



Risk Analysis of Earthquake Occurrences in Turkey

**Student: Peter Hedley
Supervisor: Thomas Kjeldsen**

**Department of Architecture and Civil Engineering
The University of Bath**

Abstract

Earthquake hazard and risk analysis can inform governmental decisions and mitigate against the catastrophic effects of an event. There are a number of methods to describe earthquakes occurrences, this study aims to compare and contrast these methodologies. This report explores the results of this analysis, concluding that the Frechet distribution is inaccurate for sparse data samples but finds little to distinguish methods of the Gamma, GEV, Log-Normal and Weibull distributions. In addition, a new method of defining hazard areas is explored, dissipating from fault-line sources instead of hazard calculation for defined parcels. In comparison to other maps, the new methodology proposed distinguishes more discrete areas of hazard and may lead to improved governmental resource allocation, potentially saving lives. The results are highly dependent on a good-data catalogue and findings indicate further resources are needed in the production of improved datasets. The hazard map is highly dependent on the topographic attenuation coefficient, it is concluded that further research into these regional coefficients is required. A risk analysis of Turkey determined Istanbul as the area of highest risk, it is concluded that aid and mitigation schemes should focus here to provide the correct post-disaster aid. This new method of stochastic earthquake analysis is proven to provide reliable results, comparable to current maps, but with improved area definition. It will show increased potential in future years as data-catalogues become more expansive.

Acknowledgements

I would like to express my thanks to Dr Thomas Kjeldsen for his kind guidance and knowledge. His detailed explanations of statistical methods and encouraging advice has made this dissertation possible. I would also like to thank Stanford University for their online Statistical Learning Course by Trevor Hastie and Rob Tibshirani and the insights that it provided.

Contents

1. Introduction	
1.1. Objectives	6
1.2. Deliverables	6
2. Literature Review	
2.1. Deterministic vs. Probabilistic Analysis of Earthquake Risk	7
2.2. Statistical Distributions of Earthquake Occurrences	7
2.3. Accounting for Missing Data	8
2.4. Goodness-of-Fit Tests	8
2.5. The Kolmogorov-Smirnov Test	9
2.6. The Anderson-Darling Test	10
2.7. Introduction to Earthquake Energy	10
2.8. Earthquake Attenuation	11
2.9. Methods of Region Specification	12
2.10. Hazard Maps	12
2.11. Risk - Consequences	13
3. Initial Investigation into Earthquake Data	
3.1. Fault Line Activity	15
3.2. Investigation into Earthquake Magnitude	15
4. Statistical Analysis of Earthquake Data	
4.1. Introduction to AMAX Series	17
4.2. Distribution Selection - Cullen and Frey	17
4.3. Estimating Model Parameters	18
4.4. Testing Distributions for Goodness-of-Fit	19
4.5. Parameter Estimation Standard-Error	20
4.6. Introduction to Return Periods	21
4.7. Individual Fault Line Stochastic Analysis	22
4.8. Stochastic vs. Empirical Mw	23
4.9. Earthquake Dissipation	24
5. Hazard Map	
5.1. Fault Line Dissipation	25
5.2. Combining Dissipation Matrices	26
5.3. Plotting the Hazard Map	27
5.4. Comparison of Maps	28
5.5. Sensitivity Analysis - Dissipation Coefficient	29
5.6. Sensitivity Analysis - Distribution Selection	30

6. Risk Map

6.1. Population Density.....	30
6.2. Risk Map Calculation.....	31
6.3. Plotting the Risk Map.....	32

7. Conclusions

7.1. Investigation of Earthquake Analysis Techniques.....	33
7.2. Comparison of Hazard Map to European SHARE Map.....	33
7.3. Hazard Map Implications.....	34
7.4. Risk Map Implications.....	34
7.5. Sensitivity and Areas for Further Investigation.....	34

8. References**9. Appendix A****10. Appendix B****Table List**

Table-1 Earthquake Economic Effect.....	6
Table-2 Literature Distributions.....	7
Table-3 Earthquake Magnitude.....	16
Table-4 Distribution Goodness-of-Fit Tests.....	19
Table-5 Distribution Parameters.....	20
Table-6 Combined Fault AMAX Fit Test.....	20
Table-7 Log-Normal Parameter Estimation.....	21
Table-8 Example Censored AMAX.....	22
Table-9 Gamma Parameter Estimation.....	23
Table-10 Mw Comparison.....	26
Table-11 Mercalli and PGA Comparison.....	28

Figure List

Figure-1 Gumbel-I vs. Gumbel -III (Roca, 1984)	7
Figure-2 Censored Distributions	8
Figure-3 K-Fold Cross Validation	8
Figure-4 Inverse Probability Transform (Distribution Pdf and Cdf, 2016)	9
Figure-5 P and S Earthquake Waves (Bolt, 1999)	10
Figure-6 Rayleigh and Love Earthquake Waves (Bolt, 1999)	11
Figure-7 Earthquake Spherical Dissipation	11
Figure-8 Hazard Definition (Wang and Wu, 2016)	12
Figure-9 European Earthquake Hazard Map, (SHARE, 2013)	13
Figure-10 Plot of Turkey Earthquake Data and Faults	14
Figure-11 Plot of Turkey Fault Line Activity	15
Figure-12 Plot of Turkey Earthquake Mw Data	16
Figure-13 Earthquake Magnitude Histogram	16
Figure-14 EDF and CDF of Empirical AMAX Mw	17
Figure-15 Bootstrap Sampling	17
Figure-16 Cullen and Frey Graph Distribution Selection	18
Figure-17 Frechet and Log-Normal Comparison	19
Figure-18 Combined AMAX all Faults Mw>4	20
Figure-19 Log-Normal and Combined Fault AMAX	21
Figure-20 Bootstrap Distribution Parameters	21
Figure-21 Distribution Return Period Estimation	21
Figure-22 Fault Mw 1000-Year Return Period	22
Figure-23 Censored Return Periods	23
Figure-24 Predicted Mw 500-Year Return Period	23
Figure-25 Maximum Recorded Fault Mw	24
Figure-26 Matrix Explanation	25
Figure-27 Matrix Dissipation	25
Figure-28 Matrix Dissipation	26
Figure-29 Offset Fault Line	26
Figure-30 Complete Dissipation Process	26
Figure-31 Earthquake Hazard Map	27
Figure-32 European Earthquake Hazard Map, (SHARE, 2013)	27
Figure-33 Dissipation Sensitivity Analysis	29
Figure-34 Distribution Sensitivity Analysis	30
Figure-35 Turkey Population Map	31
Figure-36 Earthquake Risk Map	32
Figure-37 Risk Sensitivity Analysis	32

1. Introduction

Earthquakes have long been a source of human fatalities and have a devastating impact on local infrastructure. It is estimated that globally, between the years of 1980-2008, c. \$350 billion USD of economic damage was caused by earthquakes and an estimated 385,630 people lost their lives in the same period (statisticbrain, 2016). This dissertation seeks to improve upon current understanding of these events using a new approach to produce informative and potentially life-saving data.

This document aims to investigate the risk of earthquakes to local residents and infrastructure across Turkey. A database of c. 10,000 earthquake events since 1900 provided by the SHARE European initiative (SHARE, 2013) shall be utilised in order to conduct the study. A statistical extreme value analysis of this data using R-programming software could provide an increased understanding of earthquake occurrences and better inform government decisions and prevention schemes. The estimated cost of an earthquake event in Turkey is estimated below in Table-1 (Gurenko, 2006).

Table-1 | Earthquake Economic Effect

Annual Exceedence Probability	Probable Maximum Economic Loss (\$ millions)	GDP (%)
0.5	11,406.0	6.20
5.0	3,476.0	1.90
20.0	24.5	0.01

1.1. Objectives

This dissertation shall analyse earthquake occurrences in Turkey using the following process.

- Investigate the current earthquake analysis techniques and compare.
- Create and compare an earthquake hazard map for Turkey with the European hazard map.
- Discuss implications of the hazard map on Turkey.
- Create and discuss implications of a risk map for Turkey, compare to hazard maps.
- Investigate sensitivity of earthquake risk and hazard maps to technique assumptions.
- Highlight areas in need of further investigation.

1.2. Deliverables

- Evaluation of existing hazard and risk perception methods for Earthquake events.
- Creation of novel approach to hazard calculation for the region of Turkey.
- Plot and compare new hazard map with European earthquake hazard map.
- Creation of a risk map for Turkey and comparison to hazard map.

2. Literature Review

2.1. Deterministic vs. Probabilistic Analysis of Earthquake Risk

Deterministic analysis involves an ontic assumption of earthquake occurrence. This means that the physical properties and variables of a fault system can be used to gain an understanding of earthquake occurrence. In Reid's paper of elastic rebound theory (Reid, 1911) fault pressure build up and then release was first explored. However, due to the complexity of earthquake occurrence, it was soon modelled with better accuracy using a probabilistic distribution (Jeffreys, 1938).

A probabilistic (stochastic) analysis is favoured by the statistical geophysics community with papers by (Wang and Wu, 2013; Ordaza and Arroyo, 2015; Greenhough and Main, 2008) concluding that stochastic analysis provides an improved understanding of earthquake occurrences. Stochastic analysis of earthquakes shall be further considered in this report.

The two main variables of interest when formulating a probabilistic model is to consider the return period and event magnitude. The magnitude of such an event can be quantified in different ways. Most well-known is the Richter scale, developed by Charles Richter (Richter, 1935), this has however, some flaws, for example, it is only valid for certain frequencies and distances. The proposed method of magnitude that shall be used in this paper is Mw, a moment magnitude measurement which has been shown to have better accuracy determining the size of larger earthquakes (EMPR, 2014).

2.2. Statistical Distributions of Earthquake Occurrences

The production of a stochastic model to describe earthquake occurrence is achieved through the use of probability distributions. Various stochastic distributions have been proposed in the literature and these are tabulated in Table-2.

An important factor in distribution selection is the correct description of the larger Mw earthquake events. The Gumbel distribution, for example is less able in this respect 'Gumbel I, however, increasingly deviates from the data for high-intensity values' (Roca. et al., 1983), Figure-1.

Overestimation by a magnitude of one on the Mw scale, as it is logarithmic, is a factor of 10 from the real predicted event.

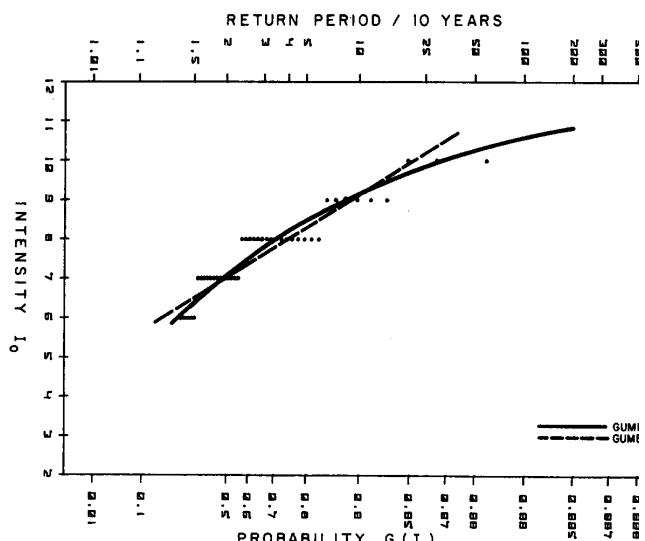


Figure-1 | Gumbel-I vs. Gumbel -III (Roca, 1984)

Table-2 | Literature Distributions

Weibull	Gamma	Log-normal	GEV	Frechet
Parvez and Ram, 1997	Yadav et al., 2010	Wang and Wu, 2014	Pasari and Dikshit, 2014	Baddari, 2016
Yadav et al., 2010				Kagan and Schoenberg, 2001
Pasari and Dikshit, 2013				SSHAC, 1997

2.3. Accounting for Missing Data

Roca (Roca, 1984) altered the time period for the extreme value series in order to produce series with fewer missing data-points by having a 10-year AMAX series. This could obviously conversely be applied on a regional basis until a sufficient quantity of data is within that region.

An approach by (Tsapanos et al., 2016) uses the assumption that a distribution can be estimated with little loss of accuracy if there is less than 25% of the data is missing - (Burton, 1979). The approach then disregards any regions with less than 10 recorded values due to a lack of data. This also seems impractical and sparse data sets would disregard large numbers of faults.

In the SHARE earthquake catalogue, the threshold of study is 4Mw and the data is left-censored past this point see Figure-2. Any data that has not been recorded must lie in this 0-4Mw range. A distribution can be fit by modifying the maximum likelihood estimation method which estimates distribution parameters based on empirical data, Equation-1.

$$L = \prod_{i=1}^k f(x_i) \prod_{i=1}^{N-k} P\{X \leq 4\} \quad (1)$$

This process of distribution fitting is statistically proven and more accurate than fitting the models with modified data. Hence this process shall be used as it is more statistically correct and will provide more accurate results.

2.4. Goodness-of-Fit Tests

To determine if a distribution is suitable for describing the data, a goodness-of-fit test must be performed. It is important to have both a method of training and testing a distribution. The two methods proposed in this paper are cross-validation and Stephen's approximation (Stephens, 1974).

Cross-validation (Mosteller and Tukey, 1968) can be performed in a number of ways, for example, leave-one-out (LOOCV), and k-fold. The principle behind these tests is to split the data randomly into a training set and a test set. This process is repeated and the mean error of fit is calculated. This process can be seen below using k-fold cross-validation, Figure-3.

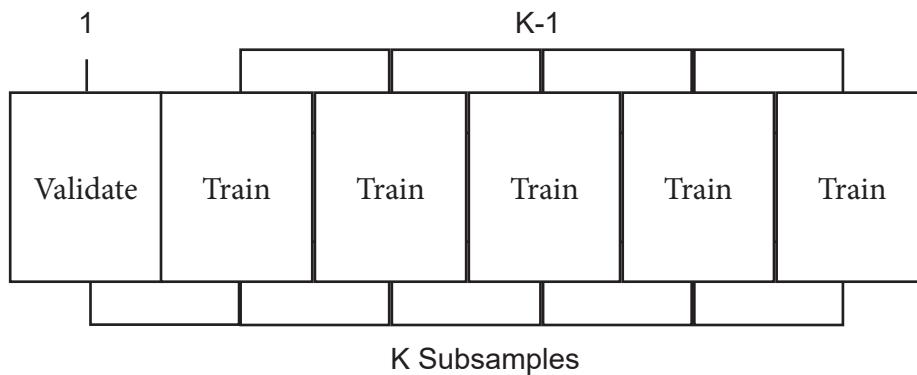


Figure-3 | K-Fold Cross Validation

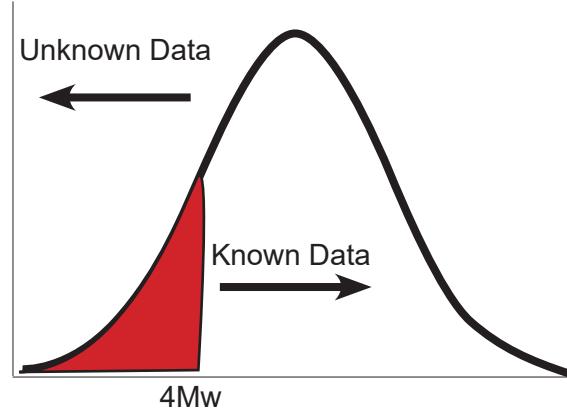


Figure-2 | Censored Distributions

Since each training data set is only $(k-1)/k$ as big as the original training set, the estimates of prediction error will typically be biased, $k=5$ or 10 provides a good bias-variance trade-off (Hastie et al., 2013).

Stephen's approximation (Stephens, 1974) is an alternative method to cross-validation which uses the inverse probability transform in an innovative way, Equation-2. If X is a continuous random variable (RV) with CDF - $F_X(x)$ and if $Y = F_X(X)$, then Y is a uniform RV on the interval $[0,1]$.

$$\begin{aligned} F_Y(y) &= \Pr [Y \leq y] \\ &= \Pr [X \leq F_X^{-1}(y)] \\ &= F_X [F_X^{-1}(y)] = y \end{aligned} \quad (2)$$

Using this method any continuous monotonic function, in this case, the random empirical distribution of X , can be converted into a uniform distribution via the use of the CDF (cumulative distribution function). This can be tested for goodness-of-fit whilst not being biased due to the pre-fit distribution parameters of scale and shape etc. (Stephens, 1974). This process is described below in Figure-4.

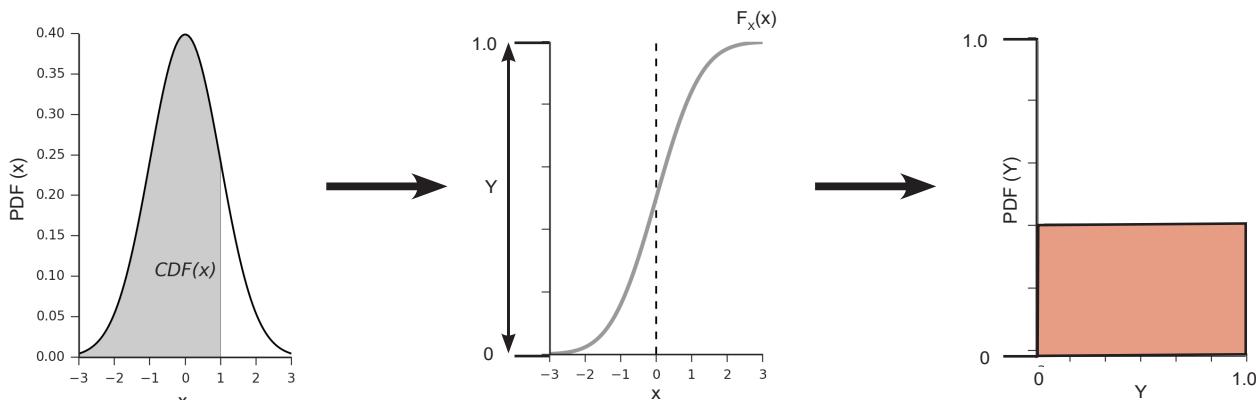


Figure-4 | Inverse Probability Transform (Distribution Pdf and Cdf, 2016)

Since the CDF of X is between 0 and 1 the function Y must also be located between 0 and 1. Also as the CDF relates to the probability of $P(X < x)$ then the following must be true Equation-3.

$$P(Y < 0.7) - P(Y < 0.6) = P(Y < 0.5) - P(Y < 0.4) \quad (3)$$

This will hold for all values on the CDF. Therefore any continuous distribution can be converted via the Probability Integral Transform to a uniform distribution. This distribution can then, in turn, be tested for goodness of fit with a number of tests that are described in the next section of this document (Kolmogorov-Smirnov, Anderson-Darling, and Cramer-Von Mises).

2.5. The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) (Kolmogorov, 1933) test is the first of the parametric tests of fit that shall be performed on the distributions. This test is mainly concerned with the shape of the fit, however other parameters such as dispersion affect this value. As an advantage of this test compared to an exact test (such as chi-squared) is that a smaller sample size is needed for an accurate value. In the specific case of this data, there are not enough observations to provide a reasonable chi-squared estimate of fit (itl.nist.gov, 2013).

The test is a representation of the maximum distance between the two sets of data, the empirical and stochastic. The KS test measures this distance across the cumulative distribution function (CDF) of the stochastic and the empirical distribution function (EDF) of the empirical data, Equation-4.

$$\text{Kolmogorov-Smirnov} = \sup |F_n(x) - F(x)| \quad (4)$$

Due to the curvature of a CDF this test is less sensitive to the tail ends of the data and since an AMAX analysis is to be performed, the Anderson-Darling test is preferable. ‘The Kolmogorov-Smirnov test is only a historical curiosity’ (Stephens and D’Agostino, 1974).

2.6. The Anderson-Darling Test

This test is designed to account for the tail ends of a cumulative distribution function, seen in Figure-3. At the extremes it can be seen that the function gradient decreases, this with respect to the Kolmogorov-Smirnov test causes uncertainty with respect to the test’s validity.

$$\text{Cramer-von Mises} = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dx \quad (5)$$

The Anderson-Darling test is similar to the Cramer-Von Mises test seen in Equation-5, which uses the distance between the hypothesised and empirical function as a measure of fit. The Anderson-Darling test modifies this to better account for the tail ends of the data, Equation-6.

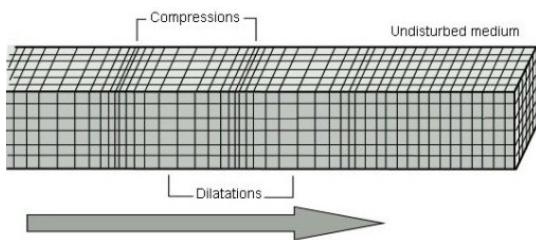
$$\text{Anderson-Darling} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1-F(x))} dx \quad (6)$$

The tails are better represented due to the weighting $F(x)(1-F(x))$ as this is largest then $F(x)=0.5$ and decreases towards the tail ends. Hence a larger weighting is applied to areas of the EDF that were not previously in both the Kolmogorov-Smirnov and Cramer-Von Mises tests (Hastie et al., 2013).

2.7. Introduction to Earthquake Energy

The energy from an earthquake event travels through the earth in different ways, surface and body waves. S and P waves (Lay and Wallace, 1995) are body waves. The S (shear) waves travel as transverse waves whereas P (primary) waves are longitudinal in nature, Figure-5.

Longitudinal P-Wave



Transverse S-Wave

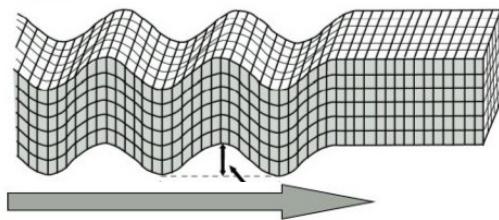


Figure-5 | P and S Earthquake Waves (Bolt, 1999)

These are generally deep waves that are able to travel below the upper mantle, sometimes refracted by the earth’s core. The body waves are accompanied by the surface Love and Rayleigh waves that cause the majority of the earthquake damage, Figure-6 (Rayleigh, 1887; Love, 1911).

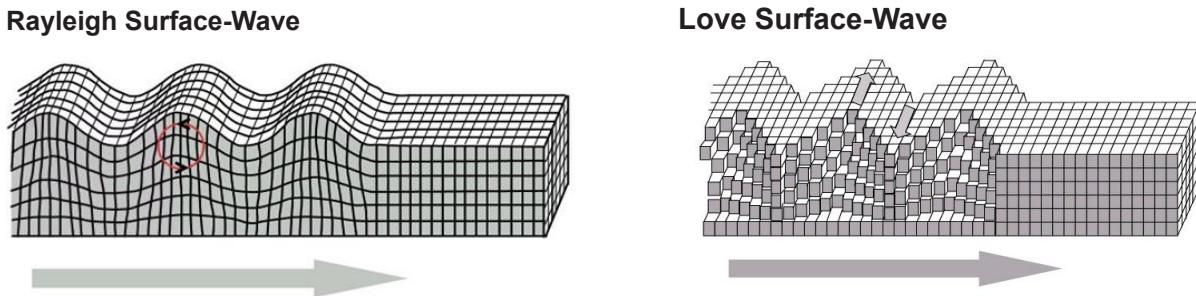


Figure-6 | Rayleigh and Love Earthquake Waves (Bolt, 1999)

The Rayleigh waves travel in a purely vertical motion, they are an accumulation of the P and S body waves. The movement caused by this wave creates an oscillating motion at the earth's surface. Love waves, on the other hand, are produced by the horizontal component of the S waves. These provide a transverse movement in the horizontal plane as can be seen in Figure-6.

2.8. Earthquake Attenuation

An earthquake event can be attributed to a source (focal point) below the ground where the energy is dissipated from a release of pressure. The point above ground at this location is known as the epicentre. Assuming that the event occurs from a point source the energy can be assumed to dissipate in a circular fashion, Figure-7. The closer to the focus the larger the energy, with the energy dissipating in a circular manner. However, this model assumes that the earth's surface is perfectly elastic and the waves are able to propagate with little resistance. This is not the case and energy is lost to friction and other forces (Shearer, 1999) called intrinsic attenuation.

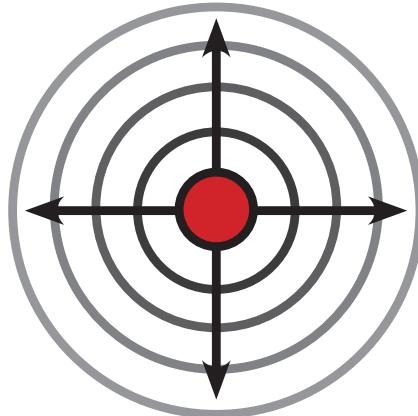


Figure-7 | Earthquake Spherical Dissipation

A purely circular dissipation is $1/r$ from the source assuming elastic conditions. As discussed this is not the case with real earthquakes. The actual dissipation is in relation to the frequency of the individual waves with higher frequencies attenuating differently (Herriaz and Espinosa, 1987). This can, however, be approximated using a decay rate of $c. 1/R^{1.1} - 1/R^{1.4}$ (Edwards et al., 2010) depending on the size of the radius and geography of the region. This, however, for the course of the document shall be approximated to a decay rate of $1/R^{1.2}$, Equation-7.

$$E_1 \times \frac{1}{R^{2(1.2)}} = E_2 \quad \longrightarrow \quad R = \left(\frac{E_1}{E_2} \right)^{\frac{1}{2(1.2)}} \quad (7)$$

Where E_1 is energy at the epicentre, E_2 is the energy at distance R from the epicentre.

This approximation assumes an event energy dissipation in all directions, however specific geographical factors will influence this. For example, attenuation across a mountainous region is higher than that in a flatter geographic setting. This, however, shall not be included in this document and the assumption of an even energy dissipation will be adopted.

2.9. Methods of Region Specification

The most simplistic approach (Greenhough and Main, 2010; Yadav et al., 2010) classifies the region of interest, in this case Turkey, and determines a stochastic distribution that encompasses the whole zone. All areas within that region are therefore assigned an equal earthquake hazard. This approach works well in small areas but is not suitable for the larger region of Turkey, as a probability map is a deliverable of this report it would not be acceptable as the sole method. However, an analysis of the whole region shall be conducted and results explored.

Another approach (Roca, 1984; Theodoros et al., 2016) is to split respective regions of earthquake activity into discrete subsections of a generally predetermined area. Both papers suggested an area dictated by $1^\circ \times 1^\circ$ longitudinal and latitudinal parcels. The probability of an earthquake occurrence and the magnitude of that event is then determined for each parcel. These are c.110km by 110km sections and in an active country such as Turkey there is enough data to merit a similar approach.

It is proposed however to modify a method in Wang's paper (Wang and Wu, 2016), Risk Assessment of Active Faults in Taiwan. This determines the population affected by creating a 50km border for each respective fault and summing the population within that area, Figure-8. This method could equally be applied to hazard perception. It is known that earthquakes are generally created by fault lines (Reid, 1911) and that this energy distributes approximately evenly from the source (Edwards et al., 2011). Therefore approximating that the points along a fault are equally probable to slip, this method can be used to calculate the hazard in the local area.

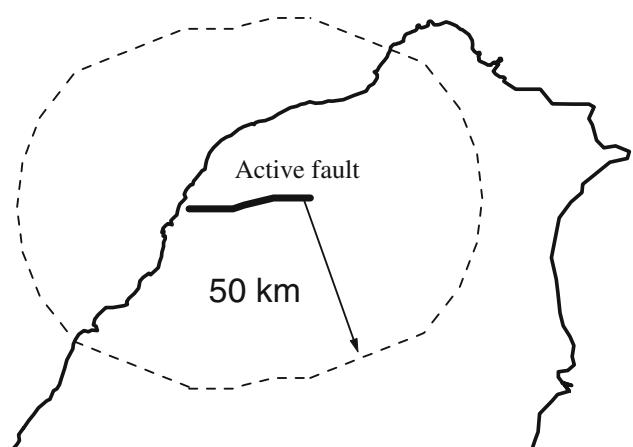


Figure-8 | Hazard Definition (Wang and Wu, 2016)

In addition to the aforementioned, the magnitude of an earthquake must be taken into account and the dissipation of that energy. It is therefore also more accurate to alter the 50km boundary depending on the earthquake magnitude. Hence the hazard can be calculated for all faults in Turkey and dissipated correctly to form a hazard map. This approach can then be compared to the results from literature and the European SHARE hazard map (SHARE, 2013).

2.10. Hazard Maps

It is crucial to pick the correct stochastic model that best describes the earthquake data. The variations in hazard obtained by the incorrect model can cost billions of dollars in planning and mitigation or worse the misuse of government resources leading to loss-of-life. This report shall, therefore, analyse the proposed stochastic models in literature and compare each distribution choice.

This model can then be used to obtain a hazard map of Turkey which only determines the probability of an earthquake event. Mapping the earthquake probability across Turkey aids in building code generation, government spending, and earthquake mitigation (EMPR, 2014). An example hazard map by the European SHARE initiative is shown in Figure-9. If the probability of an earthquake occurrence is higher then greater care is required to mitigate against it. For example, soil liquefaction and increased loading on buildings must be taken into account during the design phase. This extra loading that must be mitigated against can only be obtained from a stochastic earthquake analysis.

2.11. Risk - Consequences

Risk is obtained from both the probability of an event and the consequences of that event. Therefore an event with a low possibility of occurrence and with high consequences contains a higher risk than the inverse. There are many ways of quantifying risk, the European Seismic Hazard map (SHARE, 2013), Figure-9, solely displays a probability map and does not quantify the risk.

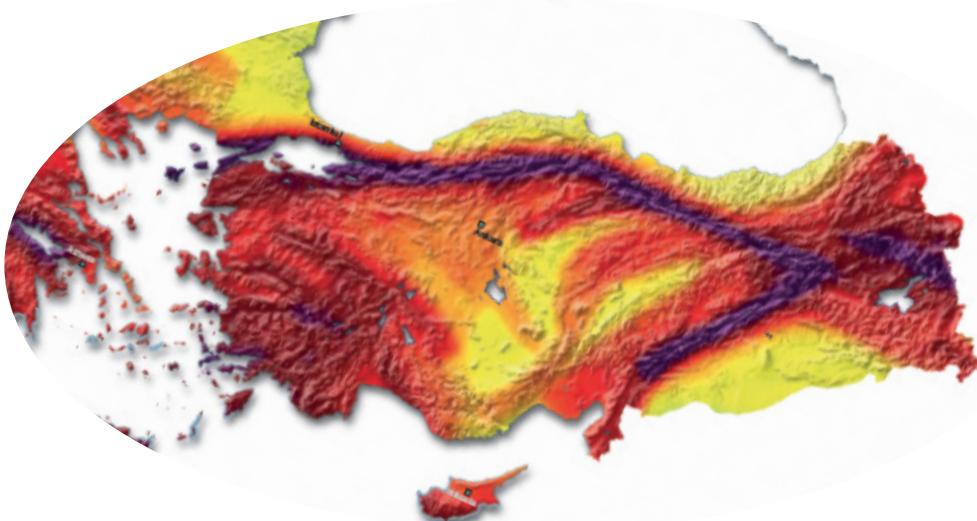


Figure-9 | European Earthquake Hazard Map, (SHARE, 2013)

An insurance company, for example, would be interested in the potential economic damage caused by an earthquake occurrence. This is achieved by estimating the damage in US dollars caused by certain magnitude earthquakes. Combining this with a hazard map enables an informed judgement of what should be charged as an insurance premium to customers in order to mitigate against the risk. However, although an interesting sphere, the monetary damage from an earthquake is of little importance in comparison to lives lost and other factors.

It seems logical to adopt a policy that concentrates entirely on saving human lives. This approach has been used by (Wang and Wu, 2016) to assess the risk factor of a fault. The theory can be expanded to a greater area and the creation of a map for Turkey. This shall be achieved by multiplying the predicted magnitude by the population density in that area. As logically the risk will increase linearly with the number of people affected by the earthquake, Equation-8.

$$\text{Earthquake Risk} = \text{Predicted Magnitude} \times \text{Population Density} \quad (8)$$

Note that the predicted magnitude is on a logarithmic scale so the equation will take this into account using 10^{Mw} . For example, on a risk map, a city with a high probability of earthquake occurrence has a higher risk than a rural location with the same probability of an event. This is because urban regions will have a higher risk with respect to loss of life.

This map shall be compared with the hazard map produced and any results discussed. A population risk map is important for government planning and event mitigation. For example, strategic locations of hospitals and aid is vital for post-event relief. If the aid is located in a danger zone it may not be accessible, so safe placement in close proximity to an event is vital.

3. Initial Investigation into Earthquake Data

The data that shall be used in this study has been obtained from a database compiled by the European SHARE initiative. The data comprises of c. 10,000 earthquake events since the year 1900 ranging from (4.0 - 7.7) Mw on the moment magnitude scale with a mean of 4.1Mw. The data covers the entirety of Europe and can be seen plotted below, Figure-10.

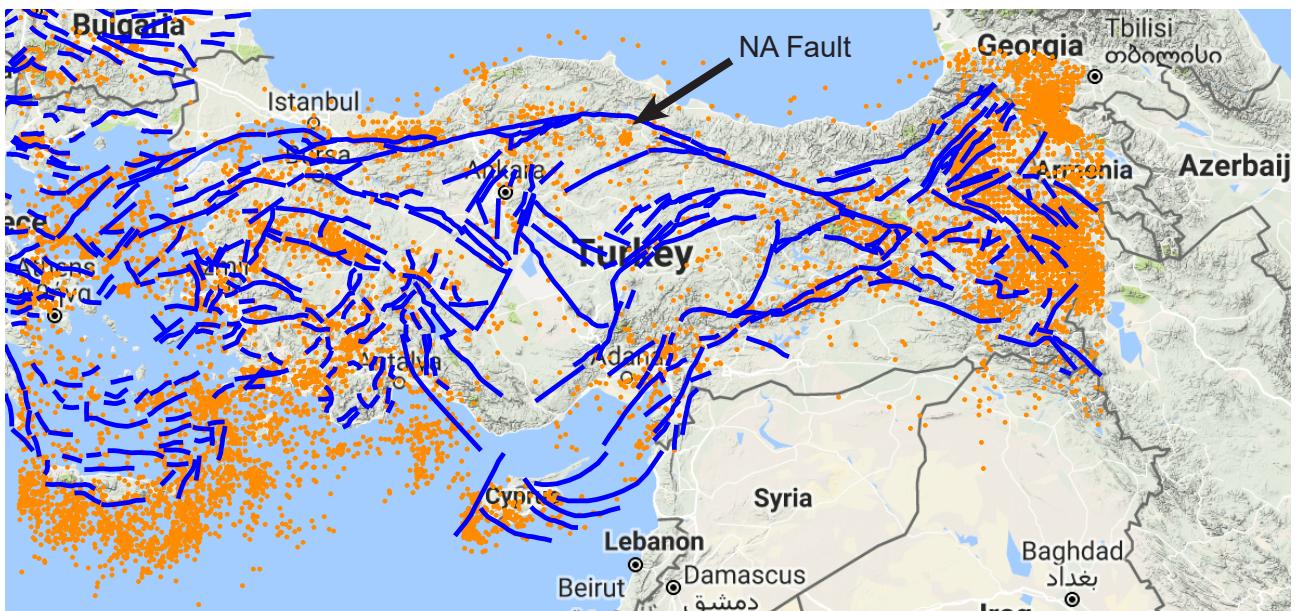


Figure-10 | Plot of Turkey Earthquake Data and Faults

The data in Figure-10 is located between c. 23° - 48° longitude and 33° - 45° latitude and earthquake events can be seen to cluster in specific areas. The more active areas are located to the extreme East and West of Turkey with the majority of events in the East.

Earthquake events are caused when two plates or sides of a fault suddenly move and the pressure build up is released. The energy release produces seismic waves which cause the ground to shake; the movement is the predominant cause of building collapse and loss of life. The focus of an earthquake is the exact point at which the faults slide and the epicentre is the point above it at ground level, the events are recorded by the nearest seismological-sensor and plotted in Figure-10.

Faults are either individual or composite seismographic sources as defined (Basili et al., 2009). Fault sources are structures that comprise of identified fault lines, yet represent a complex system with a number of aligned individual seismographic sources that cannot be separated spatially (EFEHR, 2013). The number of fault lines and the activity of those faults in a specific area will have a significant bearing on the number of earthquake occurrences. The European Facility for Earthquake Hazard and Risk (EFEHR) provides access to a known fault database for Europe. The earthquakes occur in a higher density near the crowded areas of fault lines, Figure-10 (blue).

For the purpose of this investigation, it is assumed that the events occur due to the activity of EFEHR provided fault lines. This is generally the case however some of the lesser magnitude events can be produced by fault source systems not included in the map. The faults display a similar distribution to the earthquake events, predominately located toward the East and West of Turkey with few in the centre. One large fault to note is the North Anatolian Fault. This fault is an active slip-strike fault that traverses the transform boundary of the Eurasian and Anatolian Plates. In 1999 the North Anatolian Fault produced a 7.6-moment magnitude earthquake which injured 43,953 people and caused 17,118 fatalities which emphasises the importance of earthquake mitigation schemes.

With the prior assumption, each of the 10,000 events can be assigned to a specific fault line. Each event can be attributed to a fault by a nearest neighbour approach; determining the closest fault by geographical location to the recorded earthquake event. Once the closest fault is determined, the activity of that fault is tallied and its activity increases by one per event.

$$a^2 + b^2 = c^2 \quad (9)$$

Equation 9 is used to calculate distance from an event to the nearest fault line system. Where:
 a is the difference in longitude between the closest fault and earthquake event
 b is the difference in latitude between the closest fault and earthquake event
 c is the distance in degrees between the closest fault and earthquake event

3.1. Fault Line Activity

The largest activity fault was responsible for more than 900 events since the year 1900 and the least active 0. Every fault is then colour coded in order to gain a graphical appreciation of the activity of various Turkey fault lines, Figure-11.

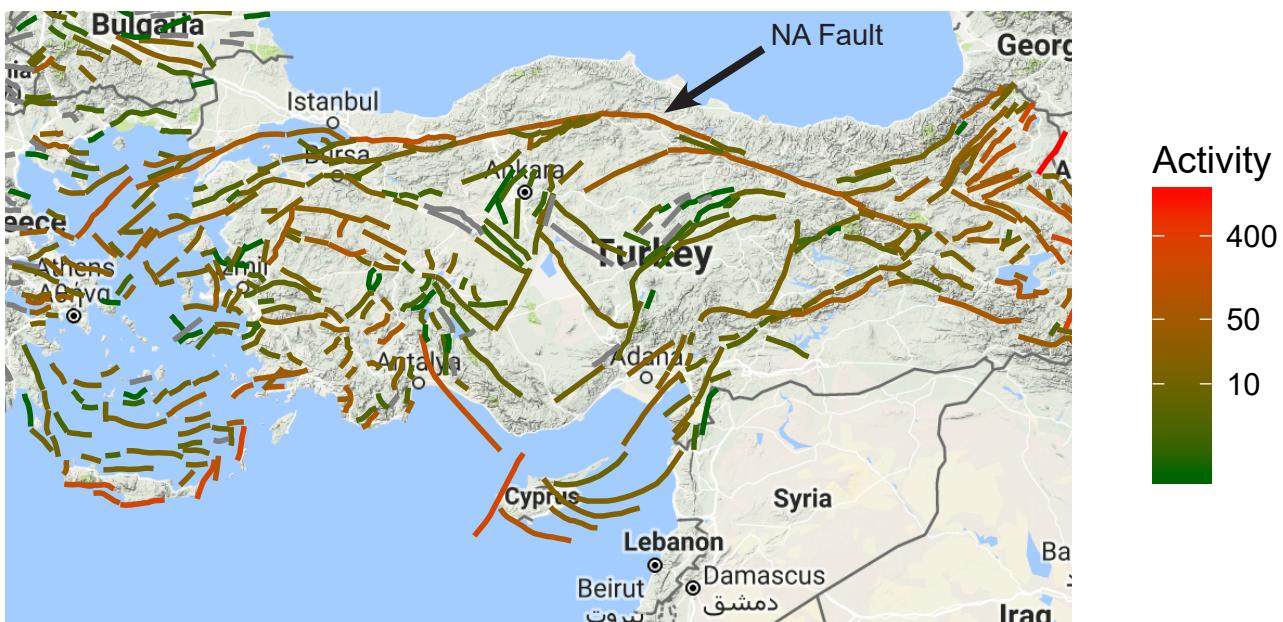


Figure-11 | Plot of Turkey Fault Line Activity

As expected the North Anatolian fault is highlighted in red as one of the more active fault lines. It can be noted from Figure-11 that the Turkish cities are generally located close to zones of high earthquake activity. This is a sad case of poor city planning prior to an understanding of earthquake occurrences which is synonymous with the number of fatalities that Turkey experiences from them. It is a topic that shall be investigated later in this document when the specific area of risk is analysed.

3.2. Investigation into Earthquake Magnitude

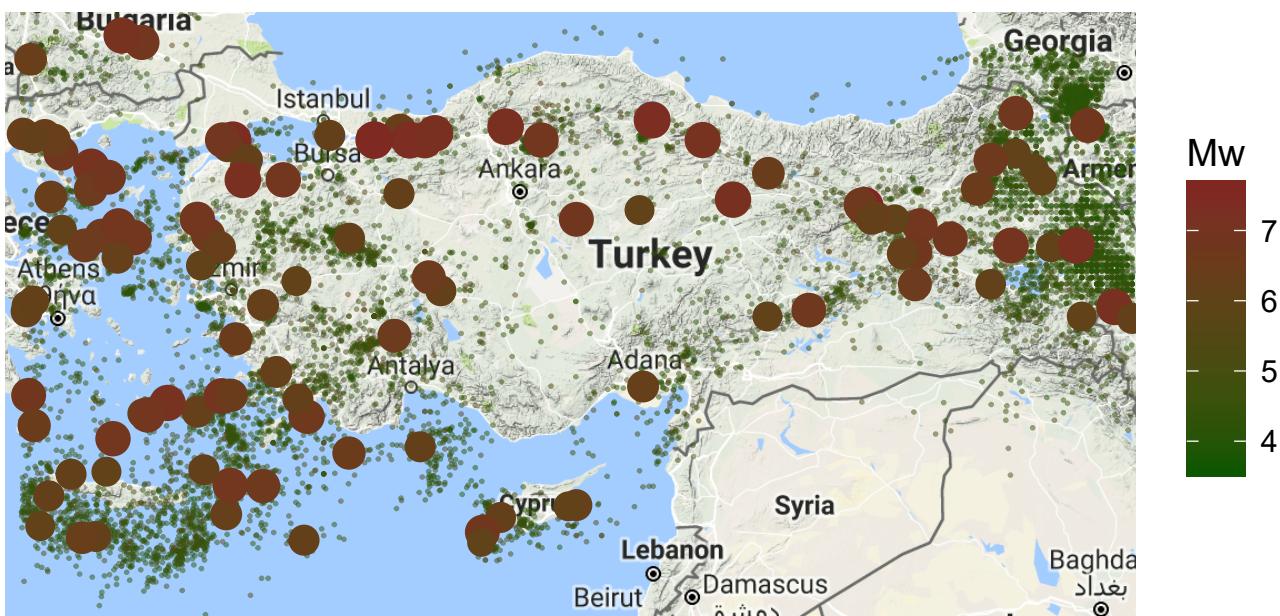
The activity of a fault, however informative can only describe the number of earthquake occurrences in an area. This may be completely removed from the actual damage that is experienced. To better consider this aspect a quantity relating to the magnitude of the earthquake events must be introduced.

The magnitude of an earthquake characterises the relative size of the earthquake. This, in turn, represents the magnitude of ground movements created by it. It is therefore important to understand both the geographical location and the magnitude of the events. A number of smaller earthquakes can be considerably less dangerous than a significant larger event, Table-3.

Table-3 | Earthquake Magnitude

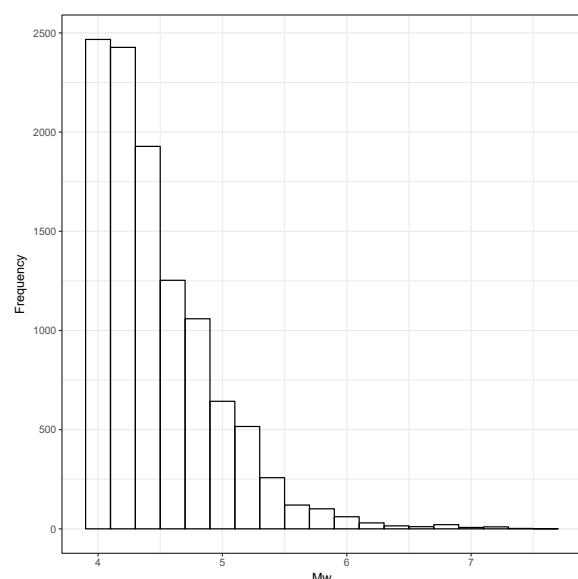
Magnitude	Earthquake Events	No. per annum
2.5 or less	Usually not felt	900,000
2.5 to 5.4	Only causes minor damage	30,000
5.5 to 6.0	Slight damage to buildings and other structures	500
6.1 to 6.9	May cause a lot of damage in very populated areas.	100
7.0 to 7.9	Major earthquake. Serious damage.	20
8.0 or greater	Great earthquake. Can totally destroy communities near the epicentre.	One every 5 to 10 years

From Table-3 (usgs.gov, 2015) as magnitude increases the damage caused also increases at an exponential rate. Also as the magnitude of an event increases it is less likely to occur and the frequency of such an event decreases. As discussed in the literature review this dissertation shall use the Moment Magnitude (Mw) scale to quantify earthquake size. Mw is more applicable to earthquakes especially the larger magnitude events (usgs.gov, 2015) than other popular scales.

**Figure-12 | Plot of Turkey Earthquake Mw Data**

Earthquake data obtained from SHARE contains both the location and magnitude of the events. The data is considered censored since it has only been collected above a threshold of 4Mw. The events are plotted with the Mw of every event contained within the colour gradient, Figure-12. Each individual area will have greater or lesser magnitude events within it depending on the proximity of highly active faults.

A histogram of Mw is plotted in Figure-13. The mode of the data is c.4Mw, although since the data is censored this is not representative, the lower the Mw the greater likelihood of an event occurrence. So the Mw<4 would have a higher occurrence frequency if they had been recorded. However, an exponential decay is still visible. With the likelihood of higher events being exponentially less than lower Mw occurrences.

**Figure-13 | Earthquake Magnitude Histogram**

4. Statistical Analysis of Earthquake Data

4.1. Introduction to AMAX Series

An Annual Maximum (AMAX) series contains the largest observed magnitude earthquake in a given calendar year. It is assumed that the value obtained for each year is independent of any values obtained in other years. The AMAX series for Turkey is plotted in Figure-14. Note that the mean of this dataset is higher than that of Figure-13; 6Mw and 4.5Mw respectively. This is because Figure-14 is an AMAX series and only the maximum recorded yearly value is included.

In this all-inclusive case of Turkey the AMAX data is available for all years (1900-2006) and so a complete AMAX series can be plotted. This, however, is not the case for an individual fault line's AMAX series where many of the years are below the 4Mw threshold and hence the exact Mw is unknown.

The AMAX series for Turkey, plotted in Figure-14 contains both the EDF and CDF of the data. This seems to follow an unusual pattern of a double peak distribution a subset of the mixed distribution family. This family generally involves the combination of a number of separate distributions that are superimposed on one another. A cautious approach is required to the data, as the less-complex models are often better, so a less complex model will be reviewed in this paper. The mixed nature of the data will need further investigation in future literature.

4.2. Distribution Selection - Cullen and Frey

The AMAX data in Figure-14 can be approximated through the use of a probability distribution. This statistically describes the data, it can be used to predict the probability of a given event and extrapolate beyond the available empirical observations. Selection of the correct statistical model to describe the empirical data shall be considered through the use of a Cullen and Frey graph which suggests a family of distributions from the kurtosis and skewness of the data.

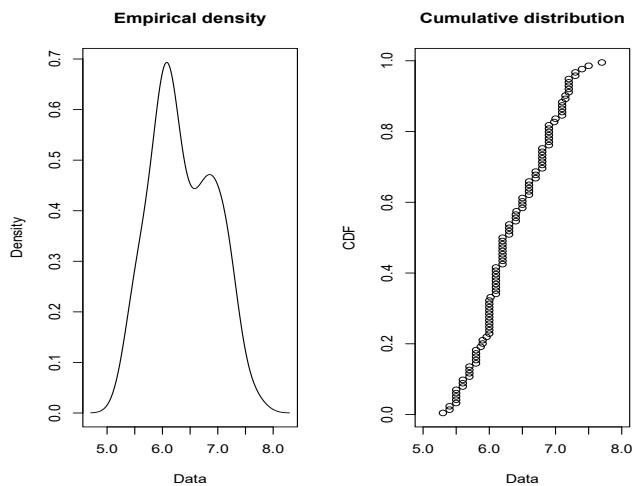
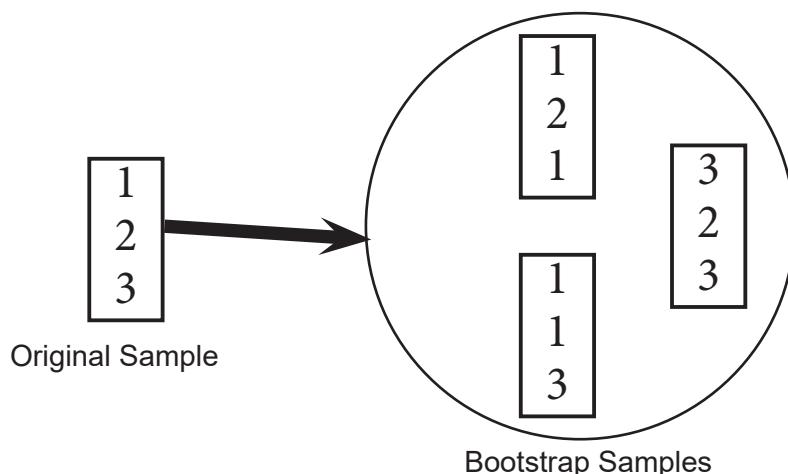


Figure-14 | EDF and CDF of Empirical AMAX Mw

As the Cullen and Frey graph is very sensitive to small sample sizes the Bootstrap procedure is used to produce extra samples from the data. Bootstrap is a statistical sampling method with replacement. This produces a range of datasets that are also plotted on the Cullen and Frey graph to determine an appropriate stochastic distribution family. The Bootstrap process is described in Figure-15.

The Turkey AMAX data Cullen and Frey graph is plotted in Figure-16 using the fitdistr R package and bootstrap sampling (Hastie et. al, 2013). The data displays a low level of skew and an equally low kurtosis validating that an exponential distribution is invalid for this dataset. This seems counter-intuitive, however, the series is an AMAX and clustered to the higher values of Mw which validates discarding the exponential distribution.

The cluster of bootstrapped points is close to the Log-Normal and Gamma lines, which are distributions suggested in the literature. Confirming that these are the correct ones to explore further. Another suggestion of the Cullen and Frey graph is a Normal distribution. This, however, will have less flexibility in fit, since there are a number of datasets, one for each fault it shall not be explored further.

4.3. Estimating Model Parameters

A number of different distributions can be used to fit the data. These shall be reviewed and the distribution that best describes the data can then be used for predictions. Each distribution requires input parameters to alter the kurtosis, skewness, and location. The parameters are estimated using Maximum likelihood Estimation (MLE), which maximises the likelihood function, Equation-10.

$$L(X_1, \dots, X_n) = \prod_{i=1}^n f(X_i; \lambda_1, \dots, \lambda_k) \quad (10)$$

Maximum Likelihood Estimation maximises the likelihood function in order to estimate the parameters. It is simpler to minimise the negative log-likelihood function instead as logs are additive, Equation-11.

$$\ln L(\theta|x) = \ln \left(\prod_{i=1}^n f(x_i; \theta) \right) = \sum_{i=1}^n \ln f(x_i; \theta) \quad (11)$$

For the Turkey AMAX series, this can be used to determine the optimal parameters. However, for the individual fault lines, there are a number of missing terms where the earthquake AMAX event for that specific year is below the threshold value of 4Mw. This is a censored sample and the maximum likelihood function must be altered to account for this, Equation-12.

$$L = \prod_{i=1}^k f(x_i) \prod_{i=1}^{N-k} P\{X \leq 4\} \quad (12)$$

Where:

$$L = \prod_{i=1}^k f(x_i) \text{ Relates to uncensored data and } \prod_{i=1}^{N-k} P\{X \leq 4\} \text{ the censored data}$$

4.4. Testing Distributions for Goodness-of-Fit

The distributions discussed in the literature are deemed suitable by the Cullen and Frey analysis, each can now be fitted to the data using the MLE method and then tested for goodness-of-fit. The three tests used, as described in the literature review section of this document are the Kolmogorov-Smirnov (KS), Cramer-Von Mises (CVM) and Anderson-Darling (AD) test. The results of each test on the fitted distributions for the AMAX data of Turkey is displayed below in Table-4.

Table-4 | Distribution Goodness-of-Fit Tests

	Weibull	Gamma	Log-Normal	GEV	Frechet
KS - Test	0.15	0.11	0.11	0.11	0.09
CVM - Test	0.32	0.19	0.18	0.18	0.13
AD - Test	1.8	1.1	1.0	1.0	0.98

The lower the test values the better the fit is as the tests measure the distance between the fitted and empirical distributions. Therefore this experiment has distinguished the Weibull distribution as the worst-fit and the Gamma, Log-Normal, GEV, and Frechet distributions as favourable distributions. The Frechet and Log-Normal distributions are plotted in Figure-17.

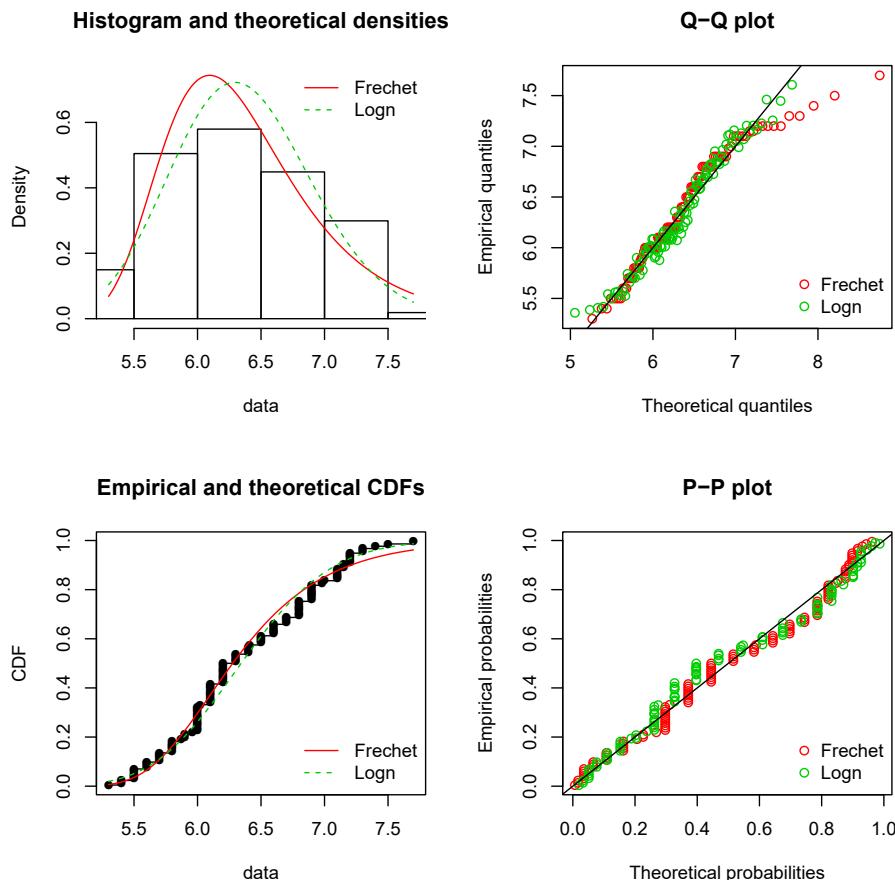


Figure-17 | Frechet and Log-Normal Comparison

The Frechet distribution, although in the testing procedure fitted well, deviates from the data at the upper tail. The Log-Normal has a worse fit in the lower Mw regions but describes the data better in the upper Mw regions. This is preferable in the case of earthquakes as Mw is a log scale and so greater emphasis should be put on the higher Mw events. In order to distinguish between the distributions with respect to goodness-of-fit tests a larger sample of data is needed.

Therefore an AMAX series was determined for each individual fault line then combined into one series containing the AMAX of the 400 faults. This is a censored sample with a number of parameters below the 4Mw threshold. In order to visualise this data in the histogram plot all data in the censored area has been removed, allowing an increased view of the 4Mw + region, Figure-18. The entire censored sample is displayed as a histogram in Figure-19. Note that there is still a mixed distribution aspect of the data, this may be an area of investigation for a future report, however, it is out of the scope of this document.

The distributions can be assessed for variance. Taking into account the number of parameters in each model, as the higher the number of parameters the larger the variance and likelihood of over-fitting the data, Table-5.

Table-5 | Distribution Parameters

	Weibull	Gamma	Log-Normal	GEV	Frechet
No. Parameters	2	2	2	3	3

As there are three parameters for both the GEV and Frechet distributions the variance is likely to be larger than that of the Log-normal, Weibull and Gamma distributions. Therefore the preference shall be to choose a distribution with the smallest number of parameters if all else is equal. The candidates are tested in Table-6 using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) tests for censored data against the combined individual fault AMAX data.

Table-6 | Combined Fault AMAX Fit Test

	Weibull	Gamma	Log-Normal	GEV	Frechet
AIC Test	33780	33870	33950	33850	33980
BIC Test	33800	33880	33970	33870	34000

The AIC and BIC tests provide an interesting comparison for this dataset compared to Table-4. The Frechet distribution that previously performed the best for the Turkey AMAX series is the worst fit on the combined individual fault line AMAX series (containing AMAX data for each of the 400 faults in Turkey). The Weibull distribution, previously the poorest fit is now the preferable distribution. This is representative of literature which is unable to specify a clearly superior distribution.

4.5. Parameter Estimation Standard-Error

The Log-Normal distribution is estimated and plotted against the combined fault line AMAX data in Figure-19. The unknown values in the sample ($Mw < 4$) is a large proportion of the data. This provides uncertainty in fitting the distribution. The missing data $P(X < 4)$ is approximated using a uniform distribution for visual purposes only. The data below 4Mw is unknown and thus the distribution of it equally unknown. The uncertainty can be quantified in the standard error of the parameter estimation by the MLE method. This can be achieved through a number of ways, one example is that the bootstrap process described in Figure-15 can be utilised to determine the standard error.

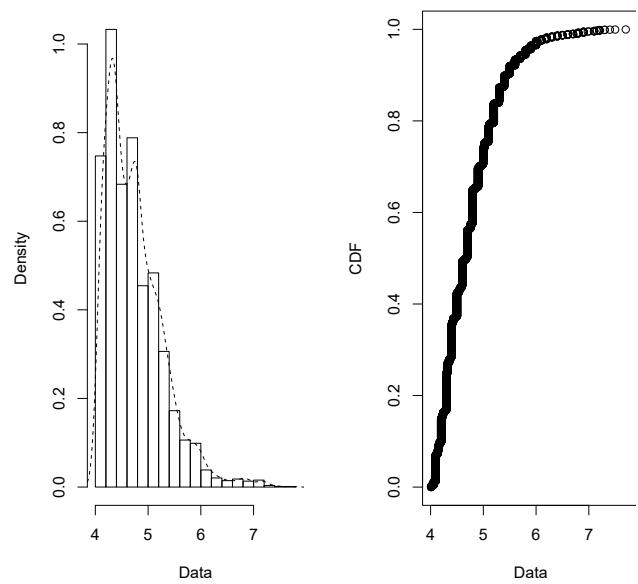


Figure-18 | Combined AMAX all Faults $Mw > 4$

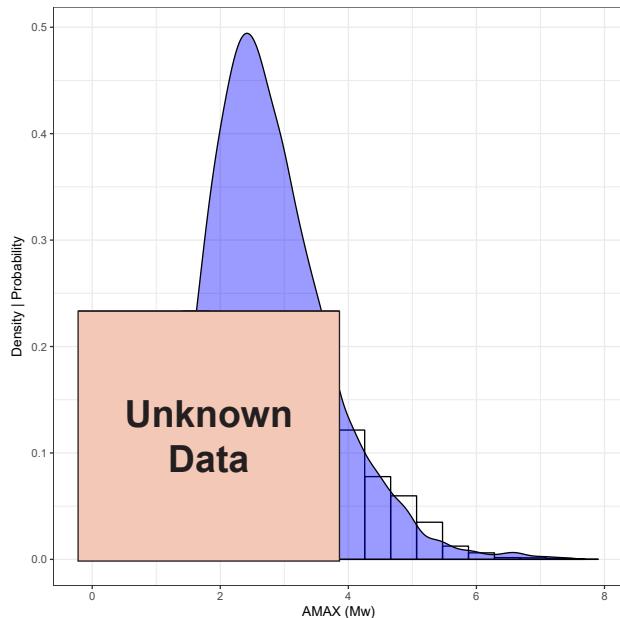


Figure-19 | Log-Normal and Combined Fault AMAX

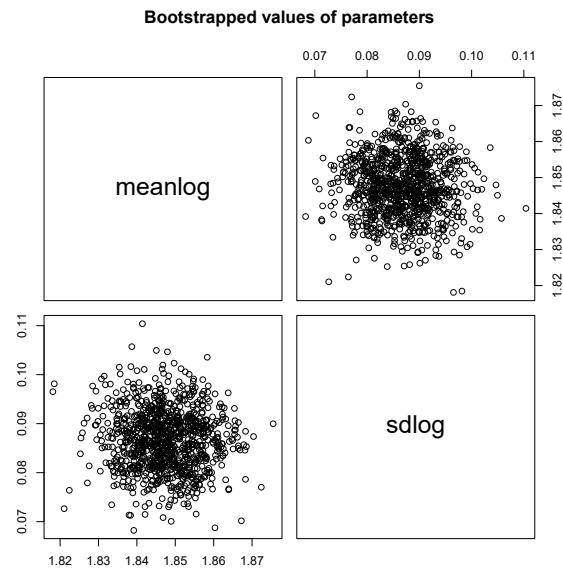


Figure-20 | Bootstrap Distribution Parameters

Using a set of 1001 separate bootstrap samples for the Log-Normal parameters Figure-20. The parameter standard error is calculated, Table-7. All distributions showed a low standard error averaging around 1% for each parameter, so there is little to distinguish between them.

Table-7 | Log-Normal Parameter Estimation

Parameter	Estimate	Std. Error
Mean Log	0.97	0.0066
Sd Log	0.32	0.0043

4.6. Introduction to Return Periods

A return period is the average time between two earthquake events that exceed a set Mw threshold and the inverse of the probability that one such event will occur in a given calendar year, Equation-13.

$$AEP = \frac{1}{T} \quad (13)$$

Where **AEP** is the annual exceedance probability and **T** the return period of an earthquake event. The return period of an event is a purely probabilistic estimate and two events may occur in a shorter time period than their return period predicts. To compare distributions the predicted magnitudes for given return periods are plotted for the non-censored country-wide AMAX data, Figure-21.



Figure-21 | Distribution Return Period Estimation

The maximum recorded earthquake in history was a 9.5 Mw earthquake in Chile (1960). The physical geography of a fault inhibits the maximum energy that it can release in an earthquake event, so a further test can be related directly to Turkey whose largest event was an 8.0 Mw in 1668. This is an approximated value from rock stress analysis but the largest event of the 1900-2006 sample is 7.7Mw so the predicted 10,000-year event should be close and likely higher than this magnitude.

The Frechet distribution here is clearly an incorrect choice and completely over-exaggerates the 10,000-year event at 10.6Mw. However, all other distributions seem a good approximation of the larger return period events, ranging from 8 - 8.8 Mw. As the empirical data extends back fewer than 500 years and is incomplete there is little evidence contrary to these approximations or to distinguish the most accurate. The majority of distributions, however, appear to approximate the empirical data well and so can be used to ascertain the earthquake hazard in Turkey.

4.7. Individual Fault Line Stochastic Analysis

Each fault line has been assigned to a subset of earthquake events via a nearest neighbour approach. Therefore every fault line can be described by a distribution. Each fault will be independent and vary according to the events that are accounted for by that particular fault.

The faults can then be assigned a Mw value for a set return period of say 1000 years. As each fault will contain a different activity and magnitude a histogram can be plotted of Mw for every fault with a 1000 year return period, Figure-22. As expected from an active earthquake region the mean Mw is 7. With few smaller values in the region of 4Mw and some very high magnitude faults which provide the largest hazard to local communities, as mentioned previously a large hazard may not correlate to significant risk.

With comparison to Figure-21, the values of Mw for individual fault lines are far higher - with one value at 12Mw. This is due lack of data and poor model fit which causes a high error in the larger return periods. One fault has an AMAX event series with a singular non-censored event, Table-8.

As mentioned previously in this report the 0-4 signifies the 'censored' data, in this case the majority of values are censored with only one recorded value at 7.1 Mw. The standard deviation for this particular fault's AMAX series is therefore high. This creates a poor prediction ability at the larger return period values, Figure-23.

The return periods for all distributions are incorrect. The Frechet distribution is the poorest estimator on this dataset with a predicted 10,000-year event of 50Mw. However, most distributions fail in this respect with GEV at 2.6Mw and the rest clustered around a 20-35Mw estimation. This is not entirely a distribution error but an incorrect estimation of the parameters by the MLE method. The average standard error for the distribution parameters is 70% of the parameter's value. With the Gamma distribution containing a parameter with over 200% standard error, Table-9.

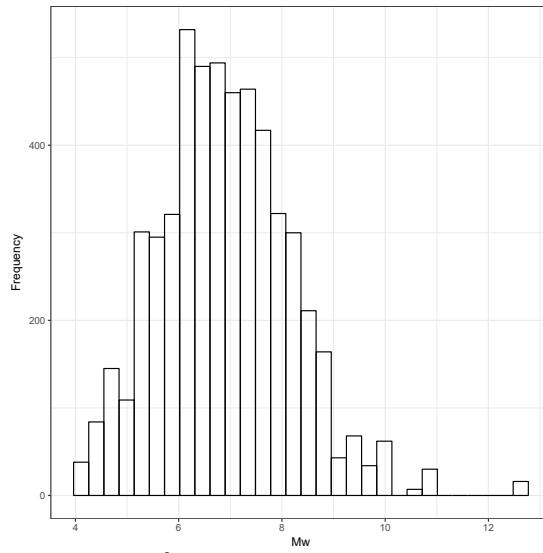


Figure-22 | Fault Mw 1000-Year Return Period

Table-8 | Example Censored AMAX

AMAX Data Value (Mw)	No. of Occurrences
0-4	106
7.1	1

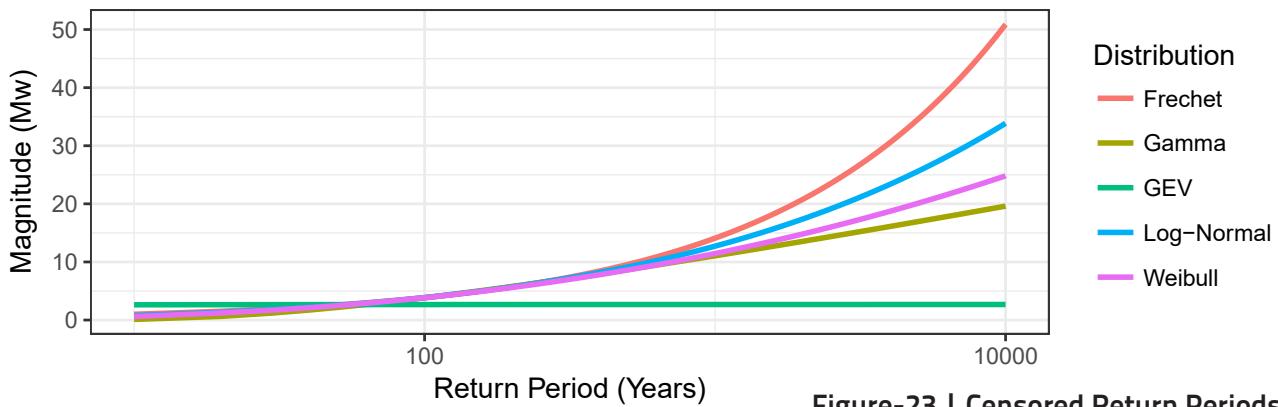


Figure-23 | Censored Return Periods

The standard error, Table-9, explains the incorrect return period predictions which are larger than the maximum return period for the entire Turkey AMAX series prediction of c.8.8 Mw. Seismographic data-catalogues are becoming more advanced and more data is available to study earthquake occurrences. Since this method of assigning data to specific faults requires a comprehensive dataset to reduce error the technique will become more prevalent in future years. At the time of writing, there are a few faults without a high enough quality dataset for this analysis to model earthquakes with lower uncertainty.

Table-9 | Gamma Parameter Estimation

Parameter	Estimate	Std. Error
Shape	0.03	0.07
Rate	0.22	0.3

4.8. Stochastic vs. Empirical Mw

The distribution that shall be used for further tests is the Gamma distribution, this distribution has marginally superior statistical test results, however, even this distribution is inaccurate at the higher Mw events with highly censored data. The data shall, therefore, be modified after prediction by a distribution to be a maximum value of 8.8Mw which is the same as the predicted 10,000-year event for the entire Turkey dataset. These modified values for each fault can be plotted in a similar manner to the activity plot in Figure-11 with a return period of 500 years, Figure-24.

This map is far more conclusive as to the faults that provide the greatest hazard than the activity map. This is because a fault with a high activity but low magnitude events is far less hazardous than a less active fault with larger magnitude occurrences. The map demonstrates the danger posed by the Anatolian fault, a very high magnitude prediction for the 500-year event. This model produced by a stochastic distribution, Figure-24, can be compared to the empirical maximum Mw recorded for each fault provided by the European SHARE initiative in Figure-25.

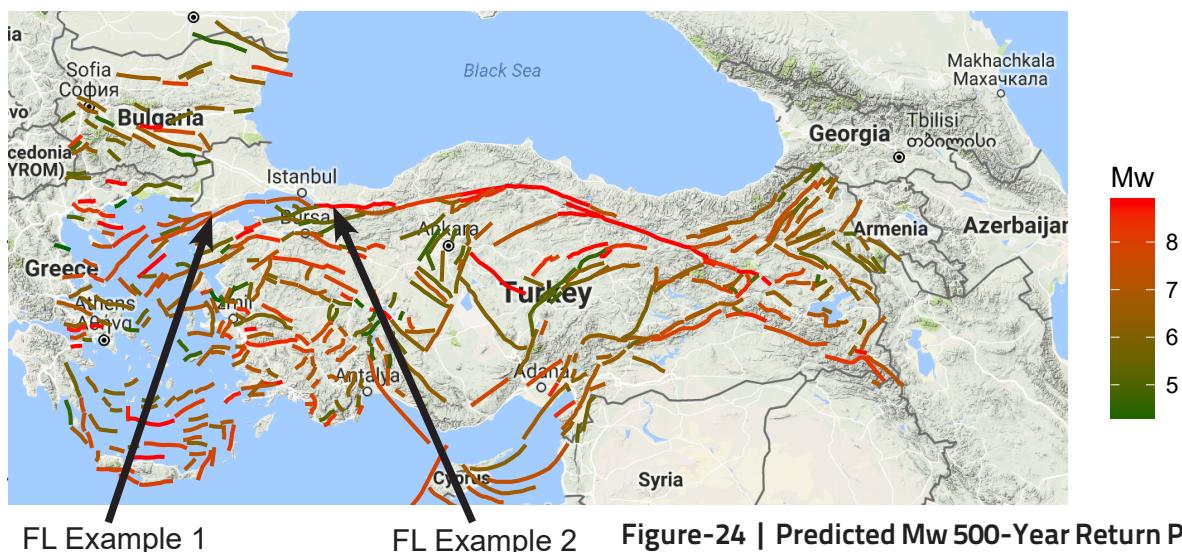
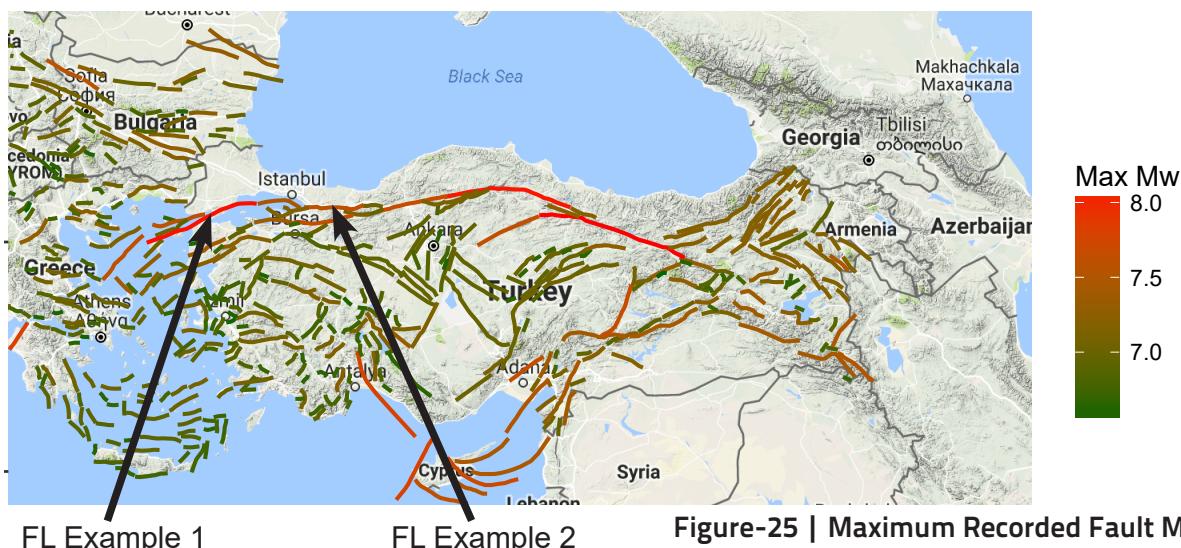


Figure-24 | Predicted Mw 500-Year Return Period

The European SHARE map is similar to the predicted Mw values from the stochastic model. However, as the predicted values are for a return period of 500 years they are generally higher than the SHARE data. A point of note is the highlighted Fault-Line Example 1 in Figure-24 compared to Figure-25. This, although still a fault of with a high magnitude c.7.5 - 8 in the stochastic data is more prevalent in the SHARE maximum Mw plot. This could be because of a large singular outlier event. An advantage of stochastic modelling is to mediate these outlier events with statistical knowledge, providing a more informed understanding than from purely empirical data.



Fault-Line Example 2 has a greater predicted hazard in the stochastic model than shown in empirical data. This has a large impact on the risk map since the fault is close to Istanbul, a highly populated city, hence the risk is larger. The lines display the risk for a fault, this however, assumes that a fault line is equally likely to slip along its length. The North Anatolian fault is over 200km in length, and so sections slip at different time periods. So this along with propagation of stresses along the fault will impact the results, it is suggested that future research in this area is needed.

4.9. Earthquake Dissipation

In order to gain a greater understanding of the hazard that each fault contributes to an area a method of dissipating the fault lines must be incorporated. From Table-3 an earthquake below a magnitude of 5.4 Mw is the threshold before major damage occurs from an event. It is assumed that even with additive effects a lower bound value of 4 is reasonable for analysis, with an energy decay rate:

$$\frac{E_1}{E_2} = 2\pi\gamma^2 \quad (14)$$

Where E is the energy in joules and $\gamma = R^{1.2}$ and R is the radius in km

As mentioned in literature the dissipation of an earthquake event is greater than purely spherical - c. $1/R^{2.4}$ with R being the distance from the source event (km) Equation-7. However, the magnitude of an earthquake is not linearly related to the energy of that earthquake see Equation-15.

$$\log(E) = 1.5Mw \quad (15)$$

Therefore to compare two magnitude events a combination of Equation-14 and Equation-15 is required. See Equation-16 below.

$$Mw_2 = Mw_1 - \frac{\log(2\pi R^{2.4})}{1.5} \quad (16)$$

5. Hazard Map

5.1. Fault Line Dissipation

In order to achieve this map, it shall be assumed that the Mw dissipates equally in all directions and that the topography has a negligible effect. This is not the case, however, for the purpose of this document these topographic effects shall be ignored. The map, although simplified, will still provide a good understanding of the earthquake activity throughout Turkey. The method of dissipating fault line energy has been conducted using the following process:

- Determine the Mw value of the fault - From the stochastic model
- Create a matrix of Turkey with each square representing a km² of area
- Assign the Fault a value of its Mw and plot on the matrix
- Calculated the radius of dissipation - until Mw is equal to 4, Equation-17.

$$R = \left(\frac{10^{1.5(Mw_1-Mw_2)}}{2\pi} \right)^{\frac{1}{24}} \quad (17)$$

Where Mw_1 is the predicted Mw for the fault at a specific return period. Mw_2 is the magnitude at distance R - in this case $Mw_2 = 4$.

No. Squares at distance 1 is 8

No. Squares at distance 2 is 16

Therefore equation relating number of squares to distance:

Where n is the distance from the middle point that is dissipated from. In this case a point on the fault line.

2	2	2	2	2
2	1	1	1	2
2	1	1	1	2
2	1	1	1	2
2	2	2	2	2

Figure-26 | Matrix Explanation

$$\text{No. Squares} = 8 \times n \quad (18)$$

- Determine the number of squares that required at that radius Figure-26, Equation-18.
- The value assigned to the cells decreases as the distance from the central point increases.
- To dissipate each fault line the end points of the fault line are moved in a circular path equal to the radius determined. The circle is split evenly depending on the location of the cells that are to be assigned at that specific radius Figure-27 with $R = 2$.

Since there are $8 \times 2 = 16$ ($8n$) cells to change in this iteration the whole circle of 2π radians must be split into 16 discrete angles.

The coordinates can then be assigned in a Cartesian reference system ready to be implemented into the matrix and the relevant cell locations. The number of segments within 2π radians decreases as the radius decreases.

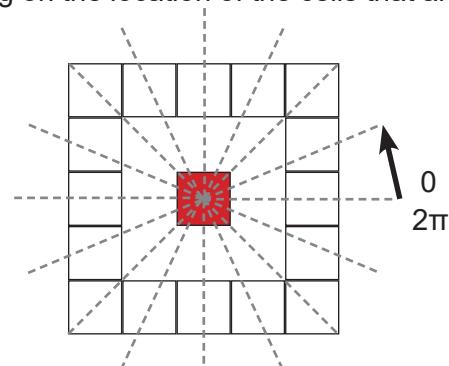


Figure-27 | Matrix Dissipation

- From these angular (polar) coordinates the x and y coordinates of the circle are assigned see Equation-19.

$$x = R\cos(\theta) \quad y = R\sin(\theta) \quad (19)$$

- These circle coordinates are placed in an array. They are now used to dissipate the end coordinates of the fault in a circular pattern, see Figure-28.

The end points are dissipated with respect to the circular path described earlier. These coordinates are then rounded to the nearest cell location. A process such as this is needed to ensure dissipation is circular and not boxlike.

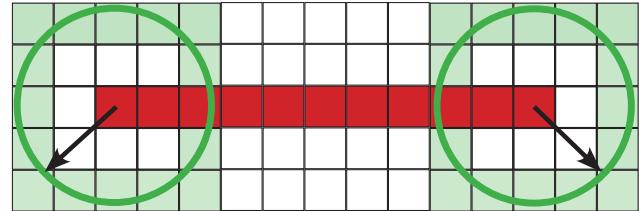


Figure-28 | Matrix Dissipation

- To save computing time the inner fault line coordinates are then offset by R to avoid re-defining cell values. See Figure-29.
- For ease of computation, the largest radius at Mw value of 4 is assigned first. The radius then decreases as Mw increases to the stochastically determined Mw for the particular fault.

For each radius the end point circles and then the offset central points are assigned to their relevant cells. With the size of each cell being small this process is generally repeated c.90 times for each of the 408 active faults.

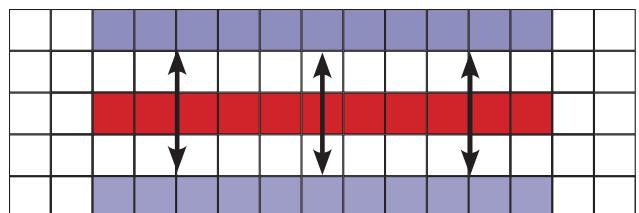


Figure-29 | Offset Fault Line

An example of the entire fault-line dissipation process can be seen below in Figure-30.

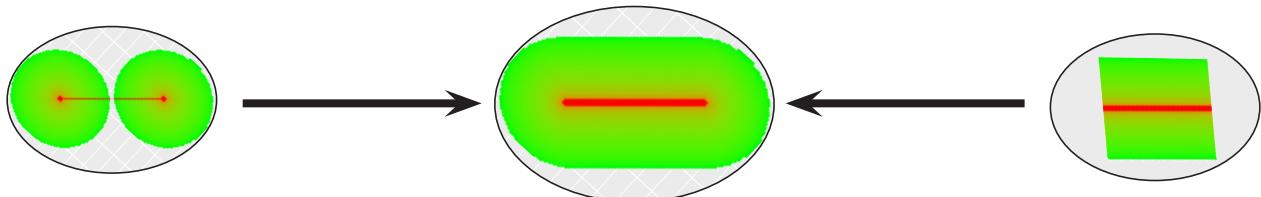


Figure-30 | Complete Dissipation Process

5.2. Combining Dissipation Matrices

There are c.400 fault lines that have a Mw above the minimum value of 4. Therefore there are c.400 individual dissipation matrices. Since the Mw scale is logarithmic a simple additive approach is not valid and must be modified. The logarithmic nature of the Mw scale is demonstrated in Table-10.

Table-10 | Mw Comparison

Mw Value	1 Mw	2 Mw	3 Mw	4 Mw	5 Mw	6Mw	7Mw	8 Mw	9 Mw	10 Mw
Mw Equivalent	$k \cdot 10^1$	$k \cdot 10^2$	$k \cdot 10^3$	$k \cdot 10^4$	$k \cdot 10^5$	$k \cdot 10^6$	$k \cdot 10^7$	$k \cdot 10^8$	$k \cdot 10^9$	$k \cdot 10^{10}$

Therefore each cell in the matrix must be combined over the active 408 faults. This is done with the use of the following formula, Equation-20.

$$\text{Combined Mw} = \log(10^{Mw_1} + 10^{Mw_2}) \quad (20)$$

5.3. Plotting the Hazard Map

Once the 400 fault-lines are dissipated and combined, the map can be plotted using the R package ggplot2, Figure-31. Fault lines are easily visible as areas of high hazard. The largest hazard is located to the North with the North Anatolian Fault producing the predominant hazard zone. In general, as expected, the areas with the least number of faults are mapped as having the smallest hazard. The majority of Turkey, however, is above a 4Mw 500-year return event. Comparison to other countries, also mapped by the SHARE initiative, Turkey is by far the most prone to earthquakes in Europe. Therefore a larger proportion of government spending should be spent on earthquake mitigation.

From the map, Figure-31 (Appendix-A), the middle and southern regions of Turkey are the least affected by earthquake occurrences. These zones may be affected by earthquakes originating in other countries, but generally, offer the least hazard. Government spending in these areas should be minimal; especially as the higher activity areas, 5Mw and above are exponentially more damaging.

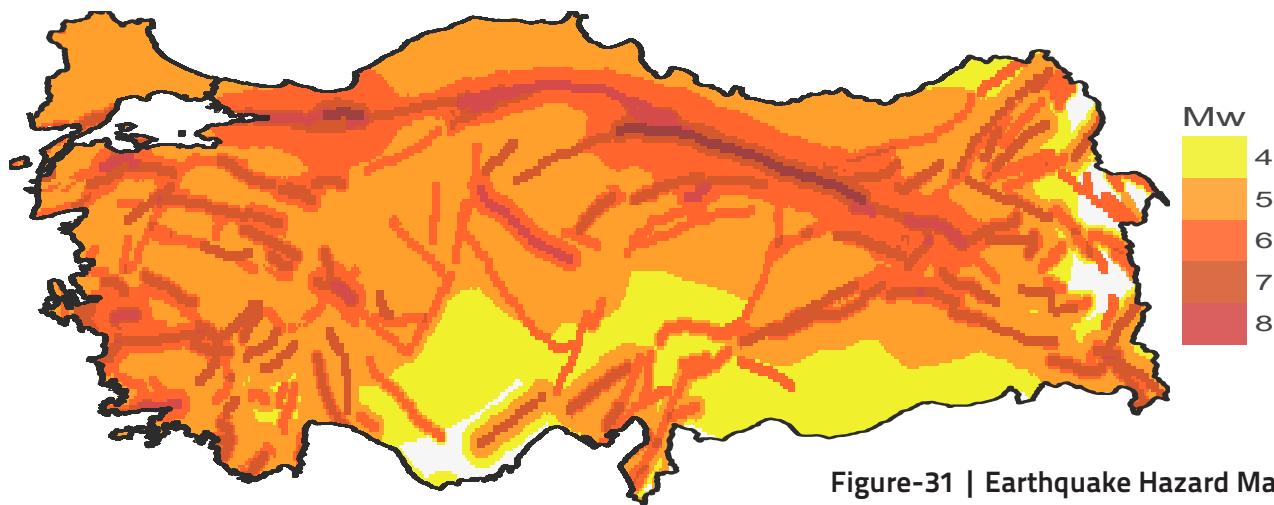


Figure-31 | Earthquake Hazard Map

The largest cities and urban development in Turkey occur close to the Mediterranean Sea, which is to the West. This zone also contains the largest hazard and so building regulations should be strictly controlled here to ensure that high magnitude events are not catastrophic. The risk element of this map in relation to population density shall be considered further in this report.

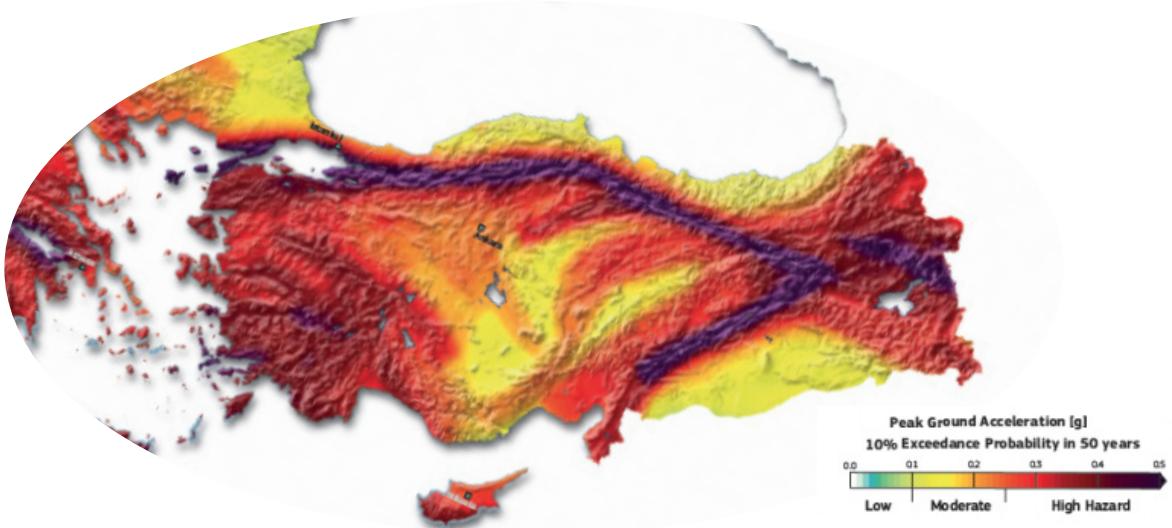


Figure-32 | European Earthquake Hazard Map, (SHARE, 2013)

5.4. Comparison of Maps

The European SHARE map, Figure-32, uses a combination of three earthquake modelling techniques:

- An Area Source Model (AS) - based on the definition of area sources for which earthquake hazard is defined individually.
- A zonation-free and kernel smoothed (SEIFA) probabilistic earthquake rate model related to seismicity and fault movement.
- A Fault Source and Background model (FSBG) that determines seismographic sources using tectonic and geophysical data. The slip-rates are then used to determine activity.

The model proposed in this paper is considered as a combination of the AS and FSBG methods of earthquake prediction. Therefore the two maps for hazard prediction can be compared. The SHARE map predicts ground acceleration in (g) (PGA) with a 10% exceedance probability in 50 years which is closely related to Figure-30, Equation-21 (Kjeldsen, n.d.).

$$T = (1 - (1 - r)^{\frac{1}{L}})^{-1} = (1 - (1 - 0.1)^{\frac{1}{50}})^{-1} \approx 500 \text{ years} \quad (21)$$

Where r is the exceedance probability, L is the duration in years, and T is the return period.

This ground motion can be compared to a magnitude on the Mercalli scale, Table-11 (ShakeMap Scientific Background, 2010). The scales, Mw and Mercalli are different but closely linked.

Table-11 | Mercalli and PGA Comparison

Magnitude	4	5	6	7	8	9	10+
PGA (g)	0.014-0.039	0.039-0.092	0.092-0.18	0.18-0.34	0.34-0.65	0.65-1.24	>1.24

The fault lines on the map created in this report clearly define the areas of highest hazard. These are significantly more prevalent than in the European SHARE map Figure-32, possibly due to the methodology used in this paper of dissipating from the fault lines. However, it is logical that the hazard above a fault line should be larger. A number of these faults are above the 8Mw level, showing that Turkey is a highly active region; this may be emphasised by the lack of data as discussed earlier. The nature of this analysis has determined some areas for future research, the hazard in the West of Turkey is highly dependent on distance to a known fault location. In comparison, the SHARE initiative prescribes an even hazard to the entire locality. If this fault-dependent distinction in this report is correct then reduced spending on earthquake mitigation could occur in these areas. It is almost as important to determine areas where fewer resources are needed as areas of high-hazard. This enables an improved use of government spending in the areas of most need.

However, focusing on the region near Istanbul, the SHARE map displays a slightly more hazard than Figure-31. This will have a large variation on the risk map and may be due to the fault line passing nearby in the Mediterranean. This paper has taken little account of earthquake propagation of submerged topography and hence the SHARE map may be more accurate in this case.

Another region where this method differs from the SHARE map is to the eastern side of Turkey. The SHARE map displays a far higher hazard in this area. This may be due to a lack of earthquake data in surrounding countries skewing the results in the outer-regions. This, however, would make a limited difference as the fault lines in that location have a low hazard, so only pockets of the region are affected. This may be an area for future investigation with a larger dataset as government money could be saved from over-designing structures to mitigate against earthquakes.

5.5. Sensitivity Analysis - Dissipation Coefficient

In Section 4.9 of this report, a dissipation co-efficient of 1.2 was deemed a good approximation of the Earthquake occurrence in Turkey. This was extracted from literature, in which the purely elastic dissipation of earthquakes (1.0) is replaced by an attenuation factor that takes into account the energy loss due to local topography. As discussed these values often vary between 1.1 and 1.3 dependent on region. Therefore a sensitivity analysis is conducted, Figure-33.

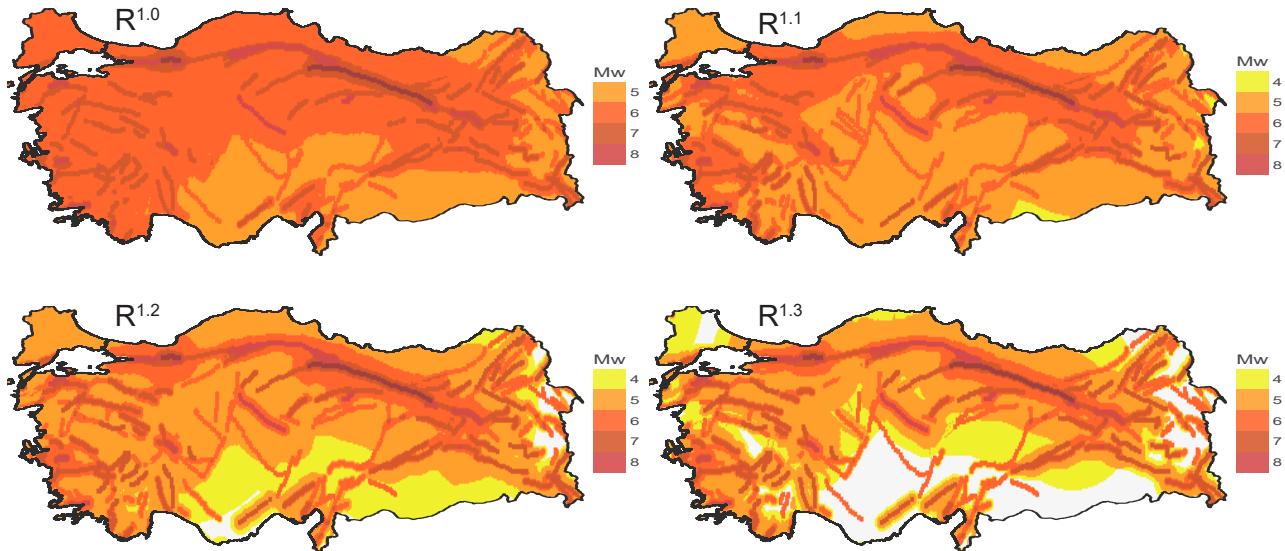


Figure-33 | Dissipation Sensitivity Analysis

The four maps display a similarity in the predicted regions of higher risk, the areas most at risk are to the north-east of Turkey with lesser-risk zones to the South and East. In all plotted maps the North Anatolian fault is the most prevalent. A point of interest is that all maps show very few areas higher than 6Mw. The attenuation between the upper magnitudes is fast, however, dissipation to lower than 4Mw takes far more distance. Therefore, many areas with few active fault lines are still prone to low-level earthquake risk from further, more active faults.

The fully elastic model - R^{1.0} shows a far higher risk for Turkey with no areas below a 5Mw for the 500 year return period. This elastic dissipation, however, is not representative and only plotted for interest. The four graphs display a large variance of prediction, with R^{1.1} mapping some areas as 5Mw where the R^{1.3} predicts lower than 4Mw. As the dissipation coefficient decreases the fault line locations become less prevalent and Mw zonation forms. A good example of this is in the west of Turkey where R^{1.3} clearly shows the fault line locations whereas R^{1.0} provides a more regional estimation (6Mw). It can, however, be assumed that the real value lies somewhere between these extremes. This is an area for further investigation as the SHARE map, which shows a regional estimation, could lead to higher predictions and spending in these lower-risk locations.

The four plotted maps, although providing a visual description of the data, are not fully representative of the predicted values. These have been plotted as discrete bins (4, 5, 6, 7, 8, 9 respectively) which contain the continuous Mw predictions for Turkey as this is easier to visualise. Therefore the changes in predictions may seem exaggerated in this experiment as each value is rounded to the closest bin. However, it is clear that correct parameter estimation is vital in the production of a hazard map. This should be performed on a regional basis and therefore the coefficient will vary across Turkey dependent on local topographic features. Estimation of these parameters are outside the scope of this document and so the R^{1.2} blanket coefficient shall be used in further analysis.

5.6. Sensitivity Analysis - Distribution Selection

The Gamma distribution was selected for further analysis in this report. However, a number of distributions were selected in literature, the Gamma, GEV, Log-Normal, Weibull, and Frechet. A comparison of the Log-Normal and Gamma distributions shall form the basis of this experiment. Each distribution, as before is used to predict a 500-year return Mw event and the results of this analysis are plotted in Figure-34. Values are again limited to a maximum 8.8Mw predicted event.

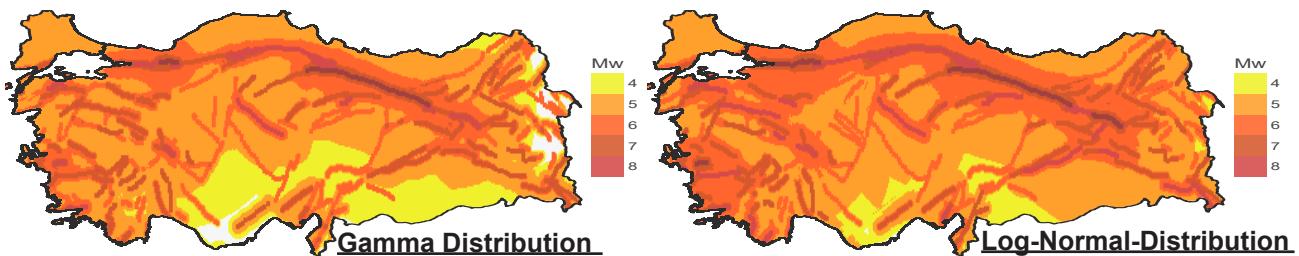


Figure-34 | Distribution Sensitivity Analysis

In the highly-censored individual-fault line experiment the Log-Normal distribution predicted a higher value than the Gamma. This is replicated in the maps produced; the Log-Normal predicts a greater hazard for the whole of Turkey. The discrete nature of the plot may, once again, skew the results. Regions surrounding the North Anatolian fault line in both distributions pose a high hazard, however, the influence of this is distributed over a larger area in the Log-Normal map.

Of particular note are the regions based to the East where the Gamma distribution predicts values <4Mw whereas the Log-Normal predicts 5Mw. The predicted Mw for fault-lines located in the region are c.6Mw, Gamma as lower in the 6Mw range and Log-Normal the higher values. Although little difference in prediction, the clustered nature of faults in the region and the circular dissipation provides a very different area estimation.

The distribution choice however, in this case, seems to impact the map little on a 500-year return period. By comparing Figure-23 and Figure-21 it is also clear that increasing the quantity of data used to produce the model will decrease this prediction difference. As discussed previously, there is not enough data at the time of writing to enable accurate prediction, however, as catalogues improve so will accuracy. The choice of distribution, apart from the Frechet, is not as vital as determining the correct regional dissipation parameter.

6. Risk Map

6.1. Population Density

As mentioned in the previous section, the hazard purely relates to the probability of an event occurrence. This document shall quantify risk by using a population density map of Turkey (Gridded Population of the World, Version 3 (GPWv3) Data Collection, 2005) in a grid format. The dataset has a resolution 2.5 arc-minutes and contains a population count for each grid cell. To quantify the risk, population data, Figure-35, can be multiplied by the predicted Mw in that location.

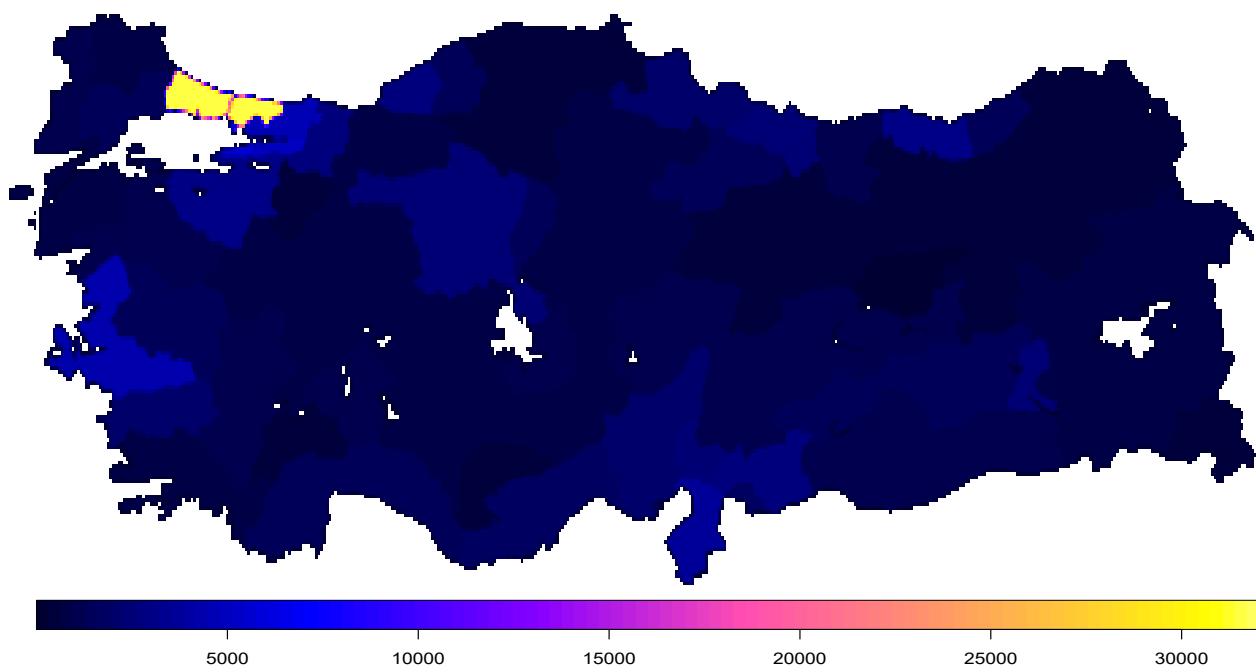


Figure-35 | Turkey Population Map

The population map shows an exceptionally high density in the region of Istanbul, this makes sense as it is a major Turkish city. However, the map does not display an equally high density around the capital city of Ankara. This could be due to Ankara being a stand-alone city and the low map resolution, but with comparison to other population maps, this seems a recurrent feature. Istanbul population is c.14.8 million with a number of other cities located in the immediate area such as Silivri and Izmit. Ankara, on the other hand, has a population of 4.6 million with little surrounding urban regions. Hence the largest consequence of an earthquake would be in the Istanbul region. The other areas of Turkey have a generally limited population and so the consequences are lower. There are a number of cities located close to the Turkish border, predominantly on the eastern side. Since these cities are located in different countries, such as Aleppo in Syria, Gyumri in Armenia, and Batumi in Georgia the risk to these locations is out of the scope of this document. It is, however, important to note that nearby earthquake occurrences in Turkey would pose a risk to these neighbouring cities.

6.2. Risk Map Calculation

The predicted M_w for a given event is calculated in the hazard map section of this report. This can be combined with the population map using the following, Equation-22.

$$Risk = \log_{10}(10^{1.5 \times M_w} \times Population\ Count) \quad (22)$$

The risk is quantified as a measure of the predicted released earthquake energy. This method takes into account the logarithmic nature of earthquake occurrence. A magnitude 8Mw causes significantly more damage than a magnitude 7Mw event, Table-3.

6.3. Plotting the Risk Map

Once the two geogrids, one containing hazard and the other population data are combined using Equation-22, the new risk map can be plotted. This map is plotted below using the ggplot2 raster function of R overlaid onto the administration boundary of Turkey, Figure-36.

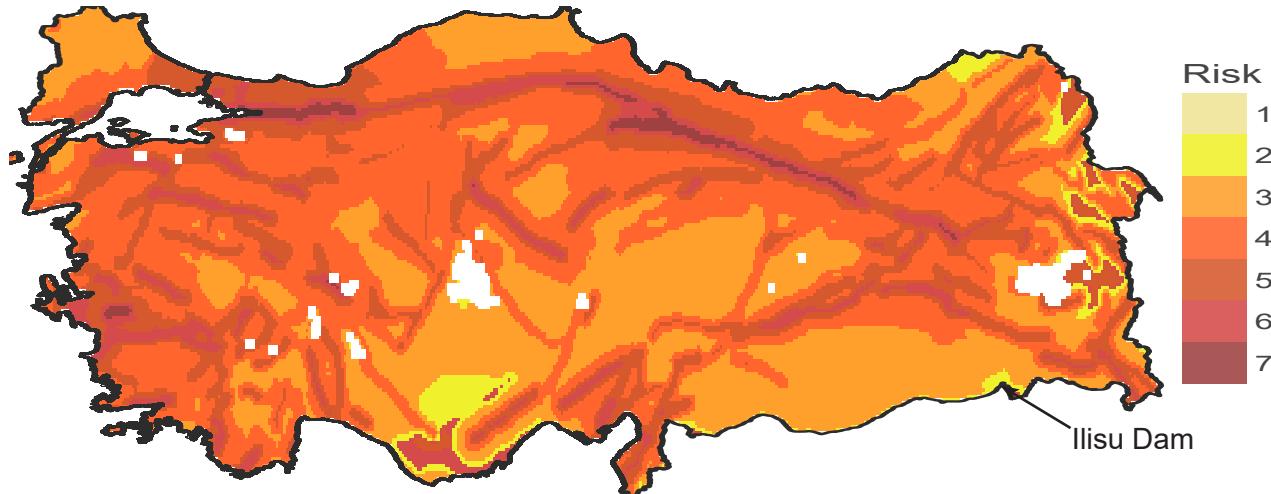


Figure-36 | Earthquake Risk Map

The Risk Map still distinguishes the areas in close proximity to the faults as high-risk. This may be due to the methodology behind the map, where the predicted Mw has a high weighting on the risk; some maps may use a larger population weighting which would alter this. However, a larger risk is calculated for the densely populated areas. Istanbul, for example, contains a far higher risk in proportion to the rest of Turkey than predicted in the hazard map. Although Istanbul is the most populated area of Turkey, in this assessment it is not the area of greatest risk. In this simplistic model, the risk is purely proportional to the population present in that area and the earthquake hazard. It does not account for the difference in infrastructure and its ability to cope with earthquakes between these areas. For example - low built wooden dwellings in a rural environment would be far less risky to residents than a poorly constructed concrete multi-storey dwelling. Not least because on a day-to-day basis rural residents are outside, which greatly reduces risk as building collapse is the main cause of loss-of-life in an earthquake event (Tulane, 2016). Variations of this risk-map can be seen in Figure-37 where population is raised to a power to account for the difference in infrastructure.

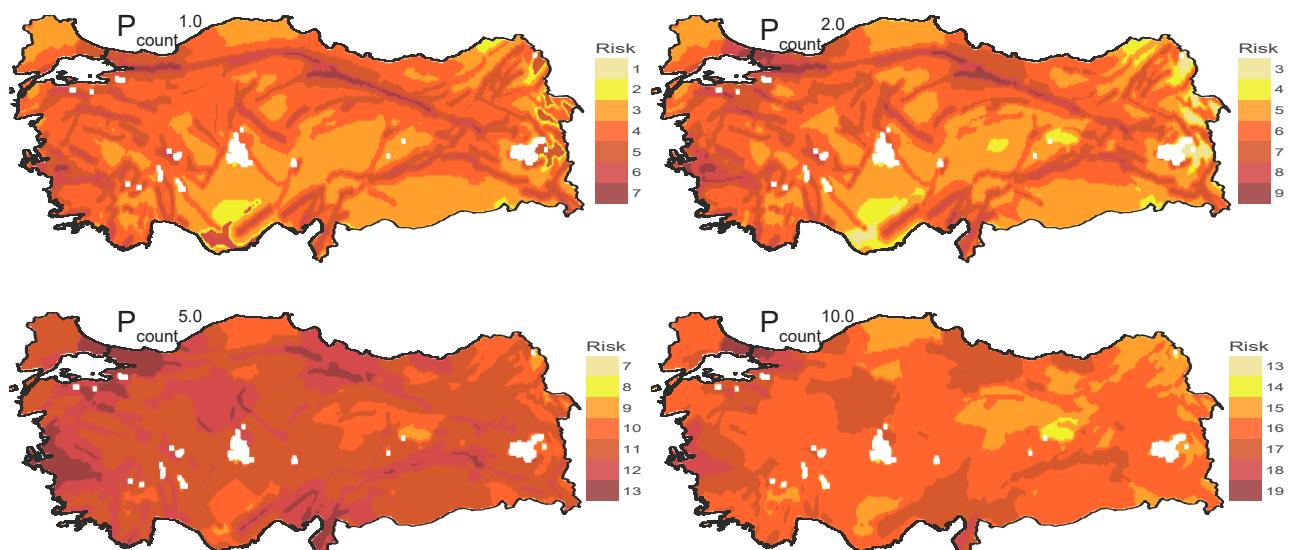


Figure-37 | Risk Sensitivity Analysis

Although the values of the risk map are not directly translatable to the number of injured civilians, there is a clear representation of the proportion of government earthquake funding that needs to be spent in relevant areas. In Figure-37 this spending will vary significantly with respect to unknown parameters such as the type of construction and working environment which increase the population component of the risk map. When population is a lower-valued variable the Mw of that area takes the greatest importance. However, when population is to the power of 5 both variables have a similar weighting and so the predicted risk is more evenly distributed throughout Turkey. When to a higher power, the population is the deciding factor and fault-line locations are less visible. It is out of the scope of this document to determine the correct factor, however to the power of 2 seems a logical approximation here and is re-plotted in the appendices (Appendix-B). This provides both a good distinction between areas of population density but also displays the hazard well.

The general areas of high-risk are located in the West. A majority of Turkey's urban areas are located on the Mediterranean Sea, with easy-access to maritime trade-routes so the population density is higher in these areas. The risk map represents this uneven population spread more effectively, assigning a larger risk to the western regions. One aspect that is not accounted for in this assessment is the risk provided by the Ilisu Dam, which is currently being constructed in the south-east of Turkey on the river Tigris. Although in a low-hazard zone would provide increased risk as there are greater consequences from its collapse, with a larger proportion of affected population, Figure-36.

The risk map also represents areas where government spending should be used for treatment facilities so the correct aid can be deployed effectively to victims. A higher population dictates more aid and a greater likelihood for loss-of-life in that area. A good recent government scheme is to encourage property owners to purchase earthquake insurance. Sadly it is estimated that only 25% of dwellings are insured in such a way (Fethiyetimes, 2010), however, the insurance industry can be predicted to allocate money to regions proportionally. This allocation of resources will closely resemble the risk map produced in this report, as the population density of a region is likely to determine the spending along with the probability of an event occurrence.

7. Conclusions

7.1. Investigation of Earthquake Analysis Techniques

This report has investigated a number of established earthquake hazard techniques, mainly focusing on probabilistic analysis. Distributions suggested in literature have been tested with regards to goodness-of-fit. The results, however, are not conclusive for the area of Turkey and so the topic remains a subject of debate and future research. Although unable to determine a superior distribution, the Frechet model has been shown to lack accuracy on highly-censored data, which is often found in earthquake catalogues. The report has emphasised the importance of up-to-date and more expansive catalogues. There is a need for more reliable data-collection, and extending below the 4Mw minimum recorded value in this report.

7.2. Comparison of Hazard Map to European SHARE Map

On comparison of the earthquake hazard map created in this report and the European SHARE map, a number of areas for further investigation have been identified. The western regions of Turkey, although a similar average hazard is predicted in both maps, the SHARE map discretises the area less effectively. The map produced in this report determines pockets of lower-hazard amongst the high-hazard areas in close proximity to fault-line locations. This is also the case for the eastern regions of Turkey. Further investigation is needed, and if proven valid, this would enable a better distribution of government funding. The correct allocation of resources is vital as money that could be saved in these regions could be better used in the higher-hazard and risk locations.

7.3. Hazard Map Implications

It is also clear that significant industry should be constructed away from these zones of high-hazard to ensure little damage occurs to the economy from an event. Major cities, such as Istanbul are unable to re-locate and only mitigation schemes are possible, such as insurance and improved building standards. However, emerging industries have greater manoeuvrability with regards to location as does planned infrastructure. An important observation is the correct placement of the new Tigris dam in an area of low hazard. This location reduces the construction cost as infrastructure in areas of low hazard are likely to experience lower seismic forces. It is suggested that industries should be encouraged to consider earthquake risk among other determining factors such as transport and availability of labour when determining a location for activities.

7.4. Risk Map Implications

The risk map highlights the areas with the greatest need for investment and infrastructure to mitigate earthquake occurrences. Unlike the hazard map which purely displays the likelihood of an event occurrence, the risk map relates this likelihood to the consequences, in this case, the affected population. Correct government planning, with respect to number of hospitals, aid locations etc. can be better informed by this map. An area identified as high-risk is the major city of Istanbul, with the highest population of c.14 million this is a region that will require substantial aid and financial support if a major event was to occur.

7.5. Sensitivity and Areas for Further Investigation

From the sensitivity analysis, there are a number of factors that affect the hazard map creation. The most prevalent, as discussed earlier is deemed to be an estimation of topographical earthquake attenuation. This varies on a regional basis dependent on the local landscape, in the case of this report a blanket estimation of 1.2 is used. Further research and improved regional estimation is required to reduce uncertainty in the hazard map.

The risk map developed in the final section of this report is subject to the proportion of consequence vs. hazard. This map is solely based on population density, an area of further study that would be beneficial is relating the population density to the number of human casualties. This, however, is another regional estimation as local building codes and resident occupations will determine the consequences of an event in that area.

8. References

- Baddari, K., Bellalem, F., Baddari, I. and Makdeche, S. (2016). Some Probabilistic and Statistical Properties of the Seismic Regime of Zemmouri (Algeria) Seismoactive Zone. *Acta Geophysica*, 64(5).
- Bolt, B. (1976). Hazards from earthquakes. *Die Naturwissenschaften*, 63(8), pp.356-363.
- Croot, E. (2010). Maximum Likelihood Estimators and least squares. pp.1-2.
- D'Agostino, R. and Stephens, M. (1986). Goodness-of-fit techniques. 1st ed. New York [etc.]: Marcel Dekker.
- Distribution Pdf and Cdf. (2016). [image] Available at: http://work.thaslwanter.at/Stats/html/_images/PDF_CDF.png [Accessed 1 Apr. 2017].
- Earthquake.usgs.gov. (2015). How much bigger is a magnitude 8.7 earthquake than a magnitude 5.8 earthquake?. [online] Available at: https://earthquake.usgs.gov/learn/topics/how_much_bigger.php [Accessed 1 Apr. 2017].
- Earthquake.usgs.gov. (2017). Earthquake Glossary. [online] Available at: <https://earthquake.usgs.gov/learn/glossary/?term=recurrence%20interval> [Accessed 1 Apr. 2017].
- Earthquake.usgs.gov. (2010). ShakeMap Scientific Background. [online] Available at: <https://earthquake.usgs.gov/data/shakemap/background.php#intmaps> [Accessed 11 Apr. 2017].
- Edwards, B., Fäh, D. and Giardini, D. (2011). Attenuation of seismic shear wave energy in Switzerland. *Geophysical Journal International*, 185(2), pp.967-984.
- EFEHR.org. (2017). European Seismic Hazard Model. [online] Available at: <http://www.EFEHR.org:8080/jetspeed/portal/hazard.psml> [Accessed 1 Apr. 2017].
- Endsley, K. (2017). Earthquakes Magnitude Scale and Classes. [online] Geo.mtu.edu. Available at: <http://www.geo.mtu.edu/UPSeis/magnitude.html> [Accessed 1 Apr. 2017].
- Greenhough, J. and Main, I. (2008). A Poisson model for earthquake frequency uncertainties in seismic hazard analysis. *Geophysical Research Letters*, 35(19).
- Gridded Population of the World, Version 3 (GPWv3) Data Collection. (2005). [Geospatial Data Presentation Form: raster digital data, map] <http://sedac.ciesin.columbia.edu/gpw/index.jsp>, New York.
- Gurenko, E. (2006). Earthquake insurance in Turkey. 1st ed. Washington, DC: World Bank, p.6.

Herraiz, M. and Espinosa, A. (1987). Coda waves: A review. *Pure and Applied Geophysics PAGEOPH*, 125(4), pp.499-577.

Itl.nist.gov. (2017). 1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test. [online] Available at: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm> [Accessed 1 Apr. 2017].

James, G., Witten, D., Hastie, T. and Tibshirani, R. (n.d.). An introduction to statistical learning. 1st ed.

Jeffreys, H. (1938). SOUTHERN EARTHQUAKES AND THE CORE WAVES. *Geophysical Journal International*, 4, pp.281-308.

Kagan, Y. and Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *Journal of Applied Probability*, 38A, pp.158-175.

Kjeldsen, T. (n.d.). Introduction to frequency estimation of extreme events.

Kolmogorov, A. and Smirnov, (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Istituto Italiano degli Attuari*, 4, pp.83-91.

M, I. (2010). Government Encourages People to Buy Compulsory Earthquake Insurance - Fethiye Times. [online] Fethiye Times. Available at: <http://www.fethiyetimes.com/property/property-a-utilities/6324-government-encourages-people-to-buy-compulsory-earthquake-insurance.html> [Accessed 11 Apr. 2017].

Mosteller, F. and Tukey, J. (n.d.). Data analysis and regression. 1st ed.

Normal Distribution. (2014). [image] Available at: https://saylordotorg.github.io/text_introductory-statistics/section_09/e7a042db29b39bb94416c06789301faa.jpg [Accessed 1 Apr. 2017].

Ordaz, M. and Arroyo, D. (2016). On Uncertainties in Probabilistic Seismic Hazard Analysis. *Earthquake Spectra*, 32(3), pp.1405-1418.

Parvez, I. and Ram, A. (1997). Probabilistic Assessment of Earthquake Hazards in the North-East Indian Peninsula and Hindu Kush Regions. *pure and applied geophysics*, 149(4), pp.731-746.

Pasari, S. and Dikshit, O. (2013). Impact of Three-Parameter Weibull Models in Probabilistic Assessment of Earthquake Hazards. *Pure and Applied Geophysics*, 171(7), pp.1251-1281.

SSHAC (Senior Seismic Hazard Analysis Committee), (1997). Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. US Nuclear Regulatory Commission Report, CR-6372, p.888.

Reid, H. (1911). The Elastic Rebound Theory of Earthquakes. *Bulletin of the Department of Geology*, 6, pp.413-444.

Richter, F. (1935). An Instrumental Earthquake Scale. *Bulletin of the Seismologic Society of America*, 25, pp.1-32.

Statistic Brain. (2017). Earthquake Statsitics - Statistic Brain. [online] Available at: <http://www.statisticbrain.com/earthquake-statistics/> [Accessed 1 Apr. 2017].

usgs.gov. (2015). USGS FAQs - Measuring Earthquakes - Moment magnitude, Richter scale - what are the different magnitude scales, and why are there so many?. [online] Available at: <https://www.usgs.gov/faq/categories/9828/3357> [Accessed 1 Apr. 2017].

Wang, J. and Wu, M. (2014). Risk assessments on active faults in Taiwan. *Bulletin of Engineering Geology and the Environment*, 74(1), pp.117-124.

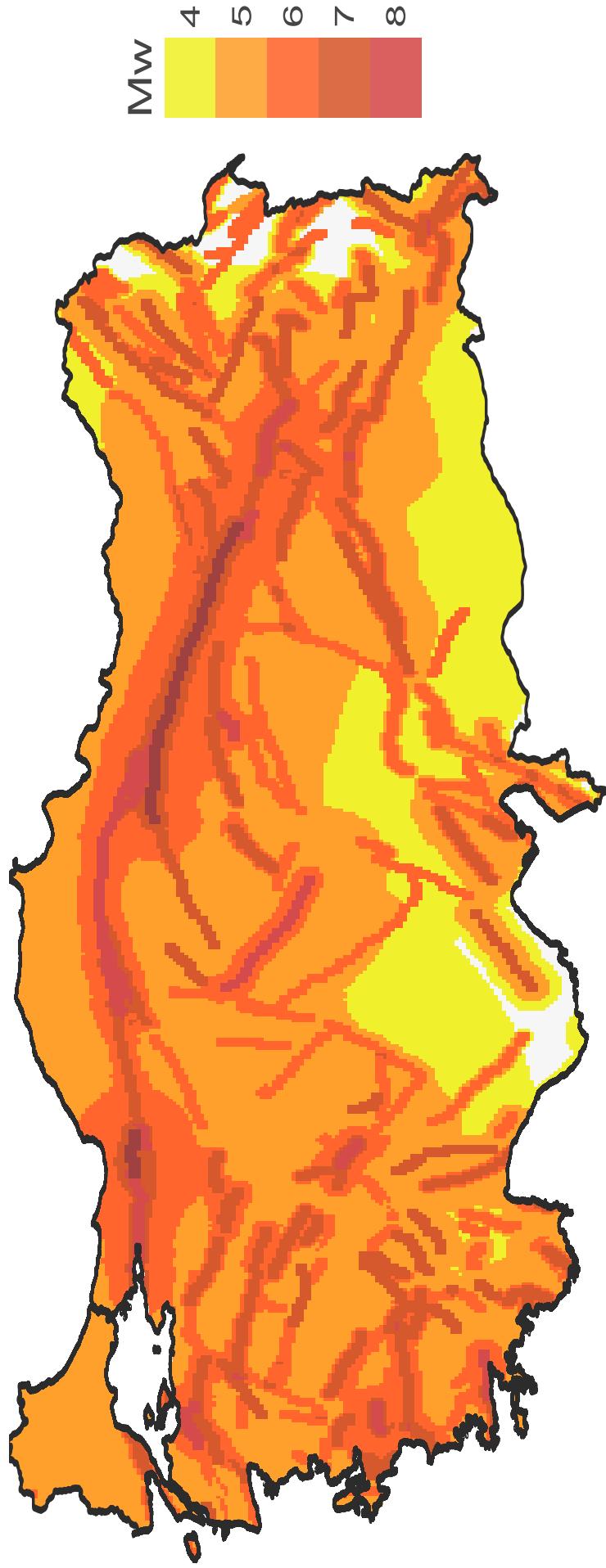
Yadav, R., Tripathi, J., Rastogi, B., Das, M. and Chopra, S. (2010). Probabilistic Assessment of Earthquake Recurrence in Northeast India and Adjoining Regions. *Pure and Applied*

Zavalov, A. (2006). Intermediate Term Earthquake Prediction. p.254.

Zivot, E. (2001). Maximum Likelihood Estimation.

9. Appendix A

Turkey Hazard Map



10. Appendix B

Turkey Risk Map

