

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique



Ecole Nationale Polytechnique
Département Génie Civil
Laboratoire de Génie Sismique et Dynamique des Structure



Mémoire de magister

Présenté par

Hamza DJOUABI

Ingénieur d'état en Génie Civil de l'École Nationale Polytechnique

THÈME

LOIS PREDICTIVES POUR
L'ESTIMATION DE LA REPONSE
SISMIQUE ET DES MOUVEMENTS
SISMIQUES FORTS

Soutenu publiquement le 06 janvier 2010 à 9h :00 (amphi 5A, génie civil)
devant le jury composé de :

Président du jury :	B. TILIOUINE	Professeur, E. N. P.
Directeur du mémoire :	M. HAMMOUTENE	Professeur, E. N. P.
Examineurs :	M. N. FARSI	Directeur de Recherche, C. G. S.
	M. BENZAIBI	Maître de Conférences, U. Blida.
	H. MOHEBEDDINE	Chargée de Cours, E. N. P.

10, Avenue Hassen Badi, BP 182, 16200 ,El-Harrach, Alger, Algérie.

A mes parents

A ma famille

A tous mes amis.

Remerciements.

Tout d'abord, je remercie Allah, le Clément et le Miséricordieux de nous avoir donné la force et le courage de mener à terme ce modeste travail de recherche.

Le présent travail n'aurait pu être mis au point sans l'aide et les orientations prodiguées par mon directeur de mémoire, monsieur le professeur M. **HAMMOUTENE**, envers lequel j'adresse tous mes remerciements et ma profonde reconnaissance.

J'exprime mes vifs remerciements à monsieur le Professeur B. **TILIOUINE**, pour les conseils éclairés et très pertinentes qu'il m'a prodigué tout au long de notre travail, ainsi que pour l'honneur qu'il me fait en présider le jury de soutenance de mon Mémoire de Magister, qu'il trouve ici l'expression de ma profonde gratitude.

Je remercie également madame H. **MOHIBEDDINE** et messieurs M. N. **FARSI** et M. **BENSAIBI** d'avoir accepté de faire partie de mon jury, et d'avoir apporté des commentaires et des critiques très constructives et qui ont grandement contribué à améliorer la qualité de mon travail de recherche tant dans la forme que dans le fond de.

Je remercie chaleureusement tous les membres de l'équipe du laboratoire LGSDS pour l'esprit d'entraide qu'on y trouve, et pour l'ambiance sympathique et scientifique qu'ils perpétuent.

À tous ceux qui ont contribué de près ou de loin à ce modeste travail, qu'ils y trouvent ici l'expression de toute ma gratitude.

Ma pensée vas à tous les enseignants qui ont contribué à ma formation depuis le primaire, ainsi qu'à tous les anonymes administratifs qui ont travaillé dans l'ombre pour qu'aujourd'hui je sois à ce stade de compétence.

Enfin, je réserve une mention de remerciements très spéciale au personnel technique du Centre National de Recherche Appliquée au Génie Parasismique (C. G. S.), notamment, monsieur M.NAILI, pour leurs aides et conseils.

Résumé :

Dans ce travail, il est formulé des lois prédictives pour l'estimation des pseudo spectres en vitesse, des pics d'accélération de sol et des durées de mouvements sismiques forts. Ces équations prédictives sont élaborées en utilisant des analyses statistiques faites sur la base d'une banque mondiale de données accélérométriques. Les résultats obtenus montrent que les lois prédictives proposées sont particulièrement intéressantes pour la prédiction réaliste de la réponse des structures et de l'intensité des mouvements forts de tremblement de terre futurs.

Mots clés : lois prédictives, pseudo spectres, pics d'accélération, durée, mouvements forts

ملخص:

في هذا العمل استعملنا طريقة التحليل التراجعي الخطي المتعدد بدف تقدير الاستجابة الطيفية للسرعة، قيمة التسارع الأقصى للتربة و مدة الحركة العنيفة . هذه العلاقات التوقعية التابعة لشدة الزلزال، المسافة البؤرية و نوعية التربة، تم استخراجها بالاعتماد على بنك بيانات عالمية تحتوي على تسجيلات لهزات أرضية عنيفة.

النتائج احصل عليها تكتسي أهمية كبيرة لاسيما فيما يخص الاستجابة الطيفية للبنىات و لتقدير الحركة الزلزالية و كذلك في تقدير الخطر الزلزالي للزلازل المتوقعة علاقات لتقدير تطورها باستعمال المتغيرات المستقلة (شدة الهزة، المسافة البؤرية و نوعية التربة) .

الكلمات المفتاحية: التحليل التراجعي، الاستجابة الطيفية، التسارع الأقصى، مدة الحركة، الحركة العنيفة.

Summary:

In this work, predictive relationships are formulated for the prediction of response spectral values (PSV), peak ground acceleration (PGA) and strong motion duration (Dv). These predictive equations are developed after statistical analysis of a worldwide strong motion accelerograms database. The obtained results show that the proposed relationships are particularly interesting for the prediction of structural responses and strong motion characteristics, and then of the seismic risk assessment of a given region during future strong earthquakes.

Key words: predictive relationships, response spectral, peak ground acceleration, strong motion, duration.

Sommaire

INTRODUCTION GÉNÉRALE.....	2
-----------------------------------	----------

CHAPITRE I : GÉNÉRALITÉS SUR LES SÉISMES.....	5
------------------------------------------------------	----------

- 1.1.	INTRODUCTION.....	5
- 1.2.	CARACTÉRISATION DES SÉISMES.	6
- 1.3.	MOUVEMENT FORT.....	7
- 1.4.	PROPAGATION DES ONDES SISMiques.....	8
	LES ONDES DE VOLUME :	8
	LES ONDES DE SURFACE :	8
- 1.5.	REPRÉSENTATION DU MOUVEMENT SISMique.	9
- 1.6.	CARACTÉRISTIQUES ENREGISTRÉS À LA SURFACE DU SOL.	10
- 1.7.	PARAMÈTRES AFFECTANT LE MOUVEMENT SISMique.....	11
	1.7.1. FACTEURS LIÉS À LA SOURCE.	11
	MAGNITUDE.....	12
	1.7.2. FACTEURS LIÉS AU TRAJET PARCOURU.	14
	1.7.3. FACTEURS LIÉS AUX CONDITIONS LOCALES.	15

CHAPITRE II : SPECTRE ET PSEUDO SPECTRE DE RÉPONSE.....	18
----------------------------------------------------------------	-----------

- 2.1.	INTRODUCTION :.....	18
- 2.2.	RÉPONSE D'UN OSCILLATEUR 1DDL	19
- 2.3.	SPECTRE DE RÉPONSE ÉLASTIQUE	20
	PRINCIPE D'ÉTABLISSEMENT.	20
- 2.4.	SPECTRE ET PSEUDO SPECTRE DE RÉPONSE :.....	22
- 2.5.	SPECTRE DE RÉPONSE DU CODE PARASISMique	27
- 2.6.	SPECTRE DE CALCUL :	29
- 2.7.	SPECTRE DE CALCUL POUR L'ANALYSE ÉLASTIQUE.....	30
	A. IMPORTANCE DE LA CONSTRUCTION.	30
	B. SÉISME PROCHE, SÉISME LOINTAIN.....	30
	C. SOLS ET SITES.	31
	D. DUCTILITÉ DE LA STRUCTURE.	32

CHAPITRE III : NOTIONS SUR LA RÉGRESSION.....	34
------------------------------------------------------	-----------

- 3.1.	ESTIMATION DES PARAMÈTRES INCONNUS D'UNE LOI DE RÉPARTITION.....	34
	3.1.1. ESTIMATIONS DE L'ESPÉRANCE MATHÉMATIQUE ET DE LA VARIANCE.....	36
- 3.2.	INTERVALLE DE CONFIANCE. PROBABILITÉ DE CONFIANCE	36
- 3.3.	TEST D'HYPOTHÈSE.....	38
	3.3.1. PRINCIPE DE TEST D'HYPOTHÈSE.....	38
	3.3.2. CONCEPTS IMPORTANTS DANS L'ÉLABORATION D'UN TEST D'HYPOTHÈSE.....	38
	SEUIL DE SIGNIFICATION D'UN TEST D'HYPOTHÈSE.....	40
	3.3.3. COMMENT EXÉCUTER UN TEST D'HYPOTHÈSE : DÉMARCHE À SUIVRE.....	41
	3.3.4. DÉMARCHE À SUIVRE DANS L'ÉLABORATION D'UN TEST D'HYPOTHÈSE.....	41
- 3.4.	RÉGRESSION LINÉAIRE SIMPLE.....	42
	3.4.1. INTRODUCTION.....	42
	3.4.2. LE MODÈLE DE LA RÉGRESSION SIMPLE.....	42

3.4.3.	CONDITION D'APPLICATION DU MODÈLE LINÉAIRE SIMPLE.	43
3.4.4.	ESTIMATION DE LA DROITE DE RÉGRESSION $E(Y) = \beta_0 + \beta_1 X$	43
- 3.5.	RÉGRESSION LINÉAIRE MULTIPLE.	44
3.5.1.	INTRODUCTION	44
3.5.2.	MODÈLE LINÉAIRE À PLUSIEURS VARIABLES EXPLICATIVES.	44
3.5.3.	HYPOTHÈSES FONDAMENTALES.	45
3.5.4.	FORMULATION D'UN MODÈLE DE RÉGRESSION MULTIPLE : IDENTIFICATION ET INTERPRÉTATION DES COMPOSANTES DU MODÈLE.	46
3.5.5.	DÉTERMINATION DE L'ÉQUATION DE RÉGRESSION : ESTIMATION DES PARAMÈTRES DU MODÈLE.	46
3.5.6.	ANALYSE DE VARIANCE EN RÉGRESSION MULTIPLE.	48
3.5.7.	TESTE D'HYPOTHÈSE EN RÉGRESSION MULTIPLE.....	50

CHAPITRE IV : BANQUE DE DONNÉES.....57

- 4.1.	INTRODUCTION.....	57
- 4.2.	ÉTUDE STATISTIQUE SUR LES ENREGISTREMENTS.[13],[14].....	57
- 4.3.	RÉPARTITION DES DONNÉES.	58
4.3.1.	CHOIX DES VARIABLES (CRITÈRES STATISTIQUE).....	59
4.3.2.	DESCRIPTION DES VARIABLES INDÉPENDANTES.	60

CHAPITRE V : PRÉSENTATION ET DISCUSSION DES RÉSULTATS POUR PSV..... 62

- 5.1.	INTRODUCTION :.....	62
- 5.2.	FORMULATION DE MODÈLE :.....	62
- 5.3.	L'ÉQUATION PRÉDICTIVE :	63
- 5.4.	DISCUSSION DES RÉSULTATS DE LA RÉGRESSION :.....	65
- 5.5.	TEST D'HYPOTHÈSE :.....	65
	SIGNIFICATION DE LA RÉGRESSION DANS SON ENSEMBLE (FISHER) :	65
- 5.6.	COMPARAISON ENTRE LES SPECTRES SIMULÉS ET CEUX DES LOIS PRÉDICTIVES:.....	75
- 5.7.	EFFET DE LA MAGNITUDE :	76
- 5.8.	EFFET DE DISTANCE :	76
- 5.9.	EFFET DE SITE :.....	77
- 5.10.	COMPARAISON AUX RÈGLEMENTS.	82
- A.	COMPARAISON ENTRE L'EUROCODE ET LA PRÉSENTE ÉTUDE	82
- B.	COMPARAISON ENTRE L'RPA ET LA PRÉSENTE ÉTUDE :	82
- C.	COMPARAISON ENTRE L'UBC ET LA PRÉSENTE ÉTUDE :	83

CHAPITRE VI : ACCÉLÉRATION MAXIMALE DE SOL 88

- 6.1.	INTRODUCTION :.....	88
- 6.2.	SPÉCIFICATION DE MODÈLE.	89
- 6.3.	DISCUTIONS DES RÉSULTATS DE RÉGRESSIONS :.....	89
6.3.1.	ANALYSE DE VARIANCE :	89
6.3.2.	ESTIMATION DES PARAMÈTRES DU MODÈLE :.....	90
6.3.3.	DÉTERMINATION DU TABLEAU DE VARIANCE.	91
6.3.4.	DÉTERMINATION DE LA VARIANCE ET DU COEFFICIENT R^2	91
6.3.5.	TESTS D'HYPOTHÈSE DE LA RÉGRESSION.	91
6.3.6.	TEST SUR LA CONTRIBUTION MARGINALE DE CHAQUE VARIABLE EXPLICATIVE AVEC LE T DE STUDENT.....	92

- 6.4.	EFFET DE LA MAGNITUDE, DISTANCE ET SITE SUR LE PGA :	92
--------	------------------------------------------------------------	----

CHAPITRE VII : DURÉE DE MOUVEMENT FORT95

- 7.1.	. INTRODUCTION :	95
- 7.2.	SPÉCIFICATION DU MODÈLE.	96
- 7.3.	DISCUTIONS DES RÉSULTATS DE LA RÉGRESSION :	97
	ANALYSE DE VARIANCE :	97
- 7.4.	TEST D'HYPOTHÈSE :	97
	SIGNIFICATION DE LA RÉGRESSION DANS SON ENSEMBLE (FISHER) :	97
	TEST DE LA CONTRIBUTION MARGINALE DE CHAQUE VARIABLE EXPLICATIVE :	98
- 7.5.	EFFET DE LA MAGNITUDE ET LA DISTANCE ÉPICENTRALE SUR DV :	100
- 8.	CONCLUSIONS.	103

BIBLIOGRAPHIE.....106

Liste des figures :

Figure (1.1) : Carte du monde montrant les principales plaques tectoniques

Figure (2.1) : Les ondes sismiques de volume et de surface

Figure (3.1) : Accélérogramme

Figure (1.4) : Schématisation d'un plan de faille (Brune 1976)

Figure (1.5) : Spectres moyens d'accélération pour différentes conditions de site selon Seed et al. (1976)

Figure (2.1) : Oscillateur simple soumis à une excitation sismique

Figure (2.5) : Construction d'un spectre de réponse élastique en accélération correspondant à un accélérogramme donné et définition d'une Force statique F_{max} « équivalente » au séisme.

Figure (2.6) : illustration du calcul du spectre de déplacement pour l'accélérogramme d'El Centro

Figure (2.7) : Les spectres de réponse en déplacement, en vitesse et en accélération pour l'accélérogramme d'El Centro sont présentés pour un taux d'amortissement $\xi = 0.02$

Figure (2.8) : Diagramme tripartite représentant les spectres de réponses combinés PSD-PSV-PSA de l'accélérogramme d'EL Centro pour $\xi = 0, 2, 5, 10$ et 20%

Figure (2.9) : Définition d'un spectre « moyen » $S_e(T)$ de code parasismique.

Figure (2.10) : Le spectre de réponse élastique $S_e(T)$ de l'Eurocode 8

Figure (2.11) : l'intensité spectrale selon Housner

Figure (2.12) : Spectres de calcul $S_e(T)$ de l'Eurocode 8 pour les tremblements de terre

Figure (3.1) Intervalle de confiance

Figure (5.1) : Comparaison de courbes simulées et courbes prédictives

Figure (5.2) : effet de la magnitude sur le PSV pour différentes classes site-distance

Figure (5.3) : effet de la distance épacentrale sur le PSV pour différentes classes site-magnitude

Figure (5.4) : Effet du site géologique sur le PSV pour différentes classes magnitude-distance

Figure (5.5) : Comparaison des spectres issus de la présente étude et ceux de l'Eurocode

Figure (5.6) : Comparaison des spectres issus de la présente étude et ceux de l'RPA

Figure (5.7) : Comparaison des spectres issus de la présente étude et ceux de la simulation.

Figure (6.1) : Variation de PGA en fonction de différentes classes (SDM)

Figure (7.1): effet de la magnitude sur la durée de mouvement fort

Figure (7.2): effet de la distance épacentrale sur la durée de mouvement fort

Liste des tableaux :

Tableau (4.1) : classes des distances

Tableau (4.2) : répartition de la banque de données

Tableau 3 : symboles utilisés pour caractériser les estimateurs

Tableau (5.1) : Résultats de régression pour le point $f=0.25\text{hz}$

Tableau (5.2) : Résultats de régression pour le point $f=0.33\text{hz}$

Tableau (5.3) : Résultats de régression pour le point $f=0.66\text{hz}$

Tableau (5.4) : Résultats de régression pour le point $f=1\text{hz}$

Tableau (5.5) : Résultats de régression pour le point $f=1.33\text{hz}$

Tableau (5.6) : Résultats de régression pour le point $f=1.5\text{hz}$

Tableau (5.7) : Résultats de régression pour le point $f=2\text{hz}$

Tableau (5.8) : Résultats de régression pour le point $f=2.5\text{hz}$

Tableau (5.9) : Résultats de régression pour le point $f=3.33\text{hz}$

Tableau (5.10) : Résultats de régression pour le point $f=5\text{hz}$

Tableau (5.11) : Résultats de régression pour le point $f=6.66\text{hz}$

Tableau (5.12) : Résultats de régression pour le point $f=10\text{hz}$

Tableau (5.13) : Résultats de régression pour le point $f=15\text{hz}$

Tableau (5.14) : Résultats de régression pour le point $f=25\text{hz}$

Tableau (5.15) : Résultats de la régression pour le pseudo spectre de vitesse

Tableau (6.1) : Coefficients de régression pour le PGA

Tableau (6.2) : Analyse de variance pour le PGA

Tableau (7.1) : résultats de la régression pour la durée du mouvement fort

INTRODUCTION GÉNÉRALE



Introduction générale.

L'étude du risque sismique d'une région donnée requiert une estimation plus au moins réaliste des variations du niveau d'intensité des mouvements forts du sol que provoqueraient des tremblements de terre futurs. Ces variations reflètent des variations de paramètres environnementaux.

Le niveau d'excitation des mouvements forts du sol reflète la mesure de l'effet moyen et des variations de toutes les variables qui contrôlent la génération et la propagation des ondes sismiques ainsi que la réponse des dépôts sédimentaires éventuels au site de construction.

Durant les dernières années, plusieurs auteurs (Abrahamson et Silva 1997, Boore et al 2003...) ont proposé des méthodes plus ou moins sophistiquées pour estimer de manière assez réaliste les mouvements sismiques sur la base d'une paramétrisation stochastique tant dans le domaine temporel que dans le domaine fréquentiel.

Des estimations fiables des paramètres caractéristiques aléatoires sont obtenues principalement à travers des relations prédictives empiriques développées sur la base d'analyses statistiques de données enregistrées durant les séismes passés.

Ces données doivent permettre la séparation des effets des conditions locales de site (S), de source (M) et de propagation (D), par des méthodes empiriques en vue de l'amélioration des procédures de contrôle du hazard sismique dans les zones de subduction des tremblements de terre. De nombreux auteurs ont ainsi proposé des lois d'atténuation, pour l'estimation des indicateurs de mouvements sismiques établies à partir d'analyses de régression effectuées sur divers banques de données d'enregistrement de mouvements forts (Pugliese and Sabetta, 1989, 1993 ; Pugliese and Sabetta, 1994 ; chang et al 2001 ; liu et Tsai 2005 ; Po-shen lin., chyi-tyi lee 2008 ; M. Massa 2008). Avec l'augmentation du nombre et de la qualité de ces enregistrements sismiques, la fiabilité de ces lois d'atténuation a été sans cesse améliorée.

Dans ce travail, on présentera la description complète des analyses statistiques, qui ont été faites sur la base d'une banque de données mondiale d'enregistrements accélérométriques, dans le but d'établir des lois prédictives pour l'estimation du niveau d'excitations sismiques traduit par le pseudo spectre de vitesse (PSV), des pics d'accélération de sol (PGA) ainsi que des durées de phases de mouvements forts du sol (Dv). La banque de données accélérométrique comprenant 411 accélérogrammes a été divisée en 45 classes de magnitude, distance épacentrale et condition local de site (M. Hammoutène 1995).

Les analyses statistiques sont basées sur des techniques de régression linéaire multiple reliant la variable dépendante (expliquée) à plusieurs variables explicatives (dans notre cas les paramètres environnementaux : magnitude, distance épacentrale, conditions de sols). La méthode des moindres

carrées est utilisée pour l'estimation des valeurs des paramètres du modèle associées aux variables explicatives. Des tests d'hypothèses sont menés sur chacun et sur l'ensemble des paramètres afin de s'assurer de leurs importances individuelles et globales vis-à-vis de la formulation donnée pour la variable expliquée.

Pour chaque variable estimée (PSV, PGA et D_v), une étude paramétrique est effectuée, afin d'apprécier l'influence de chacun des paramètres environnementaux (magnitude, distance et site).

L'analyse s'est faite pour différentes bandes fréquentielles et pour divers types de sols caractérisés par la profondeur et la vitesse des ondes S qui leur sont associées.

Les formes spectrales obtenues par les lois de régression ont été comparées avec celles associées aux différentes classes de la banque de données ainsi qu'avec celles proposées par différents codes parasismiques en vigueur dans différents pays. Les codes parasismiques considérés dans cette étude sont les Règles Parasismiques Algériennes (RPA 99 version 2003), l'Eurocode 8 et l'Uniform Building Code (UBC 1997).

Ce mémoire est divisé en huit chapitres qui forment l'ensemble du travail effectué. Le premier chapitre est consacré à la présentation des généralités sur le mouvement sismique, après un bref rappel des principales caractéristiques des séismes, le processus de propagation d'onde ainsi que la représentation du mouvement de sol y sont décrits. Le chapitre 2 est dédié à réponse structurelle : le concept de spectre de réponse, de pseudo spectre de réponse, de spectre réglementaire et le spectre de calcul y sont présentés. Les méthodologies de la régression linéaire et multilinéaire sont rappelées dans le chapitre 3. Les caractéristiques de la banque de données utilisées dans cette étude sont présentés au chapitre 4. Le chapitre 5 reprend les différentes étapes pour le développement des lois prédictives associées au spectre de réponse, l'estimation des paramètres B_i associés aux variables explicatives correspondantes sont faites en utilisant les méthodes de régressions décrits au chapitre 3. Les résultats de la régression sont discutés en terme d'analyse de variance et en terme de contribution d'ensemble et marginale des variables explicatives. Les lois prédictives pour le spectre de réponse sont alors utilisées afin d'analyser l'influence de chacun des paramètres environnementaux. Les formes spectrales prédictives sont comparées à celles proposées par les codes parasismiques considérés dans cette étude.

Les chapitres 6 et 7 sont consacrés à l'estimation des lois prédictives pour le PGA et la durée de la phase de mouvements sismiques forts D_v respectivement ainsi qu'aux études paramétriques sur l'influence des variables explicatives qui leur sont associées. Les discussions et les commentaires reflétant les résultats obtenus sont présentés.

Les principales conclusions et recommandations relatives à ce travail sont exposées dans le chapitre 8.



CHAPITRE I :

GENERALITES SUR LE MOUVEMENT SISMIQUE.

- INTRODUCTION.
- MOUVEMENT FORT.
- PROPAGATION DES ONDES SISMIQUES.
- REPRÉSENTATION DU MOUVEMENT SISMIQUE
- CARACTÉRISTIQUES ENREGISTRÉS À LA SURFACE DU SOL.
- PARAMÈTRES AFFECTANT LE MOUVEMENT SISMIQUE .
- EFFETS PARTICULIERS DES SÉISME
- ALEA SISMIQUE & RISQUE SISMIQUE DANS UNE RÉGION DONNÉE.



CHAPITRE I : GÉNÉRALITÉS SUR LES SÉISMES.

1.1. Introduction.

Les séismes majeurs sont liés à l'existence de mouvements tectoniques globaux continuellement en action qui soulèvent des chaînes de montagnes et creusent des tranchées océaniques à la surface de la terre. Ces mouvements affectent une quinzaine de "plaques" sphéroïdales et engendrent des contraintes sur les lignes de contact des plaques. Lorsque ces contraintes deviennent trop élevées et supérieures à la «résistance au frottement », une rupture brutale se produit : c'est un tremblement de terre, à la suite duquel un nouvel état provisoire de stabilité est réalisé.

Il résulte de cette description qu'il existe des régions du monde à plus ou moins haut risque sismique, suivant qu'elles sont plus ou moins proches des zones de jonctions des "plaques"

Figure (1.1).

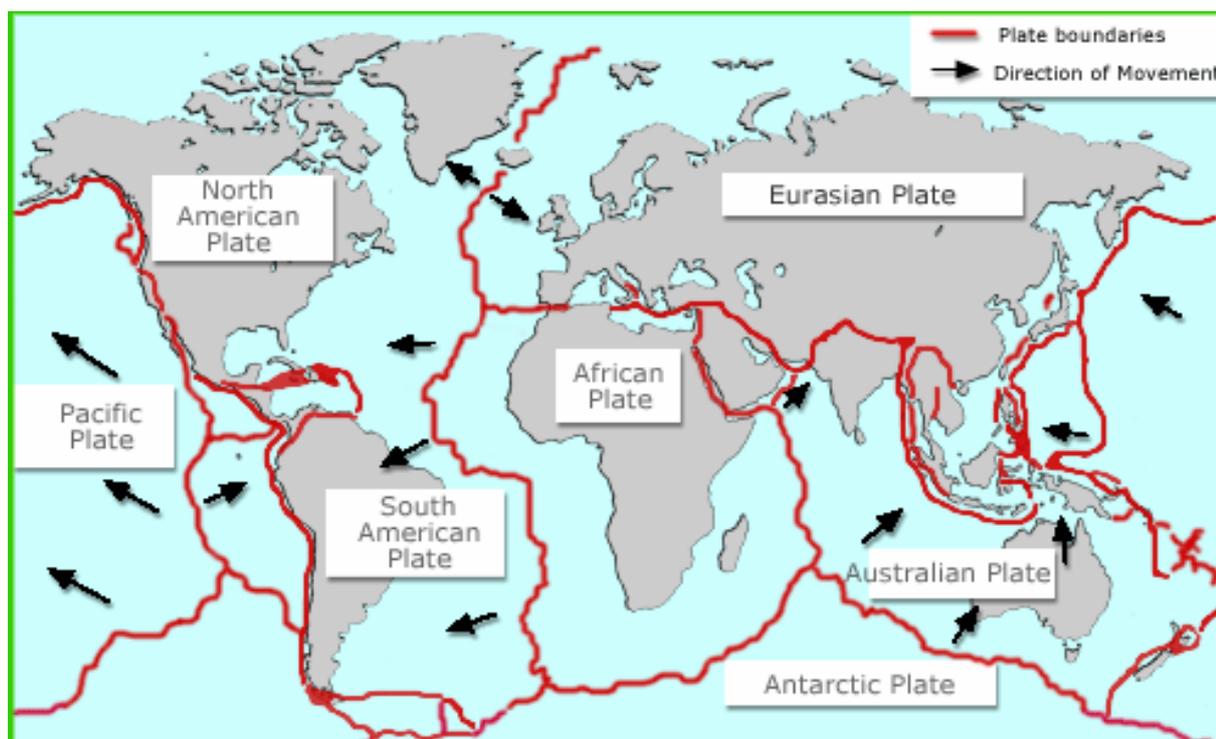


Figure (1.1) : Carte du monde montrant les principales plaques tectonique.



Les zones de jonction inter - plaques ne sont pas les seules où se passent des décrochages de failles. Des phénomènes similaires, d'amplitude moindre, ont lieu dans des failles intérieures aux plaques mentionnées. On parle alors de séismes « intra - plaques ».

Pour les séismes importants, le mouvement relatif des bords de la faille peut être apparent en surface. Il existe différents type de mouvement relatif : coulissage horizontal, coulissage vertical. Ces déplacements relatifs permanents du sol s'observent près de l'épicentre du séisme, mais le choc du tremblement de terre entraîne aussi la création et la propagation d'ondes de compression et de cisaillement dans le sol à des distances très grandes ; en fait ces ondes font le tour de la terre en s'atténuant progressivement. A la surface du globe, l'effet de ces ondes est un mouvement de déplacement vertical et horizontal du sol.

L'importance relative des composantes du mouvement du sol en un point dépend de la position du point considéré par rapport au foyer (ou « hypocentre ») du séisme. A l'aplomb de celui-ci (« épicentre »), la composante verticale est du même ordre que la composante horizontale. A plus grande distance, l'importance relative de la composante verticale diminue et vaut 50 à 70% de l'horizontale. Les composantes de rotation ont généralement des effets négligeables.

Le mouvement vertical est souvent perçu comme moins dommageable que l'horizontal, partiellement en raison de l'importance généralement moindre de cette composante, mais surtout en raison du fait que les constructions sont naturellement conçues pour reprendre l'action (verticale) de la pesanteur avec une sécurité convenable alors qu'elles ne présentent pas les mêmes ressources pour une action horizontale.

1.2. Caractérisation des séismes.

On peut caractériser la "taille" du tremblement de terre par différents paramètres.

La **magnitude M** (échelle de RICHTER) est une évaluation de l'énergie libérée au foyer du tremblement de terre. Comme telle, cette quantité n'est pas utilisable par l'ingénieur des constructions qui s'intéresse à un mouvement en surface et dans une zone géographique donnée. Le foyer peut se trouver à plusieurs kilomètres sous la surface et à plusieurs centaines de kilomètres de la zone donnée.

L'**intensité I** (échelle de MERCALLI) est une caractérisation de l'importance des effets matériels observés en surface et de la perception par la population. Cette description vaut pour une zone géographique donnée, mais est assez qualitative. L'intensité en un point dépend non seulement de la



taille du séisme (magnitude) mais aussi de la distance au foyer, de la géologie locale et de la topographie.

Un paramètre important pour l'évaluation des effets des séismes à un endroit donné est l'**accélération maximale** a_g du sol ou accélération de pointe, PGA (Peak Ground Acceleration) en anglais. Son ordre de grandeur est de 0,4 g à 0,6g en zone très sismique (Japon, Turquie) et de 0 à 0,1g en zone faiblement sismique. L'amplitude de l'accélération maximale du sol permet de se faire une idée de la résultante de force F appliquée à la construction de masse m : $F = m a_g$ si la construction est indéformable et bouge comme le sol. C'est en terme de d'accélération maximale a_g au niveau du bedrock qu'on exprime les cartes de zonation ou zonage sismique quantifiant le niveau sismique à prendre en compte pour l'application des règles de constructions parasismiques.

Déplacement maximal d_g du sol, qui donne une idée de l'ordre de grandeur du déplacement relatif du centre de gravité de la structure par rapport à la base de la structure : quelques cm en zone faiblement sismique, jusqu'à un m en zone très sismique.

La **durée** du tremblement de terre est un paramètre significatif dans les processus de fissurations et dégradations progressives des éléments d'une construction. Elle est liée à la magnitude du séisme.

Cette durée est au maximum de l'ordre de 60 s en zone très sismique, mais n'est que de quelques secondes en zone peu sismique.

1.3. 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 grandes 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 étudié les propriétés des mouvements sismiques proches de la source que pour l'étude de la variation spatiale des ondes sismiques.



1.4. 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 (fig 1.2).

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 (fig. 1.2).

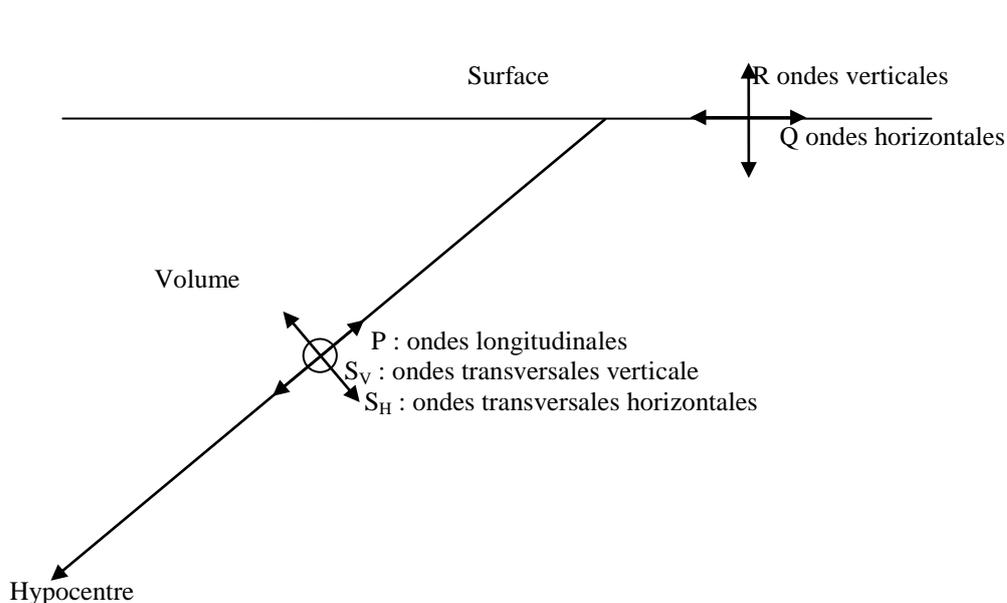


Figure (1.2) : Les ondes sismiques de volume et de surface



1.5. 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.

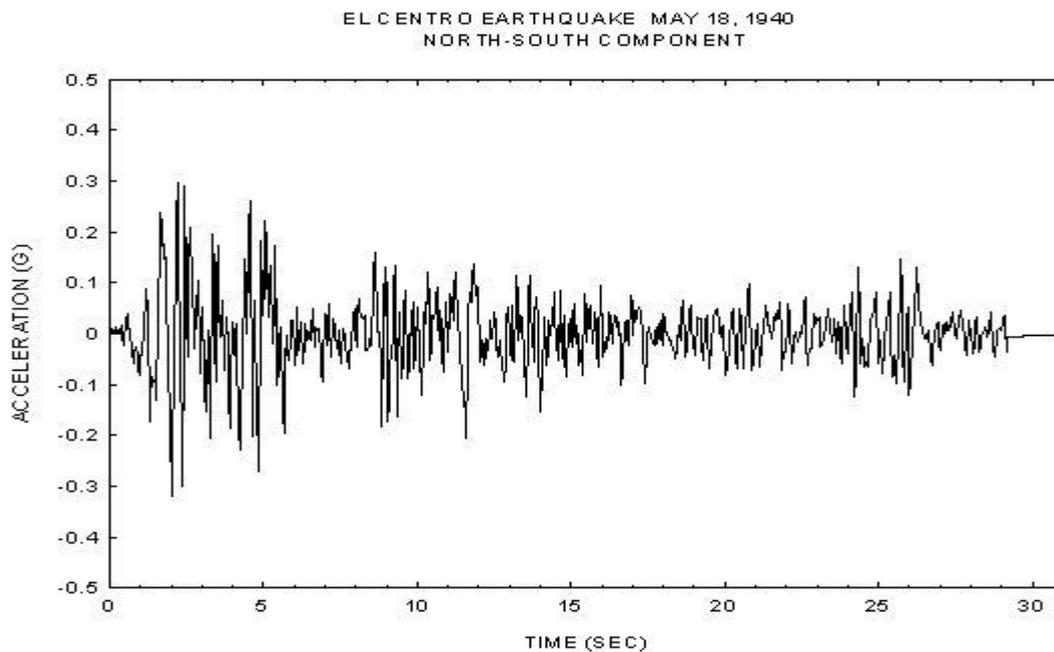


Figure (1.3) : Accélérogramme

Ainsi, sur un enregistrement accélérométrique. Un accélérogramme $x(t)$ est donc une fonction à 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.

Que peut-on tirer comme information à partir d'un accélérogramme ?

Le mouvement du sol enregistré à une station accélérométrique donnée suite à un séisme, est en général influencé (cette influence se traduit par un changement de la forme d'onde) par quatre facteurs : L'effet de la source (taille, mécanisme, directivité, nombre d'aspérités...), l'effet de la



propagation d'onde (impact de l'hétérogénéité de la croûte terrestre sur les ondes générées par la dislocation, et qui se propagent dans le volume entre la source et la station), l'effet des conditions de site local (les premières centaines de mètres du sol et/ou la topographie) et enfin l'effet de l'instrument lui-même.

De plus en plus, des données de mouvements forts seront enregistrées suite à des séismes se produisant sur des failles bien connues (par exemple, Chi-chi, Parkfield...) et donc bien instrumentées. Ces données, parfois combinées à d'autres types de données comme les données géodésiques ou télésismiques par exemple, nous fournissent d'importantes informations sur le processus de rupture des séismes. En effet, les données en champ proche (mouvements forts) sont peu contaminées par l'effet de propagation et sont plutôt dominées par l'effet de la source, à l'aide de ce genre de données, le glissement sur le plan de faille est mieux contraint pour les basses fréquences ($f < 1.5$ Hz). L'accélération du sol est très complexe et difficile à déchiffrer, alors on utilise soit la vitesse qui caractérise les détails de premier ordre de la rupture soit le déplacement qui est beaucoup plus simple à expliquer. En sismologie de l'ingénieur, les informations directes suivantes sont systématiquement lues sur un accélérogramme: Le pic d'accélération, Durée des secousses et Le contenu fréquentiel.

1.6. Caractéristiques enregistrés 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.

1.7. Paramètres affectant le mouvement sismique.

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.

1.7.1. 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 mince figure (1.4) 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.

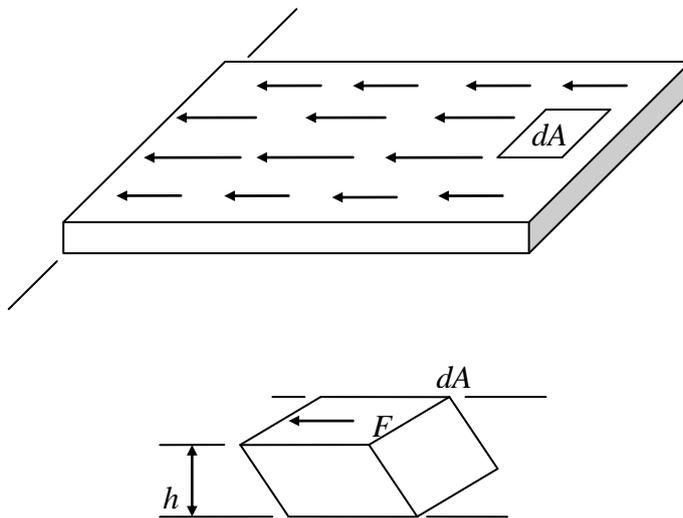


Figure (1.4) : Schématisation d'un plan de faille (Brune 1976)

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. 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. Du point de vue de 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 :

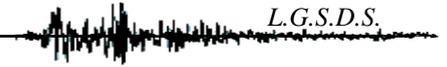
$$\log M_0 = 19,9 + M_L \quad (1.5)$$

Intensité.

L'échelle d'intensité attribuée à un séisme, caractérise de son importance en un lieu donné, d'après les manifestations ressenties 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.



1.7.2. 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.



1.7.3. Facteurs liés aux conditions locales.

Filtrage, amplification :

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 maximal 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.

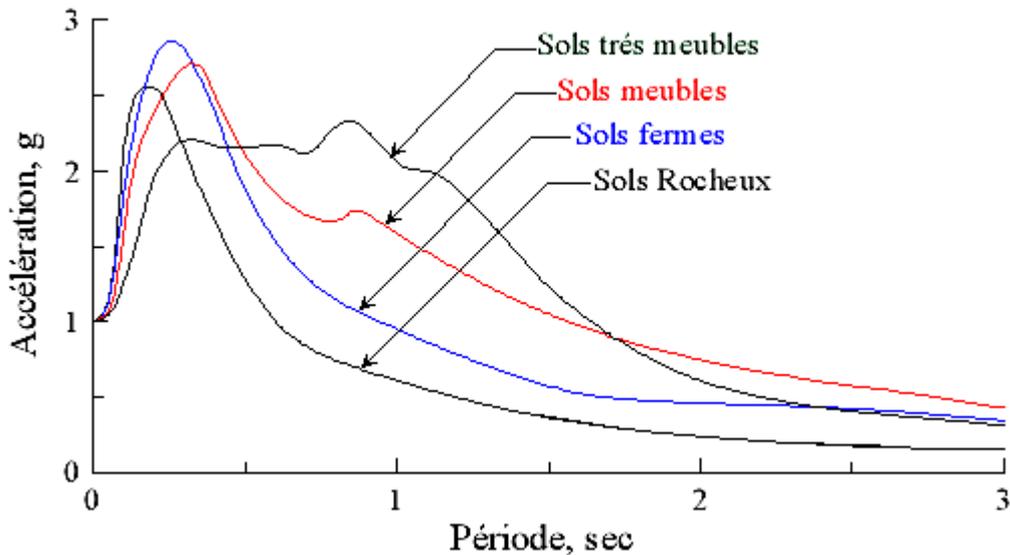
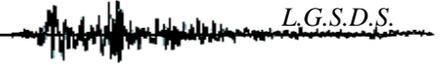


Figure (1.5) : Spectres moyens d'accélération pour différentes conditions de site selon Seed et al. (1976)

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. Il semblerait qu'en réalité, la forme du spectre de réponse dépende également de ces paramètres.

Durée des secousses: La durée de la secousse dépend du temps nécessaire à la rupture pour casser toutes les parties instables sur le plan de faille et de la propagation. Plus la surface de la faille est grande, plus la magnitude du séisme est grande et plus le temps de rupture est important. Le résultat est que la durée du mouvement fort augmente avec la magnitude du séisme. Il faut noter que si la rupture est bilatérale la durée est moins importante que si la rupture était unilatérale. En sismologie de l'ingénieur on utilise souvent la durée encadrée (« bracketed duration »), i.e., le temps entre le premier et le dernier dépassement d'un certain seuil d'accélération, souvent égal à 0.05g (Bolt, 1969).



CHAPITRE II :

SPECTRE ET PSEUDO SPECTRE DE REPONSE

- INTRODUCTION
- RÉPONSE D'UN OSCILLATEUR 1DDL
- SPECTRE DE RÉPONSE ÉLASTIQUE
- SPECTRE ET PSEUDO SPECTRE DE RÉPONSE
- SPECTRE DE RÉPONSE DU CODE PARASISMIQUE
- SPECTRE DE CALCUL

CHAPITRE II : SPECTRE ET PSEUDO SPECTRE DE RÉPONSE.

2.1. Introduction :

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 $u(t)$.

Le concept de spectre de réponse a été utilisé pour la première fois par M.A. Biot pour caractériser l'effet des tremblements de terre sur les structures. G.W. Housner a popularisé son utilisation en génie parasismique. L'importance de ce concept vient du fait que, lors d'un dimensionnement, on ne s'intéresse surtout qu'à la valeur maximum de la réponse d'une structure à un tremblement de terre.

On a pour un paramètre de réponse quelconque τ :

$$\tau_{\max} = \max (|\tau(t)|) \dots \dots \dots (2.1)$$

L'indice max désigne la valeur maximum de la réponse dans le temps. La courbe de fréquences la valeur maximum d'un paramètre de réponse quelconque (déplacement, vitesse ou accélération) en fonction de la période naturelle (ou fréquence) d'un système élémentaire est appelée un spectre de réponse.

Définissons les spectres de réponse suivants :

spectre de déplacement relatif $SD = U_{\max} = \max(|U(t, \omega_D, \xi)|) \dots \dots \dots (2.2)$

spectre de vitesse relative $SV = \dot{U}_{\max} = \max(|\dot{U}(t, \omega_D, \xi)|) \dots \dots \dots (2.3)$

spectre d'accélération absolue $SA = \ddot{U}_{\max} = \max(|\ddot{U}(t, \omega_D, \xi)|) \dots \dots \dots (2.4)$

2.2. Réponse d'un oscillateur 1DDL :

On considère un oscillateur simple lié au sol et caractérisé par sa pulsation ω et son coefficient d'amortissement ξ . Il est soumis aux effets d'un séisme, c'est-à-dire que son point d'appui se déplace avec une accélération $\ddot{u}(t)$ donnée par un accélogramme, comme par exemple celui du séisme d'El Centro du 18 mai 1940 (fig. 2.1),

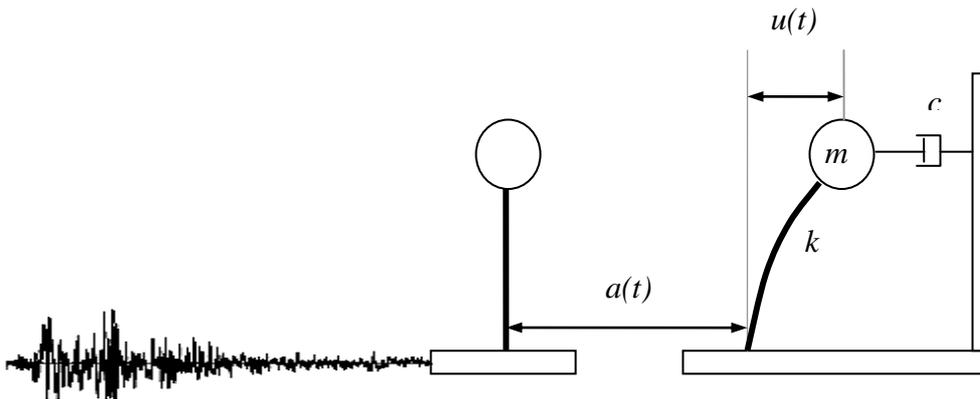


Figure (2.1) : Oscillateur simple soumis à une excitation sismique

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

$$m\ddot{u} + c\dot{u} + ku = ma(t) \quad (2.5)$$

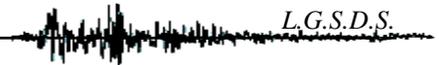
En posant $\omega_n = \frac{k}{m}$, $\xi = \frac{c}{2m}$ et $\omega_D = \omega_n \sqrt{1 - \xi^2}$, la réponse de l'oscillateur a pour expression :

$$u(t) = -\frac{1}{\omega_D} \int_0^t a(\tau) e^{-\xi\omega_n(t-\tau)} \sin[\omega_D(t-\tau)] d\tau \quad (2.6)$$

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 (2.5) 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 :



$$u(t) = \int_0^{\infty} h(\tau) a(t - \tau) d\tau \quad (2.7)$$

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 (2.1), la transformée de Fourier de la réponse impulsionnelle s'écrit :

$$H(i\omega) = \frac{-1}{(\omega_n^2 - \omega^2) + 2i\xi\omega\omega_n} = \frac{U(i\omega)}{A(i\omega)} \quad (2.8)$$

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 :

$$PSA = \omega_n^2 * PSD$$

Où la pseudo-accélération PSA constitue une bonne approximation du maximum du module de l'accélération absolue de l'oscillateur pour les amortissements faibles utilisés en pratique.

2.3. Spectre de réponse élastique.

Principe d'établissement.

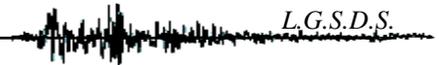
Dans une zone géographique donnée où un accélérogramme $a_g(t)$ représentatif des mouvements sismiques en base de la fondation de la structure étudiée est connu, on peut penser à simplifier la vie des ingénieurs auteurs de projet (spécialisés à ce stade en oscillateurs simples amortis) en résolvant une fois pour toutes l'ensemble des cas des divers oscillateurs simples possibles.

A cette fin, on effectue n fois l'opération suivante.

On considère un oscillateur composé d'une masse M , située à une hauteur H d'une poutre console d'inertie I et de module d'élasticité E .

La constante de rappel k de la console « ressort » est égale à : $k = 3EI / H^3$

La période propre T_1 de l'oscillateur ainsi constitué vaut:



$$T_1 = 2\pi / \omega = 2\pi (M / k)^{1/2} \Rightarrow T_1 = 2\pi \sqrt{\frac{MH^3}{3EI}}$$

En résolvant l'équation (1.6), on établit la valeur du déplacement relatif $D = u(t)$ de la masse M par rapport l'encastrement A de sa console support et de la pseudo accélération $A(t) = \omega^2 u(t)$ à chaque instant. L'amplitude du déplacement d de la masse par rapport à la base est influencée par l'amortissement du système ; on attribue à ce dernier pour la plupart des applications une valeur standard égale à 5 % de l'amortissement critique.

On retient la valeur maximale de $A(t)$ sur la durée du tremblement de terre, et on l'on associe à la période T_1 , soit $A(T_1) = \max A(t)$. On reporte $A(T_1)$ dans un diagramme $(A(T_1), T)$.

En faisant varier les paramètres qui définissent la console de référence (autres masses M , autres raideurs k , qui correspondent à d'autres périodes fondamentales $T = T_1, T_2 \dots$), on établit un ensemble de paires de valeurs $[T_i, A(T_i)]$.

L'ensemble des points $[T_i, A(T_i)]$ constitue ce qu'on appelle le **spectre de réponse élastique SA(T) en accélération** (en fait en pseudo accélération) correspondant à un accélérogramme de la zone étudiée.

L'auteur de projet qui dispose de ce spectre peut obtenir directement, pour un oscillateur simple donné dont il évalue la période propre T , la force maximale **Fmax** « équivalente » au tremblement de terre :

$$F_{max} = M * A(T_1).$$

L'équivalence mentionnée correspond au fait que le déplacement calculé sous l'application de ce F est égal au déplacement relatif maximum $D = U_{max}$ entre la masse M et la base de la console.

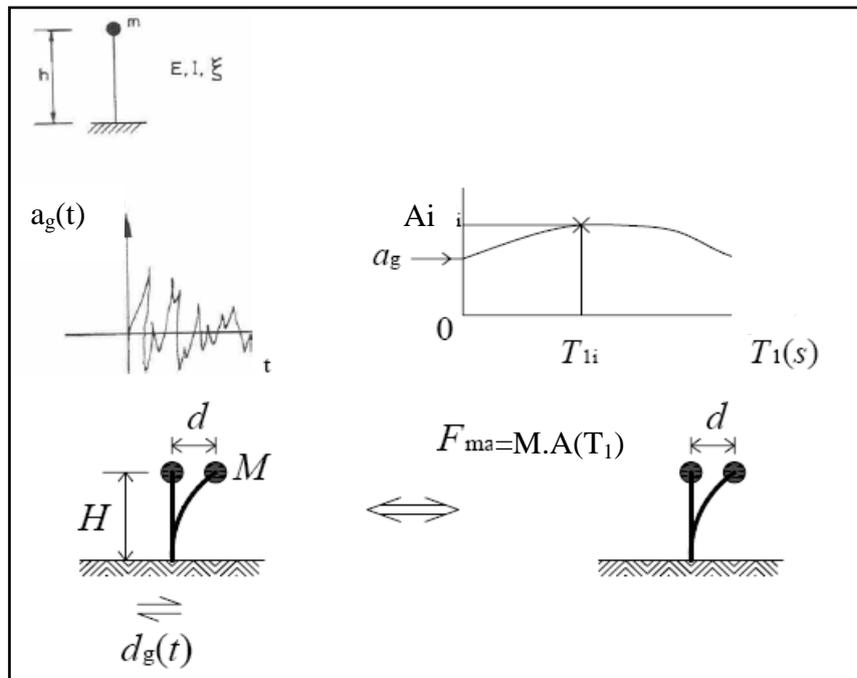
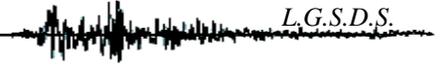


Figure (2.5) : Construction d'un spectre de réponse élastique en accélération correspondant à un accélérogramme donné et définition d'une Force statique Fmax « équivalente » au séisme.



2.4. Spectre et Pseudo spectre de réponse :

Pour un tremblement de terre donné, les spectres de réponse sont donc des fonctions de la fréquence et de l'amortissement. On les représente sous forme de graphes pour des taux d'amortissement ξ donnés sur une large bande de fréquences. Pour une fréquence naturelle et un taux d'amortissement donnés, la valeur du déplacement relatif maximum est obtenue de l'équation (1.6) pour nous donner

$$SD(\omega, \xi) = \max_t \left| -\frac{1}{\omega_D} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega_D(t-\tau) d\tau \right| = \max_t |D(t)| \quad (2.9)$$

$$D(t) = -\frac{1}{\omega_D} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega_D(t-\tau) d\tau \quad (2.10)$$

De même, pour une fréquence et un taux d'amortissement donnés, la vitesse relative maximum est obtenue en calculant le maximum absolu de la vitesse donnée par :

$$SV(\omega_D, \xi) = \max_t \left| -\int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \cos \omega_D(t-\tau) d\tau + \frac{\xi\omega}{\omega_D} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega_D(t-\tau) d\tau \right| \quad (2.11)$$

et l'accélération absolue maximum est obtenue de l'équation

$$SA(\omega_D, \xi) = \max_t \left| \frac{\omega(1-2\xi^2)}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega_D(t-\tau) d\tau + 2\xi\omega \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \cos \omega_D(t-\tau) d\tau \right| \quad (2.12)$$

Une simplification considérable peut être obtenue en notant que $\omega \approx \omega_D$, pour les faibles valeurs d'amortissement rencontrées en pratique, c'est-à-dire pour $\xi < 0.20$, et en supposant que les termes en ξ et ξ^2 sont petits. On obtient :

$$SV(\omega, \xi) = \max_t \left| \int_0^t -\ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \cos \omega(t-\tau) d\tau \right| \quad (2.13)$$

Cette expression peut encore être simplifiée en remplaçant le terme en cosinus de l'équation (2.13) par un terme en sinus -qui a la même valeur maximum bien qu'à un temps différent - afin d'obtenir la relation finale suivante :

$$PSV(\omega, \xi) = \max_t \left| \int_0^t -\ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega(t-\tau) d\tau \right| = \max_t |V(t)| \quad (2.14)$$

Où **PSV** est utilisé pour désigner le pseudo-spectre de vitesse et le distinguer du spectre exact de vitesse dénommé SV et $V(t)$ est l'historique de la pseudo-vitesse relative

$$V(t) = - \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin \omega(t-\tau) d\tau \quad (2.15)$$

On obtient de manière similaire le pseudo-spectre d'accélération

$$PSA(\omega, \xi) = \max_t \left| \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \omega \sin \omega(t-\tau) d\tau \right| = \max_t |A(t)| \quad (2.16)$$

Avec :

$$A(t) = \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \omega \sin \omega(t-\tau) d\tau = -\omega V(t) = -\omega^2 D(t) \quad (2.17)$$

En comparant les équations (2.9), (2.16) et (2.14) et en posant $PSP = SD$, on constate que :

$$PSV = \omega \cdot PSD \quad (2.18)$$

$$PSA = \omega^2 \cdot PSd = \omega \cdot PSV \quad (2.19)$$

On détermine les spectres de réponse en évaluant numériquement les intégrales des équations (2.9) et (2.12) pour un tremblement de terre et une valeur d'amortissement donnée et un grand nombre de fréquences (périodes) naturelles. Le processus d'évaluation de spectre de réponse en déplacement est illustré à la figure (2.6).

Tenant compte des relations (2.18) et (2.19), on peut calculer les pseudo-spectres de réponse par intégration numérique de l'intégrale de Duhamel dans (2.9). En général, cependant, on utilise une méthode directe d'intégration numérique des équations du mouvement pour calculer le déplacement maximum des différents systèmes élémentaires pour la détermination des spectres de réponse. On recommande d'utiliser la méthode d'intégration des fonctions linéaires par morceau. Dans ce cas, l'intégration peut être considérée comme exacte puisque l'historique de l'accélération est défini à intervalle de temps régulier et qu'une interpolation linéaire est utilisée entre les points.

Les spectres de réponse en déplacement, en vitesse et en accélération pour L'accélérogramme d'El Centro sont présentés à la figure (2.7) pour un taux d'amortissement $\xi = 0.02$. On peut constater que la différence entre SA et PSA est négligeable alors que cette différence est notable entre SV et PSV pour les périodes de vibration supérieures à 3 S.

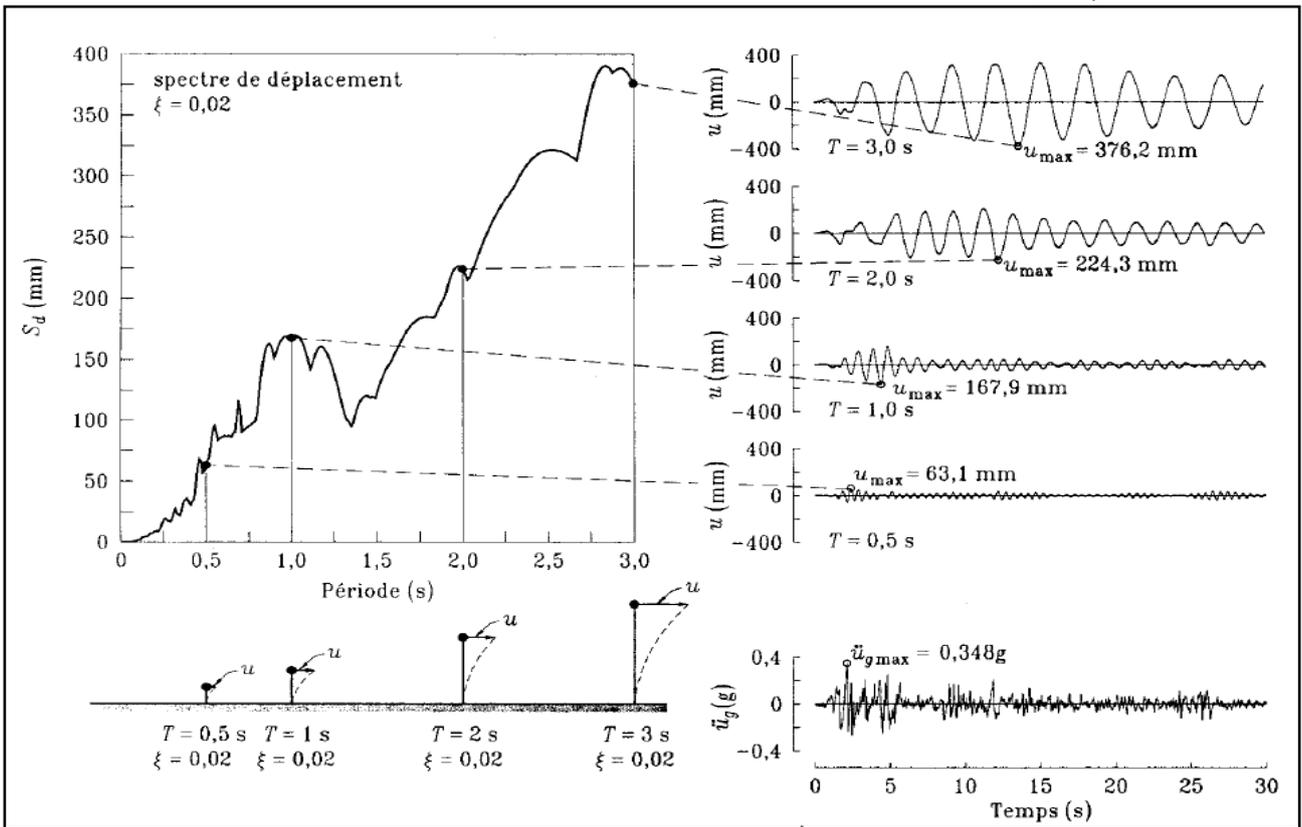


Figure (2.6) : illustration du calcul du spectre de déplacement pour l'accélérogramme d'El Centro

Pour un oscillateur élémentaire vibrant sinusoïdalement avec un déplacement maximum SD , la vitesse maximum et l'accélération maximum sont égales à PSV et PSA, respectivement.

Pour un tremblement de terre, on peut montrer que $PSV \approx SV$ et $PSA \approx SA$ sur la gamme usuelle des périodes et des amortissements. L'accélération étant, comme on l'a vu, plus précise que la vitesse. Quand $T \rightarrow +\infty$, SD tend vers le déplacement maximum du sol, SV tend vers la vitesse maximum et PSA tend vers zéro. Or, $PSV = (2\pi.PSD/T)$ tend vers zéro quand $T \rightarrow +\infty$. Donc, pour les longues périodes, le pseudo-spectre de vitesse PSV peut être considérablement différent du spectre de vitesse SV. Il faut noter, de plus, que pour un amortissement nul, $PSA = SA$ et $PSV \neq SV$ puisque l'expression de PSV contient un terme en sinus alors que l'expression de SV contient un terme en cosinus. Remarquons que l'emploi du terme pseudo - qui veut dire faux - en rapport avec la représentation approximative des spectres peut induire en erreur puisqu'il s'agit ici d'approximation et non de faux concepts.

L'avantage d'utiliser les pseudo-spectres de réponse vient des relations simples données par les équations (2.12) et (2.13). Ces relations entre PSD, PSV et PSA permettent de tracer simultanément sur un même graphique les trois spectres de réponse.

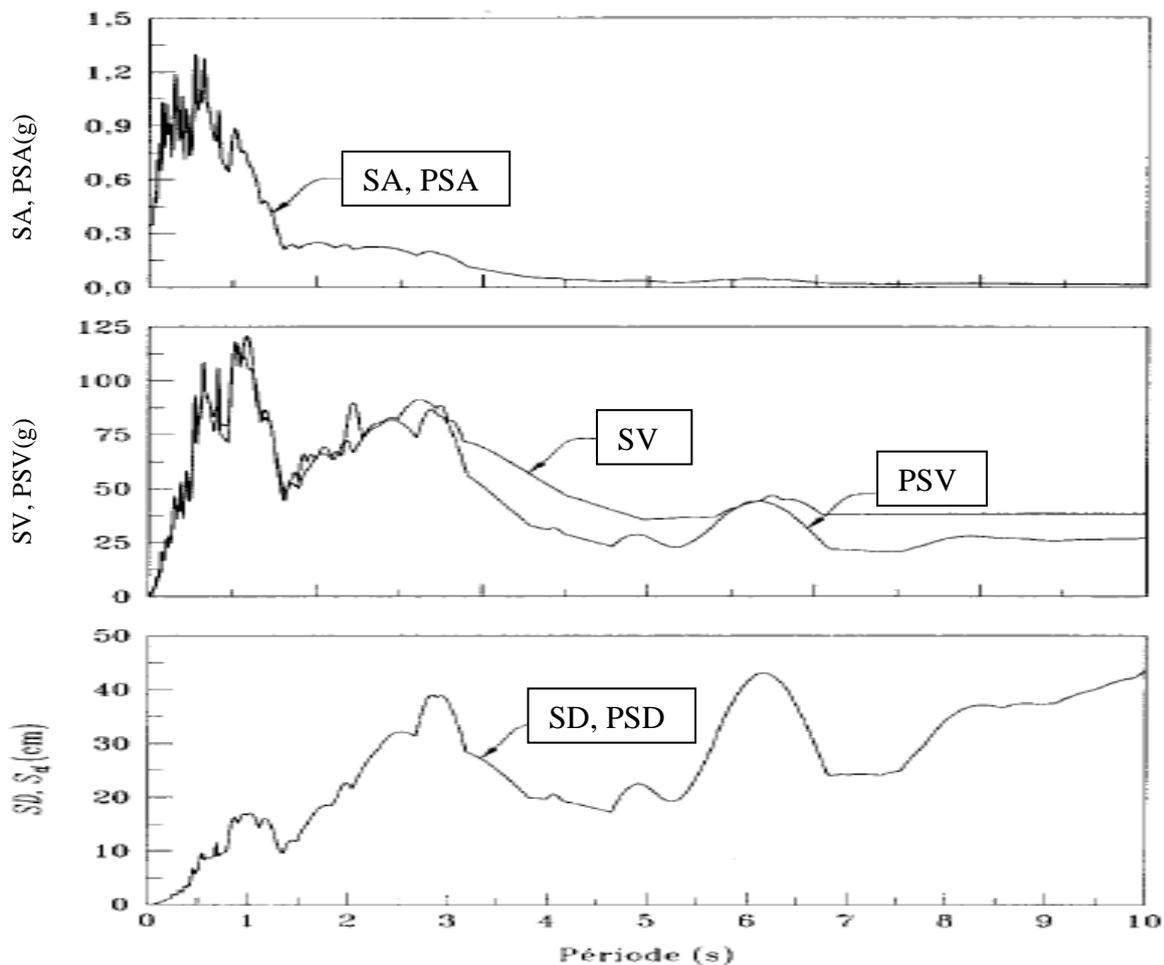


Figure (2.7) : Les spectres de réponse en déplacement, en vitesse et en accélération pour l'accélérogramme d'El Centro sont présentés pour un taux d'amortissement $\xi = 0.02$

Les spectres de PSD, PSV et PSA pour les accélérogrammes d'El Centro du tremblement de terre de l'Imperial Valley de 1940 sont illustrés à la figure (2.8) pour des taux d'amortissement $\xi = 0, 2, 5, 10,$ et 20% de l'amortissement critique. Il faut noter que la représentation sur un seul diagramme à quatre axes logarithmiques n'est possible que pour des quantités qui sont reliées par une puissance de la fréquence comme dans (2.19).

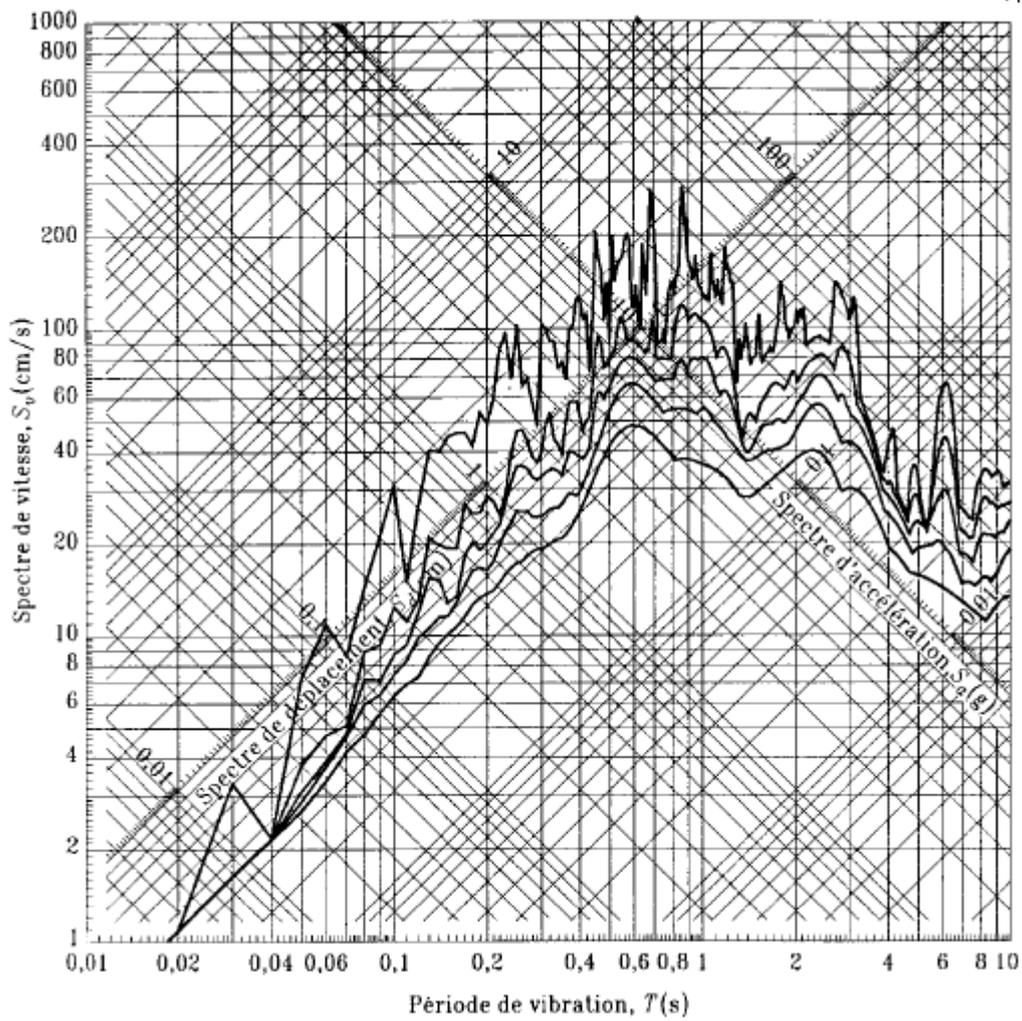


Figure (2.8) : Diagramme tripartite représentant les spectres de réponses combinés PSD-PSV-PSA de l'accélerogramme d'EL Centro pour $\xi = 0, 2, 5, 10$ et 20%

2.5. Spectre de réponse du code parasismique.

Des incertitudes affectent la connaissance de l'accélérogramme qui sera mesuré à un endroit donné lors du prochain tremblement de terre. Le "spectre de réponse en accélération A " établi comme indiqué ci-dessus correspond à un seul accélérogramme et ne peut pas bien représenter la séismicité du site.

On rencontre les incertitudes quant aux prochains séismes en établissant plusieurs spectres de réponse $A(T_1)$ qui correspondent à plusieurs accélérogrammes et en retenant comme spectre pour le code parasismique une "moyenne" de tous ces spectres – Figure (2.9).

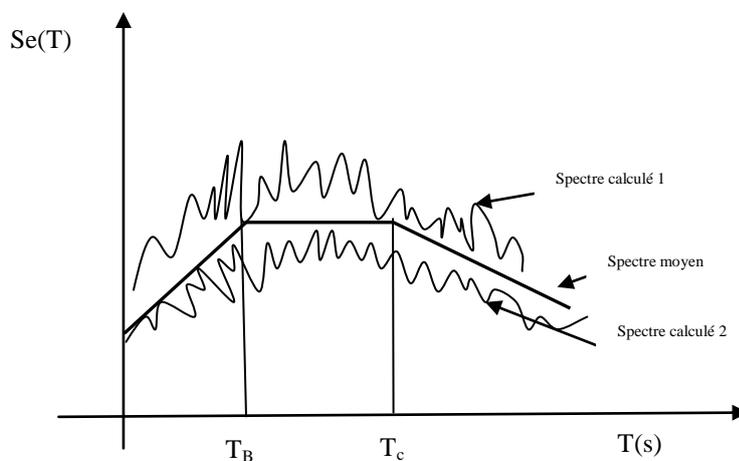


Figure (2.9) : Définition d'un spectre « moyen » $S_e(T)$ de code parasismique.

Le spectre "moyen" résulte d'une approche statistique, associée à un "jugement d'ingénieur" et au souci d'un confort d'utilisation du spectre normatif, de sorte que la forme du spectre de réponse de projet $S_e(T)$ est plus schématique que celle de chaque spectre de réponse $A(T_1)$.

On présente à la Figure (2.10) le spectre de réponse élastique $S_e(T)$ de l'Eurocode 8, où on distingue des périodes "de coin" T_B , T_C et T_D . Les codes parasismiques définissent généralement une seule schématisation de base pour le spectre de réponse élastique en accélération de référence.

L'évaluation de la déformée et des sollicitations d'une console de masse M et de raideur EI/H est effectuée comme expliqué précédemment et la force horizontale F maximale vaut:

$$F_{\max} = M.S_e(T)$$

Si on peut établir que l'amortissement est différent de la valeur standard de 5 %, le spectre de réponse élastique peut être corrigé par un coefficient η .

Le spectre des codes parasismiques est généralement établi pour une valeur standard d'amortissement égale à 5% de l'amortissement critique. Si on peut démontrer qu'une structure étudiée est caractérisée par une autre valeur d'amortissement, on modifie le spectre de base par un paramètre η .

Pour une structure très raide ($T \approx 0$ s), le pseudo accélération $S_e(T)$ est égal à l'accélération maximale $S_{ag} = S''_{dg}(t)$ du sol ($S=1$), car une structure infiniment raide suit exactement les mouvement du sol, sans avoir de déformée propre.

La réponse dynamique des structures modérément flexibles (T compris entre T_B et T_C), entraîne à une amplification des accélérations par rapport à l'accélération du sol; le pseudo accélération $S_a(T)$ est de l'ordre de 2,5 fois l'accélération maximale S_{ag} ; le contenu fréquentiel des accélérogrammes, qui est plus important dans cette gamme de périodes, favorise un aspect « résonance » dans la réponse de la structure.

- les structures très flexibles sont les moins sollicitées en terme de force, car elles fléchissent de telle sorte que $u(t) \approx d_g(t)$ et « ne voient pas passer » le séisme ; cette grande flexibilité est toutefois synonyme d'un effet P- Δ important.

On conclut que les effets d'un tremblement de terre sur une structure dépendent de sa rigidité : une structure plus rigide (période propre T faible) subit des forces plus élevées qu'une structure flexible.

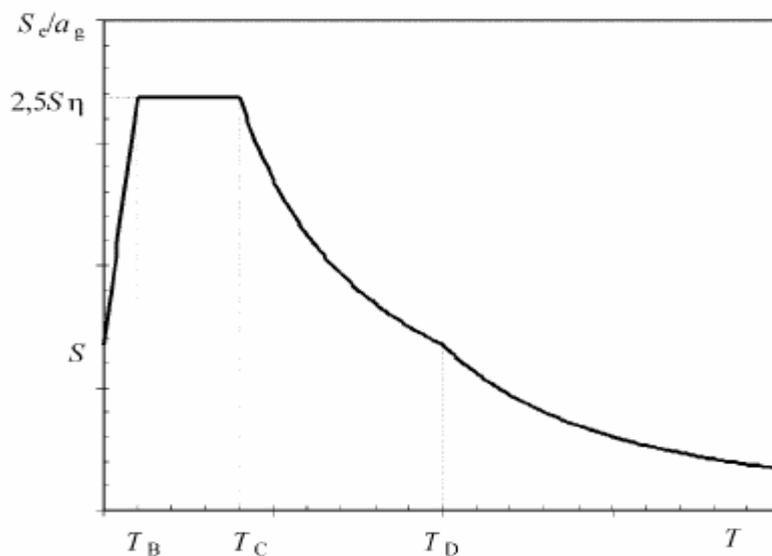


Figure (2.10) : Le spectre de réponse élastique $S_e(T)$ de l'Eurocode 8

2.6. Spectre de calcul :

Lorsqu'il s'agit de déterminer le spectre de réponse à prendre en compte pour le calcul des constructions en un site donné, il est bien entendu exclu d'utiliser un seul accélérogramme, même si par chance il a été enregistré au voisinage du site.

En effet, l'accélérogramme du séisme contre lequel on veut se prémunir n'est a priori pas prévisible. Il convient donc de déterminer un spectre de calcul qui sera l'enveloppe d'un ensemble de spectres correspondant à des accélérogrammes enregistré dans des sites comparables du point de vue de la nature du sol.

Les accélérogrammes utilisés résultent de séismes d'importance différente et les spectres de réponse que l'on en déduit ne sont pas directement comparables. Il faut donc au préalable appliquer à chaque spectre une affinité pour qu'ils aient tous la même valeur d'intensité spectrale.

Les spectres ainsi obtenus sont alors appelés « spectres normalisé »

Housner a défini l'intensité spectrale comme étant l'aire du spectre de vitesse comprise entre l'axe T et les droites d'abscisse 0.10 et 2.5 secondes cette plage de périodes couvrant la plupart des cas rencontrés.

L'intensité spectrale prend en compte la réponse de la majorité des constructions et constitue donc une mesure de l'importance de la secousse sismique.

On peut aussi normaliser les spectres en prenant en compte l'accélération maximum du sol au lieu de l'intensité spectrale.

Le spectre de calcul est obtenu en faisant l'enveloppe des divers spectres normalisé; il doit ensuite être calé en lui faisant subir une affinité pour tenir compte :

- de l'intensité probable du séisme;
- de l'importance pour la collectivité du bâtiment étudié.

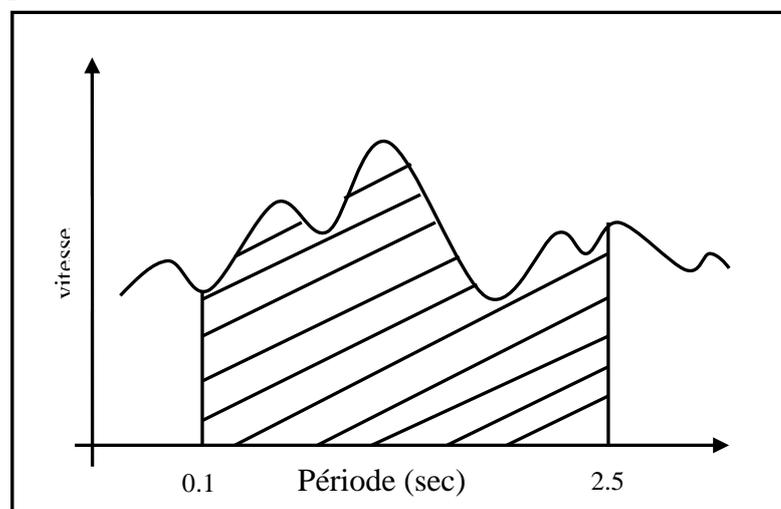


Figure (2.11) : l'intensité spectrale selon Housner



2.7. Spectre de calcul pour l'analyse élastique.

Du spectre de réponse élastique au spectre de calcul.

D'autres facteurs que ceux considérés dans l'établissement du spectre de réponse élastique $S_e(T)$ interviennent dans la réponse des structures aux tremblements de terre.

Les spectres de calcul pour l'analyse élastique $S_d(T)$ prennent en compte ces facteurs additionnels, qu'on définit aux paragraphes suivants :

a. Importance de la construction.

La définition de l'accélération maximale "de calcul" a_g résulte d'un processus statistique et correspond à l'acceptation d'un certain niveau de risque. Il en découle que l'accélération maximale de calcul a_g devrait être plus grande pour les structures considérées comme plus précieuses ou plus importantes à divers points de vue. Dans l'Eurocode 8, on définit une accélération de référence a_{gR} correspondant à un niveau standard de risque accepté ; a_{gR} est compris entre 0,05 g (0,5 m/s²) dans les zones très faiblement sismiques et 0,4 g (4 m/s²) dans les zones très sismiques. L'accélération maximale de calcul a_g est trouvée en multipliant a_{gR} par γ_1 , "coefficient d'importance" de la structure considérée :

$$a_g = \gamma_1 \cdot a_{gR} .$$

γ_1 est égal à 1 pour les bâtiments courant et vaut jusqu'à 1,4 pour les structures dont l'intégrité est vitale en cas de séisme.

b. Séisme proche, séisme lointain.

Une accélération de pointe a_{gR} donnée à un endroit donné peut être engendrée par différents types de séisme : un fort séisme dont l'épicentre est éloigné ou un séisme plus faible dont l'épicentre est proche. Le séisme réel affectant une zone est fonction de la géologie, proche et lointaine. Mais les spectres de réponse correspondant aux deux types de séisme mentionnés sont différents, parce que des ondes propagées de loin ou de près produisent des effets différents.

Dans certains codes, cette possibilité est considérée et des formes de spectres de types 1 et 2 sont définies.

Le type 1 correspond à des séismes lointains de magnitude suffisante ($M_S \geq 5,5$) pour engendrer au site de construction des accélérations significatives dont la contribution est prépondérante dans le risque sismique.

Le type 2 est à considérer si des tremblements de terre de magnitude $M_S < 5,5$ constituent le facteur prépondérant de risque.

Dans certaines régions, le spectre de calcul résulte d'une combinaison des spectres des types 1 et 2.

c. Sols et sites.

Les couches de sol présentes entre le rocher sous-jacent et la d'un bâtiment modifient la forme et les amplitudes du spectre de réponse élastique ou "alea", établies au niveau du rocher.

Un paramètre de sol S prend en compte cette influence, de sorte que l'accélération maximale à la fondation est égale à $S a_g$.

Les sites sont classifiés en types A, B, C, D et E selon des profils stratigraphiques et des valeurs de paramètres caractérisant les sols. L'influence sur le mouvement en base de la structure est significative, puisque S est compris entre 1 (sur le rocher) et 1,8 (sol très meuble). De plus, les valeurs des périodes "de coin" T_B et T_C , assez différentes selon les sites et sols et visibles à la Figure (2.12), influencent significativement le spectre. on constate que:

- plus les couches sont meubles, plus l'amplification est élevée à période égale
- l'amplification relative du site D par rapport au site A atteint pratiquement 3 pour des oscillateurs de période égale à 1 s
- la période T_C de « coin » du spectre se déplace vers la droite quand on passe de sol rocheux à sol meuble. Il apparaît clairement qu'ignorer les conditions de sol et site peut conduire à de grandes sous-estimations de l'action sismique.

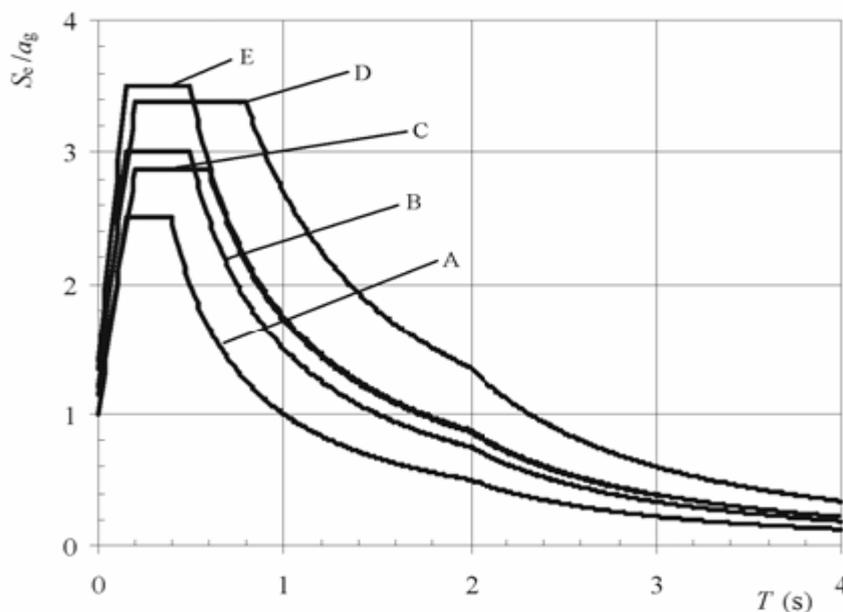


Figure (2.12) : Spectres de calcul $S_e(T)$ de l'Eurocode 8 pour les tremblements de terre



d. Ductilité de la structure.

On dit "ductile" une structure qui peut subir sans perte de résistance des déformations plastiques alternées. la ductilité peut avoir une influence positive sur l'économie d'une structure, car :

- la structure ductile est capable de subir avec succès le même déplacement qu'une structure qui répondrait de façon purement élastique, mais elle atteint ce résultat avec des éléments structuraux de section moindre ;
- les sollicitations à la fondation sont réduites.

Cette capacité à se déformer plastiquement sans perte de résistance est traduite par l'attribution d'un "coefficient de comportement" q dont la valeur dépend du type de structure résistante.

Le coefficient q intervient comme réducteur du spectre élastique $S_e(T)$ lors de la définition du spectre de calcul $S_d(T)$. La réduction est comprise entre 1,5 pour les structures peu dissipatives et 6 pour les structures très dissipatives. Le facteur q permet de tenir compte de la capacité de déformation plastique d'une structure tout en effectuant une analyse purement élastique sous un spectre $S_d(T)$.

CHAPITRE III :

NOTIONS SUR LA RÉGRESSION

- ESTIMATION DES PARAMÈTRES INCONNUS D'UNE LOI DE RÉPARTITION
- INTERVALLE DE CONFIANCE. PROBABILITÉ DE CONFIANCE
- TESTS D'HYPOTHÈSES
- RÉGRESSION LINÉAIRE SIMPLE
- RÉGRESSION LINÉAIRE MULTIPLE

CHAPITRE III : NOTIONS SUR LA RÉGRESSION.

Dans le présent chapitre on va essayer de cerner le plus grand nombre de notions et concepts nécessaires pour l'étude d'une analyse de régression,[4]

3.1. Estimation des paramètres inconnus d'une loi de répartition.

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 donnée statistique bien plus réduit : deux ou trois dizaine d'observation, souvent même moins. Cet état de choses qui est par le crée élevé et par la complexité des expériences. Des données aussi limiter sans nettement insuffisante pour retrouver la loi de répartition inconnue du variable aléatoire ; néanmoins ces données peuvent être traitées est utiliser pour obtenir certain 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 on 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 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 qu'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 lois 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 (3.1)

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 :

$$\tilde{a} = \tilde{a}(X_1, X_2, \dots, X_n) \quad (3.2)$$

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 :

$$E[\tilde{a}] = a \quad (3.3)$$

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

$$Var[\tilde{a}] = \min \quad (3.4)$$

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.

3.1.1. 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] :

$$\left. \begin{aligned} \tilde{m} &= \frac{\sum_{i=1}^n X_i}{n} \\ \tilde{D} &= \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n-1} \end{aligned} \right\} \quad (3.5)$$

3.2. Intervalle de confiance. Probabilité de confiance

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 importantes 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 $\alpha = 1 - \beta$.

Écrivons comme suit :
$$P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \beta \quad (3.6)$$

L'égalité (3.6) signifie qu'avec une probabilité égale à β la valeur inconnue du paramètre a se trouvera dans l'intervalle : $I_\beta = (\tilde{a} - \varepsilon; \tilde{a} + \varepsilon)$ (3.7)

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 (3.1).

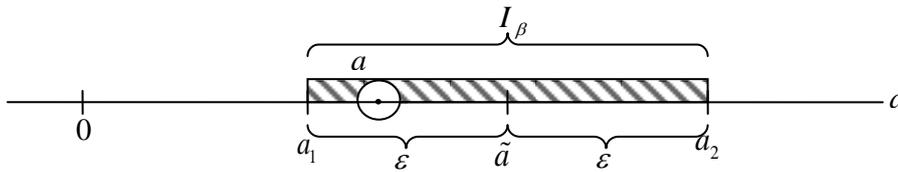


Figure (3.1) Intervalle de confiance

La probabilité β est généralement appelée *probabilité de confiance* et l'intervalle I_β (*intervalle de confiance*). 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 : $P(|\tilde{a} - a| < \varepsilon) = \beta$.

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).

3.3. Test d'hypothèse.

3.3.1. 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.

3.3.2. Concepts importants dans l'élaboration d'un test d'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 est 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 œuvre 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.

- 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 .
- 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.

3.3.3. 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

3.3.4. 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.

3.4. Régression linéaire Simple.

3.4.1. 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.

3.4.2. 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.

3.4.3. 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.

3.4.4. 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.

3.5. Régression linéaire multiple.

3.5.1. 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.

3.5.2. 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.

3.5.3. 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 :

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{pour toute } i.$$

2. Il n'existe aucune corrélation entre les erreurs c.-à-d..

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{pour tout } i \text{ et tout } j, i \neq j.$$

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 :

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

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

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 $n > k + 1$.

3.5.4. 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. Spécifier le modèle de régression multiple que l'on envisage d'analyser.
3. Prélever les observations sur les diverses 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.

3.5.5. 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 carrés.

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 » :

$$[\mathbf{X}]^T \{\mathbf{y}\} - [\mathbf{X}]^T [\mathbf{X}]\{\boldsymbol{\beta}\} = 0 \quad (0.1)$$

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 :

$$\{\mathbf{b}\} = ([\mathbf{X}]^T [\mathbf{X}])^{-1} [\mathbf{X}]^T \{\mathbf{y}\} \quad (0.2)$$

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

$$\{\hat{\mathbf{y}}\} = [\mathbf{X}]\{\mathbf{b}\} = [\mathbf{X}]([\mathbf{X}]^T [\mathbf{X}])^{-1} [\mathbf{X}]^T \{\mathbf{y}\} = [\mathbf{H}]\{\mathbf{y}\} \quad (0.3)$$

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 $\text{Vect}(\mathbf{X})$ engendrée par les vecteurs colonnes de $[\mathbf{X}]$.

$$\text{On note} \quad \{\mathbf{e}\} = \{\mathbf{y}\} - \{\hat{\mathbf{y}}\} = \{\mathbf{y}\} - [\mathbf{X}]\{\mathbf{b}\} = ([\mathbf{I}] - [\mathbf{H}])\{\mathbf{y}\} \quad (0.4)$$

le vecteur des résidus ; c'est la projection de $\{\mathbf{y}\}$ sur le sous espaces orthogonale de $\text{Vect}(\mathbf{X})$ 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[®]).

3.5.6. 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 subséquentement utile dans le calcul d'intervalles de confiance ; on pourra en déduire le coefficient de détermination multiple R^2 .

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

b. 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
$SC_{RES} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$n-(k+1)$

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

d. 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.

e. 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.

3.5.7. Teste 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'autres 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 variables 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.

a. 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 :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (\text{aucune contribution significative des } X_i)$$

$$H_1 : \text{au moins un des } \beta_j \text{ est différent de } 0 \quad (\text{au moins une variable, disant } X_j, \text{ apporte une contribution significative})$$

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 distribué

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 .

Remarques.

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.

b. 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.

c. 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 .

d. 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 $Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon_i$ au seuil de signification α . 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 $t > t_{\alpha/2; n-k-1}$.

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

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

e. 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. $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

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

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_1 : $H_0: \beta_1 = 0, \quad H_1: \beta_1 \neq 0$.

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

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

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

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

Il ne faut pas conclure que la variable explicative X_I , n'apporte aucune contribution significative. Une régression simple entre Y et X_I 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_I (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 IV : BANQUE DE DONNEES

- INTRODUCTION
- ÉTUDE STATISTIQUE SUR LES ENREGISTREMENTS
- RÉPARTITION DES DONNÉES
- CHOIX DES VARIABLES (CRITÈRES STATISTIQUE)
- DESCRIPTION DES VARIABLES DÉPENDANTES ET INDÉPENDANTES

CHAPITRE IV : BANQUE DE DONNÉES

4.1. Introduction.

Dans ce travail les données utilisées sont présentée 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 schémas 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.

4.2. Étude statistique sur les enregistrements.[13],[14]

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 épures 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.

4.3. 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 épicentrales 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 D1 à D3 comme présenté dans le tableau (4.1), catégories de sites (sol raide, dépôt d'alluvion superficielle, dépôt d'alluvion profond). Le tableau (4.2) montre le détail de cette sélection pour chaque classe.

Tableau (4.1) : classes des distances

	D1	D2	D3
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 (4.2) : répartition de la banque de données

	sol raide			Alluvion superficielle			Alluvion profonde		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
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	131			139			141		

4.3.1. 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 :

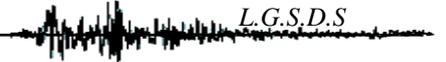
1. Faire une régression avec moindres carrés globale avec toutes les variables et éliminer les variables exogènes avec t -test qui ne sont pas dominant.

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.



4.3.2. Description des 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.

Distance :

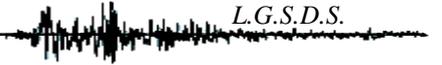
La régression a été accomplie en utilisant à la fois la distance épacentrale et distance focale, cette dernière est définie comme la plus petite distance entre la station d'enregistrement et la projection de la surface de la faille de rupture (Joyer et Boore 1981).

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.



CHAPITRE V :

PRESENTATION ET DISCUSSION

DES RESULTATS DE LA

REGRESSION

- INTRODUCTION
- FORMULATION DU MODÈLE
- L'ÉQUATION PRÉDICTIVE
- DISCUSSION DES RÉSULTATS DE LA RÉGRESSION
- COMPARAISON DE PSEUDO SPECTRES DE VITESSE SIMULÉS ET PRÉDICTIVES
- EFFET DE LA MAGNITUDE
- EFFET DE DISTANCE
- EFFET DE SITE
- COMPARAISON AU RÈGLEMENTS



CHAPITRE V : PRÉSENTATION ET DISCUSSION DES RÉSULTATS POUR PSV.

5.1. 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 le pseudo spectre de réponse de vitesse, en le décomposant en 14 points.

Quatorze lois de régression seront présentées pour l'estimation des lois de spectre de vitesse dans 14 point de fréquence, chacune de ces lois sera fonction des variables indépendantes (M caractérise la taille du séisme, D caractérise la localisation de site par rapport à la source, S_1, S_2 caractérisent la nature du site).

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

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

5.2. Formulation de Modèle :

La forme générale de la fonction adoptée pour la modélisation des atténuations des mouvements sismiques forts est représentée par l'expression :

$$f(Y) = a + f_1(M) + f_2(D) + f_3(S) + \varepsilon \quad (5.1)$$

Où Y : le paramètre à prédire, $f_1(M)$: fonction de magnitude, $f_2(D)$: fonction de la distance epicentrale.

$f_3(S)$: fonction des paramètres considérés de sol, ε : variable représentant l'incertitude.

Les formes fonctionnelles de $f(Y)$, $f_1(M)$, $f_2(D)$ et $f_3(S)$ ont été déterminées conformément à la procédures statistiques ACE « Alternating Conditional Expectation » (Breiman and Friedman, 1985 ; Sabetta and pugliese, 1987) et a abouti aux résultats suivantes :

$$f(Y) = \log(Y); \quad f_1(M) = bM; \quad f_2(D) \approx \log(D) \quad (5.2)$$

Concernant $f_1(M)$ la magnitude quadratique a été introduit (Boore et all ,1993 ; Sadigh,1993), mais le coefficient correspondant est statistiquement non significatif.

A cause de l'existence de l'atténuation géométrique, l'amortissement des matériaux et la diffraction (atténuation anélastique), la fonction $f_2(D)$ a été reformulée comme suit:

$$f_2(D) = c \log\sqrt{D^2 + h^2} + dD \quad (5.3)$$

Où le premier terme représente l'atténuation géométrique et le second représente l'inélastique atténuation. Le paramètre h représente la profondeur du séisme déterminé par la régression.

La transformation obtenue par la technique ACE pour la fonction $f_3(S)$ nous a démontré les différents effets d'amplifications et la fonction devient :

$$f_3(S) = dS_1 + eS_2 \quad (5.3bis)$$

Les variables S1 et S2 sont liées à la classification géologique du site. Ils prennent la valeur 0 pour sols rocheux, (1,0 resp) pour les sols fermes et (0,1 resp) pour les sols meubles.

La forme finale de l'équation (6.1) est :

$$\log(Y) = \beta_0 + \beta_1 M + \beta_2 \log(\sqrt{D^2 + h^2}) + \beta_3 S_1 + \beta_4 S_2 + \sigma \quad (5.4)$$

Où M représente le magnitude, D représente le distance épacentrale et σ est l'écart type du logarithme de Y.

β_0 : Théoriquement ce paramètre représente la moyenne des $\log(PSV)$ 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.

σ : 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 .

5.3. L'équation prédictive :

L'équation (5.4), en remplaçant Y par PSV en cm/sec, devient une base de données fréquentielle pour 14 fréquences, convenablement triées entre 0.25Hz et 25 Hz, pour lesquelles la réponse a été calculée avec la connaissance de données appartenant à la classe $S_i D_j M_k$.

Pour chaque fréquence, un tableau de quatre colonnes et quarante lignes, a été préparé de façon à avoir la variable dépendante à la première colonne et les trois variables explicatives M, D et S respectivement dans les autres colonnes. L'équation (5.4) devient :

$$\log(PSV) = \beta_0 + \beta_1 M + \beta_2 \log(\sqrt{D^2 + h^2}) + \beta_3 S_1 + \beta_4 S_2 \mp \sigma \quad (5.5)$$

Les symboles utilisés pour caractériser les estimateurs des différents paramètres sont consignés dans le tableau suivant :

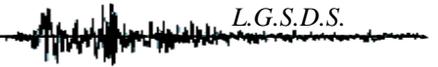
Tableau 3 : symboles utilisés pour caractériser les estimateurs

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

Chaque estimateur représente la moyenne de l'intervalle de confiance lui correspondant et l'écart type représente la fluctuation autour de la moyenne et le paramètre correspondant se situe dans le nuage de points défini par le rayon σ_i et le centre de cet intervalle β_i .

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.

Les tableaux (5.1) à (5.14) donnent les résultats de la régression pour chaque fréquence sous forme de trois sous tableaux : statistique de la régression qui mesure le degré de l'explication de la variable dépendante par les variables indépendantes en utilisant le coefficient de détermination R^2 et l'erreur type. Le deuxième tableau est pour l'analyse de variance qui donne la somme des carrés et la moyenne des carrés ainsi que le coefficient f de Fisher dont nous guide à dire que la signification globale de la régression est satisfaisante ou non. Le tableau des estimateurs dont le rôle est de voir pour chaque estimateur la valeur calculé, l'erreur type, l'intervalle de confiance et le coefficient de Student. Ce tableau a pour la détermination de la signification marginale de chaque variable en présence des autres.



5.4. Discussion des résultats de la régression :

Analyse de variance :

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 la variance de Y_i au tour de l'équation de régression (c.à.d. l'estimation de la variance des erreurs σ^2), estimation qui sera subséquentement utile dans le calcul d'intervalles de confiance ; on pourra également en déduire le coefficient de détermination multiple R^2 .

Pour évaluer la qualité de l'ajustement linéaire de l'équation de régression entre la variable indépendante et l'ensemble des variables explicatives, on recourt au coefficient de détermination multiple, noté R^2 .

Le coefficient R^2 permet d'évaluer la proportion de la variable des Y_i au tour de la moyenne qui expliquée par l'ensemble des variables explicatives retenues dans l'équation de régression.

D'après les tableaux de régression (5.1 à 5.14) les coefficients de détermination multiple sont acceptables et qui ont des valeurs autour de 90% sauf quelques fréquences (3.33hz, 10hz et 6.66hz) où il est au voisinage de 70%. Ces valeurs nous conduisent à dire que les variables explicatives (magnitude, distance et conditions de sol) ont bien expliqué la variable dépendante ($\log(\text{PSV})$).

5.5. Test d'hypothèse :

Signification de la régression dans son ensemble (Fisher) :

On veut tester l'hypothèse selon laquelle les trois variables explicatives (valeurs des magnitudes, les conditions de site et l&a distance) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\log(\text{psv})$.

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

$$H_0 : b_0 = b_1 = b_2 = b_3 = 0$$

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

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

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

Pour $k=4$, $n-k-1=35$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;4,35}=2.63$

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

D'après les tableaux d'analyse de variance on a $F > 2.63$ alors 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 b_i qui est significativement différent de zéro.

On peut dire alors que la régression des 14 équations est significative dans son ensemble.

Test de la contribution marginale de chaque variable 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.

Les tableaux des estimateurs donnent les valeurs des coefficients et leurs écart type ainsi que le coefficient de Student correspondant, par le quel on détermine la contribution marginale de chaque estimateur. Le test se fait comme suit :

au seuil de signification $\alpha=0.05$, Il s'agit alors de tester les hypothèses suivantes :

$$H_0: b_j = 0, \quad j = 1, 2, \dots, k. \quad H_1: b_j \neq 0.$$

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

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

$$n-k-1=35 \Rightarrow t_{0,025} = 2.03$$

en analysant les tableaux des estimateurs, on peut dire :

la variable S1 n'est pas significative pour les fréquences 0.25, 0.33, 0.5, 0.66, 1, 1.33, 5, 6.66, 10, 15 et 25.

La variable S2 n'est pas significative pour les fréquences 0.25, 0.33, 0.5, 0.66, 1, 1.33, 10, 15.

- Les autres variables ont une signification marginale dans toutes les fréquences.
- Toutes les variables sont significatives dans les fréquences 2, 2.5 et 3.33.

**Tableau (5.1) : Résultats de la régression pour le point $f=0.25\text{hz}$**

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple	0.853				
Erreur-type	0.355				
Observations	40.				
Analyse de variance					
	<i>Degré de liberté</i>	<i>de Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4.000	11.811	2.953	23.466	2.63
Résidus	35.000	4.404	0.126		
Total	39.000	16.215			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.754	0.416	1.814	-0.090	1.599
Variable M	0.699	0.078	8.955	0.541	0.858
Variable S1	0.177	0.139	1.275	-0.105	0.459
Variable S2	-0.067	0.142	-0.469	-0.355	0.222
Variable D	-2.638	0.470	-5.613	-3.592	-1.684

Tableau (5.2) : Résultats de la Régression pour le point $f= 0.33\text{hz}$

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple	0.855				
Erreur-type	0.345				
Observations	40				
Analyse de variance					
	<i>Degré de liberté</i>	<i>de Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	11.451	2.862	23.928	2.63
Résidus	35	4.187	0.119		
Total	39	15.638			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.812	0.405	2.00	-0.010	1.635
Variable M	0.689	0.076	9.058	0.535	0.844
Variable S1	0.194	0.135	1.436	-0.080	0.469
Variable S2	-0.055	0.138	-0.397	-0.336	0.226
Variable D	-2.647	0.458	-5.777	-3.577	-1.717

**Tableau (5.3) : Résultats de la Régression pour le point $f= 0.5\text{hz}$**

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple					0.857
Erreur-type					0.339
Observations					40

<i>Analyse de variance</i>					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	11.167	2.791	24.24	2.63
Résidus	35	4.029	0.115		
Total	39	15.197			

<i>Estimateurs</i>					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.818	0.397	2.057	0.010	1.625
Variable M	0.687	0.074	9.209	0.536	0.839
Variable S1	0.218	0.132	1.645	-0.051	0.488
Variable S2	0.032	0.136	0.238	-0.243	0.308
Variable D	-2.692	0.449	-5.98	-3.605	-1.780

Tableau (5.4) : Résultats de la Régression pour le point $f= 0.66\text{hz}$:

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple					0.841
Erreur-type					0.347
Observations					40

<i>Analyse de variance</i>					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	10.245	2.561	21.165	2.63
Résidus	35	4.235	0.121		
Total	39	14.481			

<i>Estimateurs</i>					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.629	0.407	1.542	-0.198	1.456
Variable M	0.651	0.0764	8.512	0.496	0.807
Variable S1	0.184	0.136	1.352	-0.092	0.460
Variable S2	0.067	0.139	0.484	-0.215	0.350
Variable D	-2.398	0.408	-5.203	-3.333	-1.462

**Tableau (5.5) : Résultats de la Régression pour le point $f=1.00$ Hz :**

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple					0.875
Erreur-type					0.297
Observations					40

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	10.130	2.532	28.708	2.63
Résidus	35	3.087	0.088		
Total	39	13.218			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.394	0.348	1.133	-0.312	1.101
Variable M	0.629	0.065	9.632	0.497	0.762
Variable S1	0.181	0.116	1.564	-0.054	0.417
Variable S2	0.156	0.119	1.312	-0.085	0.398
Variable D	-2.120	0.393	-5.387	-2.918	-1.321

Tableau (5.6) : Résultat de la Régression pour le point $f=1.33$ Hz :

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple					0.868
Erreur-type					0.293
Observations					40

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	9.232	2.308	26.778	2.63
Résidus	35	3.0180	0.086		
Total	39	12.250			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.3130	0.344	0.909	-0.385	1.011
Variable M	0.588	0.064	9.101	0.457	0.719
Variable S1	0.196	0.114	1.708	-0.0369	0.429
Variable S2	0.212	0.117	1.805	-0.026	0.451
Variable D	-1.904	0.3890	-4.894	-2.694	-1.114

**Tableau (5.7) : Résultats de la Régression pour le point $f=2$ Hz :**

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple	0.906				
Erreur-type	0.226				
Observations	40				

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	8.337	2.084	40.534	2.63
Résidus	35	1.799	0.051		
Total	39	10.137			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.544	0.265	2.046	0.004	1.083
Variable M	0.566	0.049	11.356	0.465	0.668
Variable S1	0.199	0.088	2.247	0.019	0.379
Variable S2	0.263	0.090	2.896	0.078	0.447
Variable D	-1.959	0.300	-6.521	-2.569	-1.349

Tableau (5.8) : Résultats de la Régression pour le point $f=2.50$ Hz :

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple	0.912				
Erreur-type	0.198				
Observations	40				

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	6.939	1.734	43.819	2.63
Résidus	35	1.385	0.039		
Total	39	8.324			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.697	0.233	2.990	0.224	1.171
Variable M	0.517	0.043	11.806	0.428	0.606
Variable S1	0.191	0.077	2.462	0.033	0.349
Variable S2	0.298	0.079	3.739	0.136	0.460
Variable D	-1.857	0.263	-7.046	-2.392	-1.322

**Tableau (5.9) : Résultats de la Régression pour le point $f=3.33$ Hz :**

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple				0.897	
Erreur-type				0.194	
Observations				40	

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	5.510	1.377	36.335	3.63
Résidus	35	1.326	0.037		
Total	39	6.837			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.792	0.228	3.472	0.329	1.256
Variable M	0.445	0.042	10.392	0.358	0.532
Variable S1	0.185	0.076	2.440	0.031	0.340
Variable S2	0.387	0.078	4.964	0.229	0.546
Variable D	-1.675	0.257	-6.495	-2.199	-1.151

Tableau (5.10) : Résultat de la Régression pour le point $f=5$ Hz :

<i>Statistiques de la régression</i>					
Coefficient de détermination multiple				0.773	
Erreur-type				0.259	
Observations				40	

Analyse de variance					
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	3.521	0.880	13.035	2.63
Résidus	35	2.363	0.067		
Total	39	5.884			

Estimateurs					
	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	1.224	0.304	4.020	0.606	1.843
Variable M	0.344	0.057	6.025	0.228	0.460
Variable S1	0.165	0.101	1.622	-0.041	0.371
Variable S2	0.405	0.104	3.888	0.193	0.616
Variable D	-1.590	0.344	-4.619	-2.289	-0.891

**Tableau (5.11) : Résultats de la Régression pour le point $f=6.66$ Hz :**

Statistiques de la régression					
Coefficient de détermination multiple					0.725
Erreur-type					0.252
Observations					40

Analyse de variance					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	4	2.479	0.619	9.732	2.63
Résidus	35	2.229	0.063		
Total	39	4.709			

Estimateurs					
	Coefficients	E-type	Statistique t	Limite inférieure	Limite supérieure
Constante	1.478	0.295	4.997	0.877	2.079
Variable M	0.319	0.055	5.757	0.207	0.432
Variable S1	0.130	0.098	1.316	-0.070	0.330
Variable S2	0.228	0.101	2.262	0.023	0.434
Variable D	-1.701	0.334	-5.088	-2.380	-1.022

Tableau (5.12) : Résultat de la Régression pour le point $f=10$ Hz :

Statistiques de la régression					
Coefficient de détermination multiple					0.697
Erreur-type					0.236
Observations					40

Analyse de variance					
	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	4	1.855	0.463	8.273	2.63
Résidus	35	1.962	0.056		
Total	39	3.818			

Estimateurs					
	Coefficients	E-type	Statistique t	Limite inférieure	Limite supérieure
Constante	1.263	0.277	4.550	0.699	1.826
Variable M	0.294	0.052	5.654	0.188	0.400
Variable S1	0.016	0.092	0.181	-0.171	0.204
Variable S2	0.012	0.094	0.1291	-0.180	0.205
Variable D	-1.616	0.313	-5.154	-2.253	-0.980

**Tableau (5.13) : Résultat de la Régression pour le point f=15 hz :**

<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0.729
Erreur-type	0.228
Observations	40

Analyse de variance

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	2.079	0.519	9.956	2.63
Résidus	35	1.827	0.052		
Total	39	3.906			

Estimateurs

	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.446	0.267	1.665	-0.097	0.989
Variable M	0.306	0.050	6.090	0.204	0.408
Variable S1	-0.014	0.089	-0.157	-0.195	0.167
Variable S2	0.117	0.091	1.285	-0.068	0.303
Variable D	-1.363	0.302	-4.504	-1.978	-0.748

Tableau (5.14) : Résultats de la Régression pour le point f=25 hz :

<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0.807
Erreur-type	0.221
Observations	40

Analyse de variance

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	4	3.216	0.804	16.388	2.63
Résidus	35	1.717	0.049		
Total	39	4.934			

Estimateurs

	<i>Coefficients</i>	<i>E-type</i>	<i>Statistique t</i>	<i>Limite inférieure</i>	<i>Limite supérieure</i>
Constante	0.0390	0.259	0.150	-0.488	0.566
Variable M	0.357	0.048	7.338	0.258	0.456
Variable S1	0.093	0.086	1.078	-0.082	0.269
Variable S2	0.281	0.088	3.165	0.100	0.461
Variable D	-1.564	0.293	-5.331	-2.160	-0.968



Les valeurs des estimateurs associés aux variables non significatives sont alors remplacées par des zéros. On obtient alors une récapitulation de ces valeurs dans le tableau (5.15) qui résume les résultats selon le modèle :

$$\log(PSV) = a + bM + c \log(\sqrt{D^2 + h^2}) + dS_1 + eS_2 \mp \sigma .$$

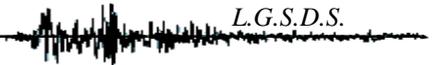
Tableau (5.15) : Résultats de la régression pour le pseudo spectre de vitesse

fréquence	Constant a	Mag.coef b	Dist.coef c	Site.coef d	Site.coef e	σ
0.25	0.7544	0.6992	-2.6377	0	0	0.355
0.33	0.8123	0.6897	-2.6476	0	0	0.345
0.5	0.8185	0.6879	-2.6926	0	0	0.339
0.66	0.6291	0.6519	-2.3982	0	0	0.347
1	0.3946	0.6297	-2.1201	0	0	0.297
1.33	0.3130	0.5883	-1.9043	0	0	0.293
2	0.5441	0.5668	-1.9593	0.1994	0.2634	0.226
2.5	0.6976	0.5171	-1.8574	0.1917	0.2984	0.198
3.33	0.7927	0.4454	-1.6756	0.1859	0.3876	0.194
5	1.2246	0.3446	-1.5905	0	0.4052	0.259
6.66	1.4784	0.3199	-1.7014	0	0.2290	0.252
10	1.2632	0.2947	-1.6169	0	0	0.236
15	0.4461	0.3063	-1.3635	0	0	0.228
25	0.0390	0.3578	-1.5646	0	0.2812	0.221

5.6. Comparaison entre les spectres simulés et ceux des lois prédictives:

La figure (5.1) montre que l'ensemble des points qui constituent la courbe prédictive de PSV sont en concordances avec ceux obtenus à partir de la base de données. Cette observation est valable pour la majorité des classes magnitude, distance et site.

Lorsque, pour une classe SDM donnée, la valeur de l'accélération est importante, l'allure du spectre est monotone et décroissante, par contre si cette valeur est faible la courbe prend une allure en forme d'une cloche. Cela a pour conséquence de resserrer les courbes en hautes fréquences et de les séparer en basses fréquences.



5.7. Effet de la magnitude :

La figure (5.2) illustre l'effet de la magnitude sur le spectre de réponse, cet effet est plus important en basses fréquences. Les valeurs spectrales correspondantes augmentent sensiblement lorsque la magnitude augmente de M1 jusqu'à M5 et ce pour un type de sol et distance épacentrale données. Le pseudo spectre de vitesse est légèrement sensible à la magnitude pour les fréquences supérieures à 6Hz. Ces phénomènes traduisant la relation qui existe entre le moment sismique et la fréquence de coin édité par (Brune 1970).

L'intervalle de fréquence autour de 6Hz joue un rôle de frontière qui sépare deux bandes fréquentielles (basses fréquences et hautes fréquences), dont les valeurs du spectre correspondant aux basses fréquences sont fortement influencées par la magnitude.

Dans un site rocheux un passage d'une magnitude M1 à M2 correspond à une augmentation de 70% à 90% des valeurs spectrales pour des fréquences inférieures à 6Hz. Le même pourcentage d'augmentation est obtenu lorsqu'on passe d'une catégorie de magnitude à une catégorie supérieure.

Pour un site ferme l'augmentation des valeurs spectrales est de 100% à 150% en passant d'une catégorie de magnitude à une catégorie supérieure. Pour un site meuble l'augmentation peut atteindre 200%.

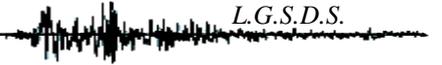
5.8. Effet de distance :

La figure (5.3) montre la comparaison des courbes représentatives des pseudo-spectres en vitesse pour les champs lointains, champs intermédiaires et proches pour une magnitude et une nature de site données.

Le fuseau des courbes spectrales représentant les différentes distances épacentrales reste pratiquement constant en allant des basses fréquences vers les hautes fréquences et ce quelque soit la magnitude.

Plus la distance épacentrale augmente, plus les amplitudes spectrales diminuent, cette observation est valable pour toutes les classes étudiées (SiD1Mj, SiD2Mj, SiD3Mj).

Les amplitudes spectrales sont au maximum 7 fois plus importantes en champ proche (classe D1) qu'en champ lointain (classe D2). En haute fréquence, ce rapport est égal à 4.



5.9. Effet de site :

La figure (5.4) illustre l'influence de la nature de sol sur les spectres de réponse en vitesse.

Cette figure montre que les valeurs des amplitudes spectrales augmentent quand on passe de sol rocheux au sol meuble pour la plage de fréquence allant de 1HZ à 10Hz environ. Les amplitudes peuvent passer du simple au double en passant du sol rocheux au sol meuble. En dehors de cette plage fréquentielle, les tendances entre les amplitudes spectrales associées aux sols fermes et meubles s'inversent tout en restant supérieur aux amplitudes spectrales obtenues sur les sols rocheux.

Les amplitudes spectrales associées au sol meubles subissent donc en filtrage passe bande (entre 1Hz et 10Hz) de sorte que les amplitudes qui sont en dehors de cette bande fréquentielle ne subissent pas une amplification.

La déamplification des spectres associées au sol meuble est tel que les amplitudes spectrales qui leurs sont associées deviennent inférieures à celles au sol rocheux pour les fréquences inférieures à 0.5Hz.

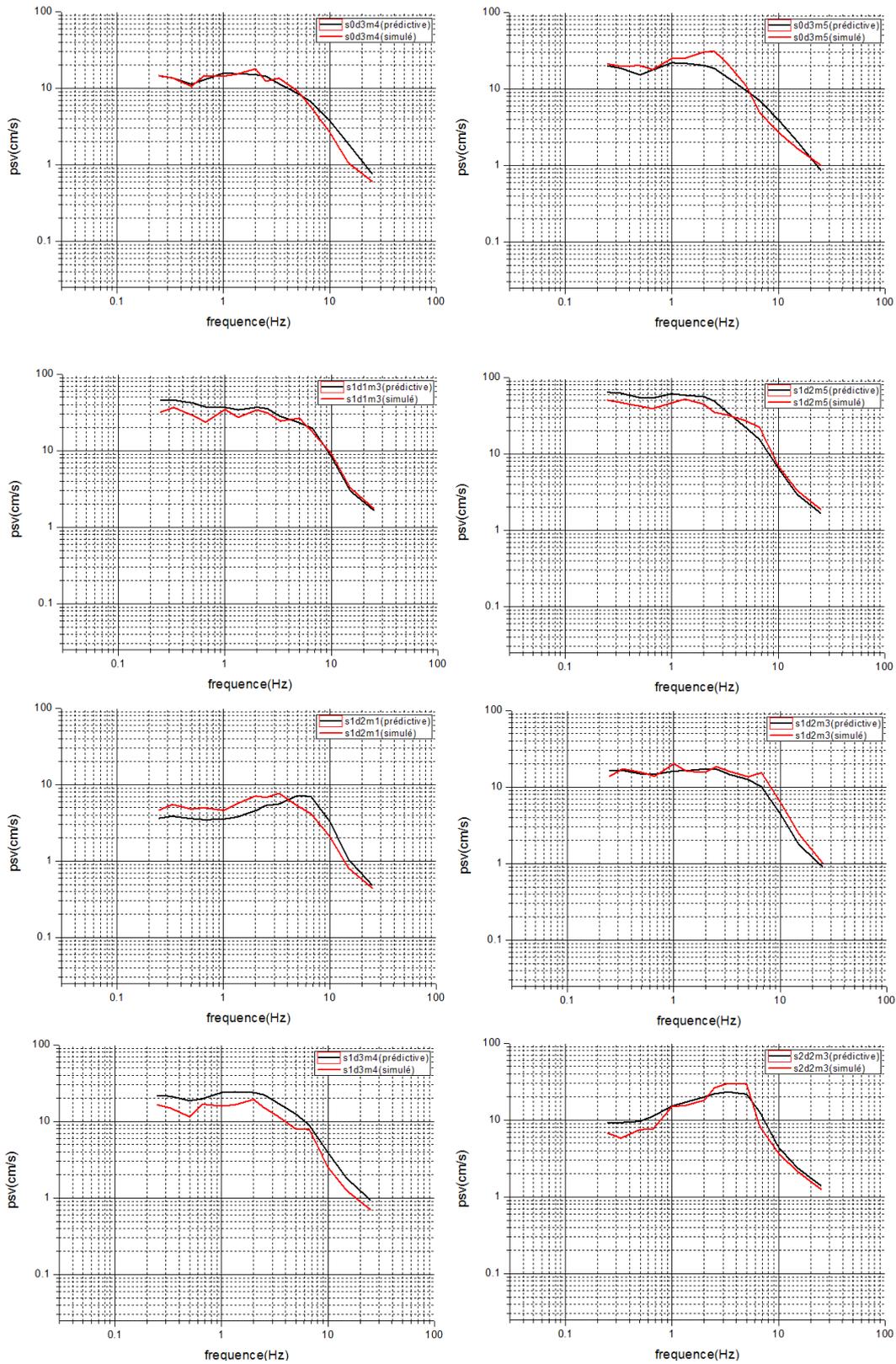


Figure (5.1) : Comparaison entre les courbes de PSV simulées et celles des lois prédictives.

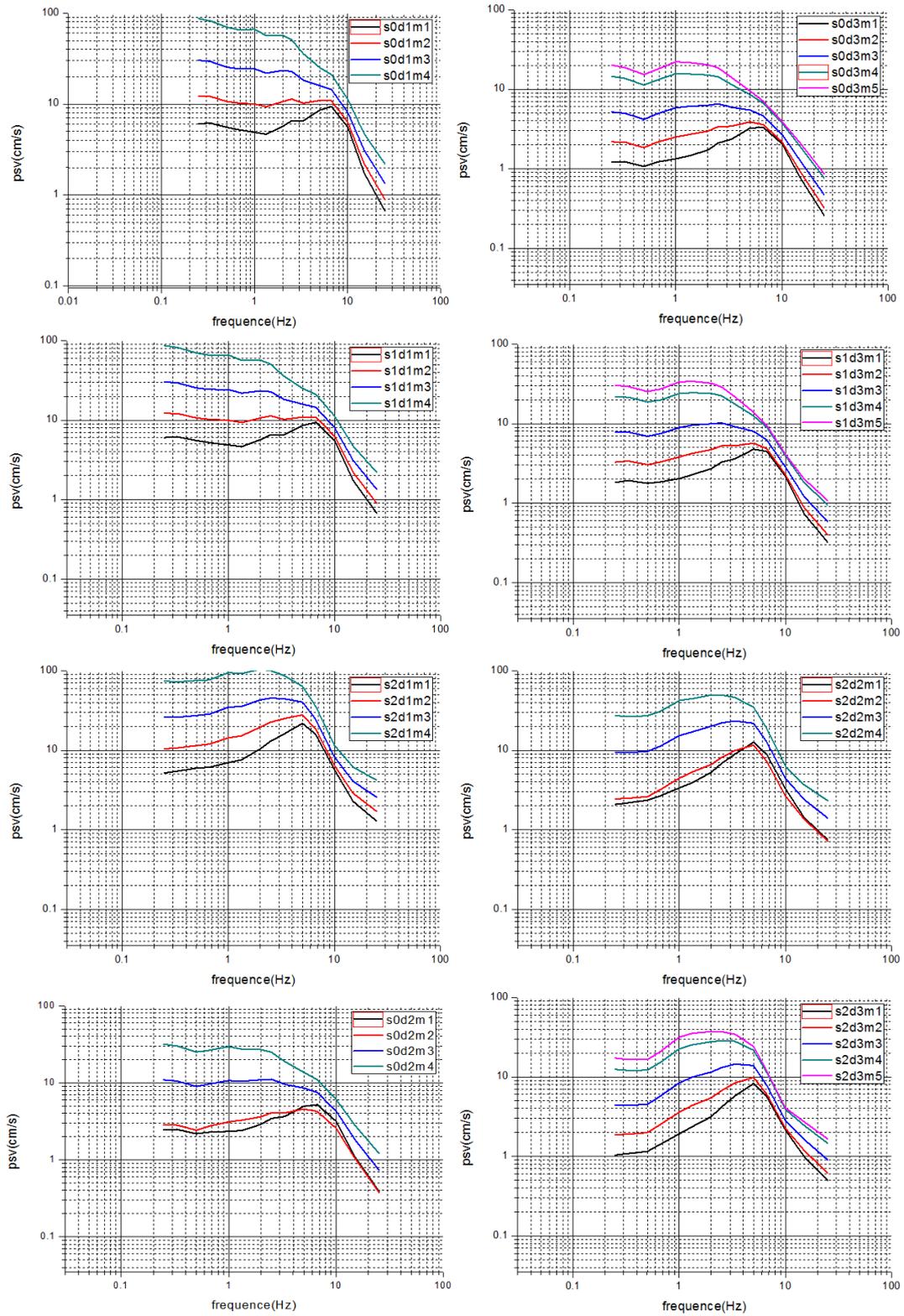


Figure (5.2) : effet de la magnitude sur le PSV pour différentes classes distance-sol.

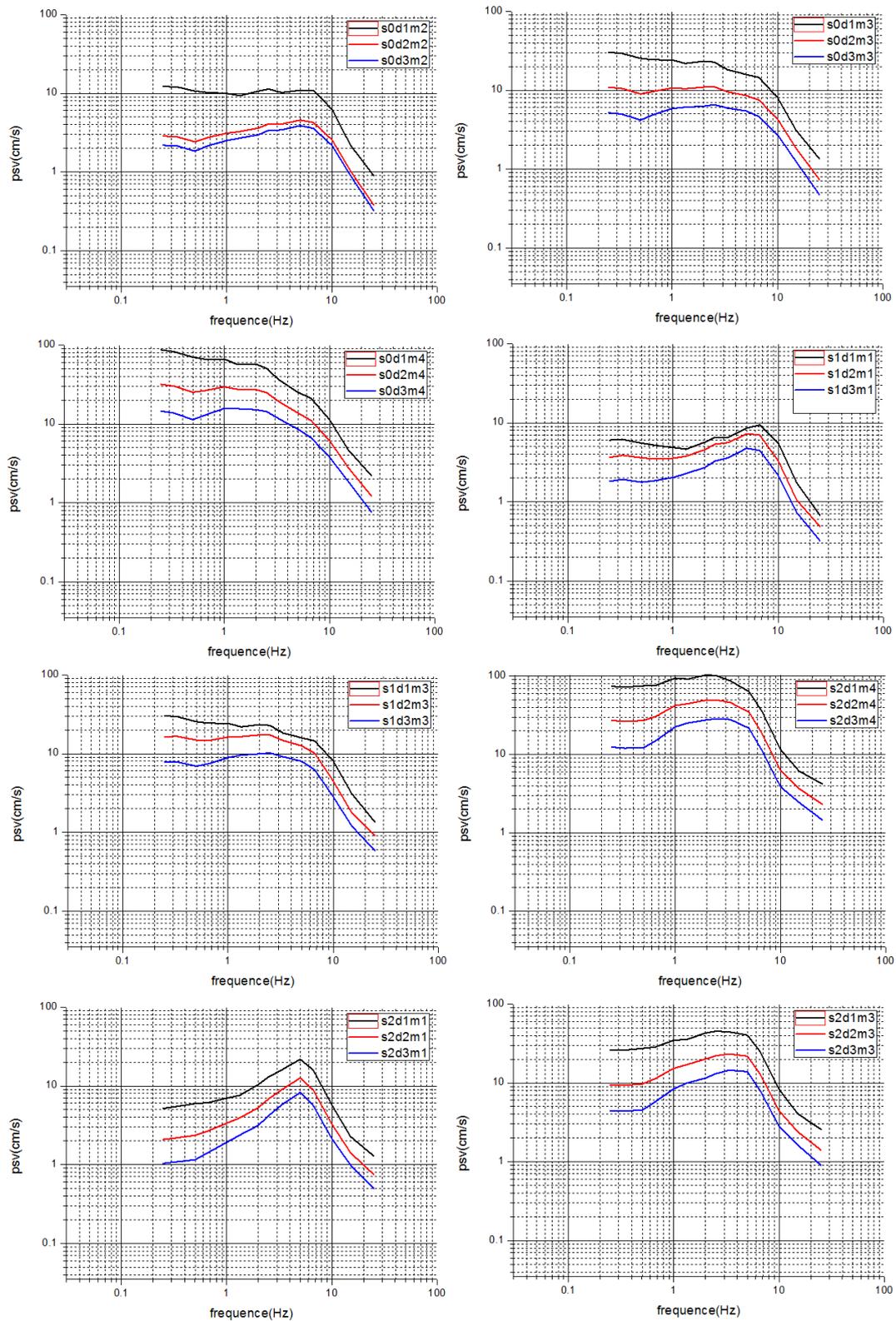


Figure (5.3) : effet de la distance épacentrale sur le PSV pour différentes classes magnitude-sol.

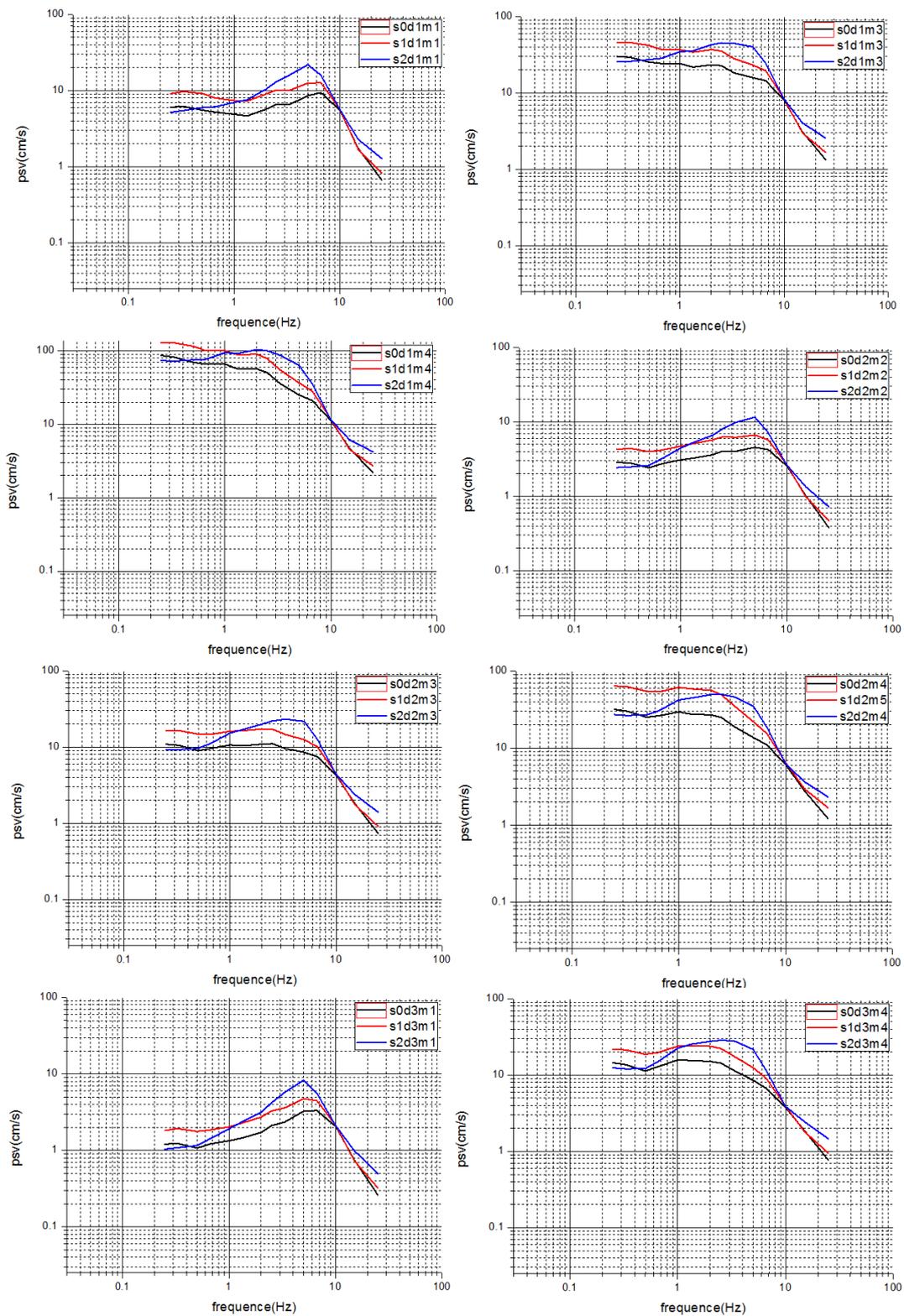
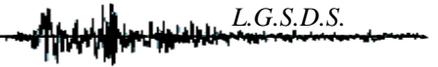


Figure (5.4) : Effet du site géologique sur le PSV pour différentes classes magnitude-distance.



5.10. Comparaison aux règlements.

a. Comparaison entre l'Eurocode et la présente étude

La figure (5.5) montre une comparaison entre les pseudo spectres de vitesses de la présente étude et ceux de l'EUROCODE 8 et de Sabetta et Pugliese (1996) pour différents sites (classes de sol), différentes distances épicentrales et différentes magnitudes. Sabetta et Pugliese ayant comparé leurs résultats au règlement EUROCODE 8, il a été jugé utile de les introduire dans cette étude comparative afin de mieux apprécier nos résultats.

La tendance globale des formes spectrales obtenues, montre que les amplitudes obtenues dans cette étude sont relativement supérieures à celles obtenues avec l'EUROCODE 8 et Sabetta.

L'intervalle de 0.25Hz à 2Hz représente la plage de bonne concordance entre les deux spectres, au-delà de la fréquence 2Hz notre spectre devient supérieur surtout pour les sols meubles.

Deux facteurs peuvent être l'origine de la supériorité de notre spectre par rapport aux autres, le premier, est l'origine de la base de données de chacun ainsi que le nombre d'enregistrements utilisés ; pour notre étude une base de données mondiale a été utilisée où on trouve une diversité importante des données.

Le deuxième facteur, qui peut être l'origine de la supériorité des valeurs spectrales de la présente étude, est la méthodologie adoptée par les différents auteurs pour l'analyse statistique.

b. Comparaison entre l'RPA et la présente étude :

la figure (5.6) montre la bonne concordance entre les spectres du RPA et ceux de la présente étude pour les différentes magnitudes et distances dans les sites rocheux et fermes. L'effet d'amplification est plus forte sur les spectres de notre étude dans les sites meubles, précisément entre la fréquence 1 Hz et 10 Hz .

On peut expliquer la bonne concordance entre les spectres prédictives et ceux du règlement algérien par la présence des enregistrements des séismes qui ont frappé l'Algérie, et ceci confirme les bons résultats trouvés par la régression.

c. Comparaison entre l'UBC et la présente étude :

Une comparaison a été faite entre le « Uniform Building Code » et la présente étude pour les trois types de sols (fig 5.7), on a constaté que :

- L'existence des enregistrements des séismes qui ont frappé les USA dans notre base de données a poussé les résultats à être plus proches au règlement (UBC) avec certaines différences, pour les sols meubles, due principalement à la variabilité des spectres de la présente étude.
- Pour les sols rocheux les spectres de la présente étude coïncident bien sur à ceux de l'UBC et sur les fréquences, idem pour les sols fermes.
- Pour les sols meubles un décalage des deux spectres causé par l'effet de filtrage sur les spectres prédictifs, ce décalage est plus apparent au niveau des basses fréquences où on trouve que les spectres issus des lois sont plus sensibles à l'amplification dynamique du sol.

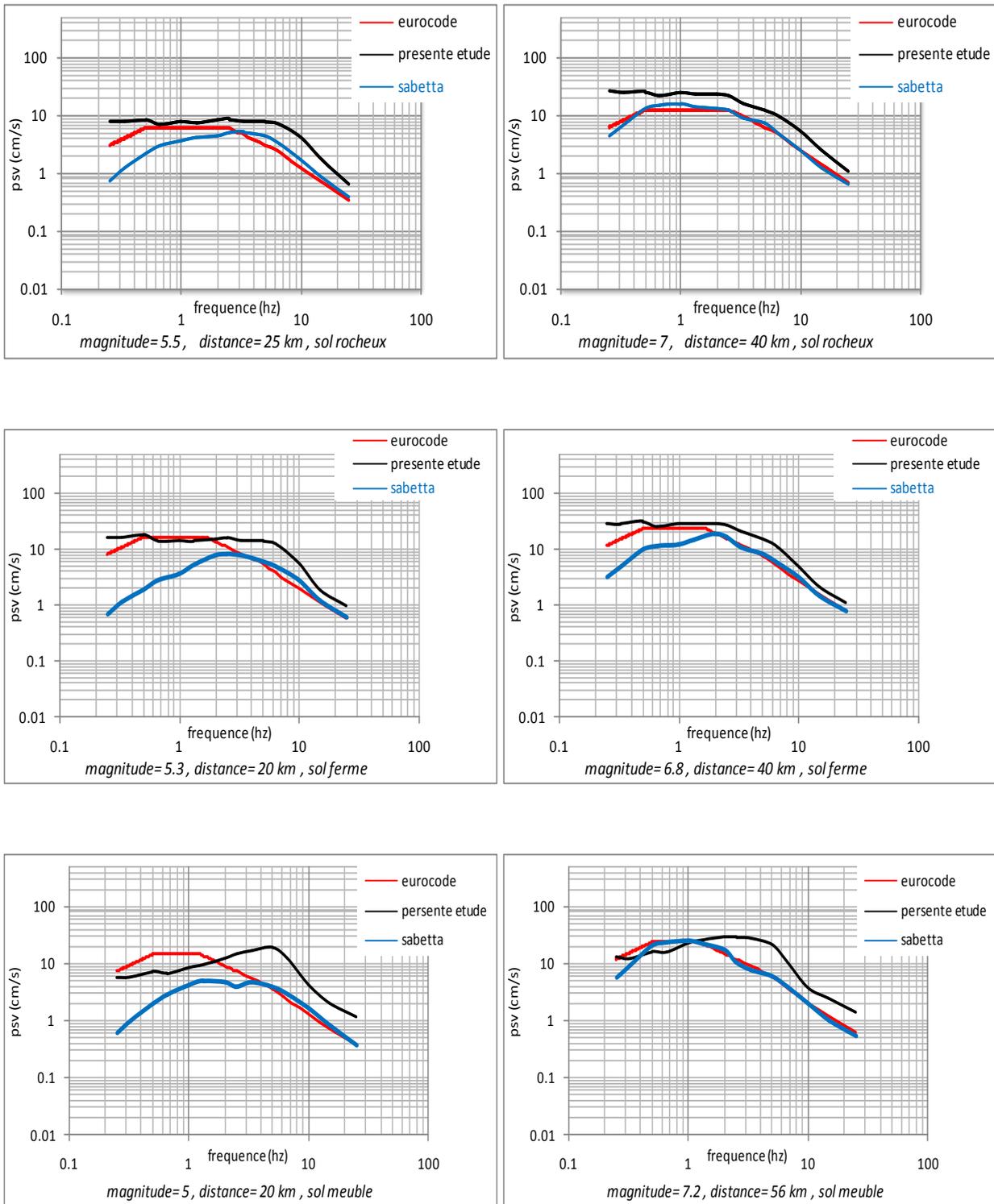
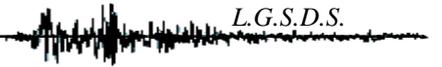


Figure (5.5) : Comparaison des spectres (PSV) issus de la présente étude et ceux de l’Eurocode

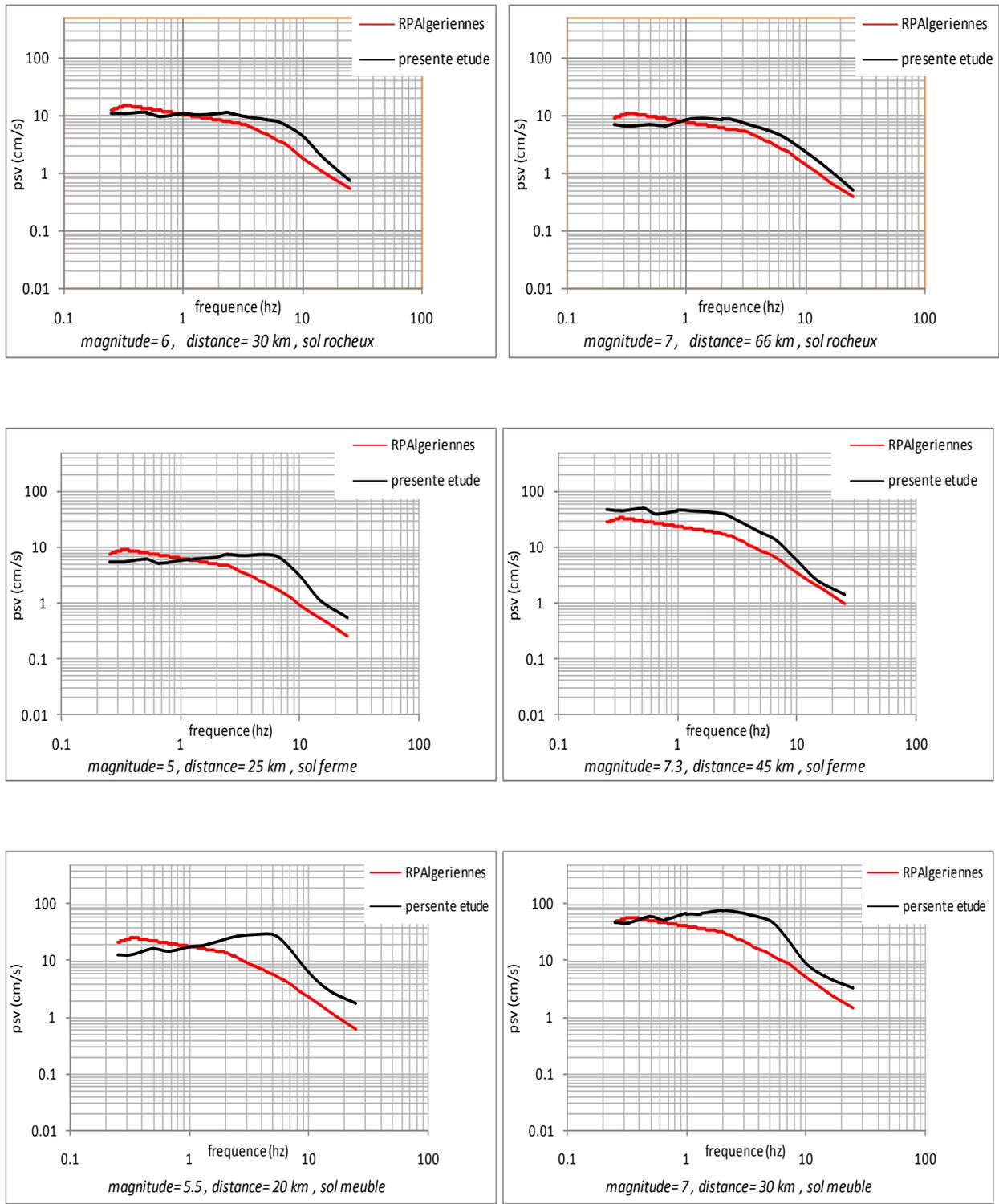
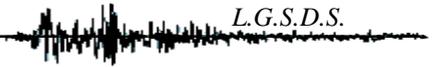


Figure (5.6) : Comparaison des spectres (PSV) issus de la présente étude et ceux de l’RPA

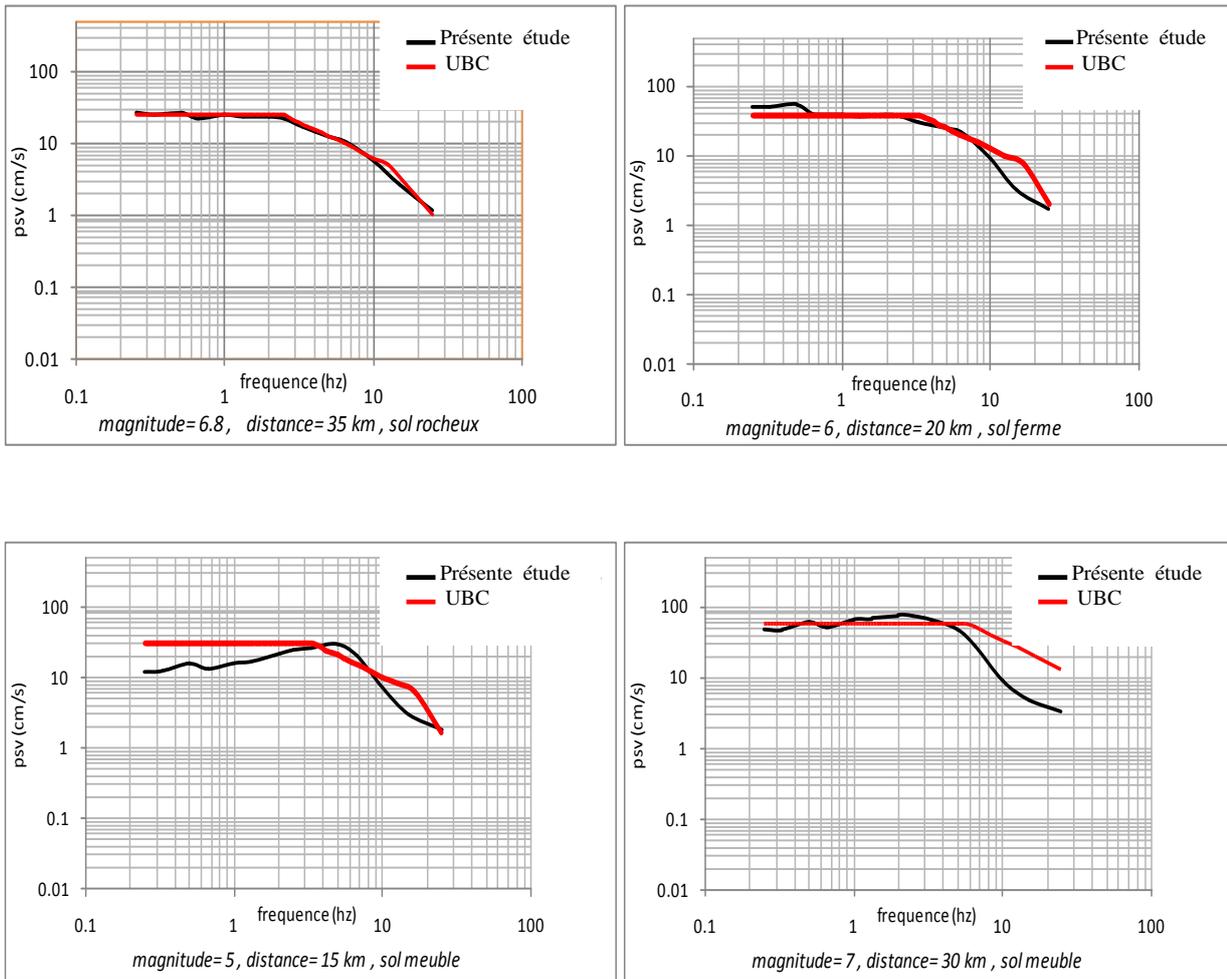
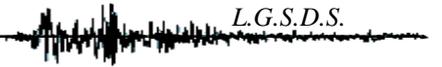


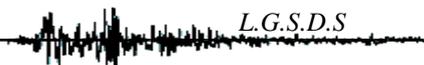
Figure (5.7) : Comparaison des spectres (PSV) issus de la présente étude et ceux de l'UBC

CHAPITRE VI :

ACCÉLÉRATION MAXIMALE

DU SOL

- INTRODUCTION
- SPÉCIFICATION DE MODÈLE
- DISCUSSIONS DES RÉSULTATS DE RÉGRESSIONS
- EFFET DE LA MAGNITUDE, DISTANCE ET SITE SUR LE PGA



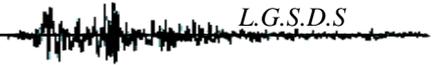
CHAPITRE VI : ACCÉLÉRATION MAXIMALE DE SOL

6.1. Introduction :

Ce qui intéresse les spécialistes en matière de prévention sismique ce sont les dégâts dans les bâtiments que peuvent causer les séismes. Il faut donc évaluer dans chaque région les dégâts les plus importants qui peuvent être causés par un séisme. Ceci revient donc à déterminer l'intensité macrosismique maximale que l'on peut ressentir dans la région en question. Les sismologues préfèrent en fait étudier l'accélération du sol qui est liée à l'intensité mais qui est une grandeur physique alors que l'intensité est une grandeur plus subjective. Déterminer le risque sismique en un endroit c'est donc déterminer l'accélération maximale que le sol est susceptible de subir lors d'un tremblement de terre. Cette accélération est en générale exprimée en fonction de l'accélération de la pesanteur qui vaut environ 9.8 m/s^2 . Cependant, on ne peut pas savoir avec précision les mouvements du sol dus à de possibles futurs séismes tant que les méthodes de prédictions ne sont pas efficaces. Les seuls éléments dont les sismologues disposent sont des statistiques sur l'occurrence et la magnitude des séismes dans une région. Ils sont donc obligés de raisonner en terme de probabilité: une évaluation de risque sismique revient à donner une accélération maximale que le sol peut subir et un pourcentage de "chance" pour que cette valeur soit atteinte au cours d'une période de référence. Ensuite, en fonction de ces valeurs, on classe la zone comme étant plus ou moins "à risques" et on trace des cartes plus globales de zonage sismique recensant le risque déterminé pour chaque zone.

Le pic d'accélération (PGA): c'est la plus grande valeur (en absolue) de l'amplitude lue sur une des composantes horizontales d'un accélérogramme. Cette valeur est appelée le PGA (Peak ground acceleration). Le PGA est important pour les spécialistes des structures parce qu'il est utilisé dans les cartes d'aléa sismique probabiliste (e.g., Trifunac et Brady, 1975a). Ce paramètre seul n'est pas suffisant pour bien caractériser le mouvement du sol. À des fréquences intermédiaires on utilise plutôt le PGV (vitesse) au lieu du PGA pour caractériser l'amplitude du mouvement du sol.

Le paramètre P_gA (ou PGV) nous informe sur l'amplitude du mouvement. D'autres paramètres sont préférés parce qu'ils nous renseignent sur d'autres caractéristiques du mouvement fort.



6.2. Spécification de modèle.

Un modèle à quartes 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 :

$$\log(PGA) = \beta_0 + \beta_1 \cdot M + \beta_2 \cdot \log\left(\frac{1}{\sqrt{D^2 + h^2}}\right) + \beta_3 \cdot S_1 + \beta_4 \cdot S_2 \pm \varepsilon$$

Où M : représente la magnitude.

S1, S2 : représente les conditions de site.

D : représente la distance épacentrale.

β_0 : ce paramètre représente la moyenne des $\log(PGA)$ 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 .

6.3. Discussions des résultats de régressions :

6.3.1. Analyse de variance :

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 la variance de PGA au tour de l'équation de régression (c.à.d. l'estimation de la variance des erreurs σ^2), estimation qui sera subséquentement utile dans le calcul d'intervalles de confiance ; on pourra également en déduire le coefficient de détermination multiple R^2 .

Pour évaluer la qualité de l'ajustement linéaire de l'équation de régressions entre la variable indépendante et l'ensemble des variables explicatives, on recourt au coefficient de détermination multiple, noté R^2 .

D'après le tableau (6.1) de la régression ci-dessous le coefficient de détermination multiple est acceptable et qui est autour de la valeur de 80%. Cette valeur nous conduit à dire que les variables

explicatives (magnitude, distance et conditions de sol) ont bien expliqué la variable dépendante ($\log(\text{PGA})$).

6.3.2. Estimation des paramètres du modèle :

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.

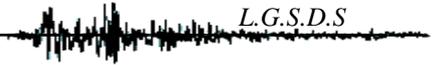
Nous résumons dans le tableau ci contre les diverses valeurs des coefficients obtenues.

Tableau (6.1) : coefficients de régression pour le PGA

Coefficients de régression
$b_0 = 2.252$
$b_1 = 0.357$
$b_2 = -1.578$
$b_3 = 0$
$b_4 = 0.282$

L'équation de régression multiple pour le PGA est donc :

$$\log(\text{PGA}) = 2.252 + 0.357.M - 1.578.\log(\sqrt{D^2 + h^2}) + 0.282.S_2 \pm 0.221$$



6.3.3. Détermination du tableau de variance.

Tableau (6.2) d'analyse de variance

Source de variance	Somme de carré	Degré de liberté	Carrés moyens
Expliquée par la régression	SCR= 3.054	4	CMR= 0.763
	SC _{RES} = 1.414	35	CM _{RES} = 0.040
Totale	SCT= 4.468	39	

6.3.4. Détermination de la variance et du coefficient R².

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

$$S^2 = CM_{RES} = 0.04$$

L'écart-type des résidus est :

$$S=0.2$$

Le coefficient de détermination est :

$$R^2 = \frac{SCR}{SCT} = 0,686$$

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

6.3.5. Tests d'hypothèse de la régression.

On va 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 :

$$CMR = 0.763 \quad CM_{res} = 0.04$$

La valeur observée pour le rapport F est donc : 18.89

Puisque $F = 18.89 > 2.9934$ 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'explication du nombre de pièces assemblées. Il y a au moins un des paramètres β_j qui est significativement différent de zéro.

6.3.6. 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.

6.4. Effet de la magnitude, distance et site sur le PGA :

La figure (6.1) résume l'effet des trois variables sur la valeur maximale de l'accélération ou on peut constater :

- La valeur de PGA est fortement liée à la magnitude du séisme, lorsque la magnitude augmente de M1 à M4 une augmentation importante de la valeur de pga sera enregistré pour toute les classes de sol et distance.
- Une atténuation de pga remarquable est enregistré si on s'éloigne de l'épicentre en gardant l'effet de la magnitude.
- Le passage de sol raide à un sol de faible raideur pour les mêmes classes distance-magnitude induit une amplification très importante de PGA.
- Une amplification de 100% (le double) est enregistré pour un sol meuble par rapport à un sol rocheux. Le facteur le plus prédominant pour la variation de PGA est la magnitude du séisme puis le type de sol.

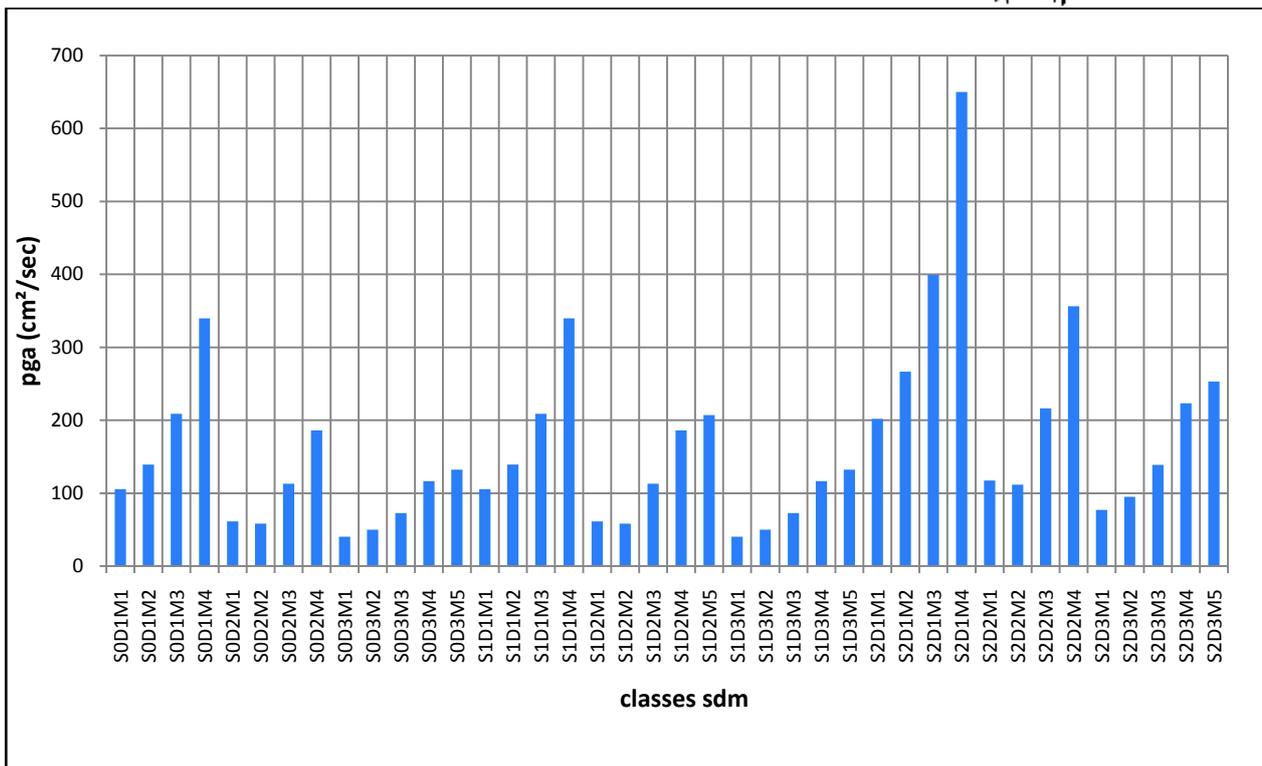
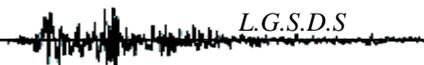


Figure (6.1) : Variation de PGA en fonction de différentes classes (SDM)

CHAPITRE VII : DURÉE DU MOUVEMENT FORT

- INTRODUCTION.
- SPÉCIFICATION DU MODÈLE.
- DISCUSSIONS DES RÉSULTATS DE RÉGRESSIONS.
- EFFET DE LA MAGNITUDE ET LA DISTANCE ÉPICENTRALE SUR DV.



CHAPITRE VII : DURÉE DE MOUVEMENT FORT

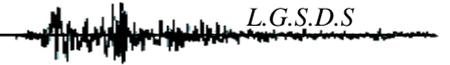
7.1. Introduction :

La durée totale d'un enregistrement ne présente qu'un intérêt pour l'étude et effets des séismes sur les structures, d'une part parce qu'elle est simplement liée à un seuil de déclenchement d'un appareil de mesure et d'autre part parce qu'elle intègre une part importante du signal non significative pour les dégâts qu'ils peuvent induire. Seule la durée de phase forte d'un séisme est considérée comme un paramètre important pour en caractériser sa nocivité. De nombreux travaux consacrés à la détermination de cette durée de phase forte, déterminante pour mesurer le potentiel endommageant d'un séisme ont donné lieu à des expressions variées de ce paramètre, ces expressions peuvent être classées en deux catégories :

La première catégorie de définitions se rattache à la notion de dépassement de seuil. La durée de phase forte d'un enregistrement est généralement la durée qui s'écoule entre la première et la dernière excursion au-dessus d'un niveau d'accélération fixé. La valeur de 0.03g est à rapprocher des travaux de AMBRASEYS et SARMA (1975) repris par PAGE (1972) 0.05g ou par Mc GUIRE et BORHARD (1979) qui proposent un seuil défini comme un pourcentage de l'accélération maximale. Ces définitions présentent deux inconvénients majeurs : le résultat est très sensible à la valeur de seuil fixé ; la nature du signal compris entre les instants qui bornent une phase faible entre deux phases fortes.

BOLT(1973) a introduit une durée uniforme en proposant de ne conserver que les intervalles pendant lesquels le seuil est dépassé. Cette définition réduit considérablement la sensibilité du résultat à la valeur du seuil et introduit la notion importante de cumul dans la définition de la durée de phase forte.

La deuxième catégorie de définitions se base sur la notion de cumul d'énergie du signal. Ce cumul s'exprime à partir de l'intensité d'Arias. La plupart des expressions de la durée qui en découlent est définie comme le temps nécessaire pour qu'une certaine fraction de l'intensité d'Arias soit atteinte. La plus connue est celle de Triffunac et Brady (1975) qui définissent la durée de phase forte correspondant à la fraction de l'intensité d'Arias comprise entre 5% et 95%. d'autres définitions avancées par Van Marcke et Lai (1980) pour qui la durée correspond à l'intervalle de temps durant lequel est spécifiée la probabilité d'excéder un certain pourcentage de la valeur maximale de l'accélération ou par Jj.Bommer et A. Martinez-Pereira (1996) pour qui la durée effective est



comprise entre un instant de départ lié à un seuil de déclenchement de l'énergie cumulée inférieur à un seuil de 10%.

Toutes ces définitions sont basées sur l'étude du signal dans le domaine temporel. Triffunac et Novikova (1995) mentionnent l'intérêt de connaître la contribution par bandes de fréquence à la durée de phase forte. Pour chaque filtre passe-bande appliqué au signal initial, une durée est obtenue sur la base d'un seuil de cumul énergétique. La durée significative est la somme des intervalles de temps pour chaque bande de fréquence.

Enfin, les mêmes auteurs ont essayé de relier la durée de phase forte aux caractéristiques sismotectoniques d'un séisme. Si des corrélations entre la durée, la magnitude et la distance épacentrale ont été proposées, Triffunac et Novikova (1995), prennent en compte la somme de trois termes dans le calcul de la phase forte : un terme lié à la durée de l'événement à la source (dépendant de la magnitude, de la surface de rupture, du mécanisme à la source et de la vitesse des ondes de cisaillement), un terme de propagation d'ondes et un terme d'effet de site.

7.2. 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 :

$$\log(Dv) = \beta_0 + \beta_1.M + \beta_2.\log(\sqrt{D^2 + h^2}) + \beta_3.S_1 + \beta_4.S_2 \pm \sigma$$

Où R : représente la distance.

M : représente la magnitude.

S_1, S_2 : représente le site géologique.

β_0 : Ce paramètre représente la moyenne des $\log(Dv)$ 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.

σ : 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 .



7.3. Discussions des résultats de la régression :

Analyse de variance :

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 la variance de DV au tour de l'équation de régression (c.à.d. l'estimation de la variance des erreurs σ^2), estimation qui sera subséquentement utile dans le calcul d'intervalles de confiance ; on pourra également en déduire le coefficient de détermination multiple R^2 .

Pour évaluer la qualité de l'ajustement linéaire de l'équation de régressions entre la variable indépendante et l'ensemble des variables explicatives, on recourt au coefficient de détermination multiple, noté R^2 .

D'après le tableau (7.1) de la régression ci-dessous le coefficient de détermination multiple est acceptable et qui est autour de la valeur de 90%. Cette valeur nous conduit à dire que les variables explicatives (magnitude, distance et conditions de sol) ont bien expliqué la variable dépendante ($\log(Dv)$).

7.4. Test d'hypothèse :

Signification de la régression dans son ensemble (Fisher) :

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

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 $k=4$, $n-k-1=35$ et au seuil de signification $\alpha=0.05$ on obtient de la table de Fisher, $F_{0.05;4,35}=2.63$

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

D'après les tableaux d'analyse de variance on a $F > 2.63$ alors 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_i qui est significativement différent de zéro, d'où la signification d'ensemble de la regression

Test de la contribution marginale de chaque variable 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.

Le tableau des estimateurs donne les valeurs des coefficients et leurs écart type ainsi que le coefficient de student correspondant, par le quel on détermine la contribution marginale de chaque estimateur. Le test se fait comme suit :

au seuil de signification $\alpha = 0.05$, Il s'agit alors de tester les hypothèses suivantes :

$$H_0: \beta_j = 0, \quad j = 1, 2, \dots, k. \quad H_1: \beta_j \neq 0.$$

Sous H_0 , l'écart réduit devient $\frac{\beta_j}{s(b_j)}$ qui est distribué selon la loi de *Student* avec $n-k-1$ degrés de liberté.

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

$$n-k-1=35 \Rightarrow t_{0.025} = 2.03$$

En analysant le tableau des estimateurs, on peut dire que :

- les variables S1 et S2 ne sont pas significative en présence des autres variables.
- Toutes les autres variables ont une signification marginale.

Tableau (7.1) : résultats de la régression pour la durée du mouvement fort.

<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0.918
Erreur-type	0.268
Observations	40

Analyse de variance

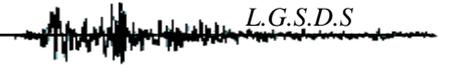
	Degré de liberté	Somme des carrés	Moyenne des carrés	Valeur critique de F
Régression	4	13.486	3.371	2.63
Résidus	35	2.515	0.0718	
Total	39	16.001		

Estimateurs

	Coefficients	Erreur	Statistique t	Limite inférieure	Limite supérieure
Constante	-1.483	0.316	-4.680	-2.127	-0.840
Variable M	0.138	0.058	2.346	0.018	0.258
Variable D	0.931	0.154	6.048	0.619	1.244
Variable S1	0.052	0.105	0.498	-0.161	0.265
Variable S2	0.063	0.103	0.612	-0.146	0.273

La formule finale après correction de la durée du mouvement fort est :

$$\mathbf{Log(DV) = -1.48 + 0.138.M + 0.931.log(\sqrt{D^2 + h^2}) \pm 0.26}$$



7.5. Effet de la magnitude et la distance épacentrale sur D_v :

Les figures (7.1) et (7.2) montre la relation entre la durée de mouvement fort et la magnitude et la distance épacentrale.

La durée du mouvement fort augmente légèrement avec la magnitude pour les champs proches ou la différence entre la durée d'un séisme de magnitude 4.5 et un autre de magnitude 7.5 ne dépasse pas les 6 secondes. Pour les champs lointains la différence pour le même cas peut atteindre les 25 secondes.

Un pourcentage de 51% à 53% de croissance de la durée de mouvement fort, en passant de la magnitude 4.5 à 7.5, est conservé quelque soit la distance épacentrale.

Alors on peut conclure que l'effet de la magnitude sur la durée du mouvement fort est presque linéaire et indépendant de la distance épacentrale. En fonction considérable qu'a pour les champs lointains et moyens.

La figure (7.2) nous montre la forte liaison entre la durée de mouvement fort et la distance épacentrale, dont la durée augmente quatre fois en passant de 20km à 100km quelque soit la magnitude.

Ce phénomène peut être expliqué en revenant à la définition énergétique de la durée de mouvement fort, dont l'énergie correspondante est de 90% . la durée qui correspond cette énergie est importante pour les séismes lointains en tenant compte l'effet de l'atténuation qui fait uniformiser l'énergie sur toute la durée de séisme. L'énergie d'un séisme proche, en absence de l'atténuation due à la distance, est généralement concentré dans une durée courte qui correspond la durée de mouvement fort.

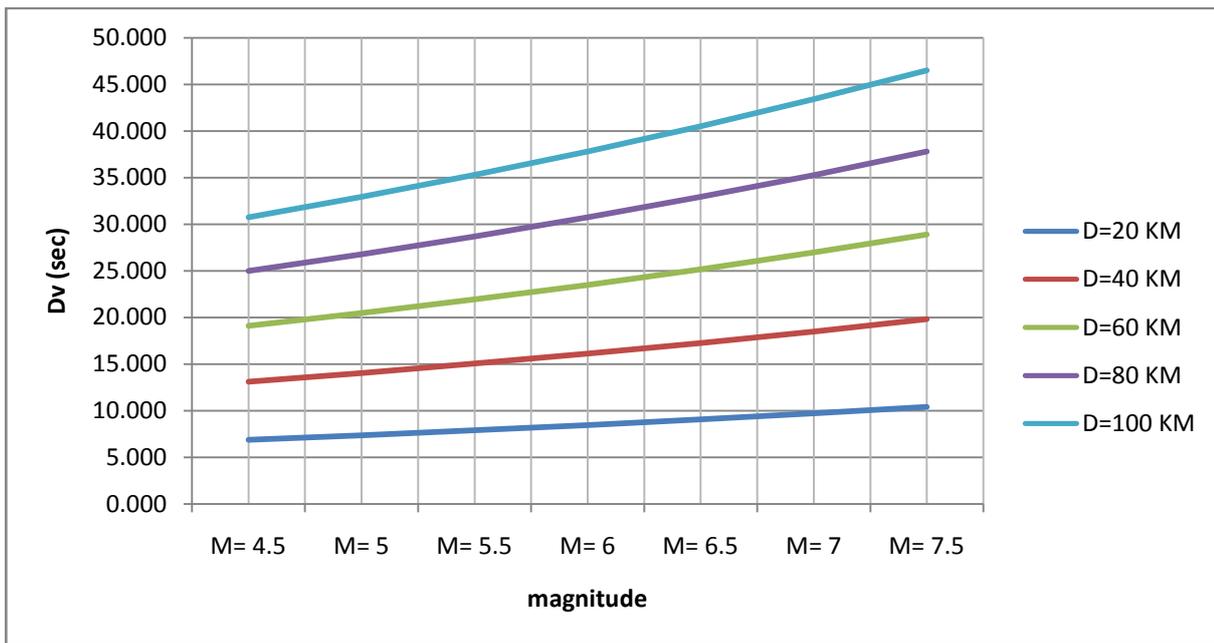
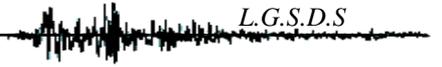


Figure (7.1): effet de la magnitude sur la durée de mouvement fort

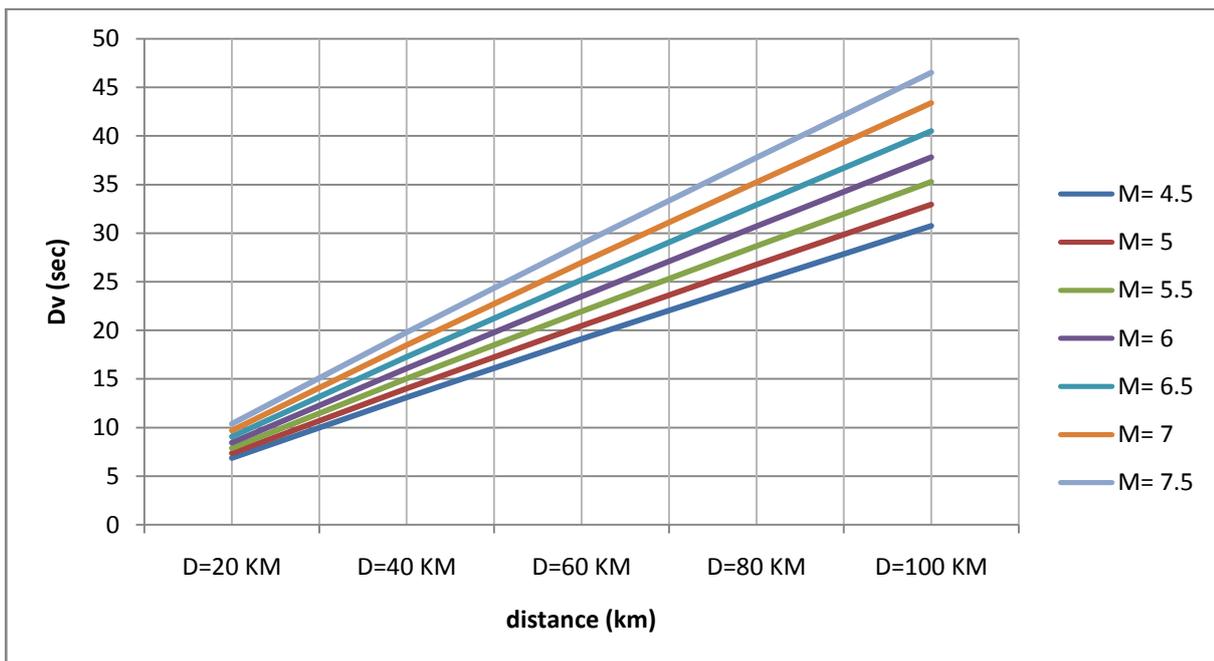


Figure (7.2): effet de la distance épiscoptrale sur la durée de mouvement fort

CONCLUSIONS

8. Conclusions.

Une analyse par régression multilinéaire a été utilisée en vue de l'estimation de pseudo spectres de vitesse (PSV), d'accélérations maximales du sol (PGA) et de durées de mouvements forts du sol (Dv).

L'analyse des lois prédictives pour l'estimation des pseudo-spectres en vitesse a conduit aux constatations suivantes :

- Une bonne concordance des courbes représentatives des spectres obtenues par régression avec celles des spectres moyens des classes correspondantes.
- De façon générale pour un sol et une distance épacentrale données on remarque que l'effet de la magnitude est plus marqué en basse fréquences ($< 5\text{Hz}$) conduisant à des amplitudes de la réponse 20 fois plus importantes pour la classe M5 que pour la classe M1. Ce rapport est réduit à 3 en hautes fréquences ($> 10\text{Hz}$). Un resserrage des courbes est constaté entre 5 et 10 Hz (fréquences intermédiaires)
- Une hiérarchie des formes spectrales est constatée aussi lorsqu'en fait varier la distance épacentrale : les amplitudes maximales sont à associer aux champs proches (D1) et les minimales aux champs lointains (D3) avec un rapport de 7 en basses fréquences et de 4 en hautes fréquences. Les formes spectrales associées au champ intermédiaire (D2) sont comprises entre celles associées au champ proche (D1) et au champ lointain (D3).
- Le niveau d'amplitudes le plus bas obtenu pour les sols est à associé aux sols rocheux (S0). En effet on remarque que quelque soit la classe magnitude-distance, les courbes représentatives du spectre de la classe (S0) se situe en deçà de celles associées aux autres catégories de sites. Le site meuble (S2) se caractérise par les niveaux d'amplitudes les plus élevés dans la plage de fréquences intermédiaires comprises entre 1 et 10 Hz environ (amplification). Ces amplitudes atteignent des rapports de 3 par rapport à celles de sites rocheux (S0) et de 2 par rapport à celle du site ferme (S1).
- La hiérarchie des amplitudes associées aux sites fermes (S1) et meubles (S2) s'inverse en basses fréquences tout en restant supérieur à celles des sites rocheux (S0). Il est toutefois à noter que les amplitudes obtenues pour les sites rocheux (S0) sont voisines à celles

associées aux sols meubles (S2) en basses fréquences et à celles associées aux sols fermes (S1) en hautes fréquences.

- Les formes spectrales résultantes de la présente étude montre une bonne adéquation avec celles données par les codes parasismiques en vigueur considéré dans ce travail (EUROCODE 8, RPA et UBC). Les amplitudes spectrales obtenues dans cette étude pour les sols S0 et S1 sont supérieures à celles proposées par l'Eurocode 8 et ce sur toute la gamme de fréquence. En ce qui concerne les sols meubles S2, la même constatation est faite pour la gamme de fréquences supérieurs à 2Hz alors que la tendance s'inverse en basses fréquences (< 2Hz). La comparaison des niveaux d'amplitudes spectrales résultants de notre étude avec celles du RPA montre globalement une inversion des niveaux d'amplitudes autour de la fréquence 1Hz et ce pour toutes les catégories de sol. A l'inverse des deux autres codes parasismiques, le code UBC propose des amplitudes spectrales très comparables à celles obtenues dans la présente étude. Dans ce dernier cas les écarts d'amplitudes vont en croissant quand on passe du sol rocheux (S0) au sol meuble (S2) avec un rapport entre 1 et 3.

Par ailleurs, les lois de régression pour le PGA ont fait ressortir de manière claire que :

- L'effet de la magnitude se traduit par une augmentation de la valeur du PGA pour les magnitudes croissantes. L'augmentation du PGA est en moyenne d'ordre 3 quelque soit la classe site-distance considérée.
- L'effet de la distance épacentrale se traduit par des valeurs de PGA en champ proche (D1) trois fois supérieur à celles obtenues en champ lointain (D3) et ce quelque soit la catégorie site-magnitude considérée.
- L'influence du site se traduit par des amplifications de la valeur du PGA de l'ordre de 200% en allant de sites rocheux aux sites meubles.

On remarque que l'effet de la magnitude (taille du séisme) est prépondérant par rapport à l'effet de la distance et du site géologique.

- L'analyse concernant la durée de phase de mouvement fort de sol a montré que celle-ci ne subit pratiquement aucune influence du type de sol. Sa valeur est plus grande en champ lointain qu'en champ proche, elle augmente avec les grandes magnitudes. La valeur de D_v croit d'environ 53% en passant de la classe de magnitude M1 à la classe de magnitude M5 et ce quelque soit la distance épacentrale et le type de sol considéré.

Les lois prédictives proposées dans le présent travail sont particulièrement intéressante pour la prédiction de la réponse des structures et de l'intensité des mouvements de sol et pour la même de l'estimation du risque sismique d'une région donnée pour des tremblements de terre futurs.

Cette étude peut être approfondie par l'intégration de nouvelles variables explicatives contrôlant de manière plus précise le processus de tremblement de terre. En effet, des variables contrôlant de manière plus fine la source sismique, la propagation des ondes et les caractéristiques des dépôts sédimentaires peuvent être intégrés à travers des modèles de régression multiple de sorte à en réduire les incertitudes et à produire une meilleure précision d'analyse du risque sismique.

Bibliographie

- [1] A. Abrahamson and W. J. Silva, 1997, *Empirical Response Spectral Attenuation Relations for Shallow Crustal Earthquakes*, *Seismological Research Letters*, Vol. 68, No. 1, 94-127.
- [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] Bolt, B. A. (Editor) (1987). *Seismic Strong Motion Synthetics*, Academic Press, Inc., Orlando, Florida.
- [6] Boore, D. M., W. B. Joyner, and T. E. Fumal (1993). *Estimation of response spectra and peak accelerations from western North American earthquakes: an interim report*, U.S. Geol. Surv. Open-File Rept. 93-509.
- [7] Boore, D. M., W. B. Joyner, and T. E. Fumal (1994). *Estimation of response spectra and peak accelerations from western North American earthquakes: an interim report-- part 2*, U.S. Geol. Surv. Open-File Rept. 94-127.
- [8] Breiman, L. and J. H. Friedman (1985). *Estimating optimal transformations for multiple regression and correlation*, *J. Am. Statist. Assoc.* 80, 580- 619.
- [9] Brune, J. N. (1970). *Tectonic stress and the spectra of seismic shear waves from earthquakes*, *J. Geophys. Res.* 75, 4997-5009.
- [10] Campbell, K. W. (1993). *Comparison of contemporary strong-motion attenuation relationships*, *Proc. International Workshop on Strong Motion Data II*, 49-70.
- [11] Clough, R.W. and J. Penzien (1975). *Dynamics of structures*, Mc Graw-Hill, New York, NY.
- [12] Commission of the European Communities (1993). Eurocode 8
- [13] Hammoutene M., Tilioune B., Bard 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.

- [14] Hammoutene M., *Simulation Nonstationnaire et Bidimensionnelle des Enregistrements Tremblements de Terre*, Thèse d'États, École Nationale Polytechnique, Alger, 1995.
- [15] Jacques Betbeder-Matibet., *Risques et aléas sismiques*, Edition Lavoisier, Paris, 2003
- [16] Joyner, W. B. and D. M. Boore (1981). *Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake*, *Bull. Seism. Soc. Am.* 71, 2011-2038.
- [17] Liu, K.S., Y.B. Tsai, *attenuation relationships of peak ground motion acceleration and velocity for crustal earthquake in Taiwan*, *BSSA*, VOL. 95, 2005, 1045-1058.
- [18] 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, Éditions TECHIP, Paris, 2001.*
- [19] Mark, W.D., *Spectral Analysis of the Convolution and Filtering of Nonstationary Stochastic Processes*, *Journal of Sound Vibration*, Vol. 11, N° 1, 1970, 19-63.7
- [20] M. Massa, P. Morasca, L. Moratto, S. Marzorati, G. Costa, and D. Spallarossa *Empirical Ground-Motion Prediction Equations for Northern Italy Using Weak- and Strong-Motion Amplitudes, Frequency Content, and Duration Parameters Bulletin of the Seismological Society of America*, June 1, 2008; 98(3): 1319 - 1342.
- [21] 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.
- [22] Novikova, E.I., M.D. Trifunac (1994). *Duration of strong ground motion in terms of earthquake magnitude, epicentral distance, site conditions and site geometry*. *Earthquake Eng. & Struct. Dynam.*, 23, 1023-1043.
- [23] Paultre, P. (2005). *Dynamique des structures application aux ouvrages de génie civil*. Lavoisier, Paris.
- [24] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., (1986). *Numerical Recipes : The Art of scientific computing*. Cambridge University Press.
- [25] Po-shen lin., chyi-tyi lee., *Ground motion attenuation relationships for subduction zone earthquakes in northeastern Taiwan*, *BSSA*, Vol. 98, 2008, 220-240.
- [26] Pousse G., Bonilla F., Cotton F., Rey J., *Génération stochastique d'accélérogrammes synthétiques, validation de méthodologie*, AFPS.
- [27] Preumont A., *Vibrations Aléatoires et Analyse Spectrale*, Presses Polytechniques et Universitaires Romandes, Lausanne, 1990.
- [28] Priestly, M. B., *Evolutionary Spectra and Nonstationary Processes*, *J. Royal Statist. Soc., Ser. B.*, N° 2, 1965, 204-237.

- [29] Règles Parasismiques Algériennes RPA99, (1999 version 2003). Centre National de Recherches Appliquées en Génie Parasismique, D.T.R.-B.C.-2.48, Alger.
- [30] Sabetta F., Pugliese A., *Estimation of response spectra and simulation of nonstationary earthquake ground motions*, BSSA, Vol. 86, 1996, 337-352.
- [31] Sabetta F., Pugliese A., *simulation of non stationary time histories scaled for magnitude, distance and soil conditions*, Proc.10th European conference on earthquake engineering, Vienna, Auustria, **1**, 247-252
- [32] Sadigh, K. R. (1993). *A review of attenuation relationships for rock site conditions from shallow crustal earthquakes in an interplate environment*, Proc. International Workshop on Strong Motion Data, II, 179-236.
- [33] Seed, H.B., C. Ugas and J. Lysmer (1976). *Site-dependent spectra for earthquake resistant design*. Bull. Seism. Soc. Amer., **66**, 221–234.
- [34] Trifunac, M.D. and A.G. Brady (1975). *A study on the duration of strong earthquake ground motion*. Bull. Seism. Soc. Amer., **65(3)**, 581–626.
- [35] Trifunac, M.D. (1978). *Response spectra for earthquake ground motion*. J. Engng. Mech. Div., ASCE, (**EM5**), 1081-1097.
- [36] Vanmarcke, E.H., S.P. Lai (1980). *Strong-motion duration and RMS amplitude of earthquake records*. Bull. Seism. Soc. Amer., **70**, 1293-1307.