Earthquake Engineering & Structural Dynamics
The Journal of the International Association for Earthquake Engineering

EARTHQUAKE ENGINEERING & STRUCTURAL DYNAMICS
Earthquake Engng Struct. Dyn. 2013; **42**:1189–1202
Published online 25 October 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/eqe.2266

A simple ground-motion prediction model for cumulative absolute velocity and model validation

Wenqi Du and Gang Wang*,†

Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong, China

SUMMARY

Cumulative absolute velocity (CAV) is an important ground motion intensity measure used in seismic hazard analysis. Based on the Next Generation Attenuation strong motion database, a simple ground-motion prediction equation is proposed for the geometric mean of as-recorded horizontal components of CAVs using mixed regression analysis. The proposed model employs only four parameters and has a simple functional form. Validation tests are conducted to compare the proposed model with the recently developed Campbell–Bozorgnia (CB10) model using subsets of the strong motion database, as well as several recent earthquakes that are not used in developing the model. It is found that the predictive capability of the proposed model is comparable with the CB10 model, which employs a complex functional form and more parameters. The study also corroborates previous findings that CAV has higher predictability than other intensity measures such as the peak ground acceleration. The high predictability of CAV warrants the use of the proposed simple model as an alternative in seismic hazard analysis. Copyright © 2012 John Wiley & Sons, Ltd.

Received 9 May 2012; Revised 21 September 2012; Accepted 1 October 2012

KEY WORDS: cumulative absolute velocity; ground motion prediction; model validation; blind test

1. INTRODUCTION

Ground motion intensity measures (IMs) are often used to represent different characteristics of the earthquake ground motions. The recently developed Next Generation Attenuation (NGA) models [1] have significantly advanced the state-of-the-art of ground-motion prediction equations (GMPEs) for spectral accelerations, which are often used as key IMs for design purposes. However, earthquake waves are complex transient time sequences. The acceleration-response spectra alone are not sufficient to characterize all important aspects of the earthquake loading. It has been found that the seismic performance of some types of structures are more closely related to the time integration of acceleration time histories, including Arias Intensity [2], cumulative absolute velocity (CAV) and its variants [3, 4], and the time rate of Arias Intensity [5]. There is a clear need to develop a new generation GMPEs to accurately characterize these IMs for engineering practice.

In this study, we focus on developing a simple ground-motion prediction model for cumulative absolute velocity (CAV). CAV is defined as the time integration of the absolute values of the acceleration as follows:

$$CAV = \int_0^{t_{tot}} |a(t)| \mathrm{d}t \tag{1}$$

where $|a(t)|$ is the absolute value of the acceleration time history, and t_{tot} is the total duration of the ground motion time history. Because small-amplitude accelerations usually do not contribute to

*Correspondence to: Gang Wang, Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.
†E-mail: gwang@ust.hk

Copyright © 2012 John Wiley & Sons, Ltd.

structural damages, several variants of CAV were also proposed to exclude small-amplitude accelerations into the time integration in Equation (1). For example, EERI (1991) proposed a standardized CAV measure, termed as CAV_{STD}, which only integrates the acceleration whose peak value exceeds a threshold value of 0.025 g within a 1-s time interval [4]. Another useful CAV variant, CAV_5, is defined to exclude acceleration values less than a threshold value of 5 cm/s [6]. These CAV variants exclude undamped high frequency component of the time history, and they appear to better represent long period characteristics of the ground motion. However, exclusion of low-amplitude acceleration from contributing to CAV variants causes relative small number of nonzero values, leading to a decreased stability in predicting these intensity measures [7]. Therefore, in this study, we will focus on the GMPE for CAV defined by Equation (1) rather than other versions.

Among a few available GMPEs for CAV, Danciu and Tselentis [8] proposed a GMPE for CAV using earthquake data in Greece. More recently, Campbell and Bozorgnia [7] provided a GMPE for CAV (defined by Equation (1)) using the NGA strong motion database (http://peer.berkeley.edu/nga/) and the same functional form as their NGA model [9], which requires a total of eight input parameters to characterize the rupture sources, travel paths and local site conditions. The model is termed as CB10 in this study. Considering that only a very limited number of CAV prediction models are available, this study will develop a GMPE for CAV on the basis of the NGA database and a simple functional form which requires only four input parameters. Validation tests will be conducted to systematically evaluate the performance of the proposed model and CB10 model. The test results demonstrate the efficiency of the proposed model and its capacity to be used as an alternative model if such simplicity is desired in engineering practice.

2. EMPIRICAL GROUND-MOTION PREDICTION MODEL FOR CAV

2.1. Strong motion database

The NGA strong motion dataset used in this study is almost identical to that used in CB10 model, except for an exclusion of 171 records, which will be explained as follows. The database includes a total of 1390 records from 62 earthquakes with moment magnitudes ranging from 4.26 to 7.9 and rupture distances ranging from 0.07 to 200 km. These ground motion data was carefully compiled to exclude aftershock records, non-free field conditions, low quality data, non-shallow-crustal events and data with missing horizontal component. In this study, the local site conditions are characterized using site classification scheme proposed by Bray and Rodriguez-Marek (termed as SGS) [10]. The SGS site classification system includes both a measure of soil stiffness and depth such that site category B is for unweathered soft rocks or shallow soils, site category C is for weathered soft rocks or shallow stiff soils, and site category D is for deep soils. Records of site categories A (hard rock) and E (soft clay) are excluded from the study because of the scarcity of the data (171 in total). Therefore, the dataset (1390 records) is only different from that used in CB10 model by limiting records to SGS site categories B (197 records), C (393 records) and D (800 records). The data distribution with respect to moment magnitude, rupture distance and site conditions is illustrated in Figure 1.

2.2. The predictive equation

Two horizontal components of as-recorded acceleration time histories are available for each record in the strong motion database. The geometric mean of CAVs obtained from these two horizontal components, termed as CAV_{GM}, are calculated by the following equation:

$$CAV_{GM} = \sqrt{CAV_1 \cdot CAV_2} \tag{2}$$

where CAV_1 and CAV_2 denote CAV values of two horizontal components of records using Equation (1). Regression analysis was performed using a random effects model [11, 12] as follows:

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

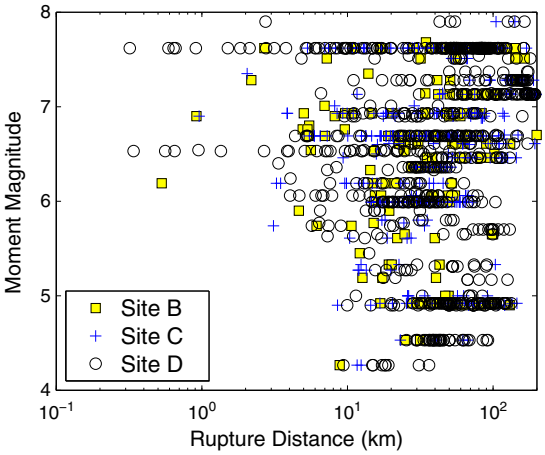


Figure 1. Moment magnitude and rupture distance distribution of records in the database.

$$\ln(CAV_{\mathrm{GM}})_{ij} = \overline{\ln(CAV_{\mathrm{GM}})_{ij}} + \eta_i + \varepsilon_{ij} \tag{3}$$

where $\ln(CAV_{\mathrm{GM}})_{ij}$ and $\overline{\ln(CAV_{\mathrm{GM}})_{ij}}$ represent the observed and the predicted logarithmic CAV_{GM} value for the jth recording and i-th event, respectively. The model divides the residual term into inter-event residual η_i and intra-event terms ε_{ij}. It is assumed that logarithmic CAV_{GM} values follow normal distribution. Therefore, the residual terms η_i and ε_{ij} are normally distributed with a mean of zero and standard deviations of τ and σ, respectively. The normality assumption is usually employed in developing GMPEs, and this assumption will be formally tested in Sections 2.3 and 3.1. Moreover, inter-event and intra-event residuals are often statistically independent variables. Therefore, the standard deviation of the total residuals $(\eta_i + \varepsilon_{ij})$ is then determined by $\sigma_T = \sqrt{\tau^2 + \sigma^2}$.

After extensive examination of many possibilities, the following functional form is used to predict the geometric means of CAVs from two as-recorded horizontal components:

$$\overline{\ln(CAV_{\mathrm{GM}})} = c_1 + c_2(8.5 - M_w)^2 + (c_3 + c_4 M_w)\ln\sqrt{{R_{\mathrm{rup}}}^2 + h^2} + c_5 S_C + c_6 S_D + c_7 F_N + c_8 F_R \tag{4}$$

where $\overline{CAV_{\mathrm{GM}}}$ is the predicted value of CAV_{GM} in the unit of g·s with g being the gravitational acceleration (9.81 m/s^2), M_w is the moment magnitude of the earthquake, the rupture distance R_{rup} measures the closest distance from the recording station to the ruptured area in km, h is a fictitious hypocentral depth in km, S_C and S_D are indicator variables representing various site conditions (e.g., $S_C = 1$ for SGS site category C and $S_C = 0$ otherwise; $S_D = 1$ for SGS site category D and $S_D = 0$ otherwise), F_N and F_R are indicator variables representing types of faulting (e.g., $F_N = 1$ for normal faulting and $F_N = 0$ otherwise; $F_R = 1$ for reserve faulting; $F_R = 0.5$ for reverse-oblique faulting and $F_R = 0$ otherwise). The number of records for normal, strike-slip, reverse-oblique and reverse faulting are 107, 526, 510 and 247, respectively. For the case of a strike-slip faulting and SGS site class B, all indicator variables should be zero. It is worth mentioning that Equation (4) is different from the simple functional form proposed for Arias Intensity [13], which includes additional terms to account for the magnitude effects for different site classes. However, statistical analysis shows that these terms are not statistically significant for the CAV model, hence, they are not included in the formulation. In addition, no magnitude dependence of fault mechanism was observed. Instead, magnitude dependence of intra-event residuals at the short distance ranges ($R_{\mathrm{rup}} < 10$ km) is accounted for by using the term associated with c_4.

The simplified model has a total of nine fitting parameters ($c_1 \sim c_8$ and h) that can be obtained by a nonlinear regression analysis such as that proposed by Joyner and Boore [14] or 'nlme' package in the statistical software 'R' [15]. The obtained coefficients are listed in Table I along with

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

Table I. Coefficients of DW12 model.

Parameter	c_1	c_2	c_3	c_4	c_5	h	c_6	c_7	c_8
Value	1.826	−0.130	−1.403	0.098	0.286	8.455	0.481	−0.155	0.095
Standard deviation of coefficients	0.157	0.017	0.134	0.019	0.037	1.166	0.033	0.116	0.108
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.188

associated standard deviations and p-values. All coefficients except c_7 and c_8 yield very small p-values; therefore, they are statistically significant. Although coefficients c_7 and c_8 have slightly higher p-values, these terms are retained because of their obvious physical interpretations to represent the styles of faulting. The inter-event and intra-event standard deviation τ and σ are slightly larger compared with the results of CB10 model [7], which employed a much more sophisticated functional form with more model parameters. A more detailed model comparison of the CB10 model with the proposed model, termed as DW12 model, will be presented in Section 3.

Figure 2(a) shows the median predicted CAV_{GM} against rupture distances for a strike-slip event with three site classes and three earthquake magnitudes. The median predicted CAV_{GM} values for site class D (deep soil) are constantly higher than these for site class C (weathered rock or shallow soil) and site class B (competent rock), manifesting the effects of site amplification on the increase of accumulative energy in seismic waves. The relative difference between sites B, C and D is not affected by the earthquake magnitude because the site class term is not coupled with the earthquake magnitude as was discussed before. Figure 2(b) displays the median predicted CAV_{GM} values against distance for different types of faulting. As is expected, reverse faulting produces a larger median CAV_{GM} than strike-slip or normal faulting.

2.3. *Inter-event and intra-event residuals*

The distributions of inter-event and intra-event residuals with respect to earthquake magnitudes and rupture distances are shown in Figure 3, and the intra-event residuals for three site classes are also illustrated in Figure 4. From these plots, no obvious biases between residuals and variables used in the regression equation can be found. Although not shown in the paper, the residuals are also not biased for different types of faulting. It is also noted that V_{S30} (the averaged shear wave velocity of the upper 30 m) is often used as a continuous variable in modern GMPEs to characterize the local site condition. However, statistical analysis shows that using V_{s30} rather than the SGS site classification scheme renders little improvement in the model efficiency in this study. Considering that V_{S30} contributes to the greatest portion of aleatory variability of the NGA ground-motion models [16] and only 30% of the stations in the NGA database have measured

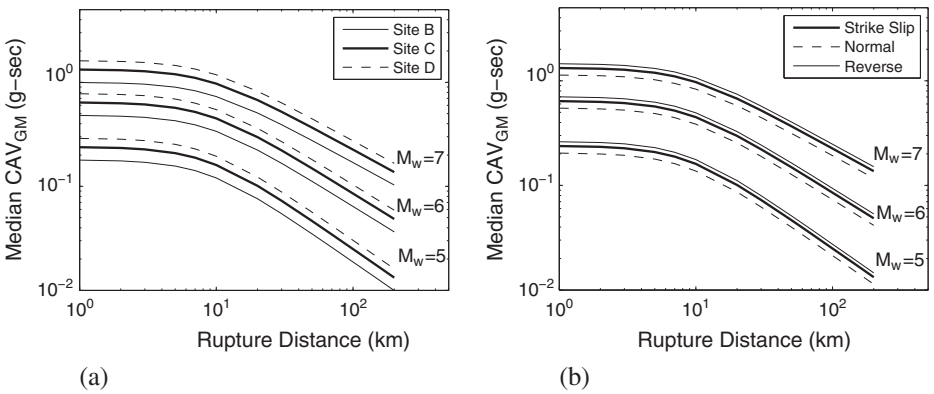


Figure 2. Median predicted values of CAV_{GM} with respect to rupture distance. (a) Three site classes, strike slip fault; (b) three fault types, site class D.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

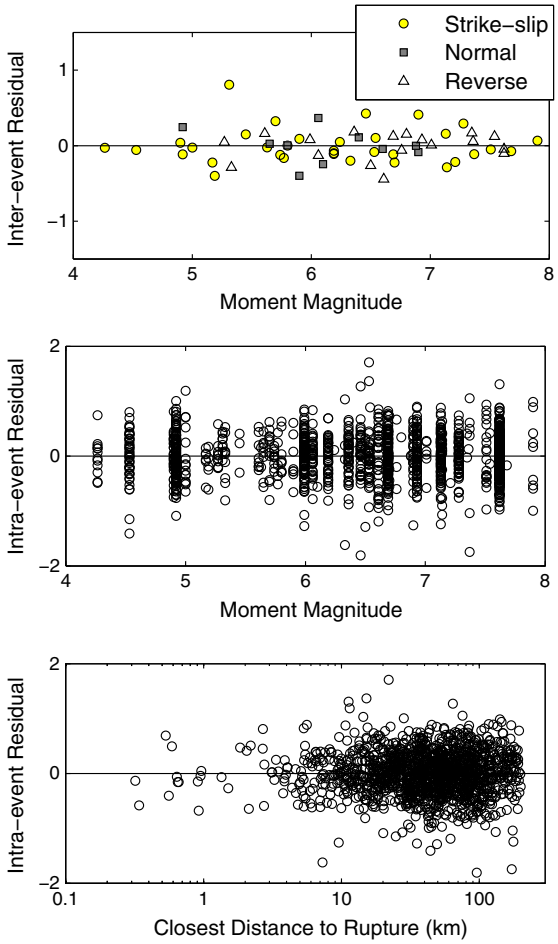


Figure 3. Distributions of inter-event and intra-event residuals with respect to moment magnitude and rupture distance.

V_{S30} values [1], we prefer to use the SGS site class as a discrete indicator for the local site conditions in the regression analysis. Consequently, we checked that the present model does not induce significant bias in terms of the intra-event residuals versus V_{s30} values. As is shown in Figure 4, the trend line is only slightly biased for the majority of the data. Considerable bias occurs only when V_{s30} values are larger than 1000 m/s, where the data is controlled by only a few records. For practical purpose, using the site categories in the present model is still reasonable and acceptable. It should also be noted that for site conditions that cannot be clearly classified using the site categories, a more complicated predictive model with a V_{s30} as an index for site conditions is preferred.

Figure 5(a) shows the intra-event residuals for different site classes versus the predicted CAV values. Because of the effect of nonlinear ground response, the standard deviation of the intra-event residuals decreases as CAV increases, especially for site classes C and D. Similarly, CB10 model shows the standard deviation of intra-event residuals in general decreases with an increasing peak ground acceleration (PGA) on rock sites [7]. Figure 5(b) shows the distribution of the standard deviation of intra-event residuals within varying CAV bins for site class C. The CAV bins are partitioned into overlapping intervals in logarithmic scale, where the point denotes the median value of each bin, and the horizontal bar indicates the range of the interval. It is designed that each bin has the same number of data, that is, 40, 70 and 100 records in each bin for site classes B, C and D, respectively. For site class B, a constant standard deviation $\sigma = 0.416$ is assigned because of the scarcity of data. For site classes C and D, a simple trilinear relationship is obtained to represent the trend of the standard deviations of the intra-event residuals:

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

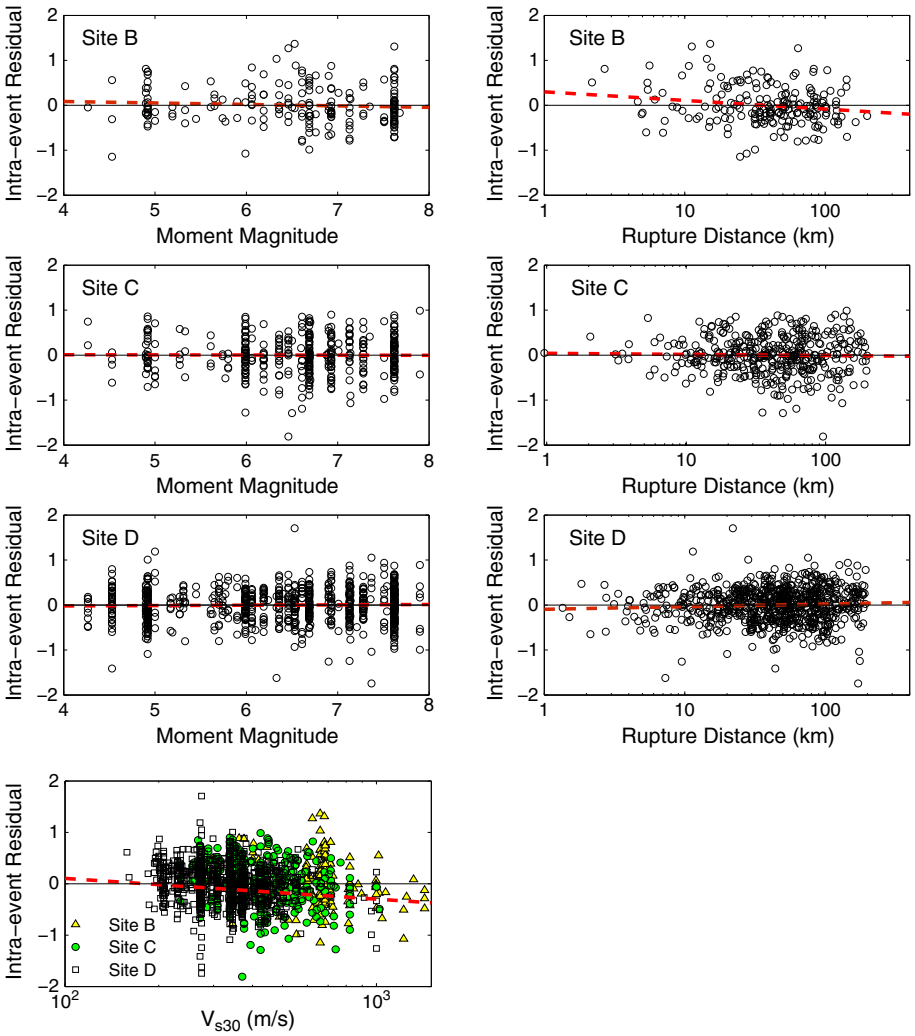


Figure 4. Distributions of intra-event residuals with respect to moment magnitude and rupture distance for three site classes.

$$\sigma = \begin{cases} a & \text{if } CAV \leqslant 0.15\,g{\cdot}s \\ a - c\cdot\ln(CAV/0.02) & \text{if } 0.15 < CAV \leqslant 1\,g{\cdot}s \\ b & \text{if } CAV \geqslant 1\,g{\cdot}s \end{cases} \tag{5}$$

where $a = 0.45$, $b = 0.37$, $c = 0.042$ for site C; $a = 0.38$, $b = 0.34$, $c = 0.021$ for site D; CAV represents the median predicted CAV_{GM} value in the unit of $g{\cdot}s$. It is noted that the standard deviations for site class D are smaller than those of site class C, and the decrease in the intra-event residuals with increasing CAV value reflects the effect of nonlinear ground response. Because of very limited number of earthquake events (62 in this study) used in the regression analysis, a constant standard deviation of inter-event residuals is recommended for all cases, that is, $\tau = 0.247$. It is noted that a slightly smaller value ($\tau = 0.196$) is proposed in CB10 model.

In the previously mentioned statistical analysis, it is assumed that the logarithmic CAV_{GM} values follow a normal distribution. Although it is a common assumption used in GMPEs, the univariate normality of $\ln CAV_{GM}$ can be formally tested using normal Q–Q plot. Q–Q plot is a method that compares two statistical distributions graphically by plotting the quantiles of two distributions against each other [17]. If the dataset follows normal distribution precisely, the data quantiles and theoretical quantiles in Q–Q plot should lie on a 45° line. The Q–Q plot test is conducted for both inter-event and intra-event residuals obtained from the present CAV model. The intra-event residual is normalized by its standard deviation σ

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

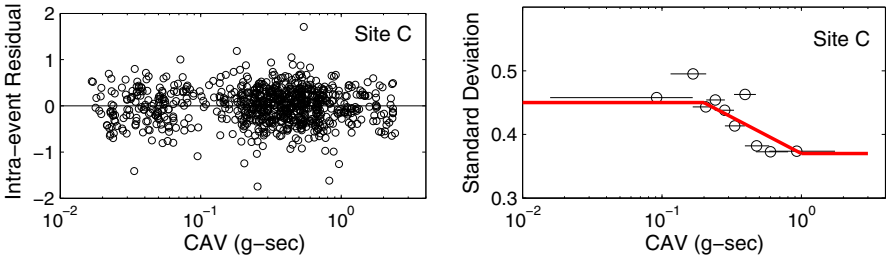


Figure 5. (a) Intra-event residual versus predicted CAV for site C; (b) trilinear relationship of intra-event standard deviations for site C.

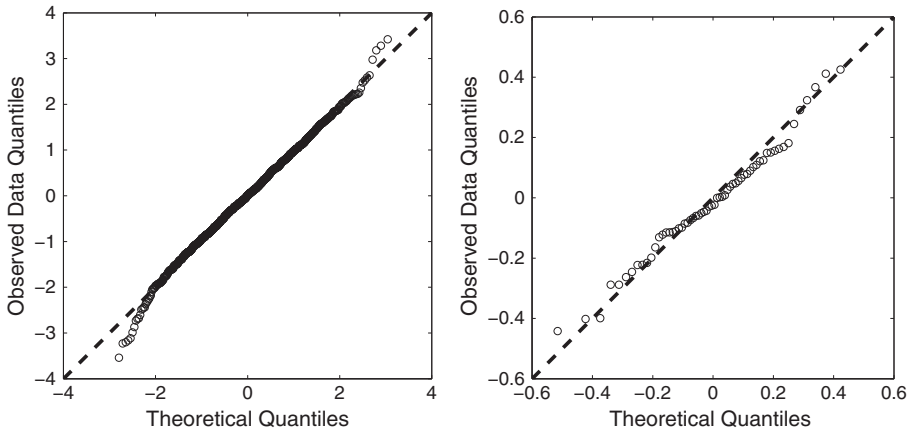
in order to compare it with the standard normal distribution $N(0,1)$. Figure 6(a) shows the Q–Q plot of the normalized intra-event residuals against the theoretical quantiles of $N(0,1)$ approximately lie on a 45° line. The normalized sample distribution is deviated from $N(0,1)$ only when the data quantile is larger than 2.5 or smaller than −2.0, that is, more than 95% of data follows the assumed theoretical distribution. Therefore, the normalized intra-event residuals can be well represented by a standard normal distribution.

For inter-event residuals, the standard deviation is independent of site conditions. Therefore, it is not necessary to normalize the inter-event residuals because of the homoscedasticity (as is also assumed in [18]). Figure 6(b) displays the normal Q–Q plot of inter-event residuals against the corresponding theoretical quantiles, where the data quantiles are directly computed using inter-event residuals and theoretical quantiles are obtained using a normal distribution with a zero mean value and the sample standard deviation. Almost all data points lie on the 45° line in the Q–Q plot, indicating that the normality of inter-event residuals can also be validated.

3. COMPARISON WITH CB10 MODEL

In this section, we systematically compare the performance of DW12 model with CB10 model. First, the input model parameters of two models are summarized in Table II. DW12 model only requires four parameters, whereas CB10 model employs a total of eight parameters, including the moment magnitude of the earthquake (M_w), types of faulting, dip angle of fault plane, depth-to-top-of-rupture (Z_{TOR}), rupture distance (R_{rup}), horizontal distance to the surface projection of rupture (Joyner–Boore distance R_{jb}), average shear-wave velocity in the top 30 m (V_{s30}), and depth to the strata where shear wave velocity $Vs = 2.5$ km/s ($Z_{2.5}$).

The predicted median CAV values from two models are schematically compared in Figure 7 for several cases of hypothetical fault ruptures. The source-to-site distance measures for CB10 model are computed by assuming that the width of the rupture plane is 10 km. The dip angle of the rupture plane is assumed



(a) Normalized Intra-event Residuals (b) Inter-event Residuals

Figure 6. The normal Q–Q plots of intra-event/inter-event residuals.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

Table II. Comparison of model parameters.

	Model parameter	CB10	DW12
Sources	Moment magnitude	M_w	M_w
	Types of faulting	NM/RV/SS	NM/RV/SS
	Dip angle of the rupture plane (°)	δ	
	Depth-to-top-of-rupture (km)	Z_{TOR}	
Travel paths	Closest distance to the rupture plane (km)	R_{rup}	R_{rup}
	Joyner-Boore distance (km)	R_{jb}	
Local site conditions	Average shear-wave velocity in top 30 m (m/s)	V_{s30}	
	Depth to $V_s = 2.5$ km/s horizon (km)	$Z_{2.5}$	
	Site category		SGS site category *

*Description of SGS site category [10]:
B: rock, most 'unweathered' California rock cases (Vs>=760 m/s or <6m of soil);
C: weathered soft rock and shallow stiff soil (<60 m of soil);
D: deep stiff Holocene or Pleistocene soil (>60m of soil and no 'soft' soils).

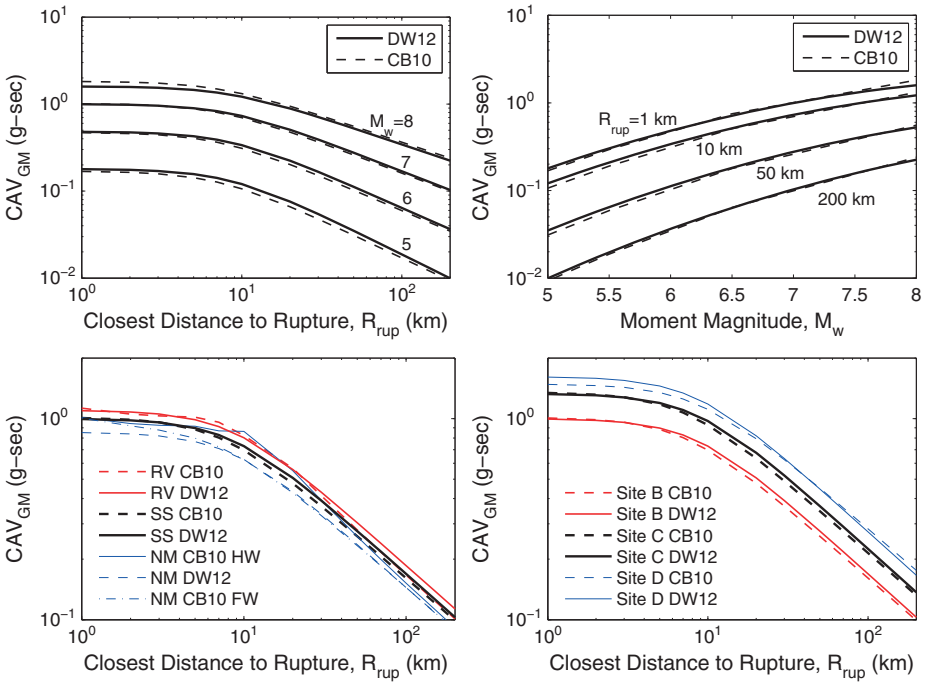


Figure 7. Comparison of median predicted values of CAV_{GM} by DW12 (this study) and CB10 models. (SS: strike-slip faulting, NM: normal faulting, RV: reverse faulting, HW: hanging wall, FW: foot wall. Unless otherwise noted, CAV_{GM} is evaluated for $M_w = 7$, vertical strike-slip faulting, $V_{S30} = 760$ m/s, $Z_{TOR} = 0$ km, and $Z_{2.5} = 2$ km. For normal faulting and reverse faulting, the dip angle is assumed to be 45° and 60°, respectively).

to be 45° for normal faulting (NM) and 60° for reverse faulting (RV). To compare with SGS site classes B, C and D, the V_{s30} values for CB10 model are assumed to be 760, 480 and 300 m/s, respectively. If not indicated otherwise, the default parameters are $M_w = 7$, strike-slip fault (SS), $R_{rup} = 10$ km and rock sites ($V_{s30} = 760$ m/s or SGS site class B). In addition, $Z_{TOR} = 0$ km and $Z_{2.5} = 2$ km are used throughout in CB10 model. It is evident from Figure 7 that two models share a high degree of agreement in terms of distance and magnitude scaling. Relatively large discrepancies can only be observed for normal faulting at the close distance range. Whereas CB10 model is able to differentiate the hanging wall/foot wall effects, and the model predictions are slightly higher than DW12 model.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

3.1. Goodness-of-fit measures

Some goodness-of-fit statistics are further utilized to provide quantitative assessment of the efficiency and accuracy of two models. Firstly we utilize the Nash–Sutcliffe model efficiency coefficient (*EC*) [19], defined in logarithmic space as follows:

$$EC = 1 - \frac{\sum_{i=1}^{n} \left(\ln Y_i - \ln \hat{Y}_i \right)^2}{\sum_{i=1}^{n} \left(\ln Y_i - \overline{\ln Y} \right)^2} \tag{6}$$

where Y_i denotes the observed CAV values, $\hat{Y}_i$ is the predicted values for the ith observation, $\overline{\ln Y}$ is the mean value of the logarithms of observed values, and n is the total number of observations. *EC* ranges between $-\infty$ to 1. When *EC* equals zero, the model efficiency is equivalent to simply using the mean of the data. It is evident that a larger *EC* implies a better agreement between observations and predictions. Compared with other widely used goodness-of-fit statistics (e.g., correlation coefficient), *EC* has the advantage to emphasize the proportional differences between the model predictions and observations, and it is less sensitive to outliers [20].

In developing the GMPE, we assumed that the logarithmic *CAV* follows normal distribution. The previous section tested the normality of logarithmic *CAV* in terms of inter-event and intra-event residuals. Furthermore, we define a metric *LH* from model residuals [21] as follows:

$$LH(z_0) = 1 - Erf\left(\frac{|z_0|}{\sqrt{2}}\right) \tag{7}$$

where *Erf* is error function given by $Erf(z) = 2/\sqrt{\pi} \int_0^z e^{-t^2} \mathrm{d}t$, and z_0 is normalized model residual $z_0 = \frac{\ln Y - \ln \hat{Y}}{\sigma_T}$. If $\ln Y$ follows normal distribution exactly, the samples of *LH* are uniformly distributed in (0, 1) [21]. Thus, the median value of *LH* equals 0.5 if the normality assumption is satisfied exactly. Therefore, it can be used to quantify the distribution of residuals such that the smaller deviation of median *LH* from 0.5, the better normality assumption is satisfied.

The observed CAV values versus predicted CAV values are shown in Figure 8 for CB10 and DW12 models. Similarity in the scattered data distribution of these two models indicates a very similar degree of predictive power for overall data. *EC* is 0.84 for DW12 model and is only slightly lower than that for CB10 model (*EC* = 0.86). Furthermore, four goodness-of-fit measures are listed in Table III for two models using the ranking scheme proposed by [21], which rates both models as rank A—the highest capability class. For a model to be qualified as rank A, a median LH (MEDLH) value should be at

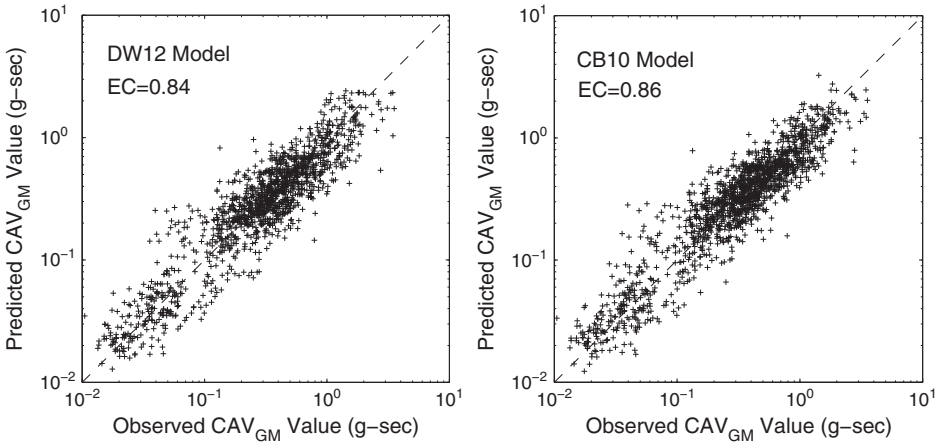


Figure 8. Distribution of observed and predicted CAV_{GM} values for CB10 and DW12 models using total dataset.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

least 0.4, and the normalized sample standard deviation (STDNR) is required to be smaller than 1.125 according to [21].

Two models are further tested using subsets of the NGA database according to the scheme proposed in [22]. The subsets divide the database into different site conditions, distance and magnitude ranges according to the criterion listed in the footnotes of Table IV. The *EC* and median *LH* values for each subset are also presented in Table IV. In general, quite similar results are obtained for two CAV models, although CB10 model does show slightly better performance overall. When the database is divided into soil and rock sites, the values of *EC* are quite similar for two cases. In terms of different distance ranges, all models have the best performance at the medium distance range where most data are available. Although there is a lack of data at small distance ($R_{\text{rup}} \leqslant 10$ km) and large distance ($100 \leqslant R_{\text{rup}} \leqslant 200$ km) ranges, CB10 model scores *EC* values about 7% higher than those of DW12 model. The superiority of CB10 model is not unexpected because it has additional parameters and much more complicated functional forms that could improve the model performance for these cases. Both models demonstrate relatively large deviation of the median *LH* values from 0.5 for the large distance subset and the footwall case. The detailed distributions of *LH* values for each subset are also compared in Figure 9. Considering all data used in each model, the distribution of *LH* values for both models closely approximate a uniform distribution. Therefore, we conclude that both models have similar global prediction accuracy and can well satisfy the underlying model assumption.

The performance of two CAV models is further compared with that of the NGA models. For this purpose, the PGA predicted from Campbell and Bozorgnia NGA model [9] (termed as CB08) are used and their associated *EC* and *LH* scores are listed in Table IV. For most cases, both CAV models achieve larger *EC* values compared with these of CB08 model. It is interesting to point out that CB10 and CB08 models utilize identical function forms and database. Yet, the predictive capacity of CB10 CAV model is consistently better than CB08 NGA model in almost all cases. On the other hand, although DW12 model uses less number of parameters and a much simpler functional form, its performance is in general better than the CB08 NGA model. The finding corroborates the

Table III. Ranking of DW12 and CB10 models.

Model name	Rank	MEDLH	MEDNR	MEANNR	STDNR	No. of records
DW12	A	0.548	0.088	0.071	0.993	1390
CB10	A	0.509	0.056	0.033	1.009	1560

Note: Goodness-of-fit measures used are: median LH values (MEDLH), and MEDNR, MEANNR, and STDNR denote the median, mean and the standard deviation of the normalized residuals, respectively.

Table IV. Goodness-of-fit results for model testing subsets

Subsets	Coefficient of efficiency (*CE*)			Median *LH*		
	CAV (CB10)	CAV (DW12)	PGA (CB08)	CAV (CB10)	CAV (DW12)	PGA (CB08)
Soil site	0.868	0.836	0.789	0.532	0.559	0.522
Rock site	0.845	0.840	0.746	0.443	0.526	0.447
Small R	0.695	0.643	0.353	0.547	0.560	0.483
Medium R	0.863	0.840	0.733	0.518	0.562	0.516
Large R	0.789	0.732	0.551	0.457	0.416	0.486
Small M	0.851	0.833	0.797	0.500	0.561	0.467
Large M	0.683	0.644	0.710	0.515	0.531	0.537
HW cases	0.789	0.736	0.722	0.528	0.540	0.562
FW cases	0.826	0.794	0.437	0.619	0.578	0.402

Note: the subsets used in the table are defined using following criteria
•Soil site ($180 \leqslant V_{s30} \leqslant 450$ m/s), rock site ($450 < V_{s30} \leqslant 1300$ m/s);
•Small R ($R_{rup} \leqslant 10$ km), medium R ($10 < R_{rup} \leqslant 100$ km), large R ($100 < R_{rup} \leqslant 200$ km);
•Small M ($5 < M_w \leqslant 6.5$), large M ($6.5 < M_w \leqslant 8$);
•HW cases (hanging-wall sites), FW cases (foot-wall sites).

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

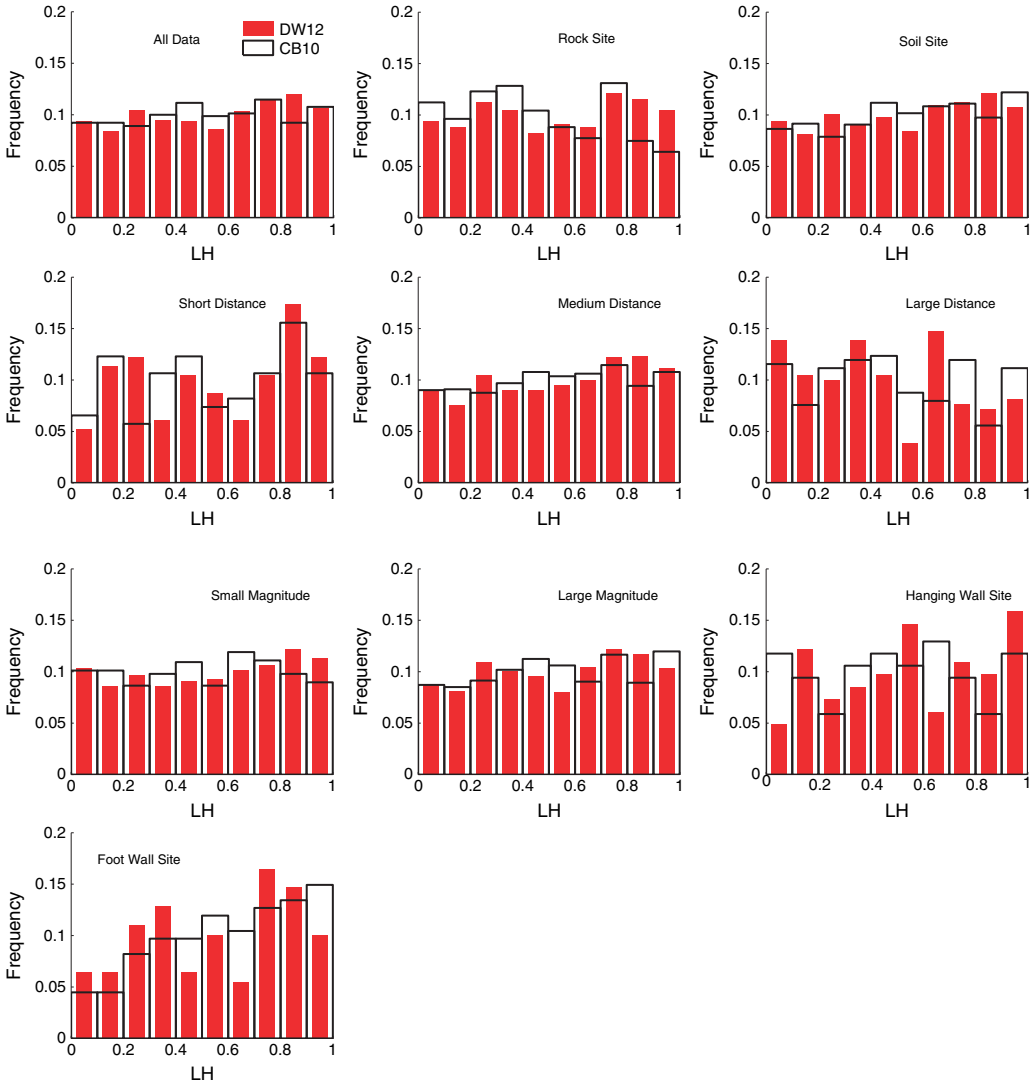


Figure 9. Comparison of LH distributions for two CAV models.

results that the residuals associated with CAV models are less than that associated with NGA models, indicating that CAV is a more predictable ground motion parameter than the spectral accelerations. Therefore, satisfactory performance can be achieved even by using a simplified GMPE as is presented in this study.

3.2. *Blind tests on recent earthquakes*

In this session, blind test will be conducted to further evaluate the performance of the models using strong motion data from nine recent earthquakes that are not included in the model regression. A total of 1599 ground motion records were compiled from four recent earthquakes in California (2003 San Simeon earthquake, 2004 Parkfield earthquake, 2005 Anza earthquake and 2010 Baja earthquake), one earthquake in New Zealand (2010 Darfield earthquake), and four earthquakes in Japan (2000 Tottori earthquake, 2004 Niigata earthquake, 2007 Chuetsu earthquake and 2008 Iwate earthquake). Detailed information of these earthquakes is listed in Table V. The selected records are all from mainshocks of the earthquakes, and the majority of events are of large magnitude. To be consistent with the developed models, only data from rupture distance within 200 km are used. Seismological information of each earthquake was systematically compiled from the strong motion databases such as CESMD [23], CESMOS [24], K-NET [25] and GeoNet [26].

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

Table V. Earthquakes used in blind test.

Earthquake name	Date (mm/dd/yyyy)	Magnitude	Hypocenter latitude (°)	Hypocenter longitude (°)	Fault mechanism	No. of recordings
San Simeon	12/22/2003	6.52	35.706	−121.102	Reverse	30
Parkfield	09/28/2004	6	35.817	−120.365	Strike-slip	89
Anza	06/12/2005	5.2	35.533	−116.578	Reverse-oblique	111
Baja	04/04/2010	7.2	32.128	−115.303	Strike-slip	136
Darfield	09/03/2010	7	−43.615	172.049	Strike-slip	63
Tottori	10/06/2000	6.61	35.275	133.350	Strike-slip	235
Niigata	10/23/2004	6.63	37.307	138.839	Reverse	365
Chuetsu	07/16/2007	6.8	37.538	138.617	Reverse	341
Iwate	06/13/2008	6.9	39.027	140.878	Reverse	229

Note: only recorded data within rupture distance of 200 km is included.

Table VI. Goodness-of-fit results for blind test.

Subsets	Coefficient of efficiency *CE*			Median *LH*		
	CAV (CB10)	CAV (DW12)	PGA (CB08)	CAV (CB10)	CAV (DW12)	PGA (CB08)
San Simeon	0.621	0.621	0.312	0.588	0.483	0.136
Parkfield	0.615	0.606	0.781	0.319	0.304	0.394
Anza	0.463	0.421	0.472	0.406	0.488	0.434
Baja	0.629	0.561	0.555	0.374	0.413	0.501
Darfield	0.809	0.816	0.804	0.437	0.522	0.464
Tottori	0.442	0.423	0.580	0.282	0.288	0.366
Niigata	0.540	0.480	0.328	0.224	0.266	0.217
Chuetsu	0.373	0.321	−0.03	0.251	0.282	0.099
Iwate	0.461	0.452	0.553	0.342	0.324	0.324

The goodness-of-fit measures are reported in Table VI on the basis of the total set of records from each earthquake. Again, two CAV models achieved quite similar results. The values of *ECs* for two CAV models approximately fall into the range between 0.4 and 0.8 for all events. In comparison, more varied predictive capacity is observed from CB08 NGA models. In particular, *EC* of CB08 NGA model is only −0.03 for Chuetsu earthquake, indicating that the predictive capacity of the CB08 model is only equivalent to using the sample average for that event. The model performance from the blind test is substantially worse than those presented in Table IV, because all the earthquakes used in the blind test are not included in the model development.

4. CONCLUSIONS

This study is motivated to develop a simplified ground-motion prediction model for the geometric mean of horizontal components of CAVs using NGA strong motion database. The present model (DW12) requires only four parameters and has a simple functional form. The model is applicable for shallow crustal earthquake events with moment magnitude 5–8 and rupture distance 0–200 km.

Systematical model validation and comparison between CB10 model and DW12 model have been conducted. The results indicate that both CAV models have 'Rank A' predictive capacity [21], and the performance of DW12 model is in par with CB10 model despite of its simplicity. On the other hand, CB10 model does demonstrate slightly better performance because of its complexity and sophistication. In addition, the efficiency of two CAV models is in general better than that of the CB08 NGA model for spectral ordinates.

The validation test results are in agreement with previous findings that the CAV has higher predictability than other intensity measures that have been investigated to date [7, 8]. The high

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

predictability of the CAV warrants the use of a simple functional form in the search of an alternative prediction model, which is needed in the logic tree framework to estimate the epistemic uncertainty in seismic hazard analysis. Besides, this simple model would have significant advantage in situations where the ground-motion parameters for a sophisticated model are unknown or highly uncertain.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from Hong Kong Research Grants Council RGC 620311 to conduct the study. We also acknowledge useful information provided by Dr. Brain Chiou. We also thank comments from Dr. Gail Atkinson and Dr. Yousef Bozorgnia to improve the quality of the paper. Ground-motion records for the blind test are obtained from the following databases: (i) Center for Engineering Strong Motion Data at http://strongmotioncenter.org/; (ii) National Strong-Motion Project at http://nsmp.wr.usgs.gov/; (iii) Consortium of Organizations for Strong-Motion Observation Systems at http://db.cosmos-eq.org/; (iv) GeoNet at http://geonet.org.nz/index.html/; and (v) K-Net at http://www.k-net.bosai.go.jp/. All these resources are in public domain.

REFERENCES

1. Power M, Chiou B, Abrahamson N, Bozorgnia Y, Shantz T, Roblee C. An overview of the NGA project. *Earthquake Spectra* 2008; **24**(1):3–21.
2. Arias A. A measure of earthquake intensity. In *Seismic Design for Nuclear Power Plants*, Hansen RJ (ed.). MIT Press: Cambridge, MA, 1970; 438–483.
3. Electrical Power Research Institute (EPRI). A criterion for determining exceedance of the operating basis earthquake, Report No. EPRI NP-5930, Palo Alto, California, 1988.
4. Electrical Power Research Institute (EPRI). Standardization of the cumulative absolute velocity, Report No. EPRI TR-100082-T2, Palo Alto, California, 1991.
5. Dashti S, Bray JD, Pestana JM, Riemer M, Wilson D. Centrifuge testing to evaluate and mitigate liquefaction-induced building settlement mechanisms. *Journal of Geotechnical and Geoenvironmental Engineering* 2010; **136**(7):918–929.
6. Kramer SL, Mitchell RA. Ground motion intensity measures for liquefaction hazard evaluation. *Earthquake Spectra* 2006; **22**(2):413–438.
7. Campbell KW, Bozorgnia Y. A ground motion prediction equation for the horizontal component of cumulative absolute velocity (CAV) based on the PEER-NGA strong motion database. *Earthquake Spectra* 2010; **26**(3): 635–650.
8. Danciu L, Tselentis G. Engineering ground-motion parameters attenuation relationships for Greece. *Bulletin of the Seismological Society of America* 2007; **97**(1B):162–183.
9. Campbell KW, Bozorgnia Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthquake Spectra* 2008; **24**:139–171.
10. Rodriguez-Marek A, Bray JD, Abrahamson NA. An empirical geotechnical seismic site response procedure. *Earthquake Spectra* 2001; **17**(1):65–87.
11. Brillinger DR, Preisler HK. An exploratory analysis of the Joyner–Boore attenuation data. *Bulletin of the Seismological Society of America* 1984; **74**(4):1441–1450.
12. Brillinger DR, Preisler HK. Further analysis of the Joyner–Boore attenuation data. *Bulletin of the Seismological Society of America* 1985; **75**(2):611–614.
13. Travasarou T, Bray JD. Empirical attenuation relationship for Arias Intensity. *Earthquake Engineering and Structural Dynamics* 2003; **32**:1133–1155.
14. Joyner WB, Boore DM. Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America* 1993; **83**(2):469–487.
15. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core team. NLME: linear and nonlinear mixed effects models. R package version 3, 2008; 1–89.
16. Abrahamson NA, Silva WJ. Summary of the Abrahamson and Silva NGA ground-motion relations. *Earthquake Spectra* 2008; **24**(1):64–97.
17. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. Prentice Hall: Upper Saddle River, New Jersey, 2007.
18. Jayaram N, Baker JW. Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America* 2008; **98**(5):2231–2243.
19. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models: Part I, a discussion of principles. *Journal of Hydrology* 1970; **10**(3):282–290.
20. Legates DR, McCabe GJ. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimaitic model validation. *Water Resources Research* 1999; **35**(1):233–241.

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe

21. Scherbaum F, Cotton F, Smit P. On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion. *Bulletin of the Seismological Society of America* 2004; **94**(6):2164–2185.
22. Kaklamanos J, Baise LG. Model validations and comparisons of the next generation attenuation of ground motions (NGA-West) project. *Bulletin of the Seismological Society of America* 2011; **101**(1):160–175.
23. CESMD strong motion database. Available from: http://strongmotioncenter.org/ [Accessed on September 2011].
24. COSMOS strong motion database. Available from: http://www.cosmos-eq.org/ [Accessed on September 2011].
25. K-NET strong motion database. Available from: http://www.k-net.bosai.go.jp/ [Accessed on September 2011].
26. GeoNet strong motion database. Available from: www.geonet.org.nz/ [Accessed on September 2011].

Copyright © 2012 John Wiley & Sons, Ltd.

DOI: 10.1002/eqe