CrossMark

ORIGINAL ARTICLE

Determination of GMPE functional form for an active region with limited strong motion data: application to the Himalayan region

Ketan Bajaj · P. Anbazhagan

Abstract Advancement in the seismic networks results in formulation of different functional forms for developing any new ground motion prediction equation (GMPE) for a region. Till date, various guidelines and tools are available for selecting a suitable GMPE for any seismic study area. However, these methods are efficient in quantifying the GMPE but not for determining a proper functional form and capturing the epistemic uncertainty associated with selection of GMPE. In this study, the compatibility of the recent available functional forms for the active region is tested for distance and magnitude scaling. Analysis is carried out by determining the residuals using the recorded and the predicted spectral acceleration values at different periods. Mixed effect regressions are performed on the calculated residuals for determining the intra- and interevent residuals. Additionally, spatial correlation is used in mixed effect regression by changing its likelihood function. Distance scaling and magnitude scaling are respectively examined by studying the trends of intraevent residuals with distance and the trend of the event term with magnitude. Further, these trends are statistically studied for a respective functional form of a ground motion. Additionally, genetic algorithm and Monte Carlo method are used respectively for calculating the hinge point and standard error for magnitude and distance scaling for a newly determined functional form. The whole

procedure is applied and tested for the available strong motion data for the Himalayan region. The functional form used for testing are five Himalayan GMPEs, five GMPEs developed under NGA-West 2 project, two from Pan-European, and one from Japan region. It is observed that bilinear functional form with magnitude and distance hinged at 6.5 M_w and 300 km respectively is suitable for the Himalayan region. Finally, a new regression coefficient for peak ground acceleration for a suitable functional form that governs the attenuation characteristic of the Himalayan region is derived.

1 Introduction

Ground motion prediction equations take a practical bend towards characterizing the present state of the practice of seismic hazard analysis for earthquake-prone regions. Advancement in the seismic network and geophysics results in incorporating various new parameters for deriving any ground motion prediction equations (GMPEs). The general procedure used in developing any GMPE is the regression analysis of the ground motion recordings either from past events or from stochastic simulation. GMPE models describe the distribution of ground motion in terms of median and logarithmic standard deviation (Strasser et al. 2009). However, the limitation of such approach is the

K. Bajaj · P. Anbazhagan (✉)
Department of Civil Engineering, Indian Institute of Science, Bangalore, India
e-mail: anbazhagan@civil.iisc.ernet.in
e-mail: anbazhagan2005@gamil.com

appropriate representation of ground motions used to develop the empirical model. Additionally, it lacks in relation to the particular seismic scenarios for which these GMPEs will be applied (Bradley 2013). Recently, a summary of the worldwide, developed GMPEs from 1964 to mid-2015 (inclusive) for estimating peak ground acceleration (PGA) and spectral acceleration (SA) was published by Douglas (2015). Despite the availability of various methods and criteria for selecting an appropriate GMPE (Cotton et al. 2006; Bommer et al. 2010; Delavaud et al. 2009) for many practical applications, there exists an important issue regarding the applicability of a GMPE developed for one region to another region.

NGA-West 2 project has developed a series of ground motions for the tectonically active region of the shallow crustal earthquakes. These GMPEs are presented by Abrahamson et al. (2014), Boore et al. (2014), Campbell and Bozorgnia (2014), Chiou and Youngs (2014), and Idriss (2014), which are recently updated under the NGA-2 project. Similarly, for the Himalayan region, which is considered as one of the most active seismic region in the world, various authors (Sharma and Bungum 2006; Das et al. 2006; Baruah et al. 2009; Nath et al. 2009; Sharma et al. 2009; Gupta 2010; Anbazhagan et al. 2013) have published different GMPEs for determining the PGA at rock level. Recently, Akkar et al. (2014) and Bindi et al. (2014) have developed new GMPEs for Pan-European region. Moreover, Zhao et al. (2016a, b, c) have developed three new GMPEs for Japan by differentiating the subduction interface earthquakes, subduction slab earthquakes, and shallow crustal and upper mantle earthquakes. Even though, tectonically, these regions are active, the attenuation characteristics of seismic waves for these regions are different. Cotton et al. (2006) describes how the source characteristic and path effect related to geometric and anelastic attenuation and site effect varies from region to region. This can be one of the reasons for many of these GMPEs to have different hinge points for magnitude and distance scaling. As the Himalayan seismotectonic is of subduction in nature, the extent of applicability of GMPEs developed for the shallow crustal region is a matter of concern. Hence, an important issue for developing a functional form that encounters the attenuation characteristic of the region is studied in this paper, using different statistical tools.

Recently developed GMPEs are based on a large number of recorded data; hence, various effects (like faulting type, site response, basin and hypocentral depth, and depth of the top of rupture, directivity, hanging-wall site locations) have been well resolved, for example, NGA-West 2 GMPEs. However, many of the seismically active regions like the Himalayan region lack in recorded strong motion data of great and large earthquakes, despite the availability of the dense seismic network. The objective of the present study is to use the statistical techniques to examine the applicability of available functional forms for the region having less recorded strong motion database. The study has been carried out by determining the residuals using the recorded and the predicted spectral acceleration at different periods. The regression on the residuals has been performed using a mixed effect regression technique to determine the applicability of different functional forms corresponding to magnitude and distance scaling. The mixed effect regression is used to determine the magnitude and distance scaling based on the residuals. For determining the intra- and interevent residuals, spatial correlation has been used in mixed effect regression by changing its likelihood function. Further, inter- and intraevents residuals have been statistically studied and used for evaluating the functional form corresponding to the distance and magnitude scaling for the limited recorded data. Distance scaling is investigated by examining the trends of intraevent residuals with distance, whereas magnitude scaling was examined by studying the trend of the interevent term with magnitude. Further, genetic algorithm and Monte Carlo method are used for determining the hinge points and to calculate the standard error in each term that governs the magnitude and distance scaling for a particular functional form. The whole procedure is applied to the active Himalayan region with a limited number of strong motion records. The functional form used for testing are five Himalayan GMPEs, five GMPEs developed under NGA-West 2 project, and three GMPEs from other active regions. Based on the analysis, a new functional form is arrived for the Himalayan region. Additionally, using the mixed effect regression along with the Monte Carlo approach, a GMPE for PGA for the Himalayan region has been developed.

2 Functional forms used in various GMPEs

With the enhancement in the database size and advancement in the simulation algorithms, new GMPEs are

being formulated with the varying degree of complexity in its functional form. Improvement in the functional form of GMPE and addition of different effects or parameters (directivity, fault type, etc.) benefited various engineering purposes, mostly for rarely observed conditions such as large magnitude and short distance events. However, a basic form of a particular GMPE must account for magnitude dependence and saturation, as well as attenuation of stress waves with distance due to spreading and material damping. In addition to this, many of the GMPEs developed worldwide include the effect of faulting type, basin and hypocentral depths and depth of the top of rupture, site response, directivity, hanging-wall site locations, and whether the event was a main shock or aftershock. Most of these parameters are only applicable for the seismic study area with a well-defined source geometry and orientation. As per Baltay and Hanks (2014), for describing these effects, a large number of explanatory parameters are necessary, which make the NGA-West 2 GMPEs more complex to understand. However, most of these parameters (hanging-wall and non-linear site amplification) can only be employed for detailed and well-studied sites. Additionally, the models for the hanging-wall, directivity, and the non-linear site response are constrained by extra data and research (Donahue and Abrahamson 2014; Kamai et al. 2014; Spudich et al. 2014). Hence, attributes of the Himalayan, NGA-West 2, Pan-European, and Japan GMPE models for active region have been discussed in this section. For hazard analysis, the commonly used Himalayan GMPE models are those of Nath et al. (2009), Sharma et al. (2009), Gupta (2010), NDMA (2010), and Anbazhagan et al. (2013). These equations will be referred with the following names here onwards; NA09, SH09, GU10, NDMA10, and AN13, respectively. Similarly NGA-West 2 models namely Abrahansom et al. (2014), Boore et al. (2014), Campbell and Bozrgnia (2014), Chiou and Young (2014), and Idriss (2014) are respectively abbreviated as AB14, BA14, CB14, CY14, and ID14 and Akkar et al. (2014), Bindi et al. (2014), and Zhao et al. (2016a, b, c) are referred to AK14, BD14, and ZH16. The above used GMPEs are developed recently with a wide range of recorded database and widely used for determination of seismic hazard for an active region.

Each of the NGA-West 2 GMPEs incorporates saturation as a function of magnitude at short distances and short periods. These five models contain the style of faulting and its dependency decreases with magnitude.

However, BA14 and ID14 have a style of faulting, which is magnitude independent. AB14, CB14, and CY14 explicitly considered hanging wall features and rupture depth term. However, BA14 accounts the hanging wall effect using Joyner–Boore distance (R_{JB}). All the five GMPEs are defined as a function of magnitude as moment magnitude or M_w. The distance metric used in AB14, CB14, CY14, and ID14 is the closest distance to the rupture plane, R_{RUP}. Whereas, BA14 used the closest distance of the horizontal projection of the rupture plane, R_{JB}. AB14, CB14, and CY14 used depth to the top of the rupture (Z_{TOR}); additionally, depth of the hypocenter, Z_{HYP} was included in CB14. BA14 included neither of the depth parameter in the GMPE functional form. AK14 and BD14 used RESORCE database for deriving the GMPE for Pan European region. For both the GMPEs, magnitude size is defined using moment magnitude or M_w and considered R_{JB}, for defining the finite fault distance metric(s). ZH16 has derived the three GMPEs using the Japan database by differentiating the subduction interface earthquakes, subduction slab earthquakes, and shallow crustal and upper-mantle earthquakes. ZH16 used moment magnitude or M_w for earthquake size and used the shortest distance from a recording station to the fault plane if a fault model is available, otherwise the hypocentral distance for distance metric(s). Additionally, ZH16 has used the geometric attenuation rate term in the GMPE functional form. In case of subduction interface earthquakes, ZH16 defined different coefficients for the magnitude scaling by differentiating deep and shallow earthquakes.

Magnitude scaling is accounted in GMPEs as both linear (CB14) and non-linear or bilinear (AB14, BA14, CY14, ID14, AK14, and BD14) functional form. BA14 used a linear functional for magnitude scaling for moment magnitude of less than 5.5 and non-linear or quadratic for magnitude more than 5.5 (check both are more than). Whereas, CB14 used piecewise linear form and CY14 used a bilinear relation with smooth transitions for different magnitude bins (see Table 1). The Himalayan GMPE models, SH09, GU10, and AN13, used a linear form, whereas NDMA10 used a non-linear (quadratic) form for magnitude scaling and NA09 used a cubical form for magnitude scaling. AK14 used the quadratic functional form for magnitude scaling with hinge at 6.75. BD14 has used a linear functional form for magnitude scaling for $M_w > 6.75$ and the quadratic for events with $M_w < 6.75$. ZH16 used linear functional form in case of subduction interface earthquakes and

shallow crustal and upper-mantle earthquakes, however non-linear or bilinear in case of subduction slab earthquakes. ZH16 used a bilinear functional form for magnitude scaling with hinge at 7.1 M_w in all the three cases. Table 1 shows the different functional forms for magnitude scaling.

Many of the GMPE models use simple, functional form as far as distance scaling is concerned. Table 2 shows the different functional form for distance adopted by various authors in NGA-West 2 and Himalayan models. The GMPE models CB14 AB14 and AK14 used function form consisting of product of a linear function of magnitude and the logarithm of the square root of the sum of squares of distance and fictitious source depth. The linear term accounts for the decrease in attenuation with respect to increase in magnitude. However, the overall intercept is negative whereas coefficient corresponding to magnitude term is positive. CY14 model accounted the variation of distance attenuation by capturing the dominant effect of body waves up to the distance of 40–70 km and Lg waves at larger distances. BA14 and BD14 accounted the distance attenuation with the apparent anelastic term along with linear function of magnitude and the logarithm of distance. Both the GMPEs used additional magnitude and distance constants which were represented as M_{ref} and R_{ref}. BA14 and BD14 respectively fixed M_{ref} to 4.5 and 5.5, and both the GMPEs used R_{ref} as 1 km. CY14 and ID14 also included the anelastic attenuation term. As far as the Himalayan GMPEs are concerned, the distance scaling is generally a logarithm of a linear combination of distances and exponential of magnitudes (NA09, NDMA10 and AN13). NDMA10 and GU10 use an anelastic attenuation coefficient for attenuation with distance to account for material damping. Most of the Himalayan GMPEs only include only the attenuation of seismic waves due to spreading, which is governed by logarithmic of distance term (see Table 2). ZH16 include both anelastic and geometric attenuation variabilities for distance scaling. ZH16 further used two coefficients for geometric attenuation rate by differentiating it from large-distance geometric attenuation rate. From Tables 1 and 2, it can be seen that different researchers have used a different functional form for magnitude and distance scaling. This variability in functional form can be accounted more clearly, if a large number of recorded strong motion data are available. For some of the regions like the Himalaya, where seismicity is distributed non-uniformly and large number of strong motion data

is not available, mixed effect regression on the residual, by including spatial variability, is one of the ways in which the best suitable functional form can be derived.

3 Different methods available for comparison of GMPEs

There are various approaches available in the literature that has been used for comparing of the GMPE. These approaches are as follows: (1) Direct comparison of median prediction of PGA from GMPEs for different regions (Campbell and Bozorgnia 2006; Abrahamson et al. 2008; Gregor et al. 2014), (2) analysis of variances (Douglas 2004), and (3) determining the consistency of data distribution with respect to GMPE using log likelihood (Delavaud et al. 2009) and Euclidian distance (Kale and Akkar 2013) concepts.

Examples of comparison of GMPEs are shown in Fig. 1(a & b). Estimated peak horizontal acceleration calculated from NA09, NDMA10, and AN13 is compared with the NGA-West-2 (AB14, BA14, CB14, CY14, and ID14), Japan (ZH16), and Pan-European model (AK14 and BD14). It can be seen from Fig. 1 (a & b) that the Himalayan GMPEs are not matching well with the NGA-2 West models over the entire range of distance and moment magnitude. The attenuation of the PGA is sharp in case of Himalayan GMPEs as compared to NGA-West 2 models, whereas NGA-west 2 models have faster distance attenuation. However, ZH16 GMPEs for subduction slab and interface earthquake is matching well with the Himalayan GMPE. From Fig. 1(a & b), it can also be seen that as far as recorded Himalayan data is concerned, NGA-West 2 and Pan-European GMPEs are under-predicting the peak ground acceleration (PGA) values. This is the direct way for comparing the different GMPEs. Another approach for comparing the GMPEs is by analyzing the variance. This approach was used by Douglas (2004) to relate the ground motion within the Europe region. In this approach, mean and variance of log of data inside particular magnitude and distance bins for two different regions and combined data for those regions have been calculated. The result has been further used in following two ways: firstly, the variance of the combined data for both the regions would be compared with the variance within the region using statistical test, and secondly, the binned results would be used to plot the mean for each magnitude-distance bin together for the pair of regions.

Table 1 Magnitude scaling function used for NGA-West 2, Himalayan, Pan-European, and Japan GMPEs

S. no.	Dominant region	GMPE	Abb.	M-scaling
1	California	Campbell and Bozorgnia (2014)	CB14	$\begin{cases} c_o + c_1 M & M \leq 4.5 \\ c_o + c_1 M + c_2(M-4.5) & 4.5 < M \leq 5.5 \\ c_o + c_1 M + c_2(M-4.5) + c_3(M-5.5) & 5.5 < M \leq 6.5 \\ c_o + c_1 M + c_2(M-4.5) + c_3(M-5.5) + c_4(M-6.5) & M > 6.5 \end{cases}$
2	California	Idriss (2014)	ID14	$a_1 + a_2 M + a_3(8.5 - M)^2$
3	California	Abrahamson et al. (2014)	AB14	$\begin{cases} a_1 + a_5(M-M_1) + a_8(8.5-M)^2 & M > M_1 \\ a_1 + a_4(M-M_1) + a_8(8.5-M)^2 & M_2 \leq M < M_1 \\ a_1 + a_4(M_2-M_1) + a_8(8.5-M_2)^2 + a_6(M-M_2) + a_7(M-M_2)^2 & M < M_2 \end{cases}$ Note: $M_1 = 6.75$, $M_2 = 5$ for PGA
4	California	Chiou and Youngs (2014)	CY14	$c_2(M-6) + \frac{c_2 - c_3}{c_n} \ln\left(1 + e^{c_n(c_M - M)}\right)$
5	California	Boore et al. (2014)	BA14	$\begin{cases} e_o + e_4(M-M_h) + e_5(M-M_h)^2 & M \leq M_h \\ e_o + e_6(M-M_h) & M > M_h \end{cases}$ Note: $M_h = 5.5$ for PGA; scaling function used is of unspecified fault type.
6	Himalayan	Nath et al. (2009)	NA09	$c_1 + c_2 M + c_3(10 - M)^3$
7	Himalayan	Sharma et al. (2009)	SH09	$b_1 + b_2 M$
8	Himalayan	Gupta (2010)	GU10	$c_1 + c_2 M$
9	Himalayan	NDMA (2010)	NDMA10	$c_1 + c_2 M + c_3 M^2$
10	Himalayan	Anbazhagan et al. (2013)	AN13	$c_1 + c_2 M$
11	Pan-European	Akkar et al. (2014)	AK14	$\begin{cases} a_1 + a_2(M-c_1) + a_3(8.5-M_w)^2 & M \leq c_1 \\ a_1 + a_7(M-c_1) + a_3(8.5-M_w)^2 & M > c_1 \end{cases}$, $c_1 = 6.75$
12	Pan-European	Bindi et al. (2014)	BD14	$\begin{cases} b_1(M-M_h) + b_2(M-M_h)^2 \\ b_3(M-M_h) \end{cases}$ $\begin{array}{l} M \leq M_h \\ M > M_h \end{array}$, $M_h = 6.75$
13	Japan	Zhao et al. (2016 a, b & c)	ZH16	Subduction interface earthquakes (ZH16SI) $\begin{array}{ll} c_{intD/intS}(m_i) & m_i \leq m_c \\ c_{intD/intS}(m_c) + d_{int}(m_i - m_c) & m_i > m_c \end{array}$ c_{intD} and c_{intS} respectively in case of deep and shallow earthquakes, $m_c = 7.1$
				Subduction slab earthquakes (ZH16SS) $\begin{array}{ll} c_{SL1}(m_i) + c_{SL2}(m_i - m_{sc})^2 & m_i \leq m_c \\ c_{SL1}(m_i) + c_{SL2}(m_i - m_{sc})^2 + d_{SL}(m_i - m_c) & m_i > m_c \end{array}$ $m_c = 7.1$ and $m_{sc} = 6.3$
				Shallow crustal and upper-mantle earthquakes (ZH16SM) $\begin{array}{ll} c_{cr}(m_i) & m_i \leq m_c \\ c_{cr}(m_c) + d_{cr}(m_i - m_c) & m_i \leq m_c \end{array}$ $m_c = 7.1$

The values of the regression coefficients and validity of a respective GMPE given in the above table can be referred from the respective paper* stands for Abbreviation

Abb. Abbreviation

Table 2 Distance scaling function used in NGA-West 2, Himalayan, Pan-European, and Japan GMPEs

S. no.	Dominant region	GMPE	Abb.	R-scaling	Notes
1	California	Campbell and Bozorgnia (2014)	CB14	Geometric attenuation $(c_5 + c_6 M)\ln\left(\sqrt{R_{RUP}^2 + c_7^2}\right)$ Anelastic attenuation $\begin{cases} c_{20}(R_{RUP}-80) & R_{RUP} > 80 \\ 0 & R_{RUP} < 80 \end{cases}$	R_{RUP} is the closest distance to the coseismic fault rupture length, and Δc_{20} is not used as this is only for Japan and China.
2	California	Idriss (2014)	ID14	$(\beta_1 + \beta_2 M)\ln(R_{RUP} + 10)$	R_{RUP} is the closest distance to the rupture surface.
3	California	Abrahamson et al. (2014)	AB14	$\begin{cases} [a_2 + a_3(M-M_1)]\ln(R) + a_{17}R_{RUP} & M > M_1 \\ [a_2 + a_3(M-M_1)]\ln(R) + a_{17}R_{RUP} & M_2 \le M < M_1 \\ [a_2 + a_3(M_2-M_1)]\ln(R) + a_{17}R_{RUP} & M < M_2 \end{cases}$ $R = \sqrt{R_{RUP}^2 + c_{4M}^2}$ $c_{4M} = \begin{cases} c_4, & M > 5 \\ c_4 - (c_4-1)(5-M), & 4 < M \le 5 \\ 1, & M \le 4 \end{cases}$	$M_1 = 6.75$, $M_2 = 5$ for PGA, R_{RUP} is the rupture distance, and c_{4M} is refereed as the fictitious depth.
4	California	Chiou and Youngs (2014)	CY14	$c_4\ln(R_{RUP} + c_5\cosh(c_6\cdot\max(M-c_{HM}, 0)))$ $+ (c_{4a}-c_4)\ln\left(\sqrt{R_{RUP}^2 + c_{RB}^2}\right)$ $+ \left\{c_{\gamma1} + \dfrac{c_{\gamma2}}{\cosh(\max(M-c_{\gamma3},0))}\right\}R_{RUP}$	R_{RUP} is the closest distance to the ruptured plane.
5	California	Boore et al. (2014)	BA14	$[c_1 + c_2(M - M_{ref})]\ln(R/R_{ref}) + c_3(R - R_{ref})$ $R = \sqrt{R_{JB}^2 + h^2}$	R_{JB} is the closest distance to the surface projection of the fault plane, $R_{ref} = 1$ and $M_{ref} = 4.5$.
6	Himalayan	Nath et al. (2009)	NA09	$c_4\ln(R_{RUP} + c_5 e^{c_6 M})$	R_{RUP} is the rupture distance.
7	Himalayan	Sharma et al. (2009)	SH09	$b_3\log\left(\sqrt{R_{JB}^2 + c_4^2}\right)$	R_{JB} is the Joyner-Boore distance.
8	Himalayan	Gupta (2010)	GU10	$c_4 R - g\log(R)$ $R = \sqrt{D_{fault}^2 + \Delta^2}$, D_{fault} as the closest distance of the fault rupture surface, $\Delta = 0.00724 \times 10^{0.507M}$	R is a distance metric with near-source saturation effect, g is the geometric attenuation factor, and Δ accounts for distance saturation.
9	Himalayan	NDMA (2010)	NDMA10	$C_4 r + C_5\ln(r + C_6 e^{C_7 M}) + C_8\log(r)f_0$ $f_0 = \max(\ln(r/100), 0)$	r is the hypocentral distance.
10	Himalayan	Anbazhagan et al. (2013)	AN13	$b\log(X + e^{c_6 M})$ $X = \sqrt{R^2 + h^2}$	R is the closest distance to the rupture, and h is the focal depth.
11	Pan-European	Akkar et al. (2014)	AK14	$[a_4 + a_5(M_w - c_1)]\ln\left(\sqrt{R^2 + a_6^2}\right)$	R is the source to site distance measure and coefficients correspond to R_{JB}; R_{epi} and R_{hypo} are given.

Table 2 (continued)

S. no.	Dominant region	GMPE	Abb.	R-scaling	Notes
12	Pan-European	Bindi et al. (2014)	BD14	$\left[c_1 + c_2\left(M-M_{ref}\right)\right]\log\left(\sqrt{R^2+h^2}/R_{ref}\right)$ $+c_3\left(\sqrt{R^2+h^2}-R_{ref}\right)$	R is the source to site distance measure, and coefficients correspond to R_{JB} and R_{hypo} is given
13	Japan	Zhao et al. (2016 a, b & c)	ZH16	$g_{int/crsl}\log(r_{i,j}) + g_{intL/crtL/slL}\log(x_{i,j}+200) + e_{int/crt/slt}x_{i,j}$ $r_{i,j}=x_{into}+x_{r,j}+\exp(c_1+c_2C_m)$ $C_m = \begin{cases} m_i, & m_i \le C_{max} \\ C_{max}, & m_i > C_{max} \end{cases}$ Here, i denotes the event and j denotes the jth record of ith event.	$x_{i,j}$ denotes the shortest distance from a recording station to fault plane if a fault plane is available, otherwise the hypocentral distance.

The values of the regression coefficients and validity of a respective GMPE given in the above table can be referred from the respective paper. It is also recommended to use the GMPE for the defined distance whether its closest distance or Joyner–Boore distance* stands for Abbreviation

Abb. Abbreviation

The detailed procedure is given in Douglas (2004). In addition to this, Bommer et al. (2010) and Stewart et al. (2015) have recommended certain criteria for selecting a GMPE for a particular tectonic region.

The information-theoretic approach was developed by Delavaud et al. (2009), which is based on quantitative assessment of GMPEs for a region, which can also be used for GMPE selection. This approach made use of log-likelihood to rank the GMPE based on the observed PGA or intensity of a region. This method was used by Nath and Thingbaijam (2011) and Anbazhagan et al. (2016) for the selection of the suitable GMPEs for Indian subcontinent. Log-likelihood (LLH) method was also used by and Zafarani and Farhadi (2017) to test the efficiency of some selected GMPEs against small-to-moderate recorded data in the Iranian plateau. Another method developed by Kale and Akkar (2013) makes use of modified Euclidean distance for ranking the GMPE. This method accounts the aleatory variability through standard deviations of GMPEs and considers the bias between the median estimations and observed ground-motion data. The bias can be identified as analogues to the residual analysis concept. Further, uncertainty in ground-motion variability could be addressed by calculating the probability distribution of the difference between the observed and calculated data. This method is different from LLH ranking, as LLH-based ranking makes use of the occurrence probability of the observed data point by using the corresponding estimation that is assumed to be log-normally distributed with median and sigma values of the GMPE (Kale and Akkar 2013). Delavaud et al.'s (2009) LLH method is used to rank the different functional form discussed above and also given in Tables 1 and 2. The ranking value of these GMPEs based on LLH approach is given as Table 4 in Appendix 1.

Even though, there are qualitative and quantitative methods (as discussed above) to select the representative GMPEs for seismic hazard analysis for a particular region; however, suitability of GMPE functional form to capture the distance attenuation and magnitude scaling is not explicitly addressed in any of the methods. Hence, in this study, GMPE functional form for magnitude and distance scaling is derived based on the mixed effect regression on the residuals. The whole methodology is applied to the recorded strong motion data for the Himalayan region. In the further sections, methodology for determining the suitable GMPE functional form is discussed.

4 Methodology

4.1 Calculation of GMPE bias and standard deviation

The methodology used in this study regarding the selection of a particular functional form corresponding to distance and magnitude scaling has been discussed in this section. The residual between the data is evaluated as

$$(R_{i,j})_k = \ln(SA_{i,j})_{data} - \ln(SA_{i,j})_k \tag{1}$$

Index i, j, and k, respectively, refers to the earthquake event, recording within the event i and a particular GMPE. Therefore, $(R_{i,j})_k$ is the residual of data from recording j for the event i, calculated using GMPE k. $(SA_{i,j})_{data}$ and $(SA_{i,j})_k$ respectively represent the spectral acceleration calculated using recorded data and median spectral acceleration calculated using k GMPE in natural logarithm.

Further, the variation in residuals corresponding to between the events (intraevent) needs to be separated from the variation within the events (interevent). To accomplish that, a mixed effect regression proposed by Abrahamson and Youngs (1992) and further extended by Jayaram and Baker (2010) by considering spatial correlation is used. The whole algorithm used is given in Appendix 2. The following equation has been used to perform the regression.

$$(R_{i,j})_k = c_k + (\eta_i)_k + (\epsilon_{i,j})_k \tag{2}$$

where c_k is the mean offset (or bias) of the data relative to GMPE k, η_i represents the event term for event i, and $\varepsilon_{i,j}$ is the intraevent residuals for recording j in event i. The event term η_i represents the mean offset of data for event i from the prediction provided by the GMPE median after adjusting the offset c_k. In other words, the intraevent residual $\epsilon_{i,j}$ is the residual after accounting for the interevent residual η_i. The event term would be helpful in providing a convenient mechanism for testing the ability of a GMPE to provide the magnitude scaling of a database. Interevent (η) and intraevent (ϵ) terms are assumed to have zero mean and τ and σ respectively as standard deviation. Hence, τ refers to the event-to-event variability and, on the other hand, σ refers to the variability in a single event. A schematic diagram illustrating the residuals and parameters is given in Fig. 2. Figure 2 explains the types of residuals and how the intraevent and interevent residuals can be extracted from

the total residuals. In this study, a database has been grouped according to the event to use the linear mixed effect. However, it can also be grouped based on the distance, but attenuation characteristic may be lost. Generally, most of the distance data is single (i.e., only for few events having same distance); hence, the mixed effect regression coefficients may not be statistically significant for that case.

Jayaram and Baker (2008) concluded that the total residuals (Eq. 1) calculated at multiple sites during a particular earthquake can be assumed to jointly follow a multivariate normal distribution. Hence, the covariance metric(s) for the total residual in the presence of spatial correlation is derived using Jayaram and Baker (2010). The algorithm used to calculate the standard deviation of the intraevent and interevent residual considering mixed effect models is given in Appendix 2. The other reason for using the mixed effect is to account for the dependency between multiple observations at a single recording site. This regression also allows estimating the repeatable biases and variances, when the database is grouped according to one or more classification factor like moment magnitude or site class.

4.2 Intraevent residual ($\epsilon_{i,j}$) trend with distance

The suitability of a respective functional form of a GMPE corresponding to distance scaling has been evaluated, and its adaptability for recorded ground motion is also studied. This can be tested by examining the trend of the intraevent residuals $\epsilon_{i,j}$ as a function of distance. $\epsilon_{i,j}$ is the remaining residual after mean error (c_k) and event term (η_i) are subtracted from the total residual (see Eq. 2). For determining the trend, a fit line has been plotted according to

$$\epsilon_{i,j} = a_R + b_R \ln(R_{i,j}) + (\kappa_R)_{i,j} \tag{3}$$

Parameters a_R and b_R are the regression parameter, and κ_R is the residual of the fit for recording j from event i. As per Scasserra et al. (2009), slope parameter b_R represents the misfit of the distance scaling in the recorded dataset with respect to the selected GMPE. The statistical significance of b_R is evaluated using t-statistic to test the null hypothesis, i.e., $b_R = 0$. This test gives the p value, which defines the significance level that null hypothesis cannot be rejected.

Fig. 1 Comparison of **a** PGA (g) with distance moment magnitude 6 M_w and **b** magnitude at distance of 100 km with recorded strong motion
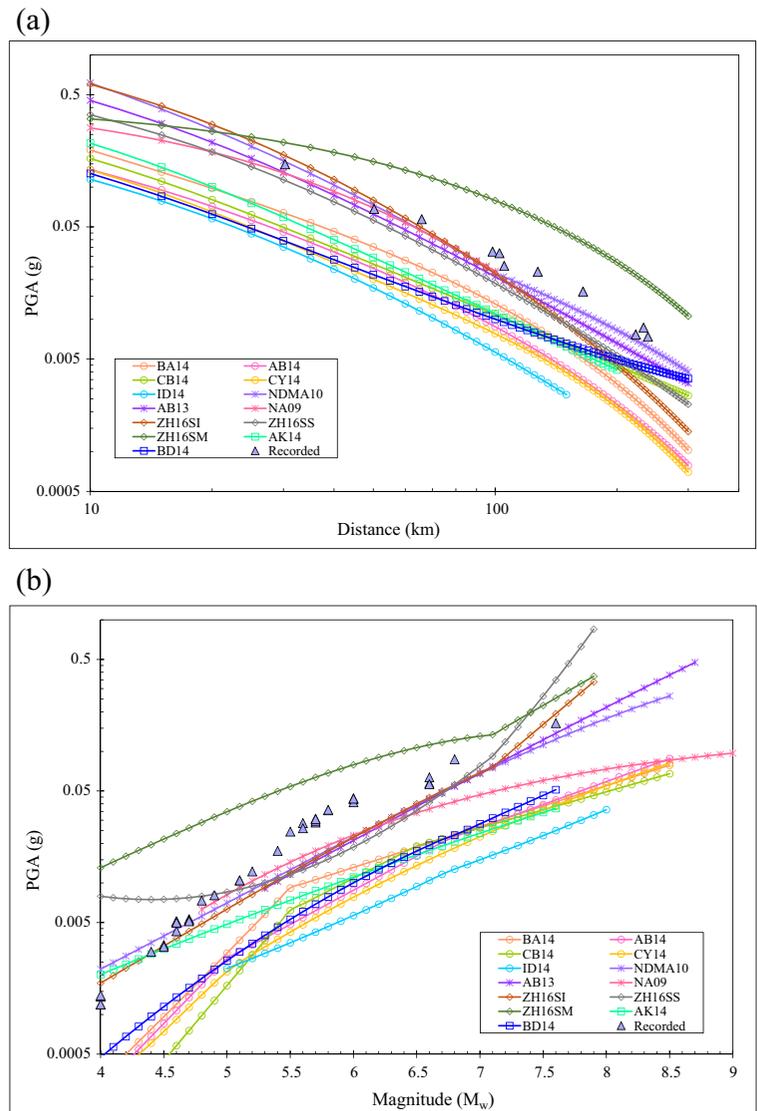
4.3 Interevent residual (η_i) trend with magnitude

Similar to distance scaling, magnitude scaling is also tested by examining the trend of the event terms versus magnitude. For evaluating the magnitude scaling, the trend of interevent term (η_i) is plotted against the magnitude and variation between the magnitudes have been studied statistically. A fit line has been plotted according to the following relation:

$$\eta_i = a_M + b_M M_i + (\kappa_M)_i \qquad (4)$$

Parameters a_M and b_M are the regression parameters, and κ_M is the residual of the fit for event i. Separate regression has been performed according to Eq. 4 for PGA and PSA (pseudo spectral acceleration). Non-zero and significant b_M term indicates that the magnitude scaling in the model does not match the data. From the t-statistics, it is determined that when the slope (b_M) of the trend line is non-zero. It is not statistically significant at 95% confidence level.

4.4 Monte Carlo and genetic algorithm method for hinge point and standard error estimation

In this section, the methodology regarding estimation of hinge point and standard errors and confidence intervals using genetic and Monte Carlo algorithm is discussed.
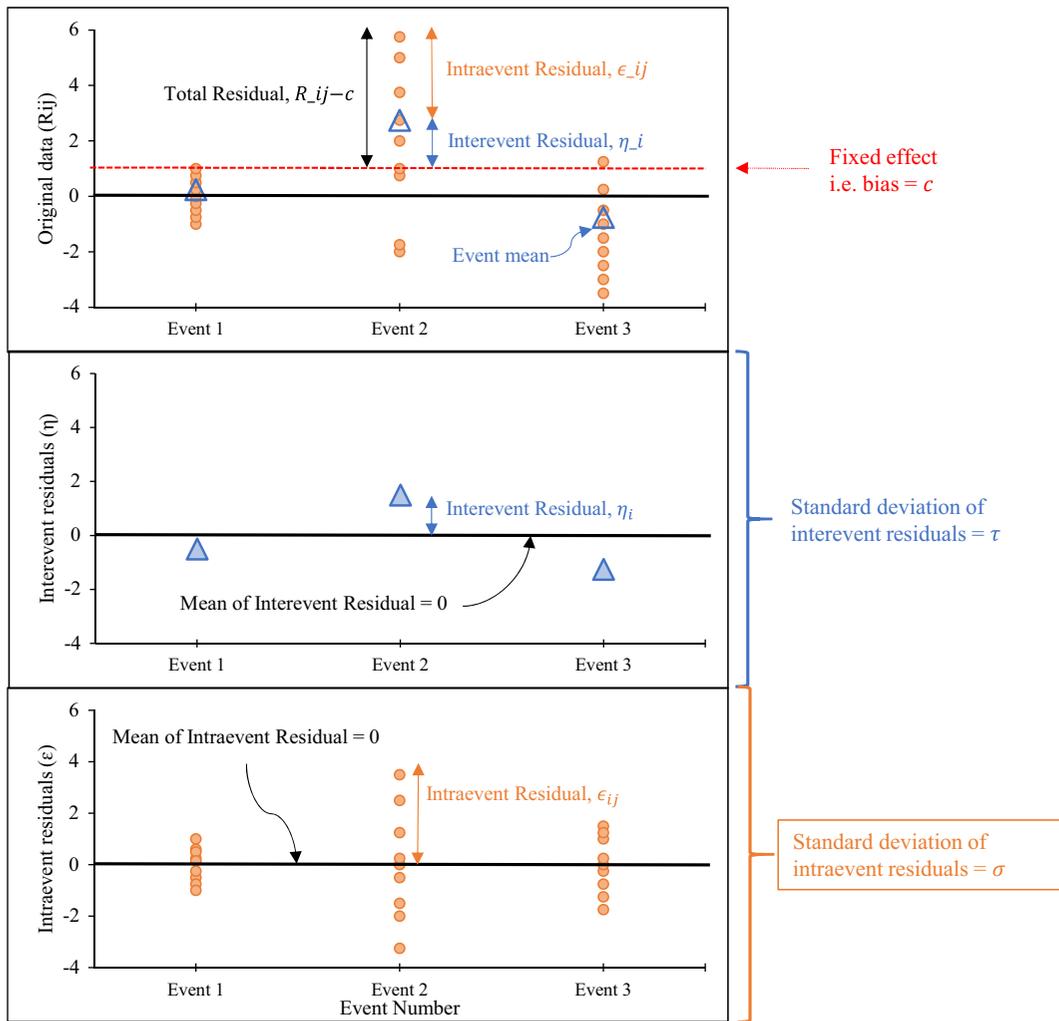
Fig. 2 Schematic diagram to illustrate the residuals and parameters in the linear mixed effect models

Recorded database is regressed against the functional form mentioned in Tables 1 and 2. In Tables 1 and 2, it can be seen that the magnitude and distance scaling is done using either three hinge or two hinged equations. The determination of hinge point based on the recorded database is very important as this is used for distinguishing the scaling of higher magnitude with lower or large distance with the shortest distance. Hence, genetic algorithm is used for determination of the hinge point. Different combinations of the magnitude and distance have been used for the hinge point determination. The best combination is the one that minimizes the mean absolute value of the residuals, defined as follows:

$$res_i = \frac{1}{N} \sum_{j=1}^{N} \left| logSA_{ij}(f) - \overline{logSA_i(f)} \right| \qquad (5)$$

where SA is the spectral acceleration, and $\overline{logSA_i(f)}$ is the average of $logSA_{ij}(f)$ over all N stations, subscript by j

$$\overline{logSA_i(f)} = \frac{1}{N} \sum_{j=1}^{N} logSA_{ij}(f) \qquad (6)$$

Therefore, the best solution is the one that results in almost identical SA for stations that recorded the same events.

In this study, a genetic algorithm (GA) is used to determine the best combination of the variables. The

GA emphasizes on a random initial population of variables, which are generated in a range defined by constraints. Variables are grouped in sets, each of which is called a string and composed of a series of characters that define a possible solution for the problem. The objective function and the constraints are used to evaluate the performance of the variables, which is represented as the fitness of each string. Mathematically, the fitness function is used to determine a value for the solution of the objective function. The GA mainly involves three operations: selection, crossover and mutation. These operations and the algorithm have been explained by Holland (1975) and Goldberg (1989). The GA tests different combinations of the variables and finds the solution that minimizes the mean value of the errors, as defined by Eq. 5, across all events. The objective function that has been used to minimize the GA is defined as follows:

$$Objective\ function = \frac{1}{M} \sum_{i=1}^{M} res_i \qquad (7)$$

where M is the number of events.

Hence, the hinge-bilinear functional form is used with a constraint on magnitude and distance. The whole procedure is shown with its applicability to the Himalayan region. Further, to test the precision of the estimated parameters, their standard errors and confidence intervals are evaluated using the Monte Carlo simulation technique (Press et al. 2007). For this purpose, 1000 samples of the PGA are simulated using the calculated coefficients for the different sets of magnitude and distance. The magnitude varies from 4.0 to 9.0 and distance from 10 to 500 km with respective interval of 0.1 M_w and 10 km. The datasets are simulated by bootstrapping the residuals (Efron and Tibshirani 1993). The simulations are performed for the entire range of magnitude and distance. The standard errors of the estimated parameters (the bootstrap estimated of the standard error) are the standard deviation of the estimated parameters of the simulated datasets (Efron and Tibshirani 1993). There are several methods to evaluate the bootstrap confidence interval. The simplest one is the standard bootstrap confidence interval that works when the estimator is normally distributed. The $100(1 - \alpha)\%$ standard bootstrap confidence interval on the estimator parameter is defined by Efron and Tibshirani (1993) as $Estimate = \pm z_{\alpha/2}\hat{se}$, where $z_{\alpha/2}$ is the upper $100_{\alpha/2}$ percentage point of the standard normal distribution and $\hat{se}$ is the bootstrap estimate of the standard error.

For each of the simulated dataset, using the mixed effect regression, new regression coefficients corresponding to each of the independent parameters is determined. Using the newly determined coefficients,

Note:- $a_{(0)}$ is the estimated vector of regression coefficients and $a_{(1)}^s$ to $a_{(1000)}^s$ is the simulated vector of regression coefficients. $c_{k(1)}$ to $c_{k(1000)}$, $\tau_{(1)}$ to $\tau_{(1000)}$ and $\sigma_{(1)}$ to $\sigma_{(1000)}$ respectively the mean offset (or bias), standard deviation in interevent (η) and intraevent (ϵ) terms calculated from Monte Carlo analysis
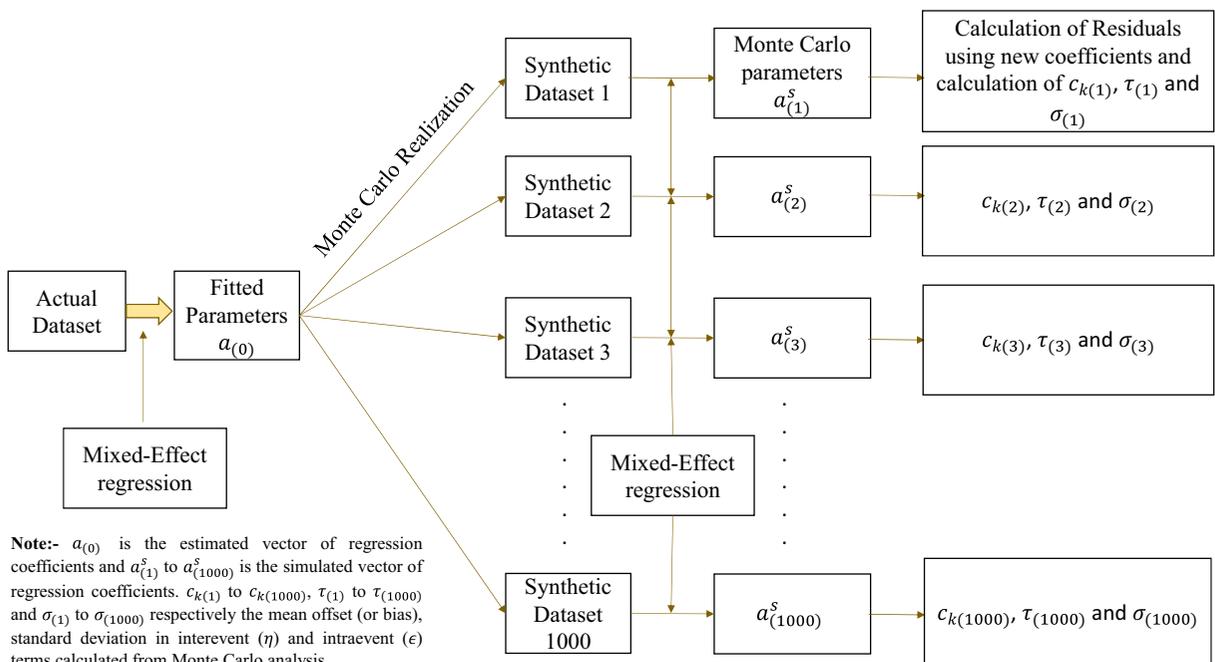
Fig. 3 Schematic diagram to illustrate variation in c_k, τ, and σ calculated using Monte Carlo technique

again residuals analysis explained above is performed. For each of the simulated dataset, c_k, τ, and σ have been calculated and compared with the original one. The schematic diagram to illustrate this is shown as Fig. 3. Further, using different sets c_k, τ, and σ, the suitable functional form is determined. In the next sections, the whole methodology is applied for the strong motion recorded database for the Himalayan region.

5 Study area and database

The tectonic framework of the Indian subcontinent is complex and spatiotemporal which covers an area of about 3.2 million km^2. The entire Himalayan arc (75° to 98° E) of around 2500 km extends from Kashmir in the northwest to Arunachal Pradesh in the northeast. The width of this arc varies from 250 to 300 km. Collision of Indian plate with the Asian plate at rate of 15 to 20 cm/ year (Bilham et al. 1998) results in building of high stress in the Indian plate. Furthermore, because of crustal shortening beside Himalayan, northern edge results in increase of the earthquake hazard, particularly in the northern part of Indian subcontinent. The Himalayan geodynamics is the reason for occurrence of the largest earthquakes (1897 Assam, 1905 Kangra, 1934 Bihar-Nepal, 1950 Assam, 2011 Sikkim, 2015 Nepal earthquake, 2016 Northeast India earthquake) in the foothills and north of the Main Boundary Thrust (MBT). Mandal et al. (2001) explained that the reactivation of the upper crustal fault due to moderate earthquakes may result in possible slip surfaces of crustal shear zones which can facilitate the upliftment of lesser as well as higher Himalaya. As per Singh et al. (2002), the 750 km long central seismic gap lies between the eastern edge of 1905 rupture zone and the western edge of the 1934 earthquake. It remains unbroken and is under high strain. A recent study by Bilham (2015) on Gorkha (2015) concluded that the northern Nepal shifted up to 7 m southward and Kathmandu was raised by 1 m. This study further describes that Gorkha (2015) earthquake failed to fully rupture the main fault beneath the Himalayas and hence a large earthquake is inevitable in the future. Currently, the Indian plate is continuously moving northwards at about 5 cm per year towards the Eurasian plate which may give rise to massive earthquakes in the near future (Kayal 2008). Hence, proper seismic studies are required for the Himalayan region to predict catastrophic hazards due to the future inevitable

earthquakes. The present study concentrates on determining the suitable functional form of GMPE for Himalayan region that would be further useful in the development of new GMPE and estimation of seismic hazard for the region.

The dataset has been collected from the strong motion instrumentation network that covers the Indian Himalayan range from Jammu and Kashmir to Meghalaya. The recorded strong motion database is available with Indian Institute of technology, Roorkee (IITR), Virtual Data Center (VDC), and Indian seismic and GNSS network (ISGN). The installation of these instruments with IITR started in November 2005 and recorded around 130 earthquakes in a span of 4 years. Detailed description of these strong motion accelerographs and data processing of the waveforms are given in Kumar et al. (2012). Before 2005, the digital processed waveform of the recorded ground motion has been collected online from Virtual Data Center and after 2005 from PESMOS (www.pesmos.in), which is maintained by the department of earthquake engineering of IITR. The database after 2013 is collected from the Indian seismic and GNSS network. Details regarding the seismographs units and broadband sensor are taken from the ISGN website (http://www.isgn.gov.in/ISGN/). The database used in this study consists of 512 ground motions recorded at rock and soil sites from 66 earthquakes with moment magnitude varying from 4 to 7.8 and hypocentral distance from 10 to 500 km. The database also included the 2015 Nepal earthquake recording. The data obtained from VDC is baseline corrected and band-pass filtered between 0.75–0.9 and 25–27 Hz. However, the waveforms obtained from PESMOS and ISGN are high-passed filtered at corner frequency of 0.1 Hz and low-pass filtered at 25 Hz using fourth order Butterworth filter. The processed acceleration time-history is further integrated to obtain the velocity time history and spectral acceleration at various time periods. The database used in this study is given in Fig. 4. From Fig. 4, it can be seen that for $M_w < 5$ strong motion database up to 200 km is considered for analysis. As for distance more than 200 km, the signal to noise ratio is very low; therefore, these data are rejected.

For most of the strong motion recordings used in this study, we do have proper information regarding R_{RUP} and R_{JB}. Hence, for distance matrix, hypocentral distance (R_{hypo}) is taken into consideration for distance scaling irrespective of the definition used in the original GMPE. From here onwards, distance is referred as

Fig. 4 Distribution of NGA-West 2 and Himalayan database with respect to moment magnitude and hypocentral distance
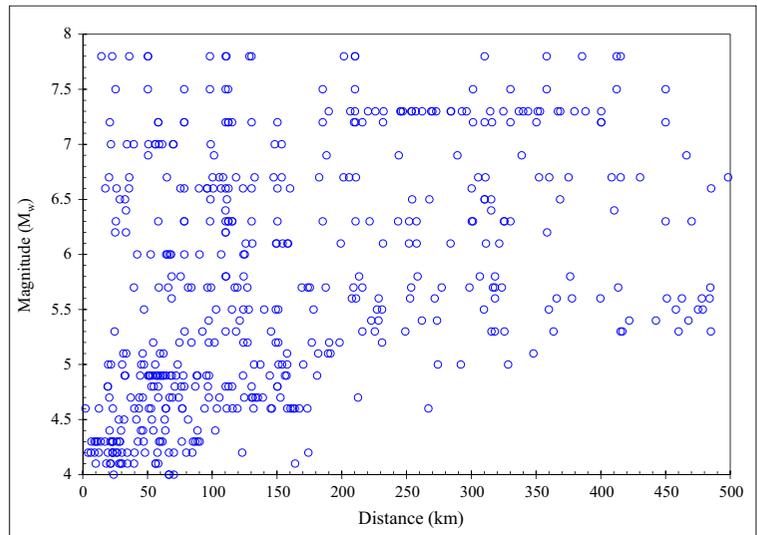
hypocentral distance. Similarly, moment magnitude is used for magnitude scaling irrespective of the definition of the magnitude used in a parent GMPE. For most of the earthquakes, the detailed fault plane solution is not available; hence, hanging wall effect is not considered in this study, and only R_{hypo} has been used for further analysis. As proper shear wave velocity or amplification factor corresponding to recorded stations is not available, shear wave velocity scaling is not considered in this study.

6 An insight of function form for the Himalayan region

From Tables 1 and 2, it can be seen that NGA-West 2, Himalayan, Pan-European, and Japan GMPEs were derived considering different functional forms. Hence, for deriving a new GMPE for the Himalayan region, the best suitable functional form of a GMPE needs to be arrived so that path and source parameters could be properly incorporated. Hence, in this section, GMPE coefficients corresponding to different functional forms (see Tables 1 and 2) have been derived using recorded ground motion database for the Himalayan region. Crossed and nested mixed effect approach is used for deriving the regression coefficients corresponding to independent parameter. Pinheiro and Bates (2000) have presented a theoretical background upon the formulation and implementation of the mixed effect models. Further, Bates (2010) has provided the optimized

methods for dealing with linear, non-linear, and generalized mixed effect models. The detailed mathematical procedure regarding the mixed effect model can be further referred from Pinheiro and Bates (2000) and Bates (2010). For implementing the mixed effect, *lme4* R package in Bates et al. (2013) is used for determining the regression coefficients for independent parameters.

The newly derived coefficients are statistically tested to determine the bias in these coefficients corresponding to different functional forms. Based on the residual analysis (using Eqs. 1 and 2) and statistical test on residuals, functional form that properly represents the Himalayan data has arrived. From Tables 1 and 2, it can be seen that there are 11 functional forms corresponding to magnitude scaling (as SH09, GU10, and AN13 have the same functional form) and 13 functional forms corresponding to distance scaling. Using all the functional form mentioned in Tables 1 and 2, respective regression coefficients have been derived using the mixed effect regression approach as explained above. After determining the coefficients corresponding to each of the functional form, residual analysis has been carried out. Using the results from residual analysis, the suitability of functional form that represents the magnitude and distance scaling for the Himalayan data has been discussed.

Figure 5 provides plots for typical example of intraevent residual (right) $\epsilon_{i,j}$ and interevent residual (left) η_i for AB14 and NA09 GMPE model as given in Tables 1 and 2. Using Eq. 2, model bias, i.e., c_k for 13 functional forms, has been calculated. The model bias

has been initially used for determining the best suitable functional form of GMPE corresponding to the Himalayan data. For example, based on analysis, c_k for AB14 is calculated as 0.0185 which means the residual in PGA value across all ground motion records is 0.0185 or the average ratio of observed and predicted PGA is equal to exponential of 0.0185, i.e., 1.02. The bias is shown in Fig. 5a which can be read as the distance between the thick black and green line. Table 3 provides the bias value in the PGA and SA at 0.2 and 1.0 s corresponding to the different functional forms of GMPEs. It can be

seen from Table 3 that the bias in PGA for the Himalayan GMPE functional forms is more as compared to the NGA-West 2, Japan, and Pan-European GMPEs. Additionally, the site-to-site standard deviation varies from 0.6 to 1.25 for different functional forms.

Based on the analysis, it is seen that the bias in PGA in case of BA14, AB14, ZH16SI, and ZH16SS is less as compared to the other three functional forms of the NGA-West 2, Japan, and Pan-European GMPEs. However, functional form of BA14 and BD14 is the same even though the bias is different. Similarly, for all the
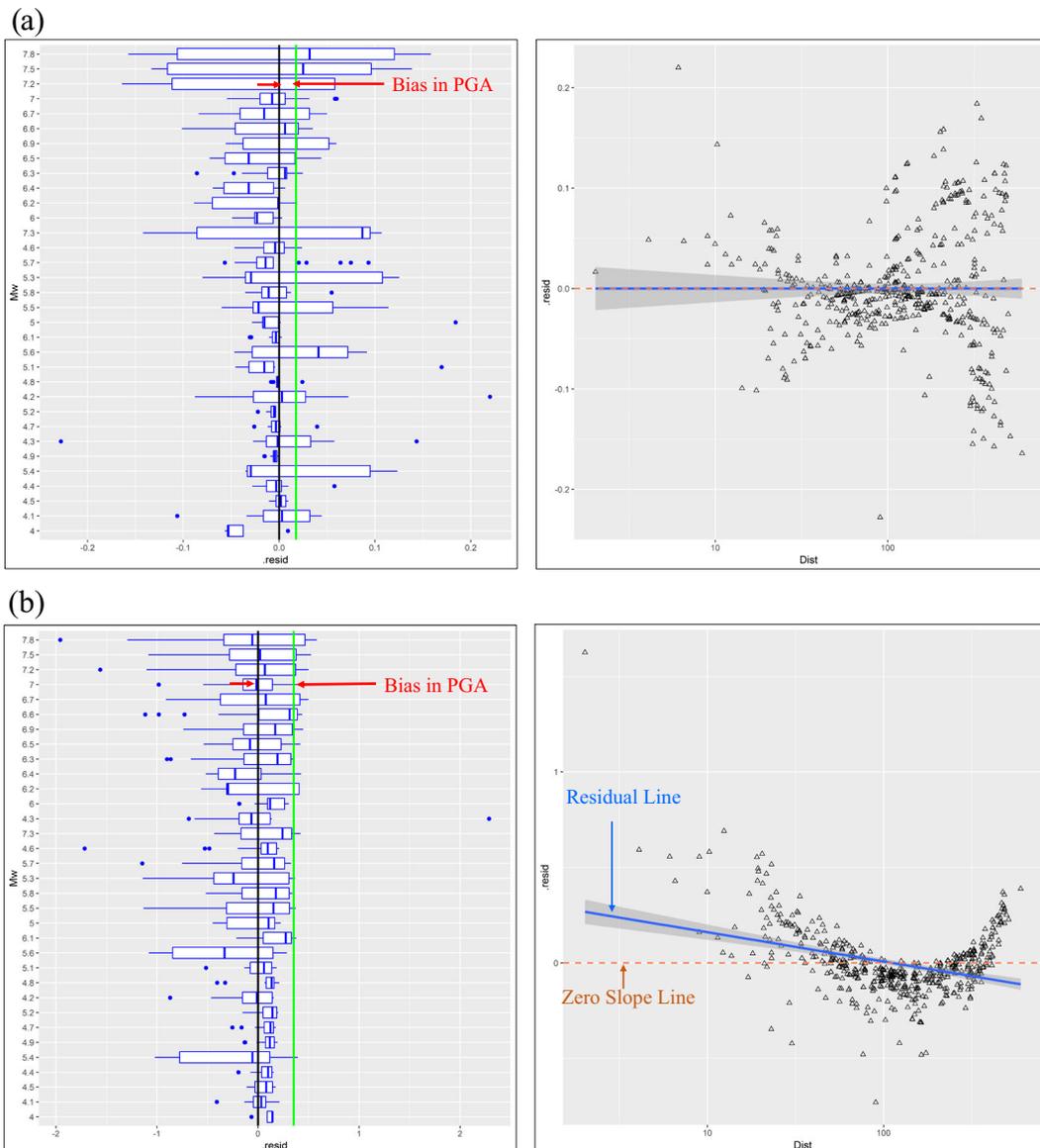
Fig. 5 Residuals for PGA for a AB14 and b NA09 models. Extreme left shows the interevent residuals with magnitude, and extreme right shows the intraevent residuals with logarithm of distance

Table 3 Summary of bias in PGA, SA (0.2 s), and SA (1.0 s) along with the 1-p value (i.e., rejection confidence of zero slope of residual line for respective functional forms of different models)

GMPE	Bias			1-p value		
	PGA	SA (0.2 s)	SA (1.0 s)	PGA	SA (0.2 s)	SA (1.0 s)
CB14	0.035	0.051	0.055	0.12	0.15	0.16
ID14	0.045	0.085	0.078	0.15	0.17	0.13
AB14	0.018	0.021	0.022	0.07	0.07	0.07
CY14	0.038	0.051	0.045	0.45	0.58	0.52
BA14	0.025	0.032	0.035	0.09	0.09	0.09
NA09	0.305	0.415	0.412	0.95	1.00	1.00
SH09	0.258	0.328	0.388	0.58	0.66	0.71
GU10	0.158	0.201	0.245	0.29	0.31	0.35
NDMA10	0.146	0.203	0.198	0.28	0.35	0.30
AN13	0.277	0.325	0.310	0.92	1.00	1.00
AK14	0.112	0.168	0.195	0.31	0.32	0.35
BD14	0.048	0.053	0.092	0.15	0.16	0.19
ZH16SI	0.034	0.042	0.056	0.13	0.13	0.13
ZH16SS	0.028	0.025	0.028	0.08	0.08	0.08
ZH16SM	0.148	0.197	0.125	0.20	0.20	0.20

Himalaya models, bias in PGA is more in the case of AN13, NA09, and SH09 as compared to NDMA10 and GU10. In some cases, like ID14, CY14, NDMA10, and AN13, bias is less for long periods. However, for AB14 and ZH16SS, bias in PGA and SA value is almost same (see Table 3). The reason for more bias in case of AN13, NA09, and SH09 may be due to non-consideration of the material damping term discussed in detail at the end. Improvement is observed in bias value in case of CY14, AN13, and NA09 as material damping term becomes less significant for long periods for the Himalayan strong motion database. One of the important things that are noticed from Table 3 is that even though the functional forms are the same, the bias value is different for example BA14 and BD14. One of the possible reasons could be the hinge magnitude as BD14 is hinged at 6.75 M_w whereas BA14 at 5.5 M_w. Hence, hinge has an important role in a functional form as it differentiates the scaling of low to moderate earthquake with the large size earthquake. This hinge magnitude determination is discussed in later sections using a genetic algorithm.

Similar observations can be seen in the case of residual variation with respect to the distance. In case of NA09 and AN13, the slope of the residual line (solid line) is not zero which means that heteroscedasticity characteristic of the residuals exists. The statistical significance of the distance dependent intraevent residuals

is examined using t-statistics to test the null hypothesis, i.e., slope of the residual line is zero. The statistical testing provides a significance level, i.e., p value, which indicates that the null hypothesis cannot be rejected. Table 3 is provided with the 1-p value which refers as a rejection confidence for a zero slope model. This p value is also one of the factors in finding the best suitable GMPE functional form for the Himalayan data. The reason for high 1 p value in case of NA09 and AN13 is explained above. Based on a variation of residuals with distance corresponding to the different magnitude, it is seen that the functional form used in NGA-West 2 and Pan-European is having more error in case of lower magnitude, i.e., $M_w < 5$. In case of the Himalayan GMPE models, mixed effect has been seen, i.e., error in both high and low magnitudes.

The Kolmogorov–Smirnov goodness-of-fit test (Ang and Tang 2007) is used for examining the cumulative distribution of the residuals so that the statistical significant departures from the residuals having a standard normal distribution can be identified. In order to demonstrate the key trends in the observed residuals as a function of the predictor variables, the mean residual and its 95% confidence interval are studied using non-parametric regression (Ruppert et al. 1995; Wasserman 2006). The non-parametric mean and its confidence intervals can be used to recognize the statistically
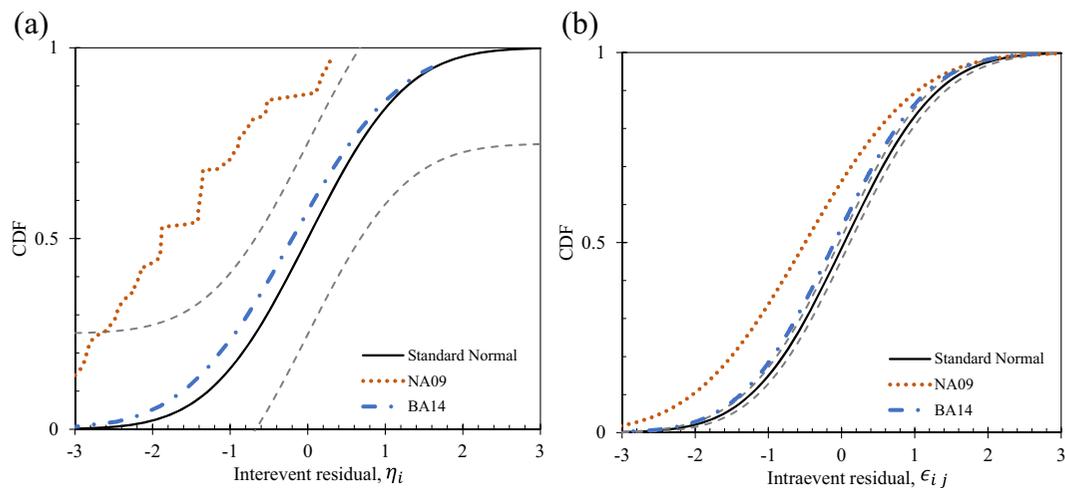
Fig. 6 Residual for SA at 0.2 s using NA09 and BA14. **a** Distribution of interevent residuals. **b** Distribution of intraevent residuals

significant biases in the functional forms. To illustrate this, the observed intra- and interevent for SA at 0.2 s from the Himalayan database with BA14 and NA09 models have been plotted in Fig. 6. It can be seen from Fig. 6a that NA09 over predicts the SA value at 0.2 s as compared to BA14. Similar observations can be seen from Fig. 6b. A same comparison has been carried out for all the 13 models (see Tables 1 and 2), corresponding to PGA and SA at 0.2 and 1.0 s. Further, the applicability of the models has been tested using the Log Likelihood method. In order to do a likelihood calculation on residual or error for different models, the probability density function needs to be chosen. The error/residuals, the difference between the predicted and observed values, are assumed to follow a normal distribution. This probability model compares the predicted and observed values to produce the final likelihood. A model that produces higher likelihood value is better. For a given set of values in a vector X, with individual observation x_i, and a set of parameter value θ, the log-likelihood function is defined as $\ln[L(\theta|X)]$ $= \sum_{i=1}^{N} \ln[g(x_i|\theta)]$, where $\ln[L(\theta|X)]$ is the logarithm of the likelihood of the set of parameters θ given the observations X, and $g(x_i|\theta)$ is the probability density function of the probability model. For determining the log-likelihood, probability density function is used to calculate the natural log of the probability of each pair of predicted and observed values. For each data point, the mean of the probability density function is the observed

value. The point, for which the probability is being calculated, given that mean, is the predicted value.

From the above statistical analysis, it can be concluded that functional form given by AB14, BA14, ZH16SI, and ZH16SS has less bias and representing the Himalayan recoded data in a much better way as compared to other GMPEs. NA09 used a three degree polynomial for magnitude scaling, this could only be feasible for very large magnitudes for the Himalayan region, i.e., $M_w > 7$. The variation of residuals with respect to magnitude is mixed; hence, either a two hinged or three hinged functional form is suitable for magnitude scaling. The hinge point will be estimated in the next section.

Another observation made is that the most of the Himalayan GMPE models are not having material damping term in the GMPE functional form and have high bias value in the PGA and SA from recorded data. Hence, material damping term $log(PGA) \propto distance$ need to be included in future for any new GMPE for better representation of the Himalayan database. As the area over which the earthquake occur increases with increasing earthquake magnitude, the effective distance becomes greater than epicentral/hypocentral distance by an amount that increases with increasing magnitude (Kramer 2003). In most of the Himalayan GMPE functional forms, this increment is exponential of distance (see NDMA10, NA09, and AN13 in Table 2, i.e., $e^{(constant)M}$); however, based on present analysis, it is seen that this is good for large magnitude but for small magnitudes using this form over-predicts the PGA

value. Hence, these GMPEs need to be used with utter care for determining the hazard values. Therefore, to account the distance attenuation for low- and high-magnitude range, either of the distance scaling given by BA14, AB14, BD14, ZH16SI, or ZH16SS will be appropriate.

7 Distance and magnitude scaling of GMPEs

In this section, the suitability of functional form for the NGA-West, Pan-European, Japan, and Himalayan GMPEs with respect to distance and magnitude has been studied in light of the recorded ground motions. From the newly determined regression coefficients, residuals have been calculated for all the functional forms as mentioned in Tables 1 and 2. Further, the residuals have been separated into intraevent residuals and

interevent residual and respectively plotted with distance and magnitude. The scaling with V_{s30} is not determined, because the proper site response study for the recorded station for the Himalayan region is not done and also including V_{s30} may result in biasing of a particular functional form of a GMPE.

7.1 Intraevent residual ($\epsilon_{i,j}$) trend with distance

The suitability of a respective functional form of a GMPE corresponding to distance scaling for the Himalayan database has been evaluated. The whole procedure for $\epsilon_{i,j}$ is explained above. Figures 7 and 8 show the trend of $\epsilon_{i,j}$ with respect to distance for PGA for suitable and non-suitable functional form for the Himalayan strong motion database. As explained above, the slope parameter b_R (Eq. 3) represents the misfit of the distance scaling in the Himalayan dataset with respect to the
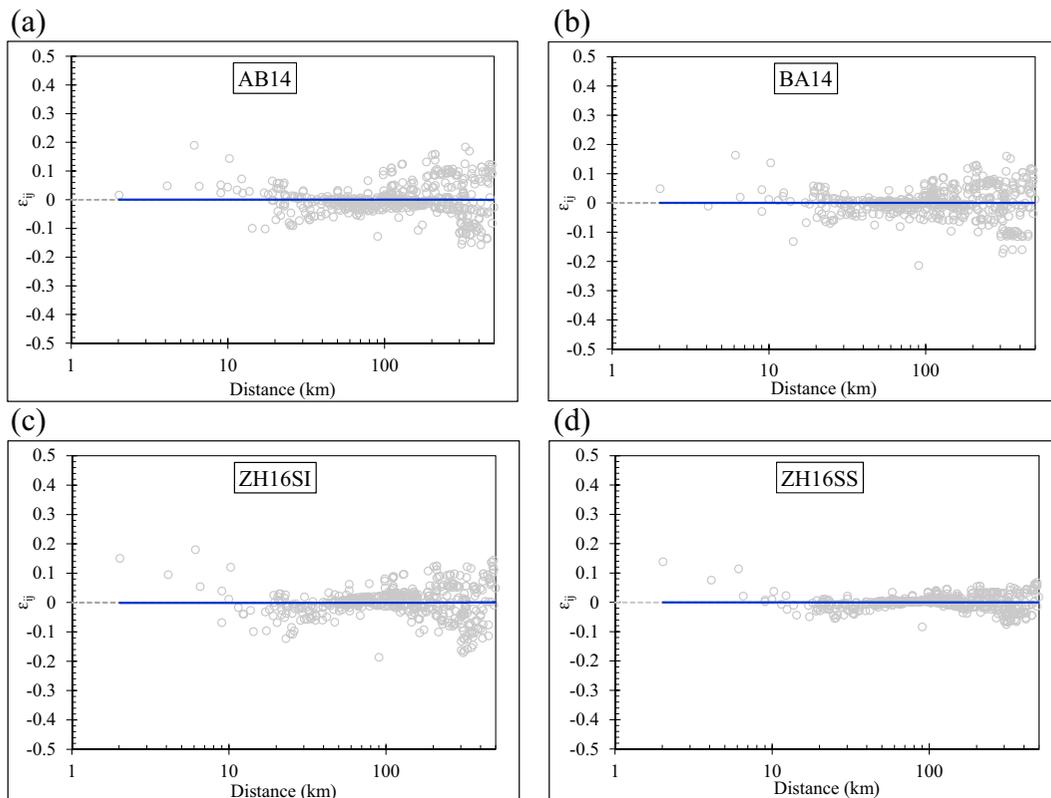
Fig. 7 Variation of intraevent residuals for the Himalayan database with distance for PGA for the suitable functional form. A typical example for (a) AB14, (b) BA14, (c) ZH16SI and (d) ZH16SS is given.
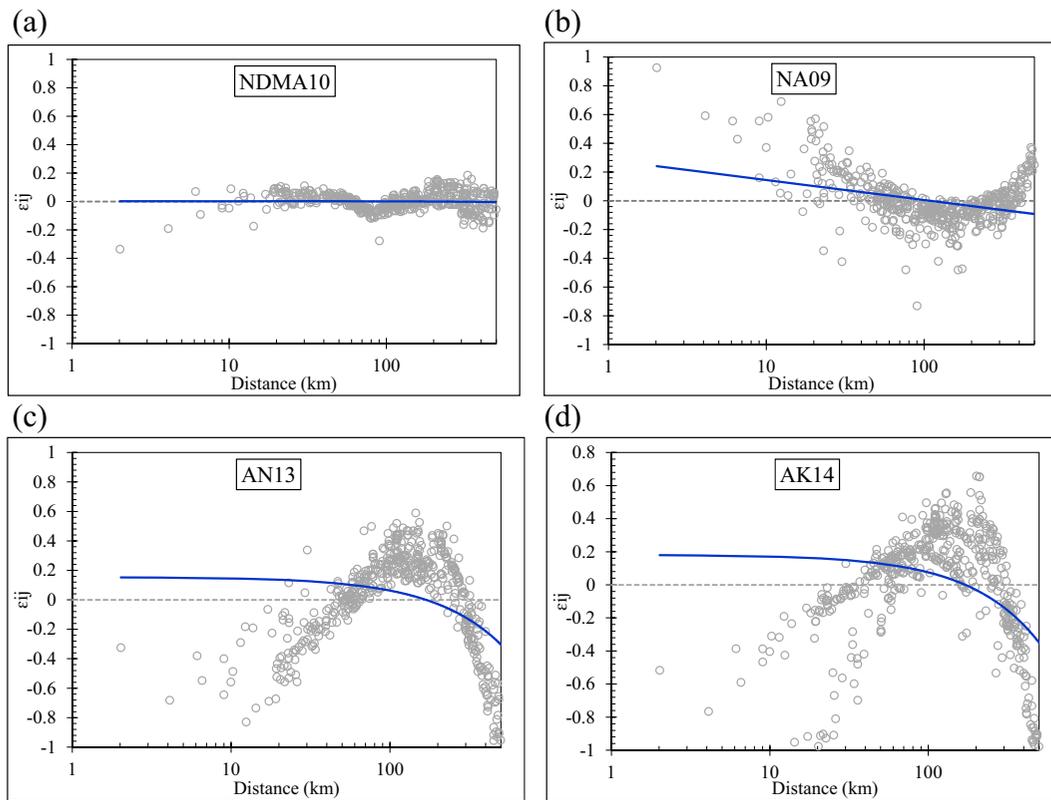
Fig. 8 Variation of intraevent residuals for the Himalayan database with distance for PGA for non-suitable functional form. A typical example for (a) NDMA10, (b) NA09, (c) AN13 and (d) AK14 is given.

selected functional form. Figures 7 and 8 also show typical median residuals within the overlapping distance bins. Similar plots for PSA at short period (0.2 s) and long period (1.0 s) have been studied. The results from Figs. 7 and 8 show the miscellaneous finding with respect to the misfits between the NGA-West 2, Pan-European, Japan, and Himalayan distance scaling with the Himalayan database. From the analysis, it is seen that the b_R value varies from − 0.005 to − 0.08, 0.02 to 0.09, and − 0.15 to − 0.25 respectively at PGA, 0.2 and 1.0 s. As b_R is less for PGA and at short period, this means that the Himalayan data is attenuating at a faster rate at long period and low at larger distance. From the *t*-statistics, it is seen from Fig. 7 that BA14, AB14, ZH16SS, and ZH16SI have the unbiased distance attention not only at short period but also at long period for the Himalayan strong motion database. The smallest and the largest b_R values are seen in ZH16SS and NA09 respectively in the Himalayan data for both long and short periods. This indicates that the distance attenuation functional form given by ZH16SS and BA14 is

following almost same trend as far as Himalayan data is concerned with AB14 and ZH16SI being the intermediate.

Similar from Fig. 8, it can be seen that even though the functional form for AN13 and NA09 is same for distance scaling, the additional coefficient of the magnitude term reduces the residual at large distances. Secondly, the exclusion of material damping term in both the equations results in heterogeneous trend of residuals. Hence, additional to regional attenuation through spreading of waves, material damping is needed to be included in GMPE functional form. The lack of three GMPEs AN13, NA09, and SH09 in estimation of attenuation of stress/seismic waves due to material composition, i.e., material damping (anelastic attenuation in other words) results in decrease of ground motion amplitude exponentially with R (Kramer 2003). The b_R indicates the attenuation of stress waves due to both spreading of waves and material damping; hence, in case of AN13 and NA09, b_R is more as compared to the NDMA10 (see Fig. 8). Similar trend is seen in both

long and short periods. From the analysis of interevent residual ($\epsilon_{i,j}$), it can be concluded that the Himalayan data is following amplitude decay with respect to distance because of both spreading and material damping. So these terms need to be included in function form of Himalayan GMPE.

7.2 Interevent residual (η_i) trend with magnitude

As explained above, similar to distance scaling, magnitude scaling is also tested by examining the trend of the event terms versus magnitude. Figures 9 and 10 show the variation of η_i with magnitude for PGA. Regression has been performed according to Eq. 4 for PGA and PSA at 0.2 and 1.0 s and given in Figs. 9 and 10. Nonzero and significant b_M term indicates that the magnitude scaling in the model does not match the data. From the visual inspection of Fig. 9, it is observed that the quadratic (ZH16SS and AB14) and linear magnitude (CB14 but with hinge) scaling is following a good trend with the Himalayan dataset. From the t-statistics, it is determined that when the slope (b_M) of the trend line is non-zero, it is not statistically significant at 95%

confidence level. Based on the Fig. 10 and t-statistics, it can be concluded that for higher magnitude, linear equation for magnitude scaling is not suitable for the Himalayan region. Even though CB14 functional form has less η_i (see Fig. 9) at higher magnitudes, a significant increase in η_i value can be seen. On testing the functional form given by NA09, it can be concluded that cubical equation from magnitude is also not suitable for the Himalayan database (see Fig. 10). On more observation is that, in case of ZH16SI and BA14 distance, functional form is capturing the attenuation properly; however, similar result is not seen for magnitude scaling for the Himalayan database. Vice versa, result is observed for CB14 and AK14 functional forms. Functional form given by AN13, GU10, and SH09 has a significant slope (i.e., b_M) of trend line. This may be because of linear equation that is used for capturing the magnitude scaling and without hinge or reference magnitude.

From the above analysis, it can be concluded that functional form given by ZH16SS and AB14 is properly capturing the attenuation and magnitude scaling for the Himalayan database. Further, it is also seen that hinge point for magnitude scaling that separates lower
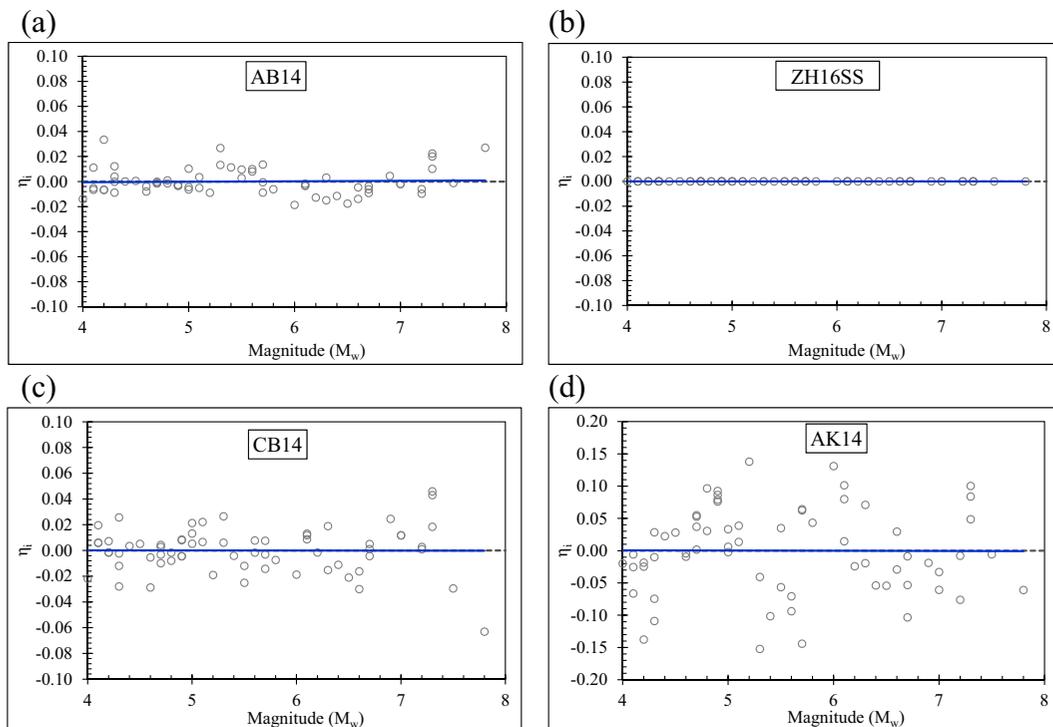
Fig. 9 Variation of interevent residuals for the Himalayan database with magnitude for PGA for the suitable functional form. A typical example for (a) AB14, (b) ZH16SS, (c) CB14 and (d) AK14 is given.

(a)

(b)

(c)

(d)
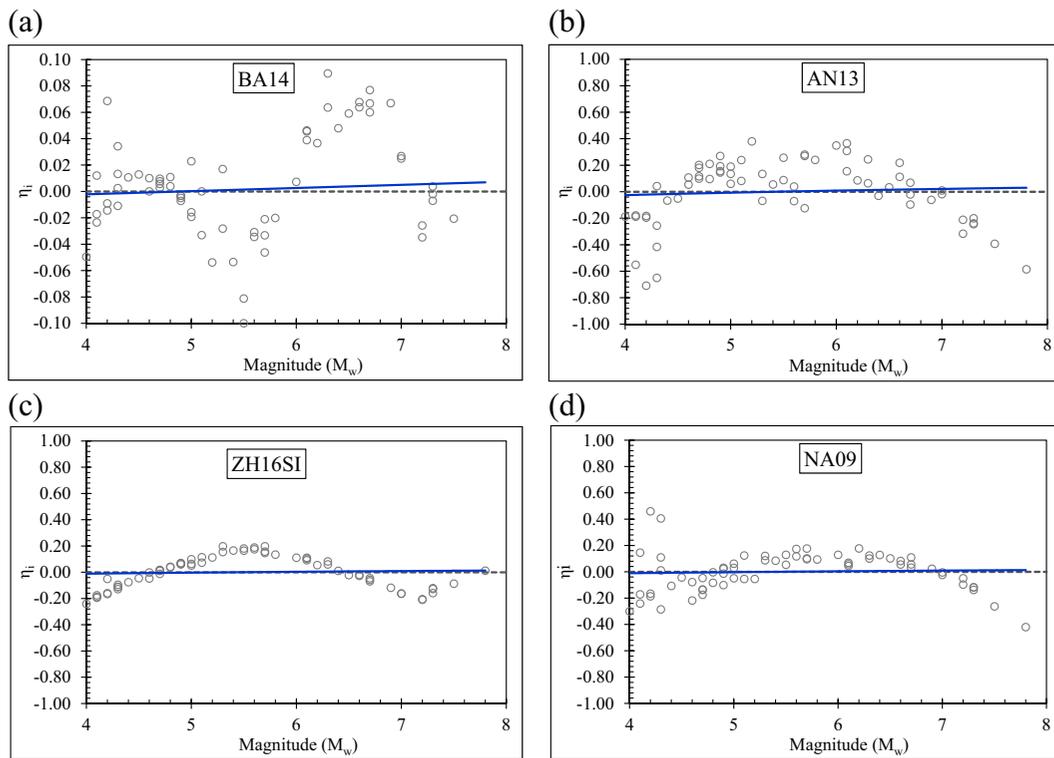
Fig. 10 Variation of interevent residuals for the Himalayan database with magnitude for PGA for non-suitable functional form. A typical example for (a) BA14, (b) AN13, (c) ZH16SI and (d) NA09 is given

magnitude to higher magnitude needs to be included in the functional form. The GMPE form that captures

source, path effects, and represents seismotectonic of the Himalayan region is given as follows:

$$\ln(SA) = \begin{cases} a_1 + a_2(M_w - M_h)^2 + \left(a_4 + a_5(M_w - M_h)\ln(R) + a_6R, & M_w < M_h \\ a_1 + a_2(M_w - M_h)^2 + a_3(M_{max} - M_w) + \left(a_4 + a_5(M_w - M_h)\ln(R) + a_6R, & M_w \geq M_h \end{cases} \qquad (8)$$

Form biased analysis (in previous section) and $\epsilon_{i,j}$ or η_i, it is seen that hinge point is much more important in magnitude scaling. Hence, in next section, genetic algorithm along with Monte Carlo method is used for determining the hinge and reference points.

7.3 Determination of hinge point and standard error

The hinge point for the magnitude scaling is calculated using GA and standard error in the coefficients using Monte Carlo approach. The bilinear hinge functional form given as Eq. 8 is used for

determining the hinge magnitude using GA. For the GA, the number of generations, number of individuals, crossover probability and mutation probability was set as 500, 50, 70, and 1%, respectively (Nakano et al. 2015). In GA, constrain is only applied to M_h in Eq. 8 and it varies from 4 to 8. The value of M_h that yielded the lowest residual value has been selected. The entire procedure is also applied to SA at 0.2 and 1.0 s. Based on the analysis; it is found that lowest residual is observed for M_h equals to 6.5.

As per Anbazhagan et al. (2015), the seismic hazard due to previous great earthquakes can be

(a)

(b)
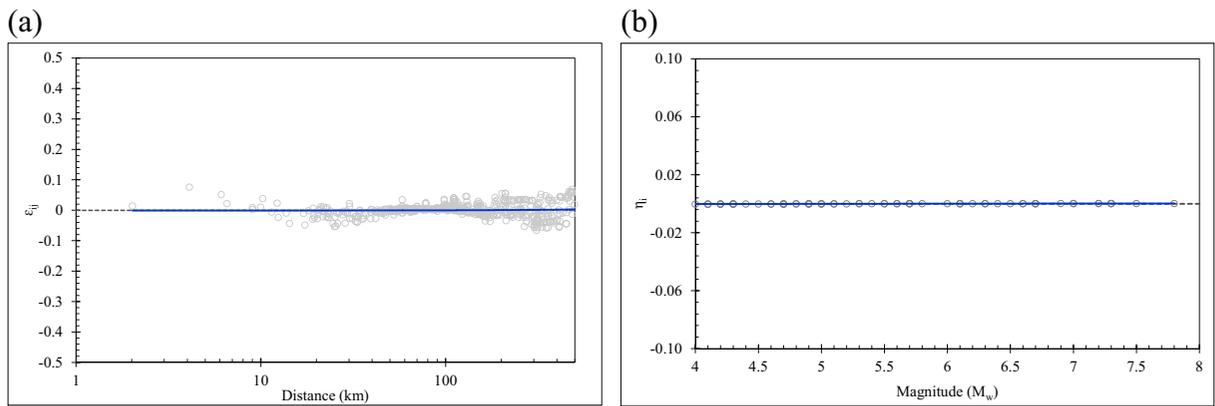
Fig. 11 Variation of **a** intraevent and **b** interevent residuals for the Himalayan database respectively with distance and magnitude for PGA for newly derived functional form

seen for more than 500 km. Hence, constraining hypocentral distance up to 300 km may not be suitable for estimation of hazard value for large earthquakes. Additionally, hinging only the magnitude may not capture the total attenuation because of great earthquake for near filed and far field. Therefore, as database is available up to 500 km for large earthquake, distance scaling is also modified. Another factor for long-distance attenuation is also added in Eq. 8. Hence, Eq. 8 is modified as

$$\ln(SA) = f(M_w) + F(R, M_w) \tag{9a}$$

$$f(M_w) = \begin{cases} a_1 + a_2(M_w - M_h)^2, & M_w < M_h \\ a_1 + a_2(M_w - M_h)^2 + a_3(M_{max} - M_w), & M_w \geq M_h \end{cases} \tag{9b}$$

$$f(R, M_w) = \begin{cases} (a_4 + a_5(M_w - M_h)ln(R) + a_6R, & R < R_h \\ (a_4 + a_5(M_w - M_h)ln(R) + a_6R + a_7ln(R + R_h), & R \geq R_h \end{cases} \tag{9c}$$

where R_h is the hinge in the hypocentral distance, and a_7 is the coefficient corresponding to long distance attenuation.

Similarly, GA is used to determine the hinge point for distance. Distance is varied from 50 to 350 km with an interval of 10 km along with magnitude. Finally, the pair of M_w and R that yielded the lowest residual value has been selected. Based on the GA, for M_w and R values at 6.5 and 300 km, respectively, lowest residual value is observed.

After determining the hinge point, the regression has been done using Eq. 9a and coefficients corresponding to independent parameters have been estimated. The equation for PGA estimation is

$$\ln(PGA) = f(M_w) + F(R, M_w) \tag{10a}$$

$$f(M_w) = \begin{cases} 5.391(0.28) + 0.121(0.01)(M_w - 6.5)^2, & M_w < 6.5 \\ 5.391(0.28) + 0.121(0.01)(M_w - 6.5)^2 + 0.356(0.025)(8 - M_w), & M_w \geq 6.5 \end{cases} \tag{10b}$$

$$f(R, M_w) = \begin{cases} -0.465(0.008) + 0.056(0.01)(M_w - 6.5)ln(R) - 0.007(0.00)R, & R < 300 \\ -0.465(0.008) + 0.056(M_w - 6.5)ln(R) - 0.007(0.0)R - 0.013(0.0)ln(R + 300), & R \geq 300 \end{cases} \tag{10c}$$

After determining the coefficients corresponding to the independent parameters, Monte Carlo method has been used for calculating the standard error in the parameters. The whole procedure is explained above and is given as a schematic diagram in Fig. 3. The residual corresponding to the newly derived equation for intra- and interevent is given in Fig. 11. Further, using Monte Carlo, c_k, τ, and σ have been calculated from residual analysis as explained above. One thousand sets of data have been produced. The variation of c_k, τ, and σ is studied. The values of c_k, τ, and σ determined are 0.02 ± 0.015, 0.52 ± 0.063, and 0.81 ± 0.036, respectively. However, one of the ways that these values can be improved is by adding more simulated data for entire range of magnitude and distance in future works.

8 Conclusion

In the present study, the compatibility of different available GMPE functional forms for active region for magnitude and distance scaling is determined to represent the Himalayan seismotectonic. Methodology used for estimating the suitable functional form has been discussed and applied to the recorded strong motion Himalayan database. Functional forms from five NGA-West 2, five Himalayan, two Pan-European, and one Japan GMPEs are tested for determining the applicability of these in representing the Himalayan attenuation character. For analysis, a mixed effect regression is used on the residuals, which is newly determined using these functional forms. The spatial correlation has been used in mixed effect regression by changing its likelihood function to determine the intra- and interevent residuals. Distance scaling was investigated by examining the trends of intraevent residual with distance, whereas magnitude scaling was examined by studding the trend of the interevent term with magnitude. Further, genetic algorithm and Monte Carlo method is respectively used for determining the hinge points and calculating the standard error in each term that governs the magnitude and distance scaling. The whole methodology is tested for the recorded strong motion database for the Himalayan region. Based on the bias analysis and variation in intra- and interevent residuals, a new functional form for the region is derived. Hinge points for magnitude and distance have significant contribution for determining the magnitude and distance scaling; hence, using genetic algorithm, these are estimated. Using a

Table 4 Ranking of GMPEs based on LLH criteria for peak ground acceleration

S. no.	GMPE	LLH score	Ranking
1	CB14	2.219	6
2	ID14	2.857	12
3	AB14	1.854	3
4	CY14	2.251	8
5	BA14	2.149	5
6	NA09	3.038	14
7	SH09	3.133	16
8	GU10	3.078	15
9	NDMA10	2.248	7
10	AN13	2.797	11
11	AK14	2.321	9
12	BD14	2.672	10
13	ZH16SI	1.783	2
	ZH16SS	2.041	4
	ZH16SM	3.019	13
14	New equation	1.701	1

derived functional form, regression coefficients for a new GMPE have been estimated for PGA for the Himalayan region. Further, Monte Carlo method has been used for calculating the standard error in these parameters and standard deviation corresponding to inter- and intraevent terms. However, the GMPE for the Himalayan region derived in this study is based on recorded database but the database used is not covering full range of magnitude and distance. This can be further improved by incorporating synthetic ground motion database in a future study.

Appendix 1

Different GMPEs mentioned in Tables 1 and 2 are ranked based on the LLH procedure. The recorded data from 2015 Nepal earthquake is used in ranking the GMPEs using LLH. Table 4 provides the ranking of GMPEs along with LLH score for peak ground acceleration.

Appendix 2

The steps used in the mixed effect algorithm are as follows:

I. Estimate the model coefficients, i.e., θ using a fixed-effects regression algorithm assuming random effect term, η equals zero.

II. Using θ, solve the variance of the residuals, σ^2 and τ^2, by maximizing the likelihood function described as follows:

$$
\begin{aligned}
C\left(\varepsilon_{ij}^{(t)}, \varepsilon_{ij'}^{(t)}\right) &= C\left(\varepsilon_{ij} + \eta, \varepsilon_{ij'} + \eta\right) \\
&= \rho\left(d_{jj'}\right)\sigma^2 + \tau^2 \forall i, j, j' \\
C\left(\varepsilon_{ij}^{(t)}, \varepsilon_{i'j'}^{(t)}\right) &= 0 \forall j, j', i \neq i'
\end{aligned}
$$

Parameter $\rho\left(d_{jj'}\right)$ denotes the spatial correlation between intraevent residuals at two sites j and j' as a function of $d_{jj'}$, the separation distance between j and j'. The detailed discussion about determining $\rho\left(d_{jj'}\right)$ is given in Jayaram and Baker (2010).

III. Given θ, σ^2, and τ^2, η_i can be calculated as follows:

$$
\eta_i = \frac{1'_{n_i,1} C_c^{-1} \varepsilon_i^{(t)}}{\frac{1}{\tau^2} + 1'_{n_i,1} C_c^{-1} 1_{n_i,1}}
$$

C_c is defined as conditional covariance metric(s) for the total residual, $\varepsilon_i^{(t)} = \left[\varepsilon_{i1}^{(t)}, \varepsilon_{i2}^{(t)}, \ldots, \varepsilon_{in_i}^{(t)}\right]$, which is the collection of total residuals at all the sites during an earthquake i. $1'_{n_i,1}$ is the transpose of the column metric(s) of ones of length n_i. The above equation is valid only if the interevent residual follows a normal distribution and the intraevent residuals at multiple sites during a given earthquake jointly follow a multivariate normal distribution

IV. Given η_i, estimate the new coefficients (θ) using a fixed effects regression algorithm for $\ln(Y_{ij}) - \eta_i$.

V. Repeat steps 2, 3, and 4 until the likelihood in step 2 is maximized and estimates for the coefficient convergence.

For more detail regarding the algorithm, refer to Jayaram and Baker (2008) and Jayaram and Baker (2010).

References

Abrahamson NA, Youngs RR (1992) A stable algorithm for regression analyses using the random effects model. Bull Seismol Soc Am 82(1):505–510

Abrahamson N, Atkinson G, Boore D, Bozorgnia Y, Campbell KW, Chiou B, Idriss IM, Silva W, Youngs R (2008) Comparisons of the NGA ground-motion relations. Earthquake Spectra 24(1): 45–66

Abrahamson NA, Silva WJ, Kamai R (2014) Summary of the ASK14 ground-motion relation for active crustal regions. Earthquake Spectra 30(3):1025–1055. https://doi.org/10.1193/070913EQS198M

Akkar S, Sandikkaya MA, Bommer JJ (2014) Empirical ground motion models for point and extended-source crustal earthquake scenarios in Europe and the Middle East. Bull Earthq Eng 12:359–387

Ang AHS, Tang WH (2007) Probability concepts in engineering: emphasis on applications in civil and environmental engineering. John Wiley and Sons, Chichester West Sussex 293 pp

Anbazhagan P, Kumar A, Sitharam TG (2013) Ground motion prediction equation considering combined data set of recorded and simulated ground motions. Soil Dyn Earthq Eng 53: 92–108

Anbazhagan P, Sreenivas M, Bajaj K, Moustafa SSR, Al-Arifi NS (2016) Selection of ground motion prediction equation for seismic hazard analysis of Peninsular India. J Earthq Eng. https://doi.org/10.1080/13632469.2015.1104747

Baruah S, Gogoi NK, Erteleva Q, Aptikaev F, Kayal JR (2009) Ground motion parameters of Shillong plateau: one of the most seismically active zones of northeastern India. Earthq Sci 22:283–291

Bilham R, Blume F, Bendick R, Gaur VK (1998) The geodetic constraints on the translation and deformation of India: Implications for future great Himalayan earthquakes. Curr Sci 74:213–229

Bilham R (2015) Raising Kathmandu. Nat Geosci 8:582–584

Bindi D, Massa M, Luzi L, Ameri G, Pacor F, Puglia R, Augliera P (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 12:391–430

Bommer JJ, Douglas J, Scherbaum F, Cotton F, Bungum H, Fäh D (2010) On the selection of ground-motion prediction equations for seismic hazard analysis. Seismol Res Lett 81(5): 783–793. https://doi.org/10.1785/gssrl.81.5.783

Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) NGA-West 2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes. Earthquake Spectra 30(3):1057–1085. https://doi.org/10.1193/070113EQS184M

Bradley BA (2013) A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow

crustal earthquakes based on foreign models. Bull Seismol Soc Am 103:1801–1822

Campbell KW, Bozorgnia Y (2006). Next Generation Attenuation (NGA) empirical ground motion models: can they be used in Europe? in Proc. of the First European Conf. on Earthq. Engineering and Seismol., Paper No. 458

Campbell KW, Bozorgnia Y (2014) NGA-West 2 ground motion model for the average horizontal components of PGA, PGV, and 5%- damped linear acceleration response spectra. Earthquake Spectra 30(3):1087–1115. https://doi.org/10.1193/062913EQS175M

Chiou BSJ, Youngs RR (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. Earthquake Spectra 30(3):1117–1153. https://doi.org/10.1193/072813EQS219M

Cotton F, Scherbaum F, Bommer JJ, Bungum H (2006) Criteria for selecting and adjusting ground-motion models for specific target regions: application to central Europe and rock sites. J Seismol 10(2):137–156. https://doi.org/10.1007/s10950-005-9006-7

Das S, Gupta ID, Gupta VK (2006) A probabilistic seismic hazard analysis of Northeast India. Earthquake Spectra 22:1–27

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

Donahue J, Abrahamson N (2014) Simulation-based hanging wall effects. Earthquake Spectra 30(3):1269–1284. https://doi.org/10.1193/071113EQS200M

Douglas J (2004) An investigation of analysis of variance as a tool for exploring regional differences in strong ground motions. J Seism 8:485–496

Douglas, J. (2015), Ground motion prediction equations 1964–2015, http://www.gmpe.org.uk

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall, New York

Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading, Massachusetts

Gregor N, Abrahamson NA, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, Chiou BSJ, Idriss IM, Kamai R, Seyhan E, Silva W, Stewart JP, Youngs R (2014) Comparison of NGA-West 2 GMPEs. Earthquake Spectra 30:1179–1197

Gupta ID (2010) Response spectral attenuation relations for inslab earthquakes in indo-Burmese subduction zone. Soil Dyn Earthq Eng 30:368–377

Holland JH (1975) Adaptation in natural and artificial systems. The University of Michigan Press, Ann Arbor, Michigan

Idriss IM (2014) An NGA-West 2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. Earthquake Spectra 30(3):1155–1177. https://doi.org/10.1193/070613EQS195M

Jayaram N, Baker JW (2010) Considering spatial correlation in mixed-effects regression and the impact on ground motions models. Bull Seismol Soc Am 100(6):3295–3303

Jayaram N, Baker JW (2008) Statistical tests of the joint distribution of spectral acceleration values. Bull Seismol Soc Am 98(5):2231–2243

Kamai R, Abrahamson NA, Silva WJ (2014) Nonlinear horizontal site amplification for constraining the NGA-West 2 GMPEs.

Earthquake Spectra 30(3):1223–1240. https://doi.org/10.1193/070113EQS187M

Kale O, Akkar S (2013) A new procedure for selecting and ranking ground-motion prediction equation (GMPEs): the Euclidean distance-based ranking (EDR) method. Bull Seismol Soc Am 103(2A):1069–1084

Kayal JR (2008). Microearthquake seismology and seismotectonics of South Asia. Springer

Kramer SL (2003) Geotechnical earthquake engineering. Pearson Education Inc., Singapore

Kumar A, Mittal H, Sachdeva R, Kumar A (2012) Indian strong motion instrumentation network. Seismol Res Lett 83:59–66

Mandal P, Padhy S, Rastogi BK, Satyanarayana VS, Kousalya M, Vijayraghavan R, Srinvasa A (2001) Aftershock activity and frequency dependent low coda Qc in the epicentral region of the 1999 Chamoli earthquake of Mw 6.4. Pure Appl. Geophys 158:1719–1735

Nath SK, Raj A, Thingbaijam KKS, Kumar A (2009) Ground motion synthesis and seismic scenario in Guwahati city, a stochastic approach. Seismol Res Lett 80(2):233–242

Nath SK, Thingbaijam KKS (2011) Peak ground motion predictions in India: an appraisal for rock sites. J Seismol 15:295–315

NDMA (2010) Development of probabilistic seismic hazard map of India. Technical report by National Disaster Management Authority, Government of India, New Delhi

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes. Cambridge University Press, Cambridge

Ruppert D, Sheather SJ, Wand MP (1995) An effective bandwidth selector for local least-squares regression. J Am Stat Assoc 90:1257–1270

Scasserra G, Stewart JP, Bazzurro P, Lanzo G, Mollaioli F (2009) A comparison of NGA ground-motion prediction equations to Italian data. Bull Seismol Soc Am 99:2961–2978

Sharma ML, Bungum H (2006) New strong ground motion spectral acceleration relation for the Himalayan region. In First European conference on earthquake engineering and seismology, p 1459

Sharma ML, Douglas J, Bungum H, Kotadia J (2009) Ground-motion prediction equations based on data from Himalayan and Zagros regions. J Earthq Eng 13:1191–1210

Singh SK, Mohanty WK, Bansal BK, Roonwal GS (2002) Ground motion in Delhi from future large/great earthquakes in the central seismic gap of the Himalayan arc. Bull Seism Soc Am 92:555–569

Spudich P, Rowshandel B, Shahi S, Baker JW, Chiou BSJ (2014) Comparison of NGA-West2 directivity models. Earthq Spectra 30(3):1199–1221. https://doi.org/10.1193/080313EQS222M

Stewart JP, Douglas J, Javanbarg M, Abrahamson NA, Bozorgnia Y, Boore DM, Campbell KW, Delavaud E, Erdik M, Stafford PJ (2015) Selection of ground motion prediction equations for the global earthquake model. Earthquake Spectra 31(1):19–45

Strasser FO, Abrahamson NA, Bommer JJ (2009) Sigma: issues, insights, and challenges. Seismol Res Lett 80:41–56

Wasserman L (2006) All of non-parametric statistics. Springer, New York, 319 pp

Zafarani H, Farhadi A (2017) Testing ground-motion prediction equations against small-to-moderate magnitude data in Iran.

Bull Seismol Soc Am 107:912–933. https://doi.org/10.1785/0120160046

Zhao JX, Jiang F, Shi P, Xing H, Huang H, Hou R, Zhang Y, Yu P, Lan X, Rhoades DA, Somerville PG, Irikura K, Fukushima Y (2016a) Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions. Bull Seism Soc Am 106:1535–1551

Zhao JX, Liang X, Jiang F, Xing H, Zhu M, Hou R, Zhang Y, Lan X, Rhoades DA, Irikura K, Fukushima Y, Somerville PG (2016b) Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions. Bull Seism Soc Am. 106:1518–1534

Zhao JX, Zhou S, Zhou J, Zhou C, Zhang H, Zhang Y, Gao P, Lan X, Rhoades DA, Fukushima Y, Somerville PG, Irikura K (2016c) Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions. Bull Seism Soc. Am. 106:1552–1569