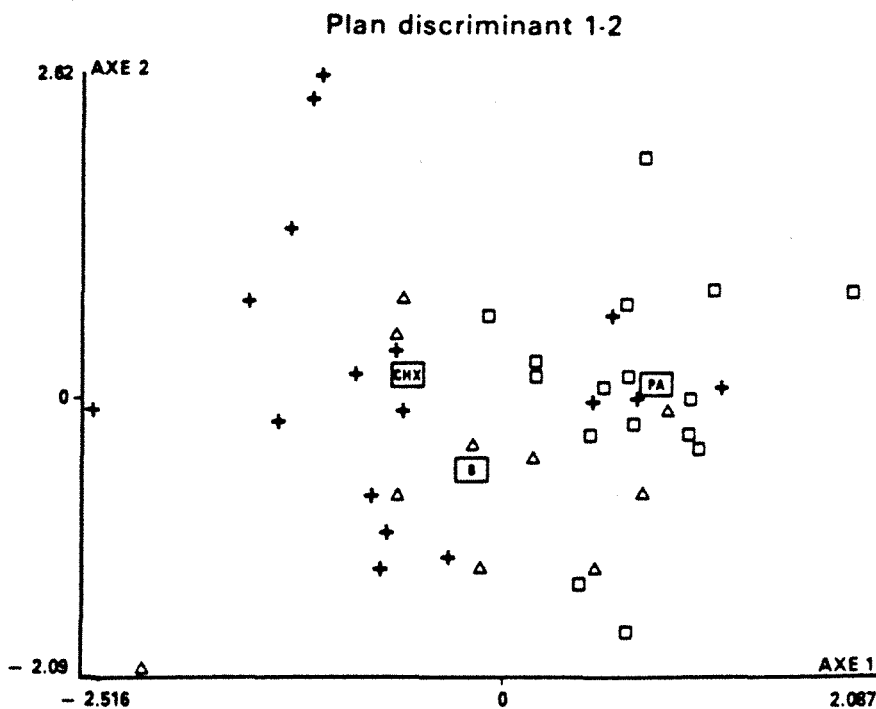
Traces Synthétiques

43 Échantillons

3 Groupes

9 Variables extraites après ACP VARIMAX sur les 39 variables initiales

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.782	0.782
2	0.218	1.000



d'amplitude, 4 composantes principales pour résumer celles sur l'autocorrélation et 2 pour résumer les paramètres calculés sur le signal analytique. L'analyse factorielle discriminante, effectuée avec ces 8 composantes principales (Figure 69), permet une séparation suivant l'axe 1 des faciès plate-forme et des faciès de plaine fluviatile ou margino-littorale. Le faciès barrière est mélangé aux deux précédents ; il n'y a aucune séparation suivant l'axe 2.

3.4.4. Analyse factorielle discriminante pas à pas

La technique du "pas à pas" consiste à introduire les n variables mesurées sur la population des échantillons, au fur et à mesure : à chaque pas, on sélectionne la variable la plus appropriée au problème étudié parmi les $(n - k)$ variables non encore sélectionnées. Le programme que nous avons utilisé base la sélection des variables sur le critère de la trace de la matrice $T^{-1}B$ (cf. paragraphe 4.2.2.). Il permet, outre de **diminuer le nombre de variables à prendre en compte**, de faire un test qui donne une idée de la fiabilité des résultats.

Ce test s'effectue en constituant, à partir de la population de départ, une sous-population que l'on appellera **échantillon de base** qui nous permettra de sélectionner progressivement les variables discriminantes et de déterminer la partition de l'espace en k classes. On définira un deuxième sous-ensemble, appelé **échantillon test**. Nous allons calculer, sur l'échantillon test, un tableau de classement : c'est-à-dire voir dans quelle région, définie à partir de l'échantillon de base, vont être affectés les individus de l'échantillon test.

On peut montrer, à condition que les effectifs de chaque groupe choisi pour l'échantillon test soient proportionnels aux probabilités a priori d'appartenir à chacun de ces groupes –il suffit de prendre un échantillon test dont les effectifs par groupe soient proportionnels aux effectifs par groupe de l'échantillon initial– que les probabilités, a posteriori, pour un nouvel individu d'appartenir à un groupe, sachant dans quel groupe il a été classé, sont données par les éléments du tableau de classement (il ne s'agit que d'un cas particulier du théorème de Bayes).

On dispose ainsi d'un critère de jugement de la fiabilité de la partition de l'espace, réalisée grâce à l'**échantillon de base**. Le pourcentage de bien classés sur l'**échantillon test** permet de connaître la **validité de cette partition du point de vue prévisionnel**.

Par la suite, il est possible d'affecter un échantillon anonyme à un groupe : l'individu anonyme est affecté au groupe dont il est le plus proche.

ANALYSE FACTORIELLE DISCRIMINANTE

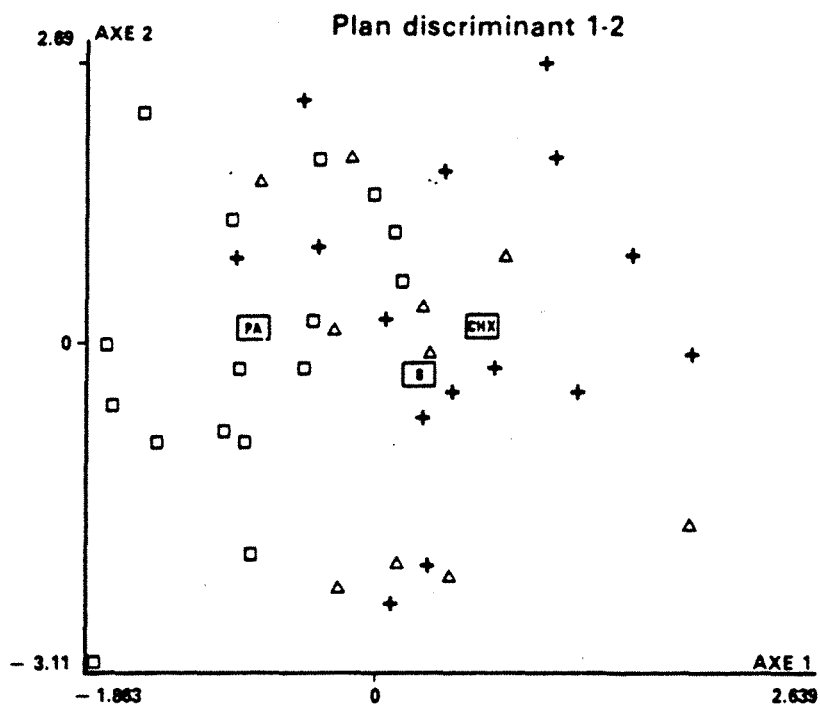
Traces Synthétiques

43 Échantillons

3 Groupes

8 Variables après avoir effectué des ACP VARIMAX sur les variables, famille par famille

AXES	POURCENTAGE D'INERTIE	CUMUL
1	0.950	0.950
2	0.050	1.000



A partir de la population des 88 traces synthétiques filtrées par le signal S1, on a constitué un échantillon de base de 62 individus et un échantillon test de 26 individus. Les individus ont été répartis en trois groupes. Les variables utilisées sont les 39 variables représentant les trois familles de paramètres.

Au pas n°10, on obtient 75,81 % de bien classés sur l'échantillon de base, et un pourcentage de bien classés de 50 % sur l'échantillon test (cf. Tableau 9 des résultats et Figure 70).

Les variables introduites sont, dans l'ordre de leur pouvoir discriminant : ETNT, TPZ2, AM1, Q5, Q2, TAM3, QW3, E1.

Le pourcentage de bien classés de l'échantillon test peut paraître mauvais ; cependant, une répartition aléatoire des individus dans les trois groupes devrait donner seulement un pourcentage de bien classés de 33 %.

Les résultats obtenus sur les 88 individus sont assez satisfaisants et nous pensons qu'ils sont fiables.

A partir de la population des 43 échantillons très caractéristiques, on a constitué un échantillon de base (35 individus) et un échantillon test (8 individus). Les 39 variables précédentes ont été utilisées.

Au pas n° 9, où ont été introduites les variables TPZ2, ETNT, FM, AM1, Q1, Q5, E9, E8, E7, on obtient un pourcentage de 80 % de bien classés sur l'échantillon de base et sur l'échantillon test de 62,50 % (Tableau 10). Le pourcentage de bien classés sur l'échantillon de base augmente au pas suivant, mais comme le pourcentage de bien classés de l'échantillon test se stabilise, nous préférons garder le pas n°9, car après, il est possible que nous n'améliorons pas réellement le pouvoir prédictif de l'analyse.

Ce résultat de 62,50 % de bien classés sur l'échantillon test est satisfaisant, ainsi que la projection du nuage initial sur le plan factoriel, qui fait apparaître un bon regroupement des individus par faciès (Figure 71).

Tableau 9 : Analyse factorielle discriminante pas à pas sur 88 traces sismiques synthétiques

Nombre d'échantillons : 88
 Nombre de groupes : 3
 Nombre de variables : 39
 Groupe 1 : plate-forme
 Groupe 2 : barrière
 Groupe 3 : chenaux

Echantillon de base :

nombre d'individus du groupe 1 : 17
 nombre d'individus du groupe 2 : 15
 nombre d'individus du groupe 3 : 30

Echantillon test :

nombre d'individus du groupe 1 : 7
 nombre d'individus du groupe 2 : 6
 nombre d'individus du groupe 3 : 13

Pas numéro : 10

Variables entrées : 30 21 24 6 26 2 3 29 13 31
 ETNT TPZ2 AM1 Q5 AM3 Q1 Q2 TAM3 QW3 E1

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

GROUPE D'ORIGINE	GROUPE D'AFFECTION		
	1	2	3
1	14	1	2
2	1	11	3
3	4	4	22

Pourcentage de bien classés : 75.81

TABLEAU DE CLASSEMENT DE L'ECHANTILLON TEST

GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	4	3	0
2	2	3	1
3	5	2	6

Pourcentage de bien classés : 50.00

Fig. 70

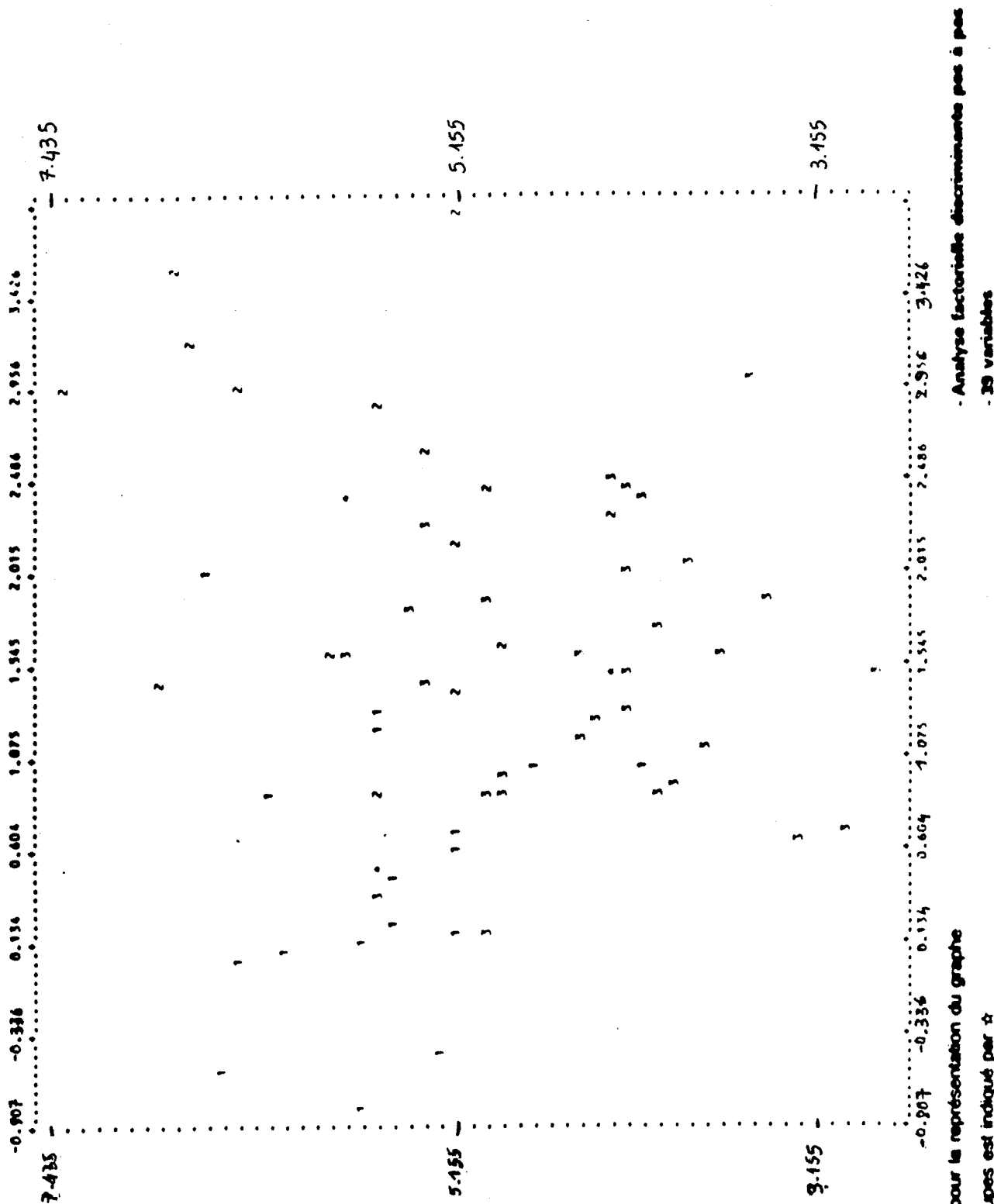


Tableau 10 : Analyse factorielle discriminante pas à pas sur 43 traces sismiques synthétiques

Nombre d'échantillons : 43
Nombre de groupes : 3
Nombre de variables : 9
Groupe 1 : plate-forme
Groupe 2 : barrière
Groupe 3 : chenaux

Echantillon de base :

nombre d'individus du groupe 1 : 13
nombre d'individus du groupe 2 : 8
nombre d'individus du groupe 3 : 14

Echantillon test :

nombre d'individus du groupe 1 : 3
nombre d'individus du groupe 2 : 2
nombre d'individus du groupe 3 : 3

Pas numéro : 9

Variables entrées : TPZ2 ETNT FM AM1 Q1 Q5 E9 E8 E7

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

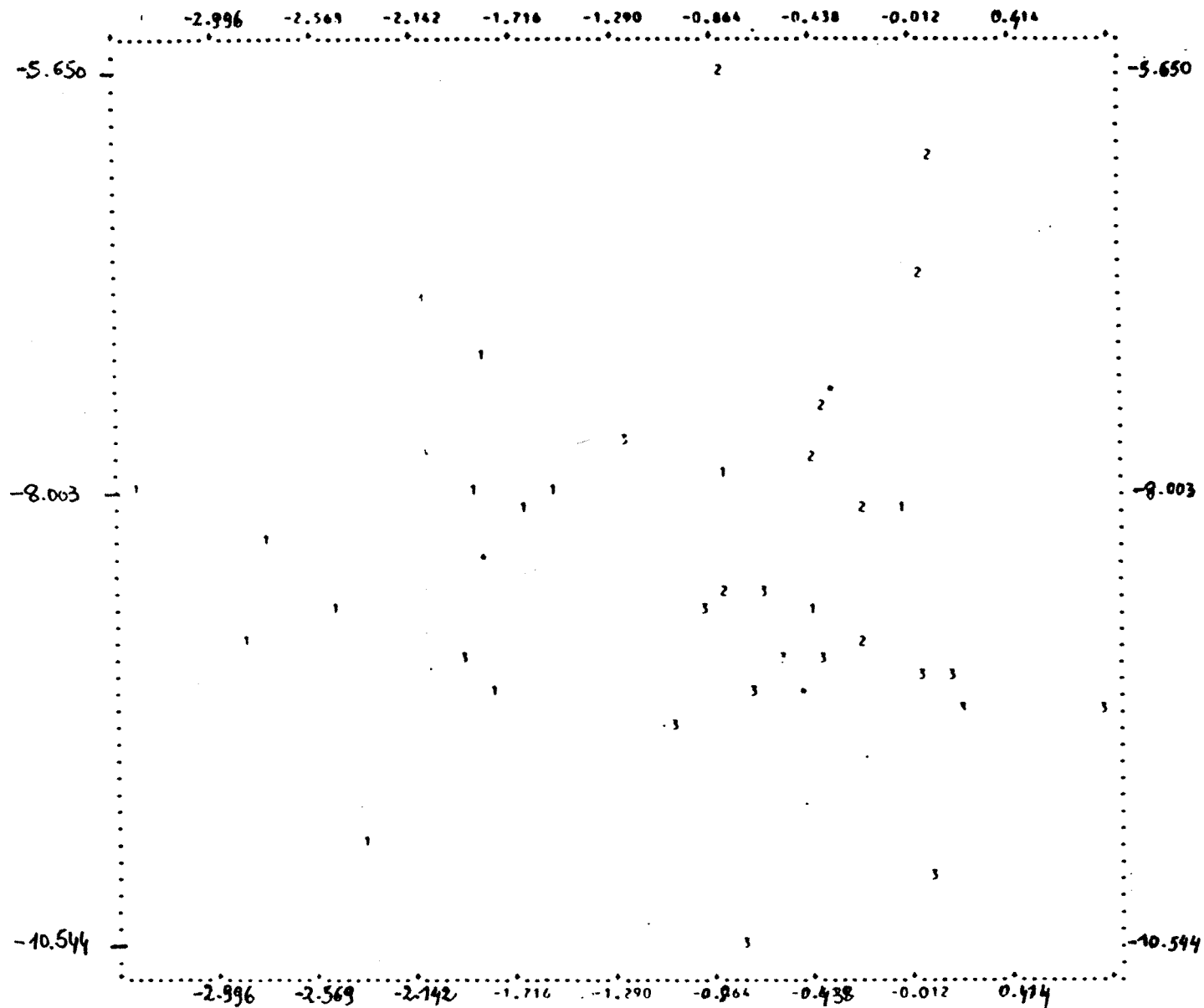
GROUPE D'ORIGINE	GROUPE D'AFFECTION		
	1	2	3
1	10	2	1
2	0	6	2
3	2	0	12

Pourcentage de bien classés : 80.00

TABLEAU DE CLASSEMENT DE L'ECHANTILLON TEST

GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	2	1	0
2	1	0	1
3	0	0	3

Pourcentage de bien classés : 62.50



Code employés pour la représentation du graphe

- La moyenne des groupes est indiqué par *
- Le recouvrement de 2 individus de groupes différents par \$
- Le recouvrement de 2 individus du groupe 1 par la lettre A
- Le recouvrement de 2 individus du groupe 2 par la lettre B
- Le recouvrement de 2 individus du groupe 3 par la lettre C

- Analyse factorielle discriminante pas à pas
- 9 variables
- 43 échantillons synthétiques
- 3 groupes
- pas n° 9 : TP22, ETNT, FM, AM1, Q1, Q5, E9, E8, E7

Fig. 71

3.5. Conclusion à l'analyse des traces synthétiques des puits du delta du Niger

Cette étude nous a amenés à calculer des paramètres sismiques à partir des représentations temporelles et fréquentielles des traces. Ces variables non seulement caractérisent la **morphologie** des traces analysées mais permettent aussi la discrimination entre des groupes de traces représentant divers paléo-environnements.

Ces paramètres sont parfois redondants mais l'analyse factorielle discriminante pas à pas s'est montrée un outil efficace pour sélectionner les paramètres les plus pertinents.

Les analyses factorielles discriminantes réalisées en ne prenant que les variables appartenant à une même famille montrent que la discrimination entre groupes n'est possible qu'avec des paramètres représentant toutes les familles. Par contre, avec l'ensemble des variables, on obtient une régionalisation du plan factoriel en trois zones correspondant aux trois groupes de faciès étudiés, et ceci avec un nombre de variables inférieur à la dizaine.

Ces études nous ont permis non seulement d'affirmer que les groupes de faciès s'individualisaient grâce aux paramètres calculés sur les traces, mais encore que la partition, réalisée au cours des analyses, avait un pouvoir prévisionnel satisfaisant. C'est-à-dire qu'on peut préjuger du faciès auquel appartient un individu anonyme avec une bonne chance de succès.

Les variables effectivement les plus discriminantes, vis-à-vis des trois groupes, se révèlent être l'énergie totale calculée sur le module du signal analytique (ETNT), le second temps de passage à zéro (TPZ2), la valeur de l'amplitude du premier rebond, AM1, le premier et le cinquième décile de la distribution des amplitudes du spectre d'amplitude (Q1, Q5). Ces variables ont toujours été introduites les premières lors des diverses analyses effectuées avec le programme d'analyse factorielle discriminante pas à pas.

Nous allons maintenant tester les paramètres proposés sur des traces sismiques **réelles**, ceci toujours dans le cadre des paléo-environnements du delta du Niger.

3.6. Caractérisation de traces sismiques réelles

La première section sismique étudiée, le profil 12 (Figure 72), qui recoupe les paléo-environnements déjà étudiés, a été calibrée puis interprétée faciologiquement avec les données d'un puits voisin.

Les intervalles prélevés sur la section pour l'analyse du faciès correspondent aux fenêtres temps double suivantes (Figure 72) :

- chenaux de marée et levées :
700 - 950 ms
800 - 1050 ms,
- chenaux et levées de plaine fluviale :
1350 - 1 600 ms,
- front de delta :
1800 - 2000 ms,
- plate-forme :
2000 - 2200 ms.

L'échantillon de base comporte 60 individus de la façon suivante :

- | | |
|---------------------------------------|-----|
| - groupe plate-forme | 15, |
| - groupe front de delta | 15, |
| - groupe margino-littoral et fluviale | 30. |

L'échantillon test comporte 16 individus répartis dans les trois groupes ci-dessus en 4,4 et 8.

Nous utilisons les 39 variables, proposées dans le paragraphe 3.4, FM, Qi, i = 1,... 9, QWi, i = 1,...9, TPZi, i = 1,... 4, AMi, i = 1,... 3, TAMi, i = 1,...3, ETNT, Ei, i = 1,...9.

Il est intéressant de remarquer que dès le pas n°3, nous obtenons un pourcentage de bien classés de 96,67 % sur l'échantillon de base, avec 93,75 % de bien classés sur l'échantillon test. Ainsi, les individus réels des divers faciès s'individualisent de façon tout à fait fiable (cf. tableau 11 des résultats).

INDIVIDUS ANONYMES

7 x 10 TR sur 4 Fen
140 20 160 4 180

AXES FACTORIELS

76 ech.

PROFIL 12

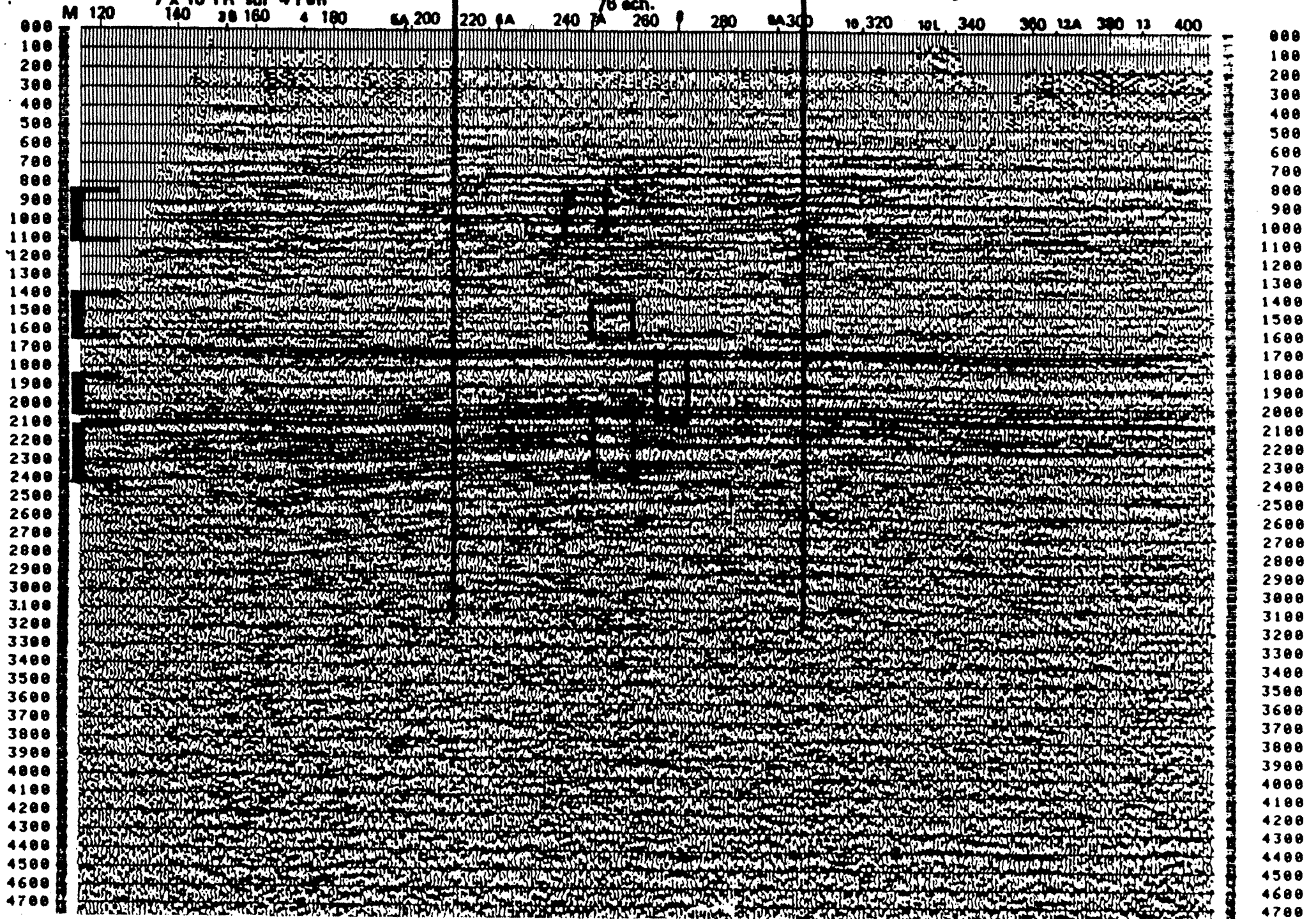


Fig. 72

Les variables sélectionnées appartiennent presque exclusivement à la famille des paramètres quantifiant l'énergie (la plupart des E_i apparaissent comme très discriminants).

Au pas n° 5, on obtient 100 % de bien classés sur l'échantillon de base, et 87,50 % sur l'échantillon test. Les résultats sont donnés dans le tableau 12 et le plan factoriel 1-2 correspond à la figure 73.

Par la suite, nous avons essayé de classer des traces réelles "inconnues" de ce profil prélevées dans les fenêtres suivantes :

1350	-	1600 ms	plaine fluvatile,
800	-	1050 ms	plaine margino-littorale,
1800	-	2000 ms	front de delta,
2050	-	2250 ms	plate-forme interne.

70 traces ont été prélevées ; on a calculé, sur les diverses fenêtres, les paramètres géophysiques que l'on a moyennés par paquet de 10 traces.

On obtient ainsi 14 individus anonymes, qui appartiennent, en fait, au groupe fluvatile et margino-littoral, 7 individus au groupe front de delta et 7 autres au groupe plate-forme.

Nous avons repris l'analyse discriminante pas à pas effectuée précédemment, en s'arrêtant au pas n°3 où les pourcentages de bien classés sur les échantillons de base et test sont déjà excellents. L'introduction d'autres variables n'apporte alors plus beaucoup d'information. Le tableau 13 donne les résultats obtenus : les individus anonymes, au pas n°3, ont tous été parfaitement classés dans le faciès auquel ils appartiennent réellement.

Ce premier profil étudié est particulièrement caractéristique du point de vue faciès sismique, mais il est tout de même rassurant de constater que l'analyse factorielle discriminante permet de retrouver les distinctions entre faciès que l'œil analyse directement sur la section. D'autre part, il est très intéressant de constater que la seule étude de caractéristiques géophysiques sur une verticale permet une bonne classification d'une portion de trace anonyme, sans tenir compte de paramètres latéraux. Ce résultat est encourageant puisqu'il confirme, sur un cas réel, la possibilité d'une discrimination entre faciès sismiques grâce à l'étude des caractéristiques géophysiques d'une portion de trace traversant un milieu géologique donné.

Par la suite, nous avons effectué une étude analogue sur une section sismique plus complexe passant par le puits U1 (Figure 74) : la section 19. Les traces réelles analysées sont voisines du puits, et on a prélevé des fenêtres dans les trois paléo-environnements principaux : margino-littoral, front de delta et plate-forme interne.

Cette section étant beaucoup moins caractéristique que la précédente, le calage des faciès des diverses fenêtres a été effectué à partir de l'étude sédimentologique du puits U1 :

- groupe plate-forme : 2440 - 2600 ms,
- groupe front de delta : 1820 - 2010 ms,
- groupe chenaux : 1080 - 1330 ms.

Tableau 11 : Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 12

Nombre de groupes : 3
 Nombre de variables : 39
 Groupe 1 : plate-forme
 Groupe 2 : barrière
 Groupe 3 : chenaux

Echantillon de base :
 nombre d'individus du groupe 1 : 15
 nombre d'individus du groupe 2 : 15
 nombre d'individus du groupe 3 : 30

Echantillon test :
 nombre d'individus du groupe 1 : 4
 nombre d'individus du groupe 2 : 4
 nombre d'individus du groupe 3 : 8

Pas numéro : 3

Variables entrées : 39 2 30
 E9 Q1 ETNT

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	13	2	0
2	0	15	0
3	0	0	30

Pourcentage de bien classés : 96,67

TABLEAU DE CLASSEMENT DE L'ECHANTILLON TEST

GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	4	0	0
2	0	4	0
3	0	1	7

Pourcentage de bien classés : 93,75

Tableau 12 : Analyse factorielle discriminante sur les traces sismiques réelles de la section 12

Pas numéro : 5

Variables entrées : 39 2 30 37 35
 E9 Q1 ETNT E7 E5

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

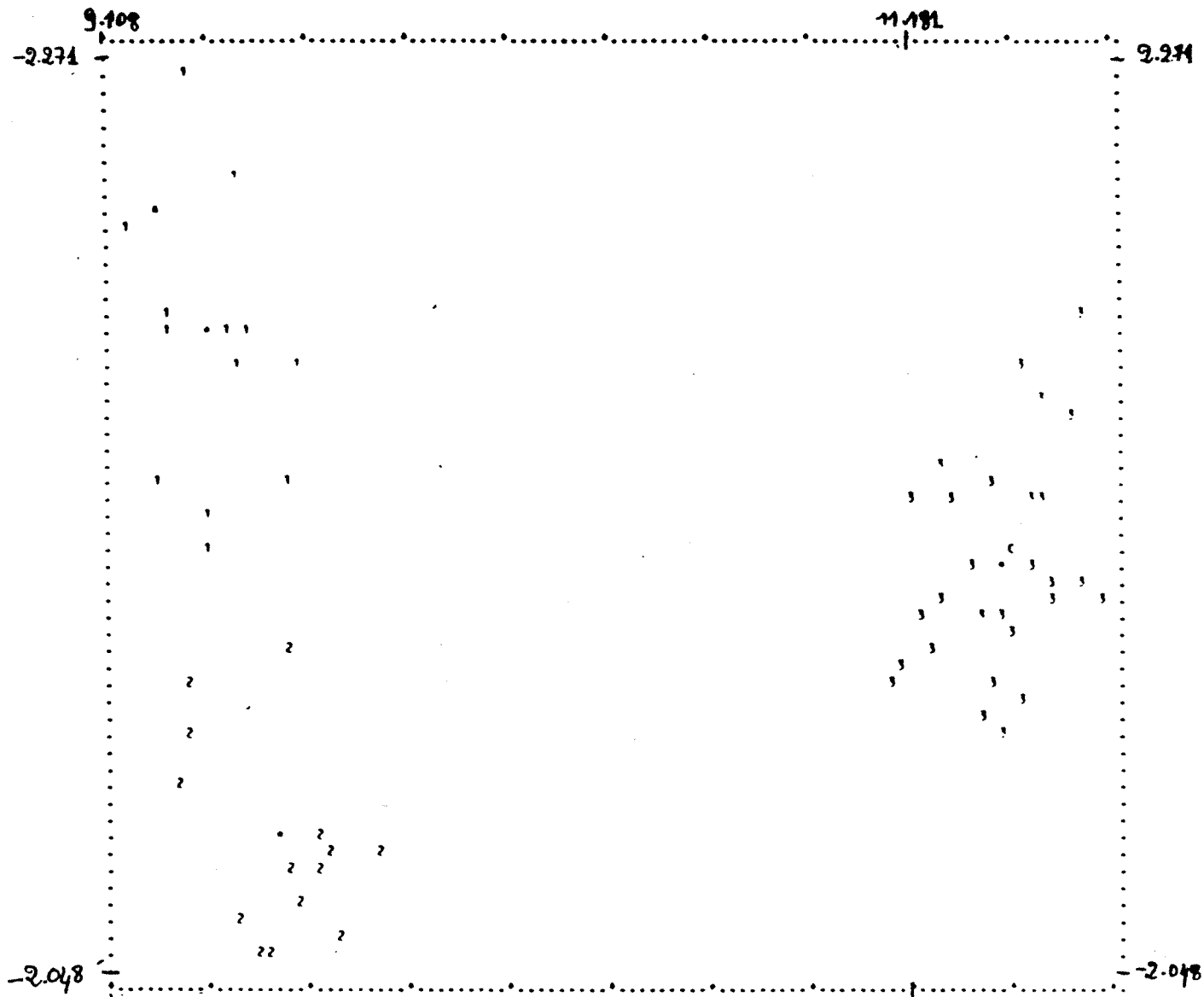
GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	15	0	0
2	0	15	0
3	0	0	30

Pourcentage de bien classés : 100,00

TABLEAU DE CLASSEMENT DE L'ECHANTILLON TEST

GROUPE D'ORIGINE	GROUPE D'AFFECTATION		
	1	2	3
1	3	1	0
2	0	4	0
3	0	1	7

Pourcentage de bien classés : 87,50



Code employes pour la representation du graphe

La moyenne des groupes est indique par :

- Le recouvrement de 2 individus de groupes differents par S
- Le recouvrement de 2 individus du groupe 1 par la lettre A
- Le recouvrement de 2 individus du groupe 2 par la lettre B
- Le recouvrement de 2 individus du groupe 3 par la lettre C

11.181

- Analyse factorielle discriminante pas a pas
- Echantillon base: 60 individus
- Echantillon test: 16 individus
- 39 variables
- 3 groupes
- pas n° 5 E 9 , Q1, ETNT, E 7 , E 5

Fig. 73

Tableau 13 : Analyse factorielle discriminante sur les traces sismiques réelles de la section 12

Affectation d'individus réels "anonymes" de la section 12

n° 1 à 14 : faciès fluviatile et margino-littoral
 n° 15 à 21 : faciès front de delta
 n° 22 à 28 : faciès plate-forme

Pas n° 3, variables E9, Q1, ETNT

AFFECTATION ET DISTANCE DES INDIVIDUS ANONYMES

NUMERO	AFFECTATION	/GP 1	/GP 2	/GP
1	3	15.235	11.011	8.86
2	3	16.232	7.724	7.165
3	3	21.401	13.588	13.617
4	3	24.234	16.363	16.236
5	3	22.677	12.012	13.454
6	3	16.535	10.844	10.705
7	3	18.608	13.056	12.024
8	3	4.300	8.950	1.311
9	3	5.289	13.267	4.351
10	3	3.971	10.517	2.814
11	3	5.911	7.175	2.987
12	3	9.968	11.221	6.087
13	3	7.101	3.142	0.823
14	3	6.315	4.648	0.746
15	2	7.625	0.981	8.017
16	2	5.849	0.048	4.924
17	2	2.520	0.798	3.397
18	2	4.030	0.469	4.386
19	2	4.567	1.717	4.861
20	2	2.199	2.013	4.052
21	2	2.742	0.399	3.730
22	1	15.160	27.657	22.575
23	1	22.390	30.085	27.171
24	1	15.594	25.920	20.579
25	1	18.912	28.377	23.859
26	1	8.016	17.839	14.478
27	1	11.528	25.002	20.234
28	1	13.488	27.497	23.273

A partir de ces portions de traces réelles, on a constitué une population de base comprenant 60 individus (20 individus dans chaque groupe) et une population test de 30 individus (10 par groupe) ; les 39 variables usuelles ont été calculées. On a soumis ce nuage de points à une analyse factorielle discriminante pas à pas.

Au pas n°5, nous obtenons des pourcentages de bien classés de 98,33 sur la population de base et de 70,00 sur la population test.(Tableau 14) Ces résultats sont satisfaisants et la répartition des points sur le plan factoriel 1.2 (Figure 75) met en évidence les différences existant entre les groupes; l'image obtenue est fiable, d'après le pourcentage de bien classés sur l'échantillon test.

Nous avons cherché, par la suite, à affecter des portions de traces de cette section aux groupes définis grâce aux traces calées au puits U1. 50 traces ont été prélevées (Figure 74), les paramètres géophysiques ont été calculés et moyennés par paquet de 10 traces, sans recouvrement. On obtient ainsi 15 individus anonymes, 5 par groupe de faciès.

Les résultats des affectations sont donnés dans le tableau 15.

La majorité des individus représentant la plate-forme (groupe 1 et individus 1 à 5) ont bien été affectés à ce groupe. Presque tous les individus du faciès "chenaux et levées" (groupe 3 et individus 11 à 15) ont été affectés au groupe chenaux. Par contre, les individus que nous pensions attribuer au faciès front de delta (groupe 2 et individus 6 à 10) ont été massivement affectés au groupe des chenaux. On peut s'interroger sur le faciès sismique réel de ces traces : en effet, au niveau étudié, il semble que la barrière soit interrompue, peut-être suite à une érosion par des chenaux.

Les résultats obtenus sur la section 19, bien que moins bons que ceux obtenus sur le premier profil (section 12), sont encourageants dans la mesure où ils confirment la possibilité d'une **discrimination fiable**, grâce à un petit nombre d'attributs, entre traces de faciès sismiques différents qui se ressemblent pourtant beaucoup à l'œil.

Fig. 74

PROFIL 19
codecap phase
mini

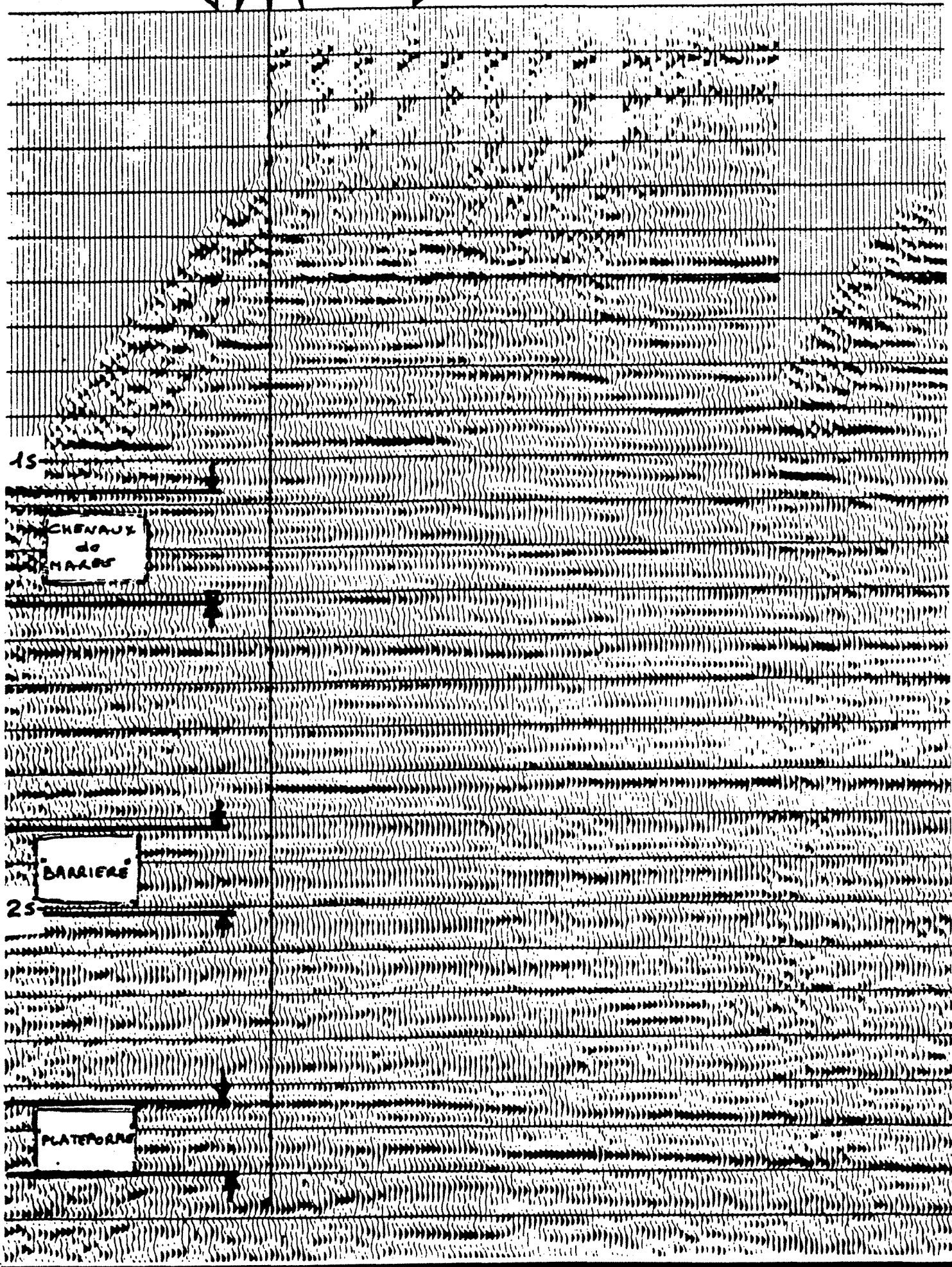
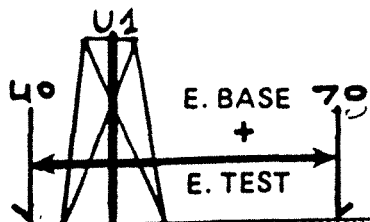


Tableau 14 : Analyse factorielle discriminante pas à pas sur les traces sismiques réelles de la section 19

Pas numéro : 5

Variables entrées : 32 36 34 23 31
 E2 E6 E4 TPZ4 E1

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

GROUPE D'ORIGINE	GROUPE D'AFFECTION		
	1	2	3
1	20	0	0
2	0	20	0
3	0	1	19

Pourcentage de bien classés : 98,33

TABLEAU DE CLASSEMENT DE L'ECHANTILLON TEST

GROUPE D'ORIGINE	GROUPE D'AFFECTION		
	1	2	3
1	7	0	3
2	1	8	1
3	0	4	6

Pourcentage de bien classés : 70,00

TRACES REELLES

CALEES AU PUIS U 1

A.F.D. pas à pas

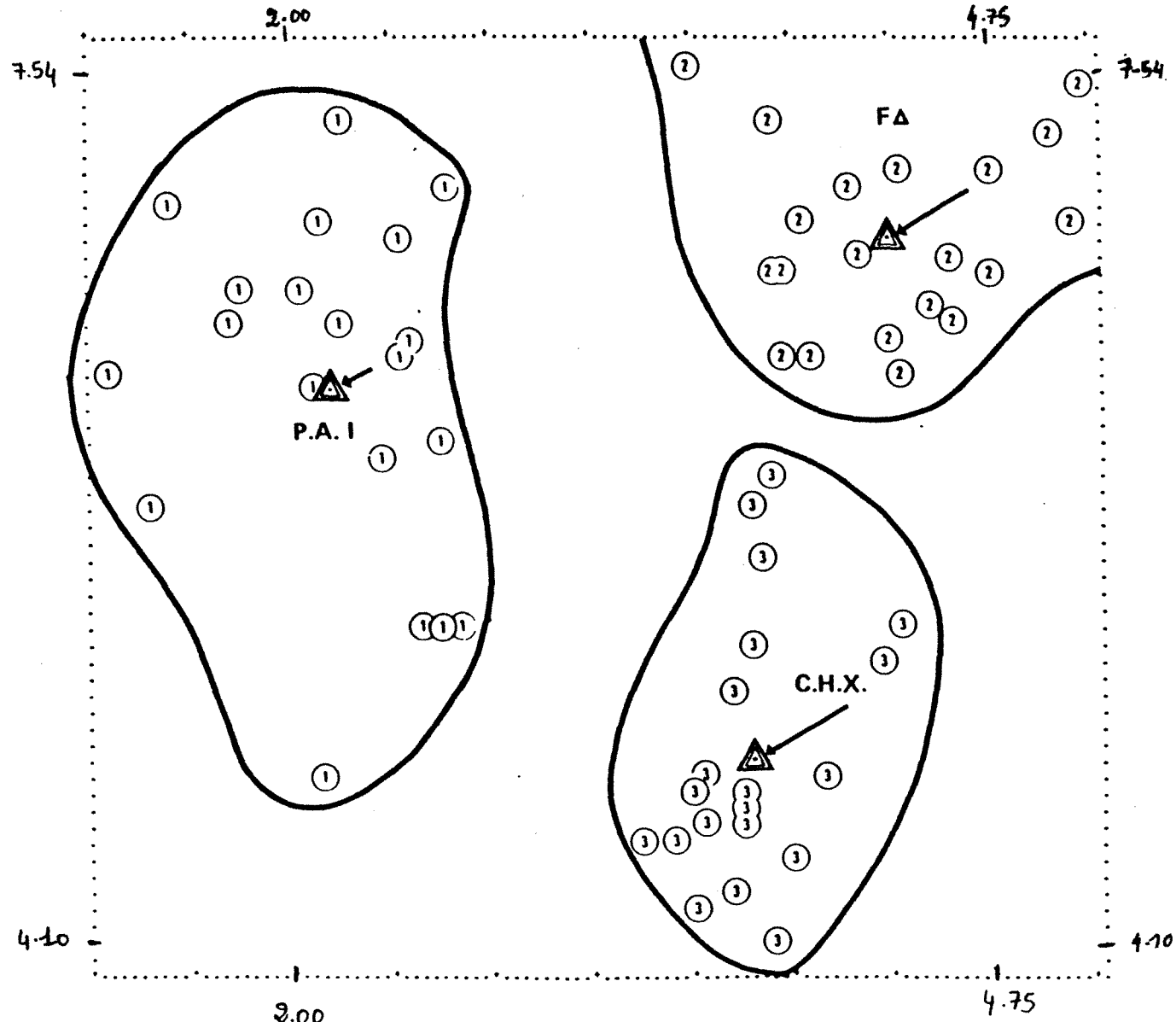
/ 3 Groupes pas n° 5

E 2

E 6

E 4

TPZ 4 E 1



% bien classés

E. BASE 98.33 %

E. TEST 70.00 %

Fig. 75

Tableau 15 : Analyse factorielle discriminante pas à pas sur les traces
sismiques réelles de la section 19

Pas n° 5

AFFECTATION ET DISTANCE DES INDIVIDUS ANONYMES

NUMERO	AFFECTATION	/GP 1	/GP 2	/GP 3
1	1	1.806	3.507	1.986
2	1	3.871	5.306	6.873
3	1	3.824	4.597	7.042
4	3	32.819	24.645	20.914
5	2	3.653	0.688	3.587
6	1	2.553	2.731	6.111
7	3	3.197	3.138	1.100
8	3	9.919	6.835	1.714
9	3	13.688	8.568	3.714
10	3	7.336	4.924	1.529
11	3	4.272	3.486	0.710
12	3	11.102	9.350	5.024
13	3	15.022	12.614	6.37
14	1	2.593	6.626	5.637
15	3	3.941	5.218	1.854

3.7. Conclusion

Afin de caractériser la morphologie des traces sismiques sur un intervalle temps d'environ 100 à 200 ms, nous avons été amenés à calculer des paramètres sur les représentations temporelles et fréquentielles de la trace sismique : enveloppe (module du signal analytique), autocorrélation et spectre de puissance.

Ces paramètres sont des mesures **globales** des caractéristiques sismiques sur la fenêtre analysée. Nous avons vu que l'autocorrélation et son spectre permettent de caractériser divers environnements de dépôts, dans le cadre de l'étude de la réponse sismique d'un modèle sédimentaire markovien de premier ordre.

D'autre part, les paramètres sismiques calculés se sont révélés efficaces pour discriminer des groupes de traces sismiques synthétiques, calculées à partir des données diagraphiques de puits du Niger dans les trois principaux environnements (plate-forme, front de delta et margino-littoral). Des résultats analogues ont été retrouvés sur des **sections sismiques réelles** présentant les mêmes faciès et environnements. Le jeu de paramètres proposés semble donc relativement complet.

Ces paramètres sont partiellement redondants mais l'analyse discriminante pas à pas nous a permis de bien discriminer les différents environnements avec un **nombre restreint** de paramètres mesurés sur les traces.

Les variables sélectionnées au cours des différentes études sont :

- ETNT, TPZ2, AM1, Q1, Q5,
- E9, Q1, ETNT,
- E2, E6, E4, TPZ4, E1.

On constate que les variables les plus discriminantes changent d'une étude de cas à l'autre et qu'aucune règle générale n'apparaît. Toutefois, parmi ces variables, on trouve toujours une ou plusieurs variables décrivant la trace (ETNT ou E_i), parfois l'autocorrélation (TPZ4 ou TPZ2) ou son spectre (Q1, Q5). Le fait qu'aucune règle générale ne se dégage de cette étude pourrait être une limite très sérieuse à la mise en œuvre de cette méthode de caractérisation du faciès.

En fait, nous verrons dans le chapitre suivant, qu'à condition de disposer d'une certaine information a priori sur la zone étudiée (par l'intermédiaire de puits, par exemple), ce problème

peut être résolu en faisant appel à une méthodologie mettant en œuvre diverses techniques statistiques multivariées.

4. MISE AU POINT D'UNE METHODOLOGIE MULTIVARIABLE POUR LA CARACTERISATION DES FACIES SISMIQUES

4.1. Introduction

La caractérisation quantitative et l'analyse automatique des faciès sismiques soulèvent deux problèmes :

- comment caractériser les portions de traces étudiées ?
- comment prendre en compte toute la richesse de l'information sismique et géologique.

A ce stade de l'étude, nous pensons avoir résolu de façon assez satisfaisante le premier problème. Les attributs sismiques calculés reprennent la quasi-totalité de l'information sismique après sommation, de façon globale et assez peu sujette au bruit. Par contre, ils sont très nombreux, probablement redondants et tous ne sont pas nécessairement pertinents pour le problème posé. Ainsi sur la figure 76, nous étudions les variations latérales d'un réservoir carbonaté. Nous disposons de deux puits de calage sur la section sismique. Le premier (puits B) présentait, au niveau de la formation étudiée, un faciès carbonaté poreux (réservoir oolithique) et le second (puits A) un faciès compact (pas de réservoir). Il fallait non seulement étudier d'éventuelles variations morphologiques des traces sismiques entre les deux puits, mais aussi chercher à savoir si les variations étaient en relation avec l'évolution des propriétés géologiques. Dans ce cas là, il était important de déterminer les paramètres sismiques les plus discriminants pour la reconnaissance des zones poreuses par rapport aux zones compactes.

Il faut donc pouvoir trier l'information sismique et intégrer des connaissances géologiques a priori (données de puits, par exemple).

Les méthodes statistiques multivariées apparaissent particulièrement adaptées à ce type de problème puisqu'elles permettent une synthèse de la totalité de l'information, la mise en évidence de ses traits majeurs et ceci éventuellement par rapport à une information pré-déterminée (analyse factorielle discriminante).

Il est étonnant de constater à quel point les méthodes classiques d'analyse des données ont peu pénétré en exploration pétrolière, en particulier pour le traitement des données sismiques. A ceci, deux explications sont possibles. Tout d'abord, les domaines d'application traditionnels des méthodes multivariées sont les sciences humaines (les premiers développements ont eu lieu dans ce contexte), les sciences biologiques (taxinomie pour la systématique des êtres vivants) ainsi que le domaine médical (caractéristiques pathologiques de certaines maladies, aide au diagnostic...). D'autre part, le traitement des données sismiques a été développé par des médecins, pour qui l'analyse des données est un monde étranger ; quant à l'interprétation des données sismiques, elle est souvent le fait de géologues, naturalistes de formation, et donc peu sensibilisés aux méthodes statistiques.

SEISMIC STRATIGRAPHIC INTERPRETATION

LATERAL VARIATIONS OF A CARBONATE RESERVOIR

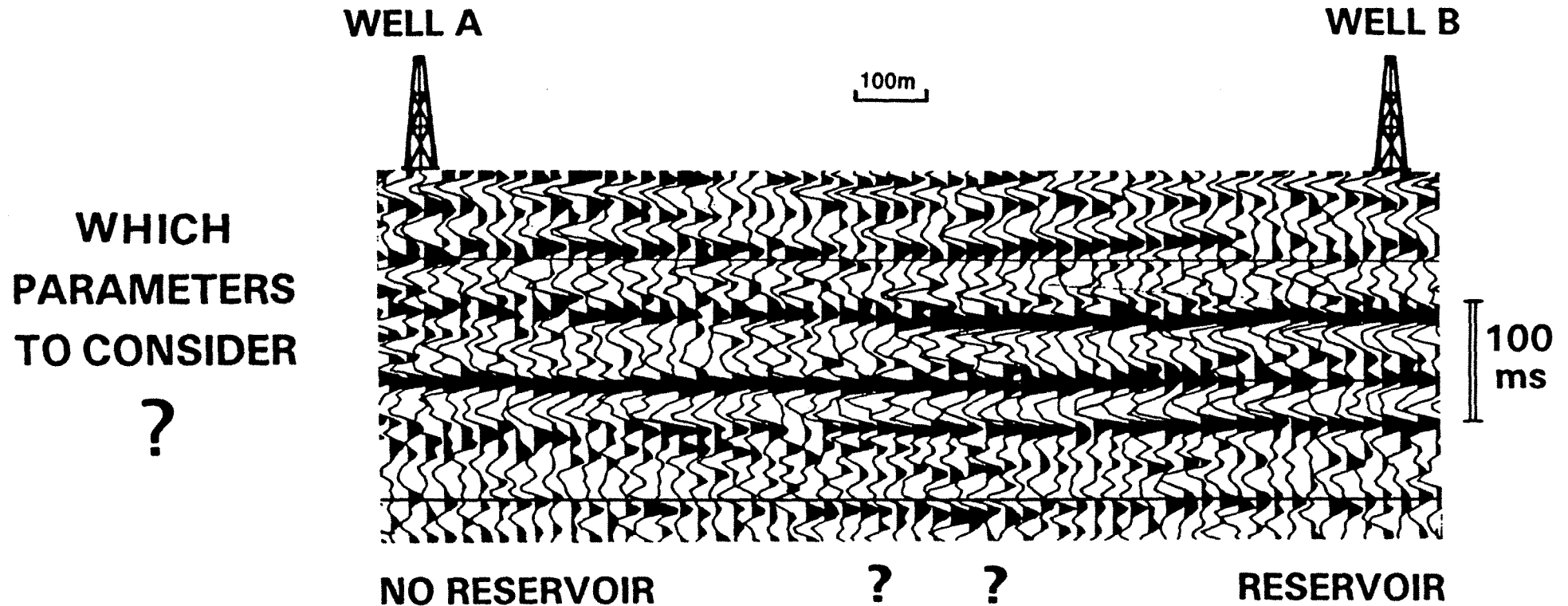


Fig. 76

Quelques applications statistiques dans le domaine pétrolier existent, mais surtout en géochimie organique ou pour le traitement et l'interprétation des enregistrements diagraphiques aux puits. Au début de ces travaux de recherche (mi-84), seuls quelques travaux (cf. analyse bibliographique, paragraphe 1) montraient des applications aux données sismiques et l'emploi d'une technique statistique multivariable était presque accessoire par rapport à l'analyse classique des traces. Or les données sismiques correspondent à une masse d'information énorme (domaines avant ou après sommation) qu'il peut être avantageux de traiter par méthodes multivariées, et ceci à tous les stades depuis l'acquisition (par exemple, détection de traces bruitées) jusqu'à l'inversion (cf. les travaux de recherche menés actuellement sur "l'inversion statistique").

Les méthodes que nous avons utilisées sont : les classifications automatiques, l'analyse en composantes principales et l'analyse factorielle discriminante (Figure 77).

4.2. Rappels sur les méthodes utilisées

Nous considérons un tableau de données de N individus sur lesquels ont été mesurées P variables.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1P} \\ x_{21} & x_{22} & \dots & x_{2P} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NP} \end{pmatrix}$$

Nos mesures peuvent se représenter :

- soit dans un espace de dimension P engendré par les P variables, les points correspondant aux individus dont les coordonnées sont les valeurs prises par les variables ;
- soit dans un espace de dimension N, dont les axes sont les individus ; les variables y sont à leur tour représentées par des points dont les coordonnées sont les différentes mesures, pour cette variable, enregistrées sur les individus.

Ces deux représentations sont implicites au tableau des données. La plupart des techniques d'analyse des données reposent sur la notion de distance entre les points (les points proches sont des points semblables), soit dans l'espace des individus, soit dans l'espace des variables. L'analyse de l'un ou l'autre de ces deux espaces est équivalente en analyse factorielle (propriété de dualité).

La notion de distance implique au préalable la définition d'une base sur l'espace de représentation.

ANALYSES MULTIDIMENSIONNELLES

ANALYSE FACTORIELLE

REPRESENTATION
de la POPULATION
dans un ESPACE $< n$



CONDENSATION
de L'INFORMATION

A.C.P.

Analyse Composantes Principales

RECHERCHE d'un ESPACE $< n$ où

L'IMAGE de la population projetée
dans cet espace est la plus voisine
de l'image initiale

CLASSIFICATION

AGREGATION ECHANTILLONS
SEMBLABLES
dans l'espace à n dimensions
des variables

A.F.D.

Analyse Factorielle Discriminante
(les échantillons appartiennent
à des GROUPES PRE DEFINIS)

RECHERCHE d'un ESPACE $< n$ où

- LES GROUPES sont les mieux SEPARES
- LES ECHANTILLONS D'UN GROUPE SONT LES MIEUX REGROUPES

La définition d'une base et d'une distance permet de former la métrique de l'espace, qui caractérise les différentes techniques d'analyse de données.

Soit S la matrice de variance-covariance associée au tableau X

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{pmatrix}$$

avec :

$$s_{kj} = \frac{1}{N} \sum_{i=1}^N (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)$$

où \bar{x}_k est la moyenne de la k^{ième} variable

$$\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{ik}$$

et \bar{x}_j la moyenne de la j^{ième} variable.

Soit R la matrice de corrélation

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

où :

$$r_{kj} = \frac{s_{kj}}{\sqrt{s_{kk} s_{jj}}}$$

Soit \bar{x} la moyenne du nuage de points :

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_p \end{pmatrix}$$

4.2.1. Analyse en composantes principales (Figure 78)

4.2.1.1. Nature du problème

Il s'agit de trouver une suite orthonormée u_1, u_2, \dots, u_l de vecteur de R^P avec $l \leq P$ pour laquelle on ait "en moyenne" (Figure 78) :

$$\|x_i - x_j\|_{R^P} \approx \|\pi_E(x_i) - \pi_E(x_j)\|_{R^P}$$

où :

$$E = E(\bar{x}/u_1, u_2, \dots, u_l)$$

est la variété affine centrée sur le barycentre \bar{x} du nuage de points et générée par les vecteurs u_1, u_2, \dots, u_l , et π_E est l'opérateur de projection orthogonale de R^P sur E .

En introduisant la notion de distorsion Δ^2 ou de déformation du nuage projeté sur l'espace réduit E :

$$\Delta^2 = \frac{1}{N^2} \sum_i \sum_j \left| \|x_i - x_j\|_{R^P}^2 - \|\pi_E(x_i) - \pi_E(x_j)\|_{R^P}^2 \right|,$$

il s'agit de trouver une suite orthonormée u_1, u_2, \dots, u_l , avec $l \leq P$, minimisant la quantité Δ^2 .

Soit J^2 l'inertie du nuage de points :

$$J^2 = \sum_{i=1}^N M_i \|x_i - \bar{x}\|_{R^P}^2$$

avec M_i , poids associé à l'individu x_i .

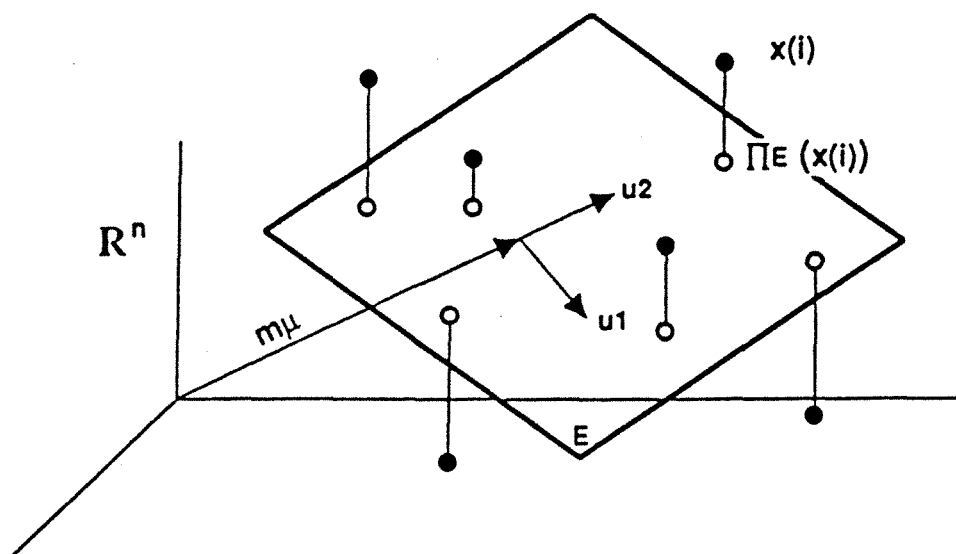
On peut montrer que le problème précédent revient à trouver une suite de vecteurs orthonormés u_1, u_2, \dots, u_l qui maximisent l'inertie $J^2_{\pi_E}$ du nuage de points projetés sur E , avec :

$$J^2_{\pi_E} = \sum_{i=1}^N M_i \|\pi_E(x_i) - \bar{x}\|_{R^P}^2$$

ou bien encore qui maximisent le pourcentage d'inertie expliquée défini par

$I^e(u_1, u_2, \dots, u_l) = \frac{J^2_{\pi_E}}{J^2}$. Le pourcentage d'inertie expliquée est d'ailleurs un indice mesurant la fidélité de l'image sur π_E par rapport à l'image initiale. Si $I^e(u_1, u_2, \dots, u_l) = 0$, alors

ANALYSE EN COMPOSANTES PRINCIPALES



PROBLÈME : Soit $m < n$
 chercher une suite orthonormée $(u_1 \dots u_m)$ pour laquelle on ait
 "en moyenne" :

$$\|x(i) - x(j)\|_{R^n} \simeq \|\Pi_E(x(i)) - \Pi_E(x(j))\|_{R^n}$$

(en moyenne $\forall i, j$)



Chercher une suite orthonormée $(u_1 \dots u_m)$ rendant
 minimale la déformation du nuage projeté



Ce sont les vecteurs propres de la matrice de
 variance-covariance du nuage de points.
 Les valeurs propres mesurent la part d'inertie portée
 par les axes factoriels

l'image projetée n'est pas meilleure que si on représentait le nuage de points par son barycentre \bar{x} . Si $I^e(u_1, u_2, \dots, u_l) = 1$, alors l'image projetée est aussi bonne que l'image initiale.

4.2.1.2. Solution au problème

La variance dans une direction u , $\sigma^2(u)$, est reliée à la matrice de covariance S par la relation suivante :

$$\sigma^2(u) = u^t \cdot S \cdot u.$$

Si u est un vecteur unitaire ($\|u\| = 1$) et s'il est vecteur propre de S associé à la valeur propre λ , alors :

$$\sigma^2(u) = \lambda.$$

La solution au problème posé est donnée par la suite de vecteurs orthonormés (u_1, u_2, \dots, u_l) qui maximisent $I^e(u_1, u_2, \dots, u_l)$.

$$\text{Or } I^e(u_1, u_2, \dots, u_l) = \frac{\sum_{i=1}^l \sigma^2(u_i)}{D^2}$$

avec D^2 la variance isotrope qui est constante ($D^2 = \text{Trace}(S)$).

$$I^e(u_1, u_2, \dots, u_l) = \frac{\sum_{i=1}^l u_i^t \cdot S \cdot u_i}{D^2}.$$

Il suffit donc de maximiser chaque terme de la somme

$$\sum_{i=1}^l u_i^t \cdot S \cdot u_i.$$

$\sigma^2(u_i)$ est maximum pour f_i , **vecteur propre de S** associé à la **valeur propre λ_i** avec :

$$\sigma^2(f_i) = \lambda_i.$$

Finalement, il suffit de calculer les valeurs propres λ_i et les vecteurs propres f_i de la matrice de variance - covariance S .

$$I^e(f_1, f_2, \dots, f_l) = \frac{\sum_{i=1}^l \lambda_i}{P}$$

$$(D^2 = \text{Trace}(S) = \sum_{i=1}^P \lambda_i).$$

4.2.1.3. Définition de l'analyse en composantes principales

L'analyse en composantes principales du nuage des individus $x(i)$ affectés des poids M_i consiste à déterminer la suite orthonormée (f_1, f_2, \dots, f_l) de l vecteurs ($l \leq P$), tels que pour toute autre suite orthonormée (u_1, u_2, \dots, u_l) , on ait :

$$I^e(f_1, f_2, \dots, f_l) \geq I^e(u_1, u_2, \dots, u_l).$$

On appelle **composante principale** ou **facteur principal** chacune des l fonctions $(\phi_1, \phi_2, \dots, \phi_l)$ définies sur \mathbb{R}^P , à valeurs dans \mathbb{R} telles que $\phi_i(x) = \langle (x - \bar{x}), \frac{f_i}{\sqrt{\lambda_i}} \rangle$.

On appelle **coordonnée principale** chacune des l fonctions $(\pi_{E_1}, \pi_{E_2}, \dots, \pi_{E_l})$ définies sur \mathbb{R}^P , à valeurs dans \mathbb{R} telles que :

$$\pi_{E_i}(x) = \langle (x - \bar{x}), f_i \rangle$$

où $E_i = E(\bar{x} / f_i)$.

Remarque : les l vecteurs (f_1, f_2, \dots, f_l) ne sont autres que les l vecteurs propres de S associés aux l premières valeurs propres, supposées rangées par ordre décroissant.

Les composantes principales $\phi_i(x)$ sont liées par la relation :

$$\pi_{E_i}(x) = \sqrt{\lambda_i} \phi_i(x) = \sigma(f_i) \phi_i(x).$$

Les nuages de points projetés soit avec ϕ_i , soit avec π_{E_i} sont tels que :

$$m\phi_i = 0$$

$$\sigma_{\phi_i} = 1$$

$$m\pi_{E_i} = 0$$

$$\sigma_{\pi_{E_i}} = \sqrt{\lambda_i}$$

La matrice de variance-covariance S peut se décomposer en :

$$S = \sum_{i=1}^P \lambda_i f_i f_i^t.$$

4.2.1.4. Représentations graphiques et interprétation

4.2.1.4.1. Construction d'une image du nuage de points initial

a - Image fidèle

Pour construire une **image fidèle**, il faut définir un seuil ρ , ($0 < \rho \leq 1$) représentant la valeur minimale du pourcentage d'inertie I^e que l'on désire retrouver dans l'image projetée sur l'espace de dimension $1 \leq P$, puis on utilise les l premiers facteurs tels que :

l soit le plus petit entier avec $I^e(f_1, f_2, \dots, f_l) \geq \rho$

ou tel que $\left(\frac{\sum_{i=1}^l \lambda_i}{\sum_{i=1}^P \lambda_i} \right) \geq \rho$.

L'image fidèle du nuage de points initial s'obtient en le projetant sur la variété linéaire $E = E(\bar{x}/f_1, f_2, \dots, f_l)$, avec l'opérateur de projection orthogonale :

$$\pi_E(x) = \begin{pmatrix} \pi_{E_1}(x) \\ \pi_{E_2}(x) \\ \pi_{E_l}(x) \end{pmatrix}$$

On peut montrer (avec l'inégalité de Bienaymé-Tchebycheff) qu'au moins 75 % de la masse du nuage de points projeté sur E se trouvent dans le pavé B de \mathbb{R}^l :

$$B = B_1 \times B_2 \times \dots \times B_l$$

avec $B_i = [-2\sqrt{\lambda_i}, 2\sqrt{\lambda_i}]$.

En conséquence, si les λ_i sont très disparates, l'image sur E présentera certes une distorsion minimale mais sera aussi très mal "cadrée" dans \mathbb{R}^l .

b - Image normalisée

Très souvent, au prix d'une distorsion peu gênante, on préfère réaliser une image "normalisée" du nuage de points en effectuant des affinités sur les axes de \mathbb{R}^l pour transformer chaque B_i en $B_i^* = [-2, +2]$.

Soit φ l'application de \mathbb{R}^P dans \mathbb{R}^1 telle que :

$$\varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \\ \varphi_l(x) \end{pmatrix}$$

Alors la projection du nuage de points sur E avec φ est une image bien cadrée car 75 % de sa masse se trouvent dans le pavé $B^* = [-2, +2]^1$.

4.2.1.4.2. Représentations graphiques

a - Représentation des échantillons

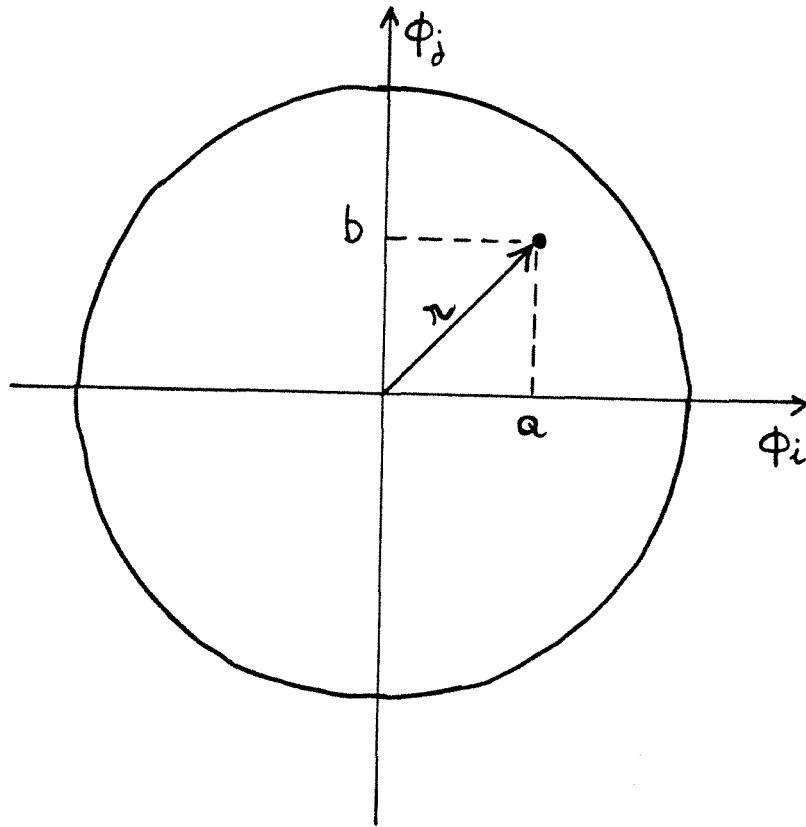
L'échantillon i est représenté dans les différents plans factoriels (ϕ_i, ϕ_j) (cf image normalisée du nuage de points) par un point dont les coordonnées sont $\phi_i(x(i)), \phi_j(x(i))$.

La quantité $I_e(\phi_i) = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$ est le pourcentage d'inertie expliquée par le plan (ϕ_i, ϕ_j) .

b - Représentation des variables

A condition de travailler en analyse en composantes principales à partir de la **matrice de corrélation R** (et non plus de la matrice de variance-covariance S) -A.C.P. réduite- les variables initiales se projettent à l'intérieur d'une hypersphère de rayon 1, chaque variable ayant sa projection à une distance de l'origine proportionnelle à la corrélation multiple avec les composantes principales qui déterminent l'espace.

Dans le cas d'un **plan factoriel**, l'hypersphère est réduite à un cercle, le cercle des corrélations.



- a : corrélation simple entre la variable X et la composante principale ϕ_i .
- b : corrélation simple entre la variable X et la composante principale ϕ_j .
- r : corrélation multiple entre la variable X et les composantes principales ϕ_i, ϕ_j .

4.2.1.4.3. Remarques sur l'interprétation des résultats de l'analyse en composantes principales

Les composantes principales ne sont pas seulement un moyen algébrique pour décrire de façon simple une réalité complexe. Leur interprétation peut permettre de mieux appréhender la réalité physique du phénomène.

a - Choix des données et de leur poids dans l'analyse

- **Choix des variables initiales**

En général, il faut éviter de rentrer un groupe de variables fortement corrélées qui accaparaient un facteur et masqueraient une grande partie de l'information.

Si certaines variables ont un ordre de grandeur très supérieur aux autres, elles auront un poids plus grand dans la détermination des composantes, ce qui est peu intéressant.

Un inconvénient est de réduire les variables (division par leur écart-type, par exemple), mais la réduction masque les différences de variabilité à ordre de grandeur égal.

Si une variable a une faible variance, il convient de lui attribuer un poids qui la rend comparable aux autres ou de l'éliminer si elle est inintéressante afin qu'elle ne crée pas un bruit de fond.

- **Point de vue des échantillons**

Il est important de détecter les valeurs "anormales" sous peine d'erreur dans l'interprétation (faire appel à des moyens statistiques classiques). L'interprétation des valeurs anormales avant leur élimination constitue un résultat en soi.

b - Analyse des tableaux et des représentations graphiques

- **Pourcentage de variance exprimée par chaque facteur et pourcentage de variance cumulée**

- au-delà du seuil de 90 %, les facteurs offriront peu d'intérêt .
- l'analyse a d'autant plus de valeur que le nombre de facteurs dont la variance cumulée atteint le seuil des 90 % est petit ; mais si un seul facteur atteint 90 %, il convient de s'intéresser aux facteurs expliquant les 10 % de variance restante.

- **Vecteurs propres**

L'examen des vecteurs propres peut mettre en évidence un "facteur taille". Tous les éléments du vecteur propre sont de même signe et de même grandeur. Les échantillons sont alors classés par ordre de grandeur le long de l'axe factoriel. Dans l'espace des variables, ceci se matérialise par un regroupement des variables à l'extrémité d'un axe. Le facteur taille n'est pas toujours intéressant. En étudiant le tableau des corrélations variable-facteur, on peut, en changeant le poids des variables dont le coefficient de corrélation est très fort, agir sur le facteur incriminé.

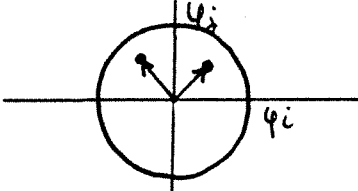
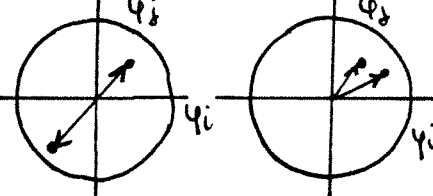
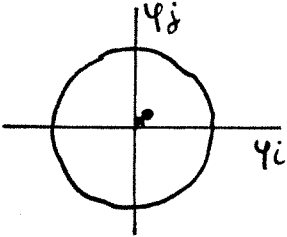
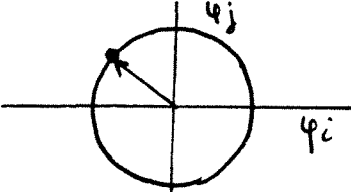
Cercle des corrélations	Interprétation
	Variables indépendantes
	Variables très corrélées
	Variable apportant peu d'information sur la structure des facteurs
	Variable influente sur les facteurs

Tableau 16 - Analyse graphique du cercle des corrélations

• **Facteur "forme"**

Les variables peuvent aussi se regrouper sur le cercle des corrélations aux deux extrémités d'un axe. Elles sont opposées mais toutes très corrélées avec le facteur, on parle de "facteur forme".

- L'étude du tableau des **corrélations simples** variables initiales-composantes principales permet de connaître les variables "expliquées" par une composante principale. L'analyse du tableau des **corrélations multiples** variables initiales-composantes principales se fait par la mise en évidence d'un brusque saut dans l'augmentation du coefficient de corrélation. La position du saut marque le facteur notable dans l'expression de la variable. Ces analyses permettent d'évaluer la contribution des facteurs à l'explication des variables.

• **Analyse graphique** (Tableau 16)

Pour une A.C.P centrée réduite (on étudie R, la matrice des corrélations), on examine le **cercle des corrélations** : le cosinus de l'angle entre deux variables traduit la corrélation existant entre ces deux variables. Si les deux variables forment un angle de 90°, elles sont indépendantes. Si les deux variables sont situées chacune à l'extrémité d'un segment passant par le centre ou si elles sont proches l'une de l'autre, elles sont fortement corrélées.

Une variable proche du centre de gravité ne donne pas d'information sur la structure des facteurs ; une variable proche du cercle des corrélations joue un rôle pour l'interprétation des facteurs.

Dans l'espace des échantillons, la représentation graphique permet de situer les échantillons les uns par rapport aux autres, de mettre en évidence leur dispersion et de les positionner par rapport aux axes. En regardant simultanément le plan des variables, on peut connaître l'influence des variables sur tel échantillon.

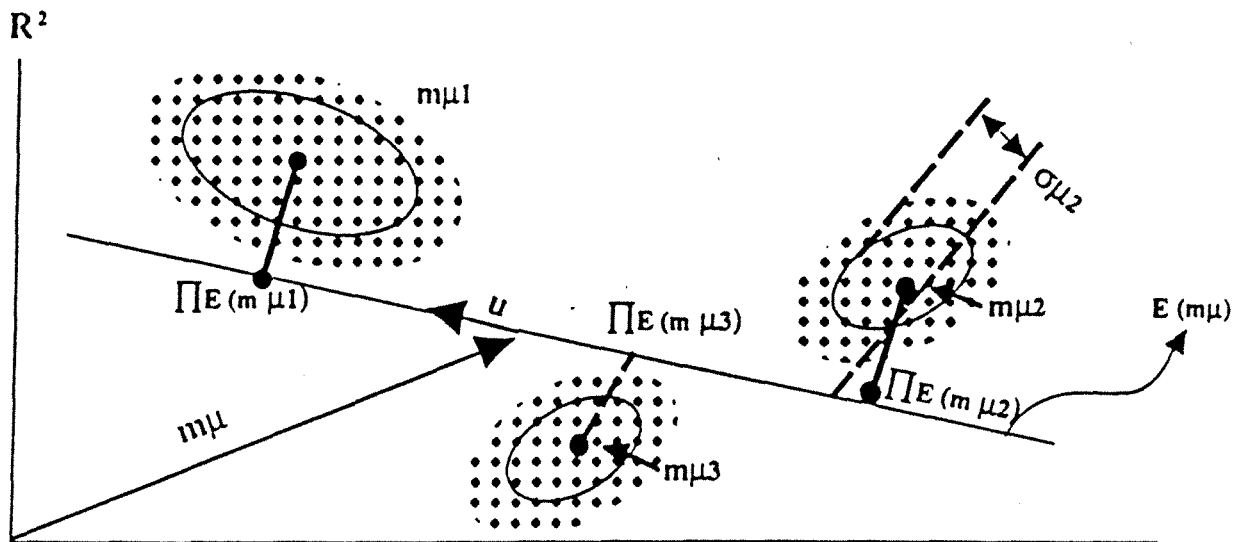
4.2.2. Analyse factorielle discriminante (Figure 79)

Une excellente revue des méthodes paramétriques (linéaire et quadratique) de la discrimination est faite dans l'ouvrage de Romeder (1973) : *Méthodes et programmes d'analyse discriminante*.

4.2.2.1. Position du problème

Soit un ensemble de N individus caractérisés par P variables. Les individus sont supposés, préalablement à l'analyse, répartis dans des **groupes définis a priori**. Le problème que l'on

ANALYSE FACTORIELLE DISCRIMINANTE



PROBLEME : Chercher une suite orthonormée ($u_1 \dots u_m$) la plus discriminante possible :

- les images $\Pi E(m\mu_i)$ sont le plus dispersées possible sur $\Pi E(R^n)$
- chacune des K classes de points est la plus rassemblée possible autour de son barycentre $\Pi E(m\mu_i)$ sur $\Pi E(R^n)$



CHERCHER DES VECTEURS u TELS QUE :

$$\frac{u^t (C_M) u}{u^t (C_\mu) u} \quad \text{MAXIMUM}$$

C_μ matrice de covariance totale

C_M matrice de covariance inter-classes (ou "between")

Théorème de HUYGENS

$$(C_\mu) = (C_M) + (C_\mu^W) \quad C_\mu^W \text{ matrice de covariance intra-classes (WITHIN)}$$



- Trouver les vecteurs propres de la matrice $C_\mu^{-1} C_M$

- Les valeurs propres mesurent le pouvoir discriminant de ces axes

se pose en discrimination est de savoir s'il est possible de différencier les classes considérées avec les variables mesurées, i.e. de discriminer les groupes a priori.

Ce problème peut se scinder en deux :

- la discrimination à but **descriptif** ;
- la discrimination à but **décisionnel**.

a - Discrimination à but descriptif

Les différentes classes d'individus étant données, on cherche à savoir si les variables mesurées permettent de les caractériser, c'est-à-dire de les individualiser. Il s'agit de déterminer le pouvoir discriminant des variables par rapport aux groupes a priori.

Ce concept de discrimination recouvre deux notions bien distinctes (la seconde impliquant la première) :

- trouver une "bonne" séparation des classes, en mettant en évidence une frontière telle que les classes se chevauchent très peu ;
- vérifier la signification statistique de cette séparation.

b - Discrimination à but décisionnel ou identification

Très souvent, ce problème fait suite au problème précédent. Une fois les classes a priori correctement discriminées par les variables mesurées, on étudie un individu sur lequel les mesures des variables sont disponibles et on s'interroge sur la classe à laquelle il appartient. Ce problème d'affectation dans l'un des groupes a priori est un problème de **classement** et non de **classification** (cf. § 4.2.3) car les classes sont définies au préalable.

Le problème à but décisionnel se rencontre lorsque l'information de la classe d'appartenance de l'individu étudié est perdue, inaccessible, ou encore inexistante (problème de la prédiction).

4.2.2.2. Discrimination linéaire

a - Notations et définitions

X : tableau des données, soient N individus $x(i)$ et P variables,

X est divisé en K classes y et **Y** est l'ensemble des classes y .

Nous supposons les individus affectés de poids m_x tous égaux à $\frac{1}{N}$. Le poids m_y de la classe y qui contient N_y individus est défini par :

$$m_y = \frac{N_y}{N}$$

On notera de la même façon la classe y et son centre de gravité.

Le nuage Y des classes est le nuage des centres de gravité affectés des poids m_y , il s'agit en fait du nuage X regroupé par paquets correspondant aux différentes classes.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x(i) \quad (\text{centre de gravité du nuage } X).$$

$$\bar{y} = \sum m_y y = \frac{1}{N} \sum N_y \sum \frac{1}{N_y} x \quad (\text{centre de gravité du nuage } Y).$$

$$\bar{y} = \bar{x}.$$

Soit u un vecteur définissant une direction de l'espace des individus \mathbb{R}^P supposé muni de la métrique euclidienne usuelle.

On peut alors définir :

- la variance totale de la direction u , notée $V_X(u)$:

$$V_X(u) = \sum_{x \in X} m_x \langle u, (x - \bar{x}) \rangle^2$$

- la variance inter-classes de la direction u , notée $V_Y(u)$:

$$V_Y(u) = \sum_{y \in Y} m_y \langle u, (y - \bar{y}) \rangle^2$$

- la variance intra-classe définie à partir de la variance de chaque classe y , dans la direction u ,

$$V_y(u), (V_y(u) = \sum_{x \in y} m_x \langle u, (x - y) \rangle^2) \text{ comme : } \sum_{y \in Y} V_y(u).$$

Ces trois quantités sont les moments d'inertie de chacun des nuages X , Y et y par rapport à leur centre de gravité respectif, le long de la direction u .

Le théorème de Huygens appliqué à ces trois quantités nous permet d'écrire :

$$V_X(u) = \sum_{y \in Y} V_y(u) + V_Y(u)$$

On retrouve l'équivalent de cette expression en considérant les matrices de variance-covariance totale T, intra-classe W, (within) et inter-classes B (between) :

$$T = \{t_{ij}\}_{\substack{i = 1, \dots, P \\ j = 1, \dots, P}}$$

$$t_{ij} = \frac{1}{N_{X \in X}} \sum (x_i - \bar{x}_i)(x_j - \bar{x}_j)$$

W (somme des K matrices d'inertie correspondant à chacun des groupes)

$$W = \{w_{ij}\}_{\substack{i = 1, \dots, P \\ j = 1, \dots, P}}$$

$$w_{ij} = \frac{1}{N_{y \in Y}} \sum_{x \in y} (x_i - y_i)(x_j - y_j)$$

$$B = \{b_{ij}\}_{\substack{i = 1, \dots, P \\ j = 1, \dots, P}}$$

$$b_{ij} = \sum_{y \in Y} \frac{N_y}{N} (y_i - \bar{x}_i)(y_j - \bar{x}_j)$$

Le théorème de Huygens s'exprime alors avec les notations matricielles :

$$T = W + B.$$

b - Recherche des axes factoriels discriminants

Le critère qui prévaut pour la définition des axes discriminants est schématisé sur la figure 79.

Il s'agit de rechercher la direction u de \mathbb{R}^P (axe factoriel ou forme linéaire) qui discrimine au mieux non pas l'ensemble X des individus mais l'ensemble Y des classes de X. La discrimination sera d'autant meilleure et aisée que :

- les barycentres des classes seront éloignés les uns des autres (variance inter-classes grande),
- les individus d'une même classe seront proches (variance intra-classe petite).

Il faut donc maximiser le rapport de deux formes quadratiques suivant :

$$\frac{u^t \cdot B \cdot u}{u^t \cdot W \cdot u}$$

Ceci revient encore à maximiser :

$$\frac{u^t \cdot B \cdot u}{u^t \cdot T \cdot u}$$

ce qui est équivalent à extraire **les valeurs propres et les vecteurs propres de la matrice $T^{-1} B$** . Les valeurs propres λ de $(T^{-1} B)$ sont toutes positives et inférieures à 1.

Le premier axe factoriel discriminant u_1 est le vecteur propre de $T^{-1} B$ correspondant à la plus grande valeur propre λ_1 , qui mesure son pouvoir discriminant. De même, on définit le deuxième axe factoriel discriminant u_2 , deuxième vecteur propre de $T^{-1} B$ et donc orthogonal à u_1 ; il constitue le meilleur facteur discriminant qui soit indépendant du premier. On extrait ainsi les axes factoriels discriminants successifs.

Lorsque $N > P > K$ (où K est le nombre de groupes a priori), il y a exactement $(K-1)$ axes discriminants.

Par contre, le nombre d'axes **significativement** discriminants est un autre problème. On peut utiliser des tests paramétriques (voir Rao, 1965) pour le résoudre.

c - Retour sur la notion de discrimination linéaire

La présentation de l'analyse discriminante que nous venons de faire repose plutôt sur des bases mécaniques. En fait, l'étude de la matrice B et du rapport $\frac{u^t \cdot B \cdot u}{u^t \cdot T \cdot u}$

revient à négliger les variations de la variance de chaque classe. La matrice W est une moyenne de ces variances qui en quelque sorte "efface" les caractéristiques propres à chaque groupe.

De façon sous jacente, nous avons donc fait l'hypothèse que les matrices de variance-covariance des K groupes sont égales ainsi que le P -uplet de variables suit une loi multivariée normale pour chaque groupe. Ces hypothèses définissent le cadre de la **discrimination linéaire**.

4.2.2.3. Méthode de classement

Il s'agit de la discrimination à but décisionnel ; nous disposons d'un individu anonyme a de \mathbb{R}^P que l'on veut affecter à l'un des groupes a priori. Il s'agit donc de trouver une partition de \mathbb{R}^P en K régions correspondant aux K classes.

La distance de a au centre y du groupe y est définie pour la métrique T^{-1} par :

$$d(a,y) = (a - y)^t \cdot T^{-1} \cdot (a - y).$$

a sera affecté à la classe y_0 telle que :

$$d(a, y_0) = \min (d(a,y) / y \in Y).$$

Les quantités $d(a,y)$ sont des fonctions quadratiques de a , mais elles ont toutes en commun le terme carré $(a^t \cdot T^{-1} \cdot a)$, on pourra donc comparer les **fonctions linéaires de a** , relatives à chaque groupe y :

$$v_y (a) = d(a,y) - a^t \cdot T^{-1} \cdot a,$$

$$v_y (a) = y^t \cdot T^{-1} \cdot (y - 2a).$$

En fait, utiliser la métrique T^{-1} pour définir $d(y,a)$ entraîne le découpage de \mathbb{R}^P en K régions définies par des **hyperplans concourants**. On est toujours dans le cadre de la **discrimination linéaire** (les lois multinormales de chacun des groupes ont même matrice de variance-covariance).

Dans le cadre de la **discrimination quadratique**, où l'hypothèse est faite que les variables suivent pour chaque groupe des lois multinormales mais de matrice de variance-covariance différente d'un groupe à l'autre, les cloisons séparatrices de l'espace \mathbb{R}^P sont des hypersurfaces du second ordre définies par des fonctions quadratiques.

D'autres procédures de classement peuvent être définies en utilisant la théorie de Bayes, à condition d'introduire les probabilités a priori d'appartenance à chacun des groupes ainsi que les coûts d'erreur suivant les groupes.

4.2.2.4. Procédures de pas à pas

a - Définition et intérêt des procédures de pas à pas

Les procédures de pas à pas permettent d'introduire les variables qui sont mesurées sur la population d'individus au fur à mesure, en fonction de leur pouvoir discriminant (par rapport aux groupes a priori). C'est d'abord la variable la plus discriminante qui est introduite, v_1 , puis on recherche une variable v_2 telle que le sous-ensemble $\{v_1, v_2\}$ soit le plus discriminant possible.

Les méthodes diffèrent :

- par le critère utilisé pour sélectionner les variables ("mesure du pouvoir discriminant d'une variable"),
- par la façon dont sont remis en cause les choix aux pas précédents pour l'étape k ; le plus souvent, au pas k , on ne remet pas en cause les $(k-1)$ variables déjà sélectionnées, ceci limite le temps de calcul ; par contre, on n'est pas assuré, au pas k , d'avoir le meilleur sous-ensemble discriminant de k variables parmi les P initiales.

Les deux points intéressants dans les procédures de pas à pas sont les suivants :

- la **réduction des "coûts" de prise en compte des variables** (problème analogue à celui de la régression linéaire multiple) ; on s'arrête au moment où l'on estime que le nombre de variables discriminantes retenues devient trop grand,
- amélioration de la **fiabilité** de la discrimination si l'on considère un petit nombre de variables ; il est toujours plus facile de trouver des cloisons séparatrices entre des groupes dans un espace de dimension élevée (beaucoup de variables) que dans un espace de faible dimension.

b - Pourcentage de bien classés

En discrimination linéaire ou quadratique, on effectue une partition de \mathbb{R}^P en K régions supposées correspondre aux K classes définies a priori.

On peut alors regarder si les individus des différents groupes se retrouvent bien dans les régions de classement correspondantes. Le pourcentage d'individus bien classés nous indique combien les groupes ont pu être séparés par la méthode utilisée. Ceci n'aide pas pour le pronostic quant à l'appartenance d'un nouvel individu à l'un des groupes.

Supposons que les échantillons d'individus de chaque groupe soient suffisamment grands pour soustraire au hasard quelques individus qui ne participeront pas à l'élaboration des régions de classement R_1, R_2, \dots, R_K :

Echantillon initial
groupe i

échantillon de base
(calcul de R_i)

échantillon test
(aide à la mesure
de la fiabilité de R_i)

Le tableau de classement de l'**échantillon test** nous fournit alors des estimations des probabilités a posteriori d'appartenance à chacun des groupes, sous une certaine hypothèse.

A partir du théorème de Bayes, et à condition que les effectifs de chaque groupe choisi pour l'échantillon test soient proportionnels aux probabilités a priori d'appartenir à chacun de ces groupes, alors les probabilités a posteriori pour un nouvel individu d'appartenir à un groupe, sachant dans quel groupe il a été classé, sont données par les éléments du tableau de classement de l'échantillon test.

Remarque : pour obtenir un échantillon test dont les effectifs de chaque groupe soient proportionnels aux probabilités a priori d'appartenir à chacun des groupes, on choisira souvent ces effectifs proportionnels aux effectifs de l'échantillon de base.

Le pourcentage de bien classés sur l'**échantillon de base** est utilisé pour sélectionner la variable la plus discriminante à chaque pas des procédures de discrimination pas à pas. C'est la variable qui maximise ce pourcentage sur l'ensemble des groupes qui est introduite.

c - Trace ($T^{-1} B$)

Ce critère, contrairement au précédent, **ne nécessite pas la définition d'une procédure de classement.**

Si l'on se reporte à l'interprétation géométrique de l'analyse discriminante, les axes discriminants sont des combinaisons linéaires des variables initiales qui maximisent l'inertie du nuage Y , calculée avec la métrique T^{-1} , relativement à son centre de gravité (on a extrait les valeurs et vecteurs propres de la matrice $T^{-1}B$).

Soient T , la matrice de covariance totale,

T_q , la matrice de covariance totale avec les q variables introduites.

Au pas q , on va chercher à maximiser :

$$\sum_{y \in Y} \left(\frac{N_y}{N} (y - \bar{y})^t \cdot T_q^{-1} \cdot (y - \bar{y}) \right),$$

ce qui revient à maximiser la trace de $(T_q^{-1} B_q)$.

Remarque : cette quantité (trace de $T_q^{-1} B_q$) peut être considérée comme un indice de séparation entre plusieurs groupes dans l'espace \mathbb{R}^p .

Pour le pas à pas, à partir des matrices T_q et B_q , on forme T_{q+1} et B_{q+1} avec les variables non encore introduites, la variable choisie est alors celle qui maximise la trace de la matrice $T_{q+1}^{-1} B_{q+1}$.

Il n'y a pas de test d'arrêt naturel, comme dans le cas du pourcentage de bien classés : la quantité trace $(T_q^{-1} B_q)$ pourra croître lors du passage du pas q au pas $(q + 1)$ sans que la discrimination soit pourtant améliorée.

d - Autres critères pour le pas à pas

D'autres critères très classiques pour le pas à pas sont basés sur des tests paramétriques pour l'analyse de la dispersion (Rao, 1965).

Le Λ de Wilks défini par :

$$\Lambda_q = \frac{|W_q|}{|T_q|}$$

au pas n° q (q variables introduites) est basé sur le test de l'égalité des vecteurs moyennes pour plusieurs groupes. Si la quantité Λ_q est suffisamment petite, il y a une différence significative entre les vecteurs moyennes des différents groupes. On retient alors au pas q l'ensemble des variables qui minimisent :

$$\frac{|W_q|}{|T_q|}.$$

En fait, $\Lambda_q = \prod_{i=1}^q v_i$

où les v_i sont les valeurs propres de $T_q^{-1} W_q$. Le critère du Λ de Wilks est donc assez proche du critère de la trace $T^{-1} B$ où l'on minimise :

$$\sum_{i=1}^q v_i.$$

Un test sur l'égalité des **moyennes conditionnelles** pour plusieurs groupes permet aussi de définir un critère pour le pas à pas.

Nous supposons que $(q-1)$ variables ont déjà été introduites et nous cherchons la q ème meilleure variable à introduire. Les matrices de variance totale T_{qq} et intra-classe W_{qq} sont décomposées de la façon suivante :

$$T_{qq} = \begin{pmatrix} T_{11} & T_{12} \\ (q-1, q-1) & (q-1, 1) \\ T_{21} & T_{22} \\ (1, q-1) & (1, 1) \end{pmatrix}$$

$$W_{qq} = \begin{pmatrix} W_{11} & W_{12} \\ (q-1, q-1) & (q-1, 1) \\ W_{21} & W_{22} \\ (1, q-1) & (1, 1) \end{pmatrix}.$$

W_{qq} et T_{qq} peuvent être considérées comme des estimations des matrices de variance-covariance respectivement sous les hypothèses H_0 (moyennes inégales) et H_1 (moyennes égales).

On remplace W_{qq} et T_{qq} par :

$$W(2/1) = W_{22} - W_{21} W_{11}^{-1} W_{12}$$

$$T(2/1) = T_{22} - T_{21} T_{11}^{-1} T_{12}$$

ce qui permet de tester les hypothèses H_0 (moyennes **conditionnelles** inégales) et H_1 (moyennes **conditionnelles** égales).

La quantité
$$\frac{T(2/1) - W(2/1)}{W(2/1)} \frac{N-K-q}{K-1}$$

suit une loi de Snedecor $\mathcal{F}(K-1, N-K-q)$.

On choisit la variable qui permet d'obtenir la plus grande valeur pour le test F.

4.2.3. *Classifications automatiques* (Figure 80)

4.2.3.1. *Introduction*

Les méthodes d'analyse factorielle déjà décrites reviennent à chercher des espaces de dimension réduite pour représenter les individus de façon à rendre l'information plus accessible ou à mieux mettre en évidence ses structures importantes. La recherche des axes engendrant ces nouveaux espaces de représentation se fait selon des critères qui varient avec le problème posé :

- minimisation de la distorsion du nuage projeté par rapport au nuage initial (A.C.P.),
- augmenter les séparations potentielles entre les groupes d'individus définis a priori (A.F.D.).

Eventuellement, l'A.F.D. permet aussi de classer des individus "anonymes" à partir de régions d'affectation définies sur des critères probabilistes. Mais l'étape prédictive de l'A.F.D. correspond à un problème de **classement** et non pas de classification.

Les techniques de classification automatique cherchent à fournir un regroupement des individus dans l'espace \mathbb{R}^P initial, en fonction de leurs proximités. Plus rarement, se pose le problème de la classification des variables ou même de la classification simultanée des variables et des individus (classification croisée).

Les résultats des classifications automatiques dépendent principalement :

- de la définition de la distance sur l'espace où sont regroupés les points (espace des individus ou espace des variables),
- de la méthode de représentation des classes qui est adoptée.

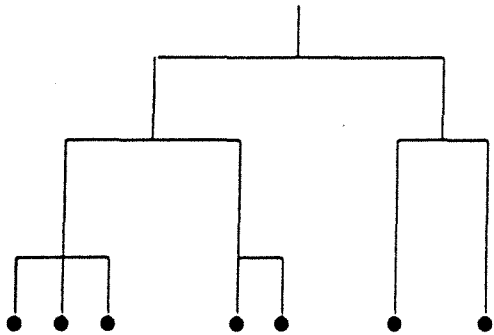
Deux grandes familles de techniques peuvent être distinguées :

- les **classifications hiérarchiques** qui peuvent être représentées sous forme d'arbres et correspondent à une suite de partitions emboîtées : ces méthodes sont riches

PRINCIPALES METHODES DE CLASSIFICATION AUTOMATIQUE

HIERARCHIQUES

Représentation en arbre



ASCENDANTE OU DESCENDANTE



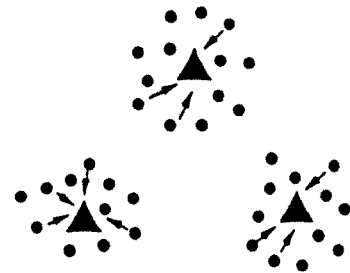
SUITE DE PARTITIONS EMBOITEES

* méthode riche mais difficultés de représentation et d'interprétation si le nombre d'individus à classer est grand

NUÉES DYNAMIQUES

Méthode Itérative

- a) Représentation d' une classe
- b) Affectation



1 PARTITION UNIQUE

* méthode rapide mais
 → il faut indiquer le nombre de groupes
 → les noyaux initiaux influent la partition finale → faire plusieurs essais

DEFINITION DE LA DISTANCE + CHOIX DE
LA METHODE DE REPRESENTATION DES CLASSES

mais si le nombre d'individus à classer est grand, on se heurte au problème de la représentation et de l'interprétation des résultats ;

- les **méthodes de partitionnement** : elles sont rapides mais il faut choisir au départ le **nombre de groupes** ainsi que les **noyaux initiaux** de la partition ; ces paramètres influencent beaucoup la partition finale ; plusieurs essais sont souvent indispensables.

Si les techniques d'analyse factorielle peuvent être considérées d'un point de vue purement géométrique ou mécanique, ceci est encore plus vraie pour les techniques de classification. La plupart des méthodes ne font appel à aucun modèle statistique (ou probabiliste), mais relèvent directement d'une approche heuristique. Ceci explique le foisonnement de ces méthodes, lié à de très nombreuses petites variantes techniques à l'origine de différents algorithmes.

4.2.3.2. *Classifications hiérarchiques*

a - Généralités

Pour classer l'ensemble des individus X , il faut se donner des indices de similarité ou de dissimilarité **entre classes** et **entre points**.

Un indice de dissimilarité entre points, $d(i,j)$, est une application dans \mathbb{R}^+ , symétrique, telle que $d(i,i) = 0, \forall i \in I$ où I est l'ensemble des points à classer.

Si d vérifie l'inégalité triangulaire, alors d est un **écart**.

Si d vérifie en plus : $d(i,j) = 0 \Rightarrow i = j$, alors d est une **distance**.

Une **distance ultramétrique** définie sur I rend tout triangle isocèle car elle vérifie la propriété suivante (plus forte que l'inégalité triangulaire) :

$$\forall i, j, k \in I, \quad d(i,j) \leq \max(d(i,k), d(k,j)).$$

Les classifications hiérarchiques peuvent être **ascendantes** ou **descendantes**.

En classification hiérarchique ascendante, si N est le nombre d'individus à classer :

- on agrège les deux points les plus proches,
- on dispose alors de $(N-1)$ classes dont $(N-2)$ sont réduites à un élément unique ; on agrège alors les **deux classes les plus proches**, on obtient ainsi $(N-2)$ classes.

On réitère le procédé jusqu'à ce que l'on obtienne une seule classe contenant tous les objets. Les méthodes ascendantes sont les plus utilisées.

Les méthodes descendantes consistent à segmenter progressivement le nuage initial de points en classes contenant un nombre d'éléments différents jusqu'à avoir N classes, toutes étant des singletons. Elles sont beaucoup plus coûteuses en temps machine.

b - Classification hiérarchique basée sur le critère de l'inertie expliquée (Ward)

Soit I l'ensemble des individus affectés des poids M_i .

\bar{x} est le centre de gravité du nuage de points.

L'inertie totale du nuage de points $V(I)$ est définie par :

$$V(I) = \sum_{i \in I} M_i d^2(i, \bar{x}).$$

Nous rappelons que :

$$D^2 = \sum_{i \in I} M_i \|i - \bar{x}\|^2$$

est la variance isotrope.

Soit P une partition sur I :

$$P = (P_1, P_2, \dots, P_K)$$

$$V(I) = \sum_{l=1}^K \mu_l d^2(y_l, \bar{x}) + \sum_{l=1}^K V_l(P_l)$$

où y_l est le centre de gravité de la classe l de masse μ_l et $V_l(P_l)$ est la variance de la classe P_l .

Remarque : il s'agit du théorème de Huygens.

A chaque étape de construction de la hiérarchie, on agrège les deux classes dont l'inertie inter-classe est minimale.

A l'étape k, en partant d'une partition en (n-k) classes P_1, P_2, \dots, P_{n-k} , on agrège les deux classes P_n et $P_{n'}$ telles que :

$$\frac{\mu_n \mu_{n'}}{\mu_n + \mu_{n'}} d^2(y_n, y_{n'})$$

soit minimale.

La classification ascendante hiérarchique par agrégation suivant la variance peut facilement se comparer avec l'analyse en composantes principales, en projetant les classes obtenues sur les axes factoriels.

c - Limitations des méthodes de classification hiérarchique

Les temps de calcul sont assez longs. D'autre part, les arbres hiérarchiques deviennent difficiles à interpréter dès que l'ensemble à classer a un grand nombre d'éléments.

Enfin, il faut signaler l'**effet de chaîne** de la classification hiérarchique : il suffit qu'il existe entre deux points même très éloignés une suite de points très proches qui les relie pour que ces deux points soient considérés comme proches.

4.2.3.3. Méthodes non hiérarchiques ou de partitionnement

a - Généralités

Ces méthodes consistent à trouver une partition de l'espace des individus en k classes, k étant défini au préalable.

Ces méthodes sont très rapides et facilement interprétables, en particulier sur de grands ensembles de données ; elles sont essentiellement itératives, l'idée centrale étant de choisir une partition initiale de l'ensemble à classer et de déplacer ses éléments d'une classe à l'autre pour obtenir une partition meilleure au sens d'un critère défini a priori.

Outre le paramètre k , le choix de la partition initiale est aussi très influent sur les résultats finaux, donc, comme il a été déjà signalé dans l'introduction, plusieurs essais de partitionnement sont souvent nécessaires pour valider la structure des groupes trouvés.

b - Méthode "K-means" ou algorithme de "réallocation-recentrage" autour des centres mobiles de gravité

On définit initialement un ensemble quelconque de k points $\{\theta_1^0, \theta_2^0, \dots, \theta_k^0\}$ correspondant au système initial des "centres d'attraction".

Ces centres permettent de définir une première partition de l'espace des individus I en $k(1)$ classes $\{I_1^1, I_2^1, \dots, I_{k(1)}^1\}$ où $k(1) \leq k$.

En effet :

soient deux centres $\theta_j^0, \theta_{j'}^0$, avec $j < j'$, le nuage I est découpé en deux morceaux :

$$I_{jj'} = \{ x / d(x, \theta_j^0) \leq d(x, \theta_{j'}^0) \}$$

$$I_{j'j} = \{ x / d(x, \theta_{j'}^0) < d(x, \theta_j^0) \}.$$

La zone attribuée à θ_j^0 se trouve définie par l'intersection :

$$Z(\theta_j^0) = \left(\bigcap_{j' > j} I_{jj'} \right) \cap \left(\bigcap_{j'' < j} I_{jj''} \right).$$

$Z(\theta_j^0)$, s'il n'est pas vide, définit l'une des classes I_j^1 , $1 \leq j' \leq k(1)$. A chacune des classes I_j^1 , on associe son centre de gravité formant ainsi une nouvelle suite des centres d'attraction $(\theta_1^1, \theta_2^1, \dots, \theta_{k(1)}^1)$. L'ensemble des centres ainsi obtenus est de cardinal inférieur ou égal à $k(1)$: en effet, il se peut qu'un même centre de gravité corresponde à plus d'une seule classe.

Puis, on réalise une **nouvelle décomposition** de I en classes dont les centres de gravité définiront les nouveaux centres d'attraction et ainsi de suite.

Le processus dont chaque pas diminue le critère :

$$\sum_{1 \leq j \leq h} \sum_{x_i \in I_j} M_i \|x_i - \bar{x}_j\|^2,$$

où h est le nombre de groupes au pas considéré, **converge** nécessairement. En fait, on espère que l'ensemble des derniers centres viendra occuper les régions à forte "densité" du nuage I.

Cet algorithme a entraîné toute une famille de méthodes basées sur le principe de "réallocation-recentrage". Ces méthodes diffèrent sur les deux niveaux suivants :

- choix ou détermination du système initial de centres d'attraction ou noyaux,
- affectation des éléments aux classes en cours de formation.

Ces variantes peuvent correspondre à des aspects algorithmiques ou statistiques (définition de critères).

Signalons par exemple la procédure de Mc Queen (1967) qui a introduit la dénomination des "K-means" : après chaque affectation d'un sommet à la classe -en cours de formation -dont le centre de gravité est le plus proche, on substitue au précédent centre d'attraction de la classe d'accueil le centre de gravité de la classe enrichie de cet élément.

Remarque : si k est le nombre de classes fixé au départ, le nombre de classes obtenu finalement résulte, en fait, de l'aléa du déroulement de l'algorithme. En pratique, il reste très proche de k .

c - Méthode des "nuées dynamiques"

E. Diday a introduit en France les méthodes "A.N.D." (algorithme des nuées dynamiques) en 1972 en y apportant une certaine formalisation et en cherchant à généraliser au maximum leur portée.

La généralisation porte d'abord sur la notion de **représentation d'une classe** qui ne se fait plus nécessairement au moyen d'un centre de gravité, mais à partir de la définition d'un "noyau" de la classe, lequel peut, par exemple, correspondre à un sous-ensemble de faible cardinal de la classe. Ce dernier, définissant un "squelette" de la classe, "épouserait" mieux sa forme que ne le ferait un centre de gravité.

Le deuxième aspect de l'extension de l'algorithme porte sur la **nature de l'espace de représentation** des données (d'ailleurs, la notion de "noyau" dépend de cet espace et du problème de reconnaissance posé).

Enfin, un problème important de métriques et critères se pose pour alimenter de façon cohérente ces développements algorithmiques.

Les deux étapes importantes de la méthode sont :

- une étape de **représentation des classes**,
- une étape d'**affectation**.

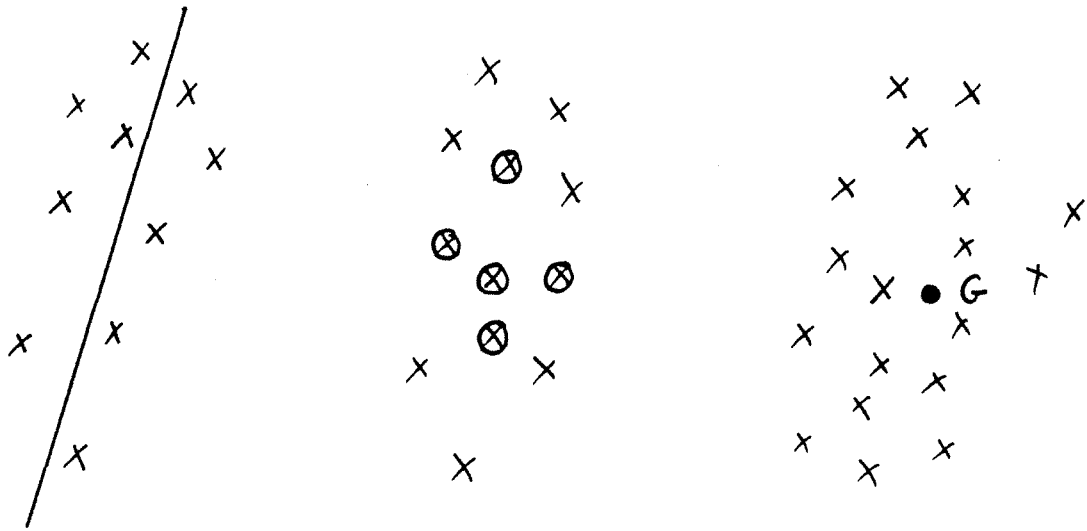
Le critère à optimiser en M.N.D. exprime l'adéquation entre une classification des objets et un mode de représentation des classes de cette classification. Le problème d'optimisation est posé en terme de recherche **simultanée** de la classification et de la représentation des classes de cette classification, parmi un ensemble de classifications et de représentations possibles, qui optimisent le critère.

Nous allons expliquer la méthode sur un exemple simple. Nous recherchons une partition en k classes d'individus (k étant donné a priori) bien agrégées et bien séparées entre elles.

• **Choix d'un mode de représentation**

L'algorithme des N.D. nécessite la définition d'un mode de représentation symbolique de tout groupe d'individus, on parle aussi de "noyau". Ce noyau peut être :

- une droite,
- un groupe de points de la population,
- un centre de gravité.



. Déroulement de l'algorithme

On se place dans le cas simple où le noyau est un point.

On part d'un choix de k noyaux estimés ou tirés au hasard parmi une famille de noyaux admissibles appelée **espace de représentation** et notée L .

Chaque point de la population est ensuite affecté au noyau dont il est le plus proche. On obtient ainsi une partition en k classes dont on calcule les noyaux. On recommence le procédé avec les nouveaux noyaux et ainsi de suite. On démontre que, sous certaines conditions, l'algorithme converge vers une position stable en améliorant, à chaque itération, un critère mathématique.

Avec des conditions portant sur les fonctions qui permettent d'affecter les points aux classes et de calculer les noyaux, cet algorithme fait décroître un critère W qui mesure l'adéquation entre les classes et leurs noyaux respectifs (autrement dit la "ressemblance" des noyaux à leur classe).

Le critère est une application $W : L_k \times P_k \rightarrow \mathbb{R}^+$

avec $L_k = \Omega^k$ ensemble des k -uplets $L (L_1, L_2, \dots, L_k)$ avec $L_i \in \Omega$,

P_k , ensemble des partitions $P = (P_1, P_2, \dots, P_k)$ à k classes de Ω ,

$$W(P,L) = \sum_{i=1}^k D(L_i, P_i) \text{ où}$$

D mesure l'adéquation du noyau L_i à la classe P_i (une petite valeur de D exprime une bonne adéquation).

A chaque itération de l'algorithme, la décroissance du critère exprime une augmentation globale de l'adéquation entre les classes et leurs noyaux.

Il est très important de noter que la méthode des nuées dynamiques, comme la plupart des méthodes de partitionnement, fournit une solution dépendant de la **position initiale** et nécessite le **choix du nombre de classes**.

Ces deux problèmes ne sont pas résolus. En général le nombre de classes est demandé à l'utilisateur et l'initialisation est faite par tirage au hasard.

Remarque 1 - La méthode des "centres mobiles" décrite précédemment (b) est un cas particulier des nuées dynamiques où \mathbb{R}^P est muni de la métrique euclidienne et où g , fonction reliant la partition à sa représentation, vaut :

$g (P_1, P_2, \dots, P_k) = (L_1, L_2, \dots, L_k)$ où L_i est le centre de gravité de chaque partie P_i , f , fonction reliant les noyaux à la partition correspondante, vaut :

$f (L_1, L_2, \dots, L_k) = (P_1, P_2, \dots, P_k)$ avec

$P_i = \{ x \in \Omega / d(x, L_i) \leq d(x, L_j) \text{ et } i < j \text{ en cas d'égalité} \}$.

Dans ce cas-là, le critère $W(P,L)$ s'exprime par :

$$W(P,L) = \sum_{i=1}^k \sum_{x_j \in P_i} p_j d^2_M(x_j, L_i).$$

$$W(P,L) = \sum_{i=1}^k I_i$$

est la somme des inerties des classes P_i par rapport à leur centre de gravité L_i .

Le critère $W(P,L)$ est donc l'inertie intra-classe de la partition P , que l'on cherche à minimiser (on **maximise** donc l'inertie inter-classe, cf. théorème de Huygens). Cette méthode a donc tendance à chercher des classes rondes, d'égaux volumes et de faible inertie. Elle est cependant très intéressante de par son efficacité pour traiter les données très abondantes.

Remarque 2 - Les méthodes de type ISODATA cherchent une partition en affectant les objets aux centres des classes dont ils sont les plus proches puis recommencent le processus en

découpant ou en regroupant les classes suivant l'importance de leur variance à l'aide de seuils. Ces algorithmes n'entrent pas dans le cadre des nuées dynamiques.

d - Recherche des classes typiques par les "formes fortes"

Le plus important en classification est d'extraire des classes **typiques**, souvent plongées dans un magma aux frontières diffuses d'individus peu caractéristiques.

Ces classes typiques ne constituent pas toujours une partition de la population.

Pratiquement, on peut les faire apparaître par la confrontation de plusieurs "bonnes" partitions au sens d'un critère à optimiser donné.

La solution donnée par la méthode des nuées dynamiques correspond à des optima locaux du critère. Elle peut différer suivant la partition ou les noyaux choisis au départ de l'algorithme. A partir de plusieurs solutions, on va donc rechercher des classes **stables** : des ensembles d'individus qui restent groupés quelque soit le tirage de départ (d'où le nom de "formes fortes").

4.2.4. Conclusions

Les méthodes que nous avons utilisées dans la méthodologie de reconnaissance automatique du faciès sismique sont essentiellement l'analyse en composantes principales, l'analyse factorielle discriminante linéaire, les classifications automatiques type "nuées dynamiques" avec un critère basé sur l'inertie intra-classe.

Ces choix seront expliqués dans l'exposé de la méthodologie ainsi que dans le paragraphe 4.4 : discussion et conclusions.

4.3. Méthodologie statistique développée

4.3.1. Contraintes à prendre en compte

Il faut déterminer des groupes de traces sismiques homogènes du point de vue morphologique et fiables du point de vue statistique. Ceci implique une présélection des variables à prendre en compte (cf. problème déjà discuté au paragraphe 4.2.2.4 de la fiabilité de l'analyse factorielle discriminante). La détermination des groupes par des méthodes de partitionnement (type nuées dynamiques) suppose connu à l'avance le nombre de groupes existant dans la population, paramètre qui est a priori inconnu par l'interpréteur sismicien.

D'autre part, la sélection des variables peut être une exigence liée à une autre raison que la seule fiabilité des analyses statistiques. En effet, supposons que nous disposions de puits sur la zone

étudiée, qui présentent des variations géologiques dont nous souhaitons connaître l'extension spatiale, au travers de l'analyse des données sismiques. Alors, non seulement nous nous intéressons aux divers faciès sismiques (variations morphologiques des traces) mais surtout aux variations qui traduisent les évolutions géologiques recherchées (cf. exemple présenté sur la figure 76). Il est certain, dans ce cas là, que les paramètres ne sont pas tous pertinents pour caractériser le faciès sismique en liaison avec les propriétés géologiques.

S'il est possible de répartir les puits en classes exprimant des discontinuités dans les propriétés géologiques étudiées, on peut ainsi avoir une indication quant au nombre de groupes qu'il est souhaitable de trouver dans la population des traces sismiques. Une phase d'apprentissage utilisant l'information géologique des puits avec l'analyse factorielle discriminante pas à pas, pour la sélection des variables, apparaît assez naturelle. Mais, il faut obligatoirement contrôler les résultats en analyse discriminante par d'autres méthodes statistiques multivariées afin de vérifier le bien-fondé de l'information a priori.

Deux catégories de problèmes existent, pour la reconnaissance automatique des faciès sismiques :

- il existe des puits (ou tout autre source d'information géologique externe) sur la zone et l'interpréteur s'attache aux faciès sismiques en vue de **prédire les évolutions géologiques** par rapport aux données observées aux puits ;
- il n'existe aucun puits sur la zone d'étude, l'interpréteur souhaite reconnaître les faciès sismiques dans un **but purement descriptif**.

Dans les deux cas, la méthodologie proposée fait appel à **plusieurs techniques multivariées** afin de valider les résultats obtenus par un faisceau d'éléments convergents.

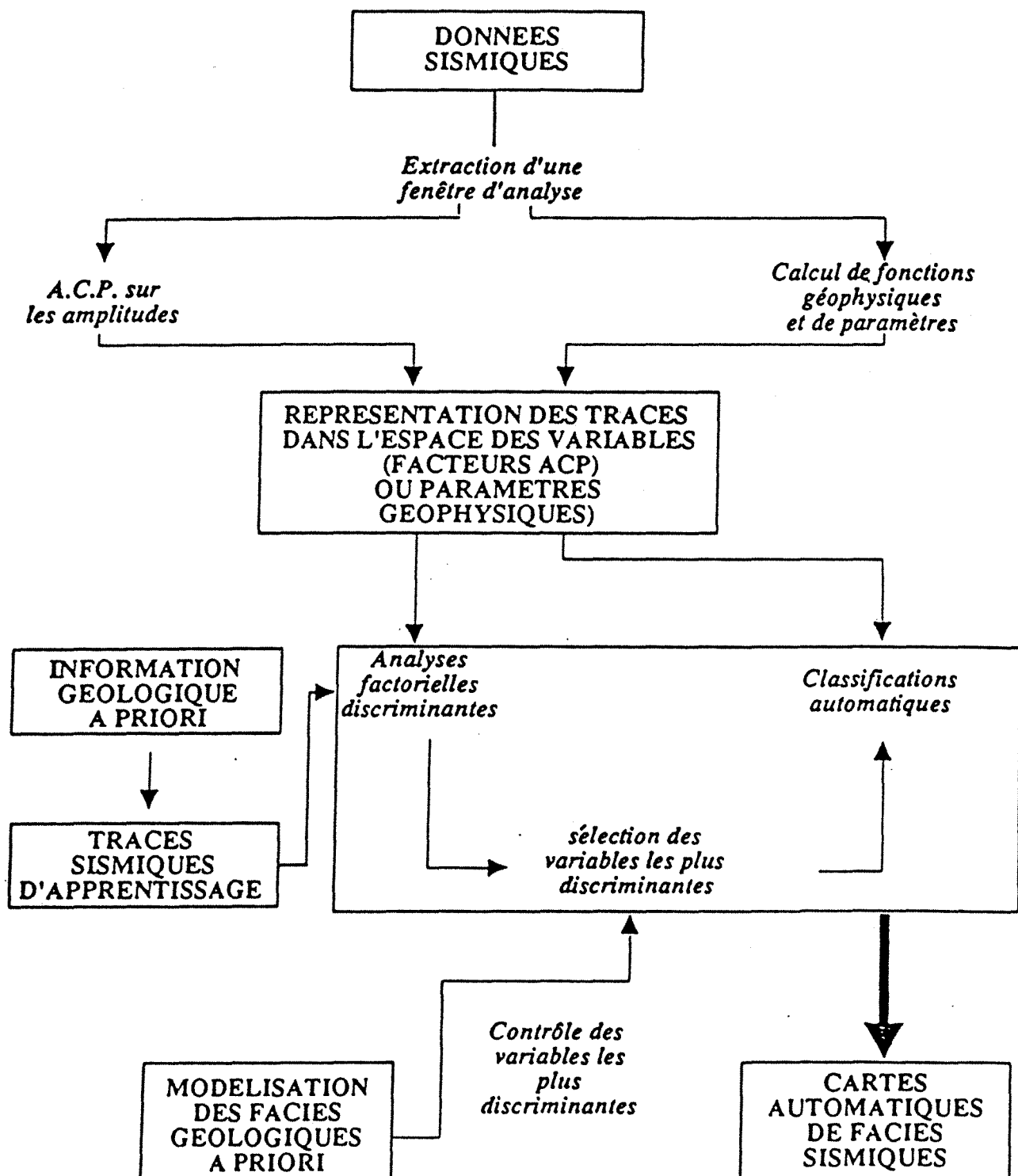
4.3.2. Cas où il existe des puits (Figure 81)

C'est la situation la plus courante d'après l'expérience acquise sur l'analyse d'une quinzaine de cas industriels. En cours d'appréciation du gisement, on dispose de quelques puits (≈ 5 à 15) et l'on s'attache aux variations géologiques majeures : mise en évidence de zones compactes dans le faciès réservoir étudié, variations du rapport sables/argiles entre les puits situés au centre d'un chenal et ceux situés sur les bords (faciès de levées, plus argileux).

En cours de développement du gisement, le nombre de puits disponibles est souvent beaucoup plus grand, mais les problèmes géologiques étudiés au travers du faciès sismique sont plus fins : variations de porosité du réservoir, présence de laminations argileuses...

Nous supposons que les puits disponibles ont été répartis en **classes de faciès géologiques**. Ceci suppose que le géologue a pu trouver des discontinuités significatives dans

METHODOLOGIE ALIAS A



les propriétés géologiques étudiées, ce qui n'est pas toujours facile. Il y a une part de subjectivité importante dans cette répartition des puits en classes qu'il est bon de garder présente à l'esprit.

Par exemple, en environnement turbiditique argilo-gréseux, on pourrait proposer trois classes de puits en fonction du rapport sables/argiles qui évolue avec la position du puits par rapport au lobe gréseux turbiditique :

- faciès bon réservoir, très gréseux (position centrale),
- faciès réservoir moyen, avec des grès moins massifs (position intermédiaire),
- faciès mauvais réservoir avec de nombreuses laminations argileuses (zone de levées du chenal).

4.3.2.1. Définition de traces sismiques d'apprentissage

Il importe de savoir si les traces sismiques de la zone d'étude présentent des variations en liaison avec les classes géologiques, puis dans ce cas là, de sélectionner les variables les plus discriminantes.

Pour cela, nous réalisons une phase d'apprentissage avec une analyse factorielle discriminante sur des groupes de traces supposées représenter les principaux faciès. A partir de chaque groupe de puits, nous définissons un ou des puits typiques et choisissons des traces adjacentes à ces puits. L'hypothèse faite est que ces traces sont typiques des faciès géologiques reconnus aux puits voisins.

Combien de traces d'apprentissage par groupe faut-il choisir ? Il faut trouver un compromis entre un nombre de traces suffisamment important pour la fiabilité statistique des analyses et rester cependant au voisinage immédiat des puits. Si des traces trop éloignées du puits interviennent dans le groupe d'apprentissage, des variations géologiques importantes (par rapport au faciès du puits) peuvent alors être intégrées dans l'analyse.

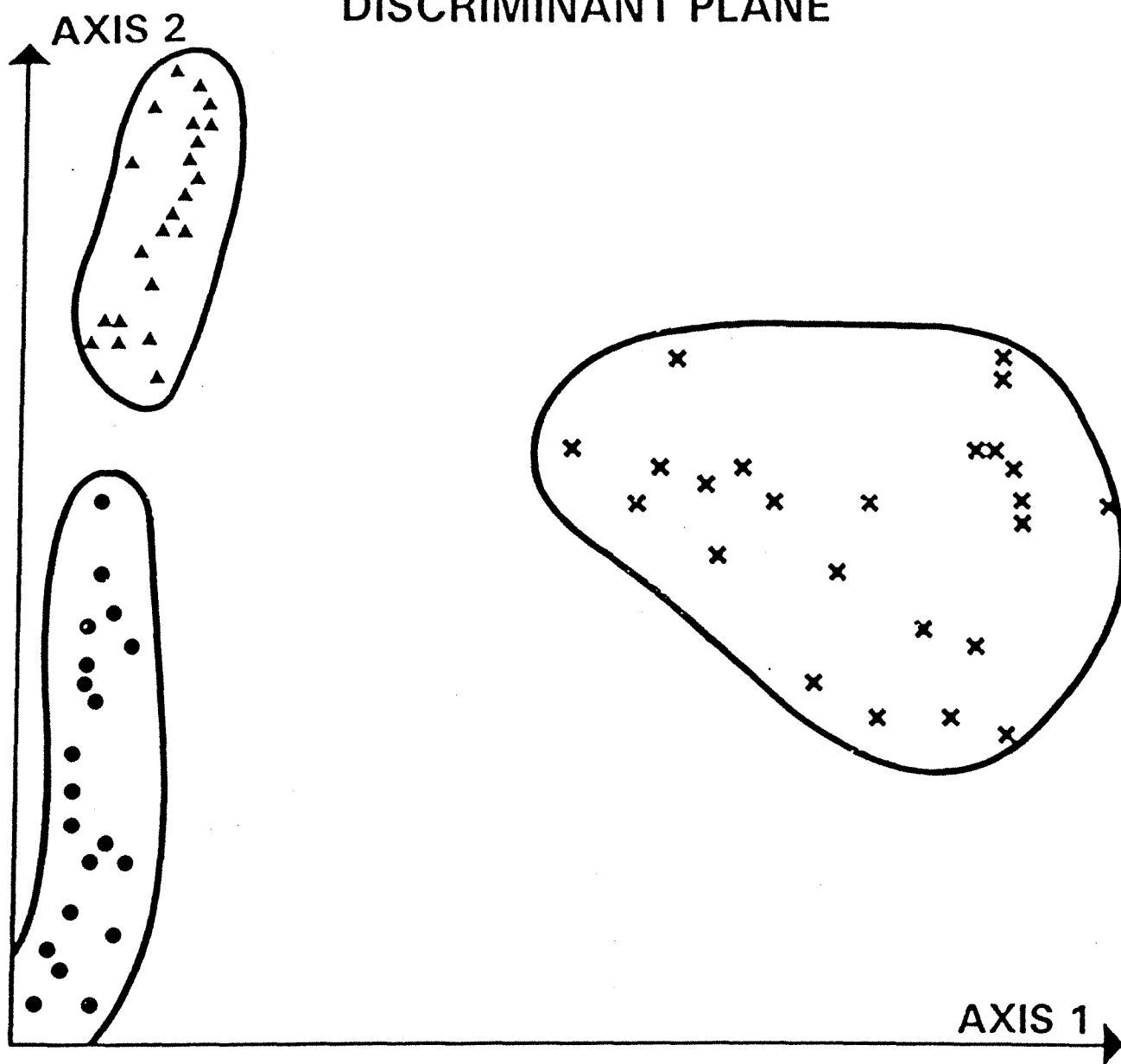
Une analyse factorielle discriminante (A.F.D.) pas à pas permet alors de choisir les variables les plus discriminantes et donne des indications quantitatives sur la séparation des groupes d'apprentissage (Figure 82) (pourcentage de bien classés sur l'échantillon de base) ainsi que sur la valeur prédictive de l'analyse (pourcentage de bien classés sur l'échantillon test). Nous rappelons que, dans le programme utilisé, le critère de sélection des variables est basé sur la trace de la matrice $T^{-1}B$.

Cette étape en analyse discriminante doit permettre de tester divers groupes d'apprentissage (en changeant par exemple les puits "typiques" retenus) et de retenir in fine les traces d'apprentissage les mieux individualisées.

CHOICE OF THE DISCRIMINANT PARAMETERS

DISCRIMINANT ANALYSIS

DISCRIMINANT PLANE



▲ _____ LEARNING TRACES NEARBY A12, SANDY WELL

● _____ LEARNING TRACES NEARBY A3, INTERMEDIATE WELL

x _____ LEARNING TRACES NEARBY A6, SHALY WELL

Une analyse en composantes principales (A.C.P.) sur les traces d'apprentissage permet de contrôler la séparation trouvée en A.F.D. avec les paramètres discriminants (Figure 83). Aucune information a priori n'est utilisée en A.C.P. Si les paramètres discriminants séparent avec fiabilité les groupes de traces d'apprentissage, on doit retrouver cette structure du nuage de points sur les plans factoriels de l'A.C.P. (Figure 83).

Si les résultats de l'analyse factorielle discriminante sont mauvais (faible séparation entre les groupes d'apprentissage), trois explications sont possibles :

- les traces d'apprentissage ne sont pas représentatives des faciès géologiques, soit que des variations latérales rapides existent entre les puits et les traces retenues, soit que les puits eux-mêmes ne soient pas réellement typiques de leur classe géologique : il peut être alors nécessaire de rediscuter de leur choix avec le géologue responsable de la zone ;
- les variations géologiques peuvent ne pas influencer la réponse sismique, qui ne permet donc pas de les mettre en évidence ;
- les variations géologiques marquent la trace sismique mais les paramètres mesurés sont impuissants à en rendre compte.

La première difficulté peut dans certains cas être levée en changeant les traces d'apprentissage. Les deux autres difficultés sont rédhibitoires pour la méthode de reconnaissance automatique des faciès sismiques.

Une fois les traces d'apprentissage choisies, les variables discriminantes sélectionnées, une étape de vérification de leur signification géologique est menée.

4.3.2.2. Validation du choix des paramètres discriminants

Les paramètres discriminants choisis sont fiables **statistiquement** mais nous souhaitons vérifier qu'ils sont bien liés aux variations géologiques qui nous intéressent.

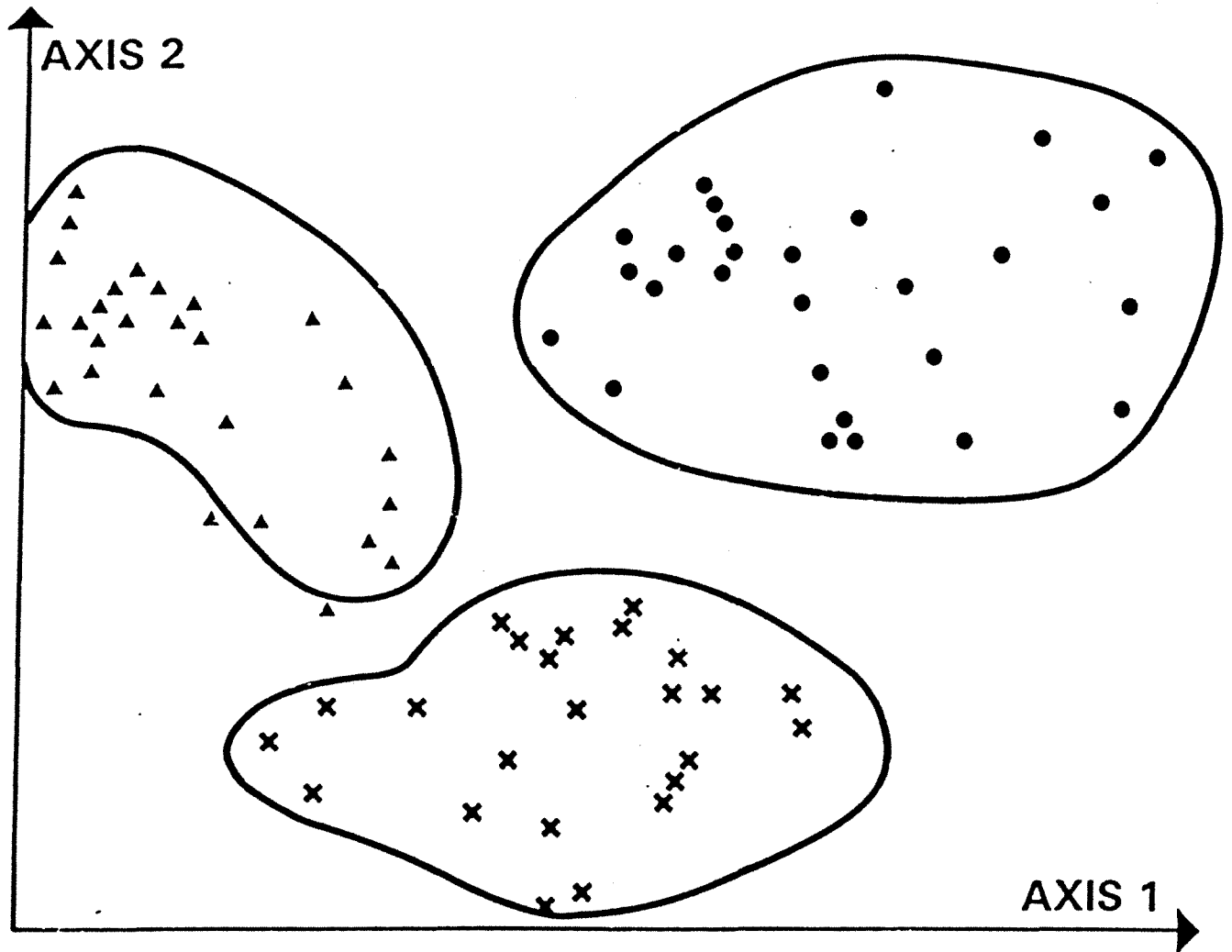
Pour cela, nous définissons à partir des puits qui présentent les variations géologiques les plus typiques, un modèle géologique représentant ces évolutions. Les couches principales sont corrélées entre les puits, les enregistrements Sonic et Densité aux puits définissent les impédances acoustiques des formations. Entre les puits et en suivant les corrélations d'horizons faites, ces impédances sont interpolées linéairement.

A partir du modèle en "impédances acoustiques", la réponse sismique synthétique est calculée par convolution 1D. Une fenêtre d'analyse est extraite qui est identique à celle prélevée sur les données réelles. Les paramètres discriminants sont alors calculés et nous vérifions que leur

VALIDATION OF THE LEARNING TRACES CHOICE

PRINCIPAL COMPONENTS ANALYSIS

PRINCIPAL PLANE 1-2



- ▲ _____ LEARNING TRACES NEARBY A12, SANDY WELL
- _____ LEARNING TRACES NEARBY A3, INTERMEDIATE WELL
- × _____ LEARNING TRACES NEARBY A6, SHALY WELL

évolution entre les puits typiques est analogue à celle qui est observée sur les traces réelles. Ils doivent être significativement différents pour les traces synthétiques adjacentes aux divers puits contrôlant la modélisation.

Dans la fenêtre temporelle analysée, plusieurs phénomènes géologiques peuvent expliquer la signature sismique. Nous pouvons essayer de détecter quel est l'événement géologique majeur qui l'explique. Pour cela, nous reprenons le modèle synthétique initial que nous modifions : par exemple, nous supprimons une des causes géologiques de variation acoustique sur l'intervalle d'étude. Nous recalculons la réponse sismique puis les paramètres discriminants et étudions comment leur comportement a été affecté par cette modification. Si leur comportement est inchangé, ce n'est pas ce phénomène qu'ils mesurent. En modifiant ainsi séparément les divers éléments géologiques intervenant dans la réponse sismique, on en détecte un certain nombre qui expliquent le comportement discriminant des paramètres sismiques sélectionnés dans la phase d'apprentissage sur les traces réelles.

Cette étape d'analyse du modèle synthétique conforte l'intérêt des paramètres sismiques discriminants pour la caractérisation géologique du faciès sismique.

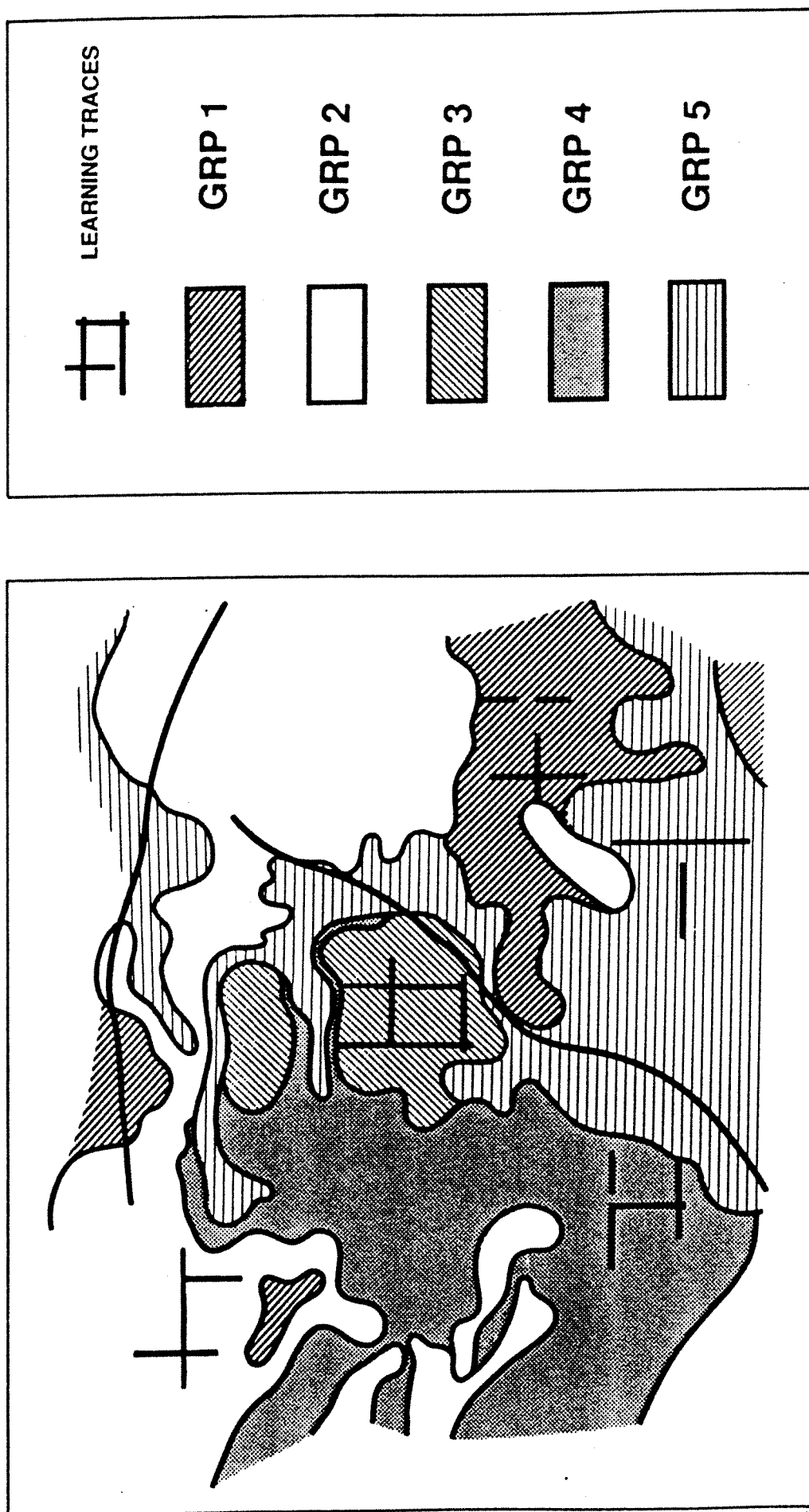
Si le comportement des variables sismiques observé sur les données réelles ne se retrouve pas sur les données synthétiques, il faut alors remettre en question la signification géologique des variations détectées (cf. paragraphe 4.4 sur les limites de la méthode). Ces variations sont détectées avec **fiabilité** mais elles ne sont probablement pas liées au problème géologique qui nous intéresse.

4.3.2.3. Etape prédictive

Les paramètres sismiques discriminants ont été sélectionnés au cours de la phase d'apprentissage et leur signification géologique a été vérifiée. Ils sont maintenant calculés sur toutes les traces sismiques qui ne sont pas intervenues dans les groupes d'apprentissage ("traces sismiques anonymes"). Ces traces sont alors affectées à l'un des groupes d'apprentissage, en fonction de leur proximité, au cours de l'étape prédictive d'une analyse factorielle discriminante **linéaire**; la répartition spatiale des affectations des traces dans les groupes est cartée. La carte de faciès sismique obtenue (Figure 84) prétend traduire les principaux faciès géologiques observés aux puits (puisque une phase d'apprentissage a eu lieu).

Les cartes donnant les probabilités de mauvaise affectation de chaque trace sont aussi représentées. Elles aident à l'interprétation de la carte des faciès sismiques, mettant en valeur les zones où l'affectation est presque certaine, ou bien, au contraire, les zones où les traces sont loin de tous les groupes d'apprentissage, ces zones sont donc moins fiables.

Fig. 84



4.3.2.4. Validation des faciès sismiques par classifications automatiques

La carte de faciès obtenue en analyse factorielle discriminante (Figure 84) repose sur une information a priori qu'il est important de contrôler : nous avons supposé que les puits se répartissaient en n groupes géologiques et que toutes les traces sismiques analysées pouvaient se rattacher à l'un de ces groupes.

Or, l'établissement de coupures géologiques à partir des données disponibles aux puits n'est pas toujours facile (nombre de classes géologiques à former...). Il n'est pas non plus certain que les puits explorent **toutes** les variations géologiques existant sur la zone couverte par les données sismiques : des "groupes géologiques" peuvent ne pas avoir été reconnus. Des variations peuvent affecter les traces et ne pas être en rapport avec les propriétés géologiques (traces bruitées). Ces traces sismiques peuvent alors constituer des classes supplémentaires.

Pour toutes ces raisons, nous effectuons des tests en classification automatique sur l'ensemble des traces sismiques, caractérisées soit par les paramètres discriminants sélectionnés sur les groupes d'apprentissage, soit par les premiers facteurs d'une analyse en composantes principales effectuée sur la totalité des paramètres sismiques (afin d'éviter l'influence du choix des traces d'apprentissage). Dans ces tests, nous faisons varier le nombre de groupes (nous rappelons que la méthode utilisée est de type "nuées dynamiques") et pour un nombre de groupes fixé, nous faisons varier les noyaux initiaux des groupes. Ce dernier test donne une idée de la stabilité des groupements trouvés, pour un nombre de groupes fixé.

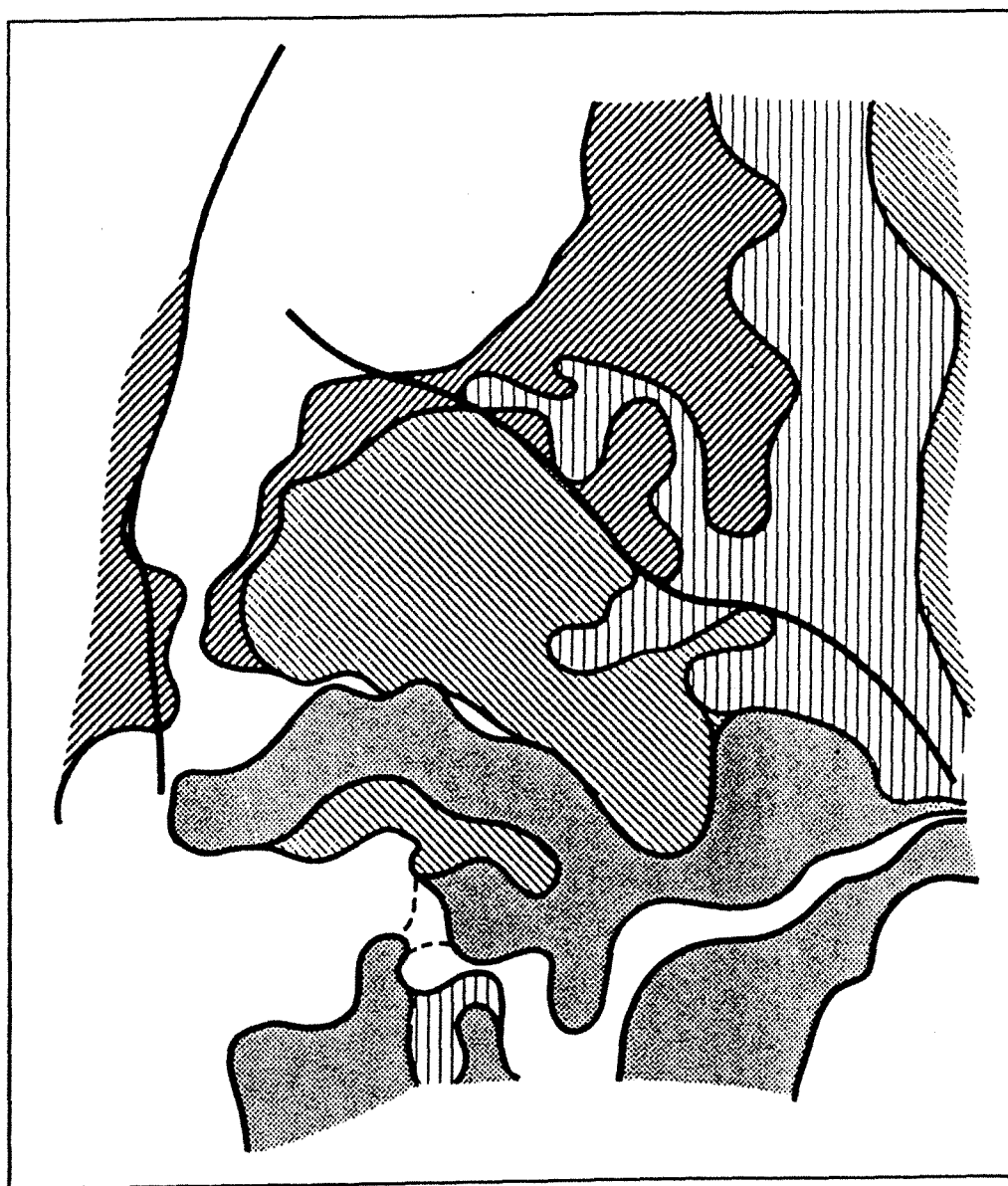
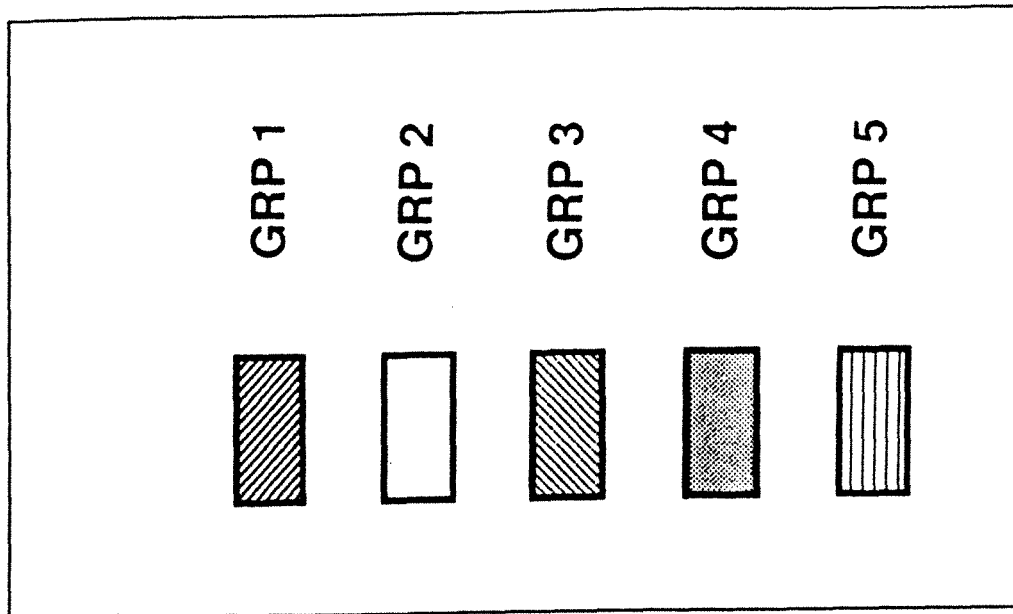
Finalement, on obtient plusieurs cartes de faciès sismiques, correspondant chacune au nombre de groupes demandé dans la classification automatique (Figure 85, 5 groupes, par exemple).

Ces cartes sont à comparer entre elles ainsi qu'à interpréter avec la carte obtenue en analyse factorielle discriminante sur n groupes d'apprentissage (Figure 84). Les zones de faciès sismiques doivent se recouper en majeure partie (Figures 84 et 85). L'analyse des groupes sismiques supplémentaires peut permettre d'interpréter des zones où les probabilités de bonne classification, en analyse factorielle discriminante, étaient assez faibles.

4.3.2.5. Synthèse de la méthodologie (information géologique disponible) (Figure 81)

L'information géologique disponible est intégrée dans le traitement des données sismiques par l'intermédiaire d'une phase d'apprentissage sur des traces sismiques, sélectionnées au voisinage des puits les plus typiques, et supposées représenter le faciès géologique.

La méthode clé est l'analyse factorielle discriminante pas à pas qui permet de sélectionner les paramètres discriminants pour le problème posé puis d'affecter les traces anonymes dans les groupes prédéfinis.



Une étape de modélisation permet de confirmer l'intérêt géologique des paramètres sismiques sélectionnés.

D'autres méthodes viennent confirmer les résultats principaux : analyse en composantes principales (pour le choix des traces d'apprentissage) et classifications automatiques (détermination des faciès sismiques).

4.3.3. Cas où il n'y a pas d'information géologique disponible

Il s'agit simplement d'établir des **cartes automatiques de faciès sismiques**, sans rechercher directement une caractérisation géologique. Sans information a priori pour guider le choix des groupes au départ (nombre et noyaux), la méthode est basée essentiellement sur des classifications automatiques des traces avec l'ensemble des variables, ou mieux avec les premiers facteurs d'une A.C.P. sur les variables (afin de limiter le nombre de variables). Plusieurs tirages des noyaux initiaux sont faits (stabilité des groupes) et nous étudions des critères quantitatifs pour déterminer le nombre de groupes qui fournit la meilleure partition, au sens statistique, de la population.

Comme nous ne sommes pas assurés que toutes les variables sont significatives ou indispensables pour caractériser la morphologie des traces et parce qu'un trop grand nombre de variables nuit aux résultats des méthodes de partitionnement, nous choisissons le meilleur regroupement pour définir des groupes de traces sismiques typiques. Par rapport à cette meilleure partition, nous prenons les traces proches du barycentre de chaque groupe.

Ces groupes constituent les groupes a priori dans une analyse factorielle discriminante pas à pas. Un petit nombre de variables discriminantes est alors sélectionné et l'étape prédictive de l'analyse fournit des affectations pour l'ensemble des traces sismiques qui doivent corroborer, la plupart du temps, les résultats de la meilleure partition. Là encore, l'A.F.D. apporte des éléments complémentaires quant à la fiabilité des regroupements.

La méthodologie sans information a priori est beaucoup plus légère que celle proposée quand une information géologique est disponible : c'est en effet l'étape de sélection des puits typiques et des traces d'apprentissage qui peut être très longue.

4.4. Discussion sur les limitations de la méthodologie proposée - Conclusions

Un premier problème est lié à la définition de la **fenêtre d'analyse**. Dans les études réalisées sur les traces sismiques **synthétiques** du delta du Niger, la taille de cette fenêtre variait afin de ne considérer sur chaque trace que la partie temporelle correspondant exactement à l'intervalle profondeur analysé. Une homothétie temporelle réalisée sur le module du signal analytique

permet de calculer les paramètres E_i de façon comparable d'une trace à l'autre. Par la suite, dans la plupart des problèmes industriels que nous avons traités, la fenêtre temporelle analysée était de **taille constante**.

En effet, le calcul des paramètres sismiques sur une fenêtre variable peut entraîner des différences sur ces paramètres, d'une trace à l'autre, qui n'expriment ni une variation des propriétés géologiques, ni une variation du caractère sismique des traces. Il s'agit d'un artefact provenant de l'analyse spectrale de fonctions observées sur des intervalles temps de **durée finie et de longueurs différentes**. L'analyse de la trace sur une fenêtre revient à multiplier la représentation temporelle infinie qu'elle pourrait avoir par un créneau. Dans le domaine spectral, ceci est équivalent à multiplier le spectre initial de la fonction par une fonction sinus-cardinal : des lobes sont introduits dont les caractéristiques sont liées à la longueur du créneau. Si cette longueur varie d'une trace à l'autre, aussi bien les paramètres calculés sur le spectre de puissance (obtenu par une transformée de Fourier rapide FFT de l'autocorrélation de la trace) que ceux calculés sur le module du signal analytique (obtenu par transformée de Hilbert et calcul dans le domaine spectral) vont être affectés et ceci sans liaison avec l'expression sismique des caractéristiques géologiques.

D'autre part, l'homothétie temporelle effectuée pour tenir compte de la longueur variable de la fenêtre paraît assez douteuse : ce n'est plus le même signal sismique qui est analysé, en particulier, le contenu fréquentiel est altéré.

Nous recommandons donc d'étudier une fenêtre temporelle de **taille constante**.

Il s'agit donc de la définir, ce qui est parfois délicat. Les variations géologiques de la zone étudiée (épaisseurs et vitesses) peuvent faire que la fenêtre temporelle correspondante est très variable. Nous proposons de choisir alors la longueur maximale définie d'après les traces sismiques synthétiques calculées aux divers puits (à partir des enregistrements Sonic et Densité). Sur certaines zones du domaine d'étude, la fenêtre sismique englobera donc de l'encaissant, c'est-à-dire des formations extérieures à l'intervalle géologique d'intérêt. En fait, ceci revient à considérer qu'au lieu de s'intéresser à l'expression sismique de la variation géologique entre les faciès A et B, on s'intéresse à la variation entre les faciès A' et B où A' est A plus son encaissant. Les indications, déduites de la sismique et concernant ce dernier problème, sont tout aussi importantes pour l'étude du problème initial.

D'autre part, à travers les études déjà réalisées, il est très rare que la définition de la taille de la fenêtre pose problème. Ses variations sont souvent de l'ordre de quelques échantillons sismiques (≈ 1 à 3), ce qui, vu le pouvoir de résolution limité des données sismiques, revient à analyser le même phénomène. Nous rappelons également que les paramètres sismiques analysés sont des mesures globales des caractéristiques sismiques : ils sont donc peu sensibles à la prise en compte ou non de quelques échantillons supplémentaires aux bornes de la fenêtre (à la différence des paramètres qui seront utilisés pour l'adaptation de la méthodologie aux courtes fenêtres temporelles).

Une autre limitation liée à la fenêtre d'analyse est sa **taille**. Les paramètres sismiques exigent l'estimation des caractéristiques spectrales des traces sismiques. La méthode utilisée est une transformée de Fourier rapide (FFT). Nous savons que l'estimation ainsi obtenue n'est fiable que si la fenêtre temporelle d'analyse du signal est suffisamment longue. Les fenêtres classiquement utilisées en traitement du signal sont de l'ordre de 500 – 1000 ms, temps double. Notre analyse sur des fenêtres de l'ordre de 100 ms est certainement moins fiable ; mais puisque nous nous intéressons à des **variations latérales** des caractéristiques sismiques (comparaison relative et non absolue des traces), l'incertitude de l'estimation spectrale sera constante d'une trace à l'autre et n'affectera pas l'étude en discrimination.

Par contre, il semble que le calcul des variables sismiques proposées (sur l'autocorrélation, son spectre, le module du signal analytique) perd sa **signification physique même** sur des fenêtres temporelles vraiment courtes ($\approx 20 - 30$ ms). Pour des problèmes fins (type "gisement"), nous ne pouvons pas appliquer la méthode telle quelle ; un autre moyen de caractérisation des traces sera adopté (cf. chapitre 6, troisième partie).

La méthodologie repose sur l'hypothèse que les caractéristiques sismiques analysées sur la fenêtre reflètent, au travers de **leurs variations**, uniquement des événements géologiques internes à l'intervalle d'étude. Or, le contenu fréquentiel de la trace sismique à un temps donné dépend non seulement des propriétés acoustiques à la profondeur correspondante, mais aussi des phénomènes d'absorption qui ont eu lieu dans les terrains sus-jacents. Par exemple, s'il existe, sur la zone étudiée, des chenaux superficiels remplis de matériel peu consolidé et érodant un encaissant beaucoup plus compact, une absorption différentielle existera latéralement entre la zone du chenal et la zone adjacente. Cette différence se retrouvera même en profondeur. Pour déduire des variations sismiques une information géologique sur l'intervalle analysé, il faut supposer que le filtrage réalisé par les terrains sus-jacents ne varie pas latéralement. Cette contrainte est forte mais il est impossible de s'en affranchir ; elle est inhérente à la nature même des données traitées.

D'autre part, peu de cas de figures existent où ces phénomènes sont contraignants d'un point de vue pratique. Ces cas sont aussi bien répertoriés (comme celui des chenaux superficiels) et attirent donc immédiatement l'attention de l'interpréteur. Cette perturbation doit alors être prise en compte dans l'analyse des résultats statistiques.

Les limitations que nous venons d'exposer sont toutes plus ou moins liées à l'analyse des données sur une fenêtre temporelle fixée. Examinons maintenant les limitations liées aux **paramètres sismiques retenus**. Les attributs que nous utilisons pour caractériser les traces ne font pas intervenir directement la **continuité latérale** des réflexions. Or ce paramètre est important dans l'analyse classique des faciès sismiques (cf. la définition des diverses unités de faciès sismiques). Mais nous pensons en tenir compte de façon indirecte.

En effet, si le faciès est très continu, la morphologie sismique est identique d'une trace à l'autre. Les traces seront alors classées ensemble. Si le faciès est continu mais si certaines traces apparaissent différentes suite à du bruit aléatoire, ces traces seront détectées dans la classification, ceci est un élément positif dans l'analyse automatique du faciès sismique. Pour les faciès très peu continus, deux cas se présentent. Soit l'absence de continuité est liée à l'absence de réflexions importantes, la zone est "sourde", les traces sont très peu contrastées. Dans ce cas là, vu l'homogénéité de ce type de faciès, toutes les traces seront affectées au même groupe et le faciès sera reconnu. L'autre cas est l'existence de réflecteurs présentant des variations latérales de continuité très fortes comme pour le faciès sismique de la zone margino-littorale (Figure 49). Les traces sismiques du faciès "levées" seront différenciées alors de celles du faciès "chenal". Le faciès sismique qui est reconnu est **plus fin** que le faciès "classique" qui est **l'enveloppe** de tous les chenaux et levées de la zone margino-littorale.

La méthodologie proposée, qui repose sur une analyse morphologique de la trace, permet de trouver des subdivisions faciologiques **plus fines** que l'approche traditionnelle ; le paramètre continuité n'apparaît important que pour regrouper (éventuellement) deux ensembles de faciès sismiques pour une analyse plus globale (faciès "levées" et "chenaux" regroupés en "complexe margino-littoral").

Examinons maintenant les limitations liées à l'**interprétation géologique** des faciès sismiques détectés. L'hypothèse qui est faite, à ce stade, est que les traces d'apprentissage, définies au voisinage de puits typiques, sont elles-mêmes représentatives des variations géologiques recherchées. Cette hypothèse n'est vérifiée qu'à l'étape de modélisation de la signification géologique des paramètres discriminants. Or, la modélisation utilisée est très simple : modèle géologique assez fruste, simple convolution 1D pour le calcul de la réponse sismique. D'autre part, les résultats issus de cette étape ne constituent pas une preuve irréfutable de la valeur géologique de la classification sismique : la réponse sismique n'est pas reliée de façon univoque aux caractéristiques géologiques.

Dans le choix des traces d'apprentissage et le nombre de groupes à constituer, il faut également se fier à l'information géologique a priori, en particulier à la classification des puits qui a été faite. Il n'est pas toujours facile d'établir une typologie des données géologiques, lesquelles s'inscrivent souvent dans un continuum. Nous pensons résoudre ce problème (choix et nombre de groupes) de façon relativement satisfaisante grâce aux tests en classification automatique qui viennent corroborer les résultats en analyse factorielle discriminante. Il reste toujours le problème du lien assez faible entre les faciès sismiques reconnus et les faciès géologiques recherchés. Aucune mesure quantitative du lien entre les paramètres sismiques et les propriétés géologiques n'est faite.

Nous allons maintenant discuter du **choix des techniques statistiques** qui est fait dans la méthodologie. Ce choix a reçu l'aval d'une équipe de statisticiens de l'INRIA (G. Celeux, Y. Lechevallier, E. Azorin-Lara, 1986) à qui nous avons confié trois fichiers de traces

sismiques correspondant à trois études différentes, pour conforter nos propres résultats statistiques et proposer éventuellement des méthodes mieux adaptées.

En présence d'information géologique (c'est-à-dire dans la plupart des cas), le problème statistique est essentiellement un problème de discrimination qui présente les particularités suivantes :

- bruit important inhérent à la sismique,
- fiabilité douteuse de l'information a priori,
- frontières floues entre les classes a priori,
- tendance des paramètres sismiques à être corrélés par construction même.

Ces particularités imposent de rechercher des règles de décision peu sensibles au bruit et permettant de détecter des erreurs de diagnostic. Elles doivent donc être simples, robustes et leur bien-fondé doit être confirmé par des analyses statistiques ne tenant pas compte de l'information a priori.

Parmi les techniques de discrimination, c'est la discrimination **linéaire** qui a été retenue pour sa robustesse (par rapport aux méthodes quadratiques ou non paramétriques) : elle est simple donc induit l'estimation d'un minimum de paramètres et autorise des tailles d'échantillons d'apprentissage plus réduites. Les résultats de la discrimination linéaire sont satisfaisants dans la plupart des cas. Cette constance dans la qualité est bien mise en évidence par Titterington et al. (1981) qui la comparent à la plupart des techniques concurrentes sur de nombreux cas.

En ce qui concerne la phase de contrôle de stabilité des résultats, nous avons le choix entre deux types de techniques : les techniques de ré-échantillonnage, type Jackknife ou Bootstrap (Efron, 1982), et les techniques de classification type nuées dynamiques (Diday, 1980).

Les premières ne nous satisfaisaient pas pour deux raisons :

- on se rend compte de plus en plus que, pour fournir des **mesures satisfaisantes de l'incertitude statistique**, les tailles d'échantillons doivent être relativement importantes ;
- ces techniques ne remettent pas en cause de manière radicale l'information a priori.

Au contraire, les techniques de classification qui n'utilisent pas d'information a priori permettent de confirmer avec certitude les résultats d'une discrimination, si ceux-ci sont stables. La technique de classification la plus adéquate doit être simple, robuste et utiliser un critère relié à celui de la discrimination linéaire. La version la plus simple des nuées dynamiques (algorithme des centres mobiles, cf. § 4.2.3.3) basée sur une minimisation de l'inertie intra-classe remplit ces conditions.

Un traitement complémentaire des partitions obtenues avec cet algorithme (méthode des formes fortes ou partition consensus) apporte aussi des éléments de validation de la structure de la population.

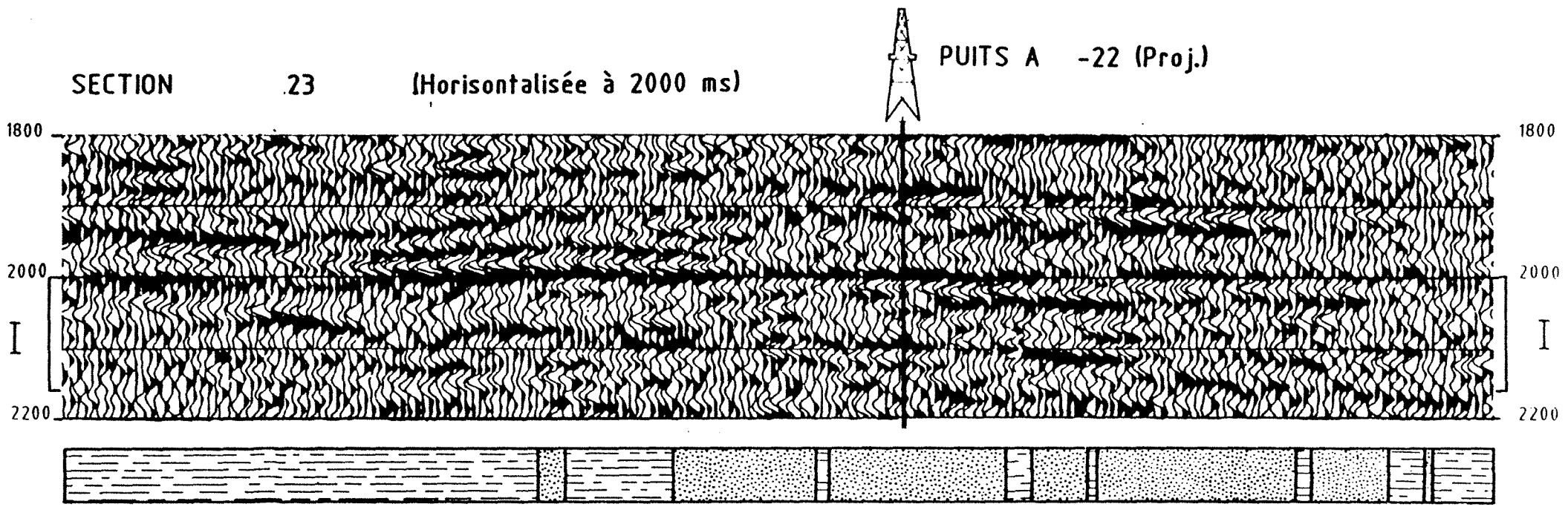
La justification des choix statistiques ayant été faite, nous montrons un exemple de type de **validation pratique** des résultats qu'il est possible de faire lorsqu'on dispose d'un **nombre de puits assez élevé**.

Trois classes de puits avaient été constituées, dans un environnement turbiditique argilo-gréseux, en fonction du rapport sables/argiles :

- faciès bon réservoir (puits A 22);
- faciès réservoir moyen (puits A 29);
- faciès réservoir mauvais (puits A 40 ou A 38).

Des étapes d'apprentissage avec trois groupes de traces sismiques mais aussi seulement avec deux (faciès bon réservoir opposé aux deux autres) ont été effectuées (Figure 86). Nous utilisons, comme moyen de contrôle des affectations proposées par le programme, la comparaison de la "qualité" des réservoirs, connue aux puits non utilisés comme référence dans l'étape d'apprentissage, avec l'affectation des traces adjacentes. Ces résultats sont repris dans le tableau 17 avec :

B pour bon réservoir, MY pour réservoir moyen, MV pour mauvais réservoir.



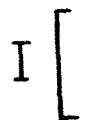
CLASSEMENT DES TRACES:



BON RESERVOIR



MAUVAIS RESERVOIR



INTERVALLE D'ETUDE

Fig. 86

Section	Puits	Classification gisement	A.F.D. 3 GRP 22, 29, 40d	A.F.D. 3 GRP 22, 29, 38	A.F.D. 2 GRP 22, 29
01	14d	MV	MY	MY	MY
	11 **	B	MY	B	B
	5 *	B	MY	MY	MY
02			PAS DE PUITES		
03	13 **	MY	MY	MY	MY
04			PAS DE PUITES		
05	32d *	B	MY	MY	MY
	34d **	MY	B	B	B
06	35d *	B	B	B	B
	39d *	MV	B	B	B
07	34d **	MY	MY	MY	MY
08	31d	MY	B	MY	B
	24d *	B	MV	MY	B
	34d *	MY	B	B	B
09	12 *	MY	B	B	B
11	39d *	MV	B	MY	B
	24d	B	B	B	B

* : puits proche d'une ligne et projeté dessus

** : puits projeté assez éloigné de la ligne correspondante

Tableau 17 : Validation pratique des résultats obtenus en reconnaissance automatique des faciès sismiques avec la connaissance géologique a priori donnée par les puits

Section	Puits	Classification gisement	A.F.D. 3 GRP 22, 29, 40d	A.F.D. 3 GRP 22, 29, 38	A.F.D. 2 GRP 22, 29
13	24d	B	B	B	B
	38d	MV	MY	Référence	MY
	3	MY	MY	MY	MY
	36d *	MY	MY	B	MY
14			PAS DE PUIITS		
15	31d	MY	B	B	B
	501 **	B	MV	B	B
	9 *	B	MV	B	B
16	22d	B	Référence	Référence	Référence
	30d *	B	MY	B	B
	35d **	B	B	B	B
21	502 *	B	B	B	B
23	22d **	B	B	B	B
	30d *	B	B	B	B
25	22d *	B	B	B	B
	37d **	B	B	B	B
27	40d	MV	Référence Référence	MY Référence	MY Référence
	29d *	MY			
	18d	MV	MY	MY	B
	502 *	B	B	B	B

* : puits proche d'une ligne et projeté dessus

** : puits projeté assez éloigné de la ligne correspondante

Tableau 17 (suite)

Dans le cas de trois groupes, la classification proposée par le programme correspond à celle du gisement dans 55 % des cas étudiés.




Même si ce résultat n'est pas très bon, il faut le comparer à celui que l'on aurait obtenu si la classification s'était faite au hasard : environ un tiers des puits bien classés seulement. La classification à deux groupes améliore un peu ces résultats (69 % de bien classés par rapport à 50 % si l'attribution s'était faite au hasard).

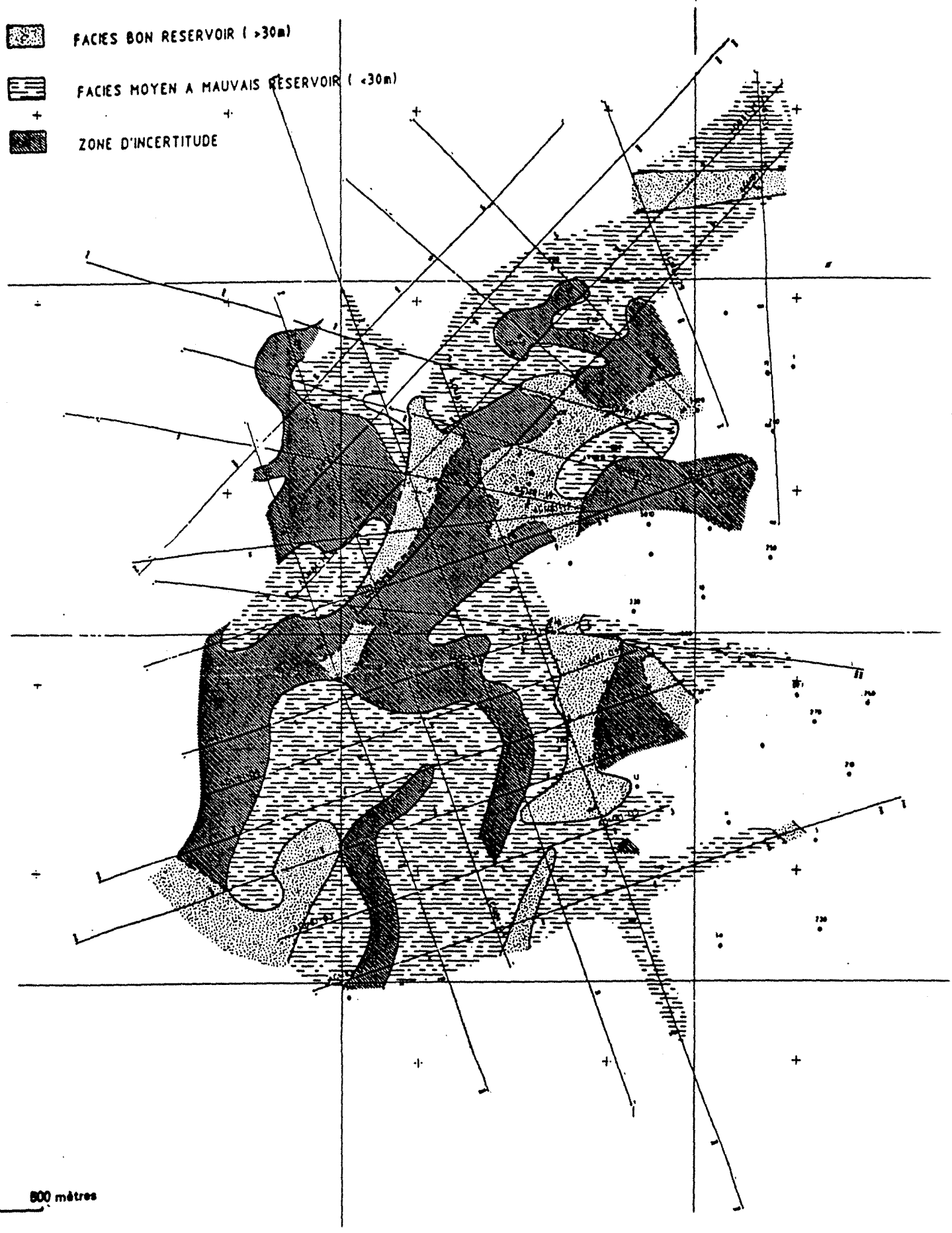
Il faut souligner, toutefois, que ces statistiques sont effectuées sur un petit nombre de puits et par conséquent ne sont pas forcément significatives. D'autre part, certains puits sont très loin des lignes sismiques sur lesquelles nous les avons projetés. Leur affectation n'est donc pas forcément celle donnée par les traces de la ligne où ils ont été projetés.

Nous avons utilisé les résultats proposés par l'analyse discriminante avec deux groupes pour réaliser une carte du faciès réservoir prédit (Figure 87). La zone grise correspond à une zone où nous avons jugé le faciès **indéterminé** : en effet, d'une trace à l'autre, l'affectation varie sans qu'il soit possible de dégager la moindre cohérence spatiale. Les seules limites dont nous sommes sûrs sont les limites "réservoir moyen à mauvais"/"bon réservoir".

Nous pouvons reprendre l'affectation proposée pour les puits en tenant compte de la zone d'incertitude et des limites de faciès proposées (Tableau 18).

Fig. 87

-  FACIES BON RESERVOIR (>30m)
-  FACIES MOYEN A MAUVAIS RESERVOIR (<30m)
-  ZONE D'INCERTITUDE



0 800 mètres

Puits	Classification gisement	Classification H.F.D. 2 GRP
3	MY/MV	MY
5	B	B
9	B	Indéterminé
11	B	MY
12	MY/MV	Indéterminé
13	MY/MV	Indéterminé
14	MV	MY
17	B	B
18	MY	Indéterminé
22	B	B
24	B	B
29	MY	MY

Tableau 18 : Validation pratique des résultats obtenus en reconnaissance automatique des faciès sismiques - opposition "bon" réservoir à réservoir "moyen à mauvais".

Puits	Classification gisement	Classification A.F.D. 2GRP
30	B	B
31	MY	Indéterminé
32	B	Indéterminé
34	MY	B
35	B	B
36	MY	MY
37	B	Indéterminé
38	MV	MY
39	MV	Indéterminé
40	B	Indéterminé
501	B	Indéterminé
502	B	B

Total des puits : 24

Puits déterminés : 14

Puits indéterminés : 10

Puits bien classés : 12

Puits mal classés : 2

Tableau 18 (suite)

Sur les 24 puits classés (Tableau 18), 14 puits n'appartiennent pas à la zone indéterminée et 12 sont correctement affectés, soit 86 %. Ce résultat est largement supérieur à ce que l'on pourrait attendre d'une classification au hasard.

Il va de soi que plus la classification géologique des puits est fiable, plus ces études pratiques de validation sont fructueuses.

La méthodologie proposée a été appliquée avec succès sur environ une dizaine de cas. Il est par ailleurs intéressant de constater que des démarches très similaires ont été proposées, depuis fin 1985, par des chercheurs d'autres sociétés. Une fenêtre d'analyse est toujours définie sur les données sismiques, au niveau de l'objectif géologique. Les portions de traces sont caractérisées par un grand nombre d'attributs, calculés sur leurs représentations temporelles et spectrales. L'analyse factorielle discriminante est la technique clé pour prendre en compte l'information géologique aux puits (Asimakopoulos, 1985, O'Connor, 1987) mais des classifications automatiques sont aussi utilisées (Thadani, 1986).

D'autres méthodes d'analyse du faciès sismique ont été proposées depuis le début de ces travaux, en particulier une approche par reconnaissance des formes de **type syntactique**. Il est déjà difficile d'obtenir avec ces méthodes des résultats convaincants sur des données sismiques **synthétiques**. Les travaux semblent peu progresser dans ce domaine et on peut penser qu'il ne déboucheront pas sur des méthodologies industrielles avant longtemps.

La voie **statistique** pour la reconnaissance automatique des faciès sismiques apparaît comme la voie opérationnelle et également la plus prometteuse pour les nouvelles applications comme les problèmes très fins de gisement. La méthodologie que nous proposons répond aux principaux objectifs que nous nous étions fixés:

- automatisation et objectivation d'une tâche manuelle,
- définition plus poussée des faciès sismiques que celle offerte par l'analyse classique,
- quantification des différences permettant d'individualiser les faciès,
- intégration plus systématique de l'information géologique connue aux puits, avec une critique possible de cette information.

Il reste que la caractérisation des traces qui est faite par les variables sismiques est mal adaptée pour des problèmes fins de géologie de réservoir (courtes fenêtres temporelles). Nous exposerons des méthodes pour adapter le procédé à ces problèmes dans la troisième partie.

Les cas d'études présentés dans le chapitre 5 de cette seconde partie cherchent à illustrer les caractéristiques et les diverses étapes de la méthodologie statistique ainsi qu'à faire comprendre les limitations que nous avons exposées.

5. ETUDES DE CAS

5.1. Introduction

Le premier cas présenté (Dogger carbonaté du Bassin de Paris) a été retenu pour son caractère particulièrement illustratif de la méthodologie. Le second cas (séries argilo-gréseuses turbiditiques du Gabon) met en valeur l'étape de validation des paramètres discriminants par modélisation géologique puis sismique.

Une information géologique est disponible dans les deux cas : c'est la situation la plus fréquemment rencontrée dans les cas que nous avons traités.

5.2. Etude des faciès sismique du Dogger carbonaté

5.2.1. *Cadre géologique*

Le champ étudié est situé à l'Est du Bassin de Paris. Une étude de sismique stratigraphique de la campagne sismique 77, la plus récente, a montré que l'ensemble oolithique à la base de la séquence régressive du Bathonien (Figure 89) s'épaissit du Sud-Ouest vers le Nord-Est. Cet épaississement n'est pas "continu". En effet, sur des "hauts fonds" bajociens, se développent des "dunes oolithiques" ; à l'abri de ces dunes, se développe une formation à calcaire oncholithique (énergie du milieu de dépôt moyenne).

Sur la ligne 12 (Figure 88), de direction Ouest-Est et où plusieurs puits existent, nous constatons une évolution du réservoir oolithique entre les puits 3 et 4.

Nous voulons positionner sur la sismique ce changement de faciès, vérifier que les puits 1 et 2 sont bien affectés et comparer avec l'interprétation classique de la campagne 77, faite en 1980 (Dumay, Kanaan).

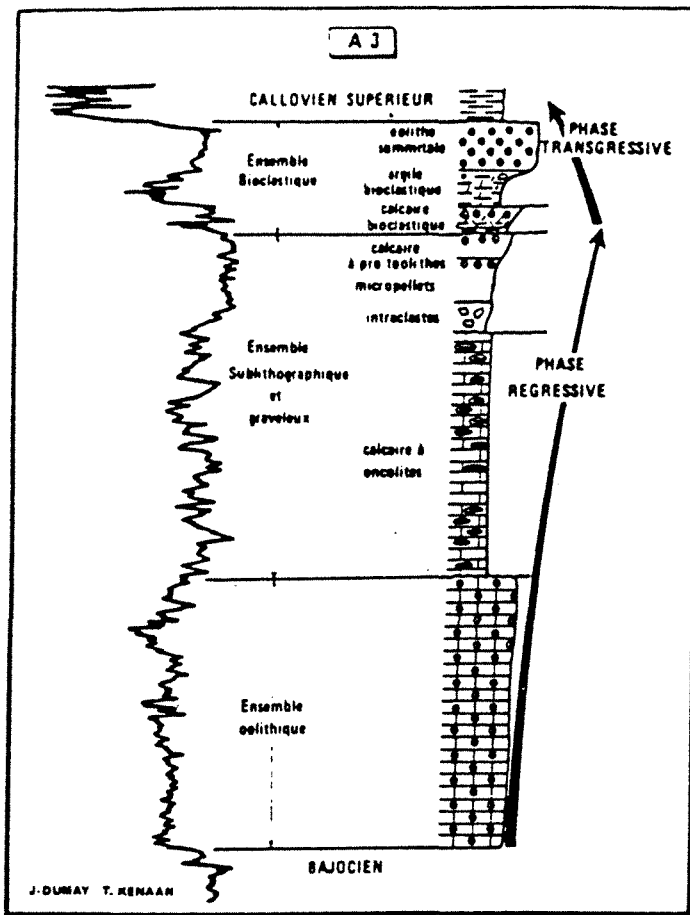
5.2.2. *Phase d'apprentissage*

5.2.2.1. *Sélection des traces de référence*

Sur la section 12, horizontalisée au toit du Callovien Inférieur, nous avons choisi comme traces de référence les traces voisines du puits 3, associées à un faciès réservoir oolithique peu épais (faciès "platier") et un faciès oncholithique développé ainsi que les traces voisines du puits 4 représentant le faciès "dune oolithique" (Figure 89).

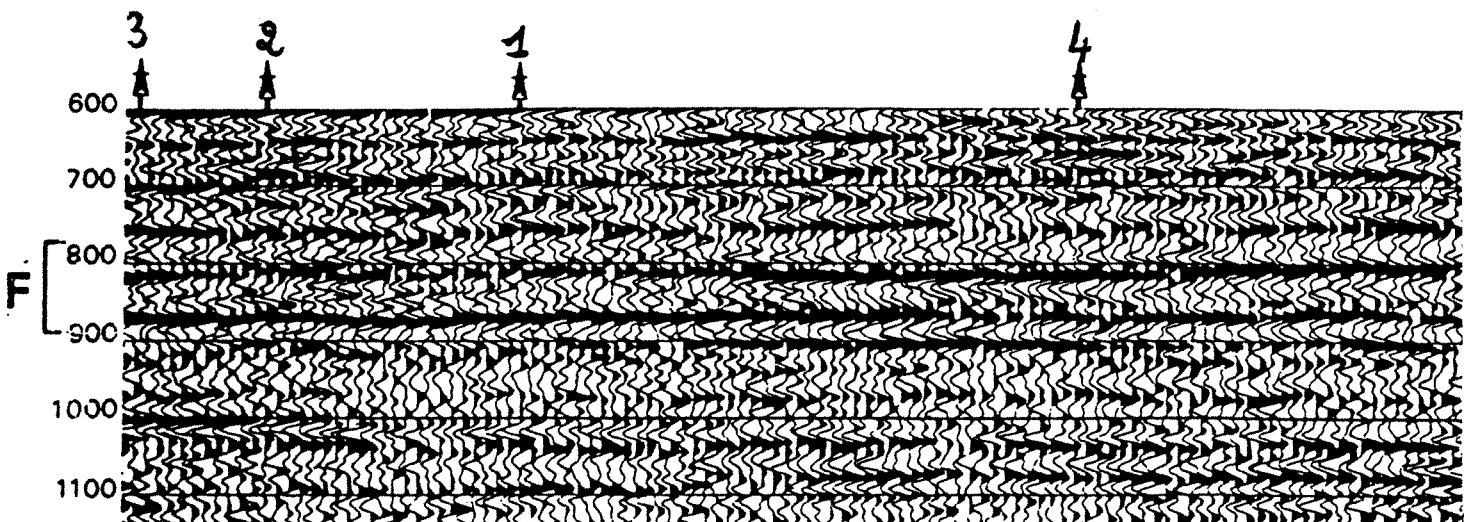
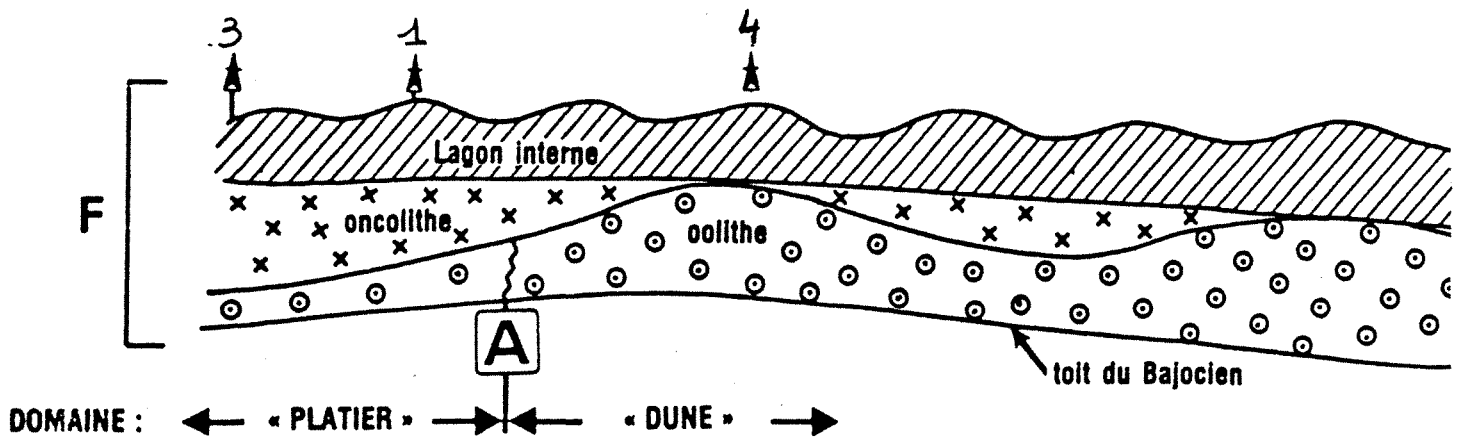
Fig. 88

DONNEES GEOLOGIQUES



F : fenêtre d'étude

S.W.

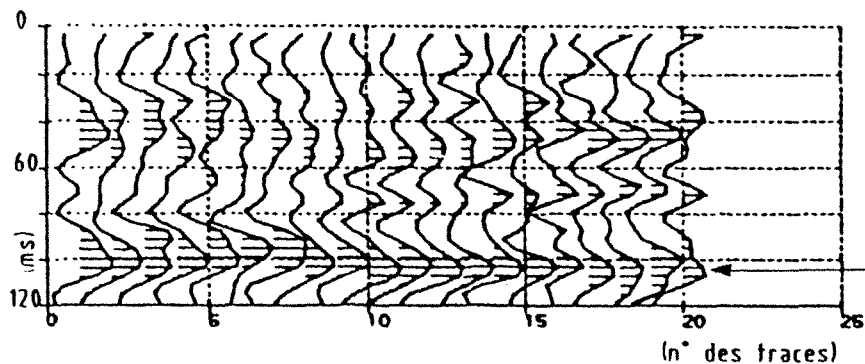


Section 12

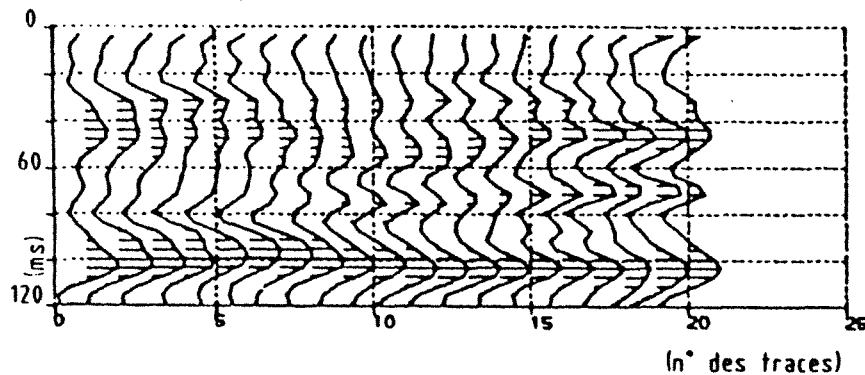
POLE PLATIER

(traces 245-226)

Traces réelles



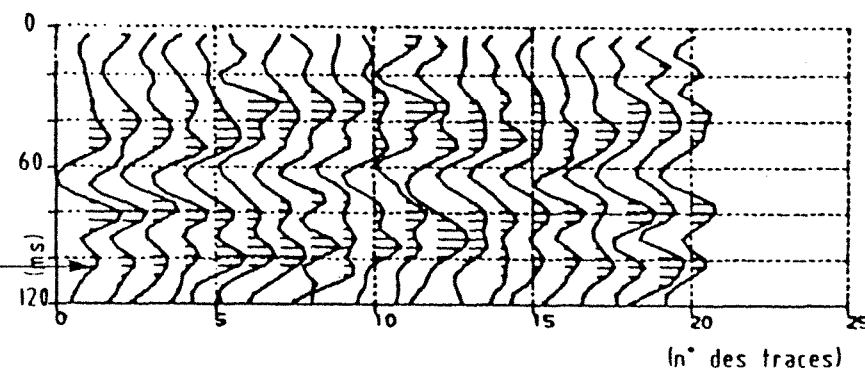
Traces moyennes



POLE DUNE

(traces 70-51)

Traces réelles



Traces moyennes

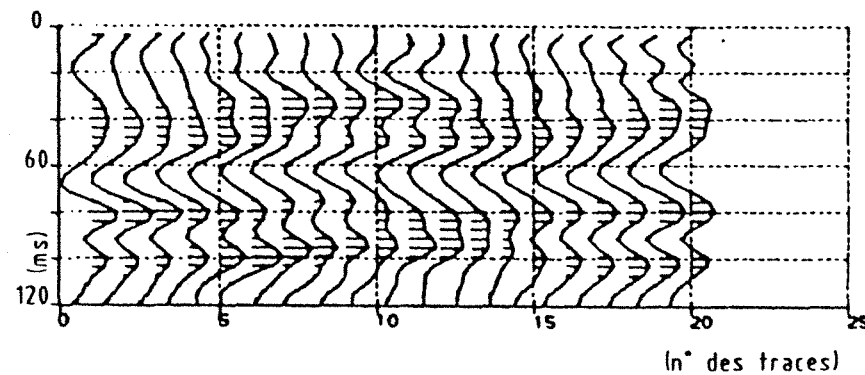


Fig. 89

La longueur de fenêtre retenue en début d'étude est de 120 ms. Le toit de la fenêtre d'étude correspond au toit du Callovien Inférieur, le mur est dans les argiles marines bajociennes (Figure 88).

Sur ces traces de référence (Figure 89), nous allons calculer les paramètres sismiques de la méthodologie. Auparavant, afin d'éliminer une partie du bruit, nous faisons une moyenne de trois traces adjacentes avec glissement d'une trace à chaque pas.

5.2.2.2. Sélection des paramètres discriminants

S'il existe une différence de faciès entre les traces de référence des deux milieux de dépôt ("dune" et "platier"), elle ne sera sensible que sur quelques variables.

Ces variables seront les variables discriminantes pour le problème posé : distinguer le faciès sismique associé aux "dunes" du faciès sismique associé au "platier". Les autres variables non seulement n'apporteront pas d'information supplémentaire mais risquent même de masquer l'information intéressante.

Une première analyse discriminante a permis de sélectionner les variables suivantes : E6, TAM1, QW5. Au premier pas, la variable la plus discriminante est la variable E6.

	Dune	Platier	% de bien classés sur l'échantillon de base
Dune	20	0	100
Platier	4	16	80

Tableau 19 : A.F.D. pas à pas sur les traces d'apprentissage des pôles "dune oolitique" et "platier oncolithique"

L'introduction de la variable TAM1 entraîne un résultat de 100 % de bien classés sur l'échantillon de base. La troisième variable introduite est la variable QW5.

Une deuxième analyse réalisée en sélectionnant d'autres traces de référence a permis de retenir les variables : E6, QW6, E3. Parmi ces variables, E6 est la première variable discriminante. Dans toutes les analyses réalisées, à elle seule, E6 permet de classer convenablement 80 % des

échantillons de référence, associée à TAM1 ou à QW5, nous parvenons à 100 % de bien classés. La discrimination sur l'échantillon d'apprentissage fournit donc d'excellents résultats.

Nous avons étudié le comportement de la variable E6 sur chacune des traces de référence des deux pôles. La figure 90 montre l'**enveloppe** des traces de référence :

- A1, enveloppe des traces du pôle platier,
- A2, enveloppe des traces du pôle dune,

ainsi que la courbe cumulée réalisée à partir de l'enveloppe :

- B1, cumul de l'enveloppe des traces du pôle platier,
- B2, cumul de l'enveloppe des traces du pôle dune.

Nous avons également représenté la position temporelle de la variable E6, qui correspond à 60 % de l'énergie de la trace.

On constate que pour le pôle dune, la variable E6 est située entre 60 et 80 ms alors que pour le pôle platier, elle est située entre 80 et 100 ms, ceci veut dire que l'énergie semble répartie uniformément sur les traces du pôle dune, alors que pour les traces du pôle platier, elle est surtout contenue en fin de trace.

Le graphique A1 qui montre l'évolution de l'enveloppe des traces du pôle platier confirme cette dernière observation. Par contre, le graphique A2 précise pour les traces du pôle dune une "concentration d'énergie" entre 60 et 80 ms. La figure 91, sur laquelle sont représentées les traces moyennées entre les deux pôles, permet de retrouver ces caractéristiques.

La figure 92 montre le comportement de la variable E6 sur cette même portion de section, on constate que vers la trace 135 nous changeons de milieu. Toutefois, il reste certaines indentations, en grisé sur la figure 92. L'association de E6 avec une autre variable discriminante devrait permettre d'atteindre un bon classement pour les traces anonymes.

5.2.2.3. Effet de la longueur de la fenêtre sur le choix des variables discriminantes

Nous avons réduit notre fenêtre d'étude pour limiter l'encaissant autour du niveau réservoir. La figure 93 représente les deux pôles de référence choisis pour une fenêtre de **60 ms**.

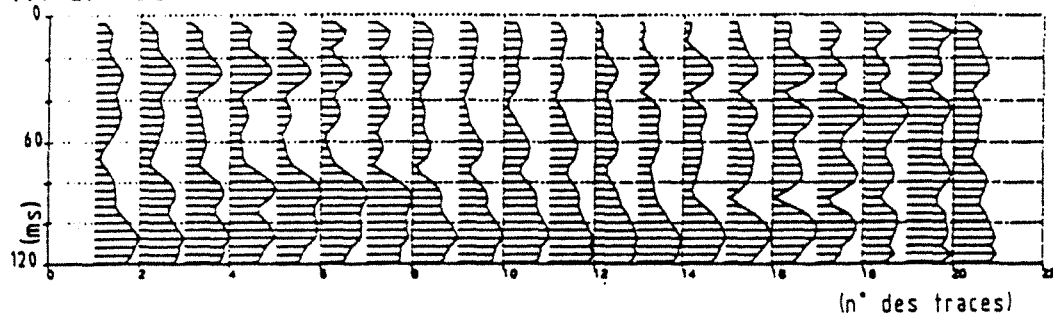
La variable E2 a été sélectionnée, ceci n'est pas étonnant car les 60 ms supprimées l'ont été à partir du toit du réservoir, d'ailleurs la figure 94 montre que la variable E2 a le même comportement que la variable E6 pour un intervalle de 120 ms.

Pour une fenêtre intermédiaire, c'est E4 qui est la première variable discriminante.

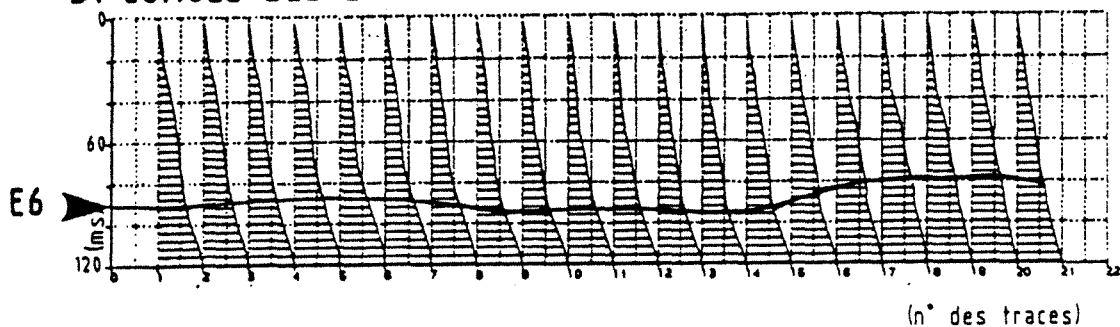
POLE PLATIER

(traces 245-226)

A1 ENVELOPPES DES TRACES



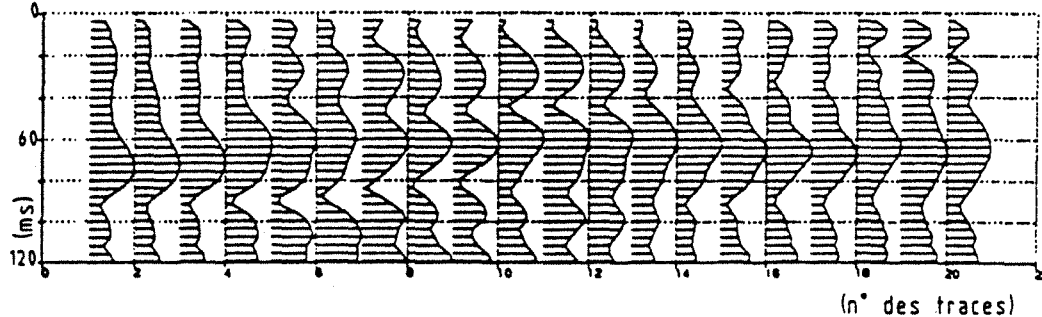
B1 CUMULS DES ENVELOPPES



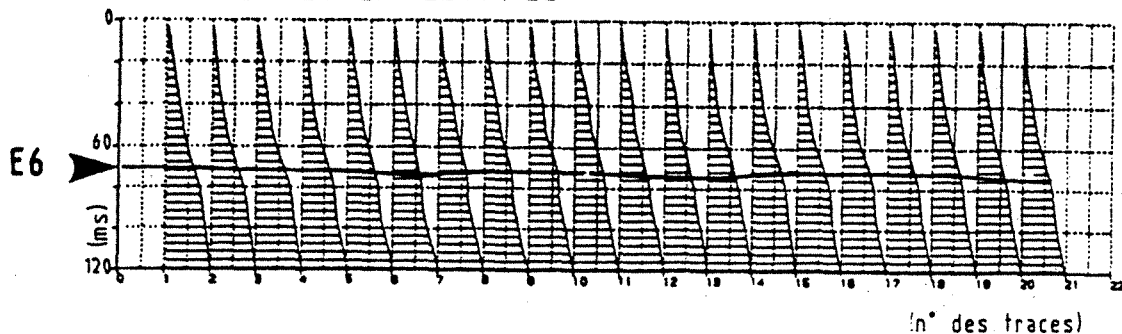
POLE DUNE

(traces 70-51)

A2 ENVELOPPES DES TRACES



B2 CUMULS DES ENVELOPPES



TRACES MOYENNES ETUDIÉES
JOIGNANT LES DEUX POLES

Section 1.2 (traces 245-51)

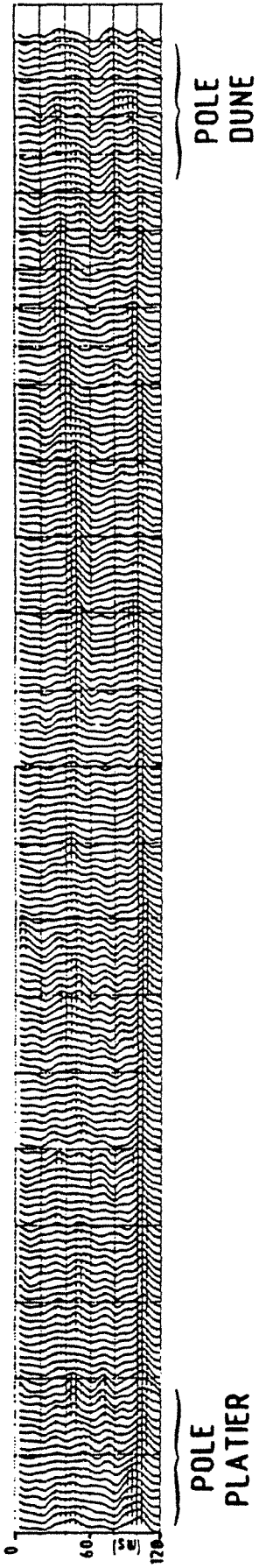


Fig. 91

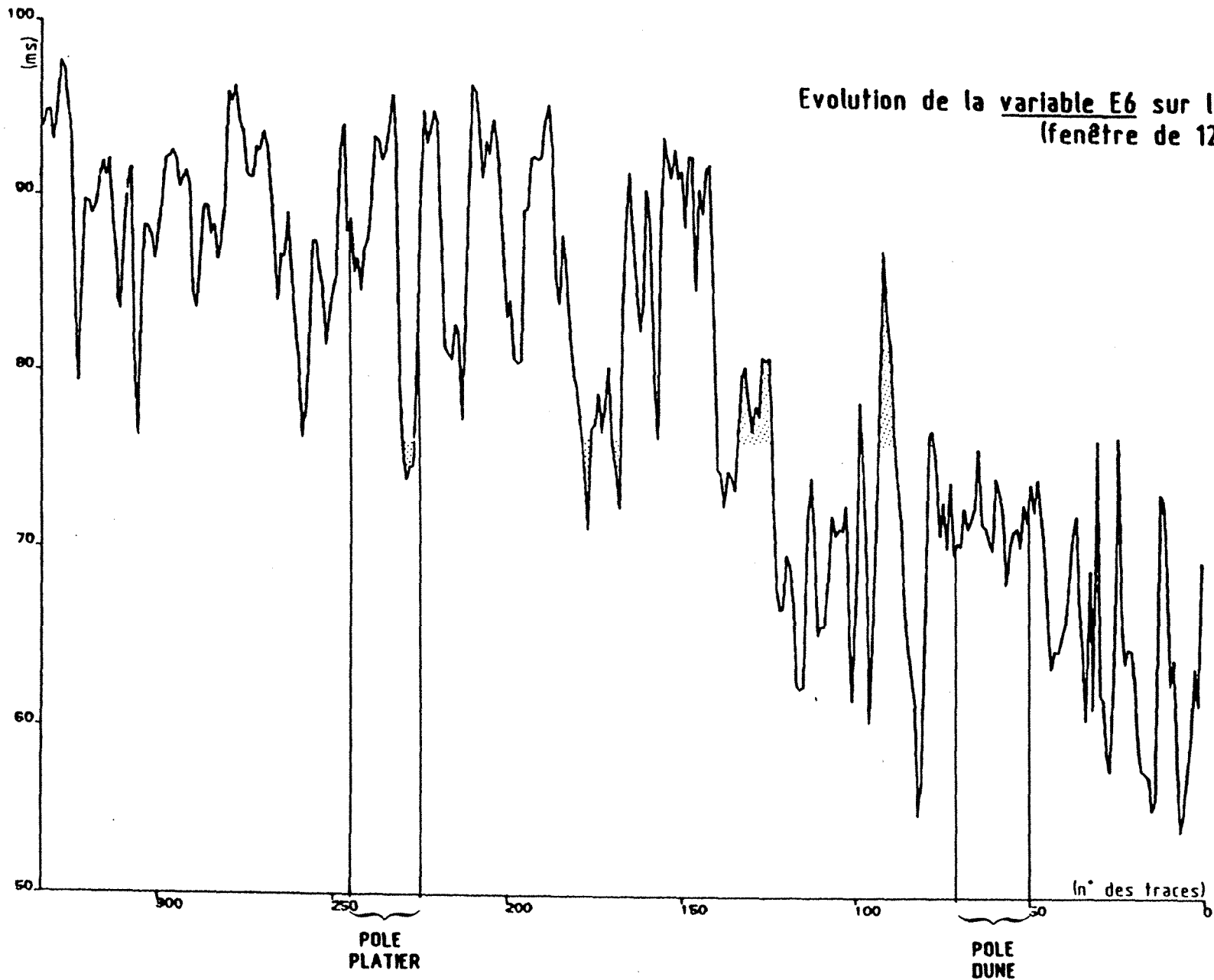
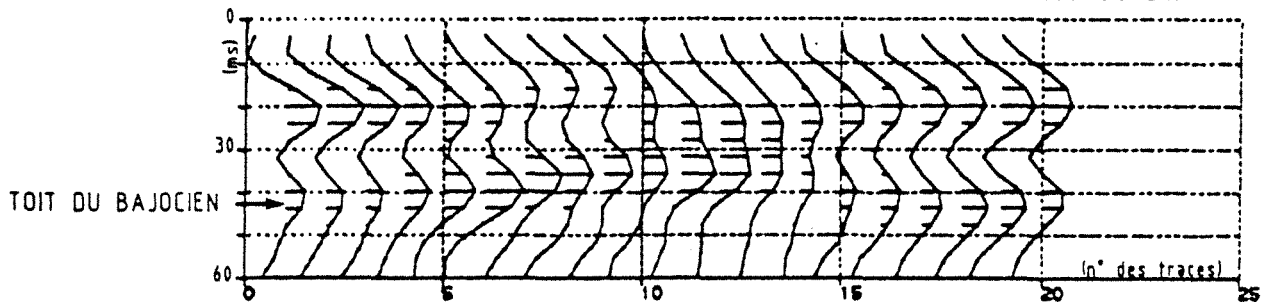


Fig. 92

TRACES SISMIQUES DE REFERENCE
pour une fenêtre de 60 ms

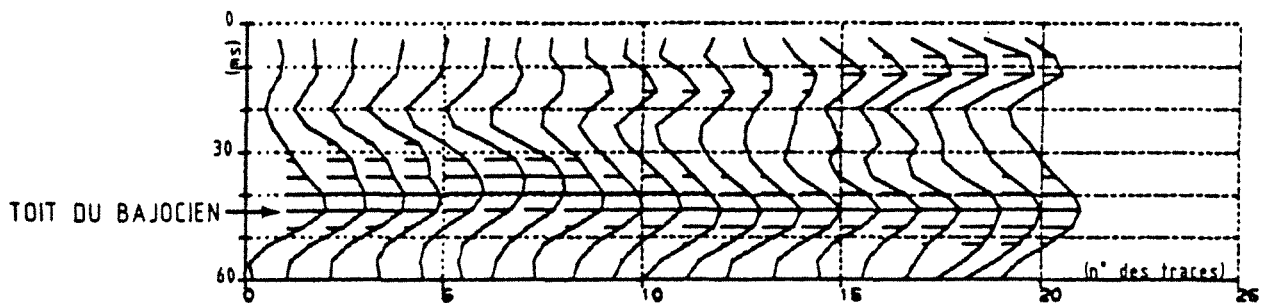
POLE DUNE

(traces 70-51)

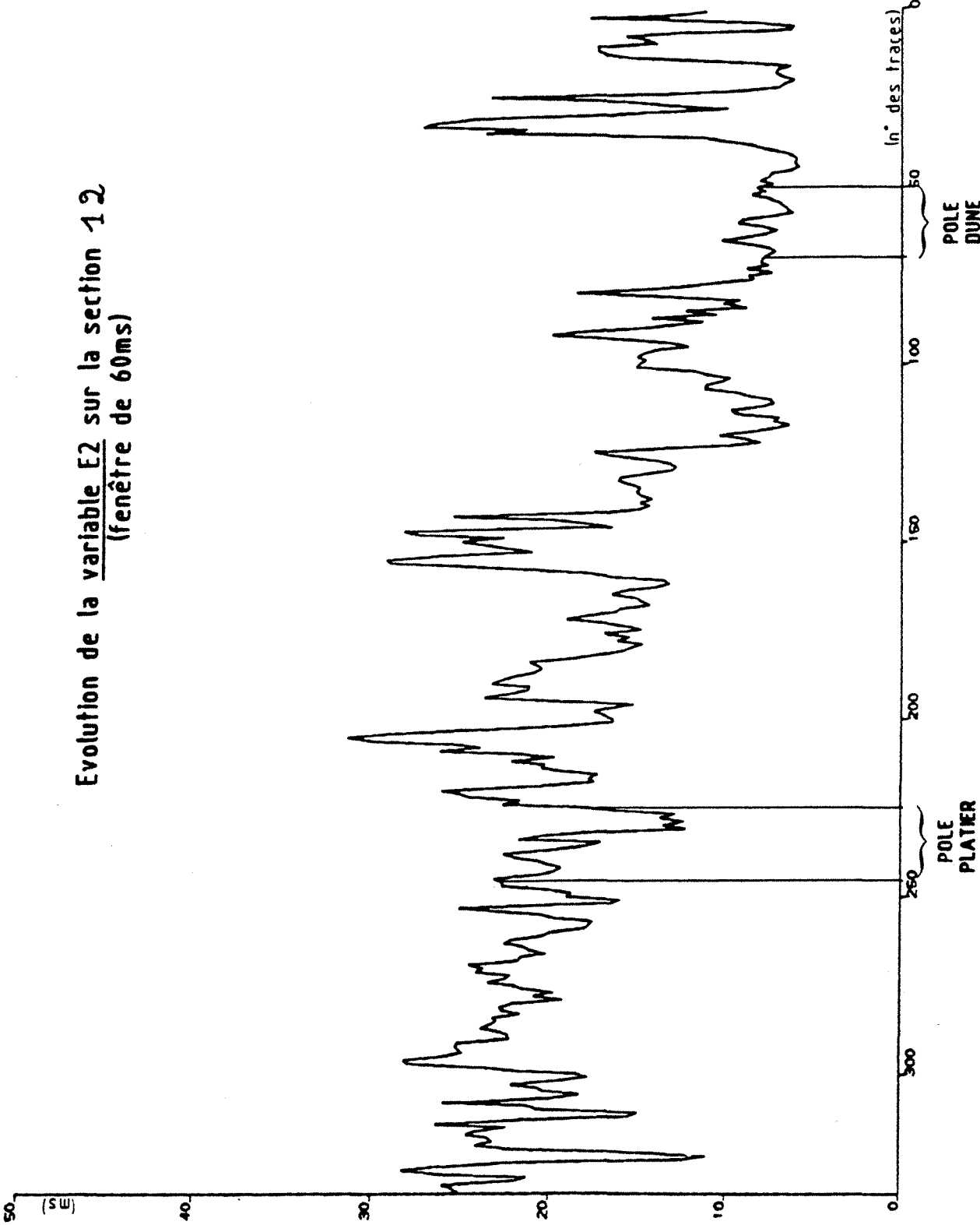


POLE PLATIER

(traces 245-226)



Evolution de la variable E2 sur la section 12
(fenêtre de 60ms)



5.2.2.4. Effet du choix des traces de référence sur la sélection des variables discriminantes

Nous avons changé les traces de référence pour contrôler l'effet du choix des pôles sur la stabilité des variables sélectionnées. Le tableau 20 résume les différents essais.

	LONGUEUR DE LA FENETRE	VARIABLES DISCRIMINANTES SELECTIONNEES		
		1ère	2ème	3ème
ANALYSE 1	120 ms	E6	TAM1	QW5
1a	80 ms	E4	Q6	QW2
1b	60 ms	E2	Q7	E9
1c	60 ms	E2		
ANALYSE 2	120 ms	E6	QW6	E3
ANALYSE 3	120 ms	E6	TAM1	QW5
ANALYSE 4	120 ms	AM1	E7	E9
ANALYSE 5	120 ms	E5	TP24	Q1

Tableau 20 : Variables discriminantes pour les pôles "dune oolithique" et "platier oncolithique"

Pour les analyses 2 et 3, nous avons décalé le groupe dune.

Pour l'analyse 4, le groupe dune de l'analyse 3 a été gardé et le groupe platier réduit.

Pour l'analyse 5, nous avons augmenté le nombre de traces du groupe platier et déplacé le groupe dune.

Au cours des analyses 2 et 3, la variable E6 reste la première variable discriminante.

Pour l'analyse 4, AM1 devient la première, mais juste derrière, on trouve E9 qui quantifie la distribution temporelle de l'énergie sur la fin de l'intervalle d'étude.

Par contre, aucune des variables discriminantes sélectionnées pour l'analyse 5 n'est apparue dans une autre analyse. En fait, si l'on examine les traces sismiques correspondant aux différents pôles dunes, on constate le risque d'englober sous un même pôle des faciès platier et des faciès dune dans l'analyse 5.

5.2.2.5. Conclusion

Compte tenu des traces de référence choisies, les variables les plus discriminantes sont E4, E6, E9, TAM1, QW5. Parmi ces cinq variables, celle qui traduit le plus la différence de faciès sismique entre les deux groupes de traces de référence est la variable E6.

Le raccourcissement par le toit de la longueur de la fenêtre ne fait que décaler la première variable discriminante, ce qui est logique du fait du mode de calcul des variables quantifiant la distribution temporelle de l'énergie.

Si nous choisissons d'autres pôles, nous retrouvons les principales variables discriminantes dans un ordre et des combinaisons qui peuvent être différents. Nous verrons l'effet de ces différentes associations dans les résultats de la phase de prédiction.

D'autre part, les estimations des pourcentages de bien classés sont très encourageantes : 98 à 100 % pour des combinaisons de trois variables, toutefois ce résultat repose sur un petit nombre d'échantillons.

5.2.3. Phase de prédiction

Au cours de cette phase, nous allons décider de l'affectation de chaque portion de trace de la section 12. Nous classons ces traces par rapport aux traces de référence (groupe platier ou groupe dune) au cours de l'étape prédictive de l'analyse factorielle discriminante.

Sur la figure 95 sont représentés les résultats des affectations pour les cinq analyses portant sur la fenêtre d'étude de 120 ms et les trois analyses avec fenêtre réduite (60 ms).

Nous constatons que les résultats des affectations sont très voisins pour l'ensemble des analyses, exception faite de l'analyse 5. Pour cette dernière, le choix du pôle dune était probablement mauvais : taille du pôle trop importante et présence de traces appartenant au groupe platier (cf. le résultat des affectations des autres analyses).

Par contre les analyses 1, 1a, 1b, 2, 3, 4, donnent des résultats très voisins. Ainsi la combinaison des variables discriminantes AM1, E7, E9 de l'analyse 4 donne des résultats très voisins de ceux obtenus par l'analyse 1 dont les variables discriminantes sont E6, TAM1, QW5.

Le phénomène important est le découpage de la section en **deux grands domaines** :

- un domaine à dominante dune de la trace 1 à la zone des traces 118, 136,
- un domaine de platier quasiment pur des environs de la trace 118 jusqu'au bout de la section.

Dans les analyses discriminantes (avec définition des pôles de référence), nous donnons une information a priori qui sert de guide au cours de la phase décisionnelle d'affectation. Par contre, dans les techniques de classification automatique, nous demandons une répartition de tous les échantillons en un certain nombre de groupes en tenant compte de certaines variables mais **sans information a priori**.

5.2.4. Phase confirmatoire à l'aide des techniques de classification

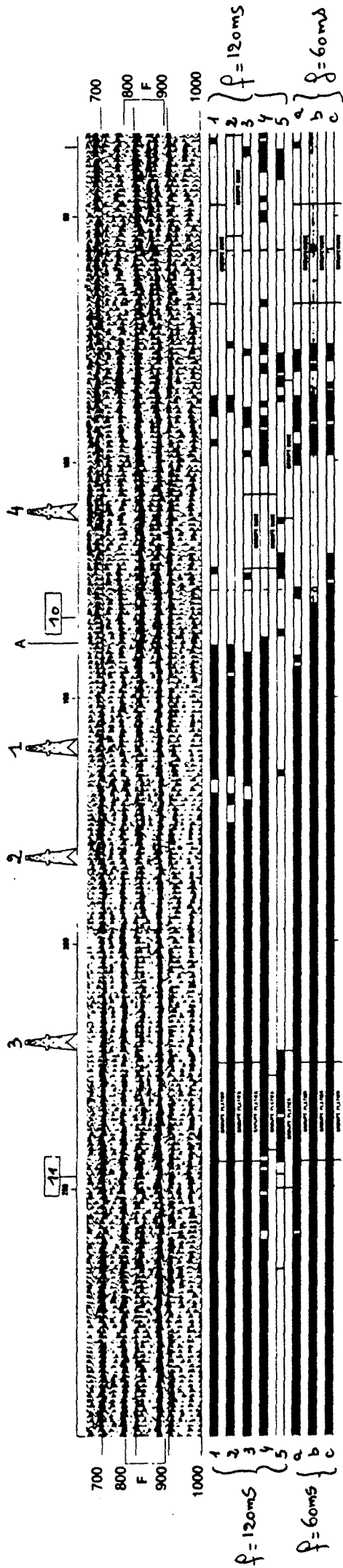
Sur la figure 95 sont représentés les résultats de deux classifications. Ces dernières ont été réalisées avec les variables suivantes :

- E6, TAM1, QW5 pour l'analyse 1 avec trois groupes,
- E6, AM1 pour l'analyse 2 avec deux groupes.

Nous retrouvons le découpage de la section en deux domaines avec la limite au niveau des traces 130, 138. Toutefois le domaine dune contient moins de traces affectées comme étant du domaine platier.

SECTION 12

(Horizontalisation à 800 ms.)



Legende

des sondages discontinués

□ sondage actif

■ sondage inactif

- 4. LIGNES DES ÉCHANGES D'ÉLÉMENTS PLASTIQUES

- 5. FRACTURE S. TYP. - (170 cm. 200 ms)

1	2	3	4	5	6	7	8	9	10	11	12
178 m	178 m	178 m	178 m	178 m	178 m	178 m	178 m	178 m	178 m	178 m	178 m
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
276 - 295	26 - 34	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213	199 - 213
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241	227 - 241
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114	87 - 114
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258	223 - 258
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78	51 - 78
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN
276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295	276 - 295
GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN	GRANITE BRUN

Fig. 95

Les résultats des classifications confortent ceux obtenus par analyse factorielles discriminante. Il s'agit maintenant de comparer ces résultats avec ce que nous connaissons de la zone d'étude.

5.2.5. Comparaison des résultats avec les interprétations antérieures

Nous avons reporté sur une carte isopache temps du réservoir oolithique, réalisée en 1980, l'affectation des traces de la section 12 (Figure 96).

Nous constatons :

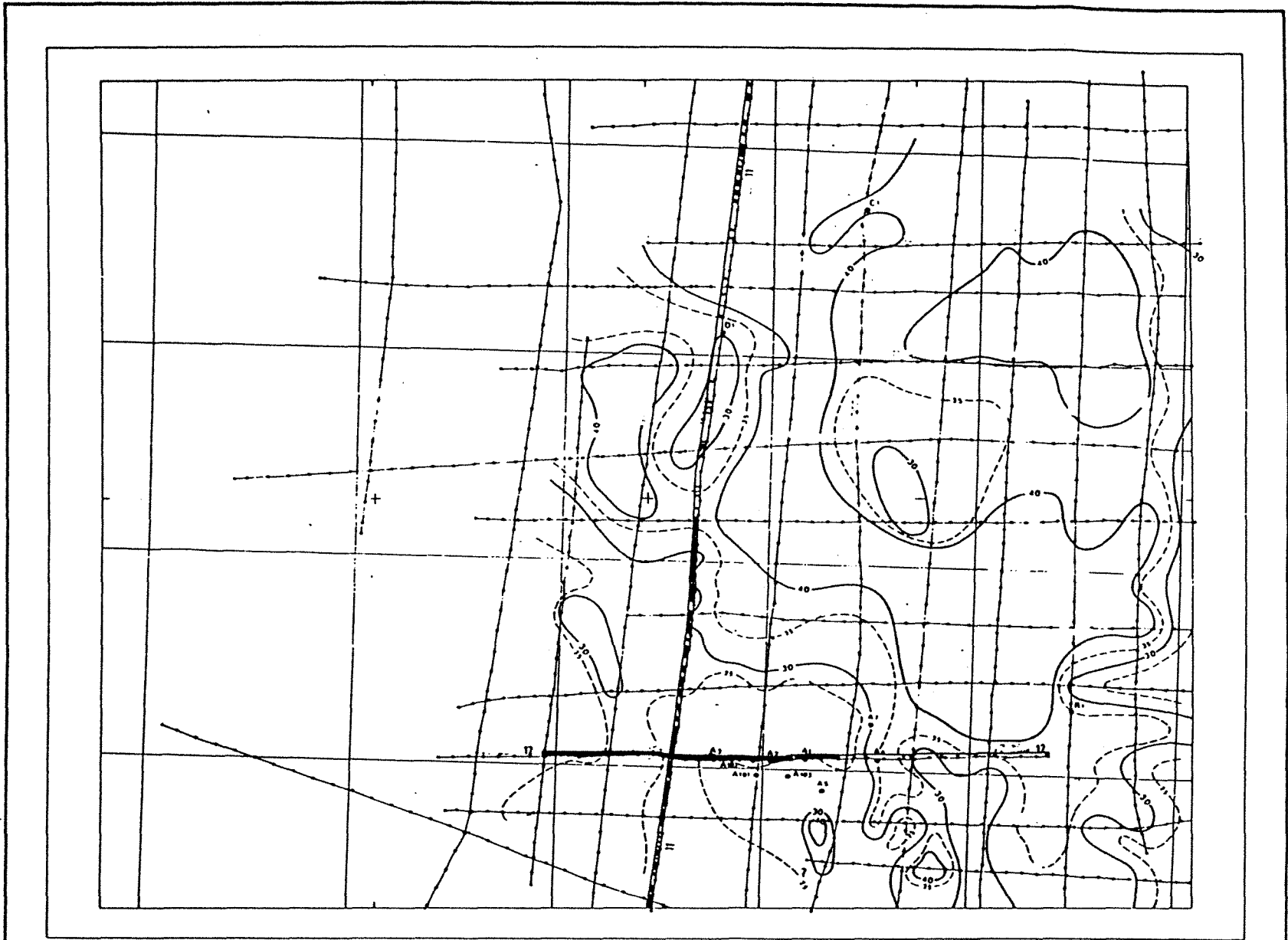
- que les traces adjacentes aux puits 1 et 2 sont correctement affectées (domaine de platier),
- que le changement de domaine correspond bien à l'épaississement oolithique (isopache > 28 ms temps double).

Pour effectuer un contrôle supplémentaire, nous avons traité une section perpendiculaire : la section 11 sur laquelle nous avons effectué une classification à trois groupes et une analyse discriminante en reprenant les pôles de l'analyse 1.

Nous retrouvons deux grands domaines, le changement s'effectuant autour de la trace 280. Toutefois, ces deux domaines apparaissent plus hétérogènes : passées de zones dune dans le domaine platier et inversement ; nous retrouvons toutefois les limites de l'étude réalisée en 1980 (Dumay, Kenaan).

Cet épaississement de l'oolithe s'accompagne d'une disparition de l'oncholithe (Figure 97), ce qui entraîne un fort contraste de vitesse (Figure 97) entre l'oolithe et le faciès lagon interne. Ce fort contraste de vitesse implique l'existence d'un pic plus marqué au toit du réservoir d'où l'explication du choix de la variable E6 comme première variable discriminante.

La méthode proposée permet bien de reconnaître les faciès sismiques (Figure 98). Les limites de faciès trouvées sont en accord avec notre connaissance de la zone ; même si un interpréteur exercé (spécialiste en stratigraphie sismique) aurait pu visuellement séparer les deux domaines, **n'importe quel** interpréteur peut appliquer cette méthodologie statistique pour réaliser **rapidement** une carte de faciès sismique, avec une grande précision de découpage et une **quantification des différences existant entre les faciès**.



Résultats des analyses discriminantes:

Traces associées au pôle PLATIER



Traces associées au pôle DUNE



CARTE ISOPAQUES-TEMPS(TO)
DE L'OOLITHE

J. Dumay -T. Kenaan (1980)

Fig. 96

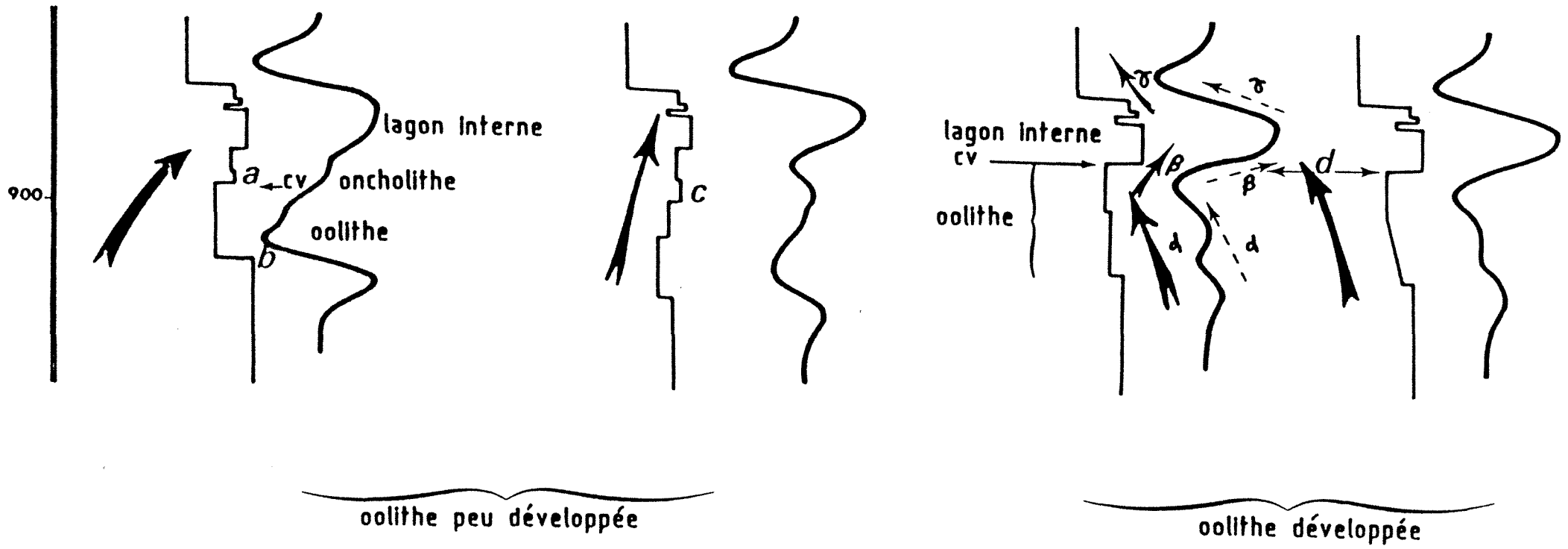
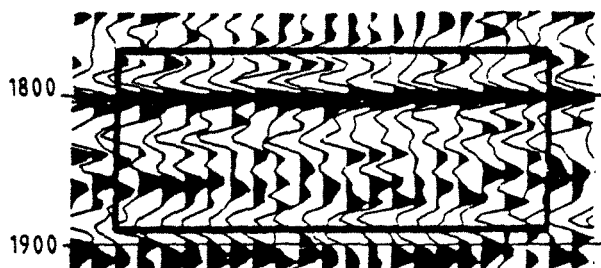


Fig. 97

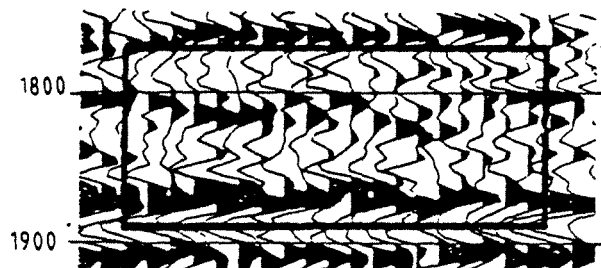
FACIES SISMIQUES SELECTIONNES
PAR LES PROGRAMMES D'ANALYSES

GROUPE DUNE

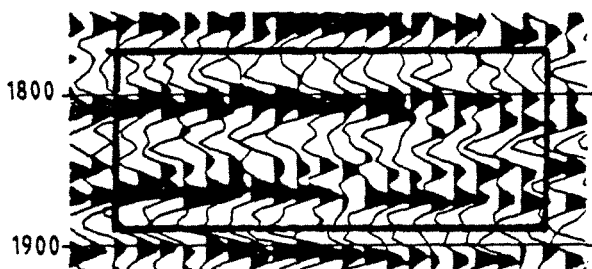
GROUPE PLATIER



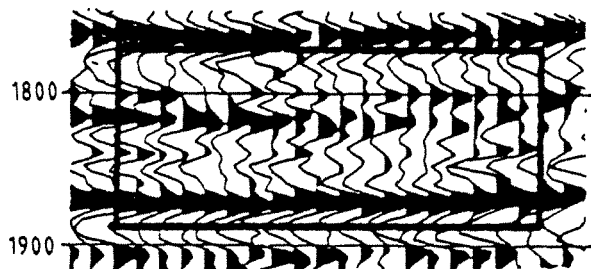
Section 11
traces 180 à 160



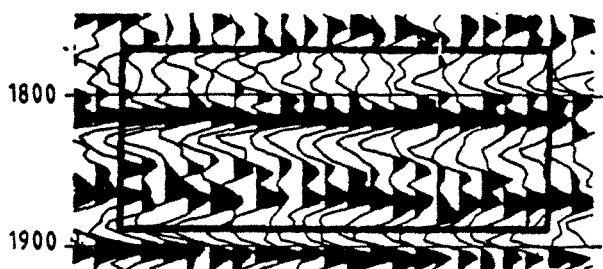
Section 11
traces 440 à 420



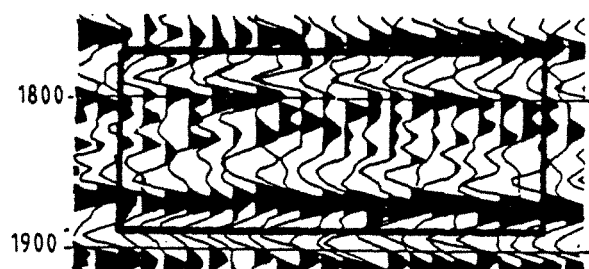
Section 12
traces 95 à 75



Section 12
traces 220 à 200



Section 12
traces 125 à 105



Section 12
traces 390 à 370

5.3. Etude des faciès sismiques en environnement turbiditique argilo-gréseux

5.3.1. Cadre géologique

La zone étudiée est une structure productrice au Gabon. Elle est limitée au Sud par un dôme de sel et au Sud-Ouest par un ensellement avec une autre structure voisine.

Les variations de faciès analysées concernent les réservoirs de l'Anguille Supérieur entre les puits T1 et M1 (Figure 99) ; ces réservoirs sont constitués de niveaux gréseux turbiditiques intercalés dans une série argileuse. La dénomination des réservoirs Anguille est celle du puits T1. De haut en bas, nous distinguerons les réservoirs R0 et R1, puis R2 et R3 ; le réservoir le plus productif à T1 est le réservoir R1, mais R2 est assez intéressant ; R3 est de qualité médiocre.

Au puits M1, outre ces unités (Figure 99), on observe le développement d'une unité réservoir supplémentaire "inter R1/R2".

Si l'on compare les réservoirs entre T1 et M1, on constate une dégradation générale vers le puits M1 :

- les réservoirs R0 et R1 demeurent isopaches mais R1 est dégradé à M1 ;
- le réservoir R1 devient complètement argileux à M1 ;
- le réservoir R3, déjà de mauvaise qualité à T1, s'affine vers M1.

Par contre, le meilleur réservoir à M1, l'unité inter R1/R2, n'existe pas à T1.

Dans cette étude, nous avons essayé de caractériser les faciès sismiques au niveau des unités inter R1/R2 et R2 qui montrent les variations les plus importantes entre les deux puits. La méthodologie statistique développée a été appliquée avec pour objectif final la prédiction des extensions de ce que l'on pourrait appeler le faciès "T" (unité R2 développée, unité inter R1/R2 absente) et le faciès "M" (unité R2 réduite, unité inter R1/R2 avec 3 mètres de sables).

5.3.2. Données sismiques et calages aux puits

5.3.2.1. Données sismiques étudiées

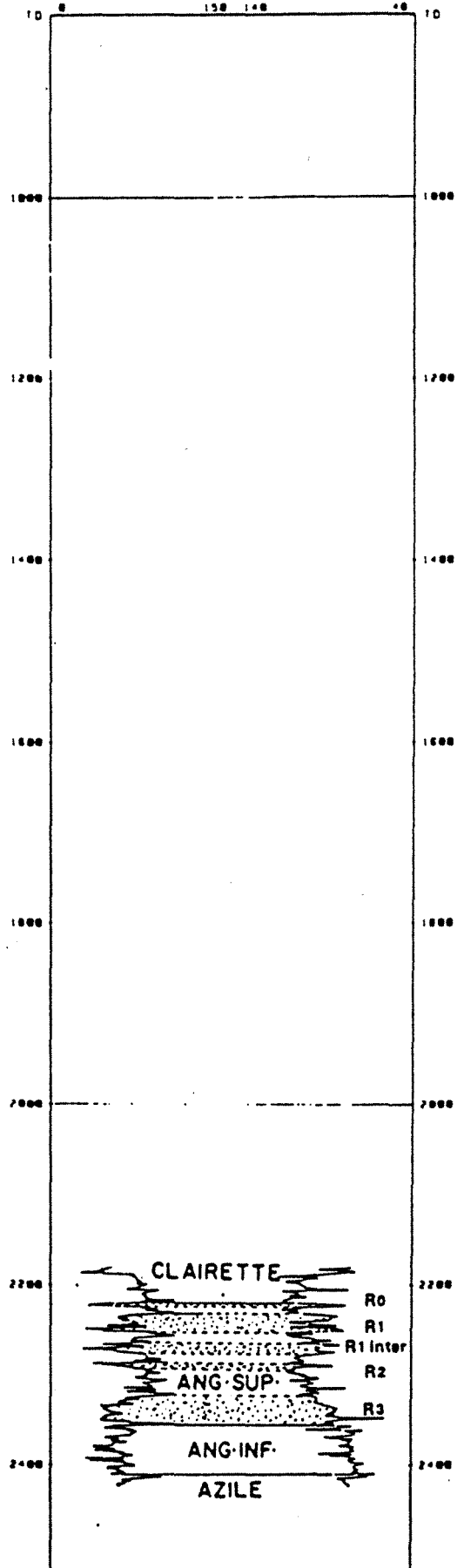
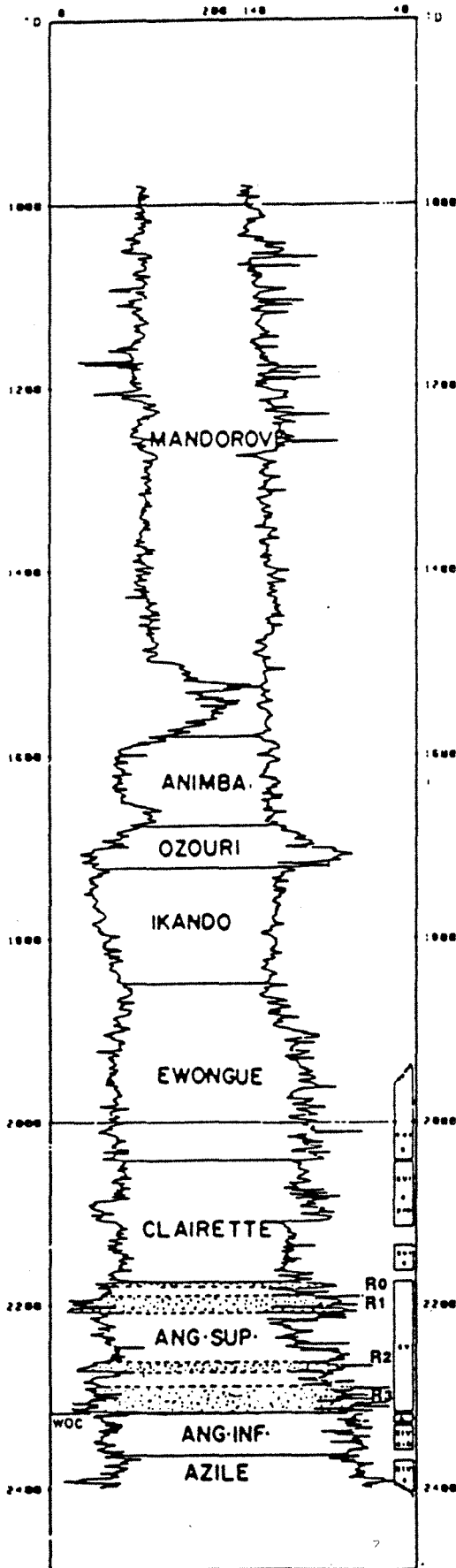
Nous avons travaillé sur la campagne sismique 3D 1982 : les lignes étudiées sont regroupées autour des puits T1 et M1 (lignes 148 à 152), puis la ligne 162 (puits T2), ainsi que les lignes 120, 122 (puits T4 proche), 108 et 110 (puits 20 g, 20 d), afin d'étudier la réponse sismique des traces à l'Ouest de l'ensellement limitant la structure.

T1

M1

GAMMA-RAY SONIC

GAMMA-RAY SONIC



Notre étude s'appuie sur l'interprétation sismique de la zone proposée par la filiale. Le calage avec les données de puits (Figure 100) utilisé pour l'interprétation des niveaux R1 à R3 est excellent aussi bien au niveau du puits T1 que du puits M1. D'autre part, des films synthétiques ont été calculés au puits T1 et montrent également un très bon accord avec la sismique. Le spectre d'amplitude du signal utilisé pour la réalisation des films a été estimé sur une fenêtre encadrant l'objectif et c'est la phase nulle qui donne le meilleur résultat.

Il faut noter qu'il y a un problème de calage pour le puits M1. En effet, au puits T1, il y a un bon accord entre le pointé sismique et la cote du toit du R3 donnée par le gisement. Par contre, au puits M1, le gisement met le toit du réservoir R3 à 3.347 m par rapport à la table de rotation du forage, alors que le calage avec la sismique du log Sonic et le pointé du R3 proposé positionnent le réservoir dans l'intervalle 3.300 - 3.350 m. Nous avons préféré tenir compte du pointé sismique proposé : la base du R3, bien calée à T1, se suit sans trop d'ambiguïtés sur la sismique vers M1. La corrélation obtenue n'est pas précise mais elle a une valeur générale sur la zone d'étude. D'autre part, à notre connaissance, il n'existe pas actuellement de corrélations diagraphiques fines entre les puits T1 et M1.

Pour faciliter les calculs ultérieurs des paramètres sismiques, nous avons horizontalisé la sismique à la base du R1 à 2.000 ms. L'analyse a porté sur une fenêtre de 140 ms, centrée sur les niveaux inter R1/R2 et R2, de 1.930 à 2.070 ms. La figure 101 montre les portions de traces, dans la fenêtre d'étude, autour des puits T1 et M1 sur la ligne 152.

A la base du R1, les réflexions sont continues, de forte amplitude vers T1, discontinues et faibles à M1. La réflexion à la base du R2, continue et de forte amplitude à T1, s'affaiblit et disparaît vers M1. Inversement, la réflexion, à la base du inter R1/R2, continue et d'amplitude moyenne à M1, disparaît vers T1.

5.3.2.2. Analyse de traces synthétiques

Nous avons calculé des traces synthétiques impulsionnelles et filtrées à partir des logs Sonic des puits M1 et T1. Les signaux utilisés sont des Ricker d'ordre 2 à phase nulle. Nous avons fait varier leur pseudo-période et retenu le signal dont le spectre d'amplitude est le plus proche de celui estimé d'après la sismique réelle sur une fenêtre encadrant les niveaux étudiés (pseudo-période 45 ms). Ces traces synthétiques sont représentées avec le Sonic sur les figures 102 (puits T1) et 103 (puits M1).

Le meilleur calage entre les films synthétiques et les traces réelles est obtenu pour la section 152 (Figure 100) ; d'autre part, c'est de cette ligne que les puits T1 et M1 sont les plus proches. Ceci suggère de choisir les traces pilotes sur cette ligne.

CALAGE AVEC LES TRACES SYNTHÉTIQUES

M1
400

M2
500

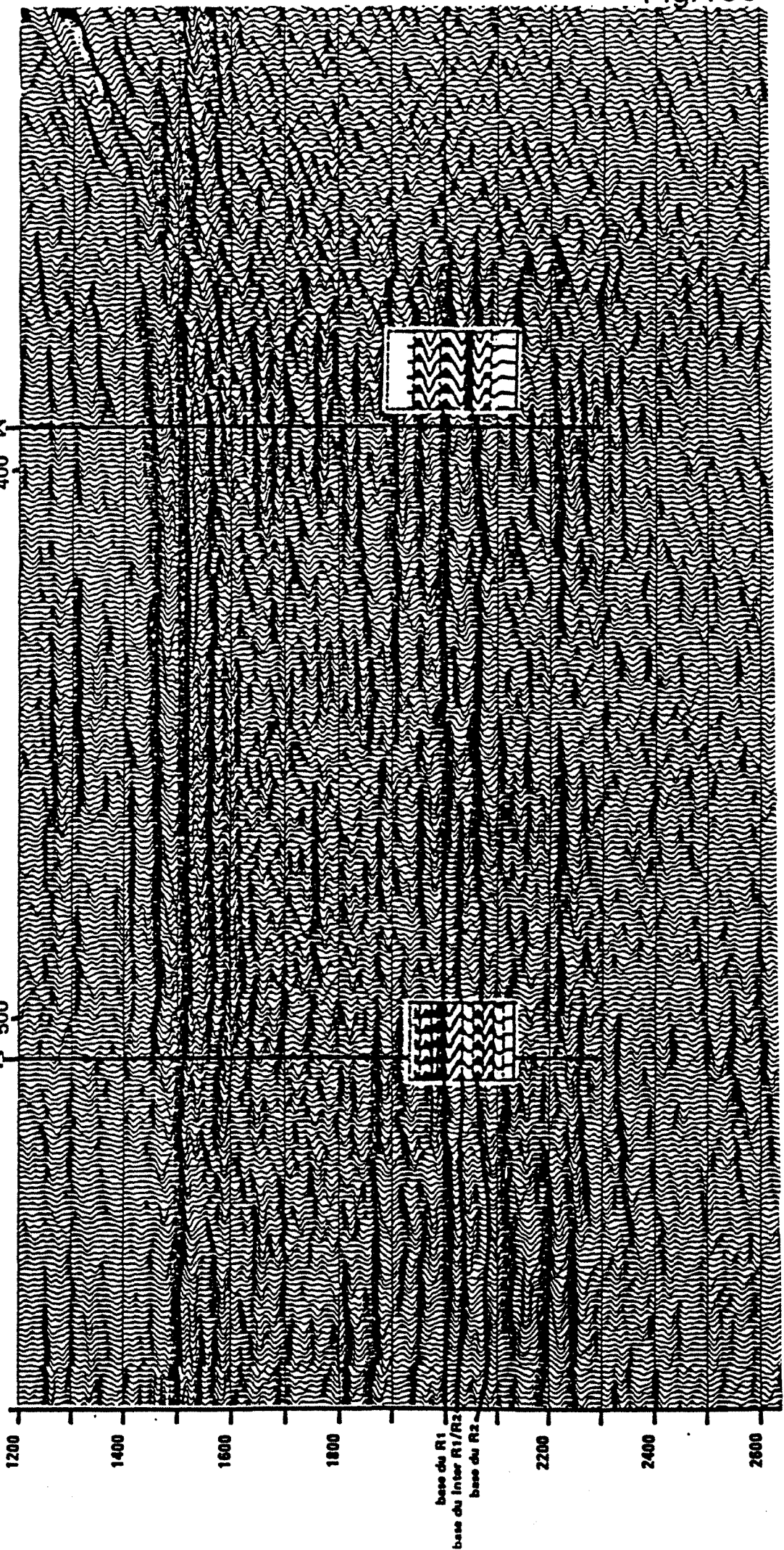


Fig. 100

LIGNE 152
HORIZONTALISATION A 2000 ms
(BASE DU R1)

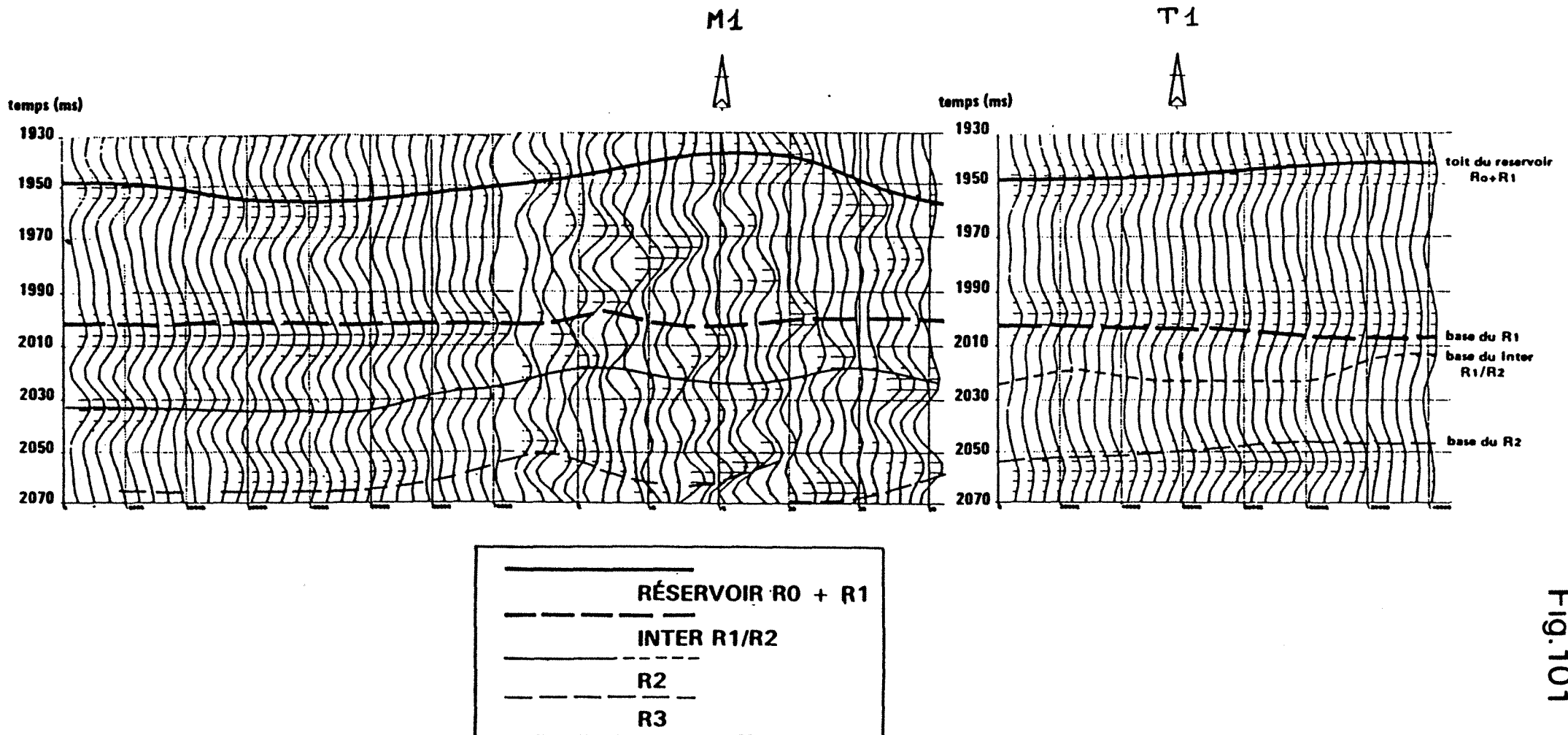


Fig. 101

T1

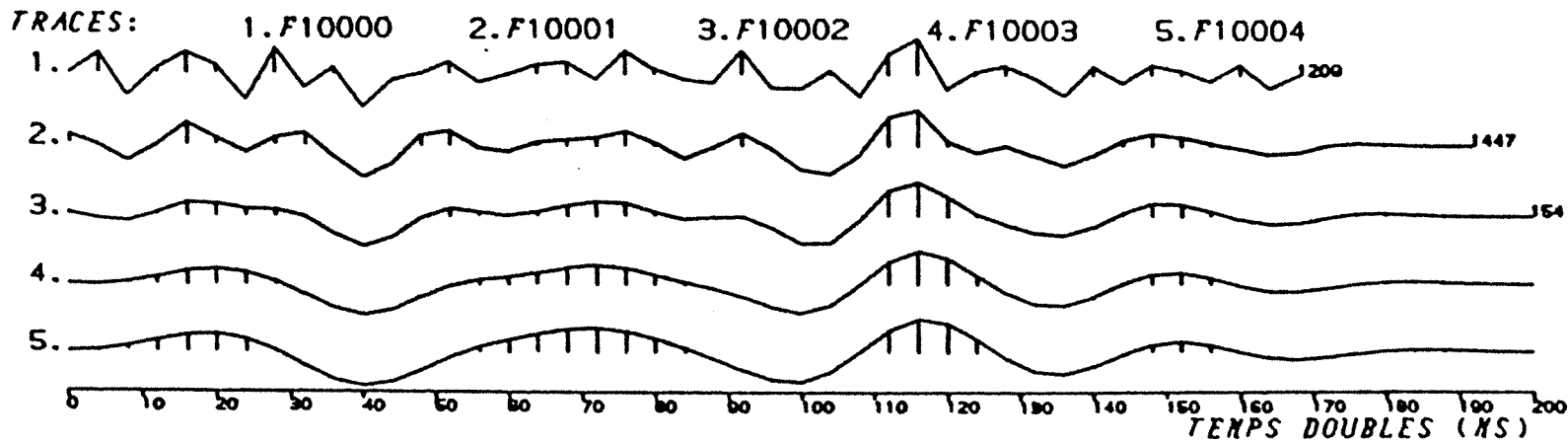
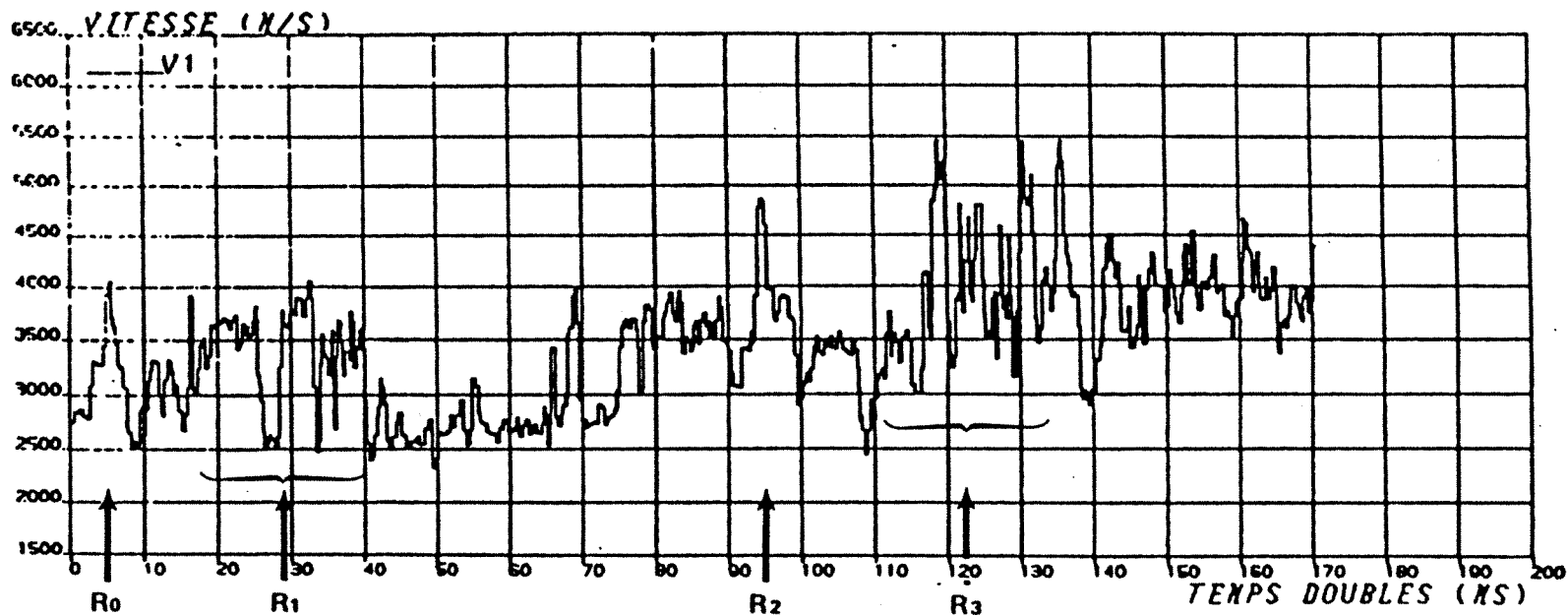
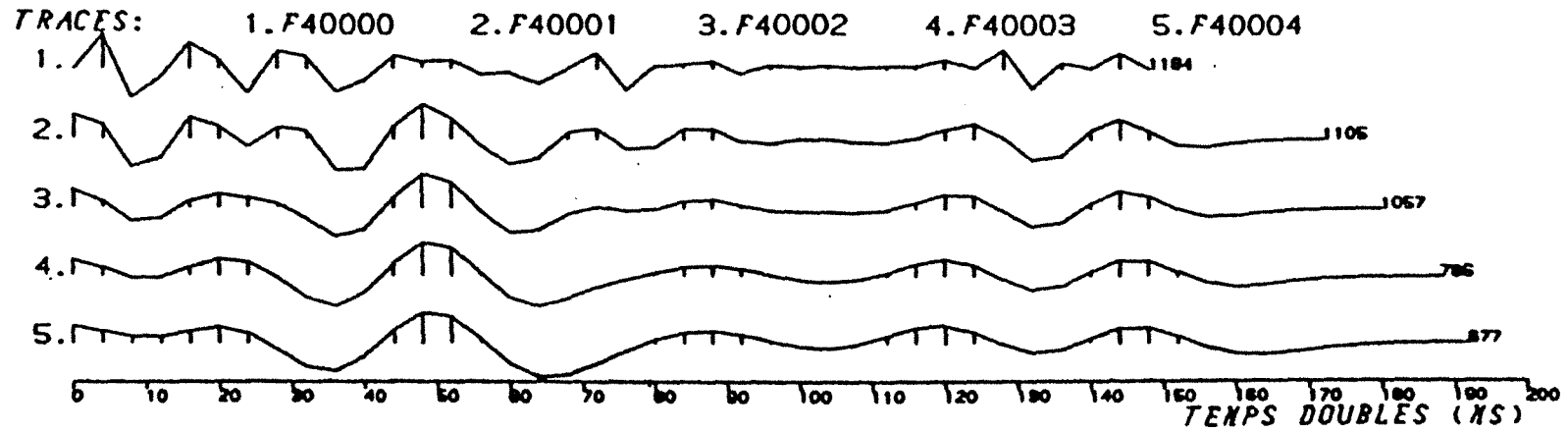
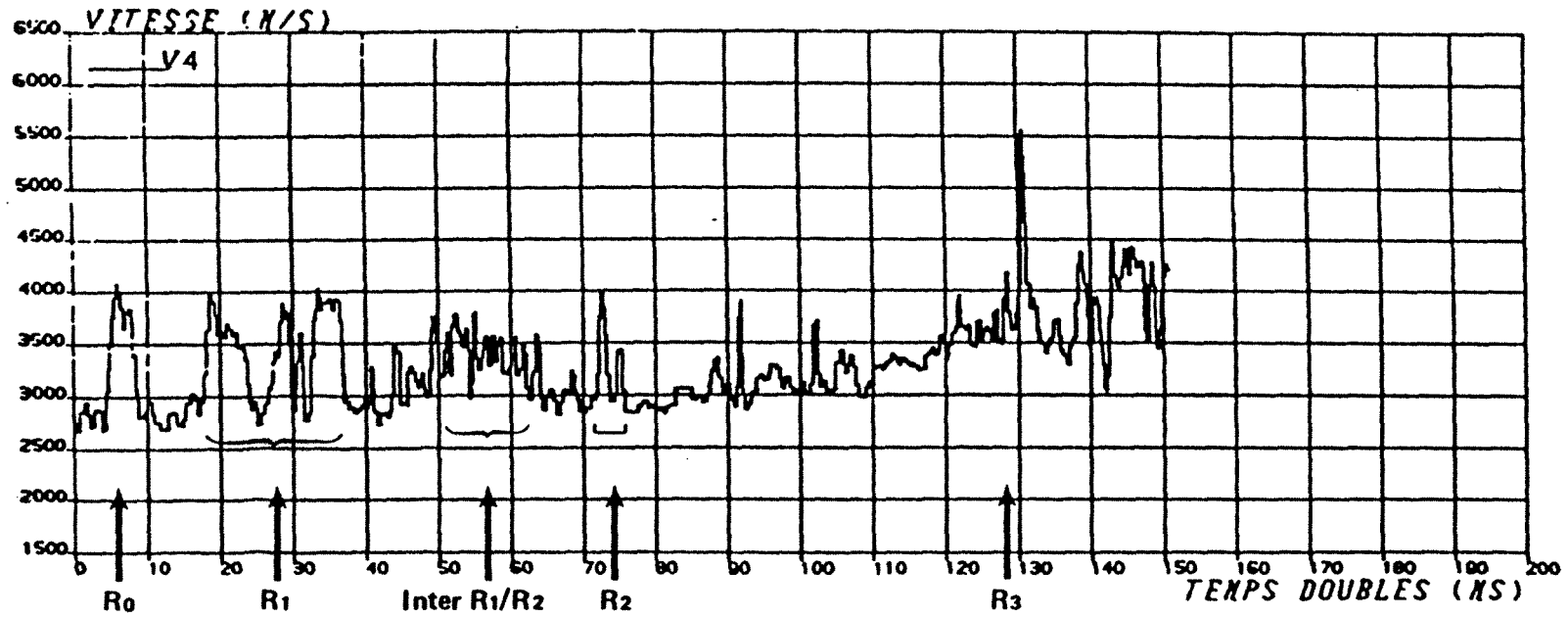


Fig. 102

M 1



IMPULSIONNEL

PP25

PP32

PP40

PP45

Fig. 103

5.3.3. Phase d'apprentissage

5.3.3.1. Choix des traces de référence

Nous avons choisi comme traces de référence des faciès sismiques étudiés, les traces de la section 152 adjacentes respectivement aux puits T1 et M1 (Figure 101). Il s'agit, en fait, de trouver environ une trentaine de traces caractéristiques et suffisamment homogènes entre elles.

Les 30 traces encadrant la projection du puits T1 sur les sections 152 (Figure 101), 151 et 153 présentent toutes des caractéristiques analogues. La section 152 ne présente donc pas, au niveau du puits T1, une anomalie locale.

D'autre part, le paquet de traces est relativement homogène ; en particulier, la moyenne des traces est bien représentative (Figure 104) de la majorité des traces (faible dispersion des caractéristiques).

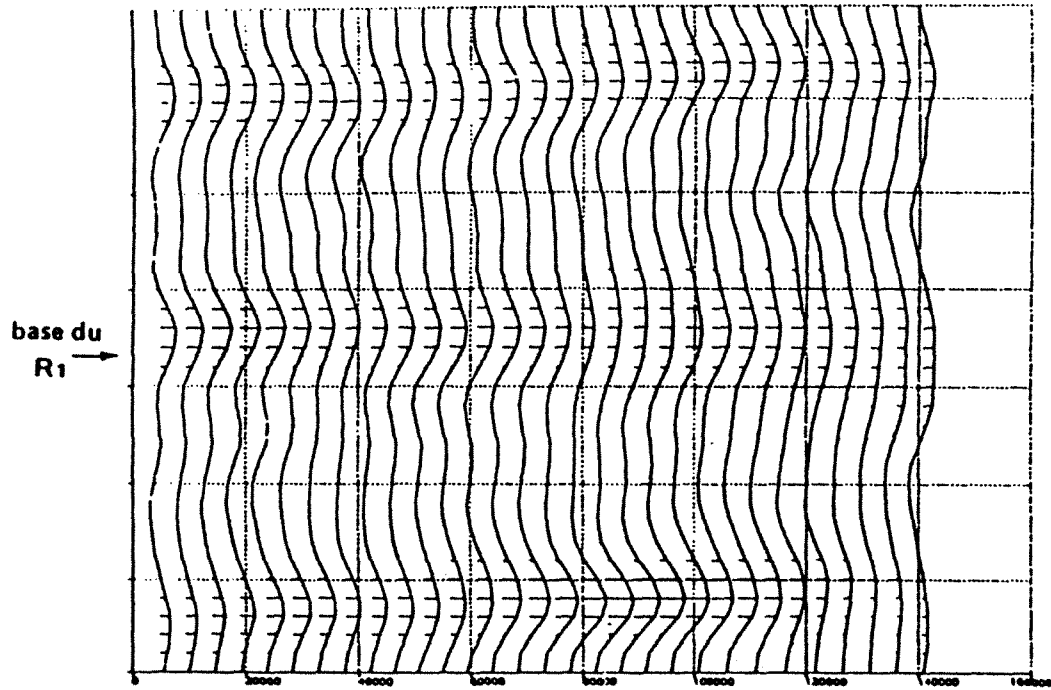
Nous avons vérifié aussi que les moyennes des traces adjacentes au puits T1 avaient un spectre et une autocorrélation comparables quelle que soit la ligne (152, 151 ou 153, Figure 105).

Par contre, dans la zone proche de M1, il est beaucoup plus difficile de trouver un ensemble de traces homogènes (Figure 101) bien que le film synthétique soit très bien calé localement (Figure 100). La moyenne des 30 traces diffère sensiblement des traces elles-mêmes (Figure 106) mais, d'une section à l'autre, (Figure 107) on retrouve les mêmes événements marquants (apparition d'un pic supplémentaire et affaiblissement de la réflexion à la base de l'intervalle d'étude).

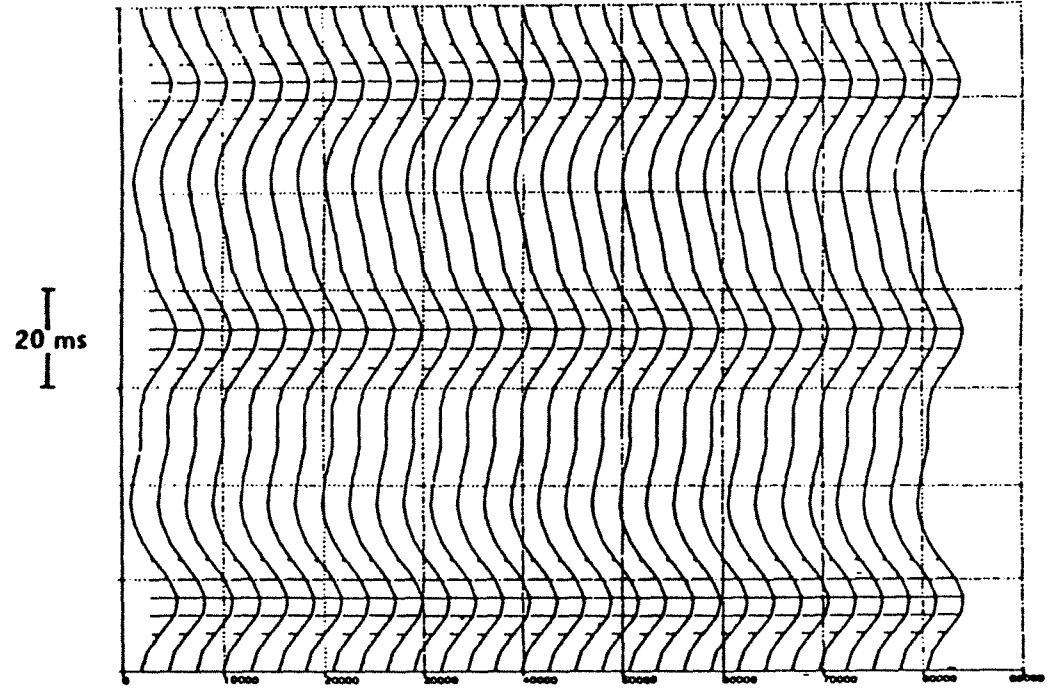
La zone pilote retenue pour le faciès M est ainsi légèrement décalée vers le Nord par rapport au puits M1, mais les traces retenues sont plus homogènes, tout en présentant les événements pointés comme caractéristiques du faciès.

Le choix des traces pilotes est un point capital de la méthodologie, ce qui justifie tous ces essais. Nous allons aborder maintenant l'étude descriptive des traces pilotes, qui peut également remettre en cause leur choix.

LIGNE 152
(PÔLE T)



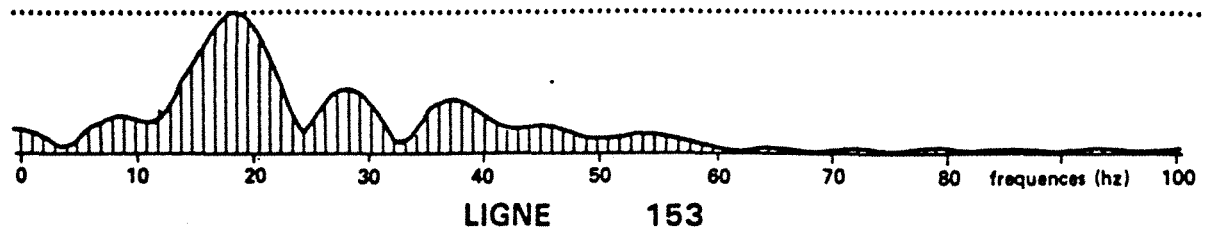
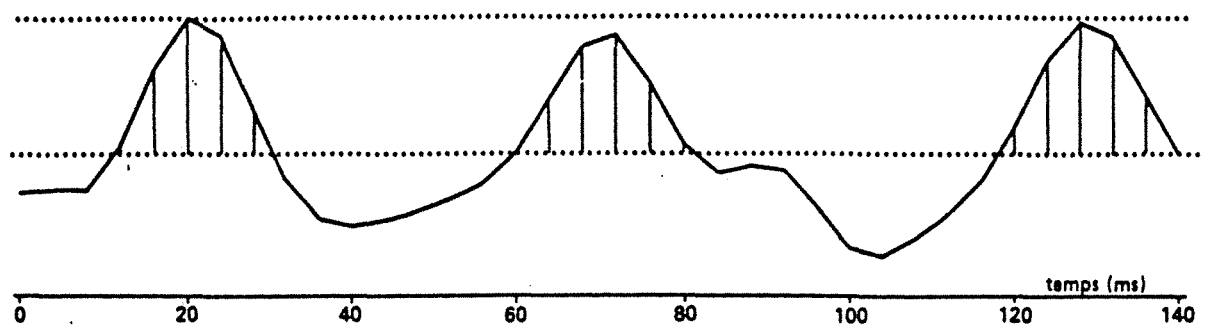
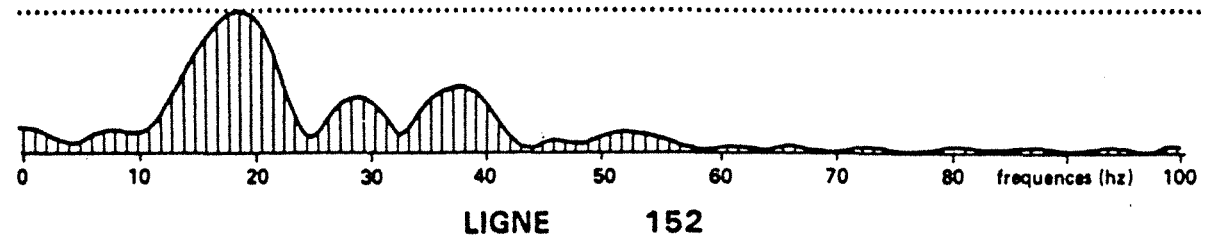
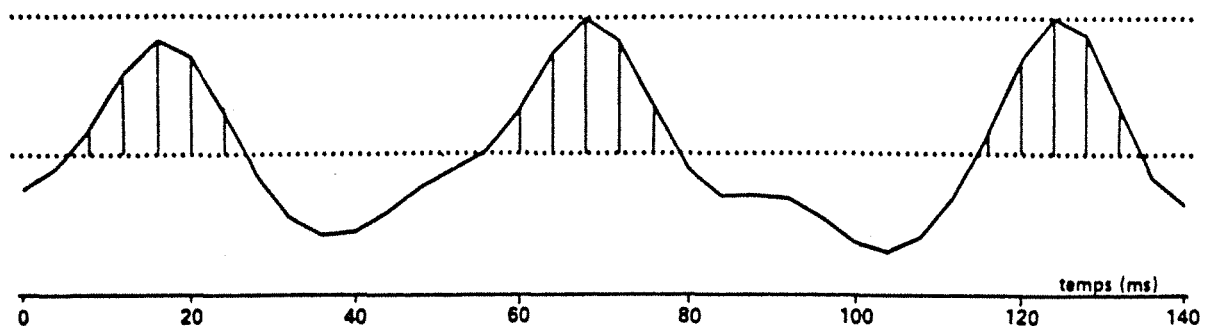
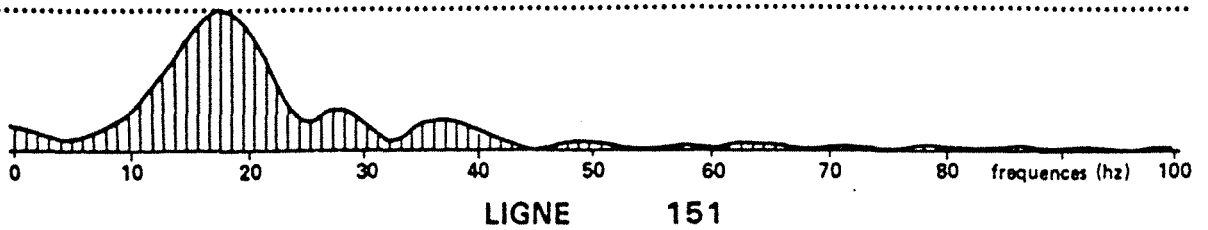
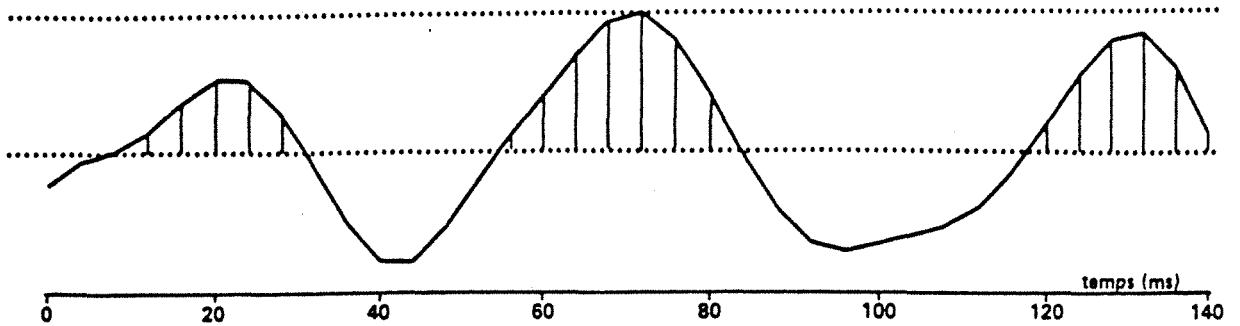
TRACES PILOTES



MOYENNE DES TRACES PILOTES

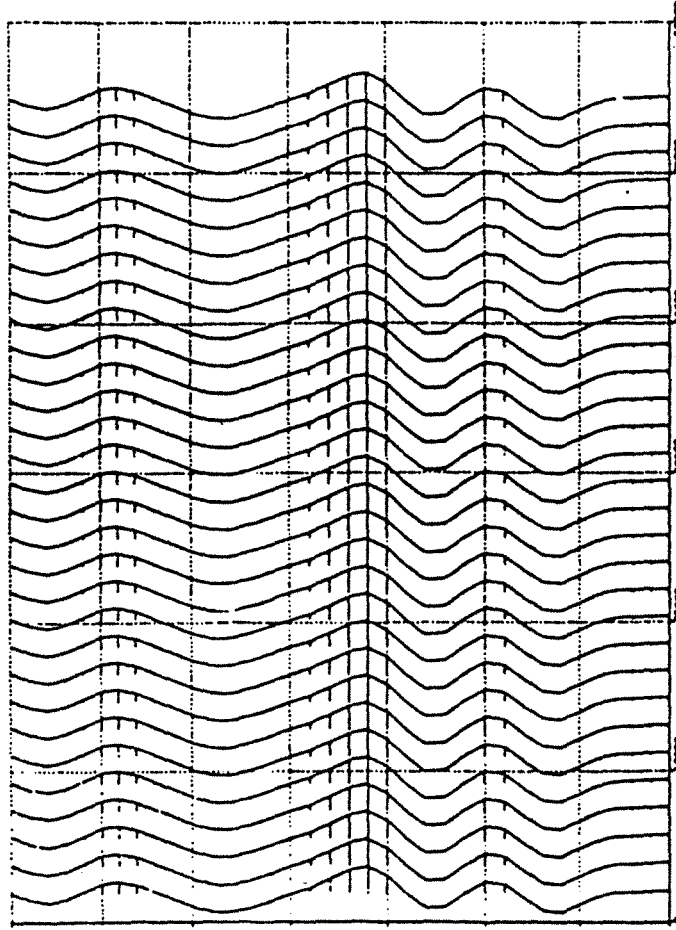
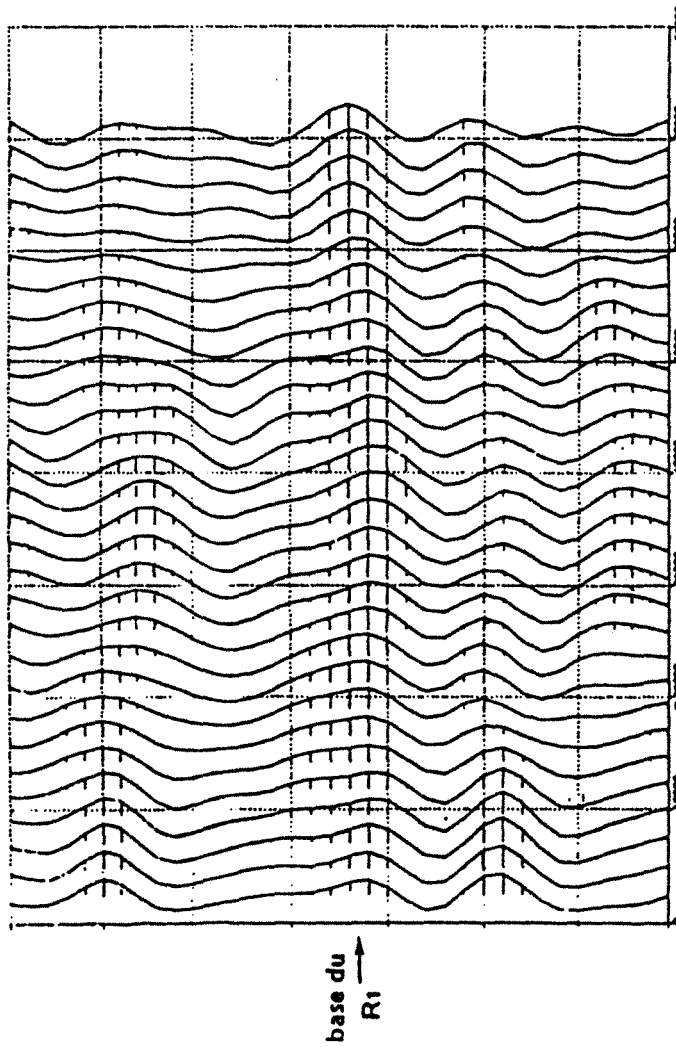
COMPARAISON DES MOYENNES DES TRACES ADJACENTES AU PUIS T1

ENTRE LES LIGNES 151 152 153



LIGNE 152

(PÔLE M)



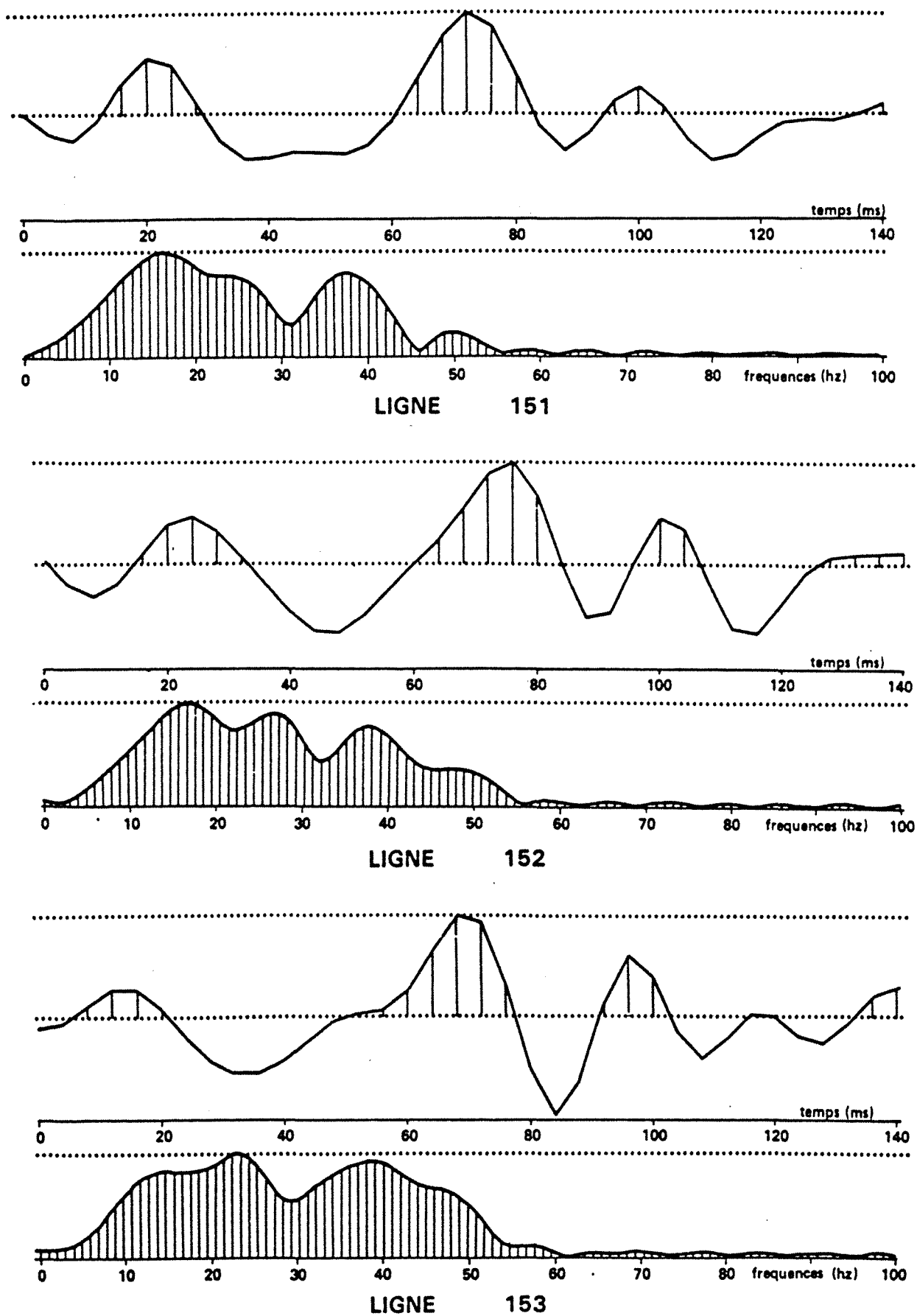
TRACES PILOTES

MOYENNE DES TRACES PILOTES

Fig.106

COMPARAISON DES MOYENNES DES TRACES ADJACENTES AU PUIS M1

ENTRE LES LIGNES 151 152 153



5.3.3.2. Etude des traces de référence

a. Calcul de l'autocorrélation et du spectre d'amplitude

Nous avons calculé le spectre d'amplitude et la fonction d'autocorrélation des traces moyennes des traces pilotes des faciès T et M (Figures 108 et 109). Les caractéristiques observées sont très proches de celles du spectre et de la fonction d'autocorrélation des traces synthétiques :

- Puits T1 : le spectre est assez mono-fréquentiel, plutôt basse fréquence (fréquence dominante 20 Hz).
L'autocorrélation présente un rebond d'assez forte amplitude vers 50 ms ($a = 0.4$).
- Puits M1 : la bande passante spectrale est plus large (fréquence dominante 30 Hz) et enrichie en hautes fréquences ; l'autocorrélation ne présente pas de rebond d'amplitude significative.

b. Calcul des paramètres sismiques

Les traces de référence sont caractérisées quantitativement au moyen de paramètres calculés sur l'autocorrélation, son spectre et le module du signal analytique.

c. Analyses factorielles

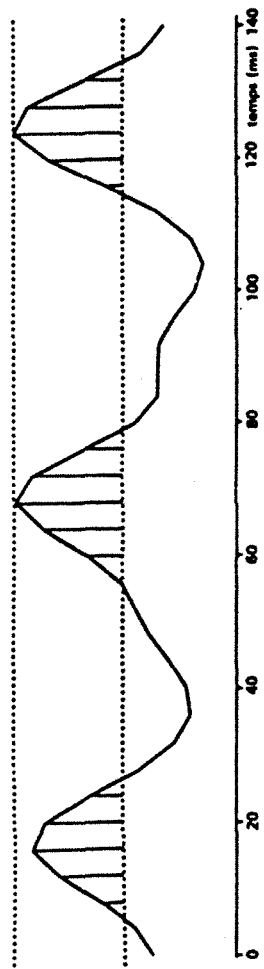
La question est de savoir s'il est possible de distinguer les deux groupes de traces pilotes grâce aux paramètres calculés. L'outil utilisé est l'analyse factorielle discriminante pas à pas ; les variables sont introduites au fur et à mesure en fonction de leur pouvoir discriminant.

Le programme mis en œuvre a l'avantage :

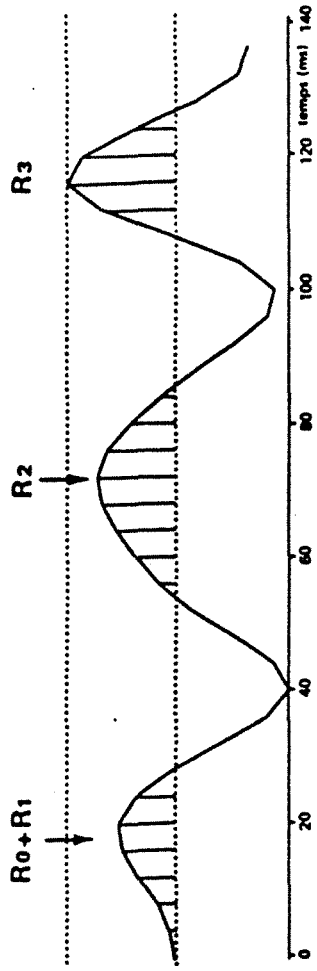
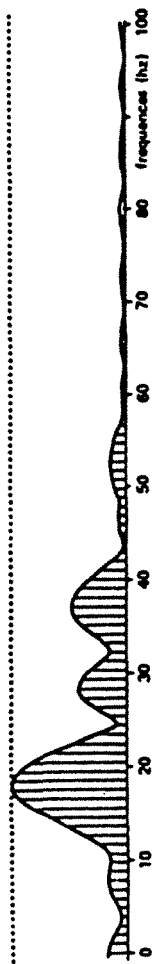
- de limiter le nombre de variables introduites (seules les plus discriminantes sont prises en compte),
- d'évaluer la fiabilité de la discrimination.

Le tableau 21 donne l'évolution de la discrimination entre les traces pilotes pour les quatre premiers pas. Au pas n°2, on constate que la discrimination est optimale. Le couple de variables sélectionnées est ETNT (énergie totale de l'intervalle d'étude) et Q8 (fréquence correspondant à 80 % de l'énergie spectrale).

FACIÈS T (152)



SPECTRE D'AMPLITUDE



SPECTRE D'AMPLITUDE

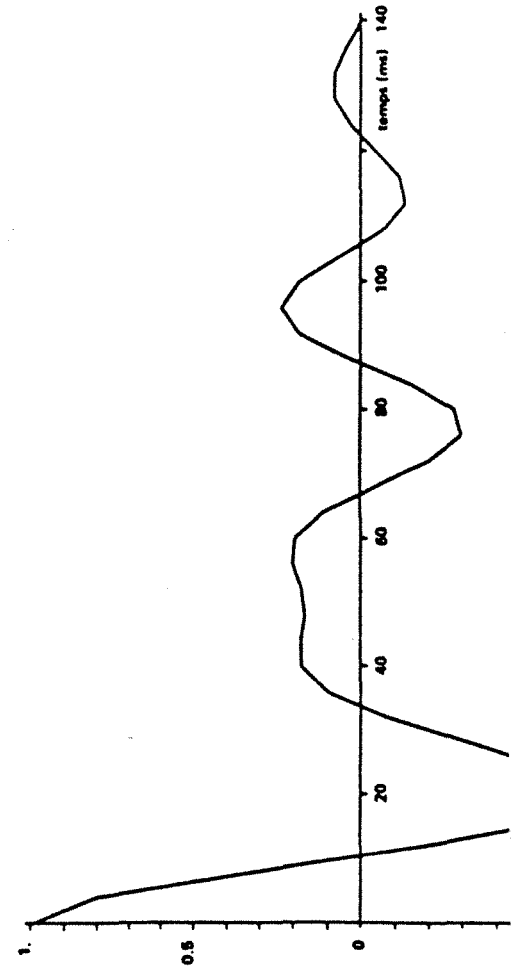
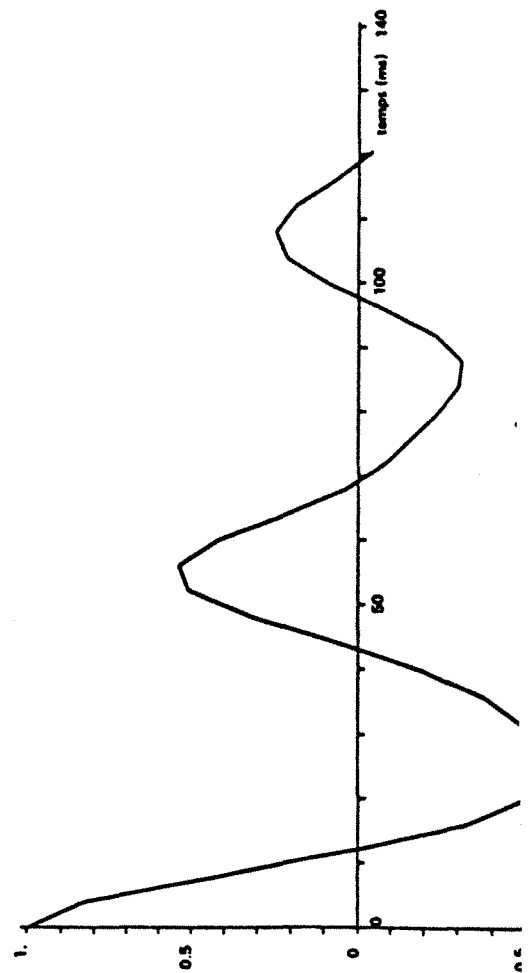
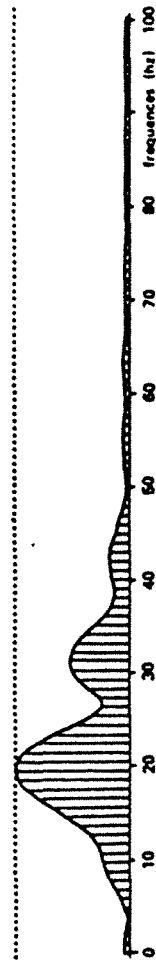
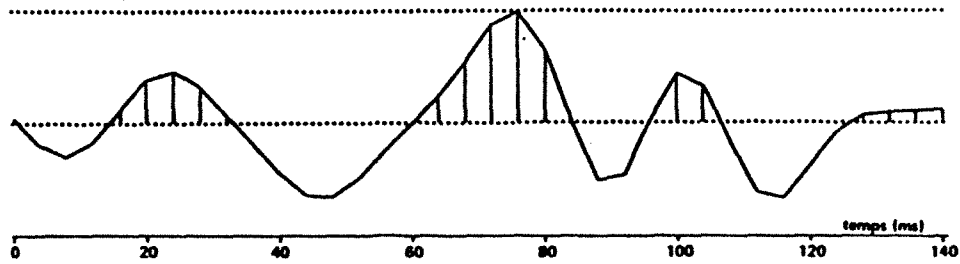


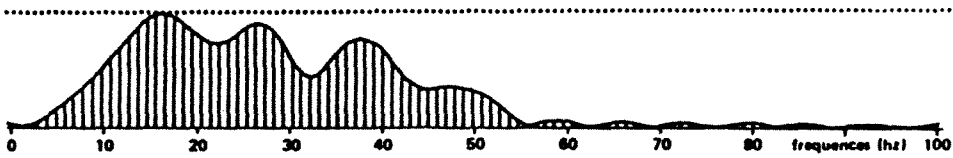
Fig. 108

FACIÈS M

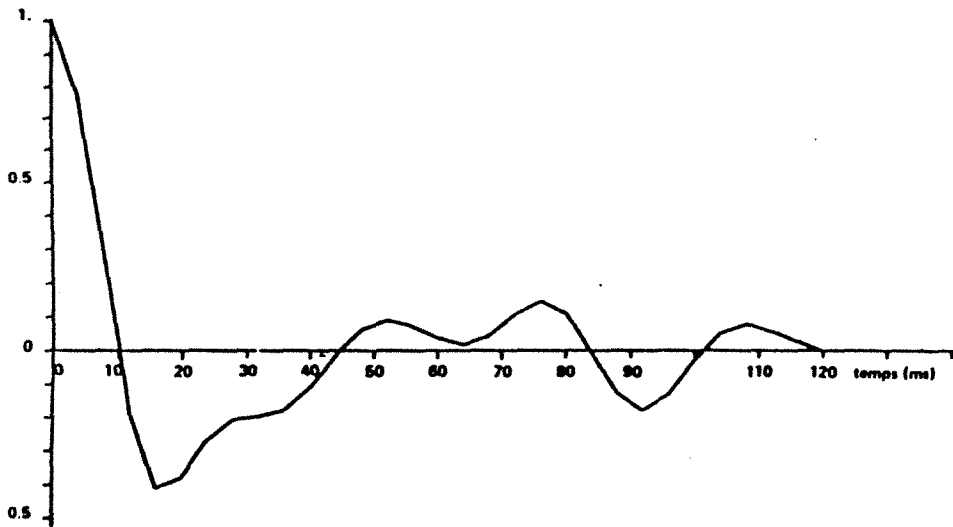
(152)



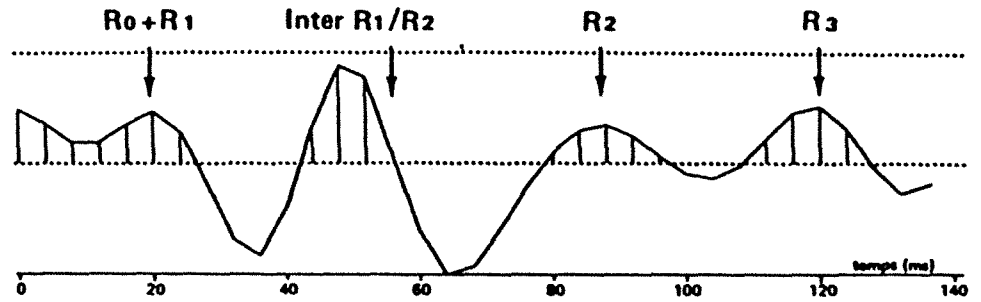
MOYENNE DES TRACES PILOTES



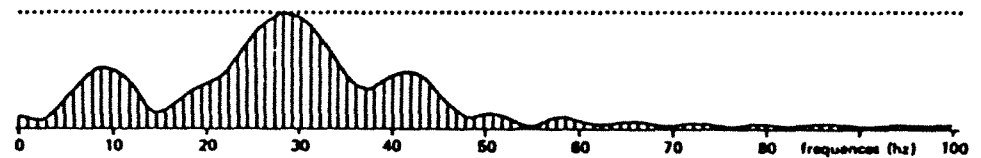
SPECTRE D'AMPLITUDE



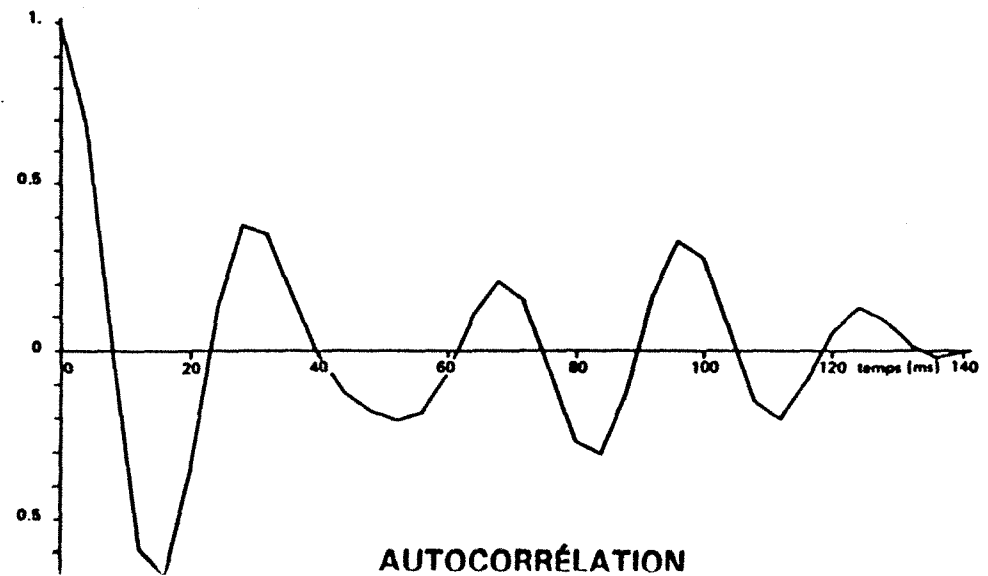
AUTOCORRÉLATION



TRACE SYNTHÉTIQUE



SPECTRE D'AMPLITUDE



AUTOCORRÉLATION

Pas	Variables entrées	% de bien classés
1	ETNT	97 %
2	ETNT, Q8	100 %
3	ETNT, Q8, AM2	100 %
4	ETNT, Q8, AM2, E7	100 %

Tableau 21 : A.F.D. pas à pas sur les traces d'apprentissage des pôles M et T

Si l'on supprime la variable ETNT, deux variables suffisent encore à assurer une discrimination optimale, c'est le couple AM1 (amplitude du rebond le plus important sur l'autocorrélation) et Q4 (fréquence correspondant à 40 % de l'énergie spectrale).

Comme les variables sont introduites au fur et à mesure, sans remettre en cause le choix des étapes précédentes, ceci prouve qu'après ETNT, la variable la plus intéressante est AM1 (variable introduite la première en l'absence d'ETNT).

Nous avons analysé l'évolution des quatre variables sélectionnées : ETNT, AM1, Q4, Q8, le long de la ligne 152 (Figures 110 à 113).

On constate que les variables sont relativement homogènes au niveau des traces pilotes. Les traces pilotes du pôle T se caractérisent par une forte énergie et un rebond sur l'autocorrélation d'amplitude importante. Au contraire, les traces proches du puits M1 montrent peu d'énergie, et les rebonds sont de très faible amplitude sur l'autocorrélation.

Quant aux variables sur le spectre, on observe, entre M1 et T1, un déplacement du spectre vers les basses fréquences (variables Q4 et Q8 plus faibles vers T1, c'est-à-dire que l'énergie du spectre est plutôt portée par les basses fréquences).

L'évolution des variables AM1, Q4 et Q8 correspond bien aux caractéristiques du spectre et de l'autocorrélation observées sur les traces pilotes (§ 5.3.3.2.a).

LIGNE 152
VARIABLE ETNT

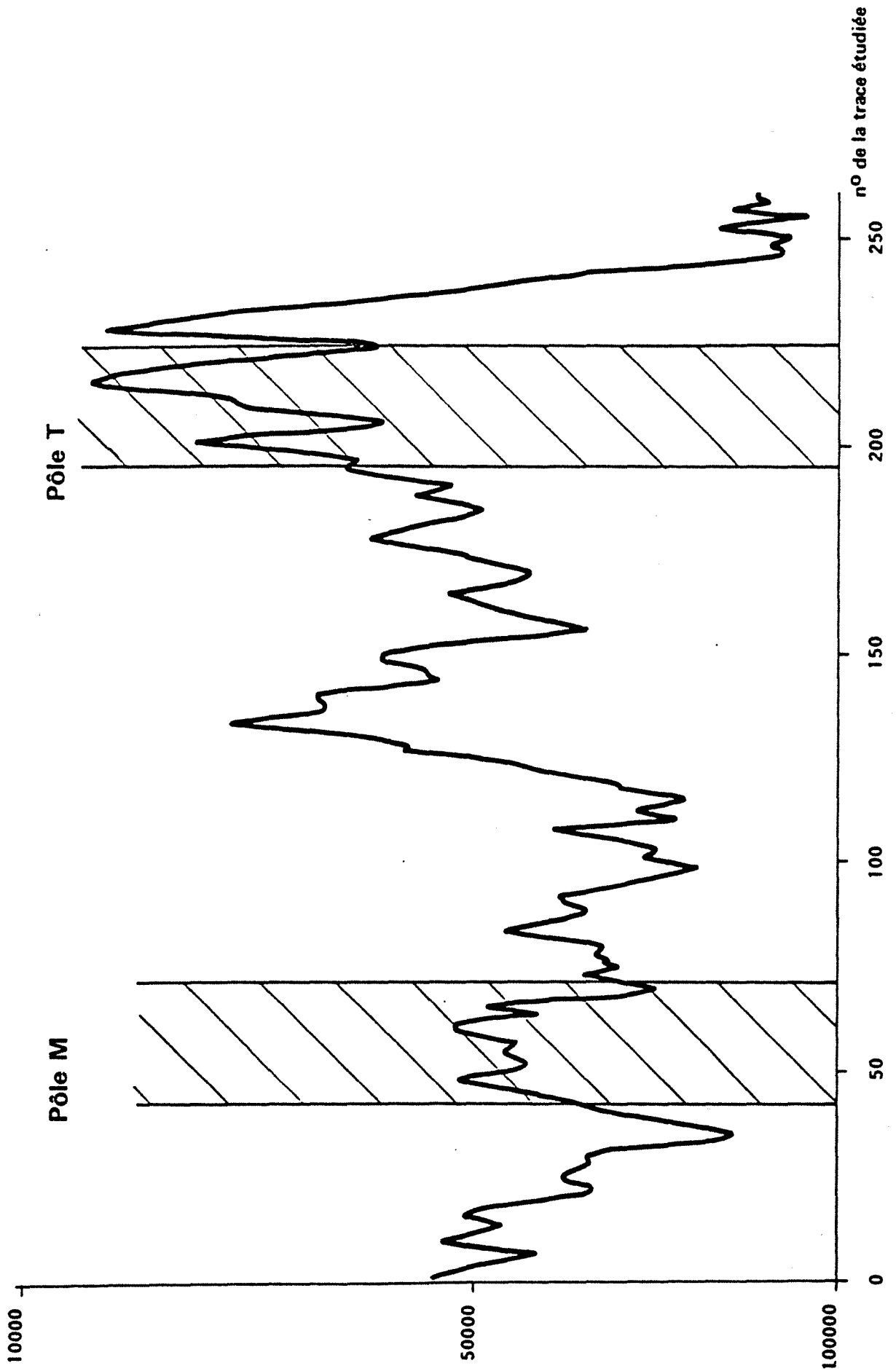


Fig.110

LIGNE 152
VARIABLE AM 1

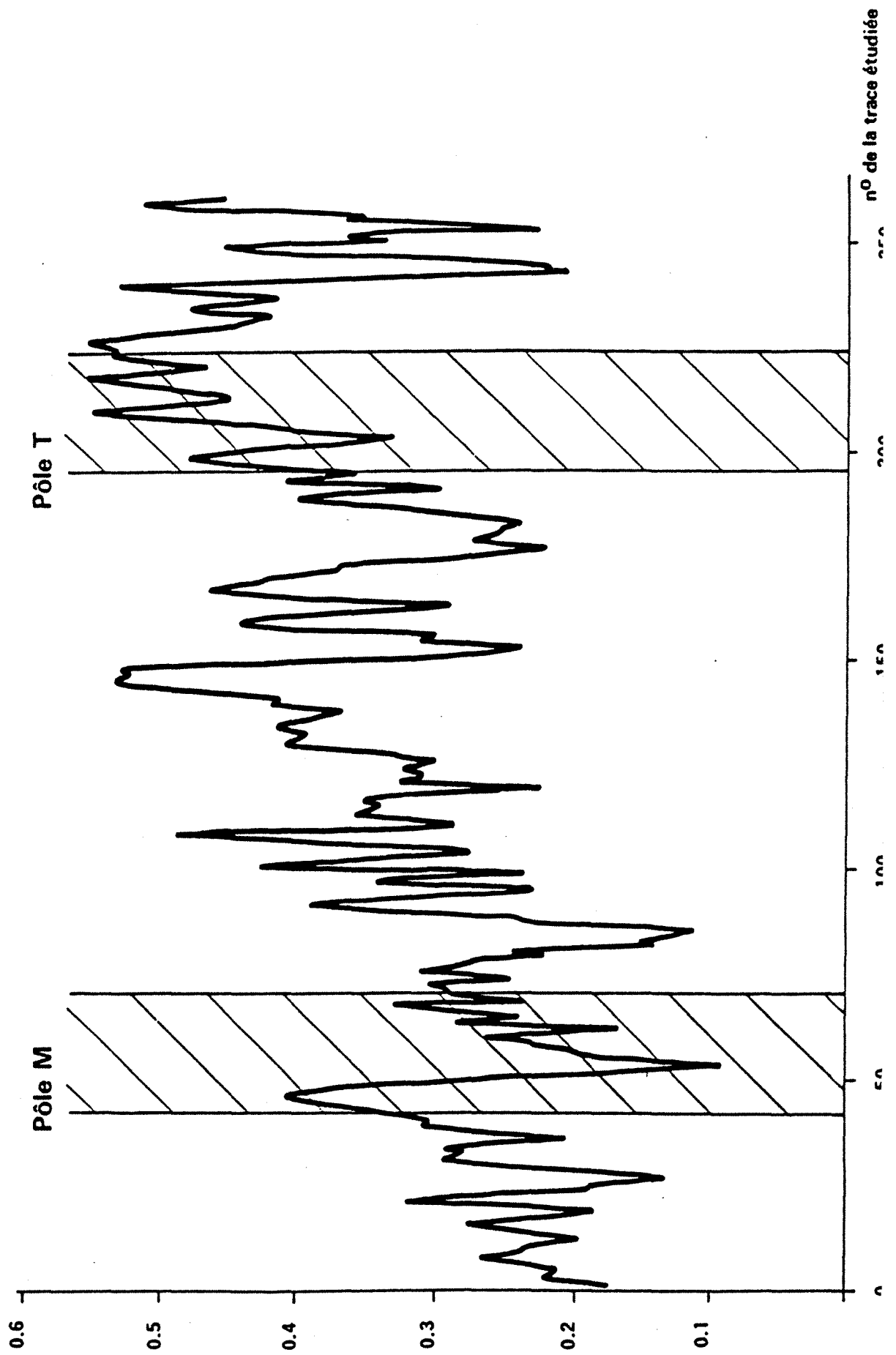


Fig.111

LIGNE 152
VARIABLE Q 4

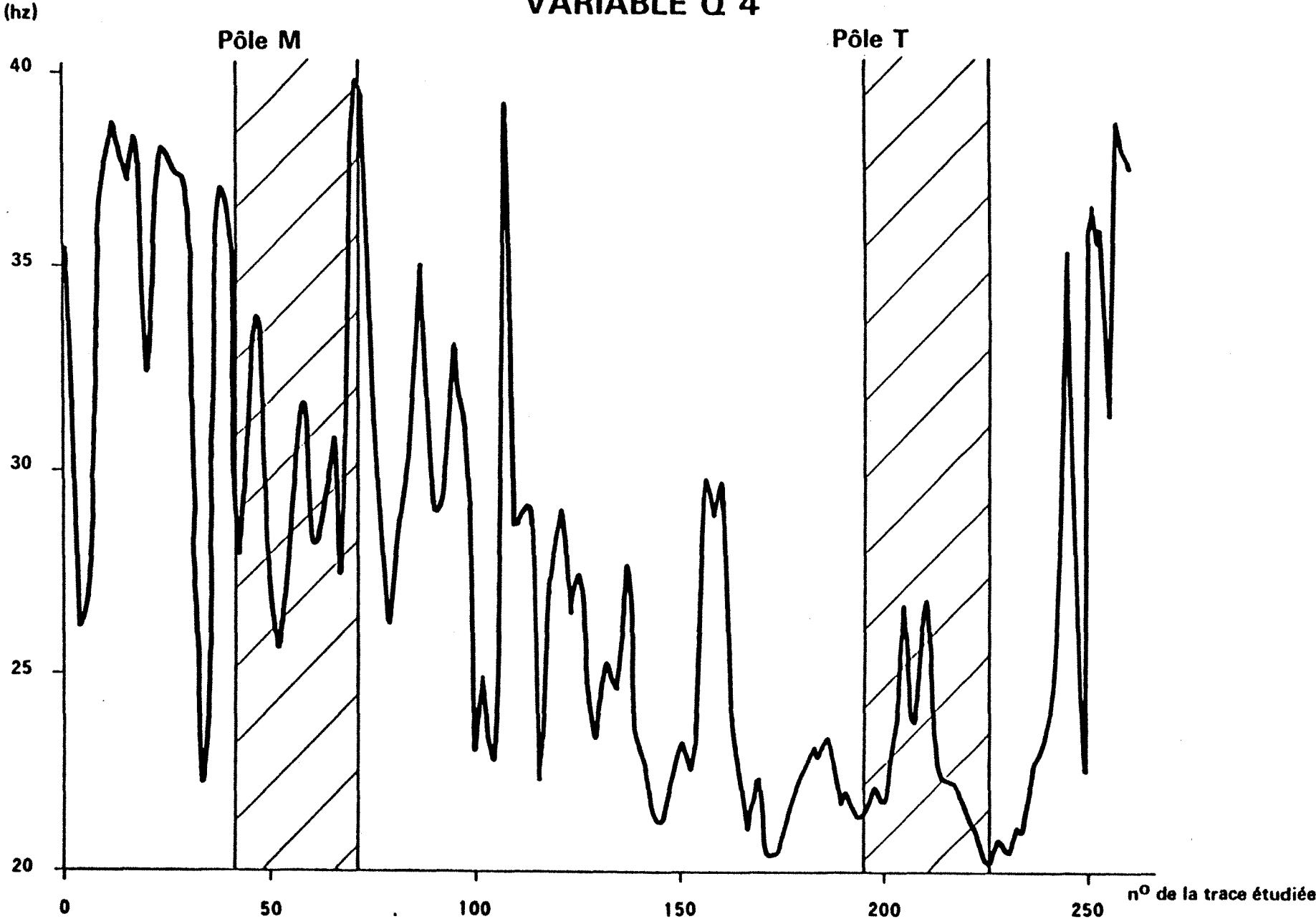
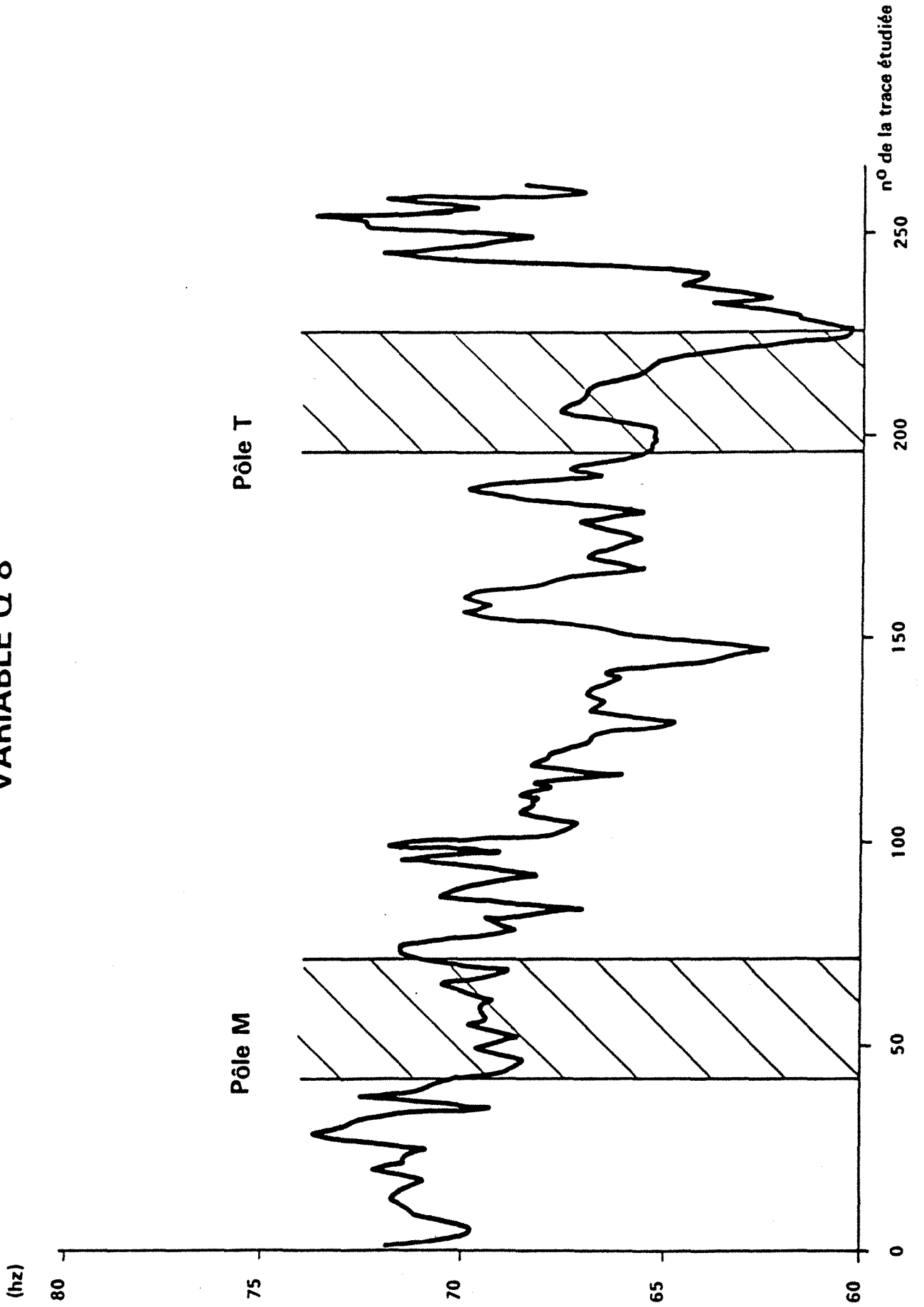


Fig. 112

Fig.113

LIGNE 152
VARIABLE Q 8



L'étude statistique réalisée assure la possibilité de caractériser le faciès sismique des traces pilotes au moyen des variables calculées ; elle permet en outre de sélectionner les paramètres les plus performants vis à vis de la discrimination souhaitée.

Il faut maintenant vérifier qu'il existe bien une liaison entre le faciès sismique caractérisé aux pôles et les évolutions géologiques. Nous avons donc réalisé des modèles à partir des données de puits.

5.3.4. Etape de modélisation pour la validation géologique des paramètres discriminants

Le modèle géologique est obtenu en corrélant les principales unités des puits T1 et M1, à partir des diagraphies Sonic et Gamma-ray. Les vitesses des couches sont calculées en interprétant et simplifiant le Sonic aux deux puits, puis par interpolation linéaire entre les deux pôles.

Nous rappelons (cf. § 5.3.2.1) que la position du R3 sur le Sonic au puits M1 correspond au calage avec la sismique (horizon suivi depuis T1).

Les sections impédances acoustiques puis traces synthétiques sont calculées à partir du modèle des vitesses (nous n'avons pas tenu compte des densités). Le signal choisi est celui utilisé pour réaliser les films synthétiques (Ricker d'ordre 2, à phase nulle, de pseudo-période 45 ms).

5.3.4.1. Premier modèle

Le premier modèle (cf. section impédance, Figure 114) est assez complet. Le log Sonic des deux puits (Figures 102 et 103) a été simplifié mais nous avons retenu le maximum d'éléments qui pourraient être significatifs pour la corrélation. Sur la figure 114, la section traces synthétiques montre l'évolution théorique de la réponse sismique entre les deux puits.

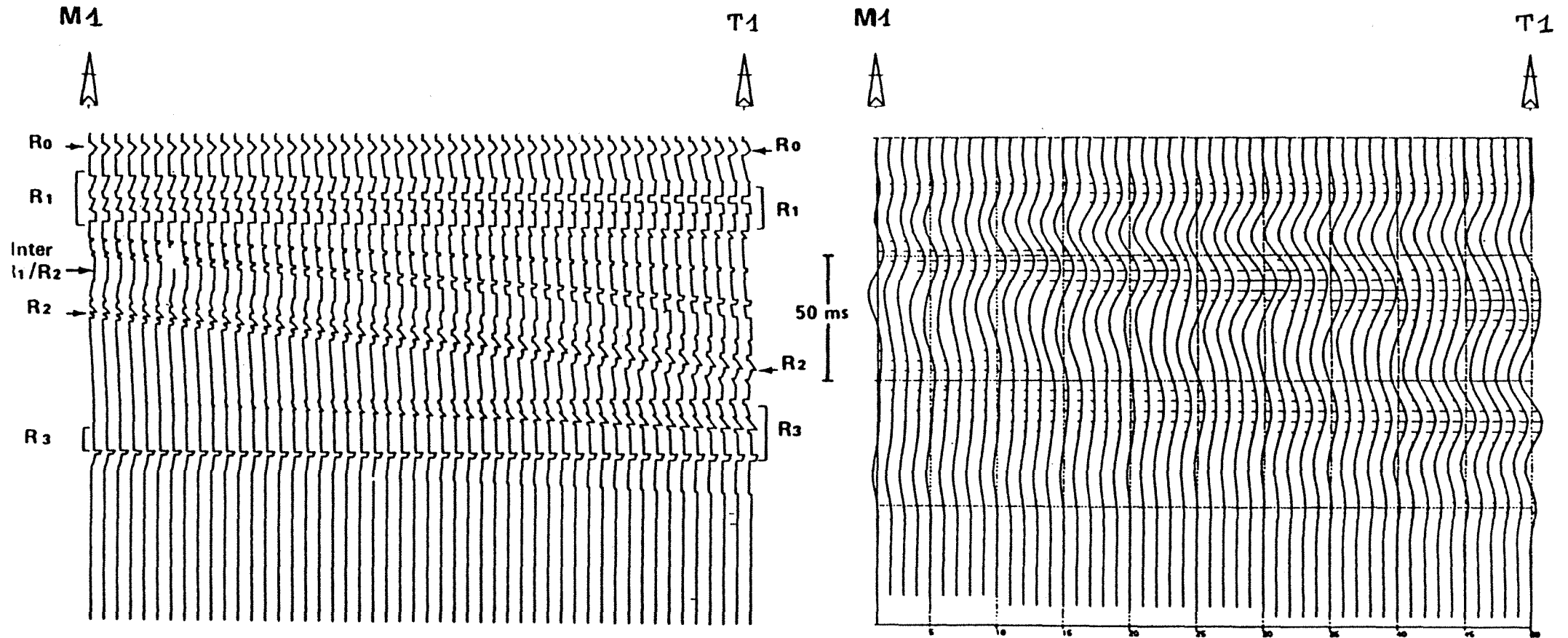
La comparaison entre les traces du modèle 1 adjacentes aux puits et les traces synthétiques calculées prouve que la schématisation des logs Sonic adoptée est acceptable (Figures 115 et 116). La correspondance est excellente pour T1 ; pour M1, le pic négatif n'apparaît que comme une inflexion sur le modèle.

Le modèle impédance (Figure 114) souligne :

- la stabilité latérale de l'ensemble R0 + R1,
- les évolutions latérales des ensembles inter R1/R2, R2 et R3.

Si l'on compare les caractéristiques des traces modèles et des traces pilotes réelles, on retrouve, dans les deux cas, pour le spectre (Figures 115, 116) un déplacement vers les basses

MODÈLE 1

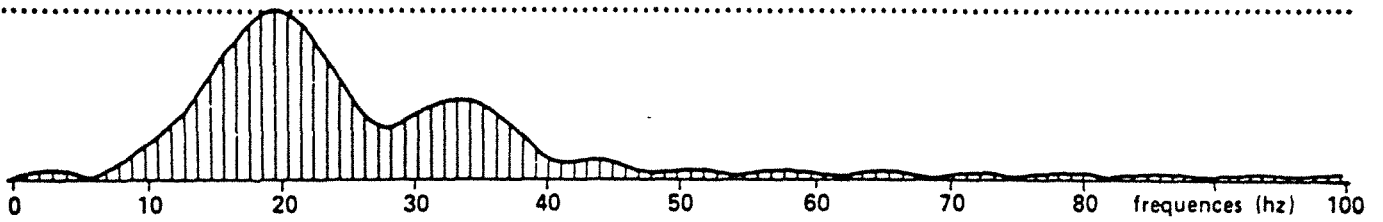
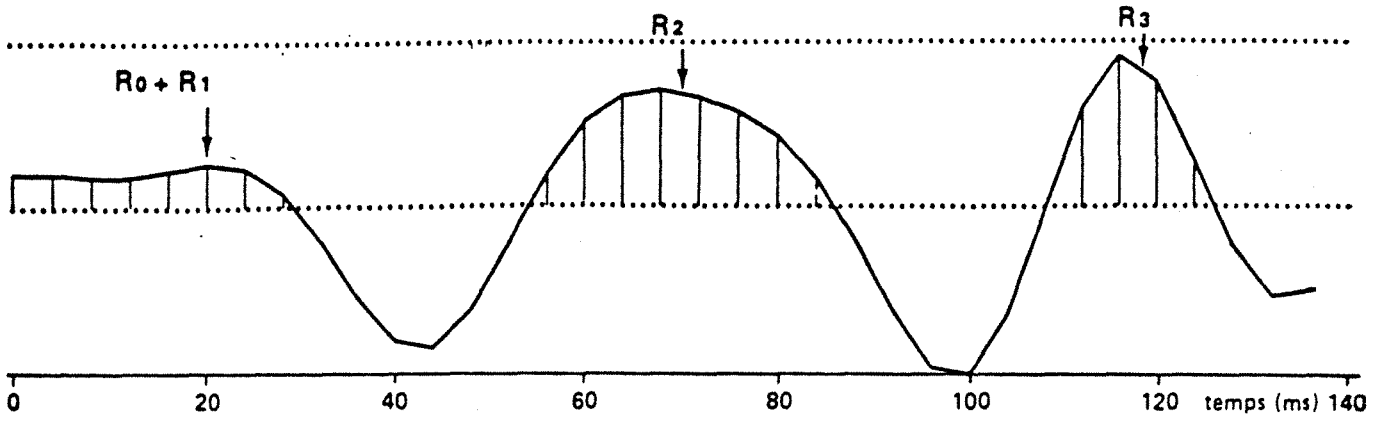


SECTION IMPÉDANCE

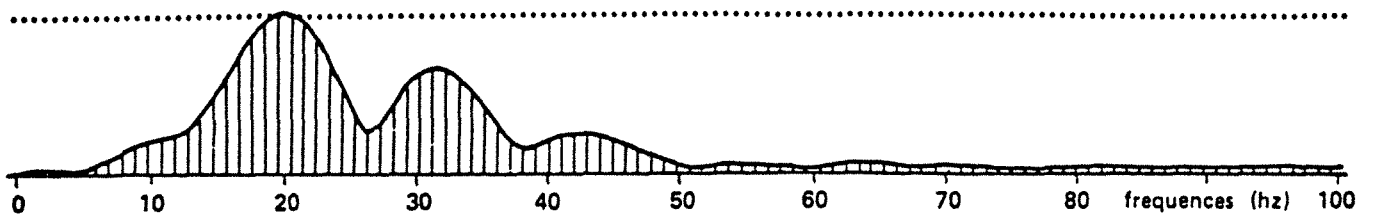
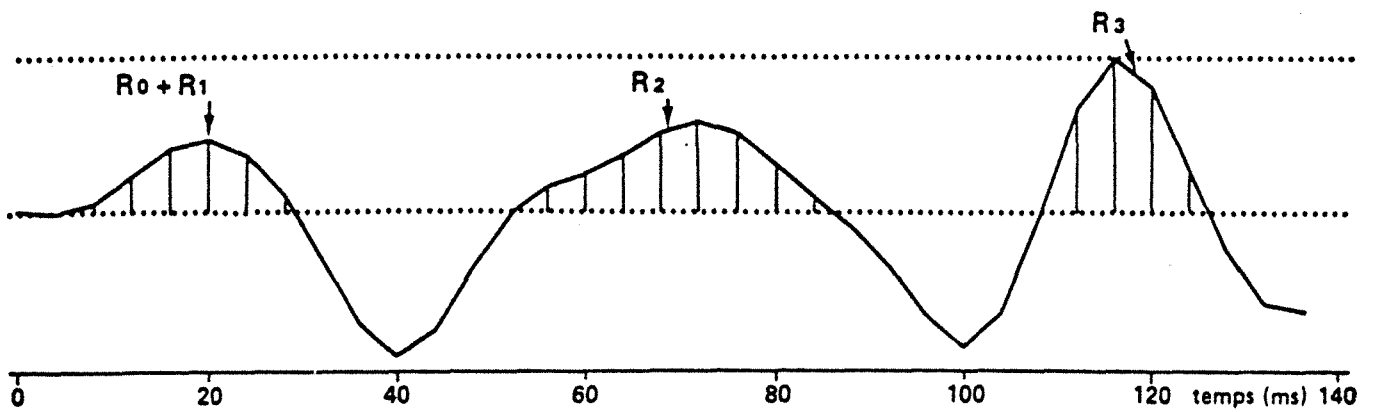
SECTION SYNTHÉTIQUE MODÈLE 1

COMPARAISON
TRACE MODÈLE/TRACE SYNTHÉTIQUE

PUITS T1



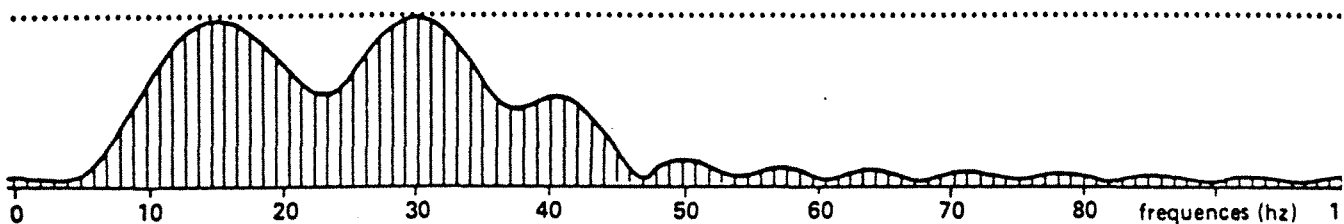
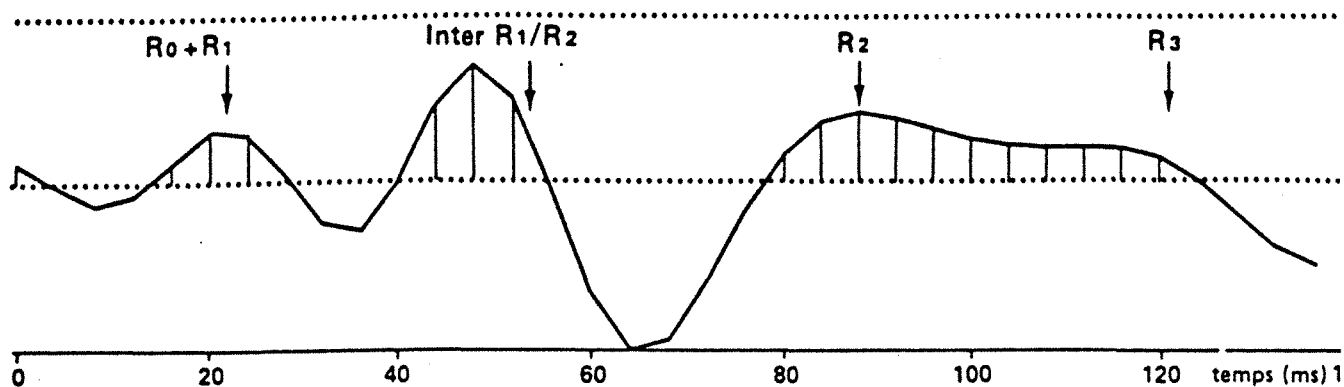
TRACE DU MODÈLE 1 ADJACENTE
AU PUIS



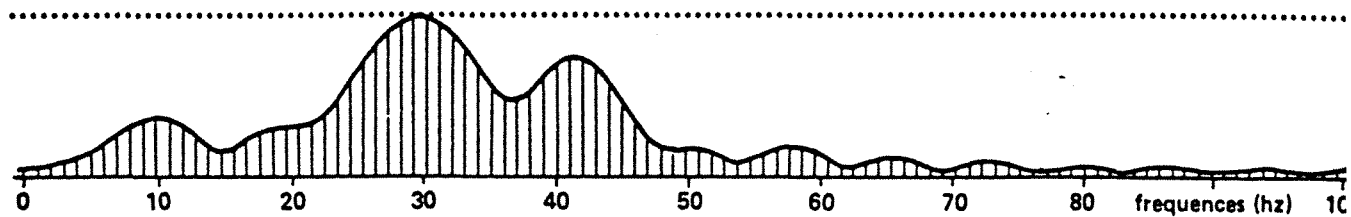
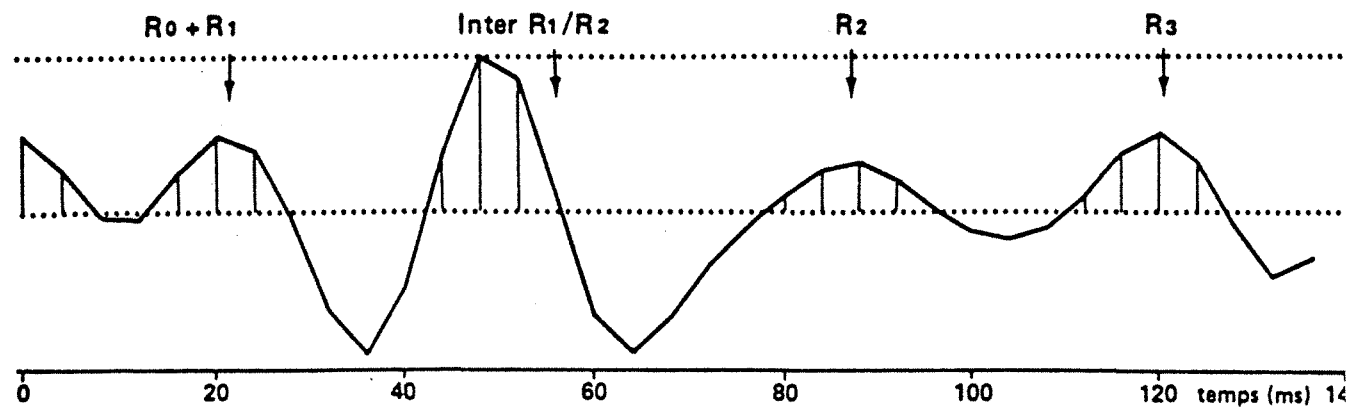
TRACE SYNTHÉTIQUE

COMPARAISON TRACE MODÈLE/TRACE SYNTHÉTIQUE

PUITS M1



TRACE DU MODÈLE 1 ADJACENTE AU PUIS



TRACE SYNTHÉTIQUE

fréquences au pôle T, et pour l'autocorrélation (Figure 117), l'existence d'un fort rebond pour le pôle T.

Les évolutions des variables ETNT, AM1, Q4, Q8 calculées sur les traces modèles sont analogues à celles observées sur la ligne 152 (Figures 118 à 121).

Maintenant que nous avons retrouvé une réponse sismique synthétique très proche de la réponse sismique réelle, il faut déterminer quel est l'événement géologique responsable du caractère sismique observé. L'ensemble R0 + R1 est assez homogène par rapport aux variations dans les unités inter R1/R2, R2 et R3.

5.3.4.2. *Influence respective des ensembles inter R1/R2, R2 et R3 sur la réponse sismique du modèle*

Nous allons analyser séparément l'influence des réservoirs inter R1/R2 + R2 et R3.

a. Modèle 2

Les variations de vitesses dans R3 au puits T1 sont remplacées par une évolution régulière jusqu'au toit de la couche rapide, comme au puits M1 (Figure 122). Dans la bande passante étudiée, la section sismique synthétique (Figure 122) est inchangée.

L'évolution des variables AM1, ETNT, Q4, Q8 calculées sur les nouvelles traces modèles est identique à celle du modèle 1.

b. Modèle 3

Les unités R0 et R1 sont maintenues constantes, la variation latérale au niveau R3 est celle du modèle initial n°1. Au puits T1, par contre, on simule une évolution de vitesse analogue à celle de M1 pour les réservoirs R2 et inter R1/R2 (Figure 123).

La section traces synthétiques est modifiée (Figure 123).

Les traces modèles adjacentes aux puits M1 et T1 se ressemblent beaucoup jusqu'à 80 ms (Figure 124), ce qui était prévisible puisque R0, R1, inter R1/R2, R2 n'évoluent plus entre les deux puits dans ce modèle. La variation au niveau R3 se traduit par la disparition vers M1 de l'arche individualisée à T1, qui se confond pour donner une réflexion beaucoup plus basse fréquence.

Les variables ne montrent pratiquement plus d'évolution entre les deux puits (Figures 125 à 128).

AUTOCORRÉLATION DES TRACES MODÈLES

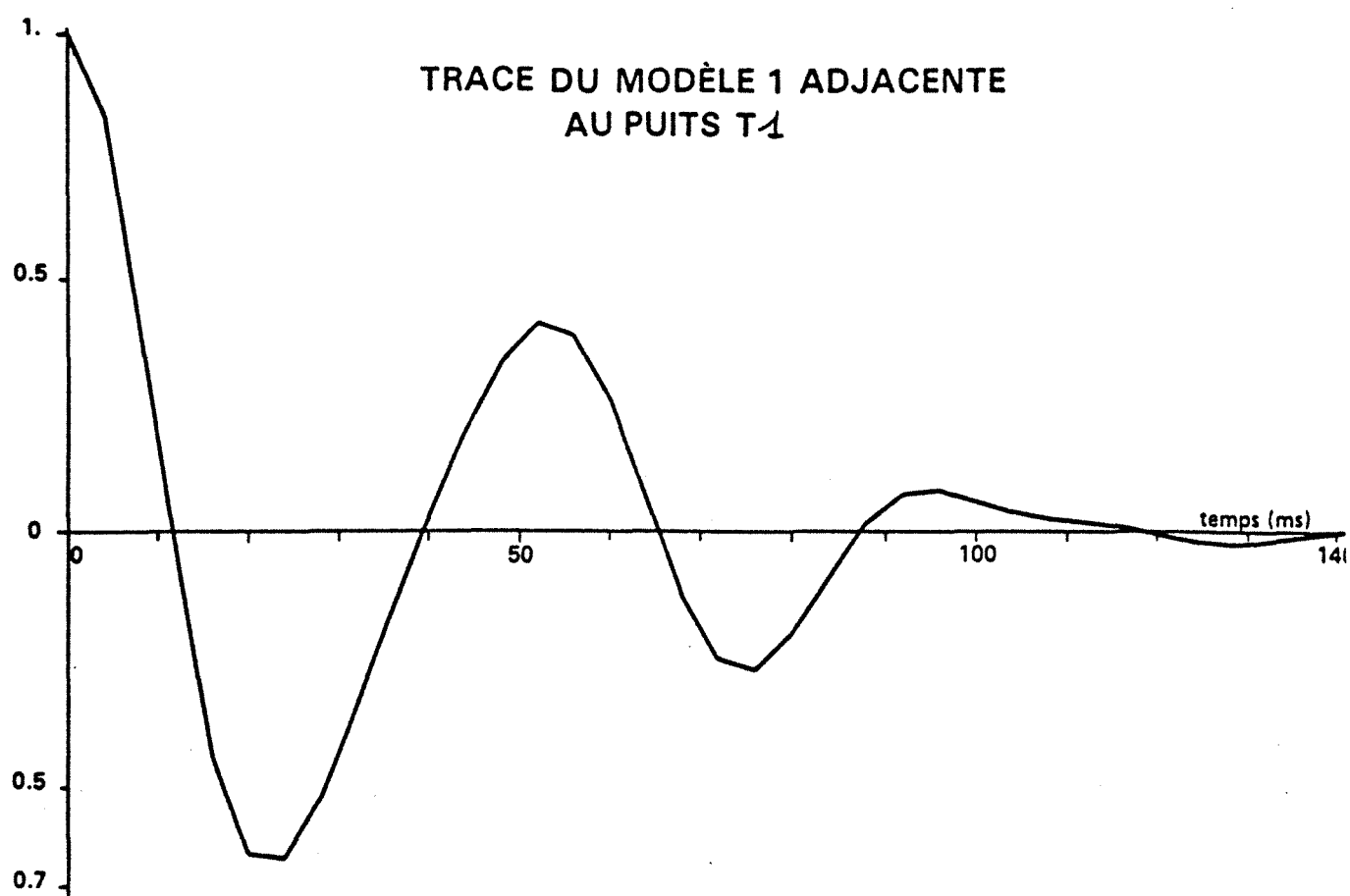
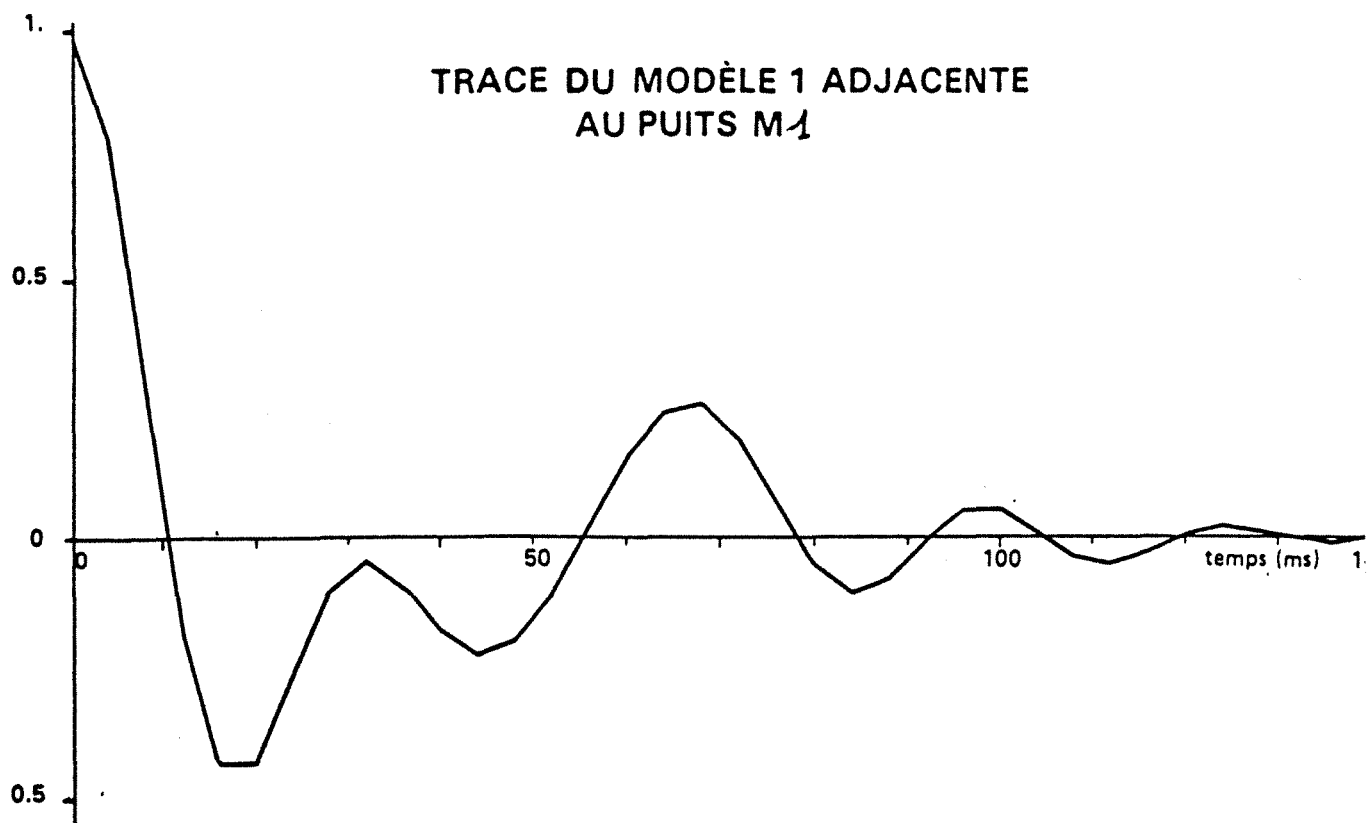
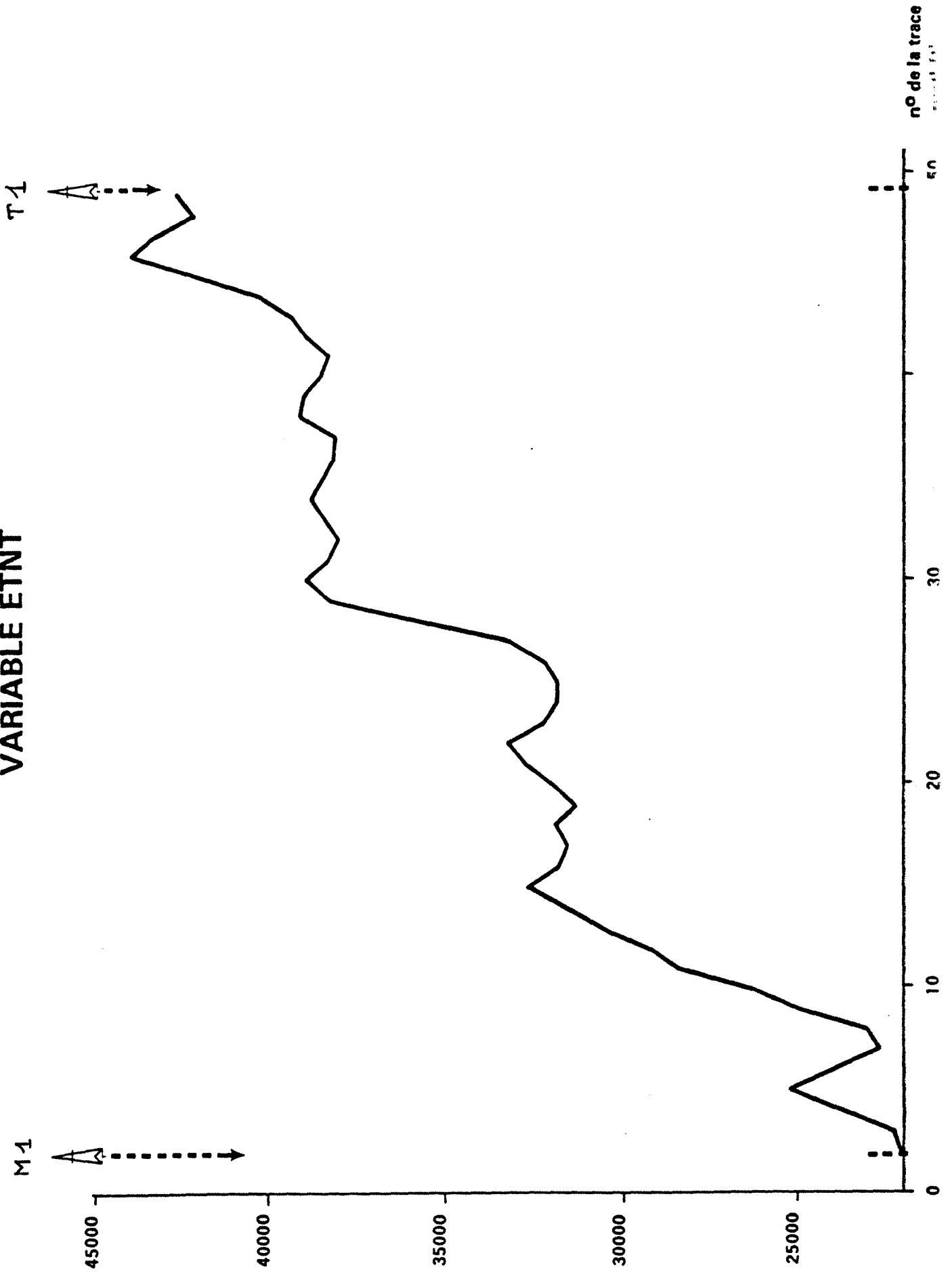


Fig.118

MODÈLE 1 VARIABLE ETNT



MODÈLE 1
VARIABLE AM 1

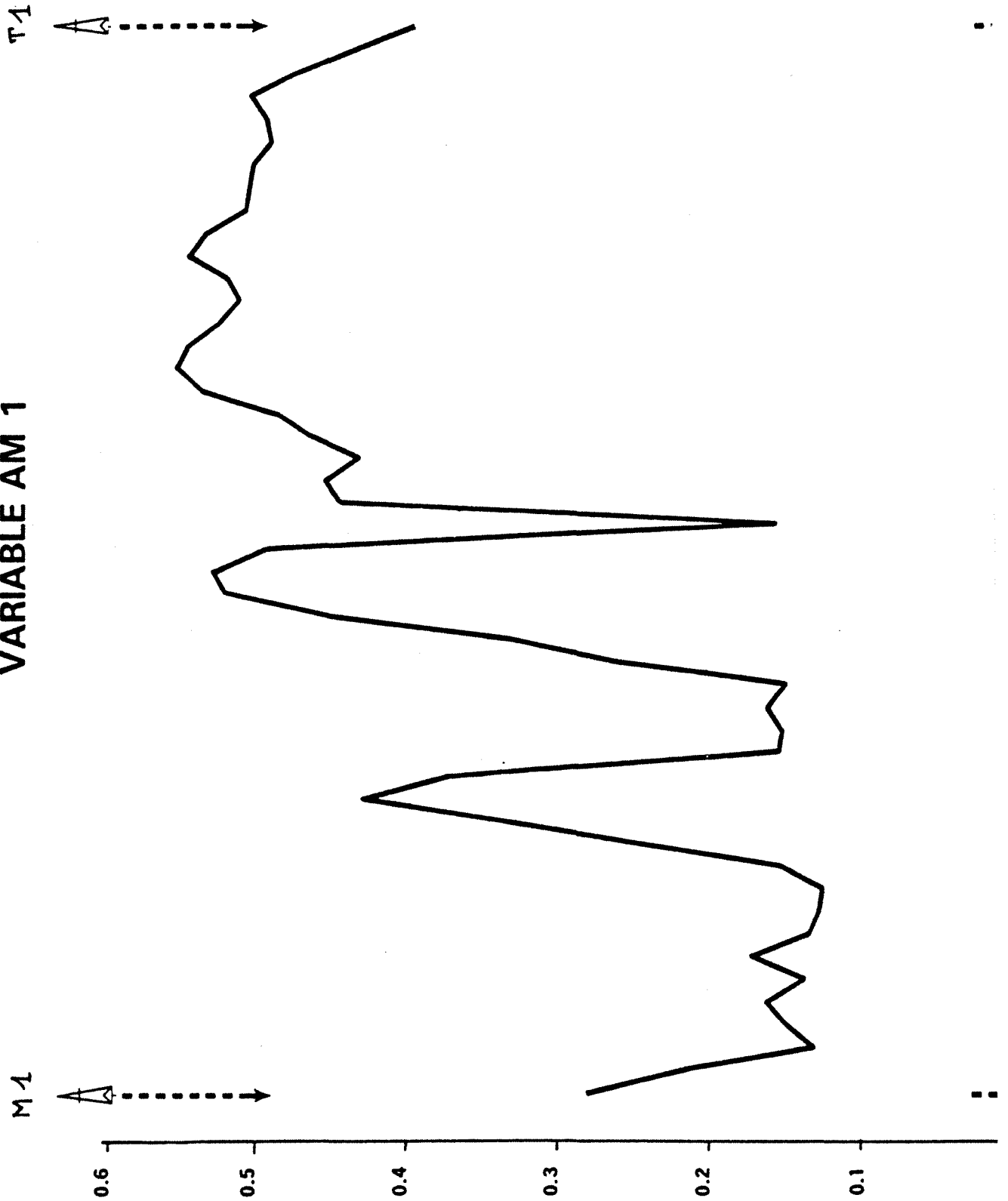
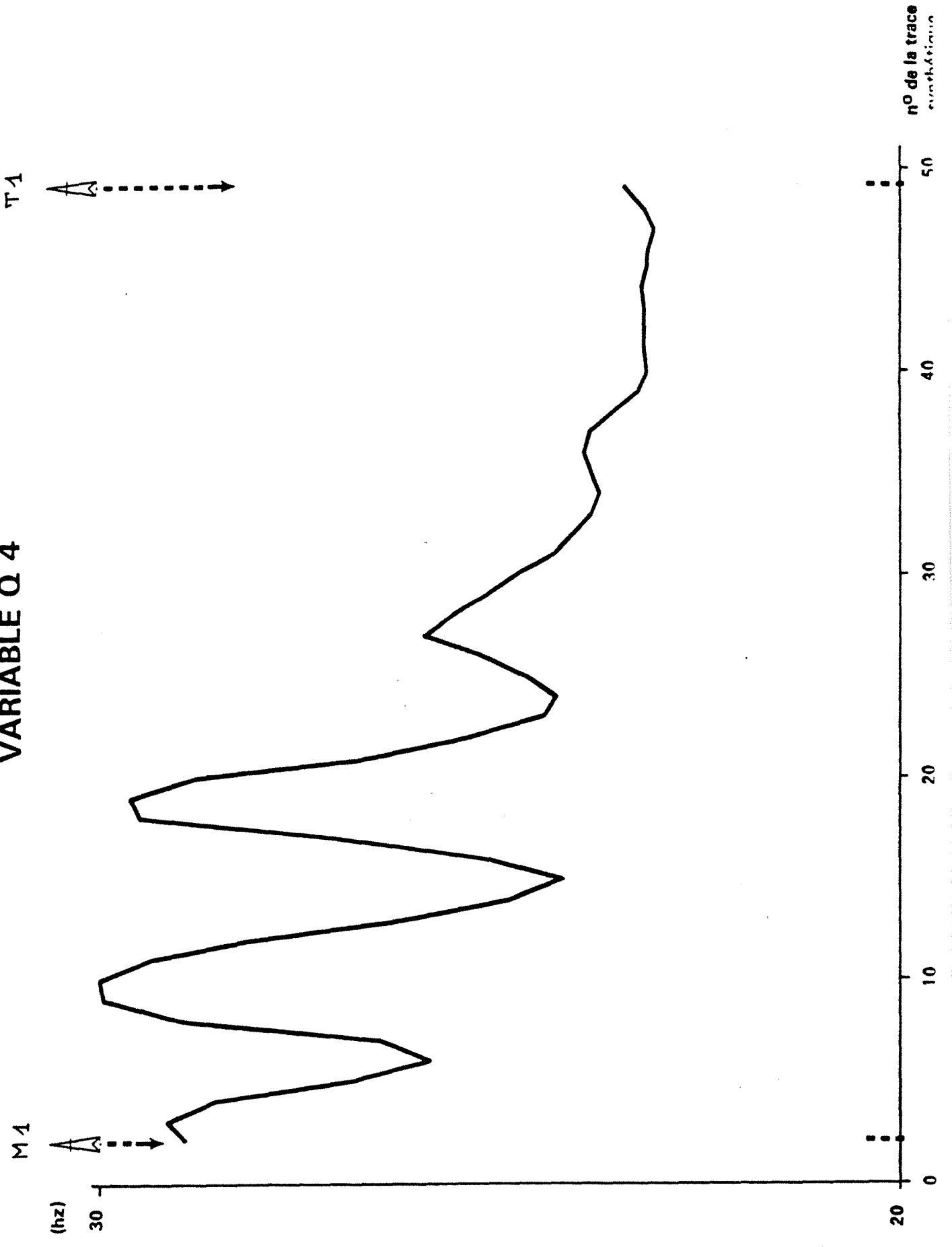
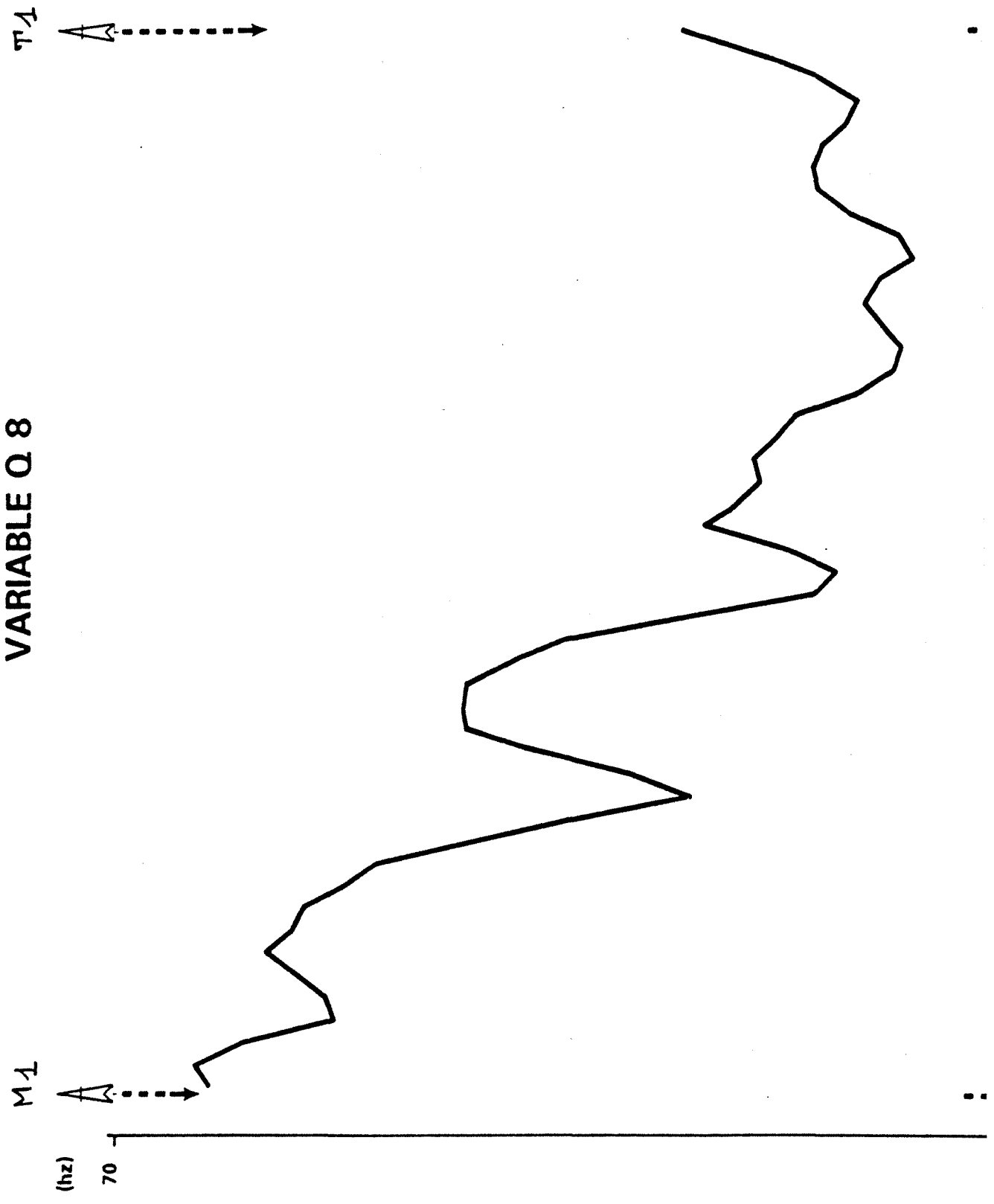


Fig.120

MODÈLE 1 VARIABLE Q 4



**MODÈLE 1
VARIABLE Q 8**



MODÈLE 2

INFLUENCE DE L'UNITÉ « R3 »

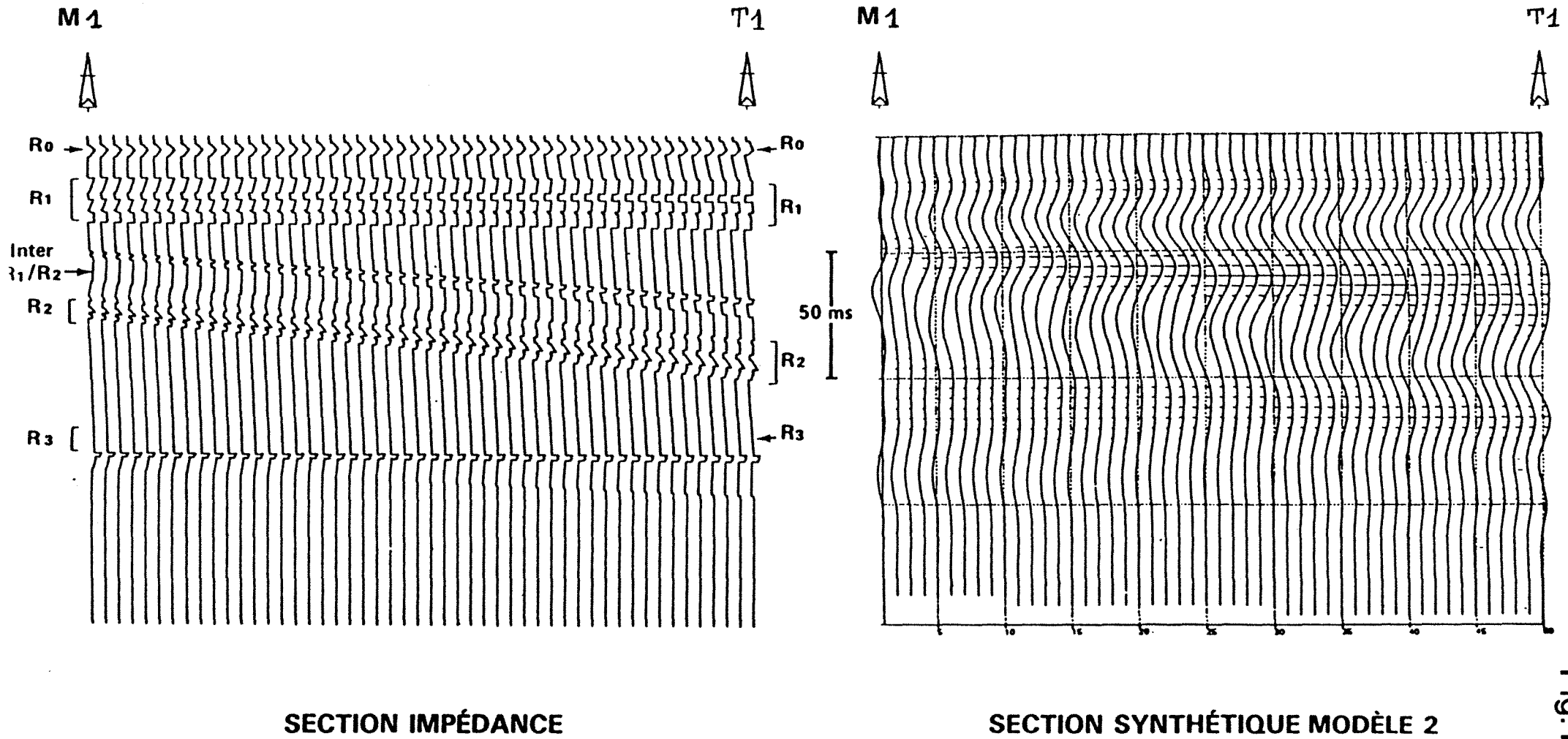
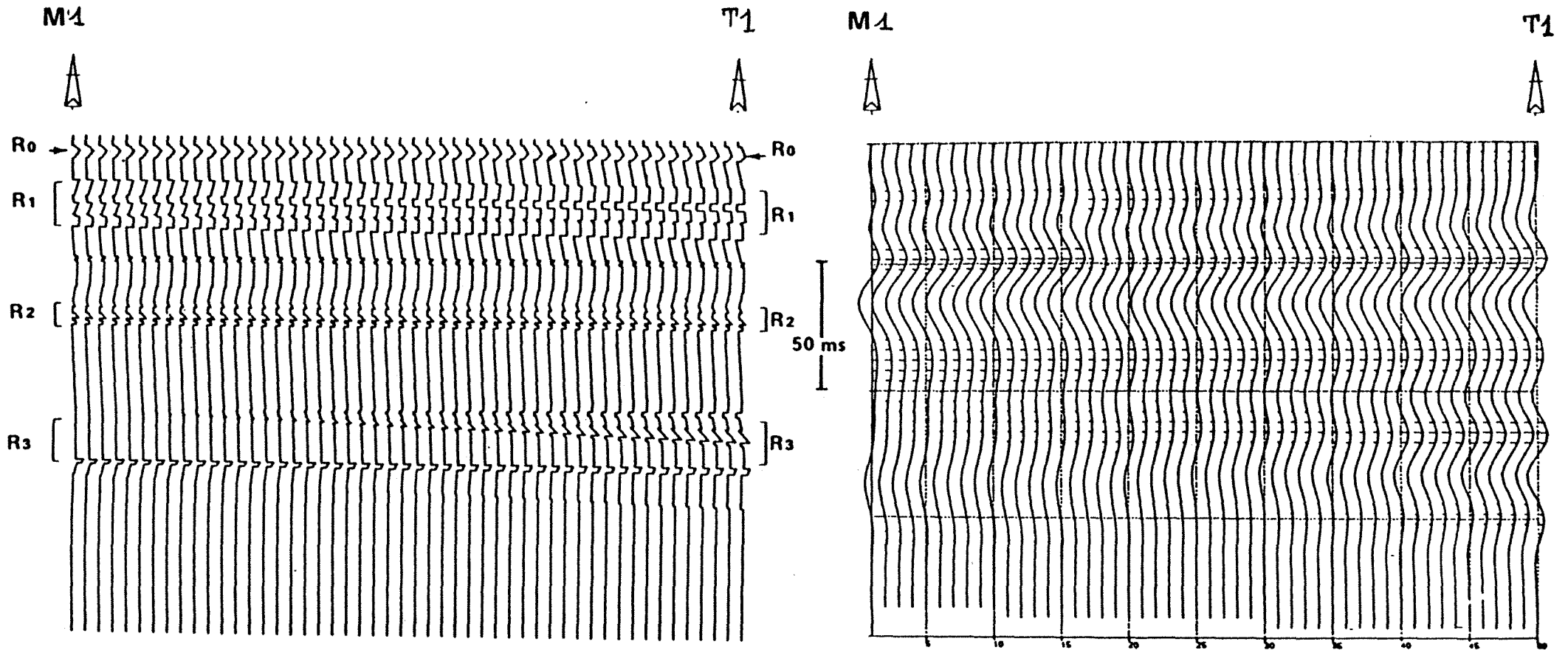


Fig. 122

MODÈLE 3

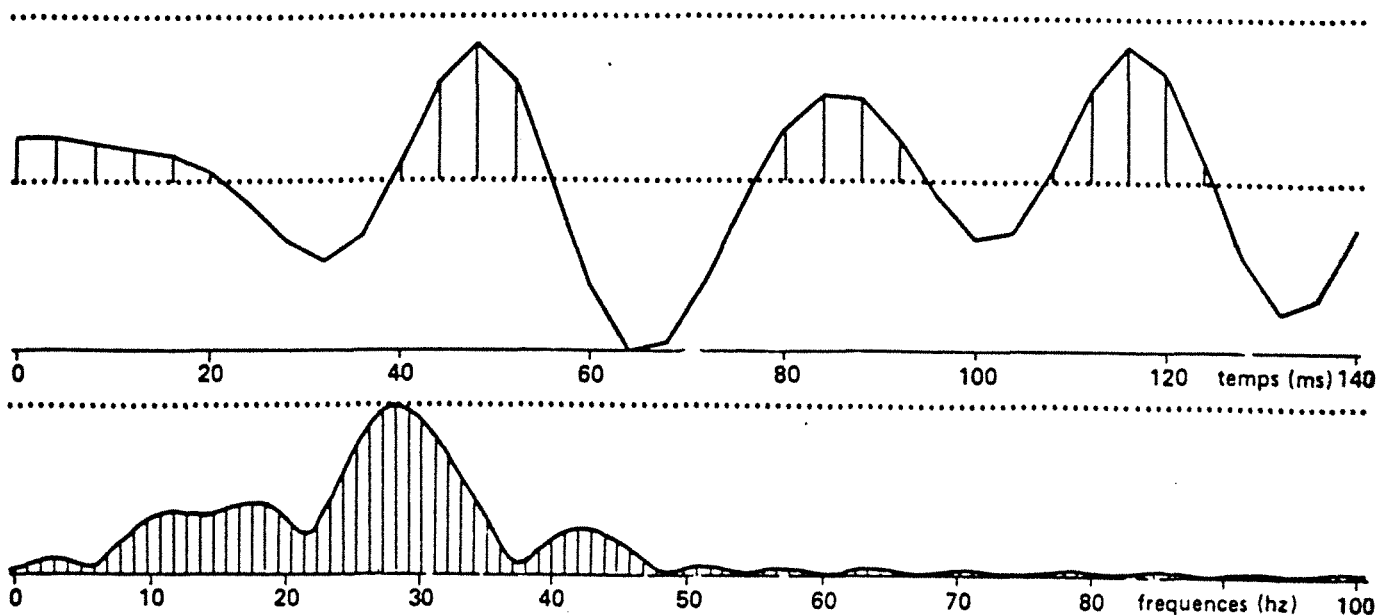
INFLUENCE DES UNITÉS INTER R1/R2 - R2



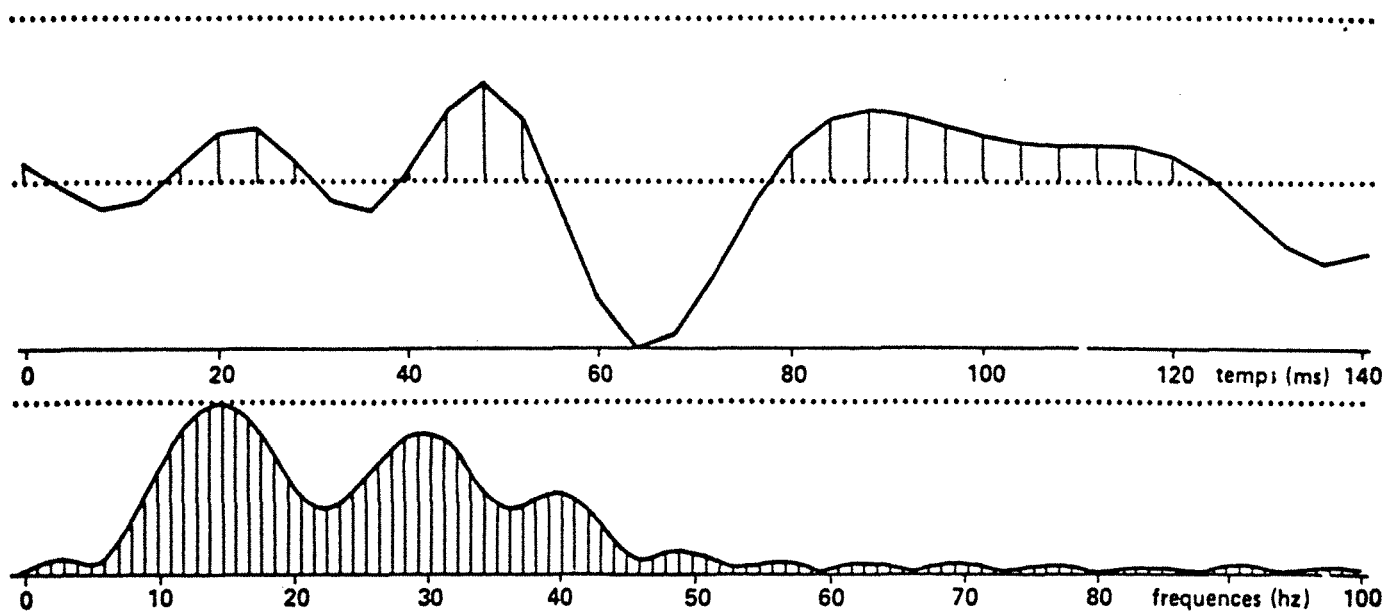
SECTION IMPÉDANCE

SECTION SYNTHÉTIQUE MODÈLE 3

MODÈLE N° 3



TRACE DU MODÈLE 3 ADJACENTE AU PUIS T₁



TRACE DU MODÈLE 3 ADJACENTE AU PUIS M₁

Fig.125

MODÈLE N° 3
VARIABLE ETNT

T1



M1



MODÈLE N° 3
VARIABLE AM 1

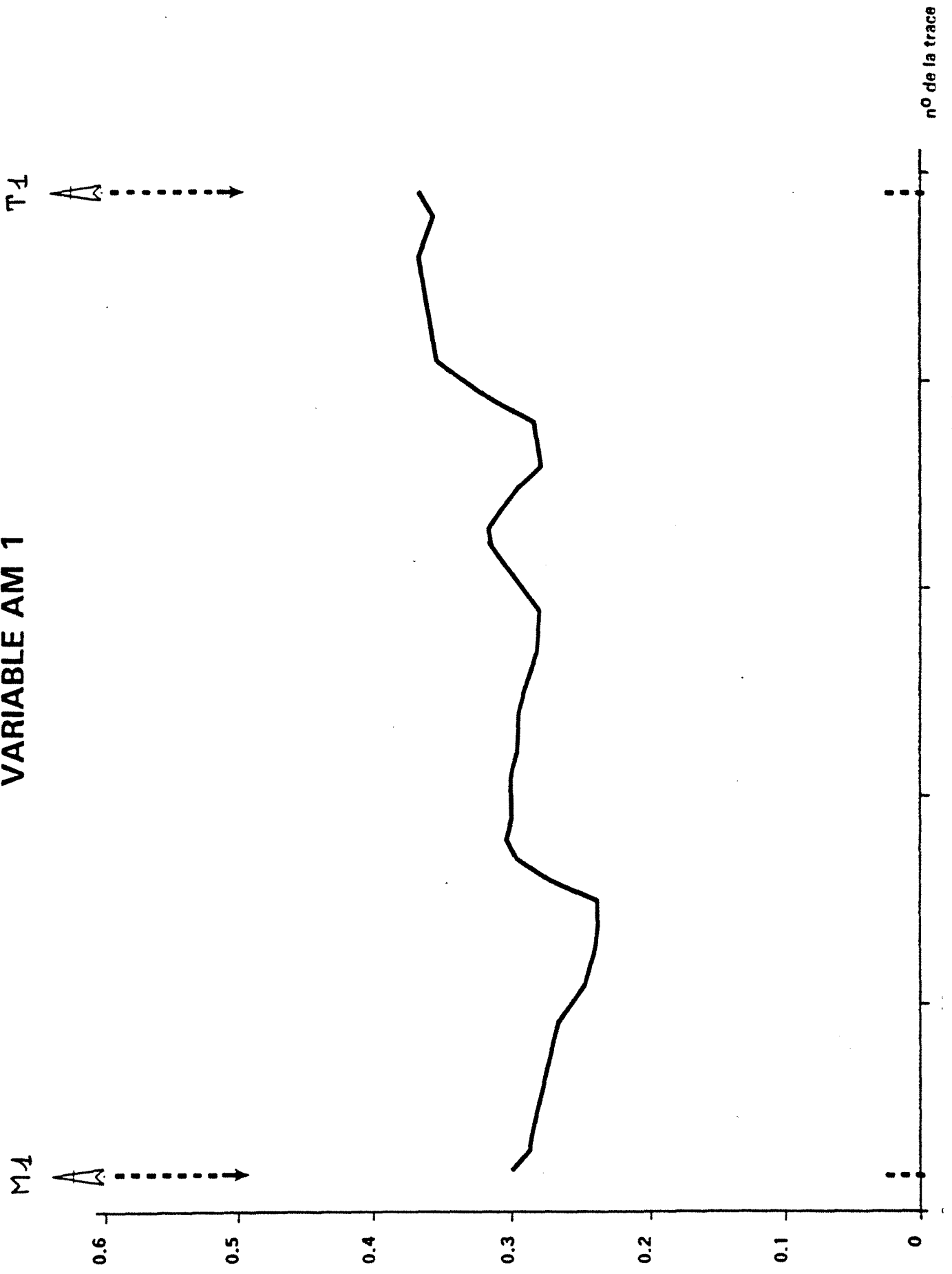


Fig.126

**MODÈLE N° 3
VARIABLE Q 4**



MODÈLE N° 4
VARIABLE Q 8

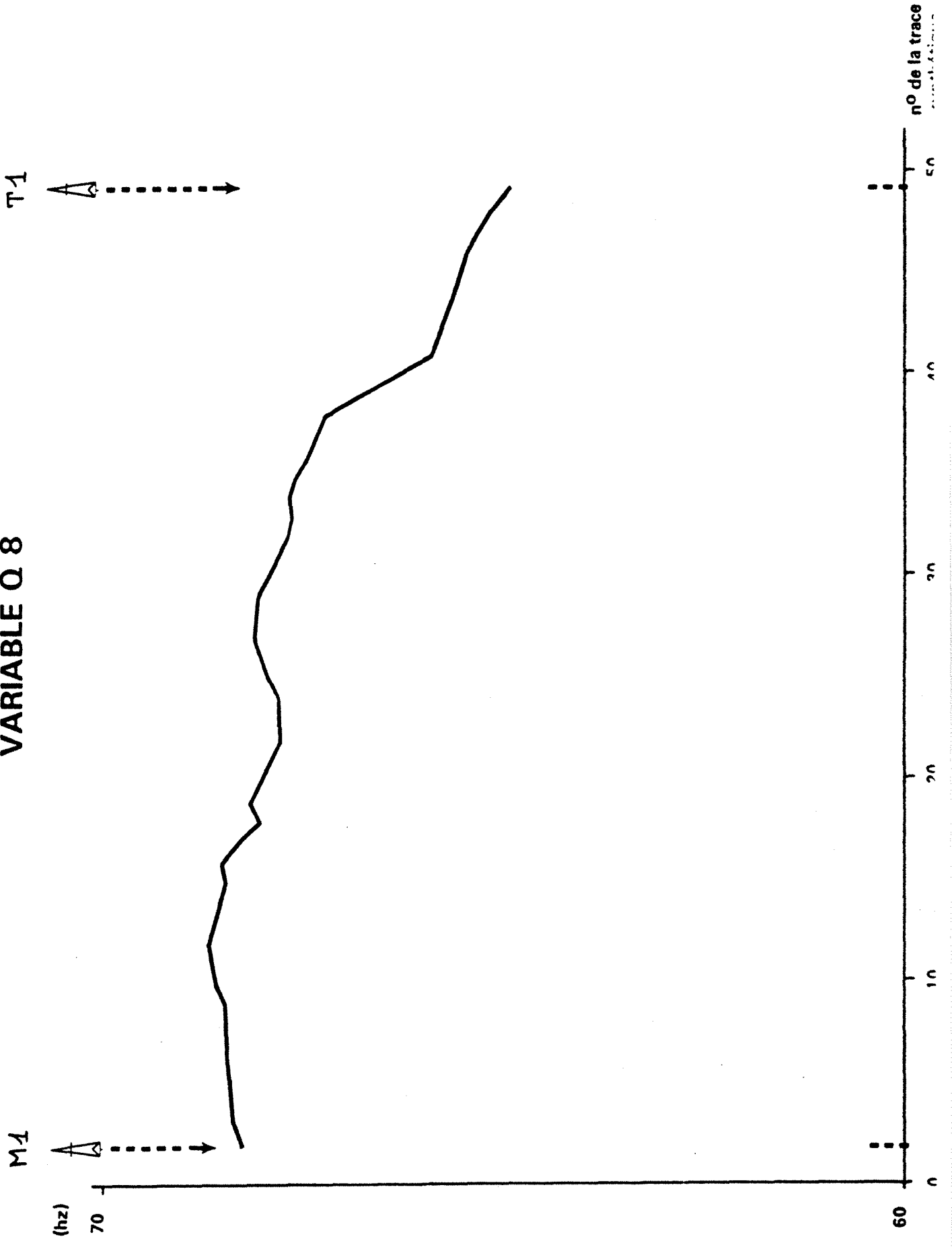


Fig.128

5.3.4.3. Conclusions de la modélisation

La modélisation nous a permis :

- de retrouver l'évolution des principales variables discriminantes entre les puits T1 et M1,
- de confirmer la liaison entre les caractéristiques sismiques et les évolutions géologiques ; en particulier, il semblerait que c'est **l'évolution latérale des unités inter R1/R2 et R2 confondues** qui rend compte de ces caractères alors que l'évolution du R3 marque très peu la réponse sismique.

5.3.5. Etape prédictive

Les deux étapes précédentes nous ont permis :

- de choisir des traces pilotes dans des zones homogènes mais de faciès sismique bien individualisé,
- de sélectionner les paramètres réellement discriminants par rapport au problème posé,
- de contrôler par modélisation que la différence de faciès sismique est liée aux variations géologiques principalement des ensembles inter R1/R2 et R2.

Nous allons maintenant passer à l'étape décisionnelle, ou prédictive, de la méthodologie employée.

Nous allons calculer sur les portions de traces inconnues, au niveau des réservoirs (1930 à 2070 ms), les variables discriminantes précédemment sélectionnées (AM1, ETNT, Q4, Q8) et classer ces traces par rapport aux traces pilotes. L'objectif est de réaliser une prédiction des limites d'extension des deux faciès analysés, dans la zone couverte par les lignes sismiques étudiées.

Nous avons auparavant réalisé des tests sur la ligne 152 afin de nous assurer de la stabilité des résultats des analyses factorielles.

5.3.5.1. Tests sur la ligne 152

Deux méthodes multidimensionnelles ont été utilisées : analyse factorielle discriminante pas à pas et classification.

Les résultats d'affectation des traces anonymes de la section 152 sont peu changés que l'on utilise les variables ETNT et Q8, ou ETNT seul. Ceci confirme la remarque du paragraphe 5.3.3.2.c sur le rôle prépondérant de ETNT dans le couple ETNT, Q8.

Si l'on retire la variable ETNT, le couple le plus discriminant est AM1, Q4 (avec rôle prédominant de AM1). Les affectations des traces anonymes sont très concordantes avec les essais précédents.

La prédiction du faciès semble identique que l'on fasse intervenir l'une ou l'autre des deux variables les plus discriminantes, ETNT ou AM1.

L'analyse factorielle discriminante part, a priori, de l'indication des groupements de traces que l'on cherche à réaliser. Il y a naturellement des tests de contrôle de la qualité de la discrimination réalisée (pourcentage de bien classés sur l'échantillon test).

Nous avons donc également réalisé des classifications : aucune hypothèse n'est émise sur les traces au départ et elles sont regroupées en fonction de leur proximité dans l'espace des variables.

Nous avons réalisé une classification à 2 groupes sur la ligne 152 avec, soit le couple de variables ETNT, Q8, soit le couple AM1, Q4. Puis nous avons fait un essai avec AM1, ETNT, Q4, Q8, à 2 ou à 5 groupes.

Nous avons remarqué que pour toutes les classifications à 2 groupes, on retrouve les zones de faciès prédites par l'analyse factorielle discriminante. On vérifie, en particulier, l'homogénéité des zones où les traces pilotes ont été sélectionnées.

Dans le cas de 5 groupes, les zones classées en faciès T restent homogènes alors que le faciès M éclate dans les divers autres groupes. La zone où les traces pilotes du faciès M ont été définies reste toutefois à peu près homogène.

5.3.5.2. Cartes d'extension des faciès étudiés

Ces différentes cartes ont été réalisées avec le souci de confirmer les résultats par plusieurs méthodes. Quatre cartes ont été réalisées à partir de l'affectation des traces anonymes des sections étudiées (lignes 108, 112, 120, 148 à 152, 164) :

- deux cartes avec les résultats de l'analyse factorielle discriminante (couples ETNT, Q8 ou AM1, Q4, Figure 129),
- deux cartes avec les résultats de la classification (variables AM1, ETNT, Q4, Q8) à 2 ou 5 groupes.

Le trait commun à ces quatre cartes (Figure 129) est la distribution du faciès T dans des formes chenalisantes, orientées Nord-Est - Sud-Ouest à l'Est de l'ensellement, et plutôt Est-Ouest à l'Ouest de cet ensellement.

Pour la classification à 5 groupes, on retrouve ces deux directions de part et d'autre de l'ensellement, mais le faciès T n'est bien développé qu'à l'Est ; le faciès M est très hétérogène. Par contre, à l'Ouest de l'ensellement, on observe le développement d'un autre faciès, classé avec T dans le cas de 2 groupes.

Les directions trouvées sont assez cohérentes avec le schéma géologique régional proposé. Sur la structure voisine, les apports turbiditiques se feraient du Sud-Est vers le Nord-Ouest mais la structure étudiée semble située à l'extrémité de cette zone d'apports et les directions de dépôt sont donc plus erratiques. La zone de dépôt serait donc plutôt, pour l'Anguille Supérieur, Est-Ouest ou même Nord-Est-Sud-Ouest, passant à Est-Ouest sur le flanc Nord de la structure voisine.

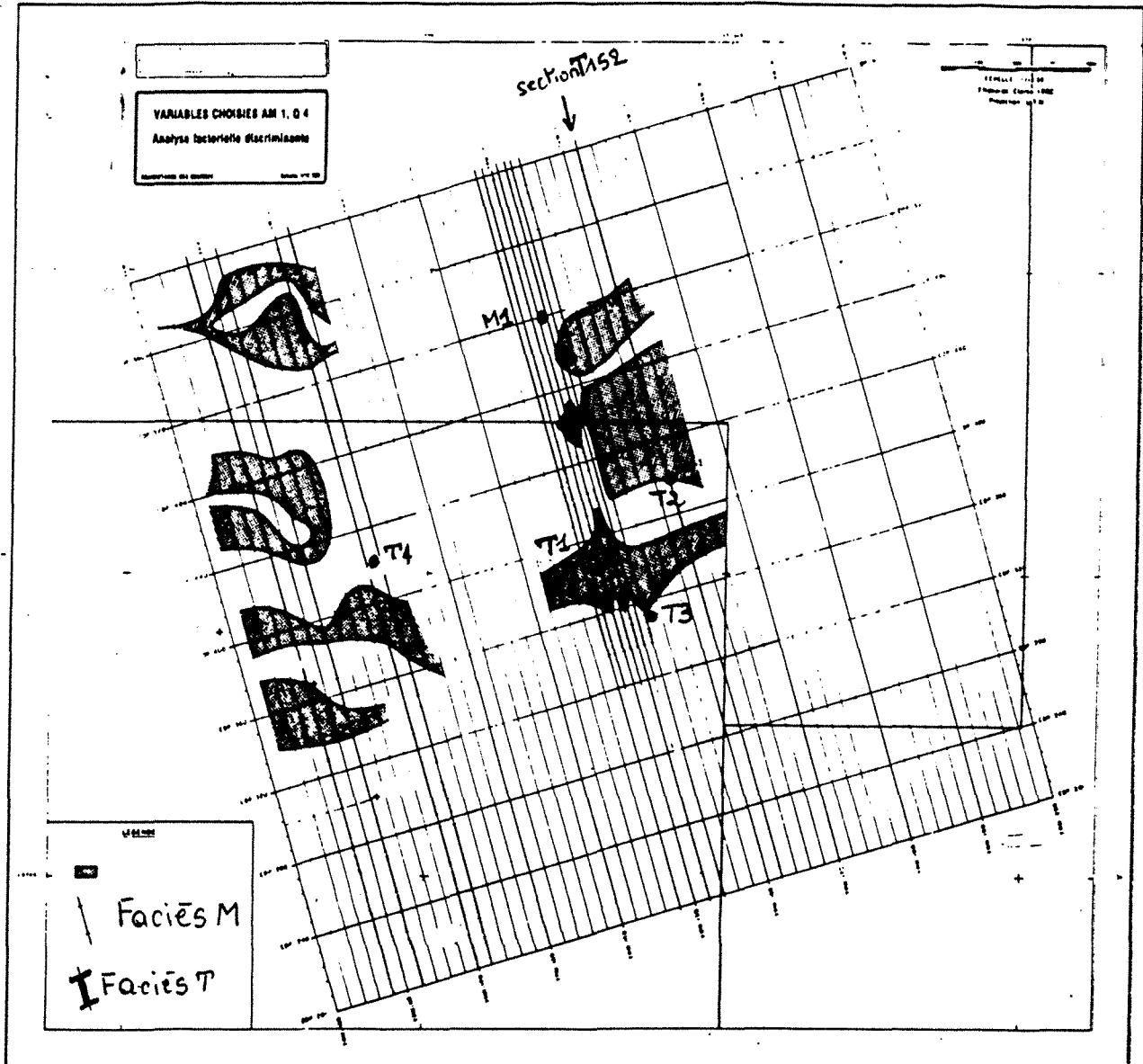
5.3.5.3. Contrôle avec les puits

Il est intéressant de contrôler les résultats prédits par la méthode avec ce que l'on sait des caractéristiques géologiques des réservoirs aux puits. Les puits T1 et T3 sont très proches, au point de vue inter R1/R2. Par contre, le puits T2 se rapproche beaucoup du puits M1 (développement de grès au niveau inter R1/R2, détérioration de la qualité réservoir du R2). Si l'on regarde les cartes de faciès à deux groupes, le puits T2 est classé en faciès T et le puits T3 est affecté à la classe M.

Pour ce qui est du puits T3, on peut penser que la qualité des données sismiques est affectée par la proximité du dôme de sel. Précédemment, nous avons remarqué que pour la classification à 5 groupes, le pôle T reste à peu près homogène alors que le pôle M a tendance à éclater. Ceci laisse à penser que nous classons un faciès T **caractéristique** par opposition à un faciès peu individualisé : en particulier, les zones bruitées (cas de T3) vont être affectées au pôle M.

Pour ce qui est de l'affectation de T2 au pôle T, alors que ses caractéristiques sont plus proches de celles de M, il faut reprendre l'image sismique au voisinage des puits T1 (lignes 148, 154) et T2 (ligne 162). Les caractères sismiques de ces zones sont bien analogues : la méthodologie utilisée "voit" donc correctement le faciès sismique (Figure 130), mais, dans le cas étudié, ce

Fig. 129



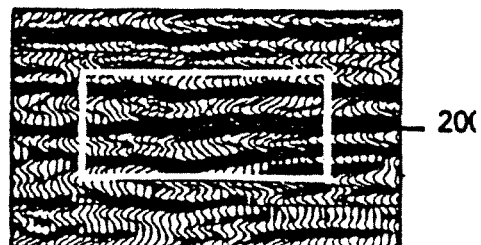
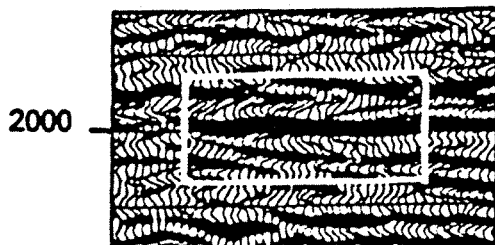
COMPARAISON DES DEUX TYPES DE FACIÈS DIFFERENCIÉS PAR L'ETUDE

FACIÈS M

FACIÈS T

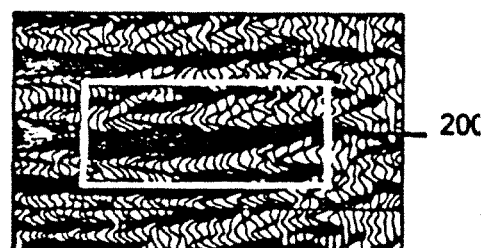
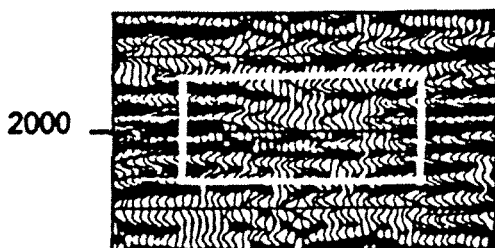
112

112



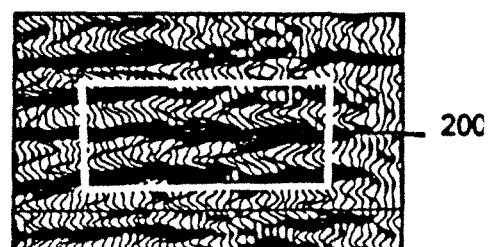
122

148



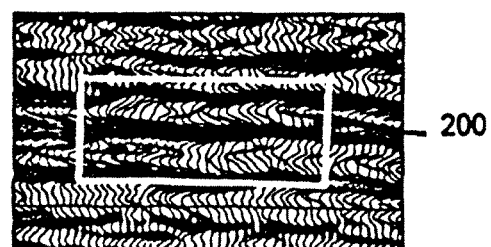
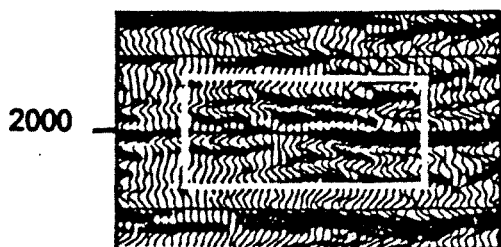
154

154



162

162



faciès sismique ne se relie pas simplement aux caractéristiques des réservoirs. Le puits T2 est proche du puits M1 mais le faciès sismique des traces adjacentes est identique à celui de T1.

5.3.6. Conclusion

Les données sismiques étudiées sont de bonne qualité et les lignes traitées très proches l'une de l'autre (information dense). D'autre part, les calages avec les données de puits sont très satisfaisants et nous disposons d'une interprétation sismique assez récente. Les conditions d'application de la méthode sont donc assez favorables. Toutefois, il faut regretter le peu de puits sur la structure ; un plus grand nombre aurait permis un meilleur contrôle du type de faciès prédit.

Les directions trouvées pour la distribution du faciès T semblent concorder avec les schémas géologiques proposés ; il est également satisfaisant de voir se développer un troisième type de faciès à l'Ouest de l'ensellement vers la structure voisine.

Les modèles réalisés à partir des données de puits semblent relier l'évolution du caractère sismique observée sur les lignes de calage (152 en particulier) avec l'évolution latérale reconnue des réservoirs inter R1/R2 et R2. En fait, l'analyse des cartes de faciès indiquerait plutôt que la méthodologie permet de déterminer des zones de faciès sismiques analogues (Figure 130) et ceci, avec une bonne fiabilité (cf. cohérence des classifications avec les analyses factorielles). Par contre, le faciès sismique appelé T ne serait pas seulement caractéristique des faciès réservoirs de T1 (on le retrouve à T2).

5.4. Conclusion aux deux études de cas présentées

Le premier cas illustre l'adaptation de la méthode pour la reconnaissance automatique du faciès : les traces sismiques regroupées ont une morphologie identique. L'analyse du faciès recoupe celle que pourrait faire un expert en interprétation sismique mais elle est plus rapide et plus objective ; de plus, elle est systématique et quantitative.

Le deuxième cas illustre le problème de la liaison entre les faciès sismiques reconnus et les propriétés géologiques. L'étape de modélisation fournit des indications supplémentaires quant à ce lien mais aucune preuve définitive. En particulier, quand le nombre de puits est très faible, il est encore plus difficile de contrôler que l'affectation des traces dans les groupes d'apprentissage reflète les propriétés géologiques (si peu de puits existent, il est difficile de remettre en question le choix des traces d'apprentissage).

TROISIEME PARTIE

**ADAPTATION DE LA METHODOLOGIE DE RECONNAISSANCE
DES FACIES SISMIQUES POUR UNE COURTE FENETRE
TEMPORELLE**

1. INTRODUCTION

Les paramètres sismiques calculés sur le spectre d'amplitude de l'autocorrélation reposent sur une estimation spectrale par la transformée de Fourier rapide. Or, cette estimation est réputée très peu fiable lorsque le nombre d'échantillons traités diminue (taille de la fenêtre d'analyse petite). On peut même s'interroger sur la signification physique de tous les traitements utilisés pour caractériser la trace, sur des fenêtres temporelles d'environ 20 ms (quelques échantillons).

Cependant, avec l'orientation de la géophysique vers des problèmes de réservoir, il devient encore plus nécessaire d'analyser le faciès sismique sur des objectifs géologiques fins et d'essayer de le caractériser en liaison avec ses propriétés géologiques.

Nous avons essayé plusieurs voies de remplacement des paramètres proposés pour la caractérisation des traces sismiques. La première voie a consisté à tester une représentation temps-fréquence particulière (celle de Wigner-Ville, Bouachache, 1982) qui semblait très bien adaptée pour une caractérisation **locale** des traces sismiques. Les résultats que nous avons obtenus étant assez décevants, aussi bien du point de vue **localisation de l'information** que **sensibilité aux variations géologiques fines**, nous avons abandonné cette voie.

Nous avons ensuite étudié simultanément une méthode d'analyse spectrale réputée pour sa robustesse vis à vis de la taille de la fenêtre analysée (méthode par maximisation de l'entropie) ainsi que la possibilité de caractériser les traces sismiques **directement** par la succession des amplitudes des échantillons.

Finalement, c'est cette dernière technique qui semble la plus fiable et la plus facile à mettre en œuvre pour la caractérisation de la trace sur une courte fenêtre. Ayant trouvé un moyen de caractériser la trace sismique, et donc le faciès, sur une **courte fenêtre temporelle**, la méthodologie statistique développée dans la seconde partie peut encore s'appliquer, avec quelques modifications liées au choix des paramètres de représentation des traces.

Après avoir exposé la méthodologie statistique modifiée, nous présenterons une application de la reconnaissance automatique des faciès dans le cadre d'un problème de gisement sur courte fenêtre.

2. UTILISATION DE LA REPRESENTATION TEMPS-FREQUENCE POUR LA CARACTERISATION DES STRUCTURES FINES SISMIQUES

2.1. Généralités sur la représentation conjointe temps-fréquence (R.C.T.F.)

2.1.1. Définition

La représentation la plus habituelle d'un signal est temporelle ; on peut accéder à ses composantes spectrales par transformation de Fourier en étudiant sa représentation harmonique. Mais ces deux représentations ne permettront pas de décrire complètement le signal : par exemple, on n'accède pas ainsi à la loi de modulation en fréquence du signal. L'insuffisance des représentations temporelles et fréquentielles provient de la décomposition d'un signal physique (appartenant à l'ensemble des fonctions de carré sommable) sur les fonctions $\{\delta(t - t_0)\}_{t_0 \in \mathbb{R}}$ pour la représentation temporelle et sur $\{e^{i2\pi\nu_0 t}\}$ pour la représentation fréquentielle. Or ces fonctions n'étant pas des signaux physiques, les décompositions sont donc purement mathématiques.

Une représentation temps-fréquence est au contraire une décomposition sur une base de signaux d'énergie finie et à bande fréquentielle limitée. Elle permet d'accéder, en particulier, à la loi de modulation en fréquence du signal (Figure 131).

Supposons que le signal $S(t)$ soit d'énergie finie.

Soit $\rho(t, \nu)$ la représentation conjointe du signal.

Elle devra vérifier :

$$\int_{\mathbb{R}^2} \rho(t, \nu) dt d\nu = E_z = \langle Z, Z \rangle$$

$$\int_{\mathbb{R}} \rho(t, \nu) d\nu = |Z(t)|^2$$

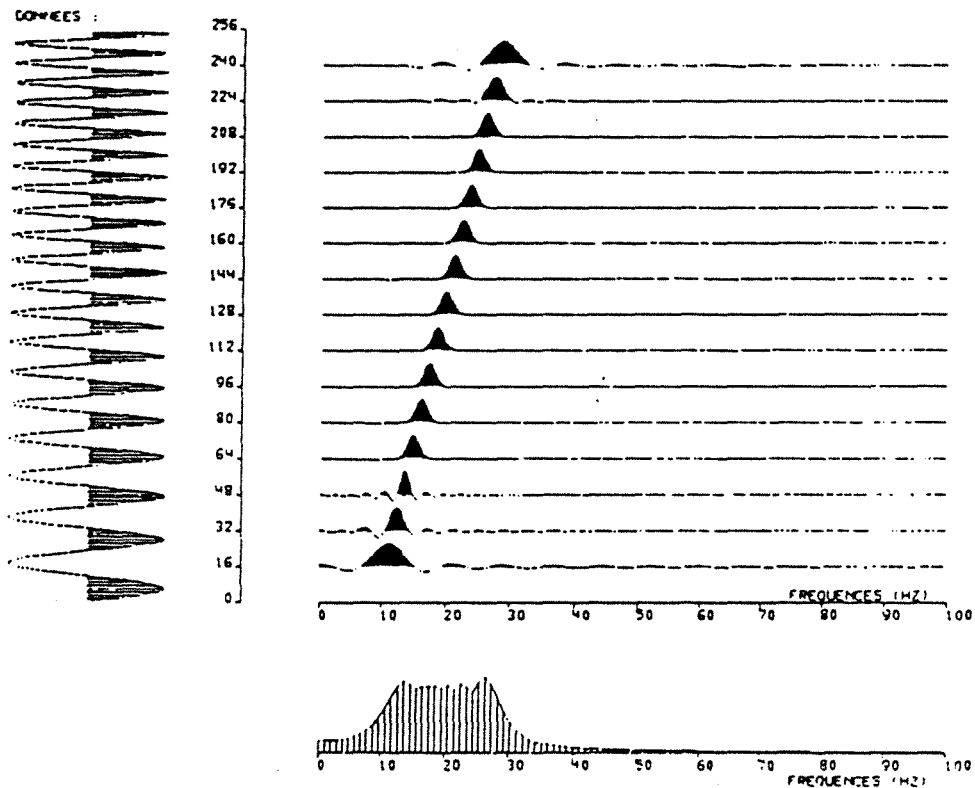
$$\int_{\mathbb{R}} \rho(t, \nu) dt = |z(\nu)|^2$$

où $Z(t)$ est le signal analytique associé à $S(t)$, et E_z l'énergie du signal Z .

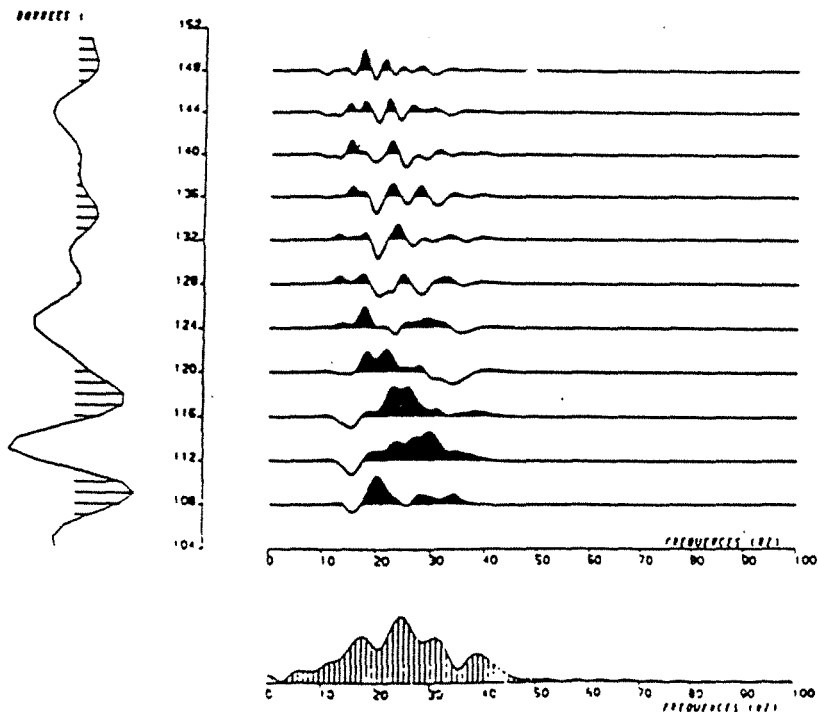
On peut ajouter la contrainte supplémentaire $\rho(t, \nu) > 0$ afin de l'interpréter comme une densité, mais cela n'est pas nécessaire ; $\rho(t, \nu) dt d\nu$ est un débit énergétique dans le plan (t, ν) .

REPRESENTATION TEMPS FREQUENCE DE WIGNER-VILLE

a/ SIGNAL MODULE LINEAIREMENT-EN FREQUENCES



b/ TRACE SISMIQUE REELLE



Ce problème a reçu de nombreuses solutions en fonction du choix effectué pour les fonctions f telles que :

$$\rho(t, \nu) = \int_{\mathbb{R}^3} f(n, \tau) e^{2i\pi n(u-t)} Z\left(u + \frac{\tau}{2}\right) Z\left(u - \frac{\tau}{2}\right) \times e^{2i\pi n t} \cdot dndud\tau$$

La représentation de Wigner-Ville (R.W.V.) correspond à $f = 1$; $\rho(t, \nu)$ est réelle mais nullement positive.

Le choix de la représentation de Wigner-Ville a été fait suite aux études sur les diverses représentations conjointes temps-fréquence faites par Bouachache (thèse de Docteur-Ingénieur, 1982).

2.1.2. Propriétés de la R.W.V.

* $\rho(t, \nu)$ est réel pour tous (t, ν) .

$$\int_{-\infty}^{+\infty} \rho(t, \nu) \cdot dt = |z(\nu)|^2, \quad \text{densité spectrale,}$$

$$\int_{-\infty}^{+\infty} \rho(t, \nu) \cdot d\nu = |Z(t)|^2, \quad \text{puissance instantanée,}$$

$$\int_{-\infty}^{+\infty} \rho(t, \nu) \cdot dt \cdot d\nu = E_Z, \quad \text{énergie du signal analytique.}$$

La R.W.V. répartit l'évolution énergétique du signal dans le plan temps-fréquence.

$$\int_{-\infty}^{+\infty} \rho(t, \nu) e^{2j\pi\nu t} d\nu = Z(t) \cdot z^*(0) : \quad \text{propriété de réversibilité.}$$

Définition de symétrie :
$$\rho(t, \nu) = \int_{-\infty}^{+\infty} z\left(\nu + \frac{\varphi}{2}\right) z^*\left(\nu - \frac{\varphi}{2}\right) e^{i2\pi\varphi t} d\varphi$$

$$z(\nu) \rightleftharpoons Z(t)$$

Pour BT (produit de la bande passante par la durée d'observation) grand devant l'unité, le lieu de concentration de $\rho(t, \nu)$ fournit la fréquence instantanée du signal.

Dans le domaine temporel, on définit la phase instantanée $\nu_i(t)$ par :

$$\nu_i(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt}, \text{ avec } Z(t) = A(t) e^{i\phi(t)},$$

et dans le domaine fréquentiel, on définit le retard de groupe par :

$$\tau_g(\nu) = - \frac{1}{2\pi} \frac{d\theta(\nu)}{d\nu}$$

avec $z(\nu)$, transformée de Fourier de $Z(t)$, s'écrivant comme :

$$z(\nu) = a(\nu) e^{i\theta(\nu)}.$$

Ces deux estimations de la fréquence instantanée donnent une valeur unique seulement si le produit bande passante-durée (BT) est grand devant 1 ; on a alors : $\nu_i = \frac{1}{\tau_g}$.

Ces estimations de τ_g et ν_i proviennent d'une **dérivation** et sont donc très sensibles au bruit alors qu'il est possible de les obtenir directement à partir de la R.C.T.F., qui, elle, est une **représentation intégrale**.

En effet, les moments d'ordre 1 de la R.C.T.F., $\mu_\rho^1(t)$ et $\mu_\rho^1(\nu)$, donnent :

$$\mu_\rho^1(t) = \nu_i(t)$$

$$\mu_\rho^1(\nu) = \tau_g(\nu).$$

Outre ces propriétés intéressantes, la représentation temps-fréquence fournit une estimation locale spectrale (Figure 131) que l'on peut essayer d'exploiter pour mettre en évidence des variations fines de la trace.

2.1.3. Etude numérique de la R.W.V.

Par changement de variable, on a :

$$\rho(t, \nu) = 2 \int_{-\infty}^{+\infty} Z(t + \theta) Z^*(t - \theta) e^{-4i\pi\nu\theta} d\theta$$

$$\rho(t, \nu) \xrightleftharpoons[2 f\theta]{} 2 Z(t + \theta) Z^*(t - \theta).$$

Le signal analytique Z associé à S est calculé comme suit :

$$S(\theta) \xrightleftharpoons[\theta\nu]{} s(\nu) \longrightarrow z(\nu) = 2 s(\nu) u(\nu)$$

$$z(\nu) \xrightleftharpoons{} Z(t) = S(t) + i QS(t),$$

u est la pseudo-fonction unité d'Heaviside, QS(t) est le signal en quadrature avec S(t). Les signaux utilisés analysés sont d'énergie finie, centrés (s(ν) = 0), causaux et s'écrivent :

$$S(t) = A(t) \pi_T(t - \frac{T}{2}) \cos \varphi(t).$$

Sous l'hypothèse BT >> 1, le signal analytique pourra s'écrire :

$$Z(t) = A(t) \pi_T(t - \frac{T}{2}) e^{i\varphi(t)}.$$

Le signal Z(t) est à "durée et bande limitées", sous l'hypothèse BT >> 1.

Donc pour t = ti, on calcule ρ(ti, ν) par voie numérique.

2.1.4. Discrétisation de la représentation de Wigner-Ville

$$\rho(n, k) = \sum_{m=-\infty}^{+\infty} Z(n + \frac{m}{2}) Z^*(n - \frac{m}{2}) e^{-i2\pi km}$$

$$\rho(n, k) = \underset{\frac{m}{2} \rightarrow k}{\text{DFT}} \left[Z(n + \frac{m}{2}) Z^*(n - \frac{m}{2}) \right]$$

$$\rho(n, k) = \underset{\theta \rightarrow 2k}{\text{DFT}} \left[Z(n + \theta) Z^*(n - \theta) \right]$$

k = fréquence,
n = temps.

Pour un meilleur pas fréquentiel, on prend la valeur entre les échantillons : $x(n + \frac{1}{2})$.

Pour cela, on calcule le signal analytique dans le domaine fréquentiel sur le double de points.

$$S(\theta) \rightleftharpoons s(\nu)$$

$$z(\nu) = 2 u(\nu) s(\nu)$$

$$u(\nu) = \text{fonction d'Heaviside} \quad \left\{ \begin{array}{l} 1 \text{ pour } \nu > 0 \\ 1/2 \text{ pour } \nu = 0 \\ 0 \text{ pour } \nu < 0 \end{array} \right.$$

$$z(\nu) \begin{array}{c} \xrightarrow{\text{inverse}} \\ \xleftarrow{\text{FFT}} \end{array} 2 Z(t) = S(t) + i H [S(t)]$$

$$\rho(t_i, \theta) = 2 Z(t_i + \theta) Z^*(t_i - \theta)$$

$$\rho(t_i, \nu) \rightleftharpoons \rho(t_i, \theta)$$

2.2. R.W.V. d'un signal modulé linéairement en fréquence : influence de la fenêtre de pondération

Un tel signal (Figure 131a) est caractérisé par :

$$A(t) = \pi_T(t) \text{ (porte),}$$

$$\phi(t) = 2\pi(f_0 t + \alpha \cdot \frac{t^2}{2})$$

$$\nu_i(t) = \alpha t + f_0 .$$

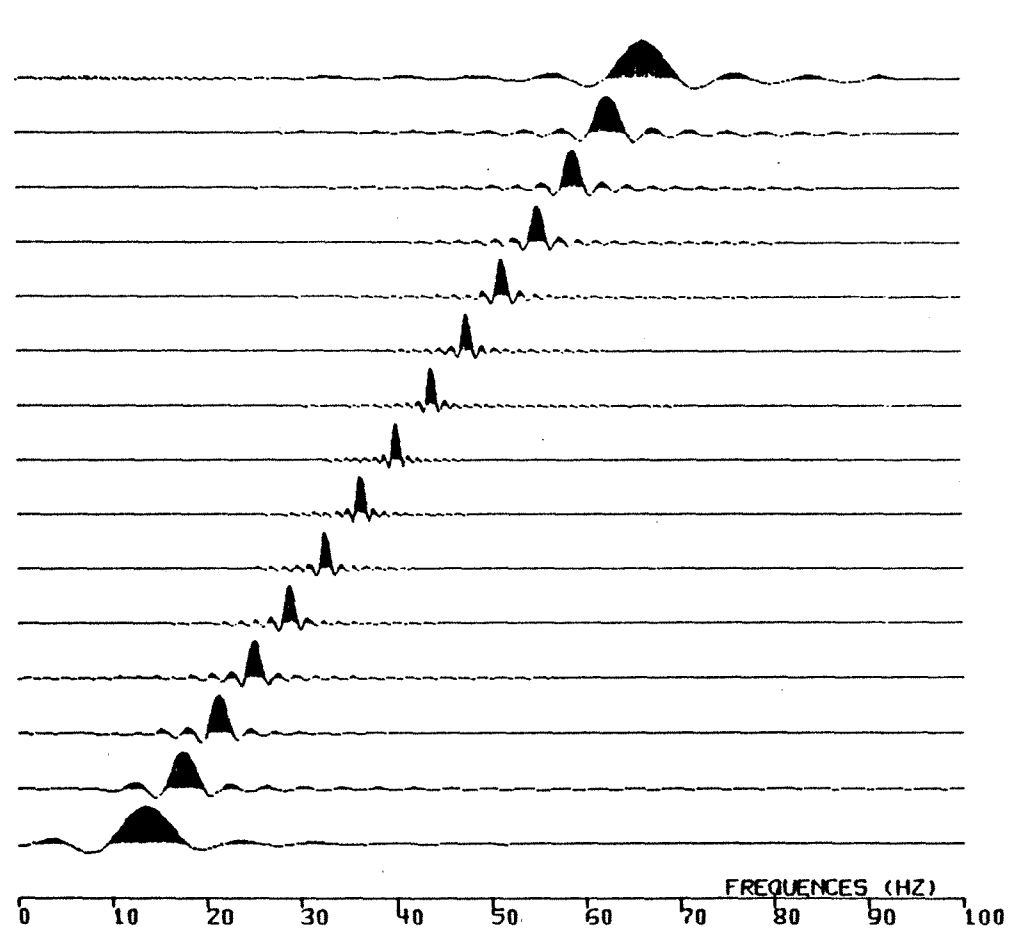
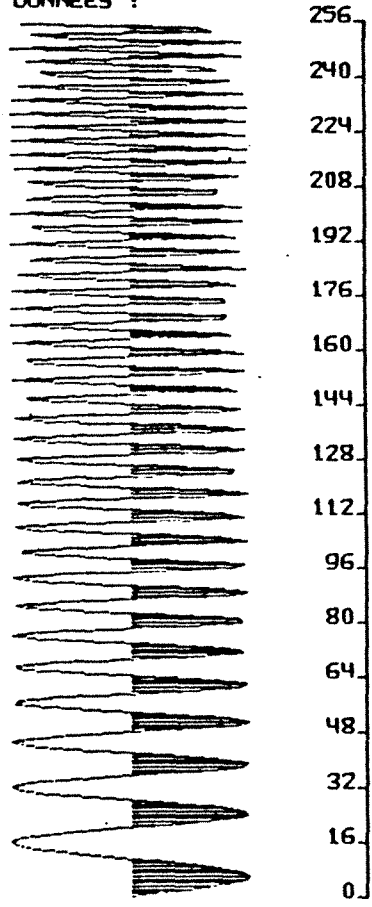
Le calcul de $\rho(t, \nu)$ fournit :

$$\rho(t, \nu) = \pi_T(t) \frac{\sin 2\pi (T - 2/t) (\nu - \nu_i(t))}{\pi(\nu - \nu_i(t))} .$$

Si on s'intéresse au cas physique où le signal $S(t)$ commence à $t = 0$, compte tenu de la propriété d'invariance par translation de la R.C.T.F., il vient :

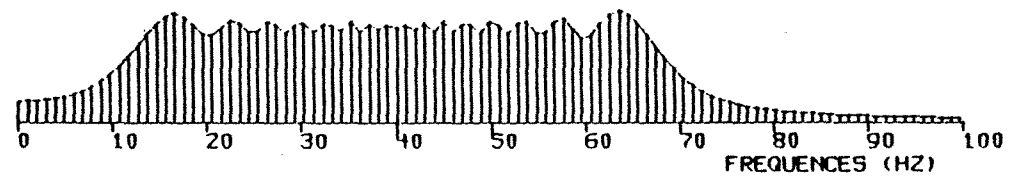
REPRESENTATION DE WIGNER-VILLE

DONNEES :



SPECTRE

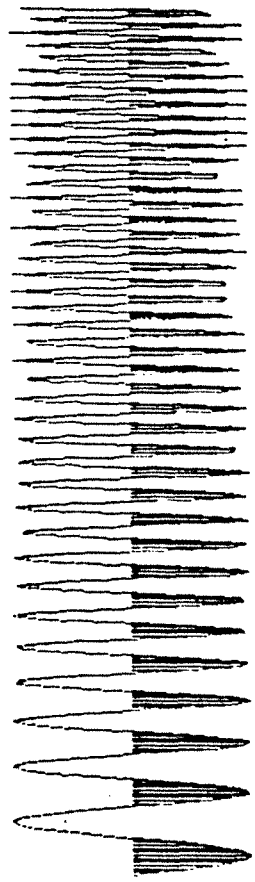
$\Delta T = 4MS$ FENETRE = 256



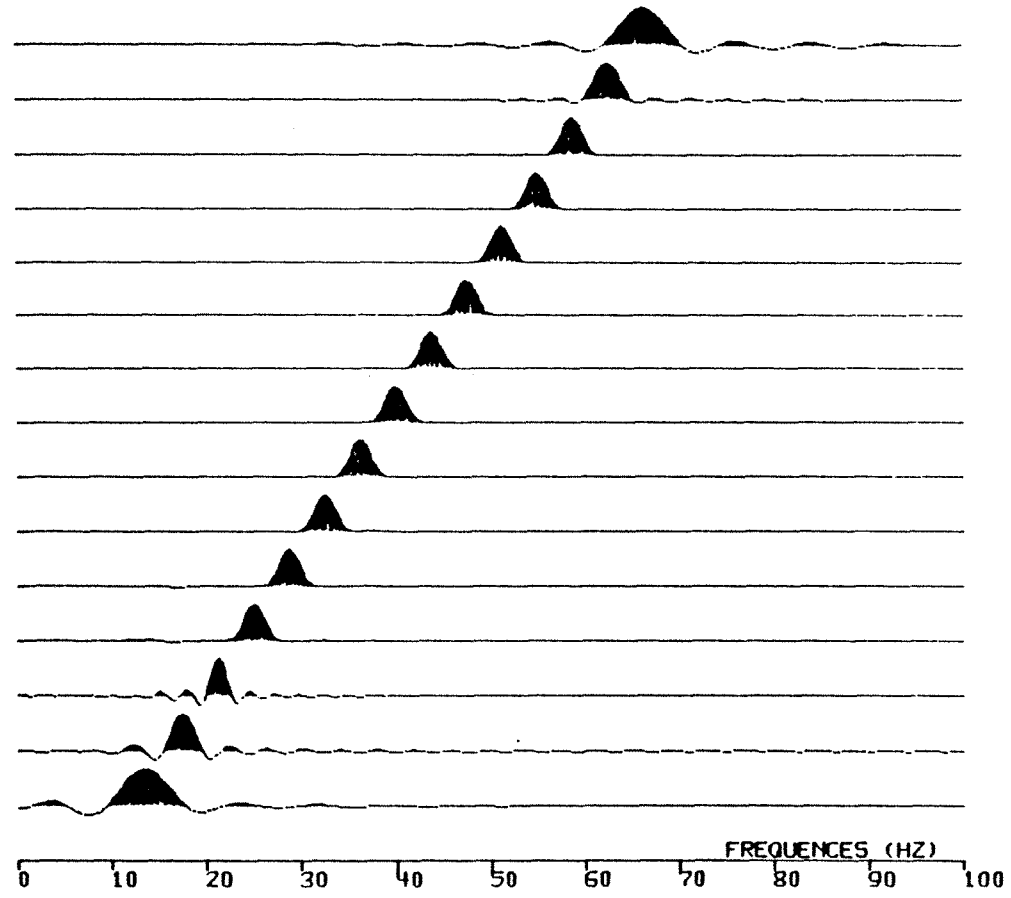
SIGNAL FM(10-70HZ) AVEC UNE LOI FM LINEAIRE

REPRESENTATION DE WIGNER-VILLE

DONNEES :

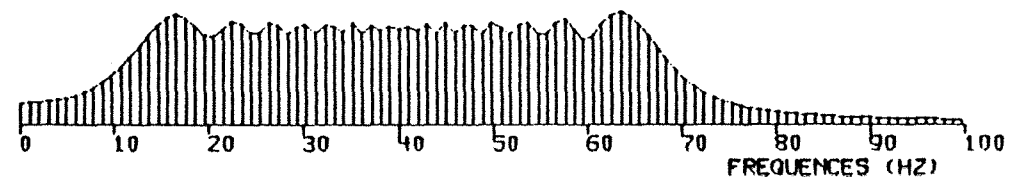


256
240
224
208
192
176
160
144
128
112
96
80
64
48
32
16
0



SPECTRE

$\Delta T = 4\text{MS}$ FENETRE = 64

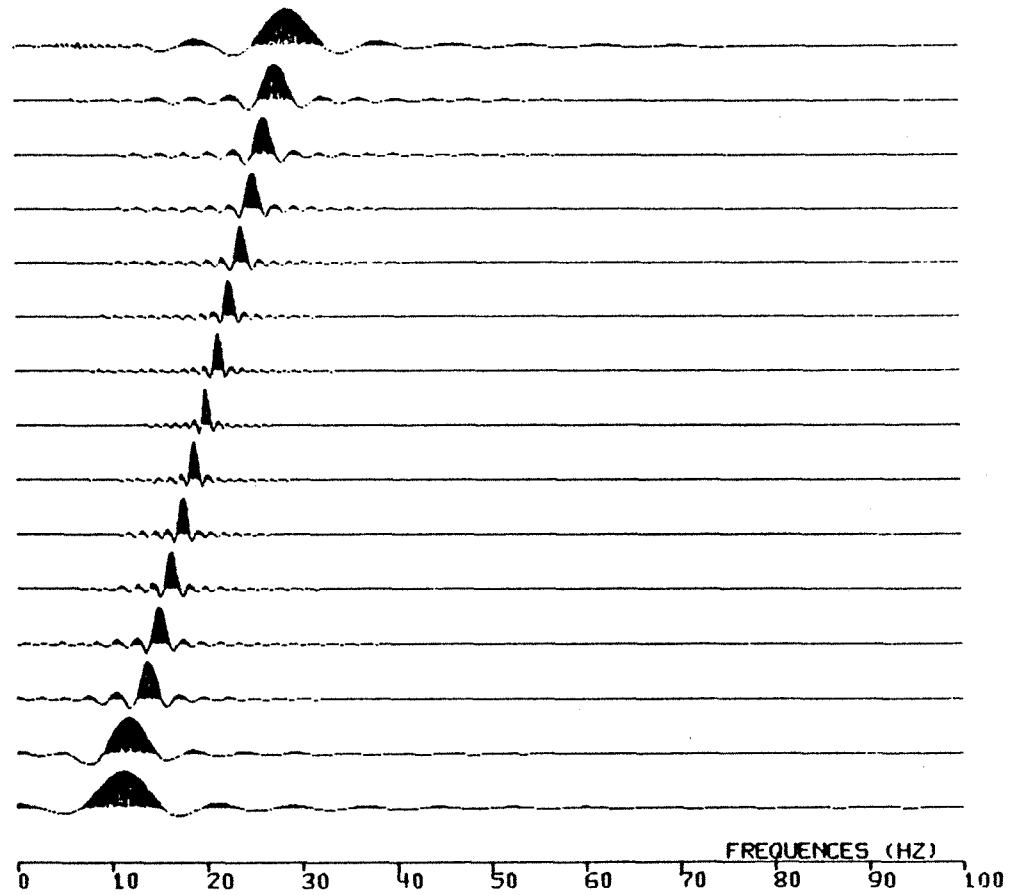
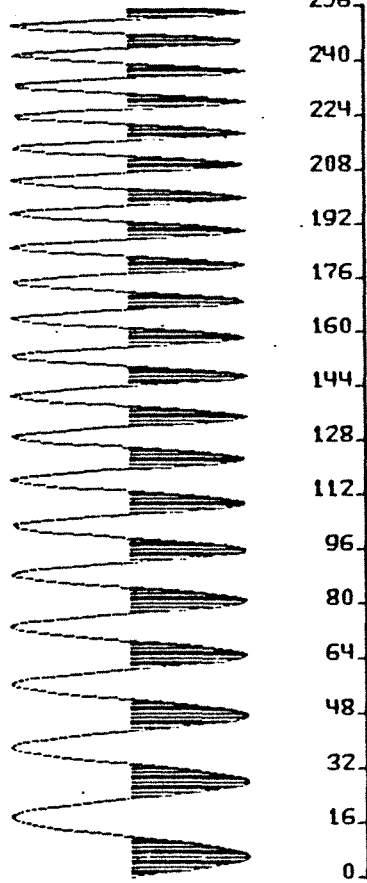


SIGNAL FM (10-70HZ) AVEC UNE LOI FM LINEAIRE

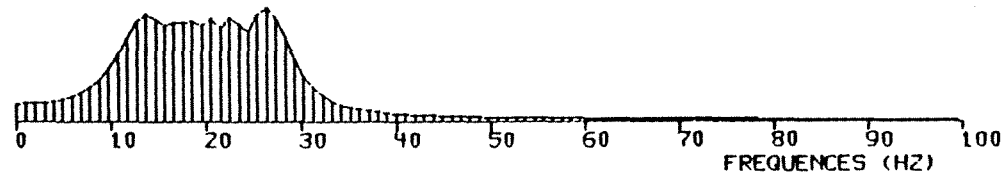
Fig. 132

REPRESENTATION DE WIGNER-VILLE

DONNEES :



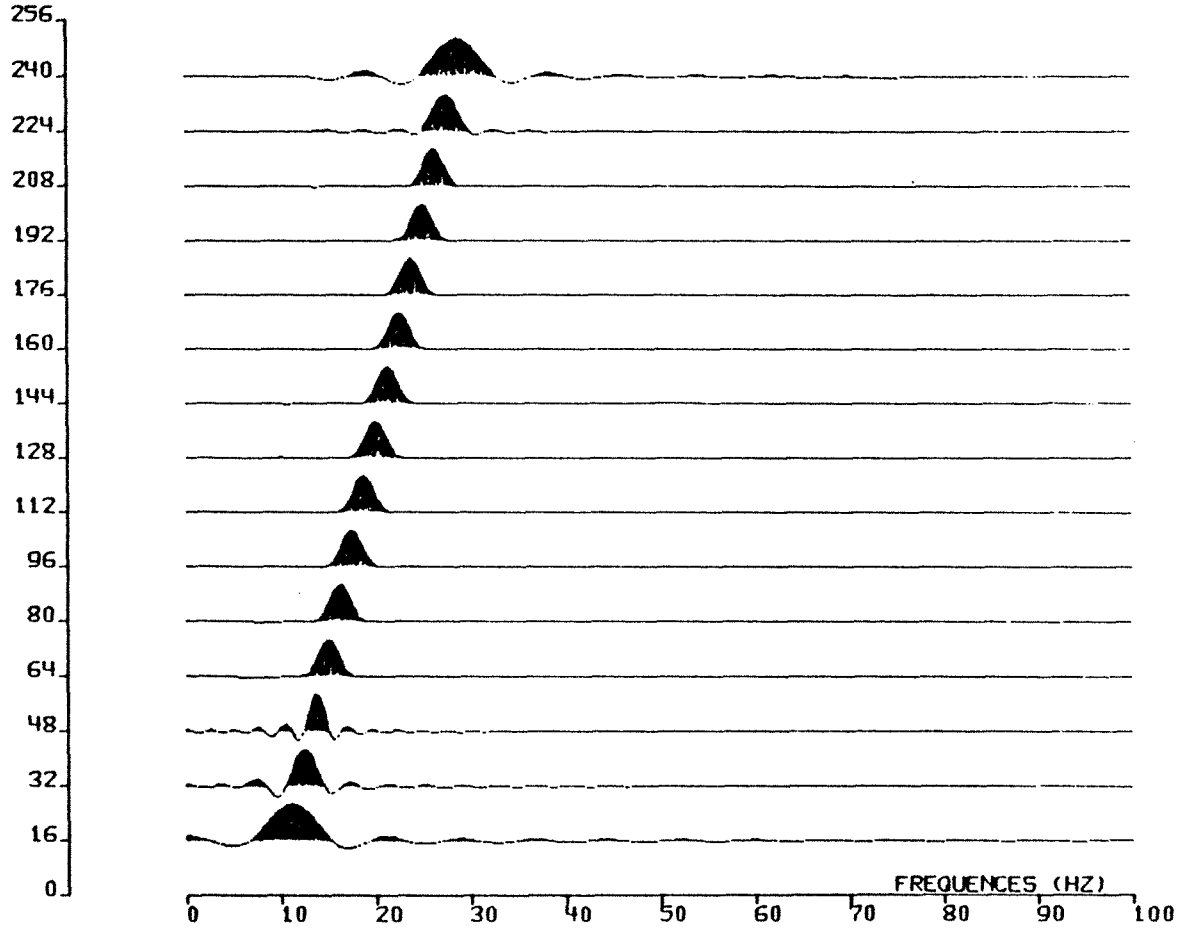
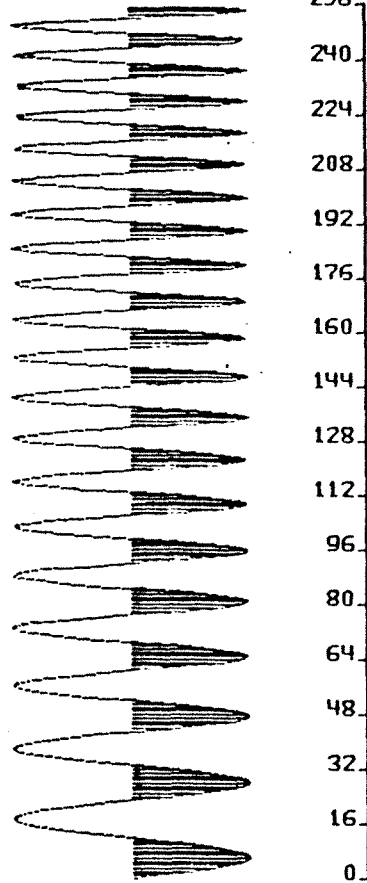
SPECTRE
AT-4MS FENETRE-256



SIGNAL FM (10-30HZ) AVEC UNE LOI FM LINEAIRE

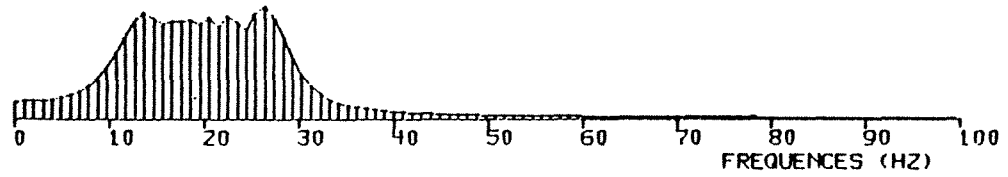
REPRESENTATION DE WIGNER-VILLE

DONNEES :



SPECTRE

$\Delta T = 4 \text{MS}$ FENETRE = 64



SIGNAL FM (10-30HZ) AVEC UNE LOI FM LINEAIRE

$$\rho(t, \nu) = \pi_T(t - \frac{T}{2}) \frac{\sin 2\pi (T - 2/(t - T/2)) (\nu - \nu_i(t))}{2 \pi(\nu - \nu_i(t))}$$

Ce signal $S(t)$ est analogue à ceux utilisés en technique vibrosismique pour la géophysique pétrolière. L'observation de la courbe obtenue (Figure 131a) fait apparaître, comme prévu, une arête le long de la fréquence instantanée, d'autant "plus pincée" que BT est grand, comme le montrent les figures 132 et 133 qui comparent deux signaux de BT différents.

Sur les figures 132a et 133a, on s'aperçoit, qu'à partir d'une coupe temps précise, la largeur à mi-hauteur des arches n'évolue plus. Ce phénomène correspond à un filtrage passe-bas dû à la multiplication du signal analytique par une gaussienne. La longueur temporelle gaussienne (nombre d'échantillons) est un paramètre du programme qui règle ce filtrage (comparer les résultats des figures 132a et 132b, 133a et 133b). La représentation de Wigner-Ville de la même trace sismique a été calculée, mais avec des poids différents : 256 et 64 échantillons (Figures 134a et 134b). Pour 256, il n'y a pas de multiplication par une gaussienne ; la surmodulation est très importante (Figure 134a). Pour un poids de 64 échantillons, cette surmodulation est supprimée et le spectre est stable (Figure 134b).

La fenêtre de pondération joue donc un rôle important pour l'élimination des surmodulations parasites.

2.3. Sensibilité et caractère local de la R.W.V.

2.3.1. Sensibilité

Nous avons fait beaucoup de tests qui ont tous mis en évidence la grande sensibilité de la R.W.V. à la moindre variation acoustique des séries dont la réponse sismique est analysée. Nous ne présentons que l'exemple suivant.

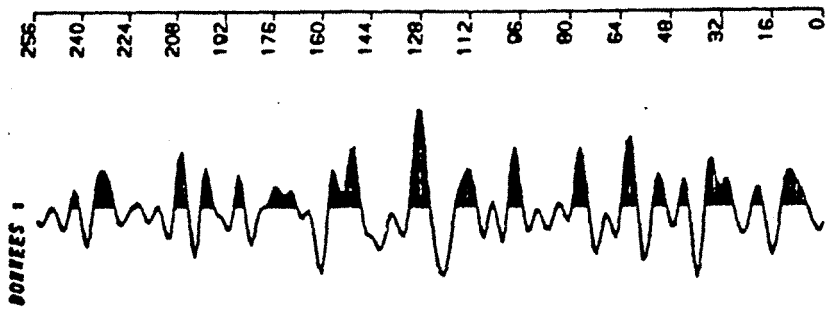
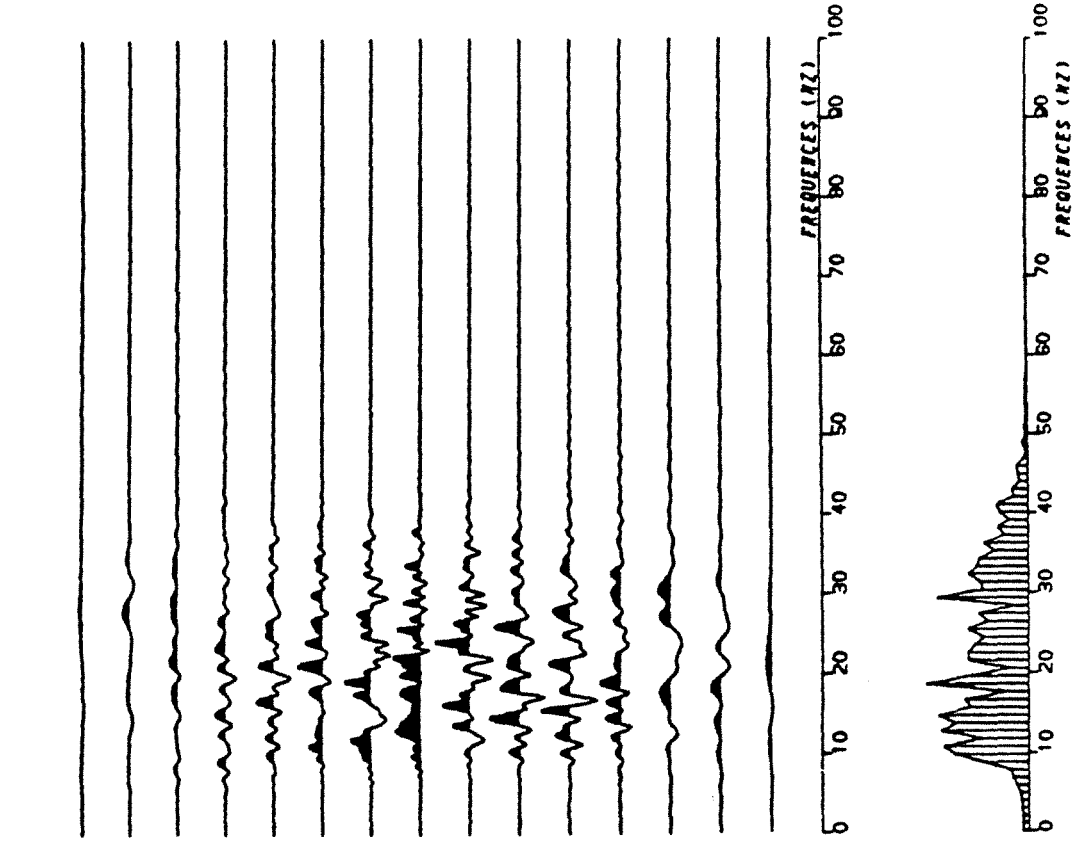
2.3.1.1. Sensibilité à une variation de vitesse (Figures 135 et 136)

On effectue les R.W.V. de deux films synthétiques de 62 échantillons chacun, au pas de 2 ms, dont les logs de vitesse ne diffèrent que par la vitesse du premier banc qui passe de 3750 ms^{-1} à 2500 ms^{-1} .

Le coefficient de réflexion de ce banc passe donc de $\pm 0,17$ à $\pm 0,35$ aux temps 22 ms et 24 ms en temps simple.

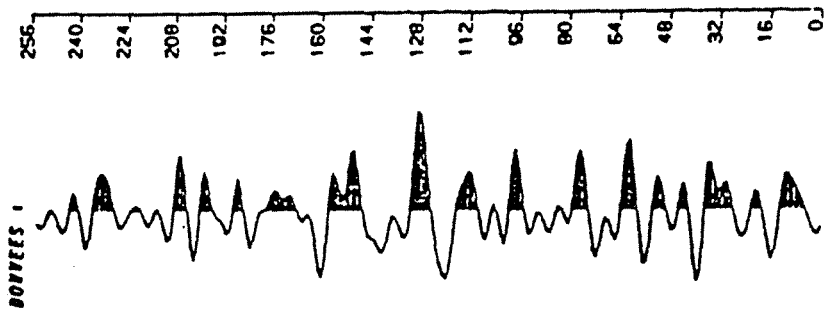
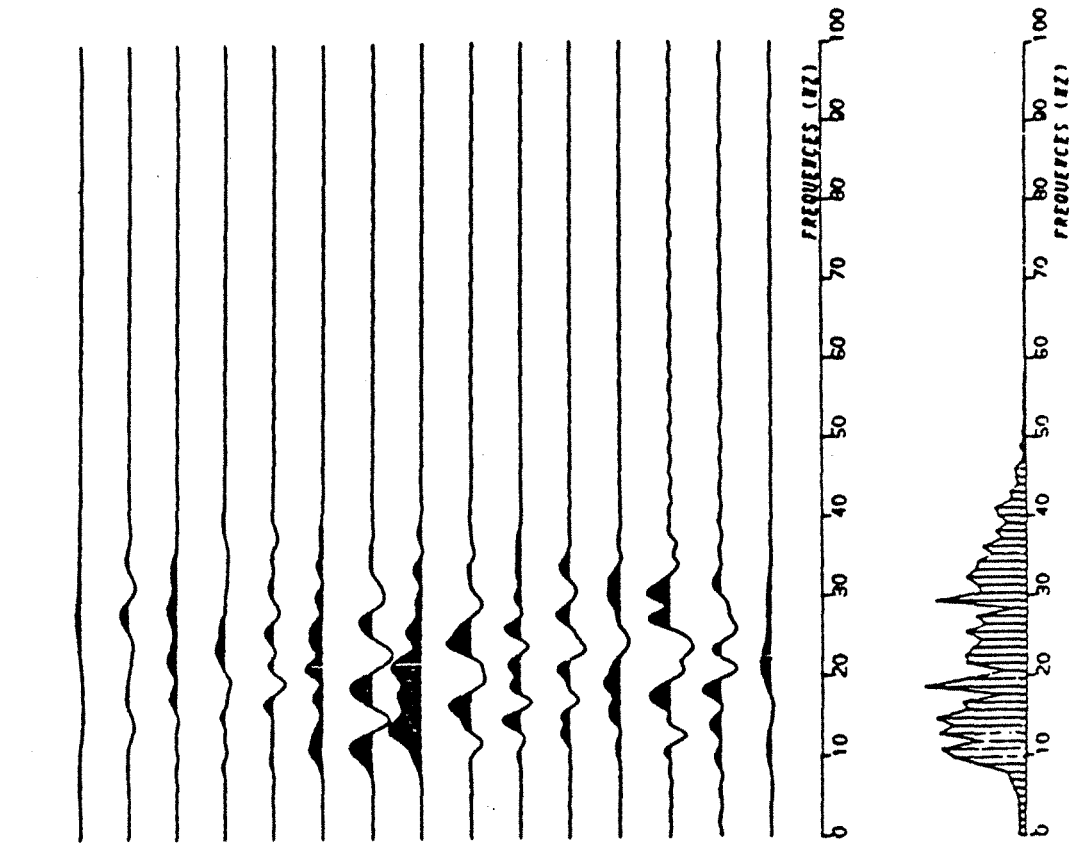
On remarque alors une nette modification de la R.W.V. du signal, le creux de coordonnées (25 Hz, 18 ms) disparaissant entièrement (Figures 135 et 136), alors que les représentations temporelle ou fréquentielle varient de façon moins significative.

REPRESENTATION DE HICHER-VILLE



SECTION : 13
PUITS : 38
TEMPS INITIAL : 1492MS
TRACE : 534
FENETRE : 256
SPECTRE

REPRESENTATION DE HIGNER-VILLE



SECTION : 13
 PUIS : 38
 TEMPS INITIAL : 1492MS
 TPAGE : 534 SPECTRE
 FENETRE : 64

Fig. 135

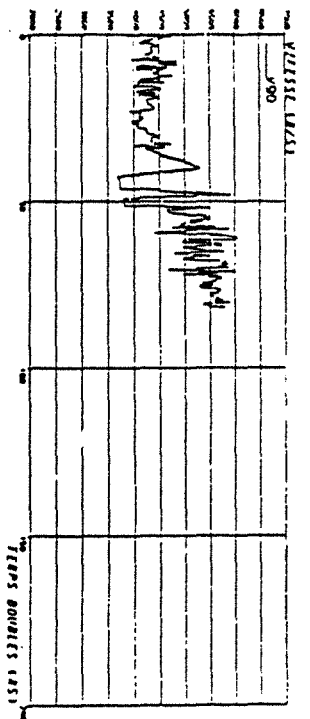
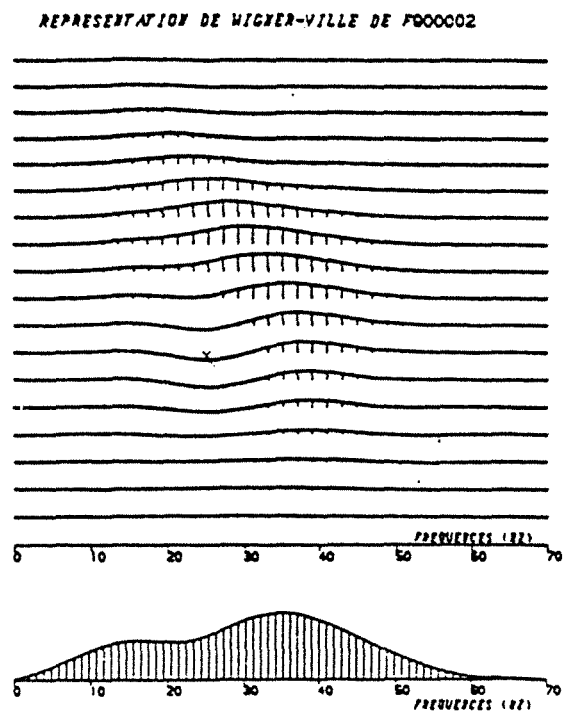
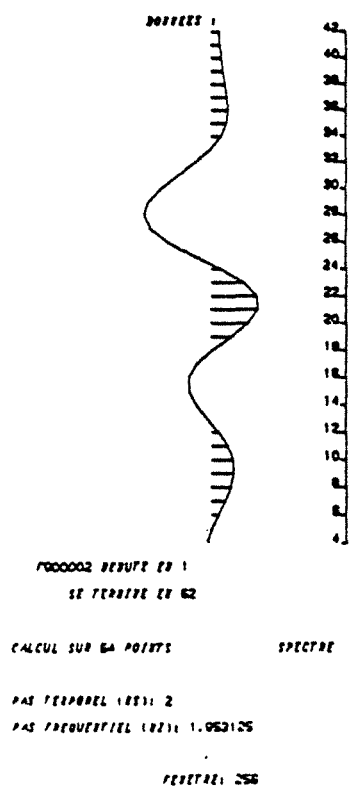
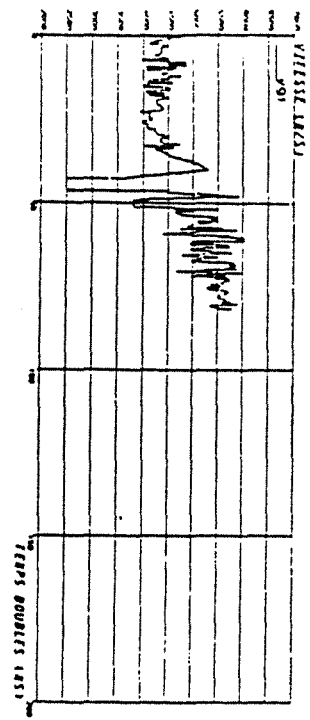
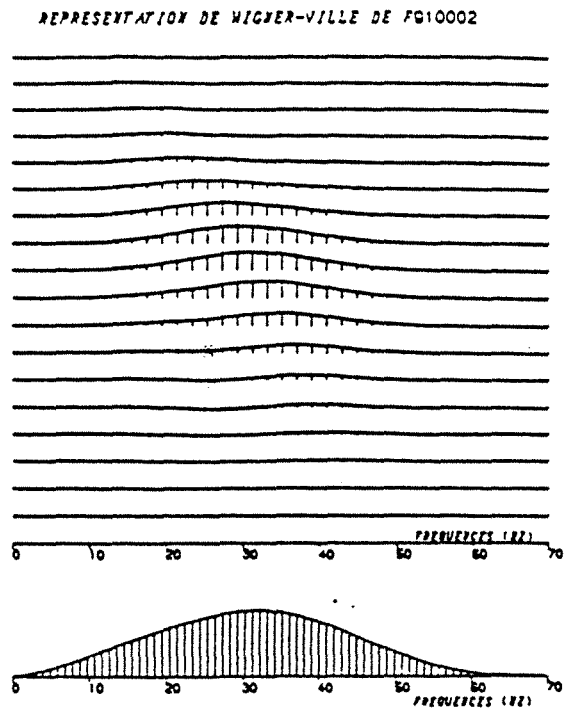
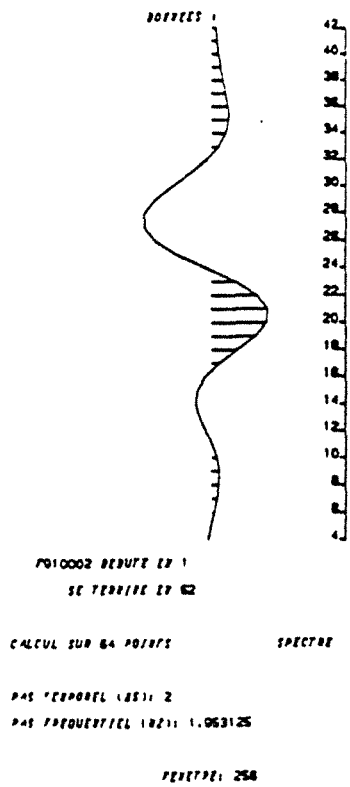
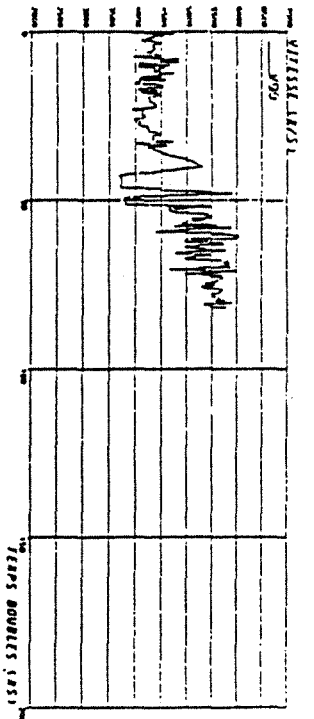
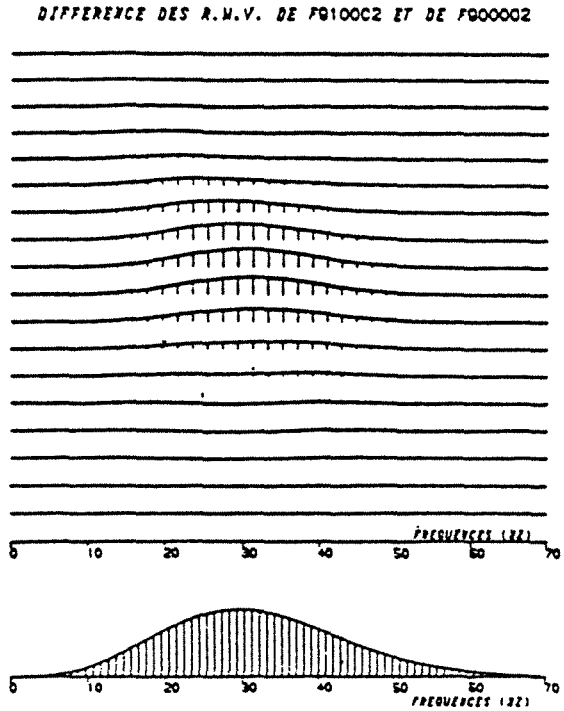
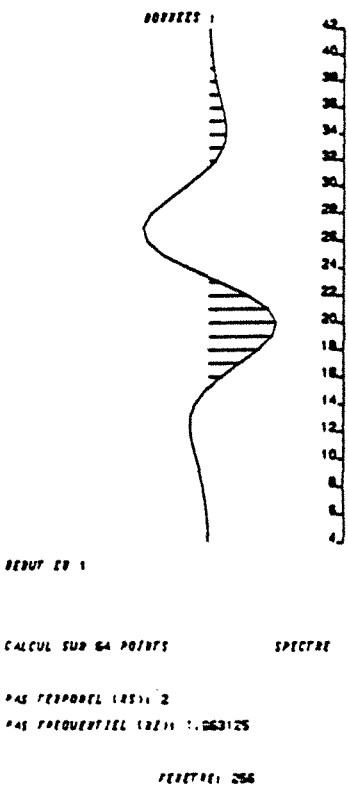
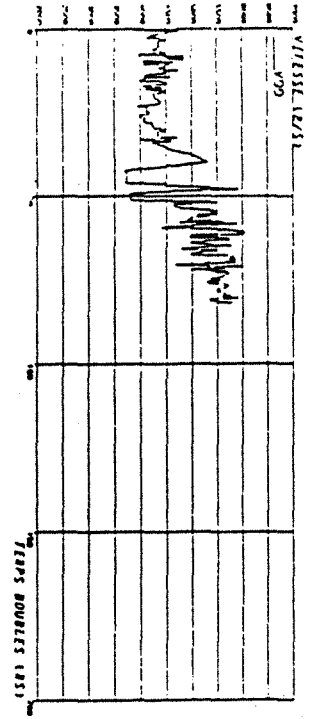
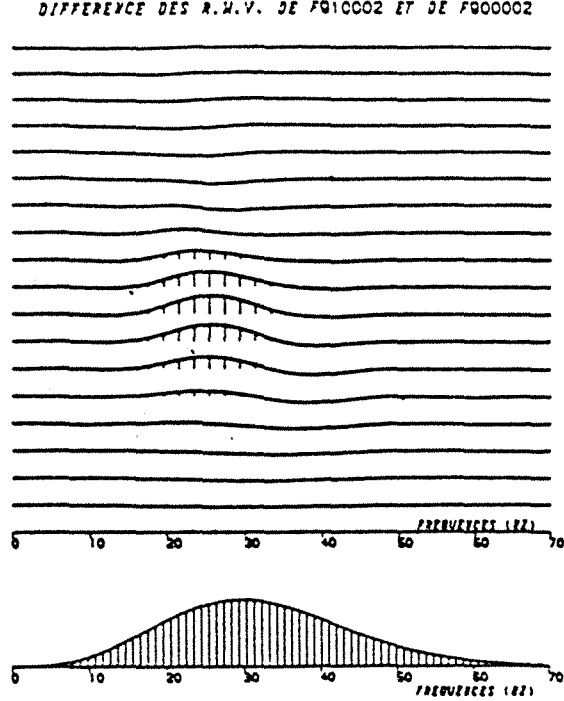
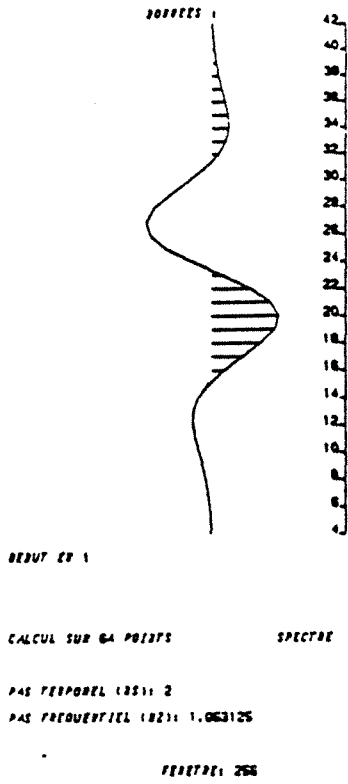


Fig.136



2.3.1.2. Sensibilité à une variation temporelle

On reprend le film synthétique précédent (avant la modification de vitesse) mais on allonge temporellement le premier banc qui passe de 4 ms à 8 ms (temps double) ce qui représente une variation d'épaisseur du banc de 8 m. On observe alors aussi la disparition du même creux sans que le signal temporel ne varie beaucoup.

Si la R.W.V. apparaît sensible à des variations de faciès géologique, il est important aussi de savoir si cette représentation permet une **localisation** de ces variations.

2.3.2. Caractère local

Là encore, beaucoup de tests ont été effectués et tous soulignent que la R.W.V. n'est que **partiellement locale**. Elle est sensible non seulement à la moindre évolution dans la fenêtre d'étude, mais aussi aux échantillons extérieurs voisins de cette fenêtre. Cette influence se caractérise par une surmodulation comme le montre l'exemple des figures 137 à 140. La même trace est analysée sur la même fenêtre au moyen de la R.W.V., mais nous avons mis à l'extérieur de cette fenêtre des échantillons différents pour compléter le signal sur une fenêtre plus large. La façon dont a été complété le signal varie d'une figure à l'autre. Elle influence nettement l'analyse R.W.V. du signal lui-même (d'un essai à l'autre, la fenêtre de pondération utilisée est identique).

Ce caractère non strictement local de la R.C.T.F. de Wigner-Ville est un peu gênant pour les fins de caractérisation géologique de la trace sismique que nous nous proposons d'étudier.

Nous avons toutefois testé l'outil sur des modèles géologiques simples, d'après les données réelles sur le Callovien Inférieur du Bassin de Paris.

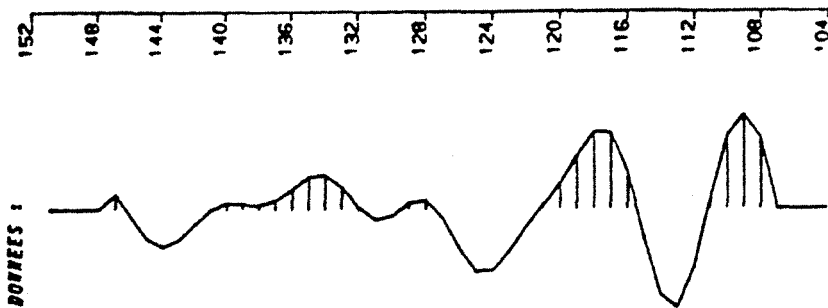
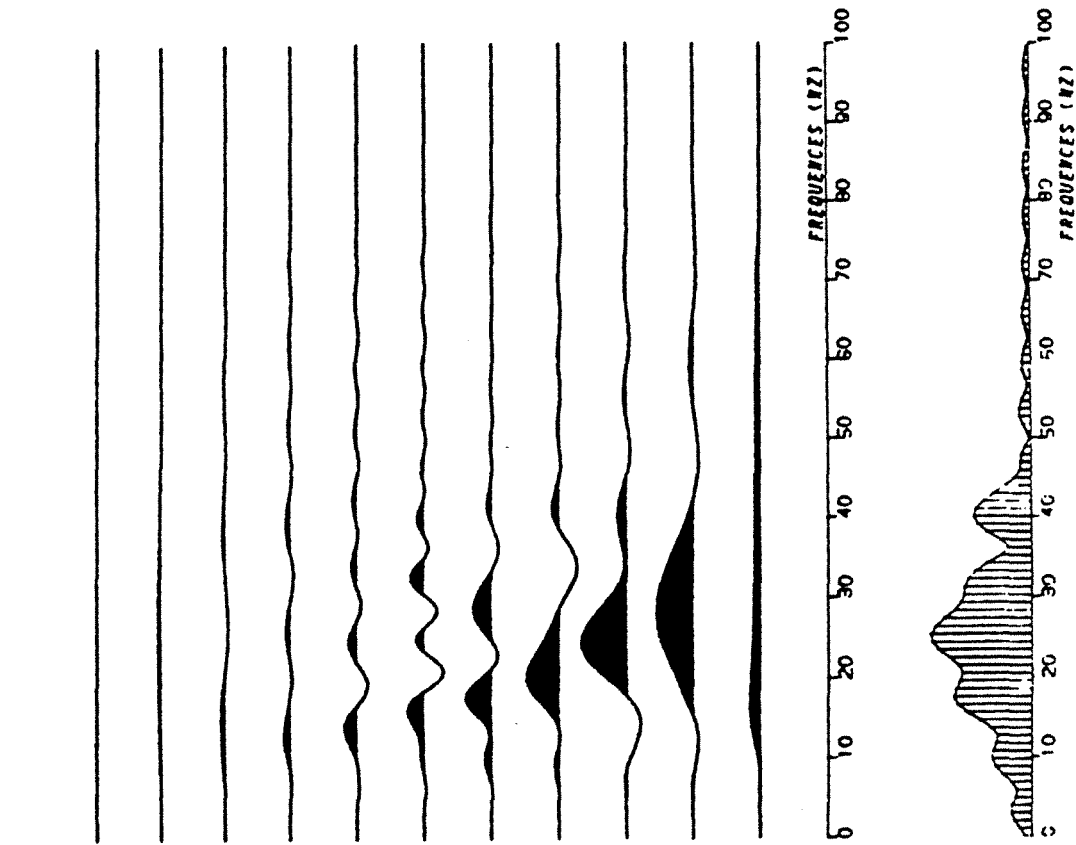
2.4. Etude des caractéristiques sismiques du Callovien Inférieur avec la R.W.V.

2.4.1. Position du problème

Le Callovien Inférieur (Figure 141), ensemble carbonaté d'environ 25 m, présente de fines variations, principalement :

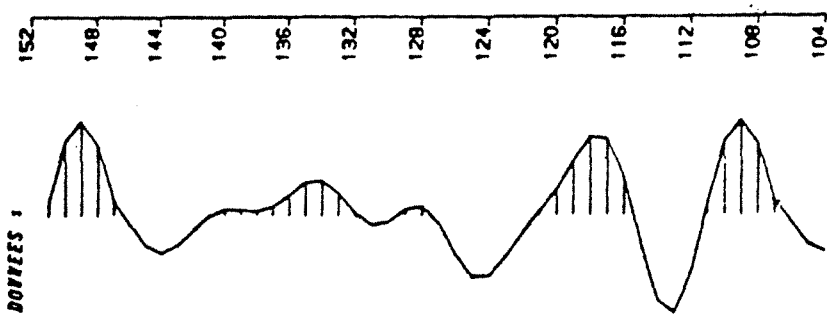
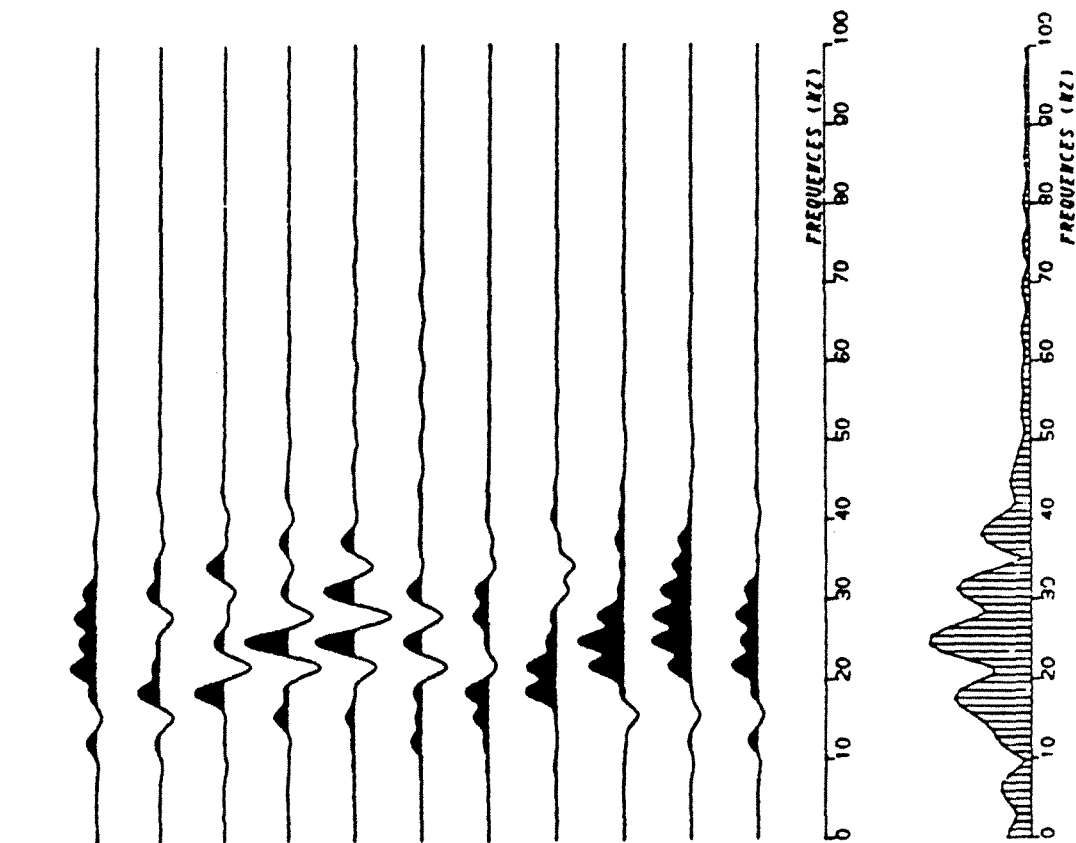
- existence ou non d'une "zone de transition", épaisse d'environ 3 m, entre le niveau oolithique sommital et le niveau des argiles bioclastiques,
- développement éventuel de niveaux plus argileux à la base des calcaires bioclastiques.

REPRESENTATION DE HIGNER-VILLE



TRACE : U22 SPECTRE

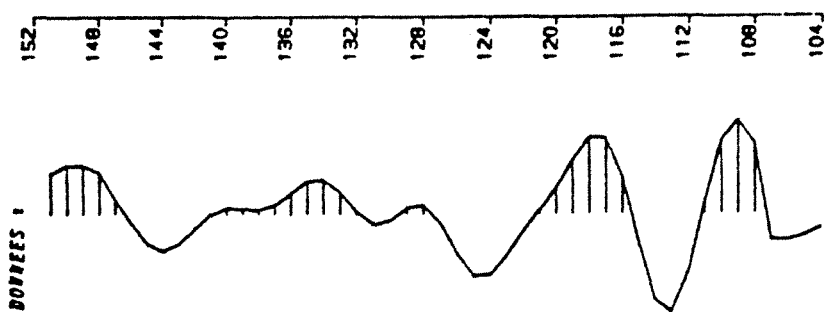
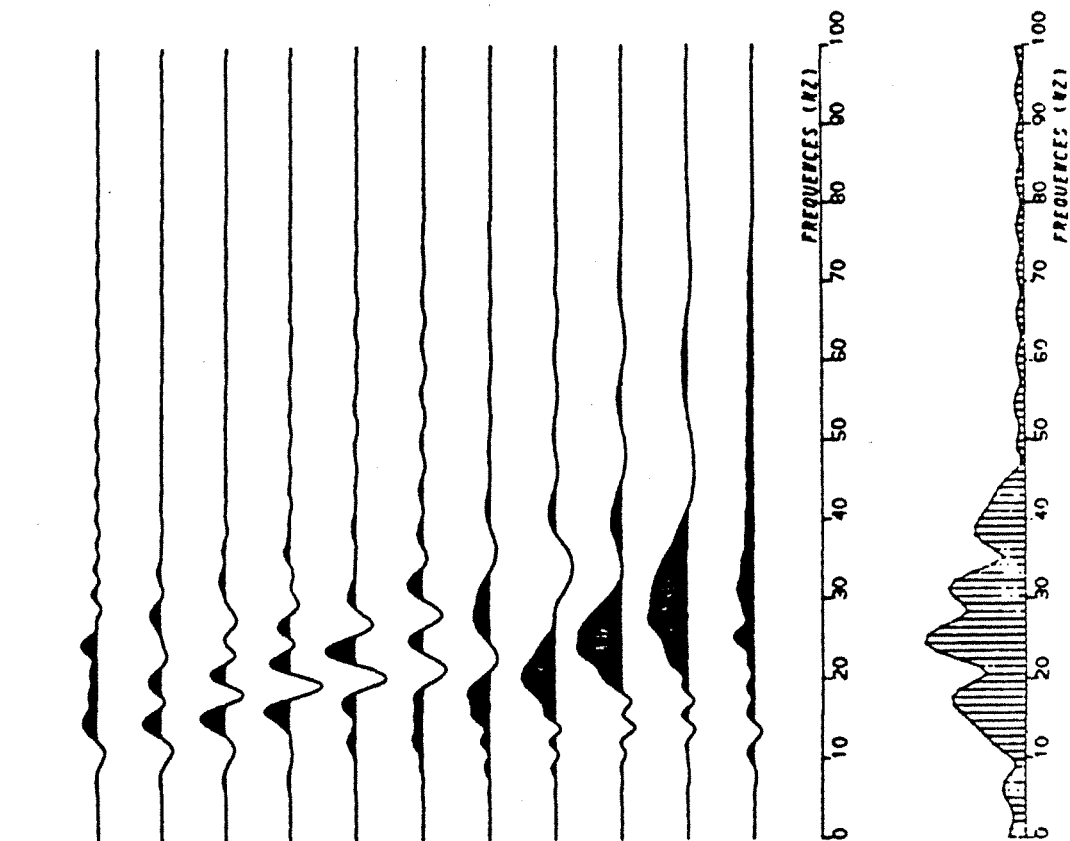
REPRESENTATION DE NIGNER-VILLE



TRACE : NXU22 SPECTRE

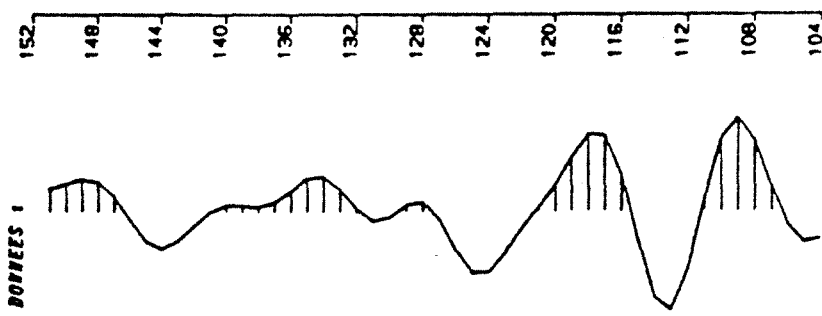
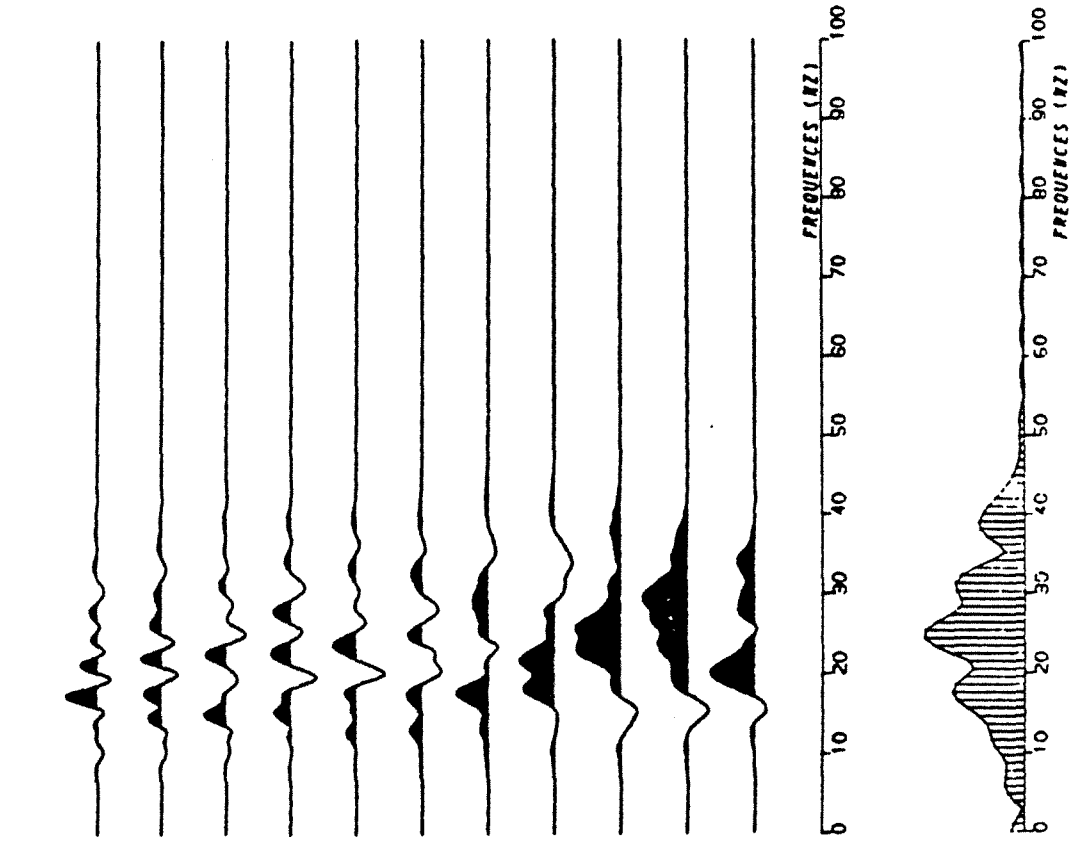
Fig.139

REPRESENTATION DE HIGNER-VILLE

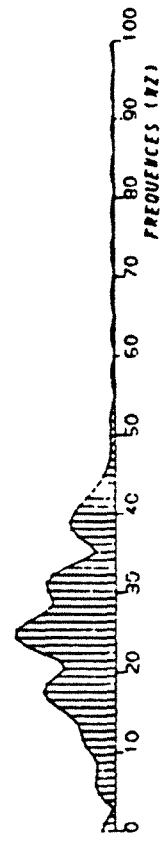


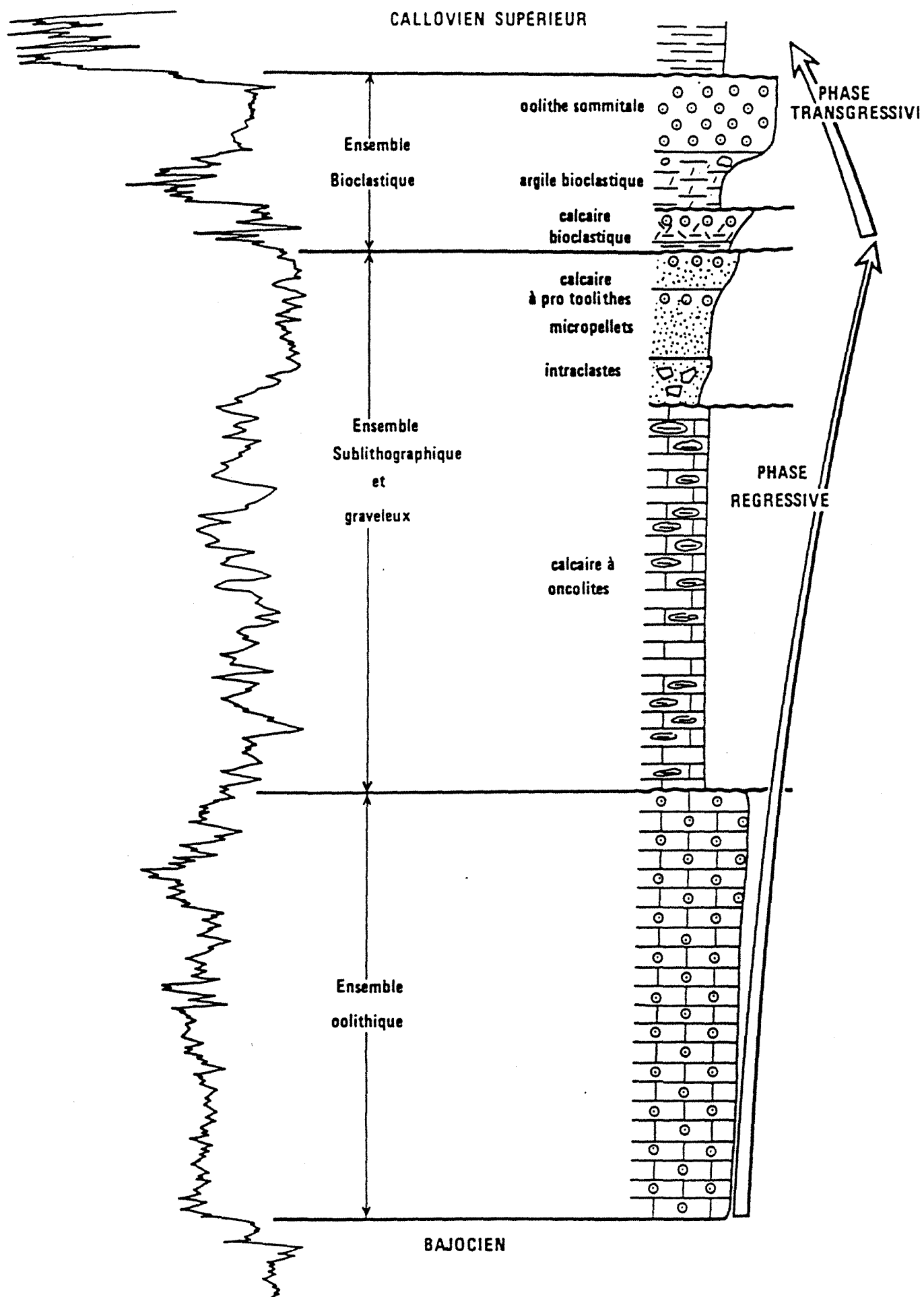
TRACE : U29-U22-U29 SPECTRE

REPRESENTATION DE WIGNER-VILLE



TRACE : U22 REEL SPECTRE





PRINCIPAUX ENSEMBLES DU BATHONIEN ET CALLOVIEN INFERIEUR

Vu les épaisseurs en question, nous nous situons bien en dessous du pouvoir de résolution de la sismique ; il ne s'agit pas de **détecter** les bancs présentant les variations décrites, mais plutôt de voir si la trace sismique ne présente pas des caractéristiques différentes à relier aux différents "patterns". L'information que nous comptons exploiter est donc **l'information interférentielle** : nous voulons essayer de la décoder en liaison avec les faciès géologiques repérés.

Il était tentant d'essayer de caractériser les différents faciès géologiques au moyen de la R.C.T.F. et d'essayer de voir s'il est possible de distinguer entre les variations d'épaisseur de la série et les variations de faciès.

Nous avons travaillé sur des données synthétiques : modèles géologiques traduisant à partir d'un modèle de référence les faciès signalés au niveau des argiles bioclastiques et des calcaires bioclastiques. Le signal (un Ricker d'ordre 2) a été choisi de telle sorte que les variations étudiées soient de l'ordre du dixième de la longueur d'onde.

Si $v_{\text{moyenne}} = 4000 \text{ m/s}$, avec un signal de pseudo-période $\approx 10 \text{ ms}$ (100 Hz de fréquence dominante), $\frac{\lambda}{10}$ vaut $\approx 4 \text{ m}$, ce qui correspond à l'ordre de grandeur des variations observées sur les données géologiques aux puits de cette zone du Bassin de Paris.

2.4.2. Description des modèles

Le modèle de base (Figure 142) comporte une barre carbonatée oolithique, un creux de vitesse correspondant aux argiles bioclastiques, une seconde barre carbonatée plus rapide (calcaires bioclastiques), une petite rampe (2,50 m) et l'ensemble sublithographique massif et rapide. Les vitesses et les épaisseurs ont été estimées en faisant une statistique sur les valeurs observées aux puits du champ (cf. tableau sur la figure 142).

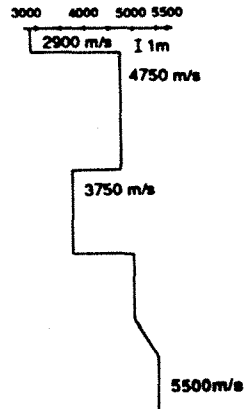
Le modèle 2 correspond au "pattern" "base argileuse" des calcaires bioclastiques (Figure 142.2). Le modèle 3 simule une zone plus argileuse au sein de cette formation (Figure 142.3) et le modèle 4 (Figure 142.4) traduit le faciès "zone de transition" pour les argiles bioclastiques.

Les traces synthétiques correspondantes sont calculées et convoluées avec le même signal que précédemment. La diminution du pic négatif du modèle 4 (Figure 142.4) traduit bien la rampe.

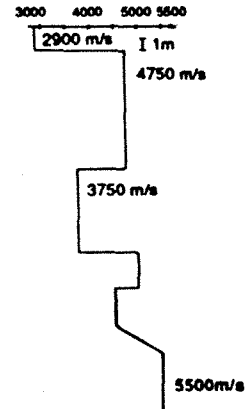
MODELES DE L'ENSEMBLE BIOCLASTIQUE

VITESSES UTILISEES:	CALCAIRE SUPERIEUR _____ 2900 m/s	CALCAIRES BIOCLASTIQUES _____ 5000m/s
	OOLITHES SOMMITALES _____ 4750 m/s	ENSEMBLE SUBLITHOGRAPHIQUE ET GRAVELEUX _____ 5500m/s
	ARGILES BIOCLASTIQUES _____ 3750 m/s	

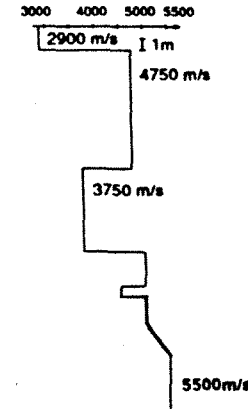
LOGS DE VITESSE



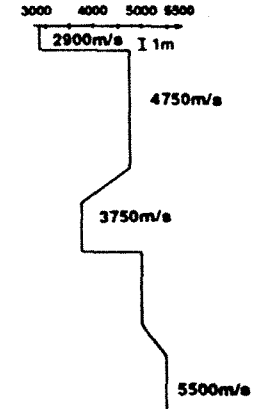
MODELE 1



MODELE 2

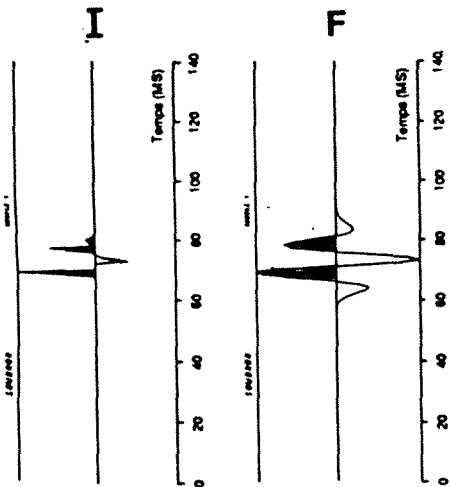


MODELE 3

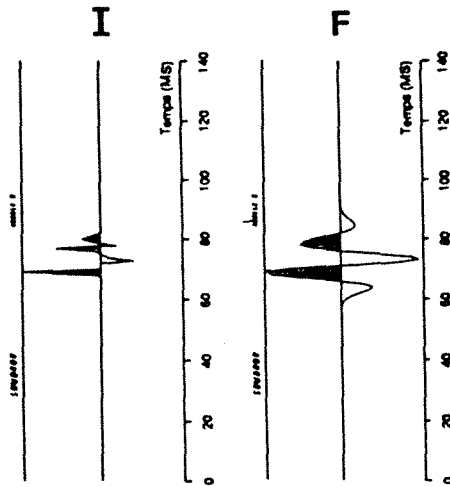


MODELE 4

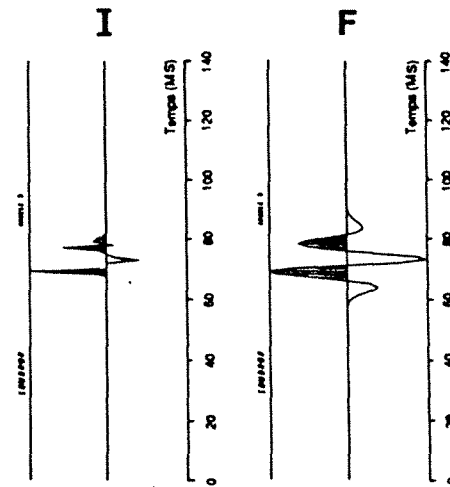
TRACES SYNTHETIQUES IMPULSIONNELLES (I) ET FILTRES (F)



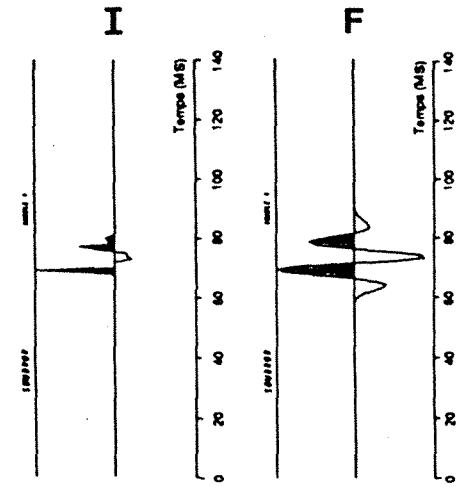
MODELE 1



MODELE 2



MODELE 3



MODELE 4

2.4.3. Calcul des R.W.V. et de leurs différences

Les représentations temps-fréquence des quatre modèles sont calculées (Figure 143) après divers essais pour le choix de la fenêtre de pondération. En fait, la pondération gaussienne introduite pour annuler les effets de bord de fenêtre s'avère inutile (traces non bruitées).

Les quatre R.C.T.F. ont toutes un aspect identique (Figure 143) avec développement d'un pic d'énergie au centre de l'intervalle d'étude (\approx coupe temps n° 9).

Les modèles sont très semblables (les différences sont métriques) et la méthode préconisée (Bruel, 1987) est d'analyser les différences normées des R.C.T.F. En prenant comme référence le modèle 1, nous avons successivement calculé et représenté la différence entre les modèles 2, 3, 4 et le modèle 1 (Figure 144).

Dans les trois cas, on constate que les différences ne sont pas nulles, on retrouve bien la très grande sensibilité de l'outil R.C.T.F. pour détecter des variations métriques.

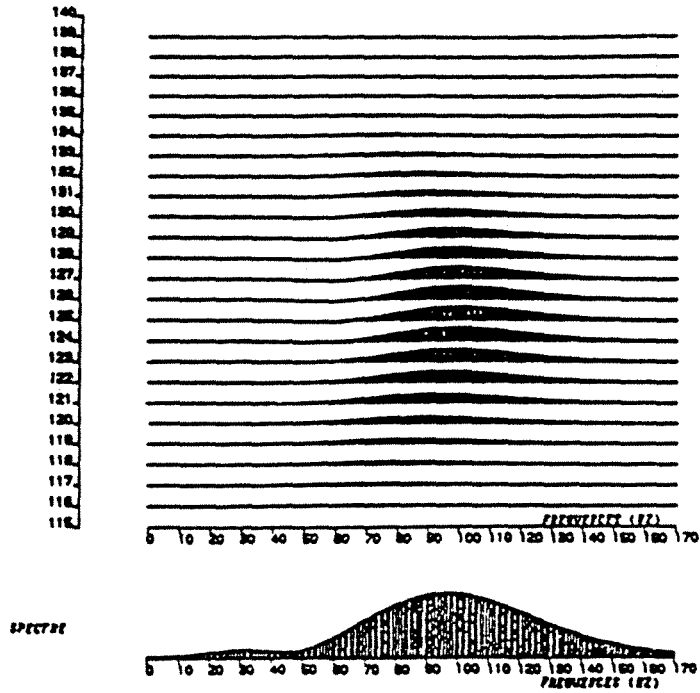
Les formes des différences (RW1-RW2) et (RW1-RW3) sont analogues, mais la seconde est beaucoup plus petite en amplitude, ceci est normal car le modèle 3 est très proche du modèle 1.

La forme de RW1-RW4 est très différente des deux autres ; les creux et les bosses sont déplacés temporellement (en particulier, apparition d'un creux négatif vers la coupe n° 8), mais l'ordre de grandeur des amplitudes est le même. On retrouve la remarque faite dans le rapport Bruel (1987) : les extrema des différences ne sont pas localisés aux coupes temps correspondant aux interfaces qui varient mais plutôt à la moitié de l'intervalle entre l'interface fixe et l'interface variable (la R.C.T.F. n'est que partiellement locale).

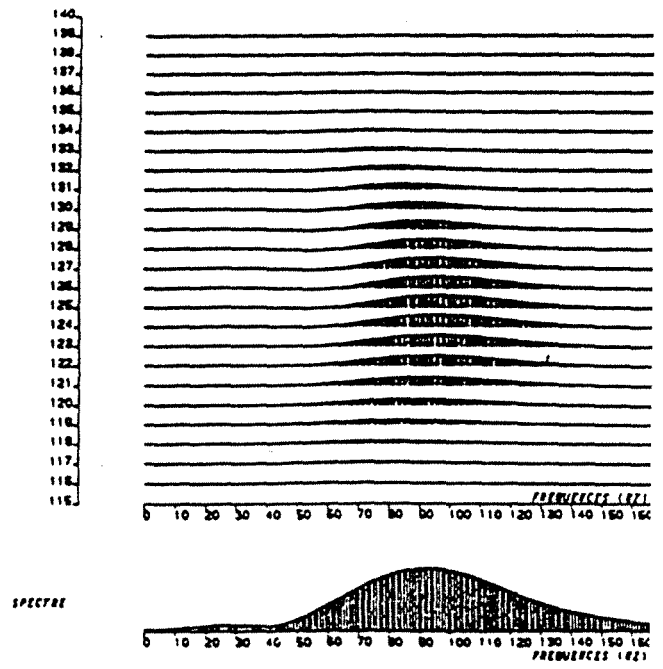
Ainsi, les différences obtenues sont caractéristiques des modèles de départ.

La question que l'on se pose est de savoir comment va évoluer la R.C.T.F. de Wigner-Ville si on introduit des variations autour des modèles représentant les faciès. En effet, le modèle 4 schématise une zone de transition d'une épaisseur de 3 m, mais cette épaisseur fluctue pour les puits qui présentent ce "pattern". Comme nous savons que la R.C.T.F. est un outil très sensible, que reste-t-il des caractéristiques observées si l'on introduit les variations réelles du "pattern" ?

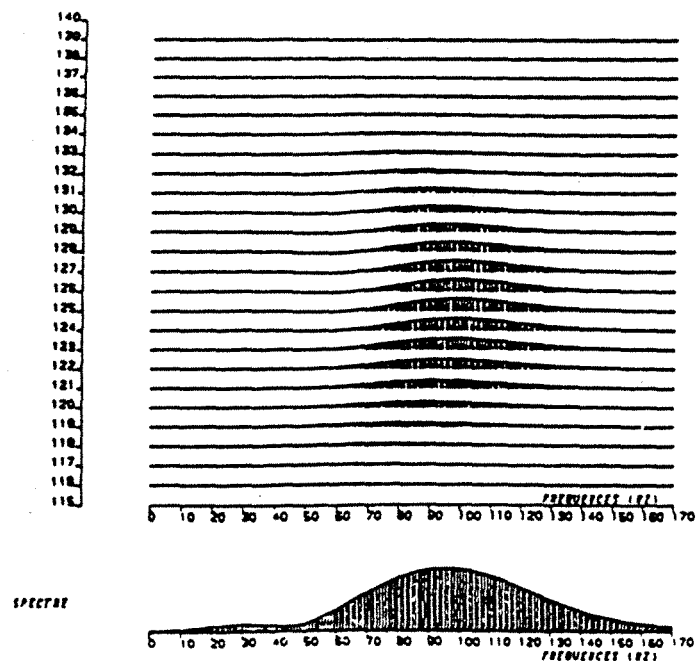
REPRESENTATION DE WIGNER-VILLE
DES MODELES DE BASE



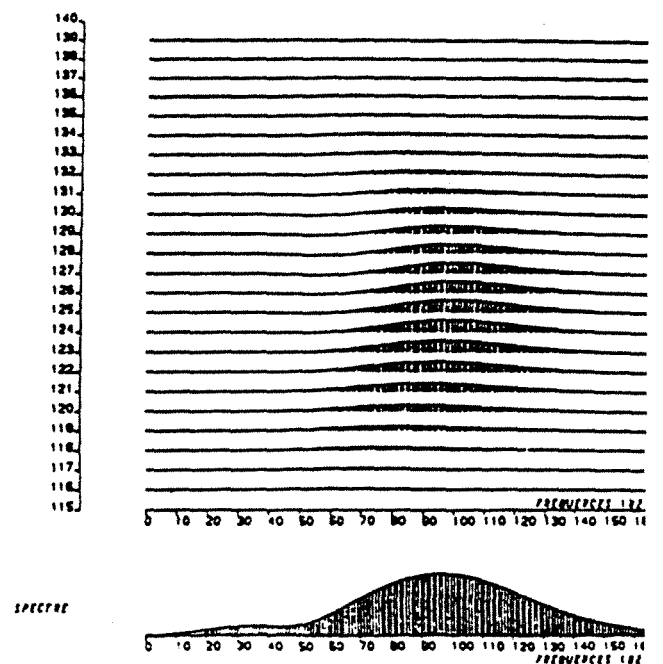
MODELE 1



MODELE 2



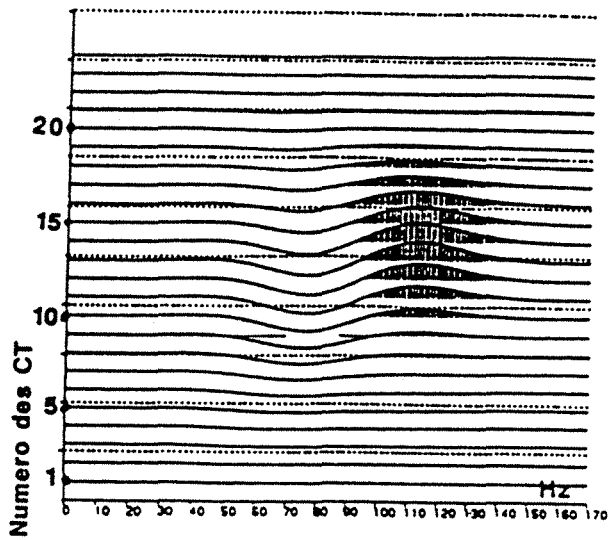
MODELE 3



MODELE 4

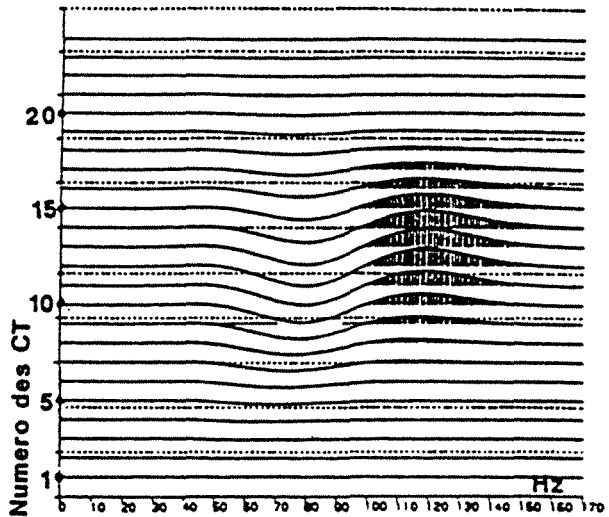
DIFFERENCES ENTRE LES MODELES DES R.C.T.F.

(REFERENCE: MODELE 1)

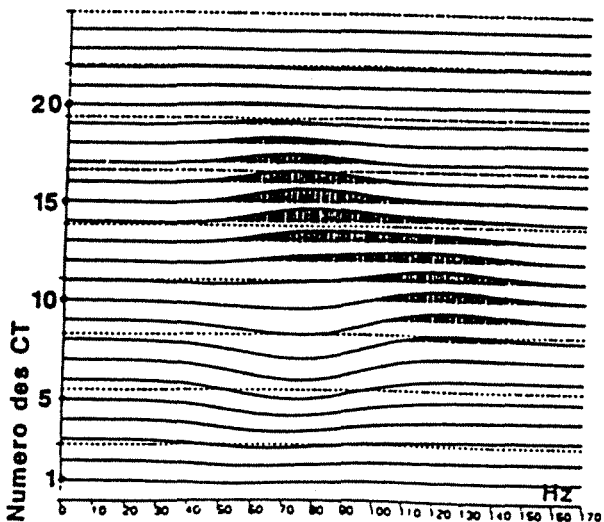


DIFFERENCE NORMEE RW1-RW2
(MODELE 1 / MODELE 2)

DIFFERENCE NORMEE RW1-RW3
(MODELE 1 / MODELE 3)



DIFFERENCE NORMEE RW1-RW4
(MODELE 1 / MODELE 4)



2.4.4. Fluctuations autour des modèles

2.4.4.1. Transition oolithes sommitales - argiles bioclastiques

La zone de transition a varié d'une épaisseur de 3 m (modèle 4) à 6 m (modèle 43). La figure 145 représente les modèles, les traces correspondantes, les différences normées de la R.C.T.F. du modèle 1 et des nouveaux modèles.

Si les deux premières réponses (Figures 145.1 et 145.2) sont proches de la réponse du modèle 4, par contre, la dernière réponse montre un glissement des pics et des creux et une évolution des amplitudes. Dans la fourchette de 2 m, l'image de la différence est à peu près identique.

2.4.4.2. Base des calcaires bioclastiques

Le modèle 2 présente une base argileuse de 3 m que l'on a fait varier entre 4 m (modèle 21), 2 m ou 1 m (modèles 22 et 23).

Les résultats sont présentés sur la figure 146. Les différences ressemblent globalement à celle entre modèles 1 et 3, mais les amplitudes deviennent très faibles pour le modèle 23 : en fait, on est très proche du modèle 1, donc la différence est très petite.

Deux images globalement différentes traduisent l'écart entre le modèle de base (1) et les modèles représentant les "patterns" étudiés.

2.4.5. Composition des modèles de base

Nous avons composé (Figure 147) le modèle "zone de transition oolithes sommitales/argiles bioclastiques" avec le modèle "base argileuse pour les calcaires bioclastiques", calculé sa R.C.T.F. et fait la différence avec la R.C.T.F. du modèle 1.

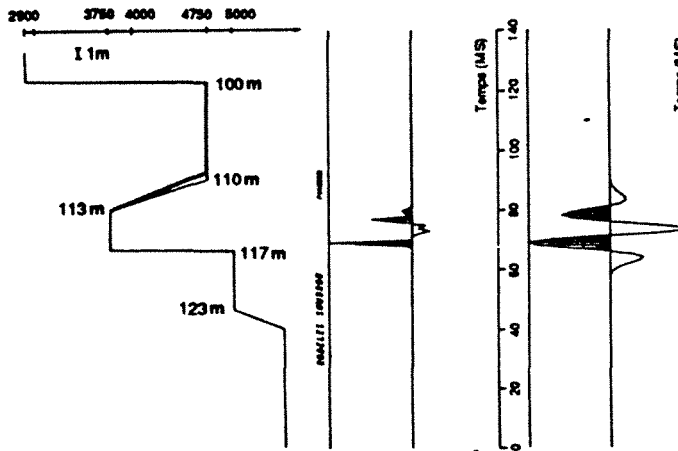
L'image obtenue est très semblable à celle correspondant à la différence entre le modèle 1 et le modèle 2. Nous avons vu que les écarts entre le modèle 1 et le modèle 2 étaient plus importants que ceux entre le modèle 1 et le modèle 4, en amplitudes, mais que la forme de ces derniers différait. Ici, la différence due aux calcaires bioclastiques, (modèle 2), "écrase" celle due à la rampe oolithes sommitales/argiles bioclastiques.

2.4.6. Comparaison entre les modèles et les traces synthétiques

Les données des modèles proviennent de l'analyse des puits disponibles sur la zone d'étude. Ces puits ont aussi été analysés du point de vue des "patterns" géologiques. Nous avons calculé les traces sismiques synthétiques des puits avec les logs d'impédances acoustiques.

VARIATION SUR LE PATTERN "RAMPE OOLITHES SOMMITALES/ARGILES BIOCLASTIQUES"

MODELE 41

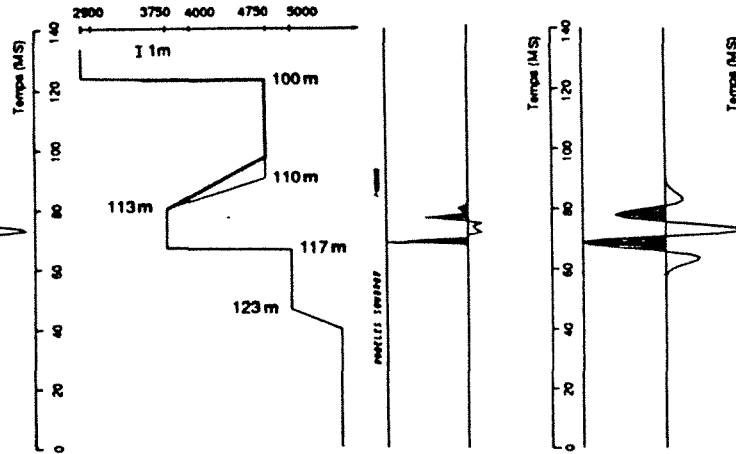


LOG VITESSE

TRACES

IMPULSION NELLE
FILTREE

MODELE 42

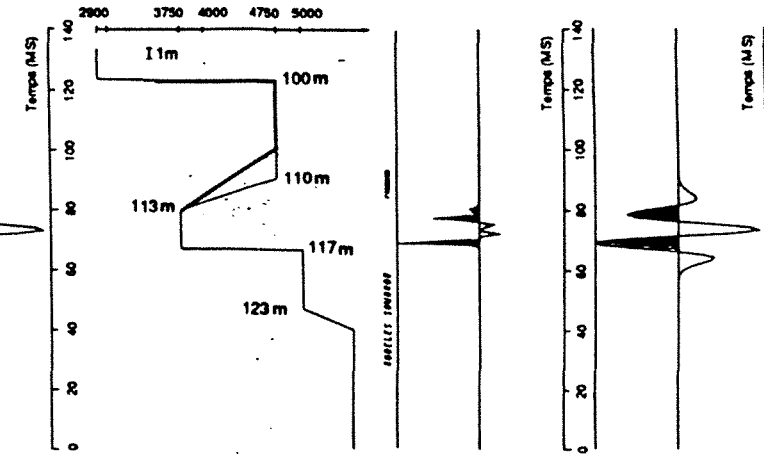


LOG VITESSE

TRACES

IMPULSION NELLE
FILTREE

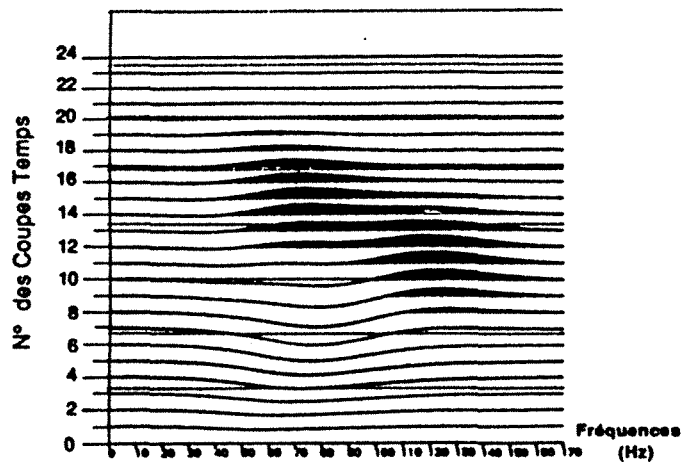
MODELE 43



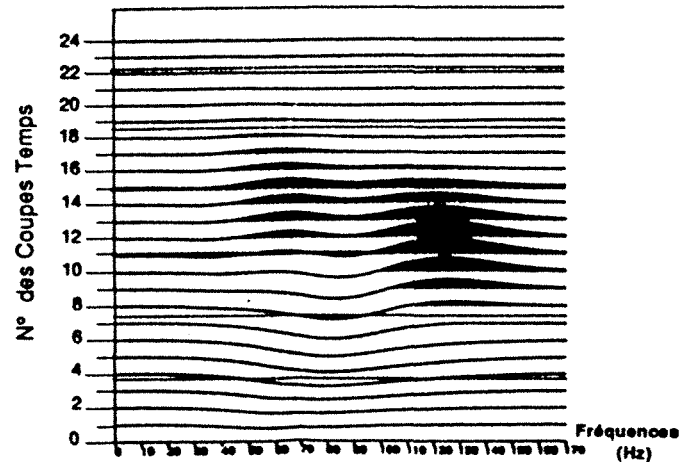
LOG VITESSE

TRACES

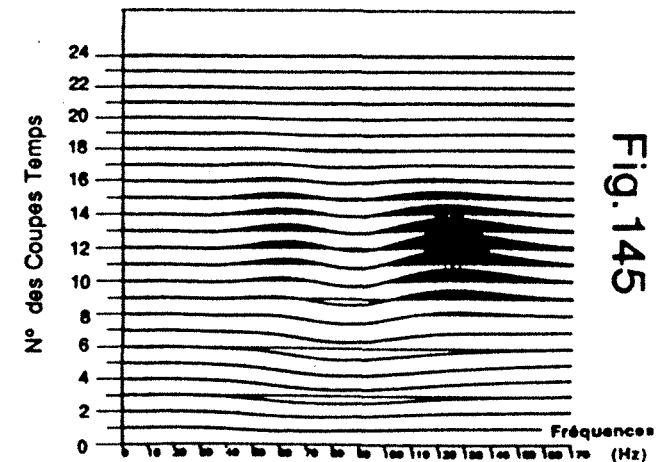
IMPULSION NELLE
FILTREE



RWV1-RWV41
(NORMEES)



RWV1-RWV42
(NORMEES)

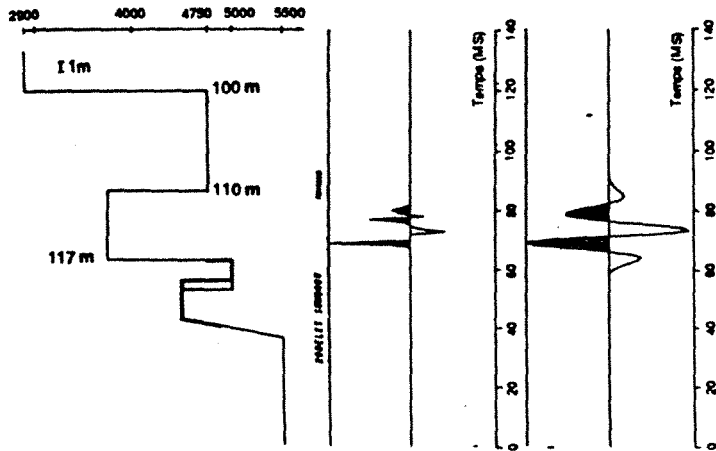


RWV1-RWV43
(NORMEES)

Fig. 145

VARIATION SUR LE PATTERN "TRANSITION CALCAIRES BIOCLASTIQUES/BATHONIEN"

MODELE 21



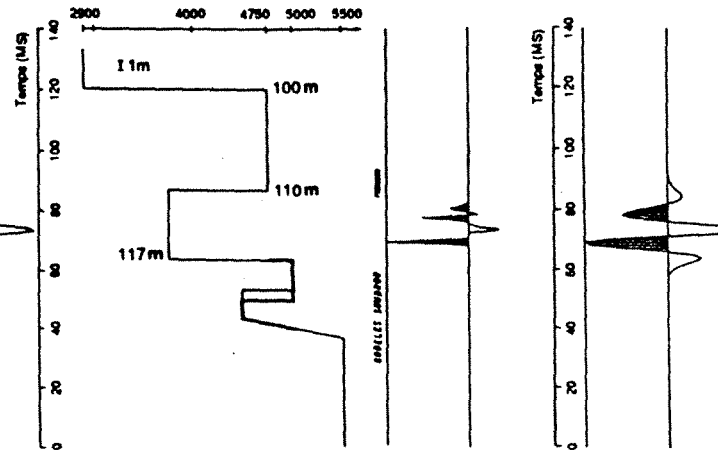
LOG VITESSE

TRACES

IMPULSION
NELLE

FILTREE

MODELE 22



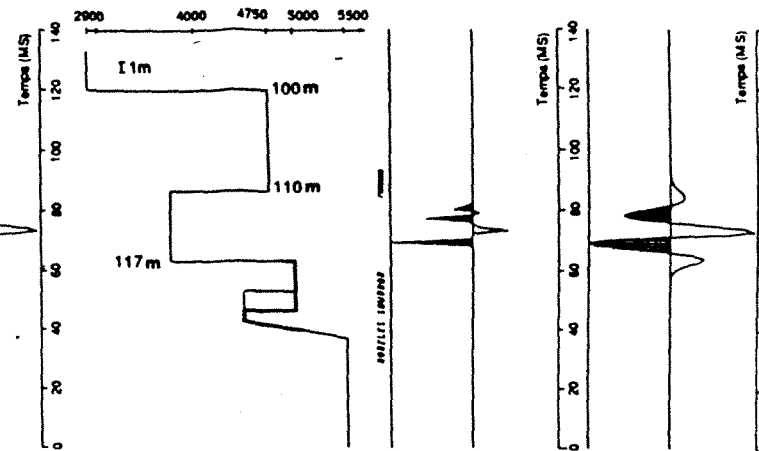
LOG VITESSE

TRACES

IMPULSION
NELLE

FILTREE

MODELE 23

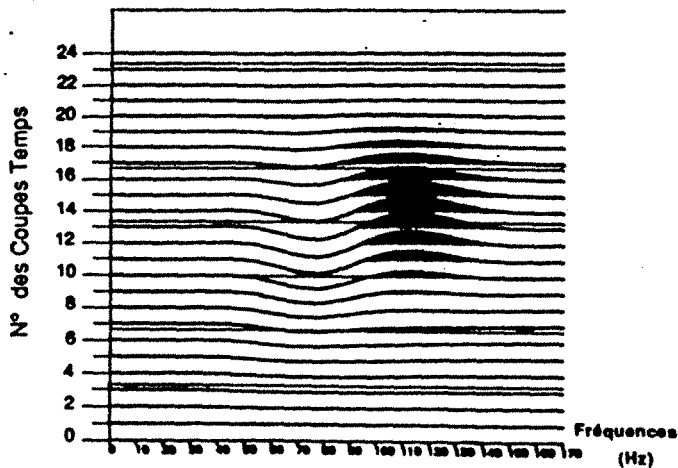


LOG VITESSE

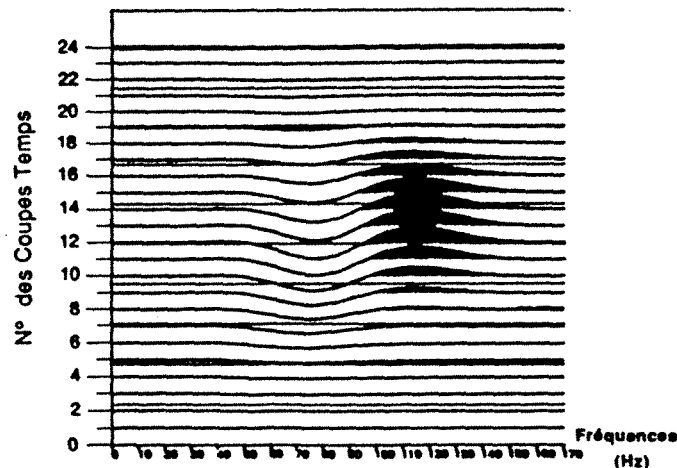
TRACES

IMPULSION
NELLE

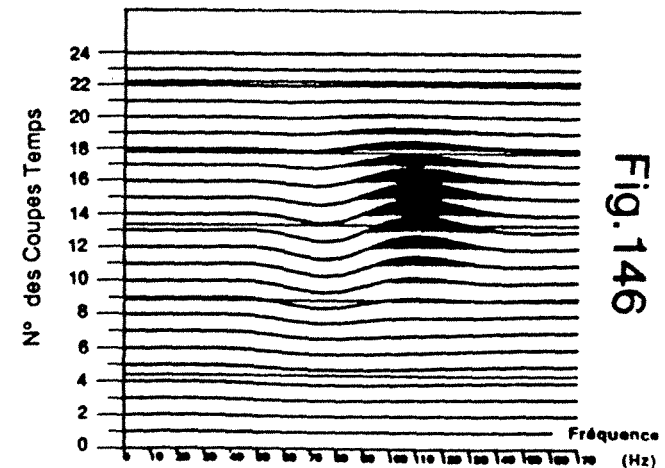
FILTREE



RWV1-RWV21



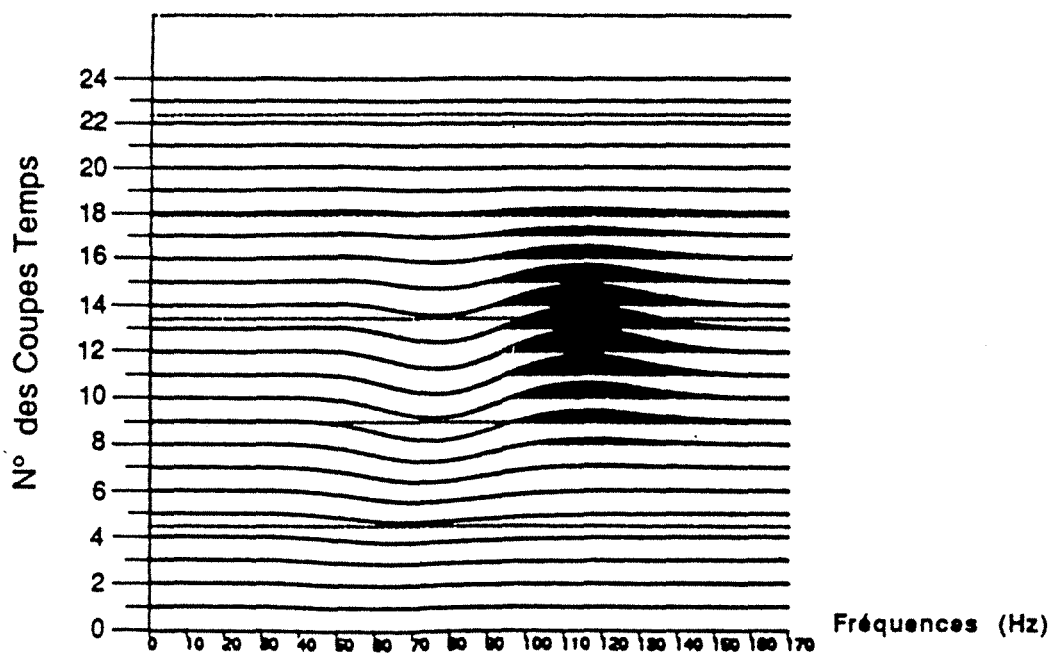
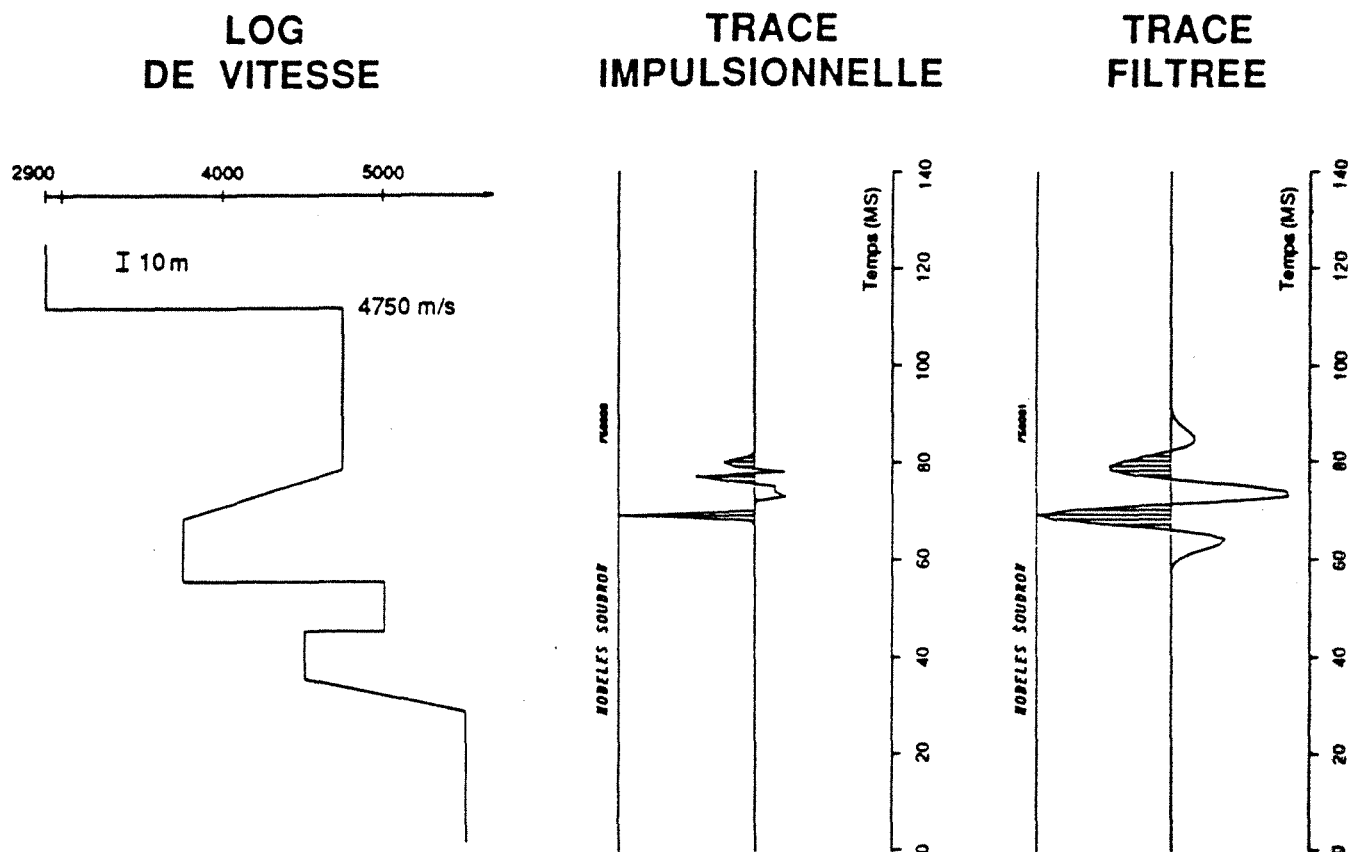
RWV1-RWV22



RWV1-RWV23

Fig. 146

COMPOSITION DES PATTERNS DE BASE (modèle 5)



DIFFERENCE DES RWV
NORMEES: RWV1-RWV5

Nous avons systématiquement calculé la différence des R.C.T.F. entre le modèle 1 et les traces synthétiques pour voir si nous retrouvions les formes caractéristiques trouvées en étudiant les différences entre les modèles.

La figure 148 montre les différences normées entre modèle 1 et puits à zone de transition (puits 110), entre modèle 1 et puits avec une base argileuse pour les calcaires bioclastiques (puits 102), entre modèle 1 et puits avec les deux "patterns" (puits 4).

Il est frappant de voir que les trois images se ressemblent beaucoup entre elles, mais ne ressemblent pas du tout aux images correspondant aux différences des modèles entre eux. Les valeurs des amplitudes des différences sont très grandes, l'écart entre les R.C.T.F. des traces synthétiques et la R.C.T.F. du modèle 1 est bien plus important que ceux entre les R.T.C.F. des divers modèles.

On ne retrouve pas du tout les images caractéristiques des faciès géologiques modèles, car la différence de caractère entre le modèle 1 et les traces synthétiques générées à partir des données de puits est considérable et "masque" tout le reste.

2.4.7. Conclusions à l'étude des modèles

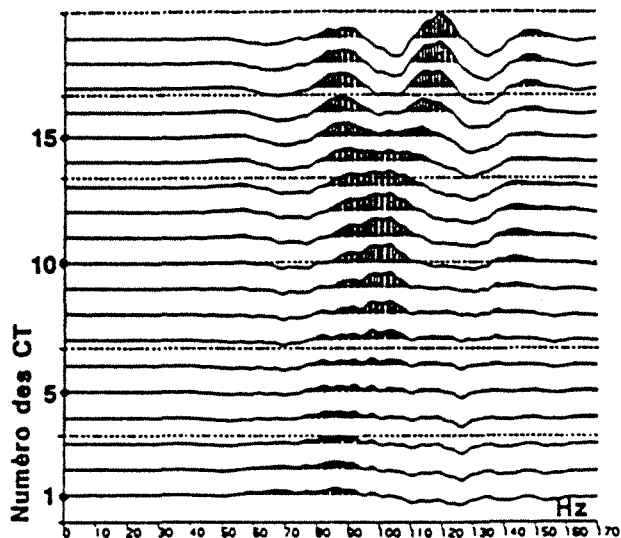
Cette étude confirme la très grande sensibilité de la représentation conjointe temps-fréquence à des modifications de l'ordre de $\lambda/10$ du modèle géologique.

S'il semblait prometteur de caractériser la R.C.T.F. de chaque "pattern" géologique détecté en réalisant un modèle synthétique du "pattern" et en l'analysant par une méthode différentielle (comparaison avec un modèle étalon), on se heurte à l'impossibilité de passer aux traces synthétiques.

Bien que les traces synthétiques soient très analogues aux traces modèles (ce qui justifie le choix des épaisseurs et des vitesses pour ces modèles), leurs R.C.T.F. sont impossibles à comparer à celles des modèles. **La méthode différentielle met en évidence plutôt l'écart entre tous les modèles "étalons" et les traces synthétiques.**

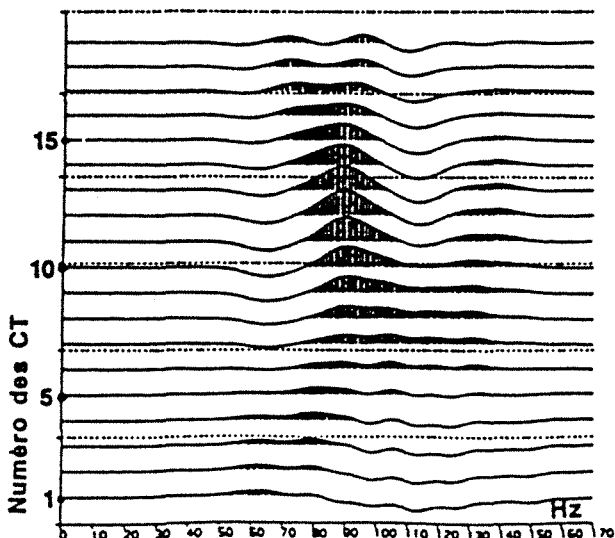
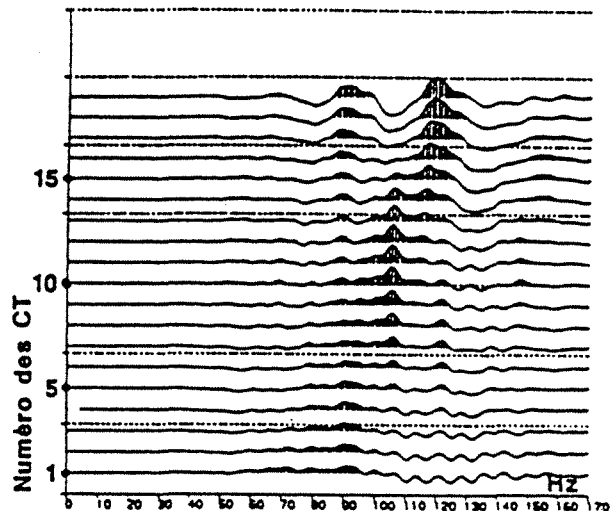
Trouver une image caractéristique du faciès "zone de transition" par une approche "modèle étalon" semble irréalisable. Il ne nous reste qu'à essayer de comparer les R.C.T.F. des traces synthétiques entre elles.

COMPARAISON MODELE DE BASE
ET TRACE SYNTHETIQUE AU PUIS



MODELE 1/ TRACE PUIS 110
(LE PUIS 110 PRESENTE LE
PATTERN "ZONE DE TRANSITION
OOS/AB")

MODELE 1/ TRACE PUIS 102
(LE PUIS 102 PRESENTE LE
PATTERN "BASE ARGILEUSE
POUR LES CB")



MODELE 1/ TRACE PUIS 4
(LE PUIS 4 PRESENTE
SIMULTANEMENT LES DEUX
PATTERNS)

2.4.8. Etude des R.C.T.F. des traces synthétiques

L'étude des traces synthétiques s'est limitée à celles calculées aux puits ayant permis de déterminer les classes de faciès géologiques. Nous avons éloigné l'analyse différentielle car les résultats précédents sont difficilement contrôlables. Nous avons préféré calculer directement, à partir des R.C.T.F., des paramètres afin d'essayer de caractériser les diverses coupes temps de la représentation.

Les maxima des différences sont localisés à partir des coupes temps n°8 ou 9 ; cela n'est pas surprenant car la R.C.T.F. de Wigner-Ville n'est que partiellement locale. Les R.C.T.F. ont été analysées sur les coupes temps n° 9 à 16 en calculant les paramètres suivants (Figure 149) :

- cinq premiers passages à zéro de la fonction $\rho(t, \nu)$,
- les deux plus grands maxima et leurs amplitudes correspondantes,
- les deux minima prédominants et leurs amplitudes.

2.4.8.1. Analyse en composantes principales des paramètres précédents

Les trois premiers facteurs expriment 81,6 % de l'inertie totale du nuage. L'axe 1 (49,9 %) exprime une certaine répartition des puits alors que l'axe 2 (19 %) traduit plutôt l'évolution des coupes $\rho(t, \nu)$ dans le temps (Figure 150).

La classification des puits que l'on pourrait obtenir sur l'axe 1 serait :

puits 4	puits 106	puits 1	puits 105
puits 7	puits 111	puits 102	puits 5
	puits 14	puits 110	
		puits 101	

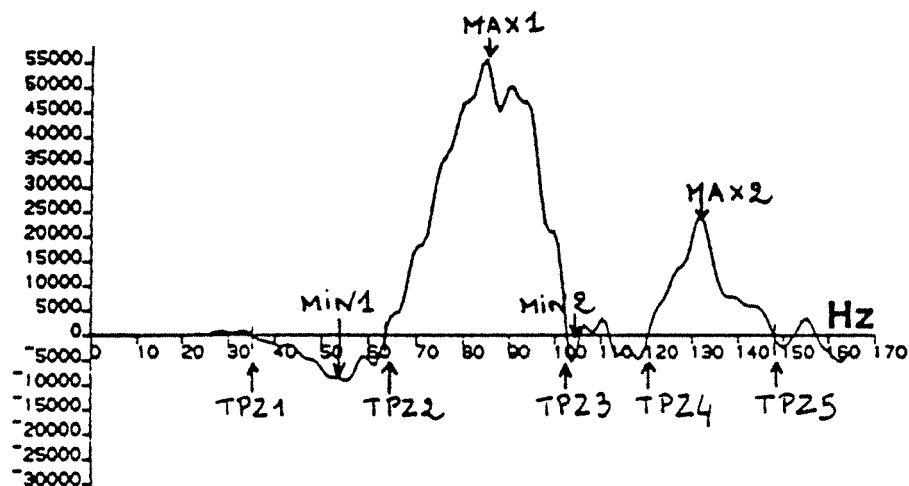
2.4.8.2. Analyse factorielle discriminante

L'idée est de répartir les puits en groupes a priori d'après les "patterns" géologiques reconnus aux puits et de voir s'il est possible d'ajuster un modèle discriminant linéaire à partir des variables calculées.

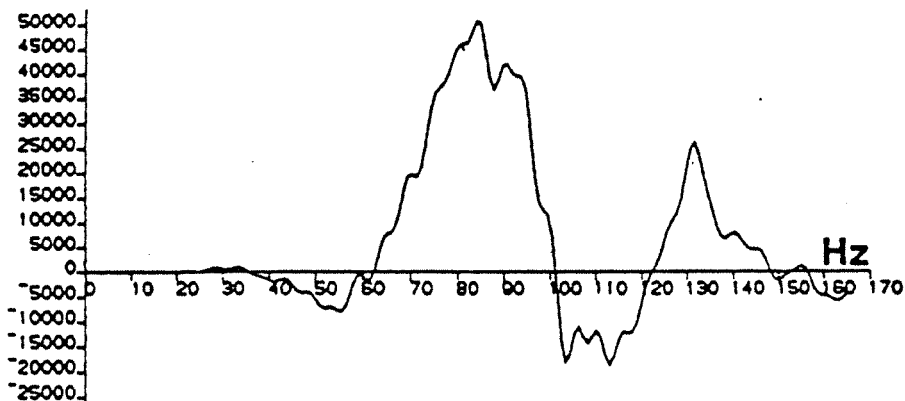
PARAMETRES MESURES SUR LES R.C.T.F.
DES TRACES AUX Puits

Ex: Puits 1

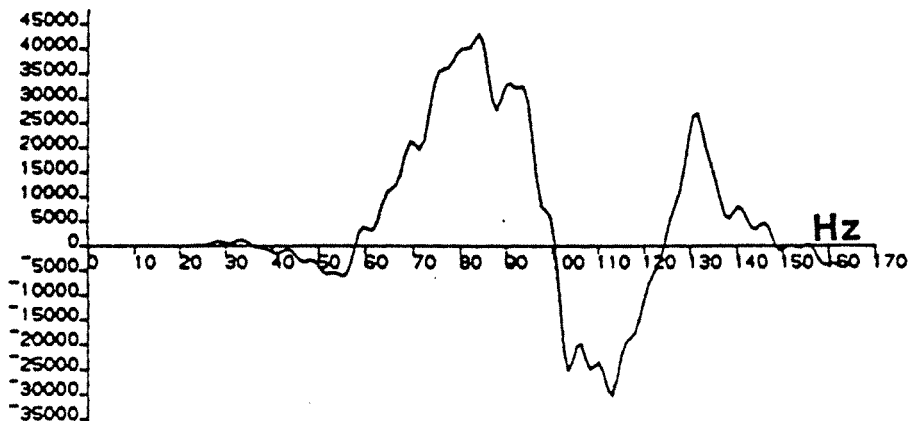
COUPE N° 10



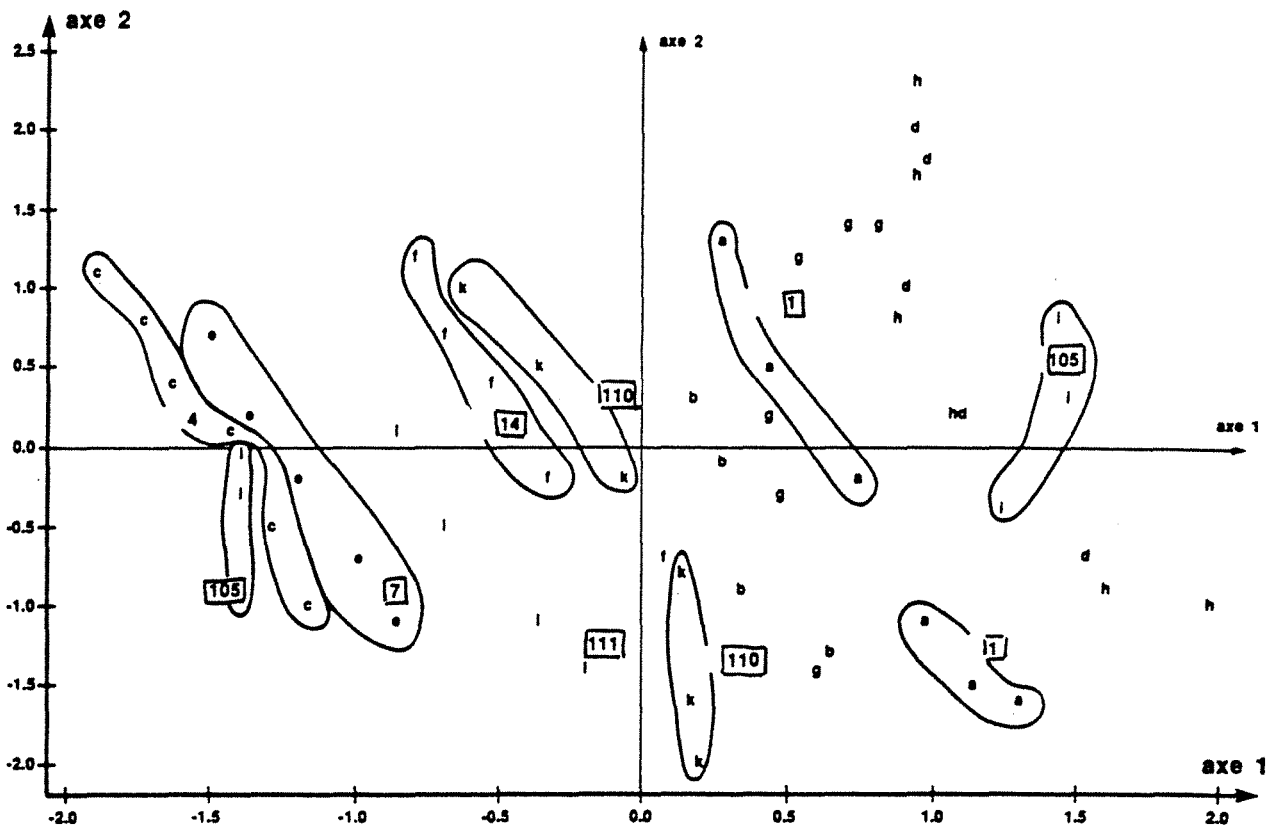
COUPE N° 11



COUPE N° 12



**ANALYSE EN COMPOSANTES PRINCIPALES
DES PARAMETRES MESURES
SUR LES R.C.T.F.**



PLAN PRINCIPAL 1-2

- *Argiles bioclastiques*

Les puits sont répartis en trois groupes : puits extrêmes atypiques (105 et 14), puits à zone de transition, puits sans zone de transition. Les variables calculées discriminent mal les groupes formés : pouvoir discriminant faible, modèle ajusté seulement à 50 %.

- *Calcaires bioclastiques*

Les puits sont répartis en d'autres groupes exprimant la classification au niveau des calcaires bioclastiques. Là encore, le modèle discriminant est mal adapté (le coefficient d'ajustement ne dépasse pas 44 %).

2.4.8.3. *Interprétation des résultats*

Les résultats sur les R.C.T.F. des traces synthétiques montrent qu'il n'est pas possible d'individualiser des caractéristiques liées aux argiles bioclastiques ou aux calcaires bioclastiques : on ne peut pas discriminer correctement les groupes formés par les différents "patterns".

Les groupes trouvés par l'analyse en composantes principales, paragraphe 2.4.8.1, ne sont pas interprétables. Il paraît clair que l'information interférentielle n'est pas utilisable si elle n'a pas été auparavant calée par une phase d'apprentissage sur des traces dont on contrôle l'interprétation géologique. Comment définir ces traces d'apprentissage ? L'idée de trouver des modèles étalons synthétiques n'a pas donné de résultats. Bien que ces modèles soient très proches des traces synthétiques générées avec les données de puits, ce qui confirme le choix des vitesses et des épaisseurs, leurs R.C.T.F. ne sont absolument pas comparables et donc impropres à la calibration.

La représentation de Wigner-Ville ne répond pas à l'objectif fixé : analyser l'information interférentielle **en liaison** avec les faciès géologiques détectés sur les logs. C'est un outil très fin pour détecter des variations (surtout si on adopte la méthode différentielle) mais contrôler la signification des évolutions du débit énergétique $\rho(t, \nu)$ est une autre affaire, et, à l'heure actuelle, sans étape d'apprentissage, on ne voit pas très bien comment on pourrait le réaliser.

2.5. Conclusion sur l'application de la R.W.V. pour la caractérisation fine de la trace

Si le physicien peut être optimiste quant aux potentialités de la R.W.V. (Bouachache, 1982, Bruel, 1987) l'avis du "géologue-géophysicien" est plus tempéré. Certes, l'outil est très

sensible et la méthodologie différentielle proposée permet de détecter des variations de l'ordre de $\lambda/10$. Par contre, pour l'extraction de nouveaux paramètres aidant à la caractérisation fine du faciès, on se heurte à deux obstacles :

- le premier est le caractère partiellement local de la R.C.T.F. (l'image est donc à interpréter partiellement globalement !!) ;
- le second est l'impossibilité de calibrer en terme de modèles géologiques les variations observées sur la R.C.T.F. Une faible dispersion autour d'un même "pattern" est à l'origine de nombreuses évolutions sur la R.C.T.F. et se traduisant globalement.

Nous avons donc abandonné cette voie et cherché à calculer les paramètres sismiques sur le **spectre de puissance** de façon **fiable sur une courte fenêtre**. Nous allons donc comparer les estimations spectrales fournies par la F.F.T. (transformée de Fourier rapide) et des méthodes type entropie maximale (censées donner une estimation plus robuste).

3. ANALYSE SPECTRALE PAR MAXIMISATION DE L'ENTROPIE

3.1. Position du problème

Les méthodes classiques d'analyse spectrale peuvent toutes se ramener à la procédure suivante :

- estimation de la fonction d'autocorrélation $C_{xx}(\tau)$,
- calcul de la densité spectrale $S_{xx}(\nu)$ à partir de $C_{xx}(\tau)$.

$$P(\lambda) = \sum_{k=-\infty}^{+\infty} C_{xx}(k) e^{-j\lambda k}$$

$P(\lambda)$ étant la fonction spectrale définie par :

$$P(\lambda) = T_e S_{xx}\left(\frac{\lambda}{2\pi T_e}\right)$$

avec T_e période d'échantillonnage du signal.

Mais il existe deux limitations :

- on ne dispose pas des valeurs exactes de la fonction de corrélation, mais seulement d'estimations (nous n'aborderons pas ce problème) ;

- la fonction de corrélation n'est connue que sur un **support limité**. Il est donc nécessaire de la **prolonger** au-delà. La méthode classique par transformation de Fourier revient à prolonger la fonction d'autocorrélation en périodisant le signal sismique à l'extérieur du support à partir des valeurs connues. Ceci entraîne un élargissement des raies (perte de résolution fréquentielle) ainsi que l'apparition de lobes secondaires. Ce problème de lobes peut être atténué par l'application de fenêtres d'apodisation mais on modifie ainsi les valeurs connues de $C_{xx}(k)$, ce qui est peu satisfaisant.

Ces problèmes sont d'autant plus sensibles que le nombre de points connus est petit.

Ainsi, sur 100 ms au pas de 2 ms (50 points), la règle du pouce $\Delta F = \frac{1}{\Delta T}$ (où ΔT est la largeur de la fenêtre temporelle) donne pour résolution fréquentielle $\Delta F = 10$ Hz.

D'autre part, on peut observer un déplacement des pics spectraux.

L'estimation des caractéristiques spectrales donnée par les paramètres de la méthodologie proposée pour la reconnaissance du faciès est donc certainement très mauvaise et peu fiable **dans l'absolu**.

Pratiquement, les bons résultats obtenus avec ces paramètres (dans plusieurs cas d'application, ils sont intervenus de façon significative pour assurer la discrimination entre les traces d'apprentissage) ne peuvent s'expliquer que par la démarche comparative utilisée : **comparaison** de la densité spectrale évaluée sur deux fenêtres de **taille identique**. Les différences éventuelles trouvées, malgré le mode d'évaluation frustré, traduisent une évolution réelle des caractéristiques de la trace.

A ce stade de la recherche, il est toutefois séduisant d'essayer les nouvelles méthodes d'analyse spectrale abondamment décrites et testées par les géophysiciens. Ces méthodes, toutes basées sur le principe de la maximisation de l'entropie, proposent une extension de la fonction d'autocorrélation compatible avec les données connues et exigeant le minimum d'hypothèses quant aux valeurs non observées de cette fonction.

Des tests sur des données synthétiques (sommées de sinusoides amorties) confirment l'intérêt de l'approche pour améliorer la résolution spectrale dans le cas de **courtes fenêtres**. Par contre, l'accord entre les auteurs se fait moins quant au choix de l'algorithme à utiliser.

Les "nouveaux" paramètres ainsi obtenus, outre qu'ils donneraient une estimation spectrale plus fiable, pourraient être utilisés directement comme vecteur attribut dans un algorithme de reconnaissance des formes. Bois (1980, 1981, 1982) a réalisé avec un certain succès une étude sur la délimitation de zones homogènes au sein d'un réservoir en représentant chaque portion

Notre étude a donc pour objectif de tester les différents algorithmes proposés pour l'analyse spectrale par la méthode de l'entropie maximale, sur des données synthétiques, de comparer les résultats avec ceux donnés par la transformation de Fourier afin d'évaluer le gain en résolution pour un petit nombre de points ainsi que le nombre de coefficients du modèle à choisir et l'algorithme le plus efficace. Après ces tests, nous avons appliqué la méthode à des données sismiques synthétiques pour un problème géologique très fin. Cette application sera discutée dans le chapitre 5.

Avant d'exposer les résultats obtenus, nous ferons une brève synthèse bibliographique sur l'analyse spectrale par la méthode de l'entropie maximale (M.E.M.) et résumerons les différences entre les algorithmes testés, du point de vue des calculs et des hypothèses sous-jacentes.

3.2. Méthode de l'entropie maximale

L'analyse spectrale est une technique fondamentale pour le géophysicien. Pendant longtemps, la transformation de Fourier a été le seul outil utilisé pour passer du domaine temporel au domaine fréquentiel. A la fin des années soixante, deux méthodes non linéaires d'estimation spectrale ont été introduites : la méthode de la vraisemblance maximale (M.V.M.) (Jenkins et Watts, 1968) et la méthode de l'entropie maximale (M.E.M.) (Burg, 1967). Toutes deux sont particulièrement intéressantes pour obtenir une résolution fréquentielle fine lorsque le nombre de données est faible.

La relation entre M.V.M. et M.E.M. a été discutée par plusieurs auteurs (Burg, 1972, Childers et Chen, 1981) : M.V.M. est la moyenne des estimations de M.E.M., pour les différents ordres M du modèle autorégressif. Par conséquent, elle est statistiquement plus stable mais la résolution fréquentielle est moins bonne. Dans ce rapport, nous ne discutons que de la méthode de l'entropie maximale (M.E.M.). Pour la méthode de la vraisemblance maximale, les lecteurs intéressés peuvent se reporter à l'ouvrage de Jenkins et Watts.

La méthode de l'entropie maximale est appelée également analyse spectrale par le modèle autorégressif (ASMA).

3.2.1. Méthode de Yule-Walker

Cette estimation spectrale par représentation de la série temporelle au moyen d'un modèle autorégressif a été suggérée pour la première fois par Yule (Ulrych, 1972, Kalouptsidis et Theodiridis, 1987). Nous rappelons que le modèle autorégressif s'écrit :

$$Y_n = \sum_{k=1}^M a_{M,k} Y_{n-k} + X_n \quad (I-1)$$

où Y_n est la fonction sortie ; X_n , la fonction entrée, est une variable aléatoire, en général un bruit blanc, appelée innovation du processus autorégressif, de moyenne nulle et de variance σ_x^2 , et les $a_{M,k}$ sont les paramètres du modèle autorégressif d'ordre M.

En multipliant par Y_{n-i} les deux membres de l'équation (I-1) et en appliquant l'opérateur espérance mathématique, on obtient facilement les équations de Yule-Walker :

$$\sum_{k=1}^M a_{M,k} \phi_{yy}(k-i) = \phi_{yy}(i) \quad \text{pour } 1 \leq i \leq M \quad (\text{I-2})$$

$$\sum_{k=1}^M a_{M,k} \phi_{yy}(k) = \sigma_x^2 + \phi_{yy}(0) \quad \text{pour } i = 0 \quad (\text{I-2)bis}$$

Les équations (I-2) et (I-2)bis peuvent se mettre sous la forme matricielle suivante :

$$\begin{bmatrix} \phi_{yy}(0) & \phi_{yy}(1) & \dots & \phi_{yy}(M) \\ \phi_{yy}(1) & \phi_{yy}(0) & \dots & \phi_{yy}(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{yy}(M) & \phi_{yy}(M-1) & \dots & \phi_{yy}(0) \end{bmatrix} \begin{bmatrix} 1 \\ -a_{M,1} \\ \vdots \\ -a_{M,M} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{I-3})$$

Il s'agit d'une matrice définie positive avec une structure de Toeplitz. La fonction d'autocorrélation est calculée le plus souvent par l'estimateur biaisé (Jenkins et Watts, 1968) :

$$\phi_{yy}(i) = \frac{1}{N} \sum_{p=1}^{N-i} (Y_{p+i})(Y_p) \quad (\text{I-4})$$

car il assure la propriété de définie positive.

La transformation en Z de la formule (I-1) nous donne :

$$Y(Z) = \frac{X(Z)}{1 - \sum_{k=1}^M a_{M,k} Z^k} \quad (I-5)$$

Nous prenons $Z = \exp[-2\pi i f \Delta t]$ et calculons le carré du module de la relation (I-5), nous avons ainsi le spectre de puissance pour la fonction Y_n :

$$S_y(f) = \frac{\sigma_x^2}{|1 - \sum_{k=1}^M a_{M,k} \exp(-2\pi i k f \Delta t)|^2} \quad (I-6)$$

où Δt est le pas temporel et la fréquence f est limitée à l'intervalle de Nyquist :

$$-1/(2\Delta t) \leq f \leq 1/(2 \Delta t).$$

Les équations normales de Yule-Walker (I-3) nous permettent de calculer les coefficients autorégressifs $a_{M,k}$, la variance de l'innovation σ_x^2 et ainsi d'estimer le spectre de puissance selon la formule (I-6).

3.2.2. Principe de l'entropie maximale

Soit un système susceptible de passer par N états différents de probabilité d'occurrence π_i . Si nous supposons a priori que les π_i sont tous égaux, nous ne disposons alors d'aucune information spéciale concernant le système. D'un autre côté, si nous pouvons préciser une valeur particulière pour π_i , alors nous gagnons une certaine quantité d'information concernant le système. Plus l'événement pour lequel nous avons spécifié une probabilité est inattendu, plus sa réalisation apportera d'information.

La relation liant quantité d'information I_i et probabilité p_i est la suivante :

$$I_i = k \ln \frac{1}{p_i}$$

où k est une constante.

L'information moyenne par unité de temps contenue dans un système est appelée **entropie**. Elle est définie par :

$$H = -k \sum_{i=1}^N p_i \ln p_i \quad (\text{I-7})$$

où N est le nombre total des événements possibles.

Ainsi l'entropie est une mesure du désordre d'un système ou bien une mesure de notre ignorance.

Van Der Bos a établi la liaison entre l'analyse spectrale par la méthode de l'entropie et la représentation autorégressive d'un processus aléatoire. En particulier, pour un processus gaussien stationnaire discret, l'entropie est donnée par (Bois, 1981) :

$$H = \frac{1}{2} \ln |\Phi_N| \quad (\text{I-8})$$

où $|\Phi_N|$ est le déterminant de la matrice d'autocorrélation de Toeplitz.

Par conséquent, maximiser l'entropie d'un processus revient à estimer les valeurs d'autocorrélation d'indice supérieur à N de façon que le déterminant de la matrice de Toeplitz soit maximum à chaque pas. Le problème soulevé avec les méthodes conventionnelles d'estimation spectrale est lié aux hypothèses faites sur les données à l'extérieur du support d'observation. En maximisant l'entropie du processus, on cherche à être consistant avec les données connues (N valeurs de l'autocorrélation) tout en faisant le minimum de suppositions possibles sur les données inaccessibles. Le calcul de la valeur $\Phi(N+1)$ se fait donc à partir des valeurs connues $\Phi(0), \Phi(1), \dots, \Phi(N)$ en maximisant $|\Phi(N+1)|$, le calcul de $\Phi(N+2)$ à partir de $\Phi(1), \Phi(2) \dots \Phi(N+1)$, et ainsi de suite. L'unicité de la solution (un seul maximum) est assurée par le fait que la matrice de Toeplitz est définie non négative.

Montrons maintenant que le principe de l'entropie maximale coïncide bien avec le modèle autorégressif. La matrice de Toeplitz pour un processus autorégressif d'ordre M+1 s'écrit :

$$\Phi_{(M+1)} = \begin{bmatrix} \Phi(0) & \Phi(1) & \dots & \Phi(M+1) \\ \Phi(1) & \Phi(0) & \dots & \Phi(M) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(M+1) & \Phi(M) & \dots & \Phi(0) \end{bmatrix} \quad (\text{I-9})$$

La dérivée de son déterminant par rapport à l'inconnue $\Phi(M+1)$ qui se situe aux deux extrémités de la diagonale non principale s'écrit :

$$\frac{\partial |\Phi_{M+1}|}{\partial \Phi(M+1)} = \det \begin{bmatrix} \Phi(1) & \Phi(0) & \dots & \Phi(M-1) \\ \Phi(2) & \Phi(1) & \dots & \Phi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(M+1) & \Phi(M) & \dots & \Phi(1) \end{bmatrix}$$

La solution de $\Phi(M+1)$ est obtenue par :

$$\det \begin{bmatrix} \Phi(1) & \Phi(0) & \dots & \Phi(M-1) \\ \Phi(2) & \Phi(1) & \dots & \Phi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(M+1) & \Phi(M) & \dots & \Phi(1) \end{bmatrix} = 0 \quad (\text{I-10})$$

Par ailleurs, reprenons l'équation normale (I-2) d'un processus autorégressif d'ordre M. Supposons connues les (M+1) première valeurs $\Phi(0) \dots \Phi(M)$ et cherchons à estimer $\Phi(M+1)$.

$$\Phi(M+1) = \sum_{k=1}^M a_{M,k} \Phi(M+1-k)$$

On remplace (I-2)bis par cette dernière expression et la forme matricielle (I-3) devient :

$$\begin{bmatrix} \Phi(1) & \Phi(0) & \dots & \dots & \dots & \Phi(M-1) \\ \Phi(2) & \Phi(1) & \dots & \dots & \dots & \Phi(M-2) \\ \vdots & \vdots & \dots & \dots & \dots & \vdots \\ \Phi(M+1) & \Phi(M) & \dots & \dots & \dots & \Phi(1) \end{bmatrix} \begin{bmatrix} 1 \\ -a_{M,1} \\ \vdots \\ -a_{M,M} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (I-11)$$

L'unicité de la solution est assurée par le fait, d'une part que nous voulons que la matrice d'autocorrélation $[\Phi]$ soit définie positive, d'autre part qu'il y a autant d'équations que d'inconnues. Ceci étant, la détermination du coefficient $\Phi(M+1)$ se fait alors en annulant le déterminant de la matrice du premier membre. C'est précisément l'expression (I-10).

Ainsi nous concluons que l'analyse spectrale d'une série temporelle par le principe de l'entropie maximale est équivalente à l'analyse spectrale par la modélisation d'un processus autorégressif en prenant le critère des moindres carrés.

3.2.3. Détermination de l'ordre du modèle autorégressif

Le choix de l'ordre du modèle autorégressif est très important pour l'estimation du spectre par la méthode de l'entropie maximale. Lorsqu'on choisit un ordre trop petit, la résolution spectrale n'est souvent pas suffisante, au contraire, un ordre trop grand crée des artefacts importants (pics supplémentaires). Plusieurs auteurs ont contribué à la recherche de l'ordre optimal du modèle autorégressif (Jenkins et Watts, 1968, Ulrych et Bishop, 1975, Akaike, Berryman,

1978, Rovelli, 1982, Wax, 1988). Certains ont montré que le choix empirique de l'ordre $M = N/2$, ou le cas échéant, $0,3 N < M < 0,5 N$, s'adapte à la plupart des cas. Par ailleurs, Berryman (1978) suggère a priori le choix de $M = 2N/\ln 2N$.

Akaike a établi un critère formel par l'intermédiaire du calcul de l'erreur de prédiction finale. Cette erreur est définie comme la moyenne quadratique de l'erreur de prédiction d'un processus stochastique par un modèle autorégressif. L'expression la plus courante de cette erreur de prédiction est donnée par :

$$e = \frac{N + (M + 1)}{N - (M + 1)} P_M \quad (I-12)$$

où N est le nombre d'échantillons et P_M la variance des résidus.

Quand M croît, la quantité P_M décroît alors que $\frac{N + (M + 1)}{N - (M + 1)}$ croît. La valeur optimale de M correspond au minimum de l'erreur de prédiction finale donnée par l'expression (I-12).

Récemment, Wax (1988) a présenté un nouveau critère pour la sélection de l'ordre du modèle autorégressif. Ce critère est basé sur la prédiction par la méthode des moindres carrés en utilisant les filtres prédictifs en treillis. Le principe est le suivant : l'ordre doit être sélectionné de telle sorte que la somme des erreurs de prédiction, selon toute la séquence de données, soit minimale. Formellement, on choisit un ordre M qui correspond à :

$$\text{avec} \quad S(m) = \frac{1}{N} \sum_{t=1}^N e_{m,t}^2 \quad \begin{matrix} \min S(m) \\ m \leq N \end{matrix}$$

$$\text{et} \quad e_{m,i} = Y_i - \sum_{j=1}^m a_j Y_{i-j}$$

Enfin, une approche qui tient compte de la corrélation a été présentée par Rovelli (1982). En effet, plus la corrélation d'une série temporelle est importante, plus l'ordre du modèle autorégressif est grand. Dans le cas où on a un bruit blanc, la corrélation entre les différents échantillons est nulle et l'ordre est bien évidemment zéro. Par ce principe, la valeur optimale de M est déterminée par :

$$M = \frac{\sum_{t=1}^{M_{\max}} |\Phi(t)|}{\Phi(0)}$$

où $\Phi(t)$ est la fonction d'autocorrélation.

La démonstration mathématique de cette expression et la détermination de M_{\max} n'ont pas été données par Rovelli : ceci reste donc à faire pour utiliser ce dernier critère.

3.3. Algorithmes d'estimation des paramètres autorégressifs

Dans le modèle autorégressif, le spectre de puissance est calculé par l'expression (I-6). Pour un ordre M du modèle autorégressif donné, il s'agit de déterminer les coefficients $a_{M,k}$ et la variance de l'innovation σ_x^2 . Il existe plusieurs algorithmes pour déterminer ces paramètres que nous présentons brièvement.

3.3.1. Algorithme de Yule-Walker

La méthode de Yule-Walker calcule les coefficients autorégressifs en utilisant les valeurs expérimentales de l'autocorrélation (cf. (I-4)) pour les $M + 1$ premiers pas.

Puis la solution des $\{a_{M,k}\}$ et de σ_x^2 est déterminée par la formule matricielle (I-3) ($M+1$ inconnues et $M+1$ équations). L'unicité de la solution est assurée par le fait que la matrice d'autocorrélation est définie positive.

En pratique, l'algorithme de Levinson est très souvent utilisé. Il consiste à calculer récursivement les coefficients autorégressifs pour l'ordre M à partir des coefficients autorégressifs précédemment déterminés pour l'ordre $M-1$ grâce à la formule suivante (Levinson, 1947) :

$$a_{M,k} = a_{M-1,k} - a_{M,M} a_{M-1,M-k} \quad (\text{II-1})$$

$$\forall k = 1, 2, \dots, M-1$$

Dans l'algorithme original de Levinson, le spectre de l'innovation P_M est calculé par une formule non récursive, mais on peut également utiliser la formule récursive :

$$P_M = P_{M-1} (1 - a_{M,M}^2).$$

Cette dernière expression présente deux avantages. D'une part, elle réduit considérablement le temps de calcul par rapport à la formule non récursive. D'autre part, elle assure une meilleure stabilité numérique, en particulier pour un ordre du modèle autorégressif important.

3.3.2. Algorithme de Burg

L'algorithme de Yule-Walker exige une estimation préalable des (M+1) premières valeurs de l'autocorrélation. Burg a proposé un algorithme pour estimer directement les coefficients autorégressifs à partir des échantillons donnés. Basée sur la notion de prédiction linéaire avant et arrière, cette méthode consiste à minimiser la somme des carrés des erreurs de prédiction dans les deux sens.

Dans le modèle autorégressif, l'erreur de la prédiction linéaire avant pour un point k+M est donnée par :

$$f_{M,k} = -Y_{k+M} + \sum_{j=1}^M a_{M,j} Y_{k+M-j} = \sum_{j=0}^M a_{M,j} Y_{k+M-j} = \sum_{j=0}^M a_{M,M-j} Y_{k+j} \quad (\text{II- 2})$$

où $a_{M,0} = -1$

et l'erreur de la prédiction linéaire arrière pour un point k est donnée par :

$$b_{M,k} = -Y_k + \sum_{j=1}^M a_{M,j} Y_{k+j} = \sum_{j=0}^M a_{M,j} Y_{k+j} = \sum_{j=0}^M a_{M,M-j} Y_{k+M-j} \quad (\text{II- 3})$$

Ainsi la somme des erreurs quadratiques de ces deux prédictions sur tous les échantillons s'écrit :

$$e_M = \sum_{k=1}^{N-M} (|f_{M,k}|^2 + |b_{M,k}|^2) \quad (\text{II- 4})$$

La méthode de Burg consiste à minimiser cette quantité en utilisant l'algorithme récursif de Levinson.

3.3.3. *Algorithme de Barrodale-Erickson*

Dans l'algorithme de Burg, on minimise la somme exprimée par (II-4) en utilisant la récursion de Levinson (II-1). Ceci impose une structure sous-jacente de Toeplitz dans le modèle (bien que l'on ne calcule pas explicitement la matrice d'autocorrélation). C'est en ce sens que cet algorithme est appelé méthode des moindres carrés sous contrainte. En utilisant la méthode de Cholesky pour décomposer la matrice d'autocorrélation, Barrodale et Erickson (1980) ont proposé un algorithme des moindres carrés **sans contrainte**.

3.3.4. *Algorithme de Marple*

Dans la méthode de Barrodale-Erickson, les équations normales sont résolues en utilisant l'algorithme de Choleski sans faire référence à la solution pour l'ordre précédent. L'algorithme proposé par Marple calcule récursivement les paramètres autorégressifs d'ordre M à partir des paramètres autorégressifs d'ordre $M-1$. La récursion s'effectue par la décomposition des matrices des équations normales en somme de deux matrices de Toeplitz, au lieu d'imposer directement la récursion de Levinson (II-1) sur les expressions (II-2) - (II-4).

3.3.5. *Récapitulation sur les algorithmes*

La méthode de Yule-Walker est la première à avoir été utilisée pour calculer les paramètres autorégressifs. Pour résoudre les équations normales, on doit préalablement calculer la fonction d'autocorrélation. L'algorithme de Levinson permet de résoudre le système d'équations récursivement pour un ordre du modèle autorégressif donné.

Les autres méthodes (Burg, Barrodale et Erickson, Marple) font appel à la prédiction mixte (avant et arrière).

La méthode de Burg minimise la somme des erreurs quadratiques sous la contrainte de l'utilisation de la récursion de Levinson ; il en résulte un temps de calcul faible.

Les méthodes de Barrodale-Erickson et de Marple sont également récursives. Cependant, la récursion de Barrodale-Erickson n'a pas pour objectif direct la résolution du système d'équations normales pour les différents ordres du modèle autorégressif mais le calcul des coefficients de l'autocorrélation. Autrement dit, pour résoudre un système d'un modèle autorégressif d'ordre $M+1$, on n'utilise pas les résultats de la solution du système correspondant à l'ordre M . Ceci étant, le temps de calcul est souvent long, en particulier pour un ordre important.

Dans le modèle de prédiction mixte (avant et arrière), la matrice d'autocorrélation a des caractères particuliers exploités dans l'algorithme de Marple. En introduisant des formules de translation temporelle, la décomposition de cette matrice en deux matrices de Toeplitz permet d'utiliser à nouveau la récursion de Levinson avec des coefficients correctifs. D'une certaine

façon, la méthode de Marple est l'algorithme de Levinson généralisé. Ainsi, le temps de calcul est du même ordre que celui de la méthode de Burg.

Nous avons comparé ces divers algorithmes entre eux sur des fonctions simples dont le spectre était connu.

3.4. Essais sur des traces synthétiques

Les traces synthétiques que nous avons générées sont des sommes de sinusoides de fréquences et phases données, tronquées par une fenêtre de longueur variable. Le spectre de ces fonctions est parfaitement connu, ce qui permet de tester les divers algorithmes entre eux et de comparer avec les résultats d'une transformée de Fourier rapide (F.F.T.).

Ces résultats sont synthétisés dans le tableau 22. Les figures 151 à 158 illustrent quelques-uns des résultats regroupés dans le tableau. Pour résumer ce tableau, on peut dire que l'algorithme de Marple est de loin le plus performant. Il est moins sensible au bruit, au déphasage, permet d'obtenir une bonne résolution spectrale sur un faible nombre de points et ceci pour des ordres relativement peu élevés. C'est aussi lui qui semble le plus robuste par rapport aux artefacts (type "splitting").

L'application des algorithmes de la méthode de l'entropie maximale aux différentes combinaisons de sinusoides et la comparaison de cette technique avec la F.F.T. nous permet de tirer les conclusions suivantes.

- La méthode autorégressive est une meilleure technique d'estimation spectrale que la F.F.T. sur une courte fenêtre temporelle. Le signal en dehors de la fenêtre est implicitement supposé périodique pour la F.F.T., ceci provoque le phénomène de Gibbs et limite la résolution fréquentielle à l'inverse de la longueur de la fenêtre. La méthode de l'entropie maximale fait implicitement une prédiction ou extension à l'extérieur de la fenêtre, la série temporelle est donc en principe infinie. La résolution fréquentielle est fonction du rapport signal sur bruit et de l'ordre du modèle autorégressif. Marple a proposé la formule suivante :

$$\Delta f = \frac{6.47}{2\pi M \Delta t [(M+1) \gamma]^{0.31}}$$

où γ est le rapport signal sur bruit (non exprimé en dB), M , l'ordre du modèle autorégressif. Ceci étant, moins un signal est bruité, meilleure sera la résolution. Pour un signal peu bruité, γ est très grand, le spectre est parfaitement résolu.

**TABLEAU 22a : PRINCIPAUX RESULTATS DES TESTS AVEC LA M.E.M.
1 COSINUS DE FREQUENCE 50 HZ
O. : ORDRE**

Nombre d'échantillons	400	40	40 (avec déphasage)	40 (avec 33 % de bruit blanc)
Algorithme				
Yule-Walker	O. de 10 à 20 : 1 pic O. optimal (Akaike) = 14	O. = 2 : pic très élargi O. = 7 : décalage du pic de 4,5 Hz O. = 15 : 1 pic sur 50 Hz O. = 25 : "splitting"		O. = 7 : 1 pic très élargi et décalé de 8.5 Hz O. = 15 : décalage de 1,5 Hz + "splitting" O. = 25 : le "splitting" augmente
Burg	O. de 2 à 25 : 1 pic O. optimal (Akaike) = 13	O. = 2 : 1 pic étroit mais décalé de 1 Hz O. = 7 : pic décalé de 5 Hz O. = 15 : "splitting" O. = 25 : optimal, moins de "splitting" et pic central sur 49,5 Hz	O. = 2 : variation des coefficients autorégressifs ainsi que de la variance de l'innovation	O. = 7 : 1 pic élargi, décalage de 3 Hz O. = 15 : décalage de 1,5 Hz
Marple	Dès O. = 2 : 1 pic très bien résolu	O. de 2 à 25 : 1 pic très étroit sur 50 Hz, pas de décalage, pas de "splitting"	O. = 2 : aucun changement dans les coefficients autorégressifs	O. = 7 : 1 pic élargi, décalage de 3 Hz O. = 15 : 1 pic très étroit, décalage de 0,5 Hz, pas de "splitting" O. = 25 : "splitting" intense
Barrodale-Erickson	Numériquement moins robuste			
FFT	1 pic sur 50 Hz, mal défini + lobes	1 pic très élargi mais non décalé, beaucoup de lobes		

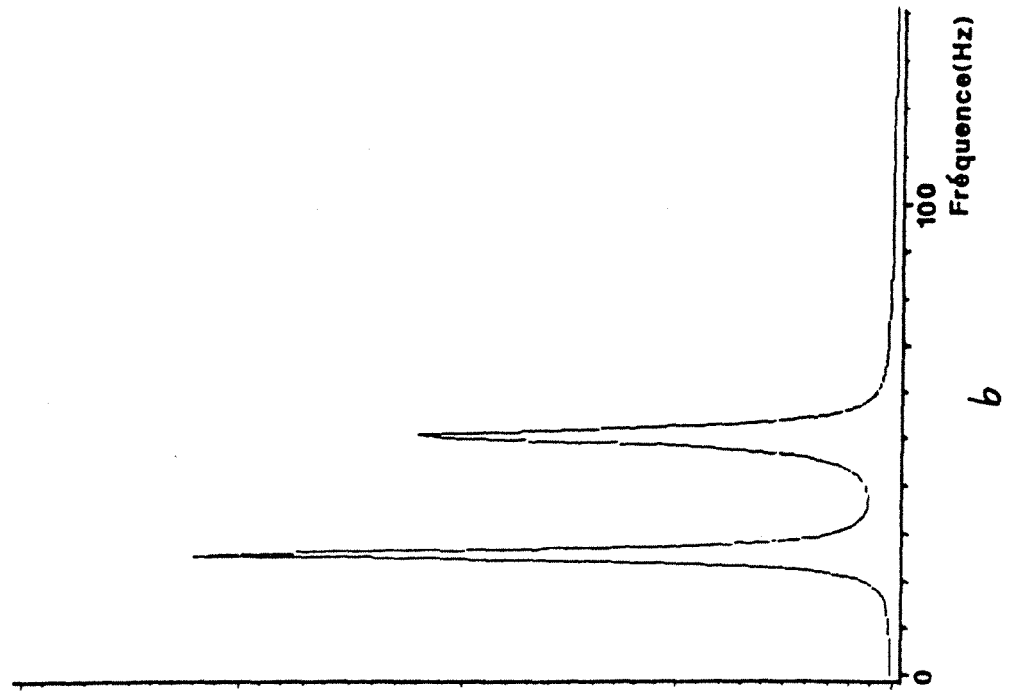
**Tableau 22 b : Principaux résultats des tests avec la M.E.M.
2 ou 3 cosinus**

Fonction Algorithme	2 cosinus : 25 Hz } + 0,25 % bruit blanc				2 cosinus : 35 Hz } + 0,25 % bruit blanc		3 cosinus : 35 Hz } + 0,25 % de bruit blanc	
	50 Hz				40 Hz		45 Hz } 50 Hz }	
Echantillons	15	25	50	100	25	50	50	100
Yule-Walker	0.= 7 : 1 pic entre 25 et 50 Hz		0.=7:1 pic entre 25 et 50 Hz 0.=15:2 pics 25 et 50,5 Hz	0. = 7 : 1 pic intermédiaire				
Burg	0.=7: décalage de 3,5 Hz (1er pic) et de 1,5 Hz (2ème pic)	0.=7: décalage de 1,5 Hz (1er pic) et 1 Hz (2ème pic)	0. = 7 : 2 pics 26,5 et 50 Hz 0. = 15 : 2 pics 25 et 50,5 très étroits	0. = 7 : 2 pics avec un décalage <1 Hz				
Marple	0. = 7 : 2 pics sans décalage	0. = 7 : 2 pics 25,5 et 50 Hz	0. = 7 : 2 pics 25,5 et 50 Hz 0. = 15 : 2 pics très étroits sur 25 et 50 Hz 0. = 2 : 1 pic sur 39 Hz 0. = 5 : 2 pics sur 28 et 47 Hz	0. = 7 : 2 pics sur 25,5 et 50 Hz	0. = 15 : pas de "splitting" mais décalage de 1,5 Hz (1er pic) et de 1 Hz (2ème pic)	0. ≤ 7 : 1 pic entre 35 et 40 Hz 0. = 15 : 2 pics aux bonnes fréquences pas de "splitting"	0. = 2 : 1 pic à 44 Hz 0. = 10 : 2 pics à 35,5 et 48,5 Hz 0. = 30 : 3 pics sur 35,5, 44,5 et 43,5 Hz	0. = 30 : 3 pics sur 35, 45 et 50 Hz, pas de "splitting"
FFT	décalage de 3 Hz (1er pic) et 8 Hz (2ème pic) + lobes	décalage de 2 Hz (2ème pic) + lobes	2 pics bien positionnés	2 pics bien positionnés	les deux pics ne sont pas séparés	les deux pics ne sont pas séparés	2 pics principaux sur 34 et 48 Hz	3 pics différenciés, mais légèrement décalés + lobes

SPECTRE DE PUISSANCE
ALGORITHME DE YULE-WALKER

N=50 , $\Delta t=2ms$

M=15



M=7

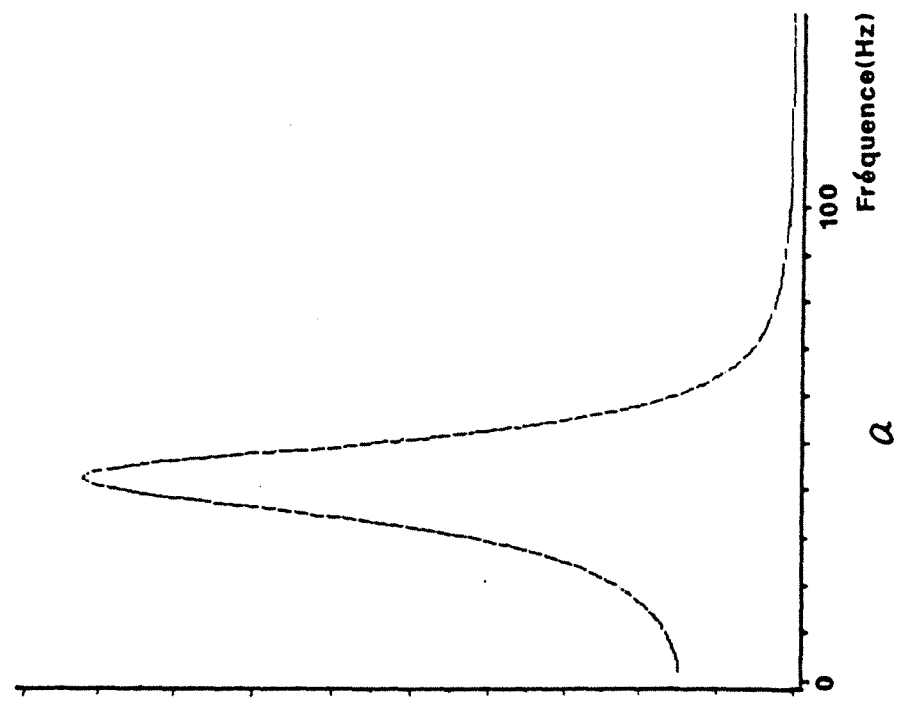


Fig.151

SPECTRE DE PUISSANCE

ALGORITHME DE BURG

$N=50$, $\Delta t=2\text{ms}$

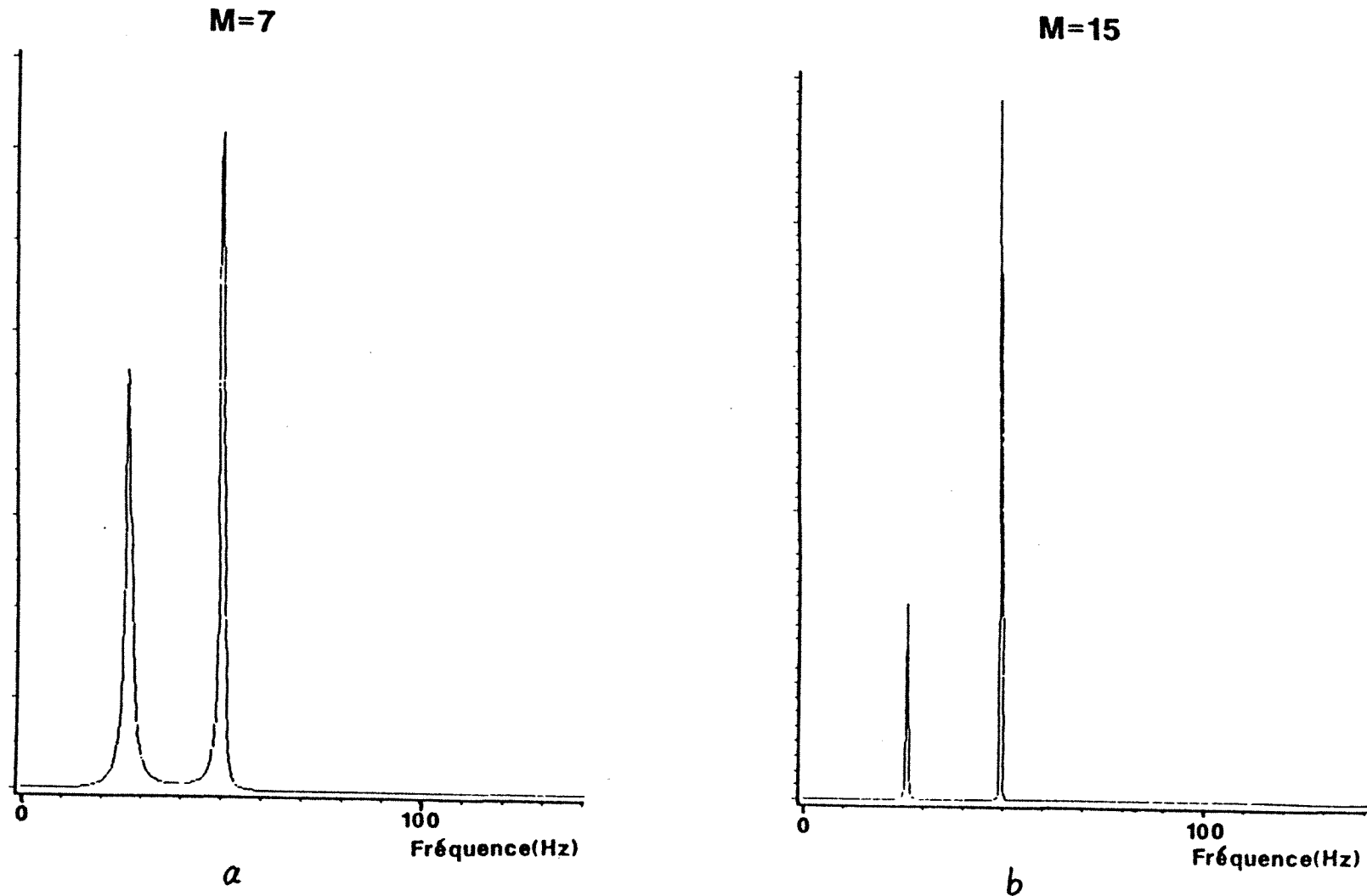
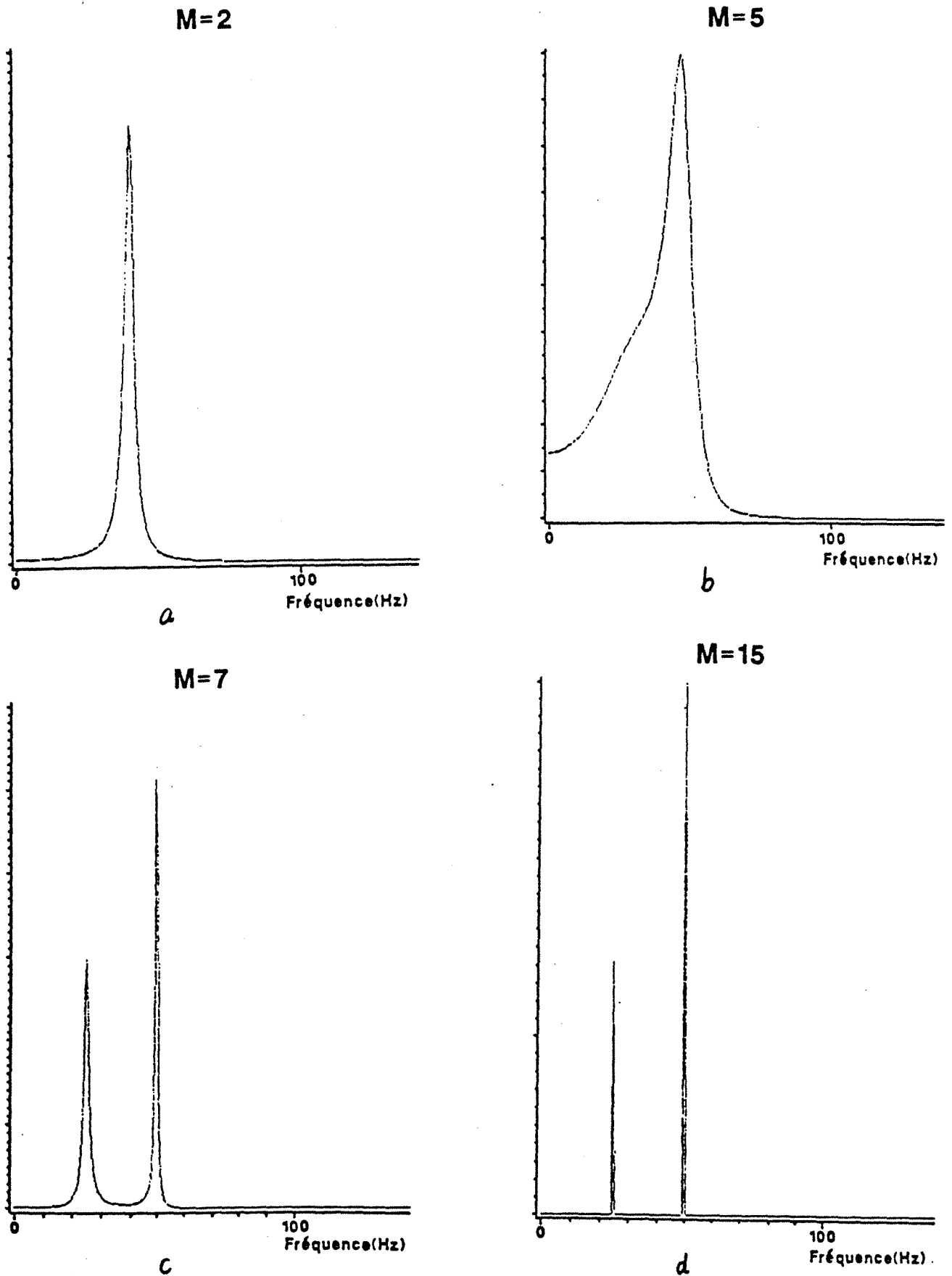


Fig. 152

SPECTRE DE PUISSANCE
ALGORITHME DE MARPLE

Fig.153

$N=50$, $\Delta t=2\text{ms}$

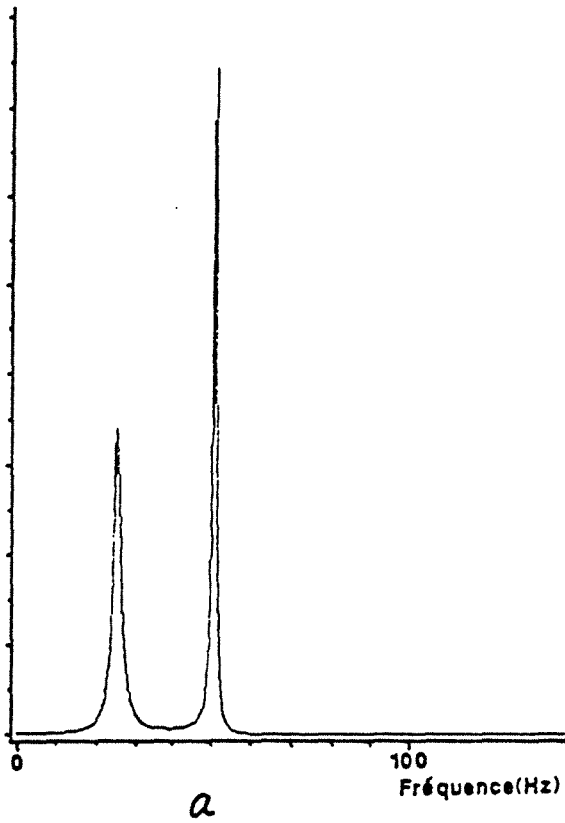


SPECTRE DE PUISSANCE

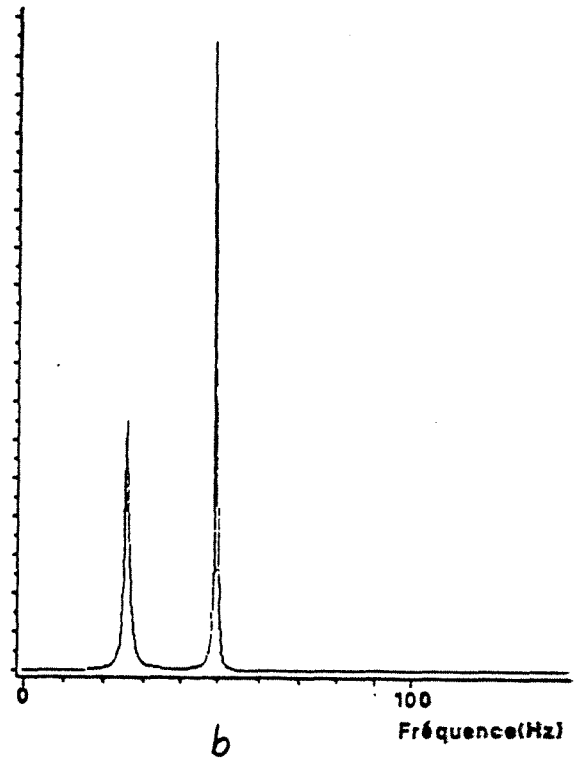
ALGORITHME DE BURG

M=7 , $\Delta t=2ms$

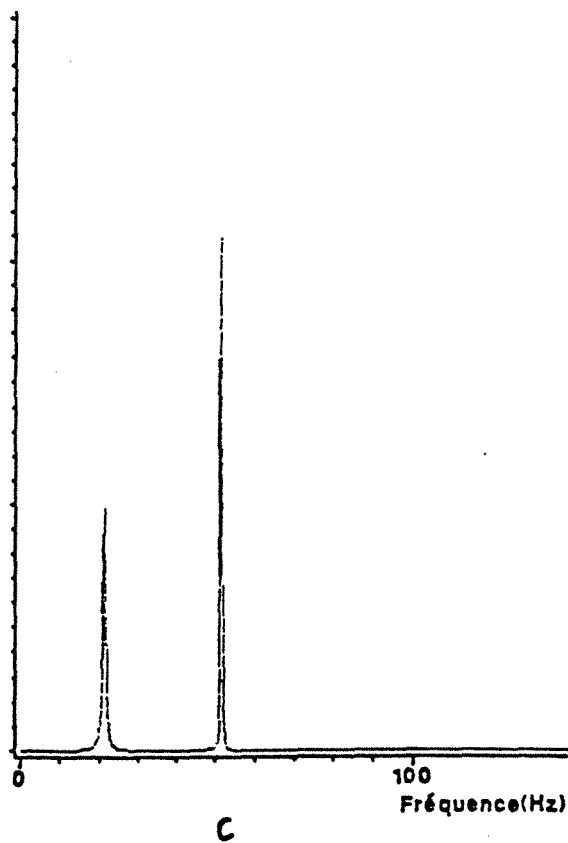
N=100



N=25

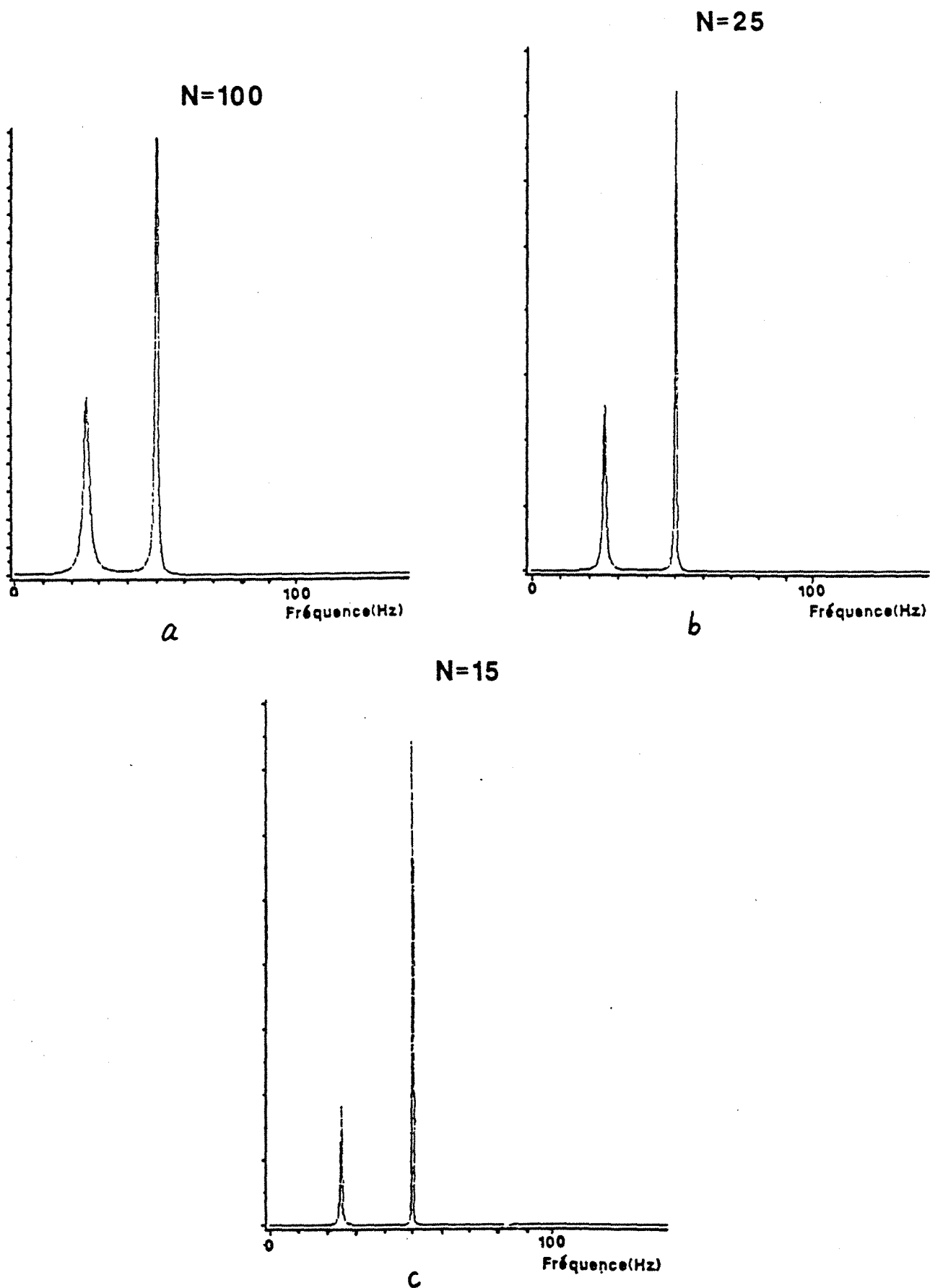


N=15



SPECTRE DE PUISSANCE
ALGORITHME DE MARPLE

$M=7$, $\Delta t=2ms$

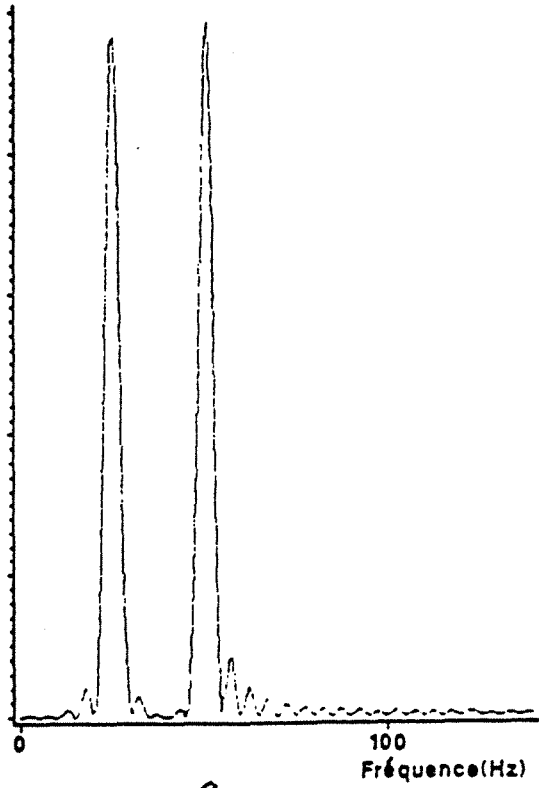


SPECTRE DE PUISSANCE

FFT

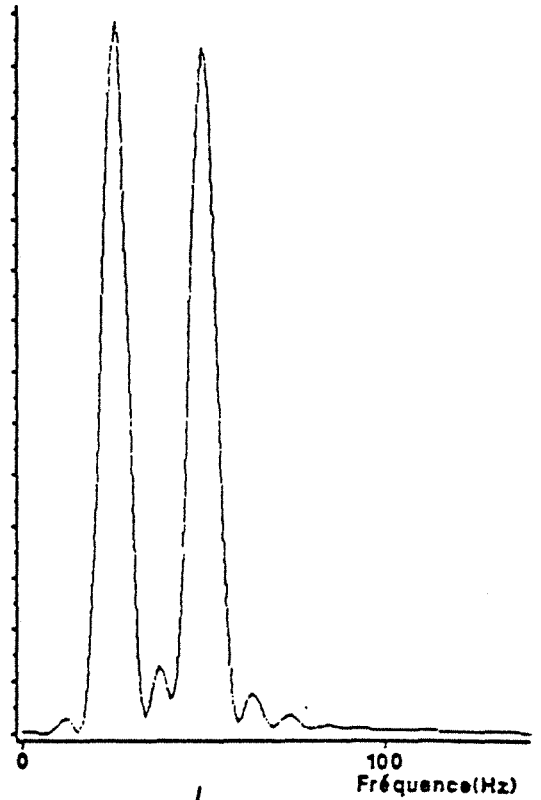
$\Delta t=2ms$, $\Delta f=0.977$

N=100



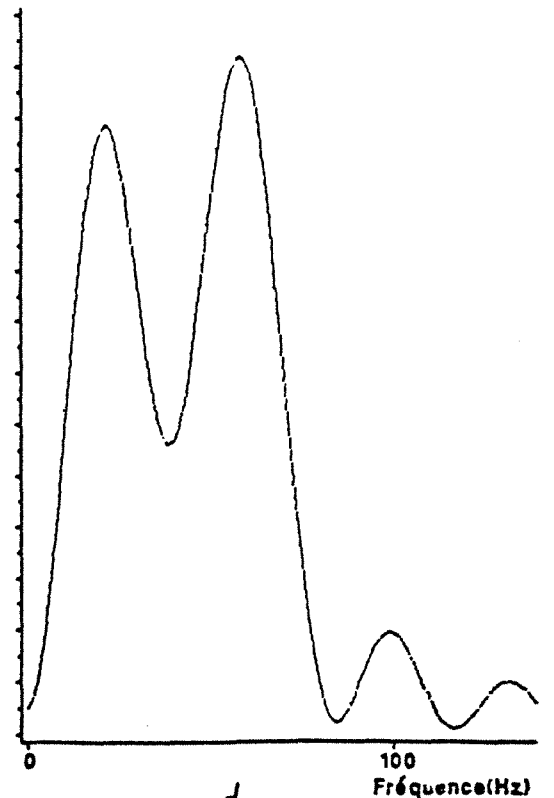
a

N=50



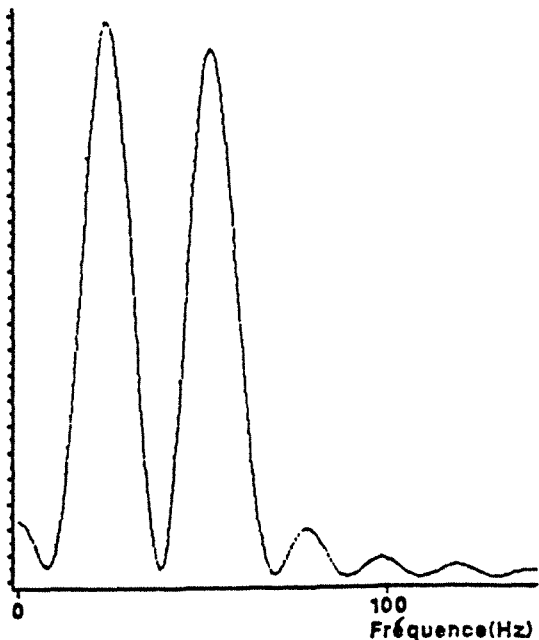
b

N=15



d

N=25



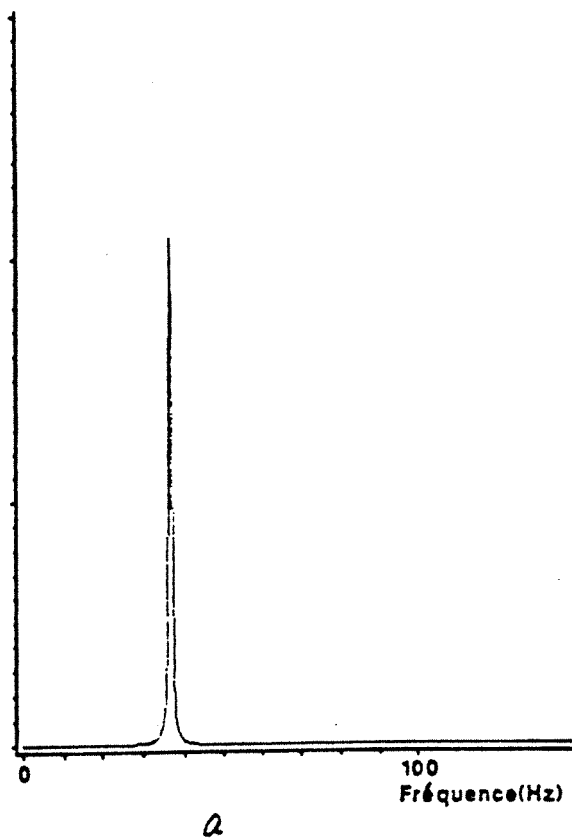
c

SPECTRE DE PUISSANCE

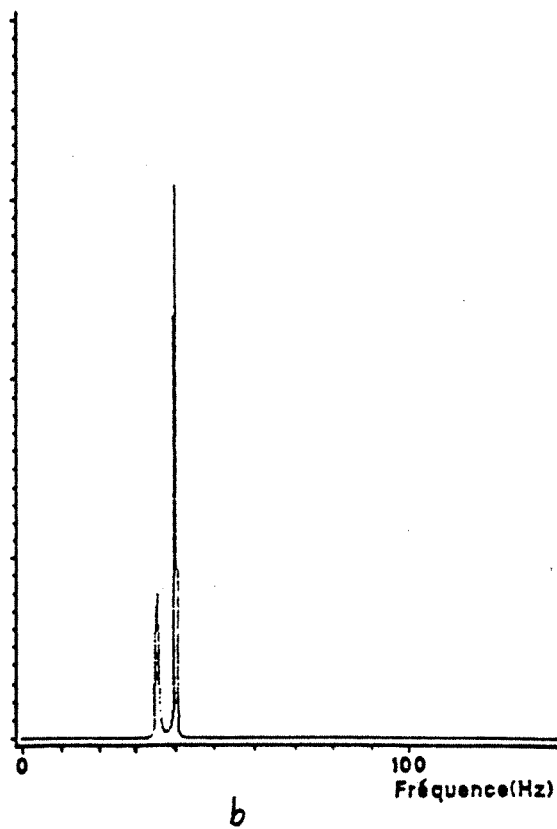
ALGORITHME DE MARPLE

$\Delta t=2\text{ms}$

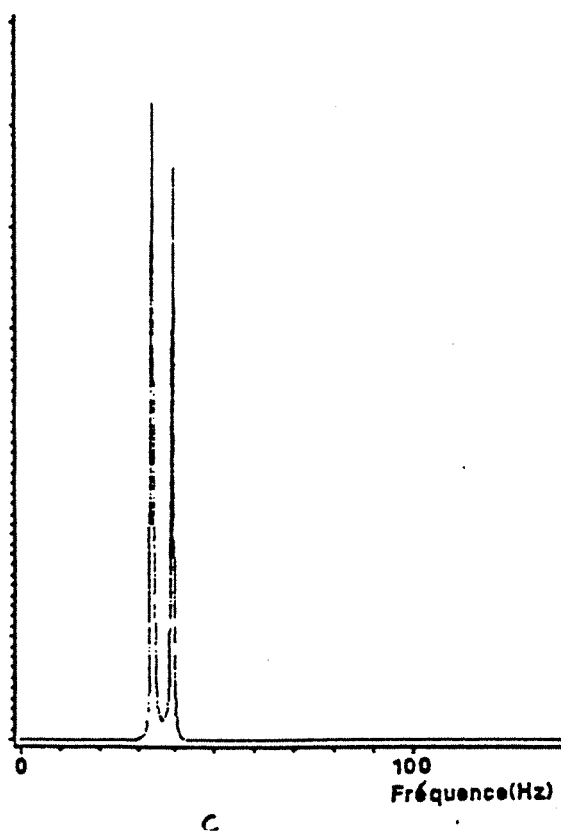
M=7 , N=50



M=15 , N=50



M=15 , N=25



SPECTRE DE PUISSANCE

FFT

$\Delta t = 2\text{ms}$, $\Delta f = 0.977$

N=50

N=100

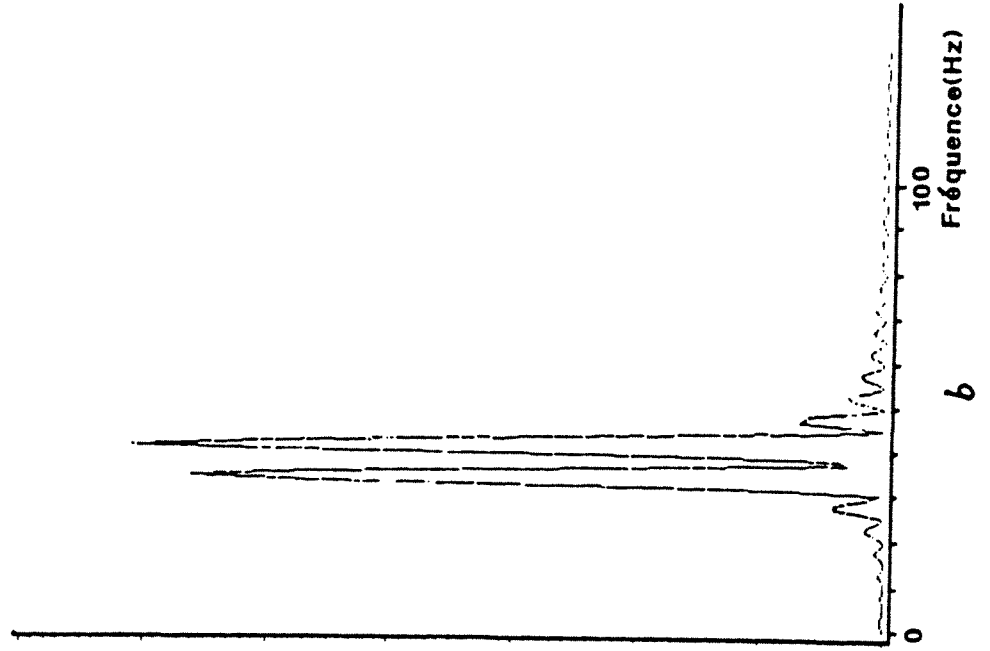
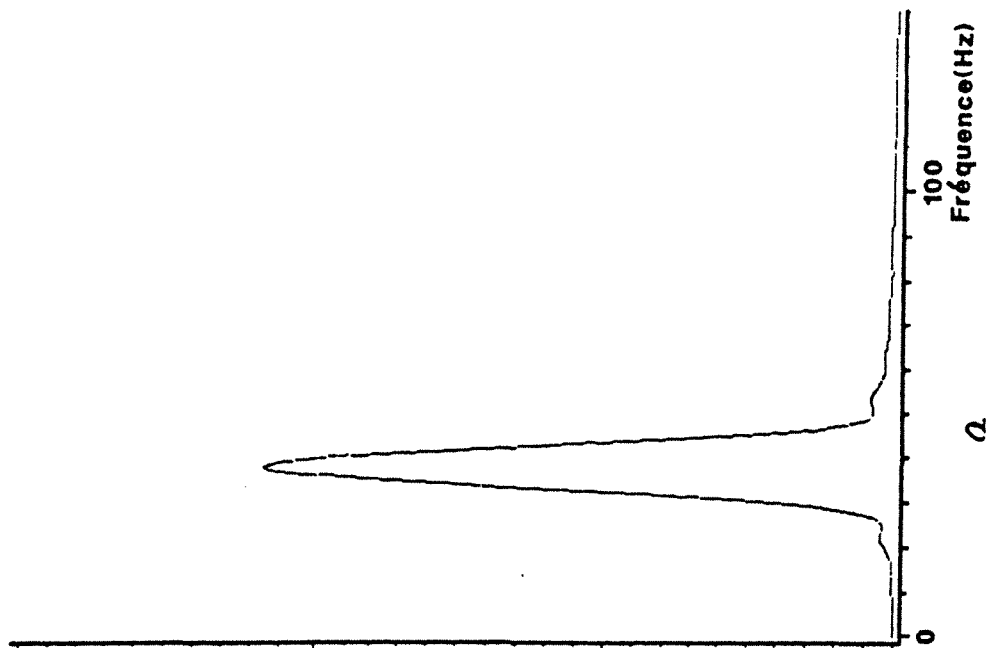


Fig.158

- Parmi les quatre algorithmes testés, c'est celui de Marple qui donne généralement les meilleurs résultats. L'algorithme de Barrodale-Erickson est numériquement moins robuste et le temps de calcul est plus long. L'algorithme de Burg est un peu moins robuste que celui de Marple, mais la différence est souvent faible. Quant à la méthode de Yule-Walker, elle est bien moins efficace que les trois autres.
- L'influence de l'ordre du modèle autorégressif est très importante dans l'estimation du spectre par la M.E.M. quel que soit l'algorithme. Mais la détermination de l'ordre est un problème très délicat. Certains auteurs le déterminent simplement à partir du nombre de données ; d'autres définissent un critère sur la fonction d'autocorrélation ou sur la variance d'estimation.
- En éliminant les cas extrêmes, l'influence du nombre de données dans l'estimation spectrale par la M.E.M. est relativement faible. Ceci veut dire que l'estimation du spectre reste satisfaisante même pour un petit nombre de points (entre 20 et 50).
- Il peut arriver que le phénomène de "splitting" soit très important, en particulier pour les ordres élevés (par rapport au nombre de données). Pratiquement, en se limitant à un ordre inférieur à la moitié du nombre d'échantillons, ce problème est secondaire.

Si l'algorithme de Marple nous semble le plus satisfaisant (faible coût de calcul, robustesse vis-à-vis de données bruitées ainsi que par rapport à la diminution du nombre de points), nous rappelons que la difficulté essentielle réside dans le **choix de l'ordre du modèle autorégressif** et qu'il n'y a pas de critère satisfaisant pour le déterminer. Nous avons pu constater à quel point la résolution fréquentielle recherchée (i.e. écart minimum entre deux pics résolus), le nombre de fréquences pures intervenant, le nombre de points de données ainsi que le phénomène de "splitting" interféraient de façon complexe. En particulier, nous avons observé que les conclusions tirées des études sur la somme de deux sinusoïdes pures n'étaient pas directement transposables au cas de trois sinusoïdes pures. On peut alors s'interroger sur la transposition de ces mêmes conclusions à des **traces sismiques réelles**. Malheureusement, nous ne pouvons pas contrôler les résultats, ne disposant que d'**estimations** de leur spectre (expression analytique inconnue).

Il est intéressant, à ce stade, de comparer les types d'essais que nous avons réalisés et nos conclusions aux études menées antérieurement.

3.5. Comparaison des résultats obtenus avec ceux des études antérieures

3.5.1. Résultats antérieurs

Les essais, toujours sur traces synthétiques, ne vont pas au-delà de la superposition de six sinusoïdes amorties et légèrement bruitées (Kane, Trivedi, 1982).

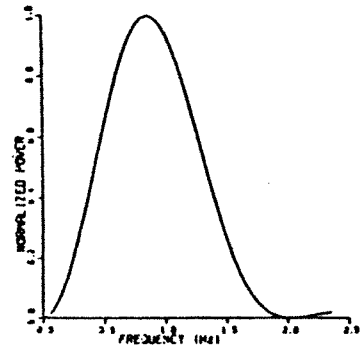
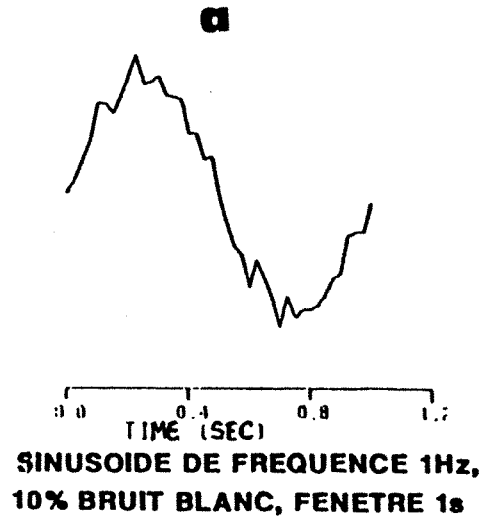
Ulrych (1972) applique la méthode de Burg sur une sinusoïde de fréquence 1 Hz bruitée de 10 % (bruit blanc), étudiée sur une fenêtre d'une seconde (une période). Il constate une amélioration spectaculaire de la définition spectrale par rapport aux résultats de la F.F.T. (Figure 159.a) ; cette amélioration semble robuste vis-à-vis de la rotation de phase de 90° (Figure 159.b). Tronquer les données (demi-période) entraîne avec la F.F.T. un décalage du spectre vers la fréquence nulle ; ce phénomène n'existe pas sur l'estimation spectrale donnée par l'algorithme de Burg (Figure 159.c).

Marple (1982) fait mention d'une série d'essais comparatifs entre l'analyse spectrale par transformée de Fourier et par la méthode de l'entropie maximale (algorithme de Burg). Le signal étudié comporte deux fréquences pures f_1 , f_2 telles que $f_1 + f_2 = 1/2 \Delta t$ (Δt étant le pas d'échantillonnage). Ni l'écart entre les deux fréquences, ni les phases ou le nombre de points ne sont précisés. Ce sont plutôt les effets du rapport signal/bruit et de l'ordre du modèle autorégressif qui sont analysés. Marple conclut que si le rapport signal/bruit est faible, le gain en résolution par rapport à la transformée de Fourier est négligeable. Au contraire, si le rapport signal/bruit est très grand, le gain en résolution est de plusieurs ordres de grandeur supérieur à ce que l'on pourrait attendre de la transformée de Fourier. Marple souligne bien que ces résultats sont expérimentaux, pour le cas de signaux sinusoïdaux ou à bande très étroite.

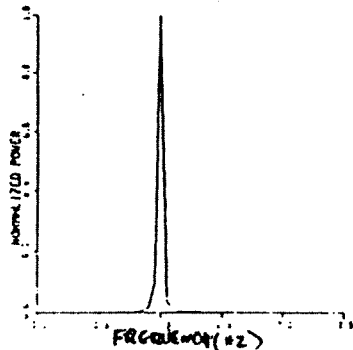
C'est également l'algorithme de Burg qui a été utilisé par Wenzel et Zurn (1980) pour juger des performances de la méthode de l'entropie. Les données synthétiques testées sont très loin du cadre de notre étude. Ils ne s'attaquent pas au problème "fenêtre courte" (fenêtre toujours supérieure à vingt fois la période du signal) mais étudient la résolution maximale qu'il est possible d'atteindre (Δf jusqu'à 5×10^{-7} Hz). Les signaux analysés sont la somme de deux ou trois sinusoïdes amorties, bruitées et déphasées. Dans le cas de deux sinusoïdes, les deux pics sont distincts jusqu'à $\Delta f = 5 \times 10^{-7}/T$ (avec T longueur de la fenêtre d'étude). Une limite plus sûre est la valeur $\Delta f T$ telle que $0,4 \leq \Delta f T \leq 0,5$ (avec la transformée de Fourier, $\Delta f T = 1$). Dans le cas de trois sinusoïdes, les auteurs montrent que la résolution donnée par l'algorithme de Burg n'est pas meilleure que celle donnée par la transformée de Fourier ($\Delta f = 1/T$).

A notre sens, ceci illustre clairement que l'analyse spectrale par la méthode de l'entropie ne présente **aucun avantage** sur une **fenêtre longue** lorsque les signaux sont à large bande, comme dans le cas des traces sismiques (plus de deux fréquences pures). L'étude de Wenzel et Zurn (1980) ne concerne pas des applications sismiques.

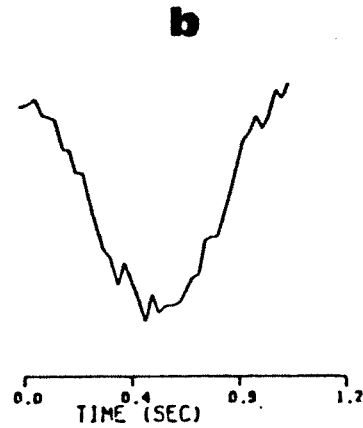
COMPARAISON ALGORITHME DE BURG/FFT (d'après ULRYCH, 1972)



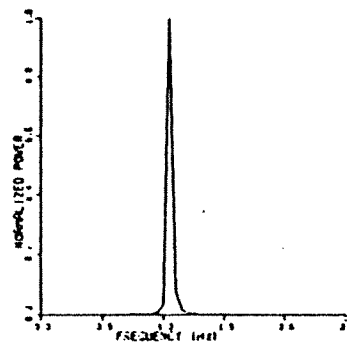
FFT



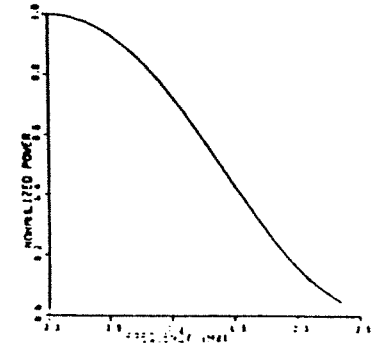
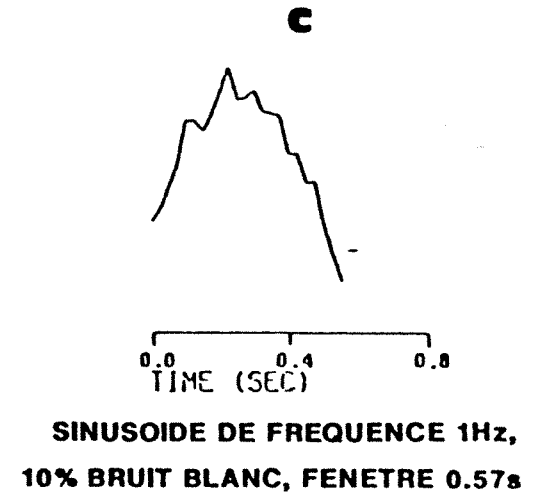
BURG



**SINUSOIDE DEPHASEE DE 90°,
FREQUENCE 1HZ, 10% BRUIT BLANC, FENETRE 1s**



BURG



FFT

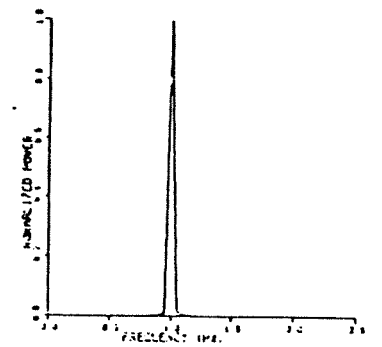


Fig. 159

Ce sont les expériences de Kane et Trivedi (1982) qui concernent vraiment le domaine sismique et se rapprochent le plus de nos tests. Les données de départ correspondent à six sinusoïdes de fréquences 5, 10, 20, 40, 80 et 160 Hz, très faiblement bruitées (bruit gaussien), de même amplitude, sur 101 points. Les algorithmes utilisés sont ceux de Burg, Barrodale et Erickson ainsi que celui de Marple.

Un premier test sur la phase ($\Phi = 0^\circ$, $\Phi = 30^\circ$, $\Phi = 60^\circ$, $\Phi = 120^\circ$) montre que son influence est négligeable.

Avec tous les algorithmes testés, les pics haute fréquence sont moins bien résolus que les pics basse fréquence (jusqu'à 30 % de décalage sur le pic à 160 Hz, Figure 160). La hauteur relative des pics est difficile à interpréter ; en fait, mieux vaut ne pas la prendre en compte dans une analyse.

Les décalages fréquentiels semblent moins importants avec l'algorithme de Barrodale qu'avec celui de Burg. Par contre, l'algorithme de Burg est plus robuste quand l'ordre augmente (moins de "splitting" ou "splitting" plus tardif).

Une autre série de tests est effectuée en rapprochant les fréquences pures constitutives de la trace synthétique. L'algorithme de Burg semble alors plus efficace que celui de Barrodale et Erickson : moins de points sont nécessaires pour résoudre correctement les bonnes fréquences.

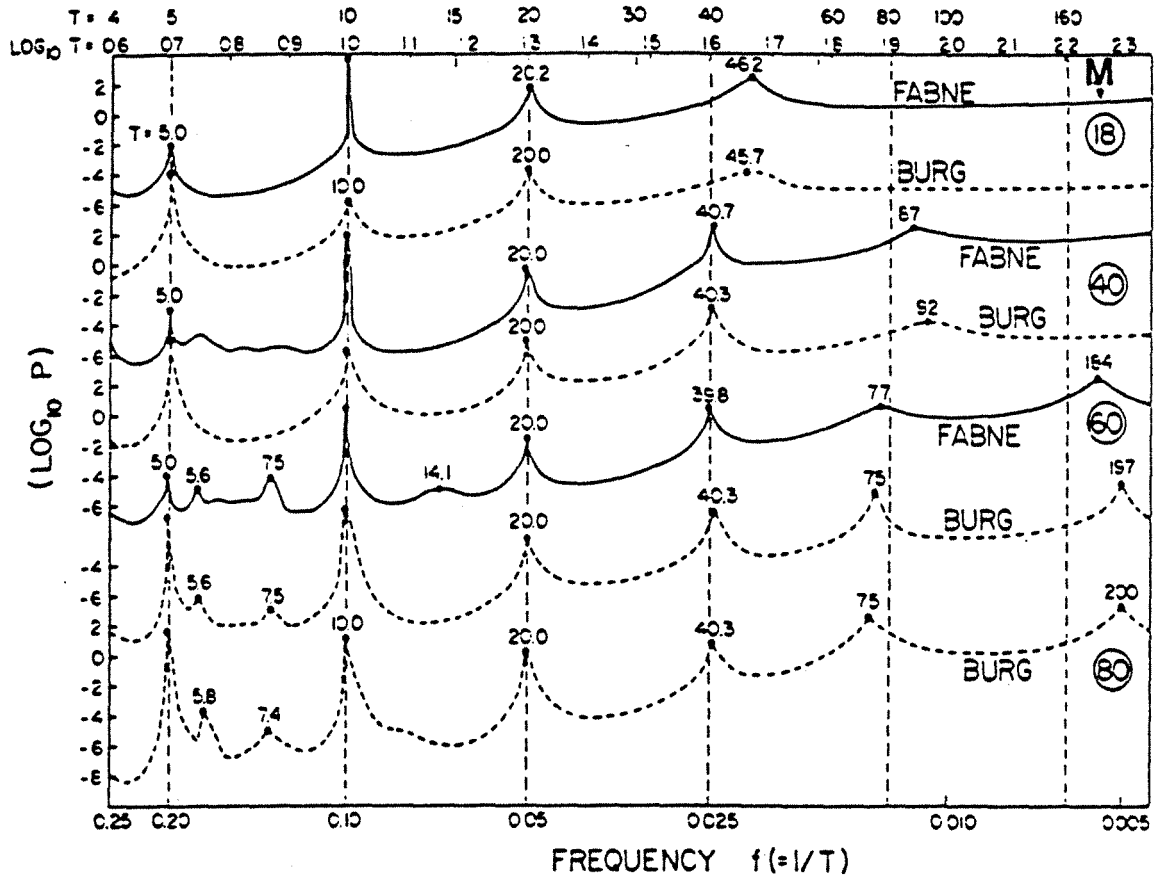
Les résultats obtenus avec l'algorithme de Marple sont très similaires à ceux obtenus avec Barrodale et Erickson .

L'étude de Kane et Trivedi (1982) souligne la difficulté de conclure quant à la supériorité de tel ou tel algorithme : cela dépend beaucoup du jeu de données. On est en droit de s'interroger sur les performances de la méthode d'entropie maximale appliquée à des données réelles pour lesquelles il est nécessaire de résoudre des **pics proches** mais qui comportent aussi des pics **largement répartis** dans la bande spectrale, et ceci sur un petit nombre de points.

Les résultats de Kane et Trivedi (1982) corroborent les difficultés que nous avons rencontrées pour interpréter nos propres résultats : interférences complexes et compromis difficile à trouver entre l'ordre du modèle, le nombre de points (taille de la fenêtre), la résolution fréquentielle recherchée et l'espacement réel des pics spectraux.

Barrodale et al. (1983) ont comparé leur algorithme avec celui de Marple sur plusieurs jeux de données. Les résultats sont identiques sur des données moyennement bruitées. Pour des données peu bruitées, l'algorithme de Barrodale et Erickson peut se révéler meilleur. Par contre, pour des données extrêmement peu bruitées (rapport signal/bruit $\rightarrow +\infty$), les deux programmes peuvent retourner des valeurs négatives pour l'erreur de prédiction (ceci nous est

ESSAIS DE LA METHODE D'ANALYSE SPECTRALE PAR L'ENTROPIE
(d'après KANE, TRIVEDI, 1982)



POINTILLE : SPECTRE BURG

PLEIN : BARRODALE-ERICKSON

N=101, 6 PERIODES(5,10,20,40,80,160)

$\phi=60\%$

0,1% BRUIT GAUSSIEN

arrivé avec l'algorithme de Barrodale-Erickson sur un cosinus pur). Vu le gain de vitesse de calcul entre les deux algorithmes, les auteurs préconisent largement l'emploi de l'algorithme de Marple et en proposent une version modifiée qui permet de parer au problème précédent ; lorsque l'erreur de prédiction devient négative (suite à l'emploi d'une "formule rapide" pour la calculer), ce sont les coefficients de l'ordre précédent qui sont retenus (même si ce n'est pas forcément l'ordre optimal).

D'autre part, il est possible d'utiliser une formule plus complète pour le calcul de l'erreur de prédiction. C'est cet algorithme de Marple révisé que nous avons implémenté et testé.

En sismique, les données sont très souvent bruitées et dans un rapport plus important que ce que Barrodale et Erickson ont testé ; c'est une raison de plus de recommander l'algorithme de Marple.

3.5.2. Comparaison des résultats antérieurs avec les nôtres

La récapitulation des travaux antérieurs met en évidence le très petit nombre de tests effectués sur des données synthétiques en rapport avec **les problèmes sismiques** : fenêtre courte, gamme de fréquences et répartition des fréquences usuelles pour un sismicien.

Ces essais sont en accord avec les nôtres pour le choix de l'algorithme (version modifiée de celui de Marple), pour l'influence du bruit et pour le gain en résolution spectrale dans le cas de signaux à bande étroite (sinusoïdes). Nous sommes aussi en accord pour remarquer que, dès que trois sinusoïdes interviennent, les résultats sont beaucoup plus complexes à interpréter.

Dans le cas d'une trace sismique réelle, a-t-on encore avantage à faire appel à cette technique d'analyse spectrale ?

Pour Kane et Trivedi, la méthode est intéressante (gain en résolution non négligeable) à condition de choisir un ordre M correct. Malheureusement, les travaux antérieurs et notre propre expérience montrent que le choix de l'ordre est totalement heuristique. Connaissant le spectre d'une fonction, il est possible de trouver un ordre qui permette d'avoir une excellente résolution spectrale, bien supérieure à celle donnée par la méthode de Fourier. Mais pratiquement, c'est justement ce spectre que l'on cherche à estimer au mieux, sans beaucoup d'information a priori.

Pour ce qui est du rapport des pics spectraux, aucun auteur n'y fait allusion à part Kane et Trivedi qui insistent sur le fait qu'il faut éviter de l'interpréter. De notre côté, nous avons constaté que, d'un ordre au suivant, la hauteur relative des pics était complètement changée, voire inversée. C'est donc un problème important, peu abordé car probablement délicat à résoudre.

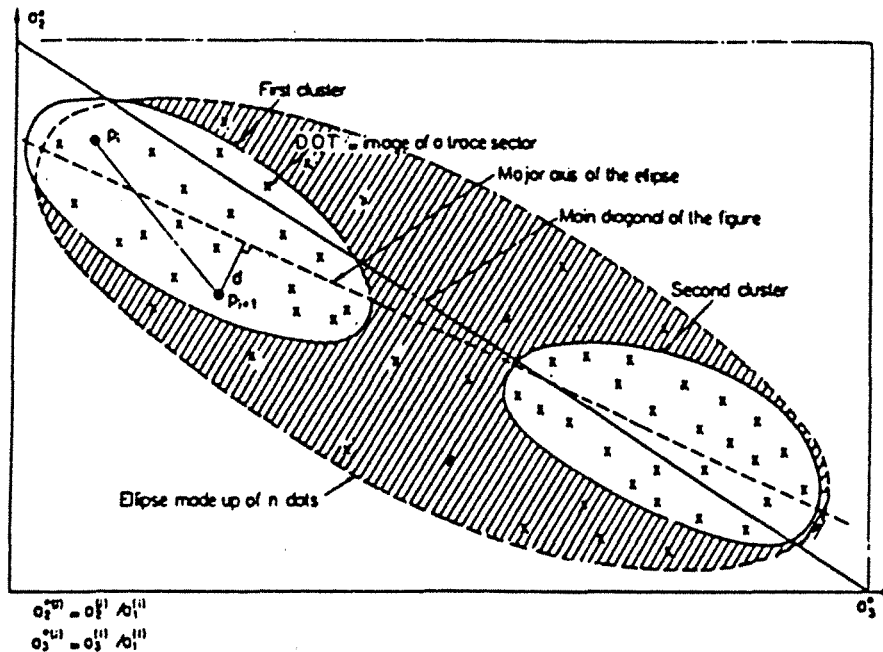
Finalement, nous avons été frappés par le contraste entre l'abondance de littérature concernant les bases théoriques de l'analyse spectrale par la méthode de l'entropie maximale ou les algorithmes possibles pour la mettre en œuvre, et l'insuffisance des tests de ces méthodes. Beaucoup d'auteurs se contentent de les appliquer sur une sinusoïde pure ; les traces synthétiques étudiées combinent rarement plus de trois fréquences. A notre connaissance, mis à part les travaux de Bois (utilisation des coefficients autorégressifs dans un algorithme de reconnaissance des formes pour classer des traces, Figure 161), aucune étude de trace sismique réelle avec l'analyse spectrale par maximisation de l'entropie n'a été entreprise.

Il faut ajouter que, pour le choix de l'ordre du modèle, des travaux plus récents permettent de proposer une solution (Ma, 1988). Il s'agit d'utiliser un critère basé sur la fonction d'autocorrélation partielle. On utilise la propriété que, pour des processus autorégressifs, les coefficients de corrélation partielle diminuent globalement selon l'ordre croissant. L'ordre M est retenu si le coefficient $a_{M+1,M+1}$ est nettement petit devant le coefficient $a_{M,M}$, en valeur absolue, ou si le coefficient $a_{M+1,M+1}$ est inférieur à une valeur faible en valeur absolue. Ce critère semble plus intéressant que les autres méthodes proposées dans la littérature.

Nous comparerons la caractérisation de la trace, sur courte fenêtre, par les coefficients du modèle autorégressif avec la technique développée dans le paragraphe suivant, sur un **modèle synthétique issu de données réelles**.

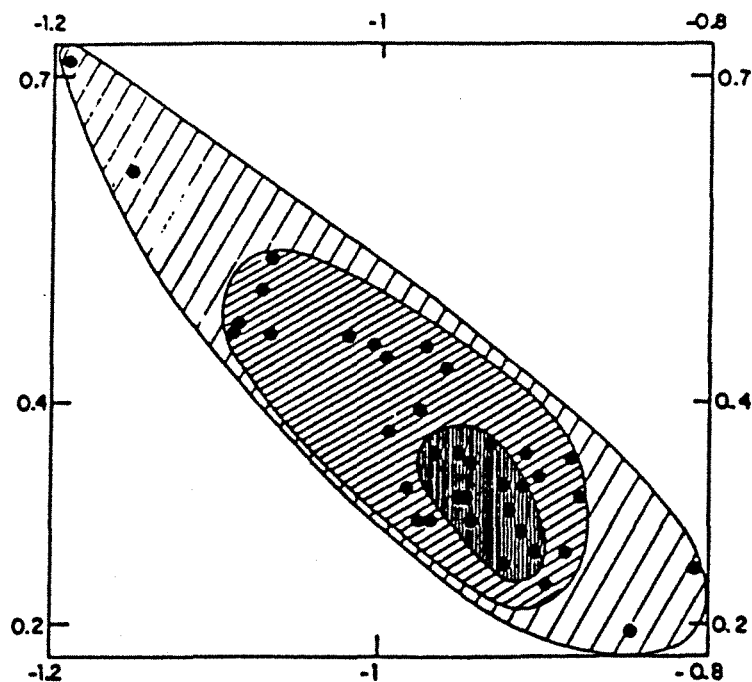
UTILISATION DES COEFFICIENTS AUTOREGRESSIFS
 DANS UN ALGORITHME DE RECONNAISSANCE DES FORMES
 (d'après BOIS 1980, 1981, 1982)

a



PLOT DES COEFFICIENTS AUTOREGRESSIFS POUR DES TRACES APPARTENANT A 2 ZONES DISTINCTES.

b



PLOT DES COEFFICIENTS AUTOREGRESSIFS SUR LES TRACES D'UNE PORTION DE SECTION.

4. CARACTERISATION DE LA TRACE PAR LA SUITE DES AMPLITUDES

4.1. Principe

L'idée sous-jacente au calcul des paramètres de la méthodologie "Alias" est de prendre en compte toute l'information sismique contenue dans la portion de trace extraite, puis de retenir les éléments importants de cette information après une phase d'apprentissage (analyse factorielle discriminante) ou de la condenser (le plus avantageusement possible).

L'information initiale étant toujours la représentation temporelle de la trace (suite de n valeurs d'amplitude échantillonnées à pas constant), il est possible de considérer comme variables les valeurs des amplitudes AMP_i ($1 \leq i \leq n$) et de représenter la trace dans IR^n , ses coordonnées étant les valeurs des amplitudes successives (Figure 162). Les traces analysées étant représentées dans un **espace multivariable, directement celui des amplitudes**, nous disposons alors d'une technique statistique très robuste pour condenser l'information : l'analyse en composantes principales qui permet d'extraire un petit nombre de nouvelles variables résumant les traits importants de l'information initiale.

Les axes factoriels extraits par analyse en composantes principales sont orthogonaux (Figure 162) (non corrélés) et l'inertie (deuxième partie, § 4.2.1) – ou l'information initiale – est dans sa majeure partie prise en compte par les premiers axes. La représentation du nuage de points – les traces sismiques – sur les premiers plans factoriels est donc une condensation avantageuse de l'information initiale (Figure 162). Il est possible de proposer une "interprétation physique" aux axes factoriels en étudiant leurs corrélations avec les variables de départ. Une forte corrélation, qu'elle soit positive ou négative, signifie que l'information prise en compte par l'axe factoriel est très influencée par l'information contenue dans cette variable.

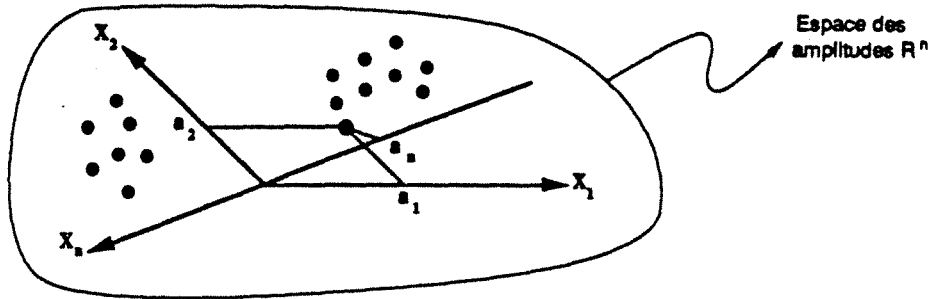
Cependant, il n'est pas toujours facile d'interpréter la signification des nouveaux axes. Ici, les facteurs vont correspondre à des **formes élémentaires de la trace sismique**, déterminées à partir de l'analyse de l'ensemble des formes présentées par toutes les traces de la zone étudiée. Chaque trace sismique peut apparaître comme une combinaison linéaire des facteurs, c'est-à-dire une somme pondérée de ces formes élémentaires. N'utiliser que les premiers axes factoriels pour représenter les traces revient à effectuer un certain **filtrage** de l'information sismique (qui correspond au pourcentage de la variance non prise en compte par ces axes).

Dans ce type de traitement des traces sismiques, il apparaît donc nécessaire de définir deux paramètres :

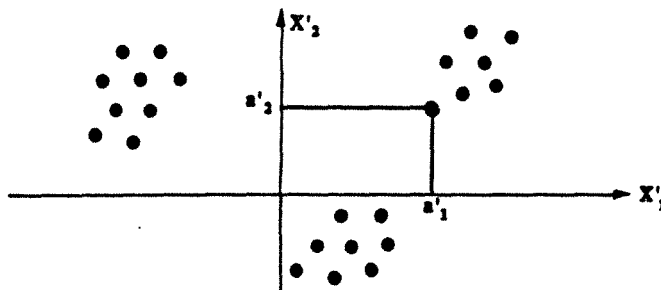
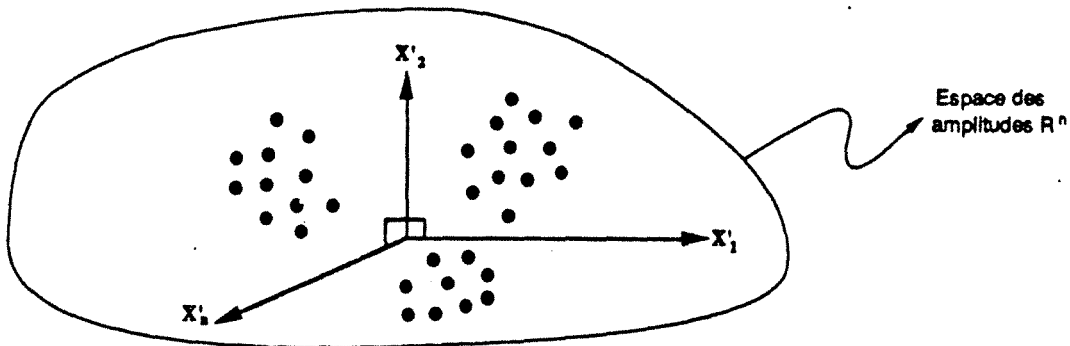
- le nombre de facteurs de l'A.C.P. à retenir pour représenter l'information sismique,

TRAITEMENT DES TRACES A PARTIR DE LEUR REPRESENTATION TEMPORELLE

**1 PORTION DE TRACE = n VALEURS D'AMPLITUDES ($a_1, a_2 \dots a_n$)
ECHANTILLONNEES A PAS CONSTANT**



ANALYSE EN COMPOSANTES PRINCIPALES (A.C.P.)



1^{ERS} PLANS FACTORIELS = CONDENSATION DE L'INFORMATION

1 TRACE = k COORDONNEES SUR LES k 1^{ERS} FACTEURS



CLASSIFICATIONS

**ANALYSES FACTORIELLES
DISCRIMINANTES**

- la **taille de la fenêtre d'analyse temporelle**, qui doit être constante (nous avons déjà rencontré ce problème dans le paragraphe 4.4, deuxième partie, concernant la discussion des limites de la méthodologie proposée pour la reconnaissance du faciès sismique).

L'extraction des facteurs de la matrice de corrélation du nuage de points est très simple et robuste. On comprend donc pourquoi il est séduisant de remplacer les paramètres sismiques usuels – calculés sur longue fenêtre – ou même les méthodes spectrales réputées robustes sur courtes fenêtres, mais complexes à appliquer, par les **premiers axes factoriels de l'A.C.P.** Le même type de traitement statistique (analyses factorielles et classifications) peut, par la suite, être appliqué à ces nouvelles variables, qui caractérisent vraiment la "morphologie" des traces.

Nous avons réalisé un premier test de faisabilité de cette méthode sur des données sismiques synthétiques, calculées par convolution 1D à partir d'un modèle géologique et des impédances acoustiques correspondantes.

4.2. Tests de faisabilité de la méthode

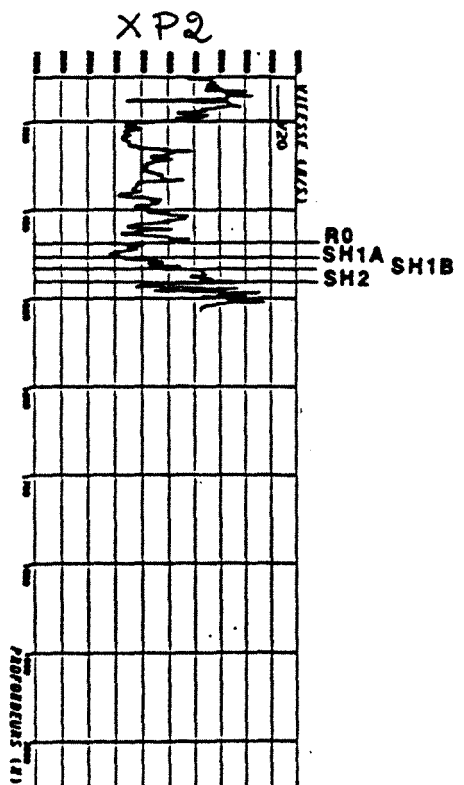
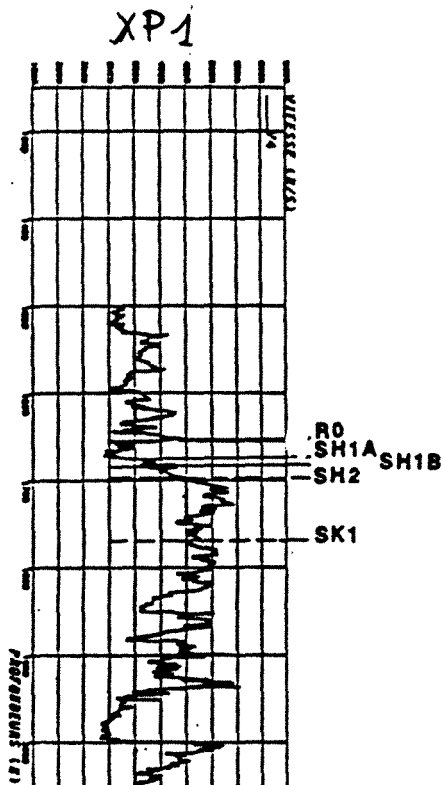
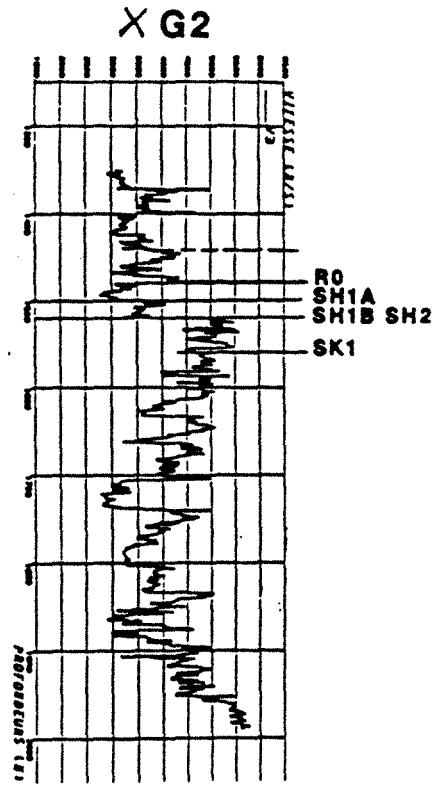
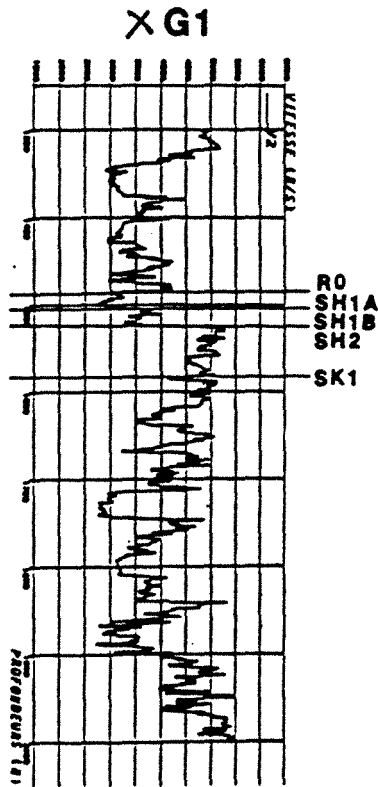
4.2.1. *Données analysées*

L'étude est située sur un permis du Bassin de Paris. L'objectif est un faciès carbonaté du Callovien Inférieur ("dalle nacrée") présentant une bonne porosité au puits XP1. Au puits XP2, la dalle nacrée est **présente mais compacte**.

Les variations de faciès du Callovien Inférieur sont nettement visibles sur les logs Sonic des 4 puits disponibles (Figure 163). Aux deux puits XG1 et XG2, le Callovien Inférieur est marneux (absence de dalle nacrée) ; à XP1, après quelques mètres de Callovien Inférieur marneux, apparaît un faciès oolithique à bonne porosité (vitesse lente par rapport au Bathonien) : c'est la dalle nacrée, faciès poreux. A XP2, la dalle nacrée est compacte (observer sur la figure 163 l'évolution progressive de la vitesse depuis le toit du Callovien Inférieur jusqu'au Bathonien).

Entre le toit du Callovien Inférieur et le toit du Bathonien, il y a environ 12 ms, temps double. On peut supposer que le Callovien Supérieur argileux ne présente pas de variation latérale. D'autre part, le dernier marqueur pointable avec sûreté est un repère oolithique légèrement au dessus du toit du Bathonien. Tout ceci fait que la fenêtre maximale d'étude (n'englobant qu'une seule évolution : celle du Callovien Inférieur) fait 40 ms, temps double. Ce cadre géologique et sismique correspond tout à fait au problème des courtes fenêtres temporelles. Nous l'avons utilisé pour construire un modèle sismique synthétique pour le test de faisabilité.

SONIC DES Puits XG1, XG2, XP1 ET XP2



R0: TOIT DU CALLOVIEN SUPERIEUR
 SH1A: TOIT DU CALLOVIEN INFERIEUR
 SH1B: TOIT DU CALLOVIEN INFERIEUR CARBONATE (FACIES DALLE NACREE)
 SH2: TOIT DU BATHONIEN

Nous avons corrélé les principaux horizons entre les puits XG, XP1 et XP2, puis interpolé les vitesses à partir des 4 logs Sonic. La section "impédances synthétiques" obtenue ainsi que la section "coefficients de réflexion" (signal utilisé : Ricker d'ordre 2, de fréquence centrale 30 Hz) sont horizontalisées sur le marqueur Xo : creux négatif à 892 ms, temps double (Figures 164 et 165).

4.2.2. Premier essai de l'analyse en composantes principales

La fenêtre étudiée fait 40 ms (toit sur le marqueur Xo), soit 21 échantillons. Les deux premiers facteurs représentent 76 % de l'inertie totale et 97 % de l'inertie est représenté par les 5 premiers axes (Figure 166).

Si l'on retient les 5 premiers facteurs (ou composantes principales) comme **variables initiales** pour un algorithme de classification, que ce soit à 3 ou 4 groupes, (Figure 167), seul le faciès XP2 est individualisé par rapport à tous les autres : l'apparition de la dalle nacrée (XP1 par rapport à XG) ne marque pas, le trait dominant est la disparition du second pic avec renforcement du premier.

Par contre, avec 3 facteurs, la classification à 3 ou 4 groupes (Figure 167) isole bien des zones différentes correspondant à l'évolution latérale progressive du faciès.

Assez curieusement, la perte de 10 % d'information initiale permet une meilleure classification ; bien que ces données synthétiques ne soient pas bruitées, les premiers axes prennent vraiment en compte les traits dominants de l'information et donc assurent une bonne classification. Introduire les axes d'ordre supérieur, en ramenant plus d'information, semble "brouiller" l'essentiel et la classification est moins bonne. Le nombre de facteurs ou composantes principales à retenir apparaît déjà comme un paramètre important dont le réglage correspond à un effet de filtrage plus ou moins prononcé.

Ce premier essai sur le modèle synthétique prouve la faisabilité de la technique proposée. L'analyse en composantes principales permet de réduire efficacement la dimension du problème en fournissant de nouvelles variables pouvant caractériser la trace dans un algorithme de reconnaissance des formes.

A partir de là, nous avons cherché à tester l'influence du pas d'échantillonnage et du nombre de facteurs retenus, ainsi qu'à diminuer la taille de la fenêtre.

SECTION IMPEDANCES ACOUSTIQUES DU MODELE 1
(EVOLUTION ENTRE X_{G2} , X_{G1} , X_{P1} ET X_{P2})

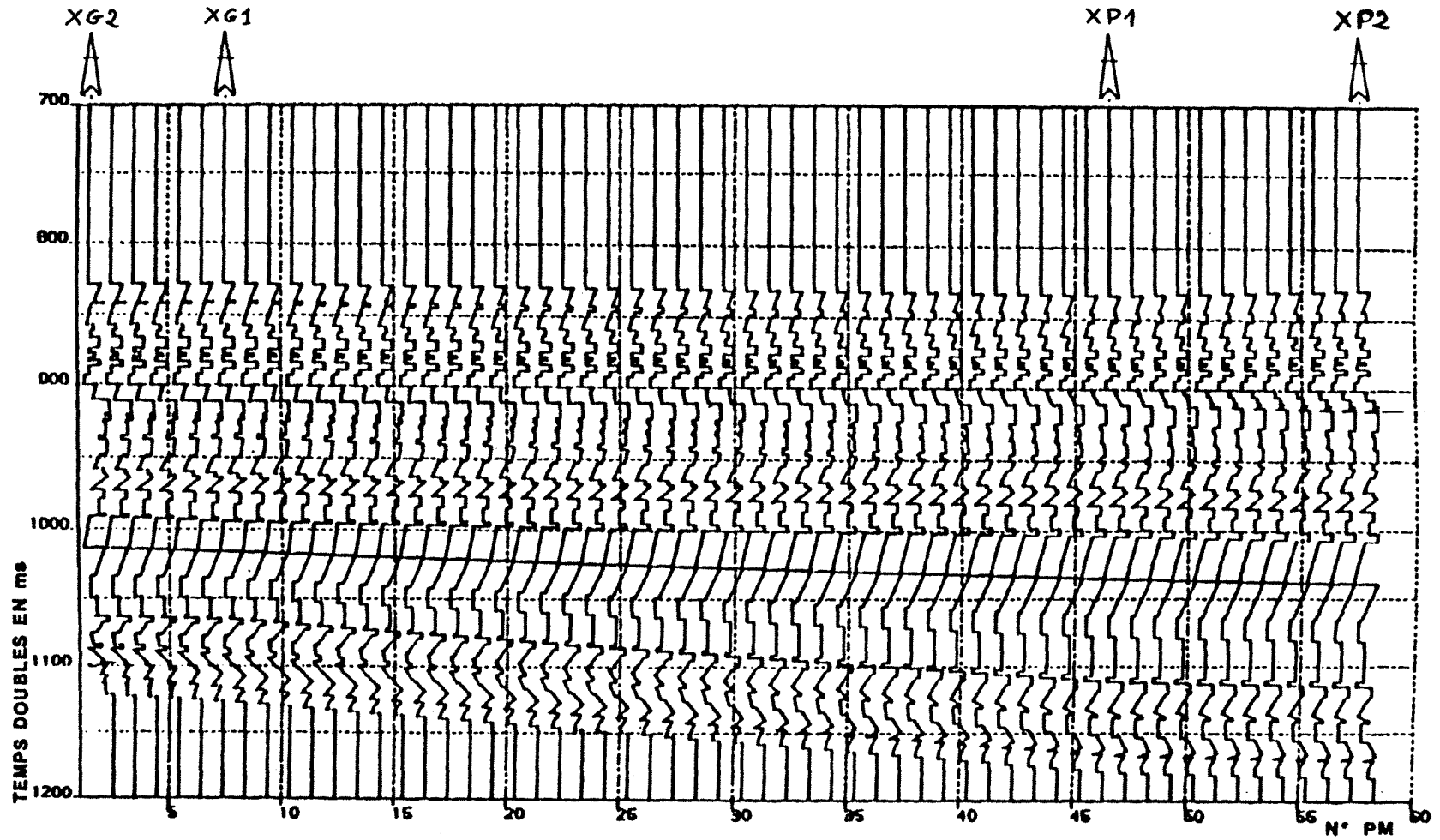


Fig. 164

MODELE SYNTHETIQUE N° 1 D'EVOLUTION ENTRE LES Puits XG2, XG1, XP1 ET XP2

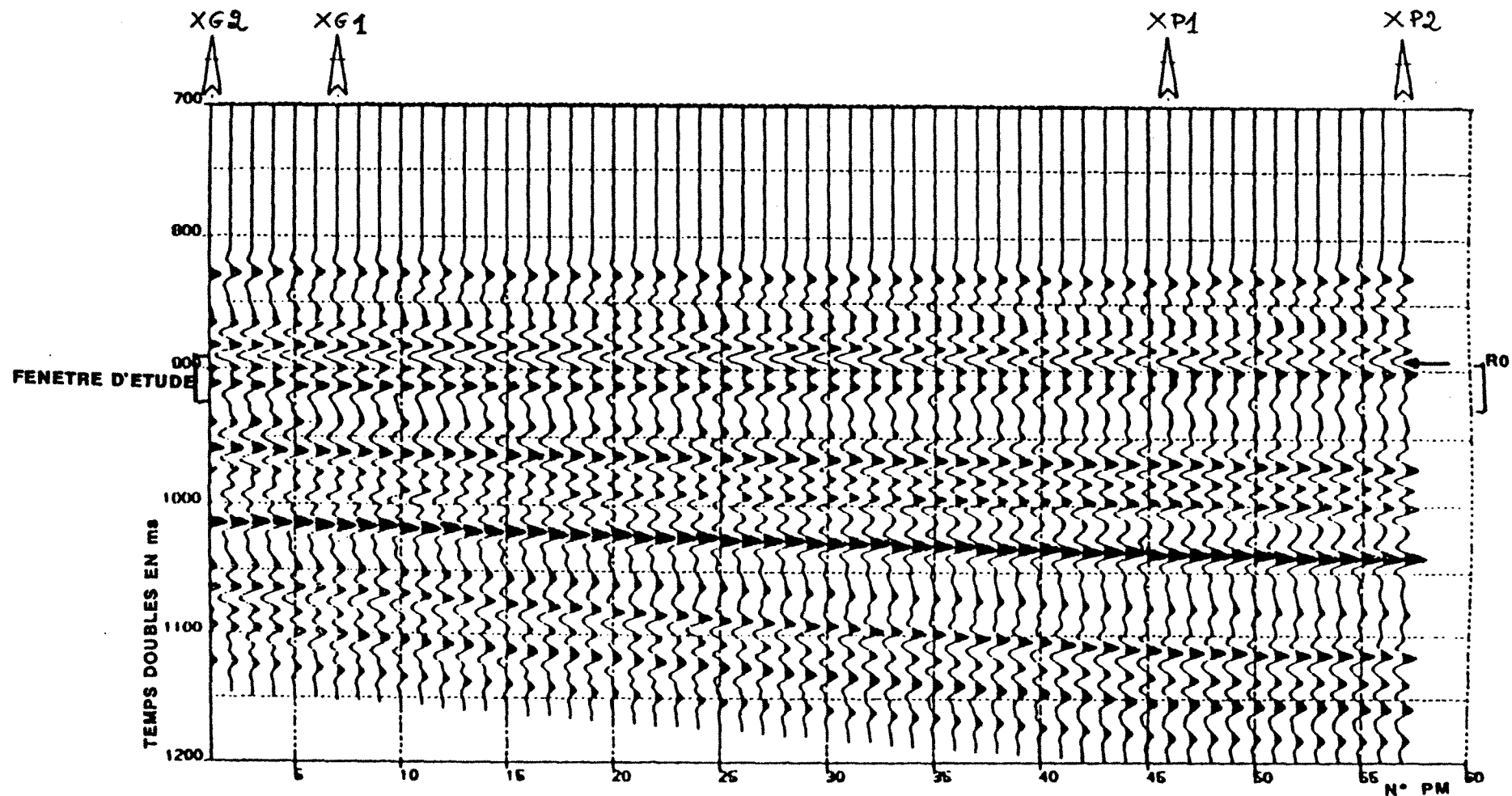


Fig. 165

ANALYSE EN COMPOSANTES PRINCIPALES SUR LA FENETRE 892-932ms

Extraction des facteurs et classifications

DISTRIBUTION DE L'INERTIE SUR LES FACTEURS

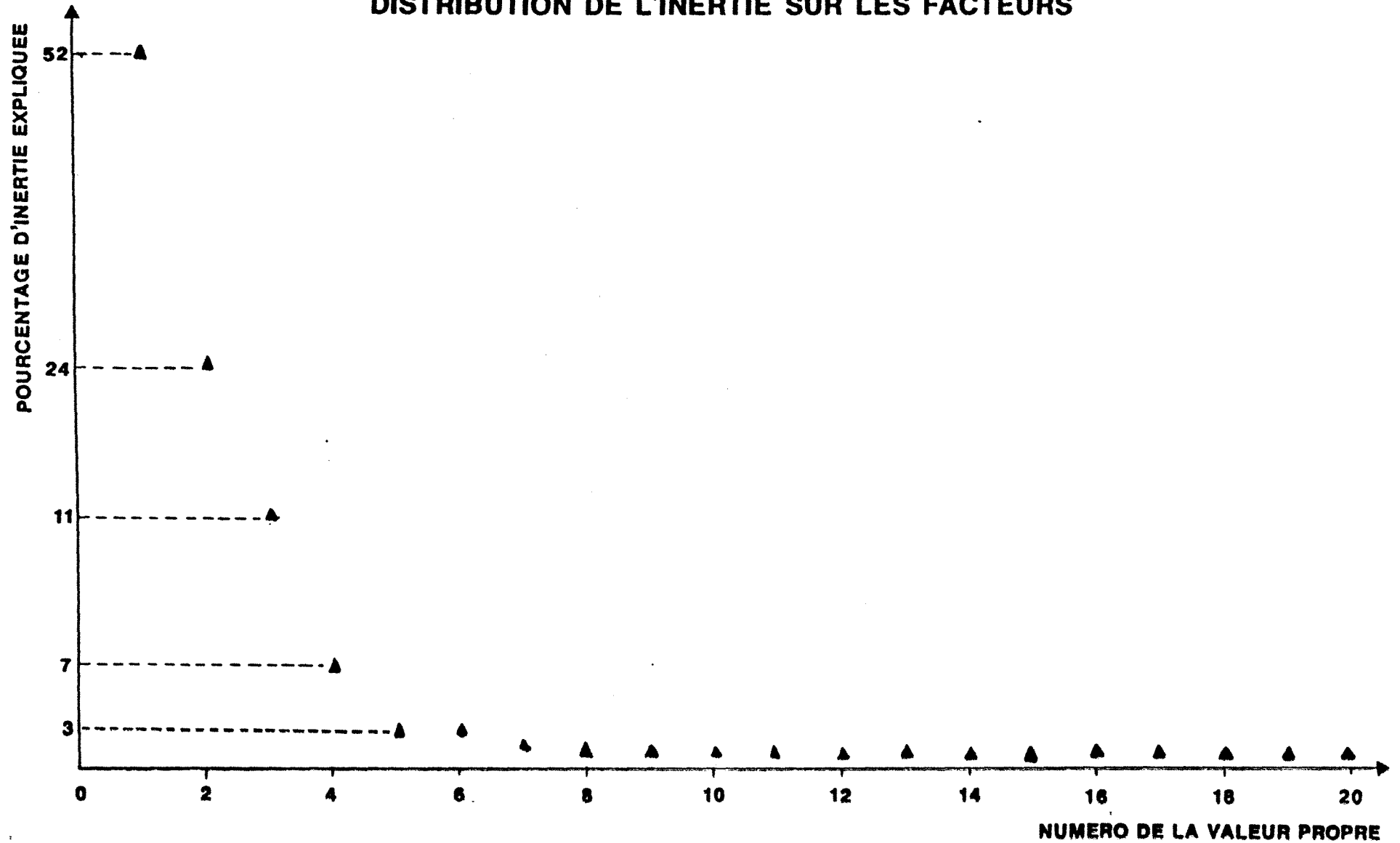
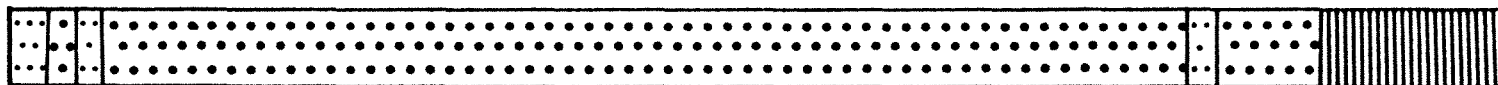
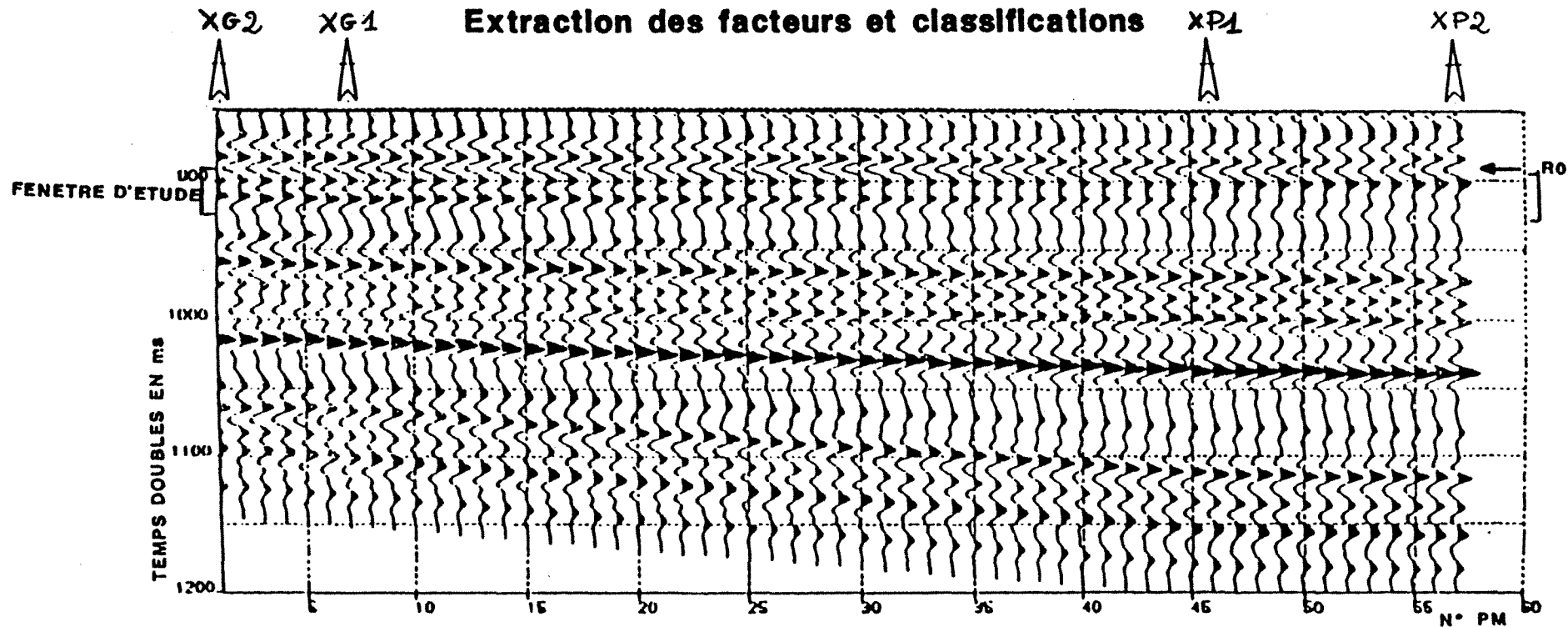
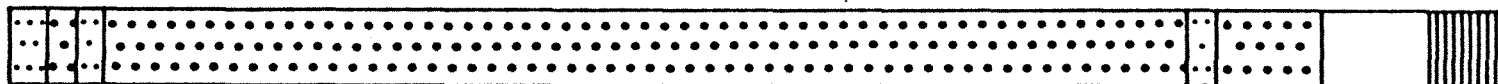


Fig. 166

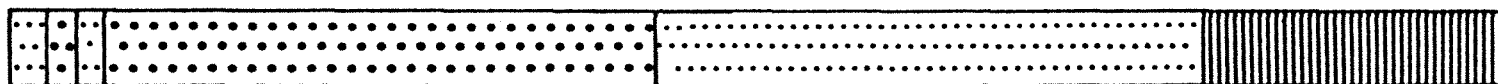
ANALYSE EN COMPOSANTES PRINCIPALES SUR LA FENETRE 892-932ms



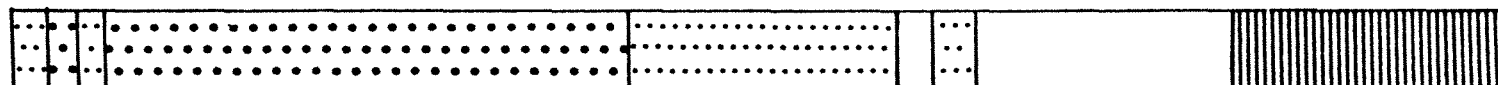
5 FACTEURS 3 GROUPES



5 FACTEURS 4 GROUPES



3 FACTEURS 3 GROUPES



4.2.3. Choix des paramètres

4.2.3.1. Influence du pas d'échantillonnage

Nous avons sous-échantillonné : passage de $\Delta t = 2$ ms à $\Delta t = 4$ ms. La distribution de l'inertie sur les facteurs est pratiquement identique (Figure 168) ainsi que les corrélations entre les facteurs principaux et les amplitudes (voir Tableau 23). La figure 169 permet de comparer les limites de faciès pour 3 ou 4 groupes et 2 facteurs dans les cas de $\Delta t = 2$ ms ou $\Delta t = 4$ ms : ces limites sont identiques.

Ces résultats n'ont rien de surprenant : la quantité d'information contenue dans la trace est déterminée par le pas d'échantillonnage à l'enregistrement, lui-même lié à la bande passante mise en œuvre. Cependant, il fallait vérifier la fiabilité de l'extraction des facteurs.

TABLEAU 23 - CORRELATIONS SIGNIFICATIVES
ENTRE FACTEURS PRINCIPAUX ET VARIABLES
INITIALES (FENETRE 892-932 ms, $\Delta t = 2$ ms ou $\Delta t = 4$ ms)

	$\Delta t = 2$ ms		$\Delta t = 4$ ms	
	CORRELATIONS > 0	CORRELATIONS < 0	CORRELATIONS > 0	CORRELATIONS < 0
FACTEUR 1	AMP 10 (0,96) AMP 20 (0,89)	AMP 14 (- 0,94) AMP 6 (- 0,85)	AMP 10 (0,91) AMP 5 (0,74)	AMP 8 (- 0,94)
FACTEUR 2	AMP 8 (0,83) AMP 2 (0,81) AMP 16 (0,81)	AMP 12 (- 0,93)	AMP 2 (0,67) AMP 9 (0,64)	AMP 6 (- 0,82)

NB. : Les variables AMP_i (4 ms) sont équivalentes aux variables AMP2_i (2 ms).

4.2.3.2. Nombre de facteurs à retenir

D'après les essais exposés dans le paragraphe 4.2.2, la classification semble meilleure (au vu de notre connaissance géologique du modèle) si l'on filtre légèrement l'information : rétention seulement d'environ 80 % de l'inertie initiale (Figure 167 à 3 facteurs et Figure 169 à 2 facteurs). 2 facteurs seront retenus ultérieurement.

ANALYSE EN COMPOSANTES PRINCIPALES SUR LA FENETRE 892-932ms

Influence du pas d'échantillonnage

DISTRIBUTION DE L'INERTIE SUR LES FACTEURS

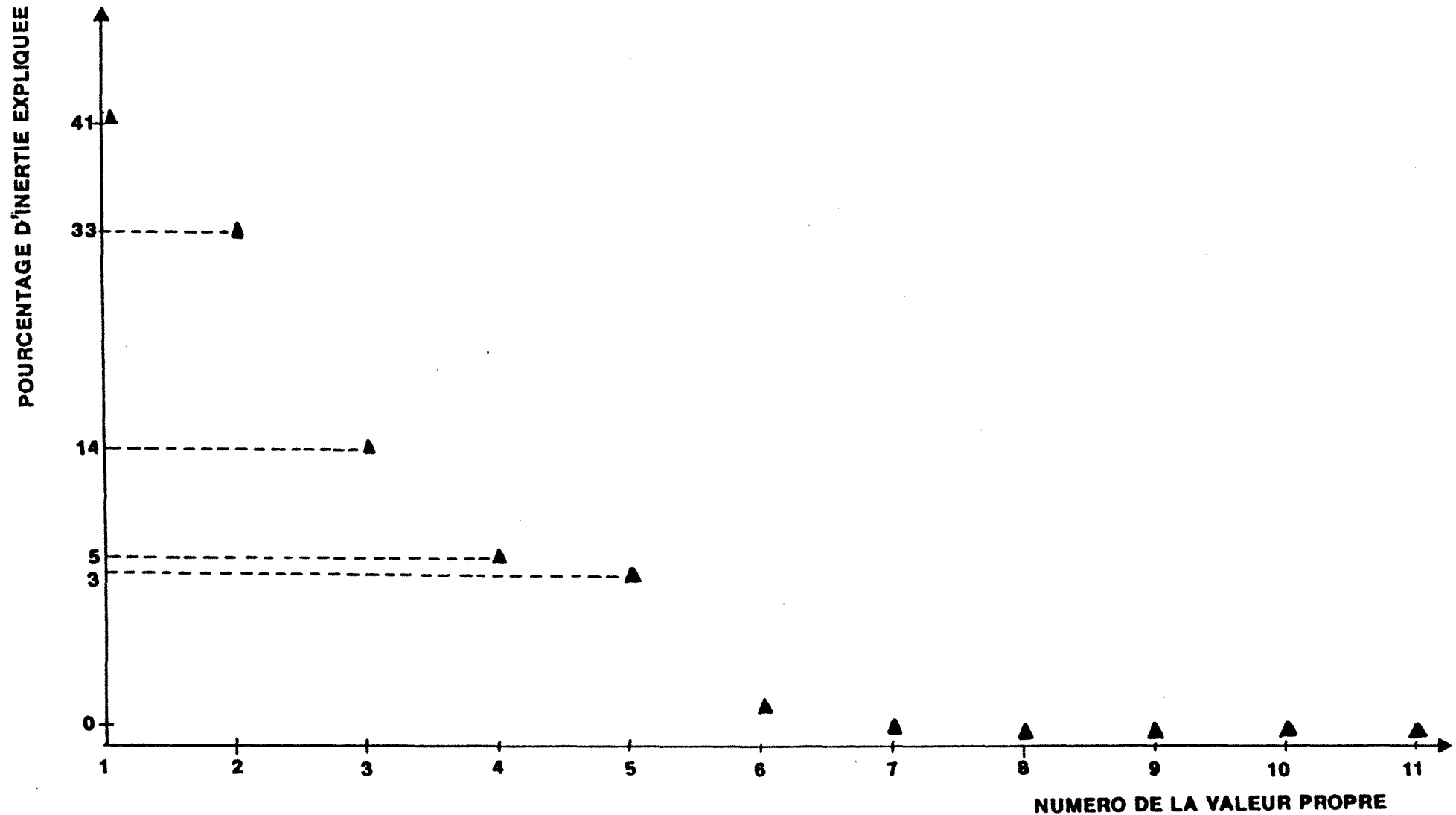
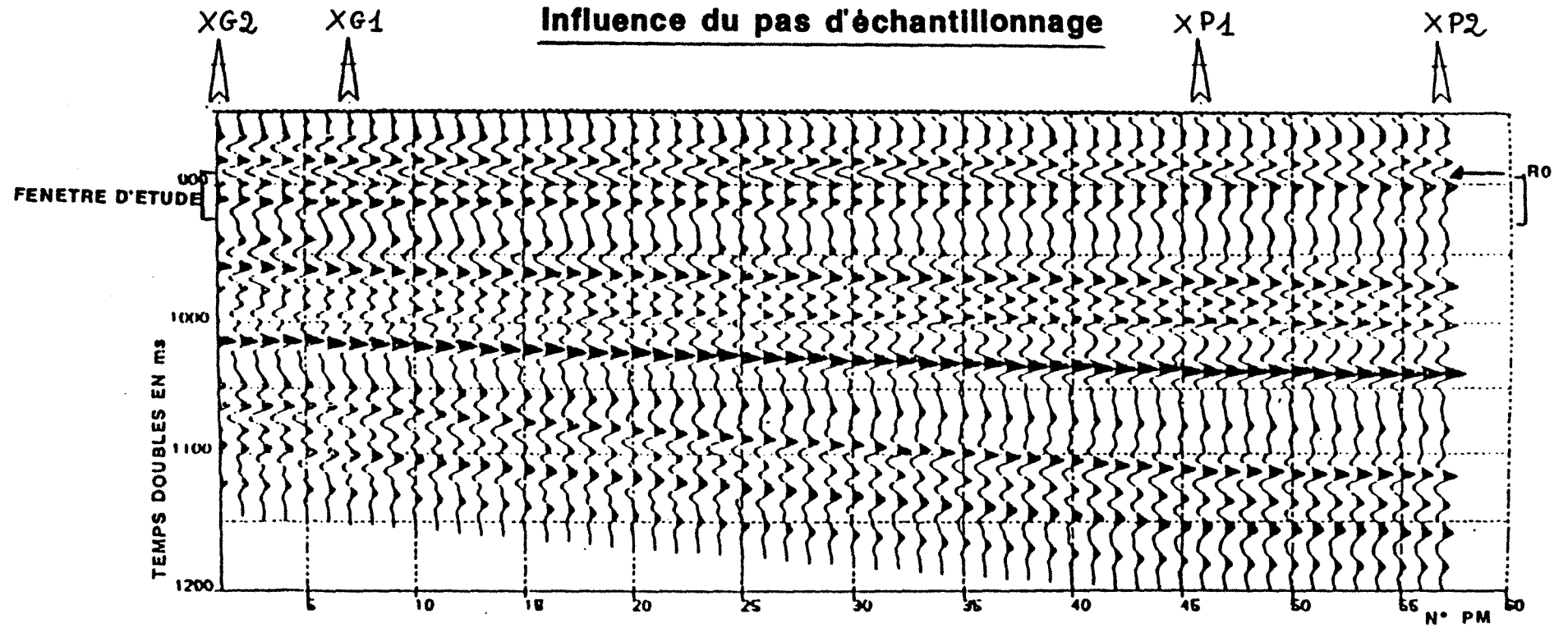


Fig. 168

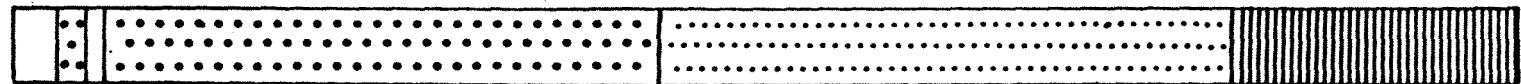
ANALYSE EN COMPOSANTES PRINCIPALES SUR LA FENETRE 892-932ms



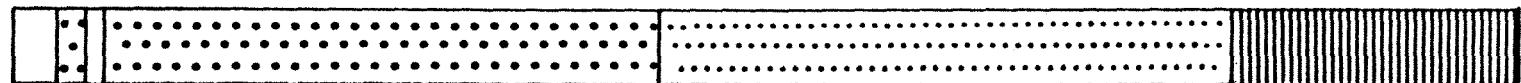
2ms 2 FACTEURS 3 CLUSTERS



4ms 2 FACTEURS 3 CLUSTERS



2ms 2 FACTEURS 4 CLUSTERS



4ms 2 FACTEURS 4 CLUSTERS

4.2.3.3. Diminution de la taille de la fenêtre

Le toit de la fenêtre extraite reste le repère oolithique Xo, la base est passée de 932 ms (essais précédents) à 920 ms (15 échantillons au pas de 2 ms) puis à 912 ms (11 échantillons au pas de 2 ms).

L'inertie expliquée avec les premiers axes (2 ou 3) augmente ce qui n'est pas surprenant puisque la quantité d'information initiale a diminué (Figure 170). Les corrélations entre les facteurs et les variables initiales (AMP_i, amplitude du ième échantillon) évoluent peu (Tableau 24).

**TABLEAU 24- CORRELATIONS SIGNIFICATIVES
ENTRE FACTEURS PRINCIPAUX ET VARIABLES
INITIALES ($\Delta t = 4$ ms)**

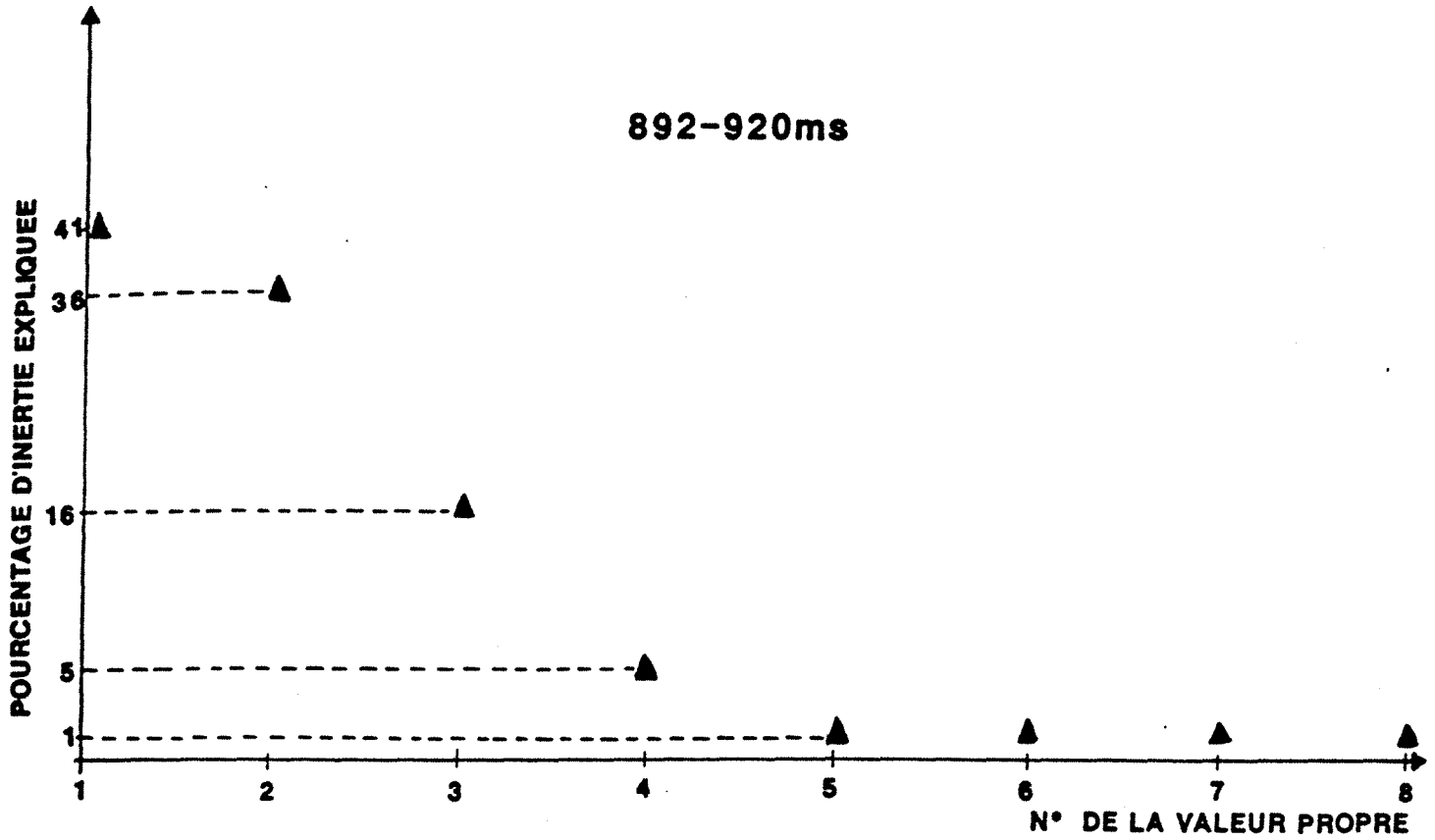
	892 - 920 ms		892 - 912 ms	
	CORRELATIONS > 0	CORRELATIONS < 0	CORRELATIONS > 0	CORRELATIONS < 0
FACTEUR 1	AMP 4 (0,91) AMP 8 (0,86)	AMP 6 (- 0,96)	AMP 4 (0,88)	AMP 6 (- 0,98)
FACTEUR 2	AMP 5 (0,90) AMP 2 (0,85)	AMP 7 (- 0,97)	AMP 2 (0,91)	

Les résultats des classifications à 3 ou 4 groupes sur les 2 premiers facteurs (Figure 171) permettent de juger de l'effet de la réduction de la taille de la fenêtre. Si, pour la fenêtre 892 à 920 ms, les résultats sont analogues à ceux de la fenêtre 892-932 ms, pour la fenêtre plus courte, on commence à voir des anomalies (irrégularités) dans la classification, comparables à celles qui étaient apparues lorsqu'on prenait un trop grand nombre de facteurs. La fenêtre retenue est donc celle de 892-920 ms (15 échantillons) qui, d'après notre connaissance du modèle, n'engloberait pas trop de Bathonien.

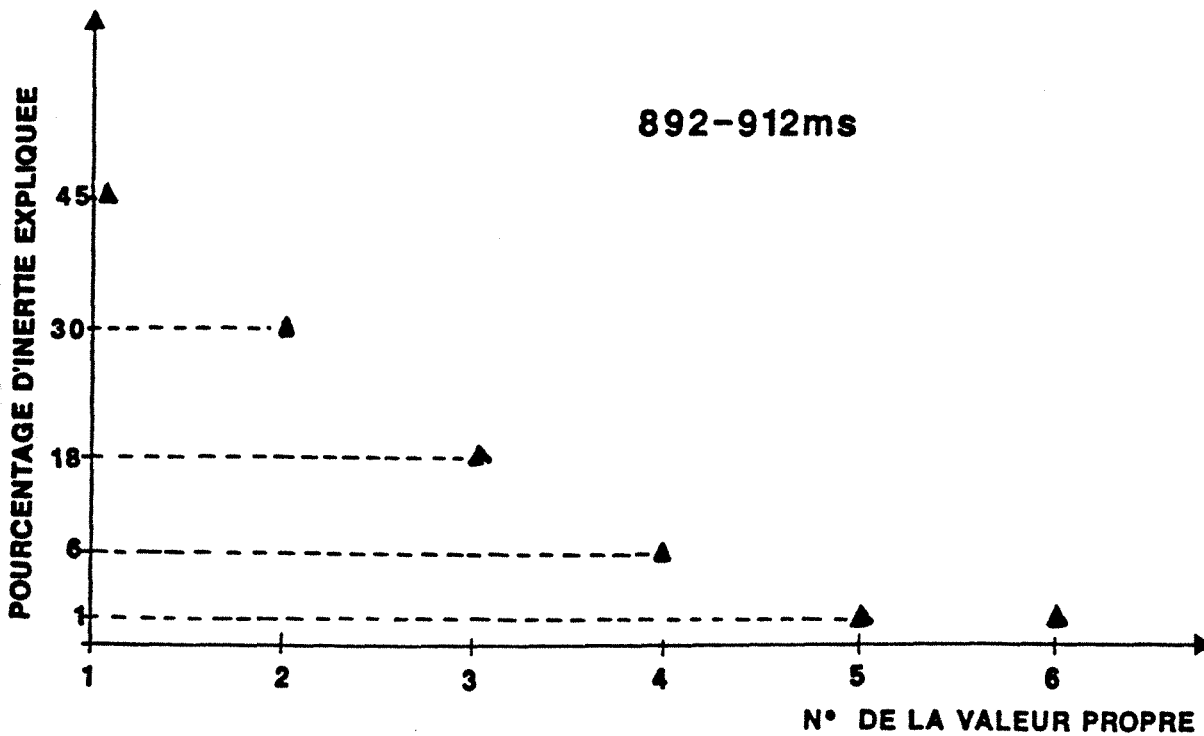
ANALYSE EN COMPOSANTES PRINCIPALES

Diminution de la taille de la fenêtre

892-920ms



892-912ms



ANALYSE EN COMPOSANTES PRINCIPALES

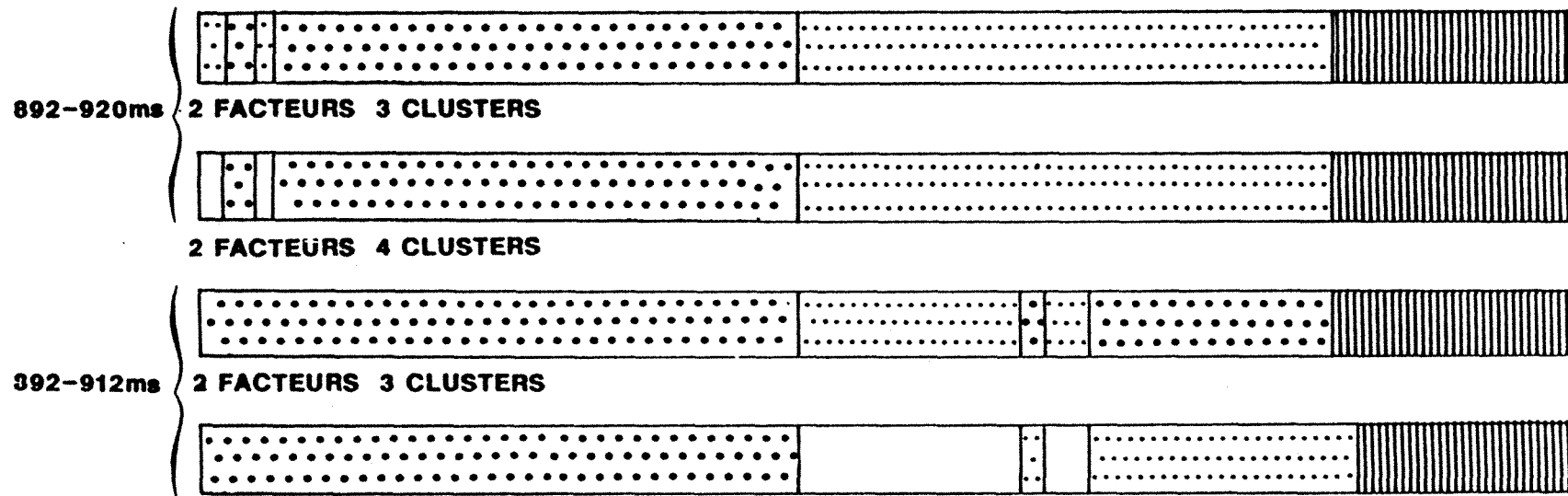
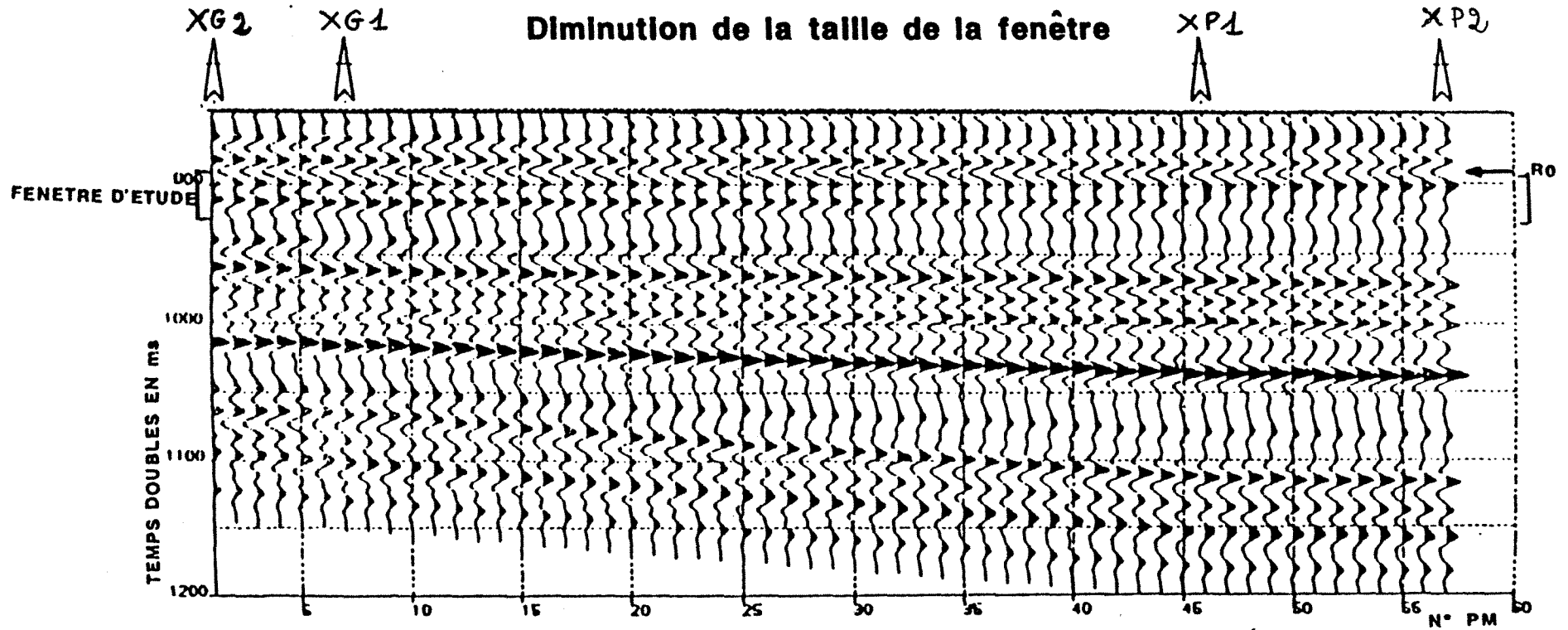


Fig. 171

4.2.3.4. Effet d'une erreur de pointé

Précédemment, nous avons étudié l'influence d'une diminution systématique de la taille de la fenêtre en supposant une extrémité au moins (toit ou mur) correctement pointée. Nous avons ainsi pu trouver une taille de fenêtre optimale. Il s'agit de voir maintenant l'influence d'une erreur de pointé aléatoire de un ou deux échantillons.

Nous avons repris le modèle synthétique, décalé de temps en temps le X_0 (de la valeur d'un ou deux échantillons en plus ou en moins, Figure 173) et extrait une fenêtre de 15 échantillons ; les traces décalées apparaissent en gras sur la figure 173.

L'inertie se distribue très semblablement sur les axes (2 facteurs $\approx 76\%$ de l'inertie totale, 3 facteurs $\approx 93\%$ - Figure 172). Le tableau 25 donne les coefficients de corrélation les plus importants entre les deux premiers facteurs et les variables initiales : d'une part, les axes ne sont pas expliqués par les mêmes variables, d'autre part, l'axe 3 est encore corrélé de façon non négligeable avec certaines variables (corrélations $\approx 0,80$ sur cet axe), ce qui n'était pas le cas précédemment (l'axe 3 contenait peu d'information).

La classification à 3 ou 4 groupes (Figure 173) donne des limites de faciès très dégradées. Toutefois, sur celle à 4 groupes, on peut reconnaître 3 zones (traces n°1 à 29, puis n° 30 à 48 et enfin 49 à 57) où un groupe prédomine, avec, dans chacune des zones, des affectations "aberrantes" qui correspondent justement aux traces décalées.

TABLEAU 25 - CORRELATIONS
ENTRE FACTEURS PRINCIPAUX ET VARIABLES
INITIALES ($\Delta t = 2$ ms, 892-920 ms,) POINTE "DEGRADE"

	CORRELATIONS > 0	CORRELATIONS < 0
FACTEUR 1	AMP 4 (0,94) AMP 11 (0,92)	AMP 15 (- 0,96) AMP 8 (- 0,84)
FACTEUR 2	AMP 9 (0,90)	AMP 13 (- 0,93)

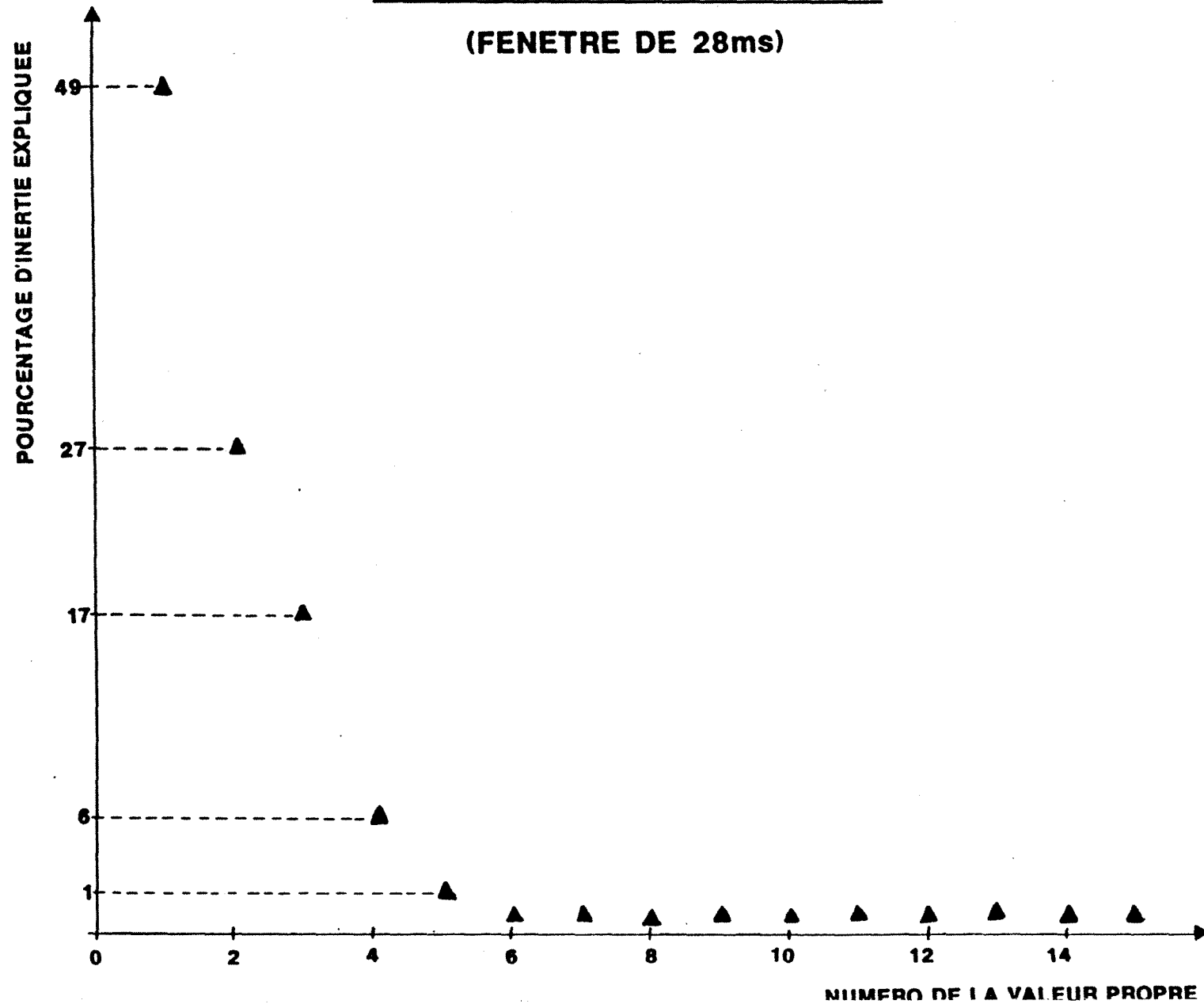
Il est certain qu'une classification de cette qualité, sur un bloc 2D ou 3D, serait difficile à interpréter du fait de l'impossibilité de dégager des zones homogènes cohérentes spatialement.

Ce test souligne l'importance d'un pointé précis et bien calé pour toute étude fine de reconnaissance automatique de faciès. Cette influence dépend, bien sûr, de la qualité des données (si les données sont dès le départ très bruitées, cela aura moins d'importance), du

ANALYSE EN COMPOSANTES PRINCIPALES

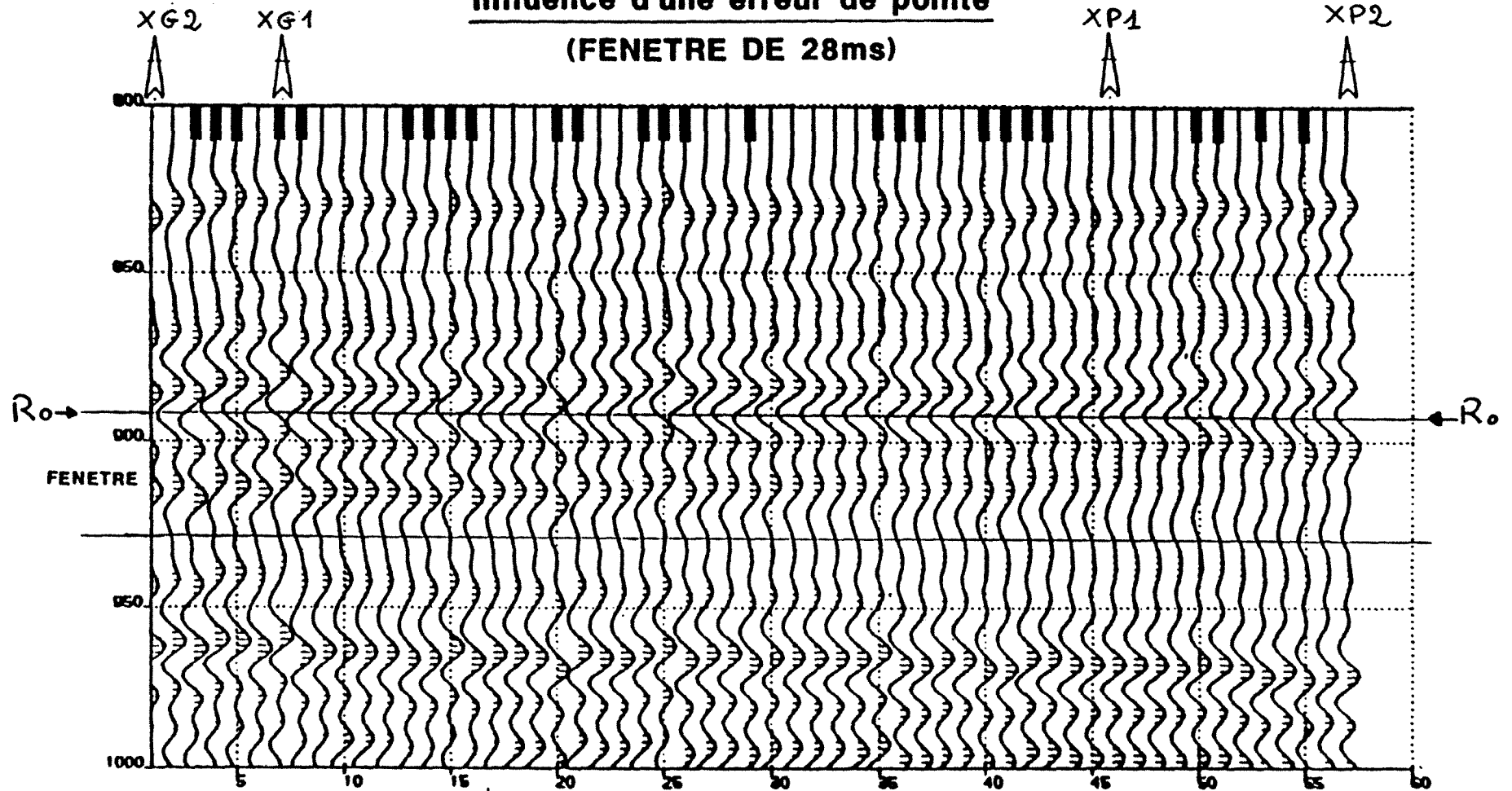
Influence d'une erreur de pointé

(FENETRE DE 28ms)



ANALYSE EN COMPOSANTES PRINCIPALES

Influence d'une erreur de pointé (FENETRE DE 28ms)



2 FACTEURS 3 CLUSTERS



2 FACTEURS 4 CLUSTERS

Fig. 173

contenu fréquentiel (variations verticales de l'amplitude très rapides ou pas), de la longueur de la fenêtre d'analyse (plus la fenêtre est courte, plus le pointé prendra de l'importance).

4.3. Conclusions

L'étude des 57 traces synthétiques du modèle d'évolution latérale entre les puits XG1, XG2, XP1 et XP2 a permis, d'une part, de démontrer la faisabilité de la méthode d'extraction des facteurs pour caractériser la trace sur une courte fenêtre, d'autre part, de dégager des préconisations quant à l'application de la méthodologie statistique de reconnaissance des faciès sur des études fines.

Si les paramètres estimés sur le spectre d'amplitude de la trace ou sur le module du signal analytique se sont montrés efficaces pour la reconnaissance automatique du faciès sur des intervalles d'environ 100 à 200 ms, leur calcul devient plus douteux sur des intervalles fins. La substitution à ces paramètres des facteurs d'une analyse en composantes principales (A.C.P.) sur les amplitudes de la trace est une technique robuste : les facteurs principaux sont des variables non corrélées (d'où l'information qu'elles portent est non redondante) qui permettent de condenser en quelques variables (2 ou 3 axes) les traits majeurs de la trace (2 ou 3 axes remplacent les n échantillons initiaux avec une perte d'information contrôlée).

Cette caractérisation de la trace est efficace puisqu'un algorithme de classification avec ces variables (les composantes principales) délimite des zones homogènes, cohérentes avec notre connaissance du modèle et ceci sur une fenêtre de l'ordre de 28 ms (15 échantillons).

La non-prise en compte de l'information contenue sur les axes d'ordre supérieur paraissait au départ peu compatible avec une reconnaissance fine de la trace. Il semblerait, avec les quelques tests que nous avons réalisés, qu'au contraire on introduit ainsi un effet de filtrage favorable aux algorithmes de classification (les groupes obtenus sont plus stables et homogènes).

L'inconvénient des facteurs par rapport aux paramètres initiaux est la difficulté de leur donner une signification physique simple. Combinaisons linéaires des paramètres initiaux, leurs corrélations avec ces variables est un moyen d'en proposer une interprétation.

Dans une étude aussi fine que celle réalisée sur le modèle synthétique, il est apparu nécessaire de faire plusieurs essais pour régler des paramètres comme **taille de la fenêtre et nombre de facteurs à retenir**.

La taille de la fenêtre n'est pas toujours une donnée du problème : dans le cas présent, seul le repère Xo est calé géologiquement et sismiquement. On ne sait pointer ni le toit du Callovien Inférieur, ni le toit du Bathonien. La taille de la fenêtre peut s'estimer en ordre de grandeur d'après les films synthétiques : avec 40 ms, on englobe déjà largement le Bathonien, il faut donc chercher à la diminuer mais dans certaines limites.

Le modèle synthétique a permis de retenir une fenêtre de 28 ms (15 échantillons) puisque nous connaissons la géologie du modèle et donc les classes de faciès à trouver.

Pour ce qui est du nombre de facteurs, nous avons vu son importance sur les résultats de la classification : l'effet de filtrage provoqué par la prise en compte des seuls premiers facteurs est souhaitable ou pas, en fonction des données. Ceci est à étudier nécessairement, sur un modèle, par exemple.

Dans le cas étudié pour l'analyse de faisabilité, il apparaît qu'il est préférable de se limiter à exploiter les deux premiers axes factoriels qui représentent 80 % de l'inertie du nuage de points (c'est-à-dire 80 % de l'information initiale totale).

Enfin, soulignons qu'avant d'effectuer toute étude de faciès avec cette méthodologie statistique, **il est indispensable d'avoir un marqueur proche de l'objectif bien calé géologiquement et bien défini sismiquement** (pointé précis). Une erreur de pointé a un effet analogue à du bruit dans la classification des traces sismiques (classification "persillée" impossible à interpréter spatialement) : il est bien évident que cet effet est d'autant plus sensible que la fenêtre d'étude est courte.

Ces quelques remarques montrent qu'il est indispensable de passer par l'analyse d'un modèle synthétique pour régler le choix des paramètres, en particulier pour une étude fine.

5. COMPARAISON DE L'UTILISATION DE LA METHODE SPECTRALE PAR MAXIMISATION DE L'ENTROPIE ET DE L'A.C.P. POUR LA CARACTERISATION DES TRACES SISMIQUES

5.1. Introduction

Nous avons vu que l'algorithme de Marple permettait d'obtenir les coefficients du modèle autorégressif d'où une estimation spectrale par la méthode de l'entropie maximale, avec fiabilité, même sur des fenêtres temporelles très courtes. Il faut toutefois avoir correctement estimé l'ordre du modèle autorégressif.

Ces coefficients peuvent remplacer les paramètres spectraux de la méthodologie statistique "fenêtres longues". Bois (1981) propose d'ailleurs les coefficients du modèle autorégressif comme paramètres dans un algorithme de reconnaissance des formes (Figure 161).

Nous les avons comparés avec la caractérisation fournie par les premiers facteurs de l'analyse en composantes principales des amplitudes de la trace, en particulier sur le modèle synthétique du Callovien Inférieur utilisé pour les tests de faisabilité de la méthode par A.C.P. (§ 4.2, troisième partie).

La fenêtre analysée fait 28 ms (15 échantillons) à partir du toit du Xo (Figure 171).

Les meilleurs résultats (par rapport à notre connaissance géologique des données) sont obtenus pour des classifications à 3 ou 4 groupes sur les 2 premiers facteurs de l'A.C.P. (Figure 170).

Vu la taille de la fenêtre, il s'agit d'un bon cas d'application de l'estimation spectrale par maximisation de l'entropie. Nous avons cherché à comparer les classifications obtenues avec les composantes principales (Figure 171) avec celles données avec les coefficients autorégressifs (Figure 175).

5.2. Choix de l'ordre du modèle

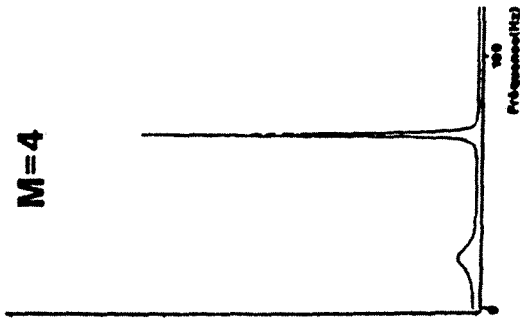
Des essais aux ordres 3, 4, 5 et 6 pour les traces n° 7, 46 et 57 (Figure 174) nous font retenir l'ordre 4.

5.3. Classifications

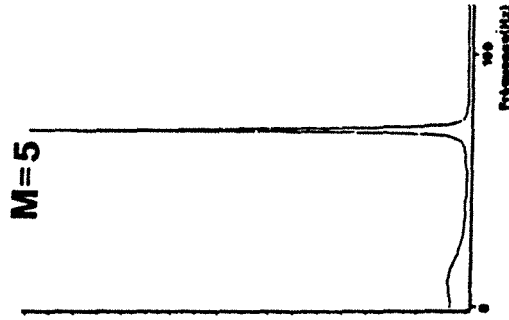
Les variables utilisées pour classer les traces synthétiques sont les 4 coefficients du modèle autorégressif, déterminés par l'algorithme de Marple.

ALGORITHME DE MARPLE

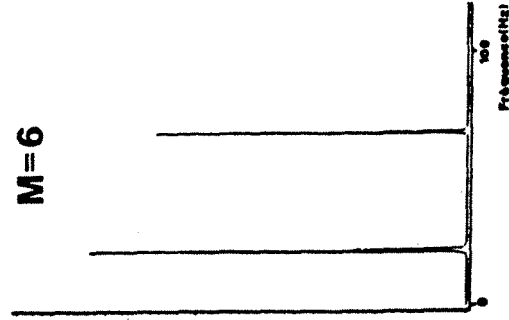
TRACE N° 7



M=4

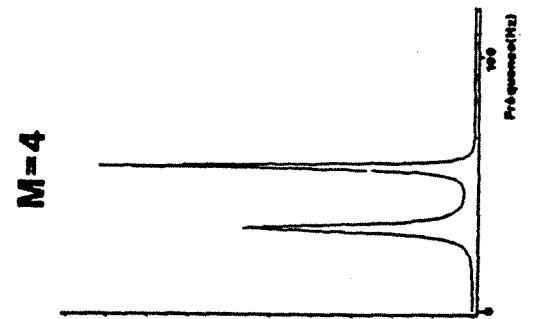


M=5

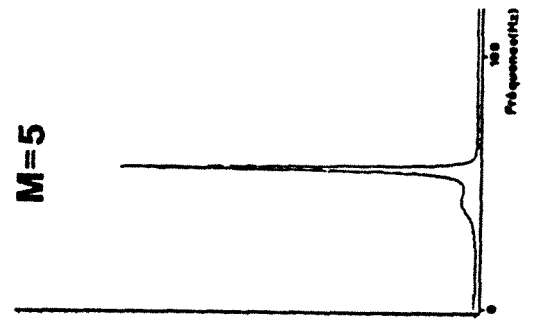


M=6

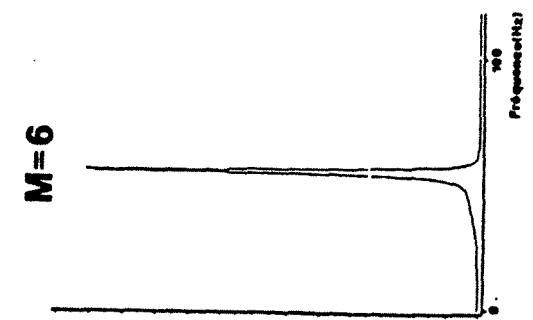
TRACE N° 57



M=4

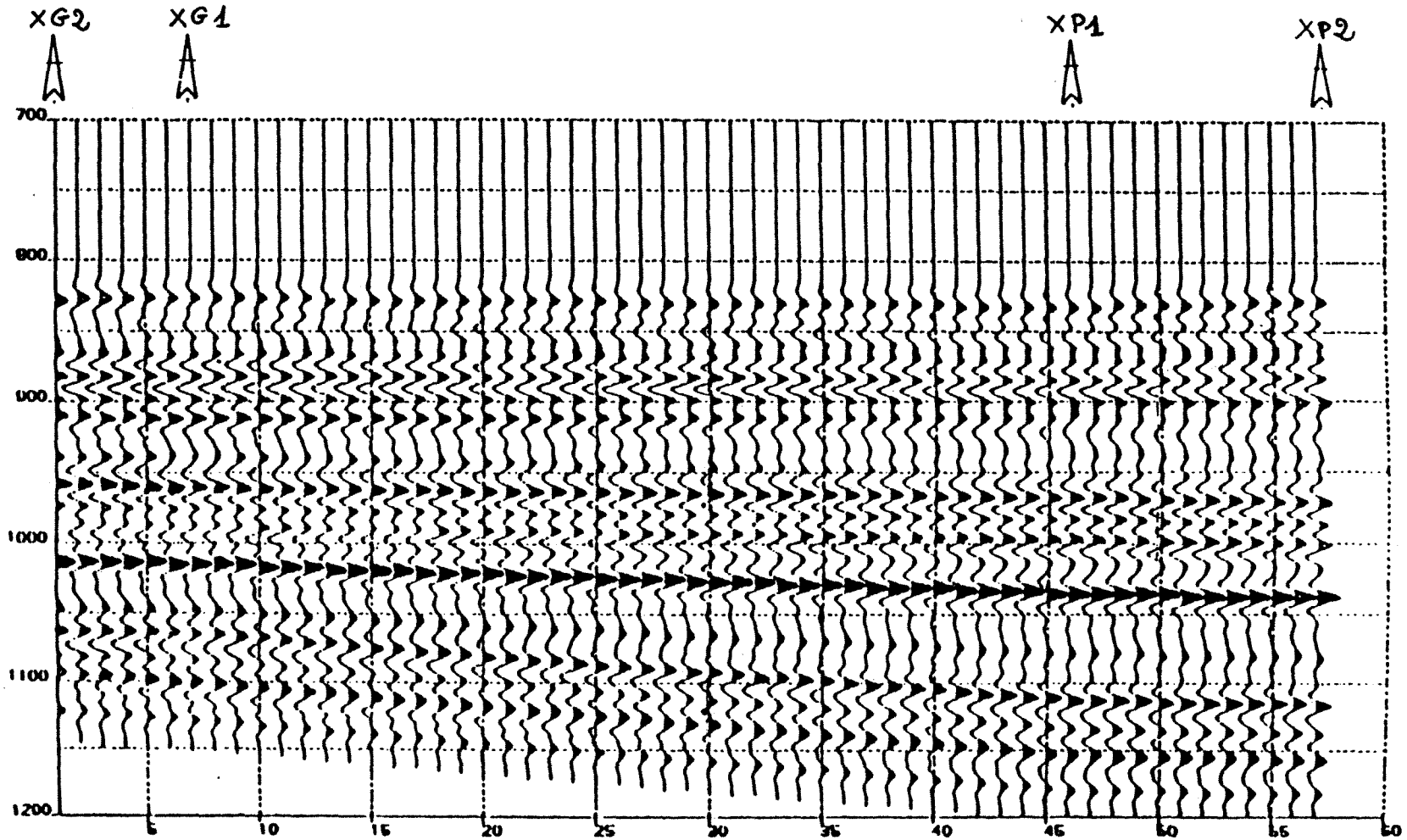


M=5

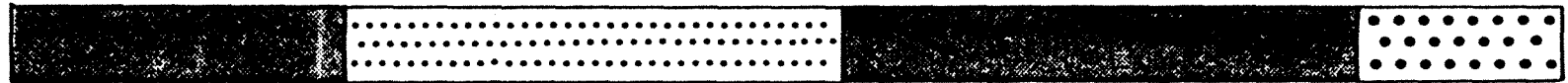


M=6

CLASSIFICATION DES TRACES AVEC LES COEFFICIENTS AUTOREGRESSIFS (M=4)



a



CLASSIFICATION 3 GROUPES VAR: AC1, AC2, AC3, AC4

b



CLASSIFICATION 4 GROUPES VAR: AC1, AC2, AC3, AC4

Des essais avec 3 ou 4 groupes ont été effectués (Figure 175). Trois groupes semblent insuffisants pour délimiter correctement les faciès : toute la zone proche de XP1 (traces n°33 à n°52) est affectée dans le même groupe que les traces proches de XG1, XG2 (n°1 à n°13). A 4 groupes, les limites observées se rapprochent plus de la classification avec les facteurs principaux (comparer la Figure 175.b et la Figure 170.a) mais il existe beaucoup plus d'anomalies dans les affectations : traces n°1 et 2 classées en faciès XP1, traces n°50 et 52 classées en XG alors, qu'à l'œil, on voit assez nettement les différences de caractère sismique.

5.4. Conclusion

Sur cet exemple synthétique du Callovien Inférieur du Bassin de Paris, les résultats des classifications sont proches de ceux obtenus avec les facteurs principaux, mais connaissant le modèle géologique initial, nous pensons que la classification avec les facteurs principaux est plus stable, donc plus fiable (cf. les anomalies signalées précédemment) comme le confirment les indices statistiques mesurant les performances des classifications.

Pour caractériser le faciès sur une courte fenêtre, l'approche mise en œuvre par Bois reste donc une possibilité mais notre façon d'extraire les facteurs principaux nous semble plus robuste (surtout si on considère les difficultés dans le choix de l'ordre du modèle autorégressif et les artefacts qu'un mauvais choix peut entraîner). Il y a peu d'hypothèses sous-jacentes pour effectuer une analyse en composantes principales, une seule matrice à inverser et la signification des axes principaux est clairement exprimée ainsi que la part d'information qu'ils contiennent. Il n'en est pas de même de l'estimation spectrale par la méthode de l'entropie maximale : la stabilité et la validité des résultats dépendent beaucoup du réglage de paramètres qui interfèrent entre eux de façon complexe.

En particulier, l'application des critères déjà cités pour le choix de l'ordre "optimal" montre des variations importantes de la valeur optimale d'une trace à l'autre. Certes, ces critères semblent assez contestables (cf. l'absence de consensus qui existe sur ce sujet), mais il n'est pas possible d'introduire, dans notre méthodologie statistique, un nombre de paramètres variable d'un individu à l'autre. Par ailleurs, les résultats obtenus par Bois avec les paramètres du modèle autorégressif ne semblaient pas très performants pour la caractérisation des traces sismiques. La méthode spectrale par maximisation de l'entropie peut être particulièrement intéressante pour certains signaux mais son application aux traces sismiques réelles soulève un certain nombre de difficultés :

- difficulté de trouver l'ordre optimal du modèle autorégressif (les critères proposés jusqu'à présent sont loin d'être satisfaisants) ;
- compromis délicat à trouver entre l'ordre, le nombre de points et la résolution fréquentielle souhaitée ;

- influence de la largeur de la bande spectrale ainsi que de la distribution des pics principaux (écarts entre ces pics) ;
- impossibilité d'interpréter la signification des hauteurs relatives des différents pics.

La complexité de l'interprétation des résultats augmente avec le nombre de fréquences pures composant le spectre. Un mauvais choix des paramètres peut conduire à des résultats tout à fait aberrants.

Pour toutes ces raisons, nous avons privilégié la méthode "analyse en composantes principales de la succession des amplitudes de la trace" pour caractériser les signaux sismiques sur de courtes fenêtres. Dans ce cadre là, nous allons présenter la méthodologie statistique modifiée.

6. METHODOLOGIE STATISTIQUE SUR "FENETRE COURTE"

6.1. Description de la méthodologie statistique

6.1.1. Introduction

Les techniques utilisées dans la méthodologie statistique développée pour l'analyse statistique des faciès sismiques sur longues fenêtres restent encore les outils essentiels. En particulier, si une information a priori est disponible (données géologiques aux puits), **l'analyse factorielle discriminante** reste la méthode clé. Des **classifications automatiques** seront aussi utilisées pour corroborer les résultats de l'analyse discriminante.

Par contre, le problème de la sélection des paramètres pertinents pour l'analyse des faciès n'est plus aussi capital. Les paramètres caractérisant la trace sur longue fenêtre sont très nombreux et fortement redondants. L'information a priori guide, au cours d'une phase d'apprentissage, le choix des paramètres pertinents. Au contraire, les composantes principales extraites à partir de l'étude de la succession des amplitudes de la trace sont décorréélées (orthogonales). Elles représentent, de façon non redondante, intégralement l'information sismique **initiale**. Cependant, il peut être intéressant de filtrer une partie de cette information pour éliminer le bruit des traces sismiques, par exemple, ou toute autre information peu corrélée à l'information géologique a priori que l'on cherche à retrouver à travers l'analyse du faciès sismique. Le problème de la sélection des paramètres pertinents est ainsi remplacé par celui du **choix du nombre de facteurs** de l'analyse en composantes principales à utiliser pour représenter l'information sismique.

Par ailleurs, **la taille de la fenêtre sismique à analyser** devient un paramètre important. La prise en compte de quelques échantillons en plus ou en moins a peu d'influence sur une fenêtre de grande taille, d'autant plus que les paramètres représentant la trace sont des mesures globales de ses caractéristiques. Sur une fenêtre de petite taille, caractérisée directement par les valeurs des amplitudes des échantillons, ceci n'est plus vrai, comme nous l'avons montré au cours de l'étude de faisabilité (§ 4.2).

L'adaptation de la méthodologie réside dans l'utilisation des techniques multivariées déjà citées (classifications automatiques et analyses factorielles discriminantes) pour guider le choix des deux paramètres suivants :

- taille de la fenêtre d'analyse (et limites),
- nombre de facteurs de l'analyse en composantes principales à utiliser pour la caractérisation des traces.

6.1.2. Cas où une information géologique est disponible

Dans la majorité des cas, les problèmes fins d'analyse du faciès sismique correspondent à des études de réservoir pour des champs au **stade du développement**. Des puits, en nombre assez important, sont donc disponibles.

Si le **nombre de puits** est assez **grand** (≈ 30), on peut utiliser les **traces sismiques synthétiques** de ces puits (calculées à partir des diagraphies Sonic et Densité, ainsi que d'un signal conforme avec le signal évalué sur les traces réelles) pour choisir les deux paramètres de représentation des traces.

Une première fenêtre d'étude est définie sur les données sismiques synthétiques à partir du calage avec les puits donnant la position temps au toit du réservoir ou d'un marqueur de référence, de façon plus générale. L'analyse en composantes principales des amplitudes est réalisée. Des tests de classification automatique des traces synthétiques avec différents nombres de composantes principales sont alors réalisés. Ces tests sont interprétés avec des indices quantitatifs permettant de mesurer la qualité des regroupements trouvés. En plus de ces indices, nous étudions aussi la façon dont les groupes de traces synthétiques permettent de retrouver les faciès géologiques. Un regroupement de traces est d'autant plus satisfaisant qu'il rassemble les traces synthétiques de puits dont les caractéristiques géologiques sont voisines.

Ces essais sont ensuite recommencés en faisant varier la taille de la fenêtre d'analyse. L'influence sur les résultats de l'analyse en composantes principales, puis sur la qualité des regroupements des traces ainsi que sur leur signification géologique, est analysée.

Cet ensemble de tests permet finalement de définir la meilleure fenêtre d'analyse ainsi que le nombre de composantes principales à utiliser. "Meilleure" se définit par rapport à la séparation statistique des groupes de traces trouvés ainsi que par rapport à la corrélation qualitative entre ces groupes et les faciès géologiques.

Ces paramètres seront ensuite utilisés pour l'étude des **traces sismiques réelles**.

Dans le cas où le **nombre de puits** est **insuffisant** pour assurer une valeur statistique aux analyses multivariées des traces synthétiques associées, le choix des paramètres se fera **directement** sur les traces réelles. La façon de procéder est analogue – boucles sur la fenêtre analysée, sur le nombre de composantes principales utilisées, sur le nombre de groupes demandés en classification automatique – mais les calculs sont beaucoup plus lourds car le nombre de traces réelles à traiter peut être très grand. On peut ne faire porter ces tests que sur une partie de la population mais la sélection des données est difficile à réaliser (aucun critère, a priori).

L'analyse des traces sismiques synthétiques des puits est préférable – lorsqu'elle est possible – car elle permet une meilleure interprétation en faciès géologiques des groupes de traces.

Après avoir défini la fenêtre temporelle d'étude des traces, cette fenêtre est extraite des données réelles, le pointé du marqueur de référence ayant été fait.

L'analyse en composantes principales est réalisée sur les amplitudes de chaque trace, et les traces sont représentées dans l'espace généré par les k premières composantes principales, k ayant été défini au cours des tests sur les traces synthétiques (Figure 162).

Des traces sismiques d'apprentissage sont alors sélectionnées au voisinage de puits appartenant à différents faciès géologiques. **L'analyse factorielle discriminante** permet d'évaluer la séparation entre les groupes de traces d'apprentissage ainsi que le pouvoir prédictif de ces classes. Le choix des traces d'apprentissage peut être remis en question par les résultats de l'analyse discriminante.

Une fois les groupes d'apprentissage définitivement fixés, l'étape prédictive de l'analyse discriminante permet d'affecter toutes les traces anonymes par rapport aux groupes a priori. Outre une carte de faciès sismiques, on obtient aussi une carte des probabilités de bonne affectation des traces, à utiliser conjointement avec la carte des faciès pour l'interprétation des zones.

Comme dans la méthodologie pour les fenêtres temporelles longues, les résultats de l'analyse discriminante sont confirmés par des **classifications automatiques** où aucune information a priori n'est utilisée.

Des essais où le nombre de groupes demandés varie permettent d'obtenir plusieurs cartes de faciès sismiques à comparer avec la carte des faciès obtenue en analyse discriminante. Les principales tendances doivent se confirmer pour assurer la fiabilité de la reconnaissance des faciès sismiques. Des groupes supplémentaires (par rapport au nombre de faciès géologiques reconnus) peuvent apparaître, traduisant soit des zones sismiques bruitées, pas nécessairement en rapport avec la zonation géologique, ou des faciès sismiques liés à des faciès géologiques qui n'ont pas été reconnus par des puits. Il peut aussi arriver que certains groupes sismiques a priori disparaissent. C'est-à-dire que les faciès géologiques, bien que différents, ne s'expriment pas par des réponses sismiques différentes. On aurait donc distingué trop de groupes a priori dans la phase d'apprentissage de l'analyse discriminante.

Si aucune corrélation n'apparaît entre les cartes de faciès en classification automatique et celle en discrimination, il faut remettre en question la signification géologique des traces sismiques. Dans ce cas là, il est préférable de retenir comme résultats de l'analyse des faciès la (ou les) carte(s) obtenue (s) en classification automatique et de se garder d'en proposer une interprétation géologique.

Les phases les plus longues de la méthodologie sont :

- la sélection des paramètres de représentation des traces (fenêtre d'étude et nombre de facteurs) sur les traces synthétiques ;
- le choix des traces d'apprentissage pour l'analyse discriminante.

Les autres phases sont rapides ; par ailleurs, la possibilité d'obtenir **rapidement** plusieurs cartes de faciès sismiques (avec ou sans information a priori, nombre de groupes sismiques variable) est un des points forts de la méthode.

6.1.3. Cas où aucune information géologique n'est disponible

Le choix des paramètres de représentation des traces se fait simultanément avec les essais de classification automatique, avec un nombre variable de groupes, **de l'ensemble des traces réelles étudiées**. La qualité des groupes trouvés ne peut se juger que d'un point de vue statistique, aucune information géologique n'étant disponible pour interpréter leur signification.

Eventuellement, pour confirmation, une étape d'analyse discriminante peut être réalisée en utilisant comme traces d'apprentissage des traces définies à proximité des centres des groupes les plus significatifs statistiquement dans les tests de classification automatique.

6.2. Discussion des limites de la méthode

La plupart des limites de la méthode sont les mêmes que celles exposées pour la méthodologie "fenêtre temporelle longue" :

- influence des terrains "hors fenêtre d'analyse" sur la réponse sismique étudiée,
- validité de l'information géologique a priori (problème, en particulier, de la détermination des coupures dans le continuum des propriétés géologiques, qui sont à l'origine de la notion de **faciès géologique**),
- choix des traces d'apprentissage (reposant sur l'hypothèse que les caractéristiques des traces sismiques au voisinage des puits reflètent les propriétés géologiques des puits),
- difficulté de définir, autrement que par des itérations sur les procédures statistiques, le nombre de groupes existant dans la population.

Là encore, la signification géologique des groupes sismiques peut être validée en réalisant un modèle sismique synthétique des principales variations géologiques, que l'on analyse indépendamment ou conjointement avec les traces synthétiques des puits (cf. application, § 6.3).

La définition de la taille et des bords (limites) de la fenêtre d'analyse prend une importance capitale dans l'analyse des courtes fenêtres. Nous avons déjà évoqué la nécessité d'un pointé précis d'un horizon marqueur pour la zone étudiée et les conséquences d'une erreur dans ce pointé, dans le paragraphe 4 concernant la caractérisation de la trace par l'A.C.P. des amplitudes.

Ce problème du **pointé précis** est intimement lié à la notion de faciès sismique **fin** pour les études à **vocation réservoir**. Il est clair que, plus l'objectif recherché est une analyse précise d'un problème géologique fin, plus un nombre important de puits est souhaitable afin de permettre, au préalable, de mettre en place le ou les réservoirs sur la section sismique et d'étudier les variations géologiques rencontrées sur le champ.

Une telle analyse se situe donc plutôt au **stade de développement d'un champ**.

Si la qualité des données sismiques ou des données de puits est insuffisante pour assurer avec fiabilité le pointé de la zone des réservoirs, il est illusoire de procéder à une analyse fine du faciès sismique.

6.3 Une application de la méthodologie statistique à un problème de réservoir

6.3.1. Cadre géologique de l'étude

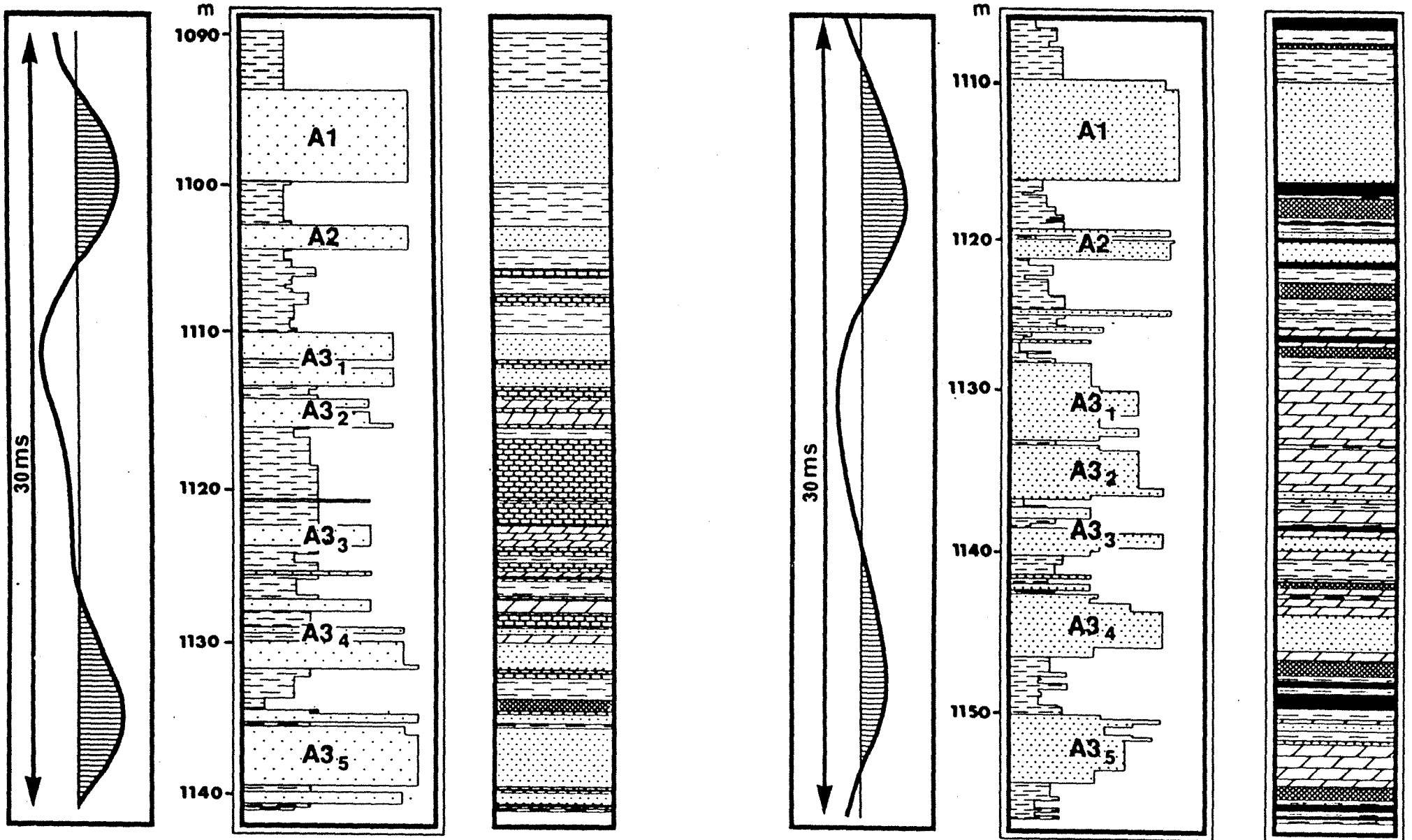
Le champ étudié produit de l'huile à partir de réservoirs carbonatés situés dans l'Albien Supérieur, qui constituent l'unité appelée A. Cette unité est découpée (Figure 176) en A1, A2 et A3 ; la zone A3 constitue le principal de la production du champ (65 %) et totalise 55 % de l'accumulation.

L'Albien Supérieur est un réservoir multi-couches typique : il y a indépendance entre le A1, le A2 et le A3.

La partie inférieure du réservoir (A3) correspond à des dépôts littoraux de haute énergie alors que la partie supérieure correspond à des dépôts intertidaux de faible énergie (A2) qui amorcent la régression généralisée du Cénomani. La continuité latérale de ces dépôts est excellente sur tout le champ, ce qui n'est pas le cas pour les dépôts du A3 (en particulier, les faciès de haute énergie sont plus développés sur les puits 3 et 1 que sur les puits 2).

WELL P6

WELL 142



SEALS

SHALES

RESERVOIRS

Fig. 176

Les faciès réservoirs sont des grès, grès dolomitiques et dolomies vacuolaires. Les meilleurs réservoirs sont des réservoirs gréseux qui ont une porosité intergranulaire élevée et une bonne perméabilité ; les dolomies vacuolaires (milieu forte énergie) ont des caractéristiques variables, fonction de la taille et des interconnexions des vacuoles provenant de la dissolution d'éléments figurés. La dolomie cristalline (dépôt de faible énergie) a une porosité intercrystalline faible qui peut être localement améliorée par la dissolution, la perméabilité est faible.

Les couvertures sont des dolomicrites, des argiles vertes et de l'anhydrite plus ou moins abondante.

La structuration halocinétique intervient dans la répartition des qualités réservoirs : les bons réservoirs sont plutôt localisés au top de la structure et au Nord-Ouest de l'accident synsédimentaire (Figure 178) ; cette structuration joue aussi en privilégiant les phénomènes diagénétiques de dissolution. Les réservoirs correspondent à des progradations du Nord-Est vers le Sud-Ouest.

L'interprétation sismique a mis en évidence un système de faille-flexure au Nord (Figure 178) lié à un changement de faciès des séries. Un grand accident synsédimentaire Nord-Est/Sud-Ouest (Figure 178) provoque un épaississement important des séries sous le A (de 60 m à plus de 100 m), son rejet est très faible au toit de l'Albien.

La définition de l'échelle de travail pour l'analyse du faciès sismique est un compromis entre le problème géologique posé et le pouvoir de résolution de la sismique.

Le problème posé par les producteurs pour ce champ est l'explication de barrières de perméabilité, au niveau du A3, mises en évidence au cours de tests et durant la production. Leur souhait serait d'obtenir des cartes de faciès géologiques, dérivés de la sismique, **pour l'ensemble A3**, pouvant expliquer les variations de porosité ou de perméabilité rencontrées. Le A3 fait entre 30 m (au top de la structure) et 60 m (compartiment affaissé Sud-Est) ; cet ensemble essentiellement carbonaté présente des vitesses rapides (en moyenne 4 000 m/s).

Calculons la tranche temps double sismique correspondante :

$$30 \leq \text{Epaisseur} \leq 60 \text{ m}$$

$$15 \text{ ms} \leq \text{Temps double (ms)} \leq 30 \text{ ms.}$$

Le pouvoir de résolution $(\approx \frac{\lambda}{2})$ vaut environ :

$$PR \approx \frac{\lambda}{2}, \text{ avec } \frac{vT}{2} = \frac{v}{2f} \text{ il vient :}$$

PR $\frac{4000}{2 \times 50} = 40$ m pour un signal dont la bande passante serait centrée sur 50 Hz.

Le pouvoir de détection $\left(\approx \frac{\lambda}{10}, \frac{\lambda}{8} \right)$ serait de l'ordre de 8 à 10 m.

Le A3 n'est pas résolu par la sismique mais il peut être détecté (ses variations peuvent influencer le signal). Par contre, sur tout le compartiment haut (au "top" et Nord-Ouest), la taille de la fenêtre qu'il faudrait étudier (15 ms) est trop fine. Avec la méthode de traitement des traces à partir de leur représentation temporelle, il est maintenant possible de descendre la taille de la fenêtre jusqu'à environ 20 ms.

L'étude ne peut donc porter exclusivement sur le A3. La totalité du réservoir A (A1 + A2 + A3) fait entre 60 m (compartiment haut) et 120 m (compartiment épaissi par l'accident synsédimentaire). La fenêtre d'étude (≈ 30 ms) rentre alors dans le cadre de la méthodologie de caractérisation du faciès.

Nous allons donc étudier **globalement l'unité A** et définir un espace géologique à partir de variables géologiques synthétiques qui seront alors compatibles avec le caractère global de la sismique, les échelles de travail dans les deux domaines (sismique et géologique) seront conformes.

L'analyse globale du A met en évidence l'individualisation nette du "top" de la structure (autour du puits P6) et de la partie Sud-Ouest autour de 233.

Sur tout le A, cette zone centrale est beaucoup plus carbonatée et se caractérise par :

- des épontes supérieures à dominante dolomitique et non pas gréseuse comme ailleurs ;
- des réservoirs plutôt gréseux que dolomitiques : en fait, les grès ne sont pas tellement plus abondants que sur les autres puits, mais il y a beaucoup moins de dolomies poreuses, ce faciès étant remplacé (peut-être pas au sens géologique) par les dolomies calcaires ;
- l'apparition du faciès "dolomies calcaires", très localisé au "top" de la structure ;
- des couvertures argileuses extrêmement peu représentées, par rapport au reste du champ.

Puis nous avons étudié **séparément** les trois principales unités, du point de vue lithologies dominantes et porosité moyenne.

En particulier, **l'étude du A3** a révélé que cet ensemble (les unités A3₁, A3₂, A3₃, A3₄ et A3₅) ont été regroupées car leur distinction est peu nette par rapport aux unités précédentes A1

ou A2) présente des variations faciologiques très importantes, la zone la plus épaisse et la plus poreuse étant au centre ("top" de la structure) et au Sud-Ouest. Les dolomies calcaires (faciès 3) se développent sur cette zone de façon presque superposable. Les réservoirs dolomitiques (faciès 4) sont nettement amincis sur la zone centrale ainsi qu'aux puits 233 et 251. Par contre, les réservoirs gréseux sont abondants au "top" de la structure et au Sud-Ouest.

Les couvertures argileuses sont très anecdotiques ($\approx 10\%$) sauf dans la zone centrale. Les couvertures dolomitiques témoignent en négatif des zones où l'épaisseur utile est forte : elles sont abondantes dans le compartiment Sud-Est.

La porosité du A3 varie beaucoup entre 15 et 30 % et ces variations se superposent très bien aux variations lithologiques. Les plus faibles porosités sont au Nord et dans le compartiment Sud-Est affaissé. La zone de bonne porosité s'étend sur celle des réservoirs gréseux et des dolomies calcaires "remplaçant" les réservoirs dolomitiques.

L'essentiel des variations du A est donc localisé au niveau du A3. Les unités A1 et A2 s'inscrivent dans une logique légèrement différente de celle du A3, mais leurs caractéristiques sont à peu près constantes sur le champ. Ceci est très avantageux pour la déduction des caractères du A3, à travers l'étude des **propriétés sismiques globales du A**. Les variations de la réponse sismique du A traduiront plutôt les variations du A3 puisque celles-ci sont prépondérantes. Le A3 présente une opposition forte entre le "top" de la structure (faciès gréseux très poreux accompagnés de dolomies calcaires), le Nord (réservoirs dolomitiques moins poreux) et la zone Sud-Est (faciès surépaissis et plus argileux).

6.3.2. Données disponibles

6.3.2.1. Données géologiques

Nous disposons de 64 puits pour lesquels des traces sismiques synthétiques pouvaient être calculées (données diagraphiques disponibles). Par contre, l'analyse géologique de seulement 40 puits était disponible.

Une analyse en composantes principales sur les données géologiques permet de visualiser les regroupements de puits exprimant les faciès géologiques majeurs (Figure 177).

Les variables analysées sont récapitulées dans le tableau 26 ;

HT est l'épaisseur totale du A ;

HU est son épaisseur utile ;

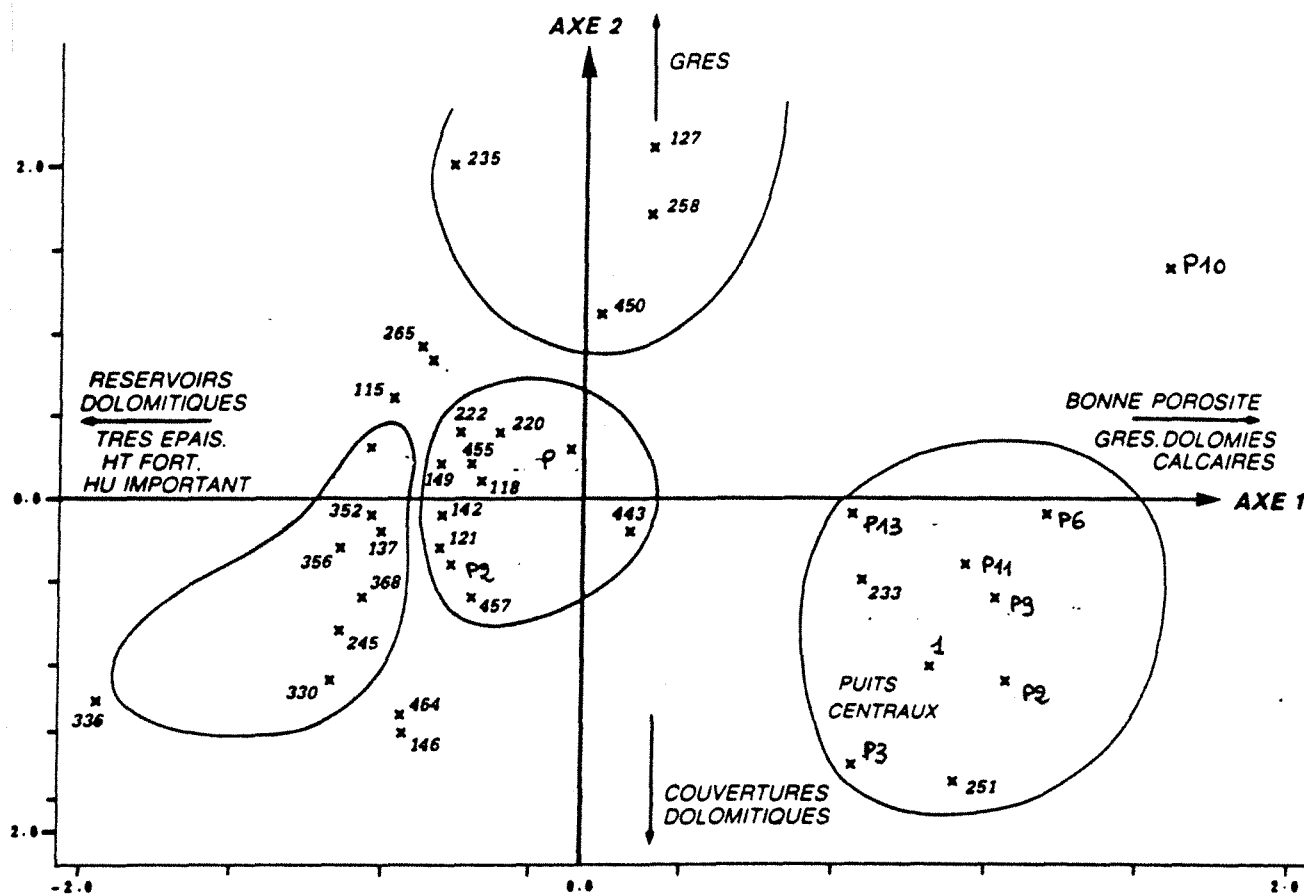
V1 à V5 sont les pourcentages des cinq lithologies dominantes (V1, couvertures dolomitiques, V2, couvertures argileuses, V3, dolomies calcaires, V4, réservoirs

dolomitiques, V5, réservoirs gréseux) ; PHI1 à PHI3 sont les porosités moyennes des unités A1, A2 et A3.

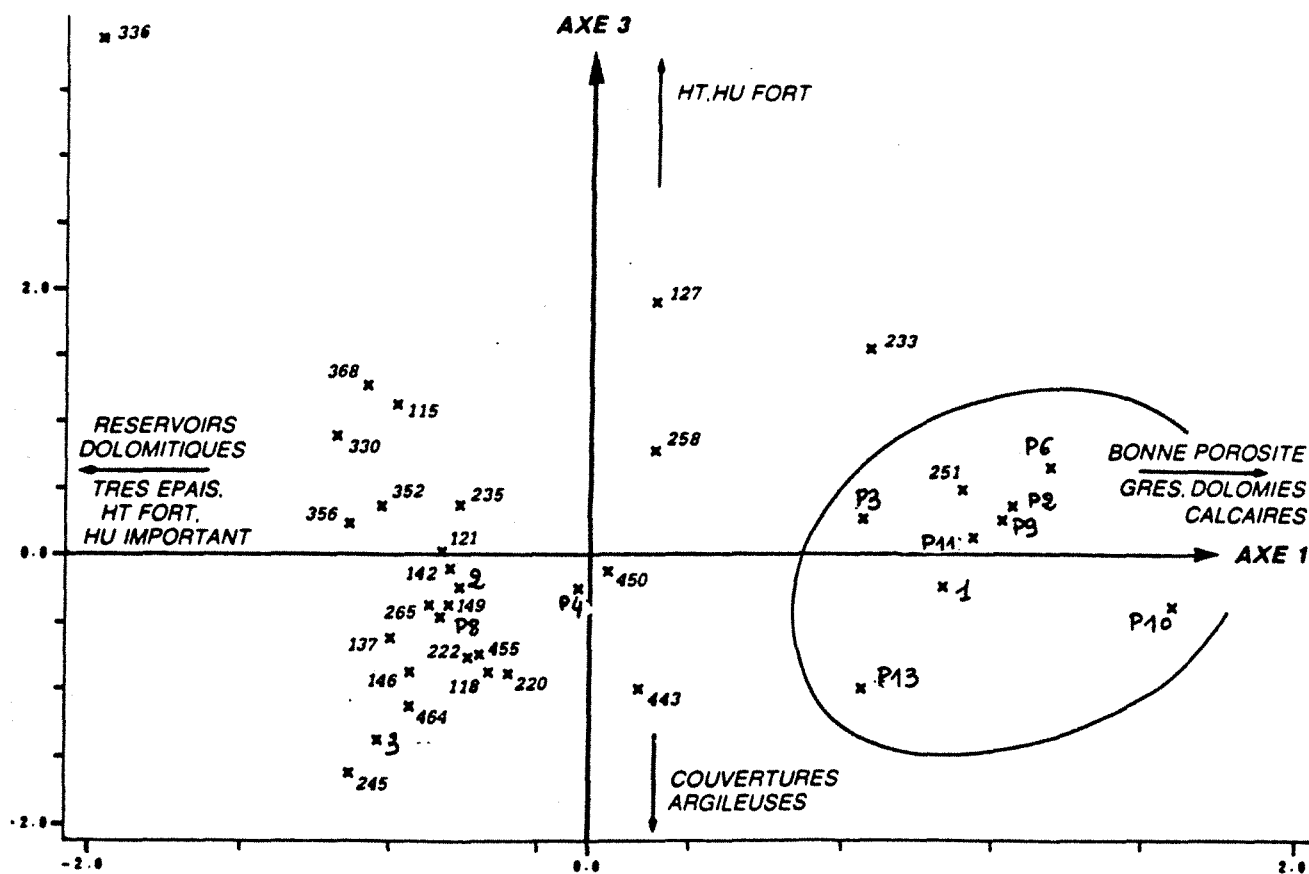
Nom de la variable	Moyenne	Ecart-type	Minimum	Maximum
HT (m)	56,3	12,8	46,0	117,1
HU (m)	30,5	8,5	18,4	60,3
V1	0,307	0,084	0,110	0,440
V2	0,094	0,050	0,000	0,180
V3	0,055	0,102	0,000	0,360
V4	0,272	0,122	0,050	0,530
V5	0,270	0,089	0,120	0,540
PHI1 (%)	21,11	2,36	14,90	26,00
PHI2 (%)	17,67	3,14	12,40	27,60
PHI3 (%)	22,07	3,85	15,70	30,30

Tableau 26- Caractéristiques des variables géologiques

PLAN FACTORIEL 1 - 2



PLAN FACTORIEL 1-3



Les 4 premières composantes principales prennent en compte 80 % de l'inertie totale du nuage de points. Le premier facteur est fortement corrélé positivement avec PHI3 (porosité du A3), V5 (grès) et V3 (dolomies calcaires), négativement avec HT, HU et V4 (réservoirs dolomitiques). Le second facteur oppose la variable V5 (pourcentage en grès) à la variable V1 (pourcentage en dolomicrites). Le troisième facteur est très corrélé positivement avec HT, HU et négativement avec V2 (couvertures argileuses). Le quatrième facteur est corrélé positivement avec PHI1.

La figure 177 présente la projection des puits sur les principaux plans factoriels de l'analyse, les plans 1-2 et 1-3 (i.e. engendrés respectivement par les nouveaux axes 1,2 ou 1,3).

Un premier groupe de puits apparaît, vers la droite de l'axe 1 ; il comprend la plupart des puits du "top" de la structure qui présentent l'association typique : dolomies calcaires, réservoirs gréseux et bonne porosité (cf. signification géologique de l'axe 1). Quelques puits sont projetés très haut sur l'axe 2 et proches du centre sur l'axe 1 : ces puits sont riches en grès (axe 2) mais ne présentent pas l'association précédente. Les puits tirés à gauche sur l'axe 1 (Figure 177) sont presque tous les puits du compartiment Sud-Est affaissé, où les réservoirs dolomitiques sont dominants, les épaisseurs totale et utile fortes (cf. signification de l'axe 1 du côté des corrélations négatives). Des points de ce nuage ont de fortes valeurs négatives sur l'axe 2 (par exemple Y336) : ce sont des puits où les couvertures de type 1 sont abondantes.

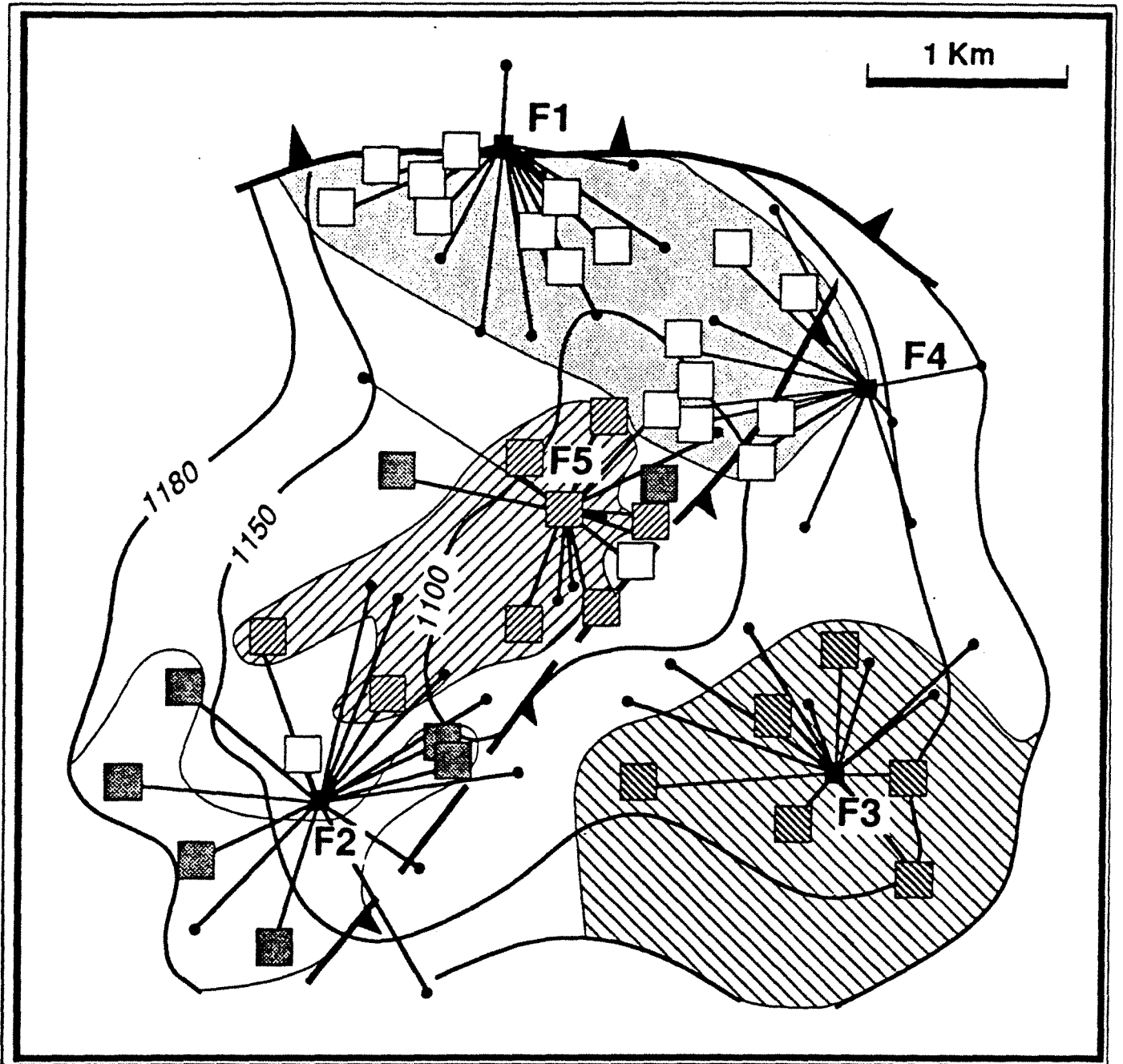
L'interprétation du plan factoriel 1-3 (Figure 177) est un peu plus difficile : l'axe 3 explique presque les mêmes variables que l'axe 2, à part la corrélation négative avec le pourcentage en couvertures argileuses.

Une classification des puits en 4 groupes, à partir des composantes principales extraites des variables géologiques, donne les résultats présentés sur la carte de la figure 178. 3 de ces groupes semblent particulièrement stables : la zone des puits du Nord-Nord-Est, la zone des puits centraux et la zone des puits du compartiment Sud-Est.





6.3.2.2. *Données sismiques*

a - Données sismiques synthétiques

Nous avons essayé de modéliser un "portrait-robot" de ces principales zones en terme d'épaisseurs constitutives du A (épaisseur totale et distribution des épaisseurs), lithologies et vitesses de tranche. Ce portrait-robot est une "moyenne" synthétique des caractéristiques des puits pour chaque groupe trouvé au cours de l'analyse géologique. Les vitesses sont déduites des Sonics, les densités n'ont pas été prises en compte, vue la différence quasi nulle entre les films synthétiques avec densité et sans densité.



4 CLUSTERS OF WELLS

GROUPS				
	1	2	3	4

La concordance entre les traces synthétiques aux puits les plus typiques et les traces modèles (i.e. traces sismiques synthétiques des portraits-robots géologiques) est assez bonne (Figure 179). Ces traces modèles interviendront dans le choix des paramètres de représentation des traces sismiques.

b - Données sismiques réelles

Les données sismiques réelles (au pas de 4 ms) correspondent à une campagne sismique 2D, acquise en 1980. Cette campagne a été entièrement retraitée afin d'améliorer la qualité des données et de permettre une interprétation fine de la zone des réservoirs.

Les points importants du retraitement sont :

- la préservation des amplitudes,
- des analyses de vitesses très serrées (1 par 500 m) sur la structure,
- une mise à phase nulle avec déconvolution stratigraphique ; l'opérateur a été calculé avec les traces impulsionnelles de plusieurs puits sur une fenêtre assez courte centrée sur l'Albien ; on a ainsi pu vérifier sa stationnarité spatiale.

Le retraitement a grandement amélioré la qualité des données. Sur le traitement antérieur (où déjà la mise en couverture multiple était déficiente, un certain nombre de points de tir manquants n'ayant pas été déclarés), aucune étude de sismique stratigraphique n'avait été entreprise, seule l'étude structurale avait été conduite avec des difficultés pour le pointé de l'Albien (plusieurs interprétations possibles). Après retraitement, le calage entre les données synthétiques et la sismique est bien meilleur, le pointé de l'Albien est devenu beaucoup moins ambigu ; le caractère sismique est bien plus marquant.

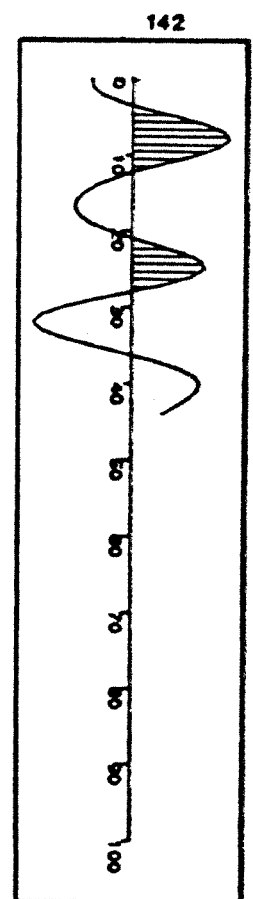
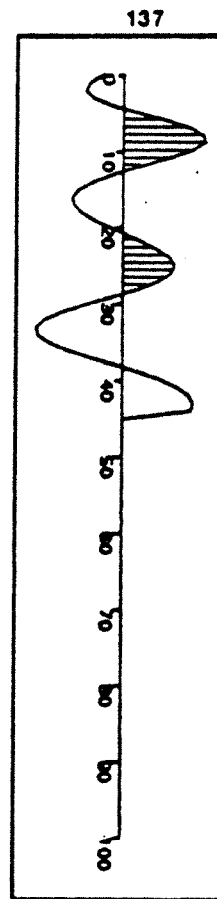
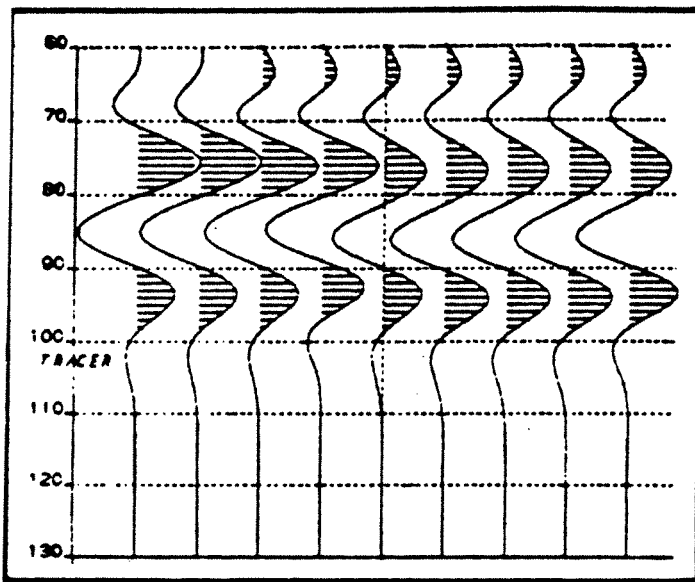
Après retraitement, l'ensemble des sections a été repointé au niveau du toit de l'Albien. Nous avons cherché à pointer le plus précisément possible en utilisant des sections à très grande échelle (40 cm/s) et **toutes les traces synthétiques** disponibles pour le **calage**. La qualité du calage (concordance entre la trace synthétique au puits et la trace sismique réelle adjacente) est indiquée dans le tableau 33 ; ceci servira ultérieurement lors de la classification simultanée des traces réelles et des traces synthétiques (Tableau 33). Dans l'ensemble, les calages sont satisfaisants. Le pointé de l'Albien est ensuite digitalisé très finement et horizontalisé à 1000 ms temps double. La fenêtre d'analyse sera extraite à partir du toit horizontalisé.

6.3.3. Etude des traces sismiques synthétiques

L'analyse des traces sismiques synthétiques permet de choisir les paramètres du traitement ultérieur des traces réelles :

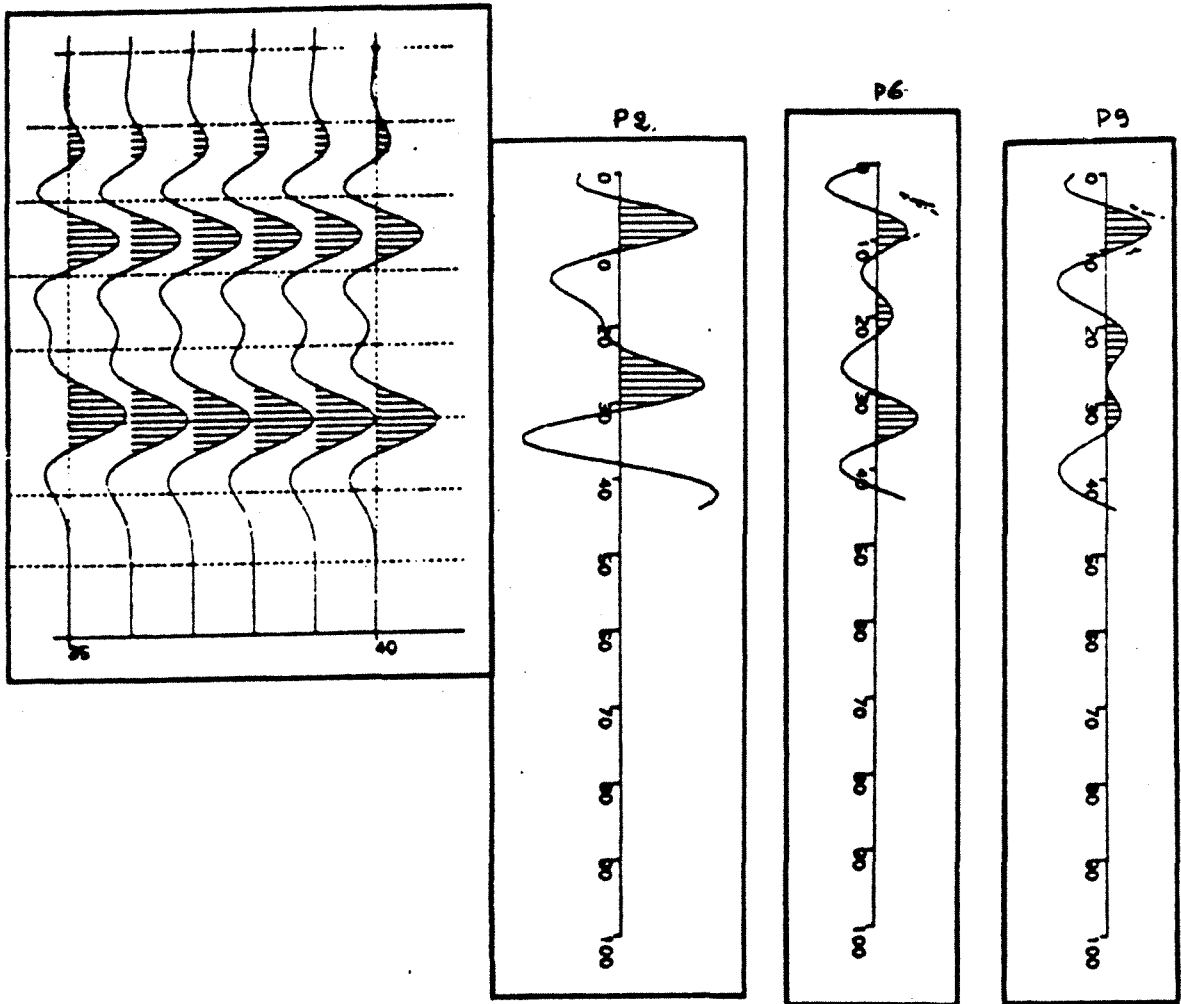
COMPARAISON TRACES SYNTHETIQUES AUX PUIITS ET TRACES MODELES POUR LES 3 FACIES PRINCIPAUX

FACIES NORD



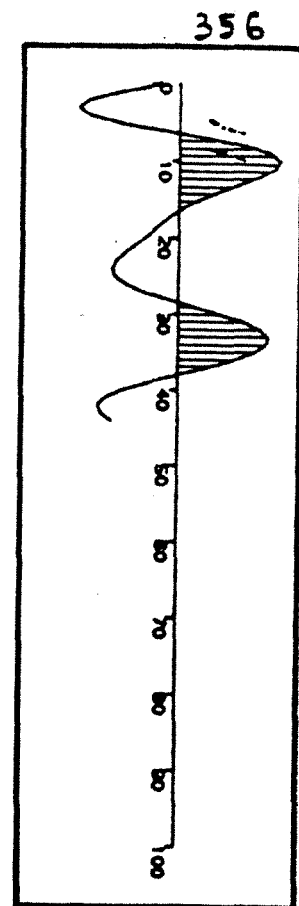
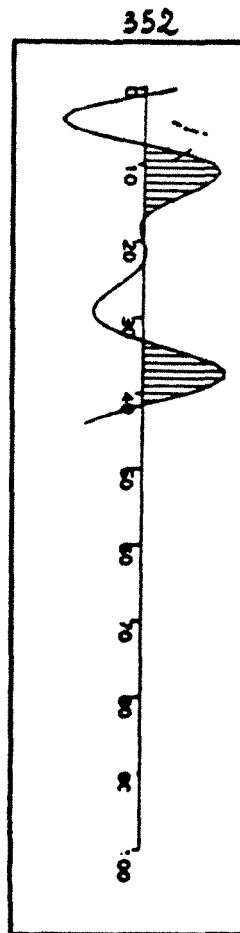
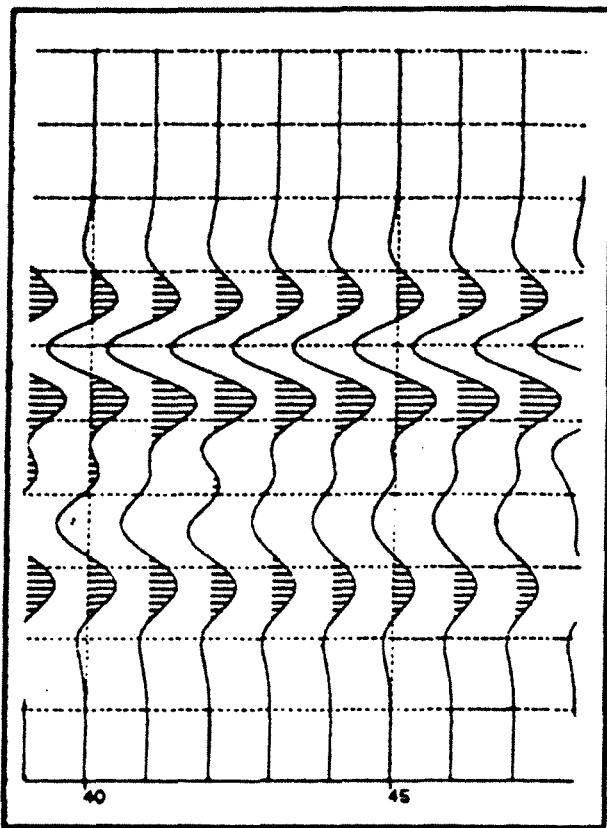
COMPARAISON TRACES SYNTHETIQUES AUX Puits ET TRACES MODELES POUR LES 3 FACIES PRINCIPAUX

FACIES TOP



COMPARAISON TRACES SYNTHETIQUES AUX PUIITS ET TRACES MODELES POUR LES 3 FACIES PRINCIPALES

FACIES SUD-EST



- taille "optimale" de la fenêtre d'analyse,
- nombre de facteurs de l'analyse en composantes principales à prendre en compte.

Cette analyse permet aussi d'étudier les recouvrements éventuels entre la classification des puits (faciès géologiques) et celle des traces adjacentes aux puits.

Les traces sismiques synthétiques étudiées sont (cf. § 6.3.2 - données disponibles) :

- les traces synthétiques des puits calculées à partir des enregistrements Sonic ;
- les traces synthétiques des modèles géologiques des trois principaux faciès (Figure 180); ces traces synthétiques sont la réponse sismique d'un modèle géologique présentant une évolution latérale entre les trois faciès (Figure 180).

Dans tous ces calculs de traces synthétiques, le signal de convolution utilisé est un Ricker d'ordre 2, de pseudo-période 28 ms, à phase nulle. La bande passante est en accord avec celle de la sismique réelle après retraitement.

6.3.3.1. Analyse des traces sismiques synthétiques des puits

a - Etude préliminaire

Nous avons repéré précisément sur tous les films synthétiques les échantillons sismiques correspondant au toit et à la base du A. Un calage précis est d'autant plus important que la fenêtre d'analyse est très fine (de l'ordre de 30 ms, temps double) et que l'A.C.P. porte directement sur la suite des amplitudes (il est nécessaire que le premier échantillon de la trace i ait la même signification que le premier échantillon de la trace $(i+j)$ car leurs amplitudes seront prises en compte dans la même variable $V1$, cf. Figure 162).

L'intervalle sismique ΔT correspondant au A varie entre 25 et 45 ms temps double ; on peut distinguer trois familles de traces synthétiques dont la distribution géographique est reportée sur la carte de la figure 181 :

$$\Delta T \approx 30 \text{ ms,}$$

$$\Delta T \approx 35 \text{ ms,}$$

$$\Delta T \geq 40 \text{ ms.}$$

Les intervalles les plus longs sont localisés dans le compartiment affaissé ce qui s'explique par les fortes épaisseurs du A et les vitesses plus lentes (sédimentation plus argileuse).

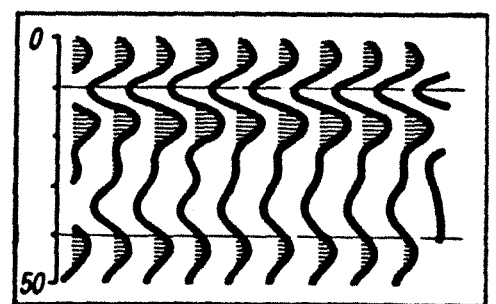
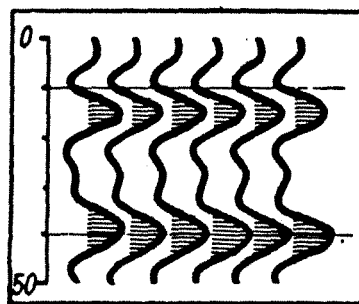
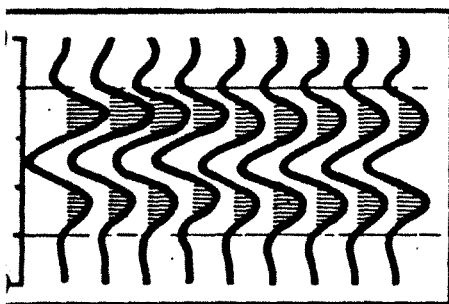
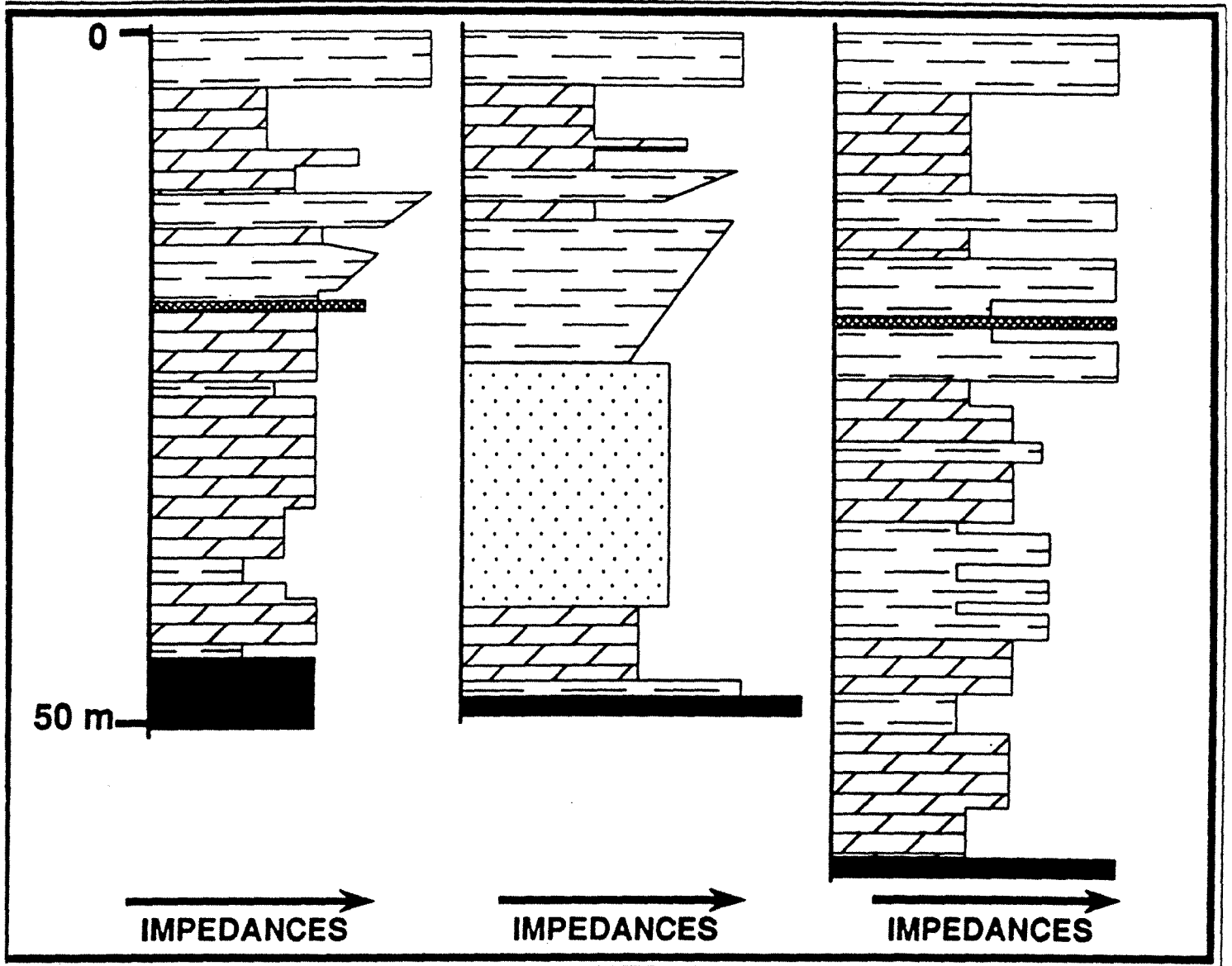
Les intervalles moyens ($\Delta T \approx 35$ ms) sont plutôt dans le coin Sud-Ouest du champ. Cependant, la plupart des puits correspondent à des intervalles voisins de 30 ms (Figure 181).

Fig.180

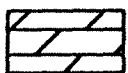
**NORTH
GROUP**

**TOP
GROUP**

**SOUTH - EAST
GROUP**



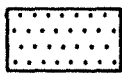
SEALS



RESERVOIRS

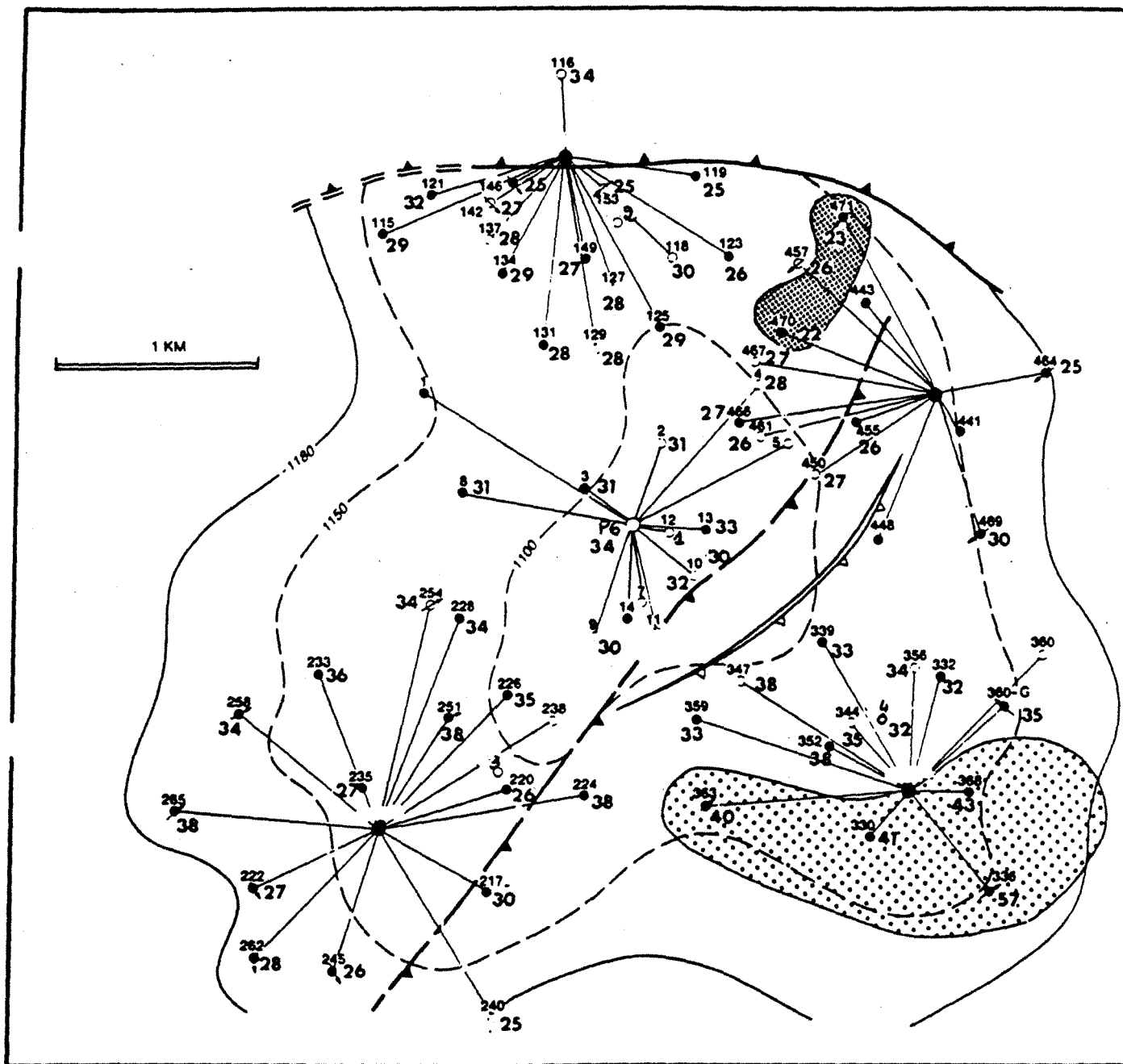


ANHYDRITE



**MIXED CALCAREOUS
DOLOMITES AND SANDSTONES**

INTERVALLES TEMPS DOUBLES CORRESPONDANT AU A1



ΔT IMPORTANT



ΔT FAIBLE

Cette analyse pose le problème du choix de la taille de la fenêtre : doit-on analyser une fenêtre plus large pour être certain d'englober tout le A ou prendre la taille la plus représentée, et quelle est l'influence de ce choix ?

Deux essais ont été réalisés : 28 ms et 35 ms.

Dans les deux cas, nous avons supprimé de l'analyse les traces synthétiques des puits 330, 336, 363, 368 (ΔT trop long par rapport aux deux fenêtres testées), celles des puits 470 et 471 (ΔT trop court). Les analyses ont donc porté sur les 58 puits restants.

b - Analyse des traces synthétiques sur une fenêtre de 28 ms

- Analyse en composantes principales

La distribution de l'inertie sur les axes de l'A.C.P. ainsi que les corrélations entre les variables initiales et les facteurs sont synthétisées dans le tableau 27 (a et b). Comme les 3 premiers axes suffisent pour prendre en compte 85,6 % de l'inertie totale du nuage de traces, les essais de classification ont été effectués dans l'espace factoriel dérivé, de dimension 3.

- Essai de classification

Un seul essai a été fait avec 5 groupes, obtenus en 2 itérations et donnant une valeur R^2 de 0,703. Les groupes trouvés sont reportés sur la carte du champ (Figure 182) et les figures 183a et 183b illustrent la morphologie des traces regroupées entre elles.

c - Analyse des traces synthétiques sur une fenêtre de 35 ms

- Analyse en composantes principales

La distribution de l'inertie sur les axes (Tableau 28 a) ainsi que les corrélations facteurs/variables initiales (Tableau 28 b) sont très comparables au cas de la fenêtre de 28 ms (Tableau 27). Les 3 premiers axes (80 % de l'inertie) sont utilisés dans la classification.

- Classification à 5 groupes

Les performances de l'algorithme sont aussi bonnes que sur la fenêtre de 28 ms, les groupes sont un peu mieux équilibrés.

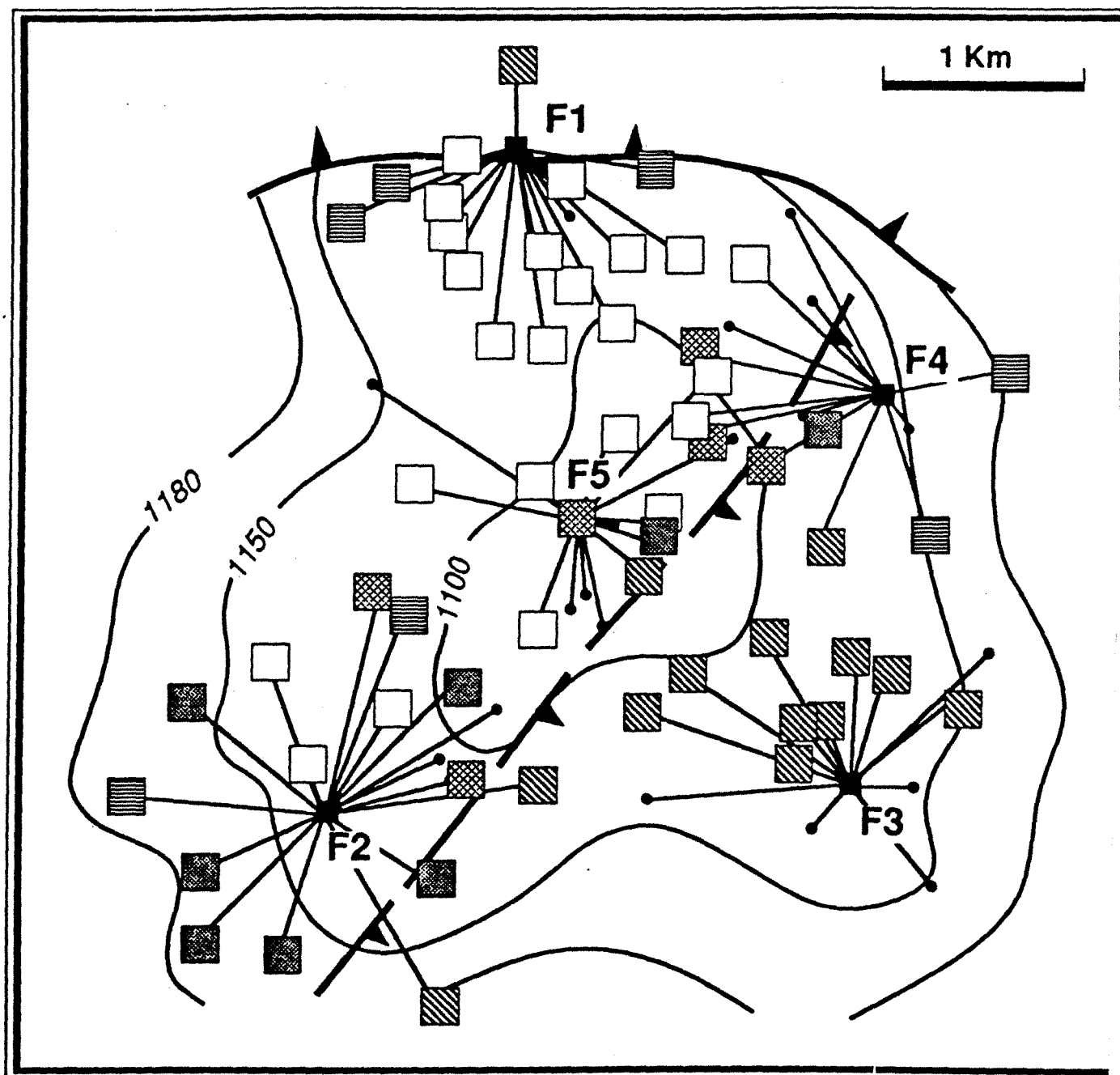
N° axe	% d'inertie	% d'inertie cumulée
1	0,4710	0,4710
2	0,2372	0,7082
3	0,1482	0,8564
4	0,0671	0,9235
5	0,0464	0,9699
6	0,0198	0,9897
7	0,0083	0,9980
8	0,0017	0,9997
9	0,0003	1
10	0,0000	1

Tableau 27a - A.C.P. sur les amplitudes de la trace-fenêtre de 28 ms.
Distribution de l'inertie sur les facteurs

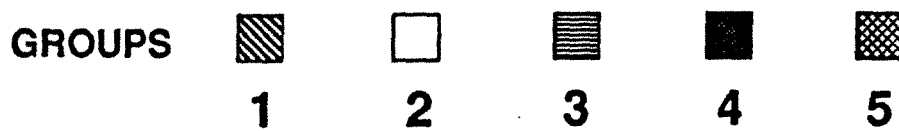
	Facteur 1	Facteur 2	Facteur 3
A1	- 0.13769	- 0.33270	0.79669
A2	0.20995	- 0.28447	0.86537
A3	0.57780	- 0.20416	0.71880
A4	0.81369	- 0.10493	0.42651
A5	0.90586	- 0.00422	0.12872
A6	<u>0.90798</u>	0.10504	- 0.12437
A7	<u>0.83978</u>	0.24635	- 0.34196
A8	0.66089	0.44761	- 0.52880
A9	0.25760	0.68754	- 0.61795
A10	- 0.30506	<u>0.76180</u>	- 0.46404
A11	- 0.66647	<u>0.63117</u>	- 0.20807
A12	- 0.82631	0.46981	- 0.02537
A13	- 0.90176	0.31982	0.08090
A14	- 0.94296	0.16346	0.13143
A15	<u>- 0.95670</u>	- 0.02619	0.12901
A16	<u>- 0.91488</u>	- 0.27478	0.06370
A17	- 0.74847	- 0.57357	- 0.07705
A18	- 0.41548	- 0.81916	- 0.25392
A19	- 0.01584	<u>- 0.90062</u>	- 0.38086
A20	0.32905	<u>- 0.83090</u>	- 0.41675
A21	0.58080	- 0.66447	- 0.37202
A22	0.73999	- 0.44396	- 0.27049
A23	0.82074	- 0.20519	- 0.14146
A24	<u>0.84360</u>	0.03002	- 0.00619
A25	<u>0.81541</u>	0.25773	0.12682
A26	0.71602	0.46828	0.25121
A27	0.51222	0.61431	0.34335
A28	0.22116	0.63371	0.36858

Tableau 27b - A.C.P. sur les amplitudes de la trace-fenêtre de 28 ms.

Corrélations entre les facteurs et les variables initiales



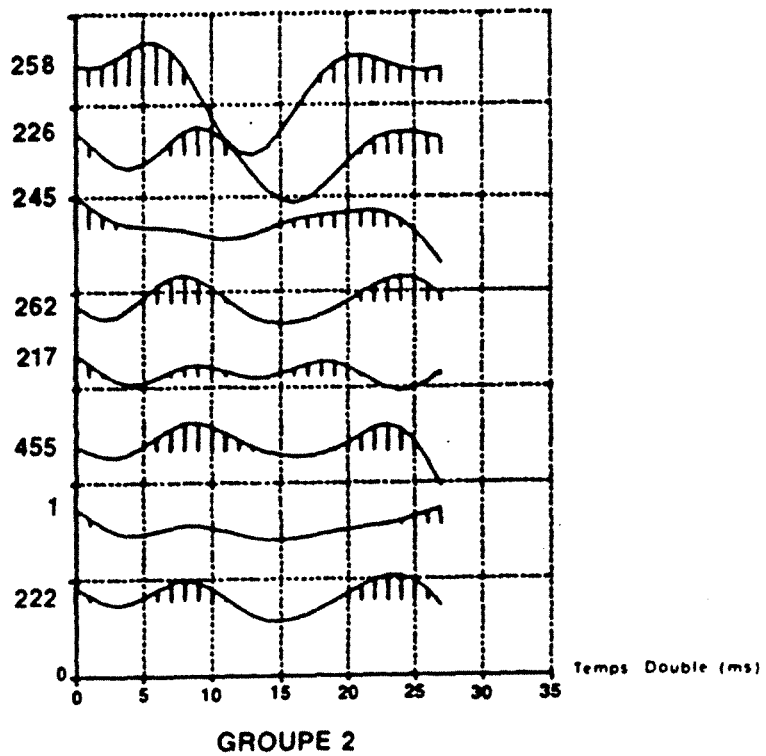
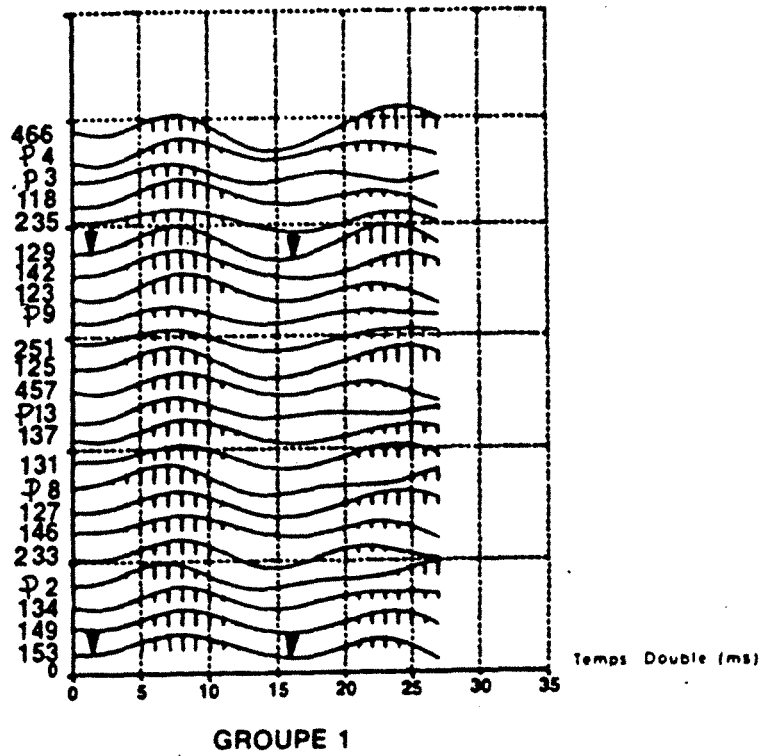
5 CLUSTERS OF SYNTHETIC TRACES



CLASSIFICATION DES TRACES SYNTHETIQUES

5 GROUPES

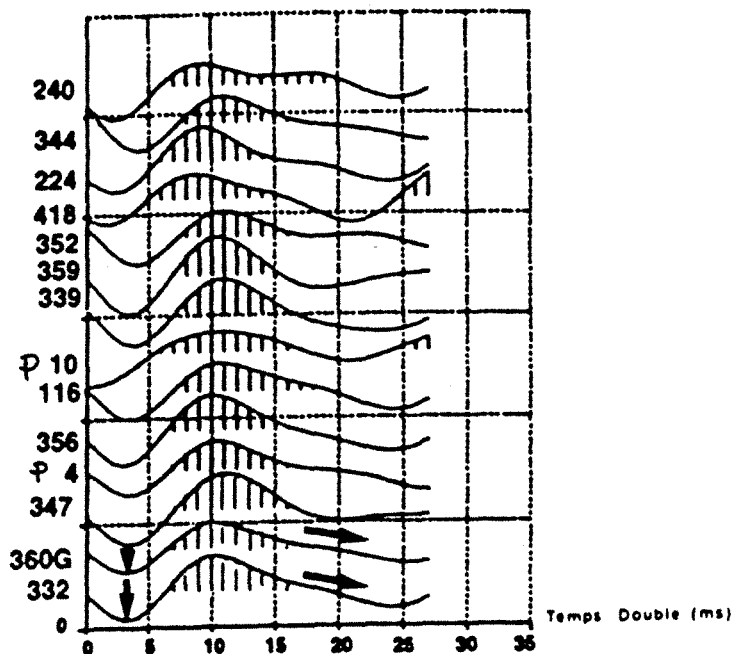
FENETRE DE 28 ms



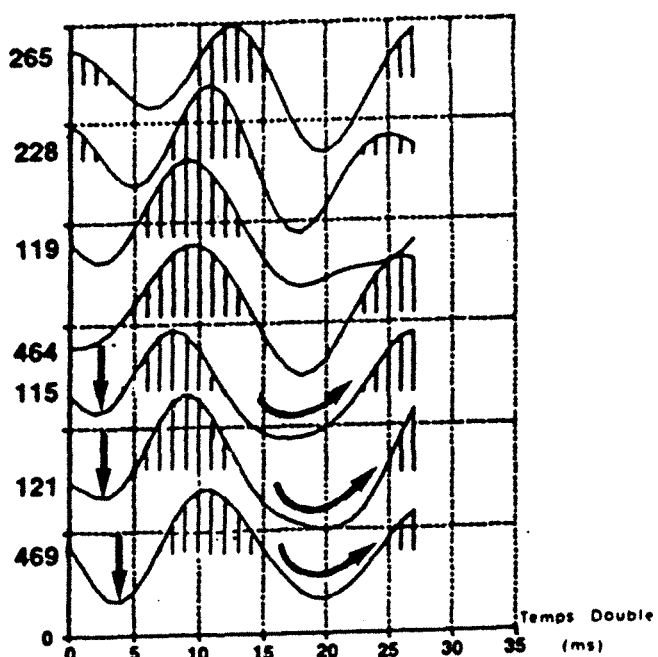
CLASSIFICATION DES TRACES SYNTHETIQUES

5 GROUPES

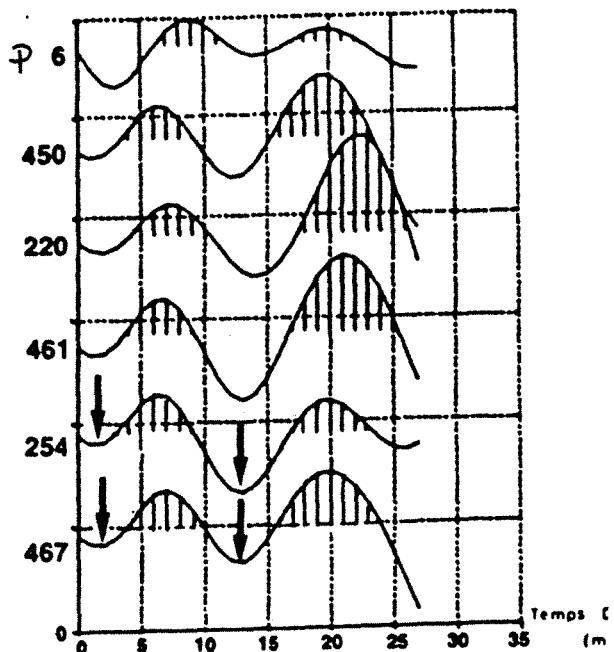
FENETRE DE 28 ms



GROUPE 3



GROUPE 4



GROUPE 5

N° axe	% d'inertie	% d'inertie cumulée
1	0,4417	0,4417
2	0,2200	0,6617
3	0,1356	0,7973
4	0,1009	0,8982
5	0,0606	0,9588
6	0,0253	0,9841
7	0,0100	0,9941
8	0,0032	0,9974
9	0,0020	0,9993
10	0,0005	0,9999

Tableau 23a - A.C.P. sur les amplitudes de la trace-fenêtre de 35 ms.
Distribution de l'inertie sur les facteurs

	Facteur 1	Facteur 2	Facteur 3
A1	- 0.15233	- 0.25981	0.59291
A2	0.18120	- 0.18323	0.70671
A3	0.53637	- 0.09371	0.67295
A4	0.76690	- 0.01119	0.50214
A5	0.86069	0.05778	0.28675
A6	0.86866	0.12821	0.07127
A7	0.81073	0.22141	- 0.15489
A8	0.64943	0.35933	- 0.41395
A9	0.27506	0.52591	- 0.67273
A10	- 0.25899	0.57110	- 0.72265
A11	- 0.60987	0.46981	- 0.56245
A12	- 0.76947	0.34810	- 0.39258
A13	- 0.84774	0.23334	- 0.26069
A14	- 0.89324	0.10991	- 0.15564
A15	- 0.91320	- 0.04672	- 0.06784
A16	- 0.88093	- 0.26091	0.00704
A17	- 0.72782	- 0.53119	0.05499
A18	- 0.40777	- 0.77056	0.05170
A19	- 0.01212	- 0.87671	- 0.00671
A20	0.34047	- 0.84962	- 0.08842
A21	0.60810	- 0.72719	- 0.16367
A22	0.78531	- 0.53763	- 0.20974
A23	0.88069	- 0.30830	- 0.21623
A24	0.91084	- 0.05759	- 0.17977
A25	0.87966	0.21232	- 0.09776
A26	0.76242	0.49393	0.03636
A27	0.52242	0.73148	0.21253
A28	0.18452	0.83079	0.37482
A29	- 0.14412	0.77265	0.47088
A30	- 0.40115	0.62407	0.50145
A31	- 0.58828	0.43667	0.48348
A32	- 0.71953	0.22314	0.42218
A33	- 0.78831	- 0.01424	0.31019
A34	- 0.76651	- 0.25928	0.14290
A35	- 0.63128	- 0.46924	- 0.05991

Tableau 28b - A.C.P. sur les amplitudes de la trace-fenêtre de 35 ms.
Corrélations entre les facteurs et les variables initiales

d - Comparaison des classifications

Trois domaines apparaissent aussi bien pour la classification sur 28 ms (Figure 182) que pour celle sur 35 ms : les puits du Sud-Est, les puits du centre de la structure (autour de la plateforme YF ainsi que 251, 235 et 233), les puits du Nord (autour du puits 2).

Seulement 25 % des traces sont affectées différemment dans les deux classifications : dans la classification des traces après analyse sur 28 ms, beaucoup plus sont rattachées au faciès du Nord, par rapport à l'analyse sur 35 ms. Si la zone Sud-Est reste constante dans les deux classifications, les différences affectent plutôt la zone haute. Nous avons construit l'intersection entre les deux classifications ("formes fortes"), technique usuelle en classification automatique pour déterminer des noyaux ou pôles vraiment stables dans une population (les individus toujours regroupés quelles que soient les conditions d'initialisation de l'algorithme). Ici, la signification est un peu différente puisque ce sont les traces synthétiques regroupées ensemble, quelles que soient les conditions d'analyse (taille de la fenêtre).

Les formes fortes des deux classifications montrent :

- un premier groupe de traces synthétiques correspondant aux traces des puits 265, 228, 469, 464, 119, 121 et 115 ;
- un second groupe comprenant les puits du Nord ;
- un troisième groupe constitué des traces synthétiques des puits du Sud-Est ;
- le quatrième groupe est un peu accessoire et comprend des puits très dispersés les uns des autres, le long de l'accident synsédimentaire.

La stabilité des "formes fortes" a été confirmée par une analyse factorielle discriminante sur les trois premiers groupes trouvés : les axes discriminants les individualisent parfaitement (100 % de bien classés sur l'échantillon de base).

Le choix des paramètres de traitement des traces – taille de la fenêtre et nombre de facteurs à prendre en compte – résulte de cette analyse.

La taille de la fenêtre (entre 28 et 35 ms) a beaucoup moins d'influence que ce que l'on aurait pu imaginer initialement. La distribution de l'inertie sur les axes principaux ainsi que les corrélations entre les variables initiales et les facteurs sont pratiquement analogues. Les résultats des classifications à 5 groupes ne diffèrent que sur 25 % des traces. On peut donc retenir une taille de fenêtre quelconque **entre 28 et 35 ms**.

Les trois premiers axes expliquent environ 80 % de l'information initiale totale et permettent dans les classifications de trouver des groupes stables et cohérents. L'espace sismique d'analyse sera donc l'espace IR^3 engendré par les **trois premières composantes principales**.

6.3.3.2. Analyse des traces sismiques modèles

Nous cherchons à valider la signification géologique des groupes de traces sismiques précédemment trouvés. Pour cela, nous classons simultanément les traces synthétiques des puits avec les réponses sismiques des modèles géologiques typiques des trois principales zones.

Une analyse en composantes principales est alors effectuée sur l'ensemble des traces modèles (60) et des traces synthétiques aux puits, sur une fenêtre de 30 ms. La distribution de l'inertie et les corrélations variables initiales/nouvelles variables sont peu affectées par l'introduction des 60 traces modèles supplémentaires (Tableau 29).

Les trois premiers axes (84,1 % de l'inertie totale) servent dans des tests de classifications à 3, 4 ou 5 groupes. Les résultats de la classification à 3 groupes sont excellents. Dans la classification des traces synthétiques des puits, on reconnaît les groupes des formes fortes, qui avaient une signification géologique, mais leur cohérence est grandement améliorée : puits au Sud-Est affectés dans le même groupe que leur portrait-robot, de même pour les puits du "top" de la structure et ceux de la zone Nord. Les puits autour de la plate-forme YF sont affectés en faciès Nord. La dispersion des groupes est la suivante :

- groupe 1	:	0,782,
- groupe 2	:	0,696,
- groupe 3	:	0,529.

Le groupe 3 (puits Sud-Est) est très compact, caractéristique déjà observée sur les traces synthétiques seules.

6.3.3.3. Conclusion à l'analyse des traces synthétiques

L'analyse des traces synthétiques permet, d'une part, de choisir les paramètres du traitement ultérieur des traces réelles, d'autre part, de voir s'il existe des recoupements entre la classification des puits et celle des traces adjacentes aux puits. Dans ce cas de figure, une interprétation géologique des faciès sismiques est possible.

Nous avons retenu une taille de fenêtre d'analyse d'environ 30 ms (influence faible entre 28 et 35 ms) et jugé que les trois premiers facteurs de l'A.C.P. sur les amplitudes de la trace étaient suffisants pour rendre compte de l'information contenue dans les traces. L'espace sismique d'étude sera donc l'espace IR^3 engendré par ces trois axes.

N° axe	% d'inertie	% d'inertie cumulée
1	0,422	0,422
2	0,236	0,658
3	0,183	0,841
4	0,111	0,952

Tableau 29a - Distribution de l'inertie sur les axes

	Corrélations > 0	Corrélations < 0
1er axe	A27 (0,84) A7 (0,83)	A16 (- 0,96)
2ème axe	A11 (0,92)	A21 (- 0,69) A4 (- 0,59)
3ème axe	A23 (0,88)	A31 (- 0,93)

Tableau 29b - Principales corrélations entre les facteurs et les variables initiales

Tableau 29 - A.C.P. sur les traces synthétiques et sur les traces modèles (signatures sismiques des modèles géologiques)

6.3.4. Etude des traces réelles

6.3.4.1. Représentation des traces dans l'espace sismique

Après moyenne mobile des traces par fenêtre glissante de 3 traces avec un recouvrement de 2, nous disposons de 13.432 traces sismiques couvrant l'ensemble du champ. Ces traces sont représentées dans l'espace des 3 premiers facteurs de l'analyse en composantes principales sur les 9 amplitudes des échantillons de la fenêtre 1.000 - 1.032 ms.

L'espace sismique prend en compte 83 % de l'information initiale totale (Tableau 30).

6.3.4.2. Classifications automatiques

Deux essais à 5 et 6 groupes ont été réalisés dont les caractéristiques sont :

- 5 groupes : $R^2 = 0,658$, 5 itérations ;
- 6 groupes : $R^2 = 0,697$, 11 itérations.

Les affectations des traces sismiques dans les 5 groupes sont reportées sur la carte de la figure 85. Nous avons ensuite interprété les cartes de faciès sismiques ainsi obtenues.

L'interprétation de la carte à 5 groupes (Figure 85) est la plus intéressante. On distingue les trois zones principales de faciès géologiques : le "top" de la structure, la zone Nord et la zone Sud-Est. Une quatrième zone apparaît dans la partie Sud-Ouest. Le cinquième groupe se retrouve assez dispersé : dans la partie Sud-Est, le long de l'accident synsédimentaire et aussi mélangé au quatrième groupe. La carte à 5 groupes en classification automatique est en accord avec les classifications géologique et sismique des traces synthétiques.

6.3.4.3. Analyse factorielle discriminante

Des groupes d'apprentissage ayant été définis au préalable, l'analyse factorielle discriminante permet de trouver les axes discriminant au mieux les traces d'apprentissage (étape descriptive) puis d'affecter un individu anonyme dans l'un quelconque des groupes a priori (étape prédictive).

N° axe	% d'inertie	% d'inertie cumulée
1	0,3330	0,3330
2	0,3068	0,6397
3	0,1947	0,8344
4	0,0974	0,9319
5	0,0515	0,9833
6	0,0147	0,9981
7	0,0017	0,9998
8	0,0002	1,000
9	0,0000	1,000

Tableau 30a - Distribution de l'inertie sur les axes

	Corrélations > 0	Corrélations < 0
1er facteur	AMP3 (0,86)	AMP7 (- 0,77)
2ème facteur	AMP5 (0,89)	AMP9 (- 0,74)
3ème facteur	AMP1 (0,82)	

Tableau 30b - Corrélations entre les principaux facteurs et les variables initiales

Tableau 30 - A.C.P. sur les 13 432 traces sismiques réelles après moyenne mobile

La classification à 5 groupes (Figure 85) détermine des zones de faciès sismique homogène cohérentes spatialement mais aussi en accord avec tous les résultats antérieurs de l'étude. A partir de cette carte, nous avons défini 5 groupes de traces sismiques d'apprentissage (Figure 84), qui correspondent aux traces classées le plus près du centre de chaque groupe (noyaux des groupes). Le nombre total de traces d'apprentissage est de 1.246 (224 pour le groupe 1, 243 pour le groupe 2, 354 pour le groupe 3, 223 pour le groupe 4 et 202 pour le groupe 5).

Les résultats de la discrimination sont excellents, de 99 à 100 % de bien classées sur les traces d'apprentissage (Tableau 31). Les groupes d'apprentissage sont réellement individualisés dans l'espace sismique puisque les axes trouvés sont très discriminants. Le groupe 3 (faciès "top structure") est toutefois un peu moins net (seulement 94 % de traces bien classées).

Les 12.186 traces qui ne sont pas intervenues dans les groupes d'apprentissage sont ensuite classées en tant que "traces anonymes" par rapport à ces groupes : chaque trace est projetée dans l'espace défini par les axes discriminants et affectée au groupe d'apprentissage dont elle est le plus proche. Nous proposons une interprétation des faciès sismiques en analyse discriminante en nous appuyant sur la probabilité de bonne classification dans les zones où la cohérence spatiale est faible. Cette carte (Figure 84) est à comparer avec celle obtenue en classification automatique (Figure 85).

Globalement, la distribution des faciès sismiques est très proche, la différence la plus grande étant la diminution de l'extension de la zone du faciès 3 ("top" de la structure) en analyse discriminante. Les affectations en discrimination sont beaucoup plus contraintes qu'en classification puisque une étape d'apprentissage a eu lieu. La carte en discrimination serait donc plus fiable qu'en classification, à condition d'avoir défini des groupes d'apprentissage cohérents. Les faciès du Sud-Est (groupe 1) et du couloir Nord, Nord-Est (groupe 2) sont très stables entre la classification et la discrimination. Le groupe 4 évolue (en négatif par rapport au groupe 3) : sa signification géologique n'est pas encore déterminée (dans l'étude des données géologiques, aucun regroupement de puits n'était apparu clairement dans cette zone). Il en est de même pour le groupe 5.

6.3.4.4. Affectation des traces synthétiques aux groupes de traces réelles d'apprentissage

Il est intéressant de savoir comment les traces synthétiques se classent dans les groupes de traces réelles afin de corroborer leur signification géologique en trouvant des regroupements analogues entre l'espace géologique, l'espace sismique synthétique et l'espace sismique réel.

Il est nécessaire auparavant d'homogénéiser traces synthétiques et traces réelles.

Groupe	Nombre d'individus	% de bien classés
1	224	99
2	243	100
3	354	94
4	223	99
5	202	100

Tableau 31 - Analyse factorielle discriminante sur cinq groupes de traces sismiques réelles

Nombre de traces analysées	Moyenne M	Ecart type	$\frac{\sigma}{M}$	Rapport $\frac{M \text{ traces réelles}}{M \text{ traces synthétiques}}$
402 (1)	39 504	7 583	0,192	28,207 (1)/(2)
38 (2)	1 400,5	-	-	29,153
12 (3)	1 355,0	-	-	(1)/(3)
21 (4)	1 466,7			27 (1)/(4)

- (1) Traces réelles
- (2) Toutes traces synthétiques aux puits
- (3) Traces synthétiques aux puits de la zone des traces réelles
- (4) En éliminant les puits où le Cénomanién peut être affecté par l'effet de bord de la fenêtre

Tableau 32 - Comparaison des amplitudes du Cénomanién entre traces réelles et traces synthétiques

Homogénéisation des données

Le contenu fréquentiel des données synthétiques est conforme à celui des données réelles (signal utilisé déduit des données réelles) ; dans les deux cas, la phase du signal est nulle. L'homogénéisation à faire porte sur les amplitudes. Pour trouver le rapport d'amplitude existant entre données réelles et données synthétiques, nous étudions l'amplitude d'un marqueur sismique important, sur les traces synthétiques et sur les traces réelles : ce marqueur ne doit pas présenter d'évolution latérale du caractère sismique sur une zone suffisamment grande pour ne pas perturber la comparaison statistique entre amplitudes réelles et amplitudes synthétiques.

Le marqueur retenu est le toit du Cénomaniens analysé sur les profils 87 et 118 ainsi que sur les 38 traces synthétiques pour lesquelles ce marqueur n'est pas perturbé par les effets de bord de fenêtre de convolution. Le tableau 32 permet de comparer les amplitudes et leur dispersion pour les données réelles et synthétiques.

L'écart-type des amplitudes réelles du Cénomaniens est assez faible ; le rapport amplitudes réelles/amplitudes synthétiques varie peu (entre 27 et 29) quelles que soient les traces synthétiques analysées. Nous avons retenu la valeur de 28 pour homogénéiser les données synthétiques avec les données réelles.

Affectation des traces synthétiques

Les axes discriminants sont les mêmes que ceux utilisés pour la carte de faciès en discrimination à 5 groupes. Les classifications des traces synthétiques sont reportées dans le tableau 33. Les puits marqués d'un astérisque sont ceux pour lesquels le calage entre la trace synthétique et les traces réelles est mauvais.

Malgré la difficulté d'homogénéiser des données réelles à des données synthétiques, on retrouve correctement affectées un certain nombre de traces synthétiques, en particulier dans les groupes de traces réelles correspondant aux trois faciès géologiques principaux. Dans le groupe 1 (faciès Sud-Est), sont affectées les traces synthétiques des puits 4, 330, 332, 336, 352, 356, 359, 363 et 368. Dans l'ensemble, on retrouve le groupe Sud-Est formé au cours de la classification des traces synthétiques.

Dans le groupe 2 (faciès Nord et Nord-Est), sont affectées, en particulier, les traces synthétiques des puits 118, 123, 137, 142, 146, 149, 455, 457, 464 et 470. Ceci est cohérent avec toutes les classifications précédentes.

Dans le groupe 3 (faciès central), on retrouve, parmi d'autres, les traces des puits P3, P9, P10 et P13, les plus typiques géologiquement du faciès "top".

Nom du puits	Groupe d'affectation	Nom du puits	Groupe d'affectation
Y. 1	5	Y. 265	4
Y. 04	1	Y. 330	1
Y. 115*	4	Y. 332*	1
Y. 116	1	Y. 336	1
Y. 118	2	Y. 339*	3
Y. 119	4	Y. 344	4
Y. 121*	4	Y. 347	3
Y. 123	2	Y. 352	1
Y. 125*	1	Y. 356	1
Y. 127	3	Y. 359	1
Y. 129	5	Y. 360*G	3
Y. 131	3	Y. 363	1
Y. 134	5	Y. 368	1
Y. 137	2	Y. 448	1
Y. 142	2	Y. 450	1
Y. 146	2	Y. 455*	2
Y. 149*	2	Y. 457	2
Y. 153	2	Y. 461	5
Y. 217	5	Y. 464	2
Y. 220	1	Y. 466	5
Y. 222*	2	Y. 467	1
Y. 224*	1	Y. 469	4
Y. 226	4	Y. 470	2
Y. 228*	4	Y. 471	4
Y. 233	4	Y. P2	4
Y. 235	4	Y. P3	3
Y. 240	3	Y. P4	1
Y. 245	2	Y. P8	4
Y. 251	5	Y. P9	3
Y. 254	5	Y. P10	3
Y. 258*	5	Y. P13	3
Y. 262	5		

* mauvais calage sismique/trace synthétique

Tableau 33 - Affectation des traces synthétiques en discrimination à cinq groupes

La validation des groupes 4 et 5 par l'affectation des traces synthétiques est plus délicate, ces groupes ne correspondant nettement ni à des groupes géologiques de puits, ni à des classes de traces synthétiques précédemment déterminées.

6.3.5. Synthèse de l'analyse du domaine sismique

L'espace sismique analysé est engendré par les 3 premiers facteurs d'une analyse en composantes principales sur les amplitudes des traces réelles, d'une fenêtre de 9 échantillons au pas de 4 ms. Un filtrage latéral par moyenne mobile est réalisé au préalable. Ces paramètres ont été choisis au cours de tests effectués sur traces synthétiques et sur deux profils se croisant (traces réelles des sections 118 et 87). Les classifications automatiques des seules traces synthétiques permettent de trouver des regroupements en accord avec la classification des puits dans l'espace géologique. La signification géologique de ces groupes est confirmée par l'étude de modèles sismiques des principaux faciès géologiques.

L'analyse des traces réelles s'est faite par classification automatique et analyse factorielle discriminante. Les cartes de faciès obtenues avec ces deux outils sont très semblables. Cinq groupes de traces réelles apparaissent nettement. Les trois premiers correspondent aux groupes déjà repérés sur les traces sismiques synthétiques et interprétés géologiquement. Ceci est par ailleurs confirmé par l'affectation de ces traces synthétiques dans les groupes de traces réelles (après leur homogénéisation). Ces 3 groupes correspondent aux faciès suivants :

- Sud-Est, zone de l'épaississement synsédimentaire,
- Nord et Nord-Est,
- "top", faciès réservoirs gréseux associés à des dolomies calcaires.

Les deux autres groupes de traces réelles ne correspondent pas de façon évidente à des groupes de traces synthétiques ou de puits.

Il existe, dans l'espace géologique, un groupe de puits qui recoupe seulement partiellement le groupe de traces sismiques réelles n°4 ; ces puits sont P8, 3, 222, 245 et 265. Par contre, on ne dispose pas de trace synthétique pour le puits 3. D'autre part, le groupe de traces synthétiques de l'espace sismique comprend des puits pour lesquels l'analyse géologique n'était pas disponible : 119, 226, 228, 469, 471. Parmi les traces synthétiques regroupées dont les puits ont été étudiés géologiquement, il ne reste que 223, 235, 265 et P8. Sur ces quatre puits, seuls P8 et 265 appartiennent aussi au groupe géologique (cf. début de ce paragraphe). L'interprétation du groupe 4 est donc délicate suite au manque d'information dans le domaine géologique.

Pour la plupart des traces synthétiques affectées au groupe 5, on ne dispose pas d'information géologique aux puits (seuls deux puits seraient disponibles). Cette interprétation serait très

précaire, c'est pourquoi nous avons préféré déclarer le groupe 5 non interprétable géologiquement.

7. CONCLUSION

L'analyse automatique du faciès sismique sur les courtes fenêtres temporelles est d'autant plus importante qu'actuellement les applications "gisement" se développent.

Le problème posé et les méthodes essayées (représentation conjointe temps-fréquence, analyse spectrale par la méthode de l'entropie maximale) nous ont conduits finalement à revoir notre conception de la caractérisation du faciès. Plutôt que d'estimer des fonctions dérivées de la représentation temporelle de la trace et de calculer sur ces estimations des paramètres caractérisant la trace, nous étudions **directement les données initiales** : la succession des valeurs d'amplitude de la trace sismique. Cette approche évite toutes les estimations intermédiaires, susceptibles d'amener du bruit dans la caractérisation de la trace, et ne prend en compte que les seules données disponibles initialement. C'est **directement un traitement multivariable** (analyse en composantes principales) de ces données qui permet de caractériser leur **morphologie** (i.e. le faciès sismique). Les composantes principales expriment des "formes de base" des traces sismiques, dont la composition pondérée reproduit la forme de toute trace sismique appartenant à la population étudiée.

L'intérêt de cette approche est lié :

- à la robustesse de l'analyse en composantes principales, surtout vu la taille de la population analysée (grand nombre de traces) ;
- à la possibilité de filtrage des données, ce filtrage pouvant être guidé par l'objectif géologique (la sélection des composantes principales à utiliser peut être guidée par l'étude des groupes de traces synthétiques aux puits détectés avec les composantes employées) ;
- à l'extension de la méthode aux données sismiques avant sommation : jusqu'à présent, la trace sismique a été étudiée après sommation mais les données sismiques avant sommation, sous réserve de les traiter spécialement, contiennent des informations supplémentaires liées aux propriétés géologiques du sous-sol (lithologies et fluides). L'étude du domaine avant sommation exige le traitement d'une masse de données énorme (au lieu d'avoir une trace pour chaque point milieu commun, on peut en avoir jusqu'à 48, en fonction de l'ordre de la couverture multiple). Il est possible de caractériser cette information par quelques paramètres, mais leur calcul peut entraîner des artefacts. Comme en caractérisation du faciès par des paramètres sur les fonctions

d'autocorrélation, spectre de puissance, module du signal analytique, (approche "fenêtre longue"), se pose le problème du choix des paramètres réellement discriminants pour le problème posé. Avec une caractérisation du type "facteurs de l'A.C.P.", le compactage des données sismiques peut avoir une **orientation géologique**. Il suffit de le faire en liaison avec une phase d'apprentissage sur des données sismiques où l'information géologique est supposée connue.

Signalons qu'après plusieurs études sur courtes fenêtres avec cette technique de caractérisation des traces, dont les résultats sont apparus très robustes et satisfaisants, nous l'avons testée sur longues fenêtres en comparaison avec la méthodologie retenue (calcul de paramètres). Nous avons retrouvé des résultats très analogues, en particulier lorsqu'il n'y a pas d'information géologique a priori (pas de puits disponible) et que les classifications automatiques sont utilisées en premier. Avec information géologique, si la fenêtre est longue (150-200 ms), le choix des facteurs de l'analyse en composantes principales est peut-être plus délicat, suite à la dispersion de l'inertie sur les axes principaux. Dans ce cas-là, la sélection des **paramètres sismiques** par la **phase d'apprentissage** sur les **traces sismiques de référence** et l'utilisation de ces paramètres discriminants pour la détection des faciès peuvent s'avérer plus efficace que les composantes principales.

Signalons enfin que la détection des groupes de traces sismiques par la méthode des composantes principales apparaît actuellement comme une étape préliminaire indispensable à certaines nouvelles méthodes d'inversion des données sismiques (Thadani, 1987, Fournier, 1989). Nous en discuterons dans le chapitre de la conclusion générale.

CONCLUSION GENERALE

Les travaux présentés dans ce mémoire s'inscrivent dans le cadre de l'**interprétation stratigraphique** des données sismiques. Plus précisément, nous nous sommes attachés à définir une méthodologie de reconnaissance automatique des **faciès sismiques** sur une fenêtre temporelle d'analyse. Nous n'avons donc pas abordé les problèmes d'analyse sismique séquentielle et nous avons supposé les principales séquences sismiques pointées et interprétées géologiquement.

Les travaux présentés ne concernent donc qu'une part limitée des problèmes rencontrés en interprétation stratigraphique, à la charnière avec l'interprétation lithologique des données sismiques. Les applications que nous avons traitées sont toutes liées à la reconnaissance du **faciès sismique fin**, en vue d'une **caractérisation lithologique** des séries pour des études de réservoir (applications à vocation "**gisement**").

En partant de l'analyse manuelle du faciès que fait l'interpréteur sismicien, nous avons constaté que l'analyse des faciès sismiques était une démarche fondamentalement **multivariable**. Plusieurs paramètres sismiques (fréquences apparentes, continuité, énergie des réflexions ...) sont pris en compte simultanément pour définir les zones homogènes de faciès sismique. Par ailleurs, l'interpréteur cherche à utiliser les caractéristiques des traces sismiques au voisinage des puits - où une information géologique est disponible - pour essayer de retrouver des zones sismiques analogues où il est possible de supposer les propriétés géologiques identiques à celles reconnues au puits.

C'est cette démarche que nous avons voulu automatiser, en espérant, outre un gain de temps, une amélioration de l'analyse en **quantifiant** l'information utile à la reconnaissance des faciès et en intégrant plus systématiquement et quantitativement l'information géologique des puits.

La démarche que nous avons suivie consiste à chercher à caractériser, le plus exhaustivement possible, les données sismiques dans le domaine après sommation, puis à mettre au point une méthodologie statistique pour traiter cette information.

Dans le cas d'une fenêtre temporelle de l'ordre de 100 ms, les paramètres sismiques utilisés pour caractériser les traces ont été calculés à partir de leurs représentations temporelles ou fréquentielles, module du signal analytique, fonction d'autocorrélation et son spectre. Ces paramètres mesurent l'énergie de la trace, sa répartition temporelle, la cyclicité éventuelle du signal sismique analysé, ses fréquences dominantes et leur importance relative. Ces paramètres physiques sont reliés de façon très complexe aux propriétés géologiques de la zone étudiée (épaisseurs et distribution des épaisseurs des bancs, contrastes lithologiques, propriétés pétrophysiques) et il serait illusoire de chercher à les relier directement à un attribut géologique. Par contre, ils caractérisent **globalement** les caractéristiques géologiques des séries concernées.

Ces paramètres sont également redondants (fortement corrélés) et tous ne sont pas discriminants par rapport au problème géologique posé. Le choix des paramètres les plus pertinents pour l'analyse des faciès est réalisé au cours de la méthodologie statistique.

La méthodologie statistique est essentiellement basée sur une étape d'apprentissage avec des traces sismiques sélectionnées au voisinage des puits les plus typiques du point de vue de leurs caractéristiques géologiques. L'hypothèse fondamentale qui est faite est que les caractéristiques sismiques de ces traces expriment les propriétés géologiques des séries rencontrées aux puits. La technique statistique clé est alors l'analyse factorielle discriminante linéaire, avec une procédure de pas à pas pour la sélection des variables les plus discriminantes. La limitation du nombre de variables prises en compte assure la fiabilité de l'analyse et permet d'évaluer la séparabilité effective entre les groupes de traces a priori. Une analyse en composantes principales vient corroborer ces résultats.

Par ailleurs, la liaison entre les variables sismiques discriminantes et les caractéristiques géologiques des séries aux puits est contrôlée avec une étape de modélisation. Un modèle des principales évolutions géologiques est réalisé à partir des corrélations entre les puits les plus typiques ; sa réponse sismique est calculée par convolution 1D. Sur ce modèle sismique synthétique, les variables sismiques discriminantes précédemment sélectionnées doivent présenter un comportement analogue à celui analysé sur les données réelles.

L'étape d'apprentissage est suivie d'une étape prédictive au cours de laquelle les traces sismiques "anonymes" sont affectées aux groupes a priori.

Les résultats obtenus en discrimination sont confirmés par des tests en classification automatique où aucune information a priori n'est utilisée. Cette phase confirmatoire est tout aussi importante que la phase d'apprentissage. En effet, la définition de coupures dans le continuum des données géologiques fait partie des points délicats de la méthode. Ces coupures sont à l'origine de la répartition des puits en classes de faciès géologiques, ces classes étant représentées ultérieurement par un (ou des) puits typique(s) au voisinage desquels seront sélectionnées les traces d'apprentissage. Les choix des coupures (nombre de classes, seuils...), puis des puits représentatifs des classes peuvent être discutables. La classification automatique est une façon intéressante de les valider.

De plus, en discrimination, le nombre de groupes a priori est défini par le nombre de faciès géologiques reconnus aux puits. Dans l'étape prédictive, les traces sismiques sont affectées nécessairement à l'un de ces groupes. Mais il est tout à fait possible que l'ensemble des puits n'ait pas reconnu **toutes les variations géologiques** existant sur le champ étudié. Dans ce cas-là, les méthodes de partitionnement, à condition de faire plusieurs essais avec différents nombres de groupes, peuvent révéler des groupes sismiques supplémentaires. Ces groupes peuvent soit traduire des faciès géologiques supplémentaires, non reconnus par des puits, soit correspondre à des caractéristiques de l'information sismique n'exprimant pas les propriétés

géologiques des séries étudiées (bruit sismique dans des zones faillées, problème à l'acquisition, etc.).

Il est donc particulièrement important de confronter et d'analyser simultanément les résultats obtenus au cours de la phase d'apprentissage et ceux de la phase confirmatoire afin de valider la carte des faciès sismiques et son interprétation géologique.

Dans le cas où aucune information géologique n'est disponible sur la zone, la méthodologie est légèrement modifiée : les techniques de classification automatique sont privilégiées par rapport à l'analyse discriminante qui ne vient que corroborer la fiabilité statistique de la séparabilité des groupes déterminés en classification automatique. La carte des faciès sismiques obtenue n'est pas interprétée géologiquement.

La méthodologie ainsi définie a été testée sur plus d'une quinzaine de champs aux stades de l'appréciation ou du développement. Elle est actuellement en cours d'industrialisation. Les résultats obtenus ont prouvé son efficacité pour la reconnaissance des faciès sismiques ainsi que l'intérêt qu'il y a à vouloir automatiser la démarche usuelle : **quantification** des différences de caractère entre les traces sismiques, prise en compte plus systématique et remise en cause éventuelle justifiée de l'information géologique a priori, détection de groupes sismiques supplémentaires, rapidité et objectivité de l'analyse.

La non prise en compte de la continuité latérale des réflexions (un des paramètres utilisés classiquement en interprétation des faciès sismiques) n'a pas posé de problème particulier pour la détection des faciès. En fait, la caractérisation morphologique verticale de la trace permet d'en rendre compte indirectement. Les difficultés posées par la redondance des paramètres sismiques mesurés et par la fiabilité discutable de l'information géologique a priori (classes de faciès géologiques...) sont maîtrisées de façon assez satisfaisante par les techniques statistiques utilisées (procédures de pas en pas en discrimination, méthodes de classification automatique).

Les points faibles de la méthode restent liés à l'interprétation géologique des faciès sismiques. L'hypothèse fondamentale que les traces sismiques voisines des puits traduisent les propriétés géologiques des séries observées aux puits n'est validée que par un modèle extrêmement simple. De plus, il n'y a pas unicité entre le faciès géologique et sa réponse sismique. L'étape de modélisation n'amène donc pas de preuves irréfutables pour l'interprétation géologique des faciès sismiques. Même si une évaluation **quantitative** de la séparation entre les groupes de traces sismiques d'apprentissage est faite, l'information géologique n'a été intégrée dans l'analyse que de façon **qualitative**.

Cependant, une prise en compte quantitative de l'information géologique (par l'utilisation, par exemple, de méthodes statistiques mesurant les corrélations entre les paramètres géologiques aux puits et les propriétés sismiques) n'est possible que si le nombre de puits disponibles est suffisant. Chaque classe de faciès géologique doit être représentée par un nombre statistiquement significatif de puits. Ceci limite l'intégration quantitative des données de puits

aux problèmes de gisement, lorsque le champ est déjà, en grande partie, développé. Les procédures de validation des résultats de l'interprétation géologique des faciès sismiques que nous avons évoquées (comparaison de la classification des traces sismiques adjacentes aux puits avec la classe du puits) ne sont d'ailleurs possibles que pour des applications "gisement".

Au stade d'appréciation d'un champ, il n'est donc guère possible d'aller au-delà de l'**intégration qualitative** de l'information géologique aux données sismiques, laquelle est réalisée dans la méthodologie statistique que nous proposons. Par contre, pour les problèmes de gisement, des méthodes statistiques visant à intégrer conjointement et quantitativement information géologique et information sismique sont en cours d'étude (Thadani, 1987, Fournier, 1989, voir Annexes).

Sous la poussée du développement des études de réservoir qui cherchent à intégrer le maximum d'information possible (information aux puits, information obtenue à partir des tests de production, information sismique ...), nous avons été amenés à adapter la méthodologie statistique de reconnaissance des faciès pour l'étude de **fenêtres temporelles très courtes** correspondant à des problèmes fins de variations géologiques. Si la méthodologie statistique n'est que peu modifiée, c'est la façon de caractériser les traces qui est redéfinie. Il n'est plus possible d'évaluer avec suffisamment de fiabilité les paramètres sismiques utilisés initialement.

Les méthodes d'analyse spectrale par maximisation de l'entropie, malgré leur réputation de robustesse vis-à-vis de la taille de la fenêtre d'analyse, ont donné des résultats décevants pour la caractérisation des faciès. Outre le problème de la détermination de l'ordre du modèle autorégressif, il n'est même pas assuré que ces méthodes soient réellement performantes dans le cas de signaux à bande large comme les signaux sismiques. La réputation de ces méthodes est presque entièrement établie sur l'étude de signaux synthétiques, à bande étroite.

La représentation conjointe temps-fréquence présente le grand avantage de mettre en valeur la loi de modulation du signal. Pourtant, les essais réalisés avec pour objectif la mise en évidence de faciès géologiques très fins ont été peu concluants. Cependant, nous pensons que cet outil d'analyse nécessiterait des études plus approfondies que celles que nous avons réalisées. Certaines compagnies pétrolières l'utilisent à l'heure actuelle avec succès pour le traitement des images sismiques (Thadani, communication personnelle), mais une meilleure caractérisation du **faciès géologique** avec cette représentation n'est pas encore prouvée (par comparaison avec d'autres moyens de caractérisation de la trace sismique).

Pour l'analyse fine du faciès sismique, nous avons préféré utiliser directement les **amplitudes de la trace**, traitées par **analyse en composantes principales** pour décorréler et filtrer éventuellement l'information. Cette méthode minimise les traitements intermédiaires qui peuvent générer des distorsions de l'information sismique. De plus, elle est très robuste numériquement. Les paramètres de la représentation des traces sont la taille de la fenêtre temporelle à analyser et le nombre de facteurs à prendre en compte dans les techniques statistiques. Ces paramètres sont déterminés préférentiellement sur les traces sismiques

synthétiques calculées à partir des données Sonic et Densité des puits. Des essais de classification automatique de ces traces synthétiques, dont les résultats sont interprétés avec la connaissance du domaine géologique, permettent d'évaluer l'influence de ces paramètres sur la classification et de les choisir.

La classification des traces sismiques est ensuite réalisée avec comme variables les k premiers facteurs de l'analyse en composantes principales qui ont été retenus.

Les applications "gisement" que nous avons réalisées montrent l'influence du pointé de la fenêtre d'analyse sur les résultats de la reconnaissance des faciès. Un décalage de quelques échantillons les affecte significativement. Les traitements des données sismiques appliqués **avant le pointé** doivent être faits avec beaucoup d'attention, en cherchant à intégrer toutes les informations de calage disponibles aux puits. Par ailleurs, il est illusoire de chercher à faire une interprétation faciologique fine d'un réservoir si son pointé est déjà douteux.

Dans le cadre des applications "gisement", la méthodologie statistique apparaît maintenant comme une phase préliminaire indispensable à certaines méthodes développées actuellement, pour extraire des données sismiques des informations géologiques **quantitatives** utilisables ultérieurement dans des modèles de réservoir (Thadani, 1987, Fournier, 1989). Ces méthodes recherchent les faciès sismiques avec une méthodologie identique à celle proposée pour les applications "gisement" puis utilisent les informations géologiques quantitatives des puits (porosités, pourcentages en lithofaciès, saturations ...) pour calibrer les traces sismiques de chaque groupe. Un intervalle de variations possibles pour le jeu de paramètres géologiques étudiés est déduit alors de chaque trace sismique. Cette information supplémentaire, par sa forte densité spatiale, améliore nettement notre connaissance des variations latérales des propriétés géologiques du réservoir. Plusieurs méthodes de calibration statistique peuvent être envisagées ("affine mapping", analyse canonique) mais quelle que soit la méthode, les résultats dépendent beaucoup de l'étape préliminaire où les traces sont regroupées, chaque groupe étant par la suite relié à un ou des puits. C'est pourquoi, dans le cadre de ces nouveaux développements, il peut être intéressant de raffiner l'analyse automatique du faciès. En particulier, une comparaison entre les résultats de méthodes paramétriques (que nous avons employées dans ce mémoire) et non paramétriques s'impose. Cette comparaison est actuellement en cours sur plusieurs jeux de données, de caractéristiques assez différentes.

Le traitement des données sismiques apparaît comme un domaine particulièrement intéressant pour l'application des méthodes statistiques multivariées. Ces méthodes permettent de prendre en compte toute la richesse de l'information sismique, de l'exploiter au mieux en intégrant l'information géologique. Depuis quelques années, on assiste au développement de ce type de travaux, développement qui devrait s'intensifier avec la demande accrue en géologie-géophysique de gisement. Nous avons déjà évoqué la voie "inversion statistique des traces sismiques", on peut citer aussi la prise en compte par des méthodes statistiques de l'information géologique **avant sommation** (particulièrement riche). Par rapport aux méthodes déjà discutées, cela revient à augmenter la dimension de l'espace de description des données

sismiques (variables supplémentaires). Cette voie, bien que moins avancée que "l'inversion statistique", est aussi très prometteuse.

ANNEXES

- Multivariate statistical analyses applied to seismic facies recognition, J. Dumay, F. Fournier, 1988.
- Extraction of quantitative geologic information from seismic data with multidimensional statistical analyses.
Part I : methodology.
Part II : a case history.
F. Fournier, 1989.

Multivariate statistical analyses applied to seismic facies recognition

Jean Dumay* and Frederique Fournier*

ABSTRACT

One of the most important goals of seismic stratigraphy is to recognize and analyze seismic facies with regard to the geologic environment. The first problem is to determine which seismic parameters are discriminant for characterizing the facies, then to take into account all those parameters simultaneously. The second problem is to be sure that there is a link between the seismic parameters and the geologic facies we are investigating.

This paper presents a methodology for automatic facies recognition based upon two steps. The first, or learning step, begins with the definition of learning seismic traces for each facies we wish to recognize. The choice of learning traces is based upon either well data or a seismic stratigraphic interpretation. A large number of seismic parameters are then computed from the

learning traces; multidimensional analyses are carried out in order to validate the choice of learning traces and to select, among all the available parameters, those that discriminate best. At this stage, a modeling step may be carried out to relate the seismic parameters to the geologic features. The second step is a predictive one which allows automatic facies recognition. We compute the previously chosen discriminant parameters on unknown seismic traces and classify the unknown traces with regard to the learning traces.

We develop the methodology and successfully apply it to two examples of reservoir facies recognition. Our main conclusion is that seismic traces contain geologic information that can be extracted by multivariate data analyses of a large number of seismic parameters. Automatic facies recognition is reliable and fast; the derived facies map has the great advantage of combining simultaneously several discriminant parameters.

INTRODUCTION

One of the most important goals of seismic stratigraphy is to define and recognize seismic facies. For instance, it is possible to distinguish three seismic facies in the regressive sequence in the deltaic environment shown on Figure 1. These facies correspond to inner platform shaly deposits at the bottom of the sequence, then delta front deposits, and, at the sequence top, deltaic plain deposits.

If an interpreter wishes to characterize the deltaic plain facies, he has to consider the energy of the reflections (low energy versus high energy for the two other facies); but if he would like to distinguish the delta front facies from the inner platform facies, he has to study an extra parameter, for instance, the apparent frequency. Indeed, the other two facies show rather high energy, but the delta front facies seems to have a higher frequency spectrum.

To separate the three main facies in this sequence, we have

to consider two parameters simultaneously. In this example, it was easy to see which parameters we could use to distinguish the different facies, but selecting parameters is not always this simple. Consider the following example (Figure 2) which comes from the study of lateral variations in a carbonate reservoir. At well B, the reservoir is very porous, but at well A, the facies lacks significant porosity. In order to define the porous facies limit, we have to know which seismic parameters characterize the porous facies; this is difficult.

To define the seismic facies, it is important to take into account all the information contained in the seismic traces; thus, we need to study simultaneously a large number of seismic parameters computed from the traces. Then we have to determine which variables, among all those we could compute, discriminate facies in each case being studied. The most efficient way to deal with these two requirements is to carry out multivariate analyses of the seismic parameters extracted from the traces.

Manuscript received by the Editor May 4, 1987; revised manuscript received January 15, 1988.

*SNEA(P) Centre Scientifique et Technique du Cami Salic, Avenue Larribau, B.P. 64018 Pau Cedex, France.

©1988 Society of Exploration Geophysicists. All rights reserved.

Statistical analyses have been applied for several years to seismic data in order to characterize a seismic facies related to a geologic environment. The techniques which have been used are conventional or multidimensional statistics. Most of the previous works use multidimensional data analyses because the techniques take a large number of seismic parameters into account simultaneously.

Conticini (1984) carried out conventional statistical analyses (cross-plots, star diagrams) on attributes of the traces such as continuity, instantaneous frequency, and analytical signal modulus. His aim was to identify the different facies encountered in fluvial deposits. Mathieu and Rice (1969) studied lateral variations in sand-shale ratio for a reservoir formation. They analyzed trace amplitudes by means of discriminant factor analyses and pointed out the possibility of grouping the traces (either synthetic seismograms computed from well data or real traces) with respect to the prevailing lithology. Hagen (1982) studied real seismic traces at a reservoir level in order to emphasize lateral variations of the formation porosity. He took into account frequencies estimated at the reservoir level. A principal components analysis reduced the number of parameters used to describe each trace. Then the new parameters, the principal components, were used in a clustering algorithm to group traces that correspond to porosity areas.

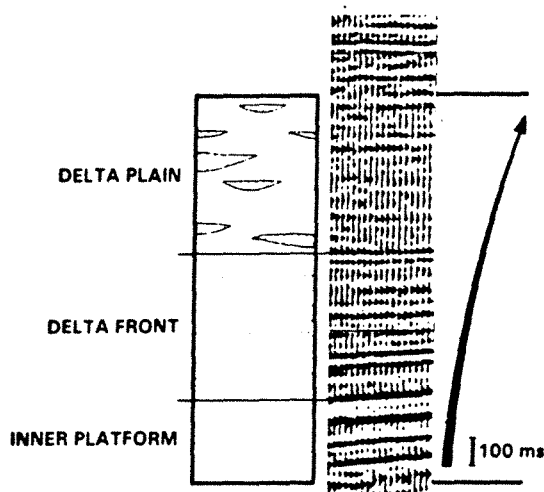


FIG. 1. Seismic stratigraphic interpretation: regressive sequence in deltaic environment. Discriminant parameters are energy and frequency.

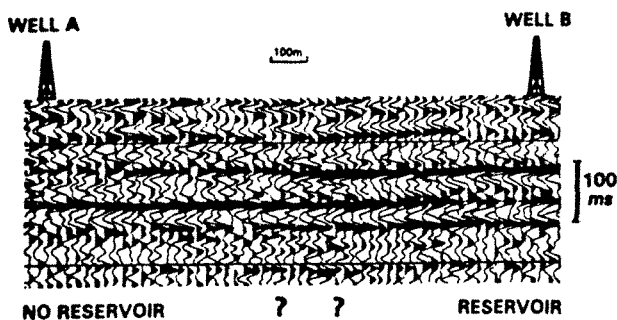


FIG. 2. Seismic stratigraphic interpretation: lateral variation of a carbonate reservoir. Discriminant parameters are not obvious.

Khattri and Gir (1976), Khattri et al. (1979), Sinvhal and Khattri (1983), Sinvhal et al. (1984) studied real and synthetic traces simultaneously. They generated synthetic sedimentary sequences by means of first-order Markov chains or by Monte Carlo simulations. Only a few lithologies were used in the studies, i.e., binary sequences with sand and shale or shale and coal. The impulse response was computed and analyzed with variables estimated from the autocorrelation function and from the power spectrum. Conventional and multidimensional analyses (histograms, Kolmogorov-Smirnov tests, discriminant factor analyses) on the variables permitted trace classification with regard to the sedimentary sequence type. The study of real seismic data, in formations with lateral sand-shale ratio variation, corroborated the value of applying discriminant analyses to the computed variables.

Bois (1980, 1981, 1982) developed two techniques based on pattern recognition: one with and one without prior learning. Bois' techniques were tested on a reservoir formation in order to define facies limits. Bois computed, on the portion of each seismic trace at the level of the reservoir, the first three coefficients of an autoregressive adjustment of the traces. After defining a pseudodistance based on these coefficients, Bois applied clustering techniques. The process with prior learning calibrated seismic facies (traces) to a nearby well. If the studied traces were near the learning traces (as defined by the pseudodistance), the seismic facies were classified with the corresponding calibrated facies.

Matlock et al. (1985) characterized the rapid facies variations in a 150 ms thick reservoir formation. Parameters were related to the frequency characteristics of the traces and seismic traces classified by comparison to reference traces located near wells where reservoir facies were found. Multidimensional statistical techniques clearly defined the limits between the different facies on the two seismic lines studied.

METHODOLOGY

Our methodology uses two types of multidimensional analyses: clustering techniques and factor analyses (Figure 3). If n denotes the number of measured seismic variables, each trace can be represented by a point in n -dimensional space, the coordinates of which are given by the values of the trace's n variables. Clustering techniques group points in an n -dimensional vector space based upon their proximity (or distance) and provide us with a number of different classes or groups of traces in this space. Factor analyses try to summarize the initial information in the most efficient way by computing a new vector space, of lesser dimension than the original, generated by linear combinations of the initial variables. The points are then projected from the original n -dimensional (variables) space to the new p -dimensional (factors) space, $p < n$ (Figure 3).

The aim of principal components analysis is to compute p orthogonal axes such that the projection of the points $x(i)$ onto the new space is similar to the initial positions of the points. The criterion for finding the new axes is to minimize the distortion which is necessarily introduced by the projection. When we carry out a discriminant factor analysis, we must a priori gather the points in classes. The criterion used to determine the new axes is that the projected points belonging to the same group are as near one another as possible and

that the group projections are as far apart as possible. The main difference between the two techniques is that in discriminant factor analysis the spatial distortion of the data may be increased in order to emphasize the differences among the a priori groups.

The methodology we have developed uses both clustering techniques and factor analyses; however, discriminant factor analysis has a dominant role. Additional information on statistical analyses is presented in the Appendix. Indeed, our aim is to characterize and distinguish the traces related to particular geologic environments.

Steps of the methodology

There are two distinct steps in the methodology, a learning step followed by a predictive step. In the learning step, we

determine if discrimination among the traces is possible calibrate the seismic data being studied with data from nearby wells. This step is necessary because of the great heterogeneity of seismic data caused by geographic variation and variation in the acquisition and seismic processing. The predictive step provides a classification of unknown seismic traces. Let us examine the two steps in detail.

Learning step.—Learning or reference seismic traces are defined for each seismic facies based on a seismic stratigraphic study or a comparison to data from nearby wells for which geologic information is available (Figure 4). Each set of traces represents a specific seismic facies. The choice of learning traces is essential to the whole study. It is necessary to validate our choices by carrying out statistical analyses such as clustering or principal components analyses on the parameters

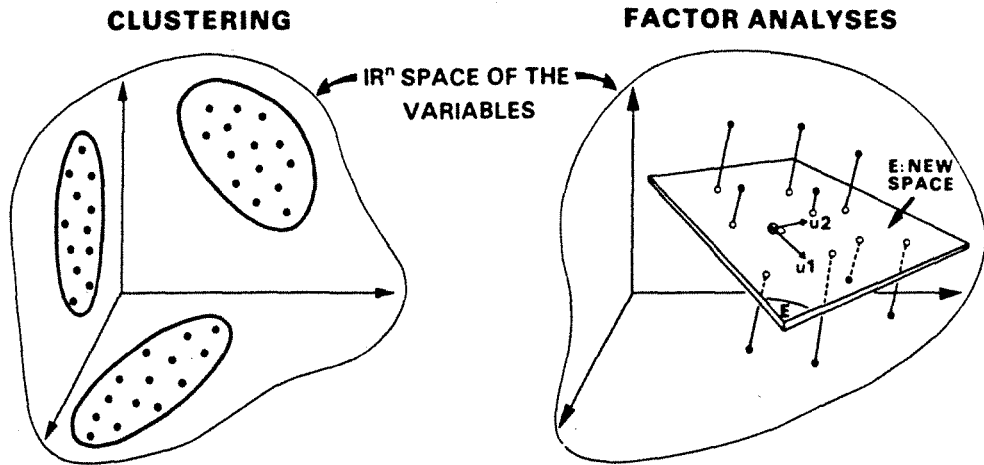


FIG. 3. Multivariate data analyses: clustering techniques and factor analyses. Clustering gathers points based on their proximity in variable space. Factor analyses compute a reduced number of variables to most efficiently represent the original information.

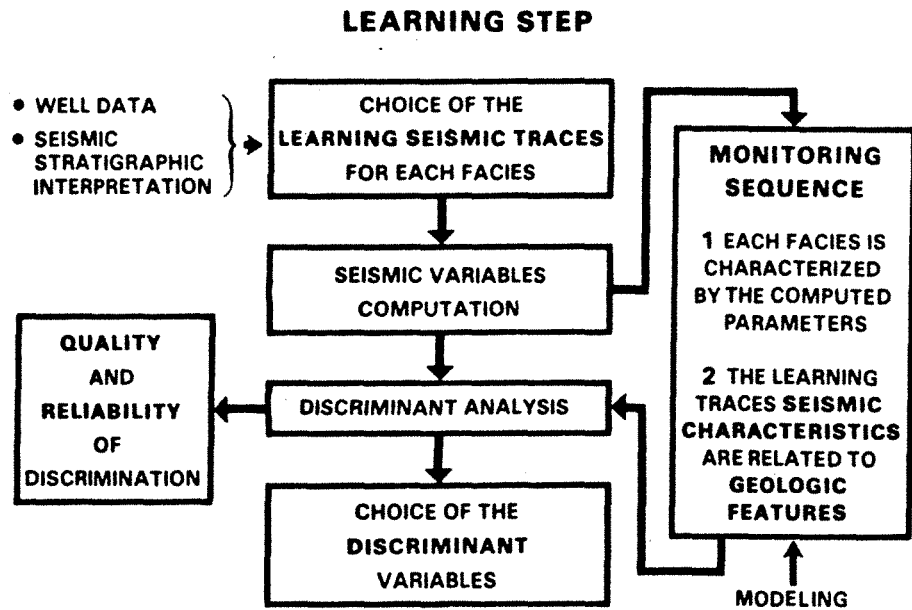


FIG. 4. Flow chart for learning step.

measured on the learning traces. Those controls allow us to evaluate the distances among the individual groups of learning traces and also to estimate the power of discrimination of the measured variables.

Then we apply discriminant factor analyses to the learning traces. The technique used is a stepwise discriminant analysis by stepwise selection of variables that are the most useful for discriminating among the a priori classes. The classes are assumed to be multivariate normal with a common covariance matrix. Variables are chosen to enter or leave the model according to the significance level of an *F* test from an analysis of covariance, where the variables already chosen act as covariates and the variable under consideration is the dependent variable. We can also evaluate the reliability and the predictive power of the analysis.

If all the analyses show that discrimination among the learning traces is possible, we carry out a modeling step to establish that the chosen discriminant parameters are related to the geologic variations we aim to characterize. If the result of all analyses is positive, we address the predictive step.

Predictive step.—We compute for unknown seismic traces, that is, traces for which we want to determine the facies, the previously chosen discriminant parameters and classify the traces with regard to the learning traces. We present two applications of the methodology.

EXAMPLE IN A CLASTIC ENVIRONMENT

Figure 5 is a time-structure map of the oil field we studied. One of the most important problems encountered in ex-

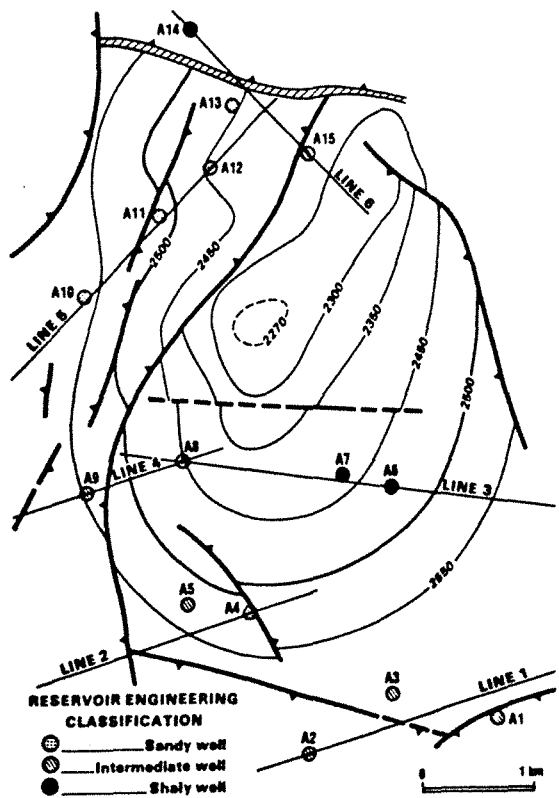


FIG. 5. Time structure map of the oil field studied.

plotting the field is very rapid and unpredictable variations in reservoir quality. The reservoir corresponds to sandy turbidites interbedded with shaly formations. Reservoir engineering studies have shown that the reservoir quality depends mainly on the sand-shale ratio of the formation. Our aim is to predict the reservoir quality variations by studying the seismic traces recorded over the west part of the structure (Figure 5).

Learning step

For each well, we know the sand-shale ratio of the reservoir formation. Reservoir engineering studies provide a well classification into three main reservoir types: sandy reservoir, intermediate reservoir, and shaly reservoir. We chose for each reservoir facies a well which we assume to be typical of the related facies. The learning traces are traces near this well. We chose well A12 to represent the sandy facies, well A3 the intermediate facies, and well A6 the shaly facies.

Next we computed for each learning trace a large number of seismic parameters. These parameters were related to the time or frequency representation of the signal and were estimated from the autocorrelation function, the power spectrum, and the analytical signal modulus. Table 1 gives a list of the computed variables. Figure 6 shows, as an example, the parameters extracted from the power spectrum.

It is possible that the three chosen wells have distinctly different reservoir facies, but that the seismic traces from near the wells are very similar or that the computed variables do not discriminate among them. In order to study these possibilities, we carried out a principal components analysis on the learning traces. This type of analysis was selected because

Table 1. Computed variables.

Autocorrelation	
AM1 } AM2 }	Positive peaks
TAM1 } TAM2 }	Times corresponding to the previous peaks
TPZ1 TPZ2 TPZ3 TPZ4	Zero crossings
Power spectrum	
FM, Q_1, Q_2, \dots, Q_9 are defined in Figure 6.	
QW_1, QW_2, \dots, QW_9 are the frequencies at which the deciles of the weighted cumulative power spectrum occur.	
Analytical signal modulus	
ET:	Total energy
E1:	Time at which 10 percent of the total energy has arrived
E2:	
⋮	
E5:	Time at which 50 percent of the total energy has arrived
E6:	
⋮	
E9:	Time at which 90 percent of the total energy has arrived

it minimizes the distortion introduced by projecting the points. If the principal planes take into account a sufficient percentage of total variance and if the projected learning traces are grouped in the same way as the facies, then we can conclude the computed variables are sufficient to characterize the three facies.

Figure 7 shows the principal plane 1-2 of the analysis; more than 55 percent of the total variance was taken into account by this plane. The axes are linear combinations of the initial variables. Note that the traces representing each facies are grouped together. Thus, we can state that there are differences among the learning traces as far as geophysical variables are concerned. By contrast, if the learning traces had not plotted

as distinct groups but were all mixed up, it might not have been possible to characterize the facies by studying the seismic traces. This validation step establishes whether or not the learning traces can be grouped appropriately by the discriminant variables but does not demonstrate that there is a link between the trace characteristics and the geologic facies. We must carry out a modeling step to establish that link.

Next, we applied a discriminant factor analysis on the learning traces (Figure 8) in order to select those parameters which provide the best discrimination. By selecting discriminating parameters, we keep the ratio of the number of traces to the number of parameters high. A high ratio guarantees that the analysis is reliable. Four variables were chosen, two variables from the power spectrum and two from the analytical signal modulus of the traces. Together they produced a discrimination of 85 percent for the learning traces.

The learning step showed that the computed variables could appropriately group the seismic traces and helped us to choose the parameters most useful in discriminating the traces

Predictive step

The unknown seismic traces from the seismic lines on the west area of the structure (Figure 5) were analyzed at the level of the reservoir. We computed the four chosen discriminant variables for these reservoir portions of the traces in order to evaluate their similarity to the learning traces. A portion of a line studied is shown in Figure 9, together with the assignment of reservoir quality given by the program. The two classifications shown in Figure 9 correspond to different combinations of variables. The results are nearly the same; thus, the statistical analyses appear to be stable. Note that the traces near well A5 have been classified as an intermediate facies and those near well A4 as a sandy facies. These results correspond to the reservoir quality known at the two wells.

We can check the classification's validity by systematically comparing the known reservoir quality in all the wells not used to define the learning traces with the predicted quality

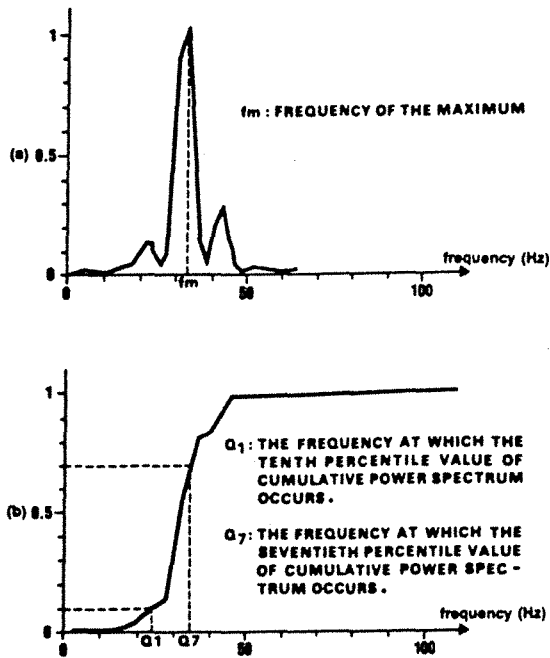


FIG. 6. Seismic variables computed from power spectrum.

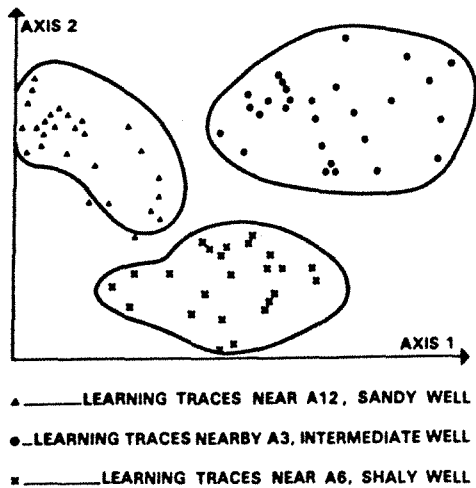


FIG. 7. Validation of the learning traces choice: principal components analysis, principal plane 1-2.

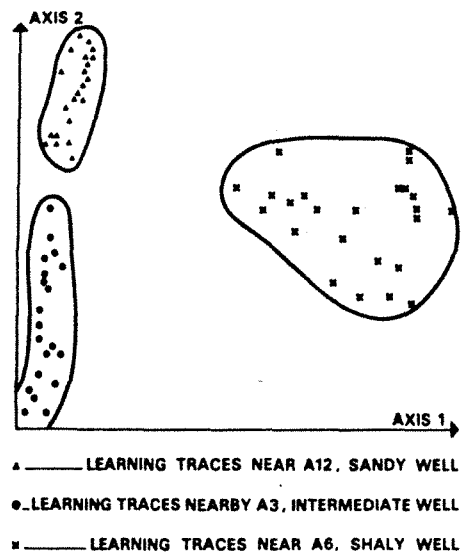


FIG. 8. Choice of the discriminant parameters: discriminant analysis, discriminant plane.

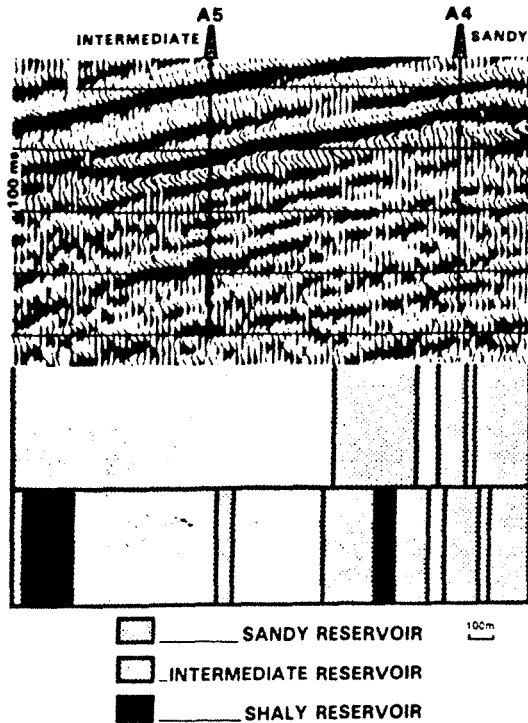


FIG. 9. Line 2: predicted quality of the reservoir between wells A4 and A5.

proposed by multivariate data analysis of the seismic traces near the wells. Table 2 shows the results of this comparison. The classification is successful in 85 percent of the wells. The results of the analysis are reported on the synthesis map (Figure 10). Only two facies were distinguished: sandy reservoir and intermediate to shaly reservoir. The dashed area corresponds to zones where the facies was indeterminate, that is, the assignment of the traces varied from one trace to the next, making it impossible to find spatial coherency. In spite of this indeterminate zone, the results we obtained seem to be reliable and allow an evaluation of the extent of the sandy reservoir facies.

EXAMPLE IN A CARBONATE SETTING

The second application is a study of rapid reservoir variation in a carbonate setting. The productive formation is an oolitic limestone (porous facies) which varies laterally to an oncolitic limestone. Figure 2 shows a seismic line linking wells B and A. At well B, the formation of interest is porous (oolitic dune facies), but at well A, the formation has very low porosity (oncolitic flat facies). Our aim is to find the porous facies limits by studying the seismic lines from the field.

Learning phase

The learning traces were chosen near wells B (dune facies) and A (oncolitic flat facies). They were analyzed for a 120 ms thick portion of the seismic traces at the level of the reservoir. For those traces, we computed variables associated with the autocorrelation, the power spectrum, and the analytical signal modulus.

Table 2. Comparison between multivariate analysis classification and reservoir engineering classification.

Seismic line	Well	Reservoir engineering classification	Multivariate analysis classification
1	A2	I	I
	A3	I	R
	A1	I	I
2	A5	I	I
	A4	S	S
3	A8	S	S
	A7	SH	SH
	A6	SH	R
4	A9	I	I
	A8	S	S
5	A11	S	S
	A12	S	R
	A10	S	S
6	A14	SH	I
	A13	I	S
	A15	I	I

Successful classification to 85 percent

S: sandy; SH: shaly; I: intermediate; R: reference

A discriminant analysis carried out on the learning traces was used to aid in the selection of a subset of variables that could discriminate between the two facies. The most important variable, which we denote as E6, corresponds to the time at which 60 percent of the total energy has arrived. We studied the changing character of this variable on the learning traces (Figure 11). Traces from the dune facies always had smaller values for E6 than did the traces of the oncolitic facies. If we look carefully at the analytical signal modulus for the traces (Figure 11), we note that the energy distribution is more uniform for the dune traces and rather focused at the end of the interval for the oncolitic flat traces.

Tests of the choice of learning traces and of the meaning of variable E6 were carried out. The first tests confirmed the choice of traces. For the second test, we reduced the interval to the lower 60 ms (Figure 12). The new discriminant variable is E2, the time at which at least 20 percent of total energy has arrived. The behavior of E2 is very similar to that of E6 in the analysis of a 120 ms interval. This test confirmed that the prevailing characteristic of oncolitic flat traces is a late arrival of most of the energy with a concentration at the base of the interval studied.

Predictive phase

As in the previous application, we assigned each unknown trace to one of the two groups. Figure 13 shows the results obtained on the portion of line linking the two reference wells. It is clear that we can distinguish two areas: a prevailing dune facies area and a prevailing oncolitic flat area. Results obtained for two sections are shown on the isopach map of the oolitic dune (Figure 14).

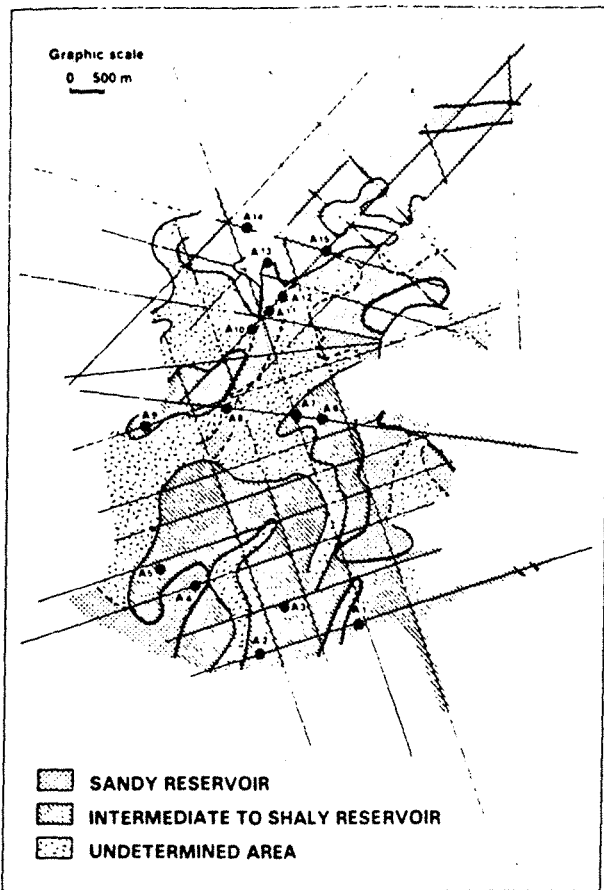


FIG. 10. Predicted reservoir facies map.

These results were confirmed by the classification techniques where no a priori information was used. We see that traces near wells C and D are correctly assigned (oncolitic flat facies) and that the transition between oncolitic flat and oolitic dune corresponds to the area of thickening oolitic facies (isopach 25–35 ms). Indeed the oolitic thickening is linked with the oncolitic limestone disappearance. A very strong acoustic contrast exists between the oolitic formation and the inner lagoon deposits. This strong impedance contrast is responsible for the appearance of a strong peak at the reservoir top (the total energy is then more uniformly distributed over the interval studied).

DISCUSSION AND CONCLUSIONS

The methodology we have described is based on a learning step with computation of seismic parameters for the learning traces. Learning is followed by multidimensional analyses and a predictive step which allows an automatic facies classification. The automatic recognition is reliable and fast; the facies map obtained combines several discriminant parameters simultaneously.

The most important problem remaining is to make the connection between the seismic facies we have characterized and the geologic facies we are trying to characterize. The physical meaning of some of the chosen discriminant variables is not entirely clear. A modeling step sometimes helps relate the geo-

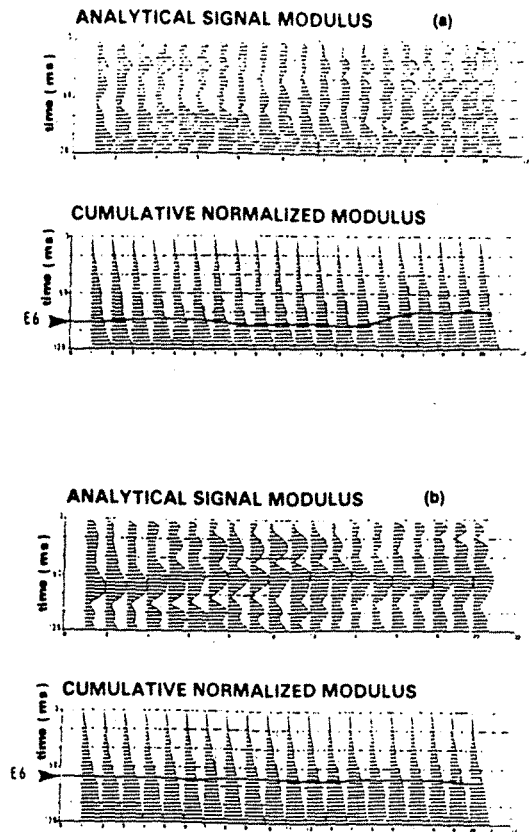


FIG. 11. Learning traces: study of the analytical signal modulus. (a) Oncolitic flat facies learning traces. (b) Dune facies learning traces.

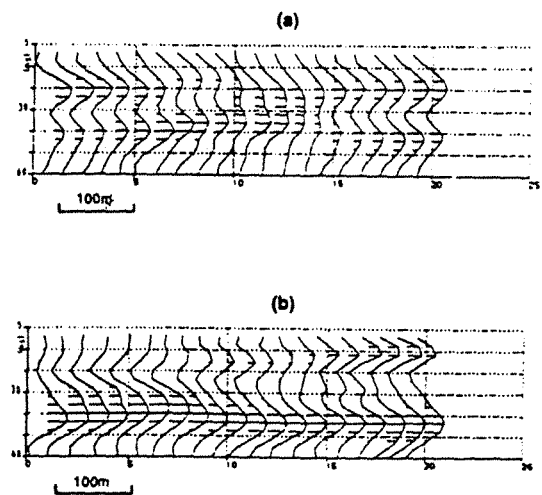


FIG. 12. Learning traces: reduced time interval. (a) Dune facies, (b) oncolitic flat facies.

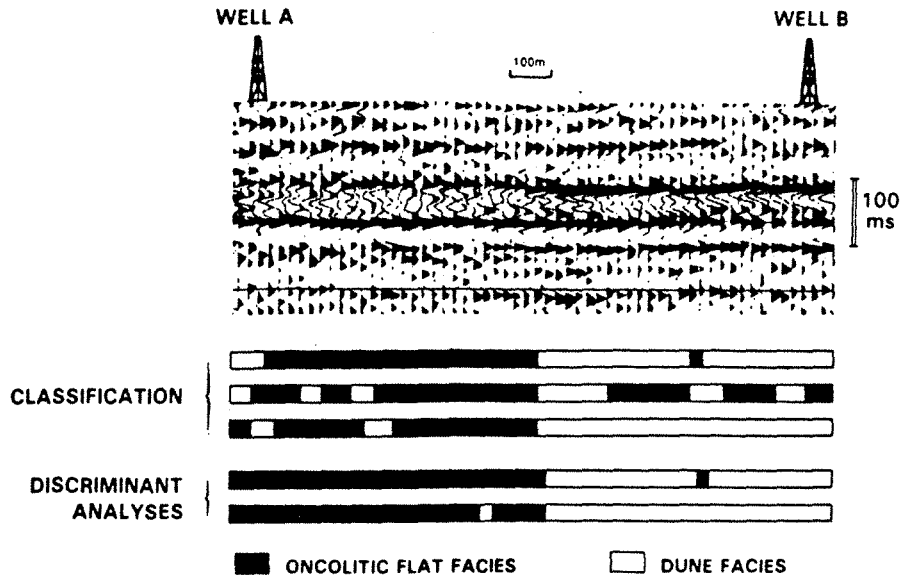


FIG. 13. Lateral variations of the carbonate reservoir; predicted quality.

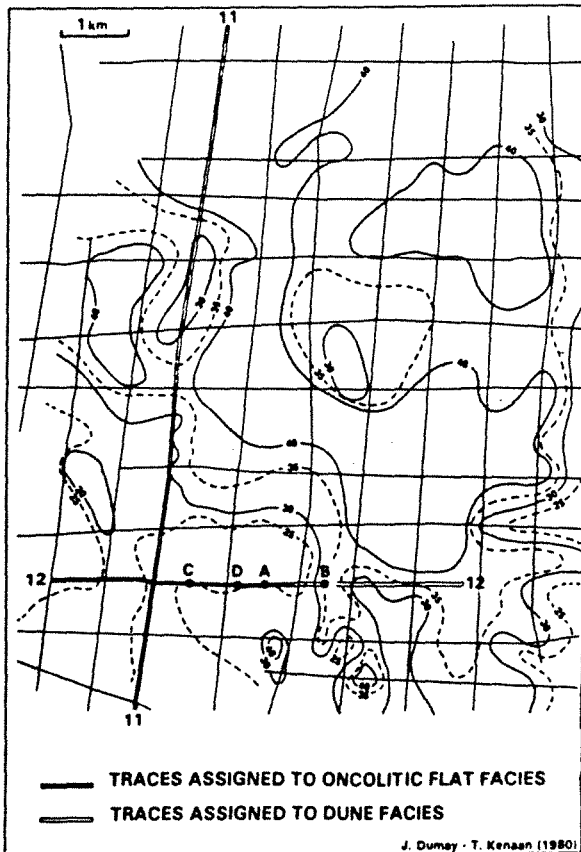


FIG. 14. Time isopach map (in milliseconds) of the oolitic facies with predicted quality of unknown traces.

logic variations to the measured seismic parameters, but this step is not always sufficient.

There are problems with the way we evaluate some of the seismic parameters; for example, the parameters computed from the spectrum. Indeed, it is very difficult to estimate the frequency characteristics of a signal for a very short time interval.

Besides interest in the methodology for direct application to petroleum exploration, we wish to emphasize that our results indicate the following:

- (1) seismic traces do contain geologic information which until now has been seldom used;
- (2) it is possible to extract this information by means of the study of seismic parameters;
- (3) multidimensional data analysis techniques can be suitably applied to this information.

The available techniques are numerous and the possibilities of applying them to geophysical data have been studied very little.

ACKNOWLEDGMENTS

We are grateful to Société Nationale Elf Aquitaine Production for permission to publish this paper.

REFERENCES

- Anderberg, M. R., 1973. Cluster analysis for applications: Academic Press Inc.
- Bois, P., 1980. Autoregressive pattern recognition applied to the delimitation of oil and gas reservoirs: *Geophys. Prosp.*, **28**, 572-591.
- 1981. Determination of the nature of reservoirs by use of pattern recognition algorithm with prior learning: *Geophys. Prosp.*, **29**, p. 687-701.
- 1982. Some comments on the application of pattern recognition to oil and gas exploration: *Geoexplor.*, **20**, 147-159.
- Conticini, F., 1984. Seismic facies quantitative analysis: New tool in stratigraphic interpretation: *54th Ann. Internat. Mtg., Soc. Expl. Geophys.*, Expanded Abstracts, 680-683.

- Hagen, D. C., 1982. The applications of principal components analysis to seismic data sets: *Geonexpl.*, **20**, 93-111.
- Khattari, K., and Gir, R., 1976. A study of the seismic signatures of sedimentation models using synthetic seismograms: *Geophys. Prosp.*, **24**, 454-477.
- Khattari, K., Sinval, A., and Awasthi, A. K., 1979. Seismic discrimination of stratigraphy derived from Monte-Carlo simulation of sedimentary formations: *Geophys. Prosp.*, **27**, 168-195.
- Mathieu, P. G., and Rice, G. W., 1969. Multivariate analysis used in the detection of stratigraphic anomalies from seismic data: *Geophysics*, **34**, 507-515.
- Matlock, R. J., McGowen, R. S., and Asimakopoulos, G. T., 1971. Can seismic stratigraphy problems be solved using automated pattern analysis and recognition?: 55th Ann. Internat. Mtg. Soc. Ex Geophys., Expanded Abstracts, 582-585.
- Rao, C. R., 1965. *Linear statistical inference and its applications*: John Wiley and Sons.
- Sinval, A., and Khattari, K., 1983. Application of seismic reflectivity data to discriminate subsurface lithostratigraphy: *Geophysics*, **48**, 1498-1513.
- Sinval, A., Khattari, K., and Awasthi, A. K., 1984. Seismic indicators of stratigraphy: *Geophysics*, **49**, 1196-1212.

APPENDIX
MULTIDIMENSIONAL STATISTICAL ANALYSES

p denotes the number of variables and n denotes the number of observations. The population is represented as a set of points in a p -dimensional space (variables space). The statistical techniques we apply to study the set of points are *clustering techniques* and *factor analyses* (Anderberg, 1973; Rao, 1965).

Clustering techniques

We group the points in k clusters in the variables space after having defined a distance between points and a similarity index.

Factor analyses

We project the points from the variables space onto a new space of *lesser dimension*. The new space has components which are linear combinations of those in variables space. Different criteria are used to compute the components of the new space depending upon the technique used. The main techniques used are principal components analysis and discriminant factor analysis.

N is the set of observations

$$N = \{(x_i, p_i), i = 1, \dots, n\},$$

where x_i is the p -dimensional observation vector and p_i is a weight.

The covariance matrix Y is

$$Y = \sum_{i=1}^n p_i (x_i - g)(x_i - g)^T,$$

where g is the mean of the n observations.

Principal components analysis.—The problem is to find a space of dimension $p' < p$ such that the distances between the projected points are closest to the distances in the original space; i.e., we must minimize the distortion introduced by projection onto the space of lesser dimension.

This problem is solved by computing u_i ($i = 1$ to p), the eigenvectors of the matrix Y . The vectors u_i generate new axes and each eigenvalue λ_i measures the percentage of original information taken into account by the new axis u_i . Generally, only a few p' axes are sufficient to represent most of the original information. Thus, it is possible to project the points onto this p' dimensional space with only a small loss of information.

Discriminant analysis.—The n observations have been gathered in k groups P_1, P_2, \dots, P_k defined a priori and $p_i = 1/i$. a_r is the mean of the points in group P_r and μ_r is its weight $\mu_r = \text{card } P_r/n$. The problem is to find a new space of dimension $p' < p$ on which to project the points such that the projected clusters' centers are as far apart as possible and so the projected points from the same cluster are as close together as possible.

T , the total covariance matrix, is the matrix Y with $p_i = 1/i$. B , the between-group covariance matrix, is given by

$$B = \sum_{r=1}^k \mu_r (a_r - g)(a_r - g)^T.$$

W , the within-group covariance matrix, is given by

$$W = \sum_{r=1}^k \mu_r Y^r,$$

with

$$Y^r = \frac{1}{\text{card } P_r} \sum_{i \in P_r} (x_i - a_r)(x_i - a_r)^T.$$

It is possible to demonstrate that:

$$T = B + W$$

The axes we search must maximize the ratio B/W , or equivalently B/T , since $T = B + W$. This problem is solved by computing eigenvectors u_i and eigenvalues λ_i of the matrix $T^{-1}B$. The vectors u_i generate the $(k - 1)$ discriminant axes and λ ($0 < \lambda_i < 1$) is a measure of the discriminant power of the corresponding axis.

Extraction of Quantitative Geologic Information from Seismic Data with Multidimensional Statistical Analyses.

Part I: Methodology

SI 2.4

Frederique Fournier, ELF-Aquitaine, France

SUMMARY

This paper presents a statistical methodology which derives quantitative geological information from seismic data by means of multidimensional statistical analyses applied both on geological and geophysical data. Well data provide global geological parameters taking into account the lithologies, porosities and thicknesses of the studied interval. These parameters generate a n multidimensional space where the wells are represented. In this space, multidimensional analyses are carried out in order to find reliable well groups with explicative variables for each significant group. The seismic traces are then analysed on the time window corresponding to the geological interval. The traces are represented by points in a k multidimensional space which is generated by the first k factors from a principal components analysis carried out on the amplitude series of the traces. Clustering and discriminant analyses allow to draw seismic facies maps, and also to confirm the geological meaning of the seismic groups by the assignment of synthetic traces in those groups. The clusters found in the two spaces are simultaneously analysed to find seismic trace groups corresponding to well groups. These groups are used to calibrate the seismic groups. The traces are then inverted towards the geological space by an affine mapping. It allows to derive bounds on each initial geological parameter and at each trace location.

INTRODUCTION

Since Exploration is increasingly devoted to the search for subtle traps and since there is a strong need from Reservoir Engineering for information between the wells, a tremendous interest is raised to obtain quantitative information from seismic data.

Seismic data are a source of potential geological information of very high spatial density, compared to well data. Reservoir models are very often built from this scarce well information, but it has been demonstrated (Alabert, Journel, 1987) that the use of extra information, even very fuzzy, but well spatially distributed, appreciably improves the reliability of the reservoir simulations.

Two approaches were developed, these last years, to predict geological information from seismic data. The first one, well illustrated in the literature, mostly deals with the evaluation of the porosity (Guidish, de Buyl, 1987, Solano, Schneider, 1987, Bell, de Buyl, Guidish, 1988) and consists in an inversion of the seismic data at the reservoir level. A process of calibration of the acoustic impedances is then carried out with porosities and impedances data available from a few wells. The results provided by this method are fairly good as long as the geological setting is simple, i.e. a unique geological parameter shows variations and the size of the reservoir is consistent with the resolution power of the seismic survey.

The other approach is statistical and has been less explored. Seismic traces are represented in a multidimensional space and processed in order to find clusters of traces (corresponding to different seismic

facies). Either these groups of traces are statistically calibrated in terms of geology (Alabert, Thadani, 1987) or the representation seismic space is calibrated through the study of the synthetic seismic response of the main geological patterns (O'Connor, 1986). Doyen (1988) has shown an application of cokriging which consistently integrates seismic data with well porosities. The interest of such an approach is that it takes into account the whole available information. It allows also to deal with complex situations where several geological features are varying simultaneously and where the reservoir may consist of different units.

The methodology presented in this paper (Figure 1) is purely statistical. It derives bounds on global geological parameters after seismic traces, by means of multidimensional statistical analyses carried out on both geological and seismic data.

STUDY OF THE GEOLOGICAL DATA (Figures 1 and 2)

Geological data are mainly provided by well information; they are characterized by their local scale (as opposed to seismic data) and their heterogeneities (direct core measures, indirect measures after a complex interpretation of logs).

First, the interval that is going to be analysed has to be defined. If the reservoir engineers wish to evaluate the porosity of a particular layer, the geophysicists must keep in mind that the geological interval has to be consistent in size with the time interval on which seismic facies analysis is possible.

I then search for geological parameters which represent the characteristics of the studied interval. These parameters come from the original geological data through averaging (porosities), simplifying (lithofacies) and totalization (thicknesses, percentages of lithofacies...). They globally represent the pattern of the reservoir and generate what I call the "geological space" of dimension n , in which a well appears as a point of which coordinates are the values of the variables taken at the well.

Since the geological properties are correlated, I can use principal components analysis to summarize in the most efficient way the major trends in the data. A new set of few axes can practically represent all the initial information. Wells are projected onto the new space. Loss of information is controlled by the number of factors used to generate the new space. In this space, wells are gathered in groups. Indexes that measure the separation between the groups are available. This allows to choose the best clustering result (as far as the number of groups and the number of factors are concerned). On this result (k groups), I carry out a discriminant analysis, using the geological parameters as variables. The stepwise discriminant algorithm selects the variables which best explain the differences between the groups and measures the reliability of the discrimination.

Multidimensional analyses in the geological space allow (Figure 1) to find reliable groups of wells and to detect the geological meaning of each group by a set of explicative features (geological portrait).

ANALYSIS OF THE SEISMIC DATA (Figure 3)

The seismic data under analysis are:

- field seismic traces,
- synthetic traces computed from well data (velocity and density logs),
- model traces expressing the seismic response of the principal geological patterns.

The time interval corresponding to the studied reservoir interval must be defined from the synthetic traces. This interval is a parameter the influence of which should be analysed for very short windows. Most reservoir engineering problems, in fact, involve very short time windows: 20 to 30 ms. Edge effects can be quite important in this case.

The best way to characterize the morphology of the traces, in this interval, is to directly use the succession of trace amplitudes. Each trace is represented in an m dimensional space (Figure 4) by a point of which coordinates on the m axes are successively the amplitudes of the m samples of the time window which is analysed. The amplitudes are highly correlated variables: the value of sample k strongly depends on the values of samples $(k-1)$ and $(k+1)$. So, I apply principal components analysis (P.C.A.) to decorrelate the initial variables and to extract the principal trends of the data (Figure 4). The loss of information, linked to the projection of the traces onto the new space of smaller dimension, is a filtering effect that must be controlled by choosing the number of factors that are used to generate the space.

The previous parameters (DT, size of time window and K , number of P.C.A. factors to keep) are chosen from test results on synthetic and model traces. The tests consist in looking at various clusterings where DT, K , and the number of groups vary. The optimal result (best separation between the groups) provides the parameter values and clues as to the geological meaning of the synthetic trace groups, when model traces of a geological pattern are classified with the synthetic traces at the wells which constitute a group in the geological space defining the same pattern.

The field seismic traces are then analysed. Before quantitative studies are undertaken, the best stratigraphic processing sequence is sought in order to produce the highest resolution possible at the reservoir level.

I use the synthetic traces to calibrate the field seismic events and to control the picking of the top and the bottom of the time window being analysed. The shorter the time interval, most important is the quality of the picking.

All the portions of traces are then projected onto the K factors space, in which clustering tests are carried out. The best result helps to define learning traces for each group (traces belonging to the centers of the groups). The learning traces are a priori information to use in a discriminant analysis (Figure 3). The descriptive step of a discriminant analysis computes axes to separate the groups, and a measure of their separability, which is related to the reliability of the next predictive step. The predictive step classifies all the traces that are not used as learning traces in the previous a priori groups; synthetic and model traces are also classified, following a homogeneization process. Seismic facies can then be obtained.

COMPARISON OF THE GEOLOGICAL AND GEOPHYSICAL SPACES AND INVERSION OF THE TRACES (Figure 5)

Now, I compare the clusters of points in the geological space (where points are wells) to the clusters in the geophysical space (where points are synthetic well traces, model traces and field traces). I decide that a group of traces which is sufficiently homogeneous and contains field traces, but also synthetic traces of wells gathered together in the geological space and model traces of the pattern representing the geological cluster, is going to be calibrated with the corresponding wells information.

If it is not possible to calibrate the seismic groups, then the seismic traces belonging to them will not be inverted.

Each trace belonging to a group for which calibration is possible is regarded as a point M on an error ellipsoid (E) around the geophysical center (Figure 5). This ellipsoid (E) is transformed into an ellipsoid (E') in the geological space, around the center of the calibration group of wells. The transformation used is an affine mapping (Thadani, 1987); the point M' , image of M in the geological domain, belongs to (E'), so I use the ellipsoid (E') to derive bounds for the different axes of the geological space. These bounds depend only on the centers and the covariances of the groups in both domains and also on the distance of the trace M to the center of its group.

From the bounds on the factors, I derive bounds on the initial geological variables, going back to the previous reference axes before the transformation by principal components analysis. I obtain maps of minimum and maximum values, for each geological parameter and at each seismic trace location.

CONCLUSION

The bounds derived from the seismic data concern global geological features of the reservoir. They provide extra information to reservoir simulations.

The statistical methodology I have described allows to take into account the whole available information. The multidimensional analyses process the parameter correlations, efficiently summarizes the trends in the data, and reveals the possible correspondences between seismic data and geological features.

This method requires a large amount of data but is still useful even in complex environments where different geological variations are interfering. The geological information that is extracted may be porosities, thicknesses or lithologies, depending on the way the wells were described in the geological domain.

ACKNOWLEDGMENTS

I am grateful to Elf Aquitaine for allowing me to publish this research. My sincere thanks are due to my colleagues and friends who revised the manuscript for their advices and their essential linguistic improvements.

REFERENCES

de Buyl, M., Ullah, S., Guidish, T., Reservoir description

from seismic lithologic modelling, part 2: substantiation by reservoir simulation, 1987: S.E.G. expanded abstracts.

de Buyl, M., Guidish, T., Bell, F., Reservoir description from seismic lithologic parameter estimation, 1988: Journal of Petroleum Technology.

Doyen, P.M., Guidish, T., de Buyl, M.H., Lithology prediction from seismic data, a Monte-Carlo approach, 1988: S.E.G. expanded abstracts.

Doyen, P.M., Porosity from seismic data, a geostatistical approach, 1988: Geophysics, vol. 53, pp 1263-1275.

Guidish, T., de Buyl, M., Well calibration of seismically derived petrophysical parameters: a sensitivity analysis, 1987: S.E.G. expanded abstracts.

Johnson, R.A., Wichern, D.W., Applied multivariate statistical analysis, Prentice-Hall, Inc.

Journel, A.G., Alabert, F., Focusing on spatial

connectivity of extreme valued attributes: stochastic indicator models of reservoir heterogeneities, 1988: S.P.E. 18324.

O' Connor, R.B., Ingraham, S.E., Chang, B., M.D. MVSA methods in the stratigraphic interpretation of seismic data, 1987: S.E.G. expanded abstracts.

Solano, M., Schneider, W.A., Rock properties estimation in the Vivian sand in northern Peru, 1987: S.E.G. expanded abstracts.

Thadani, S.G., Alabert, F., Journel, A.G., An integrated geostatistical pattern recognition technique for characterization of reservoir spatial variability, 1987: S.E.G. expanded abstracts.

Uzcategui, O., Del Pino, E., Seismic inversion for mapping porosity and thickness by using SVD and geostatistical methods, 1987: S.E.G. expanded abstracts.

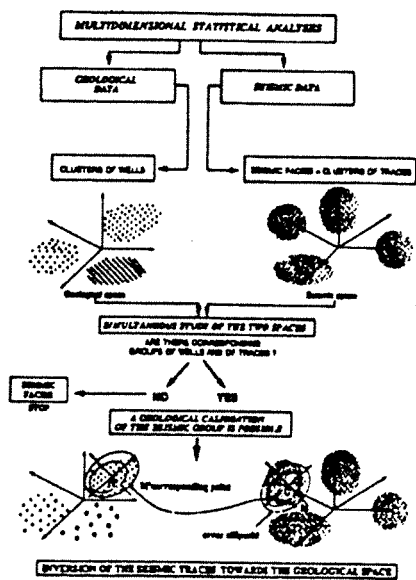


FIG. 1. Global flow chart of statistical methodology.

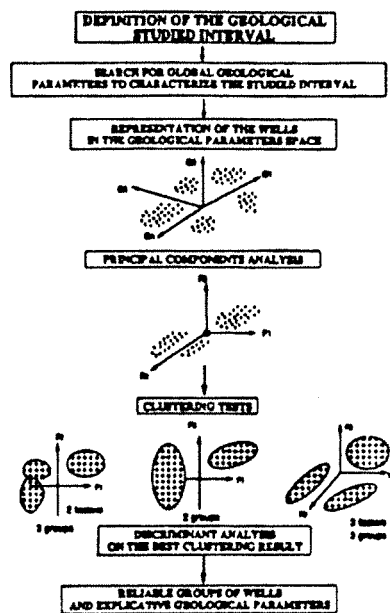


FIG. 2. Geologic data statistical processing.

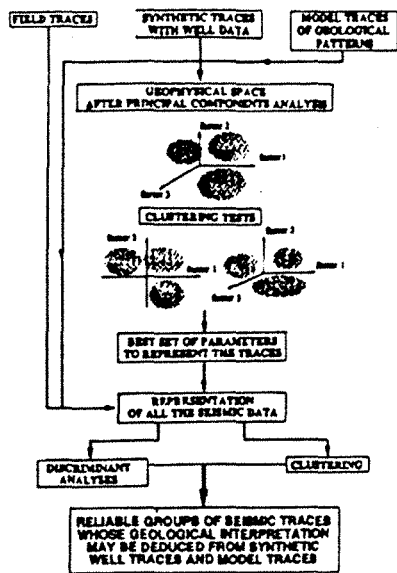


FIG. 3. Geophysical data statistical processing.

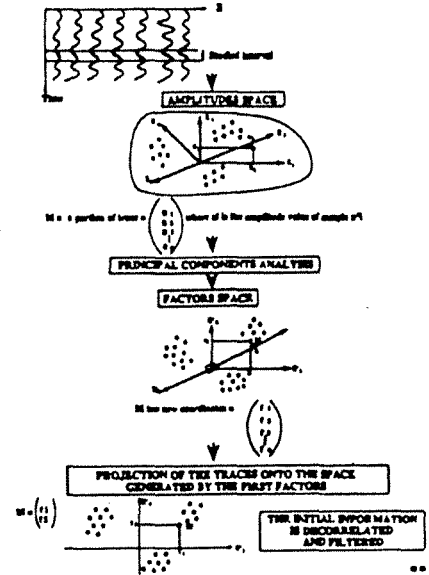


FIG. 4. Representation of traces in geophysical space.

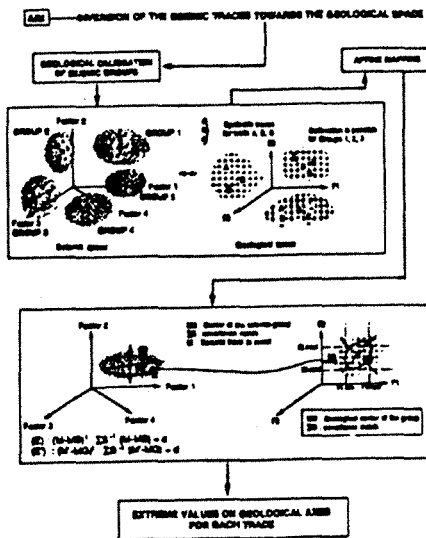


FIG. 5. Calibration and inversion of seismic traces.

Extraction of Quantitative Geologic Information from Seismic Data with Multidimensional Statistical Analyses.

Part II: A Case History

SI 2.5

Frederique Fournier, ELF-Aquitaine, France

SUMMARY

A statistical methodology to derive quantitative geological information from seismic data was applied to an oil producing field in a complex multilayered carbonate setting. The zone under analysis is of a size consistent with the resolution power of the seismic data and consists of three main units. 10 global geological parameters were defined from well information which took into account the lithologies, porosities and thicknesses of the studied interval. They generate a 10 multidimensional space in which the wells were represented and gathered in groups. Discriminant analyses have allowed to find sets of explicative parameters for each significant group. The seismic traces were then analysed on a time window corresponding to the reservoir zone. The field traces and synthetic traces computed from well logs were projected onto a multidimensional geophysical space derived from the trace amplitudes. Clusters of traces were found expressing seismic facies zones. Their geological meaning was corroborated by the assignment of the synthetic traces in the field trace groups. When it was possible, the seismic trace groups were calibrated by the corresponding well groups in the geological space. An affine mapping was then applied to derive, from the traces, bounds on the initial geological parameters. These bounds were taken into account in reservoir conditional simulations of the porosity distribution. They have improved greatly the spatial knowledge of the field.

INTRODUCTION

I have set out a statistical methodology to extract quantitative geological information from seismic data. This method is particularly adapted for complex environments. In this paper, I present an application to a multi-bed reservoir in a carbonate setting. The results will be described step by step.

GEOLOGICAL BACKGROUND

The studied reservoir A is typically multilayered (Figure 1), it is composed of three main independent units A1, A2, A3, which produce oil. Unit A3 is the most productive (50 % of the total production). The lower part of the reservoir corresponds to high energy littoral deposits while the upper part corresponds to lower energy intertidal deposits.

Reservoirs are sandstones with the best porosities, dolomitic sandstones and vacuolar dolomites. The seals are dolomitic, green shales and anhydrite.

The origin of the structure is halocinetic. A major synsedimentary fault plays a predominant role in the facies distribution. The best reservoirs are found on the top of the structure and in the North-West part of the field.

Total thickness of unit A (A1 + A2 + A3) is about 60 M in the high part of the field (West and North-West of the fault) but increases up to 110 M in the lower part.

DEFINITION OF THE STUDIED GEOLOGICAL INTERVAL

Reservoir engineers are mainly interested in the characteristics of unit A3, the most productive, which also exhibits the greatest variations of porosity. A3 thickness is around 30 M, except for the lower part of the field. In this carbonate setting, the velocities of the different lithofacies are very high (4000 m/s), the time window corresponding to A3 is quite small (15 ms). The available seismic data has a central spectrum frequency near 25 - 30 Hz which makes it impossible to pick the top of unit A3. I had therefore to consider reservoir A as a whole.

STUDY OF THE GEOLOGICAL DATA

The geological data are from well measures. The thicknesses of the three units, their effective thicknesses, porosities and lithofacies were available for every log measure. Logs analysts have defined 23 different lithofacies in the studied interval.

The methodology used represents the wells in a multidimensional geological space generated by parameters that globally characterize the studied zone. I had therefore to derive global properties from the well data. After studying impedance responses for the 23 lithofacies, while taking also into account the number of occurrences for each lithofacies, I have decided to simplify the lithofacies into 5 main lithofacies: dolomitic, argillaceous seals, calcareous dolomites, dolomitic reservoirs, sandstone reservoirs. I have also computed an average porosity for each unit and have totalized the thicknesses and effective thicknesses for interval A. This means I had 10 global parameters to characterize unit A, taking into account thicknesses, lithologies and porosities.

The wells were represented in a 10 dimensional space. Principal components analysis (P.C.A.) was then carried out. Wells were projected in the 3 dimensional first factors space (Figure 2) which explained 69.6 % of the initial information. The gathering of wells on the principal planes (Figure 2) reveals the main internal structures of the data.

Clustering tests were then carried out. Different numbers of factors were used to summarize the information and to test different numbers of clusters. The best result, as far as separation between groups is concerned, corresponds to an analysis with 5 clusters and 3 factors (Figure 3a). With this result, I have used a discriminant analysis to find out the characteristics of each group. The discriminant analysis has shown that the groups were strongly individualised. The "top" group (n°3, Figure 3) is a pattern of very porous sandstone reservoirs associated with calcareous dolomites, the effective thicknesses are not very high. The North group (n°2) shows high effective thicknesses and dolomitic reservoirs with less interesting porosities than in the top zone. For the South-West zone (n°4), the reservoirs are dolomitic, the porosities are intermediary. The South-East zone (n°1) presents typically high thicknesses, dolomitic reservoirs and many shaly seals.

For the most typical groups (North, Top, South-East), I have drawn a portrait of interval A in which the thicknesses, lithologies and the impedances resulted from averaging on the wells belonging to the group.

ANALYSIS OF THE SEISMIC DATA

Synthetic traces were computed from Sonic and Density logs at all the wells. The reflections corresponding to the top and bottom of A were then picked on these synthetic traces. The time interval corresponding to A can vary from 25 ms to 38 ms, but in most cases, it varies between 28 ms and 35 ms.

To characterize the traces on such short windows, I represent them in the factors reduced space. This follows a principal components analysis (P.C.A.) that is carried out on the series of trace amplitudes. A filtering effect is produced by the projection of the traces on the space of lesser dimension generated by the first P.C.A. factors. I then must choose two parameters, the size of the time window and the number of factors to keep.

The synthetic traces were studied with windows of 28 ms and 35 ms. In both cases, clustering tests were carried out for different numbers of P.C.A. factors (Figure 3b). The results from all the tests have pointed out that the window size (between 28 and 35 ms) had little significance on the clusters we found. The first 3 factors (85.6 % of the total initial information) were kept to generate the seismic representation space.

The gathering of the synthetic trace population in 5 groups (Figure 3b) is interesting because the synthetic trace clusters in the geophysical space correspond to the well clusters in the geological space. The "top" synthetic traces form a group as traces of the South-East zone and of the North zone do. I have computed the seismic response of the geological patterns representing the main groups of wells and classified them with the synthetic traces. They are correctly assigned to the previous groups and confirm the geological meaning of the synthetic trace groups.

With all the calibrations made between synthetic traces and well data, I was ready to analyse field seismic data.

The field seismic data come from a 2D survey shot in 1980. The resolution was not very high but a complete reprocessing greatly improved the data quality at the level of reservoir A. A special attention was paid to amplitude recovery, velocity analyses and stratigraphic deconvolution. Reprocessing enhanced the match with most of the synthetic traces and allowed to pick very precisely the event corresponding to the top of reservoir A. A 32 ms window (9 samples) was extracted and all the field traces (15000) were represented in the 3 dimensional P.C.A. factors space. The filtering effect is very similar to the one on synthetic traces (85.6 % of the initial information is taken into account). Clustering of traces were carried out for 5 and 6 clusters and maps of seismic facies were drawn.

The best result (5 clusters) was used in a discriminant analysis. The traces nearest the centers of the groups, in which they were assigned by the classification algorithm, were used as learning traces in

the discriminant stage. The descriptive step of the analysis computed discriminant axes to separate the a priori groups of learning traces. The 5 groups are highly individualized. The predictive step of the analysis has allowed the assignment of the field traces which did not belong to the learning traces. It has also allowed the classification of the synthetic traces at the wells after they had been homogenized for amplitudes. A seismic facies map was drawn (Figure 4). The seismic zones are similar to those obtained in the classification step, as far as the major trends are concerned, but this map is more reliable because I have carried out a learning stage in the discriminant analysis.

On the 5 groups, three correspond to groups already found with synthetic traces. This is corroborated by the assignment of the synthetic traces in field trace groups.

CALIBRATION OF SEISMIC GROUPS AND INVERSION OF THE TRACES

I then studied the correspondances of gatherings between the geological space and the geophysical space. For three seismic groups, there is an excellent fit between the synthetic and field traces group and the wells group. These groups correspond to the "top" pattern, the South-East facies and the North pattern. The seismic groups were calibrated by the corresponding groups of wells.

The fourth group corresponds pro parte to a group of wells. The calibration here is possible but will be less reliable. Group 5 does not fit to any cluster in the geological space and therefore will not be calibrated. Its traces will not be inverted towards the geological space.

The inversion - an affine mapping - gave for each trace extreme values (minima and maxima) on the factor axes generating the geological space. These were converted into bounds on the 10 initial geological features. At each trace location, provided that it belongs to a group for which calibration is possible, I have derived minimum and maximum values for total thickness, effective thickness, average porosities for the three units A1, A2, A3 (Figure 5) and percentages of the main lithofacies.

USING SOFT SEISMIC DERIVED INFORMATION IN RESERVOIR MODELS

I focus now on the average porosity of unit A3, the variable PHI3. (Figure 5). Conditional simulations of PHI3 distribution were made. These simulations help to visualize the potential variability of PHI3 in the field. Instead of having a unique map, which represents an average situation, I may generate several pictures for the field which honor all the initial data and their spatial structure. Simulations are a helpful tool for evaluating risks before taking decisions.

20 simulations of PHI3 were generated, conditioned only by the values of PHI3 available at the wells. The variability they exhibit is very important (Figure 6). This demonstrates that well data are not sufficient to give a good knowledge of the spatial

structure of PHI3. PHI3 is therefore not well constrained, except near the wells.

20 other simulations were generated, conditioned both by well data and by the bounds I derived from the seismic data (Figure 6). There is a drastic diminution of the variability of PHI3. Extra information, even fuzzy, greatly improves the knowledge of the PHI3 spatial structure.

CONCLUSION

I have applied the previously described statistical method in a complex environment. It was a multilayered reservoir composed of interfering lithologies where thickness and porosity varied. The velocities were very high, so the time interval corresponding to the reservoirs were very short. The available seismic data were not very broad-banded. However, the statistical method, which processed completely all the available information, provided bounds on global geological features which were of great help to constrain the reservoir models because of the high spatial density of the information.

ACKNOWLEDGMENTS

I wish to thank André Journel and François Alabert from Stanford University who warmly welcomed me last December. André gave me interesting suggestions about the future of the methodology and François has realized all the conditional simulations. In addition, I would like to thank Elf Aquitaine for permission to publish this paper.

REFERENCES

Alabert, F., Stochastic imaging of spatial distributions using hard and soft information, 1987: MSc thesis, Stanford University, 197 p.

Anderberg, M.R., Cluster analysis for applications, 1973: Probability and mathematical statistics, Academic Press.

Johnson, R.A., Wichern, D.W., Applied multivariate statistical analysis, Prentice-Hall, Inc.

Rao, C.R., Linear statistical inference and its applications, 1965, John Wiley & Sons, Inc.

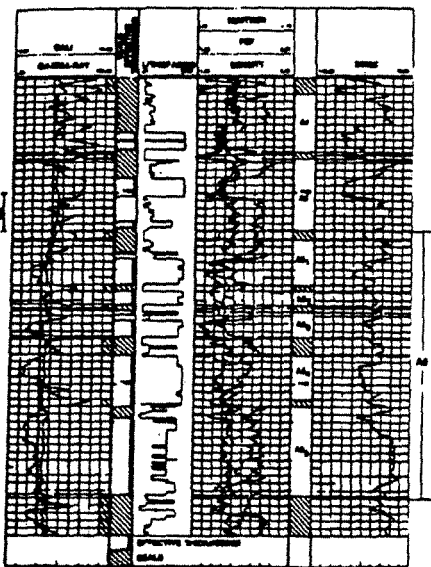


FIG. 1. Geologic description of reservoir A.

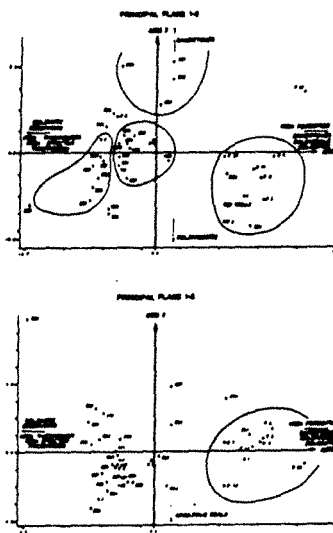


FIG. 2. Principal components analysis of geologic data.

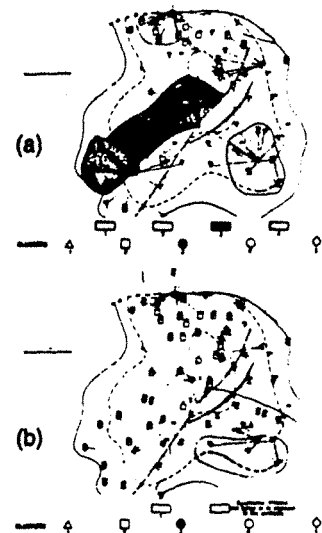


FIG. 3. Clustering tests in geologic and geophysical spaces. (a) 5 clusters of wells found in PCA factors space, and (b) clusters of synthetic traces with three PCA factors.

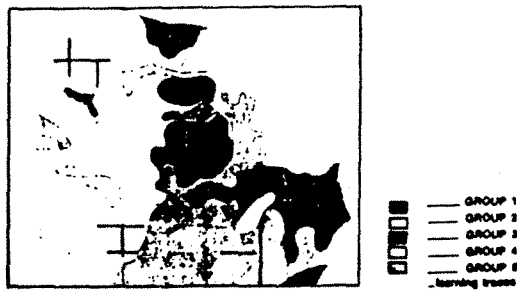


FIG. 4. Seismic facies map obtained with discriminant analysis.

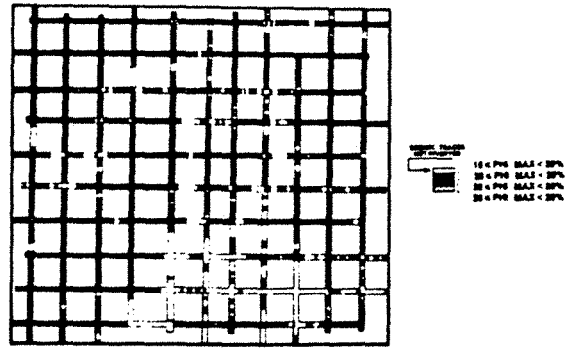


FIG. 5. PHI3 maximum value after seismic traces inversion.

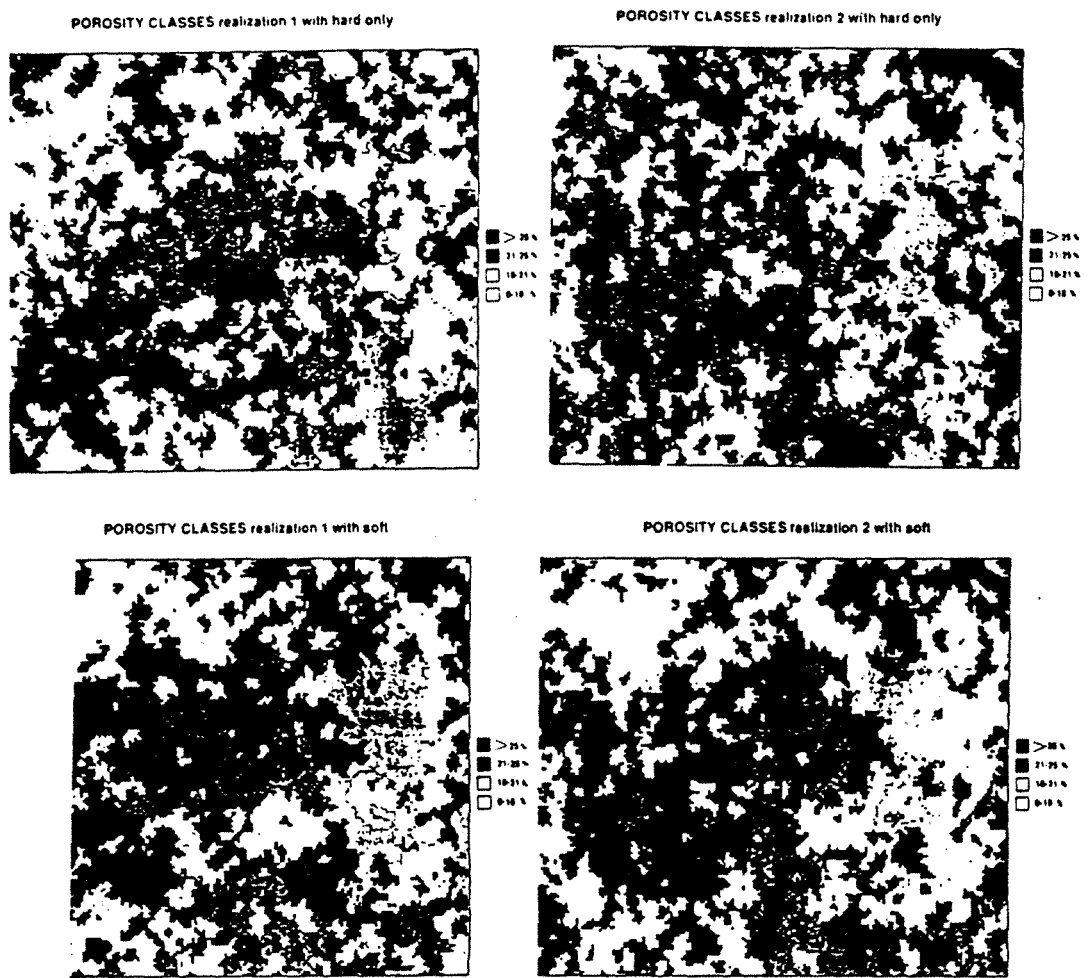


FIG. 6. Conditional simulations of PHI3 with hard data only then with hard and soft seismic derived data.

REFERENCES BIBLIOGRAPHIQUES

Akaike, H., 1970, Statistical predictor identification, *Ann. Statist. Math.*, v. 22, p. 203-217.

Akaike, H., 1976, Canonical correlation analysis of time series and the use of an information criterion, *System identification - advances and case studies*, R. R. Mehra and D.G. Lainiotis, editors, Academic Press, p. 27 - 96.

Aminzadeh, F., Chatterjee, S., 1984, Applications of clustering in exploration seismology, *Geoexploration*, 23, p. 147-159.

Anderberg, M.R., 1973, *Cluster analysis for applications*, Academic Press.

Andersen, N., 1974, On the calculation of filter coefficients for maximum entropy spectral analysis, *Geophysics*, 39, n° 1, p. 69-72.

Azorin-Lara, E., 1985, La sélection de variables en discrimination, mémoire de DEA, Université de Paris IX Dauphine.

Azorin-Lara, E., Celeux, G., 1986, Reconnaissance de faciès géologiques à partir de données sismiques. Utilisation de techniques statistiques, rapport INRIA sous contrat SNEA(P).

Barrodale, I., Erickson, R.E., 1980, Algorithms for least-squares linear prediction and maximum spectral analysis, Part I : theory, Part II : fortran programs, *Geophysics*, 45, n°3, p. 420-446.

Barrodale, I. et al., 1983, Computational experience with Marple's algorithm for autoregressive spectrum analysis, *Geophysics*, 48, n° 9, p. 1274-1286.

Benzecri, J.P., et collaborateurs, 1980, *L'analyse des données*, Dunod, Paris.

Berryman, J.G., 1978, Choice of operator length for entropy spectral analysis, *Geophysics*, 43, n° 7, p. 1384-1391.

Black, P. J., Boashash, B., 1987, An efficient real-time implementation of the Wigner-Ville distribution, *IEEE Trans.*, vol. ASSP-35, n° 11.

Bois, P., 1980, Autoregressive pattern recognition applied to the delimitation of oil and gas reservoirs, *Geophysical Prospecting*, 28, p. 572-591.

Bois, P., 1981, Prédiction linéaire et principe de l'entropie maximale, *Revue de l'Institut Français du Pétrole*, vol. 36, n° 5, p. 607-628, vol. 36, n° 6, p. 775-792.

Bois, P., 1981, Determination of the nature of reservoirs by use of pattern recognition algorithm with prior learning, *Geophysical Prospecting*, 1981, 29, p. 687-701.

Bois, P., 1982, Some comments on the application of pattern recognition to oil and gas exploration, *Geoexploration*, 20, p. 147-159.

Bois, C., 1989, Imagerie de la croûte terrestre par sismique réflexion verticales profonde : le programme ECORS, *Comptes Rendus de l'Académie des Sciences*, tome 6, n° 3, Gauthier-Villars.

Bouachache, B., 1982, Representation temps-fréquence, application à la mesure de l'absorption du sous-sol, thèse de l'I.N.P. de Grenoble.

Bouachache, B., 1983, de Bazelaire, E., 9th Colloque Grets, proceed. p.879-884.

Boashash, B., (ex Bouachache), de Bazelaire, E., 1985, On the Wigner time-frequency distribution, 10th Colloque Grets, proceed. p. 71-76.

Boashash, B., 1987, Theory, implementation and application of time-frequency signal analysis using the Wigner-Ville distribution, *Journal of Electrical and Electronics Engineering, Australia, IE Aust., and IREE Aust.*, vol. 7, n° 3.

Bouroche, J.M., Saporta, G., *L'analyse des données*, Collection Que sais-je ?, PUF.

Box, G.E.P., Jenkins, G.M., 1970, *Time series analysis, forecasting and control*, Holden-Day, San Francisco.

Brown, L. F., Fisher, W. L., 1977, Seismic stratigraphic interpretation of depositional systems : examples from Brazil rifts and pull-apart basins in Payton C. E. ed., "Seismic Stratigraphy - Application to Hydrocarbon Exploration", A.A.P.G. Memoir 26, p. 213-248.

Bruel, L., 1987, Application de la représentation conjointe temps-frequence de Wigner-Ville à des traces sismiques, rapport interne SNEA(P).

De Buyl, M.H., Guidish, T., Bell, F., 1988, Reservoir description from seismic lithologic parameter estimation, Journal of Petroleum Technology.

Burg, J.P., 1967, Maximum entropy spectral analysis, 37th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded abstracts.

Burg, J.P., 1972, The relationship between maximum entropy spectra and maximum likelihood spectra, Geophysics, 37, p. 375-376.

Burg, J.P., 1975, Maximum entropy spectral analysis, P.h.D. Thesis, Stanford University.

Celeux, G., Lechevallier, Y., 1981, Methodes de segmentation non paramétrique, RSA, vol. 30, n° 4.

Celeux, G, 1984, Approximation rapide et interprétation d'une partition centrale pour les algorithmes de classification, Rapport INRIA n° 301.

Chambre syndicale de la recherche et de la production du pétrole et du gaz naturel, Corps sédimentaires : exemples sismiques et diagraphiques, éditions Technip, Paris et Institut Français du Pétrole, Rueil-Malmaison.

Chapel, P., 1980, Geophysique appliquée, dictionnaire et plan d'étude, Masson, Paris.

Chen, C.H., 1987, Digital waveform processing and recognition, CRC Press, Boca Raton, Florida.

Chen, P.C., Galloway, R.L., Fukunaga, K., 1986, Seismic character classification based on a line-segment model, 56th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, PR/AI 1.5, p. 193-197.

Claerbout, J., 1976, Fundamentals of geophysical data processing, Mc Graw Hill, New York.

Conticini, F., 1984, Seismic stratigraphic approach to complex fluvial deposits, Seismic Research and Data Processing department, AGIP.

Conticini, F., 1984, Seismic facies quantitative analysis : new tool in stratigraphic interpretation, 54th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p.680-682.

Cooley, W.W., Lohnes, P.R., 1971, Multivariate data analysis, John Wiley and Sons, New-York.

Cordier, J.P., 1983, Les vitesses en sismique réflexion, Technique et Documentation, Lavoisier, Paris.

Davis, J.C., 1986, Statistics and data analysis in geology, second edition, John Wiley and Sons, New-York.

Devijver, P.A., Kittler, J., 1982, Pattern recognition, a statistical approach, Prentice Hall International, Inc., London.

Diday, E., et collaborateurs, 1980, Optimisation en classification automatique, rapport INRIA.

Dix, C.H., 1955, Seismic velocity from surface measurements, Geophysics, 20, p. 68-86.

Doyen, P.M., Guidish, T.M., de Buyl, M.H., 1988, Lithology prediction from seismic data, a Monte-Carlo approach, 57th Ann. Inter. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 873-876.

Doyen, P.M., 1988, Porosity from seismic data : a geostatistical approach, Geophysics, 53, p. 1263-1275.

Dumay, J., Kenaan, T.A., 1981, Interpretation stratigraphique des données sismiques de l'ensemble Bathonien-Callovien Inférieur dans le Bassin de Paris à l'aide de l'inversion des traces sismiques, thèse soutenue à l'Université Louis Pasteur de Strasbourg.

Dumay, J., Fournier, F., 1985, Caractérisation d'un environnement géologique par l'analyse statistique multidimensionnelle de paramètres sismiques, rapport interne SNEA (P).

Dumay, J., Fournier, F., 1986, Characterization of a geologic environment by multivariate statistical analyses of seismic parameters, 56th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, PR/AI 1-6, p. 197-200.

Dumay, J., Fournier, F., 1988, Multivariate statistical analyses applied to seismic facies recognition, *Geophysics*, 53, n° 9, p. 1151-1159.

Efron, B., 1982, The Jackknife, the bootstrap and other resampling plans, *Siam* n° 38.

Escofier, B., Pagès, J., 1988, *Analyses factorielles simples et multiples, objectifs, méthodes et interprétation*, Dunod, Paris.

Escudie, B., 1979, représentation en temps et fréquence des signaux d'énergie finie : analyse et observations des signaux, *Ann. Telecom.*, 34, n° 3-4.

Fournier, F., 1985, Analyse du spectre de puissance théorique obtenu à partir d'un modèle de séquence sédimentaire, rapport interne SNEA (P).

Fournier, F., 1986, Etude des faciès sismiques sur la structure de Torpille Nord-Est par analyses statistiques multidimensionnelles, rapport interne SNEA(P).

Fournier, F., Urruty, J.M., 1987, Point sur les travaux menés sur le champ d'Anguille Marine dans le cadre du projet de recherche Alias, rapport interne SNEA(P).

Fournier, F., 1988, Etude des variations lithologiques du Callovien Inférieur de Soudron au moyen de l'outil sismique, rapport interne SNEA(P).

Fournier, F., Ma, Y.Z., 1988, Analyse spectrale par maximisation de l'entropie : application aux courtes fenêtres, rapport interne SNEA(P).

Fournier, F., 1988, Adaptation de la méthodologie Alias aux "courtes fenêtres", application au permis d'Evry, rapport interne SNEA(P).

Fournier, F., 1989, Extraction d'information lithologique par statistiques multidimensionnelles sur des données sismiques, application au champ de Yanga (Congo), rapport interne SNEA(P).

Fournier, F., 1989, Extraction of quantitative geologic information from seismic data with multidimensional statistical analyses, Part 1 : methodology, Part 2 : A case history, 59th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 726-733.

Furniwal, G., Wilson, W., 1974, Regression by leaps and bounds, *Technometrics*, vol. 16, n° 14.

Godfrey, R., Muir, F., Rocca, F., 1980, Modeling seismic impedance with Markov chains, *Geophysics*, 45, n° 9, p. 1351-1372.

Hagen, D., C., 1982, The application of principal components analysis to seismic data sets, *Geoexploration* 20, p. 93-111.

Handbook of Geophysical Exploration, Pattern recognition and image processing, 1987, Helbig and Treitel, editors, Geophysical Press.

Hardenbol, J., VAIL, P.R., Ferrer, J., 1981, Interpreting paleoenvironments, subsidence history and sealevel changes of passive margins from seismic and biostratigraphy, *Oceanologica Acta*, Paris n° sp, pp. 3-44.

Harms, J.C., Tackenberg, P.L., 1972, Seismic signatures of sedimentation models, *Geophysics*, 37, n° 1, p. 45-58.

Helbig, K., Treitel, S., 1983-1985, Seismic exploration, Handbook of geophysical exploration, Geophysical Press, London.

Hileman, J.A., Embree, P., Pfleuger, J. C., 1968, Automated static corrections, *Geophysical Prospecting*, 16, p. 326-358.

Houston Geological Society, 1975, Deltas : models for exploration, editor M. Lou Broussard.

Huang, K.Y., Fu, K S., 1985, Syntactic pattern recognition for the classification of Ricker wavelets, *Geophysics*, vol. 50, n° 10, p. 1548-1555.

Huang, K.Y., Fu, K.S., Cheng, S.W., Lin, Z.S., 1987, Syntactic pattern recognition and Hough transformation for reconstruction of seismic patterns, *Geophysics*, vol. 52, n° 12, p. 1612-1620.

Huang, K.Y., Sheen, T.H., 1986, A tree automaton system of syntactic pattern recognition for the recognition of seismic patterns, 56th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, PR/AI 1.2, p. 183-187.

Jenkins, G.M., Watts, D.G., 1968, Spectral analysis and its applications, Holden-Day, San Francisco.

Johnson, R.A., Wichern, D.W., Applied multivariate statistical analysis, Prentice-Hall.

Journel, A.G., Alabert, F., 1988, Focusing on spatial connectivity of extreme-valued attributes : stochastic indicator models of reservoir heterogeneities, SPE 18324.

Kalouptisidis, N., Theodorisidis, S., 1987, Fast adaptative least squares algorithms for power spectral estimation, IEEE Trans, ASSP. 35, N° 5, p. 661-670.

Kane, R.P., Trivedi, N.B., 1982, Comparison of maximum entropy spectral analysis and least-squares linear prediction methods for some artificial samples, Geophysics, 47, n° 12, p. 1731-1736.

Khattri, K., Gir, R., 1976, A study of the seismic signatures of sedimentation models using synthetic seismograms, Geophysical Prospecting 24, p. 454-477.

Khattri, K., Sinvhal, A., Awasthi, A.K., 1979, Seismic discrimination of stratigraphy derived from Monte-Carlo simulation of sedimentary formations, Geophysical Prospecting 27, p. 168-195.

Kshirsagar, A.M., 1972, Multivariate analysis, Marcel Dekker, Inc., New-York.

Lachenbruch, P., A., 1968, On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficient, Biometrics, p. 823-834.

Lavergne, M., 1986, Méthodes sismiques, Editions Technip, Paris et Institut Français du Pétrole, Rueil-Malmaison.

Lebart, L., Morineau, A., Tabard, N., 1977, Techniques de la description statistique, Méthodes et logiciels pour l'analyse des grands tableaux, Dunod, Paris.

Lerman, I.C., Classification et analyse ordinaire des données, Dunod, Paris.

Levinson, N., 1947, The Wiener RMS (root mean square) error criterion in filter design and prediction, Journal of Mathematics and Physics, Vol. XXV, n° 4, p. 261-278.

Liu, H.H., Fu, K.S., 1982, A syntactic pattern recognition approach to seismic discrimination, Geoplotation, 20, p. 183-196.

Louis, P.R., 1976, Essai de synthèse sédimentologique des séries deltaïques du Nigéria, rapport interne SNEA(P).

Louis, P.R., 1976, Recherche des principaux caractères lithologiques et pétroliers des séries deltaïques du Nigeria, rapport interne SNEA(P).

Louis, P., Lacaze, J., Sulpice, 1978, Un delta subsident à travers quelques techniques d'exploration, rapport interne SNEA(P).

Luenberger, D.G., 1969, Optimization by vector space methods, John Wiley and Sons, New-York.

Ma, Y.Z., 1987, Filtrage géostatistique des images numériques, Thèse doctorale de l'INPL, Nancy.

Ma, Y.Z., 1988, Compléments aux essais de l'analyse spectrale par la méthode d'entropie maximale : détermination de l'ordre autorégressif et analyse spectrale par le krigeage, rapport interne SNEA(P).

Ma, Y.Z., Royer, J.J., 1988, Le comportement des pondérateurs en analyse krigeante : comparaison avec les filtres classiques utilisés en traitement d'images, Sciences de la terre, n° 28, p. 175-194, Nancy.

Marple, L., 1980, A new autoregressive spectrum analysis algorithm, IEEE, Trans. ASSP. 28, n° 4, p. 441-454.

Marple, L., 1982, Frequency resolution of Fourier and maximum entropy spectral estimates, Geophysics, 47, n° 9, p. 1303-1307.

Mathieu, P.G., Rice, G.W., 1969, Multivariate analysis used in the detection of stratigraphic anomalies from seismic data, Geophysics, 34, n° 4, p. 507-515.

Matlock, R.J., Mc Gowen, R.S., Asimakopoulos, G.T., 1985, Can seismic stratigraphy problems be solved using automated pattern analysis and recognition, 55th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 582-585.

Mc Quillin, R., Bacon, M., Barclay, W., 1979, Introduction à l'interprétation sismique, Graham et Trotman, Londres, traduction J.R. Viallix.

Mitchum, J.R., 1977, Glossary of terms used in seismic stratigraphy, A.A.P.G. Memoir 26, p. 205-212

Mitchum, J.R., Vail, P.R., Thompson, S., 1977, The depositional sequence as a basic unit for stratigraphic analysis, A.A.P.G. Memoir 26, p.53-62.

Modulad, Méthodes de classification pour l'analyse des données, support de cours, INRIA.

Morf, M., et al., 1977, Efficient solution of covariance equations for linear prediction, IEEE, Trans. ASSP. 25, n° 5, p. 429-433.

Nelson, H.R., 1983, New technologies in exploration geophysics, Gulf Publishing Co., Houston.

O'Connor, R.B., Ingraham, S.E., Chang, B., 1987, M.D. MVSA methods in the stratigraphic interpretation of seismic data, 57th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 805-809..

O'Doherty, R.F., Anstey, N.A., 1971, Reflections on amplitudes, Geophysical Prospecting 19, p. 430-458.

Paget, C., 1986, Un essai d'utilisation de la représentation temps-fréquence à la caractérisation des structures fines sismiques, rapport interne SNEA(P).

Perrodon, A., 1980, Geodynamique pétrolière, Masson, Paris.

Posamentier, H.W., Jervey, M.T., Vail, P.R., 1988, Eustatic controls on clastic deposition I, Conceptual framework in sea-level changes : an integrated approach, Tulsa, Oklahoma, SEPM Sp. Publ., n° 42, p. 109-124.

Posamentier, H.W., Vail P.R., 1988, Eustatic controls on clastic deposition II, Sequences and system tract model, in sea-level changes : an integrated approach, Tulsa, Oklahoma, SEPM Sp. Publ., n° 42.

Rafipour, B.J., 1987, Seismic response for reservoir fluid evaluation, 57th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 377-380.

Rao, C.R., 1965, Linear statistical inference and its applications, John Wiley and Sons, New-York.

Robinson, E.A., Treitel, S., 1980, Geophysical signal analysis, Prentice-Hall, Inc.

Romedor, J.M., 1973, Méthodes et programmes d'analyse discriminante, Dunod, Paris.

Rovelli, A., 1982, On the information time scale of geophysical records in maximum entropy spectral analysis, IEEE, Trans. GE-20, n° 2, p. 158-161.

Sangree, J.B., Vail., P.R., Sneider, R.M., 1988, Evolution of facies interpretation of the shelf-slope : application of the new eustatic framework to the Gulf of Mexico, OTC 5695, Houston.

Shaw, S.W., de Figueiredo, R.J.P., 1986, Artificial intelligence technique for automating seismic stratigraphic interpretation, 56th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, PR/AI 1.3, p. 187-190.

Sheriff, R.E., 1973, Encyclopedic dictionary of geophysics, Society of Exploration Geophysicists, Tulsa.

Sheriff, R.E., 1976, Inferring stratigraphy from seismic data, the American Association of Petroleum Geologists Bulletin, v. 60, n° 4, p. 528-542.

Sheriff, R.E., 1980, Seismic stratigraphy, IHRDC, Boston.

Sheriff, R.E., Geldart, L.P., 1982, Exploration seismology, Cambridge University Press, Cambridge.

Sinvhal, A., Khattri, K., 1983, Application of seismic reflection data to discriminate subsurface lithostratigraphy, Geophysics, 48, n° 11, p. 1498-1513.

Sinvhal, A., Khattri, K., Sinvhal, H., Awasthi, A.K., 1984, Seismic indicators of stratigraphy, Geophysics, 49, n° 8, p. 1196-1212.

Society of Economic Paleontologists and Mineralogists, 1970, Deltaic sedimentation modern and ancient, Special publication n° 15, editor J.P. Morgen with the assistance of R.H. Shaver.

Taner, M. T., Koehler, F., Alhilali, K. A., 1974, Estimation and correction of near-surface time anomalies, *Geophysics*, 41, p. 441-463.

Telford, W.M., Geldart, L.P., Sheriff, R.E., Keys, D.A., 1976, *Applied geophysics*, Cambridge University Press, Cambridge.

Thadani, S.G., Alabert F., Journel, A.G., 1987, An integrated geostatistical pattern recognition technique for characterization of reservoir spatial variability, 57th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, p. 372-375

Titterton and al., 1981, Comparison of discrimination techniques applied to a complex data set, J.R.S.S.A.

Ulrych, T.J., 1972, Maximum entropy power spectrum of truncated sinusoids, *Journal of Geophysical Research*, 77, p. 1396-1400.

Ulrych, T.J., Bishop, T.N., 1975, Maximum entropy spectral analysis and autoregressive decomposition, *Reviews of Geophysics and Space Physics*, vol. 13, p. 183-200.

Vail, P.R., Sangree, J.B., Mitchum, R.M., Frazier, D.E., Tood, R.G., 1977, Stratigraphic interpretation of seismic data, Mémoire AAPG n° 26 (cours EXXON).

Vail, P.R., 1987, Seismic stratigraphy interpretation using sequence stratigraphy, Part 1 : Seismic Stratigraphy interpretation procedure, in A.W. Bally eds., *Atlas of Seismic Stratigraphy*, v. 1, AAPG Studies in Geology § 27.

Van Wagoner, J.C., Posamentier H.W., Mitchum, R.M., Vail, P.R., Sarg, J.F., Loutit, T.S., Hardenbol, J., 1988, An overview of the fundamentals of sequence stratigraphy and key definitions, in sea-level changes : an integrated approach, SEPM Sp. Publ. n° 42, p. 39-45.

Velzeboer, C.J., 1981, The theoretical seismic reflection response of sedimentary sequences, *Geophysics*, 46, n° 6, p. 843-853.

Ville, J., 1948, *Theorie et application de la notion de signal analytique*, SSM 3507, Câbles et Transmissions, Paris.

Walden, A.T., Hosken, J.W.J., 1985, An investigation of the spectral properties of primary reflection coefficients, *Geophysical Prospecting*, 33, 400-435.

Waters, K.H., 1978, Reflection seismology, John Wiley and Sons, New-York.

Wax, M., 1988, Order selection for AR models by predictive least squares, IEEE, Trans. ASSP. 36, n° 4, p. 581-588.

Wenzel, F., Zurn, W., 1980, Numerical investigation of the spectral resolving power of Burg's maximum entropy method, Journal of Geophysical Research, 48, p. 121-123.

Wuenschel, P.C., 1960, Seismogram synthesis including multiples and transmission coefficients, Geophysics, 25, p. 106-129.

Wiggins, R. A., Lerner, K. L., Wisecup, R. D., 1976, Residual statics analysis as a general linear inverse problem, Geophysics, 41, p. 922-938.

Yilmaz, O., 1987, Seismic data processing, Society of Exploration Geophysicists, Investigation in geophysics n° 2, Stephen M. Doherty, editor.

Zhang, Z., Simaan, M., 1986, A knowledge-based system for automatic segmentation of a seismic section, 56th Ann. Intern. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Poster papers 2.21, p. 306-308.

