Probability and Statistics for Seismology and Structural Reliability (CENGM0078)
Coursework Assignment

Student Number: 2041710

1. DESCRIPTIVE STATISTICS, DATA ANALYSIS AND MODELLING

1.1 Measures

Firstly, the concrete data sets are considered individually. The central measures of the two samples include the mean, median and mode. The sample mean of the concrete data for storey 1 and 2 is 25.71 MPa and 20.04 MPa respectively. The median of the storey 1 sample is 25.46 MPa, whereas that of storey 2 is 19.44 MPa. Another descriptive feature of the samples is the dispersion, which describes the scatter of data around the sample mean (Nowak and Collins, 2013). The most common measure of this is variance, through which the standard deviation can be calculated. Storey 1's sample data has a variance of 103.81 MPa, whereas storey 2 has a variance of 16.07 MPa. Furthermore, the standard deviation of the storey 1 sample was found to be 10.19 MPa and for storey 2 sample this was value 4.09 MPa. Finally, the coefficient of variation is another measure of dispersion, which was found to be 0.40 MPa for storey 1 and 0.51 MPa for storey 2. It is seen that the coefficient of variation for storey 2 is greater than for storey 1. This indicates that the data observed for storey 2 are more dispersed than in storey 1.

Furthermore, the features of the two samples were explored jointly. The sample covariance is useful in determining the correlation between two observed data sets, this value from the two samples was found to be 40.3694, as this is a positive value, it indicates a a positive correlation between the observed data sets. The covariance can be further normalised with respect to the sample's standard deviation, this leads to the correlation coefficient. Limited in the range of -1 to 1, the coefficient of correlation in this case is 0.96, this is quite a high value that suggests a strong positive linear relationship between both samples. Graphical representations of the data were also studied in order to effectively communicate the features. In Figure 1, the quantile plots for both samples are illustrated (on the same scale to help comparison). For storey 1, the slope starts quite steep and continues this trend, the steeper slope implies a lower concentration of data. This means that the values of the storey 1 sample are quite different from one another. In contrast, the storey 2 quantile plot has a flatter slope, implying a higher concentration of data i.e. there are a lot of observations with the same/similar values. For storey 2, the slope is fairly flat up to about a quantile index of 0.9. Thereafter, the slope increases and thus the concentration of data is smaller. It should be noted that quantiles values can also be read off this plot. The symmetry of the data can also be explored. However, as seen the upper shape of both plots does not mirror the lower half, suggesting that the observations in both sets are not symmetrical dispersed around the median.

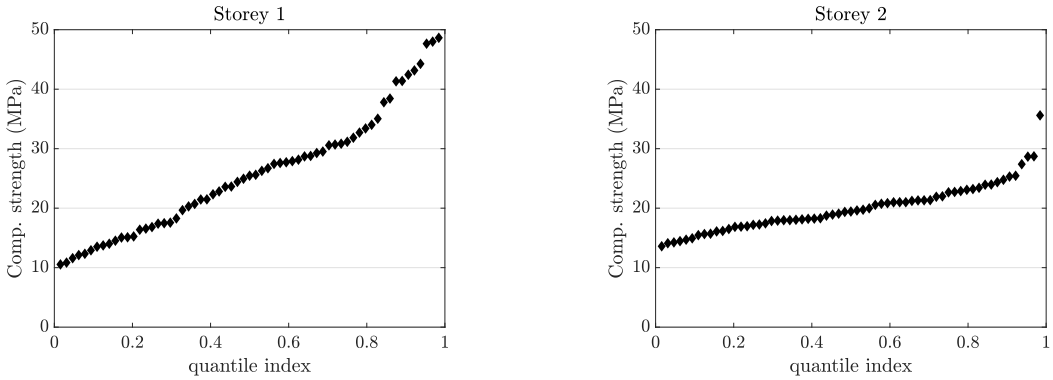

Figure 1: Quantile plots of the observed concrete data

1.2 Normal and Lognormal models (Probability plot, Chi Square and KS goodness of fit)

Firstly, the normal and lognormal models were explored using the probability plot. This aids in the selection of the most suitable model distribution, the process followed, is outlined by (Faber, 2012). Summarised as follows, the data sets were ordered, then cumulative distribution function (CDF) of both models was calculated using, $F_X = \frac{i}{N+1}$, i where data id and N is the number of data points. Thereafter, the inverse of the CDF was calculated using the *'norminv'* function in (MATLAB, 2010). Consequently, the inverse of the CDF (y-axis) was plotted

against the data (x-axis), finally a fitting process was applied using (MATLAB, 2010) and this is depicted in Figure 2. Through the probability plot, the distribution parameters may be determined. The equations of the lines are also illustrated in Figure 2. The equations for the normal model can be given by equation 1.

$$x = \mu + \sigma\Phi^{-1}(F_X) \quad 1.$$

where, x is the data values, μ and σ are the mean and standard deviation values.

On the other hand, the equation of the lognormal model can be given by equation 2.

$$\ln(x) = \ln(\eta) + \beta\Phi^{-1}(F_X) \quad 2.$$

where, x is the data values, η and β are the lognormal model parameters.

For the normal model shown in Figure 2, the plots of both storeys seem to have an S-shape and they both show a non-linear pattern. Furthermore, both plots are skewed right, although the middle points agree with the fit, the first few and last few points deviate from the line, this is more pronounced for the storey 1 data. Therefore, it can be concluded from these points that the normal distribution will not be a suitable model for both concrete samples.

On the other hand, for both samples, the lognormal probability plots in Figure 2 show a fairly strong linear pattern, a majority of the data points exhibit close agreement with the line of fit. This indicates that the concrete sample data for both storeys could be reasonably modelled with lognormal distribution.

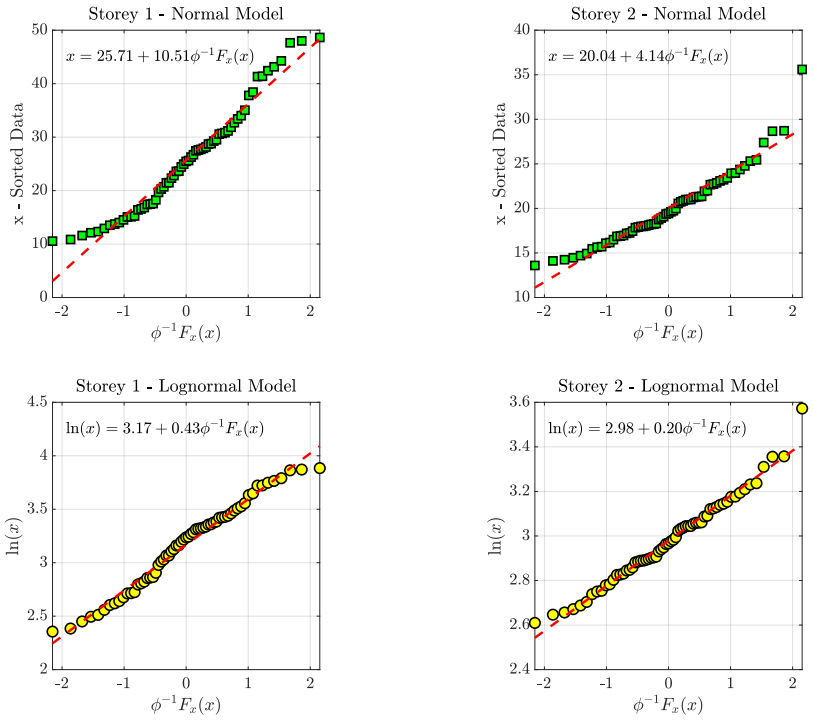

Figure 2: Probability plots of the concrete samples for the normal and lognormal models

Moreover, the models were investigated using the Chi-square test. Both samples were first binned into k intervals, with the number of intervals calculated from Sturges rule (Sturges, 1926). Some of the intervals, after discretisation had less than 5 values of data, therefore as per (Benjamin and Cornell, 2014), they were lumped with adjacent intervals. The value of k is 6 and 4 for storeys 1 and 2 respectively. Then, the observed frequencies and predicted frequencies were computed, from these frequencies, the Chi square statistic, e_m^2, was found using (Faber, 2012). The test was applied to determine whether the distributions are suitable, at a 5% significance level. As the two model parameters for each of the normal/lognormal were estimated from the sample data, the degree of freedom (df) needs to be reduced, it is given by: $df = n - 1 - m$, where n is the number of bins and m is the number of estimated parameters, 2. Hence, degree of freedom for storey 1 is 3, whereas for storey 2 it is 1. Using *'chi2inv'* from (MATLAB, 2010), which has inputs of df and the chosen significance level, the key values were calculated. The null hypothesis, H_0, is that the assumed distribution function is not in gross contradiction with the observed data, H_0 is not accepted if the value of e_m^2 is greater than the significance level (0.05).

The statistic, e_m^2 for the normal model was found to be 9.87 and the critical value was found to be 7.81, as e_m^2 is greater, the null hypothesis cannot be accepted. For storey 2, the null hypothesis is accepted as e_m^2 (2.84) is less than the critical value (3.84). As for the lognormal model of storey 1, e_m^2 (8.24) is greater than the critical value (7.81), hence H_0 is rejected. In addition, the storey 2 sample also has an e_m^2 (0.68) that is less than the critical value (3.84), so again for storey 2, H_0 is accepted. These results imply that the in the case of the storey 1 sample, both the normal and lognormal distributions are in contradiction with the data. Contrastingly, on the basis of the models, one can conclude that: the normal distribution and the lognormal distribution for storey 2, is not in gross contradiction with the observed data. The results can also be plotted on a histogram as shown in Figure 3. It can be observed that for storey 1, the normal and lognormal models are not suitable as there are big differences between the observed and expected values for some intervals (10-15.6 for storey 1 particularly). For storey 2, the normal and lognormal models seem to be a good fit of the data (smaller differences between observed/expected).

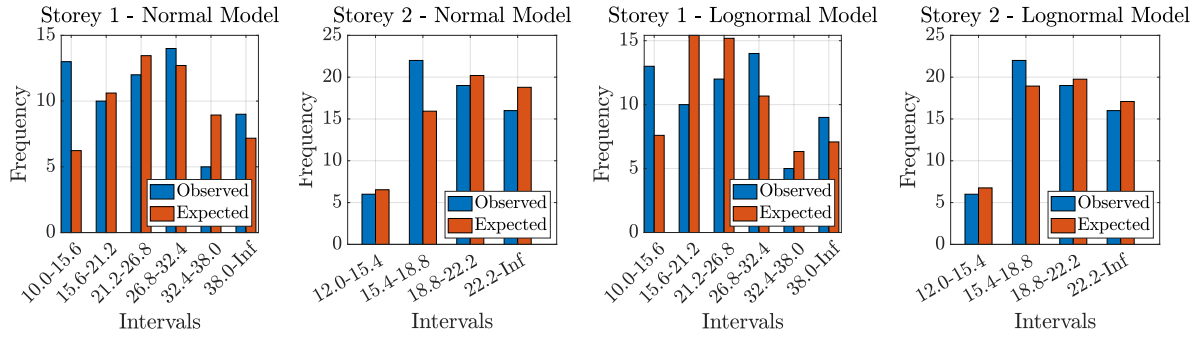

Figure 3: Expected and sample histograms of the concrete data for the normal and lognormal models.

Lastly, the Kolmogrov-Smirnov goodness test was carried out, where the null hypothesis, H_0 (observed data follows the postulated CDF), cannot be accepted if the maximum error is greater than a critical value. In this case, the sample size is 63, therefore at the 5% significance level, from tables (Faber, 2010), the critical value is obtained to be 0.17. The maximum error of the normal distribution for storeys 1 and 2 are 0.085 and 0.086, as this is lower than the critical value, H_0 cannot be rejected. The lognormal distribution resulted in a maximum error of 0.087 and 0.057 for storeys 1 and 2 respectively, again as this is lower than the critical value, H_0 cannot be rejected.

1.3 Model Comparison – sample likelihoods

Comparing the models using sample likelihoods is useful when two different model hypothesis both fall out positive which is the case here. For the storey 1 data, the sample likelihood of the normal model is 0.003, whereas that of the lognormal model is 0.008. As the value of the lognormal's is higher, it can be said that the postulation made is more probable than the postulate of the normal model. Similarly, for storey 2 data, the sample likelihood of the lognormal is 0.019, this is greater than that of the normal which was found to be 0.009. Therefore, it can be concluded that the postulation made in the lognormal distribution case is more probable than the normal model.

2. STRUCTURAL RELIABILITY

2.1 Reliability Index

The reliability index, β, can be found from both limit state functions using equations 3 and 4.

$$\beta_1 = \frac{1 - \frac{\mu_M}{M_U} - \frac{\mu_P}{P_U}}{\sqrt{\left(\frac{\sigma_M}{M_U}\right)^2 + \left(\frac{\sigma_P}{P_U}\right)^2 + 2 \cdot \frac{1}{M_U} \cdot \frac{1}{P_U} \cdot \rho_{P,M} \cdot \sigma_M \cdot \sigma_P}} \quad 3.$$

where, μ_M and μ_P are the mean and values of bending moment and axial load respectively, σ_M and σ_P are the standard deviations of bending moment and axial load (respectively). M_U and P_U is the bending moment capacity and axial load capacity respectively. The values parameters can be defined as $(\mu_P, \sigma_M, \mu_M, \sigma_M, M_U, P_U, \rho_{P,M})$ to $(1011.7, 50.16, 1843.43, 150.16, 3055.14, 4144.55)$.

$$\beta_2 = \frac{M_U \cdot P_U - P_U \cdot \mu_M - M_U \cdot \mu_P}{\sqrt{(P_U \cdot \sigma_M)^2 + (M_U \cdot \sigma_P)^2 + 2 \cdot P_U \cdot M_U \cdot \rho_{P,M} \cdot \sigma_M \cdot \sigma_P}} \quad 4.$$

Applying the above equations results in the same value reliability index of β_1 and β_2 which are both 4.34.

Here we are dealing with bending moment and applied load, which have different units, so before writing the limit state function, it is important to express the random variables in the same units.

Yes, the reliability index can also be calculated using a simulation approach, specifically, the Monte Carlo Simulation. Briefly, the process is as follows, a large number of realisations, *N*, of the basic random variables are simulated e.g. 1000. For every realisation, the limit state function is called. The number of realisations for which the limit state function is less than zero is then counted, this can be referred to as n_f. The probability of failure P_f, is then equal to n_f/N, finally, the reliability index can then be found from the value of probability of failure, i.e. $P_f = \Phi(-\beta)$.

2.2 Limit state function

Consider a mode of beam failure when the moment caused by loads is greater than the moment capacity, let *R* represent the resistance (moment capacity) and *Q* represent the demand (load). Hence, the limit state function can be defined for this failure mode as per equation 5.

$$g(R, Q) = R - Q \tag{5.}$$

The limit state, g can be treated as a boundary, if the value of g is less than 0 the structure (beam) is unsafe and if g is greater than 0, the structure is safe. No, the probability of failure will not change for definition of the limit state function, the probability of failure is based on the reliability index which after combining the random variables of both resistance and demand, will lead to the same value, regardless of the limit state definition.

2.3 Levels of reliability

There are four possible levels of reliability analysis.

- Level I: The uncertainty is modelled by a nominal value (e.g. code based).
- Level II: The uncertain parameters are modelled using mean/standard deviation values etc. Variables are assumed to be normal (e.g. the reliability index).
- Level III: An analysis considering the entire joint-distributions is done rather than just the main statistics.
- Level IV: Using a full probabilistic approach, all uncertainties are considered as well as failure consequences (e.g. casualties).

The current design codes are based on a Level I code philosophy. Practitioners approach is, first the scope is defined e.g. structural type – buildings. Next, a design space that has all the random variables is chosen. The existing design codes are then used to design the element, after which the limit states are defined. Then the statistical properties are determined and a method of reliability analysis is used and a target reliability value is defined, finally the partial safety factors are observed and selected.

2.4 One-story building

The individual reliability index of for the columns can be found from equation 6, below.

$$\beta = \frac{\mu_Z}{\sigma_Z} = \frac{\left(\mu_{M_U} - \mu_M\right)}{\sqrt{\left(\sigma_{M_U}\right)^2 + (\sigma_M)^2}} \tag{6.}$$

where, μ_z and σ_z is the mean and standard deviation value of the safety margin. The values of (μ_M, σ_M) and $\left(\mu_{M_U}, \sigma_{M_U}\right)$ are the (mean, standard deviation) of the bending moment and the bending moment capacity respectively. These were found by subbing in the given the given mean and standard deviation of the random variable *B*, into the given equation (e.g. for column 1 $\mu_{M_{U,1}} = 0.5 \cdot \mu_B$). Similarly, the bending moment values, were found by subbing in the mean value of the force *F* given previously into the available equations. Repeating this process for all three columns gives a reliability index of $4.98, 2.59$ and 3.61 for columns 1, 2 and 3 respectively.

The individual probability of failure of each column is first found as $P_{f_i} = \Phi(-\beta)$, giving values of $3.12 \cdot 10^{-7}, 0.0048$ and $1.53 \cdot 10^{-4}$ for columns 1, 2 and 3 respectively. Thereafter, when considered as a parallel system, the overall probability of failure can be given by equation 7.

$$P_f = \prod_{i=1}^{n} P_{f_i} \tag{7.}$$

where, n is the number of columns (3). Hence the overall system probability of failure is $2.28 \cdot 10^{-13}$. This is under the hypothesis that failure of all components (columns) is required for the system to fail.

The two extremes of probability of failure, in this case are brittle and ductile columns, if the column becomes completely ineffective after failure, it can be described as brittle. Whereas the ductile column can maintain its load-carrying capacity after failure. By improving the individual reliability of the columns, the probability of failure can be reduced. For example, using columns with a greater bending moment capacity will result in a greater reliability index and therefore a lower probability of failure.

3. ENGINEERING SEISMOLOGY

3.1 Catalogue completeness

For the site at latitude, longitude of 35.36, 25.12, Heraklion, Crete (Greece), two seismic zones were selected (zone 1 and 5, in the code) based on factors such as the number of earthquakes, site proximity and completeness. A sub catalogue was then created based on the zones and the main catalogue. Then, an investigation of the completeness of both zones was carried out, based on a range of moment magnitudes Mw of 1 to 6. For clarity, zone 5 from the figure will be referred to as zone 2 henceforth. For each of these magnitude values, the cumulative number of events was plotted against the observation period (time of first to last earthquake).

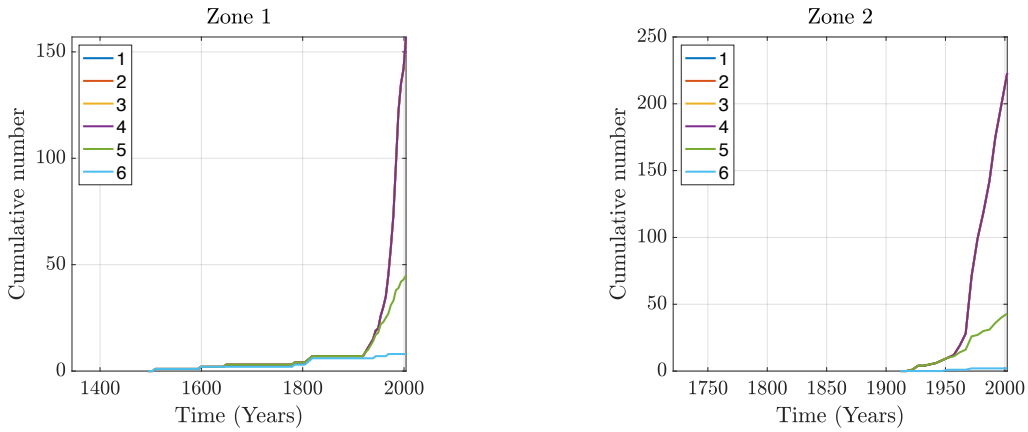

Figure 4: Cumulative number of events with respect to time

As seen in Figure 4, there is very little/no data before about until 1500 this is due to a lack of appropriate instruments, thereafter, for zone 1, the number of events recorded starts to increase slowly. Particularly, for Mw of 4 and 5, the sub catalogue shows completeness from roughly 1900 to 2000s. The sub catalogue for zone 1 is not complete due to the line remaining quite flat. Whereas for zone 2, there are no events recorded, till about 1910, for the Mw of 4, the catalogue shows completeness for the period of 1950-2000, as there is a linear increase from that time. Around the time period of 1910 to 2000, the Mw of 5 also shows completeness. The Mw of 6 remains basically flat throughout the time period of the sub catalogue, hence, it can be considered incomplete for this value of the moment magnitude.

3.2 Foreshock and Aftershock

The goal of removing foreshock and aftershock from the catalogue (declustering), is to remove any dependent earthquakes, which foreshock and aftershock are both types of. The main reason for this is, it needs to be ensured that the remaining earthquakes are independent events, according to the Poisson process. Therefore, this is important as the Poisson process is a common assumption within seismic hazard analysis. Moreover, it has been found that events that are aftershocks of the larger earthquakes term to dominate the earthquake original seismic catalog, hence the catalog may not be truly representative of what may occur in the future (Zaliapin and Ben-Zion, 2022). The removal of these inaccuracies will improve the model. Some of the declustering methods include windowing, this entails the removal of earthquakes in a time-space window around every main shock (Stiphout et al. 2012). Aftershocks cannot be distinguished by an outstanding feature in their waveforms, hence, to relate them to mainshocks, a measure of space time distance between the two is used (Stiphout et al. 2012). In more detail, for windowing, for each earthquake in the catalogue with a magnitude *M*, the following shocks are categorised as aftershocks if it occurs within a specified time interval, and distance interval (Stiphout et al. 2012). With the method of windowing, foreshocks are treated in the same way as aftershocks.

Studies have shown, declustering may have an effect on the PSHA, this may be unwanted. For example, studies have shown that the due to the fact that aftershocks can cause considerable damage, their removal of catalogues may lead to an underestimation of the seismic rates and hence the hazard (Taroni, 2021). Therefore, it should be noted that performing declustering does not mean that these dependent events are not essential for the hazard and risk.

3.3 Gutenberg Richter Relationship

The Gutenberg Richter (GR) relationship is essential for the PSHA, from the GR, one can find the annual exceedance rate of a given magnitude threshold. In this study, the moment magnitudes, Mw selected for analysis (the magnitudes the fitting is done for) range from a certain minimum magnitude to the maximum magnitude of the zone's sub catalogue. For Mmin, as this is in a European country, 4.5 was selected, this can be considered as the minimum magnitude that may cause damage to a structure on the site. Again, the moment magnitude Mw is always used for various reasons such as it is physical quantity related to the earthquake's energy, hence, saturation is avoided compared to other magnitude types.

The GR is based on an exponential distribution that can be given by: $\log_{10} N(M \geq m) = a - bm$. Where N, is the number of earthquakes greater than a particular magnitude threshold e.g. 4.5. Also, a and b are the model parameters. 10^a represents the activity rate (McGuire, 2004), or seismicity rate for zero magnitude earthquakes. The parameter b, is also known as the GR $b-value$, is close normally to unity, as this is also the slope of the GR plot. Performing the GR fitting for zones 1 and 2, the plots are shown in Figures 5(a) and 5(b), in addition, the plots also contain key information regarding the model parameters.

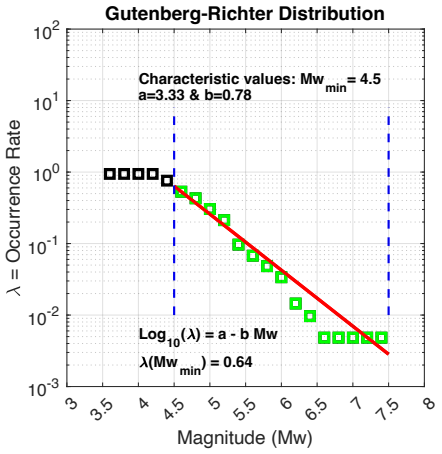

Figure 5(a): Zone 1

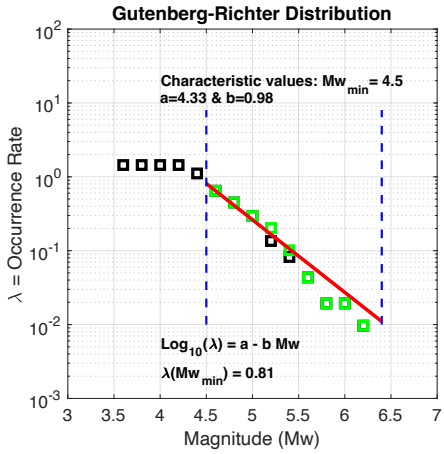

Figure 5(b): Zone 2

The annual exceedance rate of the minimum magnitude $\lambda(M_W \geq m_{w,min})$, can be read of the plots, for zone 1, this value is 0.64, whereas for zone 2, it is 0.81. This implies that zone 2 has more seismic activity than zone 1 (for the given minimum magnitude), as it has a higher earthquake occurrence. Furthermore, the values of the parameters a, b is 3.33, 0.78 for zone 1 and 4.33, 0.98 for zone 2.

Also, regarding completeness, this can be investigated again using the GR plot, when the points closely agree with the fit, it shows the catalogue completeness for those values of magnitude. As seen from Figures 5(a) and 5(b), for he case of zone 1, there is completeness from the magnitude of 4.5, up to roughly 6, after which the points start to deviate from the line. Similarly, for zone 2, Figure 5(b) indicates completeness for magnitudes ranging from 4.5 to just after 5.5, after which some of the points tend to deviate from the fitting line.

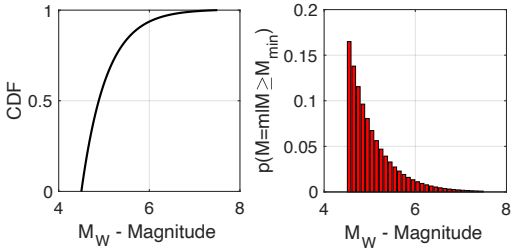

Figure 6(a): Zone 1

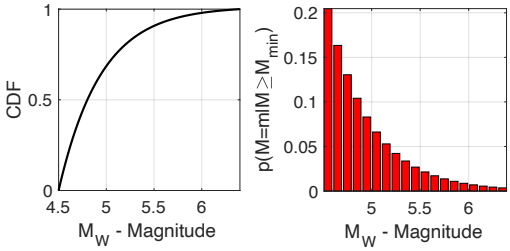

Figure 6(b): Zone 2

Moreover, the CDF of the earthquake magnitudes, given that the magnitude is in the specified range (4.5 to Mmax), can be computed using the b-value as well as Mmin and Mmax. This CDF can be further discretised in order to find the probability mass function (pmf) of the moment magnitude, the result can be seen in Figures 6(a) and 6(b). The GR also allows one to find the pmf of the magnitude, this is one of the key paramaters in the PSHA. As seen the GR aids in finding two of the parameters of the PSHA: the pmf and the annual exceedance rate.

3.4 Probability Seismic Hazard Assessment (PSHA)

To find the seismic hazard curve, the PSHA has to be carried out, this will take inputs such as the annual exceedance rate, Ground Motion Prediction equation (GMPE), and the pmf values for the magnitude and distance (from site to earthquake). For finding the values of the annual exceedance rate and magnitude pmf, see Section 3.3. The distance pmf can be found be through discretisation and this was done in MATLAB.

A GMPE model also needs to be implemented to compute the seismic hazard curve. The GMPE used for this study is given by (Akkar, 2010). Some of the inputs of this model had to be specified for the site, these include the V_{S30}, this is the average shear wave velocity of the top 30m of the site's soil. Using the USGS database (USGS, 2010), the site's V_{S30} can be found to be in the range of 300. In addition, in Heraklion, Crete, normal fault types are common (Caputo et. al 2010), hence, the fault type was selected as normal for both zones. Thereafter, the total seismic hazard curve (the two zones) of the Peak Ground Acceleration (time period of 0) was produced and is depicted in Figure 7(a). The y-axis of the plot is the exceedance rate of a specific intensity measure, and the x-axis is the spectral acceleration. From the hazard curve one can find the PGA for desired exceedance rates, e.g. it is shown the PGA value at 10% is equal to 0.05 g, 5% is 0.08 g, and 2% is 0.17g.

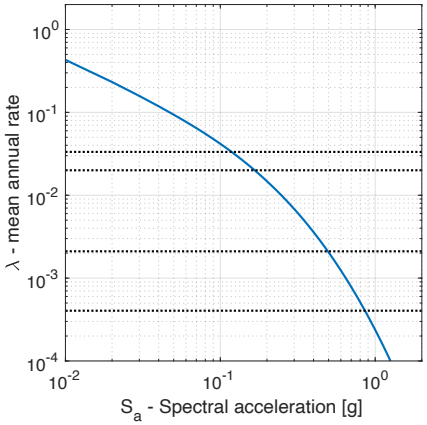

Figure 7(a): Seismic hazard curve for the PGA

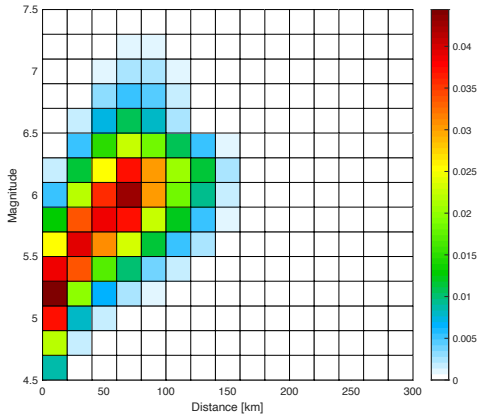

Figure 7(b): Seismic disaggregation

For the PSHA method, the classical numerical integration was chosen over simulation. The Monte Carlo based simulation does have several advantages over the classical. For example, when the intensity measure predictions have a natural form of realisations rather than probability distributions (classic), it is more convenient to use it with physics-based ground motion model (McGuire, 2004). Simulation also allows for a greater flexibility in the model, particularly with the intensity measures (Mcguire, 2004). However, some of the disadvantages of simulation is that a large number of them are required in order to produce an adequate number of the rare events that are of interest (McGuire, 2004). This is not only very time consuming but may entail great computational cost and a significant of data is produced (takes up computer memory). Hence, the classical method was used for this study, the catalogue for the seismic zones has also been provided, therefore in contrast to the Monte Carlo, a stochastic catalogue does not have to be created.

One of the key products of the PSHA is disaggregation. This term essentially describes the earthquake scenario most likely to cause $IM > im$. Disaggregation was carried out for the spectral acceleration, $S_a(3.0) = 0.01g$. The meaning of this essentially the conditional probability of an earthquake scenario (magnitude, distance), given an intensity measure exceeds a particular value (defined above) at the site. The disaggregation of this case is portrayed in Figure 7(b), the scenarios with a large probability of exceedance of the intensity measure have a dark red colour, whereas those with a lower probability have a light blue/white colour. From this figure, it can be seen that all of the hazards occur with sources that are less than 160km away from the site. Furthermore, the scenario (M, R) with the largest contributions are from $5 < M < 5.1$ at distances of $0 < R < 20\ km$, there is also another large contribution of earthquakes from $5.9 < M < 6.1$ at distances of $60 < M < 80\ km$.

3.5 Uniform Hazard Spectrum (UHS)

The uniform hazard spectrum is a useful way of aggregating the results of the PSHA calculations for a given site. The process of finding the UHS for this site is as follows, the seismic hazard curves of the site are shown in Figure 8(a), from this the spectral acceleration values for a specific exceedance rate are identified at each time period. These spectral acceleration values can then be plotted against their periods, leading to a response spectrum. Due to the spectrum ordinates all have the same mean annual rate, this plot is referred to as the uniform hazard spectrum and this is depicted in Figure 8(b) (normal axis, log-log scale).

11 vibration periods in the range of $T = 0$ to $T = 3$, were considered. The annual exceedance rate, λ, can be related to a return period T_R, by $\lambda = 1/T_R$. The values of the return period for this study as shown in Figure 8(b) are 30, 50, 475, 2475 years, these correspond to the respective exceedance rates that are indicated by the black horizontal lines in Figure 8(a). From this figure, it can be seen that the hazard (spectral acceleration) gradually increases as the return period increases. Moreover, although higher return periods generally correspond to a lower probability (exceedance), the figure shows that larger return period correspond to higher ground motions. For all the different return periods, the UHS is steep for shorter vibration periods, in this case less than $T = 0.2\ s$. All the lines for different return periods, reach a peak at the same time period, which is at $T = 0.2\ s$. After this peak, all the lines converge to reach a steady value.

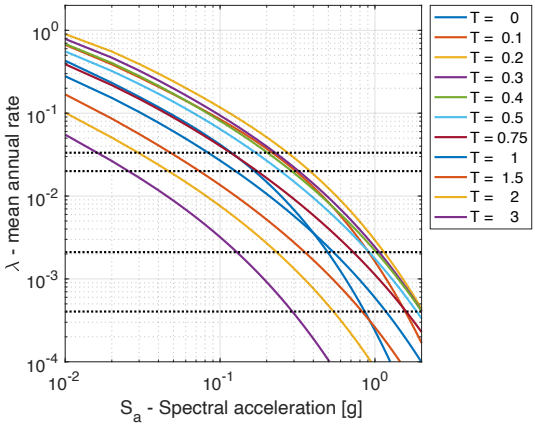

Figure 8(a): Seismic hazard curves for multiple vibration periods

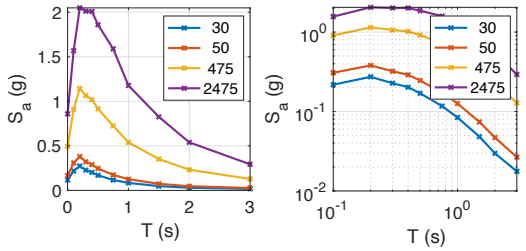

Figure 8(b): Uniform Hazard Spectrum for different return periods on a regular and log axis.

One of the common aims of the PSHA is to identify a response spectrum that can be used for structural analysis. Specifically, the UHS is also an extremely useful tool in performance-based design, through the UHS, designers may create a design spectrum that provide a safety level of a structure for different time periods.

REFERENCES

Andrzej S. Nowak and Kevin R. Collins (2013) *Reliability of Structures*. Boca Raton: CRC Press. Available at: https://search-ebscohost-com.bris.idm.oclc.org/login.aspx?direct=true&db=nlebk&AN=1763413&site=ehost-live (Accessed: 6 December 2023).

Akkar, S. and Bommer, J.J., 2010. Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East. *Seismological Research Letters*, *81*(2), pp.195-206

Benjamin, J.R. and Cornell, C.A., 2014. *Probability, statistics, and decision for civil engineers*. Courier Corporation

Faber, MH 2012, Statistics and Probability Theory : In Pursuit of Engineering Decision Support, Springer Netherlands, Dordrecht. Available from: ProQuest Ebook Central. [6 December 2023].

Caputo, R., Catalano, S., Monaco, C., Romagnoli, G., Tortorici, G. and Tortorici, L., 2010. Active faulting on the island of Crete (Greece). *Geophysical Journal International*, *183*(1), pp.111-126.

MATLAB, 2010. *version 7.10.0 (R2010a)*, Natick, Massachusetts: The MathWorks Inc

McGuire, R. K. (2004). Seismic hazard and risk analysis. Earthquake Engineering Research Institute.

Sturges, H.A., 1926. The choice of a class interval. *Journal of the american statistical association*, *21*(153), pp.65-66

Taroni, M. and Akinci, A., 2021. Good practices in PSHA: declustering, b-value estimation, foreshocks and aftershocks inclusion; a case study in Italy. *Geophysical Journal International*, *224*(2), pp.1174-1187.

U.S. Geological Survey, 2020, Earthquake Lists, Maps, and Statistics, accessed March 18, 2020 at URL https://www.usgs.gov/natural-hazards/earthquake-hazards/lists-maps-and-statistics

van Stiphout, T., J. Zhuang, and D. Marsan (2012), Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa52382934. Available at http://www.corssa.org.

Zaliapin, I. and Ben-Zion, Y., 2022. Perspectives on clustering and declustering of earthquakes. *Seismological Research Letters*, *93*(1), pp.386-401.