17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

WHY DETERMINING THE NUMBER OF CODE SPECTRUM-MATCHED RECORDS BASED ON USUAL STATISTICS IS AN ILL-POSED PROBLEM

G. Baltzopoulos[(1)], R. Baraschino[(2)], M. Giorgio[(3)], I. Iervolino[(4)]

[(1)] *Asst. Prof., Università degli Studi di Napoli Federico II, georgios.baltzopoulos@unina.it*
[(2)] *Phd student, National Technical University of Athens, roberto.baraschino@unina.it*
[(3)] *Prof., Università degli Studi di Napoli Federico II, massimiliano.giorgio@unina.it*
[(4)] *Prof., Università degli Studi di Napoli Federico II, iunio.iervolino@unina.it*

Abstract

The most advanced analytical tool in earthquake engineering is non-linear dynamic analysis, which typically entails a computer model of a structure subjected to a set of real recorded accelerograms. Although computationally demanding, this type of analysis is gradually becoming the norm in seismic design and assessment applications. Seismic reliability calculations according to the paradigm of performance-based earthquake engineering (PBEE) use dynamic analysis results to estimate the probabilistic distribution of structural response. In this context, the uncertainty in estimation of structural seismic risk depends, among other things, on the number of records used. On the other hand, modern seismic codes that espouse PBEE principles, but do not incorporate its probabilistic framework, do not require explicit estimates of structural seismic risk and acquiesce to the use of dynamic analysis for the purpose of pass/fail verifications of a single performance level, based on comparing average response vs fixed thresholds. While structural seismic reliability studies use a few tens of records, codes often require no more than seven to eleven records for obtaining an estimate of mean response, mandating that these records be selected based on criteria of compatibility with the design spectrum. The present study addresses this issue by investigating the effect of spectrum-compatible acceleration records' sample size on estimating average *single-stripe* inelastic structural response. To this end, multiple record sets are assembled, with a fixed number of records that are all selected to match-on-average the same target design spectrum, without repeating any record among sets. This exercise of obtaining multitudes of spectrum-compatible sets is repeated for various record sample sizes. The record sets are subsequently used for response-history analysis of a code-conforming inelastic frame. Examination of the results reveals that the spectral compatibility condition leads to response statistics that do not exhibit the trends expected in the case of simple random samples of various sizes. In fact, the responses obtained by using the selected spectrum-compatible records cannot be considered neither independent nor identically distributed. Thus, such statistics do not provide estimates of the intended characteristics of the underlying distribution. In other words, looking at the estimation uncertainty in the response distribution's parameters, by using statistics only suitable for cases of simple random sampling, is an ill-posed problem, since the sampled distribution may be very different from the one implicitly assumed by this approach. In conclusion, when code-based record selection is of concern, using tools of statistical inference such as the assessment of estimation uncertainty, suitable in cases of simple random sampling, to determine the necessary number of spectrum-compatible records, may be conceptually inappropriate.

Keywords: spectrum-compatible records, non-linear dynamic analysis, ground motion record selection.

1. Introduction

The use of non-linear dynamic analysis is fast becoming the norm for probabilistic seismic risk/loss analyses according to the performance-based earthquake engineering paradigm (PBEE). On the other hand, modern seismic codes are more reticent in imposing the requirement of non-linear response history analysis for design or assessment purposes, mostly reserving such high-end numerical tools for unusual or innovative structural systems or critical infrastructure. An interesting comparison, between the two cases of using dynamic analysis for probabilistic seismic risk assessment or code-based seismic design, is in the number of ground acceleration records typically involved, which determines the number of analyses to run.

On one hand, PBEE applications employ probabilistic representations of how structural seismic demand scales with shaking intensity; fitting such structure-specific models normally requires a relatively high number of non-linear runs, because a multitude of acceleration records should be used to represent each of several shaking intensities (e.g., [1,2]). Examples of such probabilistic (surrogate) models, are the, so-called, fragility functions that provide the conditional probability of the structure exceeding a performance objective given a specific value of a seismic intensity measure (IM). On the other hand, it is typical for modern seismic design codes to require performing seismic structural verification at a single (or few) intensity level(s). Running response-history analysis at a single intensity level is sometimes referred to as a single-stripe analysis. Furthermore, code-mandated design verification is usually of a binary nature; i.e., a pass/fail check is performed by comparing mean seismic response within the stripe against a preset allowable value.

Historically, modern-era seismic codes have been quite frugal with respect to the amount of required dynamic analyses, partly to avoid overtaxing the computational capabilities of the average practicing engineer and partly due to the difficulty engineers found in accessing suitable repositories of recorded strong motion. Nowadays, the latter is less of an issue, thanks to large online databases (e.g., the ESM [3] and NGA-West2 [4]), but the former is still a concern, as the state-of-the-art in numerical modelling continues to rise in complexity at a rate that rivals increasing computing power. Eurocode 8 (EC8, [5]) imposes a minimum of three design-spectrum-compatible records to be used for non-linear response-history analysis then raises that to seven if the mean response is to be considered for design verifications (maximum response is used in the case of three records). In this context, spectrum-compatibility of a record set is taken to signify having an average of spectral pseudo-acceleration ordinates, $Sa(T)$, closely matching those of the code's design spectrum. Along the same lines, the ASCE/SEI 7-16 standard [6] allows matching a spectrum corresponding to a, so-called, *risk-targeted maximum considered earthquake*, using a minimum of eleven records, which constitutes an increase from the seven required by earlier versions.

Despite the practical limitations to the number of records used in dynamic analysis, stemming from the complexity of numerical modelling and requirements of long computer times and person-hour resources, which are equally valid for both probabilistic risk assessment and code-based seismic design, consistently larger record sets are typically used in the former case with respect to the latter. This is partly due to the realization that the number of records used in the analysis, also determines the estimation uncertainty in the probabilistic description of structural response, which is then reflected onto the seismic risk metrics, such as the annual failure rate or expected annual loss. As highlighted in [7], because fragility functions are inferred from finite-size samples of structural response, they are only estimates of a hypothetical true model, and are therefore affected by estimation uncertainty.

In fact, past PBEE-oriented research has used statistical inference as a tool for discussing the number of records that ought be used for estimating the annual failure rate [8], or the distribution of structural response at a single level (stripe) of intensity [1,9]. On the other hand, statistical methods have also been employed to study the effect of the number of records under various proposals for spectral-matching and scaling of accelerograms [10] or specifically for evaluating the ASCE/SEI-7 record selection and scaling procedure [11]; i.e., in the case of code-oriented single-stripe analysis.

The present article takes a step backwards with respect to these studies and asks another question: are the adopted response statistics an appropriate tool for addressing the minimum number of records mandated by modern codes and for reconciling the disparity with the sample sizes typically used in seismic risk analyses?

To try to answer this question, the remainder of this article is structured so that first an overview of single-stripe dynamic analysis in the context of modern seismic code provisions is given, followed by a presentation of spectrum-compatible record selection. Subsequently, the responses of a simple inelastic structure to the multitude of record selections conducted is presented and their statistics are put under scrutiny to support the premise provided in the title. Finally, some concluding remarks are provided.

2. Single-stripe analysis and seismic code provisions

Structures are typically designed to enter into the range of inelastic response, during an earthquake that causes a certain shaking intensity to be exceeded. Said design usually entails satisfying specific performance objectives, such as damage reparability or safety of human life. In structural analysis, failure to meet such seismic performance goals is often defined on the basis of exceeding some threshold value of a global measure of inelastic displacement, such as roof drift ratio (maximum displacement of a building at roof level divided by total height – RDR) or maximum interstory drift ratio (IDR). Such response measures are sometimes generically termed *engineering demand parameters* (EDPs) and their failure-threshold values can be denoted by edp_f.

In seismic risk assessment, single-stripe dynamic runs of a structure's computer model, aim at estimating the probability of failure at some specific ground motion intensity measure (IM) level (im), $P\left[EDP > edp_f \,|IM = im\right]$, which can, in turn, involve estimating the conditional mean and variance of the EDP at that intensity, $\hat{\mu}_{EDP|IM=im}$ and $\hat{\sigma}^2_{EDP|IM=im}$, where the hat symbols are used to denote both the estimator and the point-estimates of the parameters of an underlying distribution. In fact, point estimates for the two parameters can be respectively obtained from the arithmetic mean, $\bar{x}$, and mean squared error, s^2, of the sample of EDP values in the stripe given by Eq. (1):

$$\begin{cases} \bar{x} = n^{-1} \cdot \sum_{j=1}^{n} edp_j \\ s^2 = (n-1)^{-1} \cdot \sum_{j=1}^{n} \left(edp_j - \bar{x}\right)^2 \end{cases}, \tag{1}$$

where n is the number of available responses (and records), edp_j, $j = \{1,..,n\}$ are the structural responses given that $IM = im$. In cases where $\hat{\mu}_{EDP|IM=im}$ is estimated as the arithmetic mean of the stripe's sample of n EDP responses, as per the equation, the *standard error* (SE) of $\hat{\mu}_{EDP|IM=im}$ can be approximated by $s/\sqrt{n}$ (approximated in the sense that the point estimate of the standard deviation is used in lieu of the true value $\sigma_{EDP|IM=im}$; [12]). This well-known result of statistical inference theory, is often used to highlight the importance of adopting *efficient* intensity measures; i.e., IMs that tend to reduce the conditional variance $\hat{\sigma}^2_{EDP}$ (henceforth omitting for brevity the condition $IM = im$, which is left implied), and consequently reduce the number of runs/records needed to maintain a desired SE for $\hat{\mu}_{EDP}$, since the required number is proportional to $\hat{\sigma}^2_{EDP}$ (e.g., [9]). The records used are typically scaled to the desired IM level and may also be selected on the basis of approximately representing the conditional distribution of spectral ordinates given $IM = im$ (e.g., [13]); in both cases each record is assumed to represent a possible manifestation of future shaking at the site and the corresponding structural response is considered a random sample.

Single-stripe analysis is also used during code-based seismic design, where dynamic analysis is typically only needed for the verification of a single performance objective, which is associated with a specific return period of the seismic actions. Whenever modern seismic codes, such as EC8 and ASCE/SEI 7, allow or require nonlinear dynamic analysis to be employed for seismic design in this manner, this usually entails comparing the average of the EDP values, which are obtained from a number of non-linear runs that use spectrum-compatible records, with a permissible value that can still be denoted as edp_f, as shown in Fig. 1. The evident similarities of this code-mandated procedure with a single-stripe analysis in PBEE context, may lead to the reflex reaction of calculating the sample statistics $\bar{x}$ and s^2 from Eq. (1) and of treating $s/\sqrt{n}$ as the SE of some $\hat{\mu}_{EDP}$, with the consequent repercussions on perceived accuracy of the estimate of the mean. However, the question begs to be asked: is the implicit assumption that $\bar{x}$ and s^2 are statistics of a simple random sample of EDPs (and therefore point estimates of the mean and variance of an underlying distribution) still valid under the *spectrum compatibility* condition imposed on the records?

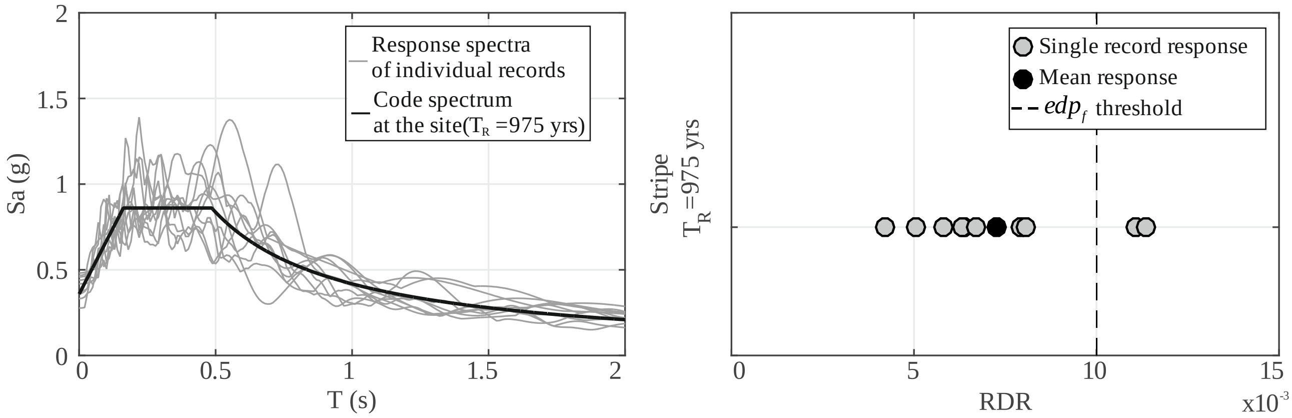

Fig. 1 – Records selected using the Italian code's EC8-style design spectrum with 5% probability of exceedance in 50 years (i.e., return period of 975 years) and corresponding stripe of structural responses.

At this point, consideration should be given to the issue of spectrum compatibility that standards, such as EC8 or ASCE/SEI 7, require for the record sets. While the codes themselves do not quantify how closely the mean spectrum of the selected records should match the target (bar some lower bound limit imposed on the mean), dedicated practice- and research-oriented software (e.g., [14–17]) tend to operate on a basis of best-fit-possible, employing various optimization algorithms, mainly limited by the size of ground motion database available. It is also noteworthy that the codes do not impose any quantitative limitation on the variability that single records may exhibit around the target. Regarding the lower bound, EC8 stipulates that the mean spectrum of the selected record set should not undercut the design spectrum by more than ten percent at any period within the range of interest. It also contains a provisional clause that mentions scaling recorded ground motions to the site-specific design peak ground acceleration (PGA), but past research advises against this for medium-to-long period structures from as early as [1] and it is not strictly necessary to do so for achieving a good overall match to the target spectrum.

3. Selection of spectrum-compatible record sets

For the purposes of this study, a ground motion pool of almost three-thousand single-component acceleration records was assembled from within the ESM and NGA-West2 databases. These records came from seventy-eight worldwide shallow crustal events of moment magnitude ranging from 5.4 to 8.0 (seventy-five events from ESM and three from NGA-West2, with no overlap between databases). From within this strong motion dataset, multiple code-spectrum-compatible sets were selected. The target spectrum considered for record selection was the one shown in Fig. 1, i.e., the Italian code's EC8-style design elastic spectrum with 5%

exceedance probability in 50 years[1] at an Italian site near the town of L'Aquila with soil class B. This spectrum is for all purposes equivalent to an EC8 spectrum.

The goodness-of-fit metric adopted for quantifying the proximity of a single scaled record's spectrum to the target spectrum is denoted as δ_j, and is given by Eq. (2):

$$\delta_j = \sqrt{\frac{1}{w} \cdot \sum_{i=1}^{w} \left(\frac{SF_j \cdot Sa_j(T_i) - Sa_{TARGET}(T_i)}{Sa_{TARGET}(T_i)} \right)^2} \quad , \tag{2}$$

where δ_j is the goodness-of-fit of the generic *j*-th record of the set, $Sa_{TARGET}(T_i)$ and $Sa_j(T_i)$ are the spectral acceleration values at period T_i of the target (code) spectrum and of the selected record, respectively, SF_j is the scale factor determined for that record and w is the total number of vibration periods considered. In this study, goodness-of-fit was evaluated at the same periods T_i as those used in [14], while the period interval, in which the spectrum matching conditions ought to be met, was $T \in [0, 2s]$, where it is implied that $Sa(T = 0s)$ denotes the PGA. In this light, scaling the records to the target PGA becomes moot, since that ordinate is also included in the matching interval.

Matching scaled records to the target spectrum was based on minimizing the sum of individual-to-target distances, according to Eq. (3):

$$\left\{ Sa_j(T_i), SF_j \right\} = \underset{Sa_j(T_i), SF_j}{\arg\min} \left\{ \sum_{j=1}^{n} \sqrt{\frac{1}{w} \cdot \sum_{i=1}^{w} \left(\frac{SF_j \cdot Sa_j(T_i) - Sa_{TARGET}(T_i)}{Sa_{TARGET}(T_i)} \right)^2} \right\} = \underset{Sa_j(T_i), SF_j}{\arg\min} \left\{ \sum_{j=1}^{n} \delta_j \right\}, \tag{3}$$

where $\{Sa_j(T_i), SF_j\}$, $i = \{1,..,w\}$, $j = \{1,..,n\}$ is the set of spectral ordinates and scale factors that together fully define the scaled record set selected. In this approach, the SF to apply to each of the accelerograms in the ground motion pool is part of the optimization process. The numerical problem of minimizing the sum of distances, $\sum_{j=1}^{n} \delta_j$, was solved using the Monte-Carlo-based algorithm proposed in [16] via a suitable modification of the code provided therein. In order to limit potential bias, that could be induced in the estimate of seismic response from scaling the records ([18,19]), a maximum admissible scale factor of ten was imposed.

It clearly emerges from the above that this algorithm tends to search for a fit of the record set to the target, by minimizing the dispersion of the single scaled records around it. On the other hand, an explicit measure of the distance between the mean of the selected records and the target, δ_m, can be provided by Eq. (4):

$$\delta_m = \sqrt{\frac{1}{w} \cdot \sum_{i=1}^{w} \left\{ \frac{1/n \cdot \sum_{j=1}^{n} \left[SF_j \cdot Sa_j(T_i) \right] - Sa_{TARGET}(T_i)}{Sa_{TARGET}(T_i)} \right\}^2} \quad . \tag{4}$$

Although the optimization process given by Eq. (3) provides a fit between the mean spectrum of the selection and the target only implicitly, since $\sum_{j=1}^{n} \delta_j$ is minimized rather than δ_m, it has the advantage of maintaining dispersion of spectral ordinates within each set as low as possible. As already mentioned, the codes have no explicit requirements for limiting said dispersion, but some authors have advocated keeping it as low as possible [14].

[1] Strictly speaking, only the PGA ordinate of the spectrum corresponding to rock site conditions has exactly that probability of exceedance; other spectral ordinates may correspond to slightly different probabilities, since the code spectrum's form is only an approximation of a uniform hazard spectrum.

The matching process provided forty-five non-overlapping sets for a series of specific sample sizes $n=\{3,5,7,10,15,17,20\}$. In other words, for every sample size n considered, the records contained in one set are never repeated in any other set of the same size; this was achieved by excluding from the pool the records already belonging to sets constructed during previous selections. The number of forty-five distinct record sets for every n, was an upper limit that was imposed due to database limitations and the need to maintain a minimum goodness-of-fit over all selections, which translates into maintaining the mean squared normalized deviation from the target spectrum, $\delta = n^{-1} \cdot \sum_{j=1}^{n} \delta_j$, below 0.40 in all cases. In Fig. 2, δ is plotted for all selected records at $n=\{3,7,20\}$ against order of selection.

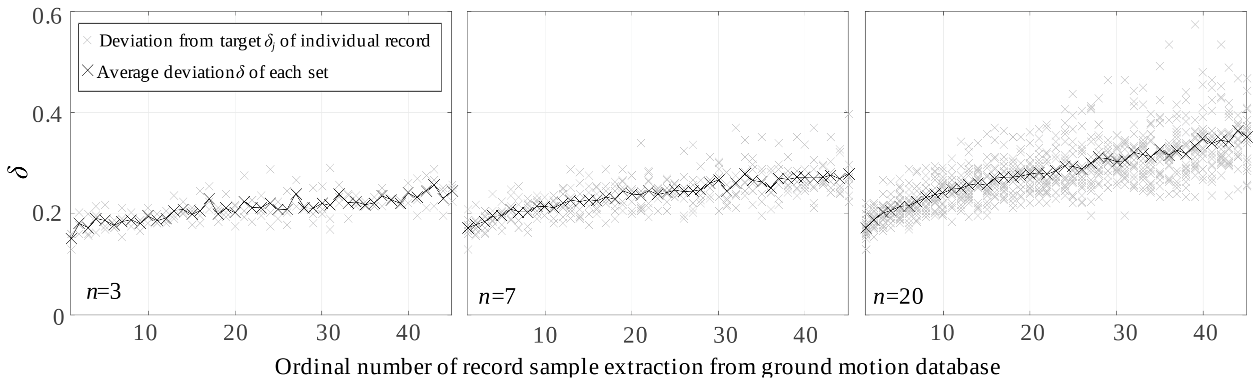

Fig. 2 – Goodness-of-fit (normalized deviation from target spectrum) for code-compatible record selections against extraction number for sample size of three, seven and twenty.

It is evident that δ tends to increase with each subsequent selection, which is to be expected, as the records that led to the best fit of all previous selections are removed from the database to avoid overlap. The fact that this trend is not strictly monotonic is due to the Monte-Carlo basis of the optimization algorithm used to implement Eq. (3). In Fig. 3, the target spectrum is shown, along with record sets consisting of three, seven and twenty ground motions, all coming from the multiple extractions described above. For sample sizes of three and seven, the selected sets exhibiting the lowest squared normalized deviation from the target δ, and those exhibiting the best fit to them mean, i.e., lowest δ_m, are shown for comparison.

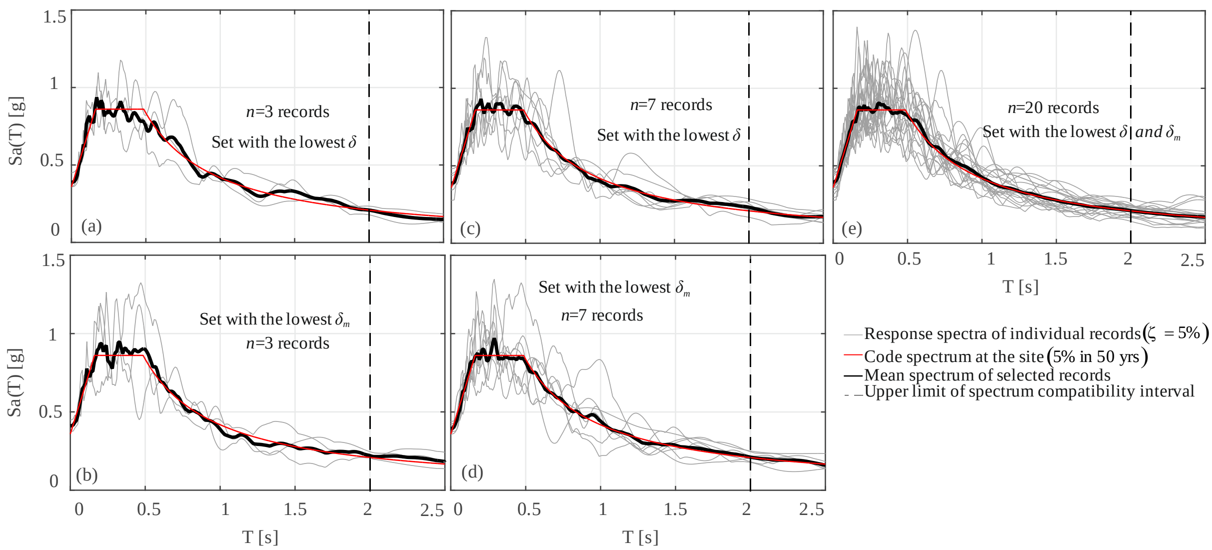

Fig. 3 – Spectrum-compatible ground motion suites composed of three records (a-b), seven records (c-d) and twenty records (e).

From Fig. 2 it can be seen that the intra-set variability of spectral shape generally tends to diminish for smaller sample sizes. It can also be observed from the same figure, that δ tends to increase going to the right (for subsequent record selections), which is an effect of database depletion and explains why the selections were limited to forty-five sets, maintaining $\delta < 0.40$. On the other hand, Fig. 3 attests to the fact that minimizing the distance of individual records from the target achieves a good fit of their mean as well, since there is hardly any discernible difference in the goodness-of-fit between panels a and b, which show the best individual record and best mean fit, respectively, for $n=3$ and likewise between c and d for $n=7$ (for twenty records δ and δ_m are minimized for the same selection). Furthermore, this figure also confirms the previous observation, that the intra-set variability of spectral ordinates tends to decrease when less records are used to achieve the same goodness-of-fit between their mean and the target spectrum.

4. Impact of the records' set size on the response statistics

The three-hundred and fifteen record suites assembled according to the procedure described in the previous section (seven set sizes times forty-five sets) were used to run response-history analysis of a non-linear numerical model of a four-storey plane, code-conforming reinforced concrete frame with first mode period $T_1 = 0.53s$ (see Fig. 4 and [20] for more information on the structure and detailing). The sample intra-set means of the IDR and RDR responses that were obtained for each record set, denoted respectively as μ_{IDR}^{intra}, μ_{RDR}^{intra}, are shown in Fig. 5. These were calculated as the arithmetic means $\bar{x}$, of the corresponding EDP, via Eq. (1).

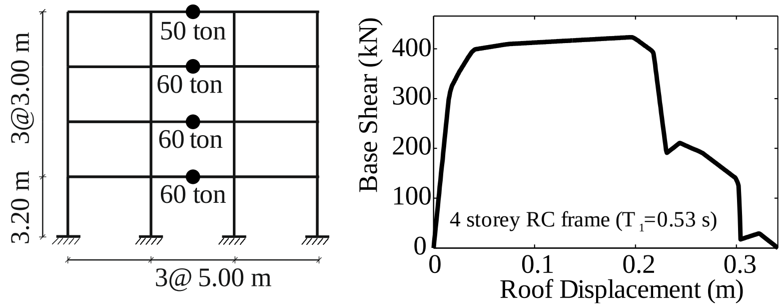

Fig. 4 – Basic dimensions and static pushover curve for the code-conforming, four-story, inelastic MDOF reinforced concrete moment-resisting frame structure used in the example.

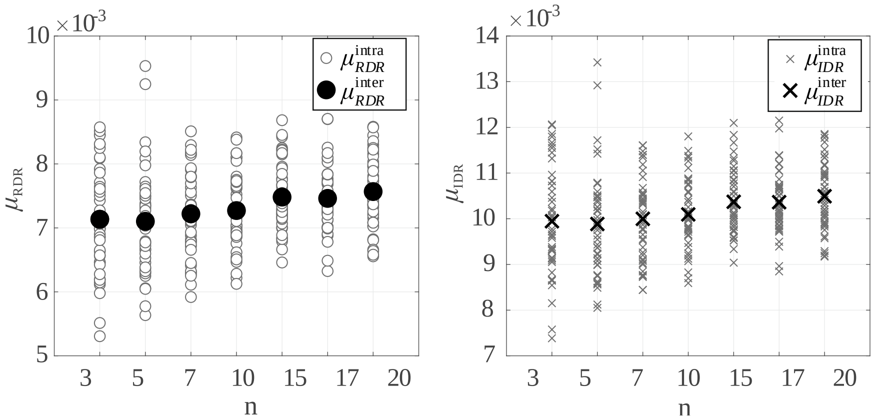

Fig. 5 – Intra- and inter-set response means (left panel RDR, right panel IDR), for the code-conforming reinforced concrete frame, plotted against sample size.

Also shown in the figure are the inter-set means, μ_{IDR}^{inter} and μ_{RDR}^{inter}, calculated as the average of the intra-set means, according to Eq. (5):

$$\mu_{EDP}^{\text{inter}} = m^{-1} \cdot \sum_{k=1}^{m} \mu_{EDP,k}^{\text{intra}}, \tag{5}$$

where m is the total number of record sets available ($m = 45$ in this case) and $\mu_{EDP,k}^{\text{intra}}$ is the arithmetic mean of IDR or RDR responses of the *k*-th record set, $k = \{1,..,m\}$. The relatively low average drift values can be attributed to overstrength and capacity design, which lead to moderate plastic rotations exclusively at the beam ends for this return period of seismic actions.

It can be noted that previous studies report that code (i.e., uniform hazard) spectrum-matched record sets tend to overestimate the central tendency of inelastic response, also when the records are scaled to a common $Sa(T)$ ordinate ([11,21,22]). Be that as it may, it can be observed from the figure that, despite some small fluctuations, the inter-set averages μ_{IDR}^{inter} and μ_{RDR}^{inter} do not appear to vary much with sample size up to $n = 10$, but do exhibit a small increase for larger sample sizes. In fact, Fig. 6 indicates that, for record sample sizes $n = \{3,7\}$, where the sets are expected to share a significant amount of records between the two size groups, the intra-set average IDR responses, μ_{IDR}^{intra}, obtained from progressively selecting new sets of increasing δ, fall around the inter-set average without exhibiting any evident strong trends with the selection order. On the other hand, for $n = 20$, there appears to be some increasing trend of μ_{IDR}^{intra} with the selection order of the corresponding record set. In fact, μ_{IDR}^{inter} for the twenty-record sets is larger than that of the lower sample sizes of thee and seven; this is a consequence of the increasing variance of spectral ordinates around the target for $n = 20$ which also affects the mean, due to the nonlinear relationship between the spectra and the inelastic response (this can also be appreciated by looking at the first couple of selections at $n = 20$, whose spectra are expected to be as close to the target as are their $n = 3$ counterparts, and whose mean is also similar). From this observation it follows that if the goal of the analysis is to evaluate the average response conditional to the code design spectrum, there may be a bias-inducing effect of the record set size.

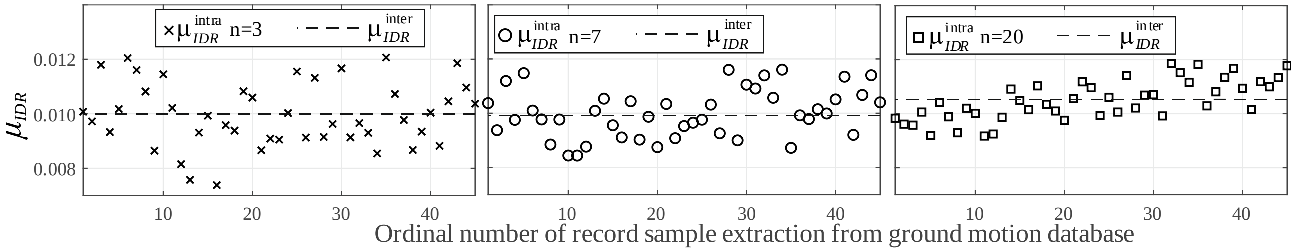

Fig. 6 – Variation of intra-set mean IDR response with selection order of the corresponding record set, for set size of three, seven and twenty records.

At first glance, this relatively stable behavior, at least the one exhibited by the smaller sample sizes, can tempt the observer to mistake the inter-set average μ_{IDR}^{inter} for an estimator of the mean of an underlying distribution. But if this were the case, one would also expect the sample standard deviations of the IDR responses from Eq. (1), s_{IDR}, to also vary randomly around some central value. However, the fact of the matter is that is not the case at all, as shown in Fig. 7. In this figure, the calculated s_{IDR} values for each set were plotted against the record selection order, as was previously done for the goodness-of-fit measure (Fig. 2) and for the means (Fig. 5). This plot is repeated for the coefficient of variation, $CoV_{IDR} = s_{IDR} / \mu_{IDR}^{\text{intra}}$. On both of these graphs, the least-square regression lines of s_{IDR} and CoV_{IDR} against selection order $m = \{1,2,..,45\}$ are displayed, with separate regressions performed for each record set size of $n = \{3,7,20\}$. Instead of a random variation around some central value, an increasing average trend with selection order is observed, where selection order is a proxy for decreasing goodness-of-fit, as shown in Fig. 2. Not only that, but the dispersion of structural response, expressed by either s_{IDR} or the normalized value CoV_{IDR}, also appears to increase, on average, with sample size; this is clearly indicated by the fact that the regression lines are arranged one underneath the other, with $n = 20$ leading with the larger average dispersions and the other two following in decreasing order with sample size.

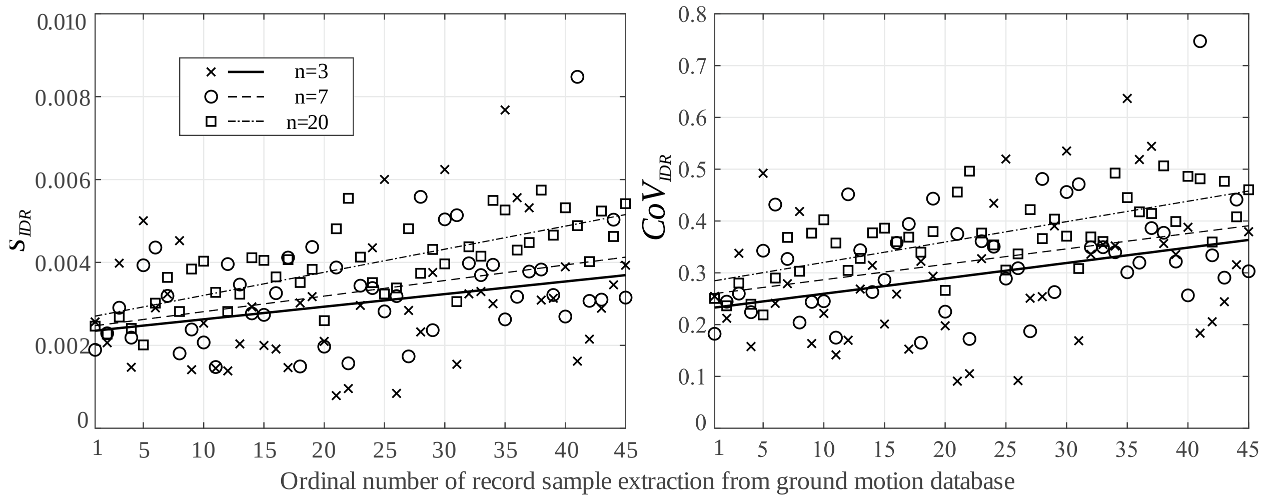

Fig. 7 – Standard deviation (left) and coefficient of variation of IDR (right) for each set of three, seven and twenty records, plotted against selection order and with a regression line displayed.

This behaviour implies a dependence of the dispersion of IDR from sample size, which means that it is not possible to predict the standard error of the average response of a larger set by using the mean squared error calculated from a smaller set via Eq. (1). An explanation for this, is that the structural responses from records selected according to the matching criteria in Eq. (3), do not represent a simple random sample of any distribution. In other words, the responses obtained from a record set of a certain size, are not independent realizations of a random variable. This is because matching the same spectrum induces a dependence between responses, due to the increased intra-set spectral similitude observed in Fig. 2 at smaller set sizes; given the important role of spectral shape in determining the distribution of inelastic-displacement-related EDPs (e.g., [23,24]) this lower dispersion in inter-set spectral shape can carry over to the responses. This correlation of spectral shape and inelastic response can also partly explain the increasing trend of s_{IDR} observed with respect to the selection order: since subsequent selections excluded all previously extracted records to avoid overlapping (i.e., avoid the presence of the same record in more than one set of the same size), the best-fit candidates are gradually removed from the selection pool. Consequently, the dispersion of the records' spectral ordinates from the target increases in subsequent selections and the dispersion of inelastic response also reflects that increase.

However, the issue remains that the simple random sample assumption is a *sine qua non* condition, both for using sample statistics, such as those of Eq. (1), as estimators of the parameters of some underlying distribution, as well as for calculating the corresponding SE. Thus, one logical conclusion that can be drawn from what evidence was drawn from this example, is that the use of statistical inference tools, only suitable in the case of simple random sampling, to determine ground motion sample size in the context of code-mandated spectrum-compatibility may be an ill-posed problem, because the underlying assumption of the corresponding structural responses being independent and identically distributed does not appear to hold.

5. Concluding remarks

The present study investigated the issue of the number of code-spectrum-compatible records used for dynamic analysis, when the objective is estimating mean inelastic structural response. Spectrum-compatibility was defined as the property of a set of scaled acceleration records, whose mean spectrum is as good a match to the target code spectrum as possible. The premise that motivated this investigation, was that the use of well-known concepts from inference theory, such as estimation uncertainty, while suitable for determining the appropriate sample size of input ground motions in other apparently similar earthquake-engineering applications, may not be applicable in this case. In the context of this investigation, a Eurocode-8 type design spectrum for an Italian site was used as reference and numerous spectrum-compatible sets, without overlap of records between them, were selected from a large pool of three thousand acceleration waveforms. This selection was repeated for various sizes of the record set, between three and twenty. It was observed that record selection based on goodness-of-fit of the mean to the target, led to less inter-set variability of the spectral ordinates for the smaller sample sizes. Non-linear dynamic analysis was performed for the numerical model of a plane four-storey inelastic frame for all base-acceleration inputs selected. Examination of the obtained structural responses revealed that response statistics as a function of sample size did not behave as expected for random samples, i.e., the responses did not appear to be independent and identically distributed. This implies that determining the number of spectrum-compatible records to use in this context via statistical tools, such as the standard error of the mean, suitable in the case of simple random sampling, is an ill-posed problem. In other words, it may be conceptually inappropriate to invoke considerations of estimation uncertainty, when dealing with code-based record selection.

6. Acknowledgements

The work presented in this paper was developed within the activities of the 2019-2021 research project DPC-ReLUIS (*Rete di Laboratori Universitari d'Ingegneria Sismica* – in Italian), funded by the Italian Department of Civil Protection of the Presidency of the Council of Ministers.

7. References

[1] Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. Earthq Spectra 1998;14:469–500. https://doi.org/10.1193/1.1586011.

[2] Baker JW. Efficient analytical fragility function fitting using dynamic structural analysis. Earthq Spectra 2015;31. https://doi.org/10.1193/021113EQS025M.

[3] Luzi L, Puglia R, Russo E, D'Amico M, Felicetta C, Pacor F, et al. The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data. Seismol Res Lett 2016;87:987–97. https://doi.org/10.1785/0220150278.

[4] Bozorgnia Y, Abrahamson NA, Al Atik L, Ancheta TD, Atkinson GM, Baker JW, et al. NGA-West2 research project. Earthq Spectra 2014;30. https://doi.org/10.1193/072113EQS209M.

[5] CEN. EN 1998-1: Eurocode 8 - Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings. European Committee for Standardization; 2004.

[6] ASCE. ASCE/SEI 7-16 Minimum design loads for buildings and other structures. American Society of Civil Engineers; 2016.

[7] Iervolino I. Assessing uncertainty in estimation of seismic response for PBEE. Earthq Eng Struct Dyn 2017;13:1834–49. https://doi.org/10.1002/eqe.

[8] Baltzopoulos G, Baraschino R, Iervolino I. On the number of records for structural risk estimation in PBEE. Earthq Eng Struct Dyn 2019;48:489–506. https://doi.org/10.1002/eqe.3145.

[9] Cornell C. Hazard, ground motions and probabilistic assessments for PBSD. In: Fajfar P, Krawinkler H, editors. Int. Work. performance-based Seism. Des. concepts Implement., 2004, p. 39–52.

[10] Hancock J, Bommer JJ, Stafford PJ. Numbers of scaled and matched accelerograms required for inelastic dynamic analyses. Earthq Eng Struct Dyn 2008;37:1585–607. https://doi.org/10.1002/eqe.827.

[11] Reyes JC, Kalkan E. How many records should be used in an ASCE/SEI-7 ground motion scaling procedure? Earthq Spectra 2012;28:1223–42. https://doi.org/10.1193/1.4000066.

[12] Mood AM, Graybill FA, Boes DC. Introduction to the Theory of Statistics. 3rd ed. McGraw-Hill; 1974.

[13] Lin T, Haselton CB, Baker JW. Conditional spectrum-based ground motion selection. Part I: Hazard

consistency for risk-based assessments. Earthq Eng Struct Dyn 2013;42:1847–65. https://doi.org/10.1002/eqe.2301.

[14] Iervolino I, Galasso C, Cosenza E. REXEL: Computer aided record selection for code-based seismic structural analysis. Bull Earthq Eng 2010;8:339–62. https://doi.org/10.1007/s10518-009-9146-1.

[15] Jayaram N, Lin T, Baker JW. A Computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. Earthq Spectra 2011;27:797–815. https://doi.org/10.1193/1.3608002.

[16] Baker JW, Lee C. An Improved Algorithm for Selecting Ground Motions to Match a Conditional Spectrum. J Earthq Eng 2017. https://doi.org/10.1080/13632469.2016.1264334.

[17] Katsanos EI, Sextos AG. ISSARS: An integrated software environment for structure-specific earthquake ground motion selection. Adv Eng Softw 2013;58:70–85. https://doi.org/10.1016/J.ADVENGSOFT.2013.01.003.

[18] Haselton, Curt B Deierlein GG. Assessing Seismic Collapse Safety of Modern Reinforced Concrete Moment-Frame Buildings, PEER Report 2007-08. Pacific Earthq Eng Res Center, Berkeley, CA 2007.

[19] Eads L. Seismic Collapse Risk Assessment Of Buildings: Effects Of Intensity Measure Selection and Computational Approach,Report No. 184. John A. Blume Earthquake Engineering Center, Stanford University, Stanford, CA., 2013.

[20] Baltzopoulos G, Chioccarelli E, Iervolino I. The displacement coefficient method in near-source conditions. Earthq Eng Struct Dyn 2015;44. https://doi.org/10.1002/eqe.2497.

[21] Lin T, Haselton CB, Baker JW. Conditional spectrum-based ground motion selection. Part II: Intensity-based assessments and evaluation of alternative target spectra. Earthq Eng Struct Dyn 2013;42:1867–84. https://doi.org/10.1002/eqe.2303.

[22] Goulet CA, Watson-Lamprey J, Baker JW, Haselton CB, Luco N. Assessment of ground motion selection and modification (GMSM) methods for non-linear dynamic analyses of structures. 4th Decenn. Geotech. Earthq. Eng. Soil Dyn. Conf. (GEESD IV), ASCE, May 18-22 2007, Sacramento, CA. ASCE Geotech. Spec. Publ. No. 181., 2008, p. 1–10. https://doi.org/10.1061/40975(318)3.

[23] Bojórquez E, Iervolino I. Spectral shape proxies and nonlinear structural response. Soil Dyn Earthq Eng 2011;31:996–1008. https://doi.org/10.1016/j.soildyn.2011.03.006.

[24] Baker JW, Cornell CA. A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. Earthq Eng Struct Dyn 2005;34:1193–217. https://doi.org/10.1002/eqe.474.

