Change in frequency of environmental catastrophes – a Bayesian analysis

Nikolaus Wehofer

Vienna University of Economics and Business

A thesis submitted at the Institute for Statistics and Mathematics in 2023

Supervising Professor – Malsiner-Walli, Gertraud, PD Mag.Dr., M.Stat.,

Co-supervising – Peter Knaus, M.Sc (WU)

Contents

Abstract **v**

1 Introduction **1**

2 Literature review **3**

3 Methodology **7**

3.1 Data description 7

3.2 Mathematical description of the model 12

4 Results **21**

4.1 The catastrophe put in numbers 21

4.2 Density distribution with probability intervals 23

4.3 Change in mean 25

5 Conclusion **31**

Bibliography **33**

Appendices **33**

A Traceplot **35**

B Trankplot **37**

Abstract

Linked to a world-wide discussion, this study investigates one aspect of the consequences of climate change by looking at the increase in frequencies of environmental catastrophes. Particularly, this paper gives an insight into how the occurrence rate of hydrological, climatological and meteorological disasters have changed up to the year 2020 and in what manner this rate will most likely develop in the intermediate future until 2050.

To answer this question, a data-set from a project governed by NASA's ESDIS department was used. The statistical framework applied to analyze the data-set is Bayesian. More precisely, an indexed Bayesian-covariate-multilevel-model was developed and used.

The study concludes that the frequency in which environmental catastrophes occur, increases in an exponential manner for all three disaster groups examined. The model finds that all three catastrophes grow exponentially at a rate of approximately 3-4% per year. This will result in a 30-fold increase for some disaster groups within the next 20 years, compared to when these catastrophes were first reliably documented.

Chapter 1

Introduction

> I was caught in the middle of a railroad track. **Thunder**.
>
> I looked around, and I knew there was no turning back. **Thunder**.
>
> My mind raced, and I thought “What could I do?” **Thunder**.
>
> And I knew there was no help, no help from you. **Thunder**.
>
> -AC/DC

When AC/DC first published their song “Thunderstruck” back in 1990, saving the environment from a climate-catastrophe was not on top of the world’s agenda. Why would it? 1990 was the year when the Berlin Wall was officially torn down and when humankind was first able to gaze at galaxies far far away with the Hubble telescope. Nonetheless, the environment was not completely irrelevant. After all, it was in 1986 when the Austrian Green party was established. However, it was different topics that occupied our minds. As we now know, the climate is changing. This is not only due to natural cycles but also as a result of human activity. These changes occur in a much more serious manner than expected.

But how exactly has our climate changed the past 50 years? And how will it continue to change? Answering these questions is the aim of this paper, narrowed down to its research question: How did the frequency of natural catastrophe occurrences change since 1950 and how will it most likely develop until 2050? Particularly, it covers the development from three environmental catastrophe-groups over time, which are hydrological, climatological, and meteorological disasters. By knowing how their occurrences have changed in the past, it is possible to model how they will most likely change in the future. To answer this research question, the paper will proceed in the following manner.

In Chapter 2 I will look at existing literature and locate this research within the academic field. Chapter 3 contains the methodology section. In this chapter, the data-set is introduced, which originates from NASA's ESDIS department. Then, each step of how the data are transformed, is discussed. Thirdly, a detailed mathematical description of the statistical model follows. This part proceeds rather slowly because it maps out every step taken towards developing the model. The resulting model is an indexed-Bayesian-covariate-multilevel-model. In the last part of the methodology section a brief look inside the statistical model is taken, ensuring that the model is able to reliably explore the posterior distribution. If you are merely interested in the results, it is possible to start reading at chapter 4. In this chapter, the model's outcome is presented and interpreted. To achieve this, the numerical output from the model is discussed and put into perspective. Afterwards, distributions of catastrophe-occurrences are plotted. Lastly, the frequency of how often environmental disasters occurred over time from 1950-2020 are plotted, including a forecast until 2050.

The most important insight from this paper is that it is possible to say that in the future *more and more* environmental catastrophes are going to happen, considering all assumptions. *More and more* can mathematically be described as an exponential increase over time, which will have a strong impact in the intermediate future. The paper's contribution lies in putting the exponential increase into exact probability intervals. To be specific, some disaster groups, compared to when catastrophes were first reliably documented, will 30-fold within the next 20 years. Or to put it in the words of AC/DC again:

> I looked around, and I knew there was no turning back.
>
> My mind raced, and I thought "What could I do?"

Chapter 2

Literature review

The aim of this paper is to investigate the changes of occurrence-frequencies for environmental catastrophes, as well as establishing a forecast for the intermediate future. This question is frequently and heavily debated in climate change discussions. At the same time however, it is a question that has been prevalent for a long period of time in the area of insurance premiums. Therefore, two very different domains of research ask a similar question.

Environmental disaster modeling in the field of risk analysis

Commonly whole departments of banks or insurances are filled with staff that is only concerned with risk analysis. To give a concrete example, imagine an insurance company that is asked to insure a holiday house at the seaside. If the premiums for this insurance does not price in an increase in the likelihood of coastal flooding, it will undervalue the risk premiums that have to be paid for the house. At the event of a flooding, the insurance group could go bankrupt, if this was a systematic error. To avoid this mistake, Grossi et al. (2005) describes the process of including the likelihood of natural hazards in the following way:

1. Identify expected environmental hazards that could negatively impact the insured object and model its likelihood.
2. Find the important vulnerabilities for the insured object.
3. Calculate the expected loss per vulnerability, accounting for the likelihood with which a natural hazard can happen.

Step 1 is also called the hazard module. To put into perspective, from all three steps displayed this thesis will be limited to building such a hazard module without proceeding to step 2 or 3. Economists are also interested in the risk of environmental catastrophes, since disasters can be the causes for economic shocks, as can be read in Hallegatte (2014). In this book the frequency of disasters is investigated for the means of finding under what circumstances environmental hazards had negative impacts on the economy. By this, an increase in the frequency of hazards is of interest. Thirdly, institutions such as the OECD publish papers and invest in research about the connection of environmental catastrophes and its link to the economy, both on the private as well as the public sector (cf. OECD (2005)). From this literature examples a limitation for this thesis can be drawn, since this paper will focus on modeling the number of natural hazards without looking at economic implications. Not having an economic focus allows this thesis to have a broader perspective on the topic without having to narrow it down towards economical questions.

Environmental disaster modeling in the ecology sector

The second big area that is trying to model and predict environmental catastrophe occurrence frequencies is the ecology sector. One model published in the magazine The Economist for example tries to capture the different likelihoods of extreme temperatures, as well as the likelihood of occurrences for floods. For this purpose, distributions and how they changed over time are calculated, which is a similar method used in this thesis (cf. The Economist (2023)). It is an analog approach, since this paper also proceeds by modeling distributions of occurrence frequencies. A difference however is that this paper goes beyond separately modeling the change of frequencies from year to year. Instead, the change in frequency is incorporated into the model. Next, in the area of environmental modeling an important question is what the effect of changes within the environment has on the frequency of natural hazards. One example that gained recent popularity is the Thwaites glacier that will – if it collapses – cause a rise of 60cm in the sea level. In comparison, in the 18th century there was an increase of 2cm, in the 19th century an increase of 6cm an in the 21st century an increase of 19cm. The effect would be devastating since its forecast predicts that in the long-term the habitable area on planet earth will drastically decrease due to permanent flooding (cf. Gertsch & Krogerus (2023)). However, most scenarios in environmental modeling focus on trend-research, that is, holding an increase in sea-level constant and then calculating its consequences. One research from 2020 concludes that if the sea level continues to increase in the same trend, it did over the past years, this will cause an exponential increase in the odds of flooding in coastal areas in America. Specifically, the odds of flooding will – according to the model – double every 5 years (cf. Taherkhani et al. (2020)). The model this thesis is concerned is limited in this aspect, since it looks at changes over time and not depended on sea-level rises.

Chapter 3

Methodology

3.1 Data description

Data origin

Finding reliable and public data can be difficult, especially when general interest is not strong enough. This used to be true for environmental topics for a long period of time. However, due to an immense increase in awareness about our environment it got easier to access trustworthy and high-quality information. One source that makes great effort in providing environmental data freely accessible is the EOSDIS, the Earth Observing System Data and Information System. EOSDIS is a program managed by NASA's ESDIS (Earth Science Data and Information System) project, which transforms satellite and stationary weather data into statistical usable data-sets (cf. Nasa ESDIS (2023)).

Data structure

The raw data from EOSDIS contain 45 categories, ranging from disaster type (such as drought, earthquake or epidemic) to details such as total deaths. To give a first insight in what data the paper is concerned with, in the following the first seven and the last seven entries for chosen variables are shown in Table 3.1 and Table 3.2.

Table 3.1: *The first 7 rows from the raw data-set for chosen variables.*

Year	Disaster Group	Disaster Type	Continent	Total Deaths
1900	Climatological	Drought	Africa	11000
1900	Climatological	Drought	Asia	1250000
1902	Geophysical	Earthquake	Americas	2000
1902	Geophysical	Volcanic activity	Americas	1000
1902	Geophysical	Volcanic activity	Americas	6000
1903	Geophysical	Mass movement (dry)	Americas	76

Table 3.2: *The last 7 rows from the raw data-set for chosen variables.*

	Year	Disaster Group	Disaster Type	Continent	Total Deaths
16119	2021	Hydrological	Flood	Asia	13
16120	2021	Hydrological	Flood	Asia	11
16121	2021	Hydrological	Flood	Africa	31
16122	2021	Biological	Epidemic	Africa	131
16123	2021	Hydrological	Flood	Europe	NA
16124	2021	Hydrological	Flood	Africa	NA

Tables give a good first impression of structural features of a data-set. However, for better understanding what the data-set contains, further work is required. The simplest way of gaining more insight is by plotting the total sums of incidents. Figure 3.1 sums up how many environmental disasters occurred over all continents, starting from 1900 until the year of 2021.

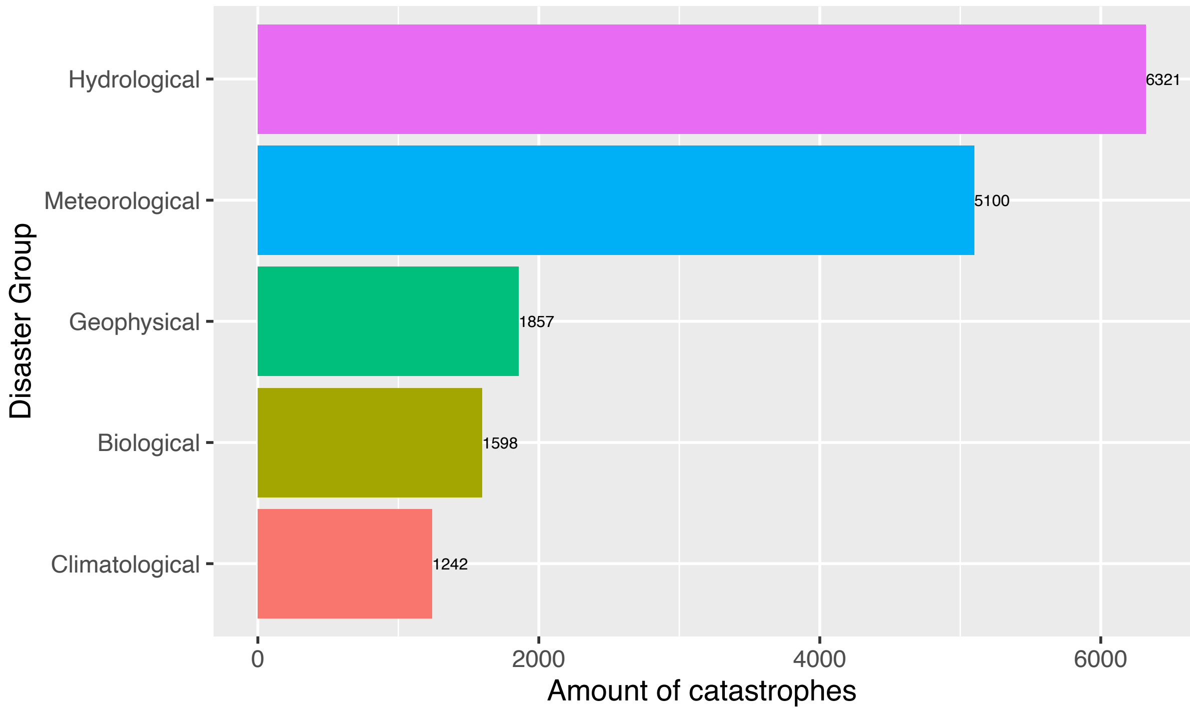

Figure 3.1: *The absolute amount of grouped natural catastrophes from 1900 until 2021.*

Taken as an example, there were a total amount of 6321 hydrological disasters over all continents for the whole recorded time span. Dividing 6321 catastrophes by the sum of years from 1900-2021 and by months, this makes an average of 4.35 catastrophes per month. This does not sound too alarming, right?

Calculating an average can be deceiving, since it does not echo information about an increase in catastrophe frequencies, neither does it reflect how these frequencies vary over decades. This big difference in frequency over time can be seen in Figure 3.2, which displays the development of the same data as in Figure 3.1 per year.

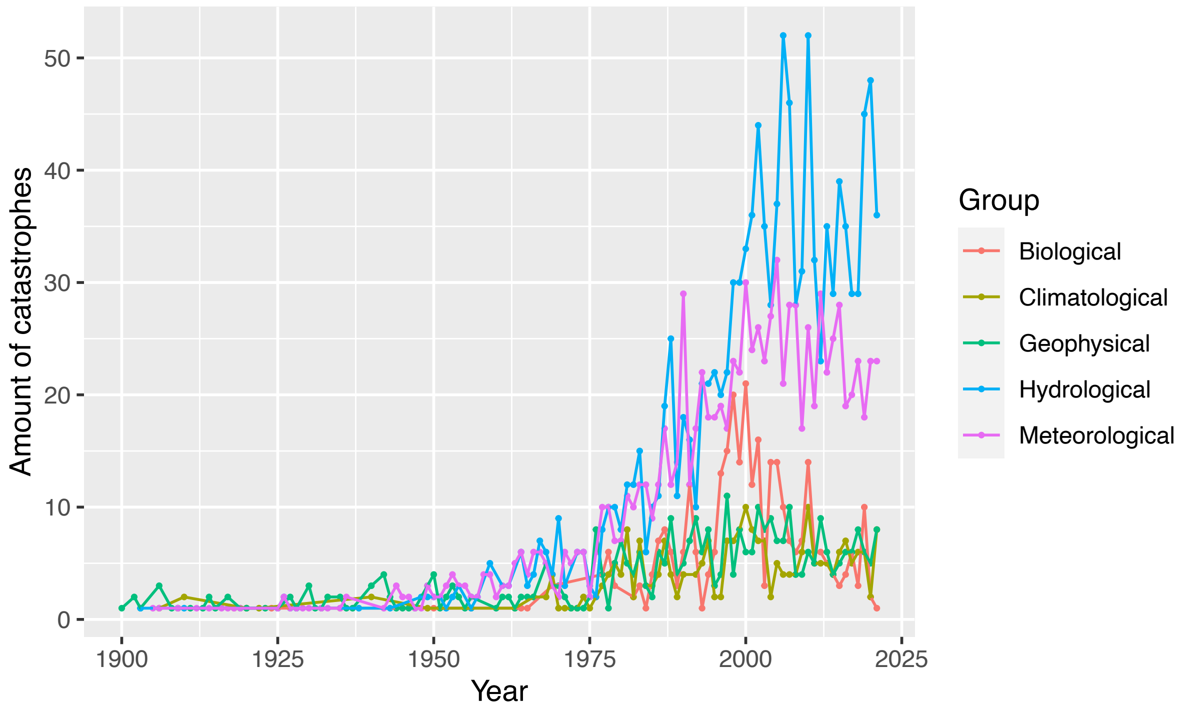

Figure 3.2: *The development of natural catastrophes per year.*

In Figure 3.2 we can see that from 1900-1950 there are very little data-points. This could either be due to few natural catastrophes before 1950 or due to poor data collection. Looking at geophysical disasters we can see that there were much less counts before 1950. Since geophysical accidents (earthquakes, volcanic outbursts) are random in their occurrence it is more likely that data collection was poor. Before modeling the data, they have to be transformed and cut, which will be the task of the next section.

Data transformation

The original data-set contains 45 variables. Not all of them are relevant for answering the research question. The first step of building the model is to choose which variables to keep. This choice is guided by the research question: how did the frequency of natural catastrophe occurrences change since 1950 and how will it most likely develop until 2050? For this purpose, only the group of climatological, hydrological and meteorological disasters are used. Geophysical accidents are excluded due to an independence from human-influence. Biological catastrophes are excluded from the analysis because the absolute frequency dropped drastically after the year 2000, probably due to an increase in international hygiene efforts and the development of efficient vaccines towards the end of the last century.

In order to avoid too small data-collections per year, 5-year intervals per disaster-group are created. Because the data set was made by NASA's EOSDIS only the continent of America[1] will be used in the analysis, as it is most likely that NASA has the best data for their own continent. Moreover, as shown in the previous section, the beginning entries in the data set were cut, since from 1900 until around 1950, as can be seen in Figure 3.2, there are little to no data.

All requirements above lead to a data-set that starts with the year 1950. From there on, data-points for each disaster group are summed up within 5-year intervals. What follows is a data-set that looks like this:

[1]Data from the whole continent of America are used, including North- Central- and South America.

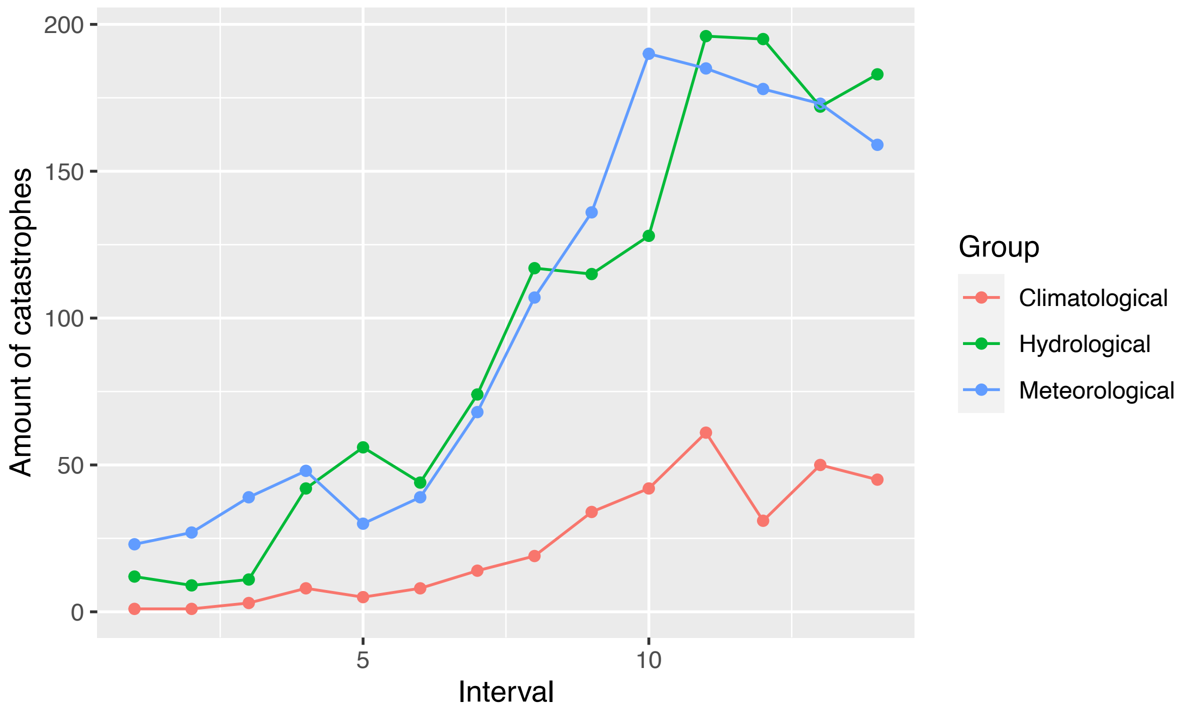

Figure 3.3: *Evolution of natural disasters in 5-year intervals for the period 1950-2020.*

Lastly it is important to briefly feature what can be found inside each disaster group, since each contains two types of disasters. Below the three groups with two disaster types are first summed up in a table and then plotted.

Table 3.3: *The climatological group consists of droughts and wildfires.*

Climatological catastrophes	Sum
Drought	770
Wildfire	470

Table 3.4: *The hydrological group consists of floods and landslides.*

Hydrological catastrophes	Sum
Flood	5545
Landslide	776

Table 3.5: *The meteorological group consists of extreme temperatures and storms.*

Meteorological catastrophes	Sum
Extreme temperature	603
Storm	4496

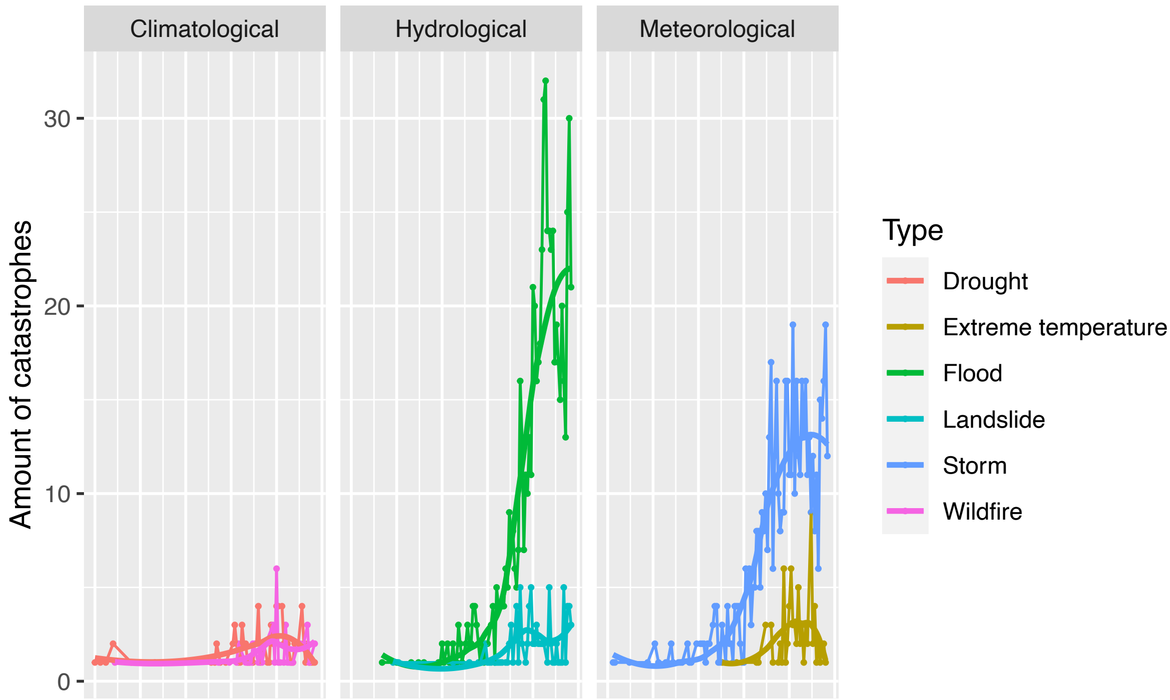

Figure 3.4: *Development of yearly grouped catastrophes in America, split up by disaster group and type.*

The tables and Figure 3.4 are important to keep in mind, since they indicate that there are differences in how often each disaster type happens within each group. In the first group of climatological disasters the curves look rather flat because all three graphics are plotted against the same y-axis scale0. This may give a false impression of overall little climatological catastrophes, which is not the case as can be seen in the 5-year interval Figure 3.3.

Figure 3.4 highlights that for several disaster types there were some intervals with poor data collection. For years, where there were to expect extreme temperatures, missing values are found. For this reason, NA-values will be replaced by their most recent value, since it is more likely that values are closer to their previous values than falling to 0. In case the first entry of a disaster-type starts with an *n/a* it is treated as a 0.

3.2 Mathematical description of the model

3.2.1 Bayesian inference

The statistical theorem applied in this thesis is Bayesian data analysis. A Bayesian model begins by assuming a set of possibilities for each event that is to be modeled. These are prior possibilities that emerge from an educated guess. Then the model updates the prior possibilities in light of data-input, by making an experiment and looking how the world behaves in an experimental context. This produces the posterior plausibilities. The process can be seen as a learning process, called Bayesian updating (cf. Richard McElreath (2020), p. 37) The mathematical description of a Bayesian models arise from Bayes' Theorem. Put into mathematical form we can write

$$P(A|B) = \frac{P(B|A) \text{ x } P(A)}{P(B)},$$

which is the same as to say that

$$Posterior = \frac{\text{Probability of the data x Prior}}{\text{Average probability of the data}}.$$

This Theorem says that the probability of any event A considering the data in B equals to the product of the relative plausibility of the data, conditional on A and the prior possibility of A, divided by the average probability of the data. The denominators' function is to standardize the posterior, to ensure it integrates to 1 (cf. Richard McElreath (2020) p. 37).

The main advantage of Bayesian statistics is that the whole process is intuitive. Simply put, Bayesian updating is assuming a probability from our prior knowledge about an event and then looking at how things are when we look into the world. This process is the same that happens inside humans all the time. Essentially, every time we learn something new or have an *aha!* moment this model has the power to explain what happened inside our minds. For example, if we assume to know a person well (prior belief), who then acts in an unexpected unfavorable way (probability of the data), we have to update our belief about this person (posterior).

3.2.2 The probability distribution

The goal is to model the amount of environmental catastrophes, depending both on the type of the disaster (d), as well as on its interval (t). To start with a probability distribution that fits this purpose, a distribution that restricts results to integers would be beneficial, since it would not make sense to expect $1/2$ catastrophes. One distribution that meets this specification is a special version of the binomial distribution, the Poisson distribution. This distribution allows to model binomial events, without having to know in beforehand the number of resulting trials N. Another important aspect of the Poisson distribution is that it only has one parameter λ

describing its shape. λ at the same time defines the expected mean and variance (cf. Richard McElreath (2020), p. 346). Let $y_{t,d}$ be the number of disasters of type d at time point t. It is modeled as

$$y_{t,d} \sim \text{Poisson}(\lambda_{t,d}), \tag{3.1}$$

where

$$\log(\lambda_{t,d}) = \alpha_d + \beta_{\text{d}} * t \quad . \tag{3.2}$$

3.2.3 The link function

To embed a linear model for λ, a link-function is required. The log-link function maps the parameter λ for positive real values onto a linear model. Restricting $y_{t,d}$ to positive counts makes sense, since it would be unreasonable to expect a negative amount of catastrophes. Taking a log-link function is an important choice, since applying it means assuming that the parameters value is the exponentiation of the linear model. Solving for $\lambda_{t,d}$ yields the inverse link of

$$\lambda_{t,d} = \exp(\alpha_d + \beta_d * t) \quad . \tag{3.3}$$

The use of a log-link for λ implies an exponential scaling growth of the predictor variables due to applying the exp(). An increase of one unit on the log scale will lead to an increasingly large unit on the un-transformed scale. Graphically this can be thought of a widening of intervals on the horizontal lines, which is visualized in Figure 3.5 below. The main takeaway is that the coefficient of beta is plotted as an exponential function on a natural scale (cf. Richard McElreath (2020), p. 318).

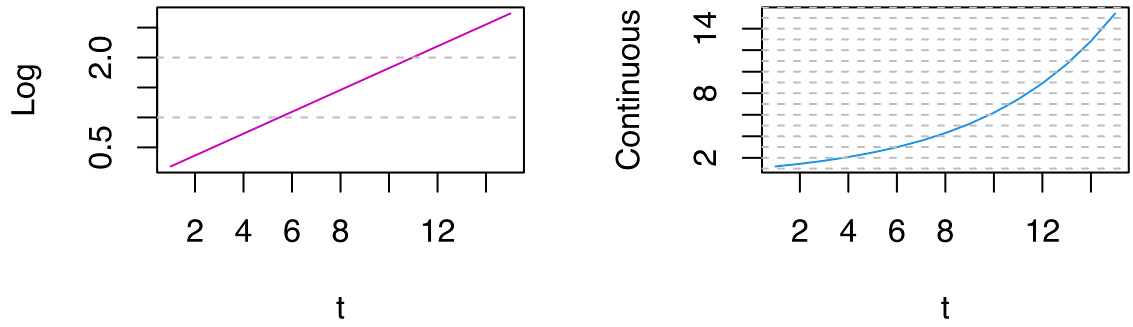

Figure 3.5: *The plotted difference between a log and a continuous scale. By using a log link function an exponential increase on a continuous scale is produced.*

3.2.4 The parameters

The model at its core is a linear regression model with an intercept α and slope β. This is extended by the disaster index d. This subscript creates an individual parameter for each disaster group, which turns λ, α and β into a disaster-specific vector. The unfolded model looks like this:

$$\begin{bmatrix} \lambda_{t,d_1} \\ \lambda_{t,d_2} \\ \lambda_{t,d_3} \end{bmatrix} = exp(\begin{bmatrix} \alpha_{d_1} \\ \alpha_{d_2} \\ \alpha_{d_3} \end{bmatrix} + \begin{bmatrix} \beta_{d_1} \\ \beta_{d_2} \\ \beta_{d_3} \end{bmatrix} * t).$$

3.2.5 Pooling: a model with memory

Before pooling, the model would treat each natural disaster independently from one-another. The frequency with which one disaster occurs would have nothing to do with the frequency another disaster occurs with. However, as can be seen in Figure 3.3 there is reason to believe that intercepts and slopes across disasters are not independent from one-another. This fact should be incorporated into the model, in order to enhance estimation of parameters. One way of integrating communication among parameters is to add pools. There will be two pools in this model: varying intercepts and varying slopes. Let α_d and β_d be the population of varying effects with

$$\alpha_d \sim \text{MVNormal} \ [0 \ , \mathbf{S}_\alpha],$$

and

$$\beta_d \sim \text{MVNormal} \ [0 \ , \mathbf{S}_\beta],$$

with a prior distribution (MVNormal), defined by a three-dimensional Gaussian distribution and a covariance matrix **S**. There are no means in the priors because they are already placed in the average treatment effects α in the linear model.

3.2.6 Covariance and correlation matrices

The parameters of this model are connected to one another by a covariance matrix **S**. It tells the model to incorporate information how all three pools move together. Let **S** be a covariance matrix of the form

$$\mathbf{S}_\alpha = \begin{pmatrix} \sigma_{\alpha,d} & 0 & 0 \\ 0 & \sigma_{\alpha,d} & 0 \\ 0 & 0 & \sigma_{\alpha,d} \end{pmatrix} \mathbf{R}_\alpha \begin{pmatrix} \sigma_{\alpha,d} & 0 & 0 \\ 0 & \sigma_{\alpha,d} & 0 \\ 0 & 0 & \sigma_{\alpha,d} \end{pmatrix}, \mathbf{S}_\beta = \begin{pmatrix} \sigma_{\beta,d} & 0 & 0 \\ 0 & \sigma_{\beta,d} & 0 \\ 0 & 0 & \sigma_{\beta,d} \end{pmatrix} \mathbf{R}_\beta \begin{pmatrix} \sigma_{\beta,d} & 0 & 0 \\ 0 & \sigma_{\beta,d} & 0 \\ 0 & 0 & \sigma_{\beta,d} \end{pmatrix},$$

where σ_d are separate standard deviations for each disaster. Let **R** be a correlation matrix of the form

$$\mathbf{R}_\alpha = \begin{pmatrix} 1 & 1 & \rho_\alpha \\ 1 & \rho_\alpha & 1 \\ \rho_\alpha & 1 & 1 \end{pmatrix}, \mathbf{R}_\beta = \begin{pmatrix} 1 & 1 & \rho_\beta \\ 1 & \rho_\beta & 1 \\ \rho_\beta & 1 & 1 \end{pmatrix}.$$

3.2.7 Hierarchical priors

Lastly, the model will make use of a multilevel-strategy. Leveling refers to the idea of not immediately defining fixed values for parameters but making the process of finding good priors part of the model. Priors for σ and ρ need to be fixed. To start with the correlation matrix **R**, a common distribution used to describe correlation-matrices is the LKJcorr() distribution that attributes probabilities to correlations between -1/+1. But how to choose the right LKJcorr()

prior? First, by plotting the priors. This will give a better understanding of what different priors mean. Then, in combination of previous knowledge about the topic and knowledge about what the prior implies, an educated guess will lead to a prior.

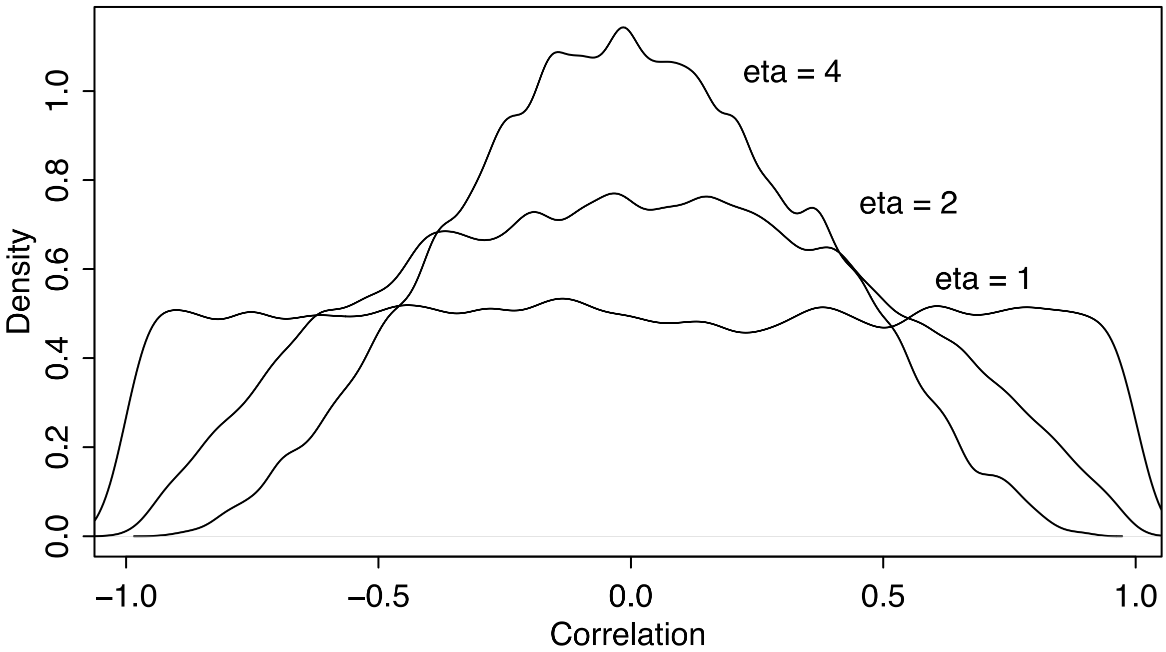

Figure 3.6: *LKJ(corr) probability density. For each eta = 1 to 3 there are three example distributions shown. Eta = 1 would indicate that all correlations are equally likely, eta = 2 would assume that priors towards more extreme values are less likely whilst eta = 4 suggests that correlations approximately from -0.5 to 0.5 are more likely.*

In Figure 3.6 three possible priors for η (eta) are displayed, giving an intuitive understanding how different values for η impact estimates for ρ. Now, by again looking at Figure 3.3 I assume the prior to be around the value of 1 because at first glance especially climatological and hydrological are moving closely together. Lastly, the hyper-priors for σ have to be fixed. For these hyper-priors I will use conventional little informative priors, which are exponentially distributed with a rate of 1. This ensures that all parameters are restricted to positive values.

3.2.8 Inside Stan

For finding the models posterior distribution the stochastic process MCMC, Markov Chain Monte Carlo will be used. In comparison to quadratic approximations the posterior distribution does not necessarily have to be normally-distributed. Instead, with MCMC it is possible to directly sample from the posterior. Beyond this advantage, it was necessary to use MCMC for producing the wanted posterior distributions. Quadratic approximations or equally commonly used algorithms would not have been able to calculate a co-variate indexed multilevel model with equally little effort (cf. Richard McElreath (2020), p 263). The model is written with STAN (Stan Development Team (2023)).

Reviewing Marcov-Chain-Monte Carlo: Chain diagnostics

For checking how well MCMC is able to explore the posterior distribution the chains used to do so should be analyzed. There are two methods to do so, which should be used complementary. The first visualization is called a trace plot, which can be seen in the first graphic in the appendix.

Trace Plot

In the trace plot some z*ig-zagging* of something that looks like white noise is displayed. This plot however, is not randomly created. It is the path every chain takes through exploring each dimension of the parameter space. Each plotted chain additionally is divided into a gray and a white area. In the gray area Markov chain is adapting to the samples. It learns where to look for the posterior in the white area, in order to more efficiently sample from the posterior distribution. As an output, that can be used for the actual sampling, only the samples to the right of the gray area are used.

Three criteria must be met for the Markov chain to be regarded as a healthy chain. Healthy means the chain is able to thoroughly explore the posterior distributions. The three criteria are:

1. Stationarity

A process is weakly stationary if $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$. Additionally, usually stationary processes are mean reverting. Taking a glance at the trace plots, these requirements seem to be met, since positive and negative amplitudes do not change over time. Neither do chains systematically wander above or below their mean. The plots that stick to the bottom of the plot are different because they are squared. They are healthy if they do not show irregular lasting highs or stay beneath the average, which does not seem to be the case.

2. Good mixing

Good mixing means that from the start on the full range of the posterior is explored. Graphically this can be seen in a quick increase in the *zig-zags*, instead of slowly building up. This also is the case.

3. Convergence

What is meant with convergence is that multiple, independent chains stick around the same region of probability. Graphically this means that independent chains stay around the same mean within a similar range in their amplitude. In the trace plot there is no chain that indicates this behavior in an alarming manner.

Trank plot

Trace plots are are a good way to quickly gain insight in how well the Marcov chain performed. However, if several chains are plotted simultaneously, as it is the case for the trace plots in the appendix, the plot can look confusing and hide pathologies within the chains. For this reason, a complementary way to plot how the chains explored the posterior, is of interest. This plot is called a trace rank plot or trank plot.

This method plots the distribution of ranked samples. This means that all samples for each parameter are divided and then ranked from 1 to the number of parameters N. From this a histogram of ranks for each chain is drawn. The Markov chain is healthy if all histograms overlap and stay within the same range. If it does so, the chains were exploring the same space efficiently. Looking at the trank plot - see appendix - this is the case. With the assuring of good working Markov-chains its results can be interpreted.

Chapter 4

Results

The focus of this section will be to convey what the model found. This means that this part will not be mathematical demanding but will rather focus on the interpretation of plots and resulting numbers. The results are presented in three steps. First, the findings will be shown numerically in order to see what the model calculated. This is presented by featuring the mean of catastrophe occurrences for all intervals. The second part will *zoom in* and look at specific means from the table in the first sections by plotting the whole distribution. Thirdly, the most important graphics are displayed, which are how the mean of environmental catastrophes changed over decades and how they will most likely develop. This will give visual insight in the exponential increase of catastrophes.

4.1 The catastrophe put in numbers

Absolute numbers of occurrences

The analysis will start by giving the models hard facts. Only in combination of knowing the numbers that create the resulting graphics, as well as seeing the plots in a scientific context, can give a thorough insight in what the charts mean. The numbers in Table 4.1 can be read as the average expected amount of environmental catastrophes within a 5-year interval.

Table 4.1: *The model mean for environmental catastrophes for previous records with a forecast until 2050.*

	Catastrophe Groups		
Interval	Climatological	Hydrological	Meteorological
1950-1955	6.52	25.72	32.45
1955-1960	7.70	30.29	37.57
1960-1965	9.31	36.52	43.87
1965-1970	10.98	43.02	50.85
1970-1975	13.14	51.10	59.06
1975-1980	15.52	60.79	68.56
1980-1985	18.40	72.61	79.93
1985-1990	22.00	85.77	92.99
1990-1995	26.39	101.98	107.75
1995-2000	31.25	121.03	125.13
2000-2005	37.48	143.81	145.83
2005-2010	44.52	170.74	169.13
2010-2015	53.04	202.41	197.59
2015-2020	63.66	240.54	229.18
2020-2025	75.79	286.57	267.19
2025-2030	90.08	340.67	311.26
2030-2035	107.90	404.21	362.11
2035-2040	128.64	481.23	420.37
2040-2045	153.39	572.12	488.83
2045-2050	183.86	680.23	569.00

It is beneficial to know how the numbers above were calculated. The procedure is that after estimating the models parameters, several thousand samples per catastrophe are drawn. These samples yield the posterior-distribution. Drawing the mean for every distribution generates Table 4.1. One possible interpretation for meteorological disasters for the last interval from 2045-2050 is the following.

> In 22 years from now for the interval 2045-2050 the most likely number of meteorological disasters occurring are about 584 incidents, based on all model assumptions. This makes an average of approximately 116 storms and extreme temperatures most likely emerging per year in America for the given interval. This results in an average of 10 disasters per month, which approximately equivalents to a catastrophe happening every third day.

One more insight emerges from the sum of environmental catastrophes for the last interval.

In 22 years for the interval 2045-2050 the most likely sum of climatological, hydrological and meteorological disasters occurring are about 1589 incidents, based on all model assumptions. This makes an average of approximately 318 disasters from all of the three categories most likely occurring per year in America for the given interval. Equivalently, that are 26 disasters per month or 0.9 disasters per day. In 22 years from now, on average almost every day an environmental catastrophe is to be expected.

4.2 Density distribution with probability intervals

Table 4.1 gives a first insight into what the model says. However, it also hides a lot of information. Each number only represents the mean of a distribution, which means that there is a whole distribution behind each number. To see what this means, in Figure 4.1 the full distribution for meteorological disasters is plotted for the 5-year interval from 2015 until 2020.

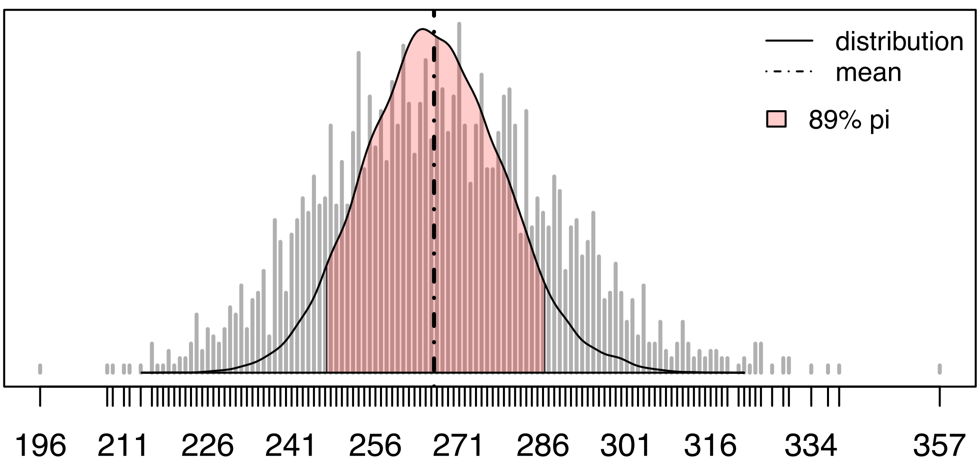

Figure 4.1: *The full density distribution for meteorological disasters for the interval from 2015 until 2020. The mean is plotted as a dashed line and the 89% interval is depicted as a shaded reddish area.*

In Figure 4.1 the distribution for the 2015-2020 interval for meteorological catastrophes is plotted as an example. The generated samples are pictured as a histogram and its density is drawn as a straight line. The mean of the distribution is plotted as a dashed line.

An important aspect of using Bayesian statistics is that it is possible to interpret Figure 4.1 in terms of probability (instead of frequency). For this reason, the reddish area in Figure 4.1 can be called probability interval[1]. One possible interpretation is the following.

> For the interval 2015-2020 with a probability of 89% there were approximately between 245-280 meteorological disasters, based on all models' assumptions. This are between 49 and 56 disasters per year on average.

To give a second example, the full density distribution for hydrological disasters for a forecast interval from 2045-2050 will be plotted.

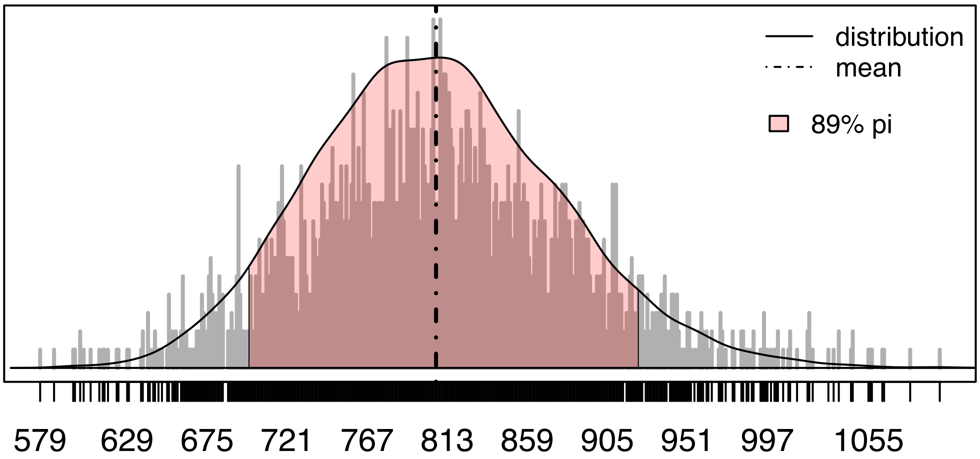

Figure 4.2: *The full density distribution for hydrological disasters for the interval from 2045 until 2050. The mean is plotted as a dashed line and the 89% interval is depicted as a shaded reddish area.*

What is of especial interest in the second figure, compared to the first distribution is that the 89% probability interval has a larger variance, put into numbers. This results mechanically from dealing with larger numbers, the model is still of 89% certainty that the reddish area will correctly predict this interval. One possible interpretation for hydrological disasters is the following.

[1] I chose a 89% probability interval to raise awareness that 95% intervals are mere convention and have no statistical value in themselves, following McElreath.

For the interval 2045-2050 with a probability of 89% there will be approximately between 700-915 natural disasters, considering all assumptions. This are between 140 and 183 hydrological disasters per year on average.

4.3 Change in mean

Thirdly, the development of means from Table 4.1 will be plotted. It was already seeable that from interval to interval there might be accelerated increases. This section will combine both previous parts and will give insight in what manner the means of environmental disasters change with the course of time.

4.3.1 Climatological disasters

The first of the three catastrophes are climatological disasters. They consist of droughts and wildfires in this data-set. How did their occurrence change over time and where does the current trend lead towards?

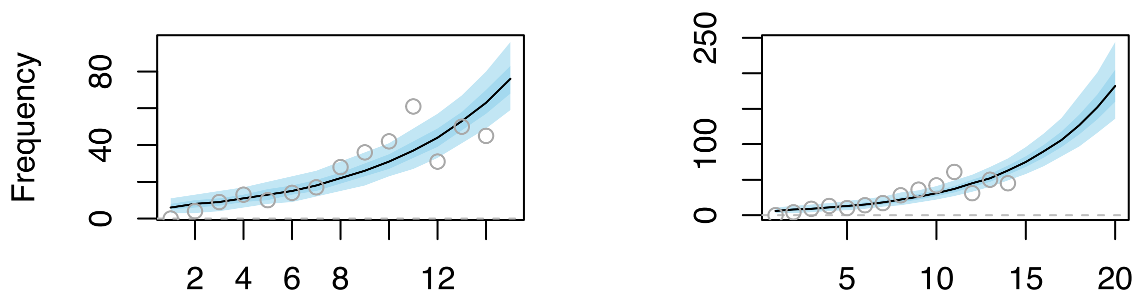

Figure 4.3: *The development of how the average amount of climatological disasters changed over time. Left for the recorded intervals and right with a 20-year forecast of the most likely future within the model.*

For both intervals close to the mean there is a dark-blue shaded area, which is the 89% probability interval, as well as the lighter-blue area, which is the 50% probability interval. Within each shaded area it is 89%/50% likely that values will be captured inside its PI. Moreover, for both graphics several dots can be seen, which are the original data-points.

Focusing on the left graph within the data from 1950-1975 (also highlighted in Figure 3.2) there was little absolute increase. From then on, the distance between each point grows. At the right plot the first 15 intervals, as well as the forecast until 2050 is drawn. The graph suggests that there will be a strong acceleration in the occurrences of environmental catastrophes. Where does this pattern emerge from?

To answer this question, it is useful to recall Table 4.1, since the plot is a combination of the points from this table with two probability intervals. In the plots the absolute difference between each value increases with an accelerating manner. The model suggests the following development, put in numbers:

Table 4.2: *The mean for climatological catastrophes for the first 3 intervals with absolute and relative increases compared to its previous value.*

Interval	Climatological	Abs. Diff	Rel. Diff.
1950-1955	6.52	NA	NA
1955-1960	7.70	1.18	1.18
1960-1965	9.31	1.61	1.21

Table 4.3: *The mean for climatological catastrophes for the last 3 intervals with absolute and relative increases compared to its previous value.*

	Interval	Climatological	Abs. Diff	Rel. Diff.
18	2035-2040	128.64	20.74	1.19
19	2040-2045	153.39	24.75	1.19
20	2045-2050	183.86	30.47	1.20

Now it is easy to see what the model assumes, an interval increase in climatological disasters with a rate ranging from 18-21%. Intuitively this does not sound like much, especially if the starting values start at approximately 4 catastrophes/year but put in long-term perspective this has devastating consequences as can be seen in Figure 4.3. One possible interpretation of the results is:

> Climatological disasters (droughts + wildfires) increase exponentially by a rate of approximately +20% per interval or equivalently +4% annually. Each year more catastrophes will occur in an ongoing accelerating manner, based on all model assumptions and data-input. To put it colloquial once: *it's bad and it will most likely get much worse.*

4.3.2 Hydrological disasters

The second catastrophe type are hydrological disasters. In this data-set they consist of floods and landslides. How did the mean of hydrological disasters change over time and in what way will it develop? The following plot can answer this question.

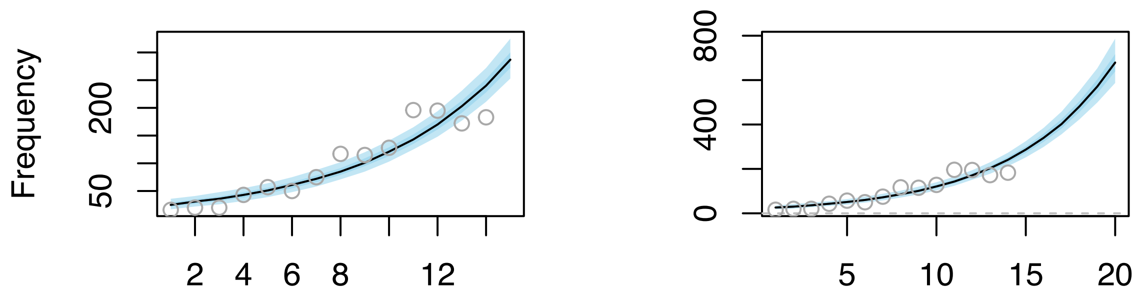

Figure 4.4: *The development of how the average amount of hydrological disasters changed over time. Left for the recorded intervals and right with a 20-year forecast of the most likely future within the model.*

The graphic looks similar to the previous one. For hydrolocial catastrophes the model also suggests an exponential increase per interval. However, a closer look at the y-axis of the graphs indicates that in absolute numbers the frequency of occurrence and its development is not the same as with climatological catastrophes, in this case climbing up to 800 disasters. As previously, the chart put into numbers.

Table 4.4: *The mean for hydrological catastrophes for the first 3 intervals with absolute and relative increases compared to its previous value.*

Interval	Hydrological	Abs. Diff	Rel. Diff.
1950-1955	25.72	NA	NA
1955-1960	30.29	4.57	1.18
1960-1965	36.52	6.23	1.21

Table 4.5: *The mean for hydrological catastrophes for the last 3 intervals with absolute and relative increases compared to its previous value.*

	Interval	Hydrological	Abs. Diff	Rel. Diff.
18	2035-2040	481.23	77.02	1.19
19	2040-2045	572.12	90.89	1.19
20	2045-2050	680.23	108.11	1.19

Again, the first and last three entries are shown. For hydrological disasters the models expected percental increase per year ranges from approximately 18-21%. One possible interpretation of the results is the following.

> Hydrological disasters (floods + landslides) increase exponentially by a rate of approximately +20% per interval or equivalently +4% annually. Each year more catastrophes will occur in an ongoing accelerating manner, based on all model assumptions and data-input.

4.3.3 Meteorological disasters

The third and last catastrophe type that is looked at closer are meteorological disasters. In this data-set this category consists of extreme temperatures and storms. This is how their occurrence changed over time and how it will most likely develop:

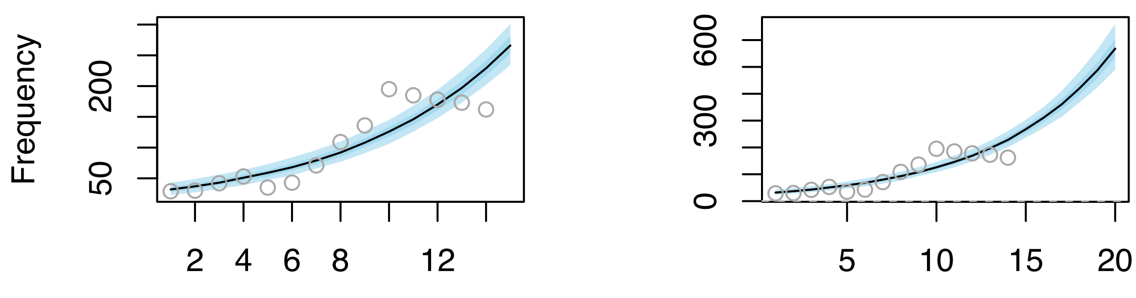

Figure 4.5: *The development of how the average amount of meteorological disasters changed over time. Left for the recorded intervals and right with a 20-year forecast of the most likely future within the model.*

Meteorological disasters change drastically over the course of time. In Figure 4.5 the original data points not always grow d'accord with the model. Nevertheless, an average for the whole time-span the models prediction of our most likely futures is an exponential increase in meteorological disasters. The numbers calculated by the model are the following.

Table 4.6: *The mean for meteorological catastrophes for the first 3 intervals with absolute and relative increases compared to its previous value.*

Interval	Meteorological	Abs. Diff	Rel. Diff.
1950-1955	32.45	NA	NA
1955-1960	37.57	5.12	1.16
1960-1965	43.87	6.30	1.17

Table 4.7: *The mean for meteorological catastrophes for the last 3 intervals with absolute and relative increases compared to its previous value.*

	Interval	Meteorological	Abs. Diff	Rel. Diff.
18	2035-2040	420.37	58.26	1.16
19	2040-2045	488.83	68.46	1.16
20	2045-2050	569.00	80.17	1.16

The model predicts an increase ranging from 16-17% per interval. Looking at the relative change, this is an similar change in comparison to previous disasters. Nonetheless, there are up to 569 expected disasters for the last interval, which is much higher in comparison to climatological catastrophes. This is owed to the fact that meteorological disasters start at a higher level to begin with. For the interval 1950-1955 there were on average approximately 3.4 climatological and 28 meteorological disasters. Due to this fact, the exponential growth in climatological disasters takes more time to "take off", whilst the exponential growth is evident much quicker for meteorological catastrophes. One possible interpretation for meteorological disasters is the following.

> Meteorological disasters (extreme temperatures + storms) increase exponentially by a rate of approximately +16% per interval or equivalently +3.2% annually. Each year more catastrophes will occur in an ongoing accelerating manner, based on all model assumptions and data-input.

Hydrological disasters, climatological disasters and meteorological disasters indicate an exponential increase in their occurrence. According to the model, this effect will be increasingly evident within the next 20 years. It is impossible to predict the future, but if this model is accurate or heads towards the right direction it gives a solid reason to worry about our future.

Chapter 5

Conclusion

This research aims to mathematically describe the change of frequencies in which environmental catastrophes occurred from 1950-2020 and how their occurrences will most likely develop in the intermediate future until 2050. By analyzing a data-set from a project governed by the NASA, this thesis has shown that three environmental catastrophe groups indicate an exponential pattern in their growth rate. This pattern will continue and result in an extremely high frequency of environmental catastrophes.

The first catastrophe type analyzed was climatological disasters. This disaster group consists of droughts and wildfires that happened between 1950-2020 on the continent of America. The stochastic model indicates that in approximately 20 years from now, climatological disasters will 40-fold in their occurrence, in comparison to when recordings started to be reliable around 1955. The second catastrophe was hydrological disasters, which consist of floods and landslides. For this type of disasters, their annual occurrence in comparison to the 60's climbed from around 6 to 40 catastrophes in 2020, a number which will presumably climb to 150 per year by 2045. In comparison to both other disasters, climatological disasters indicate the highest accelerating increase, which on average is as high as +4% per year. Lastly, meteorological catastrophes, extreme temperatures and storms, were analyzed. Along with hydrological disasters, this group will turn out to be a very common catastrophe. Within 20 years, in around 2045, on average 8 meteorological disasters per month will occur. All catastrophes added up, it is possible to say, that in 20 years from now, on average, nearly every day an environmental disaster is to be expected in America.

All in all, extreme weathers are going to be a more common part in everyday life. To put this into perspective, someone born around 2020 is possibly going to be affected by an environmental catastrophe every day by the time this person turns 20. The results are in accordance with a long chain of environmental research. They clearly give reason for concern and underpin arguments to the effect that immediate global intervention is necessary to keep mother earth habitable for future generations. Without coordinated global intervention, it should not thunderstruck us in the future that living on planet earth is going to be accompanied by a striking number of droughts, extreme temperatures, floods, landslides, storms and wildfires.

Bibliography

Gertsch, C., & Krogerus, M. (2023). *Dieser gletscher bestimmt die zukunft der menschen*. SZ. https://www.sueddeutsche.de/projekte/artikel/wissen/thwaites-gletscher-antarktis-glaziologie-meeresspiegelanstieg-klimawandel-e317022/?utm_content=antarktis_e317022&utm_medium=organic_content&deeplink=false&utm_source=googlewebstories&utm_campaign=op_social?ref=ampstories&slide_id=1

Grossi, P., Kunreuther, H., Mahdyiar, M., & Porter, B. (2005). *Catastrophe modeling - a new approach to managing risk*. Springer Science.

Hallegatte, S. (2014). *Natural disasters and climate change - an economic perspective*. Springer.

Nasa ESDIS. (2023). *EOSDIS*. https://search.earthdata.nasa.gov/search

OECD. (2005). *Policy risk in insurance - catastrophic risk and insurance*.

Richard McElreath. (2020). *StatisticalRethinking* (Second edition). Taylor & Francis Group.

Stan Development Team. (2023). *STAN*. https://mc-stan.org

Taherkhani, M., Vitousek, S., Barnard L., P., Frazer, N., Anderson R., T., & Fletcher H., C. (2020). *Sea-level rise exponentially increases coastal flood frequency*. https://rdcu.be/c6yuj

The Economist. (2023). *How to predict record-shattering weather events*. https://www.economist.com/science-and-technology/2023/02/08/extreme-weather-events-are-getting-more-frequent?giftId=296d5cf6-d1b7-45b7-9504-1930645c2421

Appendix A

Traceplot

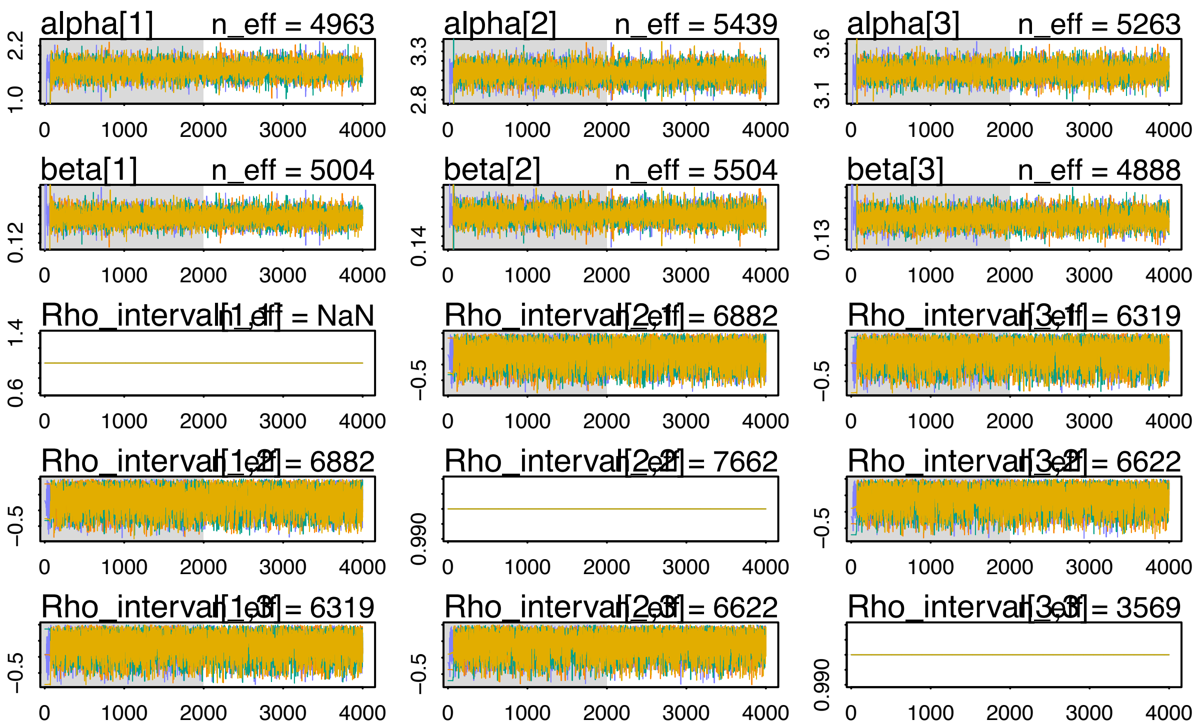

Figure A.1: *Traceplots show the path each chain takes to explore the posterior. In this graph several chains are plotted simultaneously.*

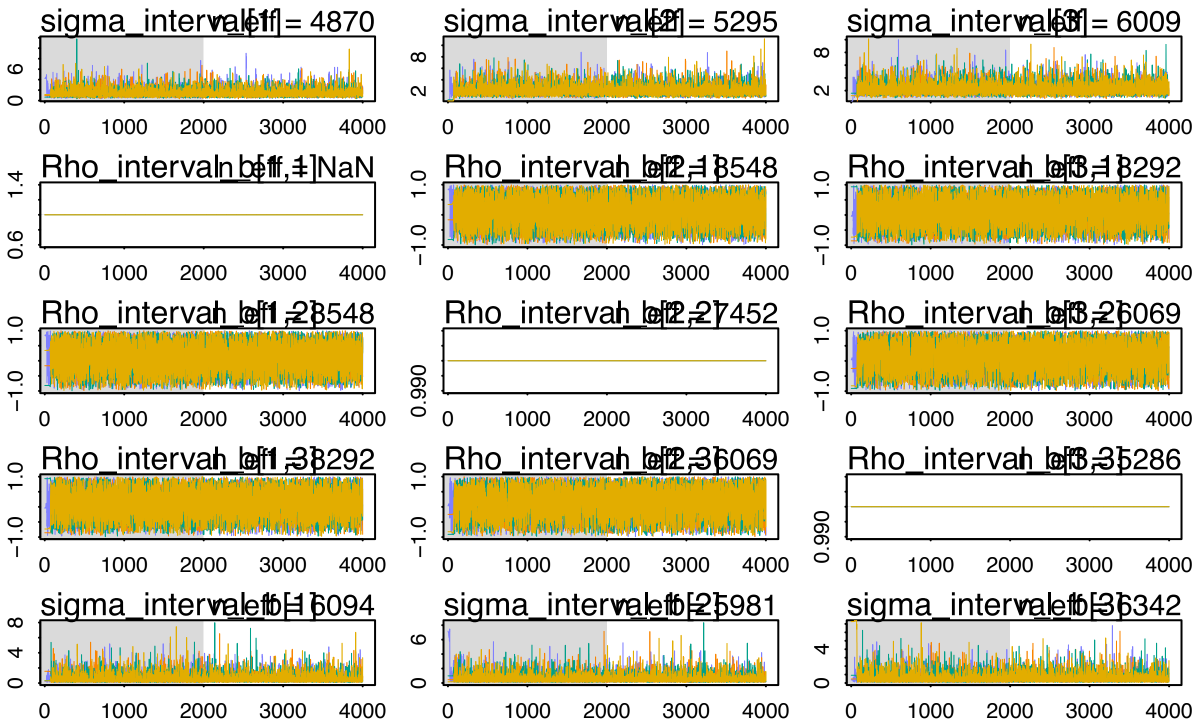

Figure A.2: *Traceplots show the path each chain takes to explore the posterior. In this graph several chains are plotted simultaneously.*

Appendix B

Trankplot

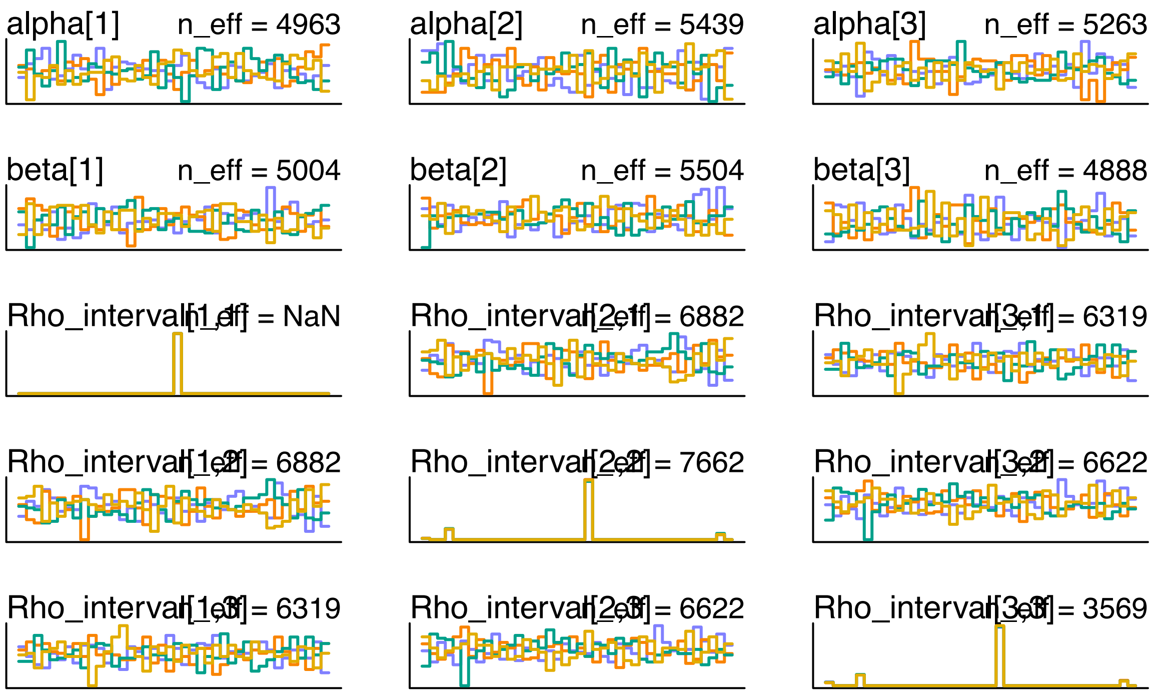

Figure B.1: *Trankplots are visualized distributions of ranked samples. This trankplot indicates can be regarded healthy, since the histograms overlap and develop quickly.*

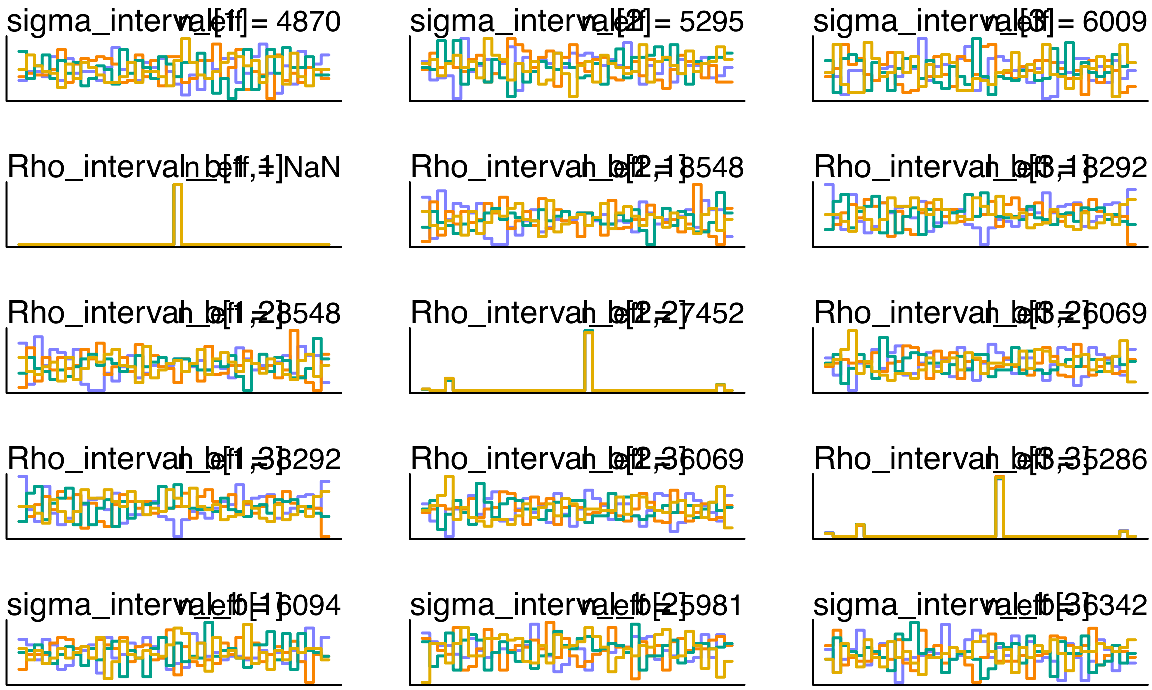

Figure B.2: *Trankplots are visualized distributions of ranked samples. This trankplot indicates can be regarded healthy, since the histograms overlap and develop quickly.*