
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
École Nationale Polytechnique



Département de Génie Civil
Laboratoire L.G.S.D.S

PROJET DE FIN D'ETUDES

EN VUE DE L'OBTENTION DU DIPLOME
D'INGÉNIEUR D'ÉTAT EN GÉNIE CIVIL

Thème

Développement des lois prédictives pour
l'estimation des paramètres stochastiques
des mouvements sismiques forts

Proposé et dirigé par :

M. M. Hammoutene

Réalisé par :

M. BELEKESSAM Samir
M. DJOUABI Hamza

Promotion : JUIN 2006

*École Nationale Polytechnique 10, Avenue Hassen Badi BP182 El-Harrach
16200 Alger (Algérie)*

Tel :213 21 52 53 01/03 — Fax :213 21 52 29 73

ملخص:

استعملنا طريقة التحليل التراجعي لهدف تقدير الخصائص الاحتمالية للهزات الأرضية العنيفة. وذلك بالاعتماد على بنك بيانات عالمية تحتوي على تسجيلات لهزات أرضية عنيفة. أربع علاقات لتقدير الخصائص الطيفية، مدة الحركة العنيفة إضافة إلى شدة أرياس، باستعمال المتغيرات المستقلة (شدة الهزة، البعد عن مركز الهزة، نوعية الأرضية) تم استخراجها خلال هذا العمل. قصد توضيح صحة العلاقات قمنا بمحاكاة منحنيات التسارع لبعض الأصناف الموافقة لخصائص معينة.

المفاتيح:

الخصائص الاحتمال، الهزات الأرضية العنيفة، شدة أرياس، مدة الحركة العنيفة، شدة الهزة، البعد عن مركز الهزة، نوعية الأرضية، الخصائص الطيفية، التحليل التراجعي

Summary:

Multiple linear regression analysis is used to estimate stochastic characteristics of seismic strong motions.

Worldwide accelerograms data bank are developed to study the attenuation of spectral parameters, Arias intensity an strong motion duration as a function of magnitude, distance and soil site conditions.

For illustration purposes, simulation results in terms of response spectra and associated artificial accelerograms are also presented for selected of strong ground motion data corresponding to specific ranges of the independent variables.

Key words: Stochastic, strong motions, Arias intensity, strong motion duration, magnitude, distance, site conditions, spectral parameters, regression analysis

Résumé :

Une analyse par régression multilinéaire est utilisée en vue de l'estimation des caractéristiques stochastiques des mouvements sismiques fort.

Quatre lois prédictives des paramètres spectraux, de l'intensité d'Arias et de durée de mouvement fort sont développées en utilisant une banque de données accélérométrique; les variables indépendantes de régression utilisées étant la magnitude, la distance épacentrale et les conditions locales de site.

A titre illustratif, sont également présentés les résultats de simulation pour diverses classes de données correspondant à des choix spécifique de variables indépendantes.

Mots clés :

Stochastique, mouvement fort, Intensité d'Arias, durée de mouvement fort, la magnitude, distance épacentrale, conditions locales, régression

Remerciements.

Tout d'abord, nous remercions Allah, le clément et le miséricordieux de nous avoir donné la force et le courage de mener à bien ce modeste travail.

On exprime nos vives gratitudee à M. le Professeur TILIOUINE, pour les conseils éclairés qu'il nous a prodigué tout au long de notre travail, ainsi que pour l'honneur qu'il nous fait en présidant le jury de soutenance de notre projet de fin d'étude.

Le présent travail n'aurait pu être mis au point sans l'aide et l'orientation prodiguées par notre directeur de thèse, M. HAMMOUTENE, envers lequel nous adressons tous nos remerciements et notre profonde reconnaissance.

Nous tenons à souligner que leurs précieux conseils et recommandation ont toujours été prodigués dans une ambiance rigoureuse et studieuse qui n'a jamais manqué d'être amicale.

On tient à exprimer nos remerciements à tous les membres du jury.

Nos remerciements vont également à M. Y.ZELMAT, et M. A.BOUDINA pour leur aide sympathique qui nous a suivi le long de ce travail. Ainsi que, M. S. BOUCHAIB et nos chers amis F. AZZI et A. LARIBIE de Génie Industriel.

Notre pensée va à tous les enseignants qui ont contribué à notre formation depuis le primaire, ainsi qu'à tous les anonymes administratifs qui ont travaillé dans l'ombre pour nous soyons à ce stade de connaissance.

Je dédie ce modeste travail

A mes parents,
A mes frères et sœurs,
A mes amis.

Samir BELKESSAM

Je dédie ce modeste travail

A mes parents,
A mes frères et sœurs,
A mes amis.

Hamza DJOUABI

TABLES DES MATIÈRES

INTRODUCTION GÉNÉRALE.....	1
CHAPITRE I :	3
GÉNÉRALITÉS SUR LES SÉISMES.	3
1. INTRODUCTION.....	3
2. PROCESSUS DE TREMBLEMENT DE TERRE.....	3
3. MOUVEMENT FORT.	5
4. PROPAGATION DES ONDES SISMIQUES.	5
5. REPRÉSENTATION DU MOUVEMENT SISMIQUE.	7
6. CARACTÉRISTIQUES ENREGISTRÉS À LA SURFACE DU SOL.	7
7. PARAMÈTRES AFFECTANT LE MOUVEMENT SISMIQUE.....	8
7.1 FACTEURS LIÉS À LA SOURCE.....	9
7.2 FACTEURS LIÉS AU TRAJET PARCOURU.	11
7.3 FACTURES LIÉS AUX CONDITIONS LOCALES.....	12
8. SPECTRE DE RÉPONSE.	14
9. ASPECTS STOCHASTIQUES DES ACCÉLÉROGRAMMES DES SÉISMES A FORTS MOUVEMENTS.	15
CHAPITRE II :	15
ÉLÉMENTS D'ANALYSE DES SIGNAUX ALÉATOIRES	15
1. PROCESSUS STOCHASTIQUE.	15
1.1 DÉFINITION.....	15
1.2 DESCRIPTION PROBABILISTE DE PROCESSUS STOCHASTIQUE.	15
1.2.1 FONCTION DE DENSITÉ DE PROBABILITÉ.	15
1.2.2 FONCTIONS MOMENTS.....	16
1.3 PROCESSUS ALÉATOIRE STATIONNAIRE.	17
1.4 ERGODICITÉ.	18
1.5 PROPRIÉTÉS DES FONCTIONS D'AUTOCORRÉLATION.	18
1.6 PROPRIÉTÉ DES FONCTIONS DE COVARIANCE.	19
2. DÉCOMPOSITION SPECTRALE D'UN PROCESSUS ALÉATOIRE.	20
2.1 CAS DES PROCESSUS STATIONNAIRES.	20
2.2 CAS DES PROCESSUS NON STATIONNAIRES.	21
CHAPITRE III :	22
CARACTÉRISATION DES MOUVEMENTS FORTS DU SOL DUS AUX SÉISMES.	22
INTRODUCTION.	22
1. CARACTÉRISATION DANS LE DOMAINE TEMPORELLE	23
1.1 CARACTÈRE TRANSITOIRE.	23

1.2 DURÉE DE MOUVEMENT FORT.....	24
1.3 CONTENUS PHASIQUES ET ENVELOPPE.....	24
1.4 INTENSITÉ D'ARIAS.....	25
2. CARACTÉRISATION DANS LE DOMAINE FRÉQUENTIEL	27
2.1 REPRÉSENTATION SPECTRALE NON STATIONNAIRE.....	27
2.1.1 DENSITÉ SPECTRALE DE PUISSANCE INSTANTANÉE.....	29
2.1.2 PROCESSUS LOCALEMENT STATIONNAIRE.....	29
2.1.3 SPECTRE ÉVOLUTIF.....	30
 CHAPITRE IV :	 33
 L'APPROCHE DE SPECTRE PHYSIQUE.....	 33
 1. INTRODUCTION.....	 33
2. SPECTRE PHYSIQUE	33
3. PRÉSENTATION DES PARAMÈTRES DU MODÈLE (SPECTRE PHYSIQUE).....	35
 CHAPITRE V.....	 39
 NOTIONS SUR LA TECHNIQUE DE RÉGRESSION.....	 39
 INTRODUCTION.....	 39
1. ESTIMATION DES PARAMÈTRES INCONNUS D'UNE LOI DE RÉPARTITION	39
1.1 ESTIMATIONS DE L'ESPÉRANCE MATHÉMATIQUE ET DE LA VARIANCE.....	41
2. INTERVALLE DE CONFIANCE. PROBABILITÉ DE CONFIANCE	41
3. TEST D'HYPOTHÈSE	43
3.1 PRINCIPE DE TEST D'HYPOTHÈSE.....	43
3.2 CONCEPTS IMPORTANTS DANS L'ÉLABORATION D'UN TEST HYPOTHÈSE.....	43
3.3 COMMENT EXÉCUTER UN TEST D'HYPOTHÈSE : DÉMARCHE À SUIVRE.....	46
3.4 DÉMARCHE À SUIVRE DANS L'ÉLABORATION D'UN TEST D'HYPOTHÈSE.....	46
4. RÉGRESSION LINÉAIRE SIMPLE	47
4.1 INTRODUCTION.....	47
4.2 LE MODÈLE DE LA RÉGRESSION SIMPLE.....	47
4.3 CONDITION D'APPLICATION DU MODÈLE LINÉAIRE SIMPLE.....	48
4.4 ESTIMATION DE LA DROITE DE RÉGRESSION.....	48
5. RÉGRESSION LINÉAIRE MULTIPLE	49
5.1 INTRODUCTION.....	49
5.2 MODÈLE LINÉAIRE À PLUSIEURS VARIABLES EXPLICATIVES.....	49
5.3 HYPOTHÈSES FONDAMENTALES.....	50
5.4 FORMULATION D'UN MODÈLE DE RÉGRESSION MULTIPLE : IDENTIFICATION ET INTERPRÉTATION DES COMPOSANTES DU MODÈLE.....	51
5.5 DÉTERMINATION DE L'ÉQUATION DE RÉGRESSION : ESTIMATION DES PARAMÈTRES DU MODÈLE.....	51
5.6 ANALYSE DE VARIANCE EN RÉGRESSION MULTIPLE.....	53
5.7 TESTE D'HYPOTHÈSE EN RÉGRESSION MULTIPLE.....	56
 CHAPITRE VI :	 62
 ESTIMATION DES PARAMÈTRES SPECTRAUX.....	 62

1. INTRODUCTION.....	62
2. BANQUE DE DONNÉES	62
2.1 INTRODUCTION.....	62
2.2 ÉTUDE STATISTIQUE SUR LES ENREGISTREMENTS.....	63
2.3 RÉPARTITION DES DONNÉES.....	64
3. PRÉSENTATION DES LOIS PRÉDICTIVES.....	65
3.1 CHOIX DES VARIABLES (CRITÈRES STATISTIQUE).....	65
3.2 DESCRIPTION DES VARIABLES DÉPENDANTS ET INDÉPENDANTE	66
 CHAPITRE VII :	 70
 PRÉSENTATION ET DISCUSSION DES RÉSULTATS DE LA RÉGRESSION.....	 70
1. ANALYSE DE RÉGRESSION SUR LA FRÉQUENCE CENTRALE.	70
2. ANALYSE DE RÉGRESSION SUR LE RAPPORT DE LA BANDE FRÉQUENTIELLE ET LA FRÉQUENCE CENTRALE.	74
3. LA RÉGRESSION SUR L'INTENSITÉ D'ARIAS.	79
4. ANALYSE DE RÉGRESSION SUR LA DURÉE DE MOUVEMENT FORT.	84
5. ILLUSTRATION DES ACCÉLÉROGRAMMES SIMULÉ ET LES SPECTRES DE RÉPONSE :.....	89
CONCLUSION	92
RÉFÉRENCES.....	91

Liste des figures.

Figure 1 : Vocabulaire et éléments caractéristique en géosismique.

Figure 2 : Les ondes sismiques de volume et de surface.

Figure 3 : Accélérogramme

Figure 4 : Schématisation d'un plan de faille.

Figure 5 : Spectres moyens d'accélération pour différentes conditions de site selon Seed (1976)

Figure 6 : Représentation d'un oscillateur simple soumis à une excitation sismique.

Figure 7 : Caractérisation de l'enregistrement de Loma-Prieta dans le domaine temporel

Figure 8 : Estimation de la puissance spectrale d'un processus au départ du carré de la réponse d'un filtre en bande étroite.

Figure 9 : Évolution temporel du contenu fréquentiel du séisme de Loma prieta.

Figure 10 : Caractérisation de séisme de Loma preita dans le domaine fréquentiel.

Figure 11 : intervalle de confiance.

Figure 12 :a) enveloppe temporelle utilisée pour la simulation.

b) accélérogramme simulé.

c) fréquence centrale en fonction de temps (Sabetta et al)

Figure 13 : comparaison des fréquences centrales.

Figure 14 : Accélérogramme simulé pour $M=5, R=20$, sol raide et spectre de réponse en accélération

Figure 15 : Accélérogramme simulé pour $M=5, R=37$, dépôts profond et spectre de réponse en accélération

Figure 16 : Accélérogramme simulé pour $M=5, R=37$, dépôts superficielle et spectre de réponse en accélération

Introduction Générale.

A l'heure actuelle, la prévention reste l'outil prépondérant pour l'amélioration de la protection des population à court et moyen termes vis-à-vis des séismes : l'état des connaissances dans le domaine de la prédiction ne permet encore aucun espoir sérieux.

Ainsi, il est nécessaire d'améliorer l'état de nos connaissances sur la caractérisation des séismes, afin de pouvoir se prémunir contre leurs effets dévastateurs en dimensionnant correctement qui y sont sujettes.

En sismologie, les lois d'atténuation sont des expressions analytiques ou graphiques qui permettent de calculer un paramètre de mouvement du sol (valeurs de pic de l'accélération, de la vitesse ou du déplacement, ordonnée spectrale, durée,.....), en fonction d'un certain nombre de paramètres caractérisant le séisme et la propagation des ondes sismiques entre la source et le site étudié, ainsi que la réponse des terrains superficiel du site à l'action des ondes. Le terme d'atténuation sous-entend que l'amplitude des paramètres du mouvement décroît quand on s'éloigne de la source, ce qui est généralement le cas, aux effets locaux d'amplification près et à l'exception toutefois de la durée.

Les lois d'atténuation sont basées sur l'exploitation statistique de données accélérométriques, d'origines géographiques et de qualités variées, obtenues pour des séismes de magnitude, de profondeurs et de mécanismes différents, enregistrés à des distances variables et sur des sites géologiques divers. De nombreux auteurs proposent ainsi des lois d'atténuation des indicateurs de mouvement sismique établies à partir d'analyse de régression effectuées sur une population d'enregistrement de mouvements forts. Avec l'augmentation du nombre et de la qualité de ces enregistrements, de nombreuses lois, dans ce sens, en été établies.

Dans le présent document, on s'est attaché à inventorier des lois d'atténuation de la durée de mouvement fort, l'intensité d'Arias, la fréquence centrale et la bande fréquentielle pour les mouvements forts. L'objectif est de fournir aux utilisateurs un ensemble de lois, établies pour les différents contextes tectoniques, qui leurs permettent d'estimer les mouvements forts qui peuvent être produits par des séismes de caractéristiques différents se déclenchant à des distance données des sites étudiés.

Chapitre I :

Généralités sur les séismes.

1. Introduction.

Les tremblements de terre ont représenté depuis toujours un des plus graves désastres de l'humanité. Leurs apparitions brutales et imprévues, la violence des forces mises en jeu, l'énormité des pertes humaines et matérielles ont marqué la mémoire des générations. Parfois, les dommages sont comparables à un budget national annuel.

De nos jours, avec l'accélération du développement industriel, la réalisation des constructions audacieuses, telles que : barrages, immeubles de grande hauteur abritant des centaines de personnes, etc. la concentration d'une grande partie de la population s'impose comme une absolue nécessité du point de vue social, économique et politique.

Ainsi l'établissement d'un projet de construction parasismique rationnelle, exige de la part de l'ingénieur l'acquisition préalable d'un minimum de connaissance concernant les caractéristiques spécifiques de l'action sismique.

2. Processus de tremblement de terre.

Le tremblement de terre consiste en un mouvement vibratoire du sol de forte amplitude de caractère aléatoire, correspondant à une libération brutale d'une partie de l'énergie qui se propage sous forme d'ondes élastiques ayant pris naissance à l'intérieur de l'écorce terrestre, ce phénomène se spécifie par une durée variable mais toujours brève. Les ondes se diffusent dans toutes les directions à partir de la source et se manifestent sous forme de déplacement du sol qui sera enregistré par des stations dotées d'instruments adéquats.

A l'origine de tels phénomènes se trouvent les déformations lentes de la croûte terrestre ou du milieu sous-jacent, qui entraînent l'accumulation de tension interne. Ces dernières finissent par dépasser, en un point donné, la résistance des roches au cisaillement, produisant à l'endroit en question une rupture qui se propage ensuite. La faille résultante constitue désormais une zone de faiblesse dans la croûte.

La création d'une faille ou, plus fréquemment, le glissement le long d'une faille existante, constitue le mécanisme générateur d'un séisme. L'endroit où se produit le séisme est appelé foyer ou *hypocentre* et la projection de ce point sur le sol est appelée *épicentre* (cf. fig. 1).

L'hypocentre peut se situer à une profondeur très variable : quelques kilomètres à plus de 100km.

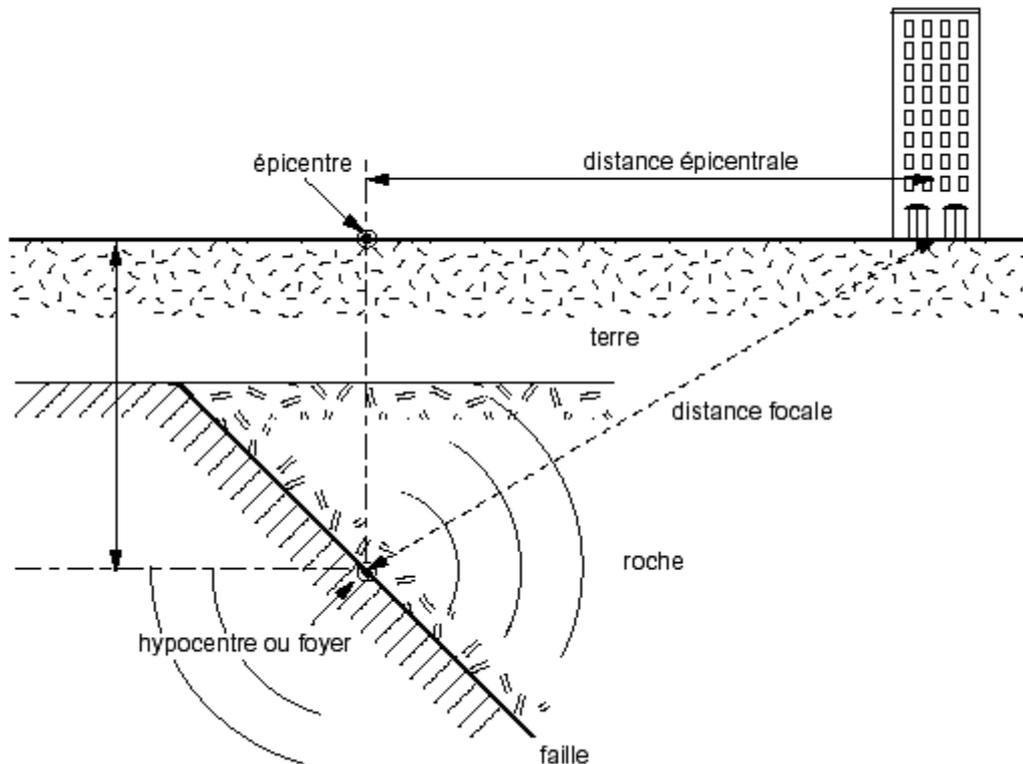


Figure 1 Vocabulaire et éléments caractéristiques en géo sismique

Les propriétés de tremblement de terre dépendent de nombreux facteurs dont notamment : la magnitude, le mécanisme au foyer et la profondeur du foyer, la distance de site à la source, la direction de propagation, la configuration des milieux traversés ainsi que les caractéristiques géologiques géotechniques et topographiques du site.

Il est bien entendu qu'à partir de certaines distances de la source allant de quelques dizaines de kilomètres pour les séismes modérés à quelques centaines de kilomètres pour les grands séismes, les mouvements sismiques n'affectent plus les constructions.

En génie sismique, l'ingénieur alloue un intérêt aux effets locaux des mouvements sismiques importants pour lesquels le mouvement est suffisamment intense pour engendrer des endommagements structuraux. Ces mouvements sismiques sont dits *mouvements forts*.

3. Mouvement fort.

L'étude analytique et expérimentale du mouvement fort a pu se développer ces dernières années grâce aux réseaux de sismographes et d'accélérographes installés dans certaines régions sismiques du monde.

L'intérêt pour une telle étude s'est accentué en raison des problèmes de sécurité se posant avec une acuité de plus en plus grande en ce qui concerne certaines installations stratégiques particulièrement sensibles telles que les ponts, les centrales nucléaires etc.

L'étude du mouvement fort en zone proche se définit comme étant la mesure et l'interprétation physique du mouvement du sol au voisinage d'une faille dans une zone où ces mouvements mettent en danger les constructions et la vie humaine.

Les enregistrements obtenus par systèmes de mesures des mouvements forts sont requis par l'ingénieur aussi bien étudier les propriétés des mouvements sismiques proches de la source que pour l'étude de la variation spatiale des ondes sismiques.

4. Propagation des ondes sismiques.

A partir du foyer la secousse sismique se propage sous forme d'ondes qui, compte tenu de l'hétérogénéité du sol, engendrent en surface un mouvement vibratoire complexe difficilement prévisible en un site donné.

On distingue deux types d'onde : les ondes de volume et les ondes de surface.

Les ondes de volume :

Elles prennent naissance au foyer et se propagent à l'intérieur du manteau terrestre sous deux formes : *les ondes longitudinales* se caractérisant par des compressions et dilatations alternées se propageant à une vitesse de 7 à 8 km/s ; *les ondes transversales* se caractérisant par une distorsion dans le plan perpendiculaire à la direction de propagation, provoquant du cisaillement et se propageant à une vitesse de 4 à 5 km/s (cf. fig. 2).

Nota : C'est la différence de vitesse de propagation des ondes longitudinales et transversales qui permet, à l'aide de l'enregistrement de plusieurs sismographes, de situer le foyer d'un séisme.

Les ondes de surface :

Elles sont générées par les ondes de volume qui arrivent à la surface et se propagent à la vitesse de 1,5 à 5 km/s. On distingue :

Les ondes de RAYLEIGH pour lesquelles les points du sol décrivent des ellipses dans le plan vertical ; elles entraînent des compressions et des cisaillements dans le sol ;

Les ondes de LOVE pour lesquelles les points du sol se déplacent tangentiellement à la surface, perpendiculairement à la direction de propagation ; elles n'entraînent que des contraintes de cisaillement (cf. fig. 2).

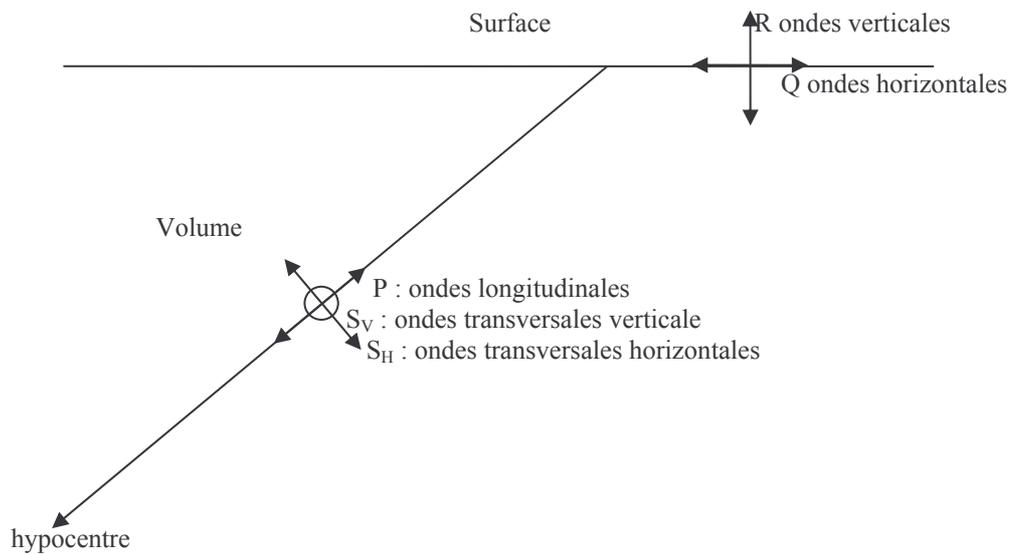


Figure 2 Les ondes sismiques de volume et de surface

5. Représentation du mouvement sismique.

Un sismographe permet de restituer sous forme de représentation analogique les mouvements sismiques sous forme de déplacement, vitesse ou accélération. Les enregistrements correspondant à l'accélération en fonction du temps sont des accélérogrammes.(cf.fig. 3)

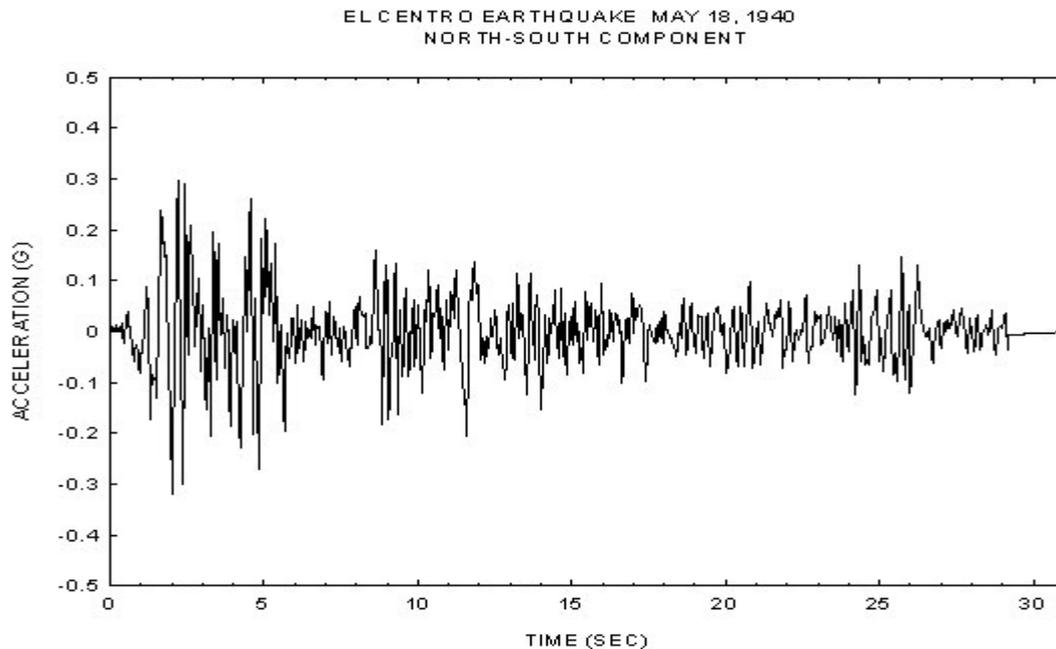


Figure 3 Accélérogramme

Ainsi, sur un enregistrement accélérométrique un accélérogramme $x(t)$ est une fonction de variations rapides et irrégulières autour de zéro dont la durée totale T peut varier de quelques secondes à quelques dizaines de secondes. Généralement, cette fonction se décompose en une phase initiale pour laquelle les accélérations sont faibles, une phase forte suivie d'une phase modérée et une phase progressive de retour à une accélération finale nulle. La durée totale d'un séisme est largement conditionnée par le seuil de déclenchement des appareils de mesures ce qui ne lui offre qu'une représentation physique limitée, d'autant plus que les phases à faibles accélérations, d'une durée variable, sont généralement sans intérêt pour l'analyse des structures. La durée de phase forte est en revanche plus significative pour conduire l'analyse sismique d'un ouvrage.

6. Caractéristiques enregistré à la surface du sol.

Dans la réalité, les choses sont encore plus complexes ; en effet, la propagation d'une onde sismique dans un milieu hétérogène provoque, pour chaque discontinuité, un système

complexe d'ondes réfractées et réfléchies, si bien qu'au niveau du sol le mouvement sismique a un aspect tout à fait aléatoire.

Cependant les mouvements vibratoires, occasionnés à la surface du sol par les séismes, présentent des caractéristiques communes et un certain nombre de paramètres sont habituellement utilisés pour les décrire.

Caractéristiques des vibrations aléatoires induites à la surface du sol par un séisme :

- Direction : Le mouvement présente simultanément des composantes verticales et horizontales qui sont indépendantes ;
- Durée : Elle est généralement comprise entre 15 et 30 s. (pour un séisme intense elle peut être de 60 à 120 s) ;
- Fréquence : Le mouvement aléatoire à bande large présente une énergie prépondérante entre 1 et 35 Hz, et provoque les effets les plus destructifs entre 1 et 10 Hz ;
- Grandeurs Maximale : Les grandeurs maximales sont :

- a_{\max} la composante max de l'accélération (généralement exprimée en nombre de g),

- L'accélération a_{rms} **Root Mean Square** (rms) ($a_{rms} = \sqrt{\frac{1}{T} \int_0^T a^2 dt}$)

où a est l'accélération enregistrée et T est la durée totale de l'enregistrement.

- v_{\max} (cm/s) la composante max de la vitesse et d_{\max} (cm) est la composante max du déplacement.

7. Paramètres affectant le mouvement sismique [2].

Le mouvement sismique en un point est affecté par des facteurs à la source, au trajet parcouru entre le foyer et le site et par des facteurs liés à des conditions locales propres au site [IDRISS, 1978]. Pour être en mesure d'évaluer la nature du mouvement sismique sur site, il est nécessaire de connaître l'influence de ces facteurs et de les quantifier. Ceci est réalisé à l'aide de relations dites, exprimant un des différents indicateurs du mouvement sismique (durée de phase forte, Intensité d'Arias, fréquence centrale et bande fréquentielle...) en fonction de l'ensemble des paramètres caractéristiques.

7.1 Facteurs liés à la source.

Ces facteurs sont fonction des dimensions de la zone de rupture, des valeurs des contraintes le long de la surface de rupture, du mécanisme de la rupture et de sa propagation le long de la faille.

D'un point de vue fondamental, ils sont caractérisés par des paramètres principaux : le moment sismique M_0 et la chute de contraintes $\Delta\tau$. L'utilisation de ces deux paramètres est limitée encore aux sismologues. Les ingénieurs préfèrent caractériser un séisme soit par sa magnitude, soit par son intensité à l'épicentre. Une certaine confusion règne dans l'utilisation de ces deux derniers paramètres du fait de la multiplicité des échelles servant à les mesurer. Ce sont pourtant eux qui servent de base aux évaluations de niveau sismique d'un site.

Moment sismique.

Il existe de nombreuses évidences indiquant que, sauf peut-être pour les séismes profonds, un séisme est engendré par un mécanisme de cisaillement pur duquel les déformations par dilatance ou contractance sont absentes. Le mécanisme à la source peut être alors représenté par un ensemble de forces constituant un double couple dont le paramètre caractéristique est M_0 , le moment sismique [BRUNE,1976]. Considérons une dalle mine (fig. 4) correspondant au plan de la faille. Le couple dû aux contraintes de cisaillement est :

$$M_0 = \iint_S \tau h \, dA \quad (1.1)$$

En désignant par G le module de cisaillement du milieu (identique au paramètre de Lamé μ de la mécanique des milieux continus) et par u le déplacement relatif des deux faces de la faille, il vient pour un milieu élastique

$$M_0 = \iint_S G u \, dA = G \tilde{u} A \quad (1.2)$$

Où \tilde{u} est le déplacement relatif moyen des deux bords de la faille et A l'aire de la surface de rupture. M_0 est généralement exprimé en dyne-cm.

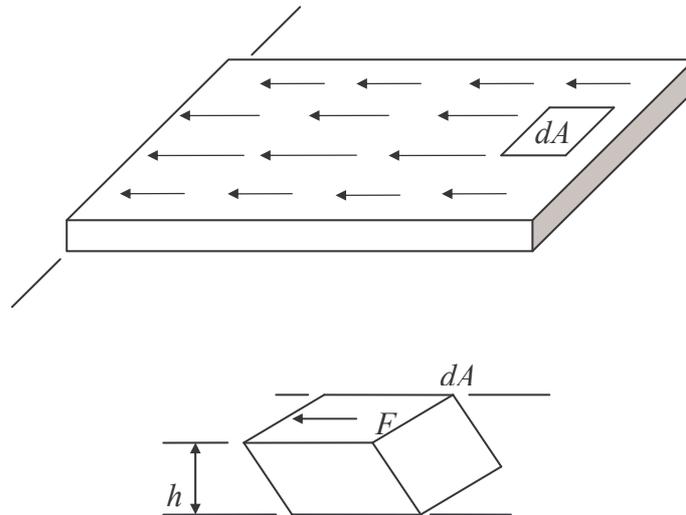


Figure 4 Schématisation d'un plan de faille (Brune 1976)

Magnitude.

La magnitude représente une mesure de l'énergie libérée par la rupture de la faille dans la zone source. Elle est calculée comme étant le logarithme d'amplitude maximale enregistrée par un sismographe de type **Wood-Anderson**, ayant une période propre de 0.8sec et un pourcentage d'amortissement critique de 80 %, situé à 100 km de l'épicentre.

D'une façon générale la magnitude est calculée par l'expression :

$$M = \log \frac{A}{T} + F(\Delta, h) + S \quad (1.3)$$

Où : A et T représente l'amplitude maximale et la période de l'onde sismique enregistrée ; F le facteur de correction, fonction de la distance épacentrale Δ et de la profondeur h et S un facteur de site. La magnitude d'un séisme est un réel exprimé en chiffres arabes.

Pour un séisme donné, il n'y a pas une magnitude unique. On distingue la magnitude des ondes de volumes M , la magnitude des ondes de surfaces M_s , la magnitude locale M_L définie par Richter. De plus, la magnitude peut varier d'une station à l'autre suivant la position géographique de celle-ci. Cela est dû à des schémas de radiation non uniformes autour de l'épicentre.

Suivant le type d'ondes considérées, la magnitude représente une mesure de l'énergie libérée dans une certaine bande de fréquences. La magnitude des ondes de surface correspond à des périodes de l'ordre de 20 secondes et celle des ondes de volume à des périodes comprises entre 1 seconde et 5 secondes. Il convient donc de se montrer prudent dans l'utilisation des magnitudes d'autant qu'il n'y a pas uniformité d'utilisation des magnitudes

entre les différents auteurs. Pour l'ingénieur, les magnitudes M et M_L sont plus significatives car elles correspondent à la plage des fréquences de vibration propre des ouvrages usuels.

Plusieurs relations empiriques ont été proposées pour relier l'énergie libérée lors d'un séisme (énergie libérée sous forme radiative E_r , à la magnitude par exemple :

$$\log E_r = 11,4 + 1,5M \quad (\text{ergs}) \quad (1.4)$$

La magnitude peut être reliée au moment sismique, cas des séismes importants, [Lomnitz ; 1974] a proposé la relation :

$$\log M_0 = 19,9 + M_L \quad (1.5)$$

Intensité.

L'échelle d'intensité attribuée à un séisme, caractérise de son importance en un lieu donné, d'après les manifestations ressentait par les personnes et les dégâts subis par les constructions.

L'intensité d'un séisme dépend du lieu où elle est mesurée et décroît généralement quand la distance épacentrale du lieu d'observation augmente. Elle est exprimée en chiffres romains. Il existe deux grandes échelles de mesure :

- L'échelle macroscopique internationale - l'échelle MERCALLI
- L'échelle macroscopique M.S.K.

7.2 Facteurs liés au trajet parcouru.

Diffraction, polarisation, conversion.

Une partie de l'énergie libérée à la source se propage par radiation d'ondes sismiques. Lors de leur trajet entre le foyer et le site à la surface de la terre, ces ondes rencontrent des hétérogénéités, des surfaces de discontinuité, des barrières naturelles. Ces obstacles ont pour effet de réfléchir et de réfracter les ondes, de les polariser et de convertir en d'autres types d'ondes : conversion d'ondes de volume en ondes de Rayleigh et en ondes de Love dans les milieux stratifiés.

Atténuation.

Par ailleurs, une partie de l'énergie transportée par ces ondes est absorbée dans les milieux qu'elles traversent du fait de l'amortissement propre, ou amortissement matériel, lié à l'hystérésis des matériaux.

Filtrage.

Tous ces phénomènes affectent la nature du mouvement sismique. L'éloignement de la source affecte le mouvement en diminuant son amplitude (diminution des accélérations, vitesses, déplacement avec la distance) et en modifiant son contenu spectral : les hautes fréquences sont filtrées et la fréquence fondamentale du mouvement abaissée. On a également noté précédemment que les magnitudes attribuées par divers organismes à un même séisme varient suivant la distribution géographique des stations à la surface de la terre. Cela reflète bien l'influence de la nature des terrains traversés.

La prise en compte de ces phénomènes est extrêmement complexe et impose en particulier de connaître la structure interne de la terre entre le foyer et le site. Celle-ci n'est généralement qu'approximativement connue et on rend compte de l'influence du trajet parcouru à l'aide d'un seul paramètre : la distance de la source au site. Suivant les auteurs cette distance peut être la distance hypocentrale, la distance épacentrale ou la plus courte distance de la faille au site. La distance hypocentrale (ou focale) est souvent mal connue. La distance épacentrale est mieux connue, mais son utilisation ne peut se concevoir que dans une zone donnée, où les profondeurs focales restent pratiquement équivalentes d'un séisme à l'autre. L'utilisation de la plus courte distance de la faille au site part de la considération simpliste que, le séisme résultant de la propagation d'une rupture le long d'un plan donné (le plan de faille), le mouvement sismique en un point sera plus fortement influencé par la plus courte distance à la zone de libération de l'énergie que par la distance au point où s'est initiée la rupture.

7.3 Factures liés aux conditions locales.**Filtrage des amplifications :**

Les conditions géologiques locales, les conditions topographiques, les interactions sol structure modifient la nature de mouvement sismique. On s'est généralement attaché à mettre en évidence l'effet de la géologie locale (nature et caractéristiques des terrains).

L'influence de la géologie locale est généralement mise en évidence à l'aide d'une classification sommaire de la nature des sites : Rocher, sols raides, intermédiaires ou très mous. La modification du mouvement sismique, lorsque l'on passe d'un type de sol à un autre, peut être examinée à l'aide des paramètres caractéristiques de ce mouvement : accélération ou vitesse maximale, spectre de réponse,... Toutes les études statistiques réalisées sur des enregistrements réels [SEED et al. ,1976 ; TRIFUNAC et BRADY, 1975 ;

JOYNER et BOORE, 1981] montre que l'accélération maximale en surface du sol est le paramètre le moins affecté par la nature et les caractéristiques des couches sous-jacentes. Les sols mous font exception à la règle, en présentant une atténuation moins rapide de l'accélération maximale avec la distance. De plus, aux fort niveaux d'accélération, leur capacité de résistance moins élevée limite les accélérations maximales qu'ils sont capables de transmettre en surface. La vitesse ou le déplacement maximale sont des paramètres beaucoup plus sensibles à la nature du sol. Plus la rigidité du sol est faible, plus grands sont les déplacements et les vitesses. Le sol se comporte comme un gros appui néoprène qui écrête les efforts, et donc les accélérations, en compensation d'une augmentation des déplacements. L'influence de la nature de sol est également mise en évidence de façon très nette en examinant les spectres de réponse à la surface du sol. SEED, UGAS et LYSMER (1976) ont réalisé une étude statistique des spectres de réponses en accélération d'enregistrements de séismes en surface de différents types de sol, répartis en quatre catégories. Leurs résultats sont présentés sur la figure 5. La fréquence prédominante du mouvement, correspondant au pic du spectre de réponse, diminue lorsque la raideur du sol diminue. Autrement dit, plus le sol est mou plus il filtre les hautes fréquences contenues dans le mouvement incident. Par ailleurs, la plage de fréquence où l'accélération spectrale reste élevée est plus étendue pour les sols à faible raideur.

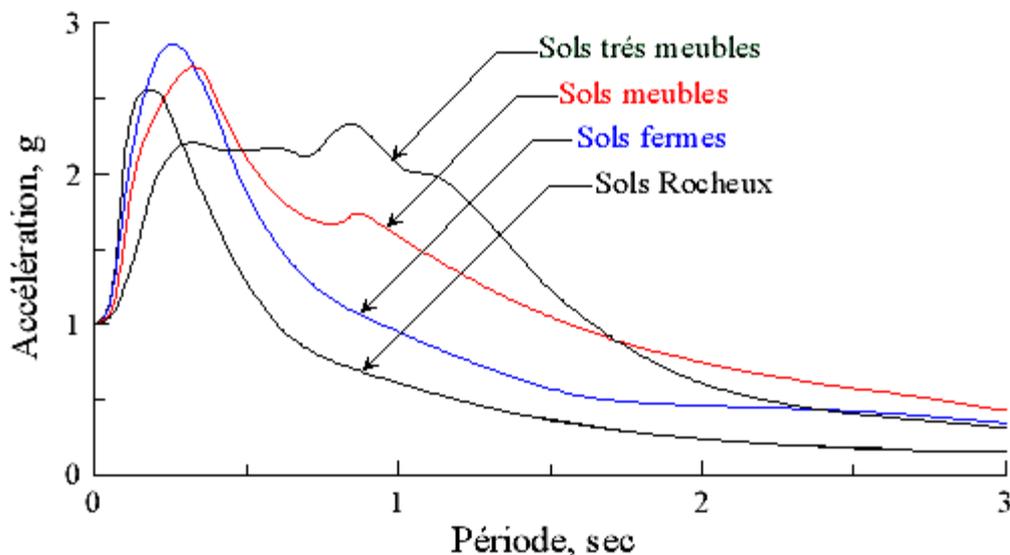


Figure 5 Spectres moyens d'accélération pour différentes conditions de site selon Seed et al. (1976)

Remarque :

Les études de [SEED et al] ont été basées sur des études statistiques ne faisant pas intervenir la magnitude. Autrement dit, la forme du spectre est indépendante de la magnitude et de la

distance à la source. Li semblerait qu'en réalité, la forme du spectre de réponse dépende également de ces paramètres.

8. Spectre de réponse.

En génie civil, la notion la plus répandue pour représenter un séisme en ingénierie sismique est le spectre de réponse d'oscillateurs linéaires. Le spectre de réponse en accélération, vitesse ou déplacement met en évidence le contenu fréquentiel du mouvement : il fournit la réponse maximale d'oscillateurs linéaires élastiques soumis à un séisme. Cette réponse se calcule simplement pour un oscillateur linéaire de masse m , de raideur k et d'amortissement c , dont le déplacement relatif par rapport à son support soumis à une excitation sismique représentée par son accélérogramme $a(t)$, sera noté $u(t)$.

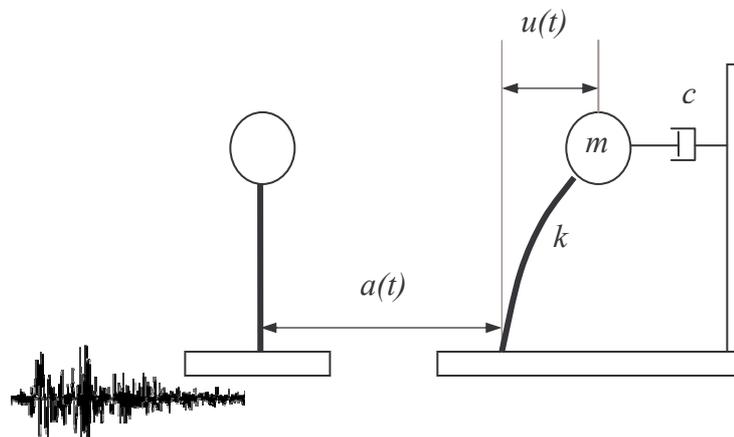


Figure 6 Représentation d'un oscillateur simple soumis à une excitation sismique

Le principe fondamental de la dynamique s'écrit pour cet oscillateur :

$$m\ddot{u} + c\dot{u} + ku = ma(t) \quad (1.6)$$

En posant $\omega_n = \frac{k}{m}$, $\xi = \frac{c}{2m}$ et $\omega_D = \omega_n \sqrt{1 - \xi^2}$, la réponse de l'oscillateur a pour expression :

$$u(t) = -\frac{1}{\omega_D} \int_0^t a(\tau) e^{-\xi\omega_n(t-\tau)} \sin[\omega_D(t-\tau)] d\tau \quad (1.7)$$

Le maximum de cette réponse constitue l'ordonnée d'un point du spectre en déplacement ayant pour abscisse la fréquence propre ou la pulsation propre de l'oscillateur ω_n .

Le calcul de cette intégrale (intégrale de Duhamel) est fastidieux et le tracé des spectres nécessite de la calculer pour un nombre de pulsations généralement très important. La résolution de l'équation (1.2) dans le domaine fréquentiel est beaucoup plus efficace.

En appelant $h(t)$ la réponse impulsionnelle (réponse à une impulsion de Dirac $\delta(t)$) d'un système linéaire et $a(t)$ l'excitation à l'entrée, la réponse du système considéré causal s'écrit :

$$u(t) = \int_0^{\infty} h(\tau) a(t - \tau) d\tau \quad (1.8)$$

Ce produit de convolution se transforme en produit simple en substituant aux expressions temporelles leur transformée de Fourier. Pour l'oscillateur représenté de la figure 3, la transformée de Fourier de la réponse impulsionnelle s'écrit :

$$H(i\omega) = \frac{-1}{(\omega_n^2 - \omega^2) + 2i\xi\omega\omega_n} = \frac{U(i\omega)}{A(i\omega)} \quad (1.9)$$

avec $U(i\omega)$ et $A(i\omega)$ les transformées de Fourier respectivement du déplacement relatif de l'oscillateur $u(t)$ et de l'excitation $a(t)$.

Cette fonction représente la fonction de transfert entre le déplacement relatif du système et l'accélération du support. Le calcul de la réponse de l'oscillateur s'effectue dans le domaine fréquentiel par l'évaluation du simple produit $H(i\omega)A(i\omega)$. Le maximum de la transformée de Fourier inverse de cette réponse constitue l'ordonnée d'un point du spectre en déplacement ayant pour abscisse la fréquence propre ou la pulsation propre de l'oscillateur ω_n . Une procédure de calcul de spectres de réponse en déplacement a été développée sur ce principe, puis validée sur un grand nombre d'exemples d'accélérogrammes naturels par comparaison des résultats avec ceux obtenus à partir de codes déjà qualifiés. Le spectre de réponse en pseudo-accélération est déduit de la relation suivante : $\tilde{a}_{\max} = \omega_n^2 d_{\max}$.

où la pseudo-accélération \tilde{a}_{\max} constitue une bonne approximation du maximum du module de l'accélération absolue de l'oscillateur pour les amortissements faibles utilisés en pratique.

9. Aspects stochastiques des accélérogrammes des mouvements forts.

L'étude des accélérogrammes des tremblements de terres à fort mouvement montre une nature stochastique évidente du signal. Cette nature aléatoire est due à deux catégories de paramètres :

- La première catégorie concerne la profondeur du foyer, la distance du site considéré à l'épicentre, la quantité d'énergie libérée dans la zone du foyer, durée pendant laquelle l'énergie se libère, la vitesse de rupture qui est variable. Ces paramètres liés aux

séismes lui-même et à sa localisation par rapport au site où se trouve la structure étudiée ont un caractère aléatoire évident.

- La seconde catégorie de paramètres à considérer est relative au fait que les ondes sismiques atteignent le site considéré après avoir subi de nombreuses diffractions, réflexions dans les couches du sous-sol. Ces phénomènes complexes font que les accélérogrammes ont une nature stochastique.

Ainsi, deux accélérogrammes de même séisme mesuré en deux points voisins dans un même site seront généralement différents. Donc, il est important de considérer les accélérogrammes comme des réalisations d'un processus stochastique, non pas comme des fonctions déterministes du temps.

Chapitre II :

Éléments d'analyse des signaux aléatoires

1. Processus stochastique [1],[6].

1.1 Définition.

Si un phénomène physique est aléatoire, alors chaque échantillon $X(t)$ de ce phénomène est le résultat de circonstances uniques qui ne se répétant pas à l'occasion d'autres mesures indépendantes de ce même phénomène. Un processus aléatoire est une famille paramétrée (le temps par exemple) de variables aléatoires. Dans le cas de plusieurs paramètres, on parle de champ aléatoire, si le paramètre est discontinu, on parle de séquence aléatoire.

La description stochastique d'un processus aléatoire est généralement obtenue à partir des lois de probabilités marginales et conjointes des variables aléatoires qui représentent le comportement du processus en des instants prescrits. Selon que ces variables sont continues ou discrètes on parle de processus aléatoire continu ou discret.

1.2 Description probabiliste de processus stochastique.

1.2.1 Fonction de densité de probabilité.

Le moyen le plus efficace pour caractériser un processus aléatoire est de spécifier ses fonctions de densité de probabilité. La première densité, fournit la structure de probabilité de la variable aléatoire $X(t)$ pour chaque valeur fixée de t , elle est définie par :

$$P_x(x, t) = \frac{P(x \leq X \leq x + dx, t)}{dx} = \frac{F(x + dx, t) - F(x, t)}{dx} \quad (2.1)$$

Où

$P(x \leq X \leq x + dx, t)$: représente la probabilité pour les valeurs de processus soient comprise dans l'intervalle $[x, x + dx]$ en t ,

$F(x, t)$: est la fonction de répartition de la variable aléatoires $X(t)$, c'est-à-dire quelle représente la probabilité que les valeurs du processus soient comprises dans l'intervalle $]-\infty, x]$ en t .

Quant aux densités d'ordres supérieures, elles décrivent la dépendance mutuelle des ordonnées de la fonction aléatoire.

$$\begin{aligned} P(x_1, t_1; x_2, t_2) \\ P(x_1, t_1; x_2, t_2; \dots; x_n, t_n) \end{aligned} \quad (2.2)$$

Toutes ces densités sont des fonctions non négatives et symétriques par rapport à leurs arguments, elles vérifient la condition de normalisation.

$$\int \int \dots \int_{\text{n fois}} P(x_1, t_1; x_2, t_2; \dots; x_n, t_n) dx_1 dx_2 \dots dx_n = 1 \quad (2.3)$$

Les densités d'ordre inférieur peuvent toujours être déduites par intégration, au départ d'une densité d'un ordre supérieur.

$$P_K = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} P(x_{K+1}, t_{K+1}; x_2, t_2; \dots; x_n, t_n) dx_1 dx_2 \dots dx_n \quad (2.4)$$

Les densités de probabilités constituent donc une description du processus aléatoire dont la précision augmente avec n . Elles offrent une classification naturelle de la complexité statistique d'un processus; en fonction de l'ordre n nécessaire à la spécification complète du processus. La classe la plus simple sera celle qui est entièrement définie par la densité d'ordre $n=1$: processus purement aléatoires.

De même, on définit les densités de probabilités conjointes de deux processus aléatoires $X(t)$ et $Y(s)$ par $P_{xy}(x_1, t_1; y_1, t_1)$ (2.5)

$P_{xy}(x_1, t_1; y_1, t_1) dx_1 dy_1$: représente la probabilité que le processus X ait une valeur dans l'intervalle $]x_1, x_1 + dx_1]$ en t_1 et Y dans $]y_1, y_1 + dy_1]$ en $s = s_1$.

1.2.2 Fonctions moments.

Les moments d'une ou plusieurs variables aléatoires sont définis comme les espérances mathématiques des différentes puissances de ces variables aléatoires. Pour une seule variable, $E[X^n]$ est le moment d'ordre n ; $E[X^m Y^n]$ est le moment conjoint d'ordre " $m+n$ " des variables X et Y . Une fois connue la distribution de probabilité, tous les moments peuvent être calculés. Du fait de la complexité engendrée par la prise en compte des moments d'ordres supérieurs, les développements pratiques sont limités en général à ces trois moments.

- **La moyenne :**

$$m_X = E[X(t)] = \int_{-\infty}^{+\infty} x P_X(x, t) dx \quad (2.6)$$

- **la variance :**

$$\sigma_X^2 = VAR[X(t)] = E[(X - m_X)^2] \quad (2.7)$$

- **La fonction d'autocorrélation :**

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 P_{XX}(x_1, t_1; x_2, t_2) dx_1 dx_2 \quad (2.8)$$

Ces trois moments sont d'une importance pratique considérable, cette importance est du fait qu'ils caractérisent entièrement le processus, ils contiennent des informations importantes (tendance, dispersion).

Pour deux processus différents $X(t)$ et $Y(t)$, on définit la fonction de **cross-corrélation** par

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \quad (2.9)$$

Par ailleurs, on définit la fonction d'**autocovariance** et de **crosscovariance** respectivement comme suit :

$$\Gamma_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - m_X(t_1)m_X(t_2) \quad (2.10)$$

$$\Gamma_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - m_X(t_1)m_Y(t_2) \quad (2.11)$$

1.3 Processus aléatoire stationnaire.

Un processus aléatoire est dit stationnaire ou strictement stationnaire, si les distributions de probabilités obtenues pour l'ensemble ne dépendent que des temps relatifs ou encore, si sa structure de probabilité soit invariante vis-à-vis d'un changement de l'origine des temps. Il en résulte que la densité de probabilité du premier ordre est donc indépendante du temps; Les densités d'ordres supérieures ne sont fonction que de la différence entre les instants considérés, sans égard pour l'origine des temps.

$$F(x_1, t_1; \dots; x_n, t_n) = F(x_1, t_1 + \tau; \dots; x_n, t_n + \tau) \quad (2.12)$$

Dans ce qui précède on peut déduire que :

$$m_X = \int_{-\infty}^{+\infty} x P(x) dx = \text{constante} \quad (2.13)$$

m_X est donc indépendante du temps.

Les fonctions d'autocorrélation $R_X(\tau)$ et d'autocovariance $\Gamma_X(\tau)$ sont aussi indépendantes du temps et ne dépendent que de la différence τ entre les instants considérés, en effet

$$E [X(t_1)X(t_2)] = R_{XX}(t_2 - t_1) = R_X(\tau) \quad (2.14)$$

$$E [X(t_1 - m_X)X(t_2 - m_X)] = \Gamma_{XX}(x_1, t_2; x_2, t_1) = \Gamma_{XX}(t_1, t_2) = \Gamma_X(\tau) \quad (2.15)$$

1.4 Ergodicité.

Considérons un processus $X(t)$ stationnaire, sa moyenne temporelle est définie par

$$S = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(t) dt = \langle X(t) \rangle_T \quad (2.16)$$

Un processus est dit ergodique, si en plus de la stationnarité, les moyennes temporelles, sur des échantillons (réalisation), et les moyennes d'ensembles sont identiques, en termes pratiques, chaque réalisation est alors complètement représentative de l'ensemble constituant le processus.

$$E[X(t)] = \lim_{T \rightarrow \infty} \langle X(t) \rangle_T \quad (2.17)$$

Un processus est dit ergodique dans la corrélation si

$$R_X(\tau) = \lim_{T \rightarrow \infty} \langle X(t)X(t+\tau) \rangle_T \quad (2.18)$$

1.5 Propriétés des fonctions d'autocorrélation.

Dans ce qui précède, nous avons toujours considéré des processus réels. Les expressions des fonctions d'autocorrélation dans le cas de processus complexes sont :

$$R_{XX}(t_1, t_2) = E [X^*(t_1)X(t_2)] \quad (2.19)$$

$$R_{XY}(t_1, t_2) = E [X^*(t_1)Y(t_2)] \quad (2.20)$$

Où * désigne le complexe conjugué.

Pour les cas des processus réels stationnaires, la fonction d'auto corrélation joint des propriétés suivantes :

- La symétrie (parité)

$$R_{XX}(\tau) = R_{XX}(-\tau) \quad (2.21)$$

Sa valeur maximale est atteinte à son origine ($\tau = 0$)

Cette propriété peut être montrée à partir du coefficient de corrélation ρ et à travers le cheminement suivant :

$$\rho(\tau) = \frac{R_{XX}(\tau) - m_X^2}{\sigma_X^2} \quad (2.22)$$

Comme les valeurs de ce coefficient sont bornées par +1 et -1, il s'ensuit l'inégalité suivante :

$$-\sigma_X^2 + m_X^2 \leq R_{XX}(\tau) \leq \sigma_X^2 + m_X^2 \quad (2.23)$$

Finalement, on retrouve la relation :

$$|R_{XX}(\tau)| \leq R_{XX}(0) \quad (2.24)$$

- Fonctions de covariance.

Les fonctions de covariance sont des fonctions d'autocorrélation particulières, elles possèdent les mêmes propriétés. En outre, si le processus ne contient aucune composante périodique, la fonction d'autocovariance diminue quand l'écart entre ses arguments augmente ($\tau \rightarrow \infty$), dans ces mêmes conditions la fonction d'autocorrélation tend vers le carré de la moyenne, ainsi on pourra avoir :

$$\lim_{\tau \rightarrow \infty} \Gamma_{XX}(\tau) = 0 \quad (2.25)$$

Et on aboutit finalement à :

$$R_{XX}(\tau \rightarrow \infty) \rightarrow m_X^2 \quad (2.26)$$

1.6 Propriété des fonctions de covariance.

En général, les fonctions de cross corrélation ne sont ni symétriques, ni paires.

- Par le biais de l'inégalité de Schwarz on déduit que la valeur absolue de la fonction de cross corrélation est bornée par la valeur de la racine carrée du produit des moyennes carrées des deux processus.

$$|R_{XX}(\tau)| \leq \sqrt{|R_{XX}(0)R_{YY}(0)|} \quad (2.27)$$

- La corrélation entre deux processus devient très faible, lorsque l'écart entre les instants est large et par conséquent, la fonction de cross corrélation tend vers le produit des moyennes de deux processus.

$$\begin{aligned} R_{XY}(\tau \rightarrow \infty) &\rightarrow m_X m_Y \\ R_{YX}(\tau \rightarrow \infty) &\rightarrow m_Y m_X \end{aligned} \quad (2.28)$$

2. Décomposition spectrale d'un processus aléatoire.

Il est important de faire une analyse spectrale, d'un processus aléatoire. Soit une fonction réelle $X(t)$, sa **transformée de Fourier** est définie par :

$$x(f) = \int_{-\infty}^{+\infty} X(t) e^{-2\pi f t} dt \quad (2.29)$$

Où $x(f)$ constitue l'amplitude spectrale de la fonction considérée. L'existence de $x(f)$ est garantie si :

- $X(t)$ est absolument intégrable.

- L'intégrale suivante à une valeur finie. $\int_{-\infty}^{+\infty} |X(t)| dt < \infty$

La notion de transformée de Fourier nous sera utile pour définir la décomposition spectrale d'un processus aléatoire.

2.1 Cas des processus stationnaires.

Dans le cas des signaux non transformables dans le domaine de Fourier (processus aléatoire stationnaire), la fonction d'auto corrélation et sa transformée de Fourier constituent une alternative très appropriée pour leur description dans le domaine spectral sur cette base, nous définissons la fonction de densité spectrale de puissance (DSP) d'un processus stationnaire $X(t)$, notée $S_{XX}(f)$, comme la transformée de Fourier de sa fonction d'auto

corrélacion, c'est-à-dire :
$$S_{XX}(f) = \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-2\pi f \tau} d\tau \quad (2.30)$$

A l'inverse, la fonction d'auto corrélation est retrouvée par la transformée de Fourier inverse

de la fonction DSP.
$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XX}(f) e^{i2\pi f \tau} df \quad (2.31)$$

Les relations (2.30) et (2.31) sont connues sous le nom de théorème de **Wiener-Kintchine**.

On note que pour $\tau = 0$, et à partir de l'équation (2.31), une relation importante

s'établit :
$$R_X(0) = E[X^2] = \int_{-\infty}^{+\infty} S_X(f) df \quad (2.32)$$

Cette dernière inégalité, montre que $S_{XX}(f)$ peut être considérée comme une mesure de la distribution dans le domaine des fréquences, de la moyenne de la valeur au carrée $E[X^2(t)]$,

ou encore, elle décrit la décomposition fréquentielle de la puissance temporelle, d'où son nom de fonction de densité spectrale de puissance.

2.2 Cas des processus non stationnaires.

La structure statistique dépend dans ce cas de la position de t le long de l'axe des temps et non du retard $\tau = t_1 - t_2$ comme dans le cas des processus stationnaires. Par conséquent, les relations de **Wiener-Kintchine** sont remplacées par les relations suivantes :

$$S_{XX}(f_1, f_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} R_{XX}(t_1, t_2) e^{-i 2\pi(f_1 t_1 - f_2 t_2)} dt_1 dt_2 \quad (2.33)$$

$$R_{XX}(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S_{XX}(f_1, f_2) e^{i 2\pi(f_1 t_1 - f_2 t_2)} df_1 df_2 \quad (2.34)$$

Chapitre III :

Caractérisation des mouvements forts du sol dus aux séismes.

Introduction.

L'appréciation d'un séisme a, jusqu'à un passé récent, été associée à des évaluations de son caractère destructif. Si cette approche a permis des progrès certains pour la compréhension des phénomènes sismiques, l'aspect extrêmement subjectif et fluctuant de cette destructivité exclut pratiquement l'obtention de renseignements quantitatifs. Ainsi s'est développé, sous l'impulsion initiale d'ingénieurs, la sismologie des mouvements forts dont l'objet se définit comme étant la mesure et l'interprétation physique des mouvements du sol au voisinage d'une faille, dans une zone où ces mouvements mettent en danger la vie ou l'activité humaine.

En effet, les mouvements forts du sol produits par les tremblements de terre, sont généralement irréguliers, ceci est dû à la nature extrêmement variable des paramètres qui caractérisent ces mouvements. Ces paramètres traduisent l'importance relative des effets de l'intensité du séisme représenté ici par la magnitude, du trajet source-récepteur des ondes sismiques émises, de la nature du sol au voisinage du point d'enregistrement ainsi que celui de la topographie du relief.

L'étude des mouvements forts en un site donné est généralement menée à travers l'analyse d'un enregistrement accélérométrique. La forme temporelle et le contenu fréquentiel d'un accélérogramme obtenu en un site donné, sont les principales caractéristiques qui traduisent l'influence des effets énumérés ci-dessus.

Les accélérogrammes sismiques présentent un aspect imprévisible en termes de lieu, de date et d'intensité. De plus, ils se caractérisent par une forme aléatoire. Leur modélisation sous forme stochastique est proche des observations.

1. Caractérisation dans le domaine temporelle [\[1\]](#),[\[6\]](#).

Comme décrit initialement, un séisme est la conséquence d'un glissement le long d'une faille préexistante entre deux compartiments rocheux solidaires précédemment. Ce glissement ou cette rupture est provoquée par un soudain déséquilibre de l'état des contraintes dans la croûte terrestre. Les ébranlements ainsi produits, donnent naissance à des ondes sismiques qui se propagent de la source aux sites d'enregistrements. Les accélérogrammes sismiques sont caractérisés dans le domaine temporel par :

1.1 Caractère transitoire.

Le mouvement sismique en un point est la somme des contributions des trois principaux groupes d'ondes, "P" compression. "S" cisaillement et "L" ondes de surfaces incluant les ondes de Rayleigh, de Love; qui atteignent le site d'enregistrement à des instants différents.

On peut observer trois parties distinctes : *la première partie* est caractérisée par des amplitudes et des périodes relativement petites associées en général aux ondes de compression. Les ondes "P" qui évoluent à des vitesses supérieures à celles des ondes cisaillement "S", sont les premières à atteindre le site considéré formant ainsi la phase des premières arrivées; *la partie centrale* dite des "mouvements forts" est influencée par une brusque augmentation des amplitudes provoquée par l'arrivée des ondes de cisaillement dont les amplitudes sont plus importantes que celles des ondes "P". Cette deuxième partie subsiste jusqu'à la libération totale de l'énergie induite par le séisme. Il ne subsiste dans *la partie finale* que les vibrations des ondes de surface qui vont persister avec des amplitudes relativement faibles avant de s'amortir lorsque le sol du site d'enregistrement retrouve son état d'équilibre.

1.2 Durée de mouvement fort.

L'utilisation de la **Cumulative RMS function** dont l'expression, pour un enregistrement $x(t)$, est :

$$CRF = \sum_{i=1}^N \sqrt{\frac{x^2(t_i)}{N-1}} \quad (3.1)$$

Où $x(t_i)$ est la valeur de l'accélération au temps $t_i = i \cdot \Delta t$.

N est associée à la durée totale $T = N \Delta t$ de l'accélérogramme.

Nous permet de représenter le taux d'accroissement de l'énergie libérée par la rupture tectonique en fonction du temps. En effet, cette fonction permet de visualiser l'apport d'énergie à chaque instant par une augmentation de sa valeur, qui se traduit géométriquement par une pente positive de la courbe représentative de cette fonction (fig.7b). La tendance générale de la CRF est d'atteindre à un instant donné un pic et décroître par la suite, ce qui signifie qu'au delà de cet instant, la pente de la fonction CRF prend des valeurs constamment négatives et l'énergie arrivant au site d'enregistrement n'a aucun apport significatif Cette énergie ne peut être associée qu'aux ondes de surface et aux vibrations libres de la terre qui vont persister avec des amplitudes relativement faibles avant de s'amortir en fin du signal. La valeur finie de l'énergie tectonique confère un caractère transitoire aux enregistrements sismiques. Donc la durée de phase forte est définie par l'intervalle compris entre les deux pics des CRFs calculée dans les deux sens (de début à la fin de l'enregistrement et dans le sens contraire)

1.3 Contenus phasiques et enveloppe.

Longtemps ignorées, les phases se sont révélées déterminantes dans l'estimation temporelle des amplitudes caractérisant les vibrations de l'énergie libérée au site. En effet, la modélisation de l'enveloppe temporelle et l'accélération sismique est directement contrôlée par deux paramètres stochastiques : le temps moyen d'arrivées des ondes sismiques au site d'enregistrement t_m et l'écart type d associé à la durée du mouvement fort, à travers la fonction de densité de probabilité des dérivés de phases .

$$p(\phi') = \frac{d^2}{2[(\phi' - t_m)^2 + d^2]^{3/2}} \quad (3.2)$$

Où : $\phi' = \frac{d\phi(f)}{df}$ dérivée par rapport à la fréquence des phases.

$$\text{Où : } d^2 = \frac{\gamma_2}{\gamma_0} - \left(\frac{\gamma_1}{\gamma_0} \right)^2 \text{ écart-type associée à la durée du mouvement fort.} \quad (3.3)$$

$$t_m = \frac{\gamma_1}{\gamma_0} \quad \text{délai moyen de groupes.}$$

$$\gamma_i = \int_{-\infty}^{+\infty} t^i \alpha^2(t) dt \quad i=0, \dots, 2 \quad (3.4)$$

Est le moment statistique d'ordre i du carré de l'enveloppe $\alpha(t)$.

L'enveloppe temporelle $\alpha(t)$ est définie par l'expression suivante :

$$\alpha(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \quad (3.5)$$

Où : $x(t)$ est un signal réel et $\hat{x}(t)$ est la transformée de **Hilbert** du signal $x(t)$.

La fonction de réparation des délais de groupes permettant de générer le contenu phasique est donnée par ma relation suivante :

$$F(\phi') = \frac{1}{2} + \frac{t_m - \phi'}{\left[(\phi' - t_m)^2 + d^2 \right]^{3/2}} \quad (3.6)$$

Les figures (7c) à (7f) illustrent respectivement les variations de l'enveloppe temporelle, les valeurs des dérivées de phases la fonction de densité de probabilité et la fonction de répartition. Ces variations ont été obtenues avec les paramètres déduits de l'enveloppe temporelle du séisme de Loma-Prieta (figure 7a) enregistré le 17/10/89, soit soit $t_m = 8.25$ sec et $d = 4.27$ sec. Il ressort de cet exemple, que la fonction de densité de probabilité des dérivées de phases pour lme accélération sismique donnée a une forme de caractéristiques identiques à celles de l'enveloppe temporelle.

1.4 Intensité d'Arias.

L'intensité d'Arias c'est l'intégrale de carrée de l'accélération $I_A = \int x(t)^2 dt$ et elle définie encore comme l'intégrale dans le domaine temporelle de l'enveloppe temporelle est l'intégrale dans le plan temps-fréquence qui seront définis dans le paragraphe suivant :

$$I_A = \iint PS(t, f) dt df = \int \alpha(t) dt \quad (3.7)$$

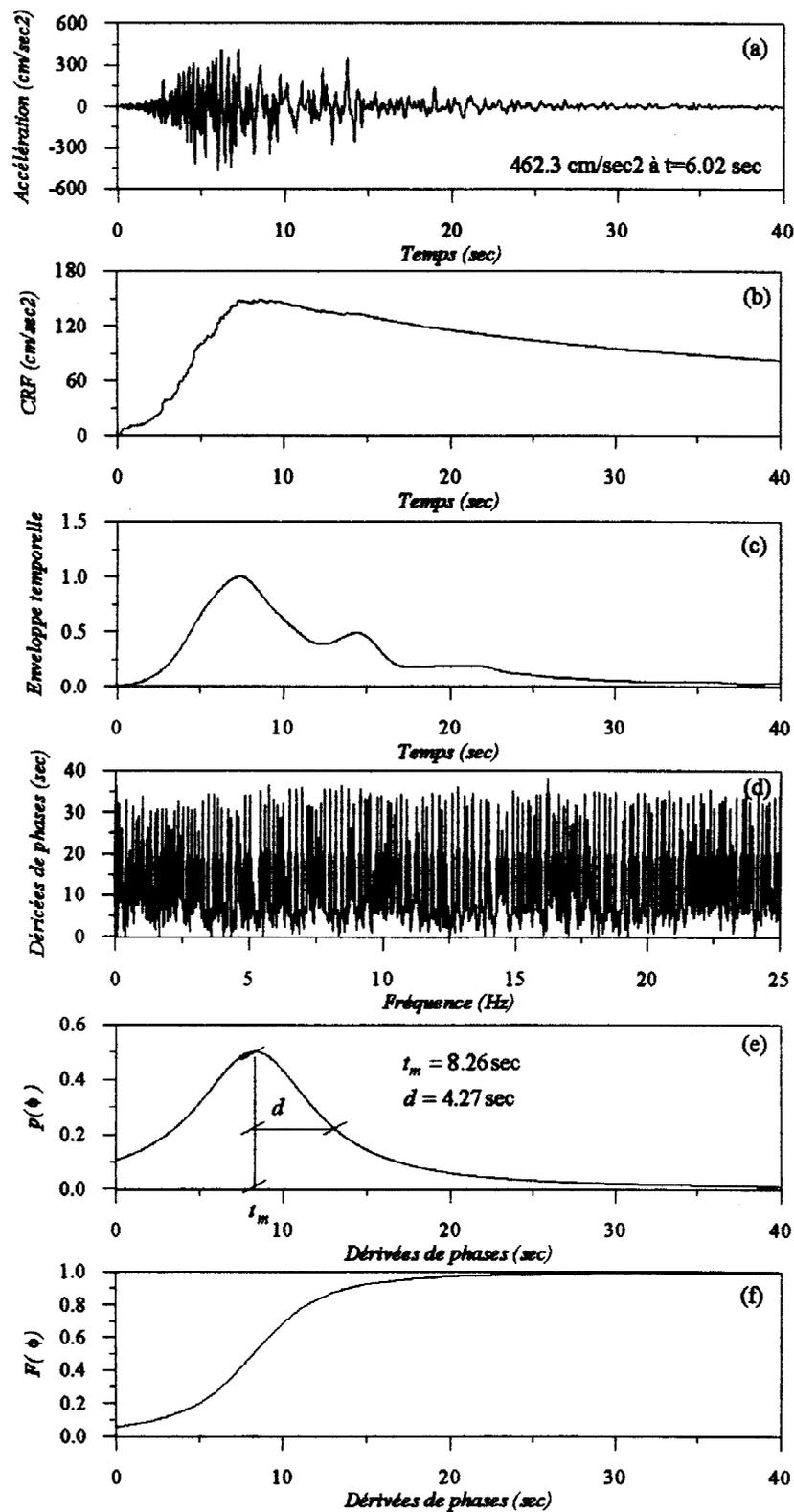


Figure 7 Caractérisation de l'enregistrement de Loma-Prieta dans le domaine temporel

- a) Enregistrement accélérométrique du séisme de Loma-Prieta.
- b) Fonction CRF
- c) Enveloppe temporelle.
- d) Valeurs des dérivées de phases.
- e) Fonction de densité de probabilité des dérivées de phases.
- f) Fonction de répartition des dérivées de phases.

2. Caractérisation dans le domaine fréquentiel [1],[6].

Les mouvements sismiques exhibent une nonstationnarité en amplitude et en contenu fréquentiel, la nonstationnarité en amplitude a été modélisée dans le paragraphe précédent par les paramètres stochastiques ; le temps moyen des délais de piques et la dispersion autour de ce temps à travers la fonction de densité de probabilité des dérivées de phases. Dans ce chapitre, nous développons les outils nécessaires pour la caractérisation nonstationnaire du contenu spectral des mouvements sismiques.

2.1 Représentation spectrale non stationnaire.

Il existe différentes approches de représentation spectrale non stationnaire, mais afin de sélectionner l'approche qui caractérise les paramètres que nous traitons dans ce travail, nous allons consacrer ce paragraphe à une analyse critique des principales approches pour la caractérisation non stationnaire des processus aléatoires. Par conséquent, nous abordons les principales techniques de caractérisation de la puissance spectrale d'un processus non stationnaire. Pour cela, considérons un filtre idéal de bande étroite, centré sur f_0 et de la largeur Δf , sa fonction de transfert est :

$$\begin{cases} H(\nu) = 1 & f_0 - \frac{\Delta f}{2} \leq |\nu| \leq f_0 + \frac{\Delta f}{2} \\ H(\nu) = 0 & \text{ailleurs} \end{cases} \quad (3.8)$$

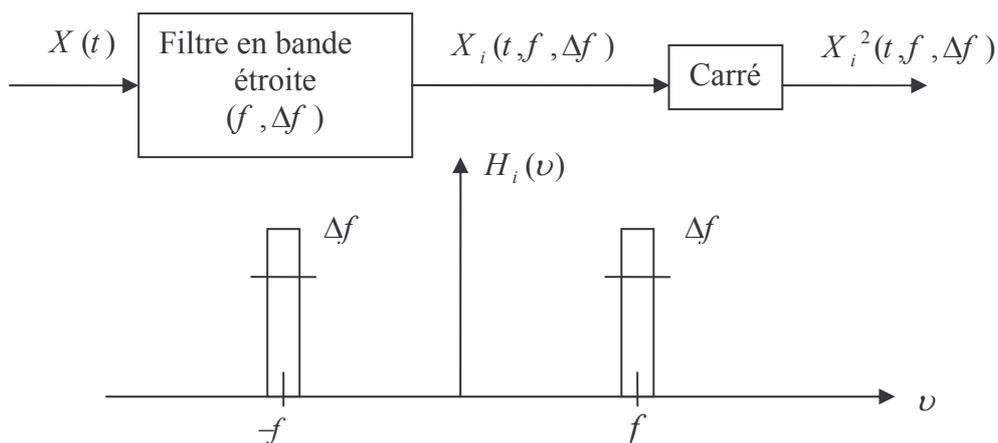


Figure 8 Estimation de la puissance spectrale d'un processus au départ du carré de la réponse d'un filtre en bande étroite.

Sa réponse impulsionnelle s'écrit :

$$h(\tau) = 2 \frac{\sin \pi \Delta f \tau}{\pi \tau} \cos 2\pi f \tau \quad (3.9)$$

Par application de la relation (2.32 cf. chapitre 2), on a

$$E \left[X_i^2(t, f_0, \Delta f) \right] = \int_{-f_0 - \Delta f / 2}^{-f_0 + \Delta f / 2} S_{XX}(f) df + \int_{f_0 - \Delta f / 2}^{f_0 + \Delta f / 2} S_{XX}(f) df \quad (3.10)$$

La relation (3.10) permet alors d'interpréter la densité spectrale de puissance comme une décomposition fréquentielle locale de la puissance du processus : la puissance moyenne dans tout intervalle de fréquence est déterminée par l'aire comprise sous la courbe de la densité de puissance spectrale entre les deux fréquences délimitant l'intervalle. Quoiqu'elle fournisse une décomposition fréquentielle de la puissance, la DSP ne nous donne aucune information au sujet de l'évolution temporelle du contenu fréquentiel du processus, d'où la nécessité de s'intéresser aux caractéristiques donnant ce type d'information.

Afin de traduire la dépendance temporelle du spectre de puissance d'un processus non stationnaire, on utilise le traitement précédant mais cette fois-ci en juxtaposant un certain nombre de filtres dans les bandes de fréquence choisies.

En utilisant l'équation (3.10) on obtient ainsi une estimation de la densité spectrale de puissance.

$$S_{XX}(t, f) = \frac{E \left[X_i^2(t, f_0, \Delta f) \right]}{2\Delta f} \quad (3.11)$$

La sortie $X_i^2(t, f_0, \Delta f)$ du filtre en bande étroite peut s'écrire sous la forme.

$$X_i(t, f_0, \Delta f) = \int_{-\infty}^{+\infty} h_i(\tau) X(t - \tau) d\tau \quad (3.12)$$

Elle représente donc une moyenne pondérée des valeurs du processus $X(t)$, la fonction de pondération $h(\tau)$ étant la réponse impulsionnelle du filtre, les tendances temporelles du processus de fréquences f_i , supérieures à $\frac{\Delta f}{2}$ seront lissées par l'opération du filtrage.

Les exigences contradictoires suivantes doivent guider le choix de Δf .

- Δf doit être petit pour avoir une bonne résolution.
- Δf doit être supérieur à f_i , valeur qui représente la plus haute fréquence des tendances non stationnaires, de manière à éviter le lissage de la non stationnarité par la convolution.

2.1.1 Densité spectrale de puissance instantanée.

Soit $R_{XX}(t_1, t_2)$, la fonction d'auto corrélation d'un processus non stationnaire, effectuons un changement de variables. $t = \frac{t_1 + t_2}{2}$; $\tau = t_2 - t_1$ (3.13)

t représente le temps moyenne et τ l'écart. Il suit de (3.13) que $t_1 = t - \frac{\tau}{2}$, $t_2 = t + \frac{\tau}{2}$.

Soit $R_{XX}(t, \tau)$ le résultat de cette transformation.

$$R_{XX}(t, \tau) = R_{XX}(t_1, t_2) = E \left[X \left(t - \frac{\tau}{2} \right) X \left(t + \frac{\tau}{2} \right) \right] \quad (3.14)$$

Le résultat $R_{XX}(t, \tau)$ à été utilisé par **Bendat** et **Priestol** pour définir, la fonction de densité spectrale instantanée $S_{XX}(t, f)$. Par analogie au cas stationnaire, la transformée de Fourier par rapport à τ de $R_{XX}(t, \tau)$ soit :

$$S_{XX}(t, f) = \int_{-\infty}^{+\infty} R_{XX}(t, \tau) e^{-2\pi f \tau} d\tau \quad (3.15)$$

Cette relation s'inverse sous la forme

$$R_{XX}(t, \tau) = \int_{-\infty}^{+\infty} S_{XX}(t, f) e^{-2\pi f \tau} df \quad (3.16)$$

Les relations (3.15) et (3.16) constituent le correspondant non stationnaire du théorème de **Wiener-Kintchine**. En faisant $\tau = 0$, il suit de (3.16) que

$$E[X^2(t)] = R_{XX}(t, 0) = \int_{-\infty}^{+\infty} S_{XX}(t, f) df \quad (3.17)$$

En intégrant (3.17) par rapport au temps, on obtient l'énergie moyenne de puissance

$$E_X = \int_{-\infty}^{+\infty} E[X^2(t)] dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S_{XX}(t, f) df dt \quad (3.18)$$

Cette dernière, nous donnée une idée intuitive de la répartition de l'énergie dans le domaine (t, f) . Néanmoins, du fait que $S_{XX}(t, f)$ peut prendre de valeurs négatives, la décomposition fréquentielle globale fournie par (3.17) ne peut être étendue à tout intervalle de fréquence comme dans le cas stationnaire.

2.1.2 Processus localement stationnaire.

Soit $X(t)$ un processus non stationnaire qui peut s'écrire sous la forme

$$X(t) = a(t)z(t) \quad (3.19)$$

Où $a(t)$ est une fonction déterministe qui varie en fonction du temps et $z(t)$ est un processus faiblement stationnaire, en effet

$$\begin{aligned} R_{XX}(t, \tau) &= E[X(t_1)X(t_2)] = E[a(t_1)z(t_1)a(t_2)z(t_2)] \\ &= a(t_1)a(t_2)E[z(t_1)z(t_2)] \\ &= a(t_1)a(t_2)R_X(t_2-t_1) \end{aligned} \quad (3.20)$$

Si les fluctuations de $a(t)$ sont très lentes comparées à celles du processus $z(t)$.

$$a(t_1)a(t_2) = a^2(t) \quad (3.21)$$

En effectuant un changement de variables $t = \frac{t_2+t_1}{2}$; $\tau = t_2-t_1$, il suit de (3.21) que :

$$R_{XX}(t, \tau) = R_{XX}(t_1, t_2) = a^2(t)R_X(\tau) \quad (3.22)$$

Ce résultat représente la fonction d'auto corrélation d'un processus localement stationnaire, largement utilisé dans la littérature. La relation (3.22) s'inverse sous la forme

$$S_{XX}(t, f) = a^2(t)S_{ZZ}(f) \quad (3.23)$$

La densité spectrale de puissance instantanée d'un processus localement stationnaire ne peut être négative.

2.1.3 Spectre évolutif [1,6].

Le spectre évolutif est une représentation harmonique plus générale, qui peut s'appliquer aux signaux s'annulant à l'infinie (cf. chapitre II), signaux périodiques et aux processus stationnaires.

$$X(t) = \int_{-\infty}^{+\infty} e^{i2\pi ft} dz(f) \quad (3.24)$$

Où $z(f)$ est une fonction uniquement déterminée par la forme de $X(t)$; En effet, si le processus $X(t)$ est stationnaire, le processus $z(f)$ est un processus à incréments orthogonaux ayant les propriétés suivantes.

$$\begin{aligned} E[dz(f)] &= 0 \\ E[dz(f_1)dz(f_2)] &= S_{XX}(f_1)\delta(f_1-f_2)df_1df_2 \end{aligned} \quad (3.25)$$

Où $\delta(\cdot)$ est la fonction de Dirac la densité spectrale de puissance $S_{XX}(f)$ de $X(t)$.

Quand le processus est non stationnaire, la représentation est encore valides mais le contenu fréquentiel varie en fonction du temps et que les amplitudes spectrales, pour deux incréments de fréquences f_1 et f_2 , sont corrélées entre elles, ce qui veut dire que $z(f)$ n'est

pas un processus orthogonal, et la décomposition précédente conduit à une représentation spectrale à deux variables fréquentielles n'ayant aucune interprétation énergétique simple.

La représentation harmonique qui permet de conserver l'interprétation du spectre en tant que répartition fréquentielle de l'énergie pour les processus non stationnaires s'écrit comme suit :

$$X(t) = \int_{-\infty}^{+\infty} a(f, t) e^{i2\pi ft} dz(f) \quad (3.26)$$

Où $z(f)$ est un processus à incréments orthogonaux et $a(f, t)$ représente une enveloppe modulant la fonction harmonique $e^{i2\pi ft}$, $a(f, t)$ doit être choisie de sorte que le module de la transformée de Fourier soit maximum à l'origine ($\theta = 0$), elle admette une représentation harmonique.

$$a(f, t) = \int_{-\infty}^{+\infty} e^{i2\pi ft} H(\theta, f) d\theta \quad (3.27)$$

Tenant compte des équations (3.25) et (3.26), l'expression de la fonction d'auto corrélation de $X(t)$, s'écrit :

$$\begin{aligned} R_{XX}(t_1, t_2) &= E[X(t_1)X^*(t_2)] \\ &= \iint a(f_1, t_1) a^*(f_2, t_2) e^{i2\pi(f_1 t_1 - f_2 t_2)} E[dz(f_1) dz^*(f_2)] \end{aligned} \quad (3.28)$$

Et, compte tenu de (3.27) et sachant que : $\int_{-\infty}^{+\infty} f(\tau) \delta(t - \tau) d\tau = f(t)$

$$R_{XX}(t, \tau) = \int_{-\infty}^{+\infty} a(f, t_1) a^*(f, t_2) e^{i2\pi f(t_1 - t_2)} S_X(f) df \quad (3.29)$$

Après avoir effectué un changement de variables.

$$R_{XX}(t, \tau) = \int_{-\infty}^{+\infty} |a(f, t)|^2 e^{i2\pi f t} S_X(f) df \quad (3.30)$$

Il suit que :

$$E[X^2(t)] = \int_{-\infty}^{+\infty} |a(f, t)|^2 S_X(f) df \quad (3.31)$$

Cette relation permet de définir le spectre évolutif comme étant la contribution de la fréquence f à la puissance instantanée.

$$S_X(f, t) = |a(f, t)|^2 S_X(f) \quad (3.32)$$

Le problème essentiel dans l'estimation de spectre évolutif, réside dans la détermination de l'enveloppe temporelle. Pour cela plusieurs procédures ont été utilisées et

qui sont basées sur les travaux de Priestley (1965), Lin (1970) et la technique du multi filtrage. Ces techniques engendrent des procédures numériques laborieuses pour atteindre une description approximative du contenu spectrale des processus non stationnaires. Du fait que le but de notre travail est d'introduire un spectre évolutionnaire qui doit avoir un sens physique et qui, si possible, devrait pouvoir être relié facilement à des paramètres mesurant le caractère destructif des séismes, l'approche que nous proposons est différente. Elle est décrite dans le chapitre suivant.

Chapitre IV :

L'approche de spectre physique

1. Introduction.

Dans ce chapitre nous définissons l'approche utilisée de spectre physique qui est défini pour un processus aléatoire $X(u)$ comme étant une représentation physique de son énergie totale moyenne dans le plan temps fréquence ; c'est une extension de la notion de DSP aux processus non stationnaires.

Le spectre physique est une décomposition en fonction du temps et de la fréquence de l'énergie libérée par le séisme. L'approche est proposée par Mark.

2. Spectre physique [1,6].

Le spectre physique est défini comme la moyenne sur échantillon du spectre de Fourier de processus $W(t-u)X(u)$.

$$S_X(f, t; W) = \frac{1}{2\pi} E \left[\left| \int_{-\infty}^{+\infty} W(t-u)X(u)e^{-i2\pi fu} du \right|^2 \right] \quad (4.1)$$

Où $X(u)$ est un processus aléatoire réel dont nous isolons la portion alentour de $u = t$, $W(t)$ est une fonction fenêtre réelle jouissant des propriétés suivantes :

- $\int_{-\infty}^{+\infty} W^2(t) dt = 1$ (4.2)
- $W(t)$ est positive au voisinage de $t=0$.
- $|W(t)|$ est petit ai dehors de ce voisinage.

L'équation (4.1) montre que le calcul du spectre physique conduit, dans le domaine spectral, à une convolution des spectres de l'enregistrement $X(t)$ avec celui de le fenêtre $W(t)$, donc implicitement à un lissage du spectre $x(f)$ du processus $X(t)$. Cela montre que le choix de la forme de le fenêtre temporelle influe grandement sur la stabilité des estimations spectrales, mais pour chaque choix de $W(t)$, $S_X(f, t; W) = S_X(-f, t; W)$.

L'application du théorème du **Parseval** au processus $W(t-u)X(u)$ conduit à une décomposition fréquentielle de la moyenne pondérée locale au voisinage de $u=t$, de la puissance instantanée $E[X^2(u)]$. La fonction de pondération étant $W^2(t-u)$, la longueur de l'intervalle ainsi que la pondération étant fixée par la longueur et la forme de $W(t-u)$.

$$\int_{-\infty}^{+\infty} W^2(t-u)E[X^2(u)]du = \int_{-\infty}^{+\infty} S_X(f, t; W)df \quad (4.3)$$

L'intégration par rapport au temps de l'équation (4.4), compte tenu de (4.2), conduit à :

$$E_X = \int_{-\infty}^{+\infty} E[X^2(t)]dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S_X(f, t; W)dfdt \quad (4.5)$$

Le volume compris sous la surface $S_X(f, t; W)$ représente l'énergie totale moyenne du processus $X(t)$, quel que soit le choix de $W(t)$.

Si l'on désigne par $w(f)$ la transformée de Fourier de $W(t)$, alors sa normalisation sera obtenue compte tenu (4.2) et par application du théorème de **Parseval**.

$$\int_{-\infty}^{+\infty} w^2(f)df = 1 \quad (4.6)$$

En tenant compte du théorème de convolution et de la condition de normalisation (4.6), on obtient, pour la relation (4.3) une forme duale dans le domaine fréquentiel.

$$\int_{-\infty}^{+\infty} |w(f-\nu)|^2 d\nu = \int_{-\infty}^{+\infty} S_X(f, t; W)dt \quad (4.7)$$

Où $x(\nu)$ est la transformée de Fourier de $X(t)$.

La relation (4.7) conduit à une décomposition temporelle de la moyenne pondérée locale au voisinage de $\nu=f$, de l'espérance du spectre énergétique $E[|x(\nu)|^2]$, la fonction de pondération étant $|w(f-\nu)|^2$, la largeur et la pondération étant fixées respectivement par la largeur de la forme de $w(f-\nu)$.

Par le biais, des équations (4.3) et (4.7), on voit que, le calcul de la valeur locale $S_X(f, t; W)$ est contrôlé par la durée nominale T de $W(t)$ dans le domaine temporel et la largeur nominale β de $w(f)$ dans le domaine spectral. Il apparaît que la résolution qu'il est possible d'atteindre dans les deux domaines n'est pas libre, le choix de la résolution dans un domaine (temps ou fréquence) fixe automatiquement, la résolution dans l'autre domaine. Cette notion de compromis entre la résolution temporelle et la résolution fréquentielle est connue sous le nom de principe d'incertitude.

La caractérisation des processus non stationnaires à l'aide du spectre physique, conduit à des procédures numériques relativement simples comparées à celles des précédents concepts. A l'issue de cette définition, nous portons notre choix sur le concept du spectre physique, qui nous semble l'outil le plus simple pour la description du caractère évolutif des processus non stationnaires.

3. Présentation des paramètres du Modèle (spectre physique)[\[6\]](#).

Pour la suite de ce travail, l'évolution temporelle du contenu spectral est décrite en utilisant le concept du spectre physique à travers la définition de trois paramètres simples dérivant de la théorie des moments spectraux (Vanmarke, 1980 ; Lai, 1982) ; ces dernières sont associés aux valeurs de ce spectre au temps t .

Paramètres fréquentiels.

- Le premier paramètre, que nous noterons **fréquence moyenne** $f^0(t)$, indique l'endroit où est concentrée la masse spectrale le long de l'axe des fréquences (centre de gravité du spectre physique).
- Le deuxième, que nous dénommerons $f^s(t)$ **dispersion**, mesure la variabilité du contenu fréquentiel, c'est à dire la largeur de bande de la fonction de densité spectrale autour de sa fréquence moyenne.

Paramètres énergétique $\alpha(t)$.

- Le troisième $\alpha(t)$ est **l'enveloppe temporelle** qui traduit la variation en fonction du temps, des enveloppes des ondes constitutives.

Afin d'approximer à chaque instant t et pour chaque fréquence f les valeurs du spectre physique nous avons choisi une forme Gaussienne dont l'expression est la suivante.

$$S(f, t) = \frac{\alpha(t)}{\sqrt{2\pi}f^s(t)} \exp\left[-0.5\left(\frac{f - f^0(t)}{f^s(t)}\right)^2\right] \quad (4.8)$$

Pour le calcul du paramètre $\alpha(t)$, on utilise l'équation suivante :

$$\alpha(t) = \sqrt{2\pi}f^s(t) \left[\int_{-\infty}^{+\infty} s_{xx}(f, t) df / \int_{-\infty}^{+\infty} \exp\left[0.5\left(\frac{f - f^0(t)}{f^s(t)}\right)^2\right] df \right] \quad (4.9)$$

Nous présentons sur la figure 9a, le spectre physique de l'enregistrement accélérométrique du séisme de Loma-Prieta. On voit que les bandes fréquentielles occupées par les spectres à différents instants, répartis le long de la durée de l'accélérogramme,

deviennent de plus en plus étroites au fur et à mesure que l'on se rapproche de la fin du signal. En effet la bande spectrale est de [0 Hz-12 Hz] dans l'intervalle de temps [0 sec-10 sec], [0 Hz- 8 Hz] entre 10 sec et 20 sec et qui devient [0 Hz-3 Hz] au-delà de 20 sec.

En utilisant l'approximation donnée par l'équation (3.65) nous montrons sur la figure 9b le spectre physique modélisé. Afin de mieux voir l'évolution de l'énergie dans le plan temps-fréquence, nous présentons sur la figure 9c le spectre physique pour t inférieure à 30 sec et inférieure à 10 Hz.

L'évolution temporelle des paramètres spectraux et énergétique définissant le spectre physique de l'enregistrement accélérométrique du séisme de Loma-Prieta est décrite par les figures 10a à 10c. La figure 10a illustre la nette décroissance de la fréquence moyenne au fur et à mesure que t augmente, cette fréquence varie entre les valeurs de 5Hz et 0Hz.

En ce qui concerne les variations de la dispersion autour de la fréquence moyenne, l'analyse de la figure 10b montre une chute progressive de 2.5 Hz vers 1Hz entre $t=0$ sec et $t=2.5$ sec, ensuite un palier de moyenne 1Hz au-delà de $t=15$ sec.

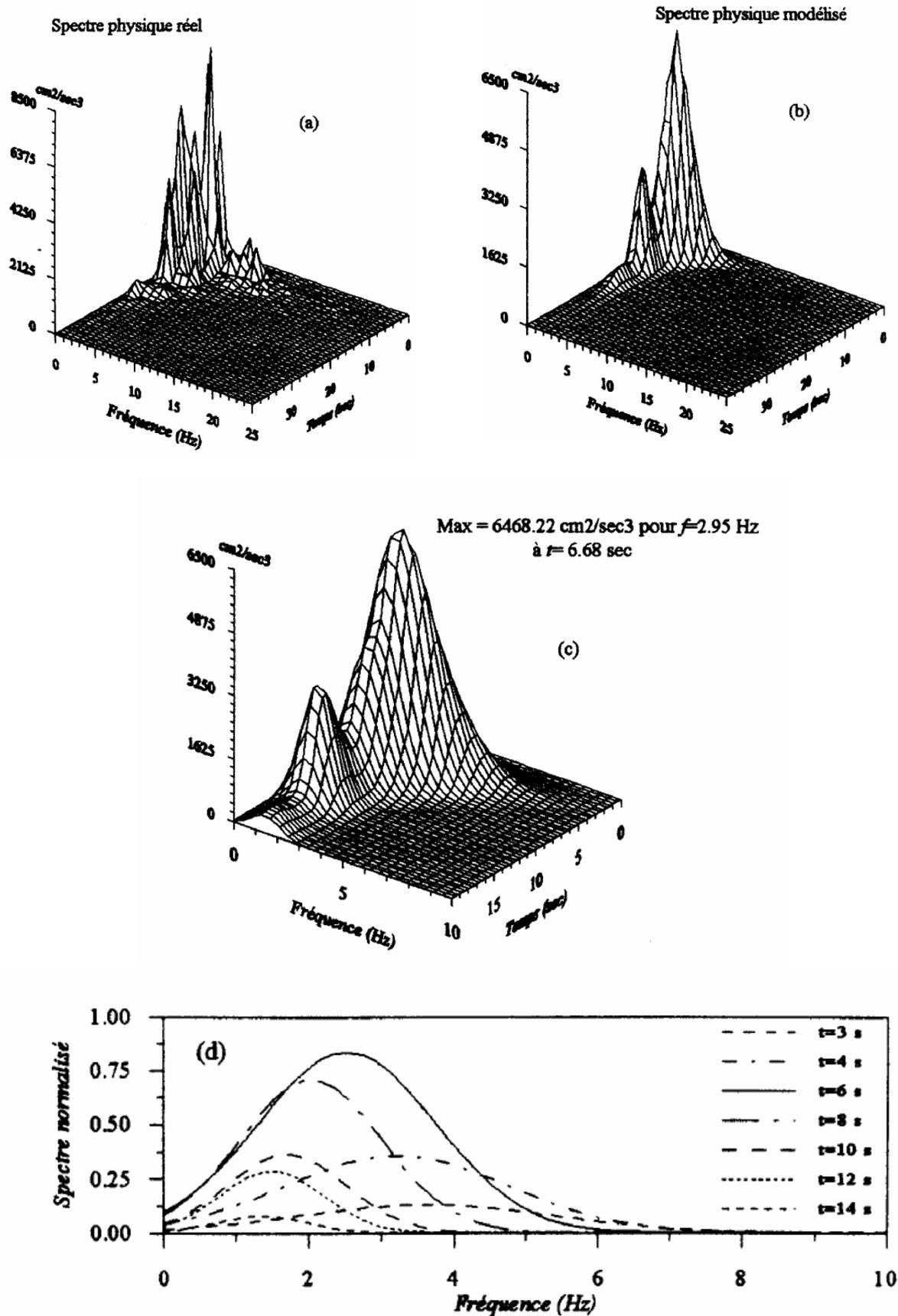


Figure 9 Évolution temporelle du contenu fréquentiel du séisme Loma Prieta

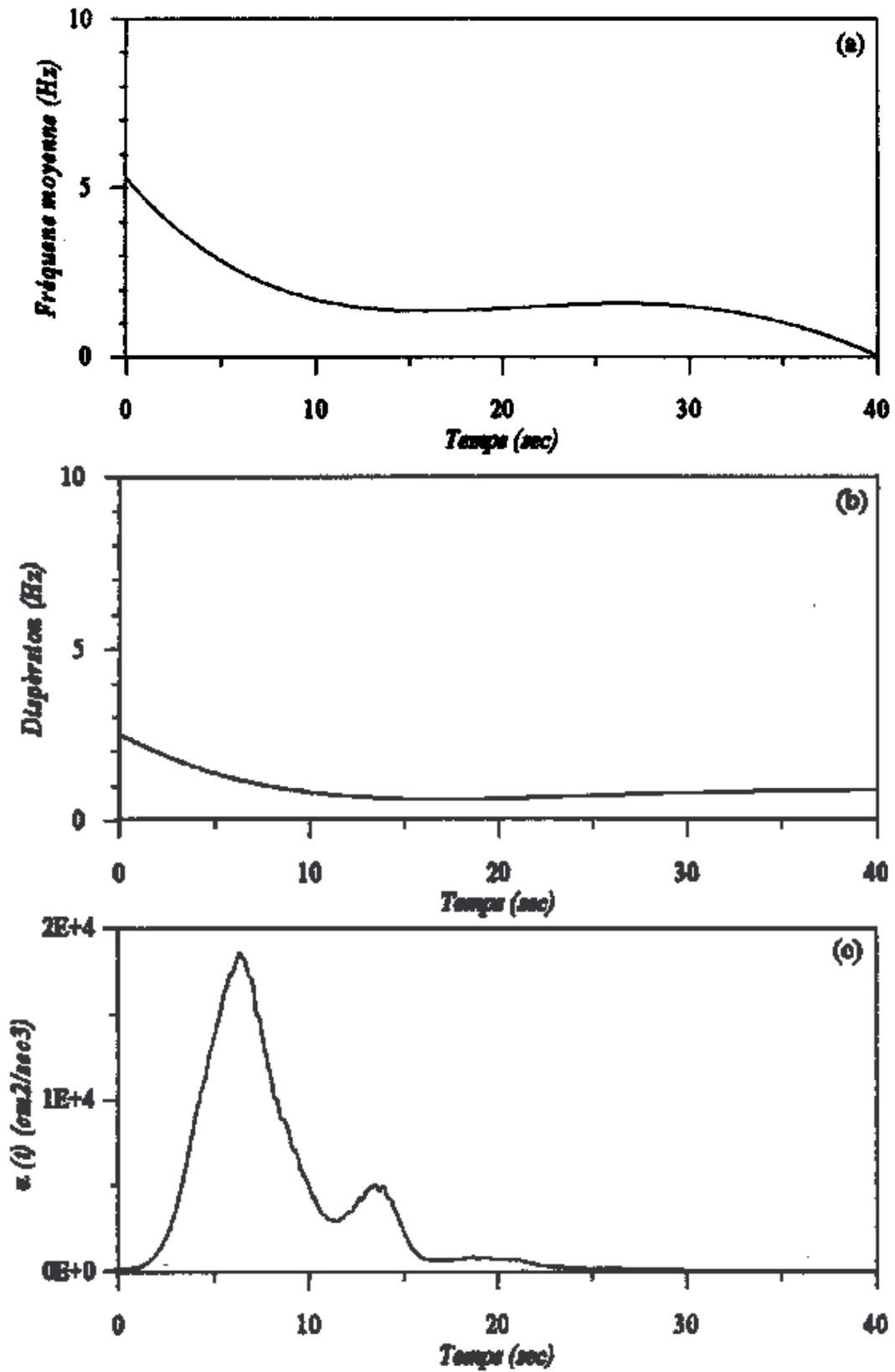


Figure 10 Caractérisation de séisme de Loma preita dans le domaine fréquentiel

Chapitre V

Notions sur la technique de régression.

Introduction.

Dans le présent chapitre on va cerner un nombre de notions et concepts nécessaire pour l'étude d'une analyse de régression,

1. Estimation des paramètres inconnus d'une loi de répartition [15].

Dans le traitement des résultats des expériences, il s'agit essentiellement de trouver les lois de répartition de grandeurs aléatoires d'après les résultats des expériences. Pour pouvoir déterminer la loi de répartition, il faut disposer de données statistiques suffisantes, plusieurs centaines au moins. Cependant dans la pratique, on a affaire à un nombre de données statistique bien plus réduit : deux ou trois dizaines d'observations, souvent même mois. Cet état de choses qu'explique par le prix élevé et par la complexité des expériences. Des données aussi limiter sont nettement insuffisantes pour retrouver la loi de répartition inconnue du variable aléatoire ; néanmoins ces données peuvent être traitées est utilisées pour obtenir certains renseignements sur la variable aléatoire étudiée. Par exemple, on peut, à partir de données statistiques limitées, déterminer, ne serait-ce qu'approximativement, des caractéristiques numériques importantes d'une variable aléatoire, à savoir l'espérance mathématique, la variance et parfois les moments d'ordre plus élevé. Dans la pratique ont connaît souvent à l'avance la forme de la loi de répartition et il ne reste qu'à déterminer certains paramètres de cette dernière. Si, par exemple, on sait qu'une variable aléatoire est normale, le traitement se réduit à la détermination de deux paramètres de cette loi : m et σ . Si l'on sait que la variable aléatoire étudiée est répartie suivant la loi de Poisson, un seul paramètre à déterminer, à savoir l'espérance mathématique a . Enfin, dans certains problèmes la forme de la loi de répartition importe peu, seules les caractéristiques numériques doivent être déterminées.

Avant tout il faut noter que la valeur d'un paramètre inconnu, calculée à partir d'un nombre restreint de données, aura toujours un caractère aléatoire. Nous appellerons *estimation* d'un paramètre cette valeur aléatoire approchée.

Soit une variable aléatoire X dont la loi de répartition contient un paramètre inconnu a . il s'agit, dans le cas général, de trouver une estimation convenable pour le paramètre a d'après les résultats de n expériences indépendantes, dans chacune desquelles la variable X a pris une valeur déterminée.

Désignons les valeurs observées de X par X_1, X_2, \dots, X_n (5.1)

Celles-ci peuvent être considérées comme n échantillons de la variable aléatoire X , c'est-à-dire n variables aléatoires indépendantes, réparties chacune suivant la même loi que X . Désignons par \tilde{a} l'estimation du paramètre a . toute estimation calculée à partir de (5.1) doit être une fonction des variables X_1, X_2, \dots, X_n :

$$\tilde{a} = \tilde{a}(X_1, X_2, \dots, X_n) \quad (5.2)$$

Et par conséquent être elle-même une variable aléatoire. La loi de répartition de \tilde{a} dépend de celle de X (et, en particulier, du paramètre inconnu a) ainsi que du nombre d'expérience n . en principe les méthodes de la théorie de probabilités permettent de trouver cette loi de répartition.

Imposons à l'estimation \tilde{a} des exigences, auxquelles elle doit satisfaire pour être considérée « bonne » en certain sens.

Il est tout naturel d'exiger que l'estimation \tilde{a} tende (converge en probabilité) vers le paramètre a lorsque le nombre d'expériences augmente. Une estimation douée de cette propriété est dite *consistante*.

Ensuite, il est désirable qu'en utilisant \tilde{a} au lieu a , on ne commette pas au moins d'erreur systématique augmentant ou diminuant l'estimation obtenue c'est-à-dire que l'on ait :

$$E[\tilde{a}] = a \quad (5.3)$$

Une estimation satisfaisante à cette condition est dite *non biaisée*.

Il est enfin désirable que l'estimation centrée choisie possède la variance minimale, c'est-à-dire

$$Var[\tilde{a}] = \min \quad (5.4)$$

Une estimation douée de cette propriété est dite *effective*.

Souvent il est difficile de satisfaire à toutes ces conditions. Par exemple, il peut se faire qu'une estimation effective existe, mais les formules de calcul sont tellement compliquées, qu'on est obligé de se contenter d'une autre estimation, dont la variance est un peu plus grande. Parfois on admet, pour simplifier les calculs, des estimations légèrement

biaisées. En tout cas, le choix d'une estimation doit être précédé d'un examen critique des possibilités.

1.1 Estimations de l'espérance mathématique et de la variance.

Soit une variable aléatoire X d'espérance mathématique m et de variance D inconnues. Les n expériences indépendantes en ont fourni les valeurs X_1, X_2, \dots, X_n . Il y a lieu de trouver les estimations consistantes et non biaisées des paramètres m et D .

Pour estimer ces paramètres il y a lieu d'utiliser les valeurs approchées (estimations) suivantes [16] :

$$\left. \begin{aligned} \tilde{m} &= \frac{\sum_{i=1}^n X_i}{n} \\ \tilde{D} &= \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n-1} \end{aligned} \right\} \quad (5.5)$$

2. Intervalle de confiance. Probabilité de confiance [15].

Dans un certain nombre de problèmes il ne suffit pas de trouver pour le paramètre une valeur numérique convenable, mais il faut estimer également sa précision et sa fiabilité. On veut savoir quelles sont les erreurs qu'entraîne le remplacement du paramètre a par son estimation ponctuelle \tilde{a} et en quelle mesure on peut être sûr que ces erreurs se trouveront dans certaines limites. Ces problèmes sont particulièrement importants lorsque le nombre de données est insuffisant. Dans ce cas l'estimation ponctuelle \tilde{a} est nettement aléatoire et des erreurs importants peuvent s'introduire lorsque l'on remplace a par \tilde{a} .

Pour évaluer la précision et la fiabilité d'une estimation, en mathématiques statistiques on utilise les notions d'*intervalle de confiance* et de *probabilité de confiance*.

Supposons que pour le paramètre a on ait expérimentalement obtenu l'estimation non biaisée \tilde{a} . Estimons maintenant l'erreur possible. A cet effet choisissons une probabilité assez grande β (par exemple $\beta = 0.9, 0.95$ ou 0.99) telle qu'un événement, avec la probabilité β , puisse être considéré comme certain et déterminons la valeur ε pour laquelle on a :

$$P(|\tilde{a} - a| < \varepsilon) = \beta$$

La gamme des valeurs possibles de l'erreur apparaissant lorsque l'on substitue \tilde{a} à a sera alors $\pm \varepsilon$; des erreurs importantes en valeur absolue apparaîtront seulement avec petite probabilité $\alpha = 1 - \beta$.

$$\text{Écrivons comme suit : } P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \beta \quad (5.6)$$

L'égalité (5.6) signifie qu'avec une probabilité égale à β la valeur inconnue du paramètre a se trouvera dans l'intervalle : $I_\beta = (\tilde{a} - \varepsilon; \tilde{a} + \varepsilon)$ (5.7)

Soulignons une particularité de l'intervalle de confiance. On emploie souvent la probabilité pour une variable aléatoire de se trouver dans un intervalle déterminé. Ici la grandeur a n'est pas aléatoire, par contre l'intervalle I_β est aléatoire sa position sur l'axe des abscisses, déterminée par le centre \tilde{a} , ainsi que la longueur 2ε de cet intervalle, car la grandeur ε se calcule en général d'après les données expérimentales. C'est pourquoi, dans le cas présent, il est préférable d'interpréter la grandeur β non pas comme la probabilité pour l'intervalle I_β de recouvrir le point a figure 11.

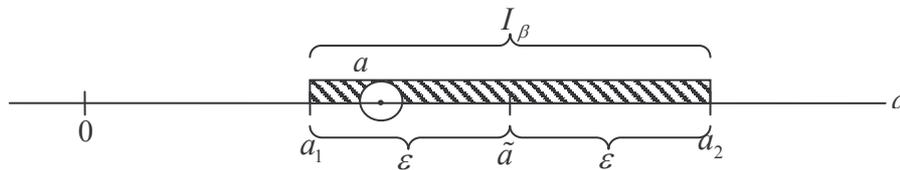


Figure 11 Intervalle de confiance

La probabilité β est généralement appelée *probabilité de confiance* et l'intervalle I_β (*intervalle de confiance*¹). Les limites $a_1 = \tilde{a} - \varepsilon$ et $a_2 = \tilde{a} + \varepsilon$, de l'intervalle I_β , sont appelées *limites de confiance*.

On peut interpréter l'intervalle de confiance comme l'intervalle des valeurs du paramètre a compatibles avec les données expérimentales et *ne les contredisant pas*. En effet, si l'on considère un événement pratiquement impossible avec une probabilité $\alpha = 1 - \beta$, les valeurs du paramètre a pour lesquelles $|\tilde{a} - a| > \varepsilon$ sont alors en contradiction avec les données expérimentales, et celles pour lesquelles $|\tilde{a} - a| < \varepsilon$ sont compatibles avec elles.

Calculons des limites de confiance a_1 et a_2 .

Supposons que pour le paramètre a ait une estimation non biaisée \tilde{a} . Si l'on connaissait la loi de répartition de la variable \tilde{a} , il serait très simple de trouver l'intervalle de confiance, il suffirait de trouver une valeur ε telle que : $P(|\tilde{a} - a| < \varepsilon) = \beta$.

¹ Sur la figure 11 l'intervalle de confiance est symétrique par rapport à \tilde{a} , comme nous allons le voir, ceci n'est pas obligatoire.

Ce qui complique le problème c'est que la loi de l'estimation \tilde{a} dépend de loi de X et par conséquent de ses paramètres inconnus (en particulier, du paramètre a lui-même).

3. Test d'hypothèse [4].

3.1 Principe de test d'hypothèse.

Les tests d'hypothèse constituent un aspect important de l'inférence (induction) statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit : Soit une population dont les éléments possèdent un caractère (mesurable ou dénombrable) et dont la valeur du paramètre, relative au caractère étudié, est inconnue. Une hypothèse est formulée sur la valeur du paramètre ; cette formulation résulte de considérations théoriques, pratiques ou encore elle est simplement basée sur un pressentiment. On veut porter un jugement sur cette hypothèse, sur la base des résultats d'un échantillon prélevé de cette population.

Il est bien évident que la statistique (variable d'échantillonnage) servant d'estimation au paramètre de la population ne prendra pas une valeur rigoureusement égale à la valeur théorique proposée dans l'hypothèse ; elle comporte des fluctuations d'échantillonnage qui sont régies par des distributions connues. Pour décider si l'hypothèse formulée est supportée ou non par les observations, il faut une méthode qui permettra de conclure si l'écart observé entre la valeur de la statistique obtenue de l'échantillon et uniquement imputable au hasard de l'échantillonnage. La construction d'un test d'hypothèse consiste effectivement à déterminer entre quelles valeurs peut varier la statistique (ou l'écart réduit), en supposant l'hypothèse vraie, sur la seule considération de hasard de l'échantillonnage. Les distributions d'échantillonnage d'une moyenne et vont être utiles dans l'élaboration d'un test statistique.

3.2 Concepts importants dans l'élaboration d'un test hypothèse.

Définissons d'abord certains concepts :

Hypothèse statistique

Une hypothèse statistique est un énoncé (une affirmation) concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population.

Test hypothèse

un test d'hypothèse (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques.

Les hypothèses statistiques qui sont envisagées a priori s'appelle l'hypothèse nulle et l'hypothèse alternative.

Hypothèse nulle (H_0) et hypothèse alternative (H_1)

L'hypothèse selon laquelle on fixe a priori un paramètre de la population à une valeur particulière s'appelle l'hypothèse nulle et est notée H_0 . N'importe quelle autre hypothèse diffère de l'hypothèse s'appelle l'hypothèse alternative (ou contre-hypothèse) et notée H_1 .

Un des aspects importants d'un test hypothèse est de convenir d'avance (avant le prélèvement de l'échantillon dans la population) à quelles conditions l'une ou l'autre des hypothèses sera considéré comme vraisemblable.

C'est l'hypothèse nulle qui est soumise au test et toute la démarche du test s'effectue en considérant cette hypothèse comme vraie. Si le test conduit, d'après les résultats de l'échantillon, au rejet de l'hypothèse nulle (elle est alors dépourvue du soutien expérimentale), nous considérons alors l'hypothèse alternative H_1 comme vraisemblable plutôt que H_0 .

Remarque. L'hypothèse nulle peut aussi affirmer que la différence entre les valeurs de deux paramètres et zéro ou affirmer que la distribution théorique des observations d'une population à une forme particulière.

La majorité des tests hypothèses que nous allons traiter vont s'effectuer à l'aide de la distribution d'échantillonnage de la statistique qui sert d'estimateur au paramètre précisé dans l'hypothèse nulle, il faut être en mesure d'établir des règles de décision qui vont nous conduire sans équivoque au non rejet de H_0 (ou au rejet). Toutefois, la décision de favoriser l'hypothèse nulle (ou l'hypothèse alternative) est basé sur une information partielle, les résultats d'un échantillon. Il est statistiquement impossible de prendre toujours la bonne décision. En pratique, ce que l'on peut faire, c'est de mettre en oeuvre une démarche qui nous permettrait, à long terme, de rejeter à tort une hypothèse nulle vraie dans une faible proportion de cas. La conclusion qui sera déduite des résultats de l'échantillon suivant la règle de décision qu'on aura adoptée, aura un caractère probabiliste ; on ne pourra prendre une décision qu'en prenant conscience qu'il y a un certain risque qu'elle soit erronée. ce risque nous est donné par le seuil de signification du test.

Seuil de signification d'un test d'hypothèse

Le risque, consenti à l'avance et que nous notons α , de rejeter à tort l'hypothèse nulle H_0 alors qu'elle est vraie (et de favoriser alors l'hypothèse alternative H_1) s'appelle le seuil de signification du test et s'énonce en probabilité comme suit :

$$\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ vraie}) = P(\text{choisir } H_1 \mid H_0 \text{ vraie})$$

A ce seuil de signification, on fait correspondre sur la distribution d'échantillonnage de la statistique (ou sur celle de l'écart réduit) une région de rejet de l'hypothèse nulle (appelée également région critique). L'aire de cette région correspond à la probabilité α . Cette région de rejet de H_0 est constituée d'un ensemble de valeurs de la statistique qui conduiront au rejet de H_0 . Si par exemple, on prend comme seuil de signification $\alpha = 0,05$, cela signifie que l'on peut admettre d'avance que la statistique (la variable d'échantillonnage) peut prendre, dans 5% des cas, une valeur vraie et ceci uniquement d'après le hasard de l'échantillonnage.

Sur la distribution d'échantillonnage correspondra aussi une région complémentaire, dite région de non rejet de H_0 (appelée également région d'acceptation) de probabilité $1 - \alpha$. La valeur observée de la statistique (ou de l'écart réduit) déduite des résultats de l'échantillon appartient, soit à la région de rejet de H_0 on favorisera alors l'hypothèse H_1 , soit à la région de non rejet de H_0 on favorisera alors l'hypothèse H_0 .

Remarque.

- a) les seuils de signification les plus utilisés sont $\alpha = 0.05$ et $\alpha = 0.01$, dépendant des conséquences de rejeter à tort l'hypothèse H_0 .
- b) La statistique qui convient pour le test est donc une variable aléatoire dont la valeur observée sera utilisée pour décider du « rejet » ou du « non-rejet » de H_0 . La distribution d'échantillonnage de cette statistique est déterminée en supposant que l'hypothèse H_0 est vraie.

3.3 Comment exécuter un test d'hypothèse : démarche à suivre.

Un test d'hypothèse comporte diverses étapes. Nous adopterons qu'en convention de travailler avec l'écart réduit (variable aléatoire dont on connaît les valeurs tabulée de la loi) qui sera distribué, suivant le cas, selon la loi normale centrée réduite ou la loi de *Student*. Au seuil de signification alpha choisi, on obtiendra directement des tables correspondantes les valeurs critiques de l'écart réduit. D'une façon générale, l'écart réduit s'exprime en unités d'écart-type de la statistique qui convient au test :

$$\text{Écart réduit} = \frac{\text{Ecart entre la statistique qui convient pour le test et la valeur du paramètre posée en } H_0}{\text{Ecart-type de la statistique}}$$

Remarque. La statistique (appelée également variable d'échantillonnage) qui est appropriée à un test d'hypothèse correspondant à une estimation non biaisée du paramètre qui est précisé dans l'hypothèse nulle. Elle sera, selon le contexte, la moyenne d'échantillon, le quotient de deux variances

3.4 Démarche à suivre dans l'élaboration d'un test d'hypothèse.

1. Formuler l'hypothèse nulle H_0 et hypothèse alternative H_1 .
2. Fixé d'avance (avant la réalisation du sondage) le seuil de signification alpha c'est-à-dire spécifier le risque de rejeter à tort une hypothèse H_0 vraie
3. préciser les conditions d'application du test. Spécifications ou non de la forme de la population échantillonnée, indication si nous sommes en présence d'un grand échantillon, si la variance de la population est connue ou inconnue, etc.
4. spécifier la statistique qui convient pour le test et définir l'écart réduit. En déduire sa distribution d'après les conditions d'application.
5. Adopter une règle de décision qui conduira au rejet ou non rejet de H_0 au seuil alpha choisi. Cette règle de décision est définie à partir des valeurs critiques de l'écart réduit.
6. Calculer la valeur numérique de l'écart réduit, valeur déduite des résultats de l'échantillon.
7. Décision est conclusion. Comparer la valeur numérique obtenue pour l'écart réduit avec la règle de décision adoptée en 5. Décidé entre les deux hypothèses formulées en 1. et conclure.

4. Régression linéaire Simple [4].

4.1 Introduction.

Dans le processus d'expérimentation technologique ou comme support à l'aide à la décision ou dans le cadre de la décision administrative, il arrive fréquemment que la conclusion et recommandation soient basées sur l'existence d'une liaison d'ordre fonctionnel ou statistique entre deux ou plusieurs variables.

Dans un cas comme dans l'autre, l'objectif est de déterminer d'abord s'il existe une relation statistique entre deux variables. Ce type d'analyse s'appelle **analyse de régression** et permet, entre autre, d'obtenir des estimations de valeurs moyennes ou des prévisions de valeurs individuelles d'une variation à partir des valeurs d'une autre variable.

4.2 Le modèle de la régression simple.

Dans diverses situations, un des objectifs est d'établir la relation statistique la mieux appropriée entre les variables concernées.

Étant donné n couples d'observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ et en supposant que la relation plausible entre X et Y est linéaire et d'ordre un, alors le modèle de régression linéaire simple s'écrit :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Où Y est la variable dépendante (ou expliquée) ayant un caractère aléatoire et dont les valeurs sont conditionnées par celle de la variable explicative X et la composante aléatoire ε ; Y_i représente la i -ème observation de Y .

β_0 et β_1 sont les paramètres du modèle de régression. X est la variable explicative (indépendante), mesurée sans erreurs ou dont les valeurs sont fixées avant expérience à des valeurs arbitraires. On la considère comme une grandeur certaine.

ε Dénote la fluctuation aléatoire non observable attribuable à un ensemble de facteurs ou de variable non pris en considération dans le modèle. Cette fluctuation aléatoire n'est pas expliquée par le modèle et se reflète la variable dépendante.

4.3 Condition d'application du modèle linéaire simple.

Le modèle simple s'appuie sur les hypothèses suivantes :

1. la courbe joignant les moyennes des distributions des Y_i pour les divers valeurs est appelée **courbe ou équation de régression**.
Lorsque ces moyennes sont alignées, la courbe de régression est alors une droite. Dans ce cas l'équation de régression est de la forme $E(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$.
2. la variance σ^2 de chaque distribution des Y_i est la même, peu importe les valeurs particulières X_i prises par la variable explicative x , c'est-à-dire $\text{var}(y_i) = \sigma^2$ pour toute les valeurs x .
3. les Y_i ne sont pas corrélé entre eux c'est-à-dire les observations de la variable indépendante ne sont pas liées avec les précédentes ni avec les suivantes.
4. pour chaque valeur X_i , les valeurs de la variable dépendante y réparties autour de l'équation de régression $E(Y_i) = \beta_0 + \beta_1 X_i$ sont distribuées d'après un loi normale.

De façon équivalente, on suppose que les ε_i sont distribuées normalement.

4.4 Estimation de la droite de régression $E(Y) = \beta_0 + \beta_1 X$

Comme dans tout travail statistique, nous nous servons des observations de l'expérience ou pour l'enquête pour calculer certaines statistiques qui servent d'estimation des paramètres.

En régression linéaire simple, l'objectif est d'obtenir une droite qui ajuste le mieux possible aux points de diagramme de dispersion. Plusieurs droites peuvent s'ajusté à un nuage de points mais parmi toutes ces droites, on veut retenir celle qui jouit d'une propriété remarquable (celle qui permet de rendre minimum la somme des carrés des écarts des valeurs observées Y_i à la droite. Cette droite est dite la droite de régression empirique.

La détermination exacte des paramètres de cette droite est pratiquement impossible mais on peut calculer des estimateurs de ces paramètres ou plus clairement on estime des intervalles qui englobes les vraies valeurs, avec $E(\text{estimateur}_{(i)}) = \beta_i$.

Plusieurs méthodes peuvent être utilisé pour l'estimation, mais la plus efficace est celle de moindre de carré qui donne la droite de régression la plus proche aux données. Le principe de cette méthode sera détaillé dans les prochains paragraphes.

5. Régression linéaire multiple [4].

5.1 Introduction

Dans le paragraphe précédent, nous avons traité du modèle de régression linéaire simple où n'intervenait qu'une seule variable explicative : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

La démarche de l'analyse consiste essentiellement à estimer les paramètres du modèle, et à déterminer la variation résiduelle.

5.2 Modèle linéaire à plusieurs variables explicatives.

Énonçons d'abord le **modèle linéaire multiple** sous sa forme la plus générale. Le mot **multiple** est introduit ici à cause de la présence de plusieurs variables explicatives dans le modèle contrairement à une seule variable explicative dans le cas du modèle linéaire simple. Le terme **linéaire** s'applique, par contre, aux paramètres du modèle et non aux variables explicatives. Les méthodes de régression étudiée seront également valides même si le modèle comporte des termes non linéaires en X .

Le modèle de régression multiple permet de décrire la liaison entre une variable dépendante Y et un ensemble de variables explicatives $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ la forme générale de cette liaison s'énonce comme suit :

Modèle de régression linéaire multiple

$$Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$$

Où

Y_i est la variable dépendante (ou expliquée) dans les valeurs sont conditionnés par celles des variables explicatives $X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(k)}$ et la composante aléatoire ε_i (non observable).

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$ sont les $(k+1)$ paramètres du modèle $X_i^{(j)}$ $j = 1, 2, \dots, k$ représentent la i ème valeur des k variables explicatives. on les considère comme des grandeurs certaines.

ε_i dénote la fluctuation aléatoire non observable attribuable à un ensemble de facteurs ou de variables non pris en considération dans le modèle ou que nous ne savons pas identifier.

5.3 Hypothèses fondamentales.

Les hypothèses fondamentales sur lesquelles repose le modèle de régression multiple se sont sensiblement les mêmes que celle du modèle linéaire simple.

Hypothèses fondamentales du modèle de régression multiple

1. On suppose que le terme ε_i est une variable aléatoire de moyenne 0 et de variance constante :
 $E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2$ pour toute i .
 2. Il n'existe aucune corrélation entre les erreurs c.-à-d..
 $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ pour tout i et tout $j, i \neq j$.
 3. les variables explicatives $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ sont des grandeurs certaines. Elles ne présentent donc pas un caractère aléatoire. Elles sont observées sans erreur ou fixées à des valeurs arbitraires.
- Et, dans le but de construire des intervalles de confiance et d'effectuer des tests d'hypothèses, on suppose
4. les fluctuations aléatoires ε_i sont distribuées normalement :
 $\varepsilon_i \sim N(0, \sigma^2)$.

La spécification de ces hypothèses nous permet de caractériser le comportement des Y_i dans le modèle $Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$. On peut en déduire ce qui suit concernant la variable dépendante Y :

Les observations sans Y_i distribuée normalement et indépendamment avec :

$$\text{moyenne } E(Y_i) = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)}$$

et variance $\text{Var}(Y_i) = \sigma^2$

Remarque : a) il existe également une condition supplémentaire dans le cas de la régression multiple pour estimer correctement les paramètres du modèle soit celle d'absence de colinéarité entre les variables explicatives.

b) Précisons également que l'estimation et les testages les paramètres du modèle de régression ne sont possibles en autant que :

le nombre d'observations excède le nombre de paramètres à estimer $n > k + 1$.

5.4 Formulation d'un modèle de régression multiple : identification et interprétation des composantes du modèle.

La démarche à suivre dans l'élaboration d'un modèle de régression multiple peut se résumer selon les étapes suivantes :

1. Identifier les variables (variable indépendante et variable explicatives) qui font le sujet d'une étude de régression multiple.
2. Spécifie le modèle de régression multiple que l'on envisage d'analyser.
3. Prélever les observations sur les divers variables en s'assurant que le nombre d'observations est supérieur au nombre de paramètres à estimer.
4. Estimer, à l'aide des observations, les paramètres du modèle.
5. Effectuer divers tests statistiques sur les paramètres du modèle pour ne retenir que les variables explicatives qui ont un effet significatif.
6. Résumer les conclusions de l'analyse en indiquant l'équation de régression obtenue, des erreurs-types de chaque coefficient, la table de variance ainsi que la valeur de coefficient de détermination.
7. Effectuer à l'aide de l'équation résultante les estimations et prévisions requises.

Cette démarche est générale et nous en indiquerons les diverses particularités à mesure que nous avancerons dans l'étude de la technique de régression multiple.

5.5 Détermination de l'équation de régression : Estimation des paramètres du modèle.

Nous avons recours à la méthode des moindres carrés pour obtenir les estimations des différents paramètres du modèle de régression multiple.

Avant ça montrant l'écriture matricielle du modèle de régression multiple :

$$\{\mathbf{y}\} = [\mathbf{X}]\{\boldsymbol{\beta}\} + \{\boldsymbol{\varepsilon}\} \quad \text{où} \quad \{\mathbf{y}\} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_k \end{pmatrix}, \quad [\mathbf{X}] = \begin{pmatrix} 1 & X_0^{(1)} & \dots & X_0^{(k)} \\ 1 & \vdots & & \vdots \\ \vdots & & \ddots & \\ 1 & X_n^{(1)} & \dots & X_n^{(k)} \end{pmatrix} \quad \text{et} \quad \{\boldsymbol{\beta}\} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

5.5.1 Estimations par Moindres carrées.

L'expression à minimiser sur $\beta \in \mathbb{R}^{p+1}$ s'écrit :

$$\begin{aligned} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 X_i^{(1)} - \dots - \beta_k X_i^{(k)})^2 &= \|\{\mathbf{y}\} - [\mathbf{X}]\{\boldsymbol{\beta}\}\|^2 \\ &= (\{\mathbf{y}\} - [\mathbf{X}]\{\boldsymbol{\beta}\})^T \cdot (\{\mathbf{y}\} - [\mathbf{X}]\{\boldsymbol{\beta}\}) \\ &= \{\mathbf{y}\}^T \{\mathbf{y}\} - 2\{\boldsymbol{\beta}\}^T [\mathbf{X}]^T \{\mathbf{y}\} + \{\boldsymbol{\beta}\}^T [\mathbf{X}]^T [\mathbf{X}]\{\boldsymbol{\beta}\} \end{aligned}$$

Par dérivation matricielle de la dernière on obtient les « équations normales » :

$$[\mathbf{X}]^T \{\mathbf{y}\} - [\mathbf{X}]^T [\mathbf{X}]\{\boldsymbol{\beta}\} = 0 \quad (5.8)$$

dont la solution correspond bien à un minimum car la matrice *hessien* $2[\mathbf{X}]^T [\mathbf{X}]$ est semi définie positive.

Nous faisons l'hypothèse supplémentaire que la matrice $[\mathbf{X}]^T [\mathbf{X}]$ est inversible, c'est-à-dire que la matrice $[\mathbf{X}]$ est de rang $(k+1)$ est donc qu'il n'existe pas de colinéarité entre ses colonnes. En pratique, si cette hypothèse n'est pas vérifiée, il suffit de supprimer des colonnes de $[\mathbf{X}]$ et donc des variables du modèle. Des diagnostics de colinéarité et des critères aident au choix des variables.

Alors, l'estimation des paramètres β_i est donnée par :

$$\{\mathbf{b}\} = ([\mathbf{X}]^T [\mathbf{X}])^{-1} [\mathbf{X}]^T \{\mathbf{y}\} \quad (5.9)$$

et les valeurs ajustées (ou estimées, prédites) de $\{\mathbf{y}\}$ on pour expression :

$$\{\hat{\mathbf{y}}\} = [\mathbf{X}]\{\mathbf{b}\} = [\mathbf{X}]([\mathbf{X}]^T [\mathbf{X}])^{-1} [\mathbf{X}]^T \{\mathbf{y}\} = [\mathbf{H}]\{\mathbf{y}\} \quad (5.10)$$

Où $[\mathbf{H}] = [\mathbf{X}]([\mathbf{X}]^T [\mathbf{X}])^{-1} [\mathbf{X}]^T$ est appelée « *hat matrix* » ; elle met un chapeau à $\{\mathbf{y}\}$.

Géométriquement, c'est la matrices de projection orthogonale dans \mathbb{R}^n sur le sous-espaces Vect(X) engendrée par les vecteurs colonnes de $[\mathbf{X}]$.

$$\text{On note} \quad \{\mathbf{e}\} = \{\mathbf{y}\} - \{\hat{\mathbf{y}}\} = \{\mathbf{y}\} - [\mathbf{X}]\{\mathbf{b}\} = ([\mathbf{I}] - [\mathbf{H}])\{\mathbf{y}\} \quad (5.11)$$

le vecteur des résidus ; c'est la projection de $\{\mathbf{y}\}$ sur le sous espaces orthogonale de Vect(X) dans \mathbb{R}^n .

Le système d'équations devient rapidement lourd à résoudre. Pour cette raison, une étude de régression multiple ne peut s'effectuer dans la plupart des cas sans avoir recours à un ordinateur (logiciel utilisé pour le calcul MATLAB®).

5.6 Analyse de variance en régression multiple.

Une analyse de régression multiple s'accompagne toujours d'un tableau d'analyse de variance. Ce tableau permettra de tester si l'ensemble des variables explicatives a un effet significatif sur la variable dépendante. Il permettra également d'en déduire l'estimation de la variance des Y_i autour de l'équation de régression (c.-à-d. l'estimation de la variance des erreurs σ^2), estimation qui sera subséquemment utile dans le calcul d'intervalles de confiance ; on pourra en déduire le coefficient de détermination multiple R^2 .

5.6.1 Décomposition de la variation dans les observations : calcul des sommes de carrés.

Cette décomposition s'effectue de façon similaire à celle effectuée dans le cas de la régression linéaire simple.

Calcul des sommes de carrés en régression multiple

- Variance totale $SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$

- Variance expliquée

par la régression :

$$SCR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

- Variance résiduelle : $SC_{RES} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

5.6.2 Calcul des degrés de liberté.

Dans le cas où l'analyse de régression comporte $(k+1)$ paramètres à estimer à l'aide de n observation, les degrés de liberté des sommes de carrés précédentes se décomposent comme suit :

Sommes de carrés	Degrés de liberté
$SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$	$n-1$
$SCR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	k
$SC_{RES} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$n-(k+1)$

5.6.3 Tableau de l'analyse de la variance.

On en déduit par la suite les carrés moyens en divisant les sommes de carrés SCR (due à la régression) et SC_{RES} (résiduelle) par les degrés de liberté respectifs. Ces divers calculs se résument dans un tableau d'analyse de variance.

Tableau d'analyse de variance

Régression multiple.

Source de variance	Somme de carrés	Degrés de liberté	Carrés moyens
Expliquée par la régression	$SCR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	k	$CMR = SCR / k$
résiduelle	$SC_{RES} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$n-k-1$	$CM_{RES} = \frac{SC_{RES}}{n-k-1}$
Totale	$SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$	$n-1$	

5.6.4 Estimation de σ^2 calcul de s^2 .

Le carré moyen résiduel, CM_{RES} est une estimation ponctuelle de σ^2 (variance des Y_i autour du modèle de régression ou variance des erreurs). On peut déduire cette estimation directement du tableau d'analyse puisque :

$$CM_{RES} = s^2 = \frac{SC_{RES}}{n-k-1}$$

L'écart-type des résidus nous donne une mesure de dispersion des Y_i autour de l'équation de régression multiple : $s = \sqrt{CM_{RES}}$ les unités associées à s sont celles de la variable dépendante Y .

Remarque. Le carré moyen résiduel est particulièrement important et intervient dans plusieurs calculs subséquents, entre autres, pour le calcul des variances des coefficients de régression ($Var(b_i)$) ainsi que dans le calcul d'intervalles de confiance sur $E(Y_i)$. Il permet d'établir la marge d'erreur dans les estimations par intervalle.

5.6.5 Coefficient de détermination R^2 .

Pour évaluer la quantité de l'ajustement linéaire de l'équation de régression entre la variable dépendante Y et l'ensemble des variables explicatives, on a recours au coefficient de détermination (ou d'explication) multiple, noté R^2 :

$$R^2 = \frac{SCR}{SCT} \quad \text{avec} \quad 0 \leq R^2 \leq 1$$

Le coefficient R^2 permet d'évaluer la proportion de la variation des Y_i autour de moyenne \bar{Y} qui est expliquée par l'ensemble des variables explicatives retenues dans l'équation de régression.

La proportion qui demeure inexpliquée est $1 - R^2$. Cette partie inexpliquée est attribuable à l'omission de variables explicatives qui pourraient contribuer de façon importante à l'explication des fluctuations de Y . à une formulation incorrecte du modèle (d'autres termes seraient requis) et à l'erreur expérimentale.

Remarque.

a) même si la valeur de R^2 est élevée, l'écart-type des résidus peut être trop grand pour assurer des estimations avec une marge d'erreur acceptable. Ceci peut être le cas si la taille de l'échantillon est petite.

b) pour apprécier pleinement l'utilité d'une équation de régression multiple, il faut lui associer le coefficient de détermination, l'écart-type des résidus, la marge d'erreur dans l'estimation des paramètres β_j ainsi que les marges d'erreur dans l'estimation des $E(Y)$ et dans la prévision de valeurs individuelles de la variable dépendante pour un niveau de confiance donné.

Cependant, le R^2 dans le contexte de la régression multiple pose le problème de la sur-paramétrisation du modèle. Plus l'on ajoute de variables explicatives, plus le R^2 augmente, même si les nouvelles variables explicatives sont très "reliées" à la variable dépendante. Pour éviter ce phénomène, on calcule le coefficient de détermination ajusté :

$$R_{adj}^2 = 1 - \frac{\frac{SCR}{k}}{\frac{SCT}{n - k - 1}}$$

ce qui permet de tenir compte à la fois de l'accroissement du nombre de variables explicatives et de la réduction de SCR par rapport à SCT .

Il faut également noter que le coefficient de détermination est le carré du coefficient de corrélation et représente donc bien un indicateur d'association linéaire.

5.7 Tests d'hypothèse en régression multiple.

Dans l'élaboration d'un modèle de régression multiple, divers tests statistiques sont requis. Les plus fréquents permettent de répondre aux questions suivantes :

- 1- Est ce que la régression est significative dans son ensemble ?
- 2- Est ce que la contribution marginale de chaque variable explicative (ou de chaque terme du modèle) est significative ? le teste correspondre permettra d'examiner si l'ajout d'une variable explicative à la suite d'autre variables explicatives déjà dans l'équation de régression apporte une contribution significative dans l'explication de la variable dépendante.
- 3- Est ce qu'un sous-ensemble de variable explicatives ont une influence sur la variable dépendante Y lorsqu'elles s'ajoutent au modèle déjà un certain nombre de variables ?

Abordons d'abord le premier test qui va permettre de répondre à la question 1.

5.7.1 Comment déterminer si la régression est significative dans son ensemble ?

Dans un modèle de régression multiple, disons un modèle comportant k variables explicatives.

$$Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$$

On veut tester si la régression est significative dans son ensemble. Les hypothèses nulle et alternatives sont alors :

$$\left\{ \begin{array}{ll} H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 & \text{(aucune contribution significative des } X_i \text{)} \\ H_1 : \text{au moins un des } \beta_j \text{ est} & \text{(au moins une variable, disant } X_j \text{, apporte} \\ \text{différent de 0} & \text{une contribution significative)} \end{array} \right.$$

Pour effectuer le teste on a recours au carrés moyen d'un tableau d'analyse de variance. Il s'agit de comparer le carré moyen dû à la régression avec le carré résiduel. La valeur observée du F est donc :

$$F = \frac{CMR}{CM_{RES}} = \frac{SCR/k}{SC_{RES}/n-k-1}$$

On obtient donc un rapport de deux variables.

Sous l'hypothèse $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, on démontre que la qualité $F = \frac{CMR}{CM_{RES}}$ est

distribuée selon la loi de Fisher avec k et $n-k-1$ degrés de liberté. Au seuil de signification α , le test peut se résumer comme suit :

Test de signification sur l'ensemble des paramètres β_i
Modèle : $Y = \beta_0 + \beta_1 X^{(1)} + \beta_2 X^{(2)} + \dots + \beta_k X^{(k)} + \varepsilon$
Hypothèses : $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ $H_1 : \text{au moins un des } \beta_j \neq 0$
Quotient des carrés moyens : $F = \frac{CMR}{CM_{RES}}$
Règle de décision : Rejeter H_0 si $F > F_{\alpha; k, n-k-1}$

Si nous rejetons H_0 , nous concluons alors que la contribution de l'ensemble des variables pour expliquer les fluctuations de la variable dépendante Y est significative au seuil α .

Remarque.

a) ce test ne permet pas toutefois de préjuger la signification particulière de tels ou tels coefficients de régression. Comme nous le verrons subséquemment, un test de *Student* permettra de déterminer si la **contribution marginale** de chaque variable explicative est significative.

b) Une façon équivalente d'établir la règle de décision est d'évaluer la probabilité, en supposant $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ vraie, pour que la quantité F soit supérieure ou égale à la valeur observée de F selon les résultats de l'échantillon : $P(F \geq F_{\text{observé}})$. Si cette valeur est très petite, les résultats de l'échantillon ne permettent pas de supporter l'hypothèse H_0 . Pour conclure, on compare cette valeur de probabilité avec le seuil α . Si $P(F \geq F_{\text{observé}}) < \alpha$, on rejette H_0 .

On considère que la valeur observée pour le F ne permet pas de supporter l'hypothèse H_0 . Cette probabilité donne en quelque sorte le degré de conviction que la décision prise est bonne. Si ce degré de conviction est plus faible que le risque choisi (α) de rejeter à tort H_0 . On considère alors que l'hypothèse H_0 est peu vraisemblable et que les résultats observés favorisent plutôt H_1 . Ainsi, en rejetant H_0 le risque de prendre une mauvaise décision est encore plus faible que celui que nous nous sommes fixé.

c) Utilisation de R^2 pour tester si la régression est significative dans son ensemble. On peut établir une relation entre la variable de Fisher (F) et le coefficient de détermination R^2 pour tester l'hypothèse $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$. SCT et $SC_{RES} = (1-R^2) \cdot SCT$; alors

$$F = \frac{SCR/k}{SC_{RES}/n-k-1} = \frac{R^2 \cdot SCT/k}{(1-R^2) \cdot SCT/n-k-1} = \frac{R^2/k}{(1-R^2)/n-k-1}$$

qui est distribué selon la loi de Fisher avec k et $n-k-1$ degrés de liberté.

5.7.2 Distribution d'échantillonnage des coefficients de régression.

Avant d'aborder le testage individuel de chaque paramètre β_j du modèle, identifions la distribution d'échantillonnage des coefficients de régression b_j . Les estimateurs b_j (estimateurs de β_j) sont des variables aléatoires qui prennent des valeurs particulières pour un ensemble particulier de données. Pour être en mesure de placer des intervalles de confiance sur les β_j ou d'effectuer des tests d'hypothèses, il faut connaître selon quelle loi sont distribuées les fluctuations d'échantillonnage des b_j .

Distribution d'échantillonnage du coefficient b_j

Sous l'hypothèse de normalité des erreurs ε_i (ou celle des Y_i), la distribution d'échantillonnage du coefficient b_j , $j = 0, 1, \dots, k$, est celle d'une **loi normale** de moyenne $E(b_j) = \beta_j$ et de variance $\sigma^2(b_j)$

On obtient une estimation de $\sigma^2(b_j)$ avec $s^2(b_j)$. L'écart-type (ou l'erreur type) de chaque coefficient de régression s'obtient de $s(b_j) = \sqrt{s^2(b_j)}$.

Dans le cas d'un petit échantillon, la fluctuation de l'écart réduit.

$$\frac{b_j - \beta_j}{s(b_j)}$$

Sont celles de la loi de *Student* avec $n-k-1$ degrés de liberté.

5.7.3 Détermination des variables des coefficients de régression.

Pour effectuer des tests statistiques sur chaque paramètre β_j du modèle de régression ou calculer des intervalles de confiance sur les β_j , on doit utiliser l'expression de l'écart

réduit :

$$\frac{b_j - \beta_j}{s(b_j)}$$

Le dénominateur de l'écart réduit est l'écart-type du coefficient b_j . Toutefois l'expression algébrique pour calculer $s(b_j)$ est relativement complexe en régressions multiple (à moins de faire appel à l'approche matricielle pour traiter les notions de régression multiple).

Les programmes informatiques permettent d'obtenir les variances $s^2(b_j)$ ainsi que l'écart-type (l'erreur-type) de chaque coefficient de régression b_j .

5.7.4 Test sur la contribution marginale de chaque explicative

On veut tester si la contribution marginale de chaque variable explicative dans l'équation de régression multiple est significative. Ce test permet d'examiner si l'ajout d'une variable explicative à la suite d'autres variables explicatives déjà dans l'équation de régression apporte une contribution significative. Le teste avec le rapport F de Fisher pour l'ensemble des variables explicatives ne permet pas de juger si certains coefficients de régression b_j s'avèrent non significatifs, rendant ainsi superflu l'utilité de certains variables explicatives dans l'équation de régression. Ce test permettra donc de juger de la pertinence de chaque variable explicative comme si elle était la dernière variable introduite dans l'équation de régression. On parle donc de contribution marginale puisque l'on tiendra compte des variables déjà dans l'équation de régression.

Test avec le t de Student.

Indiquons comment s'exécute ce test en utilisant le fait que l'écart réduit :

$$\frac{b_j - \beta_j}{s(b_j)}$$

Est distribué selon une loi de *Student* avec $n-k-1$ degrés de liberté.

Test de signification sur chaque paramètre β_j

On veut tester si la contribution marginale de chaque variable explicative X_j est significative dans le modèle

$$Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i \text{ au seuil de signification } \alpha . \text{ Il}$$

s'agit alors de tester les hypothèses suivantes :

$$H_0: \beta_j = 0, \quad j = 1, 2, \dots, k.$$

$$H_1: \beta_j \neq 0.$$

Sous H_0 , l'écart réduit devient $\frac{b_j}{s(b_j)}$ qui est distribué selon la loi de

Student avec $n-k-1$ degrés de liberté.

Règle de décision : Rejeter H_0 si $t > t_{\alpha/2; n-k-1}$.

si $t < -t_{\alpha/2; n-k-1}$.

Conclusion : si H_0 est rejeté, la contribution marginale de la variable explicative X_j est significative au seuil α

5.7.5 Signification pratique de ce test.

Apportant quelque explication additionnelles sur l'utilisation du t de *Student* pour tester la contribution marginale de chaque variable explicative.

Pour ce faire, nous comparons le modèle linéaire simple avec un modèle multiple a trois explicatives.

Modèle linéaire simple. $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

$$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$$

on utilise $t = \frac{b_j}{s(b_j)}$ pour effectuer le test. Si on ne peut rejeter H_0 , on considère alors que la variable explicative ne permet pas d'expliquer de façon significative les fluctuations dans Y .

Modèle linéaire multiple : $Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$

On veut tester la contribution marginale de X_l : $H_0: \beta_l = 0, H_1: \beta_l \neq 0$.

On utilise a nouveau $t = \frac{b_j}{s(b_j)}$ pour effectuer le test (mais avec les valeurs appropriées de

l'estimation du modèle à 3 variable explicative).

Toutefois l'application du test de *Student* à l'hypothèse

$$H_0: \beta_l = 0 \quad Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$$

Permet de tester la contribution marginale de X_l comme si elle était la dernière variable introduite dans l'équation de régression.

Si on peut rejeter les $H_0: \beta_l = 0$ dans le modèle à trois variable explicatives, alors quelle conclusion peut-on- tirer?

Il ne faut pas conclure que la variable explicative X_l , n'apporte aucune contribution significative. Une régression simple entre Y et X_l ou l'étude d'une autre forme de liaison amener une conclusion contraire.

Conclusion plus nuancée : On devra plutôt conclure que la contribution marginale de X_l (ou de toute variables X_j lorsque $H_0 = \beta_j = 0$) lorsque elle est introduite a la suite d'autres variables explicatives est non significatives. Son ajout est superflu compte tenu de la présence des autres variables explicatives déjà dans l'équation de régression. En d'autres termes, l'ajout de cette variable ne réduit pas de façon appréciable la variation non expliquée par les autres variables explicatives déjà dans l'équation de régression.

Intervalle de confiance sur β_j . La plupart des programmes informatiques donnent également l'intervalle de confiance sur chaque paramètre β_j (habituellement avec un niveau de confiance $100*(1-\alpha)\%$, l'intervalle de confiance s'écrit:

$$b_j - t_{\alpha/2; n-k-1} \cdot s(b_j) \leq \beta_j \leq b_j + t_{\alpha/2; n-k-1} \cdot s(b_j)$$

Si $\beta_j = 0$ se situe dans l'intervalle, la contribution marginale de la variable explicative X_j , compte tenu des autres variables explicatives déjà incluses dans l'équation de régression, est non significative au seuil de signification α .

Chapitre VI :

Estimation des paramètres spectraux.

1. Introduction

Divers investigations ont été menées pour le développement des lois d'atténuations servant à définir les paramètres caractéristiques des mouvements forts [Sabetta&Pugliese 1996].

Dans le cadre de ce travail, nous avons choisi de définir des lois de régression pour quatre paramètres caractéristiques des accélérations sismiques à savoir $f^0(t)$ la fréquence centrale, $f^S(t)$ la bande fréquentielle, I_a l'intensité d'Arias et en fin la durée de phase forte D_V .

Quatre lois de régression seront présentées pour l'estimation des quatre paramètres respectifs, chacune de ces lois sera fonction des variables indépendantes (M caractérise la taille du séisme, R caractérise la localisation de site par rapport à la source, S_1, S_2 caractérisent la nature du site et t caractérise le temps l'évolution temporelle de mouvement sismique fort (utilisée uniquement pour le paramètre f^0).

Une méthodologie de régression multidimensionnelle a été employée pour l'évaluation des coefficients des variables indépendantes et pour chacune des quatre lois prédictives.

L'évaluation de ces coefficients est faite sur la base d'une banque de données accélérométrique mondiale issues des différents pays.

2. Banque de données [6].

2.1 Introduction.

Dans ce travail les données utilisées sont présentées sous forme de courbes décrivant la variation des paramètres spectraux, ces paramètres caractérisent les fluctuations non stationnaires du contenu spectral des enregistrements sismiques, elles sont arrangées par classes "magnitude - distance épacentrale - conditions locales de site". Afin de permettre l'obtention d'une dispersion minimale, les paramètres sont évalués à partir des spectres

physiques optimisés (maximisés et minimisés) d'enregistrements sismiques issus d'une banque de données mondiale.

L'intensité du séisme, représentée ici par la magnitude, est caractérisée pour un séisme donné, par le mode de rupture, les valeurs respectives des dimensions de faille, des vitesses de rupture et des dislocations (valeurs des glissements finaux et des temps de montée).

La variabilité, pour une même catégorie de magnitudes, provient des différents schéma de rupture sur les différentes faille, des radiations d'ondes et des caractéristiques singulières des failles (surface, azimut, pendage. Direction du glissement).

De même, La distance épacentrale a un effet direct sur le processus d'atténuation des ondes sismiques durant leur trajet de la source au récepteur. De plus, en considérant les conditions locales du site, on prend implicitement en compte l'effet, sur les ondes sismiques, des différents remplissages alluviaux (dus à l'hétérogénéité et à l'inélasticité des matériaux) et des reliefs topographiques (interférences constructives des ondes réfléchies).

Ainsi, une fois le modèle de spectre physique caractérisé, il nous est donc possible d'obtenir les valeurs des amplitudes spectrales non stationnaires dans le plan temps-fréquence, et par voie de conséquence, de simuler les accélérations sismiques pour chaque classe considérée. Nous analysons par la suite, les variations des paramètres spectraux du modèle en fonction de la magnitude, de la distance épacentrale et des conditions locales de sites.

2.2 Étude statistique sur les enregistrements.

L'estimation du spectre physique pour une classe "magnitude-distance-site" nécessite donc inévitablement une étude statistique des enregistrements accélérométriques contenus dans cette classe. Pour l'estimation des paramètres du modèle de spectre physique, nous suggérons les étapes suivantes :

1- Pour les deux composantes horizontales enregistrées en un site donné, nous calculons d'abord les spectres physiques optimisés (le maximisé et le minimisé), dans le but de retrouver les limites supérieures et inférieures de l'énergie sismique caractérisant ainsi les deux composantes horizontales au site considéré.

2- Nous généralisons le contenu fréquentiel de chaque accélérogramme optimisé en lissant son spectre physique et en le normalisant de sorte que son maximum soit égal à l'unité.

3- Nous sélectionnons un modèle analytique approprié pour approximer à chaque instant et pour chaque fréquence le spectre physique obtenu en 1. Plusieurs épreslms telles les distributions Lognormale, gamma, bêta, de Rayleigh ou Gaussienne peuvent être considérées.

4- Nous calculons pour chaque classe “magnitude-distance-site” les valeurs, à chaque instant t , des paramètres du modèle associés à chacun des enregistrements contenus dans cette classe.

5- Enfin, nous menons une étude statistique afin d’estimer les valeurs moyennes et à 90% de confiance de ces paramètres, assurant ainsi l’ajustement du modèle de spectre physique pour la classe considérée.

2.3 Répartition des données.

La méthodologie décrite précédemment est appliquée à un grand nombre d’accélérogrammes (411) enregistrés en **Italie, Californie, Chili, Mexico, Japon, Taiwan** et **Algérie** et correspondant à diverses magnitudes, distances épacentrales et conditions locales de sites.

Les données sont divisées en 45 différentes classes correspondants à 5 intervalles de magnitudes M1 à M5 ($M_1 < 4.5$, $4.5 < M_2 < 5.5$, $5.5 < M_3 < 6.5$, $6.5 < M_4 < 7.5$, $M_5 > 7.5$). Trois intervalles de distances R1 à R3 comme présenté dans le tableau 1, et 3 catégories de sites (sol raide, dépôt d’alluvion superficielle, dépôt d’alluvion profond). Le tableau 2 montre le détail de cette sélection pour chaque classe.

	R1	R2	R3
M1	<10	10 - 20	>20
M2	<15	15 - 30	>30
M3	<20	20 - 40	>40
M4	<25	25 - 50	>50
M5	<30	30 - 60	>60

Tableau 1

	sol raide			Alluvion superficielle			Alluvion profonde		
	R1	R2	R3	R1	R2	R3	R1	R2	R3
M1	8	7	2	5	5	9	5	7	1
M2	15	5	2	8	7	8	13	2	11
M3	32	6	17	8	11	8	23	19	17
M4	14	5	15	7	29	23	4	18	18
M5	3	2	9	3
Tot1	69	23	39	28	54	57	45	46	50
Tot2				139			141		

Tableau 2

3. Présentation des lois prédictives.

3.1 Choix des variables (Critères statistique).

Supposant qu'une variable Y est fonction de plusieurs variables exogène $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ ces dernière n'explique pas Y de même degré il existe des variables dominantes et des variables non dominantes. (la contribution des variables non dominantes est minimale elle est intégrer en ε) l'objectif du choix de variables est de trouver un modèle avec le minimum de variables exogènes et un maximum d'explication à un risque, les méthodes suggérées par la théorie sont :

a) Méthodes de combinaison de toutes les régressions :

Elle consiste à faire toutes les régressions possibles (à 2 variables, à 3 variables,...) pour chaque régression, on calcul le R_{adj}^2 enfin choisir la régression dont le R_{adj}^2 est le max à un risque prés avec le minimum de variables

b) Procédure une régression progressive des variables (back ward elimination) :

Elle consiste à

1. Faire une régression avec moindres carrés globale avec toutes les variables et éliminer les variables exogènes avec t -test qui ne sont pas dominantes.

2. Faire une deuxième régression avec les nouvelles variables retenus par la première étape. Faire le t -test une deuxième fois.

Répéter la procédure jusqu'à ce qu'aucune variable n'est éliminée.

c) Méthode de sélection progressive (Forward) :

Elle consiste de faire k régression simples et choisir la variable $X^{(m_1)}$ dont le R_{adj}^2 plus élevé.

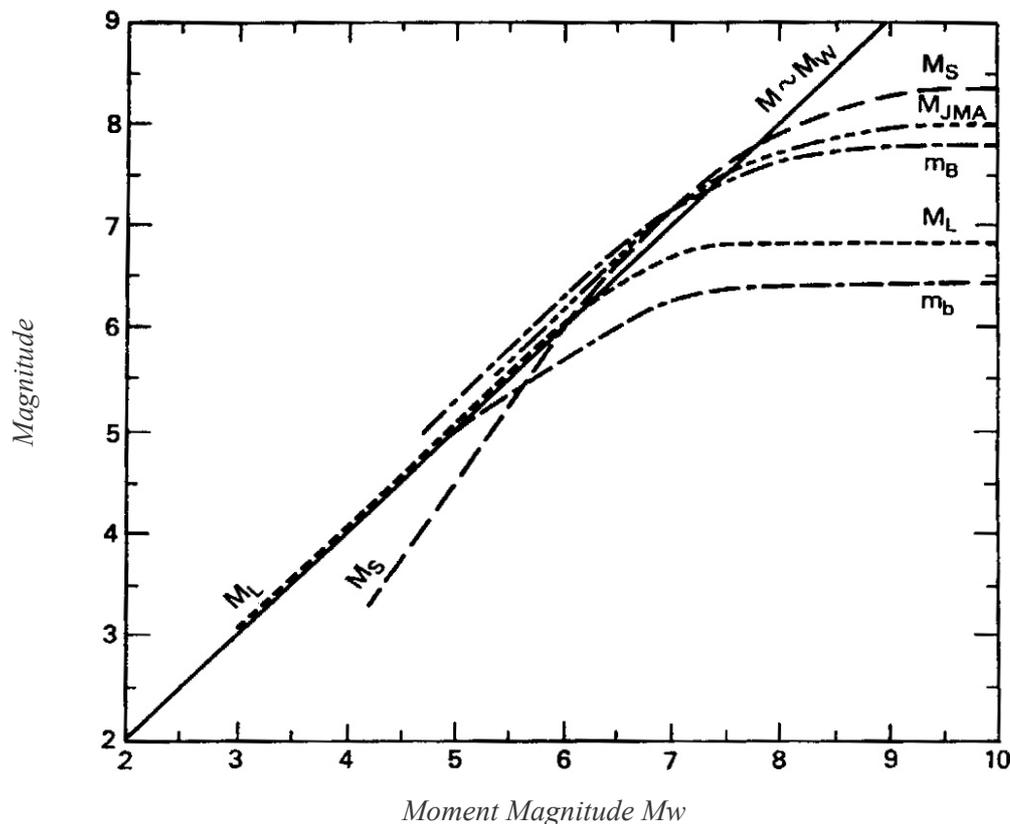
3.2 Description des variables dépendants et indépendante [14].

3.2.1 Variables indépendantes.

Les principaux paramètres mis en comptes généralement dans une loi de régression sont associés à la caractérisation de la source sismique, les effets de propagation ainsi que l'effet de site.

Magnitude :

Le paramètre magnitude, noté M dans ce travail correspond à la magnitude des ondes du surface (M_S), quand M_L et M_S sont supérieure ou égale 5,5 ($M_S > 5,5$ et $M_L > 5,5$) et à M_L quand c'est inférieure. On a pas utilise les moments magnitudes [Hanks and Kanamois 1979] de fait que les lois d'estimations du moments sismique sont pas valable pour les faibles séismes utilise dans la régression. La figure ci-dessous montre la correspondance entre les divers types de magnitudes.



Distance :

La régression a été accomplie en utilisant à la fois la distance épacentrale et distance focale, cette dernière est définie comme la plus petite distance entre la station d'enregistrement et la projection de la surface de la faille de rupture [Joyer et Borre 1981].

Condition locale de site :

Selon les formations géologiques, géotechniques [Sabetta et Pugliese 1987] et l'épaisseur de la couche de sol, les sites sont classés en trois catégories :

Sol raide (rocheux), dépôt superficielle d'alluvion ($H \leq 20m$) et dépôt profond d'alluvion ($H > 20$). Les sols raides sont caractérisés par une vitesse des ondes de cisaillement supérieure à 800m/sec et les alluvions par des vitesses entre 400 et 800 m/sec.

3..2 Variables dépendantes.

- **Paramètres Fréquentiels.**

Les paramètres fréquentiels f^0 et f^S décrivent le contenu fréquentiel de mouvement fort (cf. chapitre IV)

- **Paramètre énergétique.**

Le paramètre énergétique $\alpha(t)$ est ajusté par une fonction lognormal illustrée dans la (figure 12) définie comme suit :

$$\alpha(t) = \frac{I_A}{t \sqrt{2\pi\sigma_\alpha}} e^{-[\ln(t) - \mu]^2 / 2\sigma_\alpha^2}$$

$$\text{Où } \begin{cases} \mu^2 = \ln(t_m) + \sigma_\alpha^2 & \sigma_\alpha = \ln(t_3/t_m)/2,5 ; \\ t_1 = R / 7 ; D_v = 2 * d & t_3 = t_1 + 2,5D_v ; \\ t_m \text{ délai moyen de groupes .} \end{cases}$$

Et : I_a et D_v : sont l'Intensité d'Arias la durée de phase fort (cf. chapitre 3).

t_1 : corresponde au délai en seconde entre les ondes S et P qui est calculé en divisons la distance épacentrale R sur un facteur approximé par 7km/sec qui est tirée de la relation entre le temps d'arrivée des ondes S $t_s=R/3.2$ et des ondes P $t_p=R/5.5$ et tel que la différence $t_s - t_p = R/3.2 - R/5.5 = R/7$.

Remarque : Le choix de t_3 et σ est abouti après des tests sur des accélérogrammes réels dans le but de trouver une enveloppe temporelle avec les caractéristiques suivantes (figure12):

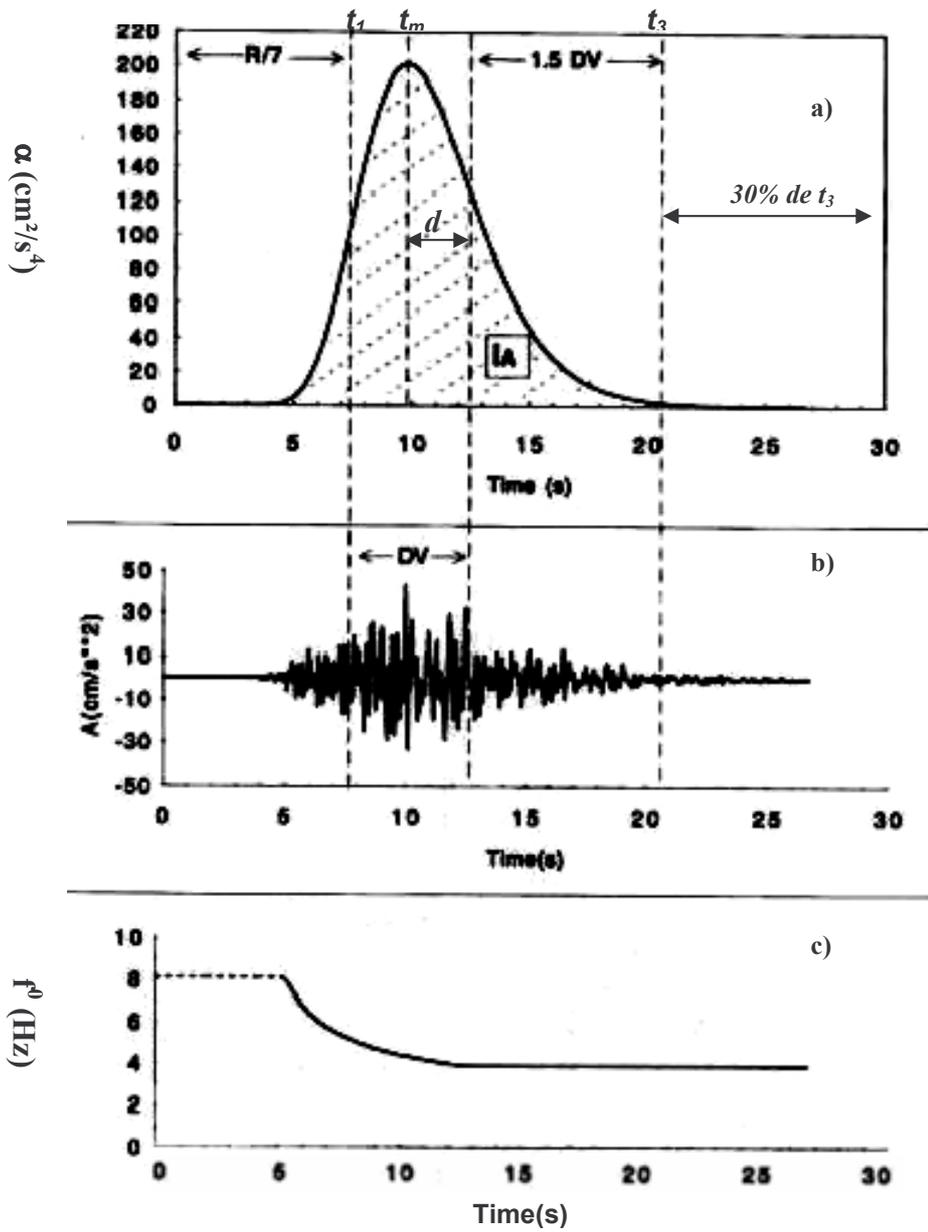


Figure 12 a) enveloppe temporelle utilisée pour la simulation ;
 b) accélérogramme simulé ;
 c) fréquence centrale en fonction de temps. (Sabetta et al.)

Les lois $\alpha(t)$, $f^0(t)$ et $f^s(t)$ utilisées comme modèle de régression dérivent d'une analyse non stationnaire effectuée par [Sabetta et al. 1986] ; les résultats trouvés indique que dans le domaine temporelle, $\alpha(t)$ se présente sous forme d'une fonction lognormal, $f^0(t)$ montre un forme exponentielle, et le rapport $\frac{f^s}{f^0}$ est indépendant du temps.

Les expressions des lois précédentes sont :

$$\left\{ \begin{array}{l} \ln[\mathbf{f}^0(\mathbf{t})] = \beta_0 + \beta_1 \cdot \ln(\mathbf{t}) + \beta_2 \cdot \mathbf{M} + \beta_3 \cdot \mathbf{S}_2 \pm \sigma \\ \frac{\mathbf{f}^s}{\mathbf{f}^0} = \beta_0 + \beta_1 \cdot \mathbf{M} + \beta_2 \cdot \mathbf{S}_1 + \beta_3 \cdot \mathbf{S}_2 \pm \sigma \\ \ln(\mathbf{D}_v) = \beta_0 + \beta_1 \cdot \mathbf{R}_v + \beta_2 \cdot \mathbf{M} + \beta_3 \cdot \mathbf{S}_1 + \beta_4 \cdot \mathbf{S}_2 \pm \sigma \\ \ln(I_a) = \beta_0 + \beta_1 \cdot \mathbf{R}_a + \beta_2 \cdot \mathbf{M} + \beta_3 \cdot \mathbf{S}_1 + \beta_4 \cdot \mathbf{S}_2 \pm \sigma \end{array} \right.$$

$$\text{Où : } R_a = \sqrt{R^2 + 5.3^2} ; \quad R_v = \sqrt{R^2 + 5.1^2}$$

σ : est estimé de l'écart type des résidus à partir des régressions.

Chapitre VII :

Présentation et discussion des résultats de la régression.

Dans ce paragraphe nous présentons les résultats de calcul avec l'analyse des erreurs et les paramètres statistiques associés.

1. Analyse de régression sur la fréquence centrale.

1.1 Spécification du modèle

Un modèle à trois variables explicatives que l'on pourrait utiliser pour expliquer les fluctuations des valeurs de la variable dépendante et interpréter chaque composante du modèle.

Le modèle à trois variables explicatives est : $\ln[\mathbf{f}^0(\mathbf{t})] = \beta_0 + \beta_1 \cdot \ln(\mathbf{t}) + \beta_2 \cdot \mathbf{M} + \beta_3 \cdot \mathbf{S}_2 \pm \sigma$

Où M : représente la magnitude

t : représente le temps

S2 : représente le site

β_0 : théoriquement ce paramètre représente la moyenne des $\ln[\mathbf{f}^0(\mathbf{t})]$ lorsque la valeur de variation de chaque variable explicative est égale à zéro. Toutefois, étant donné le champ de variation de chaque variable de ce contexte, on ne peut donner de signification concrète à β_0 n'ayant pas données dans le voisinage de zéro.

ε_i : ce terme dénote la fluctuation aléatoire non observable attribuable à l'omission dans le modèle d'autres variables explicatives ou à des composantes non identifiées dans le modèle. on le suppose distribué normalement et indépendamment avec moyenne 0 et variance σ^2 .

1.2 Estimation des paramètres du modèle : calcul des coefficient.

Les estimateurs des paramètres du modèle sont :

Paramètres	Estimateurs
β_0	b ₀
β_1	b ₁
β_2	b ₂
β_3	b ₃

Les valeurs prises par ces estimateurs sont les coefficients de l'équation de régression. Nous aurons recours au logiciel pour le calcul de ces coefficients.

Nous résumons dans le tableau ci contre les diverses valeurs des coefficients obtenues.

Coefficients de régression
$b_0=2.8451$
$b_1=-0.2067$
$b_2=-0.3272$
$b_3=-0.1659$

1.3 Équation de régression multiple.

L'équation de régression multiple est donc :

$$\ln(f^0(t)) = 2.84 - 0,209 \cdot \ln(t) - 0,165 \cdot M - 0,327 \cdot S^2$$

1.4 Détermination du tableau de variance.

Tableau d'analyse de variance :

Source de variance	Somme de carrés	Degrés de liberté	Carrés moyens
Expliquée par la régression	$SCR = 9571.104$	3	$CMR = 3190.36$
	$SC_{RES} = 10937.78$	62551	$CM_{RES} = 0.1748$
résiduelle			
Totale	$SCT = 20508.88$	62554	

1.5 Détermination de la variance et du coefficient R^2 .

Du tableau d'analyse de variance, on peut en déduire que :

$$s^2 = CM_{RES} = 0.1748$$

L'écart-type des résidus est :

$$s = 0.4123$$

Le coefficient de détermination est :

$$R^2 = \frac{SCR}{SCT} = 0,4668$$

Pratiquement nous avons 47% de la fluctuation des observations de la fréquence centrale assemblées autour de la moyenne.

1.6 Tests d'hypothèse en régression multiple.

On veut tester l'hypothèse selon laquelle les trois variables explicatives (valeurs des magnitudes, les conditions de site et le temps) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\ln(f^0)$.

Les hypothèses nulle et alternative que l'on désire tester dans le cadre d'un modèle de régression à 3 variables sont :

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$$

H1 : Au moins un des β_j est différent de 0.

On aura donc recours au rapport des carrés moyens pour effectuer le test :

$$\text{Quotient des carrés moyens : } F = \frac{CMR}{CM_{RES}}$$

Pour $k=3$, $n-k-1=62551$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;3,62551}=0,5463$.

Règle de décision : Rejeter H_0 si $F > F_{\alpha;k, n-k-1}$

D'après le tableau d'analyse de variance on a :

$$CMR = 3910,36 \quad CM_{res} = 0.1748$$

La valeur observée pour le rapport F est donc 18245,0774

Puisque $F=18245.0774 > 0.1748$ nous rejetons l'hypothèse H_0 et favorisons H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil $\alpha=0.05$, qu'au moins une des variables explicatives sinon les trois, apportent une contribution significative dans l'explication du nombre de pièces assemblées. Il y a au moins un des paramètres β_j qui est significativement différent de zéro.

1.7 Test sur la contribution marginale de chaque variable explicative avec le *t* de Student.

Dans une étude de régression multiple portant sur l'analyse de facteurs pouvant être liés à la variation de la fréquence centrale, on a obtenues les coefficients de régression et les erreurs-types de chaque coefficient.

Variabes explicatives	Coefficients de régression b_j	Erreurs-types $s(b_j)$
Coefficient b_0	2.8451	0.0087
Ln(t)	-0.2097	0.00161
M	-0.3272	0.00358
S2	-0.1656	0.00142

Testons la contribution marginale de chaque variable explicative dans le modèle $\ln[f^0(t)] = \beta_0 + \beta_1 \cdot \ln(t) + \beta_2 \cdot M + \beta_3 \cdot S_2 \pm \varepsilon$.

On utilisera donc le t de *Student* pour effectuer le test et un seuil de signification $\alpha = 0.05$, la valeurs critique est donc :

	Coefficients	Erreur-type	Statistique t	Limite inférieure pour seuil de confiance = 95%	Limite supérieure pour seuil de confiance = 95%
Constante	2,84515039	0,008729822	325,911598	2,82803992	2,86226086
Variable $\ln(t)$	-0,2097143	0,001610976	-130,17844	-0,2128719	-0,2065568
Variable M	-0,3272989	0,003581949	-91,374548	-0,3343195	-0,3202783
Variable S_2	-0,1656344	0,001427381	-116,04081	-0,1684321	-0,1628368

De cette analyse, on peut dire pour toutes les variables la contribution marginale au présence des autres variables est significatives.

1.8 Comparaisant de loi avec les données et les lois de (Sabetta et al.) :

L'équation de la fréquence centrale :

$$\ln(f^0(t)) = 2.84 - 0.209 \cdot \ln(t) - 0.165 \cdot M - 0.327 \cdot S_2 \pm 0.418$$

Et l'équation de « Sabetta&Puglièse » est :

$$\ln(f^0(t)) = 3,4 - 0,35 \cdot \ln(t) - 0,218 \cdot M - 0,15 \cdot S_2$$

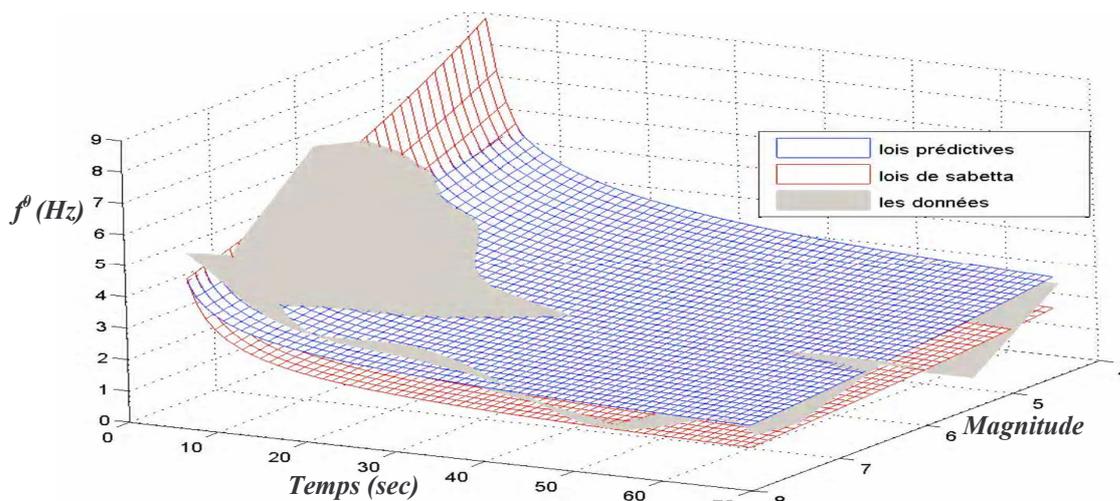


Figure 13 Comparaison des fréquences centrales

2. Analyse de régression sur le rapport de la bande fréquentielle et la fréquence centrale.

2.1 Spécification du modèle.

Un modèle à trois variables explicatives que l'on pourrait utiliser pour expliquer les fluctuations des valeurs de la variable dépendante et interpréter chaque composante du modèle.

Le modèle à trois variables explicatives est : $\frac{f^s}{f^0} = \beta_0 + \beta_1 \cdot M + \beta_2 \cdot S_1 + \beta_3 \cdot S_2 \pm \sigma$

Où M : représente la magnitude

S1, S2 : représente le site

β_0 : théoriquement ce paramètre représente la moyenne des $\frac{f^s}{f^0}$ lorsque la valeur de variation de chaque variable explicative est égale à zéro. Toutefois, étant donné le champ de variation de chaque variable de ce contexte, on ne peut donner de signification concrète à β_0 n'ayant pas données dans le voisinage de zéro.

ε_i : ce terme dénote la fluctuation aléatoire non observable attribuable à l'omission dans le modèle d'autres variables explicatives ou à des composantes non identifiées dans le modèle. on le suppose distribué normalement et indépendamment avec moyenne 0 et variance σ^2 .

2.2 Estimation des paramètres du modèle : calcul des coefficient.

Les estimateurs des paramètres du modèle sont :

Paramètres	Estimateurs
β_0	b ₀
β_1	b ₁
β_2	b ₂
β_3	b ₃

Les valeurs prises par ces estimateurs sont les coefficients de l'équation de régression. Nous aurons recours au logiciel pour le calcul de ces coefficients.

Nous résumons dans le tableau ci contre les diverses valeurs des coefficients obtenues.

Coefficients de régression
$b_0=0.312$
$b_1=-0.046$
$b_2=-0.183$
$b_3=-0.221$

2.3 Équation de régression multiple.

L'équation de régression multiple est donc :

$$\frac{f^S}{f^0} = 0,312 - 0,046.M - 0,183.S1 - 0,221.S2$$

1.4 Détermination du tableau de variance.

Tableau d'analyse de variance :

Source de variance	Somme de carrés	Degrés de liberté	Carrés moyens
Expliquée par la régression	$SCR = 0.502$	3	$CMR = 0.1674$
	$SC_{RES} = 1.267$	36	$CM_{RES} = 0.036$
résiduelle			
Totale	$SCT = 1.799$	39	

.5 Détermination de la variance et du coefficient R^2 .

Du tableau d'analyse de variance, on peut en déduire que :

$$s^2 = CM_{RES} = 0.189$$

L'écart-type des résidus est :

$$s=0.43$$

Le coefficient de détermination est :

$$R^2 = \frac{SCR}{SCT} = 0,22$$

Pratiquement nous avons 22 % de la fluctuation des observations de rapport de la fréquence centrale et la bande fréquentielle assemblées autour de la moyenne.

2.6 Tests d'hypothèse en régression multiple.

On veut tester l'hypothèse selon laquelle les trois variables explicatives (valeurs des magnitudes et les conditions de site) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\frac{f^s}{f^0}$.

Les hypothèses nulle et alternative que l'on désire tester dans le cadre d'un modèle de régression à 3 variables sont :

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$$

H1 : Au moins un des β_j est différent de 0.

On aura donc recours au rapport des carrés moyens pour effectuer le test :

$$\text{Quotient des carrés moyens : } F = \frac{CMR}{CM_{RES}}$$

Pour $k=3$, $n-k-1=36$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;3,36}=0,0075$

Règle de décision : Rejeter H_0 si $F > F_{\alpha; k, n-k-1}$

Du tableau d'analyse de variance on a :

$$CMR = 0.502 \quad CM_{RES} = 0.129$$

La valeur observée pour le rapport F est donc 4.648

Puisque $F = 4.648 > 0.0075$ nous rejetons l'hypothèse H_0 et favorisons H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil $\alpha=0.05$, qu'au moins une des variables explicatives sinon les trois, apportent une contribution significative dans l'explications du nombre de pièces assemblées. Il y a au moins un des paramètres β_j qui est significativement différent de zéro.

2.7 Test sur la contribution marginale de chaque variable explicative avec le *t* de Student.

Dans une étude de régression multiple portant sur l'analyse de facteurs pouvant être liés à la variation du rapport, on a obtenu les coefficients de régression et les erreurs-types de chaque coefficient.

Variables explicatives	Coefficients de régression b_j	Erreurs-types $s(b_j)$
Coefficient b_0	0.312	0.149
M	-0.046	0.0246
S1	-0.183	0.0731
S2	-0.221	0.074

Testons la contribution marginale de chaque variable explicative dans le modèle $\frac{f^s}{f^0} = \beta_0 + \beta_1.M + \beta_2.S_1 + \beta_3.S_2 \pm \sigma$.

On utilisera donc le t de *Student* pour effectuer le test et un seuil de signification $\alpha = 0.05$, la valeur critique est donc :

	Coefficients	Erreur-type	Statistique t	Limite inférieure pour seuil de confiance = 95%	Limite supérieure pour seuil de confiance = 95%
Constante	0,312059	0,14904	2,093754	0,00978	0,614333
Variable M	0,0464237	0,02466	1,882401	-0,00359	0,096440
Variable S1	0,1830857	0,07318	2,501709	0,03466	0,331510
Variable S2	0,2214553	0,07445	2,974408	0,07045	0,372454

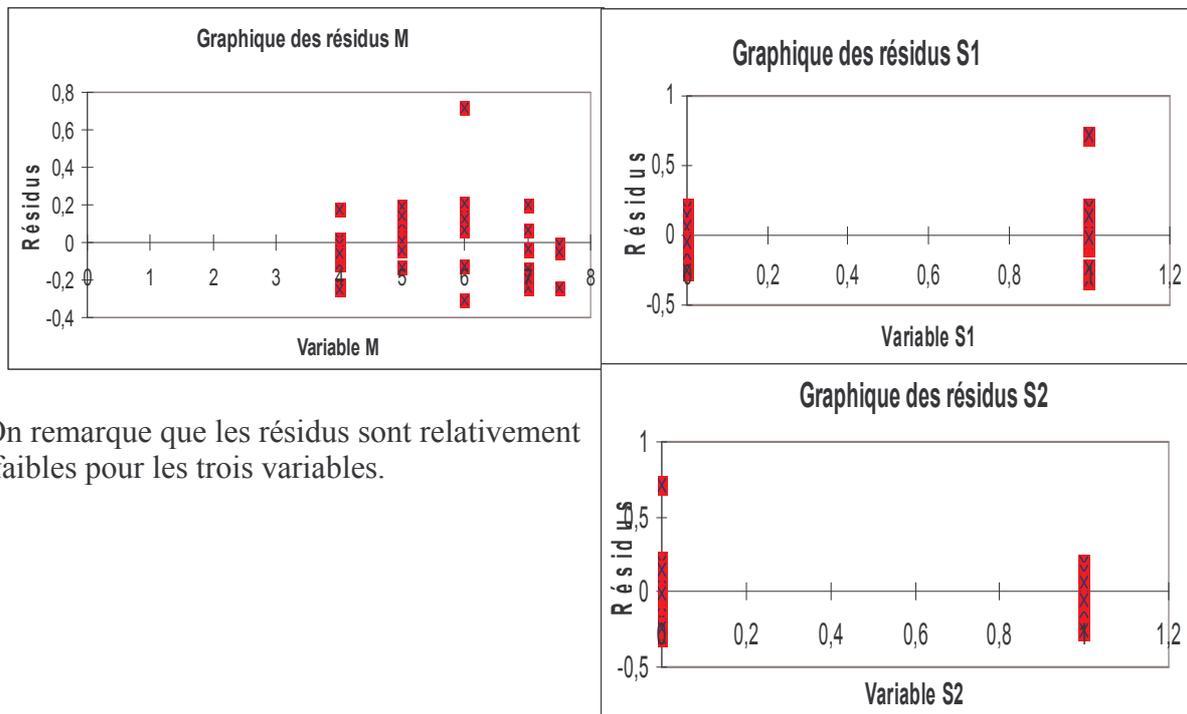
D'après le tableau ci-dessus on remarque que les variables indépendantes sont significatives sauf la variable M.

L'équation qui résulte de la régression est :

$$(f^s / f^0) = 0,312 - 0,046.M - 0,183.S1 - 0,221.S2$$

2.8 Courbes de régression et de résidus

Dans la figure ci-dessous nous présentons les graphes des résidus pour chaque variable.



On remarque que les résidus sont relativement faibles pour les trois variables.

3. La régression sur l'intensité d'Arias.

3.1 Spécification de modèle

Un modèle à quatre variables explicatives que l'on pourrait utiliser pour expliquer les fluctuations des valeurs de la variable dépendante et interpréter chaque composante du modèle.

Le modèle à quatre variables explicatives est :

$$\ln(I_a) = \beta_0 + \beta_1 \cdot R_a + \beta_2 \cdot M + \beta_3 \cdot S_1 + \beta_4 \cdot S_2 \pm \varepsilon$$

Où R_a : représente la distance.

M : représente la magnitude.

S_1, S_2 : représente le site.

β_0 : théoriquement ce paramètre représente la moyenne des $\ln(I_a)$ lorsque la valeur de variation de chaque variable explicative est égale à zéro. Toutefois, étant donné le champ de variation de chaque variable de ce contexte, on ne peut donner de signification concrète à β_0 n'ayant pas données dans le voisinage de zéro.

ε_i : ce terme dénote la fluctuation aléatoire non observable attribuable à l'omission dans le modèle d'autres variables explicatives ou à des composantes non identifiées dans le modèle. on le suppose distribué normalement et indépendamment avec moyenne 0 et variance σ^2 .

3.2 Estimation des paramètres du modèle : calcul des coefficients.

Les estimateurs des paramètres du modèle sont :

Paramètres	Estimateurs
β_0	b_0
β_1	b_1
β_2	b_2
β_3	b_3
β_4	b_4

Les valeurs prises par ces estimateurs sont les coefficients de l'équation de régression. Nous aurons recours au logiciel pour le calcul de ces coefficients.

Nous résumons dans le tableau ci contre les diverses valeurs des coefficients obtenues.

Coefficients de régression
$b_0=5.55$
$b_1=-2.12$
$b_2=1.73$
$b_3=-0.618$
$b_4=-1.206$

3.3 Équation de régression multiple.

L'équation de régression multiple est donc :

$$\ln(I_a) = 5,55 - 2,120.\ln(R_a) + 1,734.M + 0,618.S_1 - 1,206.S_2$$

3.4 Détermination du tableau de variance.

Tableau d'analyse de variance :

Source de variance	Somme de carrés	Degrés de liberté	Carrés moyens
Expliquée par la régression	$SCR = 93.91$	4	$CMR = 23.478$
	$SC_{RES} = 18.48$	35	$CM_{RES} = 0.528$
résiduelle			
Totale	$SCT = 112.40$	39	

3.5 Détermination de la variance et du coefficient R^2 .

Du tableau d'analyse de variance, on peut en déduire que :

$$s^2 = CM_{RES} = 0.528$$

L'écart-type des résidus est :

$$s=0.72$$

Le coefficient de détermination est :

$$R^2 = \frac{SCR}{SCT} = 0,83$$

Pratiquement nous avons 83 % de la fluctuation des observations de l'intensité d'Arias assemblées autour de la moyenne.

3.6 Tests d'hypothèse en régression multiple.

On veut tester l'hypothèse selon laquelle les trois variables explicatives (valeurs des magnitudes, la distance et les conditions de site) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\ln(I_a)$.

Les hypothèses nulle et alternative que l'on désire tester dans le cadre d'un modèle de régression à 4 variables sont :

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \text{Au moins un des } \beta_j \text{ est différent de 0.}$$

On aura donc recours au rapport des carrés moyens pour effectuer le test :

$$\text{Quotient des carrés moyens : } F = \frac{CMR}{CM_{RES}}$$

Pour $k=3$, $n-k-1=35$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;3,35}=2,9934$

Règle de décision : Rejeter H_0 si $F > F_{\alpha;k, n-k-1}$

Du tableau d'analyse de variance on a :

$$CMR = 23.478 \quad CM_{res} = 0.528$$

La valeur observée pour le rapport F est donc 44.444

Puisque $F = 44.44 > 2.9934$ nous rejetons l'hypothèse H_0 et favorisons H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil $\alpha=0.05$, qu'au moins une des variables explicatives sinon les trois, apportent une contribution significative dans l'explications du nombre de pièces assemblées. Il y a au moins un des paramètres β_j qui est significativement différent de zéro.

3.7 Test sur la contribution marginale de chaque variable explicative avec le t de

Student.

Dans une étude de régression multiple portant sur l'analyse de facteurs pouvant être liés à la variation de l'intensité d'Arias, on a obtenues les coefficients de régression et les erreurs-types de chaque coefficient.

Variabes explicatives	Coefficients de régression b_j	Erreurs-types $s(b_j)$
Coefficient b_0	5.557	0.85
M	1.73	0.16
S1	1.20	0.29
S2	0.61	0.28
$\ln(R_a)$	-2.12	0.4199

Testons la contribution marginale de chaque variable explicative dans le modèle $\ln(I_a) = \beta_0 + \beta_1 \cdot R_a + \beta_2 \cdot M + \beta_3 \cdot S_1 + \beta_4 \cdot S_2 \pm \varepsilon$.

On utilisera donc le *t* de *Student* pour effectuer le test et un seuil de signification $\alpha = 0.05$, les valeurs critiques sont donc :

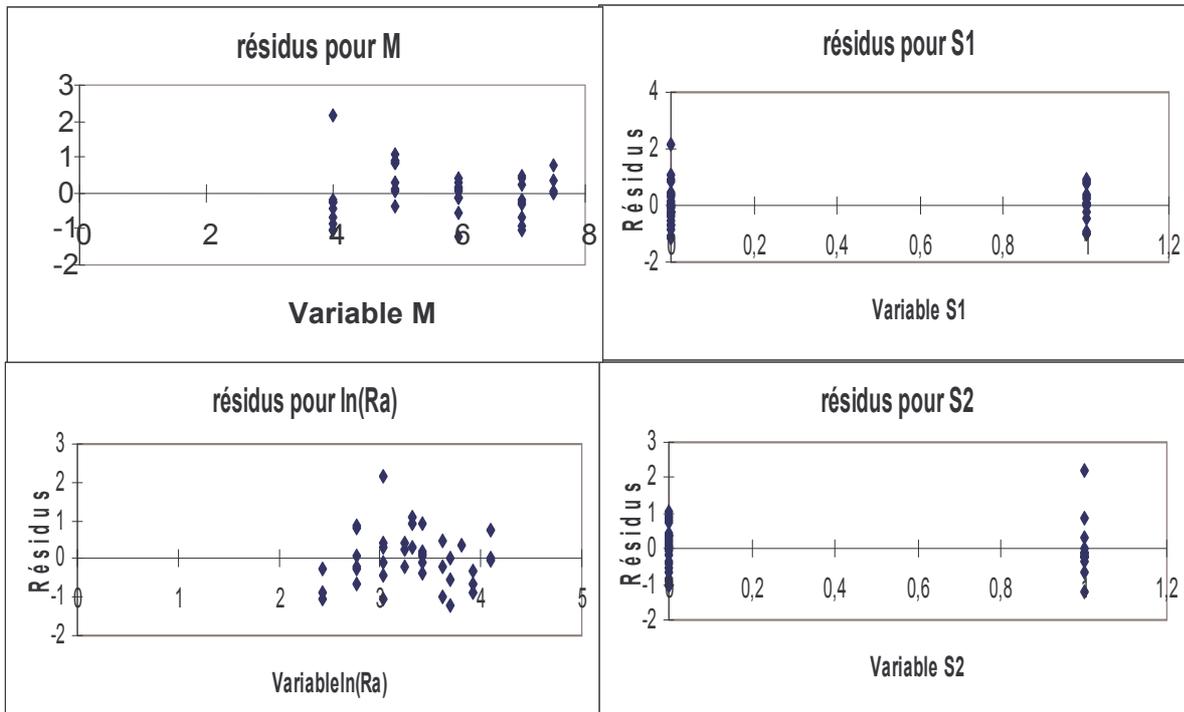
	Coefficients	Erreur-type	Statistique <i>t</i>	Limite inférieure pour seuil de confiance = 95%	Limite supérieure pour seuil de confiance = 95%
Constante	5,55704774	0,85689364	6,4851079	3,81746117	7,29663431
Variable M	1,73434316	0,16001501	10,8386278	1,40949542	2,0591909
Variable S1	1,20688966	0,29145446	4,14092022	0,61520566	1,79857366
Variable S2	0,61831649	0,28441976	2,17395758	0,04091369	1,19571929
Variable $\ln(R_a)$	-2,12024319	0,41997096	-5,0485472	-2,97282955	-1,26765683

L'équation devient :

$$\ln(I_a) = 5,55 - 2,120 \cdot \ln(R_a) + 1,734 \cdot M + 0,618 \cdot S_1 - 1,206 \cdot S_2$$

3.8 Résidus :

L'intensité d'Arias est dépendante de quatre variables dont les résidus sont illustrés ci-dessous.



4. Analyse de régression sur la durée de mouvement fort.

4.1 Spécification du modèle.

Un modèle à quatre variables explicatives que l'on pourrait utiliser pour expliquer les fluctuations des valeurs de la variable dépendante et interpréter chaque composante du modèle.

Le modèle à quatre variables explicatives est :

$$\ln(D_v) = \beta_0 + \beta_1 \cdot R_v + \beta_2 \cdot M + \beta_3 \cdot S_1 + \beta_4 \cdot S_2 \pm \varepsilon$$

Où R_v : représente la distance.

M : représente la magnitude.

S_1, S_2 : représente le site.

β_0 : théoriquement ce paramètre représente la moyenne des $\ln(D_v)$ lorsque la valeur de chaque variable explicative est égale à zéro. Toutefois, étant donné le champ de variation de chaque variable de ce contexte, on ne peut donner de signification concrète à β_0 n'ayant pas de données dans le voisinage de zéro.

ε_i : ce terme dénote la fluctuation aléatoire non observable attribuable à l'omission dans le modèle d'autres variables explicatives ou à des composantes non identifiées dans le modèle. On le suppose distribué normalement et indépendamment avec une moyenne 0 et une variance σ^2 .

4.2 Estimation des paramètres du modèle : calcul des coefficients.

Les estimateurs des paramètres du modèle sont :

Paramètres	Estimateurs
β_0	b_0
β_1	b_1
β_2	b_2
β_3	b_3
β_4	b_4

Les valeurs prises par ces estimateurs sont les coefficients de l'équation de régression. Nous aurons recours au logiciel pour le calcul de ces coefficients.

Nous résumons dans le tableau ci contre les diverses valeurs des coefficients obtenues.

Coefficients de régression
$b_0 = -0.908$
$b_1 = 0.922$
$b_2 = 0.072$
$b_3 = 0.255$
$b_4 = 0.248$

4.3 Équation de régression multiple.

L'équation de régression multiple est donc :

$$\ln(D_v) = -0,908 + 0,922 \cdot \ln(R_v) + 0,072 \cdot M + 0,255 \cdot S1 + 0,248 \cdot S2$$

4.4 Détermination du tableau de variance.

Tableau d'analyse de variance :

Source de variance	Somme de carrés	Degrés de liberté	Carrés moyens
Expliquée par la régression	$SCR = 11.24$	4	$CMR = 2.811$
	$SC_{RES} = 3.805$	35	$CM_{RES} = 0.108$
résiduelle			
Totale	$SCT = 15.04$	39	

4.5 Détermination de la variance et du coefficient R^2 .

Du tableau d'analyse de variance, on peut en déduire que :

$$s^2 = CM_{RES} = 0.108$$

L'écart-type des résidus est :

$$s = 0.32$$

Le coefficient de détermination est :

$$R^2 = \frac{SCR}{SCT} = 0,74$$

Pratiquement nous avons 74 % de la fluctuation des observations de la durée de mouvement fort assemblées autour de la moyenne.

4.6 Tests d'hypothèse en régression multiple.

On veut tester l'hypothèse selon laquelle les trois variables explicatives (valeurs des magnitudes, la distance et les conditions de site) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\ln(D_v)$.

Les hypothèses nulle et alternative que l'on désire tester dans le cadre d'un modèle de régression à 4 variables sont :

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \text{Au moins un des } \beta_j \text{ est différent de } 0.$$

On aura donc recours au rapport des carrés moyens pour effectuer le test :

$$\text{Quotient des carrés moyens : } F = \frac{CMR}{CM_{RES}}$$

Pour $k=3$, $n-k-1=35$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;3,35} = 4,9934 \cdot 10^{-10}$

Règle de décision : Rejeter H_0 si $F > F_{\alpha;k, n-k-1}$

Du tableau d'analyse de variance on a :

$$CMR = 2.811 \quad CM_{res} = 0.108$$

La valeur observée pour le rapport F est donc 25.85

Puisque $F = 25.85 > 4,9934 \cdot 10^{-10}$ nous rejetons l'hypothèse H_0 et favorisons H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil $\alpha=0.05$, qu'au moins une des variables explicatives sinon les trois, apportent une contribution significative dans l'explications du nombre de pièces assemblées. Il y a au moins un des paramètres β_j qui est significativement différent de zéro.

4.7 Test sur la contribution marginale de chaque variable explicative avec le t de

Student.

Dans une étude de régression multiple portant sur l'analyse de facteurs pouvant être liés à la variation de la durée de mouvement fort, on a obtenues les coefficients de régression et les erreurs-types de chaque coefficient.

Variabes explicatives	Coefficients de régression b_j	Erreurs-types $s(b_j)$
Coefficient b_0	-0.90	0.38
M	0.072	0.07
S1	0.255	0.13
S2	0.248	0.13
Ln(Rv)	0.92	0.19

Testons la contribution marginale de chaque variable explicative dans le modèle $\ln(D_v) = \beta_0 + \beta_1 \cdot R_v + \beta_2 \cdot M + \beta_3 \cdot S_1 + \beta_4 \cdot S_2 \pm \varepsilon$.

On utilisera donc le t de *Student* pour effectuer le test et un seuil de signification $\alpha = 0.05$, la valeurs critique est donc :

Le tableau ci-dessous contient les valeurs des coefficients des variables et le teste de signification de chaque un.

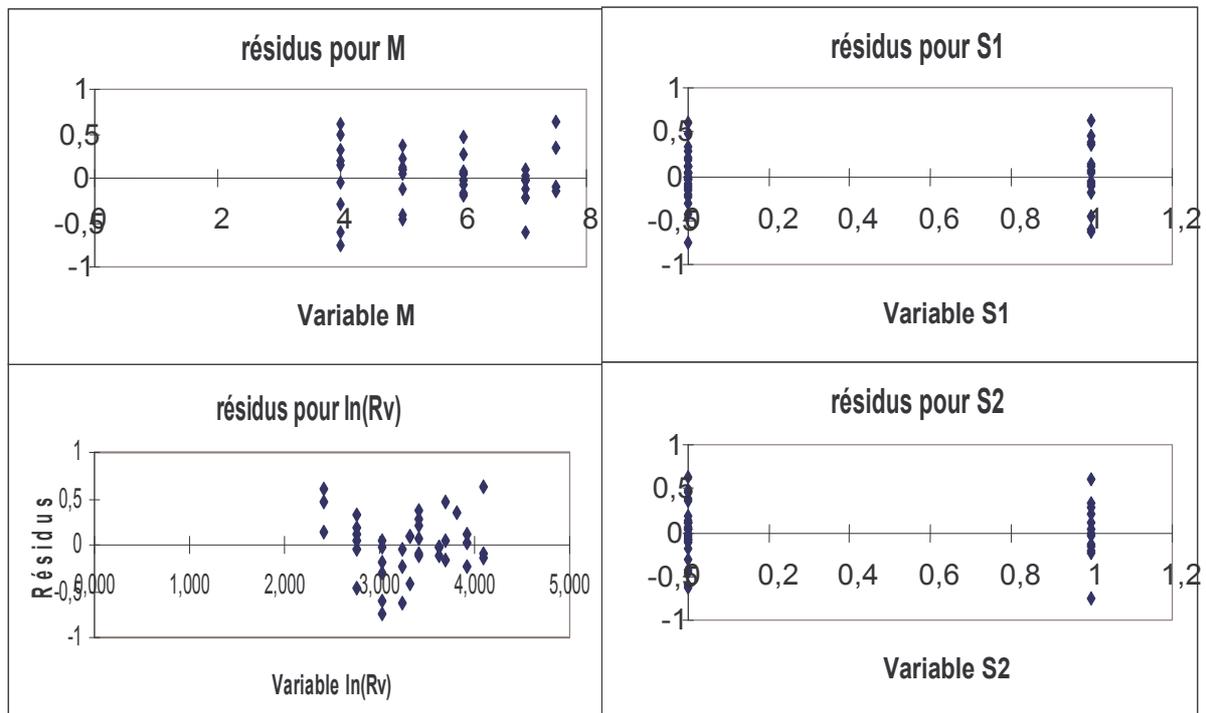
	Coefficients	Erreur-type	Statistique t	Limite inférieure pour seuil de confiance = 95%	Limite supérieure pour seuil de confiance = 95%
Constante	-0,90856218	0,38652806	-2,35057241	-1,69325585	-0,12386852
Variable M	0,07246519	0,07258187	0,99839236	-0,07488385	0,21981422
Variable S1	0,25540605	0,12903251	1,97939294	-0,00654388	0,51735598
Variable S2	0,24813158	0,13222557	1,87657785	-0,0203006	0,51656377
Variable ln(Rv)	0,92210297	0,18972222	4,86027932	0,53694639	1,30725954

On remarque que les trois variables M, S1 et S2 sont légèrement non significatives
L'équation devient :

$$\ln(D_v) = -0,908 + 0,922 \cdot \ln(R_v) + 0,072 \cdot M + 0,255 \cdot S_1 + 0,248 \cdot S_2$$

4.8 Résidus.

La figure ci-dessous présente les résidus pour la durée du mouvement fort D_v



5. Illustration des accélérogrammes simulé et les spectres de réponse :

Dans les figures ci-dessus nous présenterons à titre illustratives des accélérogrammes simulé qui indique l'utilisations des lois de régressions développer précédemment.

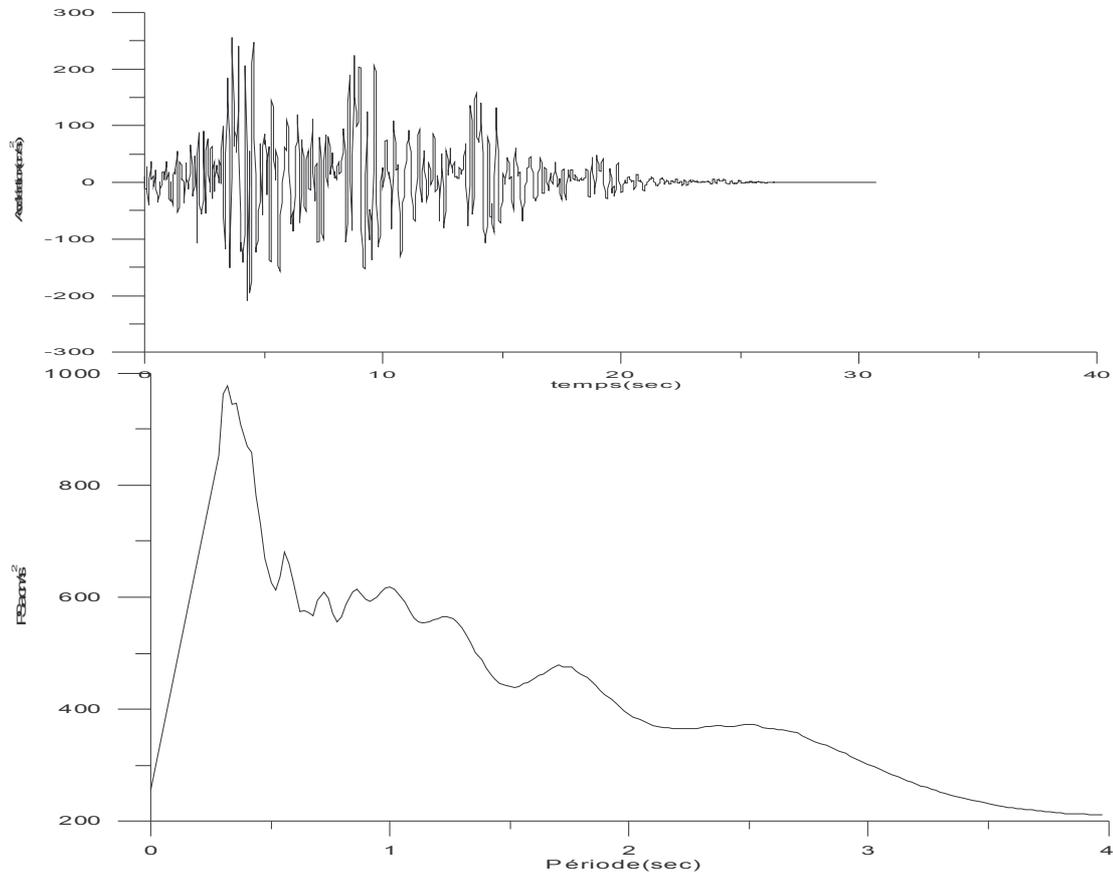
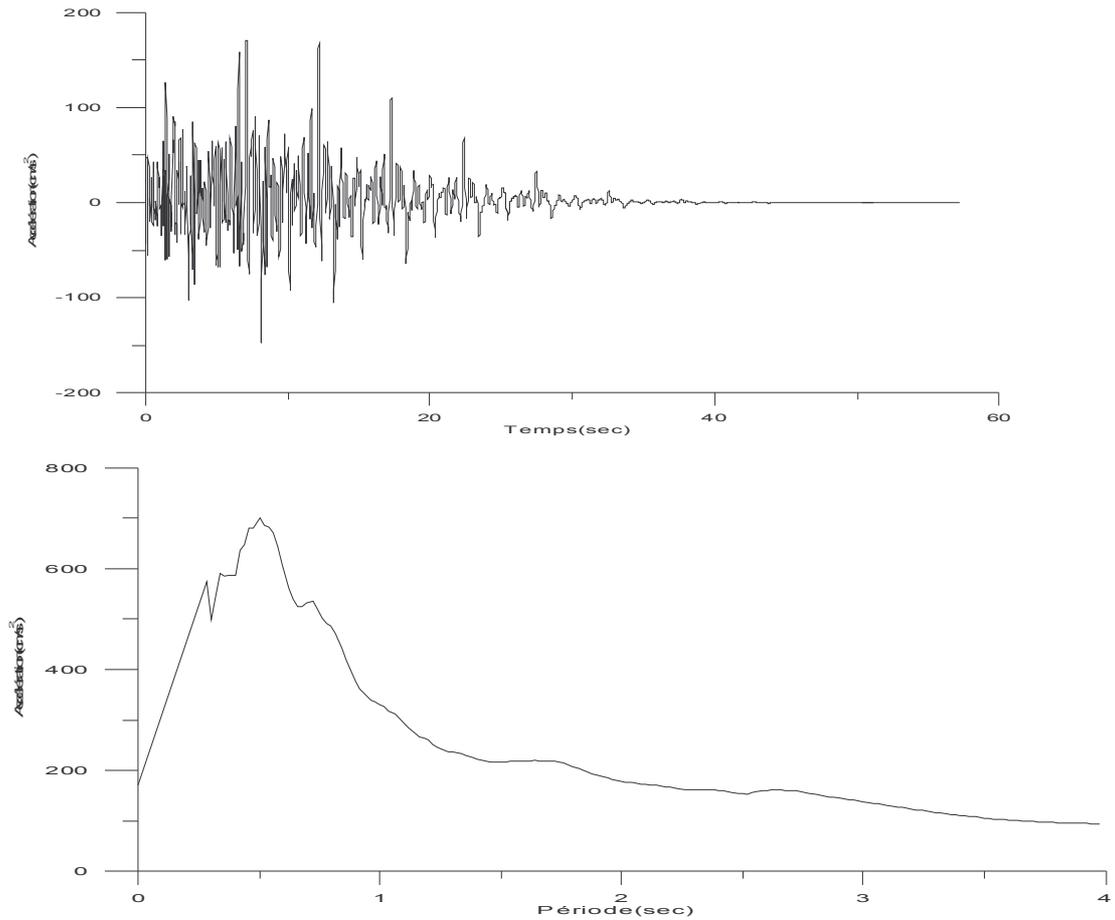


Figure 14 Accélérogramme simulé pour $M=7$, $R=37$, Sol raide
Spectre de réponse en accélération



**Figure 15 Accélérogramme simulé pour M=7, R=37, dépôts profond
Spectre de réponse en accélération**

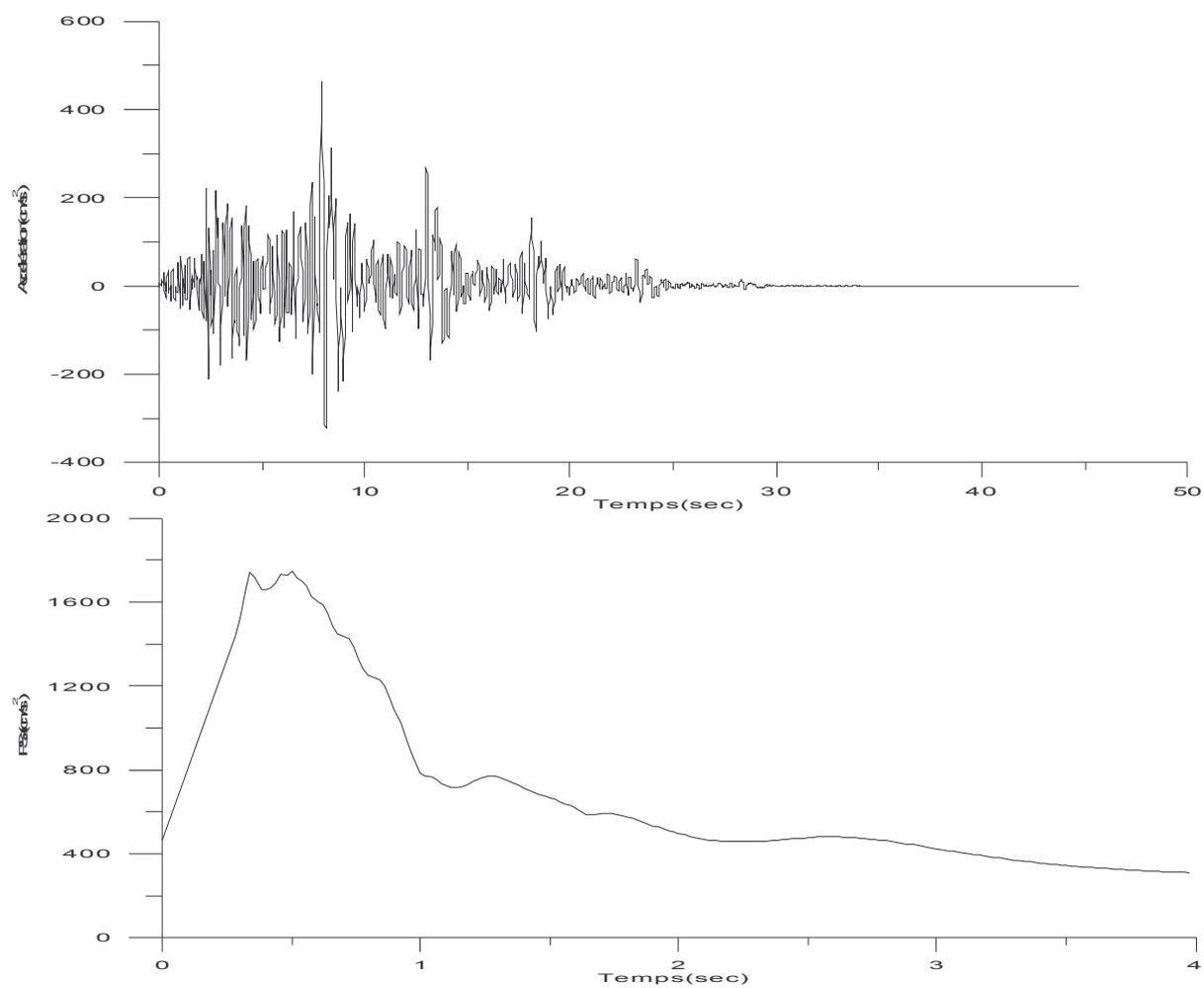


Figure 16 Accélérogramme simulé pour $M=7$, $R=37$, dépôts superficielle
Spectre de réponse en accélération

Conclusion.

Une analyse par régression multilinéaire est utilisée en vue de l'estimation des caractéristiques stochastiques des mouvements sismiques fort.

Quatre lois prédictives des paramètres spectraux, de l'intensité d'Arias et de durée de mouvement fort sont développées en utilisant une banque de données accélérométrique; les variables indépendantes de régression utilisées étant la magnitude, la distance épacentrale et les conditions locales de site.

A titre illustratif, sont également présentés les résultats de simulation pour diverses classes de données correspondant à des choix spécifique de variables indépendantes.

Il a été démontré entre autre.

1. les quatre lois de régression obtenues pour chacune des variables dépendants étudié en utilisant la banque de donnée mondiale se sont révélées statistiquement acceptables.

2. les contributions marginales des variables explicatives des quarts modèles étudiées sont généralement significatives dans l'ensemble.

3. la contribution marginale de la variable indépendante distance pour les modèles de régression. Concernant les paramètres spectraux $f^{\theta}(t), f^{\phi}(t)$ n'est pas significative (cite à dire que la variables distance peut être occultée).

4. les qualités de l'ajustement linéaire des équations de régression sont satisfaisant dans l'ensemble.

5. les tests de signification sur l'ensemble des paramètres indépendants (test de fisher) ont montré que la contribution de l'ensemble des variables pour expliquer chacun des variables dépendantes étaient significatives avec une probabilité de confiance de 95%.

6. les simulations d'accélération sismique présentées à titre illustratif montre à priori que les tendances des variables indépendantes sont respectées.

Un travail concernant la simulation des mouvements sismique forts reste à faire. Les modèles de régressions développées dans le présent travail peuvent être utilisés pour la définition des paramètres à prendre en compte dans la procédure stochastique de simulation.

Bibliographie

- [1] Aknouche H., *Identification et analyse des effets des nonstationarités des mouvements sismiques sur la réponse des structures*, Thèse de Magister, École Nationale Polytechnique, Alger, 1999.
- [2] Alain P., *Dynamique des sols*, Presses de l'école nationale des Ponts et chaussées, Paris, 1984.
- [3] Arias A., *A measure of earthquake intensity*, in *Seismic Design of Nuclear Power Plants*, R. Hansen (Editor) M.I.T. Press, Cambridge, 1970.
- [4] Baillargeon G., *Probabilités statistiques et techniques de régression*, les éditions SMG, Québec, 1989.
- [5] Hammoutene M., Tilioune B., Brad P. Y., *A two dimensional nonstationary optimized accelerogram scaled for magnitude, distance and soil conditions*, Proc. 10th World Conference on Earthquake Engineering, Madrid, 1992, 817-821.
- [6] Hammoutene M., *Simulation Nonstationnaire et Bidimensionnelle des Enregistrements des Tremblements de Terre*, Thèse d'États, École Nationale Polytechnique, Alger, 1995.
- [7] Mari J.-L., Glangeaud F., Coppens F., *Traitement du signal pour géologues et géophysiciens, Technique de base : Vol 2*, Publications de l'institut Français du pétrole, Editions TECHIP, Paris, 2001.
- [8] Mark W.D., *Power Spectrum Representation for Nonstationary Random Vibration*, Studies in Applied Mechanics 14, Random vibrations-Status and Recent Developments, The S. H Crandall Festschrift, 1986, 211-240.
- [9] Mark, W.D., *Spectral Analysis of the Convolution and Filtering of Nonstationary Stochastic Processes*, Journal of Sound Vibration, VOL. 11, N° 1, 1970, 19-63.
- [10] Mc Cann M.W., Shah H. C., *Determining Strong-Motion Duration of Earthquake*, Bulletin of the Seismological Society of America, VOL. 69, N° 4, 1979, 1253-1265.
- [11] Pousse G., Bonilla F., Cotton F., Rey J., *Génération stochastique d'accélérogrammes synthétiques, validation de méthodologie*, AFPS.
- [12] Preumont A., *Vibrations Aléatoires et Analyse Spectrale*, Presses Polytechniques et Universitaires Romandes, Laussane, 1990.
- [13] Priestly, M. B., *Evolutionary Spectra and Nonstationary Processes*, J. Royal Statist, Soc., Ser. B., N° 2, 1965, 204-237.
- [14] Sabetta F., Pugliese A., *Estimation of response spectra and simulation of nonstationary earthquake ground motions*, BSSA, VOL. 86, 1996, 337-352.
- [15] Ventsel H., *Théorie des Probabilités*, Traductions Française Éditions MIR, Moscou, 1973.