



HAL
open science

Estimation des mouvements sismiques et de leur variabilité par approche neuronale : Apport à la compréhension des effets de la source, de propagation et de site

Boumédiène Derras

► To cite this version:

Boumédiène Derras. Estimation des mouvements sismiques et de leur variabilité par approche neuronale : Apport à la compréhension des effets de la source, de propagation et de site. Tectonique. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAU013 . tel-01652115

HAL Id: tel-01652115

<https://theses.hal.science/tel-01652115>

Submitted on 29 Nov 2017

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Science de la Terre, l'Univers et l'Environnement**

Arrêté ministériel : 25 mai 2016

Présentée par

Boumédiène DERRAS

Thèse dirigée par **Pierre-Yves BARD**, Ingénieur Général Ponts, Eaux et Forêts, Institut des Sciences de la Terre (ISTerre), et codirigée par **Fabrice COTTON**, Professeur, German Research Center for Geosciences (GFZ)

préparée au sein de l'**Institut des Sciences de la Terre** dans l'**École Doctorale Terre, Univers, Environnement**

Estimation des mouvements sismiques et de leur variabilité par approche neuronale : Apport à la compréhension des effets de la source, de propagation et de site

Thèse soutenue publiquement le **11 mai 2017**,
devant le jury composé de :

M, John, DOUGLAS

DR Université de Strathclyde, Glasgow, rapporteur

M, Fabian-Louis, BONILLA

DR Université Paris-Est, IFSTTAR, Paris, rapporteur

Mme, Françoise, COURBOULEX

DR GeoAzur, Valbonne, examinatrice

Mme, Catherine, BERGE-THIERRY

DS CEA, Saclay, examinatrice

M, Philippe, GUEGUEN

DR ISTerre, Grenoble, président

M, Pierre-Yves, BARD

DR ISTerre IFSTTAR, Grenoble, directeur de thèse

M, Fabrice, COTTON

Pr GFZ, Potsdam, co-directeur de thèse



Remerciements

Plus de quatre années de thèse entre le laboratoire RISAM (Tlemcen), l'Université de Saïda, l'ISTerre (Grenoble) et GFZ (Postdam), entre la recherche scientifique et l'enseignement : le temps a passé très vite. Beaucoup de rencontres : entre Paris, Istanbul, Los Angeles et la Guadeloupe, franchement j'ai beaucoup appris.

Remercions en premier lieu, Dieu tout puissant de m'avoir permis de mener à bien et d'achever ce travail de thèse. Je souhaite remercier mon directeur de thèse Pierre-Yves Bard pour son aide et son soutien prodigué. J'adresse de chaleureux remerciements à mon co-directeur de thèse Fabrice Cotton pour ses idées et ses orientations. Sans eux, le présent travail n'aurait jamais été réalisé.

Mes remerciements vont également aux membres de jury, qui ont accepté de réviser et d'examiner le présent manuscrit. Je remercie John Douglas et Fabian-Louis Bonilla pour leurs suggestions et recommandations qui permettront, en un développement ultérieur, d'affiner les résultats. Egalement toute ma reconnaissance aux deux examinatrices Françoise Courboux et Catherine Berge-Thierry pour leurs rôles importants à l'amélioration du présent manuscrit. Enfin je remercie Philippe Gueguen d'avoir accepté de présider le jury et pour ses judicieuses recommandations. Au-delà de la qualité scientifique appréciable des membres du jury, je remercie chacun de m'avoir accordé leur attention.

Comme il me paraît injuste de faire une liste de tous ceux qui m'ont aidé, de près ou de loin, à la réalisation de cette thèse, j'ai choisi de laisser les gens se reconnaître seuls. Je vous remercie de ... m'avoir accueilli dans vos laboratoires d'étude qu'ils soient à Grenoble ou à Potsdam ; d'avoir pris le temps d'écouter mes états d'avancement, d'études, de recherche ; de m'avoir donné plusieurs gigas de données à traiter ; de m'avoir conseillé, orienté ; de m'avoir invité à manger ; de vous être intéressé à mon travail sans contrepartie ; de m'avoir soutenu.

Ce travail n'aurait pas pu voir le jour sans l'aide de différents financeurs : ressources propres de mes encadreurs, université de Saïda, programme Tassili (3MDU901), projet sinaps@, école doctorale TUE, laboratoire RISAM, ISTerre, GZF, CNRS...

Merci beaucoup à tous.

Boumédiène

Résumé

Cette thèse est consacrée à une analyse approfondie de la capacité des "réseaux de neurones artificiels" (RNA) à la prédiction des mouvements sismiques. Un premier volet important concerne la dérivation par RNA de "GMPE" (équations de prédiction du mouvement du sol) et la comparaison des performances ainsi obtenues avec celles des GMPE "classiques" obtenues sur la base de régressions empiriques avec une forme fonctionnelle préétablie (plus ou moins complexe). Pour effectuer l'étude comparative et obtenir les deux composantes inter-événements « between-event » et intra-événement « within-event » de la variabilité aléatoire, nous intégrons l'algorithme du « modèle à effets aléatoires » à l'approche neuronale. Cette approche est testée sur différents jeux de données réelles et synthétiques : la base de données compilée à partir d'événements européens, méditerranéens et du Moyen-Orient (RESORCE : Reference database for Seismic grOund-motion pRediction in Europe), la base de données NGA-West 2 (Next Generation Attenuation West 2 développée aux USA), la base de données japonaise dérivée du réseau accélérométrique KiK-net. En outre, un jeu de données synthétiques provenant d'une approche par simulation stochastique est utilisé. Les paramètres du mouvement du sol les plus utilisés en génie parasismique (PGA, PGV, spectres de réponse et également, dans certains cas, les fonctions d'amplification locales) sont considérés. Les modèles neuronaux ainsi obtenus, complètement dirigés par les données « data-driven », nous renseignent sur les influences respectives et éventuellement couplées de l'atténuation avec la distance, de l'effet d'échelle lié à la magnitude, des conditions de site et notamment la présence éventuelle de non-linéarités. Un autre volet important est consacré à l'utilisation des RNA pour tester la pertinence de différents proxies de site, au travers de leur capacité à réduire la variabilité aléatoire des prédictions de mouvement du sol. Utilisés individuellement ou en couple, ces proxies de site décrivent de manière plus ou moins détaillée l'influence des conditions de site locales sur le mouvement sismique. Dans ce même volet, nous amorçons également une étude des liens entre les aspects non linéaires de la réponse de site, et les différents proxies de site. Le troisième volet se concentre sur certains effets liés à la source : analyse de l'influence du style de la faille sismique sur le mouvement du sol, ainsi qu'une approche indirecte de la dépendance entre la magnitude et la chute de contrainte sismique.

Mots clés : Réseaux de neurones, Prédiction sismique, atténuation, base de données, effet de site, effet de source, incertitudes.

Ground-motion prediction and their variability through neural approach: Physical insight into source, path and site effects

This thesis is devoted to an in-depth analysis of the ability of "Artificial Neural Networks" (ANN) to achieve reliable ground motion predictions. A first important aspect concerns the derivation of "GMPE" (Ground Motion Prediction Equations) with an ANN approach, and the comparison of their performance with those of "classical" GMPEs derived on the basis of empirical regressions with pre-established, more or less complex, functional forms. To perform such a comparison involving the two "between event" and "within-event" components of the random variability, we adapt the algorithm of the "random effects model" to the neural approach. This approach is tested on various, real and synthetic, datasets: the database compiled from European, Mediterranean and Middle Eastern events (RESORCE: Reference database for Seismic ground-motion prediction in Europe), the database NGA West 2 (Next Generation Attenuation West 2 developed in the USA), and the Japanese database derived from the KiK-net accelerometer network. In addition, a comprehensive set of synthetic data is also derived with a stochastic simulation approach. The considered ground motion parameters are those which are most used in earthquake engineering (PGA, PGV, response spectra and also, in some cases, local amplification functions). Such completely "data-driven" neural models, inform us about the respective, and possibly coupled, influences of the amplitude decay with distance, the magnitude scaling effects, and the site conditions, with a particular focus on the detection of non-linearities in site response. Another important aspect is the use of ANNs to test the relevance of different site proxies, through their ability to reduce the random variability of ground motion predictions. The ANN approach allows to use such site proxies either individually or combined, and to investigate their respective impact on the various characteristics of ground motion. The same section also includes an investigation on the links between the non-linear aspects of the site response and the different site proxies. Finally, the third section focuses on a few source-related effects: analysis of the influence of the "style of faulting" on ground motion, and, indirectly, the dependence between magnitude and seismic stress drop.

Key word: neural networks, Ground-motion prediction, attenuation, databases, site effect, source effect, uncertainly.

Table des matières

INTRODUCTION GENERALE -----	1
CHAPITRE I	
LE FANTASTIQUE ESSOR DES DONNEES ACCELEROMETRIQUES ET SON IMPACT SUR L'EVOLUTION DE LA PREDICTION DES MOUVEMENTS SISMIQUES -----	6
INTRODUCTION -----	6
I.1. PARAMETRES DES MOUVEMENTS FORTS -----	7
I.2. ÉVOLUTION DE LA COLLECTION DES MOUVEMENTS FORTS -----	13
I.3. GMPES CLASSIQUES -----	28
I.4. ÉQUATION DE PREDICTION DU MOUVEMENT SISMIQUE (GMPE) -----	29
CONCLUSION -----	44
CHAPITRE II	
DEVELOPPEMENT D'UNE GMPE A EFFETS-ALEATOIRE DIRIGEE PAR LES DONNEES RESORCE -----	45
INTRODUCTION -----	45
II.1. APERÇU SUR LES RESEAUX DE NEURONES ARTIFICIELS -----	46
II.2. SYNTHÈSE DES TRAVAUX DÉJÀ RÉALISÉS -----	50
II.3. SITUATION DE LA PRÉSENTE THÈSE PAR RAPPORT À CELLE DE 2011 -----	56
II.4. MODÈLES À EFFETS ALÉATOIRES -----	57
II.5. PRÉSENTATION DE LA BASE DE DONNÉES RESORCE : ÉTUDE DE PREMIER ORDRE -----	63
II.6. VERS UN MODÈLE DE PREDICTION DU MOUVEMENT SISMIQUE TOTALEMENT CONDUIT PAR LES DONNÉES POUR L'EUROPE -----	65
II.7. COMPARAISON ENTRE LES 5 GMPES DÉVELOPPÉES POUR L'EUROPE ET LE MOYEN-ORIENT -----	90
CONCLUSION -----	93
CHAPITRE III	
QUEL PROXY POUR DECRIRE LES CONDITIONS DE SITE ET PREDIRE LES MOUVEMENTS SISMIQUES -----	95
INTRODUCTION -----	95
III.1. PENTE TOPOGRAPHIQUE « SLOPE » -----	96
III.2. INTRODUCTION -----	98
III.3. GROUND-MOTION DATASETS -----	100
III.4. DATA DISTRIBUTION -----	101
III.5. METHODOLOGY -----	103
III.6. HOW DO MEASURED OR INFERRED SITE-CONDITION PROXIES REDUCE GROUND-MOTION ALEATORY VARIABILITY? -----	107
III.7. IMPACT OF THE DATASET, SITE PROXY, AND QUALITY OF V_{s30} ON MEDIAN GROUND-MOTION MODELS. -----	112
III.8. APPLICATION: DEVELOPMENT OF A FULLY DATA DRIVEN GMPE BASED ON NGA-WEST 2 -----	112
III.9. THE NGA-WEST 2 ANN (ALL V_{s30} , V_{s30}): FUNCTIONAL FORM -----	114
III.10. RANGE OF APPLICABILITY -----	114
III.11. RESIDUAL ANALYSIS -----	114
III.12. DEPENDENCE OF ALEATORY VARIABILITIES WITH MAGNITUDE AND DISTANCE -----	116
III.13. COMPARISON OF THE MEDIAN ANN MODEL WITH BSSA14 -----	120
III.14. DISCUSSION AND CONCLUSIONS -----	125
III.15. ACKNOWLEDGMENTS -----	127
III.16. DATA AND RESOURCES -----	127
CONCLUSION -----	127

CHAPITRE IV

ETUDE DE LA CAPACITE DE DIFFERENTS PROXIES DE SITE A REDUIRE LA VARIABILITE ALEATOIRE DES EQUATIONS DE PREDICTION DES MOUVEMENTS DU SOL	129
INTRODUCTION	129
1^{ERE} PARTIE : V_{S30}, PENTE TOPOGRAPHIQUE, H_{800} ET F_0 : LA PERFORMANCE DE DIFFERENTS PROXYS DE SITE-CONDITION DANS LA REDUCTION DE LA VARIABILITE ALEATOIRE AU SOL ET LA PREDICTION DE LA REPONSE NON LINEAIRE DU SITE	130
IV.1. INTRODUCTION	131
IV.2. DATA SET	133
IV.3. METHODOLOGY	137
IV.4. RESULTS	140
IV.5. SUMMARY AND CONCLUSIONS	156
IV.6. DATA AND RESOURCES	157
IV.7. ACKNOWLEDGEMENTS	158
IV.8. PRINCIPAUX RESULTATS DE LA 1 ^{ERE} PARTIE	158
2^{EME} PARTIE : MODELE NEURONAL A LA DISTINCTION DES MEILLEURS PROXIES DE SITE POUR LA PRISE EN COMPTE DES EFFETS NON-LINEAIRES	159
IV.9. CONTEXTE	159
IV.10. JEUX DE DONNEES	161
IV.11. METHODOLOGIE	163
IV.12. QUELS SONT LES PROXIES DE CONDITIONS LOCALES DE SITES LES PLUS PERTINENTS?	164
IV.13. RESULTATS	169
IV.14. PRINCIPAUX RESULTATS DE LA 2 ^{EME} PARTIE	176
CONCLUSION	177

CHAPITRE V

ETUDES DES EFFETS DE LA SOURCE SUR LE MOUVEMENT SISMIQUE	179
INTRODUCTION	179
1^{ERE} PARTIE : ETUDE DE L'EFFET DU STYLE DE LA FAILLE SUR LA VARIABILITE ALEATOIRE DU MOUVEMENT SISMIQUE PAR L'APPROCHE NEURONALE : APPLICATION SUR RESORCE, NGA-WEST 2 ET KIK-NET	180
RESUME	180
V.1. INTRODUCTION	181
V.2. JEUX DE DONNEES	184
V.3. METHODOLOGIE	186
V.4. COMMENT LA PRISE EN COMPTE DU STYLE DE LA FAILLE PEUT-ELLE REDUIRE LA VARIABILITE ALEATOIRE DU MOUVEMENT SISMIQUE	190
V.5. INFLUENCE DU STYLE DE LA FAILLE SUR LA MEDIANE	191
V.6. COMPARAISON ENTRE LES IS DES TROIS STYLES DE FAILLES	194
2^{EME} PARTIE : IMPACT DU CHOIX DE LA GAMME DE LA DISTANCE SUR L'EFFET DU STYLE DE LA FAILLE : EXEMPLE DE LA BASE DE DONNEES KIK-NET	194
V.7. INTRODUCTION	194
V.8. INFLUENCE DE LA GAMME DE DISTANCE SUR LA REDUCTION DE LA VARIABILITE ALEATOIRE MENEES PAR LE STYLE DE LA FAILLE	196

V.9. DEPENDANCE ENTRE LE STYLE DE LA FAILLE ET LES PARAMETRES DE LA SOURCE, DE PROPAGATION D'ONDE ET DE LA PERIODE STRUCTURELLE----- 197

3^{EME} PARTIE : DEPENDENCE ENTRE LA MAGNITUDE ET LA CHUTE DE CONTRAINTE : QUE PEUT-ON TIRER DE L'EFFET D'ECHELLE LIE A LA MAGNITUDE -----201

ABSTRACT----- 202
V.10. INTRODUCTION ----- 202
V.11. DATA ----- 204
V.12. METHOD ----- 205
V.13. RESULTS ----- 208
V.14. CONCLUSION ----- 212
V.15. ACKNOWLEDGMENTS----- 212
V.16. DATA AND RESOURCES ----- 213
CONCLUSION ----- 213

CONCLUSIONS GENERALES ET PERSPECTIVES -----216

RAPPEL DES RESULTATS ----- 216
PERSPECTIVES----- 224

REFERENCES BIBLIOGRAPHIQUES -----227

ANNEXES-----237

ANNEXE 1 : INTERFACE GRAPHIQUE DES MODELES NEURONAUX----- 237
ANNEXE 2 : ARTICLE DE DOUGLAS ET AL. (2014) ----- 239
ANNEXE 3 : SUPPLEMENT ELECTRONIQUE DE DERRAS ET AL. (2014) ----- 256

Liste des figures

Introduction

Figure.1 Impact de la valeur de σ sur la courbe d'aléa sismique obtenue par une approche probabiliste.	1
Figure.2 Evolutions temporelles du nombre de GMPEs et des enregistrements sismiques	2
Figure.3 Evolution de σ durant les 30 dernières années (Strasser et al. 2009)	3
Figure.4 Carte isovitesse de la vallée de Grenoble donnant la classe NEHRP_Vs30 à partir de la pente topographique (slope).	3

Chapitre I

Figure.I.1 Schémas de principe de base de fonctionnement d'un accélérographe. (a) représente la mesure du déplacement de la composante horizontale et (b) la mesure de la composante verticale. m, k et c représentent la masse, la rigidité et l'amortissement de l'accélérographe, respectivement (Kramer, 1996).	8
Figure.I.2 Variation du facteur d'amplification dynamique avec le rapport des fréquences	9
Figure.I.3 Différents accélérogrammes et spectres de réponse pour la même valeur de PGA=0.17 g (Boore, 2015).	10
Figure.I.4 Exemple des traces en accélération, vitesse, déplacement et spectre de réponse de la station « MYGOO4 » du KiK-Net. L'enregistrement est celui du Tohoku (11/03/2011) avec une magnitude (JMA) de 9.0, enregistré à une distance épacentrale égale à 183 km.	11
Figure.I.5 Construction du spectre de réponse au séisme (divers k/m avec c constant) :Boore, 2015.	12
Figure.I.6 Distribution des épacentres des évènements de la base de données RESORCE (http://www.resorce-portal.eu/)	16
Figure.I.7 Structure et ressources de la base de données RESORCE (http://www.resorce-portal.eu/)	16
Figure.I.8 Distribution des stations de K-Net en bleu et KiK-Net en vert (Kashima.,2009).	20
Figure.I.9 Distribution des stations KIK-Net. Les couleurs représentent la profondeur (Aoi et al. 2004).	21
Figure.I.10 Distribution des épacentres : base de données NGA-West 1 et 2 (Ancheta et al. 2014).	23
Figure.I.11 Distribution magnitude-distance des enregistrements : NGA-West 1 et 2 (Ancheta et al. 2014).	24
Figure.I.12 Effet d'échelle lié à la magnitude : modèle de Campbell and Bozorgnia de 2008 et de 2013. La distance de rupture est $R_{rup}=10$ km, $V_{s30}=760$ m/s pour un mécanisme en décrochement (http://peer.berkeley.edu/ngawest2/wp-content/uploads/2014/04/S1-Bozorgnia_Speaker_presentation_USSD_2014.pdf).	24
Figure.I.13 Nombre d'enregistrements dans la base de données NGA-West 1 et NGA-West 2 pour différentes régions. CH: Chine, JP: Japon, Med: Méditerranée, TW: Taiwan, WNA: Nord-Ouest de l'Amérique (principalement la Californie). (Ancheta et al. 2014).	25
Figure.I.14 Nombre de sites en fonction de V_{s30} (mesurées et estimées) pour les deux bases de données NGA – West 1 et NGA-West 2 (http://peer.berkeley.edu/ngawest2/wp-content/uploads/2014/04/S1-Bozorgnia_Speaker_presentation_USSD_2014.pdf).	25

Figure.I.15 Localisation de l'épicentre : la base de données SHARE (Giardini et al. 2013).	27
Figure.I.16 Présentation de l'évolution annuelle des données de tremblements de terre et des enregistrements de mouvements sismiques (Giardini et al. 2013).	27
Figure.I.17 Distribution magnitude-distance pour les trois bases de données RESORCE (Akkar et al. 2014), KiK-Net (Dawood et al. 2016) et NGA-West 1 et 2 (Ancheta et al. 2014).	27
Figure.I.18 Distribution régionale des GMPEs (Douglas, 2016)	28
Figure.I.19 Évolution du nombre de GMPEs entre 1964 et 2016 (Douglas, 2016)	29
Figure.I.20 Evolution des GMPEs avec les enregistrements sismiques (Douglas, 2014)	29
Figure.I.21 Schéma illustratif des effets influençant le mouvement du sol à la surface (Yamada, 2013)	31
Figure.I.22 Composantes inter-événements (between-events) et intra-événements (within-events) de la variabilité aléatoire du mouvement sismique (Strasser et al. 2009)	35
Figure.I.23 Forme fonctionnelle donnée par Chiou and Youngs (2014).	39
Figure.I.24 Spectres de réponse des 5 GMPEs pour un scénario donné (http://ngawest2.berkeley.edu).	40
Figure.I.25 Comparaison entre les GMPEs élaborées pour les événements de subduction au Japon et les PGA enregistrés lors du séisme de Tohoku http://daveboore.com/daves_notes/comparisons_of_ground_motions_from_the_m_9_tohoku_earthquake_with_gmpes_v1.4.pdf	41
Figure.I.26 Comparaison de l'écart type (σ) pour $M=5$ pour des séismes à faille décrochante à une distance de rupture R_{rup} égale à 30 km et pour $V_{s30} = 760$ m/s (Gregor et al. 2014).	42
Figure.I.27 Comparaison entre l'écart type (σ) vs périodes (s) de CY2008 et CY2014 (Chiou and Youngs 2013).	42
Figure.I.28 Impact de la valeur de σ sur l'aléa sismique probabiliste (Bommer and Abrahamson, 2006).	43
Figure.I.29 Évolution de σ durant les 30 dernières années (Strasser et al. 2009)	43
Figure.I.30 Nombre de paramètres utilisés dans les GMPEs entre 1964-2009 (Bommer et al. 2010)	44
Chapitre II	
Figure.II.1 Exemple illustratif du problème de sous et sur apprentissage :biais/variance.	48
Figure.II.2 Biais et variance contribuant à l'erreur totale.	49
Figure.II.3 Organigramme illustrant le plan de l'estimation des paramètres caractérisant le mouvement sismique	51
Figure.II.4 Interface graphique sous Excel du modèle d'atténuation de la vitesse maximale du sol (ANN3).	51
Figure.II.5 Comparaison entre les σ des pseudo-spectres de réponses obtenus par le modèle neuronal et par les GMPEs classiques.	56
Figure.II.6. Variation des PGA italiennes et Turques avec la distance	64
Figure.II.7 Schéma illustratif de la distance Joyner et Boore (Scasserra et al., 2009)	65
Figure.II.8 (a) Distribution of earthquakes and records according to different magnitude intervals (magnitude step 0.5). (b) distance-Magnitude-region distribution of selected recordings. (c) Magnitude- V_{s30} -region distribution of selected recordings. (d) and (e) Distribution of the data used in terms of distance R_{JB}	

- (vector range (1 2.28 5.25 12 27.54 63.1 144.5 331.13 758.57) km and V_{s30} (interval range : 100 m/s), respectively. 70
- Figure.II.9 Distribution of PGA with R_{JB} distance for four magnitude intervals in the total range [3.6-7.6], using a magnitude bin size equal to ± 0.5 . 71
- Figure.II.10 Distribution of the style-of-faulting (normal, reverse and strike-slip) by country (Italy, Turkish and others country) 71
- Figure.II.11 Architecture of the ANN for PGA, PGV and PSA[0.01 to 4 sec] prediction. The w_{ij}^h is the synaptic weight between the i th neuron of the input layer and the j th neuron in the hidden layer, b_j^h the bias of the j th neuron in the hidden layer. Also the w_{ij}^o is the synaptic weight between the j th neuron of the hidden layer and the k th neuron in the output layer, b_j^o the bias of the k th neuron in the output layer. 72
- Figure.II.12 Variations of the between-event, within-event, and total σ values derived from the ANN model as a function of period. 80
- Figure.II.13 Distributions of the between-event and within-event residuals for PGA [(i.e, PSA(T=0 sec))] and PSA(2 sec) as a function of M_w , R_{JB} and V_{s30} . In all graph we present the dependence of between-event and within-event residuals from Italian (blue) and Turkish (red) data. We illustrate also the Mean residuals \pm one standard deviation in the different metadata ranges. The circles show average residuals, the vertical bars are the symmetrical error bars for two standard deviations; and the horizontal bars represent 1 standard deviation. The dashed lines show the $\pm\sigma$ between (or within) –event for the corresponding ground motion parameter. 80
- Figure.II.14 Ground-motion predictions as a function of the Joyner and Boore distance for four different magnitudes (color code) and two different site conditions (solid line, $V_{s30} = 300$ m/s; dashed line: $V_{s30} = 1$ km/s). Left: PGA; Right: PSA (2 sec). All plots are distances to a normal fault rupture. 82
- Figure.II.15 Magnitude-scaling ($\partial \log y / \partial M$) of predicted PGA and PGV versus M_w [from 4.5 to 6.5] at $R_{JB} = 15$ km for stiff soil, a hypocentral depth equal to 10 km and a normal style of-fault. 83
- Figure.II.16 Illustration of site effects: (a) Amplification of PGA on a soft site characterized by $V_{s30} = 300$ m/sec relatively to a rock site with $V_{s30} = 800$ m/sec) for different magnitude and distance ranges. (b) Display of the NL signature in site amplification factor between a $V_{s30} = 300$ m/sec site and a standard rock site characterized by $V_{s30} = 800$ m/sec. 84
- Figure.II. 17 Comparison of the spatial decay of PGA and PSA (2 s) predicted with the present ANN model, with those predicted by recent GMPEs, together with the corresponding RESORCE data. For a given magnitude M_w , all recordings with a magnitude equal to $M_w \pm 0.5$ and a $V_{s30} = 500 \pm 100$ m/sec, were considered $M_w = 5$ (top), $M_w = 6$ (middle) and $M_w = 7$ (bottom). The predictions are computed for $V_{s30} = 500$ m/sec, depth = 10 km and normal faulting 86
- Figure.II.18 Comparison of the period-dependence of median pseudo-spectral accelerations (5% damped) derived from the present model with those proposed in a few recent European GMPEs, The rows

correspond to different magnitude values: $M_w=5$ (top), $M_w=6$ (middle) and $M_w=7$ (bottom). The columns correspond to a relatively soft site ($V_{s30}=300$ m/sec, left) and a stiff soil / soft rock ($V_{s30}=800$ m/sec, right). For all plots, we consider $R_{JB}=30$ km, depth = 10 km and normal faulting.	87
Figure.II.19 Comparaison entre l'écart type total (σ) obtenu pas les 5 GMPEs	91
Figure.II.20 Comparaison de l'effet d'échelle lié à l'atténuation pour une faille de décrochement, $V_{s30}=270$ m/s	92
Figure.II.21 Comparaison de l'effet d'échelle lié à la magnitude pour une faille de décrochement, $V_{s30}=760$ m/s et pour $R_{JB}=30$ km.	93
Figure.II.22 Comparaison du rapport d'amplification de site pour $V_{s30}=270$ m/s en fonction de SA pour $V_{s30}=760$ m/s : $M_w=7$, $R_{JB}=100$ km et pour une faille de chevauchement.	93
Chapitre III	
Figure.III.1 Carte isovitesse de la vallée de Grenoble donnant la classe NEHRP de V_{s30} à partir de la pente topographique (slope).	96
Figure.III.2 Distribution of the four datasets considered in this study in the [Magnitude (M_w) - distance (R_{JB})] plane.	103
Figure.III.3 Cumulative distribution function of the four considered datasets (NGA-West 2 and RESORCE) with respect to the explanatory variables, i.e. R_{JB} distance (top left), magnitude M_w (top right), V_{s30} (bottom left) and Slope (bottom right).	104
Figure.III.4 Correlations of V_{s30} values (m/s) versus Slope values (m/m). for the four considered datasets boxes represent V_{s30} and Slope ranges recommended by Wald and Allen (2007) and Allen and Wald (2009) for classification into NEHRP categories (roughly equivalent to EC8 classes (Lemoine et al., 2012)) in the case of active tectonic areas	105
Figure.III.5 Structure of the Neural Networks considered for the prediction of PGA, PGV and PSA[0.01 to 4 s], in the two cases of discrete (top) and continuous (bottom) site proxies. W_{ij}^h is the synaptic weight between the i^{th} neuron of the input layer and the j^{th} neuron in the hidden layer, b_j^h the bias of the j^{th} neuron in the hidden layer. Also the W_{jk}^o is the synaptic weight between the j^{th} neuron of the hidden layer and the k^{th} neuron in the output layer, b_k^o the bias of the k^{th} neuron in the output layer	106
Figure.III.6 Sensitivity of the aleatory variability components (between-event, τ and within-event, ϕ) to the site proxies used as input for the neural network. As in Figures.III 2 and 4, the top left frame corresponds to the "All V_{s30} " NGA-West 2 dataset, the top right to the "measured V_{s30} " NGA-West 2 subset, the bottom left to the "Inferred V_{s30} " subset, and bottom right to the RESORCE dataset. The aleatory variability values are provided both on \log_{10} scale (left axis) and on \ln scale (right axis). They are provided for spectral ordinates for periods from 0.01 to 4 s, and for PGA and PGV as well.	109
Figure.III. 7 Quantifying the performance of the various continuous SCPs for the NGA-West 2 case. The two top frames display the total aleatory variability σ_t for (a) the four ANN models with continuous SCPs V_{s30}	

(All, Measured, Inferred) or Slope (All) and (b) the three "reference" ANN models without SCPs (All, Measured, Inferred). The bottom frame displays the reduction of aleatory variability (R , see equation.III.2) for the 4 ANN models displayed in (a) compared with their respective "reference models (b) 110

Figure.III.8 Robustness of the median ground motion predicted by different models. Effect of the dataset (top, NGA-West 2 RESORCE), the SCP proxy (middle, V_{s30} / Slope) and the SCP quality (bottom, measured / inferred / mix V_{s30}) measurement. The plots display the distance dependence of the spectral accelerations at three periods ($T = 0.0$ s, left column; $T = 0.2$ s, middle column; and $T = 1.0$ s, right column) for a $M_w = 6$ event, at sites with $V_{s30} = 400$ m/s and Slope=0.02 m/m. 113

Figure.III.9 Comparison of the mean binned residuals μ_M and μ_I (left), and the corresponding within-event standard deviations ϕ_M and ϕ_I (right) for the two "Measured V_{s30} " and "Inferred V_{s30} " NGA-West 2 subsets. The bins correspond to ten equal sample-size subsets with increasing V_{s30} values. The four rows correspond to spectral accelerations for four different periods, $T = 0$ s, 0.2, 1.0 and 2.0 s, from top to bottom 118

Figure.III. 10 Dependence of the between-event variability with magnitude bins (top) and of the within-event variability with distance bins (middle for moderate magnitude, stiff sites: $M_w \leq 4$, $V_{s30} \geq 300$ m/s and bottom for larger magnitudes and stiff sites: $M_w \geq 5.5$, $V_{s30} \geq 300$ m/s). The results are displayed for spectral accelerations at two periods ($T = 0.2$ s, black dashed line, and $T = 1.0$ s, black solid line), and compared with the BSSA14 model (grey lines). The dots correspond to the center of the magnitude or distance bins 119

Figure.III.11 Comparison of the distance scaling for different magnitudes (3.5 to 7.5) between the present ANN(All V_{s30}, V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for PGA at a stiff site with $V_{s30} = 500$ m/s 121

Figure.III.12 Comparison of the magnitude scaling for different spectral accelerations ($T = 0, 0.2, 1.0$ and 2.0 s, from top left to bottom right) between the present ANN(All V_{s30}, V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for a stiff site with $V_{s30} = 500$ m/s located at a R_{JB} distance of 30 km. 122

Figure.III.13 Comparison of the V_{s30} Scaling for different spectral accelerations ($T = 0, 0.2, 1.0$ and 2.0 s, from top left to bottom right) between the present ANN(All V_{s30}, V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for a site located at a R_{JB} distance of 30 km of a $M_w = 6$ event. 123

Figure.III. 14 Illustration of the non-linearity in the site term. The site amplification factor between a relatively soft site with $V_{s30} = 250$ m/s and a standard rock site characterized by $V_{s30} = 760$ m/s, is displayed as a function of the corresponding rock spectral amplitude. For the sake of simplicity, the variation of the rock motion level is considered by varying the magnitude from 3.5 to 7.5 at a fixed R_{JB} distance of 30 km. The unusual shape of BSSA14 curves is related to the fact that BSSA14 uses the rock PGA (and not the rock PSA(T)) to constrain the non-linear term. 124

Figure.III. 15 Comparison of median spectra predicted for increasing magnitudes at a stiff site $V_{S30} = 500$ m/s located at a R_{JB} distance of 30 km, between the present ANN(All V_{S30}, V_{S30}) model (solid line) and the BSSA14 model (dashed line). 125

Chapitre IV

Figure.IV.1 Distribution of the KiK-net dataset considered in this study. The three top plots correspond to the Magnitude (M_w) - distance (R_{JB}) distribution. For these three plots, the bins with different colors correspond to seven sample-size subsets with increasing PGA values from 0.001 to 1 g. On the three bottom plots, the distribution of PGA versus distance (R_{JB}) is shown. In these plots, the distribution is given by bins of Magnitude, $M_w=[4 - 8]$. On the left we take all dataset, in the middle plots the show data correspond to $V_{S30} = [100-300]$ m / s and in the right plots the datas are given for $V_{S30} > 800$ m/s. 135

Figure.IV.2 Empirical cumulative distribution function (CDF) versus the explanatory variables R_{JB} (top-left), M_w (top-right), V_{S30} (first middle-left), slope (first middle-right), H_{800} (second middle-left), f_0 (second middle-right) , depth (bottom-left) and the response PGA variable (botton-right) for the considered KiK-net data sets. 136

Figure.IV.3 Correlation between the different SCPs, with the correlation coefficient values for each pair of SCPs. The dotted line represents the extreme values of the PCS: 10% and 90% of the empirical CDF respectively. The red circles present the SCPs medians values. The extreme and medians values are presented in the Table.IV.2. 137

Figure.IV.4 Structure of the Neural Networks for PGA, PGV and PSA[0.01 to 4 sec] prediction, for one site proxy (top) and two-site proxies (bottom). The w_{ij}^h is the synaptic weight between the i th neuron of the input layer and the j th neuron in the hidden layer, b_j^h the bias of the j th neuron in the hidden layer. Also the w_{ij}^o is the synaptic weight between the j th neuron of the hidden layer and the k th neuron in the output layer, b_j^o the bias of the k th neuron in the output layer. 139

Figure.IV.5 Sensitivity of the aleatory variability components (total σ , within-event, ϕ between-event, τ) to the site proxies used as input for the neural network. The left frame corresponds to use one-SCP as input. In right part, it is for two-SCPs. The σ , ϕ , τ standard deviation values are provide both on \log_{10} scale (left axis) and on \ln scale (right axis). They are provided for spectral ordinates for periods from 0.01 to 4 s. The M_w and R_{JB} are used to obtain the reference model (without proxies in this Figure for the one-SCP and two-SCPs). 141

Figure.IV.6 Performance of various SCPs in reducing ground-motion aleatory variability. The top frames display the variance coefficient reduction (R_σ) for the total standart deviation σ . The middle frames illustre the variance coefficient reduction (R_ϕ) of the within-event standart deviation ϕ . The bottom frames show the variance coefficient reduction (R_τ) of the bewteen-event standart deviation τ . In the left frames (for top, middle and bottom frames), we use 4 ANN models with one-SCP measured V_{S30} or f_0 , H_{800} and

slope. In the right frames (for top, middle and bottom frames), we show R_σ , R_ϕ and R_τ for the 8 ANN models with two-SCPs, three-SCPs and four-SCPS configurations. 143

Figure.IV.7 Strength of the median ground motion predicted by different models. Effect of the one-SCP proxy (top, $V_{s30}/f_0/H_{800}/\text{slope}$) and the two-SCPs in the bottom row. The plots display the distance dependence of the spectral accelerations at three periods ($T=0.0$ s, left column; $T=0.2$ s, middle column; and $T=1.0$ s, right column) for a $M_w=5.5$ and depth=10 km event, at sites with $V_{s30}=460$ m/s, $f_0=4$ Hz, $H_{800}=30$ m and slope=0.04 m/m (taken from Table.IV.2: 10 % and 90 % CDF). The grey zone ($R_{JB} \leq 30$ km) should be considered less reliable since there are very few stiff site recording in this distance range (see Figure.IV.1) 146

Figure.IV.8 Site amplification factor (SR: equation.IV.3) versus R_{JB} at $T=0.0, 0.2, 1.0$ s for $M_w=6$ and focal depth=10 km. We consider one-SCP (top) and two-SCPs (bottom). The grey zone ($R_{JB} \leq 30$ km) should be considered less reliable since there are very few stiff site recording in this distance range (see Figure.IV.1) 148

Figure.IV.9 Illustration of the non-linearity in the site term. Site amplification factor (SR: equation.IV.3) versus $PGA_{r\text{stiff}}$ (90 % of CDF) at $T=0.0, 0.2, 1.0$ s for $R_{JB}=30$ km, focal depth=10 km and $M_w=[5 \text{ to } 7]$. We consider one-SCP (top) and two-SCPs (bottom). 151

Figure.IV.10 Comparison of the NL components of site amplification for different description of the loading level, for the one-SCP ANN model using V_{s30} . Each panel represents the (soft / stiff) site amplification at a given oscillator period as a function of a loading parameter. The three columns correspond to three different oscillator periods (0, 0.2 and 1.0s). The three rows correspond to different choices for the loading level: strain proxy $PGV_{\text{soft}}/V_{s30}$ on top, spectral ordinate at the stiff site $PSA_{\text{stiff}}(T)$ in the middle, and PSA_{stiff} at the 0.0 s (bottom). As in Figure.IV.9, the site amplification factors (SR : equation.IV.3) are calculated for three distances: 30 km, 50 km and 75 km, for which the variation of loading level corresponds to magnitude variations between 5 and 7 (for a fixed focal depth equal to 10 km). 152

Figure.IV.11 Similar to Figure.IV.10, but for the one-SCP model-using f_0 as site condition proxy. Stiff site corresponds to $f_0 = 11$ Hz, soft site to $f_0 = 0.4$ Hz. 153

Figure.IV.12 Similar to Figure.IV.10, but for the one-SCP model-using H_{800} as site condition proxy. Stiff site corresponds to $H_{800} = 6$ m, soft site to $H_{800} = 120$ m. 154

Figure.IV.13 Similar to Figure.IV.10, but for the one-SCP model-using slope as site condition proxy. Stiff site corresponds to slope = 0.13 m/m, soft site to slope = 0.01 m/m. 155

Figure.IV.14 Site amplification factor versus period for $R_{JB}=30, 50, 75$ km and for $M_w=5$ (top) and 6.5 (bottom) and a focal depth=10 km. We consider only the site proxy $V_{s30}-H_{800}$ ($V_{s30}=270$ m/s and $H_{800} = 6, 25$ and 120 m) to represent the soft site configurations. The stiff site is characterized by $V_{s30}=800$ m/s and $H_{800}=6$ m. 156

Figure.IV.15 Illustration de la quantification des effets de site non linéaire pour le site IWHTH04. (a) le rapport spectral surface/(fond de forage). (b) $RSR_{\text{NL-L}}$ pour un PGA_{seuil} des mouvements forts égal à 200 m/s² et pour un PGA représentant les mouvements faibles <25 cm/s². Le trait continu correspond à la médiane et les deux lignes en pointillé correspondent à \pm l'écart type. Le point rouge représente la

fréquence F_{in} qui correspond à l'intersection avec la courbe de RSR_{LN-L} médiane dont la valeur est juste inférieure à 1 (Régnier et al. 2016).	160
Figure.IV.16 Procédure de l'analyse de la persévérance des différents proxies de site	166
Figure.IV.17 Distribution de V_{s30} en utilisant les données initiales (haut), transformation en $\log_{10}(V_{s30})$ (milieu) et $BOXCOX(V_{s30})$ en bas, avec les distributions théoriques qui leurs correspondent	167
Figure.IV.18 Valeurs et tendances de $\sigma \log_{10}[RSR_{NL-L}]$ en fonction des 9 bins de fréquences. Les résultats sont obtenus avec les deux proxies V_{s30} et PGA_{surf}	170
Figure.IV.19 Courbes des médianes $\pm\sigma$ des RSR_{NL-L} adimensionnels normalisés en fréquences pour les 3 niveaux de PGA_s . (Régnier et al. 2015)	174

Chapitre V

Figure.V.1 Comparaison entre les facteurs FM des 6 GMPEs. À droite) : rapport des PSA : faille inverse sur faille de décrochement. À gauche) rapport des PSA : faille normale sur faille de décrochement.	182
Figure.V.2 Vue schématique d'une faille sismogénique avec les différents paramètres caractérisant la géométrie, le déplacement et les orientations de la faille.	183
Figure.V.3 Distribution de la magnitude du moment en fonction de la distance Joyner et Boore pour les trois jeux de données.	186
Figure.V.4 Distribution de l'accélération maximale du sol en fonction de la distance Joyner et Boore pour les trois jeux de données.	186
Figure.V.5 Distribution de l'accélération maximale du sol en fonction de la distance Joyner et Boore et du style de la faille pour les jeux de données RESORCE (à gauche), NGA-West 2 (milieu) et KiK-Net (à droite)	188
Figure.V.6 Répartition des pays par jeux de données	188
Figure.V.7 Structure des modèles neuronaux utilisés lors de cette étude pour la prédiction du mouvement sismique à la surface. Nous adoptons 6 entrées si le modèle prend en compte l'effet du style de la faille (avec 6 neurones dans la couche cachée). 3 entrées : R_{JB} , $\log_{10}(R_{JB})$ et M_w pour les modèles de références : sans effet du style de la faille dans ce cas précis nous choisissons 3 neurones cachés.	189
Figure.V.8 Sensibilité des composantes de la variabilité aléatoire (inter-événements (τ), intra-événements (ϕ) et l'écart type total (σ)) au style de la faille pour les trois modèles neuronaux élaborés à partir des trois jeux de données : RESORCE (à gauche), NGA-West 2 (au milieu) et KiK-Net (à droite). Ces composantes sont celles des ordonnées spectrales pour une période allant de 0.01 s à 2 s et pour le PGA et le PGV	192
Figure.V.9 Quantification des performances du style de la faille par modèle neuronal avec la moyenne correspondantes (μ) La performance est mesurée avec le coefficient R_c (%) pour une période comprise entre 0.01 et 2 s.	192
Figure.V.10 Effet du style de la faille sur la médiane pour les modèles : RESORCE, NGA-West 2 et KiK-Net à une magnitude égale à 5.5 pour $T=0.0$ (à gauche), $T=0.1$ s (au milieu) s et $T=1.0$ s (à droite).	193
Figure.V.11 Rapports d'amplitude spectrale ($F_{N,SS}$ et $F_{R,SS}$) en fonction de la période pour les modèles élaborés à partir des jeux de données RESORCE, NGA-West 2 et KiK-Net, avec $M_w=5.5$ et $R_{JB}=30$ km.	195

- Figure.V.12 Sensibilité des composantes de la variabilité aléatoire (inter-événements (τ), inter-événementé (ϕ) et l'écart type total (σ)) au style de la faille pour les 4 modèles neuronaux de prédiction du mouvement sismique élaborés à partir des 4 jeux de données KiK-Net : Pour toute la gamme de distance avec et sans FM (à gauche) et pour $R_{JB} < 80$ km avec et sans FM (à droite). L'ensemble des modèles comprennent les proxies de site et la profondeur focale. Les composantes de la variabilité aléatoire sont celles des ordonnées spectrales pour une période allant de 0.01 s à 4 s et pour le PGA et le PGV
196
- Figure.V.13 Quantification des performances du style de la faille par modèles neuronaux issus du jeu de données KiK-Net pour toute la gamme de R_{JB} (à gauche) et pour une distance inférieure à 80 km (à droite). Les performances sont mesurées par le coefficient R_c (%) pour une période comprise entre 0.01 et 4 s.
197
- Figure.V.14 Rapports d'amplitude spectrale ($F_{N:SS}$ et $F_{R:SS}$) en fonction de la période pour les modèles élaborés à partir des jeux de données KiK-Net pour toute la gamme de R_{JB} et pour une distance inférieure à 80 km et sur un intervalle de période allant de [0.01-4]s.
198
- Figure.V.15 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la distance pour les deux modèles des 2 jeux de données KiK-Net: $R_{JB} \leq 500$ km et $R_{JB} \leq 80$ km pour un scénario donné et pour $T = 0.1$ s.
200
- Figure.V.16 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la magnitude pour les deux modèles des deux jeux de données KiK-Net : $R_{JB} \leq 500$ km et $R_{JB} \leq 80$ km pour un scénario sismique donné et pour une période spectrale égale à 0.1 s.
200
- Figure.V.17 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la profondeur focale pour les deux modèles des deux jeux de données KiK-Net : $R_{JB} \leq 500$ km et $R_{JB} \leq 80$ km pour un scénario sismique donné et pour une période spectrale égale à 0.1 s.
201
- Figure.V.18 Distribution of the databases considered in this study in the [Magnitude (M_w) - distance (R_{hyp})] (left), [PSA @ 3.33 Hz - R_{hyp}] (right).
205
- Figure.V.19 Structure of the Neural Networks considered for the prediction of PGA, PGV and PSA[0.33 to 100 Hz].
 w_{ij}^h is the synaptic weight between the i^{th} neuron of the input layer and the j^{th} neuron in the hidden layer, b_j^h the bias of the j^{th} neuron in the hidden layer. Also the w_{jk}^o is the synaptic weight between the j^{th} neuron of the hidden layer and the k^{th} neuron in the output layer, b_k^o the bias of the k^{th} neuron in the output layer.
206
- Figure.V.20 Magnitude-dependence of ANN models based on various data selection ($R < 30$ km, $R < 40$ km, $R < 60$ km).
207
- Figure.V.21 Robustness of the ground-motion scaling derived using the ANN approach and the NGA-West 2 dataset. The graphs show the relationship between PSA at 3.33 Hz and magnitude for various subsets of the database.
209
- Figure.V.22 ANN predictive models derived with Aftershocks and Mainshocks data for 30 km.
210

Figure.V.23 Stochastic simulations of ground-motions (SMSIM : Boore, 2003) with various stress-drop models	210
Figure.V.24 Comparison between the magnitude scaling of observed (NGA-West 2) and simulated ground-motions at the distance of 30 km.	211
Figure.V.25 Magnitude-dependent decay of the seven models, for pseudo absolute acceleration spectra (PSA) at 3.3 Hz. We represent the ground-motion intensity normalized ($Z(M_w)$) at $M_w=4.5$ versus magnitude at the distance of 30 km.	211

Conclusions générales et perspectives

Figure.C.1 Variation des poids synaptiques en (%) avec la période pour les trois jeux de données (KiK-Net, RESORCE et NGA-West 2)	221
---	-----

Annexes

Figure.A.1 Interface graphique, sous Excel, du modèle d'atténuation donnant le PGA, PGV et PSA : application au set de données RESORCE (supplément électronique de l'article BEE)	237
Figure.A.2 Interface graphique sous Excel : Modèle neuronal de prédiction du mouvement sismique élaboré avec le set de données NGA-West 2.	238
Figure.A.3 Distributions of the between-event and within-event residuals for PSA(0.2 sec) and PGV according to M_w , R_{JB} and V_{S30} . In all graph we present the dependence of between-event and within-event residuals from Italian (blue) and Turkish (red) data. We illustrate also the Mean residuals -standard deviations in the different metadata ranges. The circles show average residuals, the vertical bars are the symmetrical error bars for two standard deviations; and the horizontal bars represent 1 standard deviation. The dashed lines show the $\pm\sigma$ between (or within) -event for the corresponding ground motion parameter	256
Figure.A.4 Comparison of predicted PSA at 2 sec and PGV by the present model and by other studies with recorded PSA and PGV from earthquakes with magnitude equal $M_w\pm 0.5$ and $V_{S30}=500\pm 100$ m/sec. $M_w=5$ (top), $M_w=6$ (middle) and $M_w=7$ (bottom). With the use of $V_{S30} = 500$ m/sec, depth = 10 km and normal faulting.	257

Liste des tableaux

Chapitre I

Table.I.1 Liste des séismes destructeurs du XXe siècle recensés jusqu'à 1995 (Kashima, 2009).	18
Table.I.2 Distribution de la profondeur des stations installées en fond du forage du KIK-Net Aoi et al. 2004).	21

Chapitre II

Table.II.1 Nombre d'enregistrements et d'événements par pays : base de données RESORCE	65
Table.II.2 Influence of different activation functions on σ and on kk: number of iteration which gives the maximum likelihood in the random effect algorithm	75
Table.II.3 Evolution of the overall fit of the ANN model (as measured with total σ) for PGA, PGV, PSA at 0.2 sec and PSA at 2 sec, according to the different sets of input parameters.	76
Table.II.4 Preprocessing parameters for the ANN Model	76
Table.II.5 Postprocessing parameters for the ANN Model σ , τ and ϕ standard deviations	77
Table.II.6 Synaptic weight matrices and bias vectors for ANN model	79
Table.II.7 Sensitivity of the metadata parameters, P_i (%) is the total percentage of synaptic weight corresponding to each the five parameters.	81
Table.II.8 Description of validity range of the Ground-Motion Prediction Equations for Euro-Mediterranean region used in this study. N: normal, R : reverse and SS is strike-slip. M_w and R_{JB} are used by all authors	85
Table.II.9 Comparison between between-event σ , within-event σ and total σ for the actual study and other classical GMPEs.	85
Table.II.10 Nombre d'événements sismiques, les stations et les enregistrements utilisés pour élaborer les cinq modèles, la magnitude et la distance des plages de données utilisées, les gammes d'application recommandée par les développeurs de modèles et les critères d'exclusion utilisés pour sélectionner les enregistrements utilisés dans chaque le modèle	91

Chapitre III

Table.III.1 Range of magnitude, distance and site parameters for the various NGA-West 2 and RESORCE subsets considered in this study	102
Table.III.2 Sensitivity of ground motion models to magnitude, distance and site parameters, as expressed through the total percentages of synaptic weight (P , %) corresponding to each input parameter.	112
Table.III.3 Coefficients of the final ANN(All V_{s30} , V_{s30}) model (geometric mean of horizontal components, median value, see equation.III.3 in the text)	115
Table.III.4 Values of the between-event, within-event and total variabilities for the final ANN(All V_{s30} , V_{s30}) model; the within-event variability depends on whether the V_{s30} value is measures or inferred	117

Chapitre IV

Table.IV.1 Range of magnitude, distance and site parameters for the KiK-Net subset considered in this study	133
Table.IV.2 Metadata range for site effect and for fully data-driven models applicability.	138
Table.IV.3 Variation reduction of the variance (R_{σ}) in % for the 12 ANNs types for PGA, PGV and at $T=[0.2,0.5,1.0,2.0]$ s	144
Table.IV.4 Sensitivity of ground motion models to magnitude, distance and one-SCP, as expressed through the total percentages of synaptic weight (P, %) corresponding to each input parameter.	145
Table.IV.5 Sensitivity of ground motion models to magnitude, distance, depth and two-SCPs, as expressed through the total percentages of synaptic weight (P, %) corresponding to each input parameter.	145
Table.IV.6 Données statistiques pour chaque site de la base de données sites : Nombre de sites qui ont la valeur de f_{NL} . La moyenne arithmétique de f_{NL} . L'intervalle de f_{NL} l'écart type de f_{NL} en Hz (Régnier et al. 2016).	162
Table.IV.7 Définition de la fréquence en bins (Hz).	162
Table.IV.8 Valeurs des deux coefficients kurtosis et skewness pour les 5 proxies de site en tenant compte de la base de données site pour un seuil de $PGA=200$ cm/s^2 .	168
Table.IV.9 Valeurs des deux coefficients kurtosis et skewness pour les 7 proxies de site en utilisant la base de données événements	168
Table.IV.10 Valeurs de σ et R_c pour un proxy de site et pour l'ensemble de proxies de site en utilisant la base de données site avec un seuil de $PGAs=100$ cm/s^2 .	171
Table.IV.11 Valeurs de σ et R_c pour un proxy de site et pour l'ensemble de proxies de site en utilisant la base de données site avec un seuil de $PGAs=200$ cm/s^2 .	172
Table.IV.12 Valeurs de σ et R_c pour un proxy de site et pour l'ensemble de proxies de site en utilisant la base de données site avec un seuil de $PGAs=300$ cm/s^2 .	173
Table.IV.13 Valeurs de σ et R_c pour, en plus, de l'ensemble de proxies de site le/les proxies de sollicitation sismique en utilisant la base de données événement.	173
Table.IV.14 Valeurs de σ et R_c donnés par les RSR_{NL-L} prédites adimensionnels normalisés en fréquences pour, en plus, de l'ensemble de proxies de site le/les proxies de sollicitation sismique en utilisant la base de données événement. f_{NLobs} représente f_{NL} observée.	174
Table.IV.15 Valeurs de σ et R_c donnés par les f_{NL} prédites pour l'ensemble de proxies de site et de sollicitation sismique en utilisant la base de données événement.	175
Table.IV.16 Valeurs de σ et R_c donnés par les RSR_{NL-L} prédites adimensionnels normalisés en fréquences pour, en plus, de l'ensemble de proxies de site le/les proxies de sollicitation sismique en utilisant la base de données événement. f_{NLpred} représente f_{NL} prédite.	176

Chapitre V

Table.V.1 Critère de classement des trois styles de la faille en utilisant l'angle entre la direction de glissement et l'horizontale (Rake)	183
---	-----

Table.V.2 nombre d'événements, d'enregistrements, de sites. Les intervalles de magnitude et distance pour les jeux de données RESORCE, NGA-West 2 et KiK-Net considérés dans cette étude	185
Table.V.3 Nombre d'enregistrements (brut et en pourcentage) par style de faille de décrochement (SS), normal (N) et inverse (R) pour les jeux de données RESORCE, NGA-West 2 et KiK-Net utilisés dans cette étude	185
Table.V.4 Valeurs des variables binaires pour les trois types de failles.	187
Table.V.5 Sensibilité des modèles de prédiction du mouvement sismique à la distance, magnitude et aux styles de la faille par le pourcentage du poids synaptique (P %)	191

Introduction générale



Dans toute étude d'évaluation déterministe ou probabiliste de l'aléa sismique, le choix de l'équation de prédiction du mouvement sismique (EPMS ou GMPE) est important. Si une GMPE (classiquement déterminée par régression empirique sur des jeux d'enregistrements accélérométriques) traduit correctement le comportement du mouvement sismique pour une région donnée, son utilisation doit alors conduire systématiquement à une évaluation correcte, non biaisée, de l'aléa sismique, et par voie de conséquence permettre une prévention adaptée du risque sismique.

Par ailleurs, les estimations d'aléa sismique sont aussi très sensibles au niveau des incertitudes associées à la GMPE utilisée, caractérisée par la variabilité aléatoire. La réduction de ces incertitudes est donc un objectif récurrent des travaux en sismologie de l'ingénieur, surtout celles liées à la variabilité aléatoire du signal sismique. Cette dernière est quantifiée par l'écart-type des résidus (prédictions – observations), communément connu sous le nom de sigma (σ). Un léger accroissement de σ engendre une augmentation significative des niveaux d'aléa sismique (généralement quantifiés par les ordonnées spectrales), et croissant avec la période de retour considérée, comme indiqué sur la Figure.1.

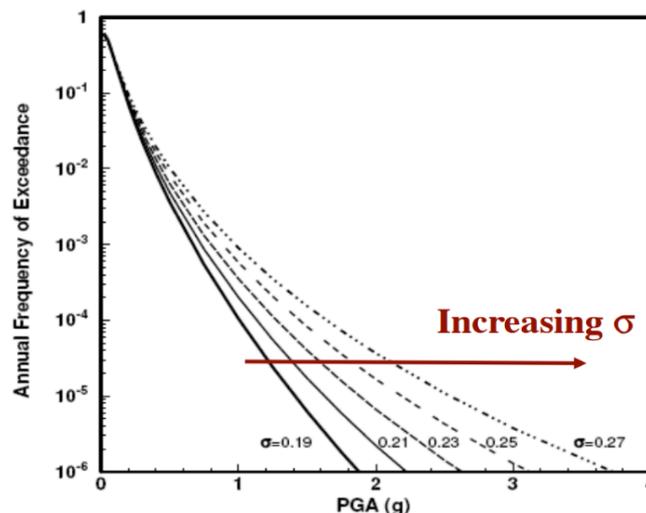


Figure.1 Impact de la valeur de σ sur la courbe d'aléa sismique obtenue par une approche probabiliste ([Bommer and Abrahamson, 2006](#))

Pour réduire la variabilité aléatoire dans les GMPEs, il faut premièrement s'assurer de la quantité et de la qualité de la base de données utilisée pour l'élaboration du modèle d'atténuation, deuxièmement choisir une forme fonctionnelle susceptible de capter les phénomènes physiques lors d'une secousse tellurique et troisièmement avoir suffisamment de paramètres sismiques (proxies) censés décrire les différents effets de la source au site d'investigation.

La qualité et de la quantité des données ont un fort impact sur la prédiction du mouvement sismique. Cela justifie les différents efforts consentis pour le déploiement d'appareils de mesures et pour des campagnes d'essais géologiques, géophysiques et géotechniques. Le fantastique essor des données accélérométriques illustré sur la Figure 2, et son impact sur l'évolution des GMPEs sont présentés en détail dans le chapitre I.

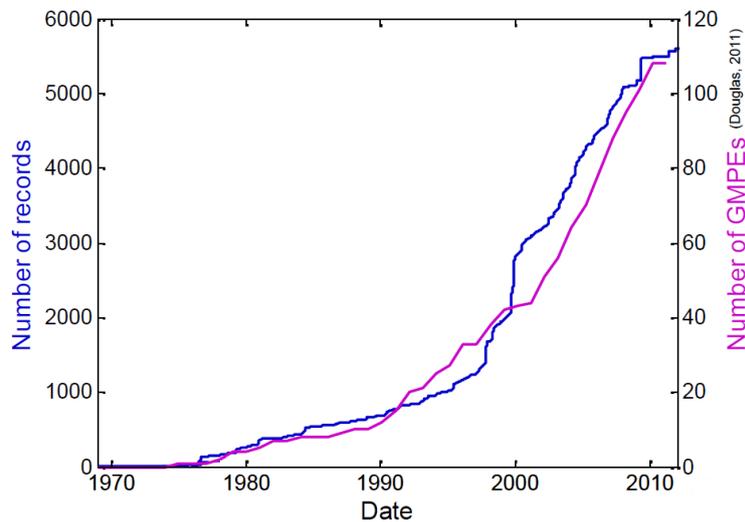


Figure.2 Evolutions temporelles du nombre de GMPEs et des enregistrements sismiques (courbe basée sur [Douglas and Edwards 2016](#))

Cette augmentation des données a contribué à l'évolution du nombre de GMPE. Cet essor important des données au cours des dernières décennies a donné naissance à plus de 450 GMPEs développées entre 1964 et 2016 ([Douglas, 2016](#)). Cependant, bien que les GMPEs deviennent de plus en plus complexes, et intègrent un nombre croissant de degrés de liberté par le biais de multiples paramètres et des formes fonctionnelles sophistiquées, elles ne conduisent qu'à une réduction timide de σ (figure.3). Plus de détails à ce sujet sont donnés dans le chapitre I.

Ce constat nous a motivé à chercher d'autres alternatives qui ne nécessitent pas d'hypothèses a priori sur les phénomènes physiques sous-jacents, tout en permettant des dépendances complexes. Pour contourner cette étape embarrassante, nous proposons au cours de cette thèse une approche totalement conduite par les données dans laquelle nous utilisons une seule et unique fonction (nommée fonction d'activation) considérée comme universelle et

parcimonieuse. Cette approche est appelée Réseaux de Neurons Artificiels (RNA). Tous ses aspects vont être présentés dans le chapitre II, où est élaboré un modèle de prédiction à partir de la base de données RESORCE.

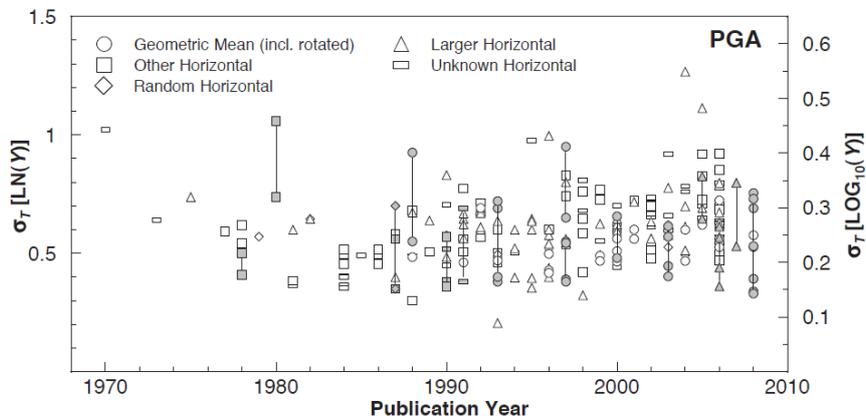


Figure.3 Evolution de σ durant les 30 dernières années (Strasser et al. 2009)

Cette approche neuronale, totalement contrôlée par les données, offre, en plus de l'établissement de modèles de prédictions sismiques, la possibilité de tester la pertinence des différents proxies utilisés pour décrire l'impact des paramètres de sites et de sources sur le mouvement sismique. L'objet du chapitre III est de faire une analyse critique des différents types de proxies de conditions de site en comparant les taux de réduction de la variabilité aléatoire qui leur sont associés dans les GMPEs, à savoir : V_{s30} mesurée, V_{s30} estimée, classe de V_{s30} mesurée, classe de V_{s30} estimée, la pente topographique « slope » (Figure.4) et sa classe. Toujours dans ce chapitre, nous comparons aussi l'effet de la base de données, type de proxy et la nature de proxy sur la médiane des GMPEs et sur la variabilité aléatoire. Dans cette partie de thèse, c'est la base de données NGA-West 2 qui est utilisée.

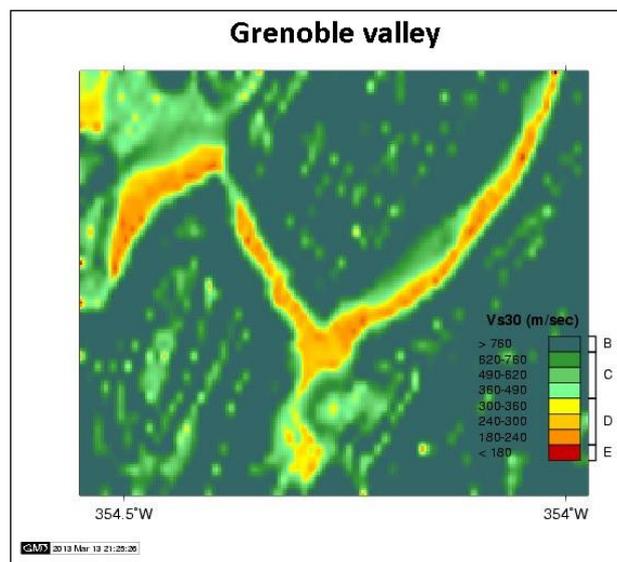


Figure.4 Carte isovitesse de la vallée de Grenoble donnant la classe NEHRP_Vs30 à partir de la pente topographique (slope) : calcul effectué à partir du : <https://earthquake.usgs.gov/data/vs30/>.

L'utilisation d'un seul proxy des conditions de site tel que V_{s30} ou la pente topographique semble insuffisante. Ces deux sont en effet, a priori, pertinents surtout pour des hautes fréquences (formation sédimentaire peu profonde), et leur utilisation pour des sols plus profonds (basses fréquences) engendre probablement une prédiction plus incertaine. Pour contourner ce problème, nous proposons dans le chapitre IV l'utilisation de plus d'un seul proxy de site dans une GMPE. Le chapitre IV se divise en deux parties. La première porte encore sur une analyse comparative des proxies de site, en ajoutant aux V_{s30} et la pente topographique déjà évoqués dans le chapitre III, la fréquence de résonance f_0 et la profondeur H_{800} à laquelle la vitesse des ondes de cisaillement "S" atteint 800 m/s. La deuxième partie s'intéresse plus spécifiquement aux non-linéarités dans la réponse de site, et à la capacité des proxies habituels à rendre compte de ces non-linéarités qui sont représentées par le rapport RSR_{NL-L} de [Régnier et al. \(2016\)](#), comparant la réponse fréquentielle sous forte sollicitation à la réponse (linéaire) sous faible sollicitation, permettant une caractérisation assez fine des réponses non-linéaires. Cette dernière étude est effectuée avec 5 proxies de sites et 2 proxies qui caractérisent le mouvement sismique (MS). Les deux parties de ce chapitre nécessitent une connaissance approfondie des conditions de site et/ou de la réponse du site, qui n'existe à l'heure actuelle que pour la seule base de données KiK-Net.

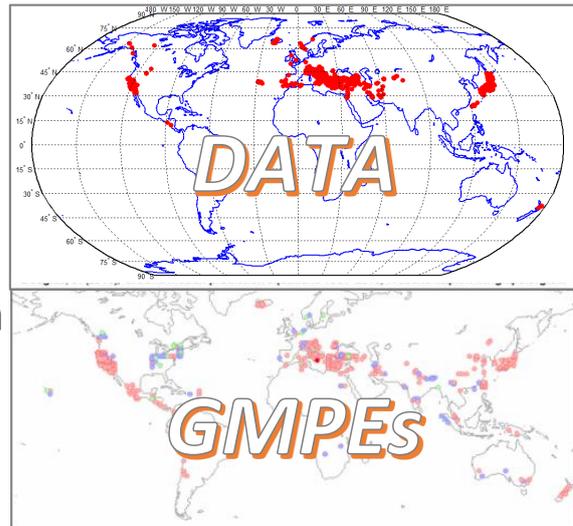
Le dernier chapitre (V) se concentre sur les paramètres et les effets de source, avec une analyse de l'impact de la prise en compte du mécanisme au foyer (FM) et de la chute de contrainte sismique ($\Delta\sigma$) [en plus de la magnitude de moment M_w , qui est plus naturellement associée aux basses fréquences] sur la variabilité aléatoire. L'étude de l'effet FM est menée avec trois jeux de données : RESORCE, NGA-West 2 et KiK-Net, car ces informations sont disponibles pour les 3 bases de données. En revanche, l'étude de l'impact et de l'interdépendance entre la chute de contrainte sismique et la magnitude est réalisée uniquement sur le jeu de données NGA-West 2. Les MS obtenues à partir de ce jeu de données représentent une référence à l'étude comparative menée entre les différents modèles de simulation à $\Delta\sigma$ variable et $\Delta\sigma$ constant.

Il est important de signaler que, tout au long de cette thèse, les prédictions des modèles neuronaux sont très souvent comparées à celles des GMPEs classiques dérivées sur les mêmes bases de données. Le but est de valider cette approche neuronale, qui est nouvelle dans la communauté des GMPEs, et rencontre un certain nombre de réticences "a priori" en raison de sa réputation de "boîte noire" à faible contenu physique. La première comparaison est présentée dans le chapitre II pour la base de données RESORCE, où les modèles obtenus par approche

neuronale sont comparés à ceux obtenus par [Akkar and Bommer \(2010\)](#), [Akkar and Çagnan \(2010\)](#) et [Bindi et al. \(2011\)](#). Un complément de cette comparaison va être donné en annexe 2 ([Douglas et al. 2014](#)). Une autre série de comparaisons est effectuée dans le chapitre III entre le modèle neuronal et le modèle de [Boore et al. \(2014\)](#) avec la base de données NGA-West 2.

Chapitre I

Le fantastique essor des données accélérométriques et son impact sur l'évolution de la prédiction des mouvements sismiques



Introduction

Depuis toujours, l'être humain se sert de la nature pour comprendre, expliquer des phénomènes et trouver des solutions parfois spectaculaires à des problèmes bien spécifiques. Nos observations, traitements, analyses et synthèses nous donnent la possibilité d'évoluer au cours du temps : création-invention-innovation. Pour arriver au stade de la compréhension, une bonne observation des phénomènes demeure indispensable. De nos jours, les observations peuvent être effectuées de manières qualitative et quantitative.

Le moyen d'observation qualitatif le plus fréquent reste l'œil humain. Par exemple, les dégâts dus à un tremblement de terre peuvent être observés, recensés et classés. Dans ce contexte, des échelles ont été créées. Les plus célèbres sont celles de Mercalli et MSK. L'échelle de Mercalli a été développée en 1902 et modifiée en 1931. Elle indique l'intensité d'un séisme (effets ressentis) sur une échelle graduée de I à XII. Tandis que, l'échelle MSK (du nom de ces inventeurs : Medvedev, Sponheuer, et Karnik) a été proposée en 1964 et comporte 12 degrés. MSK classe les séismes en fonction de leurs effets à un endroit donné. Cependant, ces échelles sont subjectives.

Pour quantifier avec exactitude le mouvement du sol (énergie et intensité), il faut avoir recours à des mesures. Pour quantifier l'énergie et l'intensité d'un séisme, nous utilisons le sismographe et/ou l'accélérographe comme appareils de mesures. Ces deux outils permettent d'enregistrer les vibrations du sol qui sont ensuite étudiées. L'enregistrement produit par les sismographes

s'appelle un sismogramme (déplacement en fonction du temps) et celui produit par un accélérographe se nomme accélérogramme (accélération en fonction du temps).

De nos jours, ces deux dispositifs sont devenus des outils indispensables lors des différentes étapes de l'analyse des mouvements sismiques : de la source sismique à l'ouvrage. Les données ainsi fournies permettent, donc de développer des modèles de prédiction des mouvements sismiques qui permettent d'avoir une bonne estimation du chargement sismique auquel ces structures sont soumises.

Le déploiement massif des sismographes et accélérographes a permis de construire différentes bases de données de plus en plus fournies. En général, le sous-ensemble qui constitue ces bases de données est l'accélération maximale du sol (PGA), la vitesse maximale du sol (PGV) et les pseudo-accélérations (PSA). De tels indicateurs permettent, alors, d'établir des équations de prédictions du mouvement sismique (GMPE : Ground-Motion Prediction Equation). Ces modèles sont conçus en utilisant, traditionnellement, des méthodes de régression empirique par ajustement aux observations expérimentales à partir des magnitudes, distances, profondeur focale, type de faille et proxies des conditions de site.

L'objectif de ce chapitre est de mesurer l'impact de l'amélioration des différentes bases de données récentes sur l'évolution des GMPEs. Pour ce faire, nous commençons notre investigation par rappeler le principe de fonctionnement des appareils de mesure des mouvements sismiques et par un aperçu sur les différentes stations d'enregistrement historiques marquant l'évolution des bases de données de mouvements forts « strong motion database ». Nous montrons ensuite avec des exemples des réseaux instrumentaux et bases de données très utilisés par la communauté et notamment les jeux de données utilisées dans ce travail de thèse. Nous clôturons ce chapitre par l'impact de la base de données, le choix de la forme fonctionnelle et les paramètres constituant une GMPE sur la réduction de la variabilité aléatoire du mouvement sismique.

I.1. Paramètres des mouvements forts

I.1.1. Appareils de mesure des mouvements forts

Un sismographe (sismomètre) permet de restituer les mouvements sismiques sous forme de déplacement, vitesse ou accélération. Le capteur de mouvement le plus utilisé en analyse dynamique est l'accélérographe (accéléromètre), qui est un appareil muni d'un dispositif mesurant l'accélération dont le principe est un oscillateur à un seul degré de liberté.

La figure.I.1 illustre le principe de base sur lequel se fonde le fonctionnement des instruments sismiques. Le système est installé dans une enceinte fixée sur la surface dont le mouvement est à étudier ; la réponse se mesure à partir du déplacement relatif de la masse : $y(t)$.

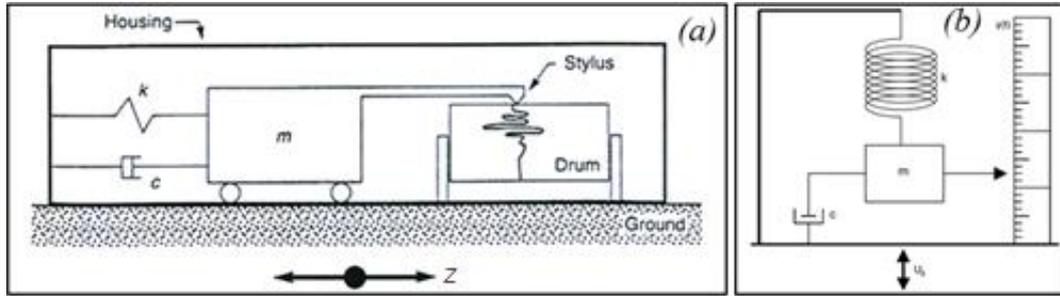


Figure.I.1 Schémas de principe de base de fonctionnement d'un accélérographe. (a) représente la mesure du déplacement de la composante horizontale et (b) la mesure de la composante verticale. m , k et c représentent la masse, la rigidité et l'amortissement de l'accélérographe, respectivement (Kramer, 1996).

Si le support sur lequel l'instrument est installé se déplace « harmoniquement » avec une amplitude d'accélération égale à :

$$\ddot{Z}(t) = \ddot{Z}_0 \cdot \sin(\bar{\omega}t) \quad (\text{I.1})$$

Le chargement effectif auquel est soumise la masse est :

$$P_{eff}(t) = -m \cdot \ddot{Z}_0 \sin(\bar{\omega}t) \quad (\text{I.2})$$

L'équation du mouvement dynamique du système est du type :

$$m \cdot \ddot{y} + c \cdot \dot{y} + k \cdot y = -m \cdot \ddot{Z}(t) = P_{eff}(t) \quad (\text{I.3})$$

m est la masse du système, c la constante d'amortissement et k la rigidité du ressort.

La réponse dynamique permanente trouvée après résolution de l'équation I.3 a pour amplitude :

$$\rho = \frac{m\ddot{Z}_0}{k} D \quad (\text{I.4})$$

Tel que :

$$D = \frac{1}{\sqrt{(1-r^2)^2 + (2\xi r)^2}} \quad (\text{I.5})$$

D représente le facteur d'amplification dynamique et est représenté graphiquement par la figure.I.2 avec un amortissement compris entre 0 et 1. r est le rapport des fréquences $r = \bar{\omega}/\omega_0$, où ω_0 représente la pulsation de résonance du dispositif qui est égale à $\omega_0 = \sqrt{k/m}$ et $\bar{\omega}$ La pulsation d'excitation sismique.

Il apparaît clairement sur cette figure que, pour un coefficient d'amortissement $\xi = 0.7$, la valeur de D est quasiment constante (égale à 1) pour un r compris entre [0 et 0.6]. C'est dans cette plage de r que les mesures sont effectuées. Dans cet intervalle, on obtient seulement la réponse du site où l'appareil est installé. Les accélérographes ont en général, une fréquence de résonance f_0 supérieure à 25 Hz. Si $\xi = 0.7$ et $f_0 = 50\text{Hz}$ donc nous pouvons enregistrer précisément le mouvement du sol pour une gamme de fréquences allant de $f = 0-25\text{Hz}$ (ex. $0 < r < 0.5$)

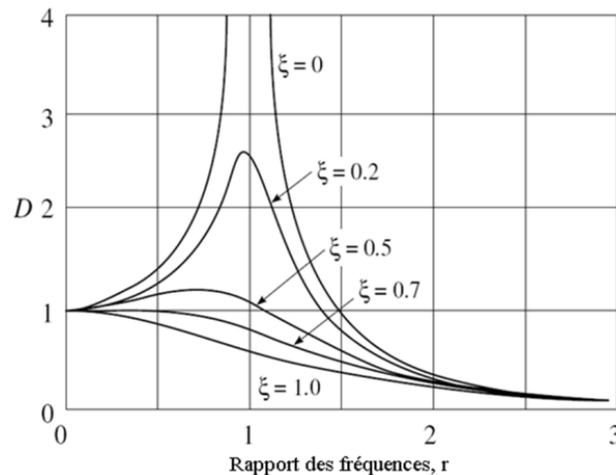


Figure.I.2 Variation du facteur d'amplification dynamique avec le rapport des fréquences

L'accélérogramme $\ddot{Z}(t)$ est donc une fonction à variations rapides et irrégulières autour de zéro dont la durée totale t_T peut varier de quelques secondes à quelques dizaines de secondes. Généralement, cette fonction se décompose en une phase initiale pour laquelle les accélérations sont faibles, une phase forte suivie d'une phase modérée et une phase progressive de retour à une accélération finale nulle ou proche du bruit de fond.

Nous pouvons extraire du signal sismique des informations sur l'accélération, la vitesse et le déplacement maximal : PGA, PGV et PGD respectivement ainsi que le spectre de Fourier (FAS) et le pseudo-spectre de réponse (PSA). Les paramètres du mouvement du sol les plus utilisés en pratique sont les PGA, PGV et les PSA. Ces trois paramètres feront l'objet des méthodes de prédiction de cette thèse.

I.1.2. Accélération maximale et spectre de réponse en accélération

L'un des paramètres les plus utilisés à l'heure actuelle pour le dimensionnement des ouvrages est l'accélération maximale (souvent noter PGA : Peak Ground Acceleration). Cette quantité correspond à la valeur maximale de l'accélération du sol atteinte, en un site donné, au cours du séisme. Bien que, très utilisé en raison de sa simplicité, le PGA décrit principalement les hautes fréquences, ce paramètre n'est pas le paramètre le plus corrélé avec les dommages.

L'exemple donné sur la figure.I.3 montre ainsi que tous les accélérogrammes ont la même valeur de $PGA=0.17\text{ g}$. En revanche, les spectres de réponse présentent des allures différentes. Cela veut dire que l'utilisation des ordonnées spectrales en plus de PGA est nécessaire.

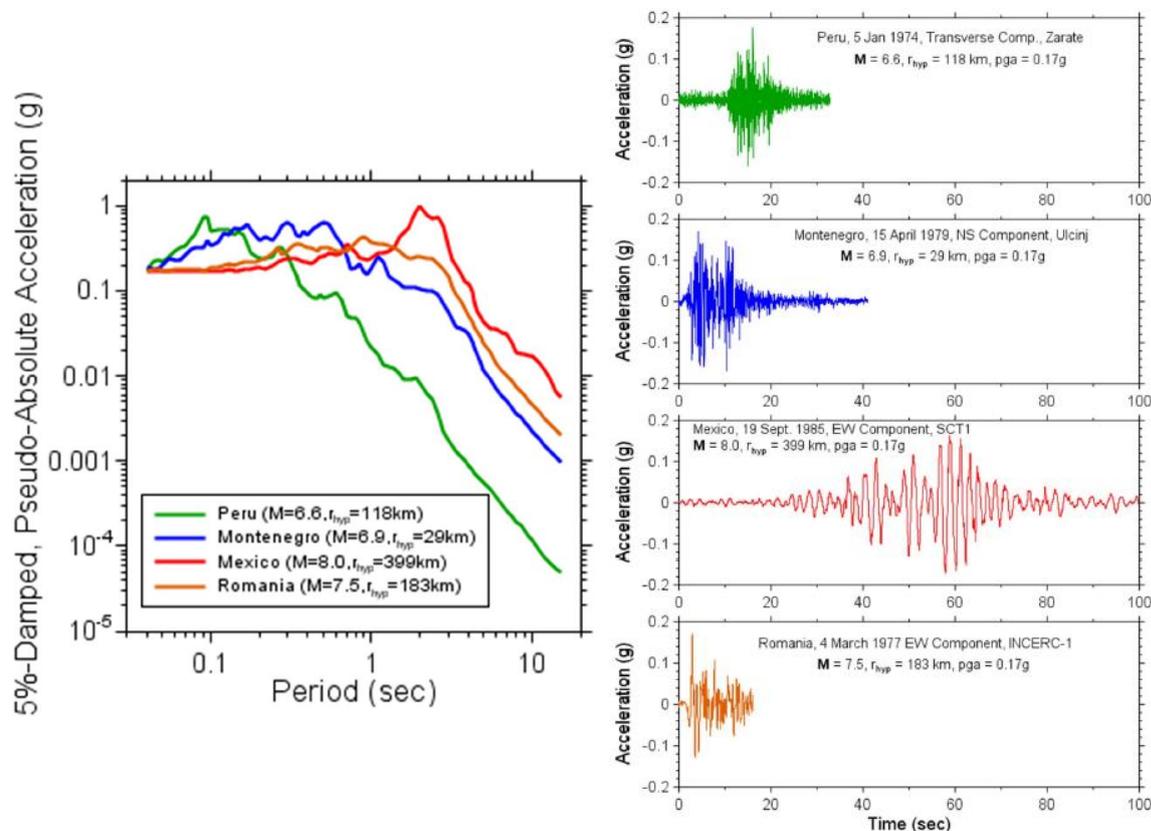


Figure.I.3 Différents accélérogrammes et spectres de réponse pour la même valeur de $PGA=0.17\text{ g}$ (Boore, 2015).

Par définition, un spectre de réponse correspond à l'amplitude maximale de la réponse d'un oscillateur à un seul degré de liberté (OSDDL) de période T et d'amortissement ξ . Lorsque l'on fait varier la période T de l'oscillateur, on associe à l'accélérogramme $a(t)$ une courbe spectrale $S_a(T, \xi)$: le spectre de réponse à la période T pour un amortissement ξ (figure.I.4). Nous montrons dans cette figure un exemple de traces en accélération, vitesse, et spectre de réponse en accélération de la station « MYGOO4 » du réseau accélérométrique KiK-Net. L'événement est celui du Tohoku (11/03/2011) avec une magnitude (JMA) de 9.0, une profondeur focale égale à 24 km et d'une distance épacentrale $R_{epi}=183\text{ km}$.

L'intérêt principal d'un spectre de réponse vient de ce qu'en première approximation, un immeuble ou un ouvrage peut être assimilé à un oscillateur à un seul degré de liberté (OSDDL). La connaissance de sa période propre T (reliée au type de structure, à la taille et aux propriétés du matériau constitutif), de son amortissement (relié aux matériaux et aux dispositions constructives) va donc permettre d'accéder aux accélérations, vitesse et déplacement maximaux

subis par le centre de gravité de l'ouvrage, et par voie de conséquence aux forces et contraintes à l'intérieur de la structure. De plus, ce spectre possède la propriété remarquable d'avoir une ordonnée à période nulle ($T = 0$ sec) égale à l'accélération maximale du sol PGA.

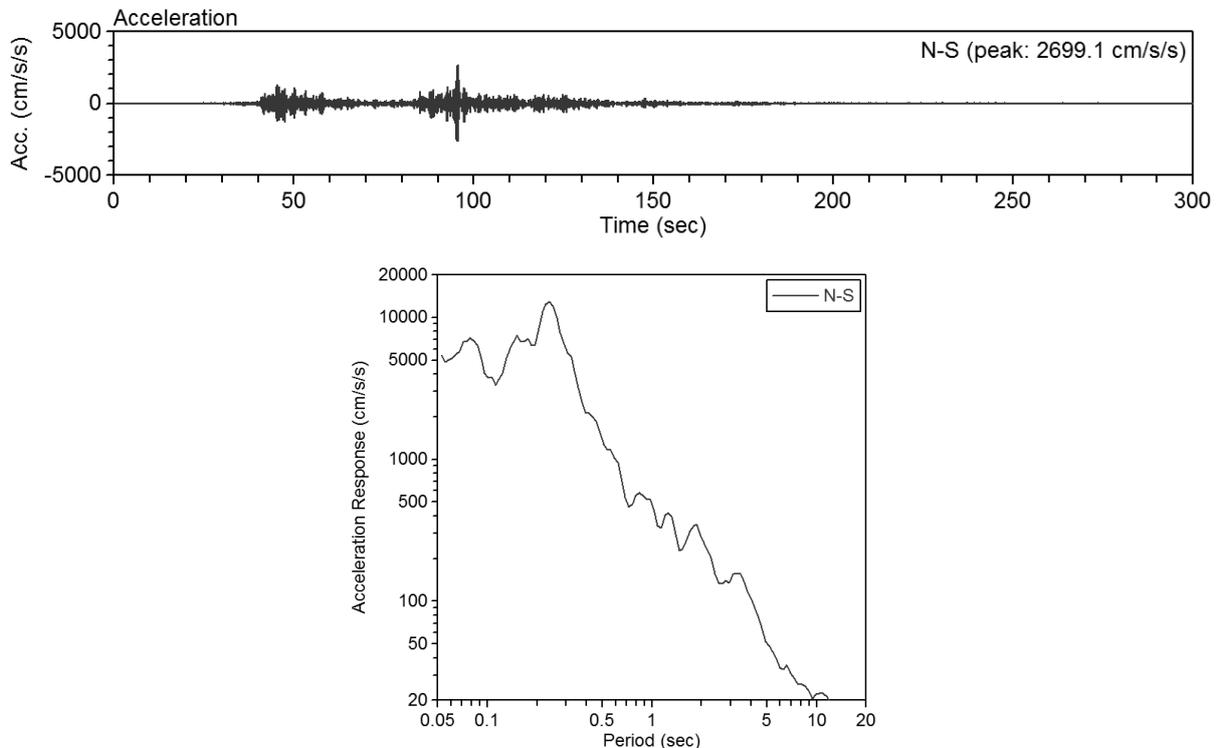


Figure.I.4 Exemple des traces en accélération, vitesse, déplacement et spectre de réponse de la station « MYGOO4 » du KiK-Net. L'enregistrement est celui du Tohoku (11/03/2011) avec une magnitude (JMA) de 9.0, enregistré à une distance épiscoptrale égale à 183 km.

De ce fait, la quasi-totalité des règles parasismiques actuelles décrivant les mouvements sismiques sous la double forme d'une accélération maximale PGA et d'un spectre de réponse en accélération élastique normalisé $S_a(T, \xi)/PGA$.

I.1.2.1. Procédure d'élaboration d'un spectre de réponse

Le spectre de réponse permet de caractériser un séisme par l'effet qu'il produit sur un équipement. Pour cela, nous calculons l'effet de l'accélérogramme (du signal sismique) sur un OSDDL ; ceci pour différentes valeurs de la fréquence de résonance et de l'amortissement. Un OSDDL se caractérise par : une masse m , une rigidité k et un coefficient amortissement ξ (figure.I.5). Sa fréquence de résonance est :

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (\text{I.6})$$

La réponse maximale de ce système (accélération maximale subie par la masse) donne un point du spectre de réponse (figure.I.5). En faisant varier la période propre (ou la fréquence de

résonance) nous obtenons la courbe de la fonction $a_{max} = f(f_0)$ (figure.I.5) qui décrit le spectre de réponse caractérisant la sévérité de l'onde sismique pour un amortissement donné. Le spectre de réponse est principalement spécifié dans les règles parasismiques pour une application dans les directions horizontales.

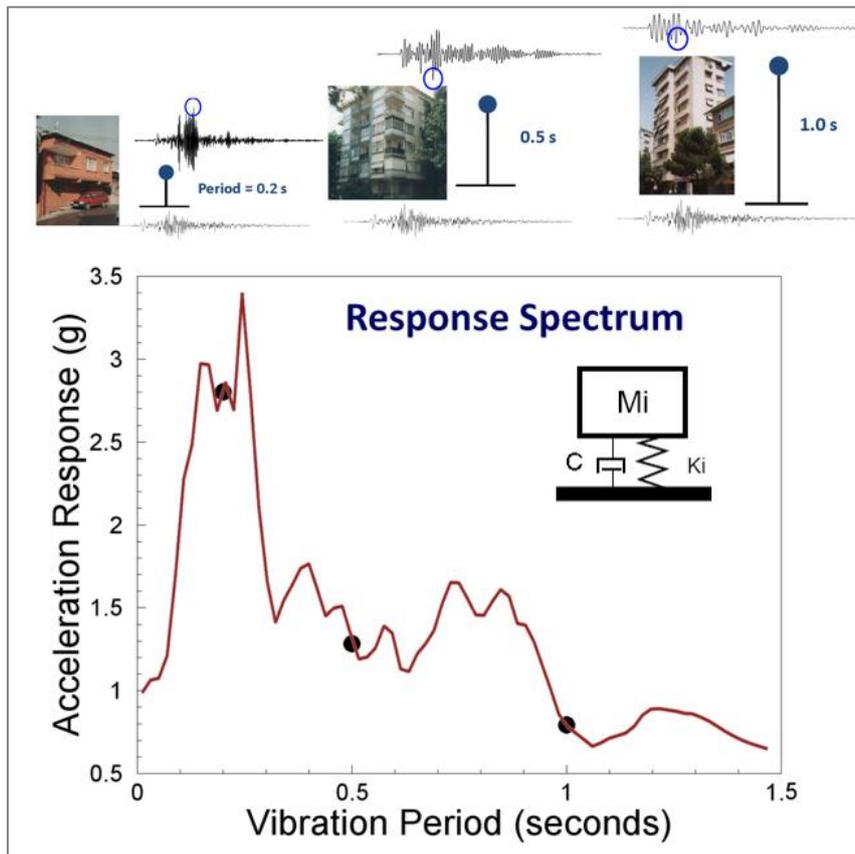


Figure.I.5 Construction du spectre de réponse au séisme (divers k/m avec c constant) : [Boore, 2015](#).

À partir du spectre de réponse, l'effort élastique maximal est défini par la relation suivante :

$$F_{s\ max} = K \cdot S_d(T, \xi) = M \cdot \omega^2 \cdot S_d(T, \xi) \quad (I.7)$$

S_d représente le spectre de réponse en déplacement.

$$\text{Dont} \quad PSA(T, \xi) = \omega^2 \cdot S_d(T, \xi) \quad (I.8)$$

$PSA(T, \xi)$ représente le pseudo-spectre en accélération.

À ces deux caractéristiques du mouvement du sol (PGA et PSA) s'ajoute un autre paramètre scalaire pertinent : PGV (vitesse maximale du sol).

I.1.3. Vitesse maximale du sol (PGV : Peak Ground Velocity)

Il existe de nombreuses utilisations de la vitesse maximale du sol V_{max} (appelé encore PGV pour : Peak Ground Velocity) en sismologie et en génie parasismique. Le PGV permet

l'estimation, entre autres, de l'intensité macrosismique et des dommages structurels ([Bradley, 2012](#)). [Wu et al \(2003\)](#) ont constaté que les dommages causés par les tremblements de terre (notamment pour les conduites enterrées) semblent beaucoup mieux corrélés avec le PGV qu'avec le PGA. Cette quantité est obtenue en intégrant une fois temporellement le signal d'accélération.

Pour des séismes destructeurs, le PGV va de quelques cm/s à plusieurs dizaines de cm/s, et peut dépasser le m/s à proximité immédiate de la rupture. Alors que le PGA correspond à la partie haute fréquence du signal, et est plus représentatif pour les ouvrages raides à courtes périodes, le PGV est représentatif pour des fréquences intermédiaires entre 1 à 2 Hz, et sera un meilleur indicateur de nocivité pour les structures les plus souples ([DGPR, 2014](#)).

Ces paramètres : PGA, PGV et les ordonnées spectrales (PSA : pseudo-spectre de réponse) sont les paramètres du mouvement du sol à prédire dans cette thèse et sont largement utilisés par les ingénieurs et les chercheurs, notamment dans les études déterministes et probabilistes de l'aléa sismique DHSA et PSHA. Dans ces deux méthodes, l'utilisation d'une GMPE adaptée est plus que souhaitable. La robustesse de la GMPE utilisée dépend essentiellement de sa forme fonctionnelle et du jeu de données utilisé pour son élaboration. C'est pour cette raison que nous présentons dans ce qui suit les différentes bases de données existantes et leur évolution temporelle.

I.2. Évolution de la collection des mouvements forts

Dans cette section, une revue bibliographique sur l'évolution quantitative et qualitative des données du mouvement fort est abordée. Trois bases de données sont présentées. Il s'agit principalement de la base de données RESORCE (Reference database for seismic ground-motion in Europe), de la base de données de [Dawood et al. \(2016\)](#) (issue principalement des données du réseau accélérométrique KiK-Net) et de la base de données NGA-West 2 (Next Generation Attenuation deuxième generation). Cette dernière représente une compilation de données récoltées de par le monde. Ces trois bases de données représentent l'outil de base de ce projet de thèse.

I.2.1. Région euro-méditerranéenne

Les tentatives de collection et compilation des données des mouvements forts de l'Europe et Moyen-Orient ont débuté dans les années 70 à Imperial Collège ([Ambraseys 1978](#)). Le travail entrepris à Imperial Collège a ensuite été soutenu par les agences gouvernementales

du Royaume-Uni et la Commission européenne ([Bommer and Douglas 2004](#)). L'objectif de ce travail était l'évaluation cohérente des tremblements de terre, des métadonnées les décrivant et le traitement uniforme des données du mouvement fort. La base de données a été diffusée par la suite via le site Internet http://www.isesd.hi.is/ESD_Local/frameset.htm (Internet Site for European Strong-Motion Data web page : ISESD). Dans ce site de la Commission européenne, les données sismiques ont été archivées sous forme de fichiers non corrigés et corrigés. Les données sont issues des tremblements de terre européens et des zones adjacentes. Cependant, la non-disponibilité des profils de site dans presque toutes les stations de cette base représente une faiblesse importante pour ce qui concerne les études des mouvements forts.

En l'absence d'études spécifiques, les métadonnées du tremblement de terre (par exemple, la localisation épacentrale, profondeur focale et magnitude) ont été principalement tirées du Bulletin du Centre sismologique international (www.isc.ac). Cependant, les stations ont été classées en utilisant la classification donnée par [Boore et al. \(1993\)](#) basée sur V_{S30} qui est la vitesse moyenne des ondes de cisaillement des 30 premiers mètres du profil de sol.

Presque tous les enregistrements de mouvements forts dans ISESD étaient filtrés en utilisant un filtre elliptique avec des fréquences de coupure passe-haut et passe-bas de 0.25 au 25 Hz respectivement. Un sous-ensemble d'ISESD a été retraité en utilisant un filtre Butterworth bidirectionnel (acausal) avec des fréquences de coupures ajustées individuellement pour chaque accélérogramme. Le sous-ensemble a été regroupé dans un CD-ROM de données de 1068 accélérogrammes ([Ambraseys et al. 2004](#)).

Les efforts de compilation des données des mouvements forts de la base de données ISESD ont été suivis par d'importants projets nationaux en Europe et dans les régions environnantes. Parmi ces projets, le projet ITalian ACcelerometric Archive Project (ITACA; <http://itaca.mi.ingv.it>; [Luzi et al. 2008](#)) et le projet Turc (T-NSMP; <http://kyh.deprem.gov.tr/>; [Akkar et al. 2010](#)). Le projet ITACA compile un total de 2182 accélérogrammes de 1004 événements ([Luzi et al. 2008](#)), tandis que T-NSMP répertorie 4607 enregistrements de mouvements forts de 2996 séismes enregistrés sur 209 stations ([Akkar and Bommer. 2010](#)). Les deux projets (ITACA et T-NSMP) ont amélioré la caractérisation des stations en réévaluant la vitesse des ondes de cisaillement des profils et les informations des colonnes du sol par des techniques d'exploration invasives ou non invasives ([Sandikkaya et al. 2010](#)). Un effort similaire a également été lancé en Grèce pour archiver les enregistrements grecs archivés par ITSAK (<http://www.itsak.gr/>; [Theodulidis et al, 2004](#)) dans le cadre du projet HEAD (HElIenic Accelerogram Database).

L'évolution continue grâce au projet Seismic Hazard HARMONIZATION in Europe Project (SHARE ; <http://www.share-eu.org>). Ce projet a été subventionné par la Commission européenne. Ce projet a permis une compilation des données de mouvements forts (Yenier et al. 2010) principalement européennes (les données ISESD, ESMD, ITACA et T-NSMP sont parmi cette base de données). Un des objectifs de ce projet était de tester les performances des GMPEs pour le calcul de l'aléa sismique en Europe. Nous reviendrons sur cette base de données plus loin dans ce chapitre lors de l'étude de l'évolution de la quantité de données associée de l'amélioration des GMPEs. Un des points faibles de la base développée au sein de SHARE est cependant l'hétérogénéité du traitement.

Pour surmonter cette faiblesse et pour améliorer les métadonnées, un projet complémentaire a été développé en 2011. Il s'agit du projet RESORCE (<http://www.resorce-portal.eu>) pour Reference database for seismic ground-motion in Europe. RESORCE représente une compilation des enregistrements des mouvements sismiques européens de grande qualité, associée à une information complète et uniforme des métadonnées associées. Le résultat est une base de données homogène. RESORCE est l'un des produits du projet SIGMA (Seismic Sol Evaluation Motion) dont l'objectif principal a été d'améliorer les méthodes d'évaluation de l'aléa sismiques en France et les régions voisines (figure I.6), avec la caractérisation réaliste des incertitudes aléatoires et épistémiques. La structure principale de RESORCE se compose de deux blocs principaux : (1) tremblement de terre, métadonnées et stations, et (2) les données accélérométriques. Intrinsèquement, ces deux blocs sont reliés les uns aux autres et assemblés avec les mêmes sources de référence. La figure I.7 résume la structure globale de la base de données RESORCE. ISESD et son sous-ensemble ESMD sont considérés comme les principales sources de métadonnées établies avant 2004 de RESORCE : M_w , coordonnées épicentrales, profondeurs focales, styles des failles, géométries de la faille, proxies des conditions de site, coordonnées des stations, différentes mesures de distances source-site, etc. Pour les métadonnées italiennes, ITACA a été utilisé. Par ailleurs, ISESD et ESMD sont adoptés aussi pour les tremblements de terres turques survenues avant 2004. Les informations des tremblements de terre et les stations de références supplémentaires, autres qu'ISESD et ESMD, (voir annexe 7 tableau 5 et 6 Akkar et al. 2014) sont principalement prises en compte pour les tremblements de terre après 2004. Plus de détails sur le sous-ensemble utilisé dans cette thèse sont donnés dans le chapitre II.

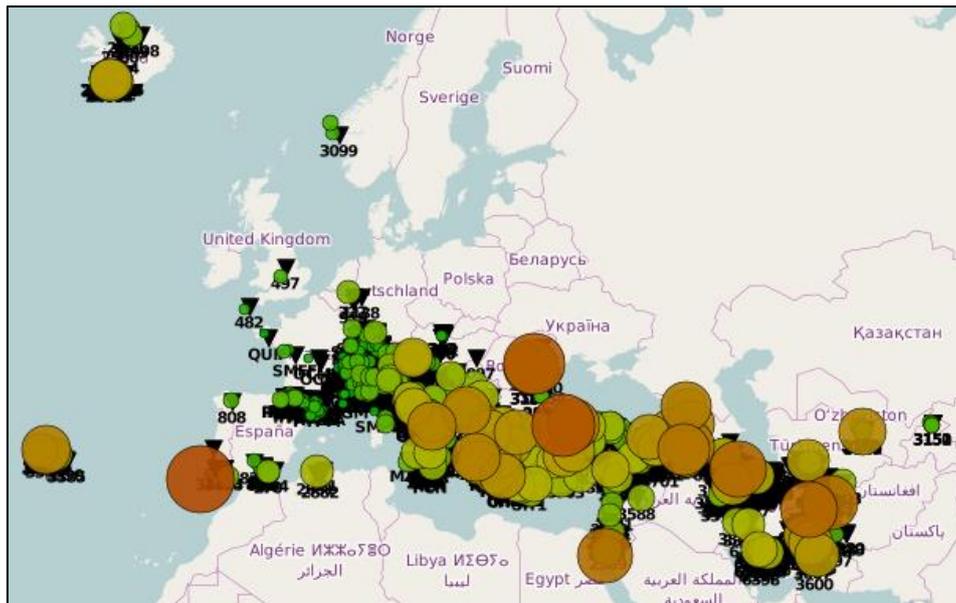


Figure.I.6 Distribution des épencentres des évènements de la base de données RESORCE (<http://www.resorce-portal.eu/>)

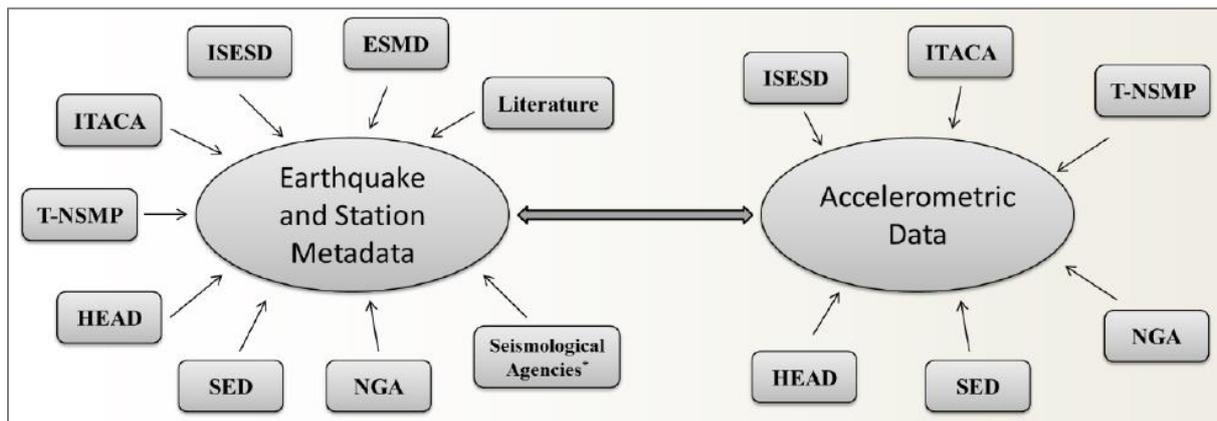


Figure.I.7 Structure et ressources de la base de données RESORCE (<http://www.resorce-portal.eu/>)

Par ailleurs, et dans l'optique de promouvoir la rapidité d'accès à l'information liée à l'événement sismique, le portail du mouvement fort rapide RAW (RRSM) a été récemment élaboré et a commencé sa diffusion en juin 2014. Le portail RRSM permet aux utilisateurs d'avoir des informations décrivant les tremblements de terre, des paramètres de l'amplitude maximale du sol, l'amplitude spectrale. Le portail permet de télécharger les formes d'onde du tremblement de terre (de la région euro-méditerranéenne) quelques minutes après l'événement sismique. Les informations des tremblements de terre sont fournies par l'EMSC. Toutes les informations sur la forme d'onde sismique sont disponibles depuis le portail d'ORFEUS (<http://www.orfeus-eu.org>) dont le traitement est entièrement automatisé et disponible en temps réel. Un retraitement hors ligne a été effectué pour tous les événements qui ont des magnitudes $M \geq 4.5$ et qui se sont produits depuis janvier 2005 et pour tous les événements de Magnitude $M \geq 3.5$ depuis janvier 2012. Le portail RRSM est le premier système européen qui fournit des

informations des mouvements forts en temps réel. Il est considéré comme un outil important pour les scientifiques et les ingénieurs. Plus d'informations sur ce portail sont données par [Cauzzi et al. \(2016\)](#).

La base de données RESORCE est utilisée dans le chapitre II de la présente thèse. Cependant, notre travail de thèse ne s'articule pas uniquement sur les données RESORCE, mais plus encore sur d'autres jeux de données tels que celui tiré du réseau accélérométrique japonais. Dans la section suivante, nous allons donner un aperçu de cette base de données.

I.2.2. Données accélérométriques japonaises

Le Japon a souvent souffert de tremblements de terre. Les principaux séismes japonais destructeurs du XXe siècle sont énumérés dans le tableau.I.1. La figure.1.8 montre les épicentres de ces mouvements telluriques.

L'observation des mouvements forts dans le domaine de l'ingénierie a commencé au Japon dans les années cinquante. Des professeurs, des chercheurs et des ingénieurs ont créé un comité afin de développer les instruments d'enregistrement de mouvements forts et de promouvoir des projets d'observation et d'études connexes. Les premiers progrès ont été réalisés en accumulant les documents (l'aspect théorique) et en améliorant la technologie instrumentale. Le premier instrument prototype a été fabriqué en 1953 et a été nommé SMAC (initiales du comité). Dans les premiers stades de l'observation du mouvement fort, la plupart des sites ont été instrumentés et ont été opérés par les Universités et les instituts de recherche nationaux. Le comité observation du mouvement fort des tremblements de terre a été créé en 1956 afin d'entretenir les instruments de mesure, de traiter les données, et de publier les enregistrements ([Kashima, 2009](#)).

Avec le tremblement de terre de Kobe (1995), la situation de l'observation de mouvement fort a subi des changements drastiques. Plusieurs projets ont été planifiés et réalisés afin de renforcer le réseau d'observation des mouvements forts. L'agence météorologique japonaise (JMA) a déployé environ 600 stations mesurant l'intensité sismique à travers le Japon. L'institut national de recherche pour les sciences de la terre et de la prévention des catastrophes (NIED) avait créé le réseau national "K-Net" avec 1000 sites d'observation. NIED construit également un autre réseau "KiK-Net". Chaque préfecture est équipée d'un système de réseau d'information, pour piloter les données de toutes les municipalités. Environ 2600 stations mesurant l'intensité

sismique ont été installées dans les municipalités. Ces réseaux mis à jour constamment peuvent diffuser rapidement les informations sismiques presque en temps réel.

Table.I.1 Liste des séismes destructeurs du XXe siècle recensés jusqu'à 1995 ([Kashima, 2009](#)).

Date	Epicentre	M _{JMA}	Casualités
01/09/1923	Southern Kanto	7.9	Mort/Perdu : >142,000, feu
23/05/1925	Northern Tajima	6.8	Mort : 428
07/03/1927	N.W. Kyoto	7.3	Mort : 2,925
26/11/1930	N. Izu	7.3	Mort : 272
03/03/1933	Off Sanriku	8.1	Mort / Perdu : >3,064, Tsunami
10/09/1943	Tottori	7.2	Mort : 1,083
07/12/1944	Off Tokaido (Tonankai)	7.9	Mort / Perdu : 1,223
13/01/1945	S. Aichi Pref. (Mikawa)	6.8	Mort : 2,306
21/12/1946	Off Nankaido	8.0	Mort : 1,330
28/06/1948	Fukui	7.1	Mort : 3,769
16/06/1964	Off Niigata	7.5	Mort : 26
16/05/1968	Off Aomori Pref.	7.9	Mort : 52, Blessé : 330
09/05/1974	Off Izu Peninsula	6.9	Mort : 30, Blessé : 102
14/01/1978	Near Izu-Oshima	7.0	Mort : 25, Blessé : 211
12/06/1978	Off Miyagi Pref.	7.4	Mort : 28, Blessé : 1,325
26/05/1983	Central Japan Sea	7.7	Mort : 104, Blessé : 163, Tsunami
14/09/1984	E. Nagano Pref.	6.8	Mort : 29, Blessé : 10
15/01/1993	Off Kushiro	7.8	Mort : 2, Blessé : 928
12/07/1993	S.E. Off Hokkaido	7.8	Mort:202, Perdu:29, Blessé :305, Tsunami
04/10/1994	E. Off Hokkaido	8.1	Blessé : 343
28/12/1994	Far Off Sanriku	7.5	Mort : 3, Blessé : 784
1/01/1995	S. Hyogo Pref. (Kobe)	7.2	Mort : >5500, Blessé : >30000

Réseau d'intensité sismique JMA : <http://www.jma.go.jp/jma/indexe.html>

L'organisme JMA est responsable des prévisions : tsunamis, activités volcaniques et tremblement de terre. Après un tremblement de terre, JMA traite immédiatement les données d'observation et publie l'information sur l'épicentre, la magnitude et la répartition de l'intensité sismique au public (à travers les médias) et aux organismes de prévention des catastrophes. Les informations des 2000 stations gouvernementales d'intensité sismique sont compilées.

Réseau K-Net : réseau Kyoshin network (<http://www.kyoshin.bosai.go.jp>)

K-NET (réseau Kyoshin) est un réseau d'accélérographes de mouvements forts qui couvre tout le territoire Japonais et se compose de plus de 1000 stations d'observation (en surface) tous les 20 km ce qui permet ainsi d'avoir des enregistrements en zone épacentrale, et ce même pour des séismes de magnitudes dépassant 7. K-NET est exploité par NIED. Il est opérationnel depuis juin 1996.

À chaque station K-NET, un accélérographe est installé. Ces derniers (K-NET95) enregistrent trois composantes et échantillonnent l'information à 100 Hz. Ils enregistrent un événement si

le mouvement perçu dépasse une valeur de $2 \cdot 10^{-2} \text{ m/s}^2$ et s'arrêtent après 30 s d'enregistrement d'un signal ne dépassant pas 10^{-3} m/s^2 . La résolution des stations est de 10^{-5} m/s^2 et celles-ci saturent si l'accélération du sol dépasse 20 m/s^2 . La réponse des instruments est plate de 0 à 30 Hz et au-delà, approximativement, elle est celle d'un filtre de Butterworth, à 3 pôles et de fréquence de coupure égale à 30 Hz. L'horloge des mesures est contrôlée par un signal GPS et est précise à 5 ms près.

Sous chacune des stations, une caractérisation géotechnique est effectuée jusqu'à 20 m de profondeur, fournissant ainsi les valeurs SPT (test de pénétration standard), les valeurs de la densité, les profils de vitesse des ondes de compression et de cisaillement et les logs sédimentaires.

Réseau KiK-Net : Kiban Kyoshin Network (<http://www.kyoshin.bosai.go.jp/>)

Des sous-ensembles de données du réseau KiK-Net sont utilisés le long de ce projet de thèse. KiK-Net représente l'avantage d'avoir un capteur en surface et un autre en profondeur. Ce dispositif nous a permis d'étudier l'influence des différents proxies des conditions de site par le biais du facteur d'amplification (surface/profondeur).

La répartition géographique des stations des deux réseaux K-Net et KiK-Net est donnée par la figure.I.8.

Les 669 couples de stations de Kik-Net (surface/profondeur) sont installés sur tout le territoire Japonais, y compris dans les grandes villes comme Osaka ou Sendai. Le réseau à un maillage de 20 km, avec des profondeurs comprises entre 100 et 2300 mètres. Les accélérographes de (KiK-Net) sont à trois composantes avec un facteur l'amortissement de 0,6 à 0,7. Le taux d'échantillonnage est de 200 Hz. L'enregistrement obtenu est de type 24 bits. La plupart des stations KiK-Net sont installées sur du rocher ou des sites sédimentaires. Il est à noter que les instruments de mesure accélérométriques KiK-net ont les mêmes caractéristiques que les instruments K-Net.

La figure.I.9 présente la distribution des stations de ce dense réseau en surface et en profondeur. Tandis que le tableau.I.2 ([Aoi et al. 2004](#)) donne la répartition du nombre de stations en fonction de leurs profondeurs.

L'ordre du déclenchement de l'enregistrement (jumelé surface/profondeur) est donné par la station localisée au fond du forage si un seuil de $2 \cdot 10^{-2} \text{ m/s}^2$ y est dépassé. Sous chacune des stations, une caractérisation géotechnique a été effectuée sur toute la profondeur du forage,

fournissant ainsi l'épaisseur, la nature géotechnique et l'âge géologique de chaque couche. Les profils de vitesse des ondes de compression et de cisaillement sont également disponibles.

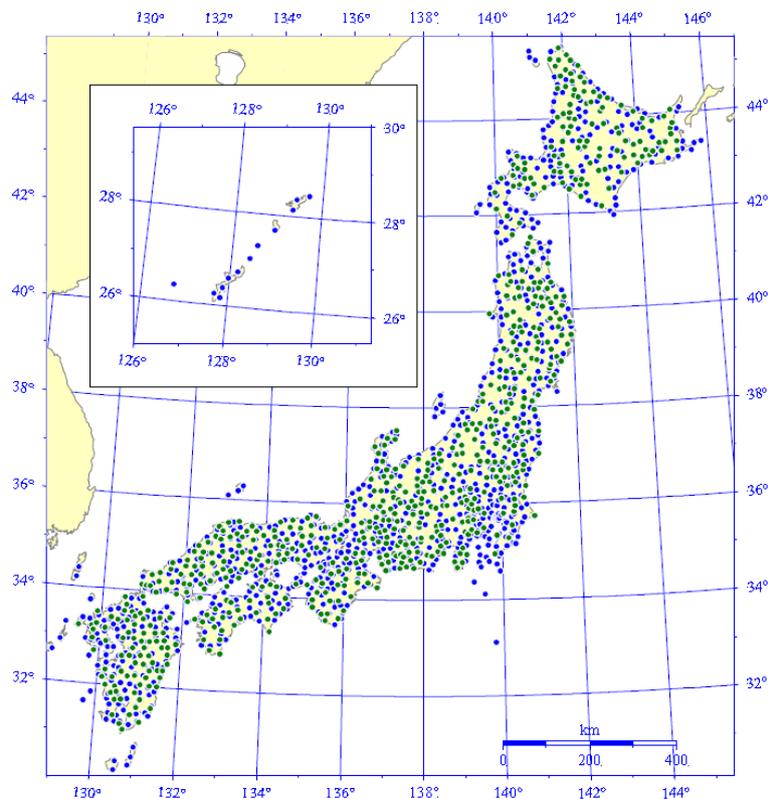


Figure.I.8 Distribution des stations de K-Net en bleu et KiK-Net en vert (*Kashima.,2009*).

L'originalité, la disponibilité gratuite et la qualité des données KiK-Net ont incité plusieurs chercheurs à utiliser des sous-ensembles de données issues de ce réseau pour, entre autres, élaborer des GMPEs. À cet égard nous pouvons citer les travaux de [Zhao et al. 2006](#), [Kanno et al. 2006](#) et de [Cotton et al. 2008](#). Récemment, [Dawood et al. 2016](#) proposent un set de données traité à l'aide d'un protocole automatisé à vocation pratique. Cet ensemble de données est téléchargeable gratuitement via le site <https://datacenterhub.org/resources/272>. Dans leurs travaux, les auteurs présentent un protocole à plusieurs étapes automatisé qui traite systématiquement environ 157000 enregistrements de mouvement fort KiK-net. Le protocole comprend la sélection de la fréquence de coupure du filtrage passe-haut et un ensemble complet de métadonnées compilé pour chaque enregistrement.

Les tremblements de terre sont également classés en événements de subduction et crustaux. Un « fichier plat » qui comporte toutes les métadonnées et les accélérations spectrales en surface et en fond de forage sont disponibles sur le site de NEEShub. Dans la 1^{ère} partie du chapitre IV, la 1^{ère} et la 2^{ème} partie du chapitre V, nous allons utiliser ce jeu de données. Par ailleurs, dans la

2^{ème} partie du chapitre IV, nous allons travailler avec le jeu de données élaboré par Régnier et al. 2016.

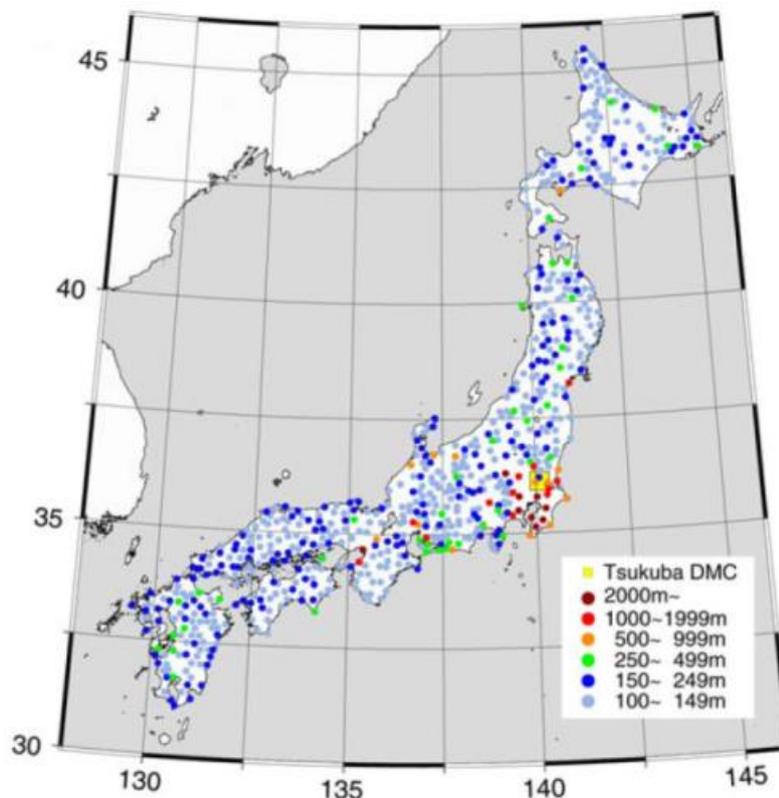


Figure.I.9 Distribution des stations KIK-Net. Les couleurs représentent la profondeur (Aoi et al. 2004).

Table.I.2 Distribution de la profondeur des stations installées en fond du forage du KIK-Net (Aoi et al. 2004).

Profondeur [m]	Nombre de stations
100-149	415
150-249	179
250-499	41
500-999	13
1000-1999	12
total	669

Il est à rappeler aussi que, les données de KiK-Net sont utilisées avec d'autres jeux de données (Europe, Chine, etc.) dans des projets et programmes internationaux. Tel que le programme NGA Next Generation Attenuation. Dans la section suivante, nous présentons la base de données NGA dans ses deux versions 1 et 2.

I.2.3. Base de données : Next Generation Attenuation West

NGA présente trois projets séparés suivant le type de l'événement sismique :

- Pour les tremblements de terre de subduction (par exemple nord-ouest du Pacifique et le nord de la Californie). Le projet est nommé NGA-Sub.
- Région continentale stable (centre et l'est des États-Unis). Le projet est appelé NGA-East.
- Pour les tremblements de terre actifs (crustaux) de faibles profondeurs (comme la Californie, le Japon et Taiwan, etc.) : NGA-West

Dans ce qui suit, nous présentons seulement la base de données NGA-West dans ces deux versions 1 et 2, car, dans cette thèse, nous nous intéressons seulement aux séismes crustaux des zones actives.

En 2003, le « Pacific Earthquake Engineering Research Center » (PEER) a lancé un grand programme de recherche pour le développement des prochaines générations de GMPEs en s'intéressant aux séismes crustaux peu profonds dans les régions tectoniques actives. Le projet appelé NGA-West 1 a été clôturé en 2008. Ce projet a fourni plusieurs résultats importants, y compris une base de données de mouvement fort et un ensemble de GMPEs.

De nombreux chercheurs, praticiens et organisations à travers le monde utilisent maintenant les modèles et la base de données NGA-West. Le programme NGA-West dans sa version primitive était original. Néanmoins, certaines insuffisances ont été identifiées au fil du temps avec les GMPEs qui en résultent :

- Manque d'événements sur le même site (Figure.I.10) : avec plus de données enregistrées sur le même site, il est possible d'analyser de multiples événements pour caractériser la variabilité de site (single station σ).
- Peu de données pour des événements faibles (Figure.I.11). De telles données sont importantes pour donner la possibilité aux régions qui disposent principalement de données de magnitude faible de comparer leurs résultats avec ceux issus de NGA.
- Surestimation des prédictions pour les magnitudes faibles $M < 5$ (Figure.I.12).

Ces insuffisances (non exclusives) ont motivé des travaux pour acquérir des données supplémentaires et incorporer des résultats de simulation via le programme NGA-West 2.

Ici, nous résumons les principaux aspects de la base de données NGA-West 2 (enregistrements et métadonnées), qui sont présentés dans le fichier plat NGA-West 2. Le fichier plat contient les informations utilisées par les développeurs des GMPEs et est fourni sur le site Web <http://peer.berkeley.edu/ngawest2/databases/>.

Le fichier plat contient quatre sources de données : (1) le tableau décrivant les sources sismiques (2) la base de données des sites (3) le tableau de propagation d'onde et (4) le catalogue d'enregistrement (PSA pour différents coefficients d'amortissement).

La base de données comprend 21336 enregistrements à trois composantes issus de 599 événements. La magnitude (M) de ces événements est entre 3.0 et 7.9, la distance la plus proche est égale à 0,05 km et la plus grande égale à 1533 km et $V_{S30} = [94 - 2100]$ m/s (bien que les données deviennent rares pour des distances > 400 km et $V_{S30} > 1200$ m/s ou < 150 m/s). La base de données comprend uniformément les ordonnées pseudo-spectre d'accélération pour des périodes structurelles T allant de 0,01 s à 20 s soit 111 périodes et 11 coefficients d'amortissement.

La figure.I.10 montre la répartition des épicentres des 599 événements. La figure.I.11 représente la distribution magnitude-distance des enregistrements sélectionnés. Nous remarquons, à partir de cette figure que, le nombre d'enregistrements dans NGA-West 2 pour des séismes de magnitude $M < 5$ est important. Ce complément d'information a amélioré la prédiction du mouvement sismique pour les séismes de magnitudes faibles, tout en limitant la sous-estimation (correction de l'effet d'échelle lié à la magnitude) qui a été présente dans les GMPEs utilisant la base NGA-West 1. Nous pouvons montrer ce point via la figure.I.12 ([Campbell and Bozorgnia 2014](#)).

En outre, les nombres de stations dans la version 2008 et la version actuelle sont illustrés dans la figure.I.13. Le nombre des enregistrements issus de la base de données actuelle est 6 fois plus grand que dans la base précédente, notamment pour le Japon, la Chine, et la Californie.

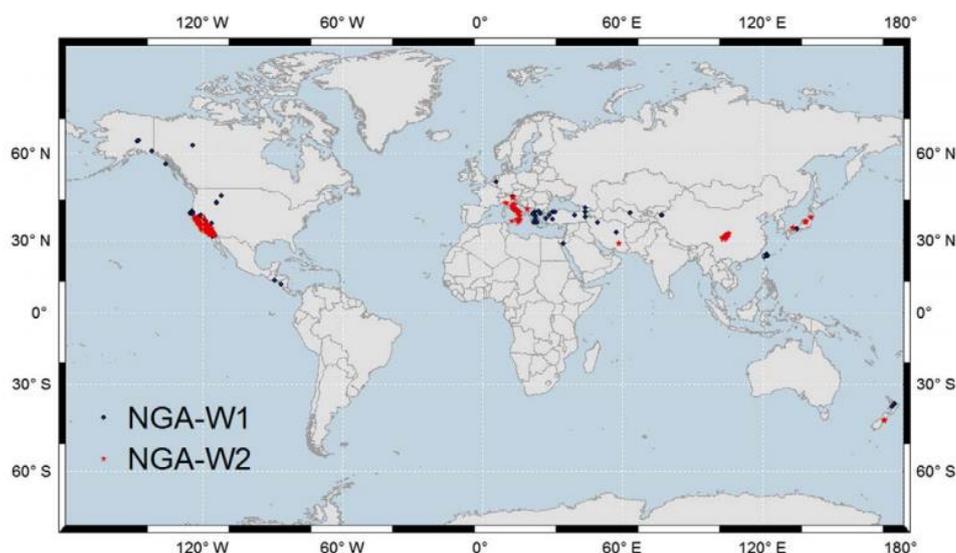


Figure.I.10 Distribution des épicentres : base de données NGA-West 1 et 2 ([Ancheta et al. 2014](#)).

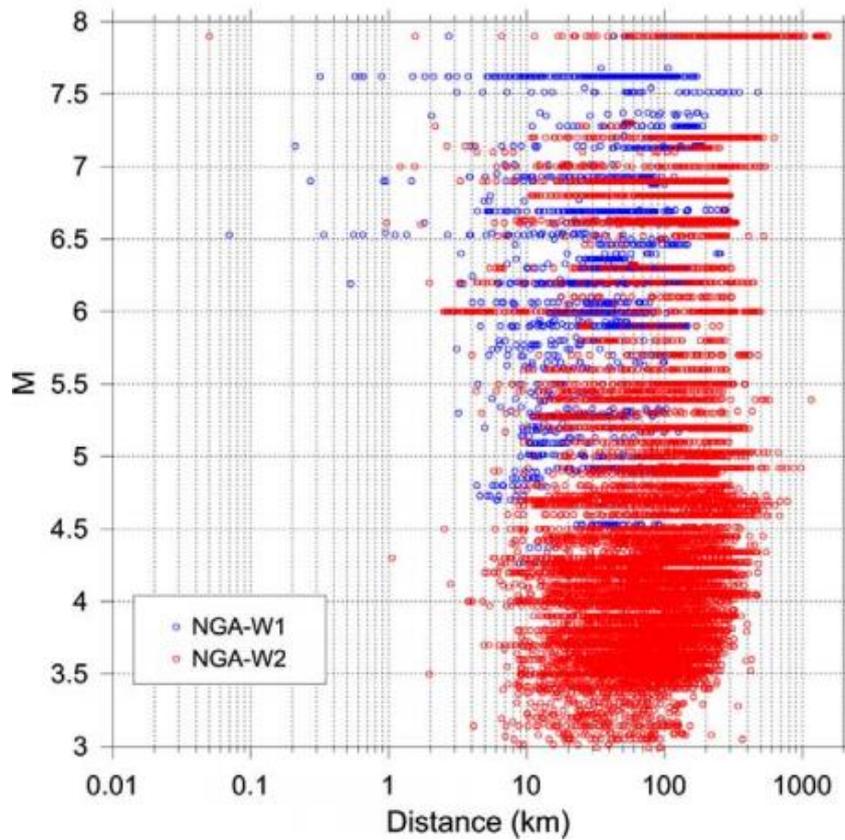


Figure.I.11 Distribution magnitude-distance des enregistrements : NGA-West 1 et 2 ([Ancheta et al. 2014](#)).

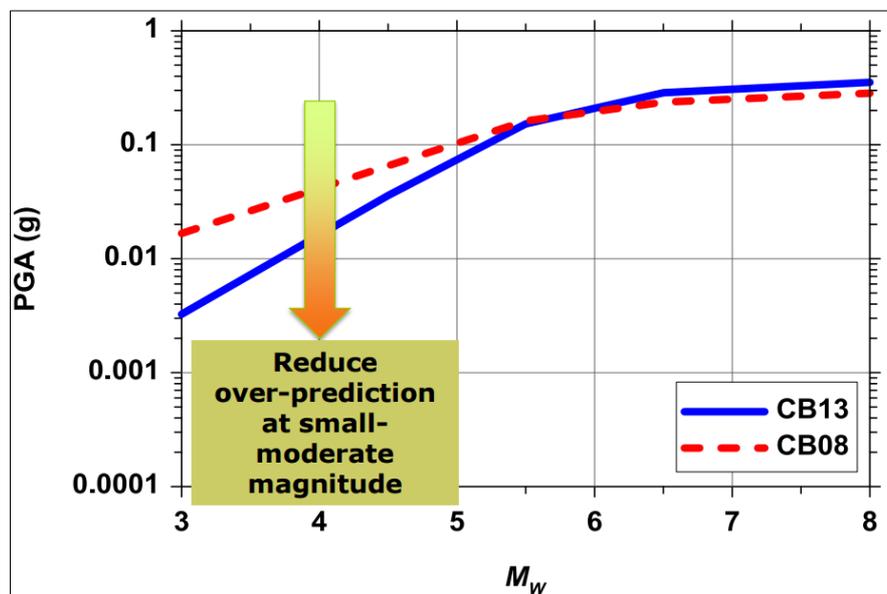


Figure.I.12 Effet d'échelle lié à la magnitude : modèle de Campbell and Bozorgnia de 2008 et de 2013. La distance de rupture est $R_{rup}=10$ km, $V_{s30}=760$ m/s pour un mécanisme en décrochement (http://peer.berkeley.edu/ngawest2/wp-content/uploads/2014/04/S1-Bozorgnia_Speaker_presentation_USSD_2014.pdf).

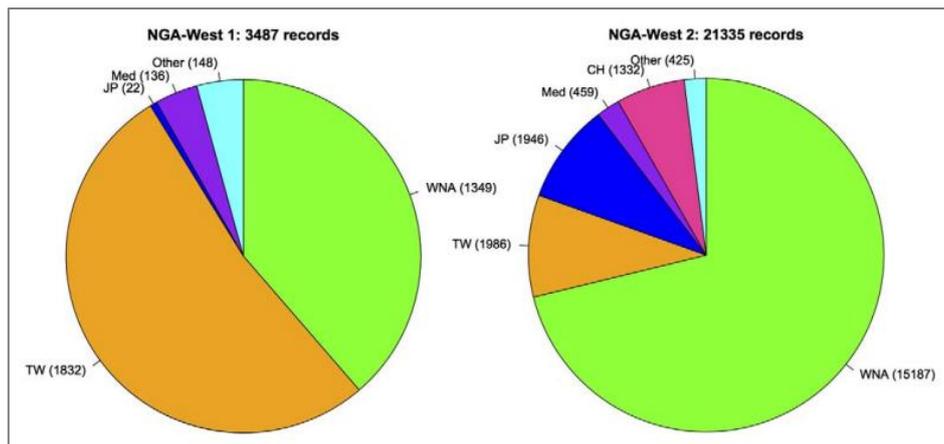


Figure.I.13 Nombre d'enregistrements dans la base de données NGA-West 1 et NGA-West 2 pour différentes régions. CH: Chine, JP: Japon, Med: Méditerranée, TW: Taiwan, WNA: Nord-Ouest de l'Amérique (principalement la Californie). (Ancheta et al. 2014).

Ce qui intéressant dans cette base de données est la disponibilité des informations tabulaires relatives aux conditions de sites telles que, V_{S30} , nature de V_{S30} (mesurée ou estimée), la valeur de la pente topographique Slope (m/m) et bien d'autres proxys de site. La figure.I.14 montre une importante augmentation des sites dans NGA-West 2 par rapport à NGA-West 1. L'augmentation est notable dans le cas où V_{S30} est mesurée. Nous remarquons aussi que la valeur moyenne de V_{S30} est de 300 m/s pour NGA-West 1 et de 400 m/s pour NGA-West 2. Il est à noter aussi que, le nombre de sites où V_{S30} est estimée reste non négligeable et mérite d'être pris en compte lors de l'élaboration des GMPEs. Nous appréhendons cette thématique dans le chapitre III.

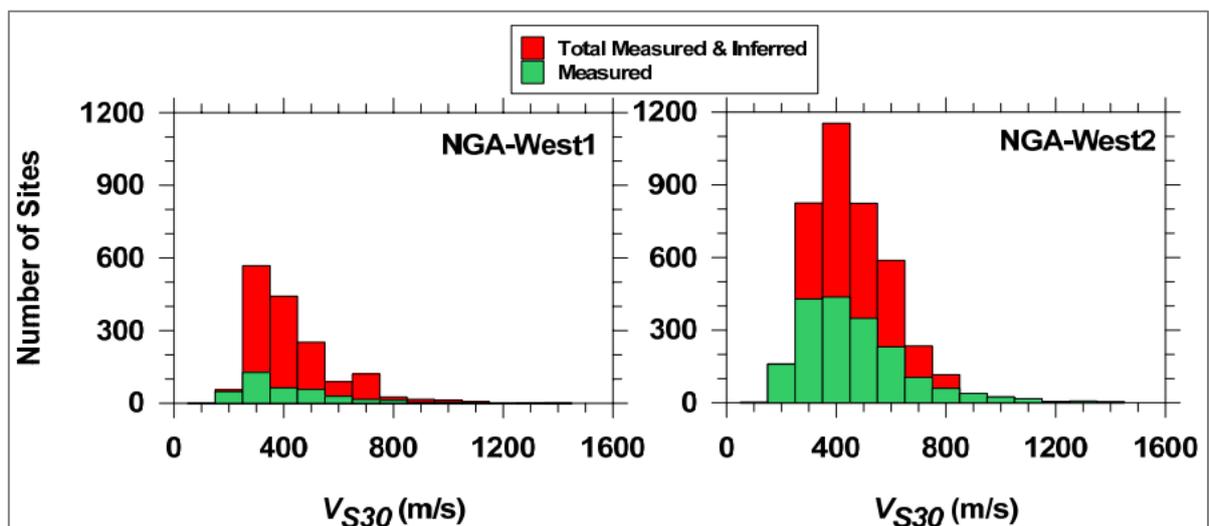


Figure.I.14 Nombre de sites en fonction de V_{S30} (mesurées et estimées) pour les deux bases de données NGA –West 1 et NGA-West 2 (http://peer.berkeley.edu/ngawest2/wp-content/uploads/2014/04/S1-Bozorgnia_Speaker_presentation_USSD_2014.pdf).

Jusqu'à présent nous avons traité l'évolution de chaque base de données séparément. Il est plus intéressant de combiner différentes bases de données. Pour ce faire, nous utilisons les résultats obtenus via le projet SHARE (Seismic Hazard Harmonisation in Europe Project).

I.2.4. Projet SHARE

Dans SHARE sont regroupées plusieurs bases de données de mouvements forts. La base de données se compose d'accélérogrammes recueillis à partir de :

- La base de données nationale Turque du mouvement fort (T-NSMP),
- La base de données NGA,
- La base de données KIK-Net,
- La base de données mondiale globale compilée par Cauzzi and Faccioli (C&F),
- Le site Internet pour les données européennes du mouvement fort (ISESD),
- La base de données accélérométrique italienne (ITACA).

La base de données couvre des tremblements de terre entre 1933 à 2009. La base de données contient un total de 2448 événements (Figure.I.15) et 14193 enregistrements ([Giardini et al. 2013](#)).

La figure.I.16 montre la répartition annuelle des tremblements de terre et les enregistrements de mouvements de terrain donnés par la base de données unifiée (SHARE). Environ 50% des événements ont eu lieu dans la dernière décennie. Là encore, plus de 50% des enregistrements de mouvements de sol appartiennent à cette période. Une concentration plus élevée des événements et des enregistrements au cours des dix dernières années peut être attribuée à l'augmentation du nombre d'instruments de mesure autour du monde. Une autre constatation importante est que, les enregistrements numériques dans la base de données sont prédominants. Un nombre important de capteurs numériques a ainsi été déployé dans les régions sismiques pendant ces dernières décennies.

Ce déploiement instrumental important se concentre cependant dans peu de régions (ex. Japon, Californie, Turquie et Italie) ce qui rend la distribution non homogène. En outre, si nous nous focalisons sur les trois bases de données (RESORCE, Kik-Net et NGA-West 2) illustrées dans la figure.I.17 ; nous constatons que les trois bases couvrent pratiquement toutes la gamme de distances. En revanche, peu de données sont enregistrées pour des séismes de $M > 7$ et pour des distances épacentrales inférieures à 10 km.

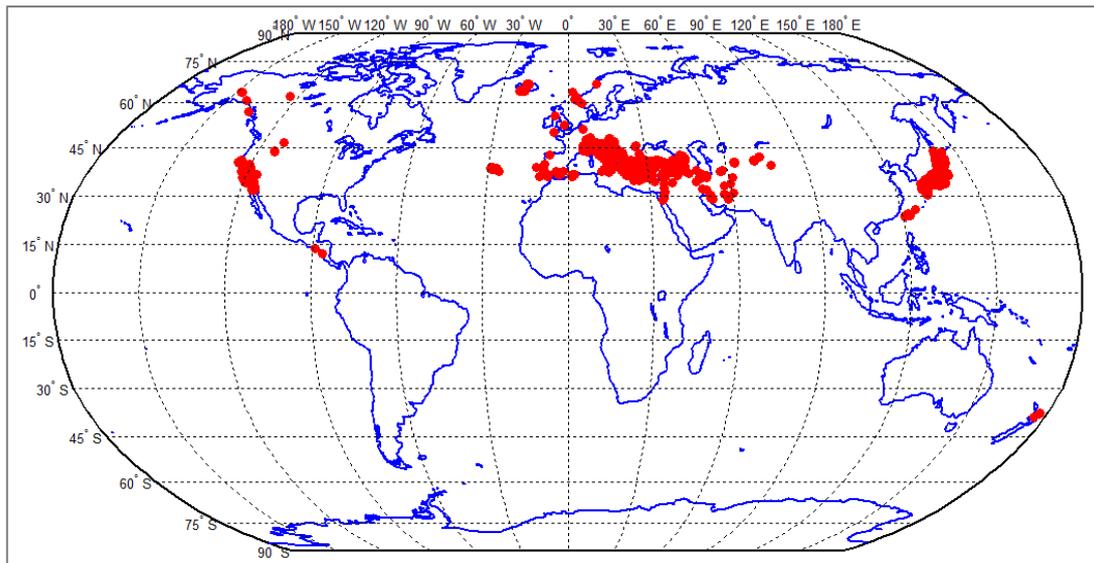


Figure.I.15 Localisation de l'épicentre : la base de données SHARE (Giardini et al. 2013).

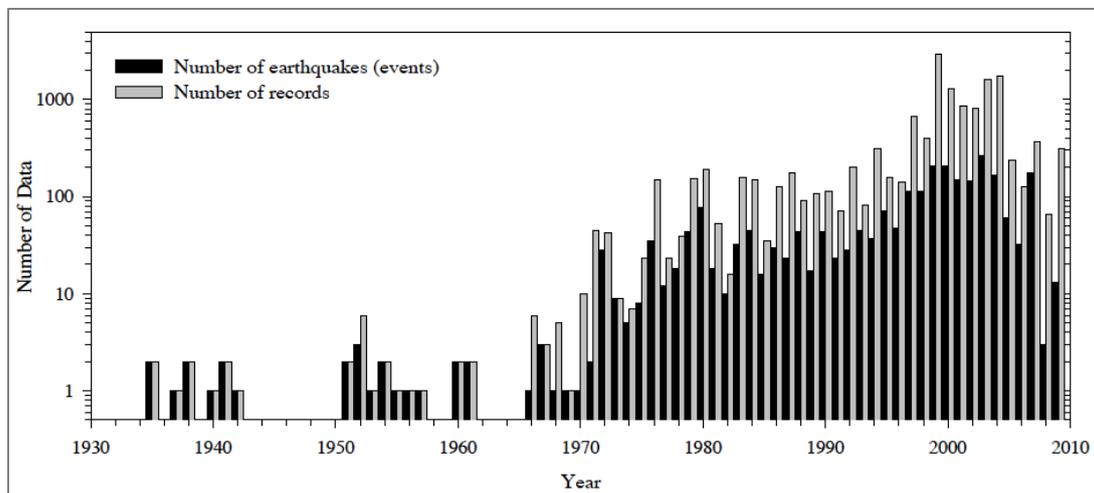


Figure.I.16 Présentation de l'évolution annuelle des données de tremblements de terre et des enregistrements de mouvements sismiques (Giardini et al. 2013).

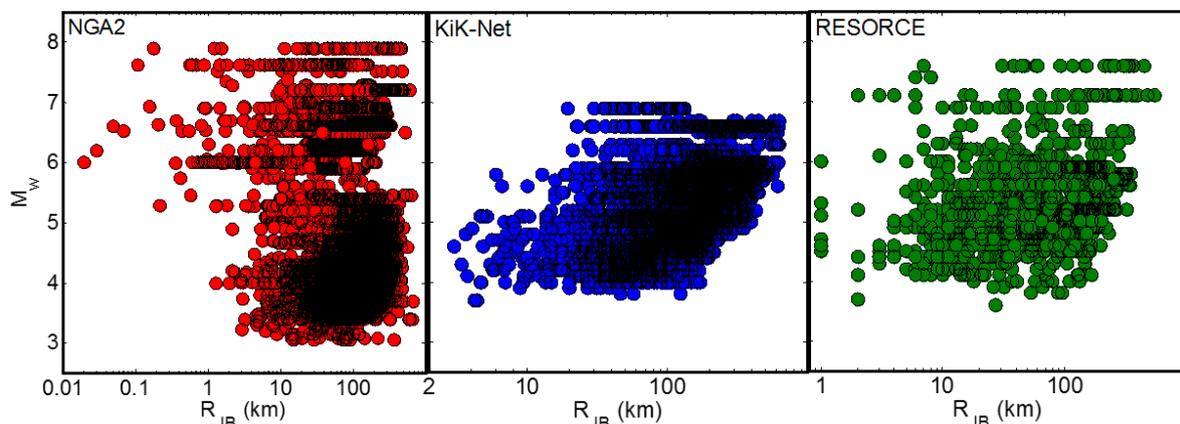


Figure.I.17 Distribution magnitude-distance pour les trois bases de données RESORCE (Akkar et al. 2014), KiK-Net (Dawood et al. 2016) et NGA-West 1 et 2 (Ancheta et al. 2014). Nous représentons ici que les données avec V_{s30} mesurées

Après avoir vu l'évolution en quantité et en qualité au fil du temps de plusieurs bases de données, il reste à présent de faire le point sur les répercussions de ce fantastique essor sur l'amélioration de la prédiction du mouvement sismique.

I.3. GMPEs classiques

Dans ce contexte, [Douglas \(2016\)](#) a recensé plus de 450 équations de prédiction du mouvement sismique établies entre 1964 et 2016 (Figure.I.18 et I.19). La figure.I.18 représente la répartition géographique des GMPEs. Il en résulte une concentration autour des régions actives telles que les USA, Japon et l'Europe. La Figure.I.19 donne l'évolution annuelle des GMPEs. L'augmentation exponentielle en nombre des GMPEs dans cette dernière décennie et l'émergence d'autres approches (en plus de la régression empirique et la simulation) présente l'essentielle remarque de cette Figure.

Par ailleurs, nous pouvons voir le lien existant entre l'évolution drastique des enregistrements accélérométriques et celles de nombres de GMPEs en utilisant les données RESORCE ([Douglas, 2016](#)). La figure.I.20 montre que, le nombre des GMPEs suit la même variation ascendante que le nombre de données (toutes magnitudes et distances confondues). Ce qui nous laisse à dire que, du point de vue quantitatif, l'augmentation du nombre de données a favorisé le développement de plus en plus de GMPEs, surtout en cette dernière décennie.

Douglas, J. (2016), Ground motion prediction equations 1964–2016, Website <http://www.gmpe.org.uk>.



Figure.I.18 Distribution régionale des GMPEs ([Douglas, 2016](#))

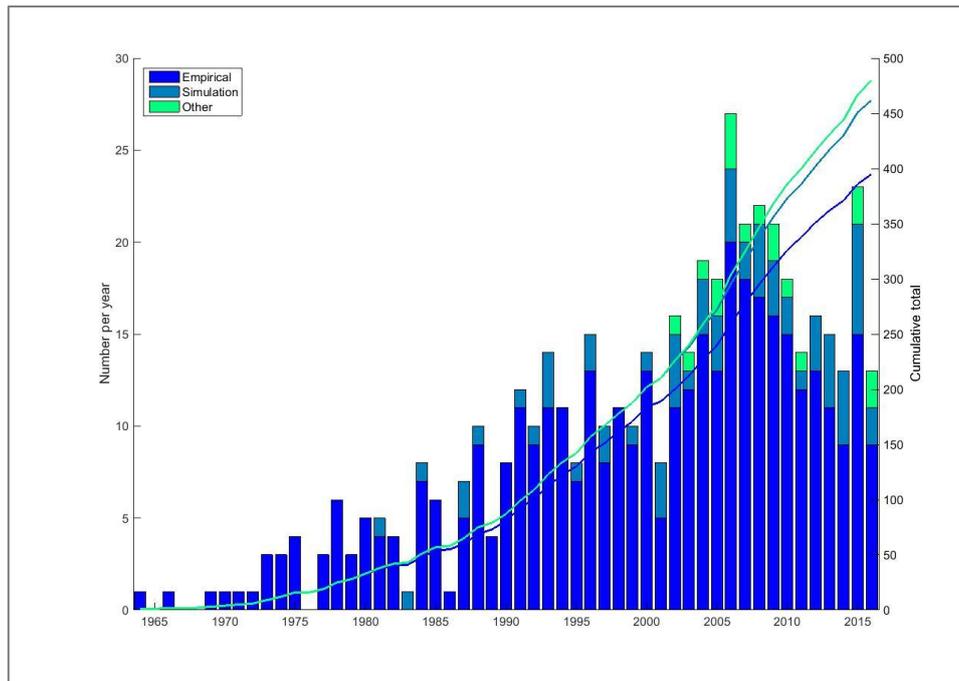


Figure.I.19 Évolution du nombre de GMPEs entre 1964 et 2016 (Douglas, 2016)

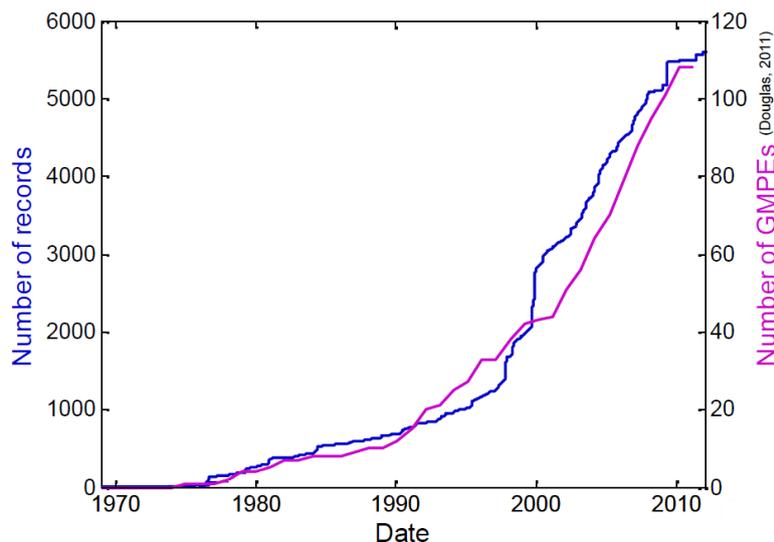


Figure.I.20 Evolution des GMPEs avec les enregistrements sismiques (Douglas, 2014)

Il reste à présent à savoir si la qualité des GMPEs suit cette quantité. C'est ce que nous allons voir ensemble par la suite. Mais avant, nous donnerons des définitions et un aperçu historique sur l'évolution des GMPEs.

I.4. Équation de prédiction du mouvement sismique (GMPE)

Les paramètres décrivant le mouvement du sol peuvent être prédits de manière empirique en fonction des métadonnées d'une base de données (ex. RESORCE, KiK-Net, NGA-West). Ce type de prédiction est assuré classiquement par des équations de prédiction du mouvement sismique (GMPE). Les GMPEs renseignent, donc, sur la valeur attendue d'un paramètre choisi (ex. PGA, PGV, PSA...) en fonction, entre autres, de la magnitude du séisme

(M), de la distance à la source sismique (R) et classe de site ou un proxy de condition de site (ex. V_{s30}), voir aussi d'autres paramètres (directivité, mécanisme au foyer, etc.).

Les modèles empiriques de prédiction du mouvement du sol permettent de calculer une ou plusieurs grandeurs caractérisant le mouvement du sol, comme celles décrites auparavant dans ce chapitre. Ces modèles sont établis pour des régions données de forte sismicité (USA, Japon, Italie...) ou de faible sismicité (France, Canada...). Les GMPEs sont utilisées essentiellement pour l'estimation de l'aléa sismique et l'amélioration de la prise en compte du chargement sismique dans les codes parasismiques en vigueur.

Les conditions géologiques, morphologiques et tectoniques ne sont pas les mêmes d'une région à une autre. Par voie de conséquence, nous pouvons distinguer trois catégories de GMPEs :

- Les équations établies à partir d'événements des régions de subduction, où les séismes peuvent d'atteindre de grandes profondeurs.
- Celles obtenues à partir des données continentales stables qui sont caractérisées par des chutes de contraintes élevées.
- Les GMPEs provenant des événements crustaux dans des zones à sismicité active.

Il est à noter que, tout au long de cette thèse, nous nous intéressons seulement à des événements crustaux avec des profondeurs focales ne dépassant pas les 30 km.

Les faiblesses qui conditionnent et freinent l'évolution et l'utilisation des GMPEs sont multiples nous citons par exemple :

- Le non-consensus autour d'une forme fonctionnelle unique engendré par des interprétations différentes des phénomènes physiques sous-jacents.
- L'absence d'un compromis autour des proxies de source-propagation d'onde-site utilisés lors de l'établissement d'une GMPE.
- La complexité de ces formes fonctionnelles (de plus en plus de facteurs explicatifs donc de degrés de liberté sont incorporés dans les formes fonctionnelles).
- L'incapacité de diminuer considérablement les incertitudes malgré le nombre de paramètres importants.
- L'absence de consensus vis-à-vis du type de magnitude et de la distance utilisés, surtout pour les failles étendues.

Nous tentons, dans les chapitres suivants, de répondre partiellement à ces préoccupations en proposant une approche, autre que les régressions empiriques basées totalement sur les données évitant ainsi le passage par une forme fonctionnelle prédéfinie. Dans ce qui suit, nous essayons de répondre à la question suivante : qu'est-ce qu'une GMPE.

I.4.1. qu'est-ce qu'une GMPE ?

L'équation de prédiction du mouvement sismique (GMPE) permet de rendre compte de la combinaison entre le mécanisme de rupture de la source sismique, de la propagation d'ondes sismiques entre la source et le site et de l'effet de site (Figure.I.21).

De manière basique, la source est assimilée (si on utilise le modèle physique de base) à un point source situé à l'hypocentre du séisme. Le modèle suppose que toute l'énergie du séisme est libérée à partir de ce point. La propagation de cette énergie est sous forme d'onde sphérique. Ces ondes sont accompagnées d'une décroissance géométrique de leur amplitude. De ce fait, cette dernière est inversement proportionnelle à la distance entre la source et le site.

Si le milieu de propagation d'ondes est élastique, seule l'atténuation géométrique est prise en compte. On note Y l'amplitude de PGA, PGV, PSA..., la magnitude M représente l'énergie libérée à la source et R la distance entre la source et le point de mesure (Lussou, 2001) :

$$Y \propto 10^{a.M} \cdot \frac{1}{R^b} \quad (\text{I.9})$$

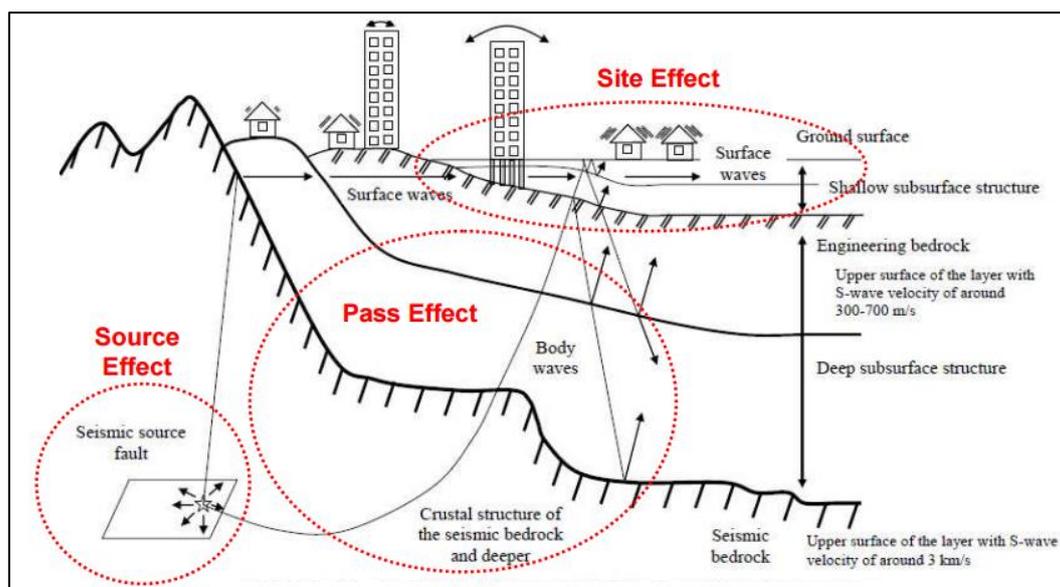


Figure.I.21 Schéma illustratif des effets influençant le mouvement du sol à la surface (Yamada, 2013)

En revanche si le milieu est anélastique, la décroissance de l'amplitude est également fonction du paramètre c qui décrit le caractère dissipatif du milieu de propagation ([Boore et al. 1982](#)) :

$$Y \propto 10^{a.M} \cdot \frac{1}{R^b} \cdot 10^{-c.R} \quad (\text{I.10})$$

On obtient, pour un milieu élastique, la relation suivante :

$$\log_{10}(Y) = a.M - b.\log_{10}(R) + d \quad (\text{I.11})$$

Dans le cas où nous considérons le comportement anélastique du milieu, nous obtenons :

$$\log_{10}(Y) = a.M - b.\log_{10}(R) - c.R + d \quad (\text{I.12})$$

a , b , c et d sont des coefficients de la GMPE. Dans la pratique, l'effet de site est pris en compte par le coefficient d (classe de site) ou par un proxy des conditions de site tel que V_{s30} .

Par exemple une relation prédictive du mouvement du sol, relative à la valeur maximale enregistrée de l'accélération (PGA) peut s'écrire :

$$\log_{10}(PGA_{ij}) = a.M_i - b.\log_{10}(R_{ij}) - c.R_{ij} + d_j \quad j = 1,2,3 \dots, n \quad (\text{I.13})$$

PGA_{ij} est l'accélération maximale pour le $i^{\text{ème}}$ séisme, enregistré à la $j^{\text{ème}}$ station. Le coefficient « a » est relié à la magnitude, « c » dépend de la distance et rend compte de l'atténuation anélastique, $-b.\log_{10}(R)$ Corresponds à l'atténuation géométrique. Dans un milieu homogène, l'atténuation égale à $\frac{1}{R}$ et est égale à $\frac{1}{R^b}$ dans un milieu stratifié d'où le coefficient b . d_j Est un coefficient lié au site sur lequel se trouve la station d'enregistrement. Et n est le nombre de sites étudiés. Les coefficients a , b , c et d sont déterminés souvent par régression. D'autres phénomènes physiques ont une influence sur le mouvement du sol et n'ont pas été pris en considération dans l'équation (I.13).

Pour la prise en compte d'un plus grand nombre de phénomènes physiques, nous présentons l'exemple de GMPE donné par [Pousse \(2005\)](#) dont la forme est la suivante :

$$\log(Y) = A_1 + A_2.M + A_3.(M_{max} - M)^{A_4} + A_5.\log(R + A_6.e^{A_7.M}) + A_8.R + F^{source} + F^{site} \quad (\text{I.14})$$

Y est le paramètre recherché, M est la magnitude et M_{max} est une magnitude de référence, R est la distance mesurée, F^{source} et F^{site} sont des variables qui tiennent compte du type de mécanisme au foyer (faille normale, inverse ou de chevauchement) et de site considéré. Plus de détails de F^{site} sont donnés ci-après. Tandis que, les A_i sont déterminés par régression.

Dans l'équation (I.14) nous remarquons que $\log(Y)$ 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 du sol augmente d'autant moins vite surtout à haute fréquence. Cet effet s'appelle aussi effet d'échelle. C'est ce à quoi fait référence le terme $(M_{max} - M)^{A_4}$. Un autre type de saturation a été perçu, qui exprime le fait que l'amplitude du mouvement en champ proche de la source ne tient compte que partiellement du rayonnement total du phénomène de rupture sismique d'une zone d'étendue finie ([Campbell, 1981](#), [Abrahamson and Siva 1997](#)). Cet effet de saturation à courte distance se traduit dans l'équation par une pente de l'atténuation géométrique qui diminue en champ proche $A_6 \cdot e^{A_7 \cdot M}$. En outre, le modèle de prédiction rend également compte d'une atténuation anélastique $A_8 \cdot R$ qui traduit l'absorption de l'énergie dans le milieu traversé par les ondes sismiques ([Pousse, 2005](#)).

L'effet de site est présent dans les GMPEs par des coefficients F^{site} . Ils définissent la nature géologique, le comportement rhéologique et dynamique du site en question. Généralement la vitesse de propagation des ondes de cisaillement (qui sont les plus destructives : leurs fréquences et modes propres de vibration sont similaires à la plupart des structures courantes) est utilisée comme indicateur de la rigidité du site. En pratique et par convention, cette vitesse est calculée sur les 30 premiers mètres de profondeur (V_{s30}). Cependant, ce proxy de site s'avère insuffisant ([Derras et al. 2012](#)), d'autres paramètres de site méritent donc d'être testés, à savoir la fréquence de résonance du site f_0 , la profondeur à une vitesse des ondes de cisaillement égale à 800 m/s H_{800} , la pente topographique et autres proxies de site. L'étude de ces proxies est l'un des objectifs de ce projet de thèse.

Les GMPEs récentes proposent de remplacer le coefficient F^{site} par une fonction donnant l'amplification sismique d'un site en tenant compte du comportement linéaire et non linéaire du sol tel que :

$$F^{site} = F_L^{site} + F_{LN}^{site} \quad (I.15)$$

Ici, nous donnons un exemple à titre illustratif de l'une des premières fonctions développées. [Choi and Stewart \(2005\)](#) ont établi empiriquement le facteur d'amplification basé sur V_{s30} et PGA_{rock} au rocher en utilisant la formule suivante :

$$F^{site} = \log(Amp) = a \cdot \log\left(\frac{V_{s30}}{V_{ref}}\right) + b \cdot \left(\frac{PGA_{rock}}{0.1}\right) \quad (1.16)$$

V_{ref} représente la vitesse de référence, a et b sont des coefficients de régression.

Dans ce même contexte, nous pouvons trouver d'autres formes fonctionnelles dans la littérature. En outre, [Boore et al. \(2014\)](#) ont ajouté un terme avec lequel ils modélisent l'effet des bassins par région.

Ces paramètres suscités (ex. M , R et V_{s30}) restent des indicateurs qui représentent partialement les différents effets sur le mouvement sismique. Les effets non pris en compte engendrent les effets aléatoires du mouvement sismique. En résumé, le paramètre caractérisant le mouvement du sol « Y » peut être écrit de la manière suivante :

$$\log(Y) = f(X_{es}, \theta) + \Delta \quad (I.17)$$

$f(X_{es}, \theta)$ est la médiane qui représente la partie expliquée de l'équation de prédiction du mouvement sismique. X_{es} représente le vecteur des variables indépendantes (magnitude, distance, style de la faille, conditions de site, etc.). θ Est le vecteur qui symbolise les coefficients de régression et Δ représente la partie inexpliquée de l'équation. Il s'agit de la variabilité aléatoire qui décrit l'ensemble des incertitudes. Δ tient compte aussi de la dispersion des données autour de la valeur médiane prédite $\log(Y)$.

Δ est usuellement décomposée en une variabilité inter-événement between-event (inter-event) ΔB et une variabilité intra-événement « within-event » ou « intra-event ») ΔW . τ et ϕ représentent les écarts types de ΔB et ΔW respectivement. B et ΔW ont des moyennes nulles, sont indépendants et normalement distribués. Sur la figure I.22 nous présentons les résidus inter-événements δB_e qui sont définis par l'équation I.18 et les résidus intra-événements δW_{es} à partir de l'équation I.20.

$$\delta B_e = \frac{1}{n_e} \sum_{e=1}^{n_e} \Delta \quad (I.18)$$

$$\Delta = (\log(Y) - f(X_{es}, \theta)) \quad (I.19)$$

$$\delta W_{es} = \Delta - \delta B_e \quad (I.20)$$

n_e représente le nombre des stations déclenchées pendant l'événement sismique e .

De ce fait, l'équation I.17 peut être écrite comme suit :

$$\log(Y) = f(X_{es}, \theta) + \delta B_e + \delta W_{es} \quad (I.21)$$

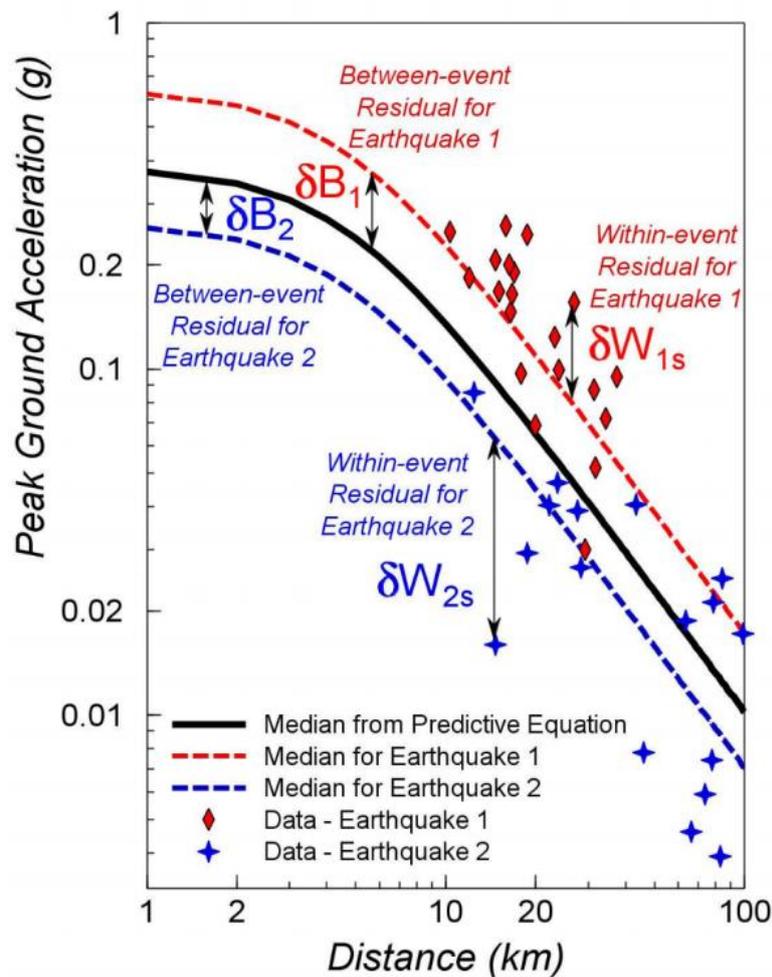


Figure.I.22 Composantes inter-événements (between-events) et intra-événements (within-events) de la variabilité aléatoire du mouvement sismique (Strasser et al. 2009)

Le calcul des deux composantes δB_e et δW_s pour les modèles de prédiction du mouvement sismique est important. Les résidus inter-événements représentent la variabilité liée à la source, et reflètent l'influence de plusieurs facteurs tels que la chute des contraintes, la variation du glissement de la faille en espace et en temps qui ne sont pas pris en compte par l'inclusion de la magnitude, le type de la faille et la profondeur focale. Les résidus intra-événements représentent, quant à eux, les variations qui sont dues, à la fois, à la source, à la propagation d'onde et à l'effet de site. Ces derniers effets représentent l'influence de certains facteurs comme l'hétérogénéité crustale, la structure géologique profonde, les couches de sol proches à la surface. Ces facteurs ne sont pas pris en compte par la distance et la classification du site basées, entre autres, sur la vitesse moyenne des ondes de cisaillements.

Il est à noter que, l'équation I.18 reste valable seulement si nous avons un nombre suffisant de sites par événement. Dans le cas échéant, l'utilisation du modèle mixte s'impose. Il s'agit d'intégrer en plus du modèle fixe, le modèle à effet aléatoire (Abrahamson and Youngs 1992).

Les écarts type des résidus inter-événements et intra-événements du modèle de prédiction sismique représentent la variabilité séisme-séisme et la variabilité enregistrement-enregistrement respectivement. L'écart type total décrivant la variabilité des mouvements sismiques peut-être écrite par :

$$\sigma = \sqrt{\tau^2 + \phi^2} \quad (\text{I.22})$$

[Boore et al. \(2014\)](#) est d'autres études récentes proposent l'écart type total suivant :

$$\sigma(M, R_{JB}, V_{s30}) = \sqrt{\tau^2(M) + \phi^2(M, R_{JB}, V_{s30})} \quad (\text{I.23})$$

C'est une forme plus complexe de l'écart type. Cette complexité suit l'expansion des données (ex. NGA-West 2) : un large éventail de régions, de magnitudes et de distances. L'utilisation d'une forme simple (l'équation I.22) avec une grande quantité de données, riche notamment en faibles magnitudes, engendre des valeurs d'écart type plus grandes. Et cela peut avoir un effet important sur les résultats des calculs d'aléa à longues périodes de retour ([Bommer and abrahamson 2006](#)). Dans le présent manuscrit, est précisé dans le chapitre III nous essayerons d'étudier les tendances de τ avec M et ϕ en fonction de M et R et de comparer ces tendances avec celles obtenues par [Boore et al. \(2014\)](#).

Les équations de prédiction du mouvement sismique sont très nombreuses. Elles sont différentes par la forme des coefficients de l'atténuation utilisée (élastique, anélastique, saturation en distance saturation en magnitude : effet d'échelle), par les paramètres qui caractérisent les phénomènes physiques pris en considération : source, mécanique au foyer (normale, inverse ou de décrochement). Les GMPEs diffèrent aussi par le type de la magnitude utilisée (magnitude locale, magnitude d'ondes de surface, et magnitude de moment) et par les définitions de la distance (distance épacentrale, distance hypocentrale, distance la plus courte à la projection en surface de la zone de rupture, distance la plus courte à la zone de rupture). Une autre différence entre les GMPEs vient des paramètres utilisés pour tenir compte de l'effet de l'amplification sismique linéaire et/ou non linéaire lithologique. Elles diffèrent également par les enregistrements sismiques et la méthode de régression ou inversion utilisés pour calculer les coefficients. Un exemple de l'utilisation de régression est donné par [Derras. \(2011\)](#).

L'illustration donnée dans [Derras. \(2011\)](#) représente seulement une des méthodes parmi les plus simples utilisées pour la détermination des coefficients des GMPEs. Les méthodes sont multiples, en plus de [Fukushima et al. \(1990\)](#) nous citons celles de [Boore and Joyner \(1982\)](#).

La plupart des GMPEs récentes utilisent la régression à l'effet mixte ([Abrahamson and Youngs 1992](#)). Cette approche est présentée d'une façon détaillée dans le chapitre II.

Les GMPEs les plus récentes et les plus développées sont celles issues du projet NGA-West 2 ([Ancheta et al. 2014](#)). En profitant de l'expansion phénoménale des données accélérométriques (NGA-West 2 contient 6 fois plus d'enregistrements que NGA-West 1), ces 05 GMPEs estiment le mouvement du sol et la variabilité aléatoire associée à des séismes crustaux. Lesdites GMPEs sont publiées dans un numéro spécial de la revue « earthquake spectra ». Par abréviations sont nommées : [Abrahamson et al.](#) (ASK 2014), [Boore et al.](#) (BSSA 2014), [Campbell and Bozorgnia](#) (CB 2014), [Chiou and Youngs](#) (CY 2014), et [Idriss](#) (IM 2014). Parce qu'elles sont récentes, nous leur consacrons une bonne partie de ce qui suit. Commençons par leurs formes fonctionnelles.

I.4.2. Modèles NGA-West 2 (médianes)

Les 5 GMPEs proposent la prise en compte de l'effet d'échelle liée à la magnitude (effet de la saturation avec la magnitude pour les séismes petits à modérer). Chacun des cinq modèles contient un facteur qui tient compte du style du mécanisme au foyer et est basé sur leur classification du mécanisme au foyer respectivement. Cette dépendance diminue avec la magnitude ; les modèles d'ASK, CB, CY réduisent ce facteur à zéro pour $M \leq 4.0-4.5$. Les modèles BSSA et IM utilisent un style du mécanisme au foyer indépendant de la magnitude. Les trois modèles (ASK, CB et CY) présentent des formes fonctionnelles explicites pour la prise en compte du comportement chevauchant et un terme pour la profondeur de rupture. Le modèle BSSA évoque implicitement la prise en compte du comportement chevauchant grâce à l'utilisation de la distance Joyner-Boore et ne contient pas de terme pour la profondeur de rupture, parce qu'ils ont constaté que cet effet n'est pas significatif pour $M > 5$. Le terme de la profondeur de rupture pour le modèle CB est la profondeur hypocentrale, qui est corrélée avec le facteur caractérisant le style du mécanisme au foyer. Tandis que, le modèle IM n'inclut pas une forme fonctionnelle pour ces caractéristiques.

Les 5 GMPEs utilisent V_{s30} comme proxy des conditions de site. Le modèle ASK adopte, pour le facteur de l'amplification de site non linéaire, les résultats de la simulation numériques de [Kamai et al. \(2014\)](#) en plus des données du NGA-West 2. Tandis que le modèle CB a retenu les résultats des simulations non linéaires de [Walling et al. \(2008\)](#). CY a utilisé des termes de sites non linéaires provenant exclusivement de l'analyse des données. Le modèle IM contient une fonction linéaire de l'amplification de site basée sur une gamme de V_{s30} plus restrictive [450-

2000] m/s. Le facteur d'amplification non linéaire de site est en fonction soit de PGA pour BSSA et CB ou de PSA au roché pour ASK et CY. Les modèles ASK, BSSA et CY intègrent également un facteur qui dépend de la vitesse des ondes de cisaillement à 1.0 km de profondeur (Z1.0) et à 2,5 km de profondeur (Z2.5) pour le modèle CB. Ces facteurs supplémentaires sont nécessaires pour mieux caractériser les effets du bassin qui ne sont pas entièrement intégrés avec V_{S30} et la forme fonctionnelle. Ce type de facteur n'est pas pris en compte dans cette thèse, car les effets du bassin nécessitent, à mon sens, à eux seuls une étude spécifique.

Sur la base de l'augmentation significative des données non californiennes, les modèles d'ASK, BSSA, CB et CY proposent des ajustements régionaux pour la réponse de site. Un ajustement régional est donné également par les modèles ASK, CB et CY pour tenir en compte de l'atténuation anélastique à longue distance entre diverses régions géographiques. Le modèle BSSA incorpore seulement l'ajustement de l'atténuation régionale comme fonction dans son modèle. Il est à noter que, dans cette thèse la distinction régionale existante entre l'Italie et la Turquie est étudiée en deux temps (chapitre II) : étude de premier ordre en se basant sur le jeu de données (RESORCE) et de deuxième ordre en analysant les résidus obtenus à l'issue de la modélisation.

En général, les 5 GMPEs proposent des termes qui captent les trois effets : effets de source, de propagation d'onde et de site. Il reste à savoir si ces 5 modèles donnent le même niveau de mouvement sismique.

Existe-t-il un consensus autour d'une forme fonctionnelle ?

Les 5 GMPEs utilisent plus au moins le même jeu de données avec différentes formes fonctionnelles et paramètres (tableau 2 dans [Gregor et al. 2014](#)). Dans la figure.I.23, nous exposons, un exemple de forme fonctionnelle. Nous choisissons celle de [Chiou and Youngs \(2014\)](#). Cette GMPE est parmi les plus complètes, mais aussi les plus complexes. Les auteurs tiennent compte de la complexité de la source (mécanisme au foyer, Hanging Wall/ Foot wall, directivité), le site est décrit à travers plusieurs paramètres, plusieurs métriques pour les effets « source-proche ». Les 4 autres formes fonctionnelles sont aussi complexes que celle de Chiou et Youngs.

Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra

Brian S.-J. Chiou^{a)} and Robert R. Youngs^{b)} M.EERI

We present an update to our 2008 NGA model for predicting horizontal ground motion amplitudes caused by shallow crustal earthquakes occurring in active tectonic environments. The update is based on analysis of the greatly expanded NGA-West2 ground motion database and numerical simulations. The updated model contains minor adjustments to our 2008 functional form related to style of faulting effects, hanging wall effects, scaling with the depth to top of rupture, scaling with sediment thickness, and the inclusion of additional terms for the effects of fault dip and rupture directivity. In addition, we incorporate regional differences in far-source distance attenuation and site effects between California and other active tectonic regions. Compared to our 2008 NGA model, the predicted medians by the updated model are similar for $M > 7$ and are lower for $M < 5$. The aleatory variability is larger than that obtained in our 2008 model. [DOI: 10.1193/072813EQS219M]

- M = Moment magnitude.
- R_{RUP} = Closest distance (km) to the ruptured plane.
- R_B = Closest distance (km) to the surface projection of rupture plane.
- R_Y = Site coordinate (km) measured perpendicular to the fault strike from the fault line, with the down-dip direction being positive.
- F_{HW} = Hanging-wall flag: 1 for $R_Y \geq 0$ and 0 for $R_Y < 0$.
- δ = Fault dip angle.
- Z_{TOR} = Depth (km) to the top of ruptured plane.
- ΔZ_{TOR} = Z_{TOR} centered on the M-dependent average Z_{TOR} (km).
- F_{RF} = Reverse-faulting flag: 1 for $30^\circ \leq \lambda \leq 150^\circ$ (combined reverse and reverse-oblique), 0 otherwise; λ is the rake angle.
- F_{NM} = Normal faulting flag: 1 for $-120^\circ \leq \lambda \leq -60^\circ$ (excludes normal-oblique), 0 otherwise.
- V_{310} = Travel-time averaged shear-wave velocity (m/s) of the top 30 m of soil.
- $Z_{1.0}$ = Depth (m) to shear-wave velocity of 1.0 km/sec.
- $\Delta Z_{1.0}$ = $Z_{1.0}$ centered on the V_{310} -dependent average $Z_{1.0}$ (m).
- DPP = Direct point parameter for directivity effect.
- ΔDPP = DPP centered on the site- and earthquake-specific average DPP .

$$\ln(y_{\text{ref}}) = c_1 + \left\{ c_{1a} + \frac{c_{1c}}{\cosh(2 \cdot \max(M_i - 4.5, 0))} \right\} F_{HW} + \left\{ c_{1b} + \frac{c_{1d}}{\cosh(2 \cdot \max(M_i - 4.5, 0))} \right\} F_{NM} + \left\{ c_7 + \frac{c_{7b}}{\cosh(2 \cdot \max(M_i - 4.5, 0))} \right\} \Delta Z_{TOR} + \left\{ c_{11} + \frac{c_{11b}}{\cosh(2 \cdot \max(M_i - 4.5, 0))} \right\} (\cos \delta_i)^2 + c_2 (M_i - 6) + \frac{c_2 - c_1}{c_n} \ln(1 + e^{c_n (c_M - M_i)}) + c_4 \ln(R_{RUP}) + c_5 \cosh(c_6 \cdot \max(M_i - c_{RB}, 0)) + (c_{4a} - c_4) \ln(\sqrt{R_{RUP}^2 + c_{2B}^2}) + \left\{ c_{72} + \frac{c_{72b}}{\cosh(\max(M_i - c_{73}, 0))} \right\} R_{RUP} + c_8 \max\left(1 - \frac{\max(R_{RUP} - 40, 0)}{30}, 0\right) \min\left(\frac{\max(M_i - 5.5, 0)}{0.8}, 1\right) e^{-c_{1a} (M_i - c_{1b})^2 \Delta DPP_i} + c_9 F_{HW} \cos \delta_i \left\{ c_{9a} + (1 - c_{9a}) \tanh\left(\frac{R_{RUP}}{c_{9b}}\right) \right\} \left\{ 1 - \frac{\sqrt{R_{RUP}^2 + Z_{TOR}^2}}{R_{RUP} + 1} \right\}$$

$$\ln(y_g) = \ln(y_{\text{ref}}) + \eta_i + \phi_1 \cdot \min\left(\ln\left(\frac{V_{310}}{1130}\right), 0\right) + \phi_2 \left(e^{\phi_3 (\min(V_{310}, 1130) - 360)} - e^{\phi_3 (1130 - 360)} \right) \ln\left(\frac{y_{\text{ref}} e^{\eta_i} + \phi_4}{\phi_5}\right) + \phi_3 \left(1 - e^{-\Delta Z_{1.0} / \phi_5} \right) + \varepsilon_g \quad (12)$$

Figure.I.23 Forme fonctionnelle donnée par Chiou and Youngs (2014).

Pour savoir si avec les 5 GMPEs nous obtenons le même niveau de mouvement sismique, nous comparons leurs ordonnées spectrales. Pour ce faire, nous utilisons l'outil distribué par PEER (<http://ngawest2.berkeley.edu>). La figure.I.24 représente les valeurs des métadonnées choisies pour effectuer ce test et les 5 spectres de réponse en accélération issus des 5 GMPEs. À travers cette comparaison nous pouvons constater l'absence d'un consensus en termes de forme, de valeurs spectrales (PSA) et contenu fréquentiel (1/T). À titre d'exemple, le modèle ASK donne une valeur de $PSA_{\max} = 0.09$ g pour $T = 0.2$ s, tandis que le modèle CB présente un $PSA_{\max} = 0.13$ g pour une période égale à 0.1 s.

On pourrait penser que les GMPEs sont devenues assez performantes et robustes pour prédire les chargements sismiques à venir. Mais est-ce bien le cas ?

Pour répondre à cette question complexe, nous présentons un exemple montrant la différence excitante entre les formes fonctionnelles et les données sismiques en termes de PGA enregistrées lors du mégaséisme de Tohoku (11/03/2011). La figure.I.25 représente différentes allures de GMPEs établies récemment pour les zones de subductions japonaises. Les

indices Zea06, Kea06, AB03, Yea97 et Gea02 représentent les GMPEs élaborées par [Zhao et al. \(2006\)](#), [Kanno et al. \(2006\)](#), [Boore et al. \(2003\)](#), [Youngs et al. \(1997\)](#) et [Gregor et al. \(2002\)](#).

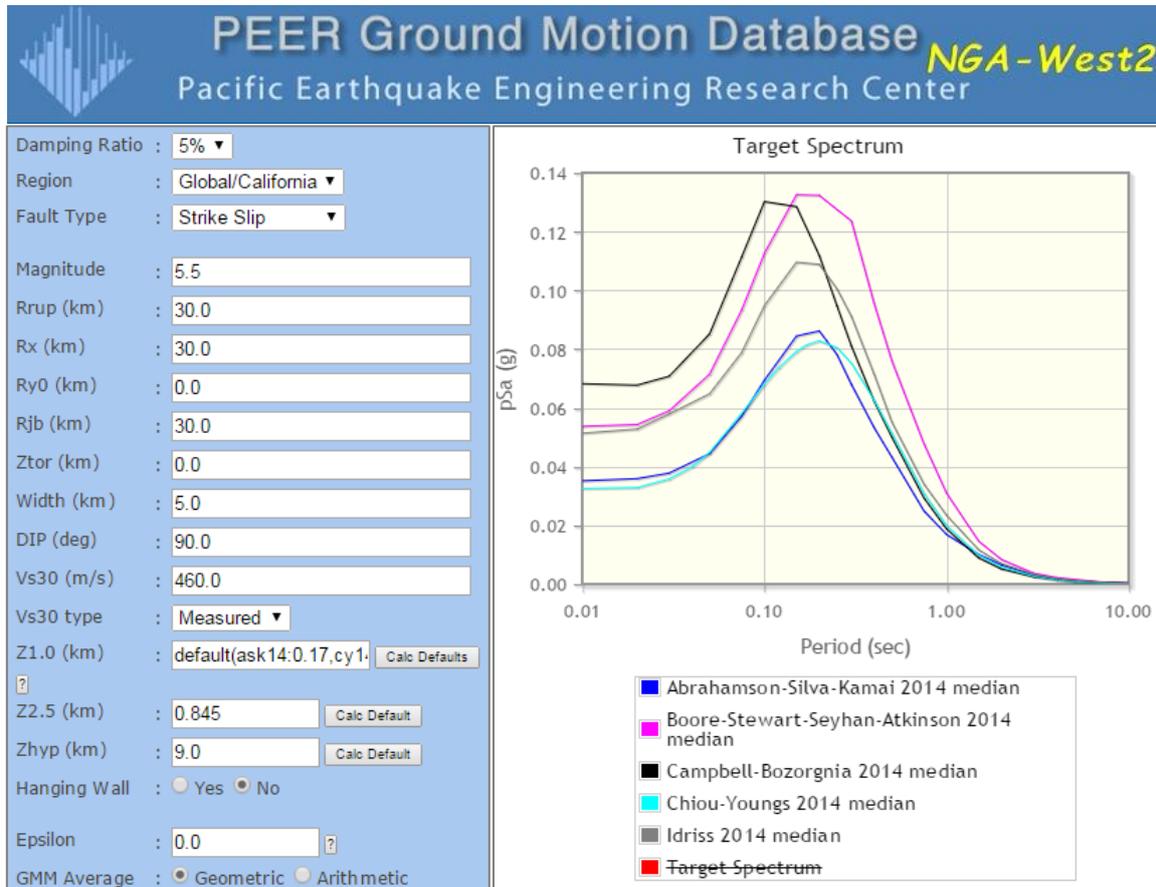


Figure.I.24 Spectres de réponse des 5 GMPEs pour un scénario donné (<http://ngawest2.berkeley.edu>).

En comparant ces modèles d'atténuation (qui, par occasion, n'ont pas le même degré d'atténuation $1/R^x$) d'une part et les enregistrements (PGA) pour différents sites d'autre part, nous pouvons remarquer que l'ensemble des modèles sous-estime le mouvement sismique en champ proche et le sur-estime en champ lointain.

Suite à cette comparaison dédiée à la partie expliquée de l'équation de prédiction, il reste à mesurer la robustesse de ces modèles. En d'autres termes, mesurer la capacité des GMPEs à intégrer le maximum d'effets pour réduire l'incertitude donnée par la partie inexpliquée de l'équation (la variabilité aléatoire).

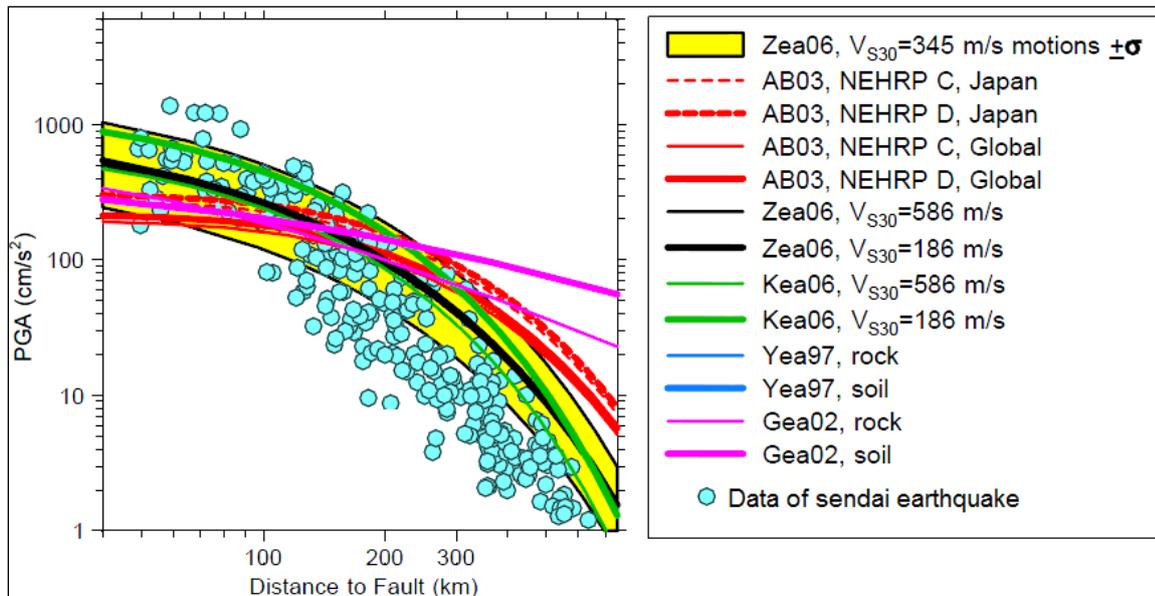


Figure.I.25 Comparaison entre les GMPEs élaborées pour les événements de subduction au Japon et les PGA enregistrés lors du séisme de Tohoku
http://daveboore.com/daves_notes/comparisons_of_ground_motions_from_the_m_9_tohoku_earthquake_with_gmpes_v1.4.pdf

I.4.3. Modèle de la variabilité aléatoire

Les 4 modèles ASK, BSSA, CB et CY supposent que la variabilité aléatoire du mouvement sismique dépend de la magnitude (la variabilité apparente diminue avec la magnitude) et de la distance (la variabilité apparente augmente avec la distance). La même constatation sera décrite au chapitre III de la présente thèse.

En outre, les modèles issus de la base de données NGA-West 2 dépendent tous de la période. La comparaison entre ces 5 modèles est illustrée sur la Figure.I.26. Il est à noter que, les valeurs de l'écart type (σ total) des modèles sont différentes.

Il est intéressant aussi de voir l'impact de la base de données et de la forme fonctionnelle sur la variabilité aléatoire apparente. Dans la figure.I.27, les auteurs comparent les deux modèles CY2014 et CY2008 ([Chiou and Youngs 2008](#)). Dans le dernier modèle, les auteurs utilisent la base de données NGA-west 1 avec moins de paramètres. Ce qui est remarquable dans cette comparaison c'est que $\sigma_{CY2008} < \sigma_{CY2014}$ pour $T < 3$ s et ce, malgré l'existence d'un modèle de variabilité aléatoire complexe.

Cette comparaison montre donc qu'on n'aura pas forcément une réduction de la variabilité aléatoire en augmentant la taille de la base de données. Le fait d'avoir plus d'informations de différents sites/régions et événements sismiques engendre plus de disparités qui ne sont pas

prises en compte par la forme fonctionnelle (même en ajoutant davantage de coefficients de régression explicites caractérisant l'effet de source, de la propagation d'onde et de site).

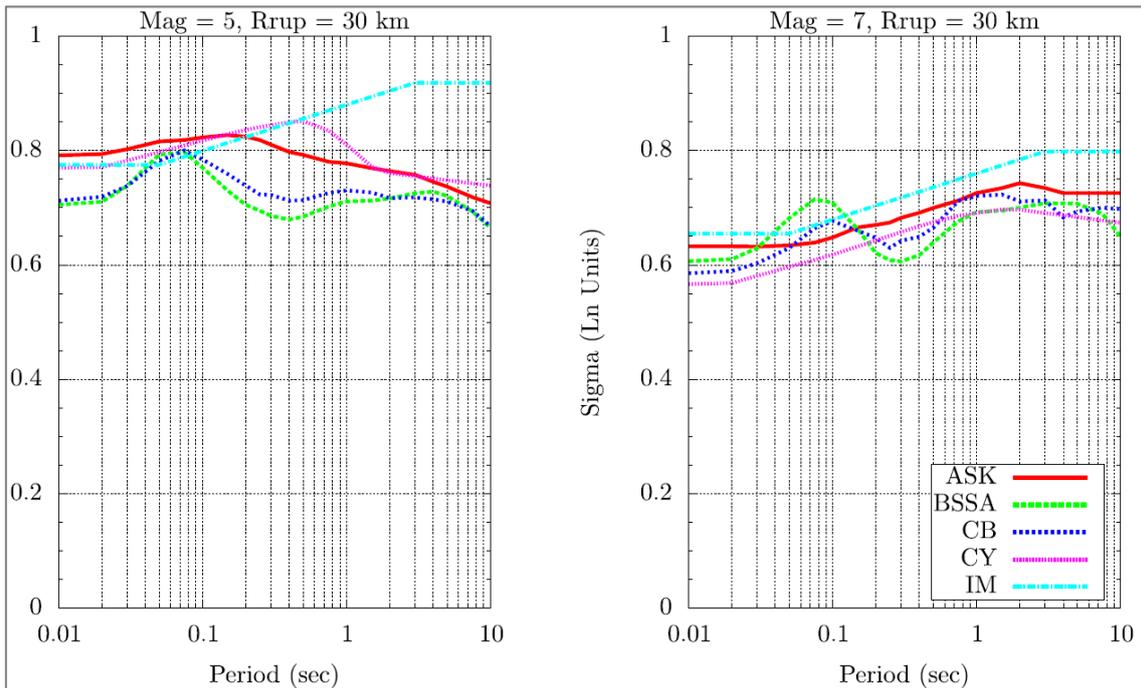


Figure.I.26 Comparaison de l'écart type (σ) pour $M=5$ pour des séismes à faille décrochante à une distance de rupture R_{rup} égale à 30 km et pour $V_{s30} = 760$ m/s (Gregor et al. 2014).

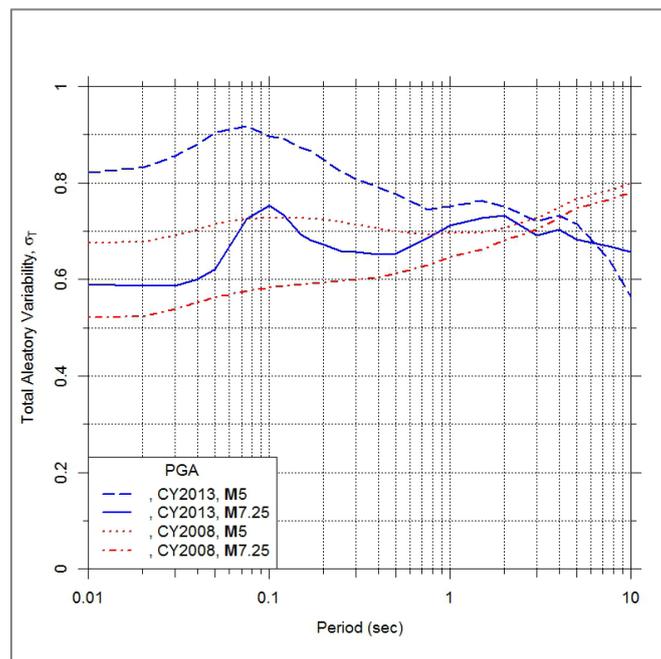


Figure.I.27 Comparaison entre l'écart type (σ) vs périodes (s) de CY2008 et CY2014 (Chiou and Youngs 2013).

L'enjeu principal reste donc de diminuer l'incertitude de ces équations. En effet une légère modification de la variabilité aléatoire a un grand impact sur l'estimation de l'aléa sismique comme c'est montré par Bommer and Abrahamson (2006) : Figure.I.28.

Conscient de l'importance de ce paramètre « σ », [Strasser, \(2009\)](#) a exposé son évolution sur les trente dernières années (Figure.I.29). Ainsi nous remarquons que la réduction n'est pas vraiment significative pendant ces 30 dernières années, et ce malgré des investissements lourds en termes de réseaux accélérométriques, formes fonctionnelles et algorithmes de prédiction. [Bommer et al. \(2010\)](#) a montré ainsi que, le nombre de coefficients utilisés pendant les 40 dernières années pour l'obtention des PGA est en nette croissance. Une croissance qui n'a pas pu réduire significativement l'écart type (figure.I.30). Le défi reste la réduction de l'incertitude dans les GMPEs pour une meilleure évaluation de l'aléa sismique.

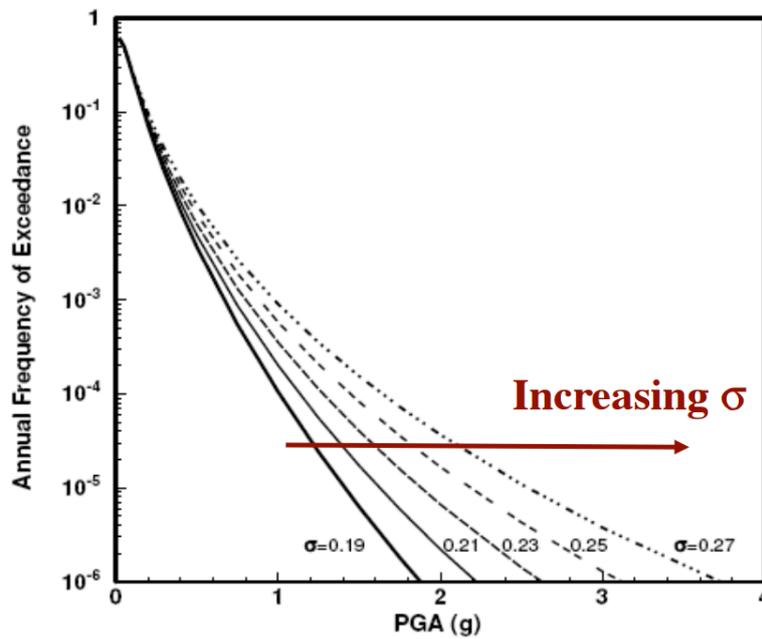


Figure.I.28 Impact de la valeur de σ sur l'aléa sismique probabiliste ([Bommer and Abrahamson, 2006](#)).

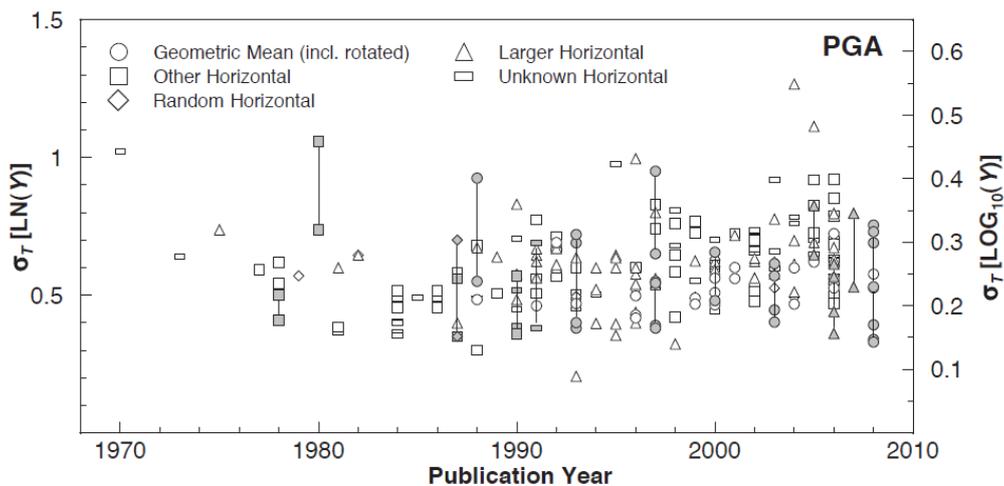


Figure.I.29 Évolution de σ durant les 30 dernières années ([Strasser et al. 2009](#))

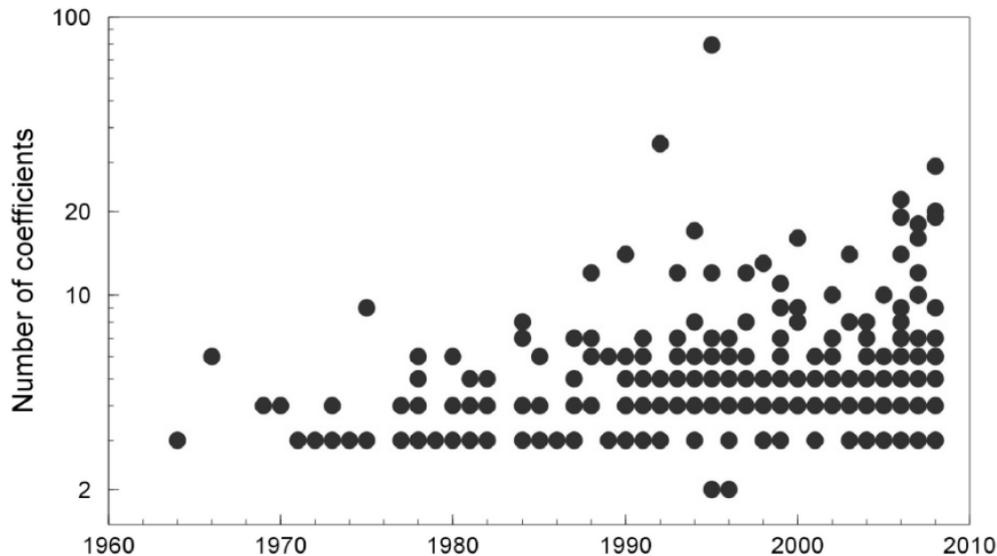


Figure.I.30 Nombre de paramètres utilisés dans les GMPEs entre 1964-2009 ([Bommer et al. 2010](#))

Conclusion

À la lumière de ce qui a été présenté dans ce chapitre, nous pouvons dire que l'amélioration de la prédiction des mouvements forts nécessite :

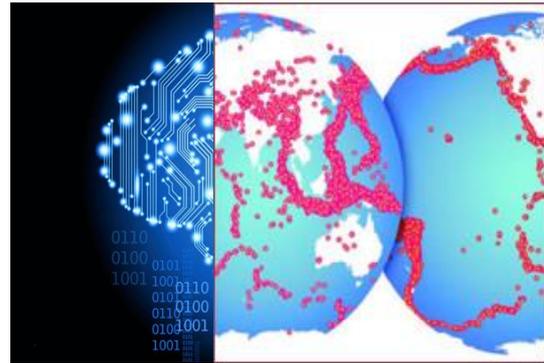
- 1) De comprendre les mécanismes qui influent sur la nature du signal en surface.
- 2) D'acquérir toujours plus de données de très haute qualité.
- 3) De caractériser au mieux les métadonnées (M_w , R , V_{s30} ...) et les paramètres du mouvement sismique recherchés (PGA, PGV et PSA, etc.)
- 4) De choisir de manière optimale la forme fonctionnelle (s'il s'agit d'une approche classique) avec des paramètres d'entrées significatifs pour bien modéliser les phénomènes physiques sous-jacents (ex. M , R et V_{s30} , région...).
- 5) Une bonne quantification des incertitudes lors de l'élaboration du modèle.

Le non-consensus autour d'une forme fonctionnelle et la complexité de ces formes (le point 4) avec la disponibilité de plus en plus de données de qualité (ex. RESORCE, KiK-Net et NGA-West 2, etc.) nous a motivé à utiliser une approche totalement dirigée par les données sans avoir à préciser une forme fonctionnelle au préalable. Cela nous ne dispense pas, bien sûr, de vérifier la physique des résultats obtenus.

Pour estimer les paramètres qui caractérisent le mouvement sismique à la surface de la Terre, nous avons choisi d'utiliser les réseaux de neurones artificiels (RNA). Dans le chapitre suivant, nous allons voir l'impact de l'utilisation des RNA sur l'évolution de la prédiction sismique en utilisant la base de données RESORCE.

Chapitre II

Développement d'une GMPE à effets-aléatoire dirigée par les données RESORCE



Introduction

Dans le chapitre précédent, nous avons vu comment l'augmentation quantitative et qualitative des données a contribué à l'évolution des équations de prédiction du mouvement sismique, en offrant aux « GMPEpists » un nombre de plus en plus important de métadonnées. Cet essor fantastique des données dans ces dernières années a donné naissance, cependant, à plus de 450 GMPEs développées entre 1964 et 2016 ([Douglas, 2016](#)). Les auteurs de ces GMPEs ont essayé de capter le maximum de phénomènes physiques régissant le mouvement sismique. Ces phénomènes qui ont été dans un temps très proche sujets inexplicés. Le but est d'avoir plus de précision de prédiction, et par conséquent, une réduction accrue de la variabilité aléatoire associée. Malheureusement, nous sommes loin de dire que nous maîtrisons tous ces phénomènes physiques. Au contraire, cet accroissement de données a donné naissance à d'autres sources d'incertitudes et à des formes fonctionnelles plus volumineuses et plus complexes. Ce constat nous a motivé à chercher d'autres alternatives moins coûteuses et qui ne nécessitent pas une grande compréhension des phénomènes physiques sous-jacents, évitant ainsi le passage par une forme fonctionnelle. Dans ce contexte, nous proposons une approche totalement conduite par les données dans laquelle nous utilisons une seule et unique fonction (nommée fonction d'activation) qui est considérée comme universelle et parcimonieuse. Cette approche est appelée Réseaux de Neurons artificiels (RNA).

Nous commençons ce chapitre par un bref aperçu sur ladite méthode neuronale et sur nos travaux antérieurs dans [Derras \(2011\)](#) en citant les étapes que nous devons suivre pour élaborer convenablement une GMPE ([Derras et al. 2012](#)). Une importance est donnée par la suite aux modèles à effets-mixtes et notamment la prise en compte des effets aléatoires en utilisant la fonction de maximisation de la vraisemblance. La démarche qui permet d'avoir les résidus inter-événements « between-event » et intra-événements « within-event » est exposée dans ce

chapitre en se focalisant, essentiellement sur, l'adaptation et l'intégration de l'approche classique d'[Abrahamson et Youngs \(1992\)](#) en utilisant la méthode neuronale comme un modèle fixe. Par la suite, nous élaborons une GMPE intégrant les effets aléatoires dans le modèle-neuronal-fixe en utilisant la base de données RESORCE ([Derras et al. 2014](#)). Ce chapitre est couronné par une étude comparative des différents GMPEs utilisant RESORCE. Les 5 modèles sujets de comparaison ont été publiés dans un numéro spécial de la revue « Bulletin of earthquake engineering » ([Douglas et al. 2014](#)).

II.1. Aperçu sur les réseaux de neurones artificiels

Les réseaux de neurones artificiels sont des systèmes de traitement de l'information dont la structure s'inspire de celle du système neuronal du cerveau humain. Un neurone artificiel est un opérateur mathématique simple, possédant une ou plusieurs entrées, une sortie qui résulte du calcul de la somme des entrées, pondérées par des coefficients (dits poids de connexions ou poids synaptiques) et du calcul d'une fonction non linéaire (dite fonction d'activation) de cette somme pondérée. Les poids synaptiques sont mis à jour en utilisant l'algorithme d'apprentissage de la rétropropagation du gradient.

Le réseau de neurones que nous avons mis en œuvre est de type Perceptron Multicouche unidirectionnel (PMCU). Ce dernier a une structure bien particulière : ses neurones sont organisés en couches successives. Chaque neurone d'une couche reçoit des signaux de la couche précédente $x(i)$ et transmet le résultat à la couche suivante O .

Prenons à titre d'exemple le $j^{\text{ème}}$ neurone de la couche cachée :

$$net(j) = \sum_{i=1}^N w_{ij} \cdot x(i) + b_j ; \quad 1 \leq i \leq N; \quad 1 \leq j \leq N_h \quad (\text{II.1})$$

N et N_h représentent le nombre de neurones dans la couche précédente et actuelle, respectivement.

La sortie du même neurone est exprimée par :

$$O(j) = f(net(j)) \quad (\text{II.2})$$

Le coefficient w_{ij} représente le poids de connexion entre le $i^{\text{ème}}$ neurone de la couche d'entrée et le $j^{\text{ème}}$ neurone de la couche cachée. f représente la fonction d'activation, tandis que b_j est le biais du neurone j .

L'algorithme de la rétropropagation du gradient est utilisé pour l'apprentissage. Cette méthode se divise en deux étapes : une phase de propagation (qui consiste à présenter une configuration

d'entrée au réseau puis à propager cette entrée de proche en proche de la couche d'entrée vers la couche de sortie en passant par la couche cachée). La deuxième phase est celle de la rétropropagation (qui consiste, après le processus de la propagation), à minimiser l'erreur commise sur l'ensemble des exemples d'apprentissage, en modifiant les poids synaptiques de la couche de sortie et de la ou des couches cachées. Cette erreur est aussi appelée fonction de coût.

Pour optimiser la vitesse de convergence, l'algorithme de BFGS (du nom de ses inventeurs : Broyden, Fletcher, Goldfarb et Shanno) est utilisé dans cette étude. BFGS fait partie des méthodes d'optimisation dite "quasi-newtoniennes". Elle est du second ordre parce qu'elle prend en considération la dérivée seconde de la fonction de coût donc plus performante que la méthode initiale de la rétropropagation. Le nombre d'itérations nécessaire pour converger est nettement plus faible et le temps de calcul est réduit ([Robitaille et al. 1996](#)).

La fonction coût utilisée dans cette étude représente la somme des différences au carré entre la réponse mesurée et celle calculée par le RNA. Ce calcul est effectué pour l'ensemble des exemples de la base de données L :

$$MSE = \frac{1}{L.M} \sum_{l=1}^L \sum_{k=1}^M (e)^2; \quad 1 \leq k \leq M; \quad 1 \leq l \leq L \quad (\text{II.3})$$

$$e = d_l(k) - y_l(k) \quad (\text{II.4})$$

M représente le nombre des neurones de sorties du réseau. $d_l(k)$ est la valeur mesurée du paramètre à expliquer k pour l'exemple l . Tandis que, $y_l(k)$ représente la valeur calculée du paramètre à expliquer k pour l'exemple l .

II.1.1. Seuil Biais/Variance (approximation/estimation) : sélection du modèle

Le compromis biais-variance est un problème central en apprentissage supervisé (ex. réseau de neurones supervisés). Idéalement, nous voulons choisir un modèle qui représente avec précision les régularités dans le jeu de données d'apprentissage (ex. jeu de RESORCE), mais qui a aussi l'aspect de généralisation des données tests. Malheureusement, il est généralement impossible d'effectuer les deux à la fois.

Les modèles avec un faible biais sont généralement plus complexes (par exemple la régression polynomiale à plusieurs degrés), mais permettent de représenter les données d'apprentissage

avec plus de précision. Cependant, ils peuvent également représenter une partie du bruit aléatoire du jeu d'apprentissage, leurs prédictions sont donc moins précises malgré la complexité supplémentaire (problème de surapprentissage). En revanche, les modèles avec un biais plus élevé ont tendance à être relativement simples (régression polynomiale à moindre degré ou même linéaire), mais peuvent produire des prédictions de variance plus faible lorsqu'ils sont appliqués au-delà de l'ensemble d'apprentissage : problème de sous-apprentissage (Figure.II.1 et II.2). Dans ce contexte, nous pouvons décomposer l'erreur (MSE) comme suit :

$$MSE = \text{Biais}^2 + \text{variance} + \text{erreur irréductible} \quad (\text{II.5})$$

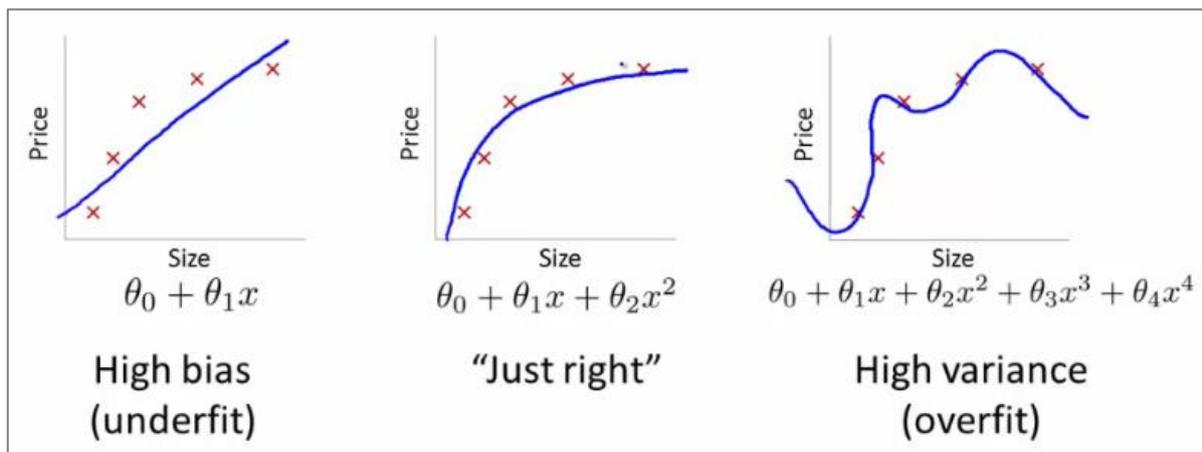


Figure.II.1 Exemple illustratif du problème de sous et sur apprentissage :biais/variance.
Source : <http://datascience.stackexchange.com/questions/361/when-is-a-model-underfitted>

L'équation II.5 se compose de :

- Le biais au carré de la méthode d'apprentissage, qui peut être vu comme l'erreur due aux hypothèses simplifiées de la méthode utilisée. Par exemple, approcher une fonction non linéaire à l'aide d'une méthode pour modèle linéaire va produire des erreurs d'estimation dues à cette hypothèse ;
- La variance de la méthode d'apprentissage, ou plus intuitivement, de combien la méthode d'apprentissage se déplace autour de sa moyenne ;
- L'erreur irréductible. Étant donné que tous les trois termes sont positifs, cela constitue une limite inférieure sur l'erreur attendue sur des échantillons tests.

Avec les réseaux de neurones, la variance augmente et le biais diminue avec le nombre de couches cachées. Comme dans le modèle linéaire généralisé, une régularisation est généralement appliquée (Figure II.2).

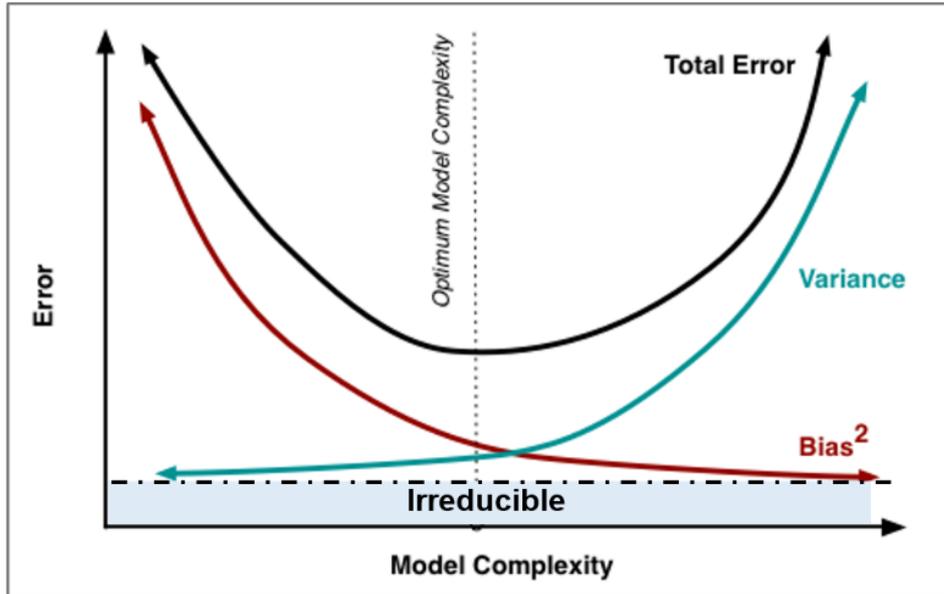


Figure.II.2 Biais et variance contribuant à l'erreur totale.

Source : <http://scott.fortmann-roe.com/docs/BiasVariance.html>

Il existe plusieurs méthodes pour obtenir le modèle optimum. La méthode utilisée dans cette thèse (et qui a été testée en la comparant avec d'autres méthodes d'optimisation dans [Derras, \(2011\)](#)) est celle de la régularisation par pénalisation des poids ([Demuth et al. 2009](#)). Cette dernière consiste à imposer des contraintes au modèle, donc à apporter une information supplémentaire sur l'évolution des poids du réseau de neurones donc sur sa complexité. Les performances des résultats obtenus par le RNA sont mesurés par l'écart type (σ) défini par :

$$\sigma = \sqrt{\gamma.MSE + (1-\gamma).MSW} \quad (\text{II.6})$$

$$MSW = \frac{1}{m} \left(\sum_{i=1}^N \sum_{j=1}^H W_{ij}^h + \sum_{j=1}^H \sum_{k=1}^M W_{jk}^o \right)^2 \quad (\text{II.7})$$

m est le nombre des poids synaptiques dans le RNA. $m=H(N+M)$

N : représente les paramètres d'entrée, H le nombre de neurones dans la couche cachée et M les neurones dans la couche de sortie. W_{ij}^h représente le poids synaptique entre la couche d'entrée et la couches cachée (h) entre le neurone i et j . Tandis que, W_{jk}^o est le poids synaptique entre la couche cachée et la couche de sortie (o) entre le neurone j et k .

Il est à noter que, La toolbox « Neural Networks » version 6.0.3 du code de calcul Matlab (R2009b) version 7.9.0 ([Demuth et al. 2009](#)) est utilisée dans ce projet de thèse pour l'élaboration de l'ensemble des modèles.

Le compromis entre la performance du modèle neuronal et la simplicité de le mettre en œuvre passe par un choix optimal du nombre de neurones. Le critère du choix est basé sur l'utilisation conjointe de σ et d'AIC ([Derras et al. 2012](#)).

Le Critère d'Information d'Akaike (AIC) est défini par la relation ([Akaike, 1973](#)) :

$$AIC(K) = N.\ln(MSE) + S.K \quad \text{si } \frac{N}{K} < 40 \quad (\text{II.8})$$

Avec N : Nombre d'exemples utilisés pour l'apprentissage,

K : Nombre des poids synaptiques dans le modèle,

S : Coefficient, généralement pris égal à 2.

II.2. Synthèse des travaux déjà réalisés

Nous continuons à travailler avec le même outil de modélisation (ANN : Artificial neural Networks) déjà utilisé dans le cadre de ma thèse de doctorat es Science soutenue en 2011 ([Derras, 2011](#)). Dans cette dernière et avec la base de données KiK-Net, nous avons étudié les deux axes de recherches suivants (Figure.II.3) :

- a. Quantification de l'amplification sismique de site par l'utilisation des rapports spectraux (BHRSR) et la détermination des paramètres significatifs du site par l'ANN1
- b. Prédiction de plusieurs grandeurs caractérisant le mouvement sismique. L'accélération et la vitesse maximale du sol (PGA et PGV) par l'ANN2 et l'ANN3 respectivement, les paramètres de nocivité (ANN4) et les ordonnées des pseudo-spectres de réponse en accélération PSA (ANN5).

Les métadonnées qui ont été utilisées sont : les paramètres caractérisant la source sismique (la magnitude du moment M_w et la profondeur focale), l'effet d'atténuation (la distance épacentrale R) et l'effet de site : vitesse moyenne des ondes de cisaillement sur z mètres de profondeur V_{sz} ([Cadet et al. 2010](#)) et la fréquence de résonance f_0 calculée à partir de la méthode H/V ([Nakamura, 1989](#)), ([Haghshenas et al. 2008](#)). Ces métadonnées représentent les entrées des modèles neuronaux ANN2-ANN5.

$$V_{sz} = \frac{Z}{\sum_i^n \frac{h_i}{V_{si}}} \quad (\text{II.9})$$

Z représente la profondeur où la vitesse moyenne est calculée, h_i est l'épaisseur de la couche numéro i et V_{si} est la vitesse des ondes de cisaillement mesurée dans la couche i .

n : nombre de couches situées entre la surface et la profondeur Z .

Il est à noter que, l'ensemble des modèles peuvent être facilement exploités à l'aide d'un tableau Excel (ex. Figure II.4).

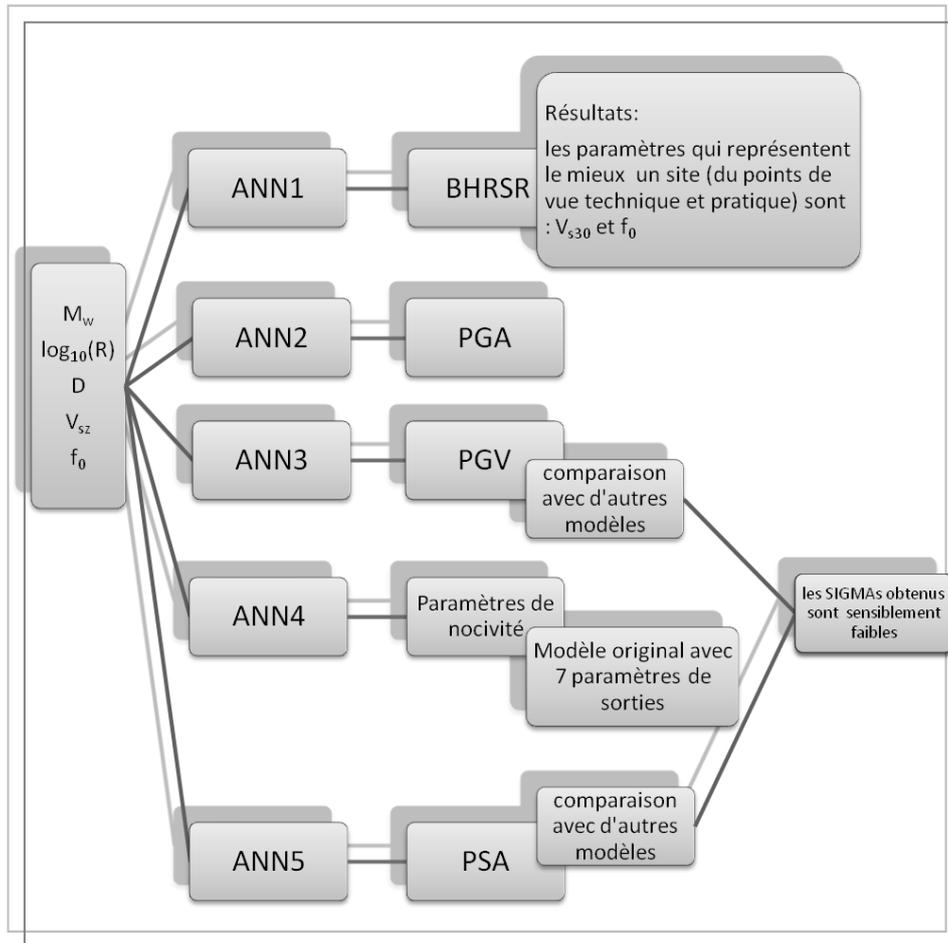


Figure.II.3 Organigramme illustrant le plan de l'estimation des paramètres caractérisant le mouvement sismique

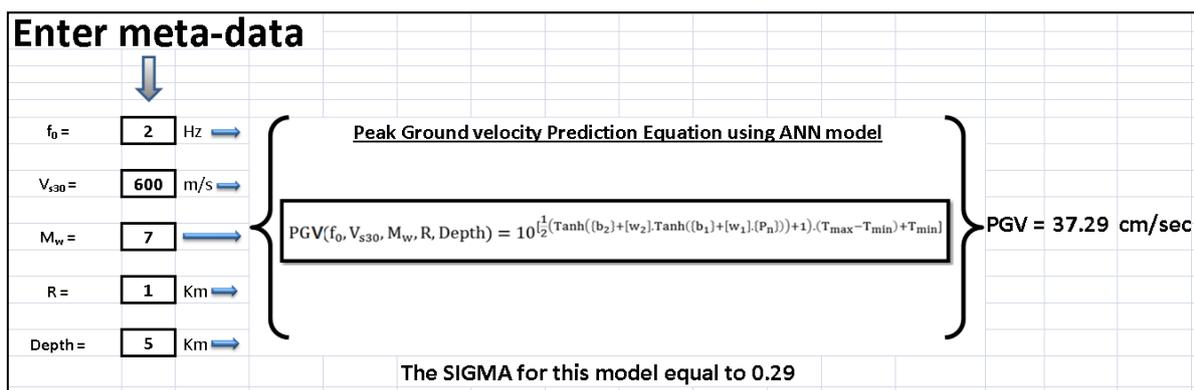


Figure.II.4 Interface graphique sous Excel du modèle d'atténuation de la vitesse maximale du sol (ANN3).

Ainsi les poids synaptiques de l'ANN2, l'ANN3, l'ANN4 et l'ANN5 sont donnés en annexe.1 de [Derras, \(2011\)](#) donnant ainsi la possibilité à une utilisation simple dans la pratique. La base de données [Cadet et al. \(2010\)](#) et [Pousse et al. \(2005\)](#) issues du réseau accélérométrique KiK-net ont servi à l'élaboration de l'ensemble des modèles neuronaux.

Pour chaque modèle les conditions suivantes doivent être vérifiées afin d'assurer leurs l'applicabilité et reproductibilité :

- 1) Le modèle ne doit pas être trop dépendant de la base de données initiale,
- 2) Le surapprentissage ne doit pas apparaître,
- 3) Il ne faut pas avoir de biais significatif et les résidus du modèle doivent suivre une loi normale à moyenne nulle
- 4) Le modèle obtenu doit représenter l'aspect physique du phénomène à modéliser.
- 5) Les tendances des formes fonctionnelles obtenues doivent être comparables aux modèles classiques.
- 6) Le domaine de validité du modèle doit suivre les données utilisées lors de l'apprentissage.

II.2.1. Quantification de l'amplification sismique par les rapports spectraux

Pour mesurer l'effet de l'amplification sismique induit par un site donné d'une manière précise, deux accélérographes doivent être installés un en surface et l'autre au rocher (ou au fond de forage caractérisant le site de référence). Le rapport des deux nous renseigne sur le taux d'amplification de site et de son contenu fréquentiel. Le KiK-Net nous donne la possibilité d'avoir ce rapport. Les caractéristiques géotechniques et géologiques de chaque profil (surface/profondeur) sont, par ailleurs, connues grâce à des tests Down-Hole.

Ce type de données nous a permis d'établir une fonction spectrale « BHRSR » qui caractérise l'effet de site en utilisant la méthode de réseau de neurones artificiels.

Le modèle neuronal élaboré nous donne le tracé complet du rapport spectral et ce à partir seulement des paramètres de site classiquement connus. Aucune forme fonctionnelle n'est exigée. Le RNA est capable de la déduire à partir de ces fonctions d'activation non linéaires et par le concept d'apprentissage.

Cette méthode donne une approximation acceptable de la fonction BHRSR, à condition de respecter les 6 conditions suscitées. L'un des problèmes connus lors de la construction du modèle neuronal, auquel il faut faire attention, est le sur apprentissage, lié au critère d'arrêt de l'apprentissage et au choix de nombre de neurones dans la couche cachée. Une méthode hybride qui combine la technique de l'arrêt prématuré et la régularisation par pénalisation des poids a

été utilisée. Ce concept nous a permis d'obtenir le nombre de neurones optimum qui est pris égal à 5. Le modèle neuronal élaboré est caractérisé par ces 4 entrées : f_0 , la profondeur du fond de forage D_{bh} , la vitesse au niveau du fond de forage $V_{z_{bh}}$ et V_{s30} .

Ce modèle présente un $\sigma=0.15$ sensiblement plus faible que celui trouvé par [Cadet et al. \(2010\)](#) égal à 0.17. En plus, le ANN1 ne présente pas de biais significatif et il suit une loi normale. Le modèle neuronal ANN1 nous a permis de :

- Générer et tracer les rapports spectraux en tenant compte de l'influence des conditions locales de site,
- Déterminer le couple de paramètres pertinents qui donne l'incertitude la plus faible et le poids le plus grand. Les 4 paramètres sujets de recherche sont « f_0 , D_{bh} , $V_{z_{bh}}$, V_{s30} ». Il en résulte que la pertinence (technico-économique) joue en faveur du couple f_0 et V_{s30} .
- Le modèle tient compte à la fois de l'amplitude des BHRSR et de son contenu fréquentiel.
- En tenant compte de l'aspect pratique et de l'influence que représentent f_0 et V_{s30} sur les BHRSR, ces deux paramètres ont été utilisés comme proxys de site lors de l'élaboration des modèles de prédiction du mouvement sismique.

II.2.2. Prédiction des grandeurs scalaires caractérisant le mouvement sismique

Les paramètres scalaires PGA, PGV, CAV (Vitesse absolue cumulative), S_1 (Intensité d'Housner), I_a (Intensité d'Arias), D_s (Durée significative), a_{rms} (accélération quadratique moyenne) ont été estimées toujours en utilisant la méthode de réseau de neurones artificiels.

Un modèle d'atténuation a été développé pour prédire les paramètres scalaires d'ingénierie caractérisant le mouvement sismique. Ces paramètres de mouvement sismique ont l'avantage de décrire mieux le potentiel de dommages du mouvement du sol pour la plupart des structures. Ces paramètres capturent ainsi les effets de l'amplitude, du contenu fréquentiel, de la durée et de l'énergie de l'enregistrement sismique. Ils représentent donc, un outil important pour l'ingénieur. L'approche neuronale permet aussi de les prédire simultanément : par un seul modèle.

Une comparaison des résultats a été effectuée entre l'ANN4 d'une part et L'ANN2 (Figure.II.3) qui donne le PGA ([Derras et al. 2012](#)) et l'ANN3 qui estime le PGV ([Derras et al. 2013](#)). Il s'est avéré que, les prédictions en termes de la médiane et l'incertitude sont les mêmes ce qui

présente un autre avantage de l'approche neuronale. Cette approche permet donc de, prédire plus d'une variable à la fois, contrairement aux modèles classiques par régression empirique qui fournit seulement une variable par équation.

Les valeurs de PGA, PGV, CAV, S_i , I_a , D_s et a_{rms} ont été estimées en utilisant M_w , R_{epi} , Depth, f_0 et V_{s30} choisis comme paramètres d'entrée ([Derras et al. 2011](#)). L'apprentissage a été réalisé par l'algorithme de la rétropropagation du gradient de deuxième ordre (BFGS).

Les variables d'entrée de l'ANN4 (Figure.II.3) qui contrôlent le PGA ont été adoptées après plusieurs tests en utilisant l'écart type (σ). Le test a montré que ce paramètre est sensible à tous les paramètres d'entrée. Un deuxième test a été élaboré afin de choisir la fonction d'activation qui donne un modèle optimum. Le résultat montre que, la configuration avec une fonction Tanh-sigmoïde pour les deux couches converge mieux vers celles enregistrées. Le troisième test concerne le nombre de neurones dans la couche cachée. Ce dernier test a montré que 20 neurones donnent des valeurs de σ et AIC minimales (compromis biais/variance). Le surapprentissage du RNA a été évité en utilisant la technique de la régularisation (équation II.6), réduit ainsi la durée de l'apprentissage. Afin de fournir à l'ingénieur un modèle simple à utiliser, le modèle neuronal ANN2 ainsi que les deux autres, ANN3 et ANN4 ont été représentés sous forme d'équations matricielles.

La validation des trois modèles a été effectuée en confrontant les valeurs estimées par le réseau de neurones et celles enregistrées sur site tout en calculant les résidus. Les résultats obtenus montrent une grande corrélation entre les PGA générés par les ANN et ceux enregistrés et montre aussi la stationnarité du processus (pas de biais significatif). Les tendances obtenues à partir de ces modèles neuronaux confirment le fait que la décroissance du mouvement sismique avec la distance dépend de la magnitude et que cette dépendance est plus significative en champ proche. Il a été remarqué ainsi la présence de l'effet d'échelle lié à la magnitude sur les paramètres de nocivité en champs proche et intermédiaire. En plus, les ANN2 ANN3 et ANN4 nous renseignent sur les effets de la profondeur, sur l'effet des conditions locales du site et sur le comportement non-linéaire du sol.

L'effet des conditions de site sur les paramètres de nocivité est confirmé. Il en résulte un comportement linéaire pour les séismes à faibles magnitudes et non linéaire pour les tremblements de terre forts avec la présence d'une déamplification.

Le modèle ANN2 a été validé en le comparant avec les modèles de [Zhao et al. \(2006\)](#), [Cotton et al. \(2008\)](#) et [Kanno et al. \(2006\)](#). Les résultats obtenus montrent une similitude entre les

tendances obtenues par l'ANN2 et celles données par les 3 GMPEs. Le σ déduit par l'ANN2 égal à 0.339 est sensiblement inférieur à ceux donnés par les modèles de Zhao (0.391), Cotton (0.353) et Kanno (0.373).

Concernent le modèle ANN3, la comparaison a été effectuée avec [Molas & Yamazaki. \(1995\)](#), [Fukushima et al \(2000\)](#), et [Kanno et al. \(2006\)](#). Les résultats révèlent une convergence entre l'ANN3 et les modèles sujets de comparaison pour les $M_w = 6$ et $M_w = 7$.

II.2.3. Prédiction des pseudo-spectres de réponse en accélération

Par la suite nous avons essayé dans [Derras \(2011\)](#) de prédire les pseudo-spectres de réponse (PSA). Comme pour les modèles déjà élaborés, nous avons démontré que l'ANN5 suit une loi normale et que le processus lié à ce modèle est stationnaire donc non biaisé.

La quantification de l'influence des deux paramètres (f_0 et Depth) outre naturellement celles des 3 autres (M_w , R_{epi} , V_{s30}) a représenté le premier objectif. Le deuxième objectif est de montrer qu'avec un seul ANN nous pouvons prédire les PSA avec des σ comparables à ceux obtenus par les méthodes classiques. Enfin, il a été montré que cette méthode d'apprentissage statistique est capable de prendre en considération les effets sous-jacents du phénomène physique, à savoir l'effet de l'atténuation des PSA avec la distance, l'effet d'échelle lié à la magnitude, l'effet de la profondeur focale et l'effet de site.

A l'issue de cette étude, nous avons montré la pertinence de f_0 notamment pour les sites profonds et que la profondeur focale a peu d'influence sur les valeurs spectrales. L'ANN5 confirme le fait que les PSA à longues périodes sont contrôlés majoritairement par la magnitude et en courtes périodes par la distance. Les σ donnés par l'ANN5 sont plus faibles que ceux obtenus par [Kanno et al. \(2006\)](#) et [Cotton et al. \(2008\)](#). Par ailleurs, la valeur de σ est comparable à celle de [Zhao et al. \(2006\)](#) : (Figure II.5).

Finalement, nous pouvons dire que, la méthode neuronale est une technique que l'ingénieur peut l'utiliser pour prédire le mouvement sismique (avec une erreur sensiblement inférieure aux GMPEs classiques) sans avoir besoin de modèle physique prédéfini combinant les effets du mécanisme de rupture de la source sismique, de la propagation d'ondes sismiques entre la source et le site et de l'effet de site.

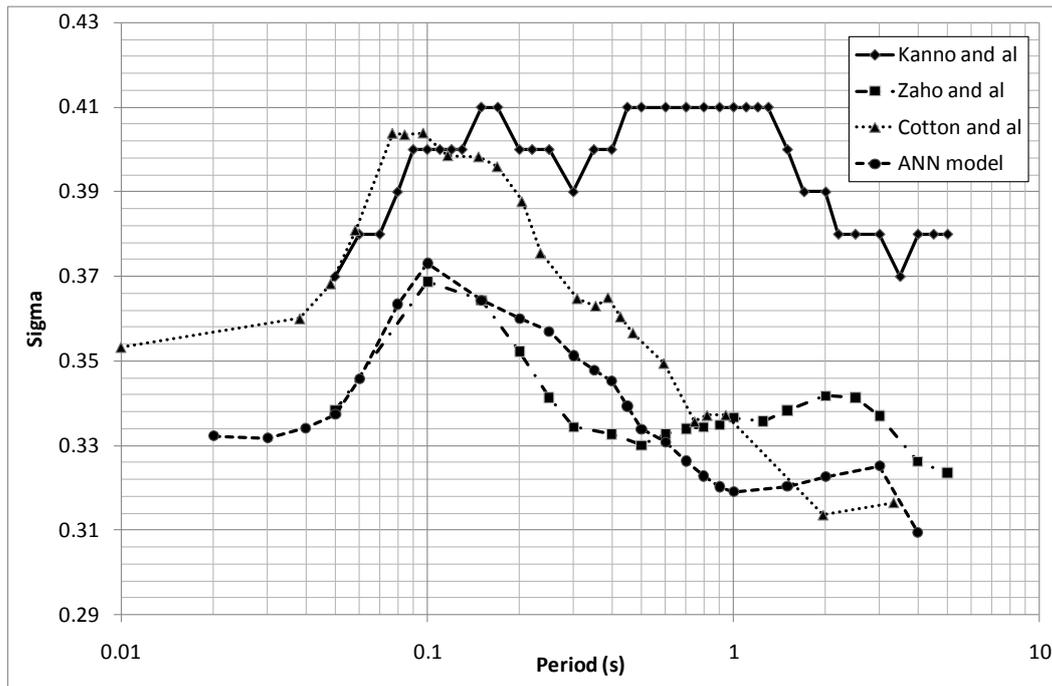


Figure.II.5 Comparaison entre les σ des pseudo-spectres de réponses obtenus par le modèle neuronal et par les GMPEs classiques.

II.3. Situation de la présente thèse par rapport à celle de 2011

Dans la présente thèse, nous utilisons le même outil de modélisation (RNA) préconisé dans [Derras \(2011\)](#), en essayant en premier lieu de tenir compte des perspectives données dans le travail de 2011, qui sont mentionnées comme suit :

On recommande pour les travaux futurs de mettre à jour la base de données KiK-Net et de combler le manque constaté en termes de mouvement fort pour les faibles distances et en termes de mouvement faible pour les grandes distances. En outre, on préconise de valider les modèles neuronaux avec d'autres bases de données de par le monde, telle que la base de données européenne, Turque « SHARE » et Californienne « NGA : Next Generation Attenuatio Projet ». Par ailleurs, il est recommandé d'introduire d'autres types de distances susceptibles de contrôler le mouvement sismique tel que la distance de (Joyner-Boore). Pour réduire le σ , on propose d'étudier la relation existante entre V_{s30} et la pente topographique « slope ».

Deuxièmement, et avec trois bases de données (RESORCE, NGA-West 2, KiK-Net) au lieu d'une seule dans [Derras \(2011\)](#), nous essayons de décomposer la variabilité aléatoire en inter-événement et intra-événement en utilisant le modèle à effet mixte (modèle fixe et à effet aléatoire). Nous avons utilisé ce type de modèle dans : [Derras et al. \(2014\)](#), [Derras et al. \(2016\)](#), [Derras et al. \(2017\)](#). En plus de l'étude de l'influence de V_{s30} et la pente topographique (modèles

continus et discrets) sur la réduction de la variabilité aléatoire, nous effectuons la même tâche avec V_{s30} mesurée et V_{s30} estimée. Et par la suite en introduisant d'autres proxys de site H_{800} et f_0 . À l'issue de ces études, nous obtenons des équations de prédiction du mouvement sismique beaucoup plus simple et interprétable que celles obtenues dans [Derras \(2011\)](#). Une autre valeur ajoutée pour cette thèse consiste à étudier la source sismique 1) l'effet de type de la faille sur le mouvement sismique (en termes d'amplitude et variabilité aléatoire) ; 2) relation existante entre Stress drop et la magnitude. En finalité, la présente thèse est une suite logique de [Derras \(2011\)](#). Dans ce qui suit nous donnons un bref aperçu sur le modèle à effet aléatoire et son intégration dans l'approche neuronale.

II.4. Modèles à effets aléatoires

Il est à noter que, cette partie est basée essentiellement sur les travaux de [Joyner et Boore \(1981\)](#), [Brillinger et Preisler \(1984\)](#), [Brillinger et Preisler \(1985\)](#), [Abrahamson et Youngs \(1992\)](#) et [Stafford, \(2014\)](#).

[Joyner et Boore \(1981\)](#) ont présenté leur approche de régression en deux étapes. L'idée était de diviser la variabilité aléatoire en deux composantes indépendantes : événement-à-événement et enregistrement-à-enregistrement. Comme son nom l'indique, l'approche de régression en deux étapes est mise en œuvre en deux temps. Tout d'abord, les paramètres qui définissent les termes d'échelle spécifiques au tremblement de terre d'un modèle sont obtenus en effectuant une analyse de régression sur un modèle de la forme ([Stafford, 2014](#)) :

$$\log(y_{ij}) = \sum_{i=1}^N \alpha_i E_i + g_l(r_{ij}, \theta_l | \beta_l) + \varepsilon_{ij}, \quad (\text{II.10})$$

Dans laquelle $g_l(r_{ij}, \theta_l | \beta_l)$ représente la forme fonctionnelle (médiane) qui est utilisée pour déduire les mouvements observés d'un événement donné i , β_l est le vecteur des coefficients de régression. N est le nombre d'événements dans la base de données. Traditionnellement, ces termes incluent l'effet d'échelle lié à la distance, ainsi que la réponse du site. Le vecteur θ_l contient toutes les autres variables qui sont utilisées afin de prédire les mouvements d'un événement donné. E_i est une variable binaire qui prend la valeur 1 pour $i^{\text{ème}}$ événement et 0 pour les autres événements. Par conséquent, α_i et β_l sont obtenus pour chaque événement dans l'ensemble de données par la régression linéaire.

Dans la deuxième étape de la régression, un modèle est ajusté aux valeurs de α_i obtenues à partir de la première étape en utilisant des termes fonctionnels qui permettent de distinguer entre les événements de différents types et tailles (équation II.11).

$$\alpha_i = g_2(M_{w,i}, \theta_2 | \beta_2) + \lambda_i, \quad (\text{II.11})$$

Dans laquelle θ_2 est un vecteur contenant des informations sur chaque événement sismique (qui pourrait inclure la profondeur : Z_{tor} , la chute de contrainte ($\Delta\sigma$), le style de la faille : FM, etc.), et λ_i est désigné comme un terme d'événement. Le terme de l'événement représente la différence entre le niveau de mouvements à partir d'un événement particulier et celui de l'ensemble des données (supposé être représentatif du niveau de la population).

Dans ce cadre, la différence totale entre une prédiction du modèle, qui est donnée par $g_1(r_{ij}, \theta_1 | \beta_1) + g_2(r_{w,i}, \theta_2 | \beta_2)$ et une observation sera égale à $\lambda_i + \varepsilon_{ij}$.

Par conséquent, cette différence totale (résiduelle) est divisée entre un terme qui représente la variabilité de l'événement à l'événement (λ_i) et un terme représentant la variabilité du site au site (ε_{ij}) à l'intérieur d'un événement donné. Dans l'hypothèse où λ_i et ε_{ij} sont indépendants, l'écart type total du modèle est donné par équation.I.22.

Il convient de noter que, dans le cas où un seul enregistrement est réalisé pour un tremblement de terre particulier :

- 1) Les termes α_i et ε_{ij} seront parfaitement corrélés (et l'augmentation d'un terme peut être exactement compensée par une diminution de l'autre), et
- 2) afin de minimiser le résidu carré, ε_{ij} doit être réglé automatiquement à zéro.

Cela signifie qu'il ne sera pas possible pour un algorithme d'identifier quelle partie de l'inadéquation (misfit) est due à un terme d'événement et quelle partie est due à un réel résidu.

Dans ce cas limité (qui n'est pas rare), l'incapacité de la fonction $g_1(r_{ij}, \theta_1 | \beta_1)$ à modéliser l'observation sera finalement poussée à travers le λ_i qui est le terme de l'événement.

Par conséquent, dans le cas où nous disposons d'un ensemble de données relativement pauvre en enregistrements par événement, la composante de la variance λ_i^2 est susceptible d'être surestimée. Le point essentiel à noter ici est que, dans l'approche en deux étapes, une estimation d'un terme d'événement est effectuée explicitement et la qualité de cette estimation dépend du nombre d'enregistrements par événement.

Le problème évoqué précédemment a été résolu (au sens statistique) par les contributions de [Brillinger et Preisler \(1984\)](#), [Brillinger et Preisler \(1985\)](#). Ce problème conceptuel que [Joyner et Boore \(1981\)](#) ont identifié et introduit est communément appelé : approche à effets aléatoires.

Dans un modèle à effets aléatoires, plutôt que de travailler avec un terme de l'événement, chaque événement sismique a un effet aléatoire associé. Ce Dernier décrit la déviation des enregistrements (pour un événement donné) de ceux d'un modèle pour toute la population. La formulation mathématique est donnée dans l'équation (II.12). Dans laquelle, plutôt que d'utiliser λ_i pour le terme de l'événement, nous utilisons à présent $\delta_{E,i}$ pour représenter l'effet aléatoire d'*i*ème événement sismique :

$$\log(y_{ij}) = f(X | \beta) + \delta_{E,i} + \varepsilon_{ij}, \quad (\text{II.12})$$

$f(X | \beta)$ Contiens tous les termes fonctionnels du modèle d'atténuation (et dont il est équivalent à $g_1 + g_2$). La matrice X englobe toutes les variables indépendantes utilisées pour la prédiction du $\log(y_{ij})$. Cette formulation peut sembler équivalente à la combinaison des deux étapes décrites par les équations (II.10) et (II.11). Mais la principale différence entre les deux approches réside dans la façon dont le paramètre $\delta_{E,i}$ est estimé par rapport à la façon dont λ_i a été estimé, et donc pour quelles valeurs de β sont finalement obtenus.

L'approche de Brillinger et Preisler (1984, 1985) obtient des coefficients de modèle en maximisant la fonction de vraisemblance, générique, donnée par équation II.13.

$$L(\beta, \sigma_E, \sigma_A | \log(y_{ij})) = \int P(\log(y_{ij}) | \beta, \delta_E, \sigma_A) P(\delta_E | \sigma_E) d\delta_E \quad (\text{II.13})$$

Cette équation peut être interprétée comme : trouver la valeur attendue de la probabilité de $\log(y_{ij})$ observée, étant donné que nous ne connaissons pas les effets aléatoires δ_E a priori.

L'approche de maximisation de la fonction de la vraisemblance est un critère d'efficacité, consiste à chercher les valeurs des coefficients à effets fixes β et les composantes de l'écart type inter-événement (σ_E) et intra-événement (σ_A) qui rendent le plus probable la réalisation $\log(y_{ij})$. La probabilité d'apparition a priori de l'échantillon en question peut alors être caractérisée par le produit des probabilités d'apparition de chacune des réalisations (puisque celles-ci sont supposées indépendantes).

Le côté gauche de l'équation donne la vraisemblance d'un ensemble de paramètres (les coefficients à effets fixes β et les composantes de la σ_E et σ_A), compte tenu des données observées. Sur le côté droit, nous intégrons le produit de la probabilité observée $\log(y_{ij})$, étant donné qu'un ensemble de paramètres des effets aléatoires est connu, sur les effets aléatoires inconnus δ_E . Par conséquent $p(\log(y_{ij})|\beta, \delta_E, \sigma_A)$ est la probabilité d'observer les données, compte tenu des paramètres du modèle et des effets aléatoires. $P(\delta_E|\sigma_E)$ est la probabilité d'avoir un ensemble d'effets aléatoires, compte tenu de la distribution relative supposée de ces effets, qui est une distribution normale de moyenne nulle avec un écart-type σ_E . Il est à rappeler que, la probabilité P est calculée en utilisant la fonction de densité de probabilité généralement symbolisée par f ,

comme mentionné précédemment, l'ensemble des paramètres à optimiser $\{\hat{\beta}, \hat{\sigma}_E, \hat{\sigma}_A\}$ sont ceux qui résultent de la valeur maximale de $L(\beta, \sigma_E, \sigma_A | \log(y_{ij}))$. Ce qui est important de noter ici est que, par opposition à l'approche en deux étapes, cette approche ne nécessite pas une estimation explicite des effets aléatoires individuels (ils ne sont pas des paramètres qui sont estimés). Cependant, une fois une estimation des paramètres du modèle a été obtenue, les estimations du maximum de vraisemblance des effets aléatoires individuels peuvent être calculées.

Par ailleurs, [Abrahamson et Youngs \(1992\)](#) ont montré que l'algorithme fourni par [Brillinger et Preisler \(1985\)](#) est instable. De mauvaises estimations des coefficients de régression pourraient être obtenues, si au début de l'analyse de régression non linéaire des estimations initiales des coefficients sont mal choisies. L'algorithme [d'Abrahamson et Youngs \(1992\)](#) a depuis été utilisé par de nombreux développeurs des GMPEs. L'avantage le plus important de l'approche [d'Abrahamson et Youngs \(1992\)](#) par rapport à l'approche en deux étapes de [Joyner et Boore \(1981\)](#) est que, l'approche à modèle à effet aléatoire aboutit à une séparation plus significative du résidu total entre les composants inter-événements et intra-événements. Dans l'exemple utilisé précédemment dans lequel un seul enregistrement a été réalisé pour un événement. Le procédé en deux étapes a alloué l'ensemble résiduel à l'événement de terme total, alors que dans le procédé d'[Abrahamson Youngs \(1992\)](#), l'estimation du maximum de vraisemblance de l'effet aléatoire s'écrit :

$$\delta_{E,i} = \frac{\sigma_E^2}{\sigma_E^2 + \sigma_A^2} \left[\log(y_{ij}) - f(X|\hat{\beta}) \right] \quad (\text{II.14})$$

Par ailleurs, et dans le cas général (équation 10 de l'article d'[Abrahamson Youngs \(1992\)](#)), $\delta_{E,i}$ deviens :

$$\delta_{E,i} = \frac{\sigma_E^2}{n_i \sigma_E^2 + \sigma_A^2} \sum_{j=1}^{n_i} \left[\log(y_{ij}) - f(X|\hat{\beta}) \right] \quad (\text{II.15})$$

[Abrahamson et Youngs \(1992\)](#) et [Joyner et Boore \(1993\)](#) ont démontré que les termes des effets aléatoires peuvent être maximisés directement sans les estimer de manière explicite. Ces termes peuvent également être obtenus par la maximisation directe de la fonction log-vraisemblance donnée par :

$$\begin{aligned} & \log(L(\beta, \sigma_E, \sigma_A | \log(y_{ij}))) \\ &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log|C| - \frac{1}{2} (\log(y_{ij}) - \mu_{\log(y)})^T C^{-1} (\log(y_{ij}) - \mu_{\log(y)}) \end{aligned} \quad (\text{II.16})$$

L'élément clé de l'équation II.16 est la matrice de covariance C qui est une matrice diagonale par blocs sa trace est égale à $\sigma_E^2 + \sigma_A^2$.

$$C = \begin{bmatrix} \sigma_A^2 I_{n_1} + \sigma_E^2 I_{n_1} & 0 & \dots & 0 \\ 0 & \sigma_A^2 I_{n_2} + \sigma_E^2 I_{n_2} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sigma_A^2 I_{n_M} + \sigma_E^2 I_{n_M} \end{bmatrix} \quad (\text{II.17a})$$

Dans l'équation II.17a I_N est la matrice identité, I_n est n×n matrice de 1. N est le nombre total d'enregistrements, n_i est le nombre de l'ième l'événement et \oplus est l'opérateur de la somme directe :

$$C = \sigma_A^2 I_N + \sigma_E^2 (I_{n_1} \oplus I_{n_2} \oplus \dots \oplus I_{n_M}) \quad (\text{II.17b})$$

La formulation de l'équation (II.17a et II.17b) suppose que, tous les enregistrements du même tremblement de terre sont corrélés avec un coefficient de corrélation égale à $\left(\frac{\sigma_E^2}{\sigma_E^2 + \sigma_A^2} \right)$ et qu'il n'y a aucune corrélation entre les enregistrements de différents événements.

Le déterminant et l'inverse de C sont donnés par :

$$|C| = \sigma_A^{2(N-M)} \prod_{i=1}^M (\sigma_A^2 + n_i \sigma_E^2) \quad (\text{II.18})$$

$$C^{-1} = \frac{1}{\sigma_A^2} I_N - \frac{1}{\sigma_A^2} \sum_{i=1}^M \frac{\sigma_E^2}{n_i \sigma_E^2 + \sigma_A^2} \quad (\text{II.19})$$

En substituant l'équation II.18 et II.19 dans II.16, nous obtenons la relation suivante :

$$\begin{aligned} & \log(L(\beta, \sigma_E, \sigma_A | \log(y_{ij}))) \\ &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} (N-M) \log(\sigma_A^2) - \frac{1}{2} \sum_{i=1}^M \log(n_i \rho \sigma_E^2 + \sigma_A^2) - \frac{1}{2\sigma_A^2} \sum_{i=1}^M \sum_{j=1}^{n_i} (\log(y_{ij}) - \bar{Y}_i)^2 \\ & \quad - \frac{1}{2} \sum_{i=1}^M \frac{n_i (\bar{Y}_i - \bar{\mu}_i)^2}{n_i \rho \sigma_E^2 + \sigma_A^2} \end{aligned} \quad (\text{II.20})$$

Dont :

$$\rho = \frac{\sigma_E^2}{\sigma_A^2} \quad (\text{II.21})$$

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \log(y_{ij}) \quad (\text{II.22})$$

$$\bar{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \log(\mu_{ij}) \quad (\text{II.23})$$

Avec $\mu_{ij} = f(X | \hat{\beta})$

Pour obtenir σ_E il suffit de résoudre :

$$\frac{\partial \log(L(\beta, \sigma_E, \sigma_A | \log(y_{ij})))}{\partial \sigma_E} = 0 \quad (\text{II.24})$$

Nous obtenons ainsi :

$$\sigma_E^2 = \frac{1}{N} \left[\sum_{i=1}^M \sum_{j=1}^{n_i} (\log(y_{ij}) - \bar{Y}_i)^2 + \sum_{i=1}^M \frac{n_i (\bar{Y}_i - \bar{\mu}_i)^2}{n_i \rho + 1} \right] \quad (\text{II.25})$$

En utilisant cette approche, l'algorithme d'estimation des coefficients à effets fixes β et les composantes de la variance σ_E et σ_A est donné par :

- 1) Estimation des coefficients de régression β en utilisant le modèle fixe (dans cette thèse il s'agit du modèle neuronal),
- 2) Estimation de σ_E et σ_A à partir des valeurs de β et des équations II.20 à II.25, et en suivant la procédure suivante :
 - a Supposition de ρ ,

- b Calcul de σ_E^2 par le biais de l'équation II.25,
 - c Calcul de σ_A^2 à partir de $\rho = \frac{\sigma_E^2}{\sigma_A^2}$ déjà supposé en a),
 - d Calcul de la valeur de la fonction de log-vraisemblance $\log(L)$ de l'équation II.20,
 - e Resupposition de ρ jusqu'à la maximisation de $\log(L)$
- 3) Estimation de $\delta_{E,i}$ par l'équation II.15,
 - 4) Calcul de nouveau β de $f(X|\beta)$ à partir de $(\log(y_{ij}) - \delta_{E,i})$ à la place de $\log(y_{ij})$ initialement utilisé en 1),
 - 5) Répétition des étapes 2, 3 et 4 jusqu'à avoir une convergence vers un maximum de $\log(L)$: valeur stable de $\log(L)$.

Généralement ce calcul itératif prend moins de 30 itérations pour se converger ([Abrahamson et Youngs \(1992\)](#)).

Cet algorithme hybride combine un modèle fixe avec un modèle à effets aléatoires. Généralement, le modèle fixe utilisé pour obtenir une GMPE est celui de la régression empirique par moindre carré. Etant donné que, l'outil de base de cette thèse est l'approche neuronale et par souci de compatibilité avec les développeurs des GMPEs, nous utilisons donc cette procédure d'[Abrahamson et Youngs \(1992\)](#) pour développer nos GMPEs et calculer les composantes de la variabilité aléatoire associée. Pour ce faire, nous sommes obligés d'adapter et d'intégrer la méthode d'[Abrahamson et Youngs \(1992\)](#) à notre modèle neuronal. Dans ce qui suit, nous donnons cette démarche. Mais avant, nous présentons la base de données REORCE utilisée dans [Derras et al. \(2014\)](#).

II.5. Présentation de la base de données RESORCE : étude de premier ordre

La collaboration et les interactions entre le laboratoire RISAM (RIsk ASsessment And Management) de Tlemcen et l'Université Dr Moualy Tahar Saida-Algérie et l'Institut des Sciences de la Terre (ISTerre) Grenoble-France, nous a permis de lancer un travail de recherche sur les GMPE neuronales en utilisant la base de données euro-méditerranéenne et de Moyen-Orient (figure I.6 et I.7).

Cette base de données, déjà évoquée dans le chapitre I, contient 5115 enregistrements et 1721 événements sismiques comme le montre le tableau II.1. Des critères de sélection sont adoptés afin d'écartier les données qui ne représentent pas à la fois des informations sur les effets de

source, de propagation d'onde et de site. Les séismes sélectionnés sont ceux où la profondeur focale ≤ 25 km (séismes crustaux), la magnitude du moment ≥ 3.5 (séismes ressentis). Les séismes dont l'information sur le mécanisme au foyer est absente sont écartés. La distance Joyner et Boore (R_{JB}) et la vitesse V_{s30} mesurée représentent l'effet de propagation d'onde et de site respectivement. Il en résulte de ce tri 1088 enregistrements de 312 événements.

La grande majorité des données sélectionnées sont d'origine italienne, 76 % en failles normales et de provenance de la Turquie 57 % en failles décrochantes (Strike-slip) figure.II.10. En outre, plus 2/3 des données sont turques et la majorité de ces données sont enregistrées en champs lointains. Inversement, la plupart des données italiennes se trouvent en champs proches (figure II.6).

Les paramètres de sélection représentent les variables explicatives du modèle neuronal sont la magnitude du moment (M_w), la distance Joyner et Boore (R_{jb}), la vitesse moyenne des ondes de cisaillement sur trente mètres de profondeur V_{s30} , la profondeur focale (Depth) et le mécanisme au foyer FM.

R_{JB} est la plus courte distance horizontale entre le site d'enregistrement et la projection verticale de la rupture comme le représente la figure II.7.

Dans [Derras et al. \(2014\)](#) nous avons essayé d'élaborer une GMPE en utilisant le modèle à effets mixtes. Dans la section suivante, nous présentons cette procédure en utilisant la base de données RESORCE.

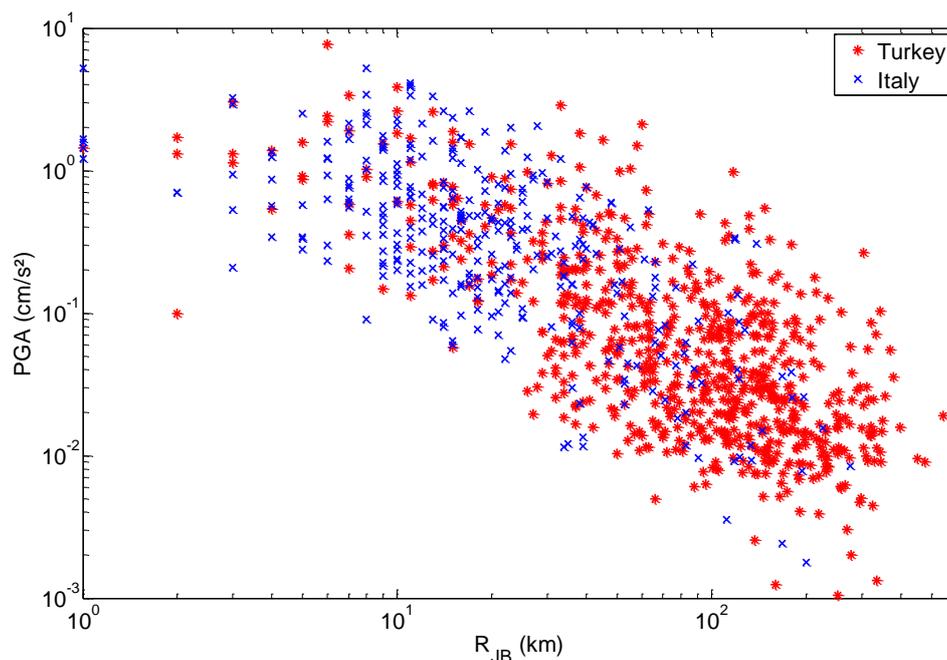


Figure.II.6. Variation des PGA italiennes et turques avec la distance

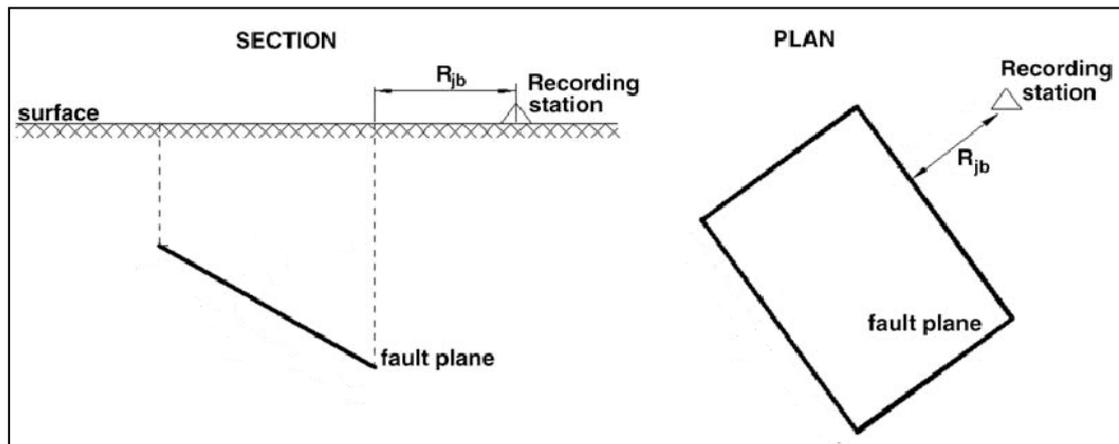


Figure.II.7 Schéma illustratif de la distance Joyner et Boore ([Scasserra et al. 2009](#))

Table.II.1 Nombre d'enregistrements et d'événements par pays : base de données RESORCE

Pays	Enregistrements	Événements	Pays	Enregistrements	Événements
Turkey	2032	763	Croatia	10	9
Italy	1390	294	Norway	10	7
Greece	498	278	Syria	10	1
Iran	396	44	Egypt	9	3
Iceland	212	48	Macedonia	9	3
Portugal	125	60	Serbia	9	9
Montenegro	58	21	Austria	7	3
Georgia	43	11	Israel	6	3
Armenia	38	13	Albania	5	4
Romania	32	4	Kyrgyzstan	5	2
Slovenia	32	14	Liechtenstein	4	1
Germany	31	9	Bulgaria	3	3
Uzbekistan	30	13	Netherlands	3	1
Algeria	28	22	United Kingdom	3	3
France	28	12	Cyprus	1	1
Spain	20	11	Hungary	1	1
Bosnia and Herzegovina	13	7	Lebanon	1	1
Switzerland	13	6			

II.6. Vers un modèle de prédiction du mouvement sismique totalement conduit par les données pour l'Europe

Les travaux de [Derras et al. \(2014\)](#) ont été regroupés et reproduits ci-dessous.

Bull Earthquake Eng
DOI 10.1007/s10518-013-9481-0

ORIGINAL RESEARCH PAPER

Towards fully data driven ground-motion prediction models for Europe

Boumediène Derras · Pierre Yves Bard ·
Fabrice Cotton

Received: 11 January 2013 / Accepted: 24 June 2013
© Springer Science+Business Media Dordrecht 2013

Abstract We have used the Artificial Neural Network method (ANN) for the derivation of physically sound, easy-to-handle, predictive ground-motion models from a subset of the Reference database for Seismic ground-motion prediction in Europe (RESORCE). Only shallow earthquakes (depth smaller than 25 km) and recordings corresponding to stations with measured V_{s30} properties have been selected. Five input parameters were selected: the moment magnitude M_W , the Joyner–Boore distance R_{JB} , the focal mechanism, the hypocentral depth, and the site proxy V_{S30} . A feed-forward ANN type is used, with one 5-neuron hidden layer, and an output layer grouping all the considered ground motion parameters, i.e., peak ground acceleration (PGA), peak ground velocity (PGV) and 5%-damped pseudo-spectral acceleration (PSA) at 62 periods from 0.01 to 4 s. A procedure similar to the random-effects approach was developed to provide between and within event standard deviations. The total standard deviation (σ) varies between 0.298 and 0.378 (\log_{10} unit) depending on the period, with between-event and within-event variabilities in the range 0.149–0.190 and 0.258–0.327, respectively. Those values prove comparable to those of conventional GMPEs. Despite the absence of any a priori assumption on the functional dependence, our results exhibit a number of physically sound features: magnitude scaling of the distance dependency, near-fault sat-

Electronic supplementary material The online version of this article (doi:10.1007/s10518-013-9481-0) contains supplementary material, which is available to authorized users.

B. Derras (✉)
Risk Assessment and Management Laboratory (RISAM),
Université Abou Bekr Belkaid, Faculté de Technologie, BP 230-13048, Tlemcen, Algeria
e-mail: boumediene.derras@ujf-grenoble.fr; b_derras@mail.univ-tlemcen.dz

P. Y. Bard
Institut de Sciences de la Terre (ISTerre), Université Joseph Fourier, CNRS, IFSTTAR,
Maison des Géosciences, BP 53, 38041 Grenoble cedex 9, France
e-mail: pierre-yves.bard@ujf-grenoble.fr

F. Cotton
Institut de Sciences de la Terre (ISTerre), Université Joseph Fourier, CNRS,
Maison des Géosciences, BP 53, 38041 Grenoble cedex 9, France
e-mail: fabrice.cotton@ujf-grenoble.fr

Published online: 14 July 2013

 Springer

II.6.1. Abstract

We have used the Artificial Neural Network method (ANN) for the derivation of physically sound, easy-to-handle, predictive ground-motion models from a subset of the Reference database for Seismic grOund-motion pRediction in Europe (RESORCE). Only shallow earthquakes (depth smaller than 25 km) and recordings corresponding to stations with measured V_{s30} properties have been selected. Five input parameters were selected: the moment magnitude M_W , the Joyner-Boore distance R_{JB} , the focal mechanism, the hypocentral depth, and the site proxy V_{S30} . A feed-forward ANN type is used, with one 5-neuron hidden layer, and an output layer grouping all the considered ground motion parameters, i.e., peak ground acceleration (PGA), peak ground velocity (PGV) and 5%-damped pseudo-spectral acceleration (PSA) at 62 periods from 0.01 to 4 s. A procedure similar to the random-effects approach was developed to provide between and within event standard deviations. The total standard deviation (σ) varies between 0.298 and 0.378 (\log_{10} unit) depending on the period, with between-event and within-event variabilities in the range 0.149 – 0.190, and 0.258 – 0.327, respectively. Those values prove comparable to those of conventional GMPEs. Despite the absence of any a priori assumption on the functional dependence, our results exhibit a number of physically sound features: magnitude scaling of the distance dependency, near-fault saturation distance increasing with magnitude, amplification on soft soils and even indications for nonlinear effects in softer soils.

Keywords: Neural networks, ground motion, RESORCE, Pseudo-Spectral Acceleration, σ .

II.6.2. Introduction

Classical Ground-motion prediction Equations (GMPEs) show significant differences in the way they model magnitude, distance and site dependencies of ground-motion parameters (e.g., [Anderson, 2000](#)). These model differences are due to an incomplete understanding of some factors (nonlinear effects, magnitude-dependent decay) that control ground-motion amplitudes. Moreover, the functional forms used for GMPEs have evolved considerably. Recent GMPEs involve more and more independent variables ([Bommer et al. 2010](#)) and the relative importance of each parameter for ground-motion predictions remains unclear.

The large amount of data collected recently by modern accelerometric networks and the development of artificial neural networks (ANN) offer the opportunity to develop “data driven” ground motion models and to evaluate the relative importance of factors which are affecting ground motions.

The ANN approach has received in recent years a growing interest from the GMPE community, concerning Peak ground motion in Europe ([Ahmad et al. 2008](#)), with the NGA database ([Alavi and Gandomi 2011](#)) or in Japan ([Derras et al. 2012](#)), along the high-speed rail tracks in Taiwan ([Kerh and Ting 2005](#)) and for Mexican subduction zone earthquakes ([García et al. 2007](#)), Peak Ground Velocity (*PGV*) in western United States ([Liu et al. 2006](#)), spectral ordinates ([Derras et al. 2010](#)).

However, ANN methods and the resulting predictive models have not yet been adopted by the engineering seismological community because of several issues presently perceived as deficiencies.

The first is the "over-fitting" problem that occurs during ANN training, which is caused by using too many unnecessary (i.e., too many degrees of freedom) neurons in the hidden layer (HL). In this case, the error on the training set is driven to a very small value, but when new data is presented to the network the error becomes large. This means the network has memorized the training examples, but it has not learned to generalize to new situations. In a recent paper ([Derras et al. 2012](#)) we have shown that the choice of an adequate regularization method helps to avoid this problem and this newly developed strategy will be adopted in this paper.

Another major concern relates to the need to disseminate ANN resulting models using simple and user-friendly tools that may replace traditional predictive analytical functional forms of GMPE's. The models developed within this study are thus implemented using ready to use

Microsoft Excel sheets, the use of which is not more complex than the one associated to the latest NGA-type GMPE's.

In this paper, the PGA , PGV and PSA are determined from five simple parameters related to magnitude, distance and site conditions. The input parameters considered here are the Joyner and Boore distance R_{JB} (km), moment magnitude M_w , the time averaged shear-wave velocity down to 30 m V_{S30} (m/s), the focal depth $Depth$ (km) and the style-of-faulting FM .

After a short presentation of the data set selection used for the present study, a section is dedicated to the presentation of the ANN approach and its specific implementation in the present application. The following section concentrates on the results of the application to the RESORCE data. The ANN has the advantage that no prior functional form is needed: the actual dependence is established directly from the data, and can therefore be used as a guide for a better understanding of the factors which control ground-motions. We then use the obtained results to discuss the respective influences of each considered input parameter and we compare this fully data-driven models with classical GMPEs previously derived from the ground motion data in Europe and the Middle East ([Akkar and Bommer 2010](#); [Akkar and Çagnan 2010](#), and [Bindi et al. 2011](#)).

II.6.3. Data set

The initial RESORCE database consists of 2520 recordings mainly in the magnitude range [3.5 - 7.6]. For this study we have eliminated all the recordings (429 in total) for which M_w is unavailable. Recordings from events with depths larger than 25 km were also removed (424), to focus on shallow crustal earthquakes. Finally, all the recordings (579) obtained at sites without measured V_{s30} values were discarded in view of minimizing the epistemic uncertainty linked with site conditions. After applying these criteria, the resulting data set consists of 1088 recordings from 320 earthquakes covering source-to-site distances up to 547 km and a magnitude range from 3.6 to 7.6 (Figure.II.8). In this database, most of the events belong to Turkey or Italy. The data from the two countries do not cover the same magnitude and distance range. Figure.II.8 indicates that the majority of Turkish earthquakes are recorded at a distance larger than 30 km (unlike Italy). Turkish events also cover a larger magnitude range than Italian earthquakes.

The site effect is considered here through the V_{s30} proxy. Only sites with measured velocities are selected in order to be able to better analyze the effect of site conditions on ground-motions.

Only very few sites (6.9 %) have a V_{S30} larger than 800 m/s, Italian sites have a slightly larger average V_{S30} (496 m/sec) than Turkish sites (389 m/sec).

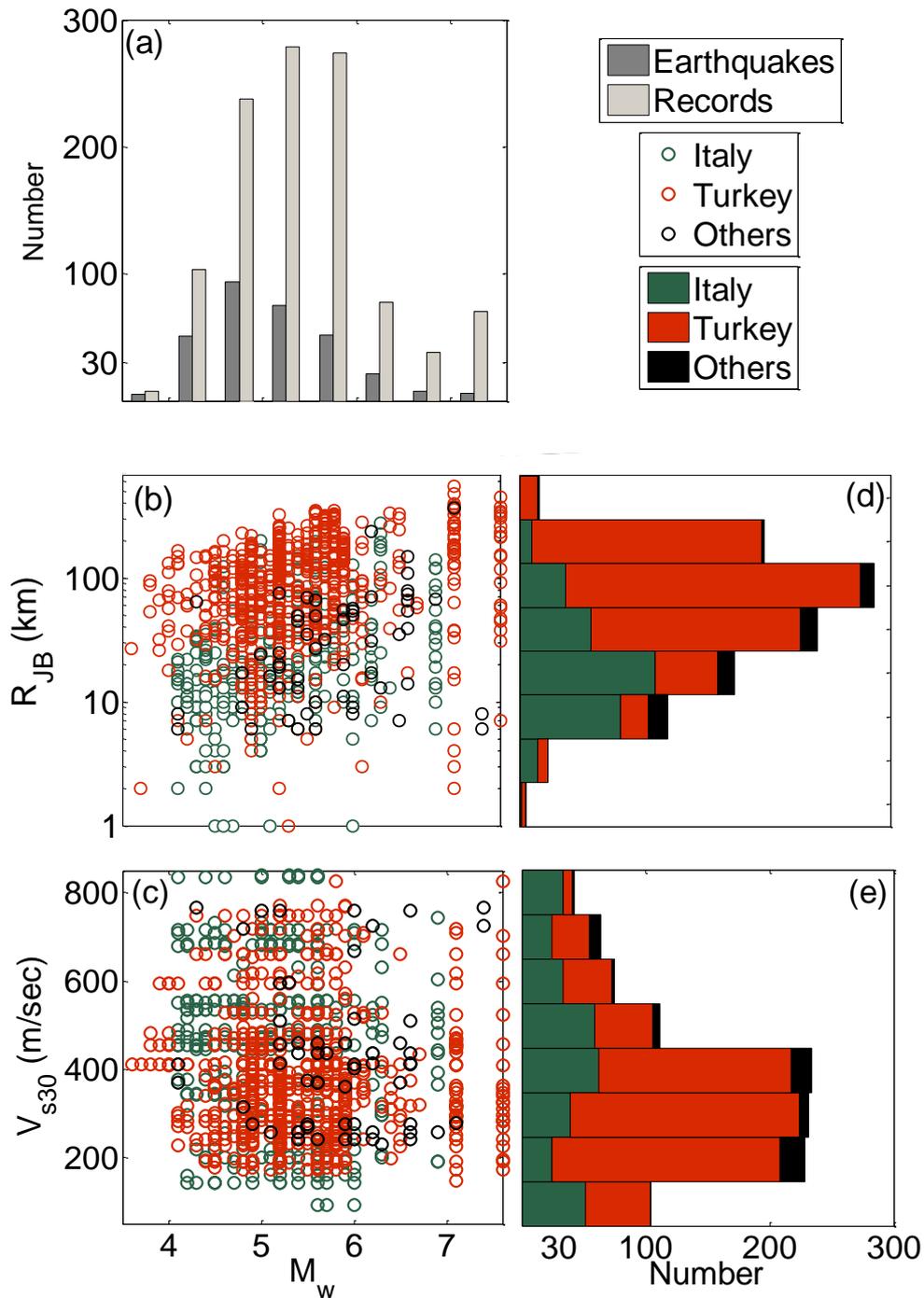


Figure.II.8 (a) Distribution of earthquakes and records according to different magnitude intervals (magnitude step 0.5). (b) distance-Magnitude-region distribution of selected recordings. (c) Magnitude- V_{S30} -region distribution of selected recordings. (d) and (e) Distribution of the data used in terms of distance R_{JB} (vector range (1 2.28 5.25 12 27.54 63.1 144.5 331.13 758.57) km and V_{S30} (interval range : 100 m/s), respectively.

The range of recorded PGA is from 10^{-3} m/s^2 to 9.69 m/s^2 . Dependence of PGA with distance is displayed in Figure.II.9 for various magnitude ranges. The figure shows clearly the decay of ground-motions with distance and with decreasing magnitude.

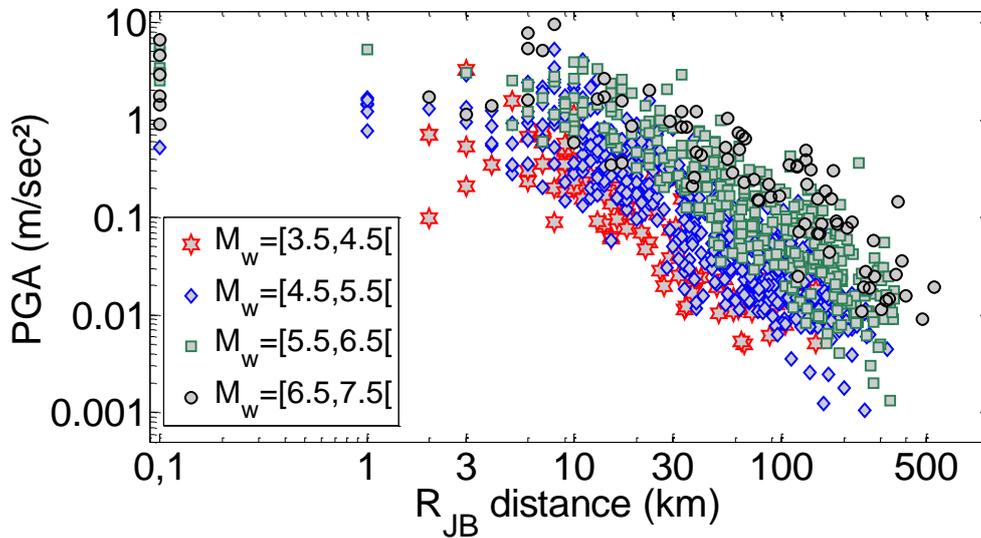


Figure.II.9 Distribution of PGA with R_{JB} distance for four magnitude intervals in the total range [3.6-7.6], using a magnitude bin size equal to ± 0.5 . The focal mechanisms (FM) of the events extracted from RESORCE are categorized into three classes (Normal: 540 records; Reverse: 93; Strike-Slip: 455), this classification is based on the plunge and rake angles (Akkar et al. 2014). The focal mechanism distribution of the dataset is shown in Figure.II.10. Most of Italian events are associated to normal faults unlike to Turkey where the style of faulting is mostly strike-slip.

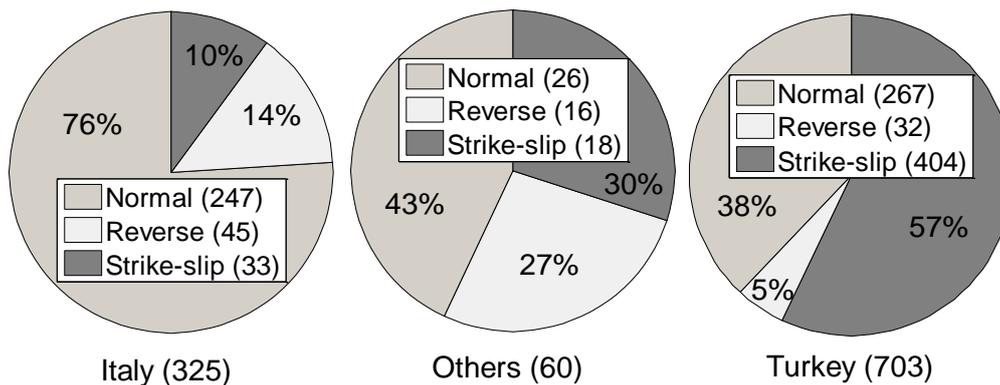


Figure.II.10 Distribution of the style-of-faulting (normal, reverse and strike-slip) by country (Italy, Turkish and others country)

We considered in this study the geometrical mean from PSA of two horizontal components (North-South and East-West).

II.6.4. Artificial Neural Network Methods with random-effects approach

An ANN is made up of interconnecting artificial neurons within input, hidden and output layers as displayed in Figure.II.11. The symbols w and b represent the synaptic weights and bias with subscripts representing the corresponding neurons between two layers respectively. The ANN type used in this work is the feed-forward ANN ([Hu and Hwang 2002](#)).

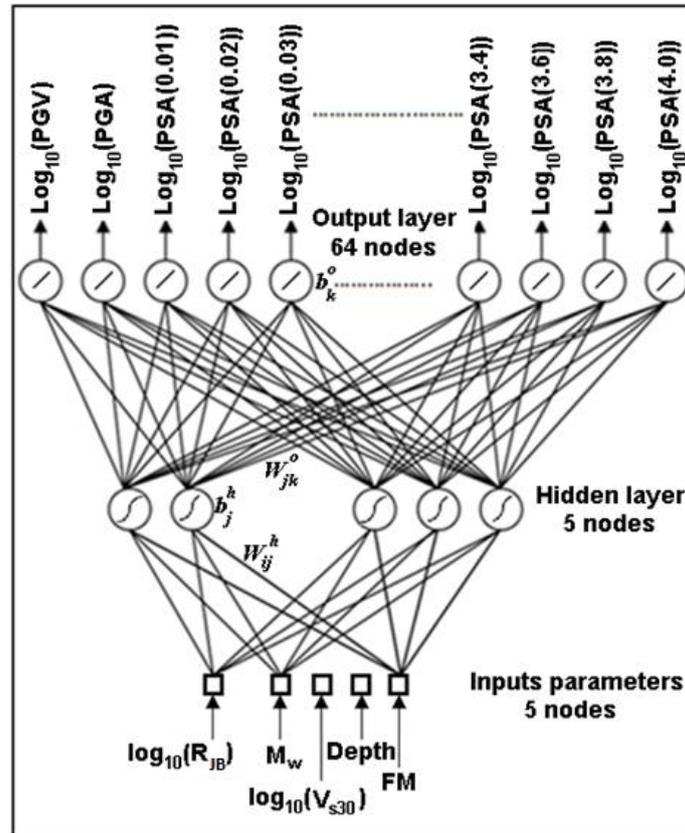


Figure.II.11 Architecture of the ANN for PGA, PGV and PSA[0.01 to 4 sec] prediction. The w_{ij}^h is the synaptic weight between the i th neuron of the input layer and the j th neuron in the hidden layer, b_j^h the bias of the j th neuron in the hidden layer. Also the w_{ij}^o is the synaptic weight between the j th neuron of the hidden layer and the k th neuron in the output layer, b_k^o the bias of the k th neuron in the output layer.

The Quasi-Newton Back Propagation technique also called “BFGS” ([Robitaille et al. 1996](#)) has been used in this work for the training phase. This method considers the second derivatives and is therefore more efficient than the original method of back-propagation. The number of iterations required to converge is significantly lower and the computation time is reduced. For smaller networks, the BFGS can be an efficient training algorithm ([Demuth et al. 2009](#)).

The performances of the results obtained by ANN are measured by the standard deviation σ value used in classical GMPEs. σ is usually decomposed into the between-event (τ) and within-event (ϕ) variabilities, which are zero-mean, independent, normally distributed random

variables with standard deviations τ and ϕ , respectively ([Al Atik et al. 2010](#)). The between-events and within-event residuals are uncorrelated, so the total standard deviation σ of the ground-motion model, σ , can be written as:

$$\sigma = \sqrt{\tau^2 + \phi^2} \quad (\text{II.26})$$

[Abrahamson and Youngs \(1992\)](#) presented a maximum likelihood approach using the expectation-maximization algorithm for the computation of model parameters and variances τ^2 and ϕ^2 . The algorithm is an iterative procedure in which the random-effects, variances, and model parameter values are computed successively. This method is now widely applied for the derivation of GMPE. In this work we have adapted this method to the neural approach. The adapted algorithm is based on the procedure described in [Abrahamson and Youngs \(1992\)](#) and can be summarized as follows:

1. Estimate an initial set of ANN model parameters: $[W_1]$, $[W_2]$, $\{b_1\}$ and $\{b_2\}$ using a fixed effects training procedure which gives the log values of PGV , PGA and of response spectrum ordinates from 0.01 sec to 4 sec, respectively (Figure.II.11),
2. Estimate τ^2 and ϕ^2 from $[W_1]$, $[W_2]$, $\{b_1\}$ and $\{b_2\}$, by maximizing the likelihood using equation (7) of [Abrahamson and Youngs \(1992\)](#),
3. Given $[W_1]$, $[W_2]$, $\{b_1\}$, $\{b_2\}$, τ^2 and ϕ^2 , estimate the random between-event term η_i (Eq. 10) in [Abrahamson and Youngs \(1992\)](#),
4. Estimate new $[W_1]$, $[W_2]$, $\{b_1\}$, $\{b_2\}$, using a fixed effects (Eq.II.27) training procedure for $(\log_{10}(Y) - \eta_i)$, $T=[0.01 \text{ to } 4]$ sec. Y represent PGV , PGA and the response spectrum of 0.01 sec to 4 sec.
5. Repeat steps 2, 3, and 4 until the likelihood in step 2 is maximized.

Using an ANN method, the remaining errors can be driven to a very small value, for instance through the use of a large number of nodes in the hidden layer. However, when new data are presented to the network the resulting error may become large. To avoid it, the regularization method is used in this paper. This method involves modifying the conventional mean sum of squares of the network errors by adding a term equal to the mean of the sum of squares of the network weights and biases ([Derras et al. 2012](#)), which de facto limits the number of degrees of freedom.

II.6.5. The ANN model

II.6.5.1. Design

The design of the ANN model requires several choices concerning the input parameters which are relevant for the outputs, the number of hidden layers and the corresponding number of neurons, and the selection of the functional form for the activation functions.

We first have chosen the independent parameters of our predictive models (inputs of the neural model). Following classical seismological studies we have considered that ground-motions are depending on magnitude and a distance metric. The hypocentral depth and style-of-faulting were added in order to test their impact on ground-motion properties. The style-of-faulting parameter corresponds to a division in three classes: normal faulting, reverse faulting and strike-slip faulting. Site effects are classically taken into account through the average shear velocity in the top 30 meters. Despite the limitations of this proxy outlined by various authors (e.g., [Mucciarelli and Gallipoli 2006](#); [Castellaro et al. 2008](#); [Assimaki et al. 2008](#); [Kokusho and Sato 2008](#); [Lee and Trifunac 2010](#); [Cadet et al. 2012](#)), it is presently the only one available for the RESORCE database. The input layer thus contains 5 input parameters, i.e., five input neurons, one for each input parameter ($\log_{10}(R_{JB})$, M_w , $\log_{10}(V_{s30})$, Depth and style-of faulting class).

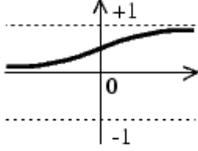
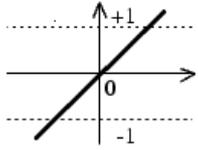
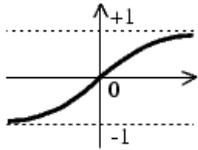
The output of the ANN are very classically $\log_{10}(PGV)$, $\log_{10}(PGA)$, and $\log_{10}(PSA)$ at 62 period values from 0.01 to 4 sec: the output layer thus contains 64 neurons. We did not include any other parameter such as PGD or severity index as CAV, Arias intensity, etc. in order to be consistent with the other papers of this special volume.

In between, a single hidden layer has been finally selected. Two or more hidden layers would have significantly increased the complexity of the model and the number of degrees of freedom, raising the issue of an "over-determination", while the σ would have shown only a marginal decrease. In addition, an ANN model with one single hidden layer has been shown to be a universal approximator of a continuous function representing the physical phenomenon ([Maass, 1997](#); [Auer et al. 2008](#)). This single hidden layer consists of five hidden neurons which is the optimal hidden neurons number ([Derras et al. 2012](#)) in order to both optimize the total σ and the Akaike Information Criterion (AIC).

The final ANN model was selected after several tests: In particular, several activation functions (between input and hidden layers, and between hidden and output layers) have been considered as listed in Table.II.2. The lowest σ value, and the lowest number of iterations as well, correspond to a hyperbolic tangent function for the former, and a linear one for the latter, which

was kept for the final model and for other sensitivity studies. The effectiveness of an ANN model to simulate highly nonlinear problems is attributed to this nonlinear activation function.

Table.II.2 Influence of different activation functions on σ and on kk : number of iteration which gives the maximum likelihood in the random effect algorithm

	Activation function of hidden layer	Activation function of the output layer	kk	Total sigma (σ)			
				PGV	PGA	PSA at 0.2 sec	PSA at 2 sec
ANN1	Log-sigmoid	Log-sigmoid 		This model does not converge			
ANN2	Log-sigmoid	Linear 	16	0.311	0.331	0.355	0.386
ANN3	Tanh-sigmoid	Tanh-sigmoid 	7	0.307	0.314	0.335	0.378
ANN4	Tanh-sigmoid	Linear	2	0.298	0.309	0.328	0.368

The “random-effects” algorithm was actually found to provide the fastest convergence, and the lowest σ values.

The way to implement these input parameters has been chosen after several tests. The evolution of the standard deviations σ obtained with different combinations of the input parameters is displayed in Table.II.3. This Table shows that the simultaneous use of the 5 parameters leads to the lowest σ , but that some parameters contribute more than others to this reduction. Despite the relatively limited gains in σ reduction obtained by the consideration of FM and depth parameters, they were finally considered in the present study in order to offer comparison with results of classical GMPEs. Five parameters, namely $\log_{10}(R_{JB})$, M_w , $\log_{10}(V_{s30})$, $Depth$ and FM , are thus used as inputs to the neural model. All the tests and the final implementation have been performed with the Matlab[®] Neural Network Toolbox[™] (Demuth et al. 2009). According to the traditional approach in the derivation of GMPE, the whole data set has been used to constrain the model, which means that the sample number for the training is the total number of available recordings and thus equal to 1088.

Table.II.3 Evolution of the overall fit of the ANN model (as measured with total σ) for PGA, PGV, PSA at 0.2 sec and PSA at 2 sec, according to the different sets of input parameters.

Models	PGA	PGV	PSA at 0.2 sec	PSA at 2 sec
$M_w + \log_{10}(R_{jb})$	0.3152	0.3188	0.3333	0.4103
$M_w + \log_{10}(R_{jb}) + FM$	0.3149	0.3186	0.3332	0.4098
$M_w + \log_{10}(R_{jb}) + Depth$	0.3145	0.3183	0.3325	0.4095
$M_w + \log_{10}(R_{jb}) + Depth + FM$	0.3141	0.3181	0.3322	0.4090
$M_w + \log_{10}(R_{jb}) + Vs30$	0.3091	0.2986	0.3290	0.3692
$M_w + \log_{10}(R_{jb}) + Vs30 + Depth + FM$	0.3090	0.2985	0.3289	0.3688

II.6.5.2. ANN model formulations

The PSA ANN model developed in this study can be implemented in simple analytical tools, providing the median \log_{10} of ground motion parameters (PGV, PSA an PSA [0.01 to 4 sec]) through the equation:

$$Log_{10}(PSA) = \frac{1}{2} \cdot \left([W_2] \cdot Tanh \left([W_1] \cdot \left(\frac{1}{2} \cdot \frac{(\{P\} - \{P_{min}\})}{(\{P_{max}\} - \{P_{min}\})} - 1 \right) + \{b_1\} \right) + \{b_2\} + 1 \right) \cdot (\{T_{max}\} - \{T_{min}\}) + \{T_{min}\} \tag{II.27}$$

in which:

The $\{P\}$ vectors the input parameter vector (5 values).

The $\{P_{min}\}$, $\{P_{max}\}$, $\{T_{min}\}$, $\{T_{max}\}$ vectors allow to normalize the input and output parameters within their actual range. Their values are listed in Table.II.4 and Table.II.5, respectively.

The $[W_i]$ matrix represents the synaptic weights, while the $\{b_i\}$ vectors represent the bias. $[W_1]$ groups the synaptic weights between the input parameters and the hidden layer, while $[W_2]$ contains the synaptic weights between the hidden layer and the output layer. $\{b_1\}$ and $\{b_2\}$ are the bias vectors of the hidden and output layers, respectively. Their values are listed in Table.II.6. The *Tanh* dependency is related to the selection of the Tanh-sigmoid activation function for the hidden layer.

Table.II.4 Preprocessing parameters for the ANN Model

	$Log_{10}(R_{JB})$ (Km)	M_w	$Log_{10}(V_{s30})$ (m/sec)	Depth (km)	Style-of-faulting
$\{P_{min}\}$	-1	3.6	1.964	0	1
$\{P_{max}\}$	2.738	7.6	3.204	25	3

Once the $\{P_{min}\}$, $\{P_{max}\}$, $\{T_{min}\}$, $\{T_{max}\}$, $[W_i]$ and $\{b_i\}$ values are tabulated, the predicted value can easily be obtained using a calculator capable of performing simple matrix operations.

Table.II.5 Postprocessing parameters for the ANN Model σ , τ and ϕ standard deviations

Period (sec)	{T _{min} }	{T _{max} }	τ	ϕ	σ
PGV	-3.849	-0.061	0.149	0.258	0.298
PGA	-2.979	0.981	0.155	0.267	0.309
0.01	-2.985	0.991	0.155	0.268	0.310
0.02	-2.986	1.008	0.157	0.270	0.313
0.03	-2.984	1.089	0.160	0.276	0.319
0.04	-2.983	1.157	0.162	0.279	0.323
0.05	-2.957	1.157	0.163	0.281	0.325
0.075	-2.946	1.261	0.165	0.284	0.328
0.1	-2.936	1.411	0.168	0.290	0.335
0.11	-2.932	1.458	0.170	0.293	0.338
0.12	-2.931	1.445	0.170	0.292	0.338
0.13	-2.929	1.453	0.169	0.292	0.337
0.14	-2.920	1.385	0.170	0.293	0.338
0.15	-2.912	1.472	0.169	0.292	0.337
0.16	-2.871	1.490	0.168	0.290	0.335
0.17	-2.859	1.491	0.168	0.289	0.334
0.18	-2.873	1.487	0.167	0.288	0.333
0.19	-2.850	1.504	0.165	0.285	0.330
0.2	-2.835	1.535	0.164	0.284	0.328
0.22	-2.782	1.409	0.164	0.284	0.328
0.24	-2.759	1.261	0.164	0.282	0.326
0.26	-2.767	1.233	0.163	0.281	0.325
0.28	-2.754	1.267	0.161	0.279	0.322
0.3	-2.685	1.269	0.161	0.279	0.322
0.32	-2.699	1.248	0.163	0.282	0.326
0.34	-2.657	1.262	0.164	0.284	0.328
0.36	-2.744	1.245	0.165	0.285	0.329
0.38	-2.720	1.213	0.166	0.286	0.330
0.4	-2.668	1.207	0.165	0.285	0.330
0.42	-2.619	1.184	0.165	0.284	0.329
0.44	-2.667	1.188	0.164	0.283	0.327
0.46	-2.715	1.188	0.164	0.283	0.328
0.48	-2.745	1.195	0.164	0.284	0.328
0.5	-2.792	1.163	0.164	0.283	0.327
0.55	-2.909	1.149	0.165	0.285	0.330
0.6	-2.953	1.226	0.166	0.286	0.331
0.65	-3.003	1.174	0.166	0.287	0.332
0.7	-3.091	1.128	0.166	0.287	0.332
0.75	-3.186	1.053	0.167	0.287	0.332
0.8	-3.243	1.036	0.166	0.287	0.331
0.85	-3.310	1.007	0.165	0.285	0.330
0.9	-3.348	1.022	0.166	0.286	0.331
0.95	-3.376	0.994	0.167	0.288	0.333
1	-3.412	0.967	0.168	0.290	0.335
1.1	-3.499	0.887	0.170	0.293	0.339
1.2	-3.599	0.786	0.172	0.297	0.343
1.3	-3.700	0.773	0.174	0.299	0.346
1.4	-3.788	0.749	0.177	0.305	0.353
1.5	-3.858	0.780	0.179	0.308	0.356
1.6	-3.920	0.736	0.180	0.310	0.359

Table.II.5 continued

1.7	-3.987	0.643	0.181	0.313	0.362
1.8	-4.037	0.611	0.183	0.316	0.365
1.9	-4.085	0.598	0.185	0.319	0.368
2	-4.124	0.587	0.184	0.319	0.368
2.2	-4.239	0.495	0.186	0.320	0.370
2.4	-4.344	0.438	0.186	0.320	0.370
2.6	-4.426	0.420	0.187	0.323	0.373
2.8	-4.470	0.428	0.188	0.324	0.375
3	-4.532	0.455	0.189	0.326	0.377
3.2	-4.600	0.466	0.190	0.327	0.378
3.4	-4.645	0.476	0.190	0.327	0.378
3.6	-4.695	0.474	0.189	0.326	0.376
3.8	-4.751	0.468	0.188	0.325	0.375
4	-4.799	0.445	0.188	0.324	0.375

II.6.6. Results of ANN model

II.6.7. Analysis of Residuals

The ANN model is obtained using the training-regularization procedure with use the random-effect model. In addition to the median, this approach gives us the aleatory uncertainly model. The resulting model is defined by the following equation:

$$\log_{10}(PSA_{ij}) = \log_{10}(P\hat{S}A_{ij}) + \lambda_i + \varepsilon_{ij} \quad (\text{II.28})$$

Where $P\hat{S}A_{ij}$ and PSA_{ij} are, respectively, the predicted and observed values, λ_i is the between-event residual for the event i , and ε_{ij} is the within-event residual for recording j of event i . The relationship between the between-event and within-event standard deviations τ and ϕ is given by (Eq.II.26) under the random-effects procedure. The variations of τ , ϕ and σ versus the period are presented in Figure.II.12. The values of the between-event variability τ are much lower than the within-event variability ϕ values, which is comparable with classical GMPE models and is indeed a consequence of the random-effects procedure. These values exhibit a general trend to increase with the period (Table.II.5).

Figure.II.13 displays the distribution of between-event residuals versus M_w , and of within-event residuals versus R_{JB} and V_{s30} , for two different ground motion parameters covering a wide period range: PGA and PSA (2 sec). Italian and Turkish residuals are displayed with different colors in order to evaluate the peculiarity of each region. The mean residuals standard deviations in different metadata ranges are also indicated in order to identify possible localized bias. No evidence was found of any obvious regional trend for between-event residuals: Italian

or Turkish data do not exhibit any specific trend with magnitude for any ground motion parameters.

The significant differences in distance distribution between Italian and Turkish data, already outlined on Figure.II.8a, make it uneasy to compare the within-event residuals for both regions.

There is a trend however for a slight under-prediction at short distances for Italy.

Table.II.6 Synaptic weight matrices and bias vectors for ANN model

[W1]=	2.648	-1.070	0.174	0.092	-0.014	{b1} =	-1.271
	-1.909	-0.535	-0.705	0.168	-0.027		1.513
	0.204	1.781	-0.080	0.014	0.062		0.591
	-0.693	0.442	0.776	-0.032	-0.163		-0.127
	0.016	0.218	-1.606	-0.042	0.026		-0.416
[W2]=	-0.511	0.071	0.221	0.169	0.171	{b2} =	-0.076
	-0.541	0.254	0.110	0.076	-0.020		-0.143
	-0.540	0.254	0.108	0.075	-0.022		-0.141
	-0.539	0.249	0.099	0.082	-0.018		-0.140
	-0.537	0.240	0.083	0.087	-0.018		-0.142
	-0.521	0.229	0.064	0.088	-0.027		-0.169
	-0.536	0.238	0.053	0.080	-0.041		-0.169
	-0.555	0.272	0.050	0.056	-0.076		-0.155
	-0.532	0.291	0.065	0.055	-0.096		-0.169
	-0.522	0.297	0.070	0.056	-0.099		-0.177
	-0.520	0.301	0.074	0.055	-0.103		-0.170
	-0.515	0.304	0.082	0.055	-0.102		-0.169
	-0.516	0.313	0.092	0.052	-0.102		-0.164
	-0.498	0.317	0.102	0.050	-0.098		-0.174
	-0.493	0.327	0.114	0.049	-0.095		-0.194
	-0.486	0.331	0.127	0.046	-0.092		-0.201
	-0.478	0.335	0.137	0.044	-0.088		-0.203
	-0.473	0.335	0.144	0.045	-0.083		-0.212
	-0.461	0.332	0.148	0.051	-0.074		-0.215
	-0.455	0.345	0.168	0.057	-0.064		-0.214
	-0.456	0.360	0.192	0.058	-0.055		-0.214
	-0.449	0.352	0.206	0.062	-0.045		-0.212
	-0.437	0.341	0.215	0.065	-0.031		-0.227
	-0.450	0.344	0.232	0.066	-0.022		-0.247
	-0.436	0.328	0.237	0.069	-0.011		-0.248
	-0.425	0.321	0.248	0.078	0.002		-0.252
	-0.416	0.315	0.256	0.082	0.015		-0.213
	-0.415	0.313	0.267	0.090	0.027		-0.216
	-0.415	0.314	0.276	0.101	0.040		-0.233
	-0.418	0.304	0.279	0.110	0.056		-0.243
	-0.412	0.294	0.280	0.108	0.066		-0.231
	-0.406	0.287	0.283	0.109	0.075		-0.220
	-0.404	0.283	0.290	0.111	0.082		-0.210
	-0.398	0.275	0.293	0.113	0.091		-0.198
	-0.383	0.253	0.298	0.122	0.116		-0.161
	-0.373	0.240	0.303	0.137	0.139		-0.141
	-0.374	0.229	0.313	0.147	0.161		-0.108
	-0.360	0.214	0.313	0.151	0.178		-0.094

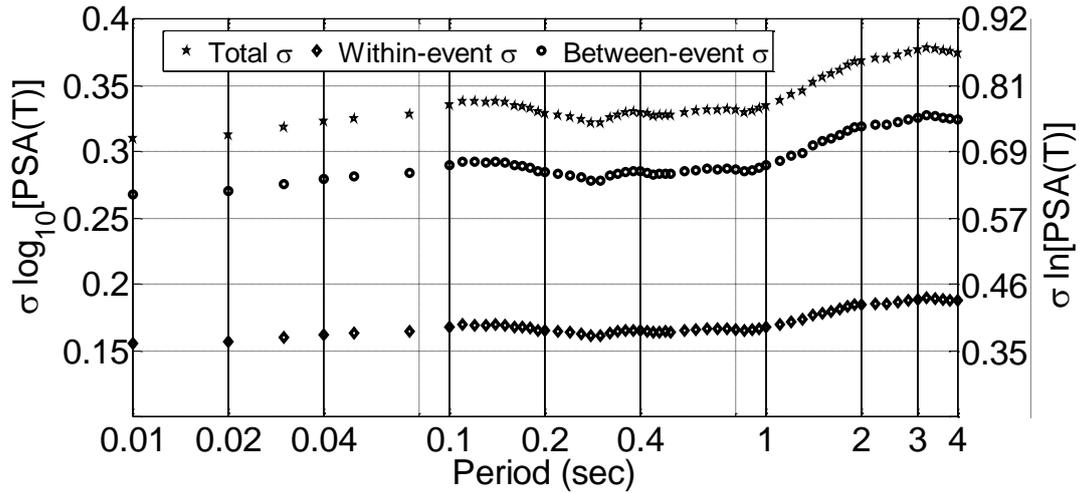


Figure.II.12 Variations of the between-event, within-event, and total σ values derived from the ANN model as a function of period.

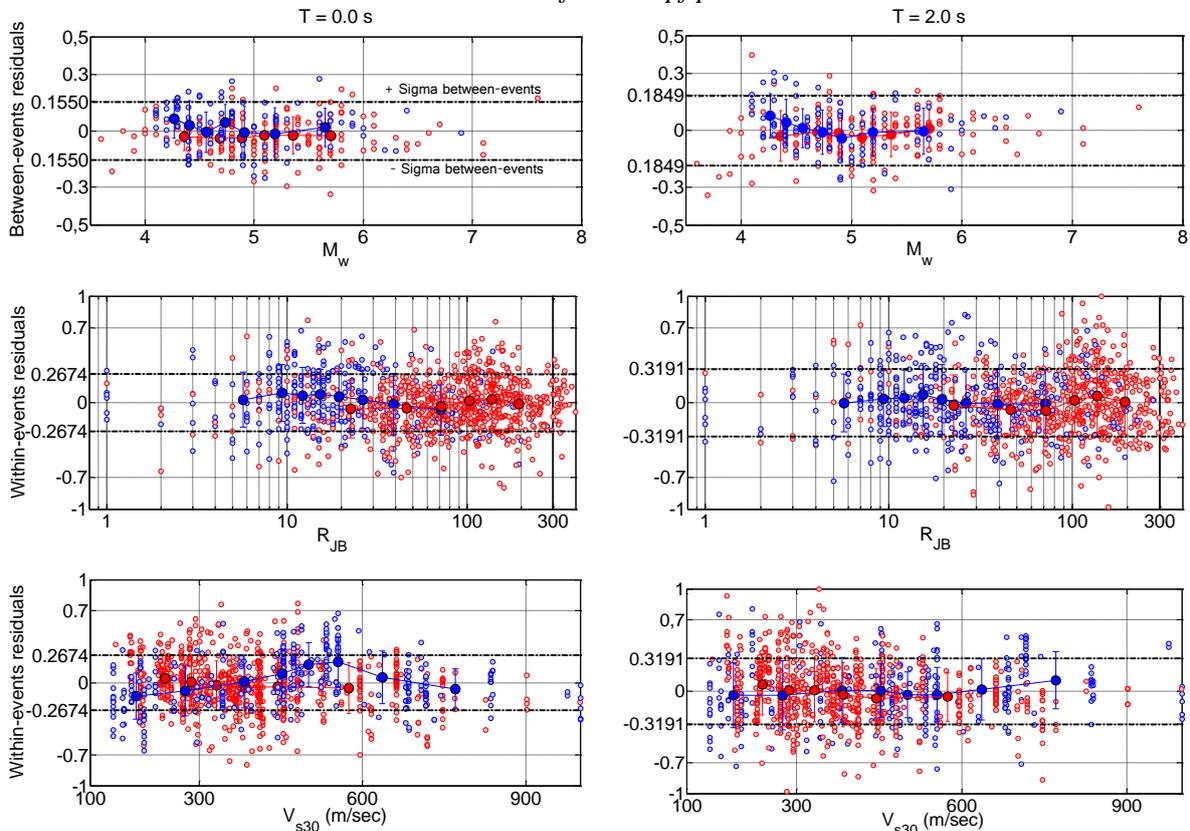


Figure.II.13 Distributions of the between-event and within-event residuals for PGA [i.e., $PSA(T=0$ sec)] and $PSA(2$ sec) as a function of M_w , R_{JB} and V_{s30} . In all graph we present the dependence of between-event and within-event residuals from Italian (blue) and Turkish (red) data. We illustrate also the Mean residuals \pm one standard deviation in the different metadata ranges. The circles show average residuals, the vertical bars are the symmetrical error bars for two standard deviations; and the horizontal bars represent 1 standard deviation. The dashed lines show the \pm sigma between (or within) –event for the corresponding ground motion parameter.

The dependency of within-event residuals with V_{s30} indicate a remarkable bias for Italian data in the range $V_{s30}=400$ m/sec to 600 m/sec, for short to intermediate period ground motion PGA , PGV and $PSA(0.2s)$ [the residuals related to PGV and $PSA(0.2s)$ are presented in the

supplementary material and [annexe 3](#)]. Such a result might be another indication that the V_{S30} proxy should not be considered as a universal proxy, and that the V_{S30} dependency should indeed be tuned regionally depending on the regional geological setting. In particular, this range of V_{S30} values might include in Italy a series of shallow soft soils overlying hard bedrock, producing short to intermediate period amplification with highly varying characteristics.

II.6.8. Key input parameters

As a complementary information to the effects of the input parameters on σ (Table.II.3), and in order to illustrate the meaning of the internal synaptic weights of the ANN approach, the total percentage of synaptic weight P_i corresponding to each the five parameters was computed by the following equation:

$$P_i = \frac{\sum_{j=1}^{N_h} |w_{ij}^h|}{\sum_{i=1}^N \sum_{j=1}^{N_h} |w_{ij}^h|} (\%) \quad (\text{II.29})$$

where the index $i \in [1 \div N=5]$ corresponds to the input parameter, and the index $j \in [1 \div N_h=5]$ correspond to the neurons of the hidden layer. The results are shown in Table.II.7.

Table.II.7 Sensitivity of the metadata parameters, P_i (%) is the total percentage of synaptic weight corresponding to each the five parameters.

Metadata parameters	P (%)
$\log_{10}(R_{jb})$	40.53
M_w	29.98
$\text{Log}_{10}(V_{s30})$	24.77
Depth	2.57
FM	2.16

As expected from the physics of ground motion, the parameter with the highest impact on ground-motions is clearly the distance $\log_{10}(R_{JB})$ followed directly by the earthquake magnitude M_w . The influence of the $\log_{10}(V_{s30})$ is also significant. The parameters with the least influence are the *Depth* and the *FM*. This ranking is similar to what has been inferred from the analysis of σ on Table.II.3, and raises the question of the actual need for including them in ground motion prediction models derived from this database.

II.6.9. PGA, PGV, and Pseudo-Spectral Accelerations ANN model curves

The attenuation curves predicted by this ANN model for *PGV* and *PSA*(2 sec) are shown in Figure.II.14 for two sites, i.e., rock ($V_{s30}=800$ m/sec) and soft soil ($V_{s30}=300$ m/sec) and four magnitude values (4, 5, 6 and 7). Unlike traditional GMPEs these dependencies are not resulting from any a priori assumptions and are fully data-driven. The shapes displayed on Figure.II.14 are however very satisfactory since they are consistent with the findings of most recent GMPEs: the magnitude scaling effect on *PGV* and *PSA* is highly non-linear in the near and intermediate fields (the ground motion parameter increase is smaller at large magnitudes than at moderate magnitudes). The results also clearly show that the decay of ground-motion with distance is magnitude-dependent: for $M_w = 7$ the behavior of *PGV* and *PSA* with distance is clearly non-linear, with saturation in the near source area (poorly constrained though, given the lack of data in this distance range), and a much smaller decay with distance than for smaller magnitudes as long as $R < 100$ km.

This figure shows also that V_{S30} does affect the ground motion. For *PGV* and *PSA* (2s) site amplification is almost the same whatever the distance and magnitude, suggesting a linear behavior at intermediate to long periods. Figure.II.14 shows that for a given magnitude and distance, *PGV* and *PSA* (2s) decrease with increasing V_{S30} .

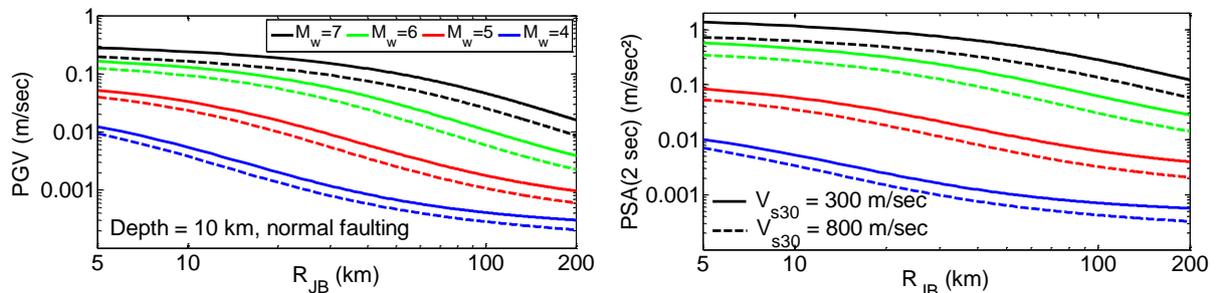


Figure.II.14 Ground-motion predictions as a function of the Joyner and Boore distance for four different magnitudes (color code) and two different site conditions (solid line, $V_{S30} = 300$ m/s; dashed line: $V_{S30} = 1$ km/s). Left: *PGA*; Right: *PSA* (2 sec). All plots are distances to a normal fault rupture.

Figure.II.15. shows the magnitude-scaling ($\partial \log_{10}(PGA)/\partial M_w$) and ($\partial \log_{10}(PGV)/\partial M_w$) versus M_w at $R_{JB} = 15$ km. The curves for both parameters exhibit a clear decrease with increasing magnitude, with a shape very similar to the one resulting from stochastic simulations (e.g. [Douglas and Jousset 2011](#)): low magnitude earthquakes show a larger magnitude-scaling factor compared to large magnitudes earthquake.

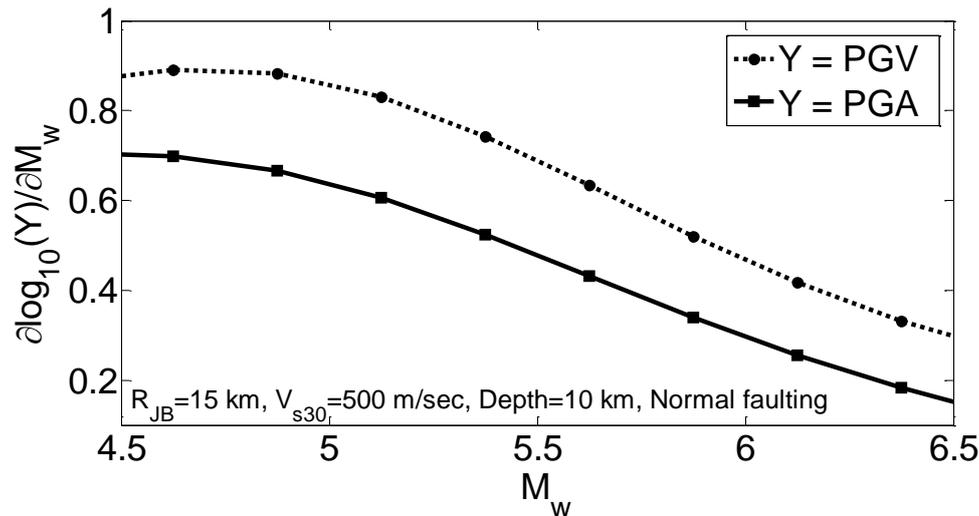


Figure.II.15 Magnitude-scaling ($\partial \log_{10}(Y)/\partial M_w$) of predicted PGA and PGV versus M_w [from 4.5 to 6.5] at $R_{JB} = 15$ km for stiff soil, a hypocentral depth equal to 10 km and a normal style of-fault.

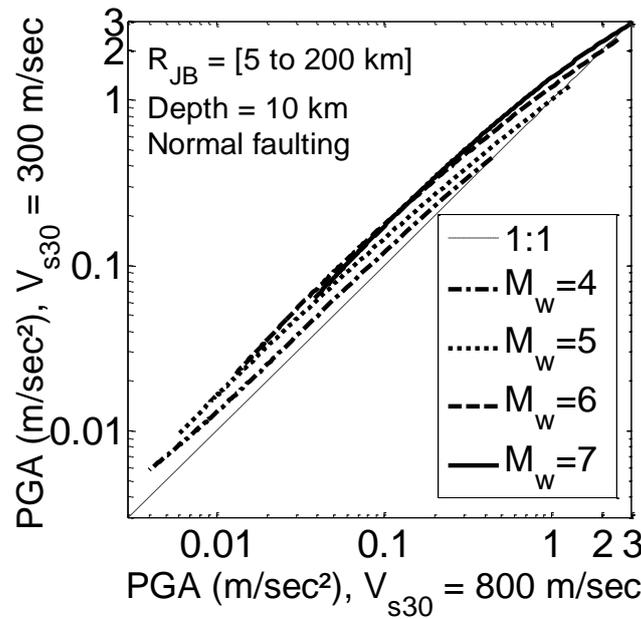
Figure.II.16 displays the variations of PGA at two different sites (soft and rock) as a function of PGA. These variations were computed for four different magnitudes (4 to 7), a fixed focal depth (10 km) and a normal faulting; Joyner and Boore distances in the range [5-200] km were considered. While the amplification associated to soft soil conditions ($V_{s30}=300$ m/sec) is apparent for all curves since they are above the 1:1 line (Figure.II.16a), the curve convexity can clearly be interpreted as some kind of non-linear (NL) behavior with a deamplification effect for magnitudes 5, 6 and 7 at large PGA levels and short distances. Figure.II.16.b displays only the site/rock PGA ratio, and thus allows to better identify the actual amplification and its decay: the amplification factor exhibits indeed a clear trend to decrease with increasing PGA values. It is premature however to draw too strong conclusions from these curves because of the small number of high PGA values in the training data set: these intriguing results however shows that the ANN approach is able to extract non-linear like behaviors which are not included in most a priori chosen GMPE functional forms because of the too poor constraint from existing datasets. In addition, a noticeable difference compared to the few presently existing GMPE non-linear models (see [Sandikkaya et al. 2013](#)), is the significant magnitude dependency of the site amplification factor, and of its decay with increasing PGA. This intriguing result might be related to the magnitude dependence of the frequency-content of ground motion, and would definitely deserve further investigations in view of improving the NL functional forms in classical GMPEs.

II.6.10. Comparison of the median ANN model with previous studies

The ANN Ground Motion Prediction Equation (Eq II.27) was compared with three other recent attenuation relations derived for the European and the Middle East region ([Akkar and](#)

[Bommer 2010](#); [Akkar and Çagnan 2010](#) and [Bindi et al. 2011](#)). Table.II.8. lists the main characteristics of the above mentioned GMPEs developed for the Euro-Mediterranean region: magnitude and distance ranges, site characterization, component definition, geographic origin of the data set. All four use the same metrics for magnitude (M_w) and distance (R_{JB}) as the present study.

(a)



(b)

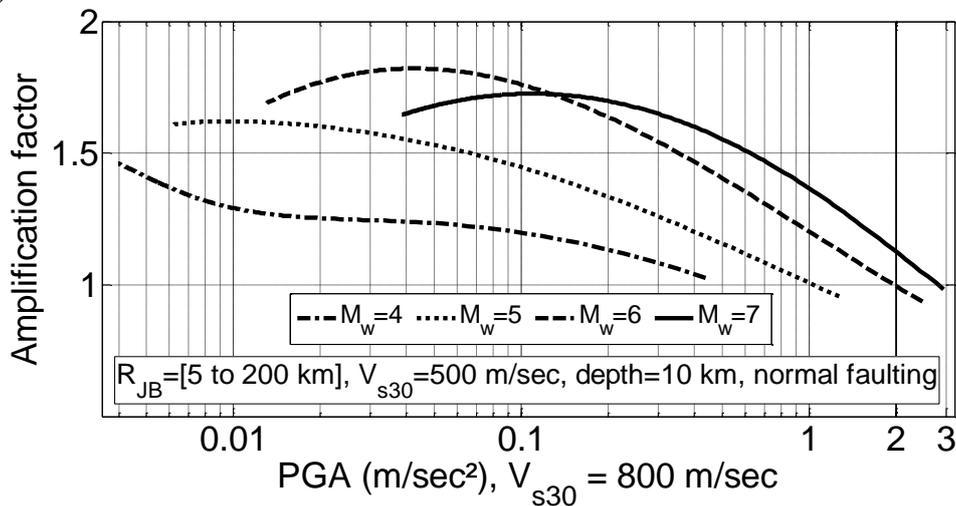


Figure.II.16 Illustration of site effects: (a) Amplification of PGA on a soft site characterized by $V_{s30} = 300$ m/sec relatively to a rock site with $V_{s30} = 800$ m/sec for different magnitude and distance ranges.

(b) Display of the NL signature in site amplification factor between a $V_{s30} = 300$ m/sec site and a standard rock site characterized by $V_{s30} = 800$ m/sec.

Table.II.9 compares the between-event variability τ , within-event variability ϕ and the total variability σ for the ANN model and for these classical GMPEs. It should be noted that the values of σ used for this comparison are those published by the authors, and that they do not

correspond to the same dataset. This comparison was made for *PGA*, *PGV*, *PSA*(0.2 s) and *PSA*(2.0 s). The model that gives the lowest σ is [Akkar and Bommer \(2010\)](#) (with a smaller magnitude range however) followed by our present model. The within-event σ of ANN is close to the one of [Akkar and Bommer \(2010\)](#) except for *PSA* (2.0).

Table.II.8 Description of validity range of the Ground-Motion Prediction Equations for Euro-Mediterranean region used in this study. N: normal, R : reverse and SS is strike-slip. M_w and R_{JB} are used by all authors

Name	M_w range	Focal depth range [km]	R_{jb} range [km]	Period range PGA, PGV	Site Class	Fault type	Horizontal componente	Main region
Akkar and Bommer (2010)	5.0-7.6	-	0-99	0.05-2 PGA PGV	3 class	N, R, SS	Geometric mean	Europe and Middle East
Akkar and Çagnan (2010)	3.5-7.6	-	0-200	0.03-2 PGA PGV	Continuous function of V_{S30}	N, R, SS	Geometric mean	Turkey
Bindi et al. (2011)	4.0-6.9	-	2.8- 100	0.03-2 PGA PGV	3 class based on sediment thickness	N, R, SS and unknown	Larger envelope	Italy
Present ANN model	4.0- 7.00	0-25	5-200	0.01-4 PGA PGV	V_{S30} Values 200-800 (m/sec)	N, R, SS	Geometric mean	Italy + Turkey

Table.II.9 Comparison between between-event σ , within-event σ and total σ for the actual study and other classical GMPEs.

GMPEs PGA	Akkar and Bommer (2010)	Akkar and Çagnan (2010)	Bindi et al. (2011)	Present ANN model
τ	0.099	0.244	0.172	0.155
ϕ	0.261	0.384	0.290	0.267
σ	0.279	0.362	0.337	0.309
GMPEs PGV	Akkar and Bommer	Akkar and Çagnan	Bindi et al.	Present ANN
τ	0.108	0.229	0.194	0.149
ϕ	0.256	0.267	0.270	0.258
σ	0.278	0.352	0.332	0.298
GMPEs PSA(0.2 sec)	Akkar and Bommer	Akkar and Çagnan	Bindi et al.	Present ANN
τ	0.108	0.242	0.209	0.164
ϕ	0.282	0.303	0.320	0.284
σ	0.302	0.387	0.332	0.328
GMPEs PSA(2.0 sec)	Akkar and Bommer	Akkar and Çagnan	Bindi et al.	Present ANN
τ	0.166	0.229	0.211	0.184
ϕ	0.283	0.267	0.308	0.319
σ	0.328	0.352	0.384	0.368

Figure.II.17 shows the PGA and $PSA(2)$ predictions for the four different GMPE as a function of the Joyner and Boore distance R_{JB} for normal faulting earthquakes of magnitudes 5, 6 and 7 (respectively) located at a 10 km $Depth$ and for a stiff site characterized by $V_{s30} = 500$ m/sec. Complementary information on the variations of $PSA(0.2)$ and PGV with magnitude and distance are given in the electronic supplement (or in [Annexe 3](#)). The predictions are compared for stiff site conditions, i.e., site classes range ($360 \text{ m/s} < V_{s30} \leq 750 \text{ m/s}$) and ($360 \text{ m/s} < V_{s30} \leq 800 \text{ m/s}$) in [Akkar and Bommer \(2010\)](#) and [Bindi et al. \(2011\)](#) models, respectively. RESORCE data corresponding to the magnitude range $M_w = 5 \pm 0.5$ (top), $M_w = 6 \pm 0.5$ (middle) and $M_w = 7 \pm 0.5$ (bottom) are also displayed in Figure.II.17. An interval equal to 500 ± 100 m/sec is taken to select these data.

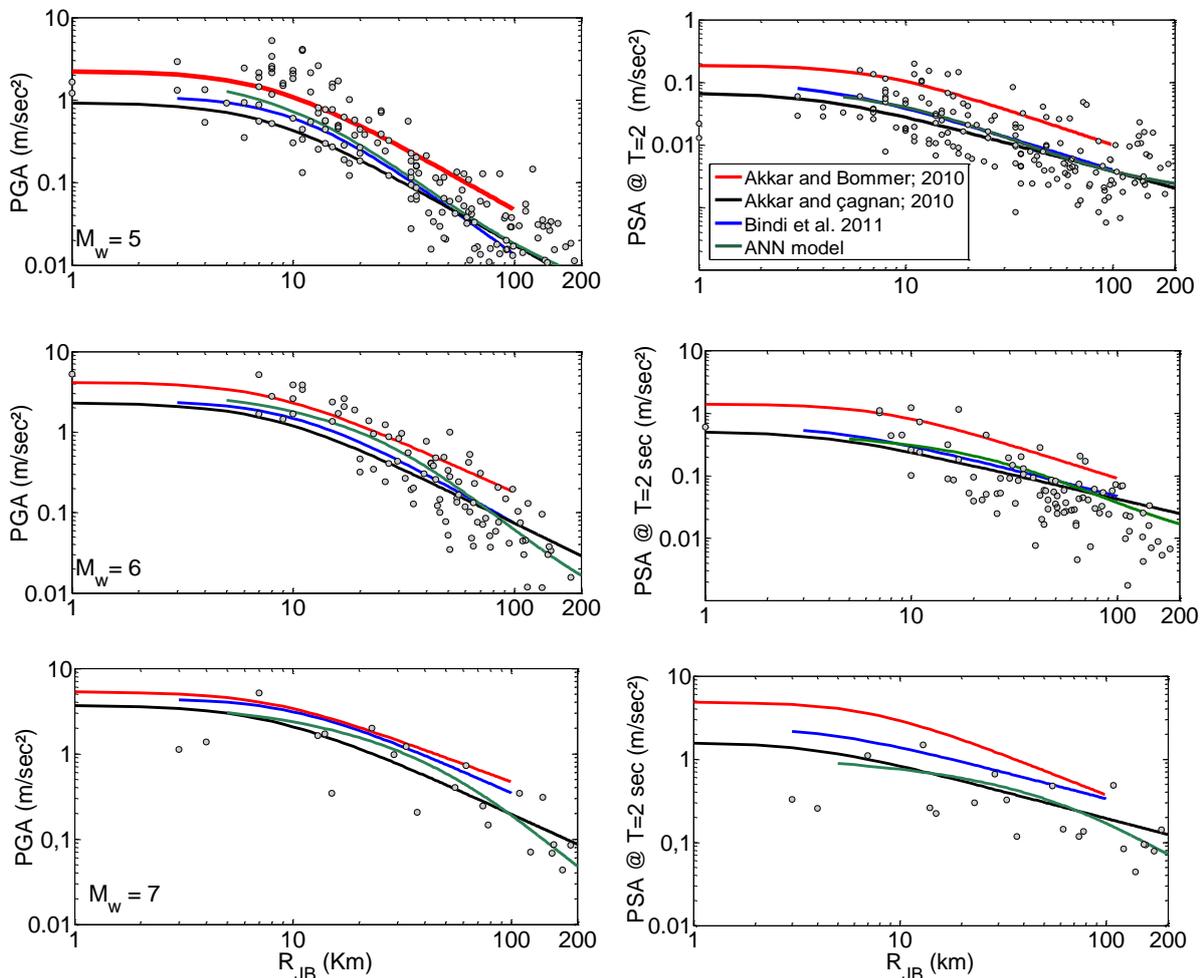


Figure.II.17 Comparison of the spatial decay of PGA and $PSA(2 \text{ s})$ predicted with the present ANN model, with those predicted by recent GMPEs, together with the corresponding RESORCE data. For a given magnitude M_w , all recordings with a magnitude equal to $M_w \pm 0.5$ and a $V_{s30} = 500 \pm 100$ m/sec, were considered $M_w = 5$ (top), $M_w = 6$ (middle) and $M_w = 7$ (bottom). The predictions are computed for $V_{s30} = 500$ m/sec, depth = 10 km and normal faulting

The distance decay shown by the ANN model (fully data driven without any a priori assumptions) is generally close to the recorded data and to models elaborated with classical methods (essentially for [Akkar and Çagnan, 2010](#) and [Bindi et al. 2011](#)). The median response spectra for $M_w=5, 6$ and 7 for normal earthquakes for soil and rock site conditions are compared in Figure.II.18. For $M_w=5$ and 6 , the predicted spectra prove similar for three models (ANN, [Akkar and Çagnan, 2010](#) and [Bindi et al. 2011](#)), while the [Akkar and Bommer 2010](#) spectra are significantly larger, especially for the $M=5$ case: this is probably due to the differences in the magnitude scaling terms. For $M=7$, the present ANN model is close to [Akkar and Çagnan 2010](#).

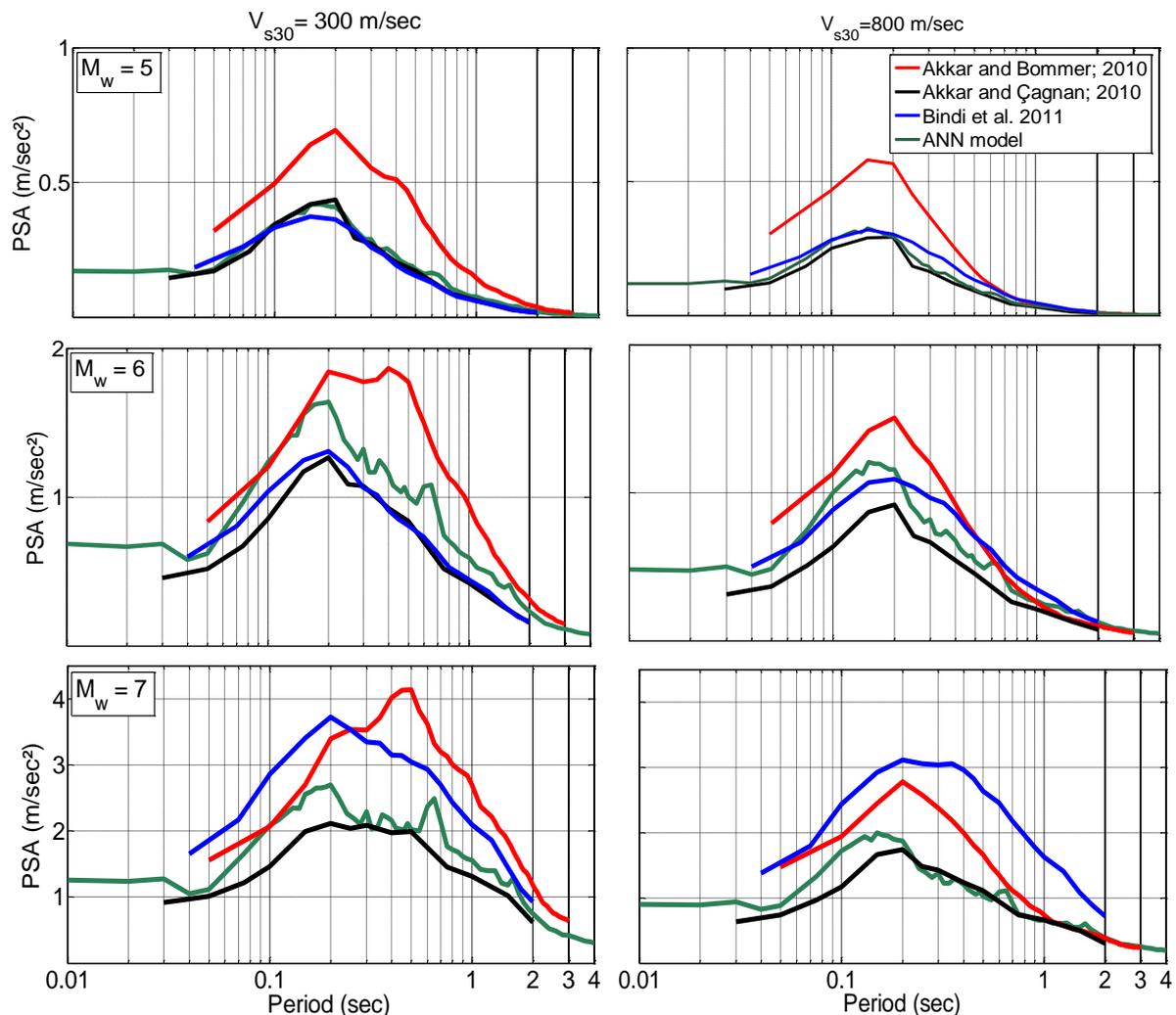


Figure.II.18 Comparison of the period-dependence of median pseudo-spectral accelerations (5% damped) derived from the present model with those proposed in a few recent European GMPEs. The rows correspond to different magnitude values: $M_w=5$ (top), $M_w=6$ (middle) and $M_w=7$ (bottom). The columns correspond to a relatively soft site ($V_{s30}=300$ m/sec, left) and a stiff soil / soft rock ($V_{s30}=800$ m/sec, right). For all plots, we consider $R_{JB}=30$ km, depth = 10 km and normal faulting.

These comparisons show that the derived ANN ground-motion model has the same ability than traditional GMPEs to predict the amplitude and variability of ground-shaking within the European area.

II.6.11. Discussion and conclusions

The ANN method has been used for the derivation of predictive ground-motion models based on a subset of the RESORCE database. Only relatively shallow earthquakes (depth smaller than 25 km) and recordings corresponding to stations with measured V_{S30} properties have been selected: the used dataset contains 1088 records from 320 earthquakes, with mainly Italian and Turkish origins, covering a magnitude range 3.6–7.6 and a distance range 1–547 km. Five input parameters were selected as they are directly available from this database: the moment magnitude M_w , the Joyner-Boore distance R_{JB} , the focal mechanism, the hypocentral depth, and the site conditions proxy V_{S30} . A feed-forward ANN type has been selected for this work, with an input layer, one 5-neuron hidden layer, and one output layer grouping all the considered ground motion parameters, i.e., peak ground acceleration (PGA), peak ground velocity (PGV) and 5%-damped pseudo-spectral acceleration (PSA) at 62 period values from 0.01 to 4 s.

The derivation used a procedure similar to the random-effects approach, so that the final standard deviation (σ) is far from being the minimum one. Its value varies between 0.298 and 0.378 (\log_{10} unit) depending on the period, with a between-event variability in the range 0.149–0.190, and a within event variability in the range 0.258–0.327. Those values prove comparable to those of conventional ground-motion prediction equations (GMPEs), unlike to the previous ANN study reported in [Derras et al. \(2012\)](#). Even though the analytical formulation is slightly more complex than for a conventional GMPE, this ANN model can very easily be implemented using a spreadsheet (as provided in the electronic supplement).

Several models and indicators were tested in order to investigate the relative importance of the various input parameters. All agree to conclude that the key parameters are, as expected, the magnitude and the distance, followed by V_{S30} . The focal mechanism and hypocentral depth exhibit only a very marginal importance. Even though this result should be moderated by the fact that these two parameters are affected by larger uncertainties for lower magnitudes, it should be taken as a warning that accounting for such parameters in GMPEs is either illusory or misleading, at least when applied to areas where the associated uncertainties are similar to those of the RESORCE database.

Despite the absence of any a priori assumption on the functional dependence, these results exhibit a number of physically sound features: magnitude scaling of the distance dependency, near-fault saturation distance increasing with magnitude, indications for nonlinear effects in softer soils. These fully data-driven outcomes corroborate results which have taken a long time

to emerge with the traditional GMPE approach involving pre-established functional forms. This encourages the use of the ANN approach which also may contribute to a better design and understanding of GMPEs functional forms especially in the case of a complex underlying physics. Our results show a significant magnitude dependence of non-linear site amplification which challenges the NL models presently used in most recent GMPEs. Another intriguing outcome of this study is the significant residuals difference exhibited by the Turkish and Italian data with respect to V_{S30} , while there are no differences with respect to distance or magnitude. This result suggests that it is worth to investigate other site condition proxies, as this difference might be an indirect hint of the regional character of this particular proxy.

Finally, one should never forget the limitations of such fully data-driven models, which are closely dependent on the chosen dataset. Such a model should NEVER be used outside the range of variability of the original dataset and we recommend not to use this ANN model outside the following ranges: M_w [4 – 7], R_{JB} [5 – 200 km], V_{S30} [200-800] m/sec. We also do not recommend the use of our model for reverse faulting style of faulting given the limited amount of data for this specific fault type in the selected database. In addition, as these limits are interdependent (for instance the maximum recorded distance depends on magnitude), another application criterion should be PGA in the range [0.01 – 10 m/s²].

II.6.12. Data and resources

The RESORCE database used in this article have been collected and disseminated by the Euro-Mediterranean Seismological Centre (EMSC) data management center at <http://jaguar.emsc-csem.org/sigma/>.

II.6.13. Acknowledgments

The authors thank S. Akkar, J. Douglas and A. Laurendeau for their generous help with computer codes of the random-effect procedure. We acknowledge the support from the Tassili program: 13MDU901 (Prédiction du mouvement sismique et estimation du risque sismique lié aux effets de site) and from the SIGMA (Seismic Ground Motion assessment) project. We also want to emphasize the background work of all strong motion network operators, without whom GMPEs and Hazard assessment studies could not exist. We also thank J. Douglas (Guest Reviewer) and anonymous reviewers for their constructive criticism and comments that helped us to improve this manuscript.

II.7. Comparaison entre les 5 GMPEs développées pour l'Europe et le Moyen-Orient

Dans cette section nous essayons d'exposer l'essentiel de l'étude comparative menée par [Douglas et al. \(2014\)](#). Vous pouvez trouver l'article complet en [annexe 2](#). Les modèles sujets de comparaison sont [Akkar et al. \(2014\)](#), [Bindi et al. \(2014\)](#), [Bora et al. \(2014\)](#), [Derras et al. \(2014\)](#) et [Hemkes et al. \(2014\)](#). Les cinq travaux ont été publiés dans un numéro spécial dans la revue Bulletin of earthquake engineering (BEE). Les cinq modèles utilisent des dérivés de la base de données RESORCE. Les cinq groupes de développeurs appliquent différents critères de sélection et d'exclusion, ce qui conduit à utiliser seulement entre 14 et 21% des données disponibles.

L'objectif de cette comparaison vise à connaître l'incertitude épistémique encore omniprésente dans l'évaluation de l'aléa sismique. Il est à noter que, les auteurs utilisent tous comme paramètres indépendants : M_w , R_{JB} , V_{s30} et FM (mécanisme au foyer). Ici deux modèles ont été obtenus avec la méthode classique : régression empirique et modèle à effet aléatoire ([Akkar et al 2014](#); [Bindi et al. 2014](#)). Le modèle [Hemkes et al. \(2014\)](#) et [Derras et al. 2014](#), par contre sont des modèles non paramétriques dérivés en utilisant des approches basées sur les données. Et [Bora et al. \(2014\)](#) ont effectué des prédictions de la réponse spectrale en accélérations en utilisant la théorie de la vibration aléatoire sur la base de modèles empiriques d'amplitude de spectre de Fourier et de la durée. Il est important de mentionner que le modèle de [Derras et al. \(2014\)](#) intègre l'effet aléatoire par maximisation de la fonction de vraisemblance.

Dans le tableau.II.10, le nombre d'événements, des stations, des enregistrements, de la gamme de magnitudes et distance utilisées pour les 5 modèles sont donnés. En outre, nous illustrons les domaines d'applications recommandés par les développeurs et les critères d'exclusion utilisés.

[Douglas et al. \(2014\)](#) a exposé dans son étude comparative, la différence existante entre ces modèles en termes de : Effet d'échelle lié à distance, - Effet d'échelle lié à la magnitude, - Effet d'échelle lié à V_{s30} , - Style de la rupture sismique, - Variabilité aléatoire et Réponse spectrale. Ces comparaisons et autres sont données dans [Douglas et al. \(2014\)](#).

Ce qui nous intéresse ici, c'est la position de notre modèle vis-à-vis de la valeur de la variabilité aléatoire et par rapport aux autres GMPEs. Dans ce contexte, nous présentons l'écart type total (σ) des 5 modèles en fonction de la période (figure.II.19).

Table.II.10 Nombre d'événements sismiques, des stations et des enregistrements utilisés pour élaborer les cinq modèles, la magnitude et la distance des plages de données utilisées, les gammes d'application recommandée par les développeurs de modèles et les critères d'exclusion utilisés pour sélectionner les enregistrements utilisés dans chaque le modèle

Model	Akkar et al.	Bindi et al. (VS30 model)	Bora et al.	Derras et al.	Hermkes et al.
Number of earthquakes (E)	221	225	369	320	279
Number of stations (S)	322	345	341	201	251
Number of records (R)	1,041	1,224	1,232	1,088	835
R/E	4.7	4.8	3.3	3.4	3.0
M_{min} to M_{max} (data used)	4.0 to 7.6	4.0 to 7.6	4.0 to 7.6	3.6 to 7.6	4.0 to 7.6
M_{min} to M_{max} (recommended)	4.0 to 8.0	4.0 to 7.6	4.0 to 7.6	4.0 to 7.0	4.0 to 7.6
R_{min} to R_{max} (km) (data used)	0 to 200	0 to 300	0 to 200	0 to 547km	0 to 200
R_{min} to R_{max} (km) (recommended)	0 to 200	0 to 300	0 to 200	5 to 200km	0 to 200
Record exclusion criteria (other than in terms of magnitude and distance)	Singly-recorded earthquakes; all three components not available; focal depth greater than 30km; sites with no measured VS30; structural period beyond usable period range defined by Akkar and Bommer (2006); events with $M_w < 5$ with fewer than 3 records to constrain SoF ratios; unknown or oblique style of faulting; not free-field.	Unknown style of faulting; sites with no measured VS30; singly-recorded earthquakes; only records with low-pass cut-off frequency lower than 20Hz and outside passband of high-pass filter all three components not available; focal depth >35 km.	Not representative of shallow crustal event; unknown style of faulting; only one horizontal component; sites with no measured VS30; poor quality record; high-pass cut-off frequency higher than Brune-source corner frequency for stress drop of 100bars.	Focal depth more than 25 km; sites with no measured VS30; unknown style of faulting	Unknown style of faulting; sites with no measured VS30; not free-field conditions; high-pass cut-off frequency higher than 0.25Hz

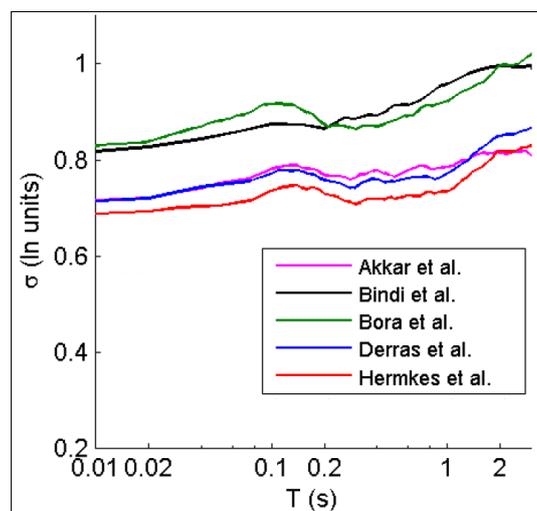


Figure.II.19 Comparaison entre l'écart type total (σ) obtenu par les 5 GMPEs

De cette figure nous pouvons constater que :

- Tous les modèles estiment le σ indépendamment de la magnitude, car il y a peu de tremblements de terre avec des grandes magnitudes dans la base de données RESORCE,
- [Bora et al. \(2014\)](#) prévoir un σ plus grand par rapport aux autres et ce en raison de son niveau élevé de σ inter-événement,
- tous les σ des 5 modèles augmentent avec la période, mais la dépendance σ -période n'est pas très significative,
- Les σ les plus faibles sont ceux de [Hemkes et al. \(2014\)](#) pour toute la gamme de périodes.

Cependant, le modèle d'Hermkes présente quelques anomalies par rapport aux autres. Ces insuffisances sont citées ici. La figure.II.20 qui représente l'effet d'échelle lié à la distance pour une magnitude 6 et une $V_{s30} = 270$ m/s et une faille de décrochement. L'analyse de cette figure montre que, le modèle de Hemkes présente un changement de pente (point d'inflexion) à partir de 40 km ce qui ne correspond pas aux observations (ex. figure.II.9 de [Derras et al. \(2014\)](#)). Le même problème est présent pour l'effet d'échelle lié à la magnitude qui est représenté dans la figure.II.21 pour une accélération spectrale (SA) à $T=1$ s vis-à-vis de la magnitude. Nous remarquons ainsi que, la tendance est inversée par rapport aux autres modèles. En outre, dans [Hemkes et al. \(2014\)](#) nous remarquons une très grande valeur du rapport d'amplification de site à $T=1$ s en comparaison avec les autres GMPEs et que la courbe est irrégulière (figure.II.22). Probablement ce modèle souffre d'un sur apprentissage.

Ces déséquilibres dans le modèle d'Hermkes le laissent à l'écart des autres modèles. Nous ne pouvons pas le considérer comme un modèle reproductible. Donc si nous écartons le modèle d'Hermkes et si nous nous basons sur l'analyse de la variabilité aléatoire nous pouvons dire que notre modèle (figure.II.19) peut être intéressant (un σ plus faible par rapport aux 3 autres) et ce pour la gamme de périodes allant de 0.1 à 1.0 s. Il est à noter que, cette gamme représente les périodes des structures courantes (bâtiments entre 1 et 10 étages). La même constatation a été trouvée par [Derras \(2011\)](#) lors de la comparaison entre le σ obtenu par notre modèle neuronal et ceux de [Kanno et al. \(2006\)](#), [Zaho et al. \(2006\)](#) et [Cotton et al. \(2008\)](#). L'analyse effectuée dans [Derras \(2011\)](#) a révélé que, notre modèle donne la plus petite valeur de σ pour l'intervalle de périodes [0.6-1.5] s.

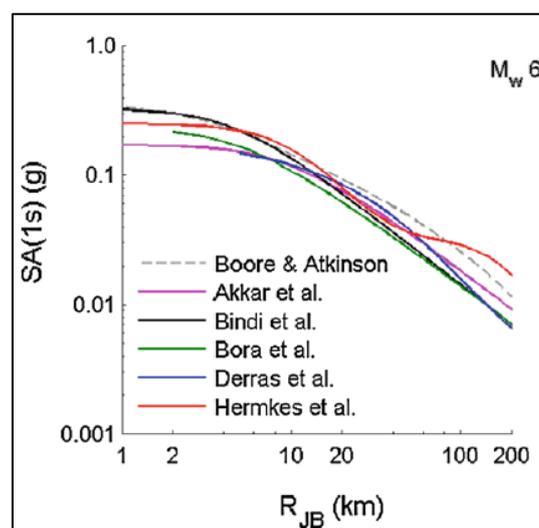


Figure.II.20 Comparaison de l'effet d'échelle lié à l'atténuation pour une faille de décrochement, $V_{s30}=270$ m/s

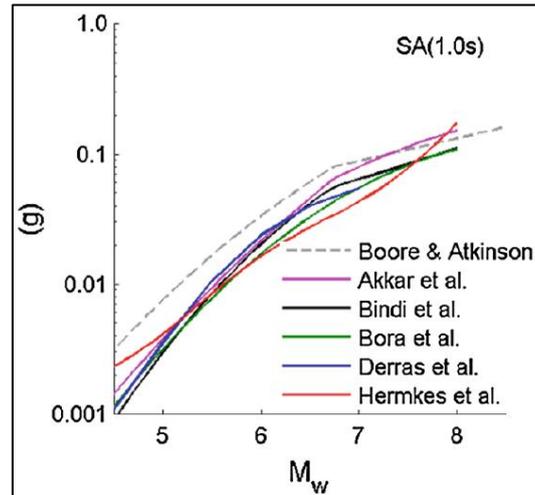


Figure.II.21 Comparaison de l'effet d'échelle lié à la magnitude pour une faille de décrochement, $V_{s30}=760$ m/s et pour $R_{JB}=30$ km.

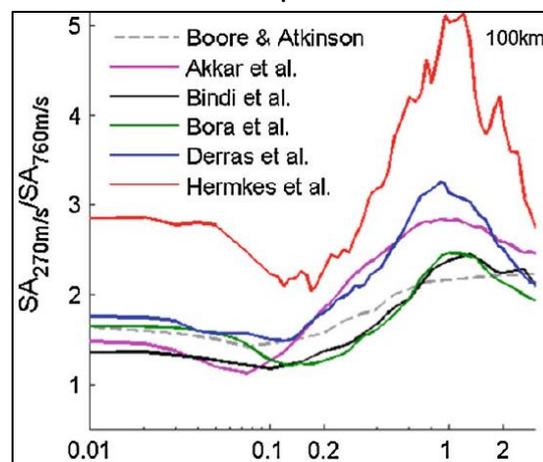


Figure.II.22 Comparaison du rapport d'amplification de site pour $V_{s30}=270$ m/s en fonction de SA pour $V_{s30}=760$ m/s : $M_w=7$, $R_{JB}=100$ km et pour une faille de chevauchement.

En revanche, il ne faut pas oublier que, le modèle [Derras et al. \(2014\)](#) présente-lui aussi 2 inconvénients majeurs. Le premier est lié à la complexité de l'équation de prédiction (médiane) équation.II.27. Et le deuxième nous le remarquons dans la figure.II.18 : là nous voyons que les valeurs du PSA pour deux périodes avoisinantes sont différentes (courbe non lissée). Cela pose problème lors de l'étude dynamique d'une structure donnée. Il est à noter, quand même, que nous avons résolu ces deux anomalies dans [Derras et al. \(2016\)](#).

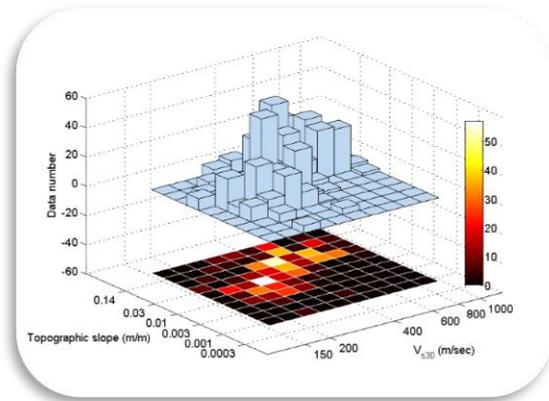
Conclusion

Dans le présent chapitre, nous avons essayé principalement de nous focaliser sur l'adaptation des modèles à effets aléatoires (largement utilisés dans la démarche classique) à l'approche par réseaux de neurones (modèle fixe). Les travaux qui ont été entrepris dans [Derras et al. \(2014\)](#) et la comparaison entre les 5 GMPEs reflète l'importance des résultats obtenus par ce type de méthodes complètement conduites par les données. Ces Résultats encourageants

trouvés dans l'article [Derras et al. \(2014\)](#) et les deux insuffisances recensées ci-dessus nous à motiver à continuer dans la même démarche en essayant d'améliorer davantage nos modèles neuronaux et essayant de réduire la variabilité aléatoire intra-événement en testant la pertinence de plusieurs proxies de site. Ces points vont être entrepris dans le chapitre suivant en utilisant les deux jeux de données RESORCE et NGA-West 2.

Chapitre III

Quel proxy pour décrire les conditions de site et prédire les mouvements sismiques



Introduction

Un modèle de prédiction du mouvement sismique pour la région euro-méditerranéenne a été élaboré dans le chapitre précédent. Ce modèle est totalement conduit par les données, il peut être utilisé pour des études de l'aléa sismique ou pour l'amélioration des codes parasismiques. Dans le présent chapitre, nous allons profiter des avantages offerts par ce type de modèle pour, 1) quantifier la pertinence des différents proxies de conditions de site dans le but de réduire la variabilité aléatoire. Les proxies concernés par cette étude sont : V_{s30} mesurée, V_{s30} estimée, les classes de V_{s30} mesurées, les classes de V_{s30} estimées, la pente topographique « slope ». Toujours dans cette première partie, nous comparons l'effet de la base de données, du type de proxy et de la nature de proxy sur la médiane des prédictions et sur leur variabilité aléatoire. 2) à l'issue des résultats de la première partie, nous élaborons une GMPE donnant la possibilité de choisir la nature de V_{s30} suivant le contexte : mesurée ou estimée.

La première partie de ce travail est réalisée avec deux jeux de données de RESORCE et de NGA-West 2. Dans la deuxième partie, seul le jeu NGA-West 2 est utilisé. Dans les deux parties nous utilisons le PGA et PGV et les ordonnées spectrales pour $T = [0 \text{ à } 4] \text{ s}$ et pour un coefficient d'amortissement égal à 5 %. Ces paramètres représentent pour cette étude les indicateurs du mouvement sismique (MS). Les métadonnées caractérisant l'effet de source et l'effet de propagation d'onde dans nos modèles sont respectivement M_w et R_{JB} . La validation de notre modèle est effectuée en le comparant à celui de [Boore et al. 2014](#).

III.1. Pente topographique « slope »

La simplicité avec laquelle est obtenue la pente topographique (Lemoine et al. 2012) et sa disponibilité, ont incité Wald and Allen (2007) puis Allen and Wald (2009) à utiliser ce paramètre comme proxy décrivant les conditions locales de site. Le but est de trouver une corrélation entre la pente topographique et V_{s30} offrant ainsi la possibilité d'obtenir la valeur et la classe de V_{s30} à partir de la valeur de la pente topographique, un exemple est donné sur la Figure.III.3 du présent chapitre (Derras et al. 2016). Le projet de Wald and Allen soutenu par l'United States Geological Survey (USGS) a pour but d'établir une carte globale donnant ainsi une indication sur V_{s30} via la pente topographique. Le site officiel du « Global V_{s30} Map Server » est : <https://earthquake.usgs.gov/data/vs30/>. Un exemple est donné dans la figure. III.1.

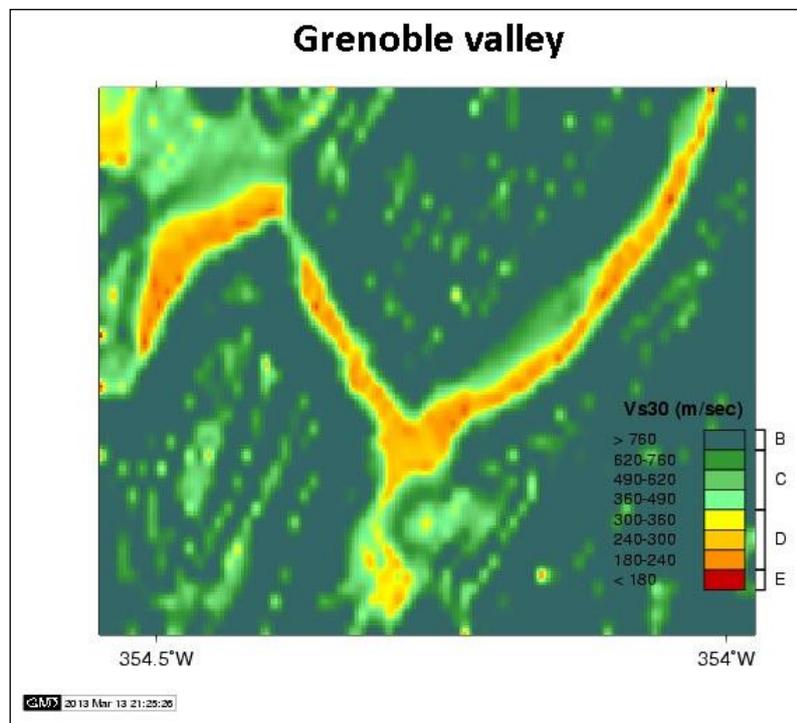


Figure.III.1 Carte isovitesse de la vallée de Grenoble donnant la classe NEHRP de V_{s30} à partir de la pente topographique (slope).

Est-ce que la pente topographique est un proxy fiable de site ? Pour répondre à cette question, nous allons le tester pour les GMPEs avec les MS déjà évoquées ci-dessus. Notons que, le présent travail est innovant du fait de l'utilisation de la pente topographique comme un proxy des conditions du site dans l'équation de prédiction du mouvement du sol. La suite de ce chapitre est un papier publié en 2016 dans la revue « Earthquake Spectra ».

Site-Condition Proxies, Ground Motion Variability, and Data-Driven GMPEs: Insights from the NGA-West2 and RESORCE Data Sets

Boumédiène Derras,^{a),b)} S.MEERI, Pierre-Yves Bard,^{c)} and Fabrice Cotton^{d),e)}

We compare the ability of various site-condition proxies (SCPs) to reduce the aleatory variability of ground motion prediction equations (GMPEs). Three SCPs (measured V_{S30} , inferred V_{S30} , local topographic slope) and two accelerometric databases (RESORCE and NGA-West2) are considered. An artificial neural network (ANN) approach including a random-effect procedure is used to derive GMPEs setting the relationship between peak ground acceleration (PGA), peak ground velocity (PGV), pseudo-spectral acceleration [$PSA(T)$], and explanatory variables (M_w , R_{JB} , and V_{S30} or $Slope$). The analysis is performed using both discrete site classes and continuous proxy values. All “non-measured” SCPs exhibit a rather poor performance in reducing aleatory variability, compared to the better performance of measured V_{S30} . A new, fully data-driven GMPE based on the NGA-West2 is then derived, with an aleatory variability value depending on the quality of the SCP. It proves very consistent with previous GMPEs built on the same data set. Measuring V_{S30} allows for benefit from an aleatory variability reduction up to 15%. [DOI: 10.1193/060215EQS082M]

INTRODUCTION

A vital component of any seismic hazard analysis is a model for predicting the expected distribution of ground motions at a site due to possible earthquake scenarios. This prediction is commonly achieved using equations derived through regression analysis on selected sets of instrumentally recorded strong-motion data, hereafter referred to ground motion prediction equations (GMPEs). The distribution of expected ground motions described by GMPEs is given in terms of median spectral amplitudes and intrinsic variability, the latter usually referred to as aleatory variability and represented by the standard deviation (σ) of the logarithmic residuals.

As near-surface site conditions can strongly affect the characteristics of ground motions, most GMPEs include a site term to capture such effects. Following the initial work of Borcherdt (1994) and colleagues in the early 1990s, the vast majority of strong-motion databases chose to characterize site conditions in terms of the average S-wave velocity V_S down

^{a)} Risk Assessment and Management laboratory (RISAM), University of Tlemcen, Algeria

^{b)} Department of Civil Engineering and Hydraulics, University of Saïda, Algeria

^{c)} Institut des Sciences de la Terre (ISTerre), University of Grenoble Alpes, CNRS, IFSTAR, Grenoble, France

^{d)} Helmholtz Centre Potsdam, German Research Center for Geosciences (GFZ), Potsdam, Germany

^{e)} Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

III.2. Introduction

A vital component of any seismic hazard analysis is a model for predicting the expected distribution of ground-motions at a site due to possible earthquake scenarios. This prediction is commonly achieved using equations derived through regression analysis on selected sets of instrumentally recorded strong-motion data, hereafter referred to ground-motion prediction equations (GMPEs). The distribution of expected ground-motions described by GMPEs is given in terms of median spectral amplitudes and intrinsic variability, the latter usually referred to as aleatory variability and represented by the standard deviation (σ) of the logarithmic residuals.

As near-surface site conditions can strongly affect the characteristics of ground-motions, most GMPEs include a site term to capture such effects. Following the initial work of [Borcherdt \(1994\)](#) and colleagues in the early 1990s, the vast majority of strong-motion databases chose to characterize site conditions in terms of the average S-wave velocity V_s down to a depth of 30m. In spite of the well-known limitations of V_{s30} as a proxy to predict local site amplification ([Castellaro et al. 2008](#); [Kokusho and Sato 2008](#); [Lee and Trifunac 2010](#); [Cadet et al. 2012](#)), it has become a standard for GMPEs development and seismic hazard assessments. Traditionally, V_{s30} is determined either by "invasive" techniques based on seismic measurements in boreholes (downhole, crosshole, or suspension logging), or by "non-invasive" techniques based on surface wave dispersion curves (SASW, MASW, microtremor array). However, such quantitative information about surface geology and V_s is still missing from many strong-motion recording sites. It is also not unusual for V_s measurements to extend to limited depths, significantly shallower than 30 m. For this reason, correlations have been developed to relate V_{s30} to surface geologic units (e.g., [Holzer et al. 2005](#)), geotechnical site categories (e.g. [Ancheta et al. 2014](#)) or *Slope* (e.g. [Wald and Allen 2007](#); [Allen and Wald 2009](#); [Lemoine et al. 2012](#)). V_{s30} values used to characterize strong-motion stations and to develop GMPEs are not systematically measured, but often simply inferred using correlations between the proxy V_{s30} and other proxies, which inherently bring some additional uncertainties.

Several studies have suggested that the ground-motions variability in the resulting GMPEs may depend on the choice of site-condition proxy (e.g. [Chiou and Young 2008](#)). Such a dependence is important as this random variability strongly influences the seismic hazard estimates, in particular for long return periods ([Bommer and Abrahamson 2006](#)). As many countries are deploying new accelerometric networks, the analysis of the impact of the site characterization on the estimated ground-motions variability is also essential to evaluate the

added-value of accurate (and more expensive) site characterization measurements compared to the use of low-cost, much less precise correlation techniques. Such comprehensive understanding of the relative efficiency of site-condition proxies (SCPs) is important to define and prioritize site characterization techniques and programs for accelerometric networks.

The aim of this paper is to investigate the impact of various SCPs choices on the resulting ground-motion aleatory variability estimation. The SCPs used here are measured V_{s30} , inferred V_{s30} and *Slope*. A number of authors have proposed other SCPs (including the site fundamental frequency f_o , the depth to a given velocity threshold, the depth of the deepest important velocity contrast): it would definitely be useful to investigate their respective impact on GMPEs (e.g. [Campbell, 1989](#), [Choi and Stewart 2005](#), [Cadet et al. 2012](#) and [Derras et al. 2012](#)), but unfortunately they are not available for the strong motion databases considered here. The analysis is performed both using discrete site classes (A/B, C and D in NEHRP) and continuous proxy values (measured/inferred $\log_{10}(V_{s30})$, and $\log_{10}(\textit{Slope})$).

Two databases are used, namely RESORCE and NGA-West 2. For both of them, only recordings from shallow crustal events are considered, and available SCPs are the local *Slope* (NGA-West 2 and RESORCE), actually measured V_{s30} values for RESORCE ([Akkar et al. 2014](#)), and either measured or inferred V_{s30} values for NGA-West 2 ([Ancheta et al. 2014](#)). Artificial Neural Network (ANN) approach including a random-effect like ([Derras et al. 2014](#)) procedure have been used for the derivation of GMPEs setting the relationship between various ground-motions parameters (*PGA*, *PGV* and 5% damped pseudo-spectral acceleration (*PSA*) from 0.01 s to 4 s) and event/station meta-parameters (moment magnitude M_w , Joyner-Boore distance R_{JB} and the various SCPs considered here). Avoiding the specification of any *a priori* functional form, ANN provide fully data-driven predictive models and allow the testing of the relative impact of independent variables (e.g. SCPs) on seismic ground-motions.

After a presentation of the datasets and the selection criteria (site proxies) used in the present study, a section is dedicated to the methodology used to obtain ground-motions models using an artificial neural network approach, and its specific implementation in the present application. The next section concentrates on the use of various ANN models with and without various SCPs, in order to answer the central question: how do measured or inferred SCPs reduce ground-motion aleatory variability ?. The last sections then present the development of a simple, fully data driven, NGA-West 2 GMPE, where users are offered

the choice between measured or estimated V_{s30} values, with the price of a larger within-event aleatory variability in the latter case. The results of this simple, three parameter GMPE are compared to a more complex NGA-West 2 model ([Boore et al. 2014](#)) in terms of the median predictions and the within- and between-events variabilities, and their M_w , R_{JB} and site dependencies.

III.3. Ground-motion datasets

III.3.1. RESORCE dataset

The RESORCE strong motion European database is fully described in [Akkar et al. \(2014\)](#). The site conditions include both V_{s30} values, and the *Slope*, derived from the Digital Elevation Model on the basis of the 30 arc s *Slope*. The present study considers only a subset of the original RESORCE database. All recordings for which M_w is unavailable, and recordings from events with a depth larger than 30 km have been eliminated, to focus on shallow crustal earthquakes ([Derras et al. 2014](#)). Only sites with measured V_{s30} values have been considered in this selection. Such selection criteria result in a total of 1028 recordings, with a M_w range [3.6–7.6] and a Joyner & Boore distance (R_{JB}) range [0.01 – 547] km. The (measured) V_{s30} values are in the range [92 - 1600] m/s, while the corresponding *Slopes* are in the range [$5 \cdot 10^{-4}$ - 0.32] m/m.

III.3.2. NGA-West 2 dataset

The NGA-West 2 project is a large multidisciplinary, research program on the Next Generation Attenuation (NGA) models for shallow crustal earthquakes in active tectonic regions. The flatfiles established by the project ([Ancheta et al. 2014](#)) contain site and source information, along with R_{JB} parameters and the corresponding GM intensity measures (IMs). [Ancheta et al. \(2014\)](#) detail the protocol used to obtain V_{s30} within the project. Only a subset of this database was used. In a first step we applied the following exclusion process ([Abrahamson et al. 2014](#)):

- Remove earthquakes not representative of shallow crustal tectonics,
- Remove recordings at R_{JB} greater than an a priori censoring distance,
- Remove recordings not representative of free-field ground-motions,
- Remove earthquakes with questionable hypocentral depths (depth ≤ 1 km),
- Remove the Wenchuan aftershocks,
- Remove recordings missing key metadata,

- Remove recordings identified as questionable (, apparent incorrect gain or spectral shape)
- Remove earthquakes with fewer than three recordings for $M_w > 5$ and earthquakes with less than ten recordings with good coverage in distance for earthquakes with $M_w < 5$.

In a second step, events with a local magnitude M_l between 3 and 5 have been added. Stations with codes I.F and S ([Ancheta et al. 2014](#)) have been excluded since they do not fit a strict free-field condition. Finally, only stations with a GMX C1 classification and an available *Slope* information have been considered. The dataset resulting from this selection process consists of 16446 recordings from 318 earthquakes and 3752 sites. The electronic supplement lists the NGA-West 2 earthquakes actually selected. In the third step, this dataset (NGA-West 2 all- V_{s30}) has been divided into two subsets according to the origin of V_{s30} , i.e. measured or inferred. These two subsets are respectively named NGA-West 2 Measured V_{s30} and NGA-West 2 inferred V_{s30} (Table.III.1). Table.III.1 also provides the number of earthquakes, records, and sites, together with the range of M_w , R_{JB} , V_{s30} and *Slope* for each subset.

III.4. Data distribution

The distribution of the datasets according to M_w , R_{JB} , and site-condition proxy (SCP) is displayed on Figures.III. 2, 3 and 4. Figure.III.2 illustrates clearly the much smaller number of recordings in the European RESORCE dataset, especially for short distances. Figure.III.3 shows the cumulative distribution function (CDF) versus R_{JB} , M_w , V_{s30} and *Slope*. The distributions with respect to V_{s30} and *Slope* are rather similar for the four datasets; one may notice however, a proportionally larger number of low V_{s30} sites for RESORCE, unexpectedly associated with a proportionally smaller number of sites with very small *Slope*, while the distributions of R_{JB} and M_w are very different, especially between RESORCE and NGA-West 2 subsets. For the three NGA-West 2 subsets, there exist however a significant difference as to the distribution with M_w : the NGA-West 2 Inferred- V_{s30} subset is dominated by small to moderate M_w events (up to 70% recordings with M_w smaller than 5), while the NGA-West 2 Measured- V_{s30} contains around 50% recordings with $M_w \geq 5$, and 40% above $M_w=6$. As will be discussed later, this distribution probably has a significant impact on resulting aleatory variability values. Figure.III.4 displays the distribution of the four data sets in the (*Slope*, V_{s30}) space, together with NEHRP site classes defined by V_{s30} intervals (m/s) and the corresponding *Slope* intervals (in m/m) as proposed by [Allen and Wald \(2009\)](#) for active tectonic areas. From this figure, we note that there are

no recordings corresponding to very soft sites (class E), thus the models developed in this study are not valid for Class E sites.

Table.III.1 Range of magnitude, distance and site parameters for the various NGA-West 2 and RESORCE subsets considered in this study

Dataset	No. Of Earthquake	Metadata parameters	Parameters range		Total No. Of recordings	Total No. Of site
			Min	Max		
NGA-West 2 all V_{s30}	318	M_w	3.05	7.9	16446	3752
		R_{JB} (km)	0.001	1533		
		V_{s30} (m/s)	89	2100		
		$Slope$ (m/m)	0.00001	0.69		
NGA-West 2 Measured V_{s30}	287	M_w	3.05	7.9	7938	1916
		R_{JB} (km)	0.001	686.84		
		V_{s30} (m/s)	89	2100		
		$Slope$ (m/m)	0.00001	0.63		
NGA-West 2 Inferred V_{s30}	303	M_w	3.05	7.9	8508	1836
		R_{JB} (km)	0.001	1533		
		V_{s30} (m/s)	196	774		
		$Slope$ (m/m)	0.00001	0.69		
RESORCE Measured V_{s30}	214	M_w	3.6	7.6	1028	228
		R_{JB} (km)	0.01	547		
		V_{s30} (m/s)	92	1598		
		$Slope$ (m/m)	0.00054	0.32		

III.4.1. Site-condition proxies

Ground-motions have to be predicted for target sites which may be described differently from a site-condition point of view. Five cases are possible:

- The site information is not available. This case is named (“without proxies”),
- The site information is described by a NEHRP site class based on an available V_{s30} measured value. This case is named Class (V_{s30}),
- The site information is described by a NEHRP site class based on the correspondence between $Slope$ and V_{s30} as proposed by [Wald and Allen \(2007\)](#) and [Allen and Wald \(2009\)](#) for active tectonic areas. This case is named (Class ($Slope$)),

- The site is described by its *Slope* value S named (*Slope*). S is the 30 arc s *Slope* ([Ancheta et al. 2014](#)),
- The site is described by a continuous V_{s30} value. This case is named (V_{s30}).

In what follows we develop GMPEs for these five cases, both for the RESORCE and the three NGA-West 2 subsets. We obtain a total of 4×5 GMPEs, which are labelled as ANN(dataset, SCP).

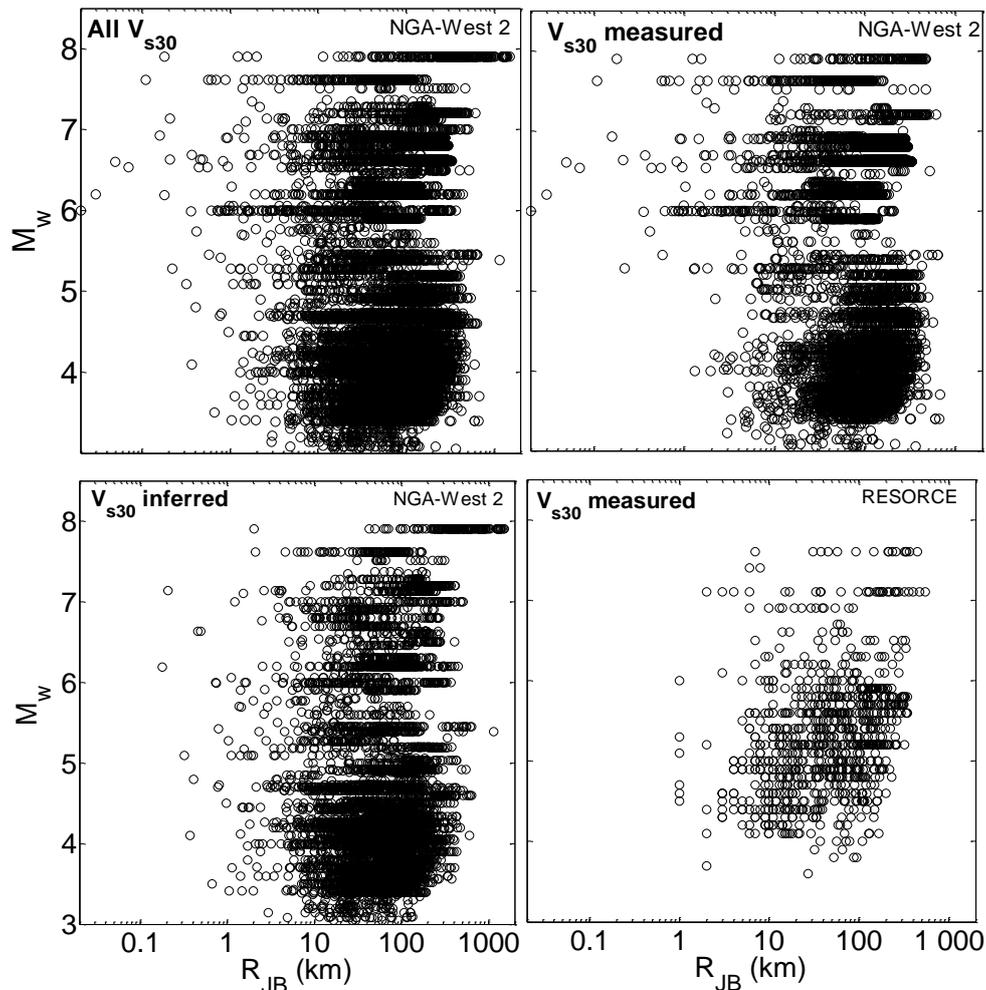


Figure.III.2 Distribution of the four datasets considered in this study in the [Magnitude (M_w) - distance (R_{JB})] plane.

III.5. Methodology

In order to facilitate the comparability with classical GMPEs, we adapted the ANN approach to the random-effects procedure proposed by [Abrahamson and Youngs \(1992\)](#). This resulted in a two-phase building process, starting with a fixed-effects model as step 1, and a random-effect like step 2.

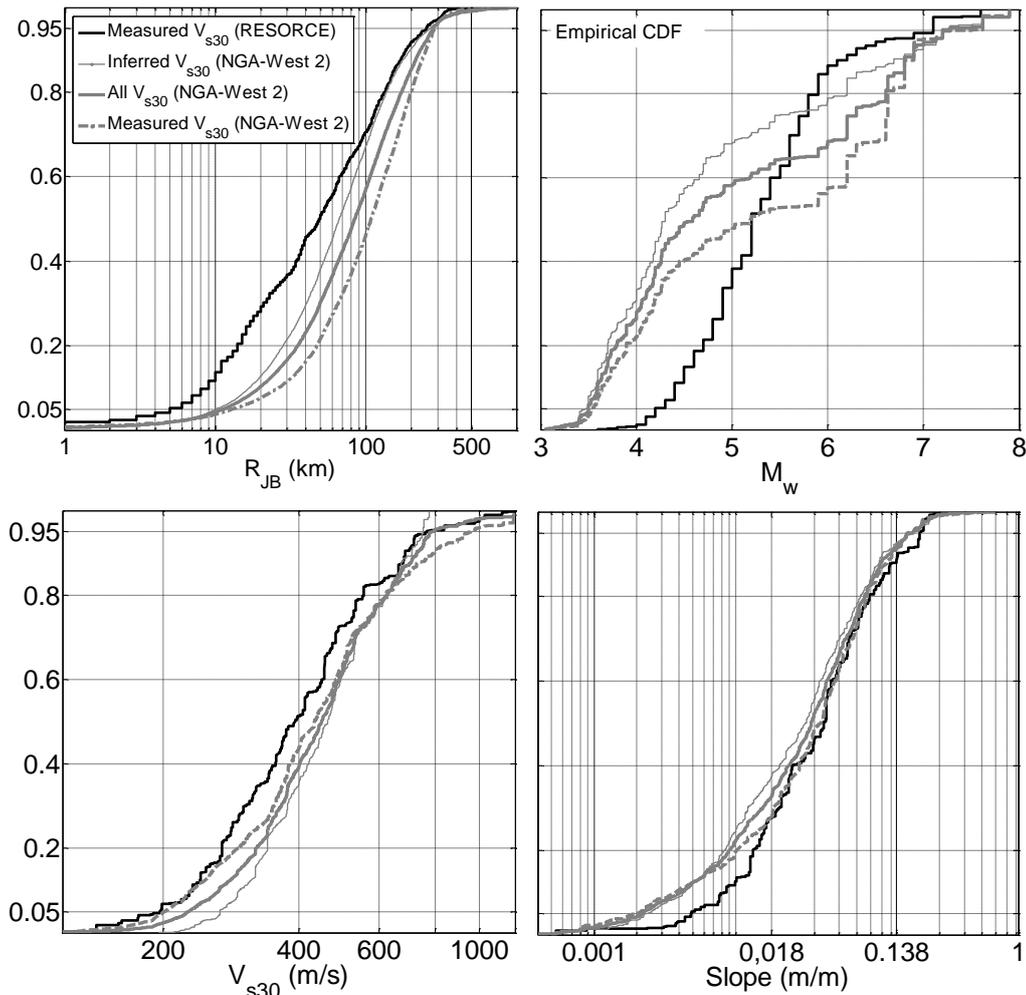


Figure.III.3 Cumulative distribution function of the four considered datasets (NGA-West 2 and RESORCE) with respect to the explanatory variables, i.e. R_{JB} distance (top left), magnitude M_w (top right), V_{s30} (bottom left) and Slope (bottom right).

Step 1: Fixed models

The input parameters are M_w , R_{JB} and either three binary parameters describing the site class, or one continuous parameter describing the SCPs (V_{s30} or Slope). The predicted ground-motions are the geometric mean horizontal component of PGA , PGV , and 5%-damped PSA at 20 periods from 0.01 to 04 s. We do not include equations for Peak Ground Displacement (PGD), which we consider to be too sensitive to the low-pass filters used in the data processing.

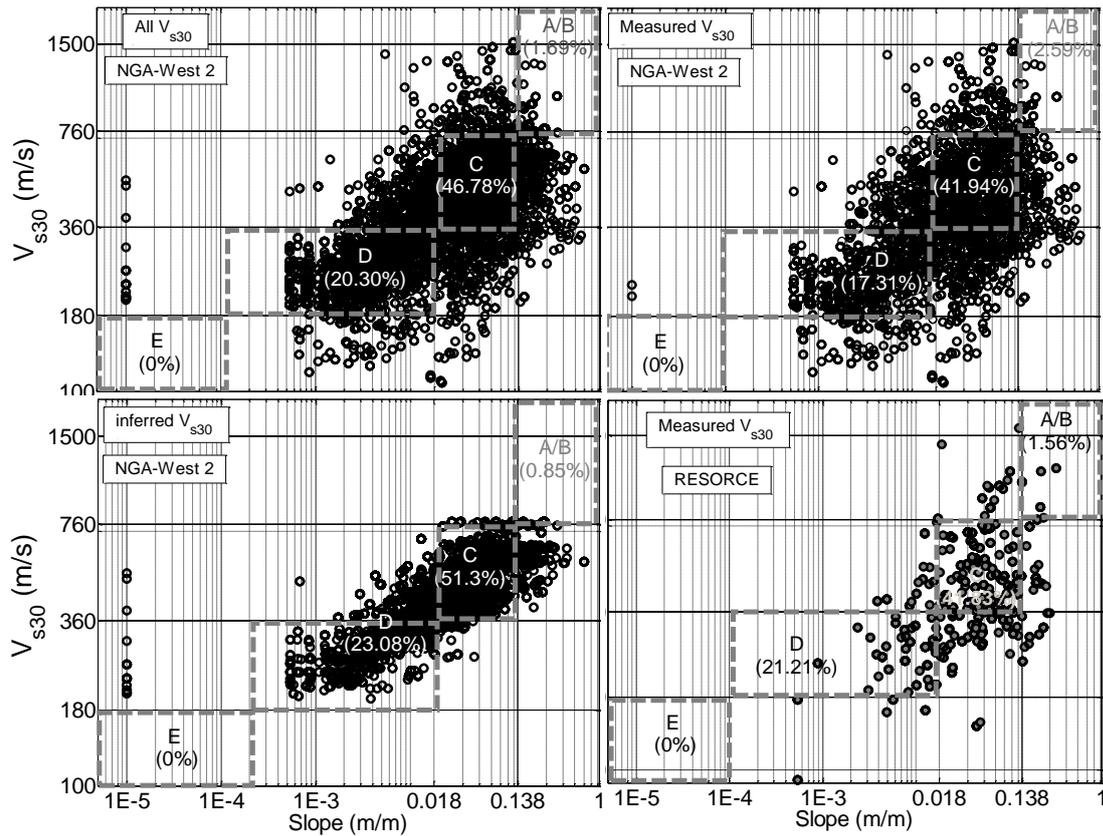


Figure.III.4 Correlations of V_{s30} values (m/s) versus Slope values (m/m). for the four considered datasets boxes represent V_{s30} and Slope ranges recommended by Wald and Allen (2007) and Allen and Wald (2009) for classification into NEHRP categories (roughly equivalent to EC8 classes (Lemoine et al., 2012)) in the case of active tectonic areas

A multilayer perceptron ANN type is used. The structure of the ANN is displayed in Figure.III.5 for both the "SCP class" and "continuous SCP" models. This structure has been taken from Derras et al. (2012; 2014). The input and output layers contain the input and output parameters, and are linked through one single hidden layer consisting of only three neurons. This small number of neurons is the optimal hidden neurons number in order to optimize both the total σ_f and the Akaike Information Criterion (Akaike 1973). The type of activation functions between input and hidden layers, and between hidden and output layers have been adopted after several tests (Derras 2010). It resulted in the choice of a "tangent sigmoid" type for the hidden layer and a "linear" one for the output layer. The quasi-Newton Back Propagation technique has been applied for the training phase (Shanno and Kettler 1970). To avoid "over-fitting" problems we chose an adequate regularization method developed by Derras et al. (2012) together with a limited number of neurons on the hidden layer. In Figure.III.5, the symbols W and b represent the synaptic weights and bias with subscripts representing the corresponding neurons between two layers (Derras 2010).

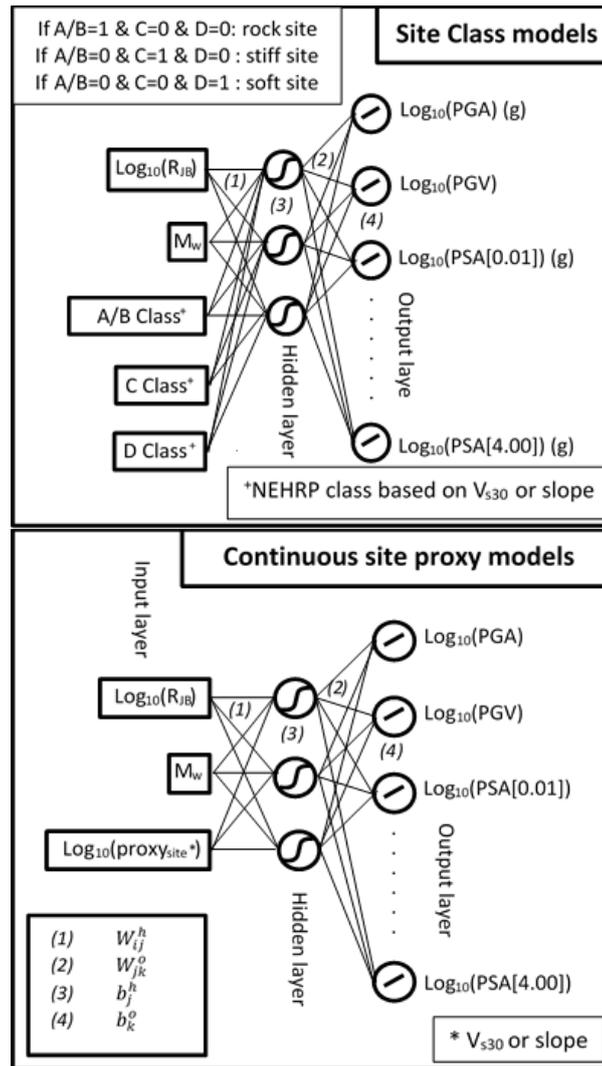


Figure.III.5 Structure of the Neural Networks considered for the prediction of PGA, PGV and PSA[0.01 to 4 s], in the two cases of discrete (top) and continuous (bottom) site proxies. W_{ij}^h is the synaptic weight between the i^{th} neuron of the input layer and the j^{th} neuron in the hidden layer, b_j^h the bias of the j^{th} neuron in the hidden layer. Also the W_{jk}^o is the synaptic weight between the j^{th} neuron of the hidden layer and the k^{th} neuron in the output layer, b_k^o the bias of the k^{th} neuron in the output layer

Twenty different ANN models were built, differing by the dataset and the SCP used in the input layer. For the four datasets, the earthquakes are described by the M_w and the source to site distance by R_{JB} . A first ANN model was considered with only those two input parameters (M_w and R_{JB}) and no SCP, in order to set the reference for quantifying the gains brought by the consideration of the various site proxies. The proxies describing the site conditions are the site classes derived from the V_{s30} values (NEHRP classes A/B, C and D, i.e., V_{s30} within the range $[\geq 760 \text{ m/s}]$, $[360-760 \text{ m/s}]$ and $[180-360 \text{ m/s}]$), the corresponding *Slope* range as proposed by [Wald and Allen \(2007\)](#). Site class models are displayed in

Figure.III.5-top. Continuous site-proxies models, i.e., $\log_{10}(V_{s30})$ or $\log_{10}(Slope)$, as displayed in Figure.III.5-bottom.

Step 2: Random-effect model

A procedure similar to the random-effects approach was then used to provide the between- and within-event σ , as described in [Derras et al. \(2014\)](#). For each of the 20 cases considered, the final ANN model is obtained using the maximum likelihood approach developed by [Brillinger and Preisler \(1985\)](#) and stabilized by [Abrahamson and Youngs \(1992\)](#). In addition to the median, this approach provides the aleatory uncertainly values. The performance of the ANN scheme is measured by the σ_t value classically used in GMPEs, which is decomposed into the between-event (τ) and within-event (ϕ) variabilities: both are zero-mean, independent, normally distributed random variables with standard deviations τ and ϕ ([Al Atik et al. 2010](#)). The between- and within-events residuals are assumed uncorrelated, so that the total σ_t at a period T of the ground-motions model can be calculated according to equation.III.1.

$$\sigma_t(T) = \sqrt{\tau(T)^2 + \phi(T)^2} \quad (\text{III.1})$$

The dependence of τ and ϕ on the prediction variables M_w, R_{JB} are presented in the following section, and form the bases of the assessment of the performance of the various SCPs.

III.6. How do measured or inferred site-condition proxies reduce ground-motion aleatory variability?

In this section, we compare the various models derived for each data set and analyze how the various SCPs reduce the ground-motion aleatory variability. The variations of τ , ϕ and σ_t versus the period are displayed in Figure.III.6 for the 4 groups (datasets) of the 5 ANN models. The Figure.III.6 shows that the between-event variability, τ , is much lower than the within-event variability, ϕ . Both τ and ϕ exhibit a general trend to increase with period for RESORCE, while a maximum is observed around 0.1 s for all NGA-West 2 data subsets, and remain stable beyond one s. Such a peak has been observed by some NGA GMPEs developers (e.g. [Chiou and Youngs 2014](#)). A possible explanation is the interaction of varying stress drop with the high-frequency damping term (κ). Such a peak is not as clearly observed for the RESORCE models. Another expected observation is the limited impact of SCPs on the τ , which shows a satisfactory separation of τ and ϕ .

Whatever the explanation behind this variable frequency dependence of τ and ϕ , the use of a SCP allows to decrease both components of the variability and, as expected, the impact of the site proxies is larger on the within-event component. However, the various site proxies exhibit a variable efficiency in decreasing the ϕ . The relevance of the SCP can be evaluated by its ability to reduce the residuals between predictions and observations (i.e., the aleatory variability σ_t). For $T > 0.2$ s (Figure.III.6), the ϕ reduction shown by models using site proxies is clear, and several observations on the performance of SCPs can be made: The largest reductions of ϕ are associated with a "measured V_{s30} " SCP. This is especially clear from the comparison of results on the three NGA-West 2 subsets, for which the largest reductions are observed for the NGA-West 2 Measured V_{s30} subset, the intermediate one for NGA-West 2 all V_{s30} and the smallest for NGA-West 2 inferred V_{s30} . Such a significant decrease is observed both for V_{s30} based site class and continuous values of V_{s30} . The smallest ϕ reductions correspond to the *slope* proxy, used either for site classification or as a continuous site parameter. This is observed for both the NGA-West 2 and RESORCE datasets. For both V_{s30} and *Slope* SCPs, continuous values allow a larger decrease of ϕ , compared to site classes. When V_{s30} is not measured but only inferred (ANN(Inferred V_{s30} , V_{s30}) case), we notice that using non-measured V_{s30} does not prove much more efficient than using the *Slope* proxy. This is the case for which the ϕ reductions brought by either SCP are the lowest.

Figure.III.7 compares the period dependence of σ_t for the seven models based on the NGA-West 2 dataset and continuous proxies, with a focus on models including SCPs on Figure.III.7(a), to be compared with the "reference models" without SCPs on Figure.III.7(b).

ANN(Inferred V_{s30} , V_{s30}) is shown to provide the greatest σ_t values among the four models. $\sigma_t(T)_{\text{ANN(Inferred } V_{s30}, V_{s30})} \geq \sigma_t(T)_{\text{ANN(All } V_{s30}, \text{Slope})}$. When considering the NGA-West 2 all V_{s30} , the V_{s30} proxy (which in this case may be either inferred or measured) leads to a smaller σ_t than the *Slope* proxy, especially in the period range [0.1 – 2 s]. The smallest σ_t is obtained for the ANN(Measured V_{s30} , V_{s30}) model, with values 15% to 20% smaller than for the ANN(Inferred V_{s30} , V_{s30}) model over the whole period range. However, Figure.III.7b shows that these large differences are not due only the SCPs: when only M_w and R_{JB} are considered, the σ_t are very significantly different from one NGA-West 2 subset to another at short and intermediate period values, while the long period values (i.e., $T \geq 0.9$ s) are about the same.

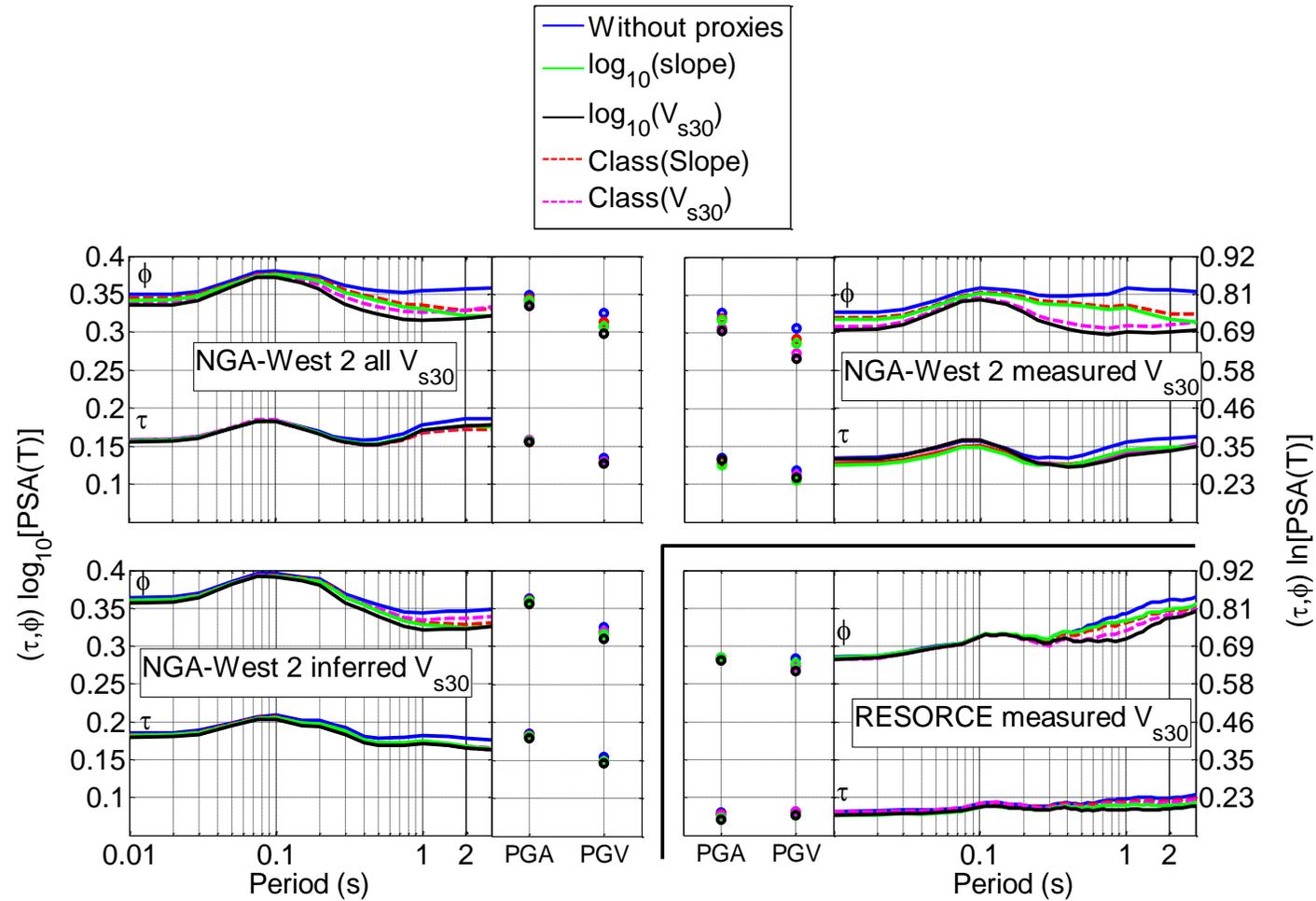


Figure.III.6 Sensitivity of the aleatory variability components (between-event, τ and within-event, ϕ) to the site proxies used as input for the neural network. As in Figures.III 2 and 4, the top left frame corresponds to the "All V_{s30} " NGA-West 2 dataset, the top right to the "measured V_{s30} " NGA-West 2 subset, the bottom left to the "Inferred V_{s30} " subset, and bottom right to the RESORCE dataset. The aleatory variability values are provided both on \log_{10} scale (left axis) and on \ln scale (right axis). They are provided for spectral ordinates for periods from 0.01 to 4 s, and for PGA and PGV as well.

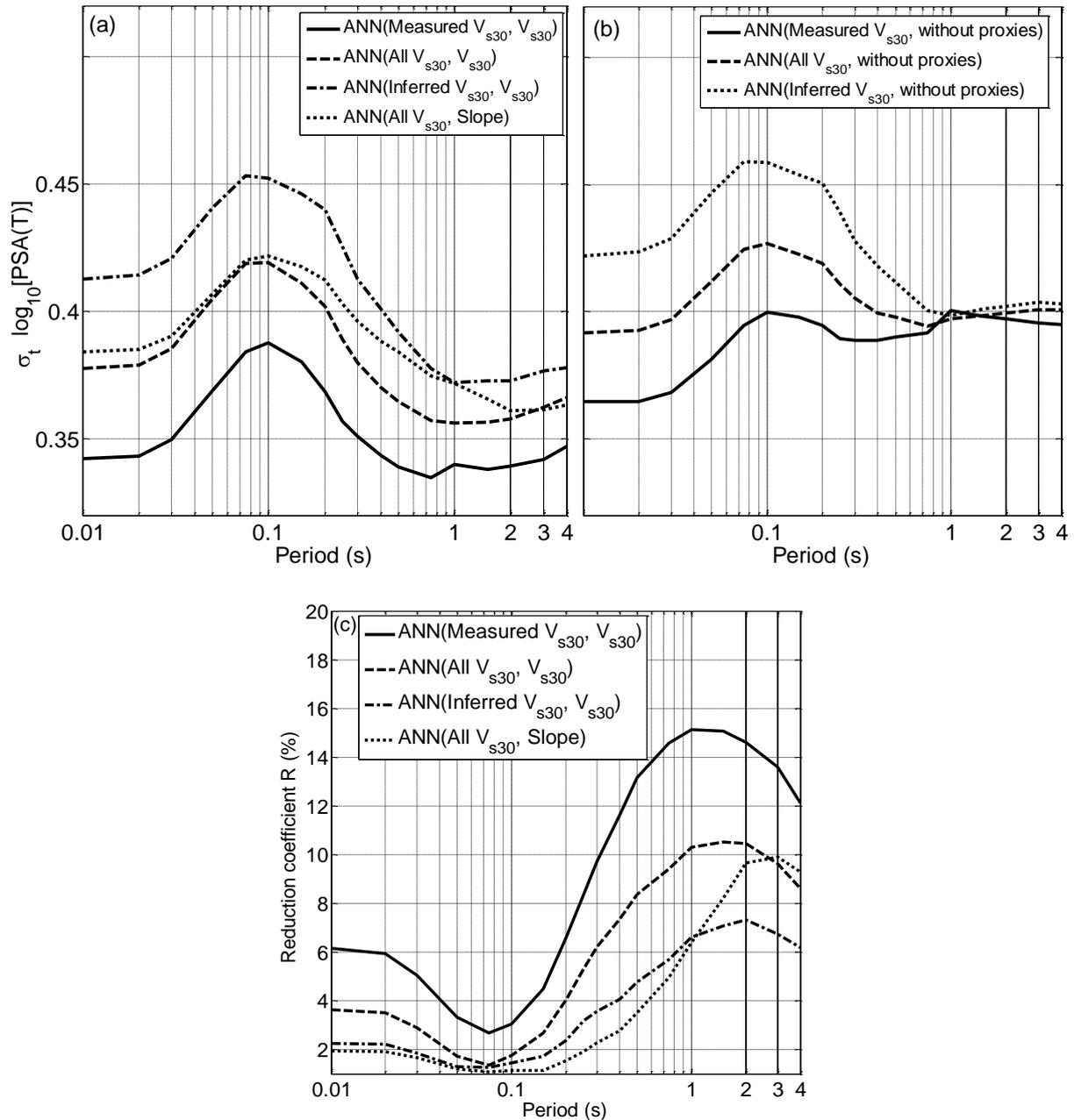


Figure.III.7 Quantifying the performance of the various continuous SCPs for the NGA-West 2 case.

The two top frames display the total aleatory variability σ_t for (a) the four ANN models with continuous SCPs V_{s30} (All, Measured, Inferred) or Slope (All) and (b) the three "reference" ANN models without SCPs (All, Measured, Inferred). The bottom frame displays the reduction of aleatory variability (R , see equation.III.2) for the 4 ANN models displayed in (a) compared with their respective "reference models" (b)

The origin of such a drastic dependence is not clear and probably includes several components:

- Different proportions of small to moderate M_w events (as shown in Figure.III.3), with more moderate events for the "inferred- V_{s30} " case, and a smaller average R_{JB} : uncertainties in source parameters (M_w and location, especially depth) are likely to be larger, and this may map in a larger variability.
- S-wave velocity measurements are very likely to have been prioritized for sites with large, well-recorded events for which uncertainties in source and R_{JB} parameters are

certainly the lowest. Therefore, the best way to measure the efficiency of a given SCP is not to look at the absolute levels of σ_t or ϕ , but at the variance reduction coefficient R as defined in Equation.III.2.

$$R(T) = \left(1 - \frac{\sigma_t(T)_{ANN(\text{dataset}, \text{site proxy})}}{\sigma_t(T)_{ANN(\text{dataset}, \text{without proxies})}} \right) \times 100; (\%); \quad (\text{III.2})$$

Where "dataset" stands for either "All V_{s30} ", "Measured V_{s30} " or "Inferred V_{s30} " dataset. The corresponding results are displayed in Figure.III.7c. They confirm that largest reductions correspond to Measured V_{s30} ; this reduction exhibits a clear dependence on period, being minimum around 3% for $T = 0.06 - 0.10$ s, and maximum around 15 % in the range $T = 1.0 - 2.0$ s. The smallest reductions correspond to the cases ANN(All V_{s30} , $Slope$) and ANN(Inferred V_{s30} , V_{s30}): they reduce to only 1% around $0.05 - 0.1$ s, and hardly reach 10% and 7%, respectively, beyond 2 s. The reduction values for the ANN(All V_{s30} , V_{s30}) case are intermediate, because they mix large reductions for sites where V_{s30} has been measured, and low reductions for sites where it has been only "inferred".

As complementary information to the effects of the input parameters on σ_t (Figure.III. 7a and 7b), Table.III.2 displays the total percentages of synaptic weights P corresponding to each of the three input parameters (M_w , R_{JB} and SCP). These P parameters are a measure of the relative importance of the individual predictor variables. They have been computed according to the procedure detailed in Derras et al. (2014, Equation.III.4), for five ANN models (the four in Figure.III.7a, and the RESORCE- V_{s30} one), where i corresponds to the input parameters. They constitute another indicator of the relative importance of each input parameter. The most efficient parameter in reducing the variance of response spectra is the R_{JB} (around 50% synaptic weight), followed directly by the earthquake M_w , M_w (around 35-40%). The P associated to the site term range from 7% to 14%: The largest SCP weights correspond to measured V_{s30} [ANN(Measured V_{s30} , V_{s30}) and ANN(RESORCE, V_{s30})] ($P_{V_{s30}} \cong 14$ %) and the smallest correspond to Inferred V_{s30} ($P_{V_{s30}} = 7$ %). In the others cases: $P_{V_{s30}ANN(\text{All } V_{s30}, V_{s30})} = 12\%$ is larger than $P_{SlopeANN(\text{All } V_{s30}, Slope)} = 11\%$. This ranking is similar to the ranking obtained from the analysis of aleatory variabilities discussed above.

Table.III.2 Sensitivity of ground motion models to magnitude, distance and site parameters, as expressed through the total percentages of synaptic weight (P , %) corresponding to each input parameter.

ANN models	P (%)		
	R_{JB}	M_w	SCP
ANN(Inferred V_{s30} , V_{s30})	55.08	38.41	6.51
ANN(All V_{s30} , <i>Slope</i>)	51.93	36.76	11.31
ANN(All V_{s30} , V_{s30})	51.73	35.82	12.45
ANN(Measured V_{s30} , V_{s30})	49.7	36.38	13.92
ANN(RESORCE V_{s30} , V_{s30})	46.64	39.72	13.64

III.7. Impact of the dataset, site proxy, and quality of V_{s30} on median ground-motion models.

As discussed above, the nature of SCP has a noticeable effect on the ground motion aleatory variability. It is therefore needed to investigate whether it modifies, or not, the median. Figure.III.8 compares the median ground-motions predicted for $M_w = 6$ on stiff soil by various ANN models considering various SCPs. The median predictions are shown to depend only weakly on the chosen dataset (RESORCE/NGA-West 2), while the type of SCP type (*Slope*/ V_{s30}) and the quality of the V_{s30} estimate (Inferred/Measured) have no influence on this median. These results indicate that the type of SCPs does not affect the median, which is fully consistent with the significantly smaller P_i associated to any SCP, compared to the M_w and R_{JB} .

III.8. Application: development of a fully data driven gmpe based on nga-west 2

Most of previous conclusions on the impact of SCPs are based on the application of the ANN approach to the NGA-West 2 dataset. Part of the engineering community is still not convinced by fully data driven, non-parametric approaches for deriving GMPEs. This section is then dedicated to the comparison of our final ANN model derived on the full NGA-West 2 dataset, with one of the NGA-West 2 GMPE classically derived using a priori functional forms. Amongst all the NGA-West 2 GMPEs, we chose BSSA14 ([Boore et al. 2014](#)) for two reasons: 1) it is simple (a minimum of metadata parameters is used in comparison with the others NGA-West 2 GMPEs); 2) In general the median ground-motions of all NGA-West 2 GMPEs are similar to one another, except around $M_w=5.5$ where the BSSA14 functional form exhibits a specific singularity ([Gregor et al. 2014](#)); therefore, a single model is sufficient for such comparison. Moreover, the BSSA14 model uses the closest R_{JB} to the horizontal projection of the rupture plane, R_{JB} , which is the same R_{JB} used in our model ANN(All V_{s30} , V_{s30}). After a

presentation of the ANN functional forms resulting from the chosen ANN architecture, the comparison with the BSSA14 model will focus on the aleatory variability and on the M_w , R_{JB} and V_{s30} scaling of median predictions.

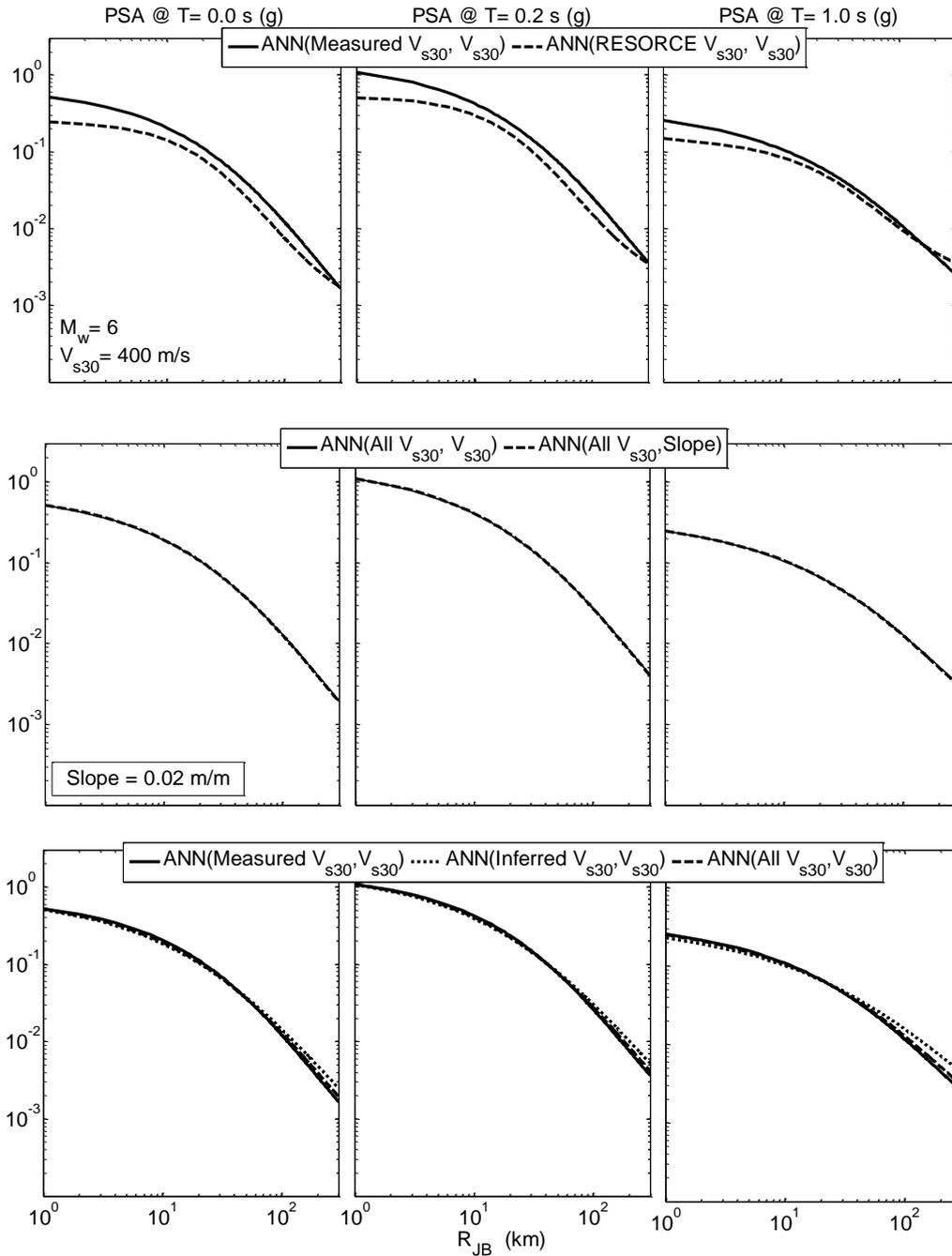


Figure.III.8 Robustness of the median ground motion predicted by different models. Effect of the dataset (top, NGA-West 2 RESORCE), the SCP proxy (middle, V_{s30} / Slope) and the SCP quality (bottom, measured / inferred / mix V_{s30}) measurement. The plots display the distance dependence of the spectral accelerations at three periods ($T = 0.0$ s, left column; $T = 0.2$ s, middle column; and $T = 1.0$ s, right column) for a $M_w = 6$ event, at sites with $V_{s30} = 400$ m/s and Slope=0.02 m/m.

III.9. The NGA-West 2 ANN (All V_{s30} , V_{s30}): functional form

The ANN(All V_{s30} , V_{s30}) model developed in this study results in a relatively simple, data driven functional form, providing the median estimate of the geometric mean horizontal component \log_{10} of ground-motions parameters ($PGA(g)$, $PSV(m/s)$ and $PSA(g)$ [0.01–4 s] through the equation.III.3.

$$\log_{10}(\hat{PSA}) = a(T).f_1 + b(T).f_2 + c(T).f_3 - d(T) \quad (III.3)$$

$$\text{Where } \begin{cases} f_1 = th[-0.321.ln(R_{JB}) + 0.134.M_w - 0.119.ln(V_{s30}) + 1.686] \\ f_2 = th[-0.072.ln(R_{JB}) + 0.554.M_w - 0.115.ln(V_{s30}) - 1.387] \\ f_3 = th[-0.014.ln(R_{JB}) + 0.141.M_w - 0.208.ln(V_{s30}) + 0.515] \end{cases}$$

The th dependency is related to the selection of the Tanh-sigmoid activation function for the hidden layer. Where a , b , c and d are listed in Table.III. 3. The coefficients for the median ground-motions and for the σ are listed in an electronic supplement titled “[DBC15 coeffs.xls](#).”

III.10. Range of applicability

The range of applicability of such fully data-driven models is closely dependent on the chosen dataset. Such models should never be used outside the R_{JB} and M_w range the original dataset. We then recommend not using this ANN model outside the following ranges: M_w [3.5–7.3], R_{JB} [3–300 km], V_{s30} [200–800] m/s. We obtained these limits on the basis of the values of the cumulative distribution function (Figure.III.3) at 5% and 95%. As real PSHA studies may need to evaluate ground-motions at distances lower than $R_{JB} = 3$ km, we recommend in such cases simply to use the values predicted for a R_{JB} value of 3 km. This is somewhat artificial, but is based on the same physical basis (i.e., near-field saturation) as the one used in classical GMPEs for the choice of the (flat) functional forms at very short distances.

III.11. Residual analysis

When performing a random-effects regression, the residuals are separated into between-event and within-event terms. The ANN(All V_{s30} , V_{s30}) model is obtained through a training-regularization procedure using the random effect model. In addition to the median, this approach allows the computation of an aleatory uncertainty model. The resulting model is defined by the following equation:

$$\log_{10}(PSA_{ij}) = \log_{10}(\hat{PSA}_{ij}) + \eta_i(M_w) + \varepsilon_{ij}(M_w, R_{JB}, V_{s30}) \quad (III.4)$$

Table.III.3 Coefficients of the final ANN(All V_{s30} , V_{s30}) model (geometric mean of horizontal components, median value, see equation.III.3 in the text)

Parameter	a	B	c	d
PGA	2.269	0.939	0.04	3.231
PGV	1.81	1.308	0.596	1.359
T 0.010	2.271	0.934	0.038	3.225
T 0.020	2.28	0.936	0.028	3.226
T 0.030	2.304	0.926	-0.009	3.213
T 0.050	2.384	0.898	-0.102	3.168
T 0.075	2.457	0.859	-0.154	3.085
T 0.100	2.456	0.843	-0.13	3.005
T 0.150	2.372	0.868	-0.08	2.913
T 0.200	2.271	0.909	0.028	2.881
T 0.250	2.189	0.962	0.117	2.894
T 0.300	2.112	1.022	0.19	2.924
T 0.400	1.983	1.141	0.261	3.014
T 0.500	1.888	1.229	0.346	3.103
T 0.750	1.719	1.411	0.482	3.314
T 1.000	1.619	1.53	0.579	3.497
T 1.500	1.458	1.722	0.711	3.792
T 2.000	1.372	1.835	0.818	4.022
T 3.000	1.326	1.958	0.929	4.377
T 4.000	1.308	2.026	1.027	4.628

Where $\log_{10}(\hat{PSA}_{ij})$ and $\log_{10}(PSA_{ij})$ are respectively, the median predicted (Equation.III.3) and observed values, η_i is the between-event residual for the event i , and ε_{ij} is the within-event residual for recording j of event i . The relationship between the between-event and within-event τ and ϕ is given by (Equation.III.1) under the random effects procedure (step 2). In this model, considering the findings of the previous section, ϕ was computed separately for the sites with measured and inferred V_{s30} , as in (Chiou and Young 2008). In this case ϕ is named ϕ_M or ϕ_I and can be determined by Equation.III.5. The value of τ remains unchanged.

$$\left\{ \begin{array}{l} \phi_M(T) = \sqrt{\frac{1}{L_M} \sum_{j=1}^{L_M} (F_{Measured,j} \cdot \varepsilon_{ij}(V_{s30M}))^2} \\ \phi_I(T) = \sqrt{\frac{1}{L_I} \sum_{j=1}^{L_I} (F_{Inferred,j} \cdot \varepsilon_{ij}(V_{s30I}))^2} \end{array} \right. \quad (III.5)$$

Where $L_{M \text{ or } I}$ are the total number of samples in each subset, with $L_M = 7938$ for measured V_{s30M} and $L_I = 8508$ for Inferred V_{s30I} (Table.III. 1) $F_{measured,j}$ is a binary variable equal to 1 if V_{s30} is measured for site j , and 0 otherwise, and vice-versa for $F_{Inferred,j}$.

The ϕ_M , ϕ_I , τ and σ_τ values are listed in Table.III. 4 for the two subsets with measured or inferred V_{s30} .

Like ϕ , the mean value of residuals, denoted μ and classically used on binned intervals for detecting possible bias or trends, also depends on the way to obtain V_{s30} . In this way μ is named μ_M and μ_I for measured and inferred V_{s30} respectively:

$$\left\{ \begin{array}{l} \mu_M(T) = \frac{1}{L_M} \sum_{j=1}^{L_M} (F_{Measured,j} \cdot \varepsilon_{ij}(V_{s30M})) \\ \mu_I(T) = \frac{1}{L_I} \sum_{j=1}^{L_I} (F_{Inferred,j} \cdot \varepsilon_{ij}(V_{s30I})) \end{array} \right. \quad (III.6)$$

In Figure.III.9 the ϕ_M and ϕ_I versus V_{s30} are presented for periods $PSA(0)$, $PSA(0.2)$, $PSA(1)$ and $PSA(2)$. The means of residuals and their corresponding σ in 10 V_{s30} bins are also indicated in order to identify possible bias. There is no obvious trend of the mean within-event residuals with V_{s30} type (measured/inferred), for any ground-motions parameters. For all cases μ_M and μ_I remains lower than 0.02 while for all periods and for most V_{s30} bins $\phi_M < \phi_I$.

III.12. Dependence of aleatory variabilities with magnitude and distance

Most NGA-West 2 GMPEs involve an increased level of complexity as to the dependence of the between- and within-event variabilities with M_w , R_{JB} and even V_{s30} . We thus investigated whether the residuals of our ANN based model exhibit such a dependence. Our approach was to bin the between-event event residuals η_i within various M_w intervals to investigate their M_w dependence, and to bin the within-event residuals ε_{ij} by R_{JB} and V_{s30} to investigate the corresponding dependence.

Table.III.4 Values of the between-event, within-event and total variabilities for the final ANN(All V_{s30} , V_{s30}) model; the within-event variability depends on whether the V_{s30} value is measured or inferred

Parameter	τ	Measured V_{s30}		Inferred V_{s30}	
		ϕ_M	σ_t	ϕ_I	σ_t
PGA	0.156	0.302	0.348	0.357	0.397
PGV	0.127	0.265	0.307	0.319	0.355
T 0.010	0.156	0.303	0.349	0.358	0.398
T 0.020	0.157	0.303	0.349	0.359	0.399
T 0.030	0.161	0.310	0.357	0.364	0.405
T 0.050	0.173	0.328	0.377	0.381	0.424
T 0.075	0.183	0.342	0.393	0.392	0.437
T 0.100	0.183	0.345	0.395	0.390	0.435
T 0.150	0.173	0.336	0.386	0.384	0.429
T 0.200	0.166	0.324	0.373	0.380	0.423
T 0.250	0.159	0.313	0.361	0.368	0.409
T 0.300	0.155	0.308	0.354	0.358	0.398
T 0.400	0.151	0.300	0.345	0.347	0.386
T 0.500	0.152	0.297	0.341	0.341	0.380
T 0.750	0.159	0.293	0.336	0.330	0.369
T 1.000	0.170	0.294	0.338	0.325	0.364
T 1.500	0.174	0.296	0.338	0.324	0.364
T 2.000	0.176	0.298	0.341	0.326	0.364
T 3.000	0.178	0.303	0.346	0.329	0.369
T 4.000	0.178	0.308	0.351	0.330	0.371

Figure.III.10 shows example plots of binned values of τ and ϕ against the respective predictive variables for two periods (0.2 and 1.0 s, respectively). We compare them with the BSSA14 model. Terms τ and ϕ are computed with the average of the binned values shown in Figure.III.10. The principal findings of this analysis are as follows: - As shown in Figure.III.10 (top), τ decreases with M_w , but relatively slightly (by about 20%), and becomes nearly constant for $M_w > 5.5$, which is the M_w range of engineering interest.

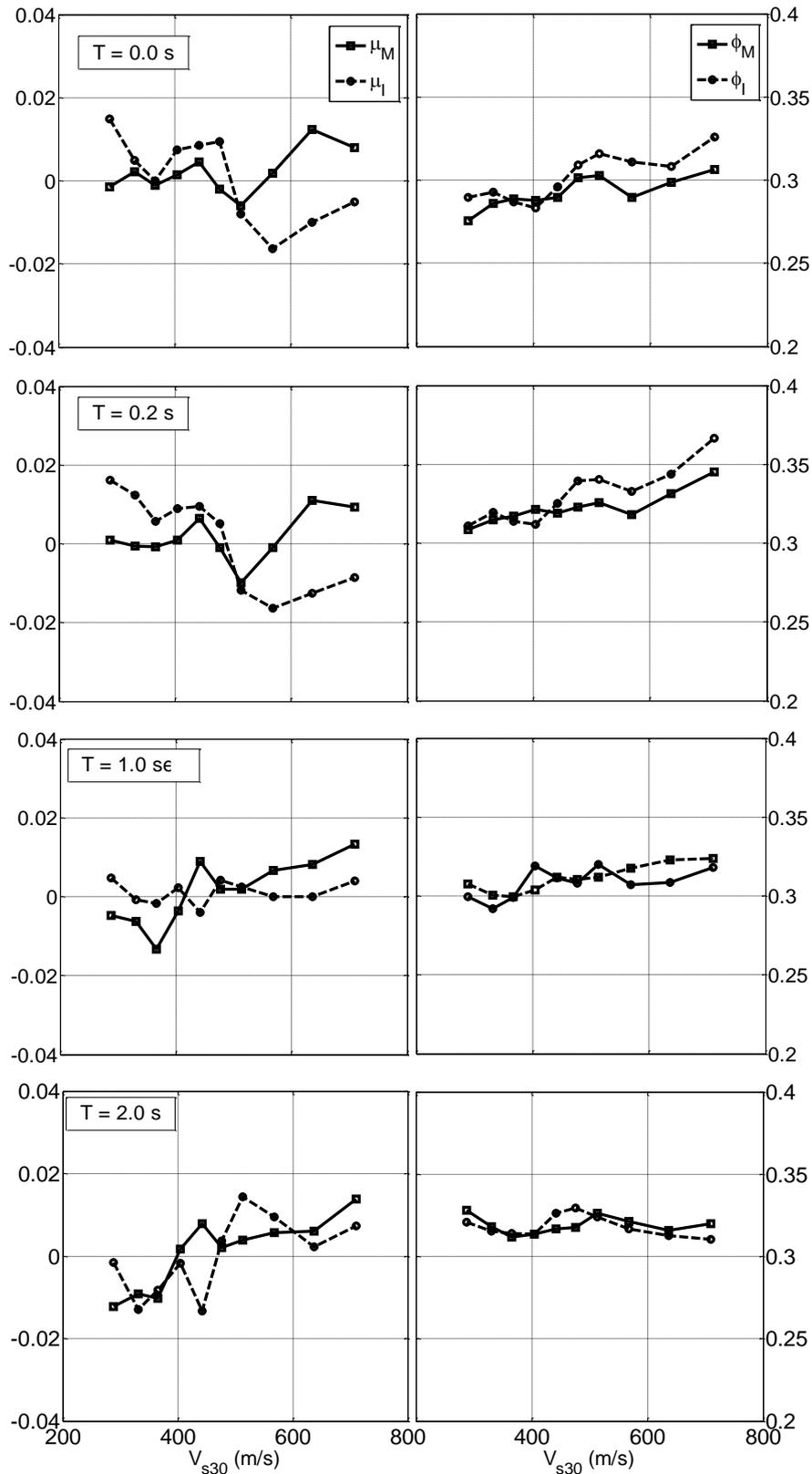


Figure.III.9 Comparison of the mean binned residuals μ_M and μ_I (left), and the corresponding within-event standard deviations ϕ_M and ϕ_I (right) for the two "Measured V_{s30} " and "Inferred V_{s30} " NGA-West 2 subsets. The bins correspond to ten equal sample-size subsets with increasing V_{s30} values. The four rows correspond to spectral accelerations for four different periods, $T=0$ s, 0.2, 1.0 and 2.0 s, from top to bottom

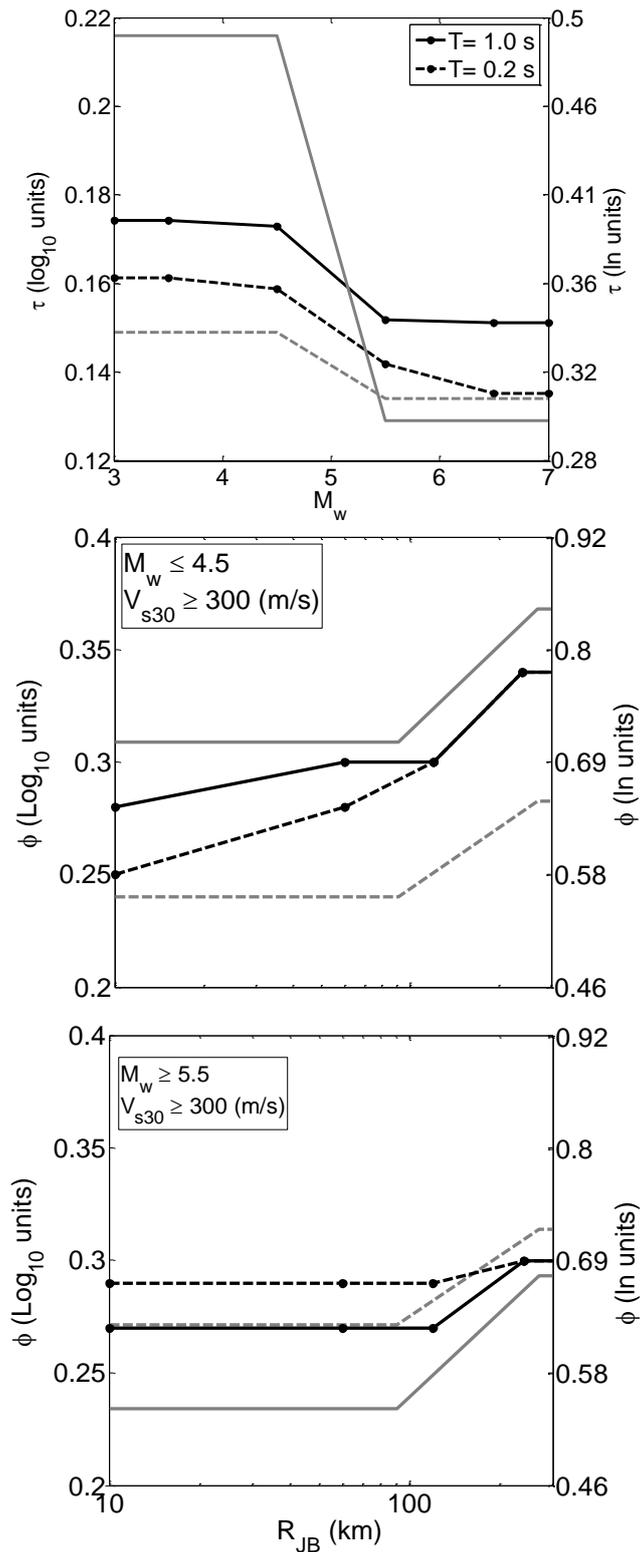


Figure.III.10 Dependence of the between-event variability with magnitude bins (top) and of the within-event variability with distance bins (middle for moderate magnitude, stiff sites: $M_w \leq 4$, $V_{s30} \geq 300$ m/s and bottom for larger magnitudes and stiff sites: $M_w \geq 5.5$, $V_{s30} \geq 300$ m/s). The results are displayed for spectral accelerations at two periods ($T=0.2$ s, black dashed line, and $T=1.0$ s, black solid line), and compared with the BSSA14 model (grey lines). The dots correspond to the center of the magnitude or distance bins

Moreover, the decrease is about the same for short and long periods, which is a major difference with BSSA14 model, which considers a larger decrease for longer periods. -The two bottom rows of Figure.III.10 display the variation of ϕ with R_{JB} , for small ($M_w \leq 4.5$) and larger ($M_w \geq 5.5$) events, considered here for stiff sites ($V_{s30} \geq 300$ m/s). There is definitely a trend for an increase of ϕ with increasing R_{JB} , but this trend is much weaker for larger events and may not be a large concern in engineering applications. The observed trends seem similar at short and intermediate periods, while BSSA14 model includes a stronger period dependence (larger ϕ for shorter periods).

As seismic hazard for active crustal regions is most often controlled by $M_w > 5.5$ events at R_{JB} less than 80km, the variation of τ and ϕ with M_w and R_{JB} should not impact very significantly the hazard estimates, at least as long as the contributing M_w do not exceed 7.

III.13. Comparison of the median ANN model with BSSA14

A comparison of the median ground-motions values from the [BSSA14](#) GMPE and our principal model ANN(All V_{s30}, V_{s30}) (Equation.III.3) is presented for a suite of deterministic cases. All the parameters used in BSSA14 which were unknown were assigned their default values (i.e., $\delta ZI.0 = 0$, unknown style of the fault).

III.13.1. Distance scaling

The median attenuation of PGA for horizontal motions from the two GMPEs is plotted in Figure.III.11 as a function of R_{JB} for M_w ranging from 3.5 to 7.5. This comparison is performed here for a $V_{s30} = 500$ m/s. The two GMPEs predict similar median ground-motions over a broad M_w range. In particular, the distance decay shown by the ANN(All V_{s30}, V_{s30}) model (fully data driven without any a priori assumptions) is very close to the model elaborated with classical methods. The comparisons between the models for the other ground-motions parameters (PGV , spectral ordinates at various periods, not shown here) are similar to the results shown in Figure.III.11 for PGA : slight differences are observed for $M_w=5.5$, the BSSA14 model provide larger predictions, which may be considered as a consequence of the specific functional form including a kink at $M_w=5.5$. The best agreement is obtained at intermediate distances, corresponding to the core of the NGA-West 2 dataset. Another interesting and physically sound result relates to the magnitude scaling effects (Figure.III.11). As expected, the PGA increase at large magnitudes is smaller than the PGA increase at moderate magnitudes.

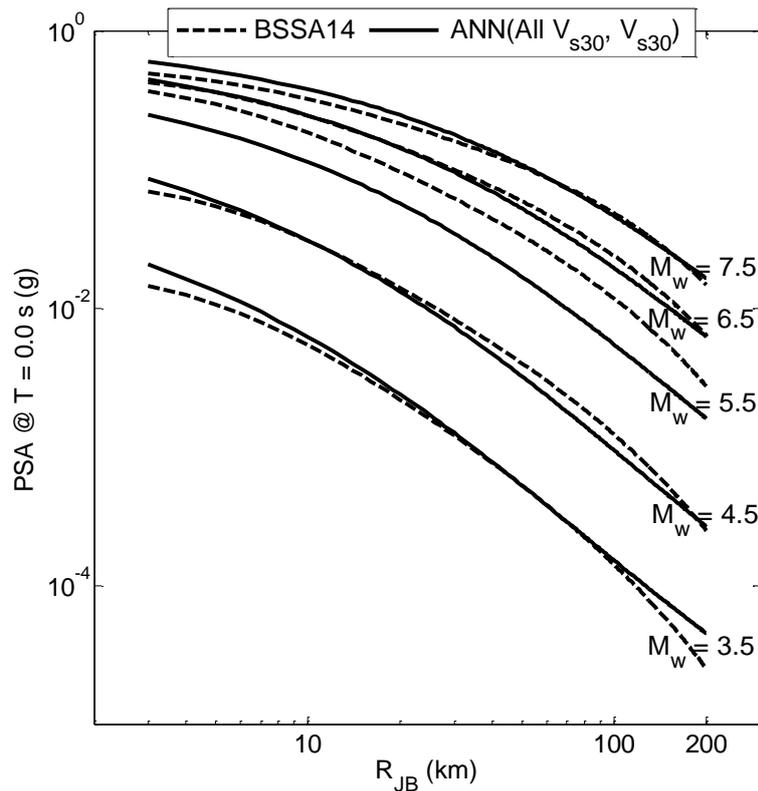


Figure.III.11 Comparison of the distance scaling for different magnitudes (3.5 to 7.5) between the present ANN(All V_{s30} , V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for PGA at a stiff site with $V_{s30} = 500$ m/s

III.13.2. Magnitude scaling

The effect of M_w scaling at a R_{JB} of 30 km is displayed in Figure.III.12, again for $V_{s30} = 500$ m/s. Plots are provided for PGA and PSA at spectral $T = 0.2, 1.0,$ and 2.0 s. The M_w scaling between the two models proves to be very similar over the wide range of $M_w=[3.5$ to $7.5]$ except in the range $M_w=5.5 / 6$: the ANN model exhibits a smooth behavior while the BSSA14 model includes a "built-in" kink at a " M_h " hinge magnitude, which ranges between 5.5 and 6.2 depending on the spectral period. This value of M_w is included in many GMPEs which is named " M_h " hinge magnitude.

III.13.3. V_{s30} scaling

Both GMPEs use the V_{s30} site proxy and can then be compared easily. Figure.III.13 shows the dependence of spectral ordinates at four period values as a function of V_{s30} for a $M_w=6$ at $R_{JB} = 30$ km. In such a case, the ground-motions remain moderate (PGA below 0.15 g). The slight convexity of the short period BSS14 curves for low V_{s30} values is linked to the slight non-linearity in the corresponding functional form. The ANN curves exhibit an overall

similar V_{s30} scaling, which however does not exhibit any convexity for low V_{s30} values. A closer look at the dependence of soft soil / rock amplification in the ANN model does indicate however both a larger amplification level for small rock motion values (as illustrated in Figure.III.14) and a significant non-linearity (i.e., decrease of amplification with increasing rock motion), which does not map in a convex shape when the amplification is plotted as a function of V_{s30} for a given rock motion level. Such a difference in the NL components of either GMPE is presently the subject of further investigations, which are however much beyond the scope of the present paper. At this stage, one may simply mention that the background of BSSA14 model (and all other NGA-West 2 models) on one hand, and the present ANN approach on the other hand, are different: the latter are only driven by the NGA-West 2 data, while the former are also based on a comprehensive set of numerical simulations ([Kamai et al. 2014](#); [Seyhan and Stewart 2014](#)) with prescribed constitutive models.

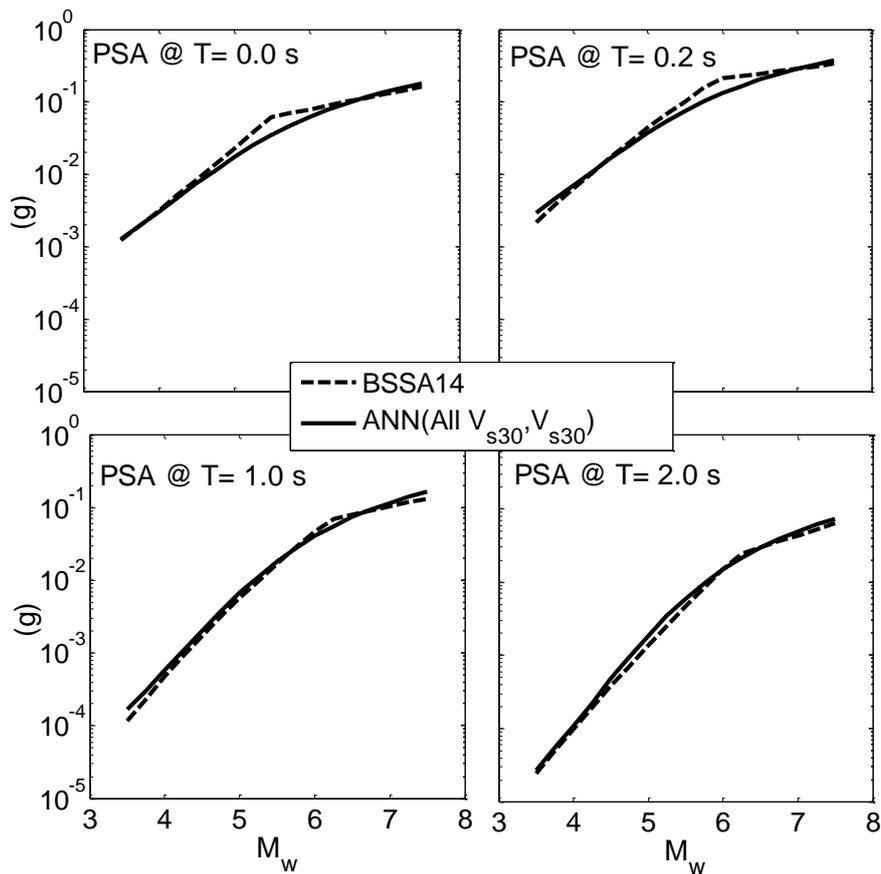


Figure.III.12 Comparison of the magnitude scaling for different spectral accelerations ($T= 0, 0.2, 1.0$ and 2.0 s, from top left to bottom right) between the present ANN(All V_{s30}, V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for a stiff site with $V_{s30} = 500$ m/s located at a R_{JB} distance of 30 km.

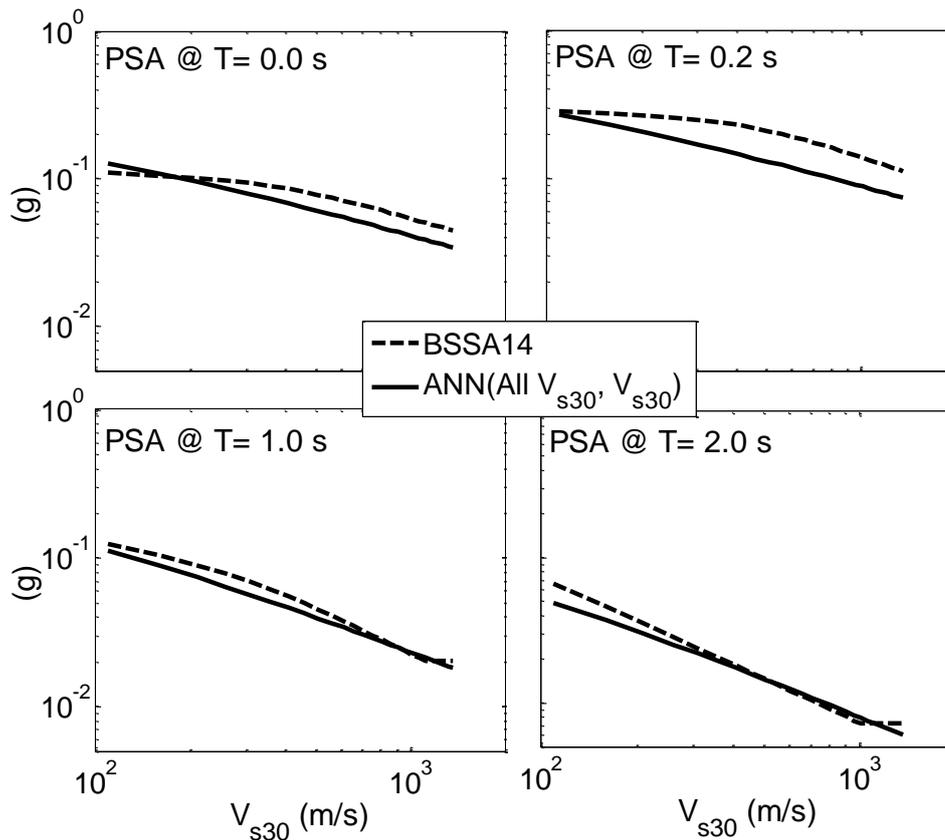


Figure.III.13 Comparison of the V_{s30} Scaling for different spectral accelerations ($T= 0, 0.2, 1.0$ and 2.0 s, from top left to bottom right) between the present ANN(All V_{s30}, V_{s30}) model (solid line) and the BSSA14 model (dashed line). This comparison is illustrated here for a site located at a R_{JB} distance of 30 km of a $M_w=6$ event.

III.13.4. Response spectra

The response spectra predicted for a M_w of 3.5 to 7.5 earthquakes at a distance $R_{JB}=30$ km on a stiff site ($V_{s30} = 500$ m/s) are plotted in Figure.III.15. There is a very satisfactory similarity between the two models for almost all M_w cases. The main differences correspond to the $M_w=5.5$ case for $T < 0.3$ s, and are linked to the "hinge magnitude" M_h specifically included in the BSSA14 model, leading to an overestimation at short periods in a narrow magnitude range.

These comparisons show that the derived ANN(All V_{s30}, V_{s30}) ground-motions model has the ability of traditional GMPEs to predict the amplitude and variability of ground-shaking within the NGA-West 2 dataset, except for the non-linear site terms, which are controlled indeed by the numerical simulations and not by the purely empirical data set.

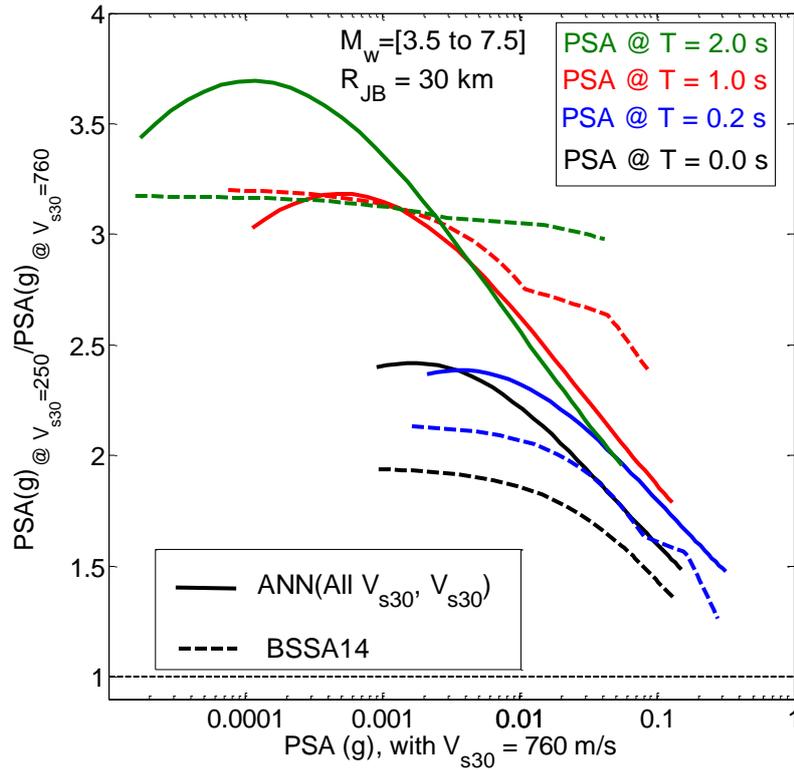


Figure.III. 14 Illustration of the non-linearity in the site term. The site amplification factor between a relatively soft site with $V_{s30} = 250 \text{ m/s}$ and a standard rock site characterized by $V_{s30} = 760 \text{ m/s}$, is displayed as a function of the corresponding rock spectral amplitude. For the sake of simplicity, the variation of the rock motion level is considered by varying the magnitude from 3.5 to 7.5 at a fixed R_{JB} distance of 30 km. The unusual shape of BSSA14 curves is related to the fact that BSSA14 uses the rock PGA (and not the rock $PSA(T)$) to constrain the non-linear term.

III.13.5. Comparison of the sigma (σ)

The period dependencies of the aleatory variability models are compared in Figure.III.16. As BSSA14 considers a very complex model where σ_t depends on M_w , distance, and even site conditions, this comparison is performed for a $M_w=4$, $M_w=7$, $V_{s30}=270$ and $V_{s30}=600 \text{ m/s}$ for BSSA14 at an $R_{JB}=30 \text{ km}$. For the ANN(All V_{s30} , V_{s30}) model, the only parameter controlling σ_t is the type of V_{s30} (measured or inferred, see Equation.III. 1 and 5). σ is found to be almost always larger for ANN compared to BSSA14, except at short periods ($T \leq 0.07 \text{ s}$) for $M_w=4$ and measured V_{s30} . The large M_w BSSA14 σ_t values are much lower than the corresponding values for ANN(All V_{s30} , V_{s30}), even for measured V_{s30} . The origin of such large differences has certainly to be investigated further, as they certainly deeply impact PSHA results. Certainly, the present simple ANN model does not detect in the NGA-West 2 dataset the very large reduction of the between-event variability beyond $M_w=5.5$ which is included in the BSSA14 model (see Figure.III.12); adding a few additional source parameters (style of faulting, hypocenter depth, ...) could help reducing the variability for large events.

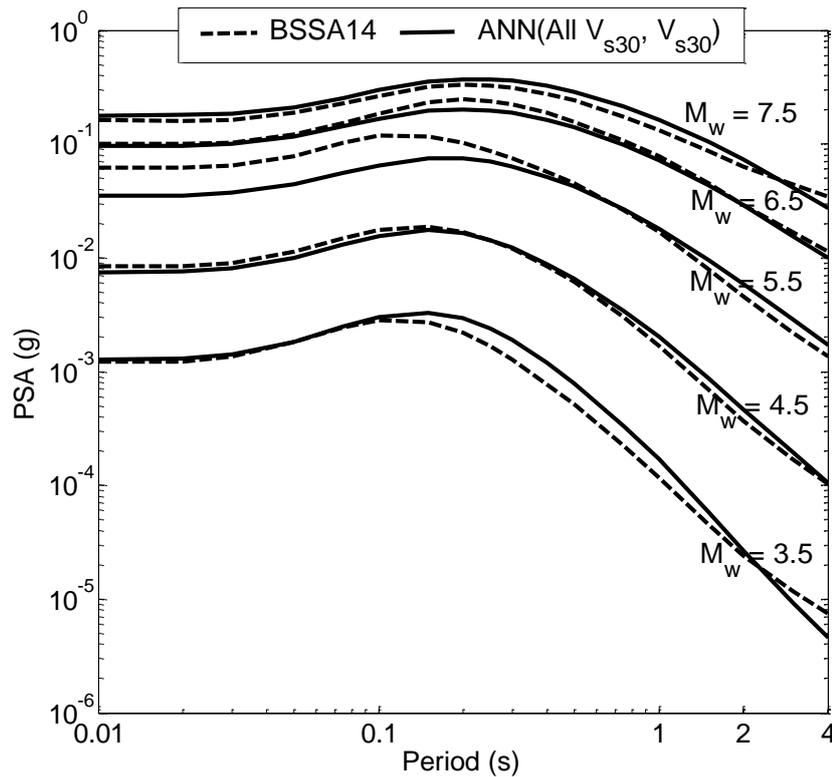


Figure.III.15 Comparison of median spectra predicted for increasing magnitudes at a stiff site $V_{s30} = 500$ m/s located at a R_{JB} distance of 30 km, between the present ANN($All V_{s30}, V_{s30}$) model (solid line) and the BSSA14 model (dashed line).

III.14. Discussion and conclusions

This series of investigations with the ANN approach on two recent, high-quality strong motion databases brought interesting insights on several much debated issues

a) Efficiency of various widely used SCPs (measured/inferred V_{s30} , *Slope*).

It turns out that the largest ϕ reduction related to site conditions is achieved when having measured V_{s30} . Inferred values (using various correlations) perform more poorly, comparably to the *Slope* proxy. When V_{s30} is measured, considering a continuous dependence on V_{s30} performs slightly better than considering a limited number of site classes based on V_{s30} intervals, as proposed in NEHRP or EC8. When V_{s30} is simply inferred, however (including using *Slope*), then introducing a continuous dependence or considering a limited number of classes provide equivalent results. It is therefore highly recommended to systematically perform S-wave measurements, firstly on strong motion networks and also on sites where hazard assessment has to be done. The latest comprehensive comparison works between various invasive and non-invasive S-wave measurement techniques concluded that the reliability and robustness of V_{s30} estimates from surface-wave, non-invasive techniques (active and/ or passive) is comparable to those of borehole measurements (e.g., [Moss, 2008](#); [Bard et al. 2010](#); [Foti et al. 2015](#); [Garofalo et al. 2015a,b](#)). This should give further encouragement to all strong motion

operators and geotechnical companies to promote such measurements. When measurements are not available, the price to pay is an increased within-event and total aleatory variability, by a factor exceeding 10% at all short and intermediate periods (up to 1 s), and decreasing to 6% at longer periods (up to 4s). Finally, it should be emphasized that the above conclusions on the respective performance of various SCPs cannot be extrapolated to other datasets. A number of authors have drawn the attention to the limitations of the V_{s30} parameter, which is only a proxy and cannot describe alone the whole physics of site amplification. A number of other proxies or groups of proxies have been proposed, coupling information on the shallow impedance and the overall sedimentary thickness (Z_{800} , Z_{1000} , Z_{2500} ; f_0 ; H and V_{SH} , amongst many other possibilities). It would thus be highly desirable to perform the same kind of investigation on other datasets offering more versatility as to the choice of SCP; it is unfortunate that most existing strong-motion datasets are exclusively oriented towards V_{s30} as the main SCP.

b) Structure and amount of the aleatory variabilities

Our results qualitatively confirm the M_w and distance dependency of σ_τ shown by [Boore et al. \(2014\)](#). We see that ϕ increases with R_{JB} , but only beyond about 100 km. At closer distances, ϕ is approximately constant with respect to distance. As discussed by [Boore et al. \(2014\)](#), this increase may be influenced by epistemic uncertainty in regional attenuation rates. τ decreases with M , but is nearly constant for $M > 5.5$. The comparative analysis of various datasets (RESORCE and three subsets of the original NGA-West 2 database), however, indicates that the σ_τ is sensitive to the dataset. In particular, for NGA-West 2, the largest variability is found for the NGA-West 2 – Inferred V_{s30} subset. The origin of such sensitivity is not fully understood yet. It probably includes a larger number of small M_w events with less constrained source parameters (size, hypocentral location), and may be also a larger stress drop variability, together with an overall lower attention to data quality for small M_w earthquakes. In any case, this indicates that one should be cautious when commenting on the overall levels of aleatory variability and deriving conclusions from small differences in the structures of the two between- and within-event components. In particular, the actual efficiency of any given "input" parameter considered as an explanatory variable, should be estimated only in the light of its ability to reduce the residuals.

c) Performance of the ANN approach and fully data driven GMPEs

The final ANN model built on the NGA-West 2 dataset combines simplicity, physical soundness and regularity. Even though it has only three explanatory variables (M_w , R_{JB} and V_{s30}), its results are very consistent with BSSA14 (which, though having the simplest functional

forms of the whole NGA-West 2 set of GMPEs, involves many more input parameters). It proves in particular successful in reproducing the correct scaling with M_w , distance (including larger decay for small magnitudes and short distance saturation), and V_{s30} . Even though it was not specifically emphasized in this paper, the ANN model does predict significant non-linearities in the site response; this is another issue which will require additional investigations with a more appropriate dataset, such as, for instance, the KiK-net data, which include a large number of sites with well constrained weak and strong motion recordings at surface and depth. One of the advantages of the ANN approach is the absence of bias associated with an a priori choice concerning the functional form; one associated drawback, however, is the impossibility to use such models outside the range of input parameters considered for the learning data set. A detailed analysis of the binned residuals also allowed to confirm the probable decrease of τ with increasing M_w , without showing any clear evidence for a dependence of ϕ with distance or V_{s30} . The large impact of small changes in σ_t , τ , and ϕ on long return period hazard is the reason behind the effort to reduce them; the application of the ANN approach to the top-quality NGA-West 2 data set draws attention to the need to further test the robustness of aleatory variability models. Meanwhile, the main outcome of the present study is the confirmation of [Chiou and Youngs \(2008\)](#) proposition to relate the amount of within-event variability to the quality of the SCP.

III.15. Acknowledgments

The authors thank Olga-Joan Ktenidou for her help and Anne Lemoine for the *Slopes* values of the RESORCE dataset. We also thank Jonathan P. Stewart and an anonymous reviewer for their constructive criticism and comments that helped us to improve this study. We acknowledge the support from the TASSILI program and the SIGMA project (<http://projet-sigma.com/>). The authors would like to thank the participants of the NGA-West 2 program for providing high quality data and stimulating ideas.

III.16. Data and resources

The datasets used in this article have been collected and disseminated by The Pacific Earthquake Engineering Research Center and Euro-Mediterranean Seismological Centre (EMSC) data management center.

Conclusion

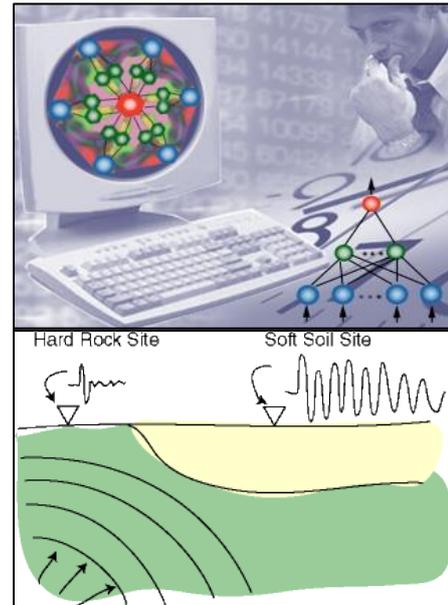
Les principaux résultats obtenus dans ce chapitre ont été cités dans la partie ‘Discussion and conclusions’ de l’article ([Derras et al. 2016](#)). Le point le plus important est peut-être le fait que

nous avons obtenu un taux maximum de réduction de la variabilité aléatoire égal à 15 % en utilisant un proxy de site lié à V_{s30} mesurée.

Pouvons-t-on améliorer encore plus la prédiction des effets de site en utilisant d'autres stratégies ? Dans la chapitre IV nous allons ainsi tester l'utilisation de plusieurs proxies de sites avec cette fois-ci un sous-ensemble des données du réseau KiK-Net.

Chapitre IV

Etude de la capacité de différents proxys de site à réduire la variabilité aléatoire des équations de prédiction des mouvements du sol



Introduction

L'utilisation d'un seul proxy de site dans une équation de prédiction du mouvement sismique peut être insuffisante. Une telle utilisation engendre une perte d'information pour quelques fréquences. V_{s30} est préconisé généralement pour représenter des formations géologiques peu profondes. Une utilisation extrapolée vers des sols plus profonds engendre probablement une prédiction non appropriée du mouvement sismique. Les incertitudes pour une telle formation seront plus grandes, rendant ainsi l'évaluation de l'aléa sismique obsolète. Pour contourner ce problème, nous testons dans ce chapitre, l'utilisation simultanée de plusieurs proxys de sites afin de réduire la variabilité aléatoire du mouvement sismique.

Pour ce faire, nous commençons par une démarche classique ([1^{ère} partie](#)) : tester différents proxys de site pour les GMPEs : en plus de V_{s30} et la pente topographique « slope » déjà évoquées dans le chapitre précédent, nous ajoutons la fréquence de résonance f_0 et la profondeur « H_{800} » à une $V_s = 800$ m/s. Pour les effets de sources et de propagation d'onde nous considérons M_w , R_{JB} et la profondeur focale « Depth ». Comme dans le chapitre II et III, le mouvement sismique est représenté par le PGA et PGV et les ordonnées spectrales entre $T = [0.01 \text{ à } 4]$ s avec un coefficient d'amortissement égal à 5 %.

La deuxième démarche ([2^{ème} partie](#)) consiste à effectuer une étude de sensibilité des différents proxys de site par le rapport ([mouvement fort]/[mouvement faible]) des rapports spectraux (site/rocher) ([Régnier et al. 2016](#)). Cela permet en principe de modéliser mieux l'effet de site non linéaire. Cette tâche est effectuée avec 5 proxys de sites et 2 proxys du mouvement

sismique. Dans la 2^{ème} partie, deux types de bases de données sont considérés : la base de données « sites » et la base de données « événements ».

La réalisation de ces deux études est effectuée avec la base de données KiK-Net et l'approche neuronale. Nous commençons dans ce qui suit par la première démarche :

1^{ère} partie : V_{s30} , pente topographique, H_{800} et f_0 : La performance de différents proxys de site-condition dans la réduction de la variabilité aléatoire au sol et la prédiction de la réponse non linéaire du site

Earth, Planets and Space

V_{s30} , slope, H_{800} and f_0 : performance of various site-condition proxies in reducing ground-motion aleatory variability and predicting non-linear site response

--Manuscript Draft--

Manuscript Number:	EPSP-D-17-00014
Article Type:	Full paper
Section/Category:	Seismology
Abstract:	The aim of this paper is to investigate the ability of various site-condition proxies (SCPs) to reduce ground-motion aleatory variability and the manner with which the SCPs capture the nonlinearity site effect. The SCPs used here are time-averaged shear-wave velocity in the top 30 m (V_{s30}), the topographical slope (slope), the fundamental resonance frequency (f_0) and the depth beyond which V_s exceeds 800 m/s (H_{800}). We considered first the performance of each SCP taken alone, and then the combined performance of the 6 SCP pairs [V_{s30} - f_0], [V_{s30} - H_{800}], [f_0 -slope], [H_{800} -slope], [V_{s30} -slope] and [f_0 - H_{800}]. This analysis is performed using a Neural Network approach including a random-effect applied on a KiK-net subset for derivation of Ground Motion Prediction Equations (GMPEs) setting the relationship between various ground motion parameters Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV) and Pseudo-Spectral Acceleration PSA(T), and Mw, RJB, focal depth and SCPs. While the choice of SCP is found to have almost no impact of the median ground motion prediction, it does impact the level of aleatory uncertainty. V_{s30} is found to perform the best of single proxies at short periods ($T < 0.6$ s), while f_0 and H_{800} perform better at longer periods; considering SCP pairs leads to significant improvements, with particular emphasis on [V_{s30} - f_0] and [f_0 -slope] pairs. The results also indicate significant non-linearity in the site terms for soft sites, and that the most relevant loading parameter for characterizing nonlinear site response is the "rock" spectral ordinate at the considered period.
Corresponding Author:	Boumediene Derras, Associate Professor Universite Abou Bekr Belkaid Tlemcen Faculte de Technologie Tlemcen, ALGERIA
First Author:	Boumediene Derras, Associate Professor
Order of Authors:	Pierre-Yves Bard, Professor Fabrice Cotton, Professor

IV.1. Introduction

Probabilistic seismic hazard analysis (PSHA) strongly relies on Ground Motion Prediction Equations (GMPEs) that quantify the amplitude of ground motion as a function of distance, magnitude and site condition proxies (SCPs). The latter are introduced to characterize the amplification effects linked to near-surface deposits. Given the variety of physical phenomena impacting the characteristics of earthquake shaking, ground motion models include a large degree of uncertainty. This uncertainty (especially the within-event aleatory variability) is strongly affected by the near-surface site conditions. An important question is the degree to which this scatter can be reduced by improvements in the way to account for the near-surface effects. The incorporation of those effects in GMPEs has gone through an evolution in the past years ([Chiou and Young 2008](#), [Seyhan et al. 2014](#), [Derras et al. 2016](#)). At the beginning, ground-motion models typically contained a scaling parameter based on site classification (e.g., [Boore et al. 1993](#)), or presented different models for “hard rock” and “soil” sites (e.g., [Campbell 1993](#), [Sadigh et al. 1997](#)). [Boore, \(1997\)](#) introduced the explicit use of the time - averaged shear - wave velocity in the top 30 m (V_{S30}). V_{S30} has become de facto a standard for the development of GMPEs and seismic-hazard assessment at national and international scales. In this way, it has been observed (e.g., [Borcherdt 1994](#)) that V_{S30} is a useful parameter to predict local site amplification in active tectonic regimes, especially when it is actually measured: [Derras et al. \(2016\)](#) showed, after [Chiou and Youngs. \(2008\)](#) that measuring V_{S30} allows a significant reduction of the aleatory variability.

V_{S30} has certainly proved to constitute a simple and efficient SCP metric, but it also proved not to be a low cost SCP, as it is far from being measured at all strong motion sites throughout the world (except in Japan). For this reason, [Wald and Allen \(2007\)](#) and [Allen and Wald \(2009\)](#) have proposed to use the topographic slope (slope) from Digital Elevation Models (DEMs) derived from remote sensing (satellite imaging) to give a first-order estimation of site classes based on V_{S30} . Many other ways to infer V_{S30} values without measuring them have been proposed, as listed in [Seyhan et al. \(2014\)](#): extrapolation from V_s measurements at depths shallower than 30m, correlations – more or less robust – with other types of information or parameters (geology, geomorphological or terrain-related proxies, geotechnical parameters). On another hand, V_{S30} alone cannot satisfactorily predict the amplification for sites underlain by deep sediments, which require knowledge of the geology to depths greater than 30 m (e.g., [Choi and Stewart 2005](#); [Luzi et al. 2011](#)). [Campbell \(1989\)](#) found that adding a parameter for depth to basement rock improved the predictive ability of empirical ground motion models. On

their side, [Cadet et al. \(2012\)](#) and [Derras et al. \(2012\)](#) used another SCP: the fundamental resonance frequency, f_0 , as determined by the horizontal-to-vertical (H/V) spectral ratio technique ([Mucciarelli 1998](#), [Haghshenas et al. 2008](#), [Bard et al. 2010](#)). As the f_0 (H/V) SCP is able to identify low frequency amplification on thick sites, its relevance may be compared with the performance of another SCP often proposed to properly account for sediment thickness, H_{800} (depth beyond which the shear wave velocity exceeds 800 m/s).

The main aim of this work is to assess the actual performance of several site condition proxies, namely V_{S30} , slope, f_0 (H/V) and H_{800} SCP, by analysing the relative decrease of the ground-motion aleatory variability each of them allow to achieve, and by investigating the benefits of considering simultaneously multiple site proxies. A secondary aim in this study is to investigate to which extent these SCPs allow to capture not only the linear, but also the non-linear nature of site amplification.

The KiK-net database used here consists of shallow crustal events recorded on sites for which several site proxies are already available: V_{S30} and H_{800} values can be directly derived from downhole measurements of V_S profile ([Dawood et al. 2014](#)), the slope values have been compiled ([Ancheta et al. 2014](#)), while f_0 values are taken from [Régnier et al. \(2013\)](#). The dataset used in this manuscript corresponds to data recorded between 1997 and December 2011. The Artificial Neural Network (ANN) approach and a random-effects like procedure ([Derras et al. 2014](#)) have been used for the derivation of GMPEs setting the relationship between various ground motion parameters (Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV) and 5% damped Pseudo-Spectral Acceleration (PSA) from 0.01 s to 4 s), and event/station meta-parameters (moment magnitude M_w , Joyner & Boore distance R_{JB} , focal depth, and site condition proxies V_{S30} , slope, H_{800} and f_0). The KiK-net data offer the unique opportunity to have for each strong motion recording a reliable measurement of the four SCPs, thus allowing a thorough and meaningful comparative assessment of the performance of each of these proxies. After a short presentation of the dataset and the selection criteria used for the present study, a section is dedicated to the presentation of the ANN models and their specific implementation for deriving GMPEs in a way similar to the classical "random effects" approach. The following section presents the results obtained for the KiK-net data, focusing on a) the respective performance of each of the four site proxies which are considered either alone or within combinations, and b) a discussion of their ability to detect and account for non-linear site response.

IV.2. Data set

The Kiban-Kyoshin network (KiK-net) is one of the two national strong motion seismograph networks developed in Japan following the 1995 Kobe earthquake. The KiK-net is a network of strong motion instruments that consists of about 700 stations with an average spacing of about 20 km distributed throughout the Japan islands ([Hayashida and Tajima 2007](#)). The KiK-net stations are each equipped with a pair of surface and downhole, sensitive 3-component digital accelerometers, allowing an empirical evaluation of the site response at each station.

The resulting dataset considered here has been compiled by Dawood et al. ([2014](#), [2016](#)) and consists of 5563 recordings from 393 sites to 249 earthquakes. This subset has been downloaded from <https://datacenterhub.org/resources/272>. The range of the M_w , R_{JB} , Depth and all SCPs are listed in Table.IV.1, which also provides the number of earthquakes, records and sites. The corresponding range of recorded PGA values spans from 10^{-4} to 0.4 g.

Table.IV.1 Range of magnitude, distance and site parameters for the KiK-Net subset considered in this study

Metadata parameters	Parameters range		Number of Earthquakes	Total Number of recordings	Total Number of sites
	Min	Max			
M_w	3.7	7.3	249	5563	393
R_{JB} (km)	3.04	1416			
Depth (km)	0.00	30.00			
V_{S30} (m/sec)	106.83	1433			
f_0 (Hz)	0.22	22.71			
H_{800} (m)	1	1000			
Slope (m/m)	0.0011	0.375			

IV.2.1. Site-condition proxies

The main goal of this paper is to investigate the impact of various SCP choices on the level of the resulting aleatory variability for ground motion models, and analyse how they actually account for non-linear site response. In that aim, we considered V_{S30} and slope, which are generally considered a priori as more relevant for short period ground-motions, and f_0 (H/V) and H_{800} , that should in principle be more suitable for long periods. Ground motion models are derived first using one single SCP [one model with each of the four values: V_{S30} or slope or f_0 or H_{800}], then using two SCPs out of the four [i.e., 6 models in total with the six pairs (V_{S30} , slope), (V_{S30} , H_{800}), (V_{S30} , f_0), (H_{800} , slope), (f_0 , H_{800}) and (slope, f_0)]. The All SCPs being

considered again through their \log_{10} values for testing the improvements potentially brought considering a twin description of site conditions. In addition, two reference ground motion models are established for comparison with each of the 10 previous ones: the first one is without any site proxy (named “without proxies”), and the second one considers simultaneously all 4 site proxies (named "all proxies") to estimate the maximum possible standard deviation reduction.

IV.2.2. Data distribution

The distribution of the dataset according to M_w , R_{JB} , focal depth, PGA and site-condition proxy (SCP) is displayed in Figures.IV.1 and 2.

Figure 1 shows the distributions of the KiK-net dataset in the magnitude-distance plane by bins of PGA, and in the PGA-distance plane by bins of M_w . The distributions are given for all site conditions (left column), but also for soft sites only (V_{S30} from 100 to 300 m/s, middle column) and stiff sites only ($V_{S30} > 800$ m/s, right column). The goal of these presentations by bins of M_w , PGA and V_{S30} , is to ensure that the data distribution is appropriate for all M_w - PGA and for soft and stiff soils. This Figure also illustrates clearly the much smaller number of recordings at distances less than 10 km in general, and less than 30 km when only stiff to rock sites are considered.

Figure.IV.2 represents the cumulative distribution function (CDF) of the used dataset versus R_{JB} , M_w , V_{S30} , topographical slope, H_{800} , f_0 , focal depth and PGA. The four SCP distributions are found to follow a lognormal distribution as well as R_{JB} and PGA, while M_w and focal depth are much closer to a normal distribution (see also Figure 16 in [Derras et al. 2012](#)). In our ANN models, we thus used the logarithm (base 10) values of all SCPs and spectral ordinates PSAs. Table.IV.2 details a fractile values for each of these metadata parameters: median value, 5% and 95% fractiles which are considered to provide the range of applicability of the models, and the 10% and 90% fractiles which will be used in the following to estimate the impact of each SCP on the site amplification factor.

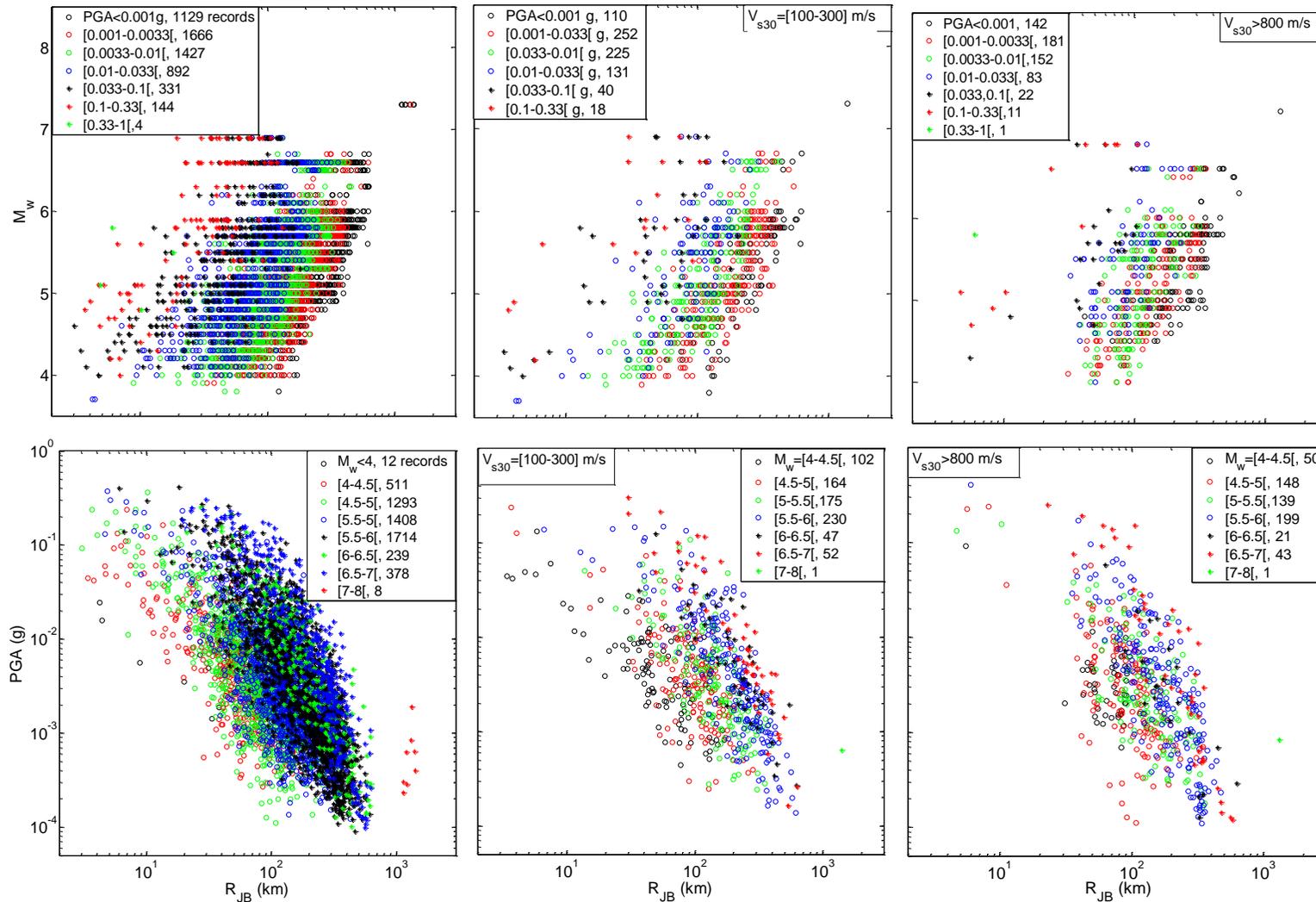


Figure IV.1 Distribution of the KiK-net dataset considered in this study. The three top plots correspond to the Magnitude (M_w) - distance (R_{JB}) distribution. For these three plots, the bins with different colors correspond to seven sample-size subsets with increasing PGA values from 0.001 to 1 g. On the three bottom plots, the distribution of PGA versus distance (R_{JB}) is shown. In these plots, the distribution is given by bins of Magnitude, $M_w=[4 - 8]$. On the left we take all dataset, in the middle plots the show data correspond to $V_{s30} = [100-300] \text{ m/s}$ and in the right plots the datas are given for $V_{s30} > 800 \text{ m/s}$.

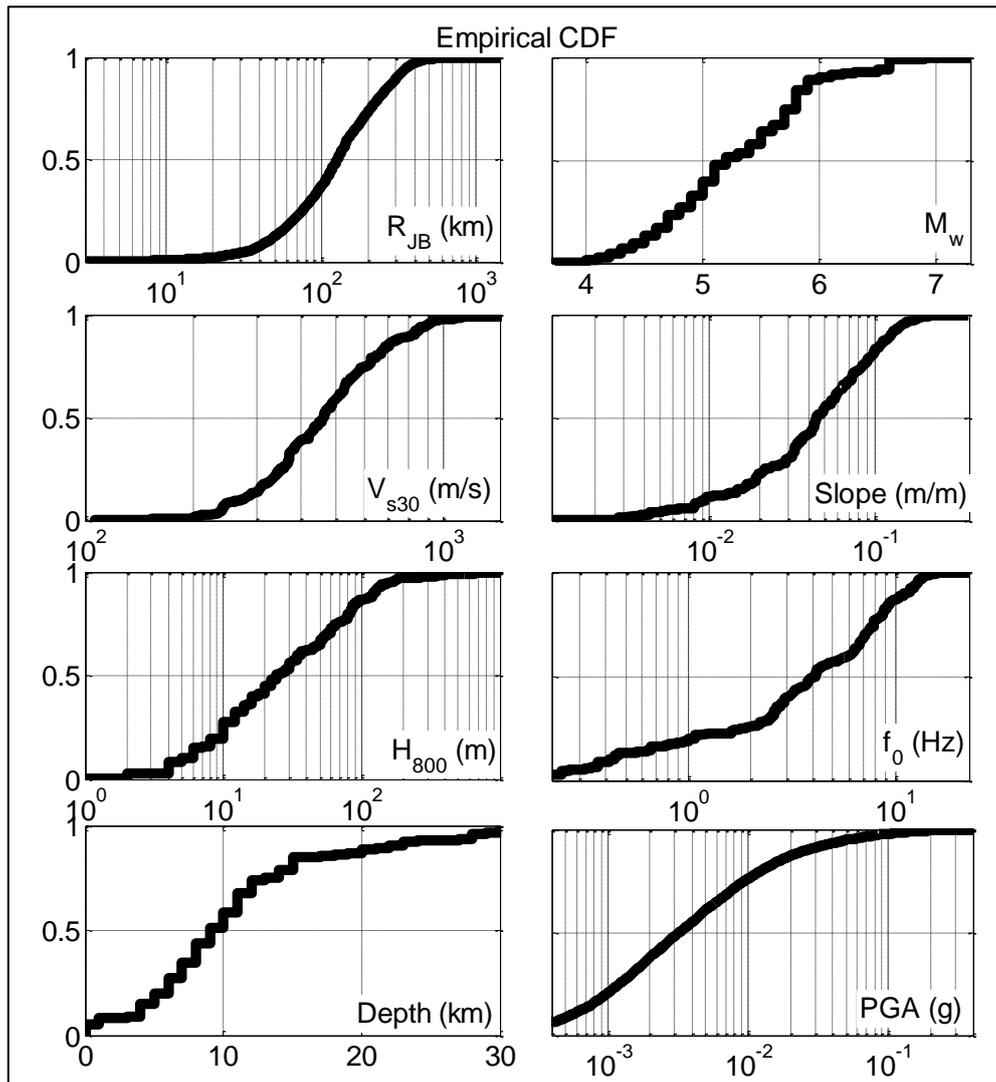


Figure.IV. 2 Empirical cumulative distribution function (CDF) versus the explanatory variables R_{JB} (top-left), M_w (top-right), V_{S30} (first middle-left), slope (first middle-right), H_{800} (second middle-left), f_0 (second middle-right), depth (bottom-left) and the response PGA variable (bottom-right) for the considered KiK-net data sets.

To ensure that the SCPs are not strongly dependent on one another, correlation plots are displayed for each pair of SCPs (Figure.IV.3) together with the corresponding correlation coefficient (R). Although some pairs do exhibit some correlation ($R_{\max} = 0.57$ between V_{S30} and f_0), the scatter is large enough for the SCPs to be considered as almost independent site parameters for the derived ANN models. The weaker correlation is found between slope and either H_{800} or f_0 . One may notice that H_{800} is negatively correlated with the three other SCPs: the larger it is, the lower are f_0 , V_{S30} and slope – as could be intuitively expected. In the same Figure are also indicated the median values and the 10% and 90% fractiles for the all SCPs.

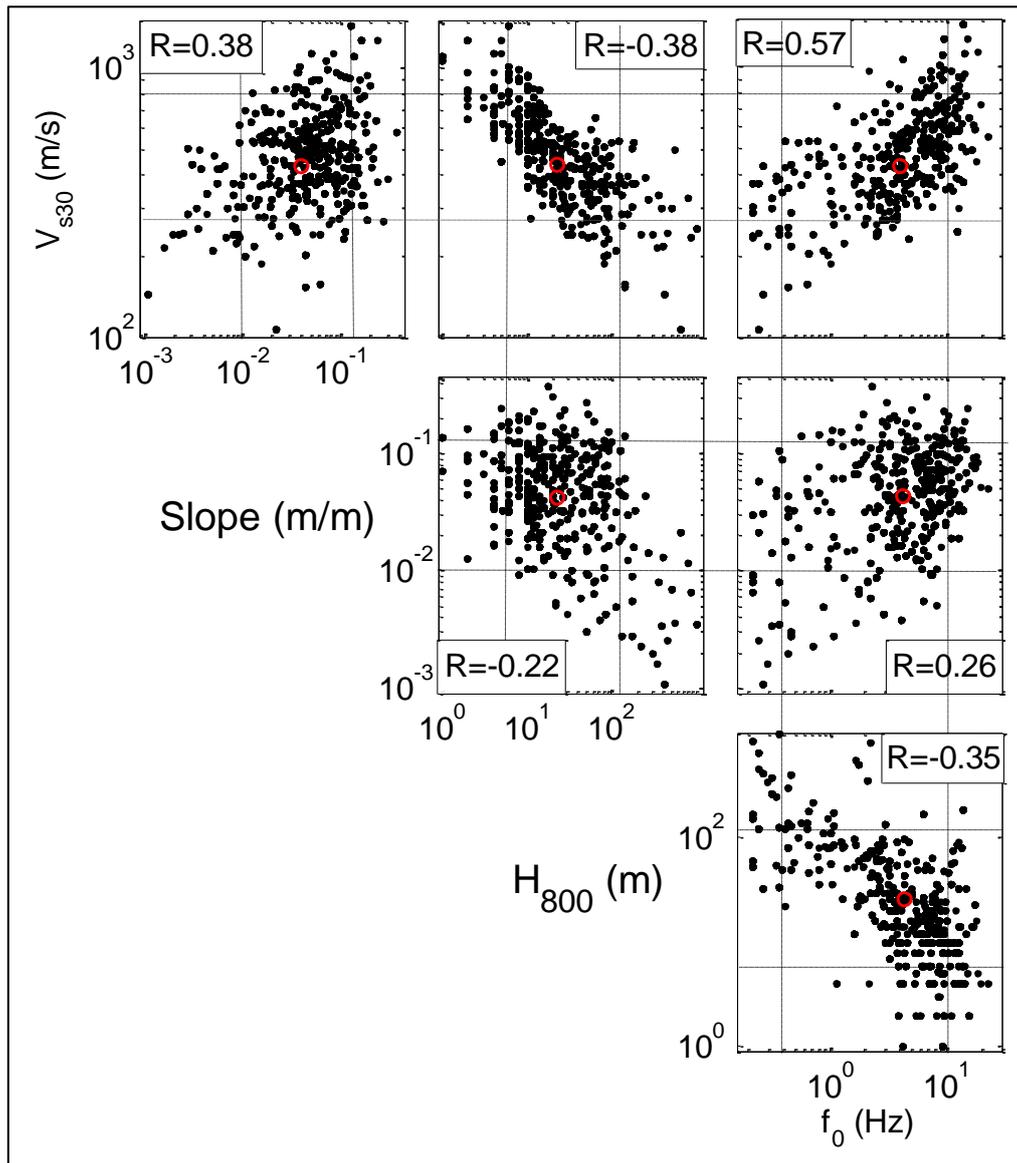


Figure.IV.3 Correlation between the different SCPs, with the correlation coefficient values for each pair of SCPs. The dotted line represents the extreme values of the PCS: 10% and 90% of the empirical CDF respectively. The red circles present the SCPs medians values. The extreme and medians values are presented in the Table.IV.2.

IV.3. Methodology

The random-effects regression algorithm, made popular within engineering seismology by [Abrahamson and Youngs \(1992\)](#), is arguably the most commonly used approach for developing empirical ground-motion models. In our ANN models, we used this type of approach in order to facilitate the comparability with classical GMPEs. The ANN has the advantage that no prior functional form is needed ([Derras et al. 2012](#)): the actual dependence is established directly from the data, and can therefore be used as a guide for a better understanding of the factors which control ground-motions. This resulted in a two-phase building process:

Table.IV.2 Metadata range for site effect and for fully data-driven models applicability.

CDF (%) Metadata parameters	Range of applicability of fully data-driven models			Extreme values for site effect analysis	
	Median			10 %	90 %
V_{S30} (m/sec)	460	235	900	270	800
f₀ (hz)	4	0.30	12	0.4	11
H₈₀₀(m)	25	2	170	6	120
Slope (m/m)	0.04	0.006	0.15	0.01	0.13
M_w	5.5	4	7	4.5	6
R_{JB} (km)	127	7	361	46	310
Depth (km)	10	1	30	4	23

IV.3.1. Fixed models

The architecture of ANN used in this work is named “feedforward Network”, consisting of a series of layers. The first layer ensures the connection with the input parameters, i.e. in our case M_w , R_{JB} and Depth and one or two (or possibly more) continuous parameter describing the SCPs (V_{S30} , slope, f_0 , H_{800}). Each subsequent layer has a connection from the previous layer. The final layer produces the network's output. A feedforward network with one hidden layer and three neurons in the hidden layer is adopted in this study. This small number of hidden neurons is the optimal number in order to optimize both the total standard deviation of residuals, σ , and the Akaike Information Criterion ([Akaike 1973](#)). Figure.IV.4 illustrates a typical architecture of the ANN-fixed models which were implemented within the MATLAB[®] Neural Network Toolbox[™] ([Demuth et al. 2009](#)). The output layer groups all the considered ground motion parameters, i.e. the classical geometric mean of the horizontal components of PGA, PGV, and 5%-damped PSA at 18 periods from 0.01 to 4 s. We did not include predictions for Peak Ground Displacement (PGD), which we consider to be too sensitive to the low-pass filters used in the data processing.

The Quasi-Newton Back Propagation technique also called “BFGS” (Broyden – Fletcher – Goldfarb - Shanno) has been applied for the training phase ([Shanno and Kettler 1970](#)). To avoid “overfitting” problems we chose an adequate regularization method involving the modification of the conventional mean sum of squares of the network errors by the addition of a term equal to the sum of squares of the network synaptic weights and bias ([Derras et al. 2012](#), [Derras et al. 2014](#)).

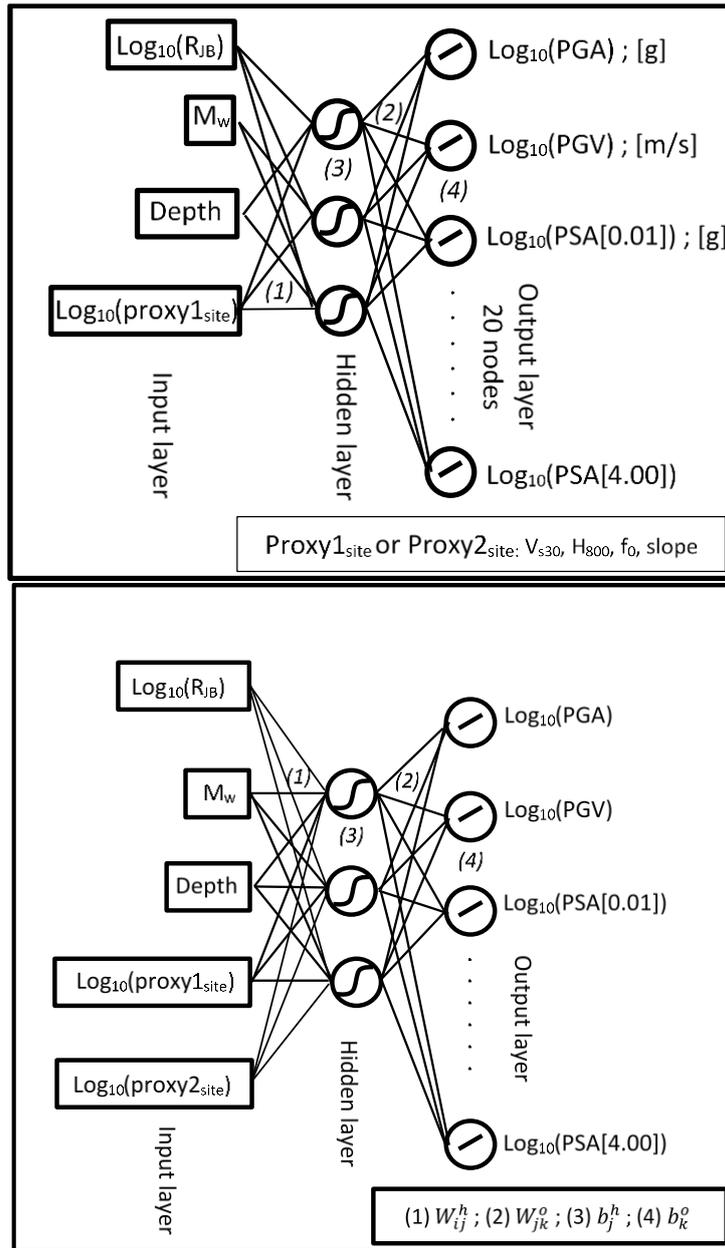


Figure.IV.4 Structure of the Neural Networks for PGA, PGV and PSA[0.01 to 4 sec] prediction, for one site proxy (top) and two-site proxies (bottom). The W_{ij}^h is the synaptic weight between the i th neuron of the input layer and the j th neuron in the hidden layer, b_j^h the bias of the j th neuron in the hidden layer. Also the W_{jk}^o is the synaptic weight between the j th neuron of the hidden layer and the k th neuron in the output layer, b_k^o the bias of the k th neuron in the output layer.

Moreover, the optimal activation functions were found to be a "tangent sigmoïd" for the hidden layer and "linear" for the output layer. Fully data driven GMPEs were developed, differing by the nature of parameters used in the input layer. The first ANN model is built on the basis of the M_w , R_{JB} and focal depth as inputs: it accounts only for source and path effects, and sets the reference to quantify the gains achieved by the consideration of the various site proxies in the other ANN models. The second pack of 4 ANN models considers only one-SCP in the input

layer, namely, V_{S30} , slope, H_{800} , f_0 ($proxy_{1site}$ in Figure.IV.4). The next 6 ANN models investigate the combined influence of pairs of SCPs ($proxy_{1site}$ and $proxy_{2site}$ in Figure.IV.4). Finally, another set of 4 ANN models combining three SCPs as input parameters, and one ANN model accounting simultaneously for the four SCPs, are developed to provide an estimate of the maximum improvement (i.e. reduction of standard deviation of residuals) which may be reached with the four considered site proxies.

IV.3.2. Random-effect model

A procedure similar to the random-effects approach was then used to provide the between- and within-event σ , as described in [Derras et al. \(2014\)](#). For each of all the considered cases, the final ANN model is obtained using the maximum likelihood approach developed by Brillinger and Preisler (1985) and stabilized by [Abrahamson and Youngs \(1992\)](#). The performance of the ANN scheme is measured by the σ value classically used in GMPEs, which is decomposed into the between-event (τ) and within-event (ϕ) variabilities: both are zero-mean, independent, normally distributed random variables with standard deviations τ and ϕ ([Al Atik et al. 2010](#)). The between- and within-events residuals are assumed uncorrelated, so that the total σ at a period T of the ground-motions model can be calculated according to equation.IV.1.

$$\sigma(T) = \sqrt{\tau(T)^2 + \phi(T)^2} \quad (IV.1)$$

IV.4. Results

IV.4.1. Performance of SCPs in reducing the aleatory variability

In this section, we compare the various models derived for KiK-Net data set and analyze how the various SCPs reduce the ground-motion aleatory variability. The variations of τ , ϕ and σ versus period are displayed on Figure.IV. 5 for all ANN models: without SCPs, one-SCP and two-SCPs.

The Figure.IV.5 shows that the between-event variability (τ) is much lower than the within-event variability (ϕ), which is consistent with the vast majority of previous GMPE models results. The total variability (σ) is found identical at very short period ($T < 0.05$ s) whatever the ANN model: none of the SCP is really efficient at high frequency. The various variability components then increase from 0.05 to about 0.15 s, and then decrease significantly as period is increasing.

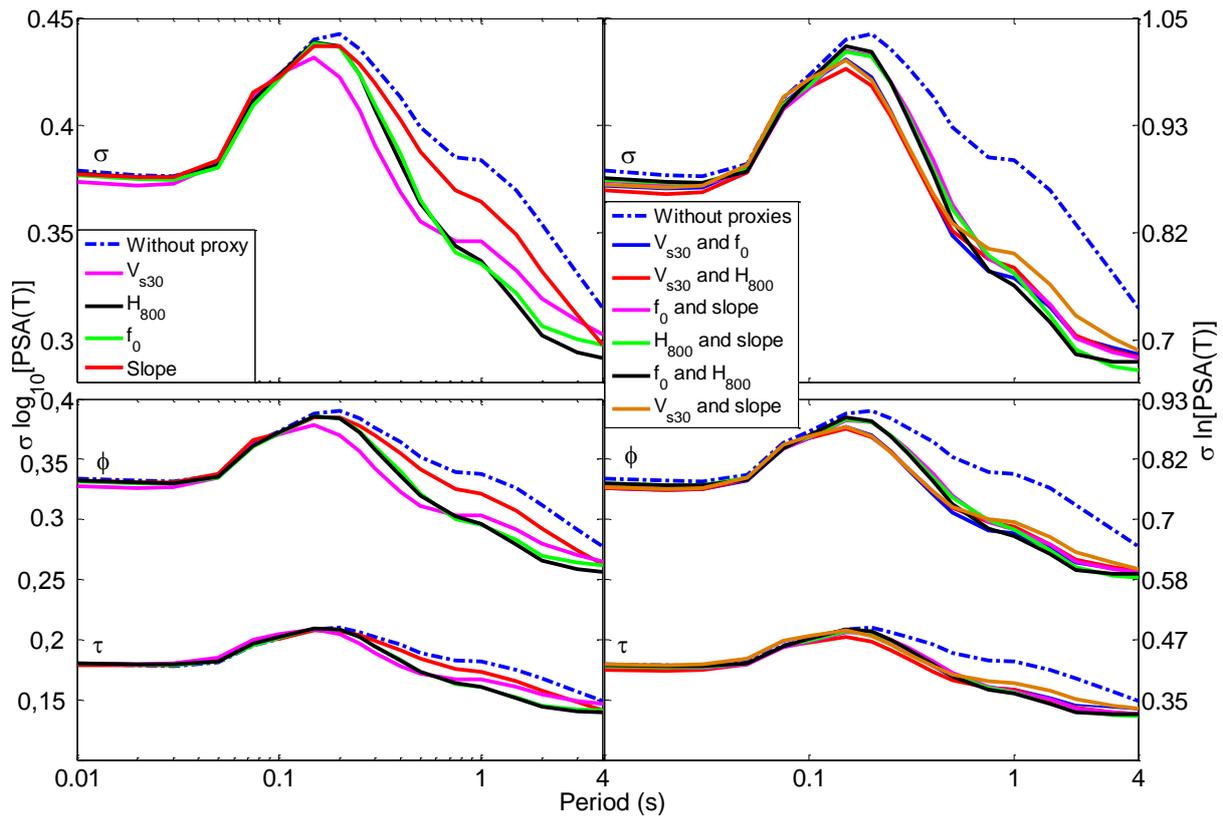


Figure IV.5 Sensitivity of the aleatory variability components (total σ , within-event, ϕ between-event, τ) to the site proxies used as input for the neural network. The left frame corresponds to use one-SCP as input. In right part, it is for two-SCPs. The σ , ϕ , τ standard deviation values are provide both on \log_{10} scale (left axis) and on \ln scale (right axis). They are provided for spectral ordinates for periods from 0.01 to 4 s. The M_w and R_{JB} are used to obtain the reference model (without proxies in this Figure for the one-SCP and two-SCPs).

A peak around 0.1 s has already been observed by some NGA GMPEs developers (e.g. [Chiu and Youngs 2014](#), [Derras et al. 2016](#)). A possible explanation is the interaction of varying stress drop with the high-frequency damping term (κ). At short to intermediate periods, i.e. for T between 0.1 and 0.7 s, one-SCP models allow reducing the within-event standard deviation compared to the reference model. The smallest ϕ is obtained for the V_{S30} SCP, followed by f_0 and H_{800} proxies which have comparable performance, while the slope proxy exhibits the poorest performance. At longer periods, f_0 and H_{800} provide the lowest ϕ values and perform better than V_{S30} . V_{S30} is thus confirmed to be relevant mainly for short to intermediate periods, as expected from the fact it samples only the shallow subsurface, while f_0 and H_{800} are more sensitive to the deep sediments and more relevant for long periods.

As expected, the two-SCP models lead to larger variance reductions, but it is interesting to notice that all pairs of proxies exhibit very similar performance. At short to intermediate periods, i.e. from 0.1 to 0.7 s, the [V_{S30} , slope] pair is found to provide the smallest values of ϕ , while for $T > 0.7$ s the "best" pair turns out to be [f_0 , H_{800}] as logically expected since both parameters are more sensitive to the bedrock depth. Interestingly enough, the [f_0 , slope] pair

exhibits a relatively good performance over the whole period range [0.1 – 4 s], while it is associated with the lowest measurement cost.

In addition, to better quantify the gains achieved by each SCP (s) model, the values of the variance reduction coefficients R_σ , R_ϕ and R_τ defined in Equation.IV.2 are presented in Figure.IV. 6. The variations of these coefficients versus period are displayed on Figure.IV.6 for all ANN models: without SCPs, one-SCP, two-SCPs, the best three-SCPs model and the single four-SCPs models. The reason for which we add these three and four SCPs cases is to obtain an estimate of the maximum possible variance reduction when many site parameters are known. The R_σ values are also listed in Table.IV.3 for a limited set of ground motion parameters (PGA, PGV and PSA at T=[0.2, 0.5, 1.0, 2.0] s).

$$\begin{cases} R_\sigma(T) = \left(1 - \frac{\sigma_{SCP}^2(T)}{\sigma_{withoutproxies}^2(T)} \right) \cdot 100; (\%) \\ R_\tau(T) = \left(1 - \frac{\tau_{SCP}^2(T)}{\tau_{withoutproxies}^2(T)} \right) \cdot 100; (\%) \\ R_\phi(T) = \left(1 - \frac{\phi_{SCP}^2(T)}{\phi_{withoutproxies}^2(T)} \right) \cdot 100; (\%) \end{cases} \quad (IV.2)$$

The obtained results confirm that the reduction of the aleatory variability becomes significant beyond $T = 0.1$ s and for PGV as well. Overall, the largest reduction is observed for the "reference" model accounting simultaneously for the four SCPs, followed by the best three SCPs model combining the use of [V_{S30} , f_0 , H_{800}]; it is noteworthy however that such "maximum possible" variance reduction does not exceed 7% for PGA and 4% for short periods around 0.08s, while it reaches 32% around 2 s. The values of these variance reduction coefficient confirm that no site proxy can be preferred over the whole frequency range. For the short periods [0.1 to 0.6 s] the best is V_{S30} – with a maximum of 21% variance reduction at 0.4 s -, while H_{800} and f_0 outperform in the [0.6, 4] s range with a peak variance reduction around 27 % at 2 s. The variance reduction obtained with the slope SCP is the lowest of all SCPs (except between 0.1 and 0.2 s), and reaches a maximum of 12% at 2 s. Similarly, the "two-SCPs" models exhibit a slightly larger variance reduction at short to intermediate period when they include V_{S30} as one of the two site proxies (the best performance being achieved by the [V_{S30} , H_{800}] pair), while the largest reduction at long period is observed for the [f_0 , H_{800}] pair, i.e. a combination of two long-period proxies, with a value of R_σ reaching 31.15 % at $T = 2.0$ s.

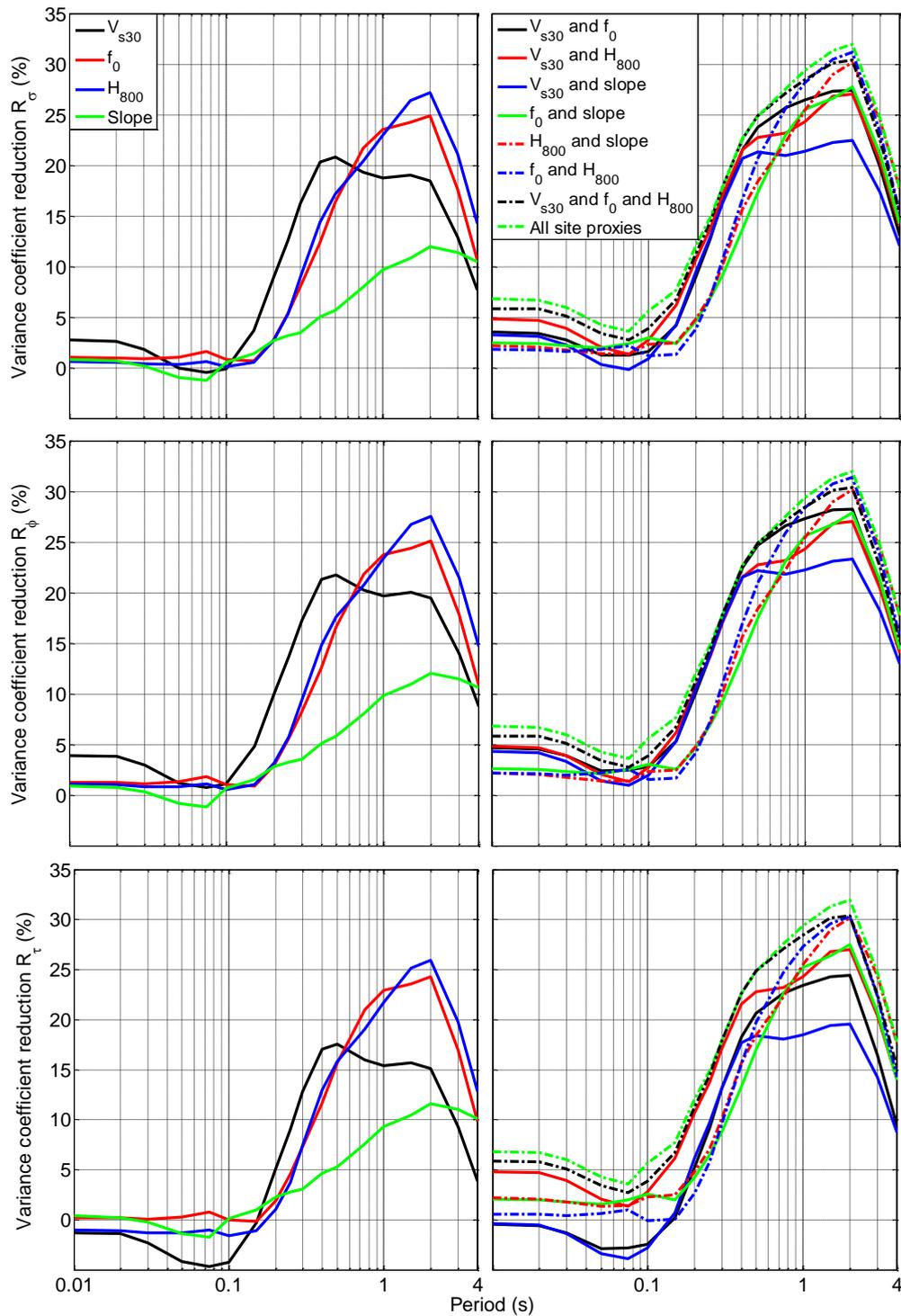


Figure.IV.6 Performance of various SCPs in reducing ground-motion aleatory variability. The top frames display the variance coefficient reduction (R_σ) for the total standard deviation σ . The middle frames illustrate the variance coefficient reduction (R_ϕ) of the within-event standard deviation ϕ . The bottom frames show the variance coefficient reduction (R_τ) of the between-event standard deviation τ . In the left frames (for top, middle and bottom frames), we use 4 ANN models with one-SCP measured V_{S30} or f_0 , H_{800} and slope. In the right frames (for top, middle and bottom frames), we show R_σ , R_ϕ and R_τ for the 8 ANN models with two-SCPs, three-SCPs and four-SCPs configurations.

It is worth noticing that site proxies also influence the between-event standard deviation τ in a very similar way they affect the within-event variability: a better site description enables a better

description of the actual dependence of the dependence on source and path parameters. It may also indicate that despite the random effects procedure, the within- and between-event variabilities are not completely independent

Another parameter used in ANN approach to measure the relevancy of each explanatory variable (and therefore of each single SCP or SCP pair) is the total percentage of synaptic weights P . These parameters P are a measure of the relative, overall importance of the individual explanatory variables, averaged for all the output ground motion parameters (thus, here, over the whole frequency range 0.01 – 4 s). They have been computed according to the procedure detailed in Derras et al. (2014, Equation.IV.4), for the 16 ANN models. Tables.IV.4 and 5 list the P values (in %) for each input variable. As expected from the data distribution, the most efficient parameter in reducing the variance of response spectra is the R_{JB} distance (synaptic weight around 41-50%), followed directly by the earthquake magnitude M_w (around 24-32 %). The P values associated to the site term ranges from 7% to 22%.

Table.IV.3 Variation reduction of the variance (R_σ) in % for the 12 ANNs types for PGA, PGV and at $T=[0.2,0.5,1.0,2.0]$ s

Ground-motion parameters ($\ln(\sigma_{i,j})$)	One-SCP				Two-SCPs				Three-SCPs	Four-SCPs		
	V_{S30}	f_0	H_{800}	Slope	(V_{S30}, f_0)	(V_{S30}, H_{800})	(f_0, slope)	(H_{800}, slope)	(f_0, H_{800})	(V_{S30}, slope)	$V_{S30}, f_0, H_{800})$	$V_{S30}, f_0, H_{800}, \text{slope})$
PGA	2.70	1.02	0.58	0.72	3.46	4.77	2.44	2.12	1.79	3.20	5.88	6.78
PGV	15.57	10.90	11.54	6.66	17.04	16.80	13.30	14.08	13.85	17.16	18.16	19.27
0.2 s	8.91	2.73	2.68	2.68	8.99	10.69	4.59	4.80	3.81	9.50	11.24	12.01
0.5 s	20.82	16.39	17.20	5.70	23.78	22.76	17.36	18.37	20.67	21.32	24.88	24.82
1 s	18.74	23.55	22.99	9.68	26.47	24.32	25.55	25.47	28.16	21.38	28.43	29.39
2 s	18.50	24.91	27.15	11.95	27.40	27.00	27.78	30.16	31.15	22.45	30.38	31.95

When only one SCP is considered, the largest SCP weights correspond to H_{800} (around 14 %) and V_{S30} (\cong 13 %) while the smallest corresponds to the slope ($P_{\text{slope}} = 7\%$). For the twin-SCPs models, the best pair is (V_{S30}, H_{800}) with $P = 22\%$. The $[f_0, \text{slope}]$ pair also performs well with $P=18\%$. This ranking is similar to the ranking obtained from the analysis of aleatory variabilities discussed above if we consider the whole period range.

Table.IV.4 Sensitivity of ground motion models to magnitude, distance and one-SCP, as expressed through the total percentages of synaptic weight (P , %) corresponding to each input parameter.

SCP	P_{RJB}	P_{Mw}	P_{Depth}	P_{SCP}
H₈₀₀	47	28	11	14
V_{S30}	47	29	11	13
f₀	49	28	12	10
Slope	50	32	11	7

Table.IV.5 Sensitivity of ground motion models to magnitude, distance, depth and two-SCPs, as expressed through the total percentages of synaptic weight (P , %) corresponding to each input parameter.

SCPs	P_{RJB}	P_{Mw}	P_{Depth}	P_{SCP1}	P_{SCP2}	$\sum P_{SCPs}$		
H₈₀₀, V_{S30}	41	24	13	12	10	22		
f₀, slope	44	25	13	15	3	18		
V_{S30}, f₀	44	26	14	10	7	17		
H₈₀₀, slope	44	29	11	12	4	16		
V_{S30}, slope	43	27	13	12	4	16		
H₈₀₀, f₀	44	27	14	9	7	16		
H₈₀₀, V_{S30}, f₀, slope	46	26	3	P_{H800} 8	P_{VS30} 9	P_{f0} 6	P_{Slope} 2	$\sum P_{SCPs}$ 25

As in the aleatory variability analysis described above, the “all proxies” model is considered for comparison. The total synaptic weight of SCPs reaches 25%, which decomposes in individual synaptic weights for each SCP ranking as for the synaptic weight of one-SCP models: the largest one is P_{VS30} , followed by P_{H800} , the poorest one is associated to the slope. When the number of SCP increases, the increase of the SCP weight is associated first with a decrease of the magnitude and RJB weights (from one-SCP to two-SCPs), and then to a drop of the relative importance of the focal depth (from two-SCPs to four-SCPs, for which P_{depth} becomes as low as 3 %): the importance of focal depth as a predictor variable gets lower and lower as the site is better described, while R_{JB} and M_w obviously remain key parameters.

IV.4.2. Impact of the various SCPs on median ground-motion models

As discussed above, the nature of SCP has a noticeable effect on the ground motion aleatory variability. We investigate here their impact on the median estimates, through a comparison of the 4 one-SCP ANN and 6 two-SCP ANN models. Figure.IV.7 displays the distance dependence of the spectral acceleration for $T=0.0, 0.2$ and 1.0 s, for the median magnitude ($M_w=5.5$), the median focal depth (10 km), and the median values of the various SCPs (i.e., $V_{S30}=460$ m/s, $f_0=4$ Hz, $H_{800}=30$ m and slope=0.04 m/m, as derived from Figure.IV.2 and Table.IV.2). Through Figure.IV.7a, we remark that the site proxy type ($V_{S30}/f_0/H_{800}/slope$) is

observed not to have any significant impact on median predictions for $R_{JB} > 30$ km (the distance beyond which the data may be considered sufficient for median site conditions, see Figure.IV.1), and a relatively limited one (up to a factor of 2) at shorter distances (gray zone, $R_{JB} < 30$ km), where the scatter may be reasonably assumed to come mainly from the small number of data.

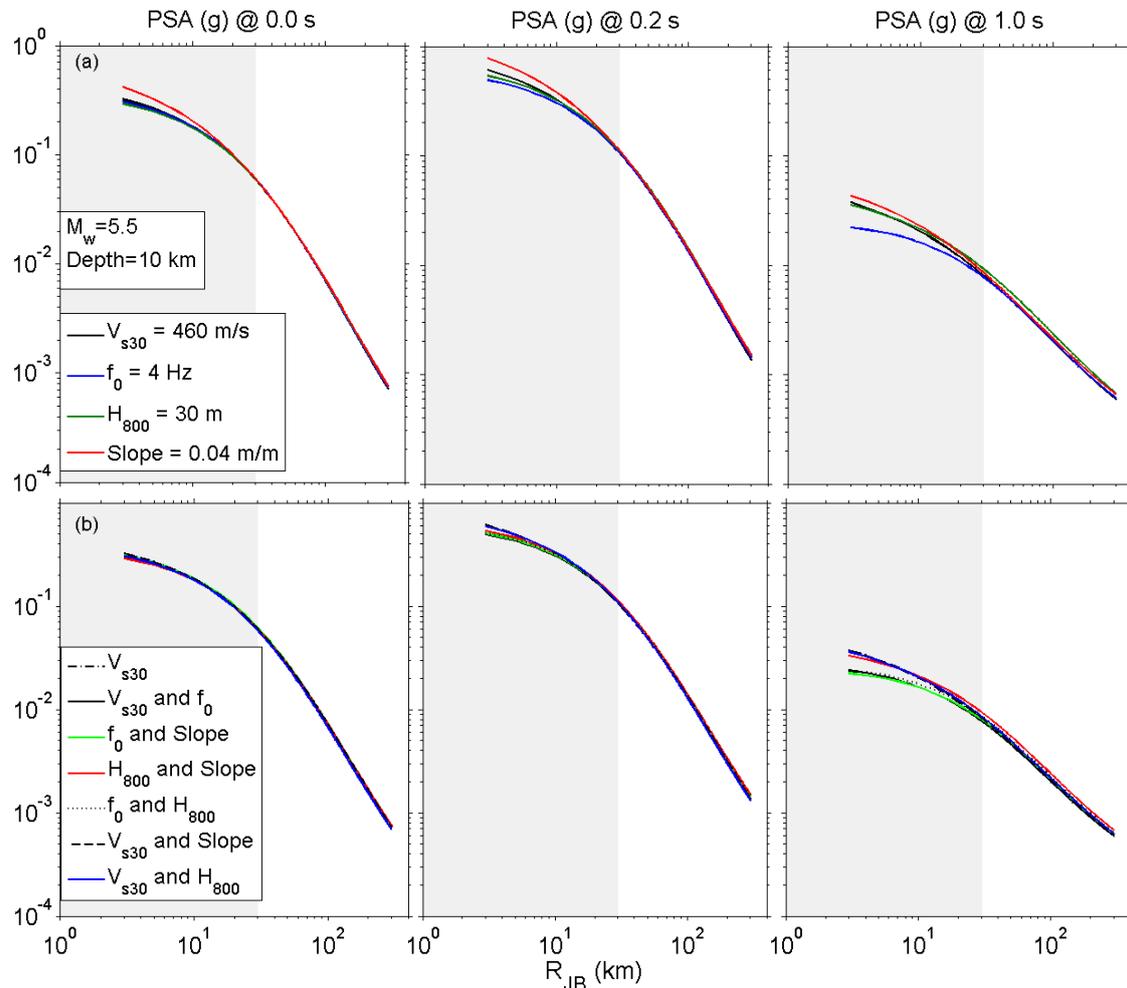


Figure.IV.7 Strength of the median ground motion predicted by different models. Effect of the one-SCP proxy (top, $V_{s30}/f_0/H_{800}/\text{slope}$) and the two-SCPs in the bottom row. The plots display the distance dependence of the spectral accelerations at three periods ($T=0.0$ s, left column; $T=0.2$ s, middle column; and $T=1.0$ s, right column) for a $M_w = 5.5$ and depth=10 km event, at sites with $V_{s30} = 460$ m/s, $f_0 = 4$ Hz, $H_{800} = 30$ m and slope=0.04 m/m (taken from Table.IV.2: 10 % and 90 % CDF). The grey zone ($R_{JB} \leq 30$ km) should be considered less reliable since there are very few stiff site recording in this distance range (see Figure.IV.1)

The two-SCP models (Figure.IV.7b) lead to similar results. All pairs of proxies exhibit very similar median predictions especially at short periods ($T=0.0$ s and $T=0.2$ s). It is however interesting to notice that, for $T = 1.0$ s at close distances (gray zone) the three pairs of proxies including the fundamental frequency lead to slightly lower predictions than the other two-SCP models. Furthermore, the comparison with the V_{s30} -SCP model highlights the fact that the type and the number (one-SCP or two-SCP) of site proxies have no influence on this median.

IV.4.3. Ground motions for "Extreme" values of site proxies

A complementary information is provided by the amount of difference in predictions for "extreme" values of the SCP. Figures.IV.8 and 9 display the "soft/stiff" Spectral Ratio (SR) for various periods ($T=0.0$ s, 0.2 s and 1.0 s: a consistent definition of "soft" and "stiff" sites was taken for all SCPs, simply by considering the SCP values corresponding to 10 % and 90 % of the CDF distributions shown in Figure.IV.2 and listed in Table.IV.2; note however that given the negative correlation between H_{800} and other proxies, the 10% fractile of H_{800} (i.e., 6 m) has been associated with the 90% fractile of V_{S30} , f_0 and slope (i.e., 800 m/s, 11 Hz and 0.13 m/m, respectively) , and vice-versa (i.e. 120 m, 270 m/s, 0.4 Hz and 0.01 m/m, respectively). Figures.IV.8 and 9 display such SR for the individual SCP (top) and two-SCPs (bottom) cases. SR has the following form:

$$SR(T) = \frac{PSA_{soft}(T)}{PSA_{stiff}(T)} \quad (IV.3)$$

Figure.IV.8 shows the sensitivity of the SR amplification factors to R_{JB} distance, at three different spectral periods (0.0, 0.2s and 1s), and for a given earthquake scenario ($M_w=6$, depth=10 km). Besides the trend of site amplification to increase with distance up to 100 km – which is related to nonlinear site response as the loading level is decreasing with increasing distance -, the site amplification is found to increase with period, as classically found in most GMPEs. Beyond 100 km the SR "saturates" at an almost constant value, which can be considered to correspond to the linear site response. The curves are displayed down to distances as short as 3 km, but they should be considered reliable only for R_{JB} values larger than 30 km (gray area), since the derived models can hardly be considered reliable for stiff sites at lower distances (see Figure.IV.1). V_{S30} and slope SCPs are found to provide the largest amplification at short periods (which remains however smaller than 60%). The situation is opposite at long period ($T = 1.0$ s) where the SCP providing the largest amplification is f_0 with amplification ranging from 2.8 to 3.2. In addition, it is clear that the site amplification predicted with the slope proxy is not very sensitive to the oscillator period. Another interesting result is that the combination of two proxies significantly increases the "soft/stiff" SR values: the amplification increase ranges from 15 % (V_{S30} to V_{S30} -slope) at short period, to 17% at intermediate periods, and 21% at long period (f_0 to f_0 - V_{S30} or f_0 -slope). The most probable explanation comes from the fact that simultaneously matching 10% and 90% fractiles for a pair of proxies corresponds to less frequent combinations, with more differentiated site conditions, than for a single proxy (as can be seen also on Figure.IV.3).

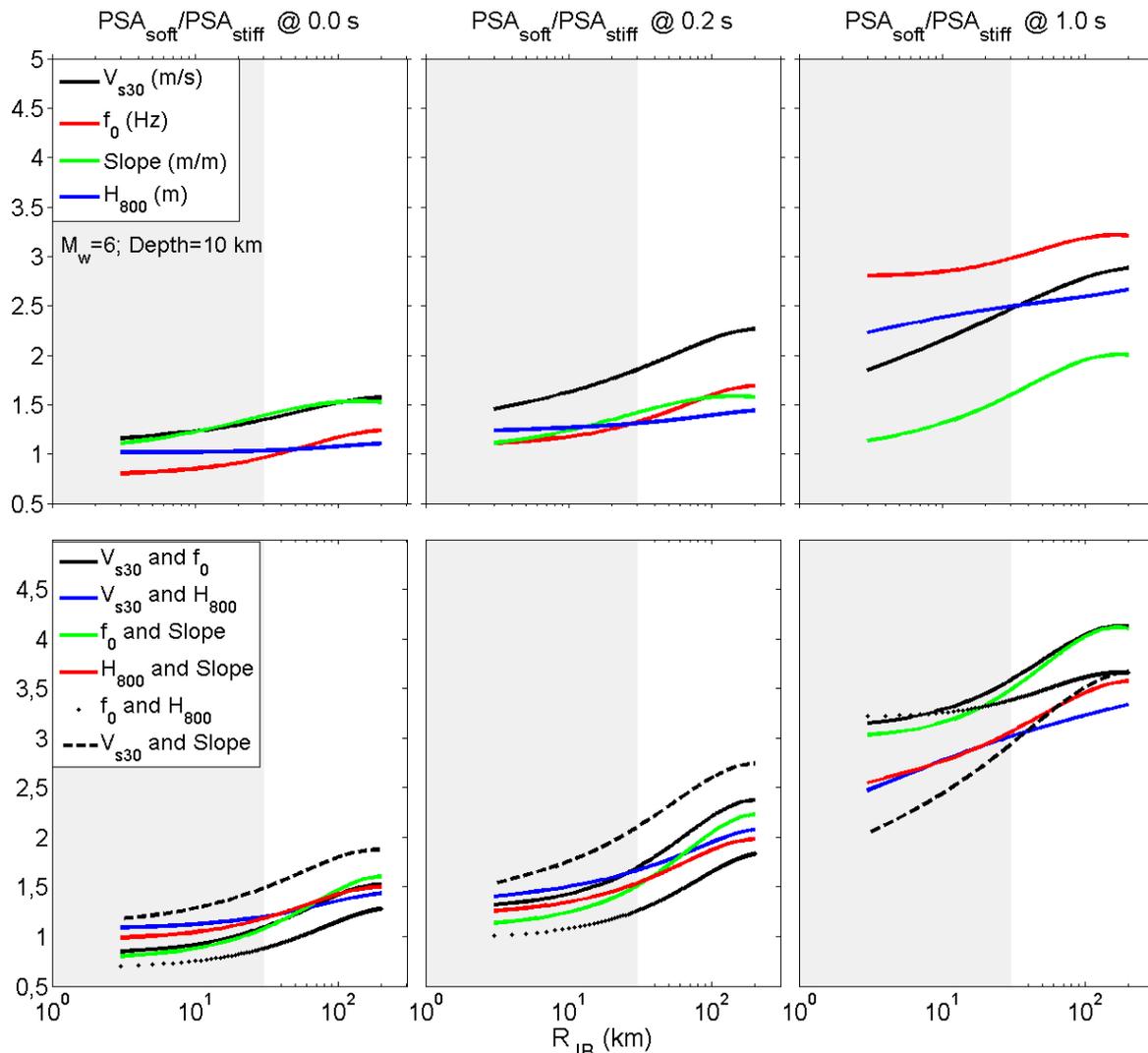


Figure.IV.8 Site amplification factor (SR: equation.IV.3) versus R_{JB} at $T=0.0, 0.2, 1.0$ s for $M_w=6$ and focal depth=10 km. We consider one-SCP (top) and two-SCPs (bottom). The grey zone ($R_{JB} \leq 30$ km) should be considered less reliable since there are very few stiff site recording in this distance range (see Figure.IV.1)

Figure.IV.9 illustrates the SR amplification factors variation versus PSA_{stiff} at $T=0.0$ s (i.e. PGA_{stiff} , one of the reference parameters that is commonly used in GMPEs to describe the dependence of the non-linear site amplification on the loading level. These curves have been established here by considering a given R_{JB} distance (30 km), a given focal depth (10 km), and a magnitude varying from 5 to 7. A closer look at the dependence of soft / stiff amplification factor does indicate a larger amplification level for small stiff motion levels, associated to a significant non-linearity (i.e., decrease of amplification with increasing loading level at the underlying bedrock). The curves displayed in Figure.IV.8 call for several comments:

- i. The amount of non-linearity depends both on the considered site proxies, and on the oscillator period.

- ii. Whatever the site proxy, a significant non-linearity can be observed at long period ($T = 1$ s), which is a somewhat unexpected result.
- iii. Among single proxy models, the one using the slope predicts similar non-linearities whatever the oscillator period, while the "long period" proxies f_0 and H_{800} are those who do not predict any significant non-linearity at short period: the predicted SR is around 1 at $T=0.0$ s, and around 1.4 around $T=0.2$ s, in both cases whatever the loading level. On the contrary, at long period ($T = 1.0$ s), the SCP providing the largest SR is f_0 , while amplification levels and their non-linear sensitivity on PGA_{stiff} are similar for V_{S30} and H_{800} proxies. This larger amplification factors for the f_0 model at $T=1.0$ s might be related to the fact that the "soft" site is characterized by a fundamental frequency of 0.4 Hz: the oscillator frequency (1 Hz) is always larger than the fundamental frequency, and one may thus expect to be systematically in the amplified frequency range, while sites with $V_{S30} = 270$ m/s or $H_{800} = 120$ m, with fundamental frequencies above 1Hz (see last column of Figure.IV.3) do exist. Correlatively, a larger reduction of the amplification with increasing loading level may be expected if the non-linear behavior affects the whole thickness of the soil deposit (see [Régnier et al. 2016](#)).
- iv. Similar observations can be done for the results with two-SCPs models. At short period ($T= 0$ and 0.2 s), the largest amplification and non-linearity is predicted when using the pair of short period proxies ($V_{S30} - \text{slope}$), while the smallest corresponds to the pair of long- period proxies ($f_0 - H_{800}$). At long period ($T= 1.0$ s), the predicted amplifications and their nonlinear component are less scattered than in the one-SCP case, the pairs including the f_0 proxy predicting however slightly larger amplifications.

These results are however partial, and should not be extrapolated too fast, as they correspond to a specific distance (and focal depth), and use the stiff-site PGA to characterize the loading level. Figures. IV.10 to 13 are thus intended to check the robustness of the results presented in Figure.IV.9, considering also other descriptions of the loading level, and other distance scenario. Only one-SCP models are considered, successively V_{S30} , f_0 , H_{800} and slope for Figures.IV 10 to 13. In each case, three different distances are considered (30, 50 and 75 km, i.e., in the range where there exist enough data within the [10% - 90%] fractile range of each SCP, and the variation of the loading level for each distance corresponds to the predictions over the magnitude range [5, 7]. Three different parameters are considered to characterize the loading level: the PGA on rock or "stiff" site as defined according to the selected site proxy, the spectral acceleration on the same rock or "stiff" site at the oscillator period considered, and

finally an estimate of the actual strain at the site: several authors ([Idriss, 2011](#); Chandra et al. [2015, 2016](#); [Gueguen, 2016](#)) proposed to use the ratio PGV/V_{S30} as a proxy to the shear strain, where PGV is the peak velocity at the site, and it has thus been tested in the present study.

In principle, if a loading parameter is relevant for nonlinear behavior, the dependency of site amplification as a function of this loading parameter should exhibit only a marginal dependency on other parameters such as magnitude, or distance, or frequency contents. Analyzing Figures.IV.10 to 13 according to this criterion clearly indicates that the lowest scatter is observed among distance and magnitude scenarios for the loading parameter "PSA_{stiff}(T)" (second row), while the largest corresponds to "PGA_{stiff}", especially for the long period site amplification.

On the light of those results, it turns out that the best ground motion parameter to be used for the characterization of the loading level in the nonlinear site amplification term of GMPEs is the spectral ordinate on rock at the considered period; the strain proxy PGV/V_{S30} may however constitute a satisfactory, alternative choice. Another major outcome of this section is the variability of the nonlinear behavior according to the site proxy selected for the GMPEs: short period non-linearity is observed preferably with short period proxies (V_{S30} and slope) and disappears when using H_{800} .

This analysis supports the use of at least two site proxies (e.g. V_{S30} - H_{800}) to predict ground-amplifications on basins. Figure.IV.14 shows the amplifications predicted by various soft site configurations ($V_{S30}=270$ m/s, $H=6$ m (10 % of the CDF), 25 m (50% of the CDF) and 120 m (90% of the CDF) and a stiff site characterized by $V_{S30}=800$ m/s (90% of the CDF) and $H_{800}=6$ m (10% of the CDF). These amplifications have been established for three R_{JB} distance (30-50-75 km), a given focal depth (10 km) and two earthquakes scenario (magnitude 5 and 6.5).

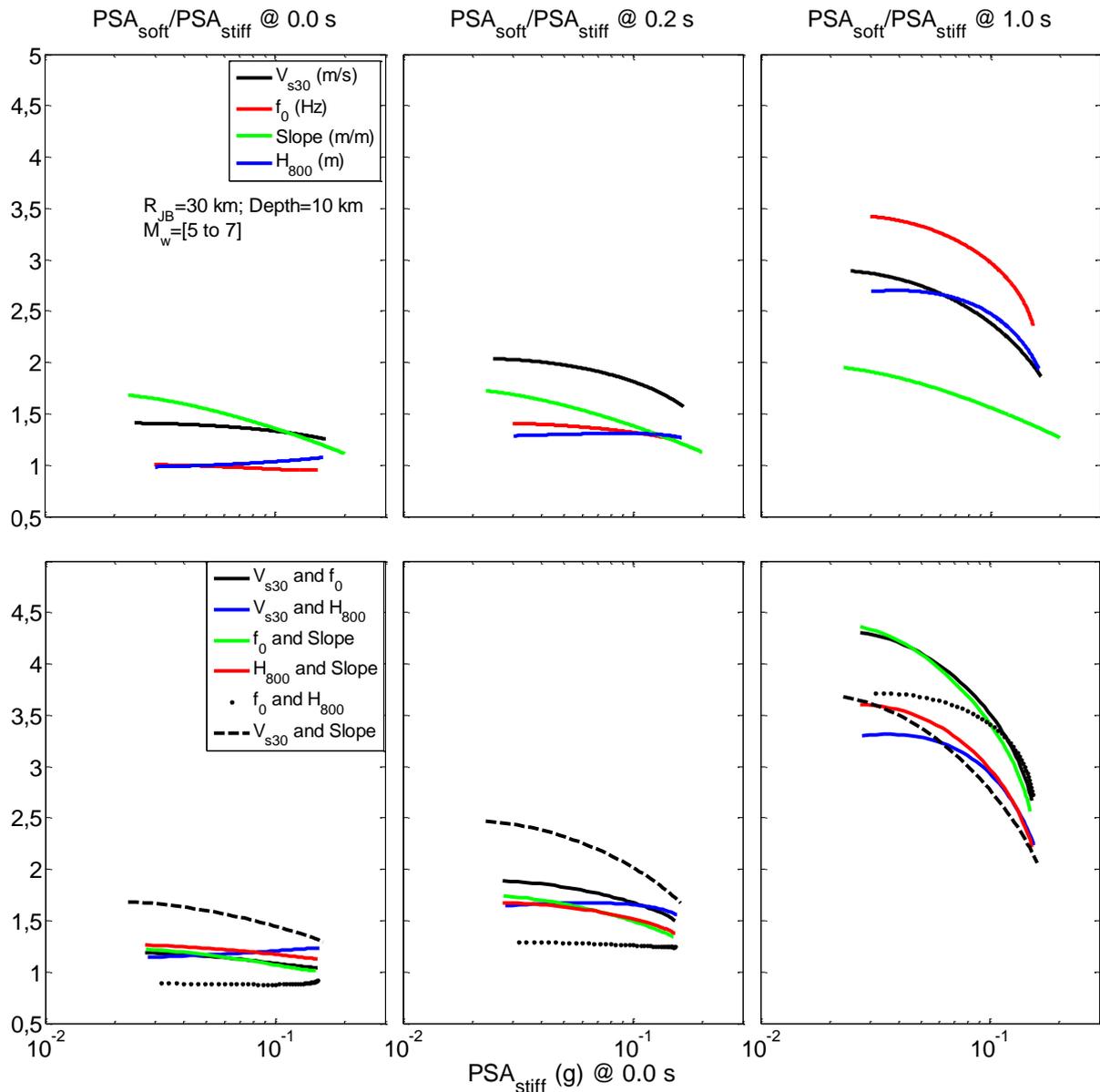


Figure.IV.9 Illustration of the non-linearity in the site term. Site amplification factor (SR: equation.IV.3) versus PGA_{stiff} (90 % of CDF) at $T=0.0, 0.2, 1.0$ s for $R_{JB}=30$ km, focal depth=10 km and $M_w=[5$ to $7]$. We consider one-SCP (top) and two-SCPs (bottom).

The results, as expected, show that the peak and period ranges of the amplifications are depending on the basin depths. These results also illustrate the nonlinearity effect: lower amplifications are observed at high frequencies ($T < 0.1$) for the largest magnitude ($M=6.5$) and the lowest distance ($R_{JB} < 30$ km). Almost all predicted amplifications are however higher than one, which means that our empirical models does not predict nonlinear soil de-amplifications.

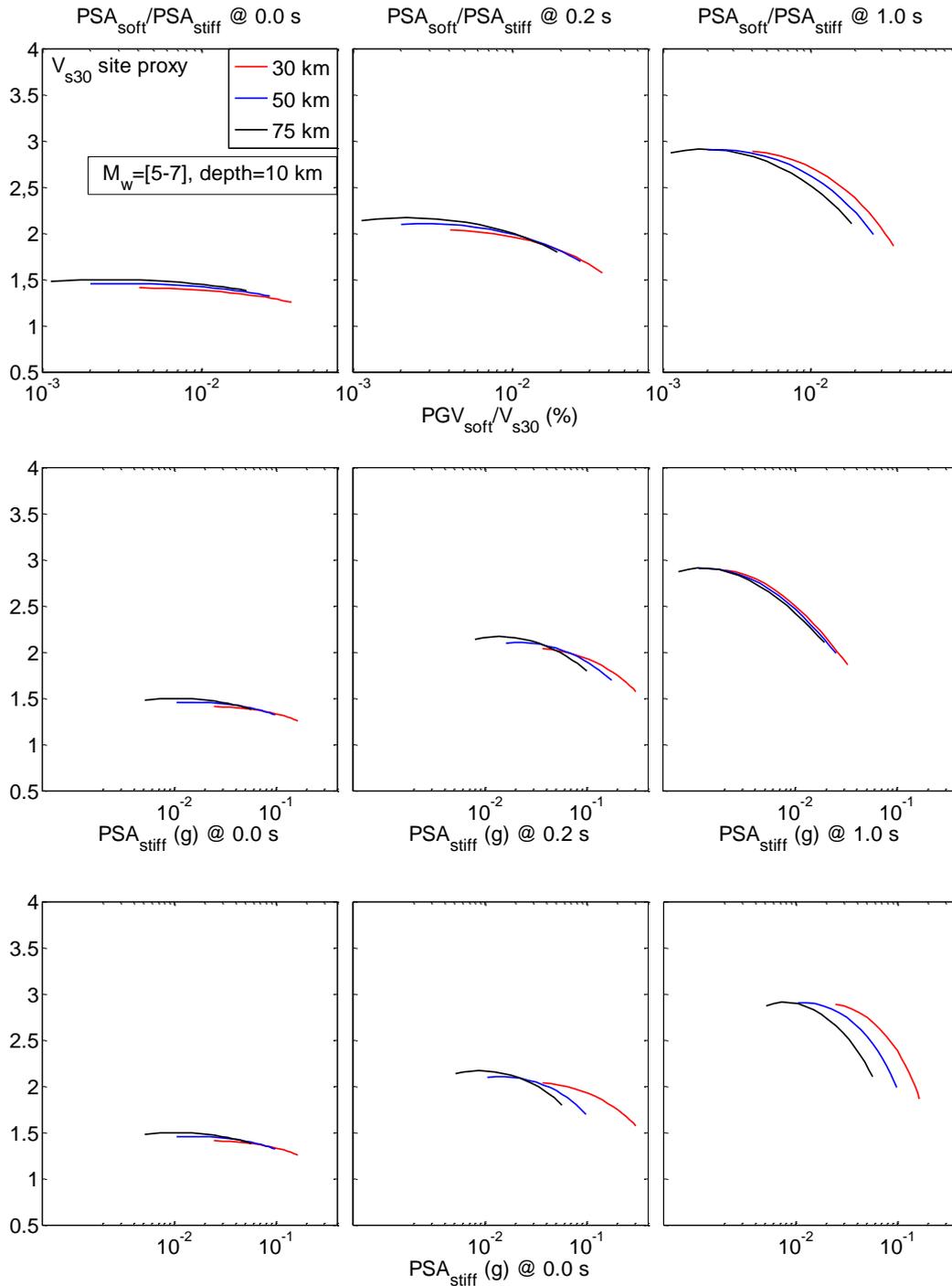


Figure.IV.10 Comparison of the NL components of site amplification for different description of the loading level, for the one-SCP ANN model using V_{S30} . Each panel represents the (soft / stiff) site amplification at a given oscillator period as a function of a loading parameter. The three columns correspond to three different oscillator periods (0, 0.2 and 1.0s). The three rows correspond to different choices for the loading level: strain proxy PGV_{soft}/V_{S30} on top, spectral ordinate at the stiff site $PSA_{stiff}(T)$ in the middle, and PSA_{stiff} at the 0.0 s (bottom). As in Figure.IV.9, the site amplification factors (SR : equation.IV.3) are calculated for three distances: 30 km, 50 km and 75 km, for which the variation of loading level corresponds to magnitude variations between 5 and 7 (for a fixed focal depth equal to 10 km).

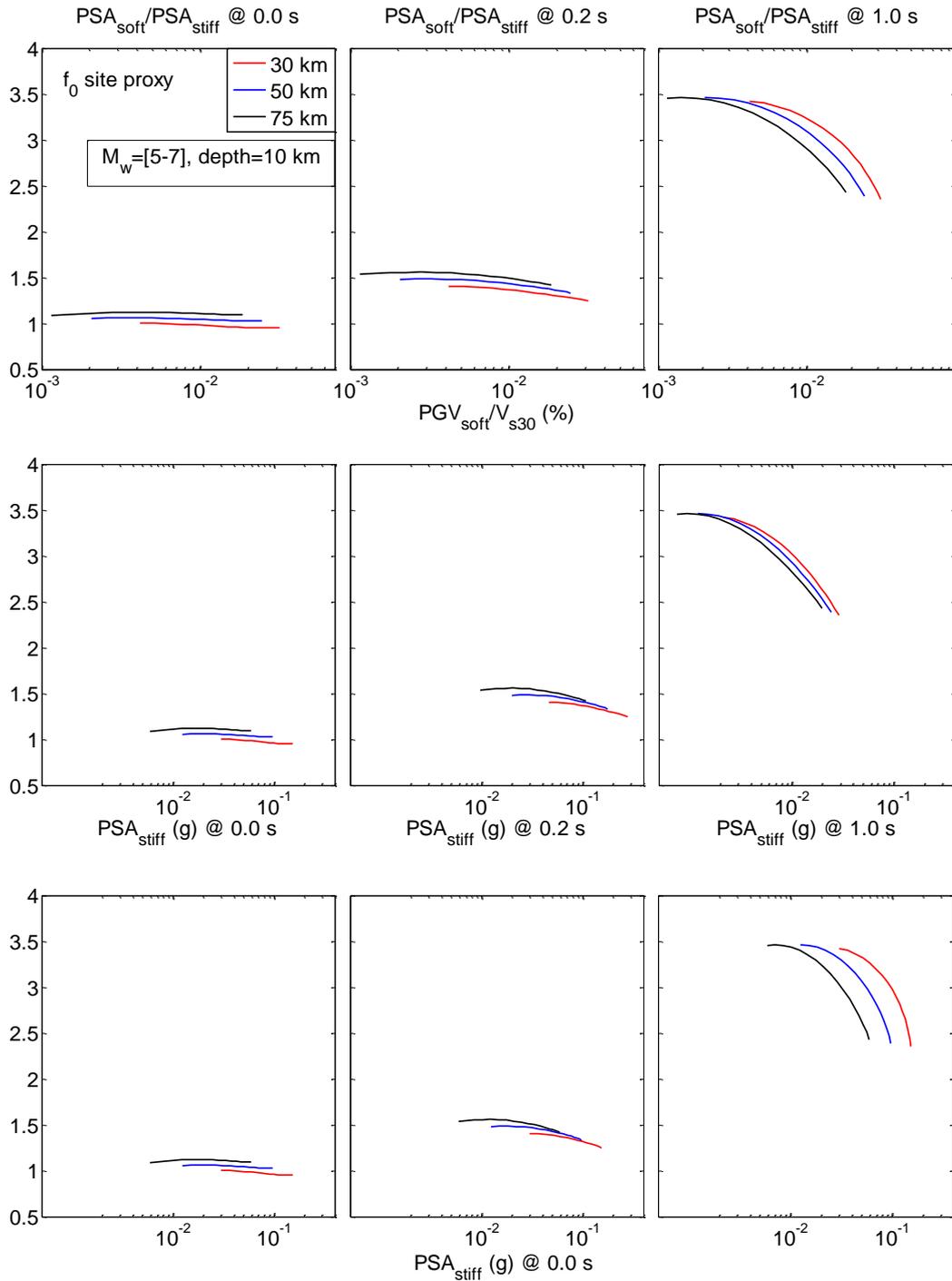


Figure.IV.11 Similar to Figure.IV.10, but for the one-SCP model-using f_0 as site condition proxy. Stiff site corresponds to $f_0 = 11$ Hz, soft site to $f_0 = 0.4$ Hz.

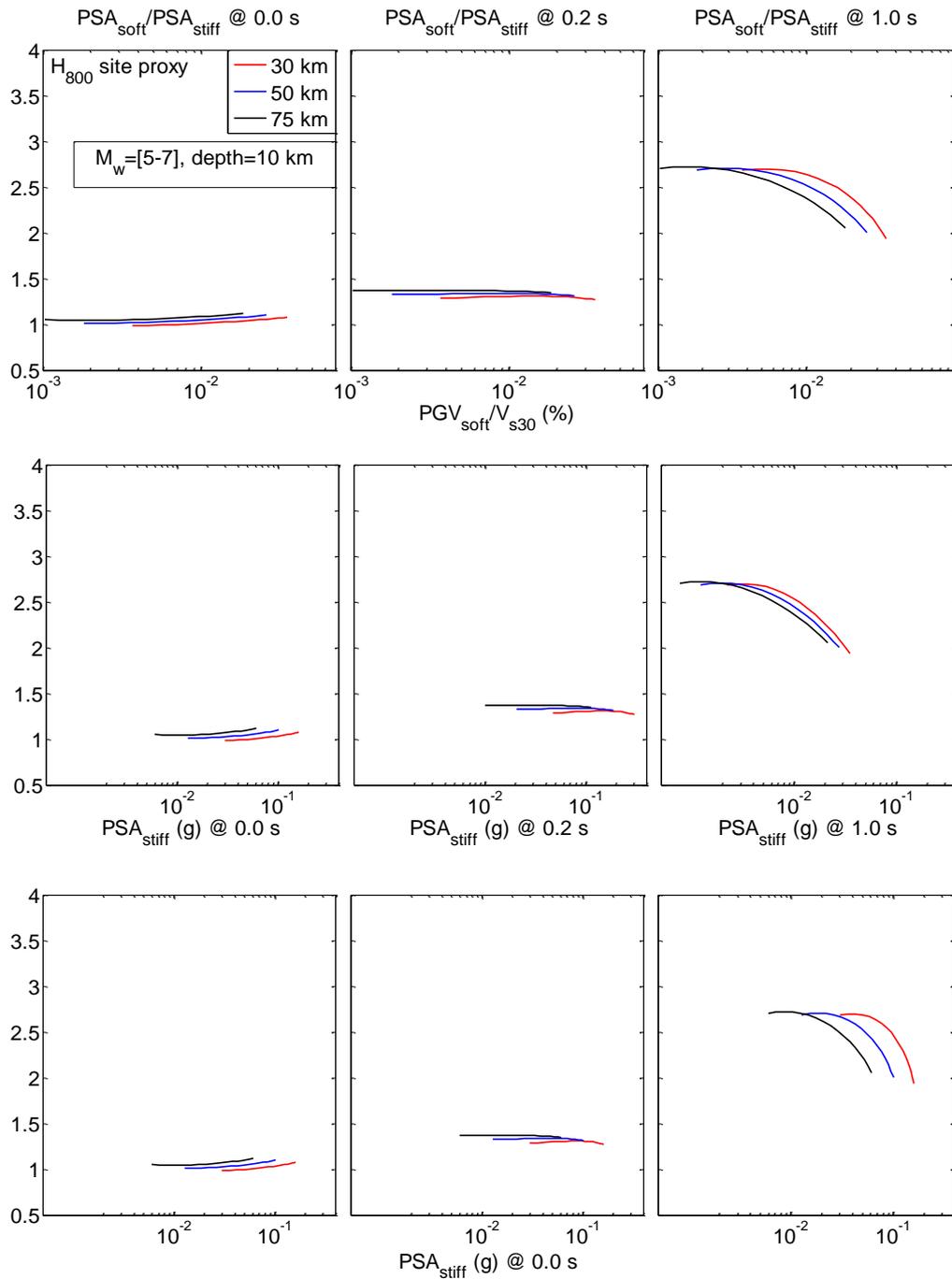


Figure.IV.12 Similar to Figure.IV.10, but for the one-SCP model-using H_{800} as site condition proxy. Stiff site corresponds to $H_{800} = 6$ m, soft site to $H_{800} = 120$ m.

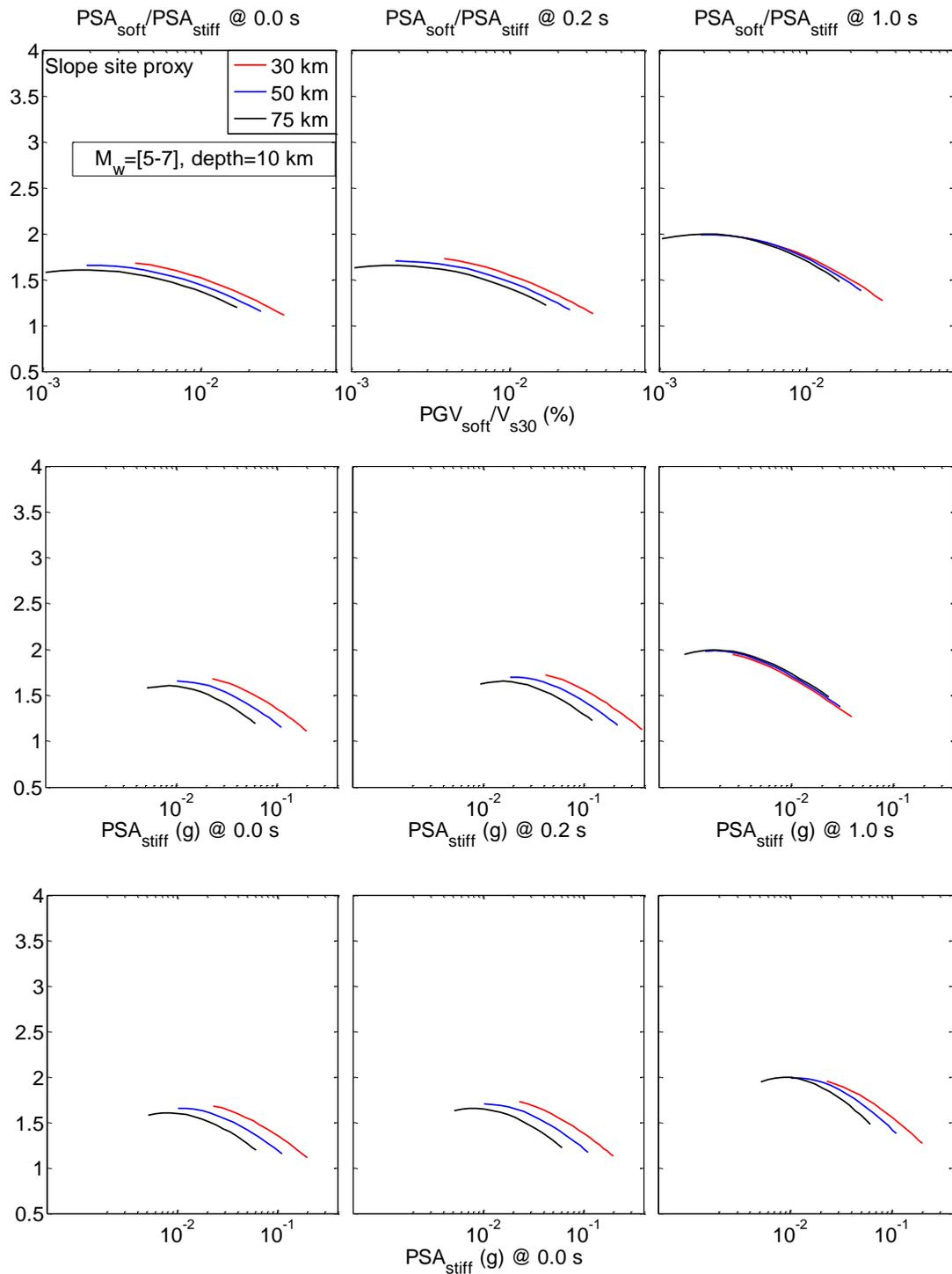


Figure.IV.13 Similar to Figure.IV.10, but for the one-SCP model-using slope as site condition proxy. Stiff site corresponds to slope = 0.13 m/m, soft site to slope = 0.01 m/m.

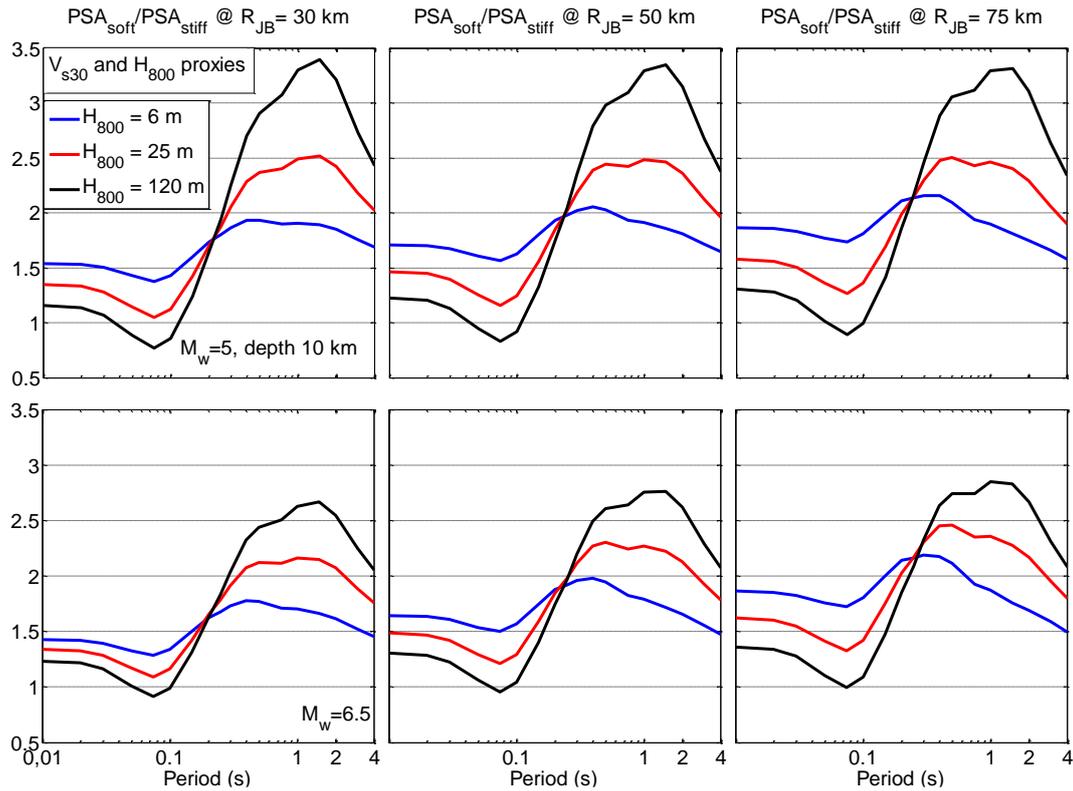


Figure IV.14 Site amplification factor versus period for $R_{JB}=30, 50, 75$ km and for $M_w=5$ (top) and 6.5 (bottom) and a focal depth=10 km. We consider only the site proxy V_{s30} - H_{800} ($V_{s30}=270$ m/s and $H_{800}=6, 25$ and 120 m) to represent the soft site configurations. The stiff site is characterized by $V_{s30}=800$ m/s and $H_{800}=6$ m.

IV.5. Summary and conclusions

The application of neural network approach to a KiK-net data set offered the possibility to test the performance of various site condition proxies to reduce the aleatory variability in ground motion prediction equations. The four available SCPs are V_{S30} and H_{800} (both derived from downhole measurements), f_0 (the fundamental frequency derived from H/V ratios and surface/downhole spectral ratios), and the slope derived from DEM data, which has been proposed as a proxy to V_{S30} values. A total of 16 neural network models were derived to describe the dependence of response spectra ordinates on moment magnitude M_w , Joyner & Boore distance R_{JB} , focal depth D and various combinations of SCPs : one without any SCP which provides the "reference case", 4 with each single SCP, 6 with the 6 possible pairs of SCPs [V_{S30} - f_0], [V_{S30} - H_{800}], [f_0 -slope], [H_{800} -slope], [V_{S30} -slope] and [f_0 - H_{800}], four with the four possible combinations of three SCPs, and one with all SCPs considered simultaneously.

When only one-SCP is used, the largest reduction of aleatory variability with respect to the "reference case" is found to be provided by V_{S30} at short to intermediate periods ($T \leq 0.6$ s), and by f_0 or H_{800} at longer periods. Amongst the four SCPs, the parameter "slope" is thus found to provide the worst performance when considered alone. However, when SCP pairs are considered, comparable performance is found whatever the pair of proxies. In particular, the

"best pairs" are found to be $[V_{S30}-H_{800}]$ at short periods and $[f_0-H_{800}]$ at long periods, while the "low cost" pair $[f_0-slope]$ provides a good compromise over the whole period range $[0.1 - 4 \text{ s}]$. None of the four tested SCP is thus "optimal" over the whole period range and all proxies show a poor contribution at high frequencies ($>5\text{Hz}$).

Otherwise, the site proxy type (slope/ $V_{S30}/H_{800}/f_0$) has no influence on this median. These results indicate that the type of SCP does not really affect the median.

Regarding site amplification, V_{S30} and slope SCPs are found to provide the largest differentiation at short periods (0, 0.2 s). At long period H_{800} and f_0 are providing the largest differentiation.

We showed also on this subset of KiK-net data that the soft to stiff site amplification exhibits a significant non-linearity, the characteristics of which are however tightly linked to the used proxy, and the parameter selected to describe the loading level. The most relevant loading parameter is found to be the spectral acceleration on rock (or "stiff" site) at the considered period, and the worst the rock (or "stiff") peak acceleration, with a satisfactory behavior for the strain proxy PGV/V_{S30} . Non-linearities are found to be systematically larger at intermediate period (1 s) than at short period (0.2, 0 s), while short-period non-linearities are better described by short period site proxies.

Otherwise, the best ground motion parameter to be used for the characterization of the loading level in the nonlinear site amplification term of GMPEs is the spectral ordinate on rock (or "stiff" site) at the considered period.

Another important result is the variability of the nonlinear according to the SCP: short period non-linearity is observed preferably with short period proxies (V_{S30} and slope) and disappears when using H_{800} .

As these results have been obtained on a specific - though large - data set (subset of KiK-net data, thus probably lacking of very soft sites), they should of course be tested for other data sets; the four SCPs are however rarely available simultaneously.

Most of the scientific and technical work has been carried out by the first author, under the scientific supervision of the two other coauthors. The redaction has been shared amongst the three authors.

IV.6. Data and resources

PSA, V_{S30} and H_{800} used in this study were collected from the KiK-net web site <https://datacenterhub.org/resources/272>. Slope have been collected and disseminated by the "The Pacific Earthquake Engineering Research Center" and f_0 from ([Régnier et al. 2013](#)).

IV.7. Acknowledgements

The authors thank Julie Régnier and Héloïse Cadet for their generous help and which provided us the f_0 of the KiK-net database. We acknowledge the support from the Sinaps project and from Haitham Dawood for his help.

IV.8. Principaux résultats de la 1^{ère} partie

Pour clôturer cette première partie, nous donnons les conclusions suivantes :

- Lorsqu'un seul proxy de site est utilisé, la plus grande réduction de la variabilité aléatoire est donnée par V_{s30} (à des courtes et moyennes périodes) et par f_0 et H_{800} à longues périodes.
- Parmi les quatre proxys de site, la pente topographique est la moins performante s'elle est utilisée seule.
- Lorsque des couples de proxys de site sont considérés, une performance à peu près comparable est trouvée, quel que soit le couple de ces proxys.
- Plus précisément, le meilleur couple est $[V_{s30}-H_{800}]$ à courtes périodes et $[f_0-H_{800}]$ à longues périodes, tandis que $[f_0-pente]$ apparaît comme un couple de proxys de faible coût avec des performances acceptables [0.1 – 4 s].
- Il s'avère que $(Pente/V_{s30}/H_{800}/f_0)$ n'ont quasiment pas d'influence sur la médiane.

En regardant l'amplification de site :

- V_{s30} et Pente fournissent la plus grande différenciation (amplification) à courte période.
- À longue période H_{800} et f_0 donnent les différenciations / amplifications les plus grandes.

Donc comme conclusion finale : Aucun des quatre proxys testés n'est optimal sur toute la gamme de périodes.

Il y a lieu de préciser à la fin de cette première partie que l'utilisation du rapport PSA_{sol}/PSA_{rock} donne, certes, l'amplification spectrale du site, mais il ne permet pas de distinguer entre l'amplification de site linéaire et non linéaire. Pour pouvoir distinguer entre ces deux effets, nous allons proposer dans la deuxième partie une autre démarche, basée sur le rapport des rapports de la réponse de site (RSR_{NL-L}). Dans ce qui suit, nous gardons le couple de proxys des conditions de site $[V_{s30}, f_0]$, trouvés comme pertinent en cette 1^{ère} partie, tout en testant d'autres proxys.

2^{ème} partie : Modèle neuronal à la distinction des meilleurs proxys de site pour la prise en compte des effets non-linéaires

IV.9. Contexte

Dans le cadre du projet [sinaps@](#) (volet 2 : effet de site & interaction sol-structure), nous essayons, comme la première partie de ce chapitre, de tester la pertinence des différents proxys de site, mais cette fois-ci avec un autre outil : les RSR_{NL-L} (Non-Linear to Linear site Response Spectral Ratio). Les RSR_{NL-L} sont censés fournir plus d'informations sur le comportement non linéaire des sols. L'analyse est effectuée toujours avec l'approche neuronale. La procédure du calcul des RSR_{NL-L} est donnée en détail par [Régnier et al. \(2016\)](#). Le RSR_{NL-L} est le rapport de spectre de Fourier (ou spectre de réponse) surface/profondeur pour des événements forts (BSR_j) et des événements faibles (BSR_i). Ce rapport est calculé pour chaque couple de mouvement faible et fort, ce qui donne un total égal à $N_{strong(th)} \times N_{weak}$. Pour un site donné, le rapport RSR_{NL-L} nommé encore $RSRs_{NL-L}$ est calculé comme suit :

$$\log_{10}(RSRs_{NL-L}) = \frac{I}{N_{strong(th)} \times N_{weak}} \left[\sum_{j=1}^{N_{strong(th)}} \sum_{i=1}^{N_{weak}} \log_{10} \left(\frac{BSR_j}{BSR_i} \right) \right] \quad (IV.4)$$

j représente l'indice des enregistrements forts ($PGA_{surface}$ supérieur par tranche de PGA). i est l'indice des enregistrements faibles (en fond de forage) qui sont compris entre 0.1 et 25 cm/s² (gal). th désigne le seuil du $PGA_{surface}$. BSR représente le rapport spectral entre la surface et le fond de forage tel que défini dans l'équation IV.5.

$$BSR = \sqrt{\frac{EW_{surf}^2 + NS_{surf}^2}{EW_{depth}^2 + NS_{depth}^2}} \quad (IV.5)$$

Dans laquelle EW_{surf} , NS_{surf} , EW_{depth} et NS_{depth} sont les spectres de Fourier (ou de réponse) lissés des composantes est-ouest (EW) et nord-sud (NS) à la surface ($surf$) et au fond de forage ($depth$).

En outre, le $RSRE_{NL-L}$ pour un site et un événement fort donné peut être calculé à partir de l'équation suivante :

$$\log_{10}(RSRE_{NL-L})_j = \frac{I}{N_{weak}} \left[\sum_{i=1}^{N_{weak}} \log_{10} \left(\frac{BSR_j}{BSR_i} \right) \right] \quad (IV.6)$$

Les $RSRE_{NL-L}$ sont calculés aussi pour 9 bins de fréquence et pour les deux rapports spectraux de Fourier et de réponse.

Comme nous pouvons remarquer à partir des deux équations IV.4 et IV.6, les deux RSR n'ont pas exactement la même définition. En effet, pour un site donné, le RSR_{NL-L} est défini pour un seul événement fort ($N_{strong(th)}=1$) avec un $PGA_{surface}$ seuil supérieur à 200 m/s^2 . Le PGA au fond de forage est inférieur à 25 cm/s^2 . j représente l'événement fort considéré et i est l'indice de l'événement faible qui varie entre 1 et N_{weak} .

Pour analyser l'effet non linéaire de site, [Régnier et al. \(2016\)](#) ont distingué entre le mouvement faible et le mouvement fort par l'utilisation du PGA en surface. Le mouvement faible est supposé correspondre à la réponse de site linéaire qui a été considérée avec un PGA au fond de forage compris entre 0.1 et 25 cm/s^2 . Tandis que, les mouvements forts sont supposés être affectés par le comportement non linéaire du sol. Trois différentes valeurs sont sélectionnées pour ce seuil de mouvement fort (enregistré en surface): 100 , 200 et 300 cm/s^2 :

- 1^{ère} sous ensemble de 164 sites avec un PGA seuil égal à 100 cm/s^2 .
- 2^{ème} sous ensemble de 76 sites avec PGA seuil égal à 200 cm/s^2 .
- 3^{ème} sous ensemble de 35 sites avec PGA seuil égal à 300 cm/s^2 .

Un exemple de RSR_{NL-L} est donné sur la figure.IV.15 : site IWTH04 et $PGA_{seuil} = 200 \text{ cm/s}^2$.

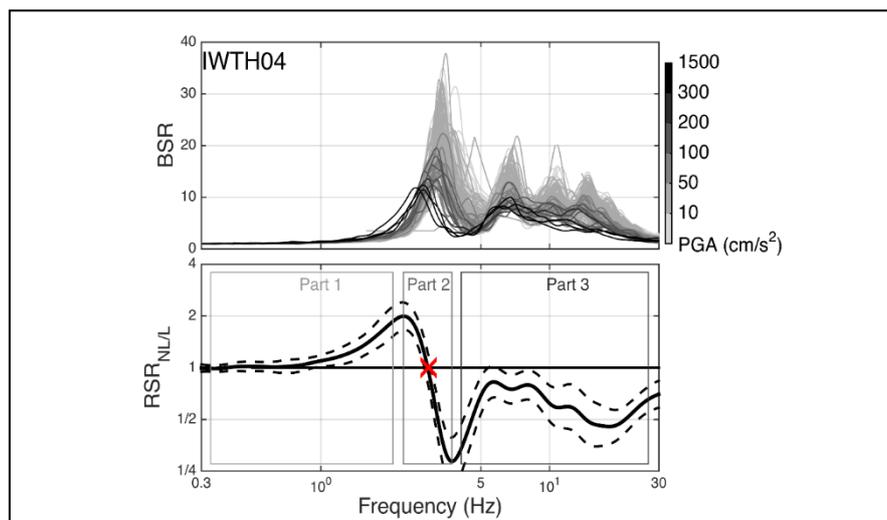


Figure.IV.15 Illustration de la quantification des effets de site non linéaire pour le site IWTH04. (a) le rapport spectral surface/(fond de forage). (b) RSR_{NL-L} pour un PGA_{seuil} des mouvements forts égal à 200 m/s^2 et pour un PGA représentant les mouvements faibles $<25 \text{ cm/s}^2$. Le trait continu correspond à la médiane et les deux lignes en pointillé correspondent à $RSR_{NL-L} \pm$ l'écart type. Le point rouge représente la fréquence F_{in} qui correspond à l'intersection avec la courbe de RSR_{NL-L} médiane dont la valeur est juste inférieure à 1 ([Régnier et al. 2016](#)).

La Figure.IV.15 représente la forme de RSR_{NL-L} qui est proche de la première dérivée de la fonction Gaussienne. Pour la majorité des sites, la courbe est divisée en 3 principales parties. La partie 1 représente une lente augmentation varie autour de 1 pour une fréquence faible jusqu'à une fréquence spécifique qui varie selon le site. Dans la partie 2 nous assistons à une diminution brutale de RSR_{NL-L} en dessous de l'unité. C'est dans cette partie 2 que, nous

définissons la fréquence f_{NL} qui est la fréquence à partir de laquelle le RSR_{SNL-L} descend en dessous de l'unité avec une signification importante. Dans la partie 3, le RSR_{SNL-L} s'approche légèrement de l'unité, mais reste inférieur à un.

La figure.IV.15 reflète la diminution de fréquence de résonance de site provoqué par la dégradation des propriétés élastiques sous l'effet de mouvement fort. Il est l'un des effets du comportement non linéaire des sols.

Nous rappelons que le but ici est de tester la pertinence de différents proxies de site à faire réduire les incertitudes lors de la prédiction du RSR_{NL-L} (soit de site ou d'événement). Afin d'atteindre cet objectif, nous utilisons les jeux de données élaborés dans [Régnier et al. \(2016\)](#) et la méthode neuronale en élaborant des équations de prédiction du rapport RSR_{NL-L} , qui est représenté par la fonction G , tel que :

$$RSR_{NL-L}(f) = G(\beta|W) \pm \varepsilon\sigma(f) \quad (IV.7)$$

G est la fonction à prédire σ représente l'écart type ε est la valeur de normalisation de σ . La fréquence structurelle est donnée par f allant de 0.2 Hz à 30 Hz. β désigne le vecteur des proxies de site et/ou de sollicitations. Tandis que, W représente les paramètres du modèle neuronal (matrices des poids synaptiques et les vecteurs des biais).

Nous commençons, donc, par définir les jeux de données utilisés, ainsi que les métadonnées choisies. Par la suite, nous exposons la démarche utilisée pour l'élaboration des modèles de RSR_{NL-L} . Enfin nous exposons les résultats obtenus.

IV.10. Jeux de données

Nous utilisons ici les jeux de données de Régnier et al. [2013](#) et [2016](#). Une sélection de sites a été effectuée par les auteurs afin d'obtenir des proxies de site et les rapports spectraux pour les deux événements faibles et forts. Plus de 50400 enregistrements de 1996 à 2014 ont été adoptés. Il est à noter que 75 % des PGA à la surface sont inférieurs à 10 cm/s², 124 sites ont enregistré au moins 3 événements avec des PGA en surface supérieure à 100 cm/s². Les BSR peuvent être soit le rapport de spectre de Fourier BSR_F soit le rapport de spectre de réponses BSR_R ([Régnier et al. 2016](#)).

Ici le fond de forage est considéré comme le site de référence. Généralement, dans ce type de site l'effet de profondeur est présent ([Duval et al. 1995](#), [Steidl et al. 1996](#); [Kokusho, 2004](#); [Cadet et al. 2012](#)). Dans notre cas la profondeur du fond de forage est comprise entre 100 m et

200 m ce qui nous permet de supposer qu'au niveau du fond de forage le rapport BSR des enregistrements forts sur le rapport BSR des enregistrements faibles est proche de l'unité. En effet, à une telle profondeur l'effet non linéaire devrait être négligeable. De cette manière, le problème des effets de profondeur est surmonté.

Les proxys sujets d'étude, sont les proxys de site (ex. V_{s30} , f_0 ...) ou/et les proxys du chargement sismique (ex. PGA, PGV...). Ces derniers sont nécessaires pour tenir compte du comportement non linéaire du sol. Pour cette raison, [Régnier et al. \(2013\)](#) ont construit deux types de bases de données : la base de données sites dans laquelle les proxys de site sont préconisés et la base de données événements qui, en plus des proxys de site, les proxys du chargement sismique sont considérés.

IV.10.1. Base de données sites

Les proxys de site disponibles dans la base de données « sites » sont, en plus de V_{s30} , f_0 et f_{NL} (Figure.IV.15), le rapport V_{s30}/V_{smin} , dont V_{smin} est la vitesse minimale du profil du sol et le gradient du profil de vitesse V_s sur 30 m de profondeur B_{30} ([Régnier et al. 2013](#)). Ces proxys, hormis f_{NL} , sont faciles à déterminer et accessibles.

Tandis que, les RSR_{SNL-L} sont les paramètres recherchés (équation IV.7). Quelques informations relatives à f_{NL} sont données sur le tableau.IV.6.

Tableau.IV.6 Données statistiques pour chaque site de la base de données sites : Nombre de sites avec la valeur moyenne arithmétique de f_{NL} ainsi que l'intervalle et l'écart type de f_{NL} ([Régnier et al. 2016](#)).

	Number of Sites with f_{NL} Values	Arithmetic Average f_{NL} (Hz)	f_{NL} Range (Hz)	Standard Deviation σ (Hz)
Subset 1 (PGA > 100 cm/s ²)	164	5.9	0.49–15.7	3.7
Subset 2 (PGA > 200 cm/s ²)	76	6.0	0.85–22.2	3.9
Subset 3 (PGA > 300 cm/s ²)	35	6.5	1.4–22.7	4.5

Les RSR_{SNL-L} sont prédits pour 9 bins de fréquences. Ces bins sont données sur le tableau.IV.7.

Tableau.IV.7 bins de fréquences (Hz) considérés pour la prédiction des RSR_{SNL-L}

0.2-0.3	0.3-0.6	0.6-1.1	1.1-1.9	1.9-3.2	3.2-5.6	5.6-9.9	9.9-17.2	17.2-30
---------	---------	---------	---------	---------	---------	---------	----------	---------

Les 5 proxys de sites (V_{s30} , f_0 , f_{NL} , V_{s30}/V_{smin} et B_{30}) et les RSR_{SNL-L} pour les 09 bins de fréquences sont les paramètres expliqués et inexpliqués, respectivement de nos modèles neuronaux. Les trois bases de données sont utilisées pour élaborer les différents modèles de prédiction de RSR_{SNL-L} .

IV.10.2. Base de données événements

Dans la base de données "événements", nous choisissons comme métadonnées les 5 paramètres de site suscités et ceux du chargement sismique qui sont susceptibles de caractériser le comportement non linéaire de site. Les proxies du chargement sismiques sont : PGA_s qui représente l'accélération maximale à la surface, PGV_s : la vitesse maximale à la surface, le proxy caractérisant la déformation PGV_s/V_{s30} ([Guéguen, 2016](#)). Il est à noter que, le PGD n'est pas considéré ici, en raison de la difficulté d'obtenir des PGD fiables à partir d'accélérogrammes. D'ailleurs c'est la raison pour laquelle nous recensons seulement 19 équations de prédiction du PGD ([Douglas, 2012](#)).

IV.11. Methodologie

Pour construire un modèle capable de prédire les RSR_{NL-L} ou les $RSRE_{NL-L}$ à partir du spectre de Fourier (RSR_{NL-L_F}) ou du spectre de réponse (RSR_{NL-L_R}), nous suivrons la procédure suivante :

- Nous fixons le nombre de bins de fréquences dont nous voulons calculer les ordonnées spectrales, ces valeurs sont données par le tableau.IV.7.
- Nous choisissons les proxies de site et du chargement sismique.
- Nous fixons le nombre de neurones dans la couche cachée. Le nombre de neurones dans la couche de sortie est le nombre de bins de fréquence : 9 neurones.
- Nous choisissons la fonction d'activation dans la couche cachée et la couche de sortie. La fonction tangente hyperbolique et la fonction linéaire sont préconisées pour la couche cachée et la couche de sortie respectivement.

Pour éviter le problème de surapprentissage, nous utilisons la technique de régularisation par modération des poids (weight decay) qui a été donnée par l'équation II.6. Il est à rappeler que la minimisation de la fonction coût est faite par l'algorithme d'apprentissage de la rétropropagation du gradient de deuxième ordre quasi-Newton (BFGS), ([Derras, 2011](#)). Cette technique est économe du point de vue espace mémoire et du temps de calcul en comparaison avec la méthode de Levenberg-Marquardt (LM). En outre, le BFGS est stable, donne des prédictions semblables, quels que soient les poids initiaux.

Pour augmenter les performances du modèle, on préfère utiliser les données d'apprentissage (entrée/sortie) normalisées entre [-1 et 1]. Cette étape est fortement recommandée pour éviter la saturation des fonctions d'activation utilisées. Les équations qui assurent la normalisation et la dé-normalisation sont A1 et A2 et A3 de [Derras et al. \(2012\)](#).

Le modèle neuronal consiste donc en une série de trois couches. La première représente les proxies de sites (bases de données sites) et ceux du site et du chargement sismique (base de données événements). La dernière couche représente les valeurs de RSR_{NL-L_F} ou RSR_{NL-L_R} en 9 bins de fréquences. Les paramètres du réseau (les valeurs des poids) sont initialisés aléatoirement. Le nombre de neurones de la couche cachée « N » égal au nombre de proxies d'entrée ([Wierenga and Kluytmans 1994](#)). L'architecture retenue est N-N-9. Plusieurs modèles de prédiction de RSR_{NL-L_F} et RSR_{NL-L_R} sont élaborés afin de répondre à la question suivante :

IV.12. Quels sont les proxies de conditions locales de sites les plus pertinents?

Pour pouvoir y répondre, nous préconisons 2 outils statistiques (déjà utilisés dans la première partie de ce chapitre). Ces outils sont l'écart type (σ) du modèle (équation.IV.8) et le taux de réduction R_c (équation IV.10) de σ du modèle par rapport à celui de la base de données (équation IV.8). La différence entre les modèles de RSR_{NL-L} élaborés réside dans la nature du proxy ou des proxies utilisé(s).

$$\sigma(f) = \sqrt{\frac{1}{M} \sum_I^M (\log_{10}(RSR_{NL-L_{obs}}(f)) - \log_{10}(RSR_{NL-L_{pred}}(f)))^2} \quad (IV.8)$$

σ intrinsèque de référence :

$$\sigma(f)_{ref} = \sqrt{\frac{1}{M} \sum_I^M (\log_{10}(RSR_{NL-L_{obs}}(f)) - \log_{10}(\text{mean}(RSR_{NL-L_{obs}}(f)_{f,obs})))^2} \quad (IV.9)$$

Coefficient de réduction :

$$R_c(f) = (1 - \frac{\sigma(f)}{\sigma_{ref}(f)}) \cdot 100 \quad (\%) \quad (IV.10)$$

M représente la taille de l'échantillon.

La procédure entreprise dans cette étude pour distinguer les meilleurs proxies de site est donnée sous forme d'organigramme schématisé par la Figure.IV.16. Dans chaque modèle neuronal nous calculons les deux quantités $\sigma(f)$ et $R_c(f)$. Les meilleurs proxies de site sont ceux qui donnent de faibles valeurs de $\sigma(f)$ et de s pourcentages de $R_c(f)$.

Dans ce qui suit nous présentons, à titre d'exemple, l'analyse effectuée sur le deuxième jeu de données « site » avec un PGA_s seuil (Th) égal à 200 m/s^2 et ensuite sur le jeu de données « événements ». Nous commençons donc par analyser le jeu de données « Sites »: métadonnées et RSR_{NL-L} .

IV.12.1. Analyse statistique des proxies de site

Cela nous amène premièrement à choisir la fonction de transformation utilisée pour les proxies de site et pour les RSR_{NL-L} . De ce fait nous proposons trois cas de figures :

- Prendre les valeurs initiales (sans transformation)
- Transformation avec \log_{10}
- Transformation avec la fonction BOX-COX.

La fonction de transformation BOX-COX (des noms de son inventeur [Box et Cox \(1964\)](#) est définie comme suit :

$$BOXCOX(data, \lambda) = \frac{data^\lambda - 1}{\lambda} \quad (IV.11)$$

Dans le cas où $\lambda=0$ $BOXCOX(data, \lambda) = \ln(data)$;

Le but de la transformation est d'obtenir une distribution normale symétrique et non aplatie. Cela permet d'éviter des estimations biaisées. Pour ce faire, nous utilisons les deux coefficients d'aplatissement Kurtosis (K) et d'asymétrie Skewness (S), qui sont définis comme suit :

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (X)^4}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (X)^2} \right)^4} \quad (IV.12)$$

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (X)^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (X)^2} \right)^3} \quad (IV.13)$$

Où X représente le paramètre sujet d'analyse et n le nombre d'échantillons (*data*) de la base de données. Donc X peut prendre trois cas de figure : $X=data$, $X=\log_{10}(data)$ ou $X=BOXCOX(data, \lambda)$. Un extrait de résultats est donné sur les deux tableaux IV.8 et IV.9.

Pour voir l'impact de l'utilisation des deux transformations sur la distribution des données, nous prenons le proxy de site V_{s30} toute en considérant la base de données « sites » avec un seuil de $PGA_s=200$ m/s². Dans la figure.IV.17 nous représentons les trois cas de figure de transformations déjà évoquées. Il en résulte que la transformation BOXCOX donne la meilleure approximation et converge vers la loi normale théorique.

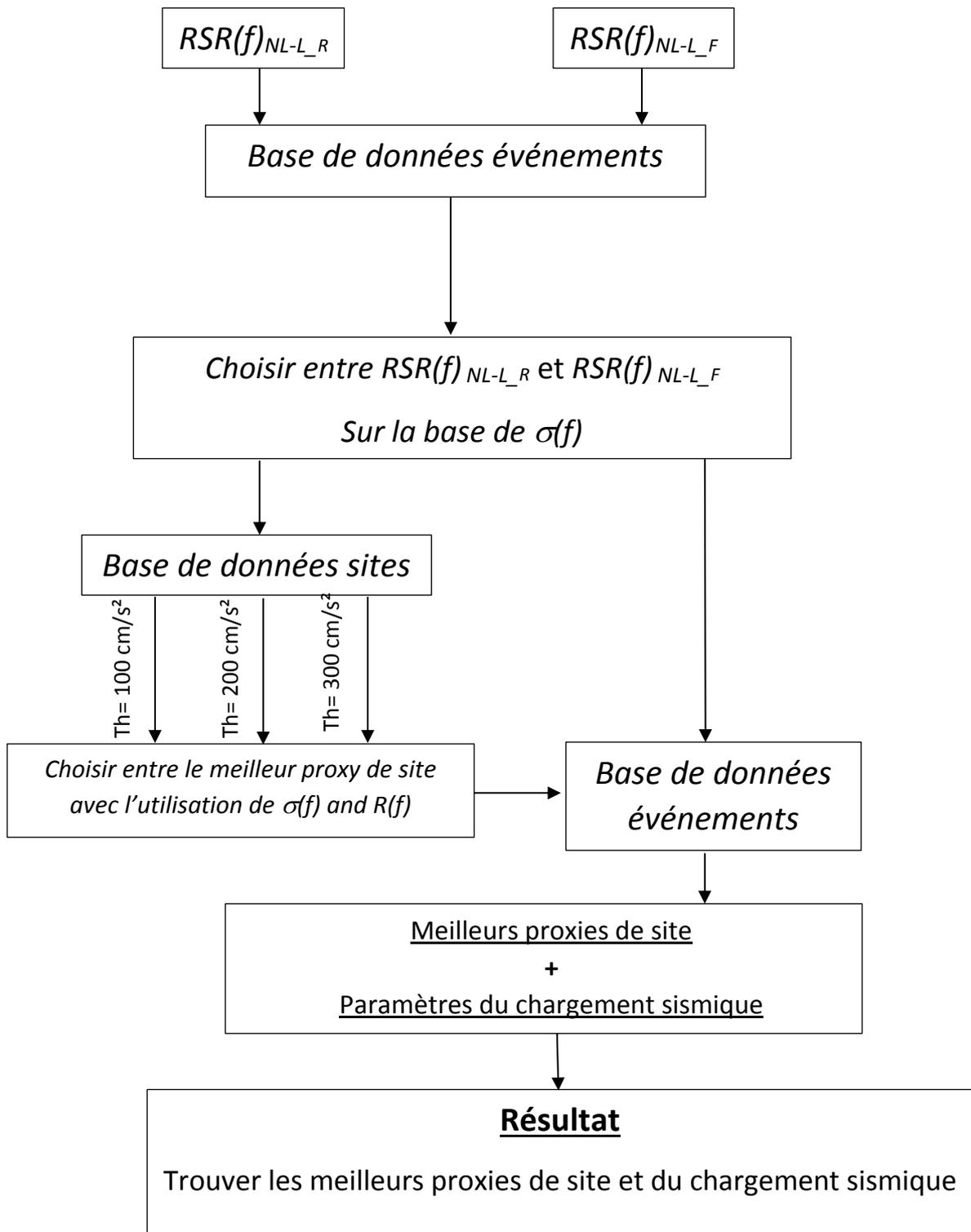


Figure.IV.16 Procédure de l'analyse de la persévérance des différents proxies de site

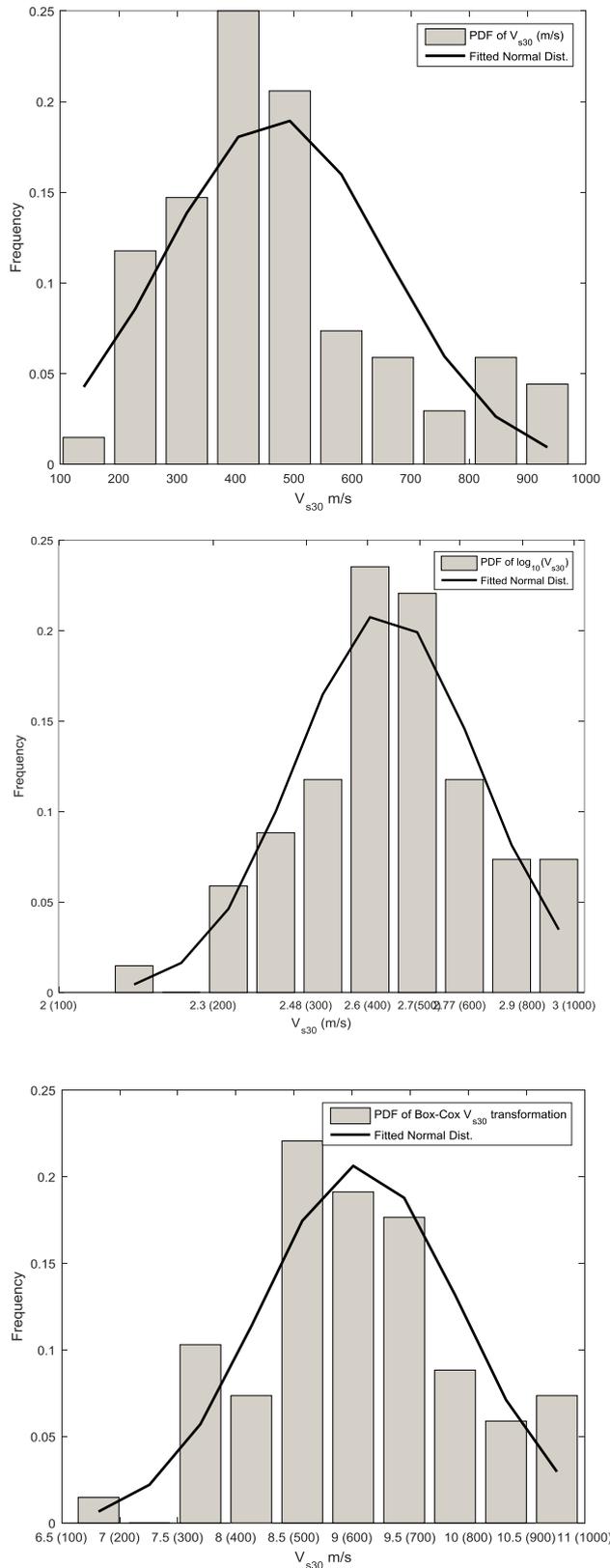


Figure.IV.17 Distribution de V_{s30} en utilisant les données initiales (haut), transformation en $\log_{10}(V_{s30})$ (milieu) et $BOXCOX(V_{s30})$ en bas, avec les distributions théoriques qui leur correspondent

Le test pour l'ensemble de proxies s'impose. Dans ce contexte, nous effectuons la même analyse que celle donnée par la figure.IV.17 en ajoutant les coefficients K et S. les résultats de

l'analyse des proxies de sites (avec la base de données « site » $PGA_s > 200$) et avec la base de données « événements » sont données par le tableau.IV.8 et le tableau.IV.9 respectivement. Nous remarquons que les proxies de site nécessitent une transformation BOXCOX, hormis pour le B_{30} .

Table.IV.8 Valeurs des deux coefficients kurtosis et skewness pour les 5 proxies de site en tenant compte de la base de données site pour un seuil de $PGA_s = 200 \text{ cm/s}^2$.

Proxies de site	Kurtosis (K idéal =3)			Skewness (S idéal = 0)			Valeur de λ (BOX-COX)	Le choix se fait sur la base de K et S et l'allure de la distribution gaussienne
	Sans transformation	BOX-COX	Log ₁₀	Sans transformation	BOX-COX	Log ₁₀		
V_{s30}	3.06	2.77	2.88	0.81	0.00	-0.14	0.12	BOX-COX
V_{s30}/V_{smin}	2.69	2.51	2.6	0.53	-0.01	-0.19	0.24	BOX-COX
B_{30}	2.26	4.19	62.68	-0.11	-1.03	-7.75	1	Sans transformation
f_{0HV}	2.86	2.01	2.34	0.79	-0.18	-0.76	0.37	BOX-COX
F_{NL}	6.57	2.94	3.01	1.38	-0.02	-0.64	0.35	BOX-COX

Table.IV.9 Valeurs des deux coefficients kurtosis et skewness pour les 7 proxies de site en utilisant la base de données événements

Proxies de site	Kurtosis (K idéal =3)			Skewness (S idéal = 0)			Valeur de λ (BOX-COX)	Le choix se fait sur la base de K et S et l'allure de la distribution gaussienne
	Sans transformation	BOX-COX	Log ₁₀	Sans transformation	BOX-COX	Log ₁₀		
V_{s30}	3.15	2.57	2.61	0.83	0.00	-0.08	0.12	BOX-COX
V_{s30}/V_{smin}	7.01	3.02	3.00	1.50	0.00	-0.05	-0.01	BOX-COX
B_{30}	2.19	3.40	85.11	-0.21	-0.84	-8.64	1	Sans transformation
f_{0HV}	3.24	2.23	2.95	0.87	-0.14	-0.90	0.41	BOX-COX
F_{NL}	2.36	2.34	3.98	0.44	-0.08	-0.97	0.54	BOX-COX
PGA_s	76.78	2.18	4.40	6.44	0.12	1.11	-0.56	BOX-COX
PGV_s	81.77	2.81	3.81	7.20	0.02	0.78	-0.31	BOX-COX
PGV_s/V_{s30}	113.45	3.01	3.36	8.26	0.00	0.41	-0.14	BOX-COX

De même, l'ensemble des bins de RSR_{NL-L_F} et RSR_{NL-L_R} pour les 4 bases de données nécessitent une transformation BOXCOX. Ce constat a été obtenu après avoir effectué tous les tests nécessaires.

La topologie des modèles neuronaux est fixée, l'analyse des jeux de données est effectuée, il reste qu'à débiter l'étude proprement dite. En premier lieu, nous commençons par choisir entre l'utilisation du spectre de Fourier ou du spectre de Réponse pour l'établissement du RSR_{NL-L} . Dans ce qui suit, nous effectuons cette analyse avec la base de données événements.

IV.13. Resultats

IV.13.1. $RSRE_{NL-L_F}$ ou $RSRE_{NL-L_R}$?

Le spectre de réponse d'un système à un seul degré de liberté est largement utilisé par les développeurs des GMPEs. Cependant, le spectre de Fourier commence à trouver son utilisation (ex. [Bora et al. 2015](#)), surtout que le passage du spectre de Fourier vers un spectre de réponse peut se faire aisément.

Le but ici est de savoir quel est le meilleur outil ($RSRE_{NL-L_F}$ ou $RSRE_{NL-L_R}$) qui génère la variabilité aléatoire la plus faible.

Pour ce faire, nous construisons deux modèles neuronaux avec les caractéristiques déjà suscitées. Pour être cohérents avec les GMPEs, nous utilisons dans ce test un proxy de site : V_{s30} et un proxy de mouvement sismique PGA enregistré en surface libre. Les sorties sont évidemment les 9 bins de $RSRE_{NL-L_F}$ ou $RSRE_{NL-L_R}$ en utilisant une fonction d'activation linéaire. La couche cachée contient 2 neurones avec une fonction tangente hyperbolique. La transformation BOXCOX est préconisée pour les couples entrées sorties.

À la fin de la phase d'apprentissage et de régularisation nous obtenons les valeurs de $\sigma[\log_{10}(RSRE_{NL-L_F})]$ et $\sigma[\log_{10}(RSRE_{NL-L_R})]$ que nous les représentons sur la figure IV.18 et ce pour les 9 bins de fréquences.

En comparant les deux courbes de la figure.IV.18, nous pouvons constater que, les deux tendances sont inversées : la valeur de $\sigma[\log_{10}(RSRE_{NL-L_R})]$ diminue avec la fréquence et $\sigma[\log_{10}(RSRE_{NL-L_F})]$ plutôt augment avec les bins de fréquences ce qui est cohérent avec la tendance obtenue dans la première partie de ce chapitre (figure.IV.5). Il est à rappeler que dans la première partie, la base de données KiK-Net a été adoptée d'où la cohérence trouvée. Cette

compatibilité s'étale aussi sur le bin de fréquences où la variabilité aléatoire est maximale. Ici c'est pour les fréquences allant de [9.9-17.2] Hz cela correspond à $T=[0.06-0.1]$ s. Tandis que, le σ dans la figure.IV.5 correspond à $T=0.15$ s. Donc le maximum de σ converge plus au moins vers le même intervalle de fréquence.

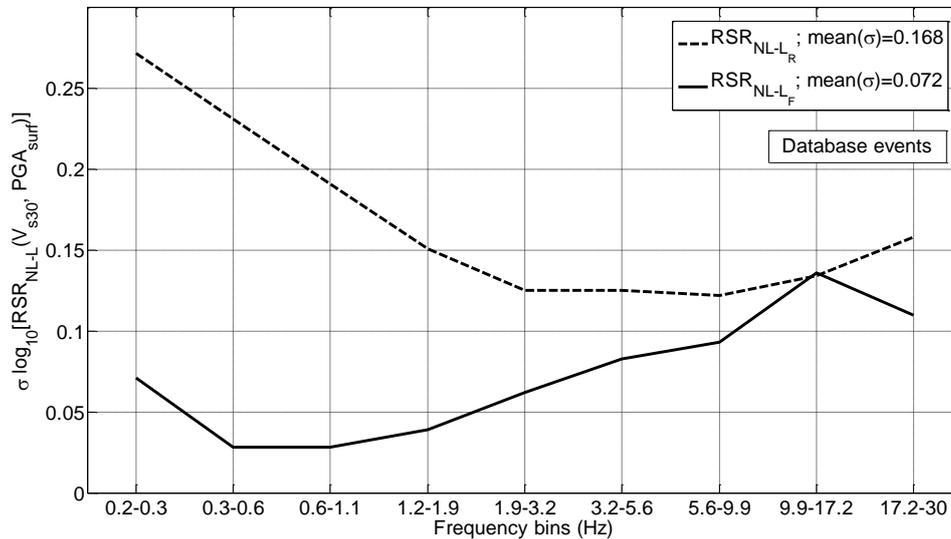


Figure.IV.18 Valeurs et tendances de $\sigma [\log_{10}(\text{RSR}_{\text{NL-L}})]$ en fonction des 9 bins de fréquences. Les résultats sont obtenus avec les deux proxies V_{s30} et PGA_s

En termes de valeurs d'incertitude, les σ obtenus à partir de $\text{RSR}_{\text{NL-L}_F}$ sont beaucoup plus faibles que ceux de $\text{RSR}_{\text{NL-L}_R}$.

Quantitativement la moyenne de $\sigma[\log_{10}(\text{RSR}_{\text{NL-L}_R})]=2.33 \times \sigma[\log_{10}(\text{RSR}_{\text{NL-L}_F})]$. Cette constatation n'est pas valable pour le bin de fréquence $=[9.9-17.2]$ Hz où les deux valeurs de σ se convergent. La distinction, en termes de variabilité aléatoire, est bien visible : le rapport $\text{RSR}_{\text{NL-L}_F}$ est mieux que $\text{RSR}_{\text{NL-L}_R}$, donc nous l'utiliserons dans ce qui suit.

IV.13.2. Contribution des proxies de site et du chargement sismique à la réduction de la variabilité aléatoire des $\text{RSR}_{\text{NL-L}}$

Dans la première partie de ce chapitre nous avons vu que, le couple de proxies de site $[V_{s30}, f_0]$ semble à la fois pertinent et économique par rapport aux H_{800} et la pente topographique « slope ». L'influence des proxies de site sur la variabilité aléatoire commence à partir de $T > 0.15$ s ($f_0 < 6.66$ Hz), dans ce domaine le $R_{c\text{max}}(0.5 \text{ Hz}) = 27\%$ pour $[V_{s30}-f_0]$. Ce dernier a été calculé par rapport au σ de référence (modèle sans proxies de site). Ici, étant donné que nous n'utilisons pas une GMPE le R_c est déterminé par rapport au σ intrinsèque (équation.IV.9). Dans le tableau IV.10, IV.11 et IV.12, nous donnons-le σ de référence (σ_{ref}) ainsi que le R_c et le σ des $\log_{10}(\text{RSR}_{\text{NL-L}})$ obtenus par les modèles neuronaux en considérant, en premier lieu, un

seul proxy de site, en deuxième lieu, les 5 proxys de site et en troisième lieu nous ajoutons le proxy f_{NL} à ces 5 proxys de site. Dans les 3 tableaux nous présentons le σ_{ref} , le R_c et le σ pour les 9 bins de fréquences et ce pour les trois bases de données respectivement.

Dans le tableau.IV.10 et pour une valeur seule de $PGA_s=100$ cm/s² le paramètre le plus pertinent est le f_{NL} pour une fréquence allant de 1.2 Hz à 30 Hz. L'utilisation des 5 proxys de site plus le f_{NL} donne un R_c moyen égal à 15 % avec une valeur maximale de $R_c=33.3$ % pour $f=[3.2-5.6]$ Hz. Si nous augmentons le seuil de PGA_s à 200 cm/s², la pertinence de f_{NL} se rétrécit pour couvrir un intervalle de [3.2-30] Hz à l'extérieur de ce domaine c'est V_{s30} qui semble le proxy qui réduit le plus la variabilité aléatoire (tableau.IV.11). L'ensemble de paramètres donne un R_c moyen égal à 22.4 % et une valeur de R_c maximale =41.5 % [5.6-9.9] Hz. Avec la dernière base de données site ($PGA_s>300$ cm/s²), le f_{NL} et B_{30} s'avèrent les proxys de site qui réduisent le plus la variabilité aléatoire dans les modèles de prédiction des RSR_{NL-L} , ce résultat est représenté sur la tableau.IV.12. L'utilisation de tous les paramètres de site accroît le R_c à une valeur égale à 49.5 % pour une des fréquences comprises entre 5.6-9.9 Hz et une valeur moyenne égale à 37.6 %. Nous enregistrons une nette amélioration par rapport à la première et la deuxième base de données.

Table.IV.10 Valeurs de σ_{ref} , σ et R_c pour un proxy de site et pour l'ensemble de proxys de site en utilisant la base de données « sites » avec un seuil de $PGA_s=100$ cm/s².

Seuil de $PGA_s = 100$ cm/s ² (nombre de sites =164)									
f_bin	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.030	0.026	0.062	0.061	0.090	0.115	0.144	0.113	0.137
Proxies de site									
$V_{s30} (\sigma, R_c)$	0.029	0.024	0.062	0.060	0.084	0.099	0.118	0.108	0.137
	2.218	5.585	0.008	1.013	7.420	14.524	18.052	4.505	0.009
V_{smin}	0.030	0.025	0.062	0.061	0.090	0.113	0.144	0.113	0.137
	0.627	0.705	0.021	0.061	0.465	1.741	0.130	0.173	0.357
V_{s30}/V_{smin}	0.029	0.025	0.062	0.061	0.089	0.110	0.142	0.113	0.137
	1.443	2.274	0.089	0.056	1.470	5.032	1.509	0.750	0.459
B_{30}	0.030	0.025	0.062	0.061	0.087	0.111	0.142	0.113	0.136
	0.597	3.839	0.472	0.276	3.478	3.455	1.398	0.362	0.953
f_{0HV}	0.029	0.023	0.062	0.060	0.083	0.103	0.130	0.111	0.136
	0.991	8.064	0.306	1.658	8.296	10.669	9.952	1.755	0.514
f_{NL}	0.030	0.024	0.062	0.060	0.082	0.087	0.115	0.105	0.136
	0.557	4.254	0.320	2.152	9.191	24.906	20.234	7.871	1.009
5 SCPs	0.029	0.023	0.061	0.055	0.074	0.091	0.113	0.107	0.133
	3.493	10.877	1.378	10.223	18.005	20.921	21.621	5.727	2.814
5 SCPs + f_{NL}	0.029	0.023	0.061	0.051	0.067	0.077	0.104	0.101	0.131
	3.908	10.897	2.604	16.715	25.618	33.275	28.189	10.612	4.824

Globalement, R_c augment avec le seuil du chargement sismique cela veut dire une réduction de la variabilité aléatoire liée au comportement non-linéarité de site.

Donc il existe une dépendance entre le chargement sismique est le comportement non-linéaire du sol. Cependant quel est le proxy du chargement sismique qui réduit le plus la variabilité aléatoire ?.

Pour y répondre et afin d'élaborer les modèles de prédiction des $RSRE_{NL-L}$, nous utilisons la base de données « événements ». En plus des 5 proxys de site utilisés précédemment, nous testons les PGA_s et les PGV_s qui sont les plus utilisés dans GMPEs pour tenir compte de la non-linéarité du sol. Les résultats de cette analyse sont donnés sur le tableau.IV.13. Malgré, l'intégration de ces deux proxys dans les modèles de prédiction des $RSRE_{NL-L}$, les résultats restent décevants. La situation s'améliore un peu en ajoutant f_{NL} .

Table.IV. 11 Valeurs de σ_{ref} , σ et R_c pour un proxy de site et pour l'ensemble de proxys de site en utilisant la base de données « sites » avec un seuil de $PGA_s=200 \text{ cm/s}^2$.

Seuil de $PGA_s = 200 \text{ cm/s}^2$ (nombre de sites =76)									
f_{bin}	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
σ_{ref}	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.030	0.032	0.049	0.065	0.108	0.141	0.184	0.141	0.179
Proxies de site									
$V_{s30} (\sigma, R_c)$	0.029	0.029	0.042	0.064	0.104	0.115	0.151	0.137	0.178
	2.508	8.061	14.746	1.715	3.861	18.487	17.943	2.995	0.339
V_{smin}	0.030	0.031	0.048	0.063	0.108	0.138	0.184	0.141	0.177
	0.156	1.046	0.741	2.471	0.018	2.119	0.059	0.588	1.340
V_{s30}/V_{smin}	0.030	0.031	0.047	0.062	0.108	0.133	0.181	0.140	0.176
	0.632	3.131	3.361	3.919	0.078	5.994	1.716	1.249	1.509
B_{30}	0.029	0.030	0.046	0.065	0.103	0.135	0.179	0.140	0.178
	1.980	6.539	6.423	0.091	4.277	4.129	2.604	0.930	0.665
f_{0HV}	0.030	0.030	0.045	0.064	0.105	0.126	0.166	0.137	0.175
	1.027	6.049	7.998	1.842	2.481	10.875	9.368	2.797	1.977
f_{NL}	0.029	0.030	0.045	0.064	0.105	0.110	0.139	0.131	0.172
	1.550	4.874	8.519	0.983	2.615	22.284	24.348	7.306	4.026
5 SCPs	0.029	0.027	0.038	0.053	0.092	0.105	0.133	0.133	0.166
	4.300	13.481	21.076	18.945	14.876	25.535	27.758	6.018	7.393
5 SCPs + f_{NL}	0.028	0.027	0.037	0.044	0.082	0.096	0.107	0.120	0.153
	4.602	14.129	23.373	31.431	24.321	32.213	41.554	15.363	14.731

À partir de ces résultats et pour pouvoir divulguer la participation du chargement sismique à la réduction de la variabilité aléatoire, nous proposons de normaliser les bins de fréquences par le f_{NL} (Figure.IV.19). Cela se traduit par l'ajout de f_m/f_{NL} comme une entrée aux modèles neuronaux de prédiction des $RSRE_{NL-L}$. Il est à noter que la fréquence f_m représente la fréquence moyenne de chaque bin $m=[1-9]$. A cet égard, nous élaborons 9 réseaux de neurones.

Table.IV.12 Valeurs de σ_{ref} , σ et R_c pour un proxy de site et pour l'ensemble de proxys de site en utilisant la base de données « sites » avec un seuil de $PGA_s=300 \text{ cm/s}^2$.

Seuil de $PGA_s = 300 \text{ cm/s}^2$ (nombre de sites =35)									
f_bin	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.017	0.035	0.050	0.069	0.120	0.132	0.203	0.170	0.190
Proxies de site									
$V_{s30}(\sigma, R_c)$	0.015	0.032	0.037	0.067	0.115	0.120	0.169	0.167	0.188
	10.864	9.040	25.846	3.607	3.970	8.857	16.584	1.554	0.800
V_{smin}	0.017	0.032	0.045	0.063	0.120	0.115	0.186	0.168	0.184
	2.721	9.693	10.071	8.449	0.076	12.726	8.170	1.175	2.940
V_{s30}/V_{smin}	0.017	0.031	0.043	0.063	0.120	0.113	0.183	0.167	0.185
	3.555	10.818	12.382	8.932	0.024	13.866	9.878	1.315	2.408
B_{30}	0.015	0.030	0.039	0.069	0.111	0.119	0.192	0.165	0.188
	11.142	13.849	20.912	0.112	7.083	9.739	5.441	2.570	0.586
f_{0HV}	0.017	0.033	0.042	0.067	0.117	0.115	0.191	0.158	0.181
	2.713	6.332	15.963	2.424	2.553	12.839	5.923	6.782	4.682
f_{NL}	0.017	0.033	0.042	0.064	0.118	0.113	0.164	0.155	0.185
	2.037	6.196	16.155	7.349	1.567	13.953	19.132	8.378	2.209
5 SCPs	0.014	0.027	0.028	0.033	0.096	0.094	0.123	0.140	0.147
	21.465	21.558	44.447	52.476	20.100	28.888	39.216	17.195	22.331
5 SCPs + f_{NL}	0.013	0.025	0.027	0.028	0.077	0.091	0.103	0.115	0.126
	24.381	27.488	45.633	59.694	35.935	30.602	49.398	31.983	33.394

Table.IV.13 Valeurs de σ_{ref} , σ et R_c pour, en plus, de l'ensemble de proxys de site le et les proxys du chargement sismique en utilisant la base de données « événements ».

Jeu de données événements (nombre d'enregistrements =3790)									
f_Bin	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.077	0.028	0.024	0.039	0.065	0.084	0.098	0.145	0.113
Proxies de site									
5 SCPs	0.077	0.028	0.024	0.038	0.064	0.081	0.095	0.140	0.112
	0.866	1.237	0.756	2.290	0.939	3.237	3.096	3.368	1.234
5 SCPs + PGA_s	0.077	0.028	0.024	0.038	0.064	0.081	0.095	0.139	0.112
	0.972	1.349	0.747	2.465	1.259	3.785	3.205	3.663	1.290
5 SCPs + PGV_s	0.077	0.028	0.024	0.038	0.064	0.081	0.095	0.140	0.112
	0.927	1.428	0.818	2.482	0.962	3.439	3.191	3.547	1.249
5 SCPs+ PGV_s/V_{s30}	0.077	0.028	0.024	0.038	0.064	0.081	0.095	0.140	0.112
	0.919	1.425	0.809	2.475	0.979	3.456	3.216	3.523	1.220
5 SCPs+PGA_s+PGV_s+PGV_s/V_{s30}	0.076	0.028	0.024	0.037	0.063	0.079	0.095	0.138	0.107
	1.926	2.029	2.510	4.578	3.693	5.485	3.500	4.882	5.199
5 SCPs+f_{NL}	0.076	0.028	0.024	0.035	0.062	0.078	0.095	0.139	0.102
	1.481	2.275	1.418	10.707	5.302	6.948	3.667	4.017	10.184
5 SCPs+PGA_s+PGV_s+PGV_s/V_{s30}+f_{NL}	0.076	0.027	0.023	0.028	0.056	0.069	0.093	0.134	0.094
	1.399	4.805	5.272	28.308	13.335	17.597	5.750	7.674	17.105

Plusieurs tests ont été effectués afin d’obtenir les proxies de site les plus pertinents. La seule combinaison qui donne des résultats satisfaisants c’est celle où nous utilisons les 5 proxies de site, le f_{NL} , les proxies du chargement sismique et le rapport f_m/f_{NL} . Le σ_{ref} le σ et R_c sont mentionnés sur le tableau.IV.14.

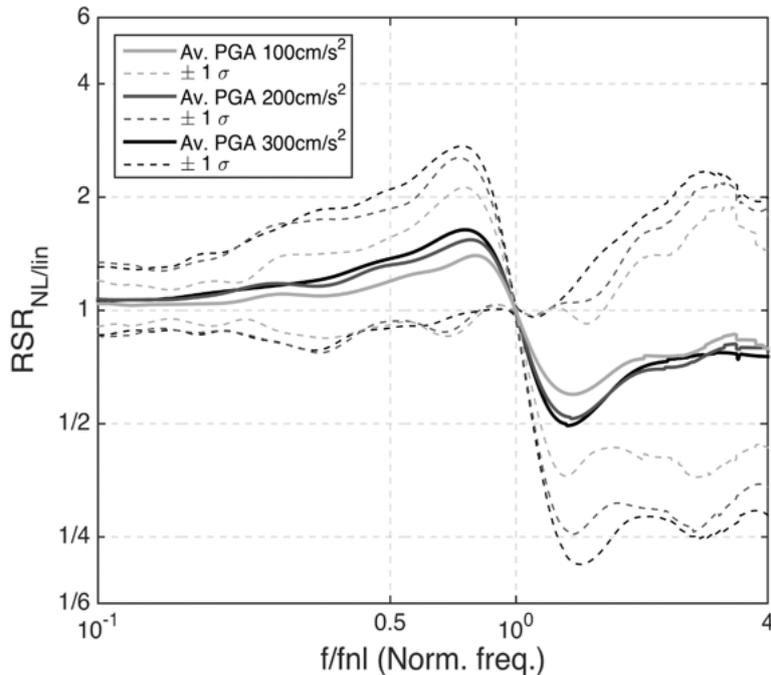


Figure.IV.19 Courbes des médianes $\pm\sigma$ des RSR_{NL-L} adimensionnels normalisés en fréquences pour les 3 niveaux de PGA_s . (Régnier et al. 2015)

Table.IV.14 Valeurs de σ_{ref} , σ et R_c données par les $RSRE_{NL-L}$ prédites en utilisant le rapport f_m/f_{NL} et l’ensemble de proxies de site et du chargement sismique avec la base de données « événements ». f_{NLobs} représente f_{NL} observée.

Jeu de données événements (nombre d’enregistrements =3790)									
f_bin	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.077	0.028	0.024	0.039	0.065	0.084	0.098	0.145	0.113
Proxies	$(f_{NLobs}, V_{s30}, B_{30}, f_0, V_{s30}/V_{smin}, PGAs, PGVs, PGVs/V_{s30}, f_m/f_{NLobs})$								
σ	0.052	0.019	0.018	0.011	0.029	0.053	0.064	0.099	0.060
R_c	32.567	33.283	27.176	70.851	54.683	37.057	35.181	31.726	47.068

Les résultats obtenus semblent très intéressants, traduits par des valeurs de coefficients de réduction grandes allant jusqu’à 71 %. En revanche, pour obtenir la valeur de f_{NL} il faut avoir le rapport RSR_{NL-L} .

Pour éviter ce cercle vicieux entre f_{NL} et RSR_{NL-L} . Nous élaborons un modèle neuronal qui peut fournir la valeur de f_{NL} sans passer par le rapport RSR_{NL-L} . Les paramètres indépendants (inputs

du réseau de neurones) susceptibles d'être sensibles à f_{NL} sont V_{s30} , B_{30} , f_0 , V_{s30}/V_{smin} , PGA_s , PGV_s , PGV_s/V_{s30} .

Dans le tableau.IV.15 nous représentons les valeurs de σ et R_c pour chaque paramètre d'entrée (proxy de site) et pour différentes combinaisons de proxys de sites.

Table.IV.15 Valeurs de σ_{ref} , σ et R_c donnés par les f_{NL} prédites pour l'ensemble de proxys de site et du chargement sismique en utilisant la base de données « événements ».

Jeu de données « événements » (nombre d'enregistrements= 5760)	f_{NL}	
	σ	R_c
σ_{ref}	0.316	-
Proxies de site		
V_{s30}	0.240	23.866
f_{0HV}	0.262	16.905
V_{smin}	0.309	2.1505
V_{s30}/V_{smin}	0.284	10.249
B_{30}	0.293	7.2680
PGA_s	0.316	0.000
PGV_s	0.311	1.563
PGV_s/V_{s30}	0.311	1.553
$V_{s30}+f_{0HV}$	0.227	28.277
$V_{s30}+B_{30}$	0.229	27.558
$V_{s30}+V_{smin}$	0.232	26.627
$V_{s30}+B_{30}+f_{0HV}$	0.224	28.969
$V_{s30}+B_{30}+f_{0HV}+V_{smin}$	0.223	29.295
$V_{s30}+B_{30}+f_{0HV}+V_{smin}+V_{s30}/V_{smin}$	0.213	32.730
SPCs +PGA_s+PGV_s+PGV_s/V_{s30}	0.203	35.773

Les résultats obtenus (tableau.IV.15) confirment que f_{NL} ne dépend pas des proxys du chargement sismique (ici PGA_s et PGV_s) figure.IV.19. En outre, les proxys de sites qui influent le plus sur la variabilité aléatoire liée à f_{NL} sont V_{s30} et f_{0HV} . Il est à noter que nous avons effectué le même test avec la base de données site avec un PGA_s seuil =100 cm/s² et nous avons trouvé un $\sigma=0.275$ plus grand que celui de la base de données événements.

À présent, nous allons voir si l'utilisation de f_{NL} prédits par la fonction $f_{NL}(V_{s30}, B_{30}, f_{0HV}, V_{smin}, V_{s30}/V_{smin})$ converge vers les $RSRE_{NL-L}$ obtenus par les f_{NL} observés. Dans le tableau.IV.16 nous donnons (σ , R_c) calculés par les f_{NL} prédites.

Table.IV.16 Valeurs de σ_{ref} , σ et R_c donnés par les $RSRE_{NL-L}$ prédites adimensionnels normalisés en fréquences pour, en plus, de l'ensemble de proxys de site le et les proxys du chargement sismique en utilisant la base de données « événements ». f_{NLpred} représente f_{NL} prédite.

Jeu de données événements (nombre d'enregistrements =3790)									
f_Bin	0.2	0.3	0.6	1.2	1.9	3.2	5.6	9.9	17.2
	-	-	-	-	-	-	-	-	-
	0.3	0.6	1.1	1.9	3.2	5.6	9.9	17.2	30.0
σ_{ref}	0.077	0.028	0.024	0.039	0.065	0.084	0.098	0.145	0.113
Proxies	(f_{NLpred}, V_{s30}, B_{30}, f_0, V_{s30}/V_{smin}, $PGAs$, $PGVs$, $PGVs/V_{s30}$, f_m/f_{NLpred})								
σ	0.063	0.023	0.021	0.031	0.043	0.061	0.071	0.098	0.0816
R_c	19.371	19.856	13.374	21.437	34.160	27.038	28.220	32.240	27.9784

En comparant les R_c du tableau.IV.14 (où les f_{NL} observées sont utilisées) et ceux du tableau.IV.16 (f_{NL} prédites), nous remarquons que, la nature du proxy f_{NL} (observée ou prédite) influe sur les (σ, R_c) des $RSRE_{NL-L}$, nous assistons si nous utilisons le f_{NLpred} à une augmentation de σ et diminution de R_c . Mais malgré cela, les résultats avec des f_{NL} prédites à partir des (V_{s30} , B_{30} , f_{0HV} , V_{smin} , V_{s30}/V_{smin}) restent satisfaisants.

IV.14. Conclusion de la 2^{ème} partie

Une grande partie a été consacrée à la présentation du rapport spectral : (mouvement fort)/(mouvement faible) qui peut être utilisé dans l'estimation des effets de site linéaires et non linéaires pour les GMPEs, SAPE (Spectral Amplification Prediction Equation), AF (Amplification factor), pour l'étude du site spécifique et semi-spécifique et pour l'analyse de la variabilité aléatoire. Le but de cette partie est de quantifier la variabilité du RSR_{NL-L} de point de vue comportement du sol : comportement linéaire et non linéaire. Avec l'approche neuronale et la transformation boxcox, nous avons essayé d'effectuer une analyse de sensibilité des paramètres qui peuvent influencer sur la variabilité aléatoire des RSR_{NL-L} . ladite analyse a été réalisée par le biais de 4 bases de données. Dans les 3 premières bases de données, nous n'avons considéré qu'une RSR_{NL-L} médiane par site avec un seuil minimum de PGA_s égal à 100, 200 et 300 cm/s^2 respectivement. Par la suite, et pour prendre en compte le chargement sismique, nous avons introduit une 4^{ème} base de données nommée « événements » : une $RSRE_{NL-L}$ par site et par enregistrement.

La première étape était de choisir entre le $RSR_{NL-L,R}$ et $RSR_{NL-L,F}$. Il en résulte de cette étape que le $RSRE_{NL-L}$ calculé à partir du spectre de Fourier donne des écarts types plus faibles que

celui fondé sur le spectre de réponses. Et que la tendance de $RSRE_{NL-L,F}$ semble comparable à celle donnée par la GMPE de la première partie de ce chapitre.

Nous avons enchaîné par l'étude de la pertinence des proxys de sites, et comment ces derniers peuvent participer à la réduction de la variabilité aléatoire. Le résultat le plus important est que, la pertinence des proxys de site est liée au seuil de PGA en surface. Un autre résultat montre que le f_{NL} est plus pertinent que les autres proxys de site et que la pertinence des autres proxys augment avec le seuil du PGA_s . Donc, pour avoir une bonne prise en compte de la variabilité aléatoire nous avons intérêt à utiliser les 5 proxys de site, à savoir V_{s30} , V_{smin} , f_{0HV} , B_{30} et surtout f_{NL} .

Les 5 proxys de site ont été donc adoptés pour l'étude effectuée sur la pertinence du proxy lié au chargement sismique avec la base de données événements. La réduction de σ obtenue par cette base de données est plus faible que celle de la base de données sites. Car avec la base de données événements nous avons ajouté en plus de l'incertitude site-site, l'incertitude enregistrement-enregistrement. Pour contourner ce problème, nous avons essayé de réduire la variabilité site-site par l'introduction en input du rapport f_m/f_{NL} . Cela a donné une réduction importante allant jusqu'à 70 % pour l'intervalle de fréquence compris entre 1.2 Hz et 1.9 Hz.

Par la suite, et pour contourner le problème de la disponibilité de f_{NL} nous avons construit un autre modèle neuronal capable de prédire le f_{NL} . La première constatation confirme, le fait que, le niveau du mouvement sismique n'a aucune influence sur le f_{NL} et que les proxys de site suffisent à eux seuls à la prédiction du f_{NL} et que l'utilisation de f_{NL} prédit pour estimer les RSR_{NL-L} donne des résultats satisfaisants (ex. $R_c=[19-34]\%$).

Conclusion

À la fin de ce chapitre nous pouvons conclure que, l'utilisation d'un seul proxy de site pour décrire son comportement via une GMPE peut être insuffisante. L'emploi de plusieurs proxys peut réduire davantage la variabilité aléatoire même s'il s'agit d'un proxy comme la pente topographique. Là aussi comme le chapitre précédant le type de proxy n'influe par vraiment sur la médiane. Dans les deux parties, nous avons constaté que le V_{s30} et f_0 sont des proxys pertinents. Plus de pertinence a été remarquée avec le f_{NL} lors de l'analyse des RSR_{NL-L} . Le modèle RSR_{NL-L} , que nous puissions obtenir avec un modèle totalement conduit par les données semble très intéressant pour les gmpistes et ce pour décrire les résidus site-site (δ_{s2s}) par une base de données sites et les résidus événement-événement (δ_{e2e}) par une base de données événements en utilisant des RSR_{NL-L} adimensionnels.

A partir de la base de données site nous pouvons construire le modèle suivant :

$$RSR_{NL-L}(f) = G(V_{s30}, f_0, f_{NL}, V_{smin}, B_{30}) \pm \delta_{s2s}(f) \quad (IV.14)$$

Et via la base de données évènements, le RSR_{NL-L} peut être obtenu par :

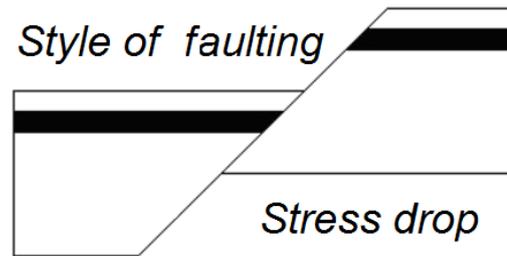
$$RSR_{NL-L}\left(\frac{f}{f_{NL}}\right) = G(V_{s30}, f_0, f_{NL}, V_{smin}, B_{30}, PGA_s, PGV_s) \pm \delta_{e2e}(f) \pm \delta_{s2s}(f) \quad (IV.15)$$

Les deux équations IV.14 et IV.15 peuvent être utilisées à la place des formes fonctionnelles utilisées habituellement dans les GMPEs pour décrire l'effet de site linéaire et non linéaire.

En revanche, pour avoir une GMPEs robuste avec des incertitudes faibles, il ne faut pas se limiter à étudier seulement les effets de site, mais aussi analyser l'effet lié à la source sismique (ex. mécanisme au foyer et chute de contrainte). C'est ce que nous allons aborder dans le dernier chapitre V.

Chapitre V

Etudes des effets de la source sur le mouvement sismique



Introduction

Dans les chapitres III et IV, nous avons essayé d'étudier la façon avec laquelle le ou les différents proxies de site participent à la prise en compte des effets de site linéaire et non linéaire et à la réduction de la variabilité aléatoire. Cette réduction touche, spécialement, la partie intra-événement « within-event », qui influe grandement sur la variabilité aléatoire totale du mouvement sismique. Cependant, l'influence du « between-event » (la variabilité inter-événement) sur la réduction de la variabilité totale est à ne pas négliger. Il est nécessaire, donc, de se focaliser aussi sur les paramètres de la source sismique susceptibles de réduire davantage la variabilité aléatoire. Il existe plusieurs proxies importants caractérisant la source sismique autre que la magnitude, qui est, cependant, un paramètre important, essentiellement, à des basses fréquences.

Dans le présent chapitre nous nous intéressons essentiellement au style de la faille (mécanisme au foyer : FM) et à la chute de contrainte sismique ($\Delta\sigma$). Afin de se focaliser seulement sur l'effet de la source sismique, nous n'effectuons l'analyse que sur des sites fermes et rocheux et pour des distances ne dépassant pas les 80 km pour l'étude du FM et 60 km pour $\Delta\sigma$. La raison pour laquelle nous choisissons un seuil de 80 km au lieu de 60 km dans la première étude est purement liée à la quantité des données disponibles dans RESORCE qui s'est avérée non représentative pour des $R_{JB} < 60$ km.

L'étude de l'effet FM est effectuée avec trois jeux de données : RESORCE, NGA-West 2 et KiK-Net. Pour ce faire, nous comparons la différence existante entre failles normales, inverses et de décrochement, en termes de leurs influences sur le mouvement sismique (MS) enregistrée à la surface. Cela est réalisé par le coefficient de réduction de la variabilité aléatoire et par les tendances des médianes. Cela permet de mesurer, ainsi, la faisabilité d'introduire le style de la faille comme un proxy de la source sismique dans les modèles de prédiction du mouvement sismique.

L'étude de la dépendance entre la chute de contrainte sismique et la magnitude est effectuée indirectement par le biais du MS. Dans cette partie de chapitre, nous utilisons principalement le jeu de données NGA-West 2.

Le long de ce chapitre, les MS sont représentées par l'accélération maximale du sol (PGA), la vitesse maximale du sol (PGV) et les Pseudo-Spectres d'Accélération (PSA) pour un amortissement égal à 5 %. La prédiction de ces MS est effectuée toujours avec la méthode des réseaux de neurones artificiels.

Nous commençons notre investigation en étudiant l'influence du type de la faille sismique (FM) sur le MS enregistrées à la surface libre, et ce pour différentes régions : [Italie, Turquie], [Californie, Taiwan] et le Japon. Ces régions sont présentées, respectivement par les jeux de données RESORCE, NGA-West 2 et KiK-Net.

1^{ère} partie : Etude de l'effet du style de la faille sur la variabilité aléatoire du mouvement sismique par l'approche neuronale : application sur RESORCE, NGA-West 2 et KiK-Net

Style-of-Faulting ground-motion variability and data-driven GMPEs. Insights from RESORCE, NGA-West 2 and KiK-Net

Proposed for published

Boumediène Derras,^{a), b), c)} Fabrice Cotton,^{d), f)} and Pierre-Yves Bard,^{c)}

Résumé

Le but de cette étude est d'étudier l'effet du style de la faille (mécanisme au foyer : FM) sur la variabilité aléatoire du mouvement sismique en utilisant des modèles de prédiction du mouvement sismique issus des trois bases de données : RESORCE, NGA-West 2 et KiK-Net. Pour éviter l'effet local de site et l'effet régional, nous utilisons des jeux de données avec des sites durs ($V_{s30} \geq 450$ m/s) et à une distance R_{JB} qui ne dépasse pas les 80 km. Pour obtenir des modèles totalement conduits par les données, l'approche neuronale est utilisée. Deux familles de modèles sont élaborées : des modèles de référence avec comme seuls paramètres

^{a)} Risk Assessment and Management laboratory (RISAM), University of Tlemcen, Algeria

^{b)} Department of Civil Engineering and hydraulic, University of Saïda, Algeria

^{c)} Institut des Sciences de la Terre (ISTerre), University of Grenoble Alpes, CNRS, IFSTAR, Grenoble, France.

^{d)} Helmholtz Centre Potsdam, German Research Center for Geosciences (GFZ), Potsdam, Germany

^{f)} Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

indépendants M_w et R_{JB} , une deuxième famille de modèles obtenus en ajoutant à ces deux paramètres, ceux du style de la faille. La comparaison entre ces deux types de modèles permet d'analyser et de quantifier l'effet du style de la faille sur la variabilité aléatoire et sur la médiane. Trois types de failles sont adoptées, à savoir FM_{SS} : faille de décrochement, FM_N : faille normale et FM_R qui désigne les failles inverses. Les paramètres du mouvement sismique prédits sont les PGA, PGV et les PSA [0.01 à 2] s pour un amortissement égal à 5 %. Seuls les séismes crustaux sont considérés. Les résultats obtenus montrent que l'impact du FM est négligeable sur la variabilité inter-événements du modèle issue de la base de données NGA-West 2 et est non négligeable pour les modèles issus des bases RESORCE et KiK-Net. Ainsi le taux de réduction (R_c) pour ce dernier modèle est de l'ordre de 15 %. La prise en compte de la profondeur focale permet d'avoir un $R_c=25$ %. Les rapports entre les amplitudes des spectres de réponse associés aux failles normales et aux failles de décrochements ($F_{N:SS}$) d'une part et des failles inverses et de décrochement ($F_{R:SS}$) d'autre part, donnent les résultats suivants : $F_{SS} > F_N$ pour les deux modèles RESORCE et NGA-West 2. $F_{SS} < F_N$ pour le modèle avec KiK-Net. L'impact de F_N augmente avec la distance et diminue avec la profondeur focale. Par ailleurs $F_R \cong F_{SS}$ pour l'ensemble des modèles. Nous constatons aussi que, le fait d'utiliser un jeu de données KiK-Net contenant des distances supérieures à 80 km et des $V_{s30} < 450$ m/s (effet régional et de site respectivement) cache l'effet réel du FM sur le mouvement sismique.

V.1. Introduction

Après la magnitude (M), la distance (R) et le/les proxies de site (ex. V_{s30}), la variable explicative la plus souvent incluse dans les GMPEs est le style de la faille (ou mécanisme au focal : FM). [Bommer et al. \(2003\)](#) montrent que, le FM est une des caractéristiques de la source sismique. Ces auteurs concluent que, l'introduction du FM comme proxy de source dans des GMPEs transforme progressive la variabilité aléatoire en variabilité épistémique. Les FMs peuvent être déterminés à partir de la configuration tectonique, des contraintes crustale, des plans de faille et des ruptures en surface. Les techniques de télédétection sont également devenues des outils puissants pour identifier les failles et leurs mécanismes. Des GMPEs qui représentent explicitement les FMs, devraient produire des estimations de l'aléa sismique plus fiables, toujours selon [Bommer et al. \(2003\)](#).

Plus récemment, les GMPEs utilisant la base de données NGA-West 2 ont un facteur qui tient compte du FM. Les modèles d'[Abrahamson et al. \(2014\)](#), [Campbell and Bozorgnia \(2014\)](#), [Chiou and Youngs \(2014\)](#) proposent de réduire ce facteur à zéro pour les magnitudes comprises

entre [4.0-4.5]. [Boore et al. \(2014\)](#) et [Idriss, \(2014\)](#) supposent que le FM est indépendant de la magnitude.

En parallèle, les GMPEs obtenues avec la base de données RESORCE : [Akkar et al. \(2014\)](#), [Bindi et al. \(2014\)](#), [Derras et al. \(2014\)](#), [Hermkes et al. \(2014\)](#) et [Bora et al. \(2014\)](#) comprennent toutes un terme de style des failles. En revanches, les tendances obtenues sont différentes d'un modèle à un autre. Dans la figure.V.1 nous exposons la comparaison qui a été effectuée par [Douglas et al. \(2014\)](#) entre ces 5 GMPEs en plus de celui de [Boore and Atkinson \(2008\)](#). Dans cette figure nous remarquons que les $F_{R:SS}$ et $F_{N:SS}$ donnés par les deux modèles de [Derras et al. \(2014\)](#) et [Hermkes et al. \(2014\)](#) dépendent de R_{JB} et M_w contrairement aux autres modèles où $F_{R:SS}$ et $F_{N:SS}$ dépendent seulement de la période. Ce non-consensus peut être dû, entre autres, à la manière dont le FM a été défini.

Pour éviter ce problème, nous adoptons, dans la présente étude, la même définition de FM pour les trois bases de données RESORCE, NGA-West 2 et KiK-Net. Cette définition est basée sur l'angle Rake (λ). Ce dernier est représenté schématiquement par la Figure.V.2. Les intervalles en matière de λ et qui servent à la détermination du type de la faille sont données par le Tableau.V.1.

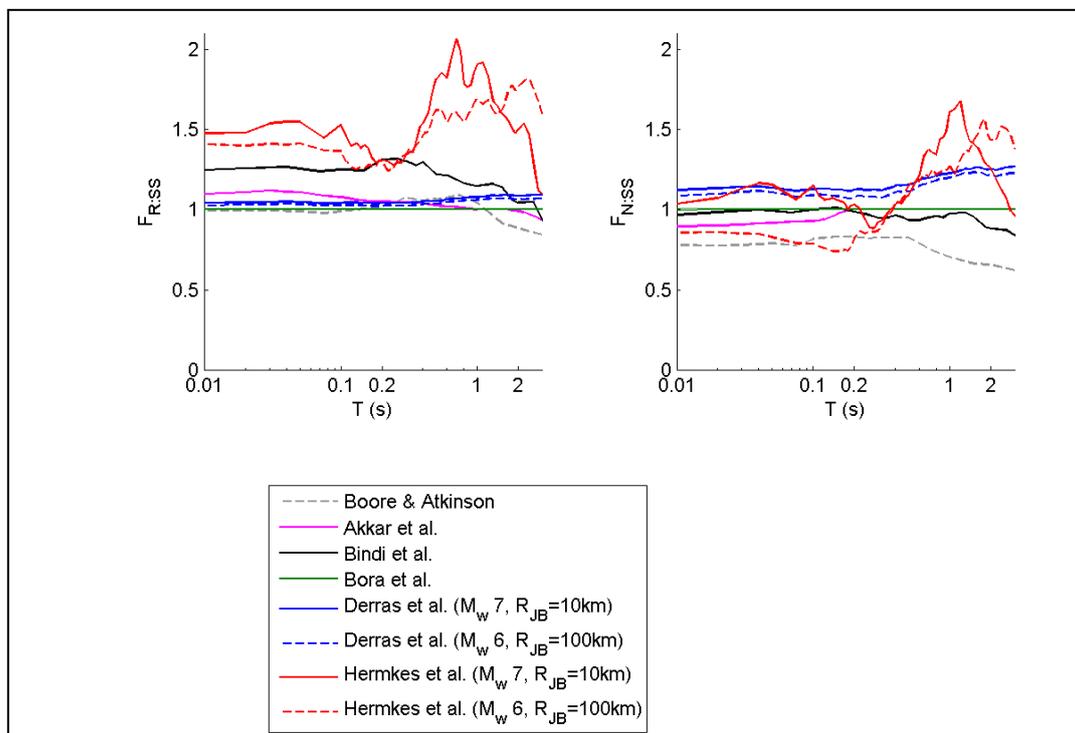


Figure.V.1 Comparaison entre les facteurs FM des 6 GMPEs. À droite) : rapport des PSA : faille inverse sur faille de décrochement. À gauche) rapport des PSA : faille normale sur faille de décrochement.

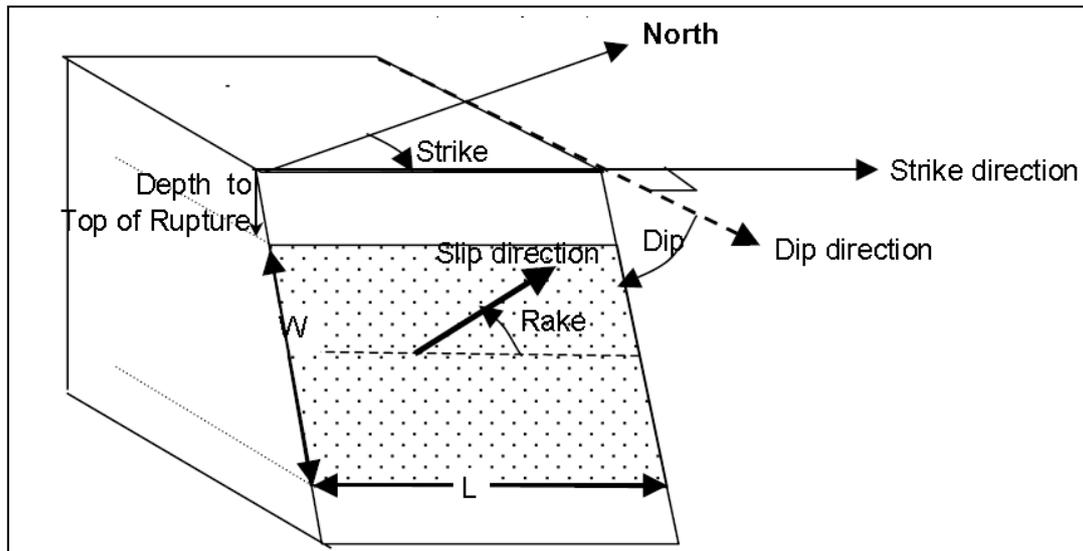


Figure.V.2 Vue schématique d'une faille sismogénique avec les différents paramètres caractérisant la géométrie, le déplacement et les orientations de la faille.

Le but de cette étude est de mesurer le taux de réduction de la variabilité aléatoire par l'introduction aux GMPEs des informations liées aux styles de la faille. Seulement trois styles de failles sont adoptés : normal, inverse et décrochant.

Nous commençons notre investigation par exposer les jeux de données utilisées au sein de cette étude. Puis nous enchaînons avec un aperçu sur l'approche neuronale utilisée à l'élaboration des équations de prédiction. L'ensemble des résultats obtenus (par les trois jeux de données), ont comme premier objectif, de mesurer l'impact de l'utilisation du FM sur la variabilité aléatoire et le taux de réduction de cette variabilité en utilisant les composantes : σ inter-événements (τ), σ intra-événement (ϕ) et σ total (σ). Le deuxième objectif est de mesurer l'influence, toujours du FM, sur la médiane de la GMPE.

Table.V.1 Critère de classement des trois styles de la faille en utilisant l'angle entre la direction de glissement et l'horizontale (Rake)

Style de la faille	Rake (λ)
Normal	$-150 \leq \lambda \leq -30$
Inverse (reverse)	$30 \leq \lambda \leq 150$
Décrochement (strike-slip)	$-180 \leq \lambda \leq -150$, $-30 \leq \lambda \leq 30$ et $150 \leq \lambda \leq 180$

La présente étude se divise en deux principaux axes. En premier, lieu et pour diminuer la contribution de l'effet de site et l'effet régional, nous proposons des jeux de données avec une vitesse V_{s30} minimale égale à 450 m/s et une distance R_{JB} maximale qui vaut 80 km. Par la suite, et pour voir l'interaction qui peut exister entre le niveau du mouvement sismique donné par les FMs et les autres effets tels que l'effet régional et l'effet de site, nous menons une analyse avec deux modèles. Dans les deux nous utilisons le maximum de proxies : R_{JB} , $\log_{10}(R_{JB})$, M_w ,

Depth, $\log_{10}(V_{s30})$, $\log_{10}(f_0)$, $\log_{10}(\text{Pente})$ en plus de FM_{SS} , FM_N et FM_R . Le premier modèle est élaboré avec toute la gamme de distances et de vitesses. Et le deuxième représente une restriction en termes de distances et vitesses. Cette partie est effectuée avec un jeu de données KiK-Net qui offre la possibilité d'avoir le maximum de proxies.

V.2. Jeux de données

Nous utilisons dans cette étude des bases de données de qualité avec des événements récents. Les jeux de données adoptés ici nous donnent la possibilité d'avoir le PGA et PGV, les PSA pour un amortissement égal à 5 % et pour $T=[0.01-2.0]$ s ainsi que, les métadonnées R_{JB} , M_w et les trois styles de failles : normale, inverse et de décrochement. Pour les conditions de site nous prenons, en premier lieu, seulement les sites dont $V_{s30} \geq 450$ m/s. Pour diminuer l'effet régional nous choisissons une distance $R_{JB} \leq 80$ km. À partir de ces données, nous essayons de prédire le MS à la surface libre et de voir l'influence du style de la faille sur ces quantités. Les jeux de données utilisées pour réaliser cette étude sont : RESORCE ([Akkar et al. 2014](#)), NGA-West 2 ([Ancheta et al. 2014](#)) et KiK-Net ([Dawood et al. 2016](#)). Les informations relatives à chaque jeu de données sont exposées sur le tableau.V.2 et V.3. Dans le tableau.V.2, nous donnons le nombre d'événements, d'enregistrements et de sites ainsi que la valeur minimale et maximale de M_w et de R_{JB} pour les trois sous-ensembles de données. Dans le tableau.V.3, le nombre brut et en % de chaque style de faille par jeu de donnée sont illustrées. De point de vue type de faille, nous pouvons remarquer que la faille normale représente plus de 57 % pour RESORCE, 60 % des failles de NGA-West 2 sont de type décrochantes et 53 % des failles de KiK-Net sont de mécanismes inverses. La distribution de la magnitude et du PGA en fonction de la distance est donnée par la Figure.V.3 et la Figure.V.4 respectivement. Il est à remarquer que le nombre d'enregistrements pour une distance inférieure à 10 km est faible, surtout pour le jeu de données RESORCE et KiK-Net.

Comme complément d'information, nous donnons la distribution du style de failles par jeu de données (Figure.V.5). Cela permet de voir le niveau du MS pour chaque type de FM et la dépendance existante entre FM et la distance. Les tendances sont plus visibles pour les données de NGA-West 2. Nous remarquons, ainsi, que les PGA enregistrés par des failles inverses et décrochantes sont plus grands que ceux donnés par une faille normale.

Table.V.2 Nombre d'événements, d'enregistrements, de sites. Les intervalles de magnitude et distance pour les jeux de données RESORCE, NGA-West 2 et KiK-Net considérés dans cette étude

Jeux de données	Nombre total d'événements	Métadonnées	Intervalle des paramètres		Nombre total d'enregistrements	Nombre total de sites
			Min	Max		
RESORCE	224	M_w	3.7	7.6	438	107
		R_{JB} (km)	0.01	80.00		
NGA-West 2	304	M_w	3.05	7.9	3695	584
		R_{JB} (km)	0.01	79.96		
KiK-Net	543	M_w	3.6	6.9	2844	198
		R_{JB} (km)	0.73	80.00		

Il est à noter que dans le jeu de données RESORCE 53% d'enregistrements appartiennent à l'Italie et 26% à la Turquie. Tandis que, la majorité d'enregistrements de NGA-West 2 sont Californiennes et 16% sont enregistrés en Taiwan (Figure.V.6). Ces sous-ensembles de données représentent la base de nos modèles neuronaux. Dans ce qui suit, nous allons décrire la méthode adoptée pour la mise en œuvre de ce type de modèles.

Table.V.3 Nombre d'enregistrements (brut et en pourcentage) par style de faille de décrochement (SS), normal (N) et inverse (R) pour les jeux de données RESORCE, NGA-West 2 et KiK-Net utilisés dans cette étude

Jeux de données	Appellation	Nombre d'enregistrements par style de faille	% d'enregistrement par style de faille
RESORCE	FM_{SS}	108	24.65
	FM_N	253	57.76
	FM_R	77	17.58
NGA-West 2	FM_{SS}	2193	59.35
	FM_N	270	07.30
	FM_R	1232	33.34
KiK-Net	FM_{SS}	567	19.94
	FM_N	754	26.51
	FM_R	1523	53.55

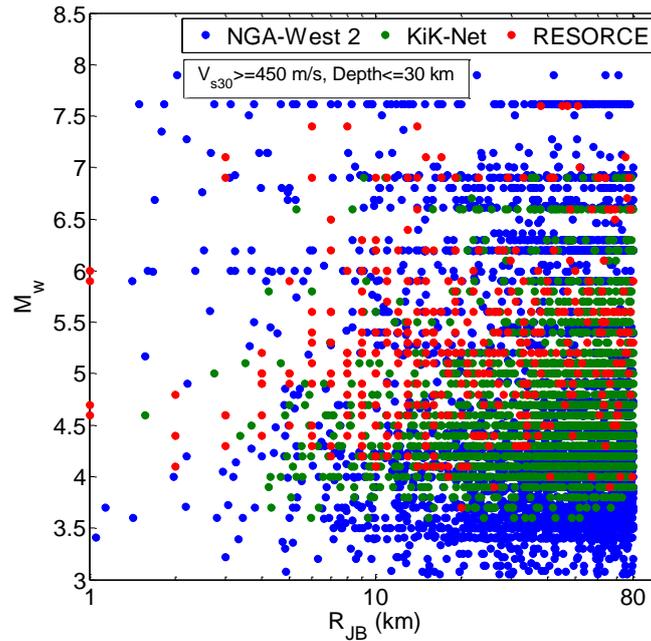


Figure.V.3 Distribution de la magnitude du moment en fonction de la distance Joyner et Boore pour les trois jeux de données.

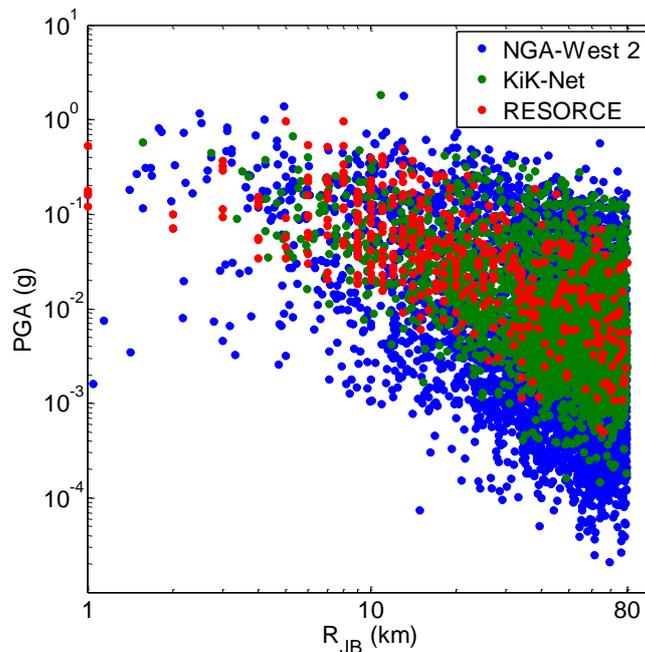


Figure.V.4 Distribution de l'accélération maximale du sol en fonction de la distance Joyner et Boore pour les trois jeux de données.

V.3. Methodologie

Il s'agit essentiellement d'élaborer des modèles permettant la prédiction du MS à la surface libre (PGA, PGV et PSA) à partir des 3 métadonnées à savoir R_{JB} , M_w et FM. Nous utilisons $\log_{10}(R_{JB})$ et R_{JB} comme variable indépendante décrivant l'atténuation géométrique et anélastique respectivement. L'effet de la source est représenté par la magnitude du moment M_w et les 3 styles possibles de faille : FM_{ss} pour le style de faille décrochante, FM_N pour la faille

normale et FM_R comme représentant de la faille inverse. Le tableau.V.4 représente la manière avec laquelle nous introduisons le style de la faille dans nos modèles. Il en résulte à la fin 6 entrées (Figure.V.7).

Table.V.4 Valeurs des variables binaires pour les trois types de failles.

Style de faille	FM_{ss}	FM_N	FM_R
Décrochante	1	0	0
Normale	0	1	0
Inverse	0	0	1

Dans la figure.V.7, nous représentons les variables à expliquer caractérisant le MS. Le procédé par apprentissage est utilisé ici pour avoir les modèles prédictifs du MS à partir des 6 proxies de source et de propagation d'onde.

L'approche classique en deux phases, utilisée pour l'élaboration d'un modèle reproductible qui permet de donner plus informations sur la variabilité aléatoire est adoptée dans cette étude. La première phase consiste à élaborer le modèle à effet fixe par l'approche neuronale et la deuxième phase comprend le modèle à effet aléatoire.

Modèle fixe

Le réseau de neurones de type Perceptron multicouche à une seule couche cachée est utilisé. Le nombre de neurones dans cette couche cachée est égal au nombre de variables d'entrée ([Wierenga and Kluytmans 1994](#)). La fonction d'activation de type Tangente hyperbolique est adoptée dans la couche cachée. Pour la couche de sortie, nous utilisons une fonction linéaire. Pour éviter le problème de surapprentissage, nous utilisons la technique de régularisation par modération des poids (weight decay). Il est à rappeler que la minimisation de la fonction de coût est effectuée par l'algorithme d'apprentissage de la rétropropagation du gradient de deuxième ordre quasi Newton (BFGS). Ce type de modèle est caractérisé par les connexions entre neurones (matrices des poids synaptiques : W_{ij}) et par les biais (vecteur : b_i) d'activation pour chaque neurone (Figure.V.7).

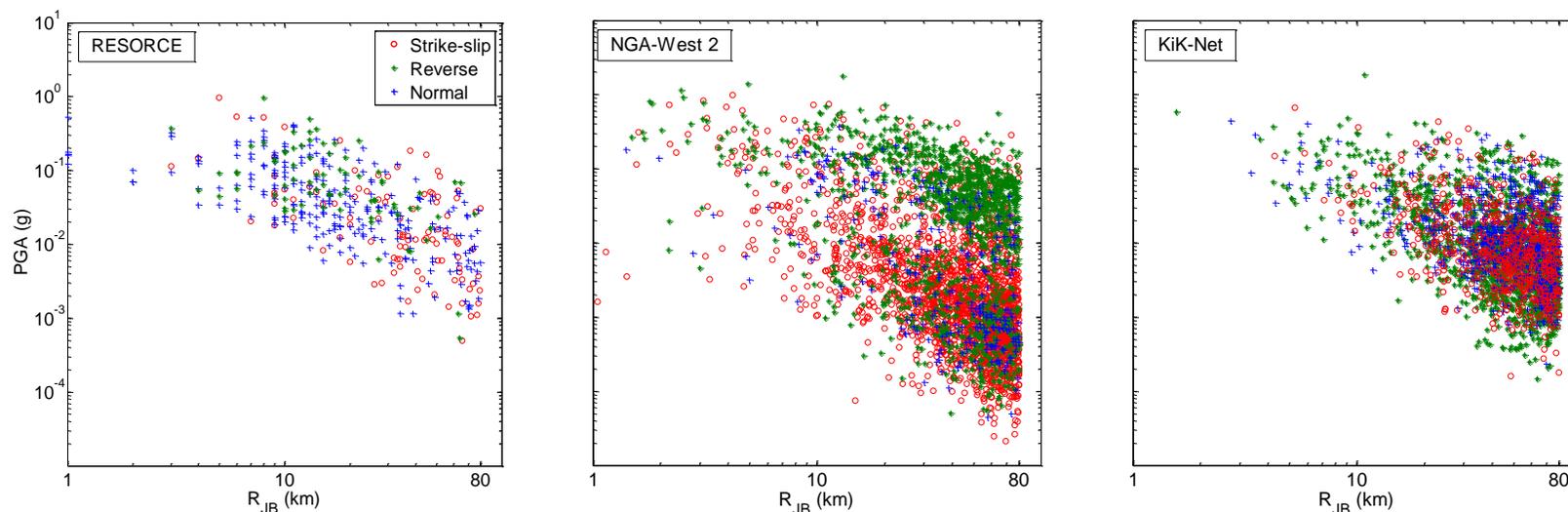


Figure.V.5 Distribution de l'accélération maximale du sol en fonction de la distance Joyner et Boore et du style de la faille pour les jeux de données RESORCE (à gauche), NGA-West 2 (milieu) et KiK-Net (à droite)

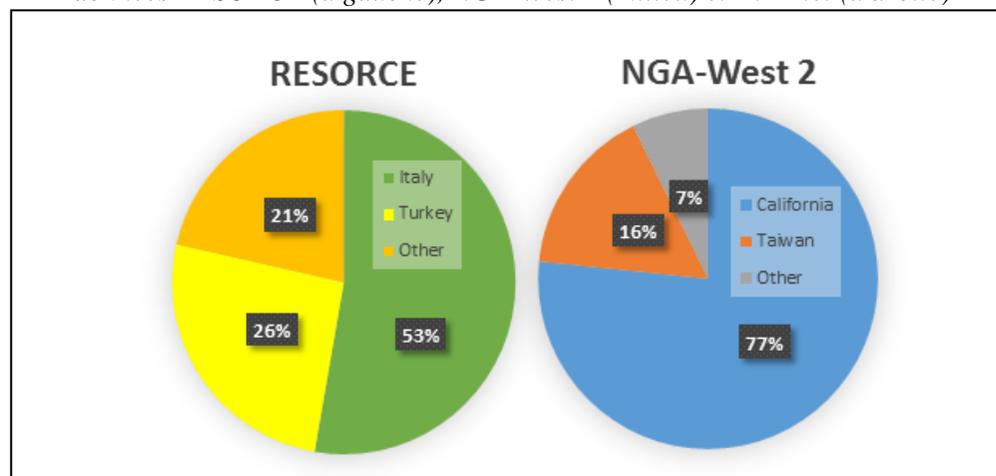


Figure.V.6 Répartition des pays par jeux de données

V.3.1. Modèle à effet aléatoire

Une procédure similaire à celle utilisée dans [Derras et al. \(2014\)](#) est reproduite dans le présent article. Nous utilisons l'approche par maximisation de vraisemblance développée par [Brillinger and Preisler \(1985\)](#) et stabilisée par [Abrahamson and Youngs \(1992\)](#). Il s'agit d'élaborer un modèle à effet aléatoire qui fournit deux quantités importantes à savoir les résidus inter-événements et intra-événements. Ces deux quantités, sont indépendantes, normalement distribuées, avec une moyenne nulle et un écart type inter-événements (τ) et intra-événements (ϕ), respectivement. La racine carrée de la somme aux carrés des deux écarts types est égale à l'écart type total (σ_i) pour une période donnée. Comme c'est représenté dans la formule suivante :

$$\sigma_i(T) = \sqrt{\tau(T)^2 + \phi(T)^2} \quad (\text{V.1})$$

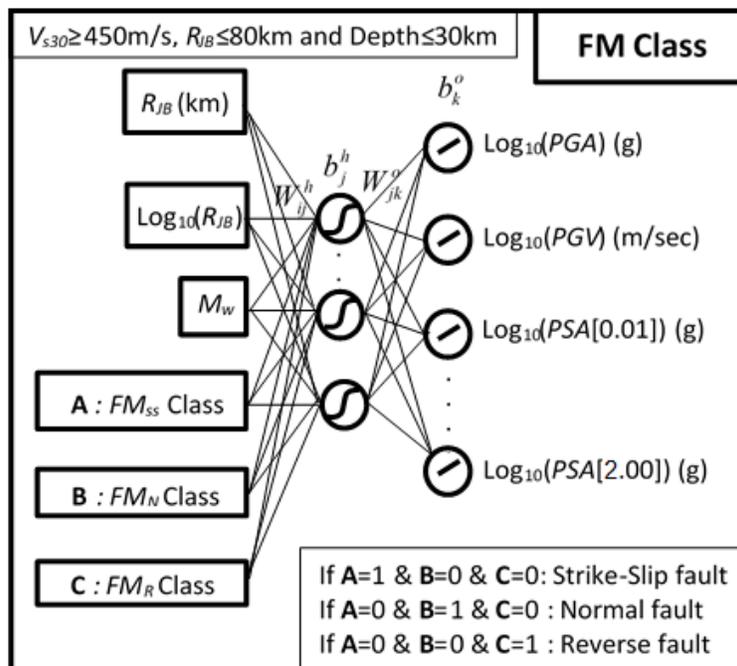


Figure.V.7 Structure des modèles neuronaux utilisés lors de cette étude pour la prédiction du mouvement sismique à la surface. Nous adoptons 6 entrées si le modèle prend en compte l'effet du style de la faille (avec 6 neurones dans la couche cachée). 3 entrées : R_{JB} , $\log_{10}(R_{JB})$ et M_w pour les modèles de références : sans effet du style de la faille dans ce cas précis nous choisissons 3 neurones cachés.

Pour quantifier l'apport de la prise en compte du mécanisme au foyer à la détermination du taux de réduction de la variable aléatoire dans les GMPEs, nous élaborons en premier lieu trois modèles (en utilisant les trois jeux de données) dits modèles de références dans lesquels les FM ne sont pas considérés. Par la suite nous élaborons les trois modèles qui intègrent les FMs comme proxy du mécanisme au foyer (Figure.V.7). En dernier lieu, et pour être cohérent avec les GMPEs classiques, nous construisons des

modèles sans restriction sur V_{s30} . Dans ces derniers nous incorporons la profondeur focale, et les trois proxies de site : V_{s30} , f_0 et la pente topographique. Les détails sont donnés plus loin dans ce chapitre.

Dans la section suivante nous examinons l'influence du style de la faille sur les deux quantités $\tau(T)$ et $\phi(T)$.

V.4. Comment la prise en compte du style de la faille peut-elle réduire la variabilité aléatoire du mouvement sismique

Dans cette section, nous analysons la manière avec laquelle le FM participe à la réduction de la variabilité aléatoire. Pour ce faire, nous comparons les modèles avec et sans la prise en compte de FM. Nous commençons par comparer les $\sigma_t(T)$, $\tau(T)$ et $\phi(T)$ (Figure.V.8). Premièrement, du point de vue variation des σ avec la période, nous remarquons que les modèles RESORCE et NGA-West 2 dépendent moins de la période en comparaison avec celui de KiK-Net. Pour ce dernier modèle, nous remarquons un fort pic autour de 0.15 s. Deuxièmement, le modèle RESORCE donne des valeurs plus petites de σ_t par rapport aux deux autres et ceux pour des périodes inférieures à 0.5 s, au-delà de cette valeur c'est les σ_t du modèle KiK-Net qui deviennent les plus faibles.

Le fait de prendre $V_{s30} \geq 450$ m/s et $R_{JB} \leq 80$ km, nous a permis d'éliminer partiellement l'effet de site et les effets des variations régionales. Nous focalisons notre investigation sur le $\tau(T)$. Dans ce contexte, toujours pour la figure.V.8, nous remarquons que les modèles RESORCE et KiK-Net donnent une réduction remarquable pour le τ inter-événements. En revanche, les valeurs avec et sans FM sont pratiquement les mêmes pour le modèle NGA-West 2.

Pour quantifier cette réduction, nous utilisons le coefficient de réduction (%) R_C donné par équation V.2. La présentation graphique de R_C en fonction de la période est donnée par la Figure.V.9.

$$R_c(T) = \left(1 - \frac{\sigma(T)_{With FM}}{\sigma(T)_{Without FM}}\right) \cdot 100; (\%) \quad (V.2)$$

Par le biais de ce coefficient R_c , nous pouvons quantifier la réduction de la variabilité aléatoire liée au style de la faille. En lisant la Figure.V.9, nous pouvons remarquer que le modèle issu de la base de données KiK-Net donne le taux de réduction le plus élevé (il est de l'ordre de 15 %) suivi par le modèle RESORCE avec une moyenne de réduction égale à 13.44 % et enfin le modèle-NGA-West 2 avec un R_c égal seulement à 2 %. Cela suggère que, l'influence de FM change d'une base de données à une autre et donc peut être d'une région à une autre. L'effet du FM au Japon ou en Italie est plus important que celui de la Californie. En outre, nous remarquons que, R_c est presque indépendant de la période en particulier pour le modèle NGA-West 2.

Une autre manière de mesurer l'importance de chaque proxy sur le MS, est d'utiliser le rapport (P%) entre la somme des éléments du vecteur des poids synaptiques $\{w_{xj}^h\}$ du proxy x et la somme des éléments de la matrice $[w_{ij}^h]$. Le tableau.V.5 donne les P% de chaque proxy d'entrée pour les trois modèles neuronaux avec FM. Ici nous calculons les P% de R_{JB} et $\log_{10}(R_{JB})$, cela permet de faire la différentiation entre le poids de l'atténuation anélastique et géométrique respectivement. En calculant P% nous mesurons l'importance globale des variables indépendantes. Ainsi pour RESORCE et NGA-West 2, l'atténuation anélastique a plus d'importance que l'atténuation géométrique, c'est l'inverse avec le jeu de donnée KiK-Net. En outre, il est à noter que, pour tous les modèles, les P%(FM_S) voisine 7% et P%(FM_R) \cong 9.5%. Tandis que, FM_N = 4.87 % pour les modèles de RESORCE et NGA-West 2 et de 12.29% pour le KiK-Net. Ces valeurs montrent encore une fois l'importance de FM pour les GMPEs obtenues à partir de la base de données KiK-Net.

Table.V.5 Sensibilité des modèles de prédiction du mouvement sismique à la distance, magnitude et aux styles de la faille par le pourcentage du poids synaptique (P %)

Modèle/Base de données	Atténuation anélastique R_{JB}	Atténuation géométrique $\log_{10}(R_{JB})$	M_w	Style de la faille		
				FM _S	FM _N	FM _R
RESORCE	28.73	15.36	34.09	7.35	4.88	9.58
NGA-West 2	28.73	15.38	34.08	7.35	4.87	9.59
KiK-Net	16.16	22.82	31.36	7.73	12.29	9.64

V.5. Influence du style de la faille sur la mediane

Comme pour la variabilité aléatoire, le style de la faille à un impact sur la forme de l'atténuation avec la distance surtout pour le modèle construit avec KiK-Net (Figure.V.10). Dans cette figure nous représentons les trois MS : PGA, PSA à T=0.1 s et à T=1.0 s en fonction de la distance. Là aussi les modèles utilisant RESORCE et KiK-Net sont plus sensibles au style de la faille que celui prédit par le NGA-West 2. Pour ce dernier cas, l'influence du FM sur le MS est négligeable (sauf pour T=1 s). Les formes fonctionnelles et les intensités engendrées par FM_R et FM_{SS} convergent pour T=0 s et à T=0.1 s. Le seul écart entre F_N et F_{SS} est à T=1 s (FM_{SS} et FM_R)>FM_N.

Si nous concentrons notre investigation pour R_{JB} = [10-70] km (ce domaine représente là où il y a le plus de données) ; nous remarquons que, le MS avec FM_N du modèle KiK-Net est grande surtout pour T=0.0 s et T=0.2 s (même constatations avec P%(FM_N)=12.29 %). Par contre, pour le modèle (RESORCE), les PSA obtenus avec le mécanisme au foyer normal sont plus petits, ce résultat confirme les valeurs de P%(FM_N)=4.88 %.

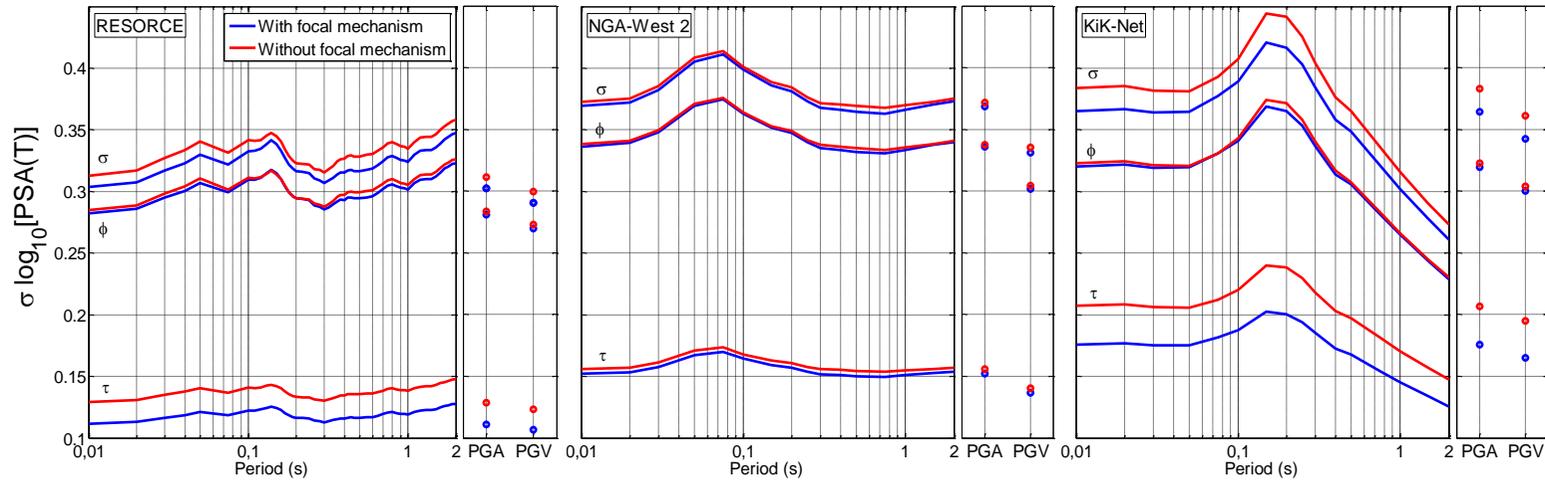


Figure.V.8 Sensibilité des composantes de la variabilité aléatoire (inter-événements (τ), intra-événements (ϕ) et l'écart type total (σ)) au style de la faille pour les trois modèles neuronaux élaborés à partir des trois jeux de données : RESORCE (à gauche), NGA-West 2 (au milieu) et KiK-Net (à droite). Ces composantes sont celles des ordonnées spectrales pour une période allant de 0.01 s à 2 s et pour le PGA et le PGV

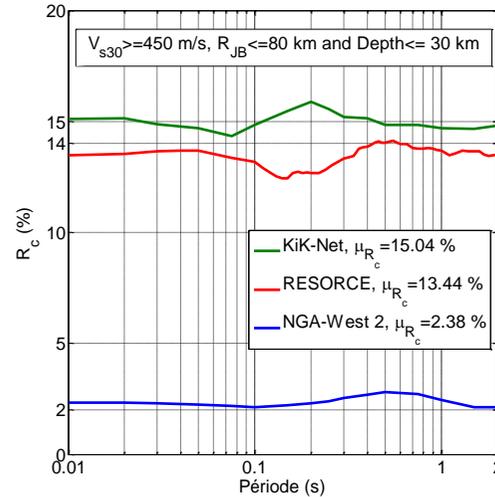


Figure.V.9 Quantification des performances du style de la faille par modèle neuronal avec la moyenne correspondantes (μ) La performance est mesurée avec le coefficient R_c (%) pour une période comprise entre 0.01 et 2 s.

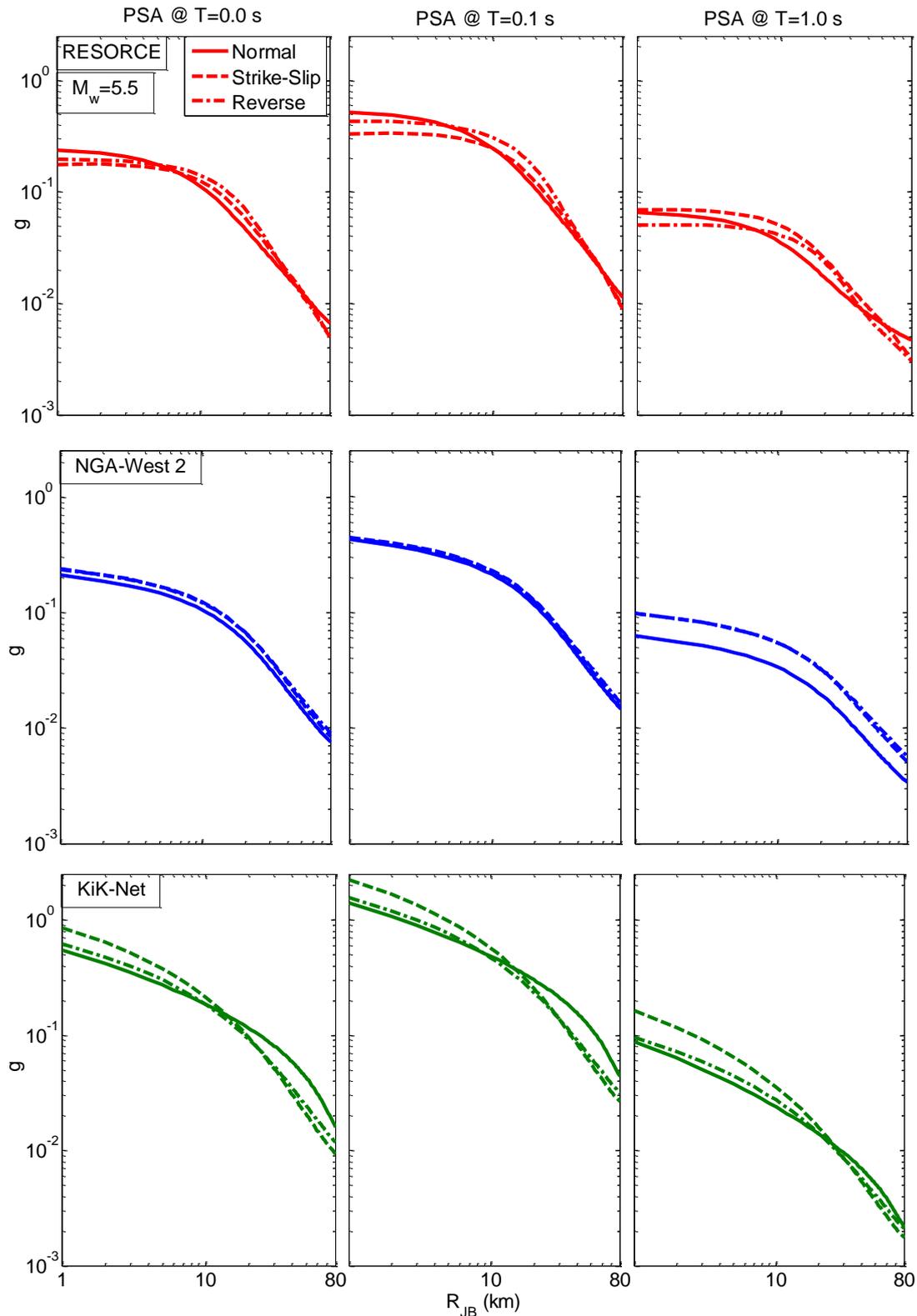


Figure.V.10 Effet du style de la faille sur la médiane pour les modèles : RESORCE, NGA-West 2 et KiK-Net à une magnitude égale à 5.5 pour T=0.0 (à gauche), T=0.1 s (au milieu) s et T=1.0 s (à droite).

Pour avoir plus de précision de l'impact des FM sur le MS, nous allons, dans la section suivante, comparer les facteurs $F_{N:SS}$ et $F_{R:SS}$ en fonction de la période.

V.6. Comparaison entre le MS des trois styles de failles

La comparaison est effectuée avec les deux rapports $F_{N:SS}$ et $F_{R:SS}$ donné par les deux formules suivantes ([Ambraseys et al. 2005](#)):

$$F_{N:SS}(T) = \frac{PSA_{FM_N}(T)}{PSA_{FM_{SS}}(T)} \quad (V.3)$$

$$F_{R:SS}(T) = \frac{PSA_{FM_R}(T)}{PSA_{FM_{SS}}(T)} \quad (V.4)$$

V.6.1. Dépendance des FMs avec la période

Dans cette section nous comparons les PSAs en utilisant les deux rapports suscités. Les trois modèles issus des bases de données RESORCE, NGA-West 2 et KiK-Net avec la prise en compte du style de la faille sont adoptés, ce qui donne 6 facteurs. Pour ce faire, nous choisissons un scénario donné : une magnitude égale à 5.5 et une distance qui vaut 30 km. Les tendances des facteurs F sont représentées sur la Figure.V.11. Nous remarquons ainsi que, les trois $F_{R:SS}$ des trois modèles convergent vers l'unité avec ± 0.15 pour RESORCE et KiK-Net et $\pm 5\%$ pour NGA-West 2, confirmant ainsi des niveaux du mouvement sismique proches engendrés par les deux types de failles, et ce pour l'ensemble du spectre. En revanche, les valeurs des PSAs produites par une faille de décrochement sont plus grandes que celles d'une faille normale surtout pour les périodes dépassant 0.1 s et ce pour les modèles RESORCE et NGA-West 2. De manière surprenante, le modèle réalisé avec le jeu de données KiK-Net fournit des $F_{N:SS}$ très importantes de l'ordre de 1.86 pour une période égale à 0.15 s, ce facteur $F_{N:SS}$ reste aussi important jusqu'à une période égale à 0.3 s.

2^{ème} partie : Impact du choix de la gamme de la distance sur l'effet du style de la faille : exemple de la base de données kik-Net

V.7. Introduction

Dans la première partie de ce chapitre, nous avons vu que le modèle le plus sensible au style de la faille est le modèle élaboré avec le jeu de données KiK-Net (un taux de réduction moyen R_c de l'ordre de 15 %). Pour cette raison, nous continuons la présente étude avec la base de données KiK-Net seulement. Cette fois-ci sans aucune restriction sur la distance et la vitesse V_{s30} . Le but est d'étudier l'impact de l'effet régional sur les PSA_{FM} .

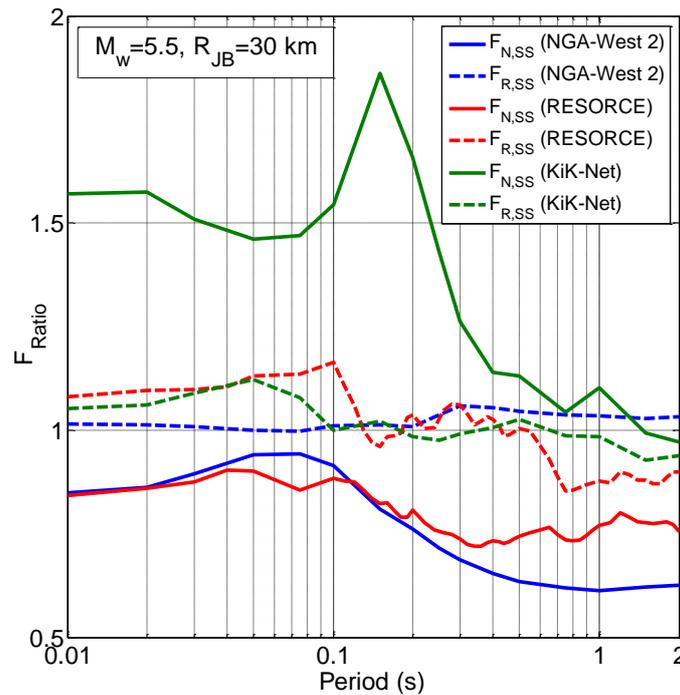


Figure.V.11 Rapports d'amplitude spectrale ($F_{N,SS}$ et $F_{R,SS}$) en fonction de la période pour les modèles élaborés à partir des jeux de données RESORCE, NGA-West 2 et KiK-Net, avec $M_w=5.5$ et $R_{JB}=30$ km.

À cet égard, nous choisissons comme proxys de site [V_{s30} , f_0 et la pente topographique]. En outre, nous adoptons le même type du modèle neuronal utilisé dans [Derras et al. 2014](#).

Quatre réseaux de neurones sont construits. Les deux premiers sont élaborés avec une distance $R_{JB}=[1-500]$ km, avec FM et sans FM nommés ANN(all R_{JB} ,FM) et ANN(all R_{JB}) respectivement. Les deux autres modèles sont construits avec une distance inférieure à 80 km (pour minimiser l'effet régional) toujours avec FM et sans FM baptisés ANN(80 R_{JB} ,FM) et ANN(80 R_{JB}) respectivement. Les métadonnées (les entrées) des deux modèles ANN(all R_{JB}) et ANN(80 R_{JB}) sont : R_{JB} pour l'atténuation anélastique, $\log_{10}(R_{JB})$ qui désigne l'atténuation géométrique, M_w , Depth (profondeur focale), [$\log_{10}(V_{s30})$, $\log_{10}(f_0)$, $\log_{10}(\text{Pente})$] comme proxys des conditions locales de site. Les modèles ANN(all R_{JB} ,FM) et ANN(80 R_{JB} ,FM) englobent, en plus, des métadonnées utilisées dans les deux premiers modèles, les trois termes binaires FM_{SS} , FM_N , FM_R caractérisant le mécanisme au foyer. Les PGA (g), PGV(m/s) et PSA[0.01-4]s représentent les paramètres du mouvement sismiques à prédire par les 4 modèles. La comparaison des résultats obtenus par ces 4 modèles permettent d'analyser l'impact du style de la faille sismique sur le MS pour les deux plages de distances $R_{JB} \leq 500$ km (avec effet régional : ER) et $R_{JB} < 80$ km (effet régional réduit). La base de données KiK-Net ([Dawood et al. 2016](#)) est encore adoptée ici pour l'élaboration des 4 modèles.

V.7.1. Jeu de données

Le jeu de données pour toute la gamme de distance contient 419 événements et 251 sites et 6317 enregistrements répartis sur 906 FM_{SS}, 2404 FM_N et 3007 FM_R. Dans le cas où $R_{JB} < 80$ km, le jeu de données comporte 328 événements, 180 sites et 1677 enregistrements distribués sur 297 FM_{SS}, 495 FM_N et 885 FM_R.

Nous commençons notre étude par l'analyse de l'impact de FM et de l'effet régional sur la réduction de la variabilité aléatoire.

V.8. Influence de la gamme de distance sur la réduction de la variabilité aléatoire menée par le style de la faille

La Figure.V.12 représente la variation des trois quantités (ϕ , τ et σ) en fonction de la période pour les 4 modèles suscités. Les ANN (AllR_{JB}) et ANN(AllR_{JB}, FM) sont représentés sur la partie de gauche. les ANN(80R_{JB}) et ANN(80R_{JB}, FM) sont illustrés sur la partie de droite.

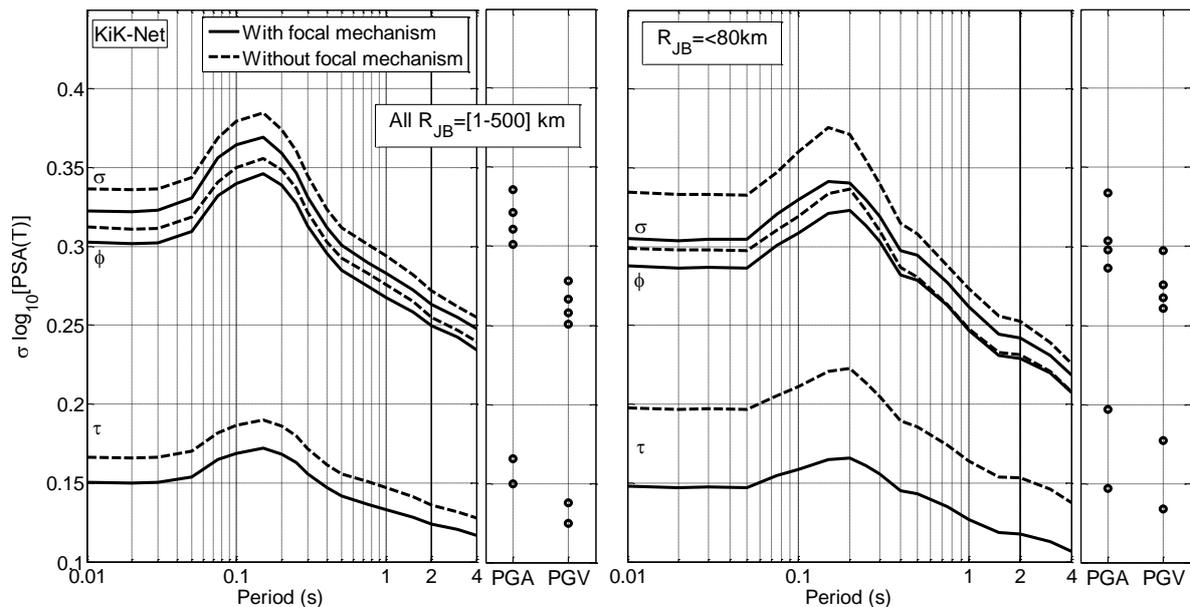


Figure.V.12 Sensibilité des composantes de la variabilité aléatoire (inter-événements (τ), inter-événement (ϕ) et l'écart type total (σ)) au style de la faille pour les 4 modèles neuronaux de prédiction du mouvement sismique élaborés à partir des 4 jeux de données KiK-Net : Pour toute la gamme de distance avec et sans FM (à gauche) et pour $R_{JB} < 80$ km avec et sans FM (à droite). L'ensemble des modèles comprennent les proxies de site et la profondeur focale. Les composantes de la variabilité aléatoire sont celles des ordonnées spectrales pour une période allant de 0.01 s à 4 s et pour le PGA et le PGV

Pour une distancer inférieur à 80 km et en introduisant le FM et les proxies de site dans le modèle neuronal nous diminuons considérablement l'incertitude et nous augmentons l'effet des FMs sur la variabilité aléatoire.

Ces résultats sont confirmés par le R_c (coefficient de réduction donnée par équation.V.2) exposé sur la Figure.V.13. Cette dernière illustre ledit coefficient vs la période pour les 4 modèles. Si nous utilisons toute la gamme de distance, nous obtenons une réduction de moins de 10 % sur le σ inter-événements. Se limiter à 80 km augmente l'effet de FM est réduit le τ de 25 % entre $T=[0.01-0.2]$ s et vers 22.5 % au-delà de $T=0.2$ s. Pourquoi dans la Figure.V.9 avons-nous trouvé une valeur de R_c plus faible (15 %) ? Dans les deux cas nous avons utilisé une base de données avec $R_{JB}<80$. La réponse à cette question est peut-être liée à la prise en compte de la profondeur focale (Depth) qui est un proxy de la source sismique et qui a un impact sur τ (Derras et al. 2012). Pour tester cette hypothèse, nous avons effectué un test sans introduire la profondeur et nous avons obtenus des R_c sur τ allant de [15.91% à 18.45%], ce qui représente pratiquement les valeurs obtenues par les modèles de la figure.V.9.

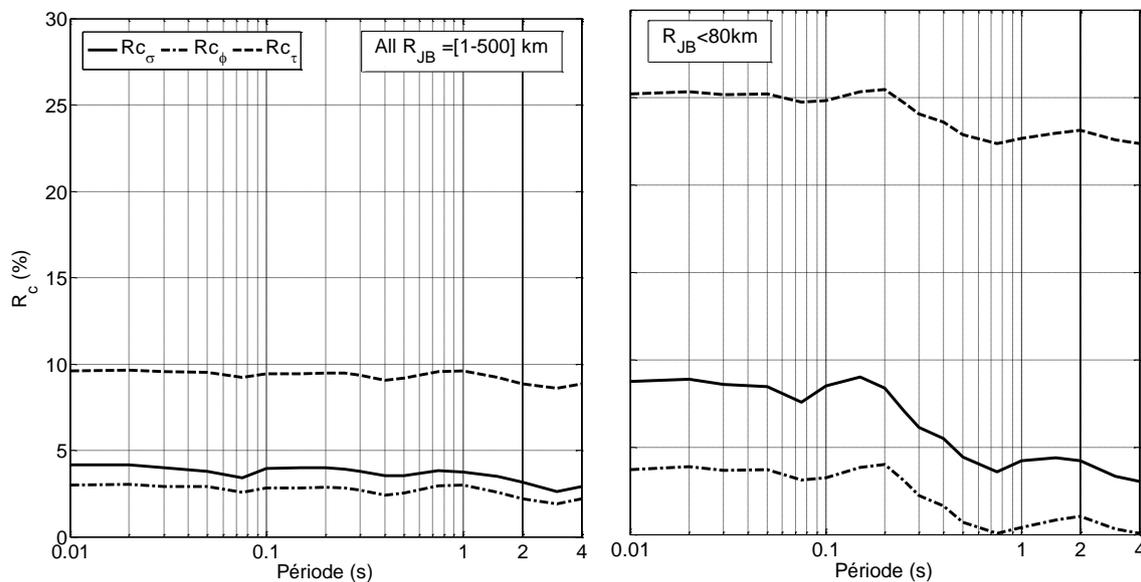


Figure.V.13 Quantification des performances du style de la faille par modèles neuronaux issus du jeu de données KiK-Net pour toute la gamme de R_{JB} (à gauche) et pour une distance inférieure à 80 km (à droite). Les performances sont mesurées par le coefficient R_c (%) pour une période comprise entre 0.01 et 4 s.

V.9. Dépendance entre le style de la faille et les paramètres de la source, de propagation d'onde et de la période structurale

Ici nous considérons seulement les deux modèles ANN(All R_{JB} ,FM) et ANN(80 R_{JB} ,FM) en vue de tester l'effet de FM (pour les deux gammes de distances $R_{JB}<500$ et $R_{JB}<80$ km) sur la médiane en illustrant les dépendances existantes entre FM et : la période, la magnitude, la distance et en fin la profondeur focale. Dans ce contexte, le rapport $F_{N:SS}$ et $F_{R:SS}$ (équation V.3 et V.4) sont utilisés. Dans nos analyses nous présentons la variation du rapport F avec les quantités ci-dessus (période...etc.) en choisissant un scénario donné.

Nous commençons notre analyse par voir la variation du rapport F avec la période (Figure.V.14). Nous remarquons que, si nous utilisons toute la gamme de distance, les rapports seront plus significatifs surtout pour $F_{N:SS}$. En revanche, les tendances des $F_{R:SS}$ sont pratiquement les mêmes dans les deux cas ($R_{JB} < 80$ km et all R_{JB}). Ces tendances sont analysées pour une distance égale à 30 km. Les valeurs des autres proxies sont choisies minutieusement : sont les valeurs correspondantes à 50 % de la fonction de densité cumulative (CDF) de chaque proxy, tout en assurant une compatibilité entre les valeurs de V_{s30} , f_0 et la pente topographique.

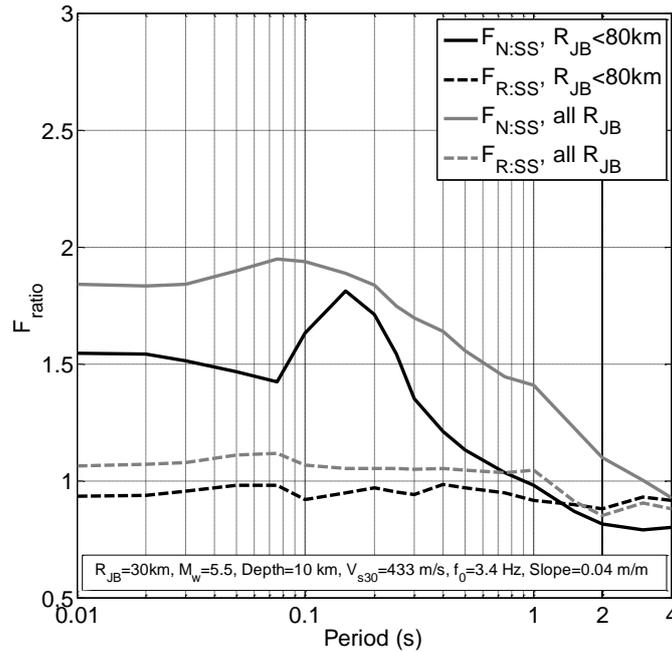


Figure.V.14 Rapports d'amplitude spectrale ($F_{N:SS}$ et $F_{R:SS}$) en fonction de la période pour les modèles élaborés à partir des jeux de données KiK-Net pour toute la gamme de R_{JB} et pour une distance inférieure à 80 km et sur un intervalle de période allant de [0.01-4]s.

V.9.1. Dépendance des FM avec la distance

Pour voir l'évolution de l'effet de FM avec la distance, nous traçons les graphes de F_{ratio} vs R_{JB} (Figure.IV.15). De 10-30 km le $F_{N:SS}[ANN(AllR_{JB},FM)] > F_{N:SS}[ANN(80R_{JB},FM)]$, les deux sont supérieur à 1 (nous n'analysons pas les distances inférieures à 10 km vu que les données ne sont pas représentatives dans cette gamme de distance). Au-delà de 30 km nous constatons que $F_{N:SS}[ANN(AllR_{JB},FM)] < F_{N:SS}[ANN(80R_{JB},FM)]$, les deux rapports trouvent leur maximum à $R_{JB}=70$ km $F_{N:SS}[ANN(AllR_{JB},FM)]=2$ et $F_{N:SS}[ANN(80R_{JB},FM)]=2.9$ et $F_{R:SS}[ANN(AllR_{JB},FM)]=F_{R:SS}[ANN(80R_{JB},FM)]$. Il est à noter que, $F_{R:SS}[ANN(AllR_{JB},FM)]$ est presque constant sur toute la gamme de distances pour une magnitude égale à 5.5. Par ailleurs, nous pouvons remarquer que l'effet de FM augmente avec la distance.

V.9.2. Dépendance des FM avec la magnitude

La variation de FM avec M_w est présentée sur la Figure.V.16. Toujours pour les deux gammes de distances All R_{JB} et 80 R_{JB} et pour une distance égale à 30 km. Les formes des 2 modèles se ressemblent. Ce qui concerne les valeurs extrêmes, $F_{N:SS}[ANN(AllR_{JB},FM)]=1.9$ à $M_w=5.5$ et $F_{N:SS}[ANN(80R_{JB},FM)]=1.95$ pour $M_w=6.5$. Les deux $F_{N:SS}$ sont comprises entre [1 et 2]. En outre, les deux $F_{R:SS}[ANN(AllR_{JB},FM)]$ et $F_{R:SS}[ANN(80R_{JB},FM)]$ donnent des valeurs comprises entre [0.5 et 1.1]. Vu la ressemblance existante entre les facteurs F du même type, nous constatons que, la gamme de distances choisie en amont n'influe pas sur la variation du FM avec la magnitude.

V.9.3. Dépendance des FM avec la profondeur focale

Dans la Figure.IV.17, nous exposons l'impact de la profondeur focale sur les F_{ratio} . D'une façon générale, les deux facteurs diminuent avec le Depth jusqu'à 20 km. Les $F_{N:SS}[ANN(AllR_{JB},FM)]$ et les $F_{N:SS}[ANN(80R_{JB},FM)]$ d'une part et les $F_{R:SS}[ANN(AllR_{JB},FM)]$ et les $F_{R:SS}[ANN(80R_{JB},FM)]$ d'autre part se convergent. Les valeurs maximales appartiennent à Depth=0 km : $F_{N:SS} \cong 3$ et $F_{R:SS} \cong 1.5$. Cet écart diminue progressivement avec la profondeur pour atteindre une valeur nulle à une profondeur de 20 km. Même remarque pour les $F_{R:SS}$ avec un facteur égal à 1.5. Là aussi la gamme de distances choisie en amont n'influe pas sur l'évolution du FM avec la profondeur focale. Cependant, la profondeur focale a un effet considérable sur le l'effet du mécanisme et l'effet de FM diminue avec la profondeur.

Il est à constater que la gamme de la distance pour une base de données choisie pour l'élaboration des GMPEs, n'influe pas pratiquement sur les variations du rapport F avec la magnitude et la profondeur focale. Par contre, la gamme de la distance du jeu de données influe considérablement sur les variations des F_{ratio} avec la période et la distance.

En conclusion, l'analyse de l'effet du style de la faille comme proxy de la source sismique sur le mouvement sismique montre que ces effets donc complexent, car dépendants d'autres facteurs (distance, profondeur, région). Les mécanismes physiques à l'origine des dépendances observées ne sont pas compris. En pratique, nos résultats suggèrent qu'il est peut-être prudent, du fait de ce manque de compréhension, de ne pas inclure ce paramètre pour le développement de modèle de prédiction du mouvement du sol.

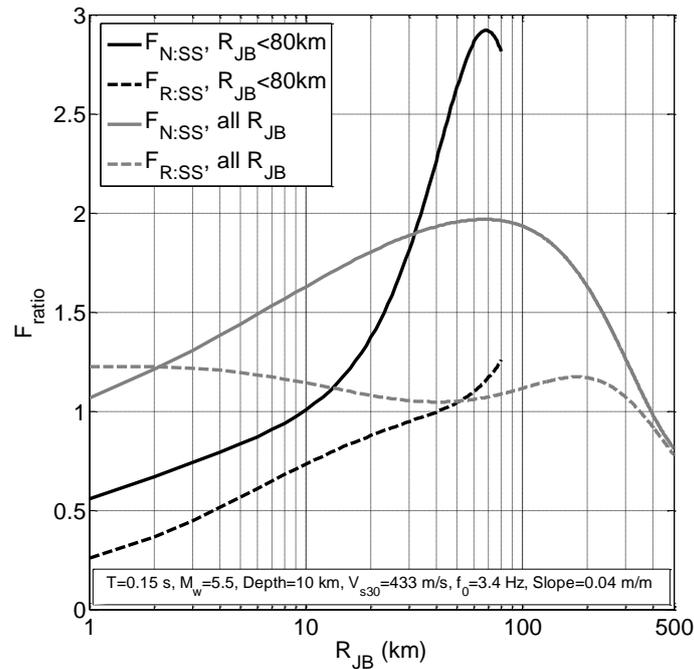


Figure.V.15 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la distance pour les deux modèles des 2 jeux de données KiK-Net: $R_{JB} \leq 500 \text{ km}$ et $R_{JB} \leq 80 \text{ km}$ pour un scénario donné et pour $T= 0.1 \text{ s}$.

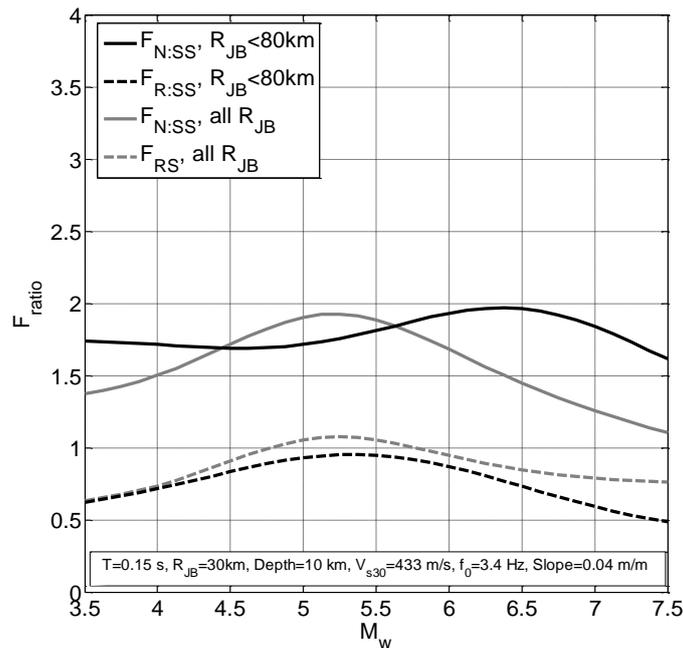


Figure.V.16 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la magnitude pour les deux modèles des deux jeux de données KiK-Net : $R_{JB} \leq 500 \text{ km}$ et $R_{JB} \leq 80 \text{ km}$ pour un scénario sismique donné et pour une période spectrale égale à 0.1s.

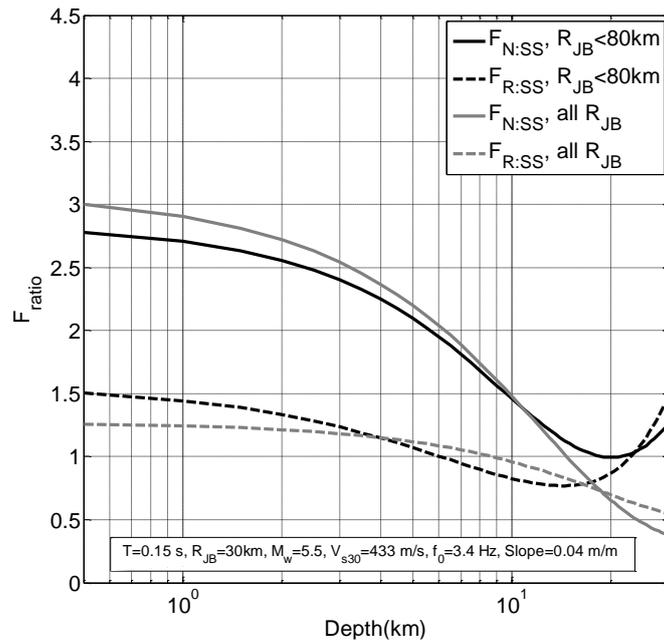


Figure.V.17 Variation des deux facteurs $F_{N:SS}$ et $F_{R:SS}$ avec la profondeur focale pour les deux modèles des deux jeux de données KiK-Net : $R_{JB} \leq 500$ km et $R_{JB} \leq 80$ km pour un scénario sismique donné et pour une période spectrale égale à 0.1s.

Dans ce qui suit, nous testons d'autres proxies de la source sismique, en analysant la dépendance existante entre la chute de contrainte sismique ($\Delta\sigma$) et la magnitude (M_w) et par conséquent voir si la variation de la chute de contrainte sismique est constante ou variable.

3^{ème} partie : Dépendance entre la magnitude et la chute de contrainte : que peut-on tirer de l'effet d'échelle lié à la magnitude

MAGNITUDE DEPENDENCE OF STRESS DROP: WHAT DOES THE OBSERVED MAGNITUDE SCALING OF GROUND-MOTIONS TELL US?

<http://hal.ird.fr/hal-01461210v1>

B. Derras^(1,2,3), F. Cotton^(4,5), S. Drouet⁽⁶⁾, P-Y. Bard⁽⁷⁾

⁽¹⁾ISTerre, University of Grenoble Alpe, Grenoble, France.

⁽²⁾RISAM laboratory, University of Tlemcen, Algeria.

⁽³⁾Department of Civil Engineering and Hydraulic, University of Saïda, Algeria.

⁽⁴⁾Helmholtz Centre Potsdam, German Research Center for Geosciences (GFZ), Germany.

⁽⁵⁾Institute of Earth and Environmental Sciences, University of Potsdam, Germany.

⁽⁶⁾Geoter International – FUGRO Group, Auriol, France.

⁽⁷⁾ISTerre, University of Grenoble Alpe, CNRS, IFSTAR, Grenoble, France.

Abstract

The behavior of earthquake stress-drop magnitude scaling has been the topic of significant debate in the earthquake source community over the past two decades. Methodologies which have been adopted by a large number of source studies require corrections for source radiation pattern, path attenuation and site amplification that ultimately introduce large uncertainties for stress-drop estimates. In this study, we adopt a different strategy: we analyze directly the ground-motions (Y) and their dependencies with magnitude (M). We first use simple stochastic models (e.g. [Boore, 2003](#)) comprised of a ([Brune, 1970](#), [Brune, 1971](#)) source spectrum and various models of magnitude-dependent stress drop. We show that magnitude-dependent stress-drop and constant stress-drop models lead to different scaling of ground-motions ($d\log Y/dM$) with frequency. Using the results of Molkenhain et al. (2014), we then analyze the magnitude dependency of NGA-West 2 ground-motions for source-site configurations where stress-drop is the key controlling factor of ground-motions (moderate distances and rock-sites). In addition, the use of a neural network method allows us to obtain fully record-driven evaluations of ($d\log Y/dM$) with frequency both for simulated and observed records. The comparison between these observed and simulated ($d\log Y/dM$) allows us to discuss the scaling of the stress-drop with magnitude. We show that, there is not the strong differences of the magnitude scaling of ground-motions between mainshocks and aftershocks. Also we observe that, constant stress-drop models and the tested magnitude-dependent models have not been able so far to reproduce the observed scaling.

Keywords: Source scaling; ground motion; stochastic models; neural networks; stress drop.

V.10. Introduction

Dynamic stress drop (“Brune” stress-drop) is a key parameter in the estimation of strong ground motion, as it influences the high-frequency level of acceleration ([Brune, 1970](#), [Brune, 1971](#)). There are many methods for simulating ground motion from earthquakes. Each has particular strengths, and each can provide estimates of ground motion. However, the simulation methods all depend, directly or indirectly, on the selected dynamic stress drop which is then a primary input parameter needed to compute ground-motion simulations (e.g. stochastic simulations).

Since the seminal work of [Aki K \(1967\)](#), stress drops have been assumed to be constant, allowing for scaling between earthquakes of different magnitudes. However, several studies have suggested a magnitude-dependence of the Brune’s stress-drop for $M_w < 4.5$ ([Edwards and Fäh 2013](#), [Rietbrock et al. 2013](#), [Drouet et al. 2015](#)). Such results generate a large epistemic

uncertainty and the inflation of hazard logic tree branches in regions where models based on small earthquakes have to be extrapolated to large earthquakes. There is then an urgent need to better constrain the magnitude scaling of stress-drops. However, classical methods to analysis stress drops are facing some practical and theoretical problems.

a) Brune's stress-drop evaluation is strongly dependent on the corner frequency measurements uncertainty and corner frequencies are difficult to measure (e.g. Cotton et al. 2013).

b) Fourier spectra spectral inversion results may be biased by trade-off between source, path and site effects including kappa the high-frequency decay of ground-motion amplitude (e.g. [Drouet et al. 2010](#)). Methodologies which have been adopted to analyze the behavior of earthquake source scaling require substantial corrections for path attenuation, site amplification, and source radiation pattern effects that ultimately yield large variances for corner frequency estimates and subsequently for stress-drop estimates ([Yoo and Mayeda 2013](#)).

c) Potential regional variations of source parameters and the heterogeneity of inversion methods complicate the analysis of stress-drops.

d) Spectral ratios of two co-located events removes common path and site effects.

In this study, we adopt a different strategy: we analyze directly the ground-motions (Y) and their dependencies with magnitude (M). Such analysis is performed both on data-driven empirical models data and physics-based, stochastic, models. Several authors (e.g. [Luco 1985](#)) have already been comparing the magnitude scaling of ground-motions predicted both by empirical models and more physics based models. Such comparisons were limited because of the range of magnitude covered by classical crustal earthquakes GMPE's (the magnitude range of these models was usually between 4.5 and 7.5), the fact that classical GMPE's functional forms depend on developers a priori choices and also the difficulty to evaluate the relative impact of source, path and site effects on ground-motions. We then develop here a more comprehensive method to analyze the dependency of ground-motions on the stress-drop magnitude-scaling:

1. We first take advantage of the recent development of the NGA-West 2 database ([Ancheta et al. 2014](#)). This database is now allowing the development of ground-motion models for magnitudes as low as 3.5. Only recordings from shallow crustal events are considered, including mainshocks as well as aftershocks.

2. We build synthetic ground-motions databases using stochastic simulations and various models of magnitude-dependent stress drop: three synthetic databases with $\Delta\tau = 2.5$ Mpa, $\Delta\tau = 5$ Mpa and $\Delta\tau = 10$ Mpa (constant stress drop) and two databases based on [Drouet et al. \(2015\)](#) and [Yoo and Mayeda \(2013\)](#) magnitude-dependent stress drop models.

3. We use the Artificial Neural Network (ANN) approach ([Derras et al. 2012](#)) to get fully data-driven empirical predictive models.

4. We then analyze the magnitude dependency of NGA-West 2 ground-motions for source-site configurations where stress-drop is the key controlling factor of ground-motions according to the recent results of [Molkenthin et al. \(2014\)](#).

Finally, the comparison between these observed and simulated ($d\log Y/dM$) allow us to discuss the scaling of the stress-drop with magnitude. We also compare the differences of the magnitude scaling of ground-motions between mainshocks and aftershocks.

V.11. Data

V.11.1. NGA-West 2 dataset

The NGA-West 2 project is a large multidisciplinary, research program on the Next Generation Attenuation (NGA) models for shallow crustal earthquakes in active tectonic regions. The flatfiles established by the project Ancheta et al. (2014) contain site and source information, along with distances parameters and the corresponding ground-motion intensity measures (IMs). Only a subset of this database was used ([Derras et al. 2016](#)).

One of the aims of our study is to compare the magnitude scaling of ground-motion for both mainshock and the aftershock events. In the NGA-West 2 database, we use the definition of Class 1 and Class 2 events described in [Wooddell and Abrahamson \(2014\)](#) to distinguish mainshocks and aftershocks.

V.11.2. Stochastic simulations of ground-motions (SSGM)

The stochastic simulations are based on the method of [Drouet et al. \(2015\)](#). A set of stochastic parameters is considered, including uncertainties on each parameter. The uncertainty is then propagated through the simulations by random sampling of the parameters distributions. The SMSIM program ([Boore, 2003](#)) is used to simulate data for M_w from 3 to 8 at epicentral distances between 1 and 250 km. Virtual fault planes are created in order to compute various distance metrics and for each magnitude-epicentral distances pairs, a number of different

configurations around the fault plane are simulated (various azimuths). The attenuation parameters considered are those [Drouet et al. \(2015\)](#), and standard rock site conditions are considered. A generic velocity profile from Cotton et al. (2006) with $V_{s30}=800$ m/s is used together with $\kappa=0.03$ s. Uncertainties on κ and on the site amplifications computed from the velocity profile are propagated ([Drouet et al. 2015](#)).

V.11.3. Data distribution

The distributions of the datasets according to M_w , R_{hyp} , and PSA are displayed on Figure.V.18. This figure illustrates the range of $M_w=[3$ to $6.5]$ and $R_{hyp}<60$ km for NGA-West 2 and $R_{hyp}<41.79$ km for SSGM. To minimize the influence of site effects we use NGA-West 2 data with $V_{s30}>500$ m/s.

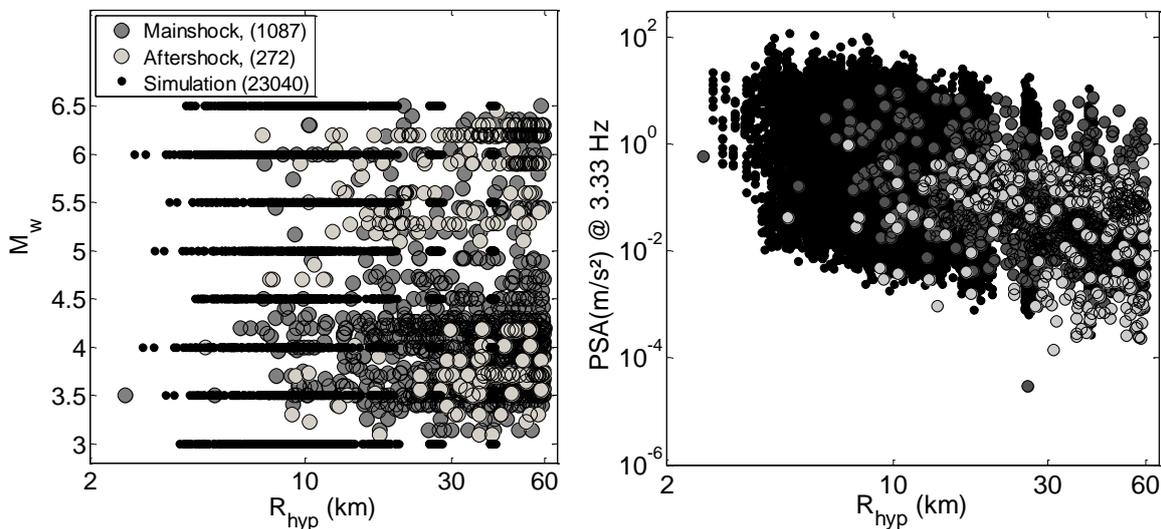


Figure.V.18 Distribution of the databases considered in this study in the [Magnitude (M_w) - distance (R_{hyp})] (left), [PSA @ 3.33 Hz - R_{hyp}] (right).

V.12. Method

V.12.1. Development of data-driven ground-motion models

The input parameters are M_w and R_{hyp} . The predicted ground-motions are the geometric mean of the horizontal components for PGA, PGV, and 5%-damped PSA at 19 frequencies from 0.33 to 100 Hz.

A feed-forward backpropagation neural network type is used. The structure of the ANN is displayed in Figure.V.19. This structure has been taken from [Derras et al. \(2012\)](#). The input and output layers contain the input and output parameters, and are linked through one single hidden layer consisting of only three neurons. This small number of neurons is the optimal hidden

neurons number in order to optimize both the total σ (Cotton et al. 2013) and the Akaike Information Criterion (Akaike 1973). The type of activation functions between input and hidden layers, and between hidden and output layers have been adopted after several tests (Derras et al. 2014). It resulted in the choice of a "tangent sigmoid" type for the hidden layer and a "linear" one for the output layer. The quasi-Newton Back Propagation technique has been applied for the training phase (Shanno et al. 1970). To avoid "over-fitting" problems we chose an adequate regularization method developed by Derras et al. (2012) together with a limited number of neurons on the hidden layer. In Figure.V.19, the symbols W and b represent the synaptic weights and bias with subscripts representing the corresponding neurons between two layers.

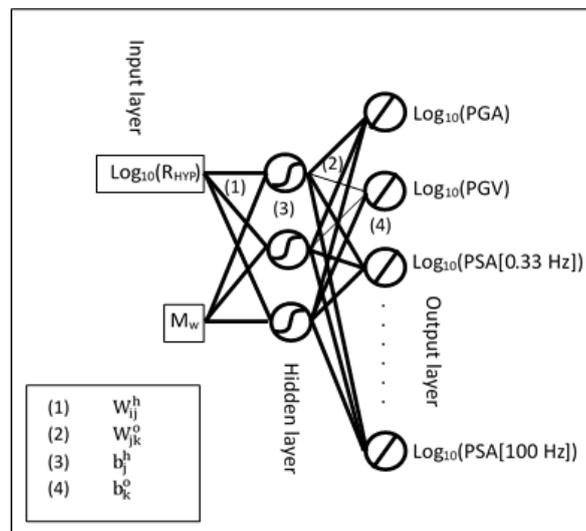


Figure.V.19 Structure of the Neural Networks considered for the prediction of PGA, PGV and PSA[0.33 to 100 Hz]. w_{ij}^h is the synaptic weight between the i^{th} neuron of the input layer and the j^{th} neuron in the hidden layer, b_j^h the bias of the j^{th} neuron in the hidden layer. Also the w_{jk}^o is the synaptic weight between the j^{th} neuron of the hidden layer and the k^{th} neuron in the output layer, b_k^o the bias of the k^{th} neuron in the output layer.

At the end, eight different ANN models were built, differing by the dataset. For the eight datasets, the earthquakes are described by the moment magnitudes M_w and the source to site distances by R_{hyp} and site effect is not modeled (also explained below only sites with $V_{s30} > 500$ m/s are considered here). The first ANN model is based on only mainshocks from the NGA-West 2 database and the second ANN relies on only aftershocks data. For the third ANN model all the NGA-West 2 data are considered (this model is the reference model for the comparison study). Using the synthetic data three ANN models are based on simulations for constant stress drop of 2.5, 5 and 10 Mpa, respectively, and two other ANN models are based on simulations for magnitude-dependent stress drop using the stress drop models of Drouet et al (2015) and Yoo and Mayeda (2013).

V.12.2. Optimal configuration to analyze the magnitude scaling

To analyze the influence of stress drop in ground-motion magnitude scaling, we shall minimize the influence of site effect, anelastic attenuation, kappa effect.

To limit site effect influence, NGA-West 2 data with V_{s30} (a widely used site effect proxy) greater than 500 m/s are selected. At first an even greater V_{s30} was used but this resulted in a too small number of data to derive ANN models. An important point is to choose the optimal frequency range to analyze the influence of stress drop on ground-motion magnitude scaling. [Molkenthin et al. \(2014\)](#) analyzed the sensitivity of response spectral amplitudes on seismological parameters using algorithmic differentiation. They found that at 2 to 7 Hz, there is a large impact of stress drop and a low impact of kappa, Q and M_w . Therefore, for this study, we select a frequency equal to 3.33 Hz for the stress-drop analyze.

[Boore et al \(2014\)](#) and [Kotha et al. \(2016\)](#) show that regional ground-motions variations start to be significant for distances greater than 60-70 km. To limit the impact of regional attenuation variations only data recorded at distances lower than 60 km have been selected. Figure.V.20 shows the magnitude scaling of ground-motions against M_w at a distance of 30 km. The choice of a different distance limit (40 and 30 km) for the data selection does not impact the observed magnitude scaling.

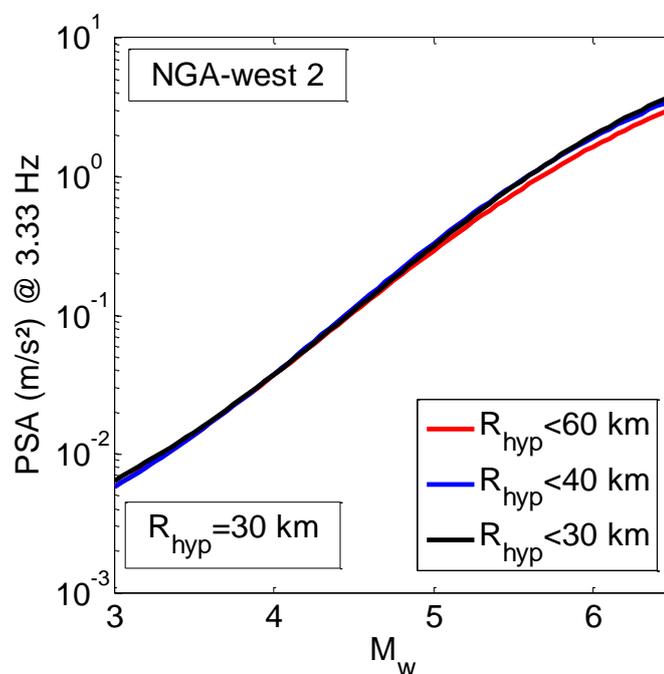


Figure.V.20 Magnitude-dependence of ANN models based on various data selection ($R < 30\text{km}$, $R < 40\text{km}$, $R < 60\text{km}$).

V.12.3. Robustness of ANN approach

Tests of the robustness of the predicted ground-motion values were carried out using different subsets of the whole database to train the ANN. Figure.V.21 compares the pseudo-spectral acceleration curves vs M_w derived when using 100%, and 25% of the whole datasets. For each run, the records of the training set were selected randomly. The results are shown here for a target distance $R_{hyp}=30$ km (which falls in the area with a lot off data). This result shows that the ground-motion model derived using the ANN method and the NGA-West2 database is stable at a distance of 30 km.

V.13. Results

V.13.1. Do aftershocks and mainshocks show a different ground-motion magnitude scaling?

Some previous studies, such as [Abrahamson et al. \(2008\)](#), have found that the median short-period ground motions from aftershocks are smaller than the median ground motions from mainshocks. In our study, the ANN predictive models derived with aftershocks and mainshocks data show the same ground-motion magnitude scaling. Again, some tests have been performed for various subsets of the NGA-West 2 database characterized by various maximum distances 30 km (Figure.V.22).

V.13.2. Comparison between the magnitude scaling of observed (NGA-West 2) and simulated ground-motions

In the last section, we have found that the mainshocks and aftershocks show a similar ground-motion magnitude scaling. For this reason, we now use all NGA-west 2 (aftershocks and mainshocks) ground-motions to compare observed and simulated magnitude scalings.

First, we show the various Brune's stress drop models used in the stochastic simulations (Figure.V.23). Two types of models have been implemented: constant stress-drop models (10, 5 and 2.5 MPa) and magnitude-dependent models. The magnitude-dependent models have been taken from the studies of [Drouet et al. \(2015\)](#) and [Yoo and Mayeda \(2013\)](#). These models both suggest an increase of stress-drop with magnitude. The Drouet and Cotton model is based on the analysis of moderate stress-drops analysis in the French Alps. This model suggest a significant stress-drop increase from magnitude 3 to 4.5. The [Yoo and Mayeda \(2013\)](#) model is suggesting a stress-drop increase from $M=4.5$ to $M=5.5$.

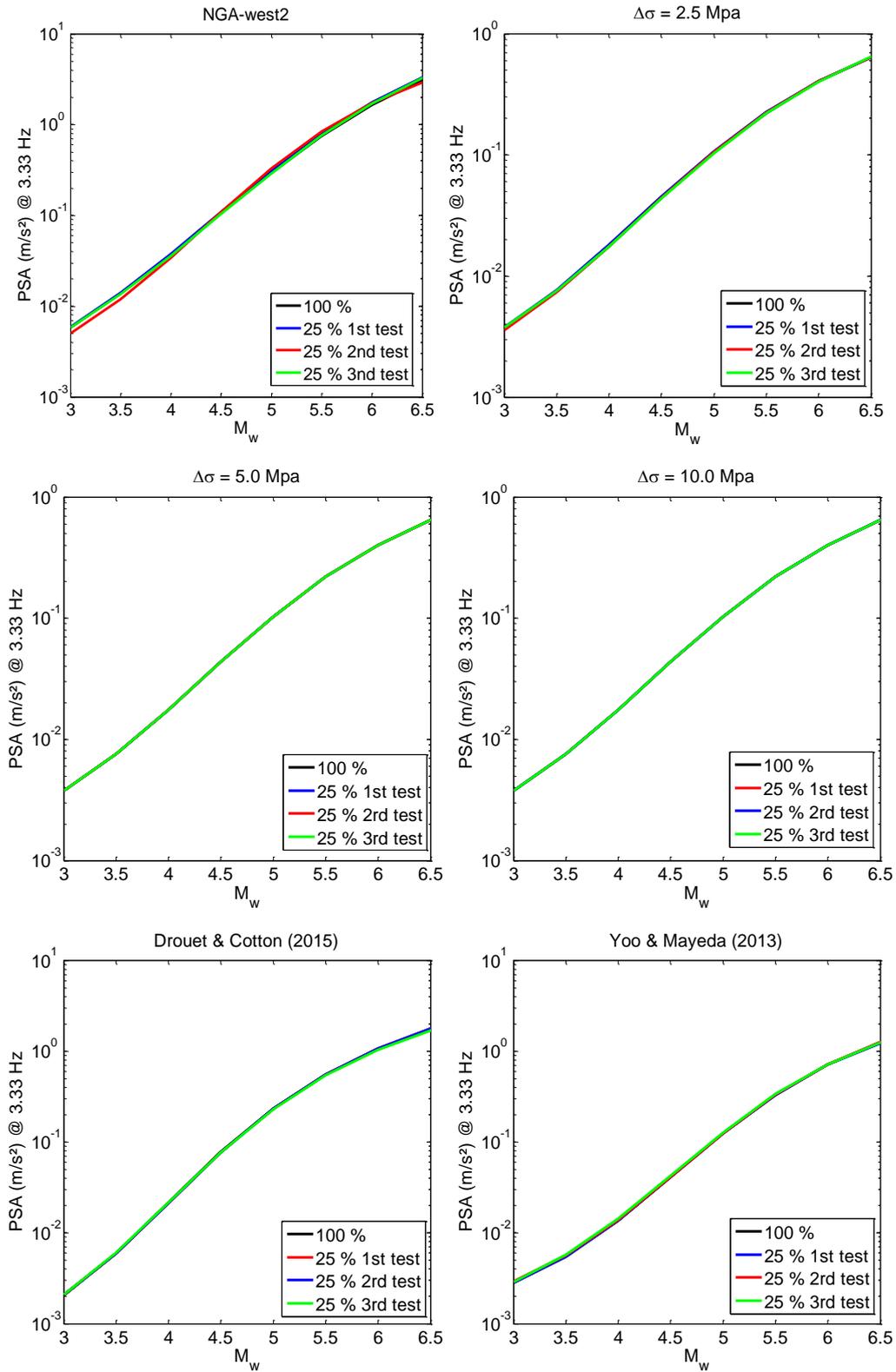


Figure.V.21 Robustness of the ground-motion scaling derived using the ANN approach and the NGA-West 2 dataset. The graphs show the relationship between PSA at 3.33 Hz and magnitude for various subsets of the database.

We then compare the observed NGA-West 2 ground-motion magnitude scaling and the scaling predicted by stochastic simulations performed with the stress-drop models shown on Figure.V.24. Figure.V.25 first illustrates the comparison with stochastic simulations performed with constant stress-drops models. None of the selected models is showing a perfect fit. The tests show that the amplitude of the stress drops has an impact both on the slope and the amplitude of the scaling.

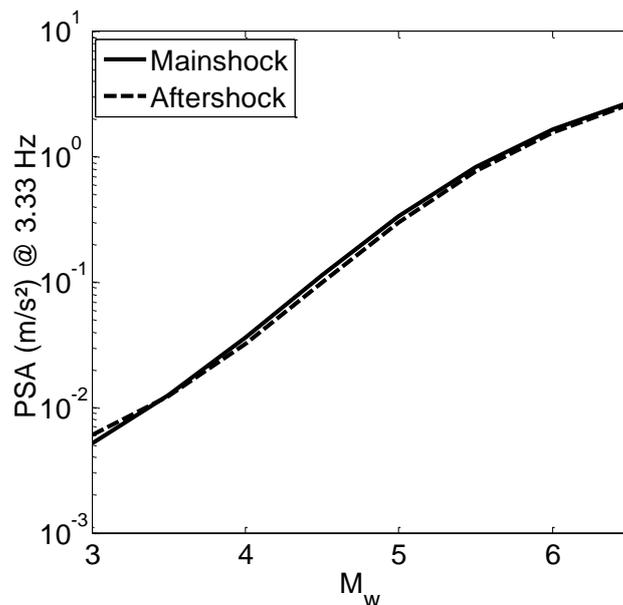


Figure.V.22 ANN predictive models derived with Aftershocks and Mainshocks data for 30 km.

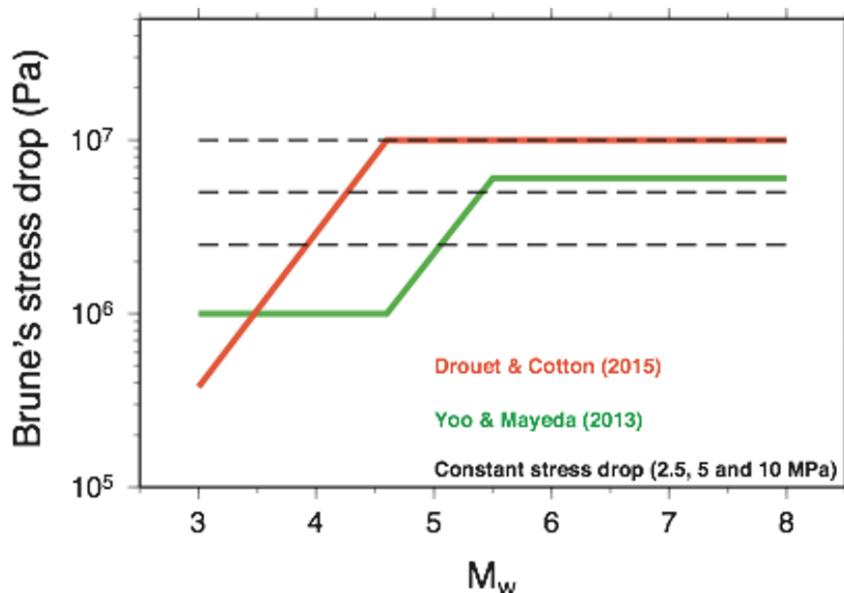


Figure.V.23 Stochastic simulations of ground-motions (SMSIM : [Boore, 2003](#)) with various stress-drop models

The closest model is obtained for a constant stress-drop $\Delta\tau=10$ Mpa. All models underestimate the magnitude scaling ($d\log Y/dM$) for low ($M<4$) and large ($M>5$) earthquakes. The tested magnitude dependent stress-drops do not show either a perfect fit. Figure.V.24 shows that the

magnitude scaling of stress-drop has a strong impact on the slope of the ground-motion magnitude-scaling.

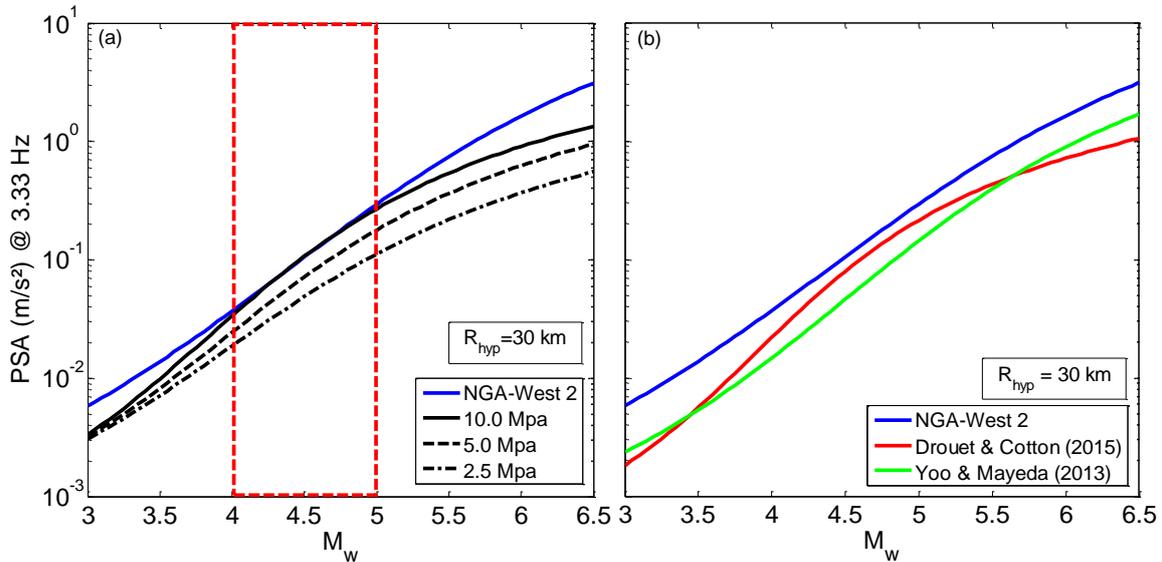


Figure.V.24 Comparison between the magnitude scaling of observed (NGA-West 2) and simulated ground-motions at the distance of 30 km.

To illustrate this, ground-motion amplitudes have been normalized by the observed or predicted amplitude at $M_w=4.5$ and a distance of 30 km. The results (Figure.V.25) shows that the Yoo et al. (2013) model show an increase of the magnitude scaling which is similar to the observed one. However, given the fact that stress-drop amplitudes have an impact on the magnitude-scaling increase (Figure.V.24a) it is not straight forward to derive a model which will fit the relationship (amplitude and slope) between $d\log Y/dM$ and the magnitude.

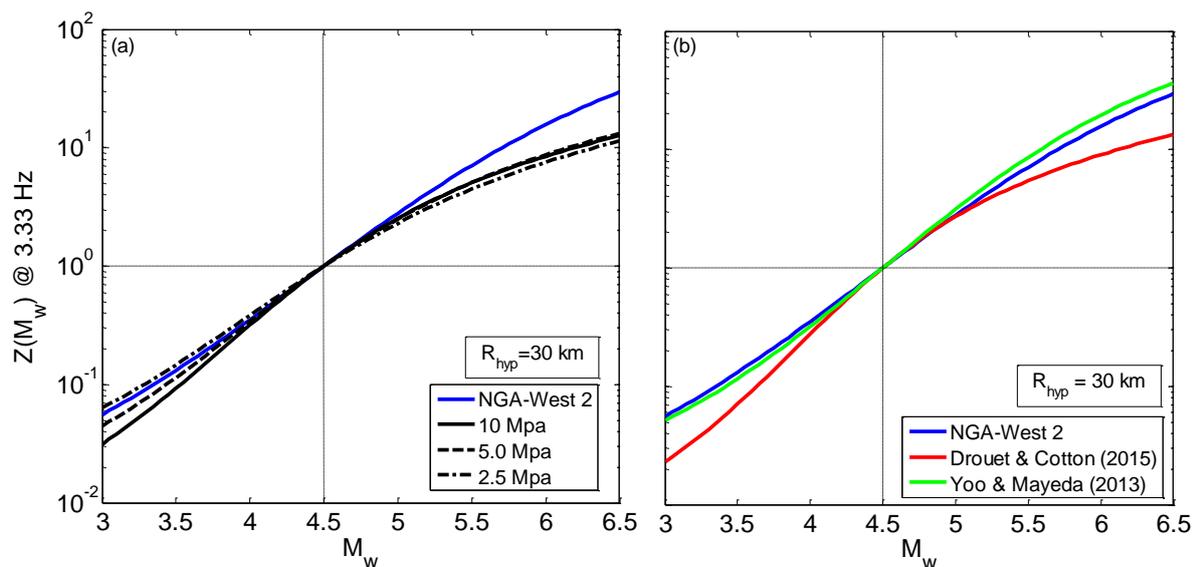


Figure.V.25 Magnitude-dependent decay of the seven models, for pseudo absolute acceleration spectra (PSA) at 3.3 Hz. We represent the ground-motion intensity normalized ($Z(M_w)$) at $M_w=4.5$ versus magnitude at the distance of 30 km.

V.14. Conclusion

We first take advantage of the recent development of the NGA-West 2 database ([Ancheta et al. 2014](#)) to analyze the magnitude scaling ($d\log Y/dM$) of ground-motions for shallow crustal earthquakes. We then analyze the magnitude dependency of NGA-West 2 ground-motions for source-site configurations ($R_{hyp} = 30$ km, $f = 3.3$ Hz) where stress-drop is the key controlling factor of ground-motions. Only sites with $V_{s30} > 500$ m/s and $R_{hyp} < 60$ km have been considered.

The ANN approach has been used in this article to analyze the ground-motions and their dependencies with magnitude (M_w). Eight ANN models were built, differing by the dataset. The first ANN model is based on only mainshocks events and the second ANN relies on only aftershocks data. For the third ANN model all the data are considered. Using the synthetic data three ANN models are based on simulations for constant stress drop of 2.5, 5 and 10 Mpa, respectively, and two other ANN models are based on simulations for magnitude-dependent stress drop using the stress drop models of [Drouet et al \(2015\)](#) and [Yoo and Mayeda \(2013\)](#).

The comparison between these observed and simulated scaling ($d\log Y/dM$) allows to discuss the scaling of the stress-drop with magnitude and compare the differences of the magnitude scaling of ground-motions between mainshocks and aftershocks. The ANN predictive models derived with aftershocks and mainshocks data show the same ground-motion magnitude scaling. These results confirm that NGA-West 2 aftershocks and mainshocks ground-motions show a similar magnitude scaling.

The comparison between these observed and simulated scaling ($d\log Y/dM$) allows also to show that the amplitude of the stress drops has an impact both on the slope and the amplitude of the scaling. The closest model is obtained for a constant stress-drop $\Delta\tau = 10$ Mpa and it fits perfectly with observations between $M_w = [4-5]$. Also we suggest that, the relationship between stress-drops and ground-motion magnitude scaling is far to be linear. Constant stress-drop models and the tested magnitude-dependent models have not been able so far to reproduce the observed scaling.

V.15. Acknowledgments

The authors would like to thank the participants of the NGA-West 2 program for providing high quality data and stimulating ideas.

V.16. Data and resources

The datasets used in this article have been collected and disseminated by The Pacific Earthquake Engineering Research Center.

Conclusion

Conclusion de la 1^{ère} partie

Les modèles de prédiction du mouvement sismique issus de trois bases de données : RESORCE, NGA-West 2 et KiK-Net ont été considérés. Pour minimiser l'impact de l'effet local de site, nous avons utilisé des jeux de données avec des sites durs ($V_{s30} \geq 450$ m/s). Avec une distance R_{JB} qui ne dépasse pas les 80 km, nous pouvons réduire l'effet des variations régionales de l'atténuation (Boore et al. 2014). Les métadonnées adoptées pour le modèle de références (sans FM) sont les paramètres décrivant l'effet de source et de propagation d'onde, à savoir M_w et R_{JB} . En ajoutant à ces deux paramètres, les coefficients décrivant les styles de la faille, nous avons élaboré un 2^{ème} modèle. La comparaison entre ces deux modèles nous a permis d'analyser l'effet du style de la faille sur la variabilité aléatoire et la médiane. La modélisation du mécanisme au foyer est définie par trois coefficients binaires regroupés en trois classes. À savoir FM_{SS} : faille de décrochement, FM_N : faille normale et FM_R pour décrire les failles inverses. Les paramètres du mouvement sismique à prédire sont les PGA, PGV, PSA [0.01 à 2] s pour un amortissement égal à 5 %. Seuls les séismes crustaux ont été considérés.

En choisissant une distance maximale égale à 80 km et $V_{s30} \geq 450$ m/s nous réduisons considérablement la variabilité intra-événement engendrée par le FM. L'analyse a été concentrée sur la variabilité inter-événement. Les résultats obtenus montrent que, l'impact du FM est négligeable sur la variabilité inter-événement du modèle issu de la base de données NGA-West 2 et que l'introduction du type de la faille dans les GMPEs Californiennes n'a pas un grand effet sur le mouvement sismique. L'influence du mécanisme au foyer sur le modèle issu de la base RESORCE reste faible. Par contre, l'effet du mécanisme au foyer est important sur la variabilité inter-événement obtenue à partir du modèle KiK-Net (Japon).

Les rapports entre les PSAs donnés par la faille normale et la faille de décrochement $F_N:SS$ d'une part et par la faille inverse et la faille de décrochement $F_R:SS$ d'autre part donnent les résultats provisoires suivants : $F_{SS} > F_N$ pour les deux modèles RESORCE et NGA-West 2, $F_N > F_{SS}$ pour le modèle avec KiK-Net. Par ailleurs $F_R \cong F_{SS}$ pour l'ensemble des modèles. Les

tendances obtenues par les différents modèles des trois bases de données ne sont pas cohérentes ce qui indique que l'effet FM sur le mouvement sismique reste mal-compris.

Conclusion de la 2^{ème} partie

Seule, la base de données KiK-Net a été utilisée dans cette partie. Nous avons adopté 4 modèles avec et sans FM pour une $R_{JB}=[1-500]$ km et avec et sans FM pour une $R_{JB}=[1-80]$ km. L'objectif était de mesurer l'impact du choix de la gamme de distance (avec et sans effet régional) sur l'effet du mécanisme au foyer :

- 1) Le modèle de prédiction construit avec une $R_{JB}<80$ km donne des incertitudes plus faibles que le modèle qui se base sur des $R_{JB}<500$ km. Le premier a un $\sigma_{(PGA)}=0.30$ et le deuxième exhibent in $\sigma_{(PGA)}=0.32$. Ces deux modèles intègrent l'effet FM.
- 2) La réduction du τ du modèle avec $R_{JB}<80$ km est plus grande qu'avec le modèle où nous utilisons une $R_{JB}<500$ km. Il est à noter aussi que le fait d'introduire la profondeur focale dans les GMPEs réduit significativement le σ inter-événements (τ), $R_c=25\%$.
- 3) Dans l'ensemble des modèles nous remarquons la présence d'une réduction de ϕ entre les modèles avec et sans FM.
- 4) Nous constatons une grande variation du rapport $F_{N:SS}$ avec la période, la distance et la profondeur focale.
- 5) En outre, les valeurs de $F_{N:SS}$ du modèle produit avec une gamme de distance inférieure à 500km sont plus grandes, d'une façon générale, que celui construit avec $R_{JB}<80$ km.
- 6) Il est à constater que le choix de la gamme de la distance, n'influe pas pratiquement sur les variations du rapport F avec la magnitude et la profondeur focale. Par contre, ce choix influe grandement sur les variations des F_{ratio} avec la période et la distance.

Nous constatons finalement que, la gamme de distances choisie en amont pour l'élaboration des GMPEs influe considérablement sur l'effet du FM-MS soit dans la partie explicite de l'équation de prédiction du mouvement sismique ou dans la partie implicite. Ces résultats confirment la difficulté à prédire l'effet du mécanisme sur le mouvement sismique.

Conclusion de la 3^{ème} partie

L'élaboration de plusieurs GMPEs a permis d'analyser l'effet de la variation de la chute de contrainte avec la magnitude sur le mouvement sismique afin d'évaluer si la chute de contrainte sismique est constante ou variable en fonction de la magnitude. Cette analyse nous a permis de conclure que :

- Les spectres de réponse des chocs principaux et des répliques sont similaires.
- Le modèle le plus proche aux observations est obtenu pour une chute de contrainte constante $\Delta\sigma = 10$ Mpa pour $M_w = [4-5]$.
- Les comparaisons suggèrent, cependant, que la relation entre la chute de contrainte et la magnitude n'est pas constante. Les modèles de la chute de contrainte constante et variable testés n'ont pas été capables de reproduire les tendances observées.

Conclusions générales et perspectives



Les travaux qui ont été réalisés durant ces années de thèse nous ont permis d'améliorer nos connaissances et compétences vis-à-vis des équations de prédiction du mouvement sismique (GMPEs) et de tester la pertinence de plusieurs proxies de site et de source en vue de réduire la variabilité aléatoire des estimations du mouvement sismique.

Avec l'utilisation de trois jeux de données : RESORCE, NGA-West 2 et KiK-Net, nous avons pu élaborer des GMPEs en utilisant l'approche neuronale. Les modèles de prédiction du mouvement sismique élaborés peuvent être utilisés directement :

- pour l'évaluation déterministe et probabiliste de l'aléa sismique,
- comme un paramètre d'entrée de la méthode spectrale modale largement utilisée dans le domaine de l'analyse dynamique des structures,
- afin d'améliorer les règles parasismiques telles que l'Eurocode 8

Par le biais de ces GMPEs, nous avons essayé de résoudre certains problèmes rencontrés dans des études antérieures (ex. [Derras, 2011](#)). Le couplage entre l'approche "par effets aléatoires" et l'approche neuronale ("effets fixes") a permis d'avoir des éléments de comparaison entre nos GMPEs totalement dirigées par les données, et les GMPE classiques, en décomposant la variabilité aléatoire en deux composantes : inter-événements et intra-événement. Par ailleurs, le reproche souvent fait aux modèles neuronaux de conduire à des équations de prédictions trop complexes est en partie résolu par la proposition d'équations analytiques pour prédire le mouvement sismique. Il est à noter, cependant, que ces types de modèles conduits par les données ne peuvent en aucun cas être utilisés hors domaine de validité (extrapolation interdite !).

Rappel des résultats

Les conclusions globales obtenues par la présente thèse sont résumées comme suit :

- a) Les incertitudes données par les différents modèles neuronaux sont de même ordre de grandeur que celles des GMPEs classiques,

- b) Les variations du mouvement sismique (PGA, PGV et PSA) avec les différents proxies (de source, de propagation d'onde et de site) obtenues par les modèles neuronaux sont comparables à celles des modèles classiques les plus récents et sophistiqués.
- c) Les tendances données par les différents modèles neuronaux en termes de sensibilité aux différents paramètres d'entrée, convergent vers celles des GMPEs classiques : même sans aucune hypothèse ou connaissance a priori sur la forme fonctionnelle, l'approche neuronale capte l'effet d'échelle lié à la magnitude, l'atténuation du mouvement sismique avec la distance (et notamment sa sensibilité à la magnitude), et les effets de site linéaires et non linéaires,

Les conclusions détaillées sont regroupées dans les points suivants :

Chapitre II

- 1) Les résultats obtenus à l'issue de l'investigation effectuée sur la base de données RESORCE révèlent, une différence significative des tendances (avec V_{s30}) des résidus entre les données italiennes et turques. Cette différence pourrait être une indication indirecte du caractère régional du proxy de site (V_{s30}). En revanche, le type de mécanisme au foyer et la profondeur focale jouent un rôle marginal sur la prédiction du mouvement sismique, contrairement à la magnitude, la distance et V_{s30} .

Chapitre III

- 2) Deux jeux de données (RESORCE et NGA-West 2) ont été utilisés afin de tester la pertinence de différents proxies de site, via leur capacité à réduire la variabilité aléatoire du mouvement sismique. Plusieurs GMPEs "neuronales" élaborées avec un seul proxy de site, sont comparées d'une part avec une GMPE "de référence" négligeant le terme de site, et d'autre part entre-elles en considérant séparément plusieurs proxys différents : V_{s30} mesurée, V_{s30} "estimée" par corrélations avec d'autres mesures ou informations, ou pente topographique. Les performances de chaque proxy de site peuvent alors être quantifiées par la réduction de la variabilité aléatoire par rapport au modèle de référence. Le proxy de site le moins performant s'avère être la pente topographique du site, qui permet cependant utile pour permettre une légère réduction à moyenne à basse fréquence dans le cas où aucune autre information sur le site n'est disponible. La manière dont V_{s30} est obtenue joue un rôle

- crucial. La mesure directe de V_{s30} conduit à une réduction de la variabilité intra-événements sensiblement plus importante que de simples "estimations" à partir d'autres informations de site. Par ailleurs, les proxys continus sont plus intéressants que les classes de site discrètes telles que proposées dans les règles NEHRP ou EC8.
- 3) La nature de la base de données a une influence considérable sur la variabilité aléatoire du mouvement sismique. Cette dernière augmente avec la période pour RESORCE, mais les grandes valeurs constatées pour les longues périodes sont, peut-être, liées au jeu de données utilisé, qui contient un nombre non négligeable d'événements de faible magnitude, pour lesquels les précisions de localisation et d'estimation de la magnitude sont moindres (ex. [Maufroy et al. 2016](#)). Dans ce contexte, la variation des poids synaptiques en (%) avec la période illustrée sur la Figure C1 pour les trois jeux de données (KiK-Net, RESORCE et NGA-West 2) montre clairement que la magnitude affecte surtout les longues périodes (basses fréquences – comme on peut s'y attendre pour la magnitude de moment...), alors qu'inversement la distance influe beaucoup plus sur les courtes périodes (hautes fréquences). Cependant, ces deux quantités ont toujours un effet plus significatif que les autres paramètres tels que V_{s30} , la profondeur focale et le mécanisme au foyer.
 - 4) La variabilité aléatoire présente toujours un maximum vers $T=0.1$ s pour les jeux de données NGA-West 2. Cela est probablement dû à la conjonction de plusieurs effets à courte période / haute fréquence : κ , chute de contrainte ($\Delta\sigma$), variabilité de la structure superficielle mal représenté par V_{s30} , et peut-être un effet "résiduel" de M_w .
 - 5) Dans le même chapitre et via le jeu de données NGA-West 2, nous avons pu confirmer que l'écart type inter-événements manifeste une certaine décroissance quand la magnitude augmente – qui peut cependant être simplement due à une moins grande précision des paramètres de source, de plus, nous remarquons que la variabilité intra-événements augmente avec la distance pour des R_{JB} supérieurs à 100 km. L'effet de l'atténuation régionale est peut-être à l'origine de cette augmentation.
 - 6) L'analyse de l'ensemble des GMPEs montre que la nature du proxy de site n'influe pas sur la médiane des estimations. Ce résultat apparaît très robuste, car il est obtenu avec les trois jeux de données RESORCE, NGA-West 2 et KiK-Net (Chapitre IV), et pour tous les proxys disponibles : V_{s30} mesurée, V_{s30} estimée, V_{s30} "indifférenciée" (mesures + estimations), P_{ente} , H_{800} et f_0 (Chapitre IV). Il reste cependant que les estimations médianes dépendent bien évidemment du jeu de données initial utilisé

pour l'élaboration d'une GMPE : il peut y avoir des différences entre le Japon, l'ouest des États-Unis ou l'Europe !

Chapitre IV

- 7) Le premier résultat obtenu dans le chapitre IV sur un jeu de données KiK-Net est que, pour de très courtes périodes (< 0.1 s) aucun des proxies de site n'arrive à réduire significativement la variabilité aléatoire. Ensuite, pour des périodes variant entre 0.1 et 0.6 s, c'est V_{s30} qui s'avère le plus performant comparé à la pente topographique, H_{800} et f_0 . Au-delà de cette période, c'est f_0 ou H_{800} qui permettent les plus grandes réductions de la variabilité aléatoire. En outre, c'est à nouveau la pente topographique qui s'avère être le proxy le moins performant parmi les quatre proxies de site. Par ailleurs, lorsque des couples de proxies de site sont considérés, leur performance est trouvée d'une part sensiblement supérieure par rapport aux cas d'un seul proxy, et d'autre part peu variables d'un couple à l'autre. Dans le détail, c'est le couple [V_{s30} - H_{800}] qui présente une performance (légèrement) meilleure à courte période et le couple [f_0 - H_{800}] à longue période ; le couple [f_0 -pente], sans jamais être quantitativement le meilleur, apparaît comme présentant un très bon rapport coût / performance sur toute la gamme [0.1 – 4 s].
- 8) Une autre série de résultats concerne la sensibilité du facteur d'amplification spectrale aux différents proxies, mesurées entre leurs valeurs pour les fractiles 10% et 90%. Ce sont V_{s30} et la pente qui fournissent la plus grande différenciation à courte période pour $T = [0 \text{ et } 0.2]$ s, alors qu'à longue période ce sont H_{800} et f_0 . Conformément aux attentes, les amplifications sont systématiquement plus importantes à période intermédiaire (1 s) qu'à courte période (0, 0.2 s). En outre, nous avons pu constater que, l'amplification de site présente une non-linéarité significative, dont les caractéristiques sont, cependant, étroitement liées au proxy utilisé et au paramètre sélectionné pour décrire le niveau du chargement. Le paramètre de chargement le plus pertinent (suivant cette analyse) est l'accélération spectrale au rocher (ou «site ferme») à la période considérée. Un autre résultat important est la variabilité de la non-linéarité selon le proxy de site utilisé : la non-linéarité à courte période s'avère plus importante quand elle est décrite avec les proxys V_{s30} et pente, et disparaît avec H_{800} . Donc pour tenir compte de la non-linéarité à courte période nous devrions (de préférence) utiliser des proxies qui sont robustes à cet intervalle de période.

- 9) Dans la deuxième partie du chapitre IV, nous avons utilisé les RSR_{NL-L} (Non-Linear to Linear site Response Spectral Ratio, mesurés entre les capteurs de surface et de profondeur du réseau KiK-net) pour tester la pertinence des différents proxies de site à prédire les non-linéarités de réponse de site. Avec l'approche neuronale, nous avons ébauché une analyse de sensibilité pour savoir quels paramètres sont les mieux placés pour prédire ces indicateurs de non-linéarité RSR_{NL-L} . Les principaux résultats sont les suivants.
- Le RSR_{NL-L} calculé à partir du spectre de Fourier donne des écarts types plus faibles que celui évalué à partir du spectre de réponse.
 - La tendance de l'écart type de $RSR_{NL-L,F}$ semble comparable à celle observée sur les GMPE : σ est maximal à courte période ($T=0.15$ s).
 - La pertinence des proxies de site est liée au niveau du PGA en surface : l'écart-type des prédictions diminue quand le seuil de PGA augmente, c.-à-d. le niveau de non-linéarité augmente.
 - L'utilisation de la fréquence caractéristique f_{NL} réduit significativement les résidus pour 6 bins fréquentiels sur 9. Une bonne prédiction des effets non linéaire demande d'utiliser de nombreux proxies de site simultanément, à savoir V_{s30} , V_{smin} , f_{0HV} , B_{30} et surtout f_{NL} .
 - La réduction de σ obtenue par la base de données événements (cette base de données contient les RSR_{NL-L} par site et par événement) est plus faible que celle de la base de données sites (seulement une moyenne de RSR_{NL-L} est donnée pour chaque site). Car dans la première nous avons ajouté en plus de l'incertitude site-site, l'incertitude enregistrement-enregistrement. Pour contourner ce problème (faible réduction de σ) nous avons songé à normaliser les fréquences par f_{NL} . Cela semble impossible avec la présente base de données, car le nombre de bins est très réduit 9. Pour cette raison nous avons essayé de réduire la variabilité site-site par l'introduction (en input) du rapport f_m/f_{NL} (f_m : fréquence d'oscillation moyenne à un bin donné). Cela a donné une réduction pour tous les bins allant jusqu'à 70 % pour l'intervalle de fréquence compris entre 1.2 Hz et 1.9 Hz.

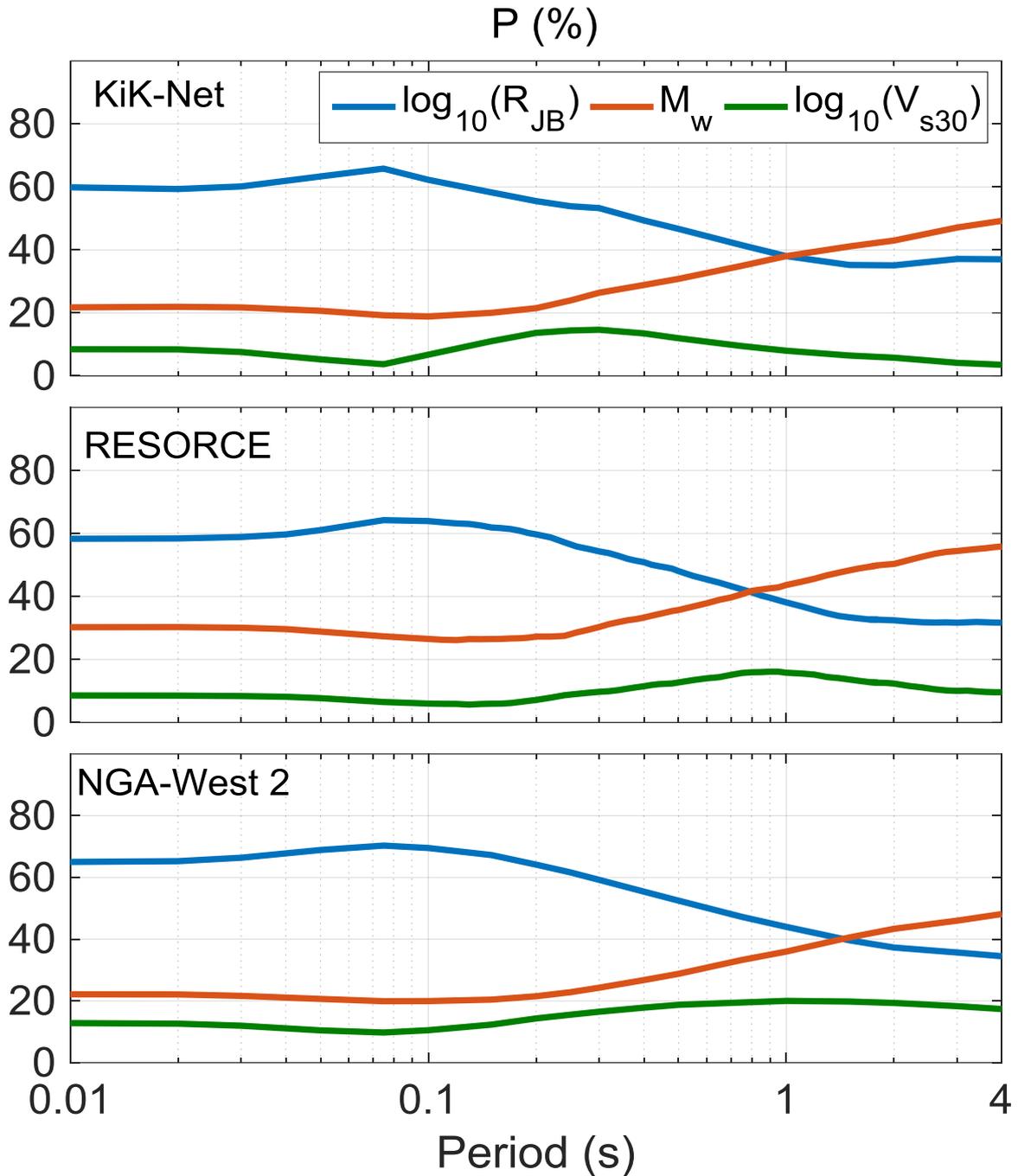


Figure.C.1 Variation des poids synaptiques en (%) avec la période pour les trois jeux de données (KiK-Net, RESORCE et NGA-West 2)

10) Il est clair que, le problème majeur dans la prédiction des RSR_{NL-L} réside dans la disponibilité de f_{NL} : comment estimer f_{NL} sans passer par le RSR_{NL-L} . Dans ce sens, nous avons construit un autre modèle neuronal pour prédire cette fréquence caractéristique f_{NL} en fonction des proxies de site voire aussi de chargement. La première constatation confirme le fait que le niveau de la sollicitation sismique n'a

aucune influence sur la valeur de f_{NL} prédite, et que l'utilisation du f_{NL} ainsi prédit donne des résultats satisfaisants ($R_c=[19-34]\%$).

Chapitre V

Dans la première partie du dernier chapitre, nous avons étudié l'effet du style de la faille (mécanisme au foyer : FM) sur la variabilité aléatoire du mouvement sismique (spécialement sur la variabilité inter-événements) et sur les dépendances existantes entre les FM d'une part et la période, la distance, la magnitude et la profondeur focale d'autre part.

- 11) Les résultats obtenus montrent que, l'impact du FM est négligeable sur la variabilité du modèle issu de la base de données NGA-West 2, ainsi l'introduction du type de la faille dans les GMPEs Californiennes est marginale. Cette influence du FM est moyenne sur le modèle RESORCE, il est souhaitable, donc, d'introduire le FM dans les GMPEs utilisant RESORCE. Par contre, le mécanisme au foyer est important sur la variabilité inter-événements (τ) obtenue à partir du modèle KiK-Net (Japon). Donc il est fortement recommandé d'introduire cet effet dans les modèles de prédictions japonais.
- 12) Le taux de réduction de τ pour les modèles issus de KiK-Net et RESORCE est de l'ordre de 15 % et 13 % respectivement.
- 13) Les rapports entre les PSAs donnés par la faille normale et la faille de décrochement $F_{N:SS}$ d'une part et par la faille inverse et la faille de décrochement $F_{R:SS}$ d'autre part donnent les résultats préliminaires suivants : $FM_{SS}>FM_N$ pour les deux modèles RESORCE et NGA-West 2, $FM_N>FM_{SS}$ pour le modèle avec KiK-Net. Par ailleurs $FM_R\cong FM_{SS}$ pour l'ensemble des modèles. Les tendances obtenues par les différents modèles des trois bases de données sont différentes, ce qui montre que l'effet du mécanisme de rupture sur le mouvement sismique reste mal compris.

L'impact significatif du FM sur le mouvement sismique prédit à partir des données japonaises nous a motivé (dans la deuxième partie) à établir une nouvelle ANN-GMPE en montrant l'influence de l'effet régional et l'effet de la profondeur focale sur la variabilité aléatoire du mouvement sismique. Cela a réduit les composantes de la variabilité aléatoire. De ce fait nous avons trouvé un $\sigma_{(PGA)}=0.30$ plus faible que celui obtenu en [Derras et al. 2012](#). Les principaux résultats liés à l'influence du FM sur la variabilité aléatoire et la dépendance existante entre le rapport F et la période, la magnitude, la distance et la profondeur focale sont donnés comme suit :

- 14) La réduction de τ avec $R_{JB} < 80$ km est plus grande qu'avec le modèle où nous utilisons une $R_{JB} < 500$ km. Il est à noter aussi que l'introduction de la profondeur focale dans les GMPEs a réduit significativement le σ inter-événements (τ), $R_c = 25\%$.
- 15) Nous constatons une grande variation du rapport $F_{N:SS}$ avec la période, la distance et la profondeur focale.
- 16) En outre, les valeurs de $F_{N:SS}$ du modèle produit avec une gamme de distance inférieure à 500 km sont plus grandes, d'une façon générale, que celui construit avec $R_{JB} < 80$ km.
- 17) Nous constatons aussi que, la gamme de distances choisie en amont pour l'élaboration des GMPEs influe considérablement sur la manière avec laquelle le mécanisme au foyer influe sur le mouvement sismique : soit pour la partie explicite de l'équation de prédiction du mouvement sismique ou pour la partie implicite.
- 18) Nous recommandons à cet égard, d'utiliser une distance maximale égale à 80 km sur un sol ferme à rocheux pour réduire l'effet de la variation régionale de l'atténuation anélastique et des effets de site est avoir ainsi une meilleure évaluation de l'effet du FM.

La troisième partie du cinquième chapitre a été consacrée à la dépendance (voir la non-dépendance) existante entre la chute de contrainte ($\Delta\sigma$) et la magnitude (M_w). Pour ce faire plusieurs bases de données issues de simulation numérique NGA-West 2 ont été utilisées. L'analyse a été effectuée toujours en analysant le mouvement sismique en surface pour étudier les variations de ce dernier avec la magnitude et par conséquent voir si la chute de contrainte sismique est constante ou variable. À $f = 3.3$ Hz, $R_{hyp} = 30$ km le $\Delta\sigma$ peut être le paramètre qui contrôle le plus le mouvement sismique. Avec des $V_{s30} > 500$ m/s et pour des distances ne dépassant pas les 60 km, huit GMPE ont été développés :

- 1^{ère} GMPE avec les chocs principaux du jeu de données NGA-West 2,
- 2^{ème} GMPE avec les répliques du jeu de données NGA-West 2,
- 3^{ème} GMPE chocs principaux et répliques confondues (NGA-West 2),
- 4^{ème} GMPE issue des simulations stochastiques avec $\Delta\sigma = 2.5$ Mpa,
- 5^{ème} GMPE élaborée à partir des simulations stochastiques avec $\Delta\sigma = 5$ Mpa,
- 6^{ème} GMPE construite par le biais des simulations stochastiques avec $\Delta\sigma = 10$ Mpa,
- 7^{ème} GMPE, en se basant sur les simulations et les travaux de Drouet et Cotton (2015),

8^{ème} GMPE, toujours avec les simulations cette fois-ci en utilisant le modèle de Yoo et al. (2013).

L'analyse de ces 8 GMPEs a permis d'avoir les conclusions suivantes :

- 19) Les PSA des chocs principaux et des répliques sont similaires et l'effet d'échelle entre PSA- M_w est le même.
- 20) La comparaison entre les PSA observés (NGA-West 2) et les simulées stochastiques permet de montrer que l'amplitude des chutes de contraintes a un impact à la fois sur la pente et l'amplitude l'effet d'échelle.
- 21) Le modèle le plus proche aux observations est obtenu pour une chute de contrainte constante ($\Delta\sigma = 10$ Mpa)
- 22) En outre, nous suggérons que la relation entre les chutes de contrainte et l'effet d'échelle de la magnitude du mouvement sismique est loin d'être complètement comprise. Les modèles de chute de contrainte constante et les modèles dépendant de la magnitude testés ici n'ont pas été capables de reproduire l'effet d'échelle observé.

PERSPECTIVES

Plusieurs pistes d'amélioration sont possibles. Ces améliorations peuvent être menées soit en se focalisant sur la problématique à traiter, soit avec l'utilisation d'autres bases de données ou bien sur des développements méthodologiques.

Thèmes de recherches

Les sujets susceptibles d'intéresser la communauté scientifique sont multiples, nous donnons ici une liste non exhaustive de propositions :

- Tester de nouveaux proxys de source et de site pour essayer de réduire encore davantage la variabilité aléatoire du mouvement sismique.
- Les résultats obtenus par une analyse probabiliste de l'aléa sismique dépendent en grande partie de la GMPE utilisée. C'est pour cette raison que nous proposons d'évaluer l'effet sur les calculs de l'aléa en utilisant l'ensemble des GMPEs obtenues dans cette thèse et de comparer les périodes de retour et les niveaux d'accélération obtenus par ces GMPEs et les GMPEs classiques.
- Puisque dans l'analyse dynamique des structures, le spectre de réponse moyen conditionnel (CMS) est souvent utilisé, nous proposons d'élaborer des CMS avec les

GMPEs issues des méthodes totalement conduites par les données et de comparer les CMS ainsi obtenus avec ceux calculés par les GMPEs classiques.

- On propose aussi de générer des spectres de réponses inélastiques nécessaires à la prise en compte du comportement non linéaire des structures pendant les séismes sévères.
- Pour essayer de réduire la variabilité aléatoire du mouvement sismique dans un site donné, nous préconisons l'utilisation d'un modèle à une seule station (single station).
- Nous pourrions envisager aussi d'étudier l'effet régional (ex. entre les Alpes et les Pyrénées) en termes de la chute de contraintes et la dépendante existante avec la magnitude et le Kappa.
- Nous souhaitons continuer à étudier l'effet de site non linéaire et notamment en utilisant les RSR_{L-NL} avec plus de bins afin d'avoir la possibilité de les normaliser avec F_{NL} .

Bases de données

- Nous pourrions aborder les sujets déjà évoqués ci-dessus avec les données du RAP (Réseau Accélérométrique permanent) ainsi qu'avec les données accélérométriques algériennes, le jour où elles seront librement accessibles.
- Dans la présente thèse, nous avons utilisé les trois bases de données « globales » les plus développées et les plus complètes. Cependant, nous avons constaté que (pour les trois bases de données) peu de données sont enregistrées pour des séismes de forte magnitude ($M > 7$) et pour une distance épacentrale inférieure à 10 km. Ce manque de données peut être comblé par l'utilisation (lors du développement des GMPEs) des résultats issus des simulations stochastiques. Dans ce sens, nous superviserons l'élaboration d'une GMPEs totalement conduite par les données en utilisant, en plus des données réelles, celles obtenues par les simulations stochastiques.
- En outre, dans le cas où nous voulons construire une équation de prédiction de l'amplification de site, les données de simulations de la réponse d'une colonne de sol (ex. par le code NOAH ou CyberQuake) peuvent être utilisées.

Outils et méthodes

L'outil principal que nous avons utilisé dans cette thèse pour obtenir les modèles d'atténuations et pour effectuer les différents tests sur les proxies de sites et de sources est le réseau de neurones artificiels. Cependant il existe d'autres méthodes complémentaires, telles

que, l'approche hybride l'Adaptive Neuro-Fuzzy Interface System (ANFIS) et l'algorithme génétique (AG) pour l'optimisation. L'utilisation de l'ANFIS à la place de la méthode de régression empirique est déjà entreprise (nous ne l'exposons pas ici). Il reste à présent à optimiser les modèles neuronaux avec ces deux méthodes.



Références bibliographiques

- Abrahamson, N. A., and Silva, W. J., 1997. Empirical Response Spectral Attenuation Relations for Shallow Crustal Earthquakes, *Seism. Res Lett* **68**, [94-127](#).
- Abrahamson, N. A., and Silva, W. J., 2008. Summary of the Abrahamson & Silva NGA ground motion relations, *Earthquake Spectra* **24**, [67-97](#).
- Abrahamson, N. A., and Youngs, R. R., 1992. A stable algorithm for regression analyses using the random-effects model, *Bull. Seismol. Soc. Am* **82**, [505-510](#).
- Abrahamson, N. A., Silva, W. J., and Kamai, R., 2014. Summary of the ASK14 Ground Motion Relation for Active Crustal Regions, *Earthquake Spectra* **30**, [1025-1055](#).
- Ahmad, I., El Naggar, M., and Khan, A. N., 2008. Neural Networks Based Attenuation of strong Motion Peaks in Europe, *Journal of Earthquake Engineering* **12**, [663-680](#).
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in *Proc. of the 2nd International Symposium on Information Theory*, Budapest, Hungary **1**, 267-281.
- Aki, K., 1967. Scaling law of seismic spectrum. *Journal of Geophysical Research* **72**, [1217-1231](#).
- Akkar, S., and Bommer, J. J., 2010. Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. *Seismol. Res Lett* **81**, [195-206](#).
- Akkar, S., and Çagnan, Z., 2010. A local ground-motion predictive model for Turkey and its comparison with other regional and global ground-motion models. *Bull Seismol Soc Am* **100**, [2978-2995](#).
- Akkar, S., Çagnan, Z., Yenier, E., Erdogan, Ö., Sandıkkaya, M. A., and Gülkan, P., 2010. The recently compiled Turkish strong-motion database: preliminary investigation for seismological parameters. *J. Seismol* **14**, [457-479](#).
- Akkar, S., Sandıkkaya, M. A., and Bommer, J. J., 2014. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull Earthq Eng* **12**, [359-387](#).
- Akkar, S., Sandıkkaya, M. A., Senyurt, M., Azari, A. S., Ay, B. Ö., Traversa, P., Douglas, J., Cotton, F., Luzi, L., Hernandez, B., and Godey, S., 2014. Reference database for seismic ground-motion in Europe (RESORCE). *Bull. Earthq. Eng* **12**, [311-339](#).
- Al Atik, L., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N., 2010. The variability of ground-motions prediction models and its components, *Seismol. Res. Lett* **81**, [794-801](#).
- Alavi, A. H., and Gandomi, A. H., 2011. Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing, *Computers and Structures* **89**, [2176-2194](#).
- Allen, T. I., and Wald, D. J., 2009. On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}), *Bull. Seismol. Soc. Am* **99**, [935-943](#).

- Ambraseys, N. N., 1978. Preliminary analysis of European strong-motion data 1965–1978. *Bull. Europ. Assn. Earthq. Eng* **4**, 17–37.
- Ambraseys, N., Douglas, J., Sarma, S. K., Smit, P. M., 2005. Equations for the Estimation of Strong Ground Motions from Shallow Crustal Earthquakes Using Data from Europe and the Middle East: Horizontal Peak Ground Acceleration and Spectral Acceleration. *Bull. Earthq. Eng* **3**, 1–53.
- Ambraseys, N., Smit, P., Douglas, J., Margaris, B., Sigbjornsson, R., Olafsson, S., Suhadolc, P. and Costa, G., 2004, Internet-Site for European Strong-Motion Data. *Bollettino di Geofisica Teorica ed Applicata* **45**, 113-129.
- Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, S-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L., 2014. NGA-West 2 database. *Earthquake Spectra* **30**, 989–1005.
- Anderson, J. G., 2000. Expected shape of regressions for ground-motion parameters on rock, *Bull Seismol. Soc. Am* **90**, S43–S52.
- Aoi, S., Kunugi, T., and Fujiwara, H., 2004. Strong-motion seismograph network operated by K-net and Kik-net, *Journal of Japan association for earthquake engineering* **4**, Special Issue.
- Assimaki, D., Li, W., Steidl, J. H., and Tsuda, K., 2008. Site amplification and attenuation via downhole array seismogram inversion: a comparative study of the 2003 Miyagi-Oki, aftershock sequence, *Bull Seism Soc Am* **98**, 301–330.
- Auer, P., Burgsteiner, H., and Maass, W., 2008. A learning rule for very simple universal approximators consisting of a single layer of perceptrons. *Neural Networks* **21**, 786–795.
- Bard, P-Y., Cadet, H., Endrun, B., Hobiger, M., Renalier, F., Theodulidis, N., Ohrnberger, M., Fäh, D., Sabetta, F., Teves-Costa, P., Duval, A. M., Cornou, C., Guillier, B., Wathélet, M., Savvaidis, A., Köhler, A., Burjanek, J., Poggi, V., Gassner-Stamm, G., Havenith, H. B., Hailemichael, S., Almeida, J., Rodrigues, I., Veludo, I., Lacave, C., Thomassin, S., and Kristekova, M., 2010. From non-invasive site characterization to site amplification: recent advances in the use of ambient vibration measurements, in *Earthquake Engineering in Europe* (M. Garevski, A. Ansal, eds.), Geotechnical, Geological, and Earthquake Engineering books, Springer Dordrecht Heidelberg London New York, 105-123.
- Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Puglia, R., Augliera, P., 2014. Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-Damped PSA at spectral periods up to 3.0 s using the RESORCE dataset, *Bull Earthq Eng* **12**, 391-430.
- Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., and Paolucci, R., 2011. Ground motion prediction equations derived from the Italian strong motion database, *Bull Earthq Eng* **9**, 1899-1920.
- Bommer, J. J., and Abrahamson, N. A., 2006. Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?, *Bull. Seismol. Soc. Am* **96**, 1967–1977.
- Bommer, J. J., and Douglas, D., 2004. Processing of European strong-motion records at Imperial College London. In: *Proceedings of the COSMOS workshop on record processing guidelines*, Richmond, California.
- Bommer, J. J., Douglas, J., and Strasser, F. O., 2003. Style-of-faulting in ground-motion prediction equations. *Bull Earthq Eng*, **1**, 171–203.

- Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., and Fäh, D., 2010. On the selection of ground-motion prediction equations for seismic hazard analysis, *Seismol. Res. Lett* **81**, [783–793](#).
- Boore, D. M., 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes—A summary of recent work, *Seismol. Res. Lett* **68**, [128–153](#).
- Boore, D. M., 2003. Simulation of ground motion using the stochastic method, *Pure Appl. Geophy* **160**, [635–675](#).
- Boore, D. M., 2015. Deterministic Seismic Ground Motions and the PEER NGA Ground Motion Prediction Equations (GMPEs). [EERI Utah Chapter Short Course on Seismic Ground Motions](#), West Jordan, Utah, USA.
- Boore, D. M., and Atkinson, G. M., 2008. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01s and 10.0s, *Earthquake Spectra* **24**, [99–138](#).
- Boore, D. M., and Joyner, W. B., 1982. The empirical prediction of ground motion, *Bull Seism Soc Am* **72**, [S43-S60](#).
- Boore, D. M., and Thompson, E. M., 2014. Path durations for use in the stochastic-method simulation of ground motions, *Bull. Seismol. Soc. Am* **104**, [2541–2552](#)
- Boore, D. M., Joyner, W. B., and Fumal, T. E., 1993. Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report, part 2. U.S. Geological Survey Open-File Report 94-127.
- Boore, D. M., Skarlatoudis, A. A., Margaris, B. N., Papazachos, C. B., and Ventouzi, C., 2009. Along-arc and back-arc attenuation, site response, and source spectrum for the intermediate-depth 8 January 2006 M 6.7 Kythera, Greece, Earthquake. *Bull. Seismol. Soc. Am* **99**, [2410–2434](#).
- Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M., 2014. NGA-West 2 Equations for Predicting Response Spectral Accelerations for Shallow Crustal Earthquakes, *Earthquake Spectra* **30**, [1057-1085](#).
- Bora, S. S., Scherbaum, F., Kuehn, N., and Stafford, P. J., 2014. Fourier spectral- and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions, *Bull Earthq Eng.* **12**, [467–493](#).
- Bora, S. S., Scherbaum, F., Kuehn, N., Stafford, P., and Edwards, B., 2015. Development of a Response Spectral Ground-Motion Prediction Equation (GMPE) for Seismic-Hazard Analysis from Empirical Fourier Spectral and Duration Models, *Bull. Seismol. Soc. Am* **105**, [2192-2218](#).
- Borcherdt, R. D., 1994. Estimates of site-dependent response spectra for design (methodology and justification). *Earthquake Spectra* **10**, [617–653](#).
- Box, G. E. P., and Cox, D. R., 1964. An Analysis of Transformations, *Journal of the Royal Statistical Society B* **26**, [211-252](#).
- Bradley, B. A., 2012. Empirical Correlations between Peak Ground Velocity and Spectrum-Based Intensity Measures, *Earthquake Spectra* **28**, [37-54](#).
- Brillinger, D. R., and Preisler, H. K., 1984. An exploratory analysis of the Joyner-Boore attenuation data, *Bull. Seism. Soc. Am* **74**, [1441-1450](#).

- Brillinger, D. R., and Preisler, H. K., 1985. Further analysis of the Joyner-Boore attenuation data, *Bull Seismol Soc Am* [75,611–614](#).
- Brune, J. N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, [75, 4997–5009](#).
- Brune, J. N., 1971. (Correction). Tectonic stress and the spectra, of seismic shear waves from earthquakes, *J. Geophys. Res* **76**, 5002.
- Cadet, H., Bard, P-Y., and Rodriguez-Marek, A., 2010. Defining a Standard Rock Site: Propositions Based on the KiK-net Database, *Bull Seismol Soc Am* [100, 172–195](#).
- Cadet, H., Bard, P-Y., Duval, A-M., and Bertrand, E., 2012. Site effect assessment using KiK-net data: Part 2—Site amplification prediction equation based on f_0 and V_{sz} . *Bull Earthq Eng* [10, 451–489](#).
- Campbell, K. W., 1989. Empirical prediction of near-source ground motion for the diablo canyon power plant site. San Luis Obispo County, California, U.S. Geological Survey Open-File Report [89-484](#).
- Campbell, K. W., 1993. Empirical prediction of near-source ground motion from large earthquakes. Proceedings International Workshop on Earthquake Hazards and Large Dams in the Himalaya, January 15–16, New Delhi, India.
- Campbell, K. W., and Y. Bozorgnia., 2014. NGA-West2 Campbell-Bozorgnia Ground Motion Model for the Horizontal Components of PGA, PGV, and 5%-Damped Elastic Pseudo-Acceleration Response Spectra for Periods Ranging from 0.01 to 10 sec, *Earthquake Spectra*, [30, 1087-1115](#).
- Campbell, K.W., 1981. Near-source Attenuation of peak horizontal Acceleration. *Bull Seismol Soc Am*, [71, 2039-2070](#).
- Castellaro, S., Mulargia, F., and Rossi, P. M., 2008. V_{s30} : Proxy for seismic amplification?. *Seism. Res. Lett* [79, 540– 543](#).
- Cauzzi, C., Sleeman, R., Clinton, J., Ballesta, J. D., Galanis, O., and Kästli, P., 2016. Introducing the European Rapid Raw Strong-Motion Database, *Seismol. Res. Lett* [87, 977-986](#)
- Chandra, J., Gueguen, P., and Bonilla, L. F., 2016. PGA-PGV/ V_s considered as a stress–strain proxy for predicting nonlinear soil response, *Soil Dyn and Earthq Eng* [85, 146–160](#).
- Chandra, J., Gueguen, P., Steidl, J. H., and Bonilla, L. F., 2015. In-situ Assessment of the G- γ Curve for Characterizing the Nonlinear Response of Soil: Application to the Garner Valley Downhole Array (GVDA) and the Wildlife Liquefaction Array (WLA). *Bull Seism Soc Am* [105, 993-1010](#).
- Chiou, B-S. J., and Youngs, R. R., 2013. Update of the Chiou and Youngs NGA Ground Motion Model for Average Horizontal Component of Peak Ground Motion and Response Spectra, [PEER Report 2013/07](#), Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.
- Chiou, B-S. J., and Youngs, R. R., 2014. Update of the Chiou and Youngs NGA Ground Motion Model for Average Horizontal Component of Peak Ground Motion and Response Spectra, *Earthquake Spectra* [30, 1117-1153](#).
- Chiou, B-S. J., Youngs, R. R., 2008. An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra* [24, 173–215](#).
- Choi, Y., and Stewart, J. P., 2005. Nonlinear site amplification as function of 30 m shear wave velocity, *Earthquake Spectra* [21, 1–30](#).

- Cotton, F., Archuleta, R., and Causse, M., 2013. What is sigma of the stress drop?, *Seismol. Res. Lett.*, [84, 42–48](#).
- Cotton, F., G. Pousse., F. Bonilla., and F. Scherbaum., 2008. On the Discrepancy of Recent European Ground-Motion Observations and Predictions from Empirical Models: Analysis of KiK-net Accelerometric Data and Point-Sources Stochastic Simulations, *Bull Seism Soc Am* [98, 2244–2261](#).
- Cotton, F., Scherbaum, F., Bommer, J. J., and Bungum, H., 2006. Criteria for selecting and adjusting ground-motion models for specific target regions: Application to Central Europe and rock sites, *J. Seismol* [10, 137–156](#).
- Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., 2014. A flatfile for the KiK-net database processed using an automated protocol. <https://datacenterhub.org/resources/272>.
- Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., 2016. A flatfile for the KiK-net Database Processed Using an Automated Protocol. *Earthquake Spectra* [32, 1281-1302](#).
- Demuth, H., Beale, M., and Hagan, M., 2009. Neural Network Toolbox™ 6. User's Guide. The MathWorks, Inc.
- Derras, B., 2010. Peak ground acceleration prediction using artificial neural networks approach: application to the Kik-Net data, *International Review of Civil Engineering* [1, 243–252](#).
- Derras, B., 2011. Contribution des données accélérométriques de KiK-Net à la prédiction du mouvement sismique par l'approche neuronale avec la prise en compte des effets de site. Thèse de doctorat Es Science. Université Aboubekr Belkaid. [Tlemcen](#), Algérie.
- Derras, B., Bard, P-Y., and Cotton, F., 2014. Towards fully data-driven ground-motion prediction models for Europe, *Bull. Earthq. Eng* [12, 495-516](#).
- Derras, B., Bard, P-Y., and Cotton, F., 2016. Site-conditions proxies, ground-motion variability and data-driven GMPEs. Insights from NGA-West 2 and RESORCE datasets. *Earthquake Spectra* [32, 2027-2056](#).
- Derras, B., Bard, P-Y., Bekkouche, A., 2011. Prédiction Des Paramètres De Nocivité Par Approche Neuronale : Exemple Des Données Kik-Net. 8^{ème} Colloque National de L'AFPS : Vers une maîtrise durable du risque sismique. Paris-France.
- Derras, B., Bard, P-Y., Cotton, F., and Bekkouche, A., 2012. Adapting the neural network approach to PGA prediction: an example based on the KiK-net data. *Bull Seism Soc Am.* [102, 1446–1461](#).
- Derras, B., Bekkouche, A., and Zendagui, D., 2010. Neuronal approach and the use of kik-net network to generate response spectrum on the surface, *Jordan Journal of Civil Engineering*, [4, 12-21](#).
- Derras, B., Cotton, F., Drouet, S., and Bard, P-Y., 2017. Magnitude dependence of stress drop: what does the observed magnitude scaling of ground-motions tell us?, 16th World Conference on Earthquake, Santiago, Chile. [hal-01461210](#).
- Derras, B., 2013. Estimation des paramétrés de nocivité par la méthode neuronale avec l'utilisation des données japonaises. 6^{ème} symposium international sur la construction en zone sismique ([siczs'2013](#)), Chlef, Algérie.
- DGPR., 2014. Risque sismique et sécurité dans les ouvrages hydrauliques. [Rapport MEDDTL-DGPR](#).

- Douglas, J., 2003. Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates, *Earth-Science Reviews* [61, 43–104](#).
- Douglas, J., 2012. Consistency of ground-motion predictions from the past four decades: peak ground velocity and displacement, Arias intensity and relative significant duration, *Bull. Earthq. Eng* [10, 1339-1356](#).
- Douglas, J., and Edwards, B., 2016. Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews* [160, 203-219](#).
- Douglas, J., 2016. [Ground motion prediction equations 1964-2016](#). Department of Civil and Environmental Engineering, University of Strathclyde Glasgow, United Kingdom.
- Douglas, J., Akkar, S., Ameri, A., Bard, P-Y., Bindi, D., Bommer, J. J., Bora, S. S., Cotton, F., Derras, B., Hermkes, M., Kuehn, N. M., Luzi, L., Massa, M., Pacor, F., Riggelsen, C., Sandikkaya, M. A., Scherbaum, F., Stafford, P. J., and Traversa, P., 2014. Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East, *Bull. Earthq. Eng* [12, 341-358](#).
- Douglas, J., and Jousset, P., 2011. Modeling the difference in ground-motion magnitude-scaling in small and large earthquakes. *Seismol. Res. Lett* [82, 504-508](#).
- Drouet, S., and Cotton, F., 2015. Regional Stochastic GMPEs in Low-Seismicity Areas: Scaling and Aleatory Variability Analysis-Application to the French Alps, *Bull. Seism. Soc. Am* [105, 1883–1902](#).
- Drouet, S., Cotton, F., and Guéguen, P., 2010. V_{S30} , κ , regional attenuation and M_w from small magnitude events accelerograms, *Geophys. J. Int* [182, 880–898](#).
- Duval, A. M., Bard, P-Y., Méneroud, J. P., and Vidal, S., 1995. Mapping site effect with microtremors In: Proc 5th Int. Conf. on Seismic Zonation, Nice, France.
- Edwards, B., and Fäh, D., 2013. A stochastic ground-motion model for Switzerland, *Bull. Seismol. Soc. Am* [103, 78–98](#).
- Foti, S., Cox, B.R., Garofalo, F., Hollender, F., Bard, P.Y., Cornou, C., Ohrnberger, M., and Sicilia, D., 2015. Uncertainties in V_s profiles from geophysical tests and their influence on seismic ground response analyses: results from the InterPACIFIC blind Test, 6th International Conference on Earthquake Geotechnical Engineering, Christchurch, New-Zealand.
- Fukushima, Y., and Tanaka. T., 1990. A new Attenuation Relation for horizontal acceleration of strong ground motion in Japan, *Bull. Seismol. Soc. Am* [80, 757-783](#).
- Fukushima, Y., Irikura, K., Uetake, T., and Matsumoto, H., 2000. Characteristics of Observed Peak Amplitude for Strong Ground Motion, *Bull. Seismol. Soc. Am* [90, 545–565](#).
- García, S. R., Romo, M. P., and Mayoral, J. M., 2007. Estimation of peak ground accelerations for Mexican subduction zone earthquakes using neural networks, *Geofísica Internacional* [46, 51-63](#).
- Garofalo, F., Foti, S., Hollender, F., Bard, P. Y., Cornou, C., Cox, B. R., Dechamp, A., Ohrnberger, M., Sicilia, D., and Vergnault, C., 2015b. Interpacific project: comparison of invasive and non-invasive methods for seismic site characterization. Part II: intercomparison between surface-wave and borehole methods, *Soil Dyn and Earthq Eng* [82, 241–254](#).

- Garofalo, F., Foti, S., Hollender, F., Bard, P.Y., Cornou, C., Cox, B.R., Ohrnberger, M., Sicilia, D., Asten, M., Di Giulio, G., Forbriger, T., Guillier, B., Hayashi, K., Martin, A., Matsushima, S., Mercerat, D., Poggi, V., and Yamanaka, H., 2015a. InterPACIFIC project: comparison of invasive and non-invasive methods for seismic site characterization. Part I: Intra-comparison of surface wave methods, *Soil Dyn and Earthq Eng* **82**, 222-240.
- Giardini, D. Woessner, J., Danciu, L., Valensise, G., Grünthal, G., Cotton, F., Akkar, S., Basili, R., Stucchi, M., Rovida, A., Stromeyer, D., Arvidsson, R., Meletti, F., Musson, Ptilakis, K., Sesetyan, K., Demircioglu, M. B., Crowley, H., Pinho, R., Douglas, J., Fonseca, J., Erdik, M., Campos-Costa, A., Glavatovic, B., Makropoulos, K., Lindholm, C., and Cameelbeeck, T., 2013. Seismic Hazard Harmonization in Europe (SHARE): Online Data Resource, doi: 10.12686/SED-00000001-SHARE.
- Gregor, N. J., W.J. Silva, I. G. Wong., and R. Youngs., 2002. Ground-motion attenuation relationships for Cascadia subduction zone mega thrust earthquakes based on a stochastic finite fault model, *Bull. Seismol. Soc. Am* **92**, 1923-1932.
- Gregor, N., Abrahamson, N. A., Atkinson, G. M., Boore, D. M., Bozorgnia, Y., Campbell, K. W., Chiou, B. S-J., Idriss, I. M., Kamai, R., Seyhan, E., Silva, W., Stewart, J. P., and Youngs, R., 2014. Comparison of NGA-West2 GMPEs, *Earthquake Spectra* **30**, 1179–1197.
- Guéguen, P., 2016. Predicting nonlinear site response using spectral acceleration vs PGV/ V_{S30} : a case history using the Volvi-test site, *Pure Appl Geophys* **173**, 2047-2063.
- Haghshenas, E., Bard, P-Y., Theodulidis, N., SESAME WP04 Team., 2008. Empirical evaluation of microtremor H/V spectral ratio, *Bull Earthq Eng* **6**, 75-108.
- Hayashida, T., and Tajima, F., 2007. Calibration of amplification factors using KiK-net strong-motion records: toward site effective estimation of seismic intensities. *Earth Planets Space* **59**, 1111-1125.
- Hermkes, M., Kuehn, N. M., Riggelsen, C., 2014. Simultaneous quantification of epistemic and aleatory uncertainty in GMPEs using Gaussian process regression. *Bull Earthq Eng* **12**, 449-466.
- Holzer, T. L., Padovani, A. C., Bennett, M. J., Noce, T. E., and Tinsley, J. C., 2005. Mapping NEHRP V_{S30} Site Classes, *Earthquake Spectra* **21**, 353–370.
- Hu, Y. H., and Hwang, J. N., 2002. [Handbook of neural network signal processing](#), Boca Raton, FL: CRC.
- Idriss, I. M., 2011. Use of V_{S30} to represent local site condition. Proceedings of the 4th IASPEI /IAEE international symposium. Effects of source geology on seismic motion, California, USA.
- Idriss, I. M., 2014. NGA-West2 Model for Estimating Average Horizontal Values of Pseudo-Absolute Spectral Accelerations Generated by Crustal Earthquakes, *Earthquake Spectra*, **30**, 1155-1177.
- Joyner, W. B., and Boore, D. M., 1981. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am* **71**, 2011–2038.
- Joyner, W. B., and Boore, D. M., 1993. Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am* **83**, 469-487.
- Kamai, R., Abrahamson, N. A., Walter, J., and Silva, M., 2014. Nonlinear Horizontal Site Amplification for Constraining the NGA-West2 GMPEs. *Earthquake Spectra* **30** 1223-1240.

- Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., and Fukushima Y., 2006. A New Attenuation Relation for Strong Ground Motion in Japan Based on Recorded Data, *Bull. Seismol. Soc. Am* **96**, [879-897](#).
- Kashima, T., 2009. Strong Motion Observation. [IISEE Lecture Note for the 2008-2009 regular course](#).
- Kerh, T., and Ting, S. B., 2005. Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system, *Engineering Applications of Artificial Intelligence* **18**, [857-866](#).
- Kokusho, T., 2004. Nonlinear site response and strain-dependent soil properties, *Curr. Sci* **87**, [1363-1369](#).
- Kokusho, T., and Sato, K., 2008. Surface-to-base amplification evaluated from KiK-net vertical array strong motion records. *Soil Dyn Earthq Eng* **28**, 707-716.
- Kotha, S. R., Bindi, D., Cotton, F., 2016. Partially non-ergodic region specific GMPE for Europe and Middle East, *Bull Earthq Eng* **14**, [1245-1263](#).
- Kramer, S., 1996. Geotechnical earthquake engineering. Prentice-Hall Civil Engineering and Engineering Mechanics Series.
- Lee, V. W., Trifunac, M. D., 2010. Should average shear-wave velocity in the top 30m of soil be used to describe seismic amplification, *Soil Dyn Earthq Eng* **30**, [1250-1258](#).
- Lemoine, A., Douglas, J., and Cotton, F., 2012. Testing the Applicability of Correlations between Topographic Slope and V_{s30} for Europe, *Bull. Seism Soc. Am* **102**, [2585-2599](#).
- Liu, B-Y., Ye, L.Y., Xiao, M. L, and Miao, S., 2006. Peak Ground Velocity Evaluation by Artificial Neural Network for West America Region, in Proceedings of the 13th International Conference on Neural Information Processing, Part II (ICONIP '06), Hong Kong [4233, 942-951](#).
- Luco, J. E., 1985. On strong ground-motion estimates based on models of the radiated spectrum. *Bull. Seism. Soc. Am*, [75, 641-649](#).
- Lussou, P., 2001. Calcul du mouvement sismique associé à un séisme de référence pour un site donné avec prise en compte de l'effet de site. Méthode empirique linéaire et modélisation de l'effet de site non linéaire, Thèse de doctorat, Université Joseph Fourier, Grenoble, France. [hal-00685419](#).
- Luzi, L., Hailemikael, S., Bindi, D., Pacor, F., Mele, F., and Sabetta, F., 2008. ITACA (ITalian ACcelerometric Archive): a web portal for the dissemination of the Italian strong motion data. *Seismol Res Lett* **79**, [716-722](#).
- Luzi, L., Puglia, R., Pacor, F., Gallipoli, M. R., Bindi, D., and Mucciarelli, M., 2011. Proposal for a soil classification based on parameters alternative or complementary to V_{s30} . *Bull Earthq Eng* **9**, [1877-1898](#).
- Maass, W., 1997. Networks of Spiking Neurons: The Third Generation of Neural Network Models, *Neural Networks* **10**, [1659-1671](#).
- Maufroy, E., Chaljub, E., Hollender, F., Bard, P-Y., Kristek, J., Moczo, P., De Martin, F., Theodoulidis, N., Manakou, M., Guyonnet-Benaize, C., Hollard N., and Pitilakis, K., 2016. 3D Numerical simulation and ground motion prediction: Verification, validation and beyond - lessons from the E2VP project, *Soil Dyn and Earthq Eng* **91**, [53-71](#).
- Molas, G.L., and F. Yamazaki., 1995. Attenuation of earthquake Ground motion in Japan including deep focus events, *Bull. Seismol. Soc. Am* **85**, [1343-1358](#).

- Molkenthin, C., Scherbaum, F., Griewank, A., Kuehn, N., and Stafford, P., 2014. A study of the sensitivity of response spectral amplitudes on seismological parameters using algorithmic differentiation, *Bull. Seismol. Soc. Am* [104, 2240–2252](#).
- Moss, R E., 2008. Quantifying measurement uncertainty of thirty-meter shear-wave velocity. *Bull. Seism Soc. Am* [98, 1399-1411](#).
- Mucciarelli, M., 1998. Reliability and applicability of Nakamura's technique using microtremors: an experimental approach. *J. Earthq. Eng* [2, 625-638](#).
- Mucciarelli, M., and Gallipoli, M. R., 2006. Comparison between V_{s30} and other estimates of site amplification in Italy. In: Proceedings 1st European conference on earthquake engineering and seismology, Geneva, Switzerland, [Paper Number: 270](#).
- Nakamura, Y., 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. Quarterly Report of Railway Technical Research Institute [30, 25–33](#).
- Pousse, G., 2005. [Analyse des données accélérométriques de K-NET et KIK-NET: implications pour la prédiction du mouvement sismique -accélérogrammes et spectres de réponse et la prise en compte des effets de site nonlinéaire](#), Thèse de doctorat, Université Joseph Fourier, Grenoble, France.
- Pousse, G., Berge-Thierry, C., Bonilla, F., and Bard, P. Y., 2005. Eurocode 8 Design response spectra evaluation using the K-net Japanese database, *J. Earthq. Eng* [9, 547–574](#).
- Régnier, J. H., Cadet, L. F., Bonilla, L. F., Bertrand, E., and Semblat, J-F., 2013. Assessing Nonlinear Behavior of Soils in Seismic Site Response: Statistical Analysis on KiK-net Strong-Motion Data. *Bull Seism Soc Am* [103, 1750-1770](#).
- Régnier, J., Cadet, H., and Bard, P.-Y., 2015. Evolution of non-linearity indicators with site proxies and ground motion level: an investigation from KiK-net data. Task 2.1.1: Strong motion and associated geotechnical parameters, plenary Meeting, ENS Cachan. France.
- Régnier, J., Cadet, H., and Bard, P-Y., 2016. Empirical quantification of the impact of non-linear soil behavior on site response. *Bull Seism Soc Am*, [106, 1710-1719](#).
- Rietbrock, A., Strasser, F. O., and Edwards, B., 2013. A stochastic earthquake ground-motion prediction model for the United Kingdom, *Bull. Seismol. Soc. Am*. [103, 57–77](#).
- Robitaille, B., Marcos, B., Veillette, M., and Payr, G., 1996. Modified quasi-Newton methods for training neural networks, *Computers chem. Engng* [20, 1133-1140](#).
- Sadigh, K., Chang, C-Y., Egan, J. A., Makdisi, F. I., and Youngs, R. R., 1997. Attenuation relationships for shallow crustal earthquakes based on California strong motion data, *Seismol. Res. Lett.* [68,180–189](#).
- Sandikkaya, M. A., Yılmaz, M. T., Bakır, B. S., and Yılmaz, Ö., 2010. Site classification of Turkish national strong-motion stations. *J Seismol* [14, 543–563](#).
- Sandikkaya, M.A., Akkar, S., and Bard, P.Y., 2013. A Nonlinear Site Amplification Model for the Next Pan European Ground Motion Prediction Equations, *Bull. Seismol. Soc. Am* [103, 19-32](#)
- Scasserra, G., Stewart, J. P., Bazzurro, P., Lanzo, G., and Mollaioli, F., 2009. A Comparison of NGA Ground-Motion Prediction Equations to Italian Data, *Bull. Seismol. Soc. Am* [99, 2961–2978](#).
- Seyhan, E., and Stewart, J. P., 2014. Semi-empirical nonlinear site amplification from NGA-West 2 data and simulations, *Earthquake Spectra* [30, 1241-1256](#).

- Seyhan, E., Stewart, J. P., Ancheta, T. D., Darragh, R. B., Graves, R. W., 2014. NGA-West2 site database. *Earthquake Spectra*, **30**, [1007-1024](#).
- Shanno, D. F., and Kettler, P. C., 1970. Optimal Conditioning of Quasi-Newton Methods. *Math. Comp* **24**, [657-664](#).
- Stafford, P.J., 2014. Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models. *Bull. Seismol. Soc. Am* **104**, [702-719](#).
- Steidl, J. H., Tumarkin, A. G., and Archuleta, R. J., 1996. What is a reference site?, *Bull. Seismol. Soc. Am.* **86**, [1733–1748](#).
- Strasser, F.O., Abrahamson, N.A., and Bommer, J.J., 2009. Sigma: Issues, Insights, and Challenges, *Seismol. Res. Lett* **80**, [40-56](#).
- Theodulidis, N., Kalogeras, I., Papazachos, C., Karastathis, V., Margaris, B., Papaioannou, Ch., and Skarlatoudis, A., 2004. HEAD 1. O: a unified Hellenic accelerogram database, *Seismol Res Lett* **75**, [36–45](#).
- Wald, D. J., and Allen, T. I., 2007. Topographic slope as a proxy for seismic site conditions and amplifications. *Bull Seism Soc Am* **97**, [1379-1395](#).
- Walling, M., W. J. Silva., and Abrahamson, N. A., 2008. Non-linear site amplification factors for constraining the NGA models, *Earthquake Spectra* **24**, [243-255](#).
- Wierenga, B., and Kluytmans, J., 1994. Neural nets versus marketing models in time series analysis: a simulation studies. Proceedings of the 23 annual conference of the European Marketing Association, Maastricht, 1139–1153.
- Wooddell, K. E., and Abrahamson, N. A., 2014. Classification of main shocks and aftershocks in the NGA-West 2 database, *Earthquake Spectra* **30**, [1257–1267](#).
- Wu, Y-M., Teng, T-L., Shin, T-C., and Hsiao, N-C., 2003. Relationship between Peak Ground Acceleration, Peak Ground Velocity, and Intensity in Taiwan. *Bull Seism Soc Am* **93**, [386-396](#).
- Yamada, T., 2013. [Uncertainty in the assessment of seismic hazard](#), The 3rd annual meeting of the seismic hazard assessment for the next generation map China-Japan-Korea cooperative program, Sendai, Japan.
- Yenier, E., Sandikkaya, M. A., and Akkar, S., 2010. Report on the fundamental features of the extended strong motion databank prepared for the SHARE project, pp. 44. Deliverable 4.1 of Seventh Framework Programme Project Seismic Hazard Harmonization in Europe (SHARE), 34 pages, Ankara.
- Yoo, SH., and Mayeda, K., 2013. Validation of Non-Self-Similar Source Scaling Using Ground Motions from the 2008 Wells, Nevada, Earthquake Sequence. *Bull Seism Soc Am*, **103**, [2508–2519](#).
- Youngs, R R., Chiou, S., Silva, W., and Humphrey, J., 1997. Strong ground motion attenuation relationships for subduction zone earthquakes, *Seismol. Res. Lett* **68**, [58–73](#).
- Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., and Fukushima, Y., 2006. Attenuation relations of strong ground-motion in Japan using site classification based on predominant period, *Bull Seism Soc Am* **96**, [898–913](#).

ANNEXES



Annexe 1 : Interface graphique des modèles neuronaux

1. Interface graphique sous Excel pour la prédiction des PSA : application à la base de données RESORCE

Les valeurs de PGA, PGV et PSA en fonction des R_{JB} , M_w , Depth, V_{s30} et FM sont obtenues par l'équation II.27. Pour rendre l'utilisation de cette équation plus facile et conviviale, une interface graphique sous Microsoft Excel a été élaborée. Cette interface donne en plus de PGA et PGV, les ordonnées spectrales en fonction de la période d'oscillation (figure.A.1).

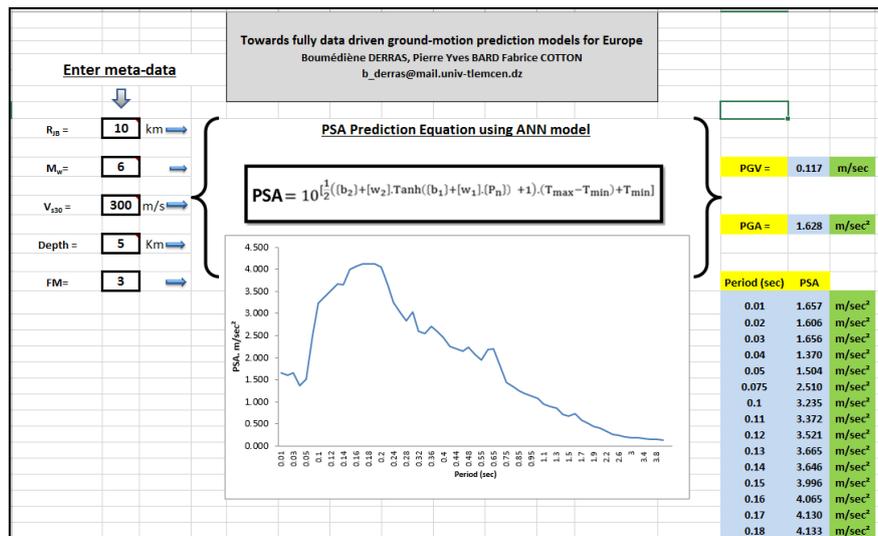


Figure.A.1 Interface graphique, sous Excel, du modèle d'atténuation donnant le PGA, PGV et PSA : application au set de données RESORCE

2. Interface graphique sous Excel pour la prédiction des PSA : application à la base de données NGA-West 2

Toujours sous Excel, qui est considéré comme un outil universel, nous donnons la possibilité aux utilisateurs d'avoir les PGA, PGV et le pseudo-spectre de réponse ainsi que la variation de ces quantités avec la distance (figure.A.2). Le code ainsi fourni est appliqué pour le set de données NGA-West 2.

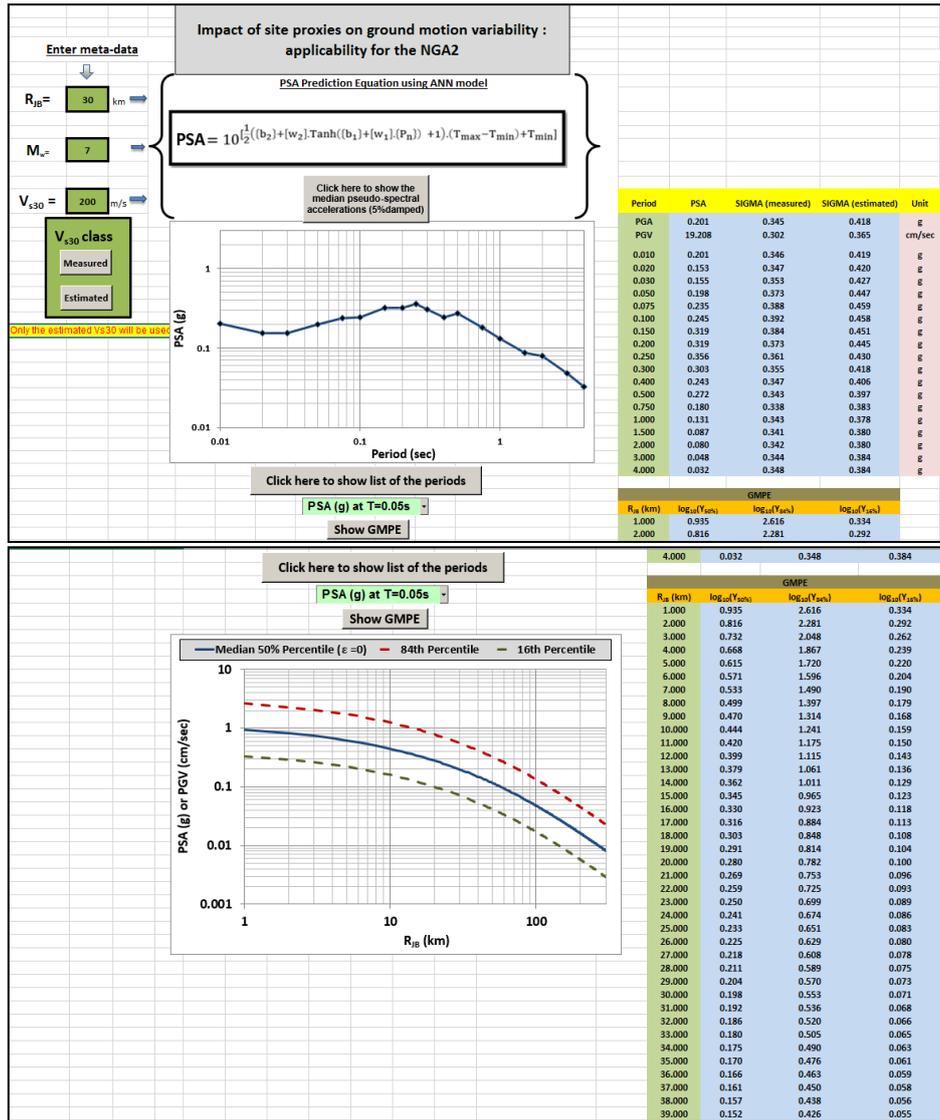


Figure.A.2 Interface graphique sous Excel : Modèle neuronal de prédiction du mouvement sismique élaboré avec le set de données NGA-West 2.

Après la spécification des R_{JB} , M_w et V_{s30} et la nature de V_{s30} (mesurée/estimée) par l'utilisateur, le code affiche les résultats sous formes tabulaires et sous forme graphique en donnant les ordonnées spectrales et les l'atténuation du mouvement sismique avec la distance (médiane et médiane $\pm\sigma$).

Annexe 2 : Article de Douglas et al. (2014)

Bull Earthquake Eng (2014) 12:341–358
DOI 10.1007/s10518-013-9522-8

ORIGINAL RESEARCH PAPER

Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East

John Douglas · Sinan Akkar · Gabriele Ameri · Pierre-Yves Bard · Dino Bindi · Julian J. Bommer · Sanjay Singh Bora · Fabrice Cotton · Boumédiène Derras · Marcel Hermkes · Nicolas Martin Kuehn · Lucia Luzi · Marco Massa · Francesca Pacor · Carsten Riggelsen · M. Abdullah Sandikkaya · Frank Scherbaum · Peter J. Stafford · Paola Traversa

Received: 16 January 2013 / Accepted: 14 September 2013 / Published online: 29 September 2013
© Springer Science+Business Media Dordrecht 2013

Abstract This article presents comparisons among the five ground-motion models described in other articles within this special issue, in terms of data selection criteria, characteristics of the models and predicted peak ground and response spectral accelerations. Comparisons are also made with predictions from the Next Generation Attenuation (NGA) models to

J. Douglas (✉)
DRP/RSV, BRGM, Orléans, France
e-mail: J.Douglas@brgm.fr

S. Akkar · M. A. Sandikkaya
Middle East Technical University, Ankara, Turkey

G. Ameri
FUGRO-Geoter, Auriol, France

P.-Y. Bard · F. Cotton · B. Derras · M. A. Sandikkaya
ISTerre, Université Joseph Fourier, CNRS, Grenoble, France

B. Derras
RISk Assessment and Management laboratory (RISAM),
Université Abou Bekr Belkaïd, Tlemcen, Algeria

D. Bindi
GFZ-German Research Center for Geosciences, Potsdam, Germany

J. J. Bommer · P. J. Stafford
Imperial College London, London, UK

S. S. Bora · M. Hermkes · N. M. Kuehn · C. Riggelsen · F. Scherbaum
Inst. Erd- und Umweltwissenschaften, Universitaet Potsdam, Potsdam, Germany

L. Luzi · M. Massa · F. Pacor
INGV, Milan, Italy

P. Traversa
EDF, Aix en Provence, France

which the models presented here have similarities (e.g. a common master database has been used) but also differences (e.g. some models in this issue are nonparametric). As a result of the differing data selection criteria and derivation techniques the predicted median ground motions show considerable differences (up to a factor of two for certain scenarios), particularly for magnitudes and distances close to or beyond the range of the available observations. The predicted influence of style-of-faulting shows much variation among models whereas site amplification factors are more similar, with peak amplification at around 1s. These differences are greater than those among predictions from the NGA models. The models for aleatory variability (σ), however, are similar and suggest that ground-motion variability from this region is slightly higher than that predicted by the NGA models, based primarily on data from California and Taiwan.

Keywords Strong-motion data · Ground-motion models · Ground-motion prediction equations · Style of faulting · Site amplification · Aleatory variability · Epistemic uncertainty · Europe · Middle East

1 Introduction

The collection of five ground-motion models presented in other articles in this special issue has similarities to the five sets of ground-motion prediction equations (GMPEs) derived during the Next Generation Attenuation (NGA) project (Power et al. 2008) and described in a special issue of *Earthquake Spectra* in 2008. Firstly, both sets of models were derived for state-of-the-art seismic hazard assessments for shallow active crustal seismicity in specific geographical regions: western North America (specifically California) for NGA, and Europe and the Middle East here. In passing it may be noted, however, that the NGA models have been shown to be applicable to Europe and the Middle East (Stafford et al. 2008). Secondly, all five GMPEs presented here were derived based on records chosen from a common strong-motion database (RESORCE, see Akkar et al. 2013c), whose compilation has similarities to the procedure followed when developing the NGA database (Chiou et al. 2008). Thirdly, careful data selections were made by each of the GMPE developers and state-of-the-art derivation techniques were followed. Lastly, the collection of GMPEs produced seeks to acknowledge the still considerable epistemic uncertainty present in the assessment of earthquake shaking (e.g. Douglas 2010). For the application of the NGA models within the USGS national hazard calculations additional branches were added to the logic-tree in certain magnitude-distance bins to capture epistemic uncertainty beyond that represented by these models (Petersen et al. 2008).

On the other hand, the collection presents significant differences with respect to the NGA models. Firstly, unlike the NGA models, which were all derived using regression analysis, generally the random-effects approach (e.g. Abrahamson and Youngs 1992), although with some coefficients fixed a priori based on physical arguments, here only two models were derived in this way (Akkar et al. 2013a,b; Bindi et al. 2013). Two of the others are non-parametric models derived using data-driven approaches (Derras et al. 2013; Hermkes et al. 2013) and the other model (Bora et al. 2013) makes predictions of response spectral accelerations using random-vibration theory based on empirical models for Fourier amplitude spectra and durations. Secondly, unlike the multi-year NGA project, which involved extensive interactions among developers and other project participants (leading to multiple iterations of the models), the models presented here were derived in a much shorter period and following limited communication among groups. Although the development of RESORCE

was funded by SHARE and SIGMA, which led to some interactions among the model developers, this special issue is principally the fruit of parallel and independent efforts (by authors in five countries) rather than a coordinated national project. This means that the differences in the approaches used are larger than for NGA. It is possible that the use of multiple approaches for the models presented in this volume more effectively captures epistemic uncertainty in terms of the centre, the body and the range of technically-defensible interpretations of the available data (USNRC 2012). Thirdly, the independent parameters used by the models presented here are: common among groups (all use only: moment magnitude, M_w ; distance to the surface projection of the fault, R_{JB} ¹; the same style-of-faulting classifications; and the average shear-wave velocity to 30 m, V_{S30}) and fewer (e.g. data were insufficient to include terms involving sediment depth, $Z_{1.0}$ or $Z_{2.5}$, or depth to the top of rupture, Z_{TOR}) than in the NGA models. This makes comparisons among the models and their use in future seismic hazard assessments easier since no adjustments for differences in independent parameters (e.g. Bommer et al. 2005) are required. Lastly, no strict model requirements were agreed at the beginning of the derivation procedure, unlike those imposed on the NGA model developers, which means that the models presented here have varying ranges of applicability in terms of, for example, magnitude and distance.

Despite the differences between the NGA project and this special issue, the NGA comparison article by Abrahamson et al. (2008) is used as a template for this article comparing the five models presented in this issue, namely those by: Akkar et al. (2013a,b) (their model using R_{JB}), Bindi et al. (2013) (their model using R_{JB} and V_{S30} directly²), Bora et al. (2013), Derras et al. (2013) and Hermkes et al. (2013). This decision means that comparisons between the figures presented here can be readily made to those shown in Abrahamson et al. (2008) because the same choices of independent parameters and the same axes and scales are used (also to help in making these comparisons the same figure numbering has been retained). Note that some of the graphs show predictions up to M_w 8, for consistency with Abrahamson et al. (2008), even though some developers do not recommend their models are applied for such large earthquakes (Table 1). To further facilitate comparisons with the NGA models, predictions from the GMPs of Boore and Atkinson (2008) are included on the figures. This NGA model was chosen from among the five because it is the most similar to those presented in this special issue through its use of R_{JB} and fewer independent variables, e.g. no terms using Z_{TOR} or $Z_{1.0}$ (or $Z_{2.5}$) are included. Because the models presented here have fewer independent parameters and the aleatory variabilities (standard deviations) of the models are all homoscedastic (uniform for all independent and dependent variables) some figures drawn by Abrahamson et al. (2008) are not relevant and are not drawn. They are replaced with figures showing other features of the models that are not covered by the other graphs, for example the influence of style of faulting on ground-motion predictions (e.g. Bommer et al. 2003).

The next section presents the data selection criteria used by the different groups. The following sections compare different aspects of the models in terms of: attenuation with distance, scaling with magnitude, style-of-faulting factors, site amplification, predicted response spectra and aleatory variability. The article ends with some brief conclusions.

¹ Akkar et al. (2013a,b) also derived GMPs using epicentral (R_{epi}) and hypocentral (R_{hyp}) distances and Bindi et al. (2013) also derived GMPs using hypocentral (R_{hyp}) distance. These are not considered in this article.

² Bindi et al. (2013) also derived GMPs using EC8 site classes rather than V_{S30} directly. These models are not considered in this article.

Table 1 Number of different earthquakes, stations and records used to derive the five models, magnitude and distance ranges of the data used, the ranges of applicability recommended by the model developers and the exclusion criteria used to select the records used to derive the model

Model	Akkar et al.	Bindi et al. (V_{S30} model)	Bora et al.	Derras et al.	Hemlkes et al.
Number of earthquakes (E)	221	225	369	320	279
Number of stations (S)	322	345	341	201	251
Number of records (R)	1,041	1,224	1,232	1,088	835
R/E	4.7	4.8	3.3	3.4	3.0
M_{\min} to M_{\max} (data used)	4.0 to 7.6	4.0 to 7.6	4.0 to 7.6	3.6 to 7.6	4.0 to 7.6
M_{\min} to M_{\max} (recommended)	4.0 to 8.0	4.0 to 7.6	4.0 to 7.6	4.0 to 7.0	4.0 to 7.6
R_{\min} to R_{\max} (km) (data used)	0 to 200	0 to 300	0 to 200	0 to 547 km	0 to 200
R_{\min} to R_{\max} (km) (recommended)	0 to 200	0 to 300	0 to 200	5 to 200 km	0 to 200
Record exclusion criteria (other than in terms of magnitude and distance)	Singly-recorded earthquakes; all three components not available; focal depth greater than 30km; sites with no measured V_{S30} ; structural period beyond usable period range defined by Akkar and Bommer (2006); events with $M_w < 5$ with fewer than 3 records to constrain SoF ratios; unknown or oblique style of faulting; not free-field.	Unknown style of faulting; sites with no measured V_{S30} ; singly-recorded earthquakes; only records with low-pass cut-off frequency lower than 20Hz and outside passband of high-pass filter all three components not available; focal depth > 35 km.	Not representative of shallow crustal event; unknown style of faulting; only one horizontal component; sites with no measured V_{S30} ; poor quality record; high-pass cut-off frequency higher than Brune-source corner frequency for stress drop of 100bars.	Focal depth more than 25 km; sites with no measured V_{S30} ; unknown style of faulting	Unknown style of faulting; sites with no measured V_{S30} ; not free-field conditions; high-pass cut-off frequency higher than 0.25Hz

2 Data selection criteria

All GMPE developers started with the same RESORCE archive, which is presented by Akkar et al. (2013c) in this special issue. At the time of model derivation this databank contained 5,882 mainly-triaxial accelerograms (from $0 \leq R \leq 587$ km) from 1,814 earthquakes (with $2.8 \leq M_w \leq 7.8$) and 1,540 different strong-motion stations. The five groups of developers applied different selection and exclusion criteria, which led to them using between only 14 and 21 % of the available accelerograms (see Table 1). The same magnitude ranges were used by all groups, except by Derras et al. (2013) who used a slightly lower minimum magnitude (3.6 rather than 4.0), to select their data and only Bindi et al. (2013) and Derras et al. (2013) varied from the distance cut-off of 200 km (using 300 and 547 km, respectively, instead). None of the RESORCE developers used selection criteria based on earthquake type (e.g. mainshock, aftershock or swarm) or considered its influence on ground motions. Consequently all types of earthquakes (including aftershocks) were selected, unlike Boore and Atkinson (2008) who exclude this type of event when deriving their NGA model and other NGA models that included terms in their models to distinguish between mainshocks and aftershocks. As discussed by Douglas and Halldórsson (2010) there is considerable doubt over the classification of European earthquakes into mainshock, aftershock and swarm and their analysis using the data and model of Ambraseys et al. (2005) suggested that the influence of earthquake type on ground motions is limited. A similar conclusion is reached by Bindi et al. (2013) after examining the residuals for their model separated into mainshock and aftershock classes. The five model databases principally comprise records from normal and strike-slip earthquakes, with a smaller number of accelerograms from reverse-faulting events. The distribution of records by style-of-faulting is reasonably uniform with respect to magnitude but the largest ($M_w > 7$) earthquakes are mainly from strike-slip earthquakes in Turkey (Kocaeli and Düzce) and Iran (Manjil). The variation in the final databases principally results from the exclusion of data based on the filters used to process the accelerograms and maximum focal depth criteria. The result of these various selection criteria are different sizes of databases used for the derivations of the five models (Table 1). All of the models were derived using roughly 1,000 strong-motion records.

One major difference between the data used by the models compared here and that used for the NGA models is the large number of poorly-recorded earthquakes. This is indicated by the mean number of records per earthquake for the five RESORCE models being between 3.0 and 4.8 (Table 1) whereas the mean number of records per earthquake for the NGA models varies between 13.1 and 27.1. This difference implies that the terms of the models related to the earthquake source (e.g. style-of-faulting terms and between-event standard deviations) are more poorly constrained than they are in the NGA models, which, as shown below, leads to significant differences in these aspects of the models. The complexity of the source modelling in some of the NGA models, however, means that these models may suffer from trade-offs, for example between the effect of Z_{TOR} and style-of-faulting.

3 Attenuation with distance

The decay with distance from the source for peak ground acceleration (PGA) and spectral acceleration for a structural period of 1s and 5 % critical damping [SA(1s)] can be seen in Fig. 1, for $V_{S30}=760$ m/s, i.e. NEHRP B/C boundary (Building Seismic Safety Council, 2009) (soft rock, Eurocode 8 class B), and in Fig. 2, for $V_{S30}=270$ m/s, i.e. NEHRP D (soft soil, Eurocode 8 class C). Generally the decay rates are similar as are the predicted ground motions,

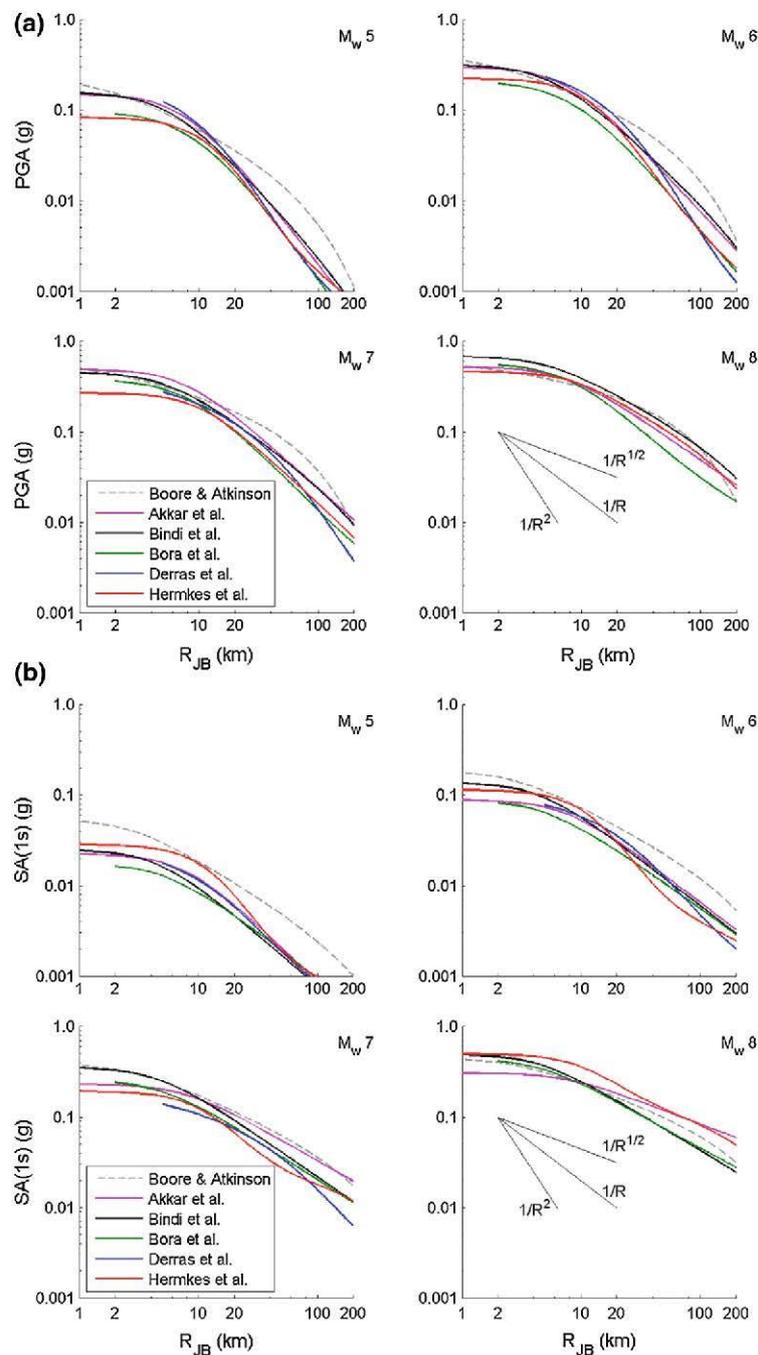


Fig. 1 Comparison of distance scaling for strike-slip earthquakes for $V_{S30} = 760$ m/s (NEHRP B/C boundary) for M_w 5 (top left), 6 (top right), 7 (bottom left) and 8 (bottom right) for **a** PGA and **b** SA(1s). The predictions from the model of Derras et al. (2013) are not shown for M_w 8 since this is outside its range of applicability. The other models are shown for this magnitude even though some developers do not recommend their application for such large events

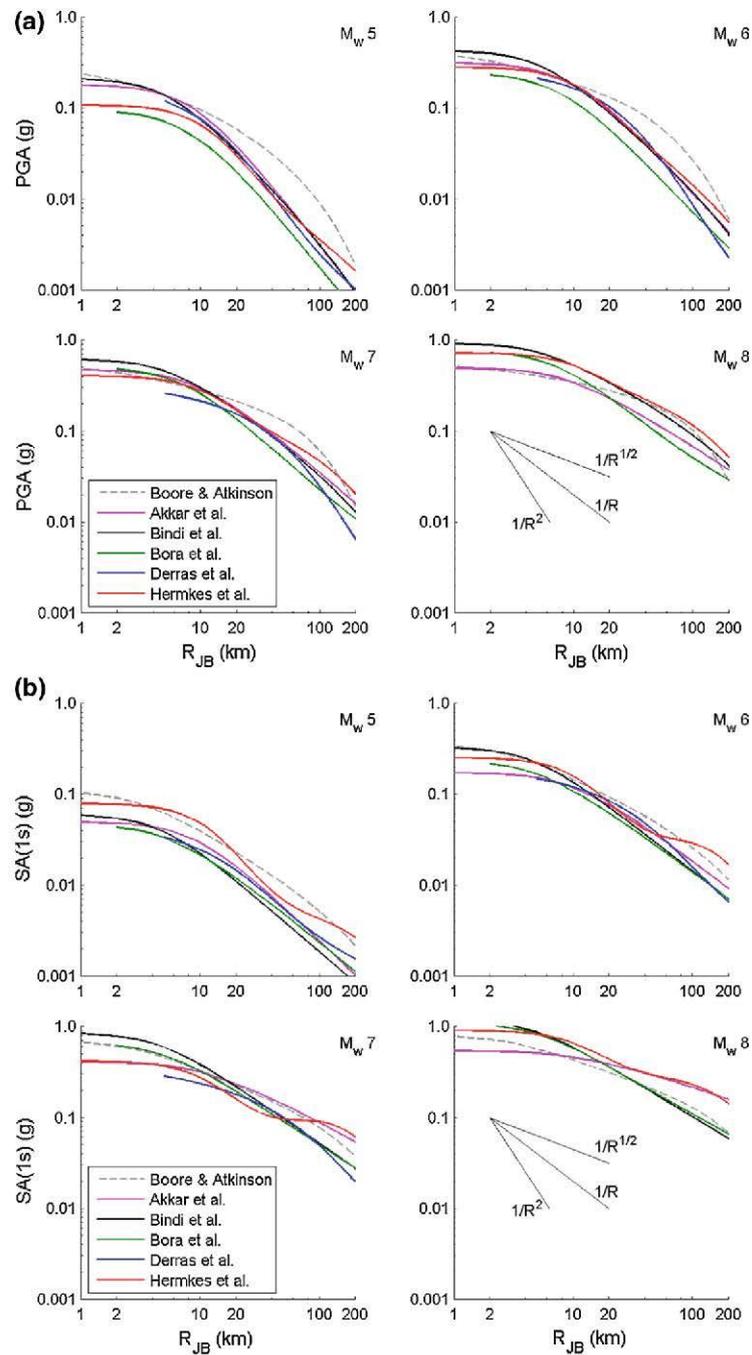


Fig. 2 Comparison of distance scaling for strike-slip earthquakes for $V_{S30} = 270$ m/s (NEHRP D) for M_w 5 (top left), 6 (top right), 7 (bottom left) and 8 (bottom right) for **a** PGA and **b** SA(1s). The predictions from the model of Derras et al. (2013) are not shown for M_w 8 since this is outside its range of applicability. The other models are shown for this magnitude even though some developers do not recommend their application for such large events

particularly for small and moderate events and PGA. Predictions from the models derived by standard regression techniques (Akkar et al. 2013a,b; Bindi et al. 2013) are comparable except at the limits of their applicability (M_w 8 and close to the source of large earthquakes, $R_{JB} < 10$ km). Bindi et al. (2013) include an anelastic attenuation³ term for short periods whereas Akkar et al. (2013a,b) tried including such a term but found that it converged to a non-physical value and hence they removed it from their functional form. Predictions from the nonparametric models show considerable variations and the model of Hermkes et al. (2013) shows a complex decay rate, with a change of slope (often flattening) starting around 50 km. Despite all models having been derived from a common original archive (even if the final databases used differed), a factor of two difference in predicted median ground motions from the models is not uncommon, except for magnitudes and distance near the centre of the available data (e.g. M_w 6).

As is becoming commonly recognised and modelled, the decay of earthquake ground motions is magnitude dependent. This effect can be seen by comparing the decay rates for M_w 5 (roughly $1/R^{1.5}$ for PGA) to those for M_w 8 (slower than $1/R$). The predicted ground motions from the RESORCE models all decay more rapidly than those from the GMPEs of Boore and Atkinson (2008), particularly for PGA, which leads to much lower predicted ground motions at moderate distances (roughly 20–100 km) from these models compared to Boore and Atkinson (2008). Boore and Atkinson (2008) note that their distance dependence for small earthquakes and long periods may be biased towards a decay that is less rapid than the true decay. The faster decay of ground motions in Italy (from where a considerable portion of the data used to develop the RESORCE models comes) than in California was previously noted by Scasserra et al. (2009).

4 Magnitude scaling

The magnitude scaling of the five models show the expected behaviour of higher scaling at long structural periods (Fig. 3). All models show nonlinear magnitude scaling with, generally, lower dependence of ground motions on magnitude for large events. This nonlinear behaviour is expected from physical models (e.g. Douglas and Jousset 2011). Some studies (e.g. Schmedes and Archuleta 2008) provide physical arguments for oversaturation of short-period ground motions for large earthquakes (i.e. ground motions that decrease as magnitude increases). This effect is not seen for any of the final RESORCE models for magnitudes within their range of applicability. However, when Akkar et al. (2013a,b) included a cubic magnitude term they found that the obtained model predicted oversaturation for $M_w > 7.25$, which they considered physically unrealistic and hence they finally adopted a functional form that did not allow such oversaturation. They note, however, that due to a lack of data from large earthquakes in Europe and the Middle East there is considerable epistemic uncertainty in magnitude scaling for $M_w > 7.5$ and hence they suggest including additional branches in a logic tree to account for this uncertainty. As for the distance decay, within the magnitude range that is well covered by data (M_w 5–7) the models predict similar spectral accelerations whereas for larger earthquakes the models differ greatly, depending on whether they are solely driven by the data or the functional form assumed. The magnitude scaling of the RESORCE models is broadly in line with that predicted by the Boore and Atkinson (2008) GMPEs, although because of the lower attenuation predicted

³ The expression ‘anelastic attenuation’ is only strictly valid for GMPEs for Fourier amplitudes and not response spectral ordinates.

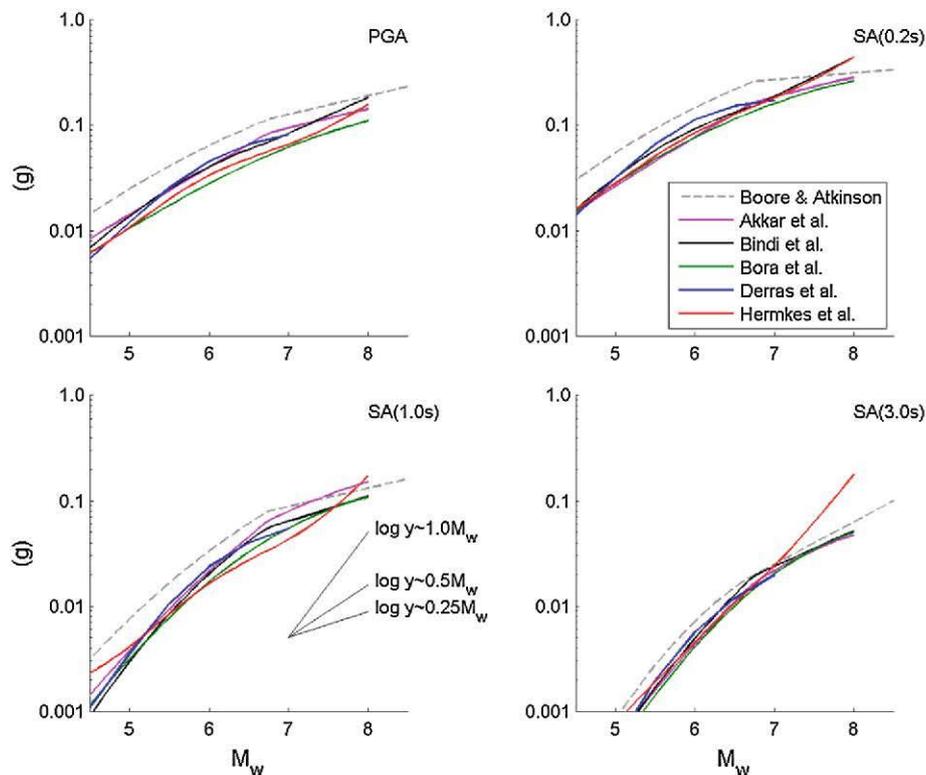


Fig. 3 Comparison of magnitude scaling of the median ground motion for strike-slip earthquakes and $V_{S30} = 760$ m/s (NEHRP B/C boundary) at $R_{JB} = 30$ km for PGA (*top left*), SA(0.2s) (*top right*), SA(1.0s) (*bottom left*) and SA(3.0s) (*bottom right*). Predictions are generally shown up to M_w 8 even though some developers do not recommend their models for such large events

by this model there is a considerable offset in the predictions at the considered distance of 30 km.

5 Style-of-faulting factors

The effect of style of faulting (faulting mechanism) on strong ground motion was highlighted by the review of [Bommer et al. \(2003\)](#), who compared predictions of the reverse-to-strike-slip spectral ratios ($F_{R,SS}$) for various GMPEs (their Fig. 3) and who also discussed the limited number of estimates of the ratio of normal-to-strike-slip motions ($F_{N,SS}$) then available. In the decade since then many more estimates of these factors have been published as part of GMPEs, including in the NGA models, but they still show considerable dispersion. Nevertheless, as shown by the example of the [Boore and Atkinson \(2008\)](#) ratios plotted on Fig. 4, reverse-faulting events are often thought to generate slightly higher amplitude motions than strike-slip earthquakes that in turn are slightly higher than motions from normal-faulting earthquakes.

Figure 4 compares $F_{R,SS}$ and $F_{N,SS}$ for the five RESORCE models [and those of [Boore and Atkinson \(2008\)](#)]. All developers, except [Hermkes et al. \(2013\)](#), assumed ratios that are independent of magnitude and distance. Using a nonparametric approach [Hermkes et al.](#)

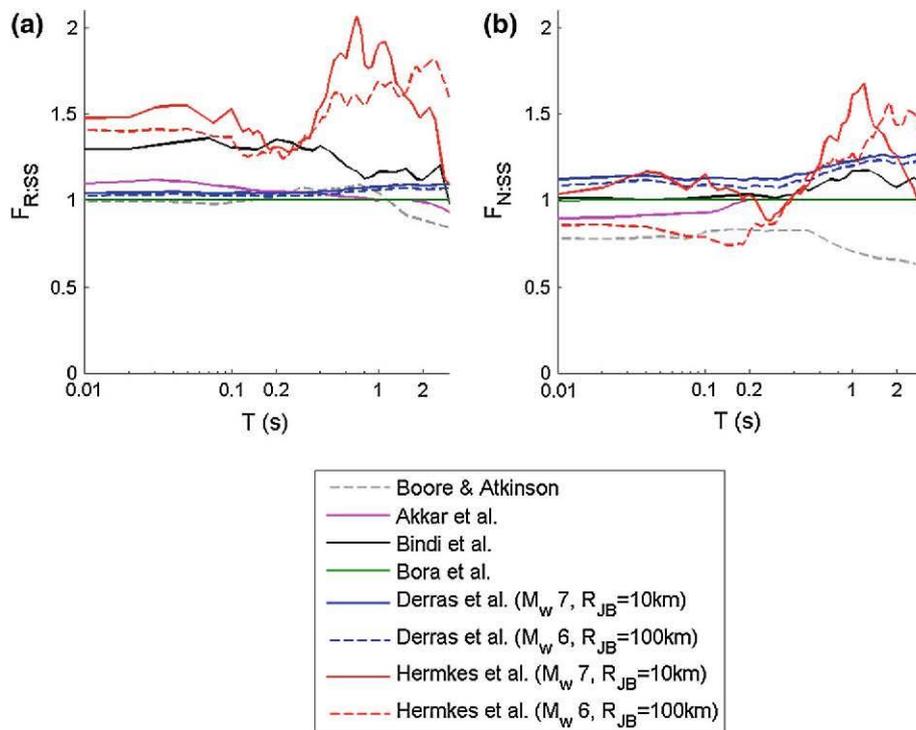


Fig. 4 Comparison of style-of-faulting factors for SA: **a** ratio of reverse to strike-slip ($F_{R:SS}$) and **b** ratio of normal to strike-slip ($F_{N:SS}$). Ratios are scenario-independent except for those of [Hermkes et al. \(2013\)](#). The predictions of [Bora et al. \(2013\)](#) are independent of the style of faulting. $F_{N:SS}$ of Akkar et al. equals unity for $T > 0.2$ s and therefore this curve is under that of [Bora et al. \(2013\)](#) for these periods

(2013) find ratios that depend weakly on these variables. These ratios are generally quite close to unity (i.e. rupture mechanism has no effect on spectral accelerations) but two models ([Bindi et al. 2013](#); [Hermkes et al. 2013](#)) show large values for $F_{R:SS} (> 1.25)$, particularly those of [Hermkes et al. \(2013\)](#), whose ratios reach over two. $F_{N:SS}$ are generally within 0.1 of unity except, again, for [Hermkes et al. \(2013\)](#) at moderate and long periods where the ratios reach 1.5. The overall observation that the style of faulting has a limited impact on spectral accelerations is in line with the findings from previous studies, including those associated with the NGA models. The usual order of which style of faulting leads to the highest and lowest motions is reversed in the model of [Derras et al. \(2013\)](#), which predicts that normal-faulting events cause higher SAs than reverse-faulting earthquakes. One possible reason for this is that only 93 of the 1,088 records used to derive this model are from reverse-faulting events (compared to 540 from normal and 455 from strike-slip earthquakes) and, in addition, each earthquake is only associated with on average 3.4 records (Table 1) and hence the style-of-faulting factors are poorly constrained. In view of this, the style-of-faulting factors implied by the model of [Derras et al. \(2013\)](#) are not recommended for application. Compared with the NGA database, RESORCE is much richer in data from normal-faulting earthquakes, e.g. less than 3% of the records used by [Boore and Atkinson \(2008\)](#) come from normal events, and consequently the estimates of $F_{N:SS}$ from the RESORCE models are much better constrained.

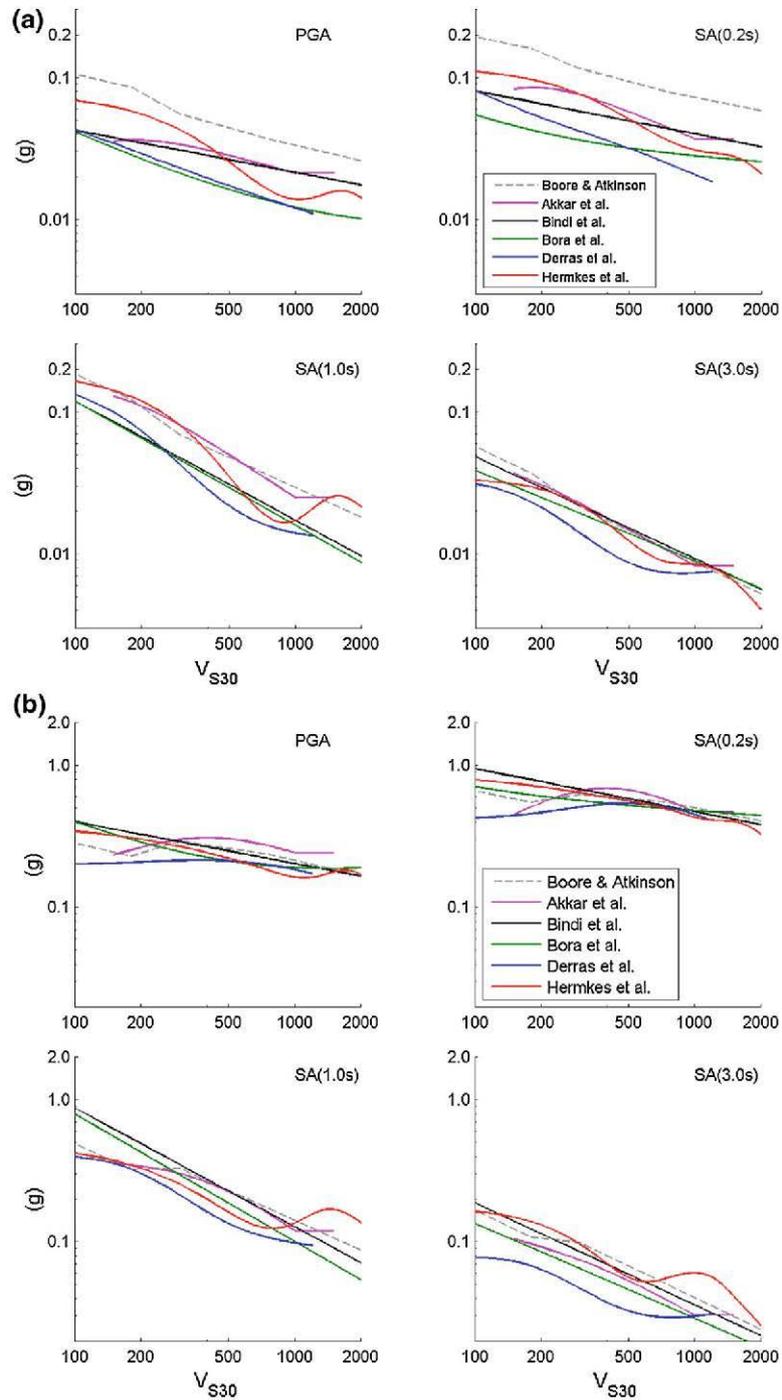


Fig. 5 Comparison of V_{S30} scaling of the median ground motion for M_w 7 strike-slip earthquakes for PGA (*top left*), SA(0.2s) (*top right*), SA(1.0s) (*bottom left*) and SA(3.0s) (*bottom right*) at: **a** $R_{JB} = 100$ km and **b** $R_{JB} = 10$ km

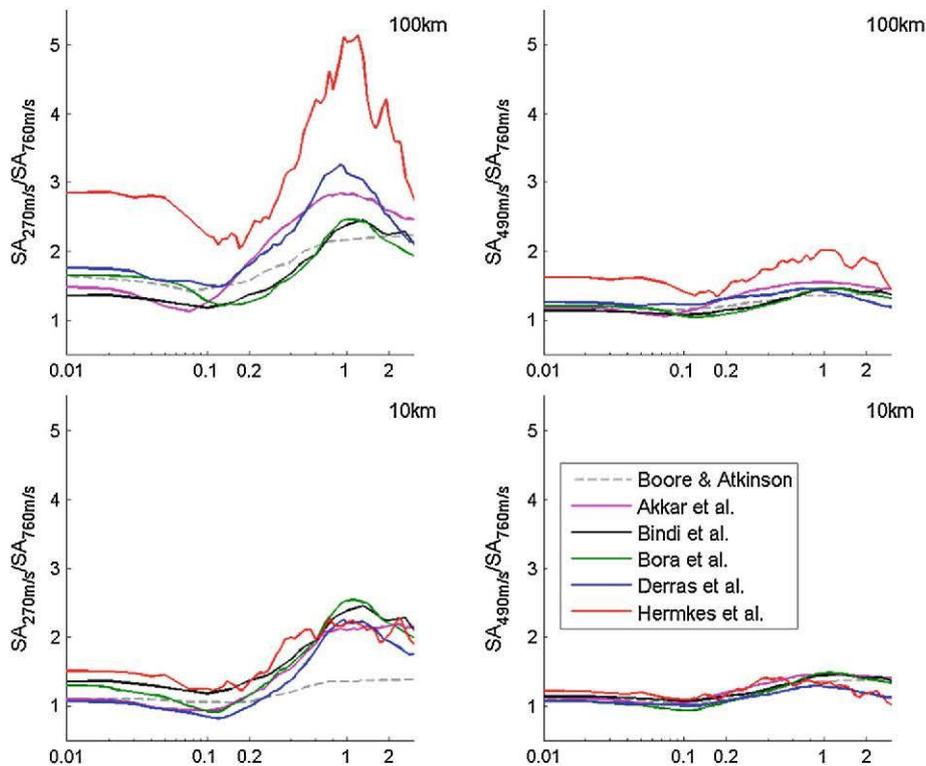


Fig. 6 Comparison of ratios between SA for $V_{S30} = 270$ m/s (NEHRP D) (*left*) and SA for $V_{S30} = 490$ m/s (NEHRP C) (*right*) to SA for $V_{S30} = 760$ m/s for $M_w 7$ (strike-slip) at $R_{JB} = 100$ km (*top*) and $M_w 7$ (strike-slip) at $R_{JB} = 10$ km (*bottom*)

6 Scaling with V_{S30}

All models (Figure 5) predict an overall inverse dependence on V_{S30} , i.e. as V_{S30} increases ground motions decrease, even if no functional form was imposed. In addition, the models predict a stronger dependence on V_{S30} for longer structural periods (Figs. 5, 6). All of the models except those of Bindi et al. (2013) and Bora et al. (2013) include nonlinear site behaviour, i.e. lower amplifications on soft soils (low V_{S30}) for stronger shaking (Figs. 5, 6). However, once again the dispersion in the predictions is quite large, particularly at longer periods.

The ratios of spectral accelerations on soft soil to rock reach their peak for a structural period of around 1s with ratios of three or even higher (up to about 5.5 for Hermkes et al. (2013) (Fig. 6), although they show considerable variation among models. Similarly the peak in the stiff-soil-to-rock ratios is at about 1s but the peak ratios are lower (around 1.5) and show smaller dispersion. These ratios are similar to those represented in a similar plot (their Figure 10) by Ambraseys et al. (2005). One difference with the NGA models, however, is that the peak amplification occurs in the NGA models at a longer period (>3 s) [see, e.g., the curves for Boore and Atkinson (2008) in Fig. 6], which could be related to soil profiles that are deeper on average in California than in Europe and the Middle East (Stewart et al. 2012) or to smaller sedimentary basins in Europe compared to California that give rise to

2D-3D basin effects at shorter periods. Also the long-period site amplifications predicted by the [Boore and Atkinson \(2008\)](#) model are generally lower than those predicted by the RESORCE models.

7 Predicted response spectra

The models all predict similar response spectra on NEHRP B/C boundary sites for M_w 5–7 at $R_{JB}=10$ km (Fig. 7); any differences in the models become apparent at large magnitudes, longer distances and for softer sites (see, e.g., Figs. 1, 2). For the largest events, the functional forms used to develop the models of [Akkar et al. \(2013a,b\)](#), [Bindi et al. \(2013\)](#), [Bora et al. \(2013\)](#) and [Hermkes et al. \(2013\)](#) allow evaluation up to M_w 8 whereas the model of ([Derras et al. 2013](#)) should not be used for such magnitudes. The periods of the plateaus in the spectra do not show strong magnitude dependency. Predictions from the GMPEs of [Boore and Atkinson \(2008\)](#) at this distance for all magnitudes fall roughly in the middle of the

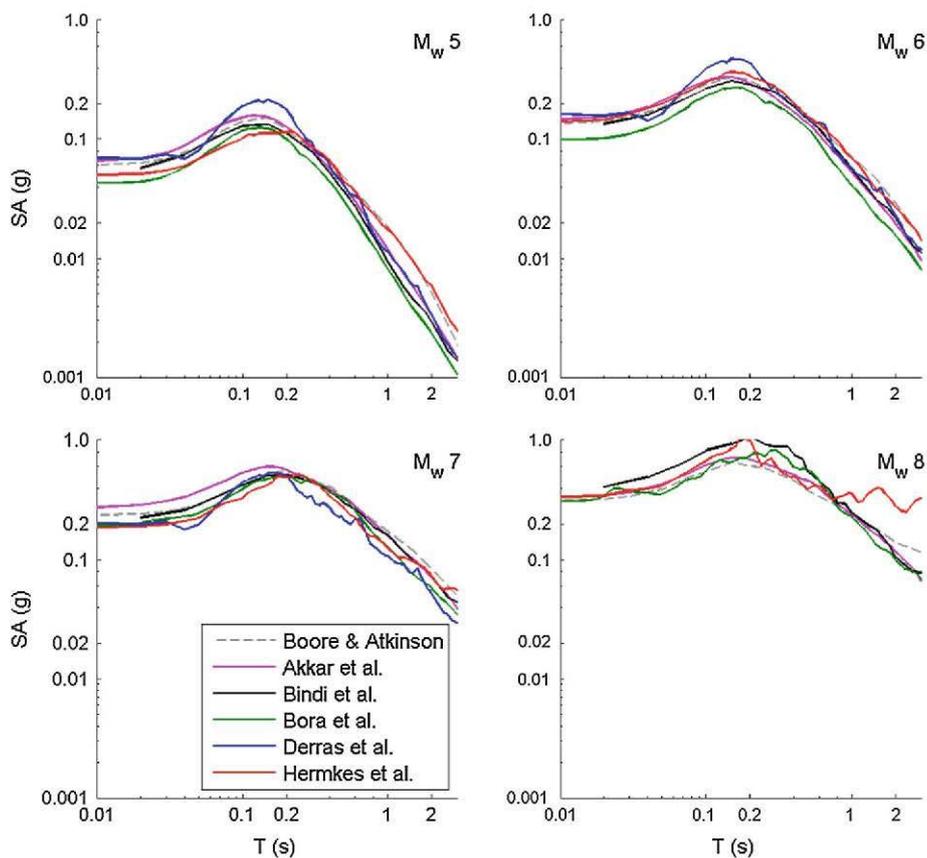


Fig. 7 Comparison of median 5% damped spectra for strike-slip earthquakes and $V_{S30} = 760$ m/s (NEHRP B/C boundary) at $R_{JB} = 10$ km for M_w 5 (*top left*), 6 (*top right*), 7 (*bottom left*) and 8 (*bottom right*). The predictions from the model of [Derras et al. \(2013\)](#) are not shown for M_w 8 since this is outside its range of applicability. The other models are shown for this magnitude even though some developers do not recommend their application for such large events

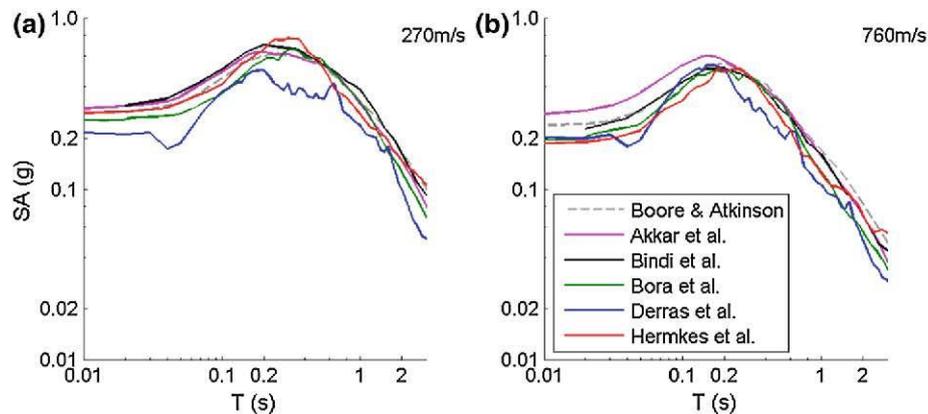


Fig. 8 Comparison of median 5% damped spectra for strike-slip earthquakes at $R_{JB} = 10$ km for $M_w 7$ and a) $V_{S30} = 270$ m/s (NEHRP D) and b) $V_{S30} = 760$ m/s (NEHRP B/C boundary)

predictions from the RESORCE models but because of the lower attenuation predicted by this model the predicted spectra for longer distances are higher than those predicted by the RESORCE models (not shown here).

The predicted spectra for soft soil sites show a much broader plateau and greater dispersion than in the predicted spectra on rock (Fig. 8), which is due to the strong long-period site amplifications predicted by some models (Fig. 6). Again a factor of two in the predicted spectral accelerations can be seen between the highest and lowest predictions.

8 Aleatory variability

As noted above, all models predict homoscedastic aleatory variability (standard deviation, sigma) and consequently only a single figure is required to summarise this aspect of the models (Fig. 9). As Akkar et al. (2013a,b) note there is limited data from larger earthquakes and consequently the apparent magnitude dependency seen within their between-event residuals may not represent the true aleatory variability at large magnitudes. Consequently they assumed magnitude-independent sigmas. Similar arguments hold for the other models. The sigmas fall into two groups: Bora et al. (2013), which has slightly higher values, and the other four models. This difference is related to higher values of the between-event (τ) standard deviations whilst the within-event (ϕ) standard deviations are similar. The sigmas show similar dependence on period with a first peak between 0.1 and 0.2s (near the plateau of predicted response spectra) and then a further increase in sigma as period increases. However, the period dependency is quite limited with less than a 20% difference between the lowest and the highest sigma.

The values of τ for the models of Akkar et al. (2013a,b), Derras et al. (2013) and Hermkes et al. (2013) are similar to those of the NGA models although slightly higher [see, e.g., the curve for Boore and Atkinson (2008) shown on Fig. 9], whereas the τ s of Bindi et al. (2013) and Bora et al. (2013) are larger. The values of ϕ of the different models are slightly (by about 0.1 ln units for moderate magnitudes) higher than those of the NGA models [again, see the curve for Boore and Atkinson (2008) on Fig. 9], which leads to overall sigmas that are also about 0.1 ln units higher. The NGA models of aleatory variability also do not show a strong period dependence. The higher estimates of aleatory

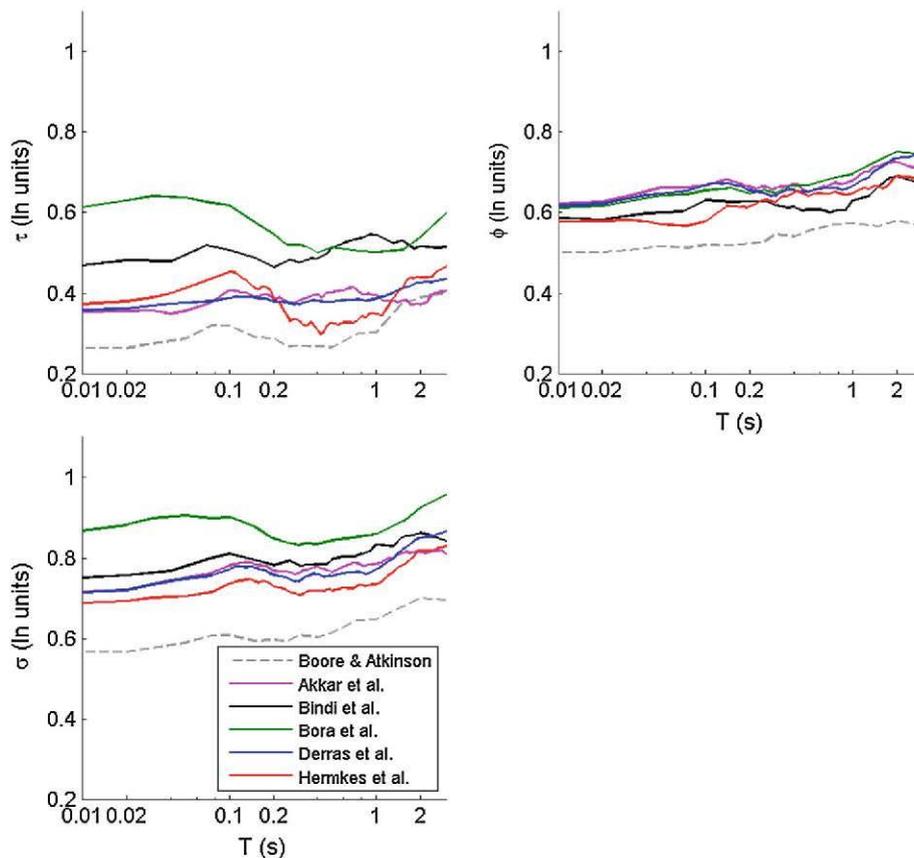


Fig. 9 Comparison of the between-event (τ), within-event (ϕ) and total (σ) standard deviations. All models have homoscedastic standard deviations

variability for the RESORCE models compared with the sigmas of the NGA models could be related to: (a) truly higher variability in ground-motion databases in Europe and the Middle East (caused by, e.g., mixing together of data from a wide geographical region with different tectonics and geology); (b) the use of more data from small earthquakes whose motions are possibly intrinsically more variable than those from large events because of, e.g., higher variability in stress drops; or (c) problems with the metadata in RESORCE, particularly for small events (or more likely a mixture of these reasons). (Insufficiently complex functional forms for the RESORCE models cannot explain this difference because it is apparent even for the non-parametric models). One aspect of the metadata that could be revisited in future models for Europe and the Middle East, particularly for applications below M_w 5.5, is the use of moment magnitude (sometimes obtained by conversions from other magnitude scales) rather than local magnitude (M_L) for the smaller earthquakes. It was shown by [Bindi et al. \(2007\)](#), for north-western Turkey, that the use of M_L for small earthquakes leads to lower estimates of between-event variability (τ) compared to using M_w . This is because corner frequencies for such earthquakes are generally higher than 1 Hz, which is the frequency range at which M_L is measured whereas M_w is measuring energy at frequencies below the corner and hence it is a poorer measure of the size of such events.

Therefore, it could be envisaged that M_L is used below, say, M_w 5.5 for the derivation of GMPEs and then in applications the local magnitude scale for that region is used to evaluate the model. Sabetta and Pugliese (1987) adopted a similar composite magnitude scale (M_L below M 5.5 and surface-wave magnitude, M_s , above this limit) when deriving their GMPEs for Italy.

9 Conclusions

In this article, various aspects of the five ground-motion models that are described in other articles in this special issue have been compared. Despite all the developers having started with the same common strong-motion archive and having used the same independent parameters, the predicted spectral accelerations from the models show significant differences, which can be related to varying data selection criteria and derivation techniques. All aspects of the models for the median ground motions (magnitude scaling, style-of-faulting factors, distance decay and site amplification) show variation from one model to the next. These differences when combined lead to variations in the predicted response spectral accelerations for scenarios of interest of more than a factor of two. These differences demonstrate that epistemic uncertainty in ground-motion prediction in Europe and the Middle East remains large and it cannot be explained by differences in the metadata of the strong-motion records used or different sets of independent parameters (e.g. hypocentral distance rather than Joyner-Boore distance or surface-wave magnitude rather than moment magnitude). One of the reasons for this large epistemic uncertainty is that a given earthquake in Europe and the Middle East is, on average, recorded by fewer strong-motion instruments than in California, Taiwan and Japan and hence the aspects of the models related to source effects are less well constrained.

The aleatory variabilities are slightly higher than those associated with the NGA models, again (e.g. Strasser et al. 2009) showing that this aspect of ground-motion modelling is stable within a narrow band (± 0.2 ln units) around 0.7 (for PGA). In particular, estimates of the within-event variability (ϕ) show little variation from one study to the next. The between-event variability (τ), however, can be significantly affected by the inclusion of data from smaller (less well-studied) earthquakes. Further studies to constrain the value of τ for European events are, therefore, recommended.

The five models presented in this volume should be of considerable value for seismic hazard assessments in Europe and the Middle East, providing both state-of-the-art predictions of spectral accelerations and a basis for quantifying epistemic uncertainty in those predictions.

Acknowledgments The work presented in this article was partially funded by the SHARE (Seismic Hazard Harmonization in Europe) Project funded under contract 226967 of the EC-Research Framework Programme FP7 and by the task 'Reference database for seismic ground motion in Europe' of the SIGMA (Seismic Ground Motion Assessment) project. We thank the personnel of the organisations operating seismological stations and freely disseminating their ground-motion data and related metadata, without which the derivation of the ground-motion models compared here would have been impossible. Finally, we thank an anonymous reviewer for their constructive comments on a previous version of this article.

References

- Abrahamson NA, Youngs RR (1992) A stable algorithm for regression analyses using the random effects model. *Bull Seismol Soc Am* 82(1):505–510
- Abrahamson N, Atkinson G, Boore D, Bozorgnia Y, Campbell K, Chiou B, Idriss IM, Silva W, Youngs R (2008) Comparisons of the NGA ground-motion relations. *Earthq Spect* 24(1):45–66. doi:10.1193/1.2924363

- Akkar S, Sandikkaya MA, Bommer JJ (2013a) Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull Earthq Eng* doi:10.1007/s10518-013-9461-4
- Akkar S, Sandikkaya MA, Bommer JJ (2013b) Erratum: Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bull Earthq Eng*. doi:10.1007/s10518-013-9508-6
- Akkar S, Sandikkaya MA, Şenyurt M, Azari AS, Ay BÖ, Traversa P, Douglas J, Cotton F, Luzi L, Hernandez B, Godey S (2013c) Reference database for seismic ground-motion in Europe (RESORCE). *Bull Earthq Eng*. doi:10.1007/s10518-013-9506-8
- Ambraseys NN, Douglas J, Sarma SK, Smit PM (2005) Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration. *Bull Earthq Eng* 3(1):1–53. doi:10.1007/s10518-005-0183-0
- Bindi D, Massa M, Luzi L, Ameri G, Pacor F, Puglia R, Augliera P (2013) Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-Damped PSA at spectral periods up to 3.0 s using the RESORCE dataset. *Bull Earthq Eng* (submitted to this issue)
- Bindi D, Parolai S, Grosser H, Milkereit C, Durukal E (2007) Empirical ground-motion prediction equations for northwestern Turkey using the aftershocks of the 1999 Kocaeli earthquake. *Geophy Res Lett* 34(L08305)
- Bommer JJ, Douglas J, Strasser FO (2003) Style-of-faulting in ground-motion prediction equations. *Bull Earthq Eng* 1(2):171–203
- Bommer JJ, Scherbaum F, Bungum H, Cotton F, Sabetta F, Abrahamson NA (2005) On the use of logic trees for ground-motion prediction equations in seismic-hazard analysis. *Bull Seismol Soc Am* 95(2):377–389. doi:10.1785/0120040073
- Boore DM, Atkinson GM (2008) Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01s and 10.0s. *Earthq Spect* 24(1):99–138
- Bora SS, Scherbaum F, Kuehn N, Stafford PJ (2013) Fourier spectral- and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions. *Bull Earthq Eng*. doi:10.1007/s10518-013-9482-z
- Building Seismic Safety Council (2009) 2009 NEHRP recommended seismic provisions For new buildings and other structures: part 1, provisions, Federal Emergency Management Agency (P-750), Washington
- Castellaro S, Mulargia F, Rossi PL (2008) V_{S30} : Proxy for seismic amplification? *Seismol Res Lett* 79(4):540–543. doi:10.1785/gssrl.79.4.540
- Chiou B, Darragh R, Gregor N, Silva W (2008) NGA project strong-motion database. *Earthq Spect* 24(1):23–44. doi:10.1193/1.2894831
- Derras B, Cotton F, Bard P-Y (2013) Towards fully data driven ground-motion prediction models for Europe. *Bull Earthq Eng*. doi:10.1007/s10518-013-9481-0
- Douglas J (2010) Consistency of ground-motion predictions from the past four decades. *Bull Earthq Eng* 8(6):1515–1526. doi:10.1007/s10518-010-9195-5
- Douglas J, Halldórsson B (2010), On the use of aftershocks when deriving ground-motion prediction equations, 9th US National and 10th Canadian conference on earthquake engineering: reaching Beyond Borders, paper ID 220.
- Douglas J, Jousset P (2011) Modeling the difference in ground-motion magnitude-scaling in small and large earthquakes. *Seismol Res Lett* 82(4):504–508. doi:10.1785/gssrl.82.4.504
- Hermkes M, Kuehn NM, Riggelsen C (2013) Simultaneous quantification of epistemic and aleatory uncertainty in GMPEs using Gaussian process regression. *Bull Earthq Eng* (this issue)
- Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ, Rukstales KS (2008) Documentation for the 2008 Update of the United States National Seismic Hazard Maps, Open-File Report 2008–1128, U.S. Department of the Interior, U.S. Geological Survey, 61 pp.
- Power M, Chiou B, Abrahamson B, Bozorgnia Y, Shantz T, Roblee C (2008) An overview of the NGA project. *Earthq Spect* 24(1):3–21. doi:10.1193/1.2894833
- Sabetta F, Pugliese A (1987) Attenuation of peak horizontal acceleration and velocity from Italian strong-motion records. *Bull Seismol Soc Am* 77(5):1491–1513
- Scasserra G, Stewart JP, Bazzurro P, Lanzo G, Mollaioli F (2009) A comparison of NGA ground-motion prediction equations to Italian data. *Bull Seismol Soc Am* 99(5):2961–2978
- Schmedes J, Archuleta RJ (2008) Near-source ground motion along strike-slip faults: Insights into magnitude saturation of PGV and PGA. *Bull Seismol Soc Am* 98(5):2278–2290. doi:10.1785/0120070209
- Stafford PJ, Strasser FO, Bommer JJ (2008) An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region. *Bull Earthq Eng* 6(2):149–177

Annexe 3 : Supplément électronique de Derras et al. (2014)

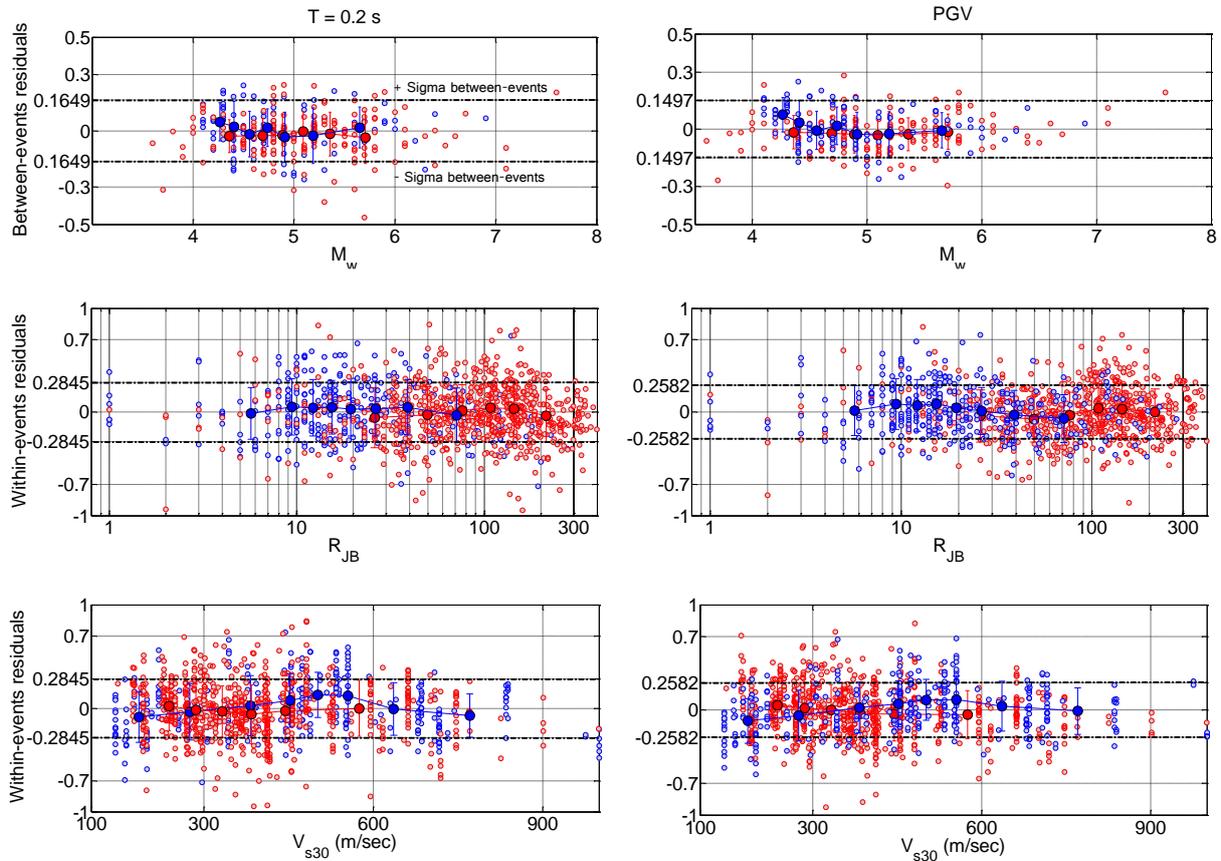


Figure.A.3 Distributions of the between-event and within-event residuals for PSA(0.2 sec) and PGV according to M_w , R_{JB} and V_{s30} . In all graph we present the dependence of between-event and within-event residuals from Italian (blue) and Turkish (red) data. We illustrate also the Mean residuals - standard deviations in the different metadata ranges. The circles show average residuals, the vertical bars are the symmetrical error bars for two standard deviations; and the horizontal bars represent 1 standard deviation. The dashed lines show the \pm sigma between (or within) -event for the corresponding ground motion parameter

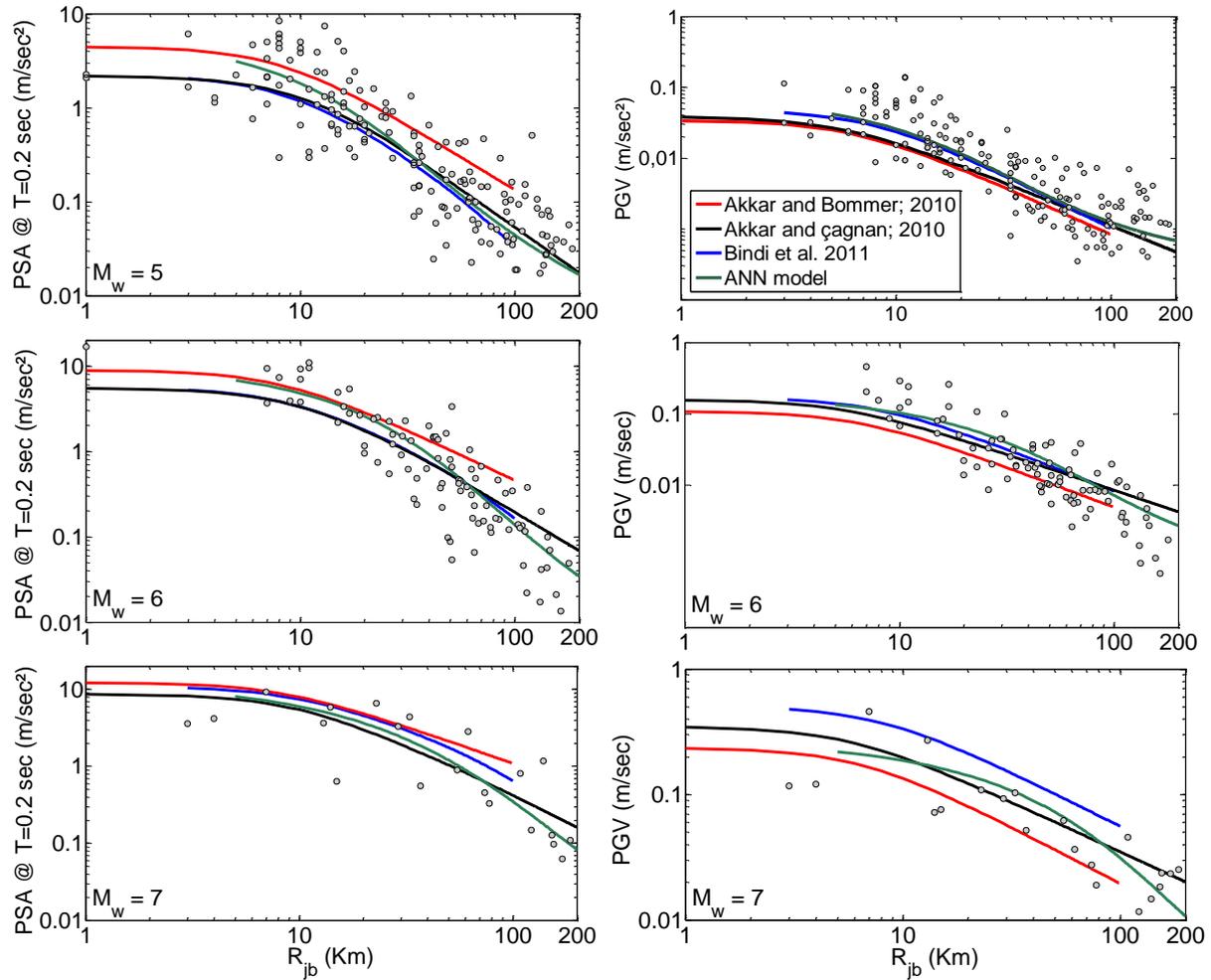


Figure.A. 4 Comparison of predicted PSA at 2 sec and PGV by the present model and by other studies with recorded PSA and PGV from earthquakes with magnitude equal $M_w \pm 0.5$ and $V_{s30} = 500 \pm 100$ m/sec. $M_w = 5$ (top), $M_w = 6$ (middle) and $M_w = 7$ (bottom). With the use of $V_{s30} = 500$ m/sec, depth = 10 km and normal faulting.

Résumé

Cette thèse est consacrée à une analyse approfondie de la capacité des "réseaux de neurones artificiels" (RNA) à la prédiction des mouvements sismiques. Un premier volet important concerne la dérivation par RNA de "GMPE" (équations de prédiction du mouvement du sol) et la comparaison des performances ainsi obtenues avec celles des GMPE "classiques" obtenues sur la base de régressions empiriques avec une forme fonctionnelle préétablie (plus ou moins complexe). Pour effectuer l'étude comparative et obtenir les deux composantes inter-événements « between-event » et intra-événements « within-event » de la variabilité aléatoire, nous intégrons l'algorithme du « modèle à effets aléatoires » à l'approche neuronale. Cette approche est testée sur différents jeux de données réelles et synthétiques : la base de données compilée à partir d'événements européens, méditerranéens et du Moyen-Orient (RESORCE : Reference database for Seismic grOund-motion pRediction in Europe), la base de données NGA-West 2 (Next Generation Attenuation West 2 développée aux USA), la base de données japonaise dérivée du réseau accélérométrique KiK-net. En outre, un set de données synthétiques provenant d'une approche par simulation stochastique est utilisé. Les paramètres du mouvement du sol les plus utilisés en génie parasismique (PGA, PGV, spectres de réponse et également, dans certains cas, les fonctions d'amplification locales) sont considérés. Les modèles neuronaux ainsi obtenus, complètement dirigés par les données « data-driven », nous renseignent sur les influences respectives et éventuellement couplées de l'atténuation avec la distance, de l'effet d'échelle lié à la magnitude, des conditions de site et notamment la présence éventuelle de non-linéarités. Un autre volet important est consacré à l'utilisation des RNA pour tester la pertinence de différents proxies de site, au travers de leur capacité à réduire la variabilité aléatoire des prédictions de mouvement du sol. Utilisés individuellement ou en couple, ces proxies de site décrivent de manière plus ou moins détaillée l'influence des conditions de site locales sur le mouvement sismique. Dans ce même volet, nous amorçons également une étude des liens entre les aspects non-linéaire de la réponse de site, et les différents proxies de site. Le troisième volet se concentre sur certains effets liés à la source : analyse de l'influence du style de la faille sismique sur le mouvement du sol, ainsi qu'une approche indirecte de la dépendance entre la magnitude et la chute de contrainte sismique.

Abstract

This thesis is devoted to an in-depth analysis of the ability of "Artificial Neural Networks" (ANN) to achieve reliable ground motion predictions. A first important aspect concerns the derivation of "GMPE" (Ground Motion Prediction Equations) with an ANN approach, and the comparison of their performance with those of "classical" GMPEs derived on the basis of empirical regressions with pre-established, more or less complex, functional forms. To perform such a comparison involving the two "between event" and "within-event" components of the random variability, we adapt the algorithm of the "random effects model" to the neural approach. This approach is tested on various, real and synthetic, datasets: the database compiled from European, Mediterranean and Middle Eastern events (RESORCE: Reference database for Seismic grOund-motion pRediction in Europe), the database NGA West 2 (Next Generation Attenuation West 2 developed in the USA), and the Japanese database derived from the KiK-net accelerometer network. In addition, a comprehensive set of synthetic data is also derived with a stochastic simulation approach. The considered ground motion parameters are those which are most used in earthquake engineering (PGA, PGV, response spectra and also, in some cases, local amplification functions). Such completely "data-driven" neural models, inform us about the respective, and possibly coupled, influences of the amplitude decay with distance, the magnitude scaling effects, and the site conditions, with a particular focus on the detection of non-linearities in site response. Another important aspect is the use of ANNs to test the relevance of different site proxies, through their ability to reduce the random variability of ground motion predictions. The ANN approach allows to use such site proxies either individually or combined, and to investigate their respective impact on the various characteristics of ground motion. The same section also includes an investigation on the links between the non-linear aspects of the site response and the different site proxies. Finally, the third section focuses on a few source-related effects: analysis of the influence of the "style of faulting" on ground motion, and, indirectly, the dependence between magnitude and seismic stress drop.