Prediction Equations for Ground-Motion Significant Durations Using the NGA-West2 Database

by Wenqi Du and Gang Wang

Abstract Significant duration is an important parameter in seismic risk assessment. In this article, new prediction equations for significant duration parameters are developed using a recently compiled Next Generation Attenuation-West2 (NGA-West2) database. The use of the greatly expanded NGA-West2 database improves the model predictions for small-to-moderate magnitude and far-source earthquake scenarios. The new model has a functional form with only four predictor variables, namely the moment magnitude ($M_w$), rupture distance ($R_{rup}$), time-averaged shear-wave velocity in the top 30 m ($V_{S30}$), and depth to the top of the rupture ($Z_{rup}$). A magnitude-dependent aleatory variability term is also proposed. The new model can be used to estimate significant durations for earthquake scenarios with moment magnitude $M_w$ from 3 to 7.9 and rupture distance up to 300 km. The proposed model has been systematically compared with some existing prediction models. In addition, empirical correlations between significant durations and spectral accelerations have been studied using the proposed model and the NGA-West2 database.

Introduction

Damage potential of ground motions is usually related to ground-motion amplitude, frequency content, duration, and other cumulative effects. In current seismic-hazard analysis, the duration parameters may not be regarded as equally important as other amplitude and frequency-content parameters, such as peak ground acceleration (PGA) and spectral accelerations (SAs). However, the ground-motion duration, associated with other amplitude parameters, has been regarded as a meaningful indicator in seismic risk assessment, especially in the geotechnical field. For instance, it is well studied that the pore pressure buildup in liquefiable soils is directly related to the number of cycles of the shaking (e.g., Seed and Lee, 1966; Green and Terri, 2005). It is also observed that long-duration ground motions would increase the liquefaction potential of saturated sands (e.g., Idriss and Boulanger, 2006).

The significance of the duration effect on structural response is still controversial. As described by Hancock and Bommer (2006), some researchers found little correlation between duration and peak structural responses (e.g., Iervolino et al., 2006), whereas other researchers observed a positive correlation between duration and cumulative damage measures (e.g., Bommer et al., 2004; Hancock and Bommer 2007). Recently, Chandramohan et al. (2016) found that structures with higher deformation capacities and rapid rates of cyclic deterioration would be more susceptible to damage under long-duration ground motions.

As summarized by Bommer and Martinez-Pereira (1999), there have been more than 30 definitions of ground-motion durations in the literature. Among these definitions, significant duration (termed as $D_s$) is one of the most widely used metrics. $D_s$ can be defined on the basis of the Arias intensity (Arias, 1970), which is given by the following equation:

$$I_A = \frac{\pi}{2g} \int_0^{t_{max}} a(t)^2 dt,$$

in which $I_A$ is the Arias intensity; $t_{max}$ is the total duration of ground-motion time history; and $g$ refers to acceleration of gravity. Significant durations are then evaluated as the time interval over which specified proportions of normalized $I_A$ are accumulated. Two kinds of significant durations are commonly used, namely the time intervals between 5%–75% and 5%–95% of $I_A$ (denoted as $D_{5-75}$ and $D_{5-95}$, respectively). Other time intervals such as 20%–80% of $I_A$ have also been used for some applications. Figure 1 shows an example of the computed significant durations using a ground-motion record from the Next Generation Attenuation-West2 (NGA-West2) database (Ancheta et al., 2014). Apparently, $D_{5-95}$ is always greater than $D_{5-75}$ for a given time history. It is also worth mentioning that a completely different definition of ground-motion duration has been proposed recently, which is based on minimizing the mismatch between the observed response spectrum and the one estimated from the random vibration theory (Bora et al., 2014, 2015).

A number of researchers have proposed prediction equations for different significant duration parameters.
NGA-West2 database (Ancheta et al., 2014). Because the NGA-West2 database consisting of 21,335 ground-motion records from a variety of worldwide earthquakes has been compiled recently, it is necessary to further explore the features of $D_s$ based on the expanded database.

The objective of this article is to develop prediction equations for significant durations ($D_{s, 75}$ and $D_{s, 95}$) using the latest NGA-West2 ground-motion database. Simple functional form employing four predictor variables is proposed based on a mixed-effects regression analysis. Then, the performance of the proposed equations is compared with some existing models. Finally, empirical correlation between the residuals of $D_s$ and SA, which is required to predict the joint distribution of multiple intensity measures (IMs), is also studied using the expanded database.

**Ground-Motion Database**

In this article, a subset of the Pacific Earthquake Engineering Research Center’s (PEER) NGA-West2 database (Ancheta et al., 2014) is selected to develop the empirical models for significant duration ($D_s$) of ground motions. The NGA-West2 database includes a total of over 21,000 three-component uniformly processed recordings, most of which were recorded in free fields. They have been used to develop the latest NGA-West2 ground-motion prediction equations (GMPEs) for shallow crustal earthquakes in active tectonic regions (Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014).

The exclusion criteria introduced by Campbell and Bozorgnia (2014) are adopted herein to select the reliable records. The criteria goal at excluding low-quality, unreliable, incomplete or poorly recorded data, aftershocks,

---

**Table 1**

<table>
<thead>
<tr>
<th>Duration Parameter</th>
<th>Predictors</th>
<th>Regions of Records</th>
<th>Number of Earthquakes Used</th>
<th>Number of Records Used</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{s, 75}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>United States</td>
<td>49</td>
<td>188</td>
<td>Trifunac and Brady (1975)</td>
</tr>
<tr>
<td>$D_{s, 75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>United States</td>
<td>2</td>
<td>30</td>
<td>Abrahamson and Silva (1996)</td>
</tr>
<tr>
<td>$D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>California and Italy</td>
<td>32</td>
<td>272</td>
<td>Hernandez and Cotton (2000)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Mexico</td>
<td>12</td>
<td>800</td>
<td>Reinoso and Ordaz (2001)</td>
</tr>
<tr>
<td>$D_{s, 75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Global</td>
<td>73</td>
<td>1559</td>
<td>Kempton and Stewart (2006, referred to as KS06)</td>
</tr>
<tr>
<td>$D_{s, 75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Global</td>
<td>114</td>
<td>2406</td>
<td>Bommer et al. (2009, referred to as BSA09)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Stable continental</td>
<td>4</td>
<td>28</td>
<td>Lee and Green (2014)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Iran</td>
<td>141</td>
<td>286</td>
<td>Yaghmaei-Sabegh et al. (2014)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$, $D_{s, 95}$</td>
<td>$R_{eq}$, $S$</td>
<td>Japan</td>
<td>1735</td>
<td></td>
<td>Ghofrani and Atkinson (2015)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Global</td>
<td>1195</td>
<td></td>
<td>Afshari and Stewart (2016)</td>
</tr>
<tr>
<td>$D_{s, 5-75}$, $D_{s, 95}$</td>
<td>$M$, $R_{eq}$, $S$</td>
<td>Global</td>
<td>311</td>
<td>13,958</td>
<td>This study</td>
</tr>
</tbody>
</table>

$M$, earthquake magnitude; $M$, moment magnitude; $R_{eq}$, epicentral distance (km); $R_{eq}$, closest distance from site to the rupture plane (km); $S$, indicator of soil types; $S$, fundamental site period (s); $V_{30}$, average shear-wave velocity of the upper 30 m (m/s); $Z_{15}$, depth to the top of rupture (km); $Z_1$, depth from the ground surface to the 1 km/s shear-wave surface (km); $Z_{15}$, depth from the ground surface to the 1.5 km/s shear-wave surface (km).
non-free-field recordings, non-shallow-crustal earthquakes, and
data with only one horizontal component. The detailed criteria
have been listed in Campbell and Bozorgnia (2013). In addition,
visual inspection indicates that some low-amplitude and
long-duration acceleration-time histories seem to be affected
by signal noise. Inclusion of these ground motions may not influ-
ence the predictions of amplitude-based IMs (e.g., NGA-West2
GMPEs), but they do influence the predictions of duration
parameters. Thus, the exclusion criteria used by Afshari and
Stewart (2016), which are based on comparing the durations
estimated from the acceleration and velocity-time histories, are
used to further exclude these noisy, unreasonable recordings
from the database. The final selected database is composed of
13,958 recordings from 311 earthquakes with moment mag-
nitudes $M_w$ from 3.05 to 7.9 and rupture distances (closest
distance from the site to the ruptured area) $R_{rup}$ ranging from 0.1
to 499.54 km. The moment magnitude and rupture distance
distribution of records contained in the database is shown in
Figure 2. The NGA-West2 flatfile containing detailed informa-
tion of these ground motions was also downloaded from the
PEER website (see Data and Resources).

Regression Analysis

A typical ground-motion duration model for the mixed-
effects regression analysis takes the form as

$$\ln(D_{GM})_{ij} = \ln(D_{GM})_{ij} + \eta_i + \epsilon_{ij},$$

in which $\ln(D_{GM})$ and $\ln(D_{GM})_{ij}$ denote the logarithm of
measured and predicted geometric mean of significant dura-
tions ($D_{GM,75}$ or $D_{GM,95}$) from two as-recorded horizontal
components; $i$ and $j$ denote the $i$th recording in the $j$th event,
respectively. $\eta_i$ and $\epsilon_{ij}$ represent the between-event residuals
and within-event residuals, which are assumed to be nor-
mally distributed with zero means and standard deviations $\tau$
and $\phi$, respectively (Abrahamson and Youngs, 1992). The
advantage of the mixed-effects model is that it can separate
the total residuals into two components: between-event and
within-event. Therefore, it has been widely used to develop
GMPEs in earthquake engineering (e.g., Foulser-Piggott and
Stafford, 2012; Du and Wang, 2013; Campbell and Bozorg-
nia, 2014).

It has been studied that many engineering applications
need an estimate of the arbitrary horizontal component of
IMs (Baker and Cornell, 2006). For an arbitrary horizontal
component of the measured $D_s$, the component-to-component
residuals can be computed as

$$\xi_{ijk} = \ln D_{s,jk} - \ln(D_{GM})_{ij},$$

in which $D_{s,jk}$ is the measured $D_s$ for the $k$th ($k = 1, 2$) hori-
zontal component of the $j$th recording and the $i$th event; $\xi_{ijk}$
denotes the component-to-component residuals of the two
horizontal components with an assumed zero mean and a stan-
dard deviation of $\phi_c$. $\phi_c$ can be calculated by the following
equation:

$$\phi_c^2 = \frac{1}{4N} \sum_{i=1}^{N} (\ln D_{s,1} - \ln D_{s,2})^2$$

(Boore, 2005), in which subscripts 1 and 2 denote the two
horizontal components of the measured $D_s$; $l$ is the recording
number index; and $N$ is the total number of recordings in the
database.

Therefore, the total standard deviation ($\sigma$) for the geo-
metric mean component and the total standard deviation ($\sigma_{ARB}$)
for the arbitrary horizontal component of $D_s$ can be
determined as follows:

$$\sigma = \sqrt{\phi^2 + \tau^2}$$

$$\sigma_{ARB} = \sqrt{\phi^2 + \tau^2 + \phi_c^2}.$$ 

The symbols $\phi$, $\tau$, and $\sigma$ are consistent with the notations
suggested by Al Atik et al. (2010).

The mixed-effect regression analysis is conducted using
the nlme package in the statistical programming software R
(Pinheiro et al., 2008). The algorithm is similar to that pro-
posed by Abrahamson and Youngs (1992). The regressed co-
efficients, between-event and within-event residuals, as well
as the values of standard statistical metrics can be obtained
by performing the nlme model in one stage.

Functional Form

To develop an appropriate functional form, the existing
prediction equations for $D_s$ can provide some useful insights.
The significant duration at the source is usually assumed to
be equivalent to the source duration $D_{source}$. For significant
duration at a given site, the effects of traveling paths, site

Figure 2. Distribution of earthquake recordings used in this
study. The color version of this figure is available only in the elec-
tronic edition.
conditions, and other seismological conditions need to be well accounted for. Magnitude scaling of the source duration has been studied based on seismological considerations (Abrahamson and Silva, 1996; Kempton and Stewart, 2006). As suggested by theoretical seismic source models (e.g., Boore, 1983), $D_{source}$ is inversely related to the corner frequency $f_c$, in the Fourier amplitude spectrum of the ground motion. Such corner frequency $f_c$ has been related to the seismic moment (Brune, 1970). Therefore, the source duration $D_{source}$ can be expressed as follows:

$$D_{source} = \frac{1}{f_c} = \frac{1}{4.9 \times 10^6 \times \left(\frac{M_0}{\Delta \sigma}\right)^{1/3}},$$

(7)

in which $\beta$ is the crustal shear-wave velocity at the source (km/s), $\Delta \sigma$ is the stress-drop index (bars), and $M_0$ denotes the seismic moment (dyn-cm). Assuming that $\Delta \sigma$ and $M_0$ are related to the moment magnitude $M_w$ through the following relationships: $\Delta \sigma = \exp(b_1 + b_2(M_w - 6))$ and $M_0 = 10^{1.5M_w + 1.65}$. Bommer et al. (2009) recast equation (7) into an equivalent, but simpler, linear functional form as $\ln D_{source} = c_0 + m_1M_w$, in which $c_0$ and $m_1$ are parameters to be determined by regression analysis. Other functional forms for the magnitude scaling have also been proposed, for example, a power $M_w$ function (Yaghmaei-Sabegh et al., 2014) based on the regression of strong-motion data in Iran.

Although the linear magnitude scaling of the logarithmic source duration is based on physical considerations, the functional form was only applied to moderate-to-large earthquakes in previous studies. For example, the Kempton and Stewart (2006) model is valid for $5 \leq M_w \leq 7.6$ and $R_{rup} \leq 200$ km, whereas the model proposed by Bommer et al. (2009) is applicable for $4.8 \leq M_w \leq 7.9$ and for $R_{rup}$ up to 100 km. Figure 3 shows the distributions of the empirical $D_{s_{75}}$ and $D_{s_{95}}$ calculated from the selected NGA-West2 database, together with linear regression lines showing the trend of the data versus moment magnitude in various $R_{rup}$ bins. The data can be clearly separated into two groups.

A linear increase of significant durations with increasing $M_w$ is evident for moderate-to-large events ($M_w \geq 5.3$) at $R_{rup}$ up to around 100 km. However, the significant durations are almost independent of $M_w$ for relatively small events ($M_w < 5.3$). Observation of the near-source duration data in Figure 3 ($R_{rup} = 0$–10 km) further implies that the source durations for small events ($M_w < 5.3$) are approximately $M_w$ independent. Direct extrapolation of linear magnitude scaling based on the moderate-to-large events would significantly underestimate durations in the small-magnitude range. In addition, the significant durations are not notably affected by moment magnitudes at far distances ($150$ km $\leq R_{rup} \leq 300$ km).

The above observations are also supported by several recent studies on ground-motion durations using the NGA-West2 database (e.g., Boore and Thompson, 2014; Afshari and Stewart, 2016).

There are several possibilities to reconcile the notable difference in the magnitude scaling of durations between small and moderate-to-large earthquake events. To cast the source duration into the theoretical expression shown in equation (7), different magnitude-dependent $\Delta \sigma$ terms are needed for small and moderate-to-large events, as proposed in Afshari and Stewart (2016). However, it is worth mentioning that the $\Delta \sigma$ term is just a stress-drop index, not the true stress drop (e.g., Allmann and Shearer, 2009). One would rather regard it as an unknown parameter that needs to be backcalculated from the empirical data. Another possible explanation is that, the $D_{source} \sim f_c^{-1}$ relationship in equation (7) is mainly based on the observational and theoretical expectations of moderate-to-large ground motions (e.g., Hanks and McGuire, 1981; Boore, 1983). For small-magnitude events, the source duration may not be well correlated with $f_c^{-1}$.

Several functional forms for the magnitude scaling were tried and tested to fit the trend of empirical data. Standard statistical metrics such as the Akaike information criterion, the Bayesian information criterion, and the log-likelihood tests were used to compare the different scaling terms. The results showed that the piecewise linear magnitude-scaling function performed better than the other magnitude scaling forms. Similar statistical tests were also performed for the scaling of other parameters, such as rupture distance, site condition, and fault category. After extensive trials and comparisons, the final functional expression is proposed:

$$\ln(D_s) = c_1 + f_{M,R} + c_6 \times \ln(V_{300}) + c_7 \times Z_{tor},$$

(8)

in which $\ln(D_s)$ denotes the natural log of signification duration ($D_{s_{75}}$ or $D_{s_{95}}$); $V_{300}$ represents the time-averaged shear-wave velocity of the upper 30 m (m/s); $Z_{tor}$ is the depth to the top of the fault rupture (km); $f_{M,R}$ represents the piecewise magnitude- and rupture distance-dependent function as follows:

$$f_{M,R} = \begin{cases} 
  c_2 \times \ln(\sqrt{R_{rup}^2 + h^2}) + c_3 \times \ln\left(\frac{R_{rup}}{150}\right); & \text{for } M_w < 5.3 \\
  c_2 \times \ln(\sqrt{R_{rup}^2 + h^2}) + c_3 \times \ln\left(\frac{R_{rup}}{150}\right) + c_4 \times (M_w - 5.3) \left[1 - \frac{\ln(\sqrt{R_{rup}^2 + h^2})}{\ln(\sqrt{150^2 + h^2})}\right]; & \text{for } 5.3 \leq M_w < 7.5, \\
  c_2 \times \ln(\sqrt{R_{rup}^2 + h^2}) + c_3 \times \ln\left(\frac{R_{rup}}{150}\right) + c_4 \times (M_w - 5.3) \left[1 - \frac{\ln(\sqrt{R_{rup}^2 + h^2})}{\ln(\sqrt{150^2 + h^2})}\right] + c_5 (M_w - 7.5) \times \ln\left(\frac{R_{rup}}{150}\right); & \text{for } M_w \geq 7.5 
\end{cases}$$

(9)
in which $M_w$ refers to the moment magnitude; $R_S$ and $R_F$ are two distance parameters defined as $R_S = \min(R_{rup}, 150)$ and $R_F = \max(R_{rup}, 150)$ in kilometers, in which $R_{rup}$ stands for the rupture distance (km); and $h$ is a fictitious hypocentral depth (km). According to the above definition, $R_S$ equals to $R_{rup}$ when $R_{rup}$ is smaller than 150 km, controlling the short-to-moderate distance scaling. $R_F$ equals to $R_{rup}$ if $R_{rup}$ is greater than 150 km, controlling the long-distance scaling. Inclusion of $R_S$ and $R_F$ in the functional form can better fit empirical data in terms of the distance scaling.

The new model employs a piecewise magnitude-scaling function delineated by $M_w = 5.3$ and 7.5, which is motivated by the empirical data in Figure 3. For small-magnitude events ($M_w < 5.3$), the predictive model is independent of $M_w$. For moderate-to-large magnitude events ($M_w \geq 5.3$), the magnitude scaling is influenced by the rupture distance.

Figure 3. Distributions of significant durations versus moment magnitudes: (a) $D_{S,75}$ and (b) $D_{S,95}$ for various $R_{rup}$ bins. The solid lines are obtained by simple linear regression. The color version of this figure is available only in the electronic edition.
The proposed model would be valid for shallow crustal earthquakes with $M_w$ between 3 and 7.9, $R_{rup}$ ranging from 0 to 300 km, and $V_{S30}$ values in the 80–2100 m/s range. Figure 4 shows the median predicted values of $D_{s5-75}$ and $D_{s5-95}$ with respect to $M_w$, $R_{rup}$, $V_{S30}$, and $Z_{tor}$, respectively. It is clearly shown that both the $D_{s5-75}$ and $D_{s5-95}$ are strongly dependent on the magnitude and distance scalings, whereas the influence of $V_{S30}$ and $Z_{tor}$ is relatively weak. Figure 4a shows that $D_s$ is dependent on $M_w$ only at short-to-moderate distance range ($R_{rup} \leq 150$ km). The significant durations greatly increase with increasing $M_w$ for moderate-to-large events ($M_w \geq 5.5$), which is physically expected because these events are generally associated with large fault dimension and long rupture duration at the sources (Dobry et al., 1978). The notable increase of duration at far distances is mainly caused by the increasing refractions and reflections of body waves over the travel path as well as the arrival of slowly propagating surface waves. Besides, the slight increase of durations at soft soil sites (smaller $V_{S30}$) is mainly due to resonance effects within soil layers. Longer duration would also be expected for smaller $Z_{tor}$ values, possibly due to the fact that the depth of buried ruptures can influence the stress drop $\Delta \sigma$ in the fault rupture process (Kagawa et al., 2004). These observations are generally consistent with previous studies (e.g., Bommer et al., 2009).

Aleatory Variability Model

The between-event and within-event residuals of the regression models for $D_{s5-75}$ and $D_{s5-95}$ are plotted versus $M_w$, $R_{rup}$, $V_{S30}$, and $Z_{tor}$ in Figures 5 and 6, respectively. In these plots, the residuals are obtained by the nlme package in R. The between-event and within-event residuals are then partitioned into 10 nonoverlapping $M_w$ bins with an increment of 0.5, and 10 nonoverlapping $Z_{tor}$ bins of 2 km in size. The within-event residuals are also partitioned into nine bins equally spaced with respect to $\ln(R_{rup})$ and $\ln(V_{S30})$, respectively. The black square symbols indicate the local means of these binned residuals, and their 95% confidence intervals are shown as dashed lines. It can be observed that there is a slightly biased trend for the between-event residuals of $D_{s5-75}$ at the small magnitude range, which is caused by reinforcing the $M_w$-independent functional form for small-magnitude events. No clear biases or trends with respect to

### Table 2

Regression Coefficients of the Proposed Significant Duration Model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$D_{s5-75}$</th>
<th>$D_{s5-95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Value</td>
<td>95% CI</td>
</tr>
<tr>
<td>$c_1$</td>
<td>-0.912</td>
<td>-1.073, -0.750</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.850</td>
<td>0.831, 0.869</td>
</tr>
<tr>
<td>$c_3$</td>
<td>1.142</td>
<td>1.091, 1.192</td>
</tr>
<tr>
<td>$c_4$</td>
<td>1.587</td>
<td>1.506, 1.668</td>
</tr>
<tr>
<td>$c_5$</td>
<td>1.726</td>
<td>0.114, 3.338</td>
</tr>
<tr>
<td>$c_6$</td>
<td>-0.066</td>
<td>-0.089, -0.044</td>
</tr>
<tr>
<td>$c_7$</td>
<td>-0.015</td>
<td>-0.023, -0.007</td>
</tr>
<tr>
<td>$h$</td>
<td>3.296</td>
<td>1.986, 4.606</td>
</tr>
</tbody>
</table>
these predictor variables can be observed in other plots. The few slightly biased points (e.g., $V_{S30}$ as 2000 m/s) are possibly caused by a paucity of data. The distributions of the residuals imply that the selected functional form can provide generally unbiased predictions for the significant durations.

As mentioned previously, the standard deviations of the between-event variability, within-event variability, and component-to-component variability are expressed as $\tau$, $\phi$, and $\phi_c$, respectively. The within-event residuals show a strong magnitude-dependent distribution; the variability of the residuals is notably larger for smaller magnitude ($M_w < 5$), as shown in Figure 7a. The component-to-component residuals were also found to be magnitude dependent (shown in Fig. 7c). Therefore, it would be desirable to empirically develop a magnitude-dependent aleatory variability model. The two kinds of residuals were first partitioned into several overlapping magnitude bins with 0.5-unit $M_w$ width. The standard deviation of within-event residuals in each bin can then be calculated using the maximum-likelihood method adopted in the Abrahamson and Youngs (1992) algorithm, whereas the component-to-component standard deviation of residuals in each bin can be computed via equation (4). Figure 7b and 7d shows the distributions of the computed standard deviations of the within-event and component-to-component residuals within varying magnitude bins, respectively. Thus, the standard deviations $\phi$ and $\phi_c$ can be estimated by the following trilinear magnitude-dependent equations:

\[
\phi = \begin{cases} 
\phi_1 & \text{for } M_w < 3.5 \\
\phi_2 & \text{for } 3.5 \leq M_w < 5 \\
\phi_3 & \text{for } M_w \geq 5 
\end{cases}
\]

\[
\phi_c = \begin{cases} 
\phi_1' & \text{for } M_w < 3.5 \\
\phi_2' & \text{for } 3.5 \leq M_w < 5 \\
\phi_3' & \text{for } M_w \geq 5 
\end{cases}
\]
\[
\begin{align*}
\phi &= \begin{cases} 
\phi_1 & M_w \leq 5 \\
\phi_2 + 2(\phi_1 - \phi_2) \times (5.5 - M_w) & 5 < M_w < 5.5 \\
\phi_2 & M_w \geq 5.5
\end{cases} \\
\phi_c &= \begin{cases} 
\phi_{c1} & M_w \leq 5 \\
\phi_{c2} + 2(\phi_{c1} - \phi_{c2}) \times (5.5 - M_w) & 5 < M_w < 5.5, \\
\phi_{c2} & M_w \geq 5.5
\end{cases}
\end{align*}
\]

in which \(\phi_i\) and \(\phi_{ci}\) \((i = 1, 2)\) are regressed parameters. The proposed trilinear variability models of the within-event and component-to-component residuals for \(D_{S_5} - 95\) are shown (solid lines) in Figure 7b and 7d, respectively. It is seen that the trilinear curves generally fit the empirical points well.

The distribution of the between-event residuals does not show such dependency, and therefore, the constant value \(\tau\) obtained by regression analysis can be directly used for various earthquake scenarios.

Table 3 presents the values of \(\tau\), \(\phi_i\), and \(\phi_{ci}\) \((i = 1, 2)\) obtained from the above analysis. The total standard deviations for geometric mean component (\(\sigma\)) and arbitrary horizontal component (\(\sigma_{ARB}\)) can then be calculated by equations (5) and (6), respectively. The values of \(\sigma\) and \(\sigma_{ARB}\) for specific scenarios are also listed in Table 3.

Figure 8 shows the median and median standard deviation predicted \(D_{S_5} - 75\) and \(D_{S_5} - 95\) for a variety of earthquake scenarios associated with the empirical data in the present database. For a given earthquake scenario with input...
parameters \( [M_w, Z_{tor}, V_{S30}] \), the empirical data are selected based on a combination of magnitude bin \([M_w - 0.25, M_w + 0.25]\), depth to the top of rupture bin \([Z_{tor} - 2 \text{ km}, Z_{tor} + 2 \text{ km}]\), and shear-wave velocity bin \([V_{S30} - 100 \text{ m/s}, V_{S30} + 100 \text{ m/s}]\). It appears that the estimates of \( D_{S,75} \) and \( D_{S,95} \) are generally in agreement with the empirical data, and then the proposed model can broadly capture the scaling of significant durations versus rupture distances.

**Comparison with Other Models**

In this section, the performance of the new model is compared with several previous studies. Three global models from Table 1 are selected (Kempton and Stewart, 2006, hereafter KS06; Bommer et al., 2009, hereafter BSA09; Afshari and Stewart, 2016, hereafter AS16). As mentioned previously, these three models were developed using the NGA-West1 or NGA-West2 strong-motion database. Some predictor variables such as \( M_w, R_{app} \), and \( V_{S30} \) are commonly used in these models. Therefore, they are appropriate to perform direct comparisons with the new model. It should be noted that for the KS06 model, only the base model without the consideration of basin or directivity effect is used herein.

Distance scalings of the median predictions of \( D_{S,75} \) and \( D_{S,95} \) are compared in Figure 9. The predictions are evaluated for a strike-slip earthquake with \( V_{S30} = 400 \text{ m/s} \),
Figure 8. Comparisons of the predicted $M_w$ and $M_w$ against empirical data for five earthquake scenarios: (a) $M_w = 4$, $Z_{cor} = 8$ km, $V_{S30} = 500$ m/s; (b) $M_w = 5$, $Z_{cor} = 8$ km, $V_{S30} = 400$ m/s; (c) $M_w = 6$, $Z_{cor} = 2$ km, $V_{S30} = 250$ m/s; (d) $M_w = 7$, $Z_{cor} = 2$ km, $V_{S30} = 400$ m/s; and (e) $M_w = 7.75$, $Z_{cor} = 0$ km, $V_{S30} = 400$ m/s. The empirical data are selected following a combination of $|M_w - 0.25, M_w + 0.25|$, $|Z_{cor} - 2$ km, $Z_{cor} + 2$ km|, and $|V_{S30} - 100$ m/s, $V_{S30} + 100$ m/s]. The color version of this figure is available only in the electronic edition.
$Z_1 = 0.3\ \text{km},$ and \( Z_{\text{tot}} = 7, 6, 4, \) and \( 0\ \text{km} \) for \( M_w = 4.5, \) 5.5, 6.5, and 7.5, respectively. These input parameters are generally in line with the estimates provided by Kaklamanos et al. (2011). Figure 10 shows the comparisons of magnitude scaling of the median \( D_{s,75} \) and \( D_{s,95} \) for \( R_{\text{rup}} = 10, 30, 100, \) and 200 km, respectively. Other parameters are used as \( V_{s30} = 400\ \text{m/s}, Z_{\text{tot}} = 0 \) km, and \( Z_1 = 0.6\ \text{km}. \)

Figures 9 and 10 show predictions by this study are generally consistent with other models, whereas some discrepancies can be observed at short distances \((<10\ \text{km})\) and far distances \((100-300\ \text{km})\). Durations predicted for small events \((M_w < 5)\) by this model and the AS16 model are systematically higher than the KS06 and BSA09 models. This is not surprising because the database used in KS06 and BSA09 contains a very limited number of small-magnitude recordings. The predictive trends of the KS06 and BSA09 models at the small-magnitude range may not be properly constrained by the database used. It has been reported that the GMPEs based on the NGA-West1 database (e.g., Boore and Atkinson, 2008; Campbell and Bozorgnia, 2008) generally overpredict the PGA values for small-magnitude events (Boore et al., 2014; Campbell and Bozorgnia, 2014). Considering the fact that Ds are negatively correlated with PGA (Bradley, 2011), the existing models using the NGA-West1 database tend to underpredict Ds for small-magnitude earthquakes. Therefore, the proposed model can better estimate significant durations at the small-magnitude range.

Figure 11 compares the total standard deviations \( \sigma \) and \( \sigma_{\text{ARB}} \) from this study with those from the other three models. As can be seen from these plots, although a magnitude-dependent variability model is adopted in this study, the total standard deviations of these models are generally in a similar range. Both this study and the AS16 model indicate that the standard deviations for smaller magnitude events are generally larger than those of larger events.
Empirical Correlation Analysis

Empirical correlations between ground-motion IMs are important in seismic-hazard analysis and the ground-motion selection process. Some researchers have studied the empirical correlations between various IMs (e.g., Baker and Jayaram, 2008; Bradley, 2011; Wang and Du, 2012, 2013; Huang and Wang, 2015). Specifically, Bradley (2011) has developed parametric equations to quantify the correlations between significant durations and other IMs, using the ground motions selected in the NGA-West1 database. It is tempting to further examine the empirical correlations between significant durations and other IMs such as PGA, peak ground velocity (PGV), and SA using the NGA-West2 database.

Similar to equation (2), current GMPEs usually assume that IMs are normally distributed in logarithmic scale, which can be shown as

$$\ln(\text{IM}_k) = \ln(\text{IM}_k^\text{p}) + \eta_k + \epsilon_k,$$

in which \(\ln(\text{IM}_k)\) and \(\ln(\text{IM}_k^\text{p})\) denote the measured and the predicted logarithmic \(k\)th IM, respectively. \(\eta_k\) and \(\epsilon_k\) represent the between-event and within-event residuals for the \(k\)th IM. Statistical tests have been performed to prove the normality of the between-event and within-event residuals associated with SA (Jayaram and Baker, 2008). The correlation coefficients between the between-event and within-event residuals for different IMs can be estimated using the well-known Pearson product-moment correlation estimator

$$\rho_{x_1,x_2} = \frac{\sum_{i=1}^{n}(x_1^{(i)} - \bar{x}_1)(x_2^{(i)} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1^{(i)} - \bar{x}_1)^2 \sum_{j=1}^{n}(x_2^{(j)} - \bar{x}_2)^2}}$$

(Ang and Tang, 2007), in which \(x_1\) and \(x_2\) are random variables; \(n\) is the total number of the considered random variables; \(\bar{x}_1\) and \(\bar{x}_2\) denote the sample mean of variables \(x_1\) and \(x_2\), respectively. Therefore, \(\rho_{\eta_1,\eta_2}\) and \(\rho_{\epsilon_1,\epsilon_2}\) representing the correlations of between-event and within-event residuals between \(\text{IM}_1\) and \(\text{IM}_2\) can be estimated via equation (13). In this study, \(\text{IM}_1\) refers to \(D_s\) and \(\text{IM}_2\) refers to PGA, PGV, and SA at different periods. For demonstration purpose, Figure 12a shows the scatter plot of the within-event residuals for \(D_{s,75}\) and PGA. The corresponding correlation coefficient obtained by equation (13) is \(-0.51\).

Under the assumptions that the between-event and within-event residuals of IMs are independent (Abrahamson and Youngs, 1992), the correlation between the total residuals \(\rho_{\xi_1,\xi_2}\) can be expressed as a combination of the between-event and within-event correlations as

$$\rho_{\xi_1,\xi_2} = \frac{1}{\sigma_1 \times \sigma_2} (\rho_{\eta_1,\eta_2} \rho_{\epsilon_1,\epsilon_2} + \rho_{\epsilon_1,\epsilon_2} \phi_1 \phi_2)$$

Figure 11. Comparisons of the total standard deviations (geometric mean component \(\sigma\) and arbitrary component \(\sigma_{\text{ARB}}\)) from this study with the other predictive models for (a) \(D_{s,75}\) and (b) \(D_{s,95}\). Note that the component-to-component variability \(\phi_1\) was not specified in the Kempton and Stewart (2006; KS06) and Afshari and Stewart (2016; AS16) models. The color version of this figure is available only in the electronic edition.

The normality of the between-event and within-event residuals associated with SA (Jayaram and Baker, 2008).
in which $\tau_k$, $\phi_k$, and $\sigma_k$ ($k = 1, 2$) are the standard deviations of the between-event, within-event, and total residuals for the $k$th IM, respectively. For each pair of IMs, the correlation between the total residuals can be calculated via equations (13) and (14).

To compute the $D_s$–SA correlations, the predicted median SA and standard deviations at each period are computed using the Campbell–Bozorgnia NGA-West2 model (Campbell and Bozorgnia, 2014, hereafter CB14). It should be noted that each recorded time history has a usable frequency range which is related to removing low- or high-frequency noises in signal processing. Thus, only SA for periods less than the inverse of the lowest usable frequency is used to compute the $D_s$–SA correlation coefficients. The number of usable records is then expected to decrease as the vibration period increases, as shown in Figure 12b.

Once the predicted median SAs of the CB14 model for each recording are computed, the total SA residuals can be calculated. Based on the provided $\tau_{SA}$, $\phi_{SA}$, and $\sigma_{SA}$ (between-event, within-event, and total aleatory standard deviations) of the CB14 model, the total SA residuals can be partitioned into between-event and within-event residuals by performing the random-effects regression (Abrahamson and Youngs, 1992). Then, the correlations between the between-event and within-event residuals for $D_s$ and SA can be computed. Finally, equation (14) can be used to calculate the correlations between the total residuals of $D_s$ and SA.

Figure 13 shows the computed empirical correlations between significant durations and SA at various periods: (a) $D_{s_{5-75}}$–SA and (b) $D_{s_{5-95}}$–SA. The parametric equations proposed by Bradley (2011) are also plotted for comparison. The color version of this figure is available only in the electronic edition.
Table 4

<table>
<thead>
<tr>
<th>Ds,5.75</th>
<th>Ds,9.05</th>
<th>PGA</th>
<th>PGV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ds,5.75</td>
<td>-0.47</td>
<td>-0.44</td>
<td>-0.39</td>
</tr>
<tr>
<td>Ds,9.05</td>
<td>-0.48</td>
<td>-0.41</td>
<td>-0.35</td>
</tr>
</tbody>
</table>

Conclusions

This study presented new prediction equations for significant durations \( D_{5,75} \) and \( D_{5,05} \) using 13,958 well-recorded ground motions selected from the latest NGA-West2 database. The equations are expressed as a function of moment magnitude \( (M_w) \), rupture distance \( (R_{rup}) \), site parameter \( (V_{530}) \), and depth to the top of the fault rupture \( (Z_{top}) \). The proposed model can be applied to shallow crustal earthquakes with moment magnitude ranging from 3 to 7.9, and rupture distance less than 300 km.

It was observed that the variability of the within-event residuals for \( M_w < 5 \) is larger than that for \( M_w > 5.5 \). Similar to the NGA-West2 GMPEs, a magnitude-dependent within-event standard deviation structure is proposed. Compared with a constant standard deviation model, the proposed trilinear model can better predict the aleatory variability of various earthquake scenarios.

The performance of the new model has been compared with three other predictive models that were developed based on the NGA database. The predictions of significant durations for earthquakes with moderate-to-large magnitude \( (M_w \geq 5) \) and \( R_{rup} \leq 100 \) km are generally consistent. However, compared with the models developed based on the NGA-West1 database, noticeable discrepancies can be observed for scenarios with small magnitudes \( (M_w < 5.5) \). This is expected, because the NGA-West2 database is abundant in small-to-moderate events. Therefore, the new model can reasonably improve the estimates of \( D_s \) for small-magnitude or far-source earthquake scenarios, owing to the use of the expanded database. We do not think that the new model is definitely superior to the other models, but rather provides an interpretation of the latest database.

Finally, the empirical correlations between \( D_s \) and other IMs such as PGA and SA were investigated. It was found that the empirical correlations using the NGA-West2 database are generally consistent with previous studies using the NGA-West1 database. Therefore, the empirical correlations are relatively insensitive to the selection of ground-motion database. We recommend the parametric \( D_s–SA \) correlation models proposed by Bradley (2011) for practical applications.

Data and Resources

Ground-motion time histories used in this study are obtained from the Pacific Earthquake Engineering Research (PEER) Next Generation Attenuation (NGA) database (http://ngawest2.berkeley.edu/), last accessed August 2015. The NGA-West2 database flatfile including source, site information as well as rupture distances are downloaded via http://peer.berkeley.edu/ngawest2/databases/ (last accessed August 2015).

Acknowledgments

The work described in this article was supported by Hong Kong Research Grants Council through General Research Fund Grant Number 162/3615 and Collaborative Research Fund Grant Number CityU9/CRF/13G. The authors thank the Pacific Earthquake Engineering Research (PEER) researchers who compiled the Next Generation Attenuation-West2 (NGA-West2) ground-motion database and make it available to the public. The authors thank two anonymous reviewers and Associate Editor Peter Stafford for their helpful comments to improve this article.

References


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 0.1 to 10 s, Earthq. Spectra 24, no. 1, 139–171.

Campbell, K. W., and Y. Bozorgnia (2013). NGA-West2 Campbell–Bozorgnia ground motion model for the horizontal components of PGA, PGV, and 5%-damped elastic pseudo-acceleration response spectra for periods ranging from 0.01 to 10 s, PEER Report No. 2013/06, Pacific Earthquake Engineering Research Center, University of California, Berkeley, California, 238 pp.


Institute of Catastrophe Risk Management
Nanyang Technological University
50 Nanyang Avenue, North Spine, Block N1-B1b-08
Singapore 639798, Singapore
wqdu@ntu.edu.sg
(W.D.)

Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology
Room 3594, Academic Building
Clear Water Bay, Hong Kong
wgbwq@ust.hk
(G.W.)

Manuscript received 3 December 2015;
Published Online 27 December 2016