Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period.

Shrey K. Shahi & Jack W. Baker
Department of Civil and Environmental Engineering
Stanford University, Stanford, CA, USA

ABSTRACT: Near-fault earthquake ground motions containing large velocity pulses are known to cause severe demands on structures and geotechnical systems, but the probability of occurrence of these pulses in future earthquakes is not well understood. Using a database of past ground motions that have been classified as containing or not containing velocity pulses, this paper develops calibrated regression models to predict the occurrence of velocity pulses in future ground motions, as well as the nature of the pulses if they do exist. The regression model selection procedure indicates that useful predictors of pulse occurrence include source-to-site geometry variables such as the length of rupture between the epicenter and location of interest, and the closest distance from the location of interest to the fault rupture. It is observed that the period of any resulting velocity pulse is related primarily to the earthquake magnitude, but other predictive parameters are also considered and discussed. Both empirical regression tests and theoretical seismology explanations are given as to why the chosen predictor variables are important and meaningful. Some comparisons with previous similar models are also presented. The resulting predictive models can be incorporated into probabilistic seismic hazard analysis calculations. These results demonstrate the potential importance of quantitatively considering occurrence of near-fault pulses, and facilitate seismic reliability calculations that explicitly consider near-fault ground motion pulses.

1 INTRODUCTION

Near fault ground motions may sometime contain a strong pulse at the beginning of the velocity time history. This pulse-like feature is primarily caused by forward directivity effects and are observed when the fault ruptures towards the site at a speed close to the propagation velocity of the shear waves (Somerville et al. 1997, Somerville 2003, Spudich and Chiou 2008). These ground motions referred as "pulse-like" in this paper place extreme demands on structures and are known to be the cause of extensive damage in previous earthquakes (e.g., Bertero et al. 1978, Anderson and Bertero 1987, Hall et al. 1995, Iwan 1997, Alavi and Krawinkler 2001, Menun and Fu 2002, Makris and Black 2004, Mavroeidis et al. 2004, Akkar et al. 2005, Luco and Cornell 2007). Pulse-like ground motions have higher elastic spectral acceleration (S_a) compared to ground motions without the pulse-like feature. The current ground motion models which are used to perform probabilistic seismic hazard analysis (PSHA) do not account for the amplification in S_a caused by these ground motions. Thus the PSHA results computed using the current ground motion models results in under-prediction of hazard at near fault sites, where pulse-like ground motion are expected. Along with amplifying S_a, pulses also cause larger inelastic multi degree of freedom (MDOF) response. Traditional intensity measures like S_a at the fundamental period of the structure are inadequate in capturing the larger nonlinear response of MDOF systems excited by pulse-like ground motions (Baker and Cornell 2008), which makes characterizing the risk from pulse-like ground motion difficult. The importance of accounting for the effect of pulse-like ground motion in design code has long been recognized, but the methods to account for pulses used currently are relatively ad-hoc. We need deeper understanding of how these pulse-like ground motions affect both the hazard and the risk before we can properly account for their effect in future design codes.

Predicting the probability of observing pulse-like ground motion at a site is an important step towards accounting for the effects of these ground motion in hazard computation. Modifications for the ground motion models to predict the ground motion intensity measure at sites when pulses are observed have been proposed in the past (e.g., Somerville et al. 1997, Spudich and Chiou 2008, Shahi and Baker 2010). With the knowledge of probability of occurance of pulse

at the site one can combine the prediction from traditional ground motion models and the modified ground motion model to come up with an estimate of hazard at the site. Equation 1 shows one possible way of combining the modified and traditional ground motion model results using the total probability theorem (e.g., Benjamin and Cornell 1970). The term $P(Sa > x|\text{pulse})$ in equation 1 is computed using the results from modified ground motion models for pulse-like ground motions and the term $P(S_a > x|\text{no pulse})$ is computed using the results from a traditional ground motion models which does not account for the effects of pulses. Note that S_a depends on the period being considered and $P(S_a > x)$ depends on parameters like magnitude, distance etc., those dependences along with the period under consideration are not explicitly included in equation 1 for brevity.

$$P(S_a > x) = P(S_a > x|\text{pulse}) \cdot P(\text{pulse at site}) + P(S_a > x|\text{no pulse}) \cdot (1 - P(\text{pulse at site})) \quad (1)$$

As discussed above, along with amplifying S_a at moderate to large periods, pulse-like ground motions also cause amplification in response of nonlinear MDOF systems and this amplified response cannot be completely characterized by using traditional intensity measures. It is believed that the response depends on the period of the pulse (T_p), and vector valued intensity measures including T_p can be used to characterize the nonlinear response of MDOF systems. It is also known that pulse in the ground motion amplifies the S_a in a narrow band of period (Somerville 2003, Shahi and Baker 2010) and this narrow band amplification is centered about the period of the pulse T_p. So along with helping in characterizing the response of nonlinear MDOF systems, estimation of T_p can help in characterizing the amplification of S_a.

In this paper we study and develop predictive equations for the probability of observing pulse-like ground motion at a site and the period of the pulse expected at a site. The predictive relationships developed in this paper can be used to improve the understanding of pulse-like ground motions and help the practitioners dealing with problems related to hazard and risk at near fault sites.

2 DATASET USED TO BUILD THE MODELS

The earthquake ground motion recordings in the NGA database (Chiou et al. 2008) were used as the base ground motion library for this study. Each ground motion in the database was classified as pulse-like or non-pulse-like using the classification algorithm proposed in Shahi and Baker (2010). The classification algorithm rotates the ground motion in all orientation and classifies the ground motion in each orientation using the wavelet based classification scheme proposed by Baker (2007). A site is classified as having observed pulse-like ground motion if ground motion in any orientation is classified as pulse-like. The database consists of 3551 ground motions, of which 169 were classified as pulse-like.

3 PROBABILITY OF PULSE

Pulse-like ground motions caused by forward directivity effects are observed at near fault sites, but not all near fault sites experience pulse-like ground motion. This makes it important to estimate the probability of observing a pulse in order to correctly do PSHA calculations for near fault sites. Forward directivity is a physical phenomenon with well known causes, but it is hard to predict the occurance of pulse-like ground motion at a site because of incomplete information about the source, site and the path of wave propagation that cause this phenomenon. Due to this lack of knowledge it is useful to develop a statistical model which agrees with the observations. We follow the approach of Iervolino and Cornell (2008) and model the occurance of pulse by a random variable (I) which takes the value 1 if pulse is observed at the site and 0 if pulse is not observed at the site (these type of variables are also called indicator variables).

3.1 *Logistic Regression*

Generalized linear models (GLMs) are generalization of the ordinary least squares regression and allows modeling variables following any distribution belonging to the exponential family of distribution (e.g., McCullagh and Nelder 1989). A distribution belongs to exponential family if it can be written in the form :

$$f(y) = exp(\eta y - \psi(\eta)) \cdot c(y) \quad (2)$$

where $f(y)$ is the probability density, η is called the natural parameter, y is the sufficient statistics, $\psi(\eta)$ is called the normalizing or the cumulant generating function, $c(y)$ is called the carrier density. Many common probability distributions like normal distribution, poisson distribution, binomial distribution and gamma distribution belong to the exponential family. GLMs uses linear model to predict the natural parameter (η) as shown in equation 3.

$$\eta = X \cdot \beta = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_p \cdot x_p \quad (3)$$

In our case we are interested in modeling the probability of pulse occurance, modeled by a random variable capable of taking the values 0 or 1 as explained above. This random variable can also be thought of as a binomial random variable which represents the number of successes (pulses) in 1 trial. Now the probability density of a binomial random variable is given

by equation 4, which can be re-written in the exponential family form as shown in equation 5.

$$f(n) = \binom{N}{n} p^n (1-p)^{N-n} \tag{4}$$

$$f(n) = exp[(log\frac{p}{1-p}) \cdot n + Nlog(1-p)] \cdot \binom{N}{n} \tag{5}$$

By comparing equations 2 and 5 one can see that the natural parameter (η) for the binomial distribution is $log(\frac{p}{1-p})$, the sufficient statistic is n, the cumulant generating function is $Nlog(1-p)$ and the carrier density is $\binom{N}{n}$. As shown in equation 3, we model the natural parameter as a linear function of predictors as shown in equation 6. Equation 7 shows the function we will use to predict the probability of pulses as a function of various parameters (X in the equation).

$$\eta = log\frac{p}{1-p} = X \cdot \beta \tag{6}$$

$$\implies p = \frac{e^{X\cdot\beta}}{1+e^{X\cdot\beta}} = \frac{1}{1+e^{-X\cdot\beta}} \tag{7}$$

where $X \cdot \beta = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_p \cdot x_p$

3.2 *Fitting of the model*

Iervolino and Cornell (2008) developed a model to predict probability of pulse occurance at a site using a dataset of pulse-like ground motion classified using the pulse classification technique of Baker (2007). Since the Baker (2007) technique classifies pulses only in the fault normal orientation, the Iervolino and Cornell (2008) model predicts the probability of observing a pulse only in the fault normal orientation. The Shahi and Baker (2010) classification technique used here classifies pulses in any arbitrary orientation and thus necessitates building a new model to predict probability of observing a pulse in any orientation at the site.

It is known that the observation of forward directivity effects depends on source-to-site geometry, so we can use this knowledge to narrow our search for predictor variables to different source-to-site geometry parameters. Several source-to-site geometry parameters have been used in past to predict directivity effects at a site. Iervolino and Cornell (2008) used many combination of these parameters to fit logistic regression and compared the Akaike's information criterion (AIC) of different models to select the one with lowest AIC. The AIC is defined by equation 8, where k is the number of parameters in the model and lnL is the log likelihood of the model; the function represents the trade off between model fit and parsimony (Akaike 1974). By increasing the number of parameters in the model one can always improve the fit to data, and thus increase log-likelihood. This improvement in fit may be due to fitting the noise in the data, and is known as over-fitting the data. Models which over-fit the data are good at reproducing the data used for fitting but are bad for prediction. Minimizing the AIC tends to select parsimonious models with sufficiently high log-likelihood and thus avoid the problem of over-fitting.

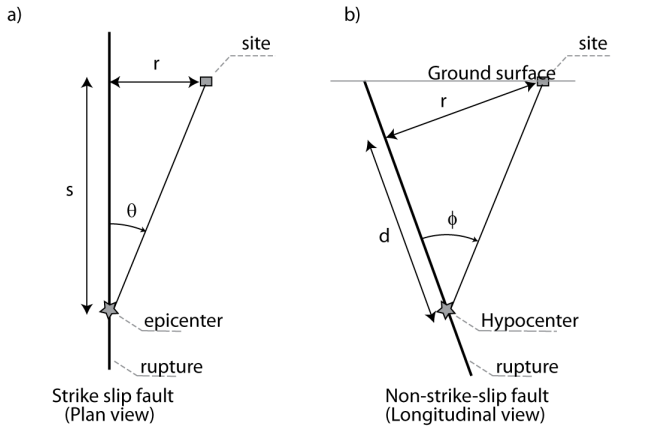

Figure 1: A diagram explaining the source-to-site parameters used to predict probability of pulse for a) Strike-slip fault and b) Non strike-slip fault. (from Shahi and Baker 2010)

$$AIC = 2k - 2lnL \tag{8}$$

Iervolino and Cornell (2008) found that closest distance from the fault (r), amount of rupture between the fault and the site (s) and the angle between the strike of the fault and line joining epicenter and the site (θ) are the best predictors for predicting probability of observing a pulse from a strike slip fault. Analogous predictors r, d and ϕ were found best for non strike slip fault. Figure 1 shows a diagram explaining these parameters. We consider the same parameters identified by Iervolino and Cornell (2008) even for this larger dataset as they seem to define the source-to-site geometry sufficiently well.

The logistic regression models were fitted by maximum likelihood method using the glm function in R (R Development Core Team 2010).

3.2.1 *Model for strike-slip faults*

The dataset contained 680 ground-motions from strike-slip earthquakes for which all the source-to-site geometry parameters were known, out of these 41 were classified as pulse-like. This dataset of 680 ground motions was used to fit the model for strike-slip earthquakes. While fitting the model, θ turned out to be an insignificant predictor when we used r,s and θ as predictors for the logistic regression. Generally insignificance of a predictor in a statistical model is interpreted to mean that the predictor does not have any predictive value. One can see that in some cases when the site lies in the unshaded region shown in figure 2, r and s can define the geometry completely and θ (which is $tan^{-1}(\frac{r}{s})$ in this case) is a redundant parameter not providing any new information. But theta

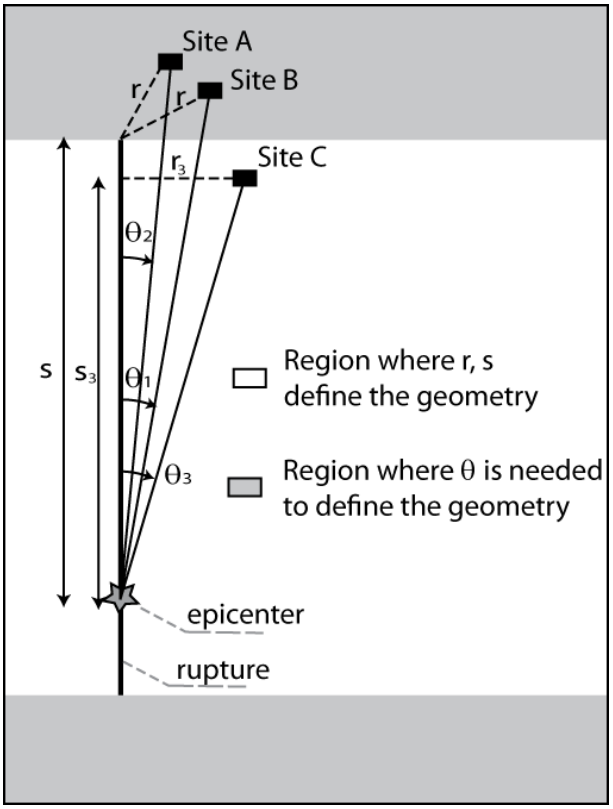

Figure 2: A diagram showing regions where r and s completely defines the source-to-site geometry and θ is redundant and the region where θ is needed along with r and s to completely define the source-to-site geometry. The figure shows two sites (A and B) which has same r and s but different θ to show the importance of θ in the shaded region.

does provide some new information when the site lies in the shaded region of figure 2, in this case two sites with same r and s can have different θ (as shown in the figure). One needs to know the complete source and site geometry information to determine the region in which a site is located but a quick heuristic check is sufficient to estimate the fraction of sites in the shaded area. If one takes the ratio of $R_{epi} \cdot cos(\theta)$ and s, where R_{epi} is the epicentral distance and s and θ are defined above, the ratio will be close to 1 when the site is located within the area where θ is redundant (unshaded region) and will be larger than 1 when the site is located in the area where θ is important (shaded region). Note that this calculation will not give us exact results because real fault geometries are more complex than the simple straight line shown in figure 1. Figure 3 shows the histogram of $\frac{R_{epi} \cdot cos(\theta)}{s}$ for sites where pulses were observed, one can see that few pulse-like ground motions lie in the region where θ is important. So we may not have enough data to constrain the relationship with respect to θ and thus θ ended up as an insignificant parameter in the regression.

As explained above, θ is insignificant because r and s alone can explain the geometry for most of the pulse-like sites. Actually any two parameters from r, s and θ can describe the geometry completely when the site is in the unshaded region shown in figure 2. So we fit logistic regression models using all possible pairs of predictors (r and s, r and θ, s and θ) and select the one with lowest AIC. The result from fitting is shown in table 1, note that the result from using all three parameters is also shown for comparison.

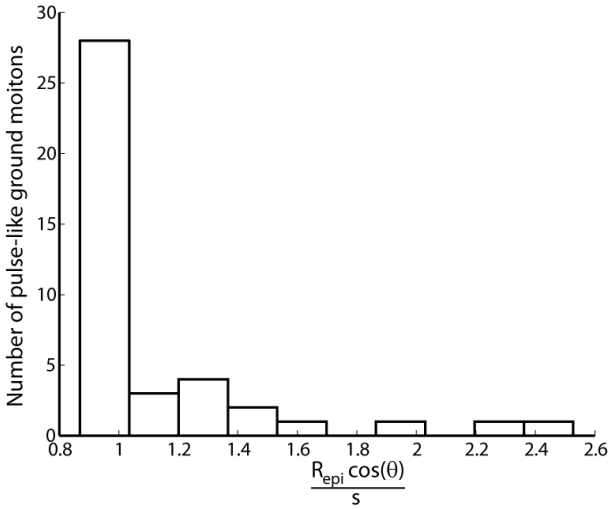

Figure 3: Histogram of $\frac{R_{epi} \cdot cos(\theta)}{s}$

Table 1: AIC for strike-slip models with different predictors.

Parameters used	AIC
R, s	140.77
R, θ	155.95
s, θ	250.61
R, s, θ	141.34

The model with lowest AIC was selected and is shown in equation 9. Here the units of r and s are km. The dataset contained r ranging from 0.07 km to 472 km and s from 0.3 km to 143 km. Since the model only depends on two parameters we can look at the contours of probability in r, s space as shown in figure 4. We have superimposed the data used for regression on figure 4 to visually verify the prediction. One can interpret the ratio of pulses to non pulses in a region as an estimate of probability of observing a pulse at a site with some r and s. Figure 4 shows that the probability contours generally follow the trend shown by the data. One can also test the model by comparing actual pulse observation from a particular earthquake with the prediction from the model, this comparison for Imperial valley fault is shown in figure 5.

$$P(pulse) = \frac{1}{1 + e^{(0.642 + 0.167 \cdot r - 0.075 \cdot s)}} \tag{9}$$

3.2.2 *Model for non-strike-slip faults*

The dataset contained 2169 non-strike-slip ground motions for which all the required parameters were known, out of these 124 were classified as pulse-like. We used the same variables selected by Iervolino and Cornell (2008) to fit the non-strike-slip model. In this case all three parameters are statistically significant. The model can be summarized by equation 10. The unit for r,d is km and ϕ is degrees. The range of r values was 0.3 km to 255 km, d ranged from 0 to 70 km and ϕ ranged from 0 to 90 degrees.

$$P(pulse) = \frac{1}{1 + e^{(0.128 + 0.055 \cdot r - 0.061 \cdot d + 0.036 \cdot \phi)}} \tag{10}$$

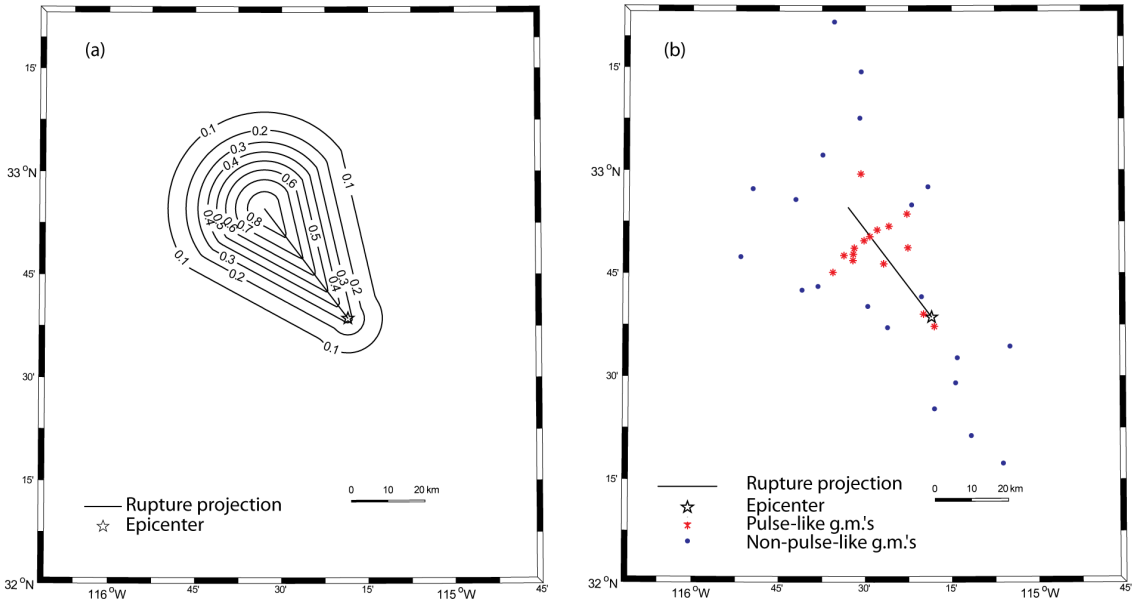

Figure 5: Map of Imperial Valley earthquake showing (a) contours of probability of pulse occurance for the given rupture, (b) sites where pulse-like ground motion was observed (from Shahi and Baker 2010).

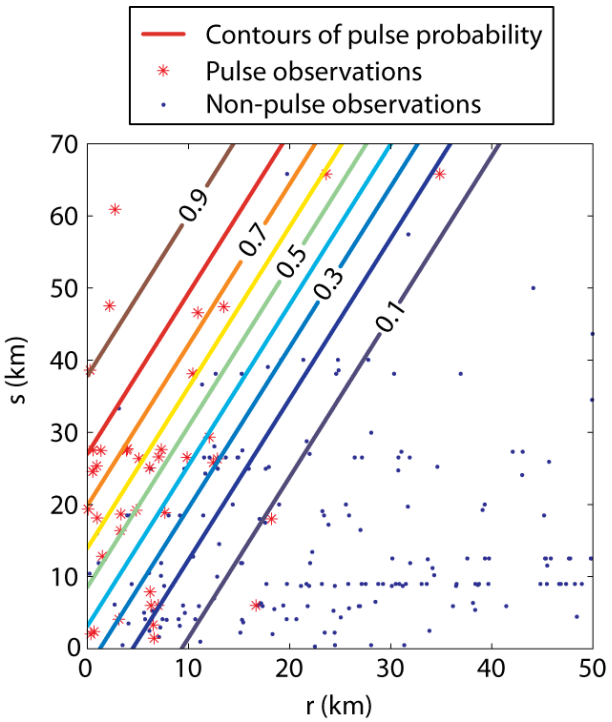

Figure 4: Contours of predicted probability of pulse due to an earthquake from a strike-slip fault at different r and s values.

Since the non-strike-slip model depends on three parameters, a simple visual verification as shown in figure 4 for strike-slip fault is difficult. But we can still check the model by comparing the actual observation of pulse-like and non-pulse-like ground motion with the model prediction. Figure 6 shows such a comparison for the Northridge earthquake. The model predicts high probabilities in regions where pulses were actually observed during the Northridge earthquake.

4 PULSE PERIOD

The period of the pulse-like feature is an important parameter, as the ratio of pulse period and the structural period can be used to determine the structure's response (Anderson and Bertero 1987, Alavi and Krawinkler 2001, Mavroeidis et al. 2004). The amplification of S_a due to presence of pulse also occurs in a small band of period close to the period of the pulse, this makes predicting pulse period an important part of hazard and risk computations. Several models have been proposed in the past for predicting the period of pulse-like ground motion (e.g., Mavroeidis and Papageorgiou 2003, Bray and Rodriguez-Marek 2004, Akkar et al. 2005). We decided to model this relationship again as the classification algorithm of Shahi and Baker (2010) used for this study identifies pulses in different orientation. With the new dataset we have information from many pulses in different orientation at the same site, something which was not available for previous studies.

4.1 *Mixed effects regression*

The dataset used for this study included many pulses from the same site which were identified in different orientations. Pulses at the same site but in different orientations share common source and site effects,

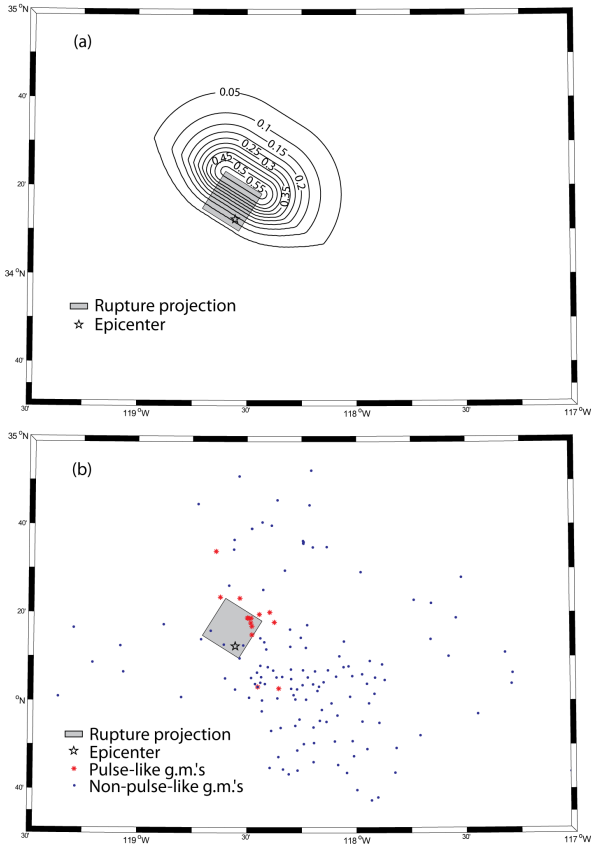

Figure 6: Map of Northridge earthquake showing (a) contours of probability of pulse occurance for the given rupture, (b) sites where pulse-like ground motion was observed (from Shahi and Baker 2010).

and this commonality introduces some correlation between periods of these pulses which must be properly accounted for while fitting the model. Mixed effect model is a popular statistical technique to capture this type of within group correlation. Equation 11 summarizes the mixed effects model used here, the term y_{ij} represents the parameter of interest in j^{th} orientation of i^{th} site, $f(\cdot)$ is the functional form used for regression, η_i is the random effect term, it represents the error common to the i^{th} site, ϵ_{ij} represent the error at i^{th} site in j^{th} orientation. We follow the convention used by Abrahamson and Youngs (1992) and call η_i the inter-event residual and ϵ_{ij} the intra-event residual.

$$y_{ij} = f(\cdot) + \eta_i + \epsilon_{ij} \tag{11}$$

In order to model the predictive relationship for T_p we find a suitable functional form for the regression ($f(\cdot)$ in equation 11) and fit the regression. The search for predictors is discussed below and the fitting was done by maximum likelihood method using the lme4 package (Bates and Maechler 2010) in R (R Development Core Team 2010).

4.2 *Search for predictors*

Most existing predictive equations for T_p model lnT_p as a linear function of magnitude (magnitude refers to the moment magnitude of the earthquake). Seismology theory indicates that the pulse period is related to the rise time of slip on the fault and the logarithm of rise time is proportional to magnitude, this justifies using a linear relationship to predict lnT_p as a linear function of magnitude (Somerville 1998, Somerville et al. 1999). Since the dataset used in our study was much bigger than those in previous studies we were in a position to systematically search for additional predictors. As the space of all possible functional forms is vast and it is impossible to exhaustively search for the best functional form, we reduced the search space by only considering linear combination of the predictors described below.

Magnitude of earthquake M, parameters describing the source to site geometry, (r,s for strike-slip faults and r, d for non-strike-slip faults) along with V_{s30} and their log, square and square root were taken as candidates to be included in the linear mixed effect model (total of $4 \times 4 = 16$ possible predictors each, for both strike-slip and non-strike slip faults). The dataset was divided into two part depending on if the source was strike-slip or not. All the pulse-like ground motions classified by the Shahi and Baker (2010) algorithm except those with $V_{s30} > 2000$ m/s were used for this study.

A stepwise regression scheme was used to select the parameters for the final model. First all mixed effect models with just a single predictor were fitted to predict lnT_p and the predictor with highest log-likelihood was selected. After selecting the first parameter, all possible two parameter models were constructed by combining the first selected parameter with each of the other parameters one by one. The parameter which increased the log likelihood the most was selected as the second parameter. This process was repeated until no parameter made a significant contribution in the model (a 95% level of significance was used as the cutoff). Note that the model with best log-likelihood has the best AIC too.

This scheme identified M, lnV_{s30} and $\sqrt{r}$ as predictors for strike-slip fault and M and r^2 as predictors for non-strike-slip faults.

4.3 *Fitting of the model*

After selecting the parameters using the forward stepwise procedure the following model was fitted to the data.

$$lnTp_{ij} = \alpha + \beta_1 \cdot M + \beta_2 \cdot lnV_{s30} + \beta_3 \cdot \sqrt{r} + \eta_i + \epsilon_{ij} \tag{12}$$

$$lnTp_{ij} = \alpha + \beta_1 \cdot M + \beta_2 \cdot r^2 + \eta_i + \epsilon_{ij} \tag{13}$$

equation 12 shows the model for strike-slip faults and equation 13 shows the model for non-strike-slip faults. In both models η_i is assumed to follow a normal distribution with mean 0 and standard deviation τ, while ϵ_{ij} are assumed to be normally distributed

Table 2: Result of mixed effects regression, for both strike-slip (SS) and non-strike-slip (NSS) faults.

	SS. (eq.12)	NSS (eq.13)	NSS (eq.14)
α	-0.41	-7.84	-7.60
β_1	0.50	1.29	1.25
β_2	-0.37	-5×10^{-5}	-
β_3	0.12	-	-
τ	0.55	0.49	0.50
σ	0.19	0.18	0.18
σ_{total}	0.58	0.52	0.53

with mean 0 and standard deviation σ. so the total standard deviation of predictions from this model is $\sqrt{\tau^2 + \sigma^2}$. The coefficients and the standard deviations for both the models are given in table 2.

The inter-event standard deviation τ is greater than the intra-event standard deviation σ for both strike-slip and non-strike-slip models. This trend is opposite to the typical results of mixed-effect models used for ground motion modeling where σ is greater than τ. This difference is due to the different ways in which the data is grouped. In case of ground motion models all the recordings from a single earthquake form a group (i.e., share a common η_i), while in the model developed here all the pulses from the same site form a group. The grouping scheme used here ensures that members of the same group share the same source, site and path and thus the within-group/intra-event standard deviation (σ) makes a lower contribution to the total standard deviation (σ_{total}) compared to the ground motion models where the groups only share the source.

4.4 *Statistical versus practical significance*

Statistical significance of a parameter in the model suggests that the parameter has some predictive power and including it in the model is generally recommended. But with large datasets, even very small differences that may be practically insignificant become statistically significant. For example consider a group of numbers which are independently and identically sampled from a normal distribution with an unknown mean μ and standard deviation of five. With this information one may want to test the hypothesis that $\mu = 0$. If the dataset consist of 10 samples the absolute value of sample average ($\overline{X}$) needs to be greater than 3.1 (i.e., $|\overline{X}| \geq 3.1$) to conclude that μ is significantly different than 0 at 95% confidence level, but with a larger sample size of say 1000 samples the significance level threshold comes down to 0.31 (i.e., $|\overline{X}| \geq 0.31$), a difference which is statistically significant but may be practically insignificant depending on the problem. The models we developed to predict T_p (equations 12 and 13) consists of parameters whose coefficients are all statistically significant, but in this section we determine whether they have any practical significance.

The models for both strike-slip and non-strike-slip earthquakes use magnitude (M) as a predictor. As discussed above, it is known that lnT_p scales linearly with M and M is the most important predictor for T_p, a fact affirmed by all of the previous predictive models proposed for T_p. The non-strike-slip model uses r^2 along with M as a predictor. Pulse-like ground motions are generally expected only when $r \leq 30km$, over this distance range the term $-5 \times 10^{-5} \cdot r^2$ varies from 0 to -0.045 ($-5 \times 10^{-5} \times 30^2$). Considering that the σ_{total} for the non-strike slip model is 0.52, the contribution r^2 term has in the model is an order of magnitude less than the inherent uncertainty in the model, a fact which suggests practical insignificance. Figure 7 shows prediction of T_p at different distances (r) made using the model using both M and r^2 (equation 13) and prediction from a model for non-strike-slip faults fitted using only M as a predictor. The difference between the predictions shown in figure 7 is small, so the term r^2 can be safely dropped from the model without loosing predictive power. The new model for non-strike slip fault is shown in equation 14. Again the inter-event and intra-event residuals are assumed to follow normal distribution with mean 0 and standard deviation τ and σ respectively. The values of the fitted parameters and standard deviations are shown in table 2.

$$lnTp_{ij} = \alpha + \beta_1 \cdot M + \eta_i + \epsilon_{ij} \tag{14}$$

The strike-slip model uses $\sqrt{r}$ and lnV_{s30} along with M as predictors. One expects pulses with higher period on soil sites when compared with pulses at rock sites due to local site effects. This effect has been discussed by Bray and Rodriguez-Marek (2004) and is also evident from the regression model for strike-slip fault which predicts that T_p decreases with increase in V_{s30}. The effect of V_{s30} term on T_p has some physical explanation and its contribution is non-trivial which justifies it being practically significant. Ignoring the $\sqrt{r}$ term from strike-slip model changes the prediction on the order of the standard deviation of the model ($0.12 \cdot \sqrt{r}$ ranges from 0 to 0.71 when r ranges from 0 to 10 km), this change is large and cannot be ignored so we decided to keep $\sqrt{r}$ in the final model for strike-slip faults. The exact cause of dependence of T_p on r is not clear but T_p may increase with distance due to attenuation of high frequency waves or loss of pulse coherence at larger distances.

The final model for strike-slip faults is given by equation 12 while the model for non strike-slip if given by equation 14.

5 CONCLUSION

Pulse-like ground motions classified by Shahi and Baker (2010) were used to fit predictive relationships for probability of observing pulse-like ground motion at a site and the period of the pulse expected at a site. Statistical techniques were used to find appropriate functional forms for the models and effort was made

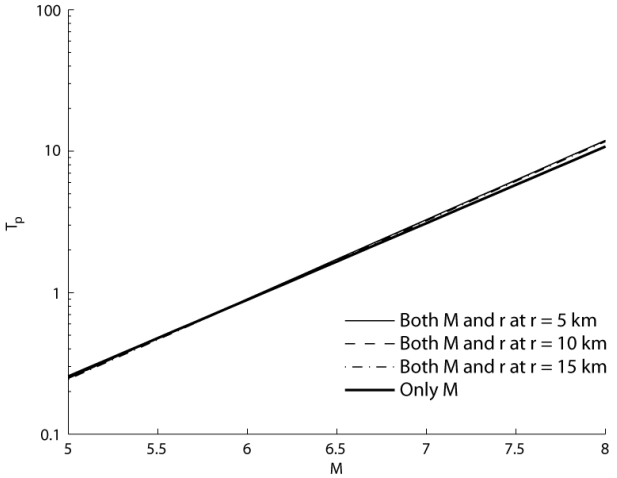

Figure 7: Comparison of predictions from models for non-strike-slip faults using M and r^2 as predictors and only M as predictor at different r values.

to develop parsimonious models which are easy to interpret and thus can lead to better understanding of the overall phenomenon.

Separate relationships were developed for strike-slip and non-strike-slip faults and these relationships were very different from each other. In case of probability of pulse model, the difference was primarily due to the difference in the geometry of the fault ruptures and the different parameters used to define the source-to-site geometry. In case of the pulse period model, different parameters appear to influence the pulse period from strike-slip and non-strike-slip faults. The period of the pulse primarily depend on the magnitude of the earthquake in both cases, but the closest distance to the fault and the V_{s30} also had an influence on the pulse period in case of strike-slip earthquakes. Along with having immediate practical use for hazard and risk estimation, the trends and differences between the predictive equations developed here can be useful in further understanding properties of pulse-like ground motions.

REFERENCES

Abrahamson, N. A. & R. R. Youngs (1992). A stable algorithm for regression analysis using the random effects model. *Bulletin of the Seismological Society of America 82*(1), 505–510.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716 – 723.

Akkar, S., U. Yazgan, & P. Gulkan (2005). Drift estimates in frame buildings subjected to near-fault ground motions. *Journal of Structural Engineering 131*(7), 1014–1024.

Alavi, B. & H. Krawinkler (2001). Effects of near-fault ground motions on frame structures. Technical Report Blume Center Report 138, Stanford, California.

Anderson, J. C. & V. Bertero (1987). Uncertainties in establishing design earthquakes. *Journal of Structural Engineering 113*(8), 1709–1724.

Baker, J. W. (2007). Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the Seismological Society of America 97*(5), 1486–1501.

Baker, J. W. & C. A. Cornell (2008). Vector-valued intensity measures for pulse-like near-fault ground motions. *Engineering Structures 30*(4), 1048–1057.

Bates, D. & M. Maechler (2010). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-33.

Benjamin, J. R. & C. A. Cornell (1970). *Probability, Statistics, and Decision for Civil Engineers*. New York: McGraw-Hill.

Bertero, V., S. Mahin, & R. Herrera (1978). Aseismic design implications of near-fault San Fernando earthquake records. *Earthquake Engineering & Structural Dynamics 6*(1), 31–42.

Bray, J. D. & A. Rodriguez-Marek (2004). Characterization of forward-directivity ground motions in the near-fault region. *Soil Dynamics and Earthquake Engineering 24*(11), 815–828.

Chiou, B., R. Darragh, N. Gregor, & W. Silva (2008). NGA project strong-motion database. *Earthquake Spectra 24*(1), 23–44.

Hall, J. F., T. H. Heaton, M. W. Halling, & D. J. Wald (1995). Near-source ground motion and its effects on flexible buildings. *Earthquake Spectra 11*(4), 569–605.

Iervolino, I. & C. A. Cornell (2008). Probability of occurence of velocity pulses in near-source ground motions. *Bulletin of the Seismological Society of America 98*(5), 2262–2277.

Iwan, W. (1997). Drift spectrum: measure of demand for earthquake ground motions. *Journal of Structural Engineering 123*(4), 397–404.

Luco, N. & C. A. Cornell (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthquake Spectra 23*(2), 357–392.

Makris, N. & C. J. Black (2004). Dimensional analysis of bilinear oscillators under pulse-type excitations. *Journal of Engineering Mechanics 130*(9), 1019–1031.

Mavroeidis, G. P., G. Dong, & A. S. Papageorgiou (2004). Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems. *Earthquake Engineering & Structural Dynamics 33*(9), 1023–1049.

Mavroeidis, G. P. & A. S. Papageorgiou (2003). A mathematical representation of near-fault ground motions. *Bulletin of the Seismological Society of America 93*(3), 1099–1131.

McCullagh, P. & J. A. Nelder (1989). *Generalized Linear Models*. New York: Chapman & Hall.

Menun, C. & 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, MA, pp. 10. Earthquake Engineering Research Institute.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Shahi, S. K. & J. W. Baker (2010). An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America* (under review).

Somerville, P., K. Irikura, R. Graves, S. Sawada, D. Wald, N. Abrahamson, Y. Iwasaki, T. Kagawa, N. Smith, & A. Kowada (1999). Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters 70*, 59–80.

Somerville, P. G. (1998). Development of an improved representation of near fault ground motions. In *SIMP 98, Seminar on Utilization of Strong-Motion Data*, Oakland, CA, pp. 20.

Somerville, P. G. (2003). Magnitude scaling of the near fault rupture directivity pulse. *Physics of the earth and planetary interiors 137*(1), 12.

Somerville, P. G., N. F. Smith, R. W. Graves, & N. A. Abrahamson (1997). Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological Research Letters 68*(1), 199–222.

Spudich, P. & B. S. J. Chiou (2008). Directivity in NGA earthquake ground motions: Analysis using isochrone theory. *Earthquake Spectra 24*(1), 279–298.