ORIGINAL PAPER

Determination of near-fault impulsive signals with multivariate naïve Bayes method

Deniz Ertuncay[1] [ORCID] **· Giovanni Costa[1]**

Abstract

Near-fault ground motions may contain impulse behavior on velocity records. To calculate the probability of occurrence of the impulsive signals, a large dataset is collected from various national data providers and strong motion databases. The dataset has a large number of parameters which carry information on the earthquake physics, ruptured faults, ground motion parameters, distance between the station and several parts of the ruptured fault. Relation between the parameters and impulsive signals is calculated. It is found that fault type, moment magnitude, distance and azimuth between a site of interest and the surface projection of the ruptured fault are correlated with the impulsiveness of the signals. Separate models are created for strike-slip faults and non-strike-slip faults by using multivariate naïve Bayes classifier method. Naïve Bayes classifier allows us to have the probability of observing impulsive signals. The models have comparable accuracy rates, and they are more consistent on different fault types with respect to previous studies.

Keywords Near-fault ground motion · Velocity pulse · Impulsive signal · Probability of pulse occurrence

1 Introduction

Areas where the distance between the site of interest and the fault line that ruptures during an earthquake are less than the length of the ruptured fault are called near-fault areas or near-fault regions (Lay and Wallace 1995). The velocity-time history of the seismic stations in near-fault regions may have large, pulse-shaped motions. These velocity waveforms have large, long period signals. A pulse-shaped signal is the representation of the effect of almost all seismic energy from the fault arriving within a short period of time. In this case, ground motion prediction equations (GMPEs) may under-determine the amplitudes of these signals. It is due to unexpectedly high amplitudes caused by directivity and fling step effects. In such cases, structures may have to deal with loads higher than predicted in building codes (Li et al. 2020).

✉ Deniz Ertuncay
  dertuncay@units.it

[1] Department of Mathematics and Geosciences, University of Trieste, via E.Weiss 4, 34128 Trieste, Italy

A velocity waveform, which contains the effect of this natural phenomena, is called an impulsive signal, a velocity pulse or a pulse shaped signal. Thanks to the denser installation of seismic stations around near-fault regions in recent years, waveforms with impulsive features are now being recorded. These signals mostly result from large magnitude earthquakes. Data centers around the world are now monitoring near-fault regions and have started to collect broadband and strong motion data.

Starting from the late 1990s, several different indicators of impulsive signals have been identified (Grimaz and Malisan 2014). Impulsive signals may have long periods depending on the magnitude of the earthquake. Such signals also have large amplitudes in velocity time history (Somerville et al. 1997). It means that energy of the earthquake is concentrated into one or few pulses. In impulsive signals, PGV/PGA ratio is higher with respect to non-impulsive signal's (Bray and Rodriguez-Marek 2004; Loh et al. 2002). They create unexpectedly high amplitudes on long periods (> 3 s) in velocity response spectra (Yang and Wang 2012).

In order to identify the impulsive signals, various methods are developed (Baker 2007; Chang et al. 2016; Ertuncay and Costa 2019; Kardoutsou et al. 2017; Mena and Mai 2011; Shahi and Baker 2014; Zhai et al. 2018). These methods are focused to analyze seismic signals by using the indicators mentioned above. Signals are analyzed both in time and frequency domain as acceleration, velocity and/or displacement signals. Shahi and Baker (2014) is one of the most used algorithms for identification of impulsive signals. The algorithm can differentiate early and late arrival pulses by analyzing the arrival of PGV. Early arrivals of PGV generally indicate directivity effects. To detect the impulsive signal, it requires horizontal components of a seismic station.

Impulsive signals can create high seismic demand on high buildings (Li et al. 2020). Spectral ratios can be locally amplified in the region where structural fundamental period is closer to the pulse period (Shahi and Baker 2011). Structure will be loaded with considerable seismic energy in few pulses in the higher modes (Kalkan and Kunnath 2006). They also create different inelastic seismic demands with respect to ordinary strong motion signal (Alavi and Krawinkler 2004; Hall et al. 1995; Iervolino et al. 2012, 2017). Impulsive signals can be destructive to various types of infrastructure, such as buildings having a behavior describable as single or multi degree of freedom systems (Guo et al. 2018), seismically isolated structures (Mazza 2018), and bridges (Antonellis and Panagiotou 2013).

The spatial distributions of impulsive signals are generalized by analyzing the physical features of the earthquakes that have created impulsive signals, spatial information, and local soil conditions of the stations which have recorded impulsive waveforms. It has been observed that impulsive signals are more likely to occur when any of the following conditions are satisfied:

1. Forward directivity (Somerville et al. 1997; Somerville 2003, 2005; Spudich and Chiou 2008),
2. Fling step effect,
3. Similar rupture velocity and shear-wave velocity of the bedrock of the site of interest.

Previous models are developed to assess the probability of pulse occurrence. Parameters are identified to explain the relation between the fault plane and site of interest and impulsiveness (Chioccarelli and Iervolino 2013; Iervolino and Cornell 2008; Shahi and Baker 2014). Probability of occurrence of impulsive signals may be implemented to probabilistic seismic hazard studies (Tothong et al. 2007). Previous methods require intensive information about the fault

and the earthquake. Precise information such as dip angle, initiation point of the fault rupture, hypocentral depth of the earthquake, and the depth of the shallowest part of the ruptured fault have to be given to the model in order to calculate the probability distributions of occurrence. However, it is impossible to provide some of them before an earthquake such as the initiation point. Furthermore, it is hard to determine these information even after an earthquake, due to scarcity of the stations or it is hard to make a site investigation after an event.

We investigate various parameters that carry information about the site of interest, earthquake and the position between the fault and the site of interest and select the most meaningful parameters by calculating the correlation coefficient. To create the model, multivariate naïve Bayes classifier method is used. Bayesian approach provides us probabilities of observing impulsive signals with given input parameters. Models are developed for strike-slip and non-strike-slip faults. We investigate the probability of occurrence for non-strike-slip faults for two different regions which are hanging wall and footwall. Application of the models for Imperial Valley and Chi–Chi earthquakes are presented.

2 Data

To create a generalized model a large dataset is collected. Crustal earthquakes are collected from various data centers. Strong-motion and broad-band stations are collected. Earthquakes with moment magnitude (M_w) bigger than 5.5 with hypocentral depth smaller than 55 km are chosen. Stations with epicentral distance less than 150 km are used. Data are collected from NGA-West 2 (Bozorgnia et al. 2014) database and data centers of National Seismological Center of Chile (Universidad De Chile 2013), The Italian Accelerometric Archive (Luzi et al. 2016; Pacor et al. 2011), Kyoshin Net and Kiban-Kyoshin Net, National Autonomous University of Mexico (SSN 2017), GeoNet (Van Houtte et al. 2017), Broadband Array in Taiwan for Seismology (Institute of Earth Sciences 1996), Disaster and Emergency Management Presidency of Turkey (AFAD 2019) and Incorporated Research Institutions for Seismology (IRIS).

We created a database with 25,376 stations. Signals that are recorded at the earthquakes that do not produce any impulsive signals are excluded. Algorithm of Shahi and Baker (2014) is used for the determination of the pulse-shaped signals. East–West and North–South component signals are required to identify impulsive signals. Among the impulsive signal-produced earthquakes, 206 of them are identified as impulsive and 5175 are identified as non-impulsive.

3 Method

To calculate the probability of occurrence of impulsive signals, the parameters that are correlated with the impulsive signals are determined, and then, they are analyzed with multivariate naïve Bayes approach (Sect. 3.1).

3.1 Multivariate naïve Bayes classifier

Bayes' theorem describes the probability of an event with a given prior information, which is the information about the event without having the data of that event. It can be formulated as in Eq. 1.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

A and B are events, and P states the probability. $P(B)$ should be bigger than zero. $P(A|B)$ and $P(B|A)$ are the conditional probabilities, which are the likelihood of events A and B occurring for a given true B and A, respectively. $P(A)$ and $P(B)$ are independent from each other. It means that event A has no effect of event B.

In Bayes classifier A will be the classes, which are impulsive and non-impulsive. B is the feature that is going to be used to determine the impulsive and non-impulsive signals. A and B will get indices and become A_j, and B_i. Binary classification will be applied on the data and A_0 and A_1 will be the probabilities of non-impulsive and impulsive signals, respectively. B_i will have the dataset input parameters. They are also called predictor features. $P(B|A)$ can be rewritten in vector form, $P(B_i|A_j)$, as:

$$P(B_i|A_j) = P(B_1|A_j) \times P(B_2|A_j) \times \cdots \times P(B_n|A_j) \tag{2}$$

The probability of class A_j with the set of values at B_i, is explained as in Eq. 2. Equation 1 can be modified with the information in Eq. 2 and rewritten as in Eq. 3.

$$P(A_j|B_{1,\ldots,n}) = \frac{P(B_1|A_j)P(B_2|A_j) \ldots P(B_n|A_j)P(A_j)}{P(B_1)P(B_2)P(B_3) \ldots P(B_n)} \tag{3}$$

Equation 3 can be put into words as, the probability of the class (A_j), for a given predictor feature combination ($B_{1,\ldots,n}$), is as a function of the probability of the predictor feature combination ($P(B_{1,\ldots,n}|A_j)$) with the given class (A_j). It is the likelihood term times the probability of the class $P(A_j)$. The prior (numerator part of Eq. 3) all divided by the evidence term (denominator part of Eq. 3), which is the combination of the predictor features ($P(B_{1,\ldots,n})$).

Sum of the prior would be 1 which is the summation of the probability of impulsive and non-impulsive probabilities. A naïve prior would use 50–50% probability of each class. In such a case, conditional probability, $P(B_n|A)$ can be written as in Eq. 4. μ and σ are the average and standard deviation of each indices of B_n.

In classes, which are impulsive and non-impulsive, the μ and σ for each feature are calculated.

$$P(B_n|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{(B_n - \mu_A)^2}{2\sigma_A^2}\right) \tag{4}$$

Since B has higher dimensions, multivariate Gaussian distribution is applied to it. Multivariate Gaussian distribution is a vector with multiple Gaussian distribution variables. Any linear combination of the variables is also Gaussian distribution. One should modify the Eq. 4, since it is actually valid for the normal distribution in 1 variable.

In normal distribution with 1 variable, $-(B_n - \mu_A)^2/2\sigma_A^2$ is a parabola. In the multivariate case, Eq. 4 should be modified with a quadratic version of B_n. Modified multivariate Gaussian distribution formula can be seen in Eq. 5.

$$P(B_n|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left(-\frac{1}{2}(B_n - \mu_A)^T \sigma^{-1}(B_n - \mu_A)\right) \tag{5}$$

Denominator of Eq. 3 is constant with the all values in the dataset. One can remove the denominator and introduce the proportionality as in Eq. 6.

$$P(A|B_1, \dots, B_n) \alpha P(A_j) \Pi_{i=1}^n P(B_n|A) \tag{6}$$

Since there are two different classes, A with the maximum probability is the answer that needs to be found. One can obtain the class by using Eq. 7.

$$A = \text{argmax}_A P(A) \Pi_{i=1}^n P(B_n|A) \tag{7}$$

Conditional probabilities and priors are assigned accordingly. After that, the model is used for making predictions for the impulsive and non-impulsive signals by giving input parameters.
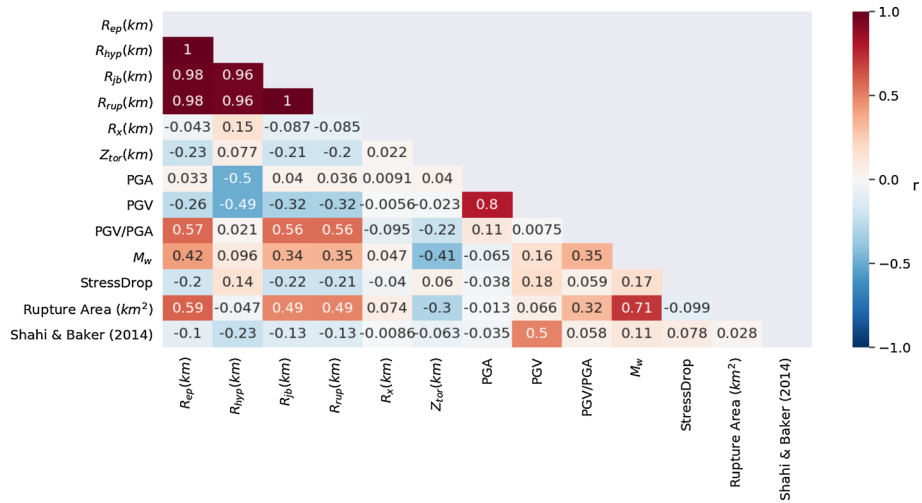
4 Parameter selection

The Pearson correlation coefficient, r, measures the linear correlation between two variables (Eq. 8), x and y. The correlation coefficient is bounded between -1 and 1. A correlation coefficient of ± 1 implies that x and y have perfect correlation (depending on the sign). 0 means no correlation. r has been calculated for each variable.

$$r = \frac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \dfrac{(\sum y_i)^2}{n}}} \tag{8}$$

In order to select the input parameters for the model of probability of occurrence of impulsive signal, r is calculated for several parameters that provide information about the fault plane, at the site of interest and between the fault and the site of interest. Signals and/or earthquakes without variables are excluded from the analysis. r are calculated for all the parameters that can be seen in Fig. 1.

Parameters that are chosen for the first stage are:

1. Epicentral distance (R_{ep}),
2. Hypocentral distance (R_{hyp}),
3. Distance to the surface projection of the rupture (R_{jb}),
4. Distance from the ruptured fault (R_{rup}),
5. The horizontal distance to the surface projection of the top edge of the rupture measured perpendicular to the fault strike (R_x),
6. Depth to top of rupture (Z_{tor}),
7. PGA,
8. PGV,
9. PGV/PGA,
10. M_w,
11. Stress drop,
12. Rupture area.

Fig. 1 Correlation coefficient between the parameters that have been used in the preliminary analysis

In Fig. 1, positive values indicate variables which are directly correlated. Conversely, negative values indicate variables which are inversely correlated. Impulsive signals are labeled as 1 and non-impulsive signals are labeled as 0. Correlation between the variables and impulsive signals are done accordingly. PGV is by far the most relevant parameter for the impulsive signals. However, GMPEs have failed to predict the amplitudes in near-fault regions. In ground motion parameters, largest amplitude of the station is selected. Difference between the actual and predicted amplitudes of PGV can be seen in Fig. 2. Soil classes are taken into account on the amplitude prediction of GMPE. Amplitudes are underestimated in most cases. Due to the unrealistic prediction of GMPE, PGV is excluded from further analysis. PGV/PGA and PGA are also excluded due to the same reason.

R_{hyp} has the second highest correlation coefficient. However, it requires the location of the hypocentral depth of the earthquake. Determination of hypocentral depth can change depending on the station distribution and the velocity structure of the subsurface that is used for the modeling. Uncertainties in hypocentral depth are relatively large. Because of that, it is also eliminated. R_{jb} and R_{rup} have the third highest correlation coefficients. We prefer R_{jb} over R_{rup}, since R_{jb} provides deeper information about the fault plane, such as its limits. R_{rup} on the other hand, does not provide vital information on the fault plane for non-strike-slip faults. R_{jb} is chosen as an input for further analysis. Other distance-related parameters are eliminated due to their relatively lower r.

Moment magnitude, M_w, is also chosen for the next step, since it is one of the most vital information in earthquake hazard studies. It is also proportional to parameters such as ground motion parameters on the site of interest.

Stress drop and ruptured fault are excluded because of their lower correlation coefficients. Stress drop is calculated with the same method that is used in NGA-West 2 database. When the fault dimensions are not available, they are estimated by using the method of Wells and Coppersmith (1994).

Chosen parameters may have very low value of r. Occurrence of an impulsive signals is scarce. To observe an impulsive signal, there are other parameters that are playing big roles such as directivity effect and rupture velocity. However, these parameters are earthquake
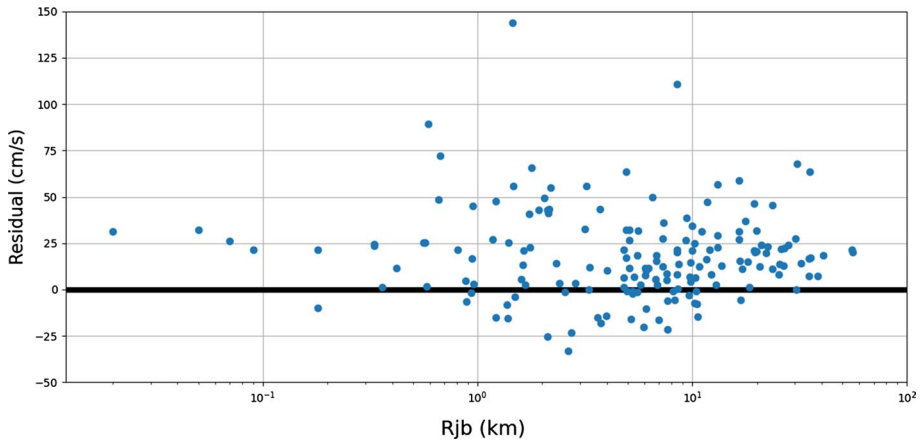
Fig. 2 Amplitude difference between measured PGV and predicted PGV of impulsive signals. GMPE of Boore et al. (2014) is used for the prediction of PGVs

dependent, and it is not convenient to use them as variables. There are several reasons for that. First of all, not all the earthquakes are well studied in terms of their kinematics. Secondly, parameters such as rupture velocity and slip-time history may differ even when using different type of observations for the same earthquake. It is not feasible to use any of these parameters when trying to create a generalized model. Because of that, we limit our parameter space with distance-related, amplitude-related and basic earthquake information-related parameters.

We also add source-to-site azimuth into our parameter space. It is used to give the azimuthal information between the fault plane and the site of interest (Kaklamanos et al. 2011). Source-to-site changes between −180° and 180° depending on the position of the site of interest. Hanging wall and footwall separation can be done by calculating source-to-site azimuth. Site of interest in hanging wall always has positive source-to-site azimuth. Graphical explanation of source-to-site azimuth can be seen in the Electronic Supplementary.

5 Results

Multivariate naïve Bayes classifier is used to model the probability of observing impulsive signals for a given fault type, M_w, R_{jb} and source-to-site azimuth. Dataset is divided into 2 types of faults, which are strike-slip and non-strike-slip.

Both models have limits in terms of R_{jb} and M_w. Distance limitation has been done for computational point of view. Far-field regions are considered to have 0% probability of observing impulsive signals regardless of the source-to-site azimuth and M_w. Probability of occurrences are calculated for maximum and minimum M_w of the dataset for each fault type. In strike-slip faults, minimum M_w that resulted with an impulsive signal is 5.7, whereas maximum magnitude is 7.9. For non-strike-slip faults, minimum and maximum M_ws are 5.4 and 7.9, respectively. There is only one single impulsive signal in $M_w = 5.4$ earthquake in non-strike-slip faults. Impulsive signals become more frequent starting from $M_w = 5.9$. Contour maps are created for M_ws with 0.1 bins. Upper and lower M_w limits

of the models are the magnitudes of the earthquakes that produced at least one impulsive signal for each fault type. Probability distribution maps can be seen in the Electronic Supplementary.

In both strike-slip and non-strike-slip faults, surface rupture may not be observed. In such cases (dip angle ≤ 90°), upper line of the ruptured fault is not located on the surface. Thus, the part over the upper line of the buried rupture is still on the hanging wall. However, on NGA-West2 dataset some of the earthquakes do not have surface rupture information. Furthermore, national data providers have not investigated it. Hence, we assume that all earthquakes have surface rupture. Upper line of the rupture is always located on the surface, and hanging wall - footwall separation has been carried out accordingly. It may create problems, especially when the deriving force of the impulsive signals is directly related with the motion of the hanging wall (e.g., fling step effect during Norcia earthquake studied by D'Amico et al. 2019). Dimensions of the ruptured fault are calculated by using the methodology of Wells and Coppersmith (1994). Epicenter of an earthquake is used as a centroid point of the fault plane.

Number of impulsive and non-impulsive signals with fault types and source-to-site azimuths can be seen in Table 1. In strike-slip faults, source-to-site azimuth values are reduced to binary format, 90° and ≠90°. In strike-slip faults, there is no footwall and hanging wall. Thus source-to-site azimuths can be used in absolute sense. Furthermore, in point source assumption, wave propagation of strike-slip faults can be explained in 4 quadrants. However, we explained faults as planes. Sites that are located at the normal direction of the ruptured fault have source-to-site azimuth of 90°.

Non-strike-slip faults, on the other hand, require the hanging wall and footwall information, since amplitudes vary on both sides of the fault plane. Positive source-to-site azimuth indicates the hanging wall, whereas negative indicates the footwall. Previous studies show that in non-strike-slip faults probability of observing impulsive signals may change on hanging wall and footwall for the same distance and azimuth angle (Chioccarelli and Iervolino 2013; Iervolino and Cornell 2008; Scala et al. 2018; Shabestari and Yamazaki 2003; Shahi and Baker 2014).
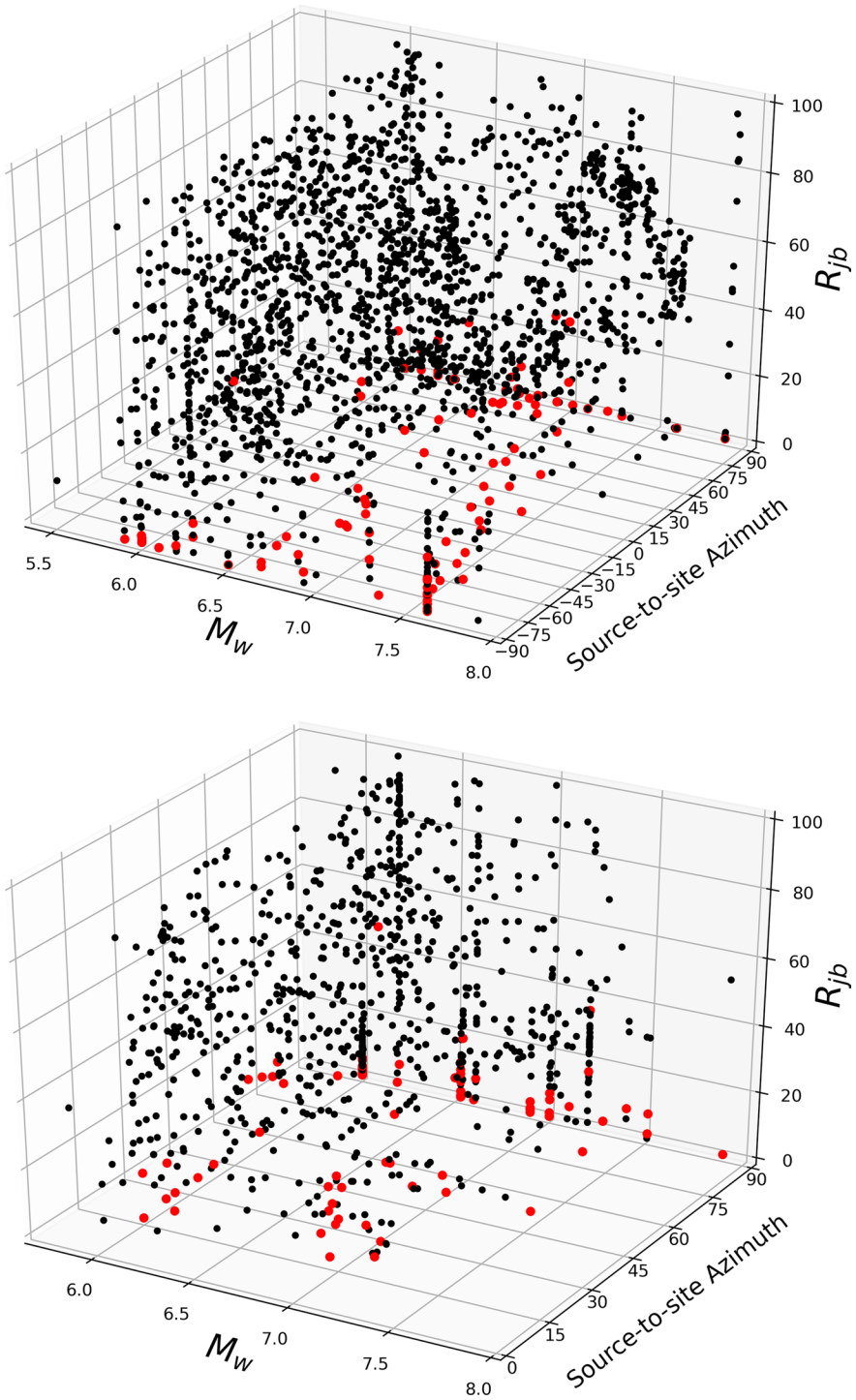
Distribution of the data points on each class is heterogeneous (Fig. 3). The heterogeneity is not only present in variable space but also in class space. Due to the heterogeneous distribution of the stations, they are not able to fully cover all source-to-site azimuth angles. Furthermore, we do not have equal number of data for M_ws. For some earthquakes, stations are sparse in terms of R_{jb} distance, whereas others have dense seismic network in all distances. To top it all, the ratio between impulsive and non-impulsive signal is low.

The heterogeneity in all parameter spaces led us to have incoherent models (see the Electronic Supplementary). To overcome the problem, we smoothed the models. Two different manipulations have been done to smooth the models. The first one is merging

Table 1 Number of impulsive and non-impulsive stations at different fault types and source-to-site azimuth bins

Type of observation	Strike-slip			Non-strike-slip					
	Source-to-site Azimuth (°)								
	[0 90)	90	(90 180)	[− 180 − 90)	− 90	(− 90 0)	[0 90)	90	(90 180)
Impulsive	27	50	10	6	36	27	14	29	7
Non-impulsive	751	286	745	676	166	656	811	220	864

Fig. 3 Distribution of data points in parameter space. Black and red dots represent non-impulsive and impulsive signals, respectively

different source-to-site azimuths (as explained above) which is applied to the data regardless of the fault type. The second smoothing process is done for having a decreasing probability of occurrence with increasing R_{jb} distance. Multivariate Gaussian distribution assures that relation between all parameters are Gaussian distributions. Depending on the data points, Gaussian distribution may have its maximum point in not at the $R_{jb} \approx 0$ but somewhere in the far away distance. It is due to station distribution around the fault planes. When there are lack of stations in and around the R_{jb} area, impulsive behavior around the rupture area may not be captured. In this case, stations are located, relatively, away from the R_{jb} area and the model tends to focus in the detected impulses. This creates Gaussian distributions with a maximum probabilities away from the R_{jb} area (Fig. 4). We would like to avoid the unrealistic models and to do that we have extended to maximum probability value to the $R_{jb} = 0$.

We developed two different models for hanging wall and footwall of non-strike-slip faults. Smoothing is applied separately to these models. In hanging wall, there is a decreasing probability of occurrence with increasing M_w between 5.4 and 6.8. We have calculated the trend of increase magnitudes from 7.1 to 7.4. We have calculated the maximum probability of occurrence for M_w between 5.4 and 6.8 and adjusted the curves accordingly. We have done the same procedure for strike-slip faults by using the M_w from 6.1 to 6.3 to calculate trend, and applied it to M_w between 5.7 and 6.0.

We calculate the accuracy of strike-slip and non-strike-slip models by using the smoothed models. Impulsive and non-impulsive signals are labeled as 1 and 0, respectively. Methodology for the accuracy rates can be seen below:

$$\tilde{x} = \begin{cases} 1, & \text{if } pred(M_w, R_{jb}, \text{source-to-site azimuth}) \geq 50 \text{ when impulsive} \\ 0, & \text{otherwise when impulsive} \\ 1, & \text{if } pred(M_w, R_{jb}, \text{source-to-site azimuth}) < 50 \text{ when non-impulsive} \\ 0, & \text{otherwise when non-impulsive} \end{cases} \quad (9)$$

Accuracy rates are calculated for strike-slip and non-strike-slip models by using Eq. 9. Results can be seen in Table 2. As one can notice, accuracy rates of impulsive signals are lower than the non-impulsive signals for all fault types. There are several reasons for the relatively lower accuracy rates. First of all, impulsive signal(s) may be suppressed by non-impulsive signals in some regions. In such cases, probabilities are less than 50%, in which
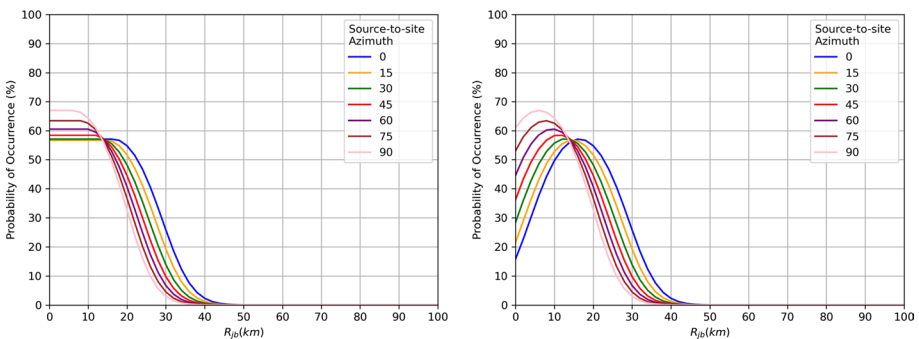
Fig. 4 Probability of observing impulsive signal curves for source-to-site azimuths of 0°, 15°, 30°, 45°, 60°, 75° and 90° for $M_w = 7.1$ in a strike-slip fault. Figure on the left is the original distribution and figure on the right is the smoothed distribution

Table 2 Accuracy rates of strike-slip and non-strike-slip faults for impulsive, non-impulsive and general cases

Accuracy rates (%)	Strike-slip	Non-strike-slip
Impulsive	36.84	41.35
Non-impulsive	98.59	97.77
General	92.94	96.10

Table 3 Accuracy rates for previous studies of Shahi and Baker (2014) and Iervolino and Cornell (2008) strike-slip and non-strike-slip faults for impulsive, non-impulsive and general cases

Accuracy rates (%)	Iervolino and Cornell (2008)		Shahi and Baker (2014)	
	Strike-slip	Non-strike-slip	Strike-slip	Non-strike-slip
Impulsive	49.25	2.47	25.37	51.82
Non-impulsive	98.19	99.92	99.33	99.05
General	96.08	96.90	96.14	97.59

output of our accuracy calculation is always 0. There are also impulsive signals which are observed due to weak soil conditions (Bradley 2012; Kobayashi et al. 2019). Rare occurrence of impulsive signals and rough estimation process of accuracy rates lead us to have relatively lower rates.

Iervolino and Cornell (2008) and Shahi and Baker (2014) have also calculated the probability of occurrence for the impulsive signals. Both of these studies have analyzed the NGA-West 2 database, in which other parameters, apart from the ones that we used in parameter selection, are calculated. It is due to the fact that, well-studied earthquakes are collected in NGA-West 2 database. Both of these studies have used the parameters of R, s and θ for strike-slip, and R, d and ϕ for non-strike-slip faults. Visual explanation of these parameters can be seen in the Electronic Supplementary. Since none of these are calculated for the earthquakes provided by other data centers, we are not able to calculate the accuracy rates for our dataset. To overcome this problem, we calculated the accuracy rates by using NGA-West 2 database data and Eq. 9. Results can be seen in Table 3.

Iervolino and Cornell (2008) has slightly higher accuracy rates in strike-slip faults, whereas it has very low accuracy rates on non-strike-slip faults. In fact, unreliability of the model for non-strike-slip faults is explicitly stated in the study. Shahi and Baker (2014) has relatively higher accuracy rate for strike-slip faults, whereas our model has noticeably higher accuracy rates for non-strike-slip faults.

We build our model by calculating the probabilities in evenly distributed sites near the fault planes. Contour maps are created by using the probability of observing impulsive signals. We choose 2 earthquakes that produced impulsive signals at least in one station. Imperial Valley (Sect. 5.1) and Chi–Chi (Sect. 5.2) earthquakes are used as example cases.
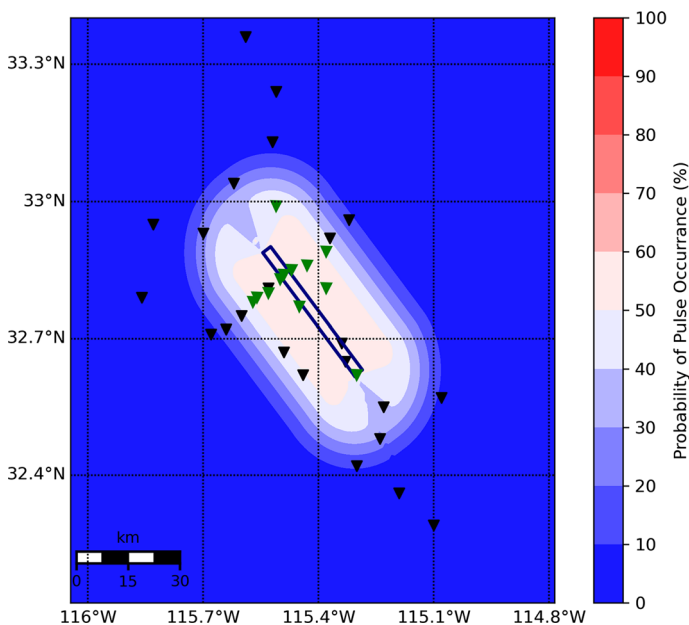
5.1 Imperial Valley earthquake

Impreial Valley earthquake ($M_w = 6.5$) has occurred on 15th of October 1979 in California, USA. It has occurred on a strike-slip fault with 80° of dip angle. Dimensions of the ruptured fault are 50 km of length and 13 km of width. Thirty-five stations have recorded the earthquake of which 12 have impulsive signals.

In Fig. 5, nine stations with impulsive signals lie on the area with more than 50% probability of occurrence. There are non-impulsive signals in the same region where impulsive signals are detected. Three non-impulsive signals where the model predicts the probability of observing impulsive signals to be between 50 and 60%. For this particular earthquake, 75% of the signals are impulsive in the region where model predicts the probability of occurrence is between 50 and 60%. 50% of the signals are impulsive in the region where model predicts the probability of occurrence is between 40% and 50%. For this particular earthquake, model underestimates in the probabilities R_{jb} distances between 0 km and 15 km and overestimates between 15 km and probability of occurrence ≈ 0%. It is worth considering that the model is created by using various earthquakes that produced impulsive signals.

5.2 Chi–Chi earthquake

Chi–Chi earthquake ($M_w = 7.6$) has occurred on 20 of September 1999 in Taiwan. Features of the fault plane are well studied for the earthquake. We use the fault plane information that is determined by Chi et al. (2001). It has occurred on a reverse fault with 30° of dip angle. Dimensions of the ruptured fault are 112 km of length and 45.5 km of width. Four hundred and twenty stations have recorded the earthquake in which 39 of them have impulsive signals.

The probability distribution map of observing impulsive signals for the Chi–Chi, Taiwan earthquake can be seen in Fig. 6. Aagaard et al. (2004) states that there is an up-dip
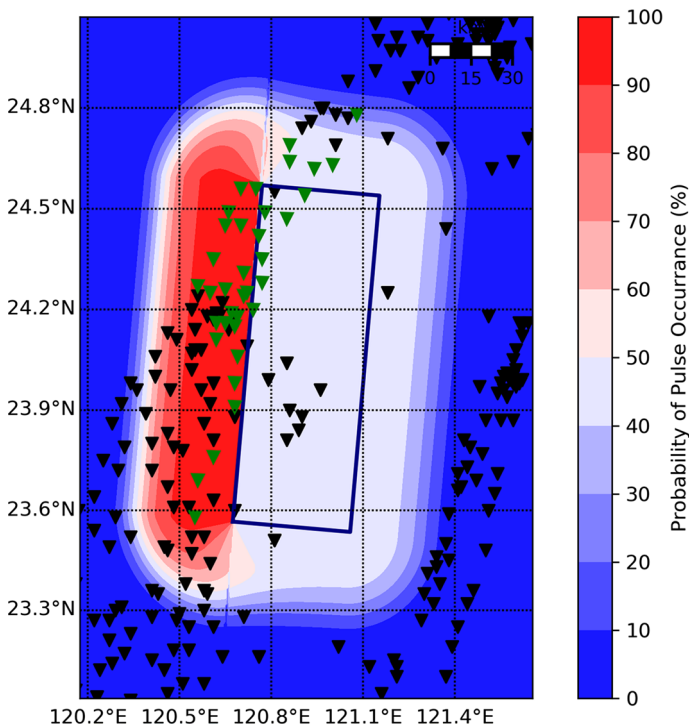

Fig. 5 Probability of observing impulsive signals for the ruptured fault in Imperial Valley earthquake with the same fault type. Green and black triangles are the impulsive and non-impulsive stations for the earthquake, respectively. Navy rectangle represents the surface projection ruptured fault

directivity effect that created impulsive signals that are located in the footwall part of the R_{jb} area. A fling step effect is also recorded (Ji et al. 2003; Kalkan and Kunnath 2006).

Most of the impulsive signals are located in the footwall part and they are located in the region with high probabilities ($\geq$70%). Largest slip is occurred in the fault plane is around north west of the rupture area and most of the rupture is occurred on the shallow part that is the west edge of the fault rupture area in Figs. 6 and 7 (Chi et al. 2001). As before-mentioned, up-dip directivity effect may create impulsive signals in footwall regions (Scala et al. 2018). Furthermore, Aagaard et al. (2004) states that the ruptured directed most of its energy toward the north to northwest part in the Chi–Chi Taiwan earthquake. The forces that described by Aagaard et al. (2004) and Scala et al. (2018) may be the driving forces to the large amount of impulsive signals around the western and north western part of the fault plane. Twelve impulsive signals lie in the hanging wall part with source-to-site azimuth close to zero. The impulsive signals are located in the area where probability of occurrence is less than 50%. Only 5 of the impulsive signals are located in the top part of the hanging wall. It can be explained by the spreading of the radiation on longer time scale for the stations placed on hanging wall (Scala et al. 2018).

53.33% of the signals are impulsive in the region with probability of occurrence between 90 and 100%. In total 40.91% of the signals are impulsive in the area with probability of occurrence bigger than 50%. Hence, model overestimates the probabilities of that area. However, one can see that there are many impulsive and non-impulsive signals located within close distances to each other. High nonlinearity of the process

Fig. 6 Probability of observing impulsive signals for the ruptured fault in Chi–Chi Earthquake with same fault type. Colors have same meaning as in Fig. 5
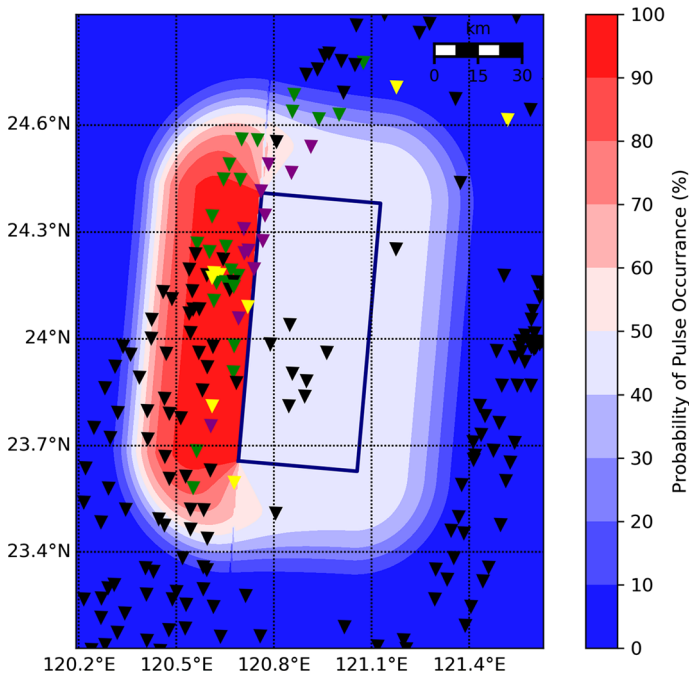
Fig. 7 Impulsive and non-impulsive signals detected by Ertuncay and Costa (2019); Shahi and Baker (2014). Green, yellow and purple triangles indicate the impulsive signals detected by Shahi and Baker (2014), Ertuncay and Costa (2019) and both of the studies, respectively. Black triangles indicates non-impulsive signals. Probability values have the same meaning as in Fig. 6

makes it harder to create a model out of it. Furthermore, determination of an impulsive signal depends on the decision making procedure of studies. Ratio of impulsive to non-impulsive signals in the R_{jb} area is between 40 and 50% which is in agreement with the prediction of the model. Heterogeneous placement of the stations also affect the model. In the end, neither our model nor previous models have high accuracy rates (Tables 2, 3).

Moreover, we analyze the impulsive signals in vertical components (Fig. 7). Shahi and Baker (2014) uses the horizontal components to detect the impulsive behavior of a given station. We use the method of Ertuncay and Costa (2019) to detect the impulsive signals since the method can analyze the components individually.

It is found that some of the impulsive signals detected on horizontal motions have also vertical impulses. Moreover, there are 6 stations located in the region, where probability of observing impulsive signals are larger than 50%, with impulsive behavior only in vertical components.

Another aspect of this earthquake is the ruptured fault area defined by Chi et al. (2001). According to the study, upper part of the rupture is ended in 0.9 km depth. Due to our assumption of having surface rupture in all earthquakes, stations with R_{jb} distance less than 1.56 km in footwall are actually located at the region where the effect of the upward movement of the fault is present. There are three impulsive signals lying in that part. We did not closely investigate these signals since the aim of this study is to create a generalized method instead of developing event-based finely tuned models.

6 Conclusions

In this study, we would like to develop a new model to predict the probability of observing impulsive signals. To do that, a vast amount of signals are collected from various data providers. Earthquakes that produced these signals are also investigated. Information about the earthquakes and the signals are collected. Earthquakes that produced impulsive signals are chosen for further analysis.

Relation between the impulsive signals and the earthquakes, ruptured faults and the stations are analyzed. We find that, it is necessary to know M_w, R_{jb}, fault type and source-to-site azimuth information to predict the probability of observing impulsive signals. To create a model for predicting the probability, multivariate naïve Bayes classifier approach is implemented to the data. The dataset is divided into two categories by using the information of the fault type. Models are developed for strike-slip and non-strike-slip faults with an assumption of surface rupture. (Hanging wall and foot walls are always separated by the upper edge of the rupture).

Comparison between our study and previous studies focusing on the generalization of the probability of observing impulsive signals are compared. We find that, even though previous models can have better accuracy rates in one of the two fault types, our model is the most consistent in terms of general accuracy. However, our model can underestimate and overestimate the probabilities depending on the given earthquake. It is due to several factors which are, i.e., (i) heterogeneous distribution of stations, (ii) strong directivity effect in several non-strike-slip earthquakes, (iii) local soil conditions, (iv) uncertainties in fault plane (e.g., assumption of the presence of the surface rupture in every non-strike-slip earthquakes) and (v) neglecting vertical impulsive signals in the model.

We applied our model in Imperial Valley (strike-slip) and Chi–Chi (non-strike-slip) earthquakes. In both cases, impulsive signals were founded in areas close to the fault. In Chi–Chi earthquake, large number of impulsive signals are located on the footwall due to rupture features of the event. On the other hand, 30th of October 2016 Norcia earthquake create many impulsive signals in the hanging wall (D'Amico et al. 2019). Spatial features of the impulsive signals may change depending on multiple parameters such as the fault mechanism and rupture feature. Correlation between impulsive signal and characteristics of the fault and rupture features may be obtained in future using our model for investigating a larger number of cases of near-field signals recorded in occasion of major earthquakes with different characteristics. Finally, it may also change in the future when new near-field signals are recorded in major earthquakes.

It is also found that vertical impulsive signals are recorded in non-strike-slip faults. It can be linked to the top part of the ruptured fault and its vertical displacement during an earthquake. Vertical impulsive signals should also be investigated in the upcoming studies by using the pulse identification methods that are capable of analyzing vertical motions.

Author Contributions DE proposed the topic, collected the data and created the models. Interpretation of the results is done by both DE and GC.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Data availability Data and materials along with the developed models can be found on dedicated GitHub repository.

References

Aagaard BT, Hall JF, Heaton TH (2004) Effects of fault dip and slip rake angles on near-source ground motions: Why rupture directivity was minimal in the 1999 Chi-Chi, Taiwan, earthquake. Bull Seismol Soc Am 94(1):155–170

AFAD (2019) AFAD—Turkey Earthquake Data Center System. Accessed 25 Oct 2019

Alavi B, Krawinkler H (2004) Behavior of moment-resisting frame structures subjected to near-fault ground motions. Earthq Eng Struct Dyn 33(6):687–706

Antonellis G, Panagiotou M (2013) Seismic response of bridges with rocking foundations compared to fixed-base bridges at a near-fault site. J Bridg Eng 19(5):04014007

Baker JW (2007) Quantitative classification of near-fault ground motions using wavelet analysis. Bull Seismol Soc Am 97(5):1486–1501

Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) Nga-west2 equations for predicting pga, pgv, and 5% damped psa for shallow crustal earthquakes. Earthq Spectra 30(3):1057–1085

Bozorgnia Y, Abrahamson NA, Atik LA, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BS-J et al (2014) NGA-West2 research project. Earthq Spectra 30(3):973–987

Bradley BA (2012) Strong ground motion characteristics observed in the 4 September 2010 Darfield, New Zealand earthquake. Soil Dyn Earthq Eng 42:32–46

Bray JD, Rodriguez-Marek A (2004) Characterization of forward-directivity ground motions in the near-fault region. Soil Dyn Earthq Eng 24(11):815–828

Chang Z, Sun X, Zhai C, Zhao JX, Xie L (2016) An improved energy-based approach for selecting pulse-like ground motions. Earthq Eng Struct Dyn 45(14):2405–2411

Chi W-C, Dreger D, Kaverina A (2001) Finite-source modeling of the 1999 Taiwan (Chi–Chi) earthquake derived from a dense strong-motion network. Bull Seismol Soc Am 91(5):1144–1157

Chioccarelli E, Iervolino I (2013) Near-source seismic hazard and design scenarios. Earthq Eng Struct Dyn 42(4):603–622

D'Amico M, Felicetta C, Schiappapietra E, Pacor F, Gallovič F, Paolucci R, Puglia R, Lanzano G, Sgobba S, Luzi L (2019) Fling effects from near—source strong—motion records: insights from the 2016 Mw 6.5 Norcia, Central Italy. Earthq Seismol Res Lett 90(2A):659–671

Ertuncay D, Costa G (2019) An alternative pulse classification algorithm based on multiple wavelet analysis. J Seismol 23:929–942

Grimaz S, Malisan P (2014) Near field domain effects and their consideration in the international and Italian seismic codes. Bollettino di Geofisica Teorica ed Applicata 55(4):717–738

Guo G, Yang D, Liu Y (2018) Duration effect of near-fault pulse-like ground motions and identification of most suitable duration measure. Bull Earthq Eng 16(11):5095–5119

Hall JF, Heaton TH, Halling MW, Wald DJ (1995) Near-source ground motion and its effects on flexible buildings. Earthq Spectra 11(4):569–605

Iervolino I, Cornell CA (2008) Probability of occurrence of velocity pulses in near-source ground motions. Bull Seismol Soc Am 98(5):2262–2277

Iervolino I, Chioccarelli E, Baltzopoulos G (2012) Inelastic displacement ratio of near-source pulse-like ground motions. Earthq Eng Struct Dyn 41(15):2351–2357

Iervolino I, Baltzopoulos G, Chioccarelli E, Suzuki A (2017) Seismic actions on structures in the near-source region of the 2016 central Italy sequence. Bull Earthq Eng 17:5429–5447

Institute of Earth Sciences, Academia Sinica, T (1996) Broadband array in Taiwan for seismology

Ji C, Helmberger DV, Wald DJ, Ma K-F (2003) Slip history and dynamic implications of the 1999 Chi–Chi, Taiwan, earthquake. J Geophys Res Solid Earth 108(B9):2412

Kaklamanos J, Baise LG, Boore DM (2011) Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. Earthq Spectra 27(4):1219–1235

Kalkan E, Kunnath SK (2006) Effects of fling step and forward directivity on seismic response of buildings. Earthq Spectra 22(2):367–390

Kardoutsou V, Taflampas I, Psycharis IN (2017) A new pulse indicator for the classification of ground motions. Bull Seismol Soc Am 107(3):1356–1364

Kobayashi H, Koketsu K, Miyake H (2019) Rupture process of the 2018 Hokkaido Eastern Iburi earthquake derived from strong motion and geodetic data. Earth Planets Space 71(1):63

Lay T, Wallace TC (1995) Modern global seismology. Elsevier, Amsterdam

Li C, Kunnath S, Zuo Z, Peng W, Zhai C (2020) Effects of early-arriving pulse-like ground motions on seismic demands in RC frame structures. Soil Dyn Earthq Eng 130:105997

Loh C-H, Wan S, Liao W-I (2002) Effects of hysteretic model on seismic demands: consideration of near-fault ground motions. Struct Des Tall Build 11(3):155–169

Luzi L, Puglia R, Russo E, D'Amico M, Felicetta C, Pacor F, Lanzano G, Çeken U, Clinton J, Costa G et al (2016) The engineering strong-motion database: a platform to access pan-European accelerometric data. Seismol Res Lett 87(4):987–997

Mazza F (2018) Seismic demand of base-isolated irregular structures subjected to pulse-type earthquakes. Soil Dyn Earthq Eng 108:111–129

Mena B, Mai PM (2011) Selection and quantification of near-fault velocity pulses owing to source directivity. Georisk 5(1):25–43

Pacor F, Paolucci R, Luzi L, Sabetta F, Spinelli A, Gorini A, Nicoletti M, Marcucci S, Filippi L, Dolce M (2011) Overview of the Italian strong motion database ITACA 1.0. Bull Earthq Eng 9(6):1723–1739

Scala A, Festa G, Del Gaudio S (2018) Relation between near-fault ground motion impulsive signals and source parameters. J Geophys Res Solid Earth 123(9):7707–7721

Shabestari KT, Yamazaki F (2003) Near-fault spatial variation in strong ground motion due to rupture directivity and hanging wall effects from the Chi-Chi, Taiwan earthquake. Earthq Eng Struct Dyn 32(14):2197–2219

Shahi SK, Baker JW (2011) An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis. Bull Seismol Soc Am 101(2):742–755

Shahi SK, Baker JW (2014) An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions. Bull Seismol Soc Am 104(5):2456–2466

Somerville PG (2003) Magnitude scaling of the near fault rupture directivity pulse. Phys Earth Planet Inter 137(1–4):201–212

Somerville PG (2005) Engineering characterization of near fault ground motions. In: NZSEE conference, pp 1–8

Somerville PG, Smith NF, Graves RW, Abrahamson NA (1997) Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. Seismol Res Lett 68(1):199–222

Spudich P, Chiou BS (2008) Directivity in NGA earthquake ground motions: analysis using isochrone theory. Earthq Spectra 24(1):279–298

SSN (2017) Servicio Sismologico Nacional

Tothong P, Cornell CA, Baker J (2007) Explicit directivity-pulse inclusion in probabilistic seismic hazard analysis. Earthq Spectra 23(4):867–891

Universidad De Chile (2013) Red sismologica nacional

Van Houtte C, Bannister S, Holden C, Bourguignon S, McVerry G (2017) The New Zealand strong motion database. Bull N Z Soc Earthq Eng 50(1):1–20

Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. Bull Seismol Soc Am 84(4):974–1002

Yang D, Wang W (2012) Nonlocal period parameters of frequency content characterization for near-fault ground motions. Earthq Eng Struct Dyn 41(13):1793–1811

Zhai C, Li C, Kunnath S, Wen W (2018) An efficient algorithm for identifying pulse-like ground motions based on significant velocity half-cycles. Earthq Eng Struct Dyn 47(3):757–771

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.