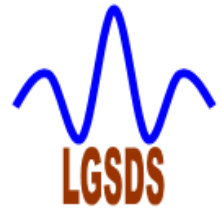


REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de L'enseignement Supérieur et de La Recherche Scientifique



Ecole Nationale Polytechnique
Ecole Doctorale Ingénierie et Environnement
Option Construction et Environnement
Laboratoire de Génie Sismique & Dynamique des Structures



Mémoire de Magister

Présenté par

AMMOUR Mahfoud

Ingénieur d'Etat en Génie Civil

Diplômé de l'Ecole Nationale Polytechnique

Thème :

**DEVELOPPEMENT D'UN MODELE STATISTIQUE
POUR LA DUREE DE PHASE FORTE
DES MOUVEMENTS SISMIQUES : CAS DES SOLS ROCHEUX**

Soutenu publiquement le 26 Janvier 2012 devant le jury composé de :

Président du jury :	M. HAMMOUTENE	Professeur, E. N. P.
Directeur du mémoire :	B. TILIOUINE	Professeur, E. N. P.
Examineurs :	M. BOUABDALLAH	Professeur, E. N. P.
	R. ILLOUL	M. Conférences A, E. N. P.

Remerciements

Je remercie tout d'abord ALLAH le tout puissant de m'avoir donné le courage de surmonter les moments les plus difficiles de ma vie et la volonté de mener à terme ce travail.

Monsieur B. TILIOUINE, professeur à l'Ecole Nationale Polytechnique, m'a fait le très grand honneur de diriger ma thèse. Son enseignement, Sa rigueur et son insatiable curiosité scientifique restent un exemple pour moi. Ses conseils éclairés et très pertinentes, la démarche scientifique qu'il m'a inspirée ainsi que ses encouragements m'ont été précieux. Je tiens à lui exprimer ici mes très sincères remerciements.

J'ai trouvé au laboratoire LGSDS, unité mixte de recherche du laboratoire du Génie civil, un environnement scientifique et un cadre de travail de qualité. Je suis très largement redevable à Monsieur Hammoutene. Professeur à l'Ecole Nationale Polytechnique. Les grandes orientations de ma recherche sont le fruit de ses conseils et des nombreuses discussions que nous avons eues. Ses encouragements sans cesse renouvelés ont été déterminants pour l'aboutissement de mes efforts. Qu'il trouve ici l'expression de ma profonde gratitude.

Nombre de mes collègues ont contribué à rendre cette période fructueuse et plaisante. Je les en remercie chaleureusement et je leur garde toute mon amitié.

Ma pensée vas également à tous les enseignants qui ont contribués 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.

Je remercie enfin et surtout ma famille et mes amis qui ont su faire preuve d'une pertinence et d'un soutien sans faille. Je leur voue toute mon affection.

DÉDICACE

A mes parents

A ma famille

A tous mes amis

ملخص

يهدف العمل الحالي إلى استحداث قانون تجريبي لدراسة تطور مدة الحركة الزلزالية العنيفة استنادا إلى عدة من المتغيرات المستقلة (طبيعة الوسط، القوة الزلزالية، البعد إلى مستوى التصدع الزلزالي) العلاقات التوقعية المتحصل عليها في هذا العمل، بالاعتماد على طريقة التحليل التراجعي (الخطى و غير الخطى) لبنك من البيانات العالمية تحوى تسجيلات تسارع لهزات ارضية عنيفة .
تعتبر مدة الحركة الزلزالية العنيفة على أنها مؤشر مهم في دراسة الزلازل و النتائج المتحصل عليها مثيرة جدا للاهتمام بشكل خاص سواء بالنسبة لمحاكاة تسجيلات تسارع اصطناعية ا و الى استجابة منشآت في بيئة عشوائية.
الكلمات المفتاحية:- مدة الحركة الزلزالية العنيفة، العلاقات التوقعية، التحليل التراجعي، و لبنك من البيانات العالمية، محاكاة تسجيلات تسارع.

Résumé

Le présent travail s'inscrit dans le cadre de l'élaboration d'une loi empirique nouvelle de l'évolution de la durée de phase forte des mouvements sismiques en fonction de la magnitude, la distance et les conditions de site. Les équations prédictives utiliseront un modèle de régression statistique linéaire et non-linéaire et seront basées sur une banque de données accélérométriques appropriée.

La durée de phase forte est un indicateur important du mouvement sismique et les résultats obtenus seront particulièrement intéressants, aussi bien pour la simulation des accélérogrammes synthétiques que pour la détermination de la réponse des structures dans un contexte aléatoire.

Mots clés : Durée de phase forte, équations prédictives, régression, banque de données, simulation des accélérogrammes.

Abstract

The present work is a part of the development of an empirical law for the evolution of the duration of strong earthquake Ground motion based on magnitude, distance and site conditions. The Predictive equations use either linear or nonlinear regression model and will be based on a suitable accelerometric data set.

The duration of strong earthquake ground motion is an important indicator of earthquake ground motion and the obtained results are particularly interesting, both for the simulation of synthetic accelerograms and for determining the response of structures in a random environment.

Keywords: Strong motion duration, Predictive equations, regression analysis, data set, simulation of synthetic accelerograms.

SOMMAIRE

Liste des figures	i
Liste des tableaux	iii
INTRODUCTION GENERAL	1
1. INTRODUCTION	2
2. MOTIVATIONS DU PROJET DE RECHERCHE	3
2.1. Synthèse bibliographique	3
2.2. Objectif de la recherche	6
3. ORGANISATION DU MEMOIRE.....	7
CHAPITRE 1- GENERALITES SUR LES SEISMES	9
1.1. INTRODUCTION	10
1.2. PROCESSUS DU TREMBLEMENT DE TERRE	11
1.3. LES DIFFERENTS TYPES D'ONDES SISMIQUES	12
1.4. QUANTIFICATION DES SEISMES A LA SOURCE ET AU LIEU D'ENREGISTREMENT.....	13
1.4.1. Magnitude et grandeurs associées	13
1.4.2. L'intensité macrosismique.....	15
1.5. MOUVEMENT A LA SURFACE DU SOL.....	16
1.6. PARAMETRES DU MOUVEMENT DU SOL.....	17
1.6.1. Indicateurs de nocivité du mouvement du sol.....	17
a) Durée de phase forte	17
b) Nombre de passages par zéro (NPZ).....	17
c) Intensité d'Arias (I_a).....	18
d) Accélération, vitesse, déplacement maximum du sol.....	19
e) Rapport Accélération maximale/Vitesse maximale (PGA/PGV).....	19
1.6.2. Spectre de réponse	20
1.7. FACTEURS AFFECTANT LE MOUVEMENT DU SOL	21
1.9.1. Facteurs liés au chemin traverse	21
1.9.2. Facteurs liés à la source	23
1.9.3. Facteurs liés aux conditions locales de site	24

CHAPITRE 2 -	LA SISMOLOGIE DES MOUVEMENTS FORTS :	
	PRINCIPES ET DONNEES	26
2.1.	INTRODUCTION	27
2.2.	ESTIMATION DES MOUVEMENTS FORTS.....	27
2.3.	DESCRIPTION EMPIRIQUE DES MOUVEMENTS FORTS.....	28
2.3.1.	Les paramètres de description du mouvement.....	28
2.3.2.	Les équations prédictives.....	29
2.3.3.	Différents types de relations prédictives.....	31
CHAPITRE 3 -	BANQUE DE DONNEES	33
3.1.	INTRODUCTION.....	34
3 .2.	SELECTION DES ENREGISTREMENTS SISMIQUES.....	34
3 .2.1.	Présentation des différentes classes : répartition des données.....	35
3 .2.2.	Présentation des enregistrements pour les sols rocheux.....	36
CHAPITRE 4 –	LA DUREE DE PHASE FORTE	38
4.1.	INTRODUCTION.....	39
4.2.	LA DUREE DU MOUVEMENT SISMIQUE.....	39
4.2.1.	Dépendance du RMS_a avec la durée.....	39
4.2.2.	Durée de phase forte du mouvement sismique.....	41
a)	La première définition.....	41
b)	La deuxième définition.....	42
c)	La troisième définition.....	43
4.3.	ANALYSE COMPARATIVE ENTRE LES DIFFERENTES MESURES DE LA DUREE.....	45
CHAPITRE 5 -	NOTIONS SUR L'ANALYSE DE REGRESSION	48
5.1	.INTRODUCTION.....	49
5.2.	ANALYSE DE REGRESSION.....	49
5.2.1.	Régression linéaire.....	50
a)	Modèle de régression linéaire multiple.....	50
b)	Conditions d'application du modèle linéaire multiple.....	51
c)	Détermination de l'équation de régression : Estimation des paramètres du modèle....	52
5.2.2.	Régression non linéaire.....	54
a)	Modèle de régression non linéaire multiple.....	54

b) Conditions d'application du modèle non linéaire multiple.....	54
c) Détermination de l'équation de régression : Estimation des paramètres du modèle....	55
5.2 .3 .Formulation d'un modèle de régression multiple : identification et interprétation des composantes du modèle.....	57
4 .2 .4.Analyse de variance en régression multiple	57
a) Décomposition de la variation dans les observations : calcul des sommes de carrés..	58
b) Calcul des degrés de liberté.....	58
c) Estimation de σ^2 calcul de s^2	59
d) Coefficient de détermination R^2	59
4 .2 .5.Test d'hypothèse et intervalles de confiance.....	60
e) Principe d'un test d'hypothèse.....	60
f) Concepts importants dans l'élaboration d'un test d'hypothèse.....	61
g) Démarche à suivre dans l'élaboration d'un test d'hypothèse	63
h) Test d'hypothèse en régression multiple.....	64
i) Intervalle de confiance β_j	68

CHAPITRE 6 - RESULTATS DES ANALYSES ET DISCUSSION.....69

6.1. INTRODUCTION.....	70
6.2. DESCRIPTION DES VARIABLES INDEPENDANTES.....	70
6.2.1. La Magnitude.....	70
6.2.2. La Distance source-site.....	70
6.2.3. Les conditions locales de site.....	71
6.3. DISCUSSION DES RESULTATS POUR LA REGRESSION LINEAIRE.....	71
6.3.1. Spécification du modèle.....	71
6.3.2. Discussion des résultats.....	73
a) Analyse de variance.....	70
b) Tests d'hypothèses.....	74
- Test de signification de la régression dans son ensemble (test de Fisher).....	74
- Test de la contribution marginale de chaque variable explicative (test de Student)..	75
6.3.3. Effets des différents paramètres environnementaux dans l'explication de la durée....	76
6.4. DISCUSSION DES RESULTATS POUR LA REGRESSION NONLINEAIRE.....	78
6.4.1. Spécification du modèle	78
6.4.2. Discussion des résultats.....	80
a) Analyse de variance.....	80

b) Tests d'hypothèses.....	81
- Test de signification de la régression dans son ensemble (test de Fisher).....	81
- Test de la contribution marginale de chaque variable explicative (test de Sudent)..	82
6.4.3. Effets des différents paramètres environnementaux dans l'explication de la durée....	83
CONCLUSION GENERALE	84
Références bibliographiques	88
Annexe I : Tests d'hypothèses en régression multiple.....	92
Annexe II : Différents codes de classification du sol	93
Annexe III : Durée de phase forte calculée selon les trois définitions	94

Liste des figures

Fig.1.1 : Carte du monde montrant les principales plaques tectoniques.

Fig.1.2 : Vocabulaire et éléments caractéristiques en géosismiques.

Fig.1.3 : Les ondes sismiques de volume et de surface.

Fig.1.4 : Comparaison des différents types de magnitudes utilisées en sismologie.

Fig.1.5 : Des enregistrements d'accélération de quelques événements sismiques.

Fig.1.6 : Différents mesures de distances utilisées en géosismiques.

Fig.1.7 : Schématisation d'un plan de faille (Brune 1976).

Fig.1.8: Spectres moyens d'accélération pour différentes conditions de site selon Seed et al. (1976).

Fig.2.1 : Evolution de l'accélération, de la vitesse et du déplacement du sol au cours du temps.

Fig.3.1 : Représentation des données pour les sols rocheux.

Fig.4.1 : Fonction RMS_a cumulative (séisme d'El-Asnam, réplique du 30101980, comp-NS).

Fig.4.2 : Durée de phase forte selon la définition de McCann et Shah (1979).

Fig.4.3 : Durée de phase forte selon la définition de Trifunac-Brady(1975).

Fig.4.4 : Durée de phase forte selon la définition de Bolt (1973.)

Fig.4.5 : Analyse graphique entre les différentes mesures de la durée.

Fig.6.1 : Effet de la distance, R_{rup} sur la durée de phase forte (cas du modèle linéaire).

Fig.6.2 : Effet de la magnitude, M_w sur la durée de phase forte (cas du modèle linéaire).

Fig.6.3 : Effet de la distance, R_{rup} sur la durée de phase forte (cas du modèle non linéaire).

Fig.6.4 : Effet de la magnitude, M_w sur la durée de phase forte (cas du modèle non linéaire).

Liste des tableaux :

Tab.3.1 : Classification du site selon (UBC).

Tab.3.2: Classes des distances.

Tab.3.3 : Répartition de la banque de données pour différentes classes de sol.

Tab.4.1 : Coefficients de corrélation pour la durée.

Tab.5.1 : Degré de liberté pour chaque somme de carrée.

Tab.5.2 : Analyse de variance en régression multiple.

Tab.6.1 : Analyse de variance en régression linéaire.

Tab.6.2 : Statistiques des estimateurs en régression linéaire.

Tab.6.3 : Analyse de variance en régression non linéaire.

Tab.6.4 : Statistique des estimateurs en régression non linéaire.

INTRODUCTION GENERALE

1. INTRODUCTION

Jusqu'à un passé récent, l'appréciation d'un séisme était liée à l'évaluation de son caractère destructeur. Le séisme n'est dangereux pour les vies humaines qu'à travers les dommages qu'il inflige aux ouvrages et aux pertes matérielles et économiques qu'il entraîne. L'aspect subjectif de son caractère destructeur exclut pratiquement l'obtention des renseignements quantitatifs.

Les dégâts causés par les tremblements de terre incitent les scientifiques à comprendre les phénomènes mise en œuvre lors des séismes pour réduire leurs effets destructeurs. Ces effets sont de plus en plus meurtriers du fait de l'urbanisation galopante de certaines villes, en particulier dans les pays en développement. Aujourd'hui, l'étude des tremblements de terre est devenue, en l'espace de quelques décennies, une véritable science ; elle regroupe plusieurs disciplines, physique, mathématiques, mécanique, géologie etc...). Cette science a fait un progrès significatif dans la compréhension de la physique de base des tremblements de terre dans le domaine particulier de l'étude de la source sismique, de plus en plus d'observations précises et abondantes sont disponibles pour étudier le phénomène à l'origine des ondes sismiques.

Aussi, sous l'impulsion des ingénieurs, s'est développée la mesure et l'interprétation physique des mouvements au voisinage d'une faille, ce qui l'on appelle la sismologie des mouvements forts .Ces mouvement produits par des tremblements de terre qui sont irréguliers et de nature aléatoire, les paramètres qui les caractérisent sont très variables ce qui rend leur description difficile. Des études expérimentales et analytiques des mouvements forts du sol ont pu être développées ces dernières années grâce à une densification plus importante des réseaux de sismographes et d'accélérographes installés dans les différentes régions sismiques du monde. Plusieurs auteurs (Abrahamson et Silva 1997, Boore et al 2003, Bommer et Martinez-Periera...) 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 des données enregistrées durant les séismes passés.

2. MOTIVATIONS DU PROJET DE RECHERCHE

2.1. Synthèse bibliographique

Le mouvement sismique fort dû au tremblement de terre est le résultat d'une dislocation complexe sur une faille sismique et la propagation des ondes sismiques à travers les couches du sol non homogènes. Les enregistrements dans le temps des mouvements forts ont été considérés dans le passé comme une simple réalisation d'échantillon d'un processus aléatoire, (Housner et Jennings, 1964). Cependant, comme chaque tremblement de terre est étudié, plus on en apprend sur le mouvement fort et le degré d'incertitude est lentement réduite.

L'état de la pratique dans la modélisation pour la prédiction du mouvement fort s'est développé le long plutôt simplifié des lignes. Cela est dû en partie de la nature complexe du processus et de notre niveau de compréhension. La modélisation du mouvement fort du sol consiste en trois parties, le paramètre décrivant la taille du tremblement de terre, le modèle du trajet source-récepteur des ondes émises, et les paramètres utilisés pour décrire l'intensité des secousses. Pour chacun de ces derniers une description simple a été utilisée. L'événement sismique est défini en fonction de la magnitude locale de Richter, et le trajet source-récepteur est exprimé par la distance épacentrale, hypocentrale, ou d'autres mesures significatives de distance. Il y a eu un certain nombre de paramètres employés pour caractériser l'intensité du tremblement de terre. Les plus généralement adoptés ont dépendu du fait que le chargement séismique doit être décrit d'une manière consistant avec la conception parasismique des structures. Pour cette raison l'accélération du mouvement est de la plupart d'intérêt aux ingénieurs et ainsi plusieurs paramètres employés pour décrire la secousse sismique sont définis en termes d'accélération du sol. La méthode la plus communément utilisée pour caractériser un mouvement sismique est par la valeur maximale absolue du mouvement pour un enregistrement donné dans le temps de l'accélération (*PGA*), de la vitesse (*PGV*) ou du déplacement (*PGD*). Ces paramètres ont été employés intensivement dans les études du mouvement fort, (Gutenberg and Richter, 1956; Blume, 1965 ; Esteva and Rosenblueth, 1964; Trifunac and Brady, 1975; Boore and al, 1978, ...). Des formes fonctionnelles standard ont été employées pour dériver des lois d'atténuation décrivant des valeurs extrêmes de mouvement sismique en fonction de la magnitude et la distance, (Idriss, 1978).

Il y a des avantages particuliers concernant les valeurs extrêmes, spécifiquement la facilité avec laquelle elles peuvent être définies et employées. Parmi ces différentes valeurs extrêmes, l'accélération maximale du sol (*PGA*) est le paramètre du tremblement de terre le plus généralement étudié. Cependant, il y a deux inconvénients de valeurs extrêmes comme une caractéristique numérique d'accélération maximale du sol. Premièrement, parce qu'elles présentent une large dispersion une fois tracés en fonction de la distance, et en second lieu sont souvent interrogées s'ils sont capables de représenter normalement l'intensité du tremblement de terre (Hanks et Johnson 1976). Il est facile de montrer par exemple que deux enregistrements accélérométriques dans temps dont la même accélération maximale peuvent avoir des durées, contenus en fréquences, et des niveaux d'énergies différents.

Il y a eu un intérêt considérable ces dernières années dans la recherche sur d'autres moyens plus améliorés pour la prédiction de l'intensité du mouvement fort lors d'un tremblement de terre. Un premier concerne était, les grandes incertitudes dans la prévision des valeurs extrêmes, et l'inadéquation de l'accélération maximale du sol (*PGA*) de corrélérer avec les dommages, sa sensibilité aux détails de la source sismique, la géologie du site au voisinage de station d'enregistrement et les petites variations de la topographie le long de la distance source-site. D'autres paramètres ont été employés pour caractériser le tremblement de terre, sont des valeurs extrêmes du spectre de réponse de l'accélération, de la vitesse, ou du déplacement, (McGuire, 1974), et l'amplitude du spectre de Fourier de l'accélération (Trifunac, 1976 ; McGuire, 1978). La tendance vers les paramètres spectraux s'est développée en se basant sur le concept de fournir un spectre de dimensionnement qui a un niveau de risque important à toutes les fréquences.

Un certain nombre d'investigateurs ont concentré leur attention sur la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a) en tant qu'une alternative qui a des caractéristiques théoriques attrayantes, (Housner, 1975 ; McCann et Shah, 1979 ; McGuire et Hanks, 1980). Sachant que (RMS_a) est fonction de la durée du mouvement fort (non pas un seul pic), ils ont suggéré que ce pourrait être un paramètre plus stable que (*PGA*). Le *RMS* accélération peut également être combiné avec la durée du mouvement fort pour représenter un niveau d'énergie beaucoup plus significatif qu'un seul pic d'accélération maximale. Afin de prédire ces paramètres d'intensité, les investigateurs ont proposé des différentes mesures de distance à la source, (Swang et al, 1981 et Bernreuter, 1981).

Ils ont suggéré que dans les champs proches la distance devrait être mesurée à partir de la zone dont le niveau d'énergie est le plus élevé, au lieu de la rapprocher en utilisant la distance la plus proche à la faille de rupture. En Concurrence avec le développement du RMS_a comme un paramètre du chargement, les chercheurs ont investigué la durée comme un autre paramètre possible qui serait davantage corrélé avec les dommages réels. Bien que les sismologues et les ingénieurs aient suspecté l'importance de la durée du tremblement terre pendant longtemps, la majeure partie du travail sur ce paramètre est relativement récente, son importance pour la liquéfaction des sols saturés est reconnue en premier, et la durée comme il a été mesurée par le « nombre de cycles équivalents » du mouvement, a été incorporé aux procédures d'évaluation du risque sismique (Seed and Idriss, 1967 et 1971; Seed et al, 1975). Les expériences d'un certain nombre de tremblements de terre de la première décennie ont confirmé que la durée de phase forte d'un séisme influence d'une manière significative le degré de dommages sur les structures et les sols. L'analyse des enregistrements d'accélération obtenus lors du tremblement de terre (Parkfield, 1966), a montré que des grandes accélérations et valeurs spectrales peuvent être associées à un peu de dommages sur les structures si la durée de la secousse est courte (Housner, 1975) : Les dommages causées au Lower San Fernando Dam lors du tremblement de terre de 1971 semblent d'être produits à la fin du mouvement sismique.

Un grand nombre de définitions de la durée de phase forte existent dans la littérature (Trifunac and Brady(1975), vanmarcke et Lai (1977), Bolt (1973), McCann et Shah (1979), perèz(1979)etc. La plupart d'entre eux tombent dans l'une des trois catégories suivantes :

- 1- La durée est une fonction de la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a).
- 2- La durée est définie comme le temps nécessaire pour qu'une certaine fraction de l'intensité d'arias soit atteinte (une définition basé sur la notion de cumul d'énergie du signal sismique).
- 3- La durée est une fonction du temps passée au-dessus d'un certain niveau d'accélération.

A l'heure actuelle, aucune définition de la durée n'a encore été universellement acceptée. Car en effet, plusieurs auteurs ont mis en relief que la signification d'une durée de mouvement fort est dépendante de l'usage qu'on veut en faire.

Enfin, plusieurs 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 entre la source –site ont été proposées, Trifunac et Novikova (1995) prennent en compte la somme de trois termes dans le calcul de la durée de 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.

2.2. Objectif de la recherche

Un tremblement de terre peut être caractérisé par de différents paramètres, chacun reflète une caractéristique particulière de la secousse sismique telle que les amplitudes maximales, le contenu en fréquences, ou l'énergie portée dans le signal sismique. Une caractérisation complète du tremblement de terre doit inclure une mesure de sa durée, ou plus spécifiquement la durée de la partie du signal sismique qui est considérée forte. La durée du mouvement sismique fort est l'un des paramètres principaux qui contrôlent les effets des dommages d'un tremblement de terre, elle joue un rôle principal dans beaucoup de problèmes de la séismologie. L'évaluation de la durée a une importance primordiale dans les problèmes liés aux phénomènes de fatigue due à un petit nombre de cycles, de la liquéfaction et du tassement.

Il n'y a aucun consensus clair par lequel, parmi des définitions multiples de la durée une est préférée, ce qui reflète probablement le fait que les différentes définitions peuvent être plus ou moins appropriées à des applications différentes. La vraie issue est que la durée elle-même n'explique d'une façon assez claire le tremblement de terre et son potentiel de causer des dommages, mais son influence nécessite d'être quantifiée en relation avec d'autres paramètres du tremblement de terre (magnitude, distance source-site et condition de site).

Le travail présenté dans ce mémoire s'inscrit dans l'optique de ce dernier paragraphe, dans le but d'établir une loi empirique pour l'estimation de la durée de phase forte des mouvements sismiques sur sols rocheux à travers une analyse de régression.

3. ORGANISATION DU MEMOIRE

Ce mémoire s'articule autour de la prédiction de la durée de phase forte du mouvement sismique sur sols rocheux, en deux volets principaux, chacun tentant d'apporter des éléments qui permettent de mieux comprendre la nature particulière du phénomène sismique.

- Le premier volet de ce travail est organisée en trois axes occupant chacun un chapitre.

Tout d'abord, nous avons commencé notre manuscrit par un chapitre introductif dans lequel une synthèse bibliographique donnant un aperçu sur l'historique du développement de la modélisation pour la prédiction du mouvement fort. Des différents paramètres pour caractériser l'intensité du tremblement de terre sont cités, sont aussi présentés les avantages et les inconvénients de chacun et des raisons pour lesquelles la durée est un paramètre très critique pour estimer le potentiel d'endommagement d'un séisme.

Nous avons évoqué au premier chapitre quelques éléments de la sismologie, tels que la naissance d'un séisme, la propagation des ondes sismiques, les paramètres caractérisant un séisme et les différents facteurs affectant le mouvement sismique, ceci dans le but de mieux comprendre le phénomène du tremblement de terre. Le deuxième chapitre présente des récentes avancées opérées par la sismologie des mouvements forts depuis ces quinze années, et donne les clés pour comprendre les aspects conceptuels de cette science et ses aspects méthodologiques. Les caractéristiques de la banque de données utilisées dans cette étude sont présentés au troisième chapitre.

Au quatrième chapitre sont exposées les trois définitions de la durée de phase forte du mouvement sismique, une analyse comparative est effectuée entre les différentes mesures de la durée. Les différentes définitions qui ont été considérées dans cette étude, sont celles proposées par : McCann et Shah(1979), Bolt (1973), Trifunac et Brady(1975).

- Le deuxième volet de ce travail décrit l'aspect statistique du traitement de la banque d'enregistrements dans le but d'établir une loi empirique reliant la durée aux autres paramètres environnementaux, et ce en deux chapitres :

Le chapitre 5 reprend les différentes méthodologies dédiées pour comprendre l'analyse de régression multiple par ses deux types linéaires et non linéaires.

Au chapitre 6 , nous avons exploité les résultats de la durée de phase forte calculés au chapitre 4 en utilisant la définition proposée par (McCann et Shah, 1979), laquelle est reliée à la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a), dans le but d'établir une loi empirique pour l'estimation de la durée de phase forte (D) du mouvement sismique à travers une analyse de régression décrite au chapitre 5. A cet effet, des formes fonctionnelles (linéaire ou non linéaire) reliant la durée aux différents paramètres environnementaux (moment magnitude, distance plus proche à la faille et conditions de site) ont été proposé. La méthode des moindres est utilisée pour l'estimation des valeurs des paramètres du modèle associés aux variables explicatives.

Les résultats de la régression sont discutés en termes d'analyse de variance et en termes de contribution d'ensemble et marginale des variables explicatives. Pour cela, des tests d'hypothèses sont menés sur chacun et sur l'ensemble des paramètres pour les deux modèles adoptées précédemment afin de s'assurer de leurs importances individuelles et globales vis-à-vis de la forme fonctionnelle adoptée (linéaire ou non linéaire) pour la variable expliquée.

Enfin, une étude paramétrique est effectuée, afin d'apprécier l'influence de chacun des paramètres environnementaux (moment magnitude, distance plus proche à la faille et condition de site), des discussions et des commentaires reflétant les résultats obtenus sont aussi présentés à la fin de ce chapitre.

Ce mémoire est clôturé par des conclusions et des recommandations qui pourront contribuer à des travaux et des recherches futures.

CHAPITRE I
GENERALITES SUR LES SEISMES

1.1. INTRODUCTION

Le globe terrestre est constitué de couches concentriques dont la plus externe est l'écorce qui se divise en écorce continentale et océanique, la croûte continentale possède une épaisseur d'environ 30 à 40 Km, allant jusque à 70 km sous certaines chaînes de montagne. La croûte océanique, quant à elle, est le plus réduit en épaisseur allant de 10 à 15 km, cette enveloppe externe est soumise à des mouvements perpétuels liés à la création de matières en certaines zones. Ces mouvement affectent une quinzaine de plaques et engendrent des contraintes sur les lignes de contact des plaques. Lorsque ces contraintes deviennent trop élevées et supérieures à la résistance au frottement, des ruptures brutales se produisent : c'est un tremblement de terre, à la suit duquel un nouvel état provisoire de stabilité est réalisé.

Ces ruptures se produisent essentiellement au niveau des zones frontalières des plaques tectoniques ou zones de jonction inter - plaques à proximité des limites entre les plaques, là où les tensions sont les plus élevées dans les roches (Fig.1.1) [2].

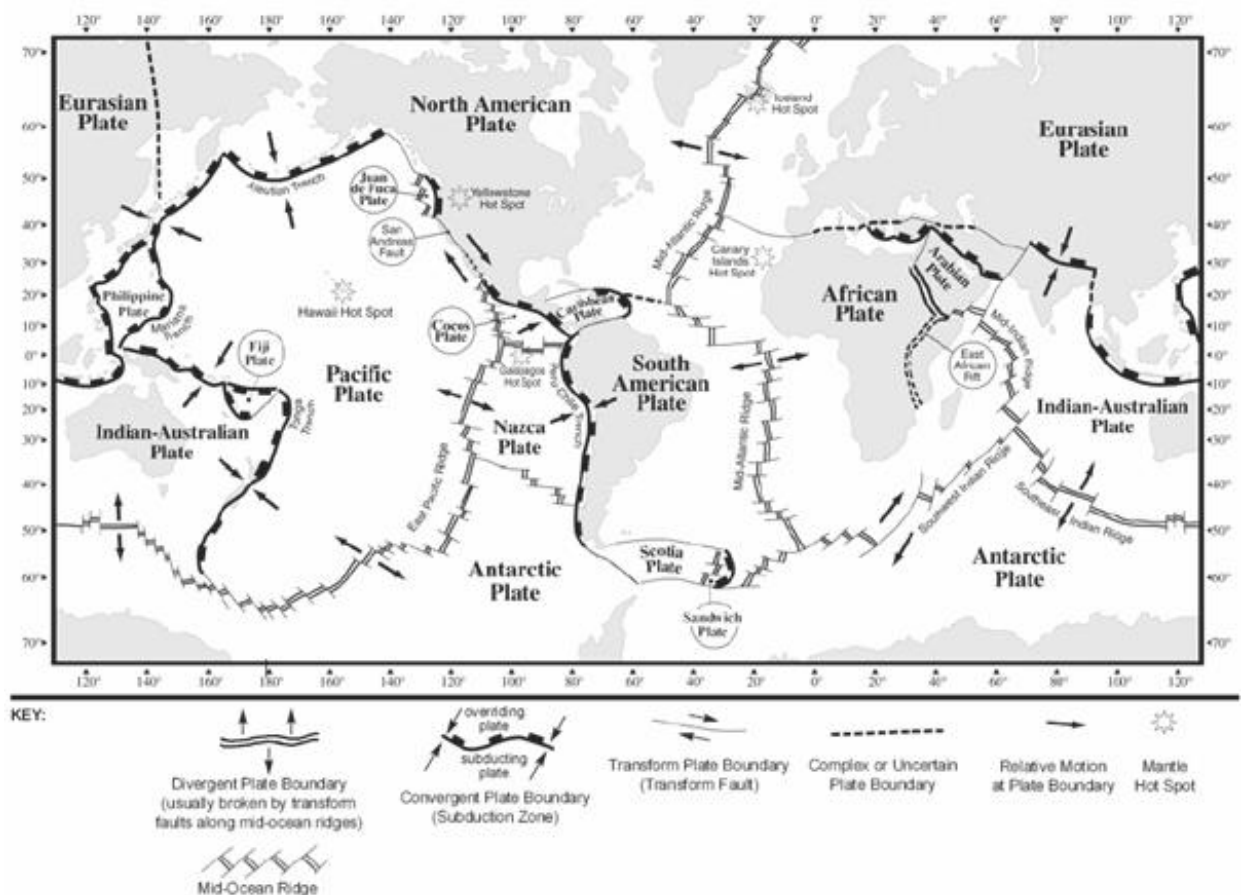


Fig. 1.1 : Carte du monde montrant les principales plaques tectoniques [2].

Cependant il faut savoir aussi, qu'il peut y avoir des phénomènes sismiques dans des failles à l'intérieur de ces plaques engendrant des mouvements de terre, on parlera 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 types de mouvements relatifs : 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.

1.2.PROCESSUS DU TREMBLEMENT DE TERRE

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

A l'origine de tels phénomènes se trouvent les déformations lentes de la croûte terrestre ou du milieu sous-jacent, qui entraînent l'accumulation de tension interne. Ces dernières finissent par dépasser, en un point donné, la résistance des roches au cisaillement, produisant à l'endroit en question une rupture qui se propage ensuite. La faille résultante constitue désormais une zone de faiblesse dans la croûte. La création d'une faille ou, plus fréquemment, le glissement le long d'une faille existante, constitue le mécanisme générateur d'un séisme. L'endroit où se produit le séisme est appelé foyer ou hypocentre et la projection de ce point sur le sol est appelée épicentre. L'hypocentre peut se situer à une profondeur très variable: de quelques kilomètres à plus de 100 km(Fig.1.2).

A partir du foyer, la secousse sismique se propage sous forme d'ondes de divers types qui compte tenu de l'hétérogénéité du sol, engendrent en surface un mouvement complexe difficilement prévisible en un site donné [15].

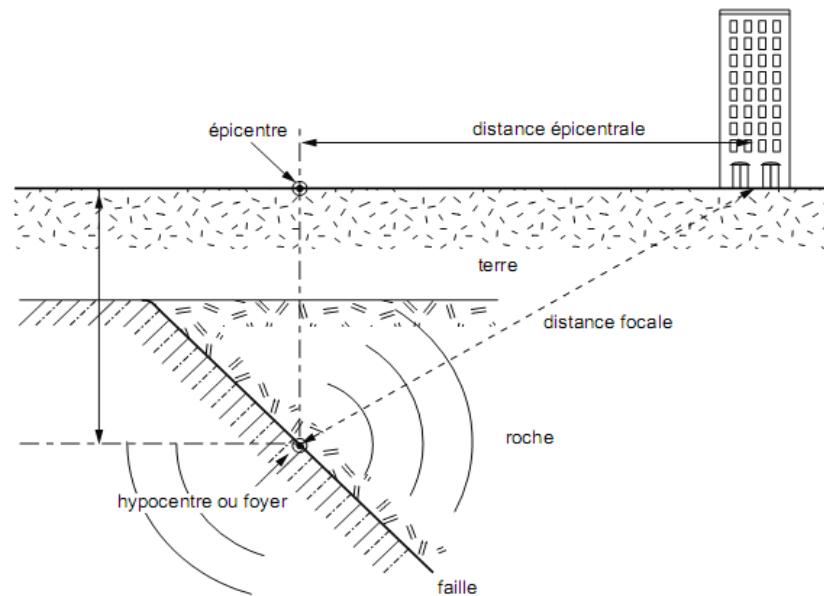


Fig.1.2 : Vocabulaire et éléments caractéristiques en géosismiques [15].

1.3. LES DIFFERENTS TYPES D'ONDES SISMQUES [15]

On distingue deux types d'ondes :

1.3.1. 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, et 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.1.3).

1.3.2. Les ondes de surface : elles sont générées par les ondes de volume qui arrivent à la surface et se propagent à une 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.1.3).

La différence de vitesse de propagation des ondes longitudinales et transversales nous permet, à l'aide de l'enregistrement de plusieurs sismographes, de situer le foyer d'un séisme.

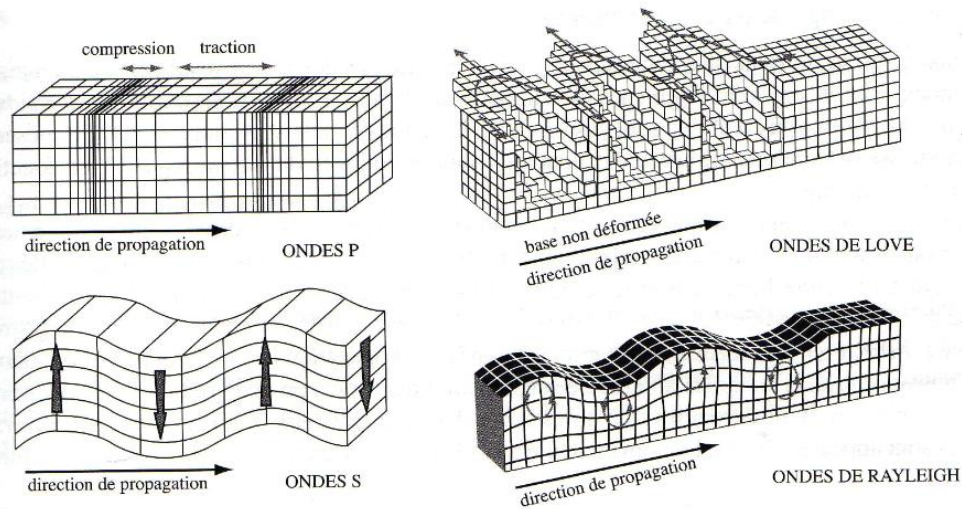


Fig.1.3 : les ondes sismiques de volume et de surface [2].

1.4. QUANTIFICATION DES SEISMES A LA SOURCE ET AU LIEU D'ENREGISTREMENT

1.4.1. Magnitude et grandeurs associées

La magnitude représente une mesure de l'énergie libérée par la rupture de la faille dans la zone source. La notion de magnitude a été introduite par C.F.Richter en 1935 en Californie pour estimer l'énergie, libérée au foyer de séismes locaux et pouvoir les quantifier. La magnitude M est une quantité logarithmique calibrée, calculée à partir de l'amplitude du signal enregistré par un sismographe, donnée par la formule :

$$M = \log \left(\frac{A}{T} \right) + f(R) + C \quad (\text{où } \log \text{ est le logarithme en bas } 10)$$

Où :

A est l'amplitude maximale du signal, T la période correspondante à A et R représente la distance à la source, C est un terme de calibration.

Les séismes sont caractériser par plusieurs types de magnitudes : des magnitudes locales M_L qui sont calculées à partir de l'amplitude des ondes S et d'un modèle d'atténuation empirique. Traditionnellement, la courbe obtenue pour la Californie par Richter (1958) lorsqu'il a défini ce type de magnitude, est utilisée moyennant des adaptations au contexte local. Globalement cette échelle est sur l'amplitude maximale observée sur les sismogrammes à courte période.

La magnitude de surface M_s est définie par rapport à l'amplitude des ondes de surface à une période de 20s. Un autre type de magnitude couramment utilisée est la magnitude des ondes de volumes m_b (ou m_B selon le type d'appareil à partir duquel la magnitude est calculée), définie pour des périodes de 1 à 10s. Toutes ces magnitudes représentent des paramètres physiques différents pour le séisme du fait de la différence des périodes observées et des distances épacentrales concernées (Kanamori, 1978). De plus ces échelles saturent pour les forts séismes, ce qui a conduit à définir une magnitude dite de moment M_w déduit du moment sismique M_0 (Fig.1.4).

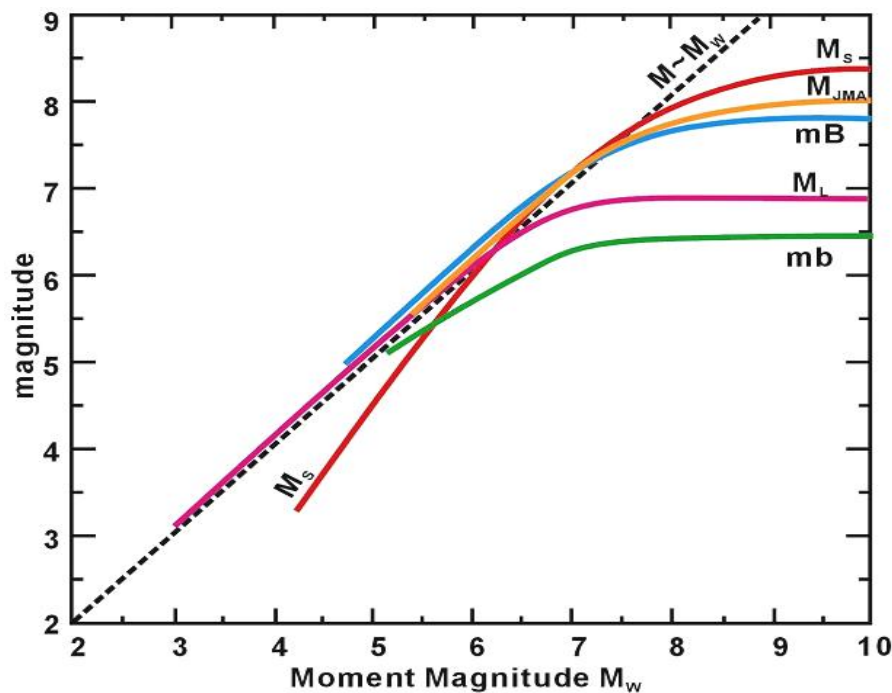


Fig1.4 : Comparaison des différents types de magnitudes utilisées en sismologie.

Pour résumer, la figure (1.4) compare les différents types de magnitudes utilisées en sismologie : M_L , M_s , m_b , M_{JMA} en fonction de la magnitude du moment M_w . La Magnitude M_{JMA} est équivalente à une magnitude de surface, mais issue d'un institut Japonais. En fin, les modèles empiriques de prédiction du mouvement du sol utilisent une magnitude pour décrire les données, dont les deux types les plus courants sont la magnitude de surface M_s et la magnitude du moment M_w qui sont généralement définis pour les séismes forts. Ceci implique donc que des conversions seront nécessaires entre ces types de magnitudes.

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 :

$$\log E_r = 11,4 + 1,5M$$

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

$$\log M_0 = 19,9 + M_L$$

1.4.2. Intensité macrosismique

La magnitude n'est pas suffisante pour définir l'importance d'un séisme sur un milieu donné. En effet, en fonction de la distance à la source et de certaines conditions locales, les effets d'un séisme peuvent être plus importants alors même que la magnitude est plus faible. L'agression du mouvement du sol sur une structure est quantifiée en fonction d'une intensité sismique, ou intensité macrosismique. Macrosismique car elle est généralement utilisée afin de regrouper les effets du séisme en zones d'intensité homogène, englobant une région et définissant des contours d'égale intensité, ou isoséistes. En général, l'intensité diminue avec la distance mais des conditions particulières telles que des effets de site ou des effets de directivité liés à la position et à l'orientation de la source peuvent dans certains cas créer des anomalies dans les cartes macrosismiques.

Plusieurs échelles d'intensité existent, les plus anciennes étaient essentiellement basées sur l'observation des effets du mouvement du sol sur l'homme et son environnement direct (vibrations des objets, ressenti en fonction de sa position et de son activité) comme celle développée par Rosi-Forel 1880, cette échelle contient dix degrés (1 à 10) ; elle a été employée pour étudier l'intensité du tremblement de terre de San Francisco 1906. Une échelle plus raffinée a été conçue en 1902 par le sismologue italien Mercalli avec une gamme de 12 degrés de I à XII, puis Mercalli Modifiée en 1956 (IMM). Plusieurs échelles ont été proposées depuis, en fonction des époques et des régions d'application, telles que : Medvedev-Sponheur-Karnik (MSK)etc.

Il est important de noter que toutes les échelles d'intensité ont permis de classer et d'évaluer l'importance des séismes historiques pour lesquels aucune donnée physique n'était disponible. En analysant les écrits historiques qui racontaient les phénomènes, il est possible de positionner grossièrement l'épicentre et d'évaluer l'importance de l'événement. L'activité d'observation des tremblements de terre est une discipline récente et le fait d'augmenter la période d'observation, même de façon empirique, contribue à améliorer la définition de l'aléa et participe ainsi à la réduction du risque sismique, en particulier dans des pays à sismicité moyenne.

Pour cela, il existe de nombreuses relations empiriques donnant l'équivalence en magnitude de l'intensité macrosismique ou épacentrale, c'est-à-dire sur la zone de l'épicentre, calées sur des séismes récents. D'autres tentatives existent pour relier le mouvement du sol et l'intensité sismique.

1.5. MOUVEMENT A LA SURFACE DU SOL

Lorsque les ondes et l'énergie qu'elles transportent atteignent un site à la surface de la terre, elles engendrent des vibrations dans trois directions deux horizontales et une verticale. Les mouvements du sol engendrés lors d'un séisme se mettent principalement sous la forme d'un *Accélérogramme* (Fig.1.5) qui représente la variation de l'accélération du sol dans une direction donnée en fonction du temps, et ce grâce à des appareils appelés accélérographes.

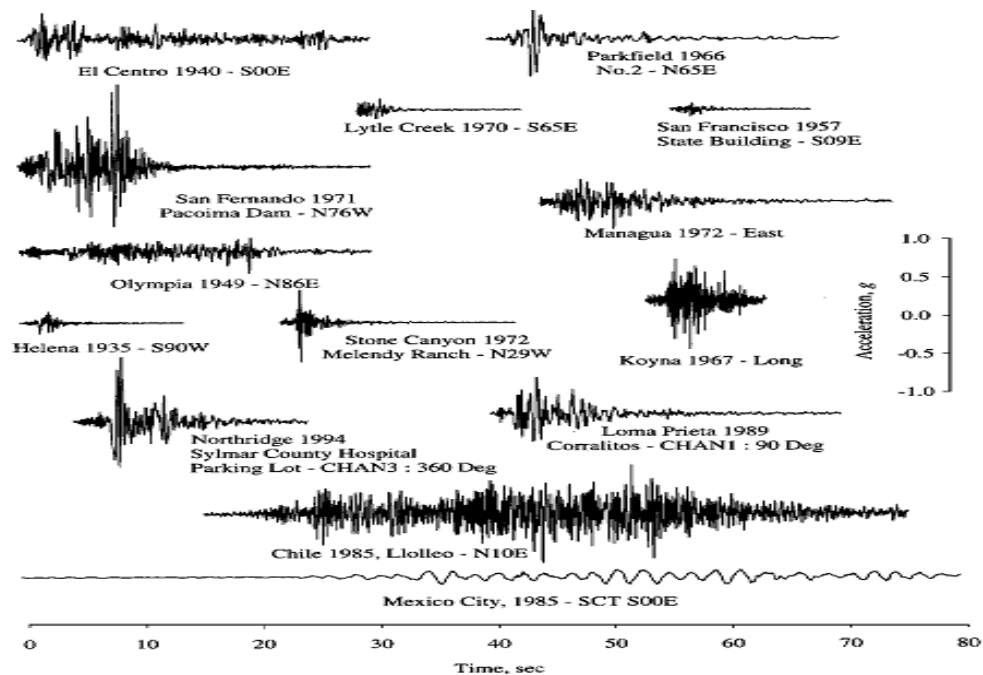


Fig.1.5 : Des enregistrements d'accélérations de quelques événements sismiques [21].

Par intégration d'un Accélérogramme, on peut obtenir les diagrammes de vitesse et de déplacement en fonction du temps. On lit sur l'accélérogramme l'accélération maximale du sol (*PGA*), paramètre important pour la définition du risque sismique, on peut également distinguer dans un accélérogramme une phase initiale pendant laquelle les accélérations sont faibles, une phase forte, une phase modérée et une phase très faible de retour progressif au repos. Les accélérogrammes constituent une donnée d'action directement utilisable en dynamique des structures [21].

1.6. PARAMETRES DU MOUVEMENT DU SOL

De la même façon que la magnitude et l'intensité d'un séisme ont été inventés de façon à ranger les séismes dans une échelle et à les classer les uns par rapport aux autres, l'importance du mouvement du sol est classée selon des paramètres. Plutôt que le séisme lui-même, ils représentent ses effets en un lieu donné, calculés à partir d'un enregistrement du mouvement du sol. Dans la littérature, un grand nombre de paramètres existent, certains empiriques d'autres plus physiques, parmi lesquels :

1.6.1. Indicateurs de nocivité du mouvement du sol

Rapidement après les premiers enregistrements du mouvement du sol, les ingénieurs-sismologues ont vite compris que des informations basiques pouvaient être extraites de ces enregistrements pour des applications en sismologie de l'ingénieur et en génie parasismique. En effet, dans les trois directions, c'est une information indispensable afin de dimensionner une structure résistante au séisme. Une réflexion a rapidement débuté afin d'en extraire des informations permettant de faire le lien direct avec l'endommagement des structures. De nombreux paramètres ont alors été proposés, afin de rechercher un indicateur de nocivité de la secousse. Les plus communément utilisés est :

a) Durée de phase forte

La durée du tremblement de terre est un paramètre significatif dans les processus de fissurations et dégradations progressives des éléments d'une construction. Ceci est particulièrement vrai dans le cas des structures ductiles conçues pour supporter des déformations inélastiques lorsqu'elles sont soumises à des forts mouvements du sol. La réponse inélastique de telle structure est sensible au nombre de cycles du mouvement fort qu'elles subissent. La durée est un facteur de destruction important en zone épiscopales, la durée est au maximum de l'ordre de 60 sec en zone très sismique, mais n'est quelques secondes en zones peu sismique.

b) Nombre de passages par zéro (NPZ)

Ce paramètre correspond au nombre de fois que l'historique de l'accélération du sol traverse l'axe des abscisses, soit 2 fois le nombre de cycles que contient l'accélérogramme.

Notons que le nombre de passages par zéro est calculé sur la durée de phase forte.

Pour évaluer l'effet du nombre de passage par zéro, il faut évaluer l'énergie et l'importance des pulsations du signal qui sont deux caractéristiques davantage révélatrices. Ceci est fait en étudiant des caractéristiques comme l'intensité.

Intensité d'Arias (I_a)

L'importance du mouvement du sol peut être décrite par plusieurs grandeurs physiques représentant l'énergie ou l'intensité du signal. La plus courante est appelée l'intensité d'Arias I_a . Arias (1970) a défini l'intensité du tremblement de terre comme la somme de l'énergie dissipée (par unité de masse) pour tous les éléments de la structure, elle est représentée par :

$$I_a = \int_0^{\infty} E d\omega$$

Où :

E , est l'énergie dissipée par unité de masse pour toute la structure, à cause du mouvement induit par le tremblement de terre.

ω , représente la fréquence de vibration de la structure. En utilisant le théorème de conservation de l'énergie de Rayleigh dans le domaine des temps et dans le domaine des fréquences et l'intensité d'Arias est définie comme la somme du carré de l'accélération, calculée sur la durée du signal, soit l'expression:

$$I_a = \frac{\pi}{2g} \int_0^{T_d} a^2(t) dt$$

Où :

I_a , est l'intensité à un coefficient d'amortissement égal zéros.

$a(t)$, est l'accélération du sol à l'instant t .

T_d , représente la durée de toute l'enregistrement d'accélération.

L'intensité d'Arias a une unité de vitesse. C'est un paramètre qui rend compte du pouvoir destructeur d'un séisme. Plusieurs corrélations existent entre la valeur de I_a et la résistance des structures, mais aussi la liquéfaction et les déclenchements de glissements de terrain. Par conséquent, des modèles empiriques d'atténuation de l' I_a existent dans la littérature qui prédit la variation de ce paramètre en fonction du moment magnitude et la distance.

c) Accélération, vitesse, déplacement maximum du sol

L'accélération maximale du sol (PGA) est la valeur la plus facile à extraire de l'accélérogramme. Elle correspond au maximum de la valeur absolue de l'accélération. Du fait de sa simplicité, elle a été très utilisée pour caractériser l'importance de mouvement du sol. Cependant, elle ne donne qu'une description partielle du mouvement, en considérant le mouvement à période nulle. La vitesse et le déplacement maximum du sol (soit PGV et PGD) s'obtient en intégrant le signal accélérométrique. Ces deux paramètres sont d'autant plus importants que les vitesses ont un pouvoir destructeurs importants sur les structures, en particulier parce que la vitesse caractérise les fréquences intermédiaires, celles qui concernent les structures. Le PGA peut s'exprimer en m/s^2 mais, cependant, on emploie bien souvent l'unité g pour exprimer l'accélération en pourcentage de l'accélération de la gravité.

Malgré ce paramètre a été largement utilisée comme une mesure d'intensité du tremblement de terre, des investigations récentes ont douté sa validité. Premièrement, des enregistrements d'accélération pour des évènements de même magnitude situés à des distances similaires ont présenté une large dispersion des valeurs de *PGA* une fois tracés en fonction de la distance et en second lieu sont souvent interrogées s'ils sont capables de représenter normalement l'intensité du tremblement de terre (Hanks et Jahnson 1976). Il est facile de montrer par exemple que deux enregistrements accélérométriques dont la même accélération maximale peuvent avoir des durées, contenu en fréquence, et des niveaux d'énergies différents. Cependant, les valeurs de *PGA* restent très utilisées en particulier pour définir des modèles de prédictions empiriques du mouvement du sol, qui interviennent en particulier dans l'élaboration des cartes d'aléa probabiliste. Un des enjeux actuels de la recherche tente de réduire l'incertitude des modèles prédictifs, en essayant d'identifier l'origine de la variabilité observée du mouvement sismique.

d) Rapport Accélération maximale/Vitesse maximale (PGA/PGV)

Le rapport de l'accélération maximale sur la vitesse maximale est souvent utilisé pour caractériser les séismes. De façon générale, un rapport *PGA/PGV* élevé indique qu'il y a beaucoup d'énergie libéré sur de courtes périodes de temps, ce qui caractérise habituellement les séismes à hautes fréquences. À l'opposé, un rapport *PGA/PGV* faible, près de l'unité, caractérise des séismes à basses fréquences.

1.6.2. Le spectre de réponse

Il n'y a pas de paramètres satisfaisant pour l'ingénieur permettant d'intégrer l'ensemble des caractéristiques du mouvement du sol qui vont influencer la réponse de la structure. C'est pour cela qu'a été introduite la notion de spectre de réponse qui permet de caractériser le mouvement du sol en fonction de la réponse d'un oscillateur à 1 degré de liberté (soit 1-DDL). Il est convenu alors d'utiliser un spectre de réponse pour définir le mouvement sismique pour l'ingénieur. En effet, il donne une information de l'accélération maximale que la structure devra supporter et il caractérise l'énergie sismique maximale. Grâce à l'utilisation de l'intégrale de Duhamel, on peut exprimer la réponse en accélération, vitesse et déplacement, d'un oscillateur à 1-DDL, de caractéristiques ω et ξ par les équations suivantes:

$$S_D = |u(t)|_{max} = \left| \frac{1}{\omega_D} \left(\int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin\omega_D(t-\tau) d\tau \right) \right|_{max}$$

$$S_V = |\dot{u}(t)|_{max} = \left| - \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \cos\omega_D(t-\tau) d\tau + \frac{\xi}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin\omega_D(t-\tau) d\tau \right|_{max}$$

$$S_A = |\ddot{u}(t) + \ddot{u}_g(t)|_{max} = \left| 2\xi\omega \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \cos\omega_D(t-\tau) d\tau + \frac{\omega(1-2\xi^2)}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin\omega_D(t-\tau) d\tau \right|_{max}$$

On peut montrer que pour des petites valeurs d'amortissement ($\xi < 0.20$), telles que celles rencontrées dans les structures de génie civil, on a: $\xi \sim \xi^2 \sim 0$ et donc $\omega \sim \omega_D$. On peut donc simplifier les équations précédentes en exprimant les relations existantes entre le déplacement, vitesse et l'accélération.

$$\dot{u}(t) \simeq - \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin\omega_D(t-\tau) d\tau = \omega u(t)$$

$$\ddot{u}(t) + \ddot{u}_g(t) \simeq \left| \omega \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin\omega_D(t-\tau) d\tau \right|_{max} = \omega^2 u(t)$$

Il est utile et pratique de définir les pseudo-spectres de réponse en vitesse et en déplacement par les relations suivantes:

$$S_{PD}(\xi, \omega) = S_A(\xi, \omega) / \omega^2$$

$$S_{PV}(\xi, \omega) = S_A(\xi, \omega) / \omega$$

Comme indiqué précédemment, cette caractérisation ne dépend pas que du mouvement du sol mais aussi de la fréquence de vibration de l'oscillateur 1-DDL et de son d'amortissement.

Ainsi, pour un Accélérogramme donné, et en supposant des valeurs discrètes de ω et ξ , il est possible de calculer les valeurs discrètes de S_A et de calculer les valeurs correspondantes des S_{PD} et S_{PV} . Cela signifie que pour un séisme donné, on peut connaître la vitesse, l'accélération et le déplacement que devra subir cette structure, représentée par un oscillateur 1-DDL. On comprend bien le spectre de réponse permet de mieux caractériser le mouvement du sol pour l'ingénieur que l'intensité macrosismique ou les autres paramètres du mouvement du sol telles que les valeurs maximales (PGA, PGV et PGD).

En effet, cette réponse spectrale montre directement la réponse d'une structure réelle d'amortissement et de fréquence donnés à un séisme. Attention cependant de ne pas oublier que l'intégrale de Duhamel est une intégrale valide en petite déformation et dans le domaine élastique du comportement de la structure. Ces spectres de réponse sont finalement des spectres élastiques et on ne peut pas évaluer le dommage final d'une structure uniquement en analysant les déformations données par le spectre de réponse du mouvement du sol. Cependant, la quantité maximale de la déformation élastique est selon certains auteurs une indication pertinente de l'intensité du mouvement du sol, sur une bande de période caractéristique de la plupart des structures. C'est pour cela que Housner proposa une mesure de l'intensité du mouvement sismique en définissant l'intégrale de la pseudo-réponse spectrale en vitesse, sur une bande de fréquence comprise entre 0.1 et 2.5 secondes. Housner proposa ainsi l'intensité spectrale par la relation:

$$SI(\xi) = \int_{0.1}^{2.5} S_{PV}(\xi, T) dT$$

1.7. FACTEURS AFFECTANT LE MOUVEMENT DU SOL

1.7.1. Facteurs liés au chemin traversé

Durant un tremblement de terre, les ondes sismiques se propagent dans un large rayon loin de la zone de rupture ce qui entraîne la modification de leurs caractéristiques en raison des réflexions et réfractions lorsqu'elles rencontrent des hétérogénéités, des surfaces de discontinuité et des barrières naturelles. Aussi, une partie de l'énergie transportée par ces ondes est absorbée dans les milieux traversés du fait de l'amortissement des matériaux. En générale, en s'éloignent de la source de rupture, l'amplitude du mouvement diminue (diminution des accélérations, vitesses et déplacements).

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.

Il y a plusieurs définitions qui ont été utilisées dans les dernières années pour mesurer effectivement la distance entre la source et la station d'enregistrement (Fig.1.5) suivant certains auteurs cette distance peut être la distance épiscopentrale, hypocentrale (ou focal), la plus proche à la faille de rupture et la distance au centre d'énergie.

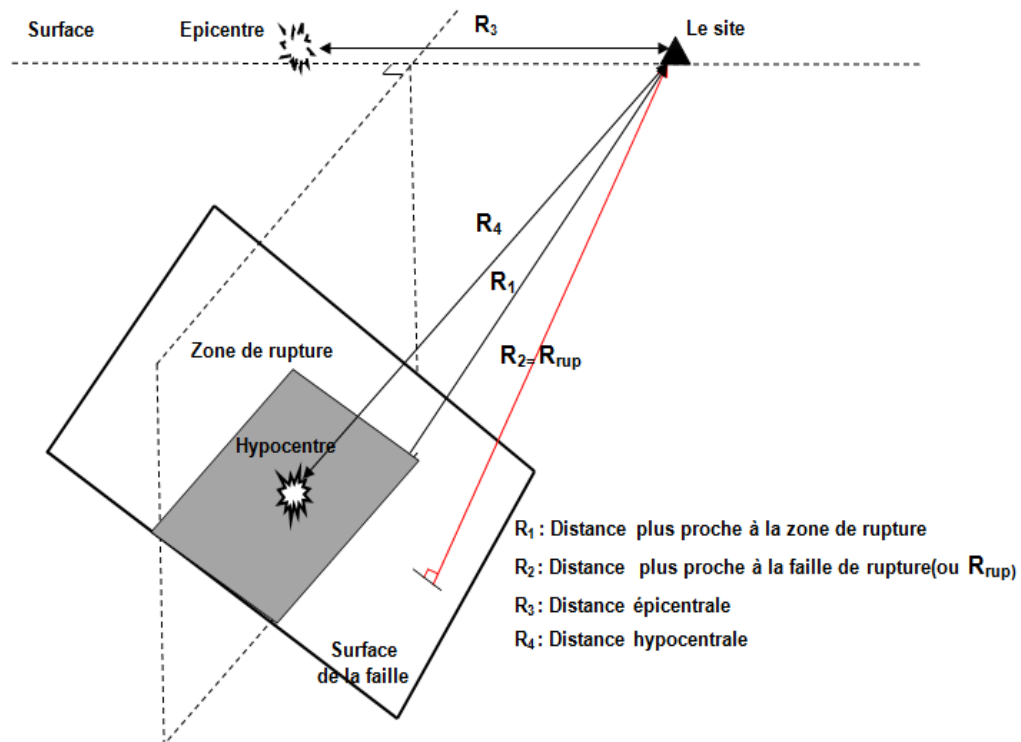


Fig.1.6 : Différents mesures de distances utilisées en géosismiques.

En champs proche ces mesures de distances peuvent différer significativement. L'utilisation de la distance épiscopentrale et hypocentrale a une avantage qu'elles peuvent être déterminées facilement, mais leur inadéquation est généralement convenu. La distance hypocentrale (ou focale) est souvent mal connue. La distance épiscopentrale 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 proche 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 (Fig.1.5).

1.7.2. Facteurs liés à la source

Le phénomène sismique représente un réajustement de la croûte terrestre par une libération brutale des contraintes tectonique cumulées .La quantité et la manière dont cette énergie est libérée sont deux facteurs très significatifs pour les dommages observés à la surface. 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 $\Delta\tau$.

Il existe de nombreuses évidences indiquant que, sauf peut-être pour les séismes profonds, un séisme est engendré par un mécanisme de cisaillement pure 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 (Fig.1.6) correspondant au plan de la faille. Le couple dû aux contraintes du cisaillement est :

$$M_0 = \iint_S \tau h dA$$

En désignant par G le module de cisaillement du milieu et par u le déplacement relatif des deux faces de la faille, il vient pour un milieu élastique :

$$M_0 = \iint_S G u dA = G \tilde{u} A$$

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.

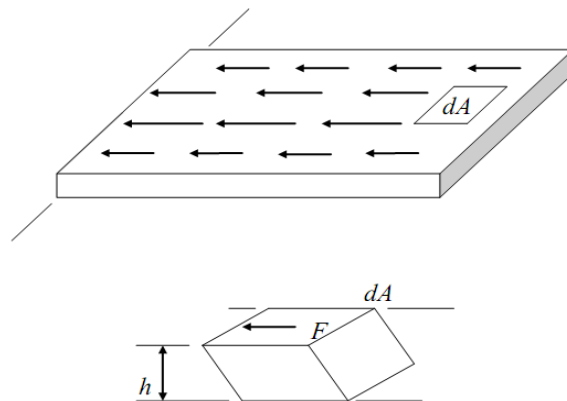


Fig. .1.7 : Schématisation d'un plan de faille (Brune 1976).

Dans un but de simplification, l'énergie à la source est parfois représentée par une mesure de la magnitude, des études récentes des lois d'atténuation des mouvements sismiques utilise le moment magnitude M_w pour caractériser la taille du tremblement de terre, (Joyer et Boore, 1982).

1.7.3. Facteurs liés aux conditions locales de site

Les conditions de site (géologiques, topographiques) ont été reconnues comme des facteurs contribuant à la distribution géographique des ruptures du sol et des endommagements structuraux durant un tremblement de terre. En effet, les enregistrements des mouvements du sol, ont montré que les variations des conditions du sol peuvent avoir une influence majeure sur le mode d'endommagement ainsi que sur l'amplitude et les caractéristiques fréquentielles des mouvements du sol à la surface.

Les enregistrements des mouvements des sols recueillis à partir de séismes passés ont contribué à la compréhension des effets de site pour une large variété de conditions géologiques. Ces enregistrements ont été utilisés comme une base d'études quantitatives de l'influence de la réponse du sol sur les caractéristiques des mouvements forts du sol. L'influence de la géologie locale est généralement mise en évidence à l'aide d'une classification sommaire de la nature de site. Seed et Al ont réalisé une étude statistique des spectres de réponse en accélération d'enregistrements des séismes en surface et ce pour différents types de sol reparties en quatre catégories (rocher, sol raide (rigide), sol pulvérulents profonds et sols mous) (Fig.1.7).

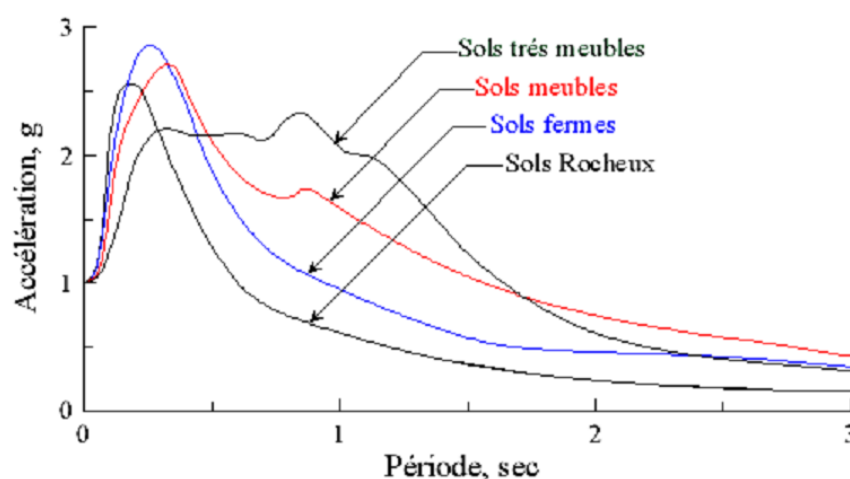


Fig.1.7: Spectres moyens d'accélération pour différentes conditions de site selon Seed et al. (1976) [15].

De cette étude, il en ressort que la fréquence prédominante du mouvement, correspondant au pic du spectre de réponse, diminue lorsque la rigidité du sol diminue. Autrement dit, plus le sol, est mou plus il filtre les hautes fréquences contenues dans le mouvement incidente, une structure rigide sera donc plus vulnérable sur rocher que sur sol mou et vice versa. Aussi, la plage de fréquence où l'accélération spectrale reste élevée et plus étendue pour les sols à faible rigidité.

CHAPITRE II
SISMILOGIE DES MOUVEMENTS FORTS :
PRINCIPES ET DONNEES

2.1. INTRODUCTION

L'un des buts de la sismologie est d'estimer le mouvement du sol en réponse à un séisme, en particulier en un endroit représentant un intérêt. Lorsque la structure est soumise à un mouvement de sol dû à un séisme, les ondes élastiques la traversent et celle-ci commence à vibrer à diverses fréquences, caractéristiques de sa propre résistance et de son agencement spatial. Elles s'étendent en général sur une plage de 0,1 à 10 Hz. Les séismes génèrent quant à eux des mouvements du sol sur une large gamme de fréquence, allant du déplacement permanent à quelques dizaines de cycles par seconde.

Les vibrations du sol correspondant aux fréquences particulières de la structure provoquent chez celle-ci une sensibilité exacerbée, et les dommages infligés alors à cette structure dépendent à la fois de ses propriétés et des caractéristiques des mouvements sismique [17]. De celui-ci, on peut extraire des informations sur les accélérations et vitesses maximales ressenties, la durée, le contenu en fréquence, l'énergie cinétique, la phase et les cohérences spatiales qui en constituent autant de caractéristiques. Les enregistrements en temps du mouvement du sol sont donc requis pour l'étude dynamique du comportement non-linéaire de bâtiments parasismiques et de structures particulières comme les ponts, les barrages [20],[21].

2.2. ESTIMATION DES MOUVEMENTS FORTS

La sismologie des mouvements forts est une science dédiée à la compréhension des vibrations provoquées par les forts séismes à courte distance des failles sismogéniques. C'est une science récente qui traite de la mesure, de l'interprétation et de l'estimation de ces mouvements. Ceux-là sont appelés mouvements forts car ce sont ces mouvements qui provoquent la destruction des infrastructures et que l'homme ressent au cours d'un séisme. Ils sont caractérisés par des amplitudes importantes et des fréquences élevées de l'ordre de l'Hertz [17].

Depuis 25 ans, tous les séismes de magnitude supérieur à 6 qui ont produit des enregistrements de mouvements forts, ont permis une avancé scientifique remarquable. Certains enregistrements ont révélé que les accélérations maximales peuvent atteindre et excéder la valeur de (3g), où (g) correspond à l'accélération maximale gravitationnelle terrestre. Il est cependant difficile de déterminer un seuil précis à partir duquel un mouvement est dit fort. Les vieux accéléromètres n'étaient pas d'une résolution suffisante pour

des valeurs inférieures à 10 cm.s^{-2} . Les instruments modernes quant à eux sont devenus beaucoup plus sensibles et enregistrent des mouvements du sol ayant une accélération maximale 100 fois inférieure. Les personnes, elles, peuvent ressentir des mouvements de faible amplitude, de l'ordre de 1 cm.s^{-2} . Enfin, les constructions non parasismiques présentent des dégâts pour des séismes modérés et pour des accélérations maximales de l'ordre 100 cm.s^{-2} , Aussi, une valeur seuil de 10 cm.s^{-2} semble convenir pour désigner comme forts mouvements d'amplitude supérieure [17].

2.3. DESCRIPTION EMPIRIQUE DES MOUVEMENTS FORTS

2.3.1. Les paramètres de description du mouvement

Ainsi qu'écrit précédemment, on peut extraire du signal sismique des informations sur l'accélération et vitesse maximales ressenties, la durée, le contenu en fréquence, l'énergie cinétique, la phase et les cohérences spatiales qui constituent autant de caractéristiques. Traditionnellement, l'évolution de l'accélération, de la vitesse et du déplacement du sol au cours du temps, l'intégrale du carré de l'accélération, la transformée de Fourier et le spectre de réponse font l'objet d'une analyse permettant de caractériser un séisme [17], [21].

Une fois l'enregistrement traité et corrigé des effets instrumentaux, des paramètres additionnels en sont extraits. La durée du mouvement fort est considérée comme importante, quelle que soit la manière dont la définit. Ainsi, on peut par exemple calculer la somme des intervalles temporels au cours desquels la trace temporelle excède une amplitude de $0.05g$ ainsi que Bolt (1969) l'a fait. Une autre définition proposée par Trifunac & Brady (1975) consiste à retenir l'intervalle de temps situé entre les instants où 5 et 95% d'une valeur cible sont atteints, laquelle valeur correspond à l'intégrale du carré de l'accélération enregistrée [9]. Bien que de même, ces deux durées conduisent à des résultats opposés lorsque la distance à la source sismique augmente. A grande distance, le pic d'accélération diminue entraînant une diminution de la durée de Bolt(1969), tandis qu'à l'inverse, l'intensité d'Arias(1970) (proportionnelle à l'intégrale du carré de l'accélération) diminue et s'étale au cours du temps, produisant un allongement de la durée de Trifunac & Brady (1975)[17].

Enfin, les spectres de réponse en accélération, vitesse ou déplacement peuvent être calculés, ils sont d'un intérêt remarquable pour le dimensionnement des structures face au risque sismique [20], [22]. La figure (2.1) illustre quelques unes des grandeurs mentionnées. La ligne supérieure montre trois enregistrements temporels du mouvement du sol et ses dérivées, et la ligne inférieure les trois contenus fréquentiels correspondants.

Entre les deux, l'allure de l'intensité d'Arias (à une constante près) est montrée. Enfin, le dernier graphique montre le spectre de réponse en accélération.

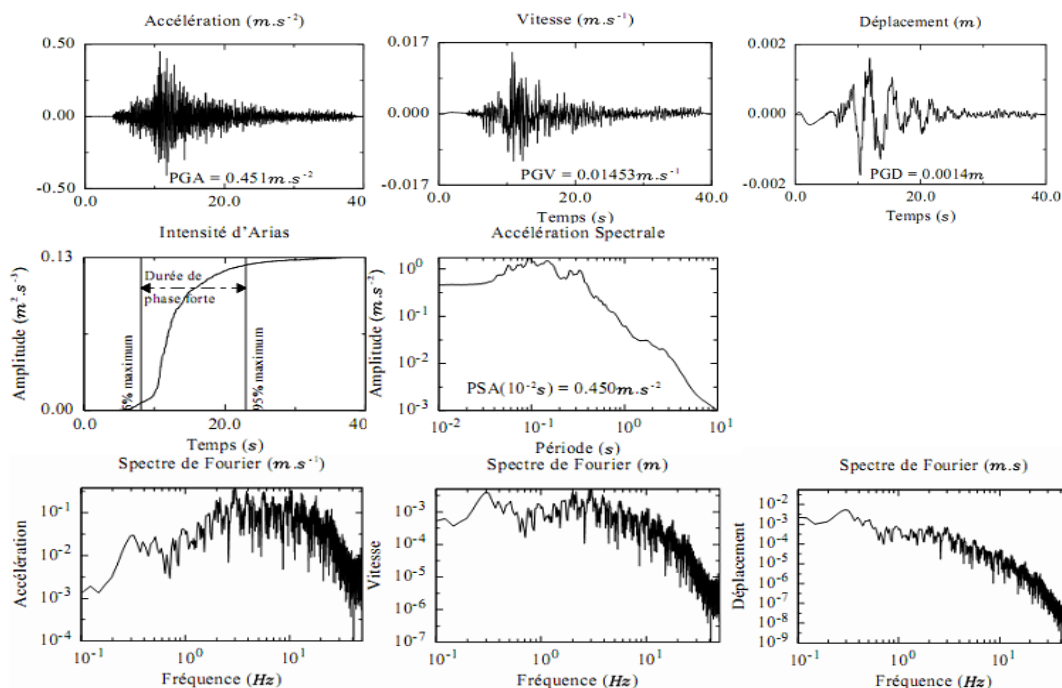


Fig.2.1 : Evolution de l'accélération, de la vitesse et du déplacement du sol au cours du temps.

2.3.2. Les équations prédictives

Les paramètres décrits au paragraphe précédent peuvent faire l'objet d'une tentative de prédiction effectuée sur la base d'enregistrements existants. Les équations prédictives de mouvement fort renseignent sur la valeur attendue d'un paramètre choisi en fonction de la magnitude du séisme, de la distance à la source sismique et du type de sol sur lequel on se place [21].

- *Le modèle du point source*

Une relation prédictive permet de rendre compte de la combinaison des trois effets (source, chemin, site) à l'aide d'un modèle physique simple pour décrire les paramètres du tremblement de terre, par exemple :

$$Y = f(M, R, S)$$

La source du séisme est assimilée à un point situé à l'hypocentre du séisme. Le modèle suppose que toute l'énergie du séisme est libérée à partir de ce point. Le principe de conservation de l'énergie impose que la propagation sphérique des ondes sismiques dans le milieu s'accompagne d'une décroissance en $1/r$ de l'amplitude de ces ondes, où r désigne la distance parcourue par l'onde.

Lorsque le milieu est anélastique, alors à cette décroissance de l'amplitude avec la distance s'ajoute un effet de dissipation qui est fonction du paramètre d'atténuation anélastique.

Par exemple, une relation prédictive du mouvement fort, relative à la valeur maximale enregistrée de l'accélération (PGA), qui adopte le formalisme du point source peut s'écrire :

$$\text{Log}(PGA_{ijk}) = aM_i + bR_j - \text{Log}(R_j) + dS_k$$

Où :

PGA_{ijk} , est l'accélération maximale pour le i^{me} séisme, enregistrée à la j^{me} station avec une k^{me} condition de site.

a , c'est un coefficient relié à la magnitude M .

b , C'est un coefficient relié à la distance et rend compte de l'atténuation anélastique.

– $\log_{10}(R)$ correspond à l'atténuation géométrique. Finalement, d est un coefficient lié au site S_k sur lequel se trouve la station d'enregistrement.

- Les affinements du modèle

Les modèles choisis généralement pour éditer les équations prédictives du mouvement fort adoptent la formulation suivante [21] :

$$\begin{aligned} \text{Log}(Y) = \beta_0 + \beta_1 M + \beta_2 (M_{max} - M)^{\beta_3} + \beta_4 \log(R^2 + \Delta(M)^2)^{1/2} + \beta_5 R \\ + F(\text{source}) + F(\text{site}) \pm (\sigma_{\log Y}) \quad (\text{Eq. 2.1}) \end{aligned}$$

Y , est le paramètre d'intérêt, M est la magnitude, R est la mesure de la distance considérée, $F(\text{site}), F(\text{source})$ sont des variables qui rendent compte du type de source et de site considérés et les coefficients β_i sont déterminés par la régression. Finalement, $\sigma_{\log Y}$ est l'incertitude associée au calcul et rend compte de la dispersion des données autour de la valeur médiane prédite.

- Au premier ordre, $\log(Y)$ est proportionnel à la magnitude, mais des études récentes suggèrent un effet de saturation avec la magnitude (Anderson, 2000, Douglas, 2003), à mesure que la magnitude augmente, l'amplitude du mouvement augmente d'autant moins vite. C'est ce à quoi fait référence le terme $(M_{max} - M)^{\beta_3}$ dans (Eq.2.1).
- Un autre type de saturation a été perçu, qui exprime le fait que l'amplitude du mouvement à proche distance de la source ne rend que partiellement compte du rayonnement total

du phénomène de rupture sismique d'une zone d'étendue finie (Campbell,1981,Abrahamson et Silva ,1997). Cet effet de saturation en distance se traduit dans les calculs par une pente d'atténuation géométrique (le terme Δ dans *Eq. 2.1*), qui diminue à proche distance.

- De plus, le modèle de prédiction rend également compte d'une atténuation anélastique ($\beta_5 R$) (*Eq. 2.1*) qui traduit absorption énergétique du milieu traversé par les ondes sismiques.
- Finalement la catégorie de site sur lequel on prédit le mouvement a son importance et reproduit la tendance qu'un sol peu rigide a toutes choses égales par ailleurs, à amplifier le mouvement observé sur un sol rigide. La classification du site fait intervenir des critères géologiques descriptifs et /ou un paramètre mathématique qui rend compte de la rapidité avec laquelle les ondes de cisaillement (les plus destructives) traversent le milieu sur une profondeur donnée. Usuellement, le critère mathématique est privilégié et on le calcul sur les 30 premiers mètres de profondeur du sol. plus précisément, on calcul la moyenne harmonique des ondes cisaillement sur cette épaisseur (V_s^{30}).

Il faut remarquer qu'on raisonne pour l'établissement de ce genre d'équation non pas avec la valeur prise par le paramètre du mouvement fort Y mais avec le logarithme de cette valeur, en raison de la distribution normale de $\text{Log}(Y)$ autour de sa médiane [19] : pour cette raison, une fois l'équation établie et dans le cadre d'une prédiction, on parle de médiane attendue et non de moyenne attendue. Cette valeur médiane n'est qu'une indication autour de laquelle, par comparaison, la valeur expérimentale réelle se situe. Cette différence entre la donnée expérimentale et la valeur médiane prédite est traitée comme aléatoire.

2.3.3. Différents types des relations prédictives

On distingue dans la littérature trois catégories de relations prédictives du mouvement fort lorsqu'on les utilise pour des études d'aléa sismique dans une région particulière [1]:

- a) Les relations provenant de données d'événements crustaux dans des zones sismiques actives, comme par exemple le Nord-Ouest du continent Américain, l'Italie, la Grèce, la Turquie ou encore l'Algérie.
- b) Celles issues de données enregistrées dans des régions continentales stables, qui sont caractérisées par des chutes de contrainte élevées, avec des valeurs élevées de Q , et donc révélant un contenu peu atténué de hautes fréquences dans les signaux.

- c) Et enfin, les relations provenant de régions de subduction, où les séismes enregistrés ont eu lieu à grande profondeur et à l'intérieur de la plaque plongeante, ou bien à plus faible profondeur et à l'interface des deux plaques [6] .

Historiquement, les estimations des paramètres de mouvement fort (accélération et vitesse maximales, spectre de réponse, durée du mouvement fort, ...) reposaient sur l'exploitation d'enregistrements existants. Ils servaient à établir une régression qui avait une valeur empirique, et cette approche est toujours assez largement utilisée dans le milieu conventionnel de l'ingénierie: Douglas (2003) a recensé ainsi plus de 200 relations prédictives de mouvement fort (accélération maximale et spectre de réponse). Mais il existe toujours un problème relatif à l'utilisation des relations prédictives éditées en des endroits du globe différents du lieu d'application. Il faut observer que les relations prédictives de mouvement fort ont toutes un caractère régional et il est inopportun d'exploiter une relation établie aux Cascades pour tirer des conséquences sur l'aléa sismique existant en Algérie par exemple : les conditions géologiques, morphologiques et tectoniques ne sont en effet pas communes aux deux régions. Ce problème d'application des relations occupe actuellement la communauté scientifique (Campbell, 2003, Scherbaum et al, 2004, Cotton et al, 2005, Scherbaum et al, 2005).

CHAPITRE III
BANQUE DE DONNEES

3.1 .INTRODUCTION

Dans ce chapitre on présentera une description complète de la méthodologie et les principaux critères à suivre pour la sélection d'un certain nombre d'enregistrements accélérométriques qui feront l'objet d'une étude statistique dans le but d'établir une loi prédictive pour l'estimation de la durée de phase forte. Les enregistrements sont issus d'une banque de données mondiale.

3 .2.SELECTION DES ENREGISTREMENTS SISMIQUES

Les enregistrements sismiques historiques ont été choisis dans la base de données mondiale [33]. Cette base de données contient des enregistrements des mouvements forts représentant plus de 160 évènements sismiques de magnitudes variant de 4.2 à 7.9. Le choix de tels enregistrements est basé sur certain un nombre de critères :

3 .2.1.Critères de choix des enregistrements

La sélection des enregistrements historiques est faite comme suit :

- (i) les enregistrements sont représentatifs de l'activité sismique dans des régions continentales différentes telles que : (Chi-Chi, Taiwan, 1999, San Francisco, 1957 ; Loma Prieta ,1989 ; Kocaeli Turkey,...etc).
- (ii) les combinaisons (M, R) retenues devaient être les séismes ayant un scénario moment magnitude (M_w) – distance plus proche à la faille de rupture (R_{rup}) représentatif du risque sismique :
 - Plusieurs échelle de magnitudes existe (§.1.4.1, chapitre I), de plus ces échelles saturent pour des forts séismes. On a choisi le moment magnitude : M_w (Hanks et Kanamori, 1979) parce que le moment magnitude représente une échelle ouverte et a un sens physique plus réaliste pour mesurer l'énergie sismique.
 - Nous avons adopté comme mesure de distance : la plus proche a la faille de rupture ce qui devrait être plus représentatifs des dégâts observés en surface, à savoir que l'énergie sismique atteint le site à partir d'un point plus proche à la faille (§.1.7.1, chapitre I).
- (iii) Un autre aspect important à considérer lors de choix des mouvements sismiques est la géotechnique du site sur lequel le sismographe était localisé. La classification du site fait intervenir des critères géologiques : C'est un paramètre mathématique qui rend compte de la rapidité avec laquelle les ondes de cisaillement traversent le milieu sur une

profondeur donnée. On calcul donc la moyenne harmonique des ondes de cisaillement sur cette épaisseur (V_s^{30}) conformément à l'expression suivante :

$$V_s^{30} = \frac{30}{\sum_{i=1}^n \frac{h_i}{v_i}}$$

Ou :

h_i et v_i désignent l'épaisseur (en mètres) et la vitesse des ondes de cisaillement pour la $i^{\text{ème}}$ couche. Plusieurs codes de classification existent dans ce sens (Annexe II). Nous avons opté pour la classification décrite par (UBC, 1997). Ce code nous permet de limiter les profils de sol en trois classes distinctes (Tab.3.1).

Tab.3.1 : Classification du site selon (UBC, 1997).

Type de profil de sol	description	La valeur moyenne de vitesse des ondes de cisaillement dans les 30 m supérieurs: V_s^{30} (m/s)
A	Roche dure	$V_s^{30} \geq 1500$ m/s
B	Roche	$V_s^{30} = 760 - 1500$ m/s
C	Sol très dense	$V_s^{30} = 360 - 760$ m/s
D	Sol raide	$V_s^{30} = 180 - 360$ m/s
E	Sol meuble	$V_s^{30} \leq 180$ m/s

3.2.2. Présentation des différentes classes : répartition des données

Le code de classification décrit précédemment (UBC, 1997) nous permet de présenter la banque de données sous une classification sommaire en trois classes selon le type du profil de sol : sol rocheux, sol raide (intermédiaire), sol meuble. Les données sont divisées en 36 sous-classes différentes correspondantes à 4 intervalles du moments magnitudes M_1 à M_4 ($M_1 < 5.5$, $5.5 < M_2 < 6.5$, $6.5 < M_3 < 7.5$, $M_4 > 7.5$), trois intervalles de distances R_1 à R_3 (Tab.3.2), trois catégories de sites (Tab.3.3).

Tab.3.2: Classes des distances.

	R_1	R_2	R_3
M_1	< 15	15 – 30	> 30
M_2	< 20	20 – 40	> 40
M_3	< 25	25 – 50	> 50
M_4	< 30	30 – 60	> 60

Tab.3.3 : Répartition de la banque de données pour différentes classes de sol.

	Sol rocheux S1			Sol intermédiaire S2			Sol meuble S3		
	R ₁	R ₂	R ₃	R ₁	R ₂	R ₃	R ₁	R ₂	R ₃
M₁
M₂	08	06	02	04	02	06	02	04	02
M₃	16	22	04	08	06	58	04	08	06
M₄	02	20	26	69	84	56	26	69	84
Tot1	26	48	32	81	92	124	32	81	92
Tot2	184			318			205		

3.2.3. Présentation des enregistrements pour le sol rocheux

La banque d'enregistrements représentatifs du profil de sol S1 (Tab.3.3) contient en totale 12 événements sismiques, comprenant 184 accélérogrammes enregistré sur des sols rocheux, représentée ici par les deux composantes horizontales. La figure (3.1.a) représente le nombre des composantes enregistrés pour chaque événement et pour chaque class du moment magnitude.

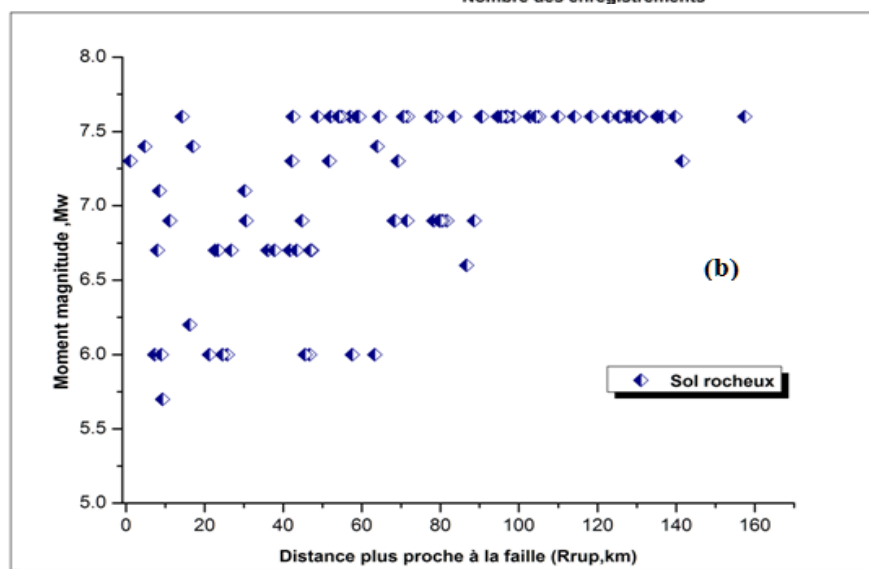
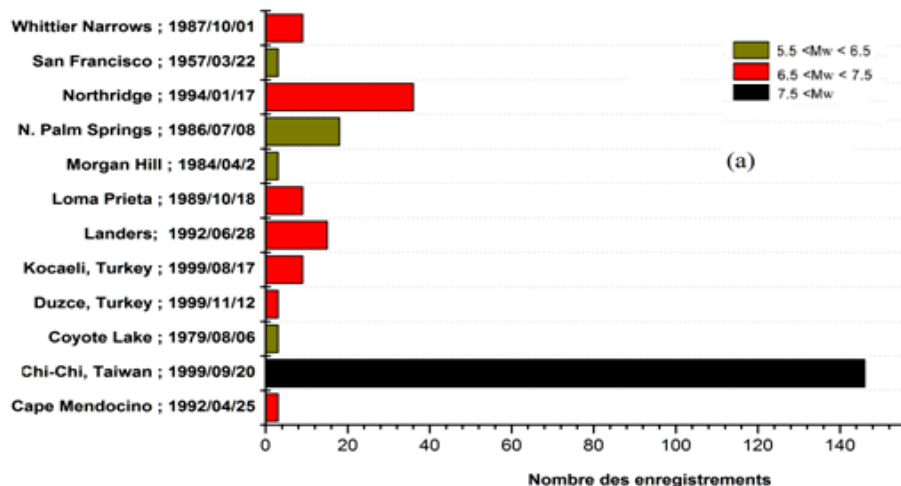


Fig.3.1 : Représentation des données pour le sol rocheux.

La figure (3.2.b) représente l'ensemble des scénarios (M_w, R_{rup}) dans la banque de données retenus pour les sols rocheux. On remarque qu'il y a une manque d'enregistrements pour la classe magnitude ($M_w \leq 5.5$), la plupart des enregistrements d'accélération se présentent pour des événements majeurs telques : Chi-Chi ,1999 ; Loma Prieta,1989 ;etc,pour un moment magnitude allant de 6.5 à 7.9 et pour une distance $R_{rup} \leq 160$ km.

En analysant la figure (3.2.b), on peut dire que : les données des mouvements forts concernant les sols rocheux semblent pratiquement biens distribuées en champs proche en terme de magnitude et distance, cependant que une dispersion a été marquée pour les faibles événements. On peut ajouter que certains événements sont très riches des enregistrements, comme celui de (Chi-Chi, Taiwan ; ,1999) qui vient d'une région très densifiée des réseaux de sismographes dotés des stations d'enregistrement modernes qui peuvent même enregistrer des scénarios en champ lointain et pour des distances très grandes, cette événement représente plus de 50 % de la banque de données.

CHAPITRE IV
DUREE DE PHASE FORTE

4.1. INTRODUCTION

Un tremblement de terre peut être caractérisé par des différents paramètres, chacun reflète une caractéristique particulière de la secousse sismique telle que les amplitudes maximales, le contenu en fréquences, ou l'énergie portée dans le signal. Une caractérisation complète du tremblement de terre doit inclure une mesure de sa durée, ou plus spécifiquement la durée de la partie de l'enregistrement sismique qui est considérée forte. La durée du mouvement fort du sol est reconnue comme étant une caractéristique importante affectant la réponse des structures, son importance pour la liquéfaction des sols saturés a été reconnue en premier (Seed et Idris ,1967 et 1971). Les degrés des dommages causés sur les structures et les sols sont en relation directe avec la durée du mouvement sismique. Des études analytiques et expérimentales ont montré et quantifié l'importance de la durée sur les dommages et les écroulements des structures (Husid, 1967), les mouvements avaient été de plus courtes durées, aucun dommage ne serai produit, par contre si la durée des mouvements sismiques avait été plus longue, on aurait sûrement vécu des inondations catastrophiques pour les habitants de la région (Seed et al.1975).

4.2. LA DUREE DU MOUVEMENT SISMIQUE

La durée totale d'un enregistrement ne présente qu'un intérêt limité pour l'étude des 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 introduire. Seule la durée de phase forte d'un séisme est considérée comme un paramètre important pour en caractériser sa nocivité. En raison du caractère non stationnaire du mouvement sismique, un problème existe :

- quelle est la partie du mouvement sismique qui est considérée comme significative pour les dégâts observés ?

L'identification de la phase forte du signale sismique est le premier problème qu'on s'adresse dans l'étude des propriétés de la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a).

4.2.1. Dépendance du RMS_a avec la durée

La racine carrée de la moyenne quadratique des accélérations du sol (RMS_a), quoique ce soit défini dans le domaine des temps ou le domaine des fréquences, elle est dépendante de

la durée d'un enregistrement sismique. Pour des situations dont lesquelles le processus est fortement non stationnaire, la durée devient un paramètre très critique.

Par seule observation visuelle, on peut voir qu'elle est dépendante de tout l'enregistrement d'accélération ou d'une partie de cet enregistrement. Pour étudier la dépendance du RMS_a avec la durée de l'enregistrement ou d'une partie bien particulière de cet enregistrement, qui représente la partie dite phase forte du mouvement sismique, considérons l'enregistrement d'accélération $x(t)$, la moyenne quadratique des accélérations du sol dont l'expression pour cet enregistrement, est :

$$RMS_a(t_m)^2 = \frac{1}{m} \sum_{i=1}^m x(t_i)^2 \quad m = 2, \dots, M$$

Où : $x(t_i)$ est la valeur de l'accélération au temps ($t_i = i. \Delta t$), l'indice M est associé à la durée total ($T_d = M. \Delta t$)e de l'accélérogramme, donc on peut déduire la racine carrée de la moyenne quadratique des accélérations du sol à chaque instant $RMS_a(t_m)$.

Le tracé des valeurs du RMS_a en fonction du temps par une représentation continue s'appelle la fonction RMS_a cumulative CRF(t_m) , le terme cumulative, veut dire le cumul en valeurs du RMS_a jusqu'à l'instant t_m .

La fonction CRF nous permet de représenter le taux d'accroissement de l'énergie libérée par la rupture tectonique en fonction du temps (Fig. 4.1). En effet, cette fonction permet de visualiser l'apport d'énergie à chaque instant par une augmentation de sa valeur, qui se traduit géométriquement par une pente positive de la courbe représentative de cette fonction .La tendance générale de la fonction CRF est d'atteindre à un instant donné un pic et décroître par la suite ,ce qui signifie qu'au-delà de cet instant ,la pente de la fonction CRF prend des valeurs constamment négatives et l'énergie arrivant au site d'enregistrement n'a aucun apport significatif. Cette énergie ne peut être associée qu'aux ondes de surface et aux vibrations libres de la terre qui vont persister avec des amplitudes relativement faibles avant de s'amortir en fin du signal. La valeur finie de l'énergie tectonique confère un caractère transitoire aux enregistrements sismiques, ce qui permet de déterminer la partie du mouvement sismique dite forte (Fig.4.1) définie par l'intervalle du temps compris entre les deux pics des CRFs calculés dans les deux sens (de début à la fin de l'enregistrement et dans le sens contraire) [22] , [23].

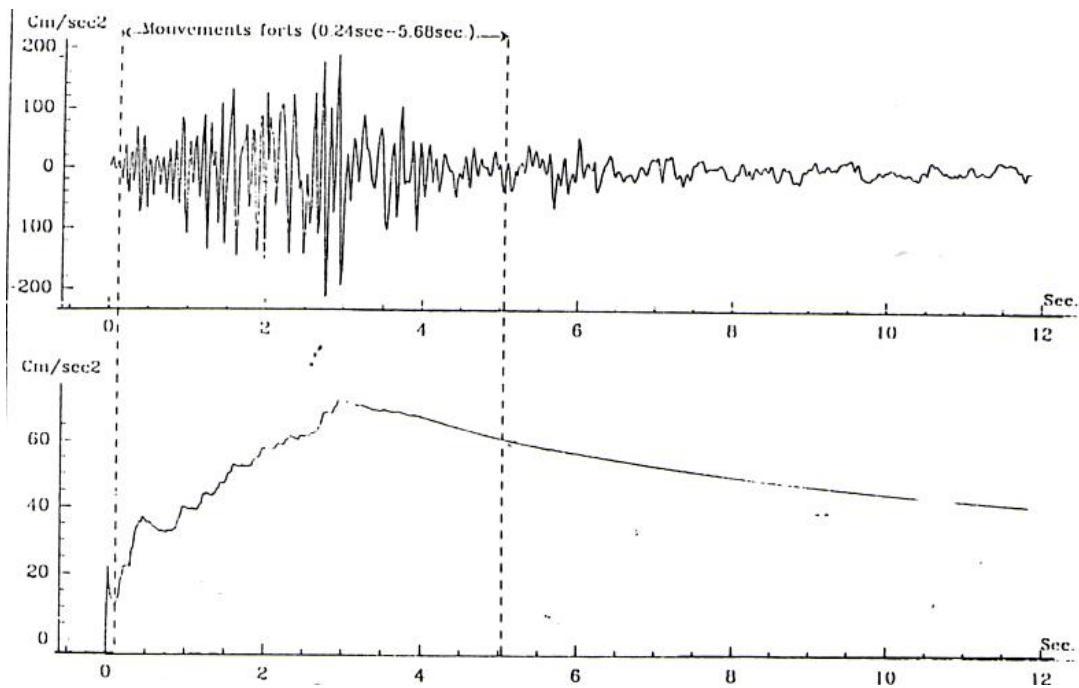


Fig.4.1 : La fonction RMS_a cumulative (séisme d'El-Asnam, réplique du 30/10/1980, composante NS) [18].

4.2.2. Durée de phase forte du mouvement sismique

De nombreux travaux consacrés à la détermination de cette durée de phase forte déterminante pour mesurer le potentiel d'endommagement d'un séisme ont donné lieu à des expressions variées de ce paramètre:

- a) **La première définition** : proposée par **McCann et Shah(1979)**, laquelle est reliée à la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a). Ces auteurs proposent que l'intensité du tremblement de terre peut être représentée, notamment par une moyenne statistique comme la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a). En identifiant la partie dite mouvement fort, la partie du signal sismique qui présente un niveau d'énergie très important. En effet, pour la définition de la durée d'un enregistrement $x(t)$, ces auteurs utilisent la "Cumulative RMS function $CRF(t)$ " dont l'expression, pour $x(t)$:

$$CRF(t_m) = \sum_{i=1}^m \sqrt{\frac{x(t_i)^2}{m-1}} \quad m = 2, \dots, M$$

Où $x(t_i)$ est la valeur de l'accélération au temps ($t_i = i \cdot \Delta t$), l'indice M est associé à la durée totale ($T_d = M \cdot \Delta t$) de l'accélérogramme.

Cette fonction représente le taux d'accroissement de l'énergie du signal en fonction du temps. Une des caractéristiques de la fonction CRF(t) est sa tendance générale à atteindre à un instant donné un pic et de décroître par la suite. En terme physique, cela signifie qu'à partir de cet instant, l'énergie arrivant au site d'enregistrement est peu significative et ne peut être associée qu'aux ondes de surface et aux vibrations libres de la terre. Ainsi, en localisant les valeurs T_2 et T_1 des instants associés respectivement aux pics de la CRF(t) de l'accélérogramme pris normalement une première fois et en inversant l'axe des temps une deuxième fois, on localise ainsi, une portion de mouvement entre ces deux instants qui est due contribution des ondes sismiques émises par la source durant le processus de rupture.

La durée ($D = T_2 - T_1$) représente la portion dite de mouvement fort. Les énergies secondaires situées de part et d'autre de cette partie ne sont pas prises en compte car étant sans signification particulière (Fig.4.2).

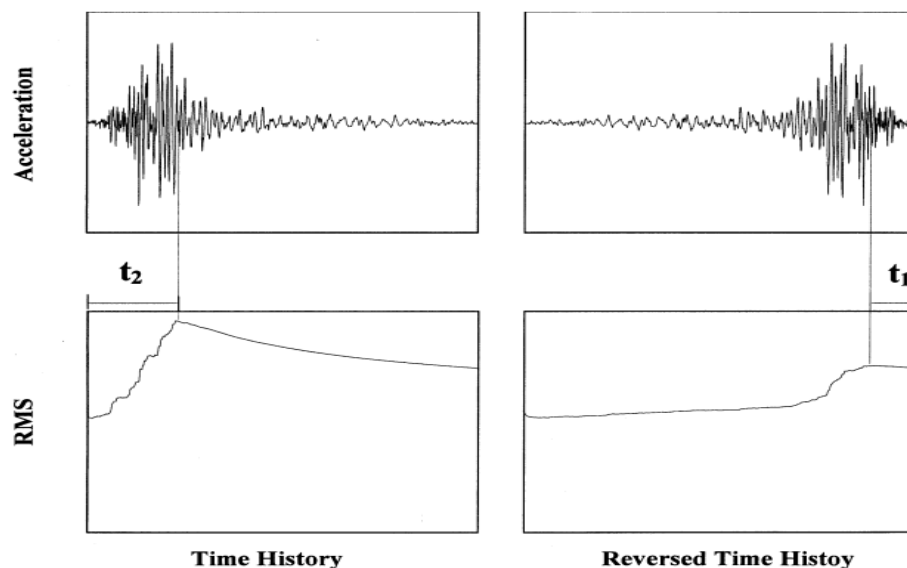


Fig.4.2: Durée de phase forte selon la définition de McCann et Shah (1979).

- a) **La deuxième définition** : se base sur la notion de cumul d'énergie du signal. Ce cumul s'exprime à partir de l'intensité d'Arias (1970). 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 et la plus connue est celle de **Trifunac et Brady (1975)** qui définissent la durée de phase forte d'un signal sismique : la partie comprise entre les instants T_1 et T_2 qui correspondent respectivement à la fraction de l'intensité d'Arias 5% et 95 % (Fig.4.3) [9].

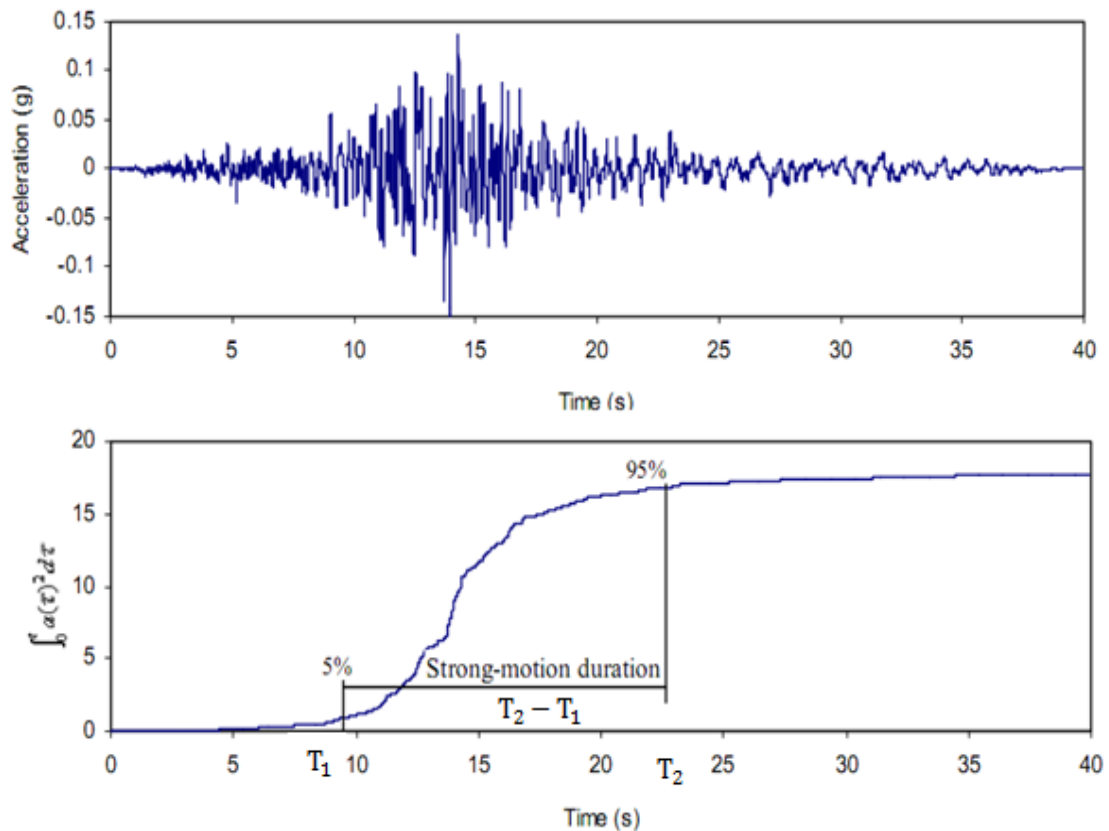


Fig.4.3 : Durée de phase forte selon la définition de Trifunac-Brady(1975).

- b) **La troisième définition :** 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) avec un seuil de (+/- 0.05g) ou par Mc Guire (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 alors 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 (Fig.4.4) [9].

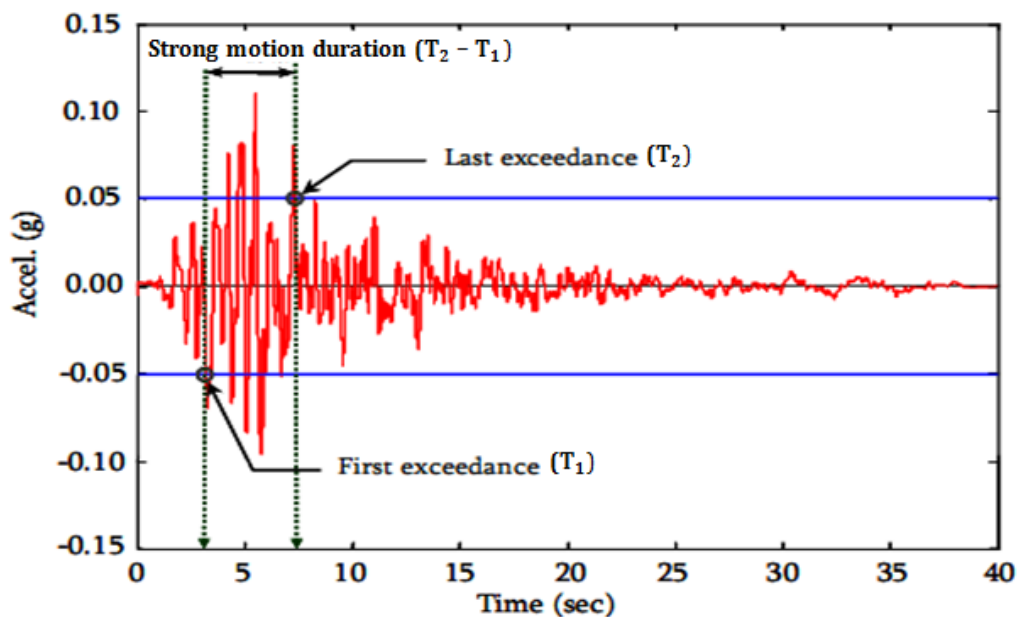


Fig.4.4 : Durée de phase forte selon la définition de Bolt (1973).

Toutes ces définitions sont basées sur l'étude du signal dans le domaine temporel. Trifunac 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, Trifunac 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.

Dans ce contexte, nous nous intéresserons aux résultats donnés par la méthode de McCann et Shah (1979), pour en développer une loi empirique reliant la durée aux différents paramètres environnementaux (dans notre cas les paramètres environnementaux : moment magnitude, distance plus proche à la faille, et les conditions locales de site) au fur et à mesure que nous avancerons dans notre mémoire.

4.3. ANALYSE COMPARATIVE ENTRE LES DIFFERENTES MESURES DE LA DUREE

Dans les sections précédentes, nous avons présenté trois définitions dédiées par la littérature pour calculer la durée qui représente la phase forte d'un signal sismique. Plusieurs auteurs ont mis en relief que la détermination de cette durée, déterminante pour mesurer le potentiel d'endommagement d'un séisme est dépendante de l'usage qu'on veut en faire : **McCann et Shah (1979)** proposent que l'intensité du tremblement de terre peut être représentée, notamment par une moyenne statistique comme la racine carrée de la moyenne quadratique des accélérations du sol (RMS_a), en identifiant la partie dite mouvement fort, la partie du signal sismique qui présente un niveau d'énergie très important. Si on est intéressé par la durée des mouvements sismiques dont les amplitudes sont au-dessus d'un niveau donné d'excitation, La définition donnée par **Bolt (1973)** peut être employée. Ainsi, si toutes les parties d'un accélérogramme ont le même niveau d'amplitudes, alors la durée de l'enregistrement coïncide avec la durée totale. Une autre définition de la durée liée à la réponse d'une structure a été proposée par **Trifunac et Brady (1975)** et concerne le temps durant lequel un certain pourcentage de l'énergie totale enregistrée est appliqué à une structure.

Toutes ces définitions ont quand même un point commun, à savoir celui de prendre en compte la partie du mouvement qui produit la réponse la plus significative d'une structure.

Les résultats de la durée de phase forte pour les différents scénarios (M_w, R_{rup}) calculés selon les trois définitions pour chaque événement sismique sont listés en annexe (I).

Pour mieux apprécier la comparaison entre les différentes définitions pour déterminer la phase forte de chaque enregistrement d'accélération, les résultats obtenus par la méthode proposée par McCann et Shah $D(M)$ avec celles des deux autres $D(T)$ et $D(B)$ sont graphiquement comparés (Fig.4.5). Les coefficients de corrélations pour ces comparaisons, sont :

Tab.4.1 : Coefficients de corrélation pour la durée.

Méthode	Coefficients de corrélation
	Durée de phase forte (D, sec)
Trifunac et Brady	0.87
Bolt	-0.25

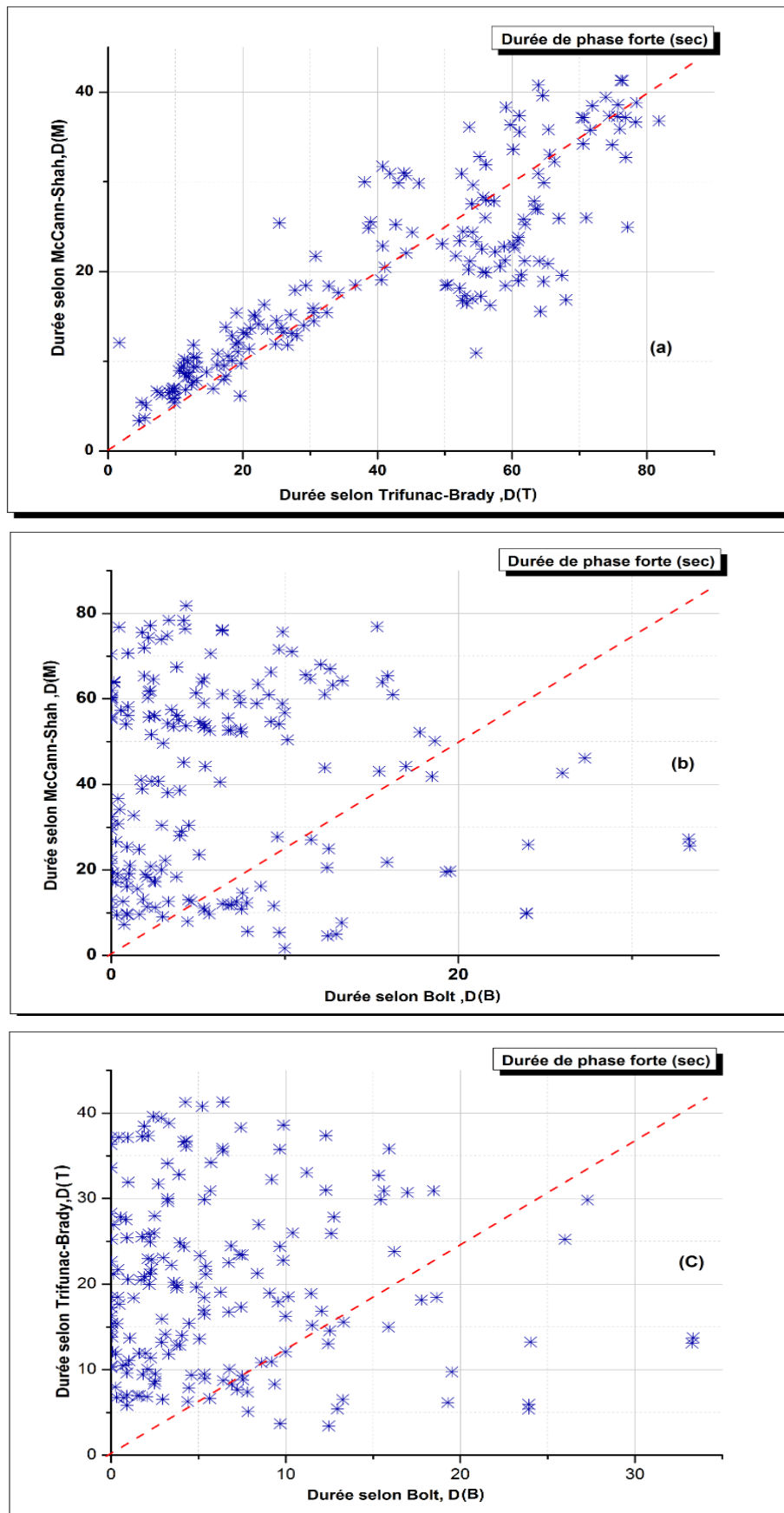


Fig.4.5 : Analyse graphique entre les différentes mesures de la durée.

À partir de ces figures, les observations générales suivantes peuvent être faites :

- La définition proposée par McCann et Shah donne une courte durée par rapport à celle définie par Trifunac et Brady (Fig.4.a), mais en comparaison avec celle de Bolt, généralement une longue durée a été obtenue (Fig.4.b).
- Pour la durée de Bolt, il y a une large dispersion sans tendance générale évidente, avec des valeurs presque nulles pour la plupart des scénarios (Fig.4.b et Fig.4.c).

En analysant le tableau en annexe (III), on voit que :

La méthode de calcul des durées selon Trifunac et Brady, $D(T)$ donne des résultats légèrement supérieurs à celles calculés par McCann et Shah, $D(M)$ pour la plupart des scénarios (Fig.4.5,a), une légère différence de 2 à 3 secondes pour des stations situées en champs proche ($R_{rup} \leq 20$ km), Ce qui peut être expliqué par des résultats nettement très proche des valeurs du RMS_a pour ces deux définitions et ceci est dû à la propriété asymptotique de la fonction RMS_a cumulative.

Les valeurs des durées calculées selon Bolt $D(B)$ sont très courtes par rapport à celles estimées par McCann-shah $D(M)$ et Trifunac-Brady $D(T)$. Ceci est vrai pour les stations situées en champs lointain. Il est difficile d'identifier pourquoi ca, mais ceci peut être dû au choix arbitraire du niveau de seuil d'accélération qui est dans notre cas de +/- 0.05g.

CHAPITRE V
NOTIONS SUR L'ANALYSE DE
REGRESSION

5.1 .INTRODUCTION

Dans le domaine des sciences appliquées, on observe fréquemment des phénomènes tels qu'il est possible de supposer l'existence d'une relation entre deux variables, par exemple : la durée de vie d'une ampoule électrique peut être liée à son rendement énergétique. Dans une étude statistique, on mesure, sur chaque unité d'un échantillon, différentes variables et on cherche s'il existe une certaine forme d'association entre elles. Le cas le plus simple est celui d'une dépendance statistique ou corrélation. *En régression*, le problème est de nature différente. On dispose de n couples (X_i, Y_i) constituant un échantillon d'observations indépendantes du couple de variables X et Y . On cherche une relation statistique pouvant exister entre la variable expliquée Y et la variable explicative X . Cette relation doit permettre de prévoir la valeur de Y pour une valeur donnée de X . Le problème est de nature dissymétrique. Au moins, trois questions se posent :

- 1) Quel est le modèle statistique le mieux adapté pour décrire la liaison entre les variables X et Y ? -Doit-on utiliser un modèle linéaire, non linéaire parabolique, exponentiel, etc. ?
- 2) En admettant comme plausible un modèle particulier, comment estimer les paramètres figurants dans ce modèle ?
- 3) Comment définir les outils permettant de calculer les valeurs prévisionnelles de la variable Y en fonction de la variable X ?

Pour répondre à ces questions, nous allons essayer dans ce chapitre de cerner un plus grand nombre de notions et concepts sur l'analyse de régression.

5.2. ANALYSE DE REGRESSION

De nos jours, les ingénieurs et les chercheurs sont fréquemment amenés à ajuster un modèle de régression par des méthodes linéaires ou non linéaire. Donc l'objectif est de déterminer 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 des valeurs moyennes ou des prévisions de valeurs individuelles d'une variation à partir des valeurs d'une autre variable. Pour ce faire ils doivent faire appel à la théorie des modèles de régression linéaire et non linéaire.

5.2.1. Régression Linéaire [7]

Étant donné n couples d'observations $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, l'objectif est d'établir la relation statistique la mieux appropriée entre les variables concernées.

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 celles de la variable explicative X et la composante aléatoire ε
 Y_i : représente la $i^{\text{ème}}$ observation de Y .

β_0 et β_1 sont les paramètres du modèle de régression.

X_i 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.

a) Modèle de régression linéaire multiple

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.

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 :

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

Où Y_i est la variable dépendante(ou expliquée) dont les valeurs sont conditionnées 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, $X_i^{(j)}$, $j = 1, 2, \dots, k$ représente 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 prises en considération dans le modèle ou que nous ne savons pas l'identifier.

b) Conditions d'application du modèle linéaire multiple

Le modèle de régression linéaire multiple est fondé sur un ensemble d'hypothèses concernant les données et la relation retenue :

1. *La linéarité* : $Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)} + \varepsilon$. La relation entre Y et l'ensemble des variables indépendantes est linéaire.
2. *Plein rang* : il n'existe pas de relation linéaire exacte entre les variables indépendantes. Cette hypothèse est nécessaire pour l'estimation des paramètres du modèle.
3. *Homoscédasticité et absence d'autocorrélation* : la variance de ε_i est constante et finie σ^2 , $Var(\varepsilon_i) = \sigma^2$ Pour tout i et ε_i n'est pas corrélé avec ε_j , $Cov(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$.
4. Les variables explicatives $X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(k)}$ sont des grandeurs certaines. elles ne représentent 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 :
5. *Distribution normale* : les fluctuations aléatoires sont distribuées normalement $\varepsilon_i \sim N(0, \sigma^2)$. Cette hypothèse n'est pas indispensable pour l'analyse.

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 Y_i sont distribuées normalement et indépendamment avec :

Une moyenne : $E(Y_i) = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \dots + \beta_k X_i^{(k)}$ une variance: $Var(Y_i) = \sigma^2$

Remarque

Précisons également que l'estimation et les testages des paramètres du modèle de régression ne sont possibles en autant que : Le nombre d'observations excède le nombre de paramètres à estimer $n > k + 1$.

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

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'ajuster 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 droite de régression empirique :

$$E(Y_i) = \beta_0 + \beta_1 X_i.$$

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 englobent les vraies valeurs, avec $E(\text{estimateur}_{(i)}) = \beta_i$. Plusieurs méthodes peuvent être utilisées pour l'estimation, mais la plus efficace est celle des moindres carrés qui donne la droite de régression la plus proche aux données :

Estimation par moindres carrés

L'écriture matricielle du modèle linéaire multiple :

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

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

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i^{(1)} - \dots - \beta_k X_i^{(k)})^2 = \|\{Y\} - [X]\{\beta\}\|^2 = (\{Y\} - [X]\{\beta\})^T \cdot (\{Y\} - [X]\{\beta\})$$

$$= \{Y\}^T \{Y\} - 2\{\beta\}^T [X]^T \{Y\} + \{\beta\}^T [X]^T [X] \{\beta\}$$

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

$$[X]^T\{Y\} - [X]^T[X]\{\beta\} = 0$$

Dont la solution correspond bien à un minimum car la matrice Hessien $2[X]^T[X]$ est semi définie positive. Nous faisons l'hypothèse supplémentaire que la matrice $[X]^T[X]$ est inversible, c'est-à-dire que la matrice est de rang $(k + 1)$ et donc qu'il ne existe pas de colinéarité entre ses colonnes. En pratique, si cette hypothèse n'est pas vérifiée, il suffit de supprimer des colonnes de $[X]$ 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 :

$$\{b\} = ([X]^T[X])^{-1}[X]^T\{Y\}$$

Et les valeurs ajustées (ou estimées, prédites) de $\{Y\}$ ont pour expression :

$$\{\hat{Y}\} = [X]\{b\} = [X]([X]^T[X])^{-1}[X]^T\{Y\} = [H]\{Y\}$$

Où : $[H] = [X]([X]^T[X])^{-1}[X]^T$ est appelée « hat matrix » ; elle met un chapeau à $\{Y\}$, géométriquement. C'est la matrice de projection orthogonale dans R^n sur les sous-espaces $Vect(X)$ engendrée par les vecteurs colonnes de $[X]$.

On note :

$$\{e\} = \{Y\} - \{\hat{Y}\} = \{Y\} - [X]\{b\} = ([I] - [H])\{Y\}$$

Le vecteur des résidus ; c'est la projection de $\{Y\}$ sur les sous-espaces orthogonales de $Vect(X)$ dans R^n .

Le système d'équation devient 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 logiciel de calcul (MATLAB).

5.2.2. Régression non linéaire

a) Modèle de régression non linéaire multiple

Malgré sa flexibilité, le modèle linéaire ne permet pas de modéliser certaines formes fonctionnelles. La forme générale d'un modèle de régression non linéaire est :

$$Y_i = h(X_i, \beta) + \varepsilon_i$$

Où :

$h(X_i, \beta)$, C'est une fonction de régression non linéaire, le terme non linéaire ici est appliqué aux paramètres β_i et non pas aux variables du modèles. Le modèle linéaire en constitue un cas particulier. Certains modèles dont l'apparence est non linéaire deviennent linéaires après une transformation par application de certains algorithmes.

a) Conditions d'application du modèle non linéaire multiple

Il est nécessaire de définir plus formellement un modèle de régression non linéaire incluant le modèle linéaire comme un cas particulier. Le modèle non linéaire s'appuie sur les hypothèses suivantes :

1. *Forme fonctionnelle* : La moyenne conditionnelle de Y_i sachant X_i est :

$$E[Y_i/X_i] = h(X_i, \beta) \quad i = 1, \dots, n$$

Où : $h(X_i, \beta)$ est une fonction deux fois continûment différentiable.

2. *Les paramètres du modèle sont identifiables* : le vecteur des paramètres dans le modèle est identifiée (estimable) s'il n'existe aucun paramètre non nul $\beta^0 \neq \beta$ tel que :

$h(X_i, \beta^0) = h(X_i, \beta)$ Pour tout X_i . Cette hypothèse correspond à la condition de plein rang dans le modèle linéaire.

3. *Nullité de la moyenne de la perturbation* : l'hypothèse implique

$$Y_i = h(X_i, \beta) + \varepsilon_i \quad i = 1, \dots, n$$

Avec $E[\varepsilon_i \setminus h(X_i, \beta)] = 0$. Cette condition énonce que la *ième* perturbation est non corrélée avec la moyenne conditionnelle de l'ensemble des observations. Elle est différente de l'hypothèse habituelle de non-corrélation entre les perturbations et les variables exogènes.

4. Comme dans le cas du modèle linéaire, on suppose que :

$$E[\varepsilon_i^2/h(X_i, \beta), i = 1, \dots, n] = \sigma^2 \quad \text{Une constant finie, et l'absence d'autocorrélation : } E[\varepsilon_i, \varepsilon_j/h(X_i, \beta), h(X_j, \beta), j = 1, \dots, n] = 0 \quad \text{pour tout } i \neq j.$$

5. *Modèle de probabilité sous jacent* : il existe une distribution de probabilité bien définie de ε_i . On suppose que ce processus génère un échantillon de variables aléatoires ε_i non corrélées, identiquement distribuées, de moyenne nulle et de variance σ^2 conditionnelle à $h(X_j, \beta)$.

b) Détermination de l'équation de régression : calcul des estimateurs

La minimisation de la somme des carrés des résidus est un problème courant de l'optimisation non linéaire. Nous avons procédé à des artifices de calculs afin de le ramener à un modèle linéaire. Cependant, pour certaines spécifications de modèles il s'avère impossible, même sous certaines hypothèses, d'utiliser les MCO (moindres carrés ordinaires). L'inspiration générale des techniques d'estimation non linéaires sont les mêmes quel que soit le type d'algorithme par exemple: l'Algorithme de Gauss-Newton.

Il s'agit des méthodes itératives dans lesquelles l'équation non linéaire est linéarisée à l'aide d'un développement limité de Taylor (à partir d'un jeu de valeurs de coefficients initiaux). Les MCO sont alors utilisés sur cette équation linéaire afin d'estimer des nouveaux coefficients. Ces nouveaux coefficients permettent, à l'aide d'un nouveau développement limité, de procéder à une nouvelle linéarisation. La procédure est arrêtée lorsque les coefficients sont relativement stables d'une itération à l'autre.

Pour que ces méthodes soient efficaces il faut que le point initial soit proche de la solution optimale car les algorithmes ne convergent pas rapidement.

Algorithme de Gauss-Newton :

Soit le modèle non linéaire : $Y = h(X, \beta) + \varepsilon$; (L'indice de l'observation i est supprimé pour simplifier la notation).

Où : X est un vecteur de variables explicatives, β est un vecteur de dimension $(K + 1)$ de paramètres à estimer. Sous les hypothèses classiques concernant ε_i , l'estimateur MCO de β qui minimise la somme des carrés des résidus :

$$S(\beta) = (Y - h(X, \beta))^T \cdot (Y - h(X, \beta))$$

Nous avons $(K + 1)$ conditions du premier ordre $\frac{\partial S}{\partial \beta} = \mathbf{0}$ soit :

$$\frac{\partial S}{\partial \beta} = -2 \frac{\partial h(X, \beta)}{\partial \beta} * (Y - h(X, \beta))$$

Avec :

$$\frac{\partial h(X, \beta)}{\partial \beta} = Z(\beta) = \begin{bmatrix} \frac{\partial h(X_1, \beta)}{\partial \beta_0} & \dots & \frac{\partial h(X_1, \beta)}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial h(X_n, \beta)}{\partial \beta_0} & \dots & \frac{\partial h(X_n, \beta)}{\partial \beta_k} \end{bmatrix}$$

A l'aide d'un développement de Taylor au voisinage de β_1 , nous pouvons approximer la i -ème observation :

$$h(X, \beta) = h(X, \beta^1) + \left[\frac{\partial h(X_i, \beta)}{\partial \beta_0} \Big|_{\beta = \beta^1} \dots \frac{\partial h(X_i, \beta)}{\partial \beta_k} \Big|_{\beta = \beta^1} \right] (\beta - \beta^1)$$

$$h(X, \beta) \simeq h(X, \beta^1) + Z(\beta^1) * (\beta - \beta^1)$$

$$\text{Soit : } Y = h(X, \beta^1) + Z(\beta^1) * (\beta) - Z(\beta^1) * (\beta^1) + \varepsilon$$

$$\text{En posant : } \bar{Y}(a^1) = Y - Z(\beta^1) * (\beta) + Z(\beta^1) * (\beta^1)$$

Nous pouvons alors construire un pseudo-modèle linéaire qui est égale à :

$$\bar{Y}(a^1) = Z(\beta^1) * (\beta) + \varepsilon$$

Donc l'estimateur MCO de ce modèle linéaire est donné par :

$$\beta^2 = \beta^1 + (Z(\beta^1)^T * Z(\beta^1))^{-1} Z(\beta^1) (Y - h(X, \beta^1))$$

Ce qui donne $(k + 1)$ nouvelles valeurs pour le vecteur $\beta = \beta^2$. Le processus est arrêté à la p -ième itération, lorsque nous observons une relative stabilité des coefficients ainsi estimés : $\hat{\beta} = \beta^p \simeq \beta^{p-1}$.

Cette méthode assez complexe connaît quelque raffinement par l'utilisation d'un développement limité de Taylor au second ordre. Des logiciels sophistiqués comme (*MATLAB*) permettent très simplement, d'utiliser ces méthodes d'estimation des modèles non linéaires.

5.2.3. 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 dépendante et explicative) qui font le sujet d'une étude de régression multiple.
2. Spécifie le modèle de régression multiple que l'on envisage d'analyser.
3. Prélever les observations sur les 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 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 les prévisions requises.

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

5.2 .4. 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 aléatoires (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 aussi 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 observations, les degrés de liberté des sommes de carrés précédentes se décomposent comme suit (Tab.4.1) :

Tab.5.1 : Degrée de liberté pour chaque somme de carrés.

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

c) Tableau de l'analyse de 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 (Tab.4.2).

Tab.5.2 : Analyse de variance en régression multiple.

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

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 ($Var(\beta_i)$) 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 qualité de l'ajustement 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 la 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.

Remarques

- 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.
- 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 *reliés* à la variable dépendante.

Pour éviter ce phénomène, on calcule le coefficient de détermination ajusté :

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

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

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

5.2.5. Tests d'hypothèses et intervalle de confiance

Les tests d'hypothèses constituent un aspect important de l'inférence (induction) statistique. Le principe général d'un test d'hypothèse peut s'énoncer comme suit :

a) Principe d'un test d'hypothèse

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 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 est 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 du hasard de l'échantillonnage. Les distributions d'échantillonnage d'une moyenne et vont être utiles dans l'élaboration d'un test statistique.

b) Concepts important 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 d'hypothèse

Un test d'hypothèse (ou test statistique) est une démarche qui a pour but de fournir une règle de décision, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques. Les hypothèses statistiques qui sont envisagées à priori s'appellent des hypothèses nulles et des hypothèses alternatives.

Hypothèse nulle (H_0) et hypothèse alternative (H_1)

L'hypothèse selon laquelle on fixe à priori un paramètre de la population à une valeur particulière s'appelle l'hypothèse nulle et elle est notée H_0 . N'importe quelle autre hypothèse Diffère de l'hypothèse s'appelle l'hypothèse alternative (ou contre hypothèse) est 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 considérée comme vraisemblable. C'est 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 expérimental), nous considérons alors l'hypothèse altérative H_1 comme vraisemblable plutôt que H_0 .

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 basée 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 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) eut 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 H_0 on favorisera alors l'hypothèse H_1 soit à la région de non rejet de H_0 de on favorisera alors l'hypothèse H_0 .

Remarques

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

c) La démarche à suivre dans l'élaboration d'un test d'hypothèse

Nous adopterons qu'en convention de travailler avec l'écart réduit (variable aléatoire dont on connaît les valeurs tabulées de la loi) qui sera distribuée, 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-types de la statistique qui convient au test :

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

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...) .Un test d'hypothèse comporte diverses étapes :

1. Formuler l'hypothèse nulle H_0 et l'hypothèse alternative H_1 .
2. Fixer d'avance (avant la réalisation du sondage) le seuil de signification alpha c'est-à-dire spécifié 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 e 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 déduisant sa distribution d'après les conditions d'application.
5. Adopter une règle de décision qui conduira au rejet ou non 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 et conclusion .comparer la valeur numérique obtenue pour l'écart réduit avec la règle de décision adopter en 5.décidé entre les deux hypothèses formulées en 1, et conclure.

d) Test 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 chaque terme du modèle) est significative ? Le test correspondant 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.

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

- **Comment déterminer si la régression est significative dans son ensemble ?**

Dans un modèle de régression multiple, disons un modèle comportant k variables explicatives $(X^{(1)}, X^{(2)}, \dots, X^{(k)})$.

$$Y_i = h(X_i, \beta) + \varepsilon_i \quad i = 1, \dots, n$$

$h(X_i, \beta)$, fonction de régression linéaire ou non linéaire aux paramètres β_j .

On veut tester si la régression est significative dans son ensemble. Les hypothèses nulle et alternative 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 test on a recours au carré moyen d'un tableau d'analyse de variance (Tab.4.2). Il s'agit de comparer le carré moyen dû à la régression avec le carré résiduel. la valeur observée du F est donc :

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

On obtient donc un rapport de deux variables.

Sous l'hypothèse $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, on démontre que la qualité :

$F = \frac{CMR}{CM_{RES}}$ est distribuée selon la loi de *Fischer* 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 : $Y_i = h(X_i, \beta) + \varepsilon_i$

Hypothèses : $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

H_1 : 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

- 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.
- 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 : $P(F \geq F_{observée})$. Si cette valeur 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 $P(F \geq F_{observée}) < \alpha$, on rejette H_0 .

On considère que la valeur observée pour le F ne permet pas de supporter l'hypothèse de H_0 . Cette probabilité donne en quelque sorte le degré de convection que la décision prise est bonne. Si ce degré de convection 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 nous nous somme fixé.

- L'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 $H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$.

On a lors :

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

Qui est distribué selon la loi de *Fisher* avec k et $(n - k - 1)$ degrés de liberté.

- **Comment savoir si contribution marginale de chaque variable explicative (ou chaque terme du modèle) est significative ?**

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 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 .

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 (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.

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 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 régression multiple (à moins de faire appel à l'approche matricielle pour traiter les notions de la régression multiple).

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

Test sur la contribution marginale de chaque variable explicative : (Test de signification sur chaque paramètre β_j)

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 test avec le rapport *F de Fisher* pour l'ensemble des variables explicatives ne permet pas de juger si certains coefficients de régression β_j s'avèrent non significatifs. Ce test permettra donc de juger 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.

Indiquons comment s'exécute *le test de Student*, 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.

Pour déterminer si la contribution marginale de chaque variable explicative X_j est significative dans le modèle de régression multiple:

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

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* $(n - k - 1)$ 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) Intervalle de confiance sur β_j

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

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

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

CHAPITRE VI
RESULTATS DES ANALYSES ET
DISCUSSION

6.1. INTRODUCTION

Ce chapitre présente les résultats obtenus dans le cadre du développement d'une loi empirique servant de l'estimation de la durée de phase forte du mouvement sismique sur un sol rocheux. Pour ce faire, nous avons proposée deux modèles linéaire et non linéaire reliant la durée aux différents paramètres (moment magnitude, distance plus proche à la faille et les conditions de site).

Nous avons suivi une méthodologie de régression multiple pour l'évaluation des coefficients des variables indépendantes et pour chacune des deux formes prédictives. L'évaluation de ces coefficients est faite sur la base d'une banque d'accélérométrie mondiale issue des différents pays. Différents tests statistiques ont été menés afin d'apprécier l'effet de chaque paramètre dans l'explication de la durée de phase forte.

6.2. DESCRIPTION DES VARIABLES INDEPENDANTES

Les principaux paramètres mis en comptes généralement dans l'établissement d'une relation prédictive pour calculer certains indicateurs de mouvements forts (dans notre cas la durée de phase forte) sont associés à la caractérisation de la source sismique, les effets de propagation ainsi que l'effet de site.

6.2.1. La Magnitude

Le mécanisme à la source peut être représenté par un ensemble de forces constituant un double couple dont le paramètre caractéristique est M_0 , le moment sismique (Brune, 1976). Ce paramètre est 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, le moment sismique à un sens physique plus réaliste pour représenter l'énergie à la source puisqu'il est directement lié aux caractéristiques de la zone de rupture sur la faille sismique. Pour des raisons de simplification. Les ingénieurs préfèrent caractériser un séisme par sa magnitude pour l'évaluation de niveau sismique d'un site. Plusieurs échelles de magnitude existent (§.1.7.2). Nous avons utilisé le moment magnitude M_w (Hanks and Kanamori 1979), c'est une échelle ouverte et donne une bonne représentation de l'énergie sismique puisqu'il est directement lié au moment sismique.

6.2.2. La Distance source-site

L'utilisation de la distance épacentrale et hypocentrale dans la formulation des lois prédictives des mouvements forts a un avantage, c'est qu'elles peuvent être trouvées

facilement, cependant leur inadéquation a été convenu, la théorie la plus réaliste est que, les ondes sismiques qui affectent beaucoup le site arrive d'un point plus proche au site plutôt que de l'hypocentre. Nous avons utilisé la distance la plus proche à la faille notée R_{rup} (Boore et al., 1978), parce qu'elle est beaucoup significatif que d'autres mesures de distance (§ 1.7.1), si on sait que la faille sismique peut s'étendre à des dizaines de kilomètres et parce qu'elle porte un sens physique plus réaliste vis-à-vis des dégâts observés en surface.

6.2.3. Les conditions locales de site

La classification du site fait intervenir des critères géologiques descriptifs et /ou un paramètre mathématique qui rend compte de la rapidité avec laquelle les ondes de cisaillement (les plus destructives) traversent le milieu sur une profondeur donnée. Usuellement, le critère mathématique est privilégié et on le calcul sur les 30 premiers mètres de profondeur du sol. Plus précisément, on calcul la moyenne harmonique des ondes de cisaillement sur cette épaisseur (V_s^{30}). la prise en compte de l'effet de site dans les équations prédictives du mouvement du sol fait intervenir des variables S_1, S_2 liées à la classification géologique du site elles prennent la valeur 0 pour les sols durs (rocheux) et (1,0 respectivement) pour les sols raide (intermédiaire) et (0,1) pour les sols mou.

6.3. DISCUSSION DES RESULTATS POUR LA REGRESSION LINEAIRE

6.3. 1. Spécification du modèle.

Le modèle linéaire que nous avons utilisé pour expliquer les fluctuations des valeurs de la durée de phase forte (D) adopte le formalisme suivant :

$$f(D) = f_1(M_w) + f_2(R_{rup}) + f_3(S) + \varepsilon \quad (Eq. 6.1)$$

Où :

$f(D)$, représente la fonction du paramètre à prédire (la durée).

$f_1(M_w)$, est la fonction du moment magnitude ; $f_2(R_{rup})$ représente la fonction de la distance et $f_3(S)$, fonction des conditions de site.

Pour l'établissement de l'équation prédictive, on considère que : $f(D) = \log(D)$, en raison de la distribution normale de $\log(D)$ autour de sa médiane [21].

Concernant $f_1(M_w)$, la magnitude quadratique a été introduit (Boore et al., 1993 ; Sadigh, 1993), mais le coefficient correspondant est statistiquement non significatif. La fonction $f_1(M_w)$ est représentée par :

$$f_1(M_w) = \beta_0 + \beta_1 M_w$$

Pour prendre en considération l'atténuation géométrique, l'amortissement des matériaux et la diffraction (atténuation anélastique) du milieu traversé par les ondes sismiques, la fonction $f_2(R_{rup})$ est formulée comme suit:

$$f_2(R_{rup}) = \beta_2 \text{Log}(R_{rup}) \quad (\text{Eq. 6.2})$$

Où :

Le terme de (Eq. 6.2) représente l'atténuation géométrique. Finalement la catégorie de site sur lequel on prédit le mouvement a son importance et la prise en compte de ses effets est représentée par la fonction dont l'expression :

$$f_3(S) = \beta_3 S_1 + \beta_4 S_2$$

Les variables S_1 et S_2 sont liés à la classification géologique du site. Ils prennent la valeur 0 pour sols rocheux, (1,0) pour les sols intermédiaires et (0,1) pour les sols meubles.

Pour notre cas, on est en présence d'un sol rocheux, et la fonction $f_3(S)$ prend une valeur nulle dans la forme fonctionnelle.

Le modèle à quatre variables explicatives que nous avons adopté pour expliquer les fluctuations des valeurs de la variable dépendante et interpréter chaque composante du modèle tient la forme suivante :

$$\text{Log}(D) = \beta_0 + \beta_1 M_w + \beta_2 \text{Log}(R_{rup}) + (\beta_3 S_1 + \beta_4 S_2) \pm \sigma \quad (\text{Eq. 6.3})$$

Où :

D , est le paramètre d'intérêt (la durée de phase forte), M_w : est le moment magnitude,

R_{rup} : représente la mesure de la distance (la distance la plus proche à la faille),

S_1, S_2 : représente les conditions géologiques.

β_0 : Théoriquement ce paramètre représente la moyenne des $\text{log}(D)$ lorsque la valeur de variation de chaque variable explicative est égale à zéro. Toutefois, étant donné le champ de variation de chaque variable, dans 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 .

Remarque

L'estimation des paramètres du modèle (Eq. 6.3) revient à résoudre un système de n équations linéaire, l'utilisation de la méthode des moindres carrés linéaires s'avère inaccessible sans avoir recours à un logiciel de calcul (MATLAB).

6.3.2. Discussion des résultats

a) 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 $\log(D)$ 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 .

En régression linéaire multiple, le coefficient de détermination multiple R^2 , nous permet évaluer la qualité de l'ajustement linéaire de l'équation de régression entre la variable dépendante et l'ensemble des variables explicatives.

Tab.6.1 : Analyse de variance en régression linéaire.

<i>Statistique de la régression</i>					
<i>Coefficient de détermination multiple R^2</i>				0.88	
<i>Erreur-Type</i>				0.373	
<i>Observations</i>				184	
<i>Analyse de variance</i>					
<i>Source de variance</i>	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>Statistique F</i>	<i>Valeur critique de F</i>
<i>Expliquée par la régression</i>	4	$SCR = 83.841$	$CMR = 20.960$	300.594	4.41
<i>Résiduelle</i>	180	$SC_{res} = 25.242$	$CM_{res} = 0.139$	-	-
<i>Total</i>	184	$SCT = 109.083$	-		

Détermination de la variance et du coefficient R^2

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

$$S^2 = CM_{res} = 0.139$$

L'écart-type des résidus :

$$S = 0.373$$

Le coefficient de détermination multiple :

$$R^2 = \frac{SCR}{SCT} = 0.88$$

On voit que le coefficient de détermination multiple est acceptable et qui a une valeur autour de 88%. Cette valeur nous permet de dire que les variables explicatives dans le modèle ont bien expliqué la variable dépendante.

b) Tests d'hypothèses**- Test de signification de la régression dans son ensemble (test de Fisher)**

On veut tester l'hypothèse selon laquelle les variables explicatives n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\text{Log}(D)$.

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

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

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

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

Quotient des carrés moyens:

$$F = \frac{CMR}{CM_{res}}$$

Pour $k = 4$, $(n - k - 1) = 179$ et au seuil de signification, $\alpha = 0.05$, on obtient de la table de Fisher la valeur critique de F .

$$F_{0.05;4,179} = 4.41$$

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

Du tableau d'analyse de variance (Tab.6.1), on a : $CMR = 20.960$, $CM_{res} = 0.139$

La valeur observée pour le rapport F (Quotient des carrés moyens) est donc : 300.594

Puisque $F > 4.41$ nous rejetons l'hypothèse H_0 et favorisons H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil 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 qui est significativement différent de zéro, d'où la signification de l'ensemble du modèle.

- Test de contribution marginale de chaque variable explicative (test de Sudent)

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. Il 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.

Pour tester la contribution marginale de chaque variable dans le modèle, nous avons recours au tableau des estimateurs (Tab.6.2). Ce tableau regroupe les estimateurs de tous les coefficients, et leurs écarts types ainsi que le coefficient de *Student* correspondant.

Tab.6.2 : Statistiques des estimateurs en régression linéaire.

<i>Statistique des estimateurs</i>					
<i>Coefficients</i>	<i>Estimation</i>	<i>Erreur type</i>	<i>Statistique T</i>	<i>Intervalle de confiance</i>	
β_0	-4.701	0.360	-13.050	-5.412	-3.99
β_1	1.034	0.056	18.152	0.921	1.145
β_2	0.192	0.036	5.322	0.120	0.262
β_3, β_4	0.00	0.00	-	-	-

Le test se fait comme suit : Au seuil de signification $\alpha = 0.05$, Il s'agit alors de tester les hypothèses suivantes :

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

$$H_1: \beta_j \text{ est différent de zéro.}$$

Sous l'hypothèse H_0 , l'écart réduit devient $b_j/S(b_j)$ qui est distribué selon la loi de *Student* avec $(n - k - 1)$ degrés de liberté.

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

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

$(n - k - 1 = 179)$, on est en présence d'un grand échantillon, donc d'après la table de la loi normale centrée réduite on trouve : $t_{0,025;179} = 1.96$.

En analysant le tableau des estimateurs, on peut dire :

Toutes les variables ont une signification marginale dans l'explication de la durée de phase forte sauf :

- Les conditions de site sont sans signification particulière puisqu' on est en présence d'un sol rocheux (les coefficients associés aux variables S_1 et S_2 ont des valeurs nulles).

La forme finale de l'équation prédictive adoptée pour la durée de phase (D) prend la forme :

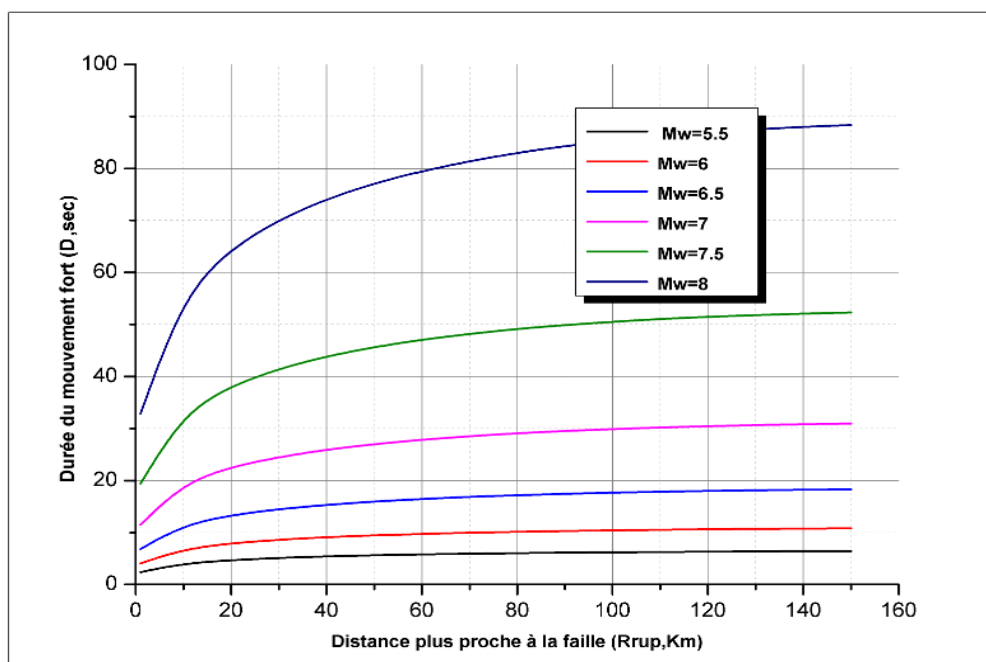
$$\text{Log}(D) = -4.701 + 1.034 * M_w + 0.192 * \text{Log}(R_{rup}) \pm 0.373$$

6.3.3 Effet des paramètres environnementaux sur la durée de phase forte :

Les figures (6.1 et 6.2) montrent respectivement l'effet de la distance la plus proche à la faille (R_{rup}) et le moment magnitude (M_w) sur la durée de phase forte (D) pour la formulation adoptée dans le cas du modèle linéaire.

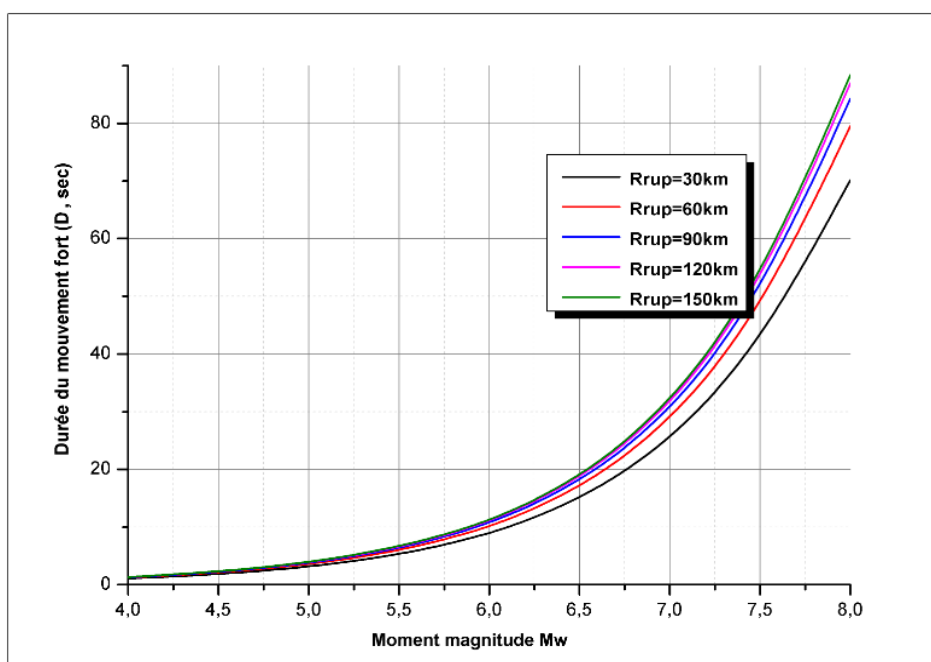
D'après la figure (6.1) On constate que :

- La durée augmente généralement avec la distance en champs proche ($R_{rup} \leq 20$), elle est d'autant plus longue à mesure que la taille de l'événement augmente, ce qui implique qu'elle a une forte dépendance avec les dimensions de la faille. Donc elle sera très proche à la durée de la source. Pour les champs lointains, pour une magnitude donnée la durée croît légèrement, elle augmente de 2 à 3 secs.
- De plus, la durée cesse d'augmenter pour des grandes distances, ce qui explique un effet de saturation avec la distance ceci est dû au fait l'énergie libérée à partir d'un point sur la faille sismique va s'atténuer avant d'arriver au site.



**Fig.6.1 : Effet de la distance, R_{rup} sur la durée de phase forte, D
(Cas du modèle linéaire).**

- la figure (6.2) montre que la durée de phase forte est étroitement liée à la taille de l'évènement sismique. Nous avons constaté que pour des magnitudes inférieures 5,5, un effet de saturation avec la distance a été perçue, la durée croît d'autant moins vite, avec l'augmentation de la magnitude, indépendamment de la distance à la faille, elle est plus longue pour les grandes magnitudes.



**Fig.6.2 : Effet de la magnitude, M_w sur la durée de phase forte, D .
(Cas du modèle linéaire).**

6.4. DISCUSSION DES RESULTATS DE LA REGRESSION NONLINEAIRE

6.4. 1. Spécification du modèle.

Le modèle non linéaire que nous avons utilisé pour expliquer les fluctuations des valeurs de la durée de phase forte (D) adopte le formalisme suivant :

$$f(D) = f_1(M_w) + f_2(R_{rup}, M_w) + f_3(S) + \varepsilon \quad (Eq. 6.4)$$

Où :

$f(D)$, est la fonction du paramètre à prédire (durée de phase forte),

$f_1(M_w)$, représente la fonction du moment magnitude.

$f_2(R_{rup}, M_w)$, Cette fonction représente l'effet combinée des deux paramètres (moment magnitude M_w et la distance la plus proche à la faille, R_{rup}) et $f_3(S)$ représente la fonction des conditions géologiques et ε , représente l'erreur type.

De même que pour le modèle linéaire, on trouve :

$$\begin{aligned} f(D) &= \log(D), \quad f_1(M_w) = \beta_0 + \beta_1 M_w \\ f_2(R_{rup}, M_w) &= \beta_2 \text{Log}(R_{rup}^2 + \Delta^2)^{1/2} + \beta_3 R_{rup} \end{aligned} \quad (Eq. 6.5)$$

Où:

Le premier terme de (Eq. 6.5) représente l'atténuation géométrique, le terme (Δ) regroupe tous les facteurs qui tendent à limiter le mouvement à proche distance (l'effet de saturation avec la distance), ce qui exprime le fait que l'amplitude du mouvement à proche distance de la source ne rend que partiellement compte du rayonnement total du phénomène de rupture sismique d'une zone d'étendue finis (Campbell, 1981, Abrahamson et Silva, 1997). Cet effet de saturation en distance se traduit dans les calculs par une pente d'atténuation géométrique (Δ) qui diminue à proche distance.

Pour tenir compte de l'effet de saturation avec la magnitude (Compbell, 1985 ; Sadigh, Bozorgina, 1994), ont analysé la dépendance du terme (Δ) avec la magnitude : ($\Delta = a \times 10^{b \times M_w}$), Les paramètres a et b seront déterminés à travers une analyse de régression non linéaire. De plus, le modèle de prédiction rend également compte d'une atténuation anélastique ($\beta_5 R_{rup}$) dans (Eq. 6.5), qui traduit l'absorption énergétique du milieu traversé par les ondes sismiques.

Finalement, la fonction $f_3(S)$ représente les différents effets d'amplification du sol. elle est décrite par :

$$f_3(S) = \beta_4 S_1 + \beta_5 S_2$$

Les variables S_1 et S_2 sont liés à la classification géologique du site. Ils prennent la valeur 0 pour sols rocheux, (1,0) pour les sols raide (intermédiaire) et (0,1) pour les sols meubles. La forme finale de l'équation (Eq. 6.4) devient :

$$\text{Log}(D) = \beta_0 + \beta_1 M_w + \beta_2 \text{Log}(R_{rup}^2 + \Delta^2)^{1/2} + \beta_3 R_{rup} + (\beta_4 S_1 + \beta_5 S_2) \pm \sigma \quad (\text{Eq. 6.6})$$

Où :

D , est le paramètre à prédire (la durée de phase forte) ; M_w , le moment magnitude

R_{rup} : représente la mesure de la distance (la distance la plus proche à la faille),

S_1, S_2 : représente les conditions géologiques et σ : représente l'écart type du $\text{Log}(D)$.

Remarque : le calcul des estimateurs des coefficients du modèle (Eq. 6.6) dans le cadre de la régression non linéaire s'avère difficile, le modèle n'est linéaire par aucune transformation. un développement de Taylor de cette fonction autour d'un point initial produira une équation intrinsèquement linéaire qui peut être estimée par moindres carrés.

6.4.2. Discussions des résultats

a) Analyse de variance

De même que pour la régression linéaire multiple, le tableau d'analyse de variance dans le cadre de régression non linéaire (Tab.6.3) permet de tester la significativité de l'ensemble des variables explicatives sur la variable dépendante. Il permettra aussi en déduire le coefficient de détermination multiple R^2 . Ce coefficient permet d'évaluer la qualité de l'ajustement de l'équation de régression non linéaire entre la variable dépendante et l'ensemble des variables explicatives.

Tab.6.3 : Analyse de variance en régression non linéaire.

<i>Statistique de la régression</i>					
<i>Coefficient de détermination multiple R^2</i>				0.77	
<i>Erreur-Type</i>				0.359	
<i>Observations</i>				184	

<i>Analyse de variance</i>					
<i>Source de variance</i>	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>Statistique F</i>	<i>Valeur critique de F</i>
<i>Expliquée par la régression</i>	6	$SCR = 1814.982$	$CMR = 383,981$	2976,596	4.41
<i>Résiduelle</i>	178	$SC_{res} = 23,016$	$CM_{res} = 0,129$	-	-
<i>Total</i>	184	$SCT = 2326,901$	-		

-Détermination de la variance et du coefficient R^2

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

$$S^2 = CM_{res} = 0,129$$

L'écart-type des résidus :

$$S = 0.359$$

Le coefficient de détermination multiple :

L'utilisation du coefficient de détermination R^2 pour évaluer la qualité de l'ajustement du modèle non linéaire décrit dans (Eq.6.6) entre la variable dépendante et l'ensemble des variables explicatives est insuffisant. Pour apprécier pleinement l'utilité de ce modèle de régression, il faut lui associer, l'écart-type des résidus et les marges d'erreurs dans l'estimation des paramètres β_j . D'après le tableau d'analyse de variance (Tab.6.3) le coefficient de détermination multiple est autour de 77 %. le modèle semble bon, l'écart type des résidus est environ de 0.359, c'est une valeur un peu grande au sens pour assurer des estimations avec une marge d'erreur acceptable.

b) Tests d'hypothèses**- Signification de la régression dans son ensemble (test de Fisher)**

On veut tester l'hypothèse selon laquelle toutes les variables explicatives présentes dans le modèle non linéaire (Eq.6.6) n'apportent, dans l'ensemble, aucune contribution significative dans l'exploitation des fluctuations de valeurs de $\text{Log}(D)$.

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

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_5 = 0$$

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

Pour effectuer le test, nous aurons donc recours au rapport des carrés moyens: Ou quotient des carrés moyens:

$$F = \frac{CMR}{CM_{res}}$$

Pour $k = 6$, $(n - k - 1) = 177$ et au seuil de signification $\alpha = 0.05$, on obtient de la table de Fisher la valeur critique de F.

$$F_{0.05;6,177} = 4.41$$

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

D'après le tableau d'analyse de variance (Tab.6.3) : $CMR = 383,981$, $CM_{res} = 0,129$

La valeur observée pour le rapport F est donc : 2976,596.

En appliquant la règle de décision, on constate que $F > 4.41$ ce qui nous conduit à rejeter l'hypothèse H_0 et favoriser H_1 .

Selon les résultats du test, il apparaît vraisemblablement, au seuil 0.05, qu'au moins une des variables explicatives sinon l'ensemble, apportent une contribution significative dans l'explications du nombre de pièces assemblées, d'où la signification d'ensemble de la régression.

- **Test de contribution marginale de chaque variable explicative (test de Student)**

Le test de *t-Student* dans le cadre d'une régression non linéaire permettra 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.

Pour test la contribution marginale de chaque variable explicative dans l'équation de régression multiple est significative nous avons recours au tableau des estimateurs (Tab.6.4) qui donne les valeurs des coefficients et leurs écart type ainsi que le coefficient de Student correspondant, par lequel on détermine la contribution marginale de chaque variable.

Tab.6.4 : Statistique des estimateurs en régression non linéaire.

<i>Statistique des estimateurs</i>					
<i>coefficients</i>	<i>Estimation</i>	<i>Erreur type</i>	<i>Statistique t</i>	<i>Intervalle de confiance</i>	
β_0	24.677	5,989	4.120	12.859	36.496
β_1	1.996	0,517	3.863	0.977	3.016
β_2	-7.831	1,599	4.900	-10.986	-4.676
a	-35.115	10,398	3.375	-55.632	-14.598
b	0.065	0,025	2.682	0.017	0.115
β_3	0.034	0,01	3.420	0.016	0.052
β_4, β_5	0.000	0.000	-	-	-

Le test se fait comme suit :

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

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

$$H_1: \beta_j \text{ est différent de zéro.}$$

Sous l'hypothèse H_0 , l'écart réduit devient $b_j/S(b_j)$ qui est distribué selon la loi de *Student* avec $(n - k - 1)$ degrés de liberté.

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

($n - k - 1 = 177$), on est en présence d'un grand échantillon, donc d'après la table de la loi normale centrée réduite on trouve : $t_{0.025; 177} = 1.96$.

En analysant les tableaux des estimateurs, on peut dire :

Toutes Les variables ont une signification marginale dans l'explication de la durée de phase forte sauf : Les conditions de site sont sans signification particulière puisqu' on est en présence d'un sol rocheux (les coefficients associés aux variables S_1 et S_2 ont des valeurs nulles).

La forme finale de l'équation prédictive :

$$\text{Log}(D) = 24.677 + 1.996 * M_w - 7.831 * \text{Log}(R_{rup}^2 + \Delta^2)^{1/2} + 0.034 * R_{rup} + 0.359$$

$$\text{Où: } (\Delta = -35.115 \times 10^{0.065 \times M_w})$$

6.4.3 .Effet des paramètres environnementaux sur la durée de phase forte :

Les figures (6.3) et (6.4) montrent l'effet du moment magnitude, M_w et la distance la plus proche à la faille, R_{rup} sur la durée de phase forte pour la formulation adoptée dans le cas du modèle non linéaire.

A partir de ces figures, on voit que :

- En champs proche la durée augmente généralement avec la distance. En passant d'une magnitude à une autre, généralement une augmentation de 2 à 3 sec a été observée. Ceci est vrai pour des événements à faible magnitude. Une augmentation de la taille de l'évènement à une vitesse de dislocation généralement constante ceci entraîne une forte augmentation de la durée.

Ce phénomène est d'autant plus important en champs proche qu'en champs lointain. Nous avons aussi remarqué que la durée tend à diminuer pour les grandes distances à cause de l'atténuation de l'énergie arrivant au site (ou effet d'atténuation à grande distance).

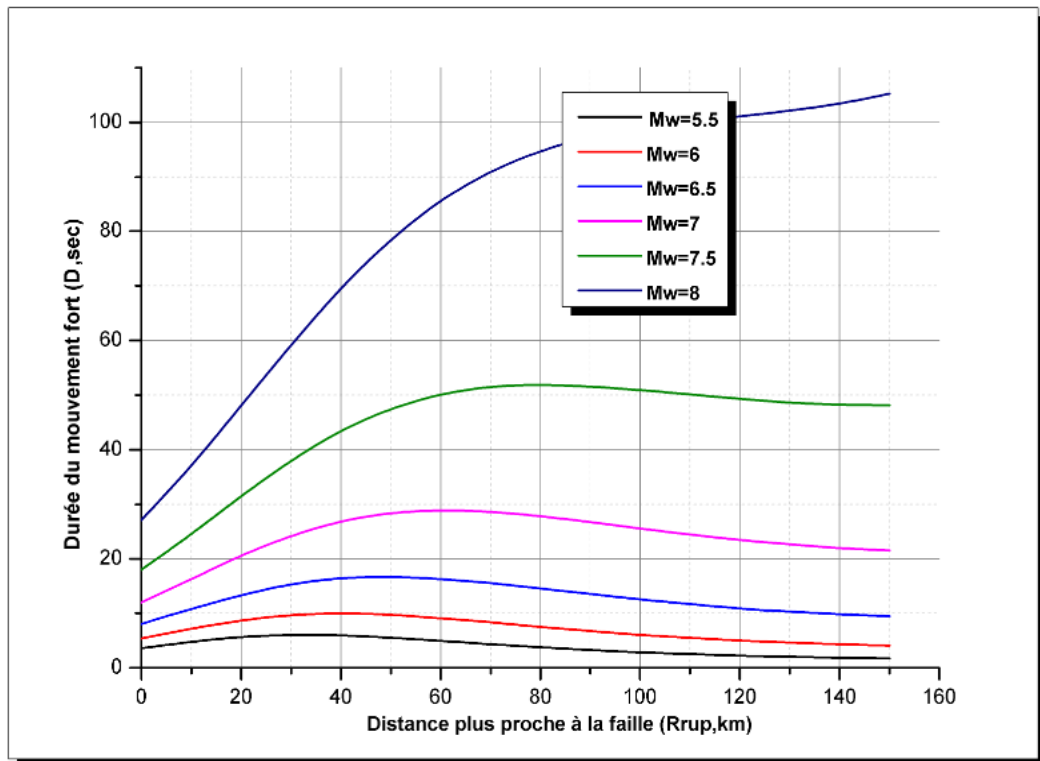


Fig.6.3 : Effet de la distance, R_{rup} sur la durée de phase forte, D (Cas du modèle non linéaire).

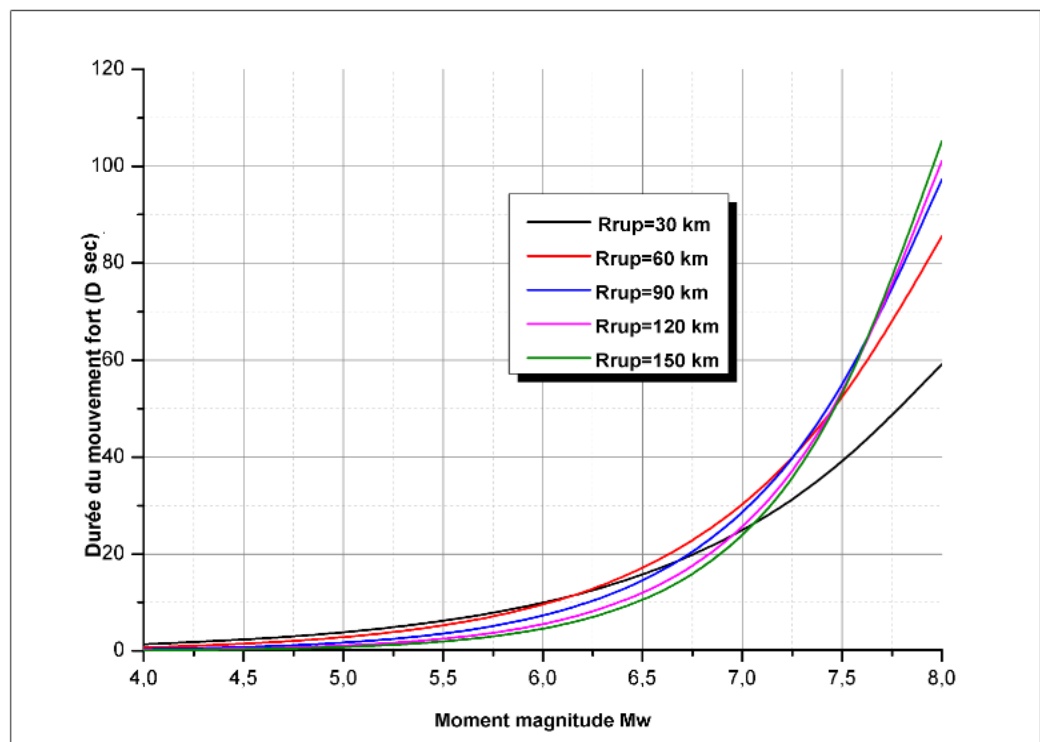


Fig.6.4 : Effet de la magnitude, M_w sur la durée de phase forte, D (cas du modèle non linéaire).

CONCLUSION GENERALE

Actuellement la caractérisation des mouvements sismiques à prendre en compte, pour les installations à risque, repose sur des lois d'atténuation. Ces lois empiriques sont déduites à partir de bases de données d'accéléromètres et servent d'estimation de quelques indicateurs de nocivité d'un séisme dans un lieu donné. Parmi lesquels, la durée de phase forte joue un rôle principal dans beaucoup de problèmes de la sismologie. L'évaluation de la durée a une importance primordiale dans les problèmes liés aux phénomènes de fatigue due à un petit nombre de cycles, de la liquéfaction et du tassement. La durée de phase forte du mouvement sismique s'intéresse beaucoup les ingénieurs pour tester les structures, l'endommagement des structures ductiles est sensible au cyclage ou à la durée de la sollicitation.

Plusieurs définitions ont été proposées dans ce sens pour estimer cette phase forte déterminante pour estimer le potentiel d'endommagement d'un séisme. Au cours de ce travail nous avons utilisé trois définitions dédiées par la littérature scientifique, sont celles proposées par : (McCann-Shah, 1979 ; Trifunac-Brady, 1975 Bolt, 1973).

L'analyse des résultats des valeurs de la durée de phase forte obtenues selon ces trois définitions pour une banque mondiale d'enregistrements accélérométriques, comportant en totale 184 scénarios (magnitude-distance) des composantes d'accélération horizontales représentatifs de l'activité sismiques des régions continentales différentes, nous amène à citer les observations suivantes :

- La durée de Bolt (1973) est très sensible au choix du seuil de d'accélération dans sa définition, généralement la durée de Bolt diminue en champs lointains, ce qui explique un effet d'atténuation de l'énergie sismique à grandes distances.
- Les durées calculées selon les définitions de McCann-Shah et Trifunac-Brady s'étaient relativement proche. Sachant que les deux définitions sont basées sur le même principe d'énergie (énergie portée dans le signal sismique), des grandes valeurs ont été enregistrées en champs lointains. Ce phénomène peut être expliqué en revenant à la définition énergétique de la durée de mouvement fort pour les séismes lointains et en tenant compte de l'effet d'atténuation qui fait uniformiser l'énergie sur toute la durée de séisme. L'énergie d'un séisme lointain, en présence de l'atténuation due à la distance, est généralement concentré dans une longue durée qui correspond à la durée de mouvement fort.

Il est tout naturel que les valeurs de la durée obtenue selon les trois définitions s'étaient pratiquement différentes, ce qui reflète la grande différence entre les théories derrière ces trois méthodes.

Au cours de ce travail, nous avons aussi 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 au site d'enregistrement. Nous avons proposée deux modèles linéaire et non linéaire qui prennent en compte l'effet des trois termes dans l'estimation de la durée de phase forte : un terme lié à durée de l'événement à la source (dépendant de la magnitude, de la surface de rupture, du mécanisme à la source), un terme de propagation des ondes et un terme d'effets de site.

Par ailleurs, l'analyse des lois empiriques développées pour la durée de phase forte obtenues suivant une méthodologie de régression selon ses deux aspects linéaire et non linéaire a fait ressortir de manière claire que :

- L'effet du moment magnitude se traduit par une augmentation de la durée de phase forte. Elle est très longue pour les grandes magnitudes, pour des faibles séismes la durée croit autant moins vite à mesure que la distance à la faille augmente. Ceci peut être traduit par un effet de saturation avec la distance.
- L'effet de la distance la plus proche à la faille se traduit par une augmentation de la durée de phase forte. En champs proche elle est très grandes pour les grandes séismes elle augmente de 10 à 15 sec en passant d'une magnitude à une autre. pour les champs lointains la durée tend à diminuer à cause de l'absorption énergétique du milieu traversé.
- L'analyse concernant la durée de phase forte (D) a montré que celle-ci ne subit pratiquement aucune influence en présence d'un sol rocheux, ca valeur augmente avec la taille de l'événement sismique, en champs proche elle dépend seulement au rayonnement total du phénomène de rupture et diminue pour les grandes distances à cause de l'atténuation anélastique du milieu.
- Des grandes valeurs des durées ont été obtenues, ce qui est en contradiction avec la nature du sol rocheux. ceci peut être interpréter par un choix simpliste du modèle point source (on n'a pas pris en compte des détails de la source sismique).

Les modèles empiriques proposées dans le présent travail pour l'évaluation de la durée de phase forte sur un sol rocheux semblent intéressantes pour l'évaluation du potentiel d'endommagement d'un séisme et même pour l'évaluation du risque sismique dans des régions données.

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 contrôlants 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 en réduire les incertitudes et à produire une meilleure précision d'analyse du risque sismique. Cette travail peut être aussi accomplir à travers l'étude de la contribution par bandes de fréquences à la durée de phase forte.

REFERENCES BIBLIOGRAPHIQUES

- [1] Abrahamson, N., Silva, W. "Summary of the Abrahamson and Silva NGA ground-motion relations".2008,*Earthquake Spectra*, Volume 24, N° 1,pp 67–97.
- [2] Aki, K., Richards. G ."Quantitative seismology: theory and methods".1980, Volume 1, W. H. *Freeman and Co.*
- [3] Anderson .J.G. "Expected shape of regressions for ground motion parameters on rock".2000, *Bulletin of the Seismological Society of America*, Vol. 90 N° .6,pp .43-52.
- [4] Anil K.Chopra. "Dynamics of structures theory and applications to earthquake engineering".1995, PRENTICE -HALL *International series.*
- [5] Arias, A. "A measure of earthquake intensity, in seismic design of nuclear power plants".1970, R. Hansen (Editor) *M.I.T. Press.*
- [6] Atkinson, G. M., Boore. D. M. " Empirical ground motion relations for subduction zone earthquakes and their application to cascadia and other regions".2003, *Bulletin of the Seismological Society of America*, Vol.93, N°.4,pp.1703-1729.
- [7] Baillargeon, G. "Probabilités statistiques et techniques de régression".1989, *les éditions SMG*, Québec.
- [8] Bolt, B. A. "Duration of strong ground motions".1973, *5th World Conference on Earthquake Engineering*, Rome, Italy.
- [9] Bommer, J.J., Martínez-Pereira, A. "The effective duration of earthquake strong motion".1999, *Journal of Earthquake Engineering*, Vol.3, N°.2, pp.127-172.
- [10] Bommer, J.J., Martínez-Pereira, A. "The prediction of strong motion duration for engineering design".1996, *11th World Conference On Earthquake Engineering*, London
- [11] Bommer, J.J., Stafford,P.J. "Empirical equations for the prediction of the significant, bracketed and uniform duration of earthquake ground motion".2009,*Bulletin of the Seismological Society of America*, Vol. 99, N° 6, pp.3217-3233.
- [12] Boore, D.M., Joyner, W.B. "The empirical prediction of ground motion".1982, *Bulletin of the Seismological Society of America*, Vol.72, N°6, pp. 43-60.
- [13] Boore, D.M., Joyner, W.B. "Methods for regression analysis of strong motion data".1993, *Bulletin of the Seismological Society of America*, Vol. 83, No. 2, pp. 469-487.
- [14] Christian, R.M. "Characteristics of the strong ground motion recording during the October 15, 1979 Imperial valley earthquake"1982,*Stanford University, Department of Civil and Environmental Engineering*, Report N°.53.

- [15] Djouabi, H. "Lois prédictives par l'estimation de la réponse sismique et des mouvements sismiques forts".2010, Thèse Magister, *École Nationale Polytechnique*, Alger.
- [16] Dorby, R., Idriss, I.M. "Duration characteristics of horizontal components of strong motion earthquake records". 1978, *Bulletin of the Seismological Society of America*, Vol. 68. N°.5, pp.1487-1520.
- [17] O'Connell, D.R.H., Akai, J.P. "Earthquake ground motion estimation .Earthquake volume of the hazards and disasters series"2002 .*Routledge, Ltd., Academic Press*.
- [18] Hammoutene, M. "Simulation non-stationnaire et bidimensionnelle des enregistrements de tremblements de terre".1995, Thèse d'États, *École Nationale Polytechnique*, Alger.
- [19] Hancock, J., Bommer, J. J. "A state of knowledge review of the influence of strong motion duration on structural damage". 2006, *Earthquake Spectra*, Vol.22, N°. 3, pp. 827–845.
- [20] J.Betbeder-Matibet. "Les phénomènes sismiques, risques et aléas sismiques, prévention parasismique".2003, *HERMES*, 3 volumes.
- [21] Kramer.S.L."Geotechnical earthquake engineering. Civil engineering and engineering Mechanics".1996, *Prentice Hall*.
- [22] McCann, M.W., Shah, H.C."RMS Acceleration and Duration of Strong Ground Motion".1980, *Stanford University, Department of Civil Engineering*, Report No. 46.
- [23] McCann, M.W., Shah, H.C. "Determining Strong-Motion Duration of Earthquake".1979, *Bulletin of the Seismological Society of America*, Vol. 69, N°.4, pp. 1253-1265.
- [24] McGuire ,R. K., Hanks, T. C. "RMS Accelerations and Spectral Amplitudes of Strong Ground Motion during the San Fernando, California, Earthquake".1980, *Bulletin of the Seismological Society of America*, Vol. 70, N°.5, pp.1907-1919.
- [25] Reinoso, E., Ordaz, M. "Duration of Strong Ground Motion during Mexican Earthquakes in Terms of Magnitude, Distance to the Rupture area and Dominant Site Period".2001, *Earthquake Engineering & Structural Dynamic*,Vol.30,pp.653-673.
- [26] Trifunac, M., Brune, J.N."Complexity of energy release during the Imperial Valley, California, earthquake of 1940".1970, *Bulletin of the Seismological Society of America*, Vol.60, N°.1, pp.137-160.
- [27] Trifunac, M. D., Brady, A. G. "A Study on the duration of strong earthquake ground Motion".1975,*Bulletin of the Seismological Society of America*. Vol. 65, N°.3, pp.581-626.
- [28] Trifunac, M.D ., Westermo, B.D. "Dependence of the duration of strong earthquake ground motion on magnitude, epicentral distance, geologic conditions at the recording station

and frequency of motion".1976,*University of Southern California, Department of Civil Engineering* , Report N°.CE 76-02,.

[29] Trifunac, M.D., Novikova, Elena.I."Duration of strong earthquake ground motion: physical basis and empirical equations". 1993, *University of Southern California, Department of Civil Engineering*, Report N°.CE 93-02.

[30] Theofanopulos, N.A., Watabe,M. "A new definition of strong motion duration and comparison with other definitions".1989, *Earthquake Engineering & Structural Dynamic*, Vol. 6, N°.1, pp.111-122.

[31] Vanmarcke, E.H., Lai, S.P. "Strong-Motion Duration and RMS Amplitude of Earthquake Records". ".1980, *Bulletin of the Seismological Society of America*, Vol. 70, N°.4, pp.1293-1207.

[32] Vanmarcke, E. H., Lai, S.P (1977). "Strong motion duration of earthquakes", *M.I.T, Department of Civil Engineering Research*, Report R 77-16.

[33] <http://peer.berkeley.edu/smcat/> PEER. (*Pacific Earthquakes Engineering Research Centre*), *Strong Motion Database*, 2004.

Test de signification sur l'ensemble des paramètres β_j : Test de Fisher

Dans un modèle de régression multiple, disons un modèle comportant k variables explicatives. On veut tester si la régression est significative dans son ensemble. Au seuil de signification α la démarche du test peut se résumer comme suit :

-Modèle :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

-Hypothèses statistiques :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{Au moins un des } \beta_j \neq 0$$

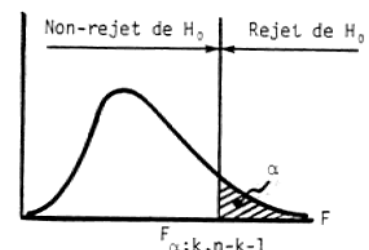
-Rapport des carrés moyens et sa distribution : il s'agit de comparer le carré moyen dû à la régression avec le carré moyen résiduel. La valeur observée du F est donc :

$$F = \frac{CMR}{CM_{RES}}$$

On obtient un rapport de deux variances. Sous l'hypothèse H_0 , on démontre que la quantité F est distribuée selon la loi de Fisher avec k et $(n-k-1)$ degrés de liberté.

-Décision et conclusion : on rejette H_0 si $F \geq F_{\alpha; k, n-k-1}$

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



Test de signification sur chaque paramètre β_j : Test avec le t -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. La démarche du test se résume comme suit : On veut tester si la contribution marginale de chaque variable explicative X_j est significative dans le modèle :

-Modèle : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$

Au seuil de signification α . Il s'agit alors de tester les hypothèses suivantes :

-Hypothèses statistiques :

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

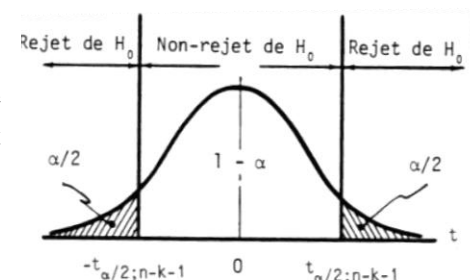
$$H_1 : \beta_j \neq 0$$

Sous H_0 , l'écart réduit devient $t = b_j/s(b_j)$ qui est distribué selon la loi de Student avec $(n-k-1)$ degrés de liberté.

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

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

-Décision et conclusion : Comparer la valeur numérique obtenue pour l'écart réduit avec la règle de décision adoptée et décider entre les deux hypothèses formulées précédemment. Si H_0 est rejeté, la contribution marginale de la variable explicative X_j est significative au seuil α (risque de rejeter tort H_0).



Code de classification	Type du profil de sol selon la valeur moyenne de Vs dans les 30m supérieurs en m/s	Description
UBC (1997)	○ A = > 1500 m/s	- Roche dure
	○ B = 760-1500 m/s	- Roche
	○ C = 360-760 m/s	- Sol très dense
	○ D = 180-360 m/s	- Sol raide
	○ E =< 180 m/s	- Sol mou
USGS (Boore et al, 1994)	○ A => 750 m/s	- Roche dure
	○ B =360-750 m/s	- Roche
	○ C =180-360 m/s	- Sol très dense
	○ D =< 180 m/s	- Sol raide
NEHRP (1994)	○ A = > 1500 m/s	- Roche dure
	○ B = 760-1500 m/s	- Roche
	○ C = 360-760 m/s	- Sol très dense
	○ D = 180-360 m/s	- Sol raide
	○ E =< 180 m/s	- Sol mou
Classification basée sur les caractéristiques géologiques	○ A : Roche. L'instrument se trouve sur roche ($V_s > 600$ m/s) ou autre formation géologique de ce type comportant une couche superficielle d'au plus 5m d'un matériau moins résistant.	
	○ B : Dépôts raides (superficiel). L'instrument se trouve sur/à l'intérieur d'un profil de sol ayant une épaisseur jusqu'à 20m et surmonte une roche.	
	○ C : Dépôts profonds de sable très dense, de gravier ou d'argile très raides sol. ayant une épaisseur d'au moins 20m et surmonte une roche.	
	○ D : Dépôts profonds de sable de densité moyenne, de gravier ou d'argile moyennement raide ayant une épaisseur d'au moins 20m et surmonte une roche.	
	○ E : Dépôts profonds de sol sans cohésion ($V_s > 150$ m/s).	

l'évènement	M_w	R_{rup}(Km)	Nom de station	Composantes	$D(M)$	$D(T)$	$D(B)$
Coyote Lake .1979/08/06	5.7	9.30	47379 Gilroy	A001	8.21	5.78	1.77
	5.7	9.30	Array #1	A001	8.18	7.30	1.84
N. Palm Springs. 1986/07/08	6	46.70	5160 Anza Fire	A002	9.52	6.77	0.31
	6	46.70	Station	A002	9.57	5.87	0.90
	6	45.60	5224 Anza - Red	A003	7.96	6.29	4.38
	6	45.60	Mountain	A003	9.08	6.54	2.96
	6	63.30	13198 Murrieta	A004	15.61	6.97	1.50
	6	63.30	Hot Springs	A004	16.18	9.62	0.90
	6	25.80	12206 Silent	A005	9.63	6.98	1.63
	6	25.80	Valley - Poppet	A005	11.45	6.84	2.04
	6	7.30	5072	A006	4.58	3.41	12.45
	6	7.30	Whitewater	A006	4.97	5.44	12.96
	6	57.60	13199	A007	11.26	8.67	2.51
	6	57.60	Winchester	A007	13.21	9.44	1.83
	6	24.60	90017 LA -	A008	9.80	7.02	0.90
	6	24.60	Wonderland	A008	7.22	6.72	0.75
6	21.20	24399 Mt	A009	11.94	8.33	6.92	
6	21.20	Wilson - CIT	A009	11.78	10.06	6.77	
6	9.00	90019 San	A010	11.64	8.32	9.36	
6	9.00	Gabriel - E	A010	5.62	5.10	7.84	
6	9.00	Grand Av	A010	5.62	5.10	7.84	
Morgan Hill .1984/04/24	6.2	16.20	47379 Gilroy	A011	10.57	8.94	5.36
	6.2	16.20	Array #1	A011	11.19	9.54	5.32
San Fernando .1971/02/09	6.6	86.60	111 Cedar	A012	11.28	10.26	0.00
	6.6	86.60	Springs. Allen	A012	13.1	10.41	0.00
	6.6	23.50	Ranch	A012	13.1	10.41	0.00
	6.6	23.50	127 Lake	A013	12.66	11.82	3.29
	6.6	23.50	Hughes #9	A013	12.66	9.38	4.61
6.7	26.80	127 Lake	A014	13.10	7.88	4.44	
6.7	26.80	Hughes #9	A014	10.88	9.32	7.50	
6.7	22.70	90017 LA -	A015	9.78	6.66	5.64	
6.7	22.70	Wonderland	A015	12.14	8.75	6.42	
6.7	22.70	Ave	A015	12.14	8.75	6.42	
Northridge .1994/01/17	6.7	36.10	24399 Mt	A016	16.24	10.84	8.59
	6.7	36.10	Wilson - CIT	A016	14.64	8.82	7.54
	6.7	36.10	Seis Sta	A016	14.64	8.82	7.54
	6.7	47.30	24310 Antelope	A017	17.46	13.84	0.06
	6.7	47.30	Buttes	A017	21.64	15.16	0.00
	6.7	37.80	24307 Leona	A018	23.62	13.6	5.06
6.7	37.80	Valley #3	A018	18.38	12.86	3.74	

	6.7	46.90	23595 Littlerock	A019	34.18	17.65	0.47
	6.7	46.90	- Brainard Can	A019	24.85	11.92	1.59
	6.7	71.50	23572 Mt Baldy	A020	28.00	12.86	3.94
	6.7	71.50	- Elementary Sch	A020	29.01	14.00	4.05
	6.7	8.00	24207 Pacoima Dam (upper left)	A021	9.94	5.96	23.92
	6.7	8.00		A021	9.84	5.40	23.92
	6.7	80.00	23598 Rancho Cucamonga - Deer Can	A022	32.45	15.45	0.039
	6.7	80.00		A022	30.48	15.43	4.46
	6.7	41.70	90019 San Gabriel - E. Grand Ave.	A023	20.57	13.05	12.42
	6.7	41.70		A023	25.00	14.55	12.51
	6.7	43.40	24644 Sandberg - Bald Mtn	A024	30.46	15.92	2.91
	6.7	43.40		A024	22.32	14.15	3.10
	6.7	68.40	23590 Wrightwood - Jackson Flat	A025	29.36	18.46	0.00
	6.7	68.40		A025	26.63	11.84	0.27
	6.9	11.20	47379 Gilroy Array #1	A026	5.42	3.68	9.68
	6.9	11.20		A026	7.68	6.53	13.26
	6.9	30.60	1032 Hollister - SAGO Vault	A027	19.06	15.40	0.35
	6.9	30.60		A027	23.17	16.30	0.00
	6.9	44.80	47377 Monterey City Hall	A028	18.95	11.93	1.98
	6.9	44.80		A028	20.09	13.23	2.90
	6.9	78.30	58338 Piedmont Jr High	A029	19.23	11.07	1.03
	6.9	78.30		A029	17.87	10.63	0.86
Loma Prieta .1989/10/18	6.9	88.60	58043 Point Bonita	A030	17.43	8.35	2.47
	6.9	88.60		A030	18.33	10.08	2.09
	6.9	81.60	58131 SF - Pacific Heights	A031	20.89	11.40	2.26
	6.9	81.60		A031	21.06	13.71	1.06
	6.9	79.70	58151 SF - Rincon Hill	A032	17.22	9.54	2.53
	6.9	79.70		A032	19.35	12.14	0.08
	6.9	68.20	58539 So. San Francisco. Sierra Pt.	A033	17.15	7.99	0.23
	6.9	68.20		A033	30.54	14.48	0.00
	6.9	80.60	58163 Yerba Buena Island	A034	1.67	12.07	0.99
	6.9	80.60		A034	12.69	10.48	0.69
Cape Mendocino. .1992/04/25	7.1	8.50	89005 Cape Mendocino	A035	19.76	9.76	19.52
	7.1	8.50		A035	19.58	6.18	19.28
Duzce Turkey .1999/11/12	7.1	30.20	1060 Lamont	A036	40.55	19.07	6.25
	7.1	30.20	1060	A036	36.70	18.51	0.40
Landers .1992/06/28	7.3	141.60	90019 San Gabriel - E Grand Av	A037	43.90	30.98	12.30
	7.3	141.60		A037	43.08	29.86	15.42

	7.3	69.20	21081 Amboy	A038	42.66	25.24	26.00
	7.3	69.20		A038	46.14	29.82	27.28
	7.3	1.10	24 Lucerne	A039	27.30	13.12	33.26
	7.3	1.10		A039	25.68	13.75	33.33
	7.3	51.70	12206 Silent Valley - Poppet Flat	A040	40.76	31.74	2.70
	7.3	51.70		A040	38.08	29.98	3.25
	7.3	42.20	22161 Twentynine Palms	A041	41.86	30.92	18.48
	7.3	42.20		A041	44.24	30.72	16.96
Kocaeli Turkey .1999/08/17	7.4	17.00	Gebze	A042	12.60	7.65	7.22
	7.4	17.00		A042	12.37	7.41	7.81
	7.4	4.80	Izmit	A043	21.86	14.99	15.88
	7.4	4.80		A043	25.96	13.24	24.02
	7.4	63.90	Maslak	A044	61.04	37.39	12.30
	7.4	63.90		A044	65.39	35.82	15.90
Chi-Chi. Taiwan .1999/09/20	7.6	77.75	ENA	A045	53.76	21.16	5.42
	7.6	77.75		A045	49.62	23.075	2.98
	7.6	57.00	HWA023	A046	54.09	17.00	5.30
	7.6	57.00		A046	52.21	23.40	7.50
	7.6	58.80	HWA026	A047	59.03	18.43	5.32
	7.6	58.80		A047	61.35	19.64	4.88
	7.6	59.26	HWA046	A048	53.00	17.31	7.44
	7.6	59.26		A048	52.64	16.76	6.73
	7.6	48.75	HWA056	A049	56.76	16.26	9.98
	7.6	48.75		A049	68.00	16.86	12.05
	7.6	96.59	ILA015	A050	63.79	26.92	0.16
	7.6	96.59		A050	57.30	27.83	0.56
	7.6	79.00	ILA024	A051	54.03	27.53	0.85
	7.6	79.00		A051	56.14	27.97	2.50
	7.6	94.77	ILA031	A052	53.53	20.23	3.59
	7.6	94.77		A052	56.10	19.88	3.77
	7.6	77.75	ILA050	A053	44.23	22.07	5.41
	7.6	77.75		A053	45.15	24.36	4.16
	7.6	90.37	ILA051	A054	70.40	37.17	0.00
	7.6	90.37		A054	60.93	18.96	9.06
	7.6	96.68	ILA052	A055	60.22	22.99	2.10
	7.6	96.68		A055	51.64	21.73	2.30
	7.6	127.53	ILA054	A056	65.34	20.88	1.90
	7.6	127.53		A056	61.73	25.84	2.16
	7.6	71.61	ILA063	A057	54.66	23.34	5.09
	7.6	71.61		A057	58.91	21.26	8.36

7.6	122.84	KAU034	A058	59.78	36.35	0.00
7.6	122.84		A058	60.16	33.59	0.00
7.6	157.43	KAU038	A059	64.54	39.62	2.43
7.6	157.43		A059	63.90	40.79	5.22
7.6	64.50	KAU047	A060	70.66	37.14	0.95
7.6	64.50		A060	73.92	39.43	2.86
7.6	52.00	KAU050	A061	75.61	37.29	1.78
7.6	52.00		A061	74.85	34.09	3.22
7.6	139.70	KAU051	A062	59.11	38.33	7.43
7.6	139.70		A062	61.09	35.57	6.40
7.6	83.58	KAU069	A063	76.79	37.18	0.45
7.6	83.58		A063	78.35	36.66	4.16
7.6	97.20	KAU077	A064	71.90	38.49	1.90
7.6	97.20		A064	76.40	41.28	4.26
7.6	102.85	KAU078	A065	74.39	37.34	2.14
7.6	102.85		A065	70.58	34.2	5.72
7.6	64.51	NSK	A066	30.78	21.71	0.39
7.6	64.51		A066	32.75	18.41	1.28
7.6	114.21	PNG	A067	55.72	28.29	0.00
7.6	114.21		A067	56.10	31.91	0.98
7.6	52.00	STY	A068	55.14	32.78	3.89
7.6	52.00		A068	54.20	29.62	3.25
7.6	98.81	TAP034	A069	61.88	21.19	2.25
7.6	98.81		A069	61.91	25.25	0.04
7.6	96.88	TAP035	A070	50.41	18.52	10.14
7.6	96.88		A070	55.51	22.52	6.73
7.6	95.60	TAP036	A071	41.08	20.49	1.69
7.6	95.60		A071	40.75	22.85	2.33
7.6	105.14	TAP051	A072	77.15	24.94	2.24
7.6	105.14		A072	71.02	25.98	10.40
7.6	125.93	TAP059	A073	66.28	32.23	9.20
7.6	125.93		A073	63.86	30.93	15.62
7.6	128.49	TAP060	A074	60.81	23.46	7.35
7.6	128.49		A074	64.67	18.92	11.45
7.6	130.91	TAP065	A075	54.10	24.40	9.66
7.6	130.91		A075	53.28	16.48	5.34
7.6	104.27	TAP067	A076	66.97	25.91	12.60
7.6	104.27		A076	58.83	22.78	9.85
7.6	135.31	TAP069	A077	58.20	20.55	0.97
7.6	135.31		A077	52.71	24.47	6.87
7.6	110.00	TAP072	A078	55.35	17.24	0.02

7.6	110.00		A078	56.00	25.99	2.46
7.6	118.44		A079	57.47	22.22	3.45
7.6	118.44	TAP075	A079	64.02	21.16	0.16
7.6	131.00		A080	60.96	23.81	16.22
7.6	131.00	TAP078	A080	63.26	27.84	12.75
7.6	130.64		A081	63.43	26.98	8.45
7.6	130.64	TAP079	A081	64.73	29.88	5.33
7.6	135.55		A082	25.40	25.40	0.90
7.6	135.55	TAP081	A082	67.39	19.56	3.76
7.6	125.50		A083	64.20	15.57	13.31
7.6	125.50	TAP103	A083	54.65	10.94	9.18
7.6	136.58		A084	38.98	25.54	1.78
7.6	136.58	TAW	A084	38.65	24.84	3.95
7.6	54.36		A085	27.11	15.19	11.51
7.6	54.36	TCU025	A085	27.76	17.92	9.56
7.6	14.34		A086	52.20	18.16	17.77
7.6	14.34	TCU046	A086	50.15	18.46	18.64
7.6	64.51		A087	60.39	22.65	0.02
7.6	64.51	TCU085	A087	55.71	19.97	2.18
7.6	136.58		A088	53.66	36.09	4.30
7.6	136.58	TTN016	A088	52.53	30.93	5.67
7.6	70.58		A089	78.44	38.83	3.32
7.6	70.58	TTN024	A089	81.81	36.82	4.32
7.6	90.63		A090	76.18	41.31	6.40
7.6	90.63	TTN028	A090	75.7	38.61	9.87
7.6	55.00		A091	65.59	33.04	11.2
7.6	55.00	TTN040	A091	71.61	35.76	9.65
7.6	54.16		A092	75.99	35.90	6.38
7.6	54.16	TTN041	A092	76.89	32.72	15.30