Bulletin of the Seismological Society of America, Vol. 109, No. 5, pp. 2073–2087, October 2019, doi: 10.1785/0120180307

Ⓔ

Assessing Predictive Capability of Ground-Motion Models for Probabilistic Seismic Hazard in Iran

by Ali Farhadi, Zoya Farajpour, and Shahram Pezeshk

Abstract We assessed the applicability of several ground-motion models (GMMs) against Iran's local data. Candidate GMMs are selected from those developed for shallow crustal regions such as Iran, Turkey, Japan, Europe and the Middle East, and the western United States. We made the evaluation database prospective to all candidate GMMs to assess their predictive capability. The evaluation database is composed of 643 records from 240 earthquakes with magnitudes ranging from 3.9 to 7.3 and Joyner–Boore distances up to 300 km. We implemented the log-likelihood method of Scherbaum *et al.* (2009), the Euclidean distance-based ranking proposed by Kale and Akkar (2013), and the multivariate logarithmic score of Mak *et al.* (2017) to evaluate the candidate models. We ranked GMMs by paying attention to the issue of score variability. To assess the score variability, we generated resampled datasets from the whole database using the cluster bootstrap technique and ranked models based on their relative performance among bootstrap samples. Overall, Sedaghati and Pezeshk (2017), Zafarani *et al.* (2018), and Farajpour *et al.* (2019) local models outperform remaining models considering the whole database over the entire frequency range. For high-seismicity regions, the Zhao *et al.* (2006) model can be used in line with the first two local models to better quantify epistemic uncertainties associated with the process of model selection. In addition to aforementioned local models, Bindi *et al.* (2014) show acceptable performance against small-to-moderate magnitude data and may be considered for estimating seismic hazard in low-seismicity regions of Iran.

Supplemental Content: Tables providing information about the evaluation database including the magnitude, the depth and the number of records, log-likelihood (LLH) and Euclidean distance-based ranking (EDR) scores of the candidate models for all representative periods, and weights that can be assigned to the candidate models based on the LLH scores, and figures illustrating the distinctness tables computed from the multivariate logarithmic scores (mvLogSs) for representative spectral periods, and residuals versus the magnitude, distance, and shear-wave velocity on the top 30 m for periods other than peak ground acceleration (PGA) and visual comparisons between observed ground-motion intensities and models' predictions for spectral periods including 0.5, 1.0, and 2.0 s.

Introduction

Selecting appropriate ground-motion models (GMMs) is an important task in modeling seismic hazard for any study region (Kulkarni *et al.*, 1984; McGuire, 2004). Typically, a group of experts select multiple GMMs to address epistemic uncertainties associated with the process of model selection. A seismic hazard modeler may follow guidelines provided in Cotton *et al.* (2006) and Bommer *et al.* (2010) to preselect a set of viable models. Then, the modeler can use empirical GMM evaluations to assess the applicability of the candidate models against evaluation data and come up with a few suitable models for populating the branches of the logic tree.

The need for assessing models' relative performances prior to carrying out seismic hazard studies is crucial for the tectonic region of Iran due to the shortage of domestic experienced experts (Zafarani and Mousavi, 2014). To meet this need, several studies have been performed in recent years to introduce suitable models for different tectonic regions within the Iranian plateau. We divide these studies into two groups. The first group evaluated the applicability of foreign

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

models for Iran, and the second group considered local models in addition to foreign models in their list of the candidate GMMs. The first group includes Shoja-Taheri *et al.* (2010) and Mousavi *et al.* (2014), who evaluated the applicability of the Next Generation Attenuation (NGA) models to the local region of Iran. Shoja-Taheri *et al.* (2010) compared three NGA models with the Iranian strong-motion data using residual analysis. They proposed models of Boore and Atkinson (2008), Campbell and Bozorgnia (2008), and Chiou and Youngs (2008) as models applicable to the Iranian plateau. The approach used in Shoja-Taheri *et al.* (2010) does not take the variability of GMMs into account and could be replaced with the newer test of goodness-of-fit measures such as the likelihood (LH) and log-likelihood (LLH) methods (Scherbaum *et al.*, 2004, 2009). Mousavi *et al.* (2014) used the LH and LLH methods of ranking to test the NGA models against a database constructed from 2012 Ahar–Varzaghan dual earthquakes and 1997 Ardabil event. They proposed that the NGA models could be confidently used for seismic hazard studies of Iran for short-to-median periods. Considering a limited range of earthquakes may limit the implications of Mousavi *et al.* (2014) in a broader range of magnitudes and distances. Moreover, the NGA models used in the studies of Shoja-Taheri *et al.* (2010) and Mousavi *et al.* (2014) are now superseded by their newer versions (Next Generation Attenuation-West2 Project [NGA-West2] models) that may perform better than their predecessors against Iran's local data.

The second group of studies on the applicability of GMMs for Iran contains studies that evaluated not only the NGA models but also other foreign models and the local GMMs. We categorized three studies including Mousavi *et al.* (2012), Zafarani and Mousavi (2014), and Zafarani and Farhadi (2017) into this group. Mousavi *et al.* (2012) used LH and LLH methods to test the applicability of models selected from three groups, including indigenous models of Iran, regional models of Europe and the Middle East, and the NGA models for the Zagros region of Iran. They suggested that local models proposed by Zafarani and Soghrat (2012) and Ghasemi, Zare, Fukushima, and Koketsu (2009; hereafter, GZFK09) perform better than other candidate models for the Zagros region. In the study of Mousavi *et al.* (2012), the best-fitting models were constructed using a portion of data that the authors considered to rank the GMMs. Zafarani and Mousavi (2014) followed a procedure similar to that of Mousavi *et al.* (2012) to perform an independent study for northern Iran. They introduced few models from each of the three mentioned categories as appropriate. In their views, the models of Ghasemi, Zare, Fukushima, and Koketsu (2009) GZFK09 and Soghrat *et al.* (2012), both developed from the local data, and the two NGA models of Abrahamson and Silva (2008) and Chiou and Youngs (2008) as well as the regional model of Akkar and Bommer (2010) are appropriate for seismic hazard studies in the northern region of Iran. The issue of similarity between the generating datasets of the local models and evaluation data still exists in the study of Zafarani and Mousavi (2014). Zafarani and Farhadi (2017) excluded the data used in the development of local models from their evaluation database and tested several GMMs against small-to-moderate magnitude data in Iran. They introduced the local model of Zafarani *et al.* (2018; hereafter, ZLLS18) as a single model consistent with the observed data. The focus of Zafarani and Farhadi (2017) was on weak motions that may limit its implications for larger magnitudes with significant impact on seismic hazard studies.

Previous studies motivated us to extend their implications and perform the current study. To this end, we took the advantage of new methodologies and currently available data to improve the implications of previous studies for the future seismic hazard assessments in Iran. For example, we improved the results of Mousavi *et al.* (2014) as well as Zafarani and Farhadi (2017) by replacing their evaluation data with the most recent evaluation database that covers a wider magnitude–distance range. In this study, we used 643 recording motions from 240 earthquakes with magnitudes ranging from 3.9 to 7.3 and distances up to 300 km as our evaluation database. The evaluation data cover a broader range of events and better serves our objectives. In addition, we made the evaluation database independent from all candidate GMMs to assess the predictive capability of the candidate models rather than their explanatory capability. Therefore, we updated previous studies including Mousavi *et al.* (2012) and Zafarani and Mousavi (2014) by paying attention to these two types of performance. In other words, our results only describe the predictive capability that is directly linked to the seismic hazard (Bindi, 2017; Mak, 2017). Moreover, we tested recently established GMMs for Iran and newer versions of both regional and the NGA models against our dataset to update previous studies.

We implemented the popular LLH method, the Euclidean distance–based ranking (EDR) proposed by Kale and Akkar (2013), and the natural extension of LLH method known as the multivariate logarithmic score (mvLogS) of Mak *et al.* (2017) to rank the candidate models. The mvLogS method is sensitive to the information provided for the variability associated with empirical models (sigma). This method measures the relative performance of competing models considering all components of sigma provided by GMM developers, whereas the LLH and EDR methods rank models based on the total sigma. This is an important feature due to recent advances in ground-motion modeling, which has resulted in developing models with more complicated hierarchical structures. Recently established GMMs have a two-layer hierarchical structure and partition the total sigma into between-event and within-event components by considering ground-motion records from the same event as correlated.

Furthermore, we improved the result of previous studies by paying attention to the issue of score variability in ranking the candidate models. We took score variability into account by using the cluster bootstrap technique for generating resampled datasets from the original evaluation database.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

The cluster bootstrap techniques enabled us to compute the distinctness index (DI) from the resampled datasets for all model pairs. The DI indicates if two models are truly different given the score variability.

We will next introduce the evaluation data and give a summary on applied GMMs. Then, we will review the data-driven evaluation methods used to evaluate the performance of selected GMMs against the evaluation dataset. Finally, we will conclude by introducing a relatively short list of appropriate GMMs most suitable for the future seismic hazard studies of Iran.

Evaluation Database

We adopted the evaluation database from two sources. The main source is a subset of data compiled by Zafarani and Soghrat (2017). Zafarani and Soghrat (2017) performed a comprehensive study to collect ground-motion records from Iranian earthquakes between 1975 and 2014. Their database represents 20% of the Iranian earthquakes and contains 2286 records from 461 events with moment magnitude from 3.9 to 7.3. We discarded those records in Zafarani and Soghrat (2017) that were used by GMM developers in Iran. This makes the test data independent from the candidate GMMs. We supplemented this database by adding some ground-motion records from the recent M_w 7.3 Kermanshah earthquake. The Kermanshah earthquake struck the province of Kermanshah in the western region of Iran on 12 November 2017 at 18:18:16 UTC (21:48:16 local time). This devastating event was felt across the Iran–Iraq border with a maximum modified Mercalli intensity of 8.0 (Babaie Mahani and Kazemian, 2018). We added 44 strong ground motion records from the Kermanshah earthquake to our evaluation database to enrich it at large magnitudes. These records were analyzed by Babaie Mahani and Kazemian (2018) and are available online. Babaie Mahani and Kazemian (2018) processed 121 accelerometers, but we used only 44 of them by considering events with closest distance to the surface projection of the fault ruptured area (R_{JB}) less than 300 km and stations with known average shear-wave velocity on the top 30 m (V_{S30}). It should be noted that adding data from recent Kermanshah earthquake will not make the evaluation database dependent on the candidate models because these data have not been used by modelers in GMM development.

Overall, we used a database of 643 records from 240 earthquakes with magnitudes ranging from 3.9 to 7.3 and distances up to 300 km to evaluate the applicability of several GMMs for the probabilistic seismic hazard in Iran. Events with magnitude below 6 contribute the most to the evaluation database with 535 ground-motion records. Many events from this range have not been used in the development of the majority of the Iranian GMMs due to the higher importance of moderate-to-large events in seismic hazard studies of Iran. However, earthquake hazard from weak-motion data could be high especially in regions close to the location of earthquakes. Because of the importance of earthquake hazard from small-to-moderate magnitude data, the NGA-West2 model developers updated the NGA models to make them applicable to smaller magnitude data, taking advantage of the rich database provided by Ancheta *et al.* (2014). According to Atkinson and Assatourians (2017), events from the 3.5–6 range are damaging especially in distances up to 50 km. Therefore, we used 14 records from three M_w 3.9 earthquakes with distances below 50 km for ranking the candidate GMMs. This magnitude is smaller than the minimum magnitude thresholds of some of the GMMs, but the minimum magnitude considered for performing seismic hazard in some parts of Iran could be as small as M_w 3.9. In low-seismicity regions such as Arvand-Shatt-Al-Arab, Arabian platform, and Persian Gulf seismotectonic provinces of Iran, magnitudes as small as M_w 4.0 could be responsible for hazard in large return periods (Zafarani and Farhadi, 2017). Zafarani and Farhadi (2017) also mentioned the need for developing more elaborate GMMs that can provide reliable predictions for small-to-moderate magnitude earthquakes. The ZLLS18 model is a single Iranian GMM that used a considerable portion of small-to-moderate magnitude data in model development. It should be mentioned that the evaluation database is independent of the ZLLS18 model. Zafarani *et al.* (2018) have not used all small-to-moderate magnitude events provided by Zafarani and Soghrat (2017) in their model to avoid dominating their generating dataset with events from small-to-moderate magnitude range.

There are 108 ground-motion records with magnitude above 6 in the evaluation database. Forty-four records of these 108 ground-motion records are from recent Kermanshah earthquake and have not been considered in the development of existing Iranian GMMs. Therefore, there are 64 ground-motion records with magnitudes larger than 6 in the evaluation database that have not been used by all GMM developers, whereas many of them were available at the time of development of most of the Iranian GMMs. These records represent an important range of earthquakes, and we will explain why they have not been used by all model developers. Fifty-seven of these 64 ground-motion records were recorded by stations with estimated V_{S30}. Availability of an estimate of this parameter rather than its measurement could be the main reason for excluding these records from the evaluations performed by GMM developers. Nonetheless, estimated values of V_{S30} are still useful and Zafarani and Soghrat (2017) approximated the site condition for stations recorded these 57 records using robust methodologies proposed by Zare *et al.* (1999) and Ghasemi, Zare, Fukushima, and Sinaeian (2009). It is worth mentioning that Ghasemi, Zare, Fukushima, and Sinaeian (2009) assessed the applicability of their proposed method by reclassifying the stations with previously determined V_{S30}. Success rates of their method for the Iranian sites are comparable with those obtained by applying the approach proposed by Zhao *et al.* (2006; hereafter, Zetal06) and are higher than the rates gained using peak periods of horizontal-to-vertical spectral ratios. Further information regarding the evaluation database

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

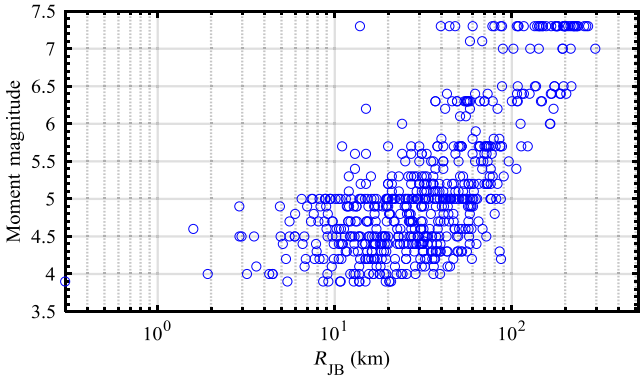

Figure 1. Magnitude–distance distribution of the evaluation database. The color version of this figure is available only in the electronic edition.

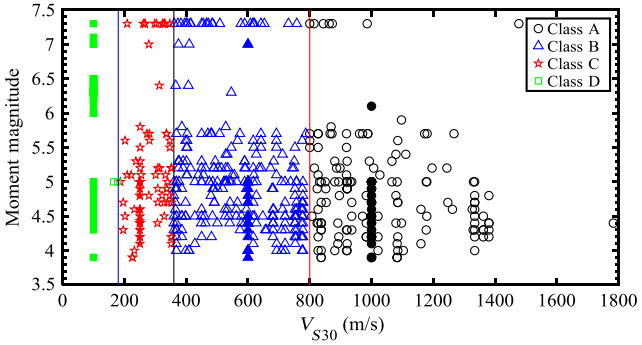

Figure 2. Distribution of magnitude data versus V_{S30}. Records with estimated value of V_{S30} are shown with filled markers. Sites are classified into four groups based on the Eurocode site classification scheme (Eurocode 8, 2004). The color version of this figure is available only in the electronic edition.

is provided in Ⓔ Table S1, available in the supplemental content to this article. Ⓔ Table S1 includes the magnitude, the depth, and the number of records per event.

Figure 1 shows the magnitude–distance distribution for the final evaluation database. According to this figure, the evaluation database is sparse within two ranges. For small-to-moderate magnitude events, the test data contain few records at distances greater than 100 km. The evaluation database is also limited for moderate-to-large magnitude data with distances up to 40 km. Figure 2 illustrates the distribution of magnitude data versus V_{S30}. In Figure 2, filled markers show recorded data at sites with an estimated value of V_{S30}. Zafarani and Soghrat (2017) grouped stations with unknown values of V_{S30} into four Eurocode 8 (2004) site classes: class A, $V_{S30} \geq 800$ m/s; class B, $360 \leq V_{S30} \leq 800$ m/s; class C, $180 \leq V_{S30} \leq 360$ m/s; and class D, $V_{S30} \leq 180$ m/s. They assigned V_{S30} equal to 1000, 600, 250, and 100 m/s to sites categorized in classes A, B, C, and D, respectively. Eurocode 8 (2004) site classification is very similar to the Iranian standard number 2800 (Building and Housing Research Centre [BHRC]) with four categories based on the V_{S30} measurements: (1) $V_{S30} \geq 750$ m/s, (2) $375 \leq V_{S30} \leq 750$ m/s, (3) $175 \leq V_{S30} \leq 375$ m/s, and (4) $V_{S30} \leq 175$ m/s.

Summary of the Considered Candidate GMMs

We selected the candidate models from five groups including indigenous models of Iran established from the local data (group 1), models developed based on Turkey earthquakes (group 2), regional models developed based on the Middle East and European data (group 3), global models based on the NGA-West2 database (group 4), and Japanese GMMs (group 5). All candidate models were developed for shallow crustal regions across the world and might be appropriate for the Iranian plateau with similar tectonic environment. Ghasemi, Zare, Fukushima, and Koketsu (2009), Sedaghati and Pezeshk (2017; hereafter, SP17), Shahidzadeh and Yazdani (2017), Zafarani *et al.* (2018), and Farajpour *et al.* (2019; hereafter, FPZ19) are considered indigenous models for Iran. Akkar and Çağnan (2010; hereafter, AC10) and Kalkan and Gülkan (2004; hereafter, KG04) are models developed for Turkey. The Kale *et al.* (2015; hereafter, KAAH15) model belongs to the first two groups and is applicable to both Turkey and Iran. Akkar *et al.* (2014; hereafter, ASB14), Bindi *et al.* (2014; hereafter, Betal14), and Kotha *et al.* (2016; hereafter, KBC16) used datasets made of events from the Middle East and European regions to establish their models. We considered four NGA-West2 models including Abrahamson *et al.* (2014), Boore *et al.* (2014), Campbell and Bozorgnia (2014), and Chiou and Youngs (2014). These are global models developed based mainly on the California data. In addition to these GMMs, we selected two other GMMs including the Zetal06 and Kanno *et al.* (2006; hereafter, Ketal06) models. These two GMMs were developed mainly for Japanese data, but their generating datasets include a good portion of ground-motion records from active crustal regions.

Table 1 gives more information about the candidate models including their magnitude–distance applicability range, the distance metric they use in their functional form, which group they belong to, dominant generating region, the definitions of component combination, and abbreviations used to represent each GMM. As is clear from Table 1, various models use different definitions of component combination. The evaluation database provides ground-motion records for geometric mean and the maximum of the two horizontal components. Therefore, the definition of peak motion used in the observed and predicted ground motions is not the same for some of the candidate GMMs. The difference between RotD50 and the GMRotI50 was on average about 3% for spectral periods below 3.0 s in Boore (2010). Van Houtte *et al.* (2017) reported the same for the difference between the geometric mean and the RotD50. Beyer and Bommer (2006) have shown that the ratio of geometric mean to GMRotI50 is near unity at all periods. Based on the mentioned studies and following Mak *et al.* (2018), we did not convert peak definitions into a single definition, to avoid

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

Table 1
Summary of the Candidate Ground-Motion Models

GMM Reference	Abbreviation	Group	Dominant Region(s)	Component	Period (s)	Moment Magnitude	Distance (km)
Abrahamson *et al.* (2014)	ASK14	4	California and western United States	PGA, PGV, PSA in RotD50	0.01–10.0	3.0–8.5	R_{rup} 0–300
Boore *et al.* (2014)	BSSA14	4	California and western United States	PGA, PGV, PSA in RotD50	0.01–10.0	3.0–8.5	R_{JB} 0–400
Campbell and Bozorgnia (2014)	CB14	4	California and western United States	PGA, PGV, PSA in RotD50	0.01–10.0	3.3–8.5	R_{rup} 0–300
Chiou and Youngs (2014)	CY14	4	California and western United States	PGA, PGV, PSA in RotD50	0.01–10.0	3.5–8.5	R_{rup} 0–300
Zafarani *et al.* (2018)	ZLLS18	1	Iran	PGA, PSA in GM	0.04–4.0	4.0–7.3	R_{JB} 0–200
Farajpour *et al.* (2019)	FPZ19	1	Iran	PGA, PSA in GM	0.01–4.0	4.5–7.5	R_{rup} 0–400
Ghasemi, Zare, Fukushima, and Koketsu (2009)	GZFK09	1	Iran	PSA in GMRotI50	0.05–3.0	5.0–7.4	R_{rup} 0–100
Sedaghati and Pezeshk (2017)	SP17	1	Iran	PGA, PGV, PSA in GM	0.05–4.0	4.7–7.4	R_{JB} 0–250
Shahidzadeh and Yazdani (2017)	SY17	1	Iran	PGA_{max}, PSA_{max}	0.05–2.5	5.0–7.3	R_{JB} 0–100
Kale *et al.* (2015)	KAAH15	1 and 2	Iran and Turkey	PGA, PGV, PSA in GM	0.01–4.0	4.0–8.0	R_{JB} 0–200
Akkar and Çağnan (2010)	AC10	2	Turkey	PGA, PGV, PSA in GM	0.03–2.0	5.0–7.6	R_{JB} 0–200
Kalkan and Gülkan (2004)	KG04	2	Turkey	PGA_{max}, PSA_{max}	0.10–2.0	4.0–7.4	R_{JB} 1.2–250
Akkar *et al.* (2014)	ASB14	3	Europe and the Middle East	PGA, PGV, PSA in GM	0.01–4.0	4.0–8.0	R_{JB} 0–250
Bindi *et al.* (2014)	Betal14	3	Europe and the Middle East	PGA, PGV, PSA in GM	0.02–3.0	4.0–7.6	R_{JB} 0–300
Kotha *et al.* (2016)	KBC16	3	Europe and the Middle East	PGA, PGV, PSA in GM	0.01–4.0	4.0–7.6	R_{JB} 0–200
Kanno *et al.* (2006)	Ketal06	5	Japan	PGA, PGV, PSA in GM	0.05–5.0	5.2–8.2	R_{rup} 0–300
Zhao *et al.* (2006)	Zetal06	5	Japan	PGA, PSA in GM	0.05–5.0	5.0–7.3	R_{rup} 0–300

GMM, ground-motion model; GMRotI50, rotation-independent average horizontal component (Boore *et al.*, 2006); max and GM, maximum and geometric mean of horizontal components, respectively; RotD50, orientation-independent nongeometric-mean measure (Boore, 2010); R_{JB}, closest distance to the surface projection of fault plane; R_{rup}, closest distance to fault plane.

introducing another factor that may affect the result. Mousavi *et al.* (2015) and Akhani *et al.* (2019) GMMs developed using soft computing techniques discussed in Fister *et al.* (2014) and Azizi *et al.* (2017) can be considered in future studies on selecting suitable GMMs for probabilistic seismic hazard in Iran.

Parameter compatibility is an important challenge that one faces in evaluating several GMMs. NGA models require a few input parameters that are not available for some events in the evaluation database. We tried to handle this challenge as much as possible by estimating unknown values using references by the NGA GMMs developers (e.g., Kaklamanos *et al.*, 2011). For example, for moderate-to-large events, we used the relations suggested by Kaklamanos *et al.* (2011) to better understand the extent of the ruptured area and to convert the Joyner–Boore distance to the rupture distance. However, we were unable to calculate extended source to site distances for small-to-moderate magnitude data because we had limited information about the extent of the ruptured area for this magnitude range. In this regard, we assumed the Joyner–Boore distance to be equal to the epicentral distance ($R_{JB} = R_{epi}$) and the closest distance to the fault ruptured area as the hypocentral distance ($R_{rup} = R_{hyp}$). These assumptions are physically consistent with the small ruptured area for small-to-moderate magnitude data, and we made them in accordance with previous experiences and arguments raised by Ambraseys *et al.* (2005), Bindi *et al.* (2006), Farhadi and Mousavi (2016), and Tavakoli *et al.* (2018). Farhadi and Mousavi (2016) modeled fault rupture model parameters as random variables within the probabilistic seismic hazard assessment using the Monte Carlo simulation approach. They noticed that the influence of the uncertainties associated with the fault rupture model parameters is negligible for small-to-moderate magnitude events.

Tests of Goodness-of-fit Measures

In this study, we used three tests of goodness-of-fit measures including the LLH method of Scherbaum *et al.* (2009), EDR method proposed by Kale and Akkar (2013), and the mvLogS of Mak *et al.* (2017) to evaluate the candidate models against the evaluation database. The LLH method is the most popular data-driven evaluation method among earthquake engineers and seismologists. This method is simple and has the advantage of indicating model relative performance by a single score. In the LLH approach, a lognormal distribution with the median and standard deviation of the GMM is assumed to determine the LLH values of all observations. Then the LLH values are normalized by the total number of observations to compute the LLH score. In other words, for an individual observation x_i, we first compute LLH value assuming $g(x_i)$ as the probability density function of the candidate GMM. Then, the LLH score is the average of the LLH values computed for all observed data from the following equation, with N representing the total number of observations:

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

$$\mathrm{LLH} = -\frac{1}{N}\sum_{i=1}^{N}\log_2(g(x_i)). \quad (1)$$

The EDR method as a novel approach can also be used in selecting and ranking of GMMs for regional or site-specific probabilistic seismic hazard assessment. The EDR index is based on the Euclidean distance (DE) with some minor modification to account for the effect of the standard deviation of the candidate GMMs and can be expressed as the square root of the expression given in the following equation:

$$\mathrm{EDR}^2 = \kappa\frac{1}{N}\sum_{i=1}^{N}\mathrm{MDE}_i^2. \quad (2)$$

MDE represents the modified DE. For discrete points, MDE parameter is defined as equation (3), with D representing the differences between the logarithms of each observed data point and corresponding predictions and $\Pr(|D| < |d_j|)$ denoting the occurrence probabilities of absolute differences, d_j, within an infinitesimal bandwidth for n discrete points

$$\mathrm{MDE} = \sum_{j=1}^{n}|d_j|\Pr(|D| < |d_j|). \quad (3)$$

In addition, κ parameter is used to measure the level of bias between the observed and estimated data as

$$\kappa = \frac{\mathrm{DE}_{\mathrm{org}}}{\mathrm{DE}_{\mathrm{cor}}}. \quad (4)$$

The $\mathrm{DE}_{\mathrm{org}}$ and the $\mathrm{DE}_{\mathrm{cor}}$ are further expanded in the following equations:

$$\mathrm{DE}_{\mathrm{org}}^2 = \sum_{i=1}^{N}(q_i - p_i)^2, \quad (5)$$

$$\mathrm{DE}_{\mathrm{cor}}^2 = \sum_{i=1}^{N}(q_i - p_{c,i})^2, \quad (6)$$

in which q_i and p_i are the natural logarithms of the ith observed and predicted data, respectively. N represents the total data number in the assembled ground-motion database. The parameter $p_{c,i}$ denotes the corrected prediction of the ith data after modifying p_i using the straight line fitted on the logarithms of the estimated and observed data.

In addition to the LLH and EDR approaches, we used a method with features similar to the LLH method known as the mvLogS of Mak *et al.* (2017) to assess the applicability of several GMMs for seismic hazard studies in Iran. The mvLogS is a natural extension of the popular LLH method and has some advantages that cover the limitations of the LLH approach. The mvLogS exploits all information provided for sigma components and addresses the correlation structure of hierarchical GMMs in evaluating models' relative performances. This is an important feature due to recent advances in ground-motion modeling, which has resulted in developing models with complicated correlation structures. In addition, this approach is less sensitive to unbalanced data and is less likely to be biased toward events with larger ground-motion records. One may obtain the mvLogS from the following equation:

$$\mathrm{mvLogS} = [N\ln(2\pi) + \ln|\boldsymbol{V}| + (\boldsymbol{q}-\boldsymbol{p})'\boldsymbol{V}^{-1}(\boldsymbol{q}-\boldsymbol{p})]/2, \quad (7)$$

in which N represents the total number of observations. $\boldsymbol{p}$ and $\boldsymbol{q}$ are the vectors of logarithmic predictions and logarithms of observed ground motions, in turn. $|\boldsymbol{V}|$ and $\boldsymbol{V}^{-1}$ are the determinant and the inverse of the covariance matrix, respectively (see Mak *et al.*, 2017, and Farhadi *et al.*, 2018, for further information).

To assess score variability, we used the cluster bootstrap technique at the event level to resample 200 datasets from the evaluation database. Cluster bootstrap can be generalized into a single value referred to as the DI that shows if two models are truly different given the score variability. DI ranges from −1.0 to 1.0 and a positive value of DI indicates that the model scores better more often than another one, given the variability of the evaluation data. In this study, we used DI values to rank candidate models instead of ranking them based on the absolute value of their final score. A model having all positive DIs is the best model and more often than not scores better than the rest of the GMMs. However, the second-best model should have a single negative DI with respect to the best model. One may compute the DI from the following equation:

$$\mathrm{DI}_{ij} = \frac{1}{N_{\mathrm{bs}}}\sum_{k}^{N_{\mathrm{bs}}}\tilde{\mathbf{1}}(s_i^{(k)}, s_j^{(k)}), \tilde{\mathbf{1}}(s_i^{(k)}, s_j^{(k)}) = \begin{cases} 1 & \text{when } s_i^{(k)} < s_j^{(k)} \\ -1 & \text{when } s_i^{(k)} > s_j^{(k)} \\ 0 & \text{when } s_i^{(k)} = s_j^{(k)} \end{cases}, \quad (8)$$

in which DI_{ij} is the DI of model i with respect to model j. N_{bs} represents the number of bootstrap samples and $s_i^{(k)}$ is the score of model i for the kth bootstrap sample. A model with better performance has a smaller score. $\tilde{\mathbf{1}}(\cdot)$ is the modified indicator function.

Results

In this section, we provided the result of comparisons for four spectral periods including PGA, 0.5, 1.0, and 2.0 s. This range covers frequencies important for engineering applications. We compared models' relative performance based on the distinctness tables computed from the LLH, the EDR, and the mvLogSs of resampled datasets. We avoid presenting the distinctness tables for representative periods because

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

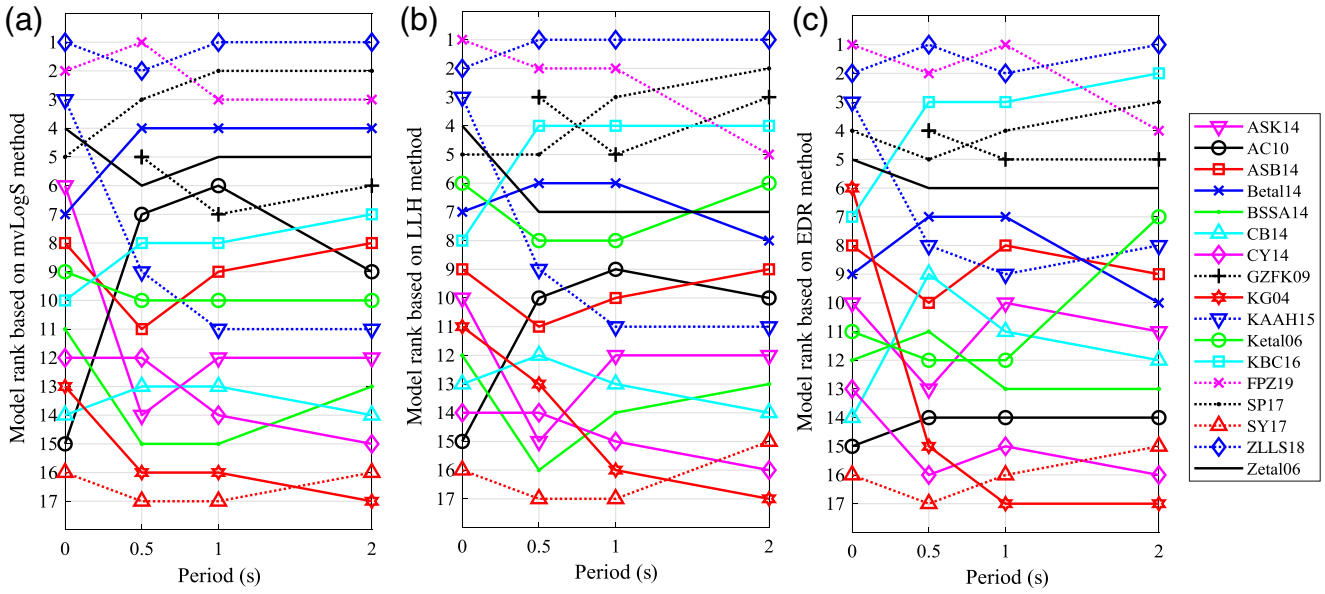

Figure 3. Model rank versus period considering the whole evaluation database and based on three different tests of goodness-of-fit measures including (a) the multivariate logarithmic score (mvLogS) of Mak *et al.* (2017), (b) the log-likelihood (LLH) method of Scherbaum *et al.* (2009), and (c) the Euclidean distance-based ranking (EDR) proposed by Kale and Akkar (2013). Models developed from considerable portion of Iranian data are plotted using dotted line style. AC10, Akkar and Çağnan (2010); ASB14, Akkar *et al.* (2014); ASK14, Abrahamson *et al.* (2014); Betal14, Bindi *et al.* (2014); BSSA14, Boore *et al.* (2014); CB14, Campbell and Bozorgnia (2014); CY14, Chiou and Youngs (2014); FPZ19, Farajpour *et al.* (2019); GZFK09, Ghasemi, Zare, Fukushima, and Koketsu (2009); KAAH15, Kale *et al.* (2015); Ketal06, Kanno *et al.* (2006); KBC16, Kotha *et al.* (2016); KG04, Kalkan and Gülkan (2004); SP17, Sedaghati and Pezeshk (2017); SY17, Shahidzadeh and Yazdani (2017); Zetal06, Zhao *et al.* (2006); ZLLS18, Zafarani *et al.* (2018). The color version of this figure is available only in the electronic edition.

models ranking according to these tables are identical to the ranking based on the scores obtained from the whole database. Ⓔ Figures S1–S4 show the distinctness tables based on the mvLogS for representative periods. Moreover, to confirm the result of the comparisons, we visually compared observed ground motions with predicted values. Zafarani and Farhadi (2017) tested several GMMs for small-to-moderate magnitude data of Iran. To eliminate the need for performing a study similar to Zafarani and Farhadi (2017), we presented the result for the whole data, small-to-moderate magnitude range, and moderate-to-large events. We evaluated models for the whole database using the three goodness-of-fit measures. However, we used the recent approach of mvLogS to compare models in various magnitude ranges. This method is sensitive to the information provided for sigma component and measures the relative performance of models considering the effects of data correlation and unbalanced data.

Models ranking based on the three goodness-of-fit measures for representative spectral periods are presented in Figure 3. In this figure, models with considerable portion of local data in their generating datasets are shown with dotted lines. The GZFK09 model is not considered in ranking GMMs for PGA because this model does not provide predictions for PGA. Based on the mvLogS method, local models outperform models from other categories for PGA. The top four models at PGA are local models of ZLLS18, FPZ19, KAAH15, and SP17. It should be noted that the KAAH15 model as the third best-fitting GMM is also valid for Turkey. The Zetal06 model with only limited data from Iran in its generating database performs better than the rest of the candidate models including the SY17 local GMM. For the short period of 0.5 s, results are relatively similar to those obtained for PGA. The first three best-fitting models are local models. Unlike PGA, the FPZ19 outperforms the ZLLS18 in the majority of resampled datasets. Compared to PGA, KAAH15 is no longer among the list of appropriate GMMs. In this period, we observed improvements in the rankings of the SP17 and Betal14 models. These models are the third and the fourth best-fitting models. The GZFK09 model is now present in the ranking list and is ranked as fifth. The Japanese model of Zetal06 is still better than a few local models and other foreign models. For the median period of 1.0 s, ZLLS18 is the best-fitting model based on the mvLogS. The SP17 model performs better than the FPZ19 model. The Betal14 model preserved its ranking at 0.5 s and is ranked as fourth for 1.0 s. Compared to 0.5 s, Zetal06 shows better performance than GZFK09 and is the fifth best-fitting model among 17 candidate GMMs. At the large period of 2.0 s, the results of ranking based on the mvLogS are akin to those obtained for the median period of 1.0 s. At 2.0 s, ZLLS18 is still the best-fitting model and outperforms the rest of the candidate models. Similar to other spectral periods, the two other local models of SP17 and FPZ19 show acceptable performances. The Betal14 and Zetal06 GMMs that represent the third and fifth groups, respectively, may be considered as appropriate in addition to the three mentioned local models.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

Figure 3 also shows the models ranking based on the LLH and the EDR methodologies over the whole frequency range. Overall, results of ranking based on the EDR and LLH methods are not significantly different from that of the mvLogS approach. The three local models including ZLLS18, SP17, and FPZ19 are among the top five models over the whole frequency range and according to all data-driven methods. In detail, the top five models are identical among all methodologies at PGA. Changes in model rank among different ranking methods are only one unit for the top five models at PGA. However, the level of difference in results between the three goodness-of-fit measures increases by increasing the spectral period. For spectral period of 0.5 s for instance, the third best-fitting model according to the mvLogS method is ranked fifth by both LLH and EDR approaches. For median-to-large periods, the KBC16 model is among top four models according to the LLH and EDR approaches, whereas this model does not show an appropriate performance according to the mvLogS. In contrast, Betal14 with the rank of four according to the mvLogS method in the same range does not perform well based on scores obtained from other ranking methods. The local model of GZFK09 has also better ranking based on the LLH and EDR approaches compared to the mvLogS method for the short-to-median periods. The LLH and EDR approaches may have favored the GZFK09 and KBC16 models because these methods are insensitive to sigma components. Figure 3 does not present scores in line with model ranking. Scores from the LLH and EDR methods are summarized in Ⓔ Tables S2 and S3, respectively. In Ⓔ Table S2, we used LLH scores to assign weights to competing models using the following equation with w representing the weight for ith GMM and K the total number of GMMs (Scherbaum *et al.*, 2009):

$$w(\mathrm{GMM}_i) = \frac{2^{-\mathrm{LLH}(\mathrm{GMM}_i)}}{\sum_{i=1}^{K} 2^{-\mathrm{LLH}(\mathrm{GMM}_i)}}. \tag{9}$$

It should be noted that scores may not replace expert opinion in evaluating GMMs or seismic hazard evaluations (Mak *et al.*, 2014). In addition, seismic hazard studies should be performed for different combinations of weights assigned to models for the branches of the logic tree to ensure that the results are not biased toward a specific model (Kale and Akkar, 2017). Furthermore, Scherbaum *et al.* (2009) proposed direct usage of obtained weights from equation (9) when an infinite amount of observed data is available. In this study, the weights computed from equation (9) are not based on infinite amount of data and we do not suggest them to be directly applied in seismic hazard studies. Therefore, the weights given in Ⓔ Table S2 still require use of experts' judgment in line with sensitivity analysis for use in seismic hazard assessment. In our opinion, these weights are useful in updating experts' prior beliefs before applying the candidate GMMs because they are based on mathematics and rely on well-established statistical tests.

Results of three goodness-of-fit measures are supplemented with residual analysis. We provided the distribution of residuals for PGA in Figures 4 and 5. We partitioned GMMs into two groups to make these figures clear. Figure 4 shows the residual distribution for models developed for Iran and Turkey, and Figure 5 does the same task for remaining candidate models. In these figures, the distribution of residuals is plotted for event terms versus magnitude and within-event residuals against distance and V_{S30}. ε_{ij} represents the within-event residuals for jth record of ith earthquake and η_i stands for event term of the ith event. According to Figure 4, there is no discernable trend in the distribution of within-event residuals versus distance and V_{S30} for all GMMs. Between-event residuals for models with superior performance at PGA show no trend, whereas models with poor performance are biased in certain magnitude ranges. SY17 as the only local model with unacceptable performance at PGA shows a pronounced negative residual trend against magnitude. This trend is less pronounced for Turkish model of KG04. The AC10 model as another Turkish model shows positive residual trend versus magnitude especially in small-to-moderate magnitude. Similarly, Figure 5 shows no trend for within-event residuals against distance and V_{S30} for remaining models selected from other groups. Based on the distribution of the between-event residuals shown in Figure 5, the NGA-West2 models with positive trend in small-to-moderate magnitude range tend to underpredict observed data, whereas they overpredict events with magnitude up to 7.0. These models show relatively no trend for magnitude 7.0 and above and seem to be reliable in this range. Models developed for Europe and Middle East show trends like those of NGA models in small-to-moderate magnitude range. These models with positive trend underpredict local data in this range. Performance of the Betal14 and ASB14 models improved for larger magnitude data. Japanese model of Zetal06 shows relatively no residual trend over the whole magnitude range, whereas Ketal06 as another Japanese model shows pronounced positive trend for small-to-moderate magnitude data and negative trend for larger events. The distribution of residuals versus magnitude, distance, and V_{S30} for other periods are shown in Ⓔ Figures S5–S10.

Figure 6 shows model ranking based on the mvLogS for two magnitude ranges and four representative periods. Following Zafarani and Farhadi (2017), we considered events with moment magnitudes up to M_w 5.0 to represent small-to-moderate magnitude data. The test data for this range contain 407 ground-motion records from 194 earthquakes with moment magnitudes ranging from 3.9 to 5 and distances below 100 km. Earthquakes in small-to-moderate magnitude range are responsible for nonnegligible contribution to the seismic hazard, specifically in short-return periods. In low-seismicity regions such as the Persian Gulf seismotectonic province, this range may contribute to the hazard in return periods up to 10,000 yr (Zafarani and Farhadi, 2017). With respect to small-to-moderate magnitude data, comparison results are not significantly different from what we obtained considering the whole data and model ranking slightly changes in some periods for a few candidate

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

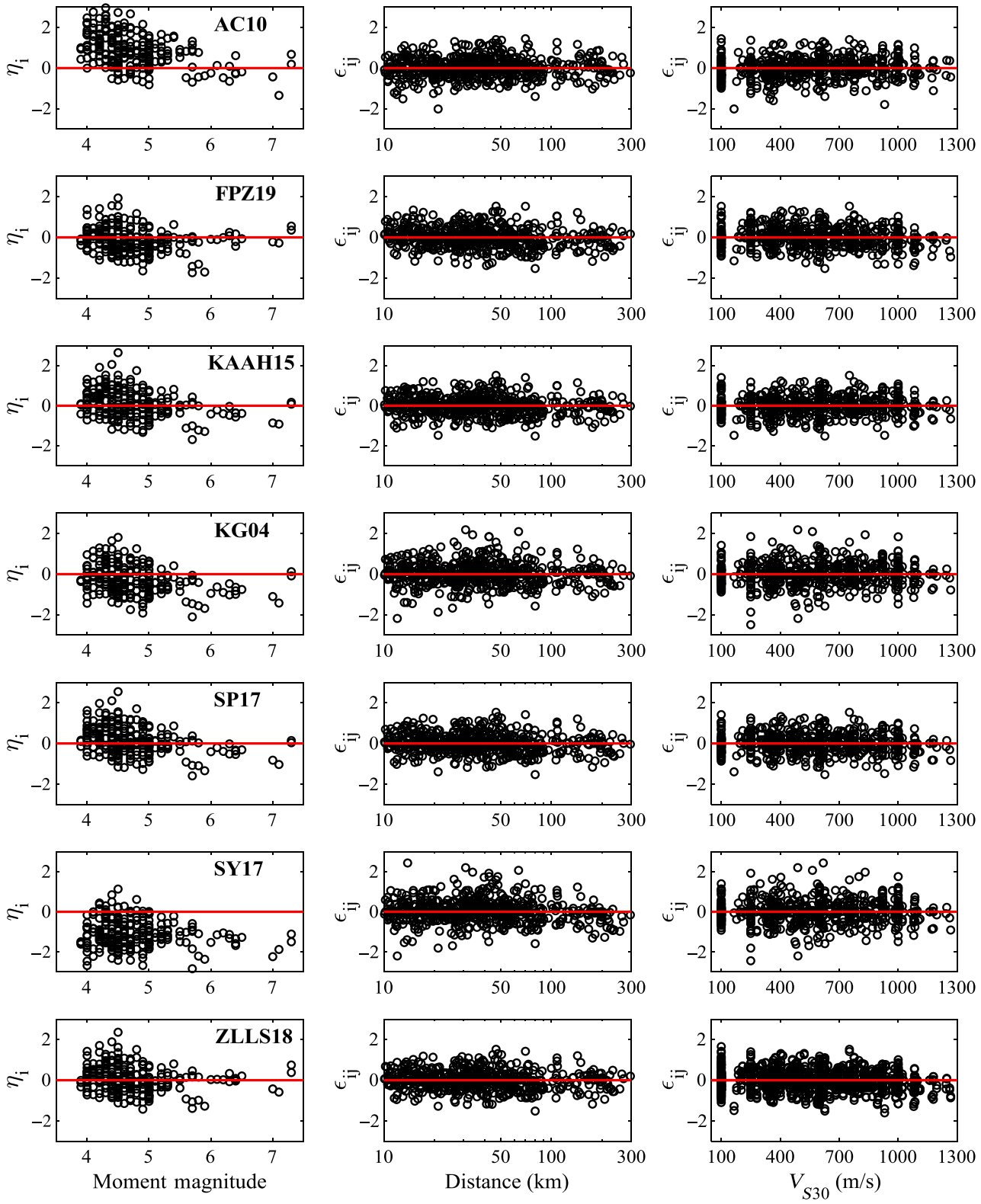

Figure 4. Distribution of residuals at PGA versus magnitude, distance, and V_{S30} for models developed for Iran and/or Turkey. The distribution of residuals against magnitude is plotted for event terms. Within-event residuals are used to plot residuals distribution versus distance and V_{S30}. The color version of this figure is available only in the electronic edition.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

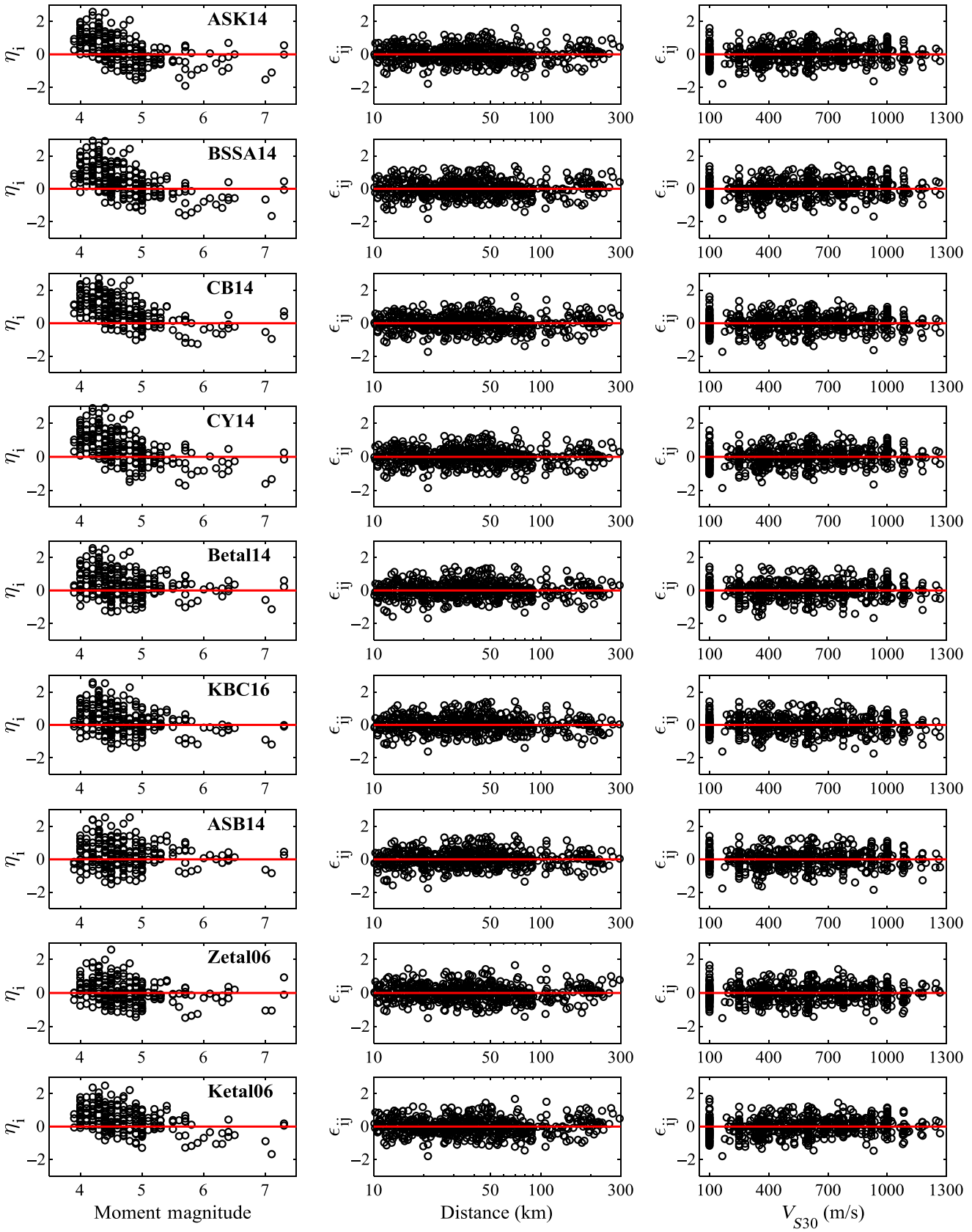

Figure 5. Distribution of residuals at PGA versus magnitude, distance, and V_{S30} for models developed for Japan, Europe and the Middle East, and the western United States. The distribution of residuals against magnitude is plotted for event terms. Within-event residuals are used to plot residuals distribution versus distance and V_{S30}. The color version of this figure is available only in the electronic edition.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

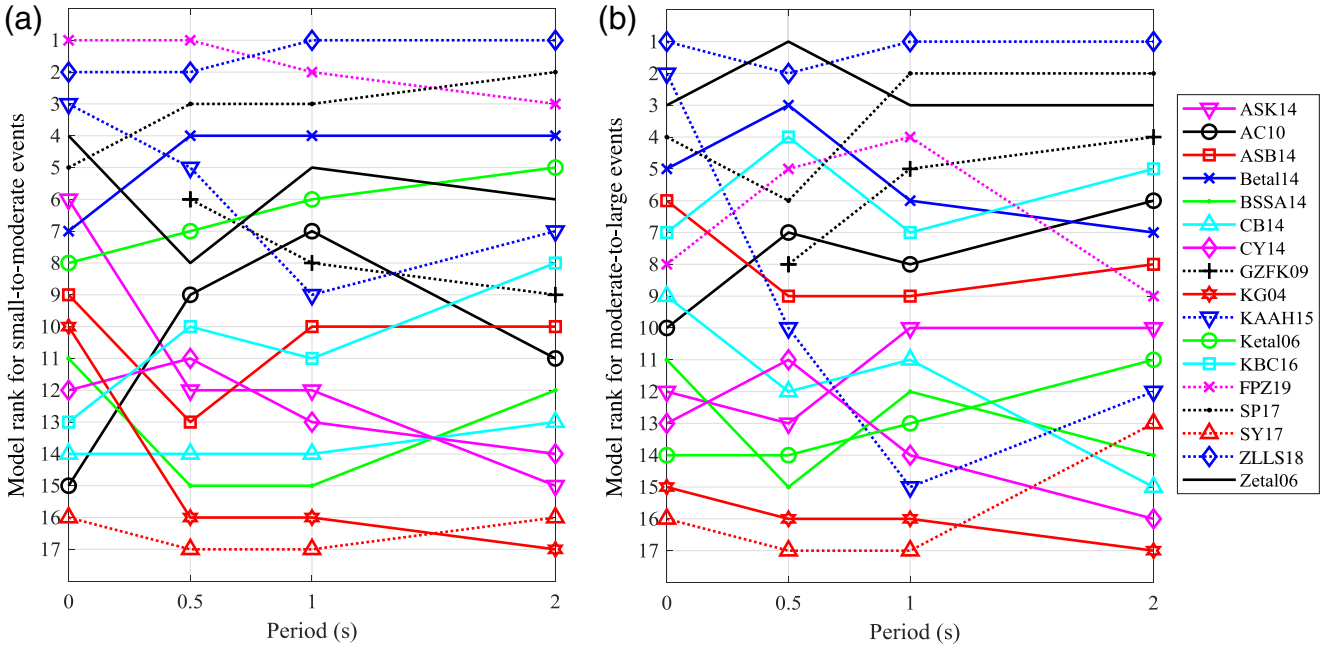

Figure 6. Model rank versus period based on the mvLogS of Mak *et al.* (2017) for (a) small-to-moderate and (b) moderate-to-large magnitude ranges. Models developed from considerable portion of Iranian data are plotted using dotted line style. The color version of this figure is available only in the electronic edition.

models. The three best-fitting models for small-to-moderate magnitude data are identical to those we found for the whole data over the entire frequency range. Overall, three local models including ZLLS18, FPZ19, and SP17 outperform other models for small-to-moderate magnitude data. FPZ19 and SP17 GMMs are valid for magnitudes down to M_w 4.5 and 4.7, respectively. Therefore, acceptable performance for FPZ19 and SP17 is an interesting result because we tested these models outside their applicability ranges. We observed a relatively poor performance for Zetal06 in the small-to-moderate magnitude range, whereas this model performed well against the whole database. This is not a surprising observation because the generating database of this model was supplemented by Iranian earthquakes in the moderate-to-large magnitude data range. Figure 6b shows the result of evaluations for representative periods considering the moderate-to-large magnitude data. The evaluation database in this range contains 236 ground-motion records from 46 earthquakes with magnitudes ranging from 5.1 to 7.3 and Joyner–Boore distances between 11 and 296 km. Model ranking based on the distinctness tables and absolute values of mvLogSs is identical for this magnitude range. However, these distinctness tables contain many elements with DI values nearly equal to zero. A DI value close to zero for model i with respect to model j occurs when model i scores better than model j in nearly half of the resampled datasets, and worse for the remaining occasions. When DI is close to zero for the two competing models, their rankings (scores) frequently change over different evaluation datasets (bootstrap samples) resampled from the original database. In such a case, results are less stable, and it is impossible to confidently determine the superior model because the two competing models are indistinguishable. Stability means relatively no fluctuation of model ranking against datasets resampled from the original database. Therefore, for moderate-to-large magnitude range results are less stable. This could be due to reduced sample size, resulting higher fluctuation in the models' relative performances among bootstrap samples and DIs closer to zero. In this range, three models including ZLLS18, Zetal06, and SP17 outperform the rest of the candidate models over the entire frequency range. The FPZ19 model seems not to perform well for PGA and the spectral period of 2.0 s. The reason for this might be the generating database (Farajpour *et al.*, 2018) of this model that includes few events from regions other than Iran. Moderate-to-large earthquakes are of great importance in high-seismicity regions. Because of this fact, we visually compared the candidate GMMs with the observed data for moderate-to-large earthquakes. To this end, we binned the data into various magnitude–V_{S30} ranges and selected two bins from moderate-to-large earthquakes with larger number of ground-motion records. We used ground-motion records with V_{S30} below 180 m/s (class D) and magnitudes between 6.0 and 6.5 as the first bin. In addition, observed data with V_{S30} between 360 and 800 m/s (class B) and magnitudes above 7.0 represent the second bin. Records of the second bin are mostly from the recent Kermanshah earthquake. GMMs are plotted for reverse faulting and the average V_{S30} and magnitude in each bin. Figure 7 shows the result of comparisons for PGA. In Figure 7, the candidate models are divided into three categories with one category including models developed using considerable portion of Iranian data. The second group includes models

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

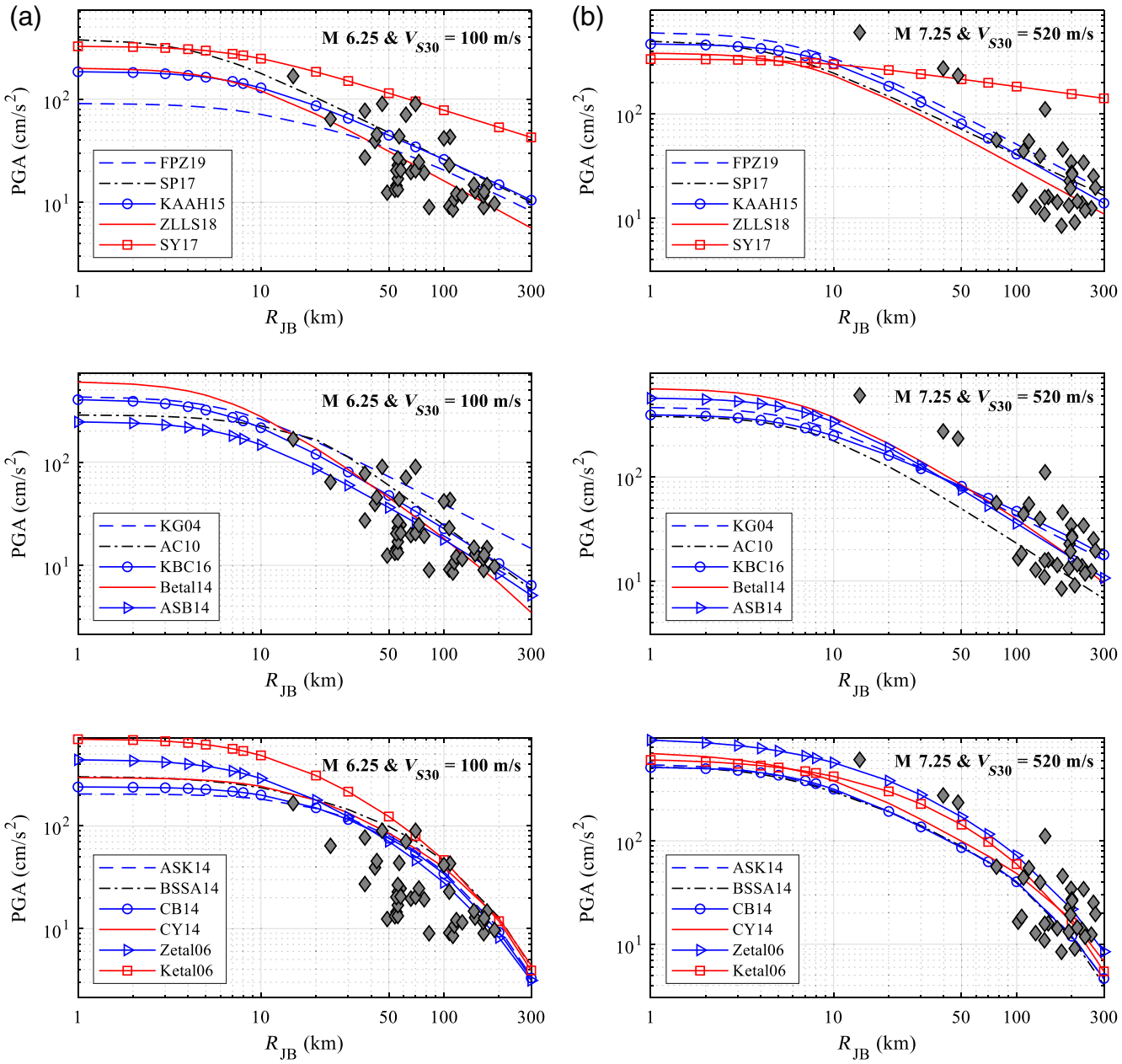

Figure 7. Visual comparison of the candidate ground-motion models (GMMs) at PGA with observed data. Filled markers show the observed data. (a) Ground-motion records (GMRs) with V_{S30} below 180 m/s and magnitudes between 6.0 and 6.5 are considered. (b) Observed data with V_{S30} between 360 and 800 m/s and magnitudes above 7.0 are considered. GMMs are plotted for reverse faulting and the average V_{S30} and magnitude in each bin. Top row includes model developed from Iranian GMRs; middle row illustrates models developed for Turkey and Europe and Middle East; and bottom row provides the visual comparisons for the Next Generation Attenuation-West2 Project (NGA-West2) models as well as the two Japanese models. The color version of this figure is available only in the electronic edition.

developed for use in Turkey and Europe and Middle East. The third group contains NGA-West2 models in addition to two Japanese models. In preparing this figure, we used the study of Beyer and Bommer (2006) to convert predictions of the KG04 and SY17 models to geometric mean. According to Figure 7, some of the local models are compatible with the observed data for both bins. Among these models, superior performance of ZLLS18 over other GMMs is obvious. Models from other categories overpredicted the observed data from ground-motion records in first bin. Foreign models' performance has improved for the second bin with larger magnitude data recorded on stiffer soil site condition. To reduce the length of the article, we did not provide the visual comparisons between observed and predicted ground-motion intensities for other periods and moved related figures to the supplemental content.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

Ⓔ Figures S11–S13 illustrate the visual comparisons for 0.5, 1.0, and 2.0 s, respectively.

Summary and Conclusion

We used a database of 643 ground-motion records from 240 Iranian earthquakes with magnitudes ranging from 3.9 to 7.3 and distances up to 300 km to evaluate the predictive capability of several GMMs for the probabilistic seismic hazard in Iran. To compare the models' predictive capability, we made the evaluation database prospective to all candidate models by excluding ground-motion records used to develop these models. We selected the candidate GMMs from five groups including local models of Iran, models from Turkey, two Japanese models, NGA-West2 global models, and models established for Europe and the Middle East. We evaluated the models' relative performance using three tests of goodness-of-fit measures including the LLH method of Scherbaum *et al.* (2009), the EDR proposed by Kale and Akkar (2013), and a new method of ranking referred to as the mvLogS of Mak *et al.* (2017). The mvLogS method exploits all information provided for sigma components and addresses the correlation structure of hierarchical GMMs in evaluating models' relative performances. In addition, this approach is less sensitive to unbalanced data and is less likely to be biased toward events with larger ground-motion records. We ranked the candidate models by paying attention to the issue of score variability. To take score variability into account, we used the cluster bootstrap technique to generate resampled datasets from the whole data. Then, we computed the DI for all model pairs from resampled datasets and ranked the candidate models.

We performed the evaluation for various magnitude ranges and different frequencies. We presented the result for PGA and three spectral periods including 0.5, 1.0, and 2.0 s. Considering the whole dataset, three local models including ZLLS18, FPZ19, and SP17 outperformed the rest of the candidate models. Shahidzade and Yazdani (2017), a local model developed based on the Bayesian updating, appears not to perform better than models established based on the conventional empirical approaches. None of the NGA models seems to be sufficient for the Iranian plateau. Betal14 developed for Europe and the Middle East and Japanese model of Zetal06 are models that we recommend above other foreign models. For the small-to-moderate magnitude events, the test data contain 407 ground-motion records from 194 earthquakes with moment magnitudes ranging from 3.9 to 5 and distances up to 100 km. Evaluation results do not change significantly for this range, and we still propose the ZLLS18, FPZ19, and SP17 local models as the most appropriate GMMs for seismic hazard studies in Iran. FPZ18 and SP17 GMMs are valid for magnitudes down to M_w 4.5 and 4.7, respectively. We tested these two models outside their applicability ranges and their acceptable performance is an interesting result. With respect to the moderate-to-large magnitude range, the results are less stable due to reduced sample size. For this range, however, three models including ZLLS18, Zetal06, and SP17 outperform the rest of the candidate models over the entire frequency range.

The result of the present study can be useful in selecting a set of appropriate GMMs for future seismic hazard and risk assessments and updating previous studies such as Ghodrati Amiri *et al.* (2013), Khodaverdian *et al.* (2016), Mousavi *et al.* (2016), and Khoshnevis *et al.* (2017), which used some GMMs not included in models proposed in this study as most suitable GMMs. Our study shows the extent to which complex foreign models such as the NGA-West2 models were successful in predicting the local observed data. Moreover, it provides insights on the ways that local models could be possibly improved by refining their models or updating their generating datasets.

Data and Resources

We used the data provided by Zafarani and Soghrat (2017) and Babaie Mahani and Kazemian (2018) as our evaluation database. These studies and the datasets they provide are available online. The MATLAB codes for the Next Generation Attenuation (NGA) models were downloaded from the Baker Research Group (https://web.stanford.edu/~bakerjw/GMPEs.html, last accessed March 2019).

Acknowledgments

This study benefited from the constructive and insightful comments of two anonymous reviewers as well as Associate Editor John Douglas.

References

Abrahamson, N. A., and W. J. Silva (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, *Earthq. Spectra* **24,** no. 1, 67–97.

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra* **30,** no. 3, 1025–1055.

Akhani, M., A. R. Kashani, M. Mousavi, and A. H. Gandomi (2019). A hybrid computational intelligence approach to predict spectral acceleration, *Measurement* **138,** 578–589.

Akkar, S., and J. J. Bommer (2010). Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean Region and the Middle East, *Seismol. Res. Lett.* **81,** 195–206.

Akkar, S., and Z. Çağnan (2010). A local ground-motion predictive model for Turkey and its comparison with other regional and global ground motion models, *Bull. Seismol. Soc. Am.* **100,** no. 6, 2978–2995.

Akkar, S., M. A. Sandıkkaya, and J. J. Bommer (2014). Empirical ground motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East, *Bull. Earthq. Eng.* **12,** no. 1, 359–387.

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit (2005). Equations for estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration, *Bull. Earthq. Eng.* **3,** 1–53.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, *et al.* (2014). PEER NGA-West2 Database, *PEER Report No. 2013/03*, University of California, Berkeley, California.

Atkinson, G. M., and K. Assatourians (2017). Are ground-motion models derived from natural events applicable to the estimation of expected

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

motions for induced earthquakes? *Seismol. Res. Lett.* **88,** no. 2A, 430–441.

Azizi, K., J. Attari, and A. Moridi (2017). Estimation of discharge coefficient and optimization of Piano Key Weirs, *Labyrinth and Piano Key Weirs III: Proc. of the 3rd International Workshop on Labyrinth and Piano Key Weirs (PKW 2017)*, Qui Nhon, Vietnam, 22–24 February 2017, p. 213.

Babaie Mahani, A., and J. Kazemian (2018). Strong ground motion from the November 12, 2017, M 7.3 Kermanshah earthquake in western Iran, *J. Seismol.* 1–20, doi: 10.1007/s10950-018-9761-x.

Beyer, K., and J. J. Bommer (2006). Relationships between median values and between aleatory variabilities for different definitions of the horizontal component of motion, *Bull. Seismol. Soc. Am.* **96,** 1512–1522.

Bindi, D. (2017). The predictive power of ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **107,** no. 2, 1005–1011, doi: 10.1785/0120160224.

Bindi, D., L. Luzi, F. Pacor, G. Franceschina, and R. R. Castro (2006). Ground-motion predictions from empirical attenuation relationships versus recorded data: The case of the 1997-1998 Umbria-Marches, Central Italy, strong-motion data set, *Bull. Seismol. Soc. Am.* **96,** no. 3, 984–1002.

Bindi, D., M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, and P. Augliera (2014). Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset, *Bull. Earthq. Eng.* **12,** no. 1, 391–430.

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fäh (2010). On the selection of ground-motion prediction equations for seismic hazard analysis, *Seismol. Res. Lett.* **81,** no. 5, 783–793, doi: 10.1785/gssrl.81.5.783.

Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion, *Bull. Seismol. Soc. Am.* **100,** no. 4, 1830–1835, doi: 10.1785/0120090400.

Boore, D. M., and G. M. Atkinson (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s, *Earthq. Spectra* **24,** 99–138.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes, *Earthq. Spectra* **30,** no. 3, 1057–1085, doi: 10.1193/070113EQS184M.

Boore, D. M., J. Watson-Lamprey, and N. A. Abrahamson (2006). Orientation-independent measures of ground motion, *Bull. Seismol. Soc. Am.* **96,** no. 4A, 1502–1511, doi: 10.1785/0120050209.

Campbell, K. W., and Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5%-damped linear elastic response spectra for periods ranging from 0.01 to 10 s, *Earthq. Spectra* **24,** no. 1, 139–171.

Campbell, K. W., and Y. Bozorgnia (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra, *Earthq. Spectra* **30,** no. 3, 1087–1115, doi: 10.1193/062913EQS175M.

Chiou, B. S.-J., and R. R. Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **24,** 173–215.

Chiou, B. S.-J., and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **30,** no. 3, 1117–1153.

Cotton, F., F. Scherbaum, J. J. Bommer, and H. Bungum (2006). Criteria for selecting and adjusting ground-motion models for specific target regions: Application to Central Europe and rock sites, *J. Seismol.* **10,** no. 2, 137–156, doi: 10.1007/s10950-005-9006-7.

Eurocode 8 (2004). *Design of Structures for Earthquake Resistance. Part 1: General Rules, Seismic Action and Rules for Buildings, EN 1998-1*, European Committee for Standardization (CEN), Brussels, Belgium, available at http://www.cen.eu/Pages/default.aspx (last accessed August 2015).

Farajpour, Z., S. Pezeshk, and M. Zare (2019). A new empirical ground-motion model for Iran, *Bull. Seismol. Soc. Am.* **109,** no. 2, 732–744, doi: 10.1785/0120180139.

Farajpour, Z., M. Zare, S. Pezeshk, A. Ansari, and E. Farzanegan (2018). Near-source strong motion database catalog for Iran, *Arab. J. Geosc.* **11,** 80.

Farhadi, A., and M. Mousavi (2016). Consideration of the rupture model uncertainties in the probabilistic seismic hazard analysis, *Soil Dynam. Earthq. Eng.* **83,** 191–204.

Farhadi, A., S. Pezeshk, and N. Khoshnevis (2018). Assessing the applicability of ground-motion models for induced seismicity application in Central and Eastern North America, *Bull. Seismol. Soc. Am.* **108,** no. 4, 2265–2277.

Fister, I., A. H. Gandomi, I. J. Fister, M. Mousavi, and A. Farhadi (2014). Soft Computing in Earthquake engineering: A short overview, *Int. J. Earthq. Eng. Hazard Mitig.* **2,** no. 2, 42–48.

Ghasemi, H., M. Zare, Y. Fukushima, and K. Koketsu (2009). An empirical spectral ground-motion model for Iran, *J. Seismol.* **13,** 499–515.

Ghasemi, H., M. Zare, Y. Fukushima, and F. Sinaeian (2009). Applying empirical methods in site classification, using response spectral ratio (H/V): A case study on Iranian strong motion network (ISMN), *Soil Dynam. Earthq. Eng.* **29,** 111–132.

Ghodrati Amiri, G., N. Khoshnevis, and S. Razavian Amrei (2013). Probabilistic assessment of earthquake damage and loss for the city of Tehran, Iran, *J. Rehabil. Civil Eng.* **1,** no. 2, 10–23.

Kaklamanos, J., L. G. Baise, and D. M. Boore (2011). Estimating unknown input parameters when implementing the NGA ground motion prediction equations in engineering practice, *Earthq. Spectra* **27,** 1219–1235.

Kale, Ö., and S. Akkar (2013). A new procedure for selecting and ranking ground-motion prediction equations: The Euclidean distance-based ranking (EDR) method, *Bull. Seismol. Soc. Am.* **103,** 1069–1084.

Kale, Ö., and S. Akkar (2017). A ground motion logic tree scheme for regional seismic hazard studies, *Earthq. Spectra* **33,** no. 3, 837–856.

Kale, O., S. Akkar, A. Ansari, and H. Hamzehloo (2015). A ground-motion predictive model for Iran and Turkey for horizontal PGA, PGV, and 5% damped response spectrum: Investigation of possible regional effects, *Bull. Seismol. Soc. Am.* **105,** 963–980.

Kalkan, E., and P. Gülkan (2004). Site-dependent spectra derived from ground motion records in Turkey, *Earthq. Spectra* **20,** 1111–1138.

Kanno, T., A. Narita, N. Morikawa, H. Fujiwara, and Y. Fukushima (2006). A new attenuation relation for strong ground motion in Japan based on recorded data, *Bull. Seismol. Soc. Am.* **96,** no. 3, 879–897.

Khodaverdian, A., H. Zafarani, M. Rahimian, and V. Dehnamaki (2016). Seismicity parameters and spatially smoothed seismicity model for Iran, *Bull. Seismol. Soc. Am.* **106,** no. 3, 1133–1150.

Khoshnevis, N., R. Taborda, S. Azizzadeh-Roodpish, and C. H. Cramer (2017). Seismic hazard estimation of northern Iran using smoothed seismicity, *J. Seismol.* **21,** no. 4, 941–964, doi: 10.1007/s10950-017-9645-5.

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

Kulkarni, R. B., R. R. Youngs, and K. J. Coppersmith (1984). Assessment of confidence intervals for results of seismic hazard analysis, *Proc. of the Eighth World Conf. on Earthquake Engineering*, San Francisco, California, Vol. 1, 263–270.

Mak, S. (2017). Measuring the performance of ground-motion models: The importance of being independent, *Seismol. Res. Lett.* **85,** no. 5, 1212–1217.

Mak, S., R. A. Clements, and D. Schorlemmer (2014). Comment on "A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method" by Özkan Kale and Sinan Akkar, *Bull. Seismol. Soc. Am.* **104,** no. 6, doi: 10.1785/0120140106.

Mak, S., R. A. Clements, and D. Schorlemmer (2017). Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting, *Bull. Seismol. Soc. Am.* **107,** no. 2, doi: 10.1785/0120160232.

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user

Mak, S., F. Cotton, M. Gerstenberger, and D. Schorlemmer (2018). An evaluation of the applicability of NGA-West2 ground-motion models for Japan and New Zealand, *Bull. Seismol. Soc. Am.* **108,** no. 2, 836–856, doi: 10.1785/0120170146.

McGuire, R. K. (2004). *Seismic Hazard and Risk Analysis*, EERI Monograph MNO-10, Earthquake Engineering Research Institute, Oakland, California, 187 pp.

Mousavi, M., A. Ansari, H. Zafarani, and A. Azarbakht (2012). Selection of ground motion prediction models for seismic hazard analysis in the Zagros region, Iran, *J. Earthq. Eng.* **16,** 1184–1207.

Mousavi, M., A. Azarbakht, S. Rahpeyma, and A. Farhadi (2015). On the application of genetic programming for new generation of ground motion prediction equations, in *Handbook of Genetic Programming Applications*, Springer, Cham, Switzerland, 289–307.

Mousavi, M., A. Farhadi, A. Azarbakht, and H. Zafarani (2016). Influence of depth uncertainty treatment on result of probabilistic seismic hazard analysis, *Sharif: Civil Eng.* **32-2,** no. 2.1, 111–118.

Mousavi, M., H. Zafarani, S. Rahpeyma, and A. Azarbakht (2014). Test of goodness of the NGA ground-motion equations to predict the strong motions of the 2012 Ahar-Varzaghan dual earthquakes in northwestern Iran, *Bull. Seismol. Soc. Am.* **104,** no. 5, 2512–2528.

Scherbaum, F., F. Cotton, and P. Smit (2004). On the use of response spectral-reference data for the selection and ranking of ground motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion, *Bull. Seismol. Soc. Am.* **94,** no. 6, 2164–2185.

Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, *Bull. Seismol. Soc. Am.* **99,** no. 6, 3234–3247.

Sedaghati, F., and S. Pezeshk (2017). Partially nonergodic empirical ground motion models for predicting horizontal and vertical PGV, PGA, and 5% damped linear acceleration response spectra using data from Iranian plateau, *Bull. Seismol. Soc. Am.* **107,** no. 4, 934–948.

Shahidzadeh, M. S., and A. Yazdani (2017). A Bayesian updating applied to earthquake ground-motion prediction equations for Iran, *J. Earthq. Eng.* **21,** no. 2, 290–324, doi: 10.1080/13632469.2016.1158754.

Shoja-Taheri, J., S. Naserieh, and H. Ghofrani (2010). A test of the applicability of NGA models to the strong ground motion data in the Iranian Plateau, *J. Earthq. Eng.* **14,** 278–292.

Soghrat, M. R., N. Khaji, and H. Zafarani (2012). Simulation of strong ground motion in northern Iran using the specific barrier model, *Geophys. J. Int.* **188,** 645–679.

Tavakoli, B., F. Sedaghati, and S. Pezeshk (2018). An analytical effective point-source-based distance-conversion approach to mimic the effects of extended faults on seismic hazard assessment, *Bull. Seismol. Soc. Am.* **108,** no. 2, 742–760.

Van Houtte, C., S. Bannister, C. Holden, S. Bourguignon, and G. McVerry (2017). The New Zealand strong motion database, *Bull. New Zeal. Soc. Earthq. Eng.* **50,** no. 1, 1–20.

Zafarani, H., and A. Farhadi (2017). Testing ground-motion prediction equations against small-to-moderate magnitude data in Iran, *Bull. Seismol. Soc. Am.* **107,** 912–933, doi: 10.1785/0120160046.

Zafarani, H., and M. Mousavi (2014). Applicability of different ground-motion prediction models for northern Iran, *Nat. Hazards* **73,** 1199–1228.

Zafarani, H., and M. Soghrat (2012). Simulation of ground motion in the Zagros region, Iran using the specific barrier model and stochastic method, *Bull. Seismol. Soc. Am.* **102,** 2031–2045.

Zafarani, H., and M. Soghrat (2017). A selected dataset of the Iranian strong motion records, *Nat. Hazards* **86,** 1307–1332.

Zafarani, H., L. Luzi, G. Lanzano, and M. Soghrat (2018). Empirical equations for the prediction of PGA and pseudo spectral accelerations using Iranian strong-motion data, *J. Seismol.* **22,** no. 1, 263–285.

Zare, M., P. Y. Bard, and M. Ghafory-Ashtiany (1999). Site characterizations for the Iranian strong motion network, *Soil Dynam. Earthq. Eng.* **18,** 101–123.

Zhao, J. X., J. Zhang, A. Asano, Y. Ohno, T. Oouchi, T. Takahashi, H. Ogawa, K. Irikura, H. K. Thio, P. G. Somerville, *et al.* (2006). Attenuation relations of strong ground motion in Japan using site classifications based on predominant period, *Bull. Seismol. Soc. Am.* **96,** no. 3, 898–913.

Department of Civil Engineering
The University of Memphis
3815 Central Avenue
Memphis, Tennessee 38152 U.S.A.
afarhadi@memphis.edu
zfrjpour@memphis.edu
spezeshk@memphis.edu

Manuscript received 6 November 2018;
Published Online 16 July 2019

Downloaded from https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/4833993/bssa-2018307.1.pdf
by University of Memphis user