Maria Sklodowska-Curie Actions (MSCA)
Innovative Training Networks (ITN)
H2020-MSCA-ITN-2018
Grant number 813137

Project number 813137

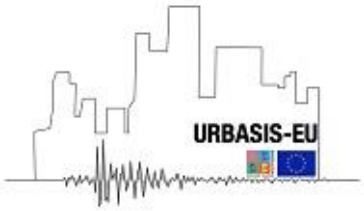
URBASIS-EU
New challenges for Urban Engineering Seismology

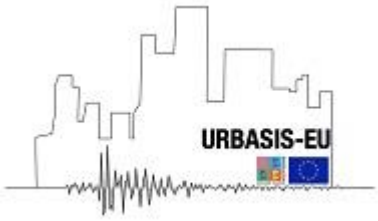
DELIVERABLE

Work Package: WP4
Number: D4.4 – Testing of Site-Amplification Models used by Ground Motion Models

Authors:	**Løviknes, Karina**	(GFZ)
Co-authors:	Cotton, Fabrice	(GFZ)
	Schorlemmer, Danijel	(GFZ)
Reviewer	Paolucci, Roberto	(POLIMI)

Approval	Management Board
Status	Final Version
Dissemination level	Public
Delivery deadline	31.10.2021
Submission date	01.02.2022
Intranet path	https://urbasis-eu.osug.fr/Scientific-Reports-157





Index

D4.4 – Testing of Site-Amplification Models used by Ground Motion Models

ABSTRACT

The testing procedure for non-linear site amplification models of Loviknes et al. (2021) are applied to two new datasets: ESM (European Engineering Strong-Motion) and NGA-West2 (Next Generation Attenuation Relationships for Western US). For the two datasets, stations from Italy and California are used, respectively. For both datasets the non-linear amplification models perform better than for the Japanese strong motion KiK-net (Kiban-Kyoshin) network tested by Loviknes et al. (2021). In particularly, several Italian stations show a downgoing trend at strong-ground motions. However, the non-linear models do not perform well with the 30m time-averaged shear-wave velocity (V_{S30}).

INTRODUCTION

The structure of near-surface geology has a strong effect on earthquake ground motions. Such variations in seismic ground motions caused by soil properties at a site are commonly referred to as site effects. For weak ground motions, the site-specific amplification is constant for any ground-motion intensity, that is, only the linear site-response is invoked. However, for large ground motions and mainly soft soils, non-linear site amplification is expected. Non-linear site effects have been shown to produce a shift of shear-wave energy towards frequencies lower than the fundamental resonance frequency of the soil column, along with a relative decrease in amplification at high frequencies (e.g. Bonilla et al., 2005; Régnier et al., 2013; Guéguen et al., 2019).

Including non-linear site effects in ground-motion prediction models (GMMs) is a challenge. 1-D numerical simulations and several observations suggest that non-linear site-amplification should decrease with increasing intensity of predicted ground motions (Bonilla et al., 2005; Field et al., 1997). However, these observations are related to a limited number of stations and earthquakes. GMMs aim to predict the probability of reaching a level of ground-motion intensity, given the properties of the earthquake source (e.g. magnitude), wave-propagation path (e.g. epicentral distance) and site (site-response proxy). Calibration of such models usually requires a high number of observations over a large range of distances, magnitudes and acceleration levels. Such large datasets have not been available until now and recent ground-motion models including non-linear site amplification components (e.g. Boore et al., 2014; Abrahamson et al., 2014) derived relying partly or fully on simulated data. One example is from Seyhan and Stewart (2014), who developed a semi empirical non-linear site-amplification model relying both on empirical observations in the ground-motion database NGA-West2 (Next Generation Attenuation Relationships for Western US, Ancheta et al., 2014) created by the Pacific Earthquake Engineering Research Center (PEER) and numerically simulated data by Kamai et al. (2014). Seyhan and Stewart (2014) analyzed non-linear site effects and calibrated their model by quantifying non-linearity as the gradient of decreasing site amplification with increasing predicted peak acceleration on rock.

Building upon the approach of Seyhan and Stewart (2014), Loviknes et al. (2021) developed a testing framework to test non-linear amplification models against observed site amplification of individual stations in the Kiban-Kyoshin (KiK-net) network in Japan. Loviknes et al. (2021) found that for most of the selected stations a simple linear amplification model scored better than the non-linear amplification models. In a similar study, Paolucci et al. (2021) also showed that although some KiK-net stations show clear signs of non-linearity, the amplification models predict a stronger de-amplification than what is observed in the data. These findings indicates that non-linear amplification models based on PGA and V_{S30} are not able to fully capture the non-linearity in KiK-net stations. In this study we will apply the testing procedure of Loviknes et al. (2021) on the European Engineering Strong-Motion (ESM) database and on Californian stations in the NGA-West 2 dataset.

DATASETS

Loviknes et al. (2021) used ground motions recorded by stations from the Japanese KiK-net network, compiled in a database by Bahrampouri et al. (2020) (BEA20), to test non-linear amplification models. In this study we will use the same testing framework on two new datasets, the ESM database and the NGA-West 2 database. We followed the same data selection criteria as in Loviknes et al. (2021) using active crustal events with hypo-central depth ≤ 35km, recorded at Joyner-Boore distance R_{JB} < 600km. A comparison of the selected records and the site distribution from the three datasets are shown in Figure 1 and Figure 2, respectively.

Engineering Strong-Motion Database

The ESM database is developed and maintained under the ORFEUS committee and supported by the EPOS (European Plate Observing System) and SERA (Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe) projects (Luzi et al. 2016, Lanzano et al., 2021). The ESM database contains uniformly processed strong-motion records from Europe and the Middle East with magnitudes in the range 3.5–8.0. Because only Italian stations recorded a sufficient number of strong-motion records (at least 4 with PGA > 0.05 g), we only use the Italian records in the ESM flatfile compiled within the ESM database. After applying the selection criteria, we obtain 12,628 records from 344 earthquakes and 1239 stations.

NGA-West 2

The ground-motion database NGA-West2 (Next Generation Attenuation Relationships for Western US) was created by the Pacific Earthquake Engineering Research Center (PEER) and is a global database with shallow crustal earthquakes in active tectonic regions (Ancheta et al., 2014). In this study we only use events recorded by stations in California, USA. After applying the selection criteria we obtain 14,125 records from 318 events and 827 stations.

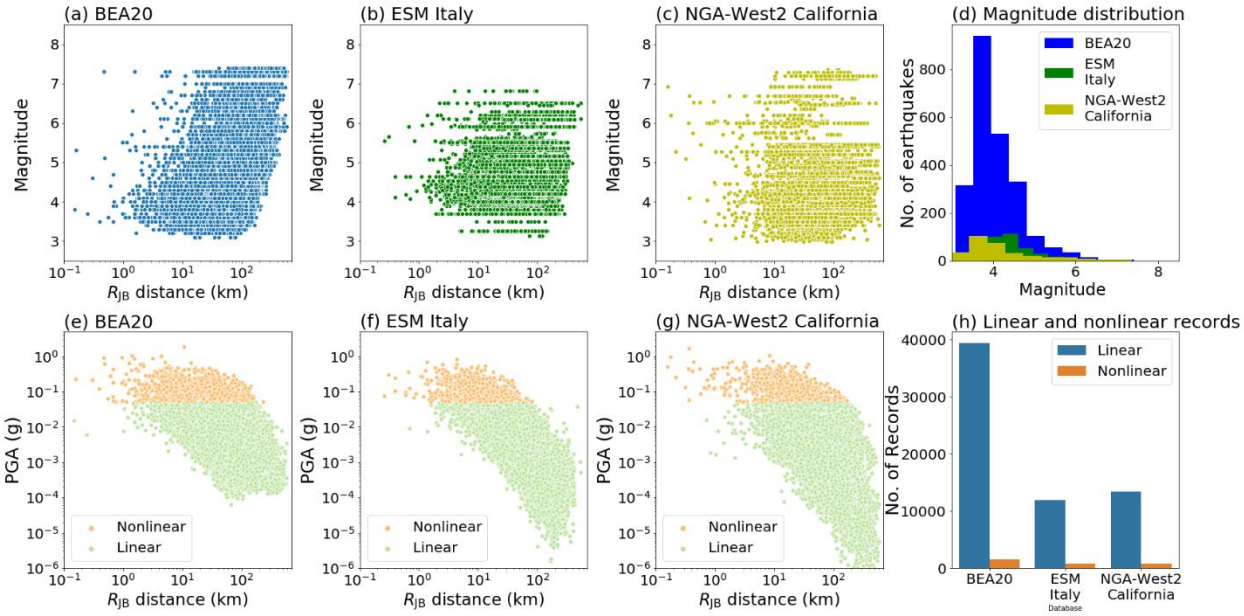


Figure 1. Data and site distribution of the selected records from the databases used in this study. Magnitude with distance (R_{JB}) distribution (a) of Bahrampouri et al. (2020) (BEA20), (b) ESM Italy and (c) NGA-West 2 California, (d) Number of Earthquakes per Magnitude for the three datatbases, Peak Ground Acceleration with distance (R_{JB}) distribution colored by whether the record is in the linear or nonlinear range (PGA > 0.05g) for (e) of Bahrampouri et al. (2020) (BEA20), (f) ESM Italy and (g) NGA-West 2 California, (h) Number of linear and nonlinear records for the three databases.

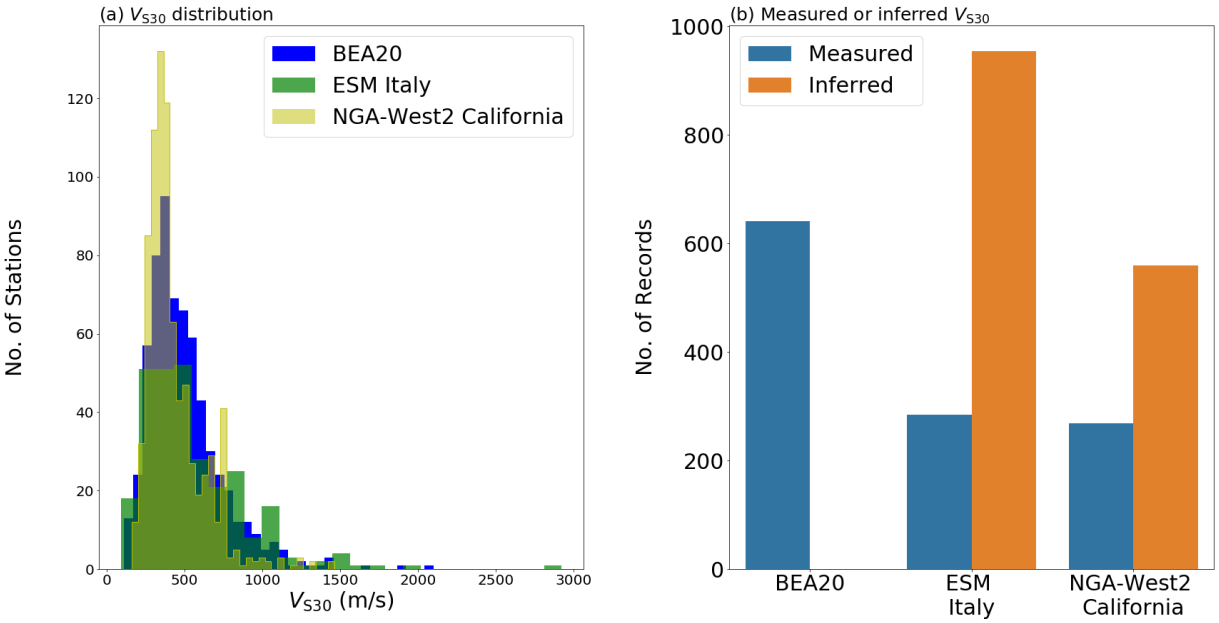


Figure 2 Site distribution of the selected records from the databases used in this study. (a) Number of stations per V_{S30} bin and (b) Number of stations with measured and inferred V_{S30} values for the three databases.

METHOD

The testing framework of Loviknes et al. (2021) consist of three parts. First a site-specific ground-motion model taking into account the linear site amplification is derived on the dataset of interest. Secondly, the residuals between the predicted linear ground motion and each observation are computed. Sites with linear behavior will then show residuals equal to zero for the full range of predicted ground motions for rock conditions. Finally, site-amplification models are tested against the residuals of individual well-recorded stations and stations grouped into site proxy bins. Each step is described in further details in the following sections.

Development of a linear ground-motion model

We use the same method and simple functional form to derive the GMMs of each dataset, this is to maintain consistency in the test and allow better comparison between the results of the different datasets. The method and functional form used by Loviknes et al. (2021) were developed by Kotha et al. (2018) The GMM of Kotha et al. (2018) differs from other contemporary models by not including any site terms. The GMM is thus only based on magnitude and distance scaling, which is captured using mixed-effects regression. A mixed effects regression model deals with hierarchical data by including both fixed-effect and random-effect terms in the regression, where fixed-effects are the explanatory variables (in this case magnitude and distance) and the random effects are the grouping factors (here event and site) (Bates et al., 2015).

The GMM was derived for the geometric mean of horizontal pseudo spectral acceleration (PSA):

$$\ln(\mathrm{PSA}) = f_{\mathrm{R}}(M_{\mathrm{W}},R_{\mathrm{JB}}) + f_{\mathrm{M}}(M_{\mathrm{W}}) + \delta \mathrm{B}_e + \delta \mathrm{S2S}_s + \delta \mathrm{WS}_{e,s} \quad (1)$$

Here, $f_{\mathrm{R}}(M_{\mathrm{W}},R_{\mathrm{JB}})$ and $f_{\mathrm{M}}(M_{\mathrm{W}})$ are the fixed effects capturing the scaling of PSAs with distance and magnitude. The between-event random effect $\delta \mathrm{B}_e$ and site-specific random effect $\delta \mathrm{S2S}_s$, quantify the event and site variability, while $\delta \mathrm{WS}_{e,s}$ is the "left-over" residual capturing the record-to-record variability. Since the GMM does not feature a fixed-effect site response based on V_{S30}, the $\delta \mathrm{S2S}_s$ captures all site-specific response and can be used as an empirical site-amplification function (Kotha et al., 2018).

We derive a linear GMM, using only records that contain linear soil response, for each dataset. The non-linear records are omitted to avoid biasing the GMM median predictions and the estimates of the $\delta \mathrm{S2S}_s$ with non-linear soil response. Loviknes et al. (2021) defined the non-linear range as recordings from soft-soil stations ($V_{\mathrm{S30}} < 760\mathrm{m/s}$) with $\mathrm{PGA}_{\mathrm{rock}} > 0.05\mathrm{g}$. However, several studies have found that V_{S30} is not an ideal proxy for non-linear site response (e.g. Loviknes et al. 2021, Derras et al., 2016; Thompson and Wald, 2016). Furthermore, in contrast to the KiK-net network, not all sites in Italy or California have measured $\mathrm{V_{S30}}$ values (Figure 2b). Instead, the V_{S30} value associated to the site is inferred using e.g. the Wald and Allen (2007) method, which has large uncertainties. In this study, we therefore omit the V_{S30} criteria and consider all strong-motion records with $\mathrm{PGA} > 0.05\mathrm{g}$ as in the non-linear range. This allow us to use records from stations with no measured V_{S30}.

It is important to point out that these definitions of nonlinear range can only tell us if a record have the potential to be nonlinear, not whether a record show nonlinear site amplification behavior. Other parameters might describe the nonlinear site response more precisely, e.g. shear strain (Gueguen et al. 2019). However, because the aim of this study is to test nonlinear site amplification models, we use PGA to define the nonlinear range because that is what is used by the models. Furthermore, the threshold 0.05 g is a somewhat arbitrary constant and nonlinear site amplification have been observed outside this range (Régnier et al., 2013). Loviknes et al. (2021) therefore performed a sensitivity test to

evaluate how the assumptions affected the result of the test. The sensitivity test showed that lowering the threshold down to PGA=0.01 g gives the linear amplification model an advantage in the test, while augmenting the threshold up to PGA=0.1 g drastically reduces the number of records available in the nonlinear range. The limitations and alternatives to the PGA threshold for defining nonlinearity is further discussed in the Discussion section.

The predicted response spectra of the three linear GMMs at 50 km R_{JB} distance for different magnitudes are shown in Figure 3a. Because the GMMs are from different databases and regions the difference between then is expected.

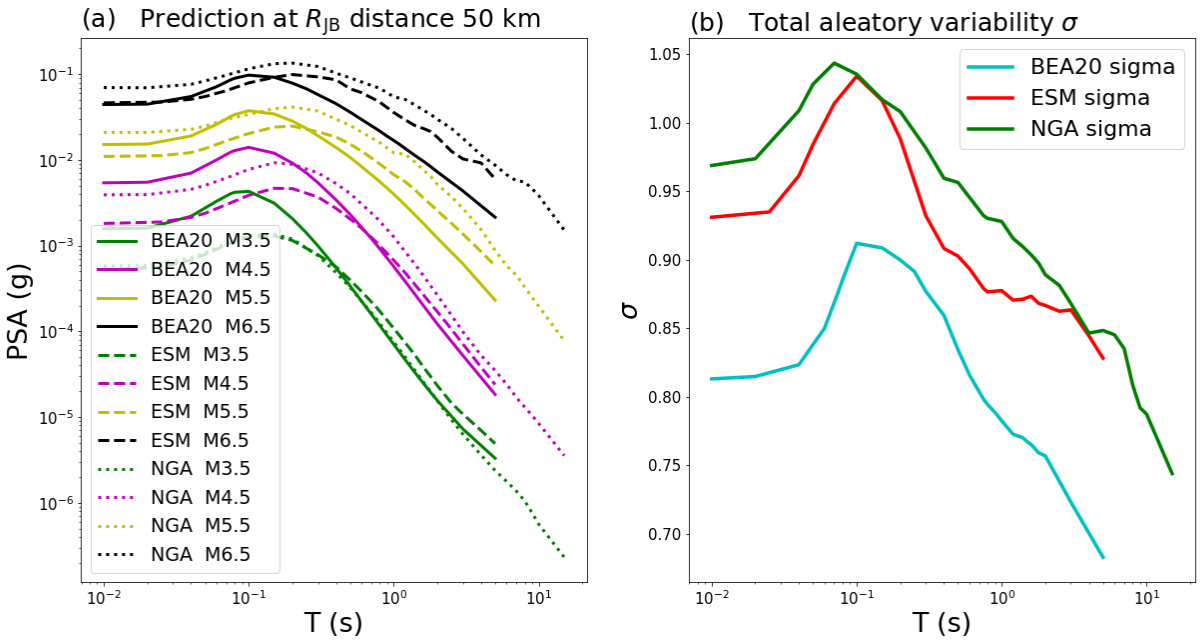


Figure 3 (a) Response spectra of pseudo spectral acceleration (PSA) at R_{JB} 50 km for different magnitudes and (b) the total aleatory variability σ for the three GMMs derived in this study. The difference between the GMMs is expected since they are derived on different datasets and for different regions. The GMM from the BEA20 dataset is derived on the highest number of records and therefore the lowest variability σ.

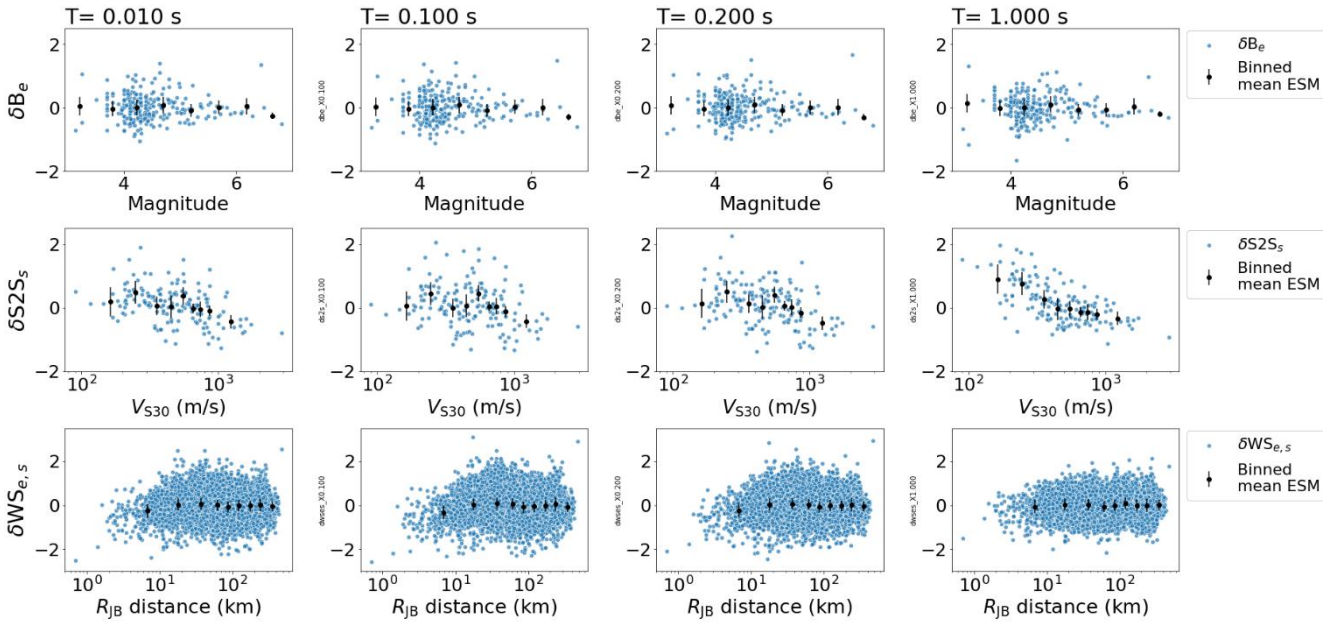


Figure 4 ESM Italy random effect residual plot from the mixed-effect regression used to develop the linear ground-motion model. (Top row, a–d) Distribution and binned mean of δB_e with magnitude M_W for each period T. (Center row, e–h) Distribution and binned mean of $\delta S2S_s$ with V_{S30} in log-scale for each period T. (Bottom row, i–l) Distribution and binned mean of $\delta WS_{e,s}$ with R_{JB} distance. The binned means are with 95% confidence interval. Because a V_{S30} site-term was not included in the fixed effects, $\delta S2S_s$ shows a downwards trend with V_{S30}. The means of δB_e and $\delta WS_{e,s}$ are centered around zero and do not show any trend with magnitude and distance, this show that the GMM regression has captured the scaling of magnitude and distance (Equation 1)

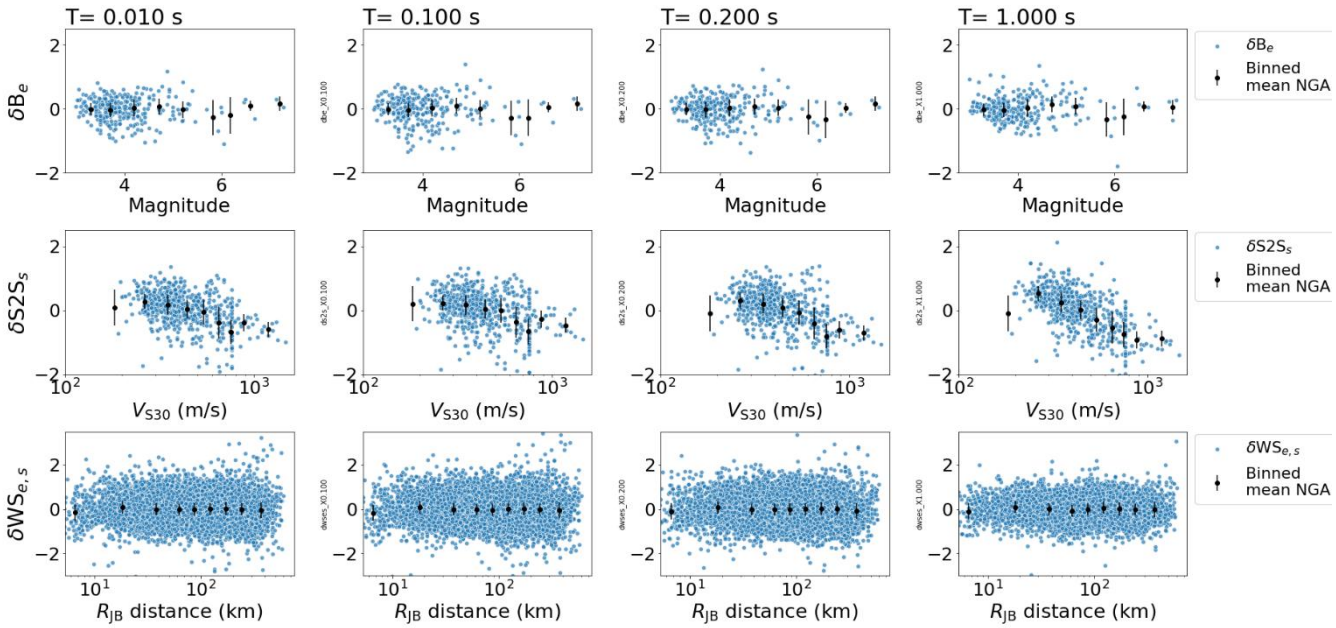


Figure 5 NGA-West 2 California random effect residual plot from the mixed-effect regression used to develop the linear ground-motion model. (Top row, a–d) Distribution and binned mean of δB_e with magnitude M_W for each period T. (Center row, e–h) Distribution and binned mean of $\delta S2S_s$ with V_{S30} in log-scale for each period T. (Bottom row, i–l) Distribution and binned mean of $\delta WS_{e,s}$ with R_{JB} distance. The binned means are with 95% confidence interval. Because a V_{S30} site-term was not included in the fixed effects, $\delta S2S_s$ shows a downwards trend with V_{S30}. The means of δB_e and $\delta WS_{e,s}$ are centered around zero and do not show any trend with magnitude and distance, this show that the GMM regression has captured the scaling of magnitude and distance (Equation 1)

The total aleatory variability of the GMMs is shown in Figure 3b. The BEA20 GMM is derived on the largest and most uniformly distributed dataset (Figure 1) and therefore has the lowest variability. Likewise, the NGA-West2 California GMM is derived on the smallest dataset and has the highest variability. To further evaluate the GMMs we perform a residual analysis to verify whether the fixed-effects have captured the ground-motion response spectra scaling with magnitudes and distances. Figures 4 and 5 show the distributions of δB_e with respect to magnitude, $\delta S2S_s$ with respect to V_{S30}, and $\delta WS_{e,s}$ with respect to distance for the GMMs of ESM Italy and NGA-West2 California, respectively. Although there are few events with higher magnitudes the mean of δB_e is close to zero for magnitudes up to 6 for both GMMs. Likewise $\delta WS_{e,s}$ have a mean consistently close to zero and no clear trend with distance. This confirms that the fixed-effects components of the GMM regressions have captured the scaling of magnitude and distance (equation 5). However, because we did not include a V_{S30} site term in the fixed effects, $\delta S2S_s$, has a negative trend with V_{S30} (Kotha et al., 2018).

Deriving the residuals

The residuals used to analyze nonlinearity are derived in two steps. First a linear GMM is developed on each linear dataset, as described in the previous section. Then each linear GMM is used to derive

the residuals for the testing. The latter step is done on the entire selected dataset, including the previously omitted data with potential to trigger nonlinear soil response at soft-soil stations.

The residuals are derived and analyzed based on the approach of Seyhan and Stewart (2014) who analyzed non-linearity in the NGA-West2 dataset and calibrated an amplification model by parameterizing non-linear site amplification as the declining trend of ground-motion record-specific (within-event) residuals with increasing level of predicted ground motion for rock conditions. Seyhan and Stewart (2014) used the GMM of Boore et al. (2014) to predict ground motion for rock conditions, setting $V_{S30} = 760\text{m/s}$, representing the V_{S30} at rock sites. The prediction on rock $\mu_{e,s}$ for an event e and site s, were then subtracted from the corresponding observation $Y_{e,s}$, to obtain the total residual $\epsilon_{e,s}$:

$$\epsilon_{e,s} = lnY_{e,s} - ln\mu_{e,s} \qquad (2)$$

The total residual $\epsilon_{e,s}$ is split using the mixed-effects regression algorithm of Bates et al. (2015) and the random effects of the events and sites are then quantified into the event and site variability:

$$\epsilon_{e,s} = \delta \mathrm{B_e} + \delta \mathrm{S2S_s} + \delta \mathrm{WS_{e,s}} \qquad (3)$$

Where $\delta \mathrm{B}_e$ and $\delta \mathrm{S2S}_s$ are the event- and site term representing the systematic deviation between the observed ground motions related to event e and site s, respectively, from the median predictions of the GMM, and $\delta \mathrm{WS}_{e,s}$ is the "left-over" residual capturing the record-to-record variability (e.g. Villani and Abrahamson, 2015; Kotha et al., 2018).

Seyhan and Stewart (2014) further used the within-event $\delta \mathrm{W}_{e,s}$ residuals, obtained by subtracting the event-term from the total residuals, to analyze nonlinearity:

$$\delta \mathrm{W_{e,s}} = lnY_{e,s} - \left[\mu_{\mathrm{rock,e,s}} + \delta \mathrm{B_e}\right] = \epsilon_{e,s} - \delta \mathrm{B_e} \qquad (4)$$

Seyhan and Stewart (2014) then parametrized non-linear site amplification as the downgoing trend of $\delta \mathrm{W}_{e,s}$, grouped by V_{S30}, with increasing predicted ground motion on rock ($\mathrm{PGA_{rock}}$). However, the method of Loviknes et al. (2021) include a few modifications to that of Seyhan and Stewart (2014).

Firstly, to avoid regional effects and better investigate the non-linear site effects, the site-specific terms are subtracted from the within-event residuals:

$$\delta \mathrm{W_{e,s}} - \delta \mathrm{S2S_s} = \epsilon_{e,s} - \delta \mathrm{B_e} - \delta \mathrm{S2S_s} = \delta \mathrm{WS_{e,s}} \qquad (5)$$

This modification is motivated by the fact that site responses of different sites can be very different from each other, even within the same V_{S30} bin. Within-event residuals $\delta \mathrm{W}_{e,s}$ grouped by V_{S30} thus features both linear and nonlinear site responses from different sites.

Because $\delta \mathrm{S2S}_s$ is derived here using a large number of records in the linear site amplification range, it can be considered as a reliable representation of empirical linear site response (Bard et al., 2020). The resulting $\delta \mathrm{WS}_{e,s}$ should therefore be distributed around zero for sites experiencing only linear soil response.

Furthermore, considering that events of the same magnitude can trigger very different levels of ground motions (e.g. Bindi et al., 2018, Bindi et al., 2019), the between-event variability is taken into account by including the event-term $\delta \mathrm{B}_e$ in the prediction on rock. The non-linearity is then investigated in the "left-over" residuals $\delta \mathrm{WS}_{e,s}$ with respect to $\mathrm{PGA_{rock}} \exp(\delta \mathrm{B}_e)$.

As described in the previous section, the GMM used to derive the residuals does not include a V_{S30} site-term, the prediction on rock is therefore derived using a rock-adjustment term α_{rock}. This rock adjustment term is the mean δS2Ss for all sites with $V_{S30} > 760$ m/s (Loviknes et al., 2020, Kotha et al., 2018). For both the GMM of ESM Italy and of NGA-West2 California, α_{rock} is derived using only sites with measured V_{S30}. This is to avoid including the uncertainty related to inferred V_{S30} values.

Testing procedure

Predictions of non-linear site-amplification models are tested against residuals derived using each dataset and the linear GMMs described in the previous sections.

Although testing of ground-motion models exists (e.g. Delavaud et al., 2009; Mak et al., 2017; Guéguen et al., 2019), reproducible testing procedures and testable models are still in development. For seismicity models and earthquake forecast models, the Collaboratory for the Study of Earthquake Predictability (CSEP) has developed advanced and community-agreed testing methods (Schorlemmer et al., 2018). CSEP aim to set an international standard for tests to be transparent, reproducible, and prospective. A future goal of CSEP is to develop testing procedures for GMMs and seismic hazard models. As a first step to include testing of site-amplification models within the standard of CSEP, Loviknes et al. (2021) developed a reproducible testing framework using residuals. Because $\delta WS_{e,s}$ is expected to contain the non-linear site response, the trend of $\delta WS_{e,s}$ with respect to $PGA_{rock} \exp(\delta B_e)$ (prediction on rock including event variability) can be used to evaluate predictions of site-amplification models.

The prediction power of the amplification models is evaluated as the deviation between the residuals and the amplification curves measured in mean absolute error (MAE):

$$MAE_S = \frac{\sum_e^N \delta WS_{e,s} - F_{e,s}}{N} \quad (6)$$

where MAE_s is calculated for each site s for N number of events e. $F_{e,s}$ is the modelled site-amplification and $\delta WS_{e,s}$ is, as defined above, the "left-over" residuals assumed to contain the site-response. However, because the MAE score does not have direct physical meaning and only measures the deviation between the residuals and the predictions of the amplification models, it is noteworthy that the model with the best score is only best in a relative sense (Mak et al., 2015).

In Loviknes et al. (2021), the non-linear site-amplification models of Seyhan and Stewart (2014) (SS14), Sandikkaya et al. (2013) (SAB13), Hashash et al. (2020) (H20) and the site-amplification model in the GMM of Abrahamson et al. (2014) (ASK14), were tested against a linear site-amplification model. The linear site-amplification model have the mean of $\delta WS_{e,s} = 0$ for every value of $PGA_{rock} \exp(\delta B_e)$. Here we test the same nonlinear and linear amplification models as Loviknes et al. (2021). All the site amplification models are functions of V_{S30} and ground acceleration on rock. An overview of the nonlinear amplification models is given in Table 1. Both SS14 and ASK14 were developed as a part of the NGA-West2 project and based on the simulations of Kamai et al. (2014). H20 was developed using the NGA-East dataset and 1-D site-response simulations from Harmon et al. (2019). SAB13 is based solely on empirical data from the SHARE (Seismic Hazard HARmonization in Europe Yenier et al., 2010) database.

Table 1 The nonlinear site amplification models tested in this study.

Non-linear model	ID	Dataset	Data type	Simulated data
Seyhan and Stewart (2014)	SS14	NGA – West2	Semi-empirical	Kamai et al. (2014)
Abrahamson et al. (2014)	ASK14	NGA – West2	Simulations	Kamai et al. (2014)
Sandikkaya et al. (2013)	SAB13	SHARE SM Databank	Empirical	
Hashash et al. (2020)	H20	NGA – East	Simulations	Harmon et al. (2019)

Figure 6 shows how the amplification models compare with the $\delta WS_{e,s}$ residuals with respect to PGA_{rock} $\exp(\delta B_e)$ for each V_{S30} bin and period. Non-linear site effects are mainly expected, and modeled, for low V_{S30} and strong ground motions, which is evident for Figure 6 where the difference between the nonlinear and linear amplification models is very small for $V_{S30} > 500$ m/s. The test was therefore performed on a subset of stations with $V_{S30} < 500$m/s and more than four records at predicted $PGA_{rock} > 0.05$g. For the same reason, the MAE score was only calculated for ground-motions with $PGA_{rock} > 0.05$g.

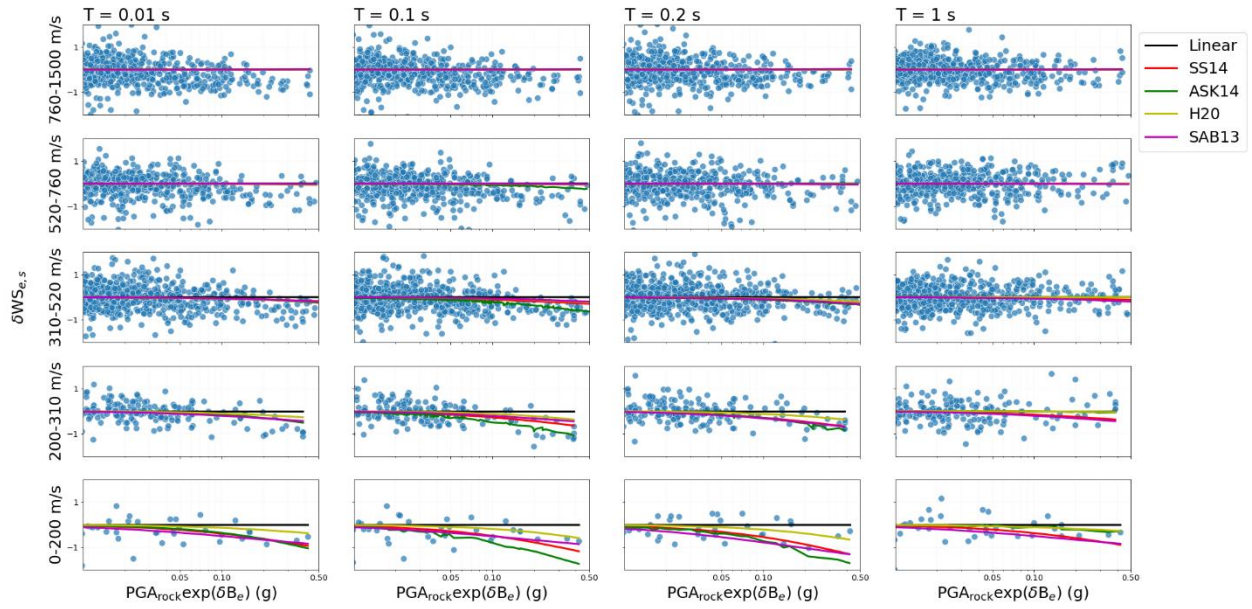


Figure 6 The ESM stations grouped by V_{S30} and the linear and non-linear site amplification models compared to $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability (PGA_{rock} $\exp(\delta B_e)$). The non-linear models are from Seyhan and Stewart (2014) (SS14), Abrahamson et al. (2014) (ASK14), Hashash et al. (2020) (H20) and Sandikkaya et al. (2013) (SAB13).

RESULTS

Using the linear GMM and residuals described in the previous sections, we apply the testing procedure to two new datasets; ESM Italy and NGA-West2. The result for each of the datasets are given in the following Section.

ESM Italy

As explained in the Method Section, the $\delta WS_{e,s}$ is the empirical site response and should be distributed around zero when linear. In Figure 6 the site response of ESM Italy has a clear downgoing trend for lower V_{S30} bins (V_{S30} < 520 m/s) and high $PGA_{rock}\exp(\delta B_e)$ (> 0.05g.), this is in contrast to the site response of the KiK-net stations from Loviknes et al. (2021) where the $\delta WS_{e,s}$ residuals had a high variability and no clear trend was observed. However, a downgoing trend also observed for T=0.01 s at high V_{S30} (V_{S30} > 520 m/s). It is therefore possible that the trends in the $\delta WS_{e,s}$ residuals are not only related to nonlinear site amplification, as also discussed by Stafford et al. (2017). This trend at low period and high V_{S30} should therefore be explored in futures studies.

Out of all the Italian ESM stations, 19 soft-soil stations (V_{S30} < 500m/s) have more than four records at predicted PGA_{rock} > 0.05g and were therefore selected for the testing of site amplification models. (Figure 7). 12 of these stations (red triangles in Figure 7) had a non-linear amplification model score better than the linear amplification model. This is more than what was found for KiK-net by Loviknes et al. (2021), where the linear amplification model scored better for most of the stations and only 5 out of 20 stations had a nonlinear model score better than the linear amplification model.

For each station selected in the test, the amplification model which has the lowest MAE score for more than half the periods (T=0.01 s, 0.1 s, 0.2 s, 1 s) is considered the best model for the station. Figures 8 and 9 show how the amplification model with the best (lowest) score compare with the site response for each period applied in test and a selection of stations where the nonlinear models scored best. A downgoing trend is observed in all these stations, but only predicted for the stations with low V_{S30} (V_{S30} < 300 m/s) in Figure 8. The stations in Figure 9 have higher V_{S30} (V_{S30} > 400 m/s) and because the nonlinear amplification models depend on V_{S30} they are not able to predict the observed de-amplification.

The remaining stations are shown in Figures S1-S5 and the scores from the test is given in Table S1 in the Supplement at the end of this report. Figure 10 show the number of times each site amplification model has the best score per period, it is clear that no model that has the best score significantly more times than the others for all periods. In addition, as with the other selected stations, the within-site amplifications variability remains large compared to the difference between the model predictions (Figure 6), and more data will therefore be needed to assess whether the differences between the models are significant.

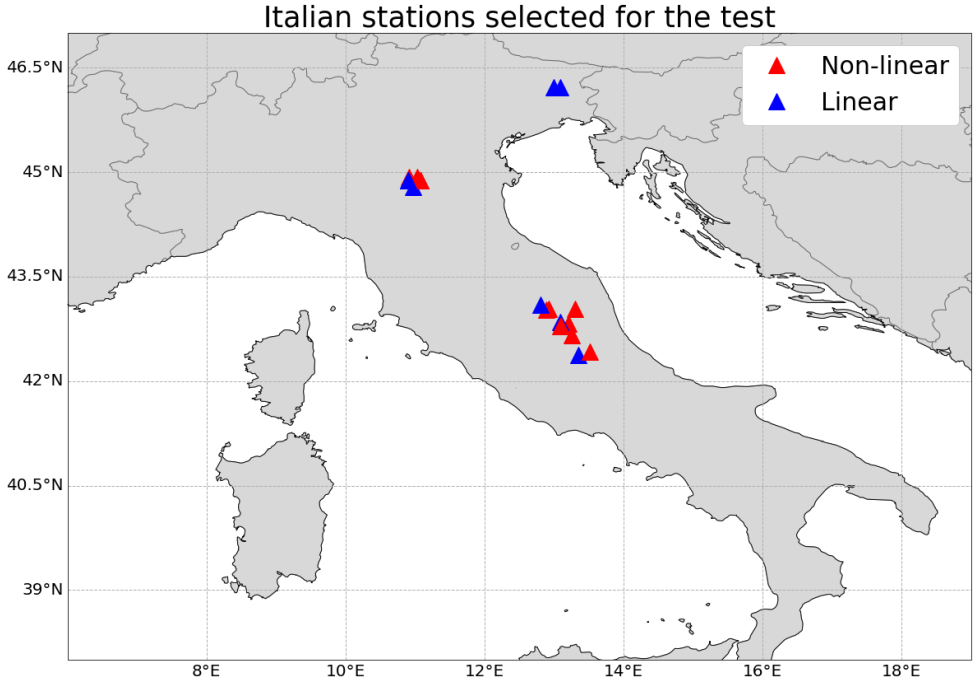


Figure 7 Map of Italy showing the location of the stations selected for the test. The blue triangles show the stations where the linear amplification model had the best score, and the red triangles show the stations where one of the non-linear amplification models had the best score.

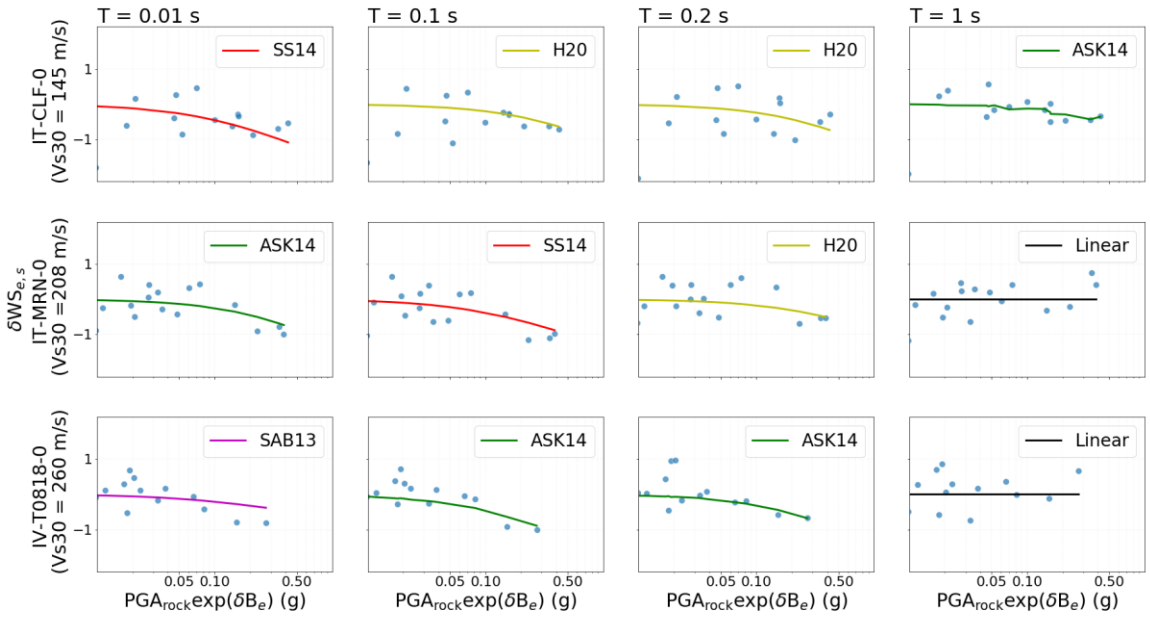


Figure 8 The ESM stations selected for the test with lowest V_{S30} (145 – 260 m/s) and the linear and non-linear site amplification models that scored best for each station compared to $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$). The non-linear models are from Seyhan and Stewart (2014) (SS14), Abrahamson et al. (2014) (ASK14), Hashash et al. (2020) (H20) and Sandikkaya et al. (2013) (SAB13).

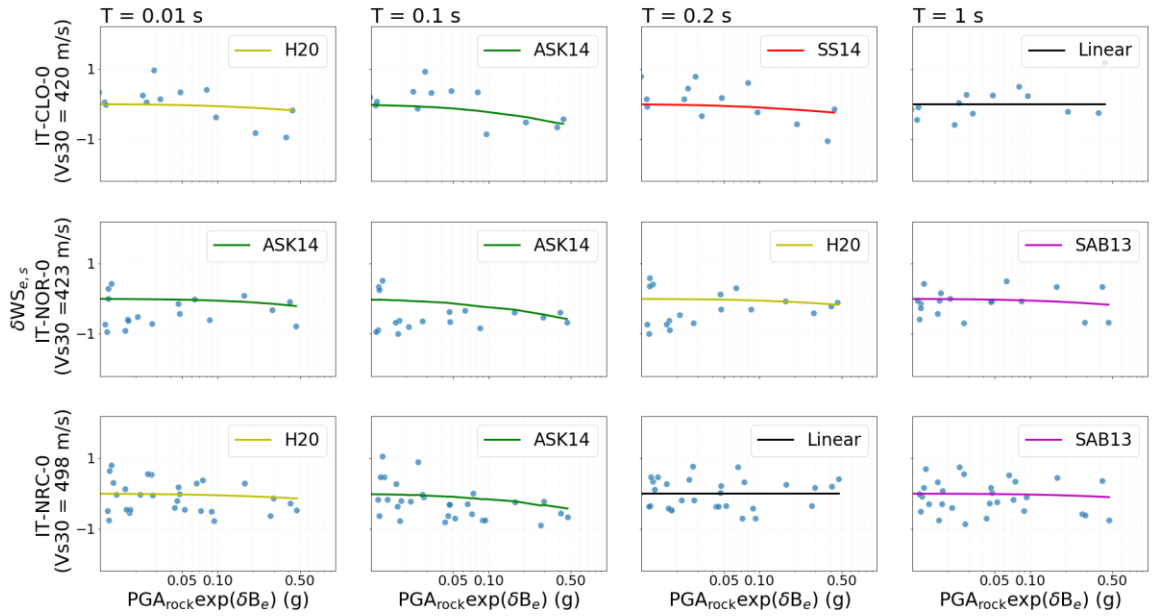


Figure 9 The ESM stations selected for the test with highest V_{S30} (420-498 m/s) and the linear and non-linear site amplification models that scored best for each station compared to $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$). The non-linear models are from Seyhan and Stewart (2014) (SS14), Abrahamson et al. (2014) (ASK14), Hashash et al. (2020) (H20) and Sandikkaya et al. (2013) (SAB13).

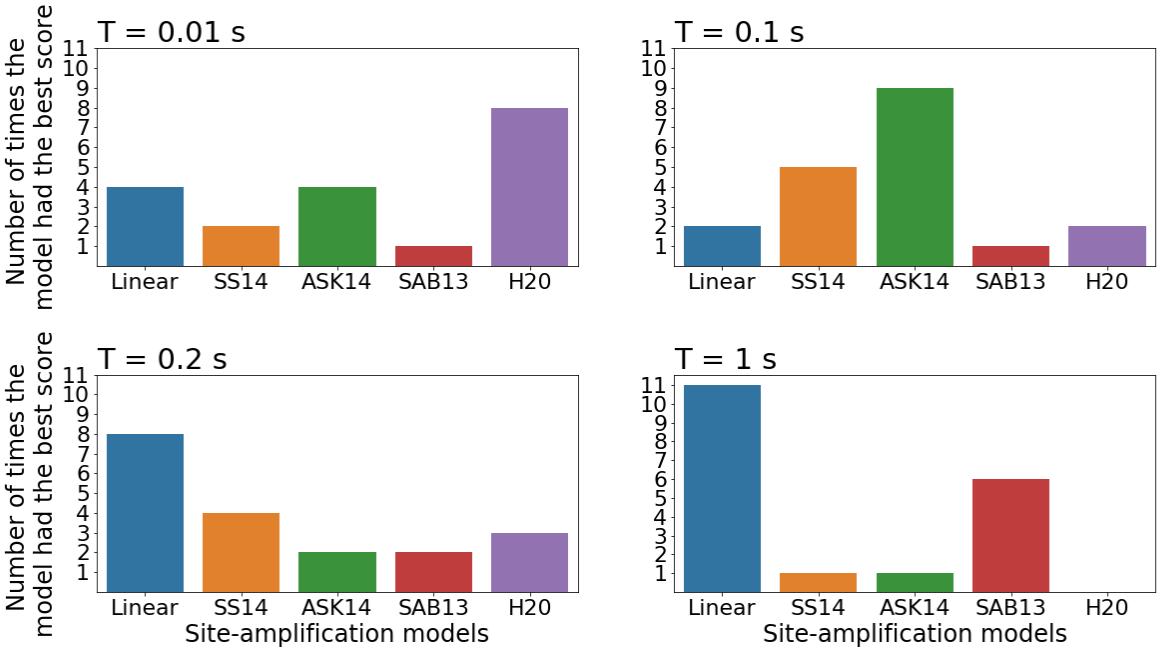


Figure 10 The number of times (stations) where each site-amplification model had the best (lowest) score for each period used in the test on the ESM Italy stations. No one of the models had the best score most times for all the periods.

NGA-West 2 California

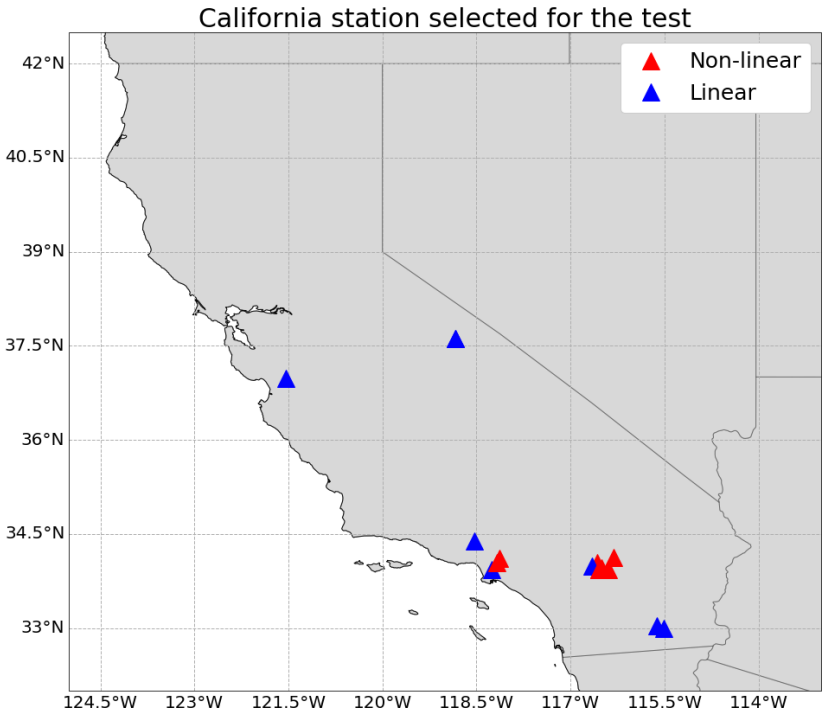


Figure 11 Map of California, USA, showing the location of the stations selected for the test. The blue triangles show the stations where the linear amplification model had the best score, and the red triangles show the stations where one of the non-linear amplification models had the best score.

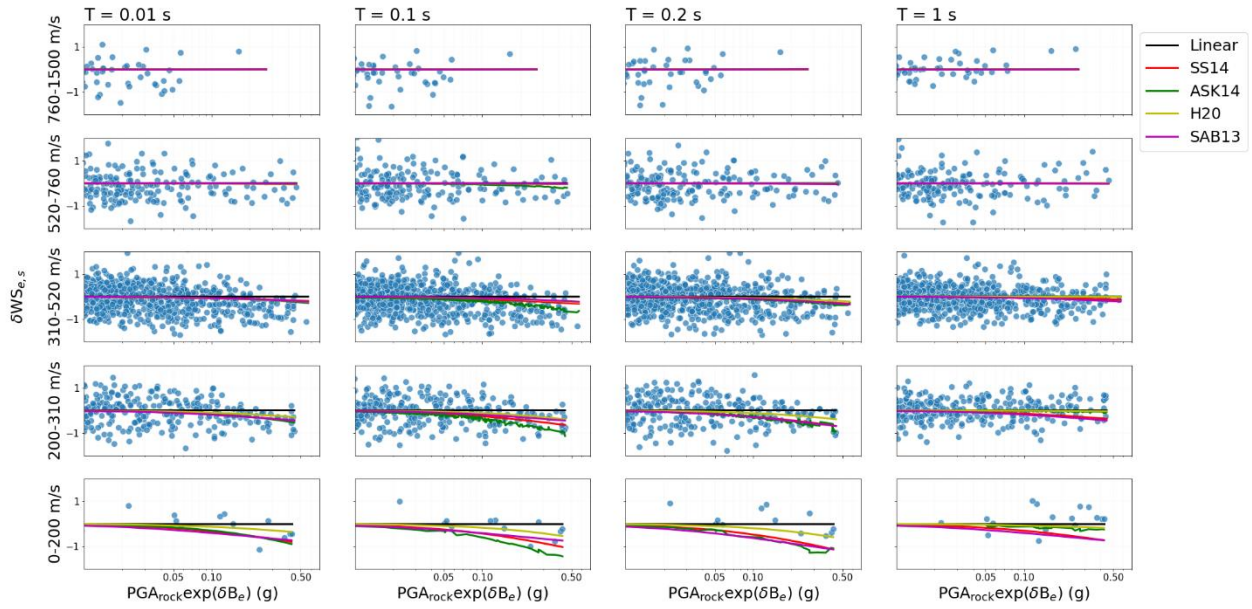


Figure 12 The NGA-West 2 California stations grouped by V_{S30} and the linear and non-linear site amplification models compared to $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$). The non-linear models are from Seyhan and Stewart (2014) (SS14), Abrahamson et al. (2014) (ASK14), Hashash et al. (2020) (H20) and Sandikkaya et al. (2013) (SAB13).

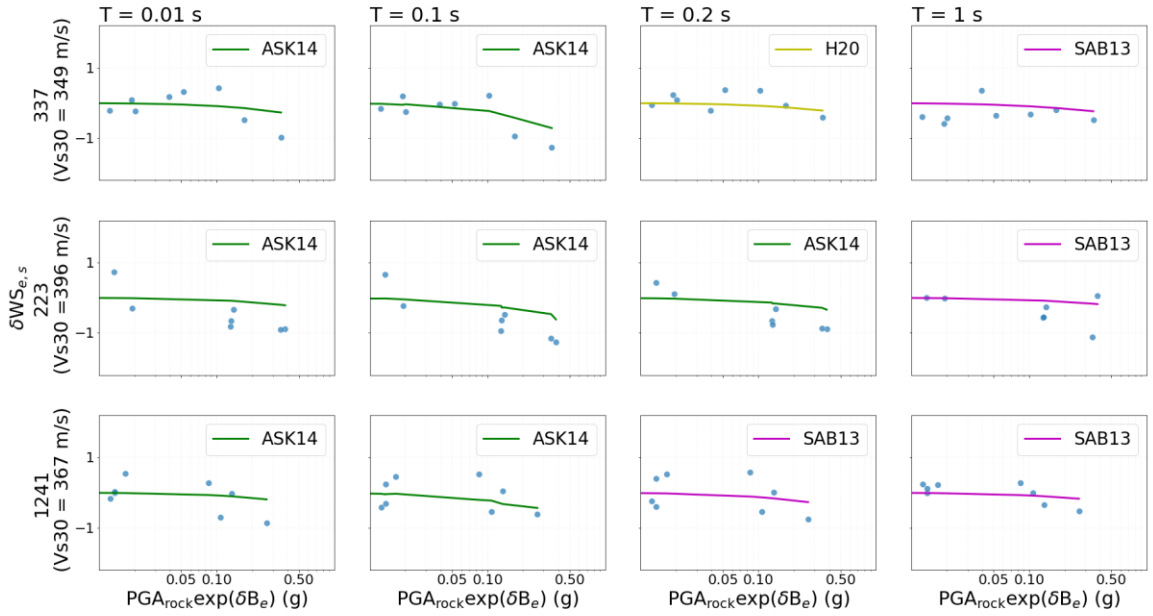


Figure 13 A selection of the NGA West 2 stations grouped the linear and non-linear site amplification models that scored best for each station, compared to $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$). The non-linear models are from Seyhan and Stewart (2014) (SS14), Abrahamson et al. (2014) (ASK14), Hashash et al. (2020) (H20) and Sandikkaya et al. (2013) (SAB13).

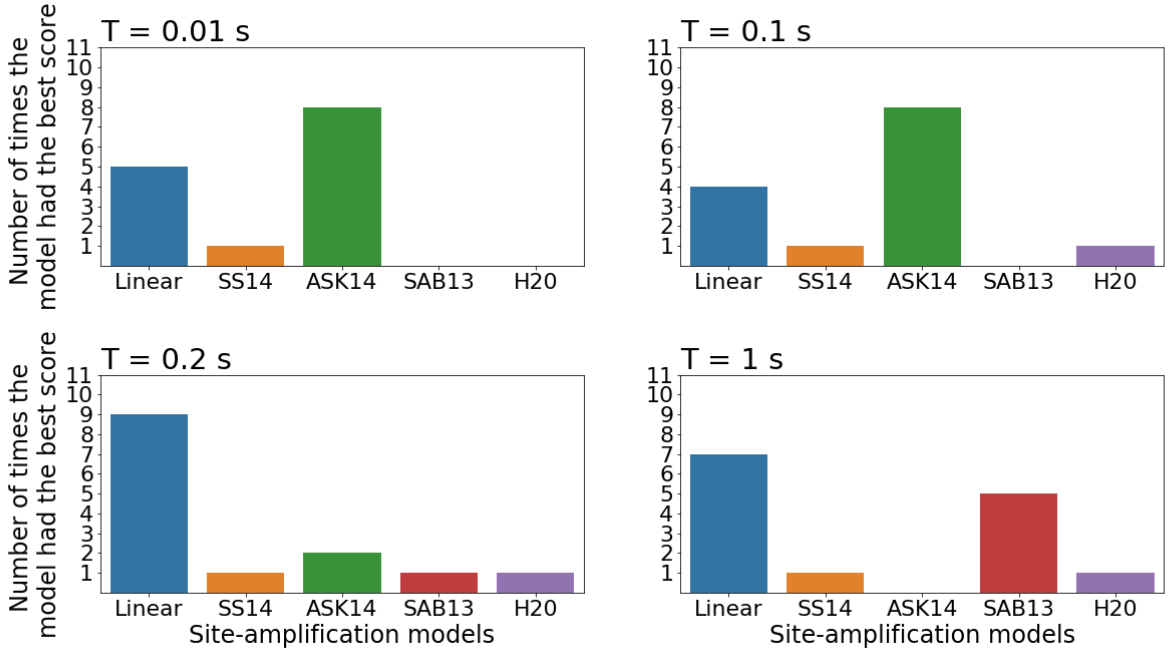


Figure 14 The number of times (stations) where each site-amplification model had the best (lowest) score for each period used in the test of the NGA-West2 California stations. No one of the models had the best score most times for all the periods.

In the NGA-West2 database there are 827 stations in California, out of these 14 stations were selected to have sufficient records for the test and are shown in Figure 11. A non-linear amplification model scored better than the linear amplification model for 6 of these stations (red triangles in Figure 11). However, most of the stations have recorded few strong-motion records and only two stations have more than 5 records at $PGA_{rock} > 0.05$ g. A clear trend is therefore difficult to observe in the residuals, even when grouped by V_{S30}, as shown in Figure 12. Some stations, however, like in Figure 13, show signs of de-amplification at high predicted $PGA_{rock} \exp(\delta B_e)$, which is also predicted by the non-linear models.

The remaining stations are shown in Figures S6-S8 and the scores from the test is given in Table S2 in the supplement. Out of the site-amplification models the ASK14 model from the GMM of Abrahamson et al. (2014) has the better score the most times for most periods ($T \leq 0.1$ s), this is shown in Figure 13. However, for higher periods ($T = 1$ s), the ASK14 model does not have the best score once, more strong-motion data is therefore needed to quantitavely evaluate which nonlinear model has the best performance consistently over the periods.

DISCUSSION

In contrast to the KiK-net stations tested by Loviknes et al. (2021), for both ESM and NGA-West 2 California, a higher fraction of the selected stations had a non-linear amplification model score better than the linear model. However, most of the non-linear site amplification models are derived partly on either European data included in the ESM database (Sandikaya et al. 2013) or on the NGA-West 2 dataset (Seyhan and Stewart 2014, Abrahamson et al. 2014), a better correlation between the observations and the observed datasets are therefore partly expected. Furthermore, KiK-net stations, are mainly installed on stiff sites with weathered rock or on thin sediment layers (Aoi et al. 2004),

which might explain why less nonlinearity is observed for these stations than for the Italian ESM and NGA-West 2 California stations. In addition, the difference in score between the nonlinear models and the linear model are for many stations small, see Table S1 and Table S2, for these cases a simple linear site amplification model might therefore still be the best choice.

Out of the three datasets tested using the testing framework of Loviknes et al. (2021), the nonlinear amplification models had the best score for most stations in the ESM Italy dataset. For the Italian stations with low V_{S30} the nonlinear model predictions and observations follow the same de-amplification trend. For stations with higher V_{S30}, however, the models do not predict the observed de-amplification. This result further suggests that V_{S30} is not a suitable proxy for predicting nonlinearity. Alternative site proxies, in particularly moving from geophysical parameters to geological and geotechnical parameters, will therefore be explored in future studies.

Another limitation of the test is how the nonlinear range is defined. In this study we define the nonlinear range as $PGA_{rock} > 0.05$g, while in Loviknes et al. (2021), a record is considered in the nonlinear range if it is recorded on a soft-soil station with $V_{S30} > 760$ m/s and has PGA > 0.05g.
However, as also discussed in the Method section, these thresholds only defines where a record have the potential to be nonlinear, and not whether the record is actually inducing nonlinear soil behavior. Nonlinear soil behavior might also occur outside this range, indeed Regnier et al. (2013) suggested that sites with shallow shear wave velocity contrast has a higher chance of triggering nonlinear soil behavior at low input values of PGA. An alternative is therefore to use other proxies to supplement or replace the PGA threshold for characterizing nonlinearity. One example is the shear strain proxy, defined by Gueguen et al. (2019) as PGV/ V_{S30}, to characterize nonlinearity. As mentioned in the Method section, the objective of this study is to test the prediction power the four nonlinear site amplification models SS14 (Seyhan and Stewart, 2014), ASK14 (Abrahamson et al., 2014), H20 (Hashash et al., 2020) and SAB13 (Sandikkaya et al., 2013). And because these models use PGA and V_{S30} to predict nonlinearity, we chose to use these proxies for characterizing nonlinear in this study. Other more specific ways to characterize nonlinearity should, however, be explored in future studies.

The main limitations for all the three datasets used on the analysis of non-linearity and the test is still the limited number of strong-motion records in the non-linear site-amplification range. This is especially clear for NGA-West 2 California where only 2 of the selected stations have more than four records. However, also for ESM and the BEA20 database, there were only 19 and 20 soft-soil stations, respectively, per database that had recorded more than four records with $PGA_{rock} > 0.05$g. The development of a new updated KiK-net and K-net database is therefore ongoing.

CONCLUSION

In this study, we used the testing framework of Loviknes et al. (2021) to test non-linear amplification models on Italian stations in the ESM database and stations in California from the NGA-West 2 database. For all three datasets the same method is used to derive a linear GMM and test non-linear amplification models using residuals between the predicted and observed ground-motions.

The nonlinear amplification models tested in this study has a better score for the Italian stations in the ESM dataset and a downgoing trend is observed in the site response of many of the stations. However, for stations with higher V_{S30} the nonlinear amplification models are not able to predict the observed de-amplification. For NGA-West 2 California there are limited numbers of strong-motion records and more data is needed to analyze nonlinearity in the stations.

We therefore argue that the conclusions of Loviknes et al. (2021), that non-linear amplification models based on PGA and V_{S30} are not fully able to capture the non-linear site response. Alternative methods to characteristic non-linear site amplification should therefore be investigated.

ACKNOWLEDGMENTS

The authors are grateful to the ORFEUS committee and the PEER center for their great work in compiling and sharing the datasets. In addition, the authors would like to thank the open-source community for the Linux operating system and the many programs used in this study. This research is funded by the European Union's Horizon 2020 research and innovation program Real-time Earthquake Risk Reduction for a Resilient Europe "RISE" project, under grant Agreement 821115 and the European Commission, ITN Marie Sklodowska-Curie URBASIS-EU project, under the grant agreement No. 813137.

REFERENCES

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

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

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. (2014). "NGA-West2 database". *Earthquake Spectra*, **30**(3):989–1005.

Bahrampouri, M., Rodriguez-Marek, A., Shahi, S., and Dawood, H. (2020). "An updated database for ground motion parameters for KiK-net records". *Earthquake Spectra*, page 875529302095244.

Bard, P. Y., Bora, S. S., Hollender, F., Laurendeau, A., and Traversa, P. (2020). "Are the Standard VS-Kappa Host-to-Target Adjustments the Only Way to Get Consistent Hard-Rock Ground Motion Prediction?" *Pure and Applied Geophysics*, **177**(5):2049–2068.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software*, **67**(1).

Bindi, D., Picozzi, M., Spallarossa, D., Cotton, F., and Kotha, S. R. (2019). "Impact of magnitude selection on aleatory variability associated with groundmotion prediction equations: Part II-analysis of the between-event distribution in central Italy". *Bulletin of the Seismological Society of America*, **109**(1):251–262.

Bindi, D., Spallarossa, D., Picozzi, M., Scafidi, D., and Cotton, F. (2018). "Impact of magnitude selection on aleatory variability associated with groundmotion prediction equations: Part I—local, energy, and moment magnitude calibration and stress-drop variability in central Italy." *Bulletin of the Seismological Society of America*, **108**(3):1427–1442.

Bonilla, L. F., Archuleta, R. J., and Lavallée, D. (2005). "Hysteretic and dilatant behavior of cohesionless soils and their effects on nonlinear site response: Field data observations and modeling." *Bulletin of the Seismological Society of America*, **95**(6):2373–2395.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M. (2014). "NGAWest2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes." *Earthquake Spectra*, **30**(3):1057–1085.

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

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**(2):1281–1302.

Delavaud, E., Scherbaum, F., Kuehn, N., and Riggelsen, C. (2009). "Informationtheoretic selection of ground-motion prediction equations for seismic hazard analysis: An applicability study using californian data." *Bulletin of the Seismological Society of America*, **99**(6):3248–3263.

Derras, B., Bard, P. Y., and Cotton, F. (2016). “Site-condition proxies, ground motion variability, and data-driven GMPEs: Insights from the NGA-West2 and RESORCE data sets.” *Earthquake Spectra*, **32**(4):2027–2056.

Field, E. H., Johnson, P. A., Beresnev, I. A., and Zeng, Y. (1997). “Nonlinear ground-motion amplification by sediments during the 1994 Northridge earthquake.” *Nature*, **390**(6660):599–602.

Guéguen, P., Bonilla, L. F., and Douglas, J. (2019). “Comparison of soil nonlinearity (in situ stress-strain relation and G/G max Reduction) observed in strong-motion databases and modeled in ground-motion prediction equations.” *Bulletin of the Seismological Society of America*, **109**(1):178–186.

Hashash, Y. M., Ilhan, O., Hassani, B., Atkinson, G. M., Harmon, J., and Shao, H. (2020). “Significance of site natural period effects for linear site amplification in central and eastern North America: Empirical and simulation-based models.” *Earthquake Spectra*, **36**(1):87–110.

Kamai, R., Abrahamson, N. A., and Silva, W. J. (2014). “Nonlinear horizontal site amplification for constraining the NGA-West2 GMPEs.” *Earthquake Spectra*, **30**(3):1223–1240.

Kotha, S. R., Bindi, D., and Cotton, F. (2016). “Partially non-ergodic region specific GMPE for Europe and Middle-East.” *Bulletin of Earthquake Engineering*, **14**(4):1245–1263.

Kotha, S. R., Cotton, F., and Bindi, D. (2018). “A new approach to site classification: Mixed-effects Ground Motion Prediction Equation with spectral clustering of site amplification functions.” *Soil Dynamics and Earthquake Engineering*, 110:318–329.

Luzi L., Puglia R., Russo E., D'Amico M., Felicetta C., Pacor F., Lanzano G., Çeken U., Clinton J., Costa G., Duni L., Farzanegan E., Gueguen P., Ionescu C., Kalogeras I., Özener H., Pesaresi D., Sleeman R., Strollo A., Zare M. (2016). The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data. Seismological Research Letters ; 87 (4): 987–997. doi:https://doi.org/10.1785/0220150278

Luzi L., Puglia R., Russo E., D'Amico M., Felicetta C., Pacor F., Lanzano G., Çeken U., Clinton J., Costa G., Duni L., Farzanegan E., Gueguen P., Ionescu C., Kalogeras I., Özener H., Pesaresi D., Sleeman R., Strollo A., Zare M. (2016). The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data. Seismological Research Letters ; 87 (4): 987–997. doi:https://doi.org/10.1785/0220150278

Loviknes, K., S. R. Kotha, F. Cotton, and D. Schorlemmer (2021). Testing Nonlinear Amplification Factors of Ground-Motion Models, Bull. Seismol. Soc. Am. 111, 2121–2137, doi: 10.1785/0120200386

Mak, S., Clements, R. A., and Schorlemmer, D. (2015). “Validating intensity prediction equations for Italy by observations.” *Bulletin of the Seismological Society of America*, **105**(6):2942–2954

Mak, S., Clements, R. A., and Schorlemmer, D. (2017). “Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting.” *Bulletin of the Seismological Society of America*, **107**(2):949–965.

Paolucci, R., Aimar, M., Ciancimino, A., Dotti, M., Foti, S., Lanzano, G., ... & Vanini, M. (2021). Checking the site categorization criteria and amplification factors of the 2021 draft of Eurocode 8 Part 1–1. *Bulletin of Earthquake Engineering*, 1-36.

Régnier, J., Cadet, H., Fabian Bonilla, L., Bertrand, E., and Semblat, J. F. (2013). “Assessing nonlinear behavior of soils in seismic site response: Statistical analysis on KiK-net strong-motion data.” *Bulletin of the Seismological Society of America*, **103**(3):1750–1770.

Sandikkaya, M. A., Akkar, S., and Bard, P. Y. (2013). “A nonlinear siteamplification model for the next pan-European ground-motion prediction equations.” *Bulletin of the Seismological Society of America*, **103**(1):19–32.

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., Mak, S., Rhoades, D. A., Gerstenberger, M. C., Hirata, N., Liukis, M., Maechling, P. J., Strader, A., Taroni, M., Wiemer, S., Zechar, J. D., and Zhuang, J. (2018). “The collaboratory for the study of earthquake predictability: Achievements and priorities.” *Seismological Research Letters*, **89**(4):1305–1313.

Seyhan, E. and Stewart, J. P. (2014). “Semi-empirical nonlinear site amplification from NGA-West2 data and simulations.” *Earthquake Spectra*, **30**(3):1241–1256.

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.” *Bulletin of the Seismological Society of America*, **104**(2):702–719.

Stafford, P. J., Rodriguez-Marek, A., Edwards, B., Kruiver, P. P., & Bommer, J. J. (2017). Scenario Dependence of Linear Site-Effect Factors for Short-Period Response Spectral Ordinates Scenario Dependence of Linear Site-Effect Factors for Short-Period Response Spectral Ordinates. *Bulletin of the Seismological Society of America*, *107*(6), 2859-2872.

Thompson, E. M. and Wald, D. J. (2016). “Uncertainty in VS30-based site response.” *Bulletin of the Seismological Society of America*, **106**(2):453–463.

Villani, M. and Abrahamson, N. A. (2015). ”Repeatable site and path effects on the ground-motion sigma based on empirical data from Southern California and simulated waveforms from the cyberShake platform.” *Bulletin of the Seismological Society of America*, **105**(5):2681–2695.

SUPPLEMENT

Description of the Supplemental Material

The supplemental material contains tables and figures from the result of the test conducted in the study. Figures S1-S5 show the ESM Italy stations with the site amplification model that scored best for each station and periods. Table S1 give the mean absolute error scores at all ESM Italy stations and periods used in the test. For the NGA-West 2 California stations Figures S6-S8 show the best-scoring amplification for each station and period, and Table S2 gives the mean absolute error scores for all the stations and periods.

ESM Italy

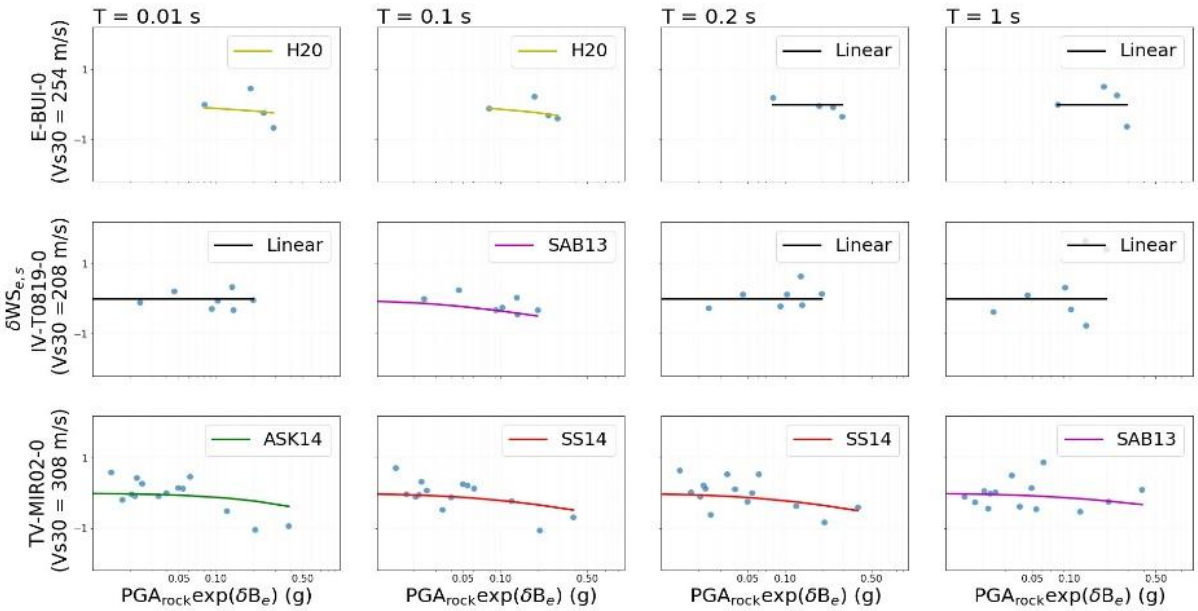


Figure S1: Italian ESM stations with V_{S30} between 208-308 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

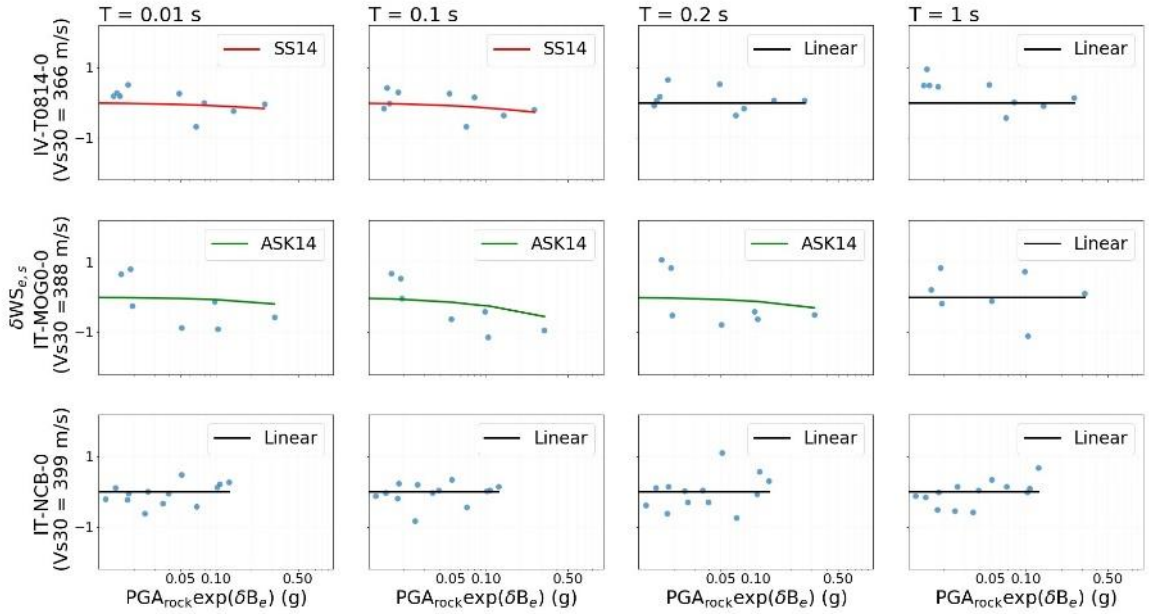


Figure S2: Italian ESM stations with V_{S30} between 366-399 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

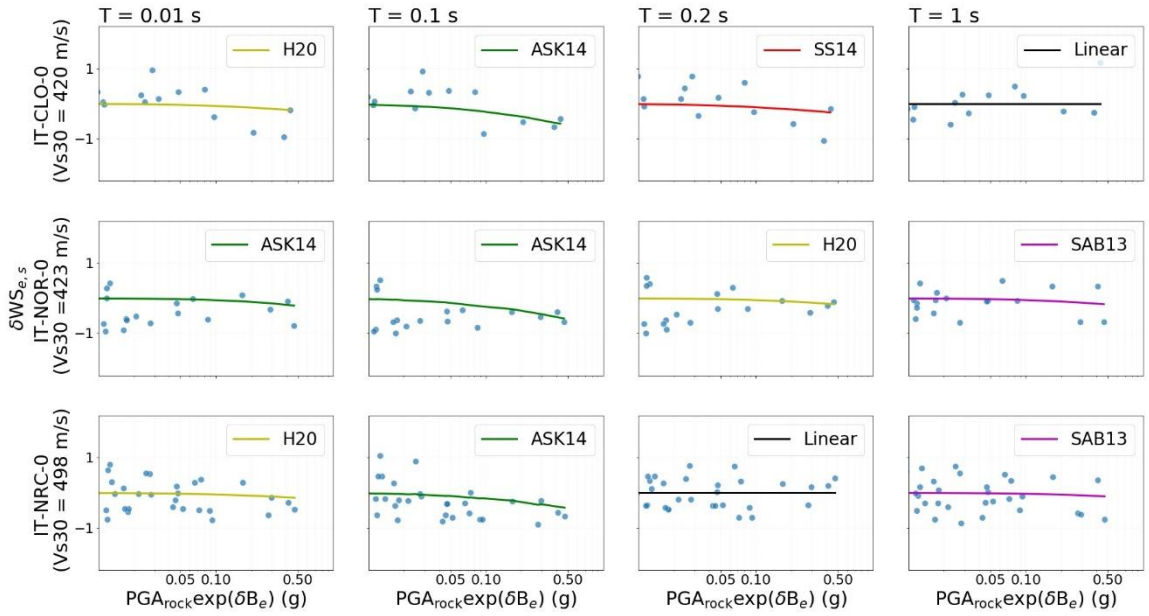


Figure S3: Italian ESM stations with V_{S30} between 420-498 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

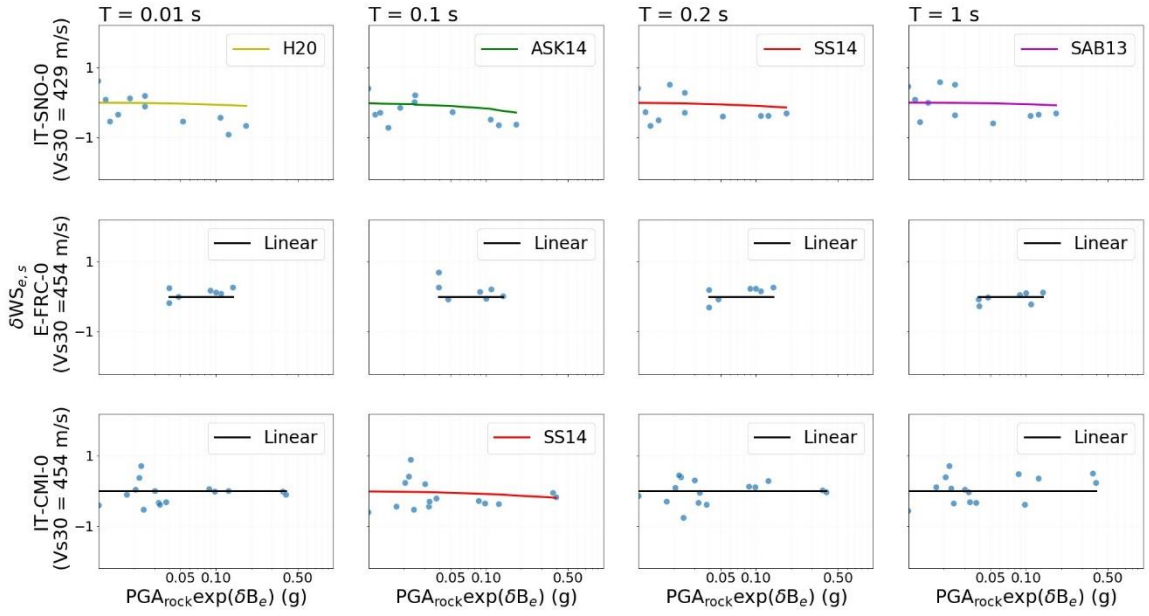


Figure S4: Italian ESM stations with V_{S30} between 429-454 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

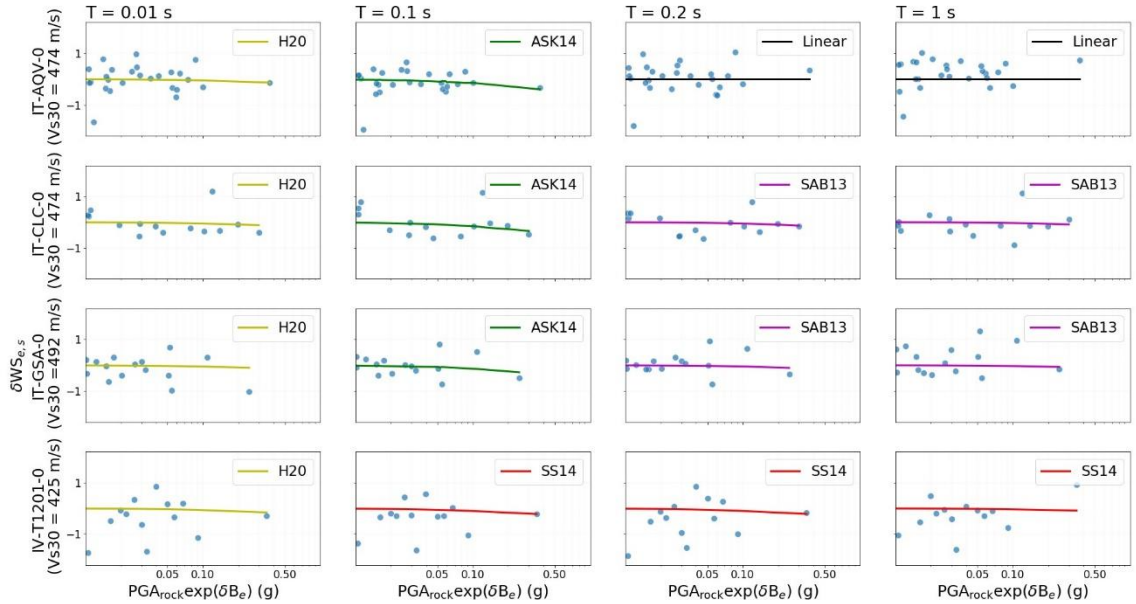


Figure S5: Italian ESM stations with V_{S30} between 425-474 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

Table S1: The mean absolute error (MAE) score of all the amplification models with the ”left-over” residuals $\delta WS_{e,s}$ for each ESM Italy station at all periods T.

Stations	Periods	Linear	SS14	ASK14	H20	SAB13	Best-scoring Model
E-BUI-0	0.01	0.339	0.337	0.321	**0.293**	0.345	H20
E-FRC-0	0.01	**0.164**	0.214	0.212	0.227	0.209	Linear
IT-AQV-0	0.01	0.353	0.340	0.336	**0.335**	0.338	H20
IT-CLC-0	0.01	0.427	0.390	0.386	**0.378**	0.391	H20
IT-CLF-0	0.01	0.567	**0.335**	0.361	0.373	0.348	SS14
IT-CLO-0	0.01	0.544	0.466	0.452	**0.449**	0.458	H20
IT-CMI-0	0.01	**0.042**	0.063	0.077	0.088	0.066	Linear
IT-GSA-0	0.01	0.675	0.666	0.664	**0.661**	0.665	H20
IT-MOG0-0	0.01	0.618	0.527	**0.525**	0.527	0.531	ASK14
IT-MRN-0	0.01	0.598	0.382	**0.344**	0.462	0.415	ASK14
IT-NCB-0	0.01	**0.308**	0.353	0.350	0.353	0.349	Linear
IT-NOR-0	0.01	0.318	0.276	**0.270**	0.273	0.273	ASK14
IT-NRC-0	0.01	0.429	0.402	0.394	**0.385**	0.399	H20
IT-SNO-0	0.01	0.635	0.573	0.575	**0.564**	0.579	H20
IV-T0814-0	0.01	0.245	**0.231**	0.238	0.232	0.234	SS14
IV-T0818-0	0.01	0.520	0.324	0.326	0.393	**0.321**	SAB13
IV-T0819-0	0.01	**0.204**	0.287	0.270	0.220	0.326	Linear
IV-T1201-0	0.01	0.488	0.444	0.436	**0.435**	0.442	H20
TV-MIR02-0	0.01	0.615	0.509	**0.493**	0.534	0.506	ASK14

E-BUI-0	0.10	0.255	0.308	0.298	**0.138**	0.194	H20
E-FRC-0	0.10	**0.107**	0.167	0.173	0.138	0.132	Linear
IT-AQV-0	0.10	0.269	0.228	**0.219**	0.240	0.247	ASK14
IT-CLC-0	0.10	0.418	0.370	**0.358**	0.378	0.385	ASK14
IT-CLF-0	0.10	0.526	0.498	0.537	**0.252**	0.418	H20
IT-CLO-0	0.10	0.565	0.430	**0.387**	0.462	0.477	ASK14
IT-CMI-0	0.10	0.239	**0.177**	0.209	0.189	0.197	SS14
IT-GSA-0	0.10	0.536	0.519	**0.511**	0.522	0.526	ASK14
IT-MOG0-0	0.10	0.777	0.627	**0.619**	0.683	0.688	ASK14
IT-MRN-0	0.10	0.666	0.308	**0.276**	0.460	0.391	ASK14
IT-NCB-0	0.10	**0.200**	0.275	0.273	0.244	0.241	Linear
IT-NOR-0	0.10	0.527	0.359	**0.315**	0.404	0.422	ASK14
IT-NRC-0	0.10	0.516	0.442	**0.408**	0.456	0.473	ASK14
IT-SNO-0	0.10	0.513	0.406	**0.400**	0.445	0.455	ASK14
IV-T0814-0	0.10	0.347	**0.276**	0.285	0.277	0.277	SS14
IV-T0818-0	0.10	0.522	0.322	**0.306**	0.378	0.356	ASK14
IV-T0819-0	0.10	0.273	0.204	0.173	0.153	**0.152**	SAB13
IV-T1201-0	0.10	0.395	**0.319**	0.338	0.340	0.349	SS14
TV-MIR02-0	0.10	0.464	0.313	**0.286**	0.363	0.353	ASK14
E-BUI-0	0.20	**0.160**	0.414	0.288	0.184	0.448	Linear
E-FRC-0	0.20	**0.225**	0.312	0.267	0.273	0.294	Linear
IT-AQV-0	0.20	**0.385**	0.401	0.396	0.396	0.402	Linear
IT-CLC-0	0.20	0.258	0.224	0.229	0.226	**0.222**	SAB13
IT-CLF-0	0.20	0.513	0.629	0.603	**0.464**	0.778	H20
IT-CLO-0	0.20	0.523	**0.428**	0.442	0.442	0.432	SS14
IT-CMI-0	0.20	**0.117**	0.232	0.184	0.186	0.224	Linear
IT-GSA-0	0.20	0.535	0.531	0.529	0.529	**0.527**	SAB13
IT-MOG0-0	0.20	0.572	**0.423**	0.480	0.490	0.437	SS14
IT-MRN-0	0.20	0.520	0.562	0.474	**0.381**	0.637	H20
IT-NCB-0	0.20	**0.564**	0.610	0.578	0.577	0.598	Linear
IT-NOR-0	0.20	0.236	0.174	**0.169**	0.169	0.177	ASK14
IT-NRC-0	0.20	**0.424**	0.449	0.440	0.443	0.452	Linear
IT-SNO-0	0.20	0.371	**0.265**	0.315	0.313	0.283	SS14
IV-T0814-0	0.20	**0.163**	0.220	0.209	0.199	0.223	Linear
IV-T0818-0	0.20	0.417	0.094	0.170	0.244	**0.087**	SAB13
IV-T0819-0	0.20	**0.260**	0.606	0.462	0.346	0.711	Linear
IV-T1201-0	0.20	0.455	**0.402**	0.406	0.409	0.407	SS14
TV-MIR02-0	0.20	0.431	0.306	**0.303**	0.336	0.306	ASK14
E-BUI-0	1.00	**0.352**	0.446	0.387	0.363	0.473	Linear
E-FRC-0	1.00	**0.127**	0.143	0.127	0.128	0.149	Linear
IT-AQV-0	1.00	**0.367**	0.380	0.367	0.367	0.387	Linear
IT-CLC-0	1.00	0.423	0.417	0.423	0.423	**0.416**	SAB13
IT-CLF-0	1.00	0.251	0.388	0.305	**0.140**	0.478	H20
IT-CLO-0	1.00	**0.474**	0.477	0.474	0.474	0.477	Linear
IT-CMI-0	1.00	**0.394**	0.429	0.394	0.395	0.452	Linear
IT-GSA-0	1.00	0.649	0.648	0.649	0.649	**0.645**	SAB13
IT-MOG0-0	1.00	**0.512**	0.531	0.512	0.513	0.544	Linear
IT-MRN-0	1.00	**0.361**	0.558	0.467	0.374	0.620	Linear
IT-NCB-0	1.00	**0.262**	0.302	0.262	0.263	0.315	Linear
IT-NOR-0	1.00	0.437	0.431	0.437	0.437	**0.425**	SAB13
IT-NRC-0	1.00	0.424	0.416	0.424	0.424	**0.410**	SAB13
IT-SNO-0	1.00	0.403	0.362	0.403	0.402	**0.346**	SAB13
IV-T0814-0	1.00	**0.167**	0.178	0.167	0.167	0.197	Linear
IV-T0818-0	1.00	**0.291**	0.427	0.326	0.300	0.477	Linear
IV-T0819-0	1.00	**0.889**	0.966	0.936	0.904	1.002	Linear
IV-T1201-0	1.00	0.510	**0.509**	0.510	0.510	0.516	SS14
TV-MIR02-0	1.00	0.440	0.430	0.441	0.439	**0.429**	SAB13

NGA-West 2 California

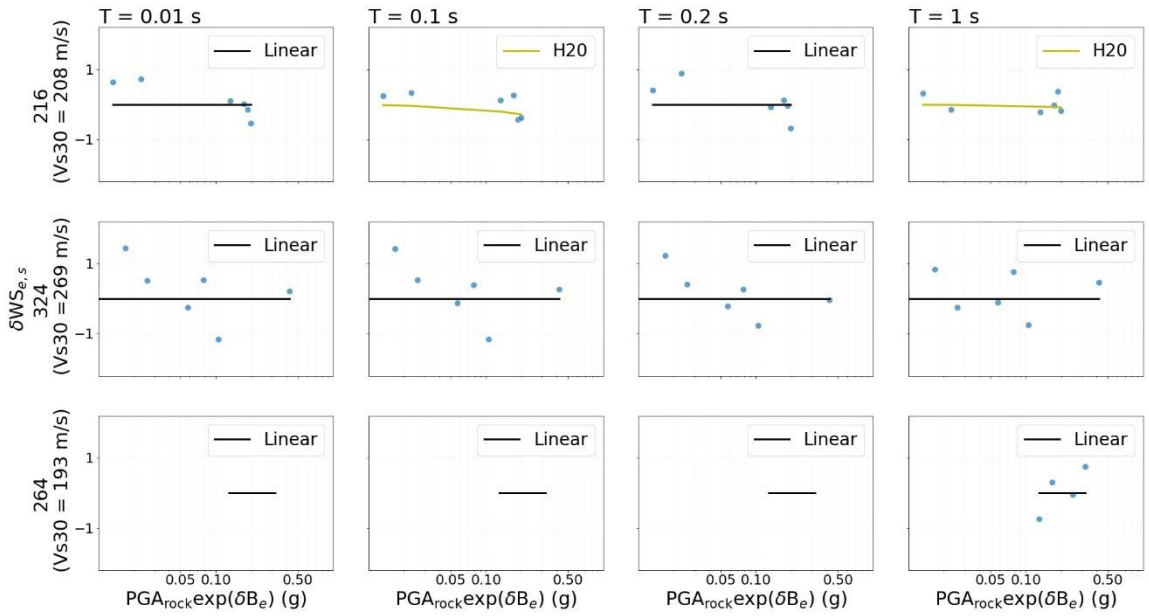


Figure S6: The NGA-West2 California stations with V_{S30} between 200-264 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

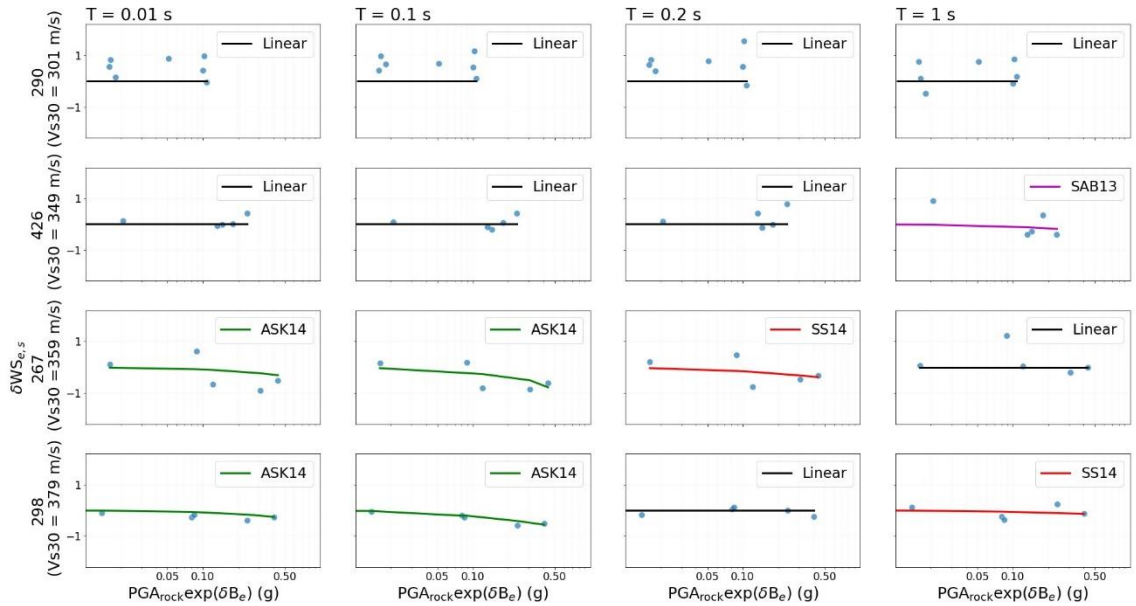


Figure S7: NGA-West2 California stations with V_{S30} between 300-379 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

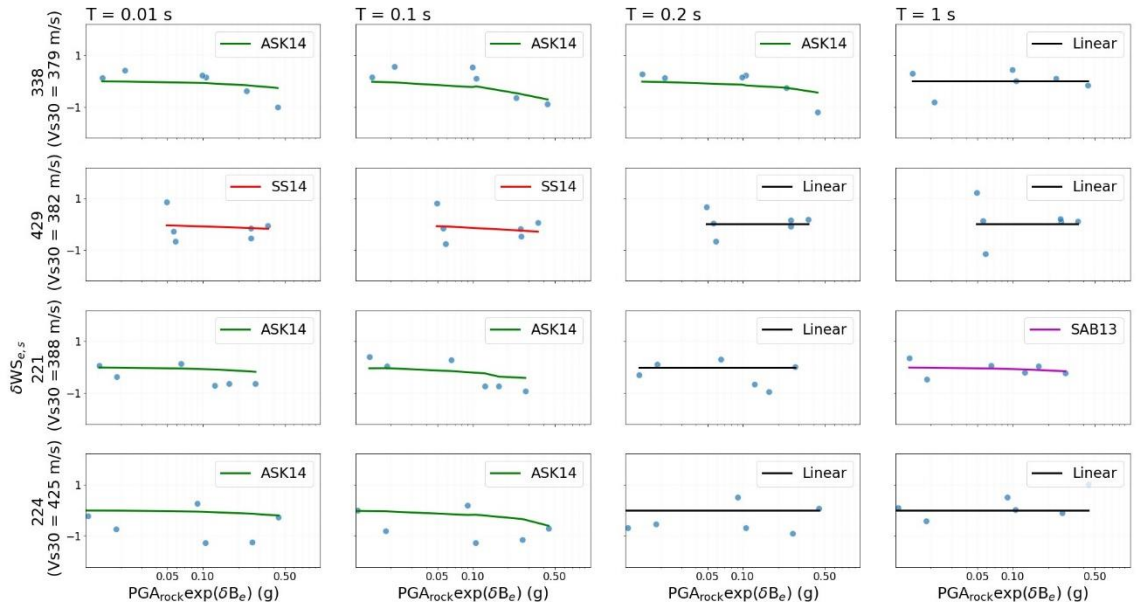


Figure S8: NGA-West2 California stations with V_{S30} between 379-425 m/s and the site-amplification model that scored best for each station. The site amplification model is compared to the site response as $\delta WS_{e,s}$ with respect to rock peak acceleration with event variability ($PGA_{rock}\exp(\delta B_e)$) for spectral periods T.

Table S2: The mean absolute error (MAE) score of all the amplification models with the ”left-over” residuals $\delta WS_{e,s}$ for each NGA-West2 California station at all periods T.

Stations	Periods	Linear	SS14	ASK14	H20	SAB13	Lowest model
216	0.01	**0.203**	0.316	0.301	0.211	0.334	Linear
221	0.01	0.531	0.458	**0.449**	0.455	0.457	ASK14
223	0.01	0.715	0.596	**0.579**	0.589	0.594	ASK14
224	0.01	0.759	0.694	**0.673**	0.678	0.687	ASK14
264	0.01	**nan**	nan	nan	nan	nan	Linear
267	0.01	0.673	0.560	**0.532**	0.565	0.553	ASK14
290	0.01	**0.582**	0.694	0.674	0.647	0.693	Linear
298	0.01	0.267	0.145	**0.130**	0.146	0.144	ASK14
324	0.01	**0.535**	0.607	0.633	0.585	0.604	Linear
337	0.01	0.552	0.500	**0.482**	0.502	0.494	ASK14
338	0.01	0.444	0.400	**0.375**	0.394	0.391	ASK14
426	0.01	**0.125**	0.239	0.241	0.217	0.237	Linear
429	0.01	0.341	**0.276**	0.283	0.279	0.280	SS14
1241	0.01	0.470	0.434	**0.426**	0.433	0.431	ASK14
216	0.10	0.297	0.454	0.720	**0.273**	0.357	H20
221	0.10	0.667	0.544	**0.451**	0.585	0.590	ASK14
223	0.10	0.895	0.697	**0.520**	0.760	0.772	ASK14
224	0.10	0.817	0.704	**0.585**	0.730	0.743	ASK14
264	0.10	**nan**	nan	nan	nan	nan	Linear
267	0.10	0.616	0.434	**0.372**	0.489	0.493	ASK14
290	0.10	**0.623**	0.826	0.953	0.722	0.761	Linear
298	0.10	0.391	0.192	**0.076**	0.260	0.266	ASK14
324	0.10	**0.486**	0.617	0.697	0.558	0.563	Linear
337	0.10	0.610	0.520	**0.413**	0.541	0.543	ASK14
338	0.10	0.547	0.475	**0.362**	0.486	0.493	ASK14
426	0.10	**0.201**	0.277	0.438	0.226	0.230	Linear
429	0.10	0.324	**0.279**	0.319	0.282	0.282	SS14
1241	0.10	0.434	0.404	**0.401**	0.411	0.412	ASK14
216	0.20	**0.230**	0.447	0.555	0.296	0.529	Linear

221	0.20	**0.479**	0.481	0.482	0.483	0.484	Linear
223	0.20	0.697	0.499	**0.477**	0.580	0.509	ASK14
224	0.20	**0.539**	0.550	0.567	0.549	0.555	Linear
264	0.20	**nan**	nan	nan	nan	nan	Linear
267	0.20	0.504	**0.354**	0.377	0.391	0.357	SS14
290	0.20	**0.767**	0.892	0.922	0.807	0.891	Linear
298	0.20	**0.107**	0.176	0.192	0.116	0.167	Linear
324	0.20	**0.318**	0.411	0.465	0.372	0.409	Linear
337	0.20	0.305	0.301	0.319	**0.273**	0.289	H20
338	0.20	0.465	0.392	**0.366**	0.412	0.382	ASK14
426	0.20	**0.334**	0.509	0.518	0.403	0.499	Linear
429	0.20	**0.232**	0.352	0.364	0.290	0.350	Linear
1241	0.20	0.474	0.444	0.450	0.453	**0.441**	SAB13
216	1.00	0.200	0.361	0.191	**0.190**	0.436	H20
221	1.00	0.131	0.118	0.131	0.130	**0.109**	SAB13
223	1.00	0.511	0.475	0.511	0.510	**0.463**	SAB13
224	1.00	**0.408**	0.433	0.408	0.408	0.453	Linear
264	1.00	**0.463**	0.743	0.564	0.512	0.765	Linear
267	1.00	**0.363**	0.407	0.363	0.365	0.431	Linear
290	1.00	**0.467**	0.527	0.467	0.471	0.555	Linear
298	1.00	0.237	**0.213**	0.237	0.235	0.236	SS14
324	1.00	**0.522**	0.581	0.553	0.529	0.600	Linear
337	1.00	0.343	0.242	0.343	0.337	**0.207**	SAB13
338	1.00	**0.175**	0.195	0.175	0.175	0.224	Linear
426	1.00	0.356	0.302	0.356	0.353	**0.285**	SAB13
429	1.00	**0.345**	0.413	0.345	0.347	0.445	Linear
1241	1.00	0.287	0.264	0.287	0.284	**0.253**	SAB13