Comparison of soil non-linearity (in situ stress-strain relation and *G/G$_{max}$* reduction) observed in strong-motion databases and modelled in ground motion prediction equations

Philippe Guéguen[1], Fabian Bonilla[2], John Douglas[3]

1 ISTerre, Université Grenoble Alpes, CNRS/IFSTTAR – France

2 IFSTTAR, Université Paris Est – France

3 University de Strathclyde, Glasgow – UK

Accepted for publication in the BSSA as a Short Note

Abstract

Earthquake ground motions are strongly affected by the upper tens of meters of the Earth's crust and consequently local site effects need to be included in any ground-motion prediction. It is increasingly common in ground motion prediction equations (GMPEs) to account for possible non-linear behavior of near-surface materials (soil). These non-linear site terms adjust observations made on soft soil sites to the ground motion expected on bedrock and hence allow these abundant soil records to be used within the regression analysis for the derivation of empirical GMPEs. These nonlinear site terms also allow rapid predictions of the expected ground motions on soil rather than requiring a site response analysis to be conducted. In this study we compare the signature on observed peak ground acceleration as a function of a strain proxy of non-linear soil behavior within four large strong-motion databases to the predicted signature from four recent GMPEs, three of which explicitly include non-linear site terms. We find that observed non-linearity in the databases, interpreted in terms of strain-stress relationships and reduction of shear modulus, is limited but that even this limited effect is underestimated by the non-linear site terms of the considered GMPEs, which suggests that predictions from these GMPEs could be biased for soft soil sites but also on bedrock. Some of this mismatch could be explained by the use of the average shear-wave velocity in the top 30m (V_{s30}) to characterize sites as well as errors in these values.

Introduction

Ground motion prediction equations (GMPEs) for active crustal regions are generally developed based on regression analysis of databases of observed strong ground motions (e.g. Douglas and Edwards, 2016). GMPEs are used in seismic hazard analysis to specify the level of ground motion expected given variables such as magnitude, distance and basic local site information. Site effects and their non-linear response are usually considered to be a key element of seismic hazard analysis. It is widely accepted that non-consolidated sediments tend to behave in a non-linear manner (e.g., Field et al., 1997; Bonilla et al., 2005). The non-linear response of superficial soil layers is characterized by a reduction in the high-frequency amplification, related to an increase of damping, and the shifting of the resonance frequency to lower frequencies, due to a reduction of the shear modulus, G (e.g., Assimaki et al., 2008; Bonilla et al., 2005; Régnier et al., 2013).

Terms associated with non-linear response of soils have recently been introduced into GMPEs (e.g., Abrahamson et al., 2014; Boore et al., 2014; Akkar et al., 2014). Uncertainties related to site effects make a significant contribution to the total uncertainties of these equations, and therefore to seismic hazard studies (Bommer and Abrahamson, 2006; Rodriguez-Marek et al., 2011). In particular, the use of *in situ* geophysical surveys to characterize the elastic properties and laboratory tests to assess the non-linear behavior parameters may result in estimation bias, affecting the GMPEs (e.g., Cabas et al., 2017). This bias may be due to differences between *in situ* and laboratory conditions, the presence of superficial layers with a significant effect (Régnier et al., 2013) or even three-dimensional geometric effects that cannot be replicated in the laboratory (e.g., Frankel et al., 2002; Assimaki et al., 2008; Sleep, 2010). In addition, the strong-motion data affected by strong soil non-linearity appeared to be insufficient in the international databases for completely empirical non-linear soil terms, which demands the use of modelling to develop such terms (e.g., Akkar et al., 2014, Zhao et al., 2015).

Thanks to recent efforts to install dense strong-motion networks and characterize local site conditions at these stations, it is now possible to interpret non-linearity *in situ* by analyzing the recorded data. The variation of G has thus been obtained from borehole data (Frankel, 1999) by measuring the velocity variation as a function of shear deformation by intercorrelation (e.g., Rubinstein and Beroza, 2005) and by seismic interferometry (Sawazaki et al., 2009; Chandra et al 2015, 2016; Guéguen, 2016). This shear strain can be calculated using a deformation proxy linking the medium's shear-wave velocity V_s to the maximum particle velocity, which is generally equivalent to the peak ground velocity, PGV as PGV/Vs (Rathje et al. 2004; Idriss, 2011). Furthermore, the peak ground acceleration (PGA) at the top of the soil column is a proxy of shear stress and the PGA *versus* PGV/V_s, and even PGA *versus* PGV/V_{s30} relationships can be associated with a stress-strain curve, i.e. an *in situ* test comparable with laboratory tests to reproduce non-linear effects (Chandra et al., 2015, 2016).

The purpose of this study is, therefore, to characterize the non-linear parameters, interpreted in terms of strain-stress relationships and reduction of shear modulus, using the international databases from which the GMPEs are derived. These parameters will be presented in the first part. In the second part, we will present the data used in this study, taken from four international databases. A final section presents a comparison of the data interpreted as strain-stress relationships, with the non-linear soil terms present in a selection of GMPEs.

***In situ* stress and strain proxies**

In the linear elastic domain, the relationship between shear strain and stress is directly proportional to G_{max}, i.e.

$$\tau = G_{max}\,\gamma \qquad (1)$$

where τ is the shear stress, γ is the shear deformation and G_{max} is the elastic shear modulus, i.e. the value under slight deformation. In the nonlinear domain, soil behavior is traditionally modelled by the following hyperbolic nonlinear model (Ishihara, 1996):

$$\tau = \frac{G_{max}\,\gamma}{1+\gamma/\gamma_r} \qquad (2)$$

with γ_r being the reference deformation and defined as G_{max}/τ_{max}, where τ_{max} is the maximum strength of the material. Assuming the propagation of a unidirectional wave in an infinitely uniformly elastic medium, the shear strain is considered according to the following equation (Newmark, 1968):

$$\gamma = V_{max}/\beta \qquad (3)$$

where V_{max} is the maximum particle velocity and β the apparent velocity of the shear waves, i.e., $\beta = \sqrt{G_{max}/\rho}$ where ρ is density. Considering maximum horizontal acceleration proportional to the shear stress (τ) and the strain according to Eq. 3, Chandra et al. (2015, 2016) used data from a vertical array to evaluate the variation of β (and therefore of G) according to the strain calculated between two sensors during seismic loading. They thus derived an *in situ* model of the nonlinear behavior of soil based on an interpretation of the experimental data in terms of strain-stress values, equivalent to the hyperbolic model (Eq. 2). On the basis of Eq. 3, Idriss (2011) suggested considering PGV/V_{s30} as the average strain over the first 30m, where nonlinearity is mainly expected to occur, PGV being comparable to V_{max} in Eq.3. Finally, by considering the shear stress proportional to acceleration and PGA (i.e., $\tau = PGA \times h \times \rho$, with h the equivalent depth), Chandra et al. (2016) confirmed the possibility of distinguishing the average behavior of different site classes (classed according to V_{s30}) according to a strain proxy, i.e. $PGA = f(PGV/V_{s30})$, using data from the Japanese networks KNET and KiK-net.

From Eq. 1, we thus obtain the *in situ* stress-strain relationship under elastic deformation, as follows:

$$PGA = G_{max}.PGV/V_{s30}/h/\rho \tag{4}$$

i.e., the maximum shear stress proxy is proportional to PGA and the shear strain proxy to PGV/V_{s30}, i.e. G is proportional to $PGA/(PGV/V_{s30})$. We can then obtain an experimental *in situ* curve characterizing the nonlinear behavior by the reduction of modulus G according to the following equation:

$$\frac{G}{G_{max}} = \frac{PGA}{PGV/V_{s30}} \Big/ \left(\frac{PGA}{PGV/V_{s30}}\right)_{max} \tag{5}$$

where $\left(\frac{PGA}{PGV/V_{s30}}\right)_{max}$ is computed for $PGV/V_{s30} < 10^{-5}\%$ corresponding to the linear elastic deformation limit (Vucetic, 1994; Johnson and Jia, 2005). Using *in situ* data, we can then explore the nonlinearity in strong-motion databases, evaluated using the shear strain proxy (Eq. 3) and the shear modulus reduction (Eq. 5). In our case, the nonlinearity is associated with the reduction of modulus *G*, and this reduction can be predicted or calculated using GMPEs (Eq. 5).

Database description

Four databases were used to test the nonlinear parameters in the data, only taking into account the parameters required for Eq. 5 as well as earthquake magnitude: the intensity measures considered were *PGA* and *PGV* and the site parameter was the V_{s30}. Data processing and information describing the source of these data are described in the original papers and the flat files.

- NGA-West2 flat file provided by *Pacific Earthquake Engineering Research Center* (Ancheta et al., 2014). The file contains 21,540 ground motion records, recorded during shallow crustal earthquakes in active tectonic regions worldwide. Two types of V_{s30} values are distinguished in this database, estimated or direct measurement, which will be discussed later.

- K-NET and KiK-net Japanese network databases (Okada et al., 2004), characterized by two different types of installations (Aoi et al., 2004). One of the advantages of the Japanese networks is the homogeneity of the metadata, characterizing the earthquakes (e.g. magnitude and locations) and the local site conditions. For K-NET, measurements were taken up to a depth of 20m, and V_{s30} was then estimated using KiK-net velocity surveys that go deeper (Boore et al., 2011). For KiK-net, V_{s30} was calculated directly from velocity profiles going from 100 up to 2008 m. For this study, K-NET records having a PGA larger than 10 cm/s^2 were collected between 1996 and end of 2016, irrespective of distance or magnitude. We use the KiK-net data processed by Regnier et al. (2013), consisting the records between 1996 and 2009, with magnitudes higher than 3 and a hypocentral depths and epicentral distances less than 150 km. We also added data from the mainshock and aftershocks of the M_w 9.0 Tohoku 2011 earthquake. Finally, we completed this database with records having PGAs larger than 100 cm/s^2 up to the end of 2016. Data processing is described in Régnier et al. (2013) and Laurendeau et al. (2013). We used a total of 178,556 records from KiK-net and 26,895 from K-NET.

- ESM (Engineering Strong-Motion) database (Luzi et al., 2016) containing data from the European networks was the final source of data. ESM was developed as part of the European NERA project, and was designed to provide end users with data from moderate and strong earthquakes in the European and Mediterranean region. The data has been quality-checked and uniformly processed, and relevant parameters, from 1969 to the present day. The 2017 flat-file was produced for the EPOS project and provided directly by the ESM facility. The ESM flat-file contains a total of 3,434 records.

The distribution of the data used in this study according to PGA and strain proxy (Eq. 4) is shown in Fig. 1 for the different databases separately. The ESM and NGA-West2 databases contain the lowest strain proxies. The ranges of *PGA* and PGV/V_{s30} are broader for NGA-West2, with many values above 1 m/s^2 for PGA and 0.1% for the strain proxy. A large number of strain proxies are below 10^{-4} %, i.e. below the linear cyclic deformation threshold (Vucetic, 1994) determined from laboratory tests. Between the linear (10^{-4} %) and volumetric (10^{-2} %) strain thresholds (Vucetic, 1994), the soil displays nonlinear elastic behavior with negligible permanent deformation. Above 10^{-2} %, the soil shows hysteretic nonlinear behavior with permanent deformation. Considering the data from the four databases, some must therefore contain nonlinear processes according to the soil models based on laboratory tests. The best-fit (linear) equations are similar, with similar slopes for three databases (ESM, KiK-Net and NGA-West2). It is also interesting to note that the coefficient of correlation R^2 for these three databases are quite high (>0.5) and suggest that these three databases are comparable and will all reproduce the equivalent strain-stress relationships. For the KNET data, however, R^2 is quite low (0.139) in the log-log representation, which suggests a poor prediction of the data by simple linear regression, suggesting an additional physical reason that we speculate is the presence of soil nonlinearities in the data. The nonlinearities in KNET data was also reported by Chandra et al. (2016), when comparing KiK-Net and KNET, who conclude that the KNET data shows the highest nonlinearity. Assuming higher nonlinearity in the KNET data, it is interesting to observe that these nonlinearities are for lower values of PGA, suggesting the inefficiency of PGA for the prediction of soil nonlinearity.

Nonlinear characterization

The nonlinearity observed in the databases selected for this study was assessed using Eq. 5. In our case, $\left(PGA/\frac{PGV}{V_{s30}}\right)_{max}$ was considered as the average of the values lower than 10^{-5} %. For KNET and the lowest values of V_{s30}, because of more data showing nonlinearity, the smallest values already contain nonlinearities and the plot is biased at low strain proxies. Since nonlinearity is highly dependent on site conditions, we separated the data according to V_{s30}, into three categories: [100-300]m/s, [400-600]m/s and [800-1,200]m/s, i.e. the approximate conditions C, B and A of Eurocode 8, respectively, in order to compare soils with *a priori* different nonlinear behavior. Fig. 2 shows the nonlinearity representation for the four databases. Means and standard deviations are indicated for different strain range values using a logarithmic scale (in %): [$<10^{-6}$], [10^{-6} - 3.5 10^{-6}], [3.5 10^{-6} - 1.2 10^{-5}], [1.2 10^{-5} - 4.3 10^{-5}], [4.3 10^{-5} - 1.5 10^{-4}], [1.5 10^{-4} - 5.3 10^{-4}], [5.3 10^{-4} - 1.9 10^{-3}], [1.9 10^{-3} - 6.6 10^{-3}], [6.6 10^{-3} - 2.3 10^{-2}], [2.3 10^{-2} – 8.1 10^{-2}], [8.1 10^{-2} - 2.8 10^{-1}] and [2.8 10^{-1} - 1].

We can see that, in spite of the large amount of data from various sources, nonlinearity characterized by the G/G_{max} reduction, is barely visible compared to the 95% and 90% reduction values of G/G_{max} in Figures 2a and 2b. This raises questions on whether nonlinearity can be incorporated empirically into GMPEs, particularly as its first effect is to reduce ground motion by increasing the energy dissipation. We also observe a slight dependency on magnitude. As expected, the decrease of G/G_{max} is greatest for the largest magnitudes, but for a given strain proxy, the range of magnitude values and G/G_{max} values is broad, regardless of V_{s30}.

For the lowest V_{s30} values, nonlinearity characterized by the variation in G/G_{max} makes a significant appearance at a strain proxy threshold of approximately 5 10^{-4} % for the Japanese data and 10^{-3} % in ESM and NGA-West2, with reduction of G/G_{max} larger than 90%. For the intermediary V_{s30} values, nonlinearity appears at around 10^{-3} % while for the highest V_{s30} values, a G reduction is visible from 5 10^{-3} % to 10^{-2} % for Japanese and other databases, respectively, i.e. nonlinear effects may also appear in stiff soils. However, care must be taken when classifying sites on the basis of V_{s30}, as certain recent studies have demonstrated visible nonlinear effects for sites with a V_{s30} greater than 800m/s but with a thin surface layer sensitive to nonlinearity (Bonilla et al., 2011; Régnier et al., 2013).

Integrating non-linearity in the GMPEs

Several GMPEs include site terms accounting for soil nonlinearity. We selected four recent GMPEs: Akkar et al. (2014), Boore et al., (2014) and Abrahamson et al. (2014) that include nonlinear site terms and Bindi et al. (2014) as reference with linear site terms. These four GMPEs provide predictions of *PGA* and *PGV* as a function of magnitude and source-to-site distance, and for different site conditions. Fig. 3 shows PGA predictions as a function of the strain proxy PGV/V_{s30}. The predictions are for magnitudes between 4 and 8 (0.5 intervals) and 50 distances logarithmically spaced between 0.1 and 300km. This unusual manner of representing ground-motion predictions as a function of strain proxy, enables visualization of how nonlinearity, interpreted as the reduction of *G* with respect to the strain proxy, is integrated in the GMPEs. It should be noted that the regression analysis used to derive each GMPE was conducted independently for PGA and PGV with different nonlinear site terms assumed for each. Also many scenarios where large PGAs and PGVs occur (M>7 and R<20km), and consequently there is a high chance of soil nonlinearity, are poorly sampled in the strong-motion databases, especially at soft soil sites. Therefore, the predictions from the GMPEs are more uncertain for these scenarios and depend strongly on the functional form adopted by the GMPE developer rather than being strongly constrained by the data.

As expected, nonlinearity is more present for soft soils (V_{s30}=100m/s) than for stiff soils (V_{s30}=1,000m/s), with an equivalent stress (i.e., *PGA*) - strain (i.e. PGV/V_{s30}) relationship that changes as deformation increases. The differences with the linear Bindi et al. (2014) GMPE are larger for soft site conditions.

We observe that the three GMPEs of Akkar et al. (2014), Boore et al. (2014) and Abrahamson et al. (2014) integrate nonlinearity differently and that using a GMPE with a linear site term (e.g. Bindi et al., 2014) may introduce a significant bias in terms of stress-strain proxies compared to previous prediction models for soft soils (V_{s30}=100 m/s or 200m/s). The curvature of the prediction increasing with strain proxy characterizes the nonlinearity accounted for by the GMPEs. Abrahamson et al. (2014) characterizes nonlinearity more strongly for soft (V_{s30}=100m/s) and intermediate soils (V_{s30}=200m/s) than Boore et al. (2014) and Akkar et al. (2014). The differences between nonlinear models challenges the way in which GMPEs consider nonlinear effects, leading to *PGAs* that are significantly different for the same magnitude-distance pairs. For example, compared to Bindi et al. (2014), the curvature of Boore et al. (2014) and Akkar et al. (2014) is not significant and the nonlinearity is considered as reducing the ground motion for equivalent strain values. These models principally use results taken from numerical modelling. This dispersion shows the high epistemic uncertainty in predicted ground motions, considering strain proxy and the G reduction, for soft soils undergoing high deformations. Nonlinear site terms in GMPEs are often introduced so that observed ground motions from soft soil sites can be reliably used, by removal of the site effects, to derive models to assess ground motions on bedrock. The ground motions implied by seismic hazard assessment using these GMPEs evaluated for bedrock conditions are subsequently used to select rock strong-motion records for input to site response analysis.

Figure 4 compares the predictions of the proxies of G/G_{max} according to the strain proxies from the four GMPEs considered in this study with the average values taken from the databases for three site classes (the class V_{s30}<100m/s is not considered because of insufficient data).

Several observations can be made from Fig. 4.

(1) Firstly, as expected, nonlinearity is more obvious for soils 100<V_{s30}<300m/s, characterized by larger shear strain proxy, which confirms that the *PGA* versus PGV/V_{s30} relationship is an efficient proxy to characterize the nonlinear *in situ* behavior of soil. These proxies and the scattering of the relationships could be reduced by integrating the occurrence time of the maximum values of acceleration, velocity and displacement (deformation) that may not occur at the same time, as suggested by Chandra et al. (2016) and Guéguen (2016).

(2) Based on the G/G_{max} reduction factor (with respect to the 95 and 90% thresholds), the ESM data seems to indicate less nonlinearity for soft soils than the other databases, particularly compared with the Japanese sites, which we know show clear nonlinear behavior (Régnier et al., 2013). However, this observation could be modulated according to the larger dispersion of the ESM data. The KNET stations show more marked nonlinear behavior than the KiK-net stations, which is in agreement with observations already reported by other authors (Aoi et al., 2004; Chandra et al., 2016) who concluded on a more pronounced nonlinear behavior for KNET than for KiK-net data, as a consequence of soil profiles beneath their stations.

(3) Compared with soil behavior based on laboratory tests and characterized by a traditional $G-\gamma$ curve, it appears that nonlinearity is limited, in spite of the large spread of data in international databases in terms of magnitude and distance, with the modulus G/G_{max} reduction only reaching 30% in the worst case for the sites most sensitive to nonlinearity (100<V_{s30}<300m/s). This observation suggests that nonlinearity effects are rare in the global databases used herein. Since the databases used in this study represent a significant proportion of strong-motion data ever recorded, this observation makes us wonder whether large (>1 %) strains could be expected during earthquakes. It also raises the question of considering nonlinearity, using modelling techniques or laboratory results, to define seismic demand, since site response may be underestimated compared to the observation. Perus and Fajfar (2014) proposed site factors between ground motions on sites characterized by low

V_{s30} and those on rock sites that overestimate the nonlinear effects in the predicting ground motion. In Perus and Fajfar (2014) few data are used and their conclusions are based on predicted values of PGA or Sa, and consequently, they recommend a careful consideration of their results since strong-motion data enable a better consideration of nonlinearity. In our study, and based on the variation of G/G_{max}, we observe a small effect of nonlinearities in predictions and observations, even for V_{s30}<300m/s, in contrast to Perus and Fajfar (2014). Chandra et al (2016) also suggested the limited effect of nonlinearities in the Japanese databases, with average accelerations on soil sites comparable to rock sites values, even for PGA>0.2g. In our study, even if GMPEs underestimate the nonlinear effects, they are very comparable to the nonlinearity contained in the database.

(4) The G/G_{max} versus PGV/V_{s30} relationship is comparable for the four databases, independently of the magnitude-distance relationship. Using the terminology of Luco and Cornell (2007) for an intensity measure of ground motion, this proxy is “efficient” for nonlinearity characterization. Fig. 5 shows shear strain proxy as a function of magnitude. Fig. 5 reveals that magnitude does not control the appearance of nonlinearity, if the latter is characterized according to shear strain proxy, confirming the representativeness of magnitude-distance criteria for predicting nonlinearity.

(5) The predictions of soil nonlinearity from the GMPEs are similar overall: underestimating the G reduction compared with the data. They are generally based on simulation techniques and do not represent the nonlinearity that can be observed in the databases. It is also interesting to observe that Bindi et al (2014) shows an increase in the curvature for the largest strains. This suggests that this GMPE implicitly includes some soil nonlinearity in its predictions, due to the underlying data, despite using linear site terms, or that various GMPE terms (such as those related to the site amplification) are not fully independent. This point could be confirmed by numerical simulation or more specific analysis of this database.

(6) For the 800<V_{s30}<1,200m/s class, the data from NGA-West2 display more non-linear behavior than those of the other databases. It is important to remember that certain V_{s30} values are possibly underestimated or not measured in the NGA-West2 database. Fig. 6 shows the reduction in *G* for NGA-West2, distinguishing between the sites with measured V_{s30} and the sites with estimated V_{s30}. A readjustment of the data to the GMPE predictions is observed for all site conditions, but particularly for stiff soil sites (800<V_{s30}<1,200m/s). This leads us to conclude that certain V_{s30} estimates are not correct in NGA-West2 meaning that some sites are incorrectly classified here.

This random 10% variation of the V_{s30} values for the 800<V_{s30}<1,200m/s class enables observation of the strong sensitivity of the nonlinearity to this parameter. We can, therefore, conclude that the consideration of nonlinearity requires detailed and precise characterization of site conditions, already mentioned for the prediction of ground motion, but all the more important if we intend to include nonlinear site terms in the equations.

Conclusions

In this project, we analyzed strong-motion data from four large databases worldwide. These data are often used by researchers to derive GMPEs, which are used to estimate earthquake ground motions for a given magnitude and source-to-site distance. Although these equations are useful for prediction of ground motions on rock, they are less efficient for prediction on soil, particularly for V_{s30}<300m/s, which reflects the sparsity of the data for this range of V_{s30} (Ktenidou et al., 2018). Indeed, such soils may display a nonlinear response due to their low resistance and a strong incident motion, as is the case for sites close to the seismic source or with strong amplification. To take such behavior into account, GMPEs have been modified to include the shear modulus reduction according to strain proxy. Description of the soil's nonlinear behavior used for numerical modeling is based on a few parameters, mainly obtained by laboratory tests, which do not represent the natural variability of soils and which neglect the propagation effects of seismic waves in the medium.

We found the characterization of V_{s30} to be essential to good prediction of the non-linear response. Soil nonlinearity, interpreted in terms of G reduction for given strain proxy values, exists and is stronger than that predicted by the GMPEs. However, unlike in the geotechnical models based on laboratory tests, the shear deformation observed in the international databases remains low, limited to a shear modulus reduction of around 30% for the softest soils. The comparison between geotechnical model and in-situ observation could be compared through numerical modeling in further studies. In addition, reduction in G with increasing strain proxy in stiff soils was also observed, which may be due to thin superficial layers that cause nonlinearity as already supported by Régnier et al. (2013) for Japanese data.

Acknowledgements

This project was partially funded by the French accelerometric network programme, within the framework of its 2017 Call of Proposal. The authors would like to thank the data providers enabling the compilation of the information in the flat-files. This study is part of the Urban Seismology project of the Earth Sciences Institute (Université Grenoble Alpes), and was supported by funding from Labex OSUG@2020 (Investissements d'avenir, ANR10-LABX56). We thank Fabrice Cotton and an anonymous reviewer for their careful and useful comments on an earlier version of this study.

Data and Resources

Authors used the technical ressources provided by the French Accelerometric Network database (RAP-DC, doi:10.15778/RESIF.RA http://data.datacite.org/10.15778/RESIF.RA) for processing the data. Flat-files of strong motion database were downloaded from European Strong Motion (ESM) flat-file http://esm.mi.ingv.it/flatfile-2017/flatfile.php) and Pacific Earthquake Engineering Research (PEER) ground-motion database (http://peer.berkeley.edu/ngawest2/) (last access: June 2018). K-NET and KiK-Net flat files were provided by references cited in the manuscript, using data downloaded from the strong-motion seismograph network services operated by the National Research

Institute for Earth Science and Disaster Resilience (NIED) seismological services (http://www.kyoshin.bosai.go.jp/).

References

Abrahamson, N. A., W. J. Silva, R. and Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthquake Spectra* **30(3)** 1025-1055.

Akkar, S., M. A.Sandıkkaya, and J. J. Bommer (2014). Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East, *Bulletin of Earthquake Engineering* **12(1)** 359-387.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B.S.-J. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue (2014). NGA-West2 database, *Earthquake Spectra* **30** 989– 1005.

Aoi S., T. Kunugi, and H. Fujiwara (2004). Strong-motion seismograph net-work operated by NIED: K-NET and KiK-net, *J. Jpn Assoc. Earthq. Eng*. **4(3)** 65–74.

Assimaki, D., W. Li, J. H. Steidl, and K. Tsuda (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(1**) 301-330.

Bindi, D., M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, and P. Augliera (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, *Bulletin of Earthquake Engineering* **12(1)** 391-430.

Bommer, J. J., and N. A. Abrahamson (2006). Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?, *Bull. Seism. Soc. Am.* **96(6)** 1967-1977.

Bonilla, L. F., R. J. Archuleta, and D. Lavallée (2005). Hysteretic and dilatant behavior of cohesionless soils and their effects on nonlinear site response: field data observations and modeling, *Bull. Seism. Soc. Am.* **95** 2373-2395.

Bonilla L.F., K. Tsuda, N. Pulido, J. Régnier, and A. Laurendeau (2011). Nonlinear site response evidence of K-NET and KiK-net records from the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space* **63** 785–789.

Boore, D., E. Thompson, and H. Cadet (2011). Regional correlations of VS30 and velocities averaged over depths less than and greater than 30 meters, *Bull. Seism. Soc. Am.* **101**(6) 3046–3059.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthquake Spectra* **30(3**) 1057-1085.

Cabas, A., A. Rodriguez-Marek, and L. F. Bonilla (2017). Estimation of site-specific Kappa (K0)-consistent damping values at KiK-Net sites to assess the discrepancy between laboratory-based damping models and observed attenuation (of seismic waves) in the field, *Bull. Seism. Soc. Am.* **107(5)** 2258-2271.

Chandra, J., P. Guéguen, J. H. Steidl, and L. F. Bonilla (2015). In situ assessment of the G–γ curve for characterizing the nonlinear response of soil: Application to the Garner Valley downhole array and the wildlife liquefaction array, *Bull. Seism. Soc. Am.* **105(2A)** 993-1010.

Chandra, J., P. Guéguen, P., and L. F. Bonilla (2016). PGA-PGV/Vs considered as a stress–strain proxy for predicting nonlinear soil response, *Soil Dynamics and Earthquake Engineering* **85** 146-160.

Douglas, J. and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation, *Earth-Science Reviews* **160** 203-219.

Field, E. H., P. A. Johnson, I. A. Beresnev, and Y. Zeng (1997). Nonlinear ground-motion amplification by sediments during the 1994 Northridge earthquake, *Nature* **390(6660)**, 599.

Frankel, A. D. (1999). How does the ground shake?, *Science* **283(5410)** 2032-2033.

Frankel, A. D., D. L. Carver, and R. A. Williams (2002). Nonlinear and linear site response and basin effects in Seattle for the M 6.8 Nisqually, Washington, earthquake, *Bull. Seism. Soc. Am.* **92(6)** 2090-2109.

Guéguen, P. (2016). Predicting nonlinear site response using spectral acceleration vs PGV/VS30: a case history using the Volvi-test site, *Pure and Applied Geophysics* **173(6)** 2047-2063.

Idriss, I.M. (2011). Use of Vs30 to represent local site condition, in *Proc. 4th IASPEI/ IAEE International Symposium. Effects of Source Geology on Seismic Motion*. August 23-26th, 2011. University of Santa Barbara California.

Ishihara K. (1996). *Soil behaviour in earthquake geotechnics*. Oxford Engineering Science Series. Oxford University Press.

Johnson, P. A., and X. P. Jia (2005). Nonlinear dynamics, granular media and dynamic earthquake triggering, *Nature* **437** 871–874.

Ktenidou, O. J., Z. Roumelioti, N. Abrahamson, F. Cotton, K. Pitilakis, and F. Hollender (2018). Understanding single-station ground motion variability and uncertainty (sigma): lessons learnt from EUROSEISTEST, *Bulletin of Earthquake Engineering* **16(6)** 2311-2336.

Laurendeau A., F. Cotton, O.-J. Ktenidou, L. F. Bonilla, and F. Hollender (2013). Rock and stiff-soil site amplification: dependency on Vs30 and Kappa (κ0), *Bull. Seism. Soc. Am.* **103(6)** 3131–48.

Luco, N. and C. A. Cornell (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, *Earthquake Spectra* **23(2)** 357-392.

Luzi L., R. Puglia, E. Russo, M. D'Amico, C. Felicetta, F. Pacor, G. Lanzano, U. Ceken, J. Clinton, G. Costa, L. Duni, E. Farzanegan, P. Guéguen, C. Ionescu, I. Kalogeras, H. Özener, D. Pesaresi, R. Sleeman., A. Strollo, and M. Zare (2016). The engineering strong-motion database: A platform to access pan-European accelerometric data, *Seismological Research Letters* **87(4)** 987-997.

Newmark, N.M. (1968). Problems in Wave Propagation in Soil and Rock, In *Symposium on Wave Propagation and Dynamic Properties of Earth Materials*, August 23-25, Univ. of New Mexico Press, Albuquerque, NM, 7-26.

Okada, Y., K. Kasahara, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara, and A. Yamamoto (2004). Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiK-net, *Earth, Planets and Space* **56(8)**.

Perus, I., and P. Fajfar (2014). Prediction of site factors by a non-parametric approach, *Earthquake*

Engng Struct. Dyn. **43** 1743–1761.

Rathje, E. M., W. J. Chang, K. H. , Stokoe, and B. R. Cox (2004). Evaluation of ground strain from in situ dynamic response, in *Proceeding of the 13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada. August 1 to 6, Paper No. 3099.

Régnier, J., H. Cadet, L.F. Bonilla, E. Bertand, and J. Semblat (2013). Assessing nonlinear behavior of soil in seismic site response: Statistical analysis on KiK-net strong motion data, *Bull. Seism. Soc. Am.* **103(3)** 1750-1770.

Rodriguez-Marek, A., G. A. Montalva, F. Cotton, and L. F. Bonilla (2011). Analysis of single-station standard deviation using the KiK-net data, *Bull. Seism. Soc. Am.* **101(3)** 1242-1258.

Rubinstein, J. L., and G. C. Beroza (2005). Depth constraints on nonlinear strong ground motion from the 2004 Parkfield earthquake, *Geophysical research letters* **32(14)**.

Sawazaki, K., H. Sato, H. Nakahara, and T. Nishimura (2009). Time-lapse changes of seismic velocity in the shallow ground caused by strong ground motion shock of the 2000 Western-Tottori earthquake, Japan, as revealed from coda deconvolution analysis, *Bull. Seism. Soc. Am.* **99(1)** 352-366.

Sleep, N. H. (2010). Nonlinear behavior of strong surface waves trapped in sedimentary basins, *Bull. Seism. Soc. Am.* **100(2)** 826-832.

Vucetic, M. (1994). Cyclic threshold shear strains in soils, *J. Geotech. Eng.* **120** 2208–2228

Zhao, J.X., J. Hu, F. Jiang, J. Zhou, Y. Zhang, X. An, M. Lu, and D. A. Rhoades (2015). Nonlinear site models derived from 1D analyses for ground-motion prediction equations using site class as the site parameter, *Bull. Seism. Soc. Am.* **105*(4)*** 2010–2022.

Figure

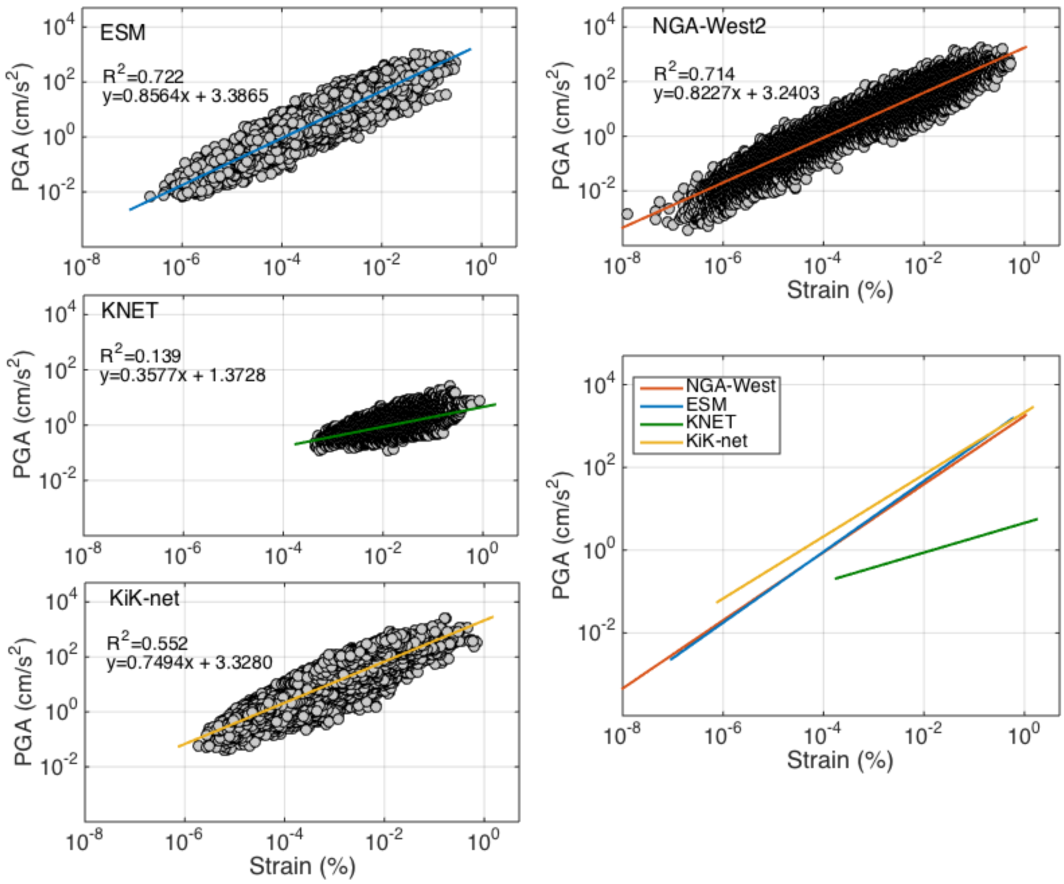
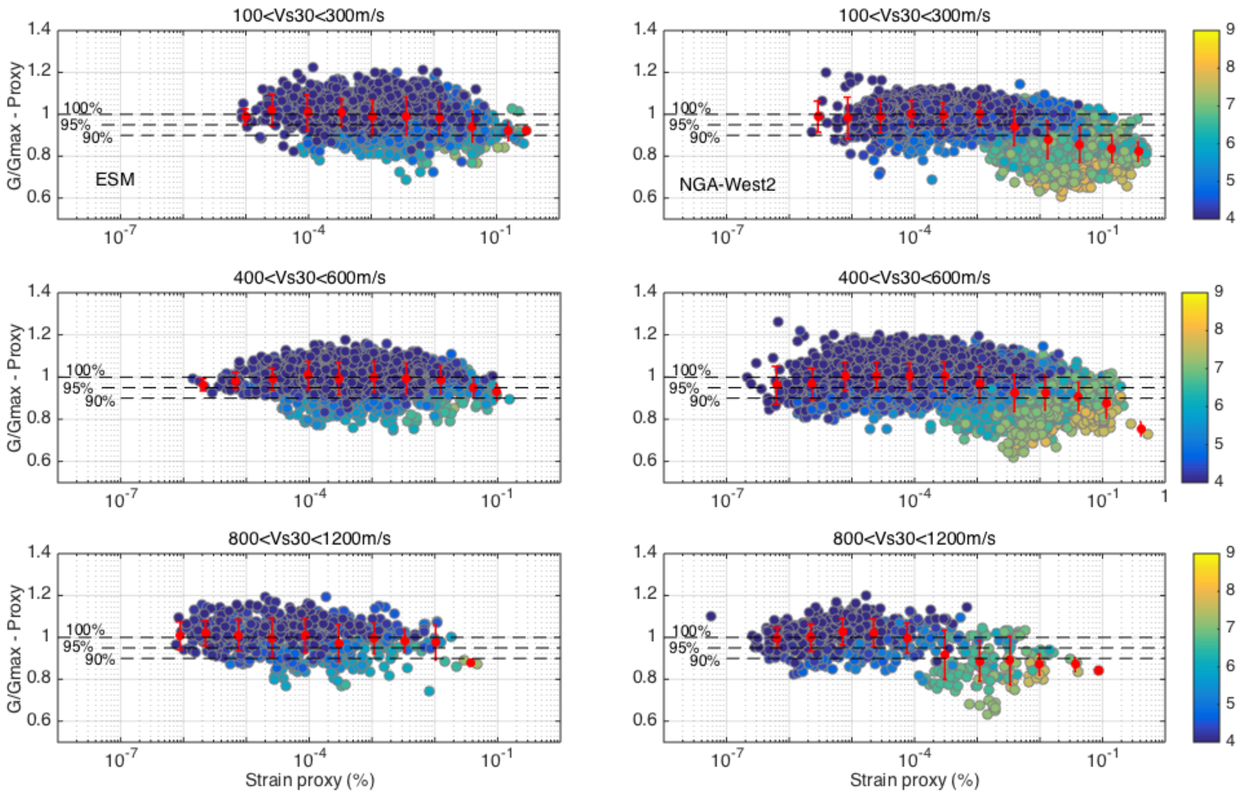


Figure 1 – Data distribution, according to the deformation estimated by PGV/V_{s30} for the four databases used. Strain proxy is given as a percentage (see text for explanations). Best-fit linear equations and coefficients of determination R^2 are given for all databases (lines).



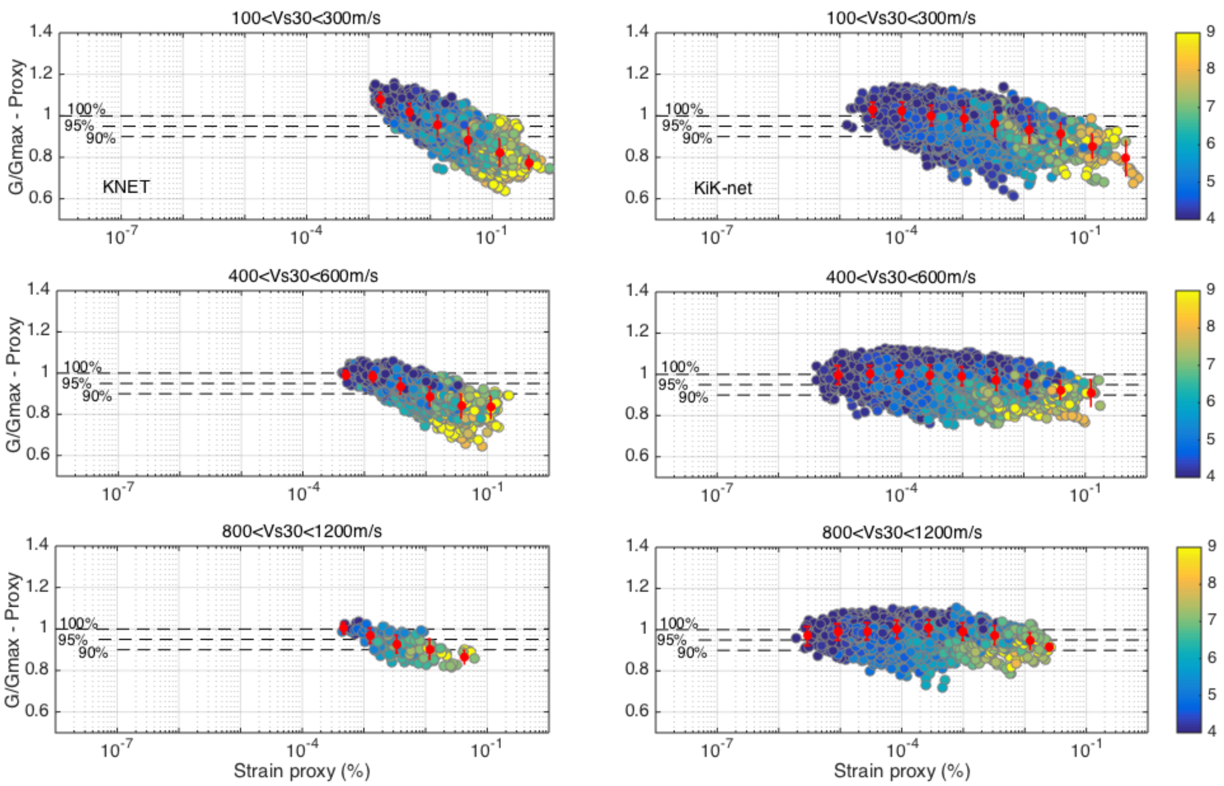


Figure 2 – Modulus *G* variation according to the strain proxy calculated by Eq. 3 and 5, for three site classes. The red symbols correspond to the average (+/- standard deviation) per strain proxy range

(see text). The color scale corresponds to magnitude. Horizontal dashed lines correspond to 100%, 95% and 90% of the values of G/G_{max}. a. ESM and NGA, b: K-NET and KiKNet.

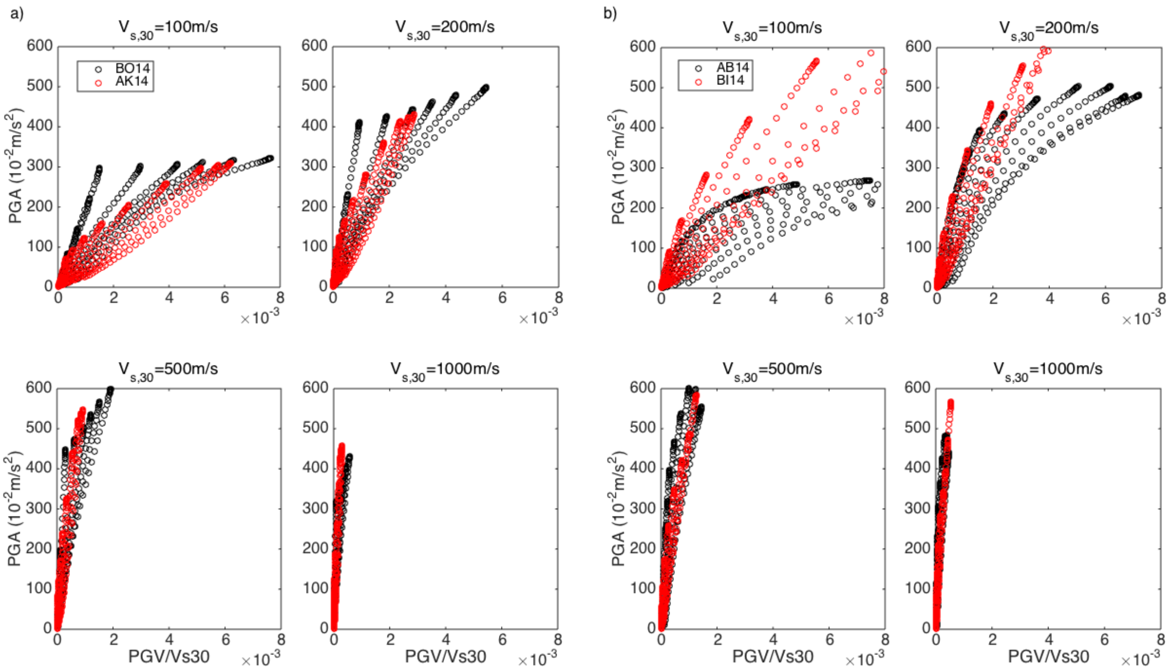


Figure 3 – Predicted *PGA* as a function of the predicted deformation. Each dot corresponds to a magnitude-distance pair for magnitudes between 4 and 8 (interval=0.5) and 50 distances between 0.1 and 300km. a) Boore et al. (2014) and Akkar et al. (2014). b) Abrahamson et al. (2014) and Bindi et al. (2014).

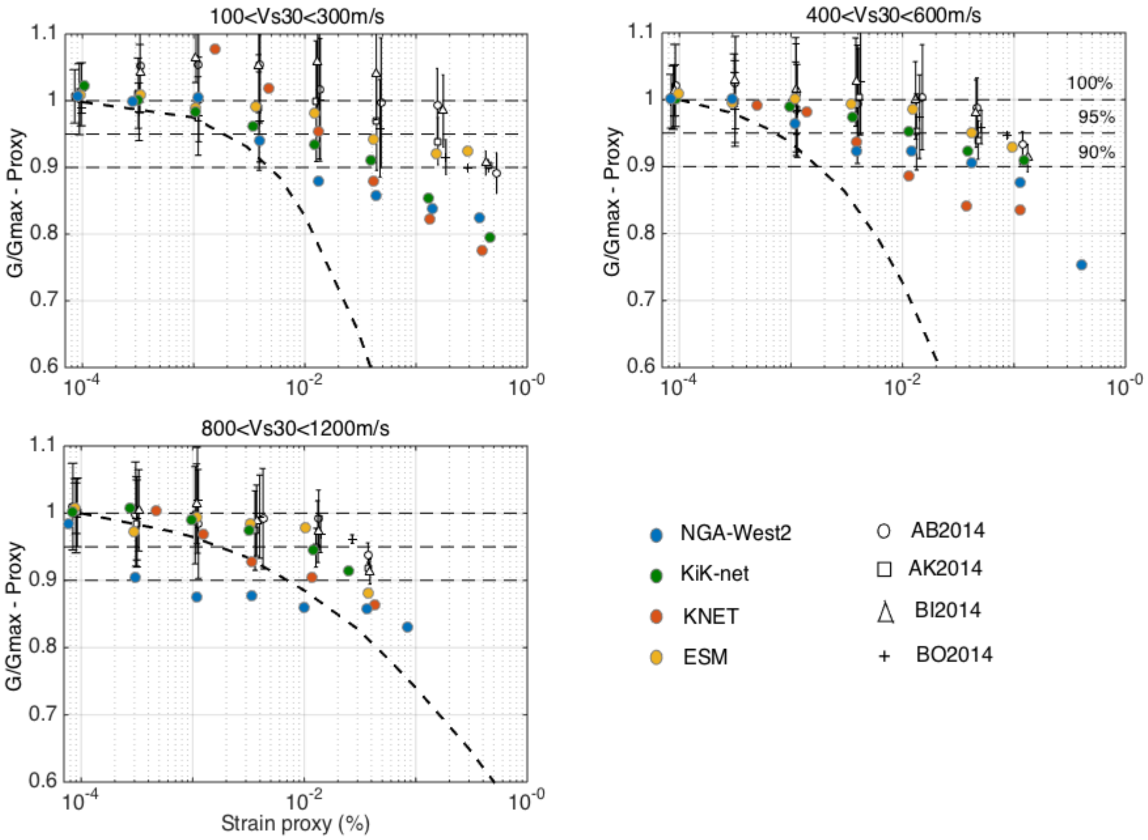


Figure 4 – Comparison of predictions of nonlinearity characterized by the proxy G/G_{max} (Eq. 5) as a function of the strain proxy (PGV/V_{s30}) according to four GMPEs (AB2014: Abrahamson et al., 2014; AK2014: Akkar et al., 2014; BI2014: Bindi et al., 2014; BO2014: Boore et al., 2014) on average values from the four databases for three site classes, for strain proxies > 10^{-4} %. Thin horizontal dashed lines correspond to 100%, 95% and 90% of the G/G_{max} values. Bold dashed lines are standard G-γ curves for clay (PI=15%), sand and rock-like soil from Zhao et al. (2015).

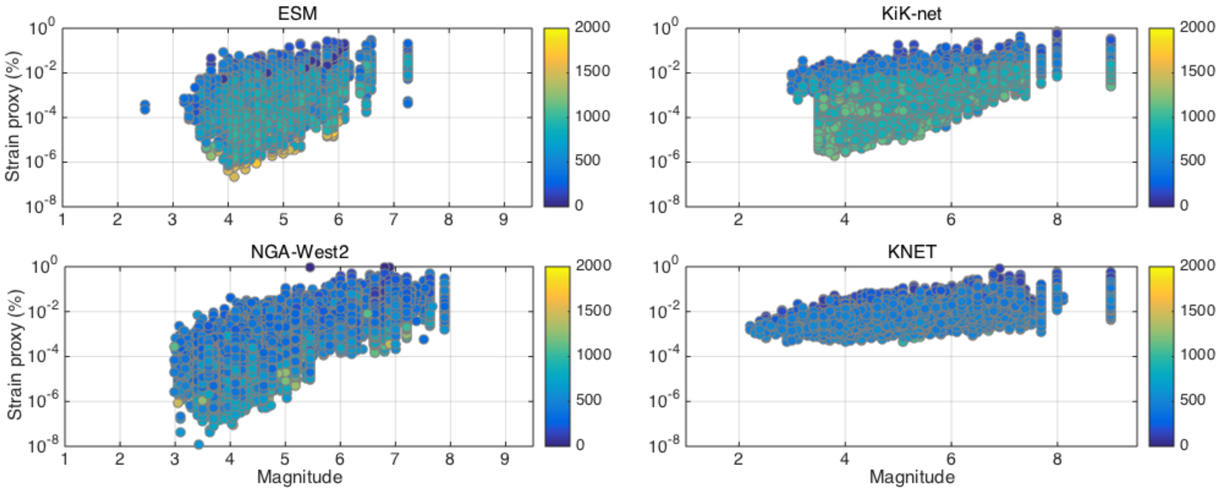


Figure 5 – Soil shear deformation (PGV/V_{s30}) as a function of earthquake magnitude. The color scale indicates V_{s30}.

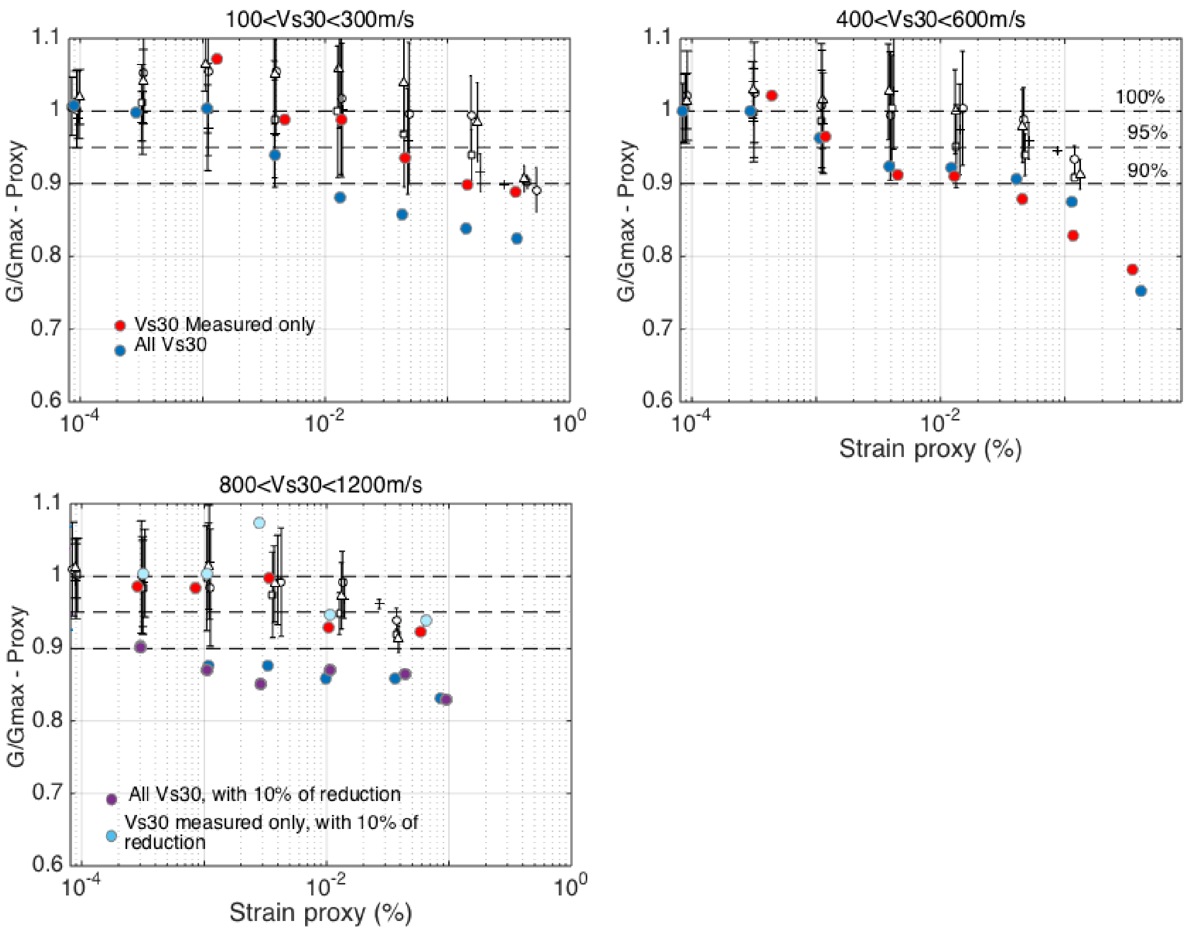


Figure 6 – Same as Fig. 4 for NGA-West2 only, distinguishing measured V_{s30} values. For $800<V_{s30}<1{,}200$m/s sites, the Vs30 values are also modified randomly by -10%.