Bulletin of the Seismological Society of America, Vol. 104, No. 5, pp. 2456–2466, October 2014, doi: 10.1785/0120130191

Ⓔ

An Efficient Algorithm to Identify Strong-Velocity Pulses in Multicomponent Ground Motions

by Shrey K. Shahi and Jack W. Baker

Abstract Ground motions with strong velocity pulses are of special concern for structural engineers. We describe an efficient and quantitative method to identify such ground motions. Previous algorithms to classify these pulse-like ground motions considered the ground motion in a single orientation, which made classifying pulses in arbitrary orientations difficult. We propose an algorithm that can identify pulses at arbitrary orientations in multicomponent ground motions, with little extra computational cost relative to a single-orientation calculation. We use continuous wavelet transforms of two orthogonal components of the ground motion to identify the orientations most likely to contain a pulse. The wavelet transform results are then used to extract pulses from the selected orientations, and a new classification criterion based on support vector machines is proposed. Because we are mostly interested in forward directivity pulses, which are found early in the time history, a criterion to reject pulses arriving late in the time history is also proposed. The procedure was used to classify ground motions in the Next Generation Attenuation-West2 database (Ancheta *et al.*, 2013). The list of pulse-like ground motions was then manually filtered using source-to-site geometry and site conditions to find the pulses most likely caused by directivity effects. Lists of both pulse-like ground motions and directivity ground motions are provided, along with the periods of the pulses and the orientations in which the pulses were strongest. Using the classification results, new models to predict the probability of a pulse and pulse period for a given future earthquake scenario are developed.

Online Material: Table of all the pulse-like ground motions identified in the Next Generation Attenuation-West2 database, and HTML files of source-to-site geometry, the recorded time history, extracted pulse, and other directivity parameters for each identified pulse.

Introduction

Pulse-like ground motions, which are often caused by forward directivity effects, can impose larger demands on structures than non-pulse-like ground motions (e.g., Housner and Hudson, 1958; Bertero *et al.*, 1978; Anderson and Bertero, 1987; Hall *et al.*, 1995; Iwan, 1997; Menun and Fu, 2002; Alavi and Krawinkler, 2004; Makris and Black, 2004; Mavroeidis *et al.*, 2004; Akkar *et al.*, 2005; Luco and Cornell, 2007). The larger potential for causing damage makes the pulse-like ground motions very important in near-fault regions, where pulses are generally expected. Empirical models are needed to quantify the hazard and risk posed by these pulse-like ground motions. To calibrate such models, we need a library of ground motions where each record is classified as pulse-like or non-pulse-like. Empirical models have been developed in the past using a version of the Next Generation Attenuation (NGA) database (Chiou *et al.*, 2008), with each ground motion classified as pulse-like or non-pulse-like (e.g., Iervolino and Cornell, 2008; Champion and Liel, 2012), which shows the importance of this line of research. Early research in the field used visually classified pulses but this approach is not reproducible and does not scale with increasing size of ground-motion databases. With the rapidly increasing size of ground-motion databases (e.g., the NGA-West2 database is over seven times the size of the NGA database), there is a need for an algorithm that is both computationally efficient and reproducible.

We propose an algorithm to classify multicomponent ground motions as pulse-like or non-pulse-like. The proposed algorithm uses the wavelet transform of two orthogonal components of a ground-motion time history to find orientations that are likely to contain strong pulses. The ground-motion recording is rotated to those specific orientations, and wavelet

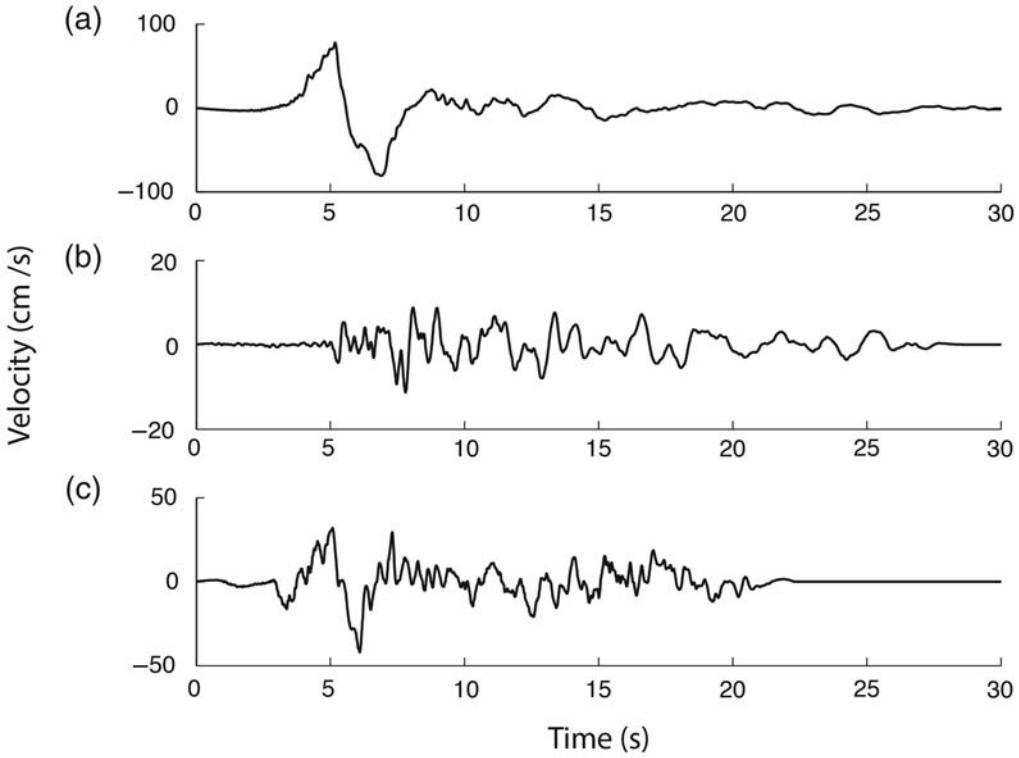

Figure 1. Different types of ground motions: (a) clear pulse (El Centro Array 4 recording from the 1979 Imperial valley earthquake), (b) clear nonpulse (Pasadena-CIT recording from the 1971 San Fernando earthquake), and (c) an ambiguous pulse (Salton Sea Wildlife Refuge recording from the 1987 Superstition Hills earthquake).

coefficients of the ground motion in those selected orientations are used to extract pulses from the recorded velocity time history. A nonlinear classification boundary developed using support vector machines (e.g., Cortes and Vapnik, 1995; Hastie *et al.*, 2001) uses the extracted pulse and the peak ground velocity (PGV) of the ground-motion record to classify the ground motion.

Classifying ground motions as pulse-like or non-pulse-like is a difficult task. There are some ground motions with clear presence or absence of visible pulses in the velocity time history that can be identified unambiguously. However, there are many ground motions that are difficult to classify as a clear pulse or a clear nonpulse by only visually examining the velocity time histories (e.g., Fig. 1c). These cases make the manual classification of pulses subjective and make this problem challenging. An algorithm to classify pulse-like ground motions should provide defensible classifications, it should check multiple orientations for pulses, and it should be computationally efficient. The proposed algorithm improves all three aspects relative to alternative approaches.

The Baker (2007) algorithm classifies a ground motion as pulse-like by examining a single component of the time-history recording (typically the fault-normal orientation). It is known that pulse-like ground motions are observed in many orientations other than fault normal (e.g., Mavroeidis and Papageorgiou, 2002; Howard *et al.*, 2005), and sometimes, due to lack of finite-fault models, the determination of fault-normal orientation may be difficult. Because of these difficulties, Baker (2007) fails to classify a ground motion as pulse-like when the pulse is in a non-fault-normal orientation or when the fault-normal orientation itself is not known. Shahi and Baker (2011a) rotated the ground motion in all nonredundant orientations (orientations spanning 180°) and classified each orientation as pulse-like or non-pulse-like using the Baker (2007) algorithm. The ground motion is then labeled as pulse-like if a pulse was found in any orientation. This method overcomes the difficulty of identifying non-fault-normal pulses and deals with cases in which the fault-normal orientation is not known, but is computationally very expensive. With the Shahi and Baker (2011a) approach, the ground motion is labeled pulse-like if even a single orientation is classified as pulse-like; therefore, using many orientations for classification increases the chance of false-positive classifications (non-pulse-like ground motions being classified as pulse-like). The proposed algorithm strikes a balance between considering just one orientation and all possible orientations while using an improved classification criteria to reduce the possibility of false-positive classifications. Additionally, we use the orthogonality of the wavelet transform in perpendicular orientations to dramatically reduce the computational expense of the classification procedure.

We use the data from the NGA-West2 database (Ancheta *et al.*, 2013) in this study. The NGA-West2 database is a collection of ground motions recorded worldwide from shallow

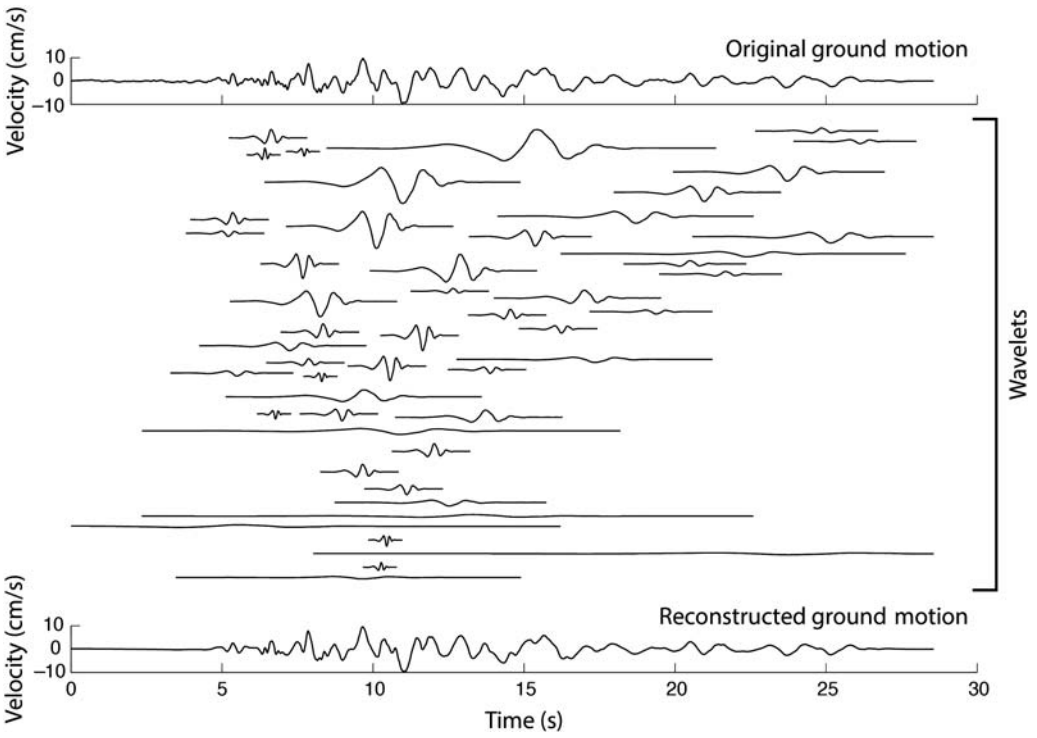

Figure 2. A ground-motion time history (original ground motion) is broken down into 50 wavelets using continuous wavelet transform. The wavelets are summed together to get the reconstructed ground motion, which is an approximation of the original ground motion. The quality of approximation improves as the number of wavelets used is increased. Pasadena-CIT Athenaeum recording, from the 1971 San Fernando earthquake, is used in this figure.

crustal earthquakes. The database has a magnitude range of 3–7.9, and the closest distance between rupture and site varies from 0.05 to 1533 km. The time-averaged shear-wave velocity in the top 30 m at the recording sites (V_{S30}) ranges from 94 to 2100 m/s. The procedure proposed here was used to classify a subset of ground motions in the NGA-West2 database that were available in 2012 (record sequence numbers 1–8611 in the database). Out of 8611 considered ground motions, 244 were classified as pulse-like. We manually filtered the 244 pulse-like ground motions using source-to-site geometry and site condition information to prepare a list of 148 pulse-like ground motions that were likely caused by directivity effects. The results of these classifications, along with improved models to predict the probability of pulse and its period for a given earthquake scenario, are provided.

Wavelet Transform for Multicomponent Ground Motion

An earthquake ground-motion time history is a nonstationary signal in both time and frequency (i.e., both the ground-motion amplitudes and frequencies change over time). The nonstationary nature of ground motion makes it difficult to find a feature like a pulse by analyzing only the time-domain representation of the signal. Because pulses have high energy in a short time interval and the energy is also concentrated in a small frequency region (i.e., the energy is carried by a coherent pulse with a well-defined frequency), a transformation of the signal in a domain that captures both time and frequency characteristics makes finding a pulse easier. Wavelet transforms provide a good representation of the signal in time and frequency domains (e.g., Mallat, 1999) and thus are very useful for the task of pulse classification (e.g., Baker, 2007).

The wavelet transform involves representing the signal as a sum of scaled and translated mother wavelets, represented as

$$\Phi_{s,l}(t) = \frac{1}{\sqrt{s}}\phi\left(\frac{t-l}{s}\right), \tag{1}$$

in which $\phi(\cdot)$ and $\Phi_{s,l}(\cdot)$ represent the mother wavelet and the scaled and translated wavelet, respectively, as a function of time (t). These wavelets are transformed in the frequency domain by changing the scale parameter s (this amounts to stretching and contracting the mother wavelet) and in the time domain by changing the parameter l (this translates the wavelet shape on the time axis). A wavelet transform breaks the signal into a sum of such scaled and translated mother wavelets as shown in Figures 2 and 3. The Daubechies wavelet of order 4 was used as the mother wavelet in this study.

The wavelet transform is similar to the Fourier transform, but the Fourier transform decomposes the signal into a sum of sine and cosine waves for which the frequency is constant with time, whereas the wavelet transform decomposes the signal into shapes that are localized in small time and frequency regions. This makes wavelets more suited for representing nonstationary signals like earthquake ground

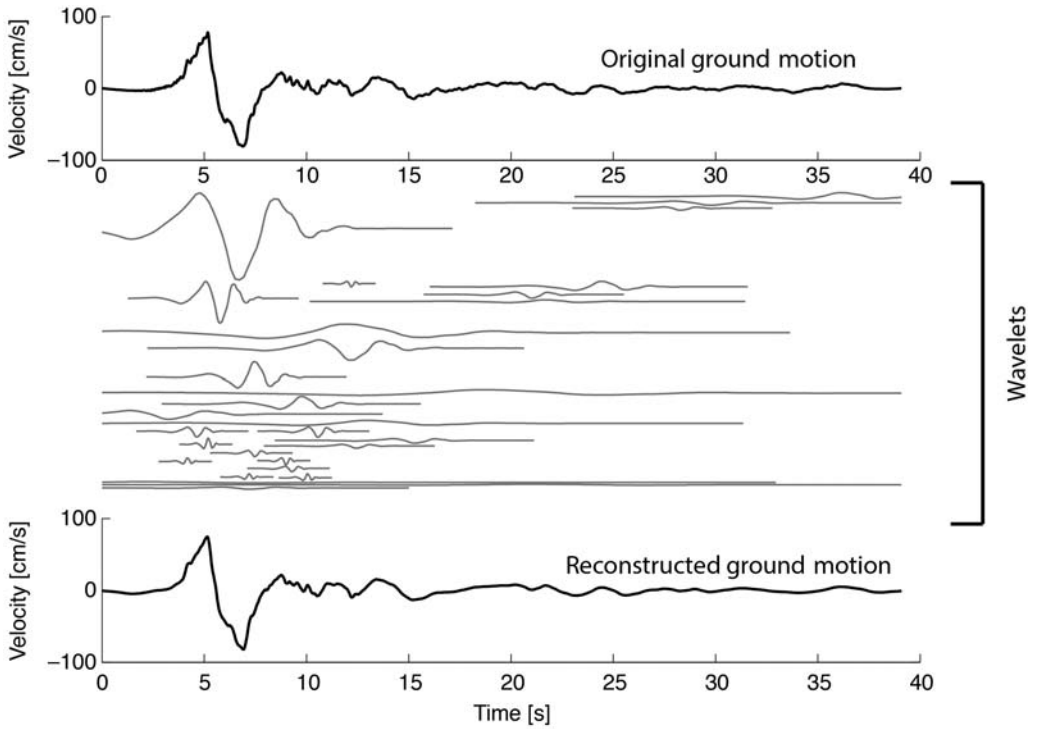

Figure 3. A pulse-like ground-motion time history (original ground motion) is broken down into 30 wavelets using continuous wavelet transform. The wavelets are summed together to get the reconstructed ground motion, which is an approximation of the original ground motion. Pulse-like ground motions are dominated by a few strong wavelets. An El Centro Array 4 recording, from the 1979 Imperial valley earthquake, is used in this figure.

motion, where the frequency characteristics of the signal change with time. The continuous wavelet transform coefficient for a signal ($f(t)$) at a particular location (l) and scale (s) can be found using the integral

$$c(s,l) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\phi\left(\frac{t-l}{s}\right)dt. \qquad (2)$$

Large-amplitude coefficients are associated with a concentration of energy in a small time and frequency range, which is characteristic of a pulse; this property allows us to search for pulses efficiently. Additionally, a linear combination of wavelets can approximate any signal. So, if we use a combination of a few wavelets to describe the pulse, we can identify a wide variety of pulse shapes.

Continuous wavelet transform coefficients from two orthogonal components of a ground motion can be combined linearly to yield the coefficients for any arbitrary orientation. We use this fact to efficiently compute wavelet coefficients in all orientations. This is illustrated below:

$$f(t,\theta) = f_1(t) \times \cos(\theta) + f_2(t) \times \sin(\theta) \qquad (3)$$

and

$$c(s,l,\theta) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} f(t,\theta)\phi\left(\frac{t-l}{s}\right)dt, \qquad (4)$$

in which $f_1(t)$ and $f_2(t)$ represent ground motions in orthogonal orientations. $f(t,\theta)$ represents the ground motion in an arbitrary angle θ away from $f_1(t)$, which can be constructed using the linear combination of $f_1(t)$ and $f_2(t)$ as shown in equation (3). Substituting $f(t,\theta)$ from equation (3) into equation (4), we get

$$c(s,l,\theta) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}[f_1(t) \times \cos(\theta)]\phi\left(\frac{t-l}{s}\right)dt + \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty}[f_2(t) \times \sin(\theta)]\phi\left(\frac{t-l}{s}\right)dt \qquad (5)$$

and

$$= c_1(s,l) \times \cos(\theta) + c_2(s,l) \times \sin(\theta), \qquad (6)$$

in which c_1 and c_2 are the wavelet coefficients corresponding to f_1 and f_2, respectively. This procedure is shown graphically in Figure 4. As the wavelet coefficient in an arbitrary orientation is $c_1\cos(\theta) + c_2\sin(\theta)$, the maximum value a coefficient can obtain in any orientation ($c_{\max}(s,l)$) can be found using

$$c_{\max}(s,l) = \max_{\theta} c(s,l,\theta) = \sqrt{c_1^2(s,l) + c_2^2(s,l)}. \qquad (7)$$

Using this procedure we only need to perform two wavelet transforms (of ground motions in orthogonal orientations) to find the maximum value of the coefficient at each scale and location over all orientations.

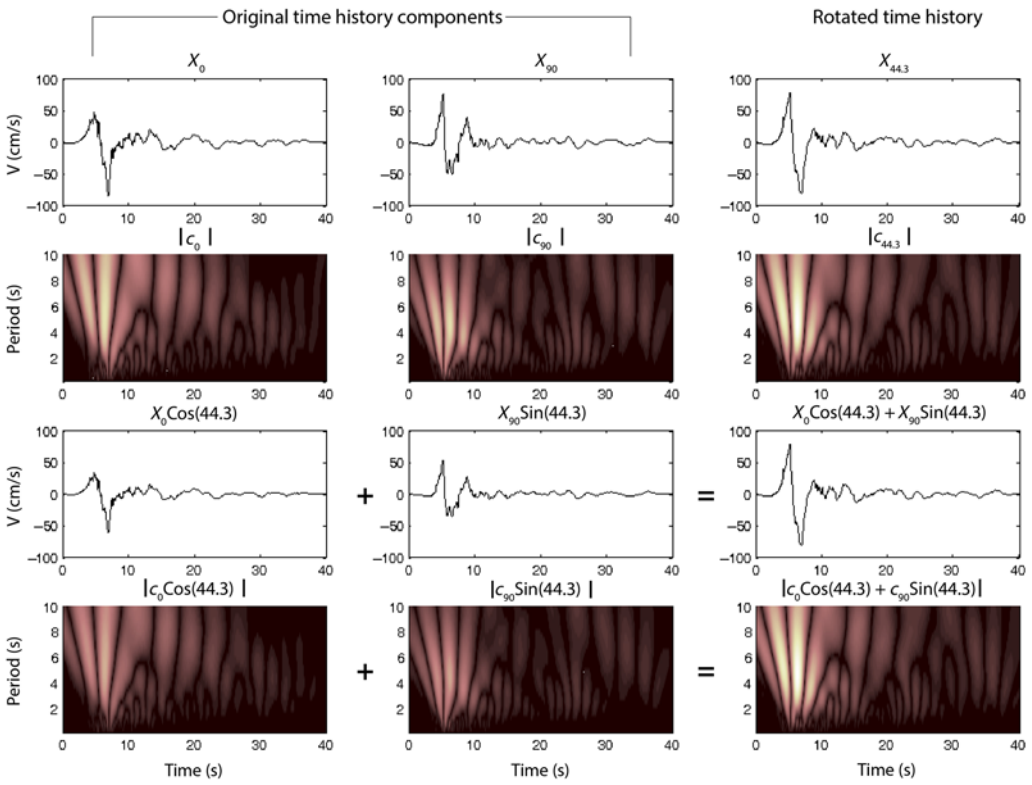

Figure 4. Two orthogonal components of a ground motion, labeled X_0 and X_{90}, are shown with the absolute value of their wavelet transform coefficients, labeled $|c_0|$ and $|c_{90}|$. The wavelet coefficients c_0 and c_{90} are combined to get an estimate of $c_{44.3}$ (coefficients for ground motion in an orientation 44.3° away from X_0). The wavelet coefficients computed directly using the rotated time histories are shown for comparison. Lighter regions indicate large coefficient amplitudes. The color version of this figure is available only in the electronic edition.

Classification Algorithm

The continuous wavelet transform coefficients are computed for two orthogonal components of the ground motion. These coefficients are then used to compute the maximum wavelet coefficients at each location and scale over all orientations using the procedure described previously. Large wavelet coefficients indicate a concentration of energy in a small time and frequency region, which is often a good indication of presence of a pulse. The Baker (2007) and Shahi and Baker (2011a) algorithms use only the wavelet with the largest coefficient to classify the ground motion as pulse-like or non-pulse-like. Sometimes, however, a wavelet other than the largest wavelet can be the dominant pulse in the ground motion. To identify these cases, we use five potential pulses for the purpose of classification. First, we select the wavelet with the largest coefficient as a potential pulse. All coefficients located in a time window of $\pm(1/2)$ the width of the selected wavelet (s) are labeled as being adjacent to the pulse. Then, we find the wavelet with the largest coefficient nonadjacent to the selected coefficient as the second potential pulse. This process is repeated to select five nonadjacent potential pulses. Choosing five potential pulses increases the chance of finding a pulse having a smaller wavelet coefficient.

For each potential pulse, we rotate the velocity time history to the orientation in which the pulse was found, as determined using equation (10):

$$c_1(s,l) = c_{\max}(s,l) \times \cos(\beta), \tag{8}$$

$$c_2(s,l) = c_{\max}(s,l) \times \sin(\beta), \tag{9}$$

and

$$\beta = \tan^{-1}\left(\frac{c_2}{c_1}\right), \tag{10}$$

in which $c_{\max}(s,l)$ represents the maximum wavelet coefficient at scale s and location l over all orientations and β represents the orientation in which this coefficient is found. We refer to the velocity time history in this orientation as the original ground motion. The selected wavelet is then subtracted from the original ground motion to yield a residual ground motion. The continuous wavelet transform of the residual ground motion is used to find the wavelet with the highest coefficient that has the same scale as the original wavelet and is located within a region of $\pm(1/2)s$ of the original wavelet (i.e., in a region adjacent to the selected wavelet). This wavelet is then added to the original wavelet

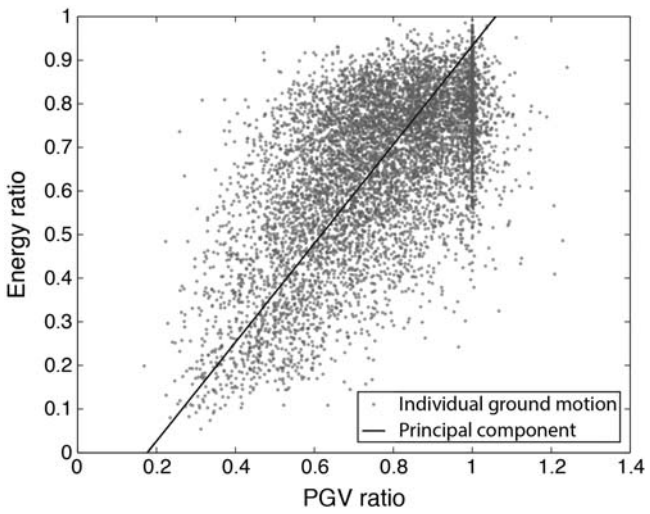

Figure 5. The principal component is shown with the energy ratio and peak ground velocity (PGV) ratio of individual ground motions. Sometimes the PGV ratio of residual to original ground motion can be greater than 1.

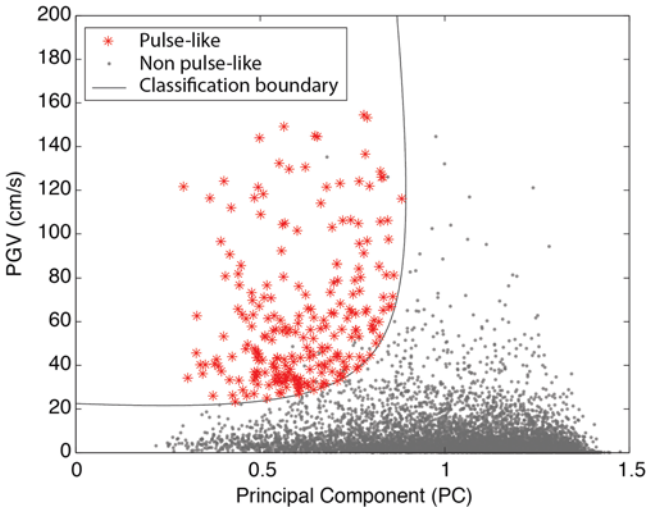

Figure 6. Classification boundary of equation (12) shown with individual ground motions. The non-pulse-like ground motions on the positive pulse indicator side of the figure were rejected by the late arrival criteria. All 8611 ground motions used in the study are shown here. The color version of this figure is available only in the electronic edition.

to refine the shape of the selected pulse. We repeat this step 10 times and use a combination of the 10 wavelets to define the shape of the extracted pulse.

After the pulse is extracted from the original ground motion, we determine whether the extracted pulse is strong enough for the record to be classified as pulse-like. We use the parameters recommended by Baker (2007) to make this determination: the energy ratio of residual and original ground motions, the PGV ratio of the residual and original ground motions, and the PGV of the original ground motion. These variables are selected as they directly measure the strength of the extracted pulse relative to the recorded ground motion. For example, a lower energy ratio means that the pulse subtracted from the original ground motion contributes significantly to the energy of the ground motion, whereas a lower PGV ratio shows that the pulse is responsible for the peak velocity of the ground motion (because removing the pulse led to a large reduction in PGV). A weak ground motion with only few cycles of shaking can also lead to low PGV and energy ratio if the cycle of the motion responsible for the PGV is extracted as the pulse. To differentiate between these weak motions and the pulse-like motions, we use the PGV of the recorded motion in the classification criteria.

We observed that the energy ratio and PGV ratio are correlated and most of the variance in these values lies along one axis, as shown in Figure 5. Principal component analysis was used to find the linear combination of the two variables that captured the most variance:

$$\mathrm{PC} = 0.63 \times (\text{PGV ratio}) + 0.777 \times (\text{energy ratio}). \quad (11)$$

The two variables (energy ratio and PGV ratio) are replaced by this linear combination, which captures most of the information expressed by the variables individually. This linear combination is referred hereafter as the principal component (PC). Using PC reduces the number of variables used in the classification procedure, making it easier to find a good classification criterion.

We manually classified 50 pulses and 50 nonpulses in the NGA-West2 database and used support vector machines to find a criterion to separate the pulses from nonpulses. Once the criterion was found, it was used to classify ground motions from the NGA-West2 database, and then the clear misclassifications were manually corrected. The classification scheme was then refitted using the manually corrected data. This fitting and manual correction was iterated until a satisfactory classification result was reached. Only clear pulses or nonpulses were used during this procedure. Ambiguous cases were not used while fitting the classification boundary. Ⓔ The distribution of the magnitude, the closest distance between site and rupture, and the PGV of the pulse-like and non-pulse-like records used to fit the classification boundary are shown in Figures S1, S2, and S3, available in the electronic supplement to this article.

A support vector classification boundary was fitted with a second-degree polynomial kernel, using PC and PGV as the independent variables. The classification boundary was used to create the following pulse indicator:

$$\mathrm{PI} = 9.384(0.76 - \mathrm{PC} - 0.0616\mathrm{PGV}) \\ (\mathrm{PC} + 6.914 \times 10^{-4}\mathrm{PGV} - 1.072) - 6.179. \quad (12)$$

The ground motion is classified as pulse-like when the pulse indicator (PI) is positive and as non-pulse-like if it is negative. PI = 0 defines the classification boundary shown in Figure 6.

The pulses caused by directivity effects arrive early in the time history. The algorithm described above identifies

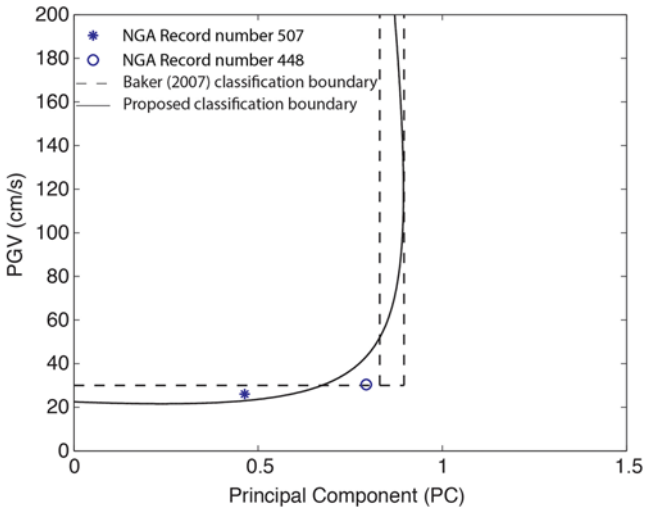

Figure 7. Comparison of the classification rule for the proposed algorithm and Baker (2007) algorithm. The two vertical lines for Baker (2007) show the maximum and minimum values the principal component (PC) can take when the pulse indicator is equal to 0.85. The PC and PGV values of two records where the proposed classification differs from the Baker (2007) classification are also shown. The velocity time histories for these records are shown in Figure 8. The color version of this figure is available only in the electronic edition.

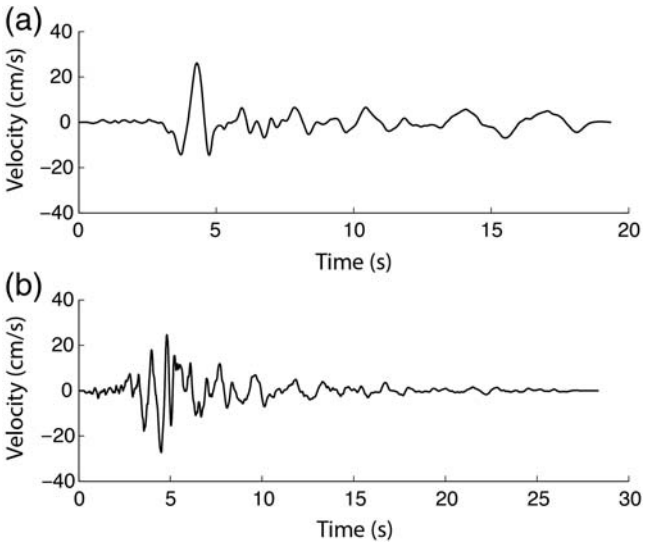

Figure 8. (a) A pulse-like ground motion with low PGV (Next Generation Attenuation [NGA] record number 507, SMART1 M01 recording from the 1986 magnitude 6.32 Taiwan earthquake). (b) A nonpulse classified as pulse-like in a few orientations by the Shahi and Baker (2011a) algorithm (NGA record number 448, Anderson Dam recording from the 1984 magnitude 6.19 Morgan Hill earthquake). For ground motion (a), the orientation with highest wavelet coefficient is shown; for ground motion (b), an orientation that the Shahi and Baker (2011a) algorithm classified as a pulse is shown.

strong pulses, regardless of their locations in the time histories. If we are interested in pulses caused primarily by directivity effects, we need an additional criterion to exclude late arriving pulses. Following the approach in Baker (2007), we distinguish between late and early pulses using the cumulative square velocity (CSV), defined as

$$\mathrm{CSV}(t) = \int_0^t V^2(u)du, \tag{13}$$

in which $V(u)$ is the velocity of the ground motion at time u. The time at which $\mathrm{CSV}(t)$ attains $x\%$ of the total CSV of the signal is represented as $t_{x\%,\mathrm{orig}}$ and $t_{x\%,\mathrm{pulse}}$ for the original ground motion and the extracted pulse, respectively. We found that the early arriving pulses had $t_{17\%,\mathrm{orig}}$ greater than the $t_{5\%,\mathrm{pulse}}$, so pulses with $t_{17\%,\mathrm{orig}} \leq t_{5\%,\mathrm{pulse}}$ were reclassified as non-pulse-like.

The pulse classification algorithm described above is used to classify each of the five potential pulses, and the ground motion is called pulse-like if any of the five potential pulses is classified as pulse-like. Sometimes more than one potential pulse can be classified as pulse-like. In such cases, the pulse with largest wavelet coefficient is chosen as the dominant pulse.

Comparison with Previous Algorithm

The algorithm proposed here differs from the Baker (2007) algorithm, as the proposed algorithm uses improved classification criteria and has the ability to handle multicomponent ground motions. The proposed algorithm also examines fewer orientations than the Shahi and Baker (2011a) modification to classify pulses in arbitrary orientations. Here, we compare the new classification criteria and the Baker (2007) classification criteria.

The proposed algorithm classifies a ground motion as pulse-like when the pulse indicator shown in equation (12) is positive, whereas the Baker (2007) algorithm labels a motion as pulse-like if

$$\frac{1}{1 + e^{(-23.3+14.6(\mathrm{PGV\ ratio})+20.5(\mathrm{energy\ ratio}))}} > 0.85 \tag{14}$$

and its PGV is greater than 30 cm/s. These classification boundaries are compared in Figure 7. Because the Baker (2007) pulse indicator does not use PC, we show the maximum and minimum value that variable PC can take when the pulse indicator shown by equation (14) is equal to 0.85. The comparison shows that the classification boundaries agree with each other to a large extent. Though the classification regions are similar, the new classification criteria is an improvement as the threshold for PGV is learned from the data, whereas in Baker (2007) it was set arbitrarily to 30 cm/s.

Most of the difference in classification results between the proposed and previous algorithms are due to the change in the PGV threshold. Also, some false-positive classifications are removed, along with false-negative classifications that can occur if only one potential pulse is used for classification. Figure 8 shows two examples in which the new algorithm's classification differs from the old one. In the first case, the ground motion has a high pulse indicator but the

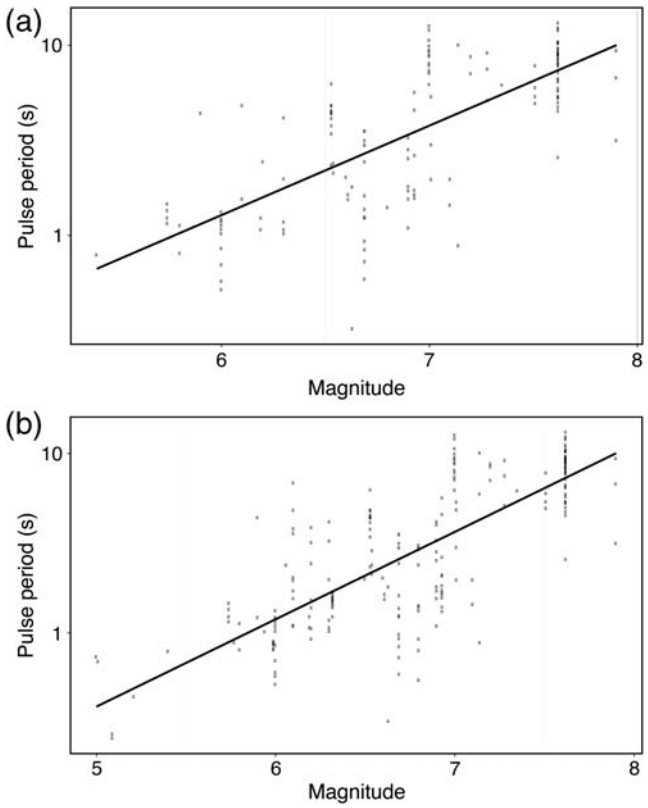

Figure 9. Regression fits to predict the period of the pulse for (a) directivity motions (equation 15) and (b) all pulses (equation 21).

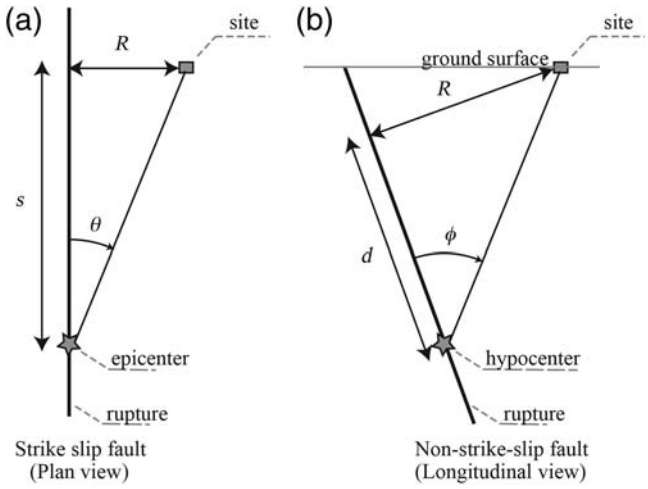

Figure 10. Parameters used to fit the logistic regressions for (a) strike slip and (b) non-strike-slip faults.

PGV is slightly below 30 cm/s, so the PGV cutoff used in Baker (2007) classifies the ground motion as non-pulse-like while the proposed classification criteria classifies it as pulse-like. In the second case, the ground motion is classified as non-pulse-like by the proposed classification algorithm; however, when we use the Baker (2007) algorithm in 180 different orientations, the ground motion is classified as a pulse in 16 of them. This is a case of a false-positive classification that is fixed by examining a smaller number of orientations. After examining several such cases, we concluded that the classification results from the proposed algorithm are consistently superior to those from the previous algorithm.

Directivity Models

Pulse-like ground motions can be caused by effects other than directivity, such as basin or soft-soil effects. To develop empirical relationships for pulses caused by directivity effects, we need to identify pulse-like ground motions caused specifically by directivity effects. To prepare a list of directivity ground motions, we manually filtered the list of pulse-like ground motions selected by the proposed algorithm to remove any ground motion that was likely not resulting from directivity effects. We mainly used the source-to-site geometry to identify directivity ground motions from the list of pulse-like ground motions. Ground motions on soft soils with multiple large cycles in the time history were also removed, as it generally indicates presence of soft-soil effects. Studies such as Chioccarelli and Iervolino (2010), Bradley and Cubrinovski (2011), and Bradley (2012) were used to aid in manual classification of ground motions from specific earthquakes. Ⓔ A list of pulse-like ground motions classified by the proposed algorithm and a manually filtered list of directivity pulses are provided in Table S1. This list of directivity ground motions was used to fit the following models.

Pulse Period

A pulse-like ground motion causes an amplification in the response spectra over a narrowband of periods centered about the pulse period (Somerville, 2003). This makes the pulse period a very important parameter. We use the period associated with the maximum Fourier amplitude of the wavelet, also known as the pseudoperiod of the wavelet (e.g., Baker, 2007), to estimate the pulse period. The quality of this estimate depends on the similarity of the shape of the mother wavelet with the shape of the pulse in the ground motion. When multiple potential pulses are classified as pulse-like, we used the period of the dominant pulse (the one with highest wavelet coefficient) to represent the pulse period at a site. The period of the velocity pulse is known to be related to source parameters like the rise time and fault dimensions, both of which increase with magnitude (e.g., Somerville, 2003). Various researchers have proposed models for the period of the pulse (e.g., Mavroeidis and Papageorgiou, 2003; Somerville, 2003; Bray and Rodriguez-Marek, 2004; Baker, 2007), and most model the log of pulse period as linearly dependent on the moment magnitude (M) of the earthquake. We refit this relationship using the directivity pulses identified here. The relationship is shown in equation (15), and the standard deviation of the residual from linear regression is given by equation (16):

$$\ln T_p = -6.256 + 1.084M \tag{15}$$

and

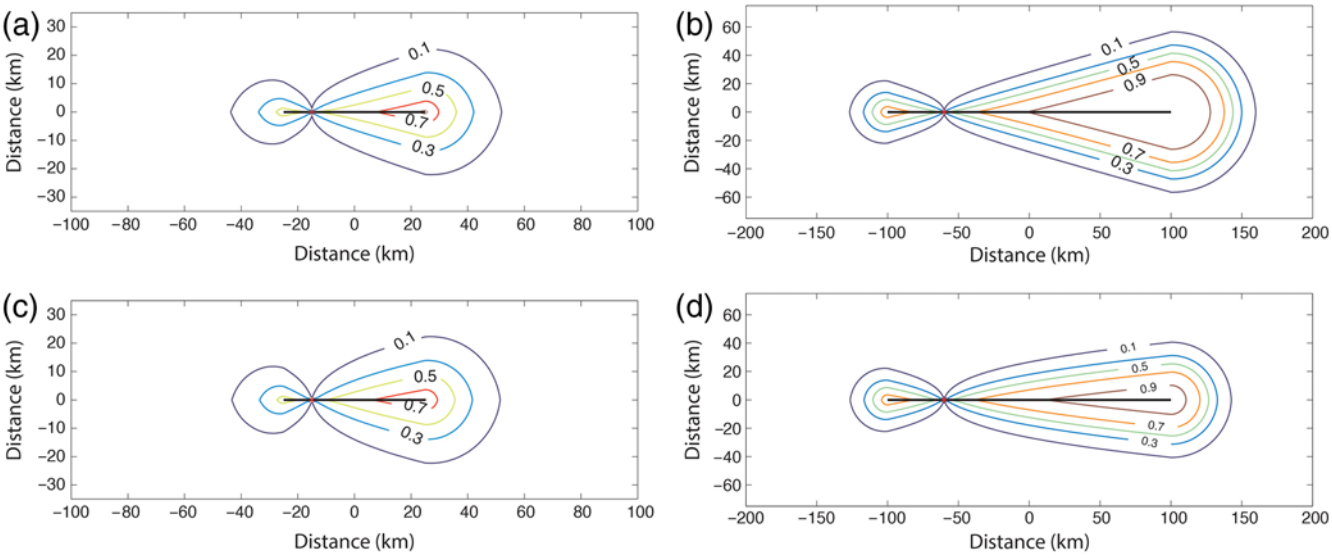

Figure 11. Probability of directivity pulse contours. Predictions from equation (17) for ruptures of length (a) 50 km and (b) 200 km. Predictions from equation (18) for ruptures of length (c) 50 km and (d) 200 km. The color version of this figure is available only in the electronic edition.

$$\sigma_{\ln T_p} = 0.61. \tag{16}$$

The regression fit is illustrated in Figure 9. This model is similar to the other published models mentioned above, and the model coefficients are similar to the self-similar scaling relationship presented in Somerville (2003), where the period of the pulse increases in proportion to the fault-rupture length.

Probability of Directivity Pulse

To account for directivity effects in hazard analysis, we need a model to predict the probability of observing a directivity pulse at a near-fault site. Several researchers have proposed models for probability of pulse at a site (e.g., Iervolino and Cornell, 2008; Shahi and Baker, 2011a). The Iervolino and Cornell (2008) model was fitted using pulses found only in fault-normal orientations, and Shahi and Baker (2011a) used pulse-like motions as a surrogate for ground motions with directivity. We refit these relationships using the new refined list of directivity pulses, as it should give a better estimate of the probability of observing directivity effects at a site. Several functional forms were explored, and special attention was given to how the predictions extrapolate for cases for which we do not have much data. We used logistic regression to model the probability of a directivity pulse at a site (see Shahi and Baker, 2011b, for detailed justification).

Iervolino and Cornell (2008) compared linear combinations of several predictor variables and found that source-to-site geometry parameters R, s, and θ for strike-slip ruptures and R, d, and ϕ for non-strike-slip ruptures resulted in the best models for predicting the probability of a pulse. These parameters are explained graphically in Figure 10. The dataset used to fit these relationships did not include many events with very long ruptures, so for cases with large s, these previous models predict very high probability of pulses even at large distances (R), which is not supported by data or by theoretical predictions of directivity. So, we tried several alternative functional forms for which the predictions extrapolate better. The models were compared using their Akaike information criterion (AIC; Akaike, 1974) and their predictions for long ruptures. A summary of the models tested in this study and their AIC are given in Table 1. For strike-slip faults, the models shown in equations (17) and (18) had the best AIC (261.02 and 261.73, respectively):

$$P(\text{directivity}|R, s, \theta, \text{strike} - \text{slip}) = \frac{1}{1 + e^{(\alpha_0 + \alpha_1 R + \alpha_2 s + \alpha_3 \theta)}} \tag{17}$$

and

$$P(\text{directivity}|R, s, \theta, \text{strike} - \text{slip}) = \frac{1}{1 + e^{(\alpha_0 + \alpha_1 R + \alpha_2 \sqrt{s} + \alpha_3 \theta)}}. \tag{18}$$

Most of the strike-slip data used in this study has the length of rupture between the epicenter and the site (s) below 100 km (only six strike-slip data points had $s \geq 100$ km). The predictions from equations (17) and (18) are similar for the smaller fault ruptures, as shown in Figure 11a,c. The predictions for larger ruptures (larger s values) are extrapolations, as we do not have much data to properly model this scenario. We choose equation (18) over (17) as equation (17) predicts high probability of pulse occurrence at large distances from the fault, which is not consistent with predictions from other directivity models developed using theoretically motivated predictors (e.g., Spudich *et al.*, 2014). For

Table 1
Details of Logistic Regression Models Tested to Predict the Probability of Directivity Pulse at a Site

Parameters	Fault Type	Akaike Information Criterion (AIC)
R, s	Strike slip	265.03
R, $\sqrt{s}$	Strike slip	263.32
R, $\ln(s)$	Strike slip	264.33
R, s, θ	Strike slip	261.02
R, $\sqrt{s}$, θ	Strike slip	261.73
R, $\ln(s)$, θ	Strike slip	263.51
R, d, ϕ	Nonstrike slip	613.39
R, d	Nonstrike slip	627.28
R, $\sqrt{d}$, ϕ	Nonstrike slip	607.10
R, d, ϕ, no intercept	Nonstrike slip	612.03
R, $\sqrt{d}$, ϕ, no intercept	Nonstrike slip	614.26

non-strike-slip faults, the model using R, $\sqrt{d}$, and ϕ had the lowest AIC and also extrapolated well, so we selected it.

The models for probability of directivity pulse are given by equations (19) and (20) for strike-slip ruptures and for non-strike-slip ruptures, respectively.

$$P(\text{directivity}|R, s, \theta, \text{strike} - \text{slip}) = \frac{1}{1 + e^{(0.7897+0.1378R-0.3533\sqrt{s}+0.020\theta)}} \quad (19)$$

and

$$P(\text{directivity}|R, d, \phi, \text{non} - \text{strike} - \text{slip}) = \frac{1}{1 + e^{(1.483+0.124R-0.688\sqrt{d}+0.022\phi)}}. \quad (20)$$

Pulse Models

Sometimes the effect of pulse-like ground motions on a structure is assumed to be similar, regardless of the cause of the pulse (e.g., Champion and Liel, 2012). If no distinction is made between directivity and nondirectivity cases, pooling the directivity and nondirectivity pulses together to fit models may be preferred. We refit the relationships to predict the probability of a pulse and its period using the entire dataset of pulse-like ground motions.

The period of a pulse can be predicted using the equations shown below:

$$\ln T_p = -6.55 + 1.12M \quad (21)$$

and

$$\sigma_{\ln T_p} = 0.57. \quad (22)$$

Equations (23) and (24) describe the probability of pulse model for strike-slip and non-strike-slip ruptures:

$$P(\text{pulse}|R, s, \theta, \text{strike} - \text{slip}) = \frac{1}{1 + e^{(0.457+0.126R-0.244\sqrt{s}+0.013\theta)}} \quad (23)$$

and

$$P(\text{pulse}|R, d, \phi, \text{non} - \text{strike} - \text{slip}) = \frac{1}{1 + e^{(0.304+0.072R-0.208\sqrt{d}+0.021\phi)}}. \quad (24)$$

The models presented in this section should only be used when the cause of pulse is not important for the study or if directivity and nondirectivity pulses are pooled together. Further, the mechanisms causing the nondirectivity pulses are not yet well understood, so appropriate caution should be used when using these empirical models to infer the potential occurrence of pulses in future earthquakes

Conclusion

An algorithm to classify multicomponent ground motions as pulse-like or non-pulse-like has been described. The proposed algorithm significantly improves upon the accuracy and the computational cost of previous algorithms. The approach uses wavelet transform results from two orthogonal components of ground motions to identify potential pulses from all orientations. These pulses are then classified using a new criterion developed using support vector machines. The algorithm was used to classify 8611 ground motions in the NGA-West2 database, and 244 pulse-like ground motions were identified. These pulse-like ground motions were then manually filtered to identify 148 ground motions most likely caused by directivity effects.

The classified ground motions were used to fit predictive models for the probability of observing a directivity pulse and its period. Limitations of the dataset were recognized and care was taken to choose functional forms that extrapolate well for cases that were poorly constrained by data. We also recognized that in some engineering applications the distinction between directivity and nondirectivity pulses may not be important. For such cases, fitting the models using a combined dataset of directivity and nondirectivity pulses may be preferred. Alternate models fitted using both directivity and nondirectivity pulses are also provided.

With increasing size of ground-motion databases, there is an ever increasing need for better and faster algorithms to process them. The speed and accuracy of the proposed algorithm allows the processing of larger amounts of data and thus helps further expand our knowledge of the properties of near-fault pulses.

Data and Resources

The earthquake ground-motion records used in the study comes from the Next Generation Attenuation-West2 database (Ancheta *et al.*, 2013). The database is accessible online at http://peer.berkeley.edu/ngawest2/databases/ (last ac-

cessed November 2013). An open source MATLAB implementation of the proposed pulse classification algorithm is available at http://github.com/shreyshahi/PulseClassification (last accessed November 2013).

Acknowledgments

Thanks to Brian Chiou, Iunio Iervolino, and Brendon Bradley for providing valuable feedback on the pulse classifications. This material is based upon work supported by the Pacific Earthquake Engineering Research (PEER) Center and also by the National Science Foundation (NSF) under NSF Grant Number CMMI 0726684. This support is gratefully appreciated. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Contr.* **19,** no. 6, 716–723.

Akkar, S., U. Yazgan, and P. Gulkan (2005). Drift estimates in frame buildings subjected to near-fault ground motions, *J. Struct. Eng.* **131,** no. 7, 1014–1024.

Alavi, B., and H. Krawinkler (2004). Behavior of moment-resisting frame structures subjected to near-fault ground motions, *Earthq. Eng. Struct. Dynam.* **33,** no. 6, 687–706.

Ancheta, T., R. Darragh, J. Stewart, E. Seyhan, W. Silva, B. Chiou, K. Wooddell, R. Graves, A. Kottke, D. Boore, T. Kishida, and J. Donahue (2013). PEER NGA-West2 database, *Technical Report 2013/03*, Pacific Earthquake Engineering Research Center, Berkeley, California.

Anderson, J. C., and V. Bertero (1987). Uncertainties in establishing design earthquakes, *J. Struct. Eng.* **113,** no. 8, 1709–1724.

Baker, J. W. (2007). Quantitative classification of near-fault ground motions using wavelet analysis, *Bull. Seismol. Soc. Am.* **97,** no. 5, 1486–1501.

Bertero, V., S. Mahin, and R. Herrera (1978). Aseismic design implications of near-fault San Fernando earthquake records, *Earthq. Eng. Struct. Dynam.* **6,** no. 1, 31–42.

Bradley, B. A. (2012). Strong ground motion characteristics observed in the 4 September 2010 Darfield, New Zealand earthquake, *Soil Dynam. Earthq. Eng.* **42,** 32–46.

Bradley, B. A., and M. Cubrinovski (2011). Near-source strong ground motions observed in the 22 February 2011 Christchurch earthquake, *Seismol. Res. Lett.* **82,** no. 6, 853–865.

Bray, J. D., and A. Rodriguez-Marek (2004). Characterization of forward-directivity ground motions in the near-fault region, *Soil Dynam. Earthq. Eng.* **24,** no. 11, 815–828.

Champion, C., and A. Liel (2012). The effect of near-fault directivity on building seismic collapse risk, *Earthq. Eng. Struct. Dynam.* **41,** no. 10, 1391–1409.

Chioccarelli, E., and I. Iervolino (2010). Near-source seismic demand and pulse-like records: A discussion for L'Aquila earthquake, *Earthq. Eng. Struct. Dynam.* **39,** no. 9, 1039–1062.

Chiou, B., R. Darragh, N. Gregor, and W. Silva (2008). NGA project strong-motion database, *Earthq. Spectra* **24,** no. 1, 23–44.

Cortes, C., and V. Vapnik (1995). Support-vector networks, *Mach. Learn.* **20,** no. 3, 273–297.

Hall, J. F., T. H. Heaton, M. W. Halling, and D. J. Wald (1995). Near-source ground motion and its effects on flexible buildings, *Earthq. Spectra* **11,** no. 4, 569–605.

Hastie, T., R. Tibshirani, and J. H. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, New York.

Housner, G., and D. Hudson (1958). The Port Hueneme earthquake of March 18, 1957, *Bull. Seismol. Soc. Am.* **48,** no. 2, 163–168.

Howard, J. K., C. A. Tracy, and R. G. Burns (2005). Comparing observed and predicted directivity in near-source ground motion, *Earthq. Spectra* **21,** no. 4, 1063–1092.

Iervolino, I., and C. A. Cornell (2008). Probability of occurrence of velocity pulses in near-source ground motions, *Bull. Seismol. Soc. Am.* **98,** no. 5, 2262–2277.

Iwan, W. (1997). Drift spectrum: Measure of demand for earthquake ground motions, *J. Struct. Eng.* **123,** no. 4, 397–404.

Luco, N., and C. A. Cornell (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, *Earthq. Spectra* **23,** no. 2, 357–392.

Makris, N., and C. J. Black (2004). Dimensional analysis of bilinear oscillators under pulse-type excitations, *J. Eng. Mech.* **130,** no. 9, 1019–1031.

Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing*, Vol. 2, Academic Press, San Diego, California.

Mavroeidis, G. P., and A. S. Papageorgiou (2002). Near-source strong ground motion: Characterizations and design issues, *U.S. National Conference on Earthquake Engineering*, Boston, Massachusetts, 21–25 July 2002, 12 pp.

Mavroeidis, G. P., and A. S. Papageorgiou (2003). A mathematical representation of near-fault ground motions, *Bull. Seismol. Soc. Am.* **93,** no. 3, 1099–1131.

Mavroeidis, G. P., G. Dong, and A. S. Papageorgiou (2004). Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems, *Earthq. Eng. Struct. Dynam.* **33,** no. 9, 1023–1049.

Menun, C., and Q. Fu (2002). An analytical model for near-fault ground motions and the response of SDOF systems, in *Proceedings, 7th U.S. National Conference on Earthquake Engineering*, Boston, Massachusetts, 21–25 July 2002, 10 pp.

Shahi, S. K., and J. W. Baker (2011a). An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis, *Bull. Seismol. Soc. Am.* **101,** no. 2, 742–755.

Shahi, S. K., and J. W. Baker (2011b). Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period, in *11th International Conference on Applications of Statistics and Probability in Civil Engineering*, Zurich, Switzerland, 1–4 August 2011, 8 pp.

Somerville, P. G. (2003). Magnitude scaling of the near fault rupture directivity pulse, *Phys. Earth Planet. In.* **137,** nos. 1/4, 201–212.

Spudich, P., B. Rowshandel, S. K. Shahi, J. W. Baker, and B. S. J. Chiou (2014). Comparison of NGA-West2 Directivity Models, *Earthq. Spectra*, doi: http://dx.doi.org/10.1193/080313EQS222M.

Department of Civil and Environmental Engineering
200 Patton Hall
Virginia Tech.
Blacksburg, Virginia 24060
shrey@vt.edu
(S.K.S.)

Department of Civil and Environmental Engineering
Stanford University
Stanford, California 94305
bakerjw@stanford.edu
(J.W.B.)

Manuscript received 22 July 2013;
Published Online 16 September 2014