PhD Qualifying exam

Marshall Pontrelli

July 24, 2020

Civilization depends on the complex and chaotic natural systems from which humanity arose. While these natural systems support life, providing the human species an environment within which remarkable civilizations have developed complex economic and societal systems, they also cause destruction and catastrophe. Humanity's painfully slow and jerking progress is a function its ability to reduce these complex systems into simpler, understandable cause and effect models of reality that allow it to increase its predictive ability and thus minimize destruction and maximize creation. Despite the simplicity of these models in relation to actual nature, this strategy has been remarkably successful. The key to understanding natural systems and thus increase the likelihood of the progress of humanity is to ask and systematically strive to answer the questions: how does order arise from chaos and what are the laws that underpin reality?

The earth is a complex environment made from many interacting dynamic systems all of which provide input to the anthropogenic systems necessary for human survival. The interests of this exam are the causes of the plate tectonic system on human systems. Plate tectonics is a model of the crust's dynamism creating earthquakes in the form of energy propagating as seismic waves. These waves go through many processes which morph them as they move through space and time until they reach human construction at the surface of the earth and interact with civil infrastructure. This interaction causes nuisance and discomfort when and where societies are prepared and death and destruction when and where they are unprepared.

This exam is composed of four disciplines: earthquake engineering, environmental statistics, digital signal processing and geomechanics. Earthquake engineering describes the set of problems that arise from earthquake wave propagation and its interaction with human infrastructure. These problems are aptly solved using combinations of tools from statistics and digital signal processing to analyze vibration data from the earth or in civil systems. These data analyses fall within a theoretical framework provided by analytical solutions in geomechanics describing material properties, wave propagation and the relevant parameters therein. Solving these problems 1) moves forward our understanding of how seismic energy affects humanity, 2) quantifies the interaction between seismic energy and civil infrastructure and 3) develops an ordered framework describing a chaotic system thus allowing for prediction useful to the advancement of human society.

Part 1: Earthquake Engineering

Marshall Pontrelli

July 24, 2020

Regional and site specific analysis of the Anchorage basin using HVSR and response spectra

Introduction

The city of Anchorage lies on the west coast of Alaska in the upper Cook inlet, a 180-mile-long bay extending inland from the Gulf of Alaska. It is within a geographic band with high seismicity and volcanism caused by the subducting Pacific plate beneath the North American plate. The City itself sits on a sedimentary basin roughly 16 by 16 km bounded by the Chugach mountains on the east, the Knik arm to the northwest and the Turnagain arm to the Southwest. Both Knik and Turnagain arms are part of the Cook inlet and Anchorage is the point at which the Cook inlet bifurcates into the respective arms (figure 1). It is one of many sedimentary basins bounded by the coastal mountain chain of western Alaska and holds the city home to roughly 40% of Alaska's total population. The basin is unbounded on the west where it meets the Cook inlet and has on this coast been subject to tidal fluctuations and estuarian environments throughout Quaternary age (Schmoll et al. 1999). This environment created a structure of marine deposits interbedded with glacial deposits coming off the Chugach mountains. The coast abutting the Knik and Turnagain arms saw extensive damage in the 1964 Good Friday earthquake, including the famous liquefaction event at Turnagain Heights (Nath et al. 2002). The basin halfspace slopes from the Chugach mountains on the east into the Cook island inlet on the west (figure 2b). Anchorage is in a tectonically active environment, sits in a basin with amplifiable soils and has a large population, thus, it is essential to quantify and understand its site response characteristics to ensure safety for the City and its residents.

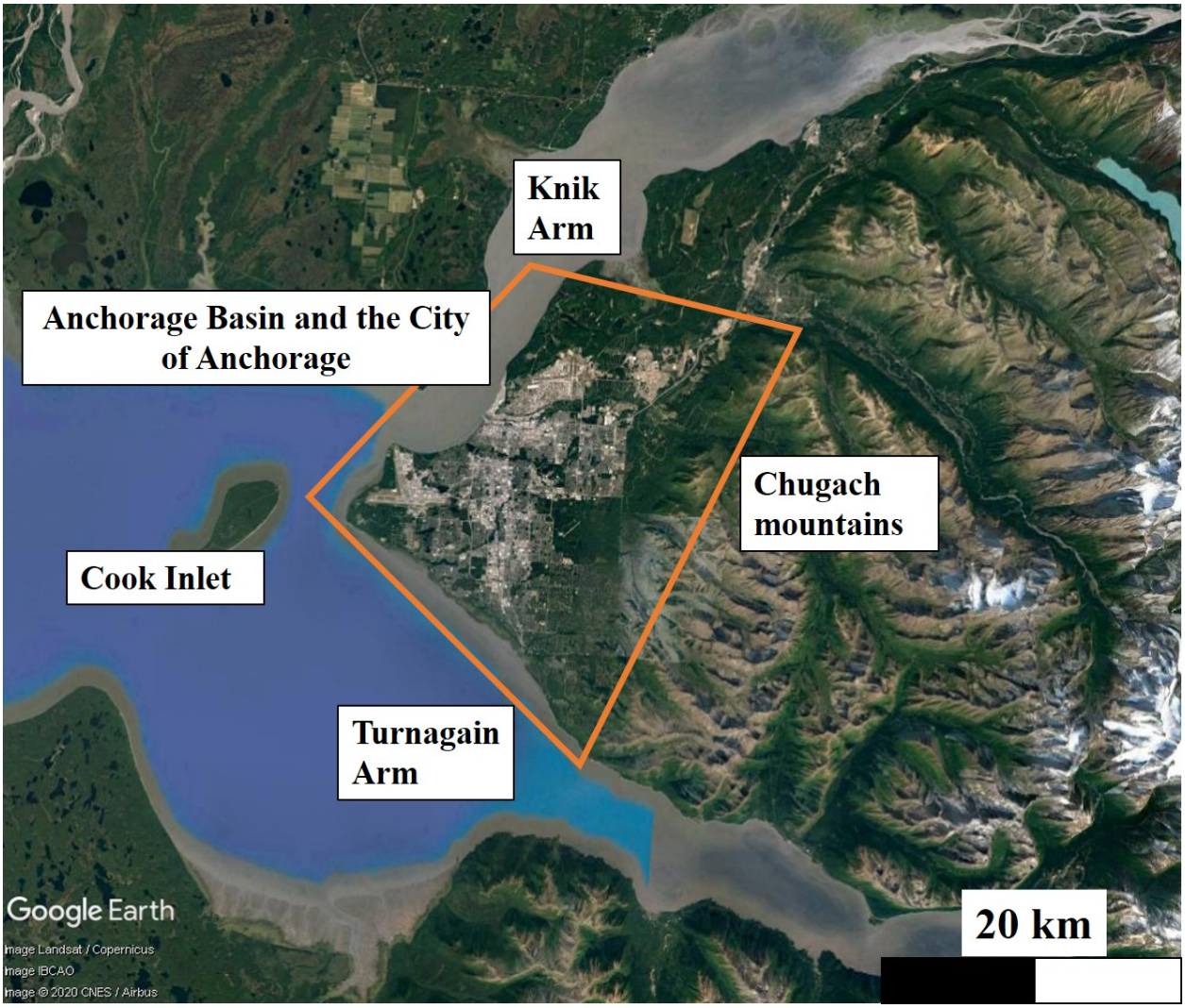


Figure 1: Map view of the Anchorage basin and surrounding area.

The glacial-marine structure forming the Anchorage basin has many subdivisions of soil units, most notable of which is the Bootlegger Cove formation, a unit highly susceptible to ground failure and which underlies most of metropolitan Anchorage (Nath et al. 2002). There are several other surficial units of glacial and marine derivation in Anchorage and for these, we refer the reader to Yehle and Schmoll, 1987. The Bootlegger Cove Formation is at the west end of the Anchorage basin abutting the Knik and Turnagain arms (figure 2a). It is composed of marine and estuarian clay and silt deposits interbedded with fluvial and glaciofluvial sand. This formation and thus its location within the City are susceptible to ground failure and experienced extensive failure in the 1964 Good Friday Earthquake (Martirosyan et al. 2002). Studies have shown that this formation has a low fundamental frequency, high 1 Hz amplification, low Vs_{30} and an NEHRP D site classification ((Martirosyan et al. 2002, Nath et al. 2002).

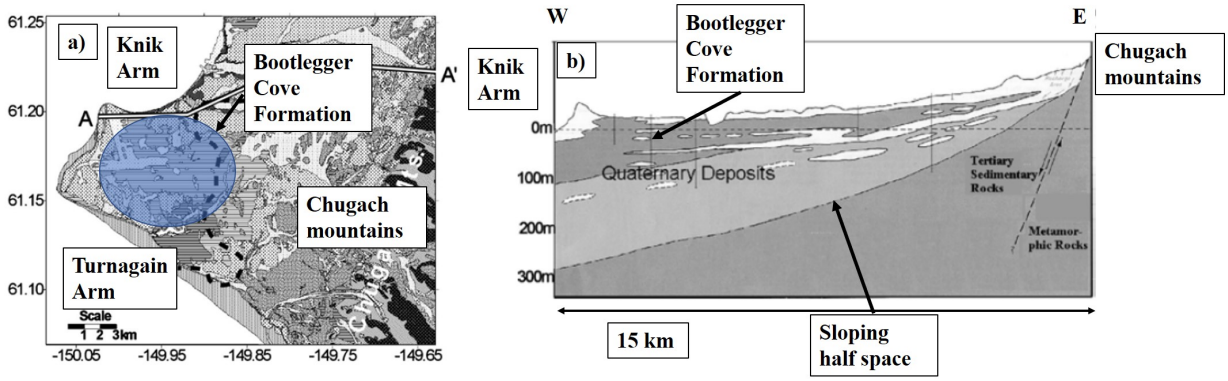


Figure 2: a) Basic surficial and bedrock geology in the Anchorage basin with Bootlegger Cove formation indicated. b). West-east transect of the Anchorage basin with sloping halfspace and Bootlegger Cove formation indicated. (Modified from Martirosyan et al. 2002).

The Anchorage basin has had extensive site response analyses performed on it due to its location in an area of high seismicity and the City's increase in growth in the last 50 years. We will summarize the results of two site response studies, Martirosyan et al. 2002 and Nath et al. 2002, each using weak motion data (avoiding non-linear effects) within a database of local and regional events from both strong and weak ground motion sensors. We look at each group's results for 1) fundamental resonance in the Anchorage basin from HVSR analysis (Nakamura, 1989), 2) site amplification at 1 Hz derived from both simple spectral ratios and HVSRs and its mapping it across the basin and 3) compiled Vs_{30} measurements and NEHRP site classification mapping within the basin. The fundamental resonances derived from HVSR analysis indicate resonance from 3-4 Hz on the east side of the basin at the edge of the Chugach mountains to around 1 Hz on the west end of the basin at the edge of the Knik arm in the Bootlegger Cove Formation (figure 3, Martirosyan et al. 2002). These results map onto the sloping depth to bedrock off the Chugach mountains into the inlet using the common relationship

$$f_0 = \beta/4d \tag{1}$$

Where β is the shear wave velocity and d is the depth to the halfspace. This relationship shows that as depth increases, f_0 decreases as it does in the Anchorage basin (figure 2b and figure 3 respectively).

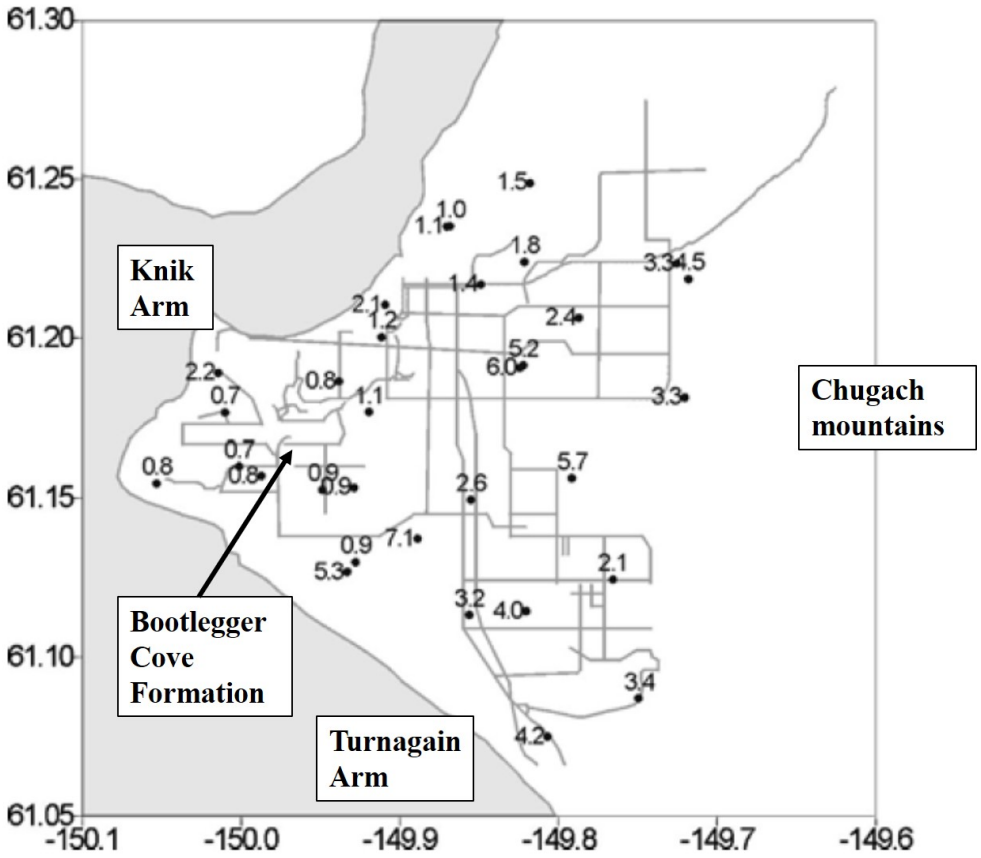


Figure 3: Fundamental resonance of the Anchorage basin derived from HVSR analysis (Martirosyan et al. 2002).

Fundamental resonance is inherently linked to depth and shear wave velocity (equation 1) and is thus also linked to site amplification. Site amplification is a function of the impedance contrast, which is a ratio of velocity and density of the overburden to the velocity and density of basement rock (equation 2).

$$\frac{\nu_s}{\nu_b} = \frac{\rho_s \beta_s}{\rho_b \beta_b} \tag{2}$$

Where ν_s and ν_b are the displacements at the surface and bedrock interface respectively, ρ_s and ρ_b are the densities of the overburden and bedrock respectively and β_s and β_b are the shear wave velocities of overburden and bedrock respectively (Haskell, 1960; Kramer, 1996). A decrease in fundamental resonance often indicates an increase in amplification because it can mean a decrease in overburden shear wave velocity. We can measure site amplification using techniques like the HVSR and the simple spectral ratio (Borcherdt 1970) with reference to a site on bedrock unaffected by site effects (Steidl et al. 1996). The researchers in Martirosyan et al. 2002 use a reference site in the Chugach mountains to compute the average of multiple spectral ratios at a station. They then average the site amplification derived from the simple spectral ratio with that derived from HVSR to come up with the site amplification factor at each frequency of the transfer functions of the SSR and HVSR. They compute an amplification factor at 1 Hz of around 3 on the west end of the basin in the Bootlegger Cove formation. This factor decreases to unity in the Chugach mountains (figure 4b). Their results are consistent with the depth to basement rock increase in figure 2b and the shear wave velocity decrease in Figure 5b. The researchers in Nath et al. 2002 compute amplifications using just HVSR. Their results show higher 1 Hz amplifications than those published in Martirosyan et al. 2002 by

about a factor of two (figure 4a), a trend that was also observed in Mexico City in Pontrelli et al. 2019 when comparing HVSR to SSR. Their HVSR amplifications concentrate on the northwest side of the basin adjacent to the Knik arm. With this exception, the overall pattern remains similar to the results of the other group.

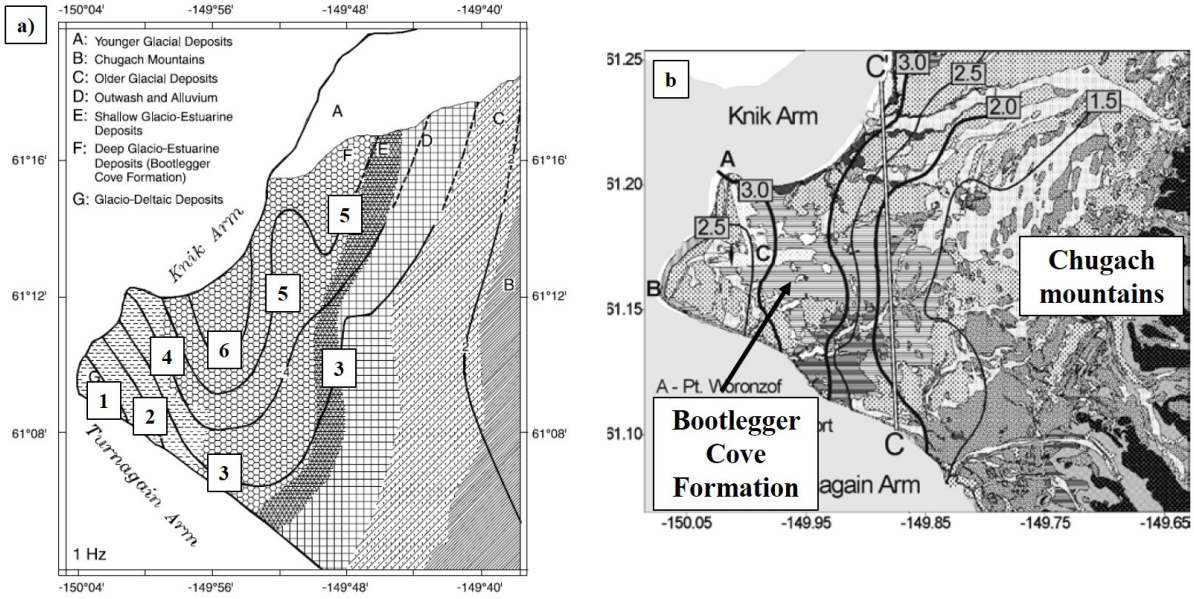


Figure 4: a) Spectral amplification at 1 Hz of the Anchorage basin computed from HVSR analysis (Nath et al. 2002). b) Spectral amplification at 1 Hz computed from average of SSR and HVSR analyses (Martirosyan et al. 2002).

After computing fundamental resonance and low frequency amplification, the researchers in Martirosyan et al. 2002 compile Vs_{30} measurements across the city and sort the sites by their NEHRP site classifications. Their results show that the basin is composed of site class C ($360m/s < Vs_{30} < 760m/s$) and site class D ($180m/s < Vs_{30} < 360m/s$) and they posit that the area at the base of the Chugach mountains is a site class B with Vs_{30} values (Vs_{30} values $760m/s < Vs_{30} < 1500m/s$), greater than those within the basin (figure 5b). The researchers identify a "transition zone" between site classes C and D. The researchers in Nath et al. 2002 indicate similar site classifications (figure 5a). In general, in the Anchorage basin, Vs_{30} measurements follow a similar trend to fundamental resonance, basin slope and 1 Hz frequency amplification with increasing site class (decreasing Vs_{30}) from the east at the base of the Chugach mountains to the west in the Bootlegger Cove formation and edge of land (figures 5a and b).

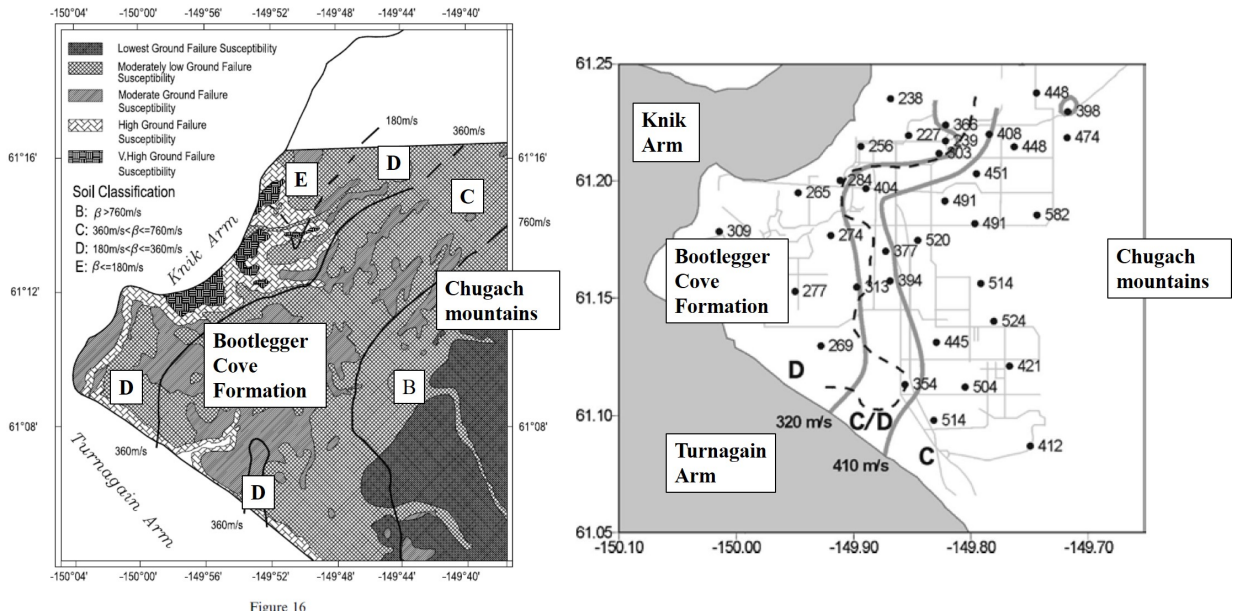


Figure 5: NEHRP site classification of the Anchorage basin derived Vs_{30} measurements (Martirosyan et al. 2002).

Analysis of November 30, 2018 Anchorage Earthquake

The magnitude 7.1 November 30, 2018 Anchorage earthquake occurred about 16 kilometers north of Anchorage due to normal faulting at a depth of 47 km (USGS event page). The earthquake caused 117 injuries, severe damage and cases of liquefaction across metropolitan Anchorage and the surrounding land totaling an estimated $25 - $50 million in damage. In this analysis, we use waveforms recorded at 28 stations within the Anchorage basin to measure site amplification using HVSR analysis and correlate the results to those published in Martirosyan et al. 2002 and Nath et al. 2002. We use these results as a regional context for a site analysis of station NP 8040 located in an area deemed by the Municipality of Anchorage as having "Very high ground failure susceptibility. We compute a theoretical transfer function at the site using the estimated shear wave velocity profile for the location and compare it to HVSR analysis results. We also classify this site as an IBC 2012 site class D, compute its design spectra and compare the design spectra to its response spectra of the Anchorage event.

Data

For the regional portion of this study, we use ground motion records at 28 strong motion accelerometer stations, 14 stations in the Alaska regional network and 14 stations in the United States National Strong-Motion Network (table 1, figure 6b). We pull three components of data from each station between the time window from 17:29:30 to 17:31:05 UTC. The relatively short duration of the event is due to the proximity of the epicenter to the City. For the site-specific portion of the study, we use the shear wave velocity map (figure 5b) from Martirosyan et al. 2002 to estimate a shear wave velocity profile and select our site of interest using the Seismic Hazard map provided by the Municipality of Anchorage

Network	Station	Latitude	Longitude
AK	K203	61.220	-149.745
AK	K204	61.176	-150.012
AK	K205	61.199	-149.916
AK	K208	61.176	-149.922
AK	K209	61.185	-149.747
AK	K210	61.129	-149.931
AK	K211	61.149	-149.858
AK	K213	61.113	-149.859
AK	K215	61.086	-149.752
AK	K216	61.098	-149.687
AK	K220	61.154	-150.055
AK	K221	61.153	-149.951
AK	K222	61.088	-149.837
AK	K223	61.234	-149.867
NP	8011	61.209	-149.786
NP	8021	61.113	-149.910
NP	8025	61.147	-149.894
NP	8027	61.161	-149.889
NP	8028	61.193	-149.782
NP	8029	61.174	-149.850
NP	8030	61.180	-149.806
NP	8036	61.178	-149.966
NP	8037	61.156	-149.985
NP	8040	61.213	-149.893
NP	8043	61.222	-149.885
NP	8047	61.189	-149.802
NP	ABBK	61.114	-149.716
NP	ALUK	61.103	-149.816

Table 1: Stations used in this study. Network "AK" is the Alaska regional network and network "NP" is the United States National Strong-Motion Network and are plotted in figure 6b.

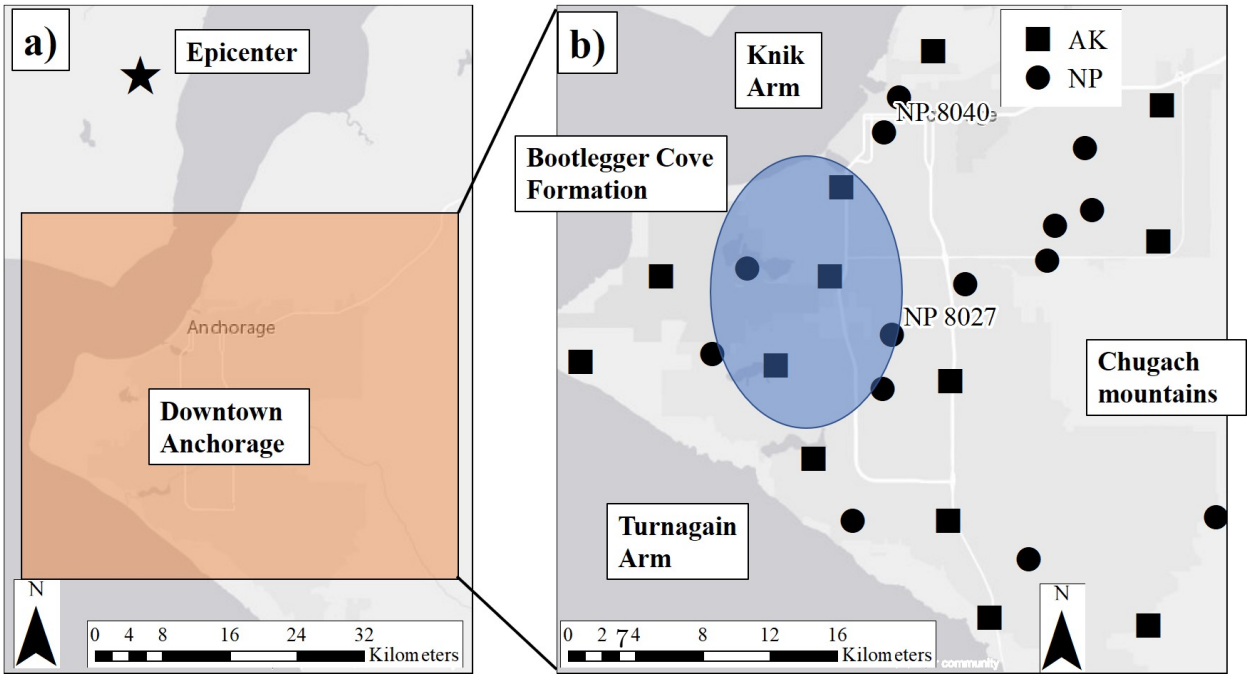


Figure 6: a) Epicenter of the 2018 Anchorage earthquake and its relation to metropolitan Anchorage. b) Stations used in this study. The time series and HVSR of the two labeled stations are presented in figure 7.

Methods

For the regional part of the study analyzing the entire Anchorage basin, we pull three components of data from 28 stations off the IRIS DMC database and then process it using Nakamura's HVSR. The data are first filtered using a 4 pole, zero phase bandpass Butterworth filter with lowcorner frequency 0.1 Hz and highcorner frequency equal to the sampling frequency fs/2 - 1, or 1 Hz below the Nyquist frequency. The highcorner serves as an anti-aliasing filter. We convert the data from counts into m/s^2 using the sensor sensitivity and then divide by g to get the data in g's and compute the PGA of each component (Figure 7). We window the data using a Hanning window the length of the data vector and compute the normalized Fourier half magnitude spectra correcting for the amplitude decrease of the Hanning window. We then smooth each component's half magnitude spectra and use a zero-phase moving average filter with a width of 0.5 Hz. We combine the horizontal components by computing their geometric mean and then compute the HVSR by dividing the combined horizontal component by the vertical component. Finally, we solve for the maximum peak amplitude and frequency and the 1 Hz amplitude of each HVSR at the 28 stations (figure 7).

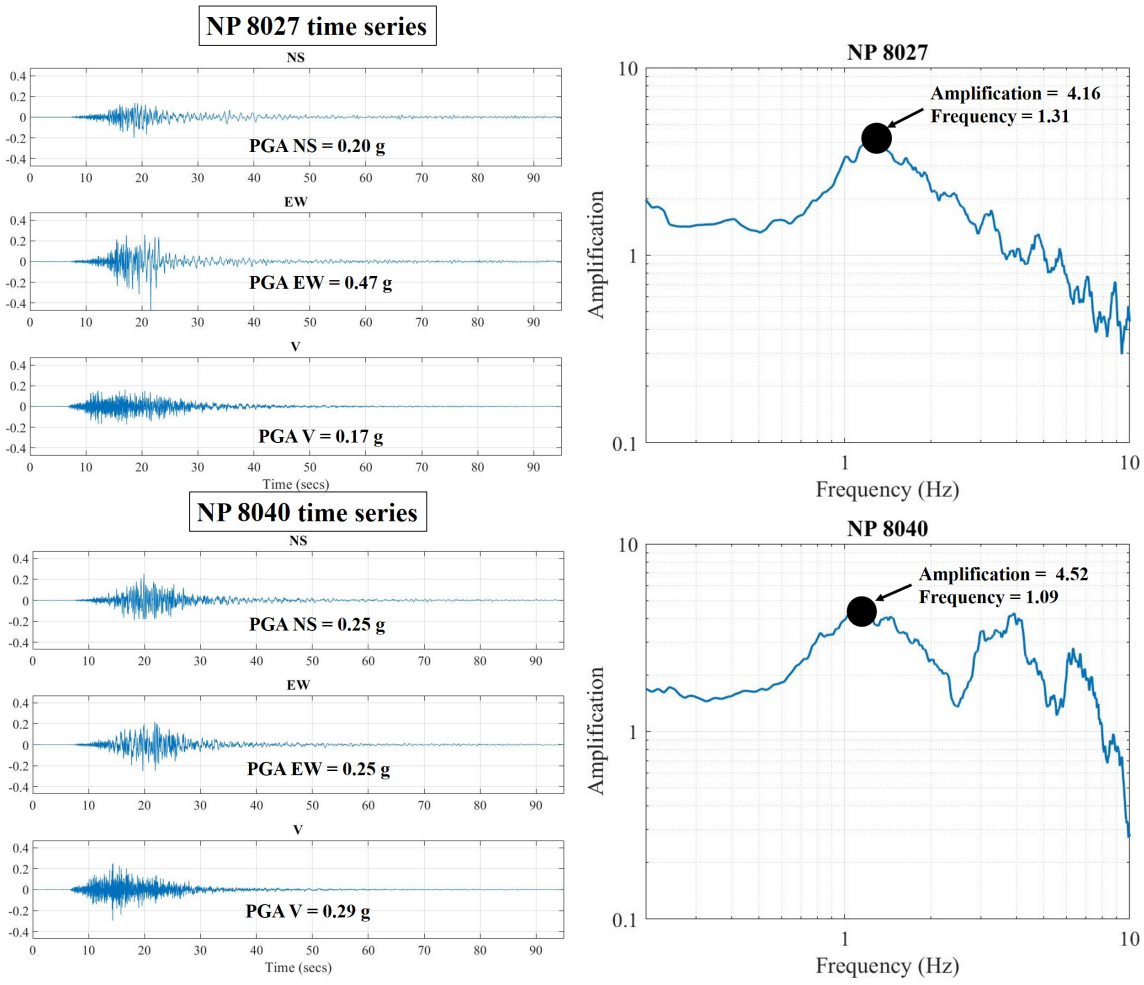


Figure 7: Waveforms from the Anchorage earthquake at two selected sites from within the Bootlegger Cove formation (NP 8027) and coastal susceptible land (NP 8040) and their corresponding HVSRs (figure 6). PGAs are indicated as are max amplitude and fundamental frequency.

For the site specific part of the study, we select a location in an area defined as "Very high ground failure susceptibility" by the Municipality of Anchorage. We compute a theoretical transfer function for the site using the Nrattle Fortran routine, which calculates the Thomson-Haskell plane SH-wave transfer function (Thomson, 1950; Haskell, 1953) for horizontally stratified, laterally homogenous layers with vertically propagating shear waves (written by C. Mueller with modification by R. Herrmann and distributed in the Boore (2005) SMSIM ground motion simulation program). We refer to this as the "SH1D" transfer function. The input parameters for Nrattle are a shear wave velocity profile (β), and corresponding depths (d), densities (ρ) and attenuations (Q) of the overburden; and the density and shear wave velocity of the basement. The outputs of Nrattle are the TTF amplification as a function of frequency in Hz. We estimate a shear wave velocity profile using the shear wave velocity map in figure 5 and add a weathered layer of higher shear wave velocity before a basement layer of 1000 m/s. We assign a damping value of 2.5% and densitiy values of $1.8g/cm^3$ for the overburden and weathered layer. We test values of depth until the fundamental resonance of the TTF matches that of the HVSR and then look for other similarities between the two.

We then use the IBC 2012 design ground motion routine to develop a design ground motion for a site class "D" site in Anchorage and compare the design ground motion to the response spectra at site NP 8040. The IBC 2012 design routine starts by selecting S_s and S_1 parameters from the USGS Maximum Credible Earthquake (MCE) maps. S_s is the risk-targeted spectral acceleration at a period of 0.2 seconds on bedrock and S_1 is the risk-targeted spectral acceleration at a period of 1.0 second for bedrock. Using the USGS MCE maps, and as was available on the Municipality of Anchorage website, we got values for S_s and S_1

$$S_s = 1.5 \tag{3}$$

$$S_1 = 0.55 \tag{4}$$

We next select the site coefficients F_a and F_v for our values of S_s and S_1 for a site class D site and get values

$$F_a = 1 \tag{5}$$

$$F_v = 1.5 \tag{6}$$

We then compute S_{MS} and S_{M1}

$$S_{MS} = F_a S_s \tag{7}$$

$$S_{MS} = 1.5 \tag{8}$$

$$S_{M1} = F_v S_1 \tag{9}$$

$$S_{M1} = 0.8250 \tag{10}$$

We then compute design spectral response parameters S_{DS} and S_{D1}

$$S_{DS} = \frac{2}{3} S_{DS} \tag{11}$$

$$S_{DS} = 1 \tag{12}$$

$$S_{D1} = \frac{2}{3} S_{M1} \tag{13}$$

$$S_{D1} = 0.55 \tag{14}$$

Finally, we compute our design response spectra for a site class D in Anchorage looping over periods from 0 to 2 seconds using

$$T_0 = 0.2 * \frac{S_{D1}}{S_{DS}} \tag{15}$$

$$T_0 = 0.11 \tag{16}$$

$$T_S = \frac{S_{D1}}{S_{DS}} \tag{17}$$

$$T_S = 0.55 \tag{18}$$

$$T_L = 16 \tag{19}$$

and for periods less than T_0

$$S_D = S_{DS}(0.4 + 0.6\frac{T}{T_0} \tag{20}$$

for periods greater than or equal to T_0 and less than or equal to T_s

$$S_D = S_{DS} \tag{21}$$

and for periods greater than T_s and less tha or equal to T_L

$$S_D = \frac{S_{D1}}{T} \tag{22}$$

and for periods greater than T_L

$$S_D = \frac{S_{D1}T_L}{T^2} \tag{23}$$

This design response spectra is shown in figure 13. Finally, we look at how our results compare to those found in a similar AIR study.

Results

i) Regional study and comparison to older studies

At the 28 stations we used within the Anchorage basin, we compute all 3 component PGAs, the HVSR peak amplification, the HVSR fundamental frequency and the 1 Hz amplification (table 2) and map the results, comparing them to the results in Martirosyan et al. 2002 and Nath et al. 2002. The PGA values for the Anchorage event are in general around 0.2 - 0.3 g's putting the event in the region of portential soil non-linearity due to high strain. The horizontal PGAs tend to be larger than the vertical PGA at each station, but there are some stations with vertical PGAs comparable to horizontal PGAs (figure 8).

Network	Station	PGA V (g)	PGA EW (g)	PGA NS (g)	Amplification	Frequency (Hz)	1 Hz amplification
AK	K203	0.231	0.295	0.252	5.06	3.41	1.15
AK	K204	0.171	0.165	0.185	2.61	1.60	2.08
AK	K205	0.183	0.264	0.242	5.09	2.05	2.55
AK	K208	0.182	0.263	0.219	5.89	1.94	2.49
AK	K209	0.159	0.190	0.183	4.26	0.59	1.83
AK	K210	0.155	0.237	0.241	4.31	2.36	2.69
AK	K211	0.267	0.463	0.360	3.82	1.94	1.94
AK	K213	0.241	0.292	0.337	8.72	3.04	1.66
AK	K215	0.356	0.323	0.564	4.35	2.61	2.47
AK	K216	0.207	0.305	0.286	4.85	8.57	1.28
AK	K220	0.367	0.254	0.326	2.05	0.37	1.91
AK	K221	0.304	0.242	0.205	4.99	2.76	3.52
AK	K222	0.175	0.252	0.204	3.30	0.74	1.92
AK	K223	0.160	0.269	0.180	4.22	2.33	2.74
NP	8011	0.245	0.330	0.180	3.24	1.85	0.94
NP	8021	0.080	0.123	0.114	4.22	3.05	2.17
NP	8025	0.094	0.266	0.166	7.11	3.12	1.62
NP	8027	0.174	0.474	0.197	4.16	1.32	3.32
NP	8028	0.112	0.210	0.157	4.46	2.27	1.41
NP	8029	0.099	0.239	0.239	6.75	3.36	1.72
NP	8030	0.155	0.227	0.293	5.31	3.48	1.84
NP	8036	0.254	0.276	0.412	5.40	2.45	1.65
NP	8037	0.227	0.278	0.361	3.80	0.94	3.50
NP	8040	0.296	0.254	0.251	4.52	1.09	3.94
NP	8043	0.634	0.392	0.440	8.18	3.37	3.68
NP	8047	0.191	0.288	0.404	5.62	4.41	1.60
NP	ABBK	0.237	0.649	0.835	5.07	4.52	1.43
NP	ALUK	0.391	0.391	0.309	6.32	1.64	2.26

Table 2: Values computed for the Anchorage earthquake.

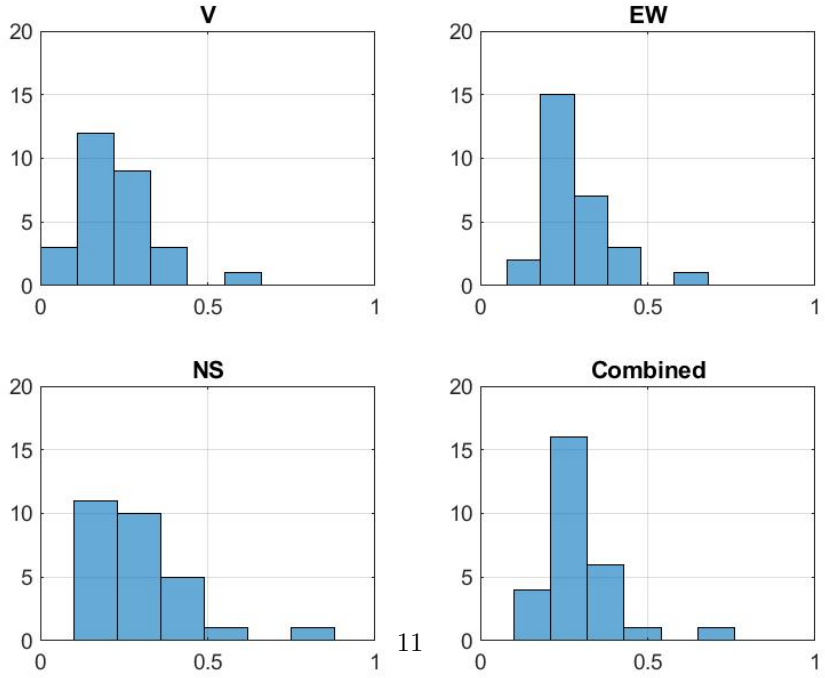


Figure 8: Histograms of the PGA at the three components of all 28 stations in g's computed and the mean of the horizontal components.

The fundamental resonances we computed using HVSR analysis decrease from the east at the base of the Chugach mountains to west in the Bootlegger Cove formation at the edge of the Knik and Turnagain arms. This is the expected result as the overburden-bedrock interface depth decreases along this trend (figure 2b). We expect that the fundamental resonances we computed in this study are slightly lower than those computed in Martirosyan et al. 2002 because the PGAs indicate that the event ground motion is in the non-linear range. Nonlinearity causes a reduction in shear modulus which in turn causes a reduction in shear wave velocity and thus a decrease in fundamental resonance. We plotted our frequencies on the frequencies computed in Martirosyan et al. 2002 to see if we could find evidence for nonlinearity (figure 9). The results are inconclusive because of the varying locations our respective stations, however, the fundamental resonance pattern, decreasing from east to west, is the same in both studies.

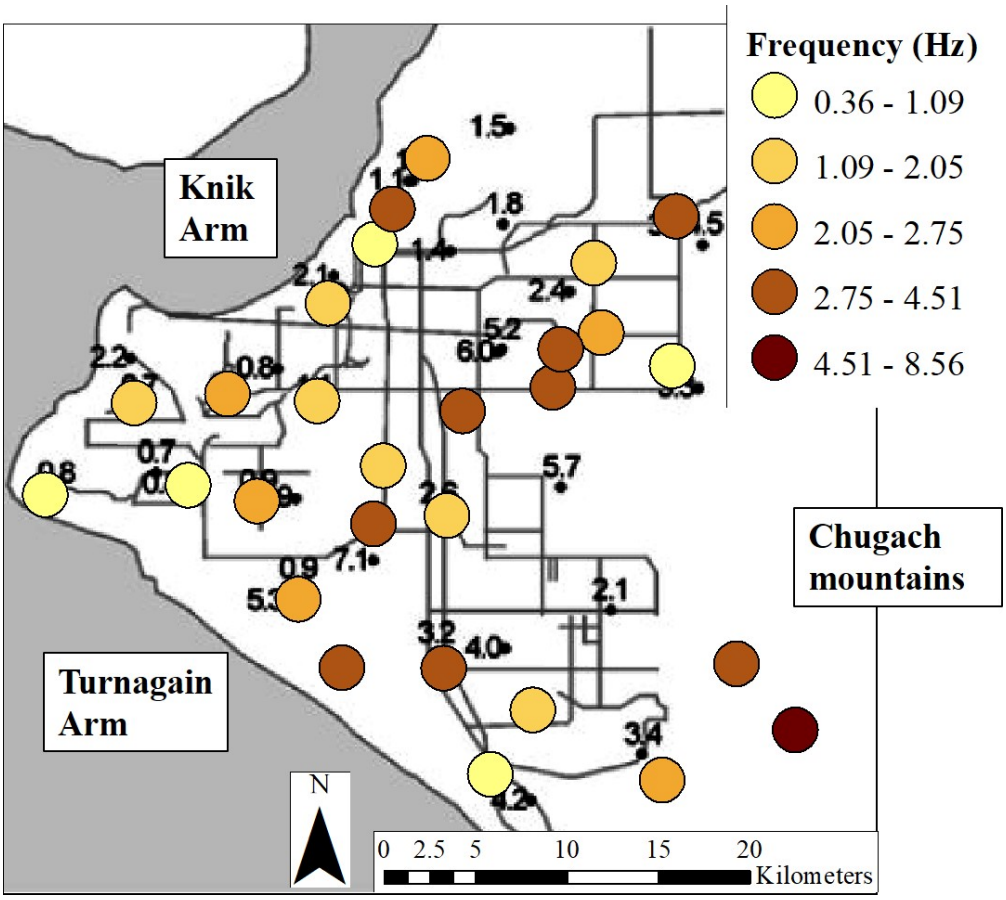


Figure 9: Fundamental resonance computed using HVSR analysis on data from the November 30, 2018 Anchorage Earthquake overlain on HVSR fundamental frequency map from Martirosyan et al. 2002.

Our HVSR amplifications computed at 1 Hz show a maximum amplification in the band of the Bootlegger Cove formation in the center-west of the basi (figure 10 a and b). This band of high 1 Hz amplification increases towards the Knik arm, which is the area the Municipality of Anchorage defines as an area of "very high ground failure susceptibility" (figure 11). The pattern of our results is like that of Nath et al. 2002 with increasing 1 Hz amplification towards the Knik arm (figure 10a). Our 1 Hz amplifications, however, are lower than those obtained in Nath et al. 2002. The 1 Hz amplifications computed by averaging simple spectral ratios and HVSRs in Martirosyan et al. 2002 are similar in magnitude to our results but have a lower resolution spatial pattern. Our results have a similar spatial pattern to Nath et al. 2002 and similar magnitude to Martirosyan et al. 2002.

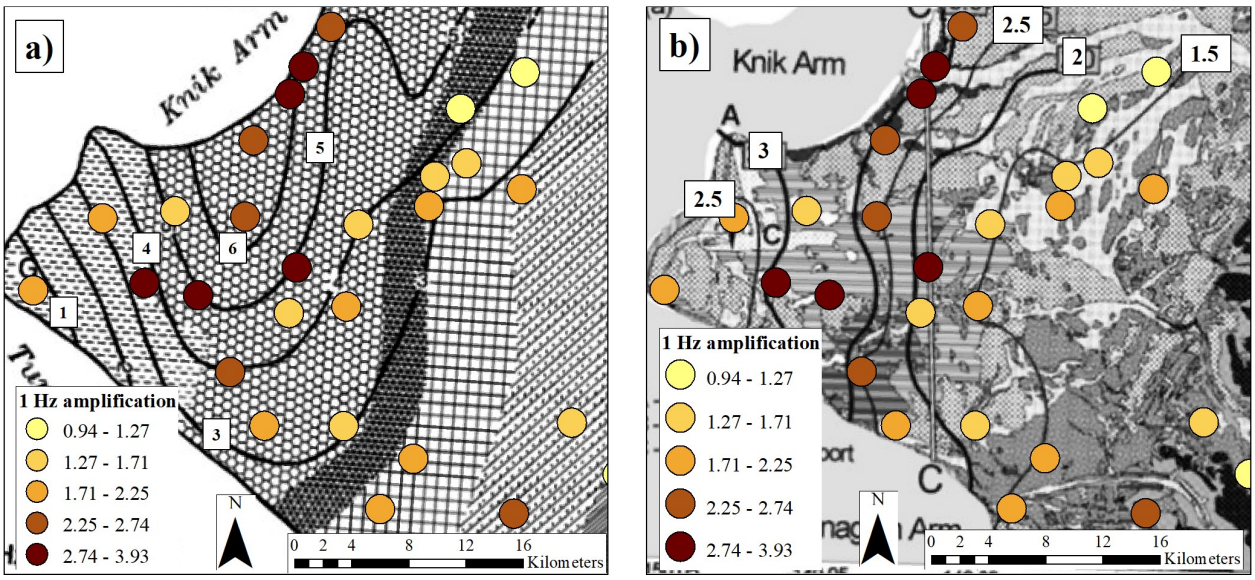


Figure 10: HVSR amplification of the November 30, 2018 Anchorage earthquake overlain on a) Nath et al. 2002 1 hz HVSR amplification and b) Martirosyan et al. 2002 1 hz HVSR and SSR amplification average.

ii) Site specific study of NP 8040

We use station NP 8040 to perform a site-specific analysis within a high risk area in Anchorage. The time series and HVSR of site NP 8040 are shown in figure 7. This station is located on the edge of the Knik arm and may be underlain by Bootlegger Cove formation or weaker coastal estuarian deposits. It is 2.5 km north up the coast from the site of the Turnagain Heights liquefaction event that occurred during the 1964 Good Friday Earthquake and just west of the Merrill Field Airport.

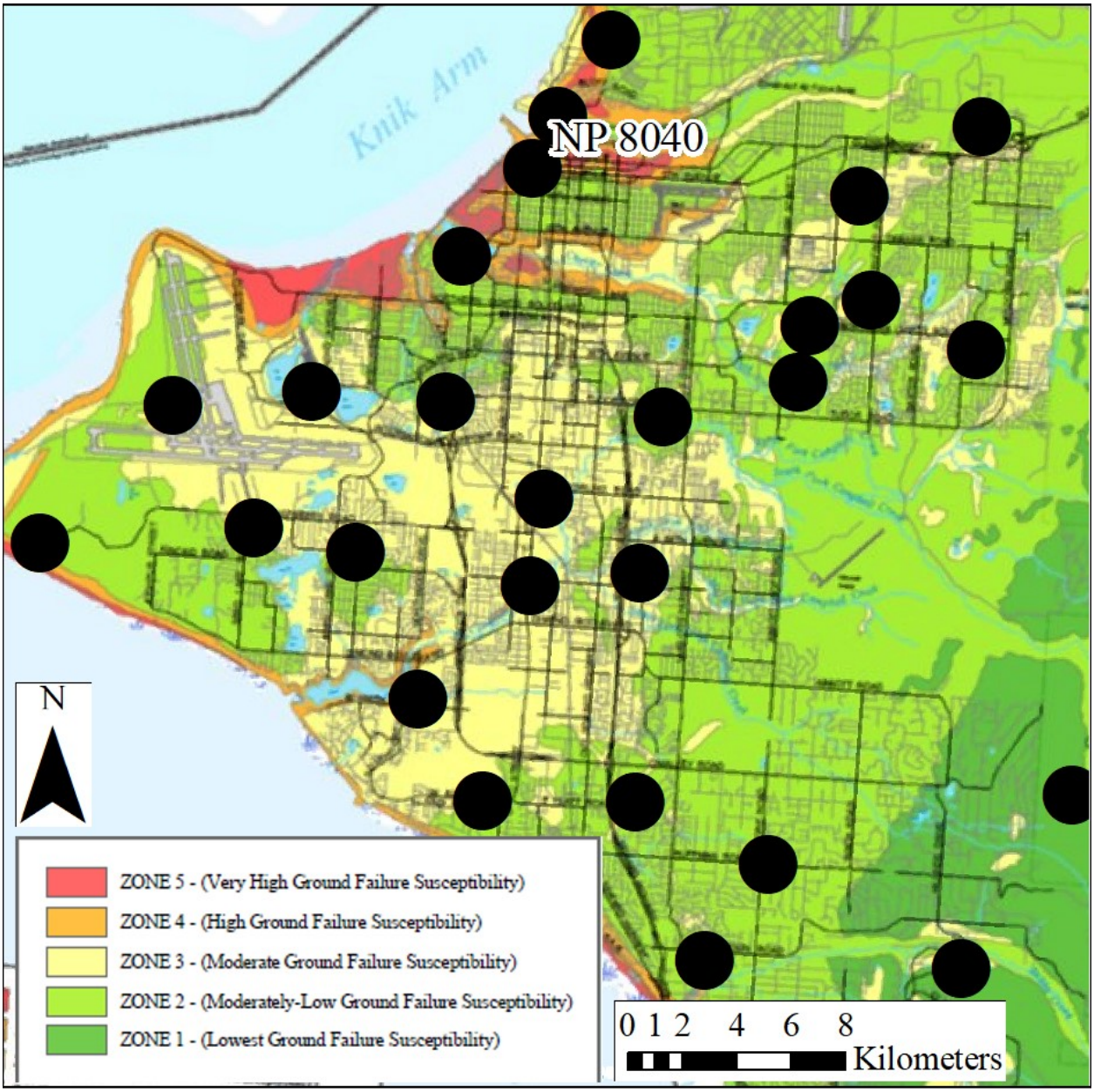


Figure 11: Station map overlying the Municipality of Anchorage seismic hazard map. Station NP 8040 is located in an area of "very high ground failure susceptibility".

Using an overburden shear wave velocity of 250 m/s from figure 5, we generated a shear wave velocity profile that also includes a 10-meter-thick weathered layer of 600 m/s (figure 12b). Using a bedrock shear wave velocity of 1000 m/s and damping of 2.5% at both overburden layers, we varied the depth of the slowest velocity layer until we got a match of the fundamental resonance of the TTF to that of the HVSR. We found that a depth of 50 meters provides the best fit. The results show that the SH1D transfer function is a good model of the HVSR at NP 8040 for the fundamental peak and the first two harmonics of the Anchorage event waveform (figure 12a). Using the "goodness of fit to the SH1D transfer function" measurement from the Thompson et al. 2012 classification system, we get a correlation of 0.58, close the threshold of a "good fit". This is also using HVSR as a proxy for the borehole empirical transfer function, likely a borehole transfer

function at this site would have a higher correlation.

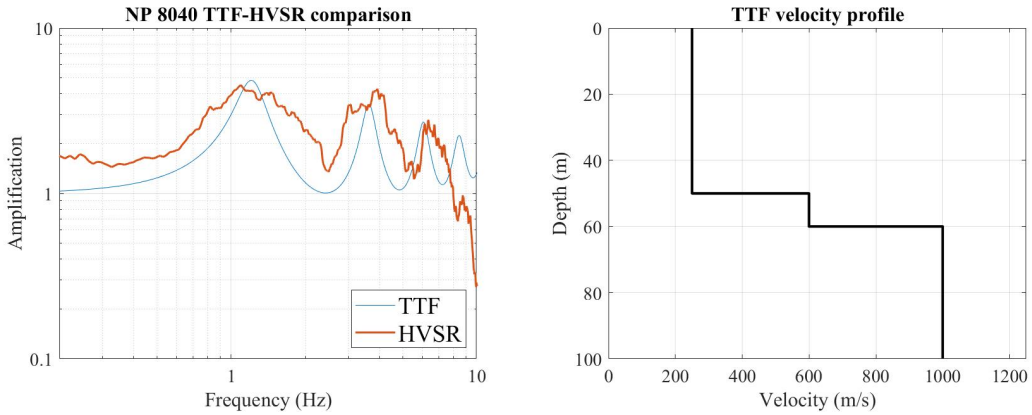


Figure 12: a) Comparison of a TTF of site NP 8040 with the HVSR at that site for the Anchorage event. b) The shear wave velocity profile used for calculating the TTF in figure a.

Using the design ground spectra computed using the steps in the results section equations 3-23, we compared the design response spectra to the response spectra of the Anchorage event waveform at NP 8040. Most of the energy is below the design response spectra, except between periods of 0.25 and 0.27 seconds. This low period amplification is likely due to the proximity of the epicenter to the basin; the waveform contains a lot of high frequency energy because less of it has attenuated away. Interestingly, the peak of the EW response spectra is at a frequency of 1.2 Hz, close to the fundamental resonance of the site. We think that this peak is likely caused by a site effect.

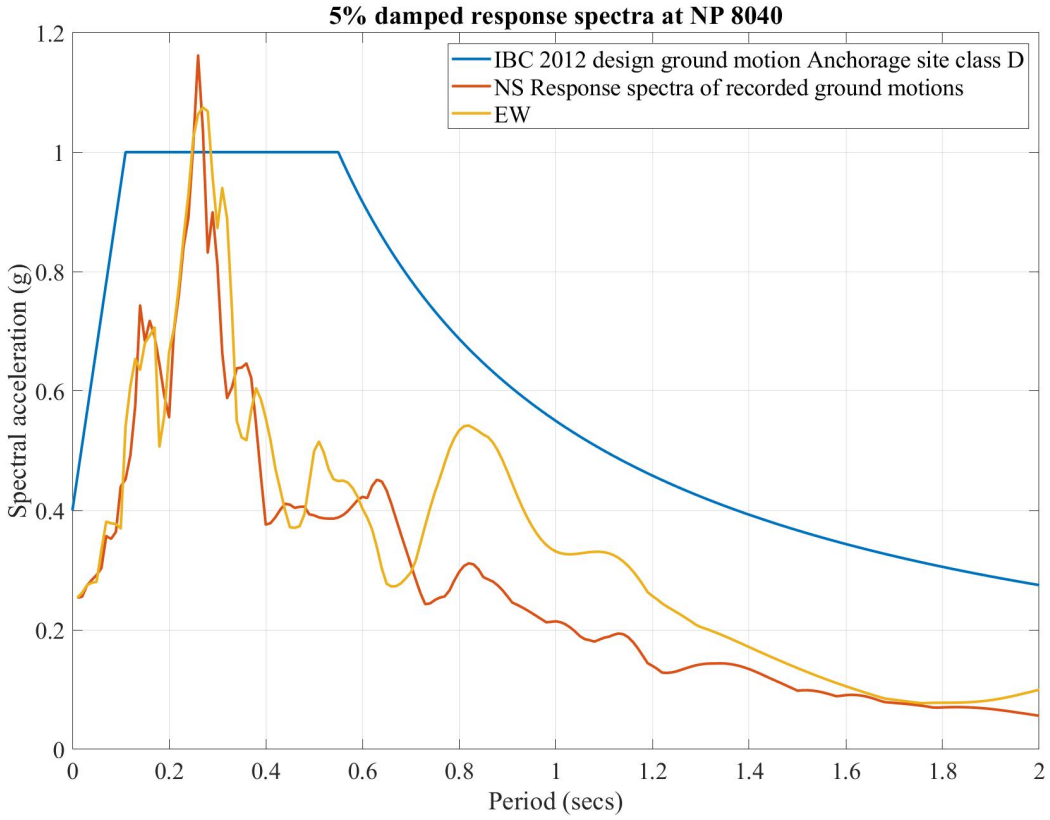


Figure 13: Response spectra of recorded ground motion of Anchorage event at site NP 8040 and the design ground motion from the IBC 2012 of an Anchorage site class D site.

Discussion

The 2018 Anchorage earthquake was a large event with an epicenter close to downtown Anchorage. We observed PGAs greater that 0.1 g which is in the region of non-linear site effects and some PGAs up to 0.8 g. Using the data from this event we performed a similar analysis to those performed by Nath et al. 2002 and Martirosyan et al. 2002 using HVSR to compute fundamental site resonance, amplification and 1 Hz amplification. Using just a single event compared to the many events used by the other two groups, we still saw similar trends in 1 Hz amplification and fundamental frequency. The 1 Hz amplification trend shows large amplifications in the Bootlegger Cove formation band increasing in amplification towards the north and the Knik arm. The trend in our study is more like that of Nath et al. 2002 than of Martirosyan et al. 2002 but the amplifications in our study are more like the latter group than the former.

We chose site NP 8040 to perform a site-specific analysis performing two calculations: 1) a theoretical SH1D transfer function and 2) response spectra for both horizontal components of the Anchorage event. For the SH1D transfer function calculation, we found that the HVSR ETF approximation is well modeled by a TTF with an overburden layer of 50 meters with shear wave velocity 250 m/s, provided in Martirosyan et al. 2002 and figure 5, a 10 meter weathered layer of shear wave velocity 600 m/s and a basement layer with shear wave velocity 1000 m/s, with a correlation 0.58. The first three harmonics are particularly modelled well the SH1D formulation. Using the design ground response spectra from equations 3-23 and classifying NP 8040 a site class D (figure 5), we computed the response spectra on the horizontal components of the Anchorage event, and found that for most periods the response spectra falls below the design response spectra except at

around 0.25 seconds. We attribute this high frequency amplification to the high frequency contained in the ground motion from the close epicenter. We note that the 0.8 second peak in the EW component spectra is likely caused by site effects. Our design and event response spectra are similar to those done by an AIR study after the event (Kianirad, 2018; figure 14).

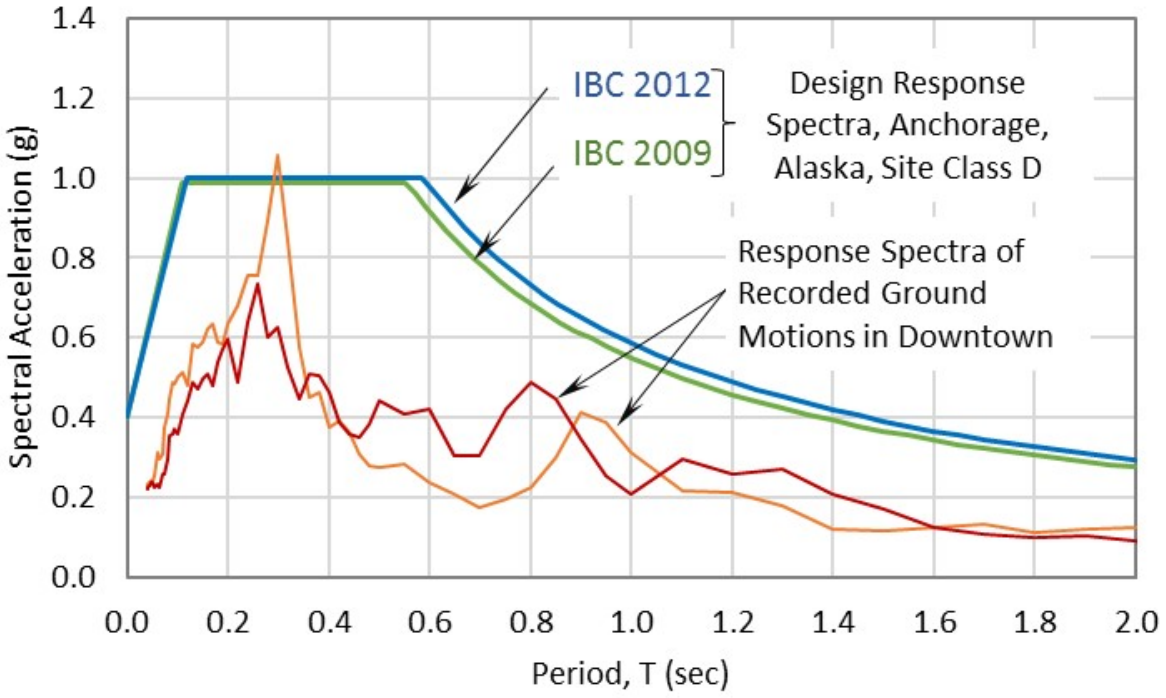


Figure 14: Response spectra computed by AIR after the Anchorage event.

Conclusion

Anchorage, Alaska is in an area of high seismicity caused by the subducting Pacific plate beneath that of the North American plate. The metropolitan area is built on a sedimentary basin with geologic units formed by the interaction of glacial and marine environments. Much of the surficial geology is governed by this environment and thus consists of interbedded glacial and glaciofluvial sands interbedded with marine clays and silts. The most notable surficial geologic unit is the Bootlegger Cove formation which lies on the west end of the City. It is susceptible to ground failure and this susceptibility increases towards the northwest and the Knik arm where the famous Turnagain Heights liquefaction event during the 1964 Good Friday Earthquake occurred. Using data from the 2018 Anchorage earthquake, we computed HVSRs and found that their 1 Hz frequency amplification is greatest in the Bootlegger Cove formation and increases towards the Knik arm. These results are like those computed by Nath et al. 2002. We also found a decrease in HVSR fundamental frequency from the base of the Chugach mountains to the Bootlegger Cove formation as the depth to basement increases to the west. These results were like those in Nath et al. 2002 and Martirosyan et al. 2002. We performed a site-specific analysis on NP 8040, a site located on the coast of the Knik arm in a region deemed by the Municipality of Anchorage to be "very susceptible to ground failure". We took known Vs_{30} measurements and added a weathered layer to the profile and varied the depth until we got a match with the HVSR fundamental peak. The HVSR and resulting TTF have a good fit between the fundamental resonance and the first two harmonics with a correlation of 0.58. Finally, we computed the response spectra at this site and found that it mostly stays below the design spectra from an Anchorage site class D site

except at a period of 0.25 seconds which we think is due to the large amount of high frequency energy in the waveforms from the close proximity of the epicenter. We also note a peak at 0.8 seconds on the EW response spectra that we attribute to site effects. These results are similar to those obtained in Kianirad, 2018. An ever-increasing precision of hazard mapping is necessary in Anchorage to prevent catastrophe from a future earthquake.

REFERENCES

Boore, D. M. (2005). SMSIM-Fortran Programs for Simulating Ground Motion from Earthquakes: Version 2.3-A of OFR 96-80-A. United States Department of the Interior. U. S Geological Survey.

Borcherdt, R.D., (1970). Effects of Local Geology on Ground Motion Near San Francisco Bay. Bulletin of the Seismological Society of America. Vol. 60, No. 1, pp. 29-61.

Haskell, N. A. (1960). Crustal Reflection of Plane SH Waves. Geophysical Research 65(12): 4147-4150.

Haskell N. A. (1953) The dispersion of surface waves on multilayered media. Bulletin of the Seismological Society of America; 72:17–34.

IBC International Building Code (2012). Intenational Code Council. Accessed 7/24/2020.
https : *//www.chino.k*12*.ca.us/cms/lib/CA*01902308*/Centricity/Domain/*5585*/icc.ibc*2012*.pdf*

Kianirad, E. (2018) Why did an M7.0 Earthquake close to Anchorage cause little damage? AIR. Accessed 7/24/2020. https://www.air-worldwide.com/blog/posts/2018/12/why-did-an-m7-0-earthquake-close-to-anchorage-cause-little-damage/.

Kramer, S.L. (1996). Geotechnical Earthquake Engineering, Prentice Hall, Upper Saddle River, N.J.

Martirosyan, A., Dutta, U., Biswas, N., Papageorgious, A., Combellick, R (2002) Determination of Site Response in Anchorage, Alaska, on the Basis of Spectral Ratio Methods. Earthquake Spectra, Vol. 18, No. 1, pp 85-104.

Municipality of Anchorage Seismic Hazard map. Building safety department. Accessed 7/24/2020.
http://wwwtst.muni.org/Departments/OCPD/development/BSD/Pages/DesignDataforAnchorage.aspx.

Nakamura, Y. (1989) A Method for Dynamic Characteristics Estimation of Subsurface using Microtremor on the Ground Surface. Railway Technical Research Institute 30(1): 25-33.

Nath, A. K., Biswas, N.N., Dravinski, M., Papageorgiou, A.S. (2002) Determination of S-wave site response in Anchorage, Alaska in the 1-9 Hz frequency band. Pure appl. geophys. Vol 159. pp 2673-2698.

Pontrelli, M.A., Baise, L.G. (2019) Assessing Site Response Complexity Using Single Station HVSR: Mexico City Case Study. Master's thesis. Tufts University.

Schmoll, H.R., Yehle, L.A., Updike, R.G. (1999) Summary of Quaternary geology of the Municipality of Anchorage, Alaska. Quaternary International. Vol 60. pp. 3-36.

Steidl JH, Tumarkin AG, Archuleta RJ. (1996) What is a reference site? Bulletin of the Seismological Society of America 1996;86(6):1733–48.

Thompson EM, Baise LG, Tanaka K, Kayen RE. (2012) A taxonomy of site response complexity. Soil Dynamics and Earthquake Engineering. 41(2012): 32-43.

Thomson WT. (1950). Transmission of elastic waves through a stratified solid. Journal of Applied Physics: 21:89–93.

Yehle, L.A., Schmoll, H.R. (1987) Surficial geologic map of the Anchorage B-7 NW quadrangle, Alaska. U.S. Geological Survey. Open-file report B7-168.

CODE LISTING

Code for processing DMC data)

```
% Qualifying exam
% Anchorage study

% Author: Marshall Pontrelli
% Date: 7/21/2019

close all
clear all


codepath = 'C:\Users\mpontr01\Desktop\HVSR\Codes';
cd(codepath)
data = zeros(1,5);
%% start
close all
clear all


% Before running this file the first time, execute the following statement
javaaddpath IRIS-WS-2.0.18.jar

network = 'NP';
statname = '8040';
starttime = '2018-11-30 17:29:30';
endtime = '2018-11-30 17:31:05';

[sampletimes,trace1,waveformfilt] = getDMCData(starttime,endtime ,statname,network,'HNZ');
fs = trace1.sampleRate;
V = trace1.data;
sensV = trace1.sensitivity;
sensVunits = trace1.sensitivityUnits;
disp(sensVunits)
%V = V(1:length(V)-1);
%%
% INPUTS
windowlen = 40;
numwin = 20;
windis = 25;
TTF = 'no';
outpath = 'no';
sav = 'no';
lowbound = 0.2;
upbound = 10;
Allplots = 'yes';
Timeplot = 'no';
IUMagplot = 'no';
AUMagplot = 'no';
IFMagplot = 'no';
AFMagplot = 'no';
HVSRplot = 'no';
Filterplot = 'no';
LowCorner = 0.1;
HighCorner = fs/2 - 1;
Npoles = 4;
width = 0.5;
%turn windows into samples for windowing calculations
```

```
[¬,trace1,waveformfilt] = getDMCData(starttime,endtime ,statname,network,'HNE');
EW = trace1.data;
sensEW = trace1.sensitivity;
sensEWunits = trace1.sensitivityUnits;
%EW = EW(1:length(EW)-1);

[¬,trace1,waveformfilt] = getDMCData(starttime,endtime ,statname,network,'HNN');
NS = trace1.data;
sensNS = trace1.sensitivity;
sensNSunits = trace1.sensitivityUnits;
%NS = NS(1:length(NS)-1);


[V] = Butter2(V, fs, 'LowCorner', LowCorner, 'HighCorner', HighCorner, 'Npoles', Npoles , ...
    'Filterplot', Filterplot);
Filterplot = 'no'; % toggle off filter plot so it doesn't plot response three times
[NS] = Butter2(NS, fs, 'LowCorner', LowCorner, 'HighCorner', HighCorner, 'Npoles', Npoles ...
    , 'Filterplot', Filterplot);
[EW] = Butter2(EW, fs, 'LowCorner', LowCorner, 'HighCorner', HighCorner, 'Npoles', Npoles ...
    , 'Filterplot', Filterplot);

%% Use sensitivity to get to acceleration
V = V / (9.81*sensV);
EW = EW / (9.81*sensEW);
NS = NS / (9.81*sensNS);
%% Create a time series plot (Output 1)]
if strcmp(Allplots, 'yes') == 1 || strcmp(Timeplot, 'yes') == 1
    timeseriesplot(NS,EW,V, fs)
end

%% Now compute PGA
PGA_V = max(abs(V));
PGA_EW = max(abs(EW));
PGA_NS = max(abs(NS));
disp(strcat('PGA V = ', {' '}, num2str(PGA_V)))
disp(strcat('PGA EW = ', {' '}, num2str(PGA_EW)))
disp(strcat('PGA NS = ', {' '}, num2str(PGA_NS)))



data(1,1) = PGA_V;
data(1,2) = PGA_EW;
data(1,3) = PGA_NS;
%% window the data
win = hann(length(V))';
V_win = V'.*win;
EW_win = EW'.*win;
NS_win = NS'.*win;

%% Compute unfiltered magnitude responses
V_mag = 4*abs(fft(V_win))/length(V);
EW_mag = 4*abs(fft(EW_win))/length(V);
NS_mag = 4*abs(fft(NS_win))/length(V);


%Computing the frequency -axis
N = length(V);
fax_binsN = (0 : N-1); %samples in NS component
fax_HzN1 = fax_binsN*fs/N; %frequency axis NS (Hz)
N_2 = ceil(N/2); %half magnitude spectrum
fax_HzN = fax_HzN1(1 : N_2);
V_mag2 = V_mag(1 : N_2);
EW_mag2 = EW_mag(1 : N_2);
NS_mag2 = NS_mag(1 : N_2);
```

```
%% create upbound and lowbound in terms of sample number
[¬, lowbound] = min(abs(fax_HzN - lowbound));
[¬, upbound] = (min(abs(fax_HzN - upbound)));



%% compute smoothed magnitude responses
window = ceil((length(V)/fs)*width); %width for smoothing filter in samples where 20 is ...
    the number of Hz on your x-axis

V_mag3 = smooth(V_mag2,window);
EW_mag3 = smooth(EW_mag2,window);
NS_mag3 = smooth(NS_mag2,window);

%% Compute geometric mean horizontals
H_mag = sqrt(EW_mag3.*NS_mag3);

%% Compute the HVSR
HV = H_mag./V_mag3;

%% Compute max and frequency
HVstat = HV(lowbound:upbound);
freqs = fax_HzN(lowbound:upbound);
[amp,I] = max(HVstat);
freq = freqs(I);
disp(strcat('Amplification = ', {' '}, num2str(amp)))
disp(strcat('f0 = ', {' '}, num2str(freq)))
data(1,4) = amp;
data(1,5) = freq;

%% compute 1 and 5 hz amplification
[¬, I] = min(abs(fax_HzN - 1)); % 1 Hz
hamp1 = HV(I);
disp(strcat('1 Hz amp = ', {' '}, num2str(hamp1)))
[¬, I] = min(abs(fax_HzN - 5)); % 1 Hz
hamp5 = HV(I);
disp(strcat('5 Hz amp = ', {' '}, num2str(hamp5)))

%% Now plot
figure

ETF = plot(fax_HzN, HV , 'Linewidth', 1.5);

xlabel('Frequency (Hz)','FontSize', 18)
ylabel('Amplification','FontSize', 18)
set(gca,'YScale', 'log', 'XScale', 'log','FontName', 'Times New Roman', 'FontSize', 14)
xlim([fax_HzN(lowbound) 10])
ylim([0.1 10])
xticks([.1 1 10])
xticklabels({'0.1', '1', '10'})
yticks([0.1 1 10 100])
yticklabels({'0.1', '1','10', '100'})
title(strcat(network, {' '}, statname))
grid on
box on
%% now save
name = strcat(network, {' '}, statname);
name = name{1};
filepath = 'C:\Users\mpontr01\Box\2020_2_summer\Qualifying exam\Earthquake Engineering\';
name = strcat(filepath,name);
saveas(ETF, name, 'jpg');


%% now calculate response spectra
zeta = 5; % 5% damping
```

```
[y, displacement, velocity, period, max1, max1V, max1D, per_maxA, per_maxV, per_maxD,...
    resp_02_secs, resp_02_secsV, resp_02_secsD ,resp_05_secs ,resp_05_secsV,resp_05_secsD, ...
        resp_1_secs, ...
    resp_1_secsV,resp_1_secsD, resp_2_secs,resp_2_secsV, resp_2_secsD, ...
        resp_5_secs,resp_5_secsV,resp_5_secsD] = Response_Spectra(NS, fs, zeta);

zeta = 5; % 5% damping
[yEW, displacement, velocity, period, max1, max1V, max1D, per_maxA, per_maxV, per_maxD,...
    resp_02_secs, resp_02_secsV, resp_02_secsD ,resp_05_secs ,resp_05_secsV,resp_05_secsD, ...
        resp_1_secs, ...
    resp_1_secsV,resp_1_secsD, resp_2_secs,resp_2_secsV, resp_2_secsD, ...
        resp_5_secs,resp_5_secsV,resp_5_secsD] = Response_Spectra(EW, fs, zeta);

zeta = 5; % 5% damping
[yV, displacement, velocity, period, max1, max1V, max1D, per_maxA, per_maxV, per_maxD,...
    resp_02_secs, resp_02_secsV, resp_02_secsD ,resp_05_secs ,resp_05_secsV,resp_05_secsD, ...
        resp_1_secs, ...
    resp_1_secsV,resp_1_secsD, resp_2_secs,resp_2_secsV, resp_2_secsD, ...
        resp_5_secs,resp_5_secsV,resp_5_secsD] = Response_Spectra(V, fs, zeta);

%% now do design response spectra
T0 = 0.11;
Ts = 0.55;
Tl = 16;
Sds = 1;
Sd1 = 0.55;
% loop over period
for ii = 1:1000
    Tn = (ii-1)*0.01;
    if Tn < T0
        Sd(ii) = Sds*(0.4+0.6*Tn / T0);
    end
    if Tn ≥ T0 && Tn ≤ Ts
        Sd(ii) = Sds;
    end
    if Tn > Ts && Tn ≤ Tl
        Sd(ii) = Sd1/Tn;
    end
    if Tn > Tl
        Sd(ii) = Sd1*Tl/Tn^2;
    end
    period(ii) = Tn;
end
figure
des = plot(period,Sd, 'linewidth',2);
xlim([0 2])
hold on


%% now plot

event = plot(period, y, 'linewidth',2);
EW = plot(period, yEW, 'linewidth',2);

xlim([0 2])
grid on
box on
xlabel('Period (secs)')
ylabel('Spectral acceleration (g)')
title('5% damped response spectra at NP 8040')
set(gca,'FontName', 'Times New Roman', 'FontSize', 18)
legend([des,event, EW], 'IBC 2012 design ground motion Anchorage site class D', 'NS ...
    Response spectra of recorded ground motions', 'EW','FontName', 'Times New Roman', ...
    'FontSize', 18)
set(gcf, 'Units', 'Normalized', 'OuterPosition', [0, 0.04, 1, 0.96]);
```

```
    name = strcat(filepath,'spectra.jpg');
    saveas(event, name, 'jpg');
```

Part 2: Environmental Statistics

Marshall Pontrelli

July 24, 2020

1) Coronavirus

i) Distributions

Given an underlying rate of infection p the number of positives in k tests will follow a binomial distribution, the PMF of which is:

$$P(Y = y) = \binom{k}{y} p^y (1-p)^{n-k}, \qquad y = 0, 1, ..., k \tag{1}$$

Where Y is the binomial random variable:

$$Y = \sum_{i=1}^{k} X_i \tag{2}$$

and X_i is the Bernoulli random variable associated with the i_{th} Bernoulli trial

$$X_i = \begin{cases} 1 & \text{if } i\text{th experiment is a success} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

(Akritas, 2019).

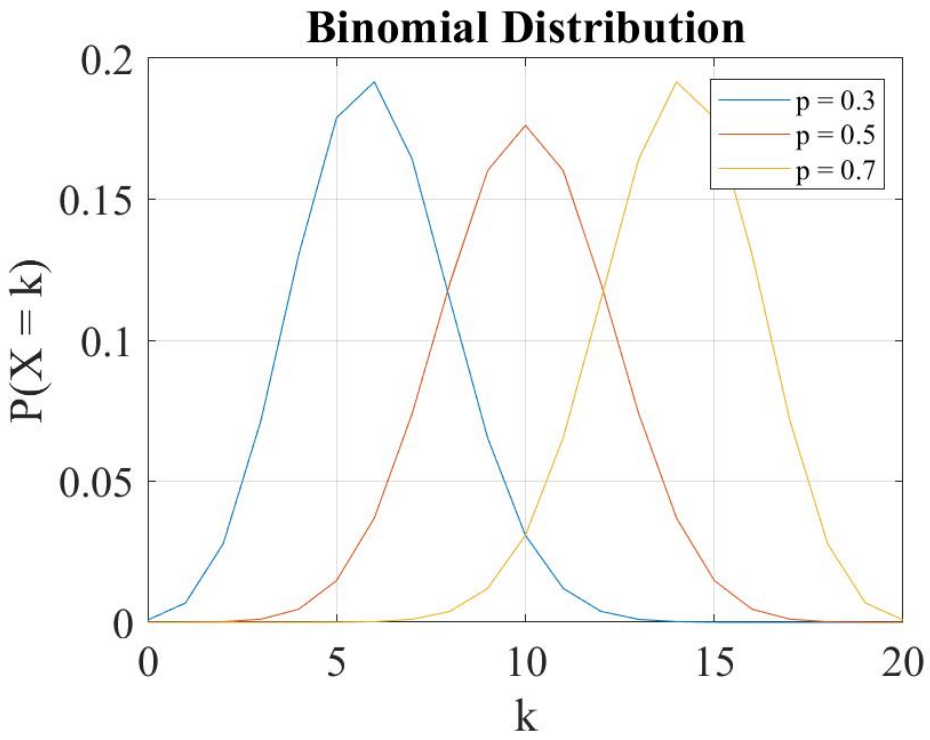


Figure 1: Examples of three binomial distributions with k = 20 and and p = 0.3, 0.5 and 0.7.

The probability of finding at least one positive if k people are tested is

$$1 - P(X = 0) \tag{4}$$

or the sum of probabilities of the pmf in figure 1 from y = 1 up to y = k.

ii) Dividing population into groups

The system of testing is as follows: divide the population N up into groups of size k and test each group by pooling their specimens and performing 1 test per group. After obtaining the first round of results, remove the groups who did not test positive, then test all of the members of each group who tested positve. The algorithm for this is to first take the underlying infection rate p and input it into the binomial distribution with group size k and solve for $1 - P(X = 0)$. This gives the probability that the group is infected.

$$P(Y = 0) = \binom{k}{0} p^y (1-p)^{n-k}, \qquad y = 0, 1, ..., k \tag{5}$$

We call this probability, the probability of a group being positive, $p2$. We now have the probability of each group being positive, and can put this back into the binomial distribution to solve for the the number of positive groups where the probability is $p2$ and the number of groups is N/k. Since the expected value of a binomial distribution is

$$E(X) = np \tag{6}$$

the expected number of tests to clear the population is

$$Expectedtests = N/k + N/k * p2 * k = N/k + p2 * N \tag{7}$$

where the first term is the total number of tests in the first round, and the second term is the expected number of groups that are infected times the number people per group. Again, this system is one round of testing groups of size k with one test per group, then a second round of all the individuals of all the groups who tested positive.

iii) President Monaco testing

If we test N = 11,000 students in a week in groups of k = 3, using equations 5 and 7, we get for an infection rate of 1%, and expected number of 3993 tests, for an infection rate of 2%, 4314 tests and for an infection rate of 5%, 4627 tests.

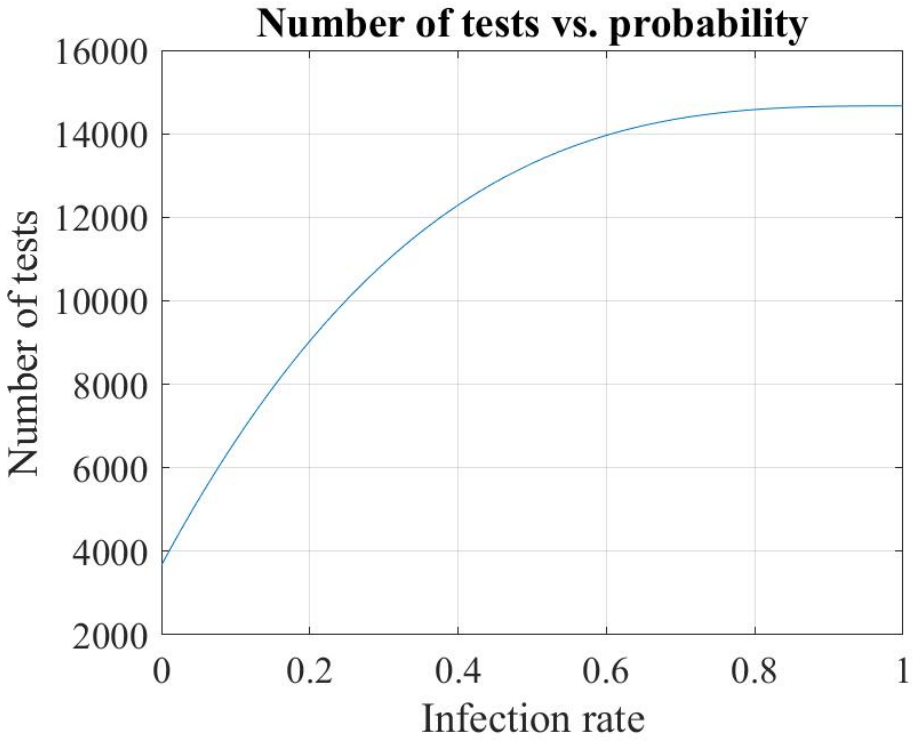


Figure 2: Plot of number of tests required vs. probability from equation 7

From figure 2 at an infection rate of 80%, the number of tests is constant.

iv) Group size

For an infection rate of 1%, we calculated an optimal group size of 11, for 2%, we calculated an optimal group size of 8 and for 5%, we calculated an optimal group size of 5.

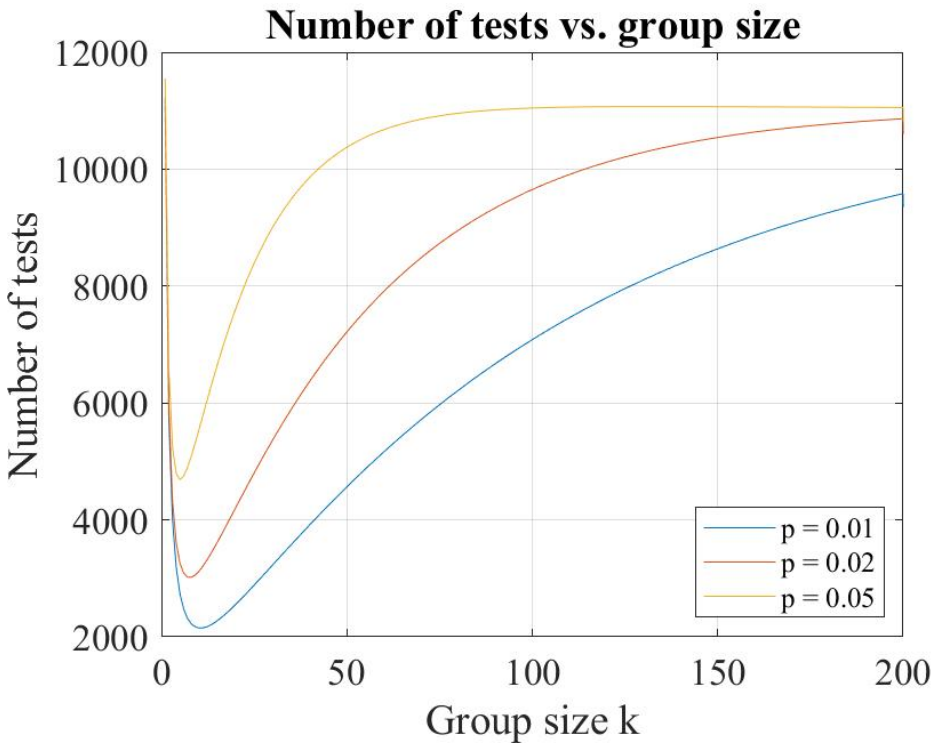


Figure 3: Plot of number of tests required vs. group size k with varying probability.

CODE LISTING

Code for binomial distribution plot)

```
% Qualifying exam
% Binomial distribution

% Author: Marshall Pontrelli
% Date: 7/20/2019

%% Begin
close all
clear all

%% Probability 0.3
n = 20;
p = 0.3;
Y = zeros(0,n+1);
for y = 1:n+1
    yy = y - 1;
    Y(y) = nchoosek(n,yy)*p^yy*(1-p)^(n-yy);
end
nx = 0:20;
p_03 = plot(nx,Y);

hold on
%% Probability 0.5
n = 20;
```

```
p = 0.5;
Y = zeros(0,n+1);
for y = 1:n+1
    yy = y - 1;
    Y(y) = nchoosek(n,yy)*p^yy*(1-p)^(n-yy);
end
nx = 0:20;
p_05 = plot(nx,Y);

hold on
%% Probability 0.7
n = 20;
p = 0.7;
Y = zeros(0,n+1);
for y = 1:n+1
    yy = y - 1;
    Y(y) = nchoosek(n,yy)*p^yy*(1-p)^(n-yy);
end
nx = 0:20;
p_07 = plot(nx,Y);

hold on

%% figure info
grid on
box on
xlabel('k')
ylabel('P(X = k)')
title('Binomial Distribution')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

legend([p_03, p_05, p_07], 'p = 0.3', 'p = 0.5', 'p = 0.7', 'FontName', 'Times New Roman', ...
    'FontSize', 12)
saveas(p_07, 'ES_dist.jpg', 'jpg');
```

Code for Presdent Monaco testing)

```
% Qualifying exam
% President Monaco testing

% Author: Marshall Pontrelli
% Date: 7/20/2019

%% Begin
close all
clear all

%%
N = 11000;
k = 3;
p_vec = 0:0.01:1;
Expect = zeros(1,length(p_vec));
for i = 1:length(p_vec)
    p = p_vec(i);
    Y = zeros(0,k+1);
    for y = 1:k+1
        yy = y - 1;
        Y(y) = nchoosek(k,yy)*p^yy*(1-p)^(k-yy);
    end
    p2 = 1 - Y(1);
    Expect(i) = N/k + p2*N;
```

```
end
fig = plot(p_vec,Expect);
grid on
box on
xlabel('Infection rate')
ylabel('Number of tests')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
title('Number of tests vs. probability')
saveas(fig, 'prob_vs_inf.jpg', 'jpg');
```

Code for optimal group size)

```
% Qualifying exam
% President Monaco testing

% Author: Marshall Pontrelli
% Date: 7/20/2019

%% Begin
close all
clear all

%%
N = 11000;
p = 0.01;
Expect = zeros(1,N);
for k = 1:200

    Y = zeros(0,k+1);
    for y = 1:k+1
        yy = y - 1;
        Y(y) = nchoosek(k,yy)*p^yy*(1-p)^(k-yy);
    end
    p2 = 1 - Y(1);
    Expect(k) = N/k + p2*N;
end
c = 1:N;
p_01 = plot(c,Expect);

hold on
p = 0.02;
Expect = zeros(1,N);
for k = 1:200

    Y = zeros(0,k+1);
    for y = 1:k+1
        yy = y - 1;
        Y(y) = nchoosek(k,yy)*p^yy*(1-p)^(k-yy);
    end
    p2 = 1 - Y(1);
    Expect(k) = N/k + p2*N;
end
c = 1:N;
p_02 = plot(c,Expect);
xlim([0 200])

hold on
p = 0.05;
Expect = zeros(1,N);
for k = 1:200
```

```
        Y = zeros(0,k+1);
        for y = 1:k+1
            yy = y - 1;
            Y(y) = nchoosek(k,yy)*p^yy*(1-p)^(k-yy);
        end
        p2 = 1 - Y(1);
        Expect(k) = N/k + p2*N;
    end
    c = 1:N;
    p_05 = plot(c,Expect);
    xlim([0 200])
    grid on
    box on
    xlabel('Group size k')
    ylabel('Number of tests')
    set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
    title('Number of tests vs. group size')
    legend([p_01, p_02, p_05], 'p = 0.01', 'p = 0.02', 'p = 0.05', 'FontName', 'Times New ...
        Roman', 'FontSize', 12,'location', 'southeast')

    saveas(p_05, 'prob_vs_gr.jpg', 'jpg');
```

2) Monte Carlo

i) Define "empirical"

"Empirical" work is work identifying relationships using quantifiable data. For example, in the case of vibration on a string, a "theoretical" model would be derived from first principles and have an analytical solution, whereas an "empirical" study would involve measuring the vibration on the string and quantifying relationships between measurable quantities. An example of an "empirical" study could be plugging in an electric guitar, playing an A, recording it, doing an fft on the recording to decompose the frequencies and determining whether the guitar A string is tuned correctly to 440 Hz. The theoretical study would arrive at the same conclusion using the tension on the string and the length of the sting. In the geosciences, this would similarly be analogized by deriving the analytical solution for vertically propagating shear waves using the density, stiffness and damping of a soil column vs. the "empirical" method of collecting data using an instrument and decomposing the frequencies from the data.

There are two challenges associated with empirical work in the geosciences: 1) we oftentimes get qualitative or categorical data and 2) patterns we observe may not map easily to useful physical quantities and may instead be patterns that don't give any insight into reality. For the challenge of categorical data, an example from classic geology is defining rocks by their "lustre" which has qualitative categories like "dull", "metallic" and "pearly". Though this information is useful, it is difficult to turn into any type of regression or make claims about because there is no continuous distribution that can be attributed to the categories. For example, our group is working on a project classifying changes in geologic environment off the coast of Massachusetts. We're using geostatistics to quantify variability in the subsurface. Some of the most important studies we have are geological investigations, many with tables containing information like "geologic age" and "geologic unit". This information does clearly denote differences, but the challenge lies in assigning statistical distributions to the categorical data. For the challenge of finding patterns with no underlying physical meaning, an example is finding a correlation between rain in Australia and the outcome of a baseball game in the U.S. This is happenstance that pops out of the data and though is easy to identify when one is dealing with obvious independent events, in the geosciences, the concepts and measurements may be more abstract and subtle and could lead the researcher to finding trends that don't mean anything in the physical world. Correlation does not equal causation. In my work, I look for trends in the shapes of frequency spectra. In the theoretical world, these trends are easily derived, however, in the messy world of data, I have found myself prone to finding patterns that may have no physical underlying meaning.

ii) "Empirical" and Monte Carlo analyses

In a world full patterns, it is often advantageous to derive parameters and define exactly what they mean. In our work, this is in the form of an analytical model. An analytical model describes relationships between input parameters and generated outputs. Models are ideal for Monte Carlo analyses because we can define each parameter with a distribution and run many simulations to observe how the output varies. Though these simulations provide a broad and relatively simple model of reality, they allow us to observe variation between definite parameters in the absence of messy noise from external parameters we would have to deal with if we were using real world data. Because we have conveniently defined our parameters, the Monte Carlo analysis of our model is 1) only as robust as the parameters we have included and 2) robust only given how we have defined our input parameter distributions. In the first instance, we may have a model of reality that has 6 parameters, it can thus only account for those 6 parameters where in reality there may be many more parameters that affect the output. In the second instance, we define the distributions that we use in our simulations, thus if we define distributions that are not as they are in nature or define them with incorrect distribution parameters (like mean or variance), we generate simulated outputs that do not model reality.

iii) Compare Pearson's r and Stedinger correlation

Using Monte Carlo analysis, and the real space correlation coefficient of 0.7, we transformed to the log-space correlation using equation 6 in Barber et al (2019). We used this to input into the equations in the appendix of Barber et al. (2019) to generate bivariate log-normal distributions with a real space correlation of 0.7. We simulated 1000 distributions each with 100 samples for each log-normal distribution and compared there Pearson's r vs. Stedinger correlation coefficient for 3 coefficients of variation: 0.5, 2 and 10 (figure 4).

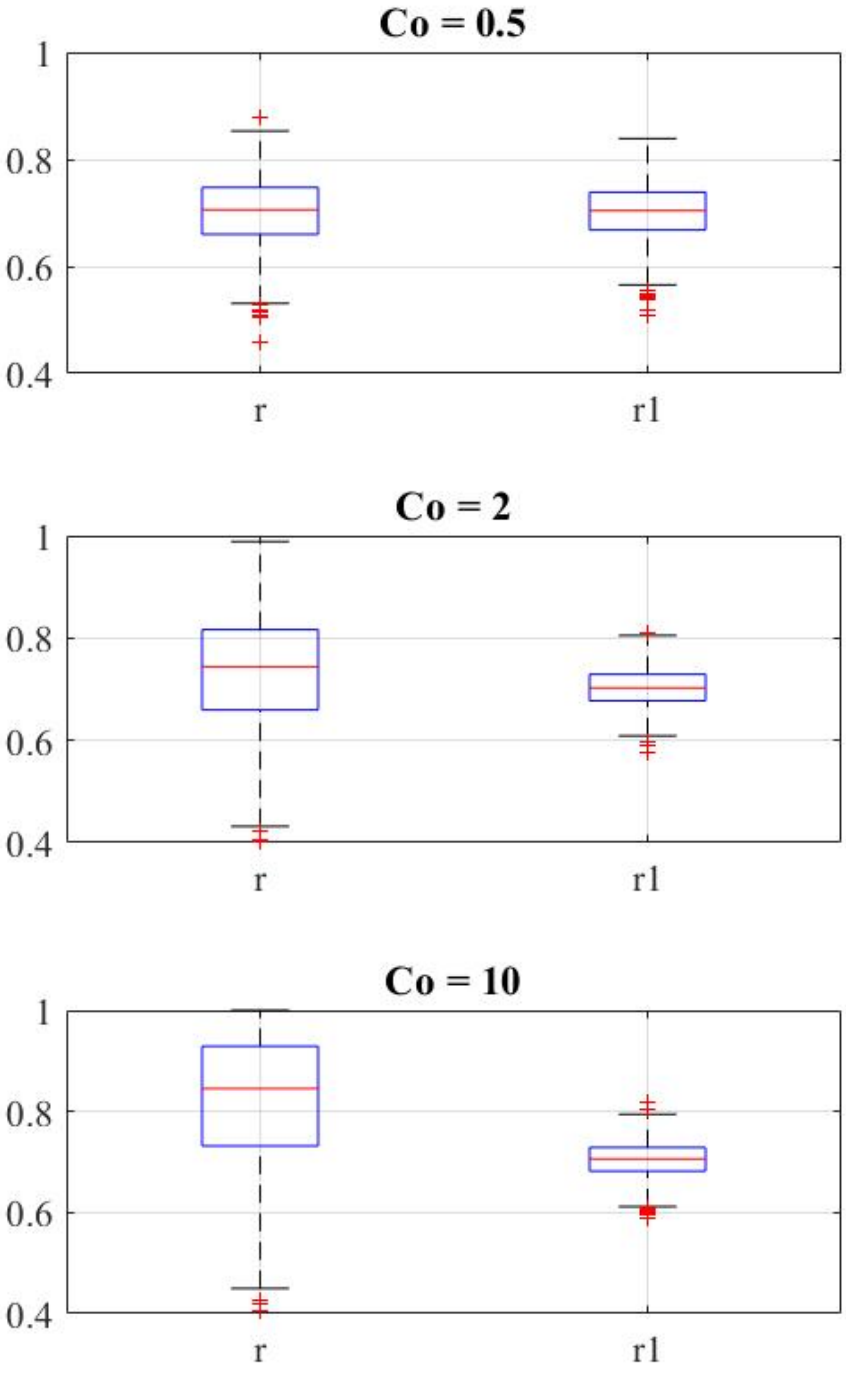


Figure 4: Results of Monte Carlo analysis with 1000 trials of 100 samples of two bivariate log normal distributions with three coefficients of variation and boxplots of the Pearson's r values and Stedinger correlation values.

These results indicate that for lognormal bivariate distributions, as the coefficient of variation increases, Pearson's r tends to overpredict the true correlation value. Thus, with more variation, the measurement estimate increasingly biases high. For my own research, given my use of Pearson's r to compare models to data, these results indicate that if both model and data are drawn from log-normal distributions, the Pearson's r estimator I use for the true correlation may be biased high if the data have significant variation.

CODE LISTING

Code for coefficient of variation = 0.5 plot)

```
% Qualifying exam
% Monte Carlo

% Author: Marshall Pontrelli
% Date: 7/20/2019

%% Begin
close all
clear all


%% From Barber et al. 2019, simulate O and S

t0 = 0;
u0 = 1;
s0 = 0.5;

% equations A1a
uu = log((u0 - t0)/sqrt(1+(s0/(u0-t0))^2));
su = sqrt(log(1+(s0/(u0 - t0))^2));

ts = 0;
us = 1;
ss = 0.5;

% Equations A1b
uv = log((us - ts)/sqrt(1+(ss/(us-ts))^2));
sv = sqrt(log(1+(ss/(us - ts))^2));

%% now solve for the log space correlation coefficient puv from eq 6 in Barber
% et al. 2019. (I'm a little short on time so I did this one by code.)

puv = 0:0.001:1;
for q = 1:length(puv)
    p(q) = (exp(puv(q)*su*sv) - 1)/(sqrt(exp(su^2)-1)*sqrt(exp(sv^2)-1));
end
[¬,I] = (min(abs(p-0.7)));
puv = puv(I);

%% Now simulate 1000 distributions of 100 samples per distribution and compute
% Pearson's r and r1

for ii = 1:1000
    % Equation A2, simulate O
    for i = 1:100
        p = rand(1);
        O(i) = t0 + exp(uu + norminv(p)*su);
    end

    % errors for equation A3
```

```
        s2e = sv^2*(1-puv^2);
        e = normrnd(0,sqrt(s2e), [1,100]);

        % Equation A3, simulate S
        for j = 1:length(O)
            S(j) = ts + exp(uv + puv * (sv/su)*(log(O(j) - t0) - uu) + e(j));
        end

        %% now compute Pearson's r
        log_O = log(O);
        log_S = log(S);
        rr = corrcoef(O, S);
        r(ii) = rr(2);

        %% Now compute r1
        v_bar = mean(log_S);
        svv2 = sum((log_S - v_bar).^2)/length(log_S);

        u_bar = mean(log_O);
        suu2 = sum((log_O - u_bar).^2)/length(log_O);

        ssuv = sum((log_O - u_bar).*(log_S - v_bar))/length(log_O);

        r1(ii) = (exp(ssuv)-1)/sqrt((exp(svv2) - 1)*(exp(suu2) - 1));


    end
    x = horzcat(r',r1');
    subplot(3,1,1)
    boxplot(x,'Labels',{'r','r1'})
    box on
    grid on
    set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
    title('Co = 0.5')
    ylim([0.4 1])
```

3) Baseball

i) Describe a method one could use to analyze these data

I will describe the method I use in section ii of this question, multiple regression. Multiple regression predicts a dependent variable, in this case 1987 salary, using a set of independent variables that can be either numerical or categorical by solving for slope coefficients and a y intercept of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... \tag{8}$$

where the number of slope coefficients is equal to the number of predictor variables. Each slope coefficient is assigned a significance value with a null hypothesis "the slope coefficient is equal to zero". A low p-value for the slope coefficient means that the predictor variable is significantly varying with the dependent variable. The model is also assigned an overall r^2 value which is interpreted as "the percentage of which the variation of the dependent variable is explained by the model". Given that the data meet some criteria, including that they are normally distributed, the researcher can develop a multiple linear regression model to predict a dependent variable and then input values into that model for each predictor variable thus obtaining a prediction of the dependent variable from those inputs. This prediction has a confidence interval with bounds dependent on the variability of the model. Models also undergo multiple collinearity from correlated independent variables. Correlated independent variables both attribute to variation in the dependent variable and causes unusual and difficult to interpret slope coefficients.

ii) Apply multiple linear regression to predict salary

We begin our multiple linear regression predicting 1987 salary using the predictor variables lifetime batting average, years in the MLB and 1986 batting average by doing a short exploratory data analysis (fig 5). First, we check for normality of the data.

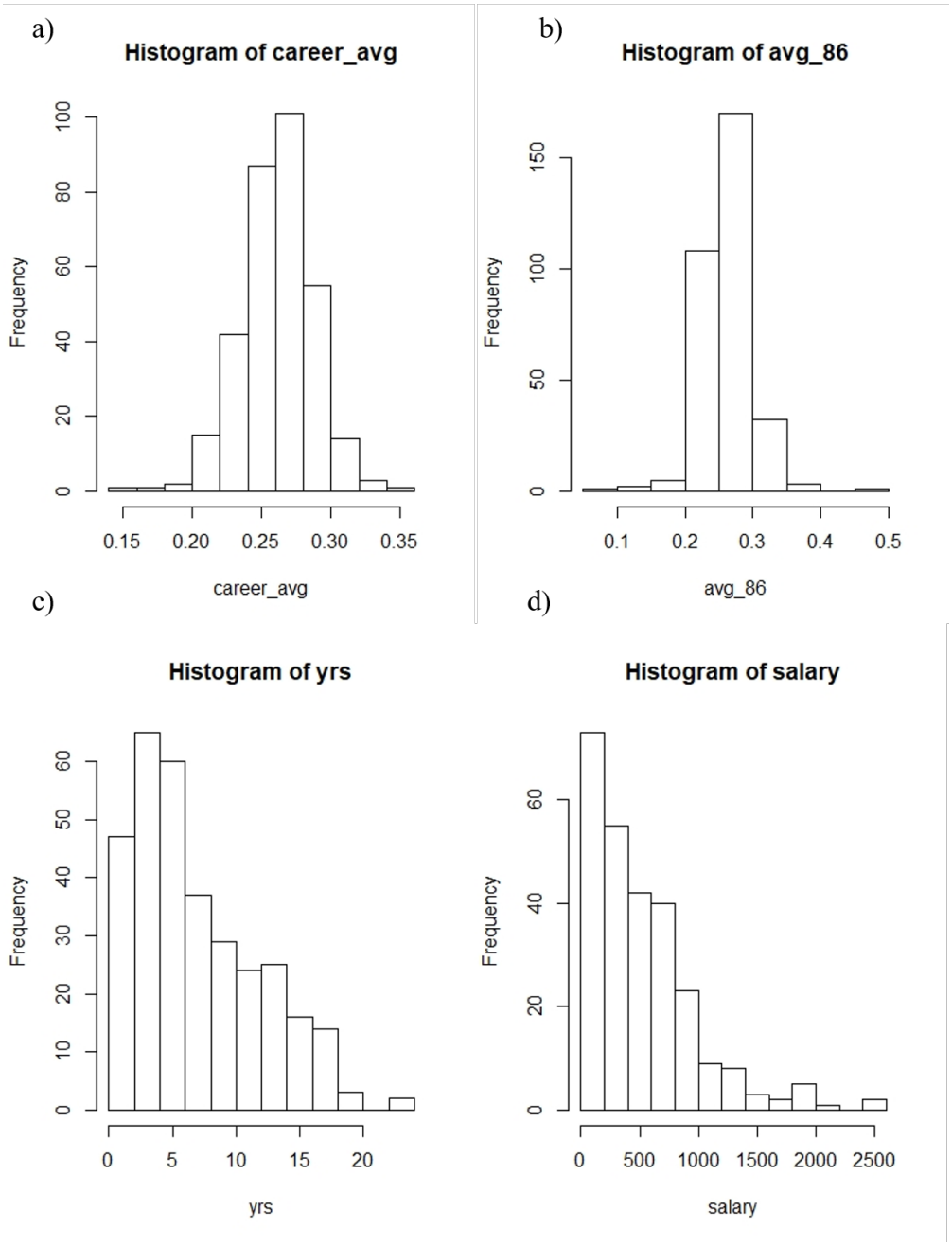


Figure 5: Histograms of the three independent variables in the multiple linear regression a) career average b)1986 average and c) career years in the MLB and the dependent variable d) salary.

Since years in the MLB and salary appear log normal, we log-normally transformed them to get the

histograms:

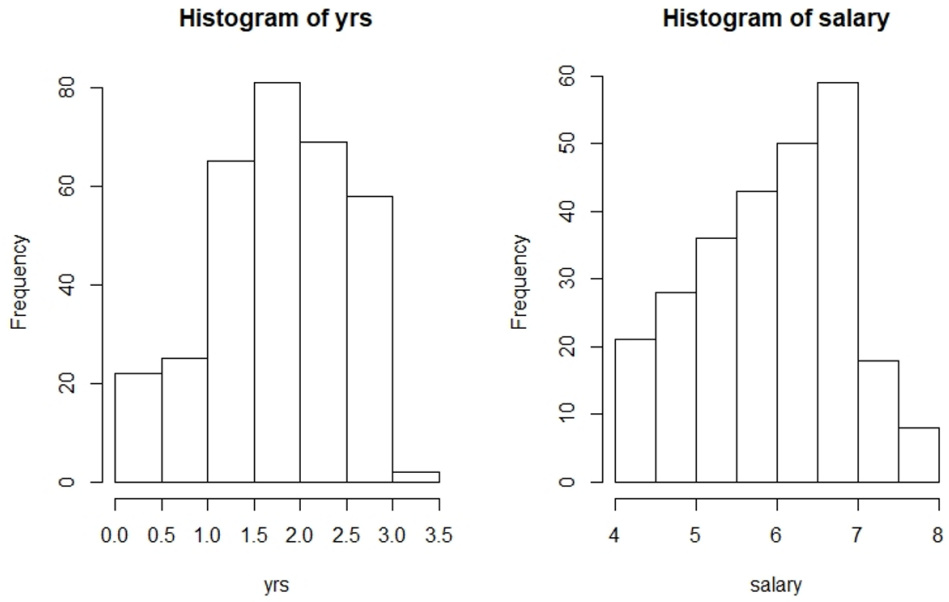


Figure 6: Replotted histograms after log-transform of a) career years in the MLB and b) salary.

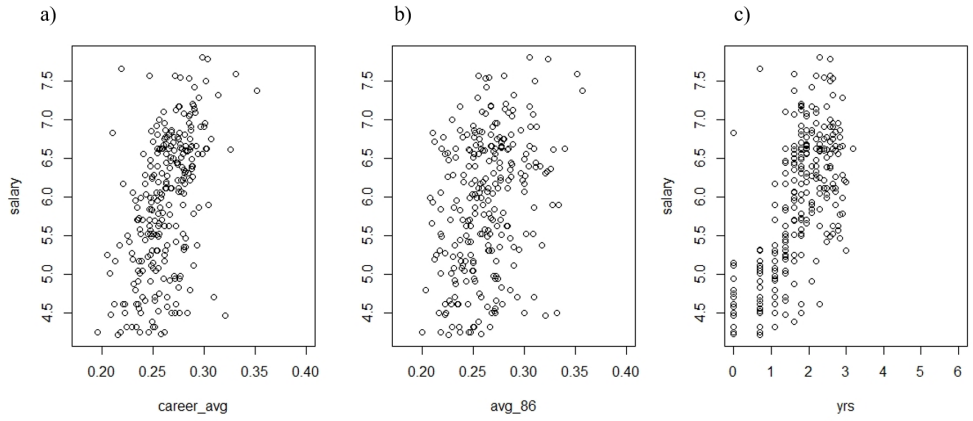


Figure 7: Scatterplots of the logarithm of salary vs the three independent variables in the multiple linear regression a) career average b)1986 average and c) log of career years in the MLB.

These plots have Pearson's correlation coefficients 0.51, 0.23 and 0.34 respectively. To look for multicollinearity, we computed Pearson's correlation coefficients between the independent variables getting values of 0.15 between 1986 average and years, 0.73 between 1986 average and career average and 0.34 between career average and years. We conclude that there is some multicollinearity between independent variables.

After exploratory data analysis consisting of two log-transforms, we performed multiple linear regression with the log of salary as the dependent variable and career average, 1986 average and log of years as the

independent variables yielding:

$$ln(salary) = 1.528 + 17.01 * career_{avg} + 0.62 * ln(years) - 4.41 * 1986_{avg} \tag{9}$$

which has a multiple r^2 value of 0.5411 and each slope coefficient and intercept with a confidence level greater than 99%. We then used this model to predict the expected salary of a player with a lifetime batting average of 0.300, with 10 years of experience and with a 0.300 batting average in 1986. From our model, this player would most likely be paid 740 thousand dollars which an upper 95% confidence interval of 977 thousand dollars and a lower 95% confidence interval of 850 thousand dollars.

ii) Do any players stand out?

To determine if any players stand out, we plotted the residuals of our model.

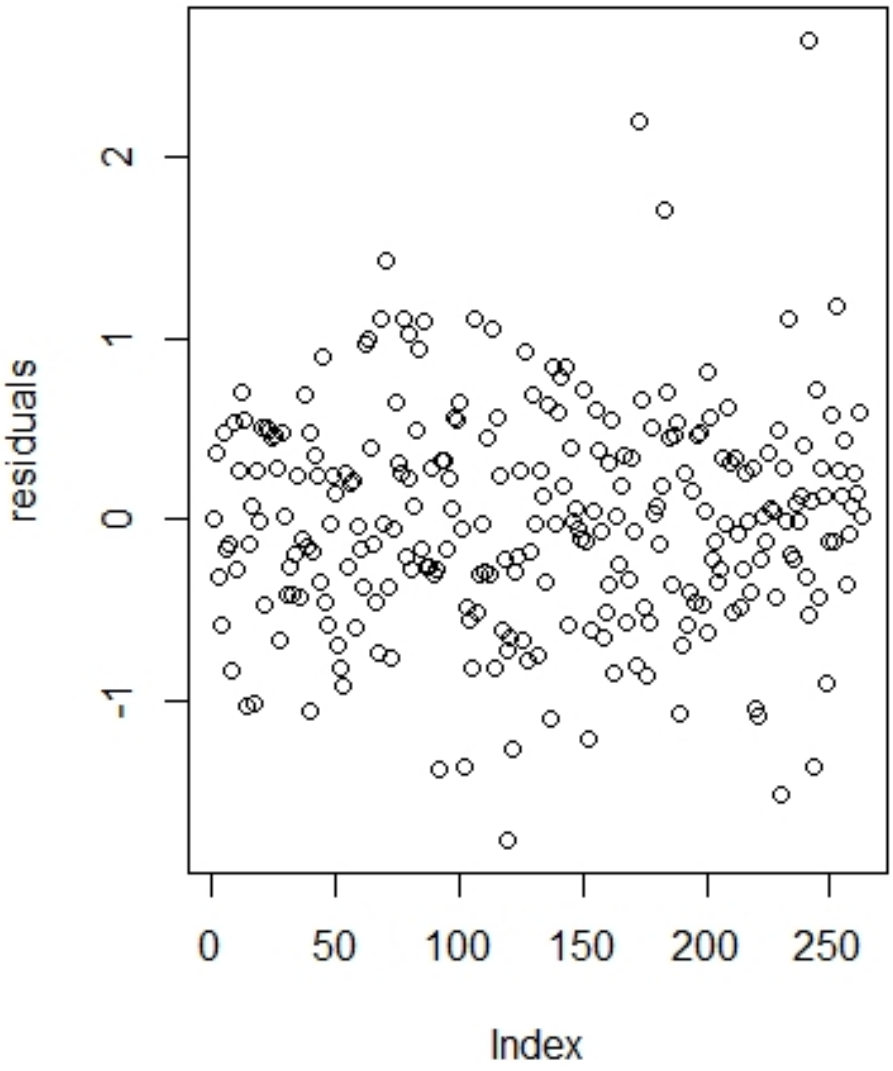


Figure 8: Plot of model residuals, the points far away from 0 are the players with the most abnormal salaries.

There are several outliers in the analysis which are the points that are far away from the 0 line. With an extremely large dataset, this method of plotting residuals is a good way to find standout data.

CODE LISTING

Code for multiple linear regression)

```
install.packages("car")
library(car)
install.packages("RColorBrewer")
library(RColorBrewer)

install.packages("corrplot")
library(corrplot)

MLB <- read.csv("C:\\Users\\mpontr01\\Box\\2020_2_summer\\Qualifying ...
    exam\\Document\\MLB_Salaries.csv",h=T)
head(MLB)
names(MLB)


# first calculate career batting average
hits <- MLB$number.of.hits.during.his.career
ABs <- MLB$number.of.times.at.bat.during.his.career
career_avg <- hits / ABs

# Now pull out how many years he played
yrs <- MLB$number.of.years.in.the.major.leagues


# Now pull out his 1986 batting average
AB_86 <- MLB$number.of.times.at.bat.in.1986
hits_86 <- MLB$number.of.hits.in.1986
avg_86 <- hits_86 / AB_86

# Now pull out salary
salary <- MLB$X1987.annual.salary.on.opening.day.in.thousands.of.dollars

# now plot each independently
# Plot career average
plot(career_avg,salary, xlim=c(0.180,.400))

# Plot number of years played in the majors
plot(yrs,salary, xlim=c(0,6))

# Plot 1986 batting average
plot(avg_86,salary, xlim=c(0.180,0.400))

# plot the pairs
pairs(¬ avg_86 + yrs + career_avg)

# compute corrcoeffs between independent and dependenr
cor(avg_86,salary, method = "pearson",use="complete.obs")
cor(career_avg, salary,method= "pearson",use="complete.obs")
cor(yrs, salary,  method = "pearson",na.omit = T,use="complete.obs")

# compute corrcoeffs for multi-collinearity
cor(avg_86,yrs, method = "pearson")
cor(avg_86,career_avg, method = "pearson")
cor(career_avg,yrs, method = "pearson")

# plot histograms
hist(avg_86)
hist(career_avg)
hist(yrs)
hist(salary)

# perform the log transform on years and salary
yrs <- log(yrs)
salary <- log(salary)
```

```
# now replot the histograms
hist(yrs)
hist(salary)



# build the model
model<-lm(salary ¬ career_avg + yrs + avg_86)
summary(model)

# Now make a prediction
predict(model,data.frame(career_avg = 0.300, avg_86 = 0.300, yrs = log(10)), interval = ...
    'confidence')
```

REFERENCES

Akritas, M. (2019) Probability and Statistics with R for Engineers and Scientists. Pearson Modern Classic.

Caitline Barber, Jonathan R. Lamontagne & Richard M. Vogel (2020) Improved estimators of correlation and R^2 for skewed hydrologic data, Hydrological Sciences Journal, 65:1, 87-101, DOI: 10.1080/02626667.2019.1686639

Part 3: Geophysical Data Processing

Marshall Pontrelli

July 24, 2020

1) Brune source spectrum

We begin with an examination of the Brune farfield ground displacement source time function given in equation 17 of Brune (1970):

$$u = f * (r/R)(\sigma/\mu)\beta t'' e^{-\alpha t''} \tag{1}$$

where, from Brune (1970) equation 18

$$t'' = t - R/\beta \tag{2}$$

To simplify the equation, we will ignore the phase shift imposed by the R/β term and will set the $f * (r/R)(\sigma/\mu)\beta$ term equal to 1, leaving the simplified Brune farfield ground displacement source time function:

$$u = te^{-\alpha t} \tag{3}$$

In his paper, Brune looks to model the source time function:

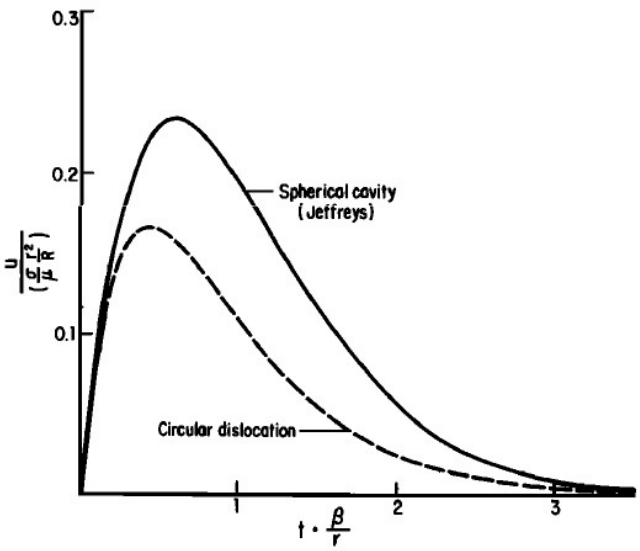


Figure 1: From Brune (1970) farfield pulse shapes. Jeffreys's model of the stress pulse on the inside of a sphere and the circular dislocation model won't be discussed here. Note that all the coefficients on the x and y axes are all equal to 1 in our simplification.

Plotting equation 3 and varying α, we get

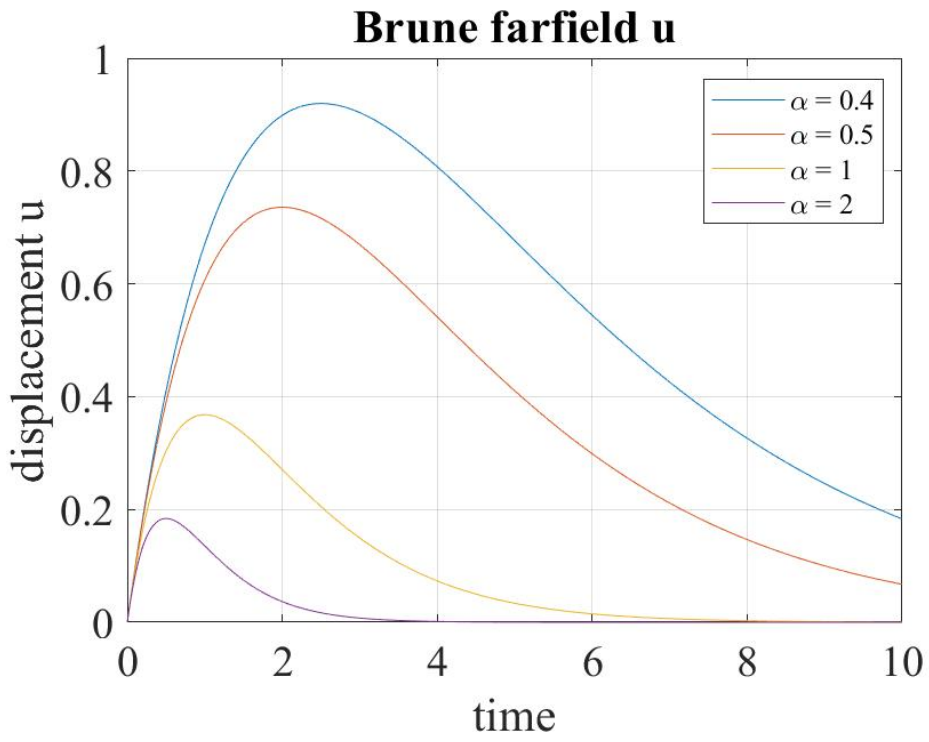


Figure 2: Plot of simplified Brune farfield ground displacement source time function from equation 3 with four values of α and all positive values of time.

a similar result to that obtained in Brune (1970). Interestingly and with relevance to the rest of this set of questions, plotting the farfield pulse shape with negative values of time yields

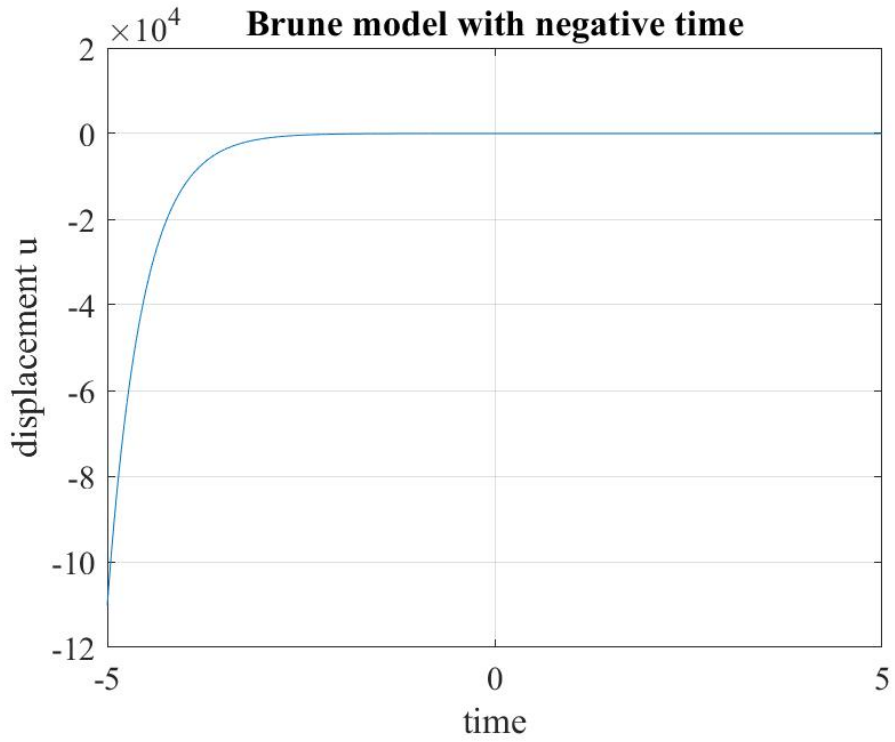


Figure 3: Plot of simplified Brune farfield ground displacement source time function from equation 3 with positive and negative values of time.

Inspecting figure 3, we see that the continuous approximation across all of time for which the Brune model allows yields unrealistic displacements as values of time become negative.

Let's now explore the Fourier transform of the Brune model and derive its source spectral shape from the time domain model in equation 3. The continuous time Fourier transform is given by

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{4}$$

where j is $\sqrt{-1}$ and ω is the circular frequency $2\pi f$ where f is the frequency in Hz. The Fourier transform takes a signal $x(t)$ in the time domain to one in the frequency domain. In our problem, we are looking to take the simplified source time function in equation 3 and convert it to the frequency domain using the Fourier transform to analyze is spectral characteristics. Let's plug equation 3 with the function u as $x(t)$ into the Fourier transform in equation 4

$$X(\omega) = \int_{-\infty}^{\infty} t e^{-\alpha t} e^{-j\omega t} dt \tag{5}$$

We recognize that this is of the form of the common Fourier transform pair:

$$u(t) t e^{-\alpha t} \iff \frac{1}{(\alpha + j\omega)^2} \tag{6}$$

where $\iff$ is performing the Fourier transform from the time domain to the frequency domain and the inverse Fourier transform from the frequency domain to the time domain. For our problem, $u(t) = 1$ thus the Fourier transform of equation 3 is

$$X(\omega) = \frac{1}{(\alpha + j\omega)^2} \tag{7}$$

We now decompose equation 7 into into magnitude ($|H(\omega)|$ and phase ($H(\Theta)$) spectra using their defintions:

$$|H(\omega)| = \sqrt{H_R^2(\omega) + H_I^2(\omega)} \tag{8}$$

$$H(\Theta) = tan^{-1} \frac{H_I(\omega)}{H_R(\omega)} \tag{9}$$

Where H_R and H_I are the real and imaginary parts of the of $X(\omega)$ respectively. We perform this calculation using the Matlab commands "abs" and "angle"

```
X_phase = angle(X(w));
X_mag = abs(X(w));
```

In our case, $X(\omega)$ in this pseudo code is equation 7 varying over ω. This yields the magnitude and phase response for the Brune farfield ground displacement source time function:

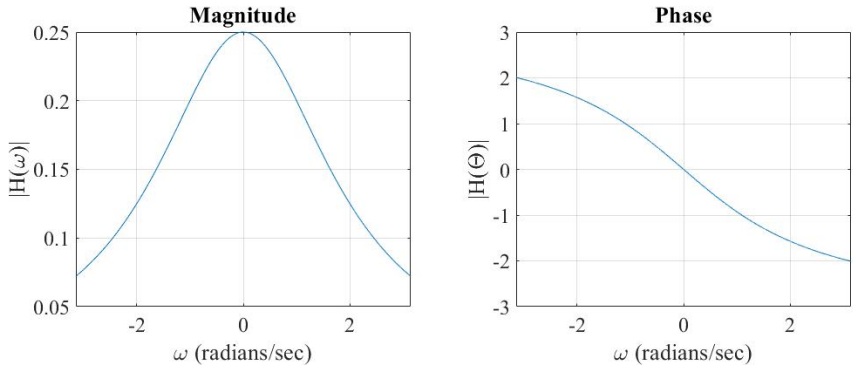


Figure 4: Plots of the magnitude and phase response of the Brune farfield ground displacement source time function from equation 3 computed using equations 8 and 9.

The result for the magnitude response in figure 4 generated from the code provided in the appendix is identical to that provided for the magnitude response in Brune (1970) equation 20, simplified:

$$|H(\omega)| = \frac{1}{\omega^2 + \alpha^2} \tag{10}$$

2) Brune farfield phase spectrum

We now explore the phase response in figure 4. For every system with a particular magnitude response, there are an infinite number of possible phase responses that can generate the magnitude response of interest. The "minimum phase response" of the system is the phase response with the property

$$\Theta(\pi) - \Theta(0) = 0 \tag{11}$$

A maximum phase system is one with the property

$$\Theta(\pi) - \Theta(0) = \pi \tag{12}$$

Minimum-phase systems have all zeros inside the unit circle whereas maximum phase systems have all zeros outside of the unit circle. A mixed-phase system is one in which some zeros lie outside of the unit circle and some inside. A minimum phase system has a stable inverse system whereas mixed-phase and maximum-phase systems may have unstable inverse systems.

We begin our study of the Brune farfield ground displacement source time function phase spectra by producing three source time functions and their corresponding phase spectra. We add a phase shift to the Brune source time function and observe how it affects the phase spectra. To add a phase shift, we rewrite equation 3:

$$u = (t - t_0)e^{-\alpha(t-t_0)} \tag{13}$$

which imposes a phase shift of t_0.

Applying the Fourier transform in equation 4 to this phase shifted Brune source model, we get:

$$X(\omega) = e^{-j\omega t_0} \int_{-\infty}^{\infty} te^{-\alpha t} e^{-j\omega t} dt \tag{14}$$

Solving the integral as we did in equation 5, our new complex spectra is

$$X(\omega) = e^{-j\omega t_0} \frac{1}{(\alpha + j\omega)^2} \tag{15}$$

The new term in front provides additional phase information to that contained in the fraction. We now compute the Brune farfield displacement source time function with three phase shifts and compute their corresponding phases responses using the Matlab code similar to the pseudo code.

```
X_phase(i) = angle(exp(-1i*ww*t0)*1/((a +1i*ww)^2));
```

The magnitude repsonses are identical for all the phase shifts.

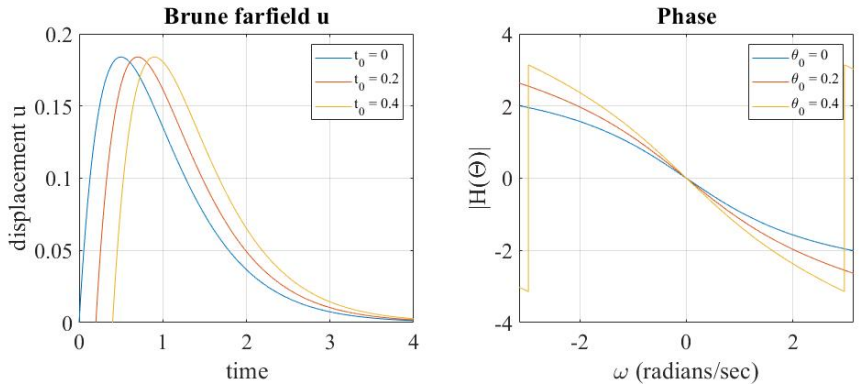


Figure 5: Plots of three Brune source time functions with different phase shifts and their corresponding phase responses.

It is clear that with no phase shift in the time domain, the blue line in figure 5, the Brune farfield ground displacement source time function phase spectra is mixed phase because equation 11 does not hold. Since mixed phase systems do not ensured a stable inverse system, the Brune source time function does not ensure a stable inverse.

3) Changes to Brune farfield source time function

The Brune source time function is difficult to interpret physically because it contains an infinite integral in the Fourier transform. Since real earthquake sources are not of function of infinite time, we propose a finite source time function with a similar shape to that of Brune's. To compute this modified source time function, we take Brune's model in equation 3 and solved on a sampled interval from 0 to 10 seconds with a sampling interval of 10,000 Hz. We then compute the magnitude and phase response of the new finite, empirical source

time function. Additionally, we computed the source time function and magnitude and phase responses for the ground velocity and acceleration source time functions which are the first and second derivatives of the displacement time function respectively.

$$v = e^{-\alpha t}(1 - \alpha t) \tag{16}$$

$$a = \alpha e^{-\alpha t}(\alpha t - 2) \tag{17}$$

where v and a are velocity and acceleration respectively. We observe amplification in magnitude and variation in phase between displacement, velocity and acceleration. In this case, the Brune source time displacement function is minimum phase, but velocity and acceleration are both mixed phase. The displacement function is minimum phase because the difference in phase between the highest and lowest frequency is 0.

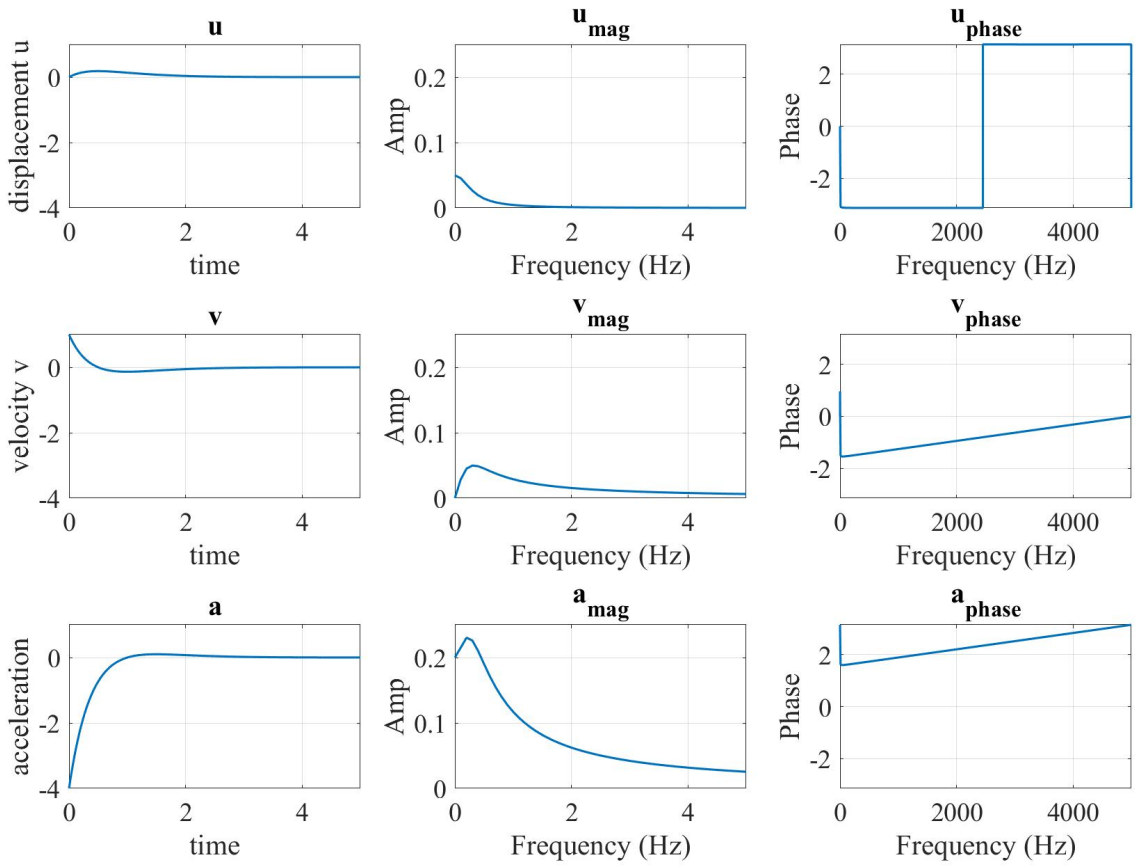


Figure 6: Plots of the finite time Brune source time functions for displacement on the top row, velocity on the second row and acceleration on the third row with corresponding magnitude and phase responses.

4) Implications of changes

The original Brune source model contains information across all time, a physically unrealistic scenario. By slightly modifying the Brune source time to make it finite in time, we maintain much of its magnitude

spectral properties and improve its phase properties. In the original model and finite model, the magnitude responses are consistent (figure 4 vs. figure 6 first row second plot). The phase responses, however, differ with the original mixed phase and the finite model minimum phase. Since the finite model is minimum phase, it ensures that the inverse system is stable whereas the original model may not have stable inverse. Having a guaranteed stable inverse allows for studies like deconvolution. Modifying the source time function to make it physical allows for invertability of earthquake data and is thus more robust than the original Brune model.

CODE LISTING

Code for figure 2)

```
% Qualifying exam
% Simplified Brune model

% Author: Marshall Pontrelli
% Date: 7/21/2019

close all
clear all
%% See eq's 1-3 in exam for simplification rationale
% This is one line of figure 2
a = 0.4;
for i = 1:100000
    t = (i-1)*0.0001;
    x(i) = t;
    uu(i) = t*exp(-a*t);
end
figure
a_04 = plot(x,uu);
grid on
box on
hold on
```

Code for figure 3)

```
% Qualifying exam
% Simplified Brune model with negative values

% Author: Marshall Pontrelli
% Date: 7/21/2019

close all
clear all
%% See eq's 1-3 in exam for simplification rationale
% This is one line of figure 2
a = 2;
for i = 1:100000
    t = (i-50000)*0.0001;
    x(i) = t;
    uu(i) = t*exp(-a*t);
end
figure
a_2 = plot(x,uu);
grid on
box on
```

```
hold on
xlabel('time')
ylabel('displacement u')
title('Brune model with negative time')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 14);
saveas(a_2, 'brune_2.jpg', 'jpg');
```

Code for figure 4)

```
% Qualifying exam
% Brune magnitude and phase response

% Author: Marshall Pontrelli
% Date: 7/21/2019

close all
clear all

%%
a = 2;
w = -pi:0.001:pi;
for i = 1:length(w)
    ww = w(i);
    X_phase(i) = angle(1/((a +1i*ww)^2));
    X_mag(i) = abs(1/((a +1i*ww)^2));
end
figure
subplot(1,2,1)
plot(w,X_mag)
xlim([-pi pi])
grid on
box on
xlabel('\omega (radians/sec)')
ylabel('|H(\omega)|')
title('Magnitude')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);
subplot(1,2,2)
plot(w,X_phase)
xlim([-pi pi])
grid on
box on
xlabel('\omega (radians/sec)')
ylabel('|H(\Theta)|')
title('Phase')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);
```

Code for figure 5)

```
% Qualifying exam
% Playing Brune magnitude and phase response

% Author: Marshall Pontrelli
% Date: 7/21/2019

% Qualifying exam
% Playing with phase

% Author: Marshall Pontrelli
```

```
% Date: 7/21/2019

close all
clear all
%% For figure 5

a = 2;
t0 = 0;
for i = 1:100000
    t = (i-1)*0.0001;
    x(i) = t;
    uu(i) = (t-t0)*exp(-a*(t-t0));
end
figure
subplot(1,2,1)
t0_0 = plot(x,uu);
xlim([0 4])
ylim([0 0.2])
grid on
box on
hold on
%%

a = 2;
t0 = 0.2;
for i = 1:100000
    t = (i-1)*0.0001;
    x(i) = t;
    uu(i) = (t-t0)*exp(-a*(t-t0));
end
subplot(1,2,1)
t0_02 = plot(x,uu);
xlim([0 4])
ylim([0 0.2])
grid on
box on
hold on

a = 2;
t0 = 0.4;
for i = 1:100000
    t = (i-1)*0.0001;
    x(i) = t;
    uu(i) = (t-t0)*exp(-a*(t-t0));
end
t0_04 = plot(x,uu);
xlim([0 4])
ylim([0 0.2])
grid on
box on
hold on
xlabel('time')
ylabel('displacement u')
title('Brune farfield u')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);
legend([t0_0, t0_02, t0_04], 't_0 = 0', 't_0 = 0.2', 't_0 = 0.4', 'FontName', 'Times New ...
    Roman', 'FontSize', 12)


%%
a = 2;
w = -pi:0.0001:pi;
for i = 1:length(w)
    ww = w(i);
    X_phase(i) = angle(exp(-1i*ww*0)*1/((a +1i*ww)^2));
```

```
end

subplot(1,2,2)
p0 = plot(w,X_phase);
hold on
%%


for i = 1:length(w)
    ww = w(i);
    X_phase(i) = angle(exp(-1i*ww*0.2)*1/((a +1i*ww)^2));
end
p02 = plot(w,X_phase);
hold on

for i = 1:length(w)
    ww = w(i);
    X_phase(i) = angle(exp(-1i*ww*0.4)*1/((a +1i*ww)^2));
end
p04 = plot(w,X_phase);
xlim([-pi pi])
grid on
box on
xlabel('\omega (radians/sec)')
ylabel('|H(\Theta)|')
title('Phase')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);
legend([p0, p02, p04], '\theta_0 = 0', '\theta_0 = 0.2', '\theta_0 = 0.4', 'FontName', ...
    'Times New Roman', 'FontSize', 12)
```

Code for figure 6)

```
% Qualifying exam
% Part 3

% Author: Marshall Pontrelli
% Date: 7/21/2019

% Qualifying exam
% Playing with phase

% Author: Marshall Pontrelli
% Date: 7/21/2019

close all
clear all

%% compute displacement
a = 2;
t0 = 0;
dt = 0.0001;
fs = 1/dt;
for i = 1:100000
    t = (i-1)*dt;
    x(i) = t;
    uu(i) = (t-t0)*exp(-a*(t-t0));
end

% Now plot
figure
set(gcf, 'Units', 'Normalized', 'OuterPosition', [0, 0.04, 1, 0.96]);
```

```
subplot(3,3,1)
t0_0 = plot(x, uu, 'linewidth', 1.5);
ylim([-4 1])
xlim([0 5])
xlabel('time')
ylabel('displacement u')
title('u')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

grid on
box on
hold on

%% Now compute the magnitude and phase response
U = fft(uu);
U_mag = 2*abs(U)/length(U);
U_phase = angle(U);

% Now compute a frequency vector
N = length(U);
fax_binsN = (0 : N-1); %samples in NS component
fax_HzN1 = fax_binsN*fs/N; %frequency axis NS (Hz)
N_2 = ceil(N/2); %half magnitude spectrum
fax_HzN = fax_HzN1(1 : N_2);
U_mag2 = U_mag(1:N_2);
U_phase2 = U_phase(1:N_2);



subplot(3,3,2)
plot(fax_HzN, U_mag2, 'linewidth', 1.5)
xlim([0 5]);
ylim([0 0.25])
xlabel('Frequency (Hz)')
ylabel('Amp')
title('u_{mag}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

subplot(3,3,3)
plot(fax_HzN, U_phase2, 'linewidth', 1.5)
xlabel('Frequency (Hz)')
ylabel('Phase')
xlim([0 5000]);
ylim([-pi pi])
title('u_{phase}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

%% compute velocity
a = 2;
t0 = 0;
dt = 0.0001;
fs = 1/dt;
for i = 1:100000
    t = (i-1)*dt;
    x(i) = t;
    uu(i) = exp(-a*t)*(1-a*t);
end
% Now plot
subplot(3,3,4)
t0_0 = plot(x, uu, 'linewidth', 1.5);
ylim([-4 1])
```

```
xlim([0 5])
xlabel('time')
ylabel('velocity v')
title('v')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

grid on
box on
hold on

%% Now compute the magnitude and phase response
U = fft(uu);
U_mag = 2*abs(U)/length(U);
U_phase = angle(U);
U_mag2 = U_mag(1:N_2);
U_phase2 = U_phase(1:N_2);

subplot(3,3,5)
plot(fax_HzN, U_mag2, 'linewidth', 1.5)
xlim([0 5]);
ylim([0 0.25])
xlabel('Frequency (Hz)')
ylabel('Amp')
title('v_{mag}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

subplot(3,3,6)
plot(fax_HzN, U_phase2, 'linewidth', 1.5)
xlabel('Frequency (Hz)')
ylabel('Phase')
xlim([0 5000]);
ylim([-pi pi])
title('v_{phase}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

%% compute acceleration
a = 2;
t0 = 0;
dt = 0.0001;
fs = 1/dt;
for i = 1:100000
    t = (i-1)*dt;
    x(i) = t;
    uu(i) = a*exp(-a*t)*(a*t - 2);
end
% Now plot
subplot(3,3,7)
t0_0 = plot(x, uu, 'linewidth', 1.5);
ylim([-4 1])
xlim([0 5])
xlabel('time')
ylabel('acceleration')
title('a')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

grid on
box on
hold on

%% Now compute the magnitude and phase response
U = fft(uu);
```

```
U_mag = 2*abs(U)/length(U);
U_phase = angle(U);
U_mag2 = U_mag(1:N_2);
U_phase2 = U_phase(1:N_2);

subplot(3,3,8)
plot(fax_HzN, U_mag2, 'linewidth', 1.5)
xlim([0 5]);
ylim([0 0.25])
xlabel('Frequency (Hz)')
ylabel('Amp')
title('a_{mag}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);

subplot(3,3,9)
fin = plot(fax_HzN, U_phase2, 'linewidth', 1.5);
xlabel('Frequency (Hz)')
ylabel('Phase')
xlim([0 5000]);
ylim([-pi pi])
title('a_{phase}')
grid on
box on
set(gca, 'FontName', 'Times New Roman', 'FontSize', 18);
saveas(fin, 'my_brune.jpg', 'jpg');
```

REFERENCES

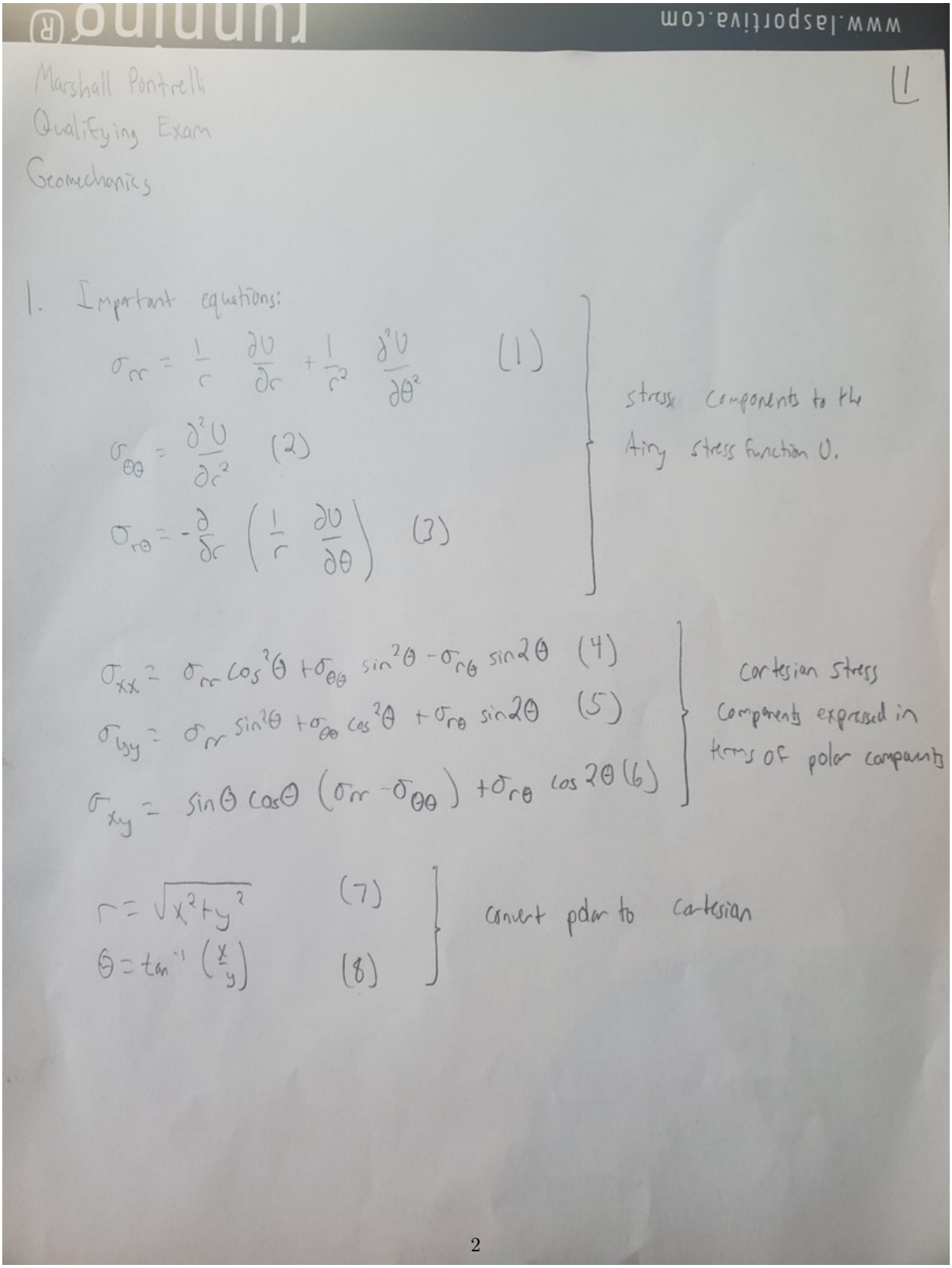
Brune, J.N., (1970) Tectonic Stress of the Spectra of Seismic Shear Waves from Earthquakes. Journal of Geophysical Research. Vol. 75, No. 26.

Part 4: Geomechanics

Marshall Pontrelli

July 24, 2020

Part 1


www.lasportiva.com
Marshall Pontrelli
Qualifying Exam
Geomechanics
1
1. Important equations:
$\sigma_{rr} = \frac{1}{r}\frac{\partial U}{\partial r} + \frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2}$ (1)
$\sigma_{\theta\theta} = \frac{\partial^2 U}{\partial r^2}$ (2)
$\sigma_{r\theta} = -\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial U}{\partial \theta}\right)$ (3)
stress components to the Airy stress function U.
$\sigma_{xx} = \sigma_{rr}\cos^2\theta + \sigma_{\theta\theta}\sin^2\theta - \sigma_{r\theta}\sin 2\theta$ (4)
$\sigma_{yy} = \sigma_{rr}\sin^2\theta + \sigma_{\theta\theta}\cos^2\theta + \sigma_{r\theta}\sin 2\theta$ (5)
$\sigma_{xy} = \sin\theta\cos\theta(\sigma_{rr} - \sigma_{\theta\theta}) + \sigma_{r\theta}\cos 2\theta$ (6)
Cartesian stress components expressed in terms of polar components
$r = \sqrt{x^2+y^2}$ (7)
$\theta = \tan^{-1}\left(\frac{x}{y}\right)$ (8)
Convert polar to cartesian

mountain

12

For our problem, the Airy stress function is

$$U = \frac{P}{\pi} r\theta\sin\theta$$

Solve for σ_{rr} using eq. (1):

$$\frac{\partial U}{\partial r} = \frac{P}{\pi}\theta\sin\theta$$

$$\frac{\partial^2 U}{\partial \theta^2} = Pr\frac{(2\cos\theta - \theta\sin\theta)}{\pi}$$

$$\text{so } \sigma_{rr} = \frac{1}{r}\frac{P}{\pi}\theta\sin\theta + \frac{1}{r^2}\frac{Pr(2\cos\theta - \theta\sin\theta)}{\pi}$$

$$= \frac{p\theta\sin\theta + 2p\cos\theta - p\theta\sin\theta}{r\pi}$$

$$\sigma_{rr} = \frac{2p\cos\theta}{r\pi}$$

solving for $\sigma_{\theta\theta}$ using eq. (2)

$$\sigma_{\theta\theta} = \frac{\partial U^2}{\partial r^2} = 0$$

$$\sigma_{\theta\theta} = 0$$

solving for $\sigma_{r\theta}$

$$\frac{\partial U}{\partial \theta} = \frac{Pr(\sin\theta + \theta\cos\theta)}{\pi}$$

$$\frac{\partial}{\partial r}\frac{1}{r}\frac{P\theta(\sin\theta + \theta\cos\theta)}{\pi} = 0$$

$$\text{so } \sigma_{r\theta} = 0$$

L3

so: $\sigma_{rr} = \frac{2p\cos\theta}{r\pi}$

$\sigma_{\theta\theta} = 0$

$\sigma_{r\theta} = 0$

now plug into eq.s 4-6

$$\sigma_{xx} = \frac{2p\cos\theta\cos^2\theta}{r\pi} + 0 - 0$$

$$= \frac{2p\cos^3\theta}{r\pi}$$

now plug eqs 7-8 in for r and θ:

$$\sigma_{xx} = \frac{2p\cos^3\left(\tan^{-1}\left(\frac{x}{y}\right)\right)}{\pi\sqrt{x^2+y^2}}$$

$$= \frac{2p}{\pi} \frac{1}{\left(\frac{x^2+y^2}{y^2}\right)^{3/2}} \frac{}{\sqrt{x^2+y^2}}$$

$$= \frac{2p}{\pi} \frac{\sqrt{x^2+y^2}}{\left(\frac{x^2+y^2}{y^2}\right)^{3/2}}$$

$$\sigma_{xx} = \frac{2p}{\pi} \frac{x^2 y}{(x^2+y^2)^2}$$ ← assumes x and y are real

now solve for σ_{yy} From eq 5:

$$\sigma_{yy} = \frac{2p\cos\theta}{r\pi}\sin^3\theta + 0 + 0$$

now plug in eqs 7-8 for r and θ:

$$\sigma_{yy} = 2p\frac{\cos\left(\tan^{-1}\left(\frac{x}{y}\right)\right)\sin^2\left(\tan^{-1}\left(\frac{x}{y}\right)\right)}{\pi\sqrt{x^2+y^2}}$$

$$\sigma_{yy} = \frac{2p}{\pi}\frac{y^3}{(x^2+y^2)^2}$$

← assumes x and y are real

now solve for σ_{xy} For eq 6

$$\sigma_{xy} = \sin\theta\cos\theta\frac{2p\cos\theta}{r\pi} + 0 + 0$$

$$= \frac{2p}{\pi}\frac{\sin\theta\cos^2\theta}{r}$$

now plug in eqs 7-8 for r and θ:

$$\sigma_{xy} = \frac{2p}{\pi}\frac{\sin\left(\tan^{-1}\left(\frac{x}{y}\right)\right)\cos^2\left(\tan^{-1}\left(\frac{x}{y}\right)\right)}{\sqrt{x^2+y^2}}$$

$$\sigma_{xy} = \frac{xy^2}{(x^2+y^2)^2}$$

← assumes x and y are real

Part 2

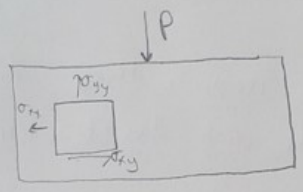
i)

2. i) To solve this problem we will use the maxwell-Betti reciprocity theorem which states: 5

"The deformation at point 1 on a linearly elastic structure due to a load at point 2 equals that at point 2 due to some load at point 1".

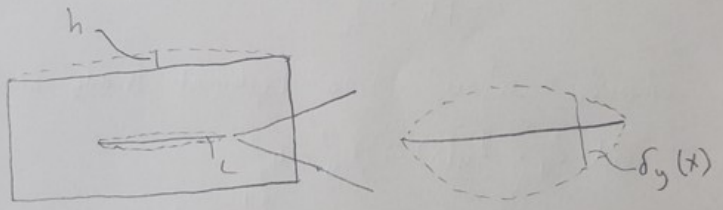
Using superpositions we separate our problem into 1) the displaced surface problem and 2) the line force P acting downward.

Problem 1:



+

Problem 2:



Let's call the point at the surface "point 1". We know that the work done at this point 1 is equal to the force from problem 1 times the displacement in problem 2 or

$$\text{Work}_{\text{point 1}} = -Ph$$

Now let's look at the line "C". We know that the work done on line "C" is equal to the force from problem 1 or the sum of the stresses over the line

$$\int_C \sigma_{ij}\, dC$$

times the displacement from problem 2, or the sum of the displacements over the line:

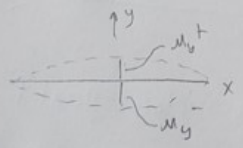
$$\int_C \delta_y(x)\, dC$$

thus, using Maxwell-Betti's reciprocity theory which states that the work done at point 1 equals the work done at line "C", we know that

$$-P_h = \int_C n_i \sigma_{ij}^{(2)} \delta_j^{(1)}\, dC$$

where $\delta_j = u_j^+ - u_j^-$, since displacement is only in the y direction, $\delta_x = 0$

and $\delta_y = \delta_y(x) = u_y^+ - u_y^-$



ii)

ii) Let's begin with the filament solution

$$\frac{\sigma_{yy}^{(2)}}{p} = -\frac{2}{\pi}\frac{y^3}{(x^2+y^2)^2} \qquad (1)$$

we will ignore σ_{xy} because $n_x \delta_y = 0$

we will ignore σ_{xx} because $\delta_x = 0$, thus $\sigma_{xx}\delta_x = 0$. Starting with eq. 1, we plug $\sigma_{yy}^{(2)}$ into the results from part i.

$$-ph = \int_C n_y \sigma_{yy}^{(2)} \delta_y^{(1)} dC$$

$$-ph = \int_C n_y \frac{-2p}{\pi}\frac{y^3}{(x^2+y^2)^2} \delta_y^{(1)} dC$$

given that $\delta_y^{(1)}(x) = \delta_0$,

$$-ph = \int_C \frac{-2p}{\pi}\frac{y^3}{(x^2+y^2)^2} \delta_0 \, dC$$

Since $\frac{-2p}{\pi}$ and δ_0 are constants, we pull them out of the integral:

$$-ph = \frac{-2p}{\pi}\delta_0 \int_C \frac{y^3}{(x^2+y^2)^2} dC$$

because we are evaluating h at x=0

$$\frac{h(0)}{\delta_0} = \frac{2}{\pi}\int_C \frac{y^3}{(x^2+y^2)^2} dC$$

now plugging in $y=d$ and $x=a$, we get

$$\frac{h(0)}{\delta_0} = \frac{2}{\pi} \int_C \frac{d^3}{(a^2+d^2)^2} \, dC$$

Since our crack has d as a constant and a varying from $-a$ to $+a$, the integral becomes:

$$\frac{h(0)}{\delta_0} = \frac{2}{\pi} \int_{-a}^{a} \frac{d^3}{(a^2+d^2)^2} \, da$$

which is:

$$\frac{h(0)}{\delta_0} = \frac{2}{\pi} \left[\frac{a/d}{1+(a/d)^2} + \arctan\left(\frac{a}{d}\right) \right]$$

$$\frac{h(0)}{\delta_0} \to 0 \quad \text{as} \quad a/d \to 0$$

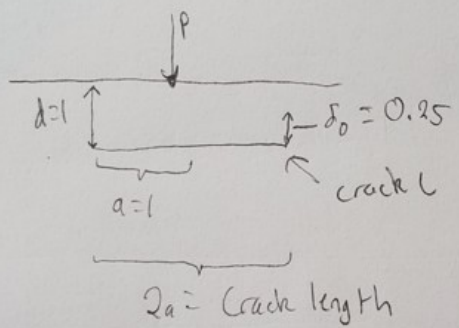
$$\frac{h(0)}{\delta_0} \to 1 \quad \text{as} \quad a/d \to \infty$$

iii)

iii) Following the same steps as problem ii), we arrive at the integral for uplift

$$h(x) = \delta_0 \frac{2}{\pi} \int_{-a}^{a} \frac{d^3}{((a-x)^2 + d^2)^2} \, da$$

We evaluate this integral numerically with $d=1$, $\delta_0 = 0.25$ on $|x| = a = 1$ along the free surface from $x=-5$ to $x=5$.



This uplift profile is in Figure 1.

As we get far away from the crack, $(a-x)^2 \to x^2$, thus the uplift decay profile decays at a power of 4.

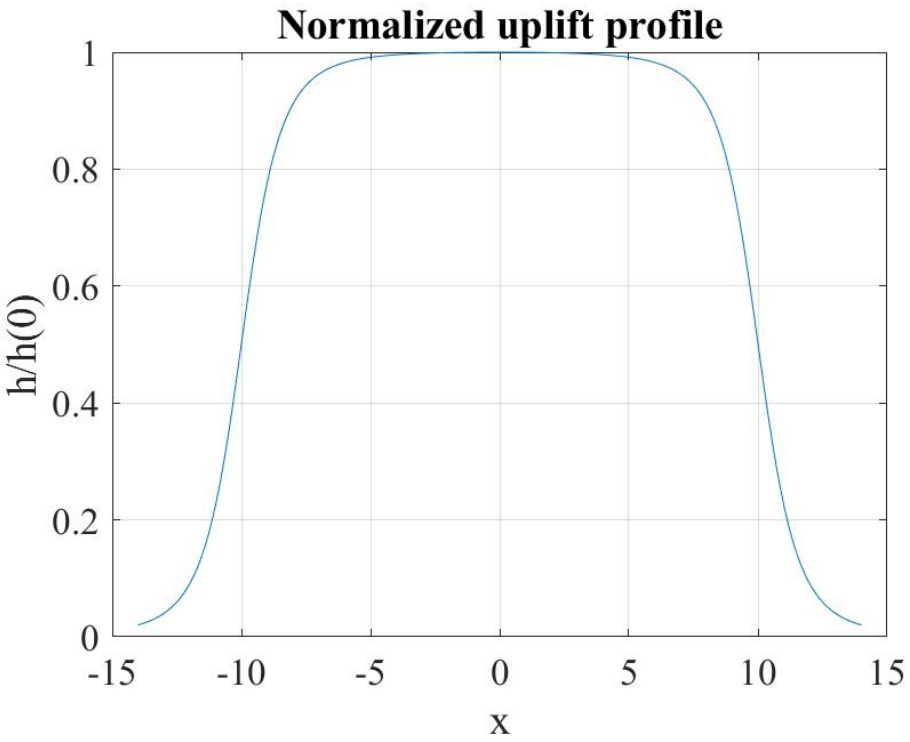


Figure 1: Uplift profile normalized by h(0) with a = 10, d = 2, and $\delta_0 = 1$.

Code for figure 1)

```
% Qualifying exam
% Geomechanics
% 2 iii

% Author: Marshall Pontrelli
% date 7/22

close all
clear all

%% inputs
a = 10; % half of the crack length
d = 2; % the depth of the crack
d0 = 1; % the value of the displacement within the crack
da = 0.01; % integral step
a_vec = -a:da:a; % summing vector over 2a, the crack length
x_vec = -a-4:da:a+4; % summing vector for x prime
integral = zeros(1,length(x_vec));
for q = 1:length(x_vec)
    x_prime = x_vec(q);
    h = zeros(1,length(a_vec));
    for i = 1:length(a_vec)
        a = a_vec(i);
        h(i) = d^3/(((a - x_prime)^2 + d^2)^2);
    end
    % now integrate using trapezoids
    int = zeros(1,length(h));
    for i = 1:length(h)-1
```

```
            int(i) = da*((h(i) + h(i+1))/2);
        end
        integral(q) = (d0*2/pi)*sum(int);
    end
    figure
    geo = plot(x_vec', integral/max(integral));
    grid on
    box on
    xlabel('x')
    ylabel('h/h(0)')
    title('Normalized uplift profile')
    set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
    saveas(geo, 'geo1.jpg', 'jpg');
```

iv)

To fabricate the data we select random values of a between 10 and 20, of d between 1 and 10 and of d_0 between 0.1 and 1 and compute the non-normalized uplift profile. From the non-normalized uplift profile, we add Gaussian noise with a mean of zero and standard deviation $d_0/10$. The final fabricated uplift profile is:

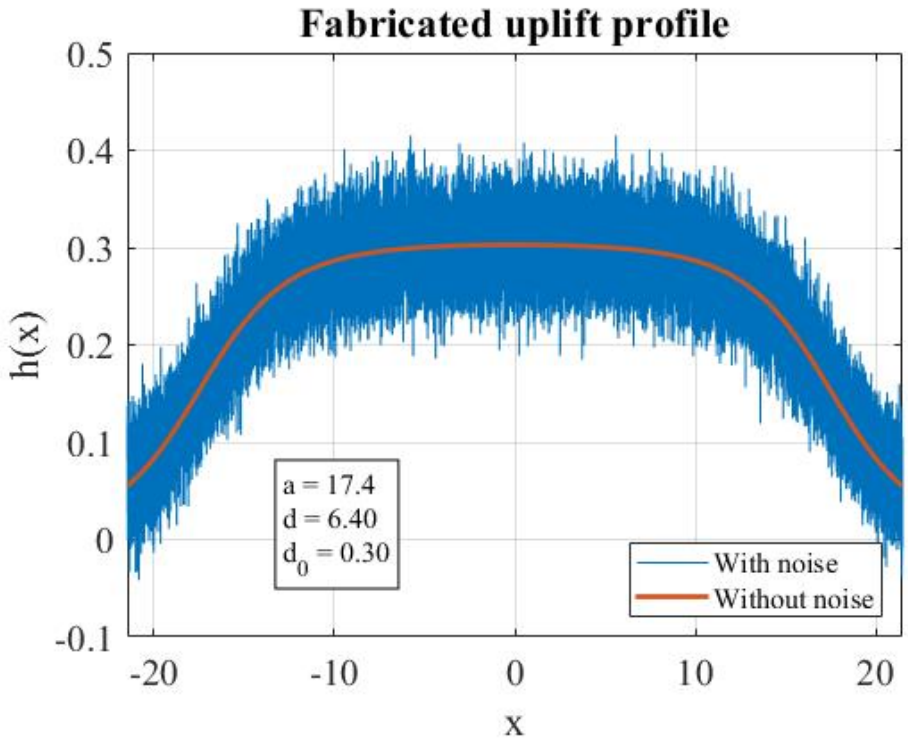


Figure 2: Fabricated uplift profile with and without noise with randomly selected a, d and d_0.

With a fabricated uplift profile, we solve for a, d and d_0 using a basic minimization algorithm that minimizes the sum of squared residuals between the fabricated data and simulated data. For simplicity of the demonstration, we keep a the same as the fabricated data and search a grid of d values between 1 and 10 and d_0 values between 0.1 and 1. The results show a non-unique minimization problem with many possible fits.

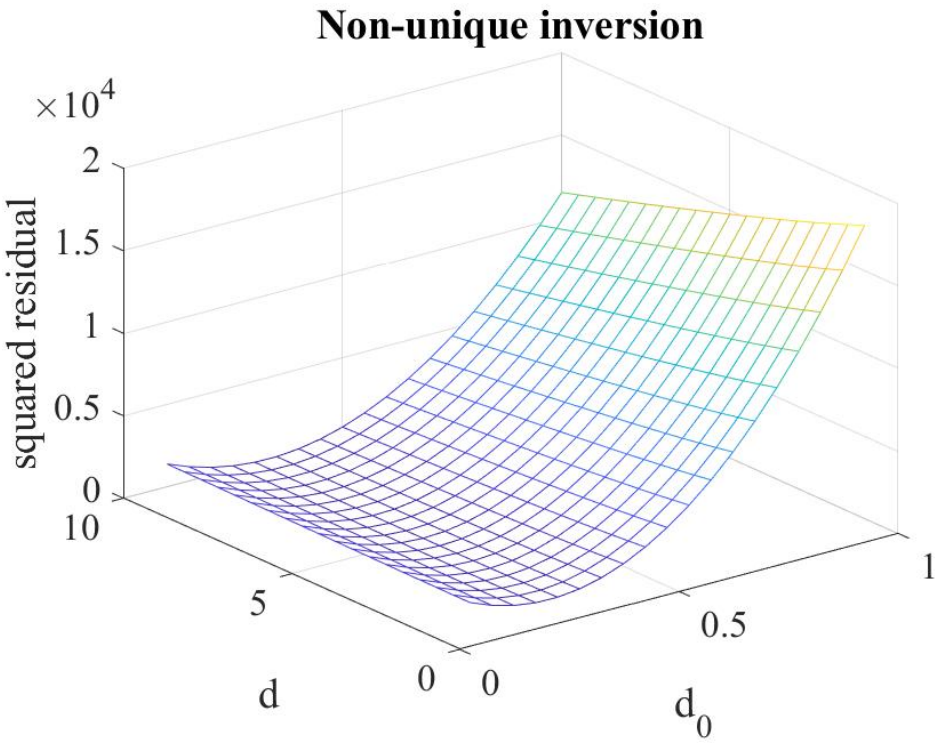


Figure 3: Results of the sum of squared residuals of a between the fabricated data in figure 2 and a set of uplift profiles generated from a grid of d values between 1 and 10 and d_0 values between 0.1 and 1.

It is clear in figure 3 that many possible d and d_0 combinations can fit the fabricated data in figure 2. As d_0 or crack width increases there are solutions that fit the data with an increase in depth. We could improve the posing of this question by providing geological constraints for possible magma sill or laccolith depths or crack thicknesses. If these constraints exist, we shrink the space of possible solutions for our inversion problem.

Code for figures 2 and 3)

```
% Qualifying exam
% Geomechanics
% 2 iv, fabricated uplift profile

% Author: Marshall Pontrelli
% date 7/22

close all
%clear all

%% first randomly selct a, d and d0
a = 10 + 10*rand(1); % half of the crack length between 1 and 10
d = 1 + 9*rand(1); % the depth of the crack between 1 and 10
d0 = 0.1 + 0.9*rand(1); % the value of the displacement within the crack between 0.1 and 1
da = 0.001; % integral step
a_vec = -a:da:a; % summing vector over 2a, the crack length
x_vec = -a-4:da:a+4; % summing vector for x prime
integral = zeros(1,length(x_vec));
for q = 1:length(x_vec)
```

```
        x_prime = x_vec(q);
        h = zeros(1,length(a_vec));
        for i = 1:length(a_vec)
            a = a_vec(i);
            h(i) = d^3/(((a - x_prime)^2 + d^2)^2);
        end
        % now integrate using trapezoids
        int = zeros(1,length(h));
        for i = 1:length(h)-1
            int(i) = da*((h(i) + h(i+1))/2);
        end
        integral(q) = (d0*2/pi)*sum(int);
    end



    %% now add some gaussian
    noise = normrnd(0, d0/10, [1, length(integral)]);
    gaus_prof = integral+ noise;

    %% Now you have the uplift profile so plot it and the noise
    figure
    noisy = plot(x_vec', gaus_prof);
    hold on
    geo = plot(x_vec', integral, 'linewidth',2);
    grid on
    box on
    xlabel('x')
    ylabel('h(x)')
    title('Fabricated uplift profile')
    set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
    xlim([-a-4 a+4])
    legend([noisy, geo], 'With noise', 'Without noise', 'FontName', 'Times New Roman', ...
        'FontSize', 12, 'location', 'southeast')
    atex = strcat('a =', {' '}, num2str(a));
    dtex = strcat('d =', {' '}, num2str(d));
    d0tex = strcat('d_0 =', {' '}, num2str(d0));
    atex = atex{1};
    atex = atex(1:8);
    dtex = dtex{1};
    dtex = dtex(1:8);
    d0tex = d0tex{1};
    d0tex = d0tex(1:10);
    str = {atex, dtex, d0tex};
    text(-10,-0.1,str, 'FontName', 'Times New Roman', 'FontSize', 12, 'EdgeColor','k', ...
        'BackgroundColor','w')
    saveas(geo, 'geo2.jpg', 'jpg');

    %% Now that you've got the fabricated data, now try to recreate it
    diff = zeros(19,19);

    for u = 1:19
        d = (u+1)/2; %loop over a grid of depths
        for v = 1:19
            d0 = (v+1)/20; %loop over a grid of crack thicknesses
            integral2 = zeros(1,length(x_vec));
            for q = 1:length(x_vec)
                x_prime = x_vec(q);
                h = zeros(1,length(a_vec));
                for i = 1:length(a_vec)
                    a = a_vec(i);
                    h(i) = d^3/(((a - x_prime)^2 + d^2)^2);
                end
                % now integrate using trapezoids
                int = zeros(1,length(h));
```

```
            for i = 1:length(h)-1
                int(i) = da*((h(i) + h(i+1))/2);
            end
            integral2(q) = (d0*2/pi)*sum(int);

        end
        diff(u,v) = sum((gaus_prof - integral2).^2); % solve for the sum squared residuals
        disp(v)
    end

end

%% Now plot heatmap
figure
y = 1:0.5:10; % vector for depth
x = 0.1:0.05:1; %vector for d0
geo3 = mesh(x,y,diff);
xlabel('d_0')
ylabel('d')
zlabel('squared residual')
title('Non-unique inversion')
set(gca, 'FontName', 'Times New Roman', 'FontSize', 16);
saveas(geo3, 'geo3.jpg', 'jpg');
```

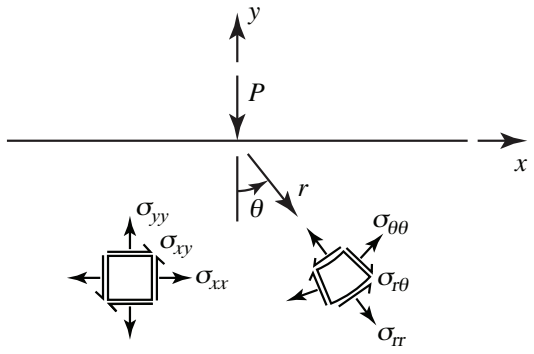


Figure 1: Figure for problem 1.

1. In class we found the appropriate Airy stress function for a line load of magnitude P normal to a half-plane was

$$U = -\frac{P}{\pi} r\theta \sin\theta$$

where the polar coordinates θ and r are as defined in the problem figure. Using the appropriate change of coordinates, show that for the Cartesian coordinates chosen in the problem, the stress components in that coordinate system are expressed as

$$\sigma_{xx} = \frac{2P}{\pi}\frac{x^2 y}{(x^2+y^2)^2}, \quad \sigma_{yy} = \frac{2P}{\pi}\frac{y^3}{(x^2+y^2)^2} \quad , \sigma_{xy} = \frac{2P}{\pi}\frac{xy^2}{(x^2+y^2)^2}$$

2. Imagine the plane-strain problem of a below-ground surface C that was once intact but is now pulled apart, leaving discontinuous displacements u_j^+ and u_j^- on each side of C.

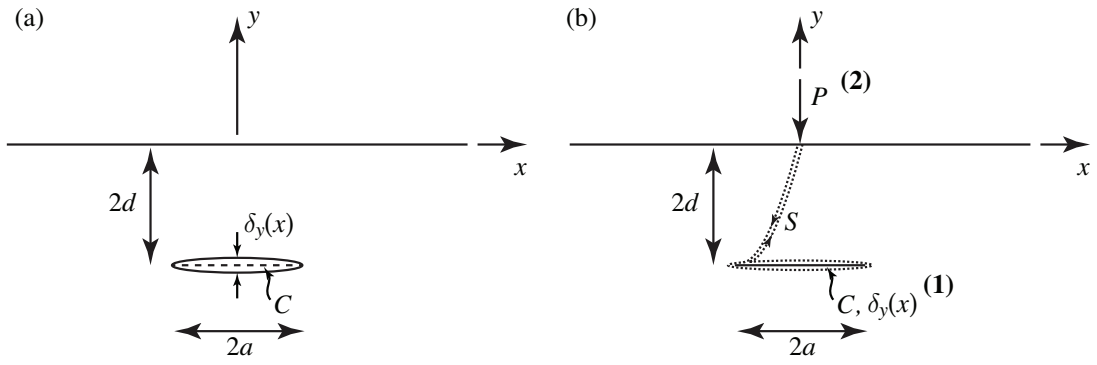


Figure 2: (a) A plane-strain crack of length $2a$ and a distribution of opening $\delta_y(x)$ lies a distance $2d$ beneath a free surface. The opening is small such that the departure of the crack walls from the straight line C is small. (b) We look to make use of the Maxwell-Betti reciproricty theorem to determine uplift using (i) a convenient choice of contour S, (ii) the problem (1) of the opened crack and (iii) the problem (2) of a vertical point force at the surface.

(i) We are interested in finding the vertical uplift h of the ground surface at the point $x = 0$, $y = 0$. One method is to use the reciprocal theorem

$$\int_S n_i \sigma_{ij}^{(1)} u_j^{(2)} dS = \int_S n_i \sigma_{ij}^{(2)} u_j^{(1)} dS$$

where (1) and (2) denote the stresses σ_{ij} and displacements u_j corresponding to the two different static problems.

We take problem (1) to correspond to the displaced surface problem and problem (2) to be that of a line force P acting downward (a known solution). Show with a convenient choice of a closed contour S (e.g., enveloping C as in Figure 2b) that the following relation holds

$$-Ph = \int_C n_i \sigma_{ij}^{(2)} \delta_j^{(1)} dC$$

where $\delta_j = u_j^+ - u_j^-$ are the components of the displacement discontinuity.

(ii) The solution to problem (2) is the Flamant solution examined in Problem **1**, for which one component of stress is

$$\frac{\sigma_{yy}^{(2)}}{P} = -\frac{2}{\pi}\frac{y^3}{(x^2+y^2)^2}$$

We may use this solution and the result in **(i)** to calculate the surface displacements, say, due a plane-strain crack at a depth d, running parallel to the surface and whose opening is given by

$$\delta_y^{(1)}(x) = \delta_0 \quad \text{on } |x| < a$$

Evaluate the integral for the surface uplift h at $x = 0$ and show that its value is

$$\frac{h(0)}{\delta_0} = \frac{2}{\pi}\left[\frac{a/d}{1+(a/d)^2} + \arctan(a/d)\right]$$

What are the first-order approximations to $h(0)/\delta_0$ as $a/d \to 0$ and $a/d \to \infty$?

(iii) Now consider the problem of determining the distribution of uplift along the surface $h(x)$. We may do so by shifting the load to an arbitrary position $x = x'$ on the surface, for which the relevant stress component is now given as

$$\frac{\sigma_{yy}^{(2)}}{P} = -\frac{2}{\pi}\frac{y^3}{((x-x')^2+y^2)^2}$$

The integral involved in calculating $h(x)$ will be a convolution, $h(x) = \int_{-1}^{+1} f(x,x')dx'$. Find $f(x,x')$. For chosen values of a, d, and δ_0, evaluate the integral numerically for various positions x to arrive to a discrete approximation for $h(x)$. Plot the uplift profile, normalized by $h(0)$. Does the uplift profile decay as a power law at large x? What is its power?

(iv) Fabricate an imaginary data set of surface uplift observations at a finite set of positions x_k. Imagine the uplift is maximum at $x = 0$ and decays as $|x|$ becomes large. In trying to infer the cause of this uplift, let's attempt to model these observations as the emplacement of a magma sill or laccolith much longer than it is wide, which corresponds roughly to the scenario sketched in Figure 2a. Using your capabilities developed in (iii) to quickly calculate the surface uplift profile $h(x)$ for a choice of a, d, and δ_0, attempt to infer the sill's depth $2d$, its width $2a$ and it's opening δ_0. Is this inverse problem well-posed? That is, discuss the likelihood for a unique set of parameters providing a best fit: i.e., can you find comparable fits for different choices of d, a, and δ_0. How would you constrain the plausible range of their values?

Marshall Pontrelli's Qualifying Exam

CEE247 Geotechnical Earthquake Engineering

July 20, 2020

Site effects in Anchorage Alaska

The November 30, 2018 Anchorage, Alaska earthquake was an important earthquake in terms of studying site response. Anchorage, Alaska had also been impacted by the 1964 event and therefore had been well studied in terms of site characterization and instrumentation.

Using existing and publicly available ground motion and site characterization data, identify a site in or around the city of Anchorage as an example of earthquake site effects. Use the information/data that you find to:

- Provide the geologic context for earthquake site response in the region and specifically the site
- Characterize the site effects for the site (from site characterization AND observations)
- Summarize the observed ground motions for the site in terms of the seismic hazard of the site (how do the ground motions compare to design ground motions)

Feel free to draw from past studies on site response in Anchorage as well as site characterization studies for the region. Feel free to use ground motion data to estimate site effects at the site of your choice, but be sure to put that site effect in context. This question is about connection geology – seismology – and hazard around the topic of site response.

Include a summary of your references as an appendix.

Time Series Analysis Qualifying Exam Question for Marshall Pontrelli, by John E. Ebel

Exploring the Brune Source Spectrum Model

The Brune source spectrum model is by far the most popular earthquake source model used by seismologists, and yet seismologists seem unaware of an aspect of this model that is not physical. In this question you will explore the unphysical problem with the Brune source model, propose a solution to fix the problem, and assess the implications of fixing the Brune source spectrum model.

To answer this question, you will need to obtain a copy of the paper on the Brune source model (JGR, vol. 75, no. 26, pp. 4997-5009, 1970).

(1) The first thing to do is to explore the unphysical nature of the Brune source model. The Brune source spectrum comes from taking the Fourier Transform of a farfield ground displacement source time function, which is given in Brune's JGR paper. Please write out the time-domain form of the Brune source and make a plot of its shape. Please explain what makes this source shape unphysical. Also, please show the mathematics (i.e., work out the mathematics of its Fourier Transform) that derives the Brune source spectral shape from this time-domain model.

(2) What is the phase spectrum of the Brune farfield ground displacement source time function? Please derive and make a plot of the phase spectrum of the model. Do an analysis to determine whether this source time function shape is minimum phase, mixed phase, or maximum phase.

(3) What changes would you make in Brune's farfield ground displacement source time function to make it be a physical shape (note that there are many possible answers here). For the change that you suggest, determine the source spectral shapes (both amplitude and phase spectra) that your modified Brune source time function would have. Compute those spectral shapes for ground displacement, ground velocity and ground acceleration. If you can, present both an analytical analysis of your modified Brune source shape as well as plots of the modified amplitude and phase spectra. Also, does your modification to the Brune source time function change whether the Brune source is minimum phase, mixed phase or maximum phase?

(4) Conclude by discussing the implications of your answers to the above questions to analyses of data using the original Brune source model. Does modifying the Brune source time function to make it physical change in any significant way the results of past analyses of earthquake data that have been carried out with the original, unphysical Brune source model?

As a PhD you may be called upon to discuss or provide opinion on a great many things that are not, strictly speaking, your area of primary training. You may even get a job doing something quite different than your current research interest. In the following three-part question, the topics and format will jump around a bit, but the point is to see how you reason through problems. This is an opportunity for us to have an in-depth discussion about reasoning with data, probability, and how to operate as an independent researcher.

Part 1: COVID testing and speaking outside your comfort zone

Public health experts think that high levels of testing (perhaps 100,000,000 a week) is necessary to truly mitigate the spread of COVID19. Some experts have suggested pooling samples, where many people's specimens are combined and then tested as a group. If the group tests negative, you have cleared many people with one test. If the group tests positive, then members of that group are tested again. Many strategies for this second step exist, but let's just assume each member of a group that tests positive is tested individually. Ignoring all of the complexities (false results, logistics of merging specimens, etc.), let's explore if this is a reasonable thing to do.

Assuming we have some idea of the underlying rate of infection p and that people are independently infected, what is the distribution of the number of positives if k tests are conducted? What is the probability of finding at least one positive if k people are tested?

Now, suppose we divide some population of N into groups of size k. What is the expected number of tests it would take to clear the entire population?

If you are President Monaco, and would like to test all 11,000 Tufts students each week, how many tests can you expect to conduct with a pool group size of three and an underlying infection rate of 1%? What if it is 2%, or 5%? At what point does pooling stop making sense?

Naturally the next question that arises is how large the groups should be? Suppose President Monaco wanted to minimize the number of tests conducted to clear every student. What is the pool size if the infection rate is 1, 2, or 5 percent, and how much capacity would we want to build (i.e. what is the associated expected number of tests)? What simplifications have you made above that may limit the accuracy of your results?

The idea of pooled testing emerged during WWII, when the American military was inducting millions of new recruits that had to be tested for STDs. Since then, the idea of pooled testing has found use in medicine, manufacturing, and electrical engineering, and sensor network testing.

Part 2: Monte Carlo Analysis and Probability Theory

As geo-scientists much of our work is empirical. Define "empirical" and describe why this might present a challenge for geo-scientists computing statistics from observed data. What are the implications for your work: *how do you know what is so and what aint so*?

One sensible solution is to do work where you know precisely what is so, perhaps because you conveniently defined it. Describe how this relates to Monte Carlo analysis, and its power and drawbacks for us as geo-scientists. How does the concept of robustness relate to the latter?

These questions can have real consequence for research. Most statistics are founded on theory, and most applications deviate from some assumptions. Suppose you were computing Pearson's correlation coefficient on nominally bivariate log-Normal data. Stedinger [1981] provided an alternative estimator for the correlation coefficient assuming bivariate log-Normality:

$$r_1 = \frac{\exp(\sigma_{UV}) - 1}{\sqrt{(\exp(\sigma_V^2) - 1)\,(\exp(\sigma_U^2) - 1)}}$$

where

$$\sigma_{UV} = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})\,(v_i - \bar{v})$$

$$\sigma_U^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2$$

$$\sigma_V^2 = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2$$

$$u_i = \ln(o_i)$$

$$v_i = \ln(s_i)$$

Here o_i and s_i are log-Normal variates. Use Monte Carlo Analysis to speak to the behavior of r_1 versus Pearson's r, assuming the true real-space correlation is 0.7. Test three cases: when the coefficient of variation of O is 0.5, 2, and 10. What do your results suggest for the reliability and bias in Pearson's r? What ramifications could this have for your research?

Part 3: Baseball and acquiring new tools

Baseball is an excellent sport for statistics because of its rigorous data collection. The attached Excel file "MLB_Salaries.xlsx" contains statistics for the 1986 Major League Baseball season. Our aim is two find which factors are most predictive of 1987 salary. Some of the features are categorical, some are numerical. Describe at least one method that is capable of dealing with these data, how you propose to implement these methods or acquire packages to do so. Basically, take me though your process for solving this problem.

Focusing for a moment on just the numerical data, use any method you like to predict salary based on performance. What is the expected salary (expected mean) of a player who has a lifetime batting average of 0.300, has played for 10 years, and batted 0.300 in 1986? What is the probability a player's 1987 salary is over a million dollars if he batted 0.300 in 1986 and has played 10 years? What is the 95% confidence interval on that probability?

Do any players standout as unusual, in terms of their salary relative to their performance? With a small dataset like this one, you could answer that problem manually. How would you do it if you had 24 features for 200,000,000 data points? If time permits, demonstrate how you would do this. This later point is particularly important to harness the information age.