Amplitude scaling of ground motions as a potential source of bias: Large-scale investigations on structural drifts

Konstantinos Theodoros Tsalouchidis | Christoph Adam

Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria

Correspondence
Christoph Adam, Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria.
Email: christoph.adam@uibk.ac.at

Abstract
The question of whether amplitude scaling of ground motion (GM) records introduces bias is controversial in earthquake engineering. In this study, this research question is formally defined with a focus on evaluating bias in determining an engineering demand parameter (EDP) as a result of nonlinear response history analyses (NLRHA) when using scaled rather than unscaled GM that have the same level of intensity. The analysed structures are 10 planar steel frame buildings ranging from low- to high-rise, where the EDP of interest is the maximum interstory drift ratio, while the structural responses range from linear to collapse. A unique contribution is the depth of the research, which employs an unprecedented number of more than 17,000 recorded GM, resulting in approximately 3.4 million NLRHA. For a thorough investigation, the most relevant intensity measures are discussed and considered, as well as novel spectral information describing the sustained vibration amplitude. The introduction of bias is examined from different points of view, using first simple and intuitive statistical methods, then machine learning techniques, and finally a novel GM selection approach. In the numerous investigations, no bias could be detected under the inherent uncertainty of the calculations. The results indicate that scaled records can be safely used in NLRHA to assess the seismic structural behaviour if spectral and scenario compatibility are ensured and it is verified that the sustained amplitude is also consistent.

KEYWORDS
amplitude scaling, bias, scale factor, nonlinear response, drift, ground-motion selection

1 | INTRODUCTION
Nonlinear response history analysis (NLRHA) is a powerful analysis tool for evaluating the seismic response of a structural model, typically using recorded ground motions (GM) as excitation and calculating the dynamic response of the model at each time step by solving the equations of motion. It is frequently used in engineering practice and academia, often as a benchmark for assessing simpler approaches, as it is considered the most accurate approach. Therefore, various leading frameworks for evaluating the behaviour of a structure under seismic excitation include a number of NLRHA, such as the incremental dynamic analysis (IDA), the cloud analysis and the multiple stripe analysis. This is also suggested...
by the popular and comprehensive performance-based methodology\textsuperscript{[7–9]} for seismic assessment, which will be the basis for the next generations of seismic codes.

The use of recorded GM is simple and intuitive, unlike alternatives such as simulated or physics-based GM. However, the scarcity of recorded GM is its main drawback and the reason for developing alternatives, especially in the case of severe earthquake events, which are particularly rare. To tackle this problem and use recorded GM corresponding to a desired level of intensity, (amplitude) scaling is commonly employed. Scaling is a uniform amplification (or shrinkage) of the recorded accelerations of a GM to increase (or decrease) their value by a constant scale factor $SF$. This is a common procedure generally used in signal processing and can be represented mathematically by the following equation,

$$a_{g,\text{scaled}}(t) = SF \cdot a_g(t)$$

where $SF$ is a scalar that determines the amplification ($SF > 1$) or shrinkage ($SF < 1$) of the recorded signal of accelerations $a_g$ of the GM, resulting in $a_{g,\text{scaled}}$.

Since amplitude scaling is easy to implement and effectively solves the problem of a specified GM excitation with the desired level of intensity, it is widely used in recent and older studies,\textsuperscript{[10,11]} as well as incorporated into design codes.\textsuperscript{[12,13]} In the IDA framework, for example, it is an essential component since a GM record is repeatedly scaled until structural collapse is reached, frequently requiring excessive scale factors. In multiple stripe analysis, scaling can be limited and used only to adjust a GM to the desired level of intensity. In cloud analysis it can be avoided altogether, at least in theory, if sufficient GM are available for the target intensities, which is rarely the case.

1.1 Questioning the validity of amplitude scaling and background

The application of scale factors to transform $a_g$ into $a_{g,\text{scaled}}$ inevitably raises the question of whether the inherent characteristics of the ‘as-recorded’ GM acceleration signal are altered in an unnatural way, leading to biased results when NLRHA is performed. This has been an ongoing debate for at least two decades\textsuperscript{[14]} From an earthquake engineering perspective, it is important to note that any potential differences between $a_{g,1,\text{scaled}}$ and $a_{g,2}$ are of interest if two conditions apply: (1) $a_{g,1,\text{scaled}}$ and $a_{g,2}$ are at the same intensity and (2) there is a systematic difference in an engineering demand parameter (EDP), i.e., a structural response of interest obtained by NLRHA for the specific structural model(s). If both parts of the above research question are true, then it can be concluded that bias is introduced when amplitude scaling is employed.

However, there are several pitfalls to consider.

The abstract notion that $a_{g,1,\text{scaled}}$ and $a_{g,2}$ are at the same level of intensity implies equality between one or more intensity measures (IMs), that is, metrics describing the intensity of the GM. The choice of IMs is of particular importance, and it is common in studies to use those that are employed in many applications (e.g., the spectral acceleration at the fundamental period of the structure, $Sa(T_1)$) or the entire list of IMs that are known to affect the specific EDP examined.

A proper selection of IMs to describe the ground excitation has been widely studied\textsuperscript{[10,15–19]} and depends strongly on various parameters such as the modal properties of the structural model, the material and the geometrical nonlinearities under consideration, the EDP of interest as well as the damage state of the structure.

Therefore, a conclusion that bias is present (or not) refers only to the specific EDP and IMs that are taken into account, while robustness is necessary and requires a wide consideration of GMs. This means that considering more or less IMs might have an impact on the conclusions and that these conclusions do not necessarily apply to other EDPs. Moreover, the above only holds for structural models similar to those studied and cannot be extended to a wider range of structures without justification.

The need for explicit studies on the legitimacy of amplitude scaling has long been recognised. Early work\textsuperscript{[4,20]} suggested that amplitude scaling does not introduce bias when its structure-specific scaling procedure is followed, and that differences in spectral shape between GM of unscaled and scaled records can explain any possible discrepancies in NLRHA. In the paper of Luco and Bazzurro,\textsuperscript{[14]} the introduction of bias was studied in terms of the nonlinear structural drift responses of multiple non-degrading single-degree-of-freedom (SDOF) systems of various periods and strength properties, as well as a 9-story steel structure modelled as a multi-degree-of-freedom (MDOF) system. One conclusion of this paper is that scaling can indeed lead to bias, but this can largely be explained by differences between the elastic response spectra of

\textsuperscript{1}Corresponding to a GM record rec1, scaled with $SF_1 \neq 1$.

\textsuperscript{2}Corresponding to an unscaled GM record rec2.
the scaled versus unscaled records.\textsuperscript{[14]} It was also pointed out\textsuperscript{[14]} that many researchers\textsuperscript{[21,22]} have approached the issue from a seismological perspective. That is, they only deal with the first part and focus on the systematic changes found in the GM characteristics (e.g., spectral shape or frequency content, duration), assuming that these differences influence the results of NLRHA. This assumption is not arbitrary, especially since systematic differences in important IMs (e.g., spectral shape) are observed when scaling is employed. Such studies shed light on these differences that tend to exist between GM records with different levels of intensities. However, it is interesting to ask whether, despite this tendency, compatibility of these IMs (e.g., spectral shapes) can be found and whether unbiased results can be obtained by NLRHA.

In this context, Baker\textsuperscript{[23]} proposed a method for detecting such scaling bias and concluded that scaling does not introduce bias when records are selected based on the GM parameter $\epsilon$ (or when the spectral shape is somehow taken into account), while significant bias is observed otherwise. These conclusions were based on the evaluation of maximum interstory drift ratios (MIDR) of a 7-story reinforced concrete (RC) moment frame building, with the structural responses involving linear and slightly nonlinear ranges, as most of the MIDR were less than 0.02. Other researchers\textsuperscript{[24–26]} have also concluded that spectral shape is critical in providing unbiased displacement responses. In particular, Hancock et al.\textsuperscript{[24]} studied the response of an 8-story MDOF RC structure when subjected to several suites of 25 records, Iervolino et al.\textsuperscript{[25]} examined displacement and cyclic response-related parameters of multiple SDOF systems, and Huang et al.\textsuperscript{[26]} used multiple bilinear SDOF systems with low to moderate inelastic behaviour.

Zacharenakiet al.\textsuperscript{[27]} investigated the full range of limit-states, including collapse, and qualitatively and quantitatively demonstrated the introduction of bias by comparing IDA and cloud analyses. They\textsuperscript{[27]} analysed six degrading SDOF systems of medium to long periods and two MDOF systems where P-\Delta effects were considered while their columns were modelled as elastic and quadrilinear springs at the beam ends captured the material nonlinearities. Analyses were performed on a set of 1015 recorded and 465 synthetic GM, without explicit consideration of spectral shape constraints in the record selection process.

More recent studies such as Davalos and Miranda\textsuperscript{[28,29]} argue that explicit consideration of spectral shape is not sufficient to avoid systematic overestimation of lateral displacement demands and collapse probabilities with increasing scaling. Specifically, bias was found to be associated with strength degradation in virtually all structural systems studied, while bias increases with decreasing period and decreasing lateral strength relative to the strength required to remain elastic.\textsuperscript{[28,29]} For this reason, other IMs were considered, such as input energy, causal parameters and damaging pulse distributions between unscaled and scaled GM sets and it was argued that differences in these IMs are the main reasons for the discrepancies in deformation demands from NLRHA.\textsuperscript{[28,29]} Similarly, Du et al.\textsuperscript{[30]} investigated the scaling limits for conditional spectrum-based GM selection and concluded that these limits have a significant impact on the distribution of the EDPs of two steel special moment frames. They recommended a scaling limit range of 3 to 5 for general use when selecting GM records from the NGA-West2 database.

1.2 Research objective and contribution

According to the requirements of the research question formally defined in this study, we investigate the introduction of bias through amplitude scaling when NLRHA of steel moment frame structures is performed to obtain the MIDR. Multiple building structures, that meet modern design standards and thus also account for seismic actions, are analysed here utilizing an unprecedented number of recorded GM (more than 17,000), resulting in a total of approximately 3.4 million NLRHA. The mechanical models of the buildings encompass the main phenomena that occur in a strong GM oscillation, such as the P-Delta phenomena, the cyclic and monotonic material deterioration at the beams and columns. The MIDR responses of interest range from linear to the collapse limit state, with the emphasis on the nonlinear and near-collapse limit states, where the discrepancies between scaled and unscaled analyses are likely to be larger.\textsuperscript{[14]}

First follows a discussion of the most important IMs, which are also considered here. Moreover, novel spectra are proposed as an intuitive, easy-to-implement, and an informative way to evaluate the sustained amplitude of the elastic SDOF oscillators, which are typically used to derive the response spectra. In this way, this study focuses on easy-to-calculate IMs and possible-to-accomplish spectral compatibility. At the same time, all expected parameters affecting the MIDR results are taken into account without resulting in computational overload. Moreover, extensive efforts are made to identify and eliminate sources of bias that potentially affect the conclusions if neglected.

The research question is examined from several angles, first by comparing clouds of unscaled analyses with those of scaled analyses without explicit consideration of spectral shape and scenario parameters. The comparisons of the clouds are performed utilizing simple and intuitive statistical methods as well as machine learning techniques. Then, spectral and scenario compatibility of scaled and unscaled sets of GM records is ensured by a recently developed GM selection
Table 1  Structural models and their first three periods of vibration

| Notation | Archetype ID | No. of stories | $T_1$ (s) | $T_2$ (s) | $T_3$ (s) |
|----------|--------------|----------------|-----------|-----------|-----------|
| Dmax4    | 3RSA         | 4              | 1.76      | 0.59      | 0.33      |
| Dmin4    | 9RSA         | 4              | 2.64      | 0.86      | 0.49      |
| Dmax8    | 4RSA         | 8              | 2.37      | 0.85      | 0.48      |
| Dmax8RBS | 4RSA         | 8              | 2.35      | 0.84      | 0.48      |
| Dmin8    | 10RSA        | 8              | 3.61      | 1.24      | 0.68      |
| Dmax12   | 5RSA         | 12             | 3.22      | 1.13      | 0.66      |
| Dmax12RBS| 5RSA         | 12             | 3.20      | 1.13      | 0.65      |
| Dmin12   | 11RSA        | 12             | 4.80      | 1.65      | 0.94      |
| Dmax20   | 6RSA         | 20             | 4.55      | 1.63      | 0.95      |
| Dmin20   | 12RSA        | 20             | 5.73      | 2.01      | 1.77      |

approach, which is adjusted here for the needs of this study. The increased efficiency and versatility of this GM selection method results in various GM sets having the same level of intensity, which is an essential requirement mentioned above. Through the unique depth of investigation, the advanced solutions to achieve spectral compatibility and the different approaches to address the research question, the current study clearly provides concrete results regarding the introduction of bias due to amplitude scaling.

2  STRUCTURAL SYSTEMS AND GROUND MOTION RECORDS UNDER INVESTIGATION

2.1  Building models

This study focuses on steel moment frame buildings ranging from low- to high-rise. Overall, two groups of 4-, 8-, 12- and 20-story planar moment frames are considered, which differ in their seismic design criteria. Detailed information on the structural systems and their suggested modelling can be found in chapter 6 of NIST GCR 10-917-8 report, \[31\] which is also followed in this study and is briefly explained below. The structural members of the frames are analytically described in the report, \[31\] designed in accordance with the strength design requirements in AISC 341-05 (2005)\[32\] and the seismic design requirements in ASCE/SEI 7-05 (2006), \[33\] while the criteria in AISC 358-05 (2005)\[34\] regarding the design of RBS connections were followed.

The first group, denoted “Dmax”, includes six buildings: one 4-story (“Dmax4”), two 8-story (“Dmax8” and “Dmax8RBS”), two 12-story (“Dmax12” and “Dmax12RBS”), and one 20-story (“Dmax20”). The second group, denoted “Dmin”, includes four buildings: one 4-story (“Dmin4”), one 8-story (“Dmin8”), one 12-story (“Dmin12”), and one 20-story (“Dmin20”), all of them shown in Table 1, which contains the notation for each building model used here, the archetype ID in the report, \[31\] the number of stories, and the first three periods of vibration.

The symmetrical plan view shown in Figure 1A is identical for all stories of all buildings. Four identical moment-resisting steel frames, peripherally to the plan, are responsible for the lateral capacity of the building. The blue area indicates the tributary area for gravity loads applied directly to one of the frames and the green area indicates the tributary area for gravity loads affecting only its lateral behaviour. Due to their symmetrical design, the evaluation of the seismic behaviour of each building is straightforwardly accomplished by a planar multi-story frame structure.

The planar structural models are implemented in OpenSees\[35\] for the requirements of this study, following the concentrated plasticity approach in which the masses representing the loads are lumped and the nonlinear behaviour of the members is modelled with two nonlinear springs at the ends of each beam and column element, as in Figure 1B.

The moment-rotation properties of the spring elements obey the modified Ibarra-Medina-Krawinkler model\[36–39\] with bilinear hysteretic response and quadrilinear backbone curve. The parameters of the deterioration model, such as the pre- and post-capping (plastic) rotation capacity, are determined using the predictive equations developed by Lignos and

\[3\] The models with the subscript “RBS” differ only in the position of the nonlinear springs of the beam elements, that is, the nonlinear spring (i.e., the plastic hinge) of each beam is modelled at the exact location of the RBS instead of the beam-column centreline
Krawinkler\cite{Krawinkler}. In order to take into account the influence of the indirect loads (i.e., green tributary area) on the modal properties and the P-Delta phenomena (i.e., second-order effects), the corresponding masses are applied on an axially rigid leaning column with zero flexural stiffness, placed in parallel to the frame (Figure 1B).

### 2.2 Ground motion records, intensity measures and spectra

For this study, as many GM records as possible were extracted from the PEER NGA-West2 database\cite{PEER} and a total of 17,150 GM were collected, covering a wide range of earthquake scenario parameters and GM intensity. These GM records form a rich database in which numerous IMs are examined in order to study their influence on scaling bias. Multiple IMs are considered whose importance is already widely recognised, such as quantities obtained from the response spectra, that is, the spectral acceleration $S_a$ at the fundamental period $T_1$, $S_a(T_1)$, and the average of spectral accelerations around $T_1$ ($S_{avg}$). Moreover, the Arias intensity (AI), the cumulative absolute velocity (CAV), the significant duration and all the seismological characteristics of magnitude (M), source-to-site distance ($R_{JB}$) as defined by Joyner and Boore, and the average shear-wave velocity in the top 30m ($V_{S,30}$) are also considered. In addition, novel spectra are introduced to include information regarding the sustained amplitude of the oscillation, as explained below.

#### Percentile spectra

In the recent study by Davalos and Miranda\cite{Davalos} on bias related to amplitude scaling, it is pointed out that the response spectrum does not provide information about other vibration amplitudes of the SDOF system that are smaller than the peak response, nor about how often these smaller amplitudes are exceeded. In some studies,\cite{Davalos} this is identified as a potential shortcoming of the response spectra in representing all the important properties of a GM.

To address this lack of information from response spectra, spectra from the 95th, 90th, 85th, and other percentiles are determined here\footnote{For every 5th percentile value until the 30th percentile, where the 100th percentile (i.e., the maximum absolute pseudo-acceleration) is the typical response spectrum of the 5% damped pseudo-accelerations of the elastic SDOF systems.}, referred to as percentile spectra. The calculation of the percentile spectra is demonstrated on the pseudo-acceleration response of the SDOF system with the period $T = 1$ s induced by the GM denoted as CHICHI.02-TCU105N, shown in Figure 2A. First, the two time instants are determined at which 1\% and 99\% of the energy is released. They are represented by the vertical lines at 30 s and 55 s in Figure 2A, similar to the calculation of the significant duration.\cite{Davalos, Davalos1} In the time interval between these time instants, the desired percentiles of the absolute values of the acceleration response are determined. For example, in Figure 2A, the horizontal dash-dotted lines correspond to the 80th percentile, that is, the areas labeled A1 and A2 contain 20\% of the accelerations, while area B contains the remaining 80\% in the time frame between 30 and 55 s. The peak acceleration indicated by the marker corresponds to the common definition of spectral acceleration.
In this way, the aim is to capture the sustained amplitude of GM that has been shown to affect structural collapse in certain cases.\cite{44,45} One advantage of the proposed percentile spectra is their straightforward computation. That is, the procedure for their calculation is the same as for the calculation of the response spectra, the only modification being that a percentile response is of interest rather than the peak response. Moreover, robust results are obtained by calculating these response quantities for the time frame of the significant response of the SDOF system rather than for the entire recorded time history. Thus, the resulting percentile response is not affected by the insignificant part of the SDOF response (e.g., the first 20 and last 15 s of the response shown in Figure 2A), whose duration can vary greatly from record to record.

In Figure 2B, the response spectra of the GM records CHICH1.02-TCU105N and NIIGATA-TCGH17NS are shown, with the NIIGATA-TCGH17NS record scaled by a scale factor of 5. The response spectra of the two GM records are very similar in the range from 0.25 to 2 s (i.e., between the vertical dashed lines). However, as shown in Figure 2C, the 80th percentile spectrum ($S_a^{80\%}$) of the CHICH1.02-TCU105N record has much larger values in the same range. This could lead to differences in the excitation of a structural model if the response is nonlinear and also if the cyclic degradation actively contributes to the behavior of the structure. With respect to the number of the percentile spectra considered and the specific percentiles (e.g., three percentile spectra for the 90th, 85th and 80th percentiles accordingly), special studies are required to determine their contribution. In this work, after computing a total of 14 percentile spectra (i.e., 95th, 90th, ..., 30th), only the 90th and 80th are considered (in Section 4.2.1) for the following two reasons: First, it is not easy to observe and control multiple spectra simultaneously, as shown in Section 4.2.1; thus, their number should be limited. Second, the percentile spectra should reflect the sustained amplitude of oscillation, that is, they should be both substantially different from the response spectrum and large enough for the oscillation to be significant.

3 | SEISMIC PERFORMANCE ASSESSMENT

3.1 | Nonlinear response history analysis, cloud analysis and incremental dynamic analysis

Exposing a structural model to a time history of support accelerations is a powerful tool for evaluating its behaviour under seismic excitation. For this reason, several frameworks are based on a number of NLRHA, such as cloud analysis and IDA, to provide a meaningful seismic assessment. In cloud analysis, a set of GM records (usually unscaled) is used as seismic
excitation for NLRHA and linear regression is performed to obtain a relationship between the IM and the EDP.\[^5\] In IDA, each record of a GM set is scaled stepwise until the structure reaches the collapse limit state. Linking the analyses for each record in the IM - EDP domain results in the “IDA curve”, and by assessing the multiple curves (for all GM records of the set) statistical evaluations of the relationship between IM and EDP are obtained. It is obvious that the concept of IDA involves amplitude scaling, which also gives rise to criticism: that is, that scaling introduces bias and that ultimately hazard inconsistent GM records are employed to assess the seismic behaviour of a structure.

Since the present study attempts to investigate the introduction of bias through amplitude scaling, it is considered suitable to perform IDA for all structures described in Section 2.1 and presented in Table I involving all 17,150 GM records. For each structure and record, the IDA curve is created using the hunt and fill algorithm,\[^2\] while multiple NLRHA, which are very dense with respect to the IM, are performed until the collapse with the lowest intensity. It should be noted that in this study, structural collapse is thoroughly examined and defined in two ways. First, the actual definition of structural collapse is sought when dynamic instability occurs, which is manifested as a flattening of the IDA curve (i.e., a tangent less than 20% of the initial stiffness\[^2\]) or non-convergence of the time-integration scheme.\[^2\] However, non-convergence might be a result of insufficient quality of the numerical code, integration steps, and even the round-off errors, and should be carefully monitored.\[^1\] Second, collapse is also defined as exceeding a sufficiently large threshold in MIDR above which the model may no longer be trustworthy. Here, an MIDR value of 0.1 is assigned, a threshold commonly adopted by many researchers to define structural collapse in ductile moment frame systems.\[^2,45\]

In all cases of this study (i.e., all structures and GM records), the MIDR threshold of collapse is reached at 0.1 and a flat line is also observed in virtually all cases. In those cases where non-convergence was observed, the computation was repeated with an insignificantly different intensity (e.g., 0.001 g higher and lower), and eventually the threshold MIDR of 0.1 was reached. In the re-calculations an excessively larger number of iterations was allowed, resulting in time-consuming computations.

A toolbox was created consisting of multiple computer programs capable of running, monitoring, post-processing and re-calculation as needed the vast number of NLRHA performed in this study. This toolbox contains multiple Matlab routines for pre- and post-processing, and OpenSees\[^35\] was used to build the structural models and perform the analyses. In order to meet the increased demands on computing power (i.e., about 3.4 million NLRHA in total and even more considering the re-calculations), the computer cluster ‘LEO HPC’ of the University of Innsbruck was used to perform the analyses in parallel utilizing its multiple cores.

### 3.2 Sensitivity of structural response on ground motion intensity

The core of this study is the database of dynamic responses of the considered structures from the numerous NLRHA and the evaluation and comparison of these results. Through these comparisons, the impact of various GM records with different amplitude scaling on the response of a single structure is demonstrated (for all structures in Table I).

When investigating the introduction of bias due to amplitude scaling, it is crucial to ensure that other sources of bias do not influence the structural responses and affect the conclusions. For example, one source of bias is the absence of important IMs in the assessment, as mentioned in Section 2.2. On the other hand, uncertainties in modelling parameters and their influence on the results are not of interest for this study, as the aim is to show intra-structure rather than inter-structure comparisons, that is, comparisons of the MIDR responses from different GM records for the same structure and not between different structural models.

As an introduction to the discussion of the response sensitivity to the GM intensity, two examples of IDAs for the Dmax12 structure are shown in Figure 3, where the circular markers represent non-collapse response and the cross markers represent collapse response. The NLRHAs that make up the IDAs in Figure 3 are numbered consecutively from 1 to the total number of NLRHAs in order to identify them. When interpreting the results of Figure 3A, several pitfalls may arise. First, when using IDA to identify collapse intensity, it is of utmost importance that the analyses are dense with respect to the IM, otherwise a range of intensities associated with collapse could be ignored, referred to as “intermediate collapse area”.\[^2\] In Figure 3A, an intermediate collapse area is found in the analyses numbered 20 to 26. If ignored, the IDA curve is identified by the analyses 1-19-27-28 instead of 1-19-20. Note that in this example the range of intensities outside the intermediate collapse area, expressed in $\text{Sa}(T_1)$, is given by the analyses 19 and 27 and is $0.175g - 0.155g = 0.02g$. Considering the analysis numbered 28 instead of 20 as the lowest intensity at which collapse occurs, the associated error in collapse intensity is $(0.192g - 0.157g)/0.157g = 22.3\%$.

In addition to the error associated with the density of NLRHA in the framework of IDA, the example of Figure 3A illustrates another potential pitfall that can occur when NLRHA is performed in general (and not necessarily in the context...
of the IDA framework). Clearly, an infinitesimal change in intensity can lead to vastly different structural responses as the collapse limit state is approached. The pairs of analyses {19, 20} as well as {26, 27} are perfect examples of the sensitivity of MIDR. In these examples, the MIDR responses vary significantly, even though the intensity is virtually the same. This pitfall can be understood by the collapse flatline described in Vamvatsikos and Cornell,\textsuperscript{[2]} that is, the plateau at maximum intensity at which the response approaches “infinity”. The more sudden this transition is (i.e., from low/linear responses to much larger/nonlinear/collapse), the more ambiguity is existent, as the same intensity can be seen to have low or high impact. In this example, the relative amplitude scaling of the GM record is 1.017 for the analyses 19 and 20 and 1.02 for the analyses 26 and 27, and therefore all IMs are virtually the same. The response spectra for analyses 19 and 20 depicted in Figure 3C graphically illustrate that the scaling difference is essentially negligible. Thus, if one of these analyses is utilized in one study instead of the other (e.g., analysis 19 instead of 20), the conclusions are very different.

In contrast, Figure 3B shows a gradual change in MIDR with increasing IM. Analyses 33, 39 and 46 have $S_a(T_1)$ of 0.162 g, 0.211 g and 0.237 g, respectively. Figure 3D illustrates their corresponding response spectra. The MIDR results of analyses 33, 39 and 46 is 0.033, 0.049 and 0.099, respectively. In other words, consecutively scaling with 1.30 and then 1.12 from the IM of analysis 33 to reach the IM of analyses 39 and 46 respectively, results in a consecutive increase of MIDR of 48% and then 100%. Even though the plateau in this case is not as sudden as in Figure 3A, it is again noted that the MIDR responses increase much more rapidly close to collapse. In these cases (Figure 3A and B), the responses are vastly different even though comparisons of the same time history are shown with small/moderate discrepancy in amplitude scaling. When comparing different GM records with similar intensities, the record-to-record variability is another source of uncertainty that is likely to amplify the discrepancies in the MIDR.

To quantitatively assess the discrepancy between two spectra, some researchers\textsuperscript{[25,28]} use the average deviation factor $\delta$, while here the metric $s$ is employed, which is based on the natural logarithms of the two spectra. The metrics $\delta$ and $s$ are given by the following equations,

$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{S_{a1}(T_i) - S_{a2}(T_i)}{S_{a2}(T_i)} \right)^2}, \quad s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \ln(S_{a1}(T_i)) - \ln(S_{a2}(T_i)) \right)^2}$$

where $S_{a1}$ and $S_{a2}$ are the spectra of interest (e.g., $S_{a2}$ might be the target spectrum) and $T_i$ are the periods in the corresponding range of interest. The choice of $s$ instead of $\delta$ is preferred in this work because it provides equivalent results.
but does not depend on the spectrum used for normalization. For these examples, where the compared spectra differ only due to scaling, $s$ is simplified to be the (absolute) logarithm of the relative amplitude scaling (i.e., $|\ln(1.017)| \approx 0.017$ for the analyses 19 and 20), regardless of period range. Note that $\delta$ is also 0.017 for this example because it is simplified as the decimal part of the relative amplitude scaling, regardless of the period range. In Iervolino et al. [25] $\delta$ values of around 0.15 are calculated between the average spectra of several GM sets and a target spectrum, arguing to reflect reasonable spectral compatibility. In contrast, Davalos and Miranda [28] compute $\delta$ values below 0.1 and also argue that spectral compatibility is ensured.

In the present study, several measures are taken to avoid false conclusions related to such problems. First, by examining a large number of analyses and GM datasets, it is expected that such discrepancies will cancel each other out. For example, comparing two large groups of analyses (e.g., in the range of hundreds or more), it is expected that the selection of analysis 19 rather than 20 from Figure 3A will be offset by the reverse selection from another GM record. However, it is common practice to select and compare sets with a limited number of GM records, as this is relevant for real-world applications. For this reason, this study demonstrates the selection and comparison of GM datasets with a relatively small number of datasets (e.g., 30) after ensuring that there are no cases as in Figure 3A. Repeating the selection and comparison process several times also helps to determine how sensitive the IM change is, as in Figure 3B, where an intensity near the collapse limit state causes a rapid increase in MIDR.

4 | THE EFFECT OF AMPLITUDE SCALING

4.1 | Comparing cloud analyses between scaled and unscaled ground motion records

As a first step, the quantification of the bias caused by amplitude scaling is examined by means of cloud analyses comparing clouds from unscaled GM records with clouds from scaled ones. These clouds describing the IM-EDP relationship are compared without explicit consideration of the GM compatibility in terms of spectral shape and scenario parameters, for example, through a GM selection.

This comparison requires an interpretation of each cloud of NLRHA, for which several possibilities are proposed in the literature [5]. In section 4.1.1 the cloud data are examined purely statistically, without introducing bias, by assuming a relationship between IM and MIDR (e.g., a commonly accepted assumption of a linear relationship between the logarithm of IM and MIDR) and thus fitting regression models that follow this relationship. In Section 4.1.2, this assumption is pursued and generalised linear models are fitted to describe the clouds considering multiple IMs.

4.1.1 | Straightforward comparison of 2D, IM-EDP clouds

First, clouds showing the relationship between an IM and the EDP (i.e., MIDR) are compared. The IM used is $S_{a_{avg}}$, the geometric mean of the spectral accelerations around $T_1$, which is generally accepted as an appropriate choice of a scalar IM to increase efficiency and sufficiency [10,16,18,19]

$$S_{a_{avg}} = \left( \prod_{i=1}^{N} S_a(c_i T_1) \right)^{1/N}$$

where $N$ is the number of periods used to compute $S_{a_{avg}}$, and the $c_i$ terms are increasing non-negative values, that is, $c_1 T_1$ is the lower bound and $c_N T_1$ is the upper bound of the range of periods considered. In this study, a period range between $0.2 T_1$ and $2.5 T_1$ is considered, where $N = 100$ and the $c_i$ values are calculated to correspond to a uniform spacing in log-periods.

The advantage of this approach is the vast number of GM records included and the straightforward comparison, which is also graphically comprehensible. Figure 4A shows the cloud for the Dmax12 structure subjected to the 17,150 unscaled GM records. For this structure, it can be observed that $S_{a_{avg}}$ of all GM have an intensity that does not exceed 0.7 g (same for all structures with the same fundamental period $T_1$ and thus the same $S_{a_{avg}}$). In order to compare unscaled with scaled clouds, it must be ensured that they refer to the same IM range. For this reason, only the $S_{a_{avg}}$ range up to 0.7 g is examined when scaling the GM records (i.e., an upper limit of 0.7 g is set). If no upper limit is set, the cloud with
FIGURE 4  (A) Unscaled cloud of the Dmax12 structure together with the 16th, 50th and 84th percentiles; (B) comparison of the 16th, 50th and 84th percentiles of unscaled and scaled clouds of Dmax12; (C) minimum, mean and maximum ratio of the scaled over the unscaled median $S_{a,avg}$ of Dmax12 for all scale factors up to 20; (D) as (C) but for the Dmax20 structure;

a scale factor of 10 would cover the IM range up to, say, 7 g. Since any GM above 0.5–0.6 g would lead to collapse, this would result in distorted collapse intensities (i.e., the median and percentiles at $MIDR = 0.1$). Note that the lower part of the IM range is always well represented regardless of the scale factor used (up to 20 in this study), because the majority of the GM records have very low intensities.

To statistically examine the cloud and express the relationship of MIDR in terms of $S_{a,avg}$, the following procedure is adopted. For each MIDR from 0 to 0.1, a small range around this MIDR is examined and the 16th, 50th and 84th percentiles of the $S_{a,avg}$ of all analyses in that range are determined. In this way, three curves are generated in the IM-EDP plot (e.g., Figure 4A for the unscaled cloud) expressing the median relationship of the MIDR with respect to $S_{a,avg}$ and indirectly the variance of this relationship (through the 16th and 84th percentiles). Figure 4A reveals that for a large proportion of the analyses, the structure remains in the elastic response domain, as indicated by an MIDR of less than about 0.01, while only 26 GM records cause the structure to collapse (i.e., $MIDR >= 0.1$). In ranges where there is not enough data (e.g., $MIDR \in [0.05 - 0.099]$), interpolation is performed to include information from higher and lower values of MIDR.

In Figure 4B the three full black curves, which correspond to the unscaled cloud of Figure 4A, are compared with the corresponding curves of a cloud with a scale factor of 2 and of 5. At first glance, the three sets of three curves appear to be the same up to an MIDR of 0.03, and at larger MIDR a slight, gradual change is observed until the collapse limit state is reached (at $MIDR = 0.1$) where the difference is greatest. Since the percentile curves of the three clouds relate to each other approximately in the same way as the median curves, it is sufficient for the following discussion to focus only on the deviations in the median curves.

To track these deviations for scale factors in the range from 2 to 20, Figure 4C illustrates the maximum, mean and minimum ratio along the MIDR, of the median $S_{a,avg}$ related to the scaled GMs to the the median $S_{a,avg}$ related to the unscaled GMs. This ratio, which is close to one (both average and maximum/minimum values) shows good agreement of the median $S_{a,avg}$ between scaled and unscaled GMs. It is readily observed that this ratio does not change significantly with increasing scale factor, as the mean values are close to one and the largest difference is approximately 20%. In Figure 4D, this ratio is depicted for the structure Dmax20, which qualitatively shows the same behaviour as for structure Dmax12, but the largest difference is around 25%, which is the largest value between all examined structures.

Examining the cloud analyses in this way suggests that scaling is not the main reason for the deviations in the MIDR, as is evident in the results with increasing scale factor (from 2 to 20). Although the ratio shown in Figure 4C and D is not
negligibly different from 1, these deviations do not increase with scaling (i.e., they are roughly the same for a scale factor of 3 and 20). Moreover, Figure 4B suggests that the response at low to medium intensities, where the structure is far away from the collapse limit state, is not affected by scaling.

It is clear that this way of comparing the response based on scaled and unscaled GM records has a number of shortcomings. First of all, the earthquake scenario parameters and IM of interest other than $S_{a_{avg}}$ are ignored, which could be the most contributing reason for the difference close to $MIDR = 0.1$, where only a small number of analyses are available for the unscaled cloud. As the scale factor increases, GM records with different scenario parameters, which naturally have lower intensities, are scaled and are therefore the ones that cause structural collapse for the scaled clouds. Furthermore, the empirically introduced upper limit (at 0.7 g for Dmax12) affects the results mainly in the collapse limit state (i.e., the limit state where the analyses are as close as possible to this upper limit), while other values could have been chosen instead (e.g., 0.5 g) delivering slightly different results. As deviations on the order of 10% to 20% are observed for all structures, further ways of analysing the data to address these shortcomings are explored below.

4.1.2 Fitting generalized linear models to compare clouds

In order to take into account the possible effect of the upper bound in intensity as well as the effect of IMs other than the $S_{a_{avg}}$, a different approach to the examination of clouds is chosen in the following. First, to avoid having to identify the appropriate range of values for each considered IM, in this approach the range of $MIDR$ is instead given as $0 \leq MIDR < 0.1$, that is, the upper boundary of the $MIDR$ is slightly below the collapse limit state. In this way, the clouds based on scaled GMs have some GMs with unrealistically large intensities, but these lead to an $MIDR \geq 0.1$ (i.e., collapse) and are therefore not included in the assessment. However, since large $MIDR$ values close to 0.1 occur more frequently due to the GM scaling, it is assumed that conclusions for the collapse limit state regarding scaling can also be drawn from these near-collapse values.

Moreover, in addition to $S_{a_{avg}}$, all IMs listed in Section 2.2 are used to create a multidimensional cloud. To be able to describe a relationship in a multiple IM-EDP domain, a generalized linear model (i.e., a regression model) is fitted to the data of the unscaled cloud. The input to this numerical model is the IMs and the output is the $MIDR$ (i.e., the prediction of the model). The accuracy of the regression model is assessed using both the unscaled and scaled clouds, so any existing difference in accuracy between the two will determine the conclusions regarding bias in the $MIDR$. In other words, if the regression model calibrated only with unscaled data can approximate structural drifts equally well for unscaled and scaled clouds, then it is assumed that no ‘hidden’ effects are introduced by scaling.

To this end, the work of Tsalouchidis et al. [46] is revisited, where linear and logistic regression models were trained to fit the collapse intensity data obtained through IDA. Since linear regression has been shown[46] to perform better than logistic regression, it is also used in this study. An approximately linear relationship between the logarithms of $Sa$ and $MIDR$ is expected,[5] and therefore the logarithms of the above mentioned IMs are utilized as a list of possible features (or predictors) for the regression models. In order to select the most comprehensive predictors of structural response from this extensive list of IMs, the regularized regression named lasso [47] is employed for the feature selection, similar to the work of Tsalouchidis et al.,[46] where detailed documentation is provided regarding the implementation of the process of regularisation. This is done separately for each structure studied here (Table 1). Overall, the IMs $S_{a}(T_1)$, $S_{a_{avg}}$ and their ratio, that is, $S_{a_{ratio}} = S_{a}(T_1)/S_{a_{avg}}$ have been found to be the best predictors, as it has been explicitly shown[17] that they predict much more accurately the collapse intensity than other spectral shape metrics. It is also noted that a non-negligible contribution to the accuracy of the prediction models comes from some non-spectral information such as CAV, AI and $R_{JB}$, similar to the outcomes in Tsalouchidis et al.[46]

After identifying the IMs to be included, the regression hyper-surface is fitted to the data of the unscaled cloud. Instead of using the entire dataset (i.e., 17,150 GM records), 70% are employed for fitting (or training), i.e., the training set, and the remaining 30% (the test set) are used for testing, i.e., evaluating the performance of the regression model on data not used for its calibration, as it is usually done. To test the regression model in its ability to accurately approximate the structural responses, the residuals of the regression

$$residuals = \ln(MIDR_{predicted}) - \ln(MIDR_{actual})$$

5 Using multiple IMs and higher order terms results in a hyper-surface, whereas using a scalar IM and simple linear regression results in a line.
are calculated. In this equation, $MIDR_{predicted}$ is the prediction of the MIDR from the prediction model and $MIDR_{actual}$ is the outcome of the NLRHA. The closer the residuals metric is to zero, the more accurate the prediction model is. In Figure 5A, each of the unscaled analyses of the test set for the Dmax12 structure is represented by its $MIDR_{actual}$ and $MIDR_{predicted}$ values, as well as the (average) regression line of this graph. The natural logarithm of the tangent of the angle of this line is the mean residual of the prediction model, whose value being close to 0 indicates that the model is not biased, that is, systematically over- or under-predicts MIDR values. In Figure 5B, the same plot is shown for the scaled cloud with a scale factor of 2 for the Dmax12 structure using the same prediction model, which was calibrated utilizing the unscaled cloud. However, the entire set of data is used as a test set here, since none of the scaled analyses were used to calibrate the prediction model (unlike in the case of the unscaled cloud). Figure 5B demonstrates that the model provides unbiased predictions for the analyses with a scale factor of 2, as the average regression line appears to have a tangent of 1. Moreover, it is clear that even limited scaling is helpful in obtaining larger MIDR values, that is, 0.04–0.07, when the structural response is highly nonlinear.

To determine whether scaling introduces a systematic bias for scale factors larger than 2 and for all structures examined in this study, Figures 6A and B show the results of the mean residual for scale factors up to 20 for the “Dmax” and “Dmin” structures, respectively (Table 1), as well as the minimum, mean and maximum residuals of all 10 structures (for comparison purposes). From these figures, it can be seen that the mean residual hardly deviates from 0 for scale factors larger than 1, while these deviations around 0 do not seem to increase systematically with increasing scale factor or...
indicate a systematic trend towards over- or under-predict the MIDR, both for the more flexible “Dmin” and the stiffer “Dmax” structures.

The analysis of the multidimensional clouds from scaled and unscaled GM records in this way leads to a similar conclusion as Section 4.1.1. Increased scale factors do not seem to have any effect on the efficiency of the prediction models calibrated with unscaled NLRHA, while the predictions of the models are practically unbiased.

4.2 Comparing spectrum and scenario compatible ground motion record sets utilizing a novel ground-motion selection scheme

The previous investigations on introducing bias by amplitude scaling did not take into account some important aspects relevant to most applications in earthquake engineering. Here, instead of comparing clouds with thousands of NLRHA (as in Section 4.1), hazard compatible GM sets with a limited number of records are selected and compared. In this way, the introduction of bias on the MIDR responses based on NLRHA, is examined when GM selection is performed to match the typical requirements of an earthquake engineering application.

The utilized GM record selection approach has been recently developed by the authors and ensures efficient selection of GM sets that achieve spectral compatibility at given spectral targets in first and second order statistics (i.e., mean and standard deviation spectra). Moreover, the scenario parameters of the GM records are controlled while the selection process is based on multi-objective optimisation and can be used to select up to three-component ground excitation sets. As shown in Tsalouchidis et al., the purpose of obtaining three-component GM sets is to perform three-dimensional excitation (i.e., using two horizontal and one vertical components), and the selection procedure can successfully meet spectral targets in all three components. For the present study, however, one-component excitation is of interest. The advantages of this GM selection procedure are utilized to obtain GM sets with specific median and standard deviation spectra, while it is adjusted for the needs of this study to also provide specific median spectra of the 90th and 80th percentile spectra (as defined in Section 2.2). In this way, the GM sets are not selected here to meet spectral targets in more than one orthogonal directions, but to meet one-component spectral targets as well as additional percentile spectral targets for the same component. Therefore, without significant computational efforts to obtain these spectra (as shown in Section 2.2) and by adjusting an already available GM selection procedure, the sustained vibration amplitude is controlled in addition to the peak responses of the elastic SDOF systems described by the response spectra.

In the following, Section 4.2.1 describes how the process of GM selection is performed to adequately examine the research question while avoiding the pitfalls associated with NLRHA addressed in Section 3. The wealth of the results and a discussion are presented in Section 4.2.2.

4.2.1 Selecting and comparing ground-motion sets

To investigate the research question, we need to compare the MIDR from GM sets that have similar attributes in terms of IM and differ only in the implemented scaling. Simultaneously, we need to investigate the extent of the differences in MIDR responses from GM sets that contain unscaled records and share the same IM, in order to determine whether amplitude scaling increases these inherent differences. To this end, four GM sets are selected (separately for each structure of Table 1), namely the reference set Set_Tgt, and the sets Set_Tgt*, Set_SF_5 and Set_SF_10, each containing 30 records. Moreover, to investigate the effect of considering the percentile spectra, three additional GM sets are selected (separately for each structure of Table 1), namely Set_Tgt*_{NapSPE}, Set_SF_5_{NapSPE} and Set_SF_10_{NapSPE}, each containing 30 records.

First, reference set Set_Tgt is created by randomly selecting 30 GM records from a given earthquake scenario, with the requirement that the distribution of MIDR reflects nonlinear and near-collapse responses. Specifically, this earthquake scenario contains records from severe earthquake events that act at a relatively short distance with moment magnitude $M \geq 6$, Joyner-Boore distance $R_{JB} \leq 80 \text{km}$ and average seismic shear-wave velocity from the surface to a depth of 30 m $V_{S,30} \in [200 \text{760}] \text{m/s}$ and relate to soils classified as category C or D according to the National Earthquake Hazards Reduction Program (NEHRP). The records that make up Set_Tgt are virtually unscaled, as each of them has a scale factor in the range of [0.5 1.5] and on average around 1. The use of limited scaling is helpful, for example, when an unscaled GM record would cause the structure to collapse and therefore a scale factor of slightly less than 1 is required. The reference

---

6To ensure highly nonlinear responses in Set_Tgt, it is aimed that the median MIDR is approximately in the range [0.04 0.05].
Set_Tgt GM set is employed as the target for the other sets. That is, the GM selection procedure of the other sets uses its median and standard deviation spectra as well as the median spectra obtained from the 90th and 80th percentile spectra. The period range in which the spectral compatibility is searched for is [0.2, 2.5] and therefore varies for each structure as it depends on the first fundamental period of vibration. Moreover, the GM records of Set_Tgt are randomly selected for each of the structures, resulting in different spectral targets.

The next GM set to be selected is Set_Tgt*, which consists of records from the same earthquake scenario as Set_Tgt, with scale factors in the range of [0.5, 1.5]. That is, Set_Tgt* has identical attributes to Set_Tgt and is used to assess the inherent differences in MIDR responses mentioned above. The next two GM sets are Set_SF_5 and Set_SF_10, which also contain 30 GM records from the same scenario, but the scale factors are around 5 and 10 (i.e., the GM records in Set_SF_5 have scale factors in the range [4.5–5.5] and those in Set_SF_10 have scale factors in the range [9.5–10.5]). In this way, two levels of scaling causing differences in MIDR responses are assessed after spectral and scenario compatibility between all these four GM sets has been established. Finally, Set_Tgt_NoPeSpe, Set_SF_5NoPeSpe and Set_SF_10NoPeSpe are selected with the same attributes as Set_Tgt*, Set_SF_5 and Set_SF_10, respectively, but without considering the percentile spectra for GM selection (i.e., only the median and standard deviation spectra of Set_Tgt are used as targets for these three GM selection procedures).

A total of 60 GM selection procedures are carried out (i.e., for a total of 10 structures, three sets are selected to match the median, standard deviation and percentile spectra of Set_Tgt and three more are selected to match only the median and standard deviation spectra). In order to present the results efficiently and meaningfully, they are shown in more detail for one structure (Dmax12) and concisely summarised for the entirety of the models examined.

Evaluating the results of the pareto-optimal solutions

As explained earlier, the selection process is based on multi-objective optimisation, thus leading to multiple optimal solutions. The paper of Tsalouchidis et al. demonstrates the advantages of multi-objective optimisation in GM selection in terms of effectively meeting the multiple selection objectives, as well as a discussion and framework for choosing between the multiple solutions. In the present study, several solutions for each GM selection process are investigated, which differ in how they meet the various spectral objectives. To rank the solutions according to the spectral objectives of this study, the metric $R_{\text{spectral}}$ of Tsalouchidis et al. is used, given by equation

$$R_{\text{spectral}} = w_{m,100} F_{m,100} + w_{\sigma,100} F_{\sigma,100} + w_{m,90} F_{m,90} + w_{m,80} F_{m,80}$$

where $F_{m,100}$ is the objective function that is minimized when the median spectrum is approaching the corresponding target. $F_{\sigma,100}$ is the objective function that is minimized when the standard deviation spectrum is approaching the corresponding target. $F_{m,90}$ and $F_{m,80}$ are the objective functions related to the medians of the 90th and 80th percentile spectra, respectively, and $w$ are the corresponding weights that determine the relative importance between the objectives. Changing the values of $w$ results in a different solution that has the best score (i.e., minimum $R_{\text{spectral}}$ value).

Figure 7 shows two optimal solutions (among hundreds) for Set_Tgt* of Dmax12, namely Set_Tgt*_A and Set_Tgt*_B, together with the target spectra obtained from Set_Tgt. The solution Set_Tgt*_A (left column) is obtained when all $w_m$ values are equal and twice as large as $w_{\sigma}$, and Set_Tgt*_B (right column) is obtained when all $w_{\sigma}$ values are equal. Figure 7A depicts the 10th, 50th and 90th percentiles of the response spectra of Set_Tgt and Set_Tgt*_A, and Figure 7C the 50th percentile of the 90th percentile spectra of Set_Tgt and Set_Tgt*_A, and Figure 7E the 50th percentile of the 80th percentile spectra of Set_Tgt and Set_Tgt*_A. Figures 7B, D, and F are equivalent to A, C and E for Set_Tgt*_B.

Figures 7A–F show that Set_Tgt*_A and Set_Tgt*_B both meet the spectral properties of Set_Tgt excellently. Moreover, it should be noted that Set_Tgt*_A and Set_Tgt*_B have 20 out of 30 GM records in common with slightly different scaling around 1 employed in each one and no GM record in common with Set_Tgt. The resulting MIDR distributions are also very similar, with a median MIDR of about 0.04 and a standard deviation of 0.02, as can be seen from Figures 7G and H. Apart from the proximity of the resulting MIDR distributions, which will be discussed in more detail below, it is also noted that the MIDR responses are particularly large, fulfilling the requirement to consider responses in the nonlinear and near-collapse state. Also, note that the GM sets Set_Tgt_NoPeSpe for the Dmax12 structure do not exhibit systematic and significant differentiation in the percentile spectra (although they were neglected) and thus show similar behaviour to Set_Tgt*. For this reason, the GM sets Set_Tgt*_NoPeSpe are not shown.

In this study, a total of 15 GM sets such as Set_Tgt*_A and Set_Tgt*_B are examined to assess the MIDR responses of Set_Tgt* and compare with Set_Tgt. The same is done for Set_SF_5 and Set_SF_10, thus Figure 8 shows one GM set (out of the 15) for Set_SF_5 and one for Set_SF_10 (with blue) for the Dmax12 structure, similar to Figures 7A–H. The GM sets
FIGURE 7  For the structure Dmax12: (A) 10th, 50th and 90th percentiles of the response spectra of Set\_Tgt and Set\_Tgt\_A; (B) as (A) for Set\_Tgt\_B; (C) 50th percentile of the 90th percentile spectra of Set\_Tgt and Set\_Tgt\_A; (D) as (C) for Set\_Tgt\_B; (E) 50th percentile of the 80th percentile spectra of Set\_Tgt and Set\_Tgt\_A; (F) as (E) for Set\_Tgt\_B; (G) empirical cumulative distribution of MIDR for Set\_Tgt and Set\_Tgt\_A of Dmax4; (H) as (G) for Set\_Tgt and Set\_Tgt\_B

Set\_SF\_5NoPeSpe and Set\_SF\_10NoPeSpe are also shown (with green). The compatibility of the response spectra as well as the percentile spectra between Set\_Tgt, Set\_SF5 (left) and Set\_SF10 (right), is excellent (Figures 8A–F) and the MIDR distributions (Figures 8G, H) appear to agree very well with the target up to the median (i.e., $MIDR = 0.04$), whereas some discrepancies are observed at larger MIDR values.

The compatibility of the response spectra of Set\_SF\_5NoPeSpe and Set\_SF\_10NoPeSpe is also excellent, while their percentile spectra exhibit noticeable differences. The 90th and 80th percentile spectra of Set\_SF\_5NoPeSpe are below the target in the period range around 6–9 s, which is about 2–3 times the fundamental period, while the lower ordinates of these spectra are in good agreement. Using the metric from Equation (2), these deviations are $s = 0.33$ and $s = 0.51$ at $T = 7$ s for the 90th and 80th percentile spectra, respectively. These lower amplitudes in the “elongated” period range can justify the noticeably lower MIDR responses of Set\_SF\_5NoPeSpe (beyond the median MIDR).

As for the Set\_SF\_10NoPeSpe, it exhibits significantly larger amplitudes of percentile spectra in the range of about 0.6–3 s (with $s \approx 0.55$), while the rest is in good agreement. It should be noted, however, that these higher amplitudes of the percentile spectra in the “higher-mode” period range have a marginal effect on the MIDR distributions (Figures 8G, H).
FIGURE 8  For the structure Dmax12: (A) 10th, 50th and 90th percentiles of the response spectra of Set_Tgt and Set_5; (B) as (A) for Set_SF10; (C) 50th percentile of the 90th percentile spectra of Set_Tgt and Set_SF5; (D) as (C) for Set_SF10; (E) 50th percentile of the 80th percentile spectra of Set_Tgt and Set_SF5; (F) as (E) for Set_SF10; (G) empirical cumulative distribution of MIDR for Set_Tgt and Set_SF5; (H) as (G) for Set_SF10.

One explanation for these results might be that the sustained amplitude of vibration is likely to play a role in highly nonlinear responses, where the effect of period elongation is also expected (i.e., a period range up to $2.5T_1$ is considered for spectral compatibility). The exact corresponding period range for the percentile spectra is not known, nor their influence and the exact percentile values that are particularly informative (i.e., the 90th, 80th or other percentiles), so further research is needed here.

In this study, more cases like those demonstrated in Figure 8 are observed, but their number is limited. On the contrary, the majority of the GM sets that do not include the percentile spectra in the selection procedure do not systematically exceed them or fall below them. This is to be expected, given the high correlation between the response spectra and the percentile spectra. Hence, spectral compatibility largely (but not completely) suggests that the sustained amplitude is also compatible.
A more detailed discussion with quantitative assessments and comparisons of the MIDR discrepancies between the GM sets can be found in Section 4.2.2. These results and discussion do not involve further comparisons between the GM sets that do or do not consider the percentile spectra because of the limited number of cases that can be directly assessed as those in Figure 8.

**Sensitivity of MIDR**

An important focus of this study is to eliminate potential sources of bias or inaccuracies. A key component of this is the examination of 15 GM sets for \( \text{Set}_T \) and comparison with the targets. To proceed with the comparison of a GM set, perfect spectral compatibility is first ensured (with \( s < 0.1 \), like in Figure 7) and then a sensitivity analysis is performed to confirm that the pitfalls mentioned in Section 3 are not present. To this end, for each GM set obtained through the GM selection process (i.e., after each of the GM records contained in it has been appropriately scaled), a uniform re-scaling of the entire set is performed twice: once with a scale factor of 1.03 and once with 0.97, that is, a uniform \( \pm 3\% \) differentiation in scaling is applied and the resulting MIDR are examined. The aim is to observe whether such marginal differentiation causes significant changes in the MIDR distributions and thus leads to unreliable conclusions when comparing such GM sets, as discussed in Section 3. Note that a uniform 3% differentiation in scaling results in a marginal loss of spectral compatibility with respect to a given target (e.g., in terms of median spectral compatibility, an increase of \( s \) of approximately 0.03 is observed).

The MIDR distributions of two GM sets, namely \( \text{Set}_T \) and \( \text{Set}_{T'} \), are presented in Figure 9A and B, respectively, which are obtained by the process of GM selection for \( \text{Set}_T \) of Dmax20. For completeness, the MIDR distribution of the corresponding target is also shown. In addition to the MIDR responses of \( \text{Set}_T \) and \( \text{Set}_{T'} \), Figure 9A and B show the MIDR distribution of the corresponding GM sets with the \( \pm 3\% \) relative scaling. In the case of \( \text{Set}_{T'} \), a huge change in the resulting MIDR distributions is observed, with a change in median MIDR of around 65%, while in the case of \( \text{Set}_{T'} \), the MIDR distribution is not significantly altered, having a change in median MIDR of around 6%. In both cases, it is noticeable that the lower values of the MIDR distribution differ less than the higher ones, which is to be expected given the considerable sensitivity of the MIDR response associated with highly nonlinear and near-collapse MIDR values.

For this reason, when a GM set such as \( \text{Set}_T \) is examined, where the median MIDR values differ more than 20% between the cases of relative scaling of \( \pm 3\% \), it is not considered here for further investigation and comparison (i.e., as one of the 15 GM sets). In most cases examined, the median MIDR values differed between the \( \pm 3\% \) relatively scaled GM sets by 5%–20%, while some exceeded the 20% threshold imposed here, such as \( \text{Set}_T \) in Figure 9A.

To illustrate this quantitatively in terms of spectral compatibility, the \( \pm 3\% \) relative scaling of a GM set such as \( \text{Set}_T \) yields \( s = \ln(1.03/0.97) = 0.06 \) between the two generated sets, that is, 0.97 \( \cdot \) \( \text{Set}_T \) and 1.03 \( \cdot \) \( \text{Set}_T \).

---

7 The structure Dmax12 that is used to demonstrate each part of the study so far does not contain a GM set in which the \( \pm 3\% \) differentiation in scaling resulted in significant MIDR changes; hence, the Dmax20 structure is shown here.
4.2.2 Results

The results of all structures are presented and discussed here. The comparisons of the MIDR distributions between Set\_T\_tgt and another GM set, for example, Set\_T\_tg^*, are conducted through $m_{\text{error}}$ and $\sigma_{\text{error}}$, given by the equations

$$m_{\text{error}} = 100 \cdot \frac{m_{\text{Set\_T\_tg}} - m_{\text{Set\_T\_tg}^*}}{m_{\text{Set\_T\_tg}}}, \quad \sigma_{\text{error}} = 100 \cdot \frac{\sigma_{\text{Set\_T\_tg}} - \sigma_{\text{Set\_T\_tg}^*}}{\sigma_{\text{Set\_T\_tg}}}$$

(6)

where $m_{\text{Set\_T\_tg}}$, $m_{\text{Set\_T\_tg}^*}$ are the medians and $\sigma_{\text{Set\_T\_tg}}$, $\sigma_{\text{Set\_T\_tg}^*}$ are the standard deviations of the lognormal distributions fitted to the MIDR responses of Set\_T\_tgt and Set\_T\_tg^*, respectively.

Figure 10A shows for each structure the $m_{\text{error}}$ of the 15 GM sets examined for each Set\_T\_tg^* (black), Set\_SF5 (red), and Set\_SF10 (blue), as well as an asterisk marker indicating the mean of these 15 $m_{\text{error}}$ values. At the top of Figure 10A, the values of $m_{\text{Set\_T\_tg}}$ (i.e., the target medians) are shown for each structure. Figure 10B is similar to Figure 10A and shows the standard deviations.

Looking first at the median discrepancies in Figure 10A, it is argued that no systematic bias is observed with Set\_SF5 and Set\_SF10. Overall, the $m_{\text{error}}$ values deviate from 0, with the lowest values reaching -20% and the largest values around 10%. Although negative $m_{\text{error}}$ values are more common than positive ones, this is considered by the authors to be insignificant and not consistent across the structures to denote a pattern. In particular, it is argued that a ±15% discrepancy in the median MIDR should be expected, justified by the large corresponding standard deviation values. For example, the
fitted lognormal distribution for the Set_Tgt of the Dmax4 structure has a median of 0.045 and a standard deviation of 0.02, where the 95% confidence intervals of the median are 0.039 and 0.052. These values correspond to approximately ±15% of m_{Set_Tgt}. As it is common to use 11 GM records in a GM set instead of 30 used here (e.g., frequently required by design codes for building structures), the expected confidence intervals would be approximately 0.035 and 0.056 (for Set_Tgt of Dmax4), which would correspond to approximately 25% of m_{Set_Tgt}. Under these uncertainties on the median MIDR, no systematic bias is observed due to amplitude scaling. Moreover, analogous results appear for all of the examined structures shown in Figure 10A, which is in contrast to relevant research\cite{14} indicating that a stronger bias is observed for stiffer structures compared to the more flexible ones.

As for the standard deviation discrepancies, very large values between -90% and 90% are observed in Figure 10B. For all structures except Dmax12, the GM sets Set_Tgt* appear to have positive (and significant) \( \sigma_{\text{error}} \) values while Set_SF5 and Set_SF10 appear to have negative (and significant) \( \sigma_{\text{error}} \) values. Overall, Figure 10B supports the understanding that large uncertainties are inherent in the MIDR distribution and that a conclusion regarding amplitude scaling bias is not obvious under these uncertainties.

### 5 SUMMARY AND CONCLUSIONS

Previous studies have found conflicting results regarding the introduction of bias into MIDR results of NLRHA due to amplitude scaling. Many researchers argued that biased MIDR distributions are to be expected when scaling is employed and explained this tendency by the difference in spectral shape between scaled and unscaled GM. Other researchers proposed methods to obtain GM record sets that take into account spectral shape compatibility and showed that bias was successfully minimized and (in most cases) negligible when their methods were applied. However, recent studies have shown that even when spectral shape is carefully considered in GM selection, biased results are often observed and MIDR distributions are likely to be overestimated.

Although the topic is attracting a great deal of research interest, no conclusive results are yet available. In the present study, the topic was investigated with the awareness that extensive analyses should be carried out in order to deepen the conclusions presented and thus avoid erroneous results. For this reason, 10 planar steel frame buildings, ranging from low to high buildings, were analysed utilizing an unprecedented number of recorded GM (more than 17,000), resulting in a total of approximately 3.4 million NLRHA. From the wealth of these data, a discussion was held on the most important IM, which were also taken into account here. Moreover, novel spectra were presented here, namely percentile spectra, as an intuitive, easy-to-implement and informative approach that can be used to assess the sustained amplitude of elastic SDOF oscillators. In this way, in addition to the commonly used response spectra (i.e., maximum responses of elastic SDOF oscillators), the spectral shape was considered in terms of the sustained amplitude of the vibrations.

The research question was examined from multiple points of view. Initially, clouds of unscaled NLRHA were compared with clouds of scaled ones. The comparisons were made in two ways, firstly using intuitive statistical metrics and secondly using machine learning techniques to create prediction models based on multiple IM. Both approaches concluded that increasing the scaling had no impact on the results of the NLRHA, even though spectral shape compatibility was not explicitly considered. However, the analyses contained many thousands of GM records, which is uncommon for a typical earthquake engineering application. For this reason, a typical GM selection process was also performed to compare multiple GM sets of 30 GM records each.

The use of sets with a limited number of GM records (leading to nonlinear structural responses) showed that considerable sensitivity of MIDR distributions can occur and marginal scaling can lead to vastly different structural responses. With respect to percentile spectra, it was found that when they exhibit systematic differences between two GM sets (e.g., in the elongated period range at around 2–2.5\( T_1 \)), they affect the corresponding MIDR distributions. However, such systematic differences were not common. For this reason, it is suggested that control of the sustained amplitude by the percentile spectra is sufficient to ensure that a GM selection procedure yields compatible GM sets (given that percentile spectra targets can be obtained), and that a GM selection that specifically considers percentile spectra (as it has been done in this work) is not necessary to ensure unbiased results.

Overall, the results suggest that scaling does not introduce bias, or at least that bias is not observed under the inherent uncertainty of the involved NLRHA calculations. However, utilizing sets of GM records with low amplitudes in combination with amplitude scaling to overcome the lack of enough GM at higher amplitudes should be employed with caution. First and second order statistics of the spectral shape need to be compatible and explicitly taken into account through a GM selection procedure, as well as a verification that the sustained amplitude of vibration is at the expected level.
ACKNOWLEDGEMENTS
The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck. The authors would like to thank the two reviewers for their thorough and critical reviews and comments, improving the quality of this work.

DATA AVAILABILITY STATEMENT
Research data are not shared.

ORCID
Konstantinos Theodoros Tsalouchidis https://orcid.org/0000-0003-4574-6503
Christoph Adam https://orcid.org/0000-0001-9408-6439

REFERENCES
1. Sextos AG. Selection of ground motions for response history analysis. *Encyclopedia Earthquake Eng*. 2014;4:3239-3247.
2. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthquake Eng Struct Dyn*. 2002;31(3):491-514.
3. Bazzurro P, Cornell CA, Shome N, Carballo JE. Three proposals for characterizing MDOF nonlinear seismic response. *J Struct Eng*. 1998;124(11):1281-1289.
4. Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. *Earthquake Spectra*. 1998;14(3):469-500. https://doi.org/10.1193/1.1586011
5. Jalayer F. Direct Probabilistic Seismic Analysis: Implementing Non-linear Dynamic Assessments. PhD thesis. Department of Civil and Environmental Engineering, Stanford University, Stanford, CA; 2003.
6. Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthquake Eng Struct Dyn*. 2009;38(8):951-972.
7. Moehle J, Deierlein GG. A framework methodology for performance-based earthquake engineering. In: *13th World Conference on Earthquake Engineering*, Vancouver, Canada; 2004:679.
8. Deierlein GG. Overview of a comprehensive framework for earthquake performance assessment. In: *Proceedings of an International Workshop on Performance-Based Seismic Design Concepts and Implementation*. Bled, Slovenia; 2004:15-26.
9. Porter KA. An overview of PEER's performance-based earthquake engineering methodology. In: *Proceedings of Ninth International Conference on Applications of Statistics and Probability in Civil Engineering*. San Francisco, CA; 2003:1-8.
10. Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure for collapse risk assessment. *Earthquake Eng Struct Dyn*. 2015;44(12):2057-2073. https://doi.org/10.1002/eqe.2575
11. Adam C, Tsantaki S, Ibarra LF, Kampenhuber D. Record-to-record variability of the collapse capacity of multi-story frame structures vulnerable to P-delta. In: *Proceedings of the Second European Conference on Earthquake Engineering and Seismology (2ECEES)*. 2014.
12. CEN. 8–Design of Structures for Earthquake Resistance–Part I: General Rules, Seismic Actions and Rules for Building. Br Stand Institute, London, UK; 2004.
13. Applied Technology Council and United States. Federal Emergency Management Agency. *Quantification of Building Seismic Performance Factors*. US Department of Homeland Security, FEMA. 2009.
14. Luco N, Bazzurro P. Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses?. *Earthquake Eng Struct Dyn*. 2007;36(13):1813-1835.
15. De Biasio M, Grange S, Dufour F, Allain F, Petre-Lazar I. Intensity measures for probabilistic assessment of non-structural components acceleration demand. *Earthquake Eng Struct Dyn*. 2015;44(13):2261-2280.
16. Kohrangi M, Vamvatsikos D, Bazzurro P. Site dependence and record selection schemes for building fragility and regional loss assessment. *Earthquake Eng Struct Dyn*. 2017;46(10):1625-1643.
17. Eads L, Miranda E, Lignos D. Spectral shape metrics and structural collapse potential. *Earthquake Eng Struct Dyn*. 2016;45(10):1643-1659.
18. Adam C, Kampenhuber D, Ibarra LF, Tsantaki S. Optimal spectral acceleration-based intensity measure for seismic collapse assessment of P-delta vulnerable frame structures. *J Earthquake Eng*. 2017;21(7):1189-1195.
19. Adam C, Kampenhuber D, Ibarra L. Optimal intensity measure based on spectral acceleration for P-delta vulnerable deteriorating frame structures in the collapse limit state. *Bull Earthquake Eng*. 2017;15:4349-4373.
20. Shome N, Cornell CA. Probabilistic seismic demand analysis of nonlinear structures. Report no. RMS-35. *Research Management System (RMS) Program, Department of Civil Engineering*. Stanford, CA: Stanford University; 1999.
21. Stewart JP, Chiou SJ, Bray JD, Graves RW, Somerville PG, Abrahamson NA. Ground motion evaluation procedures for performance-based design. *Soil Dyn Earthquake Eng*. 2002;22(9):765-772.
22. Bommer JJ, Acevedo AB. The use of real earthquake accelerograms as input to dynamic analysis. *J Earthquake Eng*. 2004;8(1):43-91.
23. Baker JW. Measuring bias in structural response caused by ground motion scaling. In: *8th Pacific Conference on Earthquake Engineering*. Singapore; 2007; Paper No 056, 8 pp.
24. Hancock J, Bommer JJ, Stafford PJ. Numbers of scaled and matched accelerograms required for inelastic dynamic analyses. *Earthquake Eng Struct Dyn*. 2008;37(14):1585-1607.
25. Iervolino I, De Luca F, Cosenza E. Spectral shape-based assessment of SDOF nonlinear response to real, adjusted and artificial accelerograms. *Eng Struct.* 2010;32(9):2776-2792.
26. Huang YN, Whittaker AS, Luco N, Hamburger RO. Scaling earthquake ground motions for performance-based assessment of buildings. *J Struct Eng.* 2011;137(3):311-321.
27. Zachareniak A, Fragiadiakis M, Assimaki D, Papadarakakis M. Bias assessment in incremental dynamic analysis due to record scaling. *Soil Dyn Earthquake Eng.* 2014;67:158-168.
28. Dávalos H, Miranda E. Evaluation of bias on the probability of collapse from amplitude scaling using spectral-shape-matched records. *Earthquake Eng Struct Dyn.* 2019;48(8):970-986.
29. Dávalos H, Miranda E. Evaluation of the scaling factor bias influence on the probability of collapse using Sa(T1) as the intensity measure. *Earthquake Spectra.* 2019;35(2):679-702. https://doi.org/10.1193/011018EQS007M
30. Du W, Ning CL, Wang G. The effect of amplitude scaling limits on conditional spectrum-based ground motion selection. *Earthquake Eng Struct Dyn.* 2019;48(9):1030-1044.
31. Kircher C, Deierlein G, Hooper J, et al. Evaluation of the FEMA P-695 Methodology for Quantification of Building Seismic Performance Factors. 2010.
32. AISC 341-05. *Seismic Provisions for Structural Steel Buildings.* American Institute for Steel Construction; 2005.
33. ASCE/SEI 7-05. *Minimum Design Loads for Buildings and Other Structures.* American Society of Civil Engineers/Structural Engineering Institute; 2006.
34. AISC 358-05. *Prequalified Connections for Special and Intermediate Steel Moment Frames for Seismic Applications.* American Institute for Steel Construction; 2005.
35. McKenna F. OpenSees: a framework for earthquake engineering simulation. *Comput Sci Eng.* 2011;13(4):58-66.
36. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake Eng Struct Dyn.* 2005;34(12):1489-1511.
37. Lignos DG, Krawinkler H, Whittaker AS. Prediction and validation of sidesway collapse of two scale models of a 4-story steel moment frame. *Earthquake Eng Struct Dyn.* 2011;40(7):807-825.
38. Lignos DG, Krawinkler H. Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading. *J Struct Eng.* 2011;137(11):1291-1302.
39. Lignos DG, Krawinkler H. Development and utilization of structural component databases for performance-based earthquake engineering. *J Struct Eng.* 2013;139(8):1382-1394.
40. Gremer N, Adam C, Medina RA, Moschen L. Vertical peak floor accelerations of elastic moment-resisting steel frames. *Bull Earthquake Eng.* 2019;17(6):3233-3254.
41. Ancheta TD, Darragh RB, Stewart JP, et al. NGA-West2 database. *Earthquake Spectra.* 2014;30(3):989-1005.
42. Perez V. Spectra of amplitudes sustained for a given number of cycles: An interpretation of response duration for strong-motion earthquake records. *Bull Seismol Soc Am.* 1980;70(5):1943-1954. https://doi.org/10.1785/0119800051943
43. Bijelić N, Lin T, Deierlein GG. Validation of the SCEC Broadband Platform simulations for tall building risk assessments considering spectral shape and duration of the ground motion. *Earthquake Eng Struct Dyn.* 2018;47(11):2233-2251.
44. Bijelić N, Lin T, Deierlein GG. Efficient intensity measures and machine learning algorithms for collapse prediction of tall buildings informed by SCEC CyberShake ground motion simulations. *Earthquake Spectra.* 2020;36(3):1188-1207. https://doi.org/10.1177/8755293020919414
45. Bijelić N, Lin T, Deierlein GG. Quantification of the influence of deep basin effects on structural collapse using SCEC cybershake earthquake ground motion simulations. *Earthquake Spectra.* 2019;35:1845-1864.
46. Tsoulosidis KT, Bijelić N, Adam C, Moschen L. IDA-based seismic collapse patterns and their predictability by generalized linear models. In: *17th World Conference on Earthquake Engineering (17WCEE).* Sendai, Japan; 2020.
47. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol.* 1996;58(1):267-288.
48. Tsoulosidis KT, Moschen L, Adam C. Efficient selection of up to three-component ground excitation sets for earthquake engineering applications. *Soil Dyn Earthquake Eng.* 2021;148:106734. https://doi.org/10.1016/j.soildyn.2021.106734

**How to cite this article:** Tsoulosidis KT, Adam C. Amplitude scaling of ground motions as a potential source of bias: Large-scale investigations on structural drifts. *Earthquake Engng Struct Dyn.* 2022;51:2904–2924. https://doi.org/10.1002/eqe.3707