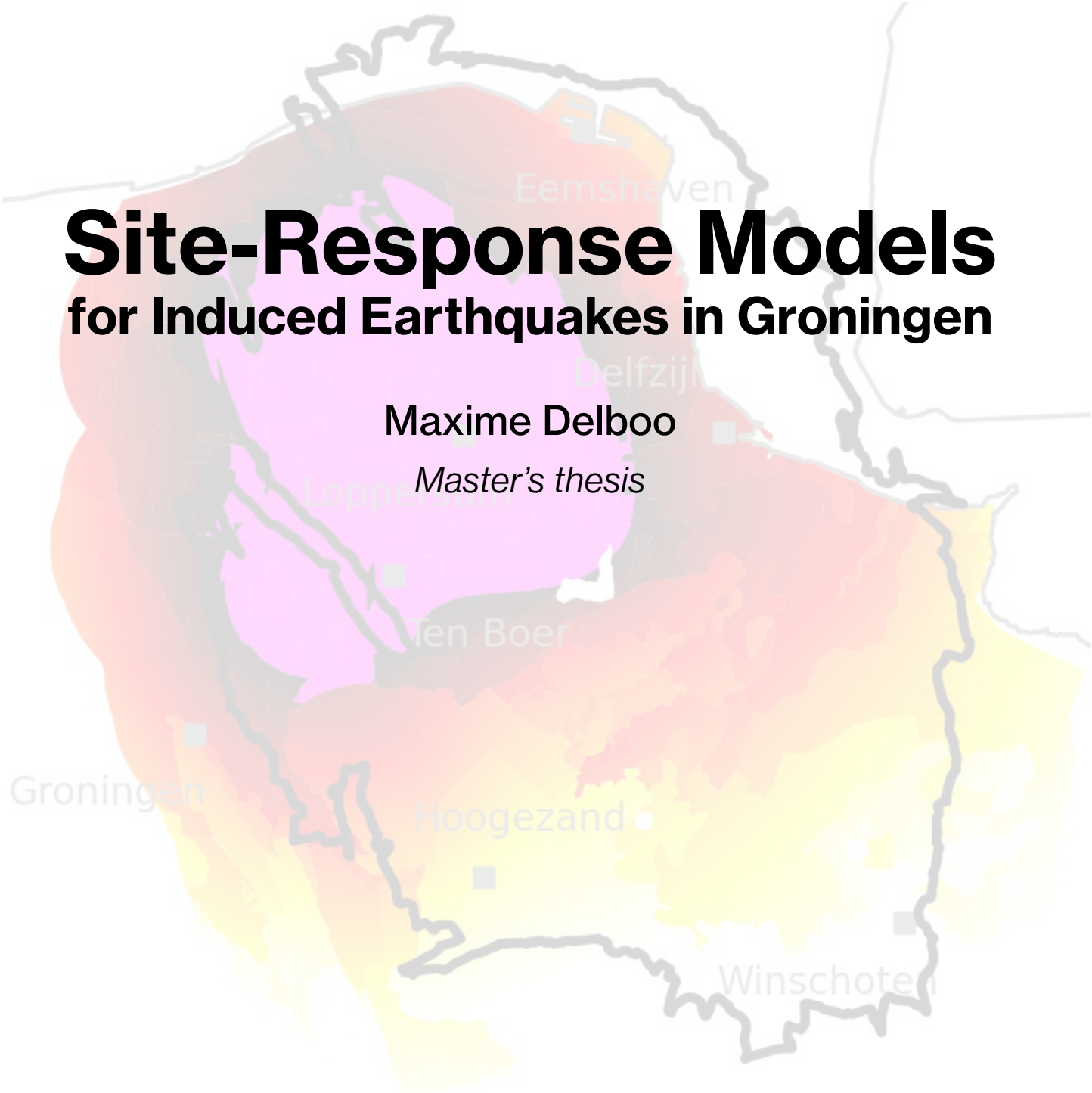


Site-Response Models
for Induced Earthquakes in Groningen

Maxime Delboo

Master's thesis

December 2024, Delft

Daily Supervisor Özge Şahin
Responsible Supervisor Dorota Kurowicka
TNO Supervisor Dirk Kraaijpoel

Thesis Committee
Dorota Kurowicka, Özge Şahin, Alexis Derumigny, Dirk Kraaijpoel

TU Delft

Delft University of Technology

Site-Response Models for Induced Earthquakes in Groningen

Author:
Maxime Delboo

Supervisors:
Dr. D. Kurowicka
Dr. Ö. Şahin
Dr. D. Kraaijpoel

December 19, 2024

Preface

You are about to read my master's thesis, which I wrote for my master degree in Applied Mathematics at the Delft University of Technology. This is the place where I grew up from being a young student, full of enthusiasm and passion for the study of mathematics, to what I am today. Fortunately for me, after around 6 years of deepening my mathematical knowledge, I have kept this enthusiasm and passion for the field. Throughout this journey, I have learned a lot, both in mathematics and in life. Perhaps the most insightful discovery that one learns along the way to becoming a master in a certain subject is that it feels like the number of things you still need to learn is an increasing function of time, but at the same time, so is your knowledge, which is frankly quite paradoxical.

This thesis started out with me contacting Dorota, my responsible supervisor, about whether she knew of any projects I could do for a master's thesis, and then one day she came back to me with this project and introduced me to Dirk and Özge. I said yes immediately and began working on the project two months later. This was quite a lot more challenging than what I expected, as I had to introduce myself to an entirely new field of earthquakes, which I knew nothing more about than the average person. Then now, 9 months later, I once again come to the same conclusion: I have now learned so much more about earthquakes, but equally so I have learned that my knowledge spans only a tiny fraction of the field.

Perhaps one can find this conclusion unsatisfactory or frustrating. In fact, it is quite a humbling, to say the least, but I also think it is beautiful. Where I find my peace is that one simply shouldn't be looking to know everything, rather that one should learn for the sake of learning. Because learning is fun! And through this journey of forced education in my high school years and slightly less forced education at university, I feel like I might just have learned how to do it all myself, and I can keep having the joy of learning throughout the rest of my life, at my own pace and about any subject I want. All this to say: I want to thank the TU Delft for my education and for learning me how to be my own teacher.

Next, I want to sincerely thank Özge and Dirk for their continued support throughout the project. They helped me day in, day out by supervising my work and giving guidance when I asked for it, or when I needed it. Because Dirk is the expert in the field of seismology, he was mostly my teacher. He was the one giving the new material to study and the one giving the assignments. Özge on the other hand, was more there to be a listening ear. She also didn't have any knowledge of seismology either, so my meetings with here were often times centered around me trying to explain what I am doing and then her giving me a reality check on whether what I am doing makes sense and seeing if I fully understand it myself. Both were extremely useful and helped me tremendously in completing a large project like a thesis.

I also want to thank Dorota. She is the one who gave me my thesis subject and although she didn't involve herself in the project on a daily basis, her office was always open to walk into and have a nice chat. Besides that, in the few instances that she did have a meeting with us, she was very helpful in guiding the project towards the finish line and setting boundaries for me.

Lastly, I want to thank TNO, who co-organized this thesis project for me. I ended up with the geophysics team, which was an amazing group of people, who were very enjoyable and interesting people to hang out with and learn new things from.

Now then, onto then thesis. Enjoy the reading!

Abstract

This thesis investigates the Ground Motion Model (GMM) for the Groningen Seismic Hazard & Risk Analysis. We look at various aspects of the model to see where improvements can be made. We start by looking at model calibration and validation, where we check to what extent the proposed model along with its parameter fits the data. Here we come to the conclusion that both the parameters and the model itself have room for further optimization to reflect the data set more accurately. In addition to this, a new model for correlations of site-response amplifications is presented. The topic of how the dependence of these quantities should be modeled is still unclear. The lack of coherent solution to this modeling problem makes the proposed model valuable, as it is simple in nature and reflects the data we have the best.

Lastly, the main improvements that come out of this research are in on the computational front, by finding a novel method for calculating average spectral accelerations, which is the main quantity used in risk assessment. This method generates the distribution of this quantity in one step using numerical integration rather than the previously used Monte Carlo method. The method speeds up computations by a factor of 550 times.

Contents

1 Introduction

The Groningen gas field, located in the northern part of the Netherlands, is one of the largest natural gas deposits in the world. Since its discovery in 1959, it has played an important role in the development of the Dutch economy, being the main provider of natural gas in the Netherlands and neighboring countries. However, the extraction of gas has led to unintended consequences. In particular, induced earthquakes have caused a lot of distress in the region and raised the question whether gas extraction should be reduced or even completely stopped. This question has been the topic of political debate for more than a decade now and has led to the permanent closure of the gas field in April 2024.

In order to make decisions about complex problems like this, it is important to have an accurate estimate of the risks involved in the subsequent extraction of gas. To this extent, many risk models have been built and refined over the years [1]–[4]. These models can be seen as a chain of submodels, ranging from estimated occurrence rates of induced earthquakes and their magnitude distribution to the economic risks of continuing extraction of gas. One of these sub-models is called the ground motion model (GMM). Ground motion models are mathematical models that predict the intensity of ground shaking during an earthquake. They relate earthquake attributes, such as magnitude, rupture distance, and local geological conditions, to the expected level of seismic shaking at a specific location. The scientific field that studies these phenomena is called statistical seismology. A complete overview of the model chain can be seen in Figure 1.

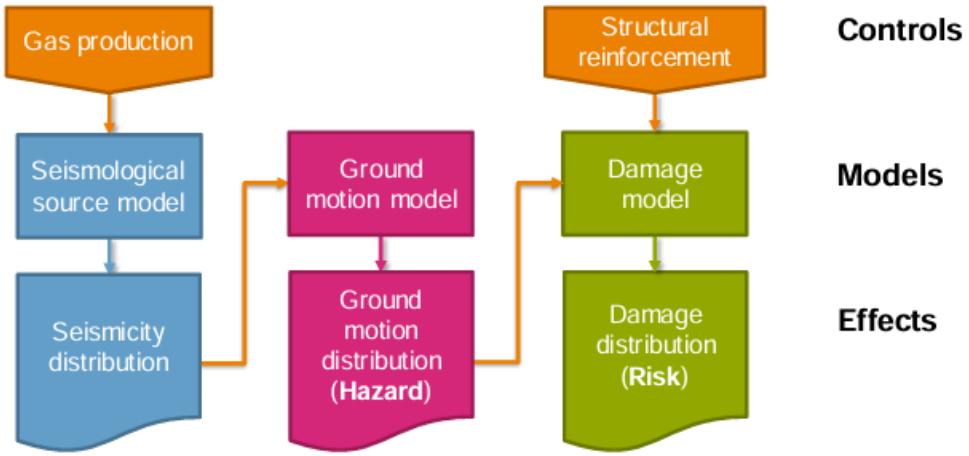


Figure 1: Visual representation of the Groningen Seismic Harzard and Risks model chain. Source: [5].

One of the most important physical quantities for estimating the shaking of buildings is called spectral acceleration. It is a measure of the maximum acceleration that a building would experience (which is directly related to the forces it feels) given some earthquake signal. One can imagine different buildings responding in different ways to the same earthquakes; therefore, there is no single spectral acceleration. Rather, there is a whole spectrum of spectral accelerations. The most simple model we can make for a building is to represent it as a one-dimensional oscillator of a certain eigenfrequency,

where we expose the oscillator to the driving force that the earthquake would cause and subsequently measure the maximum acceleration it experiences due to this force.

An important feature of the Groningen GMM is that it features a site response model. In this type of model, the region is divided into different zones based on geological characteristics to obtain an estimate of how earthquakes can affect each zone. This is done in regions where the shallow subsurface is not homogeneous enough to produce an accurate model. In Figure 2, we have a map of these zones. In the zonified model of Groningen, only a part of the subsurface is modeled to be different per zone. From the source, which is generally around 3000m deep, the subsurface is modeled to have the same properties up to a reference depth of approximately 800m. All variations in the subsurface have been absorbed into the variability of the reference model. This choice is based on known geological profiles of the area; see Kruiver et al. for more information on geological zonification [6].

Because Groningen is zonified in this way, we need to have a separate GMM for each zone. Then again, since a large part up to a reference depth is modeled to be identical, it makes more sense to make a two-stage model: we have one stage that models the spectral acceleration at a depth of 800m, which we call reference spectral acceleration, and another stage that models the zone-specific part. This is done by modeling so-called amplification functions in these zones. These take a reference spectral acceleration and give the factor by which you need to amplify (or attenuate) the reference value to get the surface value, which is called simply the spectral acceleration, or as I will often call it, the surface spectral acceleration, to avoid confusion. Figure 3 shows a schematic overview of the site-response model.

Typically, the subsurface carries the seismic waves linearly, which means that regardless of the magnitude of the wave traveling through it, it will be amplified or attenuated by a constant amplification factor. This is called linear behavior. However, some types of subsurface will 'break down' at high magnitudes, due to the heavy shaking caused by the wave, at which point they will stop showing linear behavior.

In this thesis, research has been conducted into various aspects of Groningen's ground motion model, with a particular focus on the site amplification component. The research is broad in nature and tries to look at the ground motion model from multiple perspectives. It can largely be split up into three areas:

1. **Model calibration:** In this category we will look at how we can estimate the best-fitting parameters and compare these to the parameters prescribed by the original author. To do this, we used a large dataset which features different types of earthquakes with each reference spectral acceleration and amplification values.

2. **Model validation:** In this part of the research, we look at whether some model choices are coherent with the data set and argue to which degree the model could be improved by changing aspects of it, such as the choices of probability distributions or certain functional forms for calculating parameters. This provides insights for both my own research and future research.

3. **Computational advancement:** In this area, we investigate the best and most computationally efficient ways to do risk calculations. Where the other two parts

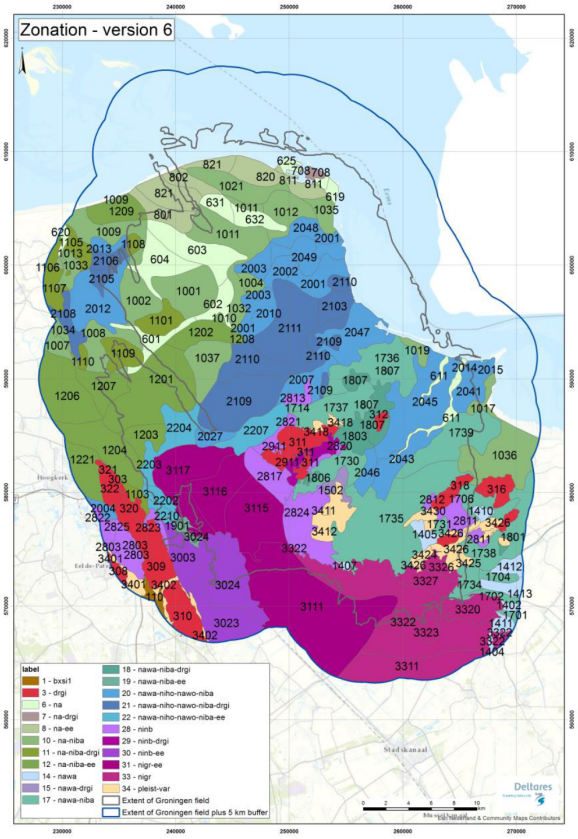
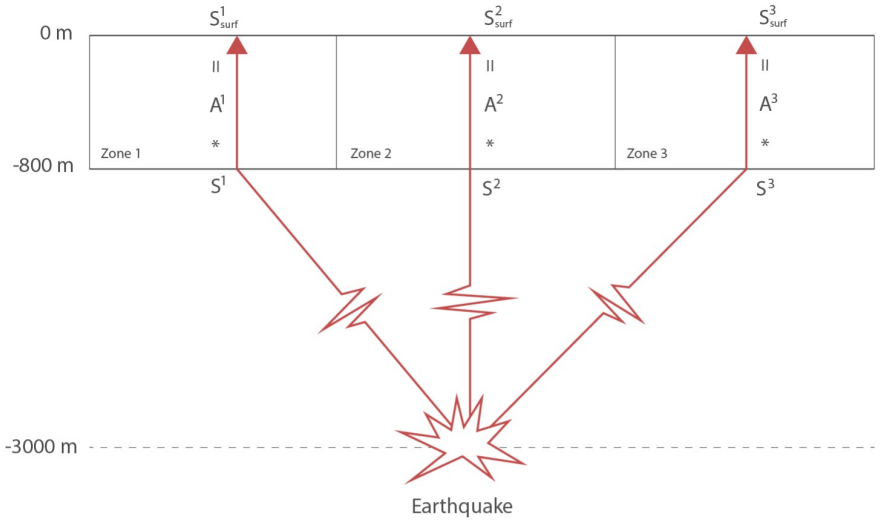


Figure 2: Zonation map for site-response model of the Groningen area. Source: [7].



S = Reference spectral acceleration

A = Amplification factor

S_surf = Surface spectral acceleration

Figure 3: Schematic overview of the site-response model

are more about making the model represent the phenomena we are investigating as accurately as possible, this part is about making sure that the model is not too computationally complex to use for risk analysis purposes. We do this by creating new methods to calculate the model and comparing them to existing ones.

The thesis will be structured such that each of these parts of the research will be featured in its own chapter. Before this though, it is necessary to give a review of the GMMs we will we working with, as well as to give insights into the data we are working with. This will be done in Chapters 2 and 3, respectively. Next, Chapters 4, 5 and 6 will feature each part of the research as described above. After this, we will look at the impact of all proposed changes that follow from this research by looking at how they affect the risk assessment. This will be done in Chapter 7. Thereafter, we will raise some critical questions in a discussion in Chapter 8. Finally, we present our conclusions in Chapter 9.

2 Review of Site-Response Models for Induced Seismicity

In this chapter, I will introduce the different ground motion models used throughout the history of modeling the seismic activities in Groningen. Before I go over the models, I take a brief excursion to formalize some notation about how I will be representing an earthquake as a mathematical object.

2.1 Earthquake data

Both for the context of simulating earthquakes and for data analysis on them, we need to know what kind of data we expect from an earthquake. As explained in the Introduction, very important quantities for earthquake engineering are the spectral accelerations, also often called the response spectrum. In theory, these are a continuous spectrum of values, but we represent them as a discrete set of points spaced such that they capture the most important frequencies that buildings are vulnerable to. Firstly, note that, rather than working with frequencies, earthquake engineers tend to work with periods, so we will be staying true to this convention. To that extent, for the Groningen ground motion model, we consider 23 periods ranging from $0.01s$ to $5s$. What periods exactly are used are of little relevance to the mathematical formulations, so we just consider periods with an index from 1 to 23, which will make the notation simpler. So in general, we will be referring to $i \in \{1, ..., 23\} := \mathbb{I}$, rather than to period $0.01s$, for example.

Subsequently, we define a ground motion measurement $\mathcal{E}$ as

$$\mathcal{E} := (\mathbf{S}, \mathbf{A}, x, y, M, R, z),$$

Here, $\mathbf{S} \in \mathbb{R}_+^d$ is the vector of spectral accelerations at the reference depth and $\mathbf{A} \in \mathbb{R}_+^d$ is the vector of amplification factors, i.e. A_i is the factor by which S_i is amplified at the surface. We restrict them to be positive because they are positive quantities by nature. The variables x and y specify a coordinate at which the measurement took place, $M \in \mathbb{R}$ specifies the magnitude of the earthquake that generated the signal, and $R \in \mathbb{R}_+$ the distance from the hypocenter of the earthquake (or sometimes the minimal distance to the rupture, because earthquakes tend to be shaped like a plane rather than a singular point). Lastly, z is the zone number where the measurement took place. As described in the Introduction, the Groningen zone is divided into 160 regions, plus two that consist completely of water, but we tend to not count those. All of these have a code consisting of 4 numbers, as in the plot of Figure 2, but for mathematical formulations, we will generally consider them as an index from 1 to 160, much the same as for periods. Since the mathematical formulations will all be symmetric w.r.t. the choice of mapping from index to zones, we do not prespecify a mapping.

2.2 History of Site Response Models for Groningen

Groningen has been modeled with a site response model since around 2017. The model has evolved quite a bit since then, but in its core has remained quite similar. For this reason, the easiest way to convey how to model works is by going through a few iterations

of the model, which gradually build up in complexity.

The responsibility for presenting the risk and hazard models associated with the extraction of gas from the Groningen field has been attributed to NAM. Consequently, NAM is the developer of the ground motion models (GMMs) under our current investigation. Over the years, they have developed model versions one through seven, denoted as GMMV1 to GMMV7.

2.2.1 GMMV2

The first GMM that is relevant to study is GMMV2 [8], which was based on the work of Dr. A. Rodriguez-Marek [9]. In this paper, he sets the foundations for the site-response model that is used to model the amplification functions to this day.

Given some earthquake $\mathcal{E}$, the goal of the models we discuss is to model amplifications A_i given the other information about the earthquake, the most important of which being the reference spectral acceleration S_i. Even though the model we are discussing is conditioned on this quantity being given, it is still important to understand how to interpret them, especially for later chapters. From the perspective of modeling the amplification factors, we can view S as

$$\ln(\mathbf{S}) \sim \mathcal{N}(\mathbf{m}, C), \tag{1}$$

i.e. S is a multivariate lognormal random variable with some mean vector $\mathbf{m}$ and covariance matrix C, although it is more commonly described in terms of its correlation matrix B. This correlation matrix B was suggested by Baker & Jayaram 2007 [10] and is now commonly used as a way to describe the correlation between spectral accelerations. Hence, when working with the reference model, we usually only work with the correlation matrix and marginal variances, which we call $\mathbf{s}$, such that $C = \mathbf{s}^T B \mathbf{s}$ These parameters are determined by other models part of the GMM. How this happens precisely is not of particular relevance; we can simply assume them as known constants.

Given this reference value, we can model the amplification of the spectral acceleration at period i as

$$A_i | S_i \sim \text{Lognormal}(\mu_i^z(S_i), \sigma_i^z(S_i)), \tag{2}$$

where S_i is the spectral acceleration at reference level for periods. μ_i^z and σ_i^z are functions of the random variables S_i, which give you the parameters of the conditional distribution. The superscript z is used to denote that the parameters are zone-specific. However, this superscript will mostly be dropped from now on, and it can be assumed that the parameters are zone-specific, unless stated otherwise.

The log-normal assumption is not specific to this model, but is very common in earthquake analysis. This is mostly because the earthquake data appears to confirm this, but one can imagine this lognormality originating from the subsurface being quite irregular and the surface values being an accumulation of many random features in the subsurface which either amplify the wave or attenuate it. If you have many independent sources of randomness multiplied together, you tend to get a lognormal random variable, much like the case where you add many sources of randomness, you tend to get normal distributions. Our knowledge of the subsurface is simply not complete. Although the behavior of a wave

traveling through the ground is deterministic, the knowledge of the ground is not precise enough to get an accurate idea of what a nearby earthquake will do to a specific building at a specific location. Therefore, the effects are described as random with a certain set of parameters for a specific region. These regions can be large or small, depending on the choice of model.

For Groningen, as we have seen, the area is divided into 160 zones based on the consistencies of the soil profile, which will each produce a different set of parameters. The GMMV2 describes specific functions for μ_i and σ_i for the conditional distribution of amplification factors A_i, given some reference spectral acceleration S_i. For μ_i we have

$$\mu_i(S_i) = a_i + b_i \ln\left(\frac{c_i + S_i}{c_i}\right) \tag{3}$$

and for σ_i we have the piecewise function

$$\sigma_i(S_i) = \begin{cases} \sigma_{1,i} & \text{for } S_i \leq \alpha, \\ \sigma_{1,i} + (\sigma_{2,i} - \sigma_{1,i}) \left(\frac{\ln(S_i) - \ln(\alpha_i)}{\ln(\beta_i) - \ln(\alpha_i)}\right) & \text{for } \alpha_i < S_i \leq \beta_i, \\ \sigma_{2,i} & \text{for } S_i > \beta_i. \end{cases} \tag{4}$$

In this model, we have 7 parameters per zone/period pair, which are $a_i, c_i \in \mathbb{R}^+$, $b_i \in \mathbb{R}$, $\alpha_i, \beta_i, \sigma_{1,i}, \sigma_{2,i} \in \mathbb{R}^+$. Rodriguez-Marek [9] proposes fixing some of these parameters to be constant across all zones and pre-specifying their values to reduce the number of free parameters. In Rodriguez-Marek [9], values of c, α_i, and β_i are given, leaving 4 free parameters to be estimated across each zone/period. The values of these parameters can be found in Appendix 1 in Table 2.

Groningen is divided into 160 zones, and the number of periods considered is 23. Therefore, the GMMV2 [8] effectively describes $23 \cdot 160 = 3680$ different random variables $A_i|S_i$ with their unique set of parameters.

Each of these parameters has a specific impact on the model. The parameters a_i, b_i and c_i determine the curve of the means around which the realizations of A_i will be centered. For values of S_i close to 0, the curve starts at a constant value of a_i and remains approximately constant until S_i gets close to the value of c_i, at which point a non-linear regime starts to emerge, which can slope both upward and downward depending on the sign of b. b will determine how strong the amplification or attenuation effect is. As stated above, c_i are fixed and were chosen by Rodriguez-Marek [9]. According to him, these parameters 'were set to constant field-wide values after a careful examination of preliminary regression results.' However, he does not share the exact methods.

The parameters α_i and β_i denote the points in the piecewise function for the standard deviation σ_i between which the standard deviation goes from $\sigma_{1,i}$ to $\sigma_{2,i}$. As stated above, α_i and β_i are also not free parameters and were chosen by Rodriguez-Marek [9] without a description of the procedure with which these values were determined.

This model aims to model an important characteristic of high-magnitude earthquake waves as they travel through soft soils: the non-linear breakdown point. This occurs when the energy of the waves can no longer be transmitted linearly through the subsurface materials. At this threshold, the intensity of the wave disrupts the usual propagation,

causing the energy transfer to become non-linear. This effect is a factor in understanding how earthquake waves are transferred through soft soils.

2.2.2 GMMV6

After the previous model GMMV4, GMMV5, and GMMV6 followed (GMMV3 was never published and was only used internally). They were all similar in the way they handled amplification functions. Therefore, we simply skip to the latest one.

The largest change in these models compared to the GMMV2 is the added dependence on magnitude and distance for the parameter a. To be precise, assume that we have an earthquake $\mathcal{E}$ in some zone z and we are considering a certain period i that is less than $0.30s$ (we again drop the indices for i for brevity). We now have

$$a(M, R) = a_0 + a_1 \log(R) + (b_0 + b_1 \log(R))(\min(M, M_{ref}) - M_{ref}) \tag{5}$$

and M_{ref} is defined as

$$M_{ref} = M_1 - \frac{\log(R) - \log(3)}{\log(60) - \log(3)}(M_1 - M_2).$$

This adds six new parameters to deal with. We have $a_0, b_0, M_1, M_2 \in \mathbb{R}^+$ and $a_1, b_1 \in \mathbb{R}$. If we set all of these to be 0 except a_0, we simply retrieve the GMMV2, which is what is done for periods of 0.3 and above. Otherwise, we let all of these be free parameters, except M_2. M_2 was chosen to be 4.5, so as to center the model in the right place [3].

This addition will add an additional dependence on M and R to the model. To be precise, we add a term scaling with the logarithm of the distance, as well as a term that scales with small magnitudes. It is important to realize that the previous model was completely independent of magnitude and distance, which may sound strange, but the nuance here is that the *conditional* distributions were specified and, obviously, the reference values are strongly dependent on magnitude and distance. However, Stafford et al. [11] found that there is a linear relationship between magnitude and distance for low periods. Therefore, Bommer et al. included these findings in GMMV5-6 [2], [3]. Note that this model is applicable for magnitudes up to 7.25 and distances up to 60 km.

In addition to this change, lower and upper bounds were also placed on the mean functions μ. For this, two parameters A_{min} and A_{max} were chosen. A_{min} is always set to 0.25 and A_{max} varies between periods and zones. For A_{max} to be relevant in the first place, we need a positive value for f_2, i.e. there needs to be an amplifying effect due to nonlinear breakdown. According to Bommer et al. [1], this is mostly relevant for longer periods, say longer than $1.5s$. We will later consider periods between $0.01s$ and $1s$, so these are not particularly relevant for us. For more information on how to calculate these, see GMMV4 [1].

2.2.3 GMMV7

GmmV7 introduced even more changes to the linear part of the model. In this version, we have

$$\begin{aligned} a(M,R) = a_0 + a_1\log(R) + (b_0 + b_1\log(R))(\min(M, M_{ref}) - M_{ref}) + \\ a_2(\ln(R) - \ln(R_{ref}))^2 + b_2(\min(M, M_{ref}) - M_{ref,2})^2 + \\ a_3(\max(M, M_{ref}) - M_{ref}). \end{aligned} \tag{6}$$

The extra terms add a quadratic dependence on magnitude and distance, as well as a term that scales with large magnitudes. The report on GMMV7 by Bommer et al. [4] does not mention any restrictions in the domain on these parameters. However, it specifies that R_{ref} is a constant equal to 10.25 and that $M_{ref,2}$ is 3.2. The other parameters are described as free, although from the parameters supplied with the report, it appears that the parameters a_1 and a_2 are 0 for periods greater than 0.5.

2.3 Period-to-period correlations for amplification factors

Recall from Eq. 1 that for the reference model, we specified a multivariate lognormal model with a given covariance matrix, but for the amplifications, only marginal distributions were specified; see Eq. 2. However, it is an important question to ask whether these should depend on each other in some way, because the dependence structure of A will heavily influence the distribution of the surface spectral acceleration vector $S \odot A = A_{surf}$, where $\odot$ is the operator for element-wise multiplication of vectors. As will be revealed later, this quantity is important for risk calculations, so it is important to model the right dependence structure for A.

In terms of what the ground motion model reports prescribe, it is rather unclear. In V6, it appears to implicitly suggest a model where all correlations between periods are 0, i.e., full conditional independence between different amplification factors, given their corresponding reference ground motions [3, p. 74]. In contrast to this, in V7 the same authors appear to suggest a model with perfect positive correlation, i.e., each period has the same residuals for their uncertain components [4, p. 172]. This sudden change in model, without justification behind why it is being modeled this way, has left TNO to question the views behind these sudden changes and has sparked a conversation between the two parties [12, Appendix C].

To this day, this confusion remains unresolved. In Chapter 5, I will look into what dependence structure the data suggests. Note the nuance here that we do not assume a priori a multivariate lognormal distribution, even though this is prescribed by the model authors [4].

2.4 Fragility modeling

The fragility model is the component of the Groningen model chain responsible for predicting the damage done to buildings, given some ground motion. Understanding how this works is quite crucial to understand my contributions regarding the model computations in Chapter 6. This is because the spectral accelerations S_{surf} that the ground motion generates are fed directly into the fragility model. Hence, if we want to make the ground

motion simulations more efficient, we need to understand exactly what is needed to apply the fragility model.

The following description is based on the in-depth explanation of H. Crowley in [13] and is a paraphrasing of her explanation in my own words. The main idea of the fragility model is that you give it some predictor for certain damage states of a building, which we call an intensity measure (IM). Given the intensity measure, it will tell you with what probability a certain building will be in that damaged state. This can be any imaginable quantity and for now, we will be treating these in full generality and hence should not be linked to the previous sections. The model can take one or multiple intensity measures. We will explain the choice of different types of IM in more detail in the next subsection.

Next, we also assume that we have models for different building types. For context, this could be a building with unreinforced masonry within a certain range of dimensions. In the latest Groningen fragility model, we have data for 35 different building types, ranging between different types of brick houses, or houses with unreinforced masonry or farmhouses, and we have knowledge of their locations within the area. For each of these buildings, we also have the so called 'displacement limits'. These represent a hard limit where if a building experiences a displacement greater than this number, the building will receive the damage type corresponding to this displacement limit. Multiple types of damage are possible, ranging from non-structural damage to collapse.

Intensity measures are used to predict the 'expected displacement' that would occur in a certain type of building, given those intensity measures. This goes as follows.

Definition 1 (Distribution of displacements)**.** Let D be the displacement response and let $\{I, b\} \in \mathbb{R}^d$, where I is the vector of intensity measures and b is a coefficient vector and $b_0 \in \mathbb{R}$. Then we define

$$\ln(D)|I \sim \mathcal{N}(\ln \eta_{D|I}, \beta_s),$$

where

$$\ln \eta_{D|I}(I) = \mathbb{E}(\ln(D)|I) := b_0 + b \cdot \ln(I) = b_0 + b_1 \ln(I_1) + ... + b_m \ln(I_m),$$

and

$$\beta_s = \sqrt{\beta_R^2 + \beta_{BB}^2 + \beta_{D|I}^2}.$$

Here, β_R and β_{BB} are predetermined model parameters, and $\beta_{D|I}$ is a variance specific to the building type.

There are several components to this model that we will briefly discuss now. Firstly, we allow a vector of intensity measures I to be a predictor. Whether to use one or multiple IMs is left open. Then, if using multiple IMs, it combines them in a hyperplane representation. This makes it such that the fragility model will work best for IMs that have an approximately linear relationship with the displacement. Instead of using a hyperplane, one could consider using other techniques, such as machine learning methods like decision trees or neural networks, as was done by Q. Kong et al. [14].

To use this model, it is necessary to have a dataset to which one can calibrate the parameters. An example of such a calibration can be found in Figure 4, where a data

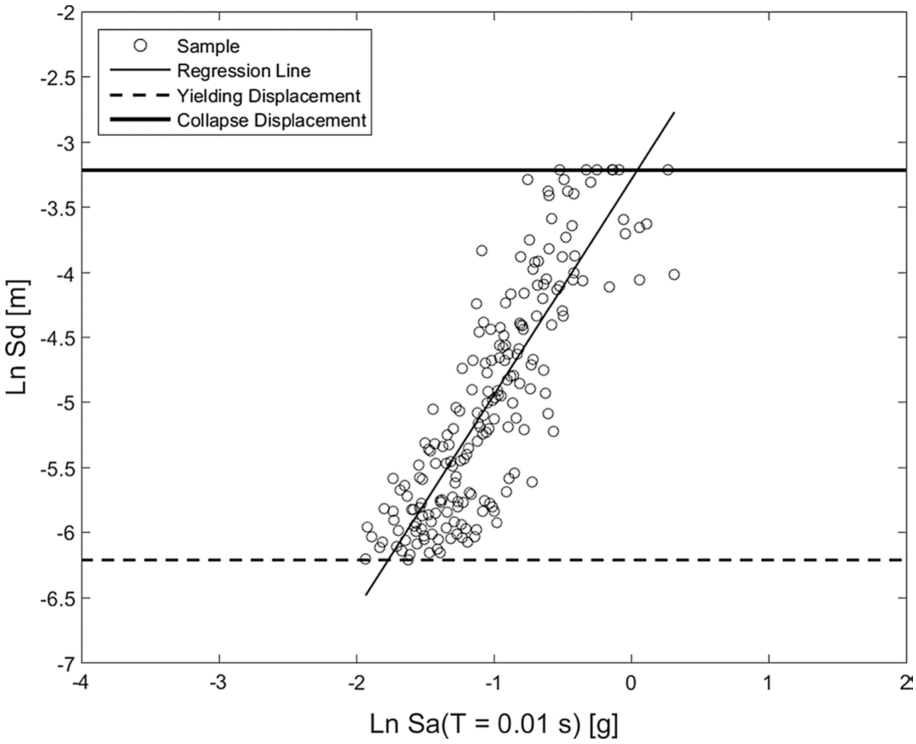


Figure 4: Example of a calibration data set for the calibration of an IM. On the x-axis we see the log of the spectral acceleration of period 0.01s being calibrated to the log displacement. Source: [13, p. 4505]

set of spectral accelerations of period 0.01s is shown with a best-fit regression line. Obtaining such a dataset is a non-trivial task. To do this, specialized structural engineering software is used to model how different earthquake signals cause certain displacements of a particular building type. The computational load is quite intense for these simulations, hence the reason that in the end, the shaking is modeled using a statistical model rather than a physical model. We will not go into how these are generated precisely. For now, simply assume that we have a data set with displacements $d_1, ..., d_n$ and for each displacement we know the intensity measures $I_i, i \in \{1, .., n\}$ that resulted in the building having this displacement. Then obtaining coefficients $b_0, ..., b_m$ can be done with a simple linear regression on the logarithms of the displacements. Then, with the same data, we determine

$$\beta^2_{D|I} = \frac{\sum_{i=1}^{n} \left[\ln(d_i) - \ln \eta_{D|I}(I_i)\right]^2}{n - (m+1)}, \tag{7}$$

which is the unbiased estimator for the variance of data with n data points and $m+1$ degrees of freedom. Lastly, β_R and β_{BB} represent, respectively, record-to-record and building-to-building variability. These can also be assumed to be known constants.

The way this model is now put to use is to calculate the probability of exceeding a certain displacement limit P_L, where L is the point where if $d \geq L$, then the building will experience some type of damage. Note that different kinds of damaged states exist, each with their own limit, as well as that all different building types have different thresholds L. The probability is then easily calculated because we defined the log-displacement as a

normal distribution earlier, so

$$P_L = 1 - \Phi\left(\frac{\ln(L) - \ln \eta_{D|I}(I)}{\beta_s}\right). \tag{8}$$

These probabilities are a fundamental component of the Groningen model chain. Together with the ground motion model, we can now answer the question 'what kind of damage can we expect if we have an earthquake of magnitude M and location (x, y)'. As you can see, the main remaining components that these risks are conditioned on are magnitude and location, so if you now integrate this model over another model for magnitude distributions and a spatial model of where we expect earthquakes, we can calculate aggregate risks over the entire region. We will not discuss the other components of the chain in detail. In order to estimate the impact of changes to the ground motion model that will be discussed in this thesis, we will, however, use the other parts of the chain. This will be essentially a black box, but since we use the same black box for all ground motion models we use as input, we can still use this as a way to measure the real world impact of the changed model.

2.4.1 Intensity measures

In the fragility model, as explained above, we have said nothing about the choice of intensity measure. The model is made in such a way that it can accept any possible set of inputs as intensity measures (IMs). It is completely up to the designer of the fragility model to determine what IMs are the most suited. One can imagine that picking things that have a higher correlation with displacement of the building, will offer more predictory value than ones with lower. For example, one can imagine that the magnitude of the earthquake will be a better predictor than, e.g., the time of day, even though both are valid IMs. The fragility models of the Groningen model chain have gone through numerous iterations. The most effective IMs have proven to be the spectral accelerations, which is why the ground-motion model focuses so heavily on these. [13]

For reference, we will briefly discuss some previous choices of IMs in older versions of the Groningen fragility model. In the paper explaining the basis of the fragility model that I explained above [13], H. Crowley, the lead scientist behind the Groningen fragility model developed by NAM, also explains the IMs used at that point in time (2019). She explains that to choose the right intensity measure, one needs to look at the most efficient ones (lowest standard deviation according to Eq. (7)) and select the ones with a low Pearson correlation with each other. The latter will ensure that you are not using 'the same information' in multiple IMs. She researched which IMs would be appropriate to use, and found that the ones that appeared to fit these criteria best were to use the spectral accelerations of two specific periods S_{surf,T_1} and S_{surf,T_2} and the metric called the 5-75% significant duration of the earthquake $D_{S5-75\%}$, which is the time between which 5% to 75% of the total energy of the earthquake's ground motion is accumulated. The choice of the two periods is such that it tries to capture the fundamental frequencies of the building type optimally. Usually, this results in two frequencies not too close together, as the closer they are, the higher their Pearson correlation will be, and the lower the predictive power of the fragility model will be overall.

2.4.2 Average spectral acceleration $\overline{S}$

A more recent development in the choice of IMs is to combine all the spectral accelerations into one predictor in the form of the average spectral acceleration. This quantity is defined as the average surface spectral acceleration over 10 periods. To be precise, we consider the periods $0.01s, 0.1s, 0.2s, 0.3s, 0.4s, 0.5s, 0.6s, 0.7s, 0.85s$ and $1s$ and to get the average spectral acceleration, we take a geometric mean in the linear domain or an arithmetic mean in the log domain. In general, we will be working with quantities in the log domain, because it is generally more readable. Therefore, we define the logarithm of the average spectral acceleration $\overline{S}$ as

$$\overline{Z} := \ln(\overline{S}) = \frac{1}{10}\sum_{i=1}^{10} \ln(S_{surf,i}), \tag{9}$$

where $S_{surf,i}$ is the surface spectral acceleration at period i. Using this as an IM will simplify the fragility model such that $\ln \eta_{D|I}$ from Definition 1 simplifies to

$$\ln \eta_{D|I} = b_0 + b_1 \overline{Z}.$$

Using this as the only IM has shown better performance than any single spectral acceleration component as IM [15]. To my current understanding, there is no clear mathematical explanation why this should be the case; however, empirically it has shown itself to be highly effective. Using this as an IM has the major advantage of being universal: it is a metric that can be used for all buildings and will generally have sufficient predictive power to estimate damage states.

We can use this universality to our advantage in the model chain. Rather than simulating all the ground motions that make up the average, we estimate the average directly. This is one of the main results of this thesis and will be discussed in Chapter 6.

Although $\overline{S}$ appears to be the best performing single IM, it can still be improved by combining it with other IMs. For example, research by E. Bojórquez et al. shows that in their context of steel frames on soft soils, it was effective to use $I = \left(\frac{\overline{S}}{S_1}, S_1\right)$ [16]. However, changing the fragility model is not the focus of this thesis.

3 Data Description

In this chapter, we study the data used for model calibration and model validation. Firstly, we will go through a brief explanation of how the data is obtained and, afterwards, we will give statistical insights into the data. In this process, the models will become more intuitive and you will get visual insights in how the site amplification works.

3.1 STRATA dataset

The dataset that is used for the calibration of the Groningen GMMs (except V7) is called the STRATA dataset, provided by the research institute Deltares [17]. The reason we call it the STRATA dataset is that it was created with software called Strata [18]. Although it is used to calibrate the GMM, it does not contain real ground-motion data. All the data in it has been generated by computed simulations. This is because the phenomenon it is trying to describe, i.e. the non-linear breakdown point of the soil, has never been actually measured in the Groningen subsurface as a consequence of an earthquake. This is because the magnitude required to reach the energy level of a breakdown has not yet occurred in the region. The models used to generate these are more physical in nature, and the known soil profiles of the area are used to generate the data. Therefore, this is the best we can do in terms of predicting the effect of larger-magnitude earthquakes.

The data set has 1.405.010 data points, divided between the 160 relevant zones. Some zones have around 500 points, some zones have in the order of 20.000 points, which has to do with the geographical area of the zone.

An obvious question regarding the use of this dataset is: If we have a way to generate data based on the geology of the region, why not use this model directly in the risk calculations? The answer to this lies in the computational complexity to create data points like in the STRATA dataset. It is many orders of magnitude more complex. Therefore, we use the STRATA data to calibrate a summary model that is probabilistic in nature and is easy to sample.

3.1.1 Examples of STRATA data in multiple GMMs

To get a sense of how these data look, we will study the plots in Figure 5. Here we see 10 plots of different periods for an arbitrarily chosen zone in the STRATA dataset. In each plot you can see a point cloud of data points. Each of these represents a simulated earthquake signal with some coordinate, magnitude, and distance that has arrived in this zone. This same earthquake also appears in the other periods' plots, but it is impossible in this way of visualizing to link them together, which is not really an issue because we do model inference separately per period. They are plotted with the reference spectral acceleration on the x-axis and the amplification on the y-axis, so we can observe the functional relationship between the two. Next, we also have the red line, which is the mean function according to the GMMV2, and we also have the 95% confidence interval around it. You can get a sense for how well the model fits by looking at these zones: approximately 95% of the data points should fall into this zone. Furthermore, certain parts of the zone being unpopulated is also a bad sign, as there should have been data points there if the model were a perfect fit. Lastly, we have the green dashed lines that are the points α_i and β_i,

that is, the points between which the variance goes from $\sigma_{1,i}$ to $\sigma_{2,i}$.

The way in which the probabilistic part of the model should be interpreted is as follows: at each point on the red line, we expect to see data above and below that point spread according to the lognormal distribution. Note that both axes are on a logarithmic scale, so that means that the spread on the plot will resemble the normal distribution, hence the symmetry in the confidence intervals.

We can also clearly observe that the mean function has a linear part (governed by the parameter a_i) and then an intermediate part where it breaks down, and finally a part where the behavior is logarithmic (which looks linear on the logarithmically scaled x axis).

We can now compare this with how the V6 model looks. We have the same plot of the same data but now with a visualization of the V6 parameters around it; see Figure 6. Observe that now the mean is no longer a single curve, but can vary for each earthquake depending on distance and magnitude. It is hard to see which orange dot corresponds to which blue dot in the earlier periods.

Besides this, we also did not show any confidence interval because they are no longer a curve, but rather vary around each orange dot separately. Because of this, we also show Figure 7, in which we have a filtered version of the previous plots, where we only include earthquakes of magnitude 5.5, making the data more concentrated. Because 5.5 is a relatively high magnitude, we see a lot of earthquakes in the nonlinear regime. Here we can also visualize the uncertainties in a relatively clear way.

3.2 Monte Carlo simulations of GMMs

The models as described in the previous chapter are very suited for Monte Carlo simulations, because of its two step nature and conditional formulation. It is easy to get a surface spectral acceleration by sampling a reference spectral acceleration and subsequently sampling an amplification factor conditioned on the reference value. Still though, the Monte Carlo simulations have one small caveat we will be going over shortly, and it also looks slightly different because of it being conditioned on magnitude and distance.

3.2.1 Site-to-site variability ϕ_{S2S}

In the process of obtaining the STRATA data, they cannot include certain epistemic sources of uncertainty. Rather than artificially try to put this into the data, they opted to leave this out and therefore calibration is done with different variances than those that are found by calibrating the model to the STRATA data. To be precise, when we calibrate the model for amplification A_i, we calibrate the function σ_i and its parameters $\sigma_{1,i}$ and $\sigma_{2,i}$. Recall that $\sigma_i(S_i)$ was described by Eq. 4. Now, we define the function $\phi_{S2S}(S_i)$ in a similar way as the piecewise function

$$\phi_{S2S}(S_i) = \begin{cases} \phi_{S2S,1,i} & \text{for } S_i \leq \alpha, \\ \phi_{S2S,1,i} + (\phi_{S2S,2,i} - \phi_{S2S,1,i}) \left(\frac{\ln(S_i) - \ln(\alpha_i)}{\ln(\beta_i) - \ln(\alpha_i)} \right) & \text{for } \alpha_i < S_i \leq \beta_i, \\ \phi_{S2S,2,i} & \text{for } S_i > \beta_i. \end{cases} \tag{10}$$

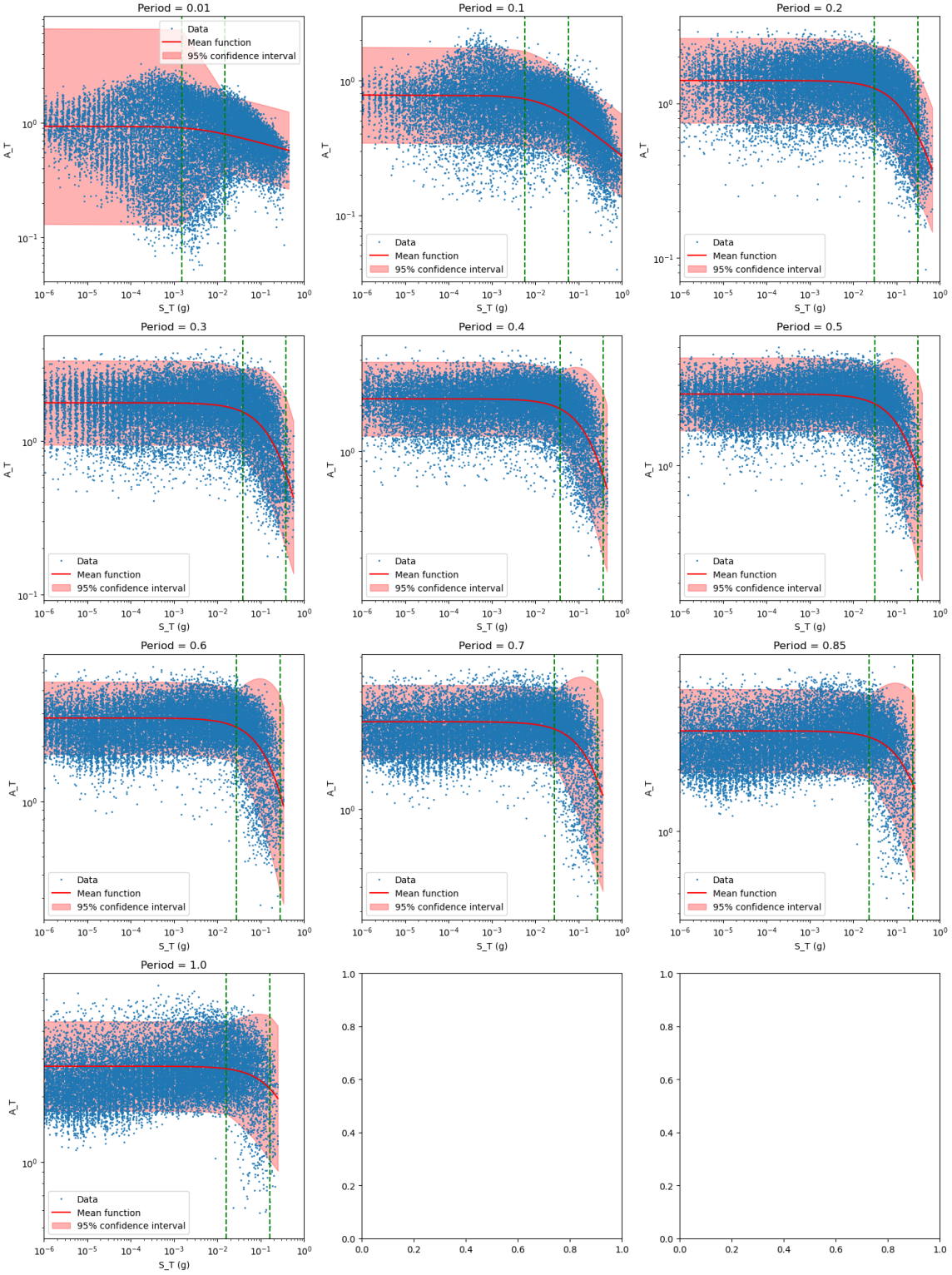


Figure 5: Plot of the data of zone 1011 from the STRATA dataset. Each point represents a certain simulated earthquake. On the x-axis we see the reference spectral acceleration of a certain period, as given in the title for each subplot, and on the y-axis we see the amplification factors in log-log scale. The red line represents the mean function according to the GMMV2 and around it we see a 95% confidence interval. The parameters used for the model are the maximum likelihood parameters for the plotted datapoints.

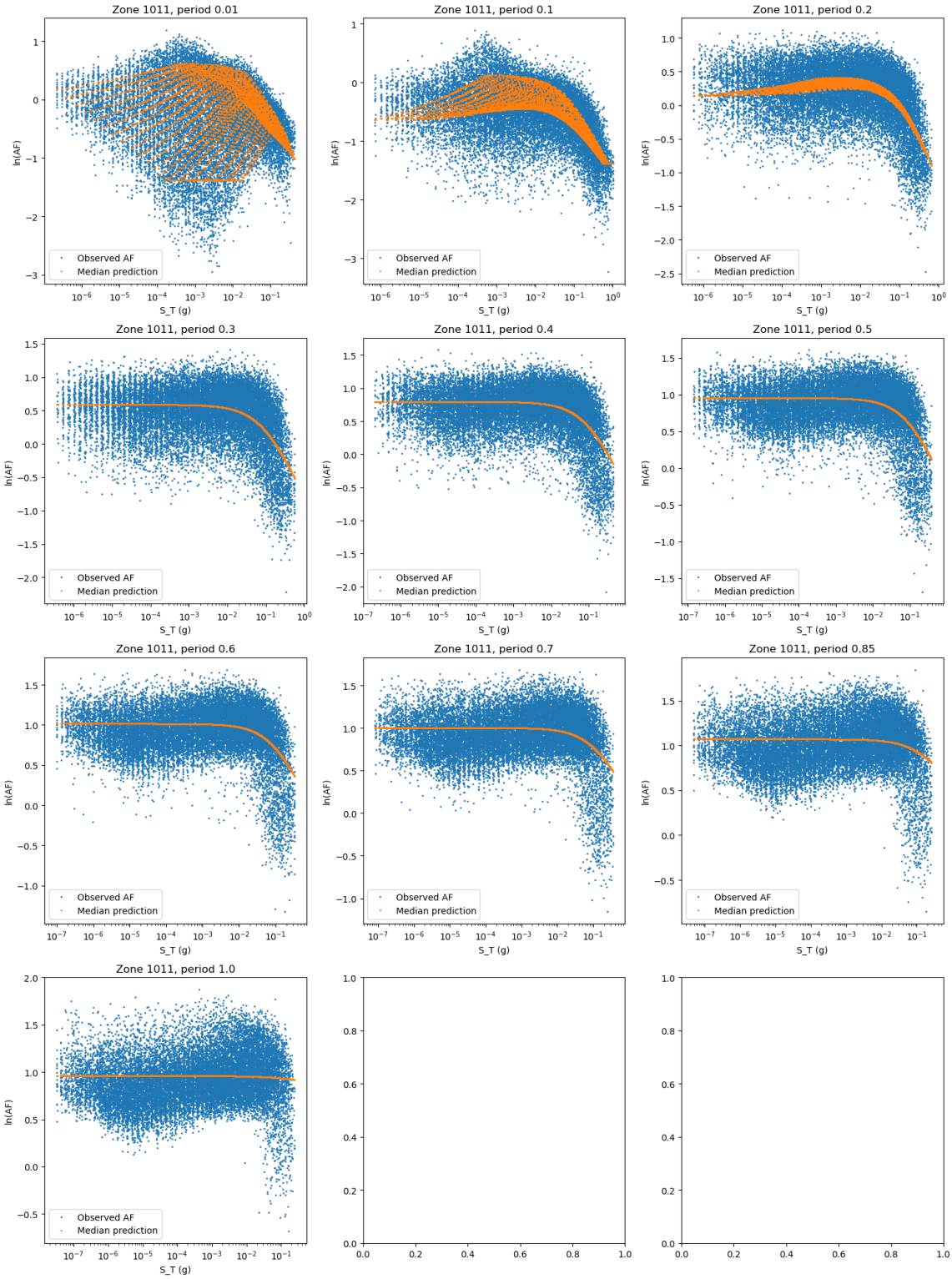


Figure 6: Plot of the data of zone 1011 from the STRATA dataset. Each blue point represents a certain simulated earthquake. On the x-axis we see the reference spectral acceleration of a certain period, as given in the title for each subplot, and on the y-axis we see the log-amplification factors. Each orange point represents the median value of a corresponding blue point either above or below it, according to the GMMV6 model. The parameters used for the model are the maximum likelihood parameters for the plotted datapoints.

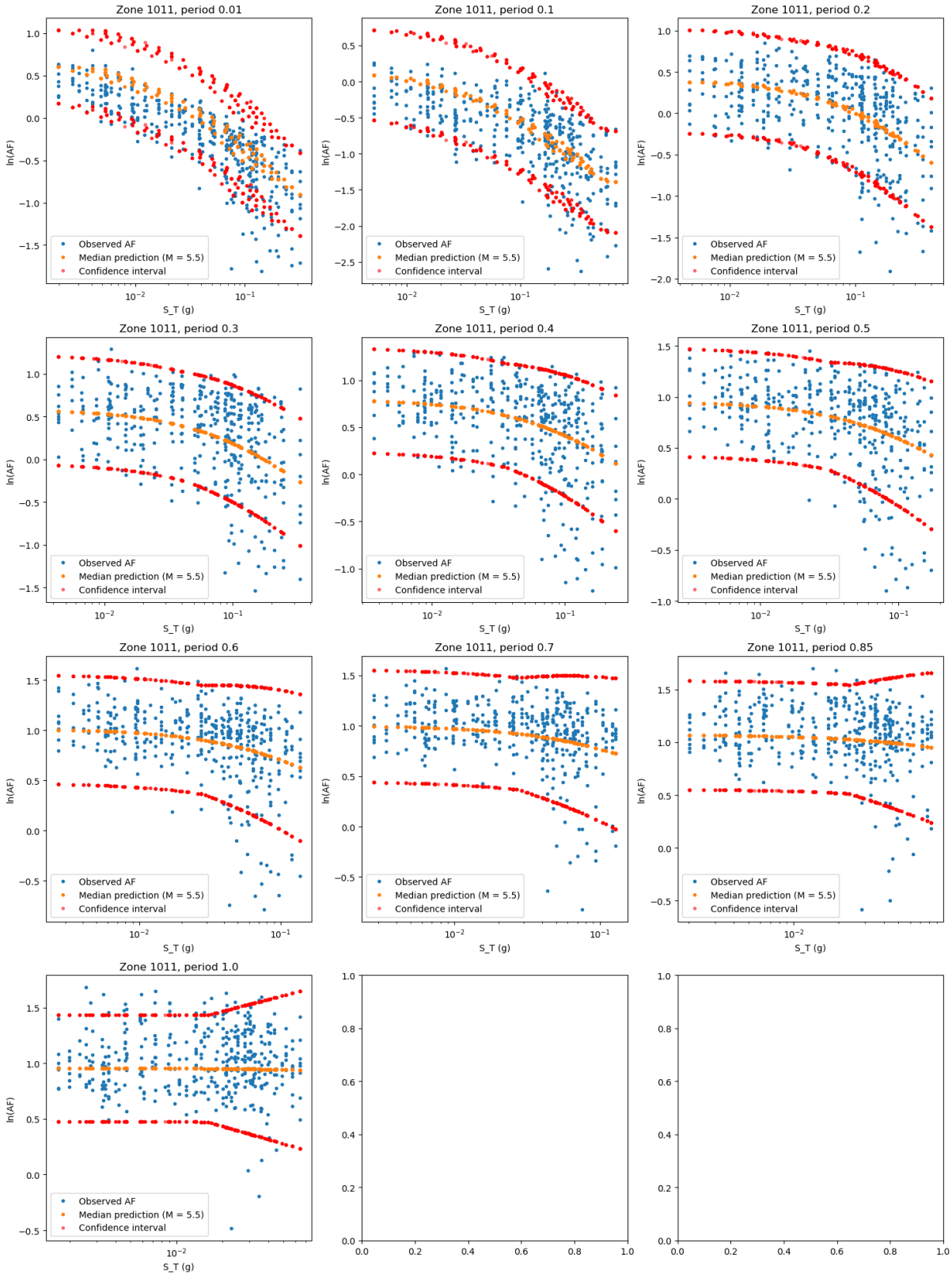


Figure 7: Plot of the data of zone 1011 from the STRATA dataset, filtered for earthquakes with magnitude $M = 5.5$. On the x-axis we see the reference spectral acceleration of a certain period, as given in the title for each subplot, and on the y-axis we see the log-amplification factors. Each orange point represents the median value of a corresponding blue point either above or below it, according to the GMMV6 model. Each orange point also has two corresponding red points representing the 95% confidence interval. The parameters used for the model are the maximum likelihood parameters for the non-filtered dataset.

Here, the parameters α_i and β_i are the same as in $\sigma_i(S_i)$. The parameters $\phi_{S2S,1,i}$ and $\phi_{S2S,2,i}$ can be found as the sum of the two sources of uncertainty: the previously discussed $\sigma_i(S_i)$ and the epistemic uncertainty σ_{MRD}, which is constant for all periods, but is different per zone. This uncertainty is further subdivided into $\sigma_{MRD,1}$ and $\sigma_{MRD,2}$ such that

$$\phi_{S2S,1,i} = \sqrt{\sigma_{1,i}(S_i)^2 + \sigma_{MRD,1}^2},$$
$$\phi_{S2S,2,i} = \sqrt{\sigma_{2,i}(S_i)^2 + \sigma_{MRD,2}^2}.$$

The epistemic uncertainties are subscripted MRD, which stands for modulus reduction and damping curves. This the part of the data generation that is responsible for the epistemic uncertainty. For a deep explanation regarding this source of uncertainty, I refer to Rodriguez-Marek [9]. Secondly, to obtain the MRD uncertainties, I refer you to GMMV4 [1, pp. 214-216]. Their calculation is of no significant importance to this thesis. We will simply use ϕ_{S2S} whenever we sample the model and use σ_i when doing calibrations.

3.2.2 Example of a simulation

To visualize how the model looks from the simulation side, we will pick a scenario and work through the steps. The scenario I am choosing is a magnitude of 4.8 at a distance of 7km. This is a medium length distance with high magnitude, so we expect quite high spectral accelerations, but these values are no particular significance. Next, for the reference spectral acceleration, there are a few branches in the model to choose from. I will choose the ones which yield the highest median, i.e. some sort of 'worst case scenario'. We will for simplicity restrict ourselves to a single period, 0.1 seconds, for demonstration purposes. We can now use the parameter files given as an attachment model V6 [3] to get the function parameters for the mean of the amplification function and its uncertainties. This is plotted in Figure 8. I chose to plot the logarithm of S and A for simplicity. In this demonstration, you can clearly see that the amplification mean is bounded from below by A_{min} as described in the model V6 section.

In this scenario, the reference model gives us that the reference spectral accelerations $\ln(S)$ is distributed as $\ln(S) \sim \mathcal{N}(\mu = -1.75, \sigma = 0.394)$. In Figure 9 you can see a demonstration of how the reference model fits into the amplification model. Here we see the density of the reference model, plotted on separate axes. The x-values of the points we simulate will be sampled according to this distribution. We then will samples a different normal distribution, determined by the amplification model, which will determine the vertical displacement from the blue line.

We can now sample reference spectral accelerations from this normal distributions and compute their amplification means and variances. Then we can sample those distributions once again to get the simulated amplification values. In Figure 10, we can see an arbitrarily sampled 1000 points according to this procedure.

Observe that the points are horizontally spread out according to the reference model, and vertically spread out according to the amplification model. This makes it so the point cloud appears stays clustered around the blue line. To demonstrate this effect even further, and to generate a plot that looks a bit more similar to the STRATA data from

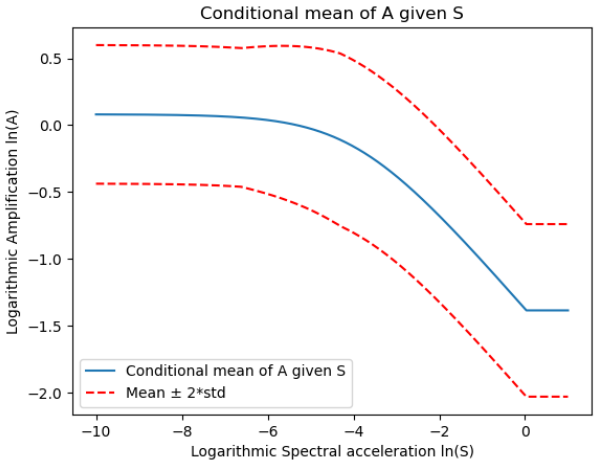


Figure 8: Amplification mean and standard deviations for a magnitude of 4.8 at a distance of 7km. On the x-axis we see the log reference spectral acceleration S and on the y-axis we see the amplification factors A. The blue line represents the mean function w.r.t. the reference spectral acceleration.

which the parameters were derived, we can shift the reference distribution a bit and add more variance. For the simulation in Figure 11, the reference model was set to have $\ln(S) \sim \mathcal{N}(-4, 2.4)$.

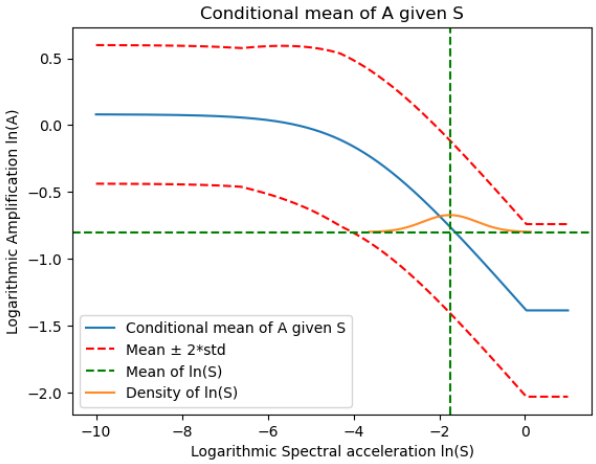


Figure 9: Amplification mean and standard deviations for a magnitude of 4.8 at a distance of 7km. On the x-axis we see the log reference spectral acceleration S and on the y-axis we see the amplification factors A. The blue line represents the mean function w.r.t. the reference spectral acceleration. Furthermore, we have plotted the density of the reference model, scaled on its own axes to the place where we expect the amplification factors to be.

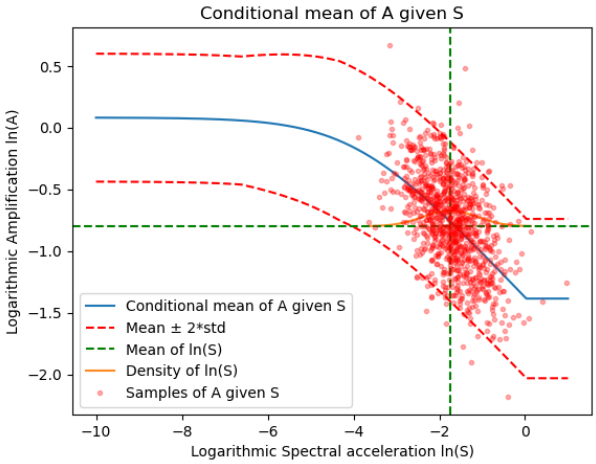


Figure 10: Amplification mean and standard deviations for a magnitude of 4.8 at a distance of 7km. On the x-axis we see the log reference spectral acceleration S and on the y-axis we see the amplification factors A. The blue line represents the mean function w.r.t. the reference spectral acceleration. Furthermore, we have plotted the density of the reference model, scaled on its own axes to the place where we expect the amplification factors to be. The red point cloud shows the values from a simulation of the site amplification model.

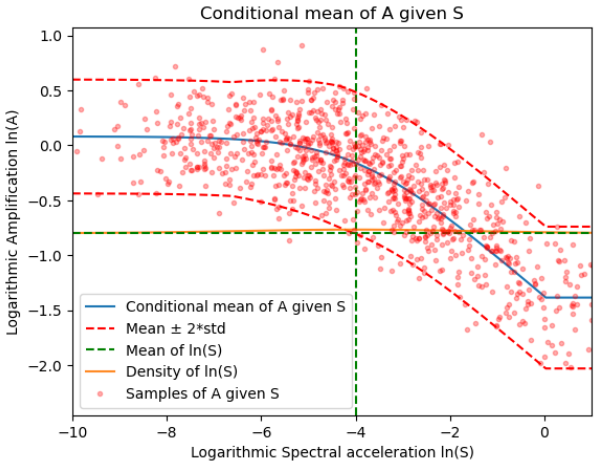


Figure 11: Simulation of the amplifications in an artificial scenario with inflated variance. On the x-axis we see the log reference spectral acceleration S and on the y-axis we see the amplification factors A. The blue line represents the mean function w.r.t. the reference spectral acceleration.

4 Model Calibration through Parameter Estimation

Throughout models GMMV4 to GMMV7, the communication around how the parameters they find are calculated is unclear. They talk mainly about regression and fixing parameters to certain values and then tuning the other ones. To be precise, if I understand the reports correctly [1]–[4], what they do is sequentially find the optimal value of a certain parameter according to some metric (which is not explained). Then they go to another parameter and optimize it to the same metric, where they now have the temporary optimum for the other parameter. Then they repeat this procedure until all parameters have been optimized a number of times. The number of times is also not explicitly stated.

This procedure seems very prone to getting stuck in local minima and in cycles, especially when optimizing a lot of parameters at the same time. Therefore, we decided to do a calibration of the model with maximum likelihood estimations in order to find parameters that fit the model better.

Note that we purely look at optimizing the parameters from a mathematical perspective. I do not claim that the other parameters are wrong, as the developers of the model undoubtedly have a lot of expert insight which I lack. On the other hand, since I have no way to replicate their method due to lack of explaining it, I found it interesting to see what a maximum likelihood analysis finds as optimal parameters.

4.1 Maximum likelihood estimation

To find suitable parameters for this model, we will use maximum likelihood estimation. The idea behind this is that we find the set of parameters that maximizes the probability of finding the particular set of observations that we have. Considering some particular period i (we drop the index for brevity), the probability of finding a particular observation under a particular set of parameters is given by

$$\mathbb{P}(A = y|S; a, b, \sigma_1, \sigma_2) = \frac{1}{A\sqrt{2\pi}\sigma(S; \sigma_1, \sigma_2)} \exp\left(-\frac{(\ln(y) - \mu(S; a, b))^2}{2\sigma(S; \alpha, \beta, \sigma_1, \sigma_2)^2}\right), \tag{11}$$

where μ and σ are defined as in equations 3 and 4 and parameters c, α and β are determined by the period i and given in Rodriguez-Marek [9]. The likelihood of finding $N \in \mathbb{N}$ observations $X := \{i \in \{1, ..., N\} : (S_i, A_i)\}$ with a certain set of parameters is given by

$$L(a, b, \sigma_1, \sigma_2; X) = \prod_{i=1}^{N} \frac{1}{A_i\sqrt{2\pi}\sigma(S_i; \sigma_1, \sigma_2)} \exp\left(-\frac{(\ln(A_i) - \mu(S_{T,i}; a, b))^2}{2\sigma(S_i; \alpha, \beta, \sigma_1, \sigma_2)^2}\right). \tag{12}$$

Maximizing this function can be numerically unstable, hence we apply a log-transformation, which increases the stability dramatically. We then get

$$\ln(L) = -\sum_{i=1}^{N} \left[\ln(A_i) + 0.5\ln(2\pi) + \ln(\sigma(S_i)) + \frac{(\ln(A_i) - \mu(S_i))^2}{2\sigma(S_i)^2}\right]. \tag{13}$$

Finding a maximum is usually done by finding a minimum to the negative of this function because many programming languages have efficient implementations for minimization problems. The minimization can be done in a fast and computationally efficient way by

using a gradient-descent method such as the *L-BFGS-B* method [19]. We can speed this procedure up dramatically by calculating an analytical expression for the gradient, which can be done manually or with symbolic math packages like SymPy. This maximum likelihood estimation has to be done for 3680 of these functions, i.e. for each period in each zone.

4.1.1 Extension to GMMV6 and GMMV7

Extending this procedure to the newer models is readily done. It is merely a matter of updating the way you calculate the mean function, given that this is the only thing that changes throughout the models. The likelihood function has stayed largely the same:

$$L(a(M,R),b,\sigma_1,\sigma_2;X) = \prod_{i=1}^{N} \frac{1}{A_i\sqrt{2\pi}\sigma(S_i;\sigma_1,\sigma_2)} \exp\left(-\frac{(\ln(A_i)-\mu(S_i;a(M,R),b))^2}{2\sigma(S_i;\alpha,\beta,\sigma_1,\sigma_2)^2}\right),$$

where $a(M,R)$ is as in Eq. 5 for GMMV6 and as in Eq. 6 for GMMV7. Although it is computationally more complex because of the increased number of parameters, estimating these parameters can be done in the same way as was done for the simpler version of the model, where you need to be careful in selecting the right selection of free parameters.

The minimization procedure was done in the same way as for the simpler model. Here it is important to stay critical and make sure the points you find are not saddle points, because this can happen when optimizing large numbers of parameters, especially if the likelihood is weakly sensitive to some parameters.

4.2 Comparing maximum likelihood estimates to given parameters

Comparing how different sets of parameters perform for a model can be done in multiple ways, but the obvious way, given that we are already in the maximum likelihood framework, is to use the likelihoods. More specifically, a common metric to compare different parameters is to calculate likelihood ratios, or the logarithm of this ratio.

We will be analyzing three sets of parameters. Firstly, the parameters supplied with the V6 model [3]. We call these *BommerV6*, named after the main model author J. Bommer. Next, we have the maximum likelihood estimates, computed with the procedure above, labeled *MLE*. Lastly, we have maximum likelihood parameters, but with the c free instead of specified by Table 2, we let it be a free parameter and optimize the parameters in a similar way. We label this set of parameters as *free c*.

We will be comparing the likelihood ratios of each set of parameters for each period and each zone. Given that there are 23 periods and 160 zones, this gives 3680 likelihood ratios, which is too many to handle individually. Therefore, it is easiest to show them in a histogram. You can find the results in Figure 12. Here we see three plots of the aforementioned differences of log-likelihoods. The first one features the *BommerV6* −

MLE, next we have *BommerV6* − *free c* and lastly we have *MLE* - *free c*

We start with looking at the first plot. We see that the distribution is skewed very negatively; in fact, there is not a single positive value. This is good news because it implies the parameters we found are indeed all better in terms of likelihood compared to the given parameters, corroborating that they are indeed optimal. The values seem to go quite negative, showing that for some periods/zones there is a large difference between the parameters. The absolute scale of these numbers is quite hard to interpret because the datasets are so large, which makes the log-likelihoods themselves either very large or very close to 0. Some values are close to 0, which means that the corresponding *BommerV6* parameters are close to optimal, but many are also very negative. Unless the author has a specific reason to not choose the MLE parameters, like any reasons stemming from expert insight, or lacking stability of the model with these parameters, I recommend that future developers of the GMM look into the effects of recalibrating the model.

It is clear that in the likelihood sense, the MLE are an optimization compared to the *BommerV6* parameters. However, it is hard to test the significance of this difference, because the zones have significantly different sizes, making their likelihoods very unbalanced in absolute size, as was explained above. Therefore, we cannot conclude that a large likelihood difference for some zone/period implies that the parameters are better than for smaller likelihood differences. Due to this ambiguity, Section 5.1.1 features some more results to establish a significant difference between the two parameter sets.

The second plot in Figure 12 showcases the results for the *free c* parameters compared to the *BommerV6* parameters. We see almost the exact same plot as before, so it is hard to see if this is an improvement.

Lastly, the *MLE* parameters versus the *free c* parameters give us an insight into which of these is better to use. Firstly, the values on the x axis are all a lot closer to 0 than above, which shows us that the improvements in likelihood are not enormous compared to the change from *BommerV6* to *MLE*. Although most values are negative, some are also positive. This is a bad sign, as including a new parameter and keeping the rest the same should only improve the likelihood values. The fact there are positive values implies that there have to be instabilities causing this. The amount of positive values is quite small, and they are also close to 0, except for 1 outlier of around 3300. This is most likely caused due to an instability in the minimization. Other than this, it seems a solid improvement, so if one double checks for instabilities and checks if the period/zone combinations with positive differences do not have obviously wrong parameters, it is most likely an improvement to switch to this version.

From this, we can conclude at the very least that it is advisable to use the maximum likelihood estimates rather than the supplied parameters, and potentially also let the c parameter be free when doing the maximum likelihood calculations. We will not be using these different parameters when doing risk calculations in subsequent chapters, simply because otherwise we would not be able to compare our results due to other changes to the old results.

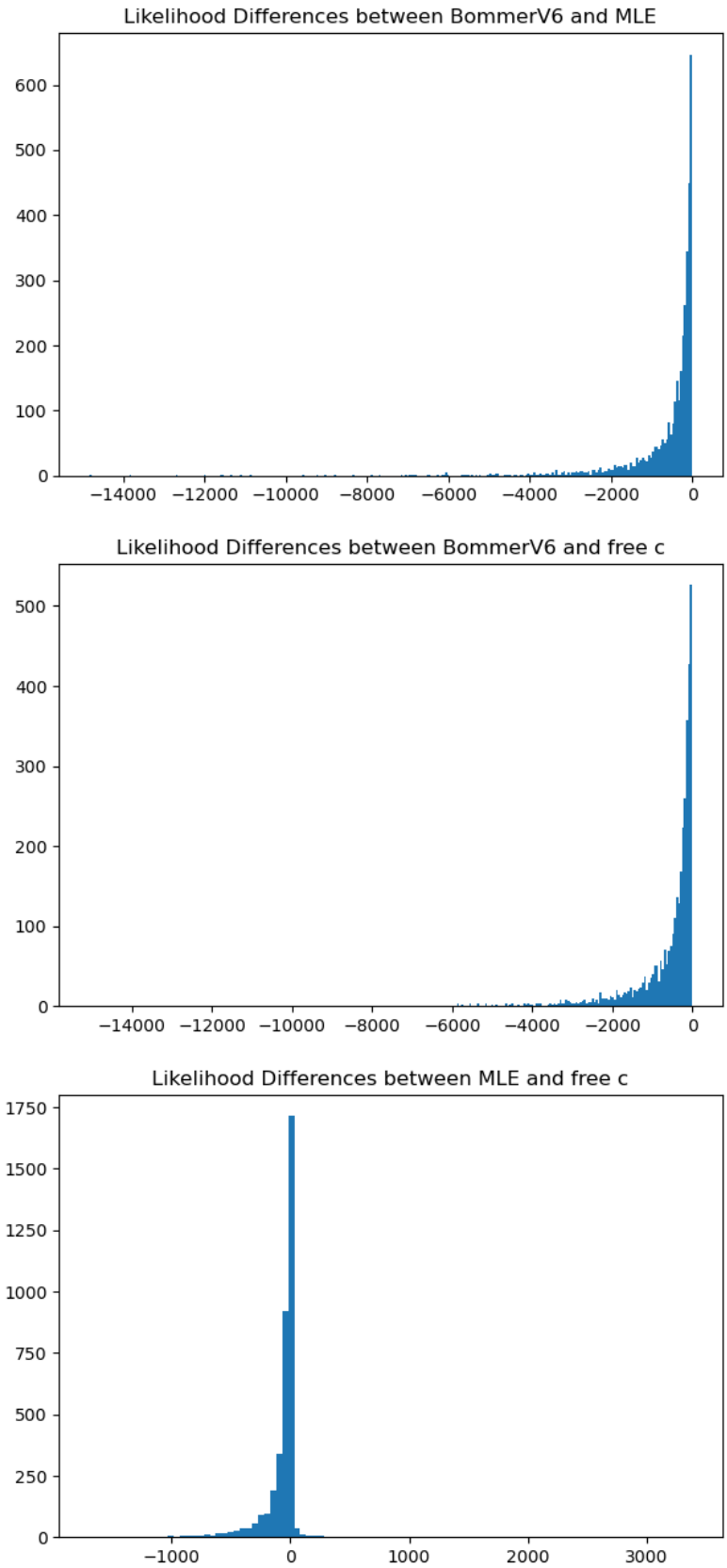


Figure 12: Different histograms of logarithmic likelihood differences for different sets of parameters. We compare parameter sets *MLE, BommerV6* and *free c*.

5 Model Validity

In this chapter we will be looking into results regarding the model validity. We will investigate two main questions:

- Does the marginal lognormality assumption for the amplification factors in Eq. 2 hold?
- What dependence structure between the periods can we find in the STRATA dataset?

The first question was raised after visual inspection of the STRATA data, which did not always look the way we expect it to. The second question was already raised in Section 2.3 and will be investigated in more detail in this chapter.

5.1 Lognormality assumption

As was said in the introduction of the chapter, it is not always visually whether the lognormal assumption is somewhat correct. Besides this, it is not an implausible suggestion that perhaps if the model's linear transfer breaks down at some point, that perhaps the lognormal assumption also breaks down after this point.

To investigate this, we start by looking at the residuals of the data. These are defined by splitting up amplification factor $\ln(A) \sim \mathcal{N}(\mu(S), \sigma(S))$ into

$$\ln(A) = \mu(S) + \sigma(S)\epsilon,$$

where ϵ is standard normally distributed. If we then calculate $\frac{\ln(A)-\mu(S)}{\sigma(S)}$, then this should therefore be a standard normal distribution, even though different data points have different means and standard deviations. We discuss a plot of the residuals of an arbitrary zone - zone 1011 - according to the V6 model. In Figure 13, we observe histograms of residuals, normalized to the standard normal distribution, done for the same zone and periods as before.

We can now clearly see that there is some skewness in these distributions, as well as some excess kurtosis in some plots. However, the other periods do look quite good. This is, of course, only a single zone, which is not at all representative of the rest of the dataset. This is merely a demonstration of how it can look. Upon inspection of more of these plots, it was found that the results from this zone do not generalize to the other ones. Sometimes the data is skewed towards the left, sometimes to the right, sometimes the low periods look fine and the higher periods are more skewed... The few consistencies in the results are that they are generally not great and generally worse in lower periods than in higher periods.

This could be happening for multiple reasons. It could be that the functional form of μ in Eq. 3 is not adequately modeling the data; it could be due to the lognormal assumption not being sufficient to map the spread. Most likely, it is even due to both not being adequate. It does not appear to have anything to do with the choice of parameters, as we see similar results for the *BommerV6* parameters as for the *MLE* parameters, introduced in Chapter 4.

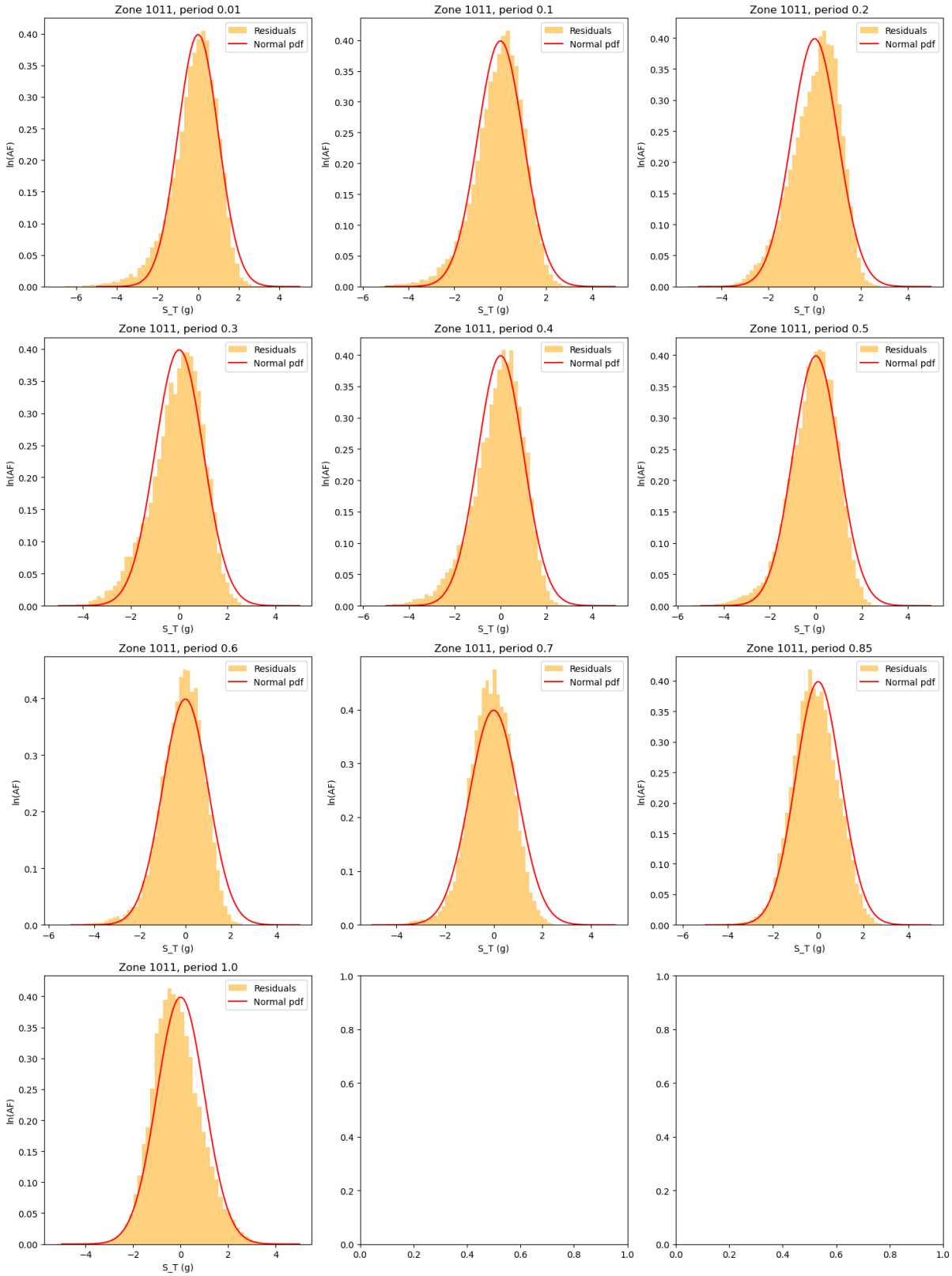


Figure 13: Standard normal residuals of zone 1011 for the maximum likelihood estimates of the GMMV6 model. Standard normal distribution is plotted as the red line for reference.

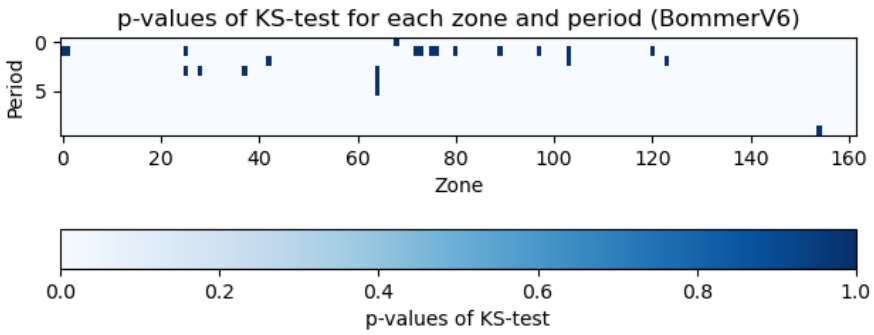


Figure 14: Colormap showing accepted p-values (> 0.05) of KS-tests for log-normality of the GMM V6 model according to the *BommerV6* parameter set.

That being said, the results do not seem too bad; there is still quite a lot of resemblance to the normal distribution. Furthermore, it would perhaps be unwise to put too much resources into perfectly recreating the STRATA data in simulations because that data in itself is also simply a reconstruction of reality and it has its own assumptions and artifacts.

5.1.1 Statistics on all zones

To get some deeper insights that are valid for the whole data set, we will now discuss some statistics on all zones.

We will start with p-values for all the parameters of all zones and periods. In Figures 14, 15 and 16, we have plots displaying whether the p-values of KS-tests for log-normality pass at confidence level 0.05 for the GMMV6 model for the three parameter sets, *BommerV6*, *MLE* and *free c* as introduced in Section 4.2. It appears that not many period-zone combinations pass the statistical tests. For the *BommerV6* parameter set, only 23 pass the test. For *MLE*, there are 162 that pass the test and for *free c* there are 176 that pass the test. This does not seem like much at first sight, but before we draw any conclusions, we can look at another test.

The following results were inspired by the observation that the quality of the estimates appears to get worse with the number of data points in the zone. Figures 17, 18 and 19 show how the average p-value over the 10 selected periods gets smaller as the zone size increases.

Interpreting these is a bit of a glass-half-full or glass-half-empty situation. We can either say that the statistical test doesn't have enough power to reject the model for lower amounts of data. Equally, you can remark that the more data there is to reject, the more

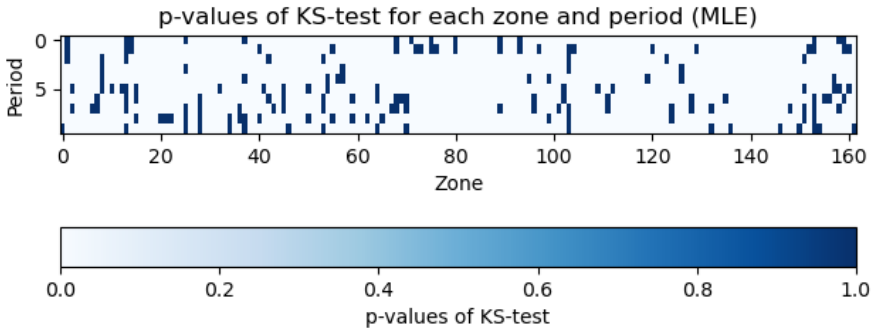


Figure 15: Colormap showing accepted p-values (> 0.05) of KS-tests for log-normality of the GMM V6 model according to the *MLE* parameter set.

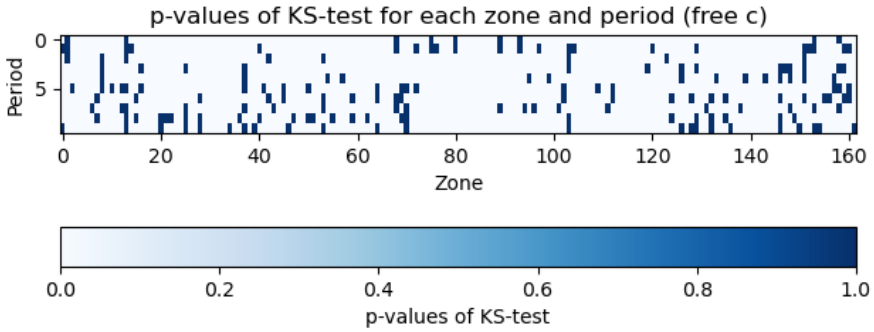


Figure 16: Colormap showing accepted p-values (> 0.05) of KS-tests for log-normality of the GMM V6 model according to the *free c* parameter set.

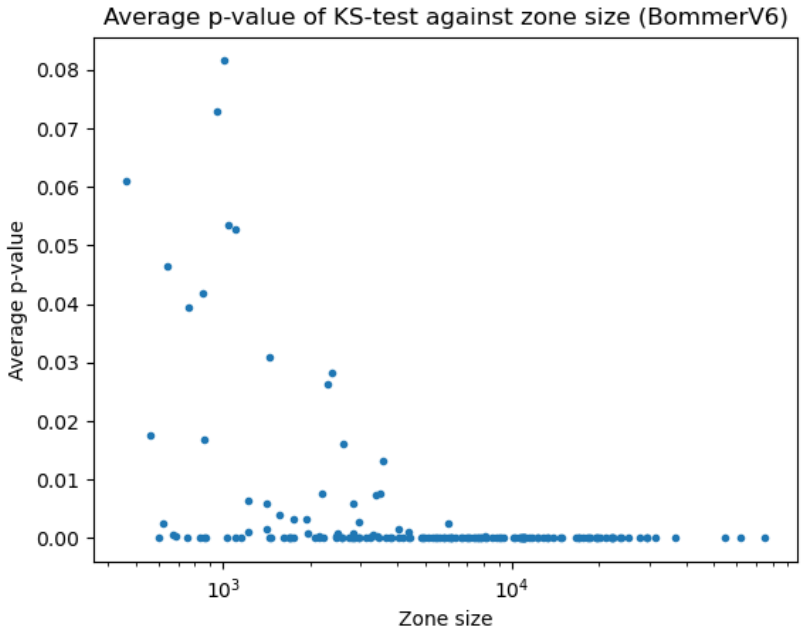


Figure 17: Scatter plot showing the average p-value of KS-test over 10 periods against the size of the zone for the *BommerV6* parameters. On the x-axis we see the zone size and on the y-axis we see the average p-value.

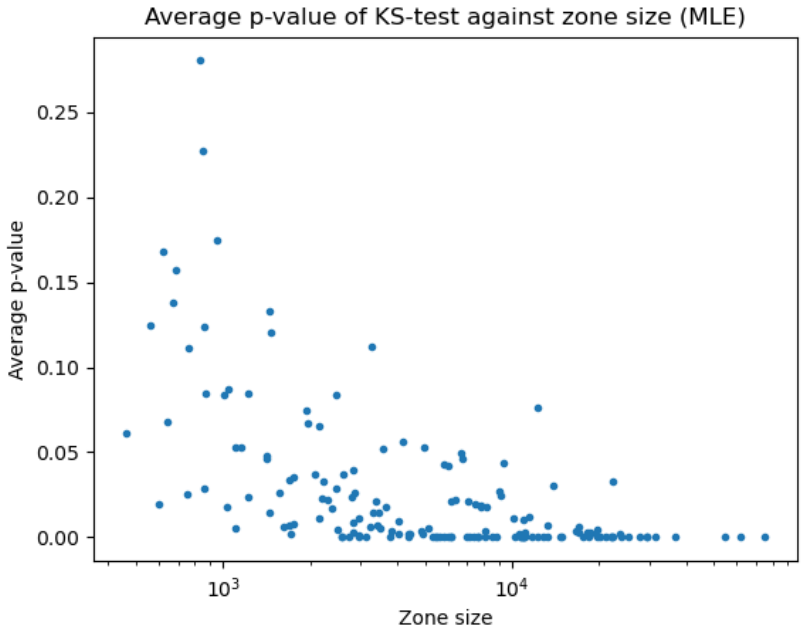


Figure 18: Scatter plot showing the average p-value of KS-test over 10 periods against the size of the zone for the *MLE* parameters. On the x-axis we see the zone size and on the y-axis we see the average p-value.

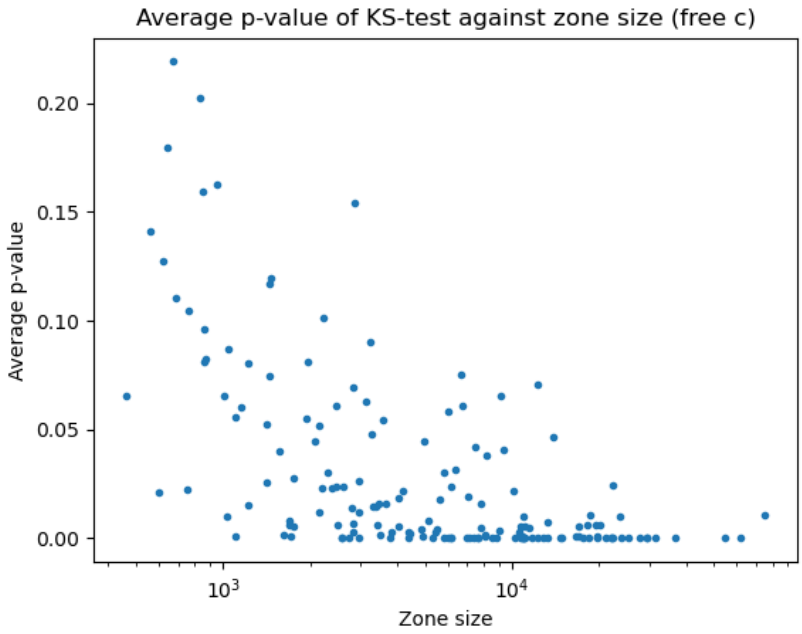


Figure 19: Scatter plot showing the average p-value of KS-test over 10 periods against the size of the zone for the *free c* parameters. On the x-axis we see the zone size and on the y-axis we see the average p-value.

tiny deficiencies will affect the outcome of the test and the less reliable it becomes. My instinct says that the second makes more sense, mostly because the visual inspections of earlier appear to give neither terrible nor perfect results, making me believe that the small imperfections will be exploited by the KS-test.

Overall, I conclude that there are multiple signs pointing to the model not being perfect, but there are also no major red flags to completely reject it. Especially considering the simplicity of the log-normal assumption, it is not appealing to reject it. Still though, there are optimizations to be made and for future iterations of the model, it is definitely wise to look at other distributions which allow for more accurate third and fourth moments.

We will not be looking into other models for the residuals in this thesis, but it should definitely be a topic for further investigation. There simply was not enough time in this project to pursue this further; rather, I decided to put the main focus into the results of Chapter 6 instead.

5.2 Dependence structure between amplification factors

As was said in Section 2.3, there appears to be ambiguity around the correlation model that should be applied to the residuals we discussed in the previous section. The instructions are unclear in the literature, and they vary from complete independence to full correlation.

We will start by looking at this in a bit of a broader manner. Firstly, we do not blindly assume multivariate log-normality. Still, we do keep the log-normal marginal dis-

tributions, but look beyond just linear correlations. This is because it could very well be that the nonlinear breakdown in the means has an effect that produces nonlinear dependencies between the surface spectral accelerations of different periods.

To investigate this, we choose to look at a transformed data of the copula space. Then, we can make a so called pairs-plot to investigate whether the copula looks Gaussian or not. The pairs plot is found in Figure 20. This pairs plot was made with all uniform data of all zones concatenated into one large dataset. We investigate all data together rather than zone by zone for two reasons: simplicity of the model and conforming to the pre-existing way of handling the dependency. Furthermore, to reliably infer a copula model, especially for a large number of dimensions, you need a large amount of data, and some zones lack data to do this analysis. Note that this plot is of a random subset of the data. To be precise, 1% of the data was randomly selected from the total data, which has around 14 million data points, which simply is too much for a scatter plot to be insightful.

In the pairs plot, the first thing we should check is the plots on the diagonal. These should be approximately uniform. For the first period, it is decently alright, but for the last 9 not really. This means that the marginal distributions we used to transform to uniform scale were not really accurate enough for this to work properly. This results in bias in the resulting copula densities, so we should take them with a grain of salt.

Ignoring the bad marginals, we can see that on the first-off diagonal, which represents neighboring periods, the dependence looks quite strong, especially for the last 5 columns. Furthermore, the contour plots for these neighboring periods look quite elliptical, suggesting a Gaussian copula model can work for these periods. In the first 4 columns, for plots lower than the third row, the dependence looks weak, but there are a few small consistent asymmetries in the dependence, likt in the second column in the last two rows, there seems to be a asymmetrical stretch upwards. In columns 5-8 on the bottom two rows, we see a shape that resembles the Clayton copula. These asymmetries hint that a based model rather than a linear correlation model can improve the model's accuracy. However, we cannot say where these effects come from. It is entirely possible that this is due to the marginals being too bad to accurately see the dependencies.

It would be interesting for further research to see the effects of improving the marginals - probably by allowing for skewed distributions rather than lognormal distributions, as suggested above in Section 5.1. Only then it would make sense to pursue a copula model. Since we do not have a better model for the marginals right now, the most reasonable choice is to stick to the previously used linear correlation model, which we will now discuss in more detail.

5.2.1 Best-fit correlation matrix

To calculate the best-fit correlation matrix, we start by taking the residuals of all the zones. As with the copula data, we once again simply concatenate this data into one dataset and then calculate the best-fit correlation matrix. The correlation matrix is found in Figure 21. Notice here that we used all the 23 periods of the data rather than only the 10 relevant ones for the fragility model.

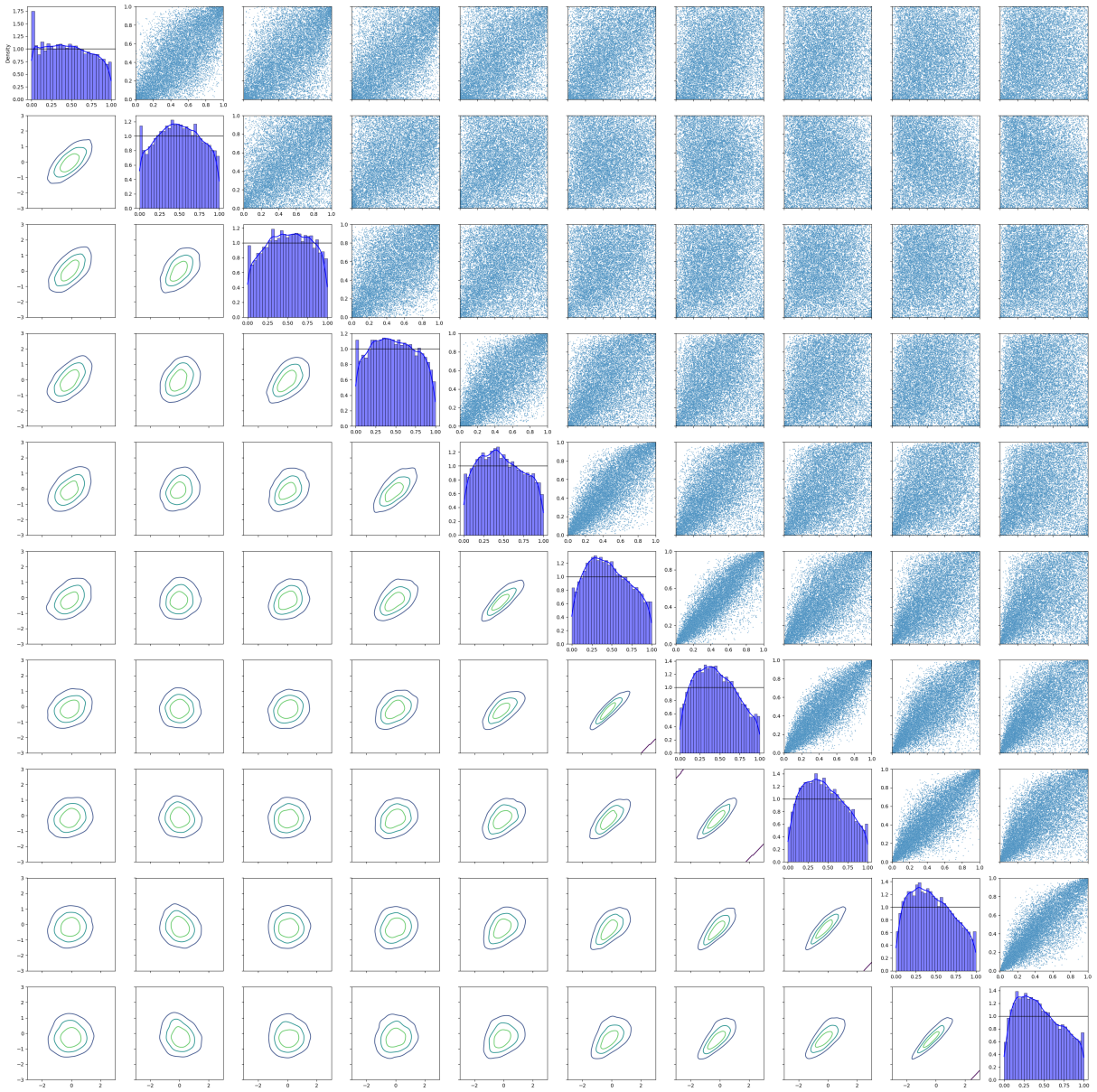


Figure 20: Pairs plot of copula-transformed data of all zones aggregated in one plot. On the diagonal we see a histogram of the copula transformed data. Above the diagonal we see a scatter plot of the data and below the diagonal we see kernel-density estimates of the data.

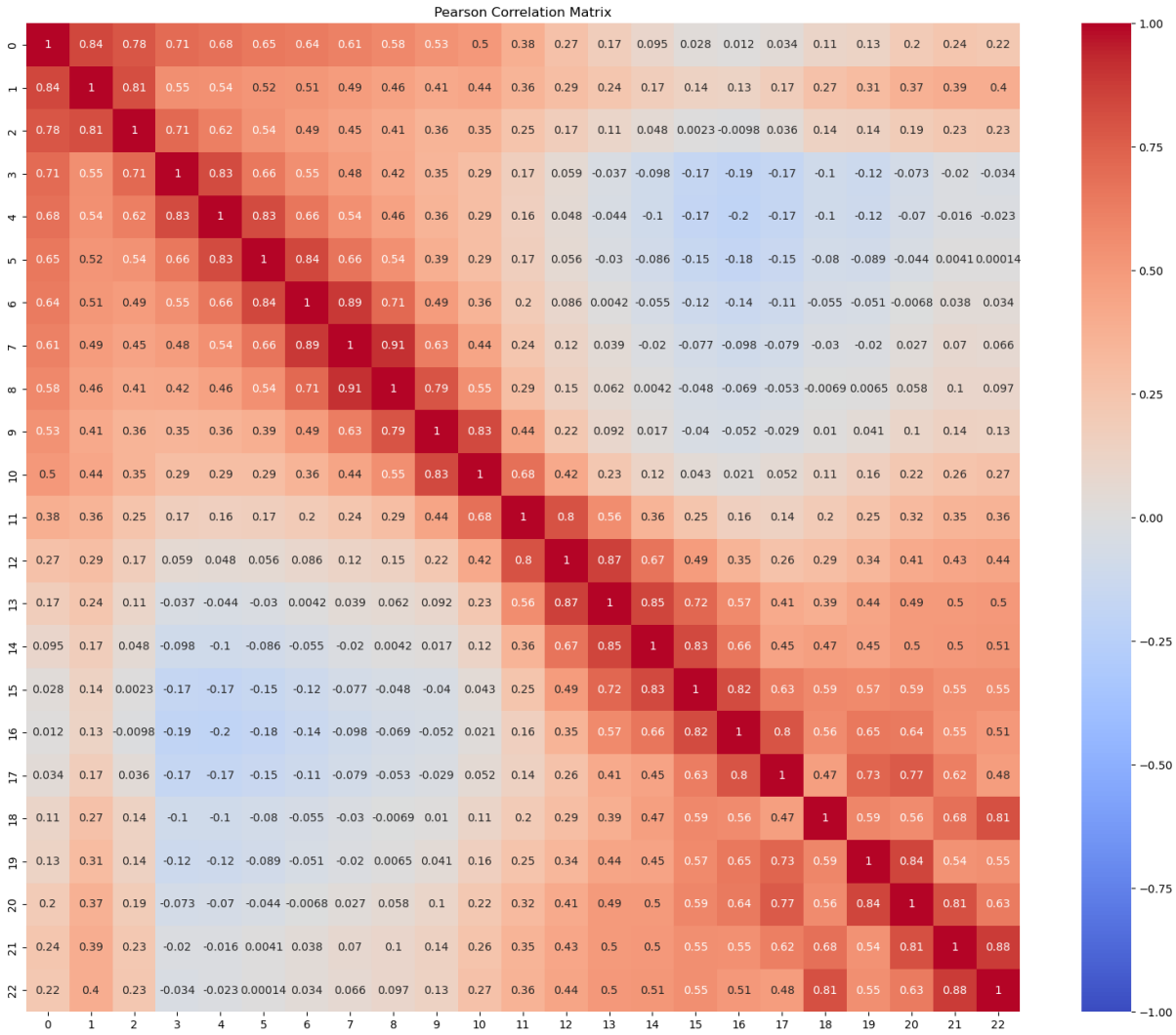


Figure 21: Correlation matrix derived from residuals of the GMMV6 model

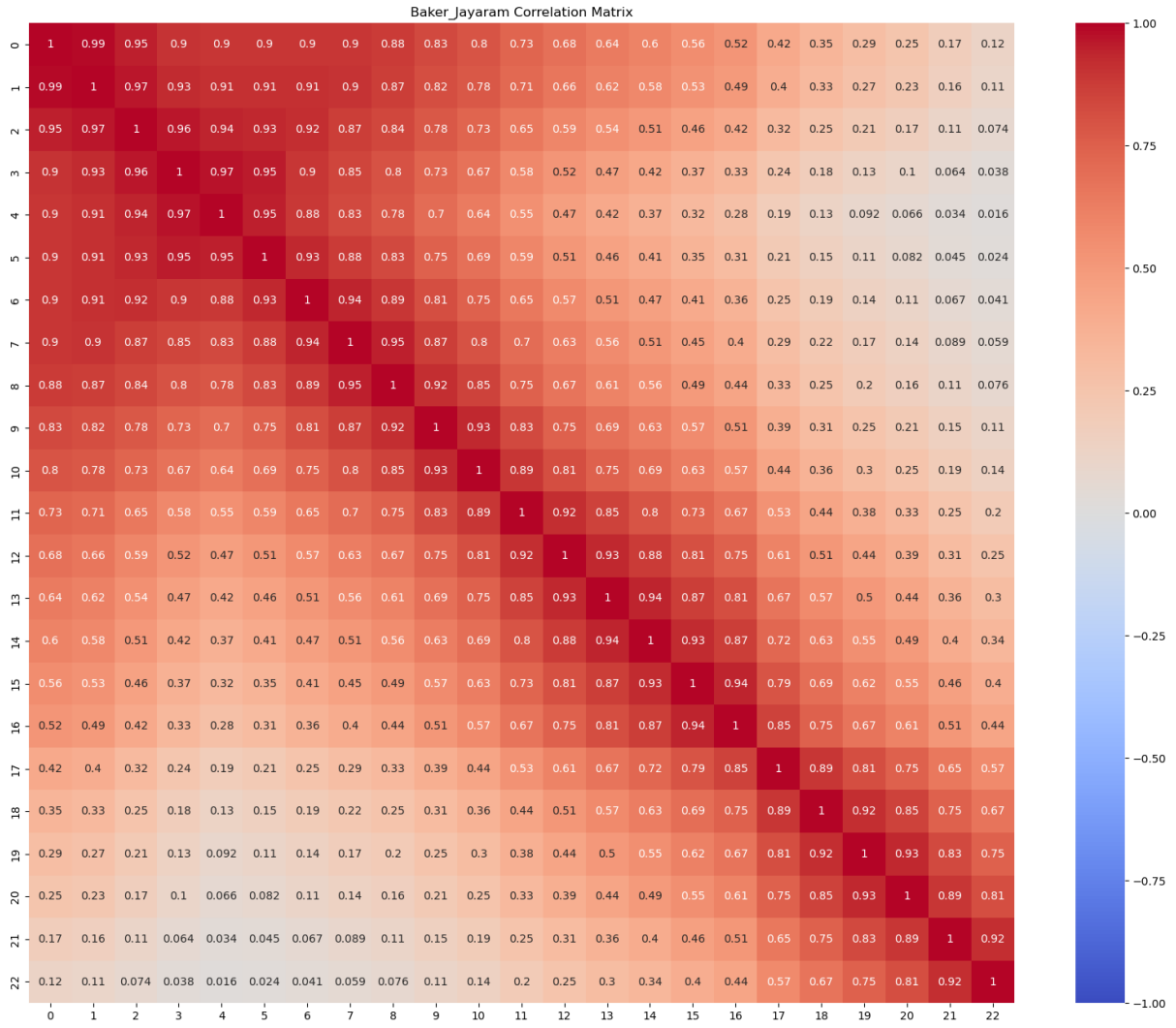


Figure 22: Baker & Jayaram correlation matrix

We can see that this matrix does not appear to be anywhere close to the two proposed matrices of full correlation (i.e. all ones in the matrix) or zero correlation (i.e. identity matrix). There is clearly some relevant correlation along the diagonal and even some negative correlation in some places. Recall that TNO suggested using a consistent model for reasons linked to the discussion in Section 8.1. With this they mean to use the same correlation in the amplifications as in the reference model. The correlation matrix they use for the reference model is plotted in Figure 22, derived by Baker & Jayaram [10]. The correlation here is clearly nowhere near the empirically derived matrix. This one is much stronger in its correlations.

Another point that is important to realize, is that even though the STRATA dataset has this correlation, that still does not mean that we would want to use it in simulations. Recall from Section 3.2.1 that for simulations, extra components of variability are added, of which we have no idea whether they bring extra correlation in reality. Therefore, we cannot conclude that this is definitely a better modeling choice compared to any of the other options. Still, I hope this sparks new insights for further research surrounding the dependence of amplification factors so we can get more definitive answers on this topic.

From this section, we can conclude that the correlations present in the STRATA data set are clearly different from any previously used model. Therefore, it is of interest to investigate how this model compares to those models, but we should still remain critical due to the STRATA dataset not being representative of a real-world scenario. In Chapter 7, we investigate the impact that the use of this correlation model has on risk calculations.

6 Computational Model Advancements

In Section 2.4.2 we saw that for risk analysis purposes, the relevant quantity that is modeled by the GMM is the average spectral acceleration. However, the simulation of the spectral acceleration is a major bottleneck in the model chain, being the only part that is reliant on Monte Carlo simulations rather than having direct information about the probability distribution. In this chapter, I will present a novel method that was created for this project to calculate the probability distribution of the average spectral acceleration.

6.1 The problem statement

As we have previously seen, an important measure for predicting earthquake damage, is the average spectral acceleration. Recall from the Section 2.4.2 that we defined the logarithm of the average spectral acceleration $\overline{Z}$ as

$$\overline{Z} = \frac{1}{10}\sum_{i=1}^{10} \ln(S_{surf,i}), \tag{14}$$

where $S_{surf,i}$ is the surface spectral acceleration at period i. Note again that i is an index which represents the i'th period in the list of 10 periods. As we know for before, we have modeled this as the spectral acceleration up to a reference depth, multiplied by an amplification factor. Therefore we can also write

$$\overline{Z} = \frac{1}{10}\sum_{i=1}^{10} \ln(S_i) + \ln(A_i),$$

where as before, S_i represents the reference spectral acceleration and A_i is the amplification factor. Recall that S_i is a lognormal variable with given parameters m_i and s_i, that are jointly lognormal over the periods with a correlation matrix B. The matrix that is currently used for this is described by the Baker & Jayaram model [10]. Furthermore A_i is lognormal given S_i with parameters defined by two functions $\mu_i(S_i)$ and $\sigma_i(S_i)$, given in (3) and (4), which as discussed before are also modeled as being jointly lognormal with some different correlation matrix R. As was discussed in Section 5.2.1, there are multiple options to choose from, ranging between zero, consistent, full or empirical correlation, the latter of which was an addition I added.

Now we define some helpful simplifications. Firstly, let $X_i := \ln(S_i)$ and $Y_i := \ln(A_i)$, so now $X_i \sim \mathcal{N}(m_i, s_i)$. Subsequently, $Y_i|X_i$ is now also also normal with parameters $\mu_i(e^{X_i})$ and $\sigma_i(e^{X_i})$. We drop the zone superscript here, just like before, but note again that the reference values X_i are independent of zone, but the amplifications Y_i are not. Furthermore, another simplification of notation will be to define function ν_i and τ_i, such that

$$\mu_i(e^{X_i}) =: \nu_i(X_i) \text{ and } \sigma_i(e^{X_i}) =: \tau_i(X_i).$$

For this to hold, we simply define ν_i and τ_i as

$$\nu_i(x) = a_i + b_i \ln\left(\frac{c_i + \mathrm{e}^x}{c_i}\right) \tag{15}$$

and for τ_i we have the piecewise function

$$\tau_i(x) = \begin{cases} \sigma_{1,i} & \text{for } S_i \leq \alpha, \\ \sigma_{1,i} + (\sigma_{2,i} - \sigma_{1,i}) \left(\frac{x - \ln(\alpha_i)}{\ln(\beta_i) - \ln(\alpha_i)} \right) & \text{for } \alpha_i < S_i \leq \beta_i, \\ \sigma_{2,i} & \text{for } S_i > \beta_i, \end{cases} \tag{16}$$

which are the functions from equations 3 and 4, but rewritten to take logarithmic arguments. A slight nuance is to recall that for GMMV6 and GMMV7, which are the versions relevant to us, the ν_i functions are also artificially bounded above and below by parameters A_{max} and A_{min}, as described in Section 2.2.2. Whether these bounds are included is irrelevant for most of the mathematical steps that follow. Anywhere it makes a difference, there will be a brief mention of it.

We can now write

$$\overline{Z} = \frac{1}{10} \sum_{i=1}^{10} X_i + Y_i := \frac{1}{10} \sum_{i=1}^{10} Z_i. \tag{17}$$

This is the quantity to be estimated. As explained before, Monte Carlo methods have been used for this up to now, due to the complex nature of the conditionality. However, we decided to step away entirely from Monte Carlo methods and attempted to model the density of $\overline{Z}$ using semi-analytical methods by means of numerical integration.

To be precise, our goal is now to find a way to model $\overline{Z}$ directly, without relying on sampling the reference values and amplification values. However, this is not a trivial task, because we do not have a closed form for the distribution of Y_i. Instead, we only know its form conditional on X_i. Getting the distribution of $X_i + Y_i$ for a given i, i.e. the marginal distribution of the surface spectral acceleration for a specific i involves a convolution integral over their joint distribution. This joint distribution is quite simple: we have that

$$f_{X_i,Y_i} = f_{X_i} \cdot f_{Y_i|X_i}.$$

However, the convolution integral is not analytically calculable because the form of $\nu_i(X_i)$ is too irregular. Numerically calculating these distributions is possible, but for each value z of $Z = X_i + Y_i$ you would have to do a numerical integration of the form

$$p_Z(z) = \int_{-\infty}^{\infty} p_{X_i,X_i}(x, z - x) dx.$$

This density is not particularly fast to compute. If you take N points to discretize x and M points to discretize z, then the complexity scales with $N * M$. Besides that, even if you have the marginals, you would still have to find a suitable way to combine them into a joint distribution. However, the idea of using the convolution integral from site-response models is not a new idea. Bazzurro and Cornell [20] took this approach in which they used a non-multivariate model, allowing them to calculate the surface densities in one step.

6.2 Method of moments

Instead of trying to find the entire joint distribution of all surface spectral accelerations, we decided to focus on the first two moments. The reason for this is that from a visual investigation of simulated datasets for the average spectral acceleration $\overline{Z}$ at different settings

for the reference model and for different zones, it appeared that the data was normally distributed. To what extent this is truly the case, we will discuss later in Section 8.2. Besides it visually appearing to work quite well, it is also a very common assumption in the study of ground motions to assume normality. Even the authors of the model (falsely) state that this is in fact the case in the GMMV6 report [3, p. 70], which goes to show that this assumption is deeply rooted in this model. For now, we note that this assumption is itself an approximation, which we will name **Approximation 1**. Assuming normality, we need only the mean and variance for the average spectral acceleration, which, as we shall see shortly, is actually feasible to find with numerical integration. The following two sections will outline the procedure to calculate these.

6.2.1 The mean of $\overline{Z}$

Finding the mean is a simple computation. We can write

$$\begin{aligned}\mathbb{E}(\overline{Z}) &= \mathbb{E}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + Y_i\right) \\ &= \frac{1}{10}\sum_{i=1}^{10} \mathbb{E}(X_i) + \mathbb{E}(Y_i) \\ &= \frac{1}{10}\sum_{i=1}^{10} m_i + \mathbb{E}(Y_i).\end{aligned}$$

Using the conditional normality of Y_i, we can now split it up as

$$Y_i = \nu_i(X_i) + \tau_i(X_i) \cdot \epsilon_i,$$

where ϵ_i is a standard normal variable and they are joint normal over the different i's with a correlation matrix R. We now get

$$\mathbb{E}(\overline{Z}) = \frac{1}{10}\sum_{i=1}^{10} m_i + \mathbb{E}(\nu_i(X_i) + \tau_i(X_i) \cdot \epsilon_i) = \frac{1}{10}\sum_{i=1}^{10} m_i + \mathbb{E}(\nu_i(X_i)).$$

The expected value that remains is not analytically solvable but is quick to calculate numerically, as it is a one-dimensional numerical integral. This quantity will return later so for future ease, we define $\overline{\nu_i} := \mathbb{E}(\nu_i(X_i))$, so finally we get

$$\mathbb{E}(\overline{Z}) = \frac{1}{10}\sum_{i=1}^{10} m_i + \overline{\nu_i}. \tag{18}$$

6.2.2 Variance of $\overline{Z}$

Calculating the variance of the average spectral acceleration is quite challenging, because it consists of many terms which will end up having some covariance with each other. The easiest way I found to show all covariances that should be taken into account is to use the law of total variance.

Theorem 2 (Law of Total Variance)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mu)$ with finite second moment. Furthermore, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then*

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|\mathcal{G})) + \mathrm{Var}(\mathbb{E}(X|\mathcal{G}))$$

A proof for this theorem can be found in many probability textbooks, such as Blitzstein [21, p. 401]. In order to use this theorem to calculate the variance of $\overline{Z}$, we need its second moment to be finite, which will make the variance finite. To show this, we start by splitting up Y_i like before to get

$$\overline{Z} = \frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i) + \tau_i(X_i)\cdot\epsilon_i.$$

Its second moment is then

$$\begin{aligned}
\mathbb{E}\left(\overline{Z}^2\right) &= \mathbb{E}\left[\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i) + \tau_i(X_i)\cdot\epsilon_i\right)^2\right] \\
&= \frac{1}{100}\sum_{i=1}^{10}\sum_{j=1}^{10} \mathbb{E}(X_iX_j) + \mathbb{E}(\nu_i(X_i)\nu_j(X_j)) + \mathbb{E}(\tau_j(X_j)\tau_j(X_j)\cdot\epsilon_i\epsilon_j) \\
&\quad + 2\,\mathbb{E}(X_i\nu_j(X_j)) + 2\,\mathbb{E}(X_i\tau_j(X_j)\epsilon_j) + 2\,\mathbb{E}(\nu_i(X_i)\tau_j(X_j)\epsilon_j).
\end{aligned}$$

With ν_i and τ_i as described in Eqs. (15) and (16), it is easy to construct bounds for each of these terms and to prove the finiteness of the second moment. We will treat it term by term. The first term is finite because it is a covariance of two joint normal RV's.

For the second term, we consider the behavior of the function $\nu_i(X_i)$ in the tails to figure out if there can be contributions diverging to infinity. If we consider

$$\lim_{x\to-\infty}\nu(x) = \lim_{x\to-\infty} f_1 + f_2\ln\left(\frac{e^x+f_3}{f_3}\right) = f_1,$$

so at negative infinity, $\nu_i(X_i)$ will never diverge. Next, at $+\infty$, we get

$$\lim_{x\to\infty}\nu(x) = \lim_{x\to\infty} f_1 + f_2\ln\left(\frac{e^x+f_3}{f_3}\right) = \lim_{x\to\infty} f_1 + f_2(x-\ln(f_3)),$$

which means that $\nu(x)$ converges to a linear relationship for increasing x. Therefore, as the reference spectral acceleration X_i increases, the amplification $\nu_i(X_i)$ will be increasingly more linear w.r.t. X_i. So, as long as X_i has finite variance, $\nu_i(X_i)$ will also have finite variance. Therefore, $\nu_i(X_i)$ is finite in variance, and the covariances between different indices i, j $\mathbb{E}(\nu_i(X_i)\nu_j(X_j))$ will be as well. Lastly, in the case of using GMMV6 and GMMV7, the mean is bounded by design, so in this case it most certainly holds that it is finite.

The third term with the τ_i-functions is very trivial, given that they have a very obvious upper bound of $\sigma_{2,i}$ and we also have $\mathbb{E}(\epsilon_i\epsilon_j) = R_{ij}$, and lastly we have the independence of all $\tau_i(X_i)$'s with all ϵ_i's. Therefore, we get

$$\mathbb{E}(\tau_i(X_i)\tau_j(X_j)\cdot\epsilon_i\epsilon_j) \leq \sigma_{2,i}\sigma_{2,j}R_{ij}.$$

The last three terms follow analogously because we have already proven bounds for each of the functional forms that are constant or linear. This means that the variance we are

trying to compute does indeed exist, and consequently we can use the law of total variance to compute it.

Recall that to use the law of total variance, we need to choose a sub-σ-algebra to condition on. For this problem, the choice is obvious: we want to use $\mathcal{G} = \sigma(X_1, ..., X_{10})$, where σ is the generator of the smallest sub-σ-algebra containing its arguments, which gives

$$\begin{aligned}
\mathrm{Var}(\overline{Z}) &= \mathbb{E}\left(\mathrm{Var}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + Y_i \middle| \mathcal{G}\right)\right) + \mathrm{Var}\left(\mathbb{E}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + Y_i \middle| \mathcal{G}\right)\right) && (19)\\
&= \mathbb{E}\left(\mathrm{Var}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i) + \tau_i(X_i)\cdot\epsilon_i \middle| \mathcal{G}\right)\right) && (20)\\
&+ \mathrm{Var}\left(\mathbb{E}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i) + \tau_i(X_i)\cdot\epsilon_i \middle| \mathcal{G}\right)\right) && (21)\\
&= \mathbb{E}\left(\mathrm{Var}\left(\frac{1}{10}\sum_{i=1}^{10} \tau_i(X_i)\cdot\epsilon_i \middle| \mathcal{G}\right)\right) + \mathrm{Var}\left(\mathbb{E}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i) \middle| \mathcal{G}\right)\right). && (22)
\end{aligned}$$

Here we use in the first term that the conditional variance of $f(X)|X$ is 0 and we use independence in the second term.

Let us now tackle these terms one by one. We start with the first term. The expression in the conditional has an expectation of 0, because of linearity and independence between X_i and ϵ_i, so therefore

$$\begin{aligned}
\mathbb{E}\left(\mathrm{Var}\left(\frac{1}{10}\sum_{i=1}^{10} \tau_i(X_i)\cdot\epsilon_i \middle| \mathcal{G}\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(\left(\frac{1}{10}\sum_{i=1}^{10} \tau_i(X_i)\cdot\epsilon_i\right)^2 \middle| \mathcal{G}\right)\right)\\
&= \frac{1}{100}\sum_{i,j=1}^{10} \mathbb{E}(\mathbb{E}(\tau_i(X_i)\tau_j(X_j)\epsilon_i\epsilon_j|\mathcal{G}))\\
&= \frac{1}{100}\sum_{i,j=1}^{10} \mathbb{E}(\tau_i(X_i)\tau_j(X_j)R_{ij}).
\end{aligned}$$

Lastly, we can use that for arbitrary X, Y with finite second moments, the covariance can be written as $\mathbb{E}(XY) = \mathrm{Cov}(X, Y) + \mathbb{E}(X)\,\mathbb{E}(Y)$, hence we can write

$$\frac{1}{100}\sum_{i,j=1}^{10} \mathbb{E}(\tau_i(X_i)\tau_j(X_j)R_{ij}) = \frac{1}{100}\sum_{i,j=1}^{10} R_{ij}\,\mathrm{Cov}(\tau_i(X_i), \tau_j(X_j)) + R_{ij}\,\mathbb{E}(\tau_i(X_i))\,\mathbb{E}(\tau_j(X_j)).$$

Here, we make an approximation, which is to set $\mathrm{Cov}(\tau_i(X_i), \tau_j(X_j)) = 0$, or equivalently, we assume $\tau_i(X_i)$ and $\tau_j(X_j)$ to be independent. We call this **Approximation 2**. The reason why this approximation works is simply because these two quantities are indeed very close to being independent. We can show this empirically by performing a range of Monte Carlo simulations for all different index combinations and for all possible parameter settings for the reference model and for all different τ_i functions for all zones. With these simulations, we can find what these covariances are empirically and we can

then look at statistics for these covariances.

Running this simulation gives that $\mathrm{Cov}(\tau_i(X_i), \tau_j(X_j))$ is at maximum 0.7% of the total covariance in this term, or equivalently

$$\frac{\mathrm{Cov}(\tau_i(X_i), \tau_j(X_j))}{\mathrm{Cov}(\tau_i(X_i), \tau_j(X_j)) + \mathbb{E}(\tau_i(X_i))\,\mathbb{E}(\tau_j(X_j))} \leq 0.007$$

for any choice of reference model, over a range of different magnitudes and distances from the epicenter. Furthermore, the average of the above quantity is approximately 0.09%. Lastly, this part of the total covariance of $\overline{Z}$ is in general also quite small compared to the other sources of covariance in the model. These give quantitative support for the use of this approximation.

It is also important to understand the need for this approximation. We make it because it makes this term very easy to compute with numerical integration. Now we need 10 unique numerical integrations to calculate $\mathbb{E}(\tau_i(X_i))$ for each i, whereas otherwise we would need to calculate a covariance for each combination of indexes, and the integrations to do so would be two-dimensional.

We now return to Eq. 22 and we continue with the second term, which is the more convoluted one of the two. We can start with an easy simplification, which is to remove the conditional expectation. All random variables inside are known with respect to the condition, so we can simply take everything out of the expectation, giving

$$\mathrm{Var}\left(\mathbb{E}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i)\middle|\mathcal{G}\right)\right) = \mathrm{Var}\left(\frac{1}{10}\sum_{i=1}^{10} X_i + \nu_i(X_i)\right).$$

The variance of this sum will give many different covariances between the terms. Consider each of these.

- $\mathrm{Cov}(X_i, X_j)$: This one is easy, it is simply equal to $s_i s_j B_{ij}$, where B is the reference model correlation matrix.
- $\mathrm{Cov}(X_i, \nu_j(X_j))$: This one is not analytically calculable. It will result in an integral that has to be solved numerically. It is feasible to compute, given that it is a two-dimensional integral over normal densities. Of the 100 terms in the sum, 55 would be unique (due to symmetry of the covariance matrix), so the computation would be 55 times a two dimensional integral. However, we can simplify it quite a lot to make the computation a lot lighter. We start by considering that X_i and X_j are joint-normal with correlation coefficient B_{ij}. Therefore we can make a Cholesky decomposition of these random variables such that

  $$X_i = m_i + s_i Z_i$$
  $$X_j = m_j + B_{ij} s_j Z_i + s_j \sqrt{1 - B_{ij}^2} Z_j,$$

  where Z_i and Z_j are uncorrelated standard normal variables. Now we can substitute the first equation into the second to get

  $$X_j = m_j + B_{ij}\frac{s_j}{s_i}(X_i - m_i) + s_j\sqrt{1 - B_{ij}^2} Z_j.$$

Obviously this expression is symmetric w.r.t. switching i and j so it also works to rewrite X_i. We have now rewritten one of the two random variables as a linear combination of the other one and some random noise term. This allows us to simplify the covariance to the following:

$$\begin{aligned}\operatorname{Cov}(X_i, \mu_j(e^{X_j})) &= \mathbb{E}\left[(X_i - m_i)(\nu_j(X_j) - \overline{\nu_j})\right] \\ &= \mathbb{E}\left[(m_i + B_{ij}\frac{s_i}{s_j}(X_j - m_j) - m_i)(\nu_j(X_j) - \overline{\nu_j})\right] \\ &= B_{ij}\frac{s_i}{s_j}\,\mathbb{E}\left[(X_j - m_j)(\nu_j(X_j) - \overline{\nu_j})\right] \\ &= B_{ij}\frac{s_i}{s_j}\operatorname{Cov}(X_j, \nu_j(X_j)).\end{aligned}$$

Now, we only need to compute 10 covariances, which are 1 dimensional integrals. Therefore, the number of computations, assuming equal grid spacings N for the old two-dimensional integral and this one-dimensional integral, is reduced by a factor $5.5N$.

- $\operatorname{Cov}(\nu_i(X_i), \nu_j(X_j))$: Recall that

$$\nu_i(X_i) = a_i + b_i \ln\left(\frac{\mathrm{e}^{X_i} + c_i}{c_i}\right).$$

Suppose now that X_i is small, then the function is practically constant. So if either X_i or X_j are in this regime, the covariance will be close to 0. On the contrary, if both are large, then the function will be rapidly converging to the linear function $a_i + b_i(X_i - \ln(c_i))$. In this case, the expectation will be approximately

$$\mathbb{E}(\nu_i(X_i)) \approx a_i + b_i(m_i - \ln(c_i))$$

and the covariance will be

$$\begin{aligned}\operatorname{Cov}(\nu_i(X_i), \nu_j(X_j)) \approx\ & \mathbb{E}[(a_i + b_i(X_i - \ln(c_i)) - a_i - b_i(m_i - \ln(c_i))) \\ & * (a_j + b_j(X_j - \ln(c_j)) - a_j + b_j(m_j - \ln(c_j)))] \\ =\ & b_i b_j\,\mathbb{E}[(X_i - m_i)(X_j - m_j)] \\ =\ & b_i b_j s_i s_j B_{ij}.\end{aligned}$$

Now, we assume linearity for $\nu_i(X_i)$ such that

$$\nu_i(X_i) = a_i + b_i(X_i - \ln(c_i)),$$

then it is normally distributed with mean $a_i + b_i(m_i - \ln(c_i))$ and variance $(b_i s_i)^2$, the latter of which can be computed as

$$(b_i s_i)^2 = \mathbb{E}((\nu_i(X_i) - \overline{\nu_i})^2).$$

Therefore, in the linear approximation, we have

$$\begin{aligned}\operatorname{Cov}(\nu_i(X_i), \nu_j(X_j)) &= \mathbb{E}((\nu_i(X_i) - \overline{\nu_i})^2)\,\mathbb{E}((\nu_j(X_j) - \overline{\nu_j})^2)B_{ij} \\ &= \sqrt{\operatorname{Var}(\nu_i(X_i))\operatorname{Var}(\nu_j(X_j))}B_{ij}.\end{aligned}$$

This is interesting because the quantity $\mathbb{E}((\nu_i(X_i) - \overline{\nu_i})^2)$ is easy to calculate for all i because once again it involves only a 1D numerical integral. Now consider this same quantity but now we assume the constant regime, i.e. $\nu_i = a_i$. Now $\mathbb{E}((\nu_i(X_i) - \overline{\nu_i})^2) = \mathbb{E}((a_i - a_i)^2) = 0$, hence the same estimator also works assuming linearity.

We now have a quantity that we know works assuming that X_i is in either of two regimes, so it is a logical choice to see how it performs in general. Obviously, this is not a fully accurate picture, as a random variable never fully in a certain regime, but only perhaps has a large part of its density in a certain region. Therefore, this approximations works better for smaller reference variances and means that are far away from the non-linear breakdown point. Still though, this will always be a decent estimate because it also happens to a sort of first-order Taylor approximation for the covariance, which we can see as follows:

$$\begin{aligned}\operatorname{Cov}(\nu_i(X_i), \nu_j(X_j)) &= \rho(\nu_i(X_i), \nu_j(X_j))\sqrt{\operatorname{Var}(\nu_i(X_i))\operatorname{Var}(\nu_j(X_j))} \\ &\approx \rho(X_i, X_j)\sqrt{\operatorname{Var}(\nu_i(X_i))\operatorname{Var}(\nu_j(X_j))} \\ &= \sqrt{\operatorname{Var}(\nu_i(X_i))\operatorname{Var}(\nu_j(X_j))}B_{ij}.\end{aligned}$$

The reason this is a first order approximation is because if you assume ν_i ad ν_j to be linear close to the mean of X_i and X_j, this approximation will be quite precise. Given the the mean function is linear in two regimes and switching between the two in the other, it appears a reasonable choice to use this quantity to estimate the covariance. This approximation we will call **Approximation 3**.

We have now considered all components, all that remains is to sum them up and divide by 100. This gives us the variance of $\overline{Z}$:

$$\begin{aligned}\operatorname{Var}(\overline{Z}) = \frac{1}{100}\sum_{i,j=1}^{10} B_{i,j}\left(s_i s_j + 2\frac{s_i}{s_j}\operatorname{Cov}(X_j, \nu_j(X_j)) + \sqrt{\operatorname{Var}(\nu_i(X_i))\operatorname{Var}(\nu_j(X_j))}\right) \\ + \frac{1}{100}\sum_{i,j=1}^{10} R_{ij}\,\mathbb{E}(\tau_i(X_i))\,\mathbb{E}(\tau_j(X_j))\end{aligned} \tag{23}$$

6.3 Comparison to Monte Carlo method

In our method for computing the average spectral acceleration $\overline{Z}$, we assumed a normal distribution, because of **Approximation 1**. Since we now have its mean and variance, given by Eqs. 18 and 23, we have all the ingredients to use this model in practice. Given that the entire idea of the method was to make computations quicker for the computer to handle, we will compare the two in terms of speed, but we also made Approximations 1-3, so we will also be investigating the impact of these. We start by investigating the latter, which will inform us on how to properly compare the two methods in a benchmarking test.

6.3.1 Accuracy

In this section, we aim te get a grasp on how accurate the numerical integration method is. Firstly, it is important to note that it is not clear-cut what the proper way should be to investigate their similarities and/or differences, as there are many statistical ways to compare two distributions. For example, one can look at the relative entropy, or Kullback–Leibler divergence, between the two distributions. Another idea could be to use Chi-squared goodness-of-fit tests and test the Monte Carlo samples to the distribution calculated with the numerical method. Rather than do this, we choose a more practical approach, which is to see the effect it has on the risk calculations. To be precise, recall that in Section 2.4 we discussed that the the average spectral acceleration was the intensity measure we use for calculating probabilities of damage states. To be precise, the fragility model specified conditional distributions, which we can now integrate over the distributions we computed for $\overline{Z}$ to get the probabilities of ending up in damage states. The approach of using the risks rather than comparing the distributions will give us direct insights in how using the new method impacts the overall risks.

Secondly, we note that the two methods for calculating the model both have different sources of error. The Monte Carlo method has an deviation from the true model in the sense that you are limited to a finite sample size, so any set of results will never show the true distribution described by the model. It will however converge to it with increased sample size. The error caused by this, we will call the **sampling error**. On the other hand, the numerical integration method has 3 built-in approximations, so it will never converge to the true distribution. We call this deviation the **methodological error**. Lastly, a numerical integral always has two extra sources of error which are unavoidable, being the **truncation error** and **discretization error**. Lastly, both suffer from an error in the risk calculation that in itself is a numerical integral, so both have a truncation error and a discretization error here as well, although this one is less relevant as it is identical for both methods.

We can see that both methods have are flawed in their own way, which makes comparing the methods hard. We have no base truth to compare the two methods to. We will however try to isolate these sources of error to gauge the impact that any of them have on the risk calculations.

To do this comparison, we will have to create a baseline to make comparisons to, otherwise there is really nothing much to say about the accuracy. To do this, we will simply take a Monte Carlo run with as much samples as we can get, which for our purposes is 1 billion samples. The model settings that we choose are: 3 different magnitudes, 3 different distances, and for the rest we fix all settings in the reference model to worst case scenarios, so highest reference means and highest reference variances. This gives 9 different instances of the model with 1 billion samples each, for which we have computed the risks of damage or probabilities of exceedence (PoE) as we will call them from now on. To get these PoE's, we integrate them over all parts of the fragility model, each resulting in a different PoE. To be specific, the fragility model has 35 building types, 6 damage states, 3 consequence branches and 3 fragility branches. For what these entail specifically, the report for the fragility & consequence model FCMV7 should be consulted [22]. All of these represent dimensions among which the PoE differ, so in the end, for our each run of this type we get $3 \cdot 3 \cdot 35 \cdot 6 \cdot 3 \cdot 3 = 17010$ different PoE's.

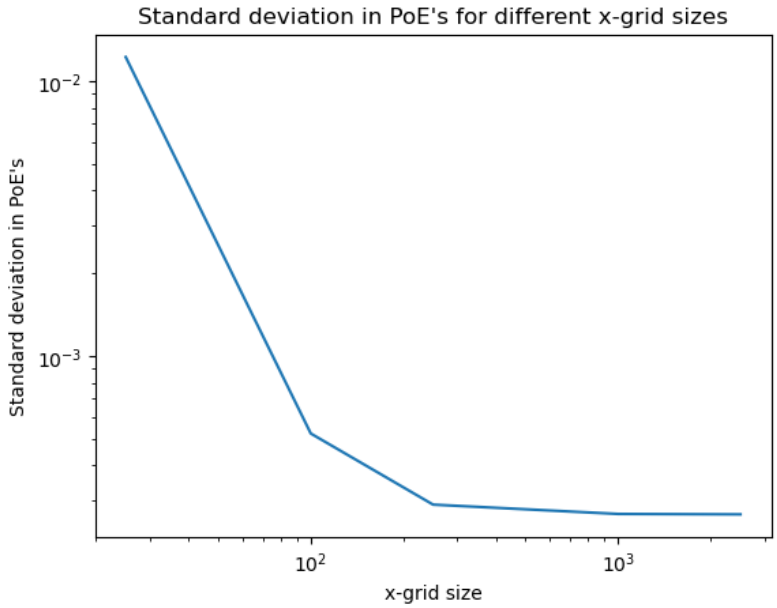


Figure 23: Standard deviations of differences between probabilities of exceedance (PoE's) for numerical integration runs against a Monte Carlo run with 1 billion samples. The numerical integration runs vary in the number of points used for the numerical integration, which can be seen on the x-axis in log-scale. On the y-axis we see the standard deviations of the difference between the PoE's for the different methods.

Now, we do a run with the exact same settings, but using the numerical integration method, but using different numbers of grid points on the to integrate over, ranging between 25 and 10000. This once again gives us the same 17010 PoE's for each number of grid points, but slightly differing due to the different sources of errors. We can now take the differences between the Monte Carlo run and each of the numerical runs and compute the standard deviations of each of these datasets. The results from that can be seen in Figure 23. Note that this is a log-log plot, so if there would be true convergence between the two datasets, we expect to see a linear relationship on this plot, which would suggest a power law scaling between the two. Instead, we see that it flattens out at a certain minimum standard deviation at 1000 and 2500 samples. The value at 250 is still 8% larger than at 1000. The most interesting thing that we can take away from this plot is that there is clearly some observable methodological error. Another take-away is that it never really makes sense to go beyond 1000 grid points, as there will be no gain in observable gain in accuracy.

The previous analysis tells us something about the convergence of the two and the model error, but it is also interesting to look at their dispersion in general. We now investigate Figure 24. On the x-axis we can see the PoE value of the numerical run and on the y-axis we see the ratio between the PoE of the Monte Carlo run and the numerical run. So this means that values close to 1 are a sign that the two methods are similar.

The first insight from the plot is that the dots for 25 numerical integration points are all over the place compared to the other ones and the other ones look quite similar, the most noticeable dissimilarity being that there is a streak of red dots ($n = 100$) on the

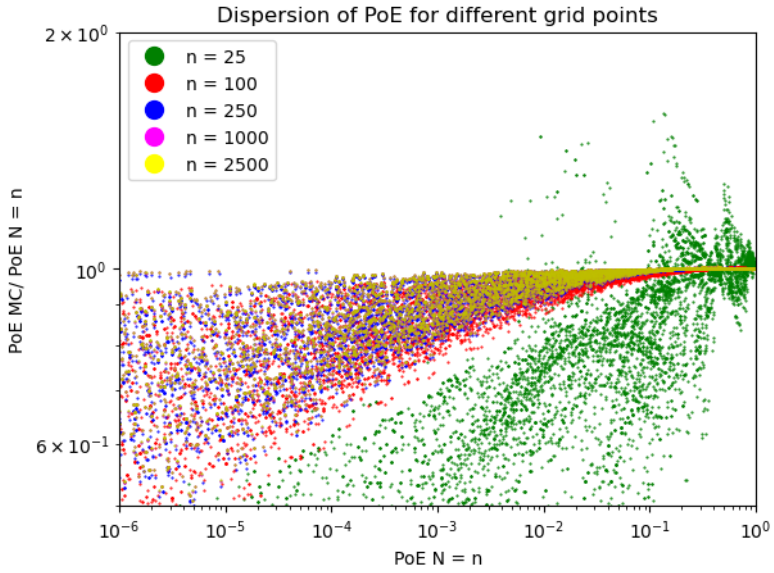


Figure 24: Dispersion between numerical integration and Monte Carlo methods. On the x-axis we see the PoE's of the numerical integration method with different numbers of grid points n and on the y-axis we see the ratio of the PoE's of a Monte Carlo run with 1 billion samples to the PoE's on the x-axis.

right of the plot that sticks out below the other three colors. This implies that for higher risks, there is a consistent difference between the two methods.

By far the most interesting observation here is the downward curvature of the points in a fan-like shape for higher number of grid points. The fan-like shape can be attributed to errors such as the sampling error and errors related to numerical integration. These are approximately constant in size, making the difference in the ratios for different points increasingly large. What is more interesting in this plot is the downward-sloping curvature. It can only be explained with methodological sources of error, as the other sources of error, such as sampling errors, should not be antisymmetric in this way. Furthermore, the numerical integration error also cannot be responsible due to the fact that it stays consistent with increasing numerical integration points.

Further investigation revealed that the methodological error responsible for this is Approximation 1, the normality assumption. The data shows a consistently negative skewness, ranging from -0.25 to 0, obtained through Monte Carlo simulation. The other methodological errors due to Approximations 2 & 3 are comparatively irrelevant in their effects, as these only generate a tiny error in the variance, compared to introducing skewness, which clearly has a much larger impact.

The way this looks visually in the worst-case scenario can be seen in Figure 25. Observe that this happens for a magnitude of 6.5, which is the highest setting. This is obviously not a coincidence, since higher magnitudes are more in the nonlinear breakdown regime, and this is the effect that will cause the skewness. The nonlinear breakdown will make it so that the higher the value of the reference spectral acceleration is, the lower the mean amplification is, thereby reducing the likelihood of sampling high surface values, and inducing a negative skewness. The plot shows the absolute worst-case scenario in terms of

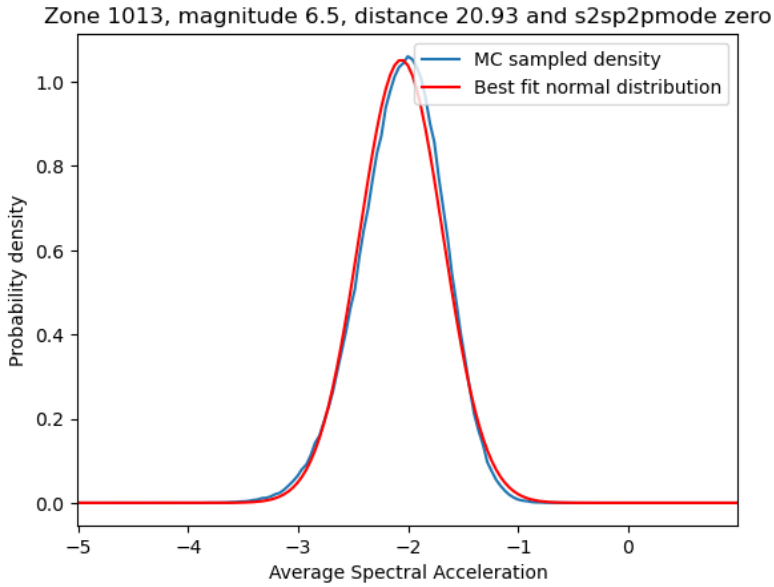


Figure 25: Plot comparing the distributions of the average spectral acceleration for the Monte Carlo simulation method and the numerical integration method in a worst-case scenario. This scenario happens for magnitude 6.5 and distance 20.93 km using the zero correlation mode. On the x-axis, we see the average spectral acceleration and on the y-axis we see the probability density of the two methods. The red curve is the best-fit normal distribution calculated with the numerical integration method and the blue curve was obtained with the Monte Carlo method.

skewness, at a value of -0.27 and, in our opinion, the fit is still very decent as the majority of the densities overlap. The main difference is in the tails of the distribution, where the Monte Carlo density will be consistently lower than the normal distribution, hence the downward curvature in Figure 24.

This raises two important questions. Firstly: How relevant is it that we observe this downward trend in the PoE's? To answer this, we start by noting that the negative skewness implies that the numerical method will always be overestimating the risks; hence it is a more cautious approach. Secondly, the overall impact on the aggregate risk calculations over the entire Groningen region is also quite limited, which we will discuss in more detail in the next chapter. It is important to realize that the higher the risk, the lower the probability of an earthquake occurring that corresponds to this risk. A standard figure in national risk policies is that the probability of a person dying in a certain risky situation should not be higher than 10^{-5}. Therefore, any risk lower than this is already inherently irrelevant. Furthermore, this risk is still conditional on things like magnitude and distance. Any probabilities we get from this sill need to be multiplied by the probability of getting an earthquake that would cause this risk. Therefore, there is a certain sweet spot in the risk where the probability of collapse is not unreasonably low and the earthquake corresponding to this risk is also not exceedingly unlikely to occur. This happens for risks around 10^{-3}, with earthquakes of around magnitude 4.5 causing these to occur, which have never happened in reality, but have an estimated probability of the order of magnitude 10^{-2}, which brings us to a general risk of around 10^{-5}. The maximum deviation from the numerical method to the Monte Carlo method at this point is around 20%.

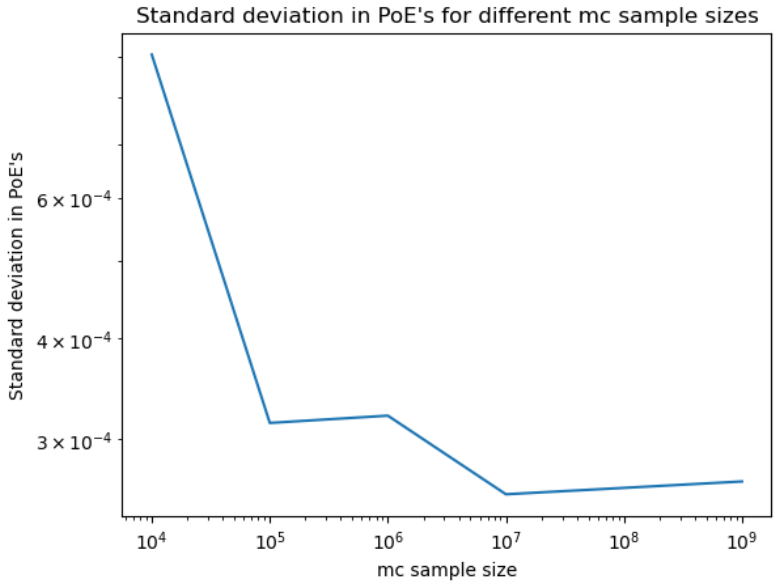


Figure 26: Standard deviations of differences between PoE's for Monte Carlo runs against a numerical integration run with 10.000 x-axis points

Another question is: Can we fix the skewness by dropping the normality assumption and working with a distribution that can account for skewness? This will be answered in the discussion in Section 8.2, where we will give an overview of a brief attempt to make such a model.

Finally, recall that earlier we established that, starting from around 250 to 1000 numerical integration grid points, the results were indistinguishable compared to the run of 1 billion Monte Carlo samples. Now we can investigate the other way around: What is the number of Monte Carlo samples needed to make the model error the dominant source of error when comparing the two methods? To do this, we can do the same procedure as before, but instead of using different runs of the numerical method and comparing it to a large Monte Carlo run, we compare a run with 10.000 x-axis grid points to Monte Carlo runs with different sample sizes. The results of the standard deviation in the distances can be seen in Figure 26.

Just like before, in a setting where the two would converge, we expect a power-law relationship between the two, which would look linear on the log-log plot. However, clearly the line flattens rapidly already at a sample size of around 100.000. It appears that 10 million points could be a marginal improvement, but anything more than this would basically have no effect in comparing the two methods.

We can also look at the same kind of dispersion plot as in Figure 24, but then again with reversed roles like above. The results can be seen in Figure 27. Note that the run for 10.000 Monte Carlo samples was removed because it had quite poor looking results and it cluttered the plot. We can see that even though the standard deviations in Figure 26 were quite similar for larger sample sizes, here we see a clear difference, in that the there are quite a few blue and red dots sticking out of the majority of the point cloud at the bottom, which shows that there is in fact accuracy to be gained from going from 100.000 points to 10 million points, although it is quite marginal.

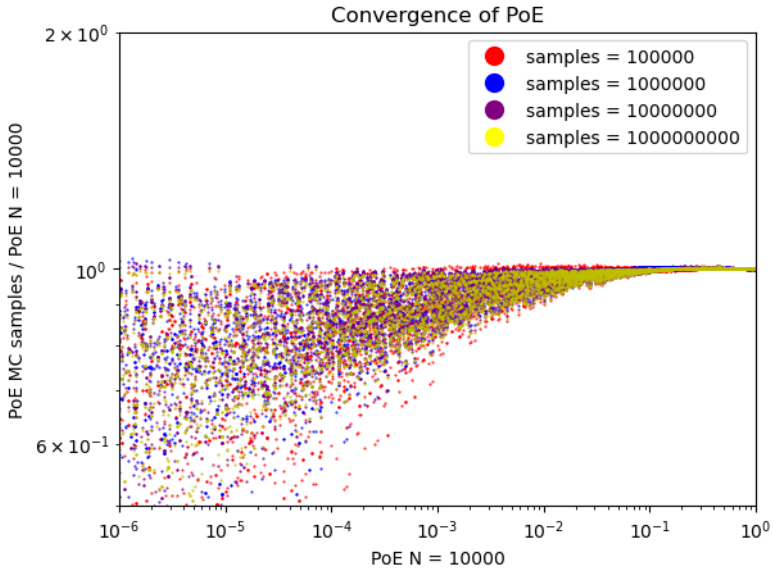


Figure 27: Dispersion between numerical integration and Monte Carlo methods. We compare the absolute risk values of the numerical to the ratios of the risks for Monte Carlo, where we vary the Monte Carlo samples per color. On the x-axis we see the PoE's of the numerical integration method with 10.000 grid points and on the y-axis we see the ratio of the PoE's of Monte Carlo run with different numbers of samples to the numerical PoE's on the x-axis. Different colors represent different numbers of MC samples, as can be seen in the legend.

6.3.2 Computation time

In order to compare the two methods, we can look at the results of the previous section. We concluded that for the numerical integration method, 1000 points for discretization is sufficient such that the model error is the dominant one. For the Monte Carlo method, we found that at least 100.000 points are needed to reach a similar accuracy.

The settings that are used for the risk calculations are to calculate the model at 50 magnitudes and 50 distances. Furthermore, unlike for the results that I have generated in the previous section, we do not preselect the 'worst-case scenario', but rather include all scenarios, of which there are a total of 32. Note that these should not be seen as reflecting some specific real-world scenario. They are actually different branches of a logic tree with two choices of four options and one option of two choices, totaling 32 permutations. These represent different sources of randomness and each choice select different variances and means of the reference ground motions. Later in the chain, these should be averaged out again in with certain weights to reproduce the effects of multiple sources of randomness. More about this logic tree can be found in [4].

Next, we also need to compute the results for all 160 zones. This, together with the fact that we previously only used 3 magnitudes and distances, makes it quite a step up for the computer to calculate. In fact, so much so that it would suddenly take multiple days to calculate the results. Therefore, it makes more sense to investigate the scaling rather than results of full runs and then extrapolate the total runtime from this.

For Monte Carlo simulations, the reasonable assumption is that e.g. including 32 scenarios multiplies the runtime time by 32, or doubling the number of magnitudes doubles the computation time. This is simply because adding these generates a new set of reference model parameters to generate results for. Simple tests like these do indeed confirm this hypothesis. With a single scenario chosen and with 5 magnitudes and 10 distances, the time to complete a run is 13.9 minutes, whereas for 10 magnitudes and 10 distances, the time was 29.5 minutes, which is approximately 2.1 times longer. Similarly, if we now include an extra scenario, the time should again be approximately double. Indeed, the run now took 54.4 minutes, or around 1.85 times longer. If we now scale up to all scenarios, 50 magnitudes and 50 distances, we get that a full run at 100.000 samples should take approximately 15 days and 7 hours, assuming the machine has sufficient memory to store all the computations, otherwise there would be extra overhead for writing/reading a disk, which would make it take even longer.

Comparing this to a full run of the numerical integration method, we get a run time of 39 minutes and 52 seconds. This is for all zones, all scenarios and the same 50 magnitudes and 50 distances, and the numerical integration points set to 1000. The results in Figure 23 suggest that even 250 points or perhaps around 500 are more than sufficient to reach the maximum accuracy of the method, which can speed up the method even more.

This means that the numerical integration method is about **553 times faster** for the settings used. Note that previously, the risk calculations used 10.000 samples rather than 100.000 samples which I am comparing my method to. So in their case, the speedup would be around 55 times.

7 Impact of Proposed Changes on Risk Assessment

So far, we have used results that are constrained by always being conditional on a certain scenario with a chosen magnitude and distance. However, the ground motion model is made to ultimately be used for risk analysis purposes. In chapter, we will present results of risk calculations that have gone through the entire model chain. Specifically, we will be investigating the impact that the proposed changes make on the final results. In the previous chapters, we proposed two such changes of which we can measure the impact:

- Using an empirical correlation model for amplification factors
- Using a numerical integration method for calculating the intensity measures

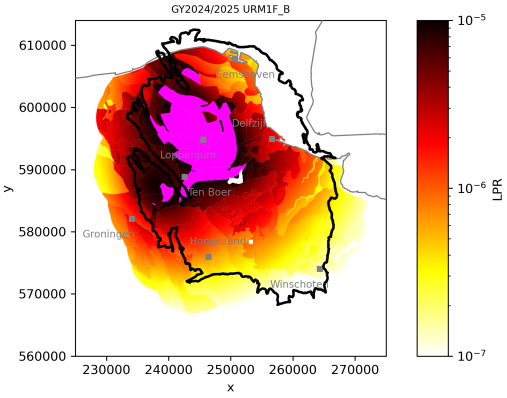
Before we review the impact, it is important to quickly explain what we can expect from these results. The Seismic Hazard- and Risk Analysis (SHRA) for Groningen is a study that evaluates the expected future seismic ground movements and the associated safety risks for residents in the Groningen earthquake region. Its goal is to support decision making for future gas extraction. To do this, the seismic source model can generate multiple scenarios depending on the gas extraction levels. For the coming results, the current scenario of no gas extraction was selected. Combining this model with the ground motion model and fragility model will in the end give us a set of heatmaps of local personal risk (LPR) levels throughout the Groningen area for any specific building type. These risks represent the probability for an individual that stays within such a building for an entire year will die. The accepted level of risk is called the Meijdam standard and is a probability of 10^{-5}, so any area above this will be of special interest. Lastly, another interesting statistic that comes out of the SHRA, is how many buildings in total are endangered, so we will be comparing this as well.

7.1 Empirical Correlation Model

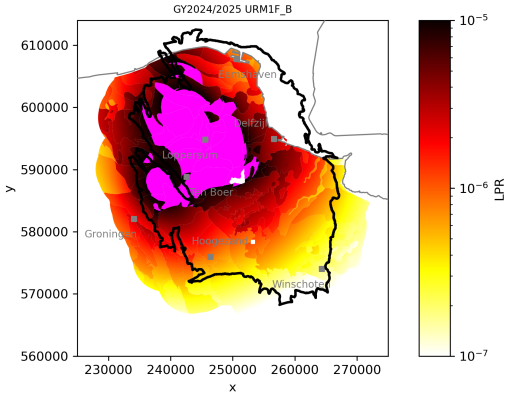
In this section, we will investigate the impact of using different correlation models. Recall that in Chapter 5 we proposed an empirically derived correlation model, rather than, e.g., the previously used full correlation model or consistent correlation model. Recall that even though we call it empirical, it is not actually driven from real-world data, but rather from a simulated data set.

The correlation model we derived had quite strong correlations for neighboring periods, just like the consistent or Baker & Jayaram model [10], but weaker correlation for periods further away. This makes us expect that the results in terms of risk will fit right in between the zero correlation mode and the consistent correlation mode.

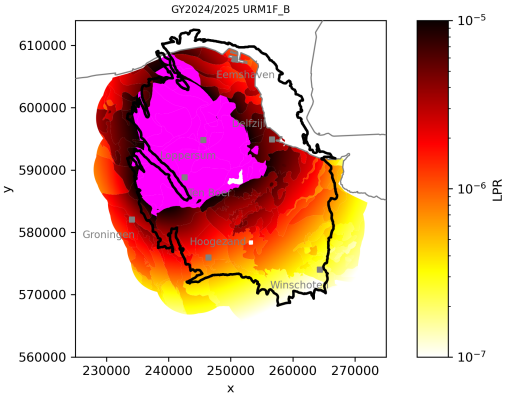
We start by looking at one of the main results from the SHRA, which is a heatmap of the Groningen area with the region where the LPR is above the Meijdam standard is indicated. Because the region exceeding is different per building type, these plots are made for a specific building type. In this case, the building type URM1F_B was selected, which represents the barn section of unreinforced masonry farmhouses. The reason specifically this building type is chosen, is because it has the highest risk of collapsing in general. For many other building types, the region exceeding the standard would be small or nonexistent. We can see such heat maps for the 4 correlation models in Figure 28.



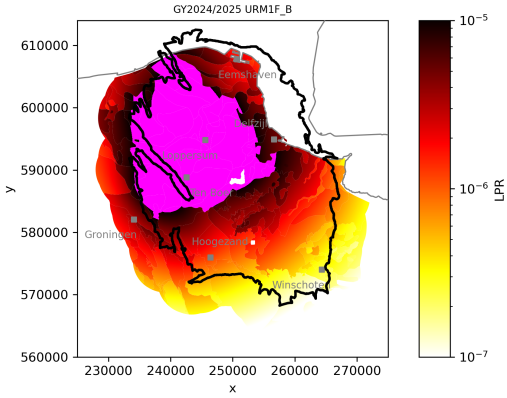
(a) Zero correlation model



(b) Empirical correlation model



(c) Consistent correlation model



(d) Full correlation model

Figure 28: Local Personal Risk (LPR) heatmaps of the Groningen area for the URM1F_B building type for gas year 2024-2025 per correlation model. The pink zone represents the area which has LPR values exceeding the Meijdam standard.

From these heatmaps, we can see that indeed the empirical model fits right in between consistent and zero correlation in terms of the size of the dangerous area. Another conclusion that can be drawn from this is that the choice of dependence model for the amplification factors is in fact quite influential in for the final risk analysis, so more research should go into finding the best dependence model.

For context, it is interesting to share how the risks evolve through time. Figure 28 were all plots for gas year 2024-2025. Since the results we present are for the currently happening scenario, which is no gas production, we can see in Figure 29 that risky region decreases quite a lot in size over time. So according to the model chain, the seismic risks decrease to be a lot smaller in 10 years due to the lack of gas production.

We now look at a more quantitative approach for comparing the correlation models. Much like the analysis we did in Section 6.3, we can compare risk values for different model by looking at their dispersion. This time however, we use LPRs that have been calculated through the entire model chain, rather than for conditional PoE's we looked at for comparing numerical integration and Monte Carlo methods. In Figure 30 we can see the dispersion relation between the consistent correlation model compared to the 3 other models. The reason we pick this as the reference for comparing the other models to, is the only one for which the fragility model is properly calibrated, about which more will become clear in Section 8.1.

From this plot, you can see that the using different correlation models effectively scales the risks, as the different colored point-clouds look very alike and look like scaled versions of each other. We also see that the LPR of the empirical model is in general approximately 50%-80% smaller than the consistent model.

7.2 Numerical Integration Method

In this section we will compare the impact of using the numerical integration method compared to using the Monte Carlo method. As we have seen in Section 6.3, the risks were consistently slightly higher than the Monte Carlo method, due to it not properly modeling the negative skew resulting from the non-linear breakdown effect. Therefore we expect to see that the numerical integration method will produce slightly higher risks.

The results that follow come from using using GMMV7 with standard settings, which is the one presented in the Groningen SHRA reports, but we computed the intensity measures with both the integral method and the Monte Carlo method.

We start by comparing the heatmaps of LPR, which can be found in Figure 31. These heatmaps are once again specifically for the building type URMF_B. Clearly, the two methods are not significantly different. In fact it is quite hard to spot any difference at all. Upon very close inspection one can observe that the numerical integration method indeed has a slightly larger zone exceeding the Meijdam standard. The fact that the two are almost identical is considered a huge success, given that the numerical integration method is significantly faster to compute.

Next, we discuss Figure 32, in which we can observe so called F-N curves, which stands

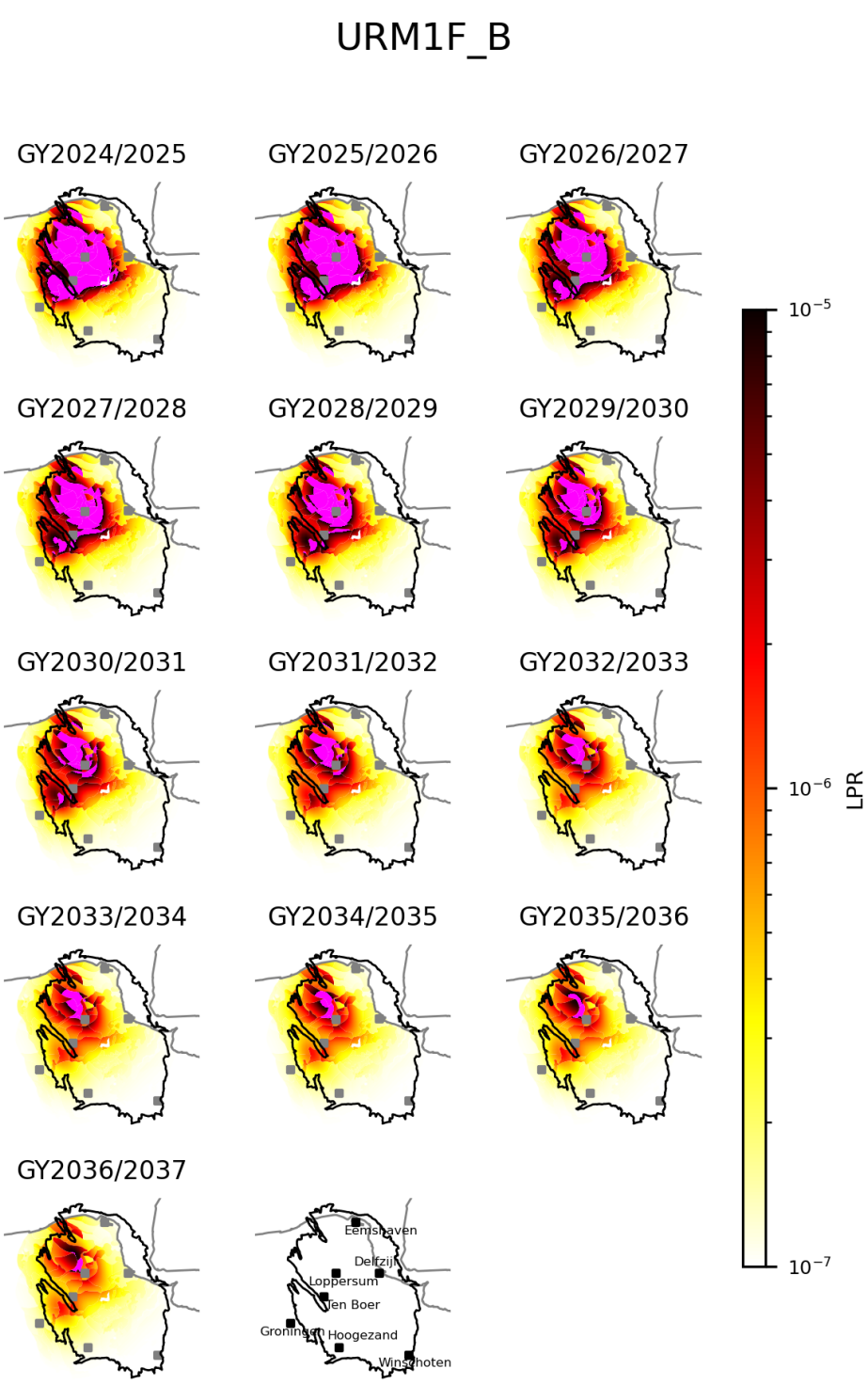


Figure 29: Overview of Local Personal Risk (LPR) heatmaps of the Groningen area for the URM1F_B building type throughout the gas years between 2024-2025 and 2036-2037 for the empirical correlation model. The pink zone represents the area which has LPR values exceeding the Meijdam standard.

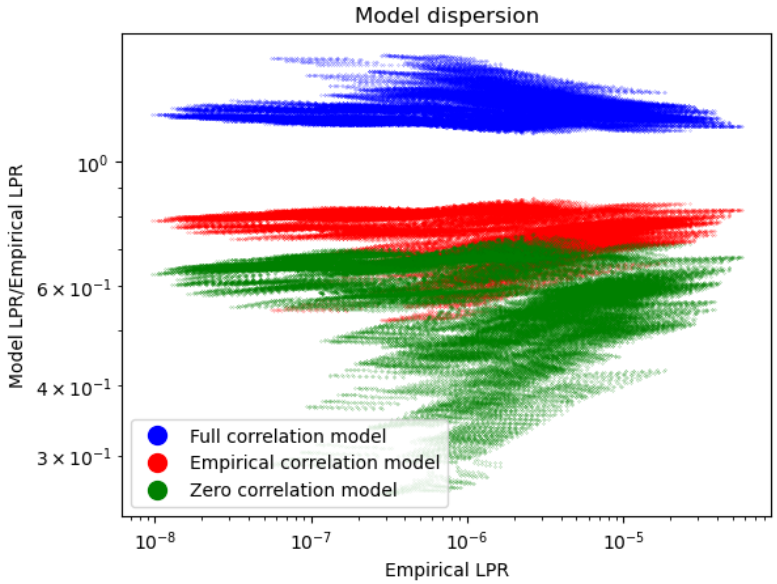


Figure 30: Dispersion relation between LPRs of different correlation models. On the x-axis we plot the LPR of the consistent correlation model and on the y-axis we have the ratio between the LPR of another correlation model compared to the consistent. Each point represents a point on the Groningen map.

for frequency vs. number. Here we can see cumulatively how the numbers of buildings that exceed a certain probability of damage, where:

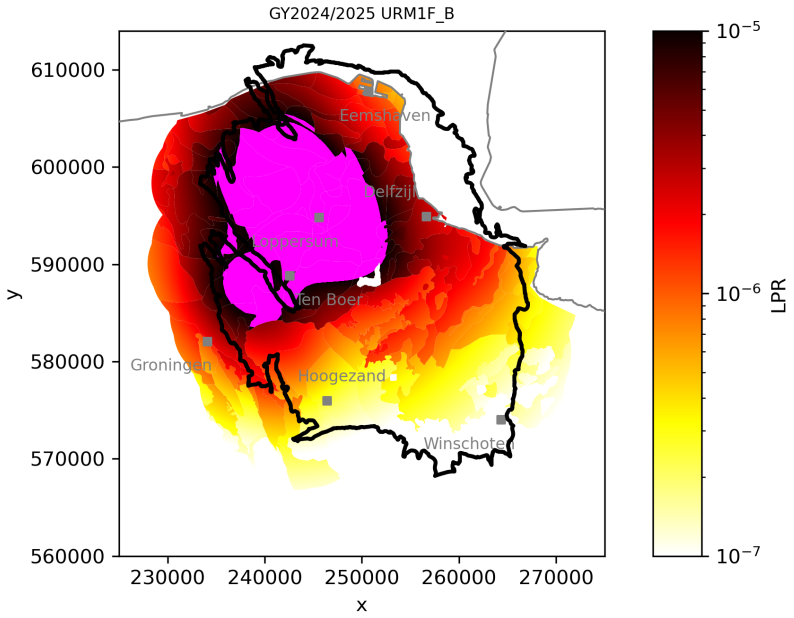
- DS1 is minor cracking and cosmetic damage without significant structural impact.
- DS2 is more pronounced cracking, potential spalling, and slight structural degradation.
- DS3 is severe structural damage with large cracks, spalling, and compromised load-bearing capacity.

The way these plots should be interpreted are as follows: say you see that at 1000 buildings there is a probability of exceedence of 0.6, then that means that 1000 buildings have around 60% chance of being damaged in this year.

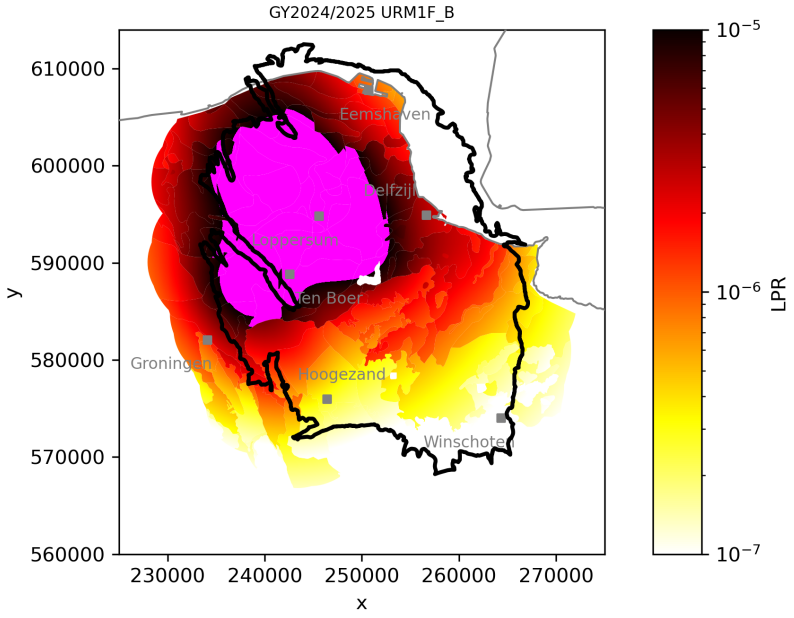
Once again, there is very little observable difference between the two methods. From this data, we can also easily derive the expected number of damaged buildings with a certain state. These can be found in Table 1. These results confirm that there is indeed very little difference between using the two methods in terms of the expected number of damaged buildings.

Table 1: Expected number of buildings (out of ∼150.000) in each damage state for the gas year 2024-2025, calculated with the two different computational methods.

	Monte Carlo	Numerical integration
DS1	918	919
DS2	27	27
DS3	7	7

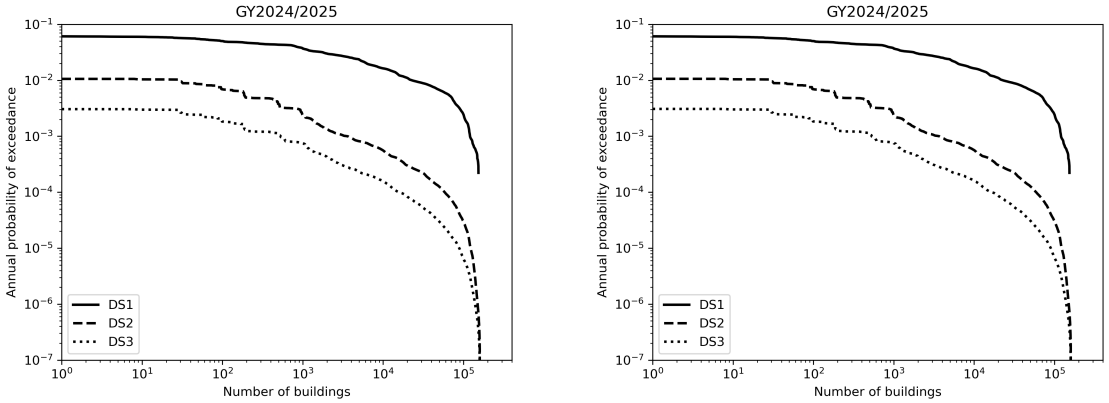


(a) Monte Carlo method



(b) Numerical integration method

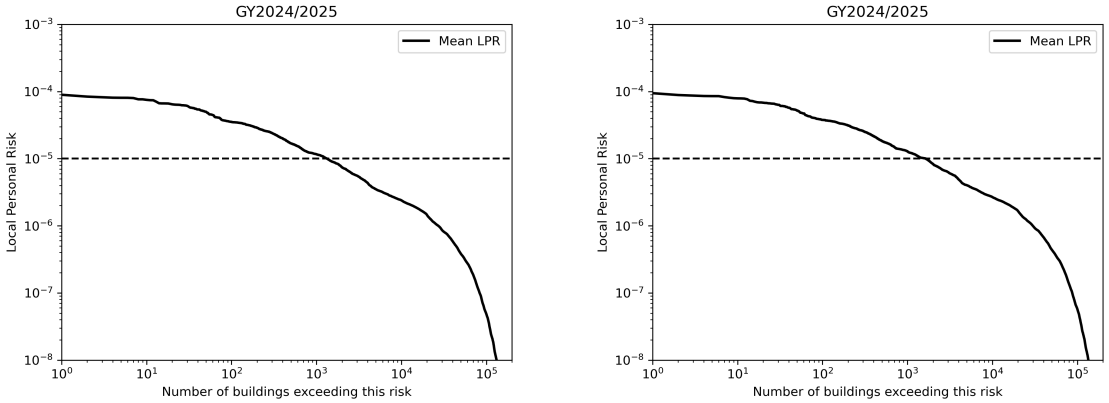
Figure 31: Local Personal Risk (LPR) heatmaps of the Groningen area for the URM1F_B building type. The pink zone represents the area which has LPR values exceeding the Meijdam standard.



(a) Monte Carlo method (b) Numerical integration method

Figure 32: F-N-curves of the amount of buildings in Groningen exceeding certain damage states, calculated using both the numerical integration method and the Monte Carlo method. On the x-axis we see the number of damaged buildings and on the y-axis we see the probability of that number of buildings being damaged, cumulatively.

Now, in Figure 33 we can see the so called mean LPR curves, where the mean is taken over all different branches of the model chain. This figure shows us the number of buildings with a certain mean LPR. The riskiest building appears to have, on average, an LPR of around 1/10.000, meaning that around once in 10.000 years, a person living in the riskiest building in Groningen can expect to have an earthquake as cause of death. The results for where the line crosses the Meijdam norm at 1/100.000 appears similar, however, for the Monte Carlo method, 1330 buildings exceed the norm and for the numerical integration method, 1640 buildings exceed the norm. This is still not that significant of a difference, since this should be viewed with a logarithmic scale in mind, just like the axes are plotted.



(a) Monte Carlo method (b) Numerical integration method

Figure 33: Mean Local Personal Risk (LPR) curves

8 Discussion

8.1 Fragility Model Improvements

It has recently been brought to attention by TNO that the calibration of the fragility model might not have been conducted in the proper way to work for different correlation models. The issue lies with the samples that are used to estimate the b_0 and b_1 coefficients in Definition 1. These are estimated with a linear regression from a certain set of data points, but whether these data points are representative for the ground motion model samples that will later be used as input in the fragility model.

To see what the issue exactly is, we need to understand the way the data for the fragility model calibration is generated. They do this by creating virtual earthquake signals that are the best possible representation of an actual earthquake with a certain set of parameters, like the magnitude and amount of energy released and the length. However, they also though of trying to make the signal be such that the correlation between the spectral acceleration components is such that it resembles the Baker & Jayaram structure [10]. This means the fragility model has the implicit assumption when using the average spectral acceleration as a predictor, that its distribution has the Baker & Jayaram correlation underlying. This means that if you use the fragility model with a correlation different than the Baker & Jayaram correlation, i.e. consistent correlation, you will not get fully accurate results. This means that any GMMV7 calculation with standard settings, which is used for the national SHRA, is also subject to this error.

The mathematical argument for why this happens is simple. Consider the exceedance probability as defined in (8). This probability is in fact conditional on a certain intensity measure distribution I. What happens in the chain is that in the end, we are not interested in the conditional probability of exceedance, but the unconditional version. That is the entire reason why we calculate the intensity measure with the ground motion models: because we want to integrate this function over the density of distribution of I. As explained before, the quantity $\ln \eta_{D|I}(I)$ is implicitly assumed to have Baker & Jayaram correlation, but so if we now do simulations of I, but use a different correlation model, we are not simulating I, but some other random variable, say I'. If we now continue to use this I' in the following calculations, we will end up integrating

$$\int_{-\infty}^{\infty} f_{\ln(D)|I=i}(d|i) f_{I'}(i) \mathrm{d}i \neq f_{\ln(D)}(d).$$

Even though this obviously does not work, this is in fact what is happening at the moment. All the results that follow discussing different correlation models will have the same error whenever we talk about results that are progressed through the chain. The size of the error introduced by using the model in this de-synchronized way is currently unknown and it should be researched how to make the fragility model compatible for future iterations of the model chain, as well as for other seismic hazard modeling research. This error has crept in the model due to the history of different experts in both of the fields of fragility modeling as well as ground motion modeling working on this model separately and not realizing that a change in the correlation model for the ground motions will implicitly affect the fragility model.

We have no gauge on how much this de-calibrated model affects the ultimate risk- and hazard analysis. That being said, it was not the purpose of this project to do anything with the fragility model, so we did not set out to fix this problem, rather we use the model as it has been used in the past. Still it is important to remark this issue so hopefully it will be resolved in the future.

Furthermore, before this issue is resolved, it is unwise to use different correlation models for the GMM than the one the fragility model is trained for. This means that the results presented in, e.g., Section 7.2 are all affected by this.

8.2 Normality of the average spectral acceleration

As was previously pointed out, Approximation 1 from Section 6.2.2 is not always the most accurate, as sometimes skewness values up to -0.25 are observed in the Monte Carlo runs. Furthermore, we say that this skewness has quite a profound effect on the risk calculations, making the numerical integration method have higher risks than the Monte Carlo method consistently.

Initially, it made sense to attempt to model this skewness with the help of numerical integrations like before, but it turned out that it is a lot more complex to calculate third moments than second moments analytically. Where the second moment has 100 terms that need to be summed up and divided, the third moment has 1000 of these terms, all of which are more complex than the second moment terms. So rather than go down this rabbit hole, it makes more sense to calculate the skewness resulting from many different choices of reference model for all the different zones and then use the Monte Carlo distribution to find the empirical skewness. From this one can find the best predictors for the skewness and build a model that takes the input parameters like reference mean, reference variance and amplification model parameters as inputs, and uses these to predict the resulting skewness.

Given the first two moments from the numerical integration model and the skewness from the hypothetical model we just described, one can then use the skew-normal distribution, which takes three input parameters - location, scale and shape - which can be derived in a one-on-one from the three moments we have. So then we basically change Assumption 1 to have a skew-normal assumption. Given one has such a model that is sufficiently good at predicting skewness, it is a simple task to adjust the model to use this distribution rather than the normal distribution.

Out of curiosity for how difficult it would be to make such a model, I made an attempt to do so. Firstly, due to the complexity of the many branches and zones, I built a model only for one zone and a select part of the reference model scenarios. The details of how this model was made are not particularly relevant, but the main gist was that it used the mean of the reference median $\overline{m}$ across the 10 periods and the mean of the b_i-parameter, $\overline{b}$ of the amplification model, see Eq. 3. The latter controls the slope of the curvature in the linear breakdown, hence my idea that it would contribute to skewness. From the different skewness values I had collected, it appeared that there was a clear linear relationship for the skewness with $\overline{b}$, given some $\overline{m}$ values. The slope of this line could then in turn be derived from the given $\overline{m}$.

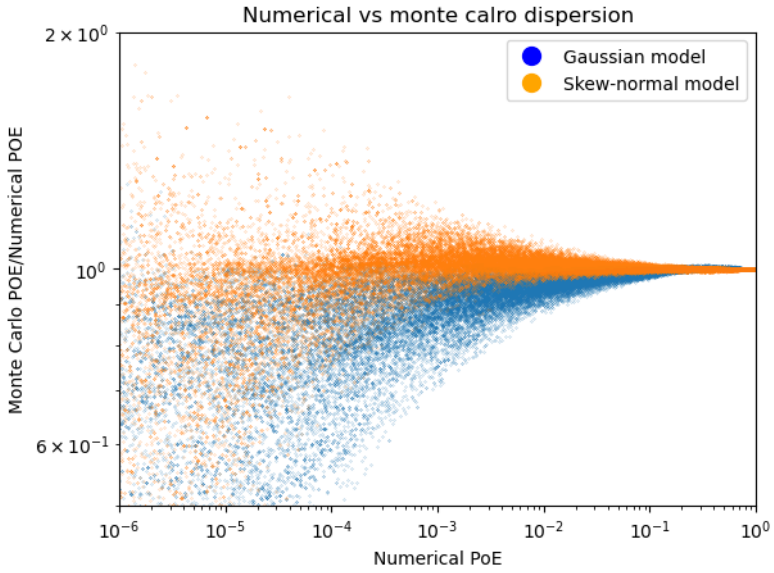


Figure 34: Comparison of dispersion using the Gaussian model vs the Skew-Normal model. On the x-axis we see the PoE's of the numerical integration methods and on the y-axis we see the ratio of the PoE's of a Monte Carlo run with 1 billion samples to the numerical PoE's on the x-axis. The orange points are the results of the skew-normal method and the blue points are from the Gaussian model.

We can now compare the effect of using this model to predict skewness, compared to the regular Gaussian model we had before. In Figure 34, we can see a dispersion plot for the two models on the data that it was trained for. At first sight, the skew model appears to work exceedingly well, reducing the maximum deviation from a ratio of 1 in the region from 10^{-3} to 1 drastically - the region we are most interested in. Furthermore, it is now a symmetric dispersion, which means we eliminated a source of model error.

However, this model is not representative for the entirety of the model chain. If we attempt to use the same model on different zones, then it sometimes works, but sometimes makes the predictions a lot worse. However, the fact that it was possible to successfully create a predictor for skewness for a subset of the model, implies that it is quite likely that there exists a simple way to do it for the entirety of the model. Perhaps by creating 160 of these simple models, or perhaps by adding in another predictor, or perhaps by using machine-learning techniques to find an optimal predictor. Due to time constraints of the project, no further attempts to calculate this were made. We do however recommend this to be implemented in future iterations of the Ground Motion Model.

9 Conclusion

In this thesis, we made progress for Groningen's Ground Motion Model on 3 fronts. Firstly, on the calibration front, we looked at generating different sets of parameters using the same data as the original model authors. Due to the ambiguity in the explanation of their methods to find these parameters, we were unable to recreate the parameters they found. In fact, we found that the maximum likelihood parameters were significantly different from the ones they present. Whether these parameters are better, simply because they have a higher likelihood, is hard to know. There can be many reasons for them choosing different parameters, coming from expert judgement-based insights. However, for future developers of the GMM, we do suggest to investigate whether it would be better to use maximum likelihood-based parameters.

Secondly, we looked into model validation. For this, we did a critical evaluation of the model's lognormal assumption. The results on this were not clearly pointing to either accepting or rejecting this assumption and we pointed out that it would be an interesting topic for further research to investigate other distributions that allow for skewness. Furthermore, we also investigated the best dependence structure for the amplification factors. We came to the conclusion that using a copula model could be superior, but for the reason of keeping the model simple, we stuck to using a matrix to model the dependency, as was done before. The empirical correlation matrix we found was significantly different from the ones used before. However, we cannot conclude that using the empirical model over any of the previously used matrices is better. Moreover, it is not advisable to change the dependence structure at the moment, because there needs to be a recalibration of the fragility model in the first place to allow other models have accurate risk assessments.

Third, we looked at a newly developed calculation method for calculating the average spectral acceleration. This method used numerical integration rather than the previously used Monte Carlo sampling technique. This method has its merits in computational speed, decreasing computation time by a factor of 553, but sacrifices some accuracy for this. Specifically, it was noted that the Monte Carlo method shows slight negative skewness, which the numerical integration method cannot reproduce. We did, however, recommend to implement a model that can include skewness, after a brief attempt at making such a model. This model showed promising results and gives a nice step towards potentially including this in future iterations of the model.

Lastly, we also looked at the impact on risk calculations of the recommended changes to the model. The results were as expected. Changing the correlation model had quite a significant impact on risk assessment. In contrast, using the numerical integration method shows very little impact on the risk assessment, which is a positive result, since it is much faster.

References

[1] J. J. Bommer, B. Dost, B. Edwards, *et al.*, *V4 Ground-Motion Model (GMM) for Response Spectral Accelerations, Peak Ground Velocity, and Significant Durations in the Groningen Field*. Jun. 2017.

[2] J. J. Bommer, B. Dost, B. Edwards, *et al.*, *V5 Ground-Motion Model for the Groningen Field*. Mar. 2018.

[3] J. J. Bommer, B. Dost, B. Edwards, *et al.*, *V6 Ground-Motion Model (GMM) for Induced Seismicity in the Groningen Field*. Dec. 2019.

[4] J. J. Bommer, B. Dost, B. Edwards, *et al.*, *V7 Ground-Motion Model for Induced Seismicity in the Groningen Gas Field*. Feb. 2022.

[5] TNO, *Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen*. Apr. 2022.

[6] P. P. Kruiver, E. van Dedem, R. Romijn, *et al.*, "An integrated shear-wave velocity model for the groningen gas field, the netherlands," *Bulletin of Earthquake Engineering*, vol. 15, no. 9, pp. 3555–3580, Sep. 2017, ISSN: 1573-1456. DOI: 10.1007/s10518-017-0105-y. [Online]. Available: https://doi.org/10.1007/s10518-017-0105-y.

[7] J. Bommer, B. Dost, B. Edwards, *et al.*, "Developing a model for the prediction of ground motions due to earthquakes in the groningen gas field," *Netherlands Journal of Geosciences*, vol. 96, s203–s213, Dec. 2017. DOI: 10.1017/njg.2017.28.

[8] J. J. Bommer, B. Dost, B. Edwards, *et al.*, *Development of Version 2 GMPEs for Response Spectral Accelerations and Significant Durations from Induced Earthquakes in the Groningen Field*. Nov. 2015.

[9] A. Rodriguez-Marek, P. Kruiver, P. Meijers, *et al.*, "A regional site-response model for the groningen gas field," *Bulletin of the Seismological Society of America*, vol. 107, pp. 2067–2077, Oct. 2017. DOI: 10.1785/0120160123.

[10] J. W. Baker and N. Jayaram, "Correlation of spectral acceleration values from nga ground motion models," *Earthquake Spectra*, vol. 24, no. 1, pp. 299–317, 2008. DOI: 10.1193/1.2857544. eprint: https://doi.org/10.1193/1.2857544. [Online]. Available: https://doi.org/10.1193/1.2857544.

[11] P. J. Stafford, A. Rodriguez-Marek, B. Edwards, P. P. Kruiver, and J. J. Bommer, "Scenario Dependence of Linear Site-Effect Factors for Short-Period Response Spectral Ordinates," *Bulletin of the Seismological Society of America*, vol. 107, no. 6, pp. 2859–2872, Oct. 2017, ISSN: 0037-1106. DOI: 10.1785/0120170084. eprint: https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/107/6/2859/3992696/bssa-2017084.1.pdf. [Online]. Available: https://doi.org/10.1785/0120170084.

[12] TNO, *Status of the TNO Model Chain Groningen per October 1, 2021 and recommendations for the public Seismic Hazard and Risk Analysis 2022, update 2022*. Apr. 2022.

[13] H. Crowley, R. Pinho, J. van Elk, and J. Uilenreef, "Probabilistic damage assessment of buildings due to induced seismicity," *Bulletin of Earthquake Engineering*, vol. 17, Aug. 2019. DOI: 10.1007/s10518-018-0462-1.

[14] Q. Kong, J. Liu, X. Wu, and C. Yuan, "Seismic fragility estimation based on machine learning and particle swarm optimization," *Buildings*, vol. 14, no. 5, 2024, ISSN: 2075-5309. DOI: 10.3390/buildings14051263. [Online]. Available: https://www.mdpi.com/2075-5309/14/5/1263.

[15] L. Eads, E. Miranda, and D. G. Lignos, "Average spectral acceleration as an intensity measure for collapse risk assessment," *Earthquake Engineering & Structural Dynamics*, vol. 44, no. 12, pp. 2057–2073, 2015. DOI: https://doi.org/10.1002/eqe.2575. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.2575. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.2575.

[16] E. Bojórquez, I. Iervolino, A. Reyes-Salazar, and S. E. Ruiz, "Comparing vector-valued intensity measures for fragility analysis of steel frames in the case of narrow-band ground motions," *Engineering Structures*, vol. 45, pp. 472–480, 2012, ISSN: 0141-0296. DOI: https://doi.org/10.1016/j.engstruct.2012.07.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141029612003604.

[17] P. Kruiver, *Personal communication*.

[18] E. Rathje and A. Kottke, *Strata*. [Online]. Available: https://www.geoengineer.org/software/strata.

[19] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, Aug. 1989, ISSN: 1436-4646. DOI: 10.1007/BF01589116. [Online]. Available: https://doi.org/10.1007/BF01589116.

[20] P. Bazzurro and C. A. Cornell, "Nonlinear soil-site effects in probabilistic seismic-hazard analysis," *Bulletin of the Seismological Society of America*, vol. 94, no. 6, pp. 2110–2123, Dec. 2004, ISSN: 0037-1106. eprint: https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/94/6/2110/3584691/2110_946_03216.pdf.

[21] J. K. Blitzstein and J. Hwang, *Introduction to Probability Second Edition*. 2019.

[22] H. Crowley and P. R., *Report on the v7 fragility and consequence models for the Groningen field*. Feb. 2020.

Appendix 1

Table 2: Model parameters that are constant for throughout all zones.

T (s)	c_T	α_T	β_T
0.01	0.004	0.0015	0.015
0.025	0.004	0.0013	0.0133
0.05	0.004	0.0009	0.0095
0.075	0.004	0.0018	0.0176
0.1	0.0188	0.0058	0.0577
0.125	0.0625	0.0118	0.1178
0.15	0.108	0.0177	0.1767
0.175	0.1715	0.0242	0.2419
0.2	0.256	0.0309	0.3086
0.25	0.5	0.0359	0.3589
0.3	0.5	0.0384	0.3837
0.4	0.5	0.0368	0.3679
0.5	0.5	0.0314	0.3142
0.6	0.5	0.0279	0.2786
0.7	0.5	0.0274	0.2739
0.85	0.5	0.0235	0.235
1	0.5	0.0159	0.1594
1.5	0.5	0.0092	0.0924
2	0.5	0.0053	0.0531
2.5	0.5	0.0033	0.0329
3	0.5	0.002	0.0201
4	0.5	0.0013	0.0126
5	0.5	0.0009	0.0088