Available online at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Use of the t-Distribution to Construct Seismic Hazard Curves for Seismic Probabilistic Safety Assessments

Eric Yee *

KEPCO International Nuclear Graduate School, Department of Nuclear Power Plant Engineering, 658-91, Haemaji-ro Seosaeng-myeon, Ulju-gun, Ulsan, 45014, South Korea

ARTICLE INFO

Article history:
Received 17 November 2016
Received in revised form
17 December 2016
Accepted 23 December 2016
Available online 24 January 2017

Keywords:
PSHA
Hazard
t-distribution
Earthquake
Seismic

ABSTRACT

Seismic probabilistic safety assessments are used to help understand the impact potential seismic events can have on the operation of a nuclear power plant. An important component to seismic probabilistic safety assessment is the seismic hazard curve which shows the frequency of seismic events. However, these hazard curves are estimated assuming a normal distribution of the seismic events. This may not be a strong assumption given the number of recorded events at each source-to-site distance. The use of a normal distribution makes the calculations significantly easier but may underestimate or overestimate the more rare events, which is of concern to nuclear power plants. This paper shows a preliminary exploration into the effect of using a distribution that perhaps more represents the distribution of events, such as the t-distribution to describe data. The integration of a probability distribution with potentially larger tails basically pushes the hazard curves outward, suggesting a different range of frequencies for use in seismic probabilistic safety assessments. Therefore the use of a more realistic distribution results in an increase in the frequency calculations suggesting rare events are less rare than thought in terms of seismic probabilistic safety assessment. However, the opposite was observed with the ground motion prediction equation considered.

1. Introduction

A common technique to evaluate the exposure of a nuclear power plant, or any other important piece of infrastructure, to earthquakes is to perform a probabilistic safety assessment for seismic events. The state-of-practice takes an interdisciplinary approach, with multiple components that designers and engineers must consider. One integral component to seismic probabilistic safety assessments is the seismic hazard curve. This curve basically plots an intensity measure, such as peak ground acceleration (PGA) or peak ground velocity, to an annual probability of exceedance (APE). Some hazard curves use a return period in place of an APE, but the name "return period" has caused some confusion to practitioners over the

* *Corresponding author.*
E-mail address: eric.yee@kings.ac.kr.
http://dx.doi.org/10.1016/j.net.2016.12.014

years. Another component is a system, structure, or component fragility curve. The fragility curve maps the conditional probability of a system, structure, or component failure due to a certain intensity measure. For the same intensity measure, the seismic hazard curve and the fragility curve are multiplied to produce a failure frequency, or failure probability curve for a specific system, structure, or component. The area under this failure probability curve can be considered the failure rate. These values are then compared to satisfy a design or regulatory minimum threshold.

Seismic hazard curves can be calculated deterministically or probabilistically. Probabilistic seismic hazard analysis is used in constructing the seismic hazard curve. Probabilistic seismic hazard analyses essentially have four steps: the first step is a characterization of the site for potential seismic sources is performed; this involves magnitude and source-to-site distance identification in a probabilistic manner; the second step is the establishment of a magnitude recurrence relationship, which addresses the temporal issues in the analyses; the third step is to prepare a ground motion prediction equation (GMPE). The GMPE takes in earthquake source characteristics and source-to-site distances to estimate a variety of intensity measure parameters. Modern GMPEs usually treat traditional intensity measures as log-normally distributed; and the fourth step uses the GMPE and temporal relationships to construct the seismic hazard curve.

A common criticism of GMPEs is the lack of data for certain conditions, such as earthquake magnitude (M) and site-to-source distance (R). As an example, Fig. 1 shows a sample database of recorded seismic events up to 2008. The figure shows a lack of data for distances < 10–20 km, and for low magnitudes generally. This presents an issue in that it is difficult for the engineer to confine or model earthquake effects for those conditions. It is generally accepted that the behavior observed at moderate distances for which there is a good amount of data should be applicable to close and far distances. However, earthquake engineers understand that observed behavior at close and far distances may not be properly represented by those at moderate distances. For example, issues such as radiation damping, surface waves, directivity, and ground nonlinearity make this problem difficult to address.

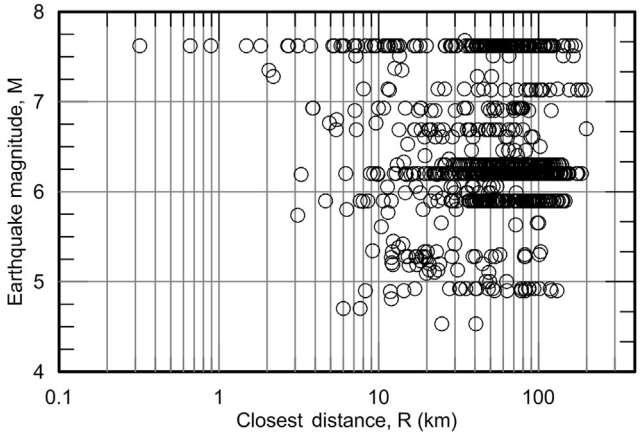

Fig. 1 – Projected number of earthquakes used by Idriss [1] up to 2008.

This paper attempts to address these data issues by assigning a Student *t*-distribution in place of the traditional normal distribution to the natural logarithm of a sample intensity measure. This requires the deconstruction of a popular GMPE to evaluate under which conditions a *t*-distribution would be applicable. A sample scenario is then presented to showcase the effects, if any, a change in distribution would have on the results of the simple probabilistic safety assessment.

2. Hazard curve construction

2.1. Deconstruction of selected GMPE

In order to determine under which conditions a *t*-distribution might be applicable, a deconstruction of a GMPE is needed. There are a variety of GMPEs all over the world, accounting for a variety of seismic strong ground motion related parameters such as faulting mechanism, local site effects through shear wave velocity in the upper 30 m (V_{S30}), hanging wall or footwall locations, and directivity effects. Of course the inclusion of more strong ground motion parameters complicates the mathematical form of the GMPE. Most GMPEs also produce estimates for a variety of intensity measures such as peak ground velocity, Arias intensity, and spectral accelerations at a range of periods. This study selected the GMPE by Idriss [1] because: (1) it is easy and simple to use; (2) the data used to construct the GMPE are readily available; (3) the data used to construct the GMPE are listed in the researchers' publication; and (4) the logic and data selection process is generally well explained because it is part of the larger Next Generation Attenuation program [2,3]. For simplicity, only PGA will be considered as the intensity measure for this study. The functional form of the GMPE for PGA where 450 m/s $\leq V_{S30} \leq$ 900 m/s and $M \leq 6.75$ is:

$$\ln(PGA) = 3.7066 - 0.1252M - (2.9832 - 0.2339M)\ln(R + 10) + 0.00047R + 0.12F \quad (1)$$

and when 450 m/s $\leq V_{S30} \leq$ 900 m/s and $6.75 < M \leq 8.5$, it is:

$$\ln(PGA) = 5.6315 - 0.4104M - (2.9832 - 0.2339M)\ln(R + 10) + 0.00047R + 0.12F \quad (2)$$

where F = faulting mechanism, $F = 0$ for strike slip and normal faulting events, and $F = 1$ for reverse faulting events. For sites with $V_{S30} > 900$ m/s, Idriss [1] recommended the following when $M \leq 6.75$:

$$\ln(PGA) = 3.5574 - 0.1252M - (2.9832 - 0.2339M)\ln(R + 10) + 0.00047R + 0.12F \quad (3)$$

and the following when $M > 6.75$:

$$\ln(PGA) = 5.4823 - 0.4104M - (2.9832 - 0.2339M)\ln(R + 10) + 0.00047R + 0.12F \quad (4)$$

Following the descriptions by Idriss [1], a database was developed containing the earthquake events, recording stations, and PGA data. These events were then compared with the Idriss GMPE [1] to calculate residuals, as shown in Fig. 2. Residuals were calculated by subtracting discrete PGA data points taken from the aforementioned database of earthquake recordings from the Idriss GMPE [1] estimated PGA, basically the difference between the observed PGA and the predicted PGA. The subscript ln is used to describe the fact that the residuals are calculated from the natural logarithm of the PGA values because, as mentioned in the Introduction, most GMPEs treat intensity measures as log-normally distributed. A positive residual would indicate that the GMPE had underpredicted that scenario and a negative residual would indicate that the GMPE had overpredicted that scenario. A plot of the residuals can provide an insight into how good the GMPE is at estimating the intensity measure of interest. Residuals relatively near or at 0 suggest a well-developed GMPE, whereas many residuals relatively far from 0 suggest the absence of a parameter from the GMPE. The Idriss GMPE [1] also provides a standard deviation dependent on the intensity measure and the earthquake magnitude, which results in a value of approximately 0.7 for PGA as an intensity measure. Fig. 2 shows that most of the data falls between 1 and 2 standard deviations of the mean near 0, suggesting a well-developed GMPE for the assumptions made.

Fig. 3 plots the PGAs from seismic events recorded at distilled recording stations against the derived GMPE for normal and reverse fault mechanisms, where mechanism 0 is for normal and strike-slip events and mechanism 1 is for reverse faulting events. The plots show seven sets of data, when $4.5 \leq M < 5$, $5 \leq M$ 5.5, $5.5 \leq M < 6$, $6 \leq M$ 6.5, $6.5 \leq M < 7$, $7 \leq M$ 7.5, and $7.5 \leq M < 8$, which is similar to how Idriss [1] grouped his data set. Earthquakes with $M < 4.5$ are typically not of engineering interest because such earthquakes rarely produce strong ground motion. Additionally, data at the time of Idriss [1] did not contain many events with $M > 8$. In each subplot of Fig. 3, the indicated earthquake magnitude is the average of all the data in the corresponding range. These plots show a lack of data at distances of < 10 km for these magnitude ranges. The lack of data for distances > 150 km is because the designer limited the GMPE to < 200 km and the observation that attenuation eventually reduces the ground motions to negligible recordings. Fig. 3 shows that the GMPE tends to underpredict PGA when $4.5 \leq M < 5$ and $6.5 \leq M < 7$, whereas the GMPE tends to overpredict when $7 \leq M$ 7.5 and $7.5 \leq M < 8$.

When comparing the results to the results provided in Idriss [1], the author was unable to reverse engineer a perfect match. Some stations could not be reasonably eliminated and some of the station descriptors listed in Idriss [1] were not identified in the updated flatfile provided by the Pacific Earthquake Engineering Research Center (PEER). This resulted in relatively more data points, mostly from the Taiwan events. Fig. 2 basically shows more points than the figures provided in the publication by Idriss [1], as well as some points having differing residuals. A few site-specific comparisons were also made, showing that a majority of the data points match; however, for each comparison, there were additional points absent from Idriss [1]. It would seem that some of the data used by Idriss [1] were different from those provided in the updated flatfile by PEER. For example, Idriss [1] states that 45 station recordings were available for sites where $V_{S30} > 900$ m/s, but the resultant database only had four station recordings available. However, because a majority of the results appeared to be similar to those published by Idriss [1], being unable to obtain a perfect match is not a significant obstacle in demonstrating the effects of modifying the underlying normal distribution assumption for intensity measures.

2.2. *Scenario*

The scenario considered for this preliminary study is shown in Fig. 4. This shows a site with two earthquake sources defined as vertical linear faults. Fault A has a closest site-to-source distance of 10 km, which can only produce earthquakes with magnitude 6.5 at a rate of 0.01 times a year. Fault B has a closest site-to-source distance of 20 km, which can only produce earthquakes with magnitude 7.5 at a rate of 0.002 times a year. This is a common example scenario and is only used here to demonstrate the effects of change in GMPE characteristic.

For the given scenario, the hazard integral is calculated using the original Idriss GMPE [1] and is plotted in Fig. 5. The hazard integral is commonly known as:

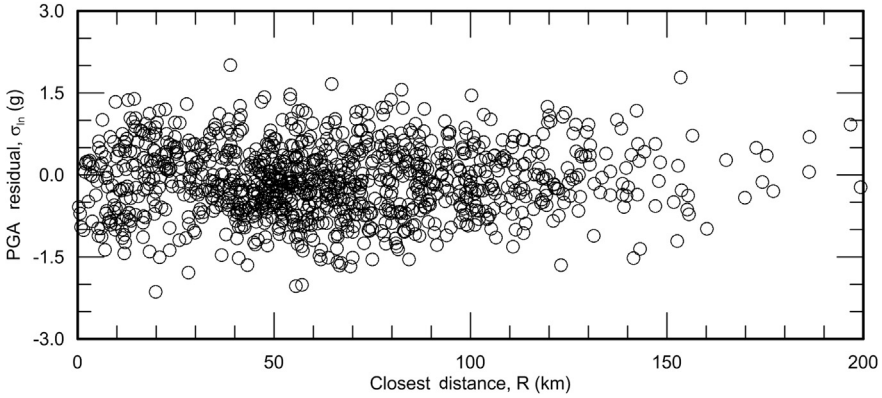

Fig. 2 – Reverse-engineered residuals using the event and station descriptors by Idriss [1].

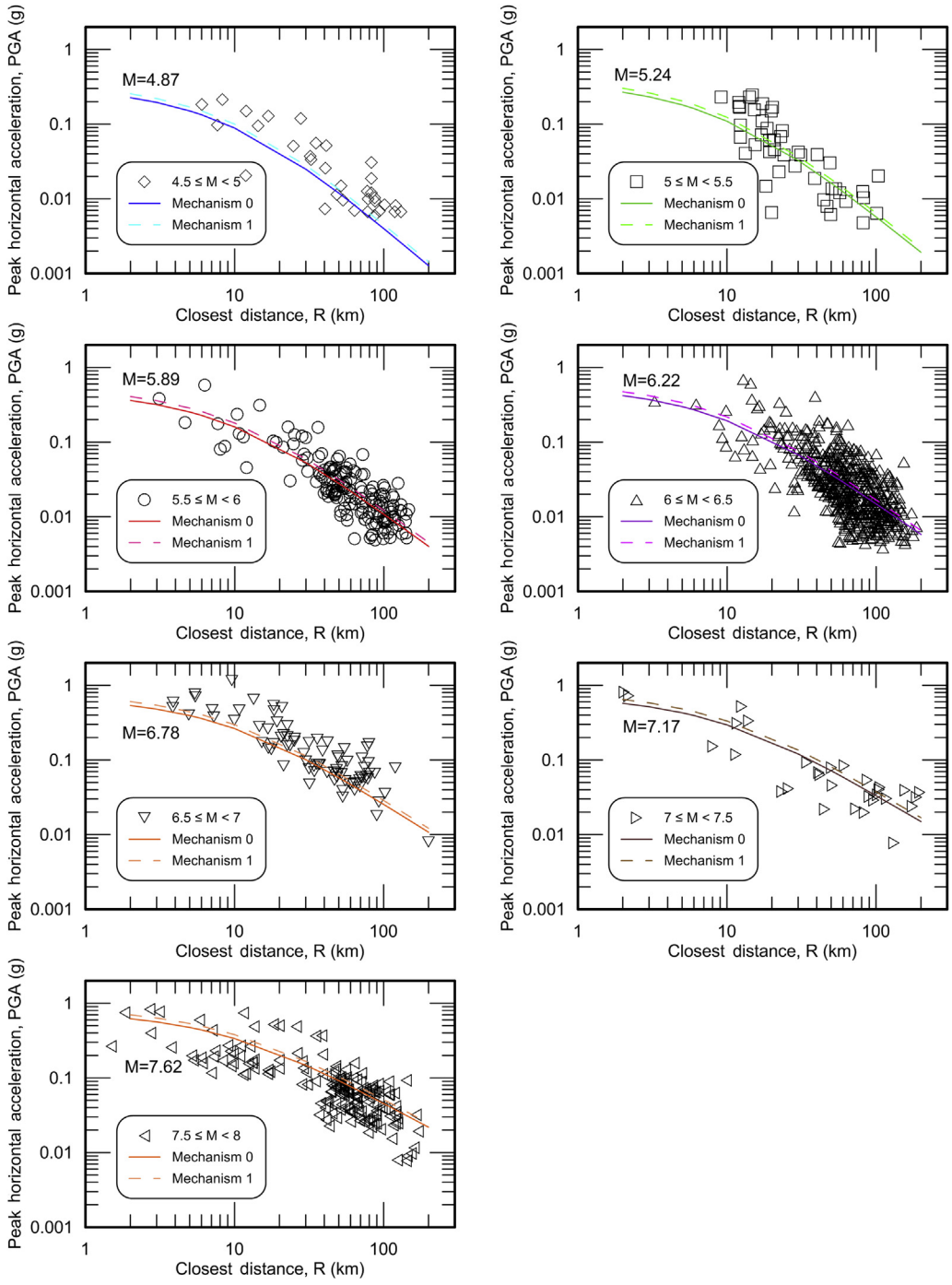

Fig. 3 – Reverse-engineered data points and the median GMPE under different seismic source conditions for Idriss [1]. GMPE, ground motion prediction equation.

$$\lambda(PGA > x) = \sum_{i=1}^{N_s} \upsilon \iint P[PGA > x|m,r]P[M = m]P[R = r]drdm \qquad (5)$$

where x = predetermined *PGA* level, λ = rate that *PGA* x is exceeded, *APE*; N_s = number of earthquake sources; υ = rate of earthquakes at the earthquake source; $P[PGA > x \mid m, r]$ = for a given magnitude m and source-to-site distance r, the probability of exceeding a *PGA* x; attenuation relationship, GMPE; $P[R = r]$ = probability that source-to-site distance is r; and $P[M = m]$ = probability that earthquake magnitude is m.

Hazard curves are calculated from the hazard integral by specifying a target *PGA* and then calculating the probability of exceeding the aforementioned level of *PGA*. This is calculated by taking the area under the lognormal distribution, which is the underlying distribution of the PGA values, given distance and magnitude. The area is typically calculated by using the z-value, which in this case is shown as:

$$\frac{\ln x - \overline{\ln PGA}}{\sigma_{\ln PGA}} \qquad (6)$$

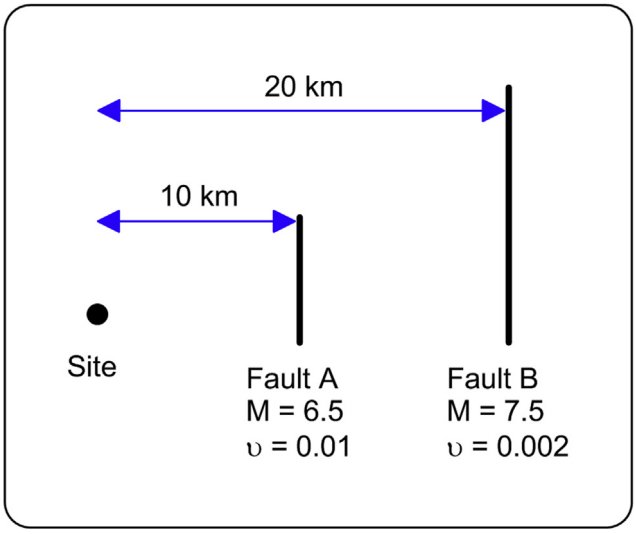

Fig. 4 – Map view of example site with two earthquake sources.

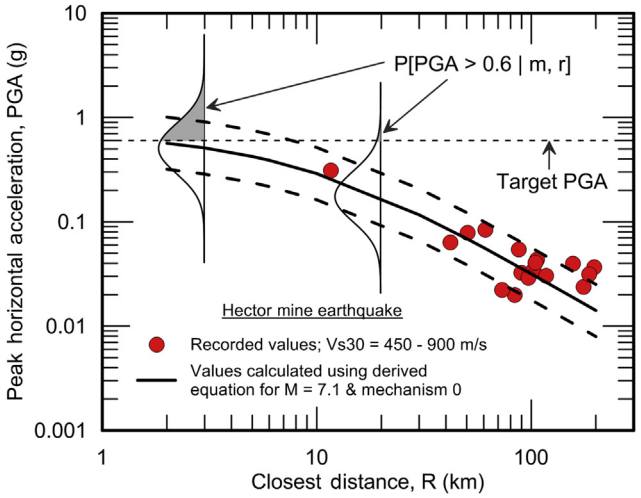

Fig. 6 – Example hazard curve calculations using Idriss [1] data for Hector Mine event.

where $\overline{\ln PGA}$ = median GMPE estimate, and $\sigma_{\ln PGA}$ = standard deviation of the log-normally distributed PGA values, usually given in the GMPE descriptor. The z-value can be used to look up the area under the normal distribution curve.

The sum of all the areas at each source-to-site distance represents the hazard level APE in the hazard curve. An example is shown in Fig. 6. As one can see, if the distribution changes, so should the hazard level. Additionally, a change in the distribution would also result in a change in the GMPE. Fig. 5 also plots the results obtained by using the *t*-distribution in the Idriss GMPE [1]. The *t*-distribution was basically implemented by estimating a t-value, which is essentially the same as a z-value for most GMPEs, by subtracting the median GMPE value from the PGA level and then dividing by a standard deviation for the *t*-distribution. The standard deviation for the unknown *t*-distribution was calculated by:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n)(n-1)}} \tag{7}$$

where x_i = data points, $\bar{x}$ = GMPE median, and n = number of points. The limits are from 0.1 g to 2 g because PGAs < 0.1 g are of little interest, whereas those > 2 g are slightly unrealistic. The area under the *t*-distribution curve was numerically estimated using the database of seismic events constructed previously. For Fault A, this meant tallying all seismic events with $6.5 \leq M < 7$ and $R \leq 10$ km; for Fault B, this meant tallying all seismic events with $7.5 \leq M < 8$ and $R \leq 20$ km. Source-to-site distance was taken as a $R \leq 10$ km and $R \leq 20$ km for Faults A and B, respectively, primarily due to a lack of data at short distances and the idea that at such close distances the differences between PGA may most likely be negligible. This resulted in nine and 48 events for Faults A and B, respectively. These values were used to estimate a standard deviation for each situation as well as to calculate the degrees-of-freedom required for estimating the area under the *t*-distribution curve.

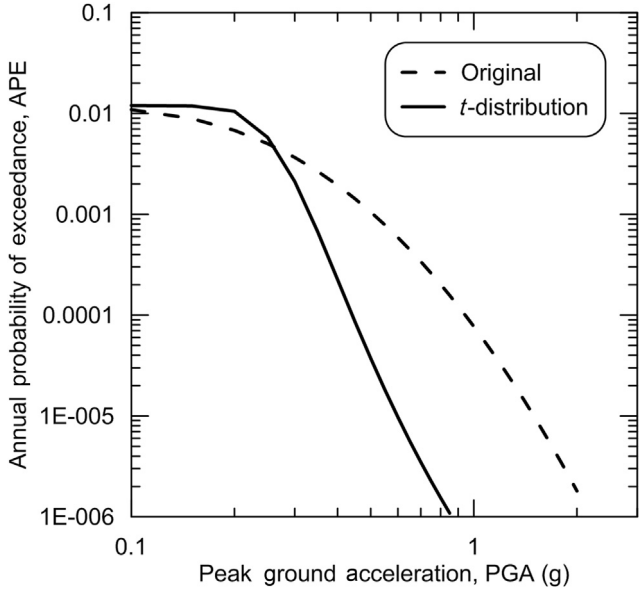

Fig. 5 – Hazard curves.

Fig. 5 shows that by changing to a Student *t*-distribution the hazard curve shows a reluctance to reach large accelerations. For accelerations < ~0.25 g, the adjusted hazard curve appears to show an increase in hazard APE, whereas accelerations > ~0.25 g, the adjusted hazard curve appears to show a decrease in hazard APE relative to the original Idriss GMPE [1] for the subject scenario. This strongly suggests that relatively strong seismic events are overestimated relative to the selected GMPE, which also suggests that the *t*-distribution used most likely had smaller tails than the log–normal distribution estimated by the GMPE. Table 1 shows that the standard deviations for the log–normal distribution were relatively larger than the standard deviations for the *t*-distribution of the natural logarithm of the PGAs. These results are slightly contrary to what was expected.

Table 1 – List of standard deviations for the earthquake events.

Seismic source	Normal distribution σ_{ln}	Student t-distribution σ_{ln}
Fault A	0.61	0.19
Fault B	0.53	0.12

3. Scenario

3.1. Fragility curve

For this scenario, we will only consider the generic fragility curve for a pressurizer. As real data are not readily available, the generic fragility curve is calculated using the procedure outlined by EPRI [4]. This procedure basically uses the following relationship:

$$f' = \Phi\left[\frac{\ln\left(\frac{a}{A_m}\right) + \beta_U \Phi^{-1}(Q)}{\beta_R}\right] \quad (8)$$

where f' = conditional probability of failure; a = given peak ground acceleration level; A_m = median ground acceleration capacity; β_U = inherent randomness (aleatory uncertainty); β_R = unknown uncertainty (epistemic uncertainty); Q = subjective probability (confidence) that the conditional probability of failure, f, is $< f'$ for a peak ground acceleration a; Φ^{-1} = inverse of Gaussian cumulative distribution; and Φ = standard Gaussian cumulative distribution. For a pressurizer, the parameters are taken from NUREG/CR-6544 [5] with $A_m = 2.5$ g, $\beta_U = 0.40$, and $\beta\beta_R = 0.30$. Additionally, $Q = 50\%$ was chosen. Fig. 7 plots the results for an example pressurizer using the aforementioned parameters.

3.2. Failure probability

The hazard curve is multiplied by the fragility curve to produce the failure probability curve, which is shown in Fig. 8.

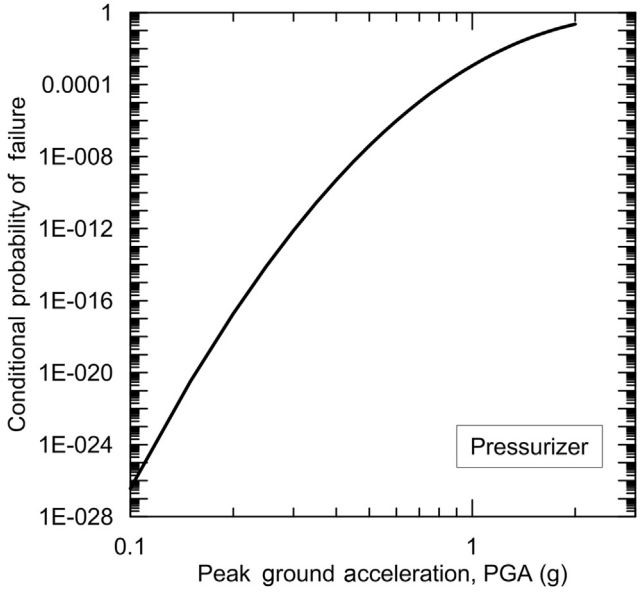

Fig. 7 – Sample fragility curve for a pressurizer.

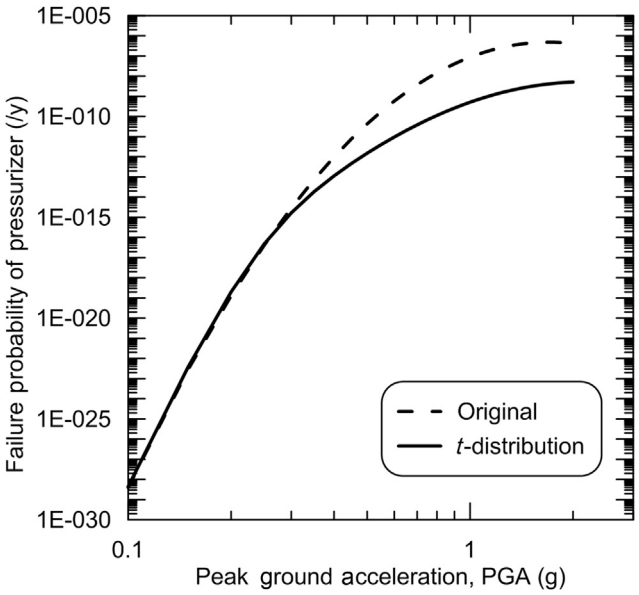

Fig. 8 – Failure probability of a sample pressurizer.

The figure shows that a change from a normal distribution to a Student t-distribution results in a lowering of the failure probability of the pressurizer for the PGA range of interest. When the area under the curve is calculated, the original GMPE results in a failure frequency of approximately 3.74×10^{-7} / year, whereas the utilization of a t-distribution results in a failure frequency of approximately 3.02×10^{-9} / year. According to this, the t-distribution has reduced the probability of failure from an earthquake at least twofold. Inclusion of PGAs < 0.1 g and > 2 g would most likely not decrease the gap in the failure frequencies.

4. Discussion

The use of a normal distribution to describe the randomness in the natural logarithm of earthquake PGA values in a GMPE is perhaps sometimes unwarranted given the data available and the environmental conditions of the data. This preliminary study attempted to address this issue by switching from a normal distribution to a Student t-distribution. A popular GMPE was selected to analyze a scenario of two faults near a site. This resulted in a seismic hazard curve that showed significantly less probability of exceedance, which was attributed to the relatively low t-distribution standard deviations calculated. The seismic hazard curves were then multiplied by a generic fragility curve for a sample pressurizer resulting, in a failure probability curve. The failure probability curve was based on a t-distribution plotted below the failure probability curve calculated from the original GMPE. Calculating the areas under the failure probability curves resulted in failure frequencies of 3.74×10^{-7}/year for the normal distribution and 3.02×10^{-9}/year for the adjusted t-distribution. Although these calculations were based on a PGA range of 0.1–2 g, it is suspected that expanding this range would not

make a significant difference in the magnitude of the difference between the two failure frequencies. However, it should be noted that these results only correspond to the two fault scenarios proposed. It is highly probable that if the two faults had different characteristics, such as $M < 6$, then the degrees of freedom would increase due to a lack of data in that magnitude range, which might result in a t-distribution with larger tails and thus a potentially higher rate of exceeding a certain PGA. This would most likely increase failure frequencies.

Conflicts of interest

The author has no conflicts of interest to declare.

REFERENCES

[1] I.M. Idriss, An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, Earthq. Spectra. 24 (2008) 217–242.
[2] M. Power, B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, C. Roblee, An overview of the NGA project, Earthq. Spectra. 24 (2008) 3–21.
[3] B. Chiou, R. Darragh, N. Gregor, W. Silva, NGA project strong-motion database, Earthq. Spectra. 24 (2008) 23–44.
[4] Electric Power Research Institute (EPRI), Seismic Fragility Application Guide, EPRI, Palo Alto, CA, 2002.
[5] A methodology for analyzing precursors to earthquake-initiated and fire-initiated accident sequences, U.S. Nuclear Regulatory Commission, Report NUREG/CR-6544 (1998).