Loviknes, K., Kotha, S. R., Cotton, F., Schorlemmer, D. (2021): Testing Nonlinear Amplification Factors of Ground-Motion Models. - Bulletin of the Seismological Society of America, 111, 4, 2121-2137.

https://doi.org/10.1785/0120200386

Testing non-linear amplification factors of ground-motion models

Karina Loviknes[1*, 2†], Sreeram Reddy Kotha[3‡], Fabrice Cotton[1, 2], Danijel Schorlemmer[1]

Corresponding author: karina.loviknes@gfz-potsdam.de
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14467 Potsdam, German

Declaration of Competing Interests: The authors acknowledge there are no conflicts of interest recorded.

*[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14467 Potsdam, Germany

†[2]University of Potsdam, Potsdam, Germany

‡[3]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, 38000 Grenoble, France

Abstract

We explore non-linear site effects in the new Japanese ground-motion dataset compiled by Bahrampouri et al. (2020). Following the approach of Seyhan and Stewart (2014), we evaluate the decrease of soil amplification according to the increasing and corresponding ground motion on surface rock ($V_{S30} = 760\,\mathrm{m/s}$). To better predict the rock ground motion associated to each record we take into account the between-event variability of the ground motion and to better evaluate the impact of non-linearity we correct observed ground motion on soil by the site-specific linear amplification. Instead of grouping the stations by site-response proxy, we focus on individual stations with several strong-motion records. We develop a framework to test recently published non-linear site-amplification models against a linear site-amplification model and also compare the results to recent building codes where non-linearity is included. The results show that the site response varies greatly from site to site, indicating that conventional site proxies, like V_{S30}, are not sufficient to characterise non-linear site response. Out of all the KiK-net stations, twenty stations are selected as having recorded sufficient data to be used in the test. Out of these twenty stations, five stations show signs of non-linearity, that is, the non-linear models performed better than the linear-amplification model for all periods T. For most sites, however, the linear site-amplification models get the best score. This suggest that, for the range of predicted rock motion considered in this study (PGA $< 0.2\,\mathrm{g}$), non-linearity may not have a sufficiently large impact on soil ground motion to justify the use of non-linear site terms in ground-motion functional forms and seismic building codes for such moderate-level shaking.

Introduction

Near-surface variations in the soil column have a strong effect on the input ground motions generated by earthquakes. The modulations of seismic ground motions by soil properties at a site are generally referred to as site effects. A weak ground motion stimulates only the linear response of the stiff soil column, i. e. the site-specific amplification is constant for any ground-motion intensity. However, for large ground motions and mostly for sites with soft soils, 1-D numerical simulations and several observations suggest that the site-specific amplification should decrease with increasing intensity of predicted ground motions for rock conditions (non-reference site method, Bonilla et al., 2005; Stewart et al., 2003; Field et al., 1997). 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, accompanied by a relative decrease in amplification at high frequencies (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. Observations related to non-linearity usually concern particular stations and earthquakes for which strong accelerations have been observed and most physics-based simulations of non-linearity have been developed to reproduce in a deterministic way these observations.

GMMs are designed to predict the probability of reaching a level (intensity) of ground motion given the earthquake-source properties (e. g. magnitude), the wave-propagation path (e. g. epicentral distance) and a simplified description of the geotechnical properties at the receiving site (site-response proxy).

GMMs need to predict the soil amplification associated to a reference "rock" ground motion at the surface. Such reference rock condition (usually defined by the time-averaged shear-wave velocity in the top 30 m of a 1-D soil column, $V_{S30} = 760\,\mathrm{m/s}$) is different from the borehole "hard"-rock conditions. A key challenge is to associate each observed soil record to its corresponding reference "rock" motion at the surface.

Calibration of such models therefore requires a large number of observations over the entire range of distances, magnitudes and acceleration levels. Such large datasets have not been available until now and recent ground-motion models which have provisioned a non-linear site-amplification component (e.g. Boore et al., 2014; Abrahamson et al., 2014; Chiou and Youngs, 2014) are 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 (SS14 from here on) 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 the SS14 model by quantifying non-linearity as the gradient of decreasing site amplification with increasing predicted peak acceleration on rock.

Non-linear simulations, following the equivalent-linear approach, compensate for the lack of data, but may however create some unintentional biases. For example, the response of sedimentary basins to strong seismic motions are often based on linear, elastic solutions incorporating frequency-independent material properties. It has been shown (e.g. Kausel and Assimaki, 2002) that this procedure may attenuate excessively the high-frequency components of motion in the waves propagating through the medium.

Fully non-linear (time-domain) site-response approaches have become increasingly popular in the recent years. However, several studies (e.g. Zalachoris and Rathje, 2015; Shi and Asimaki, 2017; Kaklamanos and Bradley, 2018) have shown that even fully non-linear 1D site response calculations do not fully capture the response at sites in the Kiban-Kyoshin network in Japan (KiK-net, Okada et al., 2004).

These considerations motivate the testing of the non-linear amplification models developed in the last years. We use the recent ground-motion dataset compiled by Bahrampouri et al. (2020)(BEA20 from here on), which consists of ground-motion data from earthquakes recorded by the KiK-net network between 1997 and 2017.

Although site amplifications in Japan have been widely studied (e.g. Zhao et al., 2015; Stafford et al., 2017; Derras et al., 2017; Sandikkaya, 2019), the large BEA20 dataset offers new opportunities to explore site effects using well-recorded events and stations.

Using this new dataset, we build upon the approach of Seyhan and Stewart (2014) to explore the empirical evidence of non-linearity and test non-linear amplification models. Taking advantage of new techniques to quantify the event and site specifics in ground-motion models (e.g. Stafford, 2014; Kotha et al., 2018), we propose a few relevant changes to the approach to better estimate the non-linearity:

1. We modify the predicted rock-motion level (peak ground acceleration on rock) by taking into account the variability specific to each event, that is, the systematic deviation of ground-motion recordings from a single event with respect to the predicted GMM median. Considering that earthquakes of the same magnitude can trigger very different levels of ground motions (Bindi et al., 2018, 2019), this modification better approximates the predicted ground motion on rock by including the effects specific to each earthquake (the event specific variability). This has also been done by Chiou and Youngs (2014) and Sandikkaya and Dinsever (2018).

2. To better investigate the non-linear site effects, we develop a ground-motion model taking into account the linear site amplification. For each observation, we then compute the difference between the observed and predicted linear ground motion. Sites with linear behavior will show residuals equal to zero for the full range of predicted rock PGA (peak ground acceleration).

3. While the gradient of SS14 is calibrated by mixing data from several sites with similar V_{S30}, we are able to investigate individual KiK-net sites that have recorded a high number of ground-motion records of earthquakes that likely have triggered non-linear soil responses. This follows the general understanding that linear and non-linear soil responses are rather site-specific, and poorly captured by V_{S30} alone (Derras et al., 2016; Thompson and Wald, 2016).

An additional goal of this paper is to develop a new GMM testing framework. We follow the vision of the Collaboratory for the Study of Earthquake Predictability (CSEP) which has developed transparent and reproducible tests of seismicity models with community-agreed testing methods, procedures and protocols, and has improved investigation on the quality of the model input data (Schorlemmer et al., 2018). For ground-motion models, although a large uncertainty in seismic-hazard assessment, testing procedures are still under development (Mak et al., 2017). Nevertheless, with upcoming new and large ground-motion observation datasets (e.g. BEA20, ESM dataset by Lanzano et al., 2019), the testing of ground-motion models can be addressed with new approaches. We build a testing procedure which allows a regular and reproducible update of the test to evaluate if the non-linear site effects predicted by existing site-amplification models are supported by high-quality empirical data. In addition to the SS14 site-amplification model, we test the non-linear site-amplification models of Abrahamson et al. (2014), Hashash et al. (2020) and Sandikkaya et al. (2013). These models are tested against a linear site amplification model.

Non-linear site response is also relevant to seismic building codes; examples are the NEHRP 2015 provisions for the US (Building Seismic Safety Council (BSSC), 2015), where site-response adjustment factors are based on SS14, and the German building code (German National Annex to Eurocode 8, Grünthal et al., 2018; Schwarz et al., 2017). The predicted rock ground motions in this study are in the same range as predictions for moderate seismic-hazard countries, hence we finally compare our results with the non-linear site amplifications proposed by the building codes of NEHRP and the German National Annex to Eurocode8.

Dataset and data selection criteria

We use ground-motion records from the Japanese Kiban-Kyoshin network (KiK-net, Okada et al., 2004) compiled into the BEA20 dataset by Bahrampouri et al. (2020). This dataset, which is the updated version of the Dawood et al. (2016) dataset, contains all KiK-net ground-motion record of earthquakes of magnitude $M_{\mathrm{W}} \geq 3$ between 1997 and 2017, which includes large events triggering non-linear soil response like the Kumamoto sequence in 2016. Compared to Dawood et al. (2016), who used earthquakes of magnitudes $M_{\mathrm{W}} \geq 4$ recorded until 2011, the number of records consequently increases from 147,282 to 351,658. Each ground-motion record in BEA20 is the outcome of an automatic processing protocol including filtering, baseline correction, and calculation of intensity measures (Bahrampouri et al., 2020). BEA20 provides a comprehensive event- and site-catalogue and several intensity measures, including spectral accelerations and Fourier amplitudes for both surface and borehole recordings. In this study, we focus on the spectral accelerations recorded at the surface. We selected a subset of the data using the following data selection criteria:

- For each record, there is an associated usable frequency bandwidth which limits the data we can use at period T. Because the usable frequency bandwidth is derived in the automatic processing, Bahrampouri et al. (2020) recommend only using records that qualify the $T \leq 1/(1.25 f_{\mathrm{low}})$ criteria in regression and residual analyzes at period T, where f_{low} is the low-pass cut-off corner frequency.

- Following the recommendation of Bahrampouri et al. (2020), we only use ground-motion records for which the usable frequency bandwidth is more than 60% of the range from zero to the Nyquist frequency. In addition, we omit ground-motion records that are flagged as having multiple wave-train arrivals.

- We only use events with depth $\leq 35\,\mathrm{km}$, recorded at $R_{\mathrm{JB}} < 600\,\mathrm{km}$. We also omit all offshore events and events whose closest recording is farther than $R_{\mathrm{JB}} = 80\,\mathrm{km}$. This is to avoid biasing the GMM median with wave propagation in oceanic crust and from possible interface events.

- We remove all events and stations with three or fewer recordings after all selection criteria are applied to ensure that the event- and site-term is derived on more than one record.

An overview of the selected stations is shown in Figure 1 and a map of Japan with the selected stations and events is given in Figure S1 in the Supplementary Material. The BEA20 dataset in comparison with the dataset of Dawood et al. (2016) and the NGA-West2 database are shown in Figure S2 in the Supplementary Material. BEA20 contains several well-recorded events and stations with many records (Figure 1b), including records in the range of non-linear soil behavior (Figure 1e); we therefore consider this dataset as an opportunity to investigate empirical site properties.

To evaluate the effects of the data selection criteria and to ensure that the final result does not depend on the assumptions made, we conduct a sensitivity test, described in the Result section.

Non-linear models

In this study, we test the non-linear component of four different site-amplification models. In addition to the aforementioned SS14, we test the amplification models by Sandikkaya et al. (2013) and Hashash et al. (2020), named SAB13 and H20 respectively, and the site-amplification model in the GMM of Abrahamson et al. (2014) (ASK14).

SS14 was developed based on empirical observations from the NGA-West2 database, appended with data from the 1-D numerical simulations of Kamai et al. (2014). SS14 consists of two additive components; a linear and a non-linear site term, and was developed for the GMM of Boore et al. (2014) (BSSA14). The site-amplification model of the ASK14 GMM was derived as a specific component of their GMM and also based on simulations of Kamai et al. (2014). SAB13 is based entirely on empirical data from the SHARE (Seismic Hazard HARmonization in Europe, Yenier et al., 2010) database and developed for pan-European GMMs. H20 was developed for central and eastern North America using 1-D site-response simulations from Harmon et al. (2019). These site-amplification models are all functions of V_{S30} and peak ground acceleration on rock ($\mathrm{PGA_{rock}}$). An overview of the models is provided in Table 1.

Method

Evaluation of the Seyhan and Stewart (2014) approach using the new dataset

Seyhan and Stewart (2014) analyzed non-linearity in the NGA-West2 dataset and calibrated their 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. As a first step of our analysis, we repeat the same procedure as Seyhan and Stewart (2014) on the BEA20 dataset to evaluate how the site amplification depends on the predicted rock ground motion.

In a typical engineering ground-motion dataset, each ground-motion observation of an event e at a recording site s is associated with the event's moment magnitude, M_{W}, the Joyner-Boore distance of the event from the recording site, R_{JB}, and the time-averaged shear-wave velocity of the top 30 m of the soil column at the recording site, V_{S30}. The ground motions are represented as period-depended spectral accelerations for periods T, $\mathrm{SA}(T)$, where the peak ground acceleration (PGA) is at $T = 0\,\mathrm{s}$.

First, for each selected record in BEA20, we use BSSA14 to predict ground motions for the record's specific M_{W}, R_{JB} and V_{S30}. We obtain the predicted ground motion on rock by setting $V_{\mathrm{S30}} = 760\,\mathrm{m/s}$, representing the V_{S30} at rock sites.

In a second step, and following Seyhan and Stewart (2014), we analyze the residuals between the observation $Y_{e,s}$ and the prediction on rock $\mu_{e,s}$ for an event e at a site s to quantify the amplification due to site effects. We subtract the rock prediction from the corresponding observation to obtain the total residual $\epsilon_{e,s}$:

$$\epsilon_{e,s} = \ln Y_{e,s} - \ln \mu_{e,s} \tag{1}$$

The total residual $\epsilon_{e,s}$ is split using the mixed-effects regression algorithm of

Bates et al. (2015), as done by e. g. Stafford (2014); Kotha et al. (2016, 2018). 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 (for example magnitude and distance) and the random effects are the grouping factors (for example event and site) (Bates et al., 2015). Here, the random effects of the events and sites are quantified by the mixed effects regression into the event and site variability:

$$\epsilon_{e,s} = \delta \mathrm{B}_e + \delta \mathrm{S2S}_s + \delta \mathrm{WS}_{e,s} \tag{2}$$

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

The within-event $\delta \mathrm{W}_{e,s}$ residuals are then obtained by subtracting the event-term from the total residuals, following Equation 5 in Seyhan and Stewart (2014) using the notation of Al Atik et al. (2010):

$$\delta \mathrm{W}_{e,s} = \ln Y_{e,s} - [\mu_{\mathrm{rock,e,s}} + \delta \mathrm{B}_e] = \epsilon_{e,s} - \delta \mathrm{B}_e \tag{3}$$

In the final step, we group the within-event $\delta \mathrm{W}_{e,s}$ residuals into V_{S30}-bins and investigate non-linearity as the trend of $\delta \mathrm{W}_{e,s}$ with respect to $\mathrm{PGA}_{\mathrm{rock}}$, see Figure 2. In this figure, the full SS14 site-response model including both linear and non-linear site-amplification components, is compared to the trend of the within-event residuals $\delta \mathrm{W}_{e,s}$ against $\mathrm{PGA}_{\mathrm{rock}}$. The non-parametric trend is given by the locally-weighted linear-regression fit (LOWESS), an algorithm developed by Cleveland (1979) for smoothing scatter plots using polynomial fit and weighted least squares. The residuals are calculated using BSSA14 on the BEA20 dataset. Figure 2 also shows that for BEA20, representing a dataset independent of the calibration dataset, SS14 does not quite follow the trend of the residuals. This misfit between the SS14 model and the KiK-net data, especially at short periods, was also reported by Seyhan and Stewart (2014).

This misfit may indicate that the linear site amplification as a function of V_{S30} is region dependent, as has also been discussed by e. g. Kotha et al. (2016). However, Sandikkaya (2019) have shown that regional variations in site amplification for Japan are likely to be insignificant. Other factors may explain those differences. Indeed, it has been shown that KiK-net stations have been installed at mostly stiff sites compared to K-net stations (Aoi et al., 2004; Zhu et al., 2020). Network installation strategies may therefore also have an impact on the dependency of amplification factors on V_{S30}.

These considerations motivate the computation of site-specific amplifications (site term) in the next paragraph and the use of site-corrected residuals to analyze the effects of non-linearity.

Proposed modifications to the Seyhan and Stewart (2014) approach

To better quantify the reference ($V_{\mathrm{S30}} = 760\,\mathrm{m/s}$) ground motion associated to each observed soil record we propose modifications to the approach of Seyhan and Stewart (2014). Restating the importance of between-event variability, it is now well known that moment magnitudes by itself cannot fully describe the complexity of earthquake rupture characteristics (e.g. Bindi et al., 2018, 2019), meaning that earthquakes of the same magnitude can produce very different levels of ground shaking.

We therefore propose to modify the Seyhan and Stewart (2014) approach by taking into account this between-event variability, by including the event-term $\delta\mathrm{B}_e$ in the prediction on rock. The non-linearity is then investigated in the residuals with respect to $\mathrm{PGA}_{\mathrm{rock}} \exp(\delta\mathrm{B}_e)$.

Seyhan and Stewart (2014) calibrated the linear and non-linear components of their site-amplification model as a function of V_{S30} from the within-event residuals. The within-event residuals, $\delta\mathrm{W}_{e,s}$, in any V_{S30}-bin contain features of both linear and non-linear site responses from several different sites. The site responses of different sites can be radically different from each other, even within the same V_{S30}-bin. We therefore suspect that the assumption that all sites in a V_{S30}-bin have similar linear and non-linear site responses, may bias the calibration of non-linear amplification models.

We propose to remove the site-specific linear site response in terms of $\delta\mathrm{S2S}_s$ from the within-event residuals. $\delta\mathrm{S2S}_s$ is arguably the most reliable representation of empirical site response, provided it is estimated from a sufficient number of site-specific ground-motion observations (Bard et al., 2020).

When $\delta\mathrm{S2S}_s$ is estimated from numerous small-magnitude events, it is safe to assume that it captures the site-specific linear soil response. Indeed, several studies (e.g. Pilz and Cotton, 2019; Thompson and Wald, 2016) have provided evidence that these terms may have also captured some of the more complex 2-D/3-D site effects at several KiK-net sites, which often evade the more prevalent numerical simulations based on simplified 1-D soil V_S-Q profile.

Therefore, to better investigate the non-linear site effects, we subtract the site-specific terms from the within-event residuals, leaving;

$$\delta\mathrm{W}_{e,s} - \delta\mathrm{S2S}_s = \epsilon_{e,s} - \delta\mathrm{B}_e - \delta\mathrm{S2S}_s = \delta\mathrm{WS}_{e,s} \tag{4}$$

This means that, for sites experiencing only linear soil response, the $\delta\mathrm{WS}_{e,s}$ should be symmetrically distributed around zero for the full range of $\mathrm{PGA}_{\mathrm{rock}}\exp(\delta\mathrm{B}_e)$.

Figure 3 shows the "left-over" $\delta\mathrm{WS}_{e,s}$ residuals with respect to $\mathrm{PGA}_{\mathrm{rock}}\exp(\delta\mathrm{B}_e)$ grouped into V_{S30}-bins for different periods T. The residuals are calculated using the BSSA14 GMM and the trend is fitted as before using the LOWESS algorithm of Cleveland (1979).

Instead of analysing non-linearity against within-event $\delta\mathrm{W}_{e,s}$ residuals with respect to $\mathrm{PGA}_{\mathrm{rock}}$ (as in Figure 2), we now analyze the non-linearity against the "left-over" $\delta\mathrm{WS}_{e,s}$ residuals with respect to $\mathrm{PGA}_{\mathrm{rock}}\exp(\delta\mathrm{B}_e)$ (Figure 3). Note that in Figure 3, only the non-linear site-amplification term of SS14 is plotted against the $\delta\mathrm{WS}_{e,s}$ residuals, while the full SS14 model is shown in Figure 2.

Comparing the corresponding residual trends between Figure 2 and 3, we see a significantly better agreement in the latter between the empirical residuals and the predictions of non-linear amplification by the SS14 model at all periods. This reasserts that both linear and non-linear site response are highly site-dependent and that V_{S30} is a poor site proxy for characterizing site amplification. With this understanding we propose that the site-specific linear site-amplification should be removed first (using $\delta\mathrm{S2S}_s$ as done here or other techniques) before investigating the residuals for non-linear site-amplification.

The modifications to the Seyhan and Stewart (2014) approach described above, should be generally applied when investigating non-linearity in new datasets. Furthermore, site amplification should be analysed using individual stations when possible, and stations grouped by V_{S30} should only be used when the database in study is not large enough to allow station-specific analysis. When grouping stations by site proxy the derived site amplification will be depended on the dataset.

Development of a linear ground-motion model

Due to the aforementioned challenges with V_{S30}, we develop a linear GMM using the same method and functional form as the GMM by Kotha et al. (2018) (K18 from here on). K18 was derived for the geometric mean of horizontal 5% damped pseudo-spectral acceleration (PSA):

$$\ln(\mathrm{PSA}) = f_{\mathrm{R}}(M_{\mathrm{W}}, R_{\mathrm{JB}}) + f_{\mathrm{M}}(M_{\mathrm{W}}) + \delta\mathrm{B}_e + \delta\mathrm{S2S}_s + \delta\mathrm{WS}_{e,s} \tag{5}$$

Here, $f_{\mathrm{R}}(M_{\mathrm{W}}, R_{\mathrm{JB}})$ is the magnitude-dependent distance scaling function, and $f_{\mathrm{M}}(M_{\mathrm{W}})$ is the magnitude scaling function, which are the fixed effects capturing the scaling of PSAs with distance and magnitude. As defined above, the event-term $\delta\mathrm{B}_e$ is the between-event random effect, the site term $\delta\mathrm{S2S}_s$ is the site-to-site or site-specific random effect, and $\delta\mathrm{WS}_{e,s}$ is the "left-over" residual capturing the record-to-record variability.

$\delta\mathrm{B}_e$, $\delta\mathrm{S2S}_s$ and $\delta\mathrm{WS}_{e,s}$ follow period-dependent normal distributions with standard deviation; τ, ϕ_{s2s} and ϕ_0, respectively:

$$\delta \mathrm{B}_e \sim \mathcal{N}(0, \tau), \tag{6}$$

$$\delta \mathrm{S2S}_s \sim \mathcal{N}(0, \phi_{s2s}) \tag{7}$$

and

$$\delta \mathrm{WS}_{e,s} \sim \mathcal{N}(0, \phi_0), \tag{8}$$

while the total aleatory variability is

$$\sigma = \sqrt{\tau^2 + \phi_{s2s}^2 + \phi_0^2}. \tag{9}$$

The distance and magnitude scaling of ground-motion observations is captured using mixed-effects regression (Bates et al., 2015) in multiple steps (see Kotha et al. (2018) for a step-by-step description). Unlike other contemporary models, K18 does not feature a fixed-effect site response based on V_{S30}. Instead, the $\delta\mathrm{S2S}_s$ captures all site-specific response and can be used as an empirical site-amplification function (Kotha et al., 2018).

Because the GMM is derived without any V_{S30} term, the predicted ground motion, PGA, is not specific to any site conditions. The predicted ground motion for rock conditions, $\mathrm{PGA}_{\mathrm{rock}}$, is therefore derived using a rock-adjustment term, α_{rock}, which is the mean $\delta\mathrm{S2S}_s$ for all sites with $V_{\mathrm{S30}} > 760\,\mathrm{m/s}$ (Kotha et al., 2018). The rock-adjustment term, α_{rock}, is added to the predicted ground motion, PGA, to obtain the predicted ground motion for rock conditions, $\mathrm{PGA}_{\mathrm{rock}}$:

$$\mathrm{PGA}_{\mathrm{rock}} = \mathrm{PGA} + \alpha_{\mathrm{rock}} \tag{10}$$

Whereas K18 is derived for the dataset by Dawood et al. (2016), we derive the new GMM from the updated BEA20 dataset while using the same method and functional form as K18. We mainly follow the same data selection criteria as used for K18 (see Dataset and data selection criteria section), but we adopt an additional criterion to omit records that might have triggered non-linear soil response at certain sites. This means that we omit all records with $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$ from stations with $V_{\mathrm{S30}} < 760\,\mathrm{m/s}$ (Guéguen et al., 2019; Régnier et al., 2013), and use the records that contain only linear soil response to derive the GMM. This is to avoid biasing the GMM median predictions and the estimates of the $\delta \mathrm{S2S}_s$ with non-linear soil response, making it a linear GMM.

The earlier omitted data, which was not used in GMM regression (from stations with $V_{\mathrm{S30}} < 760\,\mathrm{m/s}$ with $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$), will then be included in the dataset used to investigate non-linearity and to test the non-linear site-amplification models.

The residuals used to analyse non-linearity is derived in two steps; first the linear GMM is developed on the linear dataset, then the 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 non-linear soil response at stations with low V_{S30}. The residuals derived during the two processes are given in Figure S2 and S3 in the Supplementary Material.

To evaluate how well the linear GMM is developed, we analyse the distributions of $\delta \mathrm{B}_e$ with respect to magnitude, $\delta \mathrm{S2S}_s$ with respect to V_{S30}, and $\delta \mathrm{WS}_{e,s}$ with respect to distance (Figure S2). Both $\delta \mathrm{B}_e$ and $\delta \mathrm{WS}_{e,s}$ have a mean consistently close to zero and no clear trend with magnitude and distance, respectively. This confirms that the fixed effects components of the GMM regression has captured the scaling of magnitude and distance (Equation 5). $\delta \mathrm{S2S}_s$, however, have a downwards trend with V_{S30}. This is to be expected as we did not include a V_{S30} site-term in the fixed effects (Kotha et al., 2018).

To ensure that $\delta \mathrm{B}_e$ and $\delta \mathrm{S2S}_s$ appropriately quantify the event and site variability, they are only derived from events and stations with more than three records. This is applied both when deriving the linear GMM and when deriving the residuals for the non-linear site-amplification analysis and testing.

When the GMM only predicts the linear ground motion, we can use the corresponding site-term $\delta \mathrm{S2S}_s$ to correct for linear site response. The final $\delta \mathrm{WS}_{e,s}$ derived on the entire selected dataset will, by assumption, then contain any non-linear site effect. Figure 4 shows the differences between the response spectra and total aleatory variability σ of the original K18 GMM and the updated GMM derived here. The main differences between the GMMs, particularly in σ, are likely caused by the different magnitude ranges available in the two different datasets used to derive the models. The BEA20 database contains earthquakes of magnitude $M_{\mathrm{W}} \geq 3$, while Dawood et al. (2016) only included earthquakes of magnitudes $M_{\mathrm{W}} \geq 4$.

Testing of non-linear models

We built a procedure to test the predictions of non-linear site-amplification models given in Table 1 against residuals derived using the dataset and the linear GMM described in the previous sections.

Testing is largely a missing part of ground-motion modelling. The Collaboratory for the Study of Earthquake Predictability (CSEP) has developed advanced testing of seismicity models with community-agreed testing methods (Schorlemmer et al., 2018). The strategy of CSEP is to set an international standard for testing of earthquake forecasts and seismic hazard models. Tests within CSEP need to be transparent, reproducible and prospective, that is, tested against future observations. For earthquake forecast models, CSEP has long been developing and performed robust testing experiments (e.g. Schorlemmer et al., 2010; Zechar et al., 2013; Tsuruoka et al., 2012; Taroni et al., 2018). A future goal of CSEP is to develop testing procedures for GMMs and seismic hazard models. Although testing of ground-motion models exists (e.g. Delavaud et al., 2009; Mak et al., 2017; Lanzano et al., 2020), testing procedures and testable models are still in development. Ground-motion models with non-linear site terms have previously been tested by Guéguen et al. (2019) using strain as a proxy of non-linear soil response. Here we test the non-linear components of site-amplification models against the "left-over" residuals $\delta \mathrm{WS}_{e,s}$.

As described in the previous sections and by Seyhan and Stewart (2014), the residual, here $\delta \mathrm{WS}_{e,s}$, is assumed to contain the non-linear site response. The trend of $\delta \mathrm{WS}_{e,s}$ with respect to $\mathrm{PGA}_{\mathrm{rock}} \exp(\delta \mathrm{B}_e)$ (prediction on rock including event variability) can then be directly compared to non-linear site-amplification models. This is demonstrated in Figures 5, 6 and 7, where the non-linear site-amplification models are plotted with $\delta \mathrm{WS}_{e,s}$ against $\mathrm{PGA}_{\mathrm{rock}} \exp(\delta \mathrm{B}_e)$. We test the non-linear site-amplification models against a linear site-amplification model where the mean of $\delta \mathrm{WS}_{e,s} = 0$ for every value of $\mathrm{PGA}_{\mathrm{rock}} \exp(\delta \mathrm{B}_e)$. 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).

The MAE is a much used non-parametric score and chosen here for its simplicity. The score MAE_s is calculated for each site s:

$$\mathrm{MAE}_s = \frac{\sum_e^N \mid \delta \mathrm{WS}_{e,s} - F_{e,s} \mid}{N}, \tag{11}$$

where N is the number of events e recorded at site s, $\delta \mathrm{WS}_{e,s}$ is, as defined above, the "left-over" residuals assumed to contain the site-response and $F_{e,s}$ is the modelled site-amplification.

It is important to note that the MAE score only measures the deviation between the residuals and the predictions of the amplification models, and therefore does not have a direct physical meaning. The model with the best score is thus only best in a relative sense that the model with the lowest score has a smaller deviation from the residuals than a model with a higher score (Mak et al., 2015).

Because non-linear site effects are mainly expected, and modelled, for low $V_{\mathrm{S}30}$ and strong ground motions, we perform the test on a subset of stations with $V_{\mathrm{S}30} < 500\,\mathrm{m/s}$ and more than four records at predicted $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$. For the same reason, we only calculate the MAE score for ground-motions with $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$. The MAE scores of each amplification model at each of the selected stations are given in Tables 2 and S2. For each station, the model with the lowest score is evaluated as fitting the residuals best and is highlighted in bold.

Results

Following the derivation of the linear GMM, we apply the modified SS14 approach described in the previous sections to the selected dataset, including records from soft-soil sites with $\mathrm{PGA_{rock}} > 0.05\,\mathrm{g}$. We then use the residuals to test the non-linear component of the site-amplification models SS14, ASK14, SAB13 and H20. The test is performed for a subset of stations with $V_{\mathrm{S30}} < 500\,\mathrm{m/s}$ and more than four records at predicted $\mathrm{PGA_{rock}} > 0.05\,\mathrm{g}$.

Figure 5 shows how the non-linear amplification curves compare with the residuals for stations grouped by V_{S30}. Figures 6 and 7 show how the non-linear amplification curves compare with the residuals for individual stations. The decreasing amplification that is predicted by non-linear models is not emerging visually from the observed variability of amplifications when the stations are grouped according to V_{S30} (Figure 5) or when the stations with the largest number of high-acceleration records are selected (Figure 6).

Out of the 338 KiK-net stations with $V_{\mathrm{S30}} < 760\,\mathrm{m/s}$ and strong-motion records at $\mathrm{PGA_{rock}} > 0.05\,\mathrm{g}$ in the BEA20 dataset, twenty stations have recorded sufficient data to be selected for the testing exercise, that is, with more than four records at predicted $\mathrm{PGA_{rock}} > 0.05\,\mathrm{g}$. The data distribution of these stations is given in Figures S4 and S5 in the Supplementary Material. The scores of the models for each of the stations at $T = 0.01\,\mathrm{s}$ are given in Table 2, scores for all periods are given in Table S2 in the Supplementary Material. The amplification model that scores best for each station, that is the model with the lowest score, is highlighted in the tables.

Out of the 20 stations and the 4 periods used in the test, a non-linear amplification model has the lowest score 27 out of 80 times. Out of these, 5 stations have a non-linear model with the lowest score consistently for all periods T. These five stations, IBRH12, FKSH12, ISKH04, NARH01 and YMNH14, are shown on Figure 7.

However, none of the non-linear amplification models stand out with a better score significantly more times than the others or consistently for all periods.

The non-linear amplification models are therefore from here on mainly considered in comparison with the linear amplification model, and not with each other. Indeed, for most stations the linear amplification model gets the best score for the majority of the periods. That is, for 66.25% of the station-period combinations, the linear amplification model has the lowest score and therefore the best performance.

To ensure that the result does not depend on the data selection or assumptions made when deriving the linear GMM and residuals, we conduct a sensitivity test. The assumptions we have considered in the sensitivity test are:

- Only events with depth $\leq 35\,\mathrm{km}$, recorded at $R_{\mathrm{JB}} < 600\,\mathrm{km}$ are used.
- All events and stations with three or less recordings are removed after all selection criteria are applied.
- Non-linear soil behavior is only expected at $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$
  - Applied as cutoff for the dataset used to develop the linear GMM.
  - Used as criteria for the subset of stations the test is performed on.
  - Used as cutoff for the residuals the amplification models are tested on.
- The test is performed on a subset of stations with $V_{\mathrm{S30}} < 500\,\mathrm{m/s}$ where non-linear soil behavior is mainly expected.
- The subset of stations the test is performed on has at least four records at $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$.

We conduct the sensitivity test by varying these variables as described in Table 3, and evaluate how the ratio R of the number $N_{\mathrm{model},s,T}$ of stations s and periods T where each amplification model scores best, to the total number of stations and periods $N_{s,T}$:

$$R = \frac{N_{\mathrm{model},s,T}}{N_{s,T}} \tag{12}$$

Figure 8 shows the result of the sensitivity test. The ratio R is stable with depth, distance and minimum number of records used to derive the residuals. The most unstable assumption appears to be the threshold for expected non-linear soil behavior (Figure 8a, e and g). For the linear amplification model, the ratio of station-period pairs with the lowest score, stays between 0.6 and 0.8 using most criteria, but pass 0.8 when using $\mathrm{PGA}_{\mathrm{rock}} > 0.01\,\mathrm{g}$ as the non-linear range when deriving the linear GMM (Figure 8a). This is likely because more records with linear behavior are included in the test, giving the linear amplification model an advantage. However, increasing this threshold does not necessarily give the non-linear amplification models a higher ratio (Figure 8e and g), but rather decrease the number of stations used in the test. This is also the case when increasing the minimum number of records at $\mathrm{PGA}_{\mathrm{rock}} > 0.05\,\mathrm{g}$ when selecting stations for the test. These results are also supported by Figure S7 in the Supplementary Material.

In addition to varying the variables, we evaluate how the result depends on the different events by resampling the events randomly (bootstrapping), allowing one event to occur several times. This bootstrapping is performed 1000 times. The bootstrapping shows that the result has a larger variability when varying the events used for the test, with the linear amplification model having a larger variability of the ratio than the non-linear amplification models. This dependency of events is expected considering there is only a small number of events and records in the range where non-linear soil behavior is expected. The ratio R for each bootstrapping set is shown in Figure S8 in the Supplementary Material.

Although the sensitivity test and bootstrapping show that the result is generally stable, more stations are still needed to provide a ranking of the amplification models based on this test. However, the analysis shows that the complex models generally do not fit the data better than a simple linear model and that the result is stable with the different assumptions made during the data selection and in the test. Furthermore, the decrease in amplification trend predicted by the non-linear amplification models is not observed in the range of ground motions considered here. All the tests are easily reproducible with larger datasets.

Discussion

The main limitations of the test and our analysis of non-linear site effects in general, is the limited number of strong-motion records in the non-linear site-amplification range. Even with the new large BEA20 database, only 338 stations with $V_{S30} < 760\,\mathrm{m/s}$ are associated to rock peak ground acceleration larger than $0.05\,\mathrm{g}$ and only 20 stations have recorded a sufficient number of records (more than 4) in the range used in the test. Lack of data is, however, an universal issue and similar studies therefore mix stations with similar site proxies to obtain sufficient records to investigate non-linearity (e.g. Seyhan and Stewart, 2014; Guéguen et al., 2019). To improve this study, we plan the inclusion of K-net stations from the same network of KiK-net, where a high number of records with non-linearity have been reported (Chandra et al., 2016). Such a work is ongoing, but however beyond the scope of the present study (development of a transparent testing framework) because of the large amount of work needed to collect and homogeneously process K-net data.

Five stations show signs of non-linear soil behavior in that a non-linear amplification model had the best score at all periods T. These five stations all have a relatively high V_{S30} ($> 300\,\mathrm{m/s}$) considering that non-linearity is mainly predicted for low V_{S30}. Figure 7 shows that the non-linear amplification models do not predict a strong decrease in amplification for this level of V_{S30}. However, non-linear amplification can occur for sites with high V_{S30} when a high-velocity layer is covered by a shallow low-velocity layer, causing a high impedance contrast (Bonilla et al., 2011). This is in particularly relevant for KiK-net stations, which are mainly installed at stiff sites, either on weathered rock or on thin sediment layers (Aoi et al., 2004).

Along with the high variability of observed site response between the different stations, this suggests that the site proxy V_{S30} is not suitable for characterising non-linear site response. The site response of the five stations where a non-linear amplification model had the best score, should therefore be investigated further in future studies using alternative site proxies.

The dataset used in this study, despite its limitations, samples the range of rock ground motions which are considered in the seismic building codes of moderate seismic-hazard countries. Figure 9 shows that the ground motions on rock (PGA with a 475-year return period) taken into account by the French and German seismic hazard maps (Drouet et al., 2020; Grünthal et al., 2018) are fully consistent with the rock motions explored in this study (PGA $< 0.2\,\mathrm{g}$).

Recent building codes take non-linear site-amplification into account. Two examples are the NEHRP 2015 provisions for the US (Building Seismic Safety Council (BSSC), 2015) where the site-amplification factors were derived from the site-amplification model SS14. Another example is the German seismic building code, where a decrease in amplification is predicted for soft soil at PGA $> 0.1\,\mathrm{g}$ according to 1D non-linear simulations (Schwarz et al., 2017).

In Figure 10, we compare the site-amplification observed in our data with the non-linear site-amplifications proposed by the building codes of NEHRP and the EC8 German national annex. The variability of the observed response and the lack of obvious decrease of the observed amplification with increasing rock ground-motion trend, show that the data analysis does not yet confirm the need to take into account non-linear site effects for such moderate levels of ground motions on rock. For moderate seismicity areas and PGA $< 0.2\,\mathrm{g}$ the consideration of soil non-linearity to decrease amplification factors is then motivated by the results of non-linear site-response modelling or expert opinions, but is not confirmed yet by the statistical analysis of ground-motion data as done in this study. Here, the results show a high variability of responses and no evidence of a systematic decrease of soil amplification.

We also note that the models and building codes discussed in this testing exercise all represent ground motions in the response spectra domain. However, as discussed by e. g. Bayless and Abrahamson (2019) and Bora et al. (2016), the response spectra represents a response of a simple structure to the input ground motion and not directly the ground motion itself. Hence, there are some limitations to using response spectra when analyzing the physical properties of the ground motions, for example non-linear site response. Fourier spectra may therefore be a better option to analyze the physical factors controlling non-linearity, and the development and testing of amplification factors should in the future preferably be done in the Fourier domain.

Conclusion

In this study, we developed a transparent and reproducible framework to analyze non-linearity and test non-linear amplification models using residuals between ground-motion predictions and observations. We have applied this new framework on a new large dataset by Bahrampouri et al. (2020), building on the method of Seyhan and Stewart (2014). We estimate the reference (surface PGA at $V_{\mathrm{S30}} = 760\,\mathrm{m/s}$) rock motions that excites the soil column by taking into account the event-specific rock acceleration. The analysis is performed on individual stations with high number of strong-motion records, and thus avoiding mixing site responses of different sites. In order to reduce the bias of linear site amplification on calibration of non-linearity, we remove the linear site amplification using the site-specific terms of each station.

Most published site-amplification models predict a decrease of ground motion due to non-linear site effects for soft soils at accelerations above $0.05\,\mathrm{g}$. However, the actual soil data analyzed in this study do not confirm such a decrease, at least in the non-linear range $0.05\,\mathrm{g} < \mathrm{PGA} < 0.2\,\mathrm{g}$. The variability of observed amplifications remains large for such level of shaking and no systematic decrease of ground motions is observed. The testing procedure we developed shows that out of the twenty stations selected as having a sufficient number of records to be used in the test, only five stations exhibit signs of non-linearity at all periods. For most sites, however, the non-linear site-amplification models do not fit better with the observations than the linear amplification model. This does not mean that rheological non-linearities do not exist at that level of ground motion, but rather that for the level of predicted rock ground motions considered in this dataset, non-linearity is not significant enough to show a large impact on soil ground motions and justify the use of non-linear site terms in ground-motion functional forms and seismic building codes. Our results also show a large site-to-site variability which indicates that V_{S30} might not be the best site proxy to quantify linear and non-linear site-amplification models.

Data and Resources

The main dataset used was compiled by Bahrampouri et al. (2020) and downloaded from the DesignSafe repository: https://doi.org/10.17603/ds2-e0ts-c070 (last accessed September 2020). The mixed-effects regressions were performed using the LMER algorithm in the statistical software R: https://cran.r-project.org/web/packages/lme4/ (last accessed October 2020). An electronic Supplemental Material with tables and figures mainly complementing the data selection and tests conducted in the study, is available.

Acknowledgments

The authors are grateful to Mahdi Bahrampouri and Adrian Rodriguez-Marek for sharing their newly compiled dataset. The authors greatly appreciate the constructive and helpful feedback from Guest Editor Jim Kaklamanos, and reviewers Abdullah Sandikkaya and Emel Seyhan. In addition, the authors would like to thank Fabian Bonilla, Marco Pilz, Chuanbin Zhu, Graeme Weatherill and Elif Türker for valuable feedback and discussions in the early stages of this work. In particular, we want 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 Commission, ITN-Marie Sklodowska-Curie URBASIS-EU project, under grant Agreement 813137 and the 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.

References

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

Al Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components. *Seismol. Res. Lett.*, **81**, 794–801.

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

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

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

Bates, D., M. Mächler, B. M Bolker, and S. C. Walker (2015) Fitting linear mixed-effects models using lme4. *J.Stat.Software*, 1406–5823

Bayless, J., and N. A. Abrahamson (2019). An empirical model for the interfrequency correlation of epsilon for fourier amplitude spectra. *Bull. Seism. Soc. Am.*, **109**, 1058–1070.

Bindi, D., M. Picozzi, D. Spallarossa, F. Cotton, and S. R Kotha (2019). Impact of magnitude selection on aleatory variability associated with ground-motion prediction equations: Part II-analysis of the between-event distribution in central Italy. *Bull. Seism. Soc. Am.*, **109**, 251–262.

Bindi, D., D. Spallarossa, M. Picozzi, D. Scafidi, and F. Cotton (2018). Impact of magnitude selection on aleatory variability associated with ground-motion prediction equations: Part I—local, energy, and moment magnitude calibration and stress-drop variability in central Italy. *Bull. Seism. Soc. Am.*, **108**, 1427–1442.

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. Y., K. Tsuda, N. Pulido, J. Regnier, 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. 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**, 1057–1085.

Bora, S. S., F. Scherbaum, N. Kuehn, and P. Stafford (2016). On the relationship between Fourier and response spectra: Implications for the adjustment of empirical ground-motion prediction equations (GMPEs). *Bull. Seism. Soc. Am.*, **106**, 1235–1253.

Building Seismic Safety Council (BSSC) (2015). 2015a NEHRP Recommended Seismic Provisions for New Buildings and Other Structures (FEMA P-1050-

1): Volume 1: Part 1 Provisions, Part 2 Commentary, Federal Emergency Management Agency, Washington, D.C. 555 pp.

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

Chiou, B. S. J. and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, **30**, 1117–1153.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat Assoc.*, **74**, 829–836.

Dawood, H. M., A. Rodriguez-Marek, J. Bayless, C. Goulet, and E. Thompson (2016). A flatfile for the KiK-net database processed using an automated protocol. *Earthquake Spectra*, **32**, 1281–1302.

Delavaud, E., F. Scherbaum, N. Kuehn, and C. Riggelsen (2009). Information-theoretic selection of ground-motion prediction equations for seismic hazard analysis: An applicability study using Californian data. *Bull. Seism. Soc. Am.*, **99**, 3248–3263.

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

Derras, B., P.-Y. Bard, and F. Cotton (2017). V_{S30}, slope, H_{800} and f_0: performance of various site-condition proxies in reducing ground-motion aleatory variability and predicting nonlinear site response. *Earth Planets Space*, **69**, 1–21.

Drouet, S., G. Ameri, K. Le Dortz, R. Secanell, and G. Senfaute (2020). A probabilistic seismic hazard map for the metropolitan France. *Bulletin of Earthquake Engineering*, **18**, 1865–1898.

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**, 599–602.

Grünthal, G., D. Stromeyer, C. Bosse, F. Cotton, and D. Bindi (2018). The probabilistic seismic hazard assessment of Germany –version 2016, considering the range of epistemic uncertainties and aleatory variability. *Bull. Earthquake Eng.*, **16**, 4339–4395.

Guéguen, P., L. F Bonilla and J. Douglas, (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. *Bull. Seism. Soc. Am.*, **109**, 178–186.

Harmon, J., Y. M. Hashash, J. P. Stewart, E. M. Rathje, and K. W. Campbell, W. J. Silva and O. Ilhan (2019). Site amplification functions for Central and Wastern North America – Part II: Modular simulation-based models. *Earthquake Spectra*, **35**, 815–847.

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

Kaklamanos, J. and B. A. Bradley (2018). Challenges in predicting seismic site response with 1d analyses: Conclusions from 114 KiK-net vertical seismometer arrays. *Bull. Seism. Soc. Am.*, **108**, 2816–2838.

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

Kausel, E. and D. Assimaki (2002). Seismic simulation of inelastic soils via frequency-dependent moduli and damping. *J. Eng. Mech.*, **128**, 34–47.

Kotha, S. R., D. Bindi and F. Cotton (2016). Partially non-ergodic region specific GMPE for Europe and Middle-East. *Bull. Earthquake Eng.*, **14**, 1245–1263.

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

Lanzano, G., L. Luzi, V. D'Amico, F. Pacor, C. Meletti, W. Marzocchi, R. Rotondi, and E. Varini (2020). Ground motion models for the new seismic hazard model of Italy (MPS19): selection for active shallow crustal regions and subduction zones. *Bull. Earthquake Eng.*. **18**, 3487-3516.

Lanzano, G., S. Sgobba, L. Luzi, R. Puglia, F. Pacor, C. Felicetta, M. D'Amico, F. Cotton, and D. Bindi (2019). The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics. *Bull. Earthquake Eng.*, **17**, 561–582.

Mak, S., R. A. Clements, and D. Schorlemmer (2015). Validating intensity prediction equations for Italy by observations. *Bull. Seism. Soc. Am.*, **105**, 2942–2954.

Mak, S., R. A. Clements, and D. Schorlemmer (2017). Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting. *Bull. Seism. Soc. Am.*, **107**, 949–965.

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 Space*, **56**, 15–28.

Pilz, M. and F. Cotton (2019). Does the One-Dimensional Assumption Hold for Site Response Analysis? A Study of Seismic Site Responses and Implication for Ground Motion Assessment Using KiK-net Strong-Motion Data. *Earthquake Spectra*, **35**, 883–905.

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

Sahakian, V., A. Baltay, T. Hanks, J. Buehler, F. Vernon, D. Kilb, and N. Abrahamson (2018). Decomposing leftovers: Event, Path, and Site Residuals for a Small-Magnitude Anza Region GMPE. *Bull. Seism. Soc. Am.*, **108**, 2478–2492.

Sandikkaya, M. A. (2019). On linear site amplification behavior of crustal and subduction interface earthquakes in japan: (1) regional effects, (2) best proxy selection. *Bull. Earthquake Eng.*, **17**, 119–139.

Sandikkaya, M. A., S. Akkar, and P. Y. Bard (2013). A nonlinear site-amplification model for the next pan-European ground-motion prediction equations. *Bull. Seism. Soc. Am.*, **103**, 19–32.

Sandikkaya, M. A. and L. D. Dinsever (2018). A site amplification model for crustal earthquakes. *Geosciences*, **8**, 264.

Schorlemmer, D., M. J. Werner, W. Marzocchi, T. H. Jordan, Y. Ogata, D. D Jackson, S. Mak, D. A. Rhoades, M. C. Gerstenberger, N. Hirata, M. Liukis, P. J. Maechling, A. Strader, M. Taroni, S. Wiemer, J. D. Zechar, and J. Zhuang (2018). The Collaboratory for the Study of earthquake predictability: Achievements and priorities. *Seismol. Res. Lett.*, **89**, 1305–1313.

Schorlemmer, D., J. D. Zechar, M. J. Werner, E. H. Field, D. D. Jackson, T. H. Jordan and the RELM Working Group (2010). First results of the Regional Earthquake Likelihood Models experiment. *Pure Appl. Geophys.* **167**, 8–9

Schwarz, J., C. Kaufmann, and L. Abrahamczyk (2017). Bauforschungsvorhaben im bauaufsichtlichen bereich: Seismische einwirkungen für die neue generation von erdbebenbaunormen: Gefährdungskonsistente und untergrundspezifische spektren für die aktualisierte erdbebenzonenkarte. *Abschlussbericht zum Forschungsvorhaben*, **246**, 85–221.

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

Shi, J. and D. Asimaki (2017). From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses. *Bull. Seism. Soc. Am.*, **107**, 1336–1355.

Stafford, P. J. (2014). Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models. *Bull. Seism. Soc. Am.*, **104**, 702–719.

Stafford, P. J., A. Rodriguez-Marek, B. Edwards, P. P Kruiver, and J. J. Bommer (2017). Scenario Dependence of Linear Site-Effect Factors for Short-Period Response Spectral Ordinates. *Bull. Seism. Soc. Am.*, **107**, 2859–2872.

Stewart, J. P., A. H. Liu, and Y. Choi (2003). Amplification factors for spectral acceleration in tectonically active regions. *Bull. Seism. Soc. Am.*, **93**, 332–352.

Taroni, M., W. Marzocchi, D. Schorlemmer, M. J. Werner, S. Wiemer, J. D. Zechar, L. Heiniger, and F. Euchner (2018). Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy. *Seismol. Res. Lett.*, **89**, 1251–1261.

Thompson, E. M. and D. J. Wald (2016). Uncertainty in VS30-based site response. *Bull. Seism. Soc. Am.*, **106**, 453–463.

Tsuruoka, H., N. Hirata, D. Schorlemmer, F. Euchner, K. Z. Nanjo, and T. H. Jordan (2012). CSEP testing center and the first results of the earthquake forecast testing experiment in Japan. *Earth Planets Space*, **64**, 661–671.

Villani, M. and N. A. Abrahamson (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. *Bull. Seism. Soc. Am.*, **105**, 2681–2695.

Yenier, E., M. A. Sandikkaya, and S. Akkar (2010). Report on the fundamental features of the extended strong motion databank prepared for the SHARE project. *Deliverable 4.1 of seventh framework programme project seismic hazard harmonization in Europe (SHARE), Ankara*, 34–44.

Zalachoris, G. and E. M. Rathje (2015). Evaluation of one-dimensional site response techniques using borehole arrays. *Journal of Geotechnical and Geoenvironmental Engineering*, **141**, 04015053.

Zechar, J. D., D. Schorlemmer, M. J. Werner, M. C. Gerstenberger, D. A. Rhoades, and T. Jordan, (2013). Regional Earthquake Likelihood Models I: First-order results. *Bull. Seism. Soc. Am.* **103**, 787–798.

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**, 2010–2022.

Zhu, C., G. Weatherill, F. Cotton, M. Pilz, D. Y. Kwak, and H. Kawase, (2020). An open-source site database of strong-motion stations in Japan: K-NET and KiK-net (v1.0.0). *Earthquake Spectra*. doi: 10.1177/8755293020988028.

List of Authors

1. Karina Loviknes
karina.loviknes@gfz-potsdam.de
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14467 Potsdam, Germany

2. Sreeram Reddy Kotha
sreeram-reddy.kotha@univ-grenoble-alpes.fr
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, 38000 Grenoble, France

3. Fabrice Cotton
fcotton@gfz-potsdam.de
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14467 Potsdam, Germany

4. Danijel Schorlemmer
ds@gfz-potsdam.de
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14467 Potsdam, Germany

Tables

Table 1: Overview of the non-linear site-amplification models which will be 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)

∗ Next-Generation Attenuation Relationships for Western US

† Seismic Hazard HARmonization in Europe

‡ Next Generation Attenuation Relationships for Central and Eastern North-America

Table 2: The mean absolute error (MAE) score of the amplification models with the "left-over" residuals $\delta\text{WS}_{e,s}$ for each station at the period $T = 0.01\,\text{s}$. The lowest score for each station-period pair is highlighted in bold.

Stations	Linear	SS14	ASK14	H20	SAB13	Best-scoring model
CHBH14	**0.435**	0.466	0.452	0.443	0.488	Linear
FKSH12	0.164	0.152	0.151	**0.148**	0.152	H20
FKSH14	0.486	0.475	**0.466**	0.467	0.488	ASK14
FKSH21	**0.426**	0.508	0.497	0.497	0.501	Linear
IBRH12	0.484	0.469	0.468	**0.462**	0.470	H20
IBRH13	**0.359**	0.390	0.385	0.380	0.388	Linear
ISKH04	0.434	0.428	**0.426**	0.426	0.427	ASK14
KMMH14	**0.263**	0.344	0.330	0.286	0.354	Linear
KMMH16	**0.727**	0.798	0.804	0.774	0.799	Linear
NARH01	0.745	**0.658**	0.674	0.679	0.665	SS14
NGNH18	**0.624**	0.694	0.686	0.688	0.687	Linear
NGNH29	**0.884**	0.918	0.914	0.928	0.912	Linear
NIGH06	**0.370**	0.400	0.394	0.392	0.398	Linear
NIGH09	**0.384**	0.414	0.412	0.424	0.410	Linear
NIGH11	**0.699**	0.791	0.789	0.785	0.787	Linear
NIGH13	**0.702**	0.720	0.717	0.725	0.717	Linear
SMNH01	**0.389**	0.410	0.412	0.418	0.409	Linear
SZOH35	**0.469**	0.626	0.612	0.501	0.712	Linear
TTRH07	**0.202**	0.263	0.269	0.265	0.263	Linear
YMNH14	0.536	**0.499**	0.501	0.506	0.499	SS14

Table 3: Overview of assumptions made during the data selection and the test, with the values used in the test (center column) and the values applied in the sensitivity test (right column).

Assumption	Value	Sensitivity test
Maximum depth	$35\,\mathrm{km}$	$15, 25, 35\,\mathrm{km}$
Maximum distance	$600\,\mathrm{km}$	$150, 300, 600\,\mathrm{km}$
Least number of records for each event and station	3	1, 3, 10
Threshold for $\mathrm{PGA_{rock}}$ where non-linearity is expected	$> 0.05\,\mathrm{g}$	$> 0.01, 0, 02, 0.05, 0.05^{*}, 0.1\,\mathrm{g}$
Threshold for V_{S30} where non-linearity is expected	$< 500\,\mathrm{m/s}$	$< 350, 400, 500, 550, 600, 760\,\mathrm{m/s}$
The minimum number of records at $\mathrm{PGA_{rock}} > 0.05\,\mathrm{g}$	4	1, 2, 3, 4, 5, 7, 10

$*$ threshold applied for borehole data

List of Figures

Figures

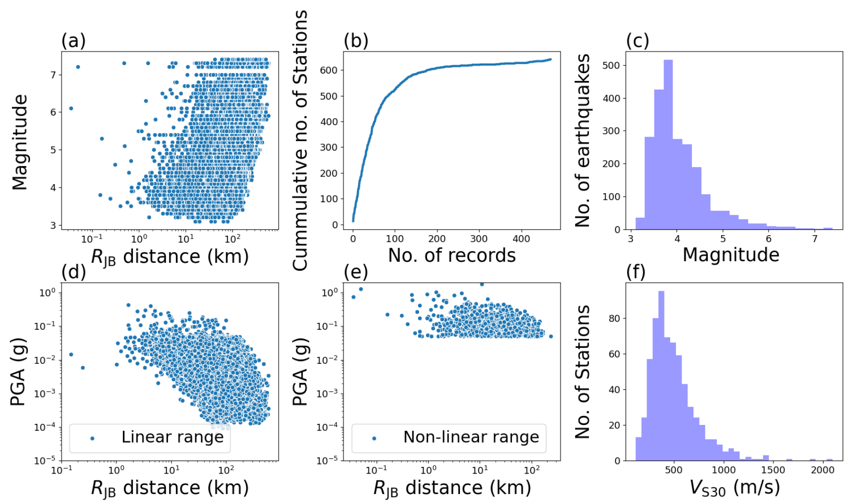

Figure 1: Data and site distribution of the selected records from the database of Bahrampouri et al. (2020) (BEA20). (a) PGA with distance (R_{JB}) and magnitude (M_{W}), (b) cumulative number of records per stations, (c) number of earthquakes per magnitude. (d) and (e) show the data defined as linear and non-linear, respectively, with PGA and distance. (e) Number of stations per V_{S30}.

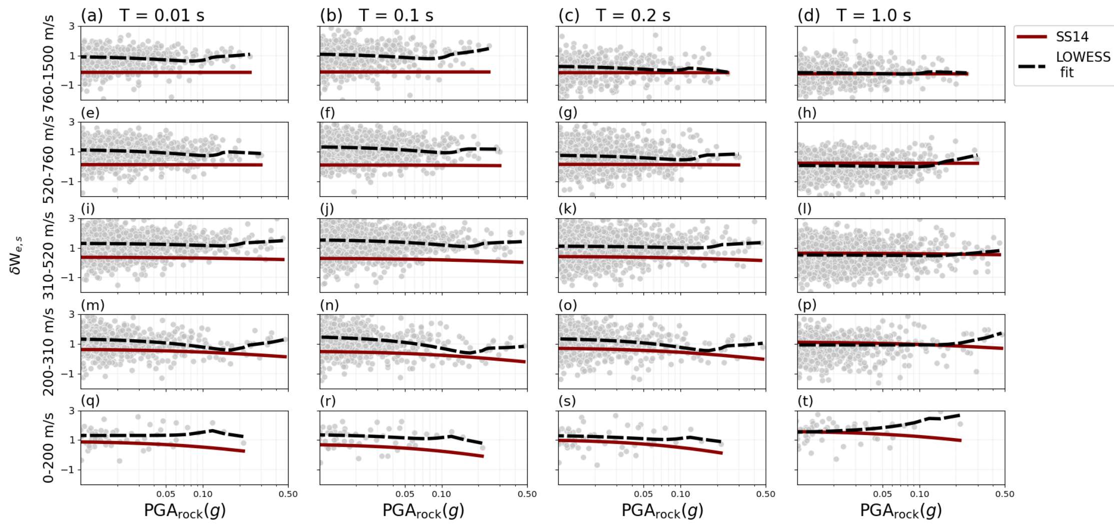

Figure 2: Variations of site-amplification given as within-event residuals ($\delta W_{e,s}$) with respect to peak acceleration for rock conditions (PGA_{rock}), following the approach of Seyhan and Stewart (2014) to investigate non-linearity. The within-event residuals $\delta W_{e,s}$ (y-axis) are grouped by V_{S30} (rows) for spectral periods T (columns), and derived using the Bahrampouri et al. (2020) database (BEA20) and Boore et al. (2014) GMM (BBS14). The range of each V_{S30}-bin is specified in the x-label. The dashed line is the LOWESS (locally weighted linear-regression) fit of the residuals and the bold line is the full SS14 site-amplification model, including both linear and non-linear site component. The large difference between the full Seyhan and Stewart (2014) model (SS14) and the trend of the residuals suggests that the relation between V_{S30} and site-amplification is site dependent.

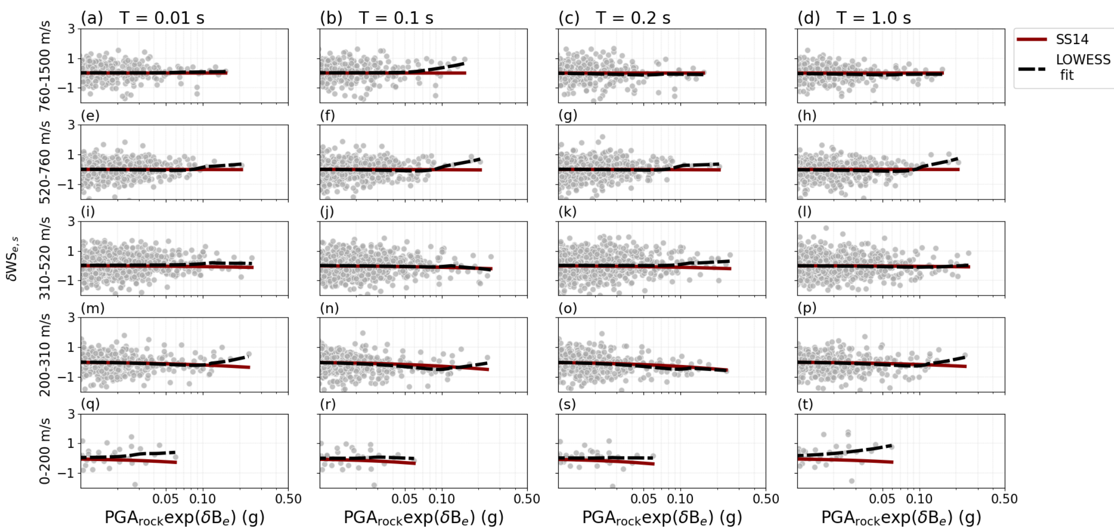

Figure 3: Variations of non-linear site-amplification given as "left-over" residuals ($\delta \mathrm{WS}_{e,s}$) with respect to rock peak acceleration with event variability ($\mathrm{PGA}_{\mathrm{rock}} \exp(\delta \mathrm{B}_e)$), after applying modifications to the approach of Seyhan and Stewart (2014) to investigate non-linearity. The "left-over" residuals $\delta \mathrm{WS}_{e,s}$ are grouped by V_{S30} (rows) for spectral periods T (columns), and derived using the Bahrampouri et al. (2020) database (BEA20) and Boore et al. (2014) GMM (BBS14). The range of each V_{S30}-bin is specified in the x-label. The dashed line is the LOWESS (locally weighted linear-regression) fit of the residuals and the bold line is the non-linear component of the Seyhan and Stewart (2014) site-amplification model (SS14). SS14 coincides better with the trend of the residuals after the site-specific linear site amplification (here using $\delta \mathrm{S2S}_s$) is removed.

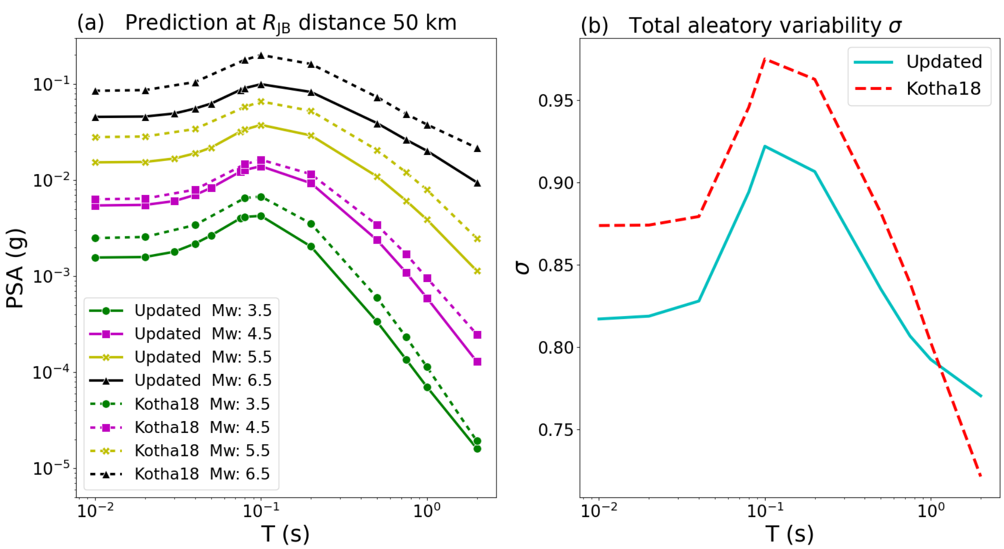

Figure 4: (a) The response spectra of Pseudo Spectral Acceleration (PSA) for the GMM of Kotha et al. (2018) (dashed lines) and the updated GMM derived here (bold lines) at $R_{\mathrm{JB}} = 50\,\mathrm{km}$ for different magnitudes. (b) The total aleatory variability σ of Kotha et al. (2018) (dashed line) and the updated GMM (bold line). The main difference between the two models are likely caused by the different magnitude ranges available in the two different datasets used to derive the models.

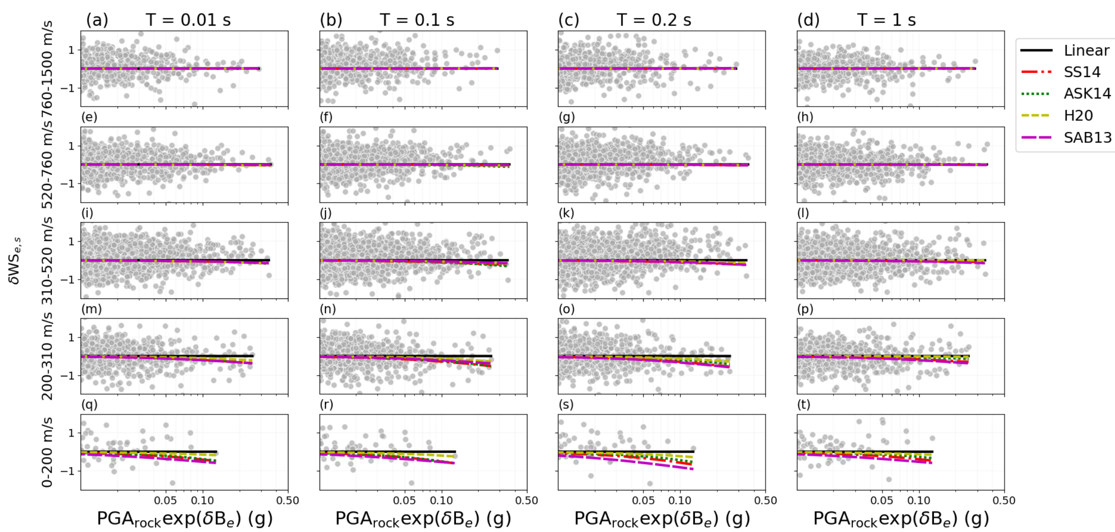

Figure 5: The site-amplification models compared to the "left-over" residuals $\delta\mathrm{WS}_{e,s}$ with respect to rock peak acceleration with event variability ($\mathrm{PGA}_{\mathrm{rock}} \exp(\delta\mathrm{B}_e)$) and stations grouped by V_{S30} (rows) for spectral periods T (columns). The range of each V_{S30}-bin is specified in the x-label. The residual plots show that the decreasing trend for low V_{S30} predicted by the non-linear site-amplification models is not observed in the residuals. The scores for each model and V_{S30}-bin are given in Table S1 in the Supplementary Material.

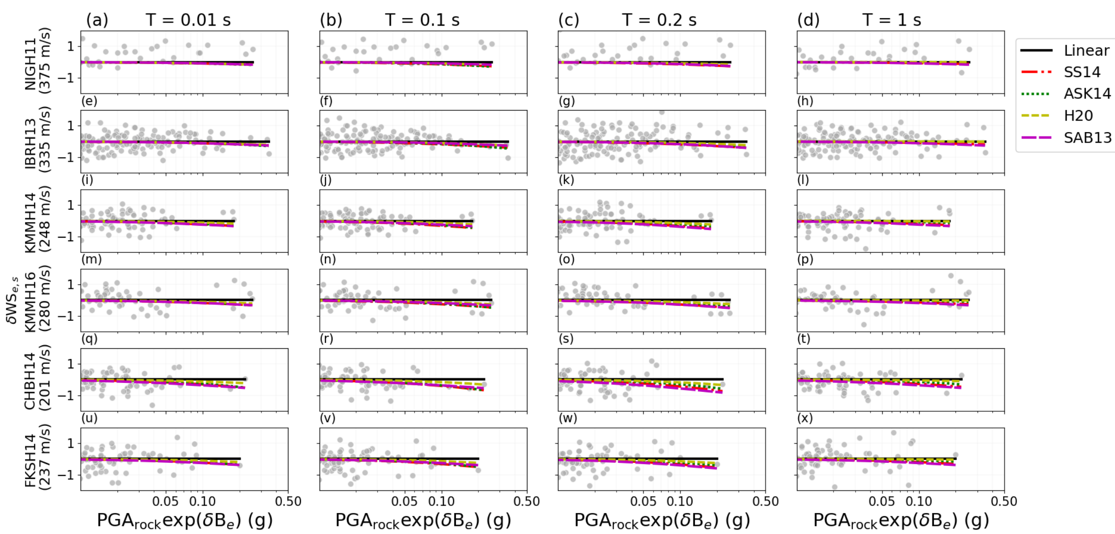

Figure 6: The site-amplification models compared to the "left-over" residuals $\delta \mathrm{WS}_{e,s}$ with respect to rock peak acceleration with event variability ($\mathrm{PGA}_{\mathrm{rock}} \exp(\delta \mathrm{B}_e)$) for spectral periods T (columns) and the selected KiK-net stations (rows) with the highest number of records. The V_{S30} of the stations are given in parenthesis under the station name in the x-label. The site-to-site variability of the residuals is high and the decreasing trend for low V_{S30} predicted by the non-linear site-amplification models is not observed in the residuals. The scores of each model for these stations at $T = 0.01\,\mathrm{s}$ are given in Table 2, and for all periods in Table S2 in the Supplementary Material.

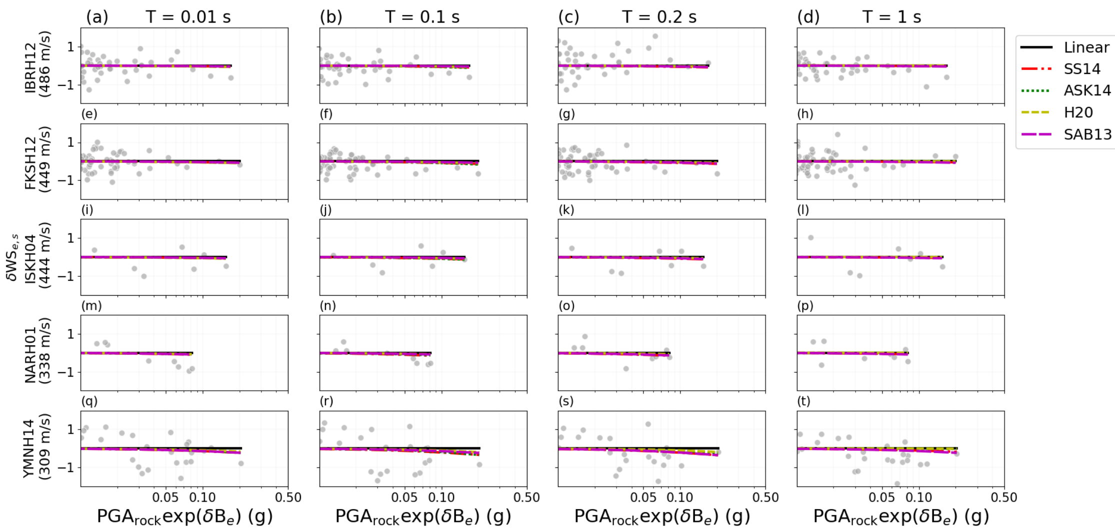

Figure 7: The site-amplification models compared to the "left-over" residuals $\delta\mathrm{WS}_{e,s}$ with respect to rock peak acceleration with event variability ($\mathrm{PGA}_{\mathrm{rock}}\exp(\delta\mathrm{B}_e)$) for spectral periods T (columns) and the selected KiK-net stations (rows) where the non-linear site-amplification models scored better than the linear site-amplification model at all periods. The V_{S30} of the stations are given in parenthesis under the station name in the x-label. The scores of each model for these stations at $T = 0.01\,\mathrm{s}$ are given in Table 2, and for all periods in Table S2 in the Supplementary Material.

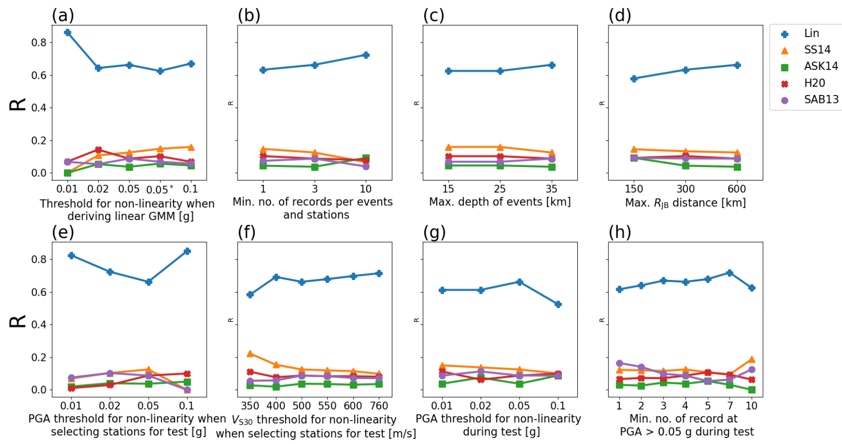

Figure 8: The ratio R of the number of stations and periods where each amplification model scores best, to the total number of stations and periods. Each subplot shows the variation of the ratio when varying the different assumptions made in the process of selecting the data and deriving the linear ground-motion model and residuals for the test. The assumptions and their values are also given in Table 3. The linear amplification model has the highest ratio for all assumptions, and the ratio of each amplification models is stable with depth, distance and minimum number of records used to derive the residuals. The ratio varies the most when applying different thresholds for expected non-linear soil behavior (a, e and g).

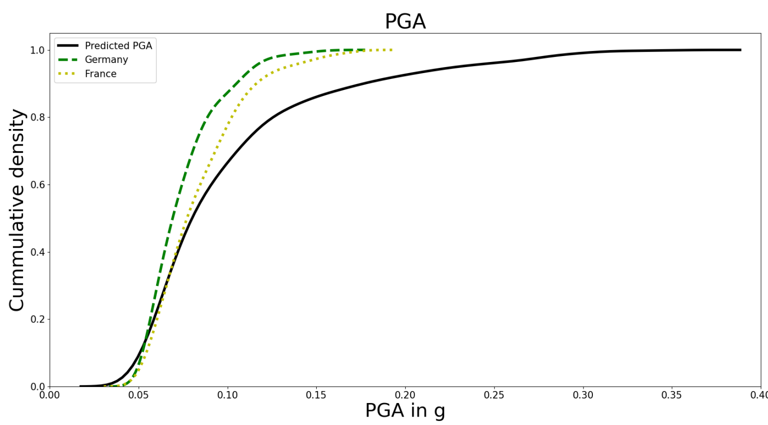

Figure 9: The cumulative density distribution of the predictions for rock conditions with PGA $> 0.05\,\mathrm{g}$ used in this study (bold line) and that of recent seismic hazard maps of Germany (dashed line) and France (dotted line) for PGA on rock with a 475-year return period. The PGA range used in this study is consistent with the predictions for Germany and France up to PGA $= 0.2\,\mathrm{g}$.

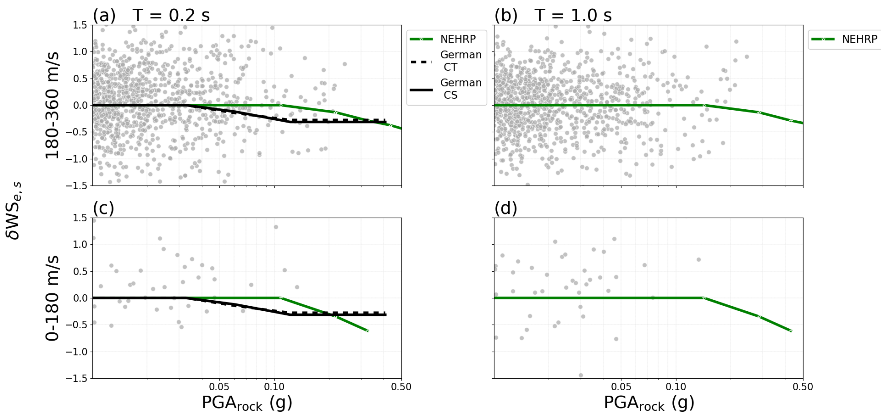

Figure 10: The site-amplification factor of the NEHRP and German seismic building codes compared to the "left-over" residual $\delta\mathrm{WS}_{e,s}$ with respect to rock peak acceleration with event variability ($\mathrm{PGA}_{\mathrm{rock}}\exp(\delta\mathrm{B}_e)$) and stations grouped by V_{S30} (rows) for spectral periods T (columns). The range of each V_{S30}-bin is specified in the x-label. The decrease of amplification predicted by the building-code site-amplification factors is not observed in the residuals.