Alex Chansky
5/13/21

National Liquefaction Loss Model Uncertainty Quantification

The goal of this project was to quantify the uncertainty associated with fragility functions constructed for an MS thesis project. In the thesis project, event-level fragility functions were constructed to estimate probabilities of exceeding cost-based damage thresholds for earthquake liquefaction loss. This project also sampled input data, quantified uncertainty, and created confidence intervals along the fragility functions using a Monte Carlo simulation.

Fragility functions are a method of fitting lognormal cumulative distribution function (CDF) curves to a dataset by using maximum likelihood estimation (MLE) on the function's parameters. This project uses the empirical bounding-failure excitation method discussed in Porter (2020). This allows failure data based on excitation measures to provide an estimate of probability failure at every excitation level. In this case, the excitation measure is defined as spatial expectation of liquefaction in an event in square kilometers. These values were derived from Zhu et al., 2017, a former project in our research group, which produces an estimated percent of each cell (LSE) across a map expected to liquefy. The event's excitation measure ("Aggregate Liquefaction Hazard" or "*Htot*") was found by the summation of the product of each cell's LSE by its area, calculated in **figure 1**. Additional excitation measures such as population expected to be exposed to liquefaction will also be considered in future work.

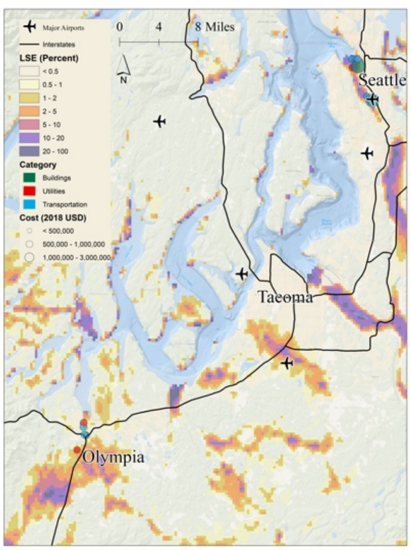

Figure 1: Htot (LSE) values for 2001 Nisqually event projected into NAD 1983 (2011) StatePlane Washington North FIPS 4601 (Meters). Htot values were calculated from this by the summation of each cell's area multiplied with its LSE value:

$$\mathrm{H}_{tot} = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{P}_{i,j}\mathrm{A}_{i,j} \text{ for } gm_{i,j} \geq gm_{thresh} \text{ and } \mathrm{P}_{i,j} \geq \mathrm{P}_{thresh}$$

Damage states of each event were determined by defining total liquefaction loss thresholds in **table 1**. Damage estimates are calculated from information in earthquake reconnaissance reports wherever liquefaction or its effects are mentioned as a damage factor. Estimation methodology is described in detail in Chansky and Baise (in progress), primarily based on cost value estimates in HAZUS MH-2.1 Technical Manual. 12 earthquakes are recorded with liquefaction loss damages and 34 events are recorded without liquefaction loss. "Observed probability of failure" for each damage state is the number of events per bin which exceed the damage state threshold divided by the total number of events in the bin, later shown in figure 2. Events are grouped to be included in the same bin with other events of similar excitation measures.

Damage State (DS)	Value Range (2018 USD)
0	< 10,000
1	>= 10,000
2	>= 1 million
3	>= 10 million
4	>= 100 million

Table 1: Damage state thresholds in 2018 USD for events with liquefaction loss estimates.

The lower ends of excitation measure (*Htot*) bin boundaries were determined by the equation, 10^(n/3), where *n* represents the bin number. For example, *Htot* bins were split apart at values 10^0, 10^0.333, 10^0.666, 10^1, 10^1.333, and so on until the maximum excitation measure is surpassed. The first bin's lower boundary was adjusted from 1 to -0.01 to include events with *Htot* values of 0 and above. *Htot* values have a minimum possible value of 0 and a maximum possible value of the total affected by an earthquake, determined as area with peak ground acceleration values by the USGS.

As previously mentioned, CDF curves are fit to the observed exceedance probability dataset using MLE. Two steps must be taken before MLE is conducted. First, the parameters used in the CDF are estimated. Their values do not matter much as the MLE process determines the most appropriate parameter values. However, if the estimated values are too far from the most appropriate values, occasionally the MLE process will go through too many trials to find appropriate values and will not succeed.

These parameters are used to calculate dependent values in the CDF curve from independent excitation measures, seen in equation 29 in Porter (2020). Theta and Beta represent the lognormal CDF parameters, median and lognormal standard deviation, while r_i represents the excitation measure at the center of each excitation measure bin. This results in p_i, or theoretical probability of failure at excitation measures at the center of each excitation measure bin.

$$p_i = \Phi\left(\frac{\ln(r_i/\theta)}{\beta}\right) \quad (29)$$

Next, the probability was calculated that at the excitation measure representing the center of each *Htot* bin, we observed the number of failures, f_i, among the number of events, n_i, the theoretical failure probability, p_i, using the binomial distribution in equation 30 of Porter (2020):

$$P[F_i = f_i] = \frac{n_i!}{f_i!(n_i - f_i)!} \cdot p_i^{f_i} \cdot (1 - p_i)^{n_i - f_i} \quad (30)$$

Finally, the product of the probabilities found at the center of each bin was maximized by allowing theta and beta to change using equation 31 of Porter (2020). In this case, MLE is a frequentist process to find quantities of interest (QOIs) theta and beta.

$$L(\theta,\beta)=\prod_{i=1}^{m_i} P\left[F_i=f_i\right] \quad (31)$$

With parameters theta and beta describing the lognormal CDF shape for each damage state, the CDF, or theoretical probability of exceeding the damage state threshold, can be plotted along the x-axis representing excitation measures (**figure 2**).

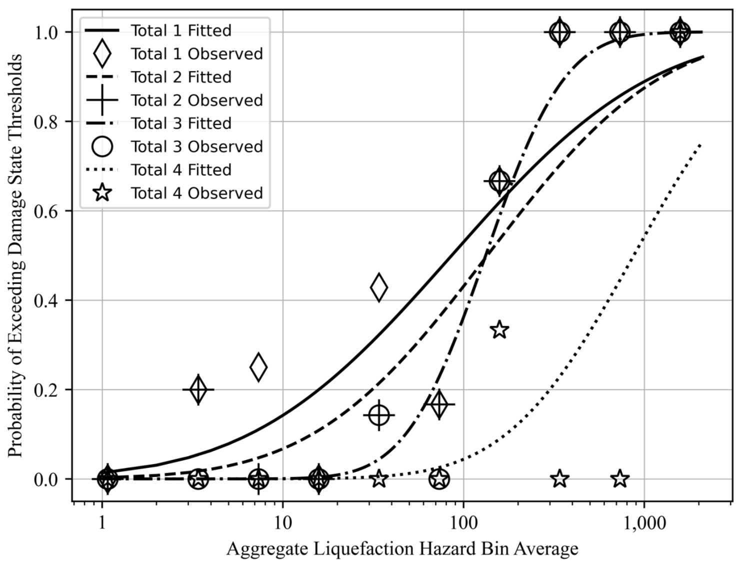

Figure 2: Probability of failure observations and resulting fragility functions using four liquefaction cost-based damage states found in table 1.

As there is some uncertainty associated with the *Htot* calculation, each event's *Htot* values used in **figure 1** were calculated as the medians of beta distributions, which were calculated and provided by Kate Allstadt of the USGS (Aug 2020, personal communication). In the next steps, a Monte Carlo simulation was used to assess uncertainty associated with the fragility function CDFs.

During Monte Carlo simulation, 1 sample was drawn randomly from the beta distribution for each event and assigned as its excitation measure. The CDF parameters were recalculated and recorded. If the excitation measure for an event changed enough to be included in a different bin, the calculated CDF parameters were estimated slightly differently. This was repeated 1,000 times. Distributions of excitation measure draws for the 2018 Anchorage event are displayed in **figures 3** and **4**.

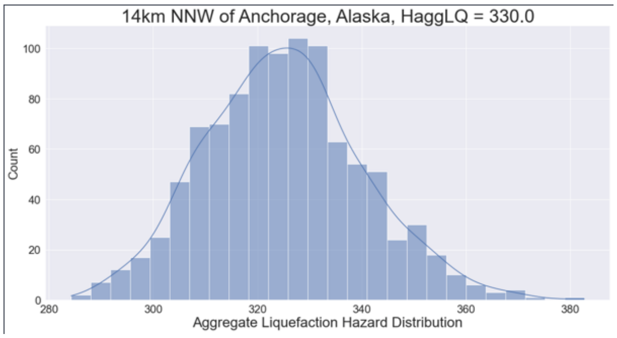

Figure 3: Histogram and probability distribution function of 1,000 draws from beta distribution of 2018 Anchorage event's aggregate liquefaction hazard.

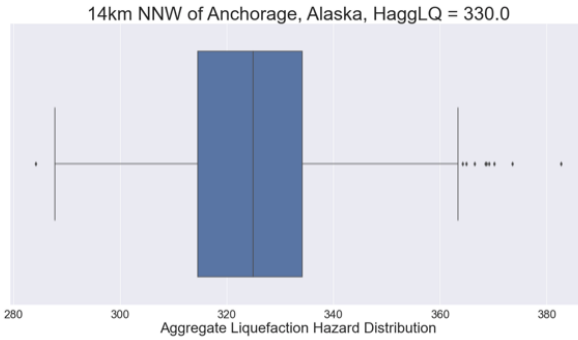

Figure 4: Box plot of 1,000 draws from beta distribution of 2018 Anchorage event's aggregate liquefaction hazard.

For iterations of sampling where CDF parameters were estimated slightly differently, the theoretical probability of exceeding damage state thresholds, or "failure" at each excitation measure was also slightly different. After 1,000 iterations, distributions of the theoretical probabilities were determined at each excitation measure (each x-value). For beginning of analysis of this data, probability distributions at a randomly selected excitation measure (292.40) were plotted in **figures 5**, **6**, and **7**.

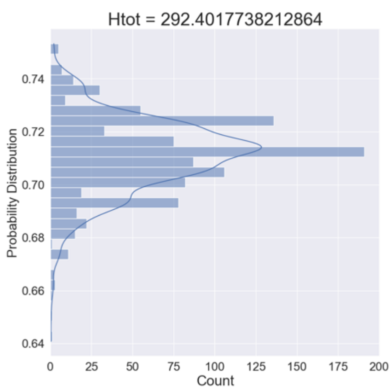

Figure 5: Histogram and distribution of resulting 1,000 fragility function probability values for DS 1 at Htot = 292.40 using 1,000 draws from each event's beta distribution.

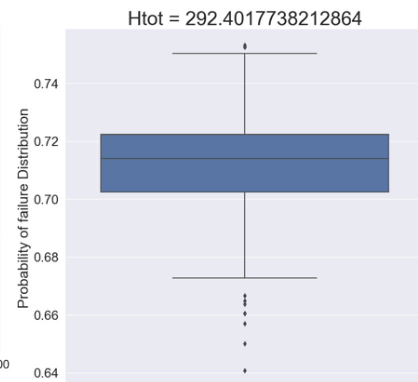

Figure 6: Box plot of resulting 1,000 fragility function probability values for DS 1 at Htot = 292.40 using 1,000 draws from each event's beta distribution.

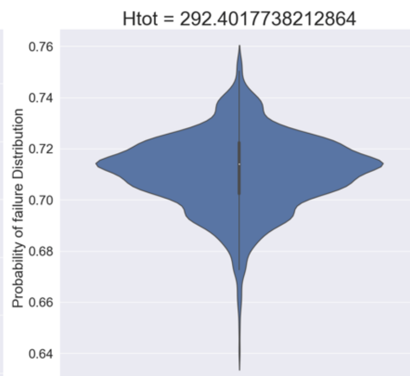

Figure 7: Violin plot of resulting 1,000 fragility function probability values for DS 1 at Htot = 292.40 using 1,000 draws from each event's beta distribution.

Using the probability distribution at each excitation measure, means and standard deviations were calculated. These were used to produce the plus or minus two standard deviation range of probabilities at each excitation measure, plotted in **figure 8**. These ranges contain approximately 95% of the 1,000 CDF values at each x-axis value for each damage state for a perfectly normal distribution. While the probability distribution is unlikely to be perfectly normal at each x-value, it is roughly normal as seen in figures such as 5, 6, and 7. Distributions on other areas of the x-axis produced similar figures.

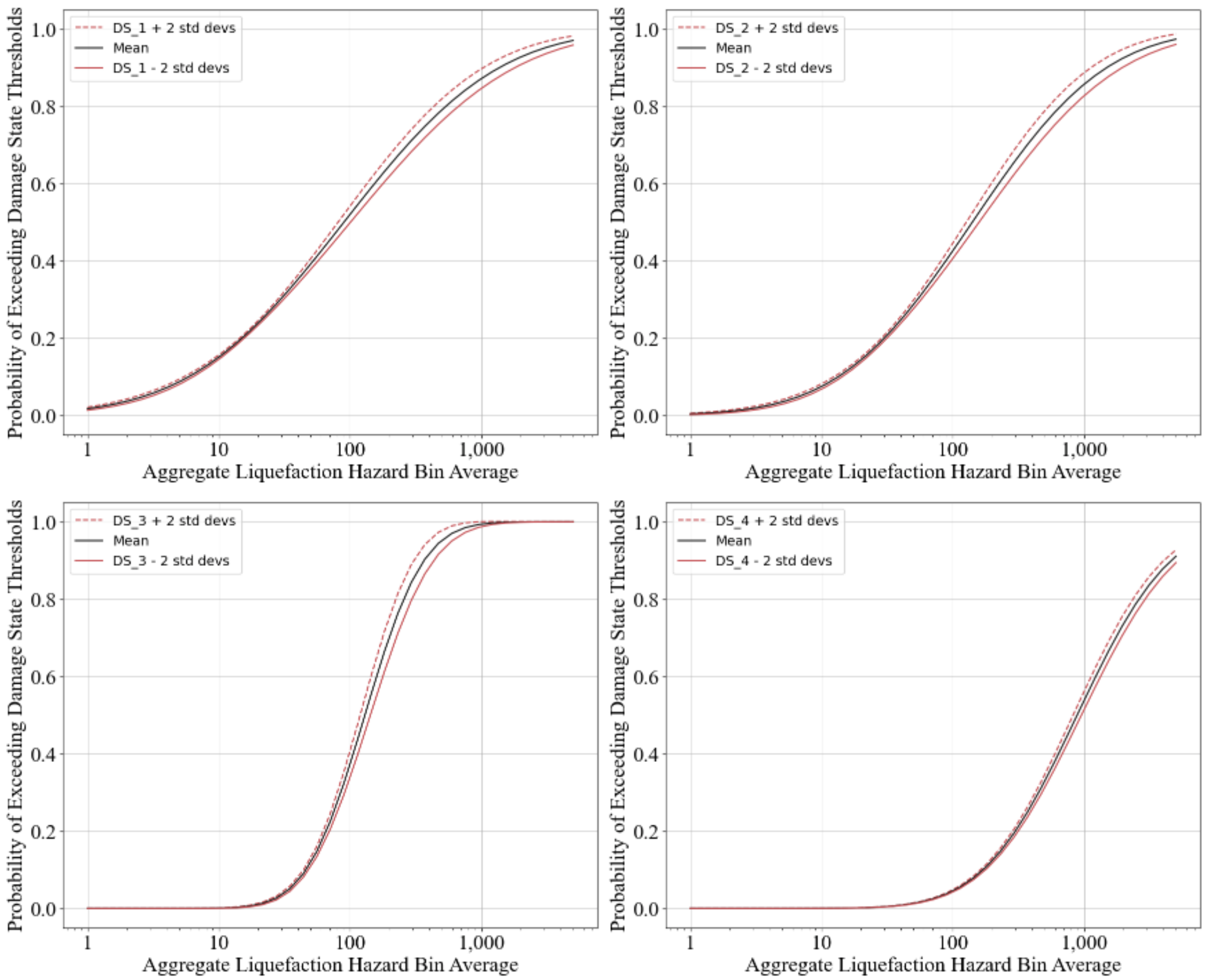

Figure 8: Mean and prediction intervals of +/- 2 standard deviations from the mean of all 1,000 resulting values along the x-axis. 95% of observations fall within two standard deviations of the mean in a perfect normal distribution.

Using the standard confidence interval equation (sample mean plus or minus the confidence level value multiplied by the standard error), confidence intervals were established at each excitation measure value using the probability distribution at that value. Confidence levels for each damage state are plotted in **figure 9**. The lines represent 97.5% and 2.5% probability that the true population mean lies below their values, meaning a 95% probability that the true mean population parameter falls between the lines.

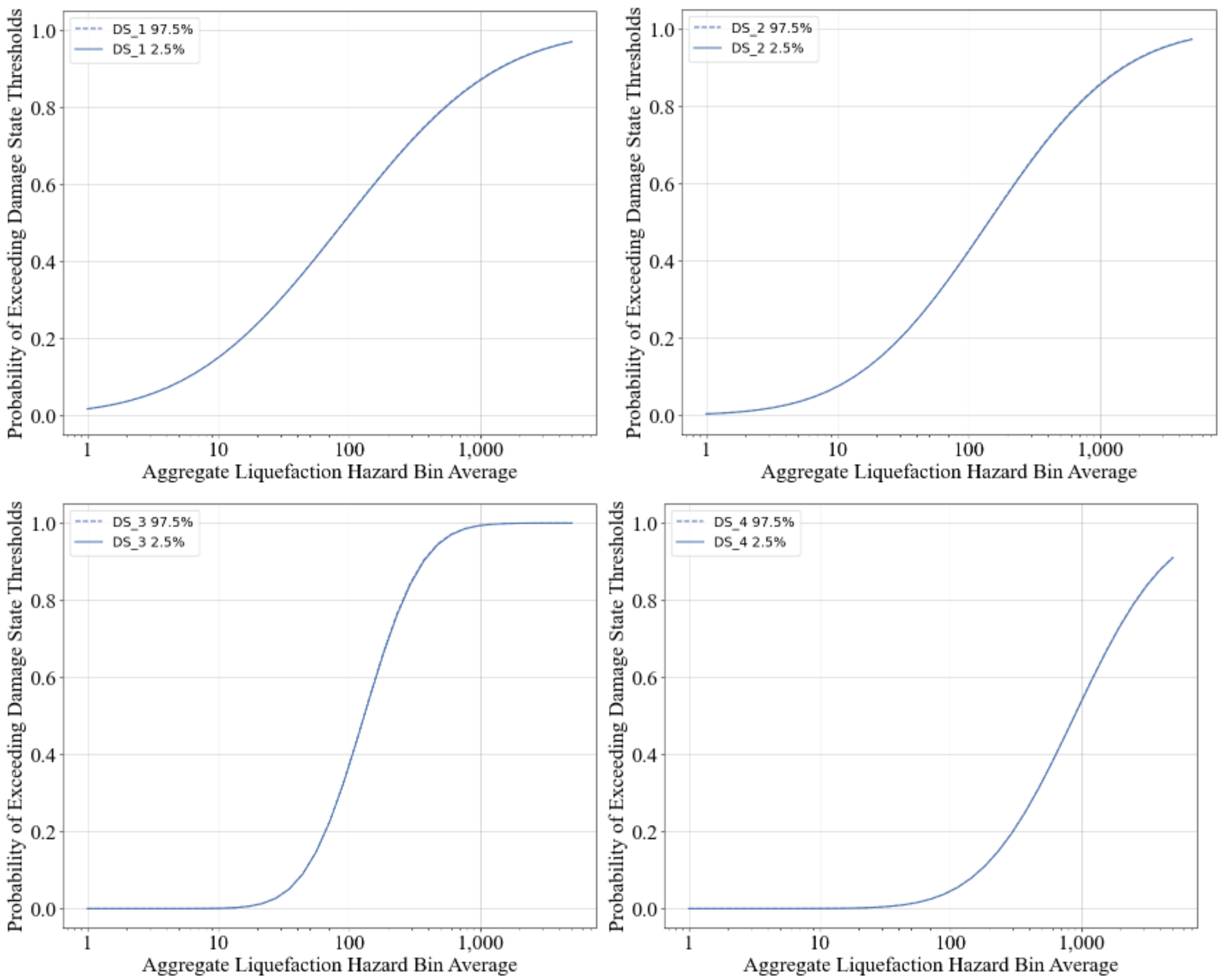

Figure 9: Confidence intervals of probability values along each fragility curve which have a 95% of containing the true probability population mean. The probability data did not have a very wide distribution at each excitation measure, resulting in a relatively narrow confidence interval.

Though excitation measure beta distributions allowed for a wide range of excitations selected for each event, most events remained in their original bins for most of the simulations. This resulted in not much change in a relatively narrow probability distribution and a very narrow confidence interval. The confidence intervals were so thin they were almost imperceptible. This implies a very high confidence on where the true mean population parameter falls. In other words, we can conclude almost exactly the theoretical probability of failure for each damage state along all excitation measure values.

For future work related to this project, these calculations will be repeated using other excitation measures such as estimated liquefaction population exposure. Another possibility is to incorporate new data in the form of new events and observe CDF parameter changes with a Bayesian process.

In this project, a thesis project was introduced which contained a liquefaction loss database for US events from 1964 – 2019 in addition to event-level fragility functions. The database was used to create damage state class thresholds by choosing round whole numbers, such as $10,000. Event-level fragility functions were created by using MLE to find lognormal CDF parameters. Monte Carlo simulations were employed to establish excitation measure distributions for each event and probability distributions along the independent variable axis of excitation measures. Finally, confidence intervals were determined along this distribution and observed for each damage state.

Notes on codes used:

A lot of code used in this project was originally written for a thesis project. All Jupyter Notebooks used are provided but much of the data organization and processing is less relevant so will not be emphasized. Points of interest to this project are (1) MLE process to determine CDF parameters, (2) Monte Carlo simulation and repeated MLE, (3) describe the data pulled from beta distributions, and (4) determining and plotting confidence intervals and data ranges. These four steps are described further in the ReadMe file for the following repository, with associated notebooks:

https://github.com/chansk/UncertProjNotebooks

Sources

- **Bird J.F., Bommer, J.J.** (2004). Earthquake losses due to ground failure. *Engineering Geology*. 75(2):147- 79.
- **Baise, L.G., Rashidian, V.** (2020). Regional Efficacy of a Global Geospatial Liquefaction Model. *Engineering Geology.* Vol. 272
- HAZUS Multi-Hazard Loss Estimation Methodology Technical Manual, Version 2.1, *Department of Homeland Security: Federal Emergency Management Agency,* Washington, DC, 2017
- **Porter, K.**, 2021. *A Beginner's Guide to Fragility, Vulnerability, and Risk*. University of Colorado Boulder, 136 pp., http://www.sparisk.com/pubs/Porter-beginners-guide.pdf.
- **Zhu, J., Daley, D., Baise, L., Thompson, E., Wald, D., and Knudsen, K.** (2015). A Geospatial Liquefaction Model for Rapid Response and Loss Estimation. *Earthquake Spectra,* 31.3, 1813-837.
- **Zhu, J., Baise, L., and Thompson, E.** (2017). An Updated Geospatial Liquefaction Model for Global Application. *Bulletin of the Seismological Society of America. Berkeley CA,* 107.3, 1365-385.