CONDITIONAL PULSE PROBABILITY FOR NEAR-SOURCE HAZARD ANALYSIS

I. Iervolino[1] and C.A. Cornell[2]

[1] *Assistant Professor, Dipartimento di Ingegneria Strutturale, Università di Napoli Federico II, Naples, Italy.*
[1] *Affiliate, Department of Civil and Environmental Engineering, Stanford University, CA, USA.*
[2] *Emeritus Professor, Department of Civil and Environmental Engineering, Stanford University, CA, USA.*
Email: iunio.iervolino@unina.it

ABSTRACT:

Near-source ground motion records affected by "directivity" may show unusual features resulting in low frequency pulses in the velocity time-history, especially in the fault-normal component. Although not all near-source recordings show pulses, such an effect is of particular interest for practitioners as it may cause the seismic demand for structures to deviate from that of, so-called, "ordinary" records. Consequently many seismology and earthquake engineering researchers have tried to parameterize the causes and the effects of directivity pulses.
In the framework of the probabilistic seismic assessment of structures in near-source conditions, a quantification of the pulse threat is required. In fact, the recently developed probabilistic seismic hazard analysis for near-source sites requires a probabilistic model for the occurrence of pulses in ground motion. Herein this issue is investigated and models are obtained via logistic regression of a set of pulse-like records from the NGA database. Analyses are limited to ground motions recorded within 30km from the source and to strike-slip events. Occurrence probability of velocity pulses is computed as conditional in respect to those factors considered by seismologists to affect the amplitude of directivity effects.

KEYWORDS: near source, directivity, velocity pulses, PSHA, strike-slip events.

1. INTRODUCTION

A site located close to the source of a seismic event may be in a geometrical configuration with respect to the propagating rupture that favors the constructive interference of the approaching waves resulting in a large velocity pulse. This phenomenon requires the rupture propagating toward the site and the alignment of the site with the slip of the fault. If these two conditions are met, the ground motion at the site may show *forward directivity* effects. In fact, directivity causes, in theory, full-cycle velocity pulses while the *fling step,* which is related to the permanent tectonic deformation at the site, is believed to cause half-cycle pulses (Bolt and Abrahamson, 2003).
Parameters considered to affect the amplitude of the pulse are related to the above-discussed rupture-to-site geometry, while empirical models positively correlating earthquake's magnitude to the period of the pulse $\left(T_p\right)$ have been proposed (e.g., Somerville, 2003); global geophysics-based directivity predictors are also available (e.g., Spudich et al., 2004).
Pulse-type records are of interest to structural engineers because they: (1) may induce unexpected demand in structures having a fundamental period equal to a certain fraction of the pulse period; and (2) such a demand may not be adequately captured by the current, best-practice, ground motion intensity measures such as first-mode spectral acceleration (S_a).
Since not all near-source ground motion records show a pulse in the velocity time-history, it may be argued that near-source records do not always induce non-ordinary seismic demand for structures. Near-source records that do not contain a pulse display virtually the same response behavior as far-field records (Tothong and Cornell, 2006). Therefore the current distinction of *far-field* and *near-source* records may not be the most practical; it should be replaced by *ordinary* versus *pulse-like* ground motions.

It is clear that it is not possible to apply the current earthquake engineering practice to the near-source, and the procedures have to be reviewed and adjusted consistently. A rational approach to the seismic risk analysis requires a probabilistic model for the occurrence of directivity effects in ground motions. The systematic deviations of pulse-like signals with respect to the ordinary imply that, in the probabilistic assessment of structures, a pulse occurrence model is required to incorporate such effects accurately in the probabilistic seismic hazard analysis (PSHA). The phenomena should also be reflected in the record selection as an input for dynamic analysis, since it should be related with the disaggregation of seismic hazard (Cornell, 2004). This issue is briefly reviewed in the following, although for a more comprehensive review the reader should refer to the paper by Tothong et al. (2007).
Assuming that all seismic sources are within 30km from a certain site of interest and given that, as discussed, not all near-source (NS) ground motions are pulse-like, the PSHA, expressed as the mean annual frequency $\left(\lambda_{S_a,NS}\right)$ of S_a exceeding a certain value (x), should be separated into two terms (Eqn. 1.1).

$$\lambda_{S_a,NS}\left(x\right)=\lambda_{S_a,NS\,\&\,pulse}\left(x\right)+\lambda_{S_a,NS\,\&\,no\,pulse}\left(x\right) \tag{1.1}$$

The second term in the right hand side, the near-source non-pulse-like, should be from, say, “ordinary” PSHA, which requires a near-source *attenuation law* computed with records not showing pulses but still coming from short source-to-site distances (i.e., within 30km). The other part should be the near-source term due to pulse-like records. This requires ground motion prediction relationships able to capture the peculiar spectral shape driven by the pulses. These so-called “narrow band attenuation laws” are currently under the attention of seismologists, e.g., the Next Generation Attenuation of Ground Motions (NGA) Project (http://peer.berkeley.edu/products/nga_project.html). In this case the attenuation law will not only depend on magnitude and distance but also on a vector of other parameters $\left(\underline{Z}\right)$, which are assumed to be meaningful to predict directivity effects. The total hazard is the linear combination of the two hazard curves *weighted* by the pulse occurrence probability as in Eqn. 1.2 and Eqn. 1.3, in which a single fault is assumed.

$$\lambda_{S_a,NS\,\&\,pulse}\left(x\right)=\nu\int_m\int_r\int_{\underline{z}}\int_{t_p}P\left[pulse\,|\,m,r,\underline{z}\right]G_{S_a|pulse,M,R,\underline{Z},T_p}\left(x\,|\,m,r,\underline{z},t_p\right)f_{T_p|\underline{Z},M,R}\;f_{\underline{Z}|M,R}\;f_{M,R}\;dt_p\;d\underline{z}\;dm\;dr \tag{1.2}$$

$$\lambda_{S_a,NS\,\&\,no\,pulse}\left(x\right)=\nu\int_m\int_r\int_{\underline{z}}\left(1-P\left[pulse\,|\,m,r,\underline{z}\right]\right)G_{S_a|no\,pulse,M,R}\left(x\,|\,m,r\right)f_{\underline{Z}|M,R}\;f_{M,R}\;d\underline{z}\;dm\;dr \tag{1.3}$$

In Eqn. 1.2 ν is the mean rate of events on the fault, M is the magnitude of the event, and R is the source-to-site distance. dt_p and $d\underline{z}$ are the integration intervals of the variables pulse period, T_p, and $\underline{Z}$, respectively. $G_{S_a|pulse,M,R,\underline{Z},T_p}$ is the complementary cumulative distribution function of S_a conditioned on M, R, $\underline{Z}$, and T_p; $f_{T_p|\underline{Z},M,R}$ is the probability density function (PDF) of T_p given M, R, and $\underline{Z}$. Similarly, $f_{\underline{Z}|M,R}$ is the conditional distribution of $\underline{Z}$ given M and R, while $f_{M,R}$ is the joint PDF of M and R. The same meaning of the symbols applies to Eqn. 1.3.
The conditional probability of having a pulse is needed to evaluate Eqn. 1.2 and Eqn. 1.3. In the following some empirical pulse probability models based on logistic regression, for strike-slip rupture data, are proposed and results discussed.

2. PULSE-LIKE RECORDS DATASET

It is a non straightforward task to ascertain whether a record shows directivity effects, i.e., a pulse in the velocity time history, and its properties such as the period T_p. Many seismologists and other earthquake science experts have engaged in this exercise but no widely accepted method is readily available. The bulk of the difficulties in identifying a pulse in the ground motion are related to the wave propagation effects and to the higher frequency content which may give an unclear picture of the directivity features. A common option is to visually analyze the waveform looking for pulses, but this method requires strong expertise in the field and may be not very efficient for short-period pulses or for small or moderate magnitude events, where the pulse may be lost in the high frequency. Above all, this method does not allow one to investigate large datasets looking for the fraction of signals showing directivity effects.

Baker (2007) analyzed extensively the NGA database and he is the only researcher the authors are aware of who has looked systematically at all records in the database. Therefore, we know which are the pulses, and also which are the non-pulses, which is crucial to develop any pulse occurrence probability model. Baker (2007) developed a method based on wavelets to assign a score, a real number between 0 and 1, to each analyzed record and to determine the pulse period. The larger the score determined the more likely the record is to show a pulse. In this way Baker (2007) has found pulse-like records in both fault-normal and fault-parallel components of the ground motions investigated. Herein only those in the fault-normal component have been considered; in particular those ground motions which have a pulse score larger or equal to 0.85 have been, arbitrarily, counted as pulse-type records. Of these records, 98 are classified as within 30km of the fault by the NGA *Flat-File* (http://peer.berkeley.edu/assets/NGA_Flatfile.xls). Six of them of them do not have a measure of the closest distance to fault rupture but their epicentral distance is within 30km; however, they still have not been included herein because in the NGA they lack information about geometry of the fault/site useful for predicting directivity. This set has also to be cleared of those considered as *late pulses*, i.e., occurring at the end of the records and, therefore, too late to be directivity caused; there are 19 of these records (J.W. Baker, written communication, 2006). The resulting dataset consists of 73 records from 23 events, and do not include *Chi-Chi* related events. Of these records, 34 are from strike-slip (SS) ruptures, which are 12 in number.

This study discusses selected models for the estimation of the pulse occurrence probability in strike-slip earthquakes. As the probability of occurrence is herein based on empirical evidence, given the pulse-like records dataset, also the complementary set of identified non-pulse records is needed. The total database considered in the following is made of SS records within 30km (in terms of closest distance to fault rupture) coming from the NGA catalog and whose characteristics have been determined via the NGA flat-file (accessed August 2006) and related documentation (http://peer.berkeley.edu/nga/NGA_Documentation.xls).

In the flat-file, the total number of SS events matching the selection requirements discussed above, and featuring records within 30km in terms of closest distance to fault rupture, is 22 (this, again, excludes *Chi-Chi* and aftershocks). The number of records from these events is 133, therefore the marginal SS pulse occurrence probability is 34/133 or 26%.

A more detailed presentation of the investigations and analogous models for non-strike-slip events is given in Iervolino and Cornell (2008).

3. DIRECTIVITY EFFECT PREDICTORS AND LOGISTIC REGRESSION

In Somerville et al. (1997), for strike-slip events, the amplitude of spectral modification of ordinary attenuation laws due to directivity in ground motions depends on $X\cos(\theta)$; where $X = s/L$ is the ratio of the distance from the epicenter to the site, measured along the rupture direction, and the fault length; θ is the angle between the fault strike and the path to the site with respect to the rupture (Figure 1 and Table 1).

Other factors which may explain directivity effects are the event's magnitude, which is correlated with the pulse period (Somerville, 2003), and the source-to-site distance. Neither of them appears explicitly in $X\cos(\theta)$, although the rupture length is related to magnitude and source-to-site distance is not independent from the geometrical configuration.

Recently, also the *s*-distance alone has been considered as a meaningful predictor of directivity and it is confirmed in the following. A rough explanation is that for large *s* the chance that the rupture evolves yielding directivity effects increases independently of the fault length. Therefore, the variables considered herein as possible covariates in the model to predict pulse occurrence are: (1) the closest distance to fault rupture (*R*); (2) the event's magnitude (*M*); (3) the length ratio *X*; (4) the θ angle; (5) the *s*-distance; and the Somerville et al. (1997) parameter $X\cos(\theta)$.

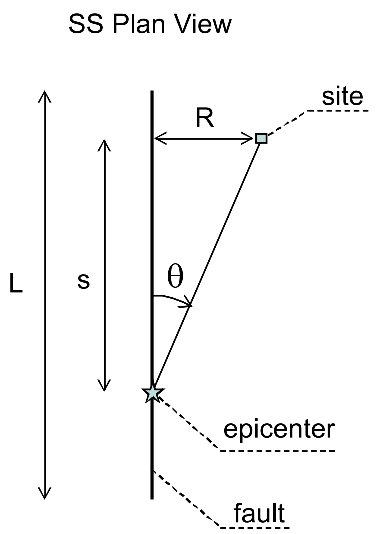

Figure 1. Near-source scheme.

Table 1. Near-source parameters in strike-slip ruptures.

Parameter	*Description*
R [km]	Closest distance of the site to fault rupture
s [km]	Distance to the site measured along the rupture
θ [deg]	Angle between the rupture and the site
L [km]	Length of the fault

To link the probability of pulse occurrence to the candidates to be directivity predictors, the *logistic* regression is used (Agresti, 2002). In fact, the occurrence of a pulse in a near-source ground motion may be represented as an indicator variable (*I*) which can assume the two values: 1 if there is a pulse in the record, or 0 if the record doesn't show a pulse. The probability of the occurrence of the pulse, is $p = P[I=1]$; the probability of the record not showing a pulse is $1-p = P[I=0]$.

Logistic regression assumes the log of the "odds ratio" to be a linear function of the predictor variable. This means $log\left(p/(1-p)\right) = \alpha + \beta z$; where *p* is the occurrence probability given *z* and $\{\alpha, \beta\}$ are the coefficients to be determined. In general, multivariate logistic regressions are of the type in Eqn. 3.1, which is written in the case of *k* predictors.

$$log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_k \tag{3.1}$$

Univariate and multivariate regression models for pulse occurrence have been investigated in this exploratory study using the *glmfit* tool, which serves to fit *generalized linear models*, in MATHWORKS – MATLAB®. Unfortunately it is not easy to determine the prediction power and to compare logistic models. One qualitative way to determine whether the logistic distribution is a good approximation of the data is to group the sample in *z*-bins and then to estimate *p* as the ratio of occurrences over the number of data within the bin; plotting these frequencies versus the fit gives a picture of the adequacy of the model. Furthermore, there is no widely-accepted direct analog to R^2 as defined for ordinary least-square regressions. Nonetheless, several logistic R^2 measures have been proposed, all of which should be reported as approximations of R^2, not as actual percents of variance explained by the model, but rather attempts to measure strength of association. Herein two of them are considered: R_E^2 (Efron, 1978) and R_{MF}^2 (McFadden, 1974).

4. RESULTS AND DISCUSSION

4.1. Univariate logistic regressions

Simple (univariate) logistic regression allows one to determine the probability trends with respect to those parameters discussed above. In Figure 2 the continuous estimated conditional occurrence probability is plotted versus each of the predictors along with the pulse observations (as coded by the values of the indicator variable defined).

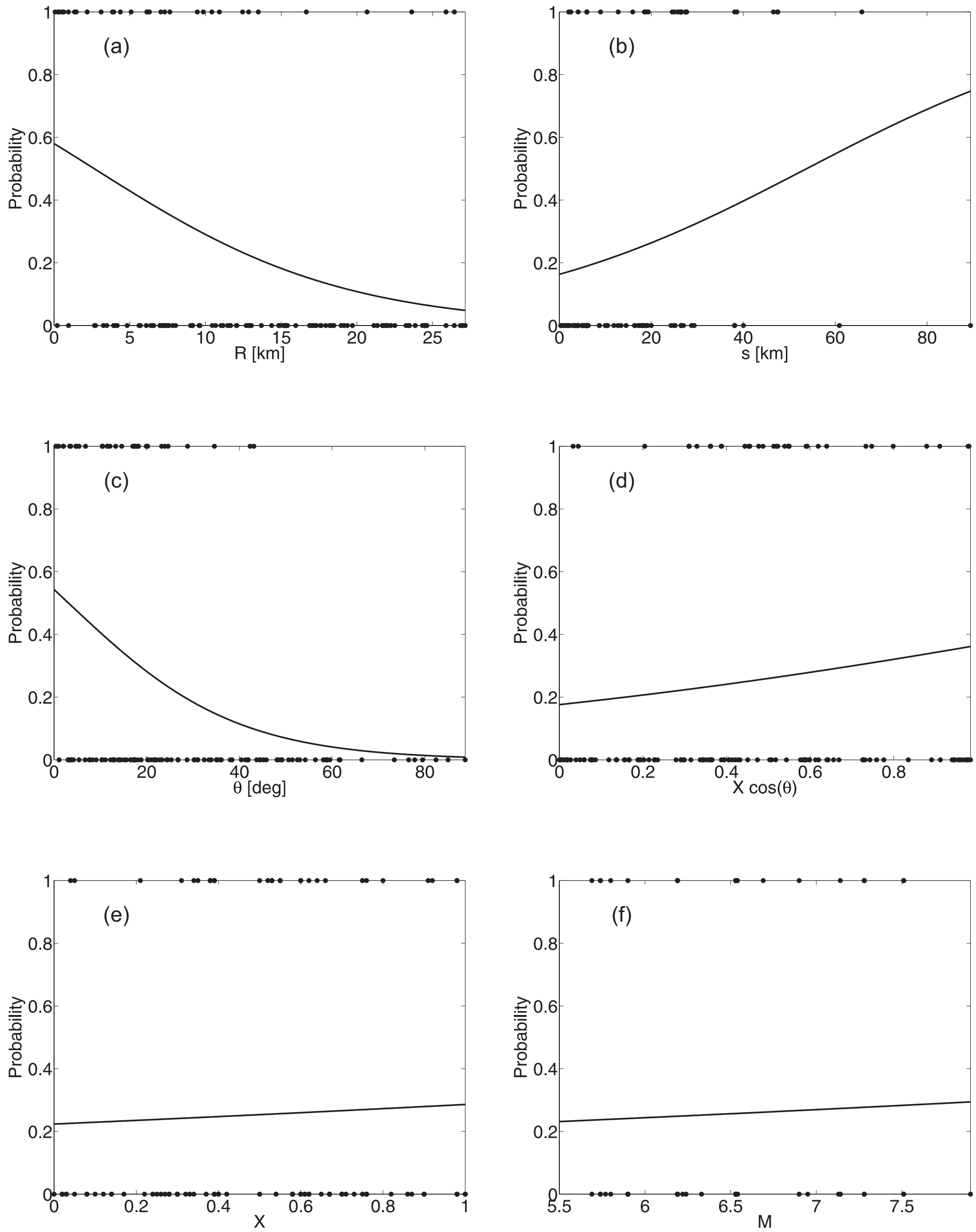

Figure 2. Univariate logistic regressions.

The covariates showing the largest explanatory power with respect to the pulse occurrence are the closest distance to fault rupture, the θ angle, and the distance measured along the rupture. In panel (a) it is possible to see the clear decreasing trend of pulse occurrence probability with *R*. The occurrence probability at zero distance is 0.58 and drops to 0.03 at 30km. The *s*-distance, panel (b), also shows some predictive power with an expected trend. The plot refers to the 0km-90km range which is the data availability interval. Between this limits the probability of observing a pulse increases from 0.16 to 0.75; however, the actual upper bound for the applicability of the model should be around 40km where the occurrence has 0.4 probability. The θ angle, panel (c), seems also significant for pulse occurrence, estimating an occurrence probability of 0.54 for a site which is sitting on the line of the rupture (in the most favorable condition to see a pulse) and drops to 0.01 for a site at orthogonally placed with respect to the fault in a way that the rupture proceeds beyond it.

Other candidates to be directivity-related parameters seem to have all small, if any, predictive power with respect to pulse occurrence probability. It can be observed in panels (d), (e), and (f) that the occurrence probability does not vary much in the data intervals for the length ratio *X* and magnitude; a slightly greater trend is shown for $X\cos(\theta)$. (Note that given these results, magnitude can, in principle, be dropped from the conditional pulse occurrence probability in Eqn. 1.2 and Eqn. 1.3.)
The coefficients, $\{\alpha,\beta\}$ for the univariate regressions shown in Figure 2, along with pseudo-R^2 measures, are given in Table 2; their values confirm the discussion given commenting the plots.

Table 2. Univariate logistic regression coefficients and scores.

Covariate	α	β	R_E^2	R_{MF}^2
R [km]	0.32347	-0.12169	0.16051	0.12467
s [km]	-1.6349	0.030403	0.040238	0.034712
θ [deg]	0.17263	-0.05545	0.12381	0.12873
M	-1.9439	0.13523	0.000848	0.000775
X	-1.2452	0.32971	0.001278	0.001522
$X\cos(\theta)$	-1.5435	0.99033	0.011531	0.013597

4.2. Multivariate logistic regressions

When constructing a multivariate regression model, choosing the appropriate covariates and terms is not a straightforward task. Including only few covariates may lead to a lower prediction power than including many terms and interactions. On the other hand a complex model, although better representing the input sample, may be less manageable and may lose in generality. Several approaches exist to build up multiple ordinary regression models once a basic set of covariates has been established, such as *forward selection* and *backward elimination* (Agresti, 2002), which also apply to the logistic case. A full quadratic model, which would include all the five basic predictor candidates, $\{R,s,\theta,X,M\}$, should have 20 terms because of the interactions and the squared variables. Because the number of pulse-like records in the dataset case is only slightly larger (34) than the number of terms in the model, there is significant risk of data over-fitting. Therefore, based on the results of the previous section, the variables showing the lowest marginal predictive power, $\{X,M\}$, have been excluded from candidates for the multivariate regressions. This does not imply that magnitude or fault length do not affect hazard analysis in the near source, but rather that they are more related to the amplitude of the directivity effects once the geometry has been determined.
For the $\{R,s,\theta\}$ set of covariates several multiple regression models have been fitted. They are linear and quadratic. Among those computed, the linear combination of the covariates is reported here, Eqn. 4.1, as it is the best performing model which includes all three covariates known given the rupture and the site. Coefficients determined for this model are reported in Table 3 along with the *pseudo-R^2* measures.

$$P\left[pulse\,|\,R,s,\theta\right]=\frac{e^{\alpha+\beta_1\cdot R+\beta_2\cdot s+\beta_3\cdot\theta}}{1+e^{\alpha+\beta_1\cdot R+\beta_2\cdot s+\beta_3\cdot\theta}} \tag{4.1}$$

Table 3. Coefficients and *pseudo-R^2* for the proposed multivariate model.

Covariates	α	β_1	β_2	β_3	R_{MF}^2	R_E^2
$\{R,s,\theta\}$	0.85925	-0.11137	0.018704	-0.04441	0.2708	0.22511

Because for the rupture's schematic geometry, θ is known given *R* and *s*, it is possible to represent this model in a three-dimensional plot, which is given in Figure 3 up to the bounds of the covariates determined by data availability. Assuming the epicenter of the event at the origin, and the rupture direction being coincident with

the *s*-axis, any point in the $\{R,s\}$ plane may be considered as a site for which the θ angle is also known $\left(\theta = \arctan\left(R/s\right)\right)$. Therefore, for that site, the pulse occurrence probability predicted according to the proposed model may be read on the vertical axis.
From the plot it is possible to observe the expected marginal trends of pulse occurrence with respect to the three covariates. The probability generally decreases with *R* and increases with *s*-distance.

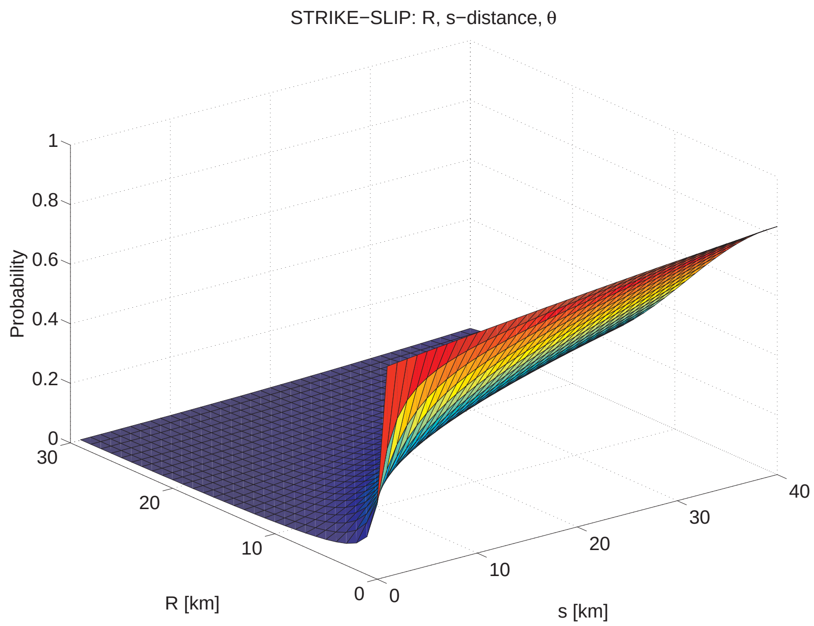

Figure 3. Selected multivariate pulse occurrence model.

5. CONCLUSIONS

Near-source issues in earthquake engineering are of concern for nonlinear assessment of structures. Due to the peculiar spectral features that ground motion may experience, the PSHA at the site requires appropriate procedures. Then, record selection for seismic structural assessment cannot follow the current far-field practice and should reflect the near-source pulse and non-pulse hazards. Both these issues call for a pulse occurrence probabilistic model. The study presented attempted to build and propose such models empirically for strike-slip events. The fundamentals of the analyses are related to the choice of the covariates (i.e., independent variables) and the determination of the response sample (i.e., the dataset).
Since PSHA refers to a specific site, the occurrence of pulses should be conditional on some parameters, available for the source-to-site configuration, which are believed to predict directivity effects. To this aim, covariates were chosen among factors identifying near-source conditions and, according to seismologists, affecting the amplitude of pulses specifically for strike-slip ruptures.
As directivity effects are generally observed most strongly in the velocity signals recorded in the direction orthogonal to the strike, the empirical dataset was made up of fault-normal rotated records. All the records from strike-slip ruptures within 30km in terms of closest distance to fault rupture (arbitrarily considered as a practical upper bound for near-source conditions) reported by the Next Generation Attenuation Project database were used (except the *Chi-Chi* related records).
Pulse-like velocity ground motions have been identified by the rational method, based on wavelets, proposed by Baker (2007). Some judgment was used to identify, and classify as non-pulse-like, those velocity recordings showing multiple low frequency cycles which are likely not related to directivity. The dataset was purged of those records for which the information regarding the covariates was not available. This alone allows one to evaluate the marginal pulse occurrence frequency, which is about 26% for the SS sample.
Simple and multiple logistic regression models have been investigated to associate pulse occurrence in the

dataset to the covariates. Pulse occurrence probability has shown, as expected, significant dependence on distance to the rupture, *R*, along the rupture, *s*, and also on the θ angle (which, in principle, is a deterministic nonlinear function of the other two parameters). Less explanatory power, if at all, for pulse-like records occurrence, was found for the event's magnitude and other pulse amplitude-related factors. Although these results may sound intuitively unexpected, it has to be recalled here that this study dealt with pulse occurrence probability alone, rather than on the prediction of the amplitude of such pulses, for which the excluded parameters do play a role.

Multivariate logistic regression models were also investigated. To avoid data over-fitting only the covariates shown to be the best predictors in the simple univariate models were considered in the multiple regressions, which have been investigated up to complete quadratic functional forms. The proposed model is the linear combination of the geometrical predictors, as there is no empirical support to use models which include interaction or quadratic terms.

Information about pulse occurrence non-strike-slip events may be found in another study of the same authors; although it may be anticipated that general conclusions found for the SS case still hold for non-strike-slip ruptures.

ACKNOWLEDGMENTS

This paper and the one referenced as Iervolino and Cornell (2008) were completed by the authors in 2007, only a few weeks before Professor C. Allin Cornell passed away. Iunio Iervolino would like to dedicate them to his memory. Allin, we miss you!

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis*, Second Edition, Wiley and Sons, NY.

Baker, J.W. (2007). Quantitative classification of near-fault ground motions using wavelet analysis, *Bull. Seism. Soc. Am.* **97,** 1486–1501.

Bolt, B.A., and Abrahamson, N.A. (2003). Estimation of strong seismic ground motions, in *International Handbook of Earthquake and Engineering Seismology*, W.H.K. Lee, H. Kanamori, P.C. Jennings, and C. Kisslinger (Editors), Academic Press, NY, Part B 983–1001.

Cornell, C.A. (2004). Hazard, ground motions and probabilistic assessment for PBSD, in *Performance based seismic design concepts and implementation*. PEER Report 2004/05, Pacific Earthquake Engineering Research Center, University of California Berkeley, CA, 39–52.

Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association,* **73**, 113–121.

Iervolino, I., and Cornell, C.A. (2008). Probability of occurrence of velocity pulses in near-source ground motions. *Bull. Seism. Soc. Am.* **98** (*in press*).

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, in *Frontiers in Economics*, P. Zarembka, (Editor), Academic Press, 105–142.

Somerville, P.G. (2003). Magnitude scaling of the near fault rupture directivity pulse, *Physics of the earth and planetary interiors*, **137**, 201–212.

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

Spudich, P., Chiou, B.S.J., Graves, R., Collins, N., and Somerville, P.G. (2004). *A formulation of directivity for earthquake sources using isochrone theory*, U.S. Geol. Surv. Open-File Rept., 2004-1268.

Tothong, P., and Cornell, C.A. (2006). *Probabilistic seismic demand analysis using advanced ground motion intensity measures, attenuation relationships, and near-fault effects*, PEER Report 2006/11, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Tothong, P., Cornell, C.A., and Baker, J.W. (2007). Explicit directivity-pulse inclusion in probabilistic seismic hazard analysis. *Earthquake Spectra*, **23**, 867–891.